Applied Network Science

RESEARCH Open Access

Heuristic methods for synthesizing realistic social networks based on personality compatibility

Daniel A. O'Neil* and Mikel D. Petty

* Correspondence: dao0030@uah.edu
University of Alabama in Huntsville, Huntsville, AL 35899, USA

Abstract

Social structures and interpersonal relationships may be represented as social networks consisting of nodes corresponding to people and links between pairs of nodes corresponding to relationships between those people. Social networks can be constructed by examining actual groups of people and identifying the relationships of interest between them. However, there are circumstances where such empirical social networks are unavailable or their use would be undesirable. Consequently, methods to generate synthetic social networks that are not identical to real-world networks but have desired structural similarities to them have been developed. A process for generating synthetic social networks based on assigning human personality types to the nodes and then adding links between nodes based on the compatibility of the nodes' personalities was developed. Two new algorithms, Probability Search and Compatibility-Degree Matching, for finding an effective assignment of personality types to the nodes were developed, implemented, and tested. The two algorithms were evaluated in terms of realism, i.e., the similarity of the generated synthetic social to exemplar real-world social networks, for 14 different real-world social networks using 20 standard quantitative network metrics. Both search algorithms produced networks that were, on average, more realistic than a standard network generation algorithm that does not use personality, the Configuration Model. The algorithms were also evaluated in terms of computational complexity.

Keywords: Social networks, Network generation, Network metrics, Personality compatibility, Probability search, Compatibility-degree matching

Introduction and motivation

Social network analysis is the study of social structures and relationships. Built from the theoretical foundation of graph theory, social networks are formal mathematical structures, consisting in their simplest form of *nodes* corresponding to actors or agents, where actors or agents may be individual people or identifiable groups of people, and *links* between pairs of nodes corresponding to relations between them, where relations may be any type of contact or connection between the actors or agents the nodes represent (Knoke and Yang, 2008) (Scott 2000).

The study and use of social networks often begins from and depends on empirical social networks. Empirical social networks are obtained directly from the real-world group or organization they represent, by the process of investigators identifying the people in the group or organization of interest and determining if the relationships to be represented in the network exist between them. Empirical social networks obtained by observation are valuable, but there are issues with them. Empirical social networks can be difficult and expensive to obtain, especially if the process for doing so is manual, and consequently relatively few in number and less than comprehensive in covering the range of possible social networks. They may not be available in the size, in terms of number of nodes or links, that an investigator needs. And while obtaining social networks from social media or other digital sources is much easier today than in the past, such empirical networks can be vulnerable to malicious recovery of private information from them using de-anonymization methods (Narayanan et al, 2011) (Narayanan and Shmatikov, 2008).

Synthetic social networks, generated algorithmically rather than obtained empirically, can mitigate these issues. Given effective social network synthesis methods, a user could produce a set of synthetic social networks, individually non-identical but collectively with specific desired structural characteristics, including size. A set of multiple social networks could be used to systematically test a network analysis or visualization tool (Staudt et al., 2017), and would allow the deliberate introduction of deviations from the defining characteristics of the class of social networks for testing purposes (Tsvetovat and Carley, 2005). In addition, synthesizing social networks is an approach to anonymization, which may protect the privacy of the individuals represented in an empirical social network (Narayanan and Shmatikov, 2009). Researchers may use the synthetic social networks without privacy concerns and freely share them with other researchers to allow repeatable experiments (Zhou et al., 2008).

However, an arbitrary or random graph is unlikely to be suitable as a synthetic social network for any particular application. To be useful a synthetic social network must "approximate certain qualities or parameters found in the empirical data" (Tsvetovat and Carley, 2005). In other words, a useful synthetic social network must possess the structural characteristics expected for the class of social networks it is intended to exemplify, without being simply a copy of one of those networks. For brevity, a synthetic social network with the structural characteristics of a desired class of social networks, perhaps as measured by suitable quantitative network metrics, will hereinafter be described as *realistic*.

A number of synthetic social generation methods exist; several important ones will be described later. Broadly speaking, the existing methods are based on replicating structural characteristics of an exemplar network. Our goal in this work was to examine whether a network generation method based instead on personality compatibility between nodes (where the nodes are assumed to correspond to persons) could be effective. Social networks based on personality compatibility can be of significant interest to organizations that must organize teams of persons to interact and work effectively, especially in challenging circumstances. We sought to develop a capability to synthesize personality-based social networks for future space exploration missions and colonies. In such missions, crew compatibility will be essential, so a capability to model social network formation and camaraderie within such circumstances could be very useful to mission planners and analysts.

Given the large number of people participating in online social networks, such as Facebook and Twitter, it is unsurprising that much current social network research tends to focus on large networks. Often, web based networks are scale free and the thousands of links and nodes tend to result in similar metrics. The research presented in this article is focused on relatively small networks with 10 to 100 nodes. The real-world networks used as exemplars are drawn from a wide range of organizations, ranging from an accounting firm to a monastery.

Two algorithms able to automatically synthesize realistic social networks using personality compatibility are described and compared in this article. The algorithms are given as input a set of nodes of the desired size. The algorithms then assign, using distinctly different methods, a personality type to each node that can be used as the basis for stochastically generating links between the nodes. Link generation between a pair of nodes depends of the relative compatibility of the personalities assigned to the two nodes. Personality type compatibilities are encoded in a personality compatibility that is an input to the generation process. Because link generation is stochastic given a personality type assignment to the nodes, multiple non-identical social networks can be generated as needed from a single assignment once a suitable assignment has been found. The algorithms have been shown to generate synthetic social networks that are significantly more realistic, in terms of their structural properties as measured by a range of standard graph metrics, than social networks generated using a standard network generation algorithm that does not use personality, the Configuration Model. The generation process has been demonstrated to work with multiple personality compatibility tables, and is thus adaptable to different personality type models.

The remainder of this article is structured as follows: Section 2 provides background information about social network analysis. Section 3 is a brief survey of important related work. Section 4 explains the social network synthesis algorithms developed in this research. Section 5 describes the software implementation of the three algorithms and discusses their execution. Section 6 reports the results of testing and comparing the algorithms, including quantitative measures. Finally, Section 7 states the conclusions of this work and suggests possible future work.

Background

This section provides background information on graph theory and social network analysis, and explains the metrics that were used to measure networks' structural similarity.

Social network analysis

The details vary by specific application, but in their simplest form, in a social network the nodes may correspond to people in a group, organization, or population of interest. The presence of a link connecting two nodes represents some relationship, such as kinship, friendship, collaboration, or information exchange, between the people corresponding to the nodes the link connects. For example, social networks are used to represent social distance in (Li et al., 2018) and information spreading in (Bouanan et al., 2018). The study of the structural properties of such social networks can provides insight into the group, organization, or population it represents. As an example, Fig. 1 shows a real world social network found to exist within a corporate law firm in the northeastern United States (Lazega 2001).

Classes of social network

Not all social networks have the same structural characteristics and properties. Social networks that represent communications in terrorist organizations might be expected to differ in structure and activity from those that represent collaborations in a scientific

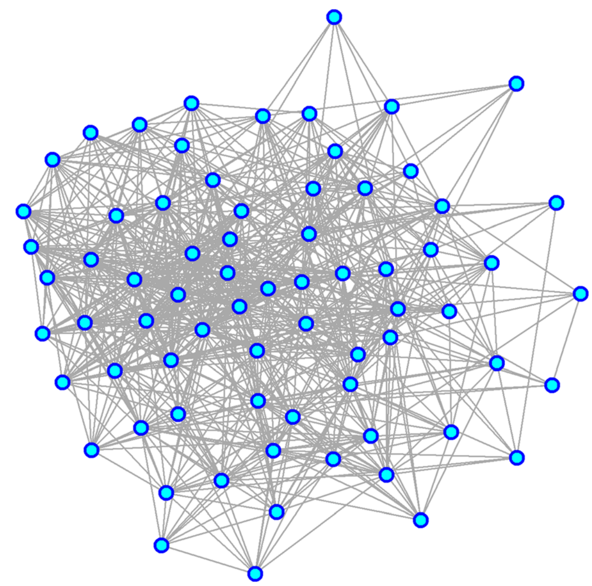

Fig. 1 Friendship within a law firm (Lazega 2001)

community. A set of social networks that represent instances of some well-defined category of group of organization will be termed a *class*. Some examples of classes of social networks are listed in Table 1; several of the examples in the table are based on (Easley and Kleinberg, 2010). The examples in Table 1 are all social networks, but intuitively they are not the same in terms of structure.

Note that in the last example in Table 1, the nodes of the social network correspond to organizations, not individual people. That example is included in order to draw attention to this distinction. This work focuses on social networks where the nodes correspond to people. The potentially different structure of an organizational-node network as compared to a people-node network will become of interest later.

A particular social network may be an element of one class, but not of another, by virtue of its structural properties. Therefore, two operations are of interest: (1) Membership; given a social network, how can it be tested for membership in a particular class of social networks? (2) Generation; given a description or example of a particular class of social networks, how can a synthetic social network that is a member of that class be generated? This work focuses on the second operation.

Table 1 Classes of social networks (Easley 2010)

Group or organization	Nodes	Possible link(s)
Terrorist organization	People	Communications Recruitment
High school student body	People	Romantic relationship Athletic teammates
Social club	People	Friendship Sponsorship
Employees of a corporation	People	Exchange of email Supervisory authority
Regional or national populace	People	Relatedness Transmission of infection
Financial system	Banks	Interbank loans Currency exchanges

Data structures and attributes for social networks

In the implementations described later, social networks were stored internally using *adjacency matrices* (Gersting 2014). More sophisticated data structures for social networks are available, but the networks used in this work were relatively small and simple adjacency matrices were sufficient. As for the attributes of the networks, two are important. First, networks may be *weighted* or *unweighted*. This work is concerned solely with the absence or presence of links, and therefore only unweighted networks were used. Second, networks may be *symmetric* or *asymmetric*. Links in symmetric networks typically represent mutual or two-way relationships, whereas links in asymmetric networks represent one-way relationships. This work is concerned solely with mutual relationships, and therefore symmetric networks were used.

Social network metrics

In this context, *metrics* are numerical measurements of a social network's structure. A wide range of different metrics are available. Graph theory provides a number of abstract metrics, sometimes known as graph invariants, that quantify some aspect of a network's structure without attaching any specific semantic meaning to the metric's values. Examples include maximal degree, girth, or vertex chromatic number (Bang-Jensen and Gutin, 2008). Social network analysis has defined additional metrics that are intended to measure something about the network that has semantic meaning in the context of the social application of the network. These metrics include centrality (Scott 2000), reciprocity (Newman 2010) (Scott & Carrington, 2011), and clustering coefficient (Easley and Kleinberg, 2010). Finally, overarching empirically-derived structural properties common to categories of networks, such scale-free and cellular, may apply to social networks (Tsvetovat and Carley, 2005). All are intended to measure in an objective and quantitative way some aspect of a network's structure that may be useful for a particular application. The intent is that realistic synthetic social networks would have metric values similar to those of the real-world social networks they were intended to mimic, without having identical structures.

Many network metrics have defined, and clearly not all could be used in this work. From those available, 20 were carefully selected to assess the similarity of real-world and synthetic social networks in this work. That selection was made in part based on the motivation of studying the social networks of future space colonies. Thus metrics that characterize information flow, integration of individuals into the network, level of camaraderie indicated by clustering, and level of influence among the individuals are of interest. Because this work used only undirected symmetric networks, only metrics suitable for those networks were considered.

The metrics selected include both standard metrics of graphs' structural characteristics (nodes, links, components, degree, radius, and eccentricity) and metrics considered to be relevant to social network structure, per (Rapoport 1957) (Freeman 1978) and (Bonacich 2007). In the former category, the number of nodes, links, and components, the network's radius and eccentricity, and the nodes' degrees fundamentally characterize a network's structure.

In the latter category, metrics found useful to study team structure and interaction were of special interest. Global clustering coefficient, average clustering coefficient, Gini coefficient, and number of communities provide some insight to the tight knit groups and the distribution of nodes among the communities. Average betweenness serves as a basis of comparison for maximum betweenness to identify the information brokers or potential bottle-necks in the network. Likewise, average closeness serves as a basis of comparison for minimum closeness to identify the nodes that are at the heart of communities. Mean path length, network radius, average eccentricity, and network diameter are geodesic distances that can be used estimating the rate of information flow across a network. Eigencentrality indicates the level of influence that a node may exert on other nodes. In some similar applications, clustering, path length, betweenness, closeness, and diameter were used in a study of information sharing and collaboration in small groups (Manso and Manso, 2010), betweenness was used in a study of interaction in programming teams (Gloor et al., 2011), density and diameter were used in a study of authorship collaboration (Gajewar and Das Sarma, 2012), eigencentrality was used in a study of leadership in social groups (Bullington 2016), and Gini coefficients have been used as a measure of inequality of participation in digital health social networks (van Mierlo et al., 2016). Table 2 lists and defines the metrics used.

Personality models

In 1923, Jung described distinct human personality types based upon his clinical observations (Jung 1971). Using Jung's ideas, in 1944 (Myers, 1962) developed a structured approach to identifying personality types and published a manual describing a personality typing process that later became known as the (Myers & McCauley, 1985) Type Indicator (MBTI) (Smathers 2003). In the MBTI typing scheme, each person is categorized on four "dichotomies" or dimensions, held to correspond to different aspects of personality. Two "preferences" or values are possible on each dichotomy, yielding a total of 16 different personality types. The four dimensions and their two preferences each are:

- Attitude (inward or outward focus); Extraversion (E) or introversion (I).
- Perceiving (information gathering) function; Sensing (S) or Intuition (N).
- Judging (deciding) function; Feeling (F) or Thinking (T).
- Lifestyle preference; Perceiving (P) or Judging (J).

Table 3(a) shows the estimated proportion of the United States population who would be categorized into each preference, with each dimension considered separately (Marioles et al. 1996; Mitchell, 1996). Table 3(b) shows the result of calculating a proportion for each personality type, based on the dimensions' proportions. A detailed description of the 16 Myers-Briggs types is beyond the scope of this article; for details see (Keirsey 1998). The important ideas here are that each person may be categorized as having one of the 16 types and that the likely compatibility of two people may be estimated from their personality types.

Critics of the MBTI personality model point to apparent problems. Metzner et al. suggested that the "rigid" dichotomies of the Jungian personality types constitute a "conceptual straight jacket" and proposed a reformulation of the dichotomies as pairs of primary and inferior psychological functions (Metzner, Burney, and Mahlberg, 1981). Additionally, McCrae and Costa commented that the MBTI lacks a neuroticism factor, perhaps because emotional instability was not part of Jung's type definitions, and it appears that Myers and Briggs believed that each personality type was positive. The lack

Table 2 Social network metrics used in this research

Metric	Definition
Nodes	Number of nodes in the network; here denoted *n*.
Links	Number of links in the network; here denoted *m*.
Components	Number of disjoint sets of connected nodes in a network.For a connected network, the value of this metric is 1.
Network density	Number of links in the network divided by the number of possible links $n \cdot (n - 1) / 2$; here denoted *p*.
Average degree	Average, or mean, of the nodes' degrees.
Standard deviation degree	Standard deviation of the nodes' degrees.
Global clustering coefficient	Ratio of closed nodes of vertices to connected triplets of nodes.
Average clustering coefficient	Average of the nodes' local clustering coefficients;the latter is the ratio of actual links to neighborsto possible links to neighbors for a given node.
Number of communities	Number of clusters in the network
Cluster Gini coefficient	Inequality of distribution of nodes among communities
Mean path length	Mean of the number of links in the shortest path betweeneach pair of nodes.
Average betweenness	Mean of the nodes' betweenness centrality values, which is the number of shortest paths between pairs of node that pass through a node.
Maximum betweenness	Maximum of the nodes' betweenness centrality values.
Average closeness	Mean of the nodes' closeness centrality values, which is the sum of the path lengths between the node and all other nodes.
Minimum closeness	Minimum of the nodes' closeness centrality values.
Average eigencentrality	Mean of the nodes' eigencentrality (also known as eigenvector centrality); the latter is a measure of the number of links each of a nodes neighbors have.
Minimum eigencentrality	Minimum of the nodes' eigencentrality.
Network radius	Minimum of the nodes' eccentricities; the latter is the maximum length of the shortest paths from a node to all other nodes.
Average eccentricity	Mean of the nodes' eccentricities.
Network diameter	Maximum of the nodes' eccentricities.

of a negative factor may make the interpretation of MBTI results easier to accept. However, it could also allow the omission of information that would be useful to employers, coworkers, counselors, and individuals (McCrae and Costa, 1989).

Nonetheless, the MBTI model is used and accepted at the U. S. National Aeronautics and Space Administration, the organization from which this work's motivating application is drawn, e.g., (Nelson and Bolton, 2008). It is also widely used in industry in the United States, including 89 of the Fortune 100 companies (Grant 2013), for applications that include increasing self-awareness to support decision analysis (Malik and Zamir, 2014) (Weiler 2017), improving team performance by explaining communication styles (Choo, Lou, Camburn, et al., 2014), identifying correlations between performance and personalities (Felder 2002) (Felder 2005) (Kiss, Kun, Kapitány, and Erdei, 2014) (Furnham and Crump, 2015a) (Furnham and Crump, 2015b), and identifying

Table 3 Personality type frequencies in the U. S. population (Marioles et al. 1996)

(a)		(b)			
E 0.463	I 0.537	ENTJ 0.045	ESTJ 0.097	INTJ 0.053	ISTJ 0.112
N 0.319	S 0.681	ENTP 0.033	ESTP 0.070	INTP 0.038	ISTP 0.081
T 0.529	F 0.471	ENFJ 0.040	ESFJ 0.086	INFJ 0.047	ISFJ 0.100
J 0.581	P 0.419	ENFP 0.029	ESFP 0.062	INFP 0.034	ISFP 0.072

correlations between professions and personalities (MH, 1977) (Freeman 2009) (Jafrani et al., 2017) (Rosati, 1993) (Capretz, 2002) (Cohen et al, 2013) (Loffredo et al, 2008) (Moutafi et al, 2007) (Emanuel, 2013).

Other personality models exist. Arguably among the best known is the Five Factor or OCEAN model. After analyzing correlations among 35 personality traits, Tupes and Christal identified five personality factors: Surgency (Extraversion), Agreeableness, Dependability (Conscientiousness), Emotional Stability (versus Neuroticism), and Culture (Openness) (Tupes and Christal, 1992) (John and Srivastava, 1999). Goldberg referred to these factors as "The Big Five" (Goldberg 1990). McCrae and Costa interpreted the factor Culture as Openness to experience (McCrae and Costa, 1987). Ruston and Irwing rearranged the first letters of the factors to form the mnemonic OCEAN (Rushton and Irwing, 2008).

Personality compatibility

The National Aeronautics and Space Administration (NASA) defines Team Risk as the risk associated with a decrease in performance and behavioral health due to inadequacy of a team's cooperation, coordination, communication, and psychosocial adaption (DeChurch et al., 2015). "Currently, NASA has no formalized process to compose mission teams from a scientific perspective, but this is an identified need for future exploration missions" (Landon 2015). Anania asserts that "crew compatibility on an interpersonal level will need to be a major factor in order to ensure optimal communication and coordination within the team" (Anania et al., 2017). Brandley and Herbert applied MBTI to their study of Information Systems teams and found that a team's personality type composition is partially related to performance (Bradley and Hebert, 1997).

Personality compatibility may play a significant role in link formation in real-world social networks. Back asserted that "personality differences influence social relationships", but noted that social network research rarely considers the effects of individual personalities (Back 2015). With that in mind, the algorithms described here both make use of inferred personality types for the people represented by the network's nodes and base the probability of a link forming between two nodes on the compatibility of the personality types associated with those nodes.

Table 4 is such a personality compatibility table for the MBTI personality types. The rows and columns are the 16 MBTI personality types. Each entry in the table is the probability of a link forming in a social network between two nodes if the nodes' associated personality types are those of the entry's row and column. Note that the table is symmetric, i.e., the two entries for two personality types are the same regardless of which type is on the row and the column. Table 4 was constructed from the personality type descriptions in (Keirsey 1998); the process for doing so is detailed in Appendix 1.

Homophily and heterophily can be modeled as likelihoods of link formation among personality types. In Table 4, values on the diagonal of the table represent a level of homophily because cells on the diagonal are the intersections of rows and columns identifying the same personality type. Values in the cells other than the diagonal represent some level of heterophily because those cells are at the intersections of rows and columns that identify different personality types.

MBTI was used in this work because of its wide application in practical settings. However, the social network generation algorithms presented later do not depend on

Table 4 Personality compatibility table for pairs of MBTI personality types

	ESTP	ISFP	ISTP	ESFP	ESTJ	ESFJ	ISTJ	ISFJ	ENFJ	INFJ	ENFP	INFP	ENTJ	INTJ	ENTP	INTP
ESTP	0.040	0.296	0.506	0.506	0.296	0.296	0.506	0.296	0.714	0.506	0.506	0.506	0.296	0.714	0.867	0.506
ISFP	0.296	0.110	0.506	0.139	0.296	0.296	0.139	0.296	0.296	0.867	0.867	0.506	0.714	0.714	0.506	0.506
ISTP	0.506	0.506	0.259	0.296	0.867	0.506	0.714	0.867	0.139	0.296	0.714	0.296	0.139	0.506	0.714	0.714
ESFP	0.506	0.139	0.296	0.460	0.506	0.867	0.714	0.506	0.506	0.296	0.296	0.714	0.506	0.139	0.296	0.296
ESTJ	0.296	0.296	0.867	0.506	0.680	0.714	0.867	0.952	0.296	0.139	0.506	0.139	0.296	0.296	0.506	0.506
ESFJ	0.296	0.296	0.506	0.867	0.714	0.840	0.867	0.714	0.296	0.506	0.506	0.506	0.714	0.051	0.139	0.139
ISTJ	0.506	0.139	0.714	0.714	0.867	0.867	0.940	0.867	0.506	0.296	0.296	0.296	0.506	0.139	0.296	0.296
ISFJ	0.296	0.296	0.867	0.506	0.952	0.714	0.867	0.940	0.296	0.139	0.506	0.139	0.296	0.296	0.506	0.506
ENFJ	0.714	0.296	0.139	0.506	0.296	0.296	0.506	0.296	0.840	0.506	0.139	0.506	0.714	0.714	0.506	0.506
INFJ	0.506	0.867	0.296	0.296	0.139	0.506	0.296	0.139	0.506	0.680	0.714	0.714	0.867	0.506	0.296	0.296
ENFP	0.506	0.867	0.714	0.296	0.506	0.506	0.296	0.506	0.139	0.714	0.460	0.296	0.506	0.506	0.714	0.296
INFP	0.506	0.506	0.296	0.714	0.139	0.506	0.296	0.139	0.506	0.714	0.296	0.250	0.506	0.506	0.296	0.714
ENTJ	0.296	0.714	0.139	0.506	0.296	0.714	0.506	0.296	0.714	0.867	0.506	0.506	0.110	0.296	0.139	0.139
INTJ	0.714	0.714	0.506	0.139	0.296	0.051	0.139	0.296	0.714	0.506	0.506	0.506	0.296	0.030	0.867	0.867
ENTP	0.867	0.506	0.714	0.296	0.506	0.139	0.296	0.506	0.506	0.296	0.714	0.296	0.139	0.867	0.110	0.714
INTP	0.506	0.506	0.714	0.296	0.506	0.139	0.296	0.506	0.506	0.296	0.296	0.714	0.139	0.867	0.714	0.250

any particular personality compatibility table or even on a particular personality model. Any personality model that satisfies the following two criteria could be used: (1) it has personality types that are discrete, or could be discretized; and (2) it provides, or enable the development of, a quantitative measure of the relative compatibility of different personality types that can be encoded as a personality compatibility table. In fact, a different personality table was used in the early stages of this work, with similar results to those reported here.

Related work

This section briefly reviews selected prior work related to generating graphs and social networks.

Real-world social networks

Social network analysis research requires real-world social networks to use as input data. First developed in the early 1980s, UCINet is a social network analysis application that calculates a variety of network metrics (Freeman 1988). UCINet includes functions for discovering cohesive subgroups in a network (Borgatti et al., 2014). An associated archive of social networks, represented as adjacency matrices, is maintained in the UCINet format (Freeman, 2009) (Freeman, 2016).

Table 5 lists the real-world social networks used in this research as source data; they are from the UCINet archive. In all but one of the networks, the nodes of the network correspond to individual people and the links to a relationship of some kind between them. (The exception is the Schwimmer Taro Exchange Network., where the nodes correspond to Orokaiva households within the Papaun village Sivepe and the links represent the mutual exchange of gifts, such as cooked taro (Schwimmer, 1979) (Schwimmer 1973).)

The real-world social networks used in this research include both symmetric and asymmetric and both unweighted and weighted networks. The new network synthesis algorithms to be described produce symmetric unweighted networks. Therefore the

Table 5 Real-world social network data sets used in this research

Real-world social network	Source	Nodes	Symmetric	Weighted
Robins Australian Bank	(Pattison et al., 2000)	11	no	no
Roethlisberger & Dickson Bank Wiring Room	(Roethlisberger and Dickson, 1939)	14	yes	no
Thurman Office	(Thurman 1979)	15	yes	no
Sampson Monastery	(Sampson 1969)	18	no	yes
Krackhardt Office CSS	(Krackhardt 1987)	21	no	no
Krackhardt High-Tech Managers	(Krackhardt 1987)	21	yes	no
Schwimmer Taro Exchange	(Schwimmer 1973)	22	yes	no
Webster Accounting Firm	(Webster 1993)	24	yes	yes
Zachary Karate Club	(Zachary 1977)	34	no	no
Bernard & Killworth Technical	(Bernard et al., 1982)	34	yes	yes
Bernard & Killworth Office	(Bernard et al., 1982)	40	yes	yes
Krebs Fortune 500 IT Department (Advice)	(Chen 2007)	56	no	yes
Krebs Fortune 500 IT Department (Business)	(Chen 2007)	56	no	yes
Lazega Law Firm	(Lazega 2001)	71	no	no

real-world networks were converted to symmetric and unweighted if necessary before being used as exemplar networks. The conversions were done in the obvious ways; if an asymmetric network had directed link(s) in either or both directions between two nodes, the converted network had an undirected link between those nodes, and if a weighted network had a weighted link of any weight between two nodes, the converted network had an unweighted link between the nodes.

Current trends in social network analysis include social networks developed from massive data sets captured from online social media and communities, such as FaceBook, Twitter, and Wikipedia; (Mislove et al., 2007), (Crandall et al., 2008), (Kwak et al., 2010), (Catanese et al., 2011), (Yang and Leskovec, 2015), and (Grandjean 2016) are examples. Common interests in careers, pastimes, politics, popular culture, and societal trends serve as the motivation for joining groups within these online communities, so personality types may be one of many factors determining how links form in real-world social networks. However, according to Krebs social networks expressed as connections via Facebook and LinkedIn can be misleading because site members may try to connect with as many people as possible and others acquiesce to the creation of apparent links with no real connection. "Two people might show to be connected but they really are not – one person was too embarrassed to turn down a 'friend request' from a total stranger. These 'false positives' tend to pollute the data of these social networking services" (Krebs 2008).

Existing models for generating synthetic social networks

Generating synthetic social networks that are more realistic than random graphs, such as those generated by the classic Erdős-Rényi $G(n, p)$ algorithm, also known as the random graph model (Erdős 1959) (Erdos and Rényi, 1960), requires attention to the properties of social networks that distinguish them from random graphs. Since 1960, several social network generation models have been developed. A selection of existing social network generation models that consider or exploit various structural characteristics of networks includes the following; each will be described following the list:

- Random graph model (Erdos and Rényi, 1960)
- Configuration model (Bollobás 1980) (Milo et al., 2003) (Newman 2003) (Viger and Latapy, 2005)
- Exponential random graph model (Holland and Leinhardt, 1981) (Frank and Strauss, 1986) (Wasserman and Pattison, 1996)
- Stochastic block model (Holland et al., 1983) (Nowicki and Snijders, 2001)
- Small world model (Watts and Strogatz, 1998)
- Preferential attachment model (Barabási and Albert, 1999)
- Popularity Similarity model (Papadopoulos et al., 2012)
- Chung-Lu graph model (Chung and Lu, 2002)
- Degree correlation dK series (Mahadevan et al., 2006)
- Block two-level Erdős Rényi model (Seshadhri et al., 2012)
- Replication of complex networks model (Staudt et al., 2017)

In random graphs, the nodes' degrees tend to follow a Poisson distribution (Bollobás 1998). This can be unrealistic; real-world networks' node degree distributions are more often non-Poisson and heavy-tailed. The configuration model extends the random graph model to address that inconsistency (Bender and Canfield, 1978) (Bollobás 1980) (Molloy and Reed, 1995) (Molloy and Reed, 1998) (Newman et al., 2001) (Milo et al., 2003) (Newman 2003) (Viger and Latapy, 2005). In the configuration model, network generation is initialized with both the number of nodes n and a specific degree sequence $K = \{k_1, k_2, \ldots, k_n\}$, where k_i is the degree of node v_i. The degree sequence K may be random variates drawn from a suitable distribution (checked to ensure that $\Sigma\ k_i$ is even), or more simply, the actual degree sequence of a real-world network serving as an exemplar of the class of networks to be generated. Given n nodes a degree sequence K, links are added by randomly connecting each node v_i to k_i other nodes, with each link uniformly possible. This produces networks with a realistic degree distribution, but if a single exemplar is used for multiple synthetic networks, all the generated networks will have the same node degrees.

The exponential random graph models (ERGM), also known as the p* model, assembles a network from subgraph structures, such as stars, triangles, paths, and cycle patterns (Wasserman and Pattison, 1996) (Snijders 2002) (Robins et al., 2007). Holland and Leinhardt developed an exponential family of probability distributions for directed graphs, which derived from empirical observations of stars (nodes with multiple links), isolates (nodes without links), and their triad census (the sixteen possible configurations of a directed triad) (Holland and Leinhardt, 1977) (Holland and Leinhardt, 1981). Frank and Strauss developed a family of distributions for directed and undirected Markov graphs wherein there existed dependence among the links (Frank and Strauss, 1986). Snijders applied Monte Carlo Markov Chains to estimate network metrics such dyads, undirected and directed two paths, and directed and undirected triangles (Snijders 2002). Hunter distinguished between ERGM and p* by associating the maximum pseudo-likelihood estimation (Wasserman and Pattison, 1996) with p* and maximum likelihood estimation (Geyer and Thompson, 1992) with ERGM (Hunter 2007).

Among the existing methods, the stochastic block model (SBM) may have the most similarity to the new methods developed in this work, and so we describe it in a bit more detail. The SBM can be used to generate networks and to detect communities within large scale networks (Holland et al., 1983) (Anderson et al., 1992) (Faust and Wasserman, 1992) (Newman and Girvan, 2004) (Bickel and Chen, 2009) (Fortunato 2010) (Decelle et al., 2011) (Abbe 2017). The set of actors or agents involved is first partitioned into B communities or clusters known as *blocks*. This partitioning is often done by manual analysis, based on observation or data. Tightly interacting groups of actors are placed into the same group. A $B \times B$ preference matrix W specifies the probabilities of link formation both within and between the blocks (Nowicki and Snijders, 2001). The probabilities may be provided manually or by automated analysis of the source data. The on-diagonal entries in W specify the probabilities of links forming between nodes in the same block, whereas the off-diagonal entries in W specify the probabilities of links forming between nodes in different blocks. If the on-diagonal probabilities are higher than the off-diagonal probabilities, then the intra-block link density will be higher than the inter-block link density; such a network is known as *assortative*. Conversely, if the off-diagonal probabilities are higher than the on-diagonal probabilities, then the resulting network will have a higher inter-block link density; such as network is known as *disassortative*. In an SBM implementation, the number of nodes in each block may be stored in an integer vector with B entries. If the blocks are assumed to be disjoint, the sum of the vector's entries is the total number of nodes in the network. To generate a synthetic network, the probability of link formation in W between each pair of nodes is used to stochastically determine if a link is formed between those nodes.

The small world model starts with a one dimensional regular ring lattice where each node has links to its k nearest neighbors (Watts and Strogatz, 1998) (Strogatz 2001). Several iterations of random rewiring produce a network with a desired density. For each node, rewiring involves stochastically determining whether an existing link is deleted or a new link is formed between the current node and another randomly selected node.

The preferential attachment model starts with a small set of nodes and then adds nodes and links in an iterative process based upon the connectivity of the nodes (Barabási and Albert, 1999) (Barabási, 2003). The number of nodes in the initial set determines the maximum degree for new nodes. In each iteration, or "time step", a new node is added to the network and then links from the new node to the existing nodes are stochastically added, up to the maximum degree. The process depends upon the existing nodes' current connectivity, which is calculated as $k = m \cdot (t \ / \ t_i)^{1/2}$ where m is the node's current degree, t is the current iteration (or time step), and t_i is the initial time step when the node was added. The probability of link being added from the new node to existing node i is k_i / $(\Sigma\ k)$ where k_i is the connectivity of node i and $(\Sigma\ k)$ is the sum of the connectivity of the other existing nodes. New nodes, and links from them to existing nodes, are iteratively added until the network has the desired number of nodes. This process produces a scale-free network.

The Popularity Similarity model bases the probability of link formation on hyperbolic distances between nodes (Papadopoulos et al., 2012). In this model, the network grows as nodes are added at successive time steps. Older (earlier added) nodes tend to be popular because they have had more time to connect to other nodes. To model similarity, new nodes are randomly placed on a circle; a node's birth time determines the radial coordinate $r_t = \ln(t)$. Two nodes, with polar coordinates $(r_s,\ \theta_s)$ and $(r_t,\ \theta_t)$, have an approximate hyperbolic distance $x_{st} = r_s + r_t + \ln(\theta_{st}/2) = \ln(st\theta_{st}/2)$ where s and t are the nodes' respective birth times. This hyperbolic distance serves as a convenient metric that represents both radial popularity and angular similarity.

The Chung-Lu model uses an exemplar degree sequence to set the probability of link formation between two nodes. For a pair of nodes, the link formation probability is proportional to the product of corresponding degrees in the sequence (Chung and Lu, 2002).

The degree correlation *dK* series model uses probability distributions for node degree correlations for subnetworks of size d to generate networks. A generated *0 K*-graph reproduces the average node degree of an exemplar network. A *1 K*-graph reproduces the degree distribution of an exemplar network. A *2 K*-graph reproduces the joint degree distribution and a $3K$-graph reproduces similar interconnectivity among triangles as an exemplar network (Mahadevan et al., 2006).

The Block two-level Erdős-Rényi model introduces community structures by generating a set of independent networks and then randomly linking nodes among the communities (Seshadhri et al., 2012). Typically, algorithms that implement this model include input parameters for nodes and density and the algorithm returns a network with the number of links based upon the density.

(Staudt et al., 2017) describes the replication of complex networks (ReCon) model that generates scalable synthetic social networks based on an exemplar network. An objective of ReCon is to generate networks of different sizes, up to 32 times larger than the exemplar. The ReCon algorithm first detects communities in the exemplar network using the parallel Louvain method. It then generates a working graph as a disjoint union of x copies of the exemplar, where x is a scaling factor. For each detected community in the working graph, the algorithm preserves the degree distribution and rewires the intra-community links through random edge switching. After rewiring the intra-community links, it rewires the inter-community links and generates links among the copies of the network (Staudt et al., 2017). In this work a realistic replica of an exemplar social network was defined as a network that has similar metric values as the exemplar. The metrics that were compared to the exemplar included sparsity, i.e. number of links versus number of nodes, the degree distribution's Gini coefficient, maximum degree, average clustering coefficient, diameter, number of connected components, and number of communities. ReCon produces replicas that are realistic under this definition because it preserves the exemplar's community structure and node degrees.

Comparison to the current work

In contrast to the algorithms reported later, with only one exception the existing social network generation methods do not use any actual or inferred attributes of the persons represented by the nodes to determine or influence the generation of links between the nodes. The exception is the stochastic block model, which uses a group attribute associated with each node to determine the probability of link formation with other nodes within the same group. None of the prior methods use personality type or compatibility, as is done in this work, to produce synthetic social networks. This idea was hinted at in (Staudt et al., 2017), which described a potential application of synthetic social networks as showing interactions that are "determined by implicit psychological and social rules", but those "rules" were not used to generate networks.

The desirable features of a synthetic social network generation algorithm include parsimony (i.e., few parameters), speed of execution, and network realism. Realism, in particular, is a very important characteristic of synthesized social networks. Realism in social networks has been defined in terms of network structural features, dynamics, and evolution (Staudt 2017). The similarity, or lack thereof, of metric values between a synthetic network and a real network is understood as a measure of realism. (Chakrabarti et al., 2004) (Leskovec et al., 2010). A quantitative assessment of realism is central to the current work.

Synthesizing social networks based on personality compatibility

This section explains the new personality-based synthetic social network generation algorithms developed in the current work. The section begins with placing the new algorithms in the context of the overall process used for network synthesis; the details of the individual algorithms in the process will follow the overview.

Synthesis process overview

Figure 2 shows the algorithms and dataflow in the network synthesis process. That process starts with a real-world social network T, which serves as an exemplar of the class of social networks to be generated. (In this work, T is any of the fourteen real-world networks listed in Table 5). Network T is input to three different algorithms. The two algorithms developed in this work, Probability Search (PS) and Configuration-Degree Matching (CDM), each construct an assignment A of personality types to the nodes of T. Both employ heuristic methods to find A, albeit in completely different ways. The resulting personality type assignment A is then input to a network generator algorithm (GNAC), which generates a set of synthetic social networks (denoted P for the PS algorithm or M for the CDM algorithm), using the personalities in A and the compatibility information in personality compatibility table C.

The challenge is to find a personality type assignment A which, when the GNAC algorithm is used with personality compatibility table C, will produce realistic synthetic social networks. A personality type assignment that produces realistic synthetic social networks will be referred to in this context as *effective*.

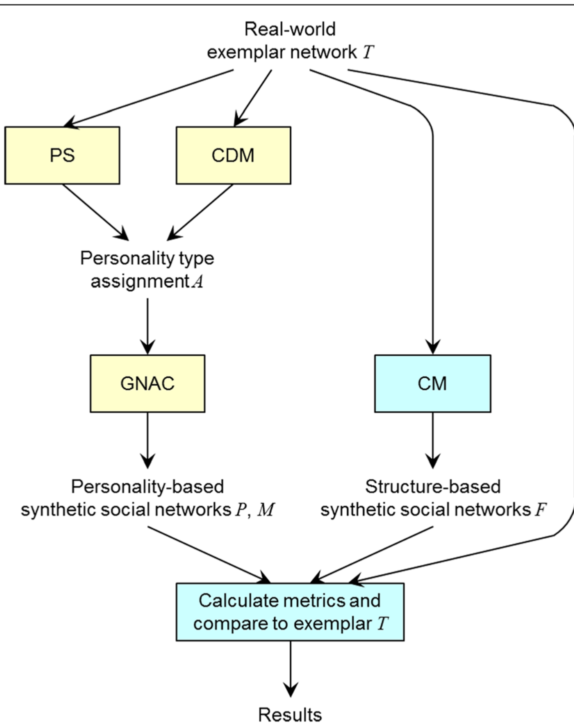

Fig. 2 Algorithms and dataflow in the network synthesis process

Real-world exemplar network T is also be input to a standard network generation algorithm, the Configuration Model (CM). CM also generates a set F of synthetic social networks based only on the structure of T and without using and personality compatibility information.

The three sets of synthetic networks are then input to a process that calculates the network metrics listed in Table 2 and compares them to the exemplar T.

Generating networks from a personality type assignment

Synthetic social networks are generated by an algorithm that considers personality compatibility by using a personality compatibility table C (e.g., Table 4) and a personality assignment A to the nodes of the network. The network generation algorithm is denoted the $G(n, A, C)$ (GNAC) algorithm, where n is the number of nodes, A is an assignment of personality types to the n nodes, and C is a personality compatibility table that includes the personality types in A. Given an assignment A of personality types to nodes and a compatibility table C, as many synthetic social networks as needed can be generated using the GNAC algorithm. They will likely differ due to the randomness in the algorithm, but they will be related in that all were produced using the same assignment A and compatibility table C.

The GNAC algorithm first determines the degree sequence of an exemplar network T. The degree sequence is used to initialize a link budget for each of the nodes in the synthetic network. The algorithm then randomly selects two triads of nodes in the synthetic network as candidates for triangles. The personality types assigned to the triads' nodes by A and the personality compatibility table C are used to find the probability of link formation between each pair of nodes in the triads, and the probabilities for each triad are summed. The triad with the larger sum is then converted into a triangle by connecting all unlinked pairs in the triad and the link budgets of any newly linked nodes are decremented. This procedure repeats until the number of triangles in the synthetic network is the same as the number of triangles in the exemplar network.

Producing the desired number of triangles typically does not completely deplete the link budgets of all of the nodes. For the nodes with remaining link budgets, the algorithm randomly selects pairs of those nodes. If the pair is not linked, then a link is formed and the nodes' link budgets are decremented. When a pair of nodes with remaining link budgets that are not already connected cannot be found, then the algorithm randomly selects nodes that have no remaining link budget. If the randomly selected node and a node needing a neighbor are not connected, the algorithm randomly adds a link between the nodes with a probability determined by the nodes' assigned personality types and the compatibility table C. The process repeats until the sum of all nodes remaining link budgets is 0, at which point the synthetic social network is returned.

In the following pseudocode, T is an exemplar network, A is a personality assignment, C is a compatibility table, $S = (V, E)$ is a synthetic network and u and v are nodes in the network. At three points in the *gnac* function links may be added to the network. The *addlink* function, shown first, is called by the *gnac* function; it adds a link between nodes u and v if they are not already connected.

```
function addlink(S, u, v, db, C, comp)
    if {u, v} ∉ E
        i ← personality type in A for node u
        j ← personality type in A for node v
        if (comp = true and random number ≤ C[i, j]) or (comp = false)
            E ← E ∪ {u, v}            # Add link between u and v
            db[u] ← db[u] – 1         # Decrement link budget for u
            db[v] ← db[v] – 1         # Decrement link budget for u
        end if
    end if
    return S
end function
```

```
function gnac(T, A, C)
    ds ← degree sequence of T
    db ← sort ds in decreasing degree order
    b ← nodes with db link budget ≥ 2
    S ← network with n nodes and 0 links
    ntriangles ← 0
    while | b | ≥ 4 and ntriangles < number of triangles in T
        do
            u1, v1, w1 ← nodes randomly selected from b without replacement
            u2, v2, w2 ← nodes randomly selected from b without replacement
        until {u1, v1, w1} ≠ {u2, v2, w2}
        triad1.p ← C[A[u1], A[v1]] + C[A[v1], A[w1]] + C[A[w1], A[u1])
        triad2.p ← C[A[u2], A[v2]] + C[A[v2], A[w2]] + C[A[w2], A[u2])
        if triad1.p ≥ triad2.p
            u, v, w ← u1, v1, w1
        else
            u, v, w ← u2, v2, w2
        end if
        if {u, v} ∉ E or {v, w} ∉ E or {w, u} ∉ E          # At least one link missing
            S ← addlink(S, u, v, db, C, false)
            S ← addlink(S, v, w, db, C, false)
            S ← addlink(S, w, u, db, C, false)
            ntriangles ← ntriangles + 1
        end if
    end while
    b ← nodes with db link budget ≥ 1
    potentialdyads ← all pairs of nodes (u, v) where u, v ∈ b, u ≠ v, and {u, v} ∉ E
    while | potentiadyads | ≥ 2
        {u, v} ← pair of nodes randomly selected from potentialdyads
        S ← addlink(S, u, v, db, C, false)
        b ← nodes with db link budget ≥ 1
        potentialdyads ← all pairs of nodes (u, v) where u, v ∈ b, u ≠ v, and {u, v} ∉ E
    end while
    b ← nodes with db link budget ≥ 1
    pend ← V – b
    while | b | ≥ 1
        u ← b[1]
        v ← node randomly selected from pend
        S ← addlink(S, u, v, db, C, true)
        b ← nodes with db link budget ≥ 1
        pend ← V – b
    end while
    return S
end function
```

The overall computational complexity of the GNAC algorithm is $O(n^4)$. To see this, consider first the function *addlink*; it does not loop over the nodes or edges and so is $O(1)$. The GNAC algorithm itself begins with some housekeeping that includes an $O(n \log n)$ sort of the nodes' degree sequence (line 3). The first main loop (lines 7–25) is over the triangles of T. A network with n nodes may have as many as $C(n, 3)$ triangles; $C(n, 3) = n!/(3!(n - 3)!) \in O(n^3)$. Within that loop, the do while loop (lines 8–11) may execute an arbitrary number of times, but on average is $O(1)$. The set membership tests (line 19) are $O(1)$ if the edge set is stored in a suitable data structure, such as an adjacency matrix. All of the remaining computation in the first main loop is also $O(1)$. Thus the second main loop is $O(n^3)$. Finding all the potential dyads the first time (line 27) is $O(n^2)$. There are potentially as many as $C(n, 2)$ such dyads; $C(n, 2) = n!/(2!(n - 2)!) \in O(n^2)$. The second main loop (lines 28–33) iterates once for each of the $O(n^2)$ dyads, and in each iteration it again finds all potential dyads $O(n^2)$, thus the second main loop is $O(n^4)$. The third and final main loop iterates at most once for each node, i.e., $O(n)$ iterations. Each iteration scans $O(n)$ nodes to find those with remaining link budgets, so the third main loop is $O(n^2)$. Thus the complexity of the GNAC algorithm as a whole is $O(n^4)$.

Probability search algorithm

The Probability Search (PS) algorithm is based on the idea that the probability of a given social network being generated algorithm from a given personality type assignment A and personality compatibility table C can be calculated. That calculation can be done in either of two ways that differ in whether or not nodes are assumed to be distinguishable. For this work, it is assumed that the nodes are uniquely identified and are thus always distinguishable from each other. This assumption is appropriate for many social network applications, where nodes correspond to specific known persons. The implication of uniquely identified nodes is that a different network, with the same connection structure (i.e., isomorphic in graph theory terminology) but connecting different specific nodes, would not be equivalent as a social network because different people would be connected.

The probability of the network will be calculated using a simple extension of the Erdős-Rényi $G(n, p)$ algorithm. In the $G(n, p)$ algorithm the probability of link formation p is constant for the entire network. In the PS algorithm's probability calculation the constant p is instead replaced for each pair of nodes with the probability of a link forming between those nodes, given a personality type assignment A and a personality compatibility table C. Let $p(i, j)$ be the probability given in C of a link being present between two nodes i and j for the personality types assigned to nodes i and j by A. The probability of a network $G = (V, E)$ being formed is therefore given by Eq. (1); we will call this the *network probability*.

$$P(G) = \prod_{i,j \in V, i \neq j} \begin{cases} p(i,j) & \text{if } \{i,j\} \in E \\ 1 - p(i,j) & \text{if } \{i,j\} \text{ not} \in E \end{cases} \tag{1}$$

Given an exemplar network T and a compatibility table C, the network probability can be used to search for the personality type assignment A that has the highest probability $P(T)$ of producing the exemplar. Once found, that personality type assignment can be used by the GNAC algorithm to generate synthetic networks that are likely to be similar to the exemplar.

In theory, the optimum personality type assignment, i.e., the assignment that has the highest possible probability of producing the given exemplar network T, could be found by methodically generating every possible personality type assignment and calculating $P(T)$ for each one. Unfortunately, this is not practical for any but the smallest networks. If a personality type scheme has k different personality types and exemplar network T has n nodes, there are k^n different possible type assignments. For the MBTI personality type scheme using in this work $k = 16$, thus for even the smallest real-world exemplar network used in this research, the Robins Australian Bank network with 11 nodes, there are $16^{11} \approx 1.76 \cdot 10^{13}$ possible personality type assignments. Calculating $P(T)$ for that many assignments at the rate of one per millisecond would require over 500 years. Thus an exhaustive search is impractical.

Instead, the new Probability Search (PS) algorithm performs a heuristic search through the space of possible personality type assignments. After generating an initial personality type assignment randomly, it iteratively changes the assignment, one node at a time. To do so, it uses *node probability*, a quantity similar to network probability, but calculated for a single node. Given a network G, a personality compatibility table C, and a personality type assignment A, the node probability of a single node i in G is given by Eq. (2).

$$P(i) = \prod_{j \in V, i \neq j} \begin{cases} p(i,j) & \text{if } \{i,j\} \in E \\ 1-p(i,j) & \text{if } \{i,j\} \text{ not} \in E \end{cases} \tag{2}$$

At each iteration, the PS algorithm selects a node i, either the node with the smallest node probability $P(i)$ under the current personality type assignment (with probability 0.95), or a random node (with probability 0.05). It then calculates $P(i)$ for that node i for each of the possible personality types, holding the network structure and other nodes' personality types fixed. The personality type that gives the highest node probability $P(i)$ is assigned to node i. This process repeats until the overall network probability improvement achieved in an iteration is less than a threshold, subject to a required minimum number of iterations. Finally, to prevent non-productive repetitive changes to the same node's personality type, when a node's personality type is changed it is added to a list of nodes excluded from adjustment in the next iteration and remains in that list for a certain number of iterations. The improvement threshold, the minimum number of iterations, and the number of iterations a node remains on the excluded list are all parameters to the algorithm. (For the results reported here, the values 0.0001, $n \cdot k \cdot 1000$, and $\lceil n / 10 \rceil$ respectively were used for those parameters. Those values were found empirically.)

In the following pseudocode for the PS algorithm, V is a set of nodes, E is a set of links, C is a personality compatibility table, A is a personality type assignment, n is the number of nodes, and k is the number of different personality types. In the pseudocode, two subroutines (functions) precede the main logic of the PS algorithm.

```
function vprob(V, E, C, A, i)
    result ←1
    for j ∈ V − {i}
        if {i, j} ∈ E
            result ← result · C[A[i], A[j]]
        else
            result ← result · (1 − C[A[i], A[j]])
        end if
    end for
    return result
end function
```

```
function gprob(V, E, C, A)
    result ←1
    for i ←1 to n − 1
        for j ← i + 1 to n
            if {i, j} ∈ E
                result ← result · C[A[i], A[j]]
            else
                result ← result · (1 − C[A[i], A[j]])
            end if
        end for
    end for
    return result
end function
```

```
function ps(T, A, C)
    excluded ← ∅
    prev ← 0
    improvement ← 1
    iterations ← 0
    A ← random assignment of personality types to n nodes
    bestA ← A
    cur ← gprob(V, E, C, A)
    bestprob ← cur
    while (improvement > 0.0001) and (iterations < n · k · 1000)
        available ← V − excluded
        if random number < 0.05
            i ← randomly selected node ∈ available
        else
            lowv ← available[1]
            lowprob ← vprob(V, E, C, lowv)
            for i ∈ available
                iprob ← vprob(V, E, C, i)
                if iprob < lowprob
                    lowv ← i
                    lowprob ← iprob
                end if
            end for
        end if
        adjustv ← lowv
        hight ← A[adjustv]
        highprob ← gprob(V, E, C, A)
        for j ← 1 to k
            A[adjust] ← j
            jprob ← gprob(V, E, C, A)
            if jprob > highprob
                hight ← j
                highprob ← jprob
            end if
        end for
        A[adjust] ← hight
        excluded ← excluded ∪ { adjustv }
        if | excluded | >⌈n / 10⌉
            excluded ← excluded − node that has been excluded the longest
        end if
        prev ← cur
        cur ← gprob(V, E, C, A)
        improvement ← cur − prev
        iterations ← iterations − 1
        if cur > bestprob
            bestA ← A
            bestprob ← cur
        end if
    endwhile
    return bestA
end function
```

The overall computational complexity of the PS algorithm is $O(n^3)$. To see this, consider first the functions *vprob* and *gprob*; *vprob* loops once over the n elements of V (lines 3–9), and so is $O(n)$, whereas *gprob* has two nested loops (lines 3–11), each over the n elements of V, and so is $O(n^2)$. The main body of PS begins with some $O(1)$ housekeeping (lines 2–9) and an $O(n^2)$ call to *gprob*. The main loop (lines 10–49) executes $O(n)$ times. Within the main loop, the search for the lowest probability vertex (lines 15–23) begins with an $O(n)$ call to *vprob* (line 16), then loops over the available nodes $O(n)$ times; within that loop is an $O(n)$ call to *vprob*, thus this portion of the main loop is $O(n^2)$. Next the search for the highest probability personality type (lines 25–35) calls *gprob* once, and then enters a while loop that iterates k times, each time calling *gprob*, which is $O(n^2)$. Because k is a constant and not a function of n, this portion of the loop is $O(n^2)$. The last part of the while loop includes two operations on the *excluded* list (lines 37 and 39) which can be accomplished in amortized $O(1)$ time if implemented as a deque, and another $O(n^2)$ call to *gprob*. Thus the complexity of the main loop, and PS algorithm as a whole, is $O(n^3)$.

Compatibility-degree matching algorithm

The Compatibility-Degree Matching (CDM) algorithm first determine the degree sequence of a given exemplar network T. It then generates a personality type assignment A in accordance with an empirical distribution based the frequency of each personality type in the U. S. population (Table 3). The columns of personality compatibility table C provides an overall compatibility of each personality type. The CDM then orders the personality types by overall compatibility and the nodes of the exemplar network T by decreasing order of degree. Using those two orderings, the CDM personality types to the nodes so that the personality types with the highest overall compatibility are assigned to the nodes with the highest degree. In the pseudocode, personality type assignment A is a vector of size n.

```
1.  function cdm(T, C)
2.      ds ← degree sequence of T
3.      w ← sort V in decreasing degree order using ds
4.      for i ← 1 to k
5.          x[i] ← sum of compatibility values in C for personality type i
6.      end for
7.      x ← sort x in decreasing order
8.      for i ← 1 to n
9.          y[i] ← random personality type generated using empirical distribution
10.     end for
11.     y ← sort y in decreasing overall compatibility order using x
12.     for i ← 1 to n
13.         A[w[i]] ← y[i]
14.     end for
15.     return(A)
16. end function
```

The overall computational complexity of the CDM algorithm is $O(n \log n)$. The n nodes are sorted (line 3), which is $O(n \log n)$. The summing of the compatibility values (lines 4–6) is $O(k^2)$, where k is the number of personality types, and the sort of the sums (line 7) is $O(k \log k)$, but for most networks $k << n$. The assignment of personality types (lines 8–10) is $O(n)$ and the sort of the assigned types (line 11) is $O(n \log n)$. The final loop (lines 12–14) is $O(n)$. Thus the complexity of the CDM algorithm as a whole is $O(n \log n)$.

Configuration model algorithm

In order to assess the effectiveness of the personality-based algorithms (PS and CDM), they were compared to an existing network generative model that was not personality-based. Two were considered for the role of baseline. Because of its abstract representation of popularity, the Popularity Similarity model (Papadopoulos et al., 2012), as implemented in the R package NetHypGeom (Alanis-Lobato et al., 2016), was examined. However, perhaps because of that model's orientation to large scale-free networks, the implementation sets certain bounds on its input parameters; in particular, the average degree must be ≥2 and the scaling exponent must be ≥2 and ≤ 3. Of the fourteen real-world networks to be used as exemplars in this work (see Table 5), only one (Zachary Karate Club) had values for these metrics that satisfied both of these bounds; the other thirteen had an average degree < 2, a scaling exponent either < 2 or > 3, or both. Thus the exemplars to be used did not seem well suited to the capabilities of the Popularity Similarity model, or its implementation.

On the other hand, the Configuration Model (CM), which was described earlier, produces synthetic networks based upon the degree sequence of an exemplar network, and does not consider personality. Because it is based on degree sequence, is usable with the exemplars. Furthermore, it is considered by some to be a standard basis of comparison: "Following the works of Barabási et al., the degree distribution has become accepted as the most fundamental network characteristic… [I]t has become a standard to compare network quantities to a null-model where the degrees of the network (the *degree sequence*) is fixed and everything else random" (Barrenas et al., 2009).

Implementation and execution

This section describes the software implementation of the algorithms and supporting functions. It also discusses their execution.

Implementation of the algorithms

The two new algorithms for finding effective personality type assignments (PS and CDM), as well as the network generator GNAC algorithm, were implemented in the R language. R is an open-source programming language and environment with powerful and extensive features for data analysis, data visualization, and statistical computing (R Core Team 2016). R also includes a full range of general purpose programming language features, including control structures, mathematical operations, and file input/output. It should be noted that for medium and large networks, the network probability value $P(G)$ computed by the PS algorithm can become quite small, as it is the product of $n(n-1)/2$ probabilities, all of which are ≤1. A computer implementation of $P(G)$ meant to handle medium and large networks must take care to avoid numeric underflow. In our implementation, we used the R gmp (GNU Multiple Precision) package for arbitrary precision arithmetic.

As already mentioned, CM is an existing algorithm for generating synthetic social networks. A prior implementation of CM in the R language is available in the R igraph package, which is a collection of R functions for network analysis and visualization (Csárdi and Nepusz, 2013). In that package function sample_degseq produces networks using CM. That function was used for this work without modification.

Execution of the algorithms

Because R is an interpreted language, R programs often execute more slowly than comparable programs written in a compiled language. In addition, the two algorithms to find effective personality type assignments (PS and CDM) both involve numerous iterations, especially the PS algorithm. Consequently, the algorithms' run times during testing and analysis were sometimes quite lengthy. To keep the executions manageable, the programs were run on supercomputers provided and supported by the Alabama Supercomputer Authority. Typical run times for the two algorithms were highly dependent on the number of nodes in the exemplar graph; for the PS algorithm the run times ranged from a few minutes for the smallest real-world network (Robins Australian Bank, 11 nodes) to several hours for the largest real-world network (Lazega Law Firm, 71 nodes). Although the algorithms' implementation code was not parallelized, scripts were used to initialize and initiate multiple instances of the programs to execute concurrently.

Results

This section reports the results of testing and comparing the PS and CDM algorithms with the Configuration Model. The comparison is in terms of quantitative measures of the generated social networks' realism.

Realism is measured by the absolute difference between the mean metrics of the synthetic networks and the network metrics of the exemplar real-world social network. The metrics used to measure realism are listed in Table 2. Smaller absolute difference is preferred. Absolute differences between the metrics of the exemplar real-world social network and the mean metrics of the synthetic networks were calculated for networks generated by the PS and CDM algorithms and compared to networks generated by the CM algorithm.

As an example of the results, Table 6 presents a comparison of the realism metrics for the assignments found by the PS and CDM algorithms for only one of the real-world exemplar networks, Bernard & Killworth Technical. (For brevity, this section presents the results for only one of the exemplars in Table 6; the complete set of results are presented in Tables 9, 10, 11, 12, 13, 14, 15,16, 17, 18, 19, 20, 21, 22 in Appendix 2.) In the table, column 1 shows the name of the metric and column 2 shows that metric's value for the exemplar social network. Columns 3–6 apply to the synthetic social networks generated by the CM algorithm, collectively denoted *F*; column 3 shows the mean metric value for the networks generated by the CM algorithm, column 4 show the absolute difference between that mean value and the exemplar metric value, column 5 shows the L^1 norm for that metric, and column 6 shows the L^2 norm for that metric. Columns 7–10 show the same for the synthetic networks generated by the PS algorithm, collectively denoted *P*, and columns 11–14 show the same for the synthetic social networks produced the CDM algorithm, collectively denoted *M*. In columns 4–6, 8–10, and 12–14, the cells' content is set in bold type to show at a glance the PS- and CDM-generated networks' realism compared to the CM-generated networks' realism. Bold indicates that the PS or CDM networks' mean metric value was closer to the exemplar than the CM networks' mean metric value.

Table 6 Realism results for the Bernard & Killwork Technical network

Metrics	T	$\overline{F}$	$\lvert T-\overline{F}\rvert$	L1(F)	L2(F)	$\overline{P}$	$\lvert T-\overline{P}\rvert$	L1(P)	L2(P)	$\overline{M}$	$\lvert T-\overline{M}\rvert$	L1(M)	L2(M)
Nodes	34.00	34.00	0.00	0.00	0.00	34.00	0.00	0.00	0.00	34.00	0.00	0.00	0.00
Links	175.00	143.63	31.37	941.00	173.03	175.00	**0.00**	**0.00**	**0.00**	175.00	**0.00**	**0.00**	**0.00**
Components	1.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
Network density	0.31	0.26	0.06	1.68	0.31	0.31	**0.00**	**0.00**	**0.00**	0.31	**0.00**	**0.00**	**0.00**
Average degree	10.29	8.45	1.85	55.35	10.18	10.29	**0.00**	**0.00**	**0.00**	10.29	**0.00**	**0.00**	**0.00**
Standard deviation degree	4.63	3.55	1.08	32.41	5.97	5.03	**0.40**	**12.12**	**2.45**	5.00	**0.37**	**11.02**	**2.31**
Global cluster coefficient	0.48	0.30	0.17	5.17	0.95	0.44	**0.03**	**0.98**	**0.20**	0.45	**0.03**	**0.77**	**0.16**
Average cluster coefficient	0.47	0.32	0.16	4.76	0.88	0.53	**0.05**	**1.61**	**0.32**	0.53	**0.06**	**1.69**	**0.33**
Mean path length	1.81	1.90	0.09	2.70	0.52	1.77	**0.04**	**1.08**	**0.21**	1.78	**0.03**	**0.95**	**0.19**
Communities	4.00	6.37	**2.37**	**75.00**	**16.82**	7.37	3.37	103.00	21.10	6.50	2.50	77.00	**16.70**
Gini coefficient	0.49	0.49	**0.01**	**1.42**	**0.31**	0.50	0.02	**1.18**	**0.27**	0.51	0.02	1.49	0.32
Average betweenness	13.32	14.81	1.48	44.53	8.51	12.75	**0.58**	**17.74**	**3.49**	12.81	**0.52**	**15.74**	**3.17**
Maximum betweenness	63.29	53.03	**10.26**	**368.03**	**75.25**	104.94	41.65	1249.56	238.61	102.52	39.23	1176.86	229.79
Average closeness	0.02	0.02	0.00	0.03	0.01	0.02	**0.00**	**0.01**	**0.00**	0.02	**0.00**	**0.01**	**0.00**
Minimum closeness	0.01	0.01	0.00	0.03	0.01	0.01	**0.00**	**0.02**	**0.00**	0.01	**0.00**	**0.02**	**0.00**
Average eigencentrality	0.53	0.59	0.06	1.82	0.38	0.50	**0.03**	**1.32**	**0.27**	0.49	**0.04**	**1.38**	**0.30**
Minimum eigencentrality	0.06	0.06	**0.00**	**0.44**	0.10	0.07	0.01	0.46	**0.10**	0.07	0.01	0.45	**0.10**
Network radius	2.00	2.13	0.13	4.00	2.00	2.00	**0.00**	**0.00**	**0.00**	2.00	**0.00**	**0.00**	**0.00**
Average eccentricity	2.88	3.08	0.19	5.79	1.33	2.79	**0.09**	**3.74**	**0.85**	2.80	**0.08**	**3.68**	**0.80**
Network diameter	4.00	3.97	**0.03**	**3.00**	**1.73**	3.47	0.53	16.00	4.00	3.47	0.53	16.00	4.00

Boldfaced numbers indicate which algorithm performed better for a particular metric

As can be seen in Table 6, for the Bernard & Killworth. Technical exemplar network both the PS and the CDM algorithms produced more realistic synthetic social networks than the CM algorithm over the majority of the network metrics.

Table 7 summarizes the overall realism results. Two realism comparisons were made: PS versus CM and CDM versus CM. Both are reported in the table. A total of 280 metric values (14 real-world social networks · 20 metrics) were calculated for each of the comparisons. The columns labeled with an algorithm's abbreviation (PS, CDM, CM) show the number of metrics where that algorithm's metric values were closer to the exemplar that the other algorithm in the comparison, and a column labeled "=" shows the number where the two algorithms' metric values were equally close. In the PS versus CM comparison, the values of 142 of the 280 metrics (~ 50.7%) for the PS networks were closer to the values of the exemplar network than those of the CM algorithm, and another 31 values (~ 11.1%) were equally close; the CM networks values were closer to the exemplar on only 107 (~ 28.2%) of the metrics. In the CDM versus CM comparison, the values of 140 of the 280 metrics (50.0%) for the DCM networks were closer to the values of the exemplar network than those of the CM algorithm, and another 35 values (~ 12.5%) were equally close; the CM networks values were closer to the exemplar on only 105 (~ 37.5%) of the metrics.

A simple hypothesis test of proportion confirms that both PS and CDM come closer to the exemplar than CM more often that can be expected from random chance. For PS versus CM we treat each of the 280 metrics as a binomial trial. A closer metric value in a PS-generated network is counted as a success, a closer metric value in a CM-generated network is counted as a failure, and equal metric values are omitted from the sample. In a right-tailed test the hypotheses are H_0: $p = 0.50$ and H_1: $p > 0.50$, so the statistical assumption is that PS is not better than CM. The level of significance is set to $\alpha = 0.05$. The sample data is $r = 142$ and $n = 142 + 107 = 249$. The results are test statistic $p = 0.570281$, $z = 2.218035$, and p-value $= 0.01326$, which is $< \alpha$, thus we reject the null hypothesis and conclude that PS outperforms CM. The same test applied to

Table 7 Realism results summary

Exemplar Real-World Social Network	PS vs. CM			CDM vs. CM		
	PS	CM	=	CDM	CM	=
Robins Australian Bank	15	4	1	14	5	1
Roethlisberger & Dickson Bank Wiring Room	9	10	1	9	9	2
Thurman Office	13	6	1	14	5	1
Sampson Monastery	10	8	2	7	9	4
Krackhardt Office CSS	9	10	1	10	9	1
Krackhardt High-Tech Managers	11	8	1	9	9	2
Schwimmer Taro Exchange	5	14	1	5	14	1
Webster Accounting Firm	9	9	2	9	9	2
Zachary Karate Club	9	8	3	10	8	2
Bernard & Killworth Technical	13	5	2	13	5	2
Bernard & Killworth Office	11	6	3	11	6	3
Krebs Fortune 500 IT Department (Advice)	9	8	3	10	7	3
Krebs Fortune 500 IT Department (Business)	7	9	4	8	7	5
Lazega Law Firm	12	2	6	11	3	6
Total	142	107	31	140	105	35

CDM versus CM has $r = 140$ and $n = 140 + 105 = 245$. The results are test statistic $p = 0.571429$, $z = 2.236068$, and p-value $= 0.01264$, which is again $< \alpha$, thus we again reject the null hypothesis and conclude that CDM outperforms CM.

To support the quantitative realism results at an intuitive level, Fig. 3 presents an example visual comparison of a real world social network with a randomly generated network and two networks that were generated using a personality compatibility table. Figure 3a shows the Robins Australian Bank social network (Pattison et al., 2000). Figure 3b shows a network that was generated using the random $G(n, p)$ algorithm. That network has the same number of nodes and network density as the exemplar real world social network. Figure 3c shows a synthetic social network generated using an assignment of personality types found by the PS algorithm. Figure 3d shows a synthetic social network generated using an assignment of personality types found by the CDM algorithm. In the figure, node communities found by the walktrap.community function in the R igraph package are depicted with bounding boxes around them. A visual inspection of the networks in the reveals what appear to be more realistic communities within Fig. 3c and d.

Conclusions and future work

This section states the conclusions of this work and suggests possible future work.

Conclusions

The PS and CDM algorithms differ from most prior work on generating synthetic social network in a significant way. Most prior algorithms do not consider the attributes of the nodes, or of the people or entities the nodes represent, when adding links; instead they are

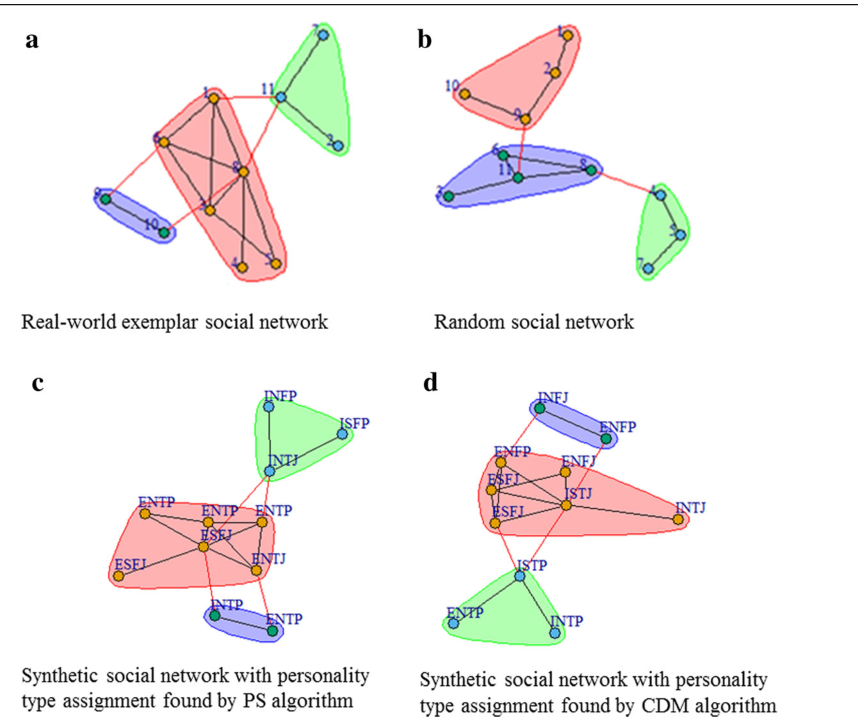

Fig. 3 Visual comparison of the real world and synthetic social networks

based on retaining or replicating some of the structural characteristics of the exemplar network in the synthetic networks. For example, CM is given a degree sequence, which may be the actual degree sequence of the real-world network serving as an exemplar (Newman 2003). In contrast, the PS and CDM algorithms use the attributes of the nodes, in particular the personality types assigned to them, as the primary driver of their calculations.

From the quantitative results, it is evident that both the PS and the CDM algorithms, which use personality compatibility information, generate more realistic synthetic social networks than the CM algorithm, which does not. The PS and CDM algorithms are quite similar in terms of realism. However, the CDM algorithm is much more computationally efficient, requiring substantially shorter execution times for large networks. Either PS or CDM could be used with small to medium exemplars; for exemplars with more than ~ 40 nodes, PS becomes impractical, at least in its current implementation.

Close examination of the results in Table 6 show that the PS and CDM both performed worst on the Schwimer Taro Exchange exemplar. It is unlikely to be a coincidence that in that network only among the fourteen exemplars the nodes correspond not to individual people, but to households, which is intuitively not as good a fit with personality-based algorithms. Thus PS and CDM, or future enhancements of them, should be considered when the nodes correspond to individual people and personality compatibility is expected to have a significant effect on whether two people have the relationship that a link represents.

Future work

Because the PS and CDM algorithms both produce personality assignments that are then input to the GNAC algorithm to generate synthetic social networks, we make two conjectures that motivate future work. First, we conjecture that any method to find an effective personality type assignment A could be combined with the GNAC algorithm to synthesize realistic social networks. Second, we conjecture that the method does not depend on a single personality type scheme, such as the MBTI scheme used in this work. Rather, we believe that any personality type scheme from which a personality compatibility table is available or can be inferred could be combined with the PS and CDM algorithms to generate realistic synthetic social networks. For example, a similar table construction process could be applied to the OCEAN personality type model, with the additional preliminary step of discretizing continuous scales for each personality factors into a finite number of discrete values or intervals.

In this work all of the social networks were treated as symmetric and unweighted. As an obvious generalization, applying these methods to asymmetric and/or weighted social networks is an opportunity for future work. Because multiple metrics generated in a single experiment are analyzed, the multiple comparison problem may be present, and suitable methods to compensate for it could be employed. Finally, the assumption in the PS algorithm that the nodes can be distinguished could be changed to consider networks that connect the same personality types in the same way, as opposed to connecting the same nodes in the same way, as equivalent. (This is analogous to color isomorphism in graph theory terms.) Changing the assumption would change the formula for calculating the network probability $P(G)$.

Finally, according to (Aiello et al., 2012), there has been considerable research aimed at predicting the overall evolution of social networks, but very few attempts to predict future connections of individual people within such networks. Within an organization, managers may wish to create a new project team or work group. The methods developed in this work could be applied to simulating the potential formation of social networks within the team or group, given a set of personality types and a compatibility table. We speculate that generating synthetic social networks using individuals' personality types has the potential to lead to a predictive or semi-predictive capability to anticipate the future social network that could emerge in a team or group. If such a capability was sufficiently reliable, managers could use its predictions when considering personnel assignments. This idea requires of careful validation, perhaps by comparing predicted social networks to actual social networks for existing teams or groups.

Appendix 1

Constructing a personality compatibility table for the MBTI

This appendix details the process used to construct a personality compatibility table for the 16 MBTI personality types. The process had these steps:

1. Identify a set of environmental factors that are important in determining personality compatibility; for this work eight such factors were identified.
2. Interpret the personality model to determine each personality type's opinion regarding each of the environmental factors.
3. Perform pair-wise comparisons of 16 MBTI personality types to determine the number of shared or consistent opinions regarding the environmental factors between each pair of personality types.
4. Scale the counts of common opinions into probabilities of link formation for the compatibility table.

In the first step, environmental factors important in determining personality compatibility were identified by examining the sources describing the personality model. Within a workplace environment, the factors that may determine compatibility of colleagues include:

- Authority; a tendency to respect or work with the chain of command.
- Communication; a tendency to value accurate and specific vernacular.
- Consideration; a tendency to respect or incorporate other people's opinions.
- Empathy; a tendency to recognize or synchronize with other people's feelings.
- Harmony; a tendency to tolerate or relieve interpersonal tensions.
- Loyalty; tendency to value relationships and defend alliances.
- Productivity; a tendency to value efficient processes or creating something.
- Rules; a tendency to follow and defend documented procedures.

The following quotations from (Keirsey 1998) illustrate the source content from which the environmental factors could be identified and the various personality types' likely opinions of them were determined. Environmental factors noted after each quotation indicate that the associated MBTI may have positive or negative attitude about those factors.

- Promoters (ESTP) "[have a] low tolerance for anxiety and are apt to leave relationships that are filled with interpersonal tensions." (Harmony, Loyalty)

- Composers (ISFP) "will put up with a lot more interpersonal tensions than other Artisans" (Harmony, Loyalty).
- Crafters (ISTP) "can be fiercely insubordinate, seeing hierarchy and authority as unnecessary and even irksome." (Authority, Rules)
- Performers (ESFP) "tolerance for anxiety is the lowest of all the types, and they will avoid worries and troubles by ignoring the unhappiness of a situation as long as possible." (Harmony, Productivity)
- Supervisors (ESTJ) "may not always be responsive to points of view and emotions of others and have a tendency to jump to conclusions too quickly." (Authority, Productivity)
- Providers (ESFJ) "tend to listen to acknowledged authorities on abstract matters, and often rely on officially sanctioned views as the source of their opinions and attitudes." (Authority, Rules)
- Inspectors (ISTJ) "Because of [being adamant about rule compliance,] they are often misjudged as having ice in their veins, for people fail to see their good intentions and their vulnerability to criticism." (Authority, Rules)
- Protectors (ISFJ) "know the value of a dollar and abhor the squandering or misuse of resources." (Productivity)
- Teachers (ENFJ) "When [they] find that their position or beliefs were not comprehended or accepted, they are surprised, puzzled, and sometimes hurt." (Communications, Harmony, Consideration)
- Counselors (INFJ) "value staff harmony and want an organization to run smoothly and pleasantly, making every effort themselves to contribute to that end." (Harmony, Consideration, Productivity)
- Champions (ENFP) "Sometimes [they] get impatient with their superiors; and they will occasionally side with detractors of their organization, who find in them a sympathetic ear and a natural rescuer." (Authority, Communication, Empathy)
- Healers (INFP) "have difficulty thinking in conditional 'if-then' terms; they tend to see things as either black or white, and can be impatient with contingency." (Communication, Empathy, Consideration)
- Fieldmarshals (ENTJ) "For the [Fieldmarshall], there must always be a reason for doing anything, and peoples' feelings usually are not sufficient reason." (Authority, Rules, Productivity)
- Masterminds (INTJ) "Colleagues may describe [Masterminds] as unemotional and, at times, cold and dispassionate, when in truth they are merely taking the goals of an institution seriously, and continually striving to achieve those goals." (Productivity, Rules)
- Inventors (ENTP) "If an [Inventor's] job becomes dull and repetitive, they tend to lose interest and fail to follow through -- often to the discomfort of colleagues." (Productivity)
- Architect (INTP) "It is difficult for an [Architect] to listen to nonsense, even in a casual conversation, without pointing out the speaker's error, and this makes communication with them an uncomfortable experience for many." (Communication, Consideration)

Based on these quotes and other similar descriptions of the personality types, their likely opinions regarding the environmental factors were determined. Table 8 shows the result. The Keirsey temperaments scheme groups the 16 possible MBTI personality types into four categories, referred to as Artisans, Guardians, Idealists, and Rationals (Keirsey, 1998); the table is organized by those categories. In the table, a 0 indicates that people of the personality type are likely to hold a low or negative opinion of the environmental factor, whereas a 1 indicates a relatively high or positive opinion.

For each pair of personality types X and Y, the number of environmental factors on which they agreed (both had 0 or both had 1 in the table) was calculated; let that value be denoted as $a(X, Y)$, with $a(X, Y) \in \{0, 1, 2, \ldots, 6\}$. (The pairwise comparison considered six environmental factors, hence six was the maximum number of possible agreements. The maximum number of agreed upon factors by any pair of two distinct personality types was actually five.) The probability of a link forming between personality types X and Y was calculated as

$$p(X, Y) = 0.5 \cdot \left(1 + \operatorname{erf}\left(\frac{(x-\mu)}{(\sigma \cdot \sqrt{2})}\right)\right)$$

where $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$ is the Gauss error function, $\mu \approx 2.9747$, and $\sigma \approx 1.8185$.

The values for μ and σ were determined empirically. The result of this formula is that $0.05 \le p(X, Y) \le 0.95$ for all personality types X and Y, leaving a small but non-zero probability (0.05) of a link forming and a small probability of link not forming (also 0.05) between any two personality types. The $p(X, Y)$ values were recorded in the personality compatibility table. The resulting personality compatibility table produced by this process and used in this work was shown earlier in Table 4.

Other methods of determining the compatibility table values are possible, of course. The synthetic social network generation algorithm will operate with any reasonable and internally consistent compatibility table.

Table 8 Inferred MBTI personality types' opinions of environmental factors

Category	Personality type		Environmental factor					
			Authority	Communication	Harmony	Loyalty	Productivity	Rules
Artisans	Promoter	ESTP	0	1	0	0	0	0
	Composer	ISFP	0	0	1	1	1	0
	Crafter	ISTP	0	1	0	1	1	1
	Performer	ISFP	1	0	0	0	0	1
Guardians	Supervisor	ESTJ	1	1	0	1	1	1
	Provider	ESFJ	1	0	0	1	0	1
	Inspector	ISTJ	1	1	0	1	0	1
	Protector	ISFJ	1	1	0	1	1	1
Idealists	Teacher	ENFJ	1	1	1	0	0	0
	Counselor	INFJ	0	0	1	1	0	0
	Champion	ENFP	0	0	0	1	1	0
	Healer	INFP	0	0	1	0	0	1
Rationals	Fieldmarshal	ENTJ	1	0	1	1	0	0
	Mastermind	INTJ	0	1	1	0	1	0
	Inventor	ENTP	0	1	0	0	1	0
	Architect	INTP	0	1	1	0	1	1

Appendix 2

Detailed realism results

The following tables report the detailed realism results for all fourteen of the real-world social networks used as exemplars.

Table 9 Realism results for the Robins Australian Bank social network

Metrics	T	$\overline{F}$	$\lvert T-\overline{F}\rvert$	L1(F)	L2(F)	$\overline{P}$	$\lvert T-\overline{P}\rvert$	L1(P)	L2(P)	$\overline{M}$	$\lvert T-\overline{M}\rvert$	L1(M)	L2(M)
Nodes	11.00	11.00	0.00	0.00	0.00	11.00	0.00	0.00	0.00	11.00	0.00	0.00	0.00
Links	16.00	12.83	3.17	95.00	18.41	16.00	**0.00**	**0.00**	**0.00**	16.00	**0.00**	**0.00**	**0.00**
Components	1.00	1.13	**0.13**	**4.00**	**2.00**	1.07	**0.07**	**2.00**	**1.41**	1.47	0.47	14.00	3.74
Network density	0.29	0.23	0.06	1.73	0.34	0.29	**0.00**	**0.00**	**0.00**	0.29	**0.00**	**0.00**	**0.00**
Average degree	2.91	2.33	0.58	17.27	3.35	2.91	**0.00**	**0.00**	**0.00**	2.91	**0.00**	**0.00**	**0.00**
Standard deviation degree	1.87	1.27	0.60	17.92	3.49	1.75	**0.12**	**8.79**	**1.93**	1.82	**0.05**	**7.46**	**1.70**
Global cluster coefficient	0.38	0.14	0.24	7.29	1.48	0.38	**0.01**	**1.27**	**0.32**	0.45	**0.07**	**2.76**	**0.61**
Average cluster coefficient	0.41	0.17	**0.24**	**7.46**	1.55	0.63	**0.23**	**6.80**	**1.29**	0.66	0.26	7.73	**1.48**
Mean path length	2.02	2.73	**0.71**	**21.20**	6.28	2.13	**0.11**	**6.55**	**2.27**	2.75	0.74	23.69	**5.83**
Communities	3.00	3.47	**0.47**	18.00	5.10	2.47	0.53	18.00	**4.24**	3.13	**0.13**	**14.00**	**4.47**
Gini coefficient	0.24	0.15	**0.10**	**2.91**	**0.66**	0.12	0.12	3.72	0.76	0.20	**0.04**	**2.85**	**0.62**
Average betweenness	5.09	6.44	1.35	53.64	11.35	5.06	**0.04**	**21.09**	**4.75**	4.52	**0.57**	**29.00**	**6.08**
Maximum betweenness	25.17	21.60	3.56	**153.90**	35.15	26.26	**1.09**	**140.33**	**31.75**	22.18	**2.99**	159.33	**34.81**
Average closeness	0.05	0.05	0.01	0.24	0.05	0.05	**0.00**	**0.11**	**0.03**	0.06	**0.00**	**0.18**	**0.04**
Minimum closeness	0.04	0.03	0.01	0.25	0.05	0.04	**0.00**	**0.11**	**0.03**	0.04	**0.00**	**0.16**	**0.04**
Average eigencentrality	0.49	0.54	**0.05**	**1.91**	0.46	0.55	0.06	1.96	**0.43**	0.56	0.07	2.11	**0.44**
Minimum eigencentrality	0.14	0.15	**0.01**	**2.08**	**0.46**	0.22	0.08	2.66	0.54	0.20	0.06	**2.05**	**0.43**
Network radius	2.00	2.67	0.67	20.00	4.69	2.10	**0.10**	**3.00**	**1.73**	2.00	**0.00**	**2.00**	**1.41**
Average eccentricity	3.09	3.78	0.69	23.73	5.26	3.01	**0.09**	**10.73**	**2.32**	2.86	**0.23**	**11.55**	**2.56**
Network diameter	4.00	4.73	0.73	26.00	6.48	3.57	**0.43**	**17.00**	**4.12**	3.57	**0.43**	**13.00**	**3.87**

Boldfaced numbers indicate which algorithm performed better for a particular metric

Table 10 Realism results for the Roethlisberger & Dickson Bank Wiring Room social network

Metrics	T	$\overline{F}$	$\lvert T-\overline{F}\rvert$	L1(F)	L2(F)	$\overline{P}$	$\lvert T-\overline{P}\rvert$	L1(P)	L2(P)	$\overline{M}$	$\lvert T-\overline{M}\rvert$	L1(M)	L2(M)
Nodes	14.00	14.00	0.00	0.00	0.00	14.00	0.00	0.00	0.00	14.00	0.00	0.00	0.00
Links	13.00	10.43	2.57	77.00	14.93	13.00	**0.00**	**0.00**	**0.00**	13.00	**0.00**	**0.00**	**0.00**
Components	6.00	6.10	**0.10**	**3.00**	**1.73**	6.43	0.43	13.00	3.61	6.10	0.10	3.00	1.73
Network density	0.14	0.12	0.03	0.85	0.16	0.14	**0.00**	**0.00**	**0.00**	0.14	**0.00**	**0.00**	**0.00**
Average degree	1.86	1.49	0.37	11.00	2.13	1.86	**0.00**	**0.00**	**0.00**	1.86	**0.00**	**0.00**	**0.00**
Standard deviation degree	1.61	1.35	0.26	7.90	1.65	1.85	**0.24**	**7.30**	**1.63**	1.74	**0.13**	**4.38**	**1.07**
Global cluster coefficient	0.64	0.16	0.48	14.44	2.76	0.45	**0.20**	**5.88**	**1.24**	0.45	**0.19**	**5.80**	**1.14**
Average cluster coefficient	0.71	0.17	0.53	16.02	3.08	0.63	**0.08**	**4.89**	**0.95**	0.64	**0.07**	**3.29**	**0.71**
Mean path length	9.34	9.49	**0.15**	**5.78**	**2.23**	9.62	0.28	13.51	3.10	9.32	**0.02**	5.84	**1.66**
Communities	7.00	7.73	0.73	22.00	5.29	7.60	**0.60**	**18.00**	**4.69**	7.60	**0.60**	**18.00**	**4.90**
Gini coefficient	0.37	0.32	**0.04**	**1.40**	**0.34**	0.34	**0.03**	**0.85**	**0.26**	0.32	0.05	1.55	0.39
Average betweenness	3.14	3.16	**0.01**	**18.57**	**4.12**	1.73	1.42	42.50	8.28	2.28	0.86	25.93	5.22
Maximum betweenness	16.00	12.84	3.16	**101.42**	**25.42**	13.31	**2.69**	159.33	35.77	15.96	**0.04**	117.17	25.69
Average closeness	0.06	0.06	**0.00**	**0.20**	**0.05**	0.08	0.02	0.64	0.13	0.07	0.01	0.33	0.07
Minimum closeness	0.04	0.04	**0.00**	**0.17**	**0.05**	0.06	0.02	0.65	0.13	0.05	0.01	0.30	0.06
Average eigencentrality	0.59	0.65	**0.05**	**1.98**	**0.43**	0.68	0.08	2.67	0.69	0.65	0.06	2.44	0.52
Minimum eigencentrality	0.20	0.21	**0.01**	**2.22**	**0.54**	0.34	0.14	4.15	0.93	0.30	0.10	3.21	0.73
Network radius	3.00	2.67	**0.33**	**12.00**	**3.46**	1.83	1.17	35.00	6.71	2.00	1.00	30.00	5.48
Average eccentricity	2.57	2.38	**0.19**	**10.50**	**2.37**	1.55	1.02	30.71	5.87	1.91	0.67	20.00	3.85
Network diameter	5.00	4.57	**0.43**	**19.00**	**4.58**	3.00	2.00	60.00	11.40	3.57	1.43	43.00	8.31

Boldfaced numbers indicate which algorithm performed better for a particular metric

Table 11 Realism results for the Thurman Office social network

Metrics	T	$\overline{F}$	$\lvert T-\overline{F}\rvert$	L1(F)	L2(F)	$\overline{P}$	$\lvert T-\overline{P}\rvert$	L1(P)	L2(P)	$\overline{M}$	$\lvert T-\overline{M}\rvert$	L1(M)	L2(M)
Nodes	15.00	15.00	0.00	0.00	0.00	15.00	0.00	0.00	0.00	15.00	0.00	0.00	0.00
Links	33.00	25.53	7.47	224.00	42.07	33.00	**0.00**	**0.00**	**0.00**	33.00	**0.00**	**0.00**	**0.00**
Components	1.00	1.07	0.07	2.00	1.41	1.00	**0.00**	**0.00**	**0.00**	1.03	**0.03**	**1.00**	**1.00**
Network density	0.31	0.24	0.07	2.13	0.40	0.31	**0.00**	**0.00**	**0.00**	0.31	**0.00**	**0.00**	**0.00**
Average degree	4.40	3.40	1.00	29.87	5.61	4.40	**0.00**	**0.00**	**0.00**	4.40	**0.00**	**0.00**	**0.00**
Standard deviation degree	2.53	1.78	0.75	22.57	4.29	3.03	**0.50**	**15.07**	**2.91**	2.94	**0.41**	**12.43**	**2.39**
Global cluster coefficient	0.52	0.25	0.27	7.96	1.51	0.47	**0.05**	**1.47**	**0.32**	0.47	**0.05**	**1.46**	**0.32**
Average cluster coefficient	0.48	0.28	**0.20**	**6.34**	**1.28**	0.73	0.25	7.49	1.38	0.71	0.23	6.90	1.29
Mean path length	1.88	2.27	0.39	11.83	3.16	1.80	**0.08**	**2.47**	**0.50**	1.86	**0.02**	**3.67**	**1.63**
Communities	3.00	3.77	**0.77**	**31.00**	**7.42**	3.87	0.87	48.00	10.10	3.63	**0.63**	35.00	8.43
Gini coefficient	0.18	0.22	**0.05**	**2.44**	**0.57**	0.28	0.11	4.03	0.80	0.28	0.11	3.48	0.76
Average betweenness	6.13	8.02	1.89	56.67	11.24	5.61	**0.53**	**17.27**	**3.53**	5.59	**0.54**	**16.73**	**3.67**
Maximum betweenness	37.25	28.64	**8.61**	**282.45**	**57.01**	47.22	9.97	312.71	64.49	47.27	10.03	307.51	65.38
Average closeness	0.04	0.03	0.01	0.14	0.03	0.04	**0.00**	**0.06**	**0.01**	0.04	**0.00**	**0.06**	**0.01**
Minimum closeness	0.03	0.02	0.01	0.18	0.04	0.03	**0.00**	**0.06**	**0.01**	0.03	**0.00**	**0.04**	**0.01**
Average eigencentrality	0.53	0.54	**0.01**	**1.08**	0.24	0.49	0.04	1.21	**0.24**	0.49	0.04	1.21	**0.24**
Minimum eigencentrality	0.11	0.12	**0.01**	**1.39**	**0.29**	0.15	0.04	1.46	**0.29**	0.15	0.04	1.68	0.32
Network radius	2.00	2.67	0.67	20.00	4.47	2.00	**0.00**	**0.00**	**0.00**	2.00	**0.00**	**0.00**	**0.00**
Average eccentricity	2.80	3.48	0.68	20.33	4.09	2.67	**0.13**	**7.13**	**1.45**	2.65	**0.15**	**5.93**	**1.31**
Network diameter	3.00	4.30	1.30	39.00	7.68	3.27	**0.27**	**8.00**	**2.83**	3.10	**0.10**	**3.00**	**1.73**

Boldfaced numbers indicate which algorithm performed better for a particular metric

Table 12 Realism results for the Sampson Monastery social network

Metrics	T	$\overline{F}$	$\lvert T\text{-}\overline{F}\rvert$	L1(F)	L2(F)	$\overline{P}$	$\lvert T\text{-}\overline{P}\rvert$	L1(P)	L2(P)	$\overline{M}$	$\lvert T\text{-}\overline{M}\rvert$	L1(M)	L2(M)
Nodes	18.00	18.00	0.00	0.00	0.00	18.00	0.00	0.00	0.00	18.00	0.00	0.00	0.00
Links	41.00	34.07	6.93	208.00	39.19	41.00	**0.00**	**0.00**	**0.00**	41.00	**0.00**	**0.00**	**0.00**
Components	1.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
Network density	0.27	0.22	0.05	1.36	0.26	0.27	**0.00**	**0.00**	**0.00**	0.27	**0.00**	**0.00**	**0.00**
Average degree	4.56	3.79	0.77	23.11	4.36	4.56	**0.00**	**0.00**	**0.00**	4.56	**0.00**	**0.00**	**0.00**
Standard deviation degree	2.09	1.55	**0.55**	**16.35**	**3.25**	2.84	0.75	22.36	4.35	2.99	0.90	26.91	5.12
Global cluster coefficient	0.26	0.20	**0.07**	**2.31**	**0.50**	0.36	0.10	2.93	0.56	0.37	0.11	3.20	0.60
Average cluster coefficient	0.29	0.21	**0.07**	**2.75**	**0.59**	0.63	0.34	10.22	1.92	0.66	0.38	11.36	2.10
Mean path length	1.97	2.15	0.18	5.33	1.06	1.83	**0.14**	**4.07**	**0.77**	1.82	**0.15**	**4.54**	**0.87**
Communities	3.00	3.67	**0.67**	28.00	7.07	3.60	**0.60**	**22.00**	**5.10**	3.83	0.83	**25.00**	**5.75**
Gini coefficient	0.07	0.19	0.12	3.52	0.82	0.18	**0.10**	**3.21**	**0.66**	0.19	**0.11**	**3.46**	**0.72**
Average betweenness	8.22	9.73	1.51	45.33	9.00	7.07	**1.15**	**34.56**	**6.54**	6.94	**1.29**	**38.56**	**7.39**
Maximum betweenness	37.62	36.81	**0.81**	**170.51**	**39.48**	78.27	40.65	1219.35	228.02	82.56	44.94	1348.16	250.19
Average closeness	0.03	0.03	0.00	**0.08**	0.02	0.03	**0.00**	**0.07**	**0.01**	0.03	0.00	0.08	0.02
Minimum closeness	0.02	0.02	**0.00**	**0.07**	**0.02**	0.03	0.00	0.09	0.02	0.03	0.00	0.10	0.02
Average eigencentrality	0.48	0.50	**0.02**	**1.32**	**0.31**	0.42	0.06	1.89	0.36	0.41	0.07	2.06	0.38
Minimum eigencentrality	0.17	0.16	**0.01**	**1.78**	**0.40**	0.23	0.06	1.79	**0.37**	0.22	0.05	1.92	0.40
Network radius	2.00	2.80	0.80	24.00	4.90	1.97	**0.03**	**1.00**	**1.00**	1.93	**0.07**	**2.00**	**1.41**
Average eccentricity	3.00	3.40	0.40	12.50	**2.56**	2.65	**0.35**	**10.50**	**2.20**	2.60	0.40	**12.39**	2.67
Network diameter	4.00	4.07	**0.07**	**6.00**	**2.45**	3.00	1.00	30.00	5.66	3.07	0.93	28.00	5.66

Boldfaced numbers indicate which algorithm performed better for a particular metric

Table 13 Realism results for the Krackhardt Office CSS social network

Metrics	T	$\overline{F}$	$\lvert T-\overline{F}\rvert$	L1(F)	L2(F)	$\overline{P}$	$\lvert T-\overline{P}\rvert$	L1(P)	L2(P)	$\overline{M}$	$\lvert T-\overline{M}\rvert$	L1(M)	L2(M)
Nodes	21.00	21.00	0.00	0.00	0.00	21.00	0.00	0.00	0.00	21.00	0.00	0.00	0.00
Links	14.00	12.50	1.50	45.00	9.75	14.00	**0.00**	**0.00**	**0.00**	14.00	**0.00**	**0.00**	**0.00**
Components	9.00	9.43	0.43	17.00	4.58	9.13	**0.13**	**4.00**	**2.00**	9.17	**0.17**	**5.00**	**2.24**
Network density	0.07	0.06	0.01	0.21	0.05	0.07	**0.00**	**0.00**	**0.00**	0.07	**0.00**	**0.00**	**0.00**
Average degree	1.33	1.19	0.14	4.29	0.93	1.33	**0.00**	**0.00**	**0.00**	1.33	**0.00**	**0.00**	**0.00**
Standard deviation degree	1.39	1.18	**0.21**	**6.23**	**1.38**	2.08	0.69	20.55	3.93	2.11	0.72	21.70	4.06
Global cluster coefficient	0.13	0.08	0.05	2.91	0.60	0.14	**0.01**	**0.69**	**0.14**	0.13	**0.01**	**0.59**	**0.13**
Average cluster coefficient	0.16	0.11	**0.05**	**3.80**	**0.76**	0.68	0.53	15.78	2.90	0.70	0.55	16.38	3.01
Mean path length	15.84	16.45	**0.61**	**41.12**	8.83	14.37	1.48	44.27	**8.77**	14.53	1.31	41.18	**8.38**
Communities	10.00	11.33	1.33	40.00	8.72	10.47	**0.47**	**14.00**	**3.74**	10.50	**0.50**	**17.00**	**4.36**
Gini coefficient	0.40	0.33	0.07	2.04	0.43	0.42	**0.02**	**0.77**	**0.17**	0.42	**0.02**	**0.80**	**0.18**
Average betweenness	3.67	3.86	**0.20**	53.33	12.11	3.90	0.23	**27.43**	**5.73**	3.51	**0.16**	**22.48**	**5.23**
Maximum betweenness	22.50	25.14	**2.64**	**315.33**	**71.60**	55.97	33.47	1004.0	186.61	54.70	32.20	966.0	182.25
Average closeness	0.04	0.06	0.01	0.66	0.24	0.04	**0.00**	**0.17**	**0.04**	0.05	**0.00**	**0.17**	**0.05**
Minimum closeness	0.03	0.04	0.01	0.52	0.19	0.04	**0.00**	**0.18**	**0.05**	0.04	**0.01**	**0.21**	**0.06**
Average eigencentrality	0.47	0.52	**0.05**	**2.33**	0.56	0.39	0.08	2.36	**0.45**	0.39	0.09	2.54	**0.48**
Minimum eigencentrality	0.11	0.17	**0.07**	**2.62**	**0.64**	0.25	0.14	4.30	0.82	0.25	0.14	4.19	0.81
Network radius	3.00	2.87	**0.13**	**14.00**	**4.00**	1.83	1.17	35.00	6.71	1.83	1.17	35.00	6.71
Average eccentricity	2.33	2.41	**0.08**	**13.00**	**3.07**	1.77	0.57	17.05	3.68	1.66	0.67	20.10	3.99
Network diameter	5.00	5.20	**0.20**	**26.00**	**6.16**	3.20	1.80	54.00	10.58	3.00	2.00	60.00	11.40

Boldfaced numbers indicate which algorithm performed better for a particular metric

Table 14 Realism results for the Krackhardt High-Tech Managers social network

Metrics	T	$\overline{F}$	$\lvert T-\overline{F}\rvert$	L1(F)	L2(F)	$\overline{P}$	$\lvert T-\overline{P}\rvert$	L1(P)	L2(P)	$\overline{M}$	$\lvert T-\overline{M}\rvert$	L1(M)	L2(M)
Nodes	21.00	21.00	0.00	0.00	0.00	21.00	0.00	0.00	0.00	21.00	0.00	0.00	0.00
Links	36.00	31.47	4.53	136.00	26.42	36.00	**0.00**	**0.00**	**0.00**	36.00	**0.00**	**0.00**	**0.00**
Components	5.00	5.00	**0.00**	**0.00**	**0.00**	5.10	0.10	3.00	1.73	5.00	0.00	0.00	0.00
Network density	0.17	0.15	0.02	0.65	0.13	0.17	**0.00**	**0.00**	**0.00**	0.17	**0.00**	**0.00**	**0.00**
Average degree	3.43	3.00	0.43	12.95	2.52	3.43	**0.00**	**0.00**	**0.00**	3.43	**0.00**	**0.00**	**0.00**
Standard deviation degree	2.14	1.88	**0.26**	**7.82**	**1.56**	2.70	0.56	16.83	3.18	2.72	0.59	17.60	3.37
Global cluster coefficient	0.50	0.19	0.31	9.23	1.70	0.37	**0.13**	**3.83**	**0.72**	0.39	**0.10**	**3.12**	**0.60**
Average cluster coefficient	0.59	0.20	0.39	11.63	2.14	0.60	**0.02**	**1.44**	**0.32**	0.63	**0.05**	**1.79**	**0.41**
Mean path length	8.98	8.79	**0.18**	**5.71**	**1.10**	8.81	**0.17**	11.42	2.45	8.67	0.30	9.13	1.68
Communities	7.00	7.73	0.73	24.00	6.48	7.63	**0.63**	**23.00**	**5.75**	7.53	**0.53**	**20.00**	**4.69**
Gini coefficient	0.44	0.41	0.03	1.48	0.34	0.42	**0.02**	**1.09**	**0.27**	0.41	**0.03**	**1.17**	**0.26**
Average betweenness	9.29	7.46	**1.83**	**57.10**	**10.97**	6.11	3.18	95.29	17.67	6.24	3.04	91.33	16.82
Maximum betweenness	44.67	27.47	**17.20**	**544.5**	**104.29**	57.55	**12.88**	**394.28**	**87.09**	65.20	20.54	636.21	127.4
Average closeness	0.03	0.03	**0.00**	**0.10**	**0.02**	0.03	0.01	0.20	0.04	0.03	0.01	0.19	0.04
Minimum closeness	0.02	0.02	**0.00**	**0.14**	**0.03**	0.03	0.01	0.18	0.04	0.02	0.01	0.16	0.03
Average eigencentrality	0.45	0.58	0.13	3.83	0.76	0.47	**0.02**	**0.96**	**0.23**	0.44	**0.01**	**0.80**	**0.16**
Minimum eigencentrality	0.04	0.19	0.15	4.56	0.91	0.15	**0.11**	**3.21**	**0.63**	0.14	**0.10**	**2.91**	**0.58**
Network radius	3.00	2.83	**0.17**	**5.00**	**2.24**	2.07	0.93	28.00	5.29	2.07	0.93	28.00	5.29
Average eccentricity	3.38	2.79	**0.59**	**19.29**	**3.62**	2.45	0.93	27.95	5.22	2.44	0.94	28.10	5.24
Network diameter	5.00	4.20	**0.80**	**28.00**	**5.29**	3.77	1.23	37.00	7.42	3.63	1.37	41.00	8.19

Boldfaced numbers indicate which algorithm performed better for a particular metric

Table 15 Realism results for the Schwimmer Taro Exchange social network

Metrics	T	$\overline{F}$	$\lvert T-\overline{F}\rvert$	L1(F)	L2(F)	$\overline{P}$	$\lvert T-\overline{P}\rvert$	L1(P)	L2(P)	$\overline{M}$	$\lvert T-\overline{M}\rvert$	L1(M)	L2(M)
Nodes	22.00	22.00	0.00	0.00	0.00	22.00	0.00	0.00	0.00	22.00	0.00	0.00	0.00
Links	39.00	35.47	3.53	106.0	20.98	39.00	**0.00**	**0.00**	**0.00**	39.00	**0.00**	**0.00**	**0.00**
Components	1.00	1.00	**0.00**	**0.00**	**0.00**	1.30	0.30	9.00	3.32	1.23	0.23	7.00	2.65
Network density	0.17	0.15	0.02	0.46	0.09	0.17	**0.00**	**0.00**	**0.00**	0.17	**0.00**	**0.00**	**0.00**
Average degree	3.55	3.22	0.32	9.64	1.91	3.55	**0.00**	**0.00**	**0.00**	3.55	**0.00**	**0.00**	**0.00**
Standard deviation degree	0.96	0.96	**0.00**	**2.90**	**0.69**	2.65	1.69	50.71	9.33	2.69	1.73	51.78	9.49
Global cluster coefficient	0.28	0.11	0.17	4.99	0.96	0.33	**0.06**	**1.77**	**0.37**	0.32	**0.05**	**1.35**	**0.29**
Average cluster coefficient	0.34	0.11	**0.23**	**6.88**	**1.30**	0.72	0.38	11.43	2.10	0.72	0.38	11.29	2.08
Mean path length	2.49	2.66	**0.16**	**5.03**	**1.22**	2.91	0.42	20.68	6.67	2.78	0.29	18.58	5.91
Communities	5.00	4.97	**0.03**	**17.00**	**4.58**	5.23	0.23	17.00	**4.36**	5.33	0.33	20.00	5.10
Gini coefficient	0.13	0.20	**0.07**	**2.43**	**0.53**	0.20	0.08	**2.29**	**0.51**	0.21	0.08	2.48	0.55
Average betweenness	15.68	17.41	**1.73**	**52.77**	**12.79**	12.95	2.73	81.91	17.33	12.86	2.82	84.64	17.09
Maximum betweenness	46.38	53.76	**7.38**	**319.93**	**80.14**	157.15	110.76	3322.93	613.21	157.83	111.44	3343.28	615.46
Average closeness	0.02	0.02	**0.00**	**0.03**	**0.01**	0.02	0.00	0.09	0.02	0.02	0.00	0.09	0.02
Minimum closeness	0.02	0.01	0.00	0.07	0.02	0.02	**0.00**	**0.03**	**0.01**	0.02	**0.00**	**0.02**	**0.01**
Average eigencentrality	0.62	0.51	**0.10**	**3.08**	**0.64**	0.33	0.29	8.69	1.59	0.33	0.29	8.67	1.58
Minimum eigencentrality	0.32	0.15	**0.16**	**4.92**	**1.01**	0.07	0.24	7.28	1.34	0.08	0.23	7.00	1.28
Network radius	3.00	3.40	**0.40**	**12.00**	**3.46**	2.17	0.83	25.00	5.00	2.20	0.80	24.00	4.90
Average eccentricity	4.09	4.40	**0.31**	**11.14**	**2.86**	3.32	0.77	23.32	4.59	3.32	0.77	23.00	4.45
Network diameter	5.00	5.40	**0.40**	**16.00**	**4.47**	4.17	0.83	25.00	5.00	4.20	0.80	24.00	4.90

Boldfaced numbers indicate which algorithm performed better for a particular metric

Table 16 Realism results for the Webster Accounting Firm social network

Metrics	T	$\overline{F}$	$\lvert T-\overline{F}\rvert$	L1(F)	L2(F)	$\overline{P}$	$\lvert T-\overline{P}\rvert$	L1(P)	L2(P)	$\overline{M}$	$\lvert T-\overline{M}\rvert$	L1(M)	L2(M)
Nodes	24.00	24.00	0.00	0.00	0.00	24.00	0.00	0.00	0.00	24.00	0.00	0.00	0.00
Links	150.0	104.7	45.30	1359.0	249.42	150.0	**0.00**	**0.00**	**0.00**	150.0	**0.00**	**0.00**	**0.00**
Components	2.00	2.00	0.00	0.00	0.00	2.00	0.00	0.00	0.00	2.00	0.00	0.00	0.00
Network density	0.54	0.38	0.16	4.92	0.90	0.54	**0.00**	**0.00**	**0.00**	0.54	**0.00**	**0.00**	**0.00**
Average degree	12.50	8.73	3.78	113.25	20.79	12.50	**0.00**	**0.00**	**0.00**	12.50	**0.00**	**0.00**	**0.00**
Standard deviation degree	5.51	3.48	2.03	60.90	11.18	5.20	**0.31**	**9.19**	**1.88**	5.22	**0.29**	**8.81**	**1.81**
Global cluster coefficient	0.81	0.46	0.36	10.67	1.95	0.71	**0.10**	**2.96**	**0.55**	0.72	**0.09**	**2.84**	**0.52**
Average cluster coefficient	0.78	0.47	0.31	9.32	1.72	0.75	**0.03**	**0.96**	**0.20**	0.74	**0.04**	**1.06**	**0.21**
Mean path length	3.48	3.48	**0.00**	**0.54**	**0.13**	3.30	0.19	5.62	1.03	3.29	0.19	5.63	1.03
Communities	5.00	4.70	0.30	37.00	8.78	4.97	**0.03**	**27.00**	**7.14**	5.07	**0.07**	**28.00**	**7.07**
Gini coefficient	0.52	0.41	0.11	3.36	0.72	0.49	**0.03**	**1.29**	**0.30**	0.48	**0.04**	**1.71**	**0.39**
Average betweenness	6.50	6.49	**0.01**	**6.25**	**1.46**	4.35	2.15	64.63	11.80	4.34	2.16	64.79	11.83
Maximum betweenness	26.80	19.42	**7.38**	**228.43**	**45.64**	17.36	9.45	286.41	55.33	18.18	8.63	260.14	51.33
Average closeness	0.03	0.03	**0.00**	**0.02**	**0.01**	0.03	0.00	0.11	0.02	0.03	0.00	0.11	0.02
Minimum closeness	0.02	0.02	**0.01**	**0.15**	**0.03**	0.02	0.01	0.21	0.04	0.02	0.01	0.22	0.04
Average eigencentrality	0.65	0.69	**0.03**	**1.09**	**0.23**	0.71	0.05	1.58	0.31	0.70	0.04	1.31	0.26
Minimum eigencentrality	0.02	0.21	0.19	5.69	1.07	0.18	**0.16**	**4.80**	**0.90**	0.18	**0.17**	**4.98**	**0.92**
Network radius	2.00	2.00	**0.00**	**0.00**	**0.00**	1.93	0.07	2.00	1.41	1.90	0.10	3.00	1.73
Average eccentricity	2.79	2.25	**0.54**	**16.13**	**2.98**	2.00	0.79	23.79	4.36	1.99	0.80	24.08	4.41
Network diameter	4.00	3.00	**1.00**	**30.00**	**5.48**	2.63	1.37	41.00	7.94	2.67	1.33	40.00	7.75

Boldfaced numbers indicate which algorithm performed better for a particular metric

Table 17 Realism results for the Zachary Karate Club social network

Metrics	T	$\overline{F}$	$\lvert T-\overline{F}\rvert$	L1(F)	L2(F)	$\overline{P}$	$\lvert T-\overline{P}\rvert$	L1(P)	L2(P)	$\overline{M}$	$\lvert T-\overline{M}\rvert$	L1(M)	L2(M)
Nodes	34.00	34.00	0.00	0.00	0.00	34.00	0.00	0.00	0.00	34.00	0.00	0.00	0.00
Links	78.00	65.63	12.37	371.00	68.61	78.00	**0.00**	**0.00**	**0.00**	78.00	**0.00**	**0.00**	**0.00**
Components	1.00	1.13	**0.13**	**4.00**	**2.00**	1.13	0.13	4.00	2.45	1.20	0.20	6.00	2.45
Network density	0.14	0.12	0.02	0.66	0.12	0.14	**0.00**	**0.00**	**0.00**	0.14	**0.00**	**0.00**	**0.00**
Average degree	4.59	3.86	0.73	21.82	4.04	4.59	**0.00**	**0.00**	**0.00**	4.59	**0.00**	**0.00**	**0.00**
Standard deviation degree	3.88	2.69	1.19	35.61	6.63	3.87	**0.00**	**6.10**	**1.38**	3.67	**0.21**	**7.49**	**1.63**
Global cluster coefficient	0.26	0.16	0.10	2.96	0.56	0.31	**0.05**	**1.63**	**0.33**	0.33	**0.08**	**2.25**	**0.44**
Average cluster coefficient	0.59	0.19	0.40	11.85	2.18	0.68	**0.09**	**2.71**	**0.51**	0.67	**0.08**	**2.49**	**0.48**
Mean path length	2.41	2.85	**0.44**	**13.27**	**4.99**	3.05	0.64	22.98	12.16	3.60	1.19	36.32	14.09
Communities	5.00	7.53	2.53	82.00	18.55	6.80	**1.80**	**58.00**	**12.41**	6.27	**1.27**	**42.00**	**9.70**
Gini coefficient	0.17	0.31	**0.14**	**4.28**	**0.87**	0.33	0.17	5.04	0.96	0.31	0.15	4.37	**0.86**
Average betweenness	23.24	25.26	2.03	**68.62**	**14.83**	22.13	**1.11**	72.03	18.05	23.14	**0.10**	72.18	17.73
Maximum betweenness	231.07	163.33	67.74	2032.19	409.13	234.37	**3.30**	**1271.06**	**277.04**	208.98	**22.09**	**1321.97**	**293.17**
Average closeness	0.01	0.01	0.00	**0.02**	**0.00**	0.01	0.00	0.03	0.01	0.01	**0.00**	0.02	0.01
Minimum closeness	0.01	0.01	**0.00**	**0.02**	**0.01**	0.01	0.00	0.03	0.01	0.01	0.00	0.03	0.01
Average eigencentrality	0.39	0.33	**0.06**	**1.79**	**0.36**	0.30	0.09	2.73	0.51	0.30	0.10	2.84	0.53
Minimum eigencentrality	0.06	0.04	**0.03**	**0.99**	**0.20**	0.03	0.03	1.01	0.20	0.03	0.04	1.07	0.21
Network radius	3.00	3.03	**0.03**	**1.00**	**1.00**	2.60	0.40	12.00	3.46	2.73	0.27	10.00	3.16
Average eccentricity	4.03	4.12	**0.09**	**6.94**	**1.79**	3.70	0.33	14.50	3.18	3.82	0.21	12.88	2.97
Network diameter	5.00	5.13	**0.13**	**8.00**	**3.16**	4.70	0.30	15.00	3.87	4.87	0.13	16.00	4.24

Boldfaced numbers indicate which algorithm performed better for a particular metric

Table 18 Realism results for the Bernard & Killworth Technical social network

Metrics	T	$\overline{F}$	$\lvert T-\overline{F}\rvert$	L1(F)	L2(F)	$\overline{P}$	$\lvert T-\overline{P}\rvert$	L1(P)	L2(P)	$\overline{M}$	$\lvert T-\overline{M}\rvert$	L1(M)	L2(M)
Nodes	34.00	34.00	0.00	0.00	0.00	34.00	0.00	0.00	0.00	34.00	0.00	0.00	0.00
Links	78.00	65.63	12.37	371.00	68.61	78.00	**0.00**	**0.00**	**0.00**	78.00	**0.00**	**0.00**	**0.00**
Components	1.00	1.13	**0.13**	**4.00**	**2.00**	1.13	0.13	4.00	2.45	1.20	0.20	6.00	2.45
Network density	0.14	0.12	0.02	0.66	0.12	0.14	**0.00**	**0.00**	**0.00**	0.14	**0.00**	**0.00**	**0.00**
Average degree	4.59	3.86	0.73	21.82	4.04	4.59	**0.00**	**0.00**	**0.00**	4.59	**0.00**	**0.00**	**0.00**
Standard deviation degree	3.88	2.69	1.19	35.61	6.63	3.87	**0.00**	**6.10**	**1.38**	3.67	**0.21**	**7.49**	**1.63**
Global cluster coefficient	0.26	0.16	0.10	2.96	0.56	0.31	**0.05**	**1.63**	**0.33**	0.33	**0.08**	**2.25**	**0.44**
Average cluster coefficient	0.59	0.19	0.40	11.85	2.18	0.68	**0.09**	**2.71**	**0.51**	0.67	**0.08**	**2.49**	**0.48**
Mean path length	2.41	2.85	**0.44**	**13.27**	**4.99**	3.05	0.64	22.98	12.16	3.60	1.19	36.32	14.09
Communities	5.00	7.53	2.53	82.00	18.55	6.80	**1.80**	**58.00**	**12.41**	6.27	**1.27**	**42.00**	**9.70**
Gini coefficient	0.17	0.31	**0.14**	**4.28**	**0.87**	0.33	0.17	5.04	0.96	0.31	0.15	4.37	**0.86**
Average betweenness	23.24	25.26	2.03	**68.62**	**14.83**	22.13	**1.11**	72.03	18.05	23.14	**0.10**	72.18	17.73
Maximum betweenness	231.07	163.33	67.74	2032.19	409.13	234.37	**3.30**	**1271.06**	**277.04**	208.98	**22.09**	**1321.97**	**293.17**
Average closeness	0.01	0.01	0.00	**0.02**	**0.00**	0.01	0.00	0.03	0.01	0.01	**0.00**	0.02	0.01
Minimum closeness	0.01	0.01	**0.00**	**0.02**	**0.01**	0.01	0.00	0.03	0.01	0.01	0.00	0.03	0.01
Average eigencentrality	0.39	0.33	**0.06**	**1.79**	**0.36**	0.30	0.09	2.73	0.51	0.30	0.10	2.84	0.53
Minimum eigencentrality	0.06	0.04	**0.03**	**0.99**	**0.20**	0.03	0.03	1.01	0.20	0.03	0.04	1.07	0.21
Network radius	3.00	3.03	**0.03**	**1.00**	**1.00**	2.60	0.40	12.00	3.46	2.73	0.27	10.00	3.16
Average eccentricity	4.03	4.12	**0.09**	**6.94**	**1.79**	3.70	0.33	14.50	3.18	3.82	0.21	12.88	2.97
Network diameter	5.00	5.13	**0.13**	**8.00**	**3.16**	4.70	0.30	15.00	3.87	4.87	0.13	16.00	4.24

Boldfaced numbers indicate which algorithm performed better for a particular metric

Table 19 Realism results for the Bernard & Killworth Office social network

Metrics	T	$\overline{F}$	$\lvert T-\overline{F}\rvert$	L1(F)	L2(F)	$\overline{P}$	$\lvert T-\overline{P}\rvert$	L1(P)	L2(P)	$\overline{M}$	$\lvert T-\overline{M}\rvert$	L1(M)	L2(M)
Nodes	40.00	40.00	0.00	0.00	0.00	40.00	0.00	0.00	0.00	40.00	0.00	0.00	0.00
Links	238.00	197.60	40.40	1212.00	223.29	238.00	**0.00**	**0.00**	**0.00**	238.00	**0.00**	**0.00**	**0.00**
Components	1.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
Network density	0.31	0.25	0.05	1.55	0.29	0.31	**0.00**	**0.00**	**0.00**	0.31	**0.00**	**0.00**	**0.00**
Average degree	11.90	9.88	2.02	60.60	11.16	11.90	**0.00**	**0.00**	**0.00**	11.90	**0.00**	**0.00**	**0.00**
Standard deviation degree	4.48	3.39	1.09	32.72	6.03	5.11	**0.63**	**19.40**	**3.83**	5.05	**0.57**	**17.59**	**3.61**
Global cluster coefficient	0.41	0.27	0.14	4.08	0.75	0.41	**0.00**	**0.30**	**0.06**	0.42	**0.01**	**0.34**	**0.08**
Average cluster coefficient	0.43	0.28	0.15	4.58	0.84	0.46	**0.03**	**0.96**	**0.20**	0.46	**0.03**	**0.87**	**0.19**
Mean path length	1.76	1.83	0.06	1.93	0.36	1.73	**0.03**	**0.96**	**0.19**	1.74	**0.03**	**0.84**	**0.17**
Communities	4.00	5.50	1.50	55.00	**12.12**	4.80	**0.80**	**42.00**	**10.86**	5.13	**1.13**	**54.00**	13.64
Gini coefficient	0.35	0.36	**0.01**	**1.51**	**0.35**	0.32	0.03	2.42	0.60	0.34	0.01	2.75	0.60
Average betweenness	14.90	16.15	1.25	37.63	7.10	14.28	**0.62**	**18.70**	**3.65**	14.41	**0.49**	**16.45**	**3.26**
Maximum betweenness	46.13	47.39	**1.27**	**188.93**	**42.15**	124.58	78.46	2353.65	456.14	118.82	72.69	2183.96	439.83
Average closeness	0.02	0.01	0.00	0.02	0.00	0.02	**0.00**	**0.01**	**0.00**	0.02	**0.00**	**0.01**	**0.00**
Minimum closeness	0.01	0.01	**0.00**	**0.02**	**0.00**	0.01	0.00	0.04	0.01	0.01	0.00	0.04	0.01
Average eigencentrality	0.58	0.60	**0.02**	**0.87**	**0.20**	0.45	0.13	3.97	0.76	0.45	0.13	3.93	0.76
Minimum eigencentrality	0.12	0.15	0.03	1.17	0.24	0.11	**0.01**	**0.56**	**0.13**	0.10	**0.02**	**0.81**	**0.19**
Network radius	2.00	2.00	0.00	0.00	0.00	2.00	0.00	0.00	0.00	2.00	0.00	0.00	0.00
Average eccentricity	2.83	2.84	**0.02**	**1.50**	**0.39**	2.53	0.29	8.88	1.76	2.59	0.24	7.10	1.48
Network diameter	4.00	3.07	**0.93**	**28.00**	**5.29**	3.03	0.97	29.00	5.39	3.00	1.00	30.00	5.48

Boldfaced numbers indicate which algorithm performed better for a particular metric

Table 20 Realism results for the Krebs Fortune 500 IT Department (Advice) social network

Metrics	T	$\overline{F}$	$\lvert T-\overline{F}\rvert$	L1(F)	L2(F)	$\overline{P}$	$\lvert T-\overline{P}\rvert$	L1(P)	L2(P)	$\overline{M}$	$\lvert T-\overline{M}\rvert$	L1(M)	L2(M)
Nodes	56.00	56.00	0.00	0.00	0.00	56.00	0.00	0.00	0.00	56.00	0.00	0.00	0.00
Links	203.00	182.53	20.47	614.00	113.61	203.00	**0.00**	**0.00**	**0.00**	203.00	**0.00**	**0.00**	**0.00**
Components	2.00	2.00	0.00	0.00	0.00	2.00	0.00	0.00	0.00	2.00	0.00	0.00	0.00
Network density	0.13	0.12	0.01	0.40	0.07	0.13	**0.00**	**0.00**	**0.00**	0.13	**0.00**	**0.00**	**0.00**
Average degree	7.25	6.52	0.73	21.93	4.06	7.25	**0.00**	**0.00**	**0.00**	7.25	**0.00**	**0.00**	**0.00**
Standard deviation degree	4.18	3.45	0.73	22.02	4.12	4.70	**0.52**	**15.57**	**3.04**	4.78	**0.60**	**17.88**	**3.34**
Global cluster coefficient	0.35	0.15	0.20	5.95	1.09	0.28	**0.07**	**2.14**	**0.40**	0.29	**0.06**	**1.81**	**0.34**
Average cluster coefficient	0.42	0.16	0.27	7.99	1.47	0.43	**0.00**	**0.46**	**0.11**	0.43	**0.00**	**0.42**	**0.10**
Mean path length	4.29	4.22	**0.06**	**1.87**	**0.37**	4.15	0.14	4.12	0.77	4.15	0.14	4.20	0.78
Communities	8.00	11.10	3.10	107.00	22.96	9.30	**1.30**	**61.00**	**13.53**	8.67	**0.67**	**68.00**	**15.75**
Gini coefficient	0.43	0.46	**0.03**	**1.63**	**0.34**	0.48	0.05	1.71	0.38	0.45	**0.02**	**1.33**	**0.30**
Average betweenness	36.32	34.61	**1.71**	**51.38**	**10.04**	32.55	3.78	113.30	21.17	32.47	3.85	115.46	21.49
Maximum betweenness	262.14	169.51	**92.63**	**2778.80**	**524.32**	457.03	194.89	5846.64	1154.39	500.81	238.67	7160.13	1355.01
Average closeness	0.01	0.01	**0.00**	**0.01**	**0.00**	0.01	0.00	0.02	0.00	0.01	0.00	0.02	0.00
Minimum closeness	0.01	0.01	0.00	**0.01**	0.00	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.00
Average eigencentrality	0.32	0.40	0.08	2.47	0.48	0.32	**0.01**	**0.75**	**0.17**	0.31	**0.01**	**0.69**	**0.15**
Minimum eigencentrality	0.02	0.05	0.03	0.80	0.17	0.03	**0.01**	**0.28**	**0.06**	0.03	**0.01**	**0.32**	**0.07**
Network radius	3.00	3.00	**0.00**	**0.00**	**0.00**	2.80	0.20	6.00	2.45	2.80	0.20	6.00	2.45
Average eccentricity	3.66	3.56	**0.10**	**4.09**	**0.89**	3.45	0.21	6.93	1.47	3.48	0.19	6.84	1.43
Network diameter	5.00	4.60	**0.40**	**12.00**	**3.46**	4.47	0.53	16.00	4.00	4.40	0.60	18.00	4.24

Boldfaced numbers indicate which algorithm performed better for a particular metric

Table 21 Realism results for the Krebs Fortune 500 IT Department (Business) social network

Metrics	T	$\overline{F}$	$\lvert T-\overline{F}\rvert$	L1(F)	L2(F)	$\overline{P}$	$\lvert T-\overline{P}\rvert$	L1(P)	L2(P)	$\overline{M}$	$\lvert T-\overline{M}\rvert$	L1(M)	L2(M)
Nodes	56.00	56.00	0.00	0.00	0.00	56.00	0.00	0.00	0.00	56.00	0.00	0.00	0.00
Links	387.00	331.40	55.60	1668.00	306.33	387.00	**0.00**	**0.00**	**0.00**	387.00	**0.00**	**0.00**	**0.00**
Components	1.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
Network density	0.25	0.22	0.04	1.08	0.20	0.25	**0.00**	**0.00**	**0.00**	0.25	**0.00**	**0.00**	**0.00**
Average degree	13.82	11.84	1.99	59.57	10.94	13.82	**0.00**	**0.00**	**0.00**	13.82	**0.00**	**0.00**	**0.00**
Standard deviation degree	5.20	3.98	1.21	36.43	6.71	5.79	**0.59**	**17.62**	**3.50**	5.76	**0.56**	**16.88**	**3.30**
Global cluster coefficient	0.49	0.24	0.26	7.69	1.41	0.35	**0.14**	**4.15**	**0.76**	0.36	**0.14**	**4.06**	**0.74**
Average cluster coefficient	0.56	0.24	0.32	9.57	1.75	0.38	**0.19**	**5.55**	**1.02**	0.38	**0.19**	**5.55**	**1.01**
Mean path length	1.90	1.86	**0.04**	**1.21**	**0.23**	1.80	0.10	3.05	0.56	1.79	0.11	3.24	0.59
Communities	3.00	5.03	**2.03**	**67.00**	**16.16**	5.60	2.60	78.00	16.85	4.80	**1.80**	**60.00**	**14.14**
Gini coefficient	0.10	0.29	**0.20**	**6.05**	**1.27**	0.39	0.29	8.71	1.63	0.27	**0.18**	**5.65**	**1.18**
Average betweenness	24.77	23.66	**1.11**	**33.36**	**6.42**	21.98	2.79	83.77	15.37	21.80	2.97	89.00	16.31
Maximum betweenness	116.33	78.13	**38.20**	**1145.89**	**215.07**	195.97	79.65	2401.24	497.39	217.56	101.24	3037.07	594.40
Average closeness	0.01	0.01	**0.00**	**0.01**	**0.00**	0.01	0.00	0.02	0.00	0.01	0.00	0.02	0.00
Minimum closeness	0.01	0.01	0.00	**0.01**	**0.00**	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.00
Average eigencentrality	0.52	0.56	**0.04**	**1.34**	**0.28**	0.40	0.11	3.39	0.65	0.38	0.14	4.11	0.76
Minimum eigencentrality	0.14	0.17	**0.03**	**1.08**	**0.23**	0.09	0.05	1.54	0.30	0.09	0.05	1.44	0.28
Network radius	2.00	2.00	0.00	0.00	0.00	2.00	0.00	0.00	0.00	2.00	0.00	0.00	0.00
Average eccentricity	2.89	2.86	**0.03**	**1.21**	**0.29**	2.75	0.14	4.20	0.86	2.71	0.18	5.41	1.10
Network diameter	3.00	3.03	0.03	1.00	1.00	3.00	**0.00**	**0.00**	**0.00**	3.03	0.03	1.00	1.00

Boldfaced numbers indicate which algorithm performed better for a particular metric

Table 22 Realism results for the Lazega Law Firm social network

Metrics	T	$\overline{F}$	$\lvert T-\overline{F}\rvert$	L1(F)	L2(F)	$\overline{P}$	$\lvert T-\overline{P}\rvert$	L1(P)	L2(P)	$\overline{M}$	$\lvert T-\overline{M}\rvert$	L1(M)	L2(M)
Nodes	71.00	71.00	0.00	0.00	0.00	71.00	0.00	0.00	0.00	71.00	0.00	0.00	0.00
Links	726.0	603.0	123.0	3690.0	675.18	726.0	**0.00**	**0.00**	**0.00**	726.00	**0.00**	**0.00**	**0.00**
Components	1.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
Network density	0.29	0.24	0.05	1.49	0.27	0.29	**0.00**	**0.00**	**0.00**	0.29	**0.00**	**0.00**	**0.00**
Average degree	20.45	16.99	3.47	103.94	19.02	20.45	**0.00**	**0.00**	**0.00**	20.45	**0.00**	**0.00**	**0.00**
Standard deviation degree	8.10	5.88	2.21	66.40	12.16	8.21	**0.12**	**5.76**	**1.41**	8.12	**0.03**	**4.85**	**1.25**
Global cluster coefficient	0.44	0.28	0.16	4.80	0.88	0.41	**0.04**	**1.06**	**0.20**	0.40	**0.04**	**1.10**	**0.20**
Average cluster coefficient	0.45	0.29	0.16	4.93	0.90	0.41	**0.04**	**1.25**	**0.23**	0.41	**0.05**	**1.34**	**0.25**
Mean path length	1.75	1.79	0.04	1.15	0.21	1.73	**0.02**	**0.73**	**0.14**	1.73	**0.02**	**0.65**	**0.12**
Communities	3.00	6.30	3.30	**99.00**	**20.95**	5.03	**2.03**	**65.00**	**16.22**	6.27	**3.27**	102.00	23.41
Gini coefficient	0.11	0.37	**0.26**	**7.74**	**1.49**	0.35	**0.23**	**7.27**	**1.47**	0.40	0.28	8.55	1.67
Average betweenness	26.30	27.64	1.34	40.28	7.45	25.45	**0.85**	**25.37**	**4.73**	25.54	**0.76**	**22.65**	**4.22**
Maximum betweenness	106.69	99.45	**7.25**	**431.28**	**86.78**	200.50	93.80	2814.09	577.96	193.7	87.00	2610.11	519.36
Average closeness	0.01	0.01	0.00	0.01	0.00	0.01	0.00	**0.00**	0.00	0.01	0.00	**0.00**	0.00
Minimum closeness	0.01	0.01	0.00	**0.01**	**0.00**	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.00
Average eigencentrality	0.45	0.54	0.10	2.86	0.54	0.42	**0.03**	**0.94**	**0.20**	0.42	**0.03**	**0.84**	**0.18**
Minimum eigencentrality	0.09	0.15	0.06	1.65	0.33	0.09	**0.00**	**0.29**	**0.07**	0.08	**0.01**	**0.37**	**0.08**
Network radius	2.00	2.00	0.00	0.00	0.00	2.00	0.00	0.00	0.00	2.00	0.00	0.00	0.00
Average eccentricity	2.75	2.66	**0.09**	**2.56**	**0.56**	2.47	0.27	8.18	1.57	2.54	0.21	6.30	1.21
Network diameter	3.00	3.00	0.00	0.00	0.00	3.00	0.00	0.00	0.00	3.00	0.00	0.00	0.00

Boldfaced numbers indicate which algorithm performed better for a particular metric

Abbreviations

CDM: Configuration-degree matching; CM: Configuration model; E: Extraversion; ERGM: Exponential random graph model; F: Feeling; GNAC: Generate network using assignment and compatibility; I: Introversion; IT: Information technology; J: Judging; MBTI: Myers briggs type indicator; N: Intuition; NASA: National aeronautics and space administration; OCEAN: Openness, conscientiousness, extraversion, agreeableness, neuroticism; P: Perceiving; PS: Probability search; ReCon: Replication of complex networks; S: Sensing; SBM: Stochastic block model; T: Thinking

Acknowledgements

The Alabama Supercomputer Authority, which is funded by the State of Alabama, provided a generous grant of supercomputer processing time to support this work. The general research topic, generating synthetic social networks, was brought to our attention by Eric W. Weisel of Old Dominion University. The anonymous reviewers of earlier versions of this article provided insightful and valuable comments that helped to substantially improve the final version. In particular, the Probability Search algorithm is based on an idea provided by one of the anonymous reviewers.

Funding

O'Neil was partially funded by the 2014 RADM Fred Lewis Postgraduate I/ITSEC Scholarship, awarded in association with the Interservice/Industry Training, Simulation and Education Conference and organized by the National Training and Simulation Association. Petty received no specific funding.

Availability of data and materials

All data and program source code described in this article is available to any interested parties. The documentation, source code, input data (the exemplar real-world social networks and compatibility table), as well as the results are available at GitHub at the following URL, https://github.com/daoneil/NetworkMetricSearch, in a directory named GenSynthNetMet.

Authors' contributions

DAO'N identified the network metrics, designed and implemented two of the algorithms (CDM and GNAC), executed the computer runs, and wrote the initial version of this article. MDP created the initial project concept, designed and implemented one of the algorithms (PS), defined the performance comparison methodology, and extensively revised this article. Both authors read and approved the final manuscript.

Authors' information

Daniel A. O'Neil works in the Office of Strategy within the Office of Strategic Analysis and Communication at the National Aeronautics and Space Administration's Marshall Space Flight Center. He develops software prototypes to demonstrate the applications of various technologies to strategic analysis, such as interactive text based scenarios, social network analysis, and 3D orbital trajectory visualization web-apps. Additionally, he integrates Microsoft SharePoint data lists via Nintex workflows. During his career spanning three decades, his employers included the Boeing Company, the U.S. Army Strategic Defense Command, and NASA. His experience includes development of real-time code for the B1-B flight training simulator, management of the development of one of the first web-based intranets, management of the development of a system-of-systems life cycle technology portfolio analysis system, and authorship of tutorials and associates demonstration code for ontology driven orbital dynamics visualization web-apps. He received a B.S. degree in Electrical and Computer Engineering in 1985 and an M.S. degree in Engineering Management in 1997 from the University of Alabama in Huntsville. He is currently a Ph.D. candidate in Modeling and Simulation at the University of Alabama in Huntsville.

Mikel D. Petty is currently Senior Scientist for Modeling and Simulation in the Information Technology and Systems Center and an Associate Professor of Computer Science at the University of Alabama in Huntsville. Prior to joining UAH, he was Chief Scientist at Old Dominion University's Virginia Modeling, Analysis, and Simulation Center and Assistant Director at the University of Central Florida's Institute for Simulation and Training. He received a Ph.D. in Computer Science from the University of Central Florida in 1997. Dr. Petty has worked in modeling and simulation research and education since 1990 in areas that include verification and validation methods, simulation interoperability and composability, and human behavior modeling. He has published over 215 research papers and has been awarded over $16.5 million in research funding. He has served on both National Research Council and National Science Foundation committees on modeling and simulation, is a Certified Modeling and Simulation Professional, and is Editor-in-Chief of the journal *SIMULATION: Transactions of the Society for Modeling and Simulation International*. He has been dissertation advisor to eight graduated Ph.D. students in four different academic disciplines (Computer Science, Modeling and Simulation, Industrial and Systems Engineering, and Computer Engineering).

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 7 July 2018 Accepted: 20 February 2019
Published online: 27 April 2019

References

Abbe E (2017) Community detection and stochastic block models: recent developments

Aiello LM, Barrat A, Cattuto C et al (2012) Link creation and information spreading over social and communication ties in an interest-based online social network. EPJ Data Sci 1(1):12

Alanis-Lobato G, Mier P, Andrade-Navarro M (2016) Manifold learning and maximum likelihood estimation for hyperbolic network embedding. Appl Netw Sci 1:10

Anania EC, Disher T, Anglin KM, Kring JP (2017) Selecting for long-duration space exploration: implications of personality. In 2017 IEEE Aerospace Conference. IEEE, Manhattan Beach, p 1–8

Anderson CJ, Wasserman S, Faust K (1992) Building stochastic blockmodels. Soc Networks 14:137–161. https://doi.org/10.1016/0378-8733(92)90017-2

Back MD (2015) Opening the process black box: mechanisms underlying the social consequences of personality. Eur J Personal 29(91):96. https://doi.org/10.1002/per.1999

Bang-Jensen J, Gutin GZ (2008) Digraphs: theory, algorithms and applications. Springer Science & Business Media, Springer-Verlag, London

Barabási AL (2003) Linked: how everything is connected to everything else and what it means. Basic Books a member of the Perseus Books Group, New York

Barabási AL, Albert R (1999) Emergence of scaling in random networks. Science 286:509–512

Barrenas F, Chavali S, Holme P, Mobini R, Benson M (2009) Network properties of complex human disease genes identified through genome-wide association studies. PLoS One 4(11):e8090

Bender EA, Canfield ER (1978) The asymptotic number of labeled graphs with given degree sequences. J Combinator Theory Ser A 24:296–307. https://doi.org/10.1016/0097-3165(78)90059-6

Bernard HR, Killworth PD, Sailer L (1982) Informant accuracy in social-network data V. An experimental attempt to predict actual communication from recall data. Soc Sci Res 11:30–66. https://doi.org/10.1016/0049-089X(82)90006-0

Bickel PJ, Chen A (2009) A nonparametric view of network models and Newman-Girvan and other modularities. Proc Natl Acad Sci 106:21068–21073. https://doi.org/10.1073/pnas.0907096106

Bollobás B (1980) A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. Eur J Comb 1:311–316. https://doi.org/10.1016/S0195-6698(80)80030-8

Bollobás B (1998) Random graphs. In: Modern graph theory. Springer, New York, pp 215–252

Bonacich P (2007) Some unique properties of eigenvector centrality. Soc Networks 29:555–564. https://doi.org/10.1016/j.socnet.2007.04.002

Borgatti SP, Everett MG, Freeman LC (2014) UCINET. In: Alhajj RRJ (ed) Encyclopedia of social network analysis and mining. Springer, New York

Bouanan Y, Zacharewicz G, Ribault J, Vallespir B (2018) Discrete event system specification-based framework for modeling and simulation of propagation phenomena in social networks: application to the information spreading in a multi-layer social network, SIMULATION: Trans Soc Model Simul Int 1 2018. https://doi.org/10.1177/0037549718776368

Bradley JH, Hebert FJ (1997) The effect of personality type on team performance. J Manag Dev 16:337–353. https://doi.org/10.1108/02621719710174525

Bullington TS (2016) Followers that lead: relating leadership emergence through follower commitment, engagement, and connectedness. Conway, University of Central Arkansas. https://uca.edu/phdleadership/files/2012/07/Bullington-Followers-that-Lead-1.pdf

Capretz LF (2002) Is there an engineering type? World Trans Eng Technol Educ 1:169–172

Catanese SA, De Meo P, Ferrara E et al (2011) Crawling facebook for social network analysis purposes. In: Proceedings of the international conference on web intelligence, mining and semantics. ACM, p 52

Chakrabarti D, Zhan Y, Faloutsos C (2004) R-MAT: a recursive model for graph mining. In: Proceedings of the 2004 SIAM international conference on data mining. Society for Industrial and Applied Mathematics, Society for Industrial and Applied Mathematics, Philadelphia, p 442–446

Chen C (2007) Social networks at Sempra Energy's IT division are key to building strategic capabilities. Glob Bus Organ Excell 26:16–24

Choo PK, Lou ZN, Camburn BA et al (2014) Ideation methods: a first study on measured outcomes with personality type. In: AASME 2014 international design engineering technical conferences and computers and information in engineering conference. American Society of Mechanical Engineers, New York, p V007T07A019–V007T07A019

Chung F, Lu L (2002) The average distances in random graphs with given expected degrees. In: Proceedings of the National Academy of Sciences, 99(25). National Academy of Sciences of the United States of America, pp 15879–15882

Cohen Y, Ornoy H, Keren B (2013) MBTI personality types of project managers and their success: a field survey. Proj Manag J 44:78–87. https://doi.org/10.1002/pmj.21338

Crandall D, Cosley D, Huttenlocher D et al (2008) Feedback effects between similarity and social influence in online communities. In: Proceedings of the 14th ACM SIGKDD international conference on knowledge discovery and data mining. ACM, New York, pp 160–168

Csárdi G, Nepusz T (2013) igraph Reference Manual, http://igraph.org/c/doc/igraph-docs.pdf

Decelle A, Krzakala F, Moore C, Zdeborová L (2011) Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. Phys Rev E 84:066106

DeChurch LA, Mesmer-Magnus JR, Center JS (2015) Maintaining shared mental models over long-duration exploration missions. NASA/TM-2015-218590. NASA, Houston

Easley D, Kleinberg J (2010) Networks, crowds, and markets: reasoning about a highly connected world. New York, Cambridge University Press

Emanuel RC (2013) Do certain personality types have a particular communication style. Int J Soc Sci Humanities 2:4–10

Erdős, P., Rényi, A., On random graphs I. Publicationes Mathematicae Debrecen. Debrecen, Hungary, Institute of Mathematics, University of Debrecen. 6, pp. 290–297

Erdos P, Rényi A (1960) On the evolution of random graphs. Publ Math 5:17–61

Faust K, Wasserman S (1992) Blockmodels: interpretation and evaluation. Soc Networks 14:5–61

Felder RM, Brent R (2005) Understanding student differences. J Eng Educ 94:57–72

Felder RM, Felder GN, Dietz EJ (2002) The effects of personality type on engineering student performance and attitudes. J Eng Educ 91:3–17

Fortunato S (2010) Community detection in graphs. Phys Rep 486:75–174

Frank O, Strauss D (1986) Markov graphs. J Am Stat Assoc 81:832–842

Freeman B (2009) Personality type and medical specialty. University of Chicago Hospital, Chicago

Freeman L (1988) Computer programs and social network analysis. Connections 11:26–31
Freeman L (2016) (2008 September 21). Datasets. Department of Sociology and Institute for mathematical Behaviorial sciences, School of Social Sciences, University of California, Irvine, Retrieved September 9
Freeman LC (1978) Centrality in social networks conceptual clarification. Soc Networks 1:215–239
Furnham A, Crump J (2015a) Personality and management level: traits that differentiate leadership levels. Psychology 6:549
Furnham A, Crump J (2015b) The Myers-Briggs type Indicator (MBTI) and promotion at work. Psychology 6:1510–1515. https://doi.org/10.4236/psych.2015.612147
Gajewar A, Das Sarma A (2012) Multi-skill collaborative teams based on densest subgraphs. In: Proceedings of the 2012 SIAM international conference on data mining. SIAM, Philadelphia, p 165–176
Gersting JL (2014) Mathematical structures for computer science: discrete mathematics and its applications. W. H. Freeman and Company, New York
Geyer CJ, Thompson EA (1992) Constrained Monte Carlo maximum likelihood for dependent data. J R Stat Soc Ser B Methodol 54(3):657-683
Gloor PA, Fischbach K, Fuehres H et al (2011) Towards "honest signals" of creativity – identifying personality characteristics through microscopic social network analysis. Procedia Soc Behav Sci 26:166–179. https://doi.org/10.1016/j.sbspro.2011.10.573
Goldberg LR (1990) An Alternative "Description of personality": The Big-five factor structure. J Pers Soc Psychol 59:1216–1229. https://doi.org/10.1037/0022-3514.59.6.1216
Grandjean M (2016) A social network analysis of twitter: mapping the digital humanities community. Cogent Arts Human 3:14. https://doi.org/10.1080/23311983.2016.1171458
Grant A (2013) Goodbye to MBTI: the fad that won't die. Psychology Today
Holland PW, Laskey KB, Leinhardt S (1983) Stochastic blockmodels: first steps. Soc Networks 5:109–137
Holland PW, Leinhardt S (1977) A method for detecting structure in sociometric data. In Social Networks (pp. 411-432). Academic Press. Retrieved from https://www.elsevier.com/books/social-networks/leinhardt/978-0-12-442450-0
Holland PW, Leinhardt S (1981) An exponential family of probability distributions for directed graphs. J Am Stat Assoc 76:33–50. https://doi.org/10.1080/01621459.1981.10477598
Hunter DR (2007) Curved exponential family models for social networks. Soc Networks 29:216–230
Jafrani S, Zehra N, Zehra M et al (2017) Assessment of personality type and medical specialty choice among medical students from Karachi; using Myers-Briggs type Indicator (MBTI) tool. J Pak Med Assoc 67:520–526
John OP, Srivastava S (1999) The big five trait taxonomy: history, measurement, and theoretical perspectives. In: Handbook of personality: Theory and research, vol 2, pp 102–138
Jung CG (1971) Psychological types. In: Volume 6 of the collected works of CG Jung. Princeton University Press, Princeton, p 169–170
Keirsey D (1998) Please Understand Me II. Prometheus Nemesis Book Company, P.O. Box 2748 Del Mar, California 92014
Kiss M, Kun A, Kapitány A, Erdei P (2014) Regression Analysis of the Effect of Personality-Career Match on the Academic Performance in Business Higher Education: An Evidence from the University of Debrecen (March 22, 2014). Tudás – Tanulás – Szabadság Neveléstudományi Konferencia, Cluj-Napoca, pp 223–227
Knoke D, Yang S (2008) Social network analysis, Second. SAGE Publications, Thousand Oaks
Krackhardt D (1987) Cognitive social structures. Soc Networks 9:109–134
Krebs V (2008) Social capital: the key to success for the 21st century organization. IHRIM J 12:38–42
Kwak H, Lee C, Park H, Moon S (2010) What is Twitte ra social network or a news media? In: Proceedings of the 19th international conference on world wide web. ACM, New York, pp 591–600
Landon LB, Vessey WB, Barrett JD (2015) Risk of performance and behavioral health decrements due to inadequate cooperation coordination, communication, and psychosocial adaptation within a team (JSC-CN-34195). NASA Conf Publ
Lazega E (2001) The collegial phenomenon: the social mechanisms of cooperation among peers in a corporate law partnership. Oxford New York, Oxford University Press, on Demand
Leskovec J, Chakrabarti D, Kleinberg J et al (2010) Kronecker graphs: an approach to modeling networks. J Mach Learn Res 11:985–1042
Li Y, Cao H, Wen G (2018) Simulation study on opinion formation models of heterogenous agents based on game theory and complex networks. SIMULATION 93(11):899–919
Loffredo DA, Opt SK, Harrington R (2008) Communicator style and MBTI extraversion-introversion domains. J Psychol Type 68:29–36
Mahadevan P, Krioukov D, Fall K, Vahdat A (2006) Systematic topology analysis and generation using degree correlations. In: SIGCOMM A (ed) Proceedings of the 2006 conference on applications, technologies, architectures, and protocols for computer communications. ACM, New York, pp 135–146
Malik M, Zamir S (2014) The relationship between Myers Briggs type Indicator (MBTI) and emotional intelligence among university students. J Educ Pract 5:35–42
Manso B, Manso M (2010) Know the network, knit the network: applying SNA to N2C2 maturity model experiments. EDISOFT SA MONTE CAPARICA (PORTUGAL) http://www.dtic.mil/dtic/tr/fulltext/u2/a546862.pdf
Marioles, N. S., Strickert, D. P., & Hammer, A. L. (1996). Attraction, satisfaction, and psychological types of couples. Journal of Psychological Type, 36, 16–27.
McCaulley MH (1977) Application of the Myers-Briggs type indicator to medicine and other health professions. Center for Applications of Psychological Type, Gainesville
McCrae RR, Costa PT (1987) Validation of the five-factor model of personality across instruments and observers. J Pers Soc Psychol 52:81. https://doi.org/10.1037/0022-3514.52.1.81
McCrae RR, Costa PT (1989) Reinterpreting the Myers-Briggs type indicator from the perspective of the five-factor model of personality. J Pers 57:17–40. https://doi.org/10.1111/j.1467-6494.1989.tb00759.x
Metzner R, Burney C, Mahlberg A (1981) Towards a reformulation of the typology of functions. J Anal Psychol 26:33–47. https://doi.org/10.1111/j.1465-5922.1981.00033.x
Milo R, Kashtan N, Itzkovitz S, et al (2003) On the uniform generation of random graphs with prescribed degree sequences. cond-mat/0312028

Mislove A, Marcon M, Gummadi KP et al (2007) Measurement and analysis of online social networks. In: Proceedings of the 7th ACM SIGCOMM conference on internet measurement. ACM, New York, pp 29–42
Mitchell WD (1996) The distribution of MBTI types in the US by gender and ethnic group. J Psychol Type 37:3
Molloy M, Reed B (1995) A critical point for random graphs with a given degree sequence. Random Struct Algoritm 6:161–180. https://doi.org/10.1002/rsa.3240060204
Molloy M, Reed B (1998) The size of the giant component of a random graph with a given degree sequence. Comb Probab Comput 7:295–305
Moutafi J, Furnham A, Crump J (2007) Is managerial level related to personality? Br J Manag 18:272–280. https://doi.org/10.1111/j.1467-8551.2007.00511.x
Myers IB (1962) The Myers-Briggs type indicator: manual. Consulting Psychologists Press, Palo Alto
Myers IB, McCauley MH (1985) Manual: a guide to the development and use of the Myers-Briggs type Indicator. Consulting Psychologists Press, Palo Alto, California
Narayanan A, Shi E, Rubinstein BI (2011) Link prediction by de-anonymization: how we won the kaggle social network challenge. In: The 2011 international joint conference on neural networks conference proceedings. IEEE Computational intelligence society, Piscataway, pp 1825–1834
Narayanan A, Shmatikov V (2008) Robust de-anonymization of large sparse datasets. In: 2008 IEEE symposium on security and privacy. IEEE Computer Society, Los Alamitos, pp 111–125
Narayanan A, Shmatikov V (2009) De-anonymizing social networks. In: 2009 30th IEEE symposium on security and privacy. IEEE computer society conference publishing services, Los Alamitos, pp 173–187
Nelson J, Bolton J (2008) Systems engineering behavior and leadership study. Johnson Space Center, National Aeronautics and Space Administration, Houston
Newman M (2010) Networks: an introduction. Oxford University Press, New York
Newman ME (2003) The structure and function of complex networks. SIAM Rev 45:167–256. https://doi.org/10.1137/S003614450342480
Newman ME, Girvan M (2004) Finding and evaluating community structure in networks. Phys Rev E 69:026113. https://doi.org/10.1103/PhysRevE.69.026113
Newman ME, Strogatz SH, Watts DJ (2001) Random graphs with arbitrary degree distributions and their applications. Phys Rev E 64. https://doi.org/10.1103/PhysRevE.64.026118
Nowicki K, Snijders TAB (2001) Estimation and prediction for stochastic blockstructures. J Am Stat Assoc 96:1077–1087. https://doi.org/10.1198/016214501753208735
Papadopoulos F, Kitsak M, Serrano MÁ et al (2012) Popularity versus similarity in growing networks. Nature 489:537
Pattison P, Wasserman S, Robins G, Kanfer AM (2000) Statistical evaluation of algebraic constraints for social networks. J Math Psychol 44:536–568. https://doi.org/10.1006/jmps.1999.1261
R Core Team (2016) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna URL https://www.R-project.org/
Rapoport A (1957) Contribution to the theory of random and biased nets. Bull Math Biophys 19:257–277. https://doi.org/10.1007/BF02478417
Robins G, Pattison P, Kalish Y, Lusher D (2007) An introduction to exponential random graph (p*) models for social networks. Soc Networks 29:173–191. https://doi.org/10.1016/j.socnet.2006.08.002
Roethlisberger FJ, Dickson WJ (1939) Management and the worker. Harvard University Press, Cambridge
Rosati P (1993) Student retention from first-year engineering related to personality type. In: Frontiers in education conference, 1993. Twenty-third annual conference. "Engineering education: renewing America's technology", proceedings. IEEE, Piscataway, pp 37–39
Rushton JP, Irwing P (2008) A general factor of personality (GFP) from two meta-analyses of the big five: Digman (1997) and mount, Barrick, Scullen, and rounds (2005). Personal Individ Differ 45:679–683. https://doi.org/10.1016/j.paid.2008.07.015
Sampson S (1969) Crisis in a cloister. Unpublished doctoral dissertation. Cornell University. https://www.uni-due.de/hummell/netzwerkbuch/ucinet/prog/UCI%20IV-%20Einzeldateien/uci4_dat.pdf
Schwimmer E (1973) Exchange in the social structure of the Orokaiva: traditional and emergent ideologies in the Northern District of Papua. London, Hurst and Co
Schwimmer E (1979) Reciprocity and structure: a semiotic analysis of some Orokaiva exchange data. Man 14:271–285. https://doi.org/10.2307/2801567
Scott J (2000) Social network analysis: a handbook, second. SAGE publications, Inc, Thousand Oaks
Scott J, Carrington PJ (2011) The SAGE handbook of social network analysis. SAGE publications, Inc, Thousand Oaks
Seshadhri C, Kolda TG, Pinar A (2012) Community structure and scale-free collections of Erdős-Rényi graphs. Physical Rev E 85. https://doi.org/10.1103/PhysRevE.85.056109
Smathers (2003) (Guide to the Isabel Briggs Myers Papers 1885–1992). University of Florida George A. Smathers Libraries, Department of Special and Area Studies Collections, Gainesville, FL. 2003. http://web.uflib.ufl.edu/spec/manuscript/guides/Myers.htm Retrieved February 28
Snijders TA (2002) Markov chain Monte Carlo estimation of exponential random graph models. J Soc Struct 3:1–40
Staudt CL, Hamann M, Gutfraind A et al (2017) Generating realistic scaled complex networks. Appl Netw Sci 2:36. https://doi.org/10.1007/s41109-017-0054-z
Strogatz SH (2001) Exploring complex networks. Nature 410:268–276. https://doi.org/10.1038/35065725
Thurman B (1979) In the office: networks and coalitions. Soc Networks 2:47–63. https://doi.org/10.1016/0378-8733(79)90010-8
Tsvetovat M, Carley K (2005) Generation of realistic social network datasets for testing of analysis and simulation tools. Carnegie Mellon University. Available at SSRN 2729296, Elsevier, Amsterdam
Tupes EC, Christal RE (1992) Recurrent personality factors based on trait ratings. J Pers 60:225–251. https://doi.org/10.1111/j.1467-6494.1992.tb00973.x
van Mierlo T, Hyatt D, Ching AT (2016) Employing the Gini coefficient to measure participation inequality in treatment-focused digital health social networks. Netw Model Anal Health Inform Bioinform 5:32
Viger F, Latapy M (2005) Efficient and simple generation of random simple connected graphs with prescribed degree sequence. In: International computing and combinatorics conference. Springer, Berlin Heidelberg, pp 440–449

Wasserman S, Pattison P (1996) Logit models and logistic regressions for social networks: I an introduction to Markov graphs and p*. Psychometrika 61:401–425

Watts DJ, Strogatz SH (1998) Collective dynamics of "small-world" networks. Nature 393:440. https://doi.org/10.1038/30918

Webster CM (1993) Task-related and context-based constraints in observed and reported relational data. PhD Thesis. University of California, Irvine

Weiler DT (2017) The effect of role assignment and personality subtypes in simulation on critical thinking development, situation awareness, and perceived self-efficacy of nursing baccalaureate students. Master's Thesis. University of Louisville

Yang J, Leskovec J (2015) Defining and evaluating network communities based on ground-truth. Knowl Inf Syst 42:181–213. https://doi.org/10.1007/s10115-013-0693-z

Zachary WW (1977) An information flow model for conflict and fission in small groups. J Anthropol Res 33:452–473. https://doi.org/10.1086/jar.33.4.3629752Zhou B, Pei J, Luk W (2008) A brief survey on anonymization techniques for privacy preserving publishing of social network data. SIGKDD explorations 10:12–22. https://doi.org/10.1145/1540276.1540279