



Applied Mathematics and Nonlinear Sciences

<http://journals.up4sciences.org>

Complex Network Theory and Its Application Research on P2P Networks

Feng Jian[†], Shi Dandan.

College of Computer Science & Technology, Xian University of Science and Technology, Xian 710054
China

Submission Info

Communicated by Juan L.G. Guirao

Received 7th May 2015

Accepted 25 October 2015

Available online 1st January 2016

Abstract

Advances in complex networks of Peer-to-Peer (P2P) networks were reviewed and summarized. The paper outlines some important topological properties such as degree, average path length and clustering coefficient at first, and then three kinds of most important network mechanism models are introduced, including random graph model, small world model and scale-free model. A simple description about research status for P2P networks based on complex networks is made from three aspects: positive research, network mechanism model, network broadcast and control. Some developing prospects of complex networks of P2P are pointed out finally. Complex network provides new ideas and methods to deal with many complex problems including P2P networks.

Keywords and phrases: Complex network; Peer to Peer network; positive research; network mechanism model

2010 Mathematics Subject Classification: 68M10

1 Introduction

Complex network theory has become an important paradigm to interpret problems in computer science, sociology, biology and many other areas. The research of complex network theory focused on aspects of the general features of network topology, topology generation mechanism, the network dynamics, and has achieved fruitful results. In recent years, peer to peer network (Peer-to-Peer, P2P) itself and its network environment both showed explosive complex growth, traditional ideas, techniques and methods for constructing P2P system are facing serious challenges, it is a urgent need to find a new perspective to understand the relationship between network structure and network behavior, thereby to improve the behavior of the network. Combined with

[†]Corresponding author.

Email address: fengjian@xust.edu.cn

advances in complex networks, P2P networks with rich application feature also attracted wide attention. This paper summarizes research outcomes of the P2P network from the perspective of complex network theory, and looks ahead for future research directions.

2 The basic theory of complex networks

2.1 Basic Metrics

Degree and degree distribution: the degree of nodes is the number of other nodes in the network connected to the node. Degree can quantitatively reflect importance of the node in the network. Degree distribution of the node is represented by distribution function; its meaning is exactly the probability that an arbitrary node has k edges, namely the ratio of the number of nodes which degree is k to the total number of network nodes in the network.

Average path length: distance between two nodes in the network is defined to the number of edges of these two nodes on the shortest path. The diameter of the network is the maximum distance between any two points; average path length of the network is the average of the distance between any two nodes in the system, which describes the degree of separation between network nodes.

Clustering coefficient: used to describe the aggregation of network nodes, namely how closely the network. Clustering coefficient of a node is the proportion of number of connected edges among all nodes adjacent to the node to number of the largest possible edges among all nodes adjacent to the node, and clustering coefficient of a network is the average of the clustering coefficients for all nodes in the network.

2.2 Complex network models

Complex network models concern about the evolution of the network topology and its dynamic process. The complex networks models which have been extensively studied currently include:

(1) Random network model

Famous mathematician Erdos and Renyi proposed a mathematical theory currently being called ER random graph model in 1959, and the model has been the basic model of complex networks research in nearly 40 years after that. It is constituted by N nodes, and any two nodes in the model can connect to each other with probability p randomly. In ER model, the edge between two nodes is independent to the presence or absence of the other edge, so there is neither clustering correlation and degree correlation, nor community structure.

(2) Small-world model

Generally, if the average distance L between any two nodes in the network increases logarithmically according to the increase of N , namely $L \sim \ln N$, that is to say when N increases rapidly, L changes relatively slowly, then the network is said to have a small world effect. The small-world network model has been proposed in 1998 [1]. Through randomly connecting each edge of regular network with probability p to a new node in the network, the model constructs a network between random network and regular network. The network been constructed has a smaller average path length and a larger clustering coefficient at the same time, and regular network and random network is the special case when p is 0 and 1 separately.

(3) Scale-free model

Degree distribution of large scale real network can be described more precisely by the power law distribution, i.e., $P(k) = k^{-\lambda}$, where k is a variable of node degree and λ is a constant independent of the scale. These networks are called scale-free networks, and there are some small numbers of hubs with high degree in this kind of networks. To explain the mechanism of its formation, in 1999, the famous BA model has been proposed [2]. Starting with a small fully connected network, the model adds a new node in each step with a certain probability, the new node should connect randomly to some node already existed in the network, the probability of the connection is proportional to degree of the selected nodes (preferential attachment).

3 Empirical Research

As a class of large-scale, self-organized, highly dynamic complex networks, accurate and complete measurement for all P2P network topology is facing great difficulties. Researching on the characteristics of their protocols and analyzing specific examples of P2P topology become alternative solution to study P2P topology characteristics.

3.1 P2P topology measurement

Current research on analysis of P2P topology features focuses on unstructured peer network measurements represented by Gnutella. Tovanovic measured early Gnutella network in 2000 [3], considered that its nodes had the small-world characteristics and the degree distribution followed power law distribution. It was only captured around 1000 peers, convincing data is limited. The literature [4] also pointed out that degree distribution of the Gnutella network accorded with power law distribution. In particular, the literature [5] pointed out that Gnutella complied with power-law distribution of $\lambda = 2.3$, the network has high fault tolerance with nodes fail randomly, but is very weak resistance to malicious attacks. Especially after node with high connectivity gets failure, it will lead overlay network to be divided easily. Xie [6] et al. divided Gnutella network into two layers, Mesh layer and Forest layer. They proved that degree grade distribution and degree frequency distribution of Mesh nodes and size distribution of trees in Forest layer all have the nature of power law.

With the continuous development of the Gnutella network, its network characteristics had undergone a fundamental change. Stutzbach et al. constructed a high-speed topology crawler [7], pointed out that degree distribution of Gnutella network follows neither power law distribution nor segmented power law distribution. Measurement found the length of the shortest path in the network was more evenly distributed; about 60% of the length was 4, and each leaf node connected to a plurality of super-nodes. Measurement of the average path length and clustering coefficient showed Gnutella network still exhibited small world characteristics. The authors analyzed the situation Gnutella topology changed over time, found that long-standing nodes in Gnutella network often tended to gradually connected to each other, together formed a stable core, and clustering coefficient within the core had an increasing trend to become larger over time. Wang Yong et al found that Gnutella network as a whole did not meet the power-law distribution; degree distribution of leaf nodes had long tail characteristics, and degree distribution of topology was in line with segmented power law [8].

Characteristics of some other P2P networks such as Edonkey have also been analyzed [9], it also reveals the networks have the characteristics of power-law, clustering and small world. Through statistic and analysis of the networks, these studies focused on testifying existed models on the P2P networks.

3.2 User behavior

With the deepening of the study, users, resources and their relationships with each other in P2P network were analyzed; especially users' inner interests have been excavated in-depth. By measuring user behavior characteristics in a large number of network, the literature [7] found user behavior was distinguish between different network, for example, offline time span of nodes in KAD and BitTorrent networks is much shorter than that of in Gnutella network; but there were also many similarities, such as online time span of nodes in these network subjected to the same distribution.

Xia Hao et al found that in the Maze system, more users downloaded popular resources and more easily formed a group with each other; and popular resources were more likely to be centralized shared by users and form groups with each other [10]. The conclusion of the study is somewhat similar to this literature [11]. The latter discovered that resources epidemic threshold impacted on community level of user network, with the increasing of resource popularity thresholds, community structure is more clear. Based on log data of the network service providers, Li et al found that [12]: ① degree distribution of resource nodes approximately followed power law distribution; ② resource network had obvious cluster structure; ③ resources among different categories showed a significant association. In the user network and the resources network formed by the measured

data, the nodes were connected together by the edges with the right values. Weighted edges reflected the relationship of interest breadth on two users. Analysis of weighting network under actual application context is the focus of further work.

4 Network mechanism model

Research of network mechanism model in complex networks theory concerns the evolution of the network topology and its impact on the statistical properties of the network. Research on P2P networks developed and deepened the existing network mechanism model.

4.1 Random network model

Kant introduced a class of "non-uniform random network model" to describe P2P networks [13], modeled the P2P networks to two layer random network consisted by backbone nodes and non-permanent nodes, and gave the generation process of the network. Limitation of this model is not considering transient nodes which account for most of the network nodes. The literature [14] used random network model too, considered the P2P network to be a two layer network which consisted by backbone layer nodes and leaf nodes, wherein connection between backbone nodes and leaf nodes was considered random bipartite graph, connection between the trunk nodes was considered generalized random graph. Carra et al modeled the evolution of P2P streaming system as a Markov process [15], gave the master equation and rate equations describing change process of the streaming media system, in which the parameters of the equation associated with specific streaming protocols. Through solving the equation can determine the relevant distribution, playback delay and other performance indicators.

4.2 Small world network model

Constructing P2P networks using small world theory can lead the P2P networks having lower average degree, smaller average path length and larger clustering coefficient, helping to improve the performance of resource location, load balancing, and to enhance fault-tolerant ability of P2P network, had become academic research focus [16]–[22].

(1) Increasing long links to form a small world

In the unstructured P2P networks, by drawing on the experience in the field of small world research Zhou Jin et al [16] inserted quick links to remote nodes into the routing table of P2P nodes in order to make the average path length shorten, thereby to improve the performance of the search algorithm.

Some studies introduced small world characteristics to a structured P2P DHT network, for example, by modifying replication and cache replacement policies of Freenet, fast connections between nodes with small-world characteristics formed in [17]; the literature [18] improved Cayley graph, proposed a P2P model based on generalized cube-connected cycle, and designed a DHT protocol to adapt dynamic changes, making the network having a small-world characteristics.

(2) Forming a hierarchical structure based on node capacity

Some studies build hierarchical P2P network model by super nodes, enabling small world characteristics. The breakthrough of this kind of questions is to assess the importance of the nodes in accordance with their difference, there are many different assessment methods combined with different actual background. Characterized by concerning the characteristics of the network in some particular aspect, such as the position of the node in the network; control ability of nodes to the dissemination of information; and contribution of node to the local structure, etc., or the integrated nature of above plurality of characteristics. For example, the literature [19] built a P2P overlay network model with hierarchical structure based on the output bandwidth. The purpose of layering was to allow the nodes with superior performance be at the upper layer to access to the data stream preferentially, so that the data flow could spread through these nodes with high-output bandwidth to members within the group as soon as possible. The structure not only has efficiency of structured search, but also has

advantages of autonomy and load balancing of non-structure.

(3) Forming community based on interest of peers

For improving the efficiency of resource search and network scalability, the concept of community is presented, it is to set up a community based on interest showed by peers, so that nodes with similar interests gather into a virtual community to enhance pertinence of information dissemination. There are two main ways to establish community: the first one is through initiatively mining share content on the known nodes to regularly detect interest of adjacent nodes, and then to establish a direct connection with nodes which have similar interests and change topology dynamically [20]; in accordance with behavior showed by user in searching resources, the second one is to determine interest of nodes based on the probability of querying similar resources recently, and to connect topology using the feedback information [21].

(4) Topology sensitive small world

In P2P networks, the situation that the network topology does not match with the actual network topology due to the dynamic changes of peers occurs frequently, so small-world networks which able to adapt to the dynamic changes of resources and can adapt to the actual network topology are concerned. [22] obtained certain practical network information by ICMP-based topology information collection method, using the idea of clustering gathered nodes which are close in real network to the same cluster, and constructed a P2P overlay network with network sensitivity and small-world characteristics to help to improve QoS and network congestion problems.

4.3 Scale-free network model

In P2P networks scale-free model has also been studied and applied. Wouhaybi proposed a fully distributed algorithm [23], it can build P2P node to short-diameter overlay topology Phenix whose degree distribution obeys a power law distribution in order to speed up resource location by taking advantage of short network diameter features of scale-free networks. Through the analysis of the creating mechanism of power law characteristics in complex networks, by learning rumor spreading mechanism in interpersonal communication, the literature [24] combined the small world and power-law distribution characteristics into P2P networks, presented search method of information resources with combination of preferential attachment mechanism and interest attenuation mechanism in rumors dissemination.

4.4 Other models

Not confined to the traditional theory of complex networks, from application-specific characteristics of P2P networks, the literature [25] tried to build an appropriate statistical model to analyze the statistics features of user behavior and its impact. Using number of shared files owned by the user as the main parameter, the model characterizes evolutionary process controlled by users' sharing behavior of P2P application. Combined with the real Internet traffic measurements, the similarities and differences of self-organizing mechanisms between the network model and the actual P2P applications were analyzed.

5 Network Transmission and Control

5.1 Propagation dynamics

Spread of P2P network resources is similar to prevalence of infectious disease in the population and the spread of viruses on the network. Mathematical models of epidemics spread can well describe the propagation characteristics of complex networks, is the basis of propagation dynamics study of the complex network. Classic propagation model is the active propagation model. But the spread of P2P file depends on the user's actions, is a kind of passive spread, so the existing propagation dynamics model cannot simulate status changes of P2P nodes in file propagation process well. The literature [26] proposed SEInR model to improve the existing SEIR model.

The improvement included: the infection were divided into n sub-categories, giving different model parameters; considered the effect of different degree values of complex networks to the propagation model.

5.2 Immunization and control

Typical current immunization strategies can be divided into random immunization, target immunization and acquaintance immunization. Random immunization does not have the efficiency and is diseconomy. For large-scale dynamic global information network such as P2P networks, target immunization method is impractical and uneconomical. Acquaintance immunization method selects all neighbor nodes randomly and without judging, and is limited to selected direct neighbor, so its actual results are severely constrained. And these immunization strategies are all aiming at generalized scale-free networks, do not have direct applicability to technology network such as P2P networks. The literature [27] proposed and implemented a decentralized target immunization strategy suitable for P2P networks by introducing dynamic adaptive target immunization mechanism for key nodes with the highest connectivity and the highest availability and for the key link with the highest load. The strategy mined and used inherent enlightening information of the actual network nodes, combined with acquaintance immunization mechanism, to improve the ability of anti-intentional attack to P2P networks.

6 Conclusions

P2P networks have large-scale, heterogeneous and complex architecture and rich dynamic characteristics, enable complex networks theory become a powerful tool for the study of generation, development and optimize mechanism of P2P networks and its network behavior. Current research of P2P network from the perspective of complex networks focuses on several aspects of the empirical analysis, network mechanism model and transmission dynamics. Extensive research is still focused on the small-world and scale-free network model based on P2P file-sharing system, the research on other kind of P2P applications is not so much, corresponding theory needs further research.

In summary, the current awareness to characteristics of P2P network topology is still immature. From behavior of P2P network topology based on the complex network theory and related technologies, there are many directions can be studied, including: ① to study different types and large-scale P2P networks by measurement in order to find the feature of each network topology and characteristic quantity of topological properties; to study the influence degree of different nodes to the evolution behavior of network topologies and the interaction between different nodes; ② through studying the problems of dynamic update, resource management, service discovery and other issues of the users in the P2P network, to explore construction mechanisms of the network topology which can reflect true network structure; ③ through establishing the network topology and behavioral models of dynamic evolution, to discuss the complex behavior of the network structure and to study the network reliability.

Acknowledgment

This work was supported in part by Shaanxi Provincial Natural Science Foundation Project (No. 2012JQ8030).

References

- [1] Watts D J, Strogatz S H. (1998), Collective dynamics of small world networks. *Nature*, 393, 440-442. doi [10.1038/30918](https://doi.org/10.1038/30918)
- [2] Barabasi A L, Albert R. (1999), Emergence of scaling in random networks. *Science*, 286, 509-512. doi [10.1126/science.286.5439.509](https://doi.org/10.1126/science.286.5439.509)
- [3] Jovanovic M A. (2001), Modeling large-scale peer-to-peer networks and a case study of Gnutella. Cincinnati: University of Cincinnati.

- [4] Ripeanu M, Foster I, Iamnitchi A.(2002), Mapping the Gnutella network: Properties of large-scale peer-to-peer systems and implications for system design. *IEEE Internet Computing Journal Special Issue on Peer-to-Peer Networking*,6(1), 50-57.
- [5] Saroiu S, Gummadi P, Gribble S.(2003), Measuring and analyzing the characteristics of napster and gautella hosts. *Multimedia Systems* 9, Berlin: Springer-Verlag,170-184. doi [10.1007/s00530-003-0088-1](https://doi.org/10.1007/s00530-003-0088-1)
- [6] Xie C, Pan Y.(2006), Analysis of large-scale hybrid peer-to-peer network topology. Kero T, ed. *Proc. of the 49th Annual IEEE Global Communications Conference*. Washington, USA: IEEE Computer Science Press,1-6. doi [10.1109/GLOCOM.2006.189](https://doi.org/10.1109/GLOCOM.2006.189)
- [7] Stutzbach D, Rejaie R, Sen S.(2008), Characterizing unstructured overlay topologies in modern P2P file-sharing systems. *IEEE/ACM Trans on Networking*,16(2), 267-280. doi [10.1109/TNET.2007.900406](https://doi.org/10.1109/TNET.2007.900406)
- [8] Wang Yong, Yun Xiao-chun, Li Yi-fei.(2008), Measuring and Characterizing Topologies of P2P Networks. *Journal of Software* 19(4), 981-992. doi [10.3724/SP.J.1001.2008.00981](https://doi.org/10.3724/SP.J.1001.2008.00981)
- [9] Le Fessant F, Handurukande S, Kermarrec A-M, et al.(2005), CClustering in peer-to-peer file sharing workloads. Voelker, G.M., Shenker, S. ed. *Proc. of IPTPS 2004*. Springer, Heidelberg, 3279, 217-226. doi [10.1007/978-3-540-30183-7_21](https://doi.org/10.1007/978-3-540-30183-7_21)
- [10] Xia Hao, Wang Yang, Di Zengru, Fan Ying.(2010), Empirical Research of Maze Based on Theory of Complex Network. *Journal of Beijing Normal University (Natural Science)*, 46(5), 647-650.
- [11] Wang Jing-Xin, Wang Yue, Li Yi-Peng, Yuan Jian, Shan Xiu-Ming, Feng Zhen-Ming, Ren Yong.(2011), Popularity based network statistical analysis in peer-to-peer application. *Acta Physica Sinica*, 60(11), 748-756. doi [10.7498/aps.60.118901](https://doi.org/10.7498/aps.60.118901)
- [12] Li Yi-Peng, Ren Yong, Yuan Jian, Wang Yue, Huang Xiao-Hong, Shan Xiu-Ming.(2011), Network statistical analysis in peer-to-peer application. *Acta Physica Sinica*, 60(5), 802-807.
- [13] Kant K.(2003), An analytic model for peer to peer file sharing networks. *Proc. of IEEE International Conference on Communications*, 3, 1801-1805.
- [14] Gaeta R, Sereno M.(2007), Random graphs as models of hierarchical peer-to-peer networks. *Performance Evaluation*, 64(9-12), 838-855. doi [10.1016/j.peva.2007.06.020](https://doi.org/10.1016/j.peva.2007.06.020)
- [15] Carra D, LoCigno R, Biersack EW.(2007), Graph based analysis of mesh overlay streaming systems. *IEEE Journal on Selected Areas in Communications*, 25(9), 1667-1677. doi [10.1109/JSAC.2007.071206](https://doi.org/10.1109/JSAC.2007.071206)
- [16] Zhou Jin, Lu Hai-Ming, Li Yan-Da.(2004), Using Small-World to Devise Routing Algorithm for Unstructured Peer-to-Peer System. *Journal of Software*, 15(6):915-923.
- [17] Zhang H, Ashish G.(2009), Using the small world model to improve Freenet performance. *Proc of IEEE INFOCOM 2004*, 555-574. doi [10.1016/j.comnet.2004.05.004](https://doi.org/10.1016/j.comnet.2004.05.004)
- [18] Wei Wen-hong, Xiao Wen-jun.(2009), A Structured P2P Overlay Network with Small-World Characteristics. *Journal of South China University of Technology (Natural Science Edition)*, 37(10), 66-72.
- [19] Hung-Chang H, Yung-Chih L, Hao L.(2009), Building small-world Peer-to-Peer networks based on hierarchical structures. *IEEE transactions on parallel and distributed systems*, 20(7), 1023-1036. doi [10.1109/TPDS.2008.173](https://doi.org/10.1109/TPDS.2008.173)
- [20] Tang Jiuyang, Zhang Weiming, Xiao Weidong, and Tang Daquan.(2006), Self-Organizing Peer-to-Peer Network Based on Community Mimicking Social Society. *Journal of Computer Research and Development*, 43(8), 1383-1390. doi [10.1360/crad20060811](https://doi.org/10.1360/crad20060811)
- [21] Kobayashi H, Takizawa H, Inaba T, et al.(2005), A self-organizing overlay network to exploit the locality of interests for effective resource discovery in P2P systems. *Proc. of The 2005 International Symposium on Applications and the Internet*, 246-255. doi [10.1109/SAINT.2005.8](https://doi.org/10.1109/SAINT.2005.8)
- [22] Ren Ping-fan, Huang Dong-jun.(2009), Network-aware structured P2P overlay based on small world. *Application Research of Computers*, 26(8), 3115-3118.
- [23] Rita H. Wouhaybi, Andrew T. Campbell.(2004), Phenix: Supporting resilient low-diameter peer-to-peer topologies. *Proc. of IEEE INFOCOM*, 108-119. doi [10.1109/INFCOM.2004.1354486](https://doi.org/10.1109/INFCOM.2004.1354486)
- [24] Tang Daquan, He Mingke, and Meng Qingsong.(2007), Research on Searching in Unstructured P2P Network Based on Power-Law Distribution and Small World Character. *Journal of Computer Research and Development*, 44(9), 1566-1571.
- [25] Shan Xiu-ming, Liu Yang, Zhang Lin, Wang Lei, Ren Yong, Yuan Jian, Wang Yao-xi.(2008), Complex network model of user-share behaviors in P2P applications. *Application Research of Computers*, 25(6), 1853-1855.
- [26] Ding Jun-ping, Cai Wan-dong.(2011), Spreading Dynamics Model Research on P2P Specific Information. *Computer Science*, 38(11), 96-99.
- [27] Huang Xin-li.(2008), A novel approach to local vaccination of P2P networks with power-law node degree distributions. *Systems Engineering-theory & Practice*, 11, 135-141.

This page is intentionally left blank