Chapter 3
Simulating an Empirically Informed Population Network of Core Discussion Ties

Michał Bojanowski, Alla Loseva, Paul Schuler, Susanne Böller, and Miranda J. Lubbers

Abstract Social cohesion and polarization can be studied with society-wide networks, however such data are not easy to collect. The alternatives—stylized networks often used in agent-based models (ABM) and small-scale empirical networks—might misrepresent societal structure at the macro-level, in terms of both topological properties and differences in network positions of social groups. We demonstrate how more realistic society-level networks can be generated from ego-network data that can be reasonably collected even for large populations. The empirical data we use come from a large-scale survey on a representative national sample about individuals' core discussion ties, collected by the Spanish Institute for National Statistics (INE). In this paper, we describe our methods and the properties of the resulting networks. We find that large-scale core discussion networks do not have small-world properties, resembling population-level familial networks rather than friendship networks. At the same time are characterized by assortativity on dimensions such as age, gender, working status and political views. We also demonstrate how model-based simulations of complete networks can be further analyzed to make inferences that go beyond personal network analysis, such as centrality and connectivity. Together with this paper, we release the simulated full networks for secondary use by ABM modelers and wider research community interested in studying cohesion in realistic networks.

M. Bojanowski (✉) · M. J. Lubbers
Autonomous University of Barcelona, Barcelona, Spain
e-mail: mbojanowski@kozminski.edu.pl

M. Bojanowski
Kozminski University, Warsaw, Poland

A. Loseva
ICS, University of Groningen, Groningen, Netherlands

P. Schuler
University of Edinburgh, Edinburgh, Scotland

S. Böller
University of Bonn, Bonn, Germany

M. A. Keijzer et al. (eds.), *Computational Social Science of Social Cohesion and Polarization*, Computational Social Sciences,
https://doi.org/10.1007/978-3-032-01373-6_3

3.1 Introduction

Social cohesion is a pivotal concept in the social sciences, reflecting the degree to which a society is characterized by absence of social conflict and polarization and presence of strong social bonds (e.g., Durkheim 1893). Social relations between groups and individuals, and social networks as their component, are the most prominent aspect of social cohesion (Schiefer and van der Noll 2017). Of particular significance are core discussion networks—3–5 typically strong ties with whom individuals discuss important matters (Marsden 1987). Structural patterns found in these networks could scale up to societal-level segregation, as well as lay out the pathways of information diffusion and polarization. The role of social networks in these macro-level phenomena is frequently studied with agent-based modeling (ABM).

To make empirically sound conclusions about cohesion, one needs to work with population networks of realistic structure. In much of the ABM literature, agents' interaction spaces are initialised using stylised networks like complete graphs, random graphs, scale-free graphs, or small-world networks (Amblard et al. 2015; Hedström & Manzo 2015). While different levels of model realism may be desirable depending on the research question, network topology can significantly affect the model results (e.g., Guilbeault et al. 2018; Rolfe 2014). Research has shown that empirical networks, drawn from real-world data, can more accurately capture complex social dynamics than stylised networks, particularly for processes like knowledge diffusion (Cointet and Roth 2007).

However, researchers have limited understanding of what network structures arise empirically at the population level, as population-level networks are scarce and notoriously hard to collect. Available large-scale networks mostly come from online networks and national registries (e.g., Chetty et al. 2022; Corten 2012; Kazmina et al. 2024; Bokányi et al. 2023). Although examining their structure is insightful, they do not necessarily represent core discussion ties.

While using stylized networks results in limited insight into empirical context and obtaining real large-scale networks is challenging, a practical alternative is to use realistic but artificially generated large-scale networks. Among the various methods available (Heß et al. 2021), inferring a sociocentric network from ego network data stands out as the most efficient one as it does not require complete data about all interdependencies in the population or even in its subset. This approach involves collecting data from individuals about their connections, such as friends or confidants, and then using a graph-generating model based, for instance, on Exponential-family Random Graph Modeling (ERGM) to infer a global network structure (Smith 2012).

So far, however, this method has only been used on small samples, which limits its potential to represent relationship patterns accurately, especially for smaller subpopulations. Here, we present a version of this strategy applied to unique, large-scale survey data from Spain, to generate empirically-informed population networks based on a wealth of relationship data. We compare the structure of our empirically calibrated networks to stylised networks sampled from classic network models. Together with this paper, we release a dataset that includes these realistic full networks.

Our simulations are valuable for two main reasons. Firstly, despite the extensive collection of personal network data, there is still a gap in understanding what these data reveal about the structure of strong sociability ties within the larger society. By simulating the networks, we can gain insights into how these data collected at the individual level reflect broader network features at the larger society level. Secondly, using the data about core discussion networks to create empirically-informed networks can enhance the realism and applicability of ABMs in the domain of opinion diffusion and polarisation (e.g., Boero and Squazzoni 2005; Bruch and Atwell 2015; Chattoe-Brown and Gabbriellini 2021; Lorenz 2021). Although we do not aim to create 'hyperrealistic' networks,[1] our study is informative in that we pinpoint where the simulated networks differ from stylized structures. We also demonstrate how analysis of distribution of such empirically-calibrated networks provides insights into relative differences between certain groups in centrality and connectivity—features which are not directly observed in egocentric data, but can be subject of inference given the partial network information available in the personal networks.

The structure of the chapter is as follows. The next section discusses the network structures used in cohesion research and argues for our approach. Then, we present the range of methods that have been used so far to insert empirical realism into ABM network environments and justify our choice of method. Then, we detail how the method works and our modeling decisions. Afterwards, we present the empirical data and the outcomes, which we discuss further in the conclusion.

3.2 Background

3.2.1 Network Structures in Cohesion Research

To approach the macro-phenomenon of cohesion, one can start by examining micro-level interpersonal networks. Core discussion partners—small egocentric networks of 3–5 people with whom individuals discuss "important matters" (e.g., Marsden 1987)—provide useful data. These core network members are typically contacted weekly and discuss personal and significant societal topics, such as politics (Klofstad et al. 2009). Therefore, core networks are particularly valuable for studying communicative processes related to social cohesion, including opinion diffusion and the emergence of shared norms.

Empirically, core discussion networks are often "small, kin-centered, relatively dense, and homogeneous" (Marsden 1987, p. 122). They typically show segregation by ethnicity and socio-economic status (SES; McPherson et al. 2001). For example, among Turkish and Moroccan minorities in the Netherlands, three-quarters of confidants are from the same ethnic group, which overlaps with categorical boundaries of

---

[1] We thank an anonymous reviewer for this notion.

SES (van Tubergen 2015). Ethnic closure is even more pronounced among women, those living in immigrant concentration areas, and individuals with strong ethnic or religious identities. A longitudinal study using US General Social Survey data reveals that core discussion networks remain segregated by race and age despite increased societal diversity over time. Moreover, young people become more isolated from older generations outside the family, and there is weak evidence of rising educational and religious homophily (Smith et al. 2014).

To study cohesion at the macro scale, one would need to work with a society-wide core discussion network. However, such networks might not exhibit the same structures as small-scale core networks. Theoretically, Blau (1974) suggested that if social networks consisted solely of core discussion network ties, they would primarily comprise small, isolated "islands" lacking overall cohesion.

Empirical large-scale discussion networks are hard to collect. The two mostly used types of representations of empirical population-level networks are online friendship networks (Chetty et al. 2022; Corten 2012) and networks based on administrative register data (Bokányi et al. 2023; Cremers et al. 2024; Kazmina et al. 2024; Menyhért et al. 2024). Studies show that these empirical networks deviate from synthetic ones (such as scale-free networks) in various ways, for example in local transitivity. One caveat is that it is unclear whether ties in these networks (e.g., between family members, next-door neighbors, colleagues, classmates) correspond to core discussion ties, as neither registers-based networks nor online friendship networks show any indication of tie strength or communication intensity.

When accessing large-scale networks is not an option, researchers can use artificially generated network structures. Most agent-based modeling studies of social influence and information diffusion use artificial networks: Only 3% of the studies published in the field between 1998 and 2015 used empirical data to inform the generated networks (Amblard et al. 2015). Researchers using artificial data typically draw on complete, random, scale-free, lattice, or small-world graphs (Albert and Barabási 2002). These networks are either grown during simulations based on agents' behavioral rules (e.g., Zachara and Piskor-Ignatowicz 2016) or formed by assuming ties between agents who are adjacent in some appropriately defined social space (e.g., Hamill and Gilbert 2009).

The choice of network topology significantly influences the outcomes of various models (Guilbeault et al. 2018; Rolfe 2014). For example, the diffusion of retirement norms differs between grid-based and random networks (Axtell 2001), and continuous opinion dynamics yield varied opinion landscapes based on network structure (Amblard and Deffuant 2004; Deffuant 2006). Rolfe (2014) shows that simple contagion models produce vastly different outcomes depending on the network used, while Guilbeault et al. (2018) find similar variations for complex contagion models. Even different types of small-world networks do not produce consistent results in models of epidemics, opinion, and cultural dynamics (Thiriot 2020).

The impact of network topology is particularly striking when comparing stylised networks to empirical ones. For example, Bearman et al. (2004) highlight differences between empirically observed sexual partner networks and those used in epidemiological models, noting variations in infection rates and discussing policy implications. Therefore, Cointet and Roth (2007) recommend using a real-world network of similar size from an adjacent domain to accurately represent real-world processes.

Empirically-based synthetic networks have been suggested as a viable compromise between data accessibility and realism of the resulting structure. Heß et al. (2021) demonstrate that such synthetic networks more closely resemble empirical ones than random networks, in their case affecting the discussed migration models. These insights highlight the importance of using realistic network topologies to improve the accuracy of social dynamics models.

3.2.2 Methods of Generating Realistic Networks

Multiple approaches have been proposed for generating realistic networks. For the sake of brevity we skip studies relying on theoretical and "feature-driven" (Lim et al. 2016) models such as ones generating scale-free (Barabási and Albert 1999) or small-world graphs (Watts and Strogatz 1998) with "plausible" generating parameters, and concentrate on models and approaches more directly focused on empirical data. We group them into the following rough categories:

1. **Methods not requiring direct social network measurement**. These approaches usually rely on multiple data sources on distributions of characteristics of the whole society: census data on demographics and household structure. Such statistics are often combined with data on population activities such as mobility and daily routine sequences. Network ties are then introduced via mechanisms such as spatial proximity, co-affiliation in schools and workplaces, or other theoretically informed procedures and their parameters are usually set by assumption or trial-and-error. Examples include Barrett et al. (2009), Burger et al. (2017), Eubank et al. (2004), Jiang et al. (2022), and Thiriot and Kant (2008). Such methods are often used in epidemiology or transportation policy planning. While based on population statistics, the network-generating part of these approaches heavily relies on theoretical assumptions.

The other two approaches involve simulating networks using a more explicit model of a target system.

2. **Methods relying on direct measurement of a complete social network**. When the full network is available, the data can be used to estimate/train a generative network model, and at a later stage, use the model to generate/sample data for the actual ABM simulations. This approach includes both generic methods utilising genetic programming or symbolic regression (Menezes and Roth 2014, 2019; Heß et al. 2021) and models specifically developed for social networks such as Exponential-family Random Graph Model (ERGM; Frank and Strauss 1986; Wasserman and Pattison 1996; Lusher et al. 2013; Snijders et al. 2006) or, if

dynamic network data is available, the Link Probability Model (McCulloh et al. 2010) or Stochastic Actor-Oriented Modeling (SAOM) (Snijders et al. 2010). The available network of the population of interest can of course be used in the simulations directly. Still, it is useful to work with a model fitted to that network and use it to produce a population of networks to consider possible variation in network structure while keeping the crucial structural features consistent with the empirical data.

3. **Methods relying on partial social network measurement**. Several methods have been proposed for scenarios when complete network data is unavailable. A variant of an ERGM can be used to infer and simulate complete networks from ego network samples (Smith 2012, 2015). Studies have shown that networks generated with this method closely resemble known complete networks in terms of diffusion dynamics and network-level cohesion (Smith and Burow 2020; Smith and Gauthier 2020). This approach works well if the network under study has a relatively low degree such that frequencies of types of simulated personal networks can be monitored and compared to the observed frequencies. As an alternative, more recent developments in the ERGM methodology allow directly fitting a subclass of these models to egocentrically-sampled data. The sub-class consists of models containing terms for which the sufficient statistics are recoverable from an egocentric sample. The list includes exogenous terms related to attribute-based degree heterogeneity, assortativity, and endogenous terms such as degree effects and triadic effects related to the distribution of edgewise shared partners (Krivitsky et al. 2011, 2019, 2022; Krivitsky and Morris 2017). We will use this method in the remainder of this article.

Adopting the Exponential-family Random Graph Model (ERGM) as the statistical framework for this study offers several compelling advantages over the other approaches. First, ERGMs are grounded in a rigorous and principled statistical foundation. These models are members of the exponential family of distributions, with properties that have been extensively studied. Their estimation relies on established statistical methods, such as Maximum Likelihood Estimation, rather than relying on less formal or ad hoc generative algorithms. This approach enables robust evaluation of model convergence and goodness of fit within a well-defined framework (Hunter et al. 2008).

Second, the coefficients of ERGMs have meaningful interpretations related to social processes of tie formation. These include concepts such as attribute-related degree heterogeneity, assortativity, mutuality, and triadic closure, making the model particularly relevant for sociological research.

Third, even when ERGMs are fitted to egocentrically sampled data, they retain the capacity to generate complete network datasets. These generated datasets exhibit properties consistent with those observed in the egocentric sample, enhancing the model's utility for analyzing and extrapolating network characteristics.

The analysis presented below applies ERGM framework to a large, representative sample from the national population of Spain, thus contributing to the growing body of literature of egocentric ERGM applications. Our model specifications not only include dyad-independent terms (e.g. in Krivitsky and Morris 2017) but also degree and triadic effects (as in Krivitsky et al. 2019).

3.3 Methods

An ERGM specifies a probability distribution over a set of all networks of specified size (number of nodes):

$$P(Y = y) = \frac{exp(\theta^T g(y))h(y)}{k(\theta)}$$

where the probability of observing a particular network y is a function of a vector of network statistics g(y), θ is a vector of model coefficients to be estimated, h(y) is the reference measure defining the baseline behavior of the model when $\theta = 0$, and k(θ) is a normalizing constant ensuring that the sum of the probabilities over all y's is equal to 1. Specifying an ERGM for the given dataset primarily consists of proposing a set of network statistics, g(y), that the researcher hypothesised to correspond to important mechanisms responsible for the network structure. Once the model is fit, the estimated values of model coefficients, θ, indicate, roughly speaking, how the probability of a tie changes given that it is to form a particular structural feature encapsulated in the network statistic, ceteris paribus other effects present in the model.

As discussed in the preceding section, our approach builds on the methodologies outlined by Krivitsky and Morris (2017) and further developed by Krivitsky et al. (2019, 2022). In a nutshell, it is an approach to estimate an ERGM but, instead of computing its sufficient statistics directly from a "complete", sociocentrically-collected network data, they are estimated based on an egocentric sample. In applying this framework to our study, we have tailored the model specifications to align with the characteristics of our dataset (see the section "Data"). We estimate two specifications. The first contains the following effects:

- Degree heterogeneity effects related to the following attributes: age, education, gender, country of birth, occupational status, political views, and religion.
- Endogeneous effects: number of nodes with degree 0 (isolates), 1, 2, 3, and 4 (5 parameters in total), number of edges with the number of shared partners equal to 0, 1, and 2 (three parameters).

The second specification extends the first with the following assortativity effects:

- Modeled with absolute difference, related to age, education, and political views.
- Modeled with categorical uniform assortativity effects related to gender, country of birth, occupational status, and religion.

Together, these specifications provide a rather detailed representation of the interplay between individual attributes and the network structure, keeping in mind that the data comes from a nationally representative survey. We estimate the models via the Maximum Pseudo-Likelihood (MPLE)[2] method using Statnet and the package 'ergm.ego' (Krivitsky 2025).

The estimated models will be used in three ways:

1. Model coefficients have substantive interpretations regarding social mechanisms related to node and tie characteristics, e.g. assortativity, reciprocity. It is then of interest to investigate which important structural effects are responsible for the presence or absence of core discussion ties.

   To further understand the probability distribution over networks, we will draw a sample of 100 networks, each of a size equal to the sample size (5094 respondents). The distribution of various properties of these networks will be utilised in the remaining two investigations:
2. We compare the ERGM-generated distribution to Erdős–Rényi graphs in features such as transitivity and average path length. Apart from each being of interest in their own right, together they constitute a fingerprint of the Small World phenomenon, i.e. the combination of high transitivity (clustering) and low average path length is a symptom of that phenomenon. In this way, we verify whether the empirically estimated network distribution has the Small World property.
3. We analyze social cohesion in the ERGM-generated network. The concept of social cohesion is often related to the general extent of the "integration" in a society. As such, a lack of social cohesion can be related to the overall prevalence of individuals who are relatively less well connected with others. We use three indicators of social cohesion: closeness centrality, betweenness centrality, and the fraction of nodes outside of the largest connected component.

Analyses were performed using the R package 'igraph' (Csardi and Nepusz 2005).

3.4 Data

3.4.1 Sample

We used data from the Spanish General Social Survey, collected by the Spanish Center for Sociological Research (CIS) between the 11th and 30th of April, 2013. The data are openly accessible via the CIS website.[3] A large, multistage, representative sample of the national adult population was drawn and interviewed face-to-face about their

---

[2] While MPLE has known limitations—primarily leads to underestimated standard errors of the coefficients—we used it because the network size and model size made the preferred MCMCMLE method computationally and technically intractable, even on High Performance Computing hardware.

[3] Encuesta Social General Española (study number 2975), see www.cis.es.

core discussion networks (realised sample size $N = 5094$; 54% response rate; see CIS 2013).

The following name generator was used to collect the data about core discussion networks:

> From time to time, most people discuss important issues with others. Thinking back over the last six months, with which people have you talked with about issues that are important to you? Please tell me the first name, initials, or nickname of these people (or family relationship).

While respondents could mention more than five people, interviewers recorded the (nick)names or initials of the first five people they nominated as well as the total number of persons a respondent answered. Follow-up information on these ties was collected on 15,580 ties, of which we excluded 157 because they were ties to children.

After the names were elicited, respondents (“egos”) were asked follow-up questions about each of the (first) five nominated persons (“alters”), such as their gender, age, country of birth, years of education, occupation, religion, and political orientation, and about the relationships among each pair of those alters, which allows us to measure transitivity. Furthermore, respondents were asked to report their own sociodemographic attributes, which allowed us to measure the similarity between respondents and their network members as well as attribute-related degree heterogeneity.

3.4.2 Measures

The interviewers assessed respondents’ *sex* as man, woman, or don’t know. Respondents were asked to report the sex of each nominated alter. Using this information, we define egos’ and alters’ sex as binary variables.

Ego’s *age* is a continuous variable, calculated as the survey year minus their birth year, while each alter’s age was reported by ego in years.

Respondents were also asked about their own *country of birth* and that of their nominated alters. Based on their responses, we defined egos’ and alters’ countries of origin as binary variables, differentiating between people born in Spain and abroad.

To define egos’ and alters’ *level of education*, we rely on responses to an open question about egos’ highest level of completed education and a closed question on each alter’s highest level of completed education. CIS used other response categories for ego than for alter. We unified the responses by transforming them into years of education, which is the typical number of years needed to complete each level of education.

Information on egos’ *occupation* is based on an open question about egos’ employment status (e.g., working, retired, unemployed, student) and, if they work, their jobs. For alters, respondents were first asked about alters’ employment status, and if they worked or had worked before, open questions enquired about alters’ jobs. CIS coded the answers to the occupation questions according to the National Classification of Occupations (CNO-11). Combining employment status and occupation, we differentiate between five categories: Unemployed, employed in a high-level job (CNO-11 levels 1–3), employed in a middle- and low-level job (CNO-11 levels 4–9), student, and inactive or other.

Respondents were further asked about their *religion* (Catholic, believer of another religion, non-believer, agnostic, or atheist) and that of their nominated alters (Catholic, believer of another religion, not religious/atheist/agnostic, don't know). Based on this information, we define egos' and alters' religion in three categories: Catholic, other religion, and no religion.

Finally, we define ego's *political orientation* based on respondents' self-placement on an 11-point left–right scale ("When talking about politics, the terms "left" and "right" are commonly used. On this card there is a series of boxes from left to right. In which box would you place yourself?"). They reported their perceptions of alters' political orientation on the same scale.

For *network structure*, respondents were asked to indicate, for each pair of alters, "Do [name person 1] and [name person 2] know each other, or do they not know each other at all, that is, they wouldn't recognise each other if they met on the street?". For those pairs of alters for which respondents reported a tie, they were additionally asked, "And do they have an especially close relationship among them, or is their relationship not especially close?". We coded alter-alter ties 1 if they were especially close and 0 if they did not know each other or did not have a close relationship, to ensure that the relationship between alters was similar in strength as ego-alter ties.

The sample consists of slightly more women than men (51% vs. 49%), reflecting their population prevalences in Spain. Respondents were between the ages of 18 and 97 (average age: 50 years) of whom the large majority was Catholic (77%) and born in Spain (91%). The average number of years of education is just over nine years, and most respondents were employed in high- (22%) or low-level employment (27%) or are inactive (26%).

Table 3.1 gives the descriptive statistics of all ego-alter pairs in the empirical data. The average number of ties respondents reported is 5.4, and most respondents have fewer than ten alters, but a few respondents nominated many more contacts. Follow-up questions were asked on a maximum of 5 alters, which had a mean number of 3.03. The set of egos is on average quite similar to the set of alters in these characteristics (see Table 3.1), except for education, employment, and religion, showing that respondents tended to nominate people who received on average three more years of education than they had received, were more often employed, and were more often Catholic. The higher educational and employment status of alters could possibly be explained by the nature of the relationships under study: People might prefer more resourceful individuals as their confidants. Since these are close ties, we see recall bias as a less likely, but potentially also valid, explanation.

For network modeling presented in the "Results" section, we imputed the missing values once using multivariate imputation by chained equations (van Buuren 2018).

Table 3.1 Summary table of descriptive statistics of the original data, at the tie level

Variables		Attributes of egos	Attributes of alters	*N* ego-alter pairs for which is available	
				Ego attribute	Alter attribute
Sex	*N* (%) man	2460 (48.3%)	7153 (47.0%)	15,657	15,222
	N (%) woman	2629 (51.7%)	8069 (53.0%)		
Age: Mean (*SD*)		48.7 (17.9)	46.3 (15.8)	15,665	15,063
Birth country:	*N* (%) abroad	447 (8.79%)	1113 (7.4%)	15,649	15,126
	N (%) Spain	4640 (91.2%)	14,013 (92.6%)		
Education: Mean (*SD*)		8.59 (4.21)	12.5 (4.7)	3169	14,554
Occupation:	*N* (%) employed high level	991 (19.6%)	4056 (27.1%)	15,539	14,942
	N (%) employed low level	1333 (26.4%)	4398 (29.4%)		
	N (%) student	258 (5.11%)	792 (5.3%)		
	N (%) unemployed	942 (18.6%)	1998 (13.4%)		
	N (%) inactive and other (non-student)	1528 (30.2%)	3698 (24.7%)		
Religion	*N* (%) Catholic	3859 (76.8%)	11,663 (79.4%)	15,439	14,682
	N (%) not religious	969 (19.3%)	2560 (17.4%)		
	N (%) other	197 (3.92%)	459 (3.1%)		
Political orientation: Mean (*SD*)		4.47 (2.09)	4.7 (2.1)	12,058	11,053
Number of ties (no upper limit): Mean (*SD*)		4.04 (4.37)	–	15,637	–

3.4.3 Group Differences and Assortativity

Two important features of real-life social networks are the variation in network structural features such as degree or closure by sociodemographic group (e.g., age, gender, or country of birth), and the tendency to associate with people who are similar to oneself in sociodemographic characteristics (assortativity). This is also the case in these data. People born in Spain, women, and highly educated individuals tend to have larger personal networks than people born abroad, men, and lower-educated individuals. The proportion of individuals with fully transitive networks is highest for inactive individuals and employed individuals with low occupational prestige (see Fig. 3.1), and lowest for individuals with high occupational prestige.

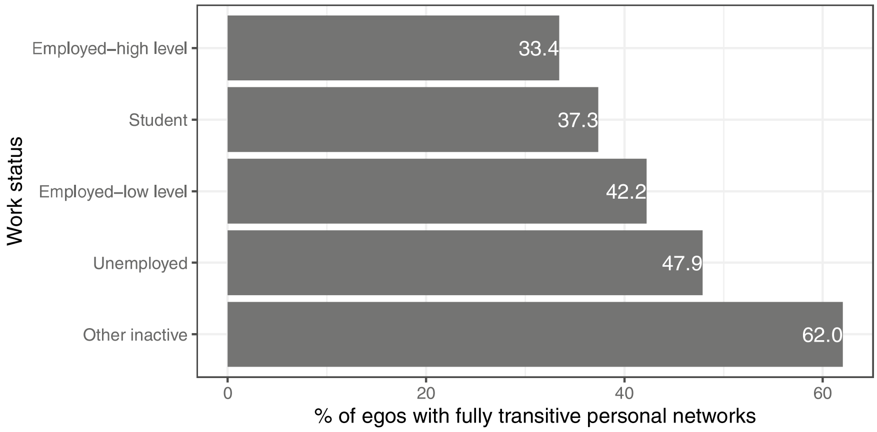

Fig. 3.1 Proportion of nodes with maximally transitive personal networks broken down by ego occupation

Regarding assortativity, individuals from different groups mix fairly well. Still, especially for higher status occupation and higher education of actors, we can see that for most categories of egos the largest fraction of alters belong to the same category, which suggests assortativity (see Fig. 3.2). Only a tiny proportion of high-level employed egos have core discussion alters who are unemployed or inactive. Similarly, the majority of alters of egos with primary education also have primary education. Assortativity is even stronger for egos with higher education who have very few alters with primary education. The ERGM we discuss in the next section will provide a more stringent test for this pattern.

3.5 Results

We first present the ERGM model estimated based on the survey data and its goodness of fit. The model was fit in two versions: on ego-alter data only, and on data including alter-alter ties. The results of the latter model are presented and discussed, unless specified otherwise. The ERGM model was then used to simulate a sociocentric network with the same size as the survey sample ($N = 5094$). The next sections compare the simulated networks with raw empirical data and pure random graphs. Finally, we describe the simulated networks in terms of social cohesion.

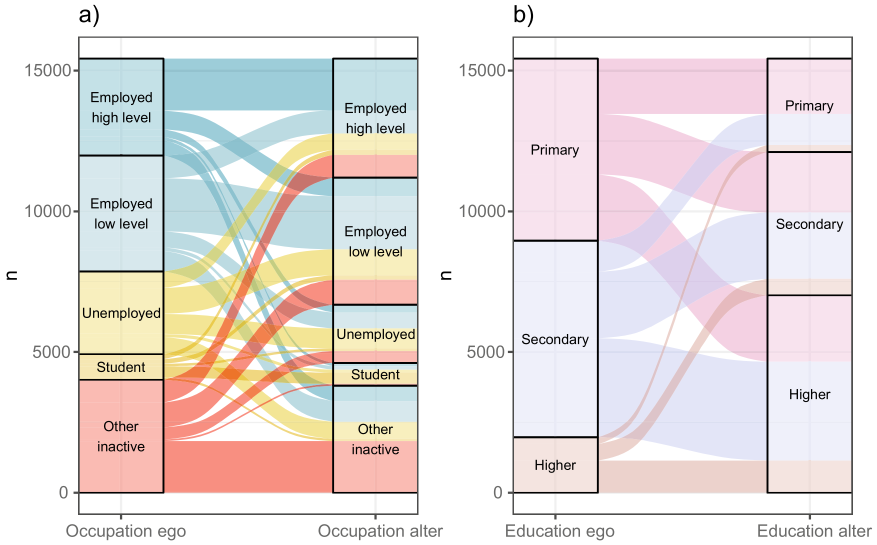

Fig. 3.2 Alluvial diagrams of ego-alter pairs by characteristics of ego (left hand side) and characteristics of alter (right hand side) for **a** work status and **b** educational level

3.5.1 ERGM

The ERGM, fitted to the Spanish data, largely confirms the descriptive patterns (see Figs. 3.1, 3.2 and Table 3.1). Older people tend to have more discussion ties, and confidants tend to be close in age to each other. Highly educated people tend to have more ties than lower-educated individuals, but these confidants do not necessarily come from the same educational backgrounds (positive distance coefficient indicating heterophily). When assortativity tendencies are accounted for, including gender assortativity, women appear to have less discussion ties. People born in Spain discuss important matters with more people, especially with Spaniards, while people born abroad tend to be surrounded by other foreign-born discussion partners. Students do not have significantly more (or fewer) confidants compared to people employed at more prestigious positions. Every other employment category has more confidants compared to people employed at more prestigious positions. At the same time, there is assortativity by occupational status. The number of confidants is not significantly related to ego's political views, however there is assortativity on this dimension. Finally, both Catholics and people of other religions have more discussion ties than nonreligious people, and alters tend to belong to the same religion category as egos.

The model fits the data reasonably well (see Fig. 3.9 in the Appendix). It slightly overestimates the number of isolates and egos who name four and more confidants, and it tends to overestimate smaller values of edgewise shared partners distribution. Substantively, this means that the model underestimates the tendency of egos to discuss important matters with only two or three people, and the tendency of alters to discuss important matters between each other (triadic closure).

Using the ERGM parameters from Fig. 3.3, we simulated 100 complete networks representing the society-wide structure of core-discussion networks in Spain on a smaller scale. To get an indication of the simulation goodness of fit, we compare the simulated networks to egocentric survey data (Fig. 3.10). We see from the transitivity distribution of egos in the simulated networks that a high number of egocentric networks have either complete transitivity or complete lack thereof (blue lines in Fig. 3.10). A smaller proportion has transitivity scores in between. This might be partly due to comparably small network sizes that only allow for specific values and increase the likelihood of having either a complete ego-network or one where no alters are connected. The egocentric transitivity distribution in the survey data follows a similar pattern, however with a much larger proportion having complete networks and fewer having scores between 0 and 1 (red line in Fig. 3.10).

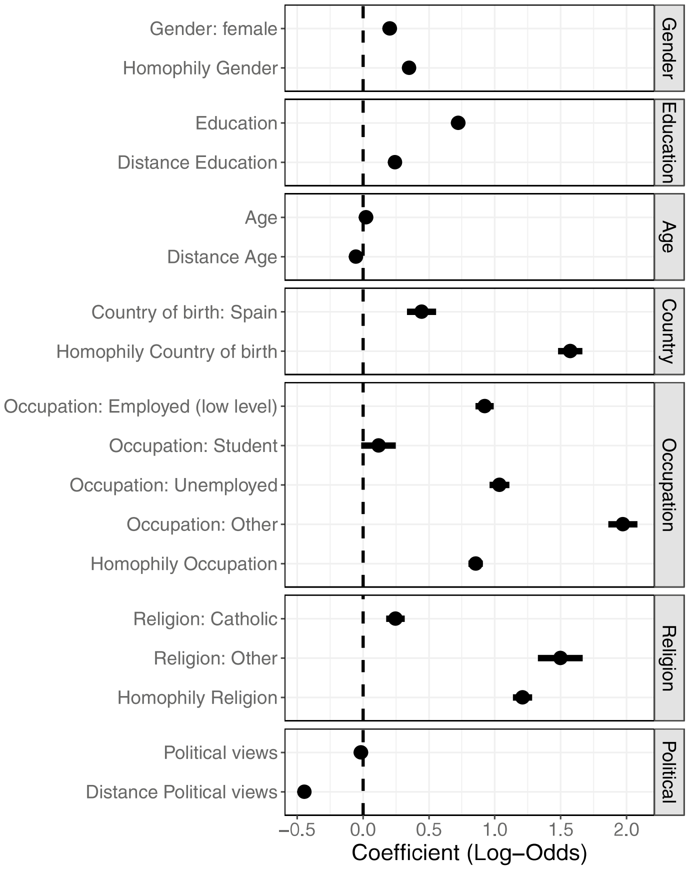

Fig. 3.3 Coefficients of node-related model terms

3.5.2 Topological Properties

Next, we compare the 100 simulated networks to 100 random networks with a similar size and density, as a ‘neutral’ null model, to evaluate structural differences, including group segregation. We see that in the empirically based simulation, more realistic network structures occur. Other than the classic ‘hairball’ of a random network, we observe more cohesive structures of many smaller clusters that are connected by few ties to form a larger component (Fig. 3.4). Zooming into network neighbourhoods (Fig. 3.5), the different smaller communities and the bridges in between become more apparent.

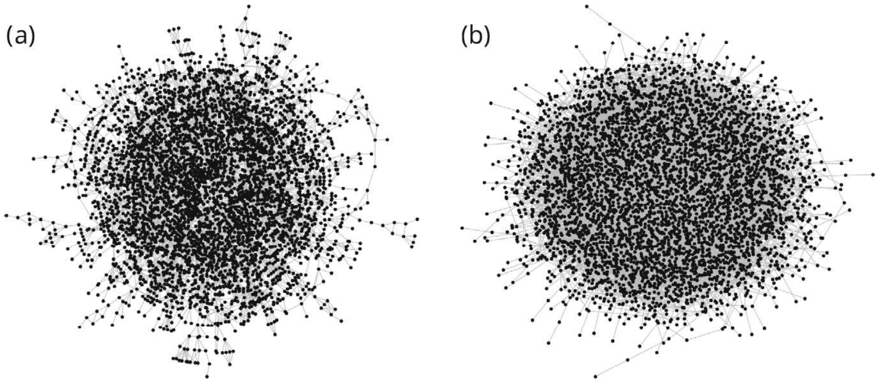

Fig. 3.4 Examples of the giant components of **a** an empirically-calibrated simulated network and **b** an Erdős–Rényi random network of the same size and density. Both figures use the stress majorization layout (Gansner et al. 2005)

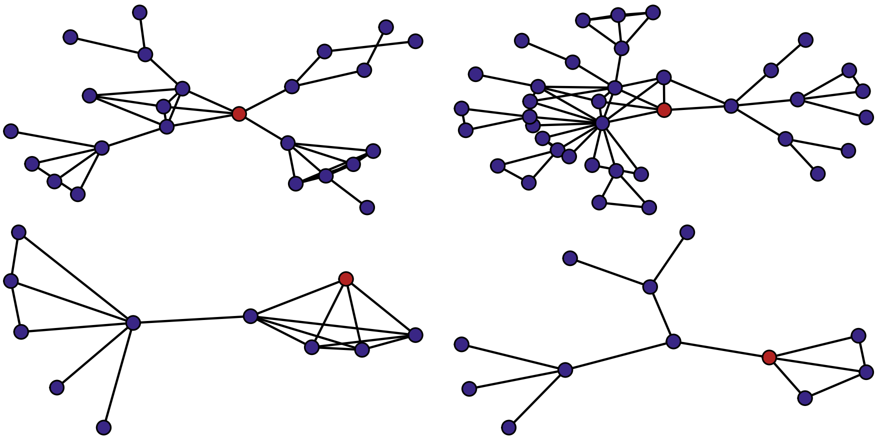

Fig. 3.5 Four examples of three-step neighborhoods in the empirically calibrated network simulation. *Note* In red, the focal node whose ego-network the neighborhood is formed around

We then compare the 'small-world' properties, i.e. the average clustering and shortest paths, to Erdős–Rényi models of similar size and density, as well as empirical large-scale networks. When alter-alter ties are not included in the model and only nodal and dyadic effects are used to calibrate the network simulation, the created networks show higher average transitivity and shorter average paths than random networks of similar size and density. In this scenario, the social network of the Spanish General Social Survey, representing the Spanish population, resembles a small-world more than expected by chance. However, when alter-alter ties are added and triadic closure effects are a part of the model, the shortest paths become much longer on average. This may be due to the lower number of isolates.

We next compare this result to complete society-level networks, namely the Dutch (Bokányi et al. 2023) and Danish register-based multilayer population network (Cremers et al. 2024). We see that the observed shortest path lengths correspond closely to the distances found among extended family members (around 10 on average in the Netherlands and 12 in Denmark). Empirical networks of neighbors, classmates, and colleagues have substantially shorter distances (around 5–6 on average in both countries). Similar short distances occur in online friendship networks, where the effective diameter (the smallest number of steps in the network at which at least 90% of all connected pairs of nodes can be reached) is around 7 (e.g., Corten 2012).

Therefore, the society-wide core-discussion network of people in Spain has higher transitivity but also longer shortest paths than random networks (in contrast to what Watts and Strogatz 1998, suggested), thus cannot be considered to have the small-world property. Comparison with register-based networks of family ties shows that our network shows analogous properties, therefore can be considered to capture relatively strong ties. This also serves as further evidence that networks of strong ties in general show less small-world properties.

Departing from the macro-comparison of ERGM-based simulated networks and random networks, we analyse potential differences in the position and embeddedness of different groups. After randomly adding node attributes with the same baseline distribution as the empirical data to the random networks, we can compare segregation patterns in empirically informed and artificial data. For gender (panel (a) in Fig. 3.6), we find higher segregation in the empirically informed network than in a random network, while for education, we find lower levels of segregation than in random networks (panel (b) in Fig. 3.6).

3.5.3 Social Cohesion by Centrality and Isolation

Finally, we interpret the simulated networks in terms of social cohesion. We compare the location of selected social groups in two ways. In the first approach, we compare the harmonic centrality and betweenness centrality scores in each group. Average closeness and betweenness centrality scores might indicate whether actors in some groups occupy more prominent positions in society. Moreover, the two indicators serve as an example of positional characteristics that are computed using a whole

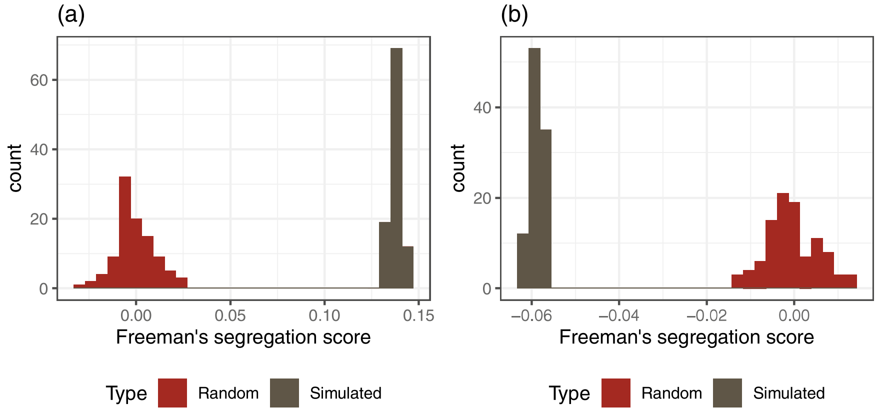

Fig. 3.6 Freeman's segregation scores for **a** gender and **b** educational levels in random networks (red) compared to empirically calibrated network simulations (gray)

network rather than personal networks only, which is impossible with the raw egocentric data we work with. In the second approach, we compare the fractions of a given social group outside of the largest connected component. We interpret this as an indicator of the extent to which members of different groups are "excluded" from the society, which points at a relative lack of social cohesion.

The differences in harmonic centrality[4] and betweenness centrality between native-born and foreign-born individuals are minimal (see Fig. 3.7). In both cases distributions have largely similar shapes suggesting that the positions of people from these two groups are by and large similar in terms of their centralities. Still, network-wide median values are slightly higher for people born in Spain. The variation in medians between the 100 simulated network is smaller than the difference between the two groups implying that the difference might not be substantively convincing, but is nevertheless statistically significant.

There are different relational structures on the periphery of the graph, outside of the largest connected component. The majority of nodes there still reside in components of sizes up to 15 (Fig. 3.13), which in some cases reflects a well-connected group of people. Nevertheless, isolates are the most frequent component pattern while dyads are also present, which means that there are people who have only one confidant, if any.

In the following, we test if any of the social categories are overrepresented in the smaller components on the periphery of the network. Figure 3.8 shows what share of the corresponding category ends up outside the largest connected component (LCC), and how these values are distributed across 100 simulated networks. If members of all categories were equally likely to appear on the network peripheries, there would

---

[4] We opted for using harmonic centrality (sum of inverse distances) rather than classical closeness centrality (inverse of the sum of distances) because the former is better at handling isolates while describing the same phenomenon—reachability of other nodes from the focal node.

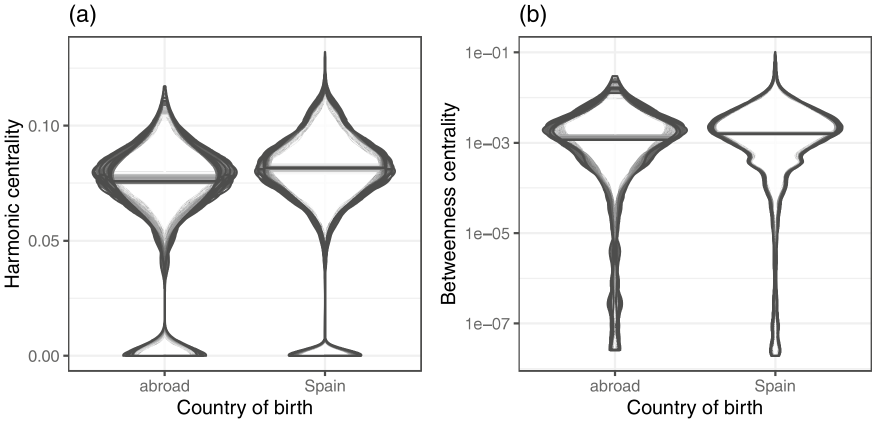

Fig. 3.7 Violin plots of **a** harmonic centrality and **b** betweenness centrality people born in Spain and abroad. Overplotted violins and horizontal lines show distributions and their medians from simulations from the fitted model

be no difference in the distribution shapes. In Fig. 3.8 we see that this is not the case. Simulations suggest that approximately 8% of people born in Spain occupy network positions outside of the largest connected component while for people born abroad this percentage is roughly twice as large. We also see that the unemployed and other people inactive on the labour market are more often in the network peripheries than the employed and the students.

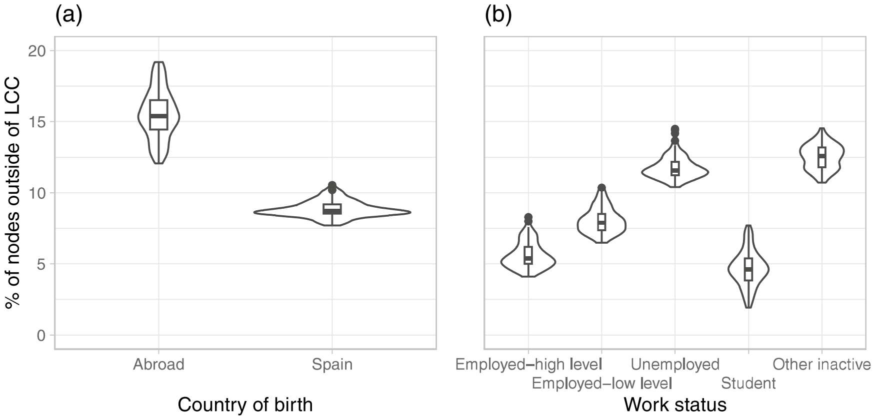

Fig. 3.8 Percentage of nodes outside of the largest connected component by **a** country of birth and by **b** work status, across 100 simulations. The Y axis is common for both panels

3.6 Discussion

The simulations presented in this study mark a significant step toward more realistic approximations of society-wide core discussion networks, especially in the context of further analysis such as with agent-based models. We contribute to the body of existing applications of Exponential Random Graph Models (ERGMs) to model egocentrically-sampled network data, by simulating society-wide networks using information on nodal, dyadic and triadic network features. We construct and analyze the properties of large-scale core discussion networks of people in Spain based on a unique ego-network dataset. We also release our empirically-calibrated simulated networks, which can hopefully serve as a baseline with included variation for agent-based models that concern core discussion networks and related phenomena.[5]

The simulated society-wide core discussion networks have notable topological properties. Comparison to random networks showed that our networks have higher transitivity but also substantially longer average shortest paths, which means that they do not exhibit small-world properties (where distances between nodes are shorter). Comparison with society-wide register-based networks showed that topologically, our networks look most similar to networks of family ties, where distances are as long. Therefore, even though previous research has shown that approximately half of the people with whom people discuss important matters are not their closest relationships (Small 2013), we find that large-scale networks of discussion ties are more similar to networks of strong family ties than of colleagues or online friends. An alternative explanation for the longer distances in our networks could be that we did not capture important discussion relationships with the collected data, or the method did not allow us to generate important macro-level structures. Should this be the case, it could be related to the fact that our simulations remain constrained by their focus on local effects. When modeling networks spanning larger populations it seems worthwhile to also empirically measure more complex network structures beyond triadic level. We believe that it remains an important challenge for future research. Importantly, longer distances appeared in the simulations made with the model version that included information about alter-alter ties and thus transitive closure. Without this data, distances in simulated networks were shorter. This has crucial implications for data collection strategies with the aim of generating large-scale networks from ego-network data. When simply using single-item questions on ego-centric network sizes, the resulting networks might misrepresent key aspects of cohesion. Should such networks be further used in models of diffusion (Valente and Vega Yon 2020) or complex contagion (Centola and Macy 2007), overestimated degree of small-worldness might bias model results. Researchers must carefully consider how to balance data collection efforts and model accuracy.

Through the ERGMs' assortativity estimates, we obtained a measure of net network segregation. This measure partially accounts for endogenous network mechanisms (Goodreau et al. 2009) and by-product effects (Smith et al. 2014). However, it

---

[5] The simulated network datasets are available at https://osf.io/v52wt/.

is crucial to recognise that these estimates may still be influenced by various mechanisms, including individuals' preferences for similarity (McPherson et al. 2001), the higher likelihood of interaction among similar others due to proximity (Kruse et al. 2016), or other network amplification effects not captured by the model (Wimmer and Lewis 2010). The inclusion of kinship relations and geographic information would be particularly valuable in future analysis and simulations.

The analysis of the distribution of complete networks simulated from the model with tools capturing aspects of social cohesion demonstrates two elements. Firstly, how the information in the egocentrically collected data can be used to make inferences about macro network features—such as betweenness and closeness scores or component sizes. Secondly, how one can make inferences about relative positions of different groups, such as natives and migrants, in a society-wide network using such seemingly limited data. Our analysis suggests that people residing in Spain but born abroad occupy positions of lower centrality and are relatively more often in the peripheries of the society-wide network compared to people who were born in Spain.

One of the primary features and limitations of the presented method is that attempts at utilizing the network information available in an egocentric sample of personal networks. It is a lot of information: about attribute mixing and assortativity and about degree distribution and certain types of social closure. This however is by no means exhaustive. Egocentric data, and thus the models for such data, are not able to represent any higher-order network features such as community structure unrelated to observed attributes or over- or under-prevalence of higher-order cycles. It is also important to measure, both ego- and alter-levels, as many nodal attributes relevant for tie formation as possible to avoid omitted variable problems.

Presented methods rely on, among other things, assumptions regarding the collected personal networks. In particular, that (1) ego-alter and alter-alter ties pretend to be the same type of social relationship, (2) all the ties are undirected and reported as such by both actors, and finally (3) that these reports are all 100% reliable. In other words, all the reports on ego-alter and alter-alter ties are taken at face value. One can argue that this is not realistic especially with respect to the alter-alter ties—for example ego's reports on ties of a socially distant alter might be less reliable than her reports on ties of a socially close alter. We believe that this asks for future research on improving the methods of data collection and incorporating reliability assessment into statistical models based on personal network data.

Another avenue of future research on the presented methods should look into applying it to other relationships than core discussion ties, like most intimate ties, acquaintanceships, or friendships. Previous research has shown that approximately half of the people with whom people discuss important matters are not their closest relationships (Small 2013), and that such relationships can change rapidly (Small et al. 2015). A comparative analysis with the simulated core discussion network could extend knowledge of society-wide relationship patterns, including social cohesion.

Acknowledgements We thank Aswathi Padmavilochanan for her contributions to the project during the BIGSSS summer school. Michał Bojanowski and Miranda Lubbers are grateful for the funding of the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (Grant Agreement No. 101020038; “A network science approach to social cohesion in European societies”; PATCHWORK). Furthermore, Miranda Lubbers is grateful for the support of the ICREA Acadèmia program (2021-2025). Paul Schuler is grateful for the funding through the Scottish Graduate School of Social and Political Science International Award scholarship (ESRC Grant Ref: ES/P000681/1). Susanne Böller is grateful for funding of the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation, 458745130). Alla Loseva is grateful for funding of the Dutch Research Council (NWO, 406.18.SW.038).

Appendix

See Figs. 3.9, 3.10, 3.11, 3.12, 3.13 and Table 3.2.

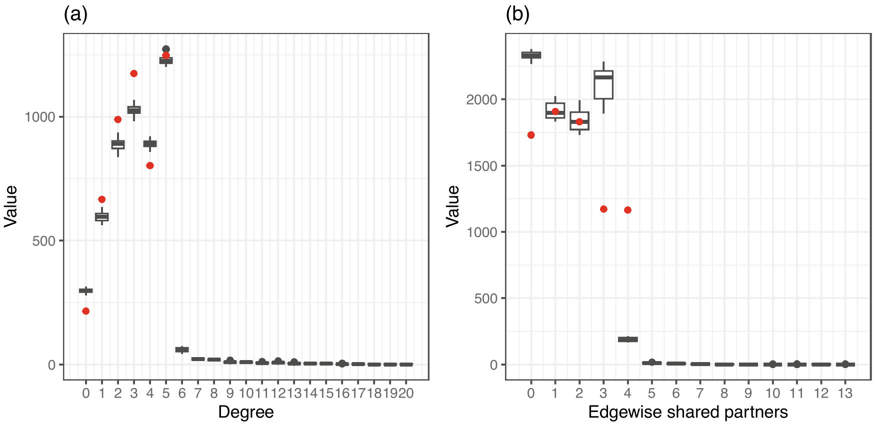

Fig. 3.9 ERGM fit for **a** degrees and **b** edgewise shared partners. Red dots indicate empirical values, boxplots show distributions over 100 simulated networks

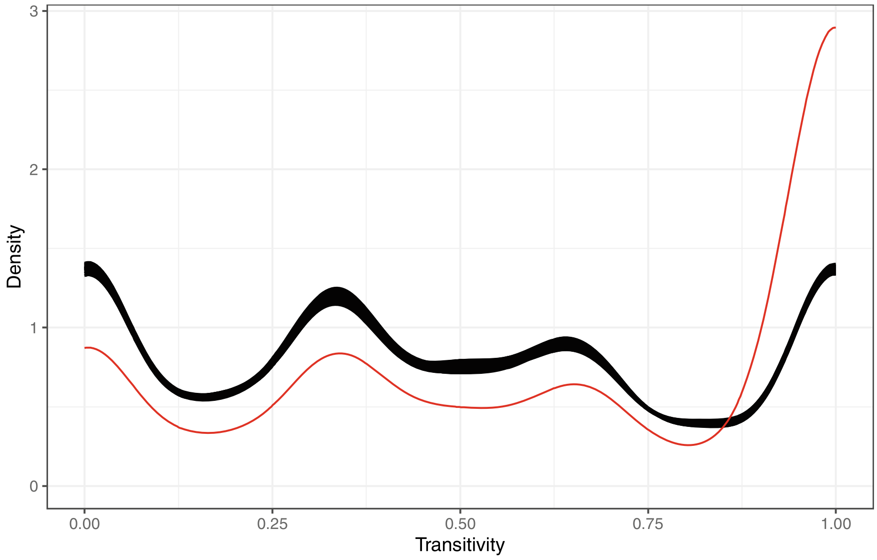

Fig. 3.10 Density plot of ego transitivity scores across 100 empirically-calibrated networks (blue lines) and the transitivity scores of the egocentric transitivity scores in the original survey data

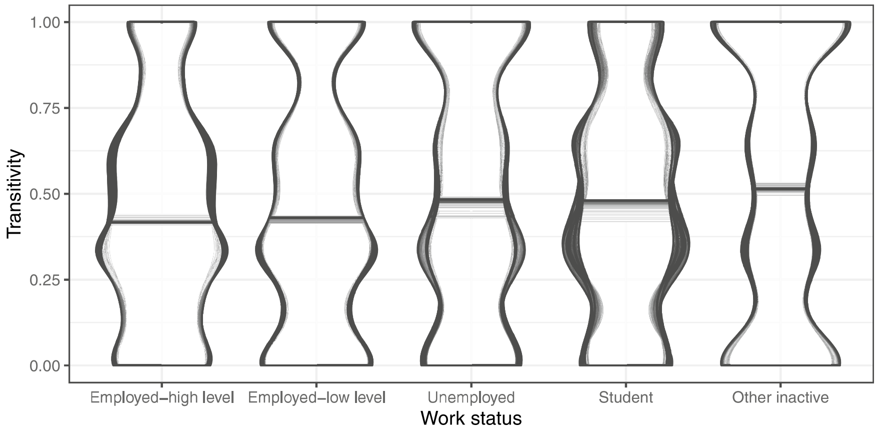

Fig. 3.11 Violin plots for transitivity by ego's employment

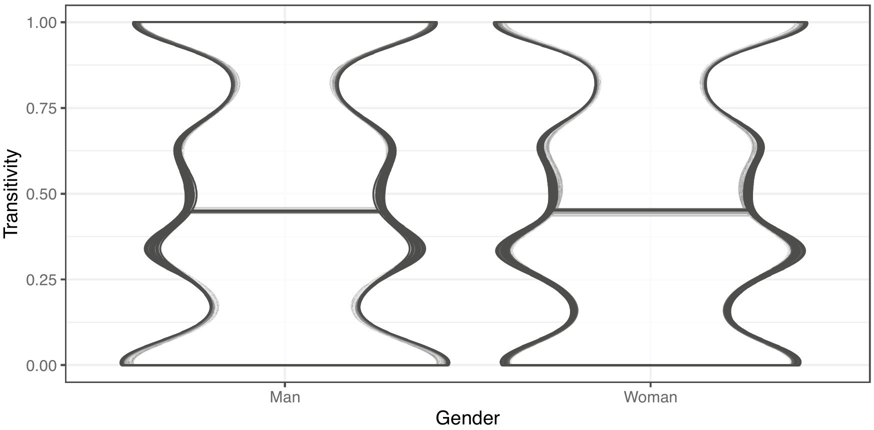

Fig. 3.12 Violin plots for transitivity by ego's gender

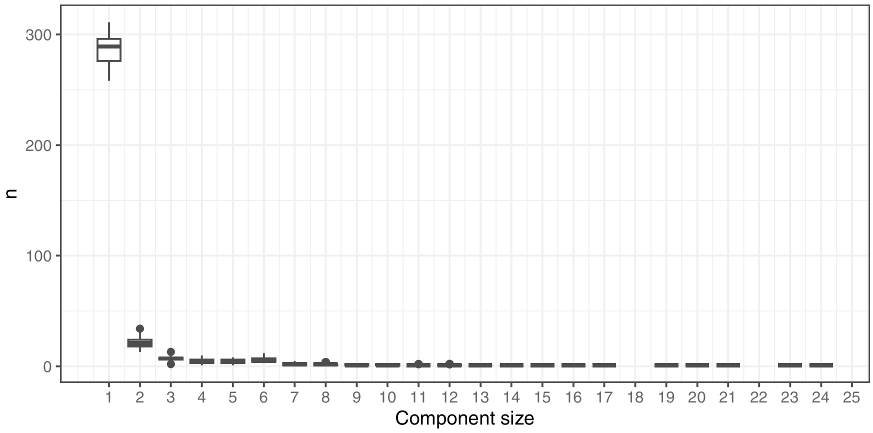

Fig. 3.13 Distribution of frequencies of smaller component sizes across 100 simulated networks

Table 3.2 ERGM results

Term	Model 1. Base			Model 2. Base + assortativity		
	β	SE	*t*	β	SE	*t*
Offset	− 8.54***	0.00	-Inf	− 8.54***	0.00	-Inf
Edges	− 81.94***	8.91	− 9.20	− 31.03***	1.53	− 20.32
Isolates	− 100.04***	10.68	− 9.37	− 40.09***	1.69	− 23.67
Degree: 1	− 72.40***	7.78	− 9.31	− 29.82***	1.26	− 23.60
Degree: 2	− 48.03***	5.11	− 9.40	− 20.28***	0.86	− 23.60
Degree: 3	− 27.39***	2.96	− 9.26	− 12.14***	0.51	− 23.66
Degree: 4	− 12.91***	1.34	− 9.66	− 6.03***	0.24	− 25.44
Edgewise shared partners: 0	− 4.76***	0.50	− 9.43	− 3.26***	0.11	− 29.17
Edgewise shared partners: 1	− 4.78***	0.38	− 12.64	− 3.64***	0.09	− 42.05
Edgewise shared partners: 2	− 2.69***	0.22	− 12.18	− 1.79***	0.05	− 35.83
Node characteristics						
Age in years	0.05***	0.01	8.70	0.02***	0.00	10.68
Education in years	1.90***	0.20	9.49	0.72***	0.03	20.66
Gender—woman (ref.: Man)	− 0.13	0.09	− 1.42	0.20***	0.04	4.64
Country of birth—Spain (ref.: Abroad)	3.13***	0.36	8.61	0.44***	0.11	4.02
Occupation—low level (ref.: High level)	2.49***	0.28	8.98	0.92***	0.07	13.32
Occupation—student	− 0.91**	0.29	− 3.08	0.12	0.13	0.90
Occupation—unemployed	3.06***	0.34	8.91	1.03***	0.08	13.69
Occupation—inactive	5.23***	0.57	9.18	1.97***	0.11	17.73
Political views (0–10)	− 0.01	0.03	− 0.39	− 0.02	0.01	− 1.60
Religion—Catholic (ref.: None)	1.93***	0.25	7.83	0.25***	0.07	3.51
Religion—other	3.67***	0.52	7.04	1.50***	0.17	8.81
Assortativity						
Distance: age in years				− 0.05***	0.00	− 24.52
Distance: education in years				0.24***	0.01	18.79
Assortativity: gender				0.35***	0.04	9.32
Assortativity: country of birth				1.57***	0.09	16.97
Assortativity: occupation				0.86***	0.05	15.56

(continued)

Table 3.2 (continued)

Term	Model 1. Base			Model 2. Base + assortativity		
	β	SE	*t*	β	SE	*t*
Distance: political views				− 0.45***	0.02	− 21.52
Assortativity: religion				1.21***	0.07	16.80

Note $N = 5094$. Asterisks denote statistical significance: $p < 0.05$ (*); $p < 0.01$ (**); $p < 0.001$ (***)

References

R. Albert, A.-L. Barabási, Statistical mechanics of complex networks. Rev. Mod. Phys. **74**(1), 47–97 (2002). https://doi.org/10.1103/RevModPhys.74.47

F. Amblard, A. Bouadjio-Boulic, C. Sureda Gutiérrez, B. Gaudou, Which models are used in social simulation to generate social networks? A review of 17 years of publications in JASSS. Winter Simulat. Conf. (WSC) **2015**, 4021–4032 (2015). https://doi.org/10.1109/WSC.2015.7408556

F. Amblard, G. Deffuant, The role of network topology on extremism propagation with the relative agreement opinion dynamics. Physica A **343**, 725–738 (2004). https://doi.org/10.1016/j.physa.2004.06.102

R. Axtell, Effects of interaction topology and activation regime in several multi-agent systems, in *Multi-Agent-Based Simulation*, ed. S. Moss, P. Davidsson (Springer, 2001), pp. 33–48. https://doi.org/10.1007/3-540-44561-7_3

A.-L. Barabási, R. Albert, Emergence of scaling in random networks. Science **286**(5439), 509–512 (1999). https://doi.org/10.1126/science.286.5439.509

C.L. Barrett, R.J. Beckman, M. Khan, V.S.A. Kumar, M.V. Marathe, P.E. Stretz, T. Dutta, B. Lewis, Generation and analysis of large synthetic social contact networks, in *Proceedings of the 2009 Winter Simulation Conference (WSC)*(2009), pp. 1003–1014. https://doi.org/10.1109/WSC.2009.5429425

P.S. Bearman, J. Moody, K. Stovel, Chains of affection: the structure of adolescent romantic and sexual networks. Am. J. Sociol. **110**(1), 44–91 (2004). https://doi.org/10.1086/386272

P.M. Blau, Presidential address: parameters of social structure. Am. Sociol. Rev. **39**(5), 615–635 (1974). https://doi.org/10.2307/2094309

R. Boero, F. Squazzoni, Does empirical embeddedness matter? Methodological issues on agent-based models for analytical social science. J. Artif. Soc. Soc. Simulat. **8**(4), 6 (2005). https://jasss.soc.surrey.ac.uk/8/4/6.html

E. Bokányi, E.M. Heemskerk, F.W. Takes, The anatomy of a population-scale social network. Sci. Rep. **13**(1), 9209 (2023). https://doi.org/10.1038/s41598-023-36324-9

E. Bruch, J. Atwell, Agent-based models in empirical social research. Sociol. Methods Res. **44**(2), 186–221 (2015). https://doi.org/10.1177/0049124113506405

A. Burger, T. Oz, A. Crooks, W.G. Kennedy, Generation of realistic mega-city populations and social networks for agent-based modelling, in *Proceedings of the 2017 International Conference of The Computational Social Science Society of the Americas* (2017), pp. 1–7. https://doi.org/10.1145/3145574.3145593

D. Centola, M. Macy, Complex contagions and the weakness of long ties. Am. J. Sociol. **113**(3), 702–734 (2007). https://doi.org/10.1086/521848

E. Chattoe-Brown, S. Gabbriellini, *How to Improve Network Science: The Potential of (Empirically Calibrated and Validated) Agent-Based Modelling* (OSF, 2021). https://doi.org/10.31235/osf.io/bym74

R. Chetty, M.O. Jackson, T. Kuchler, J. Stroebel, N. Hendren, R.B. Fluegge, S. Gong, F. Gonzalez, A. Grondin, M. Jacob, D. Johnston, M. Koenen, E. Laguna-Muggenburg, F. Mudekereza, T. Rutter, N. Thor, W. Townsend, R. Zhang, M. Bailey, N. Wernerfelt et al., Social capital I: measurement and associations with economic mobility. Nature **608**(7921), 108–121 (2022). https://doi.org/10.1038/s41586-022-04996-4

CIS—Centro de Investigaciones Sociológicas, *Estudio 2975, Encuesta Social General Española, Abril –julio 2013: Informe metodológico* (2013)

J.-P. Cointet, C. Roth, How realistic should knowledge diffusion models be? J. Artif. Soc. Soc. Simulat. **10**(3), 5 (2007). https://www.jasss.org/10/3/5.html

R. Corten, Composition and structure of a large online social network in the Netherlands. PLoS ONE **7**(4), e34760 (2012). https://doi.org/10.1371/journal.pone.0034760

J. Cremers, B. Kohler, B.F. Maier, S.N. Eriksen, J. Einsiedler, F.K. Christensen, S. Lehmann, D.D. Lassen, L.H. Mortensen, A. Bjerre-Nielsen, *Unveiling the Social Fabric: A Temporal, Nation-Scale Social Network and its Characteristics* (2024) (No. arXiv:2409.11099). arXiv. https://doi.org/10.48550/arXiv.2409.11099

G. Csardi, T. Nepusz, The Igraph software package for complex network research. Int. J. Complex Syst. **1695**, 1–9 (2005). https://igraph.org

G. Deffuant, Comparing extremism propagation patterns in continuous opinion models. J. Artif. Soc. Soc. Simulat. **9**(3), 8 (2006). https://www.jasss.org/9/3/8.html

E. Durkheim, *The Division of Labor in Society* (MacMillan, New York, 1893)

S. Eubank, H. Guclu, V.S. Anil Kumar, M.V. Marathe, A. Srinivasan, Z. Toroczkai, N. Wang, Modelling disease outbreaks in realistic urban social networks. Nature **429**(6988), 180–184 (2004). https://doi.org/10.1038/nature02541

O. Frank, D. Strauss, Markov graphs. J. Am. Stat. Assoc. **81**(395), 832–842 (1986). https://doi.org/10.1080/01621459.1986.10478342

E.R. Gansner, Y. Koren, S. North, Graph drawing by stress majorization, in *Graph Drawing*, ed. by J. Pach. Springer (2005), pp. 239–250. https://doi.org/10.1007/978-3-540-31843-9_25

S.M. Goodreau, J.A. Kitts, M. Morris, Birds of a feather, or friend of a friend? Using exponential random graph models to investigate adolescent social networks. Demography **46**(1), 103–125 (2009)

D. Guilbeault, J. Becker, D. Centola, Complex contagions: a decade in review, in *Complex Spreading Phenomena in Social Systems: Influence and Contagion in Real-World Social Networks*, ed. by S. Lehmann and Y.-Y. Ahn (Springer International Publishing, 2018), pp. 3–25. https://doi.org/10.1007/978-3-319-77332-2_1

L. Hamill, N. Gilbert, Social circles: a simple structure for agent-based social network models. J. Artif. Soc. Soc. Simul. **12**(2), 3 (2009). https://www.jasss.org/12/2/3.html

M.S. Handcock, D.R. Hunter, C.T. Butts, S.M. Goodreau, P.N. Krivitsky, M. Morris, L. Wang, K. Li, S. Bender-deMoll, C. Klumb, M. Bojanowski, B. Bolker, C. Schmid, J. Cheng, A. Karami, *ergm: Fit, Simulate and Diagnose Exponential-Family Models for Networks* (Version 4.6.0) (Computer Software, 2023). https://cran.r-project.org/web/packages/ergm/index.html

P. Hedström, G. Manzo, Recent trends in agent-based computational research: a brief introduction. Sociol. Methods Res. **44**(2), 179–185 (2015). https://doi.org/10.1177/0049124115581211

K. Heß, J. Himmelspach, O. Reinhardt, A.M. Uhrmacher, A generalized network generation approach for agent-based models. Winter Simulat. Conf. (WSC) **2021**, 1–12 (2021). https://doi.org/10.1109/WSC52266.2021.9715377

D.R. Hunter, S.M. Goodreau, M.S. Handcock, Goodness of fit for social network models. J. Am. Stat. Assoc. **103**(481), 248–258 (2008). https://doi.org/10.1198/016214507000000446

N. Jiang, A.T. Crooks, H. Kavak, A. Burger, W.G. Kennedy, A method to create a synthetic population with social networks for geographically-explicit agent-based models. Comput. Urb. Sci. **2**(1), 7 (2022). https://doi.org/10.1007/s43762-022-00034-1

Y. Kazmina, E.M. Heemskerk, E. Bokányi, F.W. Takes, Socio-economic segregation in a population-scale social network. Soc. Netw. **78**, 279–291 (2024). https://doi.org/10.1016/j.socnet.2024.02.005

C.A. Klofstad, S.D. McClurg, M. Rolfe, Measurement of political discussion networks: a comparison of two “name generator” procedures. Public Opin. q. **73**(3), 462–483 (2009). https://doi.org/10.1093/poq/nfp032

P.N. Krivitsky, *ergm.ego: Fit, Simulate and Diagnose Exponential-Family Random Graph Models to Egocentrically Sampled Network Data* (The Statnet Project, 2025). https://statnet.org. R package version 1.1.2, https://CRAN.R-project.org/package=ergm.ego

P.N. Krivitsky, M. Bojanowski, M. Morris, *Inference for Exponential-Family Random Graph Models from Egocentrically-Sampled Data with Alter–Alter Relations* (Working Paper Nos. 08–19) (National Institute for Applied Statistics Research Australia, University of Wollongong, 2019). https://niasra.uow.edu.au/workingpapers/UOW255140.html

P.N. Krivitsky, M.S. Handcock, M. Morris, Adjusting for network size and composition effects in exponential-family random graph models. Statist. Methodol. **8**(4), 319–339 (2011). https://doi.org/10.1016/j.stamet.2011.01.005

P.N. Krivitsky, M. Morris, Inference for social network models from egocentrically-sampled data, with application to understanding persistent racial disparities in HIV prevalence in the US. Ann. Appl. Stat. **11**(1), 427–455 (2017). https://doi.org/10.1214/16-AOAS1010

P.N. Krivitsky, M. Morris, M. Bojanowski, Impact of survey design on estimation of exponential-family random graph models from egocentrically-sampled data. Soc. Netw. **69** (2022). https://doi.org/10.1016/j.socnet.2020.10.001

H. Kruse, S. Smith, F. Van Tubergen, I. Maas, From neighbors to school friends? How adolescents' place of residence relates to same-ethnic school friendships. Soc. Netw. **44**, 130–142 (2016). https://doi.org/10.1016/j.socnet.2015.07.004

S.-H. Lim, S. Lee, S.S. Powers, M. Shankar, N. Imam, *Survey of Approaches to Generate Realistic Synthetic Graphs* (No. ORNL/TM-2016/3) (Oak Ridge National Lab. (ORNL), Oak Ridge, TN, USA, 2016). https://doi.org/10.2172/1339361

J. Lorenz, Data-driven agent-based modeling in computational social science, in *Handbook of Computational Social Science* (Routledge, 2021), vol. 1, pp. 150–167

D. Lusher, J. Koskinen, G. Robins, *Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications* (Cambridge University Press, 2013)

P.V. Marsden, Core discussion networks of Americans. Am. Sociol. Rev. **52**(1), 122–131 (1987). https://doi.org/10.2307/2095397

I. McCulloh, J. Lospinoso, K.M. Carley, in *The Link Probability Model: A Network Simulation Alternative to the Exponential Random Graph Model* (SSRN Scholarly Paper 2729285) (2010). https://doi.org/10.2139/ssrn.2729285

M. McPherson, L. Smith-Lovin, J.M. Cook, Birds of a feather: homophily in social networks. Ann. Rev. Sociol. **27**, 415–444 (2001). https://doi.org/10.1146/annurev.soc.27.1.415

T. Menezes, C. Roth, Symbolic regression of generative network models. Sci. Rep. **4**(1), 6284 (2014). https://doi.org/10.1038/srep06284

T. Menezes, C. Roth, Automatic discovery of families of network generative processes, in F. Ghanbarnejad, R. Saha Roy, F. Karimi, J.-C. Delvenne, B. Mitra (eds.), *Dynamics On and Of Complex Networks III* (Springer International Publishing, 2019), pp. 83–111. https://doi.org/10.1007/978-3-030-14683-2_4

M. Menyhért, E. Bokányi, R. Corten, E.M. Heemskerk, Y. Kazmina, F.W. Takes, *Connectivity and Community Structure of Online and Register-based Social Networks*(2024). arXiv. https://doi.org/10.48550/arXiv.2406.17752

M. Rolfe, Social networks and agent-based modelling, in *Analytical Sociology*, ed. G. Manzo (Wiley, 2014), pp. 233–260. https://doi.org/10.1002/9781118762707.ch09

D. Schiefer, J. van der Noll, The essentials of social cohesion: a lliterature review. Soc. Indic. Res. **132**(2), 579–603 (2017). https://doi.org/10.1007/s11205-016-1314-5

J.A. Smith, Macrostructure from microstructure: generating whole systems from ego networks. Sociol. Methodol. **42**(1), 155–205 (2012). https://doi.org/10.1177/0081175012455628

M.L. Small, Weak ties and the core discussion network: why people regularly discuss important matters with unimportant alters. Soc. Netw. **35**(3), 470–483 (2013). https://doi.org/10.1016/j.socnet.2013.05.004

J.A. Smith, Global network inference from ego network samples: testing a simulation approach. J. Math. Sociol. **39**(2), 125–162 (2015). https://doi.org/10.1080/0022250X.2014.994621

J.A. Smith, J. Burow, Using ego network data to inform agent-based models of diffusion. Sociol. Methods Res. **49**(4), 1018–1063 (2020). https://doi.org/10.1177/0049124118769100

J.A. Smith, G.R. Gauthier, Estimating contextual effects from ego network data. Sociol. Methodol. **50**(1), 215–275 (2020). https://doi.org/10.1177/0081175020922879

M.L. Small, V. Deeds Pamphile, P. McMahan, How stable is the core discussion network? Soc. Netw. **40**, 90–102 (2015). https://doi.org/10.1016/j.socnet.2014.09.001

J.A. Smith, M. McPherson, L. Smith-Lovin, 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 (2014). https://doi.org/10.1177/0003122414531776

T.A.B. Snijders, P.E. Pattison, G.L. Robins, M.S. Handcock, New specifications for exponential random graph models. Sociol. Methodol. **36**(1), 99–153 (2006). https://doi.org/10.1111/j.1467-9531.2006.00176.x

T.A.B. Snijders, G.G. van de Bunt, C.E.G. Steglich, Introduction to stochastic actor-based models for network dynamics. Soc. Netw. **32**(1), 44–60 (2010). https://doi.org/10.1016/j.socnet.2009.02.004

S. Thiriot, *Small World is not enough: Criteria for Network Choice and Conclusiveness of Simulations*(2020). https://doi.org/10.48550/arXiv.2002.03717

S. Thiriot, J.-D. Kant, *Generate Country-Scale Networks of Interaction from Scattered Statistics* (2008) https://doi.org/10.48550/arXiv.2004.01031

T.W. Valente, G.G. Vega Yon, Diffusion/contagion processes on social networks. Health Educ. Behav. **47**(2), 235–248 (2020). https://doi.org/10.1177/1090198120901497

S. van Buuren, *Flexible Imputation of Missing Data*, 2nd ed. (Chapman and Hall/CRC, 2018)

F. van Tubergen, Ethnic boundaries in core discussion networks: a multilevel social network study of Turks and Moroccans in the Netherlands. J. Ethn. Migr. Stud. **41**(1), 101–116 (2015). https://doi.org/10.1080/1369183X.2014.886955

S. Wasserman, P. Pattison, Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and $p*$. Psychometrika **61**(3), 401–425 (1996). https://doi.org/10.1007/BF02294547

D.J. Watts, S.H. Strogatz, Collective dynamics of 'small-world' networks. Nature **393**(6684), 440 (1998)

A. Wimmer, K. Lewis, Beyond and below racial homophily: documented on facebook. Am. J. Sociol. **116**(2), 583–642 (2010). https://doi.org/10.1086/653658

M. Zachara, C. Piskor-Ignatowicz, Agent-based creation and simulation of artificial social networks and the analysis of their properties. Comput. Sci. Eng. **18**(4), 34–41 (2016). https://doi.org/10.1109/MCSE.2016.46