

DOI:10.11992/tis.201507042
网络出版地址: <http://www.cnki.net/kcms/detail/23.1538.tp.20151110.1354.012.html>

复杂网络中的在线社交网络演化模型

王景丽¹, 许立波², 庞超逸³

(1. 宁波大红鹰学院 信息工程学院, 浙江 宁波 315175; 2. 浙江大学 宁波理工学院, 浙江 宁波 315000; 3. 浙江大学 宁波理工学院, 浙江 宁波 315000)

摘 要:在线社交网络是一种广泛存在的社会网络,其节点度遵循幂率分布规律,但对于其结构演化模型方面的相关研究还不多。基于复杂网络理论研究在线社交网络内部结构特征,提出一种结合内增长、外增长及内部边更替的演化模型,借助平均场理论分析该模型的拓扑特性,实验和理论分析表明由该模型生成的网络,其度分布服从幂率分布,且通过调整参数,幂率指数在 1~3,能较好地反映不同类型的真实在线社交网络的度分布特征,因此具有广泛适用性。

关键词:复杂网络;社交网络;度分布;幂率分布;演化;平均场;节点度;拓扑
中图分类号:TP393 **文献标志码:**A **文章编号:**1673-4785(2015)06-0949-05

中文引用格式:王景丽,许立波,庞超逸. 复杂网络中的在线社交网络演化模型[J]. 智能系统学报, 2015, 10(6): 949-953.
英文引用格式:WANG Jingli, XU Libo, PANG Chaoyi. Evolution model of online social networks based on complex networks[J]. CAAI Transactions on Intelligent Systems, 2015, 10(6): 949-953.

Evolution model of online social networks based on complex networks

WANG Jingli¹, XU Libo², PANG Chaoyi³

(1. Information Engineering, Ningbo Dahongying University, Ningbo 315175, China; 2. Research Center on Intelligent Computing and Data Management Ningbo Institute of Technology, Zhejiang University, Ningbo 315000, China; 3. Research Center on Intelligent Computing and Data Management Ningbo Institute of Technology, Zhejiang University, Ningbo 315000, China)

Abstract: As a widespread social network, the node degree of online social networks has been proven by many researchers to follow the power-law distribution. However, there are few studies modeling the evolution of its structure. In this paper, we propose an evolution model that combines the inside growth, outside growth, and edge replacement based on those of complex networks. The topology properties of this model are analyzed using the mean-field theory. Experiment and theoretical analyses show that the degree of a node in a network generated by the new evolution model follows the power-law distribution and that the power-law index ranges between 1 and 3. Therefore, the proposed model can better reflect the node degree distribution characteristics of different types of real online social networks and will have wide applicability.

Keywords: complex network; social network; degree distribution; power-law distribution; evolution; mean field; node degree; topology

复杂网络作为刻画和研究复杂系统结构及其动态发展的强有力工具之一,近年来迅速成为学术界研究的热点,并与计算机科学、生物学、社会学、经济学等学科交叉融合,广泛应用于诸如生物网络、传染病控制、知识网络、电力网、互联网等各种自然和社会动态网络及系统研究中.复杂网络的研究涵盖很

收稿日期:2015-07-05. 网络出版日期:2015-11-10.
基金项目:国家自然科学基金资助项目(71271191);宁波市自然科学基金资助项目(2015A610138).
通信作者:王景丽.E-mail:jingli.wang@nbdhyu.edu.cn

多方面,其中网络的演化及建模研究能够相当程度地揭示实际网络及系统的结构组成及变化规律,从而能够对系统的未来发展做出预测并提早设置改进措施。因此演化模型一直是复杂网络研究的热点之一,并提出了很多网络演化模型,以刻画现实中各种不同网络的特征及结构。

Watts 等^[1]提出小世界模型,具有短路径、高聚类 and 强传递性特征,模拟人际关系网络的演化过程。Barabási 等^[2]在研究万维网时提出基于增长和优先连接的无标度 BA 模型,连接概率与节点度成线性正比,较好地模拟出万维网中幂率度分布特征和马太效应。之后,研究者们发现幂率度分布特征广泛存在于各种现实网络,随着 Faloutsos 等^[3]发现互联网拓扑的幂率特性,AB 模型^[4]被提出应用于互联网拓扑建模,其在 BA 模型基础上增加了内部连接和重配置过程;基于现实互联网连接倾向度更高的观测现象,广义线性优先模型^[5]改进 BA 模型以模拟新节点连接对高度 hub 的优先选择性;动态优先模型^[6]基于不同层级 AS 间的关系,反映出互联网中的小世界特性;多局域世界模型^[7-8]考虑节点连接的局域特征,能较好反映节点层次性和聚类性。众多研究表明,合作网络^[9]、软件组件系统^[10]、蛋白质网络^[11]、在线社交网络^[12]、传感器网络^[13]等都呈现出了度分布幂率特性,局部事件模型被用来模拟软件组件系统的无标度特性;顶点复制模型^[14]描述知识引用网络和新陈代谢网络,新节点不完全遵循优先连接机制,而是通过模仿已有节点来构建模型。

在线社交网络如微博、交友等是社会网络的一种典型关系网络,将每个用户看作网络节点,许多研究已经表明其节点度遵循幂率分布规律,但对于其结构演化模型方面的相关研究还比较少。一方面,目前研究集中于通过抽样数据来进行数据分析,从统计角度发现拓扑特征,虽然能较好还原真实网络特性,但这种研究属于静态分析,不能模拟动态过程以及添加行为过程;另一方面,研究表明在线社交网络节点可能比一般网络更倾向于连接度大的节点,很多传统模型不能直接适用,需要进行一定的改进。

本文提出一种在线社交网络演化模型,以注册用户作为网络节点,节点间的连边就是注册用户间的关注关系,节点和连边共同构成网络,不失一般性,本文将此网络抽象成无向网络图。演化模型考虑随机增长、优先连接、内部增长及连接更替 4 种机制,采用平均场理论对模型度分布进行理论分析。理论分析和仿真表明,演化模型网络的度呈现无尺度特征,且能够模拟较大范围的幂率分布,具有广泛的适用性,并

能一定程度地模拟真实在线社交网络的度分布特征。

1 在线社交网络演化模型

本文构造无向的网络演化模型,模拟了内部增长即网络中现有节点间添加新连接、内部调整即现有节点间重连接及网络增长即新节点和现有节点的连接等机制,对于有向网络亦可以参照文中分析。具体演化步骤如下:

1) 初始化:初始网络包含 m_0 个节点,每个节点与其余 m_0-1 个节点相连;

2) 随机增长:每个时间步添加一个新节点,并连接到 m_1 个原有节点上,且满足 $m_1 < m_0$;

3) 优先连接:新节点与原有节点 i 相连接的概率 Π_i 表达式为

$$\Pi_i = \frac{k_i}{\sum_j k_j} \quad (1)$$

式中: k_i 为节点 i 的度;

4) 内部随机增长:每个时间步在网络内增加 m_2 条边,节点 i 被选为新增边的端点的概率 Π_i 按式(1)计算,且 $m_2 < m_0$;

5) 内部优先连接更替:每个时间步选择 m_3 条边进行更替,更替方法为先随机选择节点 i ,随机移除与节点 i 相连的某条边 l_{ij} ,添加以连接节点 j 和节点 r 的边 l_{jr} ,节点 r 按 Π_r 概率择优选取, Π_r 按式(1)计算。

相比一般网络,在线社交网络中的连接度大的节点可能更容易得到大量连接,因此在演化模型中,除了新节点和内部增长具有优先连接机制外,内部优先连接更替也不同于一般模型中的重连机制,后者通常是以边 l_{ir} 替代 l_{ij} ,节点 i 的度不发生变化;而内部优先连接更替是以 l_{jr} 替代 l_{ij} ,更加强化高连接度节点间的连接倾向,且为防止出现孤立节点,模型设定当被选择节点度为 1 时,进行重新选点。

2 演化模型的度分布属性

根据平均场理论对模型的度分布进行近似分析,假设节点度 k_i 随时间 t 连续变化,对 k_i 变化率求解:

随机增长:当新节点加入时,按概率 Π_i 连接到 m_1 个原节点上,则 k_i 变化率方程为

$$\frac{\partial k_i(t)}{\partial t} = m_1 \Pi_i \quad (2)$$

内部随机增长:边的端点按概率 Π_i 优先选择,边有 2 个端点,则 k_i 变化率方程为

$$\frac{\partial k_i(t)}{\partial t} = 2m_2 \Pi_i \quad (3)$$

内部优先连接更替:包含 2 个过程,首先是移除

过程,节点 i 被随机选中的概率为 $\frac{m_3}{N}$, N 为节点总数,被选择为移除边的另一端的概率为 $m_3\Pi_i$;其次是更替过程,被选择为更替边的两端的概率为 $2m_3\Pi_i$,因此整个更替过程 k_i 变化率方程为

$$\frac{\partial k_i(t)}{\partial t} = -\frac{m_3}{N} - m_3\Pi_i + 2m_3\Pi_i = m_3\Pi_i - \frac{m_3}{N} \tag{4}$$

综合式(2)~(4),得到演化模型 k_i 变化率方程为

$$\frac{\partial k_i(t)}{\partial t} = m_1\Pi_i + 2m_2\Pi_i + m_3\Pi_i - \frac{m_3}{N} \tag{5}$$

由网络节点平均度数为 $2(m_1 + m_2)$,并且对于充分大的 t , t 时刻节点总数 N 为 $m_0 + t \approx t$,因此网络节点总度数 $\sum_j k_j \approx 2(m_1 + m_2)t$,得

$$\frac{\partial k_i(t)}{\partial t} = (m_1 + 2m_2 + m_3)\Pi_i - \frac{m_3}{N} \approx \frac{(m_1 + 2m_2 + m_3)k_i(t)}{2(m_1 + m_2)t} - \frac{m_3}{t} \tag{6}$$

用变量分离法解此一阶微分方程可得

$$k_i(t) = \frac{2m_3(m_1 + m_2)}{m_1 + 2m_2 + m_3} + Ct^{\frac{m_1 + 2m_2 + m_3}{2(m_1 + m_2)}} \tag{7}$$

设节点 i 是在 t_i 时刻进入网络,则有初始条件 $k_i(t_i) = m_1$,结合式(7)可得

$$k_i(t) = \frac{2m_3(m_1 + m_2)}{m_1 + 2m_2 + m_3} + \left[m_1 - \frac{2m_3(m_1 + m_2)}{m_1 + 2m_2 + m_3} \right] \left(\frac{t}{t_i} \right)^{\frac{m_1 + 2m_2 + m_3}{2(m_1 + m_2)}}$$

网络中节点 i 的度 $k_i(t)$ 小于 k 的概率为

$$P(k_i(t) < k) = P\left(\frac{2m_3(m_1 + m_2)}{m_1 + 2m_2 + m_3} + \left[m_1 - \frac{2m_3(m_1 + m_2)}{m_1 + 2m_2 + m_3} \right] \left(\frac{t}{t_i} \right)^{\frac{m_1 + 2m_2 + m_3}{2(m_1 + m_2)}} < k\right) \\ P(t_i > \left[\frac{m_1 - a}{k - a} \right]^{\frac{m_1 + 2m_2 + m_3}{2(m_1 + m_2)}} t) \tag{8}$$

式中: $a = \frac{2m_3(m_1 + m_2)}{m_1 + 2m_2 + m_3}$ 模型假设节点是以相对时间间隔进入网络的,因此节点 i 的到达时间 t_i 具有常数概率密度:

$$P(t_i) = 1/N \approx 1/t$$

故有

$$P(t_i > \left[\frac{m_1 - a}{k - a} \right]^{\frac{m_1 + 2m_2 + m_3}{2(m_1 + m_2)}} t) = 1 - P(t_i \leq \left[\frac{m_1 - a}{k - a} \right]^{\frac{m_1 + 2m_2 + m_3}{2(m_1 + m_2)}} t) \approx$$

$$1 - \left[\frac{m_1 - a}{k - a} \right]^{\frac{2(m_1 + m_2)}{m_1 + 2m_2 + m_3}}$$

假设时间 $t \rightarrow \infty$ 时,网络节点的度具有稳态分布 $P(k)$:

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

$$\frac{2(m_1 + m_2)}{m_1 + 2m_2 + m_3} (m_1 - a)^{\frac{2(m_1 + m_2)}{m_1 + 2m_2 + m_3}} (k - a)^{-\gamma} \tag{9}$$

则度分布指数

$$\gamma = 2 - \frac{m_3 - m_1}{m_1 + 2m_2 + m_3} \tag{10}$$

上述理论分析表明,由本演化模型生成的网络具有无标度特性,且通过调整 m_1 、 m_2 和 m_3 的取值,能得到不同的分布指数。由 $-1 \leq \frac{m_3 - m_1}{m_1 + 2m_2 + m_3} < 1$,可得 $1 < \gamma \leq 3$ 。当 $m_2 = m_3 = 0$ 时, $\gamma = 3$,无标度 BA 模型成为本演化模型的一个特例。现有研究表明,绝大多数无标度网络的幂指数都在 1~3,因此本演化模型具有广泛的适用性。

3 仿真实验和讨论

仿真实验分为 2 部分:1)采用 MATLAB 软件开发仿真程序,通过程序运行的结果来分析演化模型的度统计特性、相关参数的影响以及理论分析的准确性;2)通过 MATLAB 软件分析真实社交网络数据的度统计特性,观察演化模型理论分析值的模拟程度。

分析度分布指数 γ 的表达式(10),可以看到 3 个参数的作用各不相同, m_1 和 m_3 决定 γ 取值大于 2 或者小于 2, m_2 则只影响值的大小,而 m_1 和 m_2 又影响网络的节点规模和平均度数。图 1、2、3 分别显示了 1 500 节点网络演化模型在不同参数取值下的度分布情况,演化网络度分布呈现幂率特征,且根据参数取值的不同在 1~3 之间变化,图中直线由度分布理论表达式(9)近似给出。

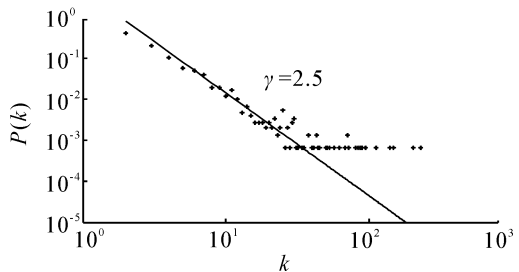


图 1 演化模型度分布 ($m_1 = 2, m_2 = 1, m_3 = 0$)

Fig.1 Degree distribution of evolution model ($m_1 = 2, m_2 = 1, m_3 = 0$)

可以看到,在 m_1 和 m_2 值不变的条件下, m_3 值越大,无标度特性越大。图 4 则显示在 m_1 和 m_3 值不变的条件下, m_2 值变化对度分布幂率的影响,分布幂率值随着 m_2 值变大而增大,但逐步接近某固定值,由式 (10) 可知,此值为 2。

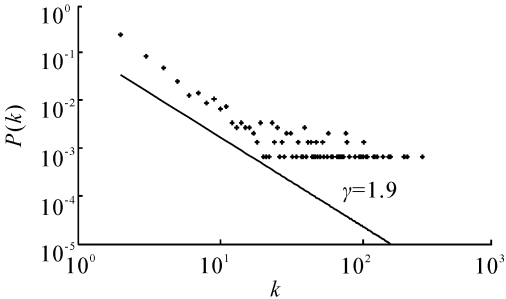


图 2 演化模型度分布 ($m_1=2, m_2=1, m_3=3$)

Fig.2 Degree distribution of evolution model ($m_1=2, m_2=1, m_3=3$)

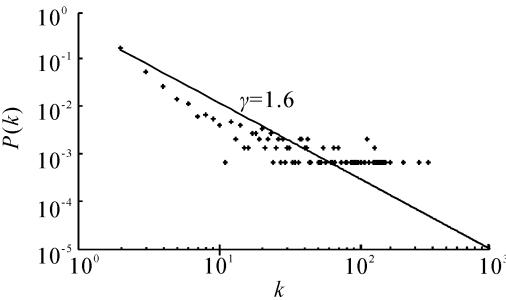


图 3 演化模型度分布 ($m_1=2, m_2=1, m_3=6$)

Fig.3 Degree distribution of evolution model ($m_1=2, m_2=1, m_3=6$)

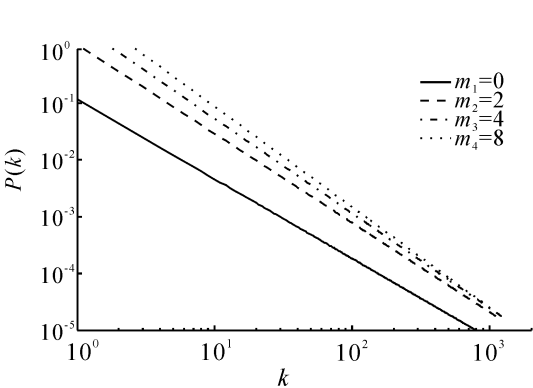


图 4 m_2 取不同值时的度分布变化 ($m_1=2, m_3=8$)

Fig.4 Degree distribution changes according to different values of m_2 ($m_1=2, m_3=8$)

采用真实在线社交网络与本文演化模型的度分布进行对比分析。图 5 显示一个美国政治家博客用户网络^[15]的度分布情况,该网络节点为博客网页,边表示为网页间的超链接,包含 1 224 个节点和 16 714 条无向边。可以看到,博客网页网络的度分布幂率在 1~2,传统很多演化模型如 BA 及其变型

模型的度分布幂率在 2~3,不能较好地表征博客网络的度分布特征。图中拟合直线幂率值为 1.4,由本文表达式 (9) 直接给出,其中参数取值 $m_1=2, m_2=0, m_3=8$,实际上有很多种参数取值方案,这里只给出其中一种仅用于说明效果。通过调整参数,演化模型能较好地模拟美国政治家博客用户网络的度分布特征。

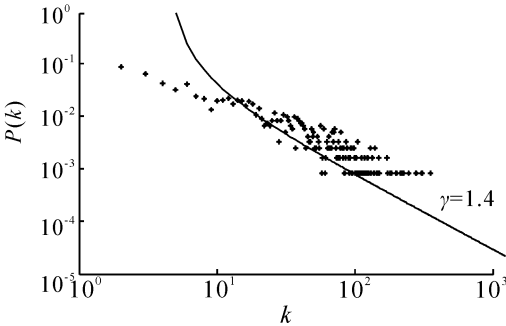


图 5 政治家博客用户网络度分布幂率

Fig.5 Power-law degree distribution network of Political blogs

图 6 显示维基选举社交网络^[15]度分布情况,该网络节点为维基用户,边表示为用户间的选举行为,包含 7 000 多节点。该网络度分布幂率也在 1~2 之间,传统模型不能较准确地表征。图中拟合直线幂率值为 1.5,由表达式 (9) 直接给出,其中参数取值 $m_1=2, m_2=0, m_3=6$,这里同样只给出一种取值方案用于说明效果。可以看到,通过调整参数,演化模型能较好地模拟维基选举社交网络的度分布特征。

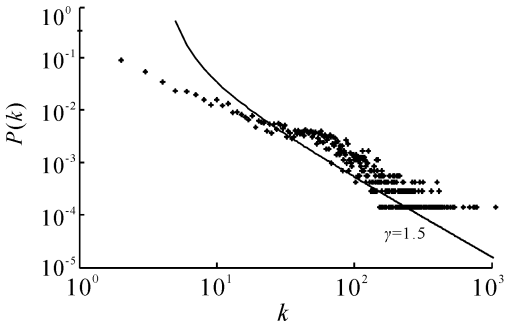


图 6 维基选举社交网络度分布幂率

Fig.6 Power-law degree distribution of Wiki vote

4 结束语

在线社交网络作为一种普遍存在的社会网络,已经引起广泛的研究并证实其具有无标度特性,但对其结构演化模型的研究相对较少。本文提出一种基于内增长、外增长及内部边更替的在线社交网络演化模型,其优先连接机制具有强化高连接度节点间的连接倾向的特征。理论分析和实验表明,该模型具有无标度特性,且通过调整参数的不同取值,度

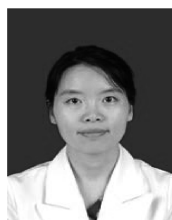
分布幂率指数在1~3,具有普遍适用性。通过和真实在线社交网络结构的模拟分析,本文提出的演化模型能较好地反映出在线社交网络的度分布特征,且能够刻画不同类型的社交网络,具有较强的实用意义。

本文演化模型是基于最基本的无向无权网络,下一步的工作考虑扩展到复杂的多属性异构网络,这些属性包括有向、权值、节点能力等。借助因素空间理论^[16-17],将多属性异构网络映射成因素空间网络,网络节点映射成因素空间对象,使其更加贴近于现实网络,通过权变、拓扑结构、动态时空等因素空间理论工具研究演化网络模型和实际在线社会网络的生成、发展及其特征。

参考文献:

- [1] WATTS D J, STROGATZ S H. Collective dynamics of 'small-world' networks [J]. *Nature*, 1998, 393 (6684): 440-442.
- [2] BARABASI A L, ALBERT R. Emergence of scaling in random networks[J]. *Science*, 1999, 286(5439): 509-512.
- [3] FALOUTSOS M, FALOUTSOS P, FALOUTSOS C. On power-law relationships of the Internet topology[J]. *ACM SIGCOMM Computer Communication Review Homepage*, 1999, 29(4): 251-262.
- [4] ALBERT R, BARABASI A L. Topology of evolving networks: local events and universality[J]. *Physical Review Letters* 2000, 85(24): 5234-5237.
- [5] BU T, TOWSLEY D. On distinguishing between Internet power law topology generators[C]//*Proceedings of INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies*. New York, NY: IEEE, 2002, 2: 638-647.
- [6] PARK S T, PENNOCK D M, GILES C L. Comparing static and dynamic measurements and models of the Internet's as topology[C]//*Proceedings of the 23rd Annual Joint Conference of the IEEE Computer and Communications Societies*. Hong Kong: IEEE, 2004, 3: 1616-1627.
- [7] CHEN Guangrong, FAN Zhenping, LI Xiang. Modeling the complex Internet topology[C]//KOCAREV L, VATTAY G. *Complex Dynamics in Communication Networks*. Berlin Heidelberg: Springer, 2005.
- [8] 田思, 李慧嘉, 赵岳. 一种新型多局域世界网络模型分析[J]. *计算机应用研究*, 2013, 30(3): 869-872.
TIAN Si, LI Huijia, ZHAO Yue. Analysis of novel multi-local world network model[J]. *Application Research of Computers*, 2013, 30(3): 869-872.
- [9] NEWMAN M E J. The structure of scientific collaboration networks[C]. *Proceedings of National Academy Sciences*, 2001, 98(2): 404-409.
- [10] MYERS C R. Software systems as complex networks: structure, function, and evolvability of software collaboration graphs[J]. *Physical Review E*, 2003, 68 (4Pt2): 046116.
- [11] JEONG H, MASON S P, BARABASI A L, et al. Lethality and centrality in protein networks[J]. *Nature*, 2001, 411 (6833): 41-42.
- [12] 王亚奇, 王静, 杨海滨. 基于复杂网络理论的微博用户关系网络演化模型研究[J]. *物理学报*, 2014, 63(20): 208902.
WANG Yaqi, WANG Jing, YANG Haibin. An evolution model of microblog user relationship networks based on complex network theory[J]. *Acta Physica Sinica*, 2014, 63(20): 208902.
- [13] 刘浩然, 尹文晓, 董明如, 等. 一种强容侵能力的无线传感器网络无标度拓扑模型研究[J]. *物理学报*, 2014, 63(9): 090503.
LIU Haoran, YIN Wenxiao, DONG Mingru, et al. Study on the scale-free topology model with strong intrusion-tolerance ability in wireless sensor networks[J]. *Acta Physica Sinica*, 2014, 63(9): 090503.
- [14] SOLE R V, PASTOR-SATORRAS R, SMITH E, et al. A model of large-scale proteome evolution[J]. *Advances in Complex Systems*, 2002, 5(1): 43-54.
- [15] 维基选举社交网络[OL/EB]. [2014-12-21]. <http://www.linkprediction.org/index.php/link/resource/data>.
- [16] 汪培庄. 因素空间与因素库[J]. *辽宁工程技术大学学报: 自然科学版*, 2013, 32(10): 1297-1304.
WANG Peizhuang. Factor spaces and factor data-bases[J]. *Journal of Liaoning Technical University: Natural Science*, 2013, 32(10): 1297-1304.
- [17] 汪培庄. 因素空间与数据科学[J]. *辽宁工程技术大学学报: 自然科学版*, 2015, 34(2): 273-280.
WANG Peizhuang. Factor spaces and data science[J]. *Journal of Liaoning Technical University: Natural Science*, 2015, 34(2): 273-280.

作者简介:



王景丽,女,1981年生,讲师。主要研究方向为复杂网络。



许立波,男,1976年生,讲师,博士。主要研究方向为复杂网络、overlay网络。