Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2019) 27: 765 – 779

doi:10.3906/elk-1806-103

Measurement of network-based and random meetings in social networks

Pranav NERURKAR*, **Madhav CHANDANE**, **Sunil BHIRUD**
Department of Computer Engineering and Information Technology, Veermata Jijabai Technological Institute,
University of Mumbai, Mumbai, India

Received: 13.06.2018 • **Accepted/Published Online:** 03.11.2018 • **Final Version:** 22.03.2019

Abstract: Social networks are created by the underlying behavior of the actors involved in them. Each actor has interactions with other actors in the network and these interactions decide whether a social relationship should develop between them. Such interactions may occur due to meeting processes such as chance-based meetings or network-based (choice) meetings. Depending upon which of these two types of interactions plays a greater role in creation of links, a social network shall evolve accordingly. This evolution shall result in the social network obtaining a suitable structure and certain unique features. The aim of this work is to determine the relative ratio of the meeting processes that exist between different actors in a social network and their importance in understanding the procedure of network formation. This is achieved by selecting a suitable network genesis model. For this purpose, different models for network genesis are discussed in detail and their differences are highlighted through experimental results. Network genesis models are compared and contrasted with other approaches available in the literature, such as simulation-based models and block models. Performance measures to compare the results of the network genesis models with baselines are statistics of networks recreated using the models. The socially generated networks studied here belong to various domains like e-commerce, electoral processes, social networking websites, peer to peer file-sharing websites, and Internet graphs. The insights obtained after analyzing these datasets by network genesis models are used for prescribing measures that could ensure continuous growth of these social networks and improve the benefits for the actors involved in them.

Key words: Generative models, random graph models, network structure, social network analysis

1. Introduction

Social networks have become ubiquitous as they have been used to represent many systems present in nature, such as the Internet, transportation systems [1], multilevel biological networks [2, 3], epidemiology networks [4, 5], social networks [6, 7], coauthorship networks [8], collective behavior [9], and political networks [10]. To model a particular system as a social network, entities of the system are shown as nodes and relationships between these entities are denoted as edges. Representing a system in the form of a social network has advantages as the established network science literature can be applied for its analysis. Empirical studies have established that the meeting processes of actors cause a network to develop certain structural features [11–14]. Behaviors of actors in the social network are a result of interactions they might have with other actors in the network. These interactions may be due to chance-based meetings with strangers or meetings that occur as a result of choice, i.e. through a network-based search process. Numerous investigations have observed that networks of different domains may develop different structures as they evolve, such as the presence or absence of central

*Correspondence: panerurkar_p16@ce.vjti.ac.in

hubs, random structures, or communities [12, 15, 16]. An intuitive reason for this could be the different meeting processes of the actors in them, which leads to the creation of such structures.

Meeting processes in a social network can be categorized broadly into two types of interactions: chance-based and choice-based. In a network genesis model, nodes may decide to create links with the other nodes in the social network that they may have met by chance. In this inquiry, such interactions are referred to as 'random' meetings. Other types of interactions are the ones where the node searches the neighborhood of the nodes to which it is already linked to identify suitable candidates for link formation. Such interactions are referred to in this work as network-based or neighborhood search-based interactions. Such network-based meetings occur as node i conducts a local search in the neighborhood of nodes that it is connected to, for the purpose of identifying a suitable node with which it wants an association. Actors may derive benefits from both chance-based and network-based meetings. The proportion of random meetings to network-based meeting varies across social networks of different domains. Understanding the interrelationship between choice and chance in a social network is important as it may provide insight into the formation of social networks and answer questions like why certain social networks exhibit certain structures.

Coevolution theory is a line of social and economic network literature that focuses on understanding the influence of meeting processes (micromechanisms) on the growth of a social network (macrooutcome) and vice versa. A number of publications in this field have examined the link between such micromechanisms on the macrooutcomes in social networks by conducting simulation studies [17–22]. However, in real social networks it is difficult to pinpoint the role of a single micromechanism such as meeting processes on the observed macrooutcome (growth of the social network) as a multitude of mechanisms interact with one another.

A second strategy to capture the effects of meeting processes on the growth of a social network is the use of social network generative (genesis) models. These are stochastic algorithms used to generate social networks or graphs. The parameters of such techniques are tuned using random parameter search or regression [11, 17]. Each generative model captures stylized facts seen in real-world networks, such as the preferential attachment model, small world model, random graph model, forest fire model, or Jackson–Rogers model [14]. The current inquiry focuses on these generative models to understand network formation processes behind social networks of various domains ranging from e-commerce to electoral processes. Different models for network genesis are discussed in detail and a unified framework for these techniques is described. Their differences are also highlighted through experimental results. A suitable model among these frameworks is identified and utilized for the task of determining the relative ratio of the meeting processes that exist between different nodes in a social network. This is the primary aim of this inquiry. A secondary objective is to use the insights obtained after analyzing socially generated datasets using network genesis models for prescribing measures that may be useful to ensure continuous growth of these social networks and improve the benefits for the actors involved in them.

Following this introduction, Section 2 provides the literature review. Section 3 highlights the results of the experimental work and the insights obtained from them. Section 4 gives the concluding remarks.

1.1. Preliminaries - network formation models

1.1.1. Barabasi–Albert - Preferential attachment

The intuition behind this model is that when new vertices enter a social network they prefer to attach to already well-connected vertices over less-connected ones. Initially in the BA model a single vertex is present that has no edges. Then, at each time step, a new vertex is created. This newly created vertex then initiates edges to existing vertices. The probability that vertex i is chosen is given by Eq. (1):

$$P[i] = k[i]^{\alpha} + a \tag{1}$$

- α = power of the preferential attachment
- a = minimum incoming edges a node should get

Thus, the probability $P(k)$ that the vertex links with k other vertices decays as a power law. The graph generated using this model has power law distribution of degrees. However, the drawback of this stochastic model is that it assumes a linear model for preferential attachment and does not consider the possibility of a nonlinear preferential attachment model [11].

1.1.2. Erdos–-Rényi random graph

These graphs are of two types, $G(n,p)$ and $G(n,m)$. $G(n,p)$ has n vertices and the probability of an edge between them is constant p. $G(n,m)$ has n vertices and m edges such that m edges are chosen uniformly at random from a set of all possible edges [11].

1.1.3. Preferential attachment and aging

This is a discrete time step model of a growing random graph. At each time step a single vertex is added and it initiates links to vertices already existing in the social network. The probability of a node k getting a newly created edge is given by $P[k]$ in Eq. (2) [11]. This model thus enriches the Barabasi–Albert model.

$$P[k] = (c * k[i]^{\alpha} + a) * (d * l[i]^{\beta} + b) \tag{2}$$

- c, d = coefficient of degree and age
- $k[i]$, $l[i]$ = in-degree and age of vertex i
- a, b = attractiveness of vertices with no adjacent edge and zero age
- α, β = preferential attachment exponent, aging exponent

1.1.4. Watts–Strogatz Model

A generative model that creates a lattice structured graph. Each node is connected to all nodes within its neighborhood. The lattice structure thus formed is rewired, i.e. edges are selected at random with a probability p and connected to nodes outside their immediate neighborhood. This is done without altering the number of nodes or edges. The rewiring procedure, creates a "small world" effect, i.e. reduction in the average path length of the graph [11].

1.1.5. J-R model

Jackson *et al.* proposed a social network generative model where nodes of the social network are allowed to form links to other nodes using a hybrid strategy that encapsulates elements of preferential attachment model and the Erdos–-Rényi model. Thus, if there are preexisting m nodes in a network then a newborn node links to $a * m$ of them chosen uniformly at random and $(1-a) * m$ using a neighborhood search strategy (choice-based links) and attaches to them. The hyperparameter a is ratio of chance-based interactions to choice-based interactions.

2. Review of the literature

Zou et al. analyzed the effect of meeting processes on a microblog social network by combining sociological theories with network science [16]. Leduc et al. observed the effect of random meetings between individuals on product referrals by word of mouth [23]. To understand the pitfalls in R&D networks between firms and academic institutes, Tomasello et al. modeled these in the form of social networks. The authors argued that the position of a firm in the network provided it with an opportunity to form strategic alliances with academic institutes. This shaped decisions about the formation of R&D alliances [24]. Ciliberto et al. empirically analyzed the effect of competition between airlines on the socially generated network of airline transportation [25]. Maggio et al. analyzed the network of relationships between brokers and institutional investors in a stock market to find how a nexus of brokers and institutional investors obtained higher returns from stock markets [26]. Thus, in the literature, several investigations have been made to understand the contribution of meeting processes in the development of alliances or relationships between nodes in a social network. However, understanding the weightage (relative ratio) of meeting processes was not within the scope of such studies.

Snijders et al. proposed a stochastic actor-oriented model (SAOM) [19] in which the network formation is believed to occur as a consequence of individual actors' actions with other actors in the social network [18]. However, in SAOMs it is difficult to measure the role of a single micromechanism such as 'chance/choice' on the formation of the social network. A second family of techniques known as network representation learning or network embedding frameworks encode network structure into low-dimensional embeddings. The literature consists of several such NRL frameworks based on matrix factorization [27–35], the word2vec (skip-gram) model of Mikolov et al. [36–45], deep convolutional neural networks [46–49], the random walk and neural network unified framework [50], hyperbolic space embedding techniques [8, 51], latent-space models [52–57], and multidimensionality reduction [58, 59]. Although these approaches capture network features such as first-order or second-order proximity but are not designed specifically to capture the interplay between meeting processes in a social network.

An alternative approach is to use network generative (genesis) models to understand the role of meeting processes in network formation. To identify a suitable generative model for a network, it must capture several characteristics exhibited by real-world social networks. Socially generated networks tend to have the average distance between a pair of nodes on the order of the log of the number of nodes. The geodesic of such networks is also on the order of the log of the number of nodes. Clustering coefficients in real networks are larger than in networks where links are generated by an independent random process but less than in networks where links are generated by a preferential attachment pattern. The clustering among neighbors of nodes is inversely related to the degree of the node [11, 17]. In socially generated networks, actors (nodes) are born over time and connect to preexisting nodes. This leads to a positive assortativity in the network as nodes prefer attaching to other nodes that are at least as old as they are.

Not all of these properties can be explained using a single network formation model. Golosovsky et al. argued that although a preferential attachment model was accepted as the most plausible generative mechanism of growing complex networks, some networks exhibit preferential attachment only for nodes with low and moderate degrees while the nodes with high degree exhibit antipreferential attachment [60]. Jackson et al. argued that there existed a bidirectional relationship between the structure of a social network and the behavior of the actors in it [61]. A pure preferential attachment or a pure random generative model does not account for this. Jackson et al. proposed a network-based search model that generated social networks with macrofeatures that were empirically known to be present in social networks, e.g., negative clustering-degree correlation, decreasing hazard rates, and positive assortativity [11, 17, 62]. The intuition behind this model is that newborn nodes create links to preexisting nodes in the network. A proportion of these are selected uniformly at random and the rest are identified by a local search of the neighborhoods of the previously selected nodes.

After an extensive review of network science-related research, it was found that only Jackson et al. [12, 15] focused on identifying the relative ratio of the meeting processes using a network genesis model on the social network. However, the authors did not focus on describing a unified framework for generative models or on utilizing the insights obtained after analyzing socially generated network datasets using genesis models. These are the original contributions of the current work.

2.1. Mathematical description of the unified network genesis model

2.1.1. Mean field approximation of degree distribution based on chance-based link formation

Consider a network where at each discrete time step a node is born and it forms m links to preexisting nodes chosen by a meeting process where all nodes have equal likelihood (uniformly at random). Mean field approximation is used to obtain the expected degree of nodes in the network at a particular time. Initially, the network is assumed to be fully connected with m nodes. The generative process of the network is such that at each time step $i, m < i < t$, a single node is born and it forms m links uniformly at random with existing nodes. Continuous time approximation is used to create a differential equation to calculate the expected degrees of node i in such a network at a particular time t.

$$d_i(i) = m \tag{3}$$

$$(dd_i(t))_{rand} = \frac{m}{t} \tag{4}$$

$$d_i(t) = m * (1 + log(\frac{t}{i})) \tag{5}$$

- $(dd_i(t))_{rand}$ is change in degree of node i with time (gain per unit time)
- $d_i(i)$ is degree of node i at time $= 0$ (initial condition)
- $d_i(t)$ is expected degree of node i at time t

2.1.2. Mean field approximation of network search-based link formation

Similarly, in the network, if a preferential attachment meeting process is observed, the calculation of the expected degree of the nodes using mean field approximation is as follows. The probability of attaching to a given node is proportional to its degree compared to the overall degree in the network. Continuous time approximation is used to obtain the expected degrees of the nodes at a particular time t. Initially, the network is fully connected with m nodes and at each time step $i, m < i < t$, a single node is added to the network. This node forms m links with existing nodes using a preferential attachment strategy. A node born at time i has initial degree

m, so $d_i(i) = m$. Then the differential of the degree of i with respect to time t, i.e. gain per unit time, is $(dd_i(t)/dt)_{pref} = m(d_i(t)/2tm)$. Solving this differential equation, the result is given in Eq. (6).

$$d_i(t) = m * [\frac{t}{i}]^{0.5} \tag{6}$$

Here, $d_i(t)$ is the expected degree of node i in the network at time t.

2.1.3. Mean field approximation of unified model

Each stochastic network generative process has two elements. The first element describes how new nodes are added to the system. The second element is how the newly added nodes form links or relationships with existing nodes of the network. Thus, a unified framework can be used to describe network generative models. The framework assumes the existence of a fully connected network with m nodes in the beginning. It assumes a stochastic process for network growth, i.e. at each discrete interval a single node is added to the system and it forms relationships (links) with existing nodes using a mathematical model. To calculate the expected degree of a node in such a system at a particular time t, mean field approximation can be applied.

The initial condition is $d_i(i) = m$, i.e. at time i node i is born and creates m links with existing nodes. It uses a mathematical model to form links with existing nodes in the network. Using a hybrid strategy where $a * m$ links are created uniformly at random and $(1 - a) * m$ links are created using a network-based search process, the gain of links per unit time by node i is given by $(dd_i(t)/dt)_{unified}$ in Eq. (7). Substituting the values of gain per unit time from previous models, Eq. (8) is obtained. Solving this differential equation, the result is given in Eq. (9) where $d_i(t)$ gives the expected degree of node i in the network at time t in the unified model.

$$(dd_i(t)/dt)_{unified} = a * (dd_i(t)/dt)_{rand} + (1 - a) * (dd_i(t)/dt)_{pref} \tag{7}$$

$$(dd_i(t)/dt)_{unified} = \frac{a * m}{t} + (1 - a) * \frac{d_i(t)}{2t} \tag{8}$$

$$d_i(t) = (m + 2am/(1 - a))(t/i)^{(1-a)/2} - 2am/(1 - a) \tag{9}$$

- $dd_i(t)/dt$ is change in degree of node i with time
- $d_i(i)$ is degree of node i at time $= 0$
- $d_i(t)$ is expected degree of node i at time t

The frequency distribution is given by Eq. (10):

$$F(d) = 1 - [\frac{(m + 2am/(1 - a))}{d + 2am/(1 - a)}]^{2/(1-a)} \tag{10}$$

The frequency distribution of the unified model in Eq. (10) is similar to the distribution of the J-R model proposed by Jackson et al. [12, 15] given in Eq. (11). The J-R model, however, has more parameters.

$$F(d) = 1 - [\frac{(d_0 + rm)}{d + rm}]^{1+r} \tag{11}$$

- d_0 is initial in-degree of a node
- r is $\frac{p_r m_r}{p_n m_n}$, i.e. number of links formed uniformly at random compared to network-based meetings
- m is links a node forms at birth

The unified model is fit to data to understand the relative ratio of the meeting processes.

Algorithm 1: Fitting uniform model to data

Result: X^2, P-value, df
Divide degree distribution P_k into 6 quantiles;
Perform binning and obtain data into equal bins;
For same graph order obtain degree distributions P_K^1 of unified model;
Perform Pearson's chi-square test and compare P_k with P_K^1 Obtain X^2, P-value, degrees of freedom *df*

3. Experimental study

3.1. Datasets

The datasets mentioned below are analyzed to understand the relative ratio of the meeting processes in the network formation process. In each dataset, a system is modeled in the form of a social network. The description of the system and the social relationship between a pair of nodes in these socially generated networks is given in Table 1. Table 2 and Table 3 provide statistical information of the datasets.

Table 1. Description of social relationships.

Sr. no.	**Dataset**	**Description**	**Social relationship** $e_{i,j}$
1	Amazon-Net	Items frequently purchased with one another on Amazon.com	Item i frequently purchased with j
2	Arxiv-Net	Citation graph of papers from Arxiv High Energy physics category	Paper i cites paper j
3	CondMat-Net	Collaboration network of scientists working on condensed matter research	Scientists i and j have collaborated
4	Epi-Net	Trust network between users on Epinion.com	User i trusts j
5	Fb-Net	Friendship network from Facebook	User i and j are friends
6	Gnut-Net	Peer2Peer file sharing network of users from Gnutella.com	User i shared file with j
7	Gow-Net	Friendship network of users from Gowalla.com	User i and j are friends
8	Slash-Net	Friendship network from users of Slashdot.com	User i and j are friends
9	Twt-Net	Followers network from Twitter.com	User i follows j
10	Wiki-Net	Voters network from Wikipedia	User i has voted for user j

3.2. Performance measures

Network genesis models, block models, and simulation models are generative models. Hence, after fitting a model to data, it is possible to simulate a complete network from it to compare the performance of the models. The below network statistics are used for the purpose of comparison.

Table 2. Description of social network datasets.

Description	Facebook	Twitter	Epinions.com	Slashdot	Gowalla	Wiki-Vote
Nodes	4039	81,306	75,879	77,360	196,591	7115
Edges	88,234	1,768,149	508,873	905,468	950,327	103,689
Ratio of nodes in largest WCC	1	1	1	1	1	0.99
Ratio of edges in largest WCC	1	1	1	1	1	1
Ratio of nodes in largest SCC	1	0.84	0.42	0.91	1	0.18
Ratio of edges in largest SCC	1	0.95	0.87	0.98	1	0.38
Avg. clustering coeff.	0.61	0.57	0.14	0.06	0.24	0.14
Fraction of closed triangles	0.26	0.06	0.02	0.01	0.007	0.05
Diameter	8	7	14	10	14	7
90th percentile effective diameter	4.7	4.5	5	4.7	5.7	3.8

Table 3. Description of social networks.

Description	ArXiv-Net	Amazon-Net	CondMat-Net	GnutellaP2P
Nodes	34,546	262,111	23,133	62,586
Edges	421,578	1,234,877	93,497	1,478,928
Ratio of nodes in largest WCC	0.97	1	0.92	1
Ratio of edges in largest WCC	1	1	0.98	1
Ratio of nodes in largest SCC	0.37	0.92	0.92	0.22
Ratio of edges in largest SCC	0.33	0.92	0.97	0.34
Avg. clustering coeff.	0.28	0.42	0.63	0.01
Fraction of closed triangles	0.05	0.09	0.107	0.001
Diameter	12	32	14	11
90th percentile effective diameter	5	11	6.5	7

- The **number of triangles** t formed from any three nodes $u, v, w \in V$ is defined in the following way:

$$t = |\{\{u, v, w\} \mid u \sim v \sim w \sim u\}| \, / \, 6 \tag{12}$$

- The **average degree** d of $G(V, E)$ is:

$$d = \frac{1}{|V|} \sum_{u \in V} d(u) \tag{13}$$

- The **relative size** of the largest connected component equals the size of the largest connected component N divided by the size of the network n:

$$N_{\mathrm{rel}} = \frac{N}{n} \tag{14}$$

- The **global clustering** c of a network is the probability that two incident edges are completed by a third edge to form a triangle:

$$c = \frac{|\{u, v, w \in V \mid u \sim v \sim w \sim u\}|}{|\{u, v, w \in V \mid u \sim v \neq w \sim u\}|} \tag{15}$$

3.3. Results and discussion

Subgraph generation models (SUGMs) [23], stochastic blocks models (SBM), the ER-model, the BA-model, and the unified model are evaluated to identify the relative ratio of meeting processes in the datasets. The comparison of their performances reveals that the BA-model, SUGM, and unified model produce networks with average degree and relative size of giant connected component less than those produced using the ER-model and SBM. However, the networks produced using the BA-model, SUGM, and unified model have higher values of global clustering coefficients and number of triangles. The advantage of the unified model is that it can be viewed as a unified generative model that factors both meeting processes, unlike the BA-model and ER-model.

3.3.1. Analysis of the datasets using unified model

The unified model was used to estimate values of the probability that a chance meeting is found to be beneficial p_r and probability that a choice meeting is found to be beneficial p_s, as given in Table 5.

Analysis of Amazon, Facebook, Twitter, and Gnutella reveal higher link creation between actors obtained through network-based meetings than chance-based meetings. The intuitive reason behind a higher ratio of network-based links in these socially generated networks is due to the friendship recommendations given by these websites. These recommendations are usually obtained by selecting potential candidates from an actor's local neighborhood. Thus, a typical actor would obtain a relatively higher ratio of friends through the network-based search process than chance meetings with strangers. For social networking websites such as Facebook and Twitter, the links created using a network-based search process are empirically found to be significantly more than those created through chance meetings with strangers. This agrees with the intuition that actors in social networking websites would form links with other actors known to them rather than complete strangers. Such networks tend to have higher clustering coefficients. Thus, both Facebook and Twitter are more effective to increase a person's friendship network than websites such as Gowalla.com.

The degree distribution of the Wikipedia electoral process provides valuable insights into the behaviors of the actors involved, i.e. voters and candidates. The high probability of the links being created by chance (random) interactions indicates that candidates seeking administrator rights would have limited success by lobbying their voters. The average clustering coefficient of Wiki-Net was ≈ 0.14. This indicates that voters supporting a single candidate were more likely to be friends of each other. Legitimacy of the electoral process should be ensured by Wikipedia by not allowing candidates access to friend lists of voters. Slashdot is a technology-related news website that allowed users to tag each other as friends or foes. Analysis of the Slashdot

Table 4. Comparison of performance of network genesis models with block models and simulation-based models.

Value	**Data**	**BA-model**	**ER-model**	**Unified model**	**SBM**	**SUGM**
Amazon-Net	t = 2.3e5 d = 3.2 Nrel = 1 c = 0.09	t = 3.4e5 d = 2.18 Nrel = 0.98 c = 0.11	t = 0.68e4 d = 4.46 Nrel = 0.96 c = 0.062	t = 2.02e5 d = 3.46 Nrel = 0.95 c = 0.12	t = 0.58e4 d = 3.116 Nrel = 0.97 c = 0.03	t = 1.96e5 d = 3.63 Nrel = 0.93 c = 0.108
Arxiv-Net	t = 1.6e3 d = 3.5 Nrel = 0.97 c = 0.05	t = 1.57e3 d = 2.89 Nrel = 0.89 c = 0.08	t = 1.08e3 d = 3.8 Nrel = 0.93 c = 0.03	t = 1.34e3 d = 3.9 Nrel = 0.94 c = 0.04	t = 0.98e3 d = 4.32 Nrel = 0.97 c = 0.02	t = 1.35e3 d = 3.85 Nrel = 0.94 c = 0.06
CondMat-Net	t = 1.4e3 d = 4.2 Nrel = 0.92 c = 0.107	t = 1.6e3 d = 3.89 Nrel = 0.94 c = 0.112	t = 0.34e3 d = 3.16 Nrel = 0.95 c = 0.08	t = 1.48e3 d = 3.87 Nrel = 0.89 c = 0.093	t = 0.45e3 d = 3.85 Nrel = 0.95 c = 0.077	t = 1.42e3 d = 3.82 Nrel = 0.96 c = 0.12
Epi-Net	t = 0.3e4 d = 2.87 Nrel = 1 c = 0.02	t = 0.29e4 d = 2.56 Nrel = 0.94 c = 0.04	t = 0.18e4 d = 3.4 Nrel = 0.87 c = 0.02	t = 0.27e4 d = 2.67 Nrel = 0.94 c = 0.03	t = 0.19e4 d = 3.68 Nrel = 0.96 c = 0.018	t = 0.29e4 d = 2.73 Nrel = 0.96 c = 0.03
Fb-Net	t = 1486 d = 2.12 Nrel = 1 c = 0.26	t = 1689 d = 1.98 Nrel = 0.96 c = 0.33	t = 1003 d = 2.09 Nrel = 0.98 c = 0.14	t = 1558 d = 1.9 Nrel = 0.95 c = 0.29	t = 987 d = 1.88 Nrel = 0.96 c = 0.16	t = 1582 d = 2.3 Nrel = 0.96 c = 0.28
Gnut-Net	t = 1.33e3 d = 3.9 Nrel = 1 c = 0.001	t = 1.85e3 d = 4.86 Nrel = 0.95 c = 0.01	t = 0.32e3 d = 6.3 Nrel = 0.98 c = \sim 0	t = 1.45e3 d = 4.25 Nrel = 0.94 c = 0.002	t = 0.33e3 d = 5.9 Nrel = 0.96 c = 0.001	t = 1.38e3 d = 5.02 Nrel = 0.98 c = 0.002
Gow-Net	t = 1.87e4 d = 4.1 Nrel = 1 c = 0.007	t = 1.98e4 d = 5.9 Nrel = 0.97 c = 0.008	t = 0.54e4 d = 6.9 Nrel = 0.98 c = 0.003	t = 1.78e4 d = 5.02 Nrel = 0.96 c = 0.005	t = 0.68e4 d = 6.9 Nrel = 0.95 c = 0.004	t = 1.83e4 d = 4.6 Nrel = 0.96 c = 0.006
Slash-Net	t = 7.7e3 d = 3.14 Nrel = 1 c = 0.01	t = 8.6e3 d = 4.89 Nrel = 0.92 c = 0.018	t = 4.3e3 d = 6.5 Nrel = 0.978 c = 0.009	t = 6.9e3 d = 4.23 Nrel = 0.95 c = 0.014	t = 4.5e3 d = 7..02 Nrel = 0.98 c = 0.008	t = 7.1e3 d = 4.56 Nrel =0.96 c = 0.016
Twt-Net	t = 0.8e4 d = 1.98 Nrel = 1 c = 0.06	t = 1.34e4 d = 1.54 Nrel = 0.92 c = 0.09	t = 0.59e4 d = 2.36 Nrel = 0.98 c = 0.03	t = 0.84e4 d = 1.96 Nrel = 0.94 c = 0.08	t = 0.52e4 d = 2.65 Nrel = 0.96 c = 0.023	t = 0.86e4 d = 1.95 Nrel = 0.93 c = 0.07
Wiki-Net	t = 0.66e3 d = 4.3 Nrel = 0.99 c = 0.05	t = 0.87e3 d = 3.25 Nrel = 0.96 c = 0.04	t = 0.38e3 d = 5.3 Nrel = 0.98 c = 0.01	t = 0.69e3 d = 4.6 Nrel = 0.96 c = 0.03	t = 0.36e3 d = 5.12 Nrel = 0.97 c = 0.012	t = 0.7e3 d = 4.8 Nrel = 0.95 c = 0.03

Table 5. Relative ratio of meeting processes in social networks estimated using the unified model.

Sr. no.	**Dataset**	p_r	p_s
1	Amazon-Net	0.2	0.8
2	Arxiv-Net	0.6	0.4
3	CondMat-Net	0.8	0.2
4	Epi-Net	0.9	0.1
5	Fb-Net	0.1	0.9
6	Gnut-Net	0.1	0.9
7	Gow-Net	0.9	0.1
8	Slash-Net	0.9	0.1
9	Twt-Net	0.3	0.7
10	Wiki-Net	0.9	0.1

friendship network reveals that actors find chance-based meetings significantly more profitable than meeting people through friends of friends, i.e. network-based meetings. This could be because a typical user of Slashdot might not have like-minded tech enthusiasts in his/her usual friend network. This observation is also validated by the low global clustering coefficient (fraction of closed triangles) ≈ 0.01 observed in the socially generated network of Slashdot. Thus, Slashdot could improve users' experience by providing friend recommendations.

Socially generated networks of the Arxiv High Energy Physics paper citation network (Arxiv-Net) indicate nearly equal importance of both chance-based and local search-based interactions in creating new links. As Arxiv does not provide recommendations to researchers, a typical researcher would browse through the large volumes of papers to identify a suitable few. Once these initial links are created, he/she would perform a search in the local neighborhood of the identified papers to select papers of interest. The analysis of the degree distribution of Arxiv-Net reveals that this strategy exists. However, to improve the experience of researchers browsing for papers through Arxiv, recommendations of papers could be provided. This may help a researcher identify suitable citations effectively.

Analysis of the degree distributions of the social networks of Epinions.com, a collaboration network of scientists working on condensed matter, and Gowalla.com reveal higher benefits from links created via random interactions with strangers. Epinions.com has actors that 'trust' each other and this forms a web of trust. As recommendations are not provided for actors to trust other actors, a typical user has to browse through the website to identify other actors whom it could trust. As trust is not transitive, a user would not trust the friends of an actor he trusts. Thus, such a social network will have low values of reciprocity of links. Epinions.com could improve the experience of its users by providing users with additional information in the form of metadata, which could help users in their decision-making. The analysis of the collaboration network [CondMat-Net] and Gowalla.com also reveals that random meetings of strangers prove profitable and lead to link creation, but as these networks expand it will become infeasible for actors to identify suitable partnerships by chance meetings. CondMat-Net has a high average clustering coefficient of ≈ 0.68; thus, even though a scientist finds a potential collaborator by random search it is highly likely that he/she would establish further collaborations. Although Gowalla.com is a social networking website, it is not suitable for increasing the reach of a friendship network compared to Facebook and hence compared to Facebook it would have lesser user engagement.

4. Conclusion

In any community, actors interact with other actors. These interactions may be premeditated or based on chance and usually have a payoff associated with it. If the payoffs are high, then a link is created. It is through such interactions that a social network is born. Quantifying the role of chance in the network formation process is difficult as multiple micromechanisms are prevalent in a social network simultaneously. However, with reasonable assumptions about the underlying network formation process, it is possible to investigate this problem using a suitable generative model.

Several generative models exist in the literature; however, they may not be able to capture the stylized facts such as small world effect, power law degree distributions, etc. seen in socially generated networks. Hence, it necessary to analyze the generative models and understand the differences in their stochastic processes using experimental results. Stochastic block models are enrichments of Erdos–Renyi random networks, while SUGMs are based on the intuition that a network is a byproduct of various subgraphs or graph statistics such as cliques, triads, dyads, etc. However, these generative models are not suitable for capturing the relative ratio of meeting processes in social networks.

The JR-model was described as a unified model that could factor both preferential attachment and random selection. Hence, it was used to estimate the meeting processes in a social network. The advantage of this approach is highlighted in this work. Once the proportion of chance and choice in network formation is established, it will be possible to make a statistically valid analysis of the effectiveness of the systems and provide recommendations for their improvements.

References

[1] Jimenez-Martinez A. Discrimination through versioning with advertising in social networks. Econ Theor 2018; 40: 1–40.

[2] Gosak M, Markovič R, Dolenšek J, Rupnik MS, Marhl M, Stožer A, Perc M. Network science of biological systems at different scales: a review. Phys Life Rev 2018; 24: 118–135.

[3] Gosak M, Markovič R, Dolenšek J, Rupnik MS, Marhl M, Stožer A, Perc M. Loosening the shackles of scientific disciplines with network science: reply to comments on network science of biological systems at different scales: a review. Phys Life Rev 2017; 24: 162-167.

[4] Jalili M, Perc M. Information cascades in complex networks. J Compl Netw 2017; 5: 665–693.

[5] Wang Z, Yamir M, Stefano B, Perc M. Vaccination and epidemics in networked populations—an introduction. Chaos Soliton Fract 2017; 103: 177-183.

[6] Jalili M, Orouskhani Y, Asgari M, Alipourfard N, Perc M. Link prediction in multiplex online social networks. R Soc Open Sci 2017; 4: 1-11.

[7] Martinčić-Ipšić S, Močibob E, Perc M. Link prediction on Twitter. PloS one 2017; 12: 1-21.

[8] Nickel M, Kiela D. Poincare embeddings for learning hierarchical representations. Adv Neur In 2017; 31: 6338–6347.

[9] Perc M, Jordan JJ, Rand DG, Wang Z, Boccaletti S, Szolnoki A. Statistical physics of human cooperation. Phys Rep 2017; 687: 1-51.

[10] Ribeiro HV, Alves LG, Martins AF, Lenzi EK, Perc M. The dynamical structure of political corruption networks. J Compl Netw 2018; 28: 1-15.

[11] Fekom M, Coolen ACC, Lopez FA, Barucca P. Exactly solvable random graph ensemble with extensively many short cycles. J Phys A-Math Theor 2018; 33: 1-15.

[12] Rogers BW, Jackson MO. Meeting strangers and friends of friends: how random are social networks? Am Econ Rev 2007; 97: 890–915.

[13] Garas A, Schweitzer F, Tomasello MV, Napoletano M. The rise and fall of R&D networks. Ind Corp Change 2017; 726: 617–646.

[14] Zenou Y, Jackson MO, Rogers BW. The economic consequences of social-network structure. J Econ Lit 2017; 55: 49–95.

[15] Wolinsky A, Jackson MO. A strategic model of social and economic networks. Netw Grps 2003; 37: 23–49.

[16] Snijders TAB. The statistical evaluation of social network dynamics. Sociol Methodol 2001; 31: 361–395.

[17] Chandane M, Bhirud S, Nerurkar P, Shirke A. Empirical analysis of data clustering algorithms. Procedia Comput Sci 2018; 125: 770–779.

[18] Kamada Y, Iijima R. Social distance and network structures. Theor Econ 2017; 12: 655–689.

[19] Bhirud S, Nerurkar P. Modeling influence on a social network using interaction characteristics. Int J Comput Mat Sci 2017; 6: 152–160.

[20] Diaconis P, Chatterjee S. Estimating and understanding exponential random graph models. Ann Stat 2013; 41: 2428–2461.

[21] Kalish Y, Lusher D, Robins G, Pattison P. An introduction to exponential random graph (p*) models for social networks. Soc Networks 2007; 29: 173–191.

[22] Snijders TAB. Markov chain Monte Carlo estimation of exponential random graph models. J Soc Struct 2002; 183: 1–40.

[23] Zhang J, Zou X, Yang J. Microblog sentiment analysis using social and topic context. PLoS One 2018; 13: 119–163.

[24] Johari R, Leduc MV, Jackson MO. Pricing and referrals in diffusion on networks. Game Econ Behav 2017; 104: 568–594.

[25] Williams JW, Ciliberto F, Cook EE. Network structure and consolidation in the us airline industry, 1990–2015. Rev Ind Organ 2017; 10: 1-34.

[26] Kermani A, Sommavilla C, Maggio MD, Franzoni F. The relevance of broker networks for information diffusion in the stock market: technical report. Nat B Eco Research 2017; 12: 1-76.

[27] Tang J, Qu M, Wang M, Zhang M, Yan J, Mei Q. Line: Large-scale information network embedding. In: Proceedings of the 24th International Conference on World Wide Web; 18–22 May 2015; Florence, Italy. New York, NY, USA: International World Wide Web Conferences Steering Committee. pp. 1067–1077.

[28] Huang X, Li J, Hu X. Label informed attributed network embedding. In: Proceedings of the Tenth ACM International Conference on Web Search and Data Mining; 6–10 February 2017; Cambridge, UK. New York, NY, USA: ACM. pp. 731–739.

[29] Huang X, Li J, Hu X. Accelerated attributed network embedding. In: Proceedings of the 2017 SIAM International Conference on Data Mining; 16–22 May 2017; Notre Dame, IN, USA. New York, NY, USA: ACM. pp. 633–641.

[30] Liao L, He X, Zhang H, Chua TS. Attributed social network embedding. arXiv preprint 2017, arXiv:1705.04969.

[31] Bandyopadhyay S, Kara H, Kannan A, Murty MN. Fscnmf: Fusing structure and content via non-negative matrix factorization for embedding information networks. arXiv preprint 2018, arXiv:1804.05313.

[32] Tsitsulin A, Mottin D, Karras P, Muller E. Verse: Versatile graph embeddings from similarity measures. In: Proceedings of the 2018 World Wide Web Conference on World Wide Web; 23–27 April 2018; Lyon, France. Geneva, Switzerland: International World Wide Web Conferences Steering Committee. pp. 539–548.

[33] Ou M, Cui P, Pei J, Zhang Z, Zhu W. Asymmetric transitivity preserving graph embedding. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 13–17 August 2016; San Francisco, CA, USA. New York, NY, USA: ACM. pp. 1105–1114.

[34] Rozemberczki B, Davies R, Sarkar R, Sutton C. Gemsec: Graph embedding with self clustering. arXiv preprint 2018, arXiv:1802.03997.

[35] Rozemberczki B, Sarkar R. Fast sequence based embedding with diffusion graphs. In: International Conference on Complex Networks; 11–13 December 2018; France. Cambridge, UK: Springer. pp. 99-107.

[36] Perozzi B, Al-Rfou R, Skiena S. Deepwalk: Online learning of social representations. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 24–27 August 2014; Washington, DC, USA. New York, NY, USA: ACM. pp. 701–710.

[37] Grover A, Leskovec J. node2vec: Scalable feature learning for networks. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 13–17 August 2016; San Francisco, CA, USA. New York, NY, USA: ACM. pp. 855-864.

[38] Sheikh N, Kefato Z, Montresor A. gat2vec: Representation learning for attributed graphs. Computing 2018; 9: 1-23.

[39] Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J. Distributed representations of words and phrases and their compositionality. Adv Neur In 2013; 23: 3111–3119.

[40] Cao S, Lu W, Xu Q. Grarep: Learning graph representations with global structural information. In: Proceedings of the 24th ACM International on Conference on Information and Knowledge Management; 19–23 October 2015; Melbourne, Australia. New York, NY, USA: ACM. pp. 891–900.

[41] Liu Q, Li Z, Lui J, Cheng J. Powerwalk: Scalable personalized pagerank via random walks with vertex centric decomposition. In: Proceedings of the 25th ACM International on Conference on Information and Knowledge Management; 24–26 October 2016; Indianapolis, IN, USA. New York, NY, USA: ACM. pp. 195–204.

[42] Pandhre S, Mittal H, Gupta M, Balasubramanian VN. Stwalk: learning trajectory representations in temporal graphs. In: Proceedings of the ACM India Joint International Conference on Data Science and Management of Data; 11–13 January 2018; Goa, India. New York, NY, USA: ACM. pp. 210–219.

[43] Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. arXiv preprint 2013, arXiv:1301.3781.

[44] Tran PV. Learning to make predictions on graphs with autoencoders. arXiv preprint 2018, arXiv:1802.08352.

[45] Wang Z, Ye X, Wang C, Wu Y, Wang C, Liang K. Rsdne: Exploring relaxed similarity and dissimilarity from completely-imbalanced labels for network embedding. Network 2018; 11: 475-482.

[46] Zhang M, Cui Z, Neumann M, Chen Y. An end-to-end deep learning architecture for graph classification. In: Proceedings of AAAI Conference on Artificial Inteligence; 2–7 February 2018; New Orleans, LA, USA: AAAI. pp. 531–538.

[47] Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. arXiv preprint 2016, arXiv:1609.02907.

[48] Chen J, Ma T, Xiao C. Fastgcn: fast learning with graph convolutional networks via importance sampling. arXiv preprint 2018, arXiv:1801.10247.

[49] Donnat C, Zitnik M, Hallac D, Leskovec J. Spectral graph wavelets for structural role similarity in networks. arXiv preprint 2017, arXiv:1710.10321.

[50] Perozzi B, Kulkarni V, Chen H, Skiena S. Don't walk, skip!: online learning of multi-scale network embeddings. In: Proceedings of the 2017 ACM International Conference on Advances in Social Networks Analysis and Mining; 1–3 August 2017; Sydney, Australia. New York, NY, USA: ACM. pp. 258–265.

[51] Desa C, Re C, Gu A, Sala F. Representation tradeoffs for hyperbolic embeddings. arXiv preprint 2018, arXiv:1804.03329.

[52] Goodreau SM. Advances in exponential random graph (p*) models applied to a large social network. Soc Networks 2007; 31: 231–248.

[53] Hoff PD, Raftery AE, Handcock MS. Latent space approaches to social network analysis. J Am Stat Assoc 2002; 64: 1090–1098.

[54] Snijders TAB. Longitudinal methods of network analysis. Enc Com Sys Sci 2009; 24: 5998–6013.

[55] Hoff PD. Dyadic data analysis with amen. arXiv preprint 2015, arXiv:1506.08237.

[56] Denny M. Social Network Analysis. Amherst, MA, USA: Academic Press, 2014.

[57] Denny M. Intermediate Social Network Theory. Amherst, MA, USA: Academic Press, 2015.

[58] Balasubramanian M, Schwartz EL. The isomap algorithm and topological stability. Science 2002; 295: 7.

[59] Roweis ST, Saul LK. Nonlinear dimensionality reduction by locally linear embedding. Science 2000; 290: 2323–2326.

[60] Snijders TAB, Block P, Stadtfeld C. Forms of dependence: comparing SAOMs and ERGMs from basic principles. Sociol Method Res 2016; 43: 672–680.

[61] Hric D, Fortunato S. Community detection in networks: a user guide. Phys Rep 2016; 659: 1–44.

[62] Spenkuch J, Cicala S, Fryer RG. Self-selection and comparative advantage in social interactions. J Eur Econ Assoc 2017; 16: 1-44.