MFSC: Mean-Field-Theory and Spreading-Coefficient Based Degree Distribution Analysis in Social Network

Chongze Lin[1] and Yi Zheng[1]
[1] Zhejiang Economic Information Center
Hangzhou, Zhejiang 310006 – P. R. China
[e-mail: zy19830814@hnu.edu.cn]
*Corresponding author: Yi Zheng

Abstract

Degree distribution can provide basic information for structural characteristics and internal relationship in social network. It is a critical procedure for social network topology analysis. In this paper, based on the mean-field theory, we study a special type of social network with exponential distribution of time intervals. First of all, in order to improve the accuracy of analysis, we propose a spreading coefficient algorithm based on intimate relationship, which determines the number of the joined members through the intimacy among members. Then, simulation show that the degree distribution of follows the power-law distribution and has small-world characteristics. Finally, we compare the performance of our algorithm with the existing algorithms, and find that our algorithm improves the accuracy of degree distribution as well as reducing the time complexity significantly, which can complete 29.04% higher precision and 40.94% lower implementation time.

1. Introduction

$\mathbf{C}$omplex networks [1], such as social network (SN) [2], often evolve heavily over time in practice, and they frequently experience topological changes during their evolution. Watts and Strogatz [3] introduced a pioneering article on this subject which deals with the structure and the dynamics of small-world networks. A small-world network is a type of mathematical graph in which most nodes are not neighbors of one another, but the neighbors of any given node are likely to be neighbors of each other and most nodes can be reached from every other node by a small number of hops or steps. Barabasi and Albert [6] proposed the model of network growth which is inspired by the emergence of the World Wide Web and is based on two basic features: growth and preferential attachment.

The degree of a node in a network is the number of connections or edges the node has to other nodes. The degree distribution $P(k)$ of a network is then defined to be the fraction of nodes in the network with degree k. A relevant property regards the degree of a node [10], is the number of its direct connections to other nodes. In real networks, the degree distribution $P(k)$ [13], defined as the probability that a node chosen uniformly at random has degree k or, equivalently, as the fraction of nodes in the graph having degree k, significantly deviates from the poisson distribution expected for a random graph [14] and, in many cases, exhibits a power law (scale-free) tail with an exponent [7]. Moreover, practical networks are characterized by correlations in the node degrees, by having relatively short paths between any two nodes [4], and by the presence of a large number of short cycles or specific motifs [5]. Thus, the research of degree distribution is beneficial to provide useful information for researchers, who investigate the interest of social members and present recommendation based on their interests.

Recently, many researchers focus on degree distribution for SN and many algorithms have been proposed [16,17,18,19,20,21,22,29,30,31,32,33,34]. For example, some algorithms [16,17] were designed for degree distribution by utilizing master equation method based on markov chain. Some approaches were based on mean-field-theory algorithm [18,19,20,21,22,34] which utilize the principle of continuity to analyze the change of node degree in complex networks. Yang *et al*. [29] utilized the friend circle to calculate the intimacy between nodes. Zhao *et al*. [30] analyzed the overlap of links between nodes for the intimacy calculation. Song *et al*. [31] proposed tracking a set of influential nodes dynamically for the spread analysis. And other social circle and link studies [32,33].

Although the existing approaches can utilize mathematical methods and social network information to solve the problem of degree distribution in social network, their research did not involve the degree distribution of changed joined members and special distributed time interval. According to the actual observation, investigation and reference [34], there is a specific SN $G = (V, E)$. For example, the club specializes for the members. When the member join the social organization, he or she will associate with other existing members. As time goes on, the time interval of member joining often follows exponential distribution (more detailed explanations are provided in Section 3.1). Some close friends may join together with this member and the size of SN is increasing with time passing. The calculated result of degree distribution depends on the number of members and edges (the relationship between members) in SN. However, this specific SN has different characteristics, which the time interval of member joining follows exponential distribution and indeterminate number of joined

members in each moment. These complicated relationships and uncertain quantity of joined members have brought many challenges to the study of degree distribution. Therefore, the existing methods are not suitable to this specific SN for these influence factors.

To address the above problems, our main goal is to design a new method for degree distribution in specific SN. In this paper, we first propose a new metric, named intimacy, which describes the comprehensive relationship between nodes. Then, we analyze the intimacy between joined node and other nodes. If the node has higher intimacy with joined node, it can also join in this SN. Further, we propose a new parameter called spreading coefficient by intimacy among nodes, which counts the number of joined nodes accurately. Spreading coefficient is good at solving the problem of complicated relationship among nodes and improving the accuracy of degree distribution. According to the characteristics of time intervals, we calculate the number of joined nodes, all nodes and edges in the end. And then, in order to improve the accuracy of degree distribution and reduce implementation time, the mean-field theory base on spreading coefficient is provided for giving an efficient method to analyze the degree distribution of this SN. Finally, we utilize Normalized Mutual Information (NMI) and real-world dataset to verify the characteristics of this network and the effectiveness of our approach. To the best of our knowledge, we are first one to utilize mean field theory with spreading coefficient for studying degree distribution. Our method aims at calculating exact degree distribution for specific SN which conforms to the actual outcomes.

The main contributions of this paper are listed as follows:

(1) We present a new metric, named intimacy, which describes the comprehensive relationship between joined node and other nodes. It consists of family affection, friendship and hobby similarity, which depicts all kinds of relationship among social members respectively. The value of intimacy reveals nodes' relationship in a certain degree.

(2) We present a new concept named spreading coefficient, which reflects the relationship between nodes and measure SN size. It consists of intimacy and threshold value, which improve the precision of degree distribution. The value of spreading coefficient reveals SN size.

(3) We propose a mean field theory and spreading coefficient based degree distribution algorithm (MFSC). The intimacy and spreading coefficient are defined to calculate the degree of nodes. Comparing with social power based degree distribution algorithm [21], MFSC can save some repeated computation to achieve the degree distribution quickly and accurately.

(4) We conduct simulations with real data on the UCINET (University of California at Irvine NETwork) [35] analysis platform to evaluate the performance of MFSC. The results show that degree distribution of this SN follows power-law distribution (power-law probability distribution is a distribution whose density function has the form, for large values of x, and $P(X > x) \sim L(x)x^{-a+1}$) and has small-world properties. **Comparing with the popular degree distribution algorithms, our method can achieve 29.04% higher precision and 40.94% lower implementation time.**

The rest of the paper is organized as follows. In Section II, we review some related studies. Section III provides the basic network model and problem statement. In order to obtain degree distribution, Section IV presents mean-field-theory and spreading-coefficient based degree distribution algorithm. In Section V, extensive simulations are proposed to compare our MFSC with other existing methods. Finally, we conclude the paper in Section VI.

2. Related Work

Recently, the research on analyzing the degree in complex networks has drawn a great deal of attention, and various kinds of algorithms have been proposed. We have reviewed some classic and recent proposed algorithms for SN. According to the evolution of complex networks, these approaches can be categorized into two types.

In the first category, Albert R. *et al* [15] reviewed the recent advances in the field of complex networks, focusing on degree distribution in order to research the statistical mechanics of network topology and dynamics. After reviewing the empirical data that motivated the recent interest in networks, the authors discuss the main models and analytical tools, covering random graphs, small-world and scale-free networks, the emerging theory of evolving networks, and the interplay between topology and the network's robustness against failures and attacks. In addition, Barabási A-L *et al* [22] developed a mean-field method to predict the growth dynamics of the individual vertices, and use this to calculate the connectivity distribution and the scaling exponents. The mean-field method can be used to address the properties of two variants of the scale-free model, that do not display power-law scaling. He W. C. *et al* [34] also gave the differential format of master equation of degree distribution and obtain its analytical solution. The obtained result $P(k)$ is the time evolution of degree distribution. $P(k)$ is composed of two terms. At given finite time, one term decays exponentially, the other reflects size effect. At infinite time, the degree distribution is the same as that of Barabási and Albert.

Based on complex networks, in the second category, Lü L. Y. *et al* [16] provided link prediction with main equation (LPME) algorithm which can be used to extract missing information, identify spurious interactions, evaluate network evolving mechanisms, and so on. Link prediction can be expressed as the degree of nodes. This article summarizes recent progress about link prediction algorithms, emphasizing on the contributions from physical perspectives and approaches, such as the random-walk-based methods and the maximum likelihood methods. On the research of markov chain, they utilize master equation to deal with degree distribution. Krapivsky P. L. *et al* [17] investigated a variety of statistical properties associated with the number of distinct degrees that exist in a typical network for various classes of networks, which is distinct degree with main equation (DDME). For a single realization of a network with *N* nodes that is drawn from an ensemble in which the number of nodes of degree *k* has an algebraic tail, $N_k \sim N/K^v$ for $k \gg 1$, the number of distinct degrees grows as $N^{1/v}$. Such an algebraic growth is also observed in scientific citation data. Based on degree distribution, Jalili M. [21] investigated the effects of social power on the evolution of opinions in model networks as well as in a number of real SN, which is social power with mean field theory (SPMF). It is a newly degree distribution algorithm. A continuous opinion formation model is considered and the analysis is performed through numerical simulation. Social power is given to a proportion of agents selected either randomly or based on their degrees.

In previous studies, researchers can solve the problem of degree distribution by different kinds of methods. Nevertheless, in the real world, there exists one specific type of SN in which the close friends of a member will join the SN at the exponential growth time interval. This SN introduces a certain challenge to degree distribution, and so far no prior research effort can address it. This paper utilizes mean field theory continuity and spreading coefficient measurability to study degree distribution under this specific SN. We also solve the problem of complicated relationships in unique node growth model. Theoretical analysis and extensive

simulations both demonstrate that our method can achieve much better performance than existing methods in term of time complexity, precision and implementation time.

3. Network Model and Problem Statement

3.1 Network Model

In scale-free network average increment model, node generally joins network evenly. However, node cannot evenly join network in reality because of the effect of natural resources' condition, environmental quality and so on. He W. C. *et al* [34] discussed the normalization of degree distribution $P(k)$ in detail. They researched linear-growth model and achieved a good result. Our research is an extension of their model.

In this paper, we discuss one specific SN which has the feature of node growth time interval follows exponential distribution. As depicted in **Fig.1(a)**, it shows the social construction for SN. It is well known that SN consists of people and relationship between people. In addition, in the network graph, we utilize node and line instead of people and the relationship.

Firstly, we show the evolution of specific SN by simulated data.

In initial stage of SN community, there are 8 nodes (a, b, c, d, e, f, g and h). Each node has relation with only one other node in community, the degree of every node is 1. Therefore, there are 8 nodes and 4 edges in this SN. If the new node joins in, it should connect with 4 former node. The process of node joining is independent.

The first step is to add a node to this community, it has relation with 4 nodes in 8 nodes. The latest community has 9 nodes and 8 edges.

In the second step, joined node with its close node added to community together, they have relation with 4 nodes in 9 nodes respectively. The updated community has 11 nodes and 16 edges.

Gradually, more nodes added to the SN, and the number of joined nodes might be different. The number of edges rises with nodes joining, there is no edge between joined node and its close node.

Therefore, the new SN community is formed as bottom of **Fig. 1(a)** after the nth step. Each step of the evolution is shown in **Table 1**.

As illustrated in the **Table 1**, the number of nodes and edges are counted after new node joining. δ denotes the spreading coefficient, β denotes the parameter of Poisson distribution, m_{n-1} denotes the number of nodes at step n-1. The number of nodes and edges will be discussed in Section 4.1.
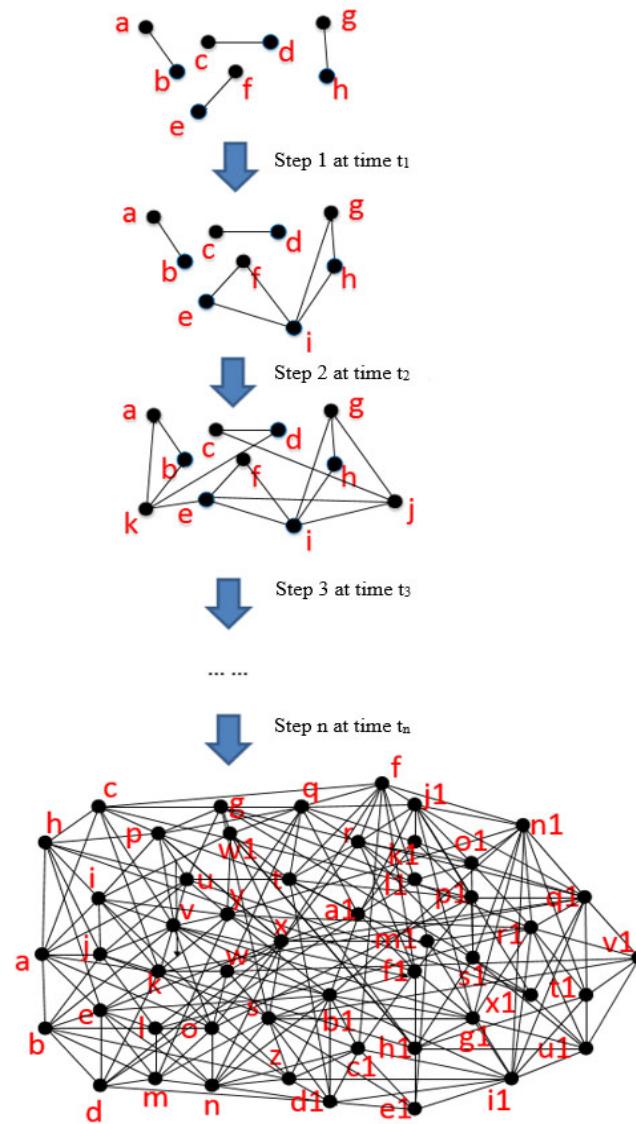
As shown in **Fig. 1(b)**, nodes added to the community with the number of times follows β parameter Poisson distribution per unit of time. The time interval between every step follows β parameter exponential distribution.

Then, we utilize web crawler to obtain the real-world data from Facebook. As one of the most popular game, there are more go clubs in Facebook, one of clubs is illustrated in **Fig. 2**. The node denotes member and the edge denotes the game between two members. If the new member joins in, he or she should play the game of go with 6 old members at the same time, it's the rule of this club. The process of member joining is independent. We analyze the formation and development of the club in detail.
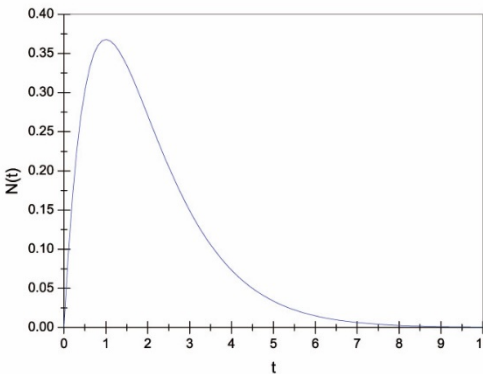
In the initial period of club, there were 30 members. Because of the game, each member played the game of go with only one other member in club at this time, the degree of every member is 1. Therefore, there were 30 members and 15 edges in this club, as shown in **Fig. 2(a)**.

At time t_1, one member joined in this club, he played the game of go with 6 members in 30 members. The latest club had 31 members and 21 edges, as indicated in **Fig. 2(b)**.

At time t_2, one member with her two close friends joined in club together, they not only played the game of go with 6 members in 31 members respectively, but also played game against each other. The updated club had 34 members and 42 edges, as depicted in **Fig. 2(c)**.
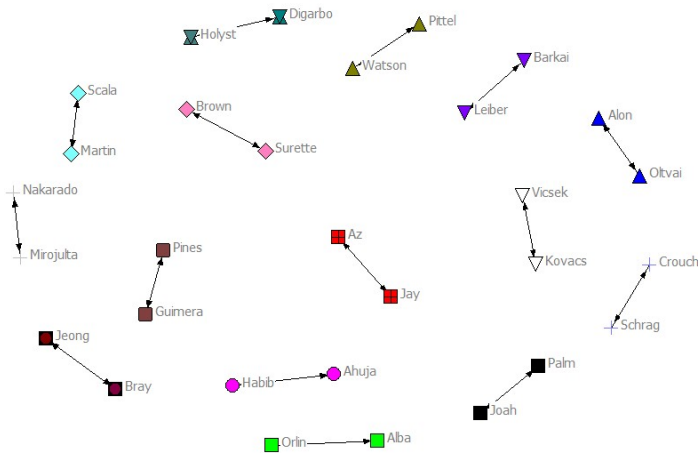

(a) The evolution of specific SN

(b) The probability of node's β times join in unit time follows Poisson distribution
Fig. 1. Node growth in specific SN

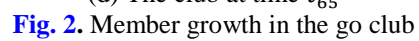**Table 1.** The number of nodes and edges at each step (the simulated data)

Step	Time	The number of joined nodes	The number of nodes	The number of edges
Step 0	t_0	0	8	4
Step 1	t_1	1	9	8
Step 2	t_2	2	11	16
…	…	…	…	…
Step n	t_n	$\delta\beta t_n$	$m_{n-1} + \delta\beta t_n$	$m_{n-1}/2 + 4\delta\beta t_n$

With time going on, more and more members joined in the club, the number of members and edges had been increasing rapidly in club. The number of added members may be different, and the game might happen between joined member and him or her friends.

At time t_{65}, one member with two of his close friends join in club together, they played the game of go with 6 members in 216 members respectively. This member played game with his friends, but his friends didn't play game with each other. The club had 219 members and 1206 edges, as indicated in **Fig. 2(d).**

(a) The initial period of club

(b) The club at time t_1

(c) The club at time t_2

(d) The club at time t_{65}

Fig. 2. Member growth in the go club

These changes for all aspects of the club are shown in **Table 2** at each time**.**

Table 2. The change of club at each time (the real-world data)

Time (t_i)	The number of joined members	Time interval ($t_i - t_{i-1}$)	The number of members	The number of edges
t_0(2016-10-12)	0	1	30	15
t_1(2016-10-13)	1	3	31	21
t_2(2016-10-16)	3	2	34	42
t_3(2016-10-18)	7	4	41	89
t_4(2016-10-22)	6	3	47	128
…	…	…	…	…
t_{61}(2017-03-31)	4	2	210	1147
t_{62}(2017-04-02)	3	5	213	1168
t_{63}(2017-04-07)	2	3	215	1180
t_{64}(2017-04-10)	1	2	216	1186
t_{65}(2017-04-12)	3	—	219	1206

As shown in **Table 2**, we have made the detail statistical analysis of data on the go club for half a year (the recorded data from 2016-10-12 to 2017-04-12). When the member join club, we recorded the time, the quantity of members and edges. The number of members added each time may be different. With more members joined the club, the number of edges were increasing rapidly. Because of the joined member might played the game of go with him or her friends, edges in the real-world data are more than the simulated data. According to our numbers, the real number of edges is 57 more than ideal value. In addition, the member may play with different player at different time. For example, Joah played the game of go with Palm during initial period, but he played the game with Crouch at time t_1.

Meanwhile, based on data analysis, the probability of members' β times join in unit time and the probability density of time interval of member joining are shown in **Fig. 3** and **Fig. 4**.
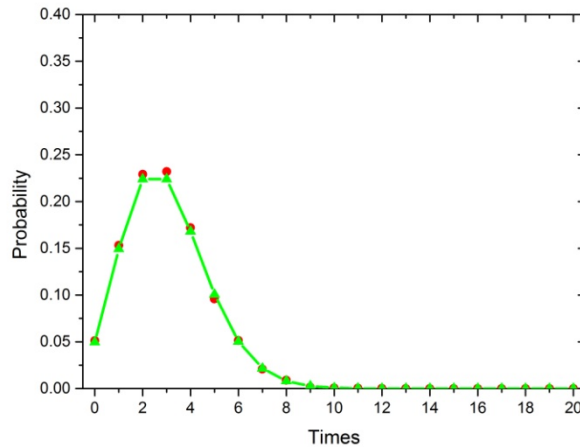

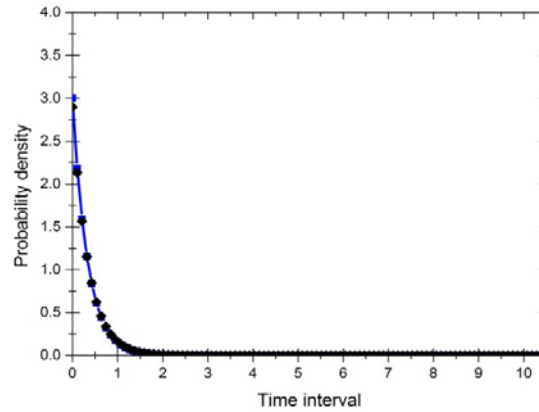**Fig. 3.** The probability of members' β times join in unit time

Fig. 4. The probability density of time interval of member joining

After passing through the analysis of a mass of recorded data, the probability of 3 members joining at unit time is high, whereas the other one is low. In **Fig. 3**, red circle curve denotes recorded value and green up triangle curve denotes β ($\beta = 3$) parameter Poisson distribution curve. The number of times member joining at unit time follows β ($\beta = 3$) parameter Poisson distribution because of two curves closely coincide with each other.

As shown in **Fig. 4**, black diamond curve denotes recorded value and blue square curve denotes γ ($\gamma = 3$) parameter exponential distribution curve. The time interval of member joining follows exponential distribution because of two curves nearly coincide with each other. Therefore, according to actual observed value, the time interval of member joining follows γ ($\gamma = 3$) parameter exponential distribution.

We have recorded the probability of 3 members joining at different unit time, as shown in **Fig. 5**.
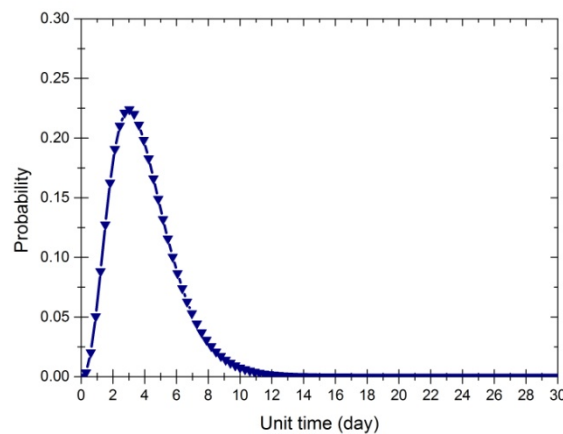

Fig. 5. Probability analysis of 3 members joining at different unit time

As depicted in **Fig. 5**, the navy down triangle curve denotes the probability of 3 members joining the club at different unit time. The probability initially grows when the unit time is less than 3 days, and decreases continuously after that. In a word, the probability of 3 members joining reaches the maximum at 3 days.

In conclusion, according to our research on the Facebook data, the time interval of member joining follows $\gamma = 3$ parameter exponential distribution, the probability of 3 members join club at 3 days is the highest. Therefore, this type of SN exists in the realistic society degree distribution analysis of this SN has certain theory significance and practical value.

3.2 Problem Statement

In the terms of online social networks (OSNs) [8], such as Facebook, Twitter or Google+, these changes are often introduced by users joining or leaving a particular group or community [9]. For example, friends connecting together, or new people making friend with each other. In SN, the community consists of social members who have close relationship. People tend to be friend with others in the same community, based on the basic assumption that a network system consists of a number of communities, among which the connections are much fewer than those inside the same community.

Let us take the fitness club in the university as an example. In the initiation stage of club, it has m_0 members and the degree of every member is 1 (member has only one connection with other member). Body builder or staff join in the club with close friends at time interval *t*, they have relations with m ($m \leq m_0$) members in former club. With club's advertising, popularity increase and word of mouth among people, the number of members is growing even faster. The growth of members follows exponential growth on the basis of massive social investigations. Therefore, the time interval of joining in club follows exponential distribution.

Under above circumstance, we must solve these problems: 1. how to describe the relationship between nodes? 2. how to count the number of joining nodes? 3. how to give the efficient degree distribution scheme?

For exploring the relationship among nodes, we present the definition of intimacy. The eligible nodes are chosen by intimacy and threshold value. In order to improve the accuracy of degree distribution, we have provided spreading coefficient to count the number of joining nodes in reality.

There have master equation method and mean field theory method to deal with the problem of degree distribution. Because of complicated relationships and unknown community structures, it is difficult to handle the problem about counting the number of joining nodes. The traditional method for degree distribution in a special SN is far from feasible.

We utilize mean field theory with spreading coefficient to settle the problem of degree distribution. At first, the probability of node degree has approximated to constant parameter and mean-field equation is described in differential equation form. Then, spreading coefficient is provided to mean-field equation for accurately depicting the change of degree distribution. At last, the approximate value of node degree probability is obtained by solving differential equations. The notations in this paper are shown in **Table 3**.

Table 3. Notations and their descriptions

Notations	Description	Notations	Description
$G = (V, E)$	this special SN	m_0	the number of initial node
m	the joining node connect with m nodes in initial node	β	the parameter of exponential distribution
Δt	the time interval of node joining	k_i	the degree of node *i*
$P(k_i)$	the probability of node *i* and new joining node	I_{il}^t	the intimacy value between node *i* and adjacent node *l* at time *t*

$N(t)$	count indicator	A_t	the intimacy set at time t
α	the threshold value which control the neighbor nodes whether they join in the community	s_k	the number of kth eligible joining nodes
δ	spreading coefficient	P_p	the precision of degree distribution
$p(k)$	the probability of node degree distribution	$P(k)$	the probability of SN degree distribution
t_T	the differences of the mean scores between test value and true value	Δk	the increment of joining nodes' degree

In the following, we present our method in detail.

4. Mean-Field-Theory and Spreading-Coefficient Based Degree Distribution Algorithm

In physics and probability theory, mean field theory studies the behavior of large and complex stochastic models by studying a simpler model [22,23]. Such models consider a large number of small individual components that interact with each other. The effect of all the other individuals on any given individual has approximated by a single averaged effect, thus reducing a many-body problem to a one-body problem. We utilize the features of mean field theory to obtain node degree distribution in the specific SN.

4.1 Exponential Growth Model

In order to study node degree distribution, we discuss exponential growth model in the following.

  A. Growth Pattern: There are m_0 nodes in SN (there are at least two persons as SN originates from interconnections among persons, we suppose m_0 is even number), the original degree value of node is one. There are several nodes join in and have a connection with m nodes in original network at interval t which is ruled by exponential distribution parameter is β.

  B. Connecting Mode: After joining new node, we denote $P(k_i)$ as connection probability of node i in original network and joining node. The relation between this node degree value k_i and $P(k_i)$ is given as

$$P(k_i) = \frac{k_i}{\sum_j k_j} \tag{1}$$

  C. Spreading Coefficient: The number of joining nodes have influenced by the relationship among nodes in a real-world situation, there are some close nodes join in together. We propose spreading coefficient δ in order to increase degree distribution accuracy. It has utilized to count the number of actual joining nodes in this SN model. This coefficient describes the number of extended joining nodes on the original basis.

  The count $N(t)$ follows a Poisson distribution on an interval [0,t], described as

$$P_{oisson}(N(t) = k) = \frac{(\beta t)^k}{k!} e^{-\beta t} \tag{2}$$

P_{oisson} denotes the probability of k nodes, β denotes the parameter of Poisson distribution. $N(t)$ denotes the number of nodes at time t and it utilized to count the number of nodes.

Then, the number of joining nodes also follow a Poisson distribution at every interval such that

$$P_{oisson}(N = k) = \frac{(\beta)^k}{k!} e^{-\beta} \tag{3}$$

In this way, there are $\delta\beta t$ nodes join in this network at time t. The network has $m_0 + \delta\beta t$ nodes at this time. The original SN has $m_0/2$ edges. The joined nodes have relation with m nodes respectively. A total of $m\delta\beta t$ edges added to the network. Therefore, the network has $m_0/2 + m\delta\beta t$ edges at time t.

4.2 Spreading Coefficient Designed for Accurate Degree Distribution

Degree distribution has studied by growth of nodes. In the specific real-world SN, the number of social members can change with the factors. The intimate relationship between social members is one of the most important factors and it will influence the research of degree distribution.

We give an example to explain the meaning of the factor. Assuming there exist a community in SN. There are M members which have a great relationship with joining member i, they will join in this community with member i. Then, the number of joining nodes will be $1+M$ and not 1. We design a new parameter to embody the character of node growth with intimate relationship, called spreading coefficient, which determines the rate of node growth in SN and increases degree distribution accuracy. We utilize spreading coefficient to count the number of joining nodes in reality, which has determined by intimacy value and threshold value.

In order to further describe internal relationship in SN, we present the definition of intimacy [28]. Specifically, we utilize the value to describe the intimate relation between nodes and the intimate relation between communities. It expresses the degree of close relationship among them. To obtain the precise refinement of relationship among each other, we add several factors of close relation to traditional weighted value. Two nodes have closer relations if they have bigger intimacy.

We can obtain the set of intimacy by intimacy, it can be expressed as

$$A_t = \left(I_{il}^t \right)_{1 \times d} \tag{4}$$

A_t denotes the intimacy set at time t, I_{il}^t denotes the intimacy value between node i and adjacent node l at time t, d denotes the number of adjacent nodes.

Intimacy describes the relationship among nodes, it consists of family affection, friendship and hobby similarity. Firstly, we obtain the intimacy set A_t by using family affection set, friendship set and hobby similarity set between initial nodes in SN.

Jiang J. *et al* [14] repeatedly crawled users in the Peking University Renren network over a period of 90 days, extracting profile visit history for 61K users, and examining issues of popularity, visitor composition, reciprocity, and latency of reciprocation. They compared user popularity distributions for latent and visible interactions, and use per-object visit counters to quantify the level of user engagement generated from user profiles, photos, and diary entries. They also studied correlation of family affection content with a user's profile popularity. Li Z. *et al* [24] showed that the measurement of social similarity based on the number of common communities has an imperfection for some cases, i.e., it ignores the fact that the members within the same community usually have different levels of local activity, which possibly results in a potentially low efficiency in terms of deliver ratio and latency due to the misalignment estimation of nodes' contact probability. Node local activity is a statistics of

encounter probability in a node's communities. Based on node local activity, they studied friendly degree in order to calculate the probability of node connecting. Obradović D. *et al* [25] firstly tried to aggregate as many articles of the domain as possible. Then, they derived a meaningful measure of social authority, based on links among blogs and articles. They utilized more articles and links to research hobby similarity of author in the Blogosphere. Finally, they enabled the approach to work over very long time periods of monitoring.

By previous studies, family affection degree set [14], friendly degree set [24] and hobby similarity set [25] are obtained in SN, and the expression of intimacy set is

$$A_t = \mu T_t + \varphi F_t + \gamma H_t \tag{5}$$

A_t denotes the intimacy degree set, T_t denotes the family affection degree set, F_t denotes the friendly degree set, H_t denotes the hobby similarity set. We will analysis the influence to community division with different μ, φ and γ in the simulation experiment.

At first, we determine nodes that can go along with joined node and of the form

$$\begin{cases} I_{il}^t > \alpha, l \in [1, G] \Rightarrow l \in c \\ I_{il}^t < \alpha, l \in [1, G] \Rightarrow l \notin c \end{cases} \tag{6}$$

The number of nodes which neighboring to join node i is G. α denotes threshold value which control the neighbor nodes whether they join in the community c. If the intimacy between joined node i and its neighbor node l is greater than threshold value α, node l can join in this community; otherwise can not join in this community.

Then, all of eligible nodes are collected and can be written as

$$s = size(C) \tag{7}$$

We denote eligible node as the node that can join in this SN together with joined node. C denotes the set of eligible nodes, *size()* denotes the function which counts the number of elements in an array C.

Then, the expression of nth spreading coefficient can be obtained as

$$\delta_n = 1 + s_n \tag{8}$$

s_n denotes the number of nth eligible joined nodes.

Due to the effect of all the other individuals on any given individual has approximated by a single averaged effect in mean field theory, total spreading coefficient is obtained by setting

$$\delta = \frac{\sum_{n=1}^{N} \delta_n}{N} \tag{9}$$

N denotes the number of add nodes to the community.

At last, the description of our method has presented in ISC Algorithm. *Comm* denotes the set of network nodes, δ denotes spreading coefficient. We utilize NMI overlapping version [27] as metrics to ensure μ, φ and γ in intimacy and threshold value. It is one of the most important entropy measures in information theory. The higher the NMI score is, the more similar the two community partitions are. If it equals 1, it means the two kinds of community partitions are exactly coincident.

In ISC algorithm, I_{max}^t denotes maximum of A at time t, I_{min}^t denotes minimum of A at time t, C_t^{l+t} denotes the array which merge eligible nodes at time t. Firstly, the intimacy of joined node and its neighboring nodes are computed. Then, the eligible nodes are selected by comparing their intimacy with threshold value. Finally, we count the number of these nodes.

	ISC Algorithm $\quad \delta=\text{ISC}(Comm)$
1:	**Input** $\quad$ Community Set *Comm*;
2:	**Output** $\quad$ Spreading Coefficient δ;
3:	$A_t = \mu T_t + \varphi F_t + \gamma H_t$;
4:	$I^t_{max} = max(A_t)$;
	$I^t_{min} = min(A_t)$;
5:	$\alpha_t = \frac{I^t_{max}+I^t_{min}}{2} : I^t_{max}$
6:	$l = 1:G$
7:	**if** $I^t_l \geq \alpha^t$ **then**
8:	$\quad C^{l+t}_t = combine\{C^l_t, l\}$
9:	**else**
10:	$\quad l{+}{+}$
11:	$\delta_n = 1 + size(C_t)$
	$\delta = \frac{\sum^N_{n=1}\delta_n}{N}$
12:	Repeat step 4 to step 11 in order to maximize NMI and obtain δ
13:	Save spreading coefficient to δ and back.

4.3 MFSC Algorithm

Based on spreading coefficient, we calculate degree distribution by utilizing mean field theory and nodes' growth pattern in the specific SN. Its main idea assumes that degree value k is continuous. Because of adding new nodes to the SN in the time interval of exponential distribution, the degree distribution of node closely related to time t. We build differential equation for node's degree k_i and time t as

$$\frac{\partial k_i}{\partial t} = a \cdot P(k_i) = a\frac{k_i}{\sum_j k_j} \tag{10}$$

Because of $\sum_j k_j = m_0 + 2m\delta\beta t$ and this SN has δ joined nodes in per interval, the increment of joined nodes' degree is obtained as $\Delta k = m\delta\beta$. Therefore, $a = m\delta\beta$ is given by

$$\frac{\partial k_i}{\partial t} = \frac{m\delta\beta k_i}{m_0+2m\delta\beta t} \approx \frac{k_i}{2t} \tag{11}$$

Because node i joined this SN at time t_i, connectivity degree of this node is $k_i(t_i) = m$. Node connectivity degree substituted into formula (11) and

$$k_i(t_i) = m\left(\frac{t}{t_i}\right)^{\frac{1}{2}} \tag{12}$$

Degree distribution of this SN is obtained as

$$p(k_i(t_i) < k) = p\left(t_i > \left(\frac{m}{k}\right)^2 \cdot t\right) \tag{13}$$

According to network expanding assumption, we can calculate probability density of t_i as

$$p_i(t_i) = \delta\beta e^{-\delta\beta t_i} \tag{14}$$

Then we can write the solution of formula (14) as

$$p\left(t_i > \left(\frac{m}{k}\right)^2 \cdot t\right) = \int_1^{t_i} \bar{p}\left(t_i > \left(\frac{m}{k}\right)^2 \cdot t\right) dt$$

$$= \int_1^{t_i} \frac{\delta\beta}{m_0+\delta\beta t} p\left(t_i > \left(\frac{m}{k}\right)^2 \cdot t\right) dt$$

$$= \frac{\delta\beta}{m_0+\delta\beta t}\left(1 - p\left(t_i \leq \left(\frac{m}{k}\right)^2 \cdot t\right)\right)$$

$$= \frac{\delta\beta}{m_0 + \delta\beta t} e^{-\delta\beta\left(\frac{m}{k}\right)^2 \cdot t} \tag{15}$$

The differential of formula (15) is

$$P(k) = \frac{\partial p(k_i(t) < k)}{\partial k}$$

$$= \frac{\delta\beta}{m_0 + \delta\beta t} e^{-\delta\beta\left(\frac{m}{k}\right)^2 \cdot t} \cdot 2\delta\beta t m^2 k^{-3} \tag{16}$$

As $k \gg m$ and $t \to \infty$, approximate value is derived by

$$P(k) = 2\delta\beta m^2 k^{-3} \tag{17}$$

After above computing, the conclusion of our method follows power-law distribution. We ingeniously utilize the characteristic of continuity in mean-field theory to simplify computation step.

On account of preparatory work, we can obtain the MFSC algorithm for degree distribution in this specific SN. In MFSC algorithm, *Comm* denotes the set of network nodes, *P* denotes degree distribution.

MFSC Algorithm   *P*=MFSC(*Comm*)

1:    **Input**   Community Set *Comm*;
2:    **Output**   Degree Distribution *P*;
3:    $\delta = \text{ISC}(Comm)$;
4:    $\frac{\partial k_i}{\partial t} = a \frac{k_i}{\sum_j k_j}$;
5:    $a = m\delta\beta$;
6:    $k_i(t_i) = m \left(\frac{t}{t_i}\right)^{\frac{1}{2}}$;
7:    $p(k_i(t_i) < k) = p\left(t_i > \left(\frac{m}{k}\right)^2 \cdot t\right)$;
      $p\left(t_i > \left(\frac{m}{k}\right)^2 \cdot t\right) = \frac{\delta\beta}{m_0 + \delta\beta t} e^{-\delta\beta\left(\frac{m}{k}\right)^2 \cdot t}$;
8:    $P(k) = 2\delta\beta m^2 k^{-3}$;
9:    Save degree distribution to *P* and back.

As depicted in MFSC algorithm, we utilize ISC algorithm to compute spreading coefficient on the foundation of intimacy among nodes firstly. Then, we use the improved mean field theory to do some research about node degree. Gradually, we obtain the degree distribution of this SN at last.

5. Performance Evaluation

In our paper, we implemented the proposed new scheme MFSC. Extensive simulations on a specific SN dataset illustrate the performance of MFSC. To compare the performance of MFSC to that of the existing algorithm, we utilize the other classic methods, such as SPMF [21], DDME [17] and LPME [16]. The comparative results show that our approach is efficient.

The experiment is divided into three parts. The first part is verified experiment, NMI and real-world dataset are utilized to verify the feasibility of the algorithm. Then, we provide degree distribution analysis with different exponential growth coefficient β and the degree of joined nodes *m* in second part. Finally, at last part of the experiment, we contrast our method to

existing method in time complexity, space complexity, precision and implementation time respectively.

5.1 Experiment Settings

A. Experimental Config and Dataset
The simulations are performed on DELL server (Intel Xeon E5-2603 1.80Ghz, 32G, 1T) by network analysis platform (UCINET) and numerical analysis platform (MATLAB).

Based on the dataset from the Pretty Good Privacy (PGP) network [26], we obtain the evaluation results of MFSC. It is a network based on trust, and the comprehension of trust networks is, nowadays, crucial to understand the complexity of the information society. It is composed of 10680 users, the initial degree value of every user is one (only one other user has relationship with this user). One user with his or her closest friend join in this SN in time *t*, new users have relationship with *m* existing users. The time *t* follows exponential distribution.

We analyze various aspects of dataset, and that includes the number of joined users, the number of added edges, the interval of user joining, the number of all users and the number of all edges in each time point. In Section 5.2, we study the number of joined users to obtain the coefficients of intimacy and spreading coefficient. Degree distribution changes with the number of added edges and the interval of user joining changes in Section 5.3.

B. Compared Algorithms
To compare the performance of MFSC to that of the existing algorithm, we utilize three other classic methods with the following algorithms as the baseline.

LPME. Lü *et al* [16] discussed link prediction model for complex network. They summarized recent progress about link prediction algorithms, emphasizing on the contributions from physical perspectives and approaches, such as the random-walk-based methods and the maximum likelihood methods.
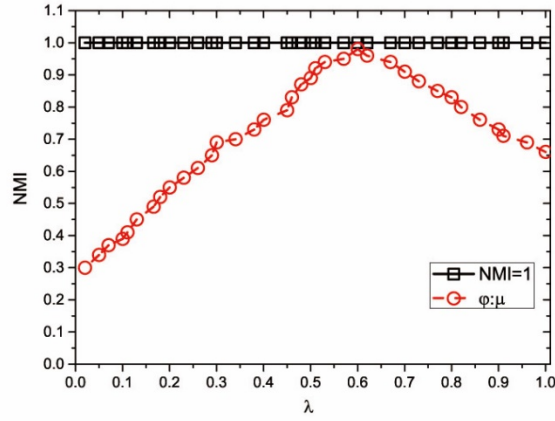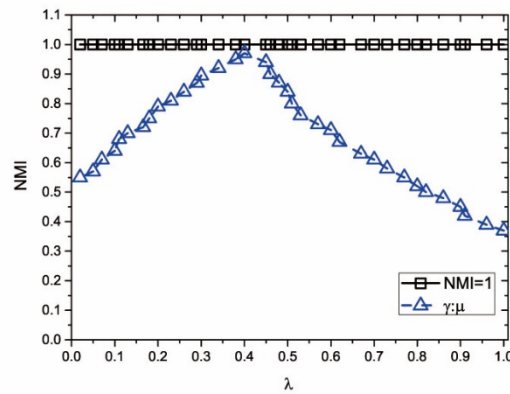
DDME. Krapivsky *et al* [17] researched a variety of statistical properties associated with the number of distinct degrees that exist in a typical network for various classes of networks. For a single realization of a network with N nodes that is drawn from an ensemble in which the number of nodes of degree k has an algebraic tail, $N_k \sim N/K^v$ for $k \gg 1$, the number of distinct degrees grows as $N^{1/v}$.

SPMF. Jalili [21] investigated the effects of social power on the evolution of opinions in model networks as well as in a number of real SN. It is a newly degree distribution algorithm. A continuous opinion formation model is considered and the analysis is performed through numerical simulation. Social power is given to a proportion of agents selected either randomly or based on their degrees.

5.2 The Verification of Power-Law Distribution

We utilize Normalized Mutual Information (NMI) overlapping version [36] as metrics to make sure μ, φ and γ in intimacy and threshold value α. It is one of the most important entropy measures in information theory. The higher the NMI score is, the more closely to the actual situation. If it equals 1, it conforms to the actual conditions.

Firstly, we get the intimacy coefficient through NMI score by **Fig. 6**.

(a) The ratio of φ and μ

(b) The ratio of γ and μ

Fig. 6. Parameter choosing for μ, φ and γ

λ(the specific value of coefficient in intimacy) in **Fig. 6(a)** shows the value of $\varphi:\mu$, λ in **Fig. 6(b)** shows the value of $\gamma:\mu$. NMI score changes as the parameter λ changes. When $\varphi:\mu = 0.6$, $\gamma:\mu = 0.4$, NMI score reach the maximum 1 respectively. Therefore, as $\mu:\varphi:\gamma = 5:3:2$, NMI score attain the maximum 1. We confirm μ, φ and γ in intimacy.

Then, from large numbers of tests, the threshold value α can be obtained ranging from 0.71-0.89. We select the representative values 0.78-0.87 to analyze an appropriate value for α. Because in this scope, the NMI score shows better than in other scopes. The simulation results are shown in **Fig. 7**.

NMI score changes as the parameter ω changes (the larger the ω is, the weaker the node strength is in its community). As $\alpha = 0.82$, NMI score will level off to 1, the threshold value is confirmed in this SN.
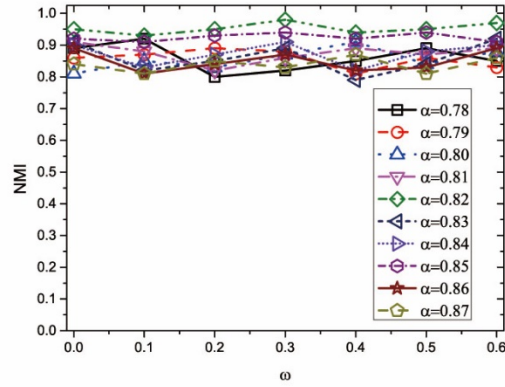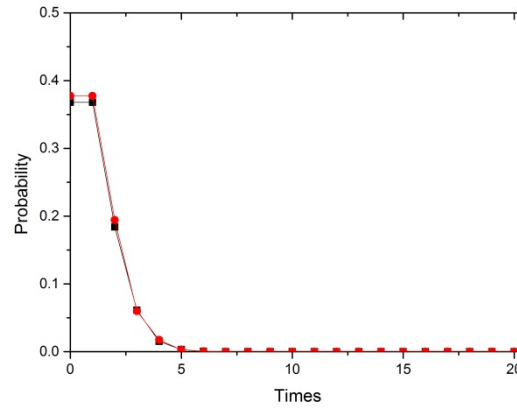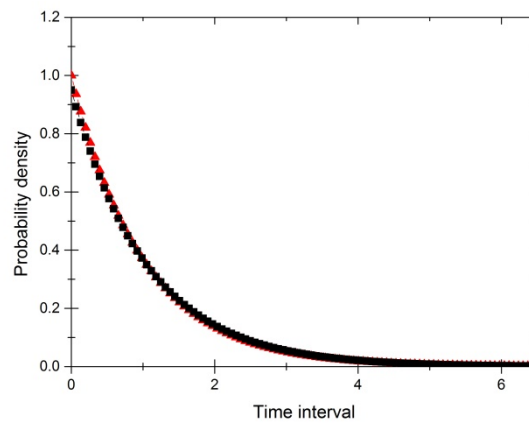
Fig. 7. Parameter choosing for threshold value

(a) The probability of members' β times join in unit time

(b) The probability density of time interval of member joining
Fig. 8. Network actual observed value

All of the parameters are obtained based on numerous simulations and we then calculate the spreading coefficient by our ISC algorithm. The spreading coefficient is obtained as $\delta = 15.26$. At last, node degree distribution has studied on the condition that the number of nodes is continuously increasing.

Meanwhile, based on long-term observational data, the probability of members' β times join in unit time and the probability density of time interval of member joining are shown in **Fig. 8**.

After pass through analyze a mass of recorded data, the probability of one time member joining at unit time is high, the other is low. In **Fig. 8(a)**, black square curve denotes recorded value and red circle curve denotes Poisson distribution curve ($\beta = 1$). The number of times member joining at unit time follows Poisson distribution ($\beta = 1$) because of two curves closely coincide with each other. Then, we study on the probability of time interval of member joining. As shown in **Fig. 8(b)**, black square curve denotes recorded value and red triangle curve denotes exponential distribution curve ($\gamma = 1$). The time interval of member joining follows exponential distribution ($\gamma = 1$) because of two curves coincide with each other. Therefore, according to actual observed value, the time interval of member joining follows γ parameter exponential distribution ($\gamma = 1$).

We have recorded the probability of one time member joining at different unit time every month in one year, as shown in **Fig. 9**.
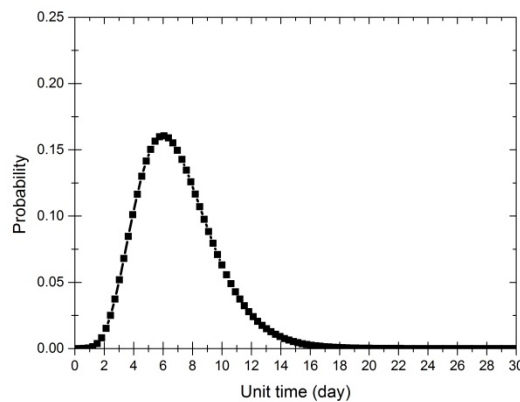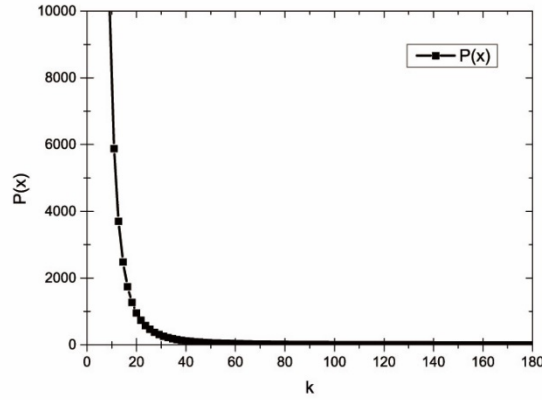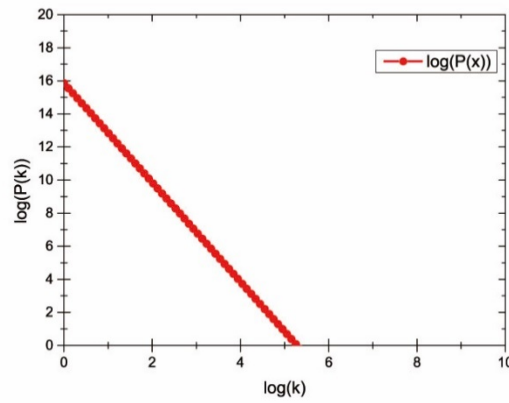

Fig. 9. Probability analysis of member joining at different unit time

As depicted in **Fig. 9**, the probability grows before 6, and decreases continuously after that. In a word, the probability of one time member joining reaches the maximum at 6. Therefore, this specific SN has one time member joining on average 6 days.

(a) Degree distribution for PGP network

(b) The logarithm of degree distribution
Fig. 10. Degree distribution for PGP network

The degree distribution for this SN is shown in **Fig. 10** on the condition that $\beta = 1$ and $m = 500$.

In **Fig. 10(a)**, degree distribution is got for this SN follows power-law distribution that a very few nodes have larger connections and the majority of nodes have fewer connections. It accords with the character of scale-free network. Meanwhile, user can meets the others through several users on the average in the SN. It reflects that PGP network has small-world properties. By evaluating the logarithm of formula (17), we can obtain **Fig. 10(b)**. It also verifies that this SN has both scale-free and small-world features by the result of a straight line.

5.3 Contrastive Analysis for Changing Parameter

We obtain the change of results through change exponential growth coefficient β and the degree of joined nodes m.

A. Different β and Same m: We change the exponential growth coefficient under the same $m = 500$ by **Fig. 11**.

In **Fig. 11**, the connection probability of nodes will increase with increasing exponential growth coefficient. Because of the time interval growth means add more nodes in the same

time, so the connection probability of nodes is increasing. It is consistent with the actual situation.
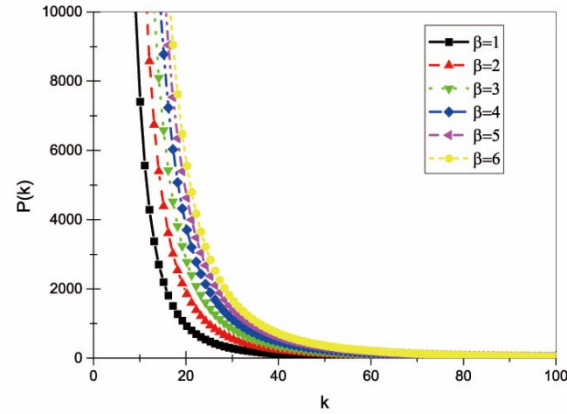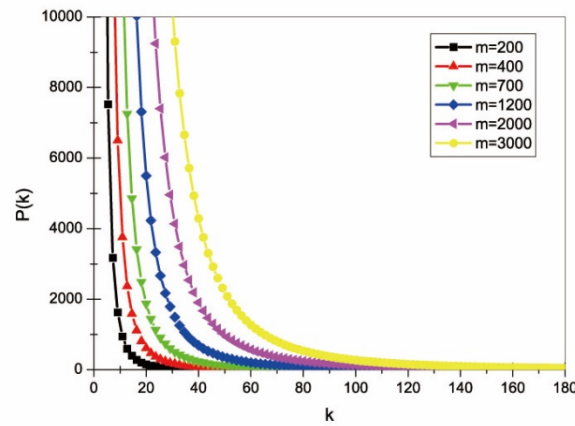

Fig. 11. Degree distribution by different β and same m

Fig. 12. Degree distribution by same β and different m

B. Same β and Different m: We change the degree of joined nodes under the same $\beta = 1$ according to **Fig. 12**.

In **Fig. 12**, the connection probability of nodes will increase with increasing degree of joined nodes. Because of *m* growth means the number of joined nodes' edge is increasing, so the connection probability of nodes is increasing. It is consistent with the actual situation.

5.4 Simulation Results and Analysis

In addition to simulations, we also compare our algorithm with existing algorithms through theoretical analysis.

Firstly, we utilize student's t-test [37] to find out whether the differences of the mean scores between test value and true value are significant. A t-test is any statistical hypothesis test in which the test statistic follows a student's t-distribution under the null hypothesis. It can be used to determine if two sets of data are significantly different from each other. A t-test is most commonly applied when the test statistic would follow a normal distribution if the value of a scaling term in the test statistic were known. When the scaling term is unknown and replaced

by an estimate based on the data, the test statistics (under certain conditions) follow a student's t distribution.

In testing the null hypothesis that the population mean is equal to a specified value ρ_0, one uses the statistic

$$t_T = \frac{\bar{x} - \rho_0}{s/\sqrt{a}} \tag{18}$$

where $\bar{x}$ is the sample mean, s is the sample standard deviation of the sample and a is the sample size. The degrees of freedom used in this test are a-1. Although the parent population does not need to be normally distributed, the distribution of the population of sample means, $\bar{x}$, is assumed to be normal. By the central limit theorem, if the sampling of the parent population is independent and the first moment of the parent population exists then the sample means will be approximately normal.

The degree distribution of this SN changes over time. Compare to real degree distribution, t_T of MFSC, SPMF [21], DDME [17] and LPME [16] is 0.948221, 0.822435, 0.754187 and 0.687728 respectively. Therefore, the computed result of our method in line with true environment approximately.

Moreover, the precision of degree distribution is defined as:

$$P_p = \frac{P(k)_T}{P(k)_R} = \frac{2\delta\beta m^2 k^{-3}}{P(k)_R} \tag{19}$$

P_p denotes the precision of degree distribution, $P(k)_T$ denotes the calculated value, $P(k)_R$ denotes the true value. The more consistent with the real world, the higher P_p is.

In formula (19), δ depends on μ, φ, γ and α. As μ, φ and γ changes, NMI score changes. If NMI score equal to 1, it conforms to the actual conditions, and then δ is accordant with the actual state. For this reason, if μ, φ and γ makes NMI score approaches 1, P_p higher. If μ, φ and γ makes NMI score keeps away from 1, P_p lower. It is same to α. β and m are determined by experimental data. The comparison of algorithms precision is depicted in **Fig. 13**.

In **Fig. 13**, the abscissa stands for time and the ordinate stands for precision. We analyze the precision of every algorithm at different time points. Because of accurate analysis of intimacy between users, we obtain subtle calculation of the number of joined users. With the increase of test time, our results basically conform with reality. On the contrary, the existing algorithms have not considered close nodes, algorithm precision decreased down gradually. Obviously, the precision of MFSC performs better than other algorithms. On average, our method is higher than SPMF with 29.04%, DDME with 44.34% and LPME with 86.71% on the precision.
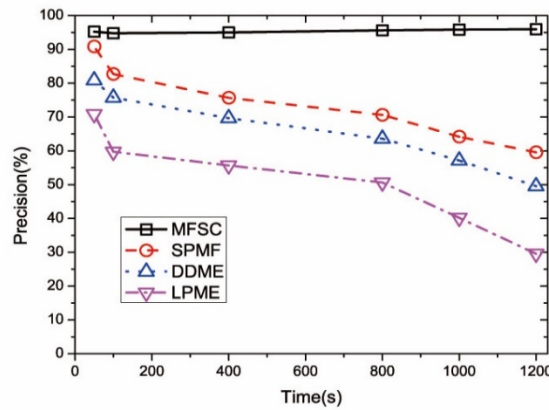

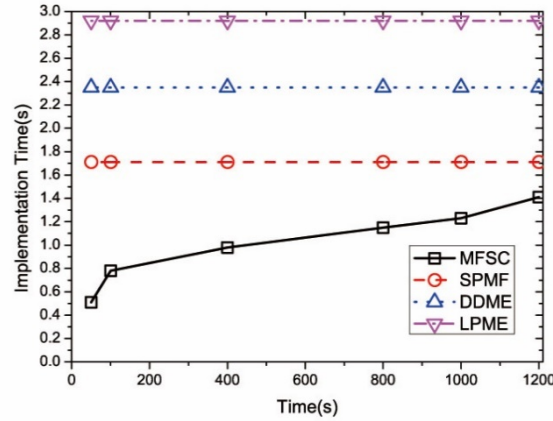**Fig. 13.** The comparison of algorithms precision

Fig. 14. The comparison of algorithms implementation time

Then, we obtain the comparison of algorithms implementation time. In this paper, we record the fired time of CPU that changes as the achievement of algorithm's program changes. The dataset have been tested for 10 times, and the average value is shown in **Fig. 14**.

As shown in **Fig. 14**, the abscissa stands for time and the ordinate stands for implementation time. In this paper, we study the implementation time for each method at every time. The implementation time of MFSC is increasing as this method needs to compute spreading coefficient. Although our method implementation time keeps increasing, we utilize optimized mean-field theory and reduce computation steps by using average effect instead of total effect. Therefore, the implementation time of our algorithm is always lower than other algorithms. We have measured that MFSC is lower than SPMF with 40.94%, DDME with 57.02% and LPME with 65.41% on the implementation time.

At last, we compare MFSC with existing algorithms for time complexity and space complexity. Time complexity of MFSC, SPMF, DDME and LPME is $O(n)$, $O(nlog_2 n)$, $2 \cdot O(nlog_2 n)$ and $4 \cdot O(nlog_2 n)$ respectively. Because of $O(n) < O(nlog_2 n) < 2 \cdot O(nlog_2 n) < 4 \cdot O(nlog_2 n)$, our algorithm is lower than other algorithms in time complexity by contrastive analysis. In addition, the space complexity of MFSC, SPMF, DDME and LPME is $O(1)$, $O(log_2 n)$, $2 \cdot O(log_2 n)$ and $4 \cdot O(log_2 n)$ separately. Therefore, the space complexity of our algorithm is lowest. The time complexity and space complexity of four algorithms are listed in **Table 4.**

Table 4. The time complexity and space complexity of four algorithms

Algorithm	The time complexity	The space complexity
MFSC	$O(n)$	$O(1)$
SPMF	$O(nlog_2 n)$	$O(log_2 n)$
DDME	$2 \cdot O(nlog_2 n)$	$2 \cdot O(log_2 n)$
LPME	$4 \cdot O(nlog_2 n)$	$4 \cdot O(log_2 n)$

In conclusion, our method performs better than other algorithms. It can achieve 29.04% higher precision and 40.94% lower implementation time.

6. Conclusion

In this paper, we propose a new degree distribution scheme for SN based on spreading coefficient, which describes the relationship among nodes and reflects network size. We utilize mean field equation with spreading coefficient to handle the problem, and conduct simulations on the UCINET analysis platform to evaluate MFSC's performance. Extensive simulations on the specific SN show that it can achieve a very good degree distribution performance in an efficient way, which can perform 29.04% higher precision and 40.94% lower implementation time. To the best of our knowledge, our work is the first one to achieve degree distribution with spreading coefficient.

7. Acknowledgement

References

[1]   Amaral L. A. N. and Ottino J. M., "Complex networks: Augmenting the framework for the study of complex systems," *The European physical journal B*, vol. 38, no. 2, pp. 147-162, March 2004. Article (CrossRef Link).

[2]   J. Leskovec, "SNAP: Stanford Network Analysis Project," February 2014. *Article (CrossRef Link)*.

[3]   Watts D. J. and Strongatz S. H., "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440-442, April 1998. Article (CrossRef Link).

[4]   M. Müller and D. W. Sun, "Directing the Self-Assembly of Block Copolymers into A Metastable Complex Network Phase via A Deep and Rapid Quench," *Physical Review Letters*, vol. 111, no. 26, pp. 268101, December 2013. Article (CrossRef Link).

[5]   Y. Li and C. Zheng, "The complex network synchronization via chaos control nodes," *Journal of Applied Mathematics*, vol. 2013, no. 2013, pp. 1-11, 2013. Article (CrossRef Link).

[6]   Barabasi A. L. and Albert R., "Emergence of Scaling in Random Networks," *Science*, vol. 286, no. 5439, pp. 509-512, October 1999. Article (CrossRef Link).

[7]   J. Z. Ji, X. J. Song, C. N. Liu, and X. Z. Zhang, "Ant colony clustering with fitness perception and pheromone diffusion for community detection in complex networks," *Physical A: Statistical Mechanics and its Applications*, vol. 392, no. 15, pp. 3260-3272, August 2013. Article (CrossRef Link).

[8]   Golder S., Wilkinson D., and Huberman B., "Rhythms of Social Interaction: Messaging Within a Massive Online Network," in *Proc. of Proceedings of the 3rd International Conference on Communities and Technologies*, pp. 1-16, 2007. Article (CrossRef Link).

[9]   Meyers L. A., Pourbohloul, Newman M. E. J. and et al, "Network theory and SARS: predicting outbreak diversity," *Journal of Theoretical Biology*, vol. 232, no. 1, pp. 71-81, January 2005. Article (CrossRef Link).

[10] G. M. Park, S. H. Kim, and H. G. Cho, "Structural Analysis on Social Network Constructed from Characters in Literature Texts," *Journal of Computers*, vol. 8, no. 9, pp. 2442-2447, September 2013. Article (CrossRef Link).

[11] X. W. Yang, Y. Guo, and Y. Liu, "Bayesian-Inference-Based Recommendation in Online Social Networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, no. 4, pp. 642-651, July 2013. Article (CrossRef Link).

[12] Z. Wang, L. F. Sun, W. W. Zhu and et al, "Joint Social and Content Recommendation for User-Generated Videos in Online Social Network," *IEEE Transactions on Multimedia*, vol. 15, no. 3, pp. 698-709, April 2013. Article (CrossRef Link).

[13] Keeling M. J. and Eames K. T. D., "Networks and epidemic models," *Journal of the royal society interface*, vol. 2, no. 4, pp. 295-307, June 2005. Article (CrossRef Link).

[14] J. Jiang, C. Wilson, X. Wang and et al, "Understanding Latent Interactions in Online Social Networks," *ACM Transactions on the Web*, vol. 7, no. 4, pp. 369-382, October 2013. Article (CrossRef Link).

[15] Albert R. and Barabási A. L., "Statistical mechanics of complex network," *Review of Modern Physics*, vol. 74, no. 1, pp. 47-97, June 2002. Article (CrossRef Link).

[16] L. Y. Lü and T. Zhou, "Link prediction in complex networks: A survey," *Physica A: Statistical Mechanics and its Applications*, vol. 390, no. 6, pp. 1150-1170, March 2011. Article (CrossRef Link).

[17] Krapivsky P. L. and Redner S., "Distinct degrees and their distribution in complex networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2013, no. 6, pp. 1-12, April 2013. Article (CrossRef Link).

[18] Newman M. E. J., "The structure and function of complex network," *SIAM Rev*, vol. 45, no. 2, pp. 167-256, August 2003. Article (CrossRef Link).

[19] Mislove A., Marcon M., Gummadi K. P. and et al, "Measurement and Analysis of Online Social Networks," in *Proc. of Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement*, pp. 29-42, October 24-26, 2007. Article (CrossRef Link).

[20] A. Dwivedi and X. H. Yu, "A Maximum-Flow-Based Complex Network Approach for Power System Vulnerability Analysis," *IEEE Transactions on Industrial Informatics*, vol. 9, no. 1, pp. 81-88, February 2013. Article (CrossRef Link).

[21] Mahdi Jalili, "Social power and opinion formation in complex networks," *Physica A: Statistical Mechanics and its Applications*, vol. 392, no. 4, pp. 959-966, February 2013. Article (CrossRef Link).

[22] Barabasi A. L., R. Albert, and H. Jeong, "Mean-field theory for scale-free random networks," *Physical A*, vol. 272, no. 1-2, pp. 173-187, July 1999. Article (CrossRef Link).

[23] R. Albert and Barabasi A. L., "Topology of Evolving Networks: Local Events and Universality," *Phys.Rev.Lett*, vol. 85, no. 24, pp. 5234-5237, May 2000. Article (CrossRef Link).

[24] Z. Li, C. Wang, S. Q. Yang and et al, "LASS: Local-Activity and Social-Similarity Based Data Forwarding in Mobile Social Networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 26, no. 1, pp. 1-15, January 2015. Article (CrossRef Link).

[25] Darko Obradovic, Stephan Baumann, and Andreas Dengel, "A social network analysis and mining methodology for the monitoring of specific domains in the blogosphere," *Social Network Analysis and Mining*, vol. 3, no. 2, pp. 221-232, June 2013. Article (CrossRef Link).

[26] M. Boguñá, R. Pastor-Satorras, A. Diaz-Guilera and A. Arenas, "Models of social networks based on social distance attachment," *Physical Review E*, vol. 70, no. 056122, pp. 1-8, November 2004. Article (CrossRef Link).

[27] A. Lancichinetti, S. Fortunato and J. Kertész, "Detecting the over-lapping and hierarchical community structure in complex networks," *New Journal of Physics*, vol. 11, pp. 033015, March 2009. Article (CrossRef Link).

[28] Y. Zheng, D. F. Zhang, and K. Xie, "An Intimacy-based Algorithm for Social Network Community Detection," in *Proc. of The 15th International Conference on Algorithms and Architectures for Parallel Processing*, pp. 764-776, August 2015. Article (CrossRef Link).

[29] X. Yang, H. Steck, and Y. Liu, "Circle-based recommendation in online social networks," in *Proc. of KDD*, pp. 1267-1275, 2012. Article (CrossRef Link).

[30] G. Zhao, X. Qian, and X. Xie, "User-Service Rating Prediction by Exploring Social Users' Rating Behaviors," *IEEE Trans. Multimedia*, vol. 18, no. 3, pp. 496-506, 2016. Article (CrossRef Link).

[31] G. Song, Y. Li, X. Chen, X. He, and J. Tang, "Influential Node Tracking on Dynamic Social Network: An Interchange Greedy Approach," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 2, pp. 359-372, 2017. Article (CrossRef Link).

[32] G. Zhao, X. Qian, X. Lei, and T. Mei, "Service Quality Evaluation by Exploring Social Users' Contextual Information," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 12, pp. 3382-3394, 2016. Article (CrossRef Link).

[33] X. Qian, H. Feng, G. Zhao, and T. Mei, "Personalized Recommendation Combining User Interest and Social Circle," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 7, pp. 1763-1777, 2014. Article (CrossRef Link).

[34] W. H. He, L. Feng, L. Y. Li, and C. Q. Xu, "Time evolution of the degree distribution of model A of random attachment growing networks," *Physica A*, vol. 384, no. 2, pp. 663-666, 2007. Article (CrossRef Link).

[35] S. P. Borgatti, M. G. Everett, and L. C. Freeman, "UCINET," *Encyclopedia of Social Network Analysis & Mining*, vol. 15, no. 7, pp. 536-544, 2014. Article (CrossRef Link).

[36] A. Lancichinetti, S. Fortunato, and J. Kertesz, "Detecting the overlapping and hierarchical community structure in complex networks," *New Journal of Physics*, vol. 11, no. 3, pp. 1-20, 2009. Article (CrossRef Link).

[37] J. F. Box, "Guinness, Gosset, Fisher, and Small Samples," *Statistical Science*, vol. 2, no. 1, pp. 45-52, 1987. Article (CrossRef Link).

Chongze Lin received his Master degree from Sichuan University, Chengdu, China, in 2008. He is currently a senior engineer at the institute of big data, Zhejiang Economic Information Center, Hangzhou, China. His research interests include social network, big data mining, E-government data management.

Yi Zheng received his Ph.D degree from college of Information Science and Electronic Engineering, Hunan University, Changsha, China, in 2016. He is currently a engineer at the institute of information research, Zhejiang Economic Information Center, Hangzhou, China. His research interests include social network, big data mining, deep learning, urban computing.