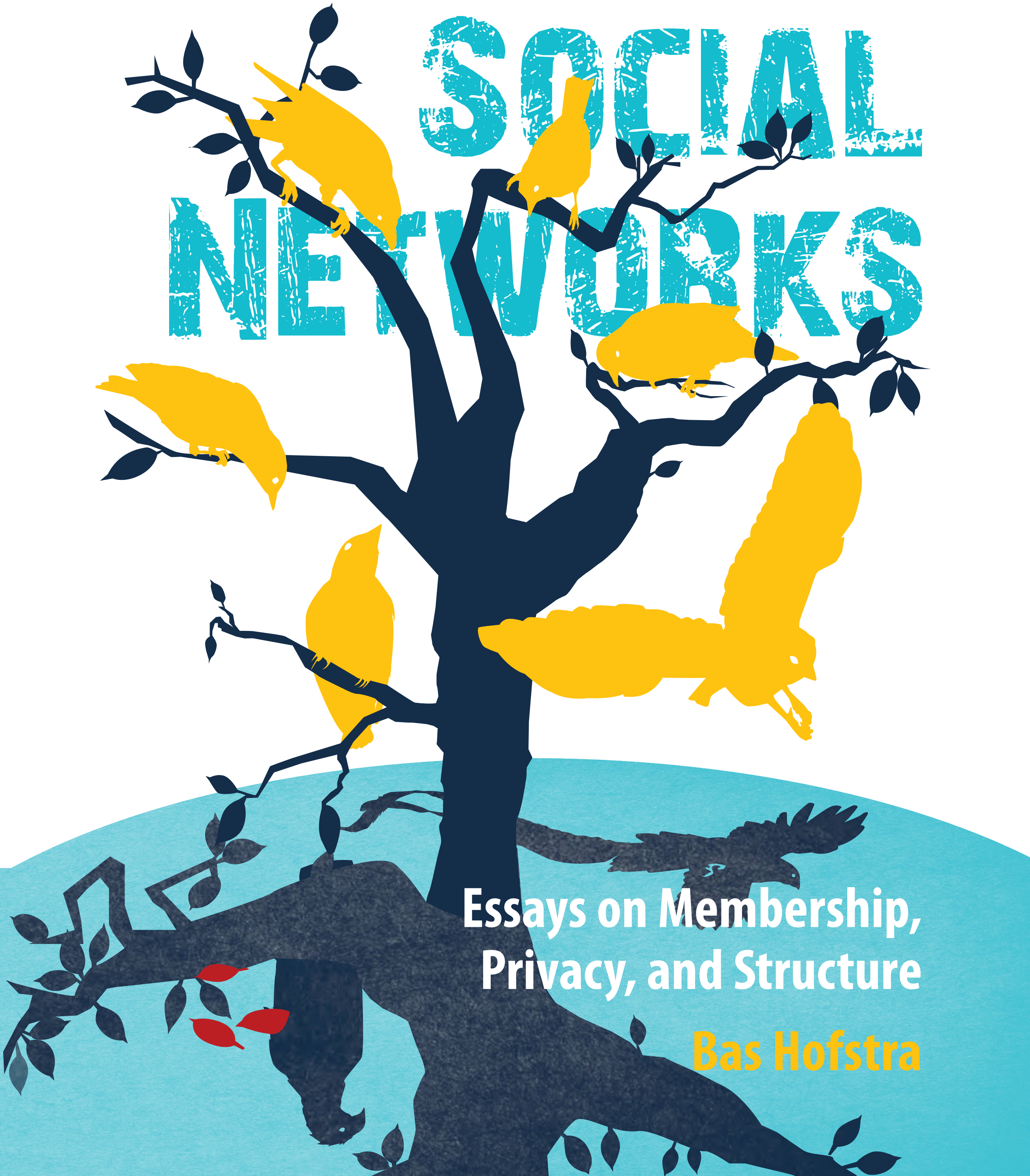
Online Social Networks



Essays on Membership, Privacy, and Structure

Bas Hofstra

Online Social Networks

Essays on Membership, Privacy, and Structure

Bas Hofstra

Online Social Networks:
Essays on Membership, Privacy, and Structure
— Bas Hofstra

Cover illustration: isontwerp.nl - Eindhoven
Cover photo: Wouter le Duc
Printing: Ridderprint BV
ISBN: 978-90-393-6808-4

This book was typeset using LaTeX by Bas Hofstra.

Online Social Networks

Essays on Membership, Privacy, and Structure

Online Sociale Netwerken

Essays over Lidmaatschap, Privacy, en Structuur

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof. dr. G.J. van der Zwaan, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op vrijdag 15 december 2017 des middags te 2.30 uur

door

Bas Hofstra

geboren op 22 oktober 1987
te Noordoostpolder

Promotor: Prof. dr. F.A. van Tubergen
Copromotor: Dr. R. Corten

This study was funded by the Netherlands Organization for Scientific Research (NWO) research talent grant [406-12-004]. This research also benefited from the support of the NORFACE research program on Migration in Europe — Social, Economic, Cultural and Policy Dynamics and from the support from "NWO middelgroot" [480-11-013] and "NWO veranderingsstudies" [481-11-004].

Voor mijn ouders

Contents

List of Tables

List of Figures

Chapter 1

A Systematic Study of Online Social Networks[1]

[1]This chapter benefited from invaluable discussions I had with Rense Corten, Frank van Tubergen, Manja Coopmans, Jesper Rözer, Maaike van der Vleuten, Wouter Quite, and Niek de Schipper.

"[The] technological revolution in mobile, Web, and Internet communications has the potential to revolutionize our understanding of ourselves and how we interact. Merton was right: Social science has still not found its Kepler. But three hundred years after Alexander Pope argued that the proper study of mankind should lie not in the heavens but in ourselves, we have finally found our telescope. Let the revolution begin..."

— Duncan J. Watts (2011: 266)

The main objective of this dissertation is to understand the *structure of online social networks* for new insights into the structure of social networks in general. What are the theoretical and empirical promises and pitfalls of such a study? I answer these questions through a collection of five self-contained, empirical studies. This first chapter outlines the societal and scientific implications of online social networks. It then synthesizes the research aims, findings, and conclusions of the five studies.

1.1 The Impact of Online Social Networks in Society

The extraordinary *rise to prominence of social media* over the last decade is a transition that has had a profound societal impact. Figure 1.1 depicts the widespread adoption of social media in the Netherlands, categorized by age group, over the last five years. It shows that nearly 95% of those aged 12 to 45 used social media at least once in 2016 (Statistics Netherlands, 2017). The prime example of such a social media platform is Facebook, which is by far the largest social network site in the world (Facebook, 2017) — 1.86 billion people use Facebook monthly as of December 2016.[2] In the Netherlands, approximately 10.4 million of those aged 15 and older were using Facebook in January 2017, covering approximately 78% of the Dutch population (Van der Veer et al., 2017).[3]

[2]There were approximately 1.23 billion *daily users* as of December 2016, using the platform about 50 minutes each day.

[3]Approximately 7.5 million *daily users*, suggesting that ∼56.3% of the Dutch use Facebook every day.

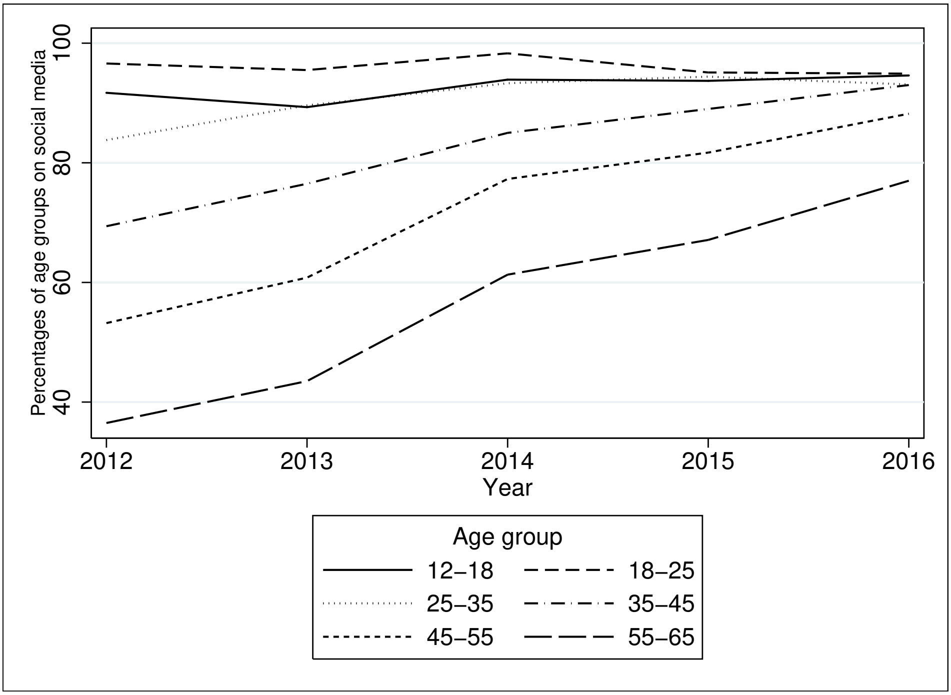


Figure 1.1: Social media use in the Netherlands by age group and year (Statistics Netherlands, 2017).

This spectacular increase in social media's popularity is made possible by the near-saturated levels of Internet penetration in Western societies, the possibilities of maintaining and sharing experiences with social relationships via social media platforms (e.g., Facebook, Instagram), and the widespread adoption of smartphones. I want to sketch three situations to specify how the presence of social media — knowingly or unknowingly — has crept into many aspects of our daily lives.

Membership. Bruce does not have membership on Facebook. One of his acquaintances announced her birthday party exclusively via Facebook. Unaware of this, Bruce missed the party and missed the opportunity to talk to many people he does not see so often. One of Bruce's friends — who is a Facebook member — attended the party and learned about a job vacancy that he is going to apply for. This knowledge would have been very beneficial for Bruce too, as he recently lost his job in a similar sector. Bruce tries to cope with his job loss. He would like to talk about it with his friends, but this is not going very well; he is unaware that many of his friends nowadays communicate via Facebook. It is difficult to reach them otherwise to ask them for help or solicit them for advice.

Privacy. Jane is a Facebook member. A few years ago, Jane attended a music festival and a friend of hers uploaded rather compromising photos shot at this festival to Jane's Facebook profile. At the time, Jane and her friends had a good laugh about it. In the present day, Jane is looking for a job and sends an application to a reputable firm. A hiring manager from this firm often checks the social media profiles of potential employees. Jane long forgot about her compromising photos, but the hiring manager looked at these photos on Jane's Facebook profile and consequently decided to not invite Jane for a job interview.

Structure. Robin is a Facebook member, and her Facebook network is a reflection of her offline contacts — nearly everyone she knows is also her friend on Facebook. Her Facebook network is rather homogeneous in terms of ethnic background. Just like herself, most of her network contacts are members of the ethnic majority group in the country. Therefore, she does not have many ties with people of another ethnicity. The negative feelings she holds towards them are, therefore, hardly challenged by positive personal encounters. She lives in an "echo-chamber:" she is increasingly surrounded by like-minded people, and members among this group reinforce each other in their negative attitudes towards people of other ethnicities via posts on Facebook and during discussions. The negative interethnic attitudes of herself and her friends become increasingly polarized.

These examples illustrate why dynamics involving social media contribute to the equal or unequal distribution of resources among society's members and the absence or presence of trust and social integration among its members. Social inequality can occur through differences between members and non-members in possibilities to mobilize social contacts for support or information (Ellison et al., 2007) and through the negative consequences of differences in privacy management of social subgroups (Roth et al., 2016). Social cohesion can be facilitated through the amount of intergroup contact individuals have in their social networks (online) and its potential consequences for intergroup trust, attitudes, and opinion polarization (Allport, 1954).

1.2 The Impact of Online Social Networks in Scientific Research

Next to its *societal impact*, the advent of social media may also have a major *impact on how we practice (social) science.* Watts' famous quote (2011: 266) — mentioned at the beginning of this synthesis — predicts a major revolution in the social sciences. He argues that the unprecedented adoption of online technologies in the last decade may (or already has) revolutionize(d) the way in which social science is practiced. He is not alone in his intuition. Others have also argued that we are in the middle of such a social science revolution (e.g., Lazer et al., 2009).

The onset of this revolution resulted from the fact that online communication leaves digital time-stamped "traces" of interactions in (often) large social networks (Golder and Macy, 2014). As part of their daily operations, many social media platforms nowadays automatically archive these interactions (Spiro, 2016), and social scientists increasingly seek to gather and analyze these data. This revolution goes alongside the rise and current popularity of "computational social science," where social scientists increasingly use methods (often) borrowed from or in collaboration with the computer sciences to analyze these "digital trace data" to answer substantive social scientific questions. Following Watts' analogy, online platforms can thus be considered our "telescope," with which we can study the many digital traces of behavior left on these platforms, and these traces give unprecedented insight into human behavior on a massive scale.

An obvious choice of where to gather and study such new digital trace data on human behavior is *Facebook*. This is because it is the dominant social media platform worldwide and is predominantly an online *friendship* network (Duggan et al., 2015), as opposed to *professional* online networks such as LinkedIn or *microblogging* websites such as Twitter. Social network sites, including Facebook, can be characterized as web-based communication platforms where individuals construct a uniquely identifiable (semi-)public profile, within which they articulate a list of other users with whom they share a relationship (boyd and Ellison 2007; Ellison and boyd, 2013).[4, 5]

[4] The terms social network(ing) site (or: SNS) and social media (platform) are used interchangeably throughout this dissertation. When I refer to the social networks that are part of these platforms, I refer to online social networks.

[5] The definition of Ellison and boyd (2013) more clearly emphasizes the *communication* and *user-generated content* aspects of social network sites as compared to that of boyd and Ellison (2007). I shortened and combined both definitions for clarity.

Prominent examples of using digital trace data from Facebook include experiments on the social contagion of happiness and voting among hundreds of thousands of test subjects (Bond et al., 2012; Kramer et al., 2014), studies on (interethnic) tie formation in large networks on Facebook (Mayer and Puller, 2008; Wimmer and Lewis, 2010), and studies on social network structure among millions of individuals (Bäckström et al., 2012; Corten, 2012). These studies illustrate that there is one particularly important observation the new "telescope" can make: the social networks that are oftentimes a major aspect of social media profiles. There are two important advantages of considering these online social networks over the social networks occasionally studied in survey research. First, online networks map *networking behavior* instead of self-reports on social contacts, which may be more susceptible to recall biases or other misperceptions. Second, online networks capture potentially *hundreds of social contacts*, as opposed to the small networks often considered in surveys.

Hence, existing hypotheses on social interactions may be tested in a new — or even more suitable — way using these new digital trace data, and existing theoretical views may be challenged by new empirical evidence. In yet other instances, the availability of new data challenges us to advance theory, as observations from the new "telescope" provide options for testing propositions on social interactions that we were unable to test before.

1.3 Linking Offline and Online Network Data

In the study of online social networks, some scholars consider online network data exclusively (e.g., Mayer and Puller, 2008; Wimmer and Lewis, 2010), while other scholars consider social media networks via surveys (e.g., Ellison et al., 2007; Van Zalk et al., 2014). However, online data often contain many observations but lack details about individual characteristics or personal attitudes. Survey data, on the other end, often include many individual-level details of fewer observations but it is often infeasible to gain insight into online social networking behavior among hundreds of friends in these data.

One substantial contribution of this dissertation is that it links offline survey data on Dutch adolescents with online (network) data from Facebook. This approach enables insight into individual characteristics, leisure time activities, attitudes, and close, small personal networks, while simultaneously observing their large online networks. Linking survey data with online (network) data has been recommended before (e.g., Stopczynski et al., 2014; Tufekci, 2014; Spiro, 2016), but, to my knowledge, I am among the first to follow this approach.

Specifically, I make use of several waves of survey data on Dutch adolescents, titled "Children of Immigrants Longitudinal Survey in Four European Countries" (CILS4EU: Kalter et al., 2013) and its follow-up, the "Children of Immigrants Longitudinal Survey in the Netherlands" (CILSNL: Jaspers and Van Tubergen, 2017). Social media's popularity is particularly high among Dutch adolescents (see **Figure 1.1**). In 2015, the vast majority of Dutch adolescents spent time on Facebook every day, with about half spending more than one hour per day on Facebook (own calculations). This makes Dutch adolescents a suitable target population for the study of online social networks.

I link these survey data with the Dutch Facebook Survey (Hofstra et al., 2015a). We — Corten, Van Tubergen, and myself — set up a project where we had coding assistants search for the CILS4EU respondents' Facebook profiles. In the vast majority of cases, we were able to successfully match respondents to Facebook profiles. These profiles contain a rich source of information about respondents' Facebook behavior, including information about themselves and what they like, some of their social interactions and textual status updates, and all of their Facebook friends (~1.1 million in total).[6]

1.4 Aims of this Dissertation

I mentioned how the prevalence of social media in our daily lives can have rather serious societal implications and that digital trace data on social networks can illuminate core puzzles in social network analysis. Notwithstanding, theory-driven empirical sociology that takes full advantage of digital trace data on social networks while simultaneously acknowledging its disadvantages (those of which I outline below) is scarce. In this dissertation, I fill in some of this knowledge gap. To be precise, I advocate the analysis of the structure of online social networks as a novel approach to the study of social network structure in general. To this end, I break down the goals of this dissertation into two overarching research aims.

[6] All of the Facebook data was publicly visible and collected via a strict procedure with password-protected files on department computers. Personal identifiers were removed from the data. The data collection, the coding procedure, and linking it with the survey data for scientific purposes were reviewed and approved by an internal review board for the social and behavioral sciences at Utrecht University (project number: FETC14-019).

The first aim is to *describe and explain the individual differences in* ***activity on social media***. Specifically, I consider activity in the form of *membership in* and *privacy on* social media. An important methodological argument for considering these dimensions is that they specify sample selection biases in online social network data (Lewis, 2015a). Who is on social media and, given membership, whose online networks can we actually observe? These dimensions are thus crucial to consider prior to the study of online social network structure. It stands to reason that these dimensions are a key substantive topic as well. The growing literature on the consequences of social network site usage (e.g., Brooks et al., 2014; Ellison et al., 2014; Hobb et al., 2016) often neglects one group, the non-members. Hence, we do not know which groups do or do not reap the potential benefits (or hazards) of membership on a social network site. Bruce, from the example at the beginning of this synthesis, suffered a lack of both maintenance of social capital (see Ellison et al., 2011) and social support (see Van Ingen et al., 2017) because of not being a Facebook member. Additionally, inherent to the rise of social media is that personal content is easily accessible to a large audience. A study of the causes behind privacy choices can identify those individuals who are less able to manage their privacy and, thus, are more susceptible to identify fraud (Acquisti and Gross, 2009; Javaro and Jasinski, 2014; Wu et al., 2014), unwanted exposure to third parties, and loss of reputation or (job) opportunities (Lewis et al., 2008a). Remember that Jane was not invited to a job interview due to her pictures on Facebook; 75% of job recruiters track potential employees' social media profiles (Roth et al., 2016).

The second aim is to *describe and explain individual differences in the* ***structure of online social networks***. The dimensions of online social network structure I study are segregation — which relates to *whom* one is tied — and its size — which specifies to *how many* one is tied. Both dimensions are related to a myriad of sociologically relevant issues. For instance, a classic argument is that diversity among weak ties — such as found on Facebook — provides novel information on job openings and is linked to labor-market outcomes (Granovetter, 1973, 1983; Lin, 1999). Literature further suggests that even superficial contact between members of different groups have the potential to reduce intergroup prejudice (Allport, 1954; Pettigrew and Tropp, 2006), something that Robin — as exemplified at the beginning of this synthesis — did not experience, due to her highly homogeneous network. Additionally, social network size is associated with health and well-being, receiving social support, and mortality risks (Wellman and Wortley, 1990; Shye et al., 1995; Smith and Christakis, 2008; Holt-Lunstad et al., 2010; Holt-Lunstad et al., 2015). Addressing this second aim requires that I — in some cases — develop new methods that assist in the study of online social network structure, which is what I indeed aim to do.

Part I: Activity on Social Media

1.5 Who Was First on Facebook?

In **Chapter 2**, I first identify a set of factors that promote membership in a social network site in general.[7] Subsequently, I study what determines early adoption of Facebook specifically. Prior research into social media membership shows differences by ethnicity and race. Asian Americans use Twitter (a *microblogging* website) less often than other ethnic and racial groups (Hargittai and Litt, 2011). Furthermore, women are more likely than men to be members of social network sites (Hargittai, 2008; Thelwall, 2008; Moule et al., 2013). Finally, membership intention in Facebook seems to be driven by others' opinions about Facebook (Cheung et al., 2011). As of yet, however, extant literature does not explain why individuals choose *specific* social network sites in contexts where there is more than one alternative.

1.5.1 Contributions

I extend these (remarkably few) studies in two ways. First, an important characteristic of social media is that their popularity is highly time-dependent. I study the adoption of Facebook in 2010 in comparison with — at that time — a far more popular Dutch platform (i.e., Hyves, which ended as a social media platform in 2013). This study context allows me to gain innovative knowledge about a process that is typically highly dynamic. Namely, I examine the causes of early adoption

[7] The empirical chapters in this dissertation are written as self-contained, standalone essays — published or (to be) submitted to scientific journals. This implies that there is overlap between the chapters. Cross-references between chapters are indicated as references to the published papers. This dissertation is cumulative and writing this dissertation over a period of four years increased my knowledge of the topics that I cover. However, this may mean that what I practice in one chapter may slightly diverge from what I practice in a later chapter.

that contributed to the rise of one of the most prominent communication innovations in the last decade (Facebook) during a unique historical context in which there was a similar but much more popular innovation (Hyves) available.

Second, I am among the first to consider whether and to what extent theories on *social contagion* affect membership on social network sites, as it has been suggested to affect social media uptake (Hargittai, 2008; Hargittai and Litt, 2011). Specifically, I consider peer influence (Brechwald and Prinstein, 2011), the phenomenon where individuals in groups increasingly resemble one another in terms of behavior over time. Peer influence has been considered across a myriad of behaviors (e.g., health behavior, school behavior: Centola et al., 2010; Geven et al., 2013), but not for social media adoption. There are, however, three reasons why such a peer influence in membership might exist. First, becoming a member is more attractive when more of a person's friends are already members (Liebowitz and Margolis, 1994). Second, individuals might imitate their friends in social network site membership (Marsden and Friedkin, 1993). Third, there might be norms within groups that push conformity in membership (Cialdini and Goldstein, 2004).

1.5.2 What Causes Membership and Early Adoption of Facebook?

In 2010, approximately 84% of Dutch adolescents were members of either Facebook or Hyves. Broken down by platform, I find that approximately 35% were members of both Facebook *and* Hyves, 61% were exclusively Hyves members, and 4% were exclusively Facebook members.

What caused individuals to be among the 84% of social network site members in 2010? I hypothesize and corroborate that adolescents who are more socially active are more likely to be members of social network sites. These adolescents engage in more leisure time activities and presumably find an outlet in social network sites to share the experiences of their busy lives. I hypothesize and confirm that exposure to and ownership of digital resources (e.g., smartphones) is positively associated with social network site membership, as digital resources provide opportunities to register with and be exposed to the platforms. These findings are consistent with the *diffusion of innovations* framework, which states that that specific lifestyles and exposure to technology promote technology adoption (Rogers, 2003). Adjusting for a number of factors, I also find that girls and Dutch ethnic majority members are more often members of a social network site than their counterparts.

What promotes early adoption of Facebook? I studied this question in terms of whether respondents were a member of Facebook *exclusively* or whether they had adopted Facebook in *addition* to Hyves. For members of ethnic minorities, Facebook had an important advantage over Hyves: Facebook is an international platform, whereas Hyves was Dutch. Many adolescents in Europe whose parents are immigrants have transnational ties (Schimmer and Van Tubergen, 2014). Facebook may have thus provided better possibilities to communicate with friends and relatives abroad for those who are members of ethnic minorities. This may be why members of the ethnic minority adopted Facebook earlier than did Dutch majority members. Additionally, I test theories on peer influence (Brechwald and Prinstein, 2011). When friends join Facebook (or Hyves), the likelihood of using Facebook (or Hyves) increases sharply, possibly because like-minded people flock together (e.g., McPherson et al., 2001), but more likely because of peer influence. Furthermore, I find some evidence for the hypothesis that adolescents adopt Facebook earlier if they have more friends who are member of the ethnic minority. Finally, I find no evidence to support my hypothesis that more-popular adolescents adopted Facebook earlier.

1.6 Who Keeps a Public Facebook Profile?

In **Chapter 3**, I describe and study the reasons behind adolescents' choice of privacy settings on Facebook. I now ask: given membership, what can we observe from these members on Facebook? Prior work into this question shows that women are more likely than men to maintain private social media profiles (Acquisti and Gross, 2006; Lewis et al., 2008a; Shin and Kang, 2016; Thelwall, 2008; boyd and Hargittai, 2010; Hoy and Milne, 2010). Younger respondents more frequently opt for private social media profiles than do older respondents (Tufekci, 2008; Litt, 2013). This prior body of work, however, does not explain *why* women and younger people opt for privacy on social media. Furthermore, those with more friends who keep private Facebook profiles are themselves more likely to maintain private profiles (Lewis et al., 2008a; Lewis, 2011). Finally, those who use Facebook more often have better Internet skills, and those who have more Facebook friends more often keep private Facebook profiles (Lewis et al., 2008a; boyd and Hargittai, 2010; Stutzman and Kramer-Duffield, 2010).

1.6.1 Contributions

I contribute to prior work on Facebook privacy in two ways. First, I aim to develop a theoretical explanation for why earlier work has consistently found that women and younger people more frequently maintain private profiles. I study whether lower levels of *generalized trust* among these groups cause them to more frequently opt for private profiles. That is, do those who place less trust in others generally (Barber, 1983; Paxton, 2007) also opt more often for private profiles on Facebook? Prior work has suggested that there are lower levels of trust among ethnic minorities and among those on lower educational tracks (Mewes, 2014; Simpson et al., 2007). Therefore, I also consider differences in privacy settings by ethnicity and education. I advance theory in privacy research by unraveling some of the mechanisms that may underlie previous findings.

Second, I study privacy *settings* rather than survey individuals about their privacy (e.g., Tufekci, 2008; Fogel and Nehman, 2009; Thomson et al., 2015). Surveying people about their privacy results in underestimation of levels of privacy behavior (Utz and Krämer, 2015) and in acquiescence biases (Kuru and Pasek, 2016). My study of privacy *settings* on Facebook — i.e., linking survey data and online data — circumvents these issues.

1.6.2 What Causes People to Choose Privacy on Facebook?

I studied Facebook privacy in terms of whether respondents' "Timelines" and "Friend lists" are publicly visible or not. Content can be posted on Timelines (e.g., photos, videos, textual status updates). Friend lists show which others one has befriended. In 2014, approximately 55% of adolescents maintained private timelines, whereas approximately 25% kept a private friend list.

What causes these privacy settings to vary from person to person? First, following theories on peer influence (Cialdini and Goldstein, 2004; Brechwald and Prinstein, 2011), I hypothesize and find associations between peers' privacy settings and respondents' Facebook privacy settings. Second, further considering the role of peer influence and social contagion, I hypothesize and confirm that groups in which more adolescents are friends with other adolescents are more likely to imitate their peers' privacy settings, presumably because behavior spreads faster and norms can be more easily monitored and enforced in more-connected groups (Coleman, 1990; Corten and Knecht, 2013). Third, those who are more popular among their peers are more likely to maintain public Facebook profiles, possibly due to a higher need for self-expression, to maintain status, or due to a higher susceptibility to risk behavior (Dijkstra et al., 2009).

This chapter suggests that girls, members of ethnic minorities, adolescents in lower educational tracks, and younger adolescents more frequently opt for private Facebook profiles. These findings are consistent with observations that these groups tend to display lower levels of trust in “most others” (Glaeser et al., 2000; Alesina and La Ferrara, 2002; Simpson et al., 2007; Mewes, 2014) and that girls and younger people more often maintain private social network site profiles (Lewis et al., 2008a; Tufekci, 2008; boyd and Hargittai, 2010). However, I find no support for my hypothesis that self-reported generalized mediates these associations.

Part II: Structure of Online Social Networks

1.7 How Segregated Are Social Networks on Facebook?

In the first part of **Chapter 4**, I study under what conditions ethnic and gender segregation occurs among weak ties as measured on Facebook. Such weak ties can be defined as social relationships that do not involve much time, emotional intensity, or intimacy (Granovetter, 1973: 1361). Prior work consistently shows that network cleavages among strong ties — i.e., relationships that do involve more time, emotional intensity, or intimacy — are formed along ethnic, gender, religious and social status lines. This finding appears in research on romantic relationships (Kalmijn, 1998; Feliciano et al., 2009; Lewis, 2013), core discussion networks (Marsden, 1988; Smith et al., 2014a), and personal friendship networks (Mouw and Entwisle, 2006; Vermeij et al., 2009; Currarini et al., 2010; Smith et al., 2014b). We do not know much, however, about how segregated people’s weaker ties are. One of the few studies on segregation among weak ties is by DiPrete et al. (2011). Using survey data, they find that Americans’ “acquaintanceship” networks (i.e., weak ties) are highly segregated along racial, political, and religious lines. Studies on segregation on Facebook — when it was still a US, within-college platform — find high levels of segregation by ethnicity and race, similar to the ethnic-racial segregation on campus (Lewis et al., 2008b; Mayer and Puller, 2008; Wimmer and Lewis, 2010; Lewis et al., 2012).

1.7.1 Contributions

I contribute to this line of research in two ways. First, I propose that the study of *online* social networks provides new opportunities to examine the segregation of large personal networks, which we thus know relatively little about. Facebook networks are particularly suited to the study of large networks, as they capture a large subset of complete *offline* networks (Ellison et al., 2011; Van Zalk et al., 2014; Duggan et al., 2015; Dunbar et al., 2015). I illustrate this new approach to the study of segregation among weak ties by considering segregation by ethnicity and gender, as previous work has consistently shown that strong-tie networks of adolescents are highly segregated according to these characteristics (Lubbers, 2003; Baerveldt et al., 2004; Vermeij et al., 2009).

Second, because previous research has exclusively focused on tie formation and segregation among core ties, there is little empirical evidence of the *determinants* of segregation among larger sets of network ties. In this chapter, I am among the first to provide such evidence. In doing so, I consider classic theories on meeting opportunities, and I elaborate on the role of relative group size (Blau 1977a, 1977b) and foci (Feld 1981, 1982, 1984), as these were important in explaining segregation among strong ties (e.g., Kalmijn and Flap, 2001; Mouw and Entwisle, 2006; Smith et al., 2014a). What do these concepts mean? Segregation in personal networks reflects the distributions of the social categories of a population, the so-called *relative group size* effect. For instance, when a society consists of 20% minority members and 80% majority members, the individuals' social networks will consist of 20% minority and 80% majority members. Additionally, individuals who share a *focus* — e.g., schools, neighborhoods, work places — will share their activities and have positive interactions and will thus likely form a tie. Foci are segregated (Feld and Carter 1999), and therefore, personal networks will resemble the structural features of foci. The question is whether and to what extent these theories predict segregation among hundreds of contacts on Facebook.

1.7.2 What Causes Segregation on Facebook?

I measured segregation as the percentage of co-ethnic and same-gender friends on Facebook. Adolescents' Facebook networks had, on average, 76.6% co-ethnic friends. Broken down by ethnicity of adolescents, the Dutch majority have by far the most-segregated networks, with 91.5% of their Facebook friends having a similar ethnicity. Turkish adolescents have, on average, 40.6% co-ethnic friends, Moroccan 28.5%, and Dutch Caribbean 9.2%. Somewhat more than half (56%) of respondents' Facebook friends have the same gender as the respondents.

Under what conditions do these patterns of segregation occur on Facebook? Using opportunity theory in the tradition of Blau (1977a) and Feld (1981), I hypothesize and find that the relative sizes of groups in society and foci are strongly associated with segregation on Facebook (adjusted for selectivity in the privacy of Facebook friend lists). The gender distribution in a population is often 50/50, whereas the distribution of ethnicities is much more unequal. Given this discrepancy, I hypothesize and confirm that gender homogeneity is lower than ethnic homogeneity in Facebook networks. Because ethnic majority members have more opportunities to meet similar others, I expect and find that the ethnic majority members, compared to ethnic minorities, have much higher levels of ethnic segregation on Facebook. Groups in society segregate over foci, and the ties that emerge within them resemble these structural features of the foci (Feld, 1981; Feld 1984). I hypothesize and find that segregation in foci is positively related to segregation on Facebook. I thus contribute to the understanding of processes that underlie segregation in large networks and simultaneously illustrate that these existing but fundamental hypotheses can be tested in novel ways using online social network data.

1.8 Are Core or Facebook Networks More Segregated?

The second part of **Chapter 4** is devoted to explaining differences in ethnic and gender segregation between core and larger networks. It asks whether and why core networks are more segregated than larger online networks. There is speculation that core networks are more segregated than larger networks (e.g., Granovetter, 1973; Putnam, 2000; Mollenhorst et al., 2008; Son and Lin, 2012), although few studies have empirically studied this pattern. One exception, however, is the study of DiPrete et al. (2011). They find that Americans' core and larger networks are equally segregated. I elaborate upon their work and am among the first to theoretically elaborate on and empirically test the conditions and mechanisms that create differences in the levels of segregation among core networks and larger Facebook networks. In doing so, I focus on theories of meeting opportunities (Blau, 1977a; Feld, 1981), homophily (Byrne, 1971), and balance (Heider, 1946).

Homophily, which is pervasive in core networks, refers to the pattern where individuals seem to inherently prefer befriending similar others (e.g., in terms of ethnicity; McPherson et al., 2001). This could be due to either a psychological preference for similar others (Byrne, 1971) or the fact that among similar pairs there are fewer cultural boundaries to overcome (Kalmijn, 1998). *Balance* refers to the tendency of triadic closure in social networks (Heider, 1946; Granovetter, 1973): when A is friends with B, and A with C, then B and C are likely to connect. This can be because of the psychological strain of individuals in an *unbalanced* network configuration (Heider, 1946) or because individuals seek opportunities for unconnected pairs in triads to become connected (Feld, 1981). Previous research has shown that homophily and balance both affect segregation in core networks (McPherson et al., 2001; Mollenhorst et al., 2011).

1.8.1 Causes of Differences in Segregation Between Core and Facebook Ties

Averaged over all of the respondents, I find that approximately three-quarters of the respondents' friends on Facebook are of a similar ethnic background, and this ratio is on par with ethnic homogeneity among core networks (which resembles the finding by DiPrete et al. [2011]). However, if I split these estimates by ethnicity, only the majority members' core networks and online networks are equally ethnically homogeneous, whereas the minority members have lower levels of ethnic homogeneity in their online than their core networks. Slightly more than half of the online network friends have the same gender as the respondents, whereas in the core networks, the ratio is well above 80%.

How do I explain these findings? In this chapter, the presence of online network data pushes me to advance theory on network formation. I do so by focusing explicitly on the *interplay* among existing theories on homophily, balance, and meeting opportunities. I theorize that Facebook networks initially mirror the features of structural meeting opportunities (as follows from opportunity theory), but only similar dyads transition into stronger bonds as time proceeds, whereas weak ties will continue to reflect the features of the meeting opportunities. This results in the pattern that core ties are more segregated than weak ties (as speculated by Granovetter [1973, 1983] and others [Blackwell and Lichter, 2004; Son and Lin, 2012]). However, it was never made explicit why dyadic similarity would foster tie strength. I theorize that this is because initial tie-investments are lower and returns on tie-investments are more likely among homogeneous pairs (Windzio and Bicer, 2013; Leszczensky and Pink, 2015) and because triadic closure is more pronounced among homogeneous triads (Feld 1997; Krackhardt and Handcock 2007). Hence, I hypothesize and corroborate that larger networks are characterized by lower gender homogeneity and that among ethnic minority groups, larger networks coincide with lower levels of ethnic homogeneity. Ethnic majority members, however, have very limited meeting opportunities to befriend dissimilar others, as reflected in core networks and larger networks that are equally homogeneous ethnically.

1.9 How Large are Social Networks on Facebook?

In **Chapter 5**, I first estimate the size of the extended social network on Facebook and, thereafter, explain individual variation in this social network size. Individual's extended social networks contain all the contacts whom individuals know on a first name basis (McCarty et al., 2001; DiPrete et al., 2011). A substantial body of prior work suggests that people have close ties with only a few others. Adults, on average, report approximately two to three core ties (McPherson et al., 2006; Hampton et al., 2011; Mollenhorst et al., 2014; Van Tubergen, 2015). Alongside this literature on the core network size, there is a growing body of literature that is developing methods to provide estimates on the extended social network size. Findings show extended network sizes within the range of 550-750 (Zheng et al., 2006; McCormick et al., 2010; DiPrete et al., 2011), and the number of friends on social media is approximately 180-200, on average (Gonçalves et al., 2011; Dunbar et al., 2015; Dunbar, 2016).

1.9.1 Contributions

I contribute to this literature methodologically as well as theoretically. The methodological contribution is that I combine a frequently used survey measure (i.e., the *network scale-up method*) on the extended social network size and the extended social network size measured as the number of Facebook friends to propose a *new measure* of the extended social network size. The extended social network size as measured via the network scale-up method is highly sensitive to respondent errors and it is unclear how large of a subset of social networks are part of the Facebook network. By combining these two measures, I contribute to an ongoing debate in the literature on how to estimate individuals' extended social network size (and variation therein). Additionally, I shed light on which individuals add a larger share of their social network contacts as Facebook friends.

The theoretical contribution is that I explain individual variation in the Facebook and extended social network sizes and move beyond current knowledge on individual variation in the number of core contacts. As of yet, there is no clear theory nor a systematic study on the causes of the extended social network size (Kadushin, 2012: 72). Therefore, I depart from classic theories on opportunities (Blau, 1977a; Feld, 1981), homophily (McPherson et al., 2001), and romantic partners (Kalmijn, 1998) and explore intuitions on the impact of education and gender to develop hypotheses on the extended social network size.

1.9.2 Causes of Social Network Sizes

How large are extended networks on Facebook and the extended social networks of the new measure? I find an average of approximately 379 Facebook friends. The extended social network size of the new measure is, on average, approximately 524, which is consistent with prior work on the extended social network size using the network scale-up method (Zheng et al., 2006; McCormick et al., 2010; DiPrete et al., 2011).

What explains individual variation in these two social network sizes? Again turning to focus theory (Feld, 1981), I hypothesized and found that those who spend more time in socially oriented foci — i.e., in bars/clubs, associations, and concerts — have larger extended networks, which is consistent with prior work showing that foci are key in the formation of strong ties (Feld, 1984; Kalmijn and Flap, 2001; Mollenhorst et al. 2014). Following opportunity and homophily theory (Byrne, 1971; Blau, 1977a; Feld, 1981), I expected and confirm that those who have a pool of potential contacts in which there are more ethnically similar people have larger social networks, as they have more possibilities to make homophilous choices (note that I make a somewhat related argument in **Chapter 4**), but only among the extended networks on Facebook. Furthermore, I hypothesized and corroborate that those in a relationship, girls, and higher-educated individuals had a larger number of Facebook friends than their counterparts. I found no such differences using the new measure of the extended social network size. However, the analyses of the new combined measure did not account for sample selections in Facebook privacy, whereas the analyses considering the number of Facebook friends did. The results suggest that there are differences in network size among those who keep a public or a private Facebook profile. The discrepancies in findings between the two measures of the extended social network size illustrate the importance of adjusting for sample selections in online social network data.

1.10 How Can We Enrich Online Social Network Data?

In **Chapter 4**, I develop a method to predict ethnicity based on names to study segregation in online networks. **Chapter 6** is a methodological study where I advance the method found in **Chapter 4**. Essentially, I first predict the most likely value of ethnicity given one's first name in networks on Facebook, and second, I show how one can test hypotheses with these predicted values for ethnicity. This type of data-enrichment is crucial for the study of online social network structure. This is because the level of individual detail in data gathered from social media networks is often lower when compared to information gathered in survey research (Golder and Macy, 2014; Spiro, 2016). Individual characteristics such as gender, ethnicity, or age are often missing in online network data (Spiro, 2016), which limits the scope of substantive questions that can be addressed using these data. Names are often among the only available indicators in online data and are a clear signal of ethnicity (Lieberson, 2000; Chang et al., 2010; Bloothooft and Onland, 2011). Therefore, I use names to predict ethnicity in online social networks. There are two studies that relate most to the procedure I propose: that of Chang et al. (2010), who use a probabilistic Bayesian approach, and, thus, the study found in **Chapter 4** (i.e., Hofstra et al. [2017]), who use a supervised learning approach to assign ethnicity based on names.

1.10.1 Contributions

I contribute to these studies in two ways. First, the two prior studies did not model the possibility of different ethnicities among people carrying the same names. I statistically take into account this uncertainty for a more realistic representation of the relationship between ethnicity and names. Second, I show how to test hypotheses with the predicted variable as an independent variable while simultaneously accounting for the uncertainty in the predicted values of this new variable. To show the promise of this approach, I provide an example of hypothesis testing. Following up on studies investigating whether or not ethnic diversity has detrimental effects on trust and social cohesion (Putnam, 2000; Van der Meer and Tolsma, 2014; Abascal and Baldassarri, 2015), I examine the relationship between ethnic homogeneity in Facebook networks and trust.

1.10.2 Tackling Uncertainty

In my procedure to predict ethnicity based on names, I account for two types of statistical uncertainty: first, for the fact that individuals with similar first names may each carry a different ethnicity and, second, for the fact that the model coefficients in the prediction model of interest (e.g., linear regressions) carry uncertainty. Consistent with recent findings on the relationship between ethnic diversity and trust (Abascal and Baldassari, 2015), I find that the predicted percentage of co-ethnic friends on Facebook is not associated with trust. This procedure is compared with two more-straightforward ways to predict ethnicity given one's first name: using a simple majority rule and the supervised learning procedure found in **Chapter 4**. The majority rule leads to false-positive statistical inferences, under the assumption that there is *no* relationship between ethnic diversity and trust. Furthermore, the confidence intervals of coefficients of the method in **Chapter 4** are narrower than the procedure of **Chapter 6**. Hence, the results of the method outlined in **Chapter 6** are less prone to false-positive results compared to the two other methods and can provide more-conservative tests of hypotheses on the potential consequences of online social network structure.

1.11 Conclusions: Have We Found Our Telescope?

1.11.1 Activity on Social Media

The first research aim was to *describe and explain individual differences in activity on social media*. **Chapters 2** and **3** examine individual differences in *membership* in and *privacy* on social media, respectively, as two key dimensions of activity. Generally speaking, socially inactive individuals, boys, ethnic minorities, those with few friends on social network sites, and those with fewer digital resources are less likely to be participants in social media and have been underrepresented in studies using public data on social media up to 2010 (i.e., the period in which I studied this question). Findings on Facebook privacy further pinpoint selectivity issues in the study of online social networks in 2014. This selectivity in privacy is the crucial pitfall in online social network analysis that should be considered. The reason for this is that membership, as opposed to privacy, became less of an issue, as membership rates among adolescents increased from 84% to approximately 95% between 2010 and 2014 — i.e., nearly everyone in this age group became a member. Also among other age groups membership rates increased sharply (see

Figure 1.1). Specifically, ethnic minorities, girls, younger, lower educated, those with more friends keeping private profiles, and unpopular individuals are more likely to be among the 25% of people who do not publicly display their Facebook networks. These groups may be underrepresented in studies that only consider online social network data. **Chapter 5** on the extended social network size on Facebook particularly illustrates that not adjusting for selectivity in Facebook privacy biases results.

To continue with Watts' (2011: 266) analogy: are digital trace data social scientists' telescope? Well, we may have "finally found our telescope" as social scientists, but the device may be somewhat more limited in *radius* — i.e., selectivity in membership (although this is less of an issue nowadays) — and *resolution* — i.e., selectivity in privacy — than it is assumed it to be. Theoretically, I found that hypotheses on social contagion and peer influence, popularity, trust, and platform-specific characteristics were key to predicting this selectivity.

1.11.2 Structure of Online Social Networks

The second research aim was to *describe and explain individual differences in the structure of online social networks.* **Chapters 4** and **5** elaborate causes of *segregation* and the *network size* on Facebook as two key dimensions of online social network structure. I provide novel tests of fundamental prior hypotheses — i.e., foci and relative group sizes — as well as new tests on the development of relationships in terms of tie strength — i.e., the interplay among opportunity, homophily, and balance. In Watts' terms, these chapters aspire to "revolutionize our understanding of ourselves and how we interact" via the use of digital trace data, while adjusting for privacy. The new "telescope" (i.e., social media platforms) enabled new tests of classic prior hypotheses on social network formation; by and large, these hypotheses remain relevant. Furthermore, the telescope provided possibilities to test new propositions, for instance, whether there are differences in segregation between strong and weak ties. In **Chapter 6**, I enrich social media data. This chapter is intended for the applied social scientist who wishes to take advantage of online data sources to answer substantive questions on online social network structure. I essentially perform an *update* to Watts' telescope to make its observations more valuable. I conclude with key findings.

First, segregated meeting opportunities prohibit intergroup friendship formation in Dutch society — among the core ties as well as among hundreds of social contacts.

Facebook networks are highly ethnically segregated, and this is mainly driven by the ethnic majority members' Facebook networks being highly segregated. Because of discrepancies in relative ethnic group sizes in Dutch society, ethnic minority members' Facebook networks are much more diverse. These same discrepancies cause gender segregation to be much lower than ethnic segregation. Additionally, segregation in foci predicts segregation on Facebook. These findings are consistent with prior work showing the importance of meeting opportunities in tie formation of core networks (e.g., Vermeij et al., 2009; Smith et al., 2014a), and I show their relevance in explaining segregation among large online networks. Core networks are more segregated by ethnicity than larger Facebook networks, but only among ethnic minority members. This suggests that, given the opportunity, tie strength may increase with dyadic similarity as a result of homophily and balance.

Second, the number of Facebook friends among adolescents is approximately 379, on average. The number of Facebook friends is higher among girls, ethnic majority members, and higher educated. The extended social network size in a combined measure of the number of Facebook friends and the network scale-up method is on average approximately 524. Theories on opportunities, homophily, romantic partners, and intuitions on gender and education predict the number of Facebook friends rather than the new estimate of the extended network size (likely resulting from not adjusting for sample selections).

Third, it is possible to enrich Facebook data and upscale the level of individual detail (e.g., in terms of ethnicity) using names, but subtle data-analytic approaches are needed, as simpler methods are susceptible to false statistical inference.

Table 1.1 shows a selection of findings in this dissertation by three key demographic characteristics: gender, ethnic background, and educational level. To sum up, online social networks can be used as a novel tool for the study of social networks in general, as I advocate throughout this dissertation. I intend to provide a (small) theoretical and empirical push forward in the field of analyzing online social networks. In conclusion, I would not label the availability of online network data and the development of this field a *revolution* but more an *evolution* towards a 21st-century empirical sociology.

Table 1.1: Summary of a selection of findings by demographic characteristics.

Demographic	Finding	Outcome	Chapter
Gender	• Girls more often member of SNS[a] than boys	Membership	**Chapter 2**
	• Girls more often opt for Facebook privacy than boys	Privacy	**Chapter 3**
	• Adolescents slightly more likely to befriend own gender on Facebook	Segregation	**Chapter 4**
	• Girls have larger Facebook networks than boys	Network size	**Chapter 5**
Ethnicity	• Ethnic minority less often member of SNS, while they adopted Facebook early	Membership	**Chapter 2**
	• Ethnic minorities more often opt for Facebook privacy than ethnic majority	Privacy	**Chapter 3**
	• Ethnic majority members' Facebook networks highly segregated	Segregation	**Chapter 4**
	• Ethnic minorities' Facebook networks smaller than that of majority members	Network size	**Chapter 5**
Education	• Lower educated more likely to be member of SNS	Membership	**Chapter 2**
	• Lower educated more often opt for Facebook privacy than higher educated	Privacy	**Chapter 3**
	• No educational level differences in segregation	Segregation	**Chapter 4**
	• Higher educated have larger Facebook networks than lower educated	Network size	**Chapter 5**

[a] SNS = social network site

1.12 Limitations and Issues For Future Research

Although I was able to describe and explain some of the key aspects of adolescents' behavior concerning social media, there are many other interesting social media behaviors, nor are the empirical chapters without limitations. Moreover, the limitations and findings of the empirical chapters raise new and intriguing research questions for future research to take up. I discuss three topics for future consideration.

1.12.1 Towards Random Samples of General Target Populations

The field of analyzing social media behavior is growing, but random samples of a general target population in a country (e.g., adults) are almost non-existent. Almost without exception, the (relatively) early studies on variation in Facebook usage (Ellison et al., 2007), privacy on Facebook (Acquisti and Gross, 2006; Tufekci, 2008; Lewis et al., 2008a), and ethnic-racial segregation on Facebook (Mayer and Puller, 2008; Wimmer and Lewis, 2010) use convenience samples from US college students. The empirical chapters of this dissertation rely on a random sample of Dutch adolescents. The presented results therefore likely better generalize to all Dutch adolescents than the findings of convenience samples of US students do to the entire US student body.

However, if we consider online platforms to be our telescope and all the planets in the universe to be the complete human population, then the study of *adolescents* looks only at *one planet* among many. The share of adults using Facebook is rapidly increasing: 79% of online American adults and 63% of online adults in the UK used social media in 2016 (Greenwood et al., 2016; Office for National Statistics, 2016). This same pattern holds for the Netherlands, where even a large share of *older* adults (e.g., 55 years or older) used social media in 2016 (see **Figure 1.1**). Novel tests of hypotheses on social media activity and the structure of online networks among random samples of adults is a next step necessary to propel this field of research. Concerning activity on social media, a starting point would be to study selectivity in adult users' membership patterns (e.g., who are among the 21% non-users of Facebook in the US?) and privacy on Facebook (e.g., what are parents' privacy strategies on Facebook and how does this relate to their children's privacy?). Questions on online social network structure among adults can also be considered. For instance, theories on segregation predict that ties formed at school during adolescence may be stable and carry over into adulthood (McPherson et al., 2001: 432). As of yet, however, this has not been directly tested. Such a question on the structure of online social networks could be answered with the data presented.

1.12.2 Towards Studies on the Consequences of the Structure of Online Social Networks

A growing body of work considers the *consequences* of the structure of online social networks. For instance, there are studies on access to social capital (Bohn et al., 2014; Brooks et al., 2014), voting (Bond et al., 2012), happiness (Kramer et al., 2014), and mortality (Hobbs et al., 2016) using social media data. However, the outcome variables in these studies are often measured from the digital trace data and lack individual-level detail, and the outcomes may be biased towards users of social media. Therefore, one of the next steps is to conduct studies on the consequences of online social network structure using the data-analytic approaches presented throughout this dissertation — i.e., linking offline and online data and adjusting for sample selection biases. I mention two puzzles on the consequences of the structure of online social networks that are possible to take up with the data presented throughout this dissertation.

First, as a direct follow-up to **Chapter 4**, one can examine how ethnic homogeneity in online networks on Facebook relates to *ethnic prejudice*. Contact theory would predict that having more *face-to-face* contact with outgroup members reduces prejudice toward them (Allport, 1954; Pettigrew and Tropp 2006). Recent work shows that also other forms of contact, such as via television news or newspapers, affect prejudice (Visintin et al., 2016). Therefore, the assumption of *face-to-face* contact needs to be studied in more detail. One potential direction for future research is whether the mechanism also applies to the relationship between ethnic segregation on Facebook and ethnic prejudice? This question is crucial to take up as societies become increasingly multi-ethnic (Castles et al., 2013).

Second, novel information is argued to easily diffuse through weak ties (Granovetter, 1973), and network positions between sub-cliques in networks enable individuals to control (i.e., "broker") flows of information (Burt, 2000). These weak ties and network positions are argued to beneficial for labor market outcomes. However, extant literature on social network effects on labor market outcomes is often limited by difficulties in measuring network structure (e.g., it is often measured for only a small set of *strong* ties), and by the issue that social network ties can be a *result* of an occupation. One can circumvent both of these issues by considering how individuals' large *online* network structure relates to the socioeconomic status of *first* jobs (i.e., reducing the network selection issues where ties are a result of an occupation).

1.12.3 Toward a Systematic Study of Multiple Social Media Platforms

Although Facebook is the most popular social network site worldwide, there are many other (oftentimes region-specific) platforms that have hundreds of millions of members — e.g., *LinkedIn*, the Chinese *Sina Weibo*, or the Russian *VKontakte*. The popularity of these platforms is volatile, and membership figures fluctuate by millions of users in short periods of time.

The observation that membership in social network sites is so unstable carries implications for researchers. Imagine that adolescents leave Facebook *en masse*, something that has been suggested (e.g., Madden et al., 2013). However, they do not delete their Facebook accounts, but the accounts are left in an idle state. If one, then, were to study, say, the number and content of status updates, the results may be biased by the inactivity of the group of adolescents in the sample. This means that we need to study activity more directly, for instance, by considering *how often* individuals use Facebook. The migration of adolescents away from Facebook would be less of an issue if one were to consider the networks from Facebook as only a snapshot of a large subset of complete offline networks. One should thus be specific about the online behaviors under consideration, on what platform, for which target population, and how the platform may bias observations by design. This can prevent findings arising as an artifact of the platform-of-choice (Lewis, 2015a).

Whereas the absolute levels of Facebook membership are increasing (Greenwood et al., 2016), there may be groups among which its popularity is decreasing. In **Chapter 2**, I provide some insight into the determinants of social media membership in a volatile situation. This presents an intriguing question regarding whether and to what extent the examined determinants are stable across platforms. In other words, is the process of migration of one social media platform to the other governed by a similar set of determinants? If so, scholars could develop *predictions* of platform popularity derived from theories on the interplay between user and platform characteristics.

Finally, scholars have recently started to collect information about the same set of respondents from multiple sources of data (including surveys and social network sites, but also mobile phone data, geo-location, etc.; see Stopczynski et al., 2014). This implies that they can compare the behaviors across platforms within the same set of respondents in detail. I would commend further efforts in this direction, because it allows scholars to answer questions and test theories on social networking behavior in an unprecedented way. One question is, for instance, which social contacts across platforms online and across contexts offline overlap and for what reason?

Part I

Activity on Social Media

Chapter 2

Who Was First on Facebook? Determinants of Early Adoption Among Adolescents[1]

Abstract: *We study what determines whether someone is an early Facebook adopter in a context in which Facebook is still relatively new compared to a far more popular Dutch SNS (Hyves). We use representative survey data among 4,363 adolescents aged 14-15. First, adolescents who participate in more leisure activities, who have more digital resources and who have more friends that are SNS members are more likely to be SNS members. Second, we hypothesize and show that for adopting communication technology that highly fluctuates in popularity and is highly time-dependent, individuals are more likely to be early Facebook adopters when the number of their friends who are Facebook members increases. Finally, non-native adolescents are also more likely to be early Facebook adopters.*

[1] A slightly different version of this chapter is published as: Hofstra, B., Corten, R., and Van Tubergen, F. (2016). Who Was First on Facebook? Determinants of Early Adoption Among Adolescents. *New Media & Society, 18*(10), 2340-2358. Hofstra wrote the main part of the manuscript and conducted the analyses. Corten and Van Tubergen substantially contributed to the manuscript. The authors jointly developed the idea and design of the study. This chapter was presented at the "Migration and Social Stratification" seminar at the ICS, at the "Dag van de Sociologie 2014" in Antwerp, and at the "1st European Conference on Social Networks" in Barcelona. I thank Jeroen Weesie for advice on methodological issues and two anonymous reviewers from *New Media & Society* for their helpful suggestions.

2.1 Introduction

In the past decade, the popularity of Social Networking Sites (hereafter: SNSs) has increased spectacularly (boyd and Ellison, 2007). Millions of users of Web services such as Facebook, Instagram and Twitter connect through virtual “friendship” networks, using them to share information, experiences, opinions and emotions.

The literature on the consequences of SNS usage for various outcomes is rapidly growing, and many studies show that SNSs play an important role in people’s lives. For instance, several studies have found that people who use SNSs more frequently experience greater well-being (Steinfield et al., 2008; Burke et al., 2011). Other studies have observed that more intense SNS users have more bridging social capital, that is, they have a greater potential to access novel information via interaction with acquaintances who are connected to different foci (Ellison et al., 2007; Brandtzæg, 2012; Brooks et al., 2014; Ellison et al., 2014).

Although the consequences of SNSs for various outcomes have been extensively studied in the literature, remarkably few studies have examined the causes of SNS membership and activity. However, it is important to examine these causes because a key characteristic of SNSs is that their popularity is highly time-dependent. For example, MySpace was founded in the US in 2003; it rapidly became popular between 2003 and 2008 but following that period, it lost many members. Facebook started in the US in 2004 and has continued to grow there; subsequently, it has spread worldwide and became popular in other nations. With more than 1.3 billion members, Facebook is now the largest SNS in the world. Given these enormous fluctuations, it is important to study variations in SNS membership. Who is joining a particular SNS and who is not? Who are the early adopters of a new SNS and who are the followers? In the current study, we aim to contribute to this underdeveloped literature on the causes of SNS membership. Specifically, we study the identities of the Netherlands’ relatively early adopters of Facebook in 2010 and 2011, and we then compare those adopters’ characteristics with the members of Hyves, a Dutch SNS that at the time was far more popular. In addition, we study what determines whether a person becomes an SNS member.

Our study therefore elaborates on the few studies of the determinants of SNS membership and activity. Prior research has shown that ethnicity and race are related to SNS membership. In the US, Asian Americans use Twitter less often than other ethnic and racial groups (Hargittai and Litt, 2011). Gender differences have also been found. Hargittai (2008), using survey data from an American college, has found that women are more likely than men to be SNS users. Based on a sample of MySpace user profiles, Thelwall (2008) has found that MySpace users are disproportionally female. Moule et al. (2013), using a US convenience sample, found that women were more likely to be SNS members than men and that being younger and having a phone promotes SNS membership. Women value maintenance of relationships on SNSs more than men do (Orchard et al., 2014), and since these are SNS' prime purposes, it stands to reason that women are more likely to be members. Using surveys of college students in Hong Kong, Cheung et al. (2011) have found that intention of Facebook usage is influenced by others' opinions of Facebook.

We elaborate and extend these earlier studies by investigating the determinants of relatively early adoption of Facebook. The setting that we use is the Netherlands between October 2010 and April 2011. We compare membership in Facebook with membership in Hyves, a Dutch SNS that was then far more popular and reached its peak membership numbers in 2010. Figure 2.1 shows the popularity of Facebook versus Hyves, in which we see large changes in the interest in both websites. We examine which people were relatively early Facebook adopters in addition to or instead of Hyves: the innovators, early adopters and a small part of the early majority in Rogers' (2003) terminology. We focus on adolescents because particularly among this subpopulation, SNSs have become an important medium for social interaction (Brandtzæg, 2012), and we can gain insight into why some adolescents select Facebook rather than Hyves. Knowledge on Facebook adoption also provides insight into boundaries to social interaction between different groups of adolescents, in the sense that some groups are more likely to come into contact with other groups via SNS membership.

Our study investigates what determines the early adoption of platforms that experience fluctuations of literally millions of members in rather short periods by contrasting membership of the popular Hyves with membership of the relatively new Facebook. Moreover, our study context allows us to gain innovative knowledge about a process that typically is highly dynamic: early adoption and the rise of one of the most prominent communication innovations in the last decade (Facebook) during a unique historical context in which there is already a rather similar innovation (Hyves) on the market. Hyves' spectacular rise and demise in the Netherlands at the hands of Facebook illustrate the volatility of media used for online interaction. From the perspective of both financial investors and SNS providers, it is crucial to gain insight into the processes that govern the dynamics of SNS membership fluctuation. Hyves was purchased in 2010 for €43.7 million,

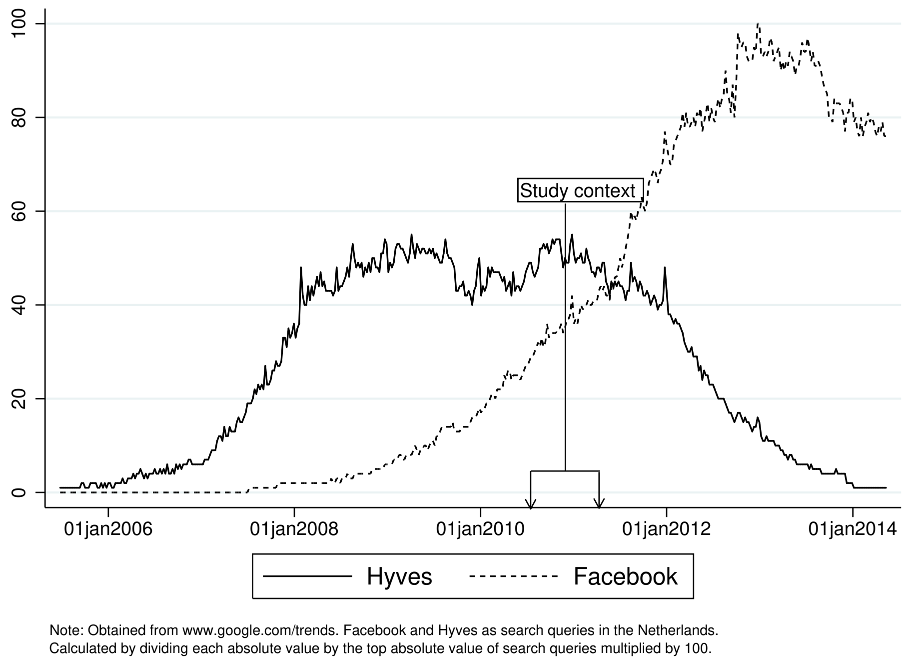


Figure 2.1: Standardized Google search queries for Facebook and Hyves in the Netherlands: 2005-2014.

whereas it depreciated to €7.7 million in 2013, for a loss of 82.4%.

We use large-scale, nationally representative data (N=4,363) about adolescents instead of the convenience samples — such as (US or UK) college students — that were often used in previous work (e.g., Hargittai and Litt, 2011; Moule et al., 2013; Orchard et al., 2014). This makes generalizable claims to a broader population on early Facebook adoption more convincing and we shed light onto potential sample selection biases in the abundance of studies that focus on consequences of SNS usage.

2.2 Theory and Hypotheses

2.2.1 The research context of social networking sites

Following boyd and Ellison (2007: 211), we define SNSs as:

> *"...web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system."*

Hyves and Facebook are good examples of SNSs and are relatively similar. Users create a profile, provide personal information and invite other users to become connected. With these connections, users can interact via personal messaging, post directly on others' personal profile pages and react to others' posts (Caers et al., 2013).

Hyves was a Dutch SNS with approximately 10 million members out of a Dutch population of 16 million, thus comprising a large portion of the Dutch population. The vast majority of users (86%) were Dutch (Corten, 2012). After its peak in 2010, Hyves became less and less popular and eventually was shut down in December 2013. By that time, Facebook had completely taken over as the dominant SNS. Dutch translations of Facebook's pages were available from May 2008. In 2010, approximately 30% of the adolescents aged 14-15 that we study were on Facebook, and in 2014, this proportion rose to more than 95% among this age group.

To understand why some Dutch people in 2010 were among the 30% of first adopters of Facebook or members of the dominant Hyves, or both, we first identify some conditions that generally affect SNS membership. Thereafter, we propose hypotheses to understand why some people adopted Facebook relatively early.

2.2.2 Membership of a social networking site

In this section, we derive hypotheses about why adolescents become SNS members, whereas others do not. We distinguish among three mechanisms.

First, SNS membership might be explained by adolescents' activity levels.[2] Following the line of reasoning by Moule et al., (2013), we assume that certain lifestyles are related to adoption of SNSs (Rogers, 2003). We assume adolescents with higher activity levels are more likely to be SNS members. Some individuals are cognitively more capable than others of pursuing a broader range of activities in their leisure time (e.g., Sullivan and Katz-Gerro, 2007) and are considered to be cultural omnivores, which is associated with having higher leisure-time activity levels such as going to the cinema, going to parties, visiting family or reading a book. We argue that SNS membership is a leisure activity engaged in by adolescents and therefore, SNS membership is more likely among adolescents who have generally higher activity levels (i.e., a combination of diversity and time spent on these activities) in their leisure time. This indicates that those adolescents are more capable of pursuing a broad range of activities and that SNS membership is one of those activities. In addition, when adolescents engage in more leisure-time activities, they might also want to share the experiences obtained from these activities with their friends. In other words, when individuals read a book, SNSs provide them with an outlet to share their opinions about the book. Highly active adolescents might have the preference to display their activities or even to coordinate those activities with peers by means of interaction on SNSs. Thus, our first hypothesis is as follows:

> **Hypothesis 1:** *Adolescents with higher activity levels are more likely to be members of SNSs than are adolescents with lower activity levels.*

A second mechanism that could explain SNS membership is that of exposure to digital resources. SNSs are digital by nature and therefore, we assume that SNS membership is more likely to occur when there are more resources at one's disposal that result in more digital connections. There are two reasons that such an influence might work. First, individuals who have more resources to be exposed to SNSs have greater likelihood of being exposed to them. We assume that greater exposure to SNSs already increases the likelihood of becoming an SNS member (Hargittai, 2008) because of the knowledge that individuals gain from this expo-

[2] An underlying assumption of our hypotheses is that individuals are goal-oriented in their behavior. We assume that, given specific attributes of adolescents, some adolescents benefit more from becoming an SNS/Facebook member. This is consistent with various more specific behavioral theories such as Ajzen's (1991) theory of planned behavior or Hedström's (2005) theory of desires, beliefs, and opportunities.

sure. For example, adolescents can be invited by online gaming friends to become an SNS member or to watch a news broadcast on the topic of SNSs. This is contingent with Rogers (2003), who argues that those who have more exposure to mass media communication are more likely to adopt new technologies. Second, when individuals have a greater ability to register as SNS members, they are more likely to be SNS members (Hargittai, 2008). This entails that a person has more resources available to actually register as an SNS member. For instance, when a person owns a smartphone or a computer for personal use, registering with an SNS is easier. Thus, we assume that the likelihood of registering as an SNS member increases with increased possibilities of doing so. We assume that when adolescents have their own smartphones, their own computers, home Internet access, gaming consoles and televisions, they are more likely to be exposed to SNSs and have more opportunities to register. We call these resources digital resources and hypothesize:

> **Hypothesis 2:** *Adolescents with more digital resources are more likely to be members of SNSs than adolescents with less digital resources.*

Rogers (2003) argues that there are social diffusion processes in adopting technologies. Following Rogers, we argue that the final mechanism that might cause an individual to become an SNS member is peer influence, which captures the tendency of friends to increasingly resemble one another based on individual characteristics (McPherson et al., 2001). In line with what Hargittai (2008) and Hargittai and Litt (2011) expect and suggest as a topic for closer investigation, we assume that an individual's SNS membership is affected by his or her friends' SNS memberships. There are three reasons why such a peer influence might exist. First, because of the social nature of SNSs, becoming an SNS member is more attractive when more of a person's friends are already members: this is the effect of network externalities. This implies that the benefits derived by individuals from using a service (e.g., SNS adoption) increase when the number of other individuals who also use this service (e.g., others' SNS adoption) increases (Liebowitz and Margolis, 1994). In other words, being SNS member is more fun when one's friends are members because the SNS provides novel ways to interact, to share content and to stay informed about one another. Second, joining an SNS might be a result of imitation. During adolescence, individuals go through important life changes and cope with many insecurities (Corten and Knecht, 2013). Consequently, adolescents look to their friends as examples of appropriate behavior (Marsden and Friedkin, 1993).

In our context, because of imitation, adolescents become SNS members if their friends are also members. Finally, there might be norms within groups that push conformity among friendship groups. In essence, this means that within friendship groups, SNS membership is a norm and friends expect membership. When more friends in an adolescent's class are SNS members, it is more likely that the adolescent will also join because of the abovementioned processes. Classes within schools are a particular attractive context to study peer influence processes because they consist of well-defined social contexts (Corten and Knecht, 2013). Furthermore, adolescents spend a large portion of their time in class and from this fact alone, adolescents could be influenced by their peers in class. Therefore:

> **Hypothesis 3:** *Adolescents with more friends in class who are SNS members are more likely to be SNS members than adolescents with fewer friends in class who are SNS members.*

2.2.3 Early adoption of Facebook

In this section, we develop hypotheses that could explain why some adolescents were among the relatively early adopters of Facebook.

Early adoption of Facebook might be the result of a social diffusion process in class, in which case friends are among the first users (Rogers, 2003). Because of peer influence processes we assume that adolescents select a particular type of SNS. Adolescents have an incentive to join Facebook if the friends with whom they interact are also Facebook members. Specific SNS membership coordination leads to the hypothesis:

> **Hypothesis 3a:** *Adolescents who had more friends in class who were early adopters of Facebook (Hyves) were themselves more likely to be early adopters of Facebook (Hyves).*

The early adopters of Facebook might also have adopted Facebook because of its international character. Unlike Hyves, Facebook is an international SNS, and therefore, it was particularly attractive to some adolescents. Specifically, we expect that adolescents with friends and family abroad were more likely to be early Facebook members and were less likely to use only Hyves. Approximately 86% of Hyves members were Dutch (Corten, 2012), whereas a maximum of 4.1% of Facebook members were Dutch. We assume that adolescents with immigrant backgrounds are more likely to have friends and family abroad. Thus, adolescents with a non-Dutch national origin — i.e., members of the ethnic minority — coordinated their SNS membership with friends or family in the country of origin and reached their goals by interacting with these friends or family members via Facebook. Therefore:

> **Hypothesis 4:** *Adolescents with a non-Dutch national origin were more likely than Dutch adolescents to be early adopters of Facebook.*

If adolescents with a non-Dutch national origin were more likely than Dutch adolescents to be early Facebook adopters, then the friends of those adolescents of other national origins might have also positively affected early adoption of Facebook, independent of national origin. Under the assumption that adolescents coordinate their SNS membership with their friends, if adolescents with a non-Dutch national origin were more likely to be early Facebook members, then adolescents with more friends of non-Dutch national origin were more likely to be Facebook members themselves. Thus:

> **Hypothesis 5:** *Adolescents who had more friends of non-Dutch national origin were more likely to be early adopters of Facebook than were adolescents with more Dutch friends.*

Elaborating further on the role of social diffusion mechanisms, age might also play a role in early Facebook membership. Facebook's initial target population was college students in the US (Caers et al., 2013) — a selective population that was approximately 18-25 years of age. After granting access to all college students, Facebook was launched among US high-school students at the beginning of 2005 (boyd and Ellison, 2007). In 2010, our period of interest, those high school students (or at least a considerable fraction of them) were likely to have made the transition to college. We assume that through a social diffusion process, adolescents in the Netherlands were "infected" and became Facebook members. This was possible thanks to connections with US college students via summer schools, internships or exchange programs, where Dutch and US college students interacted and Dutch students made contact with Facebook. In other words, Dutch college students might have been the first group of Dutch residents who were Facebook members. Given that networks are segregated by age (e.g., McPherson et al., 2001), we assume that older Dutch adolescents were more likely to be friends with older individuals such as Dutch college students (who were likely to be Facebook members) instead of or in addition to high school students. Consequently, we assume that these older friends (e.g., college students) may have positively affected adolescents' Facebook membership via the peer influence mechanisms elaborated earlier. Hence:

> **Hypothesis 6:** *Adolescents who either had older friends (H6a) or were older themselves (H6b) were more likely to be early adopters of Facebook than adolescents with younger friends and adolescents who were younger themselves.*

Early adoption of Facebook might also be caused by mechanisms other than social diffusion. The choice of a new SNS such as Facebook might be driven by the need for distinction — in essence, to differentiate oneself from other classroom peers who used Hyves. We assume that popular adolescents in particular make "risky" decisions and are more likely to explore new pathways and behaviors than do less-popular adolescents. Popular adolescents are considered attractive because of behaviors and characteristics that deviate from the behavior of their "normal" peers (Dijkstra et al., 2009), including risky behaviors such as smoking. Popular adolescents make choices that are associated with higher social status and "coolness" (Brechwald and Prinstein, 2011). These choices are most likely more distinctive than the choices of less-popular adolescents. In 2010, Facebook was a relatively new SNS and it might be that popular adolescents became Facebook members: selecting a not-yet-popular SNS might be a strategy to distinguish oneself from the majority of SNS users and from other adolescents. In other words, popular adolescents are the trendsetters, adopting the new and much more *progressive* Facebook. In addition, Facebook was a much more risky to select because outcomes in terms of social interaction were uncertain: fewer people were members. Thus:

> **Hypothesis 7:** *Popular adolescents were more likely than less-popular adolescents to be early adopters of Facebook.*

2.3 Data

We use data from the first wave of the Dutch section of the "Children of Immigrants Longitudinal Study in Four European Countries" (CILS4EU) (Kalter et al., 2013) to test our hypotheses. The data were collected in four European countries: Germany, the Netherlands, Sweden and the United Kingdom. Data were collected from October 2010 to April 2011. Data were collected among (primarily) adolescents 14-15 years of age, with an oversampling of immigrant minority youth. The survey consists of a self-completion questionnaire concerning many individual characteristics, attitudes and leisure-time activities. The survey also includes complete classroom social network data. Data collection took place at high schools and these were selected according to four strata based on educational track levels and percentage of non-Western immigrant students in schools to ensure an oversampling of non-Western immigrants, aligned with the goal of data collection (approximately 30% in the Dutch data). Research teams visited schools to give standardized instructions about how to complete the questionnaire, and researchers were present while students completed the questionnaire. In the Netherlands, 100 schools, 222 classes and 4,363 students participated. Schools' initial response rate was 34.9%. If a school refused to participate, a willing replacement school with the same characteristics was sought, which increased schools' response rate to 91.7%. In these schools, 91.1% of the pupils participated.

2.4 Measurements

2.4.1 Dependent variables

SNS membership. The first dependent variable we create is a binary variable if a respondent is a Facebook or Hyves member (1) or not (0). Respondents answered the question "Are you on Hyves: Yes/No" and "Are you on Facebook: Yes/No". If respondents answered "Yes" for Hyves, but were missing on Facebook (or vice versa), respondents score a 1 on this variable.

SNS categories. Among the respondents that are an SNS member we study if respondents are (1) only Facebook member, (2) only Hyves member, or (3) member of both SNSs. For both dependent variables, the number of observations is displayed in Table 2.1.

Table 2.1: Descriptive statistics of SNS membership and SNS categories

	Observations	Percentage
SNS membership[a]		
No	662	15.8%
Yes	3,530	84.2%
Total	3,954	100%
SNS Categories		
Facebook and Hyves	1,204	34.8%
Hyves	2,123	61.4%
Facebook	133	3.8%
Total	3,460	100%

[a] SNS = social networking site.

2.4.2 Independent variables

Activity levels. Respondents indicated how often during their leisure time they did the following eight activities “visit family”, “go to the cinema”, “go out to a café/disco/party”, “read a book”, “go to an association: sports/ music/other’, “go to a concert/ dance party”, “go to a museum” or “read a paper.” Answer categories ranged from 1 (“never”) to 5 (“every day”) and we averaged the items.

Digital resources. We measure how many of the following resources respondents reported as available to them: “personal computer”, “smartphone, e.g., iPhone or Blackberry”, “television” and “a game console, e.g., Playstation, Wii or X-Box”, ranging from 0 to 4.

Number of SNS/Facebook/Hyves members among classroom friends. Respondents answered the question: “Who are your best friends in class (you can write down a maximum of 5 friends)?” We know from these friends (see SNS membership) whether they were members of Facebook, Hyves or neither. We count the absolute number of SNS members (and members of Facebook and members of Hyves) within the respondent’s friends ranging from 0 to 5.

National origin. We construct a variable that indicates respondents’ national origin, distributed over the seven largest ethnic background groups in the Netherlands: 1 “Native Dutch”, 2 “Turkish”, 3 “Moroccan”, 4 “Surinamese”, 5 “Antillean, Aruban (including Curacao, Bonaire, Saint Eustatius and Saba)”, 6 “Other: Western (Europe or English speaking)" and 7 "Other: non-Western". We measure national origin by the country of birth of a biological parent, reported by the parent him/herself (as requested in an additional, parental survey) as a more reliable source. We obtain country of birth from the partner/spouse reported by the surveyed parent if that partner/spouse is also a biological parent. When these values are missing, we acquire the biological parents' country of birth as reported by the child. When respondents have one or more native-Dutch parent, they belong to the national origin of the other parent. When children have parents from different countries, children belong to the national origin of the mother.

Number of friends with non-Dutch national origin. Respondents indicated "Who are your best friends?" They were permitted to nominate a maximum of 5 friends both inside and outside of class, and they indicated whether those friends were "Dutch", "Turkish", "Moroccan", "Surinamese", "Antillean" or had an "Other" background. We create a variable that counts the number of friends with a non-Dutch national origin — i.e., the number of friends who are member of the ethnic minority.

Age. We construct a variable that measures the age of respondents in months. We calculate for each respondent the number of months between the date of birth and the date of the interview. We exclude the respondents 17 years of age or older from the analyses (N=13) because they are extreme outliers (> 3*SD) who might disproportionally affect our results.

Age of oldest best friend. We measure the age in years of the oldest best friends mentioned by the respondents as occupying their core networks (i.e., for the best friends mentioned). We exclude the extreme outliers out of the analyses (N=24, > 3*SD), which means that we measure the age of the oldest friend up to 25 years.

Indegree: popularity nominations in class. We construct a variable to measure respondents' popularity. Respondents answered the question, "Who are the most popular students in class (you can write down a maximum of 5 names)?" We construct a variable that indicates what percentage of classroom students mention the respondent as the most popular student. This is calculated by dividing the total number of classmates' popularity nominations by the total students in class, minus one. This is formally defined as:

$$Popularity_i = \frac{\sum_i K_{ji}}{N-1} \times 100, \tag{2.1}$$

where $Popularity_i$ is the indegree popularity of pupil i, K_{ji} indicates whether pupil j nominates pupil i as popular and N is the total number of pupils in the classroom. Hence, we acquire a standardized variable between classrooms that indicates what percentage of classroom pupils in a given class mentions the respondent as the most popular pupil.

2.4.3 Control variables

We control for respondents either being female (1) or not (0). Second, we control for high school educational track. We create dummy variables indicating the adolescents' high school track. In the Netherlands, when adolescents transition to high school, they are classified into different educational tracks, which differ in terms of level and type of education. These tracks range from 1 "lower preparatory vocational education" to 6 "university preparatory education". Thus, age is not correlated with educational level. Third, because individuals' psychological traits affect their use of SNSs (Orchard et al., 2014), we control for behavioral problems. We averaged six statements where respondents noted how often these statements were true for them, ranging from 1 "Never true" to 4 "Often true". These statements are: "I worry a lot", "I get angry quickly", "I am afraid", "I am sad" and "I feel worthless" (Cronbach's α =.745). We also control for self-esteem, ranging from 1 to 5, averaged over the following items: "I have many good qualities", "I have a lot to be proud of", "I am satisfied with myself the way I am" and "I think that I have a bright future" (Cronbach's α = .798).

Finally, we added number of best friends mentioned and the number of best friends mentioned in class to our models to control for varying network sizes. Table 2.2 shows descriptive statistics for the exogenous variables.

2.5 Hypotheses Tests

2.5.1 Analytical strategy

We perform two sets of statistical analyses to test our hypotheses. First, we estimate a random effect logistic regression for the effect of our independent variables on the binary variable SNS membership. Because our data are hierarchically structured (pupils within classes), we add a random term for a class identifier (Snijders

Table 2.2: Descriptive statistics for the independent and control variables

	Range	Mean	SD[a]	N
Independent variables				
Activity levels	1-5	2.361	2.361	4,280
Digital resources	0-4	2.671	2.671	4,252
Number of friends in class on SNSs[b]	0-5	2.993	2.993	4,109
Number of friends in class on Facebook	0-5	1.101	1.052	4,109
Number of friends in class on Hyves	0-5	2.896	1.446	4,109
National origin	-	-	-	4,363
Dutch	0-1	0.685	-	2,988
Turkish	0-1	0.061	-	266
Moroccan	0-1	0.057	-	248
Surinamese	0-1	0.039	-	169
Antillean	0-1	0.016	-	71
Other: Western	0-1	0.087	-	378
Other: non-Western	0-1	0.056	-	243
Number of friends non-Dutch national origin	0-5	1.294	1.697	4,242
Age (in months)	159-204	180.762	7.001	4,296
Age of oldest best friend	13-25	15.690	1.414	4,209
Indegree: popularity nominations in class	0-100	11.420	17.289	4,033
Control variables				
Female	0-1	0.508	-	4,358
High school educational track	-	-	-	4,347
Lower preparatory vocational	0-1	0.109	-	472
Medium/lower preparatory vocational	0-1	0.155	-	675
Medium/higher preparatory vocational	0-1	0.076	-	330
Higher preparatory vocational	0-1	0.268	-	1163
Senior general	0-1	0.198	-	859
University preparatory	0-1	0.200	-	848
Behavioral problems	1-4	2.075	0.571	4,344
Self-esteem	1-5	3.927	0.575	4,329
Number of friends nominated outside class	0-5	4.594	1.052	4,363
Number of friends nominated inside class	1-5	3.619	1.520	4,363

[a] SD = standard deviation; [b] SNSs = social networking sites.

and Bosker, 2012). Therefore, we control for class-specific tendencies in SNS membership selection. We report average marginal effects (AMEs) of the independent variables on SNS membership. AMEs are more intuitively interpreted than odds ratios (Mood, 2010). In addition, it is problematic to interpret odds ratios as substantive effects due to the unobserved heterogeneity that they reflect (cf. Mood, 2010). For dummy variables, AMEs show the difference in probability of being an SNS member between the two values, estimated over all the observed values of the other variables in the model. For categorical variables, AMEs are interpreted as the difference in the probability of being an SNS member between the categories included in the analyses and the omitted reference category, calculated over all other observed values of the independent variables. For continuous predictors, AMEs are interpreted as the probability increase or decrease in being an SNS member when the predictor variable increases with one unit, estimated over all the possible values of the variables. We use listwise deletion for missing values, which results in a loss observations of approximately 12.1% (N = 530).

Second, we estimate a multinomial logistic regression to test whether our independent variables affect being a member of Facebook only, a member of Hyves only, or a member of both SNSs as our categorical dependent variable. We use a cluster correction for a unique class identifier. In this manner, we adjust standard errors for 221 clusters, obtain robust standard errors, and reduce the likelihood of underestimated standard errors. The results for the multinomial logistic regression are found in the Appendix for Chapter 1, and the hypotheses are tested using a post-estimation technique after this initial analyses. This technique implies that we estimate AME of the independent variable of interest on a specific outcome, given that respondents are members of an SNS. Technically, this means that AMEs on specific SNS membership of a variable are divided by 1 minus the AME on the category not member of an SNS. When we do not consider AME on a specific outcome conditional upon membership, we might mis-specify our model because a considerable number of respondents are then excluded (N = 613). Thus, the AMEs of independent variables in this analysis are interpreted as the increase or decrease in average probability of being, for example, a Facebook member, given that one is a member of an SNS. We control for all variables used in the previously mentioned logistic regression: activity levels, digital resources, female, educational track, behavioral problems, self-esteem, and number of friends inside and outside of class. When we estimate the effect for being a member of Facebook (or Hyves), we predict an AME both for being member of only Facebook (or only Hyves) and for being a member of Facebook and Hyves because both categories indicate that a person is a member of Facebook (or Hyves). In addition, we predict the AME for being on Facebook (or Hyves) combined for these two categories (being a member of one SNS plus being member of both SNSs). Finally, we use listwise deletion of missing values, resulting in a loss of 15.3% (N = 667) of the observations in this analysis.

2.5.2 Membership of an SNS

The results of the random effect logistic regression are displayed in Table 2.3. At least one of the predictors differs significantly from 0 (Wald $\chi^2(20) = 308.670$; Probability $> \chi^2 = .000$).

First, the probability of being an SNS member increases when an adolescent has a higher activity level. For every additional step that an adolescent scores on the variable activity levels, the probability that he or she is an SNS member increases by .44, estimated over all observed values of the other variables. This can be considered a rather large effect, and thus, we find evidence to support H1.

When adolescents have more digital resources, they are more likely to be SNS members. The existence of one additional digital resource increases the average probability of being an SNS member by .23, averaged over all observed values of the variables. Thus, we find evidence to support H2.

Third, we find evidence to support H3; for every additional friend (with a maximum of five) in class who is an SNS member, the average probability of being an SNS member increases by .29. When one moves, for example, from zero to three friends who are SNS members, the average probability of being SNS member increases by .87. Thus, we find evidence for the substantial effect of classroom peers on SNS membership.

2.5.3 Early adoption of Facebook

The results of the post-estimation of the multinomial logistic regression analysis are shown in Table 2.4. At least one of the predictors differs significantly from 0 (Wald $\chi^2(72) = 1240.610$; Probability $> \chi^2 = .000$).

We find evidence to support H3a; the probability of being an early adopter of Facebook increases with the number of classroom friends who are early adopters

Table 2.3: Random effect logistic regression: effects of the independent variables on membership in Facebook and/or Hyves. Presented are average marginal effects (AME).

	Hyp.[a]	AME	S.E.[b]	p[c]
Independent variables				
Activity levels	H1. +	0.441	0.105	0.000
Digital resources	H2. +	0.228	0.045	0.000
Number of friends in class on SNSs	H3. +	0.289	0.056	0.000
Control variables				
Female (ref. male)		0.759	0.107	0.000
High school educational track				
Lower preparatory vocational		0.637	0.200	0.001
Medium/lower preparatory vocational		0.394	0.166	0.018
Medium/higher preparatory vocational		0.559	0.217	0.010
Higher preparatory vocational		0.351	0.143	0.014
Senior general		0.241	0.152	0.112
University preparatory		(ref.)	(ref.)	(ref.)
National origin				
Dutch		(ref.)	(ref.)	(ref.)
Turkish		-1.029	0.222	0.000
Moroccan		-1.340	0.229	0.000
Surinamese		-0.654	0.248	0.008
Antillean		0.327	0.454	0.471
Other: Western		-0.349	0.166	0.035
Other: non-Western		-0.419	0.230	0.069
Number of friends non-Dutch nat. origin		-0.037	0.043	0.393
Behavioral problems		0.269	0.099	0.006
Self-esteem		0.020	0.096	0.835
Number of friends nominated outside class		0.237	0.064	0.000
Number of friends nominated inside class		-0.110	0.056	0.052
Constant		0.092	0.058	0.000
Random part		Coef.	S.E.	
σ^{μ}		0.142	0.186	
ρ		0.006	0.016	
Log likelihood		-1,470.157		
Wald χ^2 (df)		308.570	(20)	
Prob. $> \chi^2$		0.000		
Level-1 observations		3833		
Level-2 observations		220		

[a] Hypothesis; [b] Delta-method standard errors; [c] Two-sided p-values.

of Facebook (given that one is an SNS member). When one moves from zero to five friends on Facebook, the average probability of being an early Facebook adopter increases by .06, given that one is a member of any SNS. Furthermore, when one moves from zero to five friends who are on Facebook, the average probability that one is an early member of Facebook and a member of Hyves increases by .25. Finally, when one moves from zero to five friends who are Facebook members, the average probability that one is an early Facebook adopter (only Facebook plus Facebook and Hyves) increases by .29 ($p < .001$, not reported in Table 4), given that one is an SNS member. In addition, when one moves from zero to five friends who are Hyves members, the average probability that one is a Hyves member only increases by .09. In contrast, we do not find that classroom friends who are Hyves members are related to membership of both Facebook and Hyves ($p > .05$). However, when we combine the categories (Hyves only plus Facebook and Hyves), the average probability that one is member of Hyves is .15 when one has five friends in class who are Hyves member ($p < .001$, not reported in Table 4).

The results also show that among SNS members, native Dutch have a lower average probability of being early adopters of only Facebook than are adolescents with Turkish (.09), Moroccan (.03, $p < .1$), other Western, (.04), and other non-Western backgrounds (.04). In addition, native Dutch have a lower average probability of being a member of both Facebook and Hyves than are adolescents with Turkish (.24) and Antillean backgrounds (.14), other Western national origins (.13), and other non-Western national origins (.16). Finally, when we combine the categories of Facebook members and members of both Hyves and Facebook, native Dutch adolescents are less likely to be members of Hyves than are adolescents with Turkish (.29) and Antillean (.16) backgrounds, other Western national origins (.16), and other non-Western national origin groups (.19). Because in all three cases we see that at least four out of six national-origin groups are more likely to be early Facebook adopters, we conclude that there is considerable evidence to support H4: Adolescents of non-Dutch origin were more likely to be early Facebook adopters.

Furthermore, we find very moderate evidence for H5: for every additional friend of non-Dutch national origin, the average probability of being an early Facebook adopter only increases by .01, whereas no significant relations were found for the other categories ($p > .05$).

Adolescents' age seems to be related to early adoption of Facebook only: an increase of 24 months in age increases the average probability of being a member of Facebook only by .05, which is a relatively small effect, given that one is a member of any SNS. Age is not correlated with membership of Facebook and Hyves, nor is it related to the categories Facebook and Facebook and Hyves combined ($p > .05$). Friends' ages are not related to Facebook membership for any of the three possible outcomes ($p > .05$). These findings only indicate very moderate support for Hypothesis 6a and 6b: age is positively related to early adoption of Facebook-only membership and best friend's age does not affect Facebook membership. Finally, we do not find convincing support that adolescents who are more popular are more likely to be Facebook members. On the contrary, we find a very small but significant relation that indicates that popularity negatively affects Facebook membership: when moving from the minimum popularity score to the maximum popularity score, the average probability of membership of Facebook decreases by .10. However, the fraction of classroom students who nominate the respondent as popular is neither related to Facebook and Hyves membership nor Facebook membership in total ($p > .05$).

2.6 Discussion and Conclusions

In 2014, Facebook is by far the most popular SNS in the Netherlands; more than 95% of adolescents aged 14-15 years are members. The process of Facebook's transition from being a new SNS in the Netherlands to achieving its current monopolistic status occurred over just a few years, commencing somewhere between 2007 and 2011. Hyves, which was at that time the most popular SNS in the Netherlands, suffered from Facebook's increasing popularity and was eventually terminated in December 2013. The aim of this study is to obtain more insight into Facebook's relatively early adopters during this unique historical period in the Netherlands, studying SNS membership of a nationally representative sample of adolescents in 2010-2011. At that time, approximately 30% of adolescents were on Facebook, whereas more than 90% were on Hyves.

In our study of the characteristics of these relatively early adopters of Facebook, we find that some conditions generally promote SNS membership. Given the abundance of studies on the (positive) consequences of membership and usage of SNSs such as Facebook, it seems imperative from a methodological perspective and the issue of sample selection bias to know which social categories were the focus of those studies.

Table 2.4: Post estimation of the multinomial logistic regression analysis. Average marginal effects (AME) of variables on specific SNS membership, conditional upon membership, are presented.

		Facebook			Facebook and Hyves		
	Hyp.[a]	AME	S.E.[b, c]	p[d]	AME	S.E.	p
Number of friends in class on Facebook	H3a. +	0.012	0.004	0.002	0.049	0.010	0.000
National origin							
Dutch		(ref.)	(ref.)	(ref.)	(ref.)	(ref.)	(ref.)
Turkish	H4. +	0.086	0.028	0.002	0.242	0.052	0.000
Moroccan	H4. +	0.032	0.020	0.109	-0.020	0.051	0.686
Surinamese	H4. +	-0.011	0.009	0.250	-0.051	0.048	0.288
Antillean	H4. +	0.026	0.022	0.246	0.135	0.067	0.042
Other: Western	H4. +	0.043	0.016	0.007	0.125	0.031	0.000
Other: non-Western	H4. +	0.041	0.017	0.017	0.163	0.045	0.000
Number of friends non-Dutch nat. origin	H5a. +	0.009	0.003	0.007	-0.001	0.008	0.865
Age (in months)	H6. +	0.002	0.001	0.002	0.000	0.001	0.924
Age of friends	H6. +	0.002	0.002	0.380	0.001	0.006	0.902
Indegree: popularity nominations in class	H7. +	-0.001	0.000	0.014	0.000	0.001	0.467
		Hyves			Facebook and Hyves		
Number of friends in class on Hyves	H3a. +	0.0182	0.009	0.051	-0.005	0.009	0.588
Log likelihood	-3,643.466						
Wald χ^2 (df)	1,240.610	(72)					
Prob. $> \chi^2$	0.000						
Pseudo R^2	0.100						
Observations	3696						

[a] Hypothesis; [b] Delta-method standard errors; [c] Cluster corrected for 221 classes; [d] Two-sided p-values.

How selective were the groups of users studied in those time contexts? Adolescents who are more socially active and do many things in their free time are more likely to be SNS members. We find that exposure to digital resources, such as having a computer or smartphone, is associated with SNS membership. Finally, adolescents are more likely to be SNS members when their classroom friends are also members, presumably due to peer influence processes (Brechwald and Prinstein, 2011). Thus, in 2010-2011, we investigated the selectivity of the group of SNS members.

We identified a second set of conditions that specifically promoted the early adoption of Facebook. For one group of adolescents — namely, those of non-Dutch origin — Facebook had an important advantage over Hyves: Facebook is international, whereas Hyves is Dutch. This advantage is important because many adolescents in Europe who are of immigrant origin have transnational ties (Schimmer and Van Tubergen, 2014). For adolescents of a non-Dutch background, communication with friends and relatives in the country of origin (of their parents) might have made Facebook attractive. This might be the reason that adolescents of non-Dutch national origin adopted Facebook earlier than did native Dutch adolescents.

At the same time, our study shows that social diffusion plays an important role: when classroom friends join Facebook, the likelihood of using Facebook increases sharply. Thus, when classroom friends belonged to the first 10% of Facebook users in the Netherlands, they might have affected their friends, and so forth, which possibly led to a cascade of Facebook joiners. We show that these social diffusion processes played an important role in early Facebook adoption, in line with what Hargittai (2008) and Hargittai and Litt (2011) suggested.

Differences in adoption of new SNSs among social groups can be a source of inequality. Non-natives might experience less of the positive effects of SNS membership on well-being and social capital (Ellison et al., 2007; Steinfield et al., 2008) because they are less often members of SNSs.

We framed our hypothesis on the influence of the social environment as peer influence, but to convincingly sustain causal inferences on peer influence, dynamic social network data are needed to separate influence from selection effects (Steglich et al., 2010). In our context, selection entails that individuals select their friends based on their SNS memberships, which generates a correlation between friends' memberships that may resemble influence. Altogether, to convincingly sustain causal claims, future research should use longitudinal data to study which individual characteristics determine SNS membership.

We could not study the dynamics of joining and leaving an SNS, nor study how active people are on an SNS. Follow-up research is encouraged to address these questions about dynamics. Although Facebook remains the dominant SNS in the Netherlands, its popularity might be diminishing (see Figure 2.1), and new SNSs may take over in coming years. A question for future research would be to study early leavers of Facebook: Exactly when and why do adolescents substitute Facebook for different platforms?

Chapter 3

Understanding the Privacy Behavior of Adolescents on Facebook: The Role of Peers, Popularity, and Trust[1]

Abstract: *We study whether peer influence processes, popularity, and trust predict privacy settings on Facebook. We use large-scale survey data from 3,434 Dutch adolescents combined with observed privacy behavior on Facebook. The findings show that peer influence processes play a role and that adolescents imitate the privacy settings of their peers in the classroom. Such imitation processes are particularly pronounced in highly connected classrooms. The results show that more popular adolescents in the classroom are more likely to publicly display their Facebook profiles. Furthermore, we find that low-trust groups (ethnic minorities, lower educated and younger adolescents, and girls) more frequently opt for private Facebook profiles.*

[1] A slightly different version of this chapter is published as: Hofstra, B., Corten, R., and Van Tubergen, F. (2016). Understanding the Privacy Behavior of Adolescents on Facebook: The Role of Peers, Popularity and Trust. *Computers in Human Behavior, 60*, 611-621. Hofstra wrote the main part of the manuscript, coordinated the collection of the Facebook data, and conducted the analyses. Corten and Van Tubergen substantially contributed to the manuscript. The authors jointly developed the idea and design of the study. I thank Jeroen Weesie for his advice on methodological issues and thank Jesper Rözer, Yassine Khoudja, and Wouter Quite for helpful discussions and advice on earlier drafts of this manuscript. I also thank participants in the University of Michigan Social Media Research Lab in Ann Arbor, in the online network session at the "Dag van de Sociologie" in Amsterdam, and in the online network session at the "XXXV SUNBELT" in Brighton for their feedback on this manuscript. Finally, I acknowledge the contribution of two anonymous reviewers of *Computers in Human Behavior* for their feedback on an earlier draft of this article.

3.1 Introduction

Online social media are increasingly used for the maintenance of interpersonal relations (boyd and Ellison, 2007). In early 2015, more than one billion people were members of *social networking sites* (SNSs), continuously producing terabytes of information on these platforms (Litt, 2013). This information consists of textual status updates about emotions, opinions and experiences, uploaded photos, videos and music and other highly personal content, which is typically uploaded to the personal SNS profiles of users by users.

Inherent to the unprecedented rise of SNSs is that highly personal content is more easily accessible to an increasingly expanding audience than ever before. Consequently, an unintended byproduct of sharing such personal content has thrived. As a result of sharing photos, hometowns, e-mail addresses, phone numbers, education and employment statuses on SNS profiles, SNSs are highly targeted by hackers (Wu et al., 2014), which makes it relatively easy to commit identity theft (Javaro and Jasinski, 2014). This type of theft can cause financial damage and huge personal trauma, for instance, by utilizing personal information to obtain access to credit cards and utility services, make false claims for medical services under stolen social security numbers (Acquisti and Gross, 2009), and evade law enforcement by masquerading under others' credentials (Javaro and Jasinski, 2014).

Therefore, publicly displaying content on SNSs can cause unwanted exposure to third parties, loss of reputation, and loss of (job) opportunities (Lewis et al., 2008a). Although most of these consequences are difficult to estimate, users of SNSs must decide on the use of the tools provided by SNS services to ensure protection against such types of harm. Users can typically decide with *whom* to share the content that they upload to their profiles. Facebook, for instance, has a wide spectrum of privacy settings.

Given the potentially dramatic consequences of privacy decisions on SNSs, scholars are increasingly interested in privacy behavior on SNSs. Scholars who have studied privacy behavior have consistently found that women are more likely than men to maintain private rather than public SNS profiles (Acquisti and Gross, 2006; Lewis et al., 2008a; Thelwall, 2008; boyd and Hargittai, 2010; Hoy and Milne, 2010; Shin and Kang, 2016). Younger respondents more frequently maintain private SNS profiles than do older respondents (Tufekci, 2008; Litt, 2013). There also seem to be peer influence effects: those with more friends who have private profiles on Facebook are also more likely to maintain private Facebook profiles themselves

(Lewis et al., 2008a; Lewis, 2011). Those who are more active online (Lewis et al., 2008a) and who use Facebook more often (boyd and Hargittai, 2010) are more likely to maintain private Facebook profiles. In addition, having more Facebook friends (Stutzman and Kramer-Duffield, 2010) and reporting higher Internet skills (boyd and Hargittai, 2010) are related to more private Facebook profiles. Over time, users are also more likely to switch from public to private profiles on Facebook (Stutzman et al., 2013). Finally, those who are concerned with privacy (Tufekci, 2008; Litt, 2013) and who have experienced embarrassing situations on SNSs (Litt, 2013) are more likely to maintain private profiles.

We extend this growing literature both theoretically and empirically. First, we aim to understand why prior work has consistently found that women and younger people more frequently maintain private profiles. We study whether higher levels of distrust among these groups provide an explanation. Trust has previously been linked to online privacy concerns (e.g., Fogel and Nehman, 2009; Thomson et al., 2016). To fully investigate the potential role of trust, we also consider the differences in privacy settings among ethnic groups and educational level, given that prior work has suggested that there are lower levels of trust among minorities and at lower educational tracks (Simpson et al., 2007; Mewes, 2014). Therefore, we advance theory in online privacy research by unraveling some of the mechanisms that possibly underlie previous findings by specifically considering the role of trust. Are the gender and age findings in social media privacy research a result of the differences in trust within these groups? Additionally, are other well-known trust correlates — ethnic background and education — related to online privacy? Furthermore, we also study peer influence processes and elaborate the role of social networks by considering the potential effect of popularity, given that previous studies have shown that popularity and privacy are related (e.g., Christofides et al., 2012; Utz et al., 2012). The present study focuses on adolescents (16-20 years) in high school, and using sociometric information on who are friends in high school classrooms (∼23 pupils), we construct the peer status in classrooms. We provide a novel test of the hypothesis that popularity and privacy are related; previous studies have often used the "need for popularity" (i.e., those who want to be popular) (Christofides et al., 2012; Utz et al., 2012), whereas we construct actual peer popularity as judged by the respondents' peers in class. This approach motivates the main research question of this study: *To what extent are peers' privacy settings, popularity and trust related to adolescents' privacy settings on Facebook?*

We also empirically contribute to prior work. We study privacy settings among a large and diverse sample of Dutch adolescent Facebook users in 2014 (N = 3,451).

As shown in the overview of previous studies in Table 3.1 (these studies are not exhaustive but exemplary), scholars have generally used (potentially biased) self-reported survey data to study privacy settings (e.g., boyd and Hargittai, 2010; Stutzman and Kramer-Duffield, 2010), and almost all previous studies examine privacy behavior on SNSs with convenience samples of US college students (e.g., Lewis et al., 2008a; Tufekci, 2008; Hoy and Milne, 2010).[2] We interpret the term "convenience sample" in the strict statistical sense; that is, instead of a random sample, an easily accessible sample has been used to draw a sample from a target population. Some of the target populations of the studies in Table 3.1 resemble our target population (i.e., adolescents/ young adults), but most of these studies do not use random sampling. These issues make it difficult to convincingly make generalizable claims on the factors that affect privacy behavior on SNSs. We use stratified sampling to significantly improve the generalizability of our results to the target population of adolescents. Uniquely, we use large-scale survey data (which measure social networks, popularity, and trust) and link these measures with observed behavioral data on privacy settings on Facebook. This specific design has been recommended by Tufekci (2014), and we follow this research path. An additional benefit of such behavioral instead of self-reported privacy measures is that they are less prone to underreporting or acquiescence biases (Kuru and Pasek, 2016). Therefore, this study is one of the first attempts to link large-scale survey data — a step forward with respect to the previously used small convenience samples — with the observed privacy behavior of adolescents on Facebook.

Facebook was the most popular SNS in the Netherlands among adolescents in 2014 (Hofstra et al., 2016a) and especially adolescents who are highly engaged in SNSs (Corten, 2012). In 2014, Facebook users could choose from a wide spectrum of privacy settings, providing users with several options. For instance, users can choose to hide status updates from specific Facebook friends. Facebook has a long history of changing the tools with which users can decide how public or private their profiles are (see boyd and Hargittai, 2010), and at each point that Facebook has introduced new privacy tools, the default for Facebook users has been set to "share publicly" (boyd and Hargittai, 2010). In each case, the privacy setting changes by Facebook users to a more private profile have been a deliberate

[2] Our rationale behind choosing these studies is as follows. First, we show what the standard research practices are in this line of research by choosing *five highly cited papers* on privacy in social media (i.e., Acquisti and Gross, 2006; boyd and Hargittai, 2010; Lewis et al., 2008a; Thelwall, 2008; Tufekci, 2008;). Second, we show the research practices of *five relatively recent papers* in this area (i.e., Hoy and Milne, 2010; Lewis, 2011; Litt, 2013; Stutzman and Kramer-Duffield, 2010; Stutzman et al., 2013).

choice. Many studies examine Facebook privacy in light of whether profiles are visible (e.g., Lewis et al., 2008a). Our data allow examining privacy settings more specifically, that is, whether one's friends are publicly visible to everyone (yes/no) and whether one's status updates are publicly visible to everyone (yes/no).

3.2 Theory and Hypotheses

3.2.1 Peers' Privacy and Peer Status

Following Lewis et al. (2008) and Lewis (2011), we propose that peer influence processes play an important role in adolescents' decisions to maintain private profiles on Facebook. Behavior depends heavily on the behavior of those with whom one is associated, and in particular, adolescents look to their peers to determine which behaviors are appropriate (Brechwald and Prinstein, 2011). Adolescents seek to gain social approval from their peers and avoid social exclusion by following *norms* within their groups (Cialdini and Goldstein, 2004). We assume that groups of adolescents hold norms regarding the sharing of information on SNSs and that adolescents conform to these norms; privacy concerns may be more prevalent in one group than in others. If there is a norm within a group that one should be more careful when publicly displaying SNS profile information, then one may be influenced by this norm and choose to maintain a more private profile. A person in a group that is particularly concerned with privacy may be scolded after uploading party pictures of other group members and may be told that he or she must either delete the pictures or set his profile to a more private setting so that not everyone can see the pictures. Hence, due to the actions of one individual, others experience negative externalities and implement a group sanction (Coleman, 1990): scolding or social exclusion when this person does not adhere to the Facebook privacy norm.

Table 3.1: Previous studies on privacy behavior on SNSs (non-exhaustive).

Author(s)	Aim	Data	W[a]	N	Sampling	Method	Dependent	Predictors	Rel.[b]
The five studies below are highly cited studies on privacy in social media									
Acquisti and Gross (2006)	Investigate why people show more or less information	Survey	1	294	Convenience	Correlations Difference tests Regressions	Privacy concerns	Female	+
Tufekci (2008)	Examine audience concerns, privacy worries, disclosure, and audience management behaviors	Survey	1	704	Convenience	Regressions Difference tests	Private profile	Female Age Unwanted exposure	+ - +
Lewis et al. (2008)	Analyze which factors predict privacy	Behavior	1	1,710	Convenience	Regressions	Private profile	Female Online activity Friends private Roommates private	+ + + +
Thelwall (2008)	Identify online behaviors that relate to age, network size and gender	Behavior	1	15,043	Semi-random	Difference tests	Private profile	Female	+
boyd and Hargittai (2010)	Examine how privacy practices change over time	Survey	2	495	Convenience	Difference tests	Privacy tool use	Female	+
The five studies below are relatively recent studies on privacy in social media									
Hoy and Milne (2010)	Examine gender differences in online privacy and use of personal information	Survey	1	589	Snowball	Correlations Difference tests Factor analyses	Untagging photos Selective friending Privacy tool use	Female	+
Stutzman and Kramer-Duffield (2010)	Explore the behavior of setting privacy settings to friends-only	Survey	1	494	Convenience	Regressions	Private profile	Female # Facebook friends	+ +
Lewis (2011)	Analyze the co-evolution of friendships and privacy	Behavior	>2	876	Convenience	RSiena	Private profile	Peer influence	+
Litt (2013)	Examine predictors of privacy tool use on social network sites	Survey	1	490	Semi-random	Regressions	Privacy tool use	Female Age Embarrassment	+ - +
Stutzman et al. (2013)	Understand how privacy and disclosure changed between 2005 and 2011	Behavior	>2	5,076	Convenience	Difference tests	Private profile	Time	+

[a] W = number of waves of data used; [b] Direction of associations found, + means positive association found, and – means negative association found

Schools constitute a particularly attractive study context for peer influence processes, given that they consist of clearly defined social contexts (Corten and Knecht, 2013). We distinguish among friends in class and other classmates in the classroom and assume that correlations between friends' and classmates' privacy settings and respondents' privacy settings may be found in both types of peers. Thus, we propose the following:

> **Hypothesis 1a:** *Adolescents who have a larger proportion of friends in the classroom who maintain private (open) profiles are more likely to maintain private (open) Facebook profiles.*

> **Hypothesis 1b:** *Adolescents who have a larger proportion of classmates who maintain private (open) profiles are more likely to maintain private (open) Facebook profiles.*

In addition, we hypothesize that the classroom norm for privacy on Facebook spreads more easily through class networks when more adolescents are friends with one another. That is, when more adolescents nominate each other as friends in the classroom, the fraction of classmates who maintain a private Facebook profile may be more influential in determining students' privacy settings. Previous research has shown that the density of classrooms (i.e., the fraction of classroom friends who nominate each other as friends) has a moderating effect on peer influence processes (e.g., Corten and Knecht, 2013). The reason may be, in denser classrooms, more pupils interact and, therefore, the initial propensity for a certain behavior will spread more easily through the network. Additionally, in denser classrooms, knowledge on the behavior of others spreads through the network more easily. Therefore, classroom peers can more quickly implement a group sanction when one deviates from a norm. We believe that this is also the case for the classroom norm of privacy behavior, and we hypothesize the following:

> **Hypothesis 2:** *The association between having more classmates who have private (open) Facebook profiles and maintaining a private (open) Facebook profile strengthens as the density of the classroom network increases.*

We elaborate the role of peers in the classroom and consider the potential role of peer status. Maintaining a non-private Facebook profile may be driven by the need for distinction, in essence, to differentiate oneself from other peers and to maintain status among one's peers. Research suggests that younger rather than older generations consider post-material values to be more important than material values (Inglehart and Abramson, 1994). Younger people seek means to express themselves, they want to have jobs in which they can be creative, and they value self-expression over high income. Relatedly, younger people increasingly display more narcissism than do older people (Twenge et al., 2008). We assume that self-expression is particularly strong among the most popular peers in the classroom — i.e., those who are recognized by their classmates as popular. They impress their less popular peers by breaking conventional norms, for instance, by using drugs, consuming alcohol, and showing off (Dijkstra et al., 2009). Popular adolescents display behaviors that are related to higher status and coolness (Brechwald and Prinstein, 2011). We argue that popular adolescents want to show how cool they are to as many others as possible as a tool for self-expression and to maintain their status among peers. Maintaining a public Facebook profile is a particularly good way of showing status because other peers can see the distinctive friendship choices and/or texts that one uploads when he or she publicly shows a profile. Previous studies also established a correlation between a "need for popularity" (i.e., those who *want* to be popular) and online privacy (e.g., Christofides et al., 2012; Utz et al., 2012). Thus, we propose the following:

> **Hypothesis 3:** *More popular adolescents are more likely than less popular adolescents to maintain public Facebook profiles.*

3.2.2 Generalized Trust

We also develop hypotheses concerning the role of trust in privacy behavior. Because of the potentially damaging outcomes of displaying personal information online, privacy decisions may be based on trust and expectations about the misuse of personal information disclosed online.

Various definitions of trust can be found in the literature, but a widely used trust concept is that of *generalized trust*, which can be defined as a set of "socially learned and socially confirmed expectations that people have of each other, of the organizations and institutions in which they live, and of the natural and moral social orders that set the fundamental understandings for their lives" (cf. Barber, 1983).

This definition captures that trustors form an estimate of the trustworthiness of the average person (Paxton, 2007).

Facebook privacy can be related to generalized trust. A private Facebook profile implies the adjustment of some privacy settings, which means that one closes his/her profile to the *general* audience on Facebook, possibly indicating that one generally thinks he or she "can't be too careful dealing with people." This concept relates to the trustor's estimate of the average person; Facebook users (trustors) make an assessment of the trustworthiness of generalized others.

Research shows that individual differences in generalized trust are relatively stable (Glaeser et al., 2000). Furthermore, the findings show the following: men are more likely than women to have generalized trust (Alesina and La Ferrara, 2002; Mewes, 2014); foreign-born persons and minority members are less likely than native-born people and majority members to have generalized trust (Glaeser et al., 2000; Simpson et al., 2007); lower educated people are less likely than higher educated people to trust (Mewes, 2014); and older individuals are more likely than younger individuals (Mewes, 2014) to have generalized trust.

Women are generally found to be more likely than men to maintain private profiles (Lewis et al., 2008a; boyd and Hargittai, 2010; Hoy and Milne, 2010). Possibly, women are generally less likely to trust others with personal information displayed on SNSs. Tufekci (2008) and Litt (2013) find that younger people are more likely to maintain more private profiles than are older individuals, which may be a result of younger people's being less trustful (Mewes, 2014). Nearly no results exist with regard to ethnic background and its relationship to privacy behavior on SNSs. However, previous studies have shown that those from non-native ethnic backgrounds are less likely to trust (Glaeser et al., 2000; Alesina and La Ferrara, 2002; Simpson et al., 2007) and that those from non-Western countries of origin (e.g., low-trust societies such as Turkey) display significantly lower levels of trust than do those from Western countries (Delhey and Newton, 2005). Following these earlier results, we expect the same for Facebook privacy. Thus, we propose the following:

> **Hypothesis 4a:** *Girls are more likely to maintain private Facebook profiles than boys.*
> **Hypothesis 4b:** *Adolescents with a non-Dutch ethnic background are more likely to maintain private Facebook profiles than those with a Dutch background.*

Hypothesis 4c: *Adolescents who are lower educated are more likely to maintain private Facebook profiles than those who are higher educated.*
Hypothesis 4d: *Younger adolescents are more likely to maintain private Facebook profiles than older adolescents.*

We aim to ascertain that the associations between gender, national origin, education, age and privacy run via the trust mechanism. Therefore, we propose the following:

Hypothesis 5a: *The relationship between gender and maintaining a private Facebook profile is (at least partially) mediated by generalized trust.*
Hypothesis 5b: *The relationship between ethnic background and maintaining a private Facebook profile is (at least partially) mediated by generalized trust.*
Hypothesis 5c: *The relationship between educational level and maintaining a private Facebook profile is (at least partially) mediated by generalized trust.*
Hypothesis 5d: *The relationship between age and maintaining a private Facebook profile is (at least partially) mediated by generalized trust.*

Finally, Figure 3.1 summarizes our hypotheses and the predicted associations with privacy on Facebook.

3.3 Data

We use survey data on adolescents in the Netherlands originating from the larger project entitled "Children of Immigrants Longitudinal Study in Four European Countries" (CILS4EU) (Kalter et al., 2013; Kalter et al., 2015).[3] CILS4EU followed adolescents 14-15 years of age (third-year high school pupils in the Netherlands), with an oversampling of immigrant minority youth, for three subsequent years beginning in 2010. Each year, the survey was repeated, for a large portion

[3] The data were collected in Germany, the Netherlands, Sweden and England.

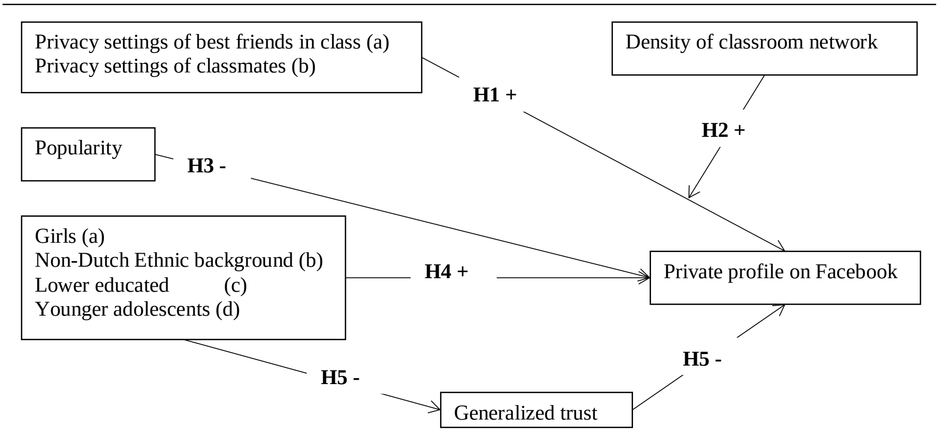


Figure 3.1: Conceptual model for the hypotheses as derived from the theory; + = positive effect hypothesized, - = negative effect hypothesized.

with the same questions. The surveys consist of self-completion questionnaires concerning many individual characteristics, attitudes and leisure time activities. The data include information on friends within classrooms. Data collection occurred at high schools.

In wave 1 (2010-2011), high schools were selected according to four strata based on educational track levels and the percentage of non-Western immigrant students in schools (controlling for strata as dummy variables does not change the results of this article). In wave 1, two classes were randomly selected per school, resulting in a total of 118 schools, 252 classes and 4,963 students who participated in the survey in the Netherlands.[4] Class composition changes between the third and fourth years are common in the Netherlands. Hence, respondents in wave 2 (2011-2012) could be scattered among multiple fourth-year classes that did not participate in wave 1. To interview as many wave 1 respondents as possible, schools were asked to provide more than the two classes initially sampled in wave 1 *if* the respondents from wave 1 were in classes different from those sampled previously. Consequently,

[4]In the first wave, N = 600 respondents who were not a part of the original sampling frame were sampled because some schools wanted to participate with more than two classes. A random sample of N = 4,363 was established in wave 1. We include as many respondents as possible in the sample for analyses, including newcomers (non-random) and the non-random sample of wave 1, to ensure a large sample size.

additional students were interviewed: 3,803 participants who participated in wave 1 were surveyed again in wave 2 (76.6%), and an additional 2,118 new respondents were surveyed in wave 2 (W2 N = 5,921).

3.3.1 Dutch Facebook Survey

The Dutch Facebook Survey (DFS) data (Hofstra et al., 2015) were collected to enrich the Dutch part of the CILS4EU survey and consist of observational data from Facebook. The data were collected between June 2014 and September 2014. In waves 3 (2012-2013) and 4 (2013-2014) of the CILS4EU survey, participants were asked about their membership in Facebook.[5] In waves 3 and 4 combined, N = 4,864 respondents indicated being a member of Facebook in at least one of these waves (W3=3,423, W4=3,595). For the project, coding assistants tracked down respondents' profiles based on the respondents' names, cities of residence and, if reported in the survey, the URLs of their Facebook profiles. This procedure occurred *after* wave 4 for the respondents who indicated being a member in wave 3 or 4. The coding assistants were personally instructed, and all followed strict coding procedures; N = 4,463 (91.8%) of the profiles were tracked.[6] Based on the tracked profiles, the privacy settings were coded — whether friend lists were publicly visible and whether timeline posts were publicly visible. We have linked the DFS with wave 2 of the CILS4EU (from which our independent variables are constructed), which is the latest licensed version of the CILS4EU and contains the latest classroom sociometric data. Of the 4,463 who were tracked in the DFS, 3,864 participated in wave 2 of the CILS4EU, which is the maximum number of observations we can analyze. Figure 3.2 graphically displays the number of observations from waves 1 and 2 and how we arrive at N = 3,864.

The collected information was publicly visible on Facebook, and we followed a strict procedure with password-protected files. All personal identifiers were removed from the data. The data collection, the coding procedure and the use of these data for scientific purposes were reviewed and approved by an internal review board.

[5] As of January 2015, two additional waves of data were collected: waves 3 and 4.

[6] We cannot distinguish profiles that were not tracked because of privacy settings that were too strict or profiles that we cannot track due to wrong or incomplete information.

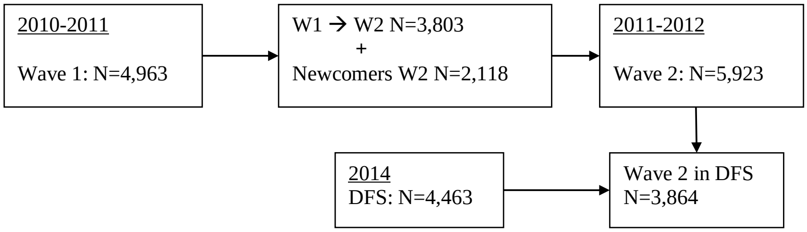


Figure 3.2: Attrition rates and maximum number of observations in the analyses.

3.4 Methods

3.4.1 Measurement of Privacy Behavior on Facebook

Based on the observational data obtained in the DFS, we code two variables that indicate whether one's Facebook profile settings are private. First, we code whether one's *timeline posts* are private (1) or not (0). Second, we measure whether one's *friend list* is private (1) or not (0). We do not distinguish between the privacy settings *visible to friends* or *visible to friends of friends* or any other custom settings chosen by the respondents on Facebook. For the timeline posts measure, the coding assistants unfold one's complete timeline and code whether *at least one* status update is publicly visible. It may be that a respondent posted publicly in 2012 but no longer posted publicly in 2013 and 2014; these respondents are coded as having a public timeline. With our data, we cannot distinguish between such cases. We capture whether one's timeline posts or relationships are visible to non-friends *in general*. We also code seven variables that indicate whether adolescents choose to disclose personal information on their Facebook profiles' information pages; whether respondents display their family, gender, relationship status, romantic interests, hometown, secondary education, and work. We also run all of our analyses described with these privacy decisions, and we do not find qualitatively different results compared with the two privacy settings described above. Table 3.2 shows that 54.7% of the respondents maintain private timeline posts and that 25.1% maintain a private friend lists on Facebook.

Table 3.2: Descriptive statistics of privacy behaviors on Facebook.

	Observations	Percentage
Timeline posts private: Yes (1)	2,437	54.65
Timeline posts private: No (0)	2,022	45.35
Timeline posts private: Total	4,459	100
Friend list private: Yes (1)	1,119	25.07
Friend list private: No (0)	3,344	74.93
Friend list private: Total	4,463	100

3.4.2 Independent Variables

We operationalize our independent variables by using wave 2 of the CILS4EU data because this wave is the most recent version of the licensed CILS4EU data and because the most recent classroom sociometric data are available in this wave of data.

Privacy Settings of Best Friends in Class. The privacy settings of best friends in class are two variables that capture the percentage of classroom friends who have their *timeline posts* and the percentage of classroom friends who have their *friend lists* private. The respondents answered the question, "Who are your best friends in the class (you can write down a maximum of 5 friends)?" We know from these friends (see the measurement of the privacy variables) whether they maintain private friend lists and timeline posts on Facebook. We count the absolute number of classroom friends who maintain a private friend list on Facebook within the respondent's friends, ranging from 0 to 5, divide this amount by the number of classroom friends indicated for this question (also ranging from 0 to 5) and multiply this number by 100. We constructed a similar measure for the percentage of classroom best friends who maintain private timeline posts. Because we limit our respondents to a maximum of five friends in class, we may not capture all friends. However, 64% of the respondents indicate fewer than five friends, and the average number of friends is 3.6. Therefore, in most cases, we have captured all of the friends of the respondents in the classroom.

Privacy Settings of Classmates. For associations between the privacy settings of non-direct classroom friends and respondents, we construct two variables that indicate the *rest* of the class's privacy preferences on Facebook. We measure the percentage of the class that maintains private posts and the percentage of the class that maintains private friend lists, excluding the privacy preferences of the best friends in the class indicated and excluding the respondents' privacy preferences.

Density. We construct a variable that measures the density of class networks, which is defined as:

$$Density_g = \frac{\sum_i X_{ij}}{5N}, \tag{3.1}$$

where $Density_g$ is the density of classroom g, i is the pupil, X_{ij} is a binary variable that indicates whether a relationship exists between pupil i and pupil j, and N is the total number of pupils in the class. We multiply N by five because respondents could maximally nominate five friends (Valente, 2010). Hence, density is the fraction of the number of ties that could have been realized in the class.

Indegree: Popularity in class. We construct a variable to measure respondents' popularity. In the second wave, respondents answered the question, "Who are the most popular students in the class (you can write down a maximum of 5 names)?" The pupils were not allowed to define themselves as popular. We acquire a comparable popularity measure between classes, which is defined as:

$$Popularity_i = \frac{\sum_i K_{ji}}{N-1} \times 100, \tag{3.2}$$

where $Popularity_i$ is the indegree popularity of pupil i, K_{ji} indicates whether pupil j nominates pupil i as popular and N is the total number of pupils in the classroom. Hence, we acquire a standardized variable between classrooms that indicates what percentage of classroom pupils in a given class indicates the respondent as the most popular pupil (Wasserman and Faust, 1994), and this measure shows sufficient discriminant validity from other dimensions of peer status (Dijkstra et al., 2010).

Gender. We measure whether the respondent is a girl (1) or a boy (0).

Ethnic Background. This variable indicates the respondents' ethnic background within one of the six largest ethnic background groups in the Netherlands: 1 "Native Dutch," 2 "Turkish," 3 "Moroccan," 4 "Dutch Caribbean," 5 "Other: Western (Europe or English speaking)" and 6 "Other: non-Western." The measure is based on the biological parents' country of birth. When the adolescent has only one native-born Dutch parent, he or she is classified as having the ethnic background of the foreign-born parent. When children have parents from different countries, they belong to the ethnic background of the mother, which is standard practice in research on ethnic background in the Netherlands (Statistics Netherlands, 2012).

Educational Track. We create dummy variables to indicate the adolescents' high school tracks. In the Netherlands, when adolescents transition to high school, they are classified into different educational tracks, which differ in terms of the level and type of education. These seven tracks range from "VMBO-basis" (lower preparatory vocational education) to "VWO-gymnasium" (university preparatory education). We combine these classes into three dummy variables: "preparatory vocational education (VMBO)," "senior general (HAVO)" and "university preparatory education (VWO)." We combine the four preparatory vocational educational tracks into one category and combine the two levels of university preparatory education into one category because the differences between these educational tracks are not substantially large. We perform robustness analyses in which we separate the highest two levels of vocational education (Dutch: VMBO-T and VMBO-GT) from the lowest two levels (Dutch: VMBO-basis and VMBO-kader). These results are in line with the results presented in the article.[7]

Age in months. This variable measures the age of respondents in months, calculated as the difference in months between the respondent's date of birth and the date when the respondent's privacy variables were obtained.

Trust. We measure generalized trust with the following standard question, which is asked in many other surveys (e.g., GSS, ESS; Nannestad, 2008): "Generally speaking, would you say that most people can be trusted (1) or that you can't be too careful in dealing with people? (0)." Because this measure is only available in waves 3 and 4, we take answers from wave 3, and if respondents were missing or did not participate in wave 3 but were not missing or did participate in wave 4,

[7] We performed robustness analyses in which we separated the highest two levels of vocational education (Dutch: VMBO-T and VMBO-GT) from the lowest two levels (Dutch: VMBO-basis and VMBO-kader). One may argue that the two highest vocational tracks are significantly different from the lowest two. For "private friend list," we found no significantly different results, and for "private timeline posts," we found that those who follow the senior general educational track were significantly less likely to maintain private timeline posts than those in the lowest two vocational tracks. Those in the university preparatory track (although marginally significant, $p = .051$) and the highest two vocational tracks were not more likely to have private timeline posts than those in the lowest two educational tracks. These results are in line with the results presented in the article.

then we take the respondents' answers from wave 4. Between waves 3 and 4, trust is relatively stable; 73% of the respondents answer equally when they answer the trust question in both waves. A score of 1 means that the respondent places trust in generalized others.

Table 3.3 shows the descriptive statistics for the independent variables for the respondents in the DFS data: 3,864 respondents participated in W2 *and* were tracked in the DFS data. This number is the maximum number of respondents who we can investigate. We show the number of respondents who have non-missing values and were tracked in the DFS data, and the % missing column indicates the percentage of missing values relative to the maximum of 3,864.

Table 3.3: Descriptive statistics for the independent variables.

	Range	Mean	SD[a]	N	missing
% Best friends' posts private	0-100	37.096	29.858	3,529	8.67%
% Best friends' friend lists private	0-100	17.056	22.818	3,529	8.67%
% Class timeline posts private	0-100	38.354	16.209	3,521	8.88%
% Class friend lists private	0-100	17.555	11.025	3,515	9.03%
Density	0.200-1	0.677	0.105	3,706	4.09%
Indegree: popularity	0-100	8.873	14.057	3,705	4.12%
Girls	0-1	0.546	-	3,719	3.75%
Ethnic background	-	-	-	3,864	0%
Dutch	0-1	0.775	-	2,996	-
Turkish	0-1	0.030	-	115	-
Moroccan	0-1	0.020	-	78	-
Dutch Caribbean	0-1	0.028	-	107	-
Other: Western	0-1	0.090	-	347	-
Other: non-Western	0-1	0.057	-	221	-
Educational track	-	-	-	3,712	3.93%
Preparatory vocational	0-1	0.486	-	1,805	-
Senior general	0-1	0.271	-	1,006	-
University preparatory	0-1	0.243	-	901	-
Age in months	201-247	223.564	7.173	3,668	5.07%
Trust	0-1	0.519	0.500	4,433	-

[a] SD = standard deviation.

3.4.3 Analytical Strategy

We perform two sets of statistical analyses to test our hypotheses. First, we estimate two logistic regression models to test whether the privacy behavior of friends and classmates (H1), popularity (H3), gender, educational level, ethnic background and age (H4) affect the tendency to have private Facebook timeline posts or a private friend list. Additionally, we interact the percentage of classmates' privacy settings and class density to test H2. Because adolescents are clustered within classes, we perform a cluster correction for 287 classes and obtain robust standard errors.

Second, we estimate two mediation models by using structural equation modeling (SEM). SEM makes it possible to simultaneously estimate models with multiple endogenous variables. To test H5, our first model estimates whether the relationships of gender, ethnic background, educational level and age with private timeline posts are (at least partially) mediated by generalized trust, and the second model analyzes these same associations with maintaining private friend lists. Generalized trust and Facebook privacy are dichotomous variables, and therefore, we use the *gsem* command in the Stata statistical software package to perform path analysis by using logistic regression (StataCorp, 2013). For both paths of H5, logistic regression is performed. We control for peers' privacy behavior and popularity and perform a correction for 287 classes.

We listwise delete the missing values of all of the variables so that we can generalize our results to the same set of 3,434 respondents, leading to an 11.1% loss of observations.

3.5 Hypotheses Tests

Table 3.4 shows the two logistic regression models for having private timeline posts and private friend lists on Facebook. We report the average marginal effects (AMEs) because they are more intuitively interpreted than are odds ratios; therefore, the effect sizes can be compared across models (Mood, 2010). The odds ratios can reflect unobserved heterogeneity and can therefore be problematic to interpret as substantive effects (cf. Mood 2010). The AMEs express how P(Y=1) changes as the predictors change: from 0 to 1 in the case of categorical or dummy variables and with a unit increase for continuous variables. The AMEs are calculated by computing a marginal effect for every observation; all of these effects are then averaged (Cameron and Trivedi, 2010).

We find evidence of the role of peers' privacy behavior in the respondents' privacy settings, but the effect sizes appear to be somewhat modest. With a two-standard-deviation increase in the percentage of classmates who have private timeline posts (32.42), the average probability of maintaining private timeline posts increases by .07, whereas classmates' private friend lists are not related to maintaining private friend lists. We find some evidence (borderline significant: one-sided p-value = .055) of the association between best friends' private timeline posts and the respondent's maintaining private timeline posts: with a two-standard-deviation increase in the percentage of friends who maintain private friend lists, the average probability of maintaining a private friend lists increases by .06. There is no significant relationship between best friends' private friend lists and respondents' private friend lists.[8]

We find evidence to support H3: popularity in class is significantly related to privacy settings. Again, the magnitude of this association is small: a two-standard-deviation increase in indegree popularity decreases the probability of maintaining private timeline posts by .08, and it decreases the probability of maintaining a private friend list by .03.

In line with H4a, we find that girls have a .07 higher probability of maintaining a private friend list on Facebook than boys have, whereas there is no significant difference in maintaining private timeline posts between girls and boys. In line with H4b, our findings show that Dutch majority adolescents have lower average probabilities of maintaining private timeline posts than do adolescents with Turkish (.11), Moroccan (.21) and Dutch Caribbean (.09) ethnic backgrounds. Additionally, native Dutch have a lower probability of maintaining private friend lists than do pupils of Turkish (.36), Moroccan (.27), Dutch Caribbean (.17), other Western (.07) and other non-Western (.20) backgrounds. Ethnic background has frequently been omitted in prior work, but our study shows that these associations, at least in the Netherlands, are rather large.

The results partly support H4c: those who are in the vocational education track in high school are slightly more likely than those in the senior general track (.04) and in the university preparatory track (.04) to maintain private timeline posts

[8] We analyzed the relationship between peers' privacy behavior and maintaining a private Hyves profile (1) or not (0) (a former Dutch SNS, see Hofstra et al., 2016a), where the time lag between the sociometric data and the privacy measure is significantly smaller, finding an AME of .002 (N = 1,029).

on Facebook. We find evidence of the role of age (H4d): adolescents who are 14.37 months older (two standard deviations) have a .10 lower average probability of maintaining private timeline posts and a .33 lower probability of maintaining private friend lists on Facebook.

We interact class peer influence and class density to test H2 (full results found in the Appendix of Chapter 3). As expected, the association between the percentage of classmates who maintain private timeline posts and the respondents' private timeline posts increases when class density increases. This model fits the data significantly better than does the model without the interaction term (LR-test: χ^2(df)=6.540(2); Prob. $> \chi^2$=.038). Figure 3.3 shows that the AME of the percentage of classmates who maintain private timeline posts on respondents' private timeline posts increases with higher density values.[9]

Table 3.5 shows the direct and indirect (via generalized trust) relationships of gender, ethnic background, educational track and age with maintaining private timeline posts and private friend lists. We once again report the AMEs, based on a self-written program, because AMEs are currently not available for *gsem* in Stata. In this program, we analytically compute the AMEs by elementary calculus and obtain standard errors by non-parametric bootstrapping (N = 1,000 bootstraps).

Girls trust less often than boys do. Adolescents from a non-native ethnic background trust less often than do adolescents from a native ethnic background. Adolescents in lower educational tracks trust less often than do those in higher tracks. Surprisingly, older adolescents trust less often than younger adolescents do.

Surprisingly, those who place trust in generalized others are significantly more likely to keep timelines private. For the indirect path coefficients, one path coefficient is multiplied by another to obtain the indirect effect. Hence, our indirect associations are in directions opposite to those we expected. For instance, the positive association between trust and private timeline posts multiplied by the negative association between a Turkish ethnic background and trust yields a negative indirect association that is significantly different from zero. Hence, the mediated associations of gender, ethnic background and educational level are in directions

[9] We estimated a logistic regression model in which we investigated having *at least one* privacy setting enabled on Facebook; an ordered logistic regression; and a linear regression model, where zero means no privacy settings enabled, one means maintaining private timeline posts or private friend lists, and two means keeping private timeline posts *and* a private friend list. In none of these analyses did we find qualitatively different results.

opposite to those we expected. Because there is a negative association between age and trust, this indirect association is in the hypothesized direction.

However, given the large number of respondents (N = 3,434), the significance of the indirect associations is somewhat unconvincing, with the smallest p-value being .026. Finally, in no case are the relations fully mediated by generalized trust.[10]

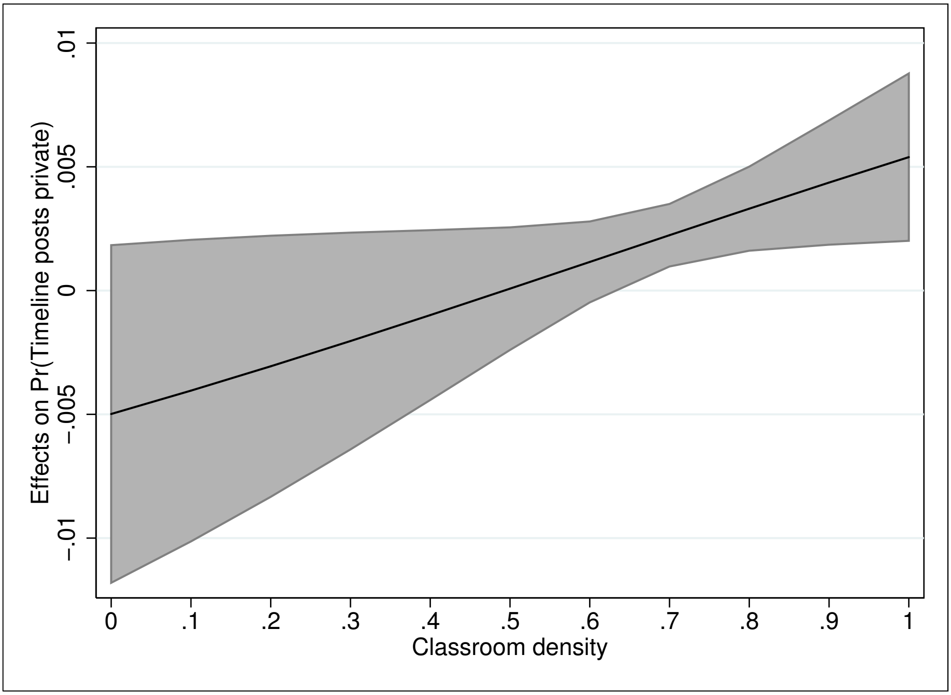


Figure 3.3: Average marginal effects of the % class timeline posts private with 95% CIs.

[10]We used the relatively new gsem command in the Stata statistical software package to estimate structural equation models with binary dependent variables. Therefore, we were not able to compute fit indices such as the χ^2, the root mean square error of approximation (RMSEA), and the comparative fit index (CFI). The existing statistical software limits us in the sense that it is currently not able to compute these fit indices.

Table 3.4: Logistic regression: associations between peers' privacy behavior, popularity, gender, ethnic background, educational level, age, and Facebook privacy. We present average marginal effects (dy/dx).

		Pr(Private timeline post)			Pr(Private friend list)		
	Hyp.[a]	dy/dx	S.E.[b]	p[c]	dy/dx	S.E.	p
% Best friends' timeline posts private	H1 +	0.001	0.000	0.110	-	-	-
% Best friends' friend lists private	H1 +	-	-	-	0.000	0.000	0.797
% Class timeline posts private	H1 +	0.002	0.001	0.005	-	-	-
% Class friend lists private	H1 +	-	-	-	0.000	0.001	0.781
Indegree: popularity	H3 -	-0.003	0.001	0.000	-0.001	0.000	0.007
Girls (ref.: boys)	H4a +	0.023	0.017	0.171	0.065	0.016	0.000
Ethnic background (ref.: Dutch)		-	-	-	-	-	-
Turkish	H4b +	0.105	0.053	0.048	0.363	0.048	0.000
Moroccan	H4b +	0.210	0.055	0.000	0.268	0.071	0.000
Dutch Caribbean	H4b +	0.086	0.052	0.097	0.168	0.051	0.001
Other Western	H4b +	-0.042	0.031	0.173	0.071	0.024	0.003
Other non-Western	H4b +	0.059	0.038	0.115	0.203	0.037	0.000
Educational track (ref.: Voc. educ.)		-	-	-	-	-	-
Senior general	H4c -	-0.043	0.019	0.026	0.021	0.019	0.260
University preparatory	H4c -	-0.040	0.022	0.063	0.033	0.023	0.159
Age in months	H4d -	-0.007	0.001	0.000	-0.023	0.002	0.000
Constant (log-odds)		6.292	1.139	0.000	29.875	2.753	0.000
N		3,434			3,434		
Wald χ^2 (df)		104.930	(12)		259.930	(12)	
Prob. > χ^2		0.000			0.000		
Log pseudolikelihood		-2,306.452			-1,658.212		
Pseudo R^2		0.024			0.147		

[a] Hypothesis; [b] Delta-method standard errors, cluster corrected for 287 classes; [c] Two-sided p-values.

Table 3.5: Structural equation models: direct and indirect associations between gender, national origin, educational level, age and Facebook privacy. We present average marginal effects (dy/dx).

	Pr(Private timeline post)			Pr(Private friend list)		
	dy/dx	S.E.[a]	p[b]	dy/dx	S.E.	p
Direct associations with privacy						
% Friends timeline posts private	0.001	0.000	0.039	-	-	-
% Friends friend lists private	-	-	-	0.000	0.000	0.760
% Class timeline posts private	0.002	0.001	0.001	-	-	-
% Class friend lists private	-	-	-	0.000	0.001	0.700
Indegree: popularity	-0.003	0.001	0.000	-0.001	0.001	0.012
Girls (ref. boys)	0.027	0.017	0.109	0.063	0.014	0.000
Ethnic background (ref. Dutch)	-	-	-	-	-	-
Turkish	0.118	0.052	0.024	0.303	0.046	0.000
Moroccan	0.240	0.070	0.001	0.228	0.060	0.000
Dutch Caribbean	0.097	0.055	0.078	0.148	0.040	0.000
Other Western	-0.040	0.030	0.183	0.069	0.022	0.002
Other non-Western	0.068	0.037	0.063	0.176	0.032	0.000
Educational track (ref. Voc.)	-	-	-	-	-	-
Senior general	-0.046	0.020	0.023	0.023	0.017	0.190
University preparatory	-0.046	0.023	0.040	0.035	0.019	0.061
Age in months	-0.007	0.001	0.000	-0.023	0.001	0.000
Trust	0.039	0.016	0.017	-0.017	0.014	0.221
	Pr(Private timeline post)			Pr(Private friend list)		
Indirect associations privacy (via generalized trust)						
Girls (ref. boys)	-0.018	0.008	0.031	0.008	0.007	0.242
Ethnic background (ref. Dutch)	-	-	-	-	-	-
Turkish	-0.044	0.021	0.038	0.019	0.017	0.258
Moroccan	-0.040	0.022	0.061	0.017	0.016	0.266
Dutch Caribbean	-0.044	0.022	0.042	0.019	0.017	0.268
Other Western	-0.004	0.006	0.430	0.002	0.003	0.561
Other non-Western	-0.036	0.017	0.030	0.016	0.014	0.250
Educational track (ref. Voc.)	-	-	-	-	-	-
Senior general	0.014	0.007	0.043	-0.006	0.005	0.258
University preparatory	0.026	0.012	0.026	-0.011	0.009	0.234
Age in months	-0.001	0.000	0.063	0.000	0.000	0.282
Observations	3,434			3,434		

[a] Delta-method standard errors, cluster corrected for 287 classes; [b] Two-sided p-values.

3.6 Discussion and Conclusions

We investigated which factors are related to the privacy settings of Dutch adolescents on Facebook in 2014. We implemented a theoretical framework consisting of peer's privacy behavior, popularity and trust. We contributed to the previous literature by simultaneously examining *multiple* privacy settings on Facebook while using unique large-scale survey data combined with behavioral data from Facebook.

In line with Lewis et al. (2008a) and Lewis (2011), we find associations between peers' privacy settings and respondents' Facebook privacy settings. However, the magnitude of these associations is somewhat small, which may be due to the large time lag between our measure of the class network (in wave 2) and the privacy settings on Facebook (measured at wave 4) — a two-year difference. Interestingly, we find that the density of the classroom friendship network moderates the associations between the influence of peers' privacy settings and the respondents' Facebook privacy settings: in more connected classes, adolescents are more likely to imitate their classmates' timeline post settings.

Why did we find no relationship between peers' and respondents' private friend lists? First, reputational damage or other negative consequences of maintaining private friend lists are not very clear for Facebook friends with public displays: for this setting, norms may be less likely to be enforced. Second, adolescents may be less likely to *know* what their peers' friend list privacy settings are. In timeline posts, with whom a post is shared is visible, whereas with friend lists, this information is not visible, which makes norm enforcement more difficult.

This study also shows that popularity, a previously omitted factor, is related to privacy settings. More popular adolescents are more likely to maintain public Facebook profiles, possibly due to a higher need for self-expression and a need to maintain their status. Popular adolescents display behaviors that are associated with higher risk (Dijkstra et al., 2009), and they are also more publicly visible on Facebook.

Our study finds (further) evidence to suggest that girls, members of ethnic minorities, pupils in lower educational tracks, and younger adolescents more frequently opt for private Facebook profiles. These results are in line with the previously found observation that these groups tend to display lower levels of trust in "most others" (Glaeser et al., 2000; Alesina and La Ferrara, 2002; Simpson et al., 2007; Mewes, 2014) and that girls and younger people also display a higher probability of maintaining private SNS profiles (Lewis et al., 2008a; boyd and Hargittai, 2010; Tufekci, 2008). In particular, we find that the differences across ethnic backgrounds are strong. Those who have an ethnic background from low-trust societies, such as those with a Turkish background (Delhey and Newton, 2005), especially display more privacy on Facebook than do native Dutch. Surprisingly, however, our mediation models do not show that the associations of gender, ethnic background, educational track and age with privacy settings are convincingly mediated by generalized trust. Contrary to our expectations, we find that those who place trust are more likely to maintain private timeline posts, possibly for two reasons. First, one may close his/her profile while generally trusting others because the actual content posted on timelines is much more sensitive. Second, the measure adopted in our study does not fully capture the more complex concept of trust. More refined measures of trust are needed in further research also because not everyone interprets “most others” in the same manner (Delhey et al., 2011). One may even speculate that a public Facebook profile that is visible to the general public is an alternative behavioral measure of self-reported trust in “most others.”

There are four limitations that warrant acknowledgement, and they pertain to the data that we used. First, our study must be replicated by using an even more representative sample. Second, to substantiate the causal inferences on peer influence, we need dynamic data on social networks and behavior to separate influence from selection effects (Steglich et al., 2010). Third, we did not study whether adolescents oscillate between settings, nor did we study the level of customization of privacy on Facebook. Further research could investigate these dynamics. Nevertheless, we analyzed far more privacy decisions (e.g., romantic interests) than presented here, and these results did not qualitatively differ from the results presented in the article. We went beyond previous studies’ findings regarding social media privacy. Finally, we restricted the respondents to a maximum of five best friends in the survey. Therefore, for a small proportion of our respondents, *indirect* friends may be included in the classmates measure. Nevertheless, research shows that, in such questions, respondents indicate their very best friends first (Marsden, 2011), and we found evidence for this phenomenon in our data; 64% of the respondents indicated less than five friends. Furthermore, this limitation did not affect our theoretical intuitions or our conclusions — we expected correlation in peers’ privacy behaviors, whether from friends or indirect friends among classmates.

SNSs are extremely volatile in terms of their popularity (see Hofstra et al., 2016a), and the privacy tools provided to users by SNS service providers frequently change. Therefore, ongoing research is needed to study the factors that predict (distinct) privacy settings on SNSs.

Part II

Structure of Online Social Networks

Chapter 4

Sources of Segregation in Social Networks: A Novel Approach Using Facebook[1]

Abstract: *Most research on segregation in social networks considers small circles of strong ties, and little is known about segregation among the much larger number of weaker ties. This article proposes a novel approach to the study of these more extended networks, through the use of data on personal ties in an online social network. We illustrate this method's potential by describing and explaining the degree of ethnic and gender segregation on Facebook among a representative survey of adolescents in the Netherlands (N = 2,810; ∼1.1 million Facebook friends). The results show that large online networks are more strongly segregated by ethnicity than by gender. Drawing on the same survey data, we find that core networks are more segregated in terms of ethnicity and gender than are extended networks. However, an exception to this pattern is personal networks of ethnic majority members, whose core networks are as segregated by ethnicity as their extended networks. Further analysis suggests this exception is due to their larger population size and the ethnic segregation of their social settings. We discuss the implications of these findings for the role of structural opportunities, homophily, and balance.*

[1]A slightly different version of this chapter is published as: Hofstra, B., Corten, R., Van Tubergen, F., and Ellison, N.B. (2017). Sources of Segregation in Social Networks: A Novel Approach Using Facebook. *American Sociological Review, 82*(3), 625-656. Hofstra wrote the main part of the manuscript, coordinated the collection of the Facebook data, and conducted the analyses. Corten, Van Tubergen, and Ellison substantially contributed to the manuscript. The authors jointly developed the idea and design of the study. I thank Manja Coopmans, Maaike van der Vleuten, Jesper Rözer, Joseph B. Bayer, Niek de Schipper, Müge Simsek, Bernie Hogan, René Veenstra, and Jeroen Weesie for advice on earlier drafts. This article benefited from discussions at the "Migration and Social Stratification seminar" in Utrecht, the "CILS4EU conference" in Stockholm, the "XXXVI SUNBELT" in Newport Beach, the "CITP/MILAB workshop" in Vienna, the "INAS 2016" in Utrecht, the "Dag van de Sociologie" in Tilburg, and the "Dutch Demography Day 2016" in Utrecht. Finally, I acknowledge the feedback of five anonymous reviewers as well as the insights of the *American Sociological Review* editors.

4.1 Introduction

One of the most consistent findings in sociological research is that strong-tie, core friendship networks tend to be homogeneously sorted (McPherson, Smith-Lovin, and Cook, 2001). Network cleavages among strong ties are formed along ethnic, gender, religious, and social status lines. This finding appears in research on romantic relationships (Kalmijn, 1998; Feliciano et al., 2009; Lewis, 2013; Anderson et al., 2014; Potârcă and Mills, 2015), core discussion networks (Marsden, 1988; Smith, et al., 2014a), and personal friendship networks (Mouw and Entwisle, 2006; Vermeij et al., 2009; Currarini et al., 2010; Wimmer and Lewis, 2010; Smith et al., 2014b).

In contrast to the abundant literature on the segregation of core networks, little is known about the segregation of weaker ties, such as those involving colleagues, neighbors, and acquaintances (Moody, 2001; DiPrete et al., 2011). For various reasons, however, it is particularly important to study segregation among weaker ties, as they relate to a myriad of sociologically relevant issues. A classic argument is that such weak ties provide novel information on job openings and hence link to labor-market outcomes and the societal distribution of wealth (Granovetter, 1973, 1983; Lin, 1999). Second, when not only core networks but also people's extended networks are homogenous — whether in terms of ethnicity, race, religion, gender, or other characteristics — intergroup trust might be undermined (Gambetta, 1988; Fukuyama, 1995), and negative intergroup attitudes may prevail, as they are not challenged by personal encounters (Allport, 1954). A rich body of literature suggests that even superficial contact (i.e., weak ties) between members of different ethnic groups has the potential to reduce intergroup prejudice (Pettigrew and Tropp, 2006). Furthermore, the co-evolution of homophilous network selection and social influence can result in "echo chambers" — in which people are increasingly surrounded by like-minded people (Halberstam and Knight, 2016) — and intergroup polarization of opinions and attitudes (Mäs and Flache, 2013).

The lack of research on diversity among weaker ties is mainly due to methodological difficulties of gathering information that includes both strong and weak ties. To our knowledge, the only study on segregation among weak ties is by DiPrete and colleagues (2011). Using the 2006 General Social Survey (GSS), they found that Americans' "acquaintanceship" networks (i.e., weak ties) are approximately as segregated as their "trust" networks (i.e., core ties) along racial, political, and religious lines. They measured weak ties using network scale-up methods, in which respondents were asked to estimate the number of people with whom they are acquainted along racial, ideological, and religious lines in various contexts (e.g., neighborhoods). This method suffers from two limitations. First, although people are asked to think about their acquaintances in specific contexts, there may be selectivity in network recall ability (Brashears et al., 2016). Second, the results may be affected by social desirability biases or "misperception or masking of behaviors and opinions that Americans think would be disapproved of by their associates" (DiPrete et al., 2011: 1272).

We propose that the study of online social networks provides new opportunities to examine the segregation of large personal networks, including networks with stronger and weaker ties. An important advantage over the scale-up method is that online networks map networking behavior up to potentially thousands of contacts, without restrictions to specific contexts. As such, online networks are less prone to recall bias and other misperceptions.

Social media has experienced a remarkable rise to prominence over the past decade and is increasingly used to maintain interpersonal relationships (Ellison and boyd, 2013). Initially, the popular belief was that such platforms, of which Facebook is the prime example, "would open up the vista of a social world that was intrinsically unlimited in size" (Dunbar et al., 2015: 39). These online networks could supposedly be used to mitigate the social segregation typically found in offline friendship networks (Rainie and Wellman, 2012; Robinson et al., 2015), as users are not restricted to exclusively befriending others from the schools they attend, from the neighborhoods in which they live, or from their offline activities.

Recent empirical research contradicts such beliefs about the promise of social media. Instead, this work shows that large online networks are indicative of an individual's complete set of offline relationships. For instance, early studies among US college students found that only .4 percent of online friendships are online-only (Mayer and Puller, 2008), and college students use online networks to maintain and strengthen their offline relationships but rarely to initiate new contacts (Ellison et al., 2007, 2011). Moreover, 80 percent of social network site users state that they use such social platforms to stay in touch with their offline ties (Subrahmanyam et al., 2008). Similarly, 77 percent of adolescents report that they befriend others online only when they have met them offline (Reich et al., 2008). In a study of Swedish adolescents, 77 percent of friends were friends both offline and online (Van Zalk et al., 2014). US adults indicate that their online network friends are predominantly family, current and past friends, neighbors, and colleagues (Duggan et al., 2015). Furthermore, the structure and size of offline and online networks are similar (Dunbar et al., 2015; Dunbar, 2016). Finally, studies conducted among US college students find similar high levels of segregation by ethnicity and race on Facebook as is found on campuses (Lewis et al., 2008b; Mayer and Puller, 2008; Wimmer and Lewis, 2010; Lewis et al., 2012).

The overlap between online and offline social networks provides a number of opportunities for sociological research: it enables the study of a large portion of overall personal networks, including small samples of core ties and far larger numbers of weaker ties. Empirically, it is relatively easy to collect data on online interactions, such as the data documented on Facebook, as the platforms generate time-stamped, digital footprints of all their users' relationships (Golder and Macy, 2014).

To illustrate this new approach to the study of segregation in social networks, we consider the Facebook networks of adolescents in the Netherlands. Because prior research shows that adolescents' strong-tie offline networks are highly segregated in terms of both *ethnicity* (e.g., Baerveldt et al., 2004; Vermeij et al., 2009) and *gender* (e.g., Shrum et al., 1988; Smith and Schneider, 2000; Lubbers, 2003), we first consider whether we find similar patterns of ethnic and gender segregation among the larger number of contacts in online networks.

Second, we examine the conditions under which ethnic and gender segregation of the extended network occur. Because previous research focuses exclusively on tie formation and segregation among core ties, there is little empirical evidence of the determinants of segregation among larger sets of network ties. In this study, we take the first steps to provide such evidence. As our theoretical point of departure, we engage classic network theories that are commonly used to explain segregation among core ties. Specifically, we consider the role of relative group size (Blau, 1977a, 1977b), foci (Feld, 1981, 1982, 1984), homophily (Lazarsfeld and Merton, 1954; Byrne, 1971), and balance (Heider, 1946; Krackhardt and Handcock, 2007), and we show their relevance in explaining segregation among hundreds of social relationships. As such, we contribute to the understanding of processes that underlie segregation in large networks while simultaneously testing existing, fundamental hypotheses in novel ways.

Third, we explore the differences in segregation between core networks and larger networks, because some scholars speculate there may be disparities in segregation among core and weaker ties (e.g., Granovetter, 1973, 1983; Putnam, 2000; Blackwell and Lichter, 2004; Mollenhorst et al., 2008; Son and Lin, 2012), although few studies have empirically studied this. Our study is among the first to elaborate and empirically test the conditions and mechanisms that create differences in the levels of segregation among core networks and larger networks.

4.2 Online Social Networks

We used a general survey of Dutch adolescents (Kalter et al., 2013, 2015) and linked these data to respondents' online Facebook networks in 2014. In doing so, we provide novel and detailed knowledge of respondents, their core networks, and their larger online networks. Previous studies of online social network segregation have used selective samples of Facebook friendships from US colleges in 2005 and 2006 (Mayer and Puller, 2008; Wimmer and Lewis, 2010). At present, Facebook friends represent a wide range of social ties, such as family members, friends, and neighbors, and Facebook is the largest social network site worldwide, with approximately one billion daily users (Facebook, 2015). Among adolescents in the Netherlands, Facebook is the most popular social network site; over 95 percent of Dutch adolescents have an account (Hofstra et al., 2016a).

Social networks emerge on Facebook when users send friendship invitations to other users, who can accept or decline the invitation. An accepted invitation shows an undirected, reciprocated friendship between two users. On Facebook, all relationships displayed on *friend lists* are indistinguishable with regard to tie strength (Lewis et al., 2008a). Nevertheless, considerable evidence suggests that the number of "best" friends does not exceed 5 to 10 people. Research also suggests that a cognitive limit prevents a person from maintaining more than approximately five *close* relationships (e.g., Zhou et al., 2005; Roberts et al., 2009; Dunbar et al., 2015). In addition, 95 percent of Americans reported fewer than six confidants (i.e., core ties) in the 1984 and 2004 GSS (McPherson et al., 2006).[2] The average of 336 friends in our online network data thus suggests that most of the sample's online friends are weak rather than strong ties. At the very least, we capture a large portion of an individual's complete personal network; thus, we clearly go beyond the small number of core social ties. This assumption is supported by

[2]There has been a debate on the increase in social isolation (i.e., having zero alters to discuss "important matters" with) of Americans between 1985 and 2004, as reported by McPherson and colleagues (2006). A 2008 erratum (McPherson et al., 2008) corrected a coding error in the original data release. In 2009 (Fischer, 2009; McPherson et al., 2009), a discussion emerged on whether the trend was a data artifact that resulted from respondent fatigue and training. Paik and Sanchagrin (2013) showed that the increase in social isolation may be attributed to substantial interviewer effects.

estimates suggesting that depending on the methods used, the average size of overall personal networks range from 150 (Hill and Dunbar, 2003) to 750 (Zheng et al., 2006) contacts.[3]

4.3 Theory and Hypotheses

A substantial literature examines how offline social ties form and why network segregation occurs (e.g., Blau, 1977a; Feld, 1981; Kalmijn, 1998; McPherson et al., 2001; Mouw and Entwisle, 2006; Kossinets and Watts, 2009; Wimmer and Lewis, 2010; Centola, 2015). Common explanations for the genesis of social segregation are relative group size, foci, homophily, and balance. We therefore focus on these factors.

Following Wimmer and Lewis (2010: 588), we use the term *homophily* for the tie-generating mechanism, which indicates a preference for the selection of similar friends, and we use the term segregation to describe the composition of a network. For clarity, we use the term homophily to indicate what is commonly called "choice" homophily (e.g., McPherson and Smith-Lovin, 1987; McPherson et al., 2001), that is, homophily net of meeting opportunities or other structural processes ("baseline" homophily). In this study, we examine segregation in social networks with regard to the ethnic and gender homogeneity found in personal social networks.

4.3.1 Meeting Opportunities: Relative Group Size and Foci

Theoretically and empirically, meeting opportunities are important in predicting strong tie formation (e.g., Kalmijn and Flap, 2001; Mouw and Entwisle, 2006; Mollenhorst et al., 2008, 2014; Vermeij et al. 2009; Wimmer and Lewis, 2010; Smith et al., 2014a). Two key dimensions of meeting opportunities are relative group size and foci (Blau 1977a, 1977b; Feld, 1981); we consider these factors because we expect they drive segregation in large online networks.

Relative group size. The relative size of a group is an important factor in friendship formation (Blau, 1977a, 1977b). Levels of personal network segregation may

[3] McCarty and colleagues (2001) found a mean total network size of 290; Hill and Dunbar (2003) found a size of 150; Zheng and colleagues (2006) found a size of 750; McCormick and colleagues (2010) found a size of 611; and DiPrete and colleagues (2011) found a median network size of 550.

reflect the distribution of social categories in a population. For instance, when a society consists of 20 percent minority members and 80 percent majority members, individuals' network contacts — of both majority and minority members — will consist of 20 percent minority and 80 percent majority members under the condition of random mixing.

In the Netherlands, ethnic groups' relative size varies, whereas the distribution of men and women is approximately 50/50 (Statistics Netherlands, 2015). Approximately 79 percent of people are "Dutch majority" members (Statistics Netherlands, 2015). In contrast, ethnic minority groups, who have an immigrant background, are much smaller in size. For instance, minority members with a Moroccan background compose approximately 3 percent of the Dutch population.

Given these differences, we first compare ethnic and gender segregation in online networks. If relative group size is important in explaining segregation, we would expect ethnic segregation in large personal networks to be higher than gender segregation, as the distribution of majority and minority populations is more unequal than the gender distribution. Large online networks will thus reflect these unequal distributions in the population. Considering these disparities at the level of the population at large, we propose the following hypothesis:

> **Hypothesis 1a:** *Larger online networks are more homogeneous by ethnicity than they are by gender.*

Second, we compare ethnic segregation between ethnic majority and minority members. When people belong to a large group, they have ample opportunities to meet members of their own group, whereas members of smaller groups are likely to develop many ties outside their own group (likely from the majority group). Therefore, we expect ethnic segregation in online networks is higher for members of the Dutch majority group than for members of an ethnic minority. We thus propose the following hypothesis:

> **Hypothesis 1b:** *Members of the ethnic majority have more homogeneous online networks than do members of the ethnic minority.*

Foci. Along with groups' relative size in a population, we consider foci and their role in friendship formation (Feld, 1981) and segregation. A focus is defined as "a social, psychological, legal, or physical entity around which joint activities are organized" (Feld, 1981: 1016). Social contexts can be represented as sets of different foci and individuals. Individuals engage in a number of different foci but not in all of them. Two individuals who engage in the same focus are thus more likely to share activities than are two individuals who do not share a focus. Sharing foci creates "positive sentiments indirectly through the generation of positively valued interaction" (Feld, 1981: 1017). Foci bring people together in mutually rewarding situations, and individuals form ties among others on whom they spend resources, such as time and emotions. Sharing a focus therefore increases the likelihood for a (friendship) tie to emerge (i.e., in the consideration of positive ties).

What aspects of foci foster dyadic similarity between individuals? Foci themselves are segregated because there is selectivity of specific groups to participate and enroll in particular foci (Feld, 1981; Feld and Carter, 1999). Hence, whereas a group's size relative to other groups is an important factor in friendship formation, these groups spread and organize in social settings in a nonrandom way. Therefore, personal networks will resemble the structural features of foci; that is, people who develop ties within foci will likely resemble one another.

Many empirical accounts illustrate that foci are segregated. In the United States and Europe, schools and school classes vary in their racial-ethnic compositions (Mouw and Entwisle, 2006; Vermeij et al., 2009; Smith et al., 2014b), and US and European neighborhoods and cities tend to be racially and ethnically segregated (Semyonov and Glikman, 2009; Lichter et al., 2015). Accordingly, scholars have found not only that many relationships are formed in the context of some sort of focus (e.g., Grosetti, 2005), but also that homogeneity in foci fosters segregation in personal networks (e.g., Feld ,1982, 1984; Kalmijn and Flap, 2001; Mollenhorst et al., 2014). This is occasionally called "inbreeding" homophily (McPherson et al., 2001).

We consider schools and classrooms to be major foci for tie formation among adolescents (McPherson et al., 2001), because adolescents spend a considerable amount of their time in these settings. These two settings do not capture *all* the foci of adolescents, and parts of the relative group size effect may be attributed to the nonrandom sorting of adolescents over foci that we do not capture in this study. Nevertheless, not all relationships originate from foci, because people may "meet 'by chance' or as a result of adjacency along some continuum" (Feld, 1981: 1018).

Foci effects are often found in segregation among core networks (e.g., Feld, 1982, 1984), and foci have similar effects on the dyadic similarity of friend *and* acquaintanceship networks that are measured by name-generating questions (Mollenhorst et al., 2008). Therefore, we assume that the foci mechanism does not vary by tie strength. We thus derive the following hypothesis:

> **Hypothesis 2:** *Ethnic and gender homogeneity in schools and classrooms predicts the ethnic and gender homogeneity of online networks.*

4.3.2 The Interplay Between Meeting Opportunities, Homophily, and Balance

Some scholars have suggested that core networks are more strongly segregated than are extended networks. Granovetter (1973: 1362), for example, states that "the stronger the tie connecting two individuals, the more similar they are" and "homophilous ties are more likely to be strong" (Granovetter, 1983: 210). Putnam (2000: 20) similarly speculates that strong relationships, which constitute "bonding" social capital, are more likely to exist among similar people, whereas weak ties, which create "bridging" social capital, are more likely to exist among dissimilar people. Son and Lin (2012: 602) argue that people with "stronger ties are more likely to share [...] commonalities" and as ties become weaker, "the ties' characteristics become dissimilar — more diverse." What follows is an explanation of the conditions that create differences in homogeneity among stronger and weaker ties, with a focus on meeting opportunities, homophily, and balance.

Two mechanisms suggest that dyadic similarity correlates with tie strength, and hence that core networks are more segregated than extended networks. The first is the homophily mechanism. According to the homophily argument, people generally prefer to befriend others similar to themselves (Lazarsfeld and Merton, 1954; Byrne, 1971; McPherson et al., 2001). Homophily exists along multiple dimensions, such as gender, ethnicity, race, education, or religion. Individuals may develop a *psychological* preference for similar friends (Byrne, 1971), which represents an enhanced degree of psychological attraction between two similar entities (Lewis, 2015b). Homophily may be driven by shared cultural norms and beliefs (Smith et al., 2014b), because shared norms can decrease the costs of investing in relationships (it takes less time to get to know one another) and increase returns on the investment (it becomes easier to interact) (Kalmijn, 1998).

Given that people have ample opportunities to select same-gender and same-ethnic friendships, homophily may be more pronounced among core ties than among weaker ties (Mollenhorst et al., 2008). One reason for this phenomenon is that stronger ties are "costly" (Windzio and Bicer, 2013), because strong ties involve more time, emotional intensity, intimacy, and reciprocal services (Granovetter, 1973: 1361), whereas "cheaper," weaker ties deplete fewer such resources. Therefore, if possible, individuals are more likely to strengthen their relationships with similar rather than dissimilar others (Windzio and Bicer, 2013; Leszczensky and Pink, 2015). Individuals perceive relationships in which they share commonalities with others to be more rewarding and less risky. People expect stable returns on investments in such relationships: it is easier to interact, and it takes less time to get to know one another because there are fewer (cultural) boundaries to overcome. Hence, dyadic similarity promotes tie strength.

The second reason why dyadic similarity would be associated with tie strength comes from the network balance mechanism. Assuming that similar dyads are more likely to be strongly connected than are dissimilar dyads, triadic closure (when A is friends with B, and A with C, then B and C are likely to connect) may occur more often among similar than among dissimilar individuals.[4] When ties are strong, unbalanced network configurations produce psychological strain for actors (Heider 1946), which leads them to close the "forbidden" triad (Granovetter, 1973). Furthermore, an individual who has two strong ties in a triad provides opportunities for the unconnected pair to befriend each other (Feld, 1981; Mollenhorst et al., 2011). Additionally, the dyadic survival of a relationship in an "isolated dyad" is lower than that of a dyad embedded in a triad, due to group identity formation, group pressure, and conflict control. These group dynamics are more likely to emerge within an embedded dyad, which creates an increased probability of triadic closure (Feld, 1997; Krackhardt and Handcock, 2007). Among embedded dyads characterized by a strong relationship, these dynamics may be even stronger, as these actors may more strongly call upon the group's identity and norms.

The composition of an individual's core friendship network, which is often limited to approximately five persons (e.g., Marsden, 1988; McPherson et al., 2006; Smith et al., 2014a), may thus be affected by homophily and balance, more so than weaker ties (which are initially formed by opportunity). In the opportunity set

[4] Balance is restored in this network configuration (when A is friends with B, and A with C, then B and C are likely to connect) under the assumption that these ties are positively signed, undirected, and of the same tie strength. For instance, a closed triad of three mutual foes is an unbalanced triad.

of network contacts, a person may have at least five similar available people with whom relationships can be strengthened. Individuals' larger networks, however, will be more likely to reflect the structural features of the population and foci. Initially, network ties mirror the features of meeting opportunities. Over time, however, ties characterized by dyadic similarity may transition into stronger ties, whereas dissimilar dyads may remain in their existing state of loosely connected weaker ties.

To examine this empirically, we first consider the number of friends individuals have in their online network. We assume that when people create online social network accounts, they start by adding close friends and contacts. This process is similar to name generators, in which people mention their closest ties first (Marin, 2004). Additionally, Facebook promotes network closure: it prompts people to become friends with the friends of their friends, which also makes it more likely that an individual's first friends on Facebook will be strong ties. When the number of online social network friends increases, an increasing number of them will likely be weaker ties.

These factors should result in lower levels of ethnic and gender homogeneity in online networks, because the relative number of strong ties is lower. However, we expect that the negative association between the number of online network friends and ethnic homogeneity pertains only to members of ethnic minorities. The opportunity set of potential contacts is often shaped such that ethnic majority members are overrepresented in public life and in foci. Therefore, members of larger ethnic majority groups have limited opportunities to befriend people from smaller groups: there are fewer such persons in the population and in the foci. Among ethnic majorities, this means core networks and larger networks largely comprise majority group members. Minority group members, in contrast, meet many dissimilar others (likely of the majority group). Although they may strengthen their relationships with the few similar minority members whom they meet (because of homophily and balance), their larger network will continue to resemble the structural features of the meeting opportunities. We therefore expect individuals with larger online networks to have lower levels of ethnic and gender homogeneity — the exception being ethnic majority groups, as the opportunities for meeting co-ethnics are so widespread for this group. Specifically, we propose the following hypotheses:

> **Hypothesis 3a:** *As online network size increases, ethnic homogeneity decreases, but only among ethnic minorities.*

> **Hypothesis 3b:** *As online network size increases, gender homogeneity decreases.*

We also provide a different test of the same arguments by directly contrasting ethnic and gender homogeneity among small, self-reported core networks with ethnic and gender homogeneity found in large online personal networks. Instead of examining the number of connections only in online networks, we compare core and large online networks directly. We thus hypothesize the following:

> **Hypothesis 3c:** *Core networks have more ethnic homogeneity than do larger online networks, but only among ethnic minorities.*

> **Hypothesis 3d:** *Core networks are more gender homogeneous than the larger online networks.*

4.4 Data and Measures

We use the second wave of survey data on adolescents in the Netherlands, which is part of a larger project titled "Children of Immigrants Longitudinal Survey in Four European Countries" (CILS4EU) (Kalter et al., 2013, 2015).[5] Although data were collected in the Netherlands, Sweden, Germany, and England, the measures we are interested in are included only in the Dutch data. In CILS4EU, adolescents age 14 to 15 years were followed for three years, starting in 2010, with a one-year time lag. The survey included data on many individual characteristics, attitudes, and information about the individuals with whom respondents associated with in their leisure time. The survey also contained sociometric data on friendships within classrooms (∼22 pupils in a classroom). The sample was stratified by the proportion of non-Western immigrants within a school. Within these strata, schools were chosen with a probability proportional to their size (based on the number of pupils at the relevant educational level).

In wave 1 (2010 to 2011), two classes were randomly selected within the schools,

[5] One can apply for data access to waves 1, 2, and 3 of the CILS4EU via the following link: https://dbk.gesis.org.

which resulted in 118 schools, 252 classes, and 4,963 Dutch pupils participating in the survey.[6] Because changes in class compositions between grades are common in the Netherlands, respondents were distributed among different classes in wave 2 (2011 to 2012) that were not part of the original sampling frame. To ensure that many wave 1 pupils also participated in wave 2, schools were asked to include more than the two classes initially sampled in wave 1 when respondents from wave 1 were in classes other than the previously sampled classes. Consequently, 2,118 new pupils were interviewed, and 3,803 of wave 1 respondents were surveyed again in wave 2 (76.6 percent; total N = 5,921). We used the second wave of the CILS4EU because it is the latest licensed data including sociometric classroom information.

4.4.1 The Dutch Facebook Survey

The Dutch Facebook Survey (DFS) enriched the Dutch part of the CILS4EU survey (Hofstra et al., 2015).[7] Data were collected between June and September 2014. Of the 4,864 respondents who indicated Facebook membership in wave 3 (2012 to 2013; N = 3,423) or 4 (2013 to 2014; N = 3,595) of the CILS4EU, 4,463 were tracked on Facebook. For respondents who kept a *public friend* list, we downloaded their complete Facebook friend lists (N = 3,252; 72.8 percent). There is selectivity in the downloaded friend lists: some respondents kept their lists private, others kept public friend lists. Girls, ethnic minority members, and unpopular adolescents are somewhat underrepresented, because they more often keep private friend lists (Hofstra et al., 2016b). Various Heckman selection-model specifications (Heckman, 1979) show that our results are insensitive to these selection biases.[8] The 3,252 respondents have a combined total of 1,158,227 friends, and

[6]In wave 1, 600 respondents who were not part of the random sampling frame were sampled because some schools wanted to participate in the survey with more than two classrooms. Therefore, a random sample of 4,363 pupils was drawn in wave 1. Because of the attrition rates between waves 1 and 2, our sample is not necessarily representative. We included as many respondents as possible in the sample for analyses, including newcomers (nonrandom) and the nonrandom sample of wave 1, to ensure a large sample size.

[7]An anonymized version of the DFS will be available in October 2017 (https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:62379).

[8]We performed robustness analyses using Heckman selection models (Heckman, 1979). Therefore, we corrected for selectivity in modeling an outcome only when a second selection equation determined that this outcome was non-missing. The errors of both equations are allowed to correlate. We correct in the selection equation for ethnicity, gender, popularity, and educational-track level. We cluster-corrected standard errors for the class cluster and school cluster, because multilevel Heckman models were computa-

Table 4.1: Overview of the relevant data sources and selections.

	N	%
Survey data (CILS4EU)		
W2 total number of respondents	5,921	100%
W2 respondents participated in W1	3,803	64.2%
W2 respondents that are newcomers	2,118	35.8%
Online network data (DFS)		
Respondents indicated being on Facebook in W3 or W4 of the survey	4,864	100%
Those respondents whose profiles were tracked on Facebook	4,463	91.8%
Those respondent keeping a public Facebook friend list	3,252	66.9%
Conditions for inclusion in the final number of cases to analyze		
W2 + Tracked on FB[a] + Keeping a public FB friend list	2,810[b]	

[a] FB = Facebook; [b] Various Heckman selection model specifications showed that our results were insensitive for selection biases.

2,810 (86.4 percent) of the respondents whose complete friend list we downloaded also participated in wave 2 of the CILS4EU.[9] This is the number of respondents for whom we present results.[10] Table 4.1 summarizes the data sources and our method of arriving at the final number of respondents.

4.4.2 Measuring Ethnic and Gender Homogeneity in Online Networks

There is no direct measure for friends' ethnic background and gender in the Facebook network. We predicted friends' gender and ethnic background based on their first names,[11] using the Dutch Civil Registration data (hereafter, DCR) for the entire Dutch population in 2010 (N = 15,785,208; Bloothooft and Schraagen,

tionally infeasible. These analyses did not provide different results than those we present here. We present the analyses that consider the clustered data. Tables are available upon request.

[9]The combined total of 1,158,227 friends is a raw count of all respondents' friendships. Respondents likely have similar friends in their online networks. Counting the unique set of friends would most likely result in a lower number.

[10]The collection and use of these data for scientific purposes were internally approved by an ethical review board for the social and behavioral sciences.

[11]We also assigned ethnicities based on name carriers' last names, following the procedure outlined in Appendix 4.1. We obtained correlations similar to those based on first names. We re-performed all analyses pertaining to ethnicity, and the results are robust when we consider last names.

2011). We obtained (1) the fraction of the name carriers and (2) the fraction of the name carriers' fathers and (3) mothers who were born in the Netherlands, Turkey, Morocco, the Dutch Caribbean, other Western countries, or other non-Western countries. Additionally, we obtained the percentage of women among the name carriers.

We matched first names in the DCR to first names in the second wave of the CILS4EU survey as a training dataset. In the CILS4EU, we measured respondents' ethnic background by classifying them into one of the six largest ethnic groups in the Netherlands (Castles et al., 2013): Dutch majority, Turkish, Moroccan, Dutch Caribbean, other Western (European or English speaking), and other non-Western. Moroccan and Turkish adolescents are children of immigrants from the low-educated labor force that was recruited by the Netherlands in the 1950s and 1960s. Dutch Caribbean adolescents originate from post-colonial countries in the Dutch Caribbean (e.g., Aruba and Suriname). Western and non-Western adolescents originate from neighboring countries such as Germany or conflict areas such as Afghanistan; these immigrant groups are relatively similar across Western European countries (Smith et al., 2014b).

We classified respondents according to their biological parents' country of birth, which is standard practice in research on Dutch ethnic minority groups (cf. Vermeij et al., 2009; Stark and Flache, 2012; Smith et al., 2014b). When students have one parent who was born in the Netherlands, the student is classified into the ethnic background of the parent who was not born in the Netherlands; if a student's parents were born in different non-Dutch countries, the student is classified according to the mother's birth country. This definition is regularly applied and used by Statistics Netherlands (Statistics Netherlands, 2012).

Combining the DCR and the CILS4EU, we developed an algorithm to estimate gender and ethnic segregation based on people's first names, which yields high correlations between the predicted and actual ethnicity and gender (this method is outlined in Appendix 4.1). We calculated the percentage of women and the percentage of each of the six ethnicities in respondents' online networks.[12] For each respondent, we assigned the percentage of same-gender friendships (i.e., the percentage of women for girls and percentage of men for boys) in their online networks. Finally, we assigned each respondent the percentage of co-ethnic ties in their online networks (e.g., the percentage of Dutch majority members among online network friends for the Dutch majority adolescents).

4.4.3 Homogeneity in Core Friendship Networks

Wave 2 of the CILS4EU has two measures that capture ethnic homogeneity and one measure that captures gender homogeneity in core friendship networks: a name generator for the five best friends *in general* (only for ethnicity), and a name generator for the five best friends *in class* (not necessarily the same friends as the former).

First, we measured the actual number of friends of Dutch, Turkish, Moroccan, Dutch Caribbean, or another ethnic background using a name-generator question. Respondents could nominate their best friends (with a maximum of five) and provide ethnic background information. From these data, we calculated the percentage of co-ethnic friends among all the close friends (co-ethnic$_{\text{FRIENDS IN GENERAL}}$). We consider *ethnically similar* friends among best friends in general. Respondents may be more accurate in reporting ethnicities of ethnically similar than ethnically dissimilar friends. Furthermore, respondents were asked to report the ethnicities of their *best friends*. Respondents may more accurately report the ethnicities of their best friends than those of acquaintances. Therefore, respondents' misreporting of alter characteristics is likely reduced to a minimum.

Second, we measured the number of best friends in a class (with a maximum of five) who were girls (which is the only core-network measure available for

[12] We also performed all of our descriptive analyses for ethnicity with an index of qualitative variation (IQV) — the inverse of network *diversity* (Agresti and Agresti, 1977). The IQV for pupil i is formally defined as follows:

$$IQV_i = 1 - [(\frac{k}{k-1})(1 - \sum_{b=1}^{k} p_b^2)], \tag{4.1}$$

where k is the number of ethnic categories and p_b is the fraction of Facebook friends in the bth category ($b = 1, ..., k$). IQV has been used in various studies to measure (ethnic) diversity in networks (e.g., Marsden, 1987; McPherson et al., 2006; Lewis et al., 2008b). In none of the analyses did the results differ from those presented in the article.

measuring gender homogeneity) and those who were of Dutch, Turkish, Moroccan, Dutch Caribbean, other Western, or other non-Western ethnic backgrounds. We calculated the percentage of co-ethnic friends and the percentage of same-gender friends among all friends in a class (co-ethnic$_{\text{FRIENDS IN CLASS}}$ and same-gender$_{\text{FRIENDS IN CLASS}}$). Because these friends themselves were respondents in the survey, they self-reported their gender and ethnicity. We constructed gender and ethnic homogeneity with respect to best friends in a class with these self-reports, and hence they do not suffer from respondents' misperceptions in alter characteristics.

4.4.4 Homogeneity in Meeting Opportunities and Number of Online Network Friends

We constructed various measures to capture ethnic and gender homogeneity in two adolescent opportunity structures, the class and the school. First, using the CILS4EU, we measured the number of *classmates* who are female and those with the six ethnic backgrounds mentioned above, excluding best friends who are mentioned in the class and respondents themselves. We calculated the percentage of same-gender and co-ethnic classmates, and we excluded the respondent and the number of best friends who are mentioned. We excluded best friends because they are included in the core-network measure, and we do not want to double-count best friends across variables. With this approach, we are better able to separate the effects between variables (same-gender$_{\text{IN CLASS}}$ and co-ethnic$_{\text{IN CLASS}}$).

Second, we measured the number of female pupils in a school (aggregated from the classes surveyed) and the number of pupils in the school from a Dutch, Turkish, Moroccan, Dutch Caribbean, other Western, or other non-Western ethnic background (measured from secondary data obtained from the Dutch inspectorate), excluding best friends who are mentioned, other classmates, and the respondent. We calculated the percentage of same-gender schoolmates (excluding the respondent, the number of best friends who are mentioned, and the number of classmates) (same-gender$_{\text{IN SCHOOL}}$). We also measured the percentage of co-ethnic schoolmates (excluding the respondent, the number of best friends, and the number of classmates) (co-ethnic$_{\text{IN SCHOOL}}$). We measured these two variables using the CILS4EU.

We also calculated the number of online network friends from respondents' Facebook friend lists using the DFS. The distribution of the number of online network friends, if it is plotted, strongly resembles the distribution plot reported by DiPrete and colleagues (2011: 1254) of the number of acquaintances reported by Americans.

4.4.5 Kinship Ties in Online Networks as a Confounding Factor

An issue with online versus offline friendship networks is that we restricted respondents to name *friends* in their self-reported core networks offline, whereas Facebook networks likely include *kin*. Therefore, when we contrast core networks with online networks, we compare two data sources of different sampling frames. Kinship ties in online networks might pull ethnic and gender homogeneity in different directions. Kin likely have a similar ethnicity as the respondent, whereas gender distribution in families is likely to be 50/50. On the one hand, the presence of kin in online networks overestimates ethnic homogeneity; on the other hand, the presence of kin among Facebook friends might lead us to underestimate gender homogeneity (see McPherson et al., 2001: 431).

We identify kinship ties in online networks in two ways, using the DFS. First, Facebook allows members to show kinship tags on their profiles. We tracked the number of kinship tags on Facebook profiles and calculated the percentage of kinship tags in the Facebook network (mean = 1.1 percent). We considered realistic tags (e.g., no granddaughters, given that we study adolescents). Individuals might not tag each family member on Facebook. Therefore, we calculated the percentage of friends in the Facebook network who share a surname with the respondent (mean = 1.7 percent). Non-kin friends may have a similar surname, which makes our analyses more conservative, because individuals with similar surnames are likely of the same ethnicity. Nevertheless, we may miss kin in online networks who are not tagged and who have different surnames. We mention where we remove kin from the online networks to avoid sampling mismatches (descriptive comparisons between core networks and the larger online networks) and where we control for these two variables (statistical tests of the hypotheses).[13]

[13] In removing kin from the Facebook homogeneity estimates, we assume that all kin are of a similar ethnic background as the respondent. We reduce the number of co-ethnic friends and the number of Facebook friends by the number of identified kin ties and calculate the percentage of co-ethnic friends on Facebook. We assume that half the kin ties are of similar gender as the respondent. We reduce the number of same-gender friends by *half* the number of identified kin and subtract the *total* kin ties that are identified from the number of Facebook friends.

Table 4.2 shows the descriptive statistics for ethnic and gender homogeneity in the large online networks (including kin), core networks, opportunity structures, and kinship ties in online networks, along with the distributions of boys and girls and ethnic groups in the data.

Table 4.2: Descriptive statistics of ethnic and gender homogeneity in large online networks, in opportunity structures, kinship ties on Facebook and the distribution of boys and girls and ethnic background.

	Min.	Max.	Mean	SD[a]	N
Online networks[b]					
Co-ethnic$_{\text{FACEBOOK}}$	0	100	76.577	32.099	2,810
Same-gender$_{\text{FACEBOOK}}$	0	100	56.087	9.745	2,809
% Female	0	100	49.453	11.475	2,810
% Dutch	0	100	86.200	15.670	2,810
% Turkish	0	100	2.304	7.649	2,810
% Moroccan	0	59.460	1.729	5.015	2,810
% Dutch Caribbean	0	54.237	1.347	3.097	2,810
% Other Western	0	57.142	3.176	2.566	2,810
% Other non-Western	0	75.676	4.234	5.025	2,810
Core networks					
Co-ethnic$_{\text{FRIENDS IN GENERAL}}$	0	100	76.218	33.730	2,810
Co-ethnic$_{\text{FRIENDS IN CLASS}}$	0	100	67.525	38.249	2,677
Same-gender$_{\text{FRIENDS IN CLASS}}$	0	100	83.175	30.227	2,677
Opportunity structures					
Co-ethnic$_{\text{IN CLASS}}$	0	100	65.710	31.743	2,690
Co-ethnic$_{\text{IN SCHOOL}}$	0	100	67.038	30.926	2,763
Same-gender$_{\text{IN CLASS}}$	0	100	50.212	22.163	2,690
Same-gender$_{\text{IN SCHOOL}}$	0	100	47.474	18.328	2,638
Number of Online Network friends	1	1067	336.853	177.702	2,810
Kinship ties on Facebook					
% kinship ties declared	0	20	1.081	1.555	2,794
% similar surname on Facebook	0	100	1.689	3.303	2,794
Girl	0	1	0.515	-	2,809
Ethnic background	-	-	-	-	2,810
Dutch	0	1	0.804	-	2,258
Turkish	0	1	0.020	-	57
Moroccan	0	1	0.015	-	42
Dutch Caribbean	0	1	0.023	-	65
Other Western	0	1	0.088	-	247
Other non-Western	0	1	0.050	-	141

[a] SD = Standard deviation; [b] These estimates of homogeneity in Facebook networks include kin.

4.4.6 Additional Confounding Factors

We adjust for the year in which respondents joined Facebook using the DFS (median = 2010). Respondents who were members for shorter periods may have been more selective in their online network friendships. Facebook membership duration and the number of Facebook friends are positively correlated (r = .250; $p < .001$).

Using the CILS4EU, we also control for educational track in high school, because such a track may be related to ethnic prejudice (Lancee and Sarrasin, 2015). When adolescents transition to high school in the Netherlands, they are placed into different tracks, which differ in their level and type of education. We measured this categorization using three dummy variables: preparatory vocational education (N = 1,358; Dutch: VMBO), senior general (N = 750; Dutch: HAVO), and university preparatory education (N = 586; Dutch: VWO). We also control for respondents' social attractiveness, which may be correlated with ethnicity (Wimmer and Lewis 2010). We measured social attractiveness by *popularity* (i.e., incoming popularity nominations from other classmates) (mean = 9.357; SD = 14.566). We calculated popularity by dividing the total number of classmates' received nominations for popularity by the total number of students in the class minus one multiplied by 100.[14]

We adjust for ethnic out-group attitudes because they may be related to ethnic homogeneity in online networks. With the survey question, "Please rate how you feel towards the following groups..." respondents used a scale ranging from 0 (negative) to 100 (positive), with 10-point intervals, to rate how positively they feel toward groups of Dutch, Turkish, Moroccan, and Dutch Caribbean ethnic backgrounds. We constructed ethnic out-group attitudes by taking the mean positivity score — on a scale from 0 to 10 — of respondents' answers to this question while excluding the respondent's own ethnic group (mean = 5.011; SD = 1.997). This variable is significantly negatively related to the percentage of co-ethnic friends online (r = –.213; $p < .001$).

We accounted for respondents' attitudes toward gender roles when we considered gender homogeneity in online networks. We captured respondents' progressiveness

[14]Indegree popularity can be formally defined as follows:

$$(\sum_i \frac{B_{ji}}{N-1} \times 100) \tag{4.2}$$

where i is the actor, B_{ji} indicates whether pupil j nominates pupil i as popular, and N is the total number of pupils in a classroom.

toward gender roles by counting (from zero to four) how many times respondents indicated that both men and women (instead of men or women) should take care of children, cook, earn money, and clean (Davis and Greenstein 2009) ($\alpha = .73$; mean = 2.689; SD = 1.352). This variable is significantly negatively related to the percentage of same-gender friends online ($r = -.072$; $p < .001$).[15, 16]

4.5 Results

4.5.1 Ethnic and Gender Homogeneity in Online Networks

In 2014, Dutch adolescents' online social networks had, on average, 76.6 percent co-ethnic friends. If everyone connected at random on Facebook in the Netherlands, the average personal network would consist of 78.6 percent Dutch, 2.4 percent Turks, 2.2 percent Moroccans, 2.9 percent Dutch Caribbean, 9.5 percent other Western individuals, and 4.3 percent individuals with other non-Western ethnic backgrounds. However, on average, the online networks in our sample consist of 86.2 percent Dutch, 2.3 percent Turks, 1.7 percent Moroccans, 1.4 percent Dutch Caribbean, 3.2 percent other Western individuals, and 4.2 percent individuals with other non-Western backgrounds (see Table 4.2).

Table 4.3 shows the ethnic homogeneity of core networks and online networks, and Table 4.4 shows these results broken down by ethnicity (both tables exclude kinship ties in online networks). The percentage of co-ethnic friends online is 76 percent; in core networks it is 76.2 percent for friends in general (co-ethnic$_{\text{FRIENDS IN GENERAL}}$) and 67.5 percent for friends in a class (co-ethnic$_{\text{FRIENDS IN CLASS}}$). The correlations between co-ethnic friendships in core and larger online networks are high: .784 (co-

[15] We also controlled for dummy variables that indicate to which stratum in the sampling frame the respondent belongs; thus, we account for some of the selectivity in the sampling strategy. In none of the analyses does this control variable lead to qualitatively different results. To keep the results parsimonious, we present the results without these variables.

[16] We furthermore controlled for dummy variables that indicate respondents' generational immigration status. Categories are, for instance, Dutch majority adolescents or adolescents who have only one foreign-born grandparent. Thus, we account for differences in immigration background. In none of the analyses does this control variable lead to qualitatively different results. To keep the results parsimonious, we present the results without these variables. For more information about generational status in the CILS4EU data, see Dollmann and colleagues (2014) and the CILS4EU Wave 2 Codebook (2016: 273).

ethnic$_{\text{FRIENDS IN GENERAL}}$) and .677 (co-ethnic$_{\text{FRIENDS IN CLASS}}$). Dutch majority members have the highest ethnic homogeneity online, 91.7 percent (co-ethnic$_{\text{FACEBOOK}}$), which resembles the homogeneity in core networks (co-ethnic$_{\text{FRIENDS IN GENERAL}}$ = 88.4 percent). The ethnic homogeneity of Turkish adolescents (co-ethnic$_{\text{FACEBOOK}}$ = 40.6 percent) is slightly higher than that of Moroccan adolescents (co-ethnic$_{\text{FACEBOOK}}$ = 28.5 percent). Ethnic homogeneity in online networks (~336 friends) mirrors ethnic homogeneity in core networks (~5 friends).

Table 4.3: Ethnic homogeneity in large online networks and ethnic homogeneity in core networks.

	Min.	Max.	Mean	SD[a]	N
Co-ethnic$_{\text{FACEBOOK}}$[b]	0	100	75.974	32.099	2,792
Co-ethnic$_{\text{FRIENDS IN GENERAL}}$	0	100	76.218	33.729	2,810
Co-ethnic$_{\text{FRIENDS IN CLASS}}$	0	100	67.525	38.249	2,677

[a] SD = Standard deviation; [b] This estimate of homogeneity in Facebook networks exclude kin.

Table 4.4: Ethnic homogeneity in large online networks and ethnic homogeneity in core networks, broken down by ethnicity.

	Dutch	Turkish	Moroccan	Dutch Carib.	Other West.	Other non-West.
Co-ethnic$_{\text{FACEBOOK}}$[a, b]	91.569	40.604	28.455	9.176	3.167	13.397
Co-ethnic$_{\text{FRIENDS IN GENERAL}}$	88.412	54.503	45.198	27.692	14.899	28.759
Co-ethnic$_{\text{FRIENDS IN CLASS}}$	79.843	31.730	19.228	12.769	11.364	19.975

[a] For the percentages of specific ethnic backgrounds within online networks broken down by ethnic background of respondents (e.g., the percentage of Moroccans in Facebook networks of Dutch majority members) we refer to the figure found in Appendix A4.2; [b] This estimate of homogeneity in Facebook networks exclude kin.

Table 4.5 shows the gender homogeneity of core networks and online networks broken down by gender (excluding kinship ties). On average, respondents have 56.3 percent same-gender friendships online. If everyone connected at random, such that the percentage of same-gender friendships on Facebook reflected the gender composition at the societal level, this number should be approximately 50 percent (Statistics Netherlands, 2015). On average, adolescents reported 83.2 percent same-gender friends in a class. Boys had slightly more same-gender friendships online than did girls (boys = 57.1 percent; girls = 55.5 percent), but boys had approximately the same percentage of same-gender friendships in a class as did girls (boys = 83 percent; girls = 83.4 percent).

Table 4.5: Gender homogeneity in large online networks and gender homogeneity in core networks, broken down by gender.

	Min.	Max.	Mean	SD[a]	N
$\text{Same-gender}_{\text{FACEBOOK}}$[b]	0	100	56.313	10.041	2,791
$\text{Same-gender}_{\text{FRIENDS IN CLASS}}$	0	100	83.175	30.227	2,677
Boys					
$\text{Same-gender}_{\text{FACEBOOK}}$	0	100	57.132	10.153	1,356
$\text{Same-gender}_{\text{FRIENDS IN CLASS}}$	0	100	82.952	30.144	1,299
Girls					
$\text{Same-gender}_{\text{FACEBOOK}}$	0	100	55.538	9.876	1,435
$\text{Same-gender}_{\text{FRIENDS IN CLASS}}$	0	100	83.384	30.314	1,378

[a] SD = Standard deviation; [b] This estimate of homogeneity in Facebook networks exclude kin.

4.5.2 Meeting Opportunities: Relative Group Size and Foci

Relative group size. We begin by examining the role of relative group size in homogeneity in online networks. We first evaluate the extent to which ethnic and gender homogeneity estimates in online networks differ from one another (Hypothesis 1a). Figure 4.1 shows the kernel density smoothed distributions for ethnic and gender homogeneity in the online networks, suggesting that online networks are more segregated by ethnicity than they are by gender.

We estimated an intercept-only multilevel model in which the intercept is the sample mean difference between the percentage of co-ethnic and same-gender friends online. This model can be specified as follows:

$$Y_{ijk} = \beta_{000} + s_{0k} + c_{0jk} + p_{0ijk}, \tag{4.3}$$

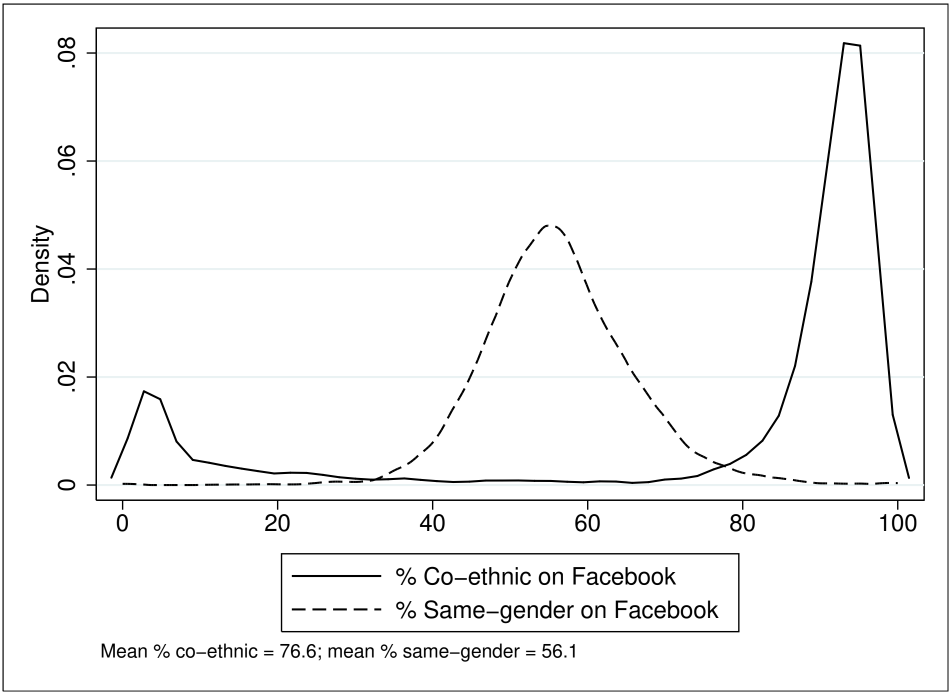


Figure 4.1: Density plots of ethnic and gender homogeneity in large online networks.

where Y_{ijk} is the difference in ethnic and gender homogeneity in online Facebook networks (co-ethnic$_{\text{FACEBOOK}}$ − same-gender$_{\text{FACEBOOK}}$) for respondent i from class j and school k; $s_{0k} \sim (0, \sigma^2_{s_{0k}})$ is the error term at the school level; $c_{0jk} \sim (0, \sigma^2_{c_{0jk}})$ is the error term at the class level; $p_{0ijk} \sim (0, \sigma^2_{p_{0ijk}})$ is the error term at the pupil level; and β_{000} is the sample mean difference in this *intercept only model.*[17, 18]

[17]Pupils from the same class may look more alike than pupils from two classes, and the proportion of variance explained at the class and school levels represents the expected correlation between two randomly selected pupils within the same class. This is defined as follows:

$$\rho_{s+c} = \frac{\sigma^2_{s_{0k}} + \sigma^2_{c_{0jk}}}{\sigma^2_{s_{0k}} + \sigma^2_{c_{0jk}} + \sigma^2_{p_{0ijk}}} \tag{4.4}$$

[18]Pupils from the same school may resemble each other more than pupils from different schools, and the expected correlation between two randomly selected pupils from the same school is defined as follows:

$$\rho_{s} = \frac{\sigma^2_{c_{0jk}}}{\sigma^2_{s_{0k}} + \sigma^2_{c_{0jk}} + \sigma^2_{p_{0ijk}}} \tag{4.5}$$

With these models, we control for class and school tendencies in the difference between ethnic and gender homogeneity in online networks (Snijders and Bosker, 2012).

The intercept of the intercept-only model significantly deviates from zero (intercept = 16.8; $p < .001$; see Appendix 4.3 for the table with the full results), suggesting that ethnic homogeneity is approximately 16.8 percent higher than gender homogeneity in online networks (Hypothesis 1a).

Second, we evaluate how the size of the ethnic majority group relative to the minority groups relates to ethnic segregation in online networks (Hypothesis 1b). We specified several multilevel regression models in order to estimate the percentage of same-ethnic friends within respondents' online networks.[19, 20, 21] We delete the missing values of the variables presented in these analyses listwise and lose 9.1 percent (N = 261) of cases in the analyses. These models can be specified as follows:

$$Y_{ijk} = \beta_{000} + \beta_{00x} X_i + s_{0k} + c_{0jk} + p_{0ijk}, \tag{4.6}$$

where Y_{ijk} is the percentage of co-ethnic friendships in the online network for pupil i from class j and school k; s_{0k}, c_{0jk}, p_{0ijk}, and β_{000} specify similar terms as in Equation 4.3; and β_{00x} is a vector for the independent variables at the pupil level (e.g., ethnic background). Table 4.6 shows the model that estimates the percentage of co-ethnic friends online for all respondents and separately for Dutch

[19]We estimated similar models with the *number* of co-ethnic and same-gender friendships on Facebook and offline instead of percentages as dependent and independent variables (we controlled for network sizes). We estimated random-effect models with either a random intercept at the class or school level. We estimated fixed-effect models with dummies for schools and classes. Finally, we estimated a model with only Dutch, Turkish, and Moroccans because we could best predict these ethnicities. In none of these analyses did the results qualitatively differ from the results presented here. Full tables are available upon request.

[20]We performed additional robustness analyses to investigate whether our main results are driven by socially isolated or highly connected adolescents on Facebook. We obtained similar results when we selected respondents with more than 50 and fewer than 750 friends on Facebook in our statistical models.

[21]For Dutch majority members, we performed analyses considering the percentage of co-ethnic neighborhood residents based on supplementary data from Statistics Netherlands. More co-ethnic neighborhood residents positively relate to the percentage of co-ethnic friends on Facebook. We could not perform similar analyses for ethnic minority members, because these data do not distinguish between the presence of various ethnic minority groups in a neighborhood.

majority and ethnic minority members (while controlling for kinship ties in the online networks).

We find relatively large effects of ethnicity on ethnic homogeneity in online networks (while controlling for ethnic homogeneity in the class and school setting). Dutch majority-group adolescents seem to have at least 31 percent more co-ethnic friendships online than do students of other ethnic backgrounds ($p < .001$). For instance, adolescents of Moroccan ethnic background have 43.5 percent fewer, and those of Turkish descent 31.8 percent fewer, co-ethnic online networks friendships than do Dutch majority members. Among ethnic minority members, students of Turkish ethnic background exhibit more ethnic segregation in online networks than do all of their ethnic minority counterparts: they have at least 9 percent more co-ethnic friends in their online networks ($p < .01$). These variables reflect the propinquity of co-ethnic individuals in the population and relate to co-ethnic friendships in the large online networks. The results thus show that the larger majority group has significantly higher levels of ethnic homogeneity than do ethnic minority group members (consistent with Hypothesis 1b).

Foci. We now consider the extent to which the homogeneity of foci relates to homogeneity in online networks. We ask whether the percentage of co-ethnic and same-gender peers in class and in school is related to homogeneity in online networks (Hypothesis 2). In addition to the results shown in Table 4.6, we estimated a multilevel regression for the percentage of same-gender ties in online networks. This model takes the form of Equation 2, but here, Y_{ijk} specifies the percentage of same-gender friendships online. We delete the missing values of the variables presented listwise in this analysis and lose 7.8 percent (N = 212) of the cases. Table 4.7 shows results of this model (with kinship ties as control variables).

Table 4.6: Multilevel model estimating the percentage of co-ethnic friends in online networks.

	All respondents			Only Dutch majority			Only non-Dutch minorities		
	Coef.	S.E.[a]	p[b]	Coef.	S.E.	p	Coef.	S.E.	p
Fixed part									
Intercept	67.935	(2.946)	***	68.491	(3.482)	***	32.998	(5.848)	***
Core-network									
Co-ethnic$_{\text{FRIENDS IN GENERAL}}$	0.083	(0.010)	***	0.052	(0.009)	***	0.120	(0.019)	***
Co-ethnic$_{\text{FRIENDS IN CLASS}}$	0.008	(0.004)	*	0.008	(0.004)	*	0.004	(0.020)	
Opportunity									
Co-ethnic$_{\text{IN CLASS}}$	0.033	(0.013)	**	0.019	(0.009)	*	0.055	(0.050)	
Co-ethnic$_{\text{IN SCHOOL}}$	0.212	(0.031)	***	0.216	(0.034)	***	0.256	(0.076)	***
Ethnicity									
Dutch	Ref.	Ref.	Ref.	-	-	-	-	-	-
Turkish	-31.820	(3.617)	***	-	-	-	Ref.	Ref.	Ref.
Moroccan	-43.554	(3.249)	***	-	-	-	-9.398	(3.557)	**
Dutch Caribbean	-57.771	(2.760)	***	-	-	-	-22.427	(3.149)	***
Other Western	-61.696	(2.431)	***	-	-	-	-25.620	(3.093)	***
Other non-Western	-56.070	(2.044)	***	-	-	-	-21.088	(2.904)	***
Number of Facebook friends	-0.002	(0.001)	*	-0.001	(0.001)		-0.007	(0.003)	***
Facebook membership (ref.: 2013)									
2012	-1.836	(1.946)		-0.472	(1.211)		-0.426	(4.625)	
2011	-2.297	(1.783)		0.417	(1.150)		-5.468	(4.113)	
2010	-2.018	(1.737)		0.578	(1.123)		-4.313	(3.919)	
2009	-2.266	(1.756)		0.166	(1.144)		-3.585	(4.044)	
2008	-1.447	(1.759)		0.091	(1.158)		-0.866	(4.085)	
2007	-2.571	(1.881)		-0.446	(1.413)		-3.136	(4.652)	
2006	-1.141	(3.028)		-0.569	(3.221)		2.532	(5.132)	
Girls (ref.: boys)	0.043	(0.250)		0.124	(0.149)		-0.661	(0.953)	
Educational track (ref.: lower voc.)									
Senior General	-0.103	(0.575)		-0.265	(0.480)		-1.370	-1.019	
University preparatory	-1.188	(0.578)	*	-0.333	(0.522)		-1.937	-1.086	*
Indegree popularity	-0.007	(0.010)		-0.010	(0.008)		0.015	(0.028)	
Ethnic outgroup attitudes	-0.188	(0.050)	***	-0.195	(0.042)	***	-0.037	(0.210)	
% kinship ties declared	0.063	(0.118)		0.004	(0.067)		0.202	(0.416)	
% similar surname on Facebook	0.280	(0.152)	*	0.091	(0.041)	*	0.906	(0.266)	***
Random part									
$\sigma^2_{s_{0k}}$ (School level)	5.246	(2.597)		7.027	(2.930)		0.000	(0.000)	
$\sigma^2_{c_{0jk}}$ (Class level)	0.000	(0.000)		0.108	(1.025)		0.000	(0.000)	
$\sigma^2_{p_{0ijk}}$ (Pupil level)	31.552	(4.468)		13.878	(1.738)		91.807	(15.139)	
Number of schools	112			101			106		
Number of classes	309			278			233		
Number of pupils	2,549			2,053			396		
Log likelihood	-8092.940			-5726.875			-1824.677		

[a] Robust standard errors, adjusted for the school-identifier; [b] One-sided p-values, * $p < .05$, ** $p < .01$, *** $p < .001$.

Table 4.6 shows that a two-standard-deviation increase in the percentage of co-ethnic classmates increases the percentage of co-ethnic friends online by 2.1 percent ($p < .01$). This relationship seems to be driven by Dutch majority members, because this variable is statistically significant for majority members ($p < .05$) but not for ethnic minority members ($p > .05$). Additionally, the percentage of co-ethnic schoolmates is associated with the percentage of co-ethnic friends in online networks for all respondents. A two-standard-deviation increase in the percentage of co-ethnic schoolmates increases the percentage of co-ethnic friends in online networks by 13.1 percent ($p < .001$). A two-standard-deviation increase in the percentage of same-gender classmates increases the percentage of same-gender friends online by 1.7 percent ($p < .001$). Additionally, a two-standard-deviation increase in the percentage of same-gender schoolmates increases the percentage of same-gender ties in online networks by 1.1 percent ($p < .001$). Given these results, we can conclude that the ethnic and gender composition of foci has a positive effect on the ethnic and gender homogeneity found in large online networks (consistent with Hypothesis 2).

4.5.3 The Interplay between Meeting Opportunities, Homophily, and Balance

Number of online network friends. We can now compare the difference in ethnic and gender homogeneity between strong versus weaker ties (Hypotheses 3c, and 3d). Before doing that, however, we will first consider the relationship between network size and ethnic and gender homogeneity in online networks (Hypotheses 3a and 3b).[22]

Table 4.6 shows that ethnic minority adolescents who have larger online networks also have a lower percentage of co-ethnic friends. For each 100 extra online network friends, the percentage of co-ethnic friends decreases by .7 percent. A two-standard-deviation increase in the number of friends decreases the percentage of co-ethnic online network friends by approximately 2.5 percent. For majority-group

[22]Correlations between degree and individual properties may happen by chance. By design, there may be a negative correlation between degree and individual-level clustering in social networks. The friends of high-degree individuals are less likely to be linked than are friends of low-degree individuals (Jackson, 2008). We tested whether the correlation between homogeneity and degree originates from design. We randomly rewired the ties of Facebook networks while keeping individual degree constant; we found no correlations between degree and homogeneity in this random mixing model. We therefore assume that H_0 is $r(\text{degree, homogeneity}) = 0$ (for more details, see Appendix 4.6).

Table 4.7: Multilevel model estimating the percentage of same-gender friends in online networks.

	Coef.	S.E.[a]	p[b]
Fixed part			
Intercept	58.901	(3.020)	***
Core-network			
Same-gender$_{\text{FRIENDS IN IN CLASS}}$	0.036	(0.005)	***
Opportunity			
Same-gender$_{\text{IN CLASS}}$	0.039	(0.010)	***
Same-gender$_{\text{IN SCHOOL}}$	0.031	(0.013)	**
Ethnicity (ref.: Dutch)			
Turkish	11.142	(1.668)	***
Moroccan	3.079	(2.985)	
Dutch Caribbean	-0.942	(1.134)	
Other Western	0.418	(0.555)	
Other non-Western	4.090	(0.830)	***
Number of Facebook friends	-0.016	(0.002)	***
Facebook membership (ref.: 2013)			
2012	-3.323	(2.855)	
2011	-3.092	(2.645)	
2010	-3.139	(2.635)	
2009	-3.345	(2.711)	
2008	-2.442	(2.795)	
2007	-2.500	(3.029)	
2006	-3.038	(4.052)	
Girls (ref.: boys)	-1.075	(0.588)	*
Educational track (ref.: lower voc.)			
Senior General	1.062	(0.529)	*
University preporatory	0.006	(0.563)	
Indegree popularity	-0.014	(0.013)	
Gender role attitudes	-0.273	(0.155)	*
% kinship ties declared	0.121	(0.157)	
% similar surname on Facebook	-0.211	(0.151)	
Random part			
$\sigma^2_{s_{0k}}$ (School level)	1.225	(0.641)	
$\sigma^2_{c_{0jk}}$ (Class level)	0.000	(0.000)	
$\sigma^2_{p_{0ijk}}$ (Pupil level)	77.957	(4.366)	
Number of schools	109		
Number of classes	302		
Number of pupils	2,598		
Log likelihood	-9361.166		

[a] Robust standard errors, adjusted for the school-identifier; [b] One-sided p-values, * $p <.05$, ** $p <.01$, *** $p <.001$.

adolescents, the number of friends and the percentage of co-ethnic friends are not related ($p > .05$). The negative association between the number of friends and the percentage of co-ethnic friends is significantly stronger for minority members than for Dutch majority adolescents ($p < .001$; tested as the product of a dichotomous variable for Dutch/non-Dutch ethnic background and the number of friends). This result suggests that a larger online network coincides with lower ethnic homogeneity only among ethnic minority members (consistent with Hypothesis 3a).

Table 4.7 shows that when the number of online network friends increases, gender homogeneity decreases. A two-standard-deviation increase in the number of online network friends decreases the percentage of same-gender friends online by 5.7 percent ($p < .001$), which suggests that gender homogeneity is stronger among smaller online networks than among larger online networks (consistent with Hypothesis 3b).

Figures 4.2 and 4.3 depict the relationships between network size and ethnic and gender homogeneity in online networks. Figure 4.2 shows that the percentage of co-ethnic friendships online decreases when the number of friends online increases for ethnic minority adolescents. For ethnic minority adolescents, the number of friends in online networks is negatively correlated with the percentage of co-ethnic friends online ($r = -.343$; $p < .001$).

Figure 4.3 shows that network size and the percentage of same-gender friendships in online networks is negatively correlated for both boys and girls ($r = -.312$; $p < .001$).

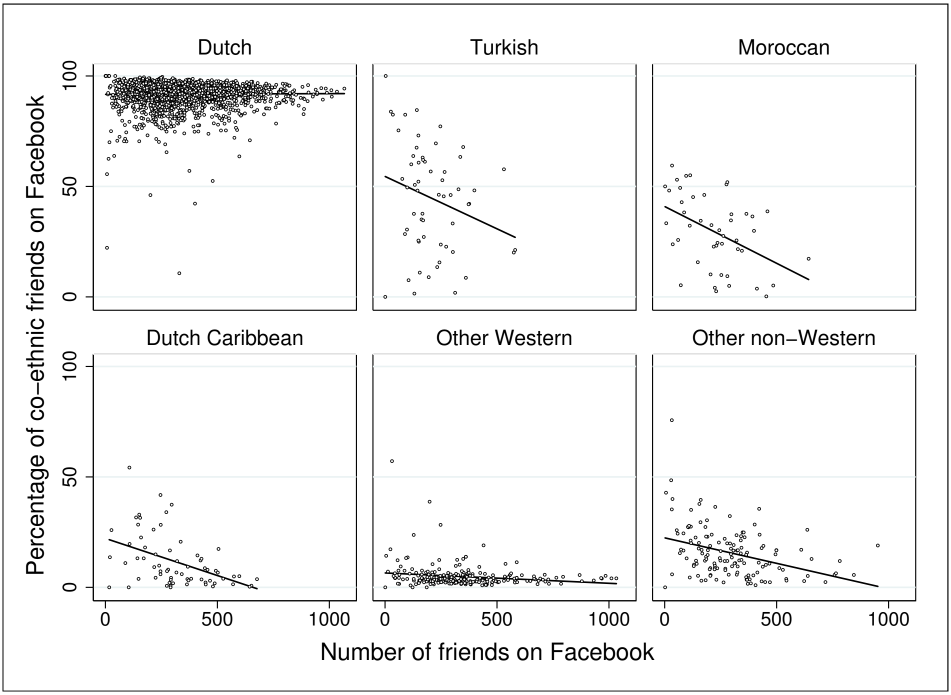


Figure 4.2: Ethnic homogeneity of large online networks by number of friends, broken down by ethnicity and including a fitted regression slope.

Self-reported core networks and online networks. Next, we contrast ethnic and gender homogeneity in self-reported core networks and larger online networks (Hypothesis 3c). To do so, we subtract ethnic homogeneity in online networks from ethnic homogeneity among friends in general (co-ethnic$_{\text{FRIENDS IN GENERAL}}$ – co-ethnic$_{\text{FACEBOOK}}$; results do not vary if we use co-ethnic$_{\text{FRIENDS IN CLASS}}$) and estimate the differences across ethnic groups between these factors in a multilevel regression model (see Table A4.2 in Appendix 4.4). Members of minority groups have at least 26 percent more co-ethnic friends among their core friends than among their online network friends than do Dutch majority members ($p < .001$; we excluded self-reported core networks because these are part of the dependent variable in this model). Hence, we find evidence supporting our argument that ethnic minority members have higher levels of ethnic homogeneity in their core networks than in their larger networks, whereas we find no such association for ethnic majority members.

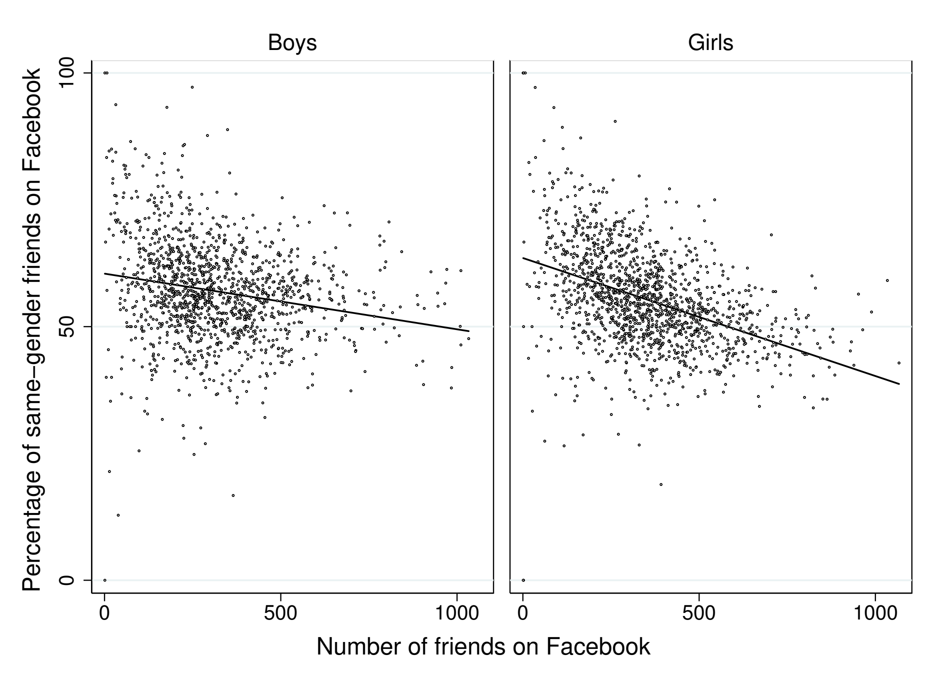


Figure 4.3: Gender homogeneity of large online networks by number of friends, broken down by gender and including a fitted regression slope.

We also consider whether gender homogeneity in core networks is higher than in online networks (Hypothesis 3d). We estimate a multilevel model where the dependent variable is the difference between the percentage of same-gender friends in the core and online networks (same-gender$_{\text{FACEBOOK}}$ – same-gender$_{\text{FRIENDS IN CLASS}}$). The intercept statistically deviates from zero (intercept = 29.995; $p < .001$; see Table A4.3 in Appendix 4.5; we excluded core networks because these are part of the dependent variable), which suggests that gender homogeneity among core ties, as measured from the survey, is approximately 30 percent higher than gender homogeneity in online networks, with all other variables kept constant. The designation of gender homogeneity in classes and school as covariates does not explain away the difference. This finding suggests that, compared to larger online networks, smaller core networks tend to be more gender homogeneous (consistent with Hypothesis 3d).

4.5.4 Confounding Factors

We examined these differences while adjusting for a variety of factors, as shown in Tables 4.6 and 4.7. Although the percentage of co-ethnic online network friends does not differ significantly between girls and boys, girls have slightly fewer same-gender online ties than do boys. Turkish minority members also have significantly more same-gender online network friendships than do Dutch majority members. Furthermore, ethnic minority adolescents who are more highly educated have fewer co-ethnic online network friendships than do their less-educated counterparts, and students in a senior general educational track have more same-gender ties online than do students in the lowest educational track. The percentage of co-ethnic friends in general and same-gender friends in a class is also positively associated with ethnic and gender homogeneity in online networks. Dutch majority members who have more positive ethnic out-group attitudes have fewer co-ethnic friends online (implying that they are more likely to connect to minority members), and adolescents who hold more progressive gender role attitudes have fewer same-gender friends online. Finally, adolescents who have more friends in their online networks with the same surnames have a higher percentage of co-ethnic friends online (consistent with the idea that kin ties increase ethnic homogeneity).

4.6 Conclusions and Discussion

We aimed to answer three key questions that are unresolved in the literature on segregation in social networks: How high are segregation levels in large online networks? Under what conditions does this segregation occur? And how can we explain disparities in segregation between core and larger networks? We show that digital footprint data from online social networks, specifically Facebook, can be used to obtain novel and robust tests of predictions derived from seminal theories of the determinants of segregation in personal networks.

The answer to our first question is that we find high levels of ethnic segregation in online networks. Averaged over all respondents, we find that approximately three-quarters of respondents' Facebook friends are of a similar ethnic background. This ratio is on par with ethnic homogeneity in core networks. However, if we split these estimates by ethnic group, only majority members' core and online networks are equally ethnically homogeneous, whereas minority members have lower levels of ethnic homogeneity in their online than in their core networks. Slightly more than half of online networks friends are the same gender as respondents, whereas in the core networks, the ratio was well above 80 percent. Second, under what conditions do these patterns of segregation occur in online networks? In the tradition of Blau (1977a, 1977b), Feld (1981), and others who have studied the role of meeting opportunities in the genesis of core ties (e.g., Kalmijn and Flap, 2001; Mouw and Entwisle, 2006; Wimmer and Lewis, 2010), we found that relative group size and social foci are strongly associated with segregation in larger personal networks. Specifically, large networks tend to mirror the structural features of the population and foci. The gender distribution in a population is often 50/50, whereas the distribution of ethnicities is much more unequal. Therefore, we hypothesized and confirmed that gender homogeneity is lower than ethnic homogeneity in online networks. Because ethnic majority members have more opportunities to meet similar others, we expected, and found, that ethnic majority members, compared to ethnic minorities, have higher levels of ethnic homogeneity in their large personal networks. Groups in society segregate over foci and the ties that emerge within them (Feld, 1981). Therefore, personal networks resemble the levels of segregation of foci. We thus hypothesized, and found, that homogeneity in foci is positively related to homogeneity among friends online.

Third, we hypothesized and corroborated that as network size increases, larger online networks are characterized by lower gender homogeneity, and that among ethnic minority groups, as online network size increases ethnic homogeneity decreases.

Our results are in line with the propositions that core ties are more segregated than weaker ties (e.g., Granovetter, 1983; Blackwell and Lichter, 2004; Son and Lin, 2012), that dyadic similarity fosters tie strength because returns on investments are more likely (Windzio and Bicer, 2013; Leszczensky and Pink, 2015), and triadic closure is more pronounced among homogeneous triads. Personal networks initially mirror the features of meeting opportunities, but over time, similar dyads are more likely to become stronger bonds, whereas weak ties will continue to reflect features of meeting opportunities. Ethnic majority members have limited opportunities to befriend dissimilar others, as reflected in core and larger networks that are equally ethnically homogeneous.

4.6.1 Limitations of this Study

Four shortcomings of this study merit acknowledgment. First, the data we used might not be perfectly representative of the overall Dutch adolescent population, due to attrition rates between waves (unit non-response) and selectivity in which respondents are more likely to maintain a public Facebook friend list (item non-response). Future studies might utilize representative samples to generalize our findings to entire adolescent populations, to other age groups, and to other nations. However, when we estimate statistical models that account for (at least some of) this selectivity, we do not find qualitatively different results than those presented by the main analyses, nor do other model specifications (e.g., fixed effects for classes or schools) lead to different results.

Second, predicting the ethnic background of online network friends by their names with our machine-learning algorithm may be an imperfect method, with the potential to misclassify individuals' ethnic backgrounds. More precise measurements of ethnicity among online network friends may be needed to establish robust evidence. One way to address this issue is to collect the birthplace of friends in the data and infer their ethnic background from these birthplaces. Nevertheless, there is a strong correlation between ethnic background and names (Mateos et al., 2011), and we find evidence for this in our data, especially for respondents of Dutch, Turkish, and Moroccan backgrounds. Limiting our analyses to these groups did not alter our results.

Third, one might argue that Facebook networks do not capture respondents' *complete* networks. There may be selectivity in the Facebook friends of the analyzed respondents, and we may have missed something specific to these friends. For instance, individuals may add to their network contacts on Facebook only others with whom they most closely relate, which could potentially bias the results toward segregation. However, at a mean network size of 336, we have, at the very least, provided insight into a *large portion* of the complete personal network, and most certainly a larger portion of networks than has previously been investigated, as further confirmed by studies on the overall size of social networks (see footnote 3 of this study).

Fourth, our measurement of kin ties likely has measurement error. Kinship tags in online networks can potentially lead to false positive kin matches, as adolescents may tag non-kin as kin. We partly solved this issue by considering realistic tags among adolescents — that is, by not considering granddaughters or grandsons.

Nevertheless, some kin tags may seem plausible (e.g., sibling tags) but are not kin. Despite this limitation, we believe that the number of kin tags on Facebook is correlated to the number of real family members present among kin tags — that is, the number of accurate kin tags exceeds the number of fictive kin tags. Some indication for this is the negative correlation between the number of kin tags and gender homogeneity online. However, future research may consider the strategic use of fictive kinship tags in online networks. Additionally, by determining kin based on shared surnames among online friends, we may miss kin who have a different surname than the respondent. However, the number of friends sharing a respondent's surname is positively correlated to ethnic and gender homogeneity on Facebook. This provides some evidence that this variable is a good proxy for the relative amount of kin in online networks. Future research may consider both parents' surnames, as we potentially miss kin ties to mothers' families since children more often share fathers' surnames. Despite these limitations, with our two measurement approaches to kin, we innovatively corrected for kinship ties in estimating segregation in large online networks.

4.6.2 Implications and Future Research

Our study elaborates the work of DiPrete and colleagues (2011), one of the few studies to investigate segregation in larger networks. DiPrete and colleagues used US survey data, whereas we used Facebook data from Dutch adolescents. Despite these differences, some of the conclusions and intuitions related to DiPrete and colleagues' findings are upheld in this study; we found similar ethnic-racial segregation in core networks and larger networks (at least when the estimates by ethnic-racial groups are not split). DiPrete and colleagues (2011: 1271) speculated about its causes, stating that meeting opportunities "do not play a strongly integrative role in contemporary [...] society." We aimed to use "imaginative strategies [...] to determine the individual and structural factors that can explain heterogeneity in segregation across individuals" (DiPrete et al. 2011: 1273). Based on expectations and considering the relative group sizes of ethnic groups and both genders (Blau, 1977a, 1977b), and the segregated nature of foci (Feld, 1981), our study confirms that meeting opportunities partially drive segregation among hundreds of contacts on Facebook.

Our findings of different levels of segregation between core networks and larger networks (e.g., Granovetter, 1973, 1983; Putnam, 2000; Blackwell and Lichter, 2004; Son and Lin, 2012) seem to contrast with DiPrete and colleagues (2011).

However, when we take into account the full range of tie-formation mechanisms, the contrast is attenuated. DiPrete and colleagues (2011) studied segregation along religious, political, socioeconomic, and racial lines — characteristics along which social settings segregate — and consequently found that both core networks and larger networks are segregated. However, we found similarity in ethnic segregation only among core ties, and we found that only among ethnic majority members core and large networks have similar levels of ethnic segregation. We specifically found that ethnic minority members have far higher levels of homogeneity among their core networks than among their larger networks, and gender homogeneity is significantly higher among core networks than among larger networks. This finding confirms the speculation (e.g., Granovetter, 1973, 1983) that weaker ties are less segregated than core ties, but it also explains the findings of DiPrete and colleagues (2011). Moreover, disparities in segregation between core ties and weaker ties occur only under specific circumstances as a consequence of the interplay among meeting opportunities, homophily, and balance.

Because our study uses data on adolescents, we should be careful in generalizing our results to a broader (adult) population. However, the tie-formation mechanisms we consider are not unique to the adolescent population, and many studies show opportunity effects on segregation in other target populations (e.g., Kalmijn and Flap, 2001; Mollenhorst et al., 2008). Therefore, we conjecture that our results generalize to different target populations, such as employees, as they do for networks measured from name generators among employees (e.g., Feld, 1982; Ibarra, 1995). Nevertheless, it may be difficult to empirically observe the patterns we found, given that adolescents' social contexts are well defined (see, e.g., Mollenhorst et al., 2008: 62) and that schools are a major focus of tie formation (Coleman, 1961; McPherson et al., 2001).

We acknowledge that cross-sectional analyses cannot be used to establish causal direction. Nevertheless, given that the relationship between homogeneity in online networks and opportunity is robust (while controlling for many confounders), we tentatively assume that similar results will be replicated using longitudinal data. We recommend that future research considers segregation in large personal networks over time and settings as a next step to obtain potential causal estimates.

Another area for future research is feedback effects: What are the effects of large personal networks online vis-á-vis core networks offline? Will weak ties characterized by dyadic similarity turn into strong ties, and what online behaviors (e.g., Facebook wall posts) will lead people out of the mode of segregated large online networks?

Finally, future research may examine whether the implications of core network segregation, such as out-group attitude formation, result from segregation in online networks. Contact theory predicts that having more contact with out-group members reduces prejudice toward them (Allport, 1954; Pettigrew and Tropp, 2006). Does this apply to the relationship between ethnic segregation in large online networks (on Facebook) and ethnic prejudice?

Chapter 5

How Large Are Extended Social Networks? Combining Evidence from Facebook and the Scale-Up Method[1]

Abstract: *Sociological literature mainly focuses on explaining individual variation in the number of core social contacts. Social networks, however, reach far beyond the small number of core contacts, yet we know little about individual variation in the size of extended social networks. In this study, we first aim to illustrate how large extended social networks are. Thereafter, we explain individual variation in the size of extended social networks and propose a new method to estimate their size. Specifically, we combine survey data among Dutch adolescents (N = 2,151) on the network scale-up method with information on their number of Facebook friends. We show that the extended social network measured as the number of Facebook friends is approximately 379. The predicted average extended social network size of our new measure is approximately 524. We hypothesize and corroborate that those who spend more time in foci have larger extended social networks. Furthermore, we hypothesize and confirm that ethnic minority members, those who have more co-ethnic classmates, girls, higher educated, and those in a romantic relationship have larger extended networks as measured via Facebook than their counterparts. We find no differences by ethnicity, gender, education, and romantic relationship status considering our new measure. We discuss the implications of these findings, elaborate their discrepancies, and suggest directions for future research.*

[1] This chapter is a manuscript in progress to be submitted to an international scientific journal. Bas Hofstra is the first author of this chapter, but the chapter presents joint work with Rense Corten, Frank van Tubergen, and Jeroen Weesie. Hofstra wrote the main part of the manuscript and coordinated the Facebook data collection. Hofstra and Weesie jointly conducted the analyses. Corten, Van Tubergen, and Weesie substantially contributed to the manuscript. The authors jointly developed the idea and design of the study. I thank Tyler H. McCormick for generously providing code for the network scale-up method from McCormick et al. (2010). Furthermore, I value the feedback of Manja Coopmans, Jesper Rözer, Robert Krause, and Jan Kornelis Dijkstra on an earlier version of this article. This study benefited from discussions at the "1st International CILS4EU User Conference" in Mannheim.

5.1 Introduction

How strongly are individuals socially connected to others? And why are some individuals socially isolated, whereas others are highly connected? These questions have been addressed in a substantial body of work devoted to explaining individual variation in the number of people with whom individuals most closely relate (e.g., Parigi and Henson, 2014; Marsden, 1987, 1988; McPherson et al., 2006).

Findings suggest that people have close ties with only a few others. One source of evidence originates from studies on people's "core discussion network." In this approach, individuals are asked with whom they discuss "important matters." Using this approach, scholars have attempted to explain individual variation in the number of core discussion partners (e.g., Burt, 1984; Marsden, 1987, 1988) and changes in social isolation (cf. McPherson et al., 2006). Findings from 1984, 2004, and 2008 (McPherson et al., 2006; Hampton et al., 2011) revealed that American adults had an average of approximately two to three core ties.[2] In 2000, 2007, and 2010 (15-45 year-olds), the Dutch, on average, reported fewer than three core discussion partners (Mollenhorst et al., 2014; Van Tubergen, 2014, 2015). If we consider the close personal contacts one step beyond these closest ties, Americans had, on average, approximately 17 alters with whom they had "trusting" relationships in 2006, i.e., people who could be considered good friends, discussion partners, or trusted for advice or with money (DiPrete et al., 2011).

Despite this interest in the size of social networks, existing sociological literature on individual variation in social network size almost exclusively focuses on the number of network contacts with whom individuals most closely relate. However, individuals' social circles reach far beyond this small circle of five core discussion partners or fifteen trusting relationships. Social networks consist of *layers of social relationships* (Dunbar, 1998; Zhou et al., 2005). First, there are the two to five *core* (e.g., McPherson et al., 2006; Mollenhorst et al., 2014) and 15-17 *trusting* social relationships (e.g., DiPrete et al., 2011; sometimes called the *sympathy group*, Dunbar, 2016). The two layers beyond that are the 50 and 150 acquaintances,

[2] There has been a controversy around the increase between 1985 and 2004 in the number of cases in which Americans had zero alters to discuss "important matters" with, as discussed in McPherson et al. (2006). An erratum (McPherson et al., 2008) corrected a coding error in the original data release. In 2009 (McPherson et al., 2009; Fischer 2009), there was a discussion over whether the trend was a data artifact resulting from respondent fatigue or training. Paik and Sanchagrin (2013) show that the increase in social isolation may be attributed to interviewer effects.

where the 150 contacts can be argued to be the number of "valued others" with which people have stable relationships (Dunbar, 2016).[3] Beyond that is something we define as the *extended social network* — i.e., the relationships beyond the 150 valued others. This layer includes all former layers and is the main focus of this study. We know little about the individual differences that explain the size of this extended social network, which thus includes both the closest contacts and the much larger set of *weaker* ties.

This research lacuna is surprising, given that the extended social network size relates to a myriad of issues that are sociologically relevant. One classic example is that having more acquaintanceship ties facilitates access to resources such as information embedded in social networks (Granovetter, 1973, 1983; Burt, 2000; Putnam, 2000; Van der Gaag and Snijders, 2005). Insight into the extended network size can thus shed light on how individuals accumulate social capital, which is in itself an understudied topic (cf. Van Tubergen and Volker, 2015). Additionally, larger networks are associated with better health and well-being, more social support, and lower mortality risks (Wellman and Wortley, 1990; Shye et al., 1995; Smith and Christakis, 2008; Holt-Lunstad et al., 2010; Holt-Lunstad et al., 2015; Hobbs et al., 2016). The extended social network size in this body of work is oftentimes measured using indirect proxies. In these studies, unavoidable measurement error in these proxies might bias relationships between the other variables and may distort conclusions. Some even argue that network size has driven the evolution of human intelligence, as it evolved as a means to survive and reproduce in increasingly complex social systems (Dunbar, 1998).[4]

But how can we explain individual variation in the extended network size? As Kadushin (2012) puts it, "[we] do not as yet have a theory or a systematic study of the causes of these variations" (p. 72). The main reason for this lack of knowledge might be that it is not at all straightforward how one can measure the extended number of contacts. Scholars acknowledge that it is challenging to measure the number of core contacts (e.g., Bearman and Parigi, 2004; Brashears, 2014). It stands to reason that measuring the number of ties beyond the closest contacts becomes even more challenging. For instance, it might be difficult to decide on a

[3]This is also the average number of stable relationships humans are argued to be able to cognitively manage based on the human brain's amygdala volume or neocortex size (Dunbar, 1993; Dunbar, 1998; Bickart et al, 2011; Kanai et al., 2011) — this is the so-called "Dunbar's Number."

[4]This argument is occasionally referred to as the "Machiavellian Intelligence Hypothesis" or the "Social Brain Hypothesis" (Dunbar, 1998).

relevant network boundary (i.e., at what distance from a person do ties need to be for inclusion in such a measure?), and individuals may find these more-distant ties more difficult to remember in surveys as compared to the closer ties.

In this study, we use a definition of extended social networks given by McCarty et al. (2001: 29) and, more recently, DiPrete et al. (2011: 1242). We consider "all the contacts whom individuals know on a first name basis to be part of the social network, such that they would have a friendly chat if they were to meet randomly." This definition includes both the core ties and the much larger set of acquaintanceship ties. This definition — knowing first names and having a chat — relates well to the societal outcomes discussed before. These contacts may grant a meaningful connection — in terms of knowledge — to other unknown people. The extended network size from this definition also identifies the people who are able to activate ties for social support, because the mere *availability* of alters is an important condition to seek them out for help or probe for advice (Small and Sukhu, 2016).

There is a growing body of literature concerned with establishing methods to provide estimates of the size of networks beyond core networks (e.g., Killworth et al., 1998; McCarty et al., 2001; McCormick et al., 2010). Contributions to the development of such measurements have provided vastly different estimates of the average network size, ranging from 108 (Killworth et al., 1998) to 5,520 (Freeman and Thompson, 1989). Methods included asking respondents whom they knew from randomly drawn pages from phonebooks (e.g., Pool and Kochen, 1978), using a summation method counting how many people respondents indicated they knew from a list of given relationships (e.g., neighbors; McCarty et al., 2001), or counting the number of Christmas cards respondents send out (Hill and Dunbar, 2003). The differences in estimates often reflect discrepancies in definitions on what constitutes the network. For instance, Hill and Dunbar (2003) studied "valued others," those contacts important enough to send a Christmas greeting card to, whereas Pool and Kochen (1978) counted all of the contacts people remembered knowing but who were not necessarily important enough to send Christmas cards to.

Here, we focus on two more recent methods that provide estimates of the extended network size beyond the close ties. First, using survey data, the "network scale-up method" provides extended network size estimates (e.g., Zheng et al., 2006; McCormick et al., 2010). In this approach, respondents are asked how many people they know from various subpopulations (e.g., people named "Kevin") to measure the social network size. One can then "scale-up" the network size by calculating the share of people one knows from a subpopulation. The network size definition often used in this approach is similar to ours. Estimates of extended network sizes using this method are within the range of 550-750 (Zheng et al., 2006; McCormick et al., 2010; DiPrete et al., 2011), and the degree distribution is highly skewed; many people have few ties, whereas few people have many ties (see DiPrete: 1254).

The second approach considers online data and measures the number of "friends" individuals have on their social media profiles (e.g., Gonçalves et al., 2011; Kanai et al., 2011; Pollet et al., 2011; Dunbar et al., 2015; Dunbar, 2016). Findings, primarily from the microblogging website Twitter (e.g., Gonçalves et al., 2011) or from the social network site Facebook (e.g., Dunbar, 2016), show that the average extended social network size as measured on social media is approximately 180-200, on average (Gonçalves et al., 2011; Pollet et al., 2011; Dunbar et al., 2015; Dunbar, 2016).

Both of these approaches, however, have their limitations. Measuring extended social networks via the network scale-up method in surveys causes recall biases (Brashears et al., 2016). Specifically, individuals often overestimate their number of contacts from smaller-sized groups, and the network scale-up method is sensitive to such biases, which renders it highly variable (McCormick et al., 2010). A limitation of measuring the extended social network size via social media is that, although online networks capture a large subset of complete social networks (e.g., Wimmer and Lewis, 2010; Duggan et al., 2015), we do not yet know *how large of a subset*. The discrepancy in findings from the number of social media friends and the number of contacts in the scale-up method suggest that there is variation in which contacts are added as friends on social media, but we do not know how this varies from person to person.

We contribute to the literature in two ways. First, we propose a methodological step forward. We take a two-step approach in which we first consider the number of friends online and, thereafter, propose a *new* measure of the extended network size. The new measure integrates the number of online friends and the network scale-up method. It starts from the conjecture that the extended network size is an unobserved, latent trait. We do observe the number of online network friends and their names. We compare the number of people from specific subpopulations whom respondents indicate they know in the scale-up method (e.g., the number of people named "Kevin") to the number of people from these subpopulations among the online networks. Assuming that the prevalence of these subpopulations is similar in the extended social network size and in online social networks, we can extrapolate how many people there are beyond those of the specific subpopulations, i.e., how many people there are in the extended social network. Our study thus contributes to the ongoing debate on how to measure individuals' extended network size. As a byproduct of our new approach, we can estimate which individuals add a larger share of their extended social network contacts as friends on social media. This latter contribution is useful for social network analysts who consider online networks as an approach to study social networks in general.

Second, we aspire to answer two substantive questions. We first aim to estimate how large the extended social network is, considering both the online social network size and the network size from our new method. Thereafter, we aim to explain individual differences in the extended social network size. The existing literature does not elaborate on how individual variation in the extended network size can be explained (and what we know is likely distorted because of measurement issues). Hence, we contribute to prior work as we move beyond hitherto knowledge on individual variation in the number of core contacts. As the theoretical point of departure, we engage classic theories on opportunities, preferences, romantic partners, and intuitions on education and gender occasionally used to explain the number of core ties (e.g., Blau, 1977a; Feld, 1981; McPherson et al., 2001). We tie them to predictions on the size of the extended network and show the relevance of fundamental prior hypotheses in the study of social network formation among weaker ties.

We link survey data on Dutch adolescents in 2012 (Kalter et al., 2015) to two data sources collected in 2014, namely, (1) to survey data measuring these respondents' network size using the scale-up method and (2) to these respondents' observed Facebook profiles, in order to measure their number of Facebook friends (Hofstra et al., 2015a; Jaspers and Van Tubergen, 2017). Because the vast majority of Dutch adolescents are Facebook members (Hofstra et al., 2016a) and Facebook is by far the largest social network site worldwide (Facebook, 2017), this sample is a suitable starting point to develop and illustrate our method. In the remainder of this study, we first develop explanations on individual variation in the extended network size. Thereafter, we elaborate our procedure to estimate the extended network size and, finally, test our explanations using this procedure.

5.2 Theory and Hypotheses

What follows is an explanation of individual differences in the extended network size. We start from theories occasionally used for explaining the emergence of core ties, i.e., theories on opportunities, homophily, and romantic partners, and we explore differences by education and gender.

5.2.1 Opportunities and Homophily

Foci. The role of opportunity is important in the genesis of core social ties (e.g., Blau, 1977a; Feld, 1981). *Opportunities* refer specifically to the possibilities individuals have to meet contacts. One dimension of opportunity theory is the concept of *foci*. A focus is a "social, psychological, legal, or physical entity around which joint activities are organized" (cf. Feld, 1981: 1016). Typical foci are associations, neighborhoods, work places, or schools. A social context contains a set of foci and individuals. In this social context, people engage in some of the foci, but not all. The idea is that those who share a focus will share activities, more so than individuals who do not share a focus. Sharing these activities creates positive sentiments and interactions between people and brings them together in reciprocally rewarding situations (Feld, 1981). In the consideration of positive ties, sharing a focus thus increases the likelihood for ties to form.

Research shows that strong ties and even many weak ties are formed in some sort of focus (Wimmer and Lewis, 2010; Hofstra et al., 2017). Many acquaintances are met, for instance, at associations, on vacation, at parties (Mollenhorst et al., 2008), at concerts (Lizardo, 2006), on campus (Wimmer and Lewis, 2010), or at school (Hofstra et al., 2017). We depart from this idea and conjecture that those who engage more often in socially and recreationally orientated foci have more opportunities to get in contact and befriend or make acquaintances with other people than those who engage less often in such foci. We thus expect the extended network size to be a function of the time individuals spend in these foci. We examine three foci in particular to capture adolescent social life: bars (e.g., going for a drink), associations (e.g., engaging in a sports club), and concerts. We hypothesize that:

> **Hypothesis 1:** *Adolescents who spend more time (a) going out, (b) in associations, and (c) visiting concerts have larger extended social networks.*

The Similarity of Potential Contacts. A substantial body of sociological literature focuses on the formation of social relationships and its relation to social network segregation (e.g., Kalmijn, 1998; McPherson et al., 2001; Mouw and Entwisle, 2006). Two of the common explanations for network segregation are structural opportunities and homophily (Kalmijn, 1998; McPherson et al., 2001; Mouw and Entwisle, 2006; Wimmer and Lewis, 2010). Results in this line of research are that racial-ethnic segregation in networks is a ubiquitous feature of social life — often caused by (the interplay of) homophily and opportunity (Wimmer and Lewis, 2010; Smith et al., 2014a). Given that racial-ethnic segregation in social networks is pervasive, we argue that the interplay of homophily and opportunity also relates to the extended network size.

Homophily refers to preferences of individuals to form relationships with similar others (McPherson et al., 2001). A second dimension of opportunity — besides the concept of foci — is the size of groups relative to other groups (Blau, 1977a). The interplay between homophily and relative group size may cause disparities in the extended network size between ethnic minority and majority members in society. When people belong to a larger group, they have more opportunities to select members from within their own group, whereas minority members have fewer possibilities to make these homophilous choices. Ties among ethnically dissimilar people are costlier because they require higher initial investments to overcome cultural boundaries (Kalmijn, 1998), i.e., it takes more time to get to know one another. Assuming that there is a limit on the investments one can make (e.g., in terms of time or emotional commitment), fewer ties are formed if the pool of potential alters includes relatively more dissimilar people. As a result, ethnic minorities establish fewer ties among their extended network than ethnic majority members. Furthermore, ties that are established — assuming that among ethnic minority groups ties are, on average, more often between dissimilar people — will be broken more frequently (cf. Smith et al., 2012).

In the Netherlands, approximately 79% of the population are so-called “Dutch majority” members (Statistics Netherlands, 2015). Much smaller in relative size are ethnic minority groups with an immigrant background. Minorities with a Turkish or Moroccan background, for instance, cover approximately 6% of the

Dutch population. Hence, Dutch majority members — for whom we assume that they prefer befriending other Dutch majority members — have ample possibilities to choose similar others as friends. Members of the ethnic minority have far fewer such opportunities and will end up with fewer contacts — the potential group from which they prefer choosing their contacts is much smaller. Hence, relative differences in group sizes will be reflected in an extended network size that is significantly larger for those people who belong to an ethnic background for which the fraction in the population is larger than for those whose fraction is smaller.

This specific interplay between homophily and group size was shown to affect the number of Twitter connections (i.e., a microblogging website: Halberstam and Knight, 2016). We thus hypothesize, based on the combination of the homophily and opportunity arguments, that:

> **Hypothesis 2:** *Dutch majority adolescents have larger extended social networks than adolescents of the ethnic minority groups.*

The Similarity of Potential Contacts in Foci. Differences in the extended network size may also result from the ethnic segregation of social settings, i.e., people of particular ethnic backgrounds self-organize themselves over foci in a non-random way. As foci increase the likelihood of a tie emerging between two people sharing a focus, this likelihood will be higher when potential contacts in foci share commonalities such as ethnic background. This is based on homophily theory, departing from the conjecture that individuals inherently prefer befriending others with whom they share commonalities (i.e., especially race or ethnicity) (McPherson et al., 2001). Hence, when there are many individuals sharing an ethnic background with an adolescent in a focus, this person will more likely form more ties, as he/she has ample possibilities to make homophilous choices. Here, we focus on the school setting as a focus of tie formation among adolescents (McPherson et al., 2001). Unfortunately, we do not have information on the ethnic composition of the foci mentioned before (e.g., associations), nor is adolescents' school-of-choice completely exogenous. Some of the relative group size effects may result from non-random sorting of adolescents over foci that we do not address. We investigate the number of potential alters in schools and school classes who share an ethnic background with the respondent, as schools are important foci for tie formation among adolescents (McPherson et al., 2001). Thus, we hypothesize:

Hypothesis 3: *Adolescents who have a greater number of co-ethnic individuals in a) their schools and b) in their school classes will have larger extended social networks.*

5.2.2 Romantic Partners.

We consider the role of romantic partners. Previous research has shown that individuals who are in romantic relationships have smaller core networks (Kalmijn, 2003, 2012; Song, 2012; Rözer et al., 2015). This social withdrawal is related to individual's limited resources to maintain core contacts — people have only a limited amount of time, energy, and emotional capacity to maintain their relationships with many close contacts besides their romantic partners (Slater, 1963; Kalmijn, 2004; Rözer et al., 2015).

However, when you befriend someone, especially when you enter a romantic relationship, he/she may introduce you to many new *acquaintances*. These new acquaintanceships are less costly to maintain than stronger ties, as they involve less emotional intensity and intimacy and less time (cf. Granovetter, 1973). The newly acquainted others introduce you to other new people, and so on. This introduces a bandwagon effect where some people are in a more advantageous position to obtain new contacts continuously as compared with others. Romantic partners thus provide ample opportunities to meet new potential contacts, under the assumption that the social networks of the romantic partners do not completely overlap (Kalmijn, 2003).

This process is related to two concepts. First, it is related to the concept of "third parties" (Kalmijn 1998) encouraging one to enter other relationships. Second, it is related to the concept of triadic closure (when A is friends with B, and A with C, then B and C are likely to connect), where the romantic partner may seek opportunities for the initially unconnected pair to connect (Feld, 1981) or may find it psychologically straining to be in an unbalanced triad (Heider, 1946). Hence, we hypothesize that:

Hypothesis 4: *Adolescents who indicate being in a romantic relationship have larger extended social networks than adolescents who are not.*

5.2.3 Education and Gender

Education. Research on "cultural omnivores" shows that some people — generally those of higher social status and education — pursue a broader range of social and cultural leisure time activities than others (e.g., Peterson, 1992; Lizardo and Skiles, 2012). Some of this discrepancy is attributed to varying cognitive capabilities; some individuals are better capable of managing a broader range of social and cultural leisure time activities than other individuals. Hence, some of the variation in the finding that those of high status and education engage in a broader range of activities is a byproduct of them being cognitively able to engage in such a diversity of activities. This may be one reason for a larger network size among higher-educated individuals — they engage in a broader range of social foci and, therefore, have more opportunities to befriend others than lower-educated individuals. The limited set of social foci we addressed before likely do not completely adjust for this indirect association.

Net of this indirect association, however, cognitive abilities are an additional factor in the formation and maintenance of social ties (e.g., Dunbar, 1993; Bickart et al., 2011). Specifically, the amygdala volume in brains is correlated with network size and reflects cognitive capabilities to keep track of all of the relationships one has (e.g., Dunbar, 1992; Gonçalves et al., 2011). We also assume that cognitive capabilities correlate with the extended network size. Specifically, we assume that there is variation in individual capacity to maintain and keep into contact with many social contacts. Following research on cultural omnivores, we consider educational level as a proxy for such capabilities. Hence, higher-educated individuals are expected to have larger extended networks. We thus hypothesize:

> **Hypothesis 5:** *Adolescents in higher educational track levels have larger extended networks than adolescents in lower educational track levels.*

Gender. A consistent finding is that women's social networks differ significantly from men's. Net of opportunity structures, women's core networks are larger (Moore, 1990; Bastani, 2007; Hampton et al., 2011; Van Tubergen, 2014) and include more kin (Marsden, 1987; Van Tubergen, 2014).

There may be various reasons why such differences occur. First, there is speculation that women are cognitively better equipped to manage larger networks than men. One indication of this is that women appear superior to men in recalling their social contacts (Brashear et al., 2016). This may be a result of circumstances that "shape males and females such that females develop a relatively greater ability to encode and recall social networks" (Brashears et al., 2016: 82). This finding is consistent with prior research showing that those in lower power situations have greater knowledge of their social networks (Simpson and Borch, 2005; Simpson et al., 2011), under the assumption that these circumstances imply lower socioeconomic positions of women compared to men (e.g., Bobbitt-Zeher, 2007; Chen and Volker, 2016). Second, men and women may differ in their sociality, may have different dispositions towards (the maintenance of) social ties (Brashears et al., 2016), or may differ in their engagement in social activities. Each of these mechanisms may explain why women may have larger extended networks than man have. Note, however, that we do not test these mechanisms directly but explore these intuitions. We hypothesize:

> **Hypothesis 6:** *Girls have larger extended social networks than boys have.*

5.3 Data

We use data on Dutch adolescents from a larger project, titled "Children of Immigrants Longitudinal Survey in Four European Countries" (CILS4EU) (Kalter et al., 2015). We use the Dutch part of the survey, as our measures of interest are included only in the Dutch part, although data were also collected in Sweden, Germany, and England. We use the second and fourth waves, because the second wave contains the latest school-level data for the total set of respondents and the fourth wave measures the network size using the network scale-up method and the number of Facebook friends. In these data, 14-15-year-old adolescents were followed for three years with a one-year time lag starting in 2010. A Dutch section of the survey continued for four additional years under the heading "Children of Immigrants Longitudinal Survey in the Netherlands" (CILSNL) (Jaspers and Van Tubergen, 2014; Jaspers and Van Tubergen, 2017). At the time of writing, six waves were collected, and the collection of a seventh wave of data is in progress. The surveys include many individual characteristics, personal attitudes, information on leisure time activities, and information about personal networks. The data are stratified by the proportion of non-Western immigrants attending schools. In these strata, schools were selected with a probability proportional to the school size using the number of students at the relevant educational track level.

At wave 1 (2010-2011), two classes were randomly chosen from within schools, resulting in a sample size of 118 schools, 252 classes, and 4,963 pupils participating in the Dutch survey. Classroom composition changes are common in the Netherlands. Therefore, respondents from wave 1 could be distributed among different classes at the time of the second wave. To ensure that many wave 1 respondents participated in wave 2 (2011-2012) as well, schools were asked to participate with all of the classes that wave 1 respondents were attending, even though they were distributed among new classes. Therefore, 2,118 new students were interviewed (attending the same class as original wave 1 respondents), and 3,803 students who participated in wave 1 participated in wave 2 as well (76.6%; total N = 5,921).[5] In wave 4 of the CILSNL, 4,073 respondents participated, of which 3,611 had also participated in wave 2 (88.7%).

5.3.1 The Dutch Facebook Survey

We link the survey data to behavioral data from Facebook using the Dutch Facebook Survey (DFS) (Hofstra et al., 2015a). The DFS enriches the Dutch part of the CILS4EU and the CILSNL. The DFS was collected between June and September 2014. Of the 4,864 respondents that indicated Facebook membership in waves 3 (2012-2013; N = 3,423) or 4 (2013-2014; N = 3,595) of the surveys, 4,473 (92%) were tracked on Facebook. For the respondents who kept a *public friend list*, we downloaded their complete Facebook friend lists (N = 3,373; 75.4% of all tracked

[5] Six hundred respondents in wave 1 were sampled who were not part of the random sampling frame. This was because some schools wanted to participate with more than the two randomly drawn classrooms. Therefore, a *random* sample of 4,363 pupils was established in wave 1. Because of attrition rates between waves 1, 2, 3, and 4, the representativeness of the sample cannot be guaranteed. We include as many respondents as possible in the sample for analyses, including newcomers (nonrandom) and the nonrandom sample of wave 1 to ensure a large sample size.

respondents on Facebook).[6, 7, 8]

5.3.2 Sample Selections

We longitudinally link two waves of survey data and behavioral data from Facebook. The total number of observations is 5,921 for wave 2, 4,073 for wave 4, and 3,373 for the DFS (i.e., for those who keep a public friend list). Some respondents keep their friend lists private on Facebook, whereas others keep public friend lists (see Hofstra et al., 2016b). We can only analyze those respondents for whom we can observe their friend list, as we measure the number of Facebook friends by counting the number of Facebook contacts among these friend lists. Potentially, we can analyze 2,387 respondents; this is the number of respondents who participated in waves 2 and 4 and for whom we can observe their number of friends on Facebook. Further deletion of cases with missing values on the variables of interest (which we outline below) leads to a final set of 2,151 respondents. This is the number of cases we continue to consider throughout our analyses (i.e., 9.9% item non-response).

Consequently, we encounter three types of selectivity in this set of cases. First, there may be respondent-selectivity in attrition; some respondents may be more willing to participate than others throughout the waves. Second, there may be selectivity in item non-response. Finally, there is selectivity in whose friend lists we can observe on Facebook. We deemed it computationally infeasible in our new approach to adjust for the sample selection biases in a selection model (e.g., Heckman, 1979). In our results, however, we do provide analyses based on the number of Facebook friends as a measure of the extended network size in which we aspire

[6] The data collection and use of the DFS for scientific goals were approved by an internal ethical review board for the social and behavioral sciences.

[7] The total number of Facebook friends we present in this article deviates from the total number of Facebook friends taken into account in Hofstra et al. (2017), although we use the same data source. In the latter article, we studied ethnic segregation and, as such, predicted the ethnic backgrounds of Facebook friends. This resulted in some Facebook friends not being coded and a smaller network size on Facebook. However, the number of friends here and in Hofstra et al. (2017) are highly correlated (r = .99), leading us to believe there was no or low selectivity in which friends were coded in Hofstra et al. (2017).

[8] Those who have their Facebook friend list private do show up in others' friend lists and are counted among the number of Facebook friends. Additionally, there is a privacy setting indicating that only members' "friends of friends" may invite them as friends on Facebook — those people are still counted in the number of Facebook friends.

to adjust for some of these sample selection biases and compare these results to the results of the analyses using our new method. Table 5.1 provides a summary of the various data sources and shows the conditions for inclusion in the set of 2,151 respondents which are considered in our new method.

Table 5.1: Overview of the used data sources and sample selections.

	N
Survey data (CILS4EU and CILSNL)	
W2 total number of respondents	5,921
W4 total number of respondents	4,073
W2 participation + W4 participation	3,611
Online network data (DFS)	
Respondents whose profiles were tracked on Facebook	4,463
Respondents keeping a public Facebook friend list	3,373
Merging the different data sources	
Participation W2 + participation W4 + Public Facebook friend list	2,387
Final conditions for inclusion in the number of cases to analyze in new procedure	
Participation W2 + participation W4 + Public friend list + No missing values	2,151

5.4 Predictor Variables

Foci. Each of the predictor variables is measured from wave 2 of the CILS4EU. Respondents were asked three questions on how often they spend time in socially oriented foci. They could indicate on a five-point scale (1-never, 2-less often, 3-once or several times a month, 4-once or several times a week, and 5-daily) how often they go out (e.g., bars/ nightclub/ etc.), spend time in associations (sport/music/ etc.), and visit concerts or DJs. Per respondent, we thus obtain three variables showing the time they spend in foci.

Ethnic Background. We categorize respondents into ethnic background groups according to the country of birth of their biological parents, which is standard practice in scholarship on Dutch ethnic groups (e.g., Vermeij et al., 2009; Smith et al., 2014b). When adolescents have one Dutch-born parent, they are categorized in the ethnic background category of the parent not born in the Netherlands. When respondents have parents born in different non-Dutch countries, they are categorized in the mother's birth country. This categorization is regularly applied and used by Statistics Netherlands (2012). We categorize respondents into one of three major ethnic background groups (Castles et al., 2013) — "Dutch ethnic majority," "Turkish and Moroccan ethnic minority," and "Other ethnic minority" — to ensure a large enough sample size across the ethnic background categories and to reduce computing time. Turkish and Moroccan youth are children of the labor force immigrants the Dutch government recruited in the 1950s and 1960s. The "Other ethnic minority" category includes Dutch Caribbean adolescents who originate from former Dutch colonies (e.g., Aruba and Suriname) and other Western and non-Western adolescents whose origin stems from neighboring countries such as Germany and Belgium or from conflict areas such as Bosnia, Iraq, and Syria.

Number of co-ethnic in Class. We measured the number of students in a *class* who share the ethnic background of the respondents: the number of Dutch majority members for those of the Dutch ethnic majority, the number of Turkish origin for those of the Turkish ethnic minority, etc. We used specific ethnicity categories for the "Other" ethnic background. For instance, we counted the number of Dutch Caribbean in a class for those of Dutch Caribbean ethnic background.

Number of co-ethnic in School. In a similar way, we calculated the number of *schoolmates* who share an ethnic background with the respondent (excluding classmates, as we aim to separate the effects of classmates and school mates). This variable was measured from secondary data that were obtained from the Dutch inspectorate of Education.

Romantic Partner. We measure whether the respondent indicated being in a romantic relationship (1) or not (0).

Educational Track-Level. When Dutch adolescents transition to high school, they are placed in different educational tracks that differ in their type of education and level (Van de Werfhorst and Van Tubergen, 2007). We measured this categorization with an ordinal variable: 1-preparatory vocational education (Dutch: VMBO), 2-senior general (Dutch: HAVO), and 3-university preparatory education (Dutch: VWO).

Gender. We measure whether respondents indicated they were a girl (1) or a boy (0). Table 5.2 displays the descriptive statistics of the predictor variables.

Table 5.2: Descriptive statistics for the predictor variables (N = 2,151).

	Min.	Max.	Mean	SD
Foci (H1)				
Going out	1	5	2.836	0.854
Associations	1	5	3.483	1.163
Concerts	1	5	1.969	0.710
Similarity of potential contacts (H2 + H3)				
Ethnicity				
Dutch	0	1	0.823	-
Arabic: Turkish and Moroccan	0	1	0.030	-
Other	0	1	0.146	-
Number co-ethnic class	0	28	14.810	6.660
Number co-ethnic school	0	2300	850.981	624.319
Romantic partners (H4)				
Partner				
Yes	0	1	0.262	-
No	0	1	0.738	-
Education and gender (H5+H6)				
Educational track level				
Vocational	0	1	0.483	-
Senior general	0	1	0.275	-
University preparatory	0	1	0.242	-
Gender				
Girls	0	1	0.544	-
Boys	0	1	0.457	-

Source: Survey data from the CILS4EU wave 2

5.5 Estimating the Extended Social Network Size

5.5.1 The Number of Friends on Facebook

On Facebook, members can send (and receive) friendship invitations to (from) other users, who can accept or decline the invitation. When accepted, a friendship tie within people's so-called *friend list* shows an undirected, reciprocated friendship between two Facebook users. Using the DFS, we measure the number of friends respondents have in their Facebook friend lists as the *extended network size on Facebook.*

5.5.2 The Network Scale-Up Method

The network scale-up method (Killworth et al., 1998) uses surveys to estimate individuals' extended social network size (McCormick et al., 2010). The method uses the following rationale. Consider a population of size N. To estimate network size, one can ask the number n randomly chosen members of the population an individual knows. Likely, the larger the population N is, the lower the probability becomes that two randomly chosen persons know one another.

The network scale-up method circumvents this issue by asking individuals about whether they know an entire set of people simultaneously. For instance, it asks individuals "How many people do you know that are named Thomas?" instead of asking which of the ~40,000 people they know in the Netherlands are named Thomas (Meertens Institute, 2016). When a respondent indicates, for example, that he/she knows 2 people who are named Thomas, one can calculate an estimate of the total network size by assuming one then knows 2/40,0000 of the entire population of persons named Thomas and that this same proportion equally applies to the entire population (e.g., 17 million in the Netherlands),

$$\frac{2}{40,000} \times (17\, million) = 850. \tag{5.1}$$

The precision of this estimate is increased by averaging responses to multiple categories of people, e.g., for a sample of first names and other disjoint subpopulations (e.g., detainees). This yields the basic scale-up estimator for the extended network size:

$$Scale\text{-}up\ degree_i = \frac{\sum_{k=1}^{K} y_{ik}}{\sum_{k=1}^{K} N_k} \times N, \tag{5.2}$$

where y_{ik} is the number of people person i knows in subcategory k, N_k is the size of subcategory k, and N is the size of the population (cf. McCormick et al., 2010). More generally, the subpopulations that are prompted to respondents are occasionally referred to as "How many X's do you know?," where the X's refer to the different subpopulations.

There are, however, three difficulties with the assumptions of the basic scale-up estimator from equation (2) (see Zheng et al., 2006; McCormick et al., 2010). First, there is the issue referred to as *barrier effects*: social ties are assumed to be formed completely at random, but often they are not (see McPherson et al., 2010). Second, respondents need to be perfectly aware of alter characteristics, but respondents may not always be aware of them. For instance, when respondents have to report how many police suspects they know, they may not be aware that some of their contacts are police suspects (note that contacts themselves may also not be aware that they are suspects). This is referred to as a *transmission error.*

Third, there are *recall errors*: respondents need to accurately answer the scale-up survey questions, but some respondents may not be able to.

McCormick and colleagues (2010) propose techniques to correct these biases (e.g., non-random latent modeling), although the data requirements to implement these techniques are rather high. Specifically, the X's asked of respondents should sum-up to approximately the same share of the general population that is surveyed. For instance, if 20% of the general population is adolescents with a Moroccan background, then 20% of the X's presented to respondents should also be adolescents with a Moroccan background. In the Netherlands, the diversity of first names (subpopulations that are oftentimes used as the X's) is high (Meertens Institute, 2016), which makes it difficult to come up with names that meet this condition. A further limitation is that even when this requirement is met for one condition (e.g., adolescence and ethnic background), the estimates could be biased when they correlate with other factors (e.g., age or geographical location).

Furthermore, the basic scale-up method assumes that a person's network is representative of the whole population: if a population consists of a fraction of people "A" (e.g., people named "Kevin"), this this is also the (expected) fraction of A's in a person's social network. In our approach, we do not use such "global population representativeness." Instead, we use "Facebook representativeness" and use a person's Facebook friend list. This implies that the (expected) fraction of A's in a person's extended network is the same as the fraction of A's in a person's Facebook friend list. We further refine the approach by (1) allowing multiple types A, B, ..., and (2) we assume that the fraction of friend types on Facebook depends on personal characteristics (gender, ethnicity, etc.) rather than being fully individual.

Next, we elaborate what version of the network scale-up method was implemented in the fourth wave of the CILSNL. The following statement (translated from Dutch) was shown to respondents:

> *The next questions are about all the people you <u>know personally in the Netherlands</u>. By knowing personally, we mean that you know the name of that person and that you would have a chat if you were to meet him or her on the street or in a shop.*

This statement implies reciprocal relationships, which makes it suitable for comparison with the number of friends on Facebook. Respondents were prompted with four names (Thomas, Kevin, Anne, or Melissa) on the question "How many people do you know personally with the following name?" They then had to indicate whether the number of people they knew fell within the numerical ranges of 0, 1, 2–5, 6–10, 11–20, 21–50, or more than 50. To ease the answering process for respondents and to reduce lack-of-response errors, the questions were asked using interval censoring. This same strategy was used by DiPrete and colleagues (2011: 1251). These Dutch names represent names of both genders from parents with either a higher or lower educational background: Anne (girl, high education), Melissa (girl, low education), Thomas (boy, high education), and Kevin (boy, low education) (see Bloothooft and Onland, 2011: 34). By using these four names, we adjust for the possibility that individuals from different societal strata know more or less of the prompted X's.[9] See Appendix 5.1 for the population numbers of these names and Appendix 5.2 for sensitivity analyses.

5.5.3 Analytical Procedure of Our New Method

Next, we problematize modeling the extended network size in our new approach and provide our solution to this problem. Table 5.3 provides a summary of the notation of this section. On Facebook, we observe the number of four first names (Thomas, Kevin, Anne, and Melissa) as well as the number of *other friends*. The crucial modeling step is our assumption that the prevalence of first names across Facebook and the extended networks is similar. This allows us to estimate how large the "other category" is in our new estimate of extended social network — i.e., how many people individuals know with *other* names than the set of four (other than Thomas, Kevin, Melissa, or Anne). Our purpose is to model $N|(F, F_{\text{NAMES}}, IN_{\text{NAMES}}, X)$, where N is the size of the extended social network (i.e., the new measure, unobserved), F is the size of the extended network on Facebook (observed), F_{NAMES} is the total number of friends with those four names in the Facebook networks (observed), IN_{NAMES} is the total number of people known personally with those four names, based on the scale-up method (an observed

[9] Respondents also had to indicate how many people they know named "Moham(m)ed." Unfortunately, this name has a disproportionately large influence on the network size. Approximately 65% of the Dutch ethnic majority indicates knowing no-one named Moham(m)ed, whereas those of Arabic origin know 6 to 10 Moham(m)ed's (the median). Furthermore, there is a high diversity of Dutch names, whereas Arabic names are more homogeneous. Hence, the probability that someone of Turkish origin knows many Moham(m)ed's is much higher than a Dutch majority member knowing many Kevin's. Therefore, the name is incomparable with the other names, and we chose to exclude it.

interval-censored count), and X is a vector of predictor variables. We describe our model as two integrated submodels: part (a) explains the network sizes (N, F) from the respondent characteristics X, and part (b) explains the network selection $(IN_{\text{NAMES}}, F_{\text{NAMES}})$ from network size (N, F) and respondent characteristics X.

For part (a), we use a random intercept bivariate Poisson distribution in which the log(rates) are predicted from (possibly different sub-vectors of) X, with network-type specific coefficients $b.F$ and $b.N$ (this implies that we use different predictor variables for the extended Facebook network size and the extended network size of our new approach), and from a bivariate normal distributed residual u. These random effects for F and N have network-type specific standard deviations and may be correlated, because both network sizes are likely affected by the (possibly unmeasured) sociability of a respondent. In other words, there likely is unexplained variance across the two network sizes which is similar within the same respondent.

For part (b), we associate the observed interval censored IN_{NAMES} with the *unobserved* count data N_{NAMES}. N_{NAMES} represents the number of people respondents know with first names *other* than Thomas, Kevin, Anne, and Melissa. N_{NAMES} is unobserved and this is the core issue our model seeks to solve. We also add another count, F_{OTHER}, which is the number of Facebook friends minus the sum of occurrences of the four first names on Facebook, or $F - F_{\text{NAMES}}$. We define e_{NAMES} as the collection of the number of four first names and other names. We assume that Fe_{NAMES} (i.e., $F_{\text{NAMES}} + F_{\text{OTHER}}$) and Ne_{NAMES} (i.e., $N_{\text{NAMES}} + IN_{\text{NAMES}}$) are stochastically independent, conditional on F, N, X, such that Fe_{NAMES} depends on (F, X) and Ne_{NAMES} on (N, X). The key assumption we make in order to assess N is that the selection processes on Facebook and for the extended network are similar. $Ne_{\text{NAMES}}|(N, X)$ and $Fe_{\text{NAMES}}|(F, X)$ are conditionally independent and multinomially distributed with unknown size N and known size F, respectively, and with common selection probabilities $p(X)$, defined as a multinomial logistic regression model, treating *other* (i.e., all names beyond the set of four first names) as the reference category. The commonness of these selection probabilities allows us to estimate N_{NAMES}. This implies that, for instance, if on Facebook the other category is responsible for 90% of the friendship choices of a particular type of respondent (given the vector of predictor variables X), we assume that this 90% similarly applies to the extended social network of our new measure.

Table 5.3: Explanation of the notation used for the new procedure to estimate the extended social network size.

Notation	Explanation
N	Extended social network size new approach
F	Extended social network size on Facebook
F_{NAMES}	Total number friends with the four first names occurring on Facebook
IN_{NAMES}	Total number of contacts with the four first names in the scale-up method
N_{NAMES}	Total number of other names than the set of four in the extended network
X	A vector of predictor variables (different for F and N)
$b.F$	Coefficients of effects of X on F
$b.N$	Coefficients of effects of X on N
u	Residual of the network sizes (correlated for F and N)
F_{OTHER}	Number of friends on Facebook minus F_{NAMES}
e_{NAMES}	The collection of the set of four names and other names
Fe_{NAMES}	Collection of four names and other names on Facebook ($F_{\text{NAMES}} + F_{\text{OTHER}}$)
Ne_{NAMES}	Collection of four names and other names extended network ($N_{\text{NAMES}} + IN_{\text{NAMES}}$)
$p(X)$	Selection probabilities for the names (similar for Facebook and new measure)

5.5.4 Application of Our New Method to the Data

Thus, we estimate three equations simultaneously in one model. First, we include an equation (submodel (a)) for the number of Facebook friends and an equation (also submodel (a)) for the new measure of the extended network size (unobserved). Second, the model includes the *selection* equation (submodel (b)) for the number of first names respondents indicate they know.

The first equation in the model predicts the number of friends on Facebook via a multinomial logistic regression for uncensored count data. The dependent variable is a count of each of the four names on Facebook and a residual count. The residual count is the number of Facebook friends minus all of the appearances of the set of four first names (i.e., $F_{\text{OTHER}} = F - F_{\text{NAMES}}$). We include all variables in this equation that are outlined as predictor variables. For those who are new to Facebook and less-frequent users, their number of friends on Facebook is more remote from their overall number of connections than for more-experienced and more-frequent Facebook users. Therefore, we control for the year in which respondents became a member of Facebook and for the number of hours respondents spend each day on Facebook. Finally, we include self-rated behavioral problems because previous research has shown that such factors relate to social network maintenance online (Ellison et al., 2007) and confound comparisons between the extended network size and the Facebook network size.[10, 11]

The second equation predicts the extended network size via a multinomial logistic regression for uncensored count data. Essentially, we predict the residual number of contacts respondents have, net of the number of contacts of the set of four names (i.e., $N_{\text{NAMES}} = N - IN_{\text{NAMES}}$). We argue that the extended network size is a function of all outlined predictor variables.

The third and final equation in the model predicts the number of friends one knows of each of the set of four first names. This equation takes the form of a multinomial logistic regression for interval censored count data because we have a collection of numerical ranges (e.g., 2-5, 6-10, etc.) for the number of X's each respondent indicates she or he knows. As covariates in this equation, we include gender, education, and ethnicity to attempt to adjust for additional homophily effects based on the set of first names. Furthermore, we include whether people have a romantic relationship and the interactions between ethnicity and gender to account for the possibility that, e.g., Dutch boys know more Kevin's than Turkish girls.

Fitting this full model is not possible with currently available standard software. For instance, even the submodel for a multinomial logistic regression based on interval censored count data cannot be fitted using standard statistical packages known to us. Consequently, to fit our model, we turned to Bayesian inference using MCMC (Markov Chain Monte Carlo simulations) using JAGS (Just Another Gibbs Sampler: Plummer, 2003).

5.6 How Large Are Extended Social Networks?

Table 5.4 provides comparisons and correlations between the number of people respondents indicate they know in the survey who have one of the four first names

[10] The number of friends on Facebook is correlated with Facebook membership duration ($r = .268$; $p < .001$; Median year of membership = 2010) and with the amount of hours spent on Facebook per day ($r = .154$; $p < .001$; Median hours per day = 1 hour or less). Facebook membership duration in years comes from the DFS, and the amount of hours spent on Facebook each day originates from wave 4 of the CILSNL.

[11] Behavioral problems is constructed as follows. Respondents could indicate on three separate questions how often they felt worried, depressed, or worthless on a four-point scale (1-often true, 2-sometimes true, 3-rarely true, and 4 never true). We calculate the mean behavioral problem score out of these three questions (Mean = 2.882; $\alpha = .752$).

and the number of occurrences of these same first names among the networks on Facebook. The number of first names known in the survey is higher than the number of occurrences of these first names among the Facebook networks.

Table 5.4: Descriptive statistics for and correlations between the number of X's mentioned in the scale-up method and in the Facebook friend lists (N = 2,151).

X's	Median X's scale-up[a]	Mean X's Facebook	Correlation[b]	p
Thomas	2-5	1.560	0.426	0.000
Kevin	2-5	1.791	0.399	0.000
Anne	2-5	1.753	0.416	0.000
Melissa	1	1.006	0.417	0.000

[a] We report the median because the number of X's are in numerical ranges (e.g., 2-5, 6-10, etc.); [b] Correlations between the categories of the scale-up questions and a count in the Facebook data.

Using the results of our analytical procedure, which are found in Table 5.7 (elaborated below), we generate a predicted extended social network size for each respondent. Essentially, we obtain a Poisson distribution of the network sizes based on the sum of products of individual characteristics and coefficients, as well as a residual term. The residual term follows a bivariate normal distribution, independently for each respondent, with mean zero, and a (co)variance matrix and their correlations (see Table 5.7, last two rows).

Table 5.5 shows comparisons between the predicted extended social network size and other, more straightforward calculations of extended social network sizes. The predicted extended social network size is approximately 524, compared to a number of Facebook friends of approximately 379. We calculated the number of contacts via the basic scale-up estimator from equation (2). Using this simple measure for the extended network size, we obtain a network size of approximately 1363. Hence, the basic scale-up estimator provides higher estimates than our predicted extended network size, whereas the extended network size measured as the number of Facebook friends is somewhat smaller than the figure our new measure provides. Figure 5.1 depicts the kernel smoothed density distributions for the predicted extended social network size and the predicted number of Facebook friends (using the same method to predict the extended social network size). It shows that even though the distributions of both network sizes look similar, the predicted extended social network size is higher.[12, 13]

Table 5.5: Descriptive statistics on the predicted extended social network size and more straightforward calculations (N = 2,151).

	25%[a]	50%	75%	Mean	SD/SE
Number of friends on Facebook	237	348	487	379.028	199.668
Basic scale-up estimator	640.087	1280.695	1600.868	1362.785	431.785
Predicted number of friends on FB	193	298	459	364.152	261.231
Predicted extended soc. net. size	264	411	661	524.034	396.487

[a] These statistics are the 25$^{\text{th}}$ percentile, median, and 75$^{\text{th}}$ percentiles of these variables in the data.

5.7 Hypotheses Tests

We take a two-step approach in which we for each hypothesis first consider the extended network size on Facebook and, thereafter, consider the analyses of our new procedure. The first main analysis is a Heckman selection model (1979) in which we regress the extended network size measured as the number of Facebook friends on our predictor- and confounding variables. Via the Heckman selection model, we correct selectivity in modeling network size only when a second selection equation determines that this social network size was non-missing. The errors of both equations are allowed to correlate. We adjust for ethnic background, gender, and educational track level in the selection equation and cluster-corrected standard errors for the school cluster to which adolescents belong. Those who are member of the ethnic minority, girls, and those in lower educational track are more likely to maintain private friend lists (Hofstra et al., 2016b), and we thus

[12] The standard error for the basic scale-up estimator is calculated using McCormick et al.'s (2010: 60) equation,

$$SE(\textit{Scale-up degree}_i) = \sqrt{\textit{Scale-up degree}_i} \times \sqrt{\frac{1 - \sum_{k=1}^{K} N_k/N}{\sum_{k=1}^{K} N_k/N}}). \tag{5.3}$$

[13] We also calculated the scale-up estimator based on geography. Respondents indicated how many people they know from one of five relatively large cities spread throughout the Netherlands (see Appendix 5.1). Here, the network size is ∼140 (SE = ∼32). A reason for why this network size estimate is low is that respondents may under-recall contacts from the large population in cities, not unlike the way respondents over-recall contacts from small subpopulations (e.g., McCormick et al., 2010).

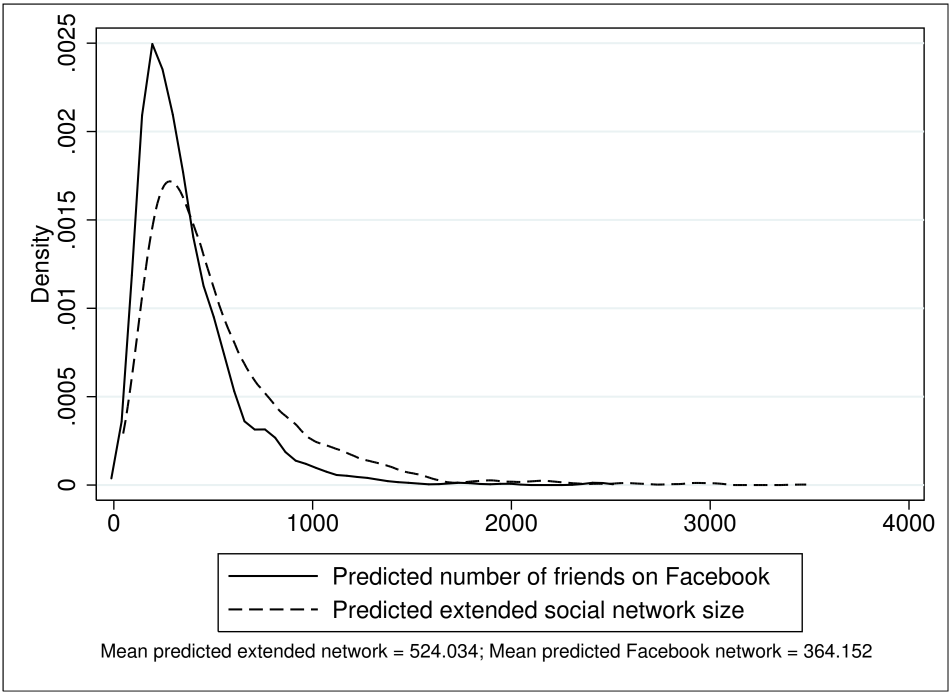


Figure 5.1: Kernel smoothed density distributions for the predicted extended social network size and the predicted number of Facebook friends.

attempt to adjust for this selectivity. The correlation (ρ) between the errors of both equations is about .962 and a Wald test of independent equations shows that $\rho \neq 0$ ($p < .000$). This non-zero correlation implies selectivity in our outcome (i.e., the number of Facebook friends). Besides this Heckman selection model, we present a simple linear regression in which we regress the number of Facebook friends on the predictor- and confounding variables to illustrate the importance of correcting for the sample selection biases. The results of this Heckman selection model are presented in Table 5.6.[14]

The second main analysis presents the results of the Bayesian estimation procedure described above. The results are presented as the mean and standard deviation of the posterior distribution of the coefficients of the model, as well as the upper

[14]Multilevel regression models that account for the clustered data structure (Snijders and Bosker, 2012; i.e., students in classes in schools) do not provide qualitatively different results than the linear regression model shown in Table 5.6.

(2.5%) and lower (97.5%) bounds of these coefficients. This allows for a Bayesian alternative to standard null-hypothesis significance testing. If the middle 95% of all coefficients do not contain zero, we can safely assume that the coefficient is non-zero. The mean values of the coefficients can be interpreted as Poisson regression coefficient estimates. A one-unit increase in a predictor increases the log count of the extended network size with the value of that coefficient. We exponentiate the mean values of the coefficients to provide an interpretable indication of effect sizes (i.e., the percent change in incident rate ratio, Long and Freeze, 2001).

The correlation between the residuals for the new measure of the extended network size and the extended network size measured as the number of Facebook friends is approximately .249. The standard deviations are indicative of the amount of unexplained variance in the respective parts of the models. The three panels of Table 5.7 show the effects of our predictors on (1) the new measure of the extended network size and (2) the extended network size measured as the number of Facebook friends, as well as (3) tests for differences in predictors between these two measures. The results for the selection part of the model (i.e., what respondent characteristics are associated with how many X's respondents know?) are found in Appendix 5.3.[15]

5.7.1 Opportunities and Homophily

Foci. We first examine whether adolescents who spend more time (a) going out, (b) in associations, and (c) visiting concerts have larger extended social networks (H1). As shown in the Heckman-selection panel of Table 5.6, adolescents who spend more time going out, in associations, and going to concerts have a larger extended network on Facebook. The magnitude of these effects are rather large. For instance, for a one-unit increase (e.g., from once a month to once a week) in going out, adolescents gain approximately 45 Facebook friends.

If we consider the analyses shown in Table 5.7, we find similar results: adolescents who spend more time going out and in associations have larger extended networks (as measured via our new procedure). Similar to the results of Table 5.6, the magnitudes of these effects are rather large: for a one-unit increase (e.g., from

[15]We also ran a Heckman selection model using the basic scale-up estimator as the dependent variable where we corrected for ethnic background, gender, and educational level in the selection equation. The full table of these results are found in Appendix 5.4. These results of this straightforward measure of the extended network size do not provide qualitatively different results from those presented in Table 5.7.

once a month to once a week) in going out and in associations, the incident ratio of the extended network size is 15% and 7.6%, respectively (e.g., $exp(0.141) = 1.151$). Furthermore, if we consider the Facebook network size in this procedure (i.e., without accounting for selection), we find that spending time in each of the three foci is positively related to the Facebook network size. Hence, the more time adolescents spend in foci, the larger their extended network sizes, both on Facebook and as calculated from our new method. These findings are consistent with Hypothesis 1.

The Similarity of Potential Contacts. Second, we consider whether Dutch majority members have larger social networks than adolescents of the ethnic minority (H2). We first consider the results of Table 5.6. The Heckman selection model shows that those who are member of ethnic minority groups have smaller extended network sizes as measured on Facebook than members of the ethnic majority. The comparison between the linear regression and the Heckman selection model shows the importance of adjusting for sample selections, as the linear regression shows no (for the Arabic ethnic groups) or contrasting (for the "other" ethnic groups) effects.

Table 5.7 shows that, in contrast to our expectation, Dutch ethnic majority members seem to have smaller extended networks in our new procedure than those who are members of an ethnic minority. If we consider the Facebook network size using our new method, we observe no statistically significant associations between ethnic background and network size (but this likely results from not correcting for sample selections). Considering the analyses using the basic scale-up estimator (see Appendix 5.4), we find that the basic scale-up estimate of the extended network size is not related to ethnic background, not in the linear regression model and not in the Heckman selection model. Hence, regarding the extended social network size on Facebook, we find evidence consistent with Hypothesis 2, whereas we find inconclusive results using our new procedure.

The Similarity of Potential Contacts in Foci. Next, we study whether having more co-ethnic classmates (a) and schoolmates (b) positively affects the extended network size (H3). Considering the results of Table 5.6, we find that those who have more ethnically similar classmates have larger extended social networks as measured on Facebook. We find no relationship between the number of co-ethnic schoolmates and the Facebook network size.

From our analyses in Table 5.7, we find no statistically significant relations between these predictors and the extended network size and the number of Facebook friends. Furthermore, the basic scale-up estimate of the extended network size is unrelated to the number of co-ethnic classmates and schoolmates (see Appendix 5.4). Concerning the relation between the number of potential contacts in foci and the extended social network size, we find some evidence consistent with Hypothesis 3, but only for the Heckman selection model that considers the extended network size as measured on Facebook.

5.7.2 Romantic Partners

We expected that adolescents who indicate being in a romantic relationship have larger extended social networks than adolescents who are not (H4). If we consider the results in Table 5.6, we find that those who are in a romantic relationship have a larger extended network size as measured on Facebook. Those in a romantic relationship have approximately 35 Facebook friends more than those we are not in a romantic relationship.

We find no statistically significant relation between the new measure of the extended social network size and having a partner in Table 5.7. Similarly, in our analyses of the basic scale-up estimator of the extended network size, we find that the basic scale-up estimate of the extended network size is not predicted by having a romantic partner. However, having a partner does seem positively related to the Facebook network size in our new procedure. Specifically, the incident ratio of the network size is approximately 16% higher for those with rather than without a partner ($exp(.148) = 1.160$). We thus find evidence consistent with Hypothesis 4 if we consider the extended social network measured as the number of Facebook friends.

5.7.3 Education and Gender

Education. Next, we expected that adolescents in higher educational track levels would have larger extended networks than adolescents in lower educational track levels (H5). In our Heckman selection analysis (Table 5.6), we find that those who are in the senior general educational track have larger social networks on Facebook than those in the lower vocational educational track. The simple linear regression model of Table 5.6 shows contrasting effects, which again shows the importance of adjusting for sample selections.

With regard to the first panel of results in Table 5.7, we find no statistically significant association between educational track level and our new measure of the extended social network size. Those in the lower educational track seem to have larger networks on Facebook than those in the highest educational track, contrary to our hypothesis (but this likely results from not adjusting for sample selections). Hence, we find some evidence consistent with Hypothesis 5, but only for the extended network size as measured on Facebook.

Gender. Finally, we expected that girls would have larger extended networks than boys (H6). The results in Table 5.6 suggest that girls indeed have larger extended networks than boys. Specifically, girls seem to have approximately 42 friends more on Facebook than boys have.

Considering our new procedure in Table 5.7, we observe that girls do not seem to have a larger extended social network than boys, but girls do have larger Facebook networks than boys. Moreover, the incident ratio for the Facebook network is approximately 9.2% higher for girls than for boys ($exp(.088) = 1.092$). There is thus some evidence in support of Hypothesis 6, but only for the extended network size measured as the number of Facebook friends.

5.7.4 Confounding Factors and Differences in Predictors

We observe that those who had a Facebook membership longer and those who spend more hours per day on Facebook have larger Facebook networks, both in Table 5.6 and in Table 5.7. This is also consistent with what one would expect concerning these confounding variables. The results of Table 5.7 suggest that Dutch majority members, girls, and those who are in a romantic relationship seem to have a larger share of their extended social network contacts among their Facebook friends than their counterparts.

Table 5.6: Maximum-likelihood estimation results of the extended network size measured as the number of Facebook friends via a linear regression and via a Heckman selection model.

	Linear regression			Heckman selection		
	Coef.	S.E.	p^a	Coef.[b]	S.E.	p
Constant	-146.450	31.581	***	-286.18	30.307	***
Foci (H1)						
Going out	57.047	4.997	***	44.542	3.961	***
Associations	24.353	3.424	***	21.604	3.528	***
Concerts	13.307	5.952	***	11.522	5.477	*
Similarity of cont. (H2+H3)						
Ethnicity						
Dutch (ref.)						
Arabic	-0.038	25.204		-254.559	30.543	***
Other	30.539	17.027	*	-76.593	20.388	***
# Co-ethnic Class	1.535	0.882	*	1.722	0.758	*
# Co-ethnic School	0.002	0.008		-0.000	0.009	
Romantic partners (H4)						
Partner (ref. No)	43.448	8.839	***	34.879	7.757	***
Education and gender (H5+H6)						
Education						
Vocational (ref.)						
Senior general	-15.856	9.994		37.720	19.849	*
University prep.	-44.158	9.943	***	8.781	15.851	
Girl (ref. Boy)	23.862	8.105	**	42.173	10.763	***
Confounders						
Membership duration	42.435	3.365	***	32.972	3.044	***
Hours FB per day	20.582	3.883	***	18.401	3.618	***
Behavioral problems	0.898	6.115		-2.797	5.254	
Observations	2113			5013		
R^2	0.228					
Log pseudolikelihood				-16983.510		

[a] One-sided p-values: * $p < .05$, ** $p < .01$, *** $p < .001$; [b] In the selection equation we adjusted for ethnic background, gender, and educational level. Boys, ethnic minority members, and lower educated are less likely to have a value on the Facebook network-size estimate. This is consistent with findings that these groups more often opt for privacy on Facebook (Hofstra et al., 2016).

Table 5.7: Posterior means, posterior standard deviations, and posterior quantiles for the Bivariate Poisson-normal distribution for the extended network size, the number of friends on Facebook, and tests for differences in predictors across extended networks and Facebook (N = 2,151; 20,000 iterations).

	New measure of extended network size					Facebook extended network size					Difference in predictors			
	Coef.	SD	2.5%	97.5%	S[a]	Coef.	SD	2.5%	97.%	S	Coef.	2.5%	97.5%	S
Constant	5.171	0.112	4.946	5.389	*	4.051	0.104	3.849	4.254	*	-1.120	-1.395	-0.834	*
Foci (H1)														
Going out	0.141	0.022	0.097	0.185	*	0.190	0.016	0.158	0.220	*	0.049	0.001	0.096	*
Associations	0.073	0.015	0.042	0.104	*	0.089	0.011	0.067	0.110	*	0.016	-0.017	0.049	
Concerts	0.041	0.027	-0.011	0.096		0.051	0.019	0.014	0.090	*	0.010	-0.047	0.068	
Similarity cont. (H2+H3)														
Ethnicity (ref.:Dutch)														
Arabic	0.681	0.150	0.381	0.967	*	-0.067	0.078	-0.217	0.080		-0.748	-1.040	-0.433	*
Other	0.209	0.074	0.062	0.353	*	0.043	0.048	-0.050	0.136		-0.165	-0.319	-0.010	*
# Co-ethnic Class	0.000	0.004	-0.007	0.007		0.005	0.003	0.000	0.010		0.005	-0.003	0.013	
# Co-ethnic School	0.000	0.000	0.000	0.000		0.000	0.000	0.000	0.000		0.000	0.000	0.000	
Romantic partners (H4)														
Partner (ref. No)	0.000	0.044	-0.087	0.086		0.148	0.028	0.092	0.203	*	0.147	0.055	0.237	*
Educ. & gender (H5+H6)														
Educ. (ref.: Uni. pr.)														
Vocational	0.084	0.049	-0.011	0.177		0.066	0.032	0.005	0.127	*	-0.017	-0.116	0.082	
Senior general	0.041	0.054	-0.067	0.149		0.068	0.034	0.003	0.136	*	0.027	-0.086	0.142	
Girl (ref. Boy)	-0.001	0.038	-0.077	0.074		0.088	0.025	0.039	0.138	*	0.088	0.005	0.169	*
Confounders														
Membership duration						0.100	0.009	0.082	0.117	*				
Hours FB per day						0.072	0.012	0.048	0.097	*				
Behavioral problems						-0.024	0.019	-0.060	0.013					
Standard deviation	0.644	0.015	0.615	0.673	*	0.547	0.009	0.530	0.565	*				
Correlation residuals FB and extended network	0.249	0.025	0.199	0.298	*									

[a] Statistically significant in the sense that the middle 95% of the 20,000 coefficients do not contain zero.

5.8 Conclusions and Discussion

Social contacts lend social support (Wellman and Wortley, 1990; Hobbs et al., 2016), can be used to solicit advice (McPherson et al., 2006), provide information (Granovetter, 1973, 1983; Burt, 2000), and grant meaningful connections to unknown social groups (Feld, 1984). However, remarkably little is known about individual variation in the size of extended social networks. We argued that this lack of knowledge is caused by uncertainty over how to measure extended social networks. We set out to address both of these issues: first, we estimated the extended social network size using the number of Facebook friends and by proposing a method which integrates the number of Facebook friends and the network scale-up method. Second, we explained how individual differences come about in these two measures of the size of the extended personal network. Uniquely, we used a combination of survey data on the network scale-up method and data from Facebook.

So, how large are extended social networks? Logically, the answer depends on the definition of the network boundary and the way it is measured. We considered all the contacts individuals know on a first name basis and with whom they would have a chat if they met randomly (McCarty et al., 2001; DiPrete et al., 2011). We found a mean of approximately 379 Facebook friends. Furthermore, using our new procedure, we predicted the extended social network size to be approximately 524, on average. Results form our integrated procedure suggest that of the total number of contacts in extended social networks, the share found in Facebook networks is higher for girls, ethnic majority members, and those in a romantic relationship. Our predictions of the extended network size using the new method are in line with prior work using solely the scale-up method among adults, showing extended network sizes in the range of 550-750 (e.g., Zheng et al., 2006; McCormick et al., 2010; DiPrete et al., 2011).

What explains individual differences in the sizes of extended social networks? We first turned to classic literature on tie-generating mechanisms among the core ties (e.g., Blau, 1977a; Feld, 1981). We hypothesized and corroborated that those who spend more time in socially oriented foci, i.e., in bars/clubs, associations, and concerts, have larger extended networks. Throughout the analyses of the number of Facebook friends and in our new procedure, participation in this set of foci was a consistent predictor of the extended social network size and the number of Facebook friends. The magnitude of these associations was rather large. These findings are consistent with prior work showing that foci are important for the emergence of strong ties (e.g., Feld, 1982, 1984; Kalmijn and Flap, 2001; Mollenhorst et al. 2014).

The results were less straightforward regarding the other hypotheses. We expected that the interplay between homophily and opportunity would affect the extended social network size. Ethnic minority members had larger networks when we considered the extended network size as measured on Facebook. The same goes for the number of co-ethnic classmates, as this number predicted the extended network size on Facebook. It is possible that a smaller pool of dissimilar alters does result in fewer realized relationships among ethnic minority members. Tie investment may be higher among dissimilar alters (see Hofstra et al., 2017, for a similar argument). Ethnic minorities seem to establish rather fewer ties rather than dissimilar ties in their extended social network measured as the number of Facebook friends. By and large, theories on relationship formation among core ties also predict the number of friends on Facebook. In the consideration of our new procedure, however, our results showed no association between ethnic background and the number of co-ethnic contacts in schools and school classes.

Those in a romantic relationship, higher educated (although not for each educational category), and girls have a larger extended network size than their counterparts, if we consider the number of Facebook friends. These findings are consistent with our hypotheses. If we compare these findings to the findings of our new procedure, we could not confirm these predictions in our new procedure.

Hence, there are discrepancies between the results of the number of Facebook friends that adjusted for the sample selections and the results of our new procedure. We draw two conclusions from these discrepancies. First, we illustrated the relevance of adjusting for sample selections in predicting the number of Facebook friends. If one were to consider Facebook-only data, this will result in fewer girls, ethnic minority members, and lower educated among observations of the extended network size. This likely will bias the analyses, as our results suggest that there are differences in network size among ethnic minorities and lower educated who show their number of Facebook friends on their profiles and those who do not. Second, our proposed measure of the extended social network size needs further improvements (those of which we outline below). We consider this new procedure as a step between prior extended social network size estimates and future measures with which scholars can test hypotheses on individual variation in the extended social network size.

5.8.1 Limitations of this Study

There are four limitations in this study that merit acknowledgement. First, data on a more general target population, such as adults, would be ideal. As of yet, however, we do not know of other samples that combine detailed survey data, including the network scale-up measurement, and behavioral data on online social networks. Furthermore, the theoretical mechanisms we describe are relatively general in nature (e.g., opportunity effects) and not limited to the adolescent population, and we view this study as a next step between the lack of systematic studies on measuring *and* explaining the extended network size and future studies considering more-general target populations. However, because 79% of US online adults use Facebook (Greenwood et al., 2016), we would recommend imaginative strategies that combine survey network data on adults and their number of online network contacts.

Second, the survey data on the network scale-up method includes only *four* names. A potential consequence of using only four names to estimate the extended social network size is that the estimates become uncertain. The prevalence of these four names was not high in the survey (see Table 5.4). Essentially, there was low variation between answers of the respondents on the question how many contacts they knew that carry these names. Trying to estimate the network size with such low variation between respondents may have decreased the precision of our estimates. Future work should ask for a larger number of names than only the four we focused on in this study (see McCormick et al., 2010 for suggestions), such that these names would cover a larger part of the total population and that variation between respondents increases.

Third, the sample selections in our new procedure were strict; we analyzed those cases who participated in two waves of survey data *and* whose Facebook friend list was public. Extending our model to account for these two selections would be ideal. Our Heckman selection analyses suggested that our findings on the number of Facebook friends were particularly affected by sample selection. The method we used for estimating the extended network size was computationally intensive. Adding an additional model equation that accounts for the sample selections may make inference infeasible. Future research should strike a compromise among the complexity of the model, the possibility of accounting for sample selectivity, and the number of simulations used (for our data, the 20,000 iterations until convergence took approximately 48 hours using the JAGS software).

Finally, we acknowledge that we have no longitudinal network data to estimate (potentially) causal relations. We especially need such dynamic network data to study the role of cumulative advantage in the extended network size. Here, we considered, for instance, an indirect proxy for such a mechanism in the form of network size differences between those in a romantic relationship and those who are not. We would commend future work that, for instance, gathers multiple waves of behavioral data on the number of friends on Facebook to study such processes more directly.

5.8.2 Implications and Future Research

What do we learn from our results? We mention four considerations for future research that arise from our study. First, the most robust finding of this study is that spending time in foci is positively related to the extended network size, in both our main analyses. This is in line with studies that have shown the importance of foci in tie formation and tie stability (e.g., Kalmijn and Flap 2001; Wimmer and Lewis, 2010; Hofstra et al., 2017). In this study, we considered three foci: going out (e.g., bars), associations (e.g., sports), and concerts (e.g., dance events). It would be interesting for future scholars to hypothesize about differences among the type of foci for the extended network size – which foci are more conducive to accumulating social contacts? For instance, how important are religious meeting places vis-à-vis sport clubs for tie formation? Or foci for individual versus team sports? Ceiling effects for the various foci could also be considered: at what point does spending extra time in foci no longer facilitate the accumulation of social contacts? For instance, spending up to eight hours a week at a sports club may be conducive to meeting new people, but any time thereafter may not.

Second, some of the mechanisms we addressed were tested using proxies. We particularly suspect that testing our hypotheses on the role of education and gender may have been problematic. We elaborated that women may recall their networks better. If we calculate the differences between the number of X's in the survey and on Facebook, the differences were consistently smaller for girls than for boys. There are two plausible explanations for this: boys over-recall names they indicate they know relative to girls (Brashears et al., 2016), or girls add a larger share of their extended network as friends on Facebook. The null-effect in our new procedure is thus far from conclusive, and we suggest that future research more closely inspect the role of gender in these network recall dynamics. Furthermore, the null-effect for education on the size of the extended social networks (using our new procedure) may be due to our consideration of time spent in three social foci. This would imply an indirect effect of education via time spent in foci on network size. This would be consistent with prior work (e.g., Peterson, 1992; Lizardo and Skiles, 2012) and would simultaneously counter the suggestion that the higher-educated are cognitively better equipped to keep track of all of their contacts. However, refined measures of cognitive capabilities are needed to study effects of cognitive abilities on the size of extended network socials in more detail.

Third, our main analyses for predicting the extended network size need additional analytical work. First, the key assumption that the relative popularity of the first names on Facebook and in the extended social networks is the same needs to be tested. Second, we could consider assumptions about the prior distributions of the extended and Facebook network sizes other than the Poisson-normal distributions we used (e.g., based on negative binomials, the power law distribution, rounded log-normal distributions). Finally, as of yet, we have been unable to provide *goodness-of-fit* indices, i.e., how well does the estimated model fit the data? Hence, further model specifications are needed.

A fourth consideration for a follow-up study is to what extent differences in the extended network size indeed translate into unequal access to resources. A possible line of inquiry is to contrast the new measyre of the extended network size and extended network size as measured on Facebook in their influence on, for instance, social support (Wellman and Wortley, 1990; Hobbs et al., 2016) or (job) information and employment success (Granovetter, 1973, 1983; Burt, 2000). One possible prediction could be that the number of friends on Facebook has a greater influence on access to social support or job information than the even larger personal network size measured in our new procedure. This is because Facebook ties may represent ties that are easier to access and maintain through the platform (Ellison et al., 2007). If scholars find this to be true, the number of friends on Facebook could be straightforward to obtain (Golder and Macy, 2014) and could be an indication of the number of relevant contacts for the study of access to resources. This last intuition was already predicted twenty years ago (cf. Lewis et al., 2008b) by Rogers (1987), as he stated that "computer-monitored data from the new media [...] can deal with the network sampling/generalizability difficulties" (p. 307). Here, we presented a study in similar vein.

Chapter 6

Predicting Ethnicity with First Names in Online Social Media Networks[1]

Abstract: *Social scientists increasingly use (big) social media data to illuminate longstanding substantive questions in social science research. However, a key challenge of analyzing such data is their lower level of individual detail compared to highly detailed survey data. This limits the scope of substantive questions that can be addressed with these data. In this study, we provide a method to upgrade individual detail in terms of ethnicity in data gathered from social media via the use of register data. Our research aim is twofold: first, we predict the most likely value of ethnicity given one's first name, and second, we show how one can test hypotheses with the predicted values for ethnicity as an independent variable while simultaneously accounting for the uncertainty in these predictions. We apply our method to social network data collected from Facebook. We illustrate our approach and provide an example of hypothesis testing using our procedure, i.e., estimating the relation between predicted network ethnic homogeneity on Facebook and trust in institutions. In a comparison of our method with two other methods, we find that our method is less prone to false-positive results. We discuss the promise of our approach and pinpoint future research directions.*

[1]This chapter is under review at an international scientific journal. Bas Hofstra is the first author of this chapter, but the chapter presents joint work with Niek de Schipper. Hofstra wrote the main part of the manuscript and coordinated the Facebook data collection. Hofstra and De Schipper jointly conducted the analyses. De Schipper substantially contributed to the manuscript. The authors jointly developed the idea and design of the study. I thank Lukas Norbutas for valuable feedback on an earlier draft of this study. Finally, I thank Rense Corten and Frank van Tubergen for the invaluable discussions we had about this study and for pinpointing the possibilities of the approach outlined in this chapter, which we first started to develop in Chapter 4.

6.1 Introduction

Research on social media is rapidly expanding in the social sciences. A query for "Facebook" — which, with more than 1.2 billion daily users, is the prime example of a social media platform (Facebook, 2017) — on the academic search engine Google Scholar provides about 145k research articles since 2015 that contain this word. Queries for critical issues such as "Climate," "Aids," "Inequality," and "DNA" each resulted in similar or fewer mentions than did "Facebook." Furthermore, studies on social media (e.g., Ellison et al., 2007; boyd and Ellison, 2008; Lewis et al., 2008b; Bond et al., 2012) are among the most highly cited articles in the social sciences. An increasing number of scientific journals cover social media, big data, and their relationship to society (e.g., *Big Data & Society*, *Social Media + Society*, *Journal of Computer-Mediated Communication*, *New Media & Society*, etc.). Even in science's most prestigious outlets — e.g., *Science*, *Nature*, and *PNAS* — there are a number of studies using big data from social media (e.g., Bond et al., 2012; Kramer et al., 2014; Bakshy et al., 2015; Hobbs et al., 2016).

This increased scholarly interest in social media is no surprise, as social media data provide scholars with novel ways to analyze human social interactions. On social media, for instance, individuals have new ways to communicate, to spread information, and to coordinate collective action (cf. Corten, 2012; see González-Bailón and Wang, 2016). Furthermore, individuals increasingly use these platforms to maintain their interpersonal social relationships (Ellison et al., 2011). In addition, computational approaches to social science make it relatively easy to collect data on online interactions, such as those documented on Facebook or Twitter, because these interactions generate digital time-stamped footprints of large social networks (Golder and Macy, 2014).

It has been argued that the networked footprints these platforms automatically archive as part of their daily operations have revolutionized social science (Lazer et al., 2009; Watts, 2011; Spiro, 2016). These features of social media data make it relatively straightforward to study larger networks that reach beyond the small number of core ties (e.g., fewer than five network contacts) and the closed social contexts (e.g., classrooms or departments) usually under consideration in the study of social relationships, i.e., in social network analyses (cf. Hofstra et al., 2017; see, e.g., Marsden, 1987; Kalmijn and Van Tubergen, 2006; Smith et al., 2014b; Hofstra et al., 2015b; Van Tubergen, 2015). These types of data are occasionally labeled "big," as they often contain information about the online behavior of millions of users (McFarland and McFarland, 2016). However, big data obtained from social media has yet to reach its full potential regarding its use in social science research.

Analyzing such online social media data comes with major challenges. One of the core challenges is that the level of individual detail in data gathered from social media is often considerably lower when compared to information gathered in survey research (Golder and Macy, 2014; Stopczynski et al., 2014; Spiro, 2016). For instance, key individual characteristics such as gender, age, ethnicity, education, or occupation are often either missing or even misreported by respondents in big data gathered online (Golder and Macy, 2014; Spiro, 2016). Individual privacy considerations may be one reason why social media data are often *broad but shallow*, as people might close their social media profiles to safeguard their personal information (e.g., Hofstra et al., 2016b). Therefore, these data may be *big*, but the level of detail is thin. This low level of detail limits the scope of the substantive questions that can be addressed when studying data obtained from social media platforms. Ethnicity and gender, for instance, are key individual characteristics by which social network analysts often study the patterns in their data (e.g., McPherson et al., 2006; Mayer and Puller, 2008; Wimmer and Lewis, 2010; Smith et al., 2014; Van Tubergen, 2015; Hofstra et al., 2017).

Hence, a key question in the growing field of analyzing online social network data is how one can use the plethora of opportunities of these data while at the same time maintaining (at least some of) the “richness” that is usually found in survey data. Effectively dealing with this issue increases the number of substantive questions one is able to answer.

These considerations motivate the aim of this study, which is twofold. First, we propose a procedure to upgrade the level of individual detail in online social network data by predicting the most likely value of ethnicity given one’s first name using register data. Second, we propose a procedure on how to test hypotheses using these “upgraded” social media data. It is well established that names are a clear signal of ethnicity. There are profound differences in how people from different ethnic backgrounds name their children (Lieberson, 2000; see, e.g., Coldman et al., 1988; Lauderdale and Kerstenbaum, 2000; Fiscella and Fremont, 2006; Chang et al., 2010; Bloothooft and Onland, 2011; Mateos et al., 2011). Scholars seem to increasingly use this empirical regularity to enrich social media data (see Cesare et al. [2017] for a recent overview). Logically, because (first) names are often among the only indicators researchers have about individuals in social media data. We follow this burgeoning line of research and thus make use of names as a signal of ethnicity.

We make two key contributions to this growing field. First, we extend prior work because we consider the possibility that those who carry the same first name can each have a different ethnicity. There are two studies that are most related to our procedure, that of Chang et al. (2010), who use a probabilistic Bayesian approach, and that of Hofstra et al. (2017), who use a supervised learning approach. Chang et al. (2010) and Hofstra et al. (2017) assign the most likely value of ethnicity to people on Facebook given their surnames (Chang et al., Hofstra et al.) and first names (Hofstra et al.). While both studies aim to validate their ethnicity predictions using a source of ground truth (Chang et al.: MySpace data; Hofstra et al.: survey data), they do not model the possibility of different ethnicities among people carrying the same names. Or, as Chang et al. (2010: 25) put it: "we [...] have not yet theoretically modeled error throughout our calculations." We statistically take this uncertainty into account for a more realistic representation of the relationship between ethnicity and (first) names. To show the promise of our approach, we directly compare the Hofstra et al. (2017) method with the method described in this study and show which method is the least prone to false-positive results.

Second, we show how to use the predicted values of ethnicity among Facebook networks in standard regression models to tests hypotheses. More specifically, we show how to test hypotheses with the predicted variable as an independent variable while simultaneously accounting for the uncertainty in the predicted values of this new variable.[2] To show the promise of our approach, we provide a toy example. In recent years, there has been a sharp increase in studies investigating the claim that ethnic diversity has detrimental effects on trust and social cohesion (Putnam, 2000; Van der Meer and Tolsma, 2014; Abascal and Baldassarri, 2015). As societies increasingly grow ethnically diverse, it is crucial to understand whether and how ethnic diversity across different contexts — e.g., in cities, neighborhoods, workplaces, and among social contacts — affects trust. As an example, we explore and engage some of the claims regarding the relationship between ethnic diversity and social cohesion. Specifically, we consider ethnic homogeneity in Facebook networks and investigate its relationship to trust in institutions. Note that we do not aim to test theoretically derived hypotheses. In tying our method to this substantive example, we merely push future work to theoretically consider the consequences of the (ethnic) composition of online (Facebook) networks. Moreover, we provide a statistically plausible method to test hypotheses with predicted indi-

[2] Data enrichment resembles issues of missing data. We specify later on how these two bodies of work relate to one another.

vidual characteristics (i.e., ethnicity) in online big data as independent variables. As an engaging starting point in doing so, we consider trust in institutions.

Therefore, we contribute to the growing field of analyzing online big data by showing how to enrich and make innovative use of such data to test hypotheses in a novel way. We aspire to make the description of the procedure as accessible as possible so that the applied empirical scientist can adopt the method with relative ease using free and open source software.

To illustrate our approach, we use three data sources: (1) survey data from a large, diverse sample of adolescents; (2) online social network data containing more than a million network members downloaded from the Facebook pages of these same adolescents; and (3) register data that capture the frequency of first names and the proportion of the name carriers and their parents who have been born in specific countries. The specific data sets we use are: (1) survey data from the "Children of Immigrants Longitudinal Survey in Four European Countries" (CILS4EU; Kalter et al., 2016) and the "Children of Immigrants Longitudinal Survey in the Netherlands" (CILSNL; Jaspers and Van Tubergen, 2017), (2) Facebook data from the "Dutch Facebook Survey" (Hofstra et al., 2015a), and (3) register data of the Dutch Civil Registration of 2010 (Bloothooft and Schraagen, 2011).

6.2 The Concept of Ethnicity in the Dutch Context

Some define ethnicity based on an individual's self-identification (e.g., Verkuyten and Kwa, 1994), such that is up to individuals to decide whether they "identify as a member of group X." Others use objective measures of ethnicity, occasionally based on parents' birth countries (e.g., Vermeij et al., 2009; Stark and Flache, 2012). Here, for simplicity, we use the regularly applied definition by Statistics Netherland, which is standard practice in research on ethnicity and social networks in the Netherlands (e.g., Vermeij et al., 2009; Statistics Netherlands, 2012). This means that we classify individuals' ethnicity by their biological parents' birth-country (cf. Vermeij et al., 2009; Stark and Flache, 2012). When individuals have one parent born in the Netherlands, we classify them as belonging to the ethnicity of the parent not born in the Netherlands, and when they have parents born in two different non-Dutch countries, we classify them in the birth country of the mother.

Because the data we use are from a sample of adolescent respondents in the Netherlands, it is informative to define the ethnic groups most salient in Dutch society. Dutch adolescents can be classified among six large ethnic origin groups (Castles et al., 2013). The first group comprises adolescents whose parents were born in the Netherlands and who are members of the Dutch majority. The second and third groups consist of children of immigrants from Turkey and Morocco. These children's parents originated from the low-educated labor force that the Netherlands recruited in the 1950s and 1960s from Turkey and Morocco or from parents who arrived more recently (e.g., because of family reunions). These two groups constitute the largest minority group in the Netherlands. Another group originates from post-colonial countries in the Dutch Caribbean (e.g., Aruba and Suriname). A fifth group originates from other Western countries (e.g., neighboring countries such as Germany), and a sixth group originates from other non-Western countries (e.g., conflict areas such as Afghanistan). These ethnic groups are rather similar across Western European countries in the type of immigrants that settled there, despite varying specific countries of origin of the ethnic groups that are present (Smith et al., 2014).

Ethnic background is thus occasionally defined by classifying individuals into one of these six large ethnic background groups in the Netherlands: Dutch, Turkish, Moroccan, Dutch Caribbean, other Western backgrounds, and other non-Western backgrounds. We also use this categorization throughout this study.

6.3 Data Sources

6.3.1 CILS4EU and CILSNL

We use the third, fourth, and fifth waves of the CILS4EU and CILSNL on adolescents in the Netherlands (Kalter et al., 2016; Jaspers and Van Tubergen, 2017).[3, 4] In the CILS4EU, adolescents were followed for three consecutive years (2010-2013). Data were collected in the Netherlands, Sweden, Germany, and England. The CILSNL followed the Dutch panel of the CILS4EU for an additional four years (2014-2017). We analyze the Dutch part of the data because our variables of interest are only included in the Dutch section. We analyze waves 3, 4, and 5

[3] One can apply for data access to waves 1, 2, and 3 of the CILS4EU via the following link: https://dbk.gesis.org.

[4] One can apply for data access to waves 4, 5, 6, and 7 of the CILSNL via the following link: https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:65866.

because these waves are the anchor of and the closest in time to the Facebook data. Essentially, the Facebook data are collected via respondents' survey answers in waves 3 and 4 (elaborated below). All surveys include detailed information on respondents' background, attitudes, and leisure time activities.

The wave 1 sample was stratified by the proportion of immigrants of non-Western origin within a school. Within these strata, schools were chosen with a probability proportional to their size (using the number of pupils in the relevant educational level), and two classes were randomly sampled within the schools. In wave 1 (2010-2011), 4,963 Dutch pupils participated.[5] There were 4,272 respondents in wave 3 (2012-2013), 4,072 in wave 4 (2013-2014), and 3,836 in wave 5 (2014-2015). In waves 3, 4, and 5, respondents could self-complete the questionnaire online, on paper, or via telephone interview.[6]

6.3.2 The Dutch Facebook Survey

The Dutch Facebook Survey (DFS) (Hofstra et al., 2015a) was collected to enrich the Dutch part of the CILS4EU survey. It consists of behavioral data obtained from Facebook.[7] The data were collected between June 2014 and September 2014. In waves 3 and 4 of the surveys, participants were asked about their membership on Facebook. In waves 3 and 4 combined, 4,864 respondents indicated that they had a membership on Facebook in at least one of these waves (wave 3 N = 3,423, wave 4 N = 3,595). Coding assistants tracked down profiles based on the respondents' names and cities of residence. A total of N = 4,463 (91.8%) Facebook profiles were tracked. From these tracked profiles, the coding assistants downloaded the complete friend lists of the respondents. These Facebook friend lists are the focus

[5]Six hundred respondents in wave 1 were sampled who were not a part of the random sampling frame because some schools wanted to participate in the survey with more than two classrooms. Therefore, a *random* sample of 4,363 pupils was drawn in wave 1. Because of attrition rates between waves 1 and 2, our sample cannot be guaranteed to be representative. We include as many respondents as possible in the sample for analyses, including newcomers (nonrandom) and the nonrandom sample of wave 1, to ensure a large sample size across waves.

[6]A minority of the pupils in the higher educational track were still in high school in wave 3. These pupils were still surveyed at their respective schools while a researcher was present.

[7]An anonymized version of the DFS is available from October 2017 onward via the following link: https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:62379.

of this study, i.e., the first names of the individuals found among these friend lists.[8]

Approximately 73% of the respondents kept a public friend list, and we collected the first names out of these friend lists. See Hofstra et al. (2016b) for a discussion of these respondents' privacy settings. We have information on complete Facebook networks of 3,352 respondents, and they had a combined total of 1,156,285 Facebook friends.[9] Together, these individuals have 52,651 unique first names.

6.3.3 The Dutch Civil Registration

The Dutch Civil Registration (DCR) data are register data of those who have Dutch nationality and were alive and living in the Netherlands in 2010 (N = 15,785,208; Bloothooft and Schraagen 2011). The Facebook networks constitute 52,651 unique first names. These first names (up to the first space or hyphen) were matched to the first names in the DCR data of those having a Dutch nationality and were living in the Netherlands in 2010. We were able to match 36,151 (68.66%) of the first names in the DFS to the DCR. These names comprise ∼92% of the total Dutch population (N = 14,447,100) and ∼95% of the respondents' total Facebook friends (N = 1,106,675). Thus, we have register data on 36,151 unique first names, and these comprise the major part of the Facebook friend lists. This is an indication that the vast majority of first names in the networks on Facebook are of sufficient quality to match them to the register data, i.e., Facebook friends seem to provide realistic first names on their Facebook profiles instead of fictive pseudonyms (e.g., "Captain Fantastic"). However, it may be that individuals provide fictive first names that match the register data (e.g., "Jane," whereas the real or legal name is "Alice"). Unfortunately, with the current data, we are unable to filter such cases among individuals' Facebook friends. Because we matched Facebook profiles to survey respondents on the basis of their names (and cities), this is not an issue for the names of the respondents themselves. The register data

[8]The collected information was publicly visible on Facebook. Coding assistants were instructed personally, and all followed strict coding procedures with password-protected files. All personal identifiers were removed from the data. The data collection, the coding procedure and the use of these data for scientific purposes were reviewed and approved by an internal review board of the Faculty of Social and Behavioural Sciences at Utrecht University (project number: FETC14-019).

[9]This total of 1,158,227 Facebook friends is a raw count of all of the friendships respondents have. A likely situation is that respondents have the same friends in their Facebook networks. Counting this unique set of Facebook friends would likely result in a lower number.

contain information on the birth country of parents of each of these first names. Table 6.1 provides a schematic overview of the three data sources used in this study.

6.4 The Misclassification Ratio

We calculate a weighted misclassification ratio for assigning ethnicity based on first names using the register data. This is done to provide an indication of how feasible it is to predict ethnicity using first names. As an example, we use a simple *majority rule* to assign ethnicity. Per first name, we assign ethnicity (i.e., a categorization into one of the six ethnicity groups mentioned before) based on the largest proportion of the name carrier's mother's country of birth (which we obtain from the register data). Next, we subtract the maximum proportion of mother's birth countries per name. For instance, if 90% of mothers of individuals named John are born in the Netherlands, we assign a Dutch ethnicity to John. However, this would mean that 10% of all John's are misclassified as Dutch. We calculate a weighted average of this number based on all of the first names. We weigh the average by how many times a name occurs in the register data. Hence, names that occur few times have less impact on this ratio than names that occur many times. The weighted misclassification rate is 1.3% (unweighted = 5.1%). This figure implies that predicting ethnicity via first names is (in our view) sufficiently accurate and can be used as such, even though it may be somewhat driven by the Dutch majority members. To make our estimates as precise as possible, we try to further adjust for misclassifications by using a procedure that accounts for further uncertainties arising from assigning ethnicity based on first names.

Table 6.1: An overview of the three data sources that are used for this study.

Name	Type	Description	Abbreviation
Children of Immigrants Longitudinal Survey in four European Countries[a]	Survey	Questionnaires to adolescents (wave 3)	CILS4EU
Children of Immigrants Longitudinal Survey in the Netherlands	Survey	Questionnaires to adolescents (waves 4, 5)	CILSNL
Dutch Facebook Survey	Behavioral	Facebook networks of respondents in the surveys	DFS
Dutch Civil Registration data	Register	Dutch register data	DCR

[a] The CILS4EU and CILSNL are the same data source under a different name, as the CILSNL continued to survey the Dutch panel of the CILS4EU four additional waves.

6.5 Outline of the Procedure

6.5.1 General Outline

Next, we outline the general procedure we use to predict ethnicity. The idea is to predict ethnicity on the basis of people's first names. This new variable (i.e., predicted ethnicity of a Facebook friend) can then be used as an independent variable for hypothesis testing. However, when one wants to test hypotheses with predicted independent variables, one should seek to take into account the uncertainty in the predicted values of this independent variable. The uncertainty about the predictions should be adjusted for in the model coefficients of interest as well.

In our specific case, we need to consider two types of uncertainty. First, the analyses should consider the possibility that different individuals who carry the same first names can each have a different ethnicity. Hence, the most likely prediction of ethnicity given one's first name is not always correct. For instance, a specific first name can be popular across more than one ethnic subgroup. Second, the analyses should consider that the parameters (i.e., the model coefficients) of the prediction model may also carry uncertainty, as the prediction model is estimated from the register data. The fact that we use *register* data that cover nearly the entire Dutch population may imply that the parameters are less uncertain than if they had been estimated from sampled data. Nevertheless, to elaborate our example and to be as precise as possible, we do adjust for uncertainty in the model parameters. In our example, the parameters of the prediction model are the conditional probabilities of one's ethnicity given one's first name.

A convenient way to consider the first type of uncertainty is to use bootstrap standard errors for hypothesis testing. A bootstrap sample is a sample from the original sample of the same size drawn *with* replacement. In our method, in each bootstrap sample, a newly predicted data set is used to obtain the parameter estimates of the hypothesis testing model of interest. The bootstrap standard error is the standard deviation of all parameter estimates obtained by estimating the model using data from the different bootstrap samples.

If we predict the ethnicity for each Facebook friend repeatedly for each bootstrap sample, we consider the possibility that the most likely ethnicity for this Facebook friend given his/her name may not be correct. Hence, each time a new bootstrap sample is used to estimate the parameters of our model of interest, we consider the possibility of different ethnicities among similar names.

We also take into account the second type of uncertainty, i.e., the fact that the estimates in our prediction model of interest carry uncertainty. For each bootstrap sample, we use different probabilities to predict ethnicity conditional upon one's first name. These conditional probabilities are obtained from a Dirichlet posterior distribution estimated from the register data (i.e., the DCR). If we use different conditional probabilities in our prediction model to predict ethnicity for each of the bootstrap samples, we account for the uncertainty of the parameters used in the prediction model.

6.5.2 Model Specification

Here, we describe the full estimation model in more detail. First, we outline the prediction model; subsequently, we describe the bootstrap procedure. For our prediction model, we assume that someone's ethnicity $y_1, ..., y_L$ is distributed according to a multinomial distribution $Mult(\theta_1, ..., \theta_L)$, where $y_1, ..., y_L$ denote indicator variables of the L ethnicity categories and $\theta_1, ..., \theta_L$ denote the probabilities of belonging to the Lth ethnicity category given one's first name. We assume $\boldsymbol{\theta} = (\theta_1, ..., \theta_L)$, the vector of conditional probabilities, to be distributed according to a Dirichlet distribution $Dir(a_1, ..., a_L)$, where $a_1, ..., a_L$ denote the parameters of the Dirichlet distribution. To obtain a distribution for $\boldsymbol{\theta}$, we consider the posterior distributions $f(\boldsymbol{\theta}|D)$, where D refers to the register data, for each occurring first name. We make use of the fact that $f(\boldsymbol{\theta}|D)$ is proportional to $f(D, \boldsymbol{\theta}) = f(D|\boldsymbol{\theta})f(\boldsymbol{\theta})$, where $f(D|\boldsymbol{\theta})$ is the density of the data (or likelihood of the data) and $f(\boldsymbol{\theta})$ is the prior distribution of $\boldsymbol{\theta}$ (see, e.g., Tu, 2017). Essentially, this means that we obtain $f(\boldsymbol{\theta}|D)$ by multiplying the density of the data with our prior distribution. Because our prior distribution is Dirichlet and the density of the data is a multiproduct of multinomial distributions, our posterior distribution $f(\boldsymbol{\theta}|D)$ is also a Dirichlet distribution, which is convenient to sample from. More specifically, $f(\boldsymbol{\theta}|D)$ is given by $Dir(b_1, ..., b_L)$, where b_1 is provided by the sum of the value on the corresponding prior parameter and the number of people with ethnicity l.

As an example, consider a posterior distribution of people named John, where L is 3 (i.e., there are three ethnic categories). When the posterior distribution $f(\boldsymbol{\theta}|D)$ is equal to $Dir(100 + 1, 300 + 1, 30 + 1)$, we have observed 100 John's in the first ethnicity category, 300 in the second, and 30 in the third. The plus 1 appears from the fact that we have used an uninformative prior $f(\boldsymbol{\theta})$.

To obtain realistic bootstrap confidence intervals for our model parameters of interest, we predict ethnicity given one's first name for each $k = 1, ..., K$ bootstrap sample using a freshly drawn $\boldsymbol{\theta}$ for each occurring name from $f(\boldsymbol{\theta}|D)$. For each of the K samples, we then obtain the regression parameters by estimating our model of interest. The bootstrap standard error of the regression parameters is then given by

$$\mathrm{SE}(\hat{\beta}) = \frac{1}{K-1}\sum_{i=1}^{K}(\hat{\beta}_i - \bar{\beta})^2. \tag{6.1}$$

Our estimation procedure is summarized in Figure 6.1. We aggregate the data because the Facebook networks are nested in individuals (i.e., Facebook friends of respondents in the surveys). Therefore, we obtain one value for the ethnic composition of Facebook networks per respondent (which we will elaborate upon later). We adjust the model parameters for confounding respondent-level variables (e.g., respondent's own gender and ethnicity).

6.5.3 Relation to Imputation of Missing Data

We briefly note how our described procedure relates to the missing data literature. Procedures of data enrichment can be seen as a missing data problem, specifically, not having certain information about subjects while still aiming to use the missing information in an analysis. In our example, ethnicity is completely unobserved. We predict ethnicity based on a prediction model, and this prediction model resembles a missing data imputation model. For an imputation model to perform well, it should be specified according to the missing data assumption. Here, we adopt the missing at random (MAR) assumption (Rubin, 1976). Essentially, this implies that the missing data can be accounted for by the observed variables. We assume that, conditional upon one's first name, we can determine ethnicity and that there is no correlation between ethnicity and other variables. MAR is a rather strong assumption, but one that is made generally in the imputation of missing data. In our case, MAR is indeed a strong assumption, and we do not expect it to hold entirely. However, we think that the imputation model is reasonable because a first name is a strong signal of ethnicity (e.g., Lieberson, 2000; Chang et al., 2010; Bloothooft and Onland, 2011; Mateos et al., 2011) and because the weighted misclassification ratio is low.

Additionally, imputation models should account for the uncertainty in their parameters (Schafer and Graham, 2002). We do this by keeping the parameters of our imputation model random (instead of fixed). Moreover, each time a new data set is imputed, the parameters of our imputation model are sampled again from the posterior distribution.

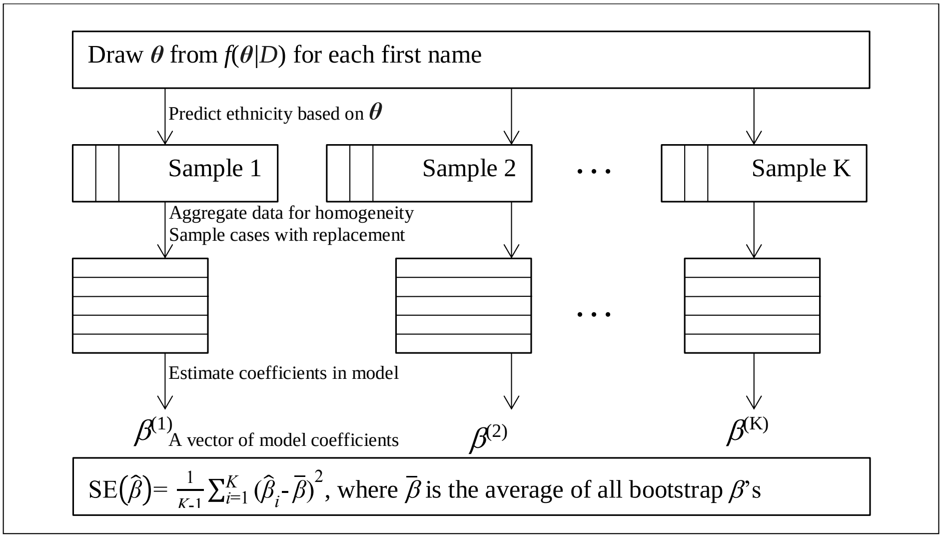


Figure 6.1: A graphical outline of the method to predict ethnicity.

6.6 Application of the Procedure to Our Data

In this section, we describe the application of our procedure. As an example, we examine to what extent we can statistically relate *trust in institutions* to a predicted measure of *ethnic diversity* in Facebook networks. There is an ongoing discussion about whether ethnic diversity is detrimental or beneficial to social trust (see Tolsma and Van der Meer, 2014; Abascal and Baldassarri, 2015).[10] As such, we think it is engaging to explore relations between a measure of ethnic diversity on Facebook and trust using our method. However, the choice of this dependent variable is arbitrary and may as well be something different, i.e., it

[10]We use trust in institutions as one dimension of *social trust*. We find that *generalized trust* (often used as a measure for social trust), i.e., whether individuals think that "most people can be trusted," correlates with trust in institutions ($r = .342$; $p < .001$). We do not use generalized trust, as it is a dichotomous dependent variable, and this makes our estimation procedure more complex and relatively slow. We acknowledge, however, that trust in institutions is only a dimension of social trust.

serves as an example. No specialized software is needed to replicate this approach. All computations of the procedure were performed using custom code written for the software package R (R Core Team, 2016), and the core code for the procedure is found in the Appendix for Chapter 6.

6.6.1 Dependent and Control Variables

The variable that we thus want to relate to our predicted values of ethnicity is *trust in institutions*. This variable is constructed as follows. In wave 5 of the CILSNL, respondents indicated on a ten-point scale on four items how much confidence (1, no confidence whatsoever — 10, a lot of confidence) they had in politicians, judges, scientists, and the police. We took the mean score out of these four to calculate trust in institutions (Mean = 6.358; SD = 1.413; Cronbach's α = .803).

Next, we construct five control variables from the third and fourth waves of the CILS4EU and CILSNL to examine whether we can isolate the effect of our newly predicted variable from confounding factors. In constructing these variables, we took the survey answers from wave 4. If these survey answers were missing, we took answers from wave 3. First, we construct the *ethnicity* of the respondents themselves. We classify respondents into one of the six largest ethnic groups in the Netherlands: "Native Dutch" (76.9%), "Turkish" (3.7%), "Moroccan" (3.5%), "Dutch Caribbean" (3.9%), "Other Western" (5.4%) and "Other non-Western" (6.6%). As we mentioned before, this is based on their parents' country of birth (Vermeij et al., 2009; Statistics Netherlands, 2012). Second, we measured whether the respondents indicated whether they were a *girl* (59.6%) or a *boy* (40.4%). Third, we measured how satisfied respondents reported being with their "life in general" (1, very dissatisfied — 10, very satisfied; Mean = 7.585; SD = 1.613). Fourth, we measured whether respondents were in a romantic relationship (38.7%) or not (62.3%). We adjust for these factors according to their availability, and the questions were posed in a similar way to respondents across the multiple waves of survey data. Finally, using dummy variables, we measured in which of the waves the respondents participated (waves, 3, 4, and 5 = 78.4%; waves 4 and 5 = 15.8%; or waves 3 and 5 = 57.5%). We listwise delete missing values across these six variables and realize a dataset consisting of 3,445 cases.

6.6.2 Predicting Ethnicity and Estimation Results

Next, we want to obtain the most likely value of ethnicity given the first name of each of the Facebook friends of a respondent. We have register data for 36,151 of the first names on Facebook (95% of all friends). These register data contain (1) the number of occurrences of each of these first names in the Netherlands and (2) for each first name, the fraction of the name carriers' mothers who were born in the six major origin countries (the Netherlands, Turkey, etc.). We thus predict the ethnicity of friends on Facebook via the birth country of the mothers. This is a small deviation from the regularly applied definition in the cases where the mothers were born in the Netherlands but the fathers were born elsewhere. To keep our method parsimonious, we consider only mothers' birth countries. For 2,208 out of the 3,445 of respondents (from which we have all values across the dependent and control variables), we can observe the first names in their Facebook friend lists, i.e., these respondents have public friend lists. This final set of 2,208 respondents have a combined total of 776,135 Facebook friends for which we want to predict ethnicity.

For each bootstrap sample, we draw the conditional probabilities for ethnicity given one's first name on Facebook using the posterior distribution obtained with the register data. Next, we count the number of friends who have the same predicted ethnicity as the ethnicity of the respondent him/herself and divide this count by the total number of friends and multiply it by 100. As such, we calculate the percentage of co-ethnic Facebook friends per respondent, or co-ethnic$_{\text{FACEBOOK}}$, as a measure of ethnic homogeneity. This means that we aggregated the predicted values for ethnicity across the friends' first names from the respondent's Facebook network.

This aggregated predicted variable is then used as an independent variable in a linear regression model — also adding the control variables mentioned before — with trust in institutions as the dependent variable. We repeat this process 10,000 times, each time bootstrapping new conditional probabilities for ethnicity and each time bootstrapping the model coefficients from the linear regression models. We obtain a distribution of 10^4 bootstrap coefficients per variable in the linear regression model. Assessing these 10^4 bootstrap coefficients, we can obtain the bootstrap confidence interval and observe whether the middle 95% of the bootstrapped coefficients are either above or below zero. If we plot these 10^4 coefficients, they visually resemble normal distributions. Addressing this bootstrap confidence interval is a non-parametric alternative to standard null-hypothesis significance testing.

Table 6.2 shows the bootstrap results of 10^4 linear regressions. We show upper and lower quantiles and the means of the coefficients. We briefly discuss the results of these regression models. First, those *with* a romantic partner seem to report *less trust* in institutions, and those who report *higher life satisfaction* report *more trust* in institutions. Those of *Turkish* and *Dutch Caribbean* ethnicity report *less trust* in institutions than members of the Dutch majority. The magnitude of these ethnicity effects seems rather high if we consider the mean of all coefficients. Those who participated in waves 4 and 5 report less trust in institutions than those who participated in waves 3, 4, and 5. Finally, we do not observe a relation between the percentage of co-ethnic Facebook friends and trust, given our bootstrapped coefficients.

Table 6.2: Regression results of co-ethnic$_{\text{FACEBOOK}}$ and control variables on trust in institutions (using 10^4 bootstrap samples and N = 2,208).

	Bootstrap coefficients				
	Mean	Lower (2.5%)	Upper (97.5%)	Sig?[a]	Pred[b]
Co-ethnic$_{\text{FACEBOOK}}$	0.113	-0.795	1.037	No	Yes
Intercept	6.058	5.198	6.909	Yes	
Wave					
Wave 3, 4, and 5	Ref.	Ref.	Ref.		
Wave 4 and 5	-0.340	-0.484	-0.195	Yes	
Wave 3 and 5	-0.017	-0.260	0.224	No	
Girls (ref. Boys)	-0.012	-0.113	0.091	No	
Romantic partner	-0.110	-0.214	-0.006	Yes	
Life satisfaction	0.088	0.054	0.123	Yes	
Ethnicity					
Dutch	Ref.	Ref.	Ref.		
Turkish	-1.277	-1.897	-0.628	Yes	
Moroccan	-0.593	-1.316	0.167	No	
Dutch Caribbean	-0.875	-1.641	-0.083	Yes	
Other Western	-0.165	-0.932	0.621	No	
Other non-Western	-0.427	-1.235	0.398	No	

[a] If the middle 95% of the coefficients do not contain zero, we can safely assume that the coefficient is non-zero; [b] Co-ethnic$_{\text{FACEBOOK}}$ is our predicted independent variable of interest.

6.7 Model Performance

Next, we provide an indication of the performance of our method. We provide confidence intervals for three methods of estimating the effects of co-ethnic$_{\text{FACEBOOK}}$ on trust in institutions. First, we provide confidence intervals based on 10^4 bootstrapped coefficients as outlined in this study. Second, we assign ethnicity based on the majority rule explained before and calculate the fraction of co-ethnic Facebook friends. We then ran a linear regression of this measure — controlling for the variables mentioned before — on trust. Third, we replicated the co-ethnic$_{\text{FACEBOOK}}$ measure of Hofstra et al. (2017). They used a training data set, where they knew the self-reported ethnic background and first names of the respondents and sought which proportions of parents' birth country in the register data correlated best with self-reported ethnicity in the survey. We regress trust in institutions on the fraction of co-ethnic Facebook friends using this measure (while controlling for the set of variables mentioned before).

Table 6.3 presents confidence intervals for effects on trust in institutions of the three methods of calculating the percentage of co-ethnic Facebook friends. It presents conservative, regular, and non-conservative confidence interval boundaries of the coefficients researchers usually consider for standard statistical significance across three panels of results.[11]

We observe that the method outlined in this study provides the most-conservative tests of the effects of the predictor on the dependent variable. Using a simple majority rule for only one predicted data set may lead more often to false-positive results. One may conclude that the more homogeneous Facebook networks are, the more trust individuals have in institutions (in the consideration of the third panel of results). However, when we take into account the two types of uncertainty incorporated in our method, i.e., the possibility of different ethnicities among similar first names and the uncertainty in the model coefficients in the prediction model, we no longer observe such a relationship. The method provided in Hofstra et al. (2017) does not seem to lead to false inferences using this example. However, the confidence intervals are more conservative using the method provided in this article. These findings suggest that using our method in the process of hypothesis testing may provide more-conservative tests of hypotheses that are less prone to false-positive results.[12]

[11] The control variables are omitted from Table 6.3. However, the control variables had qualitatively similar results over the different analyses, comparable with the coefficients found in Table 6.2.

[12] One may also argue that our procedure is more prone to false-negative results in the case of a false null-hypothesis. In the situation of null-hypothesis testing, however, one should aim for conservative hypothesis tests.

Table 6.3: A comparison of three methods for predicting ethnicity, presented are confidence intervals for effects of co-ethnic$_{\text{FACEBOOK}}$ on trust.

	Conservative		Regular		Non-conservative	
	2.5%[a]	97.5%	5%	95%	10%	90%
This study	-0.795	1.037	-0.658	0.891	-0.482	0.707
Majority rule	-0.198	1.421	-0.068	1.291	0.082	1.141
Hofstra et al. (2017)	-0.618	0.238	-0.549	0.169	-0.470	0.090

[a] These are the confidence intervals for the different methods.

6.8 Conclusions and Discussion

The aim of this study was twofold. First, we outlined a procedure to predict ethnicity in social media data using register data. Second, we showed how to use these predicted values in standard regression models to tests hypotheses. As such, we contributed to an expanding amount of prior work aiming to enrich social media data (e.g., Chang et al., 2010; Hofstra et al., 2017) based on the idea that names are a signal of individual characteristics such as ethnicity (Lieberson, 2000; Fiscella and Fremont, 2006; Bloothooft and Onland, 2011). We provided a method that accounted for (1) the possibility of people with similar names having different ethnicities and (2) uncertainty in model estimates when using predicted values as independent variables in linear regression models. We did so by bootstrapping conditional probabilities given one's first name and bootstrapping standard errors from a set of 10^4 linear regression models. The percentage of misclassifications of ethnic background using a simple majority rule was approximately 1.3%, which is relatively low and may be a further illustration of the promise of our approach.

We provided a toy example showcasing how we could predict the ethnicity of respondents' friends on Facebook (usually not readily available) and related respondents' percentage of co-ethnic friends on Facebook to trust in institutions. As such, we provided a way to illuminate longstanding substantive discussions in future research. In this example, we explored the relation between ethnic diversity and trust (e.g., Tolsma and Van der Meer, 2014; Abascal and Baldassarri, 2015). We found no significant relation between trust in institutions and the predicted values of the percentage of co-ethnic Facebook friends.

We compared the method outlined in this study with two other, more straightforward ways to predict ethnicity. First, we compared it with two other methods that assign ethnicity on the basis of first names, first, by using a simple majority rule within the register data and, second, with a supervised learning method. The results suggest that the method outlined in this study is the least prone to false-positive results. We showed that using more-straightforward ways to assign ethnicity based on first names may lead to the conclusion that the percentage of co-ethnic Facebook friends positively predicts trust in institutions. However, we showed that this result may be a false-positive finding as a consequence of not accounting for the two types of uncertainty in the analyses.

Three limitations of this study merit acknowledgement. First, there may be selection biases in these data resembling different privacy settings among Facebook users. For instance, the observed results from the Facebook data may be driven by the fact that those of non-Dutch ethnic background more often have opted for private Facebook profiles (see Hofstra et al., 2016b). According to the homophily principle (Lazarsfeld and Merton, 1954; McPherson et al., 2001), individuals prefer to befriend ethnically similar others above those of ethnically dissimilar backgrounds. When Dutch majority members more often have open profiles and they more often befriend Dutch friends on Facebook (which they do, see Hofstra et al., 2017), this will be reflected in an oversampling of Dutch majority members and Dutch first names among the Facebook friend lists. In itself, this is an interesting intuition, because then the question arises of which individuals we as researchers *by design* can study in online social media data. Future researchers in the field should be aware of such selection patterns. The goal of this paper was to outline our method, but an oversampling of typical Dutch first names may have led to insufficient variation in the ethnic homogeneity measure.

Second, we lumped "Other Western" and "Other non-Western" roots together as ethnic background categories, whereas it is only reasonable to assume that there is substantial variation in naming habits within these categories. For instance, those of Afghan origin are labeled under other non-Western ethnicities, as are those with Chinese roots. However, the naming practices between these countries vary substantially. In future research, scholars should consider how to strike a compromise between sufficient observations within ethnic-racial background categories and the precision of the ethnicity predictions. This is also highly dependent on the substantive question the researcher aspires to address. One can, for instance, decide to only study groups for which the data cells are sufficiently filled, without examining possible residual categories. Another example would be to use more-precise "Other Western" categories if one is interested in interethnic ties on Facebook between majority and minority members from neighboring countries to the Netherlands. In this study, we used these categories because studies on ethnicity in the Netherlands usually apply this categorization (Statistics Netherlands, 2015).

Third, there may be situations where first names of ethnic minorities who do not have a Dutch nationality vary from those of ethnic minorities who do. There were approximately 16.6 million inhabitants in the Netherlands in 2010. This means that the number of cases in the register data cover ~95% of all inhabitants in the Netherlands. Moreover, there were approximately 800,000 inhabitants in the Netherlands in 2010 not carrying a Dutch nationality. This may result in predictions for these first names that are less precise because we do not have register data on their mothers' country of birth. Unfortunately, we cannot adjust for these situations with the current data.

Next, we discuss implications and future research directions, and we pinpoint alternative data sources. We urge scholars to use (variants of) this method for future scientific endeavors, especially because of the growing use of online social network data and the challenges that come with it. A first key future research path we would commend is replicating the method using register data on social class. Social class is related to names as well, as parents from different societal strata name their children differently (Lieberson, 2000; Bloothooft and Schraagen, 2011). Defining social class disparities in online behavior relates to issues of digital inequalities (e.g., DiMaggio et al., 2001; Hargittai, 2002). Moreover, investigating online social network segregation or integration by social class is an understudied area that directly merits further investigation. One way to do this is to obtain register data on names and information on parents' educational background about these names. Another feature of first names is that they vary in their popularity *over time*. This may provide possibilities to study age cleavages in online social networks, i.e., one may aim to predict age or age-cohort based on first names.

A second future research endeavor would be to directly test hypotheses using "upgraded" social media data. Here, we provided an example using trust in institutions as the dependent variable. However, another example that may directly relate to ethnic homogeneity on Facebook (predicted via first names on Facebook) is *ethnic prejudice*. Literature suggests that even superficial contacts between members of different ethnic groups potentially reduce intergroup prejudice (Allport, 1954; Pettigrew and Tropp, 2006). Implementing our method using ethnic prejudice instead of trust in institutions as the dependent variable would be a direct test of the hypothesis of whether ethnic diversity among Facebook friends hampers ethnic prejudice.

Finally, we pinpoint data sources other than the ones we used. Using these data sources, scholars can replicate our method across different contexts and using different sources of data. Gathering network data from social media is relatively easy (Golder and Macy, 2014), and our predicted values of ethnicity are not limited to Facebook (nor are they to limited to ethnicity). For instance, the Application Programming Interface of Twitter (a microblogging website) is straightforward to access. A consideration for future scholars is the extent to which Twitter users use aliases instead of real (first) names. Another data source would be the networks found on LinkedIn, which tend to be related to the professional, work-related networks people have. An interesting question would be to what extent diversity in these networks by, for instance, social status is related to labor market outcomes (e.g., Granovetter, 1973, 1983). Register data that contain information on names are also not unique to the register data from the Netherlands that we used. For instance, the US Census Bureau provides a list of first names in the US (occurring > 100 times) and the percentage of these first name carriers' racial backgrounds (United Status Census Bureau, 2014). Another context would be Germany, were Statistics Germany provides lists of first names by region and gender (Statistics Germany, 2016). We described a method to predict the ethnic background of members in Facebook networks — which is unavailable in many cases (Spiro, 2016; Cesare et al., 2017) — on the basis of people's first names, but our procedure is not limited to specific individual characteristics. We recommend future replication efforts of our procedure on different individual characteristics and in different national contexts.

Appendices

Appendices Chapter 2

Table A2.1 on the next page in this appendix are the full results for the multinomial logistic regression on which the postestimations in Chapter 2 Table 2.4 are based.

Table A2.1: Multinomial logistic regression showing the effect of independent variables on the contrast between SNS membership categories. Facebook and Hyves as a category is the reference outcome.

Facebook and Hyves (ref.)	Hyves			Facebook			No member		
	Coef.	S.E.[a]	p[b]	Coef.	S.E.	p	Coef.	S.E.	p
Number of friends in class on Facebook	-0.265	0.052	0.000	0.182	0.105	0.084	-0.234	0.065	0.000
Number of frineds in class on Hyves	0.052	0.044	0.232	-0.347	0.101	0.001	-0.256	0.067	0.000
National Origin									
Dutch	(ref.)	(ref.)	(ref.)	(ref.)	(ref.)	(ref.)	(ref.)	(ref.)	(ref.)
Turkish	-1.304	0.234	0.000	1.115	0.441	0.011	0.469	0.238	0.049
Moroccan	0.052	0.249	0.834	1.002	0.544	0.065	1.431	0.270	0.000
Surinamese	0.277	0.255	0.278	-0.470	0.757	0.534	0.764	0.311	0.014
Antillean	-0.663	0.290	0.022	0.433	0.585	0.459	-0.594	0.529	0.262
Other: Western	-0.653	0.135	0.000	0.799	0.368	0.030	0.052	0.183	0.775
Other: non-Western	-0.818	0.194	0.000	0.667	0.417	0.110	-0.097	0.256	0.706
Number of friends of non-Dutch national origin	-0.012	0.036	0.732	0.237	0.094	0.012	0.044	0.046	0.335
Age (in months)	-0.004	0.006	0.479	0.046	0.017	0.005	-0.003	0.010	0.772
Average oldest best friend	-0.008	0.028	0.780	0.051	0.066	0.439	-0.043	0.043	0.315
Indegree: popularity nominations in class	0.000	0.002	0.937	-0.019	0.008	0.014	-0.019	0.004	0.000
Activity levels	-0.209	0.088	0.017	-0.324	0.236	0.171	-0.542	0.143	0.000
Digital resources	-0.171	0.039	0.000	-0.236	0.101	0.020	-0.318	0.056	0.000
Female	-0.294	0.081	0.000	-0.871	0.251	0.001	-1.008	0.125	0.000
High school educational track									
Lower preparatory vocational	(ref.)	(ref.)	(ref.)	(ref.)	(ref.)	(ref.)	(ref.)	(ref.)	(ref.)
Medium/lower preparatory vocational	-0.240	0.173	0.164	-0.131	0.350	0.707	0.110	0.253	0.663
Medium/higher preparatory vocational	-0.516	0.190	0.007	-0.038	0.527	0.942	-0.228	0.352	0.518
Higher preparatory vocational	-0.503	0.153	0.001	-0.319	0.319	0.318	-0.040	0.239	0.867
Senior general	-0.533	0.157	0.001	0.040	0.331	0.904	0.059	0.231	0.800
University preparatory	-0.682	0.161	0.000	0.256	0.351	0.466	0.208	0.244	0.395
Behavioral problems	-0.046	0.084	0.580	-0.122	0.226	0.588	-0.330	0.130	0.011
Self-esteem	0.057	0.079	0.469	0.129	0.198	0.515	-0.008	0.114	0.941
Number of friends nominated inside class	0.106	0.051	0.037	0.008	0.096	0.937	0.184	0.068	0.007
Number of friends nominated outside class	-0.122	0.063	0.055	-0.419	0.142	0.003	-0.329	0.077	0.000
Constant	3.450	1.227	0.005	-7.954	3.535	0.024	5.579	2.070	0.007
Observations	3,696								

[a] Robust standard errors, cluster corrected for 220 clusters; [b] Two-sided p-values.

Appendices Chapter 3

Table A3.1 is the full table with the results for the interaction models between peers' privacy settings and the density scores within classes. Table A3.1 refers back to the results in Chapter 3 Figure 3.3, where we test whether the percentage of peers' privacy settings in class have a stronger effect in more connected classes.

Table A3.1: Logistic regression: associations of the interaction between peers' privacy behavior in the class and class density (H2). Odds ratios are presented.

	Pr(Private timeline post)			Pr(Private friend list)		
	O.R.[a]	S.E.[b]	p[c]	O.R.	S.E.	p
% Best Friends' timeline posts private	1.003	0.002	0.117	-	-	-
% Best Friends' friend lists private	-	-	-	0.999	0.002	0.803
% Class timeline posts private	0.978	0.016	0.180	-	-	-
% Class friend lists private	-	-	-	1.010	0.034	0.766
Density	0.119	0.123	0.039	0.613	0.588	0.610
% Class timeline posts private X Density	1.046	0.025	0.056	-	-	-
% Class timeline friends private X Density	-	-	-	0.983	0.047	0.715
Indegree: popularity	0.986	0.003	0.000	0.992	0.003	0.009
Girls (ref.: boys)	1.103	0.078	0.166	1.494	0.149	0.000
Ethnic background (ref.: Dutch)	-	-	-	-	-	-
Turkish	1.507	0.362	0.088	6.678	1.671	0.000
Moroccan	2.537	0.750	0.002	4.183	1.483	0.000
Dutch Caribbean	1.435	0.325	0.111	2.576	0.685	0.000
Other Western	0.842	0.106	0.173	1.545	0.218	0.002
Other non-Western	1.266	0.202	0.140	3.048	0.579	0.000
Educational track (ref.: Voc. educ.)	-	-	-	-	-	-
Senior general	0.824	0.066	0.016	1.139	0.135	0.273
University preparatory	0.828	0.075	0.036	1.254	0.178	0.110
Age in months	0.972	0.005	0.000	0.868	0.011	0.000
Observations	3,434			3,434		
Wald χ^2 (df)	111.87	(14)		265.47	(14)	
Prob. $> \chi^2$	0.000			0.000		
Log pseudolikelihood	-2,303.18			-1,656.43		
Pseudo R^2	0.026			0.148		

[a] Odds ratios; [b] Delta-method standard errors, cluster correction for 287 classes; [c] Two-sided p-values.

Appendices Chapter 4

Appendix 4.1. Predicting Respondents' Facebook friends' Ethnicity and Gender using First Names

We predicted friends' gender and ethnic backgrounds based on their first names. The Facebook networks contain 52,651 unique first names. We matched these (up to the first space or hyphen) to the first names in the Dutch Civil Registration data. We matched 36,151 (68.66 percent) names to the Dutch Civil Registration data (comprising 91.52 percent of names in the Dutch population; N = 14,447,100). These names covered 1,106,675 of respondents' total Facebook friends (95.54 percent). We matched these 36,151 names to the first names of the survey respondents from whom we knew ethnic backgrounds and gender. The number of matched first names is N = 5,613 (94.8 percent of the total N of wave 2).

Next, a threshold that at least 50 percent of the first-name carriers must be female to assign the respondent to a female gender provided the highest correlation between real and assigned gender (r = .96, p < .001). We wrote an algorithm that automatically searched which thresholds of the fraction of the first-name carriers and their parents' country of birth yielded the highest correlations possible between real and assigned ethnicities. It did so by correlating real ethnicity with all possible permutations of the thresholds in steps of 1 percent. The highest possible correlations obtained were .66 (Dutch, 86.5 percent of the cases correctly labeled as Dutch), .87 (Turkish, 98.6 percent labeled correctly), .80 (Moroccan, 98.1 percent labeled correctly), .47 (Dutch Caribbean, 97.4 percent labeled correctly), .21 (other Western, 90.6 percent labeled correctly), and .46 (non-Western, 94.7 percent labeled correctly). Using these threshold configurations, we assigned gender and ethnic backgrounds to respondents' Facebook friends. All of respondents' friends were assigned a gender, and 99.81 percent of Facebook friends were assigned an ethnic background, given our thresholds. If we only run our analysis for Dutch, Turks, and Moroccans — which have the highest correlations — our results do not qualitatively differ from the results presented in the main text of the article.

Appendix 4.2. Specific Ethnic Backgrounds in Online Networks, Broken Down By Ethnic Background of Respondents

Figure A4.1 below shows the relative amount of specific ethnic backgrounds that are present in respondents' Facebook networks, broken down by respondents' own ethnic background.

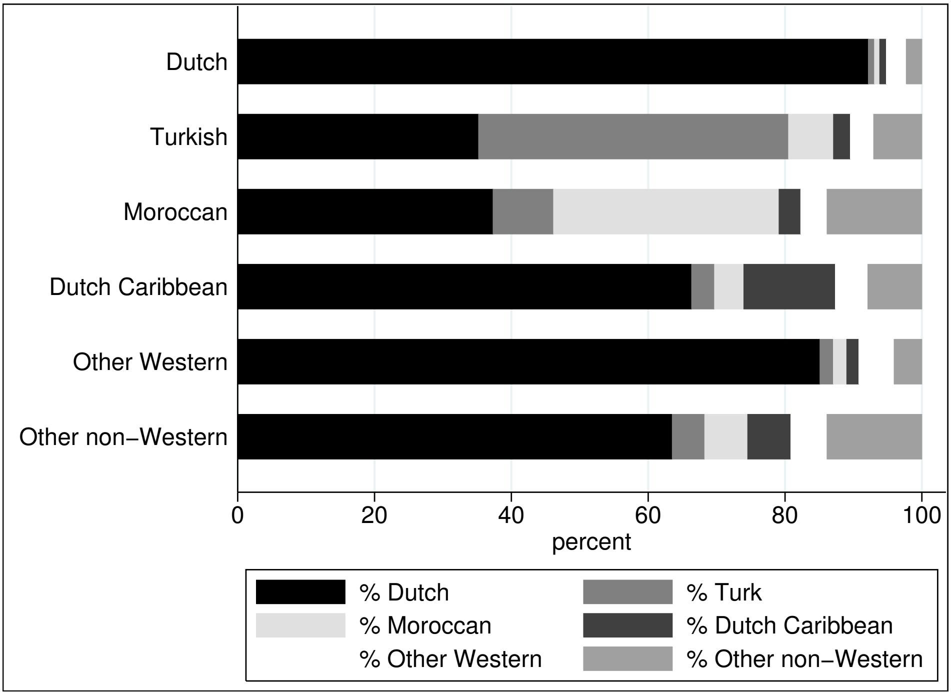


Figure A4.1: Ethnic segregation of social networks online, broken down by ethnicity.

Appendix 4.3. Full Multilevel Regression Table for Tests of Hypothesis 1a

The results found in Table A4.1 show the full results for the test of Hypothesis 1a. Here, we test whether the level of ethnic segregation is higher than that of gender segregation.

Table A4.1: Multilevel model estimating the difference between the percentage of co-ethnic and same-gender friends in online networks.

	Intercept only (H1a)		
	Coef.	S.E.[a]	p[b]
Fixed part			
Intercept	16.851	(1.515)	***
Random part			
$\sigma^2_{s_{0k}}$ (School level)	185.033	(44.113)	
$\sigma^2_{c_{0jk}}$ (Class level)	3.629	(124.899)	
$\sigma^2_{p_{0ijk}}$ (Pupil level)	1016.715	(45.640)	
Number of schools	114		
Number of classes	315		
Number of pupils	2,690		
Log likelihood	-13216.774		

[a] Robust standard errors, adjusted for the school-identifier; [b] One-sided p-values, * p <.05, ** p <.01, *** p <.001.

Appendix 4.4. Full Multilevel Regression Table for Tests of Hypothesis 3c

The results found in Table A4.2 on the next page show the full results for the test of Hypothesis 3c. Here, we test whether the level of ethnic segregation in core networks and Facebook networks differs between ethnic minority and majority members.

Table A4.2: Multilevel model estimating the difference between ethnic segregation in core and online networks.

	Co-ethnic$_{\text{FRIENDS IN GENERAL}}$ - Co-ethnic$_{\text{FACEBOOK}}$		
	Coef.	S.E.[a]	p^{b}
Fixed part			
Intercept	-13.164	(6.136)	*
Core-network			
Co-ethnic$_{\text{FRIENDS IN GENERAL}}$	-	-	-
Co-ethnic$_{\text{FRIENDS IN CLASS}}$	-	-	-
Opportunity			
Co-ethnic$_{\text{IN CLASS}}$	0.004	(0.038)	
Co-ethnic$_{\text{IN SCHOOL}}$	0.155	(0.051)	**
Ethnicity			
Dutch	Ref.	Ref.	Ref.
Turkish	26.375	(5.171)	***
Moroccan	27.692	(5.561)	***
Dutch Caribbean	32.342	(4.668)	***
Other Western	25.326	(3.785)	***
Other non-Western	29.889	(4.607)	***
Number of Facebook friends	0.003	(0.003)	
Facebook membership (ref.: 2013)			
2012	1.843	(4.865)	
2011	1.210	(4.699)	
2010	-0.107	(4.722)	
2009	0.642	(4.814)	
2008	-1.757	(4.672)	
2007	-0.693	(5.337)	
2006	-3.755	(6.975)	
Girls (ref.: boys)	0.551	(0.804)	
Educational track (ref.: lower voc.)			
Senior General	2.189	(0.928)	**
University preparatory	0.647	(1.127)	*
Indegree popularity	-0.071	(0.033)	*
Ethnic outgroup attitudes	-0.761	(0.260)	**
% kinship ties declared	-0.323	(0.337)	
% similar surname on Facebook	-0.170	(0.175)	
Random part			
$\sigma^2_{s_{0k}}$ (School level)	0.000	(0.000)	
$\sigma^2_{c_{0jk}}$ (Class level)	2.643	(4.490)	
$\sigma^2_{p_{0ijk}}$ (Pupil level)	387.398	(28.982)	
Number of schools	112		
Number of classes	309		
Number of pupils	2,549		
Log likelihood	-11220.572		

[a] Robust standard errors, adjusted for the school-identifier; [b] One-sided p-values, * $p <.05$, ** $p <.01$, *** $p <.001$.

Appendix 4.5. Full Multilevel Regression Table for Tests of Hypothesis 3d

The results found in Table A4.3 on the next page show the full results for the test of Hypothesis 3d. Here, we test whether the level of gender segregation is stronger in core networks than it is in Facebook networks.

Table A4.3: Multilevel model estimating the difference between gender segregation in core and online networks.

	Same-gender$_{\text{FRIENDS IN CLASS}}$ - Same-gender$_{\text{FACEBOOK}}$		
	Coef.	S.E.[a]	p[b]
Fixed part			
Intercept	-29.991	(7.390)	***
Core-network			
Same-gender$_{\text{FRIENDS IN CLASS}}$	-	-	-
Opportunity			
Same-gender$_{\text{IN CLASS}}$	0.027	(0.03)	
Same-gender$_{\text{IN SCHOOL}}$	0.047	(0.046)	
Ethnicity			
Dutch	Ref.	Ref.	Ref.
Turkish	19.791	(5.266)	***
Moroccan	16.665	(6.377)	**
Dutch Caribbean	3.218	(4.045)	
Other Western	1.271	(2.214)	
Other non-Western	4.921	(3.830)	
Number of Facebook friends	-0.010	(0.004)	**
Facebook membership (ref: 2013)			
2012	-3.543	(6.967)	
2011	-2.803	(6.326)	
2010	-1.098	(6.356)	
2009	-2.031	(6.527)	
2008	1.233	(6.746)	
2007	10.760	(10.483)	
2006	3.549	(10.683)	
Girls (ref.: boys)	-1.670	(1.528)	
Educational track (ref: lower voc.)			
Senior General	5.468	(1.604)	**
University preparatory	-1.001	(2.014)	
Indegree popularity	0.082	(0.050)	*
Gender role attitudes	0.809	(0.485)	*
% kinship ties declared	-0.420	(0.348)	
% similar surname on Facebook	0.455	(0.366)	
Random part			
$\sigma^2_{s_{0k}}$ (School level)	0.000	(0.000)	
$\sigma^2_{c_{0jk}}$ (Class level)	50.935	(14.677)	
$\sigma^2_{p_{0ijk}}$ (Pupil level)	856.391	(38.042)	
Number of schools	108		
Number of classes	301		
Number of pupils	2,595		
Log likelihood	-12508.024		

[a] Robust standard errors, adjusted for the school-identifier; [b] One-sided p-values, * p <.05, ** p <.01, *** p <.001.

Appendix 4.6. The Correlation between Degree and Ethnic and Gender Homogeneity under Random Mixing

The number of friends (degree) one has is an individual property as well as a function of the network. Therefore, correlations between degree and individual properties could happen by chance. These correlations are not substantive but originate from design. It implies that we cannot simply assume that $H_0 = 0$, because it might be that $H_0 \neq 0$; we cannot conveniently test whether the degree-effect statistically deviates from zero. We first need to examine whether this "correlation-by-design" also exists between homogeneity and degree. Therefore, we model what the correlation between homogeneity and degree would be when respondents are randomly tied to others, while keeping their degree constant.

Per respondent, we randomly sampled K friends from a vector containing the complete gender/ethnic distribution of ties among Facebook friends. K is the number of friends of a given respondent. We sampled without replacement within and over respondents, because one cannot befriend the same person twice and the same tie cannot exist twice. Via this procedure, we generated random realizations of our observed network. Figures A4.2 and A4.3 plot the observed and baseline (one exemplary realization) levels of homogeneity against degree. There does not seem to be a correlation between baseline ethnic homogeneity and degree ($r = .047$; $p = .433$ for Dutch; $r = -.007$; p = .867 for non-Dutch) and gender homogeneity and degree ($r = -.014$; $p = .452$).

Hence, in this study, we assume that for homogeneity and degree, the null hypothesis is that there indeed is no correlation present between homogeneity and the number of friends individuals have. Therefore, we test the alternative hypothesis against this assumed null hypothesis — whether the degree-effects statistically deviate from zero, and we have shown that doing so is statistically plausible given our empirical context.

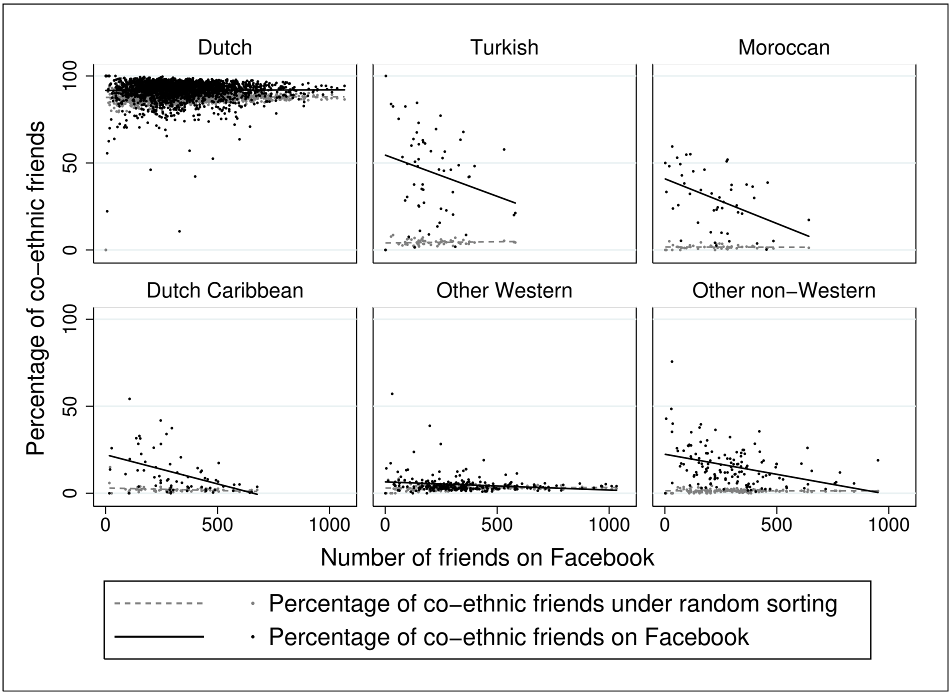


Figure A4.2: Observed and baseline ethnic homogeneity of large personal networks on Facebook by number of friends on Facebook, broken down by ethnicity and including a fitted regression slope.

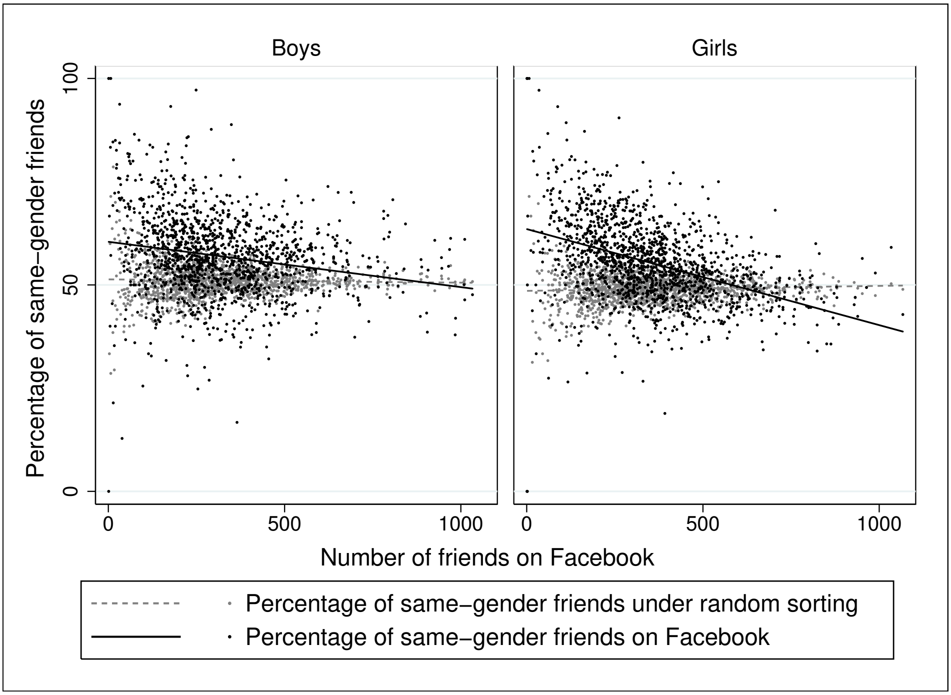


Figure A4.3: Observed and baseline gender homogeneity of large personal networks on Facebook by number of friends on Facebook, broken down by gender and including a fitted regression slope.

Appendices Chapter 5

Appendix 5.1. Populations in the Scale-up Categories in the Netherlands

Table A5.1 provides an overview of the total populations in the Netherlands that are used as the X's in our scale-up method questions in the survey.

Table A5.1: The scale-up first names and cities populations in 2014 (Population in the Netherlands in 2014=16,829,289).

	X's in the population[a,b]
Thomas	40,538
Kevin	23,162
Anne	29,720
Melissa	11,706
Moham(m)ed	13,443
Groningen	198,317
Utrecht	328,164
Maastricht	122,488
Den Haag	508,940
Zwolle	123,159

[a] Firstname population estimates are from Meertens Institute (2016); [b] City estimations are from Statistics Netherlands (2015).

Appendix 5.2. Additional Sensitivity Checks on the Data

Because the network scale-up method is sensitive to outliers we examined the "seriousness" of respondents' answers to the questionnaire. Specifically, we investigate whether there are "straighlining" respondents — i.e., respondents that tick the same box for each of a set of items belonging together. We calculated for items that belong to a battery of questions (e.g., several items measuring "health behavior") whether there were respondents that had a standard deviation of zero on their answers to these items. If a respondent has a standard deviation of zero it means that he/she ticked the boxes of all items in a measurement similarly. In total there were seven batteries of questions, containing 36 items that were not part of different routing options in the questionnaire. Of the 4,073 respondents in wave 4, four respondents (about 0.1%) had a row standard deviation of zero on all seven batteries of questions.

In a less rigid sample selection of scoring a standard deviation of zero over, for instance, three batteries of questions, we end up with the same four respondents. A visual inspection of these respondents' answers to other survey questions also showed careless responses. We did not consider these four respondents in the analyses, as they likely provided inaccurate answers that may disproportionately affect our results.

Appendix 5.3. Results for the Selection part of The Estimation Procedure

Table A5.2 on the next page shows the associations between selected covariates and the number of X's one indicates to know in the survey (i.e., the selection part of our estimation procedure). For instance, those of Turkish and Moroccan origin know less people named "Thomas" than those of Dutch origin, as do members of other ethnic minority groups. The results are as one would expected; those of higher education know more X's of high status (Thomas and Anne), and those of lower education know more X's of low status (Kevin and Melissa). This selection model is applied both for the Facebook data and for the unobserved extended network data (for the extended network, the observed data are measured by interval censoring the unobserved count data).

Table A5.2: Bayesian posterior means and posterior coefficients for the selection submodel, a multinomial logistic regression for the number of friends with names Thomas, Kevin, Anne, Melissa, and other is the reference category, conditional on the size of the Facebook network and the extended network (N = 2,151).

	Thomas				Kevin				Anne				Melissa			
	Mean	2.5%	97.5%	S[a]	Mean	2.5%	97.5%	S	Mean	2.5%	97.5%	S	Mean	2.5%	97.5%	S
Constant	-5.015	-5.077	-4.951	*	-5.583	-5.654	-5.511	*	-5.188	-5.253	-5.126	*	-6.353	-6.439	-6.260	*
Ethnicity																
Dutch (ref.)																
Turkish/Mor.	-1.931	-2.359	-1.540	*	-1.135	-1.448	-0.836	*	-1.942	-2.347	-1.560	*	-0.361	-0.676	-0.068	*
Other min.	-0.365	-0.481	-0.253	*	-0.085	-0.186	0.012		-0.539	-0.659	-0.416	*	-0.058	-0.190	0.075	
Girl (ref. Boy)	-0.228	-0.284	-0.170	*	-0.135	-0.192	-0.077	*	0.007	-0.047	0.060		0.167	0.100	0.236	*
Girl*Turk./Mor.	0.873	0.362	1.388	*	0.372	-0.039	0.788		1.000	0.508	1.498	*	-0.181	-0.597	0.256	
Girl*Other min.	0.044	-0.115	0.199		-0.052	-0.189	0.091		0.175	0.026	0.330	*	-0.048	-0.225	0.125	
Education																
Vocational	-0.484	-0.549	-0.419	*	0.468	0.400	0.539	*	-0.196	-0.262	-0.131	*	0.591	0.504	0.676	*
Senior general	-0.221	-0.290	-0.151	*	0.180	0.098	0.258	*	-0.067	-0.135	0.005		0.389	0.292	0.489	*
Uni. prep.																
Partner (ref.: no)	0.006	-0.052	0.067		0.039	-0.017	0.097		-0.030	-0.086	0.026		0.059	-0.006	0.129	

[a] The coefficients can be considered to be statistically significant if zero is not contained in the interval q(2.5)-q(97.5).

Appendix 5.4. A Straightforward Method to Estimate the Social Network Size Using the Basic Scale-Up Estimator

Table A5.3 on the next page shows a linear regression and a Heckman selection model that regresses the basic scale-up estimate of the extended social network size on the predictor variables mentioned before. The correlation of the residuals in the outcome and the selectivity equations is about .011. We observe that estimating the extended social network size via this straightforward measure does not suffer very much from sample selection biases (as reflected in the low correlation between the residuals). In the Heckman selection model, we cluster-corrected standard errors for the school cluster to which adolescents belong. Multilevel regression models that account for the clustered data structure do not provide qualitatively different results than the simple linear regression model from Table A5.3.

Table A5.3: Maximum-likelihood estimation results of the basic scale-up estimator via a linear regression and a Heckman selection model.

	Linear regression			Heckman selection		
	Coef.	S.E.	p^a	Coef.[b]	S.E.	p
Constant	589.565	141.946	***	602.498	145.173	***
Foci (H1)						
Going out	140.964	32.167	***	140.966	32.246	***
Associations	79.503	22.173	***	79.504	18.406	***
Concerts	93.562	38.544	**	93.562	52.821	*
Similarity of cont. (H2+H3)						
Ethnicity						
Dutch (ref.)						
Arabic	-208.534	165.052		-198.234	160.279	
Other	-13.694	110.030		-9.630	108.429	
# Co-ethnic Class	0.493	5.722		0.493	8.016	
# Co-ethnic School	0.062	0.051		0.062	0.076	
Romantic partners (H4)						
Partner (ref. No)	-38.138	57.000		-38.145	59.611	
Education and gender (H5+H6)						
Education						
Vocational (ref.)						
Senior general	-112.618	64.138	*	-114.368	71.008	
University prep.	-218.062	63.531	***	-220.274	91.151	*
Girl (ref. Boy)	-81.874	50.155		-83.028	61.467	
Observations	2151			5693		
R-squared	0.042					
Log pseudolikelihood				-21752.030		

[a] One-sided p-values: * $p < .05$, ** $p < .01$, *** $p < .001$; [b] In the selection equation we adjusted for ethnic background, gender, and educational level. Boys, ethnic minority members, and lower educated are less likely to have a value on the basic scale-up estimate.

Appendices Chapter 6

The R-code below is the core of the code for the method that we use to predict ethnicity. Some data handling before this code is specific to the data we used. We excluded this part to keep the code parsimonious. In the first section of the code, we write a function for only one bootstrap sample. In the second section, we write a loop around this bootstrap function. In the third section, we can adjust the number of bootstraps we run and the number of computer-cores we want to use. The more bootstraps specified, the longer the computation process will take. However, the more cores one uses, the faster the process will converge.

```
##############################################################
# Authors:  BH & NdS                                         #
# Paper:    Bas Hofstra and Niek de Schipper                 #
#           Predicting Ethnicity in Online Social Networks  #
# Date:     February/March 2017                              #
# Tasks:    Bootstrapped coefficients of conditional         #
#            probabilities to                                #
#           predict ethnicity in Facebook networks           #
#           Simultaneously perform 10k linear regressions    #
#           as a novel way to test hypotheses                #
##############################################################

##############################################################
# 1.    Function for one bootstrap sample                    #
#        core code                                           #
##############################################################

# This is the main function for one bootstrap
bootstrapB <-  function(surveyData, propMatrix, rowSumsCount,
                        count, namesData, friendNetworks){

for(i in 1:nrow(count)){
  if(rowSumsCount[i]){
    propMatrix[i, ] <- count[i, ]
  }else{
  propMatrix[i, ] <- rdirichlet(n = 1, alpha = count[i, ])
  }
}

# Input the matrix for proportion per name
propMatrixName <- data.frame(voornaam = namesData$voornaam,
                             propMatrix, stringsAsFactors = FALSE)
```

```
#merge together
total <- inner_join(friendNetworks, propMatrixName, by = "voornaam")

#sample ethnicity according to propMatrixName
X <-    apply(total[, 3:8], 1, FUN = function(w)
        sample_int_R(6, 1, prob = w))
X <-    data.frame(X, total$userID)

aggregateData <- dcast(X, total.userID ~ X, fun.aggregate = length,
                       value.var = "total.userID")

#merge this data to the survey data
data <- inner_join(aggregateData, surveyData,
        by = c("total.userID" = "userID"))

#add the total number of friends for every resp
data$nfriends <- rowSums(data[, 2:7])

#friends with the same ethnicity as respondent
homoEthnicity <- as.numeric(data[cbind(seq_len(nrow(data)),
                                       match(data$ethnicity,
                                       names(data)))])

#create homogeneity measure
data$homoGeneity <- homoEthnicity / data$nfriends

#standardize variables # Not neccessary for now
#data$satisfaction <- scale(data$satisfaction)
#data$homoGeneity <- scale(data$homoGeneity)
data$ethnicity <- as.factor(data$ethnicity)

# This is the linear regression model
bootstrapIndices <- sample(1:nrow(data), replace = TRUE)
fit <- lm(trust ~ wave + sex + relation + satisfaction
            + homoGeneity + relation + ethnicity,
            data = data[bootstrapIndices, ])
return(fit$coefficients)
}

#############################################################
# 2.    Execute the function from step                      #
#       1 "numberOfSamples" times                           #
#       this function is the analysis for one core          #
#############################################################

analysis <- function(numberOfSamples, surveyData, propMatrix,
```

```
                     rowSumsCount, count, namesData, friendNetworks){
  results <- matrix(NA, numberOfSamples, 12)
  for(i in 1:numberOfSamples){
    results[i, ] <- bootstrapB(surveyData, propMatrix, rowSumsCount,
                               count, namesData, friendNetworks)
  }
  return(results)
}

############################################################
# 3. Call this function for 3 cores                        #
############################################################

#set the number of bootstrap samples for each core
# Calculate how much you want, 3 cores * 3334 =
# 10002 bootstrapped coefficients
# Multiple core to spead up the process
numberOfSamples <- 3334

# Set the seed of the multicore simulation
# If you wan to perform this on multiple computers at once to save time,
# seed need to be different for each computer
seed <- 435136

# Execute these 4 lines to obtain results
cl <- makeCluster(3)      #run on 3 cores, more is faster
registerDoParallel(cl)
bootstrapCoefficients <- foreach(i = 1:3, .combine = rbind,
                        .options.RNG = seed, .packages = c("dplyr",
                        "reshape2", "wrswoR", "gtools"))
                        %dorng% analysis(numberOfSamples,
                        surveyData, propMatrix,
                        rowSumsCount, count, namesData,
                        friendNetworks)
stopCluster(cl)
show(bootstrapCoefficients) # results are found here
```

Nederlandse Samenvatting

Nederlandse Samenvatting

Deze dissertatie heeft als doel te laten zien hoe de analyse van *online sociale netwerken* nieuwe inzichten kan bieden in sociale netwerken in het algemeen. Ik laat zien wat de theoretische en empirische kansen, inzichten en uitdagingen van een dergelijk onderzoek zijn.

De Maatschappelijke en Wetenschappelijke Impact van Online Sociale Netwerken

Ongeveer 95% van de Nederlandse bevolking tussen de 12 en 45 jaar gebruikte in 2016 een sociale netwerksite. Facebook is het grootste sociale media platform wereldwijd; het had in 2016 ongeveer 1.86 miljard maandelijkse gebruikers. Ongeveer 10,4 miljoen (78%) van de Nederlandse bevolking gebruikt Facebook omstreeks januari 2017, waarvan 7,5 miljoen (56%) dagelijks. Deze ongekende populariteit van sociale media heeft invloed op een aantal aspecten van ons dagelijks leven. Onderzoek laat bijvoorbeeld zien dat verschillen in de hoeveelheid tijd dat iemand sociale media gebruikt, van invloed kunnen zijn op de hoeveelheid hulp die iemand van zijn omgeving ontvangt. Daarnaast kunnen verschillen in de manier waarop men met privacy op sociale media omgaat leiden tot meer of minder blootstelling aan hackers of reputatieschade. Verschillen in de mate waarop men met gelijkgestemden op sociale media omgaat, kunnen leiden tot een zogeheten "echo-kamer", waarin meningen in toenemende mate kunnen polariseren. De manier waarop men gebruik maakt van en zich gedraagt op sociale netwerksites heeft implicaties voor de sociale ongelijkheid en sociale cohesie in een samenleving.

Naast deze maatschappelijke implicaties heeft de opkomst van sociale media ook *de manier waarop we wetenschap bedrijven* beïnvloed. Watts (2011: 266) — in een beroemde quote — zegt dat gegevens van sociale media weleens "een telescoop" zouden kunnen zijn voor sociaalwetenschappers. Dit omdat sociale media platformen vaak automatisch alle *digitale voetsporen* opslaan die men online achterlaat. Wetenschappers gebruiken in toenemende mate deze digitale voetsporen om substantieve (sociaal)wetenschappelijke vragen te beantwoorden. Prominente voorbeelden in deze lijn zijn studies over sociale invloed in stemgedrag onder miljoenen Facebookgebruikers, studies naar interetnische contacten onder grote netwerken op Facebook en studies naar de netwerkstructuur onder miljoenen individuen. Deze studies laten zien dat één eigenschap van sociale media in het bijzonder de interesse van sociaalwetenschappers wekt — de sociale netwerken die vaak een belangrijk onderdeel zijn van deze platformen. Dit is vrij logisch, aangezien er twee voordelen zijn van het analyseren van online sociale netwerken ten opzichte van sociale netwerken gemeten via vragenlijsten. Ten eerste meten online sociale netwerken daadwerkelijk *gedrag* in plaats van zelf-gerapporteerde, mogelijk sociaalwenselijke gegevens zoals in vragenlijsten. Ten tweede zijn online sociale netwerken *groter* en brengen ze vaak honderden contacten meer in kaart dan de kleinere netwerken (bijv. 5 tot 10 contacten) die ontstaan uit vragenlijsten.

De voordelen van het bestuderen van online sociale netwerken ten opzichte van netwerken in vragenlijstonderzoek bieden kansen om bestaande hypothesen op een nieuwe — wellicht zelfs betere — manier te toetsen. Daarnaast zorgen gegevens van online sociale netwerken ervoor dat er theoretische vooruitgang kan worden geboekt, aangezien deze gegevens ons soms in staat stellen hypothesen te toetsen die voorheen lastig te toetsen waren.

Data over Offline en Online Sociale Netwerken

Sommige studies analyseren exclusief online netwerkgegevens. Andere studies bestuderen weer alleen vragenlijstgegevens over netwerken op sociale media. Beide benaderingen hebben nadelen; online netwerkgegevens zijn op individueel niveau (bijv. persoonskenmerken) vaak onvoldoende gedetailleerd. Daartegenover bieden vragenlijstgegevens vaak geen inzicht in de honderden sociale relaties die men op sociale media onderhoudt. Een substantiële bijdrage van deze dissertatie is dat ze offline gegevens van duizenden Nederlandse adolescenten koppelt aan gegevens die verzameld zijn via het sociale media platform Facebook. Op deze manier verkrijg ik zowel gedetailleerd inzicht in individuele kenmerken en attitudes als informatie over goede vrienden uit de vragenlijst, terwijl ik daarnaast de grote netwerken online observeer.

Overkoepelende Onderzoeksvragen

Het doel van deze dissertatie is het analyseren van online sociale netwerken om nieuw inzicht te verkrijgen in sociale netwerken in algemene zin. Om dit doel te bereiken stel ik twee overkoepelende onderzoeksvragen.

De eerste onderzoeksvraag is: *in welke mate zijn er verschillen tussen individuen wat betreft* ***activiteiten op sociale media*** *en hoe zijn deze te verklaren?* De activiteiten die ik bestudeer zijn *lidmaatschap* en *privacy* op sociale media. Een methodologisch argument om deze twee vormen van activiteit te onderzoeken is dat zij *steekproef selectie fouten* specificeren. Namelijk, wie zitten er überhaupt op sociale media en, gegeven lidmaatschap, wiens sociale netwerken kunnen we observeren? Naast dit methodologische argument is het vanzelfsprekend dat deze vormen ook vanuit een maatschappelijk perspectief relevant zijn. Dit geldt zowel voor lidmaatschap — gebruik van sociale media heeft effecten heeft op de mate waarin je hulp krijgt uit je sociale omgeving — als voor privacy — wie zijn de mensen die vatbaarder zijn voor identiteitsfraude, ongewilde blootstelling aan derde partijen en reputatieschade?

De tweede onderzoeksvraag is: *in welke mate zijn er verschillen tussen individuen wat betreft de **structuur van online sociale netwerken** en hoe zijn deze te verklaren?* De dimensies van structuur van online sociale netwerken die ik bestudeer zijn segregatie — met wie is men vrienden online? — en grootte — met hoeveel mensen is men vriend? Beide dimensies relateren aan een aantal sociologisch relevante uitkomsten. Twee klassieke argumenten zijn dat diversiteit in netwerken verband houdt met kansen op de arbeidsmarkt en vooroordelen ten opzichte van anderen. Netwerkgrootte houdt verband met gezondheid, welzijn, sociale hulp en mortaliteit. Inherent aan deze tweede onderzoeksvraag is dat ik nieuwe methoden ontwikkel die mij in staat stellen deze onderzoeksvraag te beantwoorden.

Hoofdstuk 2: Wie Was het Eerst op Facebook?

In **Hoofdstuk 2** identificeer ik eerst een set factoren die lidmaatschap op sociale netwerksites bevorderen. Daarna bestudeer ik wat *early adoption* van Facebook veroorzaakt.

Ik draag op twee manieren bij aan de bestaande (overigens opvallend beperkte hoeveelheid) literatuur op dit vlak. Ten eerste, een belangrijke eigenschap van sociale media is dat de populariteit van sociale media-platformen periode-afhankelijk is. Ik bestudeer lidmaatschap op Facebook in 2010, en vergelijk dit met lidmaatschap op Hyves, een vergelijkbaar *Nederlands* platform dat in 2010 veel populairder was dan Facebook. Dit zorgt ervoor dat ik inzicht krijg in de oorzaken van vroeg lidmaatschap op één van de meest prominente communicatieplatformen van het laatste decennium (Facebook), terwijl er al een vergelijkbaar en veel populairder platform op de Nederlandse markt was. Ten tweede, ik ben de eerste die peer-invloed bestudeerd in lidmaatschap op sociale media. Peer-invloed is het fenomeen waarin het gedrag van individuen in een groep meer op elkaar gaat lijken naarmate de tijd vordert.

Wat zorgt ervoor dat men lid wordt van een sociaal media-platform? Ik voorspel en vind dat degenen die sociaal actiever zijn — bijv. zij die vaker lid zijn van een sportclub — een grotere kans hebben om lid te zijn van een sociale netwerksite. Waarschijnlijk komt dit doordat deze personen in sociale media een uitlaatklep vinden om ervaringen uit hun leven te delen. Daarnaast voorspel en vind ik dat individuen met meer digitale hulpbronnen vaker lid zijn van sociale media. Deze bevindingen zijn consistent met het concept van *diffusie van innovaties.* Dit concept specificeert dat specifieke leefstijlen en blootstelling aan technologie veroorzaakt dat deze technologie geadopteerd wordt. Tot slot vind ik dat leden van de Nederlandse etnische meerderheid en meisjes vaker lid zijn dan hun tegenhangers.

Wat zorgt voor *early adoption* van Facebook? Voor leden van een etnische minderheid (bijv. degenen met Marokkaanse ouders) in Nederland had Facebook een belangrijk voordeel ten opzichte van Hyves. Facebook is namelijk een internationaal platform, daar waar Hyves Nederlands was. Veel adolescenten in Europa hebben transnationale banden. Facebook zorgt derhalve voor betere mogelijkheden om met familieleden in het buitenland te communiceren dan Hyves. Dit kan een reden zijn waarom leden van een etnische minderheid eerder Facebook gebruikten dan leden van de etnische meerderheid. Daarnaast heb ik de rol van peer-invloed onderzocht. Wanneer vrienden lid zijn van Facebook (of Hyves), dan neemt de kans op Facebook-lidmaatschap (of Hyves) aanzienlijk toe.

Hoofdstuk 3: Wie Kiest er voor Privacy op Facebook?

In **Hoofdstuk 3** beschrijf en onderzoek ik de oorzaken van privacy-instellingen op Facebook. In andere woorden — als men eenmaal lid is, wat kunnen we zien van deze leden op Facebook?

Er zijn twee manieren waarop ik voortbouw op voorgaand onderzoek. Ten eerste ontwikkel ik een theoretische verklaring voor de consistente bevinding dat jongere mensen en vrouwen vaker kiezen voor meer privacy op sociale media. Ik bestudeer of het hebben van minder vertrouwen onder deze groepen misschien een rol speelt. Met andere woorden — hebben zij die minder vertrouwen in anderen hebben ook vaker privé-profielen op sociale media? Eerder werk suggereert dat leden van een etnische minderheid en degenen met lagere opleidingsniveaus ook minder vertrouwen rapporteren. Daarom onderzoek ik ook of er verschillen in privacy zijn wat betreft etnische achtergrond en opleidingsniveau. Ten tweede bestudeer ik daadwerkelijke privacy *instellingen*, daar waar ander onderzoek vaak kijkt naar zelf-gerapporteerde privacy op sociale media. Als men wordt ondervraagd over privacy overschat men vaak zijn privacy. Het onderzoeken van *instellingen* voorkomt dit probleem.

Wat zijn de oorzaken van privacy op Facebook? Op basis van theorieën over peer-invloed voorspel en vind ik dat degene die meer individuen met een afgeschermd Facebook-profiel in zijn/haar sociale omgeving heeft, er zelf er ook vaker voor kiest om zijn/haar profiel af te schermen. Opmerkelijk is dat deze relatie sterker wordt naarmate meerdere individuen in deze sociale omgeving met elkaar bevriend zijn. Dit komt waarschijnlijk doordat gedrag in hechtere groepen sneller wordt verspreid en normen gemakkelijker gewaarborgd en afgedwongen kunnen worden. De bevindingen in dit hoofdstuk suggereren verder dat meisjes, leden van een etnische minderheid, lager opgeleiden en jongere mensen vaker voor privacy op Facebook kiezen. Deze bevindingen zijn consistent met eerder onderzoek dat aantoont dat deze groepen ook minder vertrouwen rapporteren.

Hoofdstuk 4: Hoe Gesegregeerd Zijn Netwerken op Facebook?

Hoofdstuk 4 heeft twee doelen — ten eerste het beschrijven en verklaren van gender- en etnische segregatie van zwakke banden op Facebook, en ten tweede het verklaren van verschillen in segregatie tussen sterke banden en zwakke banden. De sterkte van sociale relaties wordt bepaald door de hoeveelheid tijd, emotionele intensiteit en intimiteit die men investeert in een sociale relatie.

Er is een drietal manieren waarop ik voortbouw op eerder onderzoek. Ten eerste stel ik voor dat het analyseren van *online* sociale netwerken nieuwe kansen biedt voor het onderzoeken van segregatie van zwakke banden — iets waar we vrij weinig van weten. Facebook-netwerken zijn specifiek hiervoor een goed instrument, aangezien zij een groot deel van totale offline-netwerken in kaart brengen.

Omdat, ten tweede, voorgaand onderzoek zich vooral gericht heeft op relatieformatie onder sterke banden, weten we vrij weinig over de oorzaken van segregatie in grotere netwerken. In dit hoofdstuk toon ik aan wat deze oorzaken zijn. Specifiek bestudeer ik de rol van ontmoetingskansen in de vorm van *relatieve groepsgrootte* en *foci*. Relatieve groepsgrootte specificeert het mechanisme waarin persoonlijke netwerken een afspiegeling zijn van de verdeling van verschillende groepen (bijv. de beide genders of etniciteit) in de populatie. Het foci-mechanisme specifieert het proces waarin mensen die een sociale context delen (bijv. scholen of buurten) een grote kans hebben een sociale relatie met elkaar aan te gaan. Omdat foci vaak gesegregeerd zijn, zullen persoonlijke netwerken een afspiegeling zijn van de samenstelling van de foci. Beide mechanismes hebben effecten op segregatie onder sterke banden. De vraag is echter óf en in welke mate deze mechanismes segregatie in grote netwerken op Facebook voorspellen.

Ten derde wordt in de literatuur gespeculeerd dat sterke banden minder divers zijn dan zwakke banden. Weinig studies hebben dit echter empirisch onderzocht en de condities waaronder dit patroon kan ontstaan zijn niet gespecificeerd. In dit hoofdstuk toets ik nieuw- ontwikkelde hypotheses over het verschil in segregatie tussen sterke en zwakke banden. Hierin analyseer ik de rol van ontmoetingskansen, *homophily*, en balans. Homophily (of: *voorkeuren*) refereert aan het mechanisme waarin individuen een inherente voorkeur hebben om relaties aan te gaan met anderen die op hen lijken. Balans refereert aan het sluiten van triades (of: transiviteit) in netwerken: wanneer A en B vrienden zijn, en A en C vrienden zijn, hebben B en C een hogere kans om vrienden te worden.

Wat voorspelt segregatie op Facebook? Om deze vraag te beantwoorden heb ik de relatieve verdeling van verschillende groepen in de Nederlandse populatie en sociale contexten onderzocht. De verdeling van beide genders in de populatie is 50/50, maar de verdeling van etnische groepen is vaak ongelijker. Gegeven deze discrepantie, vind ik een hogere etnische dan gender-segregatie op Facebook. Omdat leden van de etnische meerderheid per definitie meer potentiële contacten hebben met dezelfde etniciteit, verwacht en vind ik dat leden van de etnische meerderheid een hogere mate van segregatie op Facebook hebben dan leden van een etnische minderheid. Ook verdelen verschillende groepen in een populatie zich vaak op een onwillekeurige manier over foci en de structurele samenstellingen van deze foci beïnvloeden de samenstelling van sociale netwerken. Daarom vind ik een sterk effect van de segregatie in foci op de mate van segregatie in Facebook-netwerken.

Wat verklaart verschillen in segregatie tussen sterke en zwakke banden? Ik beargumenteer dat grote Facebook-netwerken in eerste instantie een afspiegeling zijn van de structurele kenmerken van ontmoetingskansen. Over een langere periode echter, hebben paren die gekenmerkt worden door homogeniteit — bijvoorbeeld tussen twee individuen met dezelfde etniciteit — een hogere kans om een sterke band te krijgen dan paren die minder op elkaar lijken. Dit is een gevolg van investeringen in sociale relaties die lager en opbrengsten die stabieler zijn onder homogene paren. Daarnaast is transitiviteit sterker onder homogene triades. Daarom voorspel en vind ik onder zwakke banden lagere gender-segregatie dan onder sterke banden en zijn zwakke banden van leden van een etnische minderheid minder etnisch gesegregeerd dan hun sterke banden. Omdat leden van de etnische meerderheid erg gelimiteerde kansen hebben om etnische minderheden te ontmoeten, gaat dit mechanisme niet voor hen op; zowel hun sterke en zwakke banden zijn etnisch zeer homogeen.

Hoofdstuk 5: Hoe Groot Zijn Netwerken op Facebook?

In **Hoofdstuk 5** schat ik ten eerste hoe groot sociale netwerken op Facebook zijn. Hierna verklaar ik individuele variatie in deze netwerkgrootte.

Er zijn twee manieren waarop ik voortbouw op eerdere literatuur. Ten eerste lever ik een *methodologische* bijdrage doordat ik een eerdere vragenlijstmeting (de zogenaamde *opschaalmethode*) combineer met het aantal vrienden op Facebook voor een accuratere schatting van het aantal zwakke banden. Ten tweede lever ik een *theoretische* bijdrage doordat ik de individuele variatie in het aantal zwakke banden schat. Tot nu toe is er geen duidelijk theoretisch kader noch is er een systematische studie naar de oorzaken van het aantal zwakke banden. In dit hoofdstuk gebruik ik klassieke theorieën over ontmoetingskansen, voorkeuren en romantische partners en onderzoek ik de rol van educatie en gender om nieuwe hypothesen te ontwikkelen over individuele verschillen in het aantal zwakke banden.

Wat bepaalt netwerkgrootte? Ik voorspel en vind dat degenen die meer tijd doorbrengen in sociale foci (bijv. sportclubs) grotere netwerken hebben. Daarnaast voorspel en vind ik dat leden van de etnische meerderheid grotere Facebooknetwerken hebben, waarschijnlijk doordat deze personen een hogere kans hebben om potentiële contacten met dezelfde etnische achtergrond te ontmoeten. Ook hebben zij die zich in foci met meer etnisch gelijke mensen bevinden grotere Facebooknetwerken. Tot slot, degenen in een romantische relatie, met een hoger opleidingsniveau en meisjes hebben grotere Facebooknetwerken.

Hoofdstuk 6: Hoe Kunnen We Online Data Verrijken?

Hoofdstuk 6 is een methodestudie waarin ik allereerst de meest waarschijnlijke etniciteit van personen voorspel op basis van hun voornaam. Hierna laat ik zien hoe men hypothesen kan toetsen met deze voorspellingen voor etniciteit op basis van voornamen. Dit type dataverrijking is cruciaal voor studies naar de structuur van online sociale netwerken, omdat de gedetailleerdheid van onlinegegevens op individueel niveau vaak veel lager is dan in traditioneel vragenlijstonderzoek. Individuele kenmerken zoals gender of etniciteit ontbreken vaak in gegevens van online sociale media.

Ik bouw op twee manieren voort op voorgaande literatuur. Ten eerste neem ik in mijn analyses de statistische onzekerheid mee die voortkomt uit het feit dat ik namen gebruik om etniciteit te voorspellen voor een meer realistische schatting van etniciteit. Ten tweede laat ik zien hoe men hypotheses kan toetsen met voorspelde variabelen, rekening houdend met de statistische onzekerheid die voortkomt uit deze voorspellingen. Om de potentie van deze nieuwe methode te laten zien onderzoek ik de relatie tussen voorspelde waarden van etnische homogeniteit in Facebooknetwerken en vertrouwen.

De bevindingen laten zien dat mijn voorspelde waarde van etnische homogeniteit op Facebook niet samenhangt met vertrouwen en dit is consistent met recent onderzoek. Een vergelijking met twee simpele methoden om etniciteit te voorspellen op basis van voornamen laat zien dat mijn nieuwe methode het minst vatbaar is voor vals-positieve resultaten, resulterend in conservatieve testen van hypothesen over de potentiële consequenties van netwerkstructuur online.

Conclusies

De eerste onderzoeksvraag was: *in welke mate zijn er verschillen tussen individuen wat betreft* ***activiteiten op sociale media*** *en hoe zijn deze te verklaren?* **Hoofdstukken 2** en **3** onderzoeken deze individuele verschillen in *lidmaatschap* en *privacy* op sociale media. Mijn dissertatie laat zien dat degenen die minder sociaal actief zijn, jongens, etnische minderheden, degenen met weinig vrienden op sociale media, en degenen met weinig digitale hulpbronnen waarschijnlijk ondervertegenwoordigd waren in studies die publieke data van sociale media tot 2010 gebruikten (de tijdsperiode waarin ik dit onderzocht heb). Bevindingen over Facebook-privacy laten ook serieuze selectiviteit zien. Deze selectiviteit in privacy is potentieel schadelijker dan selectiviteit in lidmaatschap, omdat lidmaatschap in Facebook sterk is gegroeid (van 84% in 2010 tot 95% in 2014), daar waar ongeveer 25% hun Facebook-netwerken afschermt. Specifiek vind ik dat etnische minderheden, meisjes, jongere adolescenten, lager opgeleiden, degenen die meer vrienden hebben die kiezen voor een afgeschermd profiel, en degenen die minder populair zijn een hogere kans hebben om tot de 25% te behoren die hun profiel afschermen op Facebook. Deze groepen zijn ondervertegenwoordigd in studies die gebruik maken van online sociale netwerkdata. Met name **Hoofdstuk 5**, over het aantal zwakke banden, laat zien dat het cruciaal is om rekening te houden met deze selectiviteit in privacy.

De tweede onderzoeksvraag die ik stelde was: *in welke mate zijn er verschillen tussen individuen wat betreft de **structuur van online sociale netwerken** en hoe zijn deze te verklaren?* **Hoofdstukken 4** en **5** onderzoeken de oorzaken van de segregatie en grootte van netwerken op Facebook. Hier heb ik zowel klassieke hypotheses over netwerkformatie op een nieuwe manier getoetst als theoretische vooruitgang geboekt. Over het algemeen concludeer ik dat de kernhypotheses over netwerkformatie — bijvoorbeeld over ontmoetingskansen — ook netwerkformatie op Facebook voorspellen. Ik laat zien dat zowel de relatieve grootte van groepen in de populatie als de structurele kenmerken van sociale contexten de samenstelling en grootte van netwerken beïnvloeden. Daarnaast heb ik een nieuw theoretisch mechanisme beschreven waarom er tussen homogene paren vaker een sterke band ontstaat dan onder heterogene paren. In **Hoofdstuk 6** stel ik een nieuwe methode voor om online data te verrijken. Dit hoofdstuk maakt voor toegepaste sociale wetenschappers inzichtelijk hoe men de gevolgen van online netwerkstructuren potentieel kan onderzoeken.

Discussie

Ondanks dat ik een aantal cruciale dimensies van sociale media heb onderzocht, is een aantal aspecten in deze dissertatie onderbelicht gebleven en is er een aantal beperkingen aan het onderzoek dat ik hier presenteer. Ik beschrijf hier drie onderwerpen die cruciaal zijn voor toekomstig onderzoek.

Ten eerste — de databronnen die vaak worden gebruikt om sociale media te analyseren zijn bijna zonder uitzondering onwillekeurige steekproeven onder studenten. Deze dissertatie gebruikt een representatieve steekproef onder adolescenten en is dus een stap voorwaarts ten opzichte van voorgaand onderzoek. Echter, volwassenen zijn het afgelopen decennium ook massaal lid geworden van sociale media.

Het is daarom cruciaal dat toekomstig onderzoek zich gaat richten op activiteit op sociale media (bijv. privacy) en online netwerkstructuur (bijv. segregatie) onder volwassenen.

Ten tweede — deze dissertatie focust exclusief op de oorzaken van activiteit en structuur van online sociale netwerken. Een cruciale stap voor toekomstig onderzoek is om de *gevolgen* van de structuur van online sociale netwerken te onderzoeken. Ondanks dat een groeiend aantal wetenschappers zich hiermee bezighoudt, zijn de afhankelijke variabelen in dit onderzoek vaak gemeten uit de onlinegegevens en ontbreken details op individueel niveau. Onderzoek dat gebruik maakt van dezelfde combinatie van vragenlijst- en onlinegegevens zoals in deze dissertatie zou een belangrijke en innovatieve stap voorwaarts zijn. Men kan bijvoorbeeld denken aan het onderzoeken van de relatie tussen etnische segregatie op Facebook en etnische vooroordelen, als een direct vervolg op studie gepresenteerd in **Hoofdstuk 4**. In een maatschappij die in toenemende mate multi-etnisch is en waar tegelijkertijd meer mensen hun contacten onderhouden op sociale media, is dit een belangrijk onderwerp om meer inzicht in te verkrijgen.

Tot slot — wetenschappers zouden zich kunnen richten op andere platformen dan Facebook, aangezien er vele andere (vaak regio-specifieke) sociale media platformen zijn. De populariteit van deze platformen is vaak zeer volatiel, zoals ik in **Hoofdstuk 2** heb laten zien. Het zou interessant zijn om te zien of de determinanten van lidmaatschap overlappen; is het proces van migratie van het ene platform naar het andere altijd het resultaat van dezelfde combinatie van oorzaken? Wetenschappers verzamelen steeds vaker gegevens uit verschillende bronnen, zoals vragenlijsten, online sociale netwerken, maar ook mobiele telefoondata en geo-locaties, van dezelfde groep respondenten. Dit impliceert dat we, met deze gecombineerde gegevens, gedrag over platformen binnen dezelfde groep individuen kunnen vergelijken. Dit is een cruciale stap voor toekomstig onderzoek, aangezien zulke projecten theorieën over sociale netwerkformatie op een zeer innovatieve manier kunnen toetsen. Een vraag die bijvoorbeeld beantwoord zou kunnen worden is welke sociale contacten tussen online platformen en offline contexten overlappen en om welke reden.

References

References

Abascal, M. and Baldassarri, D. (2015). Love Thy Neighbor? Ethnoracial Diversity and Trust Reexamined. *American Journal of Sociology, 121*(3), 722-782.

Acquisti, A. and Gross, R. (2006). Imagined Communities: Awareness, Information Sharing, and Privacy on the Facebook. In Danezis, G. and Golle, P. (Eds.), *Privacy Enhancing Technologies* (pp. 36-58). Berlin, Heidelberg: Springer.

Acquisti, A. and Gross, R. (2009). Predicting Social Security Numbers from Public Data. *Proceedings of the National Academy of Sciences, 106*(27), 10975–10980.

Agresti, A. and Agresti, B.F. (1977). Statistical Analysis of Qualitative Variation. In Schuessler, K.F. (Ed.), *Sociological Methodology* (pp. 204-237). San Francisco: Jossey-Bass.

Ajzen, I. (1991). The Theory of Planned Behavior. *Organizational Behavior and Human Decision Processes, 50*(2), 179-211.

Allport, G.W. (1954). *The Nature of Prejudice.* Reading: Addison-Wesley.

Alesina, A. and La Ferrara, E. (2002). Who Trusts Others? *Journal of Public Economics, 85*(2), 207-234.

Anderson, A., Goel, S., Huber, G., Malhotra, N., and Watts, D.J. (2014). Political Ideology and Racial Preferences in Online Dating. *Sociological Science, 1*, 28-40.

Bäckström, L., Boldi, P., Rosa, M., Ugander, J., and Vignat, S. (2012). Four Degrees of Separation. *Working Paper.* Retrieved May 1, 2017 (arxiv.org/abs/1111.4570)

Baerveldt, C., Van Duijn, M.A.J., Vermeij, L., and Van Hemert, D.A. (2004). Ethnic Boundaries and Personal Choice. Assessing the Influence of Individual Inclinations to Choose Intra-Ehnic Relationships on Pupils' Networks. *Social Networks, 26*(1), 55-64.

Bakshy, E., Messing, S., and Adamic, L.A. (2015). Exposure to Ideologically Diverse News and Opinion on Facebook. *Science, 348*(6239), 1130-1132.

Barber, B. (1983). *The Logic and Limits of Trust.* New Jersey: Rutgers University Press.

Bastani, S. (2007). Family Comes First: Men's and Women's Personal Networks in Tehran. *Social Networks, 29*, 357-374.

Bearman, P. and Parigi, P. (2004). Cloning Headless Frogs and Other Important Matters: Conversation Topics and Network Structure. *Social Forces, 83*(2), 535-557.

Bickart, K.C., Wright, C.I., Dautoff, R.J., Dickerson, B.C., and Feldman Barret, L. (2011). Amygdala Volume and Social Network Size in Humans. *Nature Neuroscience, 14*, 163-164.

Blackwell, D.L. and Lichter, D.T. (2004). Homogamy Among Dating, Cohabiting and Married Couples. *Sociological Quarterly, 45*(4), 719–37.

Blau, P.M. (1977a). *Inequality and Heterogeneity: A Primitive Theory of Social Structure.* New York: Free Press.

Blau, P.M. (1977b). A Macrosociological Theory of Social Structure. *American Journal of Sociology, 83*(1), 26-54.

Bloothooft, G. and Schraagen, M. (2011). Name Fashion Dynamics and Social Class. *Working paper.* Retrieved August 8, 2016 (www.let.uu.nl/∼Gerrit.Bloothooft)

Bloothooft, G. and Onland, D. (2011). Socioeconomic Determinants of First Names. *Names, 59*(1), 25-41.

Bobbitt-Zeher, D. (2007). The Gender Income Gap and the Role of Education. *Sociology of Education, 80*(1), 1-22. doi:

Bond, R.M., Fariss, C.J., Jones, J.J., Kramer, A.D.I., Marlow, C., Settle, J.E., and Fowler, J.H. (2012). A 61-Million-person Experiment in Social Influence and Political Mobilization. *Nature, 489*, 295-298.

boyd, d.m. and Ellison, N.B. (2007). Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication, 13*(1), 210-223.

boyd, d.m. and Hargittai, E. (2010) Facebook Privacy Settings: Who Cares? *First Monday, 8*(2).

boyd, d.m. and Crawford, K. (2012). Critical Question for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon. *Information, Communication and Society, 15*(5), 662-679.

Brandtzæg, P.B. (2012). Social Networking Sites: Their Users and Social Implications — A Longitudinal Study. *Journal of Computer-Mediated Communication, 17*(4), 467-488.

Brashears, M.E. (2014). “Trivial” Topics and Rich Ties: The Relationship Between Discussion Topic, Alter Role, and Resource Availability Using the “Important Matters” Name Generator. *Sociological Science, 1*, 493-511.

Brashears, M.E., Hoagland. E., and Quintane, E. (2016). Sex and Network Recal Accuracy. *Social Networks 44*, 74-84.

Brechwald, W.A. and Prinstein, M.J. (2011). Beyond Homophily: A Decade of Advances in Understanding Peer Influence Processes. *Journal of Research on Adolescence, 21*(1), 166-179.

Brooks, B., Hogan, B., Ellison, N.B., Lampe, C., and Vitak, J. (2014). Assessing Structural Correlates to Social Capital in Facebook Ego Networks. *Social Networks, 38*, 1-15.

Burke, M., Kraut, R., and Marlow, C. (2011). Social Capital on Facebook: Differentiating Uses and Users. *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems*, 571-58.

Burt, R.S. (1984). Network Items and the General Social Survey. *Social Networks, 6*(4), 293-339.

Burt, R.S. (2000). The Network Structure of Social Capital. *Research in Organizational Behavior, 22*, 345-423.

Byrne, D.E. (1971). *The Attraction Paradigm.* New York: Academic Press.

Caers, R., De Feyter, T., De Couck, M., Stough, T., Vigna, C., and Du Bois, B. (2013). Facebook: A Literature Review. *New Media & Society, 15*(6), 982-1002.

Cameron, A.C. and Trivedi, P.K. (2010). *Microeconometrics Using Stata.* Stata Press: College Station.

Castles, S., De Haas, H., and Miller, M.J. (2013). *The Age of Migration: International Population Movements in the Modern World.* 5th ed. New York: The Guilford Press.

Centola, D. (2010). The Spread of Behavior in an Online Social Network Experiment. *Science, 329*(5996), 1194-1197.

Centola, D. (2015). The Social Origins of Network and Diffusion. *American Journal of Sociology, 120*(5), 1295-1338.

Cesare, N., Grant, C., and Nsoesie, E.O. (2017). Detection of User Demographics on Social Media: A Review of Methods and Recommendations for Best practices. *Working paper.* Retrieved February 13, 2017 (arxiv.org/abs/1702.01807)

Chang, J., Rosenn, I., Backstrom, L., and Marlow, C. (2010). ePluribus: Ethnicity on Social Networks. *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, 18-25.

Chen, Y. and Volker, B. (2016). Social Capital and Homophily Both Matter for Labor Market Outcomes - Evidence from Replication and Extension. *Social Networks, 45*, 18-31.

CILS4EU. *Children of Immigrants Longitudinal Survey in Four European Countries. Codebook. Wave 2 - 2011/2012, v2.3.0.* Mannheim University. 2016. Found at: www.cils4.eu

Cheung, C.M.K., Chiu, P.Y., and Lee, M.K.O. (2011). Online Social Networks: Why do Students Use Facebook? *Computers in Human Behavior, 27*(4), 1337-1343.

Christofides, E., Muise, A., and Desmarais, S. (2012). Hey Mom, What's on Your Facebook? Comparing Facebook Disclosure and Privacy in Adolescents and Adults. *Social Psychological and Personality Science, 3*(1), 48-54.

Cialdini, R.B. and Goldstein, N.J., 2004. Social Influence: Compliance and Conformity. *Annual Review of Psychology, 55*, 591–621.

Coldman, A.J., Braun, T., and Gallagher, R.P. (1998). The Classification of Ethnic Status Using Name Information. *Journal of Epidemiology and Community Health, 42*(4), 390-395.

Coleman, J.S. (1961). *The Adolescent Society.* New York: Free Press.

Coleman, J.S. (1990) *Foundations of Social Theory.* Cambridge: Harvard University Press.

Corten, R. (2012). Composition and Structure of a Large Online Social Network in the Netherlands. *PLoS ONE, 7*(4), e3476.

Corten, R. and Knecht, A. (2013). Alcohol Use Among Adolescents as a Coordination Problem in a Dynamic Network. *Rationality and Society, 25*(2), 146-177.

Currarini, S., Jackson, M.O., and Pin, P. (2010). Identifying the Roles of Race-Based Choice and Chance in High School Friendship Network Formation. *Proceedings of the National Academy of Sciences, 107*(11), 4857-4861.

Davis, S.N. and Greenstein, T.N. (2009). Gender Ideology: Components, Predictors, and Consequences. *Annual Review of Sociology, 35*, 87-105.

Delhey, J. and Newton, K. (2005). Predicting Crossnational Levels of social trust: Global Pattern or Nordic Exceptionalism? *European Sociological Review, 21*(4), 311-327.

Delhey, J., Newton, K., and Welzel, C. (2011). How General Is Trust in "Most People"? Solving the Radius of Trust Problem. *American Sociological Review, 76*(5), 786-807.

Dijkstra, J.K., Lindenberg, S., Verhulst, F.C., Ormel, J., and Veenstra, R. (2009). The Relation Between Popularity and Aggressive, Destructive, and Norm-Breaking Behaviors: Moderating Effects of Athletic Abilities, Physical Attractiveness, and Prosociality. *Journal of Research on Adolescence, 19*(3), 401-413.

Dijkstra, J.K., Cillessen, A.H.N., Lindenberg, S., and Veenstra, R. (2010) Basking in Reflected Glory and Its Limits: Why Adolescents Hang Out With Popular Peers. *Journal of Research on Adolescence, 20*(4), 942-958.

DiPrete, T.A., Gelman, A., McCormick, T.H., Teitler, T., and Zheng, T. (2011). Segregation in Social Networks Based on Acquaintanceship and Trust. *American Journal of Sociology, 116*(4), 1234–1283.

DiMaggio, P., Hargittai, E., Neuman, W.R., and Robinson, J.P. (2001). Social Implications of the Internet. *Annual Review of Sociology, 27*, 307-336.

Dollmann, J., Jacob, K., and Kalter, F. (2014). Examining the Diversity of Youth in Europe: A Classification of Generations and Ethnic Origins Using CILS4EU Data (Technical Report). Retrieved May 6, 2016 (www.mzes.uni-mannheim.de)

Duggan, M., Ellison, N.B., Lampe, C., Lenhard, A., and Madden, M. (2015). Social Media Update 2014. Pew Research Center. Retrieved July 5, 2016 (www.pewinternet.org).

Dunbar, R.I.M. (1993). Co-Evolution of Neocortex Size, Group Size and Language in Humans. *Behavioral and Brain Sciences, 16*, 681-735.

Dunbar, R.I.M. (1998). The Social Brain Hypothesis. *Evolutionary Anthropology, 6*(5), 178-190.

Dunbar, R.I.M., Arnaboldi, V., Conti, M., and Passarella, A. (2015). The structure of Online Social Networks Mirrors Those in the Offline World. *Social Networks, 43*, 39-47.

Dunbar, R.I.M. (2016). Do Online Social Media Cut Through the Constraints that Limit the Size of Offline Social Networks. *Royal Society Open Science, 3*, 150292.

Ellison, N.B., Steinfield, C., and Lampe, C. (2007). The Benefits of Facebook "Friends": Social Capital and College Students' Use of Online Social Network Sites. *Journal of Computer-Mediated Communication, 12*(4), 1143-1168.

Ellison, N.B., Steinfield, C., and Lampe, C. (2011). Connection Strategies: Social Capital Implications of Facebook-enabled Communication Practices. *New Media & Society, 13*(6), 873-982.

Ellison, N.B. and boyd, d.m. (2013). Sociality Through Social Network Sites. In Dutton, W.H. (Ed.), *Oxford Handbook of Internet Studies* (pp. 150-172). Oxford: Oxford University Press.

Ellison, N.B., Vitak, J., Gray, R., and Lampe, C. (2014). Cultivating Social Resources on Social Network Sites: Facebook Relationship Maintenance Behaviors and Their Role in Social Capital Processes. *Journal of Computer-Mediated Communication, 19*(4), 855-870.

Facebook. (2015). Facebook Company Info. Retreived July 29, 2015 (newsroom.fb.com)

Facebook. (2017). Facebook Newsroom. Retrieved 13 March, 2017 (https://newsroom.fb.com)

Feld, S.L. (1981). The Focused Organization of Social Ties. *American Journal of Sociology, 86*(5), 1015-1035.

Feld, S.L. (1982). Social Structural Determinants of Similarity Among Associates. *American Sociological Review, 47*(6), 797-801.

Feld, S.L. (1984). The Structured Use of Personal Associates. *Social Forces, 62*(2), 640-652.

Feld, S.L. (1997). Structural Embeddedness and Stability of Interpersonal Relationships. *Social Networks, 19*, 94-95.

Feld, S.L. and Carter, W.C. (1999). Foci of Activity as Changing Contexts for Friendship. In Rebecca, G.A. and Allan, G. (Eds.), *Placing Friendships in Context* (pp. 136-152). Cambridge: Cambridge University Press.

Feliciano, C., Robnett, B., and Komaie, G. (2009). Gendered Racial Exclusion Among White Internet Daters. *Social Science Research, 38*(1), 39-54.

Fiscella, K. and Fremont, A.M. (2006). Use of Geocoding and Surname Analysis to Estimate Race and Ethnicity. *Health Services Research, 41*(4), 1482-500.

Fischer, C.S. (2009). The 2004 GSS Finding of Shrunken Social Networks: An Artifact? *American Sociological Review, 74*(4), 657–69.

Fogel, J. and Nehman, E. (2009). Internet Social Network Communities: Risk Taking, Trust, and Privacy Concerns. *Computers in Human Behavior, 25*(1), 153-160.

Freeman, L. and Thompson, C.L. (1989). Estimating Acquantanceship Volume. In M. Kochen (Ed.), *The Small World* (pp. 147-158). Norwood, NJ: Ablex Publishing Corporation.

Fukuyama, F. (1995). *Trust.* New York: Free press.

Van der Gaag, M. and Snijders, T.A.B. (2005). The Resource Generator: Social Capital Quantification with Concrete Items. *Social Networks, 27*(1), 1-29.

Gambetta, D. (1988). *Trust: Making and Breaking Cooperative Relations.* New York: Blackwell.

Geven, S., Weesie, J., and Van Tubergen, F. (2013). The Influence of Friends on Adolescents' Behavior Problems at School: The Role of Ego, Alter and Dyadic Characteristics. *Social Networks, 35*, 583-592.

Glaeser, E.L., Liabson, D.I., Scheinkman, J.A., and Soutter, C.L. (2000). Measuring Trust. *Quarterly Journal of Economics, 115*(3), 811-846.

Golder, S.A. and Macy, M.W. (2014). Digital Footprints: Challenges and Opportunities for Online Social Research. *Annual Review of Sociology, 40*, 129-152.

Gonçalves, B., Perra, N. and Vespignani, A. (2011). Modeling Users' Activity on Twitter Networks: Validation of Dunbar's Number. *PLoS ONE, 6*(8), e22656.

Gonzàles-Bailòn, S. and Wang, N. (2016). Networked Discontent: The Anatomy of Protest Campaigns in Social Media. *Social Networks, 44*, 95-104.

Granovetter, M.S. (1973). The Strength of Weak Ties. *American Journal of Sociology, 78*(6), 1360-1380.

Granovetter, M.S. (1983). The Strength of Weak Ties: A Network Theory Revisited. *Sociological Theory, 1*, 201-233.

Greenwood, S., Perrin, A., and Duggan, M. (2016). Social Media Update 2016. Retrieved March 21, 2017 (http://www.pewinternet.org/2016/11)

Grosetti, M. (2005). Where do Social Relations Come From?: A Study of Personal Networks in the Toulouse Area of France. *Social Networks, 27*(4), 289-300.

Halberstam, Y. and Knight, B. (2016). Homophily, Group Size, and the Diffusion of Political Information in Social Networks: Evidence From Twitter. *Journal of Public Economics, 143*, 73-88.

Hampton, K.H., Session, L.F., and Her, E.J. (2011). Core Networks, Social Isolation, and New Media. *Information, Communication & Society, 14*(1), 130-155.

Hargittai, E. (2002). Second Level Digital Divide: Differences in People's Online Skills. *First Monday, 7*(4).

Hargittai, E. (2008). Whose Space? Differences Among Users and Non-Users of Social Network Sites. *Journal of Computer-Mediated Communication, 13*(1), 276-297.

Hargittai, E. and Litt, E. (2011). The Tweet Smell of Celebrity Success: Explaining Variation in Twitter Adoption Among a Diverse Group of Young Adults. *New Media & Society, 13*(5), 824-842.

Heckman, J.J. (1979). Sample Selection Bias as a Specification Error. *Econometrica, 47*(1), 153-161.

Hedström, P. (2005). *Dissecting the Social: On the Principles of Analytical Sociology.* Cambridge: Cambridge University Press.

Heider, F. (1946). Attitudes and Cognitive Organization. *The Journal of Psychology, 21*(1), 107-112.

Hill, R.A. and Dunbar, R.I.M. (2003). Social Network Size in Humans. *Human Nature, 14*(1),53-72.

Hobbs, W.R., Burke, M., Christakis, N.A., and Fowler, J.H. (2016). Online Social Integration is Associated with Reduced Mortality Risk. *Proceedings of the National Academy of Sciences, 113*(46), 12980-12984.

Hofstra, B., Corten, R., and Van Tubergen, F. (2015a). Dutch Facebook Survey: Wave 1 [dataset and codebook]. Found at: easy.dans.knaw.nl.

Hofstra, B., Corten, R., and Buskens, V. (2015b). Learning in Social Networks: Selecting Profitable Choices Among Alternatives of Uncertain Profitability in Various Networks. *Social Networks, 43*, 100-112.

Hofstra, B., Corten, R., and Van Tubergen, F. (2016a). Who was First on Facebook? Determinants of Early Adoption among Adolescents. *New Media & Society, 18*(10), 2340-2358.

Hofstra, B., Corten, R., and Van Tubergen, F. (2016b). Understanding the Privacy Behavior of Adolescents on Facebook: The Role of Peers, Popularity and Trust. *Computers in Human Behavior, 60*, 611-621.

Hofstra, B., Corten, R., Van Tubergen, F., and Ellison, N.B. (2017). Sources of Segregation in Social Networks: A Novel Approach Using Facebook. *American Sociological Review, 82*(3), 625-656.

Holt-Lunstad, J., Smith, T.B., and Layton, J.B. (2010). Social Relationships and Mortality Risk: A Meta-Analytic Review. *PLoS Medicine, 7*(7), e1000316.

Hoy, M.G. and Milne, G. (2010). Gender Differences in Privacy-Related Measures for Young Adult Facebook Users. *Journal of Interactive Advertising, 10*(2), 28-45.

Ibarra, H. (1995). Race, Opportunity, and Diversity of Social Circles in Managerial Networks. *The Academy of Management Journal, 38*(3), 673-703.

Van Ingen, E., Rains, S.A., and Wright, K.B. (2017). Does Social Network Site Use Buffer Against Well-Being Loss When Older Adults Face Reduced Functional Ability. *Computers in Human Behavior, 70*, 167-177.

Inglehart, R. and Abramson, P.R. (1994). Economic Security and Value Change. *The American Political Science Review, 88*(2), 336-354.

Jackson, M.O. (2008). *Social and Economic Networks.* Oxfordshire: Princeton University Press.

Jaspers, E. and Van Tubergen, F. (2014). Children of Immigrants Longitudinal Survey in the Netherlands (CILSNL) - Wave 4. Full version v4.0.0. DANS.

Jaspers, E. and Van Tubergen, F. *Thematic Collection: Children of Immigrants Longitudinal Survey in the Netherlands (CILSNL)*. DANS. 2017. Found at: easy.dans.knaw.nl.

Javaro, J.N. and Jasinski, J.L. (2014). Identity Theft and Social Networks. In Marcum, C.D. and Higgens, G.E. (Eds.), *Social networking as a criminal enterprise* (pp. 69-86). Florida: CRC Press.

Kadushin, C. (2012). *Understanding Social Networks: Theories, Concepts, and Findings.* Oxford: Oxford University Press.

Kalmijn, M. (1998). Intermarriage and Homogamy: Causes, Patterns, Trends. *Annual Review of Sociology, 24*, 395-421.

Kalmijn, M. and Flap, H. (2001). Assortative Meeting and Mating: Unintended Consequences of Organized Settings for Partner Choices. *Social Forces, 79*(4), 1289-1312.

Kalmijn, M. (2003). Shared Friendship Networks and the Life Course: An Analysis of Survey Data on Married and Cohabiting Couples. *Social Networks, 25*(3), 231-249.

Kalmijn, M. and Van Tubergen, F. (2006). Ethnic Intermarriage in the Netherlands: Confirmations and Refutations of Accepted Insights. *European Journal of Population, 22*(4), 371-397.

Kalmijn, M. (2012). Longitudinal Analyses of the Effects of Age, Marriage, and Parenthood on Social Contacts and Suppert. *Advances in Life Course Research, 17*(4), 177-190.

Kalter, F., Heath, A.F., Hewstone, M., Jonsson, J.O., Kalmijn, M., Kogan, I. and Van Tubergen, F. (2013). Children of Immigrants Longitudinal Survey in Four European Countries (CILS4EU). GESIS Data Archive, Cologne, ZA5353, Data file Version 1.0.0.

Kalter, F., Heath, A.F., Hewstone, M., Jonsson, J.O., Kalmijn, M., Kogan, I., and Van Tubergen, F. (2015). Children of Immigrants Longitudinal Survey in Four European Countries (CILS4EU). GESIS Data Archive, Cologne, ZA5353, Data file version 2.1.0.

Kalter, F., Heath, A., Hewstone, M., Jonsson, J.O., Kalmijn, M., Kogan, I., and Van Tubergen, F. (2016). *Children of Immigrants Longitudinal Survey in Four European Countries (CILS4EU) — Full version.* ZA5353, GESIS Data Archive, Cologne, Data file Version 3.1.0.

Kanai, R., Bahrami, B., Roylance, R., and Rees, G. (2011). Online Social Network Size is Reflected in Human Brain Structure. *Proceedings of the Royal Society B*,1-8.

Killworth, P.D., Johnsen, E.C., McCarty, C., Shelley, G.A. and Bernard, R.H. (1998). A Social Network Approach to Estimating Seroprevalence in the United States. *Social Networks, 20*, 23-50.

Kossinets, G. and Watts, D.J. (2009). Origins of Homophily in an Evolving Social Network. *American Journal of Sociology, 115*(2), 405-450.

Krackhardt, D. and Handcock, M.S. (2007). Heider vs Simmel: Emergent Features in Dynamic Structures. In Airoldi, E., Blei, D.M., Fienberg, S.E., Goldenberg, A., Xing, E.P., and Zheng, A.X. (Eds.), *Statistical Network Analysis: Models, Issues and New Directions. Vol 4503, Lecture Notes in Computer Science* (pp. 14-27). Berlin: Springer Berlin Heidelberg.

Kramer, A.D.I., Guillory, J.E., and Hancock, J.T. (2014). Experimental Evidence of Massive-scale Emotional Contagion Through Social Networks. *Proceedings of the National Academy of Sciences, 111*(24), 8788-8790.

Kuru, O. and Pasek, J. (2016). Improving Social Media Measurement in Surveys: Avoiding Acquiescence Bias in Facebook Research. *Computers in Human Behavior, 57*, 82-92.

Lancee, B. and Sarrasin, O. (2015). Educated Preferences or Selection Effects? A Longitudinal Analysis of the Impact of Educational Attainment on Attitudes Towards Immigrants. *European Sociological Review, 31*(4), 490-501.

Lauderdale, D.S. and Kestenbaum, B. (2000). Asian American Ethnic Identification by Surname. *Population Research and Policy Review, 19*, 283–300.

Lazarsfeld, P.F. and Merton, R.K. (1954). Friendships as a Social Process: A Substantive and Methodological Analysis. In Berg, M (Ed.), *Freedom and Control in Modern Society* (pp. 18-66). New York: Van Nostrand.

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A., Brewer, D., Christakis, N., Contracter, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M.W., Roy, D., and Van Alstyne, M. (2009). Computational Social Science. *Science, 323*(5915), 721-723.

Leszczensky, L. and Pink, S. (2015). Ethnic Segregation of Friendship Networks in School: Testing a Rational-Choice Argument of Differences in Ethnic Homophily Between Classroom- and Grade-Level Networks. *Social Networks, 42*, 18-26.

Lewis, K., Kaufman, J., and Christakis, N. (2008a). The Taste for Privacy: An Analysis of College Student Privacy Settings in an Online Social Network. *Journal of Computer-Mediated Communication, 14*(1), 79-100.

Lewis, K., Kaufman, J., Gonzalez. M., Wimmer, A., and Christakis, N. (2008b). Tastes, Ties, and Time: A New Social Network Dataset Using Facebook.com. *Social Networks, 30*(4), 330-342.

Lewis, K. (2011). The Co-evolution of Social Network Ties and Online Privacy Behavior. In Trepte, S. and Reinecke, L. (Eds.), *Privacy Online: Perspectives on Privacy and Self-Disclosure in the Social Web* (pp. 91-10). Berlin, Heidelberg: Springer.

Lewis, K, Gonzalez, M., and Kaufman, J. (2012). Social Selection and Peer Influence in an Online Social Network. *Proceedings of the National Academy of Sciences 109*(1), 68-72.

Lewis, Kevin. (2013). The Limits of Racial Prejudice. *Proceedings of the National Academy of Sciences, 110*, 18814-18819.

Lewis, K. (2015a). Three Fallacies of Digital Footprints. *Big Data & Society*, 1-4.

Lewis, K. (2015b). How Networks Form: Homophily, Opportunities, and Balance. In Scott, R.A. and Kosslyn, S.M. (Eds), *Emerging Trends in the Social and Behavioral Sciences: An Interdisciplinary, Searchable, and Linkable Resource* (pp. 1-14). New York: John Wiley & Sons.

Lichter, D.T., Parisi, D., and Taquino, M.C. (2015). Toward a New Macro-Segregation? Decomposing Segregation Within and Between Metropolitan Cities and Suburbs. *American Sociological Review, 80*(4), 843-873.

Lieberson, S. (2000). *A Matter of Taste: How Names, Fashions, and Culture Change.* New Haven & London: Yale University Press.

Liebowitz, S.J. and Margolis, E. (1994). Network Externality: An Uncommon Tragedy. *Journal of Economic Perspective, 8*(2), 133-150.

Lin, N. (1999). Social Capital and Status Attainment. *Annual Review of Sociology, 25*, 467-487.

Litt, E. (2013). Understanding Social Network Site Users' Privacy Tool Use. *Computers in Human Behavior, 29*(4), 1649-1656.

Lizardo, O. (2006). How Cultural Tastes Shape Personal Networks. *American Sociological Review, 71*(5), 778-807.

Lizardo, O. and Skiles, S. (2012). Reconceptualizing and Theorizing "Omnivorousness": Genetic and Relational Mechanisms. Sociological Theory, 30(4), 263-282.

Long, J.S. and Freeze, J. (2001). *Regression Models for Categorical Dependent Variables Using Stata.* Stata Corporation: College Station.

Lubbers, M.J. (2003). Group Composition and Network Structure in School Classes: A Multilevel Application of the P* Model. *Social Networks, 25*(4), 309-332.

Madden., Lenhart, A., Cortesi, S., Gasser, U., Duggan, M., Smith, A., and Beaton, M. (2013). Teens, Social Media, and Privacy. Retrieved March 22, 2017 (http://www.pewinternet.org)

Marin, A. (2004). Are Respondents More Likely to List Alters with Certain Characteristics?: Implications for Name Generator Data. *Social Networks, 26*(4), 289-307.

Marsden, P.V. (1987). Core Discussion Networks of Americans. *American Sociological Review, 52*(1), 122-131.

Marsden, P.V. (1988). Homogeneity in Confiding Relations. *Social Networks, 10*(1), 57-76.

Marsden, P.V. and Friedkin, N.E. (1993). Network Studies of Social Influence. *Sociological Methods & Research, 22*(1), 127-151.

Marsden, P.V. (2011). Survey Methods for Network Data. In Scott, J. and Carrington, P.J. (Eds.), *The Sage Handbook of Social Network Analysis* (pp. 370-388). London: Sage Publications.

Mäs, M. and Flache, A. (2013). Differentation Without Distancing. Explaining Bi-Polarization of Opinions Without Negative Influence. *PloS One, 8*(11), e74516.

Mateos, P., Longley, P.A., and O'Sullivan, D. (2011). Ethnicity and Population Structure in Personal Naming Networks. *PLoS One, 6*(9), e22943.

Mayer, A. and Puller, S.E. (2008). The Old Boy (and Girl) Network: Social Network Formation on University Campuses. *Journal of Public Economics, 92*(1-2), 329-347.

McCarty, C., Killworth, P.D, Bernard, H.R., Johnsen, E.C., and Shelley, G.E. (2001). Comparing Two Methods for Estimating Network Size. *Human organization, 60*(1), 28-39.

McCormick, T.H., Salganik, M.J., and Zheng, T. (2010). How Many People Do You Know? Efficiently Estimating Personal Network Size. *Journal of the American Statistical Association, 105*(489), 59-70.

McFarland, D.A., Lewis, K., and Goldberg, A. (2015). Sociology in the Era of Big Data: The Ascent of Forensic Social Science. *The American Sociologist, 47*(1), 12-35.

McFarland, D.A. and McFarland, H.R. (2016). Big Data and the Danger of Being Precisely Inaccurate. *Big Data & Society*, 1-4.

McPherson, M., Smith-Lovin, L. (1987). Homophily in Voluntary Organizations: Status Distance and the Composition of Face-to-Face Groups. *American Sociological Review, 52*(3), 370-379.

McPherson, M., Smith-Lovin, L. and Cook, J.M. (2001). Birds of A Feather: Homophily in Social Networks. *Annual Review of Sociology, 27*(1), 415-444.

McPherson, M., Smith-Lovin, L., and Brashears, M.E. (2006). Social Isolation in America: Changes in Core Discussion Networks over Two Decades. *American Sociological Review, 71*(3), 353-375.

McPherson, M., Smith-Lovin, L., and Brashears, M.E. (2008). ERRATA: Social Isolation in America: Changes in Core Discussion Networks over Two Decades. *American Sociological Review, 73*(6), 1022.

McPherson, M., Smith-Lovin, L., and Brashears, M.E. (2009). Models and Marginals: Using Survey Evidence to Study Social Networks. *American Sociological Review, 74*(4), 670-681.

Van der Meer, T. and Tolsma, J. (2014). Ethnic Diversity and Its Effects on Social Cohesion. *Annual Review of Sociology, 40*, 459-478.

Meertens Institute. (2016). De Nederlandse Voornamenbank. Retrieved September 6, 2016. (http://www.meertens.knaw.nl/nvb/)

Mewes, J. (2014). Gen(d)eralized Trust: Women, Work, and Trust in Strangers. *European Sociological Review, 30*(3), 373-386.

Mollenhorst, G., Volker, B., and Flap, H. (2008). Social Contexts and Personal Relationships: The Effect of Meeting Opportunities on Similarity for Relationships of Different Strength. *Social Networks, 30*(1), 60-68.

Mollenhorst, G., Volker, B., and Flap, H. (2011). Shared Contexts and Triadic Closure in Core Discussion Networks. *Social Networks, 33*(4), 292-302.

Mollenhorst, G., Volker, B., and Flap, H. (2014). Changes in Personal Relationships: How Social Contexts Affect the Emergence and Discontinuation of Relationships. *Social Networks, 37*, 65-80.

Mood, C. (2010). Logistic Regression: Why We Cannot Do What We Think We Can Do,and What We Can Do About It. *European Sociological Review, 26*(1), 67-82.

Moody, J. (2001). Race, School Integration, and Friendship Segregation in America. *American Journal of Sociology, 107*(3), 679–716.

Moore, G. (1990). Structural Determinants of Men's and Women's Personal Networks. *American Sociological Review, 55*(5), 726-735.

Moule, R.K., Pyrooz, D.C., and Decker, S.H. (2013). From 'What the F # @ % is a Facebook?' to 'Who Doesn't Use Facebook?': The Role of Criminal Lifestyles in the Adoption and the Use of the Internet. *Social Science Research, 42*(6), 1411-1421.

Mouw, T. and Entwisle, B. (2006). Residential Segregation and Interracial Friendship in Schools. *American Journal of Sociology, 112*(1), 394-441.

Nannestad, P. (2008). What Have We Learned From Generalized Trust, If Anything? *Annual Review of Political Science, 11*, 413-436.

Office for National Statistics. (2016). Internet Access - Households and Individuals: 2016. Retrieved 20 April, 2017 (https://www.ons.gov.uk)

Orchard, L.J., Fullwood, C., Galbraith, N., and Morris, N. (2014). Individual Differences as Predictors of Social Networking. *Journal of Computer-Mediated Communication, 19*(3), 288-402.

Paik, A. and Sanchagrin, K. (2013). Social Isolation: An Artifact. *American Sociological Review, 78*(3), 339-360.

Parigi, P. and Henson II, W. (2014). Social Isolation in American. *Annual Review of Sociology, 40*, 153-171.

Paxton, P. (2007). Association Memberships and Generalized Trust: A Multilevel Model Across 31 Countries. *Social Forces, 86*(1), 47-76.

Peterson, R.A. (1992). Understanding Audience Segmentation: From Elite and Mass to Omnivore and Univore. *Poetics, 21*(4). 243-258.

Pettigrew, T.F. and Tropp, L.R. (2006). A Meta-Analytic Test of Integroup Contact Theory. *Journal of Personality Social Psychology, 90*(5), 751-783.

Plummer, M. (2003). JAGS: A Program of Bayesian Graphical Models Using Gibbs Sampling. Pollet, T.V., Roberts, S.G.B., Dunbar, R.I.M. (2011). Use of Social Network Sites and Instant Messaging Does Not Lead to Increased Offline Social Network Size, or to Emotionally Closer Relationships with Offline Network Members. *Cyberpsychology, Behavior, and Social Networking, 14*(4), 253-258, Potârcă, G. and Mills, M. (2015). Racial Preferences in Online Dating across European Countries. *European Sociological Review, 31*(3), 326-341.

Pool, I. deS. and Kochen, M. (1978). Contacts and Influence. *Social Networks, 1*, 5-51.

Putnam, R.D. (2000). *Bowling Alone: The Collapse and Revival of American Community.* New York: Simon & Schuster.

Rainie, L. and Wellman, B. (2012). *Networked: The New Social Operating System.* Cambridge: The MIT Press.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from: www.R-project.org/.

Reich, S.M., Subrahmanyam, K, and Espinoza, G. (2008). Friending, IMing, and Hanging Out Face-to-Face: Overlap in Adolescents' Online and Offline Social Networks. *Developmental Psychology, 48*(2), 356-368.

Roberts, S.G.B., Dunbar, R.I.M, Pollet, T.V., and Kuppens, T. (2009). Exploring Variation in Active Network Size: Constraints and Ego Characteristics. *Social Networks, 31*, 138-146.

Robinson, L., Cotton, S.R., Ono, H., Quan-Haase, A., Mesch, G., Chen, R., Schulz, J., Hale, T.M., and Stern. M. (2015). Digital Inequalities and Why They Matter. *Information, Communication & Society, 18*(5), 569-582.

Rogers, E.M. (2003). *Diffusion of Innovations.* New York: Free Press.

Roth, P.L., Bobko, P., Van Iddekinge, C.H., and Thatcher, J.B. (2016). Social Media in Employee-Selection-Related Decisions: A Research Agenda for Uncharted Territory. *Journal of Management, 42*(1), 269-298.

Rözer, J.J., Mollenhorst, G., and Volker, B. (2015). Romantic Relationship Formation, Maintenance and Changes in Personal Networks. *Advances in Life Course Research, 23*, 86-97.

Ruben, D.B. (1976). Inference and Missing Data. *Biometrika, 63*(3), 581-592.

Schimmer, P. and Van Tubergen, F. (2014). Transnationalism and Ethnic Identification among Adolescent Children of Immigrants in the Netherlands, Germany, England, and Sweden. *International Migration Review, 48*(3), 680-709

Schafer, J.L. and Graham, J.W. (2002). Missing Data: Our View of the State of the Art. *Psychological Methods, 7*(2), 147-177.

Semyonov, M. and Glikman, A. (2009). Ethnic Residential Segregation, Social Contacts, and Anti-Minority Attitudes in European Societies. *European Sociological Review, 25*(6), 693-708.

Shin, W. and Kang, H. (2016). Adolescents' Privacy Concerns and Information Disclosure Online: The Role of Parents and the Internet. *Computers in Human Behavior, 54*, 114-123.

Shrum, W., Cheek, N.H., and. Hunter, S.MacD. (1988). Friendship in School: Gender and Racial Homophily. *Sociology of Education, 61*(4), 227-239.

Shye, D., Mullooly, J.P., Freeborn, D.K., and Pope, C.R. (1995). Gender Difference in the Relationship Between Social Network Support and Mortality: A Longitudinal Study of an Elderly Cohort. *Social Science and Medicine, 41*(7), 935-947.

Simpson, B. and Borch, C. (2005). Does Power Affect Perception in Social Networks? Two Arguments and an Experimental Test. *Social Psychology Quarterly, 68*(3), 278-287.

Simpson, B., McGrimmon, T., and Irwin, K. (2007). Are Blacks Really Less Trusting than Whites? Revisiting the Race and Trust Question. *Social Forces, 86*(2), 525-552.

Simpson, B., Markovsky, B., and Steketee, M. (2011). Power and the Perception of Social Networks. *Social Networks, 33*, 166-171.

Slater, P.E. (1963). On social regression. *American Sociological Review, 28*(3), 339-364.

Small, M.L. and Sukhu, C. (2016). Because They Were There: Access, Deliberation, and the Mobilization of Networks for Support. *Social Networks, 47*, 73-84.

Smith, A. and Schneider, B.H. (2000). The Inter-ethnic Friendships of Adolescent Students: A Canadian Study. *International Journal of Intercultural Relations, 24*(2), 247-258.

Smith, J.A., McPherson, M., and Smith-Lovin, L. (2014a). Social Distance in the United States: Sex, Race, Religion, Age, and Education Homophily among Confidants, 1985 to 2004. *American Sociological Review, 79*(3), 432-456.

Smith, S., Maas, I., and Van Tubergen, F. (2012). Irreconcilable Differences? Ethnic Intermarriage and Divorce in the Netherlands,1995-2008. *Social Science Research, 41*(5), 1126-1137.

Smith, S., Maas, I., and Van Tubergen, F. (2014b). Ethnic Ingroup Friendships in Schools: Testing the By-Product Hypothesis in England, Germany, the Netherlands and Sweden. *Social Networks, 39*, 33-45.

Snijders, T.A.B. and Bosker, R.J. (2012). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling.* London: Sage Publishers.

Son, J. and Lin, N. (2012). Network Diversity, Contact Diversity, and Status Attainment. *Social Networks, 34*(4), 601-613.

Song, L. (2012). Raising Network Resources While Raising Children? Access to Social Capital by Parenthood Status, Gender, and Marital Status. *Social Networks, 34*(2), 241-252.

Spiro, E.S. (2016). Research Opportunities at the Intersection of Social Media and Survey Data. *Current Opinion in Psychology, 9*, 67-71.

Stark, T.H. and Flache, A. (2012). The Double Edge of Common Interest. *Sociology of Education, 85*(2), 179-199.

StataCorp. (2013). Stata Statistical Software: Release 13. College Station. Texas: StataCorp LP.

Statistics Germany. (2016). Liste der Häufigen Vornamen 2015. Retrieved August 16, 2016 (www.govdata.de)

Statistics Netherlands. (2012). Migranten, Vreemdelingen en Vluchtelingen: Begrippen op het Terrein van Asiel en Buitenlandse Migratie. Retrieved June 16, 2015 (www.cbs.nl)

Statistics Netherlands. (2015). CBS Statline: Bevolking; kerncijfers. Retrieved March 23, 2015 (statline.cbs.nl)

Statistics Netherlands. (2017). CBS Statline: Internet: Toegang, Gebruik en Faciliteiten. Retrieved April 1, 2017. (statline.cbs.nl)

Steglich, C.E.G., Snijders, T.A.B., and Pearson, M. (2010). Dynamic Networks and Behavior: Separating Selection from Influence. *Sociological Methodology, 40*(1), 329-392.

Steinfield, C., Ellison, N.B., and Lampe, C. (2008). Social Capital, Self-Esteem, and Use of Online Social Network Sites: A Longitudinal Analysis. *Journal of Applied Developmental Psychology, 29*(6), 434-445.

Stopczynski, A., Sekara, V., Sapiezunski, P., Cuttone, A., Madsen, M.M., Larsen, J.E., and Lehmann, S. (2014). Measuring Large-Scale Social Networks with High Resolution. *PLoS ONE, 9*(4), e95978.

Stutzman, F. and Kramer-Duffield, J. (2010). Friends Only: Examining a Privacy-Enhancing Behavior in Facebook. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '10)*, 1553-1562.

Stutzman, F., Gross, R., and Acquisti, A. (2013). Silent Listeners: The Evolution of Privacy and Disclosure on Facebook. *Journal of Privacy and Confidentiality, 4*(2), 7-41.

Subrahmanyam, K., Reich, S.M., Waechter, N., and Espinoza, G. (2008). Online and Offline Social Networks: Use of Social Networking Sites by Emerging Adults. *Journal of Applied Developmental Psychology, 29*(6), 420-433.

Sullivan, O. and Katz-Gerro, T. (2007). The Omnivore Thesis Revisited: Voracious Cultural Consumers. *European Sociological Review, 23*(2), 123-137.

Thelwall, M. (2008). Social Networks, Gender, and Friending: An Analysis of MySpace Member Profiles. *Journal of the American Society for Information Science and Technology, 59*(8), 1321-133.

Thomson, R., Yuki, M., and Ito, N. (2015). A Socio-Ecological Approach to National Differences in Online Privacy Concern: The Role of Relational Mobility and Trust. *Computers in Human Behavior, 51*, 285-292.

Tu, S. (2017). The Dirichlet-Multinomial and Dirichlet-Categorical Models for Bayesian Inference. *Working paper*. Retrieved February 24, 2017 (people.eecs.berkeley.edu/∼stephentu)

Van Tubergen, F. (2014). Size and Socio-Economic Resources of Core Discussion Networks in the Netherlands: Differences by National-Origin Group and Immigrant Generation. *Ethnic and Racial Studies, 37*(6), 1020-2042.

Van Tubergen, F. (2015). Ethnic Boundaries in Core Discussion Networks: A Multilevel Social Network Study of Turks and Moroccans in the Netherlands. *Journal of Ethnic and Migration Studies, 41*(1), 101-116.

Tufekci, Z. (2008). Can You See Me Now? Audience and Disclosure Regulation in Online Social Network Sites. *Bulletin of Science Technology & Society, 28*(20), 20-36.

Tufekci, Z. (2014). Big Questions for Social Media Big Data: Representativeness, Validity and Other Methodological Pitfalls. *Proceedings of the 8th International AAAI Conference on Weblogs and Social Media*, 505-514.

Twenge, J.M., Konrath, S., Foster, J.D., Campbell, W.K., and Bushman, B.J. (2008). Egos Inflating Over Time: A Cross-Temporal Meta-Analysis of the Narcissistic Personality Inventory. *Journal of Personality, 76*(4), 875-902.

United States Census Bureau. (2014). Frequently Occurring Surnames from the Census 2000. Retrieved August 16, 2016 (www.census.gov)

Utz, S., Tanis, M., and Vermeulen, I. (2012). It is All About Being Popular: The Effects of Need for Popularity on Social Network Site Use. *Cyberpsychology, Behavior, and Social Networking, 15*(1), 37-42.

Utz, S. and Krämer, N.C. (2015). The Privacy Paradox on Social Network Sites Revisited: The Role of Individual Characteristics and Group Norms. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace, 3*(2).

Valente, T.W. (2010). *Social Networks and Health: Models, Methods, and Applications.* Oxford: Oxford University Press.

Van der Veer, N., Boekee, S., and Peters, O. (2017). Nationale Social Media Onderzoek 2017. *Newcom Research & Consultancy.* Retrieved April 2, 2017 (www.newcom.nl/socialmedia2017)

Vermeij, L., Van Duijn, M.A.J., and Baerveldt, C. (2009). Ethnic Segregation in Context: Social Discrimination Among Native Dutch Pupils and Their Ethnic Minority Classmates. *Social Networks, 31*(4), 230-239.

Verkuyten, M. and Kwa, G.A. (1994). Ethnic Self-Identification and Psychological Well-Being among Minority Youth in the Netherlands. *International Journal of Adolescence and Youth, 5*(1-2), 19-34.

Visintin, E.P., Voci, A., Pagotto, L., and Hewstone, M. (2016). Direct, Extended, and Mass-mediated Contact With Immigrants in Italy: Their Associations With Emotions, Prejudice, and Humanity Perceptions. *Journal of Applied Social Psychology, 47*(4), 175-194.

Watts, D.J. (2011). *Everything is Obvious: How Common Sense Fails Us.* New York: Crown Business.

Wellman, B. and Wortley, S. (1990). Different Strokes from Different Folks: Community Ties and Social Support. *American Journal of Sociology, 96*(3), 558-588.

Van de Werfhorst, H. and Van Tubergen, F. (2007). Ethnicity, Schooling, and Merit in the Netherlands. *Ethnicities, 7*(3), 416-444.

Wimmer, A. and Lewis, K. (2010). Beyond and Below Racial Homophily: ERG Models of a Friendship Network Documented on Facebook. *American Journal of Sociology, 116*(2), 583-642.

Windzio, M. and Bicer, E. (2013). Are We Just Friends? Immigrant Integration Into High- and Low-Cost Social Networks. *Rationality and Society, 25*(2), 123-145.

Wu, S., Chou, M., Tseng, C., Lee, Y., and Chen, K. (2014). Detecting In-Situ Identity Fraud on Social Network Services: A Case Study on Facebook. *Proceedings of the 23rd International World Wide Web Conference*, 401-402.

Van Zalk, M.H.W., Van Zalk, N., Kerr, M., and Stattin, H. (2014). Influences Between Online-Exclusive, Conjoint and Offline-Exclusive Friendship Networks: The Moderating Role of Shyness. *European Journal of Personality, 28*, 134-146.

Zheng, T., Salganik, M.J., and Gelman, A. (2006). How Many People Do You Know in Prison? *Journal of the American Statistical Association, 101*(474), 409-423.

Zhou, W.X., Sornette, D., Hill, R.A., and Dunbar, R.I.M. (2005). Discrete Hierarchical Organization of Social Group Sizes. *Proceedings of the Royal Society B, 272*, 439-444.

Acknowledgements

Acknowledgments

Conducting the research in this dissertation was an intellectual journey that shaped me for the rest of my career. Only at some point in the far future will I be able to look back and fully comprehend how much I learned. There are many people who steered, pushed, or talked me back onto the right track throughout this journey if I tended to stumble, go astray, or fail.

First, and foremost, I am forever indebted to my supervisors, Rense Corten and Frank van Tubergen. Thank you for providing the opportunity to set out on this journey. Rense, your recurring question “but what do these results mean in terms of the theory?” will probably hunt me for the rest of my career. This is not a bad thing, however, as your remarks about theoretical arguments that did not hold up to scrutiny have, without any doubt, helped me to become a much better scientist. Frank, your insights into how to write the most relevant paper and your curious-meets-pragmatic feedback helped me mature as a sociologist and scientist. You must also be a record-holder for fastest feedback ever given on a paper: 34 minutes. This is a true story, and in case one is wondering: it was super useful. The combination of the expertise you both have made that I often felt lucky to have you as supervisors. I highly value the way in which we worked together and like to believe we gave the field of online social networks a small, sociological push forward. You are the reason that I felt immensely intellectually enabled during my time as a Ph.D.-candidate. If there is unfulfilled potential in this dissertation it is solely due to my own shortcomings and not due to the guidance that I had.

There are two colleagues without whom I probably wouldn’t have made it all the way through: Manja and Maaike. Manja, your bright mood, numerous times of reading my work, and talent for putting things into perspective is priceless. Maaike, your personal involvement, awesome sense of humor, and the way in which you supported me during difficult times is an inspiration. I am confident that we’ll remain friends, independent of where our careers take us.

Two other people who influenced my scientific thinking in a positive way are Vincent Buskens and Jeroen Weesie. Vincent, your MA-thesis supervision helped me to build trust in the relevance of sociological explanations. Jeroen, you often helped me out with questions about methodology and pushed me to contemplate on the underlying substantive questions and theories. I am glad that you are co-author on the fifth chapter in this book.

There are many other colleagues who made doing a Ph.D. such a good experience. Jesper, I always enjoyed (and will keep enjoying) our diners, city walks, and philosophical discussions. Nikki, our trip to a conference in California is among the best experiences I had during my Ph.D. Müge, I really enjoyed working with you on the CILS4EU and CILSNL projects and having you as a colleague, thanks for your brightness. Yassine, our coffee breaks and discussions about careers in science were always a very welcome distraction. Sara, Sanne, and Pascale, I was lucky to work with you and to follow-up on your work on the CILS4EU and CILSNL projects. I learned a lot from your rigor and cheerfulness during extremely busy periods and enjoyed your company at conferences. Vincenz, you were only a short-term roommate, but you still managed to provide some grounded advice on science and life in general, thank you for that. Antonie, having you as a co-teacher for MoSI was a great experience for a starting (and oftentimes stressed-out) teacher. I learned a lot from your eye for details. Lukas, I also greatly enjoyed being a co-teacher with you for "Internet, Social Media, and Networks," hopefully you and Rense will keep on improving this neat (and necessary) course over the next few years. Danique, even though I was pretty busy during my last Ph.D.-year I enjoyed being your roommate.

The larger ICS community was extremely helpful. I am indebted to the many participants in the "Migration and Social Stratification" seminars for their invaluable comments on my (oftentimes prematurely submitted) work. I also want to thank Robert Krause, Nynke Niezink, Jacob Dijkstra, Jan Kornelis Dijksra, and René Veenstra for helping me improve my work as part of the ICS forum days. Furthermore, the support staff in Utrecht was always there to provide practical help. Thank you Ellen, Bärbel, Marjet, Pim, and Babs. I value(d) the advice of and talks with our former and current Ph.D.-mentors, Mariëlle and Els. Finally, I want to express gratitude to my reading committee, Mariette de Haan, Delia Baldassarri, Justus Uitermark, Arnout van de Rijt, and Rens Vliegenthart, for taking the time and energy to assess my dissertation.

During my Ph.D., Nicole Ellison hosted me at the University of Michigan School of Information for a couple of months. Nicole, thank you for hosting me. Your feedback on our paper made it exponentially more relevant. Joseph, thanks for showing me around in Ann Arbor. It is intriguing how our research interests overlap, even though we come from different disciplines. It is only fitting that we started to work on a project together.

The words that I spend here are insufficient to describe how much I owe to my friends and family. Tycho, your lectures on the value of teaching always pushed me to step-up my game. More importantly, your advice on careers and life in general is among the best I've had. Niek and Wouter, we once set out to study sociology and later wrote our BA-thesis together. I find it incredibly cool that you are my paranimphs almost a decade later. Niek, your humor was (and is) an inspiration to look at things more lightly. It's pretty cool that we wrote the sixth chapter of this book together. Wouter, your honesty and quick wit made me question my arguments more than once, but in a very positive, constructive way. Job, Jan, Tito, Ouren, Michel, Ouren, Raphael, and Arjun, it goes without saying that most of these things apply to you as well. Countless nights out, gaming sessions, holidays, festival visits, BBQs, and discussions on science, popular culture, and life helped me recharge when necessary.

Rianne and Hans, thank you for always asking whether I have fun in what I do. Tim, your reminders to do fun stuff was key to actually do so. Joke, your interest in what I do and company is always a joy. Jorn and Dineke, I'm proud to call you my brother and sister(-in-law). Thank you for your unwavering support and enthusiasm. It is an inspiration to see you raise Alesha into a beautiful person. Papa en mama, you are, without a doubt, the two people who invested the most in me and without whom this book would not exist. You support me in my ambition, keep me grounded in reality, and constantly remind me what the true value of hard work is. Therefore, there are only two people to whom I can dedicate this book: my parents.

Finally, my partner in crime and life. Merel, it must not always be easy to live with someone who can get so invested in the abstract world of science and ideas. However, your dedication to make me the best version of myself and your unconditional love feels limitless. You also invaluably influenced the content of this book, as your bright mind never stopped challenging me to reflect and act upon the societal value of my work. I have never met someone who is so true to herself and I can only wish to ever be such a good person as you are. Thank you for enabling me. I cannot wait to start our next journey.

Curriculum Vitae

Curriculum Vitae

Bas Hofstra was born in Emmeloord in the Noordoostpolder, the Netherlands, on October 22nd, 1987. In 2011 he obtained his Bachelor's degree in Sociology from Utrecht University. His Bachelor thesis was nominated for a thesis prize. In 2013, he completed his Research Master's degree in Sociology at Utrecht University. He graduated *cum laude*, won the best student award based on his GPA, and won the Peter G. Swanborn Research Master Thesis Award. In that same year, he started his Ph.D. at the Interuniversity Center for Social Science Theory and Methodology (ICS) at the Department of Sociology at Utrecht University under the supervision of Rense Corten and Frank van Tubergen. As part of his Ph.D. he visited the University of Michigan School of Information for two months to collaborate with Nicole Ellison. During his Ph.D., Hofstra taught Models for the Analysis of Social Interaction (Bachelor-level), Internet, Social Media, and Networks (Master-level), and supervised Bachelor- and Master-theses. Moreover, he helped collecting the fourth through sixth wave of the CILSNL project and coordinated data collection from Facebook and the Dutch social network site Hyves. His work, among others, has appeared in *Social Networks*, *New Media & Society*, and *American Sociological Review*. As of August 2017, he works as a Postdoctoral Fellow at Stanford University in the USA.

Publications and working papers by the author

Publications and working papers by the author

Hofstra, B., Corten, R., and Buskens, V. (2015). Learning in Social Networks: Selecting Profitable Choices Among Alternatives of Uncertain Profitability in Various Networks. *Social Networks, 34*, 100-112.

Hofstra, B., Corten, R., and Van Tubergen, F. (2016). Who Was First on Facebook? Determinants of Early Adoption Among Adolescents. *New Media & Society, 18*(10), 2340-2358. (**Chapter 2** of this dissertation)

Hofstra, B., Corten, R., and Van Tubergen, F. (2016). Understanding the Privacy Behavior of Adolescents on Facebook: The Role of Peers, Popularity and Trust. *Computers in Human Behavior, 60*, 611-621. (**Chapter 3** of this dissertation)

Hofstra, B., Corten, R., Van Tubergen, F., and Ellison, N.C. (2017). Sources of Segregation in Social Networks: A Novel Approach Using Facebook. *American Sociological Review, 82*(3), 625-656. (**Chapter 4** of this dissertation)

Hofstra, B., Corten, R., Van Tubergen, F., and Weesie, J. (2017). How Large Are Extended Social Networks? Combining Evidence from Facebook and the Scale-Up Method. *Work in progress.* (**Chapter 5** of this dissertation)

Hofstra, B. and De Schipper, N.C. (2017). Predicting Ethnicity with First Names in Online Social Media Networks. *Under review.* (**Chapter 6** of this dissertation)

Hofstra, B. (2017). Soort Zoekt Soort op Facebook? *Forthcoming*, Demos.

Quite, W., Hofstra, B., Knigge, A., and De Schipper, N.C. (2013). De Werking van Sociaal Kapitaal in het Statusverwervingsproces in Nederlands. *Mens & Maatschappij, 88*(4), 400-425.

Bayer, J.B. and Hofstra, B. (2017). Curation to Variation in Human Social Networks. *Work in progress.*

ICS Dissertation Series

ICS Dissertation Series

The ICS-series presents dissertations of the Interuniversity Center for Social Science Theory and Methodology. Each of these studies aims at integrating explicit theory formation with state-of-the-art empirical research or at the development of advanced methods for empirical research. The ICS was founded in 1986 as a cooperative effort of the universities of Groningen and Utrecht. Since 1992, the ICS expanded to the University of Nijmegen and since 2017 to the University of Amsterdam (UvA). Most of the projects are financed by the participating universities or by the Netherlands Organization for Scientific Research (NWO). The international composition of the ICS graduate students is mirrored in the increasing international orientation of the projects and thus of the ICS-series itself.

1. C. van Liere. (1990). *Lastige leerlingen. Een empirisch onderzoek naar sociale oorzaken van probleemgedrag op basisscholen.* Amsterdam: Thesis Publishers.
2. Marco H.D. van Leeuwen. (1990). *Bijstand in Amsterdam, ca. 1800–1850. Armenzorg als beheersings- en overlevingsstrategie.* ICS-dissertation, Utrecht.
3. I. Maas. (1990). *Deelname aan podiumkunsten via de podia, de media en actieve beoefening. Substitutie of leereffecten?.* Amsterdam: Thesis Publishers.
4. M.I. Broese van Groenou. (1991). *Gescheiden netwerken. De relaties met vrienden en verwanten na echtscheiding.* Amsterdam: Thesis Publishers.
5. Jan M.M. van den Bos. (1991). *Dutch EC policy making. A model-guided approach to coordination and negotiation.* Amsterdam: Thesis Publishers.
6. Karin Sanders. (1991). *Vrouwelijke pioniers. Vrouwen en mannen met een 'mannelijke' hogere beroepsopleiding aan het begin van hun loopbaan.* Amsterdam: Thesis Publishers.
7. Sjerp de Vries. (1991). *Egoism, altruism, and social justice. Theory and experiments on cooperation in social dilemmas.* Amsterdam: Thesis Publishers.
8. Ronald S. Batenburg. (1991). *Automatisering in bedrijf.* Amsterdam: Thesis Publishers.
9. Rudi Wielers. (1991). *Selectie en allocatie op de arbeidsmarkt. Een uitwerking voor de informele en geïnstitutionaliseerde kinderopvang.* Amsterdam: Thesis Publishers.
10. Gert P. Westert. (1991). *Verschillen in ziekenhuisgebruik.* ICS-dissertation, Groningen.
11. Hanneke Hermsen. (1992). *Votes and policy preferences. Equilibria in party systems.* Amsterdam: Thesis Publishers.
12. Cora J.M. Maas. (1992). *Probleemleerlingen in het basisonderwijs.* Amsterdam: Thesis Publishers.

13. Ed A.W. Boxman. (1992). *Contacten en carrière. Een empirisch-theoretisch onderzoek naar de relatie tussen sociale netwerken en arbeidsmarktposities*. Amsterdam: Thesis Publishers.
14. Conny G.J. Taes. (1992). *Kijken naar banen. Een onderzoek naar de inschatting van arbeidsmarktkansen bij schoolverlaters uit het middelbaar beroepsonderwijs*. Amsterdam: Thesis Publishers.
15. Peter van Roozendaal. (1992). *Cabinets in multi-party democracies. The effect of dominant and central parties on cabinet composition and durability*. Amsterdam: Thesis Publishers.
16. Marcel van Dam. (1992). *Regio zonder regie. Verschillen in en effectiviteit van gemeentelijk arbeidsmarktbeleid*. Amsterdam: Thesis Publishers.
17. Tanja van der Lippe. (1993). *Arbeidsverdeling tussen mannen en vrouwen*. Amsterdam: Thesis Publishers.
18. Marc A. Jacobs. (1993). *Software: Kopen of kopiëren? Een sociaal-wetenschappelijk onderzoek onder PC-gebruikers*. Amsterdam: Thesis Publishers.
19. Peter van der Meer. (1993). *Verdringing op de Nederlandse arbeidsmarkt. Sector- en sekseverschillen*. Amsterdam: Thesis Publishers.
20. Gerbert Kraaykamp. (1993). *Over lezen gesproken. Een studie naar sociale differentiatie in leesgedrag*. Amsterdam: Thesis Publishers.
21. Evelien Zeggelink. (1993). *Strangers into friends. The evolution of friendship networks using an individual oriented modeling approach*. Amsterdam: Thesis Publishers.
22. Jaco Berveling. (1994). *Het stempel op de besluitvorming. Macht, invloed en besluitvorming op twee Amsterdamse beleidsterreinen*. Amsterdam: Thesis Publishers.
23. Wim Bernasco. (1994). *Coupled careers. The effects of spouse's resources on success at work*. Amsterdam: Thesis Publishers.
24. Liset van Dijk. (1994). *Choices in child care. The distribution of child care among mothers, fathers and non-parental care providers*. Amsterdam: Thesis Publishers.
25. Jos de Haan. (1994). *Research groups in Dutch sociology*. Amsterdam: Thesis Publishers.
26. K. Boahene. (1995). *Innovation adoption as a socio-economic process. The case of the Ghanaian cocoa industry*. Amsterdam: Thesis Publishers.
27. Paul E.M. Ligthart. (1995). *Solidarity in economic transactions. An experimental study of framing effects in bargaining and contracting*. Amsterdam: Thesis Publishers.
28. Roger Th. A.J. Leenders. (1995). *Structure and influence. Statistical models for the dynamics of actor attributes, network structure, and their interdependence*. Amsterdam: Thesis Publishers.
29. Beate Völker. (1995). *Should auld acquaintance be forgot...? Institutions of communism, the transition to capitalism and personal networks: The case of East Germany*. Amsterdam: Thesis Publishers.

30. A. Cancrinus-Matthijsse. (1995). *Tussen hulpverlening en ondernemerschap. Beroepsuitoefening en taakopvattingen van openbare apothekers in een aantal West-Europese landen.* Amsterdam: Thesis Publishers.

31. Nardi Steverink. (1996). *Zo lang mogelijk zelfstandig. Naar een verklaring van verschillen in oriëntatie ten aanzien van opname in een verzorgingstehuis onder fysiek kwetsbare ouderen.* Amsterdam: Thesis Publishers.

32. Ellen Lindeman. (1996). *Participatie in vrijwilligerswerk.* Amsterdam: Thesis Publishers.

33. Chris Snijders. (1996). *Trust and commitments.* Amsterdam: Thesis Publishers.

34. Koos Postma. (1996). *Changing prejudice in Hungary. A study on the collapse of state socialism and its impact on prejudice against gypsies and Jews.* Amsterdam: Thesis Publishers.

35. Jooske T. van Busschbach. (1996). *Uit het oog, uit het hart? Stabiliteit en verandering in persoonlijke relaties.* Amsterdam: Thesis Publishers.

36. René Torenvlied. (1996). *Besluiten in uitvoering. Theorieën over beleidsuitvoering modelmatig getoetst op sociale vernieuwing in drie gemeenten.* Amsterdam: Thesis Publishers.

37. Andreas Flache. (1996). *The Double edge of networks. An analysis of the effect of informal networks on cooperation in social dilemmas.* Amsterdam: Thesis Publishers.

38. Kees van Veen. (1997). *Inside an internal labor market: Formal rules, flexibility and career lines in a Dutch manufacturing company.* Amsterdam: Thesis Publishers.

39. Lucienne van Eijk. (1997). *Activity and well-being in the elderly.* Amsterdam: Thesis Publishers.

40. Róbert Gál. (1997). *Unreliability. Contract discipline and contract governance under economic transition.* Amsterdam: Thesis Publishers.

41. Anne-Geerte van de Goor. (1997). *Effects of regulation on disability duration.* ICS-dissertation, Utrecht.

42. Boris Blumberg. (1997). *Das Management von Technologiekooperationen. Partnersuche und Verhandlungen mit dem Partner aus Empirisch-Theoretischer Perspektive.* ICS-dissertation, Utrecht.

43. Marijke von Bergh. (1997). *Loopbanen van oudere werknemers.* Amsterdam: Thesis Publishers.

44. Anna Petra Nieboer. (1997). *Life-events and well-being: A prospective study on changes in well-being of elderly people due to a serious illness event or death of the spouse.* Amsterdam: Thesis Publishers.

45. Jacques Niehof. (1997). *Resources and social reproduction: The effects of cultural and material resources on educational and occupational careers in industrial nations at the end of the twentieth century.* ICS-dissertation, Nijmegen.

46. Ariana Need. (1997). *The kindred vote. Individual and family effects of social class and religion on electoral change in the Netherlands, 1956–1994*. ICS-dissertation, Nijmegen.
47. Jim Allen. (1997). *Sector composition and the effect of education on wages: An international comparison*. Amsterdam: Thesis Publishers.
48. Jack B.F. Hutten. (1998). *Workload and provision of care in general practice. An empirical study of the relation between workload of Dutch general practitioners and the content and quality of their care*. ICS-dissertation, Utrecht.
49. Per B. Kropp. (1998). *Berufserfolg im Transformationsprozeß, Eine theoretisch-empirische Studie über die Gewinner und Verlierer der Wende in Ostdeutschland*. ICS-dissertation, Utrecht.
50. Maarten H.J. Wolbers. (1998). *Diploma-inflatie en verdringing op de arbeidsmarkt. Een studie naar ontwikkelingen in de opbrengsten van diploma's in Nederland*. ICS-dissertation, Nijmegen.
51. Wilma Smeenk. (1998). *Opportunity and marriage. The impact of individual resources and marriage market structure on first marriage timing and partner choice in the Netherlands*. ICS-dissertation, Nijmegen.
52. Marinus Spreen. (1999). *Sampling personal network structures: Statistical inference in ego-graphs*. ICS-dissertation, Groningen.
53. Vincent Buskens. (1999). *Social networks and trust*. ICS-dissertation, Utrecht.
54. Susanne Rijken. (1999). *Educational expansion and status attainment. A cross-national and over-time comparison*. ICS-dissertation, Utrecht.
55. Mérove Gijsberts. (1999). *The legitimation of inequality in state-socialist and market societies, 1987–1996*. ICS-dissertation, Utrecht.
56. Gerhard G. Van de Bunt. (1999). *Friends by choice. An actor-oriented statistical network model for friendship networks through time*. ICS-dissertation, Groningen.
57. Robert Thomson. (1999). *The party mandate: Election pledges and government actions in the Netherlands, 1986–1998*. Amsterdam: Thela Thesis.
58. Corine Baarda. (1999). *Politieke besluiten en boeren beslissingen. Het draagvlak van het mestbeleid tot 2000*. ICS-dissertation, Groningen.
59. Rafael Wittek. (1999). *Interdependence and informal control in organizations*. ICS-dissertation, Groningen.
60. Diane Payne. (1999). *Policy making in the European Union: An analysis of the impact of the reform of the structural funds in Ireland*. ICS-dissertation, Groningen.
61. René Veenstra. (1999). *Leerlingen – klassen – scholen. Prestaties en vorderingen van leerlingen in het voortgezet onderwijs*. Amsterdam: Thela Thesis.
62. Marjolein Achterkamp. (1999). *Influence strategies in collective decision making. A comparison of two models*. ICS-dissertation, Groningen.
63. Peter Mühlau. (2000). *The governance of the employment relation. A relational signaling perspective*. ICS-dissertation, Groningen.

64. Agnes Akkerman. (2000). *Verdeelde vakbeweging en stakingen. Concurrentie om leden.* ICS-dissertation, Groningen.
65. Sandra van Thiel. (2000). *Quangocratization: Trends, causes and consequences.* ICS-dissertation, Utrecht.
66. Rudi Turksema. (2000). *Supply of day care.* ICS-dissertation, Utrecht.
67. Sylvia E. Korupp (2000). *Mothers and the process of social stratification.* ICS-dissertation, Utrecht.
68. Bernard A. Nijstad (2000). *How the group affects the mind: Effects of communication in idea generating groups.* ICS-dissertation, Utrecht.
69. Inge F. de Wolf (2000). *Opleidingsspecialisatie en arbeidsmarktsucces van sociale wetenschappers.* ICS-dissertation, Utrecht.
70. Jan Kratzer (2001). *Communication and performance: An empirical study in innovation teams.* ICS-dissertation, Groningen.
71. Madelon Kroneman (2001). *Healthcare systems and hospital bed use.* ICS/NIVEL-dissertation, Utrecht.
72. Herman van de Werfhorst (2001). *Field of study and social inequality. Four types of educational resources in the process of stratification in the Netherlands.* ICS-dissertation, Nijmegen.
73. Tamás Bartus (2001). *Social capital and earnings inequalities. The role of informal job search in Hungary.* ICS-dissertation, Groningen.
74. Hester Moerbeek (2001). *Friends and foes in the occupational career. The influence of sweet and sour social capital on the labour market.* ICS-dissertation, Nijmegen.
75. Marcel van Assen (2001). *Essays on actor perspectives in exchange networks and social dilemmas.* ICS-dissertation, Groningen.
76. Inge Sieben (2001). *Sibling similarities and social stratification. The impact of family background across countries and cohorts.* ICS-dissertation, Nijmegen.
77. Alinda van Bruggen (2001). *Individual production of social well-being. An exploratory study.* ICS-dissertation, Groningen.
78. Marcel Coenders (2001). *Nationalistic attitudes and ethnic exclusionism in a comparative perspective: An empirical study of attitudes toward the country and ethnic immigrants in 22 countries.* ICS-dissertation, Nijmegen.
79. Marcel Lubbers (2001). *Exclusionistic electorates. Extreme right-wing voting in Western Europe.* ICS-dissertation, Nijmegen.
80. Uwe Matzat (2001). *Social networks and cooperation in electronic communities. A theoretical-empirical analysis of academic communication and internet discussion groups.* ICS-dissertation, Groningen.
81. Jacques P.G. Janssen (2002). *Do opposites attract divorce? Dimensions of mixed marriage and the risk of divorce in the Netherlands.* ICS-dissertation, Nijmegen.
82. Miranda Jansen (2002). *Waardenoriëntaties en partnerrelaties. Een panelstudie naar wederzijdse invloeden.* ICS-dissertation, Utrecht.

83. Anne Rigt Poortman (2002). *Socioeconomic causes and consequences of divorce.* ICS-dissertation, Utrecht.
84. Alexander Gattig (2002). *Intertemporal decision making.* ICS-dissertation, Groningen.
85. Gerrit Rooks (2002). *Contract en conflict: Strategisch management van inkooptransacties.* ICS-dissertation, Utrecht.
86. Károly Takács (2002). *Social networks and intergroup conflict.* ICS-dissertation, Groningen.
87. Thomas Gautschi (2002). *Trust and exchange, effects of temporal embeddedness and network embeddedness on providing and dividing a surplus.* ICS-dissertation, Utrecht.
88. Hilde Bras (2002). *Zeeuwse meiden. Dienen in de levensloop van vrouwen, ca. 1850–1950.* Amsterdam: Aksant Academic Publishers.
89. Merijn Rengers (2002). *Economic lives of artists. Studies into careers and the labour market in the cultural sector.* ICS-dissertation, Utrecht.
90. Annelies Kassenberg (2002). *Wat scholieren bindt. Sociale gemeenschap in scholen.* ICS-dissertation, Groningen
91. Marc Verboord (2003). *Moet de meester dalen of de leerling klimmen? De invloed van literatuuronderwijs en ouders op het lezen van boeken tussen 1975 en 2000.* ICS-dissertation, Utrecht.
92. Marcel van Egmond (2003). *Rain falls on all of us (but some manage to get more wet than others): Political Context and Electoral Participation.* ICS-dissertation, Nijmegen.
93. Justine Horgan (2003). *High performance human resource management in Ireland and the Netherlands: Adoption and effectiveness.* ICS-dissertation, Groningen.
94. Corine Hoeben (2003). *LETS' be a community. Community in local exchange trading systems.* ICS-dissertation, Groningen.
95. Christian Steglich (2003). *The framing of decision situations. Automatic goal selection and rational goal pursuit.* ICS-dissertation, Groningen.
96. Johan van Wilsem (2003). *Crime and context. The impact of individual, neighborhood, city and country characteristics on victimization.* ICS-dissertation, Nijmegen.
97. Christiaan Monden (2003). *Education, inequality and health. The impact of partners and life course.* ICS-dissertation, Nijmegen.
98. Evelyn Hello (2003). *Educational attainment and ethnic attitudes. How to explain their relationship.* ICS-dissertation, Nijmegen.
99. Marnix Croes en Peter Tammes (2004). *Gif laten wij niet voortbestaan. Een onderzoek naar de overlevingskansen van joden in de Nederlandse gemeenten, 1940–1945.* Amsterdam: Aksant Academic Publishers.
100. Ineke Nagel (2004). *Cultuurdeelname in de levensloop.* ICS-dissertation, Utrecht.

101. Marieke van der Wal (2004). *Competencies to participate in life. Measurement and the impact of school.* ICS-dissertation, Groningen.
102. Vivian Meertens (2004). *Depressive symptoms in the general population: A multifactorial social approach.* ICS-dissertation, Nijmegen.
103. Hanneke Schuurmans (2004). *Promoting well-being in frail elderly people. Theory and intervention.* ICS-dissertation, Groningen.
104. Javier Arregui (2004). *Negotiation in legislative decision-making in the European Union.* ICS-dissertation, Groningen.
105. Tamar Fischer (2004). *Parental divorce, conflict and resources. The effects on children's behaviour problems, socioeconomic attainment, and transitions in the demographic career.* ICS-dissertation, Nijmegen.
106. René Bekkers (2004). *Giving and volunteering in the Netherlands: Sociological and psychological perspectives.* ICS-dissertation, Utrecht.
107. Renée van der Hulst (2004). *Gender differences in workplace authority: An empirical study on social networks.* ICS-dissertation, Groningen.
108. Rita Smaniotto (2004). *'You scratch my back and I scratch yours' versus 'Love thy neighbour'. Two Proximate Mechanisms of Reciprocal Altruism.* ICS-dissertation, Groningen.
109. Maurice Gesthuizen (2004). *The life-course of the low-educated in the Netherlands: Social and economic risks.* ICS-dissertation, Nijmegen.
110. Carlijne Philips (2005). *Vakantiegemeenschappen. Kwalitatief en kwantitatief onderzoek naar gelegenheid- en refreshergemeenschap tijdens de vakantie.* ICS-dissertation, Groningen.
111. Esther de Ruijter (2005). *Household outsourcing.* ICS-dissertation, Utrecht.
112. Frank van Tubergen (2005). *The integration of immigrants in cross-national perspective: Origin, destination, and community effects.* ICS-dissertation, Utrecht.
113. Ferry Koster (2005). *For the time being. Accounting for inconclusive findings concerning the effects of temporary employment relationships on solidary behavior of employees.* ICS-dissertation, Groningen.
114. Carolien Klein Haarhuis (2005). *Promoting anti-corruption reforms. Evaluating the implementation of a World Bank anti-corruption program in seven African countries (1999–2001).* ICS-dissertation, Utrecht.
115. Martin van der Gaag (2005). *Measurement of individual social capital.* ICS-dissertation, Groningen.
116. Johan Hansen (2005). *Shaping careers of men and women in organizational contexts.* ICS-dissertation, Utrecht.
117. Davide Barrera (2005). *Trust in embedded settings.* ICS-dissertation, Utrecht.
118. Mattijs Lambooij (2005). *Promoting cooperation. Studies into the effects of long-term and short-term rewards on cooperation of employees.* ICS-dissertation, Utrecht.

119. Lotte Vermeij (2006). *What's cooking? Cultural boundaries among Dutch teenagers of different ethnic origins in the context of school.* ICS-dissertation, Utrecht.
120. Mathilde Strating (2006). *Facing the challenge of rheumatoid arthritis. A 13-year prospective study among patients and cross-sectional study among their partners.* ICS-dissertation, Groningen.
121. Jannes de Vries (2006). *Measurement error in family background variables: The bias in the intergenerational transmission of status, cultural consumption, party preference, and religiosity.* ICS-dissertation, Nijmegen.
122. Stefan Thau (2006). *Workplace deviance: Four studies on employee motives and self-regulation.* ICS-dissertation, Groningen.
123. Mirjam Plantinga (2006). *Employee motivation and employee performance in child care. The effects of the introduction of market forces on employees in the Dutch child-care sector.* ICS-dissertation, Groningen.
124. Helga de Valk (2006). *Pathways into adulthood. A comparative study on family life transitions among migrant and Dutch youth.* ICS-dissertation, Utrecht.
125. Henrike Elzen (2006). *Self-management for chronically ill older people.* ICS-dissertation, Groningen.
126. Ayşe Güveli (2007). *New social classes within the service class in the Netherlands and Britain. Adjusting the EGP class schema for the technocrats and the social and cultural specialists.* ICS-dissertation, Nijmegen.
127. Willem-Jan Verhoeven (2007). *Income attainment in post-communist societies.* ICS-dissertation, Utrecht.
128. Marieke Voorpostel (2007). *Sibling support: The exchange of help among brothers and sisters in the Netherlands.* ICS-dissertation, Utrecht.
129. Jacob Dijkstra (2007). *The effects of externalities on partner choice and payoffs in exchange networks.* ICS-dissertation, Groningen.
130. Patricia van Echtelt (2007). *Time-greedy employment relationships: Four studies on the time claims of post-Fordist work.* ICS-dissertation, Groningen.
131. Sonja Vogt (2007). *Heterogeneity in social dilemmas: The case of social support.* ICS-dissertation, Utrecht.
132. Michael Schweinberger (2007). *Statistical methods for studying the evolution of networks and behavior.* ICS-dissertation, Groningen.
133. István Back (2007). *Commitment and evolution: Connecting emotion and reason in long-term relationships.* ICS-dissertation, Groningen.
134. Ruben van Gaalen (2007). *Solidarity and ambivalence in parent-child relationships.* ICS-dissertation, Utrecht.
135. Jan Reitsma (2007). *Religiosity and solidarity - Dimensions and relationships disentangled and tested.* ICS-dissertation, Nijmegen.
136. Jan Kornelis Dijkstra (2007) *Status and affection among (pre)adolescents and their relation with antisocial and prosocial behavior.* ICS-dissertation, Groningen.

137. Wouter van Gils (2007). *Full-time working couples in the Netherlands. Causes and consequences.* ICS-dissertation, Nijmegen.
138. Djamila Schans (2007). *Ethnic diversity in intergenerational solidarity.* ICS-dissertation, Utrecht.
139. Ruud van der Meulen (2007). *Brug over woelig water: Lidmaatschap van sportverenigingen, vriendschappen, kennissenkringen en veralgemeend vertrouwen.* ICS-dissertation, Nijmegen.
140. Andrea Knecht (2008). *Friendship selection and friends' influence. Dynamics of networks and actor attributes in early adolescence.* ICS-dissertation, Utrecht.
141. Ingrid Doorten (2008). *The division of unpaid work in the household: A stubborn pattern?.* ICS-dissertation, Utrecht.
142. Stijn Ruiter (2008). *Association in context and association as context: Causes and consequences of voluntary association involvement.* ICS-dissertation, Nijmegen.
143. Janneke Joly (2008). *People on our minds: When humanized contexts activate social norms.* ICS-dissertation, Groningen.
144. Margreet Frieling (2008). *'Joint production' als motor voor actief burgerschap in de buurt.* ICS-dissertion, Groningen.
145. Ellen Verbakel (2008). *The partner as resource or restriction? Labour market careers of husbands and wives and the consequences for inequality between couples.* ICS-dissertation, Nijmegen.
146. Gijs van Houten (2008). *Beleidsuitvoering in gelaagde stelsels. De doorwerking van aanbevelingen van de Stichting van de Arbeid in het CAO-overleg.* ICS-dissertation, Utrecht.
147. Eva Jaspers (2008). *Intolerance over time. Macro and micro level questions on attitudes towards euthanasia, homosexuality and ethnic minorities.* ICS-dissertation, Nijmegen.
148. Gijs Weijters (2008). *Youth delinquency in Dutch cities and schools: A multilevel approach.* ICS-dissertation, Nijmegen.
149. Jessica Pass (2009). *The self in social rejection.* ICS-dissertation, Groningen.
150. Gerald Mollenhorst (2009). *Networks in contexts. How meeting opportunities affect personal relationships.* ICS-dissertation, Utrecht.
151. Tom van der Meer (2009). *States of freely associating citizens: Comparative studies into the impact of state institutions on social, civic and political participation.* ICS-dissertation, Nijmegen.
152. Manuela Vieth (2009). *Commitments and reciprocity in trust situations. Experimental studies on obligation, indignation, and self-consistency.* ICS-dissertation, Utrecht.
153. Rense Corten (2009). *Co-evolution of social networks and behavior in social dilemmas: Theoretical and empirical perspectives.* ICS-dissertation, Utrecht.
154. Arieke J. Rijken (2009). *Happy families, high fertility? Childbearing choices in the context of family and partner relationships.* ICS-dissertation, Utrecht.

155. Jochem Tolsma (2009). *Ethnic hostility among ethnic majority and minority groups in the Netherlands. An investigation into the impact of social mobility experiences, the local living environment and educational attainment on ethnic hostility*. ICS-dissertation, Nijmegen.
156. Freek Bucx (2009). *Linked lives: Young adults' life course and relations with parents*. ICS-dissertation, Utrecht.
157. Philip Wotschack (2009). *Household governance and time allocation. Four studies on the combination of work and care*. ICS-dissertation, Groningen.
158. Nienke Moor (2009). *Explaining worldwide religious diversity. The relationship between subsistence technologies and ideas about the unknown in pre-industrial and (post-)industrial societies*. ICS-dissertation, Nijmegen.
159. Lieke ten Brummelhuis (2009). *Family matters at work. Depleting and enriching effects of employees' family lives on work outcomes*. ICS-dissertation, Utrecht.
160. Renske Keizer (2010). *Remaining childless. Causes and consequences from a life course perspective*. ICS-dissertation, Utrecht.
161. Miranda Sentse (2010). *Bridging contexts: The interplay between family, child, and peers in explaining problem behavior in early adolescence*. ICS-dissertation, Groningen.
162. Nicole Tieben (2010). *Transitions, tracks and transformations. Social inequality in transitions into, through and out of secondary education in the Netherlands for cohorts born between 1914 and 1985*. ICS-dissertation, Nijmegen.
163. Birgit Pauksztat (2010). *Speaking up in organizations: Four studies on employee voice*. ICS-dissertation, Groningen.
164. Richard Zijdeman (2010). *Status attainment in the Netherlands, 1811-1941. Spatial and temporal variation before and during industrialization*. ICS-dissertation, Utrecht.
165. Rianne Kloosterman (2010). *Social background and children's educational careers. The primary and secondary effects of social background over transitions and over time in the Netherlands*. ICS-dissertation, Nijmegen.
166. Olav Aarts (2010). *Religious diversity and religious involvement. A study of religious markets in Western societies at the end of the twentieth century*. ICS-dissertation, Nijmegen.
167. Stephanie Wiesmann (2010). *24/7 negotiation in couples transition to parenthood*. ICS-dissertation, Utrecht.
168. Borja Martinovic (2010). *Interethnic contacts: A dynamic analysis of interaction between immigrants and natives in Western countries*. ICS-dissertation, Utrecht.
169. Anne Roeters (2010). *Family life under pressure? Parents' paid work and the quantity and quality of parent-child and family time*. ICS-dissertation, Utrecht.
170. Jelle Sijtsema (2010). *Adolescent aggressive behavior: Status and stimulation goals in relation to the peer context*. ICS-dissertation, Groningen.
171. Kees Keizer (2010). *The spreading of disorder*. ICS-dissertation, Groningen.

172. Michael Mäs (2010). *The diversity puzzle. Explaining clustering and polarization of opinions.* ICS-dissertation, Groningen.

173. Marie-Louise Damen (2010). *Cultuurdeelname en CKV. Studies naar effecten van kunsteducatie op de cultuurdeelname van leerlingen tijdens en na het voortgezet onderwijs.* ICS-dissertation, Utrecht.

174. Marieke van de Rakt (2011). *Two generations of crime: The intergenerational transmission of convictions over the life course.* ICS-dissertation, Nijmegen.

175. Willem Huijnk (2011). *Family life and ethnic attitudes. The role of the family for attitudes towards intermarriage and acculturation among minority and majority groups.* ICS-dissertation, Utrecht.

176. Tim Huijts (2011). *Social ties and health in Europe. Individual associations, cross-national variations, and contextual explanations.* ICS-dissertation, Nijmegen.

177. Wouter Steenbeek (2011). *Social and physical disorder. How community, business presence and entrepreneurs influence disorder in Dutch neighborhoods.* ICS-dissertation, Utrecht.

178. Miranda Vervoort (2011). *Living together apart? Ethnic concentration in the neighborhood and ethnic minorities' social contacts and language practices.* ICS-dissertation, Utrecht.

179. Agnieszka Kanas (2011). *The economic performance of immigrants. The role of human and social capital.* ICS-dissertation, Utrecht.

180. Lea Ellwardt (2011). *Gossip in organizations. A social network study.* ICS-dissertation, Groningen.

181. Annemarije Oosterwaal (2011). *The gap between decision and implementation. Decision making, delegation and compliance in governmental and organizational settings.* ICS-dissertation, Utrecht.

182. Natascha Notten (2011). *Parents and the media. Causes and consequences of parental media socialization.* ICS-dissertation, Nijmegen.

183. Tobias Stark (2011). *Integration in schools. A process perspective on students' interethnic attitudes and interpersonal relationships.* ICS-dissertation, Groningen.

184. Giedo Jansen (2011). *Social cleavages and political choices. Large-scale comparisons of social class, religion and voting behavior in Western democracies.* ICS-dissertation, Nijmegen.

185. Ruud van der Horst (2011). *Network effects on treatment results in a closed forensic psychiatric setting.* ICS-dissertation, Groningen.

186. Mark Levels (2011). *Abortion laws in European countries between 1960 and 2010. Legislative developments and their consequences for women's reproductive decision-making.* ICS-dissertation, Nijmegen.

187. Marieke van Londen (2012). *Exclusion of ethnic minorities in the Netherlands. The effects of individual and situational characteristics on opposition to ethnic policy and ethnically mixed neighbourhoods.* ICS-dissertation, Nijmegen.

188. Sigrid M. Mohnen (2012). *Neighborhood context and health: How neighborhood social capital affects individual health.* ICS-dissertation, Utrecht.
189. Asya Zhelyazkova (2012). *Compliance under controversy: Analysis of the transposition of European directives and their provisions.* ICS-dissertation, Utrecht.
190. Valeska Korff (2012). *Between cause and control: Management in a humanitarian organization.* ICS-dissertation, Groningen.
191. Maike Gieling (2012). *Dealing with diversity: Adolescents' support for civil liberties and immigrant rights.* ICS-dissertation, Utrecht.
192. Katya Ivanova (2012). *From parents to partners: The impact of family on romantic relationships in adolescence and emerging adulthood.* ICS-dissertation, Groningen.
193. Jelmer Schalk (2012). *The performance of public corporate actors: Essays on effects of institutional and network embeddedness in supranational, national, and local collaborative contexts.* ICS-dissertation, Utrecht.
194. Alona Labun (2012). *Social networks and informal power in organizations.* ICS-dissertation, Groningen.
195. Michał Bojanowski (2012). *Essays on social network formation in heterogeneous populations: Models, methods, and empirical analyses.* ICS-dissertation, Utrecht.
196. Anca Minescu (2012). *Relative group position and intergroup attitudes in Russia.* ICS-dissertation, Utrecht.
197. Marieke van Schellen (2012). *Marriage and crime over the life course. The criminal careers of convicts and their spouses.* ICS-dissertation, Utrecht.
198. Mieke Maliepaard (2012). *Religious trends and social integration: Muslim minorities in the Netherlands.* ICS-dissertation, Utrecht.
199. Fransje Smits (2012). *Turks and Moroccans in the Low Countries around the year 2000: determinants of religiosity, trend in religiosity and determinants of the trend.* ICS-dissertation, Nijmegen.
200. Roderick Sluiter (2012). *The diffusion of morality policies among Western European countries between 1960 and 2010. A comparison of temporal and spatial diffusion patterns of six morality and eleven non-morality policies.* ICS-dissertation, Nijmegen.
201. Nicoletta Balbo (2012). *Family, friends and fertility.* ICS-dissertation, Groningen.
202. Anke Munniksma (2013). *Crossing ethnic boundaries: Parental resistance to and consequences of adolescents' cross-ethnic peer relations.* ICS-dissertation, Groningen.
203. Anja Abendroth (2013). *Working women in Europe. How the country, workplace, and family context matter.* ICS-dissertation, Utrecht.
204. Katia Begall (2013). *Occupational hazard? The relationship between working conditions and fertility.* ICS-dissertation, Groningen.
205. Hidde Bekhuis (2013). *The popularity of domestic cultural products: Cross- national differences and the relation to globalization.* ICS-dissertation, Utrecht.

206. Lieselotte Blommaert (2013). *Are Joris and Renske more employable than Rashid and Samira? A study on the prevalence and sources of ethnic discrimination in recruitment in the Netherlands using experimental and survey data.* ICS-dissertation, Utrecht.
207. Wiebke Schulz (2013). *Careers of men and women in the 19th and 20th centuries.* ICS-dissertation, Utrecht.
208. Ozan Aksoy (2013). *Essays on social preferences and beliefs in non-embedded social dilemmas.* ICS-dissertation, Utrecht.
209. Dominik Morbitzer (2013). *Limited farsightedness in network formation.* ICS-dissertation, Utrecht.
210. Thomas de Vroome (2013). *Earning your place: The relation between immigrants' economic and psychological integration in the Netherlands.* ICS-dissertation, Utrecht.
211. Marloes de Lange (2013). *Causes and consequences of employment flexibility among young people. Recent developments in the Netherlands and Europe.* ICS-dissertation, Nijmegen.
212. Roza Meuleman (2014). *Consuming the nation. Domestic cultural consumption: Its stratification and relation with nationalist attitudes.* ICS-dissertation, Utrecht.
213. Esther Havekes (2014). *Putting interethnic attitudes in context. The relationship between neighbourhood characteristics, interethnic attitudes and residential behaviour.* ICS-dissertation, Utrecht.
214. Zoltán Lippényi (2014). *Transitions toward an open society? Intergenerational occupational mobility in Hungary in the 19th and 20th centuries.* ICS-dissertation, Utrecht
215. Anouk Smeekes (2014). *The presence of the past: Historical rooting of national identity and current group dynamics.* ICS-dissertation, Utrecht.
216. Michael Savelkoul (2014). *Ethnic diversity and social capital. Testing underlying explanations derived from conflict and contact theories in Europe and the United States.* ICS-dissertation, Nijmegen.
217. Martijn Hogerbrugge (2014). *Misfortune and family: How negative events, family ties, and lives are linked.* ICS-dissertation, Utrecht.
218. Gina Potarca (2014). *Modern love. Comparative insights in online dating preferences and assortative mating.* ICS-dissertation, Groningen.
219. Mariska van der Horst (2014). *Gender, aspirations, and achievements: Relating work and family aspirations to occupational outcomes.* ICS-dissertation, Utrecht.
220. Gijs Huitsing (2014). *A social network perspective on bullying.* ICS dissertation, Groningen.
221. Thomas Kowalewski (2015). *Personal growth in organizational contexts.* ICS-dissertation, Groningen.
222. Manu Muñoz-Herrera (2015). *The impact of individual differences on network relations: Social exclusion and inequality in productive exchange and coordination games.* ICS-dissertation, Groningen.

223. Tim Immerzeel (2015). *Voting for a change. The democratic lure of populist radical right parties in voting behavior.* ICS-dissertation, Utrecht.
224. Fernando Nieto Morales (2015). *The control imperative: Studies on reorganization in the public and private sectors.* ICS-dissertation, Groningen.
225. Jellie Sierksma (2015). *Bounded helping: How morality and intergroup relations shape children's reasoning about helping.* ICS-dissertation, Utrecht.
226. Tinka Veldhuis (2015). *Captivated by fear. An evaluation of terrorism detention policy.* ICS-dissertation, Groningen.
227. Miranda Visser (2015). *Loyality in humanity. Turnover among expatriate humanitarian aid workers.* ICS-dissertation, Groningen.
228. Sarah Westphal (2015). *Are the kids alright? Essays on postdivorce residence arrangements and children's well-being.* ICS-dissertation, Utrecht.
229. Britta Rüschoff (2015). *Peers in careers: Peer relationships in the transition from school to work.* ICS-dissertation, Groningen.
230. Nynke van Miltenburg (2015). *Cooperation under peer sanctioning institutions: Collective decisions, noise, and endogenous implementation.* ICS-dissertation, Utrecht.
231. Antonie Knigge (2015). *Sources of sibling similarity. Status attainment in the Netherlands during modernization.* ICS-dissertation, Utrecht.
232. Sanne Smith (2015). *Ethnic segregation in friendship networks. Studies of its determinants in English, German, Dutch, and Swedish school classes.* ICS- dissertation, Utrecht.
233. Patrick Präg (2015). *Social stratification and health. Four essays on social determinants of health and wellbeing.* ICS-dissertation, Groningen.
234. Wike Been (2015). *European top managers' support for work-life arrangements.* ICS- dissertation, Utrecht.
235. André Grow (2016). *Status differentiation: New insights from agent-based modeling and social network analysis.* ICS-dissertation, Groningen.
236. Jesper Rözer (2016). *Family and personal networks. How a partner and children affect social relationships.* ICS-dissertation, Utrecht.
237. Kim Pattiselanno (2016). *At your own risk: The importance of group dynamics and peer processes in adolescent peer groups for adolescents' involvement in risk behaviors.* ICS-dissertation, Groningen.
238. Vincenz Frey (2016). *Network formation and trust.* ICS-dissertation, Utrecht.
239. Rozemarijn van der Ploeg (2016). *Be a buddy, not a bully? Four studies on social and emotional processes related to bullying, defending, and victimization.* ICS-dissertation, Groningen.
240. Tali Spiegel (2016). *Identity, career trajectories and wellbeing: A closer look at individuals with degenerative eye conditions.* ICS-dissertation, Groningen.
241. Felix Tropf (2016). *Social Science Genetics and Fertility.* ICS-dissertation, Groningen.

242. Sara Geven (2016). *Adolescents problem behavior in school: the role of peer networks.* ICS-dissertation, Utrecht.
243. Josja Rokven (2016). *The victimization-offending relationship from a longitudinal perspective.* ICS-dissertation, Nijmegen.
244. Maja Djundeva (2016). *Healthy ageing in context: Family welfare state and the life course.* ICS-dissertation, Groningen.
245. Mark Visser (2017). *Inequality between older workers and older couples in the Netherlands. A dynamic life course perspective on educational and social class differences in the late career.* ICS-dissertation, Nijmegen.
246. Beau Oldenburg (2017). *Bullying in schools: The role of teachers and classmates.* ICS-dissertation, Groningen.
247. Tatang Muttaqin (2017). *The education divide in Indonesia: Four essays on determinants of unequal access to and quality of education.* ICS-dissertation, Groningen.
248. Margriet van Hek (2017). *Gender inequality in educational attainment and reading performance. A contextual approach.* ICS-dissertation, Nijmgen.
249. Melissa Verhoef (2017). *Work schedules, childcare and well-being. Essays on the associations between modern-day job characteristics, childcare arrangements and the well-being of parents and children.* ICS-dissertation, Utrecht.
250. Timo Septer (2017). *Goal priorities, cognition and conflict: Analyses of cognitive maps concerning organizational changes.* ICS-dissertation, Groningen.
251. Bas Hofstra (2017). *Online Social Networks: Essays on Membership, Privacy, and Structure.* ICS-dissertation, Utrecht.

The structure of social networks is crucial for obtaining social support, for meaningful connections to unknown social groups, and to overcome prejudice. Yet, we know little about the structure of social networks beyond those contacts that stand closest to us. This lack of knowledge results from a survey-research tradition in which solely strong social ties are mapped. This dissertation overcomes this issue by embracing a new feature of contemporary social life: the fact that individuals overwhelmingly maintain their social relationships online. The "digital footprints" of interactions left online enable scholars to test old and new theories on the structure of social networks in innovative ways. In this spirit, the goal of this dissertation is to understand the structure of online social networks for new insights into the structure of social networks in general. What are the theoretical and empirical promises and pitfalls of such a study? Bas Hofstra answers these questions through five empirical chapters in which he links offline survey data on Dutch adolescents with online network data from Facebook.

BAS HOFSTRA *obtained his Research Master's degree in sociology and social research at Utrecht University, graduating cum laude. He conducted the present study as part of his Ph.D. research at the Interuniversity Center for Social Science Theory and Methodology (ICS) and the Department of Sociology of Utrecht University.*