Social Networks 84 (2026) 78–86

Contents lists available at ScienceDirect

Social Networks

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

Collecting a large number of alters in egocentric network research: A comparative analysis of three approaches

Miguel A. González-Casado [a,*], Alejandro Cruzado Rey [b], Miroslav Pulgar Corrotea [c], Christopher McCarty [d], José Luis Molina [c], Angel Sánchez [a,e]

[a] Grupo Interdisciplinar de Sistemas Complejos (GISC), Departamento de Matemáticas, Universidad Carlos III de Madrid, Leganés, 28911, Spain
[b] Universidad Loyola Andalucía, Sevilla, 41704, Spain
[c] Grupo de investigación en Antropología Fundamental y Orientada (GRAFO), Departamento de Antropología social y cultural, Universitat Autònoma de Barcelona, Bellaterra, 08173, Spain
[d] Department of Anthropology, University of Florida, Gainesville, 32605, USA
[e] Instituto de Biocomputación y Física de Sistemas Complejos (BIFI), Universidad de Zaragoza, Zaragoza, 50018, Spain

ARTICLE INFO

Dataset link: this Github repository

Keywords:
Name generators
Personal networks
Egocentric networks
Network structure
Network variability

ABSTRACT

This article presents an analysis of the impact of the number of alters elicited in an ego network on the structural properties of those networks. There continues to be debate about the pros and cons of eliciting a fixed number of alters for each respondent versus allowing the respondent to list as many or few alters as they would like. This article explores a random assignment of respondents to three treatment groups – (1) a fixed number of alters set at 30, (2) a variable number of alters up to 45, and (3) a variable number of alters up to 45 with a 20 alter minimum. The results indicate that, from a non-structural perspective, all levels of emotional proximity, interaction contexts, genders, and ages are consistently sampled across the three treatment groups. At the structural level, the behavior of individual metrics is also largely similar. However, the most significant differences arise in the collective behavior of structural metrics—specifically, in their correlation structure, the amount of redundant information each variable provides, and the diversity and interpretability of the observed structural variability. When a data collection strategy constrains network size, it reduces the sparsity of the correlation matrix, effectively decreasing the number of independent global variables needed to describe network structure and making these global variables less interpretable. In other words, networks constructed with a survey that limits size tend to be more similar to each other, exhibiting less structural diversity and yielding differences that are harder to interpret. However, we discuss how these differences may simply be mathematical artifacts, without necessarily implying a clear advantage in choosing one treatment over another. Finally, we argue that the field needs a targeted study to answer whether the differing numbers of alters listed is a function of network size.

1. Introduction

Over the past two decades, research on egocentric networks has grown exponentially in terms of both publications and citations, and is now widely used not only in the social sciences but also in fields such as public health and medicine. The practice of using the compositional and structural characteristics of the social context surrounding an individual (hereafter, 'ego') to predict outcomes for that ego is widely accepted. In some cases, the individuals surrounding an ego (hereafter, 'alters') can be harvested from different sources, such as Meta, LinkedIn and X. For most studies, however, the alters must be elicited from the ego through a series of questions. With this approach the goal is to elicit a set of alters that represents the social context that may impact ego for one or more outcome variables. The process of eliciting alters is largely the same across such studies. Ego is asked a question, or set of questions, to elicit the names of alters (people they know), these are called name generators. Ego is then asked to provide information about each alter using questions that align with the goals of the study, called name interpreters. Finally, in most ego network studies egos are asked to evaluate the existence of ties between alters. There are known concerns about the accuracy of cognitive reporting in social network data collection from ego (e.g., Bernard et al., 1979, 1984; Brashears and Quintane, 2015; Marin, 2004). Nonetheless, previous research has shown that ego-reported information about their own social contacts and the ties between them tends to be highly reliable (e.g., Moody

* Corresponding author.
E-mail address: migugonz@pa.uc3m.es (M.A. González-Casado).

https://doi.org/10.1016/j.socnet.2025.07.004
Received 17 February 2025; Received in revised form 17 July 2025; Accepted 17 July 2025
Available online 3 September 2025

et al., 2007; Green et al., 2014; McEvily, 2014; McCarty et al., 2019). This reliability increases when respondents are asked to report on a relatively large number of alters and to evaluate systematically the ties among them. Cognitive representations of social ties may not fully overlap with behavioral interaction data (e.g., Barrat and Cattuto, 2015; Mastrandrea et al., 2015), but both capture distinct and complementary dimensions of social life.

Regarding the collection of ego network data, although the methodology is well established, significant differences exist in the number of alters elicited and, thereafter, in the richness of the compositional and structural information available. To account for this situation, Perry et al. (2018) distinguished between 'focused' and 'expansive' approaches to the study of egocentric networks. The first selects a focus of interest (e.g., health) to collect relevant alters. The number of alters these approaches elicit is typically reduced (5–10). Even if the number of alters elicited may be large (e.g., >20) in some designs, the alter–alter tie evaluation considers a subsample of about 5 alters. The second approach is interested in exploring the complete foci of interaction in which the ego is active (e.g., Stulp, 2021; Bolíbar et al., 2015; Lubbers et al., 2010, 2021; Maya-Jariego and González-Tinoco, 2023; Bidart and Lavenu, 2005; Vacca et al., 2018). To achieve this goal, researchers adopt different name generator strategies, from listing multiple contexts or foci of interaction to elicit alters to the use of a single free generator but with a large number of alters (20 to 45). The latter is the approach we want to examine here, specifically the effect of fixing or not the number of alters. This question is relevant because the richness of compositional and structural information provided by the "expansive" approach allows the computation and comparison of multiple measures that, to our knowledge, have not been systematically explored to date.

There is considerable research on what constitutes a good name generator. Researchers may use any of the four distinct types of name generators typically used in the literature (exchange-based, content-based, affect-based, and interaction-based), combining them in a multiple name generator approach, and even randomly selecting alters to ease the alter–alter pair evaluation task (Perry and Roth, 2021). Most of the research is focused on the phrasing of the question to elicit alter names (Bidart and Charbonneau, 2011) or whether multiple questions should be used versus one global question (Marin and Hampton, 2007). There have been some attempts to elicit alter names using prompts to sample alters from memory (McCarty et al., 1997) and subsequent research testing the potential biases using that approach (Brewer, 1997). With the introduction of software to collect and analyze ego networks, there are debates about whether it is better to elicit names straight from memory, the approach used in Egonet (McCarty and Smith, 2007), or a visual interface to collect alter data such as is used with more recent tools like Network Canvas (Birkett et al., 2021) and Vennmaker (Michael Kronenwett, 2007). Less research has addressed an ongoing debate about the number of alters that best represents an ego's network, and whether that number of alters should vary for each ego. There is broad consensus that the size of ego networks varies across egos. Some people have many alters in their networks and some have fewer. Several network studies on network size demonstrate this variability (McCarty et al., 2001; Dunbar, 1993; Killworth et al., 1990; Lubbers et al., 2019; van Willigen et al., 1995). Size of ego networks also varies depending on the boundary definition a researcher uses. For any given person, their total network of all alters is defined by some connection, such as people you know or people you are close to.

However, researchers are often interested in a subset of the total network, such as the network that provides emotional support, or the network of alters who engage in risky behavior. These numbers also vary for each ego. There have been a few studies that examined the effect of the number of alters elicited on the structural properties of those networks. McCarty et al. (2007) used data from two studies where egos were each asked to name 45 alters, and also provided an evaluation of the ties between each pair of alters. The authors used various approaches to examine the structural properties of those networks with all 45 alters, considered by the study to be the best structural representation, and subsets of those alters at various levels. They found that most of the structural properties for subsets of 20 or more alters were not significantly different than those properties at size 45. This was particularly true if the subset was randomly selected from a larger list. Below size 20 the structural properties of the networks were different than they were at size 45. Golinelli et al. (2010) conducted a similar study to estimate the time savings associated with random samples of alters from a larger list.

Neither of the studies above investigated whether a fixed or variable number of alters was a better approach. They did demonstrate that the structural properties of personal networks varied with the number of alters in the network and identified a threshold at which those structural properties changed dramatically. Neal and Neal (2017) did a review of articles in the American Journal of Community Psychology that used social network analysis. In their review they concluded that using a fixed number of alters could lead to errors. They presented a hypothetical case where adding two additional alters to an ego network of four alters dramatically changed the structural properties of the ego network. They suggest that there is a consensus in the network community against using a fixed number of alters. It is worth noting that a network of four or six alters is well below the 20-alter threshold (McCarty et al., 2007) identified in their study of ego network structure, and thus subject to structural differences based on small changes in the number of alters. Maya Jariego (2018) disagreed with the Neal and Neal conclusion that ego network data collection should avoid a fixed number of alters approach. He summarizes several reasons why a fixed number of alters may be desirable in some ego network studies. The fixed alter approach may be particularly valuable when examining the variability of the structural properties of personal networks across respondents. Asking respondents to list alters as a measure of network size is a flawed approach, while other methods to measure personal network size exist. There is little certainty why a respondent stops listing alters, but there are at least three possibilities, (1) the respondent has listed all alters, (2) the respondent has listed a subset of alters and cannot remember all alters or (3) the respondent has listed a subset of alters and refuses to list the remaining alters due to fatigue or a lack of cooperation to complete the task. The potential effects of the second or third possibilities may be exacerbated when respondents are free to list (or not) more than 20 alters.

As can be seen, contrary to the conclusions of Neal and Neal (2017), there is no consensus in the network community on whether it is better for ego to determine the number of alters they list, or if the number of alters should be fixed by the researcher for each ego. To our knowledge, there has not been a systematic comparison of the structural properties of networks produced by different name generators. Campbell and Lee (1991) performed a comparison of networks constructed using different name generators, but they focused only on non-structural properties, like age, race, gender, etc. and Eagle and Proeschold-Bell (2015) explored other methodological considerations. In our study, nonetheless, we focus on the non trivial effects that the choice of a data collection strategy imprints on the structure of the egocentric network reconstructed. We collect ego network data from 298 individuals, divided into three treatment groups. In each treatment group, data is collected using a different operationalization of the name generator. We compare systematically the networks obtained in the three treatment groups, using the methodology developed in González-Casado et al. (2024), and discuss the implications of the choice of one treatment over another. Our main hypothesis is that the more a treatment constrains the size of the networks collected, the higher the correlations between different structural variables used to describe the network, and consequently, the lower the structural variability across those networks. We develop this idea in detail in the following sections.

2. Data and methods

2.1. Data collection

As defined above, in an ego network nodes represent alters, and edges indicate the existence of a personal relationship between them. Accordingly, in our study, each respondent reported the composition of their own ego network, the relationships between their alters, and additional required data. Respondents were divided into three treatment groups based on the survey used to collect the data. The data was obtained through surveys administered via a computer or mobile phone interface, with respondents accessing the survey through a provided link that could be completed on any electronic device. Completing the survey typically took between 15 and 60 min, with variations depending on the treatment group. We will analyze these differences in survey completion times when discussing the results. The sample was taken from the volunteers data base of the IBSEN project (IBSEN, 2018). Respondents were offered a payment of 5 euros for completing the survey, and entered into a 50-euro raffle among all respondents. The panel consists of 2500 individuals, 70% of whom are undergraduate and graduate students from UC3M University (Madrid), aged between 18 and 30 years. The gender balance is discussed in the Supplementary Material. The typical response rate is 1 in 6, depending on the survey. The research project was approved by the corresponding Ethics Committee. It is worth noting that our sample is relatively homogeneous and belongs to a specific demographic and cultural group, which limits the generalizability of our findings, regarding human behavior, to other populations. However, the goal of our study is not to draw conclusions about specific human behaviors, but rather to evaluate methodological approaches. For this purpose, we do not require three representative samples of the general population, but three groups that are comparable to one another. If this comparability holds, we can reasonably argue that the observed differences are due to the data collection procedures rather than to demographic or cultural variations, regardless of the sample's homogeneity. That said, any extrapolation of behavioral findings from our study should be made with caution.

Within each treatment group we used a different survey to elicit the names of the alters in the respondent's ego network. Specifically, respondents were randomly assigned to a treatment group and asked the following question: *Write **X** names of people you know by their first name and vice versa, people with whom you have had contact at least in the last two years and whom you can contact if necessary. This can be any type of person you know in any field (studies, work, family, associations or clubs, churches, sporting events, travel, etc.)*. Here, 'vice versa' implies a reciprocal relationship (the people ego names also know ego by their first name). The only difference between the treatment groups was the number X—the number of names the respondent was asked to provide:

- **Free Choice**: Unlimited names (up to 45). We have 98 respondents in this treatment group.
- **Minimum 20**: At least 20 names. We have 95 respondents in this treatment group.
- **Fixed 30**: 30 names. We have 105 respondents in this treatment group.

Our survey also included additional questions that were identical for all respondents, regardless of the treatment group. Specifically, we collected:

- **Gender**: Man/Woman/Other/Prefer not to answer
- **Age**
- **Survey Time**: Time taken by the respondent to complete the survey

Furthermore, for each alter name provided by the respondent, we also asked for their gender and age, as well as the following information:

- **Emotional Proximity**: *How emotionally close do you feel to [alter name] (from 1 to 5, with 1 being the minimum and 5 being the maximum emotional closeness)?* Only integer numbers were allowed.
- **Context**: *How did you meet [alter name]?* Respondents were given the following options: Family / Primary or Secondary Education / Higher Education / Job / Common Hobbies / Through another person / Neighbor / Other.

Finally, respondents were asked the following question: *Do [Alter i] and [Alter j] know each other in such a way that they could contact or meet without you being present?* They were given three response options: Yes, Maybe, or No. This question was asked for every possible pair of alters from the names elicited in the first question. Thus, if a respondent provided N alter names, they were required to evaluate $(N^2 - N)/2$ potential relationships between alters. With this information, we reconstructed a network for each respondent, where each node represents an alter, and a link between two alters exists if the respondent answered *Yes* to the previous question. An alternative analysis was conducted in which *Maybe* responses were also considered as existent links, yielding equivalent results. The main findings of this alternative analysis are summarized in Section S5 of the Supplementary Material. We include as well in Section S4 of the Supplementary Material a preliminary analysis of the gender composition of our sample.

2.2. Methodology

The implicit assumption is that our sample of *real* ego networks is statistically equivalent across all the participants, irrespective of the treatment. When a specific name generator is used, we obtain an *empirical* representation of the real ego network, aiming to capture its relevant characteristics. However, the choice of a data collection strategy directly influences the information collected, shaping the empirical network reconstructed from the real network. Our surveys differ in the number of alters required from respondents. While this difference may seem minor, it can have significant implications for the data collected. For instance, it is well known that larger networks tend to have lower densities, and this relation between density and network size is far from trivial (Friedkin, 1981). Consequently, when using a survey that elicits a larger number of alters, we expect to observe networks with lower densities. Importantly, this does not necessarily mean that the real network has a lower density. Instead, it reflects the effect of our methodological choice: the empirical ego network we construct will exhibit lower density. This distinction is crucial throughout the paper—we are comparing empirical networks without direct knowledge of the real underlying ego network.

The effect of varying the number of alters requested by the survey extends well beyond network density. Several other structural properties – such as subgraph centrality metrics (Estrada and Rodriguez-Velazquez, 2005) and community structure – are significantly influenced by changes in network size. Moreover, this variation affects not only individual structural variables but also the relationships between them. In other words, changes in network size affect the correlation patterns among structural variables, shaping the range of behaviors observable in the dataset. We will discuss this idea in detail below.

Network size also impacts aspects beyond structural properties. For instance, if too few alter names are elicited, there is a risk of misrepresenting interaction contexts. A small number of alters may disproportionately belong to a limited set of contexts, such as family or work, whereas a larger number of alters may introduce additional contexts, such as former classmates or neighbors. Similarly, when respondents provide fewer names, they may prioritize individuals with a high level of emotional proximity, potentially skewing the dataset. Other factors, such as age and gender distributions, could also be affected.

With these issues in mind, we divide our analysis into two parts. First, we examine the effects of the survey on variables unrelated to network structure, such as emotional proximity, interaction contexts, age, and gender. We refer to this as the non-structural analysis. Next, we investigate in detail how the survey influences the structural properties of the ego network.

2.2.1. Non-structural analysis

In our analyses, we aim to compare networks with one another. To enable this comparison between networks beyond structural properties, we use the available data to define some variables at the ego network level. We define the following variables:

- **Gender Homophily**: Computed as the proportion of alters within the ego network that share the same gender with ego.
- **Min/Max/Mean/Median/Standard Deviation of the age difference between the ego and all the alters**: 5 variables to characterize the distribution of age differences between the ego and the rest of alters. Besides, we analyze the age of the ego separately.
- **Levels of emotional proximity in the network**: Computed as the proportion of alters within each level of emotional proximity. When interpreting these values, one needs to take into account the differences in network sizes between the three treatment groups.
- **Context Heterogeneity**: Computed as the number of unique contexts that appear in the network over 8, the maximum number of contexts that can appear.
- **Context Presence**: Computed as the proportion of alters within each context. When interpreting the values, one needs to take into account the differences in network sizes between the three treatment groups.

For this part of the analysis we will explore simply how these variables are distributed across networks depending on the survey used.

2.2.2. Structural analysis

The structural analysis is based on the methodology presented in González-Casado et al. (2024). Here, we will summarize the key points necessary to understand the results presented, while referring the reader to the original reference for further (technical) details. Our starting point is a collection of adjacency matrices representing our ego networks. Our goal is to understand the differences in structures displayed by these networks and use this information to compare different treatment groups.

To this end, we compute a comprehensive set of 41 structural metrics, drawing from well-established sources in Network Theory and Social Network Analysis, such as Wasserman and Faust (1994), Estrada (2012), and Newman (2018). These metrics capture different aspects of network structure, including connectivity, closure, local and global centralities, distances, community and subgroup structures, and structural holes. A complete list of these variables, along with their mathematical definitions and interpretations, is provided in the Supplementary Material of González-Casado et al. (2024). We calculate these metrics for each network, constructing a dataset containing the values of these structural variables for every network. However, as discussed in the original article, some of these metrics may provide redundant information about the network structure. This means that, although our dataset includes many metrics, not all of them are necessary. To address this, we analyze correlation patterns among these metrics and apply a dimensionality reduction technique based on Factor Analysis (Mulaik, 2009; Buja and Eyuboglu, 1992). Simply put, this procedure groups together structural metrics that provide redundant information into a single composite metric. For instance, if correlation analysis reveals that metrics such as density, clustering coefficients, and transitivity convey similar information, we combine them into a single metric, which we arbitrarily name cohesion. This procedure yields a small number of global structural metrics, defined as linear combinations of the original 41 metrics, allowing us to explain and interpret the spectrum of ego network structures in our dataset. However, as described in González-Casado et al. (2024), we also need to understand how these global variables relate to the original structural metrics to interpret variations in network structure in simple terms. Since these global variables are expressed as linear combinations of the original metrics, each metric is linearly correlated with certain global variables. To summarize these relationships, we use the loadings matrix, which contains the correlation coefficients between the original metrics and the global structural variables. This matrix enables us to group similar metrics together and associate them with each global variable. Consequently, instead of analyzing 41 different variables, we can focus on a few key dimensions that capture the essential structural features of ego networks, thereby improving interpretability. With this summarized dataset, comparing treatment groups becomes more straightforward.

While the above analysis should be sufficient to compare networks across groups by examining how structural behaviors differ in terms of these global variables, a complementary analysis can further clarify these differences. We can classify networks within each group based on their structure, allowing us to describe the diversity of networks collected using each survey using a few representative ego networks that contain the key characteristic structural properties. By comparing these representative samples across groups, we can effectively summarize the main structural differences between treatment groups. For technical details, the reader is again referred to González-Casado et al. (2024).

3. Results

3.1. Non-structural analysis

In Fig. 1, we present a comparison of all the non-structural variables discussed in the previous section between treatment groups. Overall, we find that the behavior of these variables is fairly similar across groups. However, it is relevant to interpret the small observed differences. To do so, it is helpful to consider the distribution of network sizes for each group, shown in Fig. 2, panel (a) (distribution of the number of nodes). For the Fixed 30 group, all networks contain exactly 30 nodes. For the Free Choice group, the distribution is wide, with most networks having between 9 and 32 nodes (first and third quartiles) and a median of around 17.5 nodes. There are 23 small networks, with a size of less than 10 nodes; and 14 networks with the maximum size imposed by the survey, 45. The rest lie between 10 and 44 nodes. For the Minimum 20 group, the distribution is bounded between 20 and 45 nodes. Notably, this distribution is narrower, with most respondents naming between 20 and 25 alters (first and third quartiles, median of 21), and fewer respondents reaching larger values for the number of nodes. In this case it is worth noticing that the behavior is not equivalent to 'collapsing' the smaller networks in the Free Choice to 20 nodes while keeping the others. We interpret this behavior as being influenced by a sense of a 'completed task'. Since respondents are required to name at least 20 alters, many stop naming additional individuals once they reach this threshold, even if their network contains more people. This contrasts with the Free Choice group, where there is no explicit lower bound, and the absence of such a threshold likely avoids this psychological effect. Furthermore, for the Minimum 20 we expect to observe intermediate behavior – both in structural and non-structural variables – between the Free Choice and Fixed 30 groups, but closer to Fixed 30. The observed limited size variability is likely to impose correlation structures similar to those observed in Fixed 30 networks. We will explore this idea further in the following section.

Focusing on non-structural variables again (Fig. 1), we first observe differences in the time respondents took to complete the survey. As expected, individuals in the Fixed 30 group spent more time because

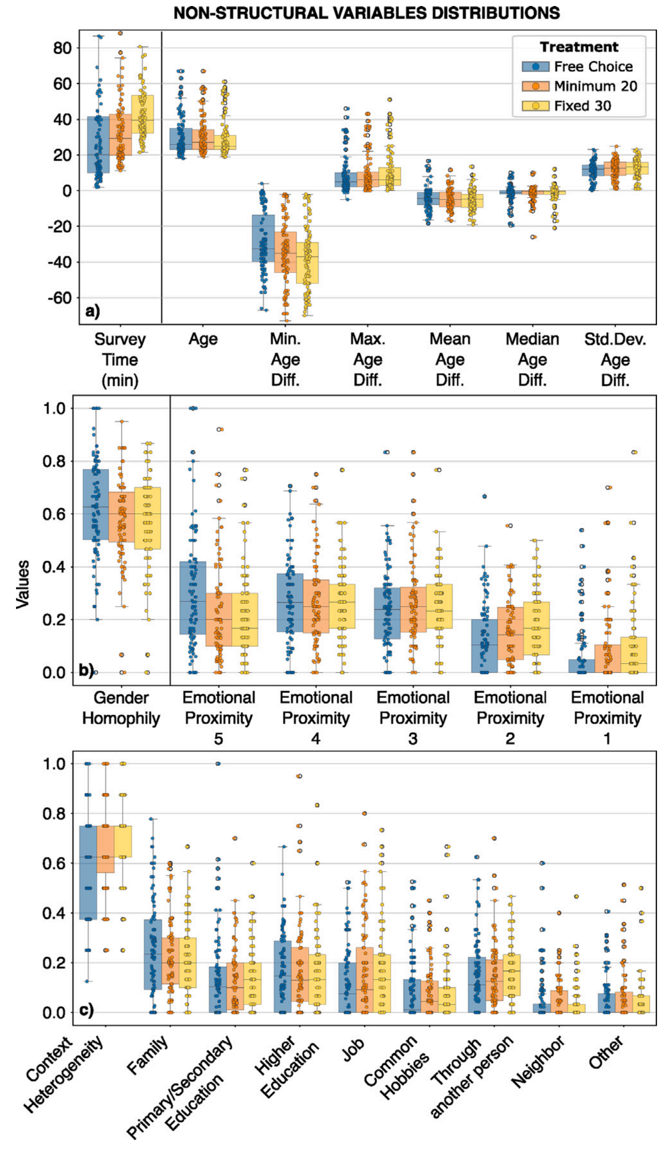

Fig. 1. Distributions of individual non-structural metrics for the networks in each treatment group.

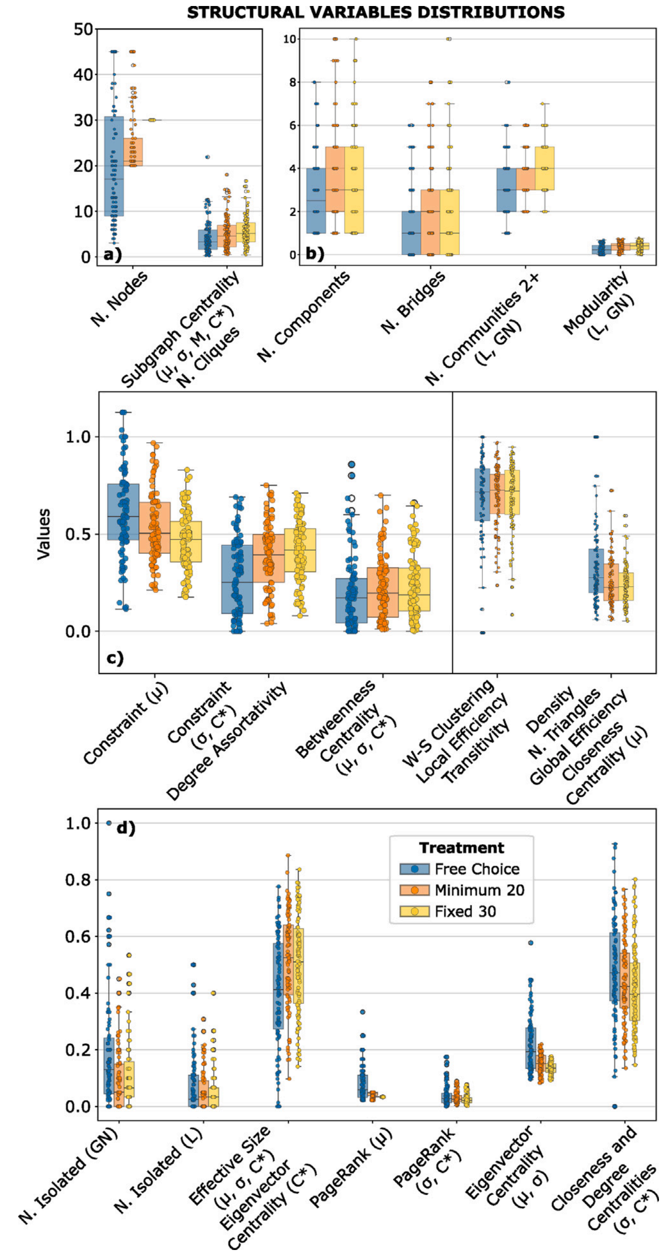

Fig. 2. Distributions of individual structural metrics for the networks in each treatment group. A subset of the original 41 metrics is shown, selected to represent different structural aspects. On the x-axis, the groups of variables that exhibit the same behavior are indicated below each distribution. The displayed distribution corresponds to the first variable listed within each group.

their networks are larger and they were required to evaluate all possible relationships among the named alters. Regarding age and the variables that characterize the distribution of age differences within each network, we find that most variables behave similarly across the three groups, except for the minimum age difference, which is larger (in absolute value) for the Fixed 30 group. We hypothesize that respondents tend to name people closer to them first, who are likely to be in the same or a nearby age group. However, as they are required to name more people, individuals from more distant age groups, such as grandparents or bosses, are increasingly included. This hypothesis was not explored further in this study. Gender homophily is slightly higher for the Free Choice group. It is well established that personal relationships tend to exhibit gender homophily, specially when these relationships are close to the ego (Rivera et al., 2010). Since respondents name fewer people in the Free Choice group, it is likely that the named individuals are closer to them and, consequently, share the same gender. This is further supported by the distribution of emotional proximities: in the Free Choice group, the proportion of alters with stronger emotional proximity is larger, as respondents name fewer, but closer, individuals. Conversely, the Fixed 30 group shows a higher proportion of weaker ties within egocentric networks. Finally, regarding the heterogeneity of contexts, networks in the Free Choice group are less heterogeneous, with a higher proportion of alters belonging to the respondents' families or higher education contexts. If respondents name people closer to them, it is reasonable that family members would be named first. Moreover, since our sample includes a significant proportion of university students, it is also likely that many alters belong to the higher education context. Despite these differences, it is important to emphasize that they are relatively minor. Across all groups, the networks provide a reasonable representation of emotional proximity levels, genders, contexts of interaction, and age groups. While the choice of data collection strategy does influence these distributions, the effects are not exaggerated, and thus this analysis does not point to a clear preference of one survey over the others.

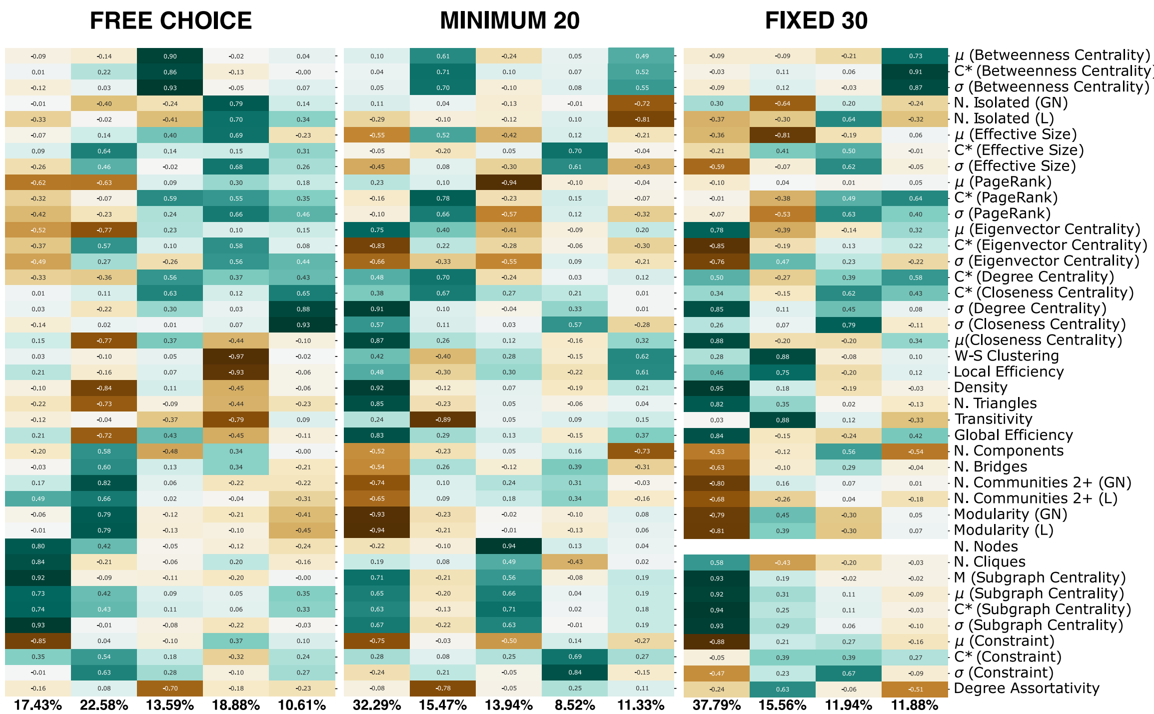

Fig. 3. Loadings Matrix for each treatment group. Columns in this matrices represent each global structural variable defined as a linear combination of the 41 structural metrics. Rows are these 41 structural metrics we measure in the ego networks. Each element of the matrix represents the correlation coefficient between each metric and each global variable. Below each variable we display the percentage of the variance of the data explained by that variable.

All these results are in agreement with the preliminary non-structural analysis performed by Campbell and Lee (1991).

3.2. Structural analysis

As explained above, for each network we measure a collection of 41 properties that capture different aspects of network structure. The most straightforward approach is to first compare how these variables are distributed across treatment groups, ignoring for the moment their collective behavior. Since many of these variables exhibit similar patterns, Fig. 2 displays the distribution of a representative subset, chosen to illustrate the different structural aspects measured. Overall, we find that, at the individual level, all variables behave similarly across the three name groups. However, as we will see, the key difference between groups is not captured by the individual behavior of these variables but rather by their collective behavior—specifically, by the correlation patterns among them. Despite the overall similarities, some differences are worth noting. As expected, for all variables, networks generated by the Minimum 20 survey exhibit an intermediate behavior between those of the other two groups. Focusing on specific metrics and comparing Free Choice networks with Fixed 30 networks, we find that Free Choice networks generally have a lower number of cliques, lower subgraph centrality metrics, fewer communities and components, lower modularity, and lower variance and centralization of the constraint and effective size metrics. They also exhibit higher values in global cohesion metrics such as density, triangle density, global efficiency, and average closeness centrality, as well as in average constraint, number of isolated nodes, average PageRank and Eigenvector centralities, and variance and centralization of closeness and degree centralities. Finally, some variables behave similarly across all three groups, including betweenness centrality metrics and local cohesion measures such as the Watts–Strogatz clustering coefficient, local efficiency, and transitivity.

All observed differences can be traced back to variations in network size across groups, since Fixed 30 networks are generally larger than Free Choice networks. In simple terms, certain structural metrics are properly normalized with respect to network size. This is the case for local cohesion variables and betweenness centrality metrics, meaning that their numerical values can be interpreted independently of network size. However, most other variables either lack explicit normalization or remain sensitive to network size even when normalization is applied. As a result, larger networks tend to exhibit more communities and components, higher subgraph centrality values, and lower global cohesion metrics, among other effects. This dependence of individual metrics on network size is well-documented in the literature and should not come as a surprise to the reader. However, as our analysis will suggest, this dependence has far-reaching implications beyond the individual behavior of variables. It fundamentally alters the internal correlation structure among variables—that is, their joint behavior. We will now discuss this in depth.

To analyze the collective behavior of these variables, the most straightforward approach is to compute the correlation matrix between them. The three correlation matrices for each group are presented in section S2 of the Supplementary Material. As explained in the methodology, some groups of metrics provide redundant information about network structure. This redundancy can be extracted from the correlation matrices using any dimensionality reduction technique. By applying this procedure, we group together structural metrics that convey redundant information into a single composite metric, effectively reducing the 41 structural metrics to a smaller set of global variables that summarize network structure. The loadings matrix for each group, which contains the correlation coefficients between the original metrics and these global structural variables, is shown in Fig. 3. We observe that networks generated by the Free Choice survey can be effectively described by five global structural variables. The first groups together all subgraph centrality metrics and the number of cliques, capturing the network's subgroup structure. The second combines all metrics related to community structure and global cohesion. The third primarily reflects betweenness centrality metrics. The fourth includes metrics related to Eigenvector and PageRank centralities, effective size, and local cohesion. Finally, the fifth is associated with centralization and the variance of closeness and degree centrality metrics. The amount of variance explained by each of these global variables is relatively

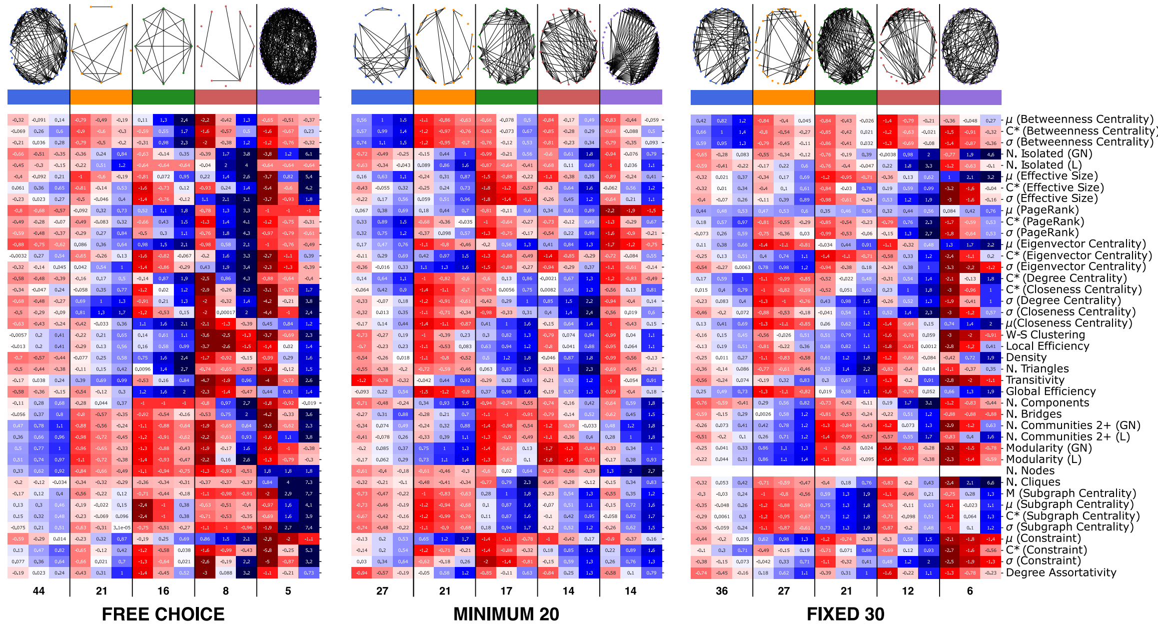

Fig. 4. Heat-Map of the values of the structural metrics that characterize each cluster in the 5-cluster partition for each treatment group. Each row in the heatmap represents one of the 41 metrics measured for each PN. The columns are divided into five clusters (colors). Within each cluster, there are three values displayed for each row. The central value represents the mean value of the standardized metric for that particular cluster. The values on the right/left sides indicate the upper/lower limits of the 95% confidence interval. On top of each cluster we present the network closer to the centroid of the cluster (thus, the network that is representative of the cluster). In the bottom of each cluster we display the number of networks that belong to that cluster. Networks are not necessarily distributed homogeneously between clusters because some regions of the space of structural variability may be denser than others. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

balanced. In other words, networks generated by the Free Choice survey can be described by five global structural variables, each summarizing a distinct and interpretable aspect of network structure and capturing a similar amount of variance.

This contrasts with the other two groups. Examining the loadings matrix of the Fixed 30 group, we see a single dominant global variable that explains almost 40% of the dataset's variance and groups together most of the individual metrics, including subgroup structure, community structure, global cohesion, and some Eigenvector centrality metrics. The remaining variance is captured by three dimensions, each grouping far fewer metrics: one related to local cohesion, another to degree and closeness centrality variance, PageRank, and effective size, and a third to betweenness centrality. The loadings matrix for the Minimum 20 group falls between these two extremes but is clearly closer to the Fixed 30 case, where a single global variable dominates, grouping most structural metrics together.

This result can be understood by revisiting the correlation matrices. In simple terms, when network size is not constrained by the survey (as in the Free Choice case), variables correlated with network size can vary independently. For instance, if we do not fix network size, networks can exhibit independent variation in subgraph centrality and community structure metrics. While both are correlated with network size—since larger networks tend to have larger subgroups and communities—these two aspects remain independent of each other. Thus, knowing a network's subgroup structure does not provide direct information about its community structure. However, when network size is constrained (as in the Minimum 20 and Fixed 30 cases), the correlation matrix changes. Essentially, the individual correlations that network size had with each metric now propagate, effectively coupling them. This means that previously independent metrics become correlated. Returning to the earlier example, community and subgroup structures are no longer independent—knowing one now provides information about the other. As a result, the correlation matrix becomes less sparse, reducing the number of global variables required to describe the data. This explains why, in the Fixed 30 case, a single dimension groups together subgraph centrality, community structure, global cohesion, and Eigenvector centrality: all these variables are strongly influenced by network size, and once size is fixed, they become mathematically coupled. Consequently, the amount of redundant information among individual metrics increases, meaning that different surveys alter the informativeness of certain variables in describing network structure.

In practical terms, our results suggest that networks generated by the Free Choice survey exhibit greater structural diversity, with a larger number of independent dimensions (i.e., the network structure can vary across more independent factors). Moreover, these independent dimensions are more interpretable, as they group together structurally related metrics and explain similar proportions of variance. By constraining network size through a specific survey, we limit the possible independent variations in structural metrics.

Despite these theoretical insights, it is not immediate to interpret how these different behaviors translate into specific structural differences in the observed networks. To better visualize these differences, we classify networks for each group, generating a taxonomy of structural types, as specified in the methodology. This approach divides the spectrum of structures observed into distinct regions characterized by specific structural properties. By selecting representative networks from each region, we can compare groups based on a small set of exemplary networks. Specifically, we choose five networks (this number is selected arbitrarily) for each group to make the comparison.

In Fig. 4 we depict the five paradigmatic networks for each group along with a diagram that conveys the structural metrics that characterize each presented network. Free Choice networks can be divided into these five groups: networks with a distinct community structure, cohesive networks, networks with central nodes (betweenness centrality metrics), small and sparse networks, and networks with subgroup structure (subgraph centrality metrics). These clusters correspond directly to different combinations of values in the five global structural variables identified earlier. In contrast, for the Minimum 20 and Fixed 30 groups, representative networks are primarily characterized by specific combinations of density, subgraph centrality, and community structure. For both groups we find networks with a distinct community structure, networks with high density and subgraph centrality but weak community structure, mixed-type networks, and networks with high subgraph centrality and community structure but low density. Besides, we also find networks with central nodes. These types illustrate how the dimensions of community structure, subgraph centrality and cohesion are no longer independent, as in the Free Choice group.

4. Discussion

In this work, we compared ego networks constructed using three different operationalizations of a name generator, where the only difference is the number of alters elicited from the respondent. We analyzed both the structural differences between these networks and differences in other non-structural variables. The three surveys produce networks that are highly similar in terms of age and gender composition, emotional proximity, and context of interaction. While some differences are expected due to variations in network size, all levels of emotional proximity, interaction contexts, genders, and ages are sampled similarly across the three treatment groups. At the individual level, structural variables also behave similarly. Although structural differences are more pronounced than non-structural ones, the overall similarity between the three groups remains. All observed differences can be traced back to the correlation patterns between the number of nodes and other structural variables.

The most significant differences among the three groups emerge in the collective behavior of structural metrics—in the correlation structure among them, the amount of redundant information each variable provides, and the diversity and interpretability of the observed structural variability. Overall, we find that, for the ego networks constructed using the Fixed 30 and Minimum 20 surveys, the constraint on network size induces correlations between variables that were previously independent. This reduces the sparsity of the correlation matrix, effectively decreasing the number of independent global variables needed to describe network structure and making these global variables less interpretable. While the Free Choice survey produces five easily interpretable global variables, each capturing distinct, intuitive aspects of network structure, the Fixed 30 and Minimum 20 surveys produce a single dominant dimension that explains most of the variance, making structural variation more difficult to interpret and visualize. This conclusion is reinforced by our structural taxonomy analysis, which shows that networks from the Fixed 30 and Minimum 20 groups are intuitively more similar to each other than to those from the Free Choice group. This similarity arises because these surveys constrain structural variability, leading to a lower-dimensional space of possible network structures.

However, caution is needed when interpreting these findings. Apart from the relatively homogeneous population of our panel, the greater structural diversity and better interpretability of networks from the Free Choice group do not necessarily indicate a methodological advantage; they may simply be mathematical artifacts resulting from a wider distribution of network sizes. If real ego networks indeed vary widely in size – from as few as 5–6 nodes to more than 45 – then surveys that restrict network size artificially limit the spectrum of observable structures, introducing biases into data collection. This raises a fundamental question: How effectively do survey methods capture the variability in personal networks size? If the variation in numbers of alters elicited with a Free Choice approach is not a function of actual network size, but something else, such as varying levels of cooperation or ability to recall, then the more varying levels of structural measures are a mathematical artifact. The field needs a targeted study to answer whether the differing number of alters listed is a function of network size. Such a study could involve a free choice elicitation and one or more independent assessments of network size, such as using the network scale-up method (Laga et al., 2021).

This reveals a trade-off: A survey that constrains ego network size inherently limits structural variability and diversity. However, if network size is left unrestricted, responses may be subject to biases—such as the inclusion of non-compliant respondents who underreport their network size. Our non-structural analyses suggest that valid networks can still be recovered despite these constraints, but we cannot be certain. Regardless of the choice of a data collection strategy, the structural information recovered is shaped and potentially biased by its design. If network size is predetermined, the range of mathematically possible structural behaviors is constrained, meaning individuals with different ego networks may appear similar when analyzed using a survey that enforces a fixed number of alters. However, if network size is not fixed, then underreporting becomes a possibility, and small empirical ego networks may not accurately reflect real small networks.

In sum, among the three groups, the Fixed 30 and minimum 20 appeared very similar, and given this result, we would not recommend the option of a minimum 20 with a variable maximum strategy. There is a trade-off regarding structural variability when comparing a Free Choice strategy to a Fixed Choice strategy. It is up to the researcher to decide if the increased structural variability from a Free Choice strategy represents actual variability in the networks, the potential introduction of bias, or if the lower structural variability introduced by the Fixed Choice approach is acceptable. This article clearly shows the consequences on structural measures of free versus fixed numbers of alters. It does not provide a clear recommendation. The researcher must determine whether they think the variation in network size from free choice generators reflects real variation in size or something else. Indeed, the field could use a study dedicated to answering that question. It is not clear to us how such a study would be designed. Note that both approaches generate variability across egos. It is important to note that these observations are only for ego networks with larger (>20) alters. Ego network studies constrained to <10 alters typically yield much lower structural variation. Some studies, such as those focused on a particular type of social support, will often yield smaller numbers of alters, and a Free Choice approach may be more appropriate. Alternatively, asking about a larger number of alters and then identifying those that provide that support provides an opportunity to understand how support networks are embedded in larger networks.

In any case, we have demonstrated that, from a theoretical and mathematical perspective, changes in network size induced by the choice of survey have profound, non-trivial effects on the reconstructed structure, extending beyond the individual behavior of each structural metric and surpassing what has been studied so far. This work highlights the effectiveness of the methodology developed in González-Casado et al. (2024) for analyzing the structural diversity of a group of networks. Besides, it enables detailed comparisons between groups of networks, and allows us to detect and characterize non-trivial structural differences with mathematical precision. We believe that such comparisons, as those reported here, and the description of the consequences of choosing one data collection strategy over another will prove themselves useful for researchers in the field.

CRediT authorship contribution statement

Miguel A. González-Casado: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Alejandro Cruzado Rey:** Writing – review & editing, Software, Data curation. **Miroslav Pulgar Corrotea:** Writing – review & editing, Investigation. **Christopher McCarty:** Writing – original draft, Supervision, Methodology, Investigation, Funding acquisition, Conceptualization. **José Luis Molina:** Writing – review & editing, Supervision, Methodology, Investigation, Funding acquisition, Conceptualization. **Angel Sánchez:** Writing – review & editing, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization.

Acknowledgments

M.A.G.-C. acknowledges support from the Comunidad de Madrid, Spain through the grants for the hiring of pre-doctoral research personnel in training (reference PIPF-2023/COM-29487). M.A.G.-C. and A.S. acknowledge support from grant PID2022-141802NB-I00 (BASIC) funded by MCIN/AEI/10.13039/501100011033 and by 'ERDFA way of making Europe'. M.A.G.-C., M.P.C, C.M, J.L.M. and A.S. acknowledge support from grant MapCDPerNets—Programa Fundamentos de la Fundación BBVA 2022.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.socnet.2025.07.004.

Data availability

All necessary data for replication can be found in this Github repository.

References

Barrat, A., Cattuto, C., 2015. Face-to-face interactions. Soc. Phenom.: From Data Anal. Model. 37–57.
Bernard, H.R., Killworth, P., Kronenfeld, D., Sailer, L., 1984. The problem of informant accuracy: The validity of retrospective data. Annu. Rev. Anthr. 495–517.
Bernard, H.R., Killworth, P.D., Sailer, L., 1979. Informant accuracy in social network data IV: A comparison of clique-level structure in behavioral and cognitive network data. Soc. Netw. 2 (3), 191–218.
Bidart, C., Charbonneau, J., 2011. How to generate personal networks: Issues and tools for a sociological perspective. Field Methods 23 (3), 266–286.
Bidart, C., Lavenu, D., 2005. Evolutions of personal networks and life events. Soc. Netw. 27 (4), 359–376.
Birkett, M., Melville, J., Janulis, P., Phillips II, G., Contractor, N., Hogan, B., 2021. Network canvas: Key decisions in the design of an interviewer-assisted network data collection software suite. Soc. Netw. 66, 114–124.
Bolíbar, M., Martí, J., Verd, J.M., 2015. Just a question of time? The composition and evolution of immigrants' personal networks in catalonia. Int. Sociol. 30 (6), 579–598.
Brashears, M.E., Quintane, E., 2015. The microstructures of network recall: How social networks are encoded and represented in human memory. Soc. Netw. 41, 113–126.
Brewer, D.D., 1997. No associative biases in the first name cued recall procedure for eliciting personal networks. Soc. Netw. 19 (4), 345–353.
Buja, A., Eyuboglu, N., 1992. Remarks on parallel analysis. Multivar. Behav. Res. 27 (4), 509–540.
Campbell, K.E., Lee, B.A., 1991. Name generators in surveys of personal networks. Soc. Netw. 13 (3), 203–221.
Dunbar, R.I., 1993. Coevolution of neocortical size, group size and language in humans. Behav. Brain Sci. 16 (4), 681–694.
Eagle, D.E., Proeschold-Bell, R.J., 2015. Methodological considerations in the use of name generators and interpreters. Soc. Netw. 40, 75–83.
Estrada, E., 2012. The Structure of Complex Networks: Theory and Applications. American Chemical Society.
Estrada, E., Rodriguez-Velazquez, J.A., 2005. Subgraph centrality in complex networks. Phys. Rev. E— Stat. Nonlinear, Soft Matter Phys. 71 (5), 056103.
Friedkin, N.E., 1981. The development of structure in random networks: an analysis of the effects of increasing network density on five measures of structure. Soc. Netw. 3 (1), 41–52.
Golinelli, D., Ryan, G., Green Jr., H.D., Kennedy, D.P., Tucker, J.S., Wenzel, S.L., 2010. Sampling to reduce respondent burden in personal network studies and its effect on estimates of structural measures. Field Methods 22 (3), 217–230.
González-Casado, M.A., Gonzales, G., Molina, J.L., Sánchez, A., 2024. Towards a general method to classify personal network structures. Soc. Netw. 78, 265–278.
Green, Jr., H.D., Hoover, M.A., Wagner, G.J., Ryan, G.W., Ssegujja, E., 2014. Measuring agreement between egos and alters: understanding informant accuracy in personal network studies. Field Methods 26 (2), 126–140.
IBSEN, 2018. Bridging the gap: From individual behaviour to the socio-technical MaN. (Accessed 07 February 2025) https://participantes.ibsen-h2020.eu/about.
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.
Laga, I., Bao, L., Niu, X., 2021. Thirty years of the network scale-up method. J. Amer. Statist. Assoc. 116 (535), 1548–1559.
Lubbers, M.J., Molina, J.L., Lerner, J., Brandes, U., Ávila, J., McCarty, C., 2010. Longitudinal analysis of personal networks. The case of argentinean migrants in Spain. Soc. Netw. 32 (1), 91–104.
Lubbers, M.J., Molina, J.L., Mccarty, C., 2021. How do migrants' processes of social embedding unfold over time? Glob. Netw. 21 (3), 529–550.
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.
Marin, A., 2004. Are respondents more likely to list alters with certain characteristics?: Implications for name generator data. Soc. Netw. 26 (4), 289–307.
Marin, A., Hampton, K.N., 2007. Simplifying the personal network name generator: Alternatives to traditional multiple and single name generators. Field Methods 19 (2), 163–193.
Mastrandrea, R., Fournet, J., Barrat, A., 2015. Contact patterns in a high school: a comparison between data collected using wearable sensors, contact diaries and friendship surveys. PloS One 10 (9), e0136497.
Maya Jariego, I., 2018. Why name generators with a fixed number of alters may be a pragmatic option for personal network analysis. Am. J. Community Psychol. 62 (1–2), 233–238.
Maya-Jariego, I., González-Tinoco, E., 2023. Use of a hierarchical deconstruction procedure for the classification of personal networks: Exploring nested groups around you. Soc. Netw. 73, 20–29.
McCarty, C., Bernard, H.R., Killworth, P.D., Shelley, G.A., Johnsen, E.C., 1997. Eliciting representative samples of personal networks. Soc. Netw. 19 (4), 303–323.
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.
McCarty, C., Killworth, P.D., Rennell, J., 2007. Impact of methods for reducing respondent burden on personal network structural measures. Soc. Netw. 29 (2), 300–315.
McCarty, C., Lubbers, M.J., Vacca, R., Molina, J.L., 2019. Conducting Personal Network Research: A Practical Guide. Guilford Publications.
McCarty, C., Smith, M., 2007. Egonet. (Accessed 07 February 2025) http://sourceforge.net/projects/egonet/.
McEvily, B., 2014. Do you know my friend? Attending to the accuracy of egocentered network data. In: Contemporary Perspectives on Organizational Social Networks. Emerald Group Publishing Limited, pp. 295–313.
Michael Kronenwett, M., 2007. Vennmaker. (Accessed 07 February 2025) https://www.vennmaker.com.
Moody, J., et al., 2007. To tell the truth: Measuring concordance in multiply reported network data. Soc. Netw. 29 (1), 44–58.
Mulaik, S.A., 2009. Foundations of Factor Analysis. CRC Press.
Neal, Z.P., Neal, J.W., 2017. Network analysis in community psychology: Looking back, looking forward. Am. J. Community Psychol. 60 (1–2), 279–295.
Newman, M., 2018. Networks. Oxford University Press.
Perry, B.L., Pescosolido, B.A., Borgatti, S.P., 2018. Egocentric Network Analysis: Foundations, Methods, and Models. Vol. 44, Cambridge University Press.
Perry, B.L., Roth, A.R., 2021. On the boundary specification problem in network analysis: An update and extension to personal social networks. Pers. Netw. 431–443.
Rivera, M.T., Soderstrom, S.B., Uzzi, B., 2010. Dynamics of dyads in social networks: Assortative, relational, and proximity mechanisms. Annu. Rev. Sociol. 36 (1), 91–115.
Stulp, G., 2021. Collecting large personal networks in a representative sample of dutch women. Soc. Netw. 64, 63–71.
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.
Wasserman, S., Faust, K., 1994. Social network analysis: Methods and applications.
van Willigen, J., Kedia, S., Chadha, N., 1995. Personal networks and sacred texts: Social aging in Delhi, India. J. Cross- Cult. Gerontol. 10, 175–198.