Soc. Netw. Anal. Min. (2015) 5:29
DOI 10.1007/s13278-015-0270-4

ORIGINAL ARTICLE

Social ties, homophily and extraversion–introversion to generate complex networks

Faraz Zaidi[1,2] · **Muhammad Qasim Pasta**[2,3] · **Arnaud Sallaberry**[4] · **Guy Melançon**[5,6]

Received: 21 September 2014 / Revised: 14 April 2015 / Accepted: 8 June 2015 / Published online: 27 June 2015

Abstract Many interconnected systems and particularly social interactions can be modeled as networks. These networks often exhibit common properties such as high clustering coefficient, low average path lengths and degree distributions following power-law. Networks having these properties are called small world-scale free networks or simply complex networks. Recent interest in complex networks has catalysed the development of algorithmic models to artificially generate these networks. Often these algorithms introduce network properties in the model regardless of their social interpretation resulting in networks which are statistically similar but structurally different from real world networks. In this paper, we focus on social networks and apply concepts of social ties, homophily and extraversion-introversion to develop a model for social networks with small world and scale free properties. We claim that the proposed model produces networks which are structurally similar to real world social networks.

Keywords Social networks · Small world networks · Scale free networks · Network generation models

✉ Faraz Zaidi
farazzaidi@yahoo.com; faraz@pafkiet.edu.pk

Muhammad Qasim Pasta
mqpasta@pafkiet.edu.pk

Arnaud Sallaberry
arnaud.sallaberry@lirmm.fr

Guy Melançon
guy.melancon@labri.fr

[1] Levich Institute and Physics Department, City College of New York, New York, NY, USA

[2] Karachi Institute of Economics and Technology, Karachi, Pakistan

[3] Usman Institute of Technology, Karachi, Pakistan

[4] LIRMM, Université Paul Valéry Montpellier, Montpellier, France

[5] CNRS UMR 5800 LaBRI, Talence, France

[6] INRIA Bordeaux – Sud-Ouest, Talence, France

1 Introduction

Many interconnected real world systems can be modelled as networks including social networks. Social networks can be defined as a set of people, or groups of people interacting with each other (Scott 2000; Wasserman and Faust 1994). Graphically, these networks can be represented by using a set of *nodes* and *edges*, where *nodes* represent people and *edges* represent their interactions.

Many researchers have studied different structural properties of social networks. Two of these properties that gained considerable importance (Sebastian and Schnettler 2009) are Small World (Watts and Strogatz 1998) and Scale free (Barabási and Albert 1999) properties. A small world network is defined by two structural metrics, small average geodesic distance and high clustering coefficient. A scale free network (Barabási and Albert 1999) is defined has having a degree distribution following power law i.e. a few nodes have a very high number of connections (degree) and lots of nodes are connected to a few nodes only.

A number of models have been proposed to artificially generate networks with both these properties (Dorogovtsev and Mendes 2002; Holme and Kim 2002; Klemm and Eguiluz 2002) (see Sect. 3 for more references). These models do not incorporate domain specific structural properties for social networks. As a result, when compared to real world social networks, the networks generated using

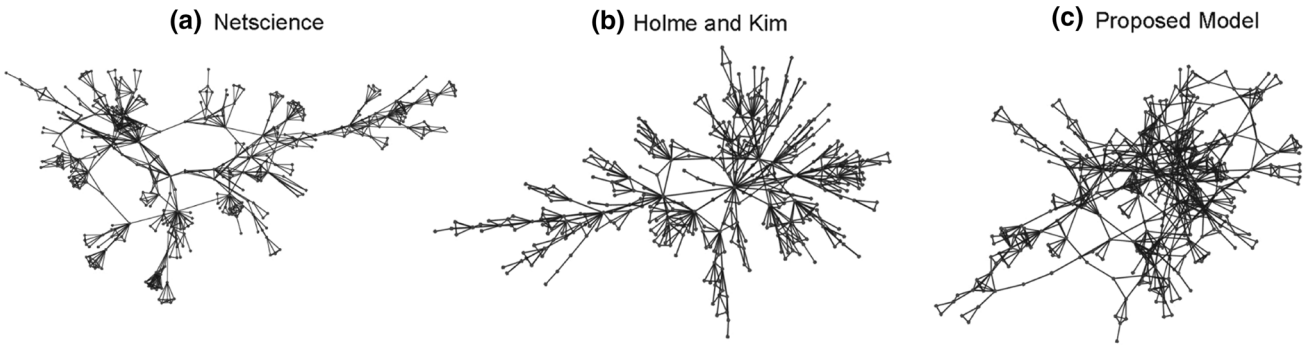


Fig. 1 Comparing Netscience co-author network with (Holme and Kim 2002) and the Proposed Model. Netscience **a** and Proposed network model **c** both have cliques of size greater than 3 whereas only triads are present in **b**

existing models are structurally different. For example, Fig. 1a shows a co-authorship network[1] and equivalent size networks generated using (Holme and Kim 2002) (Fig. 1b) and the proposed model (Fig. 1c). A basic structural motif in social networks is the presence of cliques of varying sizes in real networks. Figure 1b shows the equivalent size network generated using the model of (Holme and Kim 2002) which only has triads and cliques of bigger sizes are completely absent. Another important feature missing in artificially generated networks is the connectivity patterns of how different individuals and groups of individuals connect to each other to form a bigger society. We review a number of social traits from sociology to outline different structural characteristics a model should accommodate in order to generate networks which are structurally similar to real world networks. We extensively review different models and highlight their differences with real world networks in Sect. 3.

The study of network generation models is very useful as they can be used to construct networks with desired structural properties that mimic real world networks (Krivitsky et al. 2009). These networks can then serve as benchmarks and test beds to facilitate various experimental and empirical studies. These models are also useful for simulation studies to examine different network processes such as epidemic spreading, influence mining (Hussain et al. 2013) and formation of community structures (Badham and Stocker 2010). Researchers have also used these models to test various network sampling methods (Kurant et al. 2011) as these models can generate networks with different sizes and varying structural properties. Furthermore, these models provide us with insight about the underlying structures present in a network facilitating the task of understanding and analyzing these networks (Scott 2011).

In this paper, we propose a new model to generate social networks with small world and scale free properties where we incorporate a number of structural elements derived from the concepts of social ties, homophily and extraversion-introversion. We discuss how small social groups interact to form large social networks (Simmel and Wolff 1950), how an individual's intrinsic preferences and characteristics (Rapoport 1957) result in forming new acquaintances and how the extroverts are able to interconnect small social groups in our society. Incorporating these important elements in the proposed network generation model, we show that the generated networks are indeed small world and scale free as well as structurally similar to real world social networks.

The rest of the paper is organized as follows: In Sect. 2, we discuss a number of different social properties and argue that combining these concepts, we can understand the characteristics required for a real world social network. In Sect. 3, we discuss a number of different models to generate small world and scale free networks. We then present the proposed network generation model in Sect. 4 and provide experimental results in Sect. 5. Finally we conclude and give possible future directions of our research in Sect. 7.

2 Structure of social networks

In this section, we discuss a number of concepts from the domain of sociology in an attempt to better understand how social networks in the real world are structured.

Social ties People in the real world are linked to each other through social ties (Wasserman and Faust 1994). The simplest form of a tie is *Dyad* (Simmel and Wolff 1950) where two people are linked to each other. This is considered as the unit of studying relationships in a social network. *Triads* are relationships between three people and

[1] NetScience Network is a co-authorship network of scientists working on network theory and experiments, as compiled by Newman in May (2006). The biggest connected component is considered for visualization here which contains 379 nodes and 914 edges.

have been the focus of many social network studies (Wasserman and Faust 1994). *Groups* of larger size are also possible but since a variety of relationships can form in them, they are less stable (Simmel and Wolff 1950) and often less studied in sociology. They are often identified by their dense connectivity and clear bounds forming a cluster.

Due to dense interconnectivity, these ties are termed as *strong ties* (Krackhardt 1992) where nodes that are loosely connected to each other are said to have *weak ties* (Granovetter 1973). Each of us in the society has these weak ties along with strong relationships. These weak ties or acquaintances are important for developing new relationships and possibly joining new social communities or clusters. There is a fine mix of both these weak and strong ties that exist in our society and both should be considered to develop a model to generate artificial social networks.

Homophily An important human characteristic is *homophily*, tendency of actors or entities to associate with other actors or entities of similar type (Rapoport 1957; Rapoport and Horvath 1961). Homophily helps to explain why you know the people that you do, because you all have something in common. A model based on these ideas was proposed by Rapoport who called it *Random Biased Nets*. The idea was to modify the traditional random model of networks such that it incorporates social behaviors. Rapoport also concluded that we occasionally do things that are derived entirely from our intrinsic preferences and characteristics, and these actions may lead us to meet new people who have no connections to our previous friends at all. Although these actions might appear to be random, but can be justified as having strong social background with logical explanations. We limit our study to address this characteristic and refer it as random connectivity pattern. In the light of homophily and social dynamics, we can conclude that new connections between people are formed based on two properties, random connectivity and homophily.

Extraversion–introversion It is interesting to note that in our society, we come across people that are well known and famous, and then there are people who have very few friends and contacts. These ideas are the direct implication of the human trait of extraversion–introversion (Jung 1921). Extroverts, who are open to meeting new people and developing new relationships are expected to have high degree of connectivity in a social network as compared to Introverts, who tend to be more reserved, less outgoing, and less sociable.

An important use of this human characteristic is to explain the scale free degree behavior of social networks. A famous person is likely to become more famous as compared to a person who is not well known in the social community. Termed as the principle of *Preferential Attachment* (Barabási and Albert 1999), it explains the growth behavior of networks with power law degree distribution. The idea is that nodes having high degree, have a high probability of attracting more new connections. Thus a model to generate a social network must take this property into consideration as well.

Observations and inferences In our society, we not only form relations with individuals, but with groups of people (called social groups) as well. These relations are defined by particular circumstances, interests or some context like our school, work place, family (Rapoport and Horvath 1961; Granovetter 1973) and can be explained by homophily. These groups are densely connected to each other often forming a clique where every individual in the social group is connected to each other. Our society is built using these social groups or cliques and we can call them 'Building Blocks' of our society.

Each of these 'building block' or 'social group' (Simmel and Wolff 1950) is like a small cluster joined to each other by people belonging to more than one group (Watts 2003; Burt 2005). When these small clusters have many connections to each other, they form bigger size clusters. Bigger size clusters or small groups when connected loosely to each other, form a social network (Simmel and Wolff 1950). The size of social groups or small clusters in a network, vary to a large extent, and so does the number of clusters. Both these parameters depend largely on how the individuals and their ties evolve in a society.

Addressing the principle of Preferential Attachment, for every node in a social group, extroverts have higher connectivity with other people. For example, in a group representing the actors playing in the same movie, the famous actors who have acted in many movies will have many connections, and the actors who are starting their career, or are not so well known will have only a few connections.

Finally, we look at the society on the whole where we consider the average path length of the networks. Low average path length can be realized by random connectivity of nodes, where Watts and Strogatz (1998) used this method to have low average path lengths in small world networks. Kasturirangan argues that low average path lengths are not due to the random connectivity of nodes, but due to multiple scales in a network. These scales are formed due to the presence of high degree nodes which are responsible for reducing the overall average path length of a network Kasturirangan (1999).

Combining all these principles, we can conclude that the important elements to capture in the structure of a social network are:

1. Social networks consists of many small groups that are densely connected within themselves forming cliques. This represents the connectivity within a social group.
2. These groups overlap due to individuals having multiple affiliations. Some groups have many overlaps

which creates large size communities or clusters. These connections represent how small social groups connect to form a social network.

3. A certain degree of randomness exists where we occasionally do things that are derived entirely from our intrinsic preferences and characteristics. These actions lead us to meet new people who have no connections to our previous friends at all. This represents that nodes not only connect preferentially, but due to some extent, randomly also.
4. The random connectivity pattern and the presence of high degree nodes creating multiple scales is responsible for the low distances between any two people on average.
5. Every small group of people has a few Extroverts and many Introverts, where introverts are only connected to a few people, usually in their social group and extroverts are responsible for interconnecting people from different social groups and the society at large.

We incorporate all these principles in the proposed model. We discuss the details of the proposed model in Sect. 4.

3 Related work

We first describe the two ground breaking network models to generate small world and scale free networks. We then discuss network models proposed to generate small world-scale free networks.

The small world model (Watts and Strogatz 1998) starts with a ring of n vertices in which each vertex is connected to its k nearest neighbors, for a given k. Then, each edge is rewired with a given probability p by choosing randomly a new vertex to connect. Since the neighbors are connected to each other in a regular graph, the overall clustering coefficient is very high. On the other hand, the average path length is very low as vertices are only connected to their neighbors. Randomly rewiring a few edges connects nodes lying at long distances, which reduces the overall average path length. Most of the nodes remain connected to their neighbors, resulting in high clustering coefficient whereas the average path length is reduced, giving us the properties of a small world network. It is important to note that these networks do not have scale free degree distribution. Since every vertex in the network initially has a fix k degree, random rewiring of only a few vertices does not effect the overall behavior of the degree distribution. More formal studies of this model have been conducted with interesting results (Dorogovtsev and Mendes 2000).

The scale free model (Barabási and Albert 1999) explains how the scale free degree distribution emerges in real world networks. To begin, a disconnected graph of n vertices is created. At every time step t, a new vertex v with m edges is added to the network. These edges are connected to existing vertices with the probability proportional to the degree of the nodes in the network. This preferential bias in the connectivity is termed as preferential attachment as new nodes prefer to attach to high degree nodes. Mathematical results for scale free graphs have been studied by several researchers such as (Bollobás and Riordan 2002; Bollt and ben Avraham 2008).

Most of the early research tried to unify these two models to generate networks with both small world and scale free networks. For example, Holme and Kim (2002) modified the well known Barabasi and Albert model (Barabási and Albert 1999) to obtain graphs that are small world as well as scale free. A Triad formation step is added where every preferentially added node is also probabilistically connected to a randomly selected neighbor of the node it preferentially chose. This results in the creation of triads in the network increasing the overall clustering coefficient. A drawback of this model is that it does not generate cliques of larger size, since by construction, the algorithm only enforces triads. Figure 1 shows three network drawn using Tulip software (Auber 2010). The first network (Fig. 1a) is the well known co-authorship network of researchers working on network theory (Newman 2006). Figure 1b shows a network drawn using (Holme and Kim 2002) with approximately the same number of nodes and edges. Figure 1c shows the network drawn using the proposed model. The absence of larger size cliques in Fig. 1b can be easily noticed.

The idea of introducing triads used by Holme and Kim (2002) is similar to the model proposed by Dorogovtsev et al. (2002) where every new node added to the network is connected to both ends of a randomly chosen link where one of the nodes of this link is selected through preferential attachment. Similar behavior is obtained in terms of connectivity as lots of triads are created and the absence of large size cliques remains a drawback. Preferential attachment principle is used to represent the extroverts in the network as explained in Sect. 2.

These models inspired Jian-Guo et al. (2005) to introduce another similar model. The network starts with a triangle and a new node with two edges is added to the network in each iteration. The other two ends of these two edges connect to two different nodes say n1 and n2 which are already connected to each other forming a triad. Nodes n1 and n2 are both selected on the basis of preferential attachment. This differs from the previous two models where the second node is preferentially chosen. Structurally, the network thus generated still miss bigger size cliques. Since the new node is connected to two nodes preferentially, the random connectivity pattern also gets ignored as described in Sect. 2.

Fu and Liao (2006) proposed another extension to Barabási and Albert (1999) which they called the

Relatively Preferential Attachment method. In each step, the newly introduced node in the network connects to a node w with preferential attachment, the nodes in the immediate neighborhood of w have higher probability of connecting to this new node as compared to other nodes. The newly added nodes can have m edges which differs from the previsouly discussed models. The value of m is chosen as an initial parameter which remains constant throughout the execution of the algorithm. As a result, dense groups appear in the network which are not necessarily cliques. For values of m greater than 2, lots of triads appear in the network increasing the overall clustering coefficient but structurally dense social groups with weak ties do not appear in the network as described in Sect. 2.

Klemm and Eguiluz (2002) proposed a model where each node of the network is assigned a state variable. A newly added node is in *active* state and keeps attaching links until eventually deactivated. At each time step, a new node is added to the network by attaching a link to each of the z active nodes. The new node is set as *active*. One of the existing nodes is deactivated where the probability of a node being deactivated is inversely proportional to its degree. To reduce the average path length of the entire graph, at every step, for each link of the newly added node, it is decided uniformly at randomly whether the link connects to the active node or it connects to a randomly selected node. Again the model does not impose any other constraints to form cliques and the random connectivity pattern is induced probabilistically.

Catanzaro et al. (2004) present a model incorporating assortativity (Newman 2002) in generated networks. New nodes are added to the network based on preferential attachment and a new edge is added between previously existing nodes chosen on the basis of their degree. This introduces links between similar degree nodes forcing assortative mixing behavior of social networks. The model is innovative as it allows addition of new links between previously existing nodes. Triads are created by links added between similar degree nodes but they do not follow the connectivity pattern described in Sect. 2 where nodes from different social groups overlap to connect small social groups.

Newman et al. (2002) study models of the structure of social networks with arbitrary degree distributions also including networks with degree distributions following power-law. They use the idea to generate affiliation networks similar to co-authorship networks using random bipartite graphs. Guillaume and Latapy (2006) also propose a model based on bipartite networks. The authors use two disjoint sets called *bottom* and *top*. At each step, a new *top* node is added and its degree d is sampled from a prescribed distribution. For each of the d edges of the new vertex, either a new *bottom* vertex is added or one is picked among the pre-existing ones using preferential attachment. The bipartite graph is then projected as a unipartite graph to obtain a small world and scale free network. Figure 2a shows such a bipartite random graph and Fig. 2b shows its unipartite projection. These models produce very dense networks as the unipartite versions are projections of n-partite networks. Furthermore, random connectivity results as a by-product of the initially generated random bipartite networks rather than as a social trait.

Wang and Rong (2008) proposed a slightly different model, which is still a modified form of the preferential attachment model. Instead of adding one node in each step, rings of n nodes are added in each step. Two of these newly added nodes connected to the existing network using preferential attachment. Since n remains fixed through out the process, uniform size cliques can be found in the generated networks. Furthermore these cliques do not overlap as discussed in Sect. 2 but are connected to other cliques through newly added links differing the structure of the generated network.

Another class of models has been proposed where connectivity among nodes is determined based on social and demographic attributes (Boguñá et al. 2004; Wong et al. 2006; Badham and Stocker 2010; de Almeida et al. 2013; Pasta et al. 2014) rather than structural metrics. These models use the notions of social spaces and nodal attributes to determine the similarity among nodes and create links. For example Pasta et al. (2014) proposes a model to generate networks similar to facebook datasets using nodal and structural attributes at the same time. Nodal attributes include such as age, gender, class year, major and residence and structural attributes include node degree and friend-of-a-friend.

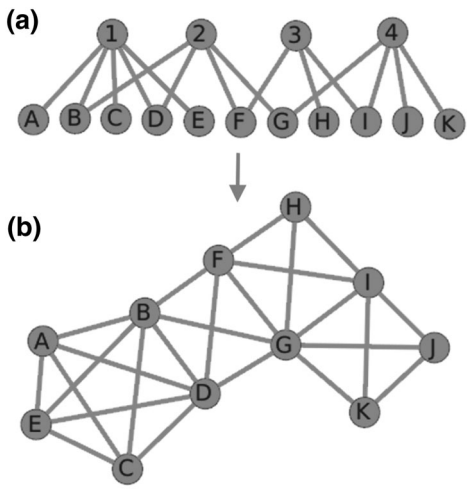


Fig. 2 **a** Randomly drawn bipartite graph, **b** unipartite Projection of graph **a**

Evolutionary network models with aging nodes have also been proposed in the literature such as (Dorogovtsev and Mendes 2000; Zhu et al. 2003; Geng and Wang 2009; Wen et al. 2011). For example Wen et al. (2011), study the dynamic behavior of local-world evolving networks with aging nodes. Newly added nodes connect to previously existing nodes based on strength-age preferential attachment. Networks thus generated exhibit power-law degree distribution, high clustering coefficient and small world properties.

Other models have also been proposed based on the local-world phenomena (Pan et al. 2006; Sun et al. 2007; Wang et al. 2009; Wen et al. 2011) where nodes only consider information from there neighbourhood to determine formation of new links in contrast to earlier discussed models that assume the presence of global information. For example, Wang et al. (2009) investigate a local preferential attachment model to generate hierarchical networks with degree distribution following power-law.

Another class of network generation models address the issue of generating graphs with community structures (Condon and Karp 1999; Virtanen 2003; Lancichinetti and Fortunato 2009; Moriano and Finke 2013; Zaidi 2013; Sallaberry et al. 2013; Pasta et al. 2013). These models generate networks with well defined communities. Since in this paper, we do not try to generate networks having community structures, we do not discuss these papers further.

Another class of graphs models, the exponential random graph models (ERGMs) have gained a lot of popularity (Frank and Strauss 1986; Snijders et al. 2006; Robins et al. 2007). These models are used to test, to what extent nodal attributes and structural dependencies describe structure of a network measured using frequency of degree distribution, traids and geodesic distances (Toivonen et al. 2009). The possible ties among individuals are modelled as random variables, and assumptions about dependencies among these random tie variables determine the general form of the exponential random graph model for the network (Robins et al. 2007). An important difference between network generation models and ERGMs is that network models try to explain how a network is generated, whereas ERGMs do not explicitly explain any network generation process (Toivonen et al. 2009).

We have cited a large number of network generation models above but the related work is by no means exhaustive. We have omitted several citations simply because either they do not generate small world-scale free networks, or because they target some specific structural property other than clustering coefficient, short geodesic distance and degree distribution following power-law. Surveys, reports and comparative analysis for different network generation models can be found in Dorogovtsev and Mendes (2002), Newman (2003), Jackson (2005), Fortunato (2010), Badham and Stocker (2010), Toivonen et al. (2009), Goldenberg et al. (2010), Pasta et al. (2014).

4 Proposed model

There are three basic steps in the model which are discussed in detail in the sections below. In the first step, we introduce what we call building blocks in the network. Our society is composed of many small social groups, which can be represented by cliques of various sizes in a network. This is different from various models described earlier, where one node or fixed number of nodes are added at a time to the network. Adding cliques to represent social groups of the society introduces densely connected nodes resulting in high clustering coefficient of the entire network. These groups act as the building blocks of our society as described earlier in Sect. 2.

The next step is to determine how to join these disconnected cliques to form a society, a social network (Simmel and Wolff 1950). From the property of Extraversion-Introversion, we know that there are people with many social contacts as well as people with only a few contacts. People with many social contacts belong to multiple social groups. This idea leads us to define the number of groups every individual belongs to. Nodes representing extroverts belongs to many different cliques, whereas intorverts only belong to only one or a few cliques. This number is drawn from a scale free degree distribution ensuring that the final node degrees follow the power law degree distribution.

In the final step, we simply merge two nodes from different groups into a single node as shown in Fig. 4. As a result, two cliques are combined with a single node being part of the two cliques representing the extroverts of the society that belong to multiple social groups as explained in Sect. 2.

As the network is built from cliques and the connections are directed by scale free degree distribution, we get a network with high clustering coefficient and degree distribution following power law. The average path length of the overall network remains low due to two connectivity patterns, the random connections and multiple scales in the degree distribution as explained in Sect. 2 .

We explain the details of the proposed algorithm below. The following mathematical notations are used throughout the explanation: $G(V, E)$ represents an undirected multigraph where V is a set of n nodes and E is a set of e edges. The graph G is initially empty and the nodes and edges are added as the algorithm progresses. $\mathcal{C}$ represents a set of cliques such that $\mathcal{C} = \{C_1, C_2, \ldots, C_k\}$ are different cliques each comprising of several nodes.

4.1 Step 1: building blocks

We start by adding cliques of variable sizes to an empty graph G. This is different from the existing network generation models where a single node or a triad is added in a single step. Recall from Sect. 2, we identified cliques as one of the fundamental patterns present in networks. These cliques represent the small social groups of our society as described in Sect. 2.

The algorithm takes as parameter, the number of cliques to be generated (k), the minimum (min) and the maximum size (max) of the cliques to be generated. A uniform random distribution is used to determine the size of each clique C_i to be added to the graph G such that nodes and edges of the clique become members of V and E respectively. G becomes a graph comprising of $\mathcal{C} = \{C_1, C_2, \ldots, C_k\}$ as shown in Fig. 3.

If we use a random number generator, for large values of k, the distribution will be uniformly spread and we will have the same number of cliques for all possible size values. In real networks, this might not be the case as often, cliques of large sizes are rare compared to cliques of small sizes. To take the correct decision, it is important to understand what type of network we are trying to generate. If the network to be generated is expected to have cliques of varying sizes uniformly distributed, the random generation will serve well our purpose. On the other hand, if we expect that all the cliques will have the exact same size, the *min* and *max* parameters can be set to that exact value to have all the cliques of the exact same size. And in the case where we expect a non-uniform distribution of different sizes, we can draw the different sizes of cliques using the type of distribution we require our final network to follow.

We consider the example of co-authorship network and explain how these values effect the algorithm. We use $k = 10$, min $= 1$ and max $= 5$ and a random generation for the size of the cliques. After the execution of this step, we get a network as shown in Fig. 3. The idea of introducing cliques, comes from the work of (Newman et al. 2002; Guillaume and Latapy 2006) where affiliation networks and the bipartite structure has been identified as an important structural property of the way, the Author network is constructed in the real world. As explained in Sect. 3, projecting a bipartite graph as a unipartite graph creates cliques of different sizes. This is a better representation of the structure of the society rather than to introduce triads as most of the models do to achieve high clustering coefficient. This phenomena was explained in detail in Sect. 2.

Note that the size of the cliques can be forced to be exactly 3, in which case we would have forced the presence of only triads just as the other network generation models presented in Sect. 3. Due to the presence of cliques (or triads), the average clustering coefficient of the entire graph increases as compared to a random graph which is a fundamental property to identify a small world network.

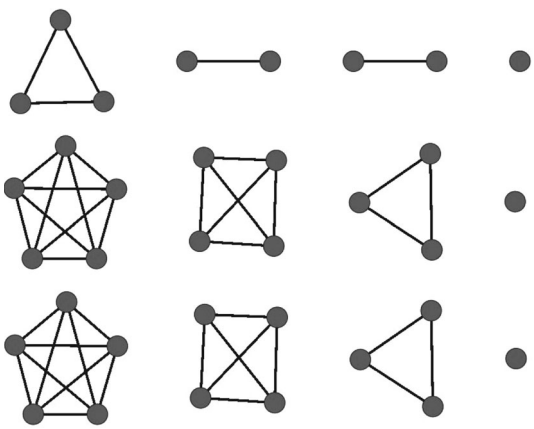


Fig. 3 Step 1: Network after execution of step 1 with min $= 1$, max $= 5$ and $k = 10$

4.2 Step 2: determine number of merges

Since our goal is to enforce a scale free degree distribution on the generated network G, we generate a power law degree distribution using a power law function. We associate this distribution on the nodes of graph G as an attribute and call this as open connections OC. These OC values determine how social an individual is, as based on these values two nodes are merged into a single node as shown in Fig. 4. Recall from Sect. 2, extroverts are responsible for connecting small social groups into a large connected society. This attribute helps us determine the extroverts and introverts of a social group which eventually helps us to connect the initially added cliques to G into a single connected network.

Note that this number of merges of OC values are directly proportional to the final node degree. If a few nodes are merged with many nodes, these nodes will end up with many connections and thus the scale free degree distribution will appear in the network.

An important variation to this step can be the assignment of an equal value to all nodes. As a result, the network produced will have only small world properties,

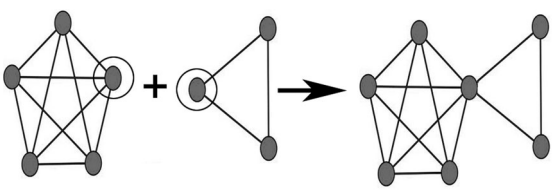


Fig. 4 Merging two nodes from two different cliques so that a node becomes part of two cliques

having high clustering coefficient and small average path length. The equal value assignment will ensure that the degree of all the nodes is approximately equal and thus the final degree distribution will not follow a power law, rather a uniform distribution.

4.3 Step 3: merge nodes

Finally, based on the number of merges assigned in the previous step, we merge two nodes to build a connected network. In case, where two nodes of different building blocks are selected and that are already connected to each other by some other node, multiple overlaps appear. This results in two small groups densely connected to each other and sparsely connected to nodes from other groups.

Two cliques can be combined by considering that one or more than one individual belongs to two different cliques, and these nodes play the role of combining two cliques (see Fig. 4) representing extroverts of a social group as discussed earlier.

Merging two nodes creates connections between previously disconnected cliques. Moreover, the merged node plays the role of a bridge between the two small clusters. In terms of the degree, the node gets many new connections. Higher the number of merges for a node, the more it gets connections and higher would be its node degree. This is the reason why we draw the number of merges from a power law function, as a result, the final degree distribution follows a power law.

An important decision while merging two nodes say n_1 and n_2 with OC values oc_1 and oc_2 is, how to decide the oc_n for the new node n. We experimented with the following different methods:

- Max: Assign the new node the maximum of the two OC values $oc_n = Max(oc_1, oc_2)$
- Min: Assign the new node the minimum of the two OC values $oc_n = Min(oc_1, oc_2)$
- Avg: Assign the new node the average of the two OC values $oc_n = Avg(oc_1, oc_2)$
- Rand: Assign the new node one of the two OC values randomly $oc_n = Rand(oc_1, oc_2)$

Assigning maximum value forces the degree distribution of the network to take a more linear decay as most of the low degree nodes disappear quickly from the network and lots of high degree nodes are left for connectivity. On the other hand, assigning minimum value removes the few nodes with very high degree and the characteristic long tail in the degree distribution disappears from the network. A similar behavior is observed with the average assignment as the long tail disappears and the average node degree increases with this assignment. The best results are obtained by a random assignment as nodes with high and low degree are equally removed and thus the overall degree distribution follows scale free behavior. We show the experimental results using the random method in Sect. 5.

5 Experimental results and discussion

The first set of experiments shows the behavior of the proposed model to generate networks with small world and scale free properties, i.e. high clustering coefficient, small geodesic distance and degree distribution following power-law. The model takes as input three parameters, the number of cliques C_k to be generated, the minimum *min* and maximum *max* sizes for the cliques to be generated. We ran simulations with $C_k =$ 2000,4000,6000,8000,10,000, min $= 1$ and max $= 9, 11, 13$ giving us 15 possibilities and the results are averaged over 5 runs for each of these settings. Figures 5, 6 and 7 shows the clustering coefficients,

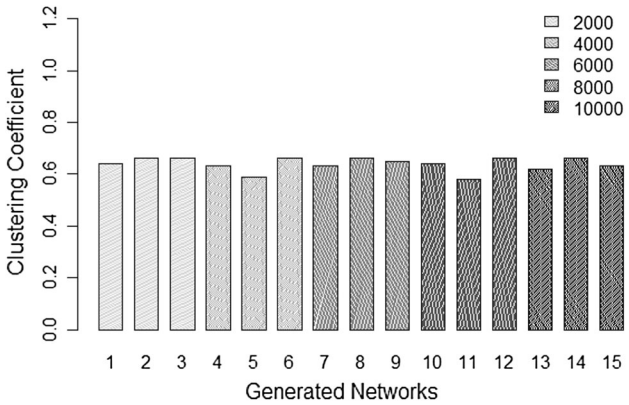


Fig. 5 Averaged clustering coefficients for the generated networks using the proposed model

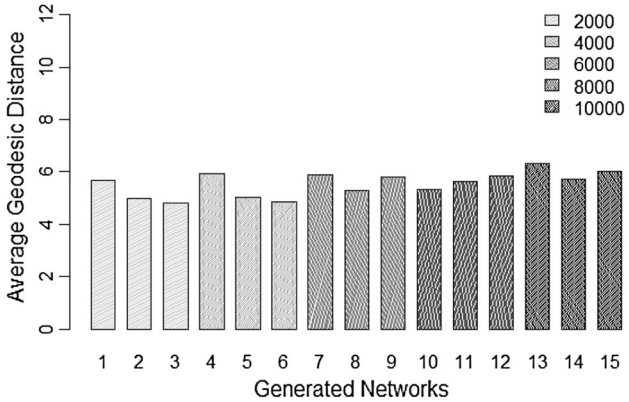


Fig. 6 Averaged geodesic distances for the generated networks using the proposed model

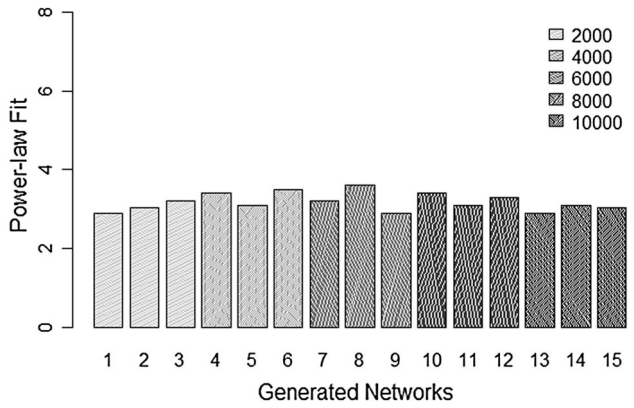


Fig. 7 Averaged power-law fit demonstrating that the degree distributions of the generated networks follows power-law

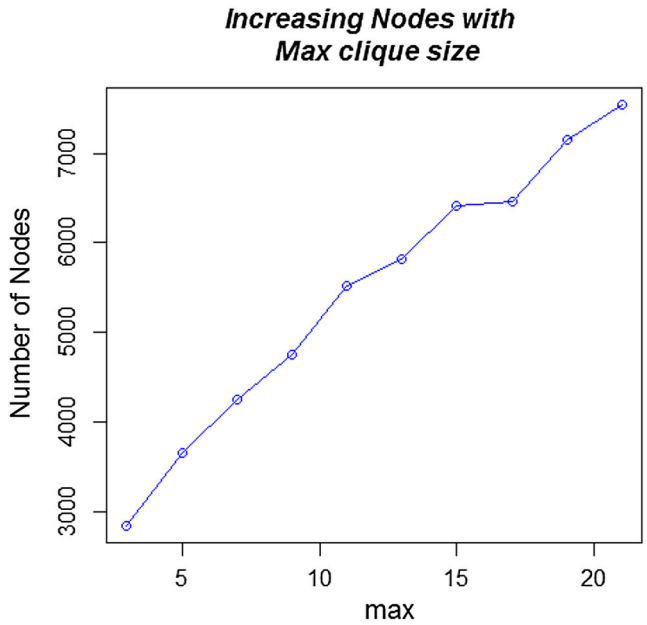
geodesic distances and power-law fit for the degree distributions of the generated networks.

Figure 5 shows that for all the generated networks, clustering coefficient remains around 0.6 which is comparably very high as compared to an equivalent random network. Geodesic distances (Fig. 6) range between 4 and 7 for all the generated networks and power-law fitting constants (Fig. 7) hover around 3.0 indicating a scale free degree distribution. Thus all the generated networks have small world and scale free networks.

The second set of experiments demonstrate the behavior of increasing the max parameter while keeping the other parameters constant. We used 10 values with max $= 3, 5, 7, 9, 11, 13, 15, 17, 19, 21$ keeping the other two parameters as $C_k = 4000$ and min $= 1$. Figure 8 (Left) clearly shows the linear relationship between increasing *max* values and the number of nodes in the generated network. Figure 8 (Right) shows the clustering coefficient, average geodesic distance and power-law fitting constant for these generated graphs. High clustering coefficients, low average geodesic distances and power-law fitting constants between 2 and 4 clearly show that the generated networks have small world and scale free properties.

6 Comparison with other models and real networks

Figure 1 shows the comparison of a real world co-authorship network to an existing and the newly proposed model. One of the major difference between the real world model and the existing models in general is the idea of the building blocks used in the proposed model. These building blocks or variable sized cliques to mimic the real world phenomena of densely connected individuals, which in turn results in high clustering coefficients for the generated



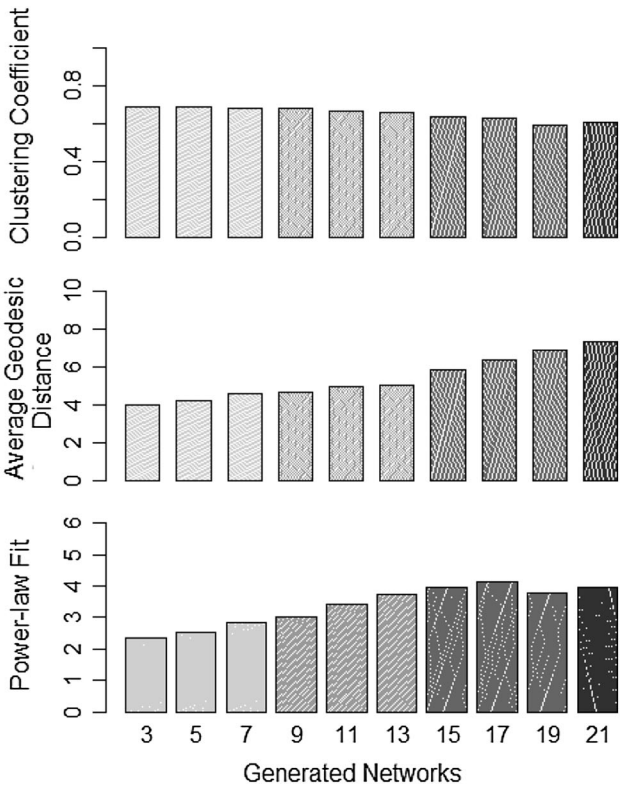


Fig. 8 *Left* Number of nodes plotted against increasing values of parameter *max*. *Right* Clustering coefficient, average geodesic distance and power-law fit corresponding to the generated graphs are shown

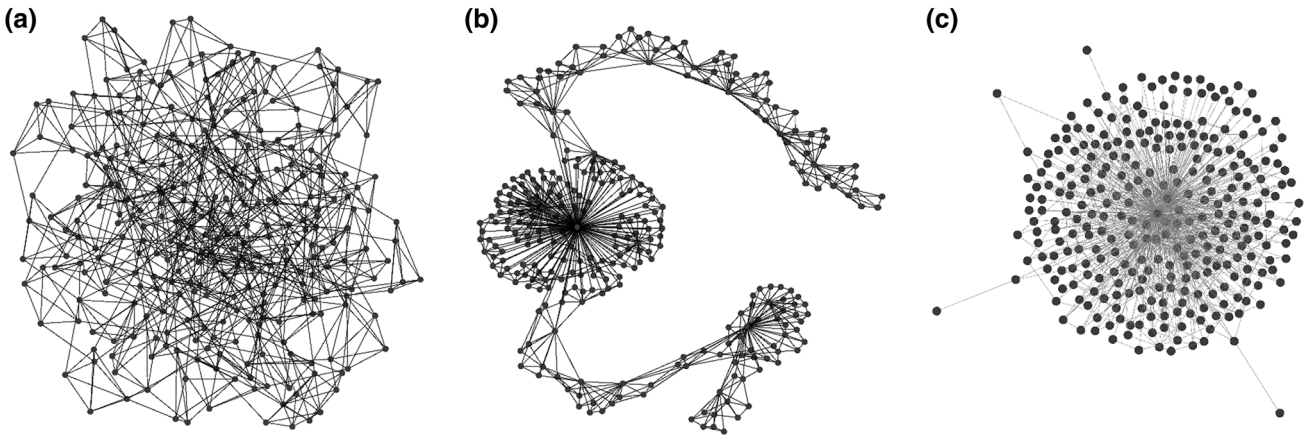


Fig. 9 Networks generated equivalent to NetScience co-author network using different models. **a** Wang and Rong Model mostly contains cliques of size 5. **b** Klemm and Eguiluz model generates a network with a long chain like structure and enforces triads where larger size cliques are missing. **c** Catanzaro et al. generates a network with very low clustering coefficient when compared to the real network

networks. Most of the existing network models (Holme and Kim 2002; Dorogovtsev and Mendes 2002; Liu et al. 2005; Fu and Liao 2006) try to force the presence of triads in a network to increase clustering coefficients which deos not necessarily mimic the social structure of real world networks. The network generated using the proposed model not only looks structurally similar to the real co-author network, but also has approximately the same clustering coefficient, geodesic distance and the degree distribution follows power law.

Figure 9 shows networks generated using Wang and Rong model (2008) (Fig. 9a), Klemm and Eguiluz model (Klemm and Eguiluz 2002) (Fig. 9b) and Catanzaro et al. (2004) (Fig. 9c). Wang and Rong model uses fixed size cliques to grow the evolving network as a result of which, nodes with varying connectivity are rare and sparse. Klemm and Equiluz modifies the preferential attachment model by forcing the formation of triads to increase the networks's clustering coefficient but does not introduce cliques or densely connected individuals in the network which differs largely from a real network as shown in Fig. 9b. The model of Catanzaro et al. preferentially connects nodes based on their degree which results in assortative networks but does not introduce triads or cliques in the network as a result of which, the generated networks have low clustering coefficients as absence of triads and cliques can be seen in Fig. 9c.

Table 1 shows the comparison of three different real networks with approximately equivalent size networks generated using the proposed model. All the generated networks exhibit small world and scale free properties with short geodesic distances, high clustering coefficients and the power-law coefficients around 3.0. It is important to note that since the proposed model uses minimal parameters, it is not able to generate networks that are exactly equivalent to real networks with respect to different network metrics.

The three networks used for comparative analysis are the NetScience co-author network, Condensed Matter co-author network and IMDB Actor network. For the NetScience network (Newman 2006), only the biggest connected component was considered containing 379 nodes. The Condensed Matter (Newman 2001) network containing 15,876 nodes and a small subset of the IMDB actor

Table 1 Comparing different real world networks with generated networks using the proposed model

Network	Nodes	Edges	Geodesic distance	Clustering coefficient	Power-law coefficient
NetScience co-author	303	873	5.11	0.65	3.55
Proposed model	379	914	6.04	0.74	3.35
Condensed matter Co-author	15,876	42,416	5.38	0.61	3.81
Proposed model	16,439	43,419	6.62	0.53	3.57
IMDB actor	7640	277,029	2.94	0.87	5.71
Proposed model	7367	288,764	5.64	0.67	3.74

network containing 7640 nodes were the other network used.

7 Conclusion and future research

In this paper, we have studied the concepts of social ties, homophily, extraversion-introversion as important properties for the structure of social networks. We use these concepts to present a model to generate complex networks with small world and scale free properties. We discussed a number of network generation models that successfully generated small world and scale free networks but produced structurally different networks as compared to real world networks. Results show that the proposed model indeed generates networks that are structurally similar to real world networks as compared to the other existing models.

We intend to extend our study to other types of networks such as biological networks (Cannataro et al. 2010), online social networks (Lewis et al. 2008), transportation networks (Ducruet and Zaidi 2012; Guimera et al. 2005) and human communication patterns (Bourqui et al. 2008). Although these networks also have small world and scale free properties but they are again structurally different from social networks and thus we need to modify the proposed model to mimic the behavior of these other types of networks.

References

Auber D (2003) Tulip—a huge graph visualization framework. In: Mutzel P, Jünger M (eds) Graph drawing software, mathematics and visualization series. Springer, Berlin

Badham J, Stocker R (2010) A spatial approach to network generation for three properties: degree distribution, clustering coefficient and degree assortativity. J Artif Soc Soc Simul 13(1):11

Barabási AL, Albert R (1999) Emergence of scaling in random networks. Science 286(5439):509–512

Boguñá M, Pastor-Satorras R, Díaz-Guilera A, Arenas A (2004) Models of social networks based on social distance attachment. Phys Rev E 70(5):056122

Bollobás B, Riordan OM (2002) Mathematical results on scale-free random graphs. In: Bornholdt S, Schuster HG (eds) Handbook of Graphs and Networks: From the Genome to the Internet. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, FRG. doi:10.1002/3527602755.ch1

Bollt EM, ben Avraham D (2004) What is special about diffusion on scale-free nets? New J Phys 7:26

Bourqui R, Zaidi F, Gilbert F, Sharan U, Simonetto P (2008) Vast 2008 challenge: social network dynamics using cell phone call patterns. In: IEEE symposium on visual analytics science and technology, 2008

Burt RS (2005) Brokerage and closure. Oxford University Press, Oxford

Cannataro M, Guzzi PH, Veltri P (2010) Protein-to-protein interactions: technologies, databases, and algorithms. ACM Comput Surv 43:1:1–1:36

Catanzaro M, Caldarelli G, Pietronero L (2004) Assortative model for social networks. Phys Rev E (Statistical, Nonlinear, and Soft Matter Physics) 70(3):1–4

Condon A, Karp RM (1999) Algorithms for graph partitioning on the planted partition model. Random Struct Algorithms 18(2):116–140

de Almeida ML, Mendes GA, Viswanathan GM, da Silva LR (2013) Scale-free homophilic network. Europ Phys J B 86(2):1–6

Dorogovtsev S, Mendes J (2000) Exactly solvable small-world network. Europ Phys Lett 50(1):1–7

Dorogovtsev SN, Mendes JFF (2000) Evolution of networks with aging of sites. Phys Rev E 62(2):1842–1845

Dorogovtsev SN, Mendes JFF (2002) Evolution of networks. Adv Phys 51:1079–1187

Ducruet C, Zaidi F (2012) Maritime constellations: a complex network approach to shipping and ports. Maritime Policy Manag 39(2):151–168

Fortunato S (2010) Community detection in graphs. Phys Rep 486(3):75–174

Frank O, Strauss D (1986) Markov graphs. J Am Stat Assoc 81(395):832–842

Fu P, Liao, K (2006) An evolving scale-free network with large clustering coefficient. In ICARCV IEEE, pp 1–4

Geng X, Wang Y (2009) Degree correlations in citation networks model with aging. Europhys Lett 88(3):38002

Goldenberg A, Zheng AX, Fienberg SE, Airoldi EM (2010) A survey of statistical network models. Found Trends Mach Learn 2(2):129–233

Granovetter M (1973) The strength of weak ties. Am J Sociol 78(6):1360–1380

Guillaume J-L, Latapy M (2006) Bipartite graphs as models of complex networks. Phys A Stat Mech Appl 371(2):795–813

Guimera R, Mossa S, Turtschi A, Amaral LAN (2005) The worldwide air transportation network: anomalous centrality, community structure, and cities' global roles. Proc Nat Acad Sci USA 102(22):7794–7799

Holme P, Kim BJ (2002) Growing scale-free networks with tunable clustering. Phys Rev E 65:026107

Hussain OA, Anwar Z, Saleem S, Zaidi F (2013) Empirical analysis of seed selection criterion in influence mining for different classes of networks. In: Cloud and green computing (CGC), 2013 third international conference on IEEE, pp 348–353

Jackson MO (2005) A survey of network formation models: stability and efficiency. Cambridge University Press, Cambridge

Jung CJ (1921) Psychologischen typen, volume Translation H.G. Baynes, 1923. Rascher Verlag, Zurich

Kasturirangan R (1999) Multiple scales in small-world networks. In Brain and Cognitive Science Department, MIT

Klemm K, Eguiluz VM (2002) Growing scale-free networks with small world behavior. Phys Rev E 65:057102

Krackhardt D (1992) The strength of strong ties: the importance of philos in networks and organization. In: Nohria N, Eccles RG (eds) Networks and organizations. Harvard Business School Press, Boston

Krivitsky PN, Handcock MS, Raftery AE, Hoff PD (2009) Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models. Soc Netw 31(3):204–213

Kurant M, Gjoka M, Butts CT, Markopoulou (2011) Walking on a graph with a magnifying glass: stratified sampling via weighted random walks. In: Proceedings of the ACM SIGMETRICS joint international conference on measurement and modeling of computer systems, pp 281–292. ACM

Lancichinetti A, Fortunato S (2009) Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities. Phys Rev E 80(1):016118

Lewis K, Kaufman J, Gonzalez M, Wimmer A, Christakis N (2008) Tastes, ties, and time: a new social network dataset using facebook.com. Soc Netw 30(4):330–342

Liu J-G, Dang Y-Z, Wang Z (2005) Multistage random growing small-world networks with power-law degree distribution. Chin Phys Lett 3(3):746

Moriano P, Finke J (2013) On the formation of structure in growing networks. arXiv preprint arXiv:1301.4192

Newman MEJ (2001) Scientific collaboration networks. I. Network construction and fundamental results. Phys Rev E 64(1):016131. doi:10.1103/PhysRevE.64.016131

Newman MEJ (2002) Assortative mixing in networks. Phys Rev Lett, pp 89–20

Newman MEJ (2006) Finding community structure in networks using the eigenvectors of matrices. Phys Rev E 74(3):036104

Newman MEJ, Watts DJ, Strogatz SH (2002) Random graph models of social networks. Proc Natl Acad Sci USA 99(Suppl 1):2566–2572

Newman MEJ (2003) The structure and function of complex networks. SIAM Rev 45:167

Pan Z, Li X, Wang X (2006) Generalized local-world models for weighted networks. Phys Rev E 73(5):056109

Pasta MQ, Jan Z, Sallaberry A, Zaidi F (2013) Tunable and growing network generation model with community structures. In: social computing and applications, 2013 third international conference on, pp 233–240

Pasta MQ, Zaidi F, Rozenblat C (2014) Generating online social networks based on socio-demographic attributes. J Complex Netw 2(4):475–494

Rapoport A (1957) Contribution to the theory of random and biased nets. Bull Math Biophys 19:257–277

Rapoport A, Horvath WJ (1961) A study of a large sociogram. Behav Sci 6(4):279–291

Robins G, Pattison P, Kalish Y, Lusher D (2007) An introduction to exponential random graph (p) models for social networks. Soc Net 29(2):173–191

Sallaberry A, Zaidi F, Melançon G (2013) Model for generating artificial social networks having community structures with small-world and scale-free properties. Soc Netw Anal Min 3:597–609

Schnettler S (2009) A structured overview of 50 years of small-world research. Soc Netw 31(3):165–178

Scott JP (2000) Social network analysis: a handbook. SAGE Publications, London

Scott J (2011) Social network analysis: developments, advances, and prospects. Soc Net Anal Min 1:21–26

Simmel G, Wolff KH (1950) The sociology of Georg Simmel / translated and edited with an introduction by Kurt H. Wolff. Free Press, Glencoe Ill

Snijders TA, Pattison PE, Robins GL, Handcock MS (2006) New specifications for exponential random graphs models. Sociolog Methodol 36(1):99–153

Sun X, Feng E, Li J (2007) From unweighted to weighted networks with local information. Phys A Stat Mech Appl 385(1):370–378

Toivonen R, Kovanen L, Kivel M, Onnela J-P, Saramki J, Kaski K (2009) A comparative study of social network models: network evolution models and nodal attribute models. Soc Netw 31(4):240–254

Virtanen S (2003) Properties of nonuniform random graph models. Research Report A77, Helsinki University of Technology, Laboratory for Theoretical Computer Science, Espoo, Finland

Wang L-N, Guo J-L, Yang H-X, Zhou T (2009) Local preferential attachment model for hierarchical networks. Phys A Stat Mech Appl 388(8):1713–1720

Wang J, Rong L (2008) Evolving small-world networks based on the modified ba model. Computer science and information technology, international conference on, pp 143–146

Wasserman S, Faust K (1994) Social network analysis: methods and applications. Cambridge University Press, Cambridge

Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world' networks. Nature 393:440–442

Wattsc DJ (2003) Six degrees: the science of a connected age, 1st edn. W. W. Norton & Company, New York

Wen G, Duan Z, Chen G, Geng X (2011) A weighted local-world evolving network model with aging nodes. Phys A Stat Mech Appl 390(21):4012–4026

Wong LH, Pattison P, Robins G (2006) A spatial model for social networks. Phys A Stat Mech Appl 360(1):99–120

Zaidi F (2013) Small world networks and clustered small world networks with random connectivity. Soc Netw Anal Min 3(1):51–63

Zhu H, Wang X, Zhu J-Y (2003) Effect of aging on network structure. Phys Rev E 68(5):056121