Social Networks 83 (2025) 173–185

Contents lists available at ScienceDirect

Social Networks

journal homepage: www.elsevier.com/locate/socnet

Who's in your extended network? Analysing the size and homogeneity of acquaintanceship networks in the Netherlands

Beate Völker[a,*], Bas Hofstra[b], Rense Corten[c], Frank van Tubergen[c,d]

[a] VU Amsterdam & Netherlands Institute for the Study of Crime and Law Enforcement, the Netherlands
[b] Radboud University, the Netherlands
[c] Utrecht University, the Netherlands
[d] Netherlands Interdisciplinary Demographic Institute, KNAW/University of Groningen, Netherlands

ARTICLE INFO

Keywords:
Network scale-up method
Extended networks
Segregation
Homogeneity

ABSTRACT

This study advances the understanding of the size and homogeneity of personal networks, focusing on extended networks that encompass both core discussion ties and the broader array of acquaintances. While previous research has primarily examined these dimensions within small, strong-tie networks, knowledge about extended networks remains limited. Using data from the Dutch Network Size Survey (2021), a representative survey of the Dutch adult population, this study provides novel insights into the size, gender, and educational homogeneity of extended networks, as well as individual variation across these dimensions. Employing the Network Scale-Up Method (NSUM) with an extensive set of scale-up items, we find a median extended network size of 446 and a mean size of 518. Substantial variation exists across individuals, with larger networks associated with being employed, having more household members, being younger, possessing greater resources (e.g., income, wealth), and attaining higher levels of education. Additionally, our findings reveal significant gender and educational segregation within extended networks. These results shed light on the structure of extended networks and highlight the social stratification of network size and homogeneity.

Introduction

Since the emergence of network research, scholars have identified size and homogeneity being among the key parameters of personal networks. Personal network size refers to the total number of individuals one knows in person (Killworth et al., 1990). In any given society, the quantity of connections that people have reflects their level of social interconnectedness, and, ultimately, the social cohesion of that society (Milgram, 1967; Dodds, Muhamad and Watts, 2003). At the micro-level, personal network size is linked to various desirable outcomes. For instance, larger networks are associated with increased access to social capital (Van Tubergen and Völker, 2015), social support (Arenson et al., 2021), improved mental health and well-being (e.g., Stokes, 1983; Penninx et al., 1999; Wang, 2016; Bidart, 2020), and reduced risk of mortality (Schutter et al., 2022; Laugesen et al., 2018; Domènech-Abella et al., 2019). In summary, network size is a crucial factor in a range of individual and societal outcomes, making it a key feature of a network's structure.

A second key element of a personal network's structure is its homogeneity – the range of others in a network who resemble the focal individual. Studying network homogeneity provides insights into the extent to which different social groups are connected. Seminal studies have shown that core networks are homogeneous regarding education, gender, ethnicity, and race (McPherson, Smith-Lovin and Cook, 2001; DiPrete et al., 2011; Smith, McPherson and Smith-Lovin, 2014; Hofstra et al., 2017; Jeroense et al., 2024). These outcomes reflect a dissociation from unfamiliar others, where groups from different walks of life neither meet nor mingle. This is potentially detrimental to social cohesion; it induces echo chambers, prejudice, and discrimination.

Previous research has attempted to describe and explain variations in personal network size and homogeneity across social groups, space, and place (Parigi and Henson, 2014). Most of these studies, however, focused on the *stronger* part of a person's network. Scholars have extensively examined the 'core discussion networks' of individuals, which capture the alters to whom people turn to discuss important matters (Marsden, 1987), and friendship networks of adolescents in school (Mouw and Entwisle, 2006). Relatedly, research has been done on people's 'sympathy' and 'affinity' network, which reflects people's

* Corresponding author.
E-mail address: bvolker@nscr.nl (B. Völker).

https://doi.org/10.1016/j.socnet.2025.05.007
Received 18 June 2024; Received in revised form 26 May 2025; Accepted 29 May 2025
Available online 27 June 2025

strong tie network beyond the inner core (Arnaboldi et al., 2017; Mac Carron, Kaski and Dunbar, 2016). Prior work studied how the size of these strong-tie networks changed over time (McPherson, Smith-Lovin and Brashears, 2006; Paik and Sanchagrin, 2013) and varies across social groups (DiPrete et al., 2011; Van Tubergen, 2014).

While research on the number and homogeneity of strong ties is thus rather well-established, far less attention has been paid to *extended* social networks, which include both stronger ties as well as a far larger number of weaker ties to acquaintances (see, e.g., Vacca et al., 2018). We contribute to the growing field of studies on extended social networks in two ways.

First, we provide new empirical evidence on the average and distribution of extended network size, as well as individual-level correlates of extended network size among adults in the Netherlands. As research on extended networks is rather scarce, there is a need to provide a solid empirical foundation for patterns observed across different contexts and samples. By providing new evidence from the Dutch Network Size Survey (DNSS, 2021), a representative survey of the Dutch adult population, our study contributes to prior work which focused on different samples, such as the US (Clay-Warner et al., 2022; Feehan and Salganik, 2016; Killworth et al., 1998; McCarty et al., 2001; McCormick et al., 2010), Spain (Lubbers et al., 2019), Chile (Contreras et al., 2019), and Dutch youth (Hofstra et al., 2021). While Jeroense et al. (2024) also studied the Netherlands, their focus is on ethnic group differences in network size.

Second, we study gender- and educational homogeneity patterns in extended networks. Although gender and education represent key sociological dimensions along which networks may be segregated, little is known about the extent to which such segregation occurs in extended networks. In addressing this important question, we elaborate on the few studies that examined homogeneity in extended networks. The seminal study by DiPrete et al. (2011) examined patterns of overall levels of segregation among acquaintances in the US along different dimensions (race, political ideology, religion). In our study, we address a different question, namely, how individual-level characteristics are related to knowing disproportionately many people from different genders and educational levels. A similar individual-level approach was taken by Hofstra et al. (2017) and Jeroense et al. (2024), who, however, focused on extended online (Facebook) networks of youth and on ethnic homogeneity in extended networks, respectively. We build on and extend their work by focusing on gender and educational homogeneity in extended networks for the general Dutch adult population.

Studying network size and homogeneity simultaneously is important because social relationships that ultimately determine a network's size and its homogeneity do not emerge in a vacuum. For instance, if more similar ties are accessible for certain social groups, their networks might grow larger and become more homogeneous at the same time, thus jointly influencing network size and composition. Likewise, when higher-educated individuals have larger networks, these extended networks might predominantly consist of others who are also highly educated, thereby coupling network size with educational homogeneity. By studying network size and homogeneity together, our study enhances the understanding of potentially interrelated inequalities in personal networks. As a point of departure to explain extended network size and homogeneity, we employ theories (e.g., opportunities, preferences) commonly used to explain the size and homogeneity in core networks. In doing so, we follow prior work that takes a similar approach (Hofstra et al., 2017, 2021; Lubbers et al., 2019; Jeroense et al., 2024).

Theory

Background

One of the earliest studies on the size of extended networks is Gurevitch's (1961) analysis of the diaries of 18 individuals, documenting their encounters over 100 days between November 1959 and February 1960. In his footsteps, a variety of methods have been established to estimate the size of people's extended network (Lubbers, Molina and Valenzuela Garcia 2019), such as contact diaries (Fu, 2005), the number of Christmas cards people send (Hill and Dunbar, 2003), connections on social media (Dunbar et al., 2015), and self-reported data on the number of ties with family, neighbors and other types of relations (Van Tubergen et al., 2016; Bernard et al., 1990).

A well-established approach for estimating the size of individuals' extended social networks in a population is the Network Scale-Up Method (NSUM, see Clay-Warner et al., 2022; Feehan and Salganik, 2016; Kilworth et al., 1990; McCarty et al., 2001; McCormick et al., 2010). This method asks respondents questions of the type 'How many people do you know in category X?' where 'X' can represent a specific name (e.g., 'Elisa') or social category with a known population size (e.g., 'women who gave birth in the past year'). By aggregating the responses and relating them to the known size of the categories, the scale-up method provides an estimate of people's extended network size. Importantly, all studies using NSUM report significant variation across individuals in their extended network, in line with the correlates earlier studies found with larger network sizes: income, education, being employed, and being native-born are generally associated with larger extended networks. Age and gender effects vary across studies (see the above-mentioned studies for NSUM and see the classic studies on determinants on network size, e.g., Marsden, 1987; Moore, 1990; Hill and Dunbar, 2003).

A few studies have used NSUMs to examine the *homogeneity* of extended networks. DiPrete et al. (2011) found that extended networks in the U.S. are highly segregated in terms of race, religion, and political ideology. Interestingly, no evidence was found for stronger social integration by the extended network compared to the core discussion network. In other words, network segregation exists not only among people's stronger ties. The weaker ties in extended networks, which are supposed to bridge social divides, are as strongly segregated as core networks. These patterns above are all dominated by racial and educational homogeneity. However, DiPrete et al. (2011) analysed their data at the group level, establishing patterns for the populations, rather than how individual-level characteristics are related to knowing disproportionately many people from different subgroups. Hofstra et al. (2017) and Jeroense et al. (2024) do take an individual-level approach. For online social connections based on Facebook data, Hofstra et al. (2017) observed that the extended networks of youth are strongly segregated by ethnicity and gender. Interestingly, especially among ethnic majority members, core networks were as ethnically homogeneous as extended networks. Extended networks of minority youth were more ethnically diverse than their core networks. Similarly, Jeroense et al. (2024) demonstrate for the Dutch context that the extended networks of the ethnic majority members are more ethnically homogenous than those of ethnic minority members.

There is little systematic theory on the determinants of the size and homogeneity of extended networks (Kadushin, 2012). Lubbers et al. (2019) and Hofstra et al. (2021), both studying extended networks, assumed that the processes underlying the formation of extended networks largely resemble those of the formation of core discussion networks. Studies explaining variation in size and homogeneity of core discussion networks have often relied on a *choice-constraint* approach (Fischer et al., 1977), assuming that social relations are the result of individual choices made under social constraints.

In terms of constraints, the literature on core discussion networks argues that *opportunities* to meet individuals are key (Blau, 1977). Feld (1981):(1016) coined the term 'focus' to describe such contexts: "social, psychological, legal, or physical entity around which joint activities are organised". Important examples of such contexts include workplaces (Zheng et al., 2006), associations (Mollenhorst et al., 2008a), neighbourhoods (Fischer et al., 1977; Kadushin and Jones, 1992, Mollenhorst et al., 2008b), and educational settings (Hofstra et al., 2017; Wimmer and Lewis, 2010). It is theorised that individuals who participate more frequently in a diverse range of such 'foci' tend to have larger and more diverse networks. However, and as mentioned already, when settings are highly segregated - for example, by education - increasing participation may lead to larger networks without increasing their educational diversity. In our study, we cannot account for the segregation of social settings, implying that we do not hypothesize the influence of a given setting on network homogeneity and size.

Social relationships are also influenced by individual choices. A well-known mechanism driving network homogeneity is *homophily* – the preference to associate with and befriend similar individuals (McPherson et al., 2001). Homophily results in more homogeneous networks, but it also constrains the formation of ties to dissimilar individuals, potentially limiting overall network size. Another mechanism is *preferential attachment,* which captures the idea that people with greater resources are seen as more attractive by others, which increases their social connections (Barabási and Albert, 1999, Barabási, 2016). These resources include one's status, prestige, and social capital (thereby giving rise to the 'Matthew effect,' i.e., people preferentially attaching to those having already larger networks). Finally, developing and maintaining social relations requires *ability*. Individuals with greater cognitive and physical abilities may find it easier to connect with others and are also better equipped to maintain a larger network of social relationships (Brashears et al., 2016; Hofstra et al., 2021).

Hypotheses

We rely on these four mechanisms to derive hypotheses on (a) network size and (b) gender/education homogeneity of networks. We start with the hypotheses on network size.

Following the opportunity mechanism, we expect to see that employment and the number of individuals in the household shape people's extended network size. Individuals who are *employed* are exposed to a larger number of people, including colleagues, business partners, and cafeteria staff, among others. Although occupations and job sectors in the Netherlands are somewhat segregated by gender and education, we assume that employment as such primarily impacts network size rather than gender/educational diversity. After all, even in a relatively segregated sector, there are many opportunities to contact diverse groups. We also expect that the number of *household members* matters. Individuals who live in larger households, whether with family or others, are exposed to more people in their daily lives. Households may act as network multipliers, as each household member introduces unique connections. These "strong ties" can, in turn, serve as linkages to other, weaker ties.

Furthermore, we expect that a person's *age* is related to network size. This association is driven by both the ability and opportunity mechanisms. People of an older age may face significant barriers in connecting with others, both mentally and physically (Wrzus et al., 2013, Cornwell et al., 2014). Because of physical conditions, they may be less mobile and more restricted in their activities and opportunities to meet others. In addition, declining cognitive skills may hamper the maintenance of many ties. Both gender and education may be related to network size as well. Because of gender roles, women are generally expected to put more effort into social relations (Lubbers et al., 2019, Brashears, Hoagland, Quintane, 2016). This normative expectation would imply that women are more socially active, increasing their opportunities to connect as well as being more cognitively able to establish and maintain larger networks. Based on these mechanisms, we expect to see that women have larger extended networks than men. In addition, we also expect that higher-educated individuals tend to have larger networks, partly due to their greater cognitive and physical ability, and their higher activity and participation in different foci (Hofstra et al., 2021).

Last but not least, we hypothesise that people who possess more *resources* have more extensive networks. Based on the preferential attachment mechanism, we assume that those with higher income or wealth are seen as having more status and being more attractive as a member of someone's network (Letki and Mieriņa, 2015), which results in a larger extended network.

In addition to theorising about network size, we also develop hypotheses on gender and educational homogeneity in extended networks. It is well-established that strong-tie networks are segregated by gender and education — a pattern driven by both segregated meeting opportunities and homophily (McPherson et al., 2001). In extended networks, the homophily mechanism may be weaker, as relationships tend to be more instrumental. Nevertheless, people may still prefer interactions with similar others. Moreover, institutional segregation and gender- and education-specific social activities further increase opportunities to form homogenous ties. Therefore, we expect to find evidence of gender and educational homogeneity within extended networks similarly as it has been found in smaller networks.

Hence, we argue that network size is largely shaped by mechanisms of opportunity and ability, and, insofar as (perceived) resources are involved, by the mechanism of preferential attachment.

Homogeneity in (extended) networks is likewise fueled by the opportunity mechanism as well as by homophily, but less so by the ability mechanism. As to preferential attachment, this might also influence homogeneity, in cases where the similarity of a contact is beneficial.

In summary, we expect to see larger extended networks among those who are employed (**H1**), who have more household members (**H2**), who are younger (**H3**), who possess greater resources (e.g., income, wealth) (**H4**), women (**H5**), and those with higher education (**H6**). We also expect that, in extended networks, women have more ties to women than to men (**H7**), and that higher (lower) educated individuals have more ties to higher (lower) educated individuals than lower (higher) educated individuals have (**H8**).

Methods

Data

The Dutch Network Size Survey (hereafter: DNSS) data was collected in 2021 by a professional fieldwork agency that employs a continuous representative (on both the national and regional level) panel of about 37,000 members from all over the Netherlands. The panel is constantly monitored and refreshed to ensure its representativeness (in terms of age, gender, migration background, education, and urban-rural areas), and members are recruited randomly from municipality registers or large address databases. A representative subset of N = 1472 respondents from the panel was invited to participate in a survey on their social networks and signed a consent form stating that their answers would be matched to their background information (e.g., postal codes, education, and so forth). Respondents filled in the survey in May 2021 and did not receive a monetary participation incentive. Yet, panel members do obtain savings points when finishing surveys, which they can exchange for gift cards or donations to charities. The survey was to be completed online (using any means: mobile, laptop, etc.) and lasted about 10–15 min. The cooperation rate among the 1472 invited panel members approached 90 %, with a realised sample for this study of N = 1325. The full questionnaire (see also Appendix 1), all raw and cleaned data (anonymous), and the annotated code to replicate the results are found on our companion website to this paper: https://bhofstra.github.io/netsize_dutch/.

The network scale-up method

The scale-up method is increasingly considered a promising method to measure the size and composition of extended networks (Baum and Marsden, 2023). With this method, respondents report the number of persons they know in various subpopulations. Since the prevalence of these subpopulations is known, the network size of the respondent can be 'scaled up' (Freeman and Thompson, 1989). The method has usually been applied to estimates of hidden or unknown populations, such as the number of victims in an earthquake (Bernard et al., 1991), or the number of heroin users (Kadushin et al., 2006).

While estimating network size is a vital step in estimating the size of hidden populations, the method also offers the possibility to study network size and composition in a general population through specific alter characteristics. Unlike other data used for estimating network parameters, the method does not rely on the delineation of individual network members but collects the numbers of people known with a specific characteristic, so-called aggregated relational data (or ARD). Respondents answer questions of the type 'How many individuals do you know who are called Jim?' The answers constitute the data, sometimes called the 'Xs,' in the analysis. To not confuse 'Xs' with independent variables, sometimes called X-variables, we refer to the Xs as 'alter categories' in this study.

For example, a respondent is asked how many twins she knows. If she knows three twins, from a total number of 10 million twins in a population of 330,000,000, it would be estimated that she knows about 3/10 million of the population, which is (3/10,000,000) x 330,000,000 = 99. Additionally, when asking respondents about a variety of alter categories – e.g., people known with specific, gender-typed names or who attained specific educational degrees –one can use this information to measure network segregation through alter characteristics inherent to that alter category. Here, we do so as well for gender (i.e., gender-typed names as alter categories) and education (knowing alters in specific educational tracks). By doing so, we follow Jeroense et al. (2024), who similarly measure the ethnic composition of extended networks through NSUM questions on ethnically typed names.

The calculation presented above is often called the "basic scale-up estimator" (e.g., see Killworth et al., 1998). This calculation becomes more precise if more questions of this type are added and when responses are then combined.

Network Scale-up instrument in the DNSS

In the DNSS questionnaire, respondents read the following NSUM-prompt:

In the following, we ask you several questions about persons you know. With "know" we mean persons in the Netherlands you know personally by name and with whom you would have a chat if you met them. For all questions: if you know a person with that feature, report how many persons in your network have that feature.

We survey respondents about two different batteries of alter category items with this prompt. The first battery consists of 9 items about knowing someone who "in the last 12 months" graduated from university, university of applied science, or tertiary vocational education, had a son/daughter, gave birth to twins, was infected with the COVID-19 virus, or knowing someone who owns an electric vehicle, owns a scooter, or is vegan. The second 34 items are about whether and how many others they know carrying a specific *name* (e.g., "Sophie," "Daan," "Noor," and so forth). We selected these names and categories based on figures provided by the Dutch Meertens Institute (in 2023, www.mertens.knaw.nl/nvb/), where the prevalence of first names in the Dutch population register is monitored. To achieve a broad range of names, we selected the most popular name of every fifth birth year between 1950 and 2014 (or the second or third most popular names to avoid repetition of names). Respondents in both batteries indicated first whether they knew someone with that feature (yes/no) and then reported the number of persons that they knew carrying that feature. See Appendix 2 for our list of NSUM categories and their population sizes, and where our known population sizes were sourced from during the time of the survey.

Network size estimation strategy

NSUMs assume that a person's social network represents the population, which is not without challenges. In general, three problems may bias NSUM estimates (McCormick, Salganik and Zheng, 2010). *Transmission errors* occur if a respondent does not know whether a person she knows belongs to a certain group (e.g., not knowing whether someone gave birth). *Barrier effects* occur since networks are not random or a perfect reflection of the population. People know many other people who are similar to themselves, but many fewer who are different (we capitalise on this for our metrics of homogeneity). Finally, *recall errors* are made in very large or very small populations: respondents overestimate the number of alters from smaller populations, but underestimate the numbers from larger populations (Zheng, Salganik, and Gelman, 2006).

By carefully curating a selection of items, we have already attempted to address some of these issues. First, by selecting items that crosscut social groups by age (e.g., the names Cornelis versus Sem, corresponding to names more prevalent in older versus younger age groups), social class (e.g., different educational levels), and ethnic backgrounds (e.g., the name Mohammed), we partly circumvent the issue that some social groups know more others in one versus another category – i.e., circumventing that for some groups networks are considerably larger than for other groups. Second, we have a large battery consisting only of names, likely reducing *transmission errors* as individuals may be more likely to remember names than, for instance, alter' specific educational degrees. In the main analysis of the paper, we present findings based on a network size estimation (detailed later) in which we only use names (though results are qualitatively similar when we use other alter categories as well).

To err on the side of caution, we follow Zheng et al. (2006), who propose Bayesian methods to estimate network size to take account of barrier effects. This procedure includes a random effect to regularise degree estimates. It also incorporates barrier effects via an overdispersion parameter from a negative binomial distribution. The Zheng et al. (2006) model has a built-in indeterminacy that requires renormalisation adjustments. Here, we follow Baum and Marsden (2023) and include all selected subpopulations in the adjustment set (again, the results do not qualitatively change when we choose different alter categories for the renormalisation adjustment).

On which subpopulations do we base our primary network size analyses? We follow McCormick et al. (2010): p. 67) and choose a specific interval for the prevalence of names in the population to minimise recall errors. Specifically, we select 18 alter categories as names that comprise about .08 % and .22 % of the population. We adopt a slightly broader range than McCormick et al.'s recommended interval of .1–.2 %. We do so because our range is the closest percentage range to McCormick and colleagues that ensures a gender-balanced set of names (.1–.2 %: 4 women and 2 men, .08–.22 %: 6 women and 6 men). Different subgroup selections (e.g., in- or exclusion based on the correlation between subpopulation size and average respondent recall, or including all subgroups) do not lead to different results as to our hypothesis tests.

To estimate network size using this procedure, we use the *networkscaleup* R package (Laga, Bao, and Niu, 2023). We refer to their documentation for more details on the estimation procedure. We use a Gibbs-Metropolis algorithm, extract starting values from Killworth's (1998) MLE model, and run 5000 iterations of the algorithm (warmup = 500) from which we obtain an average network size per respondent. We detail several alternative network size estimation scenarios (other renormalisation groups, other subgroup selections) in Appendix 3, none of which substantively change the conclusions of the analyses presented in the main text.

Network homogeneity

To measure gender and educational network homogeneity at the individual level, we use the name and education alter categories from our NSUM instrument. We do not use the NSUM estimation method for network size described above, but instead use elements of the basic scale-up estimator discussed earlier to measure respondent-level homogeneity. We count the number of known alters in total and those that have specific characteristics (i.e., number of women/men names, educational tracks) to calculate the percentage of alters with these characteristics. This is to compare these percentages between respondents. We do so following Jeroense and colleagues' (2024) approach to measuring homogeneity in extended networks.

For gender homogeneity, we first divide the numerical response to each name by that name's population size (to normalise the count for population size in that subgroup). We then sum these numbers for all names and for female and male names separately. We subsequently divide this normalised sum of female (or male) names multiplied by 100 by the normalised sum of all names for the *percentage of women* (or men) in the alter category answers (ranging from 0 to 100; we provide a formal representation in Appendix 4). This also serves to assign the *percentage of same-gender network ties* for women and men respondents.

We normalise by subpopulation size because otherwise, raw counts in larger subgroups disproportionally influence our homogeneity metric. For instance, suppose a respondent knows four men named Daniel out of a subpopulation of 1000,000, two men named Tom out of 100,000, five women named Marieke out of 1000,000, and one woman named Emma out of 100,000. Without normalisation, women would appear to constitute 50 % of the network, as each name carries equal weight. However, knowing individuals from smaller subpopulations provides a stronger indication of exposure to specific group characteristics (here: gender). With normalisation, Emma (female) and Tom (male) – both from smaller subpopulations – receive greater weight as indicators of contact with women and men, respectively. Consequently, the normalised statistic yields a lower percentage of women, 38.5 % in this example.

Normalisation is particularly important in contexts where women's names exhibit greater diversity, resulting in smaller subpopulation sizes, while men's names cluster in fewer, larger groups. This naming pattern is evident in the top-50 newborn names in 2017, where boys' names had larger subpopulations than girls' names in 42 out of 50 cases (Meertens Instituut, 2025). In sum, normalisation by subpopulation size provides a more accurate reflection of actual exposure to men and women in personal networks.

Note also that these metrics are not similar to calculating these percentages based on the *total* number of, for instance, women known in extended networks, which consists of far more alters with far more names, and, consequently, women (or men). We assume the *variation* in the percentage of women known in the survey instrument versus the percentage based on the total number of women in extended networks to be substantively similar, however. Note that our set of NSUM names is larger for women (18) than for men (16): if we drop the female names with either the largest or smallest two subpopulations to render the numbers of name items gender-balanced, this does not qualitatively alter our results in our regression analyses.

For education, we use a similar approach, by using the *number in a university, the number in a university of applied sciences*, and the *number in a secondary vocational education* from the NSUM questions. We sum the number of others known that attained a university or university of applied science degree, together capturing tertiary education in the Netherlands. As such, we know the *number in tertiary education* and the *number in secondary vocational education*. We then calculate the *percentage in tertiary education* and the *percentage in secondary vocational education* by dividing them by the total sum of contacts in all the education levels and multiplying by 100. We also calculate the p*ercentage of same-education* based on this same metric. (Note, we do not normalise for population size here as we ask about the entire population of after-high-school educational categories.)

Independent variables

We construct a vector of independent variables that we then relate to our estimated network size and network homogeneity. Respondents were asked to report their current employment status (e.g., self-employed, retired, working for government, currently a student, working in industry, and so forth). We construct a binary variable that indicates whether respondents WORK (yes/no) (**H1**). We also measure how many people live in the household that the respondent lives in (HOUSEHOLD SIZE, **H2**). We include age in four categories (18–30, 31–45, 46–65, >65) (**H3**). We measure resources by using respondents' income and home value (**H4**). Respondents reported their income in seven intervals in the survey (e.g., minimum income [<€14,100], more than twice the modal income [>€87,100], and so forth). We created a categorical variable, INCOME, that indicated whether respondents earn up to the modal income (1), earn a modal income or more (2), or do not want to report or do not know their income, or are missing altogether (3). We include this last category to not further decrease our case count. As another proxy for wealth, we measured MEAN NEIGHBOURHOOD HOUSE VALUES (Dutch "WOZ-waarde") based on tax records divided by 100,000. We do so by considering the postal codes indicating where they live. To do so, we merge the postal code information of Statistics Netherlands (Statistics, 2021) with respondents' four-digit postal codes available in the survey data. That data contains neighbourhood information on housing (house prices based on tax records) and its population. Finally, we used indicators for biological sex, where respondents reported to be a woman (yes/no, BIOLOGICAL SEX) (**H5**, **H7**). We also create an indicator for whether respondents completed a university degree or a university of applied science degree (bachelor's or master's) (yes/no, TERTIARY EDUCATION) (**H6**, **H8**).

Handling of missing data

Among the N = 1262 respondents that have gave valid responses to all of our NSUM questions (95.2 percent of N = 1325 participating respondents), there are only a few item missings among our independent variables. We therefore opted for listwise deletion, resulting in an analysis sample of 1249 respondents. This is 98.9 % of N = 1262 respondents without missing on our NSUM questions, or 94.3 % of all N = 1325 participating respondents. Table 1 reports the descriptive findings of our independent variables.

Analytical setup

We run a series of regression models of the size and homogeneity of extended networks on our independent variables to test our hypotheses. For network size, these regression models take the form of linear regression models with the log of extended network size as the dependent variables and negative binomial models with unlogged acquaintanceship network size (the latter found in Appendix 5) as dependent variables, as the distributions are right-skewed (see Fig. 1). We mostly discuss the results of the linear regression with logged network size, and only note where there are differences between the linear and negative

Table 1
Descriptive statistics of independent and control variables.

	Mean/Proportion	SD	Min.	Max.	N
Working (H1)	0.53		0	1	1249
Household size (H2)	2.13	1.12	1	8	1249
Age (H3)					
18–30	0.17		0	1	208
31–45	0.18		0	1	230
46–65	0.41		0	1	512
> 65	0.24		0	1	299
Income (H4)					
< = modal	0.50		0	1	627
> modal	0.35		0	1	440
Unknown	0.15		0	1	182
House value (H4)	2.67	0.89	0.95	9.96	1249
Woman (H5, H7)	0.51		0	1	1249
Tertiary education (H6, H8)	0.37		0	1	1249

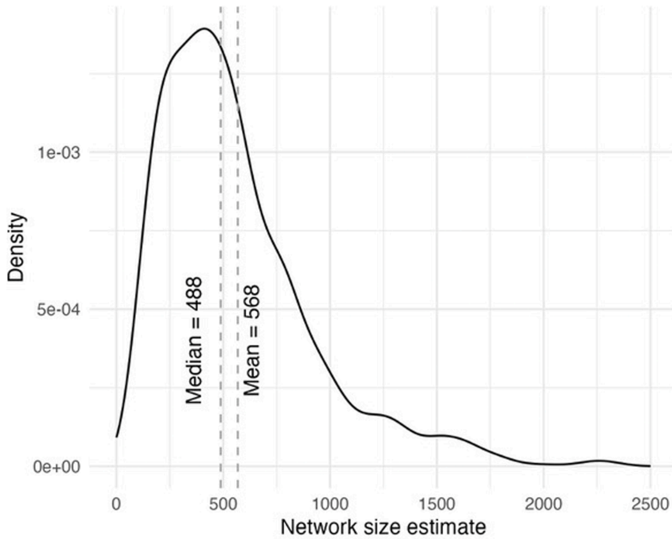

Fig. 1. Distribution of extended network size estimate.

binomial regressions. We drop the five respondents with the largest network sizes (i.e., the extreme outliers). We then run a series of linear regression analyses of the percentage of known others for gender and education. We start with network size and then discuss network homogeneity.

Results

Network size and homogeneity estimates

Fig. 1 depicts the density of our extended network size estimate described above. We find a median network size of 488 and a mean network size of 568 (interquartile range = 428.646).

A much-cited finding in the literature on large-scale networks is that many large networks have heavy-tailed degree distributions. In many cases, such distributions are found to conform to a "power law," indicative of "rich-get-richer" dynamics in the network formation process (Albert and Barabási, 1999), although for extended social networks, a log-normal form is more common (e.g., McCormick et al., 2010; Corten, 2012). A common method to visualise heavy-tailed (degree) distributions is by plotting the complementary cumulative distribution function (CCDF; see, e.g., Barabási, 2016). Fig. 2 plots the CCDF of our degree distribution and shows that the degree distribution in our data is indeed better approximated by a log-normal distribution (the fitted line shown)

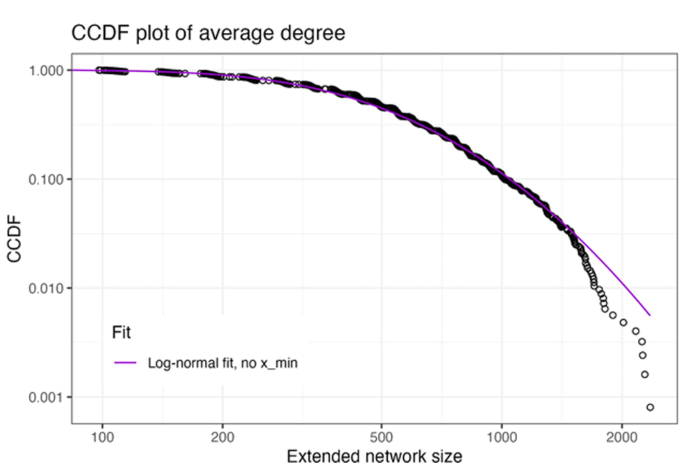

Fig. 2. Complementary cumulative distribution function plot of extended network size, with fitted log-normal curve

in line with previous findings in the literature on extended networks using NSUMs.

Table 2 reports (group-specific) indices for gender and educational homogeneity. Respondents know, on average, more women (61.1 %) than men (38.9 %) as deduced from our NSUM alter categories. Women know more women (65.6 %) than men (34.4 %), while men, surprisingly, know more women (56.3 %) than men (43.7 %). Yet, women know more women than men do, and men know more men than women do, corresponding to our hypotheses. The percentage of same-gender ties based on our alter categories for men is 43.7 %, and for women is 65.6 %. Furthermore, respondents know more others who recently attained degrees in tertiary education (65.2 %) than in secondary vocational education (34.8 %). Yet, those with tertiary education degrees know more in tertiary education (80.9 %) than in secondary vocational education (19.2 %). Those without tertiary education degrees know about as many others in either tertiary (52 %) or secondary vocational (48 %) education, rendering their networks far more heterogeneous. Extended networks of those with tertiary education degrees contain about 81 % of similarly educated, and for those without tertiary education degrees, this is about 48 %. These bivariate results correspond with our hypotheses.

Determinants of extended network size

The regression results for extended network size are found in Table 3, where we present the results for logged network size. We present results for a negative binomial regression in Appendix 5; none of its results differ from those presented. The model presented in Table 3 explains about 11.6 % of the variation in the logged network size outcome.

The following results merit attention. First, and in line with (**H1**), respondents who are employed have significantly larger extended networks than those who are not in paid employment ($p < .001$). If we consider the logged network size results in Table 3, those who work have networks about 18.5 % larger than those who do not work (exp(0.16)= 1.185). Our second hypothesis (**H2**) conjectured that those who live in larger households have larger acquaintanceship networks, which is supported by our data for both the linear and negative binomial regression ($p < .01$). For each additional household member, extended networks grow with about 8.3 % (exp(0.08)= 1.08). Our third hypothesis (**H3**) was that older people have smaller extended networks. This conjecture is also supported by our data. We find that the younger two categories (18–30, 31–45) have networks significantly larger (at least $p < .05$) than those in the oldest one (>65 years). Those aged 18–30 and 31–45 have networks 37.7 % (exp(0.32)= 1.377) and 17.4 % (exp (0.16)= 1.174) larger, respectively, than those aged 65 years or older.

We find mixed evidence for our fourth hypothesis (**H4**) that people who possess more resources have larger extended networks. On the one hand, those with modal incomes or lower (versus higher than modal incomes) do not have smaller network sizes ($p = .066$), while those in neighborhoods with lower house values do have significantly smaller logged extended network sizes ($p < .01$). As such, one finding vis-a-vis resources does not reach statistical significance, whereas the other

Table 2
Gender and educational homogeneity in alter categories.

	% Women	% Men
All respondents	61.1	38.9
Women	65.6	34.4
Men	56.3	43.7
	% Tertiary education	**% Secondary vocational education**
All respondents	65.2	34.8
Tertiary education: yes	80.9	19.2
Tertiary education: no	52.3	47.7

Table 3
Linear regression of (logged) network size on our independent variable.

	Log(Network size)			
	B	SE	95 % CI	p
Intercept	5.61	0.08	5.46 – 5.77	**< 0.001**
Working (**H1**)	0.16	0.04	0.07 – 0.24	**< 0.001**
Household size (**H2**)	0.08	0.02	0.05 – 0.11	**< 0.001**
Age (**H3**)				
18–30	0.32	0.06	0.19 – 0.44	**< 0.001**
31–45	0.16	0.06	0.04 – 0.29	**0.012**
46–65	0.10	0.05	−0.00 – 0.20	0.051
> 65 (ref.)				
Income (**H4**)				
< = modal	−0.08	0.04	−0.16 – 0.01	0.066
> modal (ref.)				
Unknown	−0.08	0.06	−0.19 – 0.04	0.184
House value (**H4**)	0.05	0.02	0.01 – 0.09	**0.008**
Women (**H5**)	0.02	0.04	−0.05 – 0.09	0.634
Tertiary education (**H6**)	0.08	0.04	0.00 – 0.16	**0.049**
Observations	1244			
R^2 / R^2 adjusted	0.123 / 0.116			

does. Notably, this is the only finding where our alternative network size estimation scenarios (explained in Appendix 3) slightly vary: in the other four network size estimation scenarios, higher income does relate to larger extended networks, whereas house value does not. As such, we conclude that we find mixed evidence for hypothesis 4.

We hypothesised (**H5**) that women would have larger extended networks than men. Our findings, however, provide no evidence for this hypothesis ($p = .634$). Finally, we find evidence in support of our sixth hypothesis (**H6**): people with higher educational levels have more extended networks. Specifically, we find that those with tertiary education degrees have networks about 8.3 % (exp(0.08)= 1.083) larger than those without higher education degrees.

Determinants of gender and educational homogeneity

Next, we discuss the results for the hypotheses on homogeneity in extended networks. Table 4 presents the findings for gender. We hypothesised (**H7**) that in extended networks, women have more ties to women than men have, and vice versa. Two findings merit attention here. The first panel of results shows that indeed women indicate that they know more women, and men know more men in our NSUM alter categories. Women know about 9 percentage points more women than men do ($p < .001$). And, likewise, men indicate to know about 9 percentage points more men than women do (i.e., the positive “Women (H7)” coefficient would simply flip to a minus sign if we were to regress the percentage of men on our covariates). Interestingly, women have a significantly higher percentage of same-gender ties than men in our NSUM survey instrument. Women’s alters consist of 22 percentage points more same-gender ties than men, or, phrased differently, men know about 22 percentage points fewer men than women know women. In sum, this supports our seventh hypothesis. (Note that the number of respondents in these homogeneity analyses is lower than in Table 1, as knowing at least someone on the NSUM items used is a prerequisite to calculating a homogeneity score.)

We further hypothesised that higher educated individuals have more ties to higher educated individuals than lower educated individuals have in extended networks (**H8**), and vice versa. Table 5 presents the findings to test this hypothesis. The first panel shows that those with tertiary education know more others with recent degrees in tertiary education than those without tertiary education: about 25 percentage points more ($p < .001$). Vice versa, those without tertiary education degrees know about 25 percentage less with recent tertiary education degrees. The second panel shows that those with tertiary education have about 30 percentage points more same-education ties than those without tertiary education degrees ($p < .001$). These findings provide evidence for Hypothesis 8.

Conclusions and discussion

This study aimed to contribute to an understanding of the size and the homogeneity of personal networks. While previous research has primarily focused on these dimensions within smaller strong-tie networks, less is known about the size and composition of extended networks, which include not only core ties but also the much larger number of acquaintances. We provide novel evidence on the size of extended networks and their gender and educational homogeneity, and the variation in these dimensions. Specifically, we examine how individual characteristics, such as employment, household size, and educational level, explain why some people have larger or more homogeneous extended networks. Drawing on a representative survey of the Dutch adult population, we employ the Network Scale-Up Method (NSUM), utilising a large number of alter categories as scale-up items. These items include questions about knowing people with specific (men or women) names, as well as knowing individuals who use electric cars, ride scooters, or have recently obtained a tertiary degree. Four conclusions can be drawn from the results.

First, we find a median network size of 488 and a mean network size of 568. These figures are slightly lower than those reported in earlier studies conducted in the United States, which found a median network size of 610 (McCormick et al., 2010) and 550 (DiPrete et al., 2011), as

Table 4
Linear regression of percentage of women (percentage of men is the inverse of the percentage of women result) and percentage of same-gender on covariates.

	% Women				% Same-gender			
	B	SE	95 % CI	p	B	SE	95 % CI	p
Intercept	49.83	3.04	43.85 – 55.80	**< 0.001**	42.44	3.05	36.45 – 48.44	**< 0.001**
Working	−0.45	1.65	−3.69 – 2.80	0.787	1.81	1.66	−1.44 – 5.06	0.275
Household size	0.95	0.62	−0.27 – 2.18	0.127	0.06	0.63	−1.17 – 1.29	0.923
Age	49.83	3.04	43.85 – 55.80					
18–30	2.40	2.39	−2.29 – 7.08	0.316	1.10	2.40	−3.61 – 5.80	0.647
31–45	1.07	2.46	−3.75 – 5.89	0.662	0.68	2.46	−4.16 – 5.51	0.784
46–65	1.67	1.98	−2.21 – 5.54	0.399	1.95	1.98	−1.94 – 5.84	0.326
> 65 (ref.)								
Income								
< = modal	1.62	1.62	−1.55 – 4.80	0.317	−1.52	1.62	−4.70 – 1.67	0.351
> modal (ref.)								
Unknown	1.98	2.15	−2.24 – 6.20	0.357	−1.04	2.16	−5.27 – 3.20	0.631
House value	0.32	0.76	−1.17 – 1.82	0.673	−0.12	0.76	−1.62 – 1.38	0.877
Women (**H7**)	9.09	1.34	6.47 – 11.71	**< 0.001**	22.23	1.34	19.60 – 24.86	**< 0.001**
Tertiary education (**H8**)	4.15	1.50	1.21 – 7.10	**0.006**	0.41	1.51	−2.54 – 3.36	0.785
Observations	1238				1238			
R^2 / R^2 adjusted	0.052 / 0.044				0.190 / 0.184			

Table 5
Linear regression of percentage in tertiary education (percentage in secondary vocational is the inverse of the percentage in tertiary in education result) and percentage of same-education on covariates.

	% Tertiary education						% Same-education					
	B	SE	95 % CI			*p*	B	SE	95 % CI			*p*
Intercept	57.38	6.60	44.43	–	70.33	**< 0.001**	34.63	6.69	21.50	–	47.76	**< 0.001**
Working	−3.24	3.39	−9.90	–	3.42	0.340	6.49	3.44	−0.26	–	13.24	0.060
Household size	−3.75	1.27	−6.25	–	−1.25	**0.003**	−0.32	1.29	−2.86	–	2.21	0.803
Age												
18–30	8.75	4.94	−0.95	–	18.45	0.077	6.56	5.01	−3.27	–	16.40	0.190
31–45	−0.79	5.59	−11.77	–	10.19	0.887	7.37	5.67	−3.77	–	18.50	0.194
46–65	−5.54	4.67	−14.71	–	3.63	0.236	1.91	4.73	−7.39	–	11.20	0.687
> 65 (ref.)												
Income												
< = modal	−4.01	3.36	−10.62	–	2.60	0.234	1.44	3.41	−5.25	–	8.14	0.672
> modal (ref.)												
Unknown	−4.61	4.70	−13.84	–	4.62	0.327	−3.56	4.76	−12.91	–	5.79	0.455
House value	3.11	1.55	0.07	–	6.15	**0.045**	2.11	1.57	−0.98	–	5.19	0.180
Women	2.34	2.81	−3.19	–	7.86	0.407	3.20	2.85	−2.40	–	8.80	0.262
Tertiary education (**H8**)	24.71	3.07	18.68	–	30.75	**< 0.001**	30.09	3.11	23.98	–	36.21	**< 0.001**
Observations	609						609					
R^2 / R^2 adjusted	0.023 / 0.015						0.206 / 0.193					

well as in Spain, where the median was 536 (Lubbers et al., 2019). However, our findings are higher than those from Chile, which had a median network size of 288 (Contreras et al., 2019). Higher numbers were observed in two earlier studies in the Netherlands. Hofstra et al. (2021), who relied on data among youth in the Netherlands, found a median network size of 892, while Jeroense et al. (2024), using data on the Dutch population between 16 and 45 years of age, found a median network size of 778.

Second, we find considerable variation between people in their network size. In line with the literature on extended networks, degrees in our data show a heavy-tailed distribution, best approximated by a log-normal distribution. We presented results for one main network size estimate but ran five estimation scenarios that either included or excluded some alter categories (exclude a *very* large subpopulation, only include small subpopulations, adjusting for over- and underrecall, and so forth). Although the point estimates for network size varied quite considerably, the correlations between the scenarios were high (at least $r = .89$). And most importantly, the five network size estimation scenarios did not matter for the tests of our hypotheses. As such, the considerable variation we find between people in their extended network sizes seems insensitive to our five estimation procedures.

Third, in line with expectations, we find that larger extended networks are associated wiabout 8 employed (H1), having more household members (H2), being younger (H3), possessing greater resources such as income and wealth (H4, either house value or income in our sensitivity analyses) and having higher levels of education (H6). The hypothesis that women have larger networks than men (H5) was not confirmed. Our findings on being employed, education, and socioeconomic status (here, however, in terms of house value rather than income) largely align with other literature on correlates of extended network size (see Background section), whereas household size has not been previously studied. In contrast to our study, Jeroense et al. (2024) found no statistically significant difference by education and age in their study on the size of extended networks in the Netherlands. Overall, the findings from our study generally align with the mechanisms of tie formation and maintenance, which highlight the importance of opportunities to establish connections, as well as factors such as homophily, preferential attachment, and individual abilities. Perhaps the most striking finding is that employment is linked to having networks at least 19 % larger compared to being unemployed. This underscores the significance of the workplace as a setting that facilitates the expansion of personal networks.

Additionally, each extra household member is associated with a growth in network size of about 8 %, while being between 18 and 30 and 31–45 years of age is associated with networks that are, respectively, 38 % and 17 % larger than the networks of people who are 65 years or older. The first finding suggests that household members may serve as important bridges through which additional network ties are formed. Furthermore, individuals with tertiary education have networks about 8 % larger compared to those without tertiary education. This difference may be attributed to their greater cognitive and physical abilities, as well as their higher levels of activity and participation in various social settings. Higher income and living in a wealthier neighbourhood are also associated with having more network members, yet only in one set of analyses. These findings regarding differences in the size of extended networks across social categories offer intriguing insights into the social stratification of network ties.

Fourth, we find evidence that extended networks are segregated in terms of gender and education. Specifically, we observe that same-gender (H7) and same-education (H8) ties are more prevalent in extended networks than different-gender and different-education ties. Interestingly, women have a significantly higher percentage of same-gender ties in their extended networks compared to men. Women's extended networks consist of 22 % more same-gender ties than men's networks, or, phrased differently, men know about 22 % fewer men than women know women. Regarding education, we find that individuals with higher education tend to have fewer network ties outside their highly educated circles, whereas those with lower education have a more educationally diverse network. Van der Laan et al. (2023), using population-wide data in the Netherlands, report considerable educational segregation in terms of meeting opportunities created by neighbourhoods, schools, family- and work settings. Here, we find that extended networks are also segregated by education, where especially higher educated individuals are predominantly surrounded by similarly educated others.

These findings add to previous literature, which, so far, has primarily focused on the homogeneity of strong-tie networks. This field has shown that core discussion networks are quite homogeneous in terms of gender, education, and ethnicity/race. Our findings reveal that even in larger networks, which include not only friendships and other strong ties but also a far greater number of acquaintance ties, homogeneity remains evident. Among the few studies on the homogeneity of extended networks, DiPrete et al. (2011) report similar findings for the U.S., noting significant segregation in acquaintance networks by race, religion, and ideology. This segregation in extended networks may be driven by a combination of homophily processes and segregated activity spaces or social foci.

Despite these important findings, our study has several limitations that could stimulate future research. First, we were unable to measure the diversity of work settings. Future studies could explore how the characteristics of workplaces, such as their level of diversity, influence the network ties formed in these environments, thus more precisely capturing opportunity structures. Second, our reliance on cross-sectional data limits our ability to draw firm conclusions about causal relationships. The absence of longitudinal data in the field of extended networks is a significant gap, and such data would be invaluable not only for establishing causality but also for examining how the size and homogeneity of extended networks evolve. For instance, it would be intriguing to examine the extent to which the observed age effect is influenced by factors such as higher rates of divorce among older individuals (adjusting, of course, for time at 'risk'). Third, our study did not account for tie strength. When identifying a network member based on specific characteristics, these alters could represent either strong or weak ties. Future research could study how network size correlates with the strength of ties within individuals' networks. Fourth, our network size estimate did not adjust for transmission error – i.e., respondents are unaware of alter characteristics – in our network size estimations. Yet, as we base our main results on a network size estimate based on names, transmission errors may have been less of an issue. Given that our substantive results are quite similar across sensitivity analyses that include NSUM items other than names, we carefully assume that transmission errors were low in these cases as well. Yet, we commend future work that accounts for transmission error.

Finally, in some cases, the prevalence of alter categories in a population might be too rough a benchmark for scaling up. This is because the use of names or certain outcomes (e.g., attending university, eating vegan) might differ starkly on regional or even local levels. On the other hand, in the era of the internet, our access to others is less geographically bound than ever (yet network niches may result from strong choice options on the internet). Although our dispersion parameters in our network size estimation should partly account for this, future research could inquire into these regional differences as well as into the various tie strengths in the networks. For instance, if individuals mention knowing a person in a given alter category, one could add a scale for the relationship strength to that alter. After all, strong ties are the ties that provide more support, but weak ties provide new information. It would be intriguing to understand how the share of stronger and weaker ties differs across social categories and how this is related to success in society.

All in all, our study highlights the significant differences in both the size and homogeneity of extended networks in terms of gender and education and is the first to do so for the full Dutch adult population. To conclude, this underscores the fruitfulness of the NSUM approach in revealing such differences, as well as the ability of the choice-constraint framework to explain these differences more consistently.

CRediT authorship contribution statement

Völker Beate: Writing – review & editing, Writing – original draft, Project administration, Investigation, Conceptualization. **Bas Hofstra:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **Rense Corten:** Writing – review & editing, Writing – original draft, Methodology, Formal Analysis, Data curation, Conceptualization. **Frank van Tubergen:** Writing – review & editing, Writing – original draft, Conceptualization.

Acknowledgements

We thank the anonymous reviewers for their thorough and helpful comments.

Appendix 1. - Questionnaire: How large is your network?

Thank you for participating in this survey that I&O is conducting in cooperation with the science magazine QUEST and Utrecht University. You will be asked several questions about positions and first names. We will use your answers to estimate the size of your network.

We also ask you to provide some basic information about your origins and health. These data are used only to estimate your network. They will be used only for this survey and processed completely anonymously. Data that could be combined to trace your identity will not be released by I&O.

Below you can fill in that you agree to the processing of these data.

☐ I agree to the processing of these data.

In the following, we ask some questions about people you know. By 'know' we mean people who live in the Netherlands, whom you know personally, by name and with whom you can have a chat if you meet. For all questions, if you know a person with the characteristic in question, please indicate how many people in your network have this characteristic.

Question 1: Do you know anyone in the last year, in 2020, who…. (please do not include yourself)

By 'know' we mean people who live in the Netherlands, whom you know personally, by name and with whom you could have a chat if you met.

Yes/no/no answer number

1 Is a university graduate (bachelor's/master's degree)?

2 Has obtained an HBO diploma

3 Obtained an intermediate vocational (MBO) degree

4 Has given birth to a daughter/son

5 Has given birth to twins

6 Bought a house in one of the four major cities (Amsterdam, Rotterdam, The Hague, Utrecht)

7 Tested positive for the coronavirus.

Question 2: Do you know anyone….(please do not count yourself)

By 'know' we mean people who live in the Netherlands, whom you know personally, by name and could have a chat with if you met.

Yes/no/no answer number

1 Drive an electric car

2 Rides a scooter

3 Eats vegan (i.e. no meat, fish, dairy, eggs)

Question 3: Do you know anyone with the following name (your own name does not count) and, if yes, how many persons do you know?

By 'know' we mean people who live in the Netherlands, whom you know personally, by name and with whom you could have a chat if you met.

Name Yes/no/no answer number Name Yes/no/no answer number

Name	Yes/no/no answer	number	Name	Yes/no/no answer	number
Sophie			Daan		
Julia			Sem		
Sanne			Thomas		
Lisa			Max		
Laura			Kevin		
Maria			Johannes		
Linda			Dennis		
Johanna			Jeroen		
Monique			Jan		
Ester			Marcel		
Anna			Cornelis		
Elisabeth			Hendrik		
Cornelia			Petrus		
Wilhelmina			Willem		
Amira			Ali		
Samira			Mohammed		
Sara			Noor		

Question 4:
Have you tested positive for coronavirus since the beginning of the pandemic? yes/no/no answer
Question 5:
What country were your parents born in?
Father:
Netherlands Elsewhere, specify
no answer
Mother:
Netherlands, Elsewhere, specify:
no answer
Question 6:
Do you drive an electric car?
Question 7:
Do you have a scooter which you can use as a means of transportation?
Question 8:
Do you eat vegan (no meat, fish, cheese, eggs)?
(Questions 6–8: Anywhere: yes/no/no answer)
This was the last question. Many thanks for your participation!

Appendix 2. Alter Categories and Their Population Sizes

Table A1
Alter categories and their associated population sizes in the Netherlands, 2021.[a]

Alter category	Population	Source	Alter category	Population	Source
1 University	84957	CBS	23 Wilhelmina	98208	Meertens
2 Univ. of Applied Science	75214	CBS	24 Amira	1386	Meertens
3Secondary Vocational	145600	CBS	25 Samira	2186	Meertens
4 Daughter/Son	168066	CBS	26 Sara	11640	Meertens
5 Twins	2500	CBS	27 Daan	22704	Meertens
6 COVID−19 Infection	1558549	RIVM	28 Sem	13276	Meertens
7 Electric car	273259	CBS	29 Thomas	40543	Meertens
8 Scooter	460618	CBS	30 Max	17024	Meertens
9 Vegan	261000	CBS	31 Kevin	23167	Meertens
10 Sophie	15276	Meertens	32 Johannes	307032	Meertens
11 Julia	16350	Meertens	33 Dennis	36411	Meertens
12 Sanne	27394	Meertens	34 Jeroen	49182	Meertens
13 Lisa	21200	Meertens	35 Jan	186746	Meertens
14 Laura	25681	Meertens	36 Marcel	35973	Meertens
15 Maria	334502	Meertens	37 Cornelis	134956	Meertens
16 Linda	29955	Meertens	38 Hendrik	118610	Meertens
17 Johanna	266522	Meertens	39 Petrus	86500	Meertens
18 Monique	39481	Meertens	40 Willem	102296	Meertens
19 Ester	2692	Meertens	41 Ali	4213	Meertens
20 Anna	136296	Meertens	42 Mohammed	5003	Meertens
21 Elisabeth	110231	Meertens	43 Noor	4517	Meertens
22 Cornelia	112807	Meertens			

[a] CBS = Statistics Netherlands, Meertens = Meertens Institute First Name Bank.
RIVM= Netherlands Institute for Public Health and the Environment

Appendix 3. Alternative Network Size Estimation Scenarios

To ascertain that our results are not influenced by either the renormalization groups or subgroup selection, we estimate network size in four alternative specifications to the one presented in the main text.

Subgroup selection

NSUM network sizes vary depending on the selected alter categories and associated under- and overrecall. In Fig. A1, we first plot all subpopulation sizes (x-axes) against the average recall (y-axis) of respondents. We observe that the high correlation in panel A ($r = .91$) seems mostly driven by the large subpopulation of individuals who carried COVID-19 and the high average reporting of respondents to know others in that subpopulation. In the second panel B we drop the COVID-19 subpopulation (i.e., selecting the <500 K subpopulation sizes) to observe how this affects the correlation and find that the correlation drops to $r = .34$. Especially by selecting the smaller subpopulations (i.e., names) do we attain a high correlation ($r = .89$) between the average mentions of the respondents on these NSUM items and the subpopulation size (panel C). Do these three different subgroup selections and associated correlations between subpopulation size and average respondent answers matter for our hypotheses on network size? When we estimate network size (similar to the Zheng et al., 2006 model, with an overdispersion parameter, and including all groups for renormalization) we find that regressing these network size estimates on our covariates hardly matters for our hypotheses' tests (and in the single case it does, we explained it in the main text).

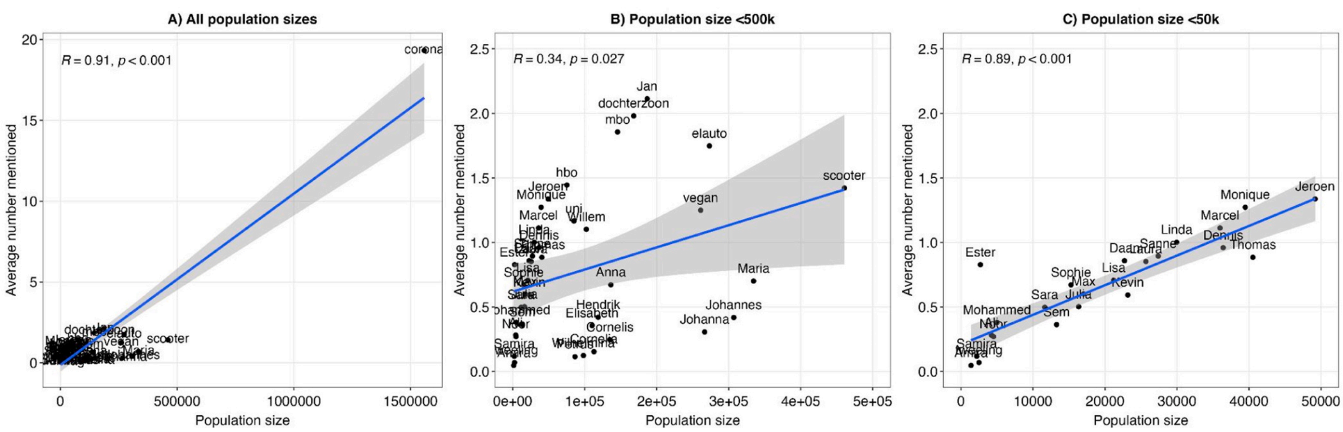

Figure A1. Average number of items mentioned by the different population sizes

Renormalization groups

We see that there is quite some deviation from the correlation line for some subgroups in Fig. A1, Panel B. We also estimate network size by using some of these as renormalisation groups. Specifically, we do so for "HBO" (i.e., university of applied sciences), "MBO" (i.e., secondary vocational education), "having a child", "electric cars", "vegan", and "Jan" as one adjustment group (overrecalled), and "Johanna", "Johannes", and "Maria" as another adjustment group (underrecalled). Here too, our results do not qualitatively differ compared with of our hypothesis tests.

Though the median and average network size deviate in these five scenarios (the main one presented in the paper, and the four described here) correlations between these five different estimation scenarios are at least larger than $r = .89$.

Appendix 4. Definition of Gender Homogeneity

If n_{ij} is the numerical response by respondent i to each name and P_j is the population size of name j, we can normalize each name response by its population size:

$$n'_{ij} = \frac{n_{ij}}{P_j}$$

If W is the set of women's names and M is the set of men's names and if we then sum over all names, we obtain the normalized number of women, men, and total names for each respondent:

$$N'_m = \sum_{j\in W,i} n'_{ij}$$

$$N'_m = \sum_{j\in M,i} n'_{ij}$$

$$N'_{total} = \sum_{j,i} n'_{ij}$$

And, finally, we compute the relative percentage of women and men for each respondent:

$$H'_w = \frac{N'_w}{N'_{total}} \times 100$$

$$H'_m = \frac{N'_m}{N'_{total}} \times 100$$

Both the relative p*ercentage of women* H'_w and the r*elative percentage of men* H'_m also serve as the *percentage of same-gender network ties* for women and men respondents, respectively.

Appendix 5. Negative Binomial Regression of Network Size

Table A1
Negative binomial regression of network size on our independent variable

	Network size			
	IRR	SE	CI	*p*
Intercept	347.17	26.57	298.37 – 404.03	**< 0.001**
Working (**H1**)	1.14	0.05	1.05 – 1.23	**0.002**
Household size (**H2**)	1.07	0.02	1.04 – 1.11	**< 0.001**
Age (**H3**)				
18–30	1.40	0.08	1.24 – 1.57	**< 0.001**
31–45	1.19	0.07	1.06 – 1.34	**0.005**
46–65	1.11	0.06	1.01 – 1.23	**0.031**
> 65 (ref.)				
Income (**H4**)				
< = modal	0.93	0.04	0.86 – 1.01	0.085
> modal (ref.)				
Unknown	0.96	0.05	0.86 – 1.07	0.447
House value (**H4**)	1.05	0.02	1.01 – 1.09	**0.019**
Women (**H5**)	0.99	0.03	0.93 – 1.06	0.813
Tertiary education (**H6**)	1.08	0.04	1.00 – 1.16	**0.044**
Observations	1244			
R^2 adjusted	0.176			

References

Albert, R., Barabási, A.-L., 1999. Emergence of scaling in random networks. Science 286 (5439), 509–512.
Arenson, M., Bernat, E., De Los Reyes, A., Neylan, T.C., Cohen, B.E., 2021. Social support, social network size, and suicidal ideation: a nine-year longitudinal analysis from the mind your heart study. J. Psychiatr. Res. 135, 318–324.
Arnaboldi, V., Conti, M., Passarella, A., Dunbar, R.I.M., 2017. Online social networks and information diffusion: the role of ego networks. Online Soc. Netw. MedMac Carron 1, 44–55 (ia).
Barabási, A.L., Albert, R., 1999. Emergence of scaling in random networks. Science 286 (5439), 509–512.
Barabási, A.-L., 2016. Network science. Cambridge University Press.
Baum, D.S., Marsden, P.V., 2023. Uses and limitations of dichotomous aggregate relational data. Soc. Netw. 74, 42–61.
Bernard, H.R., Johnsen, E.C., Killworth, P.D., McCarty, C., Shelley, G.A., Robinson, S., 1990. Comparing four different methods for measuring personal social networks. Soc. Netw. 12 (3), 179–215.
Bernard, R.H., Johnsen, E., Killworth, P., Robinson, S., 1991. Estimating the size of an average personal network and of an event subpopulation: some empirical results. Soc. Sci. Res. 20 (2), 109–121.
Bidart, C. (2020). *Living in Networks. The Dynamics of Social Relationships*. Cambridge University Press. Cambridge, UK.
Blau, P.M., 1977. Inequality and heterogeneity: A primitive theory of social structure, 7. Free Press, New York, pp. 677–683.
Brashears, M.E., Hoagland, E., Quintane, E., 2016. Sex and network recall accuracy. Soc. Netw. 44, 74–84.
Clay-Warner, J., Kawashima, T., Edgemon, T.G., 2022. Measure of personal network size using the known population method: a methodological guide. Am. J. Public Health 112 (5), 747–753.
Contreras, D., Otero, G., Díaz, J.D., Suárez, N., 2019. Inequality in social capital in Chile: assessing the importance of network size and contacts' occupational prestige on status attainment. Soc. Netw. 58, 59–77.
Cornwell, B., Schumm, L.P., Laumann, E.O., Kim, J., Kim, Y.J., 2014. Assessment of social network change in a national longitudinal survey. J. Gerontol. Ser. B Psychol. Sci. Soc. Sci. 69 (_2), S75–S82.
Corten, R., 2012. Composition and structure of a large online social network in the Netherlands. PLoS ONE 7 (4).
DiPrete, T.A., Gelman, A., McCormick, T., Teitler, J., Zheng, T., 2011. Segregation in social networks based on acquaintanceship and trust. Am. J. Sociol. 116 (4), 1234–1283.
Dodds, P.S., Muhamad, R., Watts, D.J., 2003. An experimental study of search in global networks. Science 301, 827–879.
Domènech-Abella, J., Mundó, J., Haro, J.M., Rubio-Valera, M., 2019. Anxiety, depression, loneliness and social network in the elderly: longitudinal associations from the Irish longitudinal study on ageing (TILDA). J. Affect. Disord. 246, 82–88.
Dunbar, R.I.M., Arnaboldi, V., Conti, M., Passarella, A., 2015. The structure of online social networks mirrors those in the offline world. Soc. Netw. 43, 39–47.
Feehan, D.M., Salganik, M.J., 2016. Generalizing the network scale-up method: a new estimator for the size of hidden populations. Sociol. Methodol. 46 (1), 153–186.
Feld, S.L., 1981. The focused organization of social ties. Am. J. Sociol. 86 (5), 1015–1035.
Fischer, C.S., Jackson, R.M., Stueve, C.A., Gerson, K., McCallister Jones, L., with, Baldassare, M., 1977. Networks and places: social relations in the urban setting. Free Press, New York.
Freeman, L. & C. Thompson (1989) Estimating acquaintanceship volume. Pp1470158 in M. Kochen (ed) The Small World. Norwood, NJ:Ablex.
Fu, Y., 2005. Measuring personal networks with daily contacts: a single-item survey question and the contact diary. Soc. Netw. 27 (3), 169–186.
Gurevitch, M. (1961). *The social structure of acquaintanceship networks (Doctoral dissertation, Massachusetts Institute of Technology)*.
Hill, R.A., Dunbar, R.I., 2003. Social network size in humans. Hum. Nat. 14 (1), 53–72.
Hofstra, B., Corten, R., Van Tubergen, F., Ellison, N.B., 2017. Sources of segregation in social networks: a novel approach using facebook. Am. Sociol. Rev. 82 (3), 625–656.
Hofstra, B., Corten, R., van Tubergen, F., 2021. Beyond the core: who has larger social networks? Soc. Forces 99 (3), 1274–1305.
Jeroense, T., Hofstra, B., Spierings, N., Tolsma, J., 2024. Size and ethnic homogeneity of extended social networks in the Netherlands: differences between migrant groups and migrant generations. Int. Migr. https://doi.org/10.1111/imig.13252 (online first).
Kadushin, C., 2012. Understanding social networks: theories, concepts, and findings. Oxford University Press.
Kadushin, C., Jones, D.J., 1992. Social networks and urban neighborhoods in New York city. City Soc. 6 (1), 58–75.
Kadushin, C., Killworth, P., Bernard, H., Beveridge, A., 2006. Scale-up methods as applied to estimates of heroin use. J. Drug Issues 36 (2), 417.
Killworth, P.D., Johnsen, E.C., Bernard, H.R., Shelley, G.A., McCarty, C., 1990. Estimating the size of personal networks. Soc. Netw. 12 (4), 289–312.
Killworth, P.D., McCarty, C., Bernard, H.R., Shelley, G.A., Johnsen, E.C., 1998. Estimation of seroprevalence, rape, and homelessness in the United States using a social network approach. Eval. Rev. 22 (2), 289–308.
Laga, I., Bao, L., Niu, X., 2023. A correlated network scale-up model: finding the connection between subpopulations. J. Am. Stat. Assoc. 118 (543), 1515–1524.
Laugesen, K., Baggesen, L.M., Schmidt, S.A.J., Glymour, M.M., Lasgaard, M., Milstein, A., Ehrenstein, V., 2018. Social isolation and all-cause mortality: a population-based cohort study in Denmark. Sci. Rep. 8 (1), 4731.
Letki, N., Mieriņa, I., 2015. Getting support in polarized societies: income, social networks, and socioeconomic context. Soc. Sci. Res. 49, 217–233.
Lubbers, M.J., Molina, J.L., Valenzuela-García, H., 2019. When networks speak volumes: variation in the size of broader acquaintanceship networks. Soc. Netw. 56, 55–69.
Mac Carron, P., Kaski, K., Dunbar, R., 2016. Calling dunbar's numbers. Soc. Netw. 47, 151–155.
Marsden, P.V., 1987. Core discussion networks of Americans. Am. Sociol. Rev. 122–131.

McCarty, C., Killworth, P.D., Bernard, H.R., Johnsen, E.C., Shelley, G.A., 2001. Comparing two methods for estimating network size. Hum. Organ. 60 (1), 28–39.
McCormick, T.H., Salganik, M.J., Zheng, T., 2010. How many people do you know?: efficiently estimating personal network size. J. Am. Stat. Assoc. 105 (489), 59–70.
McPherson, M., Smith-Lovin, L., Cook, J.M., 2001. Birds of a feather: homophily in social networks. Annu. Rev. Sociol. 27 (1), 415–444.
McPherson, M., Smith-Lovin, L., Brashears, M.E., 2006. Social isolation in america: changes in core discussion networks over two decades. Am. Sociol. Rev. 71 (3), 353–375.
Meertens Instituut. (2025). Nederlandse Voornamenbank: Topnamen land Nederland 2017. Accessed on March 12, 2025 (https://nvb.meertens.knaw.nl/topnamen/land/Nederland/2017).
Milgram, S., 1967. The small world problem. Psychol. Today 22, 61–67.
Mollenhorst, G., Völker, B., Flap, H., 2008b. Social contexts and core discussion networks: using a choice approach to study similarity in intimate relationships. Soc. Forces 86 (3), 937–965.
Mollenhorst, G., Völker, B., Flap, H., 2008a. Social contexts and personal relationships: the effect of meeting opportunities on similarity for relationships of different strength. Soc. Netw. 30 (1), 60–68.
Moore, Gwen, 1990. "Structural determinants of men's and women's personal networks. Am. 7ociological Rev. 726–735.
Mouw, T., Entwisle, B., 2006. Residential segregation and interracial friendship in schools. Am. J. Sociol. 112 (2), 394–441.
Paik, A., Sanchagrin, K., 2013. Social isolation in america: an artifact. Am. Sociol. Rev. 78 (3), 339–360.
Parigi, P., Henson, W., 2014. Social isolation in america. Annu. Rev. Sociol. 40, 153–171.
Penninx, B.W., Van Tilburg, T., Kriegsman, D.M., Boeke, A.J.P., Deeg, D.J., Van Eijk, J.T.M., 1999. Social network, social support, and loneliness in older persons with different chronic diseases. J. Aging Health 11 (2), 151–168.
Schutter, N., Holwerda, T.J., Comijs, H.C., Stek, M.L., Peen, J., Dekker, J.J., 2022. Loneliness, social network size and mortality in older adults: a meta-analysis. Eur. J. Ageing 19 (4), 1057–1076.
Smith, J.A., McPherson, M., Smith-Lovin, L., 2014. Social distance in the United States: sex, race, religion, age, and education homophily among confidants, 1985 to 2004. Am. Sociol. Rev. 79 (3), 432–456.
Statistics Netherlands 2021: Geographical data per pstal code. https://www.cbs.nl/nl-nl/dossier/nederland-regionaal/ visited September 2023.
Stokes, J.P., 1983. Predicting satisfaction with social support from social network structure. Am. J. Community Psychol. 11 (2), 141.
Vacca, R., Solano, G., Lubbers, M.J., Molina, J.L., McCarty, C., 2018. A personal network approach to the study of immigrant structural assimilation and transnationalism. Soc. Netw. 53, 72–89.
Van der Laan, J., de Jonge, E., Das, M., Te Riele, S., Emery, T., 2023. A whole population network and its application for the social sciences. Eur. Sociol. Rev. 39 (1), 145–160.
Van Tubergen, F., 2014. Size and socio-economic resources of core discussion networks in the Netherlands: differences by national-origin group and immigrant generation. Ethn. Racial Stud. 37 (6), 1020–1042.
Van Tubergen, F., Völker, B., 2015. Inequality in access to social capital in the Netherlands. Sociology 49 (3), 521–538.
Van Tubergen, F., Al-Modaf, O.A., Almosaed, N.F., Al-Ghamdi, M.B.S., 2016. Personal networks in Saudi Arabia: the role of ascribed and achieved characteristics. Soc. Netw. 45, 45–54.
Wang, X., 2016. Subjective well-being associated with size of social network and social support of elderly. J. Health Psychol. 21 (6), 1037–1042.
Wimmer, A., Lewis, K., 2010. Beyond and below racial homophily: ERG models of a friendship network documented on facebook. Am. J. Sociol. 116 (2), 583–642.
Wrzus, C., Hänel, M., Wagner, J., Neyer, F.J., 2013. Social network changes and life events across the life span: a meta-analysis. Psychol. Bull. 139 (1), 53.
Zheng, T., Salganik, M.J., Gelman, A., 2006. How many people do you know in prison? Using overdispersion in count data to estimate social structure in networks. J. Am. Stat. Assoc. 101, 409–423.