

# Social influence and spread dynamics in social networks

Xiaolong ZHENG<sup>1</sup>, Yongguang ZHONG<sup>2</sup>, Daniel ZENG<sup>1</sup>, Fei-Yue WANG (✉)<sup>1</sup>

1 State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

2 Department of Management Science and Engineering, Qingdao University, Qingdao 266071, China

© Higher Education Press and Springer-Verlag Berlin Heidelberg 2012

**Abstract** Social networks often serve as a critical medium for information dissemination, diffusion of epidemics, and spread of behavior, by shared activities or similarities between individuals. Recently, we have witnessed an explosion of interest in studying social influence and spread dynamics in social networks. To date, relatively little material has been provided on a comprehensive review in this field. This brief survey addresses this issue. We present the current significant empirical studies on real social systems, including network construction methods, measures of network, and newly empirical results. We then provide a concise description of some related social models from both macro- and micro-level perspectives. Due to the difficulties in combining real data and simulation data for verifying and validating real social systems, we further emphasize the current research results of computational experiments. We hope this paper can provide researchers significant insights into better understanding the characteristics of personal influence and spread patterns in large-scale social systems.

**Keywords** social networks, spread dynamics, social influence, computational experiment

## 1 Introduction

Social networks are graphs of people and their relationships, such as Email communication, scientific collaboration, social tagging, diffusion of innovation, and even epidemic

spreading. In these graphs, individuals and their relationships are represented by the nodes and links of the network, respectively. We can understand the structure and dynamic evolution of social systems by analyzing these graphs. In the past few years, this field has attracted a great deal of interest [1–6]. Since social influence and information spread in social systems is easily reproduced in such networks, this field has been recently addressed in the literature [7–11], as indicated by the large and rapidly increasing number of papers devoted to it.

Traditional research on social influence and spread dynamics in social networks has been constrained in accuracy, breadth, and depth due to its reliance on self-report data [12,13]. These studies generally involve both a small number of persons and some static time points. The detailed behavior of many physiological and psychological processes is still largely unknown [13]. As a result, the analytical results have been limited to examining small, well-bounded populations, involving a small number of snapshots of interaction patterns [14].

However, this situation has been changing with the rapid growth of Web 2.0 and mobile communication technologies. Recently, a large number of mobile-based applications and online social systems such as Facebook, Myspace, and Twitter have emerged. In these systems, social relationships between users play an important role in dictating their behavior. Users can induce their friends to behave in a similar way explicitly or implicitly via their social influence, e.g., sharing interesting ideas and even rumors. This information can spread through a network just like an infectious disease. Identifying social influence and understanding the spread patterns

Received October 19, 2011; accepted February 7, 2012

E-mail: feiyue.wang@ia.ac.cn

in social networks are of tremendous interest from an empirical analysis point of view [4,9,12].

Based on these large-scale data sets, researchers have discovered surprising social phenomena including the small-world effect and scale-free property. These results are different to classical hypotheses. The tendency to move toward the formulation of simplified models and their quantitative analysis has been instrumental in this change. The classical models used by scientists to describe social influence and spread dynamics in social networks are considered too simplified to describe any real situation [13]. Therefore, a crucial step in this perspective is the comparison with empirical data which should be primarily intended as an investigation into whether the trends seen in real data are compatible to plausible macroscopic/microscopic models of the networks, and whether they are self-consistent or require additional elements. This has inspired scientists in this field to develop new reasonable models.

Traditionally, researchers validate these models through simple simulation. However, simple simulation cannot solve all problems due to the complexity of real-world social systems. As such, computational experimentation is now being applied more often. This emerging approach can be utilized to deal with validating the complex social systems.

Despite various reviews on social networks, relatively little material has provided a comprehensive survey on the network presentation, characteristic measurement, social modeling, and computational experiments for identifying the influence between behaviors and spread patterns in large-scale social systems. This brief survey addresses precisely this issue. The framework of this paper is presented in Table 1. Section 2 describes and compares some empirical studies of social influence and spread dynamics in social networks, including network construction, measures of network, and some significant empirical results. In Section 3, we discuss in detail some significant social models in this field from the macro- and micro-level. Section 4 introduces a theoretical framework for computational experiments. In Section 5, we conclude this

paper and give an outlook over this field.

## 2 Empirical studies

Empirical studies can provide critical insight into understanding the characteristics of social influence and spread dynamics in social networks. However, this is largely an empirical matter, requiring network data combined with information about the attributes of individuals, group affiliations, sharing information or ideas and their related activities. In this section, we systematically present the methods of existing empirical studies from a network perspective, including network construction, measures of the network, and some significant empirical results that we have obtained from real-world social systems.

### 2.1 Network construction

Most real-world social systems can be represented as social networks with a set of nodes representing individuals and links the relationships between them. In the classical approach, the method of network construction is very simple. Nodes in these networks always belong to the same type and links are often undirected and unweighted. This will not explain certain real-world phenomena very well, since significant information is removed in the construction process. In the past few years, with the development of the social network research domain, more complex and feasible network construction methods have been developed. For instance, we can construct social networks in which nodes belong to two different types and links are both directed and weighted [15–18]. In these networks, nodes may represent people of different genders, occupations, locations, or ages and links represent different friendships, communications, or enmities [3,19,20]. One can also construct social networks with hyperedges [21]: links that join more than two nodes together.

According to our investigation, we generally classify existing networks construction methods into five basic classes: undirected networks, directed networks, weighted networks, bipartite graphs, and hypergraphs. These five basic network construction methods are compared in Table 2.

**Table 1** The framework of main content in this paper

	Main content	Section
Introduction	Background of this field	1
Empirical studies	Network construction	2.1
	Measures of network	2.2
	Empirical results	2.3
Social models	Macro-level models	3.1
	Micro-level models	3.2
Computational experiments	Theoretical framework	4
Concluding remarks	Conclusions and outlook	5

**Table 2** Five basic network construction methods

Types of networks	Main characteristics
Undirected network	Links are undirected
Directed network	Links are directed
Weighted network	Links are weighted
Bipartite graph	Two different types of node
Hyper-graph	Links may join more than two nodes together

## 2.2 Measures of network

With the rapid development of graph theories, researchers have developed several important measures of networks to quantitatively describe the inherent properties of social networks. Examples of these measures include: degree distribution, clustering coefficient, shortest path length, and community structure.

*Degree and strength distribution:* the degree  $k_i$  of node  $i$  is defined as the number of links incident with  $i$ . The strength  $s_i$  in a weighted network denotes the sum of weights of the corresponding links:  $s_i = \sum_j w_{ij}$ , where  $w_{ij}$  represents the weight between node  $i$  and  $j$ . A large number of previous empirical results demonstrate that not all nodes in a social network have the same degree or strength [5,15,22,23]. For an unweighted network, this difference can be characterized by the degree distribution  $P(k)$ , which gives the probability that a randomly selected node has exactly  $k$  edges. For a weighted network, we can use the strength distribution  $P(s)$  to describe the probability of selecting a node with strength  $s$ .

*Clustering coefficient:* this is a measure of network transitivity, which characterizes the fraction of members in circles of friends or acquaintances that an individual knows. Researchers have given several definitions of the clustering coefficient from different perspectives [1,21,24,25]. The most popular definition of clustering coefficient was developed by Watts and Strogatz [1]. In this definition, if a node with degree  $k_i$  is selected randomly in the graph  $G$ , then at most  $k_i(k_i - 1)/2$  links may exist between them. The clustering coefficient  $C_i$  of node  $i$  is computed as follows:

$$C_i = 2E_i/[k_i(k_i - 1)], \quad (1)$$

where  $E_i$  is the number of edges that actually exist between these  $k_i$  neighbors of node  $i$ . For a weighted network, we can use the weighted clustering coefficient proposed by Barthélemy et al. [25]. The weighted clustering coefficient  $C_i^w$  of node  $i$  in a weighted network is defined as

$$C_i^w = \frac{1}{s_i(k_i - 1)} \sum_{j,k} \frac{w_{ij} + w_{ik}}{2} a_{ij} a_{ik} a_{jk}. \quad (2)$$

*Shortest path length:* the shortest path length between node  $i$  and  $j$  is the shortest of the possible paths connecting nodes  $i$  and  $j$ . The maximum value of  $d_{ij}$  is called the diameter of the graph. The average shortest path length  $L$  of a network is defined as the mean of the shortest path lengths over all pairs of nodes [1]

$$L = \frac{1}{N(N-1)} \sum_{i \neq j} d_{ij}, \quad (3)$$

where  $N$  is the number of nodes. To avoid the divergence of  $L$ , researchers have proposed an alternative measure: efficiency

$E$  [26], which is defined as

$$E = \frac{1}{N(N-1)} \sum_{i \neq j} 1/d_{ij}. \quad (4)$$

*Community structure:* a community is also called a group, cluster, cohesive subgroup, or module [27]. The first network formalizations of this concept were proposed by Wasserman and Faust [28]. This measure attempts to obtain some clusters of groups of nodes that have a high density of connections within the group, and a lower density of connections between the groups [29]. In real social networks, we do not know the number of existing communities. To address this problem, many researchers [27,30–34] have developed several useful measures for detecting community structure. These measures can be utilized to deal with different types of networks such as heterogeneous networks. The most popular measure, modularity  $Q$ , was proposed by Newman and Girvan [30]. In this measure, if we want to divide a network into  $c$  communities, we can calculate  $Q$  from the symmetric  $c \times c$  mixing matrix  $E$  whose main diagonal elements  $e_{ii}$  represent the ratio of links between nodes in the same community  $i$  while other elements  $e_{ij}$  capture the ratio of links between nodes in different communities  $i$  and  $j$ . The measure can be calculated as follows:

$$Q = \sum_i \left[ e_{ii} - \left( \sum_j e_{ij} \right)^2 \right] = \text{Tr}E - \|E\|^2. \quad (5)$$

Several other measures have also been used, including the node/edge betweenness (the number of shortest paths going through a node/edge) [28,30], motif (a pattern of interconnections at a number significantly higher than that of random graph), and graph spectra (for discovering the presence of cohesive subgroups and other local patterns) [29]. These measures can help us to quantify the key influence factors and dynamic patterns in our real-world social systems.

## 2.3 Empirical results

Our real-world social systems embody a large number of interesting phenomena. Most of them have not been perceived by humans until now. Previous reviews on social networks [28] or diffusion of innovations [35] have given us some significant empirical results, including strong/weak ties and structural holes. However, these results were built on small social networks. Huge real-world social networks, with millions of nodes and links, may exhibit different patterns. In this subsection, we present and compare significant empirical results.

*Small-world effect:* the small-world effect is characterized by small average shortest path lengths between pairs of nodes and relatively high clustering coefficients. The most famous

empirical study on small-world phenomenon was conducted by Milgram [36], who gave us the concept of “six degrees of separation”. This study concluded that any two randomly selected persons in the USA are on average separated by only six intermediate individuals. Dodds et al. [37] conducted a similar worldwide analysis and found that the diameter of these social networks ranged from 4 to 7. These studies have some limitations: relying much on individual motivation to participate [37] or constraining the networks to a specific dynamic behavior [38]. To solve this problem, researchers have recently attempted to use massive already-recorded log data sets and conducted significant empirical studies. Research objects of existing studies include auctions [39], movie actors [1], scientific collaboration [40], citations [41], and communication [42]. Empirical results have further emphasized that the small-world effect characterizes most of our real-world social systems.

*Scale-free property:* traditional research, until a few years ago, had always considered that almost all nodes were topologically equivalent: like in regular lattices or in random graphs. The distribution of the network is binomial or Poisson at the limit of large graph size. However, recent massive empirical studies on real-world social networks have shown that the distribution  $P(k)$  of these networks follows a power-law shaped distribution,  $P(k) \sim k^{-\alpha}$ , which significantly deviates from that of traditional research. Examples include: traditional epidemic social networks [15], online social networks [18,43,44], Internet news [45], and mobile communication networks [46]. This demonstrates that in these networks there exist a few nodes linking to many other nodes, and a large number of nodes with poorly connected elements.

*Community structure:* many real-world social networks also display a strong community structure [27,34,47,48]. In these networks, there exists a heterogeneous connecting structure, which is mainly characterized by the presence of nodes which are more densely interconnected than the rest of the network. This community structure reflects in general the self-organization of individuals to achieve some intentions. Traditional studies mainly focus on static community detection [34]. However, most real-world social networks tend to evolve gradually, due to frequent changes in the activity and interaction of their individuals [49]. The communities inside a dynamic network may grow or shrink, and the community membership of the individuals shifts regularly [50]. As such, a number of researchers pay more interest to identifying critical events that characterize the evolution of communities and membership of individuals such as author communities in the blogosphere [51], mobile subscriber networks [52], and re-

search communities [40]. These studies have determined that many real-world social networks exhibit self-similar properties and the community size distribution follows a power-law.

*Cascading behavior:* cascading behavior is a social phenomenon when an idea or action is widely spread and adopted through social influence [53]. This has been studied for many years by large number of researchers from several different domains [35,54]. A growing number of empirical studies suggest that diverse phenomena in physical world, including obesity [55], happiness [56], ideas [57], and many other behaviors and affective states [58,59], can spread from person to person. Recently some researchers have conducted many significant empirical studies in cyberspace. Adar et al. [60] and Gruhl et al. [61] extracted cascades in the blogosphere and found that while information propagates between blogs, examples of genuine cascading behavior appeared relatively rarely. Leskovec et al. [62,63] further analyzed cascades in the blogosphere and viral marketing. They find some novel cascading patterns and observed that a cascade on  $n$  nodes follows a Zipf distribution:  $P(n) \propto n^{-2}$ . Anagnostopoulos et al. [64] and Aral et al. [65] analyzed the causes of correlation in social networks and distinguished influence-based contagion from homophily-driven diffusion in dynamic networks. These findings are essential to both our understanding of the mechanisms that drive contagion in networks and our knowledge of how to propagate them in diverse domains.

---

### 3 Social models

Empirical findings have initiated a revival of network modeling, since the models proposed in mathematical graph theory have turned out to be far from the empirical observations. In past few decades, social modeling for the process of social influence and spread dynamics in social networks has been studied in many areas, such as the spread of epidemics, the diffusion of technological innovations, and the effect of “word of mouth” in the promotion of new products. A large number of models have been developed. In this section, we classify these existing models into two types: macro-level models and micro-level models. The macro-level models mainly encompass SIR (susceptible/infective/removed) model and Bass model, which mainly focus on capturing the spread behaviors at the population level. The micro-level models aim to reveal the personal behavior integrating with the topological structure of social networks. These micro-level models include the famous preferential attachment (PA) model, threshold model, cascade model, and competitive model. The key

**Table 3** Key characteristics and limitations of existing social models

Levels	Models	Key characteristics	Limitations
Macro-level	SIR	Dividing populations into three groups	These models assume a homogenous mixed population where individuals have equally probability to be influenced by others.
	Bass	Influence via both mass media and word-of-mouth	
Micro-level	PA	Preferential attachment	These models cannot be evaluated easily. More specific issues such as the influence of correlations and dynamic rewiring should be further considered.
	Threshold	Node specific threshold	
	Cascade	Adopting probability depends on neighbor influence	
	Competitive	Influence maximization game	

characteristics and limitations of these models are presented in Table 3.

### 3.1 Macro-level models

Macro-level models can help us to evaluate the extent of an epidemic or information spreading qualitatively at different spread phases or across diverse groups. This field in the past few decades has attracted great attention including epidemiologists and economists. Many models have been developed and applied across diverse domains. Most of these interesting models are extended or generalized from two basic models: SIR model and Bass model. In this subsection we mainly focus on these models and present their specific mechanisms.

*SIR models:* the most famous and classical SIR model was originally developed to describe the spread dynamics of epidemics and is now applied to several domains. This model divides the population into three groups: susceptible ( $S$ ), infectious ( $I$ ), and removed ( $R$ ).  $S$  indicates that people who are not infected by the virus but can be easily infected.  $I$  represents individuals who are infectious and can transmit the virus to others.  $R$  denotes persons in the group who have recovered from the disease or are now dead. This model is based on the hypothesis that any person who belongs to  $S$  has the uniform probability  $\beta$  to become infected and infected people recover and becomes immune at rate  $\gamma$ . In the limit of populations, this model is represented by the following differential equations:

$$\frac{ds}{dt} = -\beta is, \quad \frac{di}{dt} = \beta is - \gamma i, \quad \frac{dr}{dt} = \gamma i, \quad (6)$$

where  $s(t)$ ,  $i(t)$ , and  $r(t)$  are the fractions of individuals in  $S$ ,  $I$ ,  $R$  respectively and the sum of  $s(t)$ ,  $i(t)$ , and  $r(t)$  is equal to 1.

The classic SIR model described above assumes that the population is fully mixed and all individuals are in contact with the others and transmit the disease with same probability. Newman [66] improved on this model and made several modifications. In Newman's model, the probability  $T_{ij}$  that an infected node  $i$  transmits the virus to node  $j$  follows a distri-

bution function, which can be determined by several significant parameters [66]. The resulting models are equivalent to uniform bond percolation on the same network with edge occupation probability. We can further define a generation function [24] to obtain the distribution of finite outbreak sizes, the critical transmittability, and relative final size of an epidemic. Moreno and Vázquez [67] developed networks with correlations between the degrees of nodes. Ancel et al. [68] and Newman [66,68] introduced networks with different types of nodes. Wang et al. [69] also gave a model to describe the spread behavior with non-uniform transmission. These network models are not only applied to analyze the spread dynamics of epidemics, but also to study the social influence of rumors [8,70].

*Bass model:* Bass model [71] is another classical macro-level mathematical model for the underlying diffusion of innovation in society. The Bass model assumes that potential adopters of an innovation are influenced by two means of communication: mass media and word of mouth [72]. As such, this model has two key characteristics [72–74]: one captures the rate of individuals convinced by mass media (external influence), and the other describes the ratio by which individuals are influenced by word-of-mouth communication (internal influence).

In the Bass model, at each step  $t$ , individuals are convinced by the mass media with probability  $p$  and by word-of-mouth communication by fraction  $q$ . If we assume that the ratio of individuals who will adopt the innovations or ideas is  $R(t)$ , then the differential equation of this model takes the form

$$R(t) = R(t-1) + p(1 - R(t-1)) + qR(t-1)[1 - R(t-1)]. \quad (7)$$

In this form, the first term of the right side presents the ratio of individuals who have adopted the innovations at time  $t-1$ . The second term describes the fraction of people who have not yet adopted but have been convinced by the mass media. The third term captures the rate of the imitation process that innovations spread by word-of-mouth communication. The Bass model can also be represented in a continuous

time form

$$dR(t)/dt = (1 - R(t))[p + qR(t)]. \quad (8)$$

Recently, some extensions of the Bass model have been applied in several domains such as pricing and advertising [63,75].

### 3.2 Micro-level models

The SIR and Bass models described in Section 3.1 both assume a homogeneous mixing population where individuals have equal probability to be influenced by one another. This does not accord with many empirical findings in heterogeneous networks. Based on previous work, researchers have developed some reasonable models at the micro-level aiming to describing the key influence factors and spread patterns of each individual.

*Preferential attachment models:* the empirical studies discussed in Section 2.3 demonstrate that many large networks are scale-free, that is their degree distribution follows a power-law. Previous random graph models cannot reproduce this feature. Barabási and Albert [5] and other researchers [76–78] introduced their networks and later were generally called preferential attachment (PA) models due to their growth mechanisms of preferential attachment.

The network model developed by Barabási and Albert [5] (BA model) starts with a fixed number of randomly connected nodes. At each time instant the network grows with the addition of new nodes. For each newly added node, new edges are added between it and some old nodes. The nodes to receive new edges are chosen following a linear preferential attachment rule, that is, the probability  $P(k)$  of an old node receiving a new edge is proportional to its degree  $k$ , that is,  $P(k) \sim k$ . Replacing the linear preferential attachment of the BA model, Krapivsky, Redner, and Leyvraz [76] use a non-linear preferential attachment rule. When choosing the nodes to which a new node connects, the probability  $P(k)$  depends on  $k^\alpha$ ,  $P(k) \sim k^\alpha$ .

Both models discussed above have a common characteristic: their preferential attachment rules depend only on the degree of the old node. However, in applications such as reference networks, aging occurs: the authors rarely cite very old papers. Dorogovtsev and Mendes [77] proposed an extended model in which the probability  $P(k)$  is dependent not only on the degree  $k$  of the old node but also on its age  $\tau$ , that is,  $P(k) \sim k\tau^{-\beta}$ , where  $\beta$  is a tunable parameter.

*Threshold models:* although they reproduce the scale-free property well, the PA models described above cannot reflect

the inherent influence behavior between nodes. In this condition, we should consider another suitable micro-level social influence and spread models such as the threshold model.

The earliest threshold models were developed by Granovetter [79] and Schelling [80] in 1978; they later became the foundation for a large body of work in sociology [81–83]. In these models, nodes can be classified into two types: active or inactive. An inactive node  $m$  is influenced by its neighbor  $n$  with a weight  $w_{n,m}$ , where  $\sum_n w_{n,m} \leq 1$ . Node  $m$  selects a threshold  $\theta_m$  on the uniform interval  $[0, 1]$ . Given an initial set of active nodes, the influence process can be divided into finite discrete steps: if nodes are active at step  $t - 1$ , then they will still remain active at step  $t$ ; and if the total weight of a nodes active neighbors is at least  $\theta_m$ :  $\sum_{n_a} w_{n_a,m} \geq \theta_m$ , then node  $m$  will be activated.

In the threshold models described above, different thresholds  $\theta_m$  indicate different potential tendencies of nodes to be influenced (i.e., accepting a new product or an idea) by their neighbors. Kempe et al. [84,85] recently introduced a general threshold model. This model focused more on the cumulative effect of a nodes' influence than that of previous models. Each node in this model has a monotone activation function and a threshold. One node becomes active when the value of the function is no less than the threshold. Easley and Kleinberg [86] further gave a heterogeneous threshold model. This model starts from a set of initial adopters. At each step, each node evaluates its decision according to its own threshold rule and switches to one behavior if the value is no less than its threshold. This model can explain the diversity of spread behavior in real-world heterogeneous networks.

*Cascade models:* many real-world social systems exhibit cascade behavior. However, previous network models including PA models and threshold models cannot capture this characteristic. Due to this, researchers recently have proposed several cascade models. Of these, the independent cascade model (ICM) proposed by Goldenberg et al. [87] has been paid more and more attention. In this model, node  $m$  first becomes active at step  $t - 1$  and has only one chance to activate its inactive neighbor  $n$  with a constant probability  $p_n(m)$ . If  $m$  succeeds,  $n$  will become active at step  $t$ . If multiple neighbors of  $n$  are active at step  $t - 1$ , activation attempts will be executed randomly.

Based on the ICM, Kempe et al. [84,85] further proposed a general cascade model (GCM) and a decreasing cascading model (DCM). In the GCM, the probability that node  $m$  succeeds in activating its neighbor  $n$  depends the extent of  $n$ 's neighbors that have already tried to activate it. Specifically, given an incremental function  $p_n(m, S) \in [0, 1]$ , where  $S$  and

$m$  are disjoint subset of  $n$ 's neighbors.  $m$  succeeds in activating  $n$  with probability  $p_n(m, S)$ . Due to the phenomena of "marketing-saturated", DCM further defines a non-increasing function  $p_n(m, S)$  in  $S$ , i.e.,  $p_n(m, S) \geq p_n(m, T)$  whenever  $S \subseteq T$ . This condition indicates that node  $m$ 's probability of succeeding in activating node  $n$  decreases with the number of nodes that have already attempted to activate  $n$ .

*Competitive models:* following the cascade model [85,87] described above, Bharathi et al. [11] introduced a competitive model for the influence maximization game. In this model, each edge is allocated an activation probability  $P_e$  and each node has one of these two states: active or inactive. In the active state, the node is labeled by a color denoting that it has been activated.

This model is further augmented by adding the information of activation time for each activation attempt. When node  $m$  becomes active at time  $t$ , it will attempt to activate each first-level inactive neighbor  $n$ . If  $m$  succeeds in activating  $n$ ,  $n$  will become active with the same color as  $m$  at time  $t + T_{mn}$ , where  $T_{mn}$  are independent continuous random variables following an exponential distribution. Subsequently, the active node  $n$  will attempt to activate its first-level inactive neighbors, and so forth. In this influence maximization game, each player selects a set  $S_i$  of at most  $k_i$  nodes. A node selected by multiple players will be labeled by the color of one of these players. This process unfolds as described above until no new activation occurs. Similar models have also been considered recently by Lotker et al. [88] and Dubey et al. [89]. These models provide specific mechanisms that illustrate how to interact with each other when multiple objects are competing within a social network.

## 4 Computational experiments

Section 3 above presented several types of social models for social influence and spread dynamics in social networks. Traditionally, researchers validate these models through simple simulation. This method may be valuable and ethical when examining policies dealing with matters of life and death, such as in epidemiology and terrorism [90]. However, due to the difficulties of testing real systems that are inherently open, dynamic, complex, and unpredictable, simple simulations cannot solve all the problems such as how to combine real data and simulation data for verifying and validating real social systems.

To deal with these problems, in recent years, computational experiments with artificial systems and more sophisticated simulation techniques were developed by Wang et al.

[90–92]. Based on this work, Zheng et al. [93] further proposed a theoretical framework of computational experiments which can be applied in diverse domains such as the science of team science. This framework shown in Fig. 1 mainly encompasses artificial system construction, computational experiment design, observation and evaluation, and feedback and modification. At the level of artificial systems construction, the fundamental research issues include agent interactions, the construction of interacting environments and rules. At the computational experiment design level, we should be concerned with the specification of objectives, task assignment, and experimental parameter selection. At the observation and evaluation level, we should focus on emergence-based observation, validation of objectives, and evaluation of social models. The feedback and modification of solutions for complex systems from the observation and evaluation levels can be used to improve artificial systems.

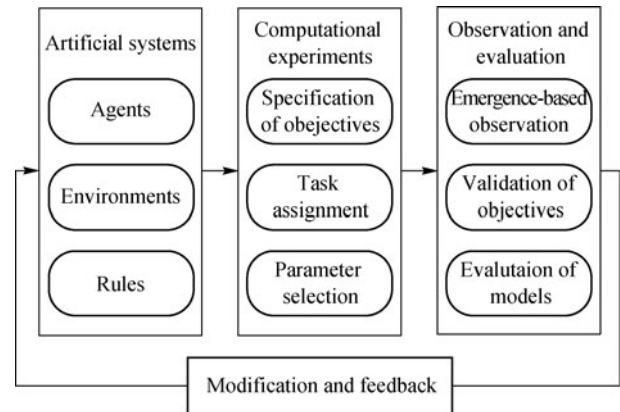


Fig. 1 The theoretical framework of computational experiments

From an implementation viewpoint, to solve those problems, such as incomplete or unavailable real-world data [90], we can utilize parallel execution [91], by executing one or more artificial systems running in parallel with a real system and employing adaptive control methods for the experiments. Through comparison, evaluation, and interaction with artificial systems, we can provide more perfect strategies for managing the target social systems.

## 5 Concluding Remarks

This paper has systematically presented existing significant results in the field of social influence and spread dynamics in social networks, ranging from empirical studies and social models to computational experiments. The section on empirical studies mainly focuses on network construction, measures of networks and significant empirical results. Several

significant social models in the section of social models have been introduced and compared at both macro- and micro-levels. The macro-level models encompass SIR and its generalized models, and Bass models. We have compared the specific mechanisms of important micro-level models, including the preferential attachment models, threshold models, cascade models and competitive models. Due to the difficulties in evaluation and validation of these models, we have also explained a theoretical framework for computational experiments.

We believe that this paper can help researchers keep up with the latest results in this field and may offer some insight into understanding related motivations, connections, and open problems. In the future, the rapid growth of the Internet, mobile, and cloud computing technologies and their applications will offer us a great deal of data sources. These massive network data sets will provide us with more reliable evidence for studying social influence and spread dynamics in social networks. Existing theoretical models can then be validated and more reasonable models will be developed. These models can be efficiently evaluated on computational experiment platforms and help us to comprehensively understand the evolution of real-world social systems and to make more reasonable strategies for controlling and managing these social systems.

**Acknowledgements** This work was supported in part by the following grants: The National Natural Science Foundation of China (Grant Nos. 71103180, 91124001, 60921061, 90924302, 70890084, 91024030, 60875028, and 40901219); The Humanities and Social Sciences General Education Project of Chinese Ministry of Education (10YJA630021); The Soft Science Research Projects of Qingdao City (10-3-3-40-43-zhc); The Chinese Academy of Sciences (2F09N06).

## References

- Watts D J, Strogatz S H. Collective dynamics of “small-world” networks. *Nature*, 1998, 393(6684): 440–442
- Song C, Havlin S, Makse H A. Self-similarity of complex networks. *Nature*, 2005, 433(7024): 392–395
- Kossinets G, Watts D J. Empirical analysis of an evolving social network. *Science*, 2006, 311(5757): 88–90
- Centola D. The spread of behavior in an online social network experiment. *Science*, 2010, 329(5996): 1194–1197
- Barabási A L, Albert R. Emergence of scaling in random networks. *Science*, 1999, 286(5439): 509–512
- Cui K, Cao Z, Zheng X, Zeng K, Zheng M. A geospatial analysis on the potential value of news comments in infectious disease surveillance. In: *Proceedings of the 6th Pacific Asia Conference on Intelligence and Security Informatics*. 2011, 85–93
- Lind P G, da Silva L R, Andrade J S Jr, Herrmann H J. Spreading gossip in social networks. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 2007, 76(3): 036117
- Moreno Y, Nekovee M, Pacheco A F. Dynamics of rumor spreading in complex networks. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 2004, 69(6): 066130
- Bakshy E, Hofman J M, Mason W A, Watts D J. Everyone’s an influencer: quantifying influence on twitter. In: *Proceedings of the 4th ACM International Conference on Web Search and Data Mining*. 2011, 65–74
- Chierichetti F, Lattanzi S, Panconesi A. Rumor spreading in social networks. *Theoretical Computer Science*, 2011, 412(24): 2602–2610
- Bharathi S, Kempe D, Salek M. Competitive influence maximization in social networks. In: *Proceedings of the 3rd International Conference on Internet and Network Economics*. 2007: 306–311
- Eagle N, Pentland A, Lazer D. Inferring friendship network structure by using mobile phone data. *Proceedings of the National Academy of Sciences of the United States of America*, 2009, 106(36): 15274–15278
- Castellano C, Fortunato S, Loreto V. Statistical physics of social dynamics. *Reviews of Modern Physics*, 2009, 81(2): 591–646
- Lazer D, Pentland A, Adamic L, Aral S, Barabási A-L, Brewer D, Christakis N, Contractor N, Fowler J, Gutmann M, et al. Computational social science. *Science*, 2009, 323(5915): 721–723
- Zheng X, Zeng D, Sun A, Luo Y, Wang Q, Wang F Y. Network-based analysis of Beijing SARS data. In: *Proceedings of the 2008 International Workshop on Biosurveillance and Biosecurity*. 2008: 64–73
- Newman M E J, Watts D J, Strogatz S H. Random graph models of social networks. *Proceedings of the National Academy of Sciences of the United States of America*, 2002, 99(Suppl 1): 2566–2572
- Liljeros F, Edling C R, Amaral L A N, Stanley H E, Aberg Y. The web of human sexual contacts. *Nature*, 2001, 411(6840): 907–908
- Leskovec J, Backstrom L, Kumar R, Tomkins A. Microscopic evolution of social networks. In: *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2008, 462–470
- Szell M, Lambiotte R, Thurner S. Multirelational organization of large-scale social networks in an online world. *Proceedings of the National Academy of Sciences of the United States of America*, 2010, 107(31): 13636–13641
- Zheng X, Zeng D, Cao Z, Wang Q, Wang F. Evolutionary patterns on SARS networks. In: *Proceedings of the 2009 International Workshop on Biosurveillance and Biosecurity*. 2009
- Newman M E J. The structure and function of complex networks. *SIAM Review*, 2003, 45(2): 167–256
- Albert R, Barabási A L. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 2002, 74(1): 47–97
- Barrat A, Barthélemy M, Vespignani R P S. The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences of the United States of America*, 2004, 101(11): 3747–3752
- Newman M E J, Strogatz S H, Watts D J. Random graphs with arbitrary degree distributions and their applications. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 2001, 64(2): 026118
- Barthélemy M, Barrat A, Pastor-Satorras R, Vespignani A. Characterization and modeling of weighted networks. *Physica A: Statistical Mechanics and Its Applications*, 2005, 346(1–2): 34–43
- Latora V, Marchiori M. Efficient behavior of small-world networks. *Physical Review Letters*, 2001, 87(19): 198701
- Tang L, Liu H. Community detection and mining in social media. *Synthesis Lectures on Data Mining and Knowledge Discovery*, 2010, 2(1): 1–137
- Wasserman S, Faust K. *Social Networks Analysis: Methods and Applications*. Cambridge: Cambridge University Press, 1994
- Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang D U. Complex



- networks: structure and dynamics. *Physics Reports*, 2006, 424(4–5): 175–308
30. Newman M E J, Girvan M. Finding and evaluating community structure in networks. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 2004, 69(2): 026113
  31. Newman M E J. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences of the United States of America*, 2006, 103(23): 8577–8582
  32. Deng C, Zheng S, He X F, Yan X F, Han J W. Mining hidden community in heterogeneous social networks. In: *Proceedings of the 3rd International Workshop on Link Discovery*. 2005, 58–65
  33. Santo F. Community detection in graphs. *Physics Reports*, 2010, 486(3–5): 75–174
  34. Takaffoli M, Sangi F, Fagnan J, Z̄arane O R. Community evolution mining in dynamic social networks. *Procedia-Social and Behavioral Sciences*, 2011, 22: 49–58.
  35. Rogers E M. *Diffusion of Innovations*. New York: Free Press, 1995
  36. Milgram S. The small-world problem. *Psychology Today*, 1967, 1(1): 61–67
  37. Dodds P S, Muhamad R, Watts D J. An experimental study of search in global social networks. *Science*, 2003, 301(5634): 827–829
  38. Costa L D F, Oliveira O N, Travieso G, Rodrigues F A, Villas Boas P, Antiquera L, Viana M P, Correa Rocha L. Analyzing and modeling real-world phenomena with complex networks: a survey of applications. *Advances in Physics*, 2011, 60(3): 329–412
  39. Wang J C, Chiu C C. Recommending trusted online auction sellers using social network analysis. *Expert Systems with Applications*, 2008, 34(3): 1666–1679
  40. Palla G, Barabasi A L, Vicsek T. Quantifying social group evolution. *Nature*, 2007, 446(7136): 664–667
  41. Bilke S, Peterson C. Topological properties of citation and metabolic networks. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 2001, 64(3): 036106
  42. Guimerà R, Danon L, Díaz-Guilera A, Giralt F, Arenas A. Self-similar community structure in a network of human interactions. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 2003, 68(6): 065103
  43. Zheng X, Li H, Sun A. Exploring social dynamics in online bookmarking systems. In: *Proceedings of 2008 International Workshop on Social Computing*. 2008, 390–391
  44. Mislove A, Marcon M, Gummadi K P, Druschel P, Bhattacherjee B. Measurement and analysis of online social networks. In: *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement*. 2007, 29–42.
  45. Wang Y, Zeng D, Zheng X, Wang F. Analyzing online media as complex network. *Complex Systems and Complexity Science*, 2008, 6(3): 11–21
  46. Onnela J P, Saramäki J, Hyvönen J, Szabó G, Lazer D, Kaski K, Kertész J, Barabási A L. Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Sciences of the United States of America*, 2007, 104(18): 7332–7336
  47. Arenas A, Danon L, Díaz-Guilera A, Gleiser P M, Guimerà R. Community analysis in social networks. *The European Physical Journal B - Condensed Matter and Complex Systems*, 2004, 38(2): 373–380.
  48. Girvan M, Newman M E J. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, 2002, 99(12): 7821–7826
  49. Newman M E J, Park J. Why social networks are different from other types of networks. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 2003, 68(3): 036122
  50. Backstrom L, Huttenlocher D, Kleinberg J, Lan X. Group formation in large social networks: membership, growth, and evolution. In: *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2006, 44–54
  51. Lin Y R, Chi Y, Zhu S, Sundaram H, Tseng B L. Facetnet: a framework for analyzing communities and their evolutions in dynamic networks. In: *Proceedings of the 17th International Conference on World Wide Web*. 2008, 685–694
  52. Wu B, Ye Q, Yang S, Wang B. Group CRM: a new telecom CRM framework from social network perspective. In: *Proceedings of the 1st ACM International Workshop on Complex Networks Meet Information & Knowledge Management*. 2009, 3–10
  53. Bikhchandani S, Hirshleifer D, Welch I. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy*, 1992, 100(5): 992–1026
  54. Fowler J H, Christakis N A. Cooperative behavior cascades in human social networks. *Proceedings of the National Academy of Sciences of the United States of America*, 2010, 107(12): 5334–5338
  55. Christakis N A, Fowler J H. The spread of obesity in a large social network over 32 years. *New England Journal of Medicine*, 2007, 357(4): 370–379
  56. Fowler J H, Christakis N A. The dynamic spread of happiness in a large social network. *British Medical Journal (Clinical Research Ed.)*, 2008, 337: a2338
  57. Singh J. Collaborative networks as determinants of knowledge diffusion patterns. *Management Science*, 2005, 51(5): 756–770
  58. Christakis N A, Fowler J H. The collective dynamics of smoking in a large social network. *New England Journal of Medicine*, 2008, 358(21): 2249–2258
  59. Rosenquist J N, Murabito J, Fowler J H, Christakis N A. The spread of alcohol consumption behavior in a large social network. *Annals of Internal Medicine*, 2010, 152(7): 426–433
  60. Adar E, Adar E, Adamic L A. Tracking information epidemics in blogspace. In: *Proceedings of 2005 IEEE/WIC/ACM International Conference on Web Intelligence*. 2005, 207–214
  61. Gruhl D, Guha R V, Liben-Nowell D, Tomkins A. Information diffusion through blogspace. In: *Proceedings of the 13th International Conference on World Wide Web*. 2004, 491–501
  62. Leskovec J, McGlohon M, Faloutsos C, Glance N S, Hurst M. Patterns of cascading behavior in large blog graphs. In: *Proceedings of 7th SIAM International Conference on Data Mining*. 2007
  63. Leskovec J, Adamic L A, Huberman B A. The dynamics of viral marketing. *ACM Transactions on the Web*, 2007, 1(1): 5
  64. Anagnostopoulos A, Kumar R, Mahdian M. Influence and correlation in social networks. In: *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2008, 7–15
  65. Aral S, Muchnik L, Sundararajan A. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences of the United States of America*, 2009, 106(51): 21544–21549
  66. Newman M E J. The spread of epidemic disease on networks. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 2002, 66(1): 016128
  67. Moreno Y, Vázquez A. Disease spreading in structured scale-free networks. *The European Physical Journal B: Condensed Matter and Complex Systems*, 2003, 31(2): 265–271
  68. Ancel L W, Newman M E J, Martin M, Schrag S. Applying network theory to epidemics: control measures for *Mycoplasma pneumoniae* outbreaks. *Emerging Infectious Diseases*, 2001, 9(2): 204–210
  69. Wang J, Liu Z, Xu J. Epidemic spreading on uncorrelated heterogeneous networks with non-uniform transmission. *Physica A: Statistical Mechanics and its Applications*, 2007, 382(2): 715–721

70. Zanette D. Dynamics of rumor propagation on small-world networks. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 2002, 65(4): 041908
71. Bass F M. A new product growth for model consumer durables. *Management Science*, 1969, 15(5): 215–227
72. Mahajan V, Muller E, Bass F M. New product diffusion models in marketing: a review and directions for research. *Journal of Marketing*, 1990, 54(1): 1–26
73. Jackson M O. *Social and Economic Networks*. Princeton: Princeton University Press, 2008
74. Kempe D. *Structure and Dynamics of Information in Networks*. 2011
75. Young H P. The evolution of conventions. *Econometrica: Journal of the Econometric Society*, 1993, 61(1): 57–84
76. Kravivsky P L, Redner S, Leyvraz F. Connectivity of growing random networks. *Physical Review Letters*, 2000, 85(21): 4629–4632
77. Dorogovtsev S N, Mendes J F F. Evolution of reference networks with aging. *Physical Review E: Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, 2000, 62(2): 1842–1845
78. Zheng X, Zeng D, Li H, Wang F. Analyzing open-source software systems as complex networks. *Physica A: Statistical Mechanics and its Applications*, 2008, 387(24): 6190–6200
79. Granovetter M. Threshold models of collective behavior. *American Journal of Sociology*, 1978, 83(6): 1420–1443
80. Schelling T. *Micromotives and macrobehavior*. New York: Norton, 1978
81. Macy I W. Chains of cooperation: threshold effects in collective action. *American Sociological Review*, 1991, 56(6): 730–747
82. Macy M W, Willer R. From factors to actors: computational sociology and agent-based modeling. *Annual Review of Sociology*, 2002, 28(1): 143–166
83. Berger E. Dynamic monopolies of constant size. *Journal of Combinatorial Theory Series B*, 2001, 83(2): 191–200
84. Kempe D, Kleinberg J M, Tardos É. Influential nodes in a diffusion model for social networks. In: *Proceedings of the 32nd International Colloquium on Automata, Languages and Programming*. 2005, 1127–1138
85. Kempe D, Kleinberg J, Tardos V. Maximizing the spread of influence through a social network. In: *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2003, 137–146
86. Easley D, Kleinberg J. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. New York: Cambridge University Press, 2010
87. Goldenberg J, Libai B, Muller E. Using complex systems analysis to advance marketing theory development. *Academy of Marketing Science Review*, 2001, 2001(9): 1–19
88. Lotker Z, Patt-Shamir B, Tuttle M R. Publish and perish: definition and analysis of an n-person publication impact game. In: *Proceedings of the 18th Annual ACM Symposium on Parallelism in Algorithms and Architectures*. 2006, 11–18
89. Dubey P, Garg R, De Meyer B. Competing for customers in a social network: the quasi-linear case. In: *Proceedings of the 2nd International Workshop on Internet and Network Economics*. 2006, 162–173
90. Wang F Y, Carley K M, Zeng D, Mao W. Social computing: from social informatics to social intelligence. *IEEE Intelligent Systems*, 2007, 22(2): 79–83
91. Wang F Y. Toward a paradigm shift in social computing: the ACP approach. *IEEE Intelligent Systems*, 2007, 22(5): 65–67
92. Wang F Y. Computational experiments for behavior analysis and decision evaluation in complex systems. *Journal of System Simulation*, 2004, 16(5): 893–897
93. Zheng X, Ke G, Zeng D, Ram S, Lu H. Next-generation team-science platform for scientific collaboration. *IEEE Intelligent Systems*, 2011, 26(6): 72–76



Xiaolong Zheng is an assistant professor in the State Key Laboratory of Management and Control for Complex Systems at the Institute of Automation of the Chinese Academy of Sciences. His research interests include complex networks, social computing, and data mining. Specifically, he is interested in social dynamics, social influence, spread dynamics, and opinion and behavior mining in social media.



Yongguang Zhong is a professor in the Department of Management Science and Engineering at Qingdao University in China. His current research interests include system dynamics, operations management, and reverse logistics.



Daniel Zeng is a professor in the State Key Laboratory of Management and Control for Complex Systems at the Institute of Automation of the Chinese Academy of Sciences. His research interests include multi-agent systems, collaborative information and knowledge management, recommender systems, and automated negotiations and auctions. Specifically, he is interested in social computing, information and security informatics, spatio-temporal data analysis, and online surveillance.



Fei-Yue Wang is a professor in the State Key Laboratory of Management and Control for Complex Systems at the Institute of Automation of the Chinese Academy of Sciences. His current research interests include social computing, parallel control and management, web and services science, and agent-based intelligent systems.