ORIGINAL ARTICLE

Information diffusion modeling and analysis for socially interacting networks

Pawan Kumar[1] · Adwitiya Sinha[2]

Abstract

Social network analysis provides innovative techniques to analyze interactions among entities by emphasizing social relationships. Diffusion in the social network can be referred to spread of information among interconnected nodes or entities in a network. The rate and intensity of diffusion depend upon network topology and initialization of network parameters. Individual nodes act as source of motivation for others in the diffusion process. The epidemic model is one of the basic diffusion models that helps in analyzing the transmission of infectious disease from one person to another through social connections. This can be further generalized for other socially connected platforms involving information exchange. In our research, we have proposed a diffusion methodology for tracking the rate with which information spread over underlying social interaction structure, with variation in time and other social parameters. In addition to forward state transitions, recoverable transition is also proposed, which allows a node currently under influence of incoming information, to revert back to previous state of perception. The proposed model also assists in predicting the fraction of population getting diffused over real and large-scale complex network for specific temporal domain.

Keywords Social network analysis · Information diffusion · Complex networks · Centrality measure · Epidemic model · State transitions

1 Introduction

Social network analysis (SNA) uses theoretical concept of graphs and networks as basic foundation to analyze the social interaction structure and relationships. Such analysis can be carried out in terms of friendship, kinship, organizations, states, web sites, or citations between publications, disease transmission and other social activities among connected entities in the network (Li and Chen 2007; Rosa and Giua 2013; Broecheler et al. 2010; Doo and Liu Sept. 2014; Zhang and Moura 2014). Graph contains nodes, which are represented as individuals or actors, and edges denote relationships between them. A network having more connection to other networked nodes promotes better sharing of ideas along with spread of information within the same network. Isolated or self-loop is considered to have no significance in network analysis (Wang et al. 2015; Fatima et al. 2013; Bo and Liu 2010). SNA has various applications in the following area, for instance diffusion of innovations, community detection, communication in complex domains such as criminal networks, demography, economic, sociology, human ecology, social media, security, news searching, business and marketing, opinion mining, intrusion detection, link prediction, recommendation system, dynamics of gossip/rumors, characteristic of person for different objective, advertising, information broadcasting, finding the potential customer, etc. (Mahdi et al. 2010; Niu et al. 2013; European Union 2010; Kuo et al. 2011). SNA captures the dynamically changing pattern of social networks and allows content sharing, long-term forecasting, trend prediction, etc., using certain specific diffusion modeling. The content like text, image, audio, video, etc., spread over the web is same as of epidemic model. SNA is further classified into micro-, meso- and macro-levels. Micro-level explores with an individual or

✉ Adwitiya Sinha
  mailtoadwitiya@gmail.com

  Pawan Kumar
  pawankumar.063@gmail.com

[1] DGAQA, Ministry of Defence, Government of India,
  New Delhi, India

[2] Department of Computer Science and Engineering, Jaypee
  Institute of Information Technology, Noida, Uttar Pradesh,
  India

a small set of individuals. Meso-level is normally applied to random networks or scale-free networks. In random network, an edge is formed by some probability value p from a set of n vertices. It is denoted by $G(n, p)$ Erdos–Renyi model. In scale-free networks, highest degree nodes are referred as hubs that follow power law degree distribution. Macro-level analysis, generally carried out over a large population, is mainly used in social behavioral sciences and economics.

1.1 Research motivation

A social network is inherently dynamic and continuously evolving in nature. Such evolution is attached with the spread of information within the network. Such networks consist of thousands or millions of nodes based on properties of networking, for instance traffic system, activity of financial market, World Wide Web (www), etc. The real-world analysis is performed on the basis of observing behavior from theoretical concepts of such systems. The density of network connections affects the process of manner in which information propagates. The adaptability of a node (also called entity or agent) toward a new behavior or technology relies on number of neighbors (i.e., node degree), associated cost, clustering coefficient, diameter, centrality, thresholds and communication media. Small diameter in the strongly connected network would lead to nodes being more beneficial (Fouad et al. 2012). Normally, a fraction of a sample population is randomly selected as initiator node to spontaneously adopt and spread the social behavior. Overall, this motivates to perform the analysis of the spread of information to a larger portion of the population within a given frame of time. This phenomenon is termed as information diffusion. Diffusion in the context of social network is the study of complex networks having irregular structure that dynamically changes with respect to time. This relies upon network structure and parameter thresholds. The extent to which a new behavior spreads throughout a society depends not only on its relative attractiveness or payoff, but also on underlying social structure.

Transmission of disease follows micro-level and scale-free network idea. Some diseases spread at a higher rate and infect a large population in community, while some disease spread at a lower rate. It mainly depends on type of disease, immunity power, transmission rate and network parameters. In the real graph, information propagation depends on degree distribution and node characteristics. The epidemic disease in the community spread similar as the information propagates from one node to another in social media. In the context of social media, some person forward the information to others, while some do not forward the information. Active person on online social media forward/share information quickly. The epidemic model follows social media information propagation feature. Due to internal immunity power, some persons have less effect of infection or we can say that they are susceptible to that infectious disease and revert back to their healthy state. Social media and epidemic model have similar network structure.

1.2 Research problem

Several approaches have been developed for addressing the problem of modeling diffusion for social structures. This may include adoption of ideas, launch of new products or technological innovation, spread of viral diseases, etc. (Kuo et al. 2011; Fouad et al. 2012; Krishna Kumar and Geethakumari 2013; Wang et al. 2013; Sato et al. 2012; Jackson 2008; Kumar and Sinha 2016; Sahal 1979; Giamberardino et al. 2017; Kandhway and Kuri 2017). Diffusion model has also been used in marketing for predicting the lifecycle and demand forecasting of a new product so as to improve the decision-making process. Diffusion can be mathematically defined as an optimization problem with single objective of optimizing the goal of spreading information and capturing the rate of information dispersion. Such methods can be explored by different approximation and randomization algorithms (Mahdi et al. 2010). In realistic social networks, individual entities simultaneously store various pieces of information during communication. Due to rapid increase in internet and mobile devices, use of social network becomes extremely large (Jiang et al. 2014). For the brand campaigns and communication, diffusion modeling assists company personnel to discover potential customer zones with the help of social media (Mini and Jacob 2014). Higher connectivity in social media interprets better promotion of products in the market. Various types of new information are posted everyday, i.e., launching of new item, some deal advertisement, rumors, etc., thereby exposing the dynamics of information diffusion to be affected by several factors (Jiang et al. 2014).

Diffusion model assists in analyzing friendship network as collaboration in different areas to find out the central (key) player or the main actor. Spread of information in the social network is conceptually same as in the dispersion of viral infection in epidemic model (Bo and Liu 2010). The process of diffusion is mainly governed by the innovation of several methods and models. Diffusion refers to the pattern and rate of propagation of some innovated idea though certain medium among the individuals in a given period of time. An innovation is an idea that has social relevance. Innovation, communication medium, time, social system (media) are important factors in the diffusion process. It explains in what manner, what way and at what rate new ideas or technological events get diffused over a given population.

Our objective is to develop a novel diffusion model and apply it over real-world data set to elaborate its impact on the basis of different parameters. The proposed diffusion model would be capable of analyzing the rate of information or

communication spread in generated network structure and predicting the outcome of the network after some instant of time. The study also reveals the influence of variation in network parameters required for the spread of information. Further, prediction of the state of network diffusion is analyzed on the basis of its early spreading strategy.

2 Research work flow

Social networking often produces a huge amount of profile-centric social data, thereby forming complex network structure emerging as an area for understanding the propagation of information and marketing innovation. Moreover, our research establishes the fact that infections may occur at any time in future, thus exhibiting dynamism in the information propagation pattern. The research highlights of our work innovation can be summarized as follows:

- Our work assumption centered on the fact that the information propagation in social media inherently resembles epidemic model for diffusion.
- Inevitably, larger degree nodes are known to act as rapid expediters for news propagation. Therefore, such nodes are initially selected to trigger the procedure for spreading information in the network.
- Our proposed model combines the advantages of existing models by introducing diffusion to operate in possibilities of two phases

    – First phase is a cyclic process that allows nodes to get influenced temporarily by the ongoing diffusion process; or may choose to revert back to the initial state
    – The second phase, however, offers permanency toward the final acceptance of the diffused information in the network

    This provides more realistic simulation of human inclinations toward propagation of information in the real world.

- Owing to the flexibility in transitions, our model exhibits adaptability, while incorporating the dynamism in user behavior
- Our proposed algorithm operates on the giant component (i.e., largest connected component) of the real graphs to analyze the impact of diffusion at mega-scale. Various networking parameters are considered, including degree centrality, density, path length and cluster coefficient.

3 Related work

Information in the social network is diffused in the similar way as infection spread in the epidemic model. Infectious diseases are caused by pathogenic microorganism such as bacteria, viruses, parasites or fungi. The disease can spread directly or indirectly from one person to another. An infectious disease is termed contagious if it is easily transmitted from one person to another.

Among several research initiatives in diffusion process, the benchmark epidemic models are the predominant ones, which includes Bass, SIS, SIR and SIRS models. SIR categorizes the host with in a population as susceptible (if previously exposed to disease), infected (if currently colonized by the pathogen) and recovered (if they had successfully cleared the infection). SIR system is an excellent example of a damped oscillator, which means the inherent dynamics contain a strong oscillatory component, but the amplitude of these fluctuations declines over time as the system equilibrates. Fraction of infectiveness oscillates with decreasing amplitude as it settles toward the equilibrium (Giamberardino et al. 2017). In the Bass model, infection initially grow at a rapid rate, but the rate of growth tapers off and finally declines with time. The channel of communication is mass media or interpersonal communication. This depicts the S-shape curve pattern. Transition is allowed to move from one state to another, without a permissible backward move, termed as one-way transition process. Some instances include movie seen, caught flu, product used, etc. In the SIS model, there are two states susceptible and infected. It allows nodes to change behavior over time, allowing it revert back to its previous state. In this model, the need for steady state arises. It means nodes in the network get infected by some rate, get recovered by some another rate. Steady state becomes essential to get established such that the rate of change in infection remains constant. An individual can be infected multiple times throughout their life with no apparent immunity. Susceptible nodes become infectious immediately when it becomes in contact with an infected node. The recovery from infection is followed by an instant return to the susceptible pool. In the modified SIS model, people having more interactions are exposed to infection rapidly. Hence, the rate of infection (rate of diffusion) largely depends on the degree distribution. SIRS model extends SIR and SIS models as two behavioral extremes where immunity is either lifelong or simply does not occur. An intermediate assumption is that immunity lasts for a limited period before warning such that the individual is once again susceptible.

The authors in Li and Chen (2007) introduced Nash and Pareto rule for nodes or agents that can follow rules for adoption of information or behavior in network. It depends on choice whether they have to change their value or not.

The agents collaborate to obtain maximum payoff value which is considered a relevant network outcome. Agents decide to adopt a technology while considering the outcome (or payoff) computed by neighbor agents and finally comparing with its own current value. Nearest neighbors of an agent contribute in deciding whether or not to change the payoff. This process is repeated for all agents in the network. At any time, agents adopt the higher payoff value. Results established the fact that agents using the Pareto rule are more co-operative, but for agents following Nash rule behave in a selfish manner. In Rosa and Giua (2013) authors introduce a scenario where each node in the social network is inclined towards the behavior of its neighbor nodes. By following a threshold mechanism, nodes periodically decide to accept, leave or maintain the innovation. An individual adopts behavior if some neighbors have already been adopted. While in epidemic model, an entity adopts a behavior with some probability if at least one of its neighboring entities has already adopted the same. In linear threshold, model innovation begins with seed (or initial set), i.e., a group of entities spontaneously accepting the innovation. Gradually, nodes representing individual nodes transit between active and non-active states. When active neighbors of node are above than a predefined threshold, then the non-active nodes become active. Active nodes are those nodes which take part in diffusion process. The nodes are allowed to perform transition from inactive state to active, but not vice versa. This leads to adoption of innovation in a permanent mode called progressive model. However, authors are permitted to voluntarily change decision regarding the suitability of progressive model is not suitable who changes their behavior as per time such as habits may change. Periodically, each entity updates its state by considering its neighbors' state and eventually decides either to become active or inactive. Social cohesion is an important factor to comprehend the collective behavior of socially connected entities (or nodes). The proposed system is subdivided into different categories: during and after the network initiation phase. System behaves as a progressive model in the seeding time period T_s. Every node in the initial state belonging to a seed set is allowed to be active for time period $t \in [0, T_s]$, independent of their neighbors. For $t > T_s$, a node become active if their outcome value is above than the threshold and its neighbors become middleware in adoption process. During this time period, each node can update their state accordingly. This would result in non-progressive evolution as network transforms dynamically with change in topology or influence weight. Authors concluded that threshold mechanism is most suitable for individual behavior and social influence phenomena while epidemic models are mainly used for mass behavior or community.

The diffusion of competitive phenomena such as product adoption and political election is through social networking (Broecheler et al. 2010). It is presumed that an individual has to choose one product from a set of competitive products. It uses Weighted Generalized Annotated Programs (WGAPs) and can learn from past historical data together with a set of integrity constraint. It also uses the Most Probable Interpretation (MPI) mechanism. Each node is assigned some value for their occurrence. The MPI has most likely outcome among competitive process and highest probability of correct interpretation. The certainty of spread of information from one vertex to another depends on the knowledge about the vertex itself and type of links between the vertices. The WGAP graph is further partitioned so as to split the problem domain into smaller fragments and solve till convergence. However, this strategy becomes infeasible for large social network. Further, in Doo and Liu (2014) authors introduce uniform distribution of heat diffusion, in which every node gets the same quantity of heat (or influence) from its adjacent nodes. This is followed with propagation of heat intensity equally to each of neighboring nodes with low values. The extent of heat or influence dispersed among its neighbors can be analyzed from the number of out-degree of that node. Authors propose computation of diffusion probability for measuring influence between pair of nodes, which can make a distinction among nodes that possess a higher level of interactivity and help to reveal factors that are responsible for diffusion in a social network. In some cases, higher degree nodes not always interact or spread information to their neighbors. Incentive phenomenon can help to promote new products. Active nodes represent group of people that adopt newly launched products. With limited incentives, the goal is often set to attract large number of people who take part in diffusion. In the cascading viral marketing, subsets of individual were identified who adopts new idea, newly launched product or some novel innovation that eventually triggers more people in future to get influenced.

In Zhang and Moura (2014), authors introduce the Susceptible Infected Susceptible (SIS) framework from epidemic model to analyze the manner in which network connections effect spreading of ideas, opinion, and innovation in the social network media. It follows Markov process and influence as neighbor-to-neighbor infections. For the steady state, the adjacency matrix shows variation at each interval for a fixed time period. The agents have two possible states: as a spreader and as an adopter. In diffusion process, spreaders spontaneously become adopters and vice versa through peer-to-peer influence as that of SIS model so that structure of the network changes. Each agent is allowed to interact directly with every other agent in the network. Agents are susceptible to infection and have the same value for all the agents with or without involvement of peer influence. For the maximization of probability of outcome value, the focus is to minimize the number of infected nodes. Diffusion rate increases as a number of spreaders increase. On the basis

of threshold value of an agent, the behavior is decided whether to remain in its previous state or upgrade its state. Another research conducted in Wang et al. (2015) specifies that information or influence among individuals that pass through cascading mechanism maximizes their utility value through uncertainty. Uncertainty generally occurs due to incompleteness in the knowledge. Therefore, change in value of uncertainty in user decision can affect diffusion. Adoption of behavior is directly related to the rate of infection or nature of neighbors. Larger degree nodes act as rapid expediters for spreading news in network topology. The authors aim at discovering nodes that can spread information at a faster rate with lower degree that learn from their informed neighbors and rapidly forward the same piece of information to remaining neighbors. This process continues from large degree nodes (acting as hubs) to other nodes in the network. Uncertainty factor was modeled using the game theory. For each action user, payoff depends on location-specific data in the network, i.e., degree, cost and their preference. In yet another research (Fatima et al. 2013), authors introduce a diffusion model with multiple objectives for nonlinear and complicated activities that support information exchange. In the social network, various information spread at the same time. On the basis of ability to diffuse the information, individuals can be further differentiated. Every piece of information has its own importance, type and associated constraints for gaining objectives for each individual in the network. In the competitive environment, user can choose only one piece of information and can revert back to stick to previous information. In multi-objective environment, few individuals in network always gain more information quickly instead of how diffusion process starts, depending upon its centrality in the network.

The pattern of sharing ideas publicly to socially connected users is highlighted in Bo and Liu (2010). It follows epidemic phenomena for dissipation of information among individuals. In the initial phase, a person does not know about the information (susceptible) and expect to obtain information with some probability when it comes in contact of infected (or diffused) user; or no longer interested in getting information. Author introduced infectiousness as a function of time. As the effect of viral diseases gradually decreases with time, people are no longer concerned of getting new information. In Mahdi et al. (2010), authors introduce the ratio of number of adopters to non-adopters as the rate of diffusion degree. The real social network lies between small word and scale-free network. The ties between nodes are considered highly central in accelerating the diffusion process. Hub nodes attract large number of relations within network. New nodes are added to higher degree nodes according to preferential attachment rule. An innovation is initially adopted by a few individuals, which relatively increases with time and finally attains stability.

Diffusion process stops at the level where no more nodes in network interested to adopt the innovation. Similarly, method of reverse diffusion works until all nodes become non-adopters. The spreading of information depends on degree of diffusion parameter. A cascading mechanism was introduced in Niu et al. (2013) on various sub-graphs by partitioning the network for analyzing its characteristics. Authors collected data set from online activities of users over web. A message is sent to all followers, while user is actively involved in some prescribed activities. If a user is influenced by another user to follow him, new connection gets added to the cascading sub-graph, which happens to follow power law distribution. Cascading mechanism with few initiators eventually grow into bigger size with diffusion process. Node popularity and similarity affect information diffusion. The authors also established that the existence of external influence or hidden nodes plays an important role while diffusion. Further, in Jiang et al. (2014) authors introduce the game theoretic approach that focuses on user behavior and compares it with the machine learning-based method. User with new information is called as mutant through their learning, interaction and decision making. Player can reproduce its own strategy under some rule and condition among the population. New information is initiated from a single or small set of users. The process of information diffusion also relies on the manner that other users choose to forward their information. Users relay information with higher probability if their neighbors also forward the same information. They observe strategy to maintain neighbors' information using payoff matrix. User attraction depends on type of information, i.e., recent topics or advertisement. In complete network, each communicating user is connected to every other user. The authors analyzed the manner in which information transmits to other users in a different group as in Facebook or Google Plus. In yet another research in Zaffar et al. (2014), authors introduced the impact of interrelationship in the social network and the manner it affects the decision making and information diffusion as in propagation of virus, spreading of epidemic diseases, flow of knowledge, etc. Network parameters help in identifying the core node in the diffusion process. The authors of Yagan et al. (2012) introduce information diffusion in various closely interacting networks through web. In physical information network, information spreads among the population via some direct communication media as in SIR (Susceptible Infected Recovered) epidemic model. An entity is either susceptible, i.e., it is not informed, or infectious, i.e., it is aware of information and capability of diffusing it among neighbors. Moreover, the state recovered gives interpretation of the fact that entity is not interested in information. Diffusion in one network may also affect other networks. New information spreads to some fraction of population in a starting phase. Physical and social networks are described by random graph

with different topologies. Information spread is studied in case of a large fraction of nodes in a connected network as compared to a disjoint network.

In Hong-wei et al. (2011), information diffusion curve is analyzed that depends on the structure of network. The rate is affected by core elements of the innovation and exchange of knowledge in innovative actors. Weak ties act as intermediation between different set of actors. Transfer of knowledge flow occurs by node interaction. Strong ties found between nodes exhibit similarity in terms of various social network characteristics. In Jiang and Jiang (2015), the diffusion of information is modeled using multi-agents such as actors, communication media and content with information. Diffusion is a collective behavior of social actors for one-to-one and many-to-one interaction. A set of agents involved in closed form of interaction maximizes the influence. Information cascading behavior, highlighted in Kuo et al. (2011), exhibits a higher rate of information diffuse each time a socially connected user interacts. Top influenced users are considered to be more interested in the diffusion process. According to the proposed method, process begins with identifying root users. These users influence other connected users and initiate calculation of influence for each user. This iteratively progresses till influence parameter for all nodes is calculated. Node can become active from inactive state in the first attempt with independent cascade model. In another recent research carried out in Fouad et al. (2012), an approximation algorithm is illustrated that determines the scope of sharing of information. Users share their private or public information with others, keeping in consideration the intensity of risk accompanying with them through web. Users can either accept or reject the information coming from other users in social network. The average number of intermediates between any two randomly selected users is approximately six, as established by six-degree separation phenomenon. While sharing of private information among some intended neighbors, chances for information leakage to unwanted users persist. Furthermore, classification method is used to identifying the trusted and un-trusted set of such users and grant access by predefined utility function.

The spread of misinformation and associated impact of mis-campaign is highlighted in Krishna Kumar and Geethakumari (2013), especially for networks using evolutionary game theory for large information-bearing nodes in the dynamic environment. The online social network provides a platform for real-time news spread by sharing. Misinformation is defined as unintentionally spread of false or inaccurate information. The authors derive a threshold value that determines whether a node adopts information or transmits to other neighbors who have not adopted the same information. The decision capability of the people changes time to time. Strategy with higher payoff (cost and benefit ratio) would spread among population by learning or infection.

The number of infected population increases and rate of spread become exponential. Once information is accepted by certain number of users, it becomes available to all users, and thus, adoption mechanism persists. It is primarily used in analysis, future prediction, preplanning strategies due to semantic cyber-attacks. In Wang et al. (2013), authors described novel ways of predicting the popularity of news. The authors analyzed the source set to estimate the influenced number of users. The users in same group (having similar properties) possess higher probability to make direct friends, hence sharing strong influence like spreading of disease within a community by some infectious agent, etc. The study established the fact that rate of spreading information depends on distance and the number of users. The users with less distance value are considered to be the super spreaders that represent the peak growth rate. In Sato et al. (2012), authors describe information spread within community that allows each node to receive content from all other nodes in the network. It is also assumed that a person can simultaneously belong to more than community and communicate between different communities through direct communication channel. Nodes that are close to each other share similar information in network. This established the fact that social-centric networking maximizes the efficiency of information diffusion from different communities. The hub nodes, in each community, act as sources for rapid diffusion leading to maximum communication, thereby enhancing the spread of information.

In Kumar and Sinha (2016), authors illustrate real-time information propagation in Twitter social network on the basis of network parameters such as degree, centrality, clustering coefficient, density, diameter, average path length. Real-time diffusion in the emergency communication system is for broadcasting and multi-casting news. On the basis of tweet messages, authors find out most occurring domain, recurring word pairs, top word occurrences in a particular community for a graph so that we have the idea about the most popular events in the network. In Aditya Prakash et al. (2014), authors emphasize over large-scale social graph, wherein finding infectious nodes is a typical problem without knowing the source nodes. The authors applied novel immunization mechanism so as that infection can be controlled and tracked down. In yet another research in Kempe and Tardos (2003), authors describe the propagation of ideas, information and influence in the social network as word of mouth phenomena. By giving free samples, rewards, discount more number of peoples, were discovered to be attached for the marketing of new product. In research conducted in Cho et al. (2011), the authors have illustrated the significance of location-based online social networks in providing data for understanding the pattern of mobility in human behavior. The experimental outcome showed decrease in degree of friendship as distance in the network

increases. The authors in Cheng et al. (1805) modeled the sharing of information through cascading phenomena in online social network. The results highlighted that information depicting the user behavior is adopted quickly as compared to broadcast method. In another innovative work conducted in Lamprier et al. (2016), the authors described information diffusion among users in the network as an iterative process. It finds the impact of content and most influential nodes in network. The authors in Yadav et al. (2019) describe behavior of people having more than one social media profile. They share similar information on different social networks to gain popularity. It is available to all users in their list in a very less time. In the epidemic model, disease can also spread from one person to another in different community domains.

Bass model describes how the infection spread in community. Prediction and forecasting are very important. Forecasting and prediction are necessary and important for handling critical situation such as disease incurred through mass infection and flood. Our proposed model helps in finding the stage where the effect of infection is maximum. By knowing this, we apply recovery media to reduce the effect of infection. From the past data, we are able to get total recovery time, rate of spread, fraction of susceptible and infected population in the community. Early finding of results in epidemic model is similar to finding popularity in social media. Transmission of disease gradually attains peak point. To find out potential customers who adopt the information in social media is similar to find out total infected population in community. The epidemic model follows social network pattern to forecast the evolution of disease such as measles, mumps, typhoid fever, smallpox, common cold, chicken pox, diphtheria, influenza, the Severe Acute Respiratory Syndrome (SARS), dengue fever, repeated infection diseases and effect of computer virus in the field of network and internet technology. Our proposed model follows principles of epidemic models. Information propagation in online social network and real-world network is similar as disease spread in epidemic model. Our objective is to develop a novel diffusion model and apply it over real-world data set to elaborate its impact on the basis of different parameters. The proposed diffusion model would be capable of analyzing the rate of information or communication spread in generated network structure and predicting the outcome of the network after some time.

4 Network model and assumptions

For experimental analysis purpose, fifteenth-century Florentine marriages dataset is considered, which form networked relations among different communities (Fig. 2). This helps to analyze the impact of interaction among small-sized communities in real world. For analyzing behavior of the nodes in large-scale complex social graphs, our study is extended over Facebook and email networks (Fig. 13), as well. The social interactions between the different families or online societies inherently behave similar to spread of disease in an epidemic model. All the nodes refer to social entities, families, friends or people. An edge in general case refers to an association, which could be marriage in case of Florentine network, friendship in Facebook and communication in email network, respectively. All the isolated nodes and self-loop edges are eradicated from network dataset, so as to find the largest connected or giant component in the network modeled by the dataset. Our experimentation computes various parameters like average path length, average degree, betweenness centrality, diameter, sub-graphs and cluster coefficient (Jackson 2008). Further, our proposed algorithm is applied and extent of information diffusion is examined. Some nodes are designated as sources that initiate the communication followed by computation of other parameters, like threshold value, cascading, mass media nodes according to our requirement. Finally, information spread is analyzed from one node to another during the given time instant. Our study is also extended with case study with real-world complex interactions.

5 Proposed diffusion model

Our proposed diffusion model is based on Bass and SIS model, which is predominantly employed for simulating and analyzing the spread of diseases in a large-scale network. Such networks are often comprised of huge number of socially interacting people. Bass model forecast adoption of new idea, information, product, innovation into a social system for a long term. It determines the number of users who are likely to adopt a piece of information at certain instant of time. It estimates the growth rate of information in social network. Our model incorporates the uni-directional flow of Bass model and extends with reversible behavior that toggles between healthy and susceptible states.

Bass model refers to event which cannot go revert back such as seen a movie. The task is either happened or it will not happen. In the context of online social media, some persons will forward the information to another one, while some do not forward. Epidemic disease such as flu is either a person have effect of flu or do not have any effect of flu. But due to changing behavior of human, technology is also possible to adopt the information, product or innovation idea in the future. The intermediate state helps in taking decision to individual to adopt permanently or go back to its previous state. The vaccination or recovery media decides the time period to get recover from infection.

The proposed model is effective in predicting information diffusion in social media and spread of infectious disease in community to find out the optimal value of susceptible and infected number of people during infection time period. It mainly helps in monitoring and studying the effect of epidemic disease in the community, time taken till most of the nodes become infected and what fraction of total population gets infected. Prediction can help in taking decision to overcome or avoid from bad circumstances. Our proposed model can also be used to find out time and how many computer systems are affected by computer virus. In an organization, most of the systems are connected to each other via some topology.

In todays' scenario, most of the people are connected to each other via social relationship, personal attachment, neighborhood and other factors. People, designated as node, act as a source for information diffusion in social network or community. The nodes that are connected by a greater number of nodes transmit the information very fast. Some people permanently or temporarily adopt new technology or idea. The feature and usefulness of technology, idea or information help people in taking decision for acceptance or refusal. In the epidemic model, initially all people are in healthy state, but when they come in contact with infected person they susceptible to that infection. During this, some will revert back to healthy state and some get infected. In the context of social media, we can say that person finally adopt the information.

6 Model assumptions

Our proposed diffusion model operates with three fundamental states that correspond to healthy, susceptible and infected condition (Fig. 1).

- Healthy state refers to the present state of information possessed by an individual entity (modeled as node or agent). During any instant, a healthy node is liable to be infected by some rate if it is directly connected to at least one infected agent. Healthy nodes are those nodes, which initially do not have any diseases.
- Susceptible state refers to vulnerability of nodes toward certain infection. However, the nodes in susceptible state could become infected after some time due to connection to some other infected nodes (i.e., nodes already diffused with infection). Susceptible nodes are supposed to carry infectious disease and hence can transmit it to other healthy nodes. However, the susceptible nodes on medication could return back to normal or healthy state.

- Infected state refers to a permanency in infection that cannot be reverted back to any other state. The nodes once in the infected state are considered permanently diseased and cannot be recovered back to healthy state.

In context of social information network, infected state has special significance. It refers to the permanent adoption of certain event/product-specific information, while susceptible state refers to temporary adoption of ideas with the chances of reverting back to the previous state of information with certain probability. Such revert transition occurs in case, a customer uses the sample of a newly launched product, but ultimately rejects it due to poor service, improper functioning, etc., finally returning back to usage of previous product. Furthermore, if the customer is highly satisfied with the newly launched product, then he accepts the new product for future usage, thereby permanently discarding any product being presently used. Therefore, diffusion rate can be defined as the rate by which information spread and reach to the connected neighbors. Recovery rate can be defined as the rate by which a node can move back to its previous behavior. It refers to the rate with which it does not adopt any new behavior or information. This creates an intermediate stage in diffusion process and acts as decision making for accepting (i.e., adopting) and rejecting of certain behavior or information. The transitions might occur due to some external events or decisions made by the potential users (or nodes). Adoption rate is the rate at which a node permanently decides to adopt the new behavior or information. Once permanently adopted, reverting back to initial or previous state becomes infeasible. It is denoted as the final stage of adopting new behavior or incoming information. A node becomes diffused, only when it comes in contact or have connection(s) with infectious peers. Information tends to propagate across network through the links connecting the nodes. Initial propagation may escalate owing to the number of initial spreaders, which, however, becomes constant after reaching the steady state. It is assumed that except the initial propagators, none of the nodes are affected by the diffusion process. Moreover, as information diffusion begins, susceptible state offers decision making toward temporary or permanent transition.

In information diffusion process, the rate of spread caused by a node to disperse information, it will depend upon its popularity in the network. In the context of chosen real graphs, diffusion rate signifies the spread of interactivity among communities. Recovery rate is the rate of denial of social interaction. Adoption rate is the rate at

Fig. 1 Proposed diffusion model

A : Healthy/Initial
B : Susceptible/Intermediate/Vulnerable
C : Infected/Final/Influenced/Stable
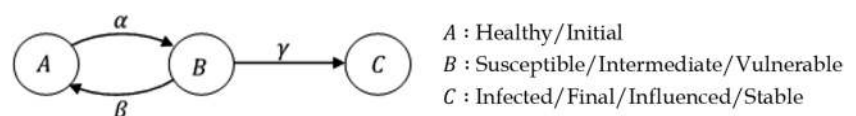
which associativity is established and become permanent. The communication in real world is a social relationship between humans.

During those times, the family with higher degree was considered more central and rose in power. Such nodes became key player in the networked society. The information diffuses in other families through marriage relationship. The interaction among the families increases with time as each family passes information to their neighbors. As an apparent visual interpretation from Fig. 2, Medici became popular among all families due to having higher degree and centrality value. In this context, diffusion probability is defined as the rate of interactivity among social communities. Recovery probability is defined as the rate of denial of social interaction. Further, adoption probability is the rate at which an associativity with respect to a social community is established and becomes permanent (i.e., changes cannot be reverted). The symbols used in the modeling and analysis of proposed diffusion model are highlighted in Tables 1 and 2, with their interpretation.

7 Model description and derivation

We assume that population is fixed over a certain period of time. During time period, no new communities are formed. Some of the network parameters are predetermined as, $p = 0.3$, $\lambda = 0.5$, $\beta = 0.4$. Further, value of α for modeling real data set is considered as $d_i/E[d]$. Also, we have:

$$E[d] = \frac{\sum_{i=1}^{i=n} d_i}{n} \tag{1}$$

Here, for real network graph $G_F(n, e)$, n and e represent number of communities and association or interaction. Now taking an example of a complete graph in which all nodes are connected to each other. For such a graph G of n nodes, degree of each node is $(n-1)$. So, $d_i = n - 1$
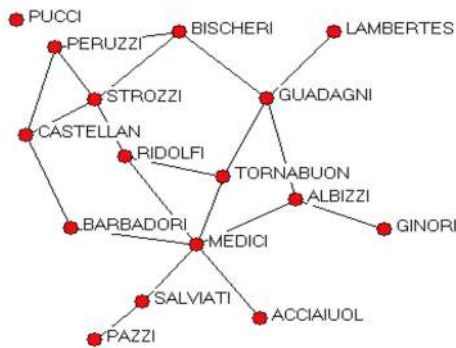
Fig. 2 Fifteenth-century Florentine marriages network graph

Table 1 Basic notations for diffusion model

Symbols	Interpretation
$G_F(\mathcal{N}, \mathcal{E})$	Available network/graph for diffusion analysis
$\mathcal{N}$	Vertex set of $G_F(\mathcal{N}, \mathcal{E})$ such that $|\mathcal{N}| = n$
$\mathcal{E}$	Edge set of $G_F(\mathcal{N}, \mathcal{E})$ such that $|\mathcal{E}| = e$
$G_C(\mathcal{N}_C, \mathcal{E}_C)$	Giant component graph
$\mathcal{N}_C$	Vertex set of $G_C(\mathcal{N}_C, \mathcal{E}_C)$ such that $|\mathcal{N}_C| = nc$
$\mathcal{E}_C$	Edge set of $G_C(\mathcal{N}_C, \mathcal{E}_C)$ such that $|\mathcal{E}_C| = ec$
n_i	i^{th} vertex such that $n_i \in \mathcal{N}$
e_{ij}	Edge connecting i^{th} and j^{th} vertex such that $e_{ij} \in \mathcal{E}$
α	Spread of interactivity (diffusion probability)
β	Recovery probability
γ	Adoption probability
λ	Control/decay rate of diffusion
t	Time variable
p	Initially interacting population (spontaneous influence rate)
$F_1(t)$	Fraction of susceptible population by time t
$F_2(t)$	Fraction of influenced people by time t
d_i	Degree of a node i
$E[d]$	Expected degree of all nodes/data set

$$E[D] = \frac{\sum_{i=1}^{i=n}(n-1)}{n} = \frac{n(n-1)}{n} = n - 1 \tag{2}$$

$$\alpha(t) = \frac{\frac{d_i}{E[d]}}{(1 + \lambda t)} \tag{3}$$

Hence, when $d_i = E[d]$ and $\alpha(t)$ becomes equal to $\frac{1}{1+\lambda t}$. Finally, the rate of diffusion for network $G_F(n, e)$ w.r.t time t can be dynamically formulated by Eq. (3):

(1) Transition from state A to B and B to A

Transition occurs initially, and intermediate states can be modeled using SIS model to find out the rate of change susceptible population with respect to time, as in the following equation:

Table 2 Composite notations for diffusion

Symbols	Interpretation
$\frac{dF_1(t)}{dt}$	Rate of susceptible population at time t
$\frac{dF_2(t)}{dt}$	Rate of adoption with respect to time t (state transition)
$1 - F_2(t)$	Fraction of people those are not in recovered state
$1 - \gamma$	Probability of recovering (refrain from final state transition)

$$\frac{dF_1(t)}{dt} = \big(1 - F_1(t)\big)\big(\alpha(t)F_1(t) + p\big) - \beta F_1(t) \tag{4}$$

Nodes in the network eventually get suscepted by some rate and get influenced by another rate. Hence, a steady state is observed, which can be obtained by differentiating $F_1(t)$ with respect to t and equating it to zero, i.e., $\frac{dF_1(t)}{dt} = 0$ represents the change in susceptibility rate parameter w.r.t time is constant. Also, new interaction is being balanced by the denial rate. This id is explained in the following equations:

$$\big(1 - F_1(t)\big)\big(\alpha(t)F_1(t) + p\big) - \beta F_1(t) = 0 \tag{5}$$

$$F_1(t)^2 \alpha(t) + F_1(t)(p + \beta - \alpha(t)) - p = 0 \tag{6}$$

On solving quadratic Eq. (6), we obtain:

$$F_1(t) = \frac{((\alpha(t) - \beta) - p) + \sqrt{(p + \beta - \alpha(t))^2 + 4\alpha(t)p}}{2\alpha(t)} \tag{7}$$

The value of initially interacting population (spontaneous influence rate) is negligible. Its value does not affect our analysis result. For better results, we have neglected p. If we drop the value of p, i.e., spontaneous rate of acquiring information, from Eq. (7), the following relation is obtained as, $F_1(t) = 0$ and $F_1(t) = 1 - \beta/\alpha(t)$.

$1 - \left(\frac{\beta}{\alpha(t)}\right) > 0$ means that recovery rate is more than infection rate; however, if $\beta < \alpha(t)$: this implies that a node is vulnerable toward catching infection rapidly than it recovers. This will lead to $F_1(t) > 0$.

(ii)   Transition from state B to C

Now, using the Bass model, in order to interpret the transition from intermediate to final state, we get:

$$\frac{dF_2(t)}{dt} = \big[F_1(t) + (1 - \beta)F_2(t)\big]\big[1 - F_2(t)\big] \tag{8}$$

Let $F_2(t) = x$, $F_1(t) = q$ and $(1 - \beta) = \gamma$. Now on solving the above equation, we get:

$$\frac{dx}{dt} = (q + \gamma x)(1 - x) \Rightarrow \frac{dx}{(q + \gamma x)(1 - x)} = dt$$

$$\Rightarrow \left\{ \frac{\gamma}{(q + \gamma)(q + \gamma x)} + \frac{1}{(q + \gamma)(1 - x)} \right\} dx = dt \tag{9}$$

On integrating Eq. (9), the fraction of population getting diffused by time t is derived as in Eq. (10):

$$\frac{1}{(q + \gamma)} \int \left\{ \frac{\gamma}{(q + \gamma x)} + \frac{1}{(1 - x)} \right\} dx = \int dt$$

$$\frac{1}{(q + \gamma)} \{\ln(q + \gamma x) - \ln(1 - x)\} = t$$

$$\frac{1}{(q + \gamma)} \left\{ \ln\left(\frac{q + \gamma x}{1 - x}\right) \right\} = t$$

$$ln((q + \gamma x)/(1 - x)) = t(q + \gamma)$$

$$\frac{q + \gamma x}{1 - x} = e^{t(q+\gamma)} \Rightarrow x = 1 - (q + \gamma x)e^{-t(q+\gamma)}$$

$$x + \gamma x e^{-t(q+\gamma)} = 1 - q e^{-t(q+\gamma)} \Rightarrow x = \frac{1 - q e^{-t(q+\gamma)}}{1 + \gamma e^{-t(q+\gamma)}}$$

$$F_2(t) = \frac{1 - F_1(t)e^{-(F_1(t)+\gamma)t}}{1 + \gamma e^{-(F_1(t)+\gamma)t}} \tag{10}$$

Algorithm 1 depicts the procedure used in the modeling and implementing the diffusion process. Giant component refers to unique largest connected component. Self-loop and isolated degree nodes have no relevance because they are no connected to any other node or vertex. Algorithm 2 in appendix section highlights the procedure for identifying the giant component in the graph.

Algorithm 1: Information Diffusion Process

Input: $G_F(\mathcal{N}, \mathcal{E}), \beta, p, \lambda, t$
Procedure:
1. $t :=$ Simulation time
2. $G_F(\mathcal{N}, \mathcal{E}) :=$ input network graph taken
3. $G_C(\mathcal{N}_C, \mathcal{E}_C) :=$ Giant_Component $(G_F(\mathcal{N}, \mathcal{E}))$
4.   while $t \leq T$ do:
5. $\alpha(t) := 1/(1 + \lambda t)$
6.   calculate $F_1(t)$
7.   for $F_1(t) < 1$ and $t \leq T$ do:
8.   calculate $F_2(t)$
9.   end-for
10. end-while

Output: $F_1(t), F_2(t)$

In diffusion process, propagated information may be adopted by some nodes in the network. Nodes already acquiring information may prefer to propagate further at a rate defined by its own popularity. The derived parameter $F_2(t)$ signifies the extent to which information would spread in the network after t units of functional lifetime of the social network is elapsed.

8 Result analysis

For our experimentation, Florentine network is used that typically highlights a small community (Fig. 2). It highlights the network generated from the dataset of marriage ties among renaissance Florentine families. The relations are marriage alliances between the families that illustrate social interaction and association. It is the network of marriages, establishing inter-community links among families in the Florence during 1430s. Each node represents a family, and links represent marriage relationship among families. The marriage is considered as pathways of communicating information, sharing business deals and making political dealings.

Our research considers the rate of spread of information or disease will depend on degree centrality value, which is the measure of the size of direct neighborhood. We have considered nodes belonging to three categories, the one having relatively with minimum, average and maximum centrality value to analyze the behaviors as initiator nodes in the network for information diffusion. On the basis of centrality measurement, nodes are divided into three categories, namely maximum (Medici node), average (Albizzi, etc.) and minimum (Lambertes, etc.), respectively. Different initiator nodes in the diffusion process assist in analyzing the manner in which interaction or information flow is affected in each case in the network (Fig. 2). Our study reveals the impact of information diffusion rate over the network. The maximum degree node diffuses the information very fast as compared to average and minimum degree node. Higher degree nodes connect various other nodes so that most of the communication or interaction is carried out by these nodes in the network. Nodes with higher degree and centrality value are beneficial for triggering the diffusion process.
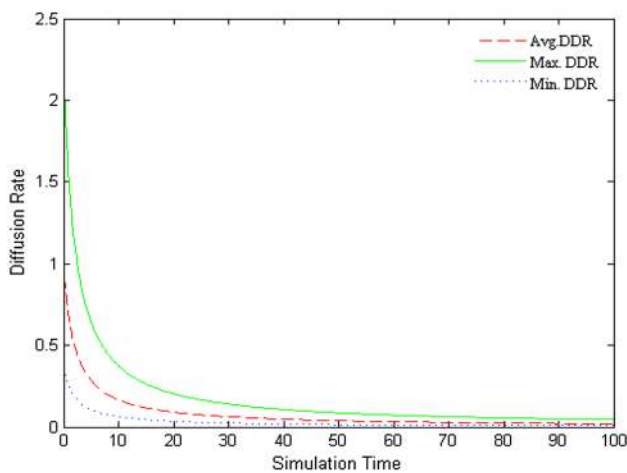
Fraction of population in each case responsible for diffusion is computed at certain regular time instants. Figure 3 illustrates how information or communication would spread in social community network by choosing nodes belonging to each category. Initially, in the context of graph highlighted as Fig. 2, the communication or interactions is apparently higher in the network for exchanging new ideas, events or maybe some valuable information.

This is similar to the case of epidemic model, in which the infection among the population increases as time advances. It is also affected by the number of people who are initially infected. Rate of spread of infection is directly dependent on initiator node and average degree. Each community wants to receive new information in the network. After some time, the interaction among the communities decreases, possibly because members have already interacted with each other and transited from the initial phase. Finally, interaction among members, hence in the context of epidemic model, the effect of infection becomes stable. Figure 4 illustrates the fraction of population in the network who has eventually interacted with different communities to adopt new information or idea. Individuals are in the susceptible state, i.e., people have used some technology/product, i.e. gained information, and they desire to revert back to their previous state. Such a reversion in transition occurs owing to dissatisfaction with the results of new product or comparatively better experience with some previous product. As in epidemic model people are said to be in susceptible state, when they are likely to become infected on getting in contact of some infectious agent(s); otherwise might have chances to return back to healthy state. Therefore, result interprets that the interaction among communities about adopting new information or behavior decreases exponentially with time. In the starting phase of disease transmission, a large number of populations are infected by epidemic disease such as
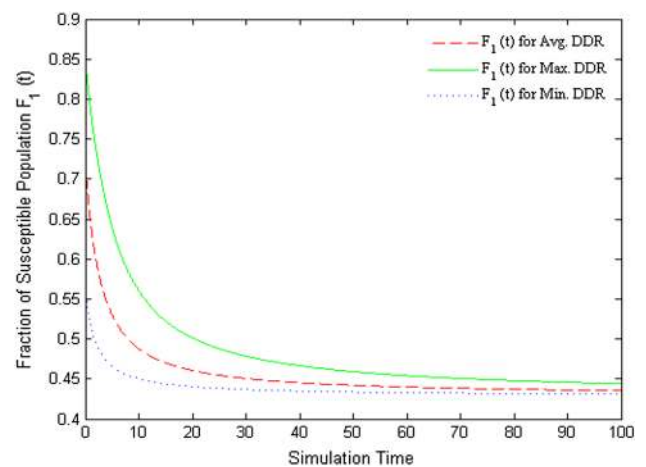

Fig. 3 Diffusion rate in social community

Fig. 4 Fraction of susceptible population in network

Fig. 5 Fraction of adopted population in network

Fig. 6 Diffusion rate (changing degree & simulation time)
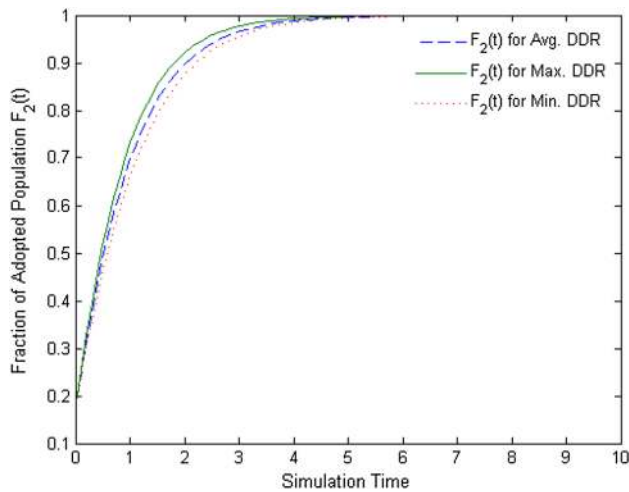
common cold. Due to internal immunity power of human, some people will get infected, while some are not infected. For the short duration of time, more people are susceptible to that particular disease.

Figure 5 illustrates fraction of population in the network who finally adopted some information or conception (maybe regarding some newly launched product, etc.). As in case of epidemic model, people have chances to recover from diseased state even after initially getting infected. The adoption of information becomes permanent to members in the community when they cannot revert back to their previous state. The effective number of recovered people (in other words, population finally adopting the new information) increases exponentially with time. Interaction among communities increases resulting in adopting of new information originating from initiator nodes. This is similar as in the case of epidemic model, where fraction of recovered population increases as certain recovery media or vaccination is provided. Initially, impact of recovery media over infectious population is less, which eventually gets stable over time. Once person get infected by some infectious disease, it has some recovery time duration. For that time, instant person will recover from infection and come in recovery zone.

Figure 6 illustrates information diffusion rate dynamically varying with degree and simulation time in the network. At each time, instant information diffusion depends mainly on selected nodes, called initiator nodes. Such nodes could be the ones with high centrality measure in the network.

In the context of epidemic model, when a person comes in contact of infectious person, then it also will become infected. This process continues till a large portion of community is infected. At this time, instant rate of infection is very high and rate of change in recovery is less.

Figure 7 illustrates fraction of susceptible population in network within a given time frame with a chosen rate of
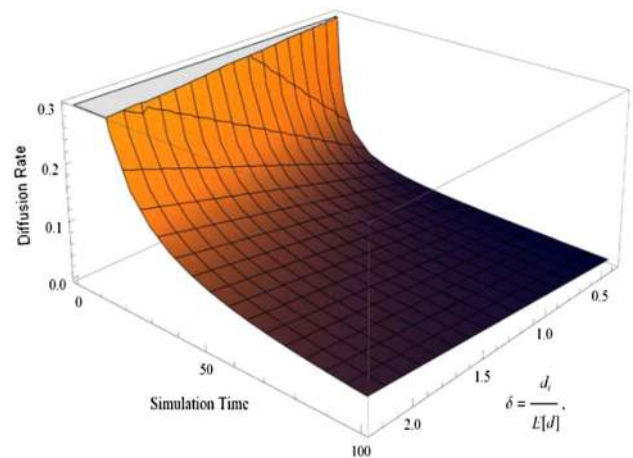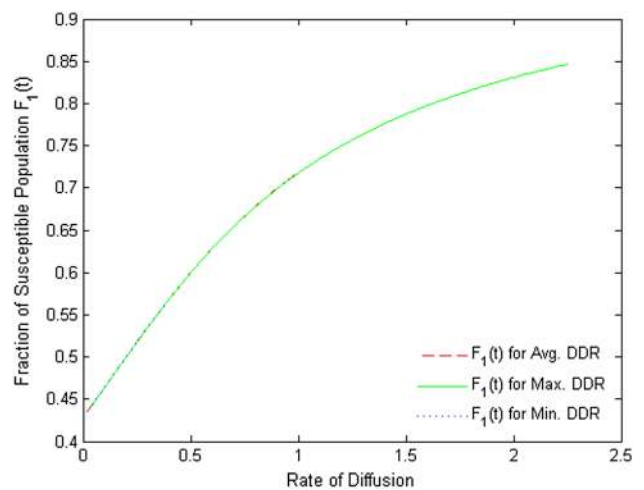

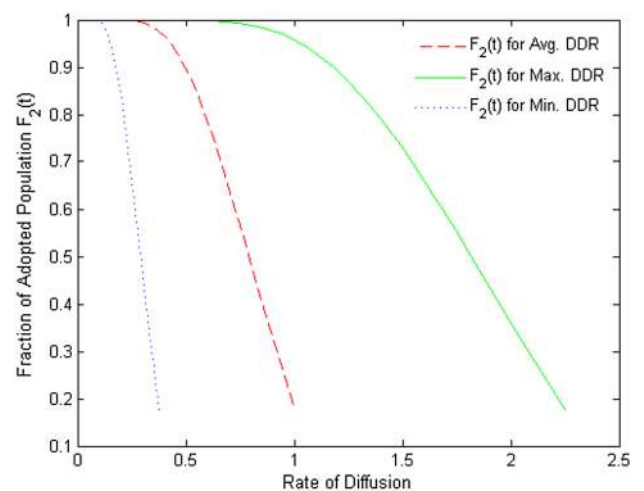**Fig. 7** Fraction of susceptible population with diffusion rate

Fig. 8 Fraction of adopted population with diffusion rate
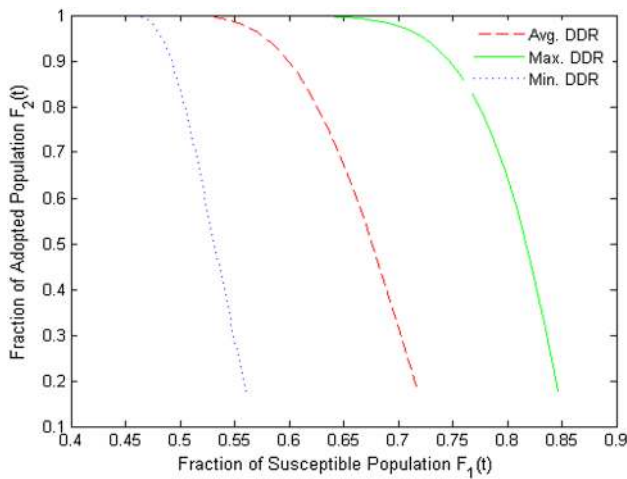
Fig. 9 Fraction of adopted population with susceptible population
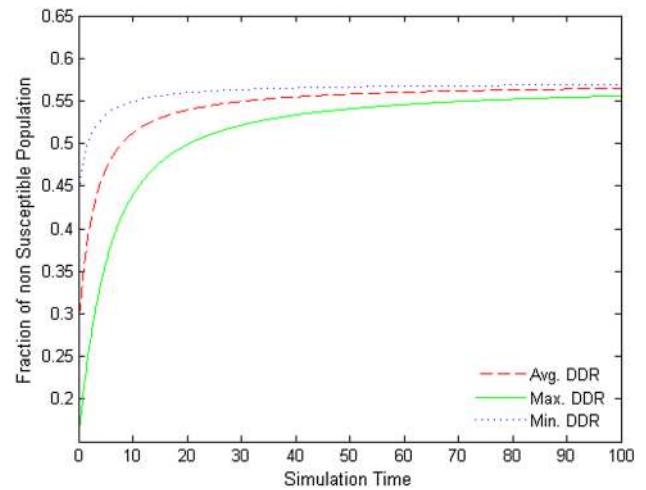

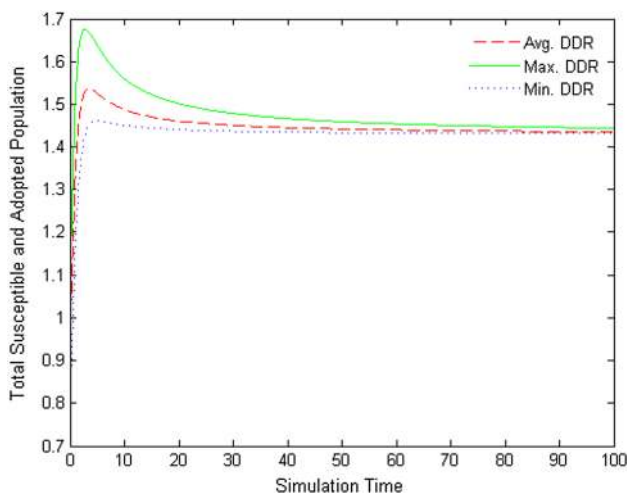**Fig. 11** Fraction of non-susceptible population

Fig. 10 Fraction of adopted and susceptible population
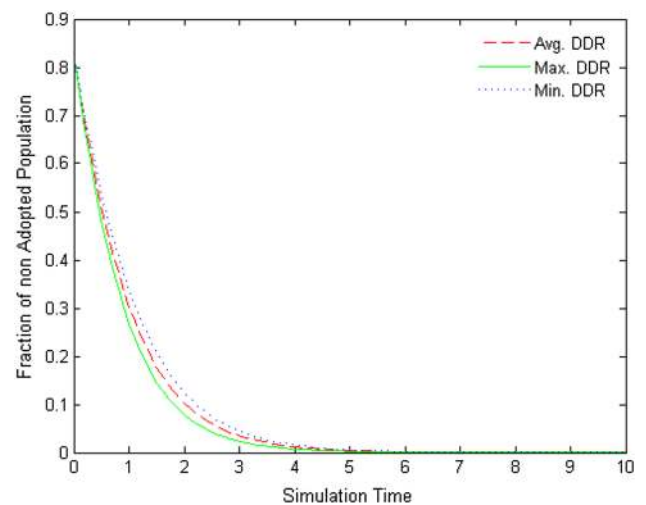

Fig. 12 Fraction of non-adopted population

diffusion. Initially, the susceptible population is lower, but the graphical trend gradually shows an exponential behavior. The communication among families increases exponentially as network connections become denser. Figure 8 illustrates fraction of adopted population in the network. The trend is initially high as the information is new for all users that gradually lowers down. New information is acquired by most of people in network like as in epidemic model most of the person are vulnerable to disease initially.

Figure 9 highlights the adopted and susceptible population in the network. The adopted population decreases with respect to susceptible population. The interest of users about information decreases as time advances. Figure 10 illustrates that susceptible and adopted population in the network have peak value in the starting phase of diffusion. After some time period, total adopted population in the

network becomes constant. During infection time domain, most of the people are influenced by infection. Initially, diffusion increases rapidly, i.e., the rate with which new information about an innovation shared among users is high. But after some period of time, due to better usability of new product, affluence of the people increases toward the better one. Figure 11 shows the fraction of non-susceptible population in network that increases and finally becomes stable. It is evident from the graph that, as the diffusion process dominates, the chances of moving toward final or influenced state becomes constant. Transmission of epidemic diseases continue from one person to another in community. The contacts/ relationships/ links act as spreader medium for disease. Higher degree node acts as initiator node and continues till the last node in network. Figure 12 illustrates the fraction of population that do not

choose to adopt the recently launched product. Initially, less number of nodes are in diffused state. However, when a new node comes in contact of a diffused node, the probability of getting influenced becomes high.
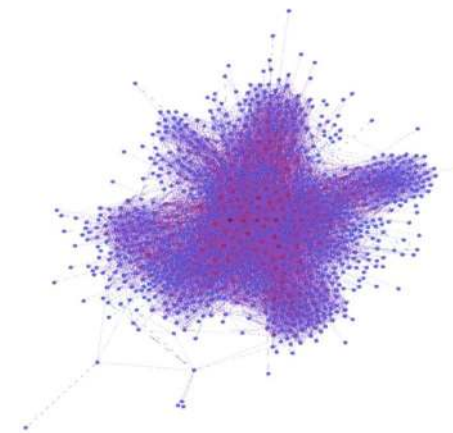
9 Large-scale complex social networks: a case study

This section includes an extended case study over large-scale real-world complex social networks from real-world domain. Figure 13a highlights Facebook network that shows social circles formed by friends' profiles and interactions (McAuley and Leskovec 2012). The dataset consists of 4039 nodes and 88,234 edges with average degree being 21.84. From Facebook online social network graph, it is clear that links/relationships among people are higher. Figure 13b refers to social graph generated from email data of a European research institute (Yin et al. 2017; Leskovec et al. 2007). The graph is composed of anonymized information regarding profiles interacting with each other within the institute mailing network. A link between two nodes in the email graph signifies existence of at least one communication through mail. Experimentation reveals an average degree of 25.44 with 1005 nodes and 25,571 associations. Both the graphs in Fig. 13 are built using Gephi on the datasets of the respective graphs being extracted from Stanford network of datasets, SNAP. A comparative analysis is performed between the small-scale real graph of marriage ties highlighted in Fig. 2 with the complex graphs of Fig. 13.

The small-scale real graph and large-scale real-world complex social networks have similar characteristics. They follow the same diffusion pattern. As information propagates from one node to another in social network like as infection spread in the community. Figure 14 describes the manner that information or communication spreads in real and online social network. The rate information diffusion in online social network decreases with advancement in time interval that follows similar nature as that of epidemic model in which infection in a community decreases with respect to time. The rate of information diffusion for Florentine network is comparatively higher than Facebook and email network. It is clear that information propagates very fast in online social network.

Figure 15 illustrates how the communication or adoption of ideas occurs in social network. Initially, people adopt information on the basis of some features, advantage or popularity of information. As the time interval increases, number of people forwards the information to others at higher rate so that a number of people who are in intermediate state decrease with time. Similarly, in the epidemic model some people will become infected and some revert back
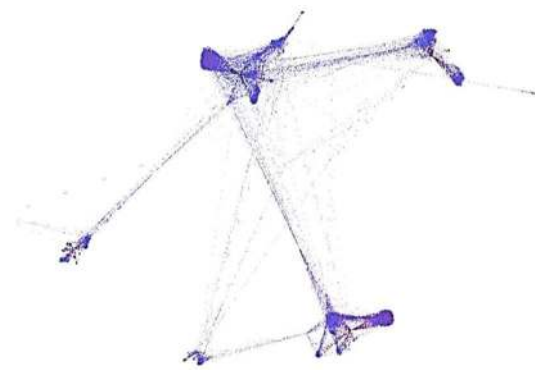
(a) Facebook Ego Network

(b) Email Network from European Research Institute

Fig. 13 Large-scale complex social networks
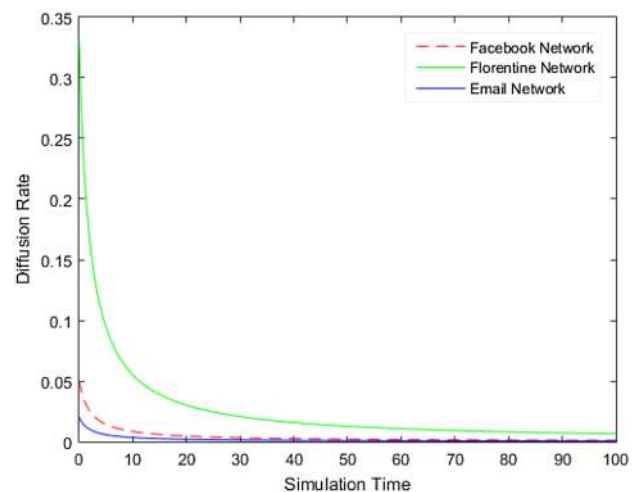

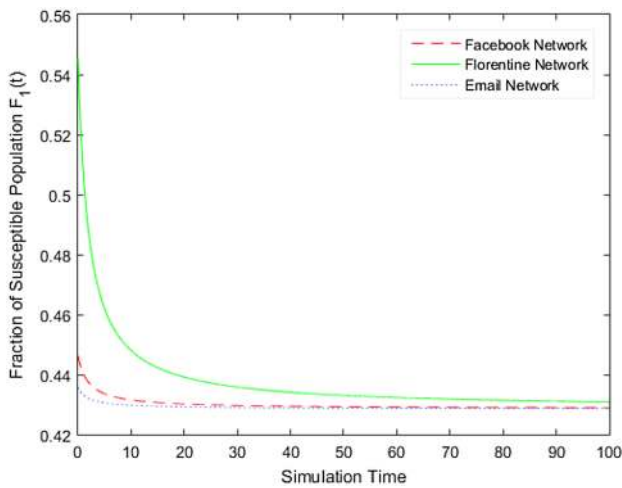**Fig. 14** Diffusion rate in social network

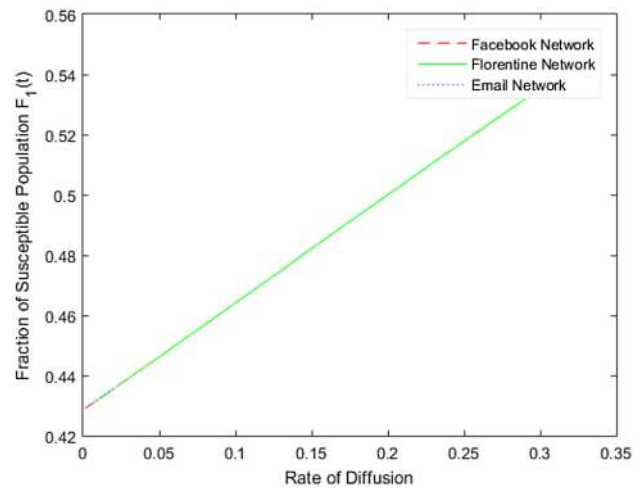Fig. 15 Fraction of susceptible population in network

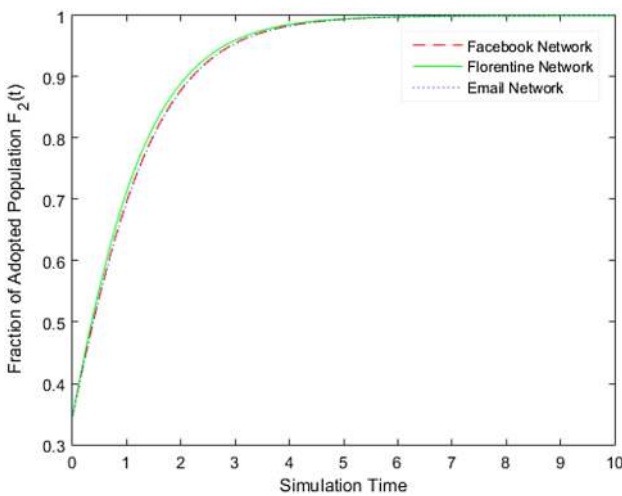Fig. 17 Fraction of susceptible population with diffusion rate

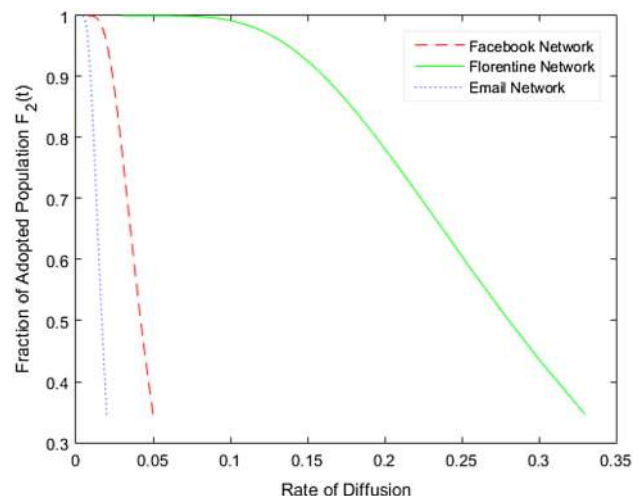Fig. 16 Fraction of adopted population in network

Fig. 18 Fraction of adopted population with diffusion rate

to its initial state. The number of susceptible populations for Florentine network is apparently higher as compared to email and Facebook network.

Figure 16 illustrates people who finally adopted the behavior or information. After certain time interval, it follows exponential nature and will become constant. A number of adopted population in Florentine network are higher in comparison with email and Facebook network with respect to time because the communication among nodes in small size graph occurs at a higher rate.

Figure 17 describes the number of susceptible people in the social network with in a time frame with a chosen rate of diffusion. It follows approximately same linear rate of susceptible population in all networks considering diffusion rate. The trend for all networks is overlapped, exhibiting a similar behavior. The Facebook, Florentine and email

network have similar characteristics. Information diffusion in real graph follows the same trend as in social network.

Figure 18 illustrates the trend of adoption for information that gradually decreases with respect to the rate of diffusion, due to decrement of information popularity in network. Initially, it is high but eventually decreases with information spread rate. Total adopted population in Florentine network is found to be comparatively higher than other networks. Figure 19 describes the relation between adopted and susceptible people in the social network. The adopted population with respect to susceptible population in Florentine network is lesser in comparison with Facebook and email network. Rate of information diffusion is associated with the degree of network.

Figure 20 illustrates that adoption of information or diffusion attains higher peak value for Florentine network than
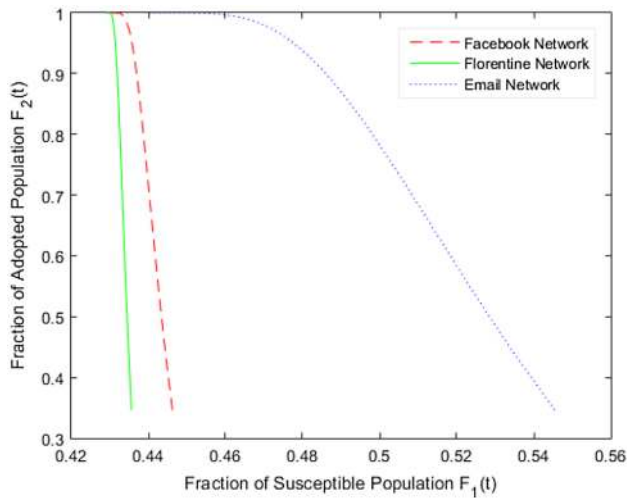
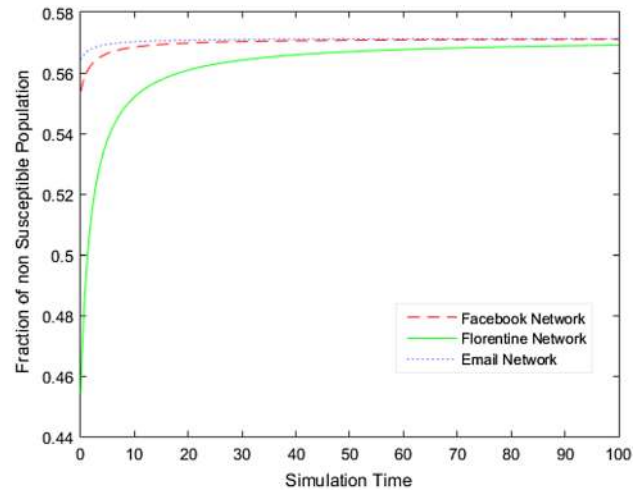**Fig. 19** Fraction of adopted population with susceptible population

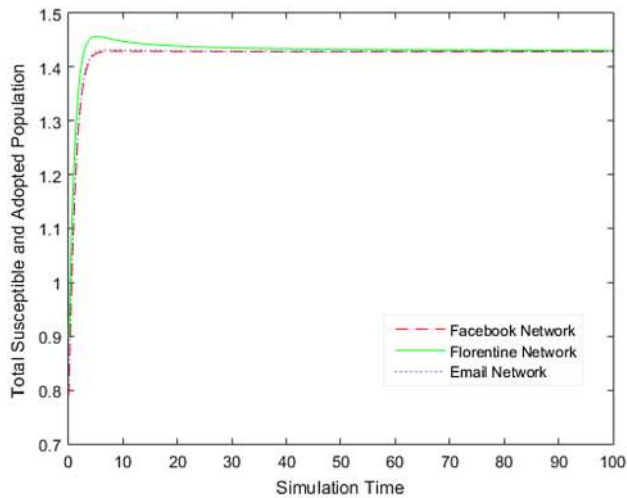**Fig. 21** Fraction of non-susceptible population

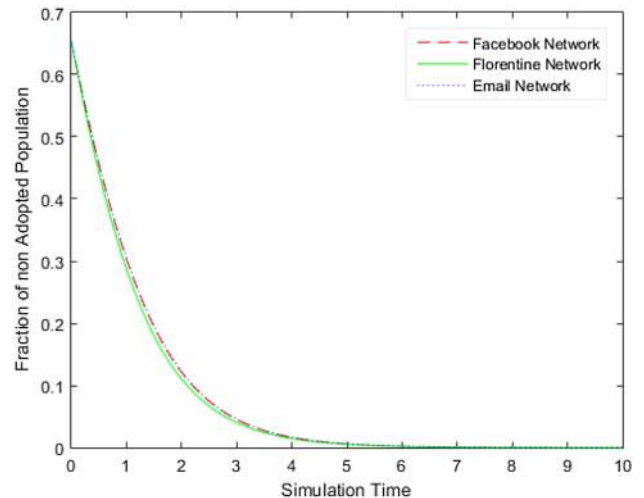**Fig. 20** Fraction of adopted and susceptible population

Fig. 22 Fraction of non-adopted population

other networks in the starting phase, after some time it will become constant. Figure 21 describes that non-susceptible population are higher for online social networks as compared to Florentine network. Figure 22 describes that total non-adopting population for Florentine network is less as compared to email and Facebook network. Therefore, the resultant outcome validates the authenticity of our proposed diffusion model and novel algorithmic approach as the network size and density spans from smaller to larger number of nodes.

10 Conclusion and future scope

Diffusion in social network can be defined as the spread of information from one node by sharing social connections between them. It provides analysis of pattern and rate of spread of new behavior or ideas about product launching events, innovating idea, etc. Such networks may span over several social entities or nodes; therefore, a careful choice of network parameters and their initial values plays a vital role in diffusion of information. Any diffusion process requires certain nodes those act central in the network. Nodes having high centrality value normally act rapid initiators for spreading information to connected peers. In our research, we have proposed model for diffusion, motivated from

epidemic model. This assisted in modeling social behavior within close communities. In order to designate central nodes, degree centrality, being the most realistic measure of neighborhood size, has been taken into consideration.

Our proposed diffusion model computes parameter $F_2(t)$ to compute the rate at which information would diffuse in the social network after t time units is elapsed. One of the predominant application of our proposed approach would assist in analyzing the progress in viral marketing toward successfully launching market propagandas. Our model is effective in predicting/forecasting information diffusion in social media and spread of infectious disease in community to find out optimal value of susceptible and infected number of person during infection time period. Prediction/forecasting can help in taking decision to overcome or avoid from bad circumstances. Our proposed model can also be used to find out time and how many computer systems are affected by computer virus. In an organization, most of the systems are connected to each other via some topology. The proposed model can be extended by incorporating more intermediate states so as to model the realistic nature of the social networks. Increase in number of states results in additional transitions abiding some constraints. Also, weighted centrality function could be designed to reveal nodes that offer better selection of initiators to participate in the diffusion process. Hence, different versions of improved models could be designed to capture the trend in social interaction networks.

Appendix

This section introduces the execution procedure of identifying giant component in the graph. The algorithm 2 reveals the giant connected sub-graph, $G_C(\mathcal{N}_C, \mathcal{E}_C)$ for the initial graph $G_F(\mathcal{N}, \mathcal{E})$.

Algorithm 2: Finding Giant Component

Input: $G_F(\mathcal{N}, \mathcal{E})$
Procedure:
1. $\mathcal{N}_C := \mathcal{N}$
2. $\mathcal{E}_C := \mathcal{E}$
3.   for each $e_{ij} \in \mathcal{E}$ do:
4.   if $i = j$ then:
5. $\mathcal{E}_C := \mathcal{E}_C - \{e_{ij}\}$ // Remove self-loop
6.   end-if
7.   end-for
8.   for each $n_i \in \mathcal{N}$ do:
9.   if $d_i = 0$ then:
10. $\mathcal{N}_C := \mathcal{N}_C - \{n_i\}$ // Remove isolated nodes
11.   end-if
12.   end-for
13. $G_C(\mathcal{N}_C, \mathcal{E}_C) \leftarrow G_F(\mathcal{N}, \mathcal{E})$

Output: $G_C(\mathcal{N}_C, \mathcal{E}_C)$

References

Aditya Prakash B, Vreeken J, Faloutsos C (2014) Efficiently spotting the starting points of an epidemic in a large graph. Knowl Inf Syst 38(1):35–59

Bo X, Liu L (2010) Information diffusion through online social networks. In: IEEE international conference on emergency management and management sciences (ICEMMS), Beijing, China, Aug 2010

Broecheler M, Shakarian P, Subrahmanian VS (2010) A scalable framework for modeling competitive diffusion in social networks. In: IEEE international conference on social computing, privacy, security, risk and trust, Minneapolis MN, Aug 2010.

Cheng J, Kleinberg J, Leskovec J, Liben-Nowell D, State B, Subbian K, Adamic L (2018) Do Diffusion protocols govern cascade growth?. arXiv preprint arXiv:1805.07368, May 2018.

Cho E, Myers SA, Leskovec J (2017) Friendship and mobility: user movement in location-based social networks. In: Proceedings of the 17th ACM SIGKDD international conference on knowledge discovery & data mining, pp 1082–1090, Aug 2011

Di Giamberardino P, Compagnucci L, De Giorgi C, Iacoviello D (2017) Modeling the effects of prevention and early diagnosis on HIV/AIDS infection diffusion. IEEE Trans Syst Man Cybern Syst 99:1–12

Doo M, Liu L (2014) Probabilistic diffusion of social influence with incentives. IEEE Trans Serv Comput 7(3):387–400

Fatima I, Fahim M, Lee Y-K, Lee S (2013) MODM: multi-objective diffusion model for dynamic social networks using evolutionary algorithm. Springer, New York

Fouad MR, Elbassioni K, Bertino E (2012) Modeling the risk & amp: utility of information sharing in social networks. In: IEEE international conference on privacy, security, risk and trust (PASSAT) and social computing (SocialCom), Amsterdam, Netherlands, pp 441–450, Sept 2012

Jackson MO (2008) Social and economics networks. Princeton University Press, Princeton, pp 1–520

Jiang Y, Jiang JC (2015) Diffusion in social networks: a multiagent perspective. IEEE Trans Syst Man Cybern Syst 45(2):198–213

Jiang C, Chen Y, Liu KJR (2014) Evolutionary dynamics of information diffusion over social networks. IEEE Trans Signal Process 62(17):4573–4586

Kandhway K, Kuri J (2017) Using Node centrality and optimal control to maximize information diffusion in social networks. IEEE Trans Syst Man Cybern Syst 47(7):1099–1110

Kempe D, Kleinberg J, Tardos E (2003) Maximizing the spread of influence through a social network. In: Proceedings of the 9th ACM SIGKDD international conference on knowledge discovery & data mining, pp 137–146, Aug 2003

Krishna Kumar KP , Geethakumari G (2013) Information diffusion model for spread of misinformation in online social networks. In: IEEE international conference on advances in computing, communications and informatics (ICACCI), Mysore, India, pp 1172–1177, Aug. 2013

Kumar P, Sinha A (2016) Real-time analysis and visualization of online social media dynamics. In: IEEE international conference on next generation computing technologies, Dehradun, pp 362–367, Oct 2016

Kuo TT, Hung SC, Lin WS, Lin SD, Peng TC, Shih CC (2011) Assessing the quality of diffusion models using real-world social network data. In: IEEE international conference on technologies and applications of artificial intelligence (TAAI), Chung-Li, China, pp 200–205, Nov 2011

Lamprier S, Bourigault S, Gallinari P (2016) Influence learning for cascade diffusion models: focus on partial orders of infections. Soc Netw Anal Min 6(1):93

Leskovec J, Kleinberg J, Faloutsos C (2007) Graph Evolution: Densification and Shrinking Diameters. ACM Trans Knowl Discov Data 1(1):1-es

Li Y, Chen Z (2007) Diffusion of Innovations in a Small World Network. In: IEEE international conference on wireless communications, networking and mobile computing, Shanghai, China, Sept 2007.

Mahdi K, Torabi S, Safar M (2010) Diffusion and reverse diffusion processes in social networks: analysis using the degree of diffusion. In: IEEE international conference on Ubi-media computing (U-Media), Jinhua, July 2010

McAuley J, Leskovec L (2012) Learning to discover social circles in ego networks. In: NIPS

Mini U, Jacob P (2014) Information diffusion through social media-analysing brand promotion through social network analysis tools-a business case Study. In: IEEE conference on IT business, industry and government (CSIBIG), pp. 1–4, Indore, India, March 2014

Niu J, Huang S, Shu L, Stojmenovic I (2013) Activities information diffusion in chinese largest recommendation social network: patterns and generative model. In: IEEE international conference on global communications (GLOBECOM), Atlanta GA, pp 3083–3088, Dec 2013

Rosa D, Giua A (2013) A non-progressive model of innovation diffusion in social networks. In: IEEE conference on decision and control, florence, Italy, Dec 2013

Sahal D (1979) The temporal and spatial aspects of diffusion of technology. IEEE Trans Syst Man Cybern Syst 9(12):829–839

Sato Y, Shimokawa H, Ata S, Oka I (2012) Towards social networking: a proof-of-concept model. In: IEEE international conference on privacy, security, risk and trust (PASSAT) and social computing (SocialCom), Amsterdam, Netherlands, pp 526–531, Sept 2012

Shi H-w, Zhao A-m, Th H, Bressers A, Deboer C (2011) The optimization of the knowledge diffusion processes and an innovation service relationship model in the context of social networks. In: IEEE international conference on industrial engineering and engineering management (IE&EM), Changchun, China, pp 1418–1423, Sept 2011

European Union (2010) Social network overview: current trends and research challenges, Nov 2010.

Wang F, Wang H, Xu K, Wu J, Jia X (2013) Characterizing information diffusion in online social networks with linear diffusive model. In: IEEE international conference on distributed computing systems (ICDCS), Philadelphia PA, pp 307–316, July 2013

Wang Y, Vasilakos AV, Ma J, Xiong N (2014) On studying the impact of uncertainty on behavior diffusion in social networks. IEEE Trans Syst Man Cybern Syst 45(2):185–197

Yadav S, Sinha A, Kumar P (2009) Multi-attribute identity resolution for online social network. SN Applied Sciences 1:1653

Yagan O, Qian D, Zhang J, Cochran D (2012) Information diffusion in overlaying social-physical networks. In: IEEE annual conference on information sciences and systems (CISS), Princeton NJ, pp 1–6, March 2012

Yin H, Benson AR, Leskovec J, Gleich DF (2017) Local higher-order graph clustering. In: Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining

Zaffar MA, Kumar RL, Zhao K (2014) Impact of interorganizational relationships on technology diffusion: an agent-based simulation modeling approach. IEEE Trans Eng Manag 61(1):68–79

Zhang J, Moura JMF (2014) Diffusion in social networks as sis epidemics: beyond full mixing and complete graphs. IEEE J Select Top Signal Process 8(4):537–551