

Research Article

Understanding the Dynamics of Knowledge Building Process in Online Knowledge-Sharing Platform: A Structural Analysis of Zhihu Tag Network

Yongning Li ¹, Lun Zhang ², and Ye Wu ^{3,4}

¹School of Systems Science, Beijing Normal University, Beijing 100875, China

²School of Arts and Communication, Beijing Normal University, Beijing 100875, China

³School of Information Engineering, Putian University, Putian 351100, China

⁴Center for Computational Communication Research, Beijing Normal University, Zhuhai 519087, China

Correspondence should be addressed to Ye Wu; wuye@bnu.edu.cn

Received 25 March 2022; Revised 30 April 2022; Accepted 4 May 2022; Published 28 May 2022

Academic Editor: Peican Zhu

Copyright © 2022 Yongning Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Through structural analysis of 8-year tag networks from online knowledge-sharing platforms, this study finds that, with the scale of tag networks growing quickly, the growth trend of number edges indicates that tag network follows densification law. The clustering coefficient and the average shortest path of the network show that the rapid growth of network size does not bring about the compartmentalization of the knowledge network, and the degree distribution of tag networks shows a truncated power-law distribution. According to the structural characteristics of the tag network, this study proposes a tag network model based on the BA model. Based on the preference attachment, the triadic closure mechanism is employed to construct the edges between the old nodes, which revises the limitation that the BA model only connects edges between old and new nodes. The results show that the simulation model matches the actual tag network structure well. The generation mechanism of the tag network model provides a reference for understanding the knowledge construction process of the online knowledge-sharing platform to a certain extent.

1. Introduction

With the success of crowdsourced platforms, such as Wikipedia, Stack Overflow, Quora, and GitHub, a class of researchers are driven toward understanding the dynamics of knowledge building on these platforms [1]. The online knowledge-sharing system is an “all ask all” knowledge-building system, where users spontaneously ask and answer questions, and most of the platforms offer the opportunity for users to add knowledge tags to questions in the process of asking.

Compared with the traditional expert-driven knowledge production mode, the users’ self-organized knowledge construction process is more distributed and diversified, which generates different driving forces for the development of the knowledge domain. The investigation of the dynamic development of knowledge network is an important basic work to understand the process of knowledge production and construction, and it will help in the exploration of the

development trend of knowledge domain, knowledge innovation, and other issues [2].

Although collaborative knowledge building platforms are the most popular knowledge product mode, limited research has been done from the perspective of knowledge tag networks. Being composed of a set of concepts and interrelationships, knowledge can be effectively represented in terms of a network or a graph [3]. Therefore, our research aimed to explain the online knowledge building process from the perspective of network analysis and provide a reasonable explanation of the mechanism of knowledge network generation.

2. Related Works

To explore the development mechanism and research trends of knowledge domains, numerous researchers construct keyword cooccurrence networks using keywords of articles

as knowledge elements and then analyze the characteristics and evolution of networks. As important components of text content, keywords can present the core idea of academic articles. The analysis of topological features and evolution of keyword cooccurrence networks help in the realization of in-depth content analysis [4]. Zhu et al. [5] observed that the keyword networks are small worlds based on the network clustering coefficient and average distance among reachable nodes in a network, and the betweenness centrality is used to conduct preliminary studies on how to detect research hotspots of a discipline based on keyword networks. Yi and Choi [6] used the keywords of articles as an alternative knowledge element and studied the structural characteristics of keyword networks to better understand scientific knowledge. By observing keyword networks snapshots over time, Behrouzi et al. [7] utilized link prediction methods to foresee the future structures of these networks, which help in the prediction of future scientific research trends.

The knowledge tags provide a concise overview of the important content and key points of a question. The relationship between knowledge tags and questions is equivalent to that between keywords and articles. Several research works focused on tag networks with increased data availability on social media platforms. Feng et al. [8] conducted a structural analysis of knowledge tag network based on the motif structure and observed that the core nodes of the tag network have a strong attraction to other nodes; thus, a large number of knowledge tags are distributed at the periphery of very few center tags. Zhang et al. [9] described the content characteristics of the knowledge frontier of online knowledge-sharing platforms based on the theory of collaborative knowledge construction and assessed the knowledge frontier inclusiveness of online knowledge-sharing platforms. Chen and Xing [10] proposed an approach to automatically mine technology landscape from Stack Overflow question tags. Structured knowledge of technologies can emerge from the tagging practices of millions of online users considered together.

Although tag data in the online platform have been studied for a long time, a limited number of works aim to investigate the knowledge-building process through the evolution of tag networks. Many of the previous studies on the tag networks of question-and-answer (Q&A) platforms focus on tag recommendations [10, 11], whereas the mechanisms of knowledge tag network generation are not explored. The collaborative knowledge building of the online knowledge-sharing system forms a Collaborative Authoring Environment for online community participation by multiple people through the “question-answer” form, which is an important way for the public to exchange, share, and collaborate on knowledge. Users can browse, discuss, and produce content freely and openly and can create new questions and tags by asking questions, enabling rapid growth of knowledge in the system. Knowledge collaborative building based on an online knowledge-sharing system is the process of coevolution of individual knowledge and group knowledge [13] and realizes knowledge building in the modern sense. The generation of the knowledge tag network is the result of user knowledge collaborative production. By

adding new knowledge tags to the system and adding associations between tags, users support the knowledge network to continue to develop. The change in the knowledge tag network reflects the changes in the audience’s knowledge concerns and the dynamic process of knowledge evolution. The investigation of the mechanism of such networks can provide references for understanding knowledge production and predicting knowledge development trends.

To explain the mechanisms of knowledge-tag network generation, the degree distribution can be an important reference property. If the log-distribution of the node degree follows a power law [14], then the graph is a scale-free network [15], which can be found in numerous complex phenomena in the real world. Barabasi and Albert proposed the BA model [16], and they suggested that the power-law distribution of degree is a consequence of two generic mechanisms: (i) continuous network expansion by the addition of new nodes and (ii) newly arriving nodes tending to connect with already well-connected nodes, known as “preferential attachment.” Although the BA model is one of the most classic and applicable models available [17–19], and there are many, it still has several limitations that are inapplicable compared with numerous real-world networks. The actual network often has certain non-power-law characteristics, such as exponential truncation and small variable saturation [20]. A number of authors have subsequently published more extensive simulation results based on the BA model.

The improvement of the BA model can be approximately divided into two directions. One way to improve the BA model is to add new information dimensions to conform to different realistic systems. Another way is that the mechanism of connecting edges is slightly adjusted to conform to the diversity systems’ characteristics.

First, the simulation rules could be adjusted by introducing new variables or parameters into the model. For example, Bianconi and Barabási [21] proposed a fitness model that reflects the basic properties of most real systems, in which the nodes compete for links with other nodes; thus, a node can acquire links only at the expense of the other nodes. Xiang and Zhao proposed a modified BA model in which connecting decisions of new nodes are motivated by different proximities [22]. The simulation results showed that degree distribution still follows power laws, and the peripheral nodes are less dependent on core actors in accessing external knowledge.

Second, different from the BA model, which only add edges when a new node arrives, numerous networks where new nodes are added are found in the real world, whereas new connections are made between existing nodes inside the network. Although the degree distributions of nodes in these networks also take on a power-law-like form, their generative patterns are more complex than those of BA models. For example, to adjust the connecting rule [23] proposed models of developing and decaying networks with undirected links showing scaling behaviors. In addition to new links connecting new sites and old ones, links between old sites may appear or break [24]. This also involves extending the BA model by allowing the number of newly added links

to be random, under some mild assumptions on its distribution law. The modified model can create new nodes with a high degree at any iteration, which seem to be capable of simulating the temporal behavior of real networks more realistically.

Following the second improvement approach, one well-known change in the connecting edges mechanism is that Holme and Kim extend the standard scale-free network model to include a “triad formation step” [25]. They formulated that when a new node v is added to the network, v will connect to an old node, m , and, following the preference attachment mechanism, node v has probability P to connect to the neighbors of m . They found that, with the BA model and triadic closure mechanism, this model possessed the same characteristics as the standard scale-free networks, like the power-law degree distribution and the small average geodesic length, but with high clustering at the same time.

Triadic closure is a natural mechanism to make new connections, especially in social networks [26]. Suppose that two people have a mutual friend in the social network; the likelihood of them becoming friends in the future increases. This mechanism has been reported as the most common structural constraint [27]. It can explain many salient features of empirical social networks, including numerous closed triangles between acquaintances and fat-tailed degree distributions [28]. This mechanism brings dense network edge connection and can be one of the reasons for the network community structure. As a keyword network clustered by topic, the tag network also has a prominent community structure. The edge connection feature of the practical principle of the tag network may also be in line with the triadic closure mechanism. For example, when tag A is connected to tag B and tag C simultaneously, tag B and tag C are also more likely to be associated in the perspective of semantic dimension.

In this research, we will first describe the essential structural characteristics of the tag network, which is constructed by data from an online knowledge-sharing platform. We proposed a tag network simulation model based on the BA model and triadic closure mechanism. For complex and large-scale networks, the exploration of network characteristics and the simulation of the generative mechanisms are significant. The network analysis results reveal the evolutionary features of the knowledge network and help us understand the process of online knowledge building.

3. Method

3.1. Data. Zhihu is the largest online Q&A platform in China [9]. The same as most Q&A platforms, Zhihu allows users to add multiple tags to their questions, similar to the keywords in an article (see in Figure 1). Users can add tags that they have built themselves or select old tags that have already been built by other users. In the analysis of knowledge networks, the multiple tags appearing in the same question could be considered to have cooccurrence relationships [3]. This study used 74,761 tags contained in 1,520,254 questions from January 1, 2011, to December 31,



FIGURE 1: An example of the relationship of question and tags.

2018, in Zhihu. The number of tags and the cooccurrence relationships are cumulatively calculated every two months, and the line shows that the cooccurrence relationships have an obvious increasing trend in the last few months (see Figure 2).

To demonstrate the dynamic development process of the network, this study first splits 8 years of data into 48 time periods based on a time window of 2 months. In each time period, the tags appearing in the same problem were connected to build an undirected tag network. Finally, 48 network slices in total were constructed.

3.2. Tag Network Characteristics. First, this study calculates the network characteristics of the Zhihu tag network. We counted the number of nodes, the number of edges, clustering coefficient, and average shortest path of the largest connected component for each network slice (Table 1). As shown in Figures 3(a) and 3(b), the tag network size developed slowly in the first 10 time periods, and the number of nodes exhibited a significant upward trend around the 10th to 20th time periods. The number of nodes was relatively stable in the 30th to 40th time periods, followed by a very sharp increase in the 40th period. The number of nodes fell back in the last two time periods. These trends were consistent with the business strategy and development of Zhihu in China. From Figures 3(c) and 3(d), the tag network is a relatively dense network. Although the node size of the network is increasing, the network does not become compartmentalized as a result, and the average shortest path and clustering coefficients of the network remained at a relatively stable level, regardless of the slow decline.

In addition, Leskovec et al. [29] observed that as a network grows, its diameter decreases over time, suggesting that the network “shrinks” or becomes denser, which challenged the existing belief that online social networks evolve with a constant average degree and a slowly growing diameter. We calculated the effective diameter of each network slice. The network diameter is the maximum node distance [30]. Numerous real networks have small diameters, indicating small worlds. However, diameter is not always the best metric, because it is difficult to compute, and it is prone to outlier effects [31]. Thus, the effective diameter of each network slice was calculated (Table 1). A given natural number d represents the effective diameter of the network when the ratio of the shortest paths between pairs of nodes in the network is less than or equal to d reaching 0.9 [29]. As shown in Figure 4, the effective diameter of the network slices showed a slowly decreasing trend.

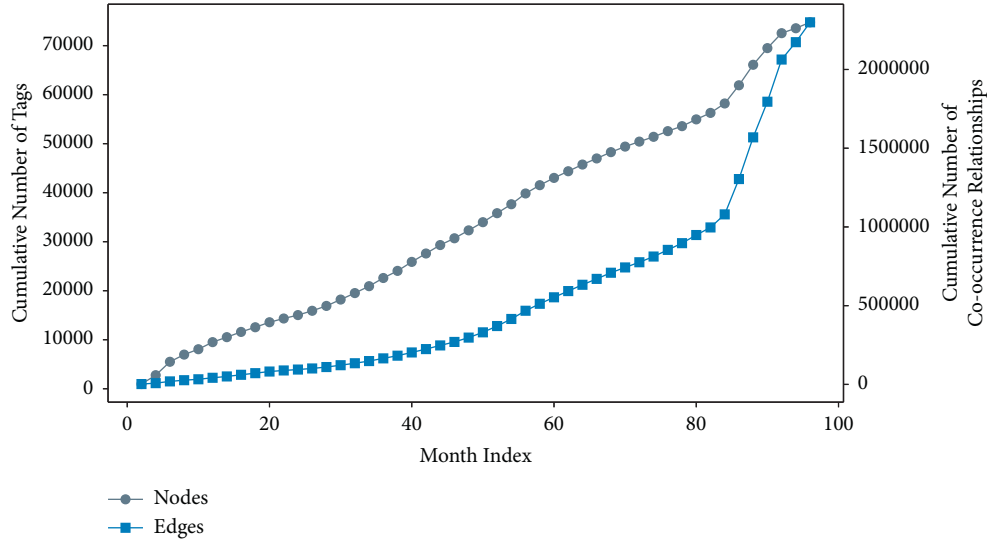


FIGURE 2: The cumulative number of tags (left) and relationships (right).

In real world, most systems experience a slow decrease in diameter due to the rapid growth in the number of edges. The growth of the numbers of nodes and edges shows a power function relationship ($e(t) \propto n(t)^\alpha$). Leskovec et al. [29] called this phenomenon densification law. Here, we constructed a complete network with 8 years of data and counted the numbers of nodes and edges of the network every two months. Figure 2 shows the cumulative numbers of nodes and edges of the network. Using the cumulative number of nodes at each moment as the horizontal coordinate and the cumulative number of edges as the vertical coordinate, the numbers of nodes and edges of the network almost approached a straight line in double logarithmic coordinates (see Figure 5). A linear regression model fitted to the scattered points yielded a slope of 1.66, an intercept of -1.92 , and a goodness of fit of 0.98.

Figure 6 shows the overall degree distribution of the network with 74,761 nodes in the double logarithmic coordinate. The horizontal axis denotes the value of degree, and the vertical axis represents the frequency of degree. The result of degree distribution showed that, as a whole, the growth of the tag network may be roughly in line with the preference attachment mechanism; that is, the nodes that newly joined the network had a higher probability to connect to the large-degree nodes, which led to the eventual appearance of the “rich get richer” power-law distribution of the network. However, the degree distribution of the tag network deviated from the power-law distribution at the tail end, which reflects that after the node scale of the network grew to a certain extent, the growth of edges of large-degree nodes was close to saturation, and the growth of their edges was limited. We counted the degree distribution during the growth of network on a bimonthly basis and screened out the top 100 nodes in degree rank. A total of 232 large-scale nodes were screened out at 48 time points. Next, we screened out a total of 28 nodes which ranked in the top 100 from the beginning of the network until the 48th time point. As shown in Table 2, these nodes are some broad and abstract

concepts, such as “life,” “movie,” “law,” “educate,” and “psychology.”

We fitted the tag network degree distribution by the power-law package of Python. Figure 7 shows that the degree distribution fit of the network was closer to a truncated power-law distribution (the red line) than to a standard power-law distribution (the blue line). Truncated power-law distribution is a common alternative to the asymptotic power-law distribution because it naturally captures finite-size effects [32]. Several measured social networks do not follow a power-law degree distribution [33] and are best fitted by an exponentially truncated power-law distribution. Clauset et al. [34] gave the basic truncated power-law functional form $f(k)$ (equation (1)) and the appropriate normalization constant C (equation (2)) such that $\int_{x=x_{\min}}^{\infty} C f(k) dx = 1$ for the continuous case. The distribution is $P(k) = C f(k)$, where k is the degree of node. The fitting results showed that the degree distribution of the tag network fitted the truncated power-law distribution with $\alpha = 2.12$ and $\lambda = 0.0003$.

$$f(k) = k^{-\alpha} e^{-\lambda k},$$

$$C = \frac{\lambda^{1-\alpha}}{\Gamma(1-\alpha, \lambda k_{\min})}. \quad (1)$$

4. Simulation

4.1. Model. The BA model is a classic model in the field of complex networks, and its simple mechanism can explain the power-law phenomenon in real networks; many simulation research of realistic mechanisms are based on the BA model [35, 36]. The two basic mechanisms of the BA model are as follows: (1) new nodes are constantly added to the network, and (2) the newly added nodes are more inclined to connect with large-degree nodes. However, for the knowledge tag network, such a mechanism is different from the

TABLE 1: The structural characteristics of 48-network slice.

No.	Start time	End time	The number of nodes	The number of edges	Effective diameter	Clustering coefficient	Average shortest path
1	2011.1.1	2011.3.1	957	2177	7	0.78	4.60
2	2011.3.1	2011.5.1	2260	5918	6	0.74	4.47
3	2011.5.1	2011.7.1	3997	12007	6	0.72	4.36
4	2011.7.1	2011.9.1	3009	8938	6	0.75	4.30
5	2011.9.1	2011.11.1	2586	7230	6	0.77	4.56
6	2011.11.1	2012.1.1	3580	11703	5	0.75	4.03
7	2012.1.1	2012.3.1	3321	11477	5	0.75	3.93
8	2012.3.1	2012.5.1	3711	14402	5	0.73	3.83
9	2012.5.1	2012.7.1	3661	14979	5	0.74	3.64
10	2012.7.1	2012.9.1	3975	15619	5	0.73	3.69
11	2012.9.1	2012.11.1	3212	10869	5	0.73	3.86
12	2012.11.1	2013.1.1	3141	9686	5	0.73	3.84
13	2013.1.1	2013.3.1	3747	11747	5	0.75	3.87
14	2013.3.1	2013.5.1	4553	15140	5	0.72	3.82
15	2013.5.1	2013.7.1	5609	19307	5	0.72	3.82
16	2013.7.1	2013.9.1	5995	21057	5	0.72	3.81
17	2013.9.1	2013.11.1	6495	23355	5	0.71	3.81
18	2013.11.1	2014.1.1	7611	28986	5	0.72	3.75
19	2014.1.1	2014.3.1	7597	29690	5	0.71	3.71
20	2014.3.1	2014.5.1	8984	35372	5	0.69	3.74
21	2014.5.1	2014.7.1	9388	37597	5	0.69	3.72
22	2014.7.1	2014.9.1	9970	40907	5	0.69	3.70
23	2014.9.1	2014.11.1	9655	40485	5	0.70	3.72
24	2014.11.1	2015.1.1	10876	49149	5	0.69	3.68
25	2015.1.1	2015.3.1	12032	59976	5	0.68	3.59
26	2015.3.1	2015.5.1	13868	71970	5	0.67	3.60
27	2015.5.1	2015.7.1	15083	81705	5	0.67	3.57
28	2015.7.1	2015.9.1	17056	92339	5	0.67	3.59
29	2015.9.1	2015.11.1	16116	83700	5	0.68	3.63
30	2015.11.1	2016.1.1	15729	80021	5	0.67	3.71
31	2016.1.1	2016.3.1	15737	77483	5	0.66	3.78
32	2016.3.1	2016.5.1	16423	77236	5	0.66	3.84
33	2016.5.1	2016.7.1	15876	73728	5	0.66	3.86
34	2016.7.1	2016.9.1	16338	78030	5	0.66	3.81
35	2016.9.1	2016.11.1	15389	69413	5	0.67	3.88
36	2016.11.1	2017.1.1	15484	68291	5	0.66	3.92
37	2017.1.1	2017.3.1	16477	76527	5	0.66	3.89
38	2017.3.1	2017.5.1	17804	89337	5	0.65	3.83
39	2017.5.1	2017.7.1	17678	90674	5	0.64	3.80
40	2017.7.1	2017.9.1	20174	107228	5	0.63	3.82
41	2017.9.1	2017.11.1	20452	101939	5	0.62	3.90
42	2017.11.1	2018.1.1	24633	164669	5	0.62	3.67
43	2018.1.1	2018.3.1	35577	392030	4	0.59	3.35
44	2018.3.1	2018.5.1	40559	479375	4	0.57	3.35
45	2018.5.1	2018.7.1	40270	445866	4	0.57	3.38
46	2018.7.1	2018.9.1	42434	507571	4	0.56	3.31
47	2018.9.1	2018.11.1	29927	275901	4	0.60	3.43
48	2018.11.1	2018.12.31	29652	300727	4	0.59	3.45

actual knowledge tag production process. First, from the perspective of knowledge building, new knowledge will be continuously produced, and new associations will also be generated between existing old knowledge concepts in the knowledge space. Therefore, for a tag network, the connection between nodes is generated when a new node joins and also between old nodes at any time. Regarding how to construct connected edges between old nodes, this study draws on the mechanism of triadic closure. For node A , if

nodes B and C are both neighbors of A but B and C are not connected, it is likely that connected edges will be generated between B and C in the subsequent moments. Second, although a “preferential attachment” exists in the tag network, that is, large-degree nodes (such as the more commonly used concepts with broader semantics) are more easily connected, when the network scale increases to a certain extent, this advantage will gradually weaken. In view of the shortcomings of the BA model and the knowledge building

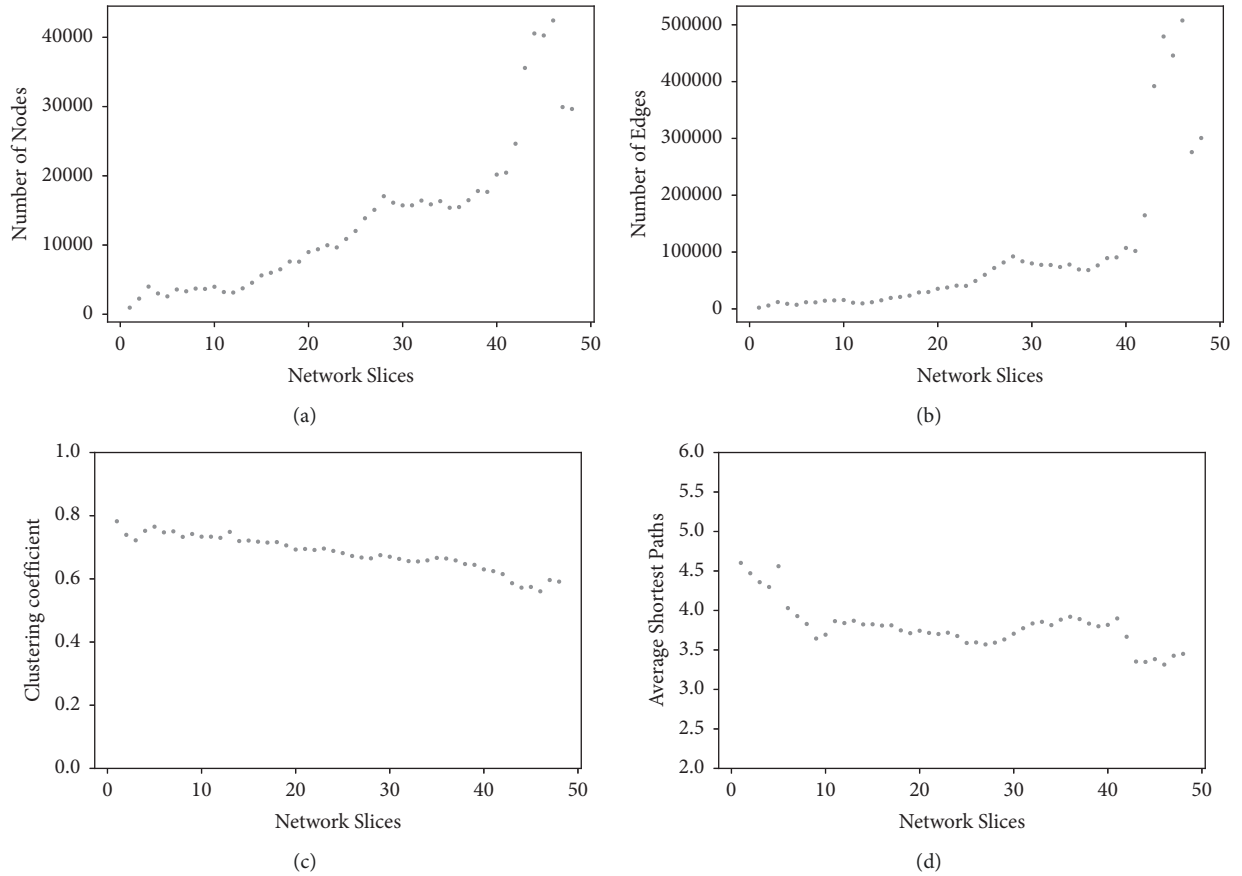


FIGURE 3: The characteristics of network slices: (a) the number of nodes, (b) the number of edges, (c) the clustering coefficient, and (d) the average shortest path of the largest connected component.

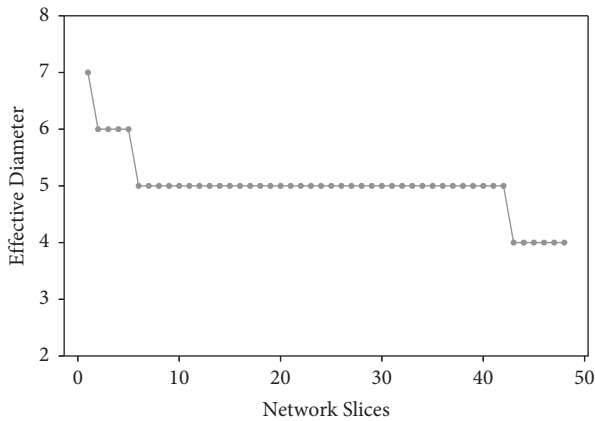


FIGURE 4: The effective diameters of network slices.

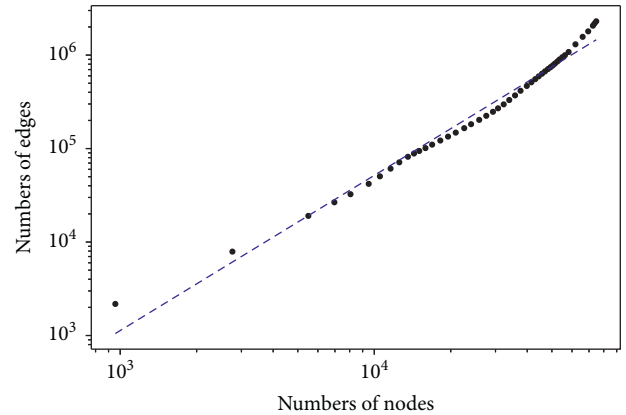


FIGURE 5: Number of edges versus number of nodes in log-log scales.

process, this paper proposed a tag network model based on the BA model and aimed to present the knowledge of the tag network generative mechanism of online knowledge-sharing platforms.

Based on these features, the specific algorithm of model generation is as follows (see Figure 8):

Step 1: a single node without edges exists in the initial network.

Step 2: the action will be selected between “add new node” and “add edge of old nodes” based on the current probability P . That is, there is the probability P to “add new node” and probability $(1 - P)$ to “add edge of old nodes.” P is a function about the numbers of nodes and edges of the network ($P = f(n, e)$, where n is the number of nodes, and e is the number of edges). The details of this function will be explained later.

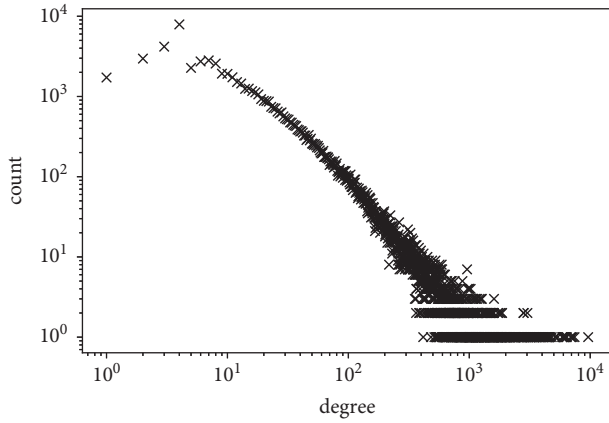


FIGURE 6: Degree distribution.

TABLE 2: Tags with high degree.

Degree	Tag
9620	Life
7416	Movie
6996	Law
6908	Educate
6399	Psychology
6108	How to think/evaluate X
5923	The internet
5386	History
5174	Healthy
5167	Music
4803	Start a business
4587	Delicacy
4556	Study
4484	Survey questions
4431	Travel
4297	Emotion
3991	Japan
3588	Interpersonal communication
3585	Design
3554	Culture
3460	Lifestyle
3174	Female
3161	Invest
3083	The life
2978	Philosophy
2837	Common sense of life
2831	English
2723	Read

Step 3: if the action “add new node” is selected, then a new node v is added, and a node m will be selected based on its degree from the current G . The larger the node degree is, the easier it is to be selected. Edges are then added to nodes v and m . Otherwise, if the action is “add edge of old nodes,” the algorithm will also select a node m based on its degree from the current G . Then, a node mn will be selected from the second-order neighbors of m , which is not connected with m . Edges are added to nodes m and mn afterward.

Step 4: steps 2 and 3 are repeated the until the number of nodes reaches the target number N .

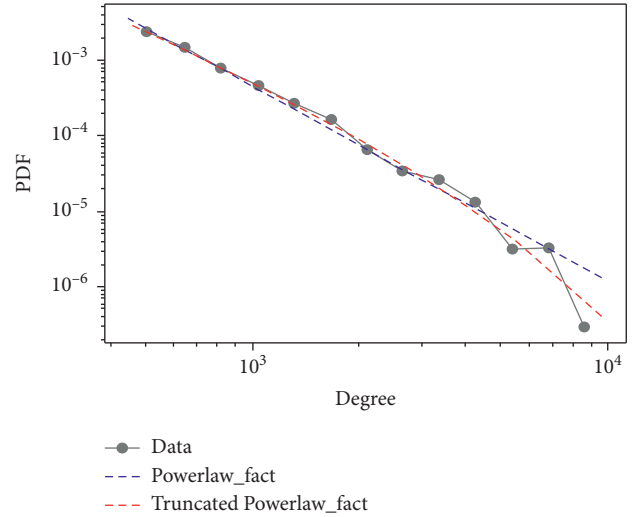


FIGURE 7: Degree distribution and fitting results.

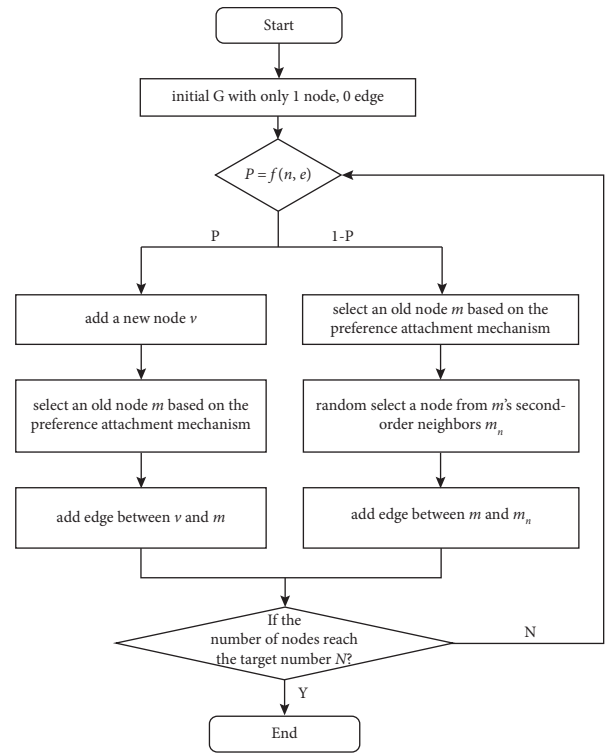


FIGURE 8: Simulation flow chart.

In this algorithm, the probability P determines whether the current time step will add new nodes or edges to the network. P is the probability of adding new nodes, and $1 - P$ is the probability of adding edges between existing nodes. Through the observation of the growth of the numbers of nodes and edges in the actual network, the growth of the number of edges in the network was mainly affected by the number of existing nodes and edges in the current network. The probability P should be correlated with the density in the current network. Therefore, we constructed a probability P_{t+1} , which is about the new node addition probability to

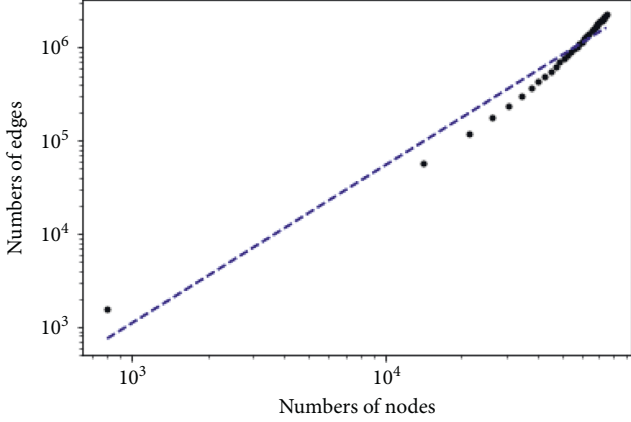


FIGURE 9: Number of edges versus number of nodes in log-log scales.

the network at time $t + 1$ based on the number of nodes n_t and the number of edges e_t in the network at time t , as shown in the following equation:

$$P_{t+1} = a * \left(\frac{e_t}{n_t} + 1 \right)^b. \quad (2)$$

In this study, the values of parameters a and b in formula (2) are obtained by fitting the actual data. We create a loss function f_{loss} that calculates the difference between the number of edges in a simulated network at parameters a and b and the number of edges in an actual network of the same network size. The loss is shown in formula (3), and t is the index of every two months. E_t is the actual number of edges in time t , E'_t is the number of edges in simulation network at the current time t , and $\log E_t$ is used as the denominator to balance out the impact of network size growth. The smaller the loss function is, the closer the performance of the simulated network is to the connection of the actual network, so the optimal solutions for parameters a and b can be obtained. Here we limit the range of a to $[1, 15]$ and the range of b to $[-10, 10]$ and then use the bisection method to find the optimal solution of b in the case of traversing a with a step length of 0.001.

$$f_{loss} = \sum_t \frac{|E_t - E'_t|}{\log E_t} \quad (t \in [1, 48]). \quad (3)$$

Furthermore, the degree growth of large-degree nodes in the tag network is not infinite. Therefore, when the degree of a concept has reached a certain threshold during model construction, its advantage in the degree-based preferential attachment mechanism needs to be weakened. Here, we assumed that when the degree of a node in the network reaches threshold H , the node with probability p_h does not increase its degree when calculating the current selected probability.

4.2. Simulation Results. By constructing the model mechanism proposed in Section 4.1, this study simulated the network generation process in which the network grew from

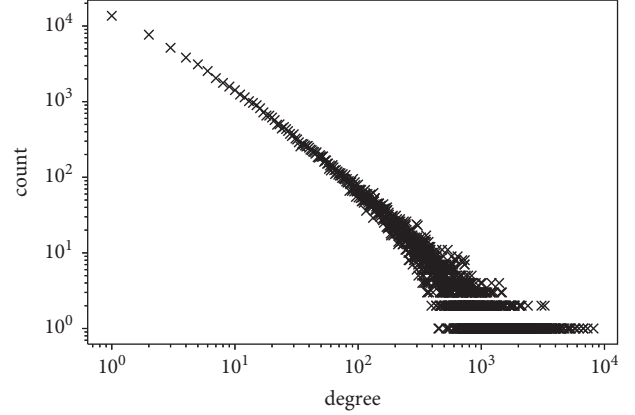


FIGURE 10: Degree distribution.

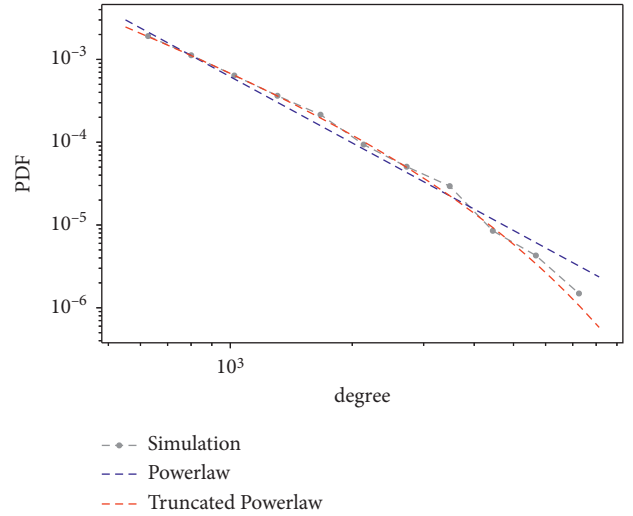


FIGURE 11: Degree distribution and fitting results.

an initial 1 node to 74,761 nodes. In this model, we fixed the parameters that determine the probability P as $a = 5.51$ and $b = -0.19$. In addition, given the situation of the tag network, we applied the degree threshold values of $H = 2000$ and $p_h = 0.69$.

Figure 9 shows the growth of the numbers of nodes and edges in the simulated network. The horizontal and vertical coordinates are the numbers of nodes and edges, respectively. With the passage of time steps, the distributions of the numbers of nodes and edges in the double logarithmic coordinate can be fitted by a straight line. The fitting slope was 1.68, the intercept was -2.01 , and the goodness of fit was $R^2 = 0.96$. Thus, the power function relationship ($e(t) \propto n(t)^\alpha$) between the numbers of nodes and edges was maintained in the simulated network, and the fitting slope and intercept were relatively close to the fitting results of the actual network.

Figure 10 displays the degree distribution of the simulated network, which is similar to the degree distribution image of the real network. In the double logarithmic coordinate, the degree distribution of the simulated network was a truncated power-law distribution with heavy tail,

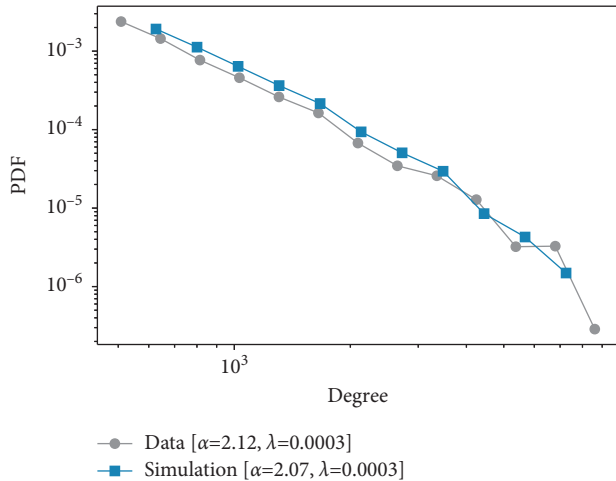


FIGURE 12: Comparison of degree distribution of real network (left) and simulation network (right).

which is a common phenomenon in empirical networks ([16], (Zhihai Rong*, Zhi-Xi Wu, Xiang Li, Petter Holme & Guanrong Chen (2019), *Heterogeneous cooperative leadership structure emerging from random regular graphs*, *Chaos*, vol.29, pp.103103)). When we fitted the degree distribution image (see in Figure 11), the degree distribution of the simulated network was closer to the truncated power-law distribution (red line). In addition, the fitting parameters $\alpha = 2.07$ and $\lambda = 0.0003$ were very close to the actual fitting parameters of the tag network (see in Figure 12).

5. Conclusion

Despite the enormous and recent interest in large-scale network data and the range of interesting patterns identified for static snapshots of graphs, relatively little work has been conducted on the properties of the time evolution of real graphs [29]. This paper introduced the knowledge tag network characteristics of online knowledge-sharing platform and simulated the generation model based on the classic BA model.

First, the results showed that the tag network exhibits a very rapid growth in scale, but it is not a fragmentation with a rapidly growing number of nodes. On the contrary, the effective diameter and clustering coefficient of the network showed a slowly declining trend. The edges of the network were very dense, and the numbers of nodes and edges of the network showed a relationship close to a power function over time, which indicates that the tag network followed the densification law.

Second, the degree distribution of the tag network followed a truncated power-law distribution. The link mechanism in tag networks also followed the “rich get richer” preference attachment mechanism. In the tag network, the edges among nodes imply that those knowledge concepts are semantically related, and the nodes with large degree are often generalized and broad concepts. Therefore, with the development of tag networks, the degree advantage of the addition of an edge stage will weaken, which explains why the degree distribution of the knowledge tag network was

closer to the truncated power-law distribution than to the power-law distribution. The fitting results showed that the degree distribution of tag network fitted the truncated power-law distribution with $\alpha = 2.12$ and $\lambda = 0.0003$. Then, this study proposed a network generation model applicable to tag networks. The model is based on the BA model with the addition of the edge linkage mechanism between old nodes, which is more consistent with the actual process of knowledge building and can make the network generate a dense network structure. A truncated power-law fit of the node degree distribution of the simulated network was obtained with $\alpha = 2.07$ and $\lambda = 0.0003$, which were close to those of the real network degree distribution. Therefore, the simulation model proposed in this study can explain the growth mechanism of the real tag network to a certain extent.

Finally, this study investigated the mechanism of tag network generation in online knowledge platforms, and this work will help us deepen our knowledge and understanding of the online knowledge building process. The model proposed in this study can provide a general adaptation to the current online knowledge-sharing platform by adjusting model parameters. This model uses probability P to balance the relationship between network edge addition and point addition. The parameters of P come from the fitting of actual data, which means that the function has strong expansibility and can be simulated according to the historical data of different platforms. In addition, the parameters of the P function in the model are determined by fitting the historical data of the network, and, in the process of searching for parameters, the algorithm uses the idea of binary search to reduce the time complexity significantly while maintaining the accuracy of the results. Even large-scale network data can be calculated in a relatively brief time. These characteristics make the model utility to adapt to different data platforms. It is useful to predict the growth scale of tag networks in the future based on the information of the network at present and provide a reasonable reference for the future knowledge platform construction.

In the future, the research on the generation mechanism of the tag network should be expanded from two dimensions. One is to broaden the research platform and research objects. The other is to use the simulation network as the basic framework to carry out the research on the network structure and network efficiency by the generation mechanism. The research platform and research objects should not be limited to the knowledge tags and the keywords. Topic and many other texts content also have research value. In addition, future research could pay attention to the information communication effect or other problems combined with the network generation mechanism, which could give an in-depth understanding of the relationship between network structure and network function [37].

Data Availability

The data used to support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant no. 11875005) and the Doctoral Interdisciplinary Foundation of Beijing Normal University (BNUXKJC2115). The authors would like to express their sincere thanks to Tianqi Wu for his valuable guidance on the data analysis of this study.

References

- [1] A. A. Verma, S. R. S. Iyengar, S. Setia, and N. Dubey, "KDAP: an open source toolkit to accelerate knowledge building research," in *Proceedings of the 16th International Symposium on Open Collaboration*, August 2020, NY, USA.
- [2] X. Liu, F. Ma, X. Chen, B. Chen, and Q. Jia, "Structure and evolution of knowledge network - concept and research review," *Information Science*, vol. 29, no. 6, pp. 801–809, 2011.
- [3] E. K. Tokuda, C. H. Comin, and L. da F. Costa, "Modeling the Perspectives for Scientific Advancement," 2021, <https://arxiv.org/abs/2103.08692>.
- [4] H. Li, H. An, Y. Wang, J. Huang, and X. Gao, "Evolutionary features of academic articles co-keyword network and keywords co-occurrence network: based on two-mode affiliation network," *Physica A: Statistical Mechanics and Its Applications*, vol. 450, pp. 657–669, 2016.
- [5] D. Zhu, D. Wang, S.-U. Hassan, and P. Haddawy, "Small-world phenomenon of keywords network based on complex network," *Scientometrics*, vol. 97, no. 2, pp. 435–442, 2013.
- [6] S. Yi and J. Choi, "The organization of scientific knowledge: the structural characteristics of keyword networks," *Scientometrics*, vol. 90, no. 3, pp. 1015–1026, 2012.
- [7] S. Behrouzi, Z. Shafaeipour Sarmoor, K. Hajsadeghi, and K. Kavousi, "Predicting scientific research trends based on link prediction in keyword networks," *Journal of Informetrics*, vol. 14, no. 4, Article ID 101079, 2020.
- [8] X. Feng, S. Hu, J. Li, and Y. Wu, "Research on evolution characteristics of tagging networks based on complex motif - the case of Q&A learning community Zhihu," *Information Science*, vol. 38, no. 9, pp. 56–62, 2020.
- [9] L. Zhang, Y. Li, and Y. Wu, "Mapping knowledge space: collaborative knowledge building on online knowledge sharing platforms," *Journalism and Communications*, vol. 28, no. 1, pp. 52–70+127, 2021.
- [10] C. Chen and Z. Xing, "Mining technology landscape from stack overflow," in *Proceedings of the 10th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement*, pp. 1–10, Ciudad Real Spain, September 2016.
- [11] A. K. Saha, R. K. Saha, and K. A. Schneider, "A discriminative model approach for suggesting tags automatically for stack overflow questions," in *Proceedings of the 2013 10th Working Conference on Mining Software Repositories (MSR)*, pp. 73–76, CA, USA, May 2013.
- [12] P. Singh, R. Chopra, O. Sharma, and R. Singla, "Stackoverflow tag prediction using tag associations and code analysis," *Journal of Discrete Mathematical Sciences and Cryptography*, vol. 23, no. 1, pp. 35–43, 2020.
- [13] J. Kimmerle, J. Moskaliuk, A. Harrer, and U. Cress, "Visualizing CO-evolution OF individual and collective knowledge," *Information, Communication & Society*, vol. 13, no. 8, pp. 1099–1121, 2010.
- [14] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and analysis of online social networks," in *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement*, pp. 29–42, CA USA, October 2007.
- [15] S. Milgram, "The small world problem," *Psychology Today*, vol. 2, no. 1, pp. 60–67, 1967.
- [16] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.
- [17] X. Xiao, H. Chen, and P. Bogdan, "Deciphering the generating rules and functionalities of complex networks," *Scientific Reports*, vol. 11, no. 1, pp. 1–15, 2021.
- [18] S. Ai, S. Hong, X. Zheng, Y. Wang, and X. Liu, "CSRT rumor spreading model based on complex network," *International Journal of Intelligent Systems*, vol. 36, no. 5, pp. 1903–1913, 2021.
- [19] C. Zhao, Y. Wang, Y. Gong, S. Brown, and R. Li, "The evolution of the port network along the Maritime Silk Road: from a sustainable development perspective," *Marine Policy*, vol. 126, Article ID 104426, 2021.
- [20] H. Wei, J. Wang, and T. Wang, "An improved B-A model for scale-free network evolution model," *Journal of Physics: Conference Series*, vol. 1486, no. 2, Article ID 022034, 2020.
- [21] G. Bianconi and A.-L. Barabási, "Competition and multi-scaling in evolving networks," *Europhysics Letters*, vol. 54, no. 4, pp. 436–442, 2001.
- [22] X. Xiang and J. Zhao, "The effect of social proximity on degree distribution of Scale-free network," in *Proceedings of the International Academic Conference on Frontiers in Social Sciences and Management Innovation (IAFSM 2018)*, pp. 242–247, Chongqing, China, March 2019.
- [23] S. N. Dorogovtsev and J. F. F. Mendes, "Scaling behaviour of developing and decaying networks," *Europhysics Letters*, vol. 52, no. 1, pp. 33–39, 2000.
- [24] S. Sidorov and S. Mironov, "Growth network models with random number of attached links," *Physica A: Statistical Mechanics and Its Applications*, vol. 576, Article ID 126041, 2021.
- [25] P. Holme and B. J. Kim, "Growing scale-free networks with tunable clustering," *Physical Review*, vol. 65, no. 2, Article ID 026107, 2002.
- [26] G. Bianconi, R. K. Darst, J. Iacovacci, and S. Fortunato, "Triadic closure as a basic generating mechanism of communities in complex networks," *Physical Review*, vol. 90, no. 4, Article ID 042806, 2014.
- [27] G. Kossinets and D. J. Watts, "Empirical analysis of an evolving social network," *Science*, vol. 311, no. 5757, pp. 88–90, 2006.
- [28] P. Klimek and S. Thurner, "Triadic closure dynamics drives scaling laws in social multiplex networks," *New Journal of Physics*, vol. 15, no. 6, Article ID 063008, 2013.
- [29] J. Leskovec, J. Kleinberg, and C. Faloutsos, "Graph evolution," *ACM Transactions on Knowledge Discovery from Data*, vol. 1, no. 1, pp. 2–es, 2007.
- [30] J. Leskovec, J. Kleinberg, and C. Faloutsos, "Graphs over time: densification laws, shrinking diameters and possible explanations," in *Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining*, pp. 177–187, Chicago Illinois USA, August 2005.
- [31] J. Niu, J. Peng, L. Shu, C. Tong, and W. Liao, "An empirical study of a Chinese online social network--renren," *Computer*, vol. 46, no. 9, pp. 78–84, 2013.

- [32] J. L. Cardy and I. Peschel, "Finite-size dependence of the free energy in two-dimensional critical systems," *Nuclear Physics B*, vol. 300, pp. 377–392, 1988.
- [33] M. E. J. Newman, "Scientific collaboration networks. I. Network construction and fundamental results," *Physical Review*, vol. 64, no. 1, Article ID 016131, 2001.
- [34] A. Clauset, C. R. Shalizi, and M. E. J. Newman, "Power-law distributions in empirical data," *SIAM Review*, vol. 51, no. 4, pp. 661–703, 2009.
- [35] Y. Mao, Z. Rong, and Z.-X. Wu, "Effect of collective influence on the evolution of cooperation in evolutionary prisoner's dilemma games," *Applied Mathematics and Computation*, vol. 392, Article ID 125679, 2021.
- [36] J. Shih, "Education, Asia-pacific society for computers in education," in *Proceedings of the 27 the International Conference on Computers in Education*, Taiwan, China, November 2019.
- [37] P. Zhu, L. Cheng, C. Gao, Z. Wang, and X. Li, "Locating multi-sources in social networks with a low infection rate," *IEEE Transactions on Network Science and Engineering*, no. 1, 1 page, 2022.