

朋友关系网络的实证统计研究

张 恺, 马忠军, 李科赞

(桂林电子科技大学数学与计算科学学院 桂林 541004)

【摘要】为了探究节点的动力学行为如何对网络拓扑结构造成影响,构建了4个节点相同但拓扑结构不同的朋友网络。对比研究发现,普通朋友网络是一个小世界网络,但随着朋友关系亲密度的增加,网络拓扑结构由连通到不连通,平均度和集聚系数在减少而平均路径长度在增加。此外,前3个网络的度相关性呈现负相关。

关键词 复杂网络; 有向网络; 实证研究; 朋友网络; 统计特征; 拓扑结构

中图分类号 O231.1

文献标志码 A

doi:10.3969/j.issn.1001-0548.2014.03.003

An Empirical Statistical Investigation on Friendship Networks

ZHANG Kai, MA Zhong-jun, and LI Ke-zan

(School of Mathematics & Computing Science, Guilin University of Electronic Technology Guilin 541004)

Abstract The dynamic behavior of nodes can significantly influence the topological structure in a network. In order to study this process, four friendship networks with same nodes and different topologies have been constructed. Comparison reveals that 1) the normal-friendship network is a small-world network, 2) the connected structure becomes disconnected along with the increasing of relationship intimacy, and 3) both average degree and clustering coefficient are becoming smaller while the average path length is getting larger. In addition, the degree correlations of the first three networks are negative correlated.

Key words complex networks; directed network; empirical study; friendship network; statistical characteristics; topological structure

复杂网络的研究始于20世纪60年代提出的ER随机图模型,直到小世界网络模型^[1]及无标度网络模型^[2]的提出,才真正令复杂网络的研究成为一个令人瞩目的领域。用复杂网络的观点和矩阵、图论等数学工具来研究社会中存在的各种关系,近几十年来一直是复杂网络中的研究热点^[3-4]。

朋友网作为一种典型的社会网络,受到了复杂网络领域诸多学者的关注。顾名思义,朋友网是根据人们之间的朋友关系所建立的网络,以人为节点,两人之间若有朋友关系则连接一条边。一些学者之前对朋友网的研究也获得了许多成果^[5-7]。本文将朋友网络的统计描述及各网络之间的数据对比等作为研究重点。

随着科技的发展,关于朋友网络的研究主要集中于在线交友网站中的好友关系。如美国的Facebook^[8], Twitter^[9]; 中国的QQ^[10], 微博^[11], 人人网^[12]等,都取得了显著的成果。广大学者之所以将研究的兴趣集中于在线网络,而不是真实网络,

原因在于现实朋友网络的数据不易进行大范围的有效收集且成本较高,而在线交友网络却拥有大量易收集的数据。尽管每个人都能感受到虚拟世界的好友关系与现实生活中的朋友关系很不一样,但在在线朋友关系网络却展现了很多与现实朋友关系网络一致的特点。上面的一些文献也很好验证了这一点,更多在线网络的相关内容可参考文献[13-15]。

以上文献虽然取得了不错的成果,但关于真实朋友网络的研究并不多见(文献[6]是研究真实朋友网络的并不多见的成果之一),且实证网络的研究中又往往将研究对象限于一个固有网络,没有涉及在网络节点不变的情况下节点动力学行为是否影响网络拓扑结构这一问题。本文以具有31个节点的真实朋友关系网为实例,通过研究发现,对于不同亲密度的朋友关系网络,对应的网络拓扑结构亦会明显不同。通过对网络各项统计特征的对比研究,揭示了这一变化的内在原因。

收稿日期: 2012-10-26; 修回日期: 2013-10-14

基金项目: 国家自然科学基金(61004101, 11162004); 广西自然科学基金(2013GXNSFAA019006, 2012GXNSFAA053006)

作者简介: 张恺(1987-), 男, 硕士生, 主要从事复杂网络和混沌同步等方面的研究。

1 朋友关系网络的构建

朋友关系网是社会与社交网络的一种, 在近年来的研究中愈发受到关注。本文主要研究的对象是基于一个班级内同学之间的朋友关系所构成的真实网络。

本文研究的数据来源于真实的调查问卷, 问卷对象选择桂林电子科技大学本科部的一个班级。具体操作流程如下: 1) 设置问卷。问卷上有全班同学的姓名, 采用具名调查方式。按照两人之间亲密程度, 将朋友关系划分为如下4个层次: “普通朋友” “比较好的朋友” “很好的朋友” 以及 “最好的朋友”; 填写选项只需选择 “是(用√表示)” 或 “否(用×表示或直接不填写)” 即可。2) 对象选择。选取一个容量为36人的班级作为调查对象。3) 具体操作。选择在课间休息时间发放和回收问卷, 为了在一定程度上保证问卷答案的真实性, 需要同学们在填写时不要相互交谈, 避免调查过程中出现 “舞弊” 现象。

以上得到的只是原始数据, 尚需要进行一些必要的处理, 具体说明如下: 1) 缺席问题。调查时没有到场的同学, 其数据自然作废, 相应的也要删除其他同学的问卷中与缺席者相关的数据, 以保证数据的完整性。2) 随意填写现象。有极少同学只填写 “最好的朋友” 一栏, 且对全部同学皆选择 “是” 这一选项, 其数据明显不符合标准, 也予以删除。3) 虽然在实际中得到的网络都是有向的, 但为了方便和统一, 如无特别说明, 采用的都是无向图(即每条边皆按双向, 单向亦按双向设定)。

最终得到31人的有效问卷数据, 以下的分析皆基于本次调查所得到的结果。由于将朋友关系划分为4个层次, 故最终可以得到4个不同的网络。这些网络之间并不是毫无关系的, 因为所有网络具有相同的节点。文章的研究重点集中在随着朋友关系亲密程度的增强, 网络拓扑结构与统计性质所发生的变化。

为了对文中的朋友关系网络在直观上有一个初步的了解, 首先给出它们的拓扑关系图, 为了使图像清晰简洁, 忽略网络的有向性, 故得到的图形皆为从有向图基础上得到的无向图。得到4种情况下的网络拓扑结构图如图1所示。

从图1中可以直观地观察到, 4个网络图首先在拓扑结构上就有了显著的变化: 从复杂变得简单, 从连通变成不连通等。改变源自个人与他人之间朋友关系的亲密程度发生了变化, 随之引发了整个网

络拓扑结构的变化。

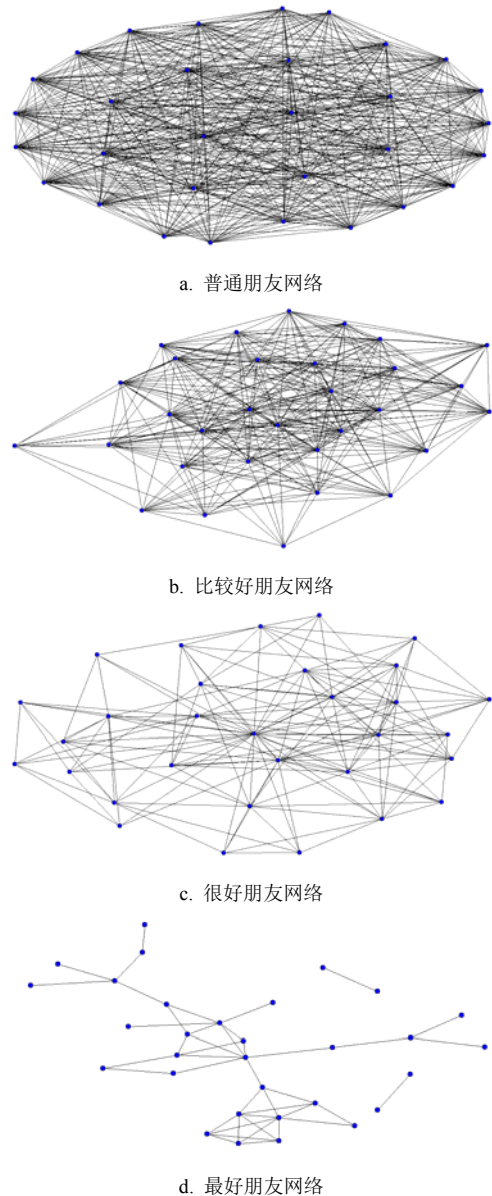


图1 网络拓扑结构图

为了叙述方便, 分别将图1a~图1d四个网络称为 “网络1”、“网络2”、“网络3” 和 “网络4”。

2 网络的基本统计特征

随着近些年对复杂网络研究的逐步深入, 学者提出了大量概念和度量方法^[16-19], 用于衡量一个网络的复杂性。

2.1 无向网络的度与平均度

度(degree)是复杂网络的一个基本且重要的统计特征。第*i*个节点的度 k_i 指的是与该节点相连接的其他节点的数目, 其数值可以在一定程度上反映该节点在整个网络中的重要程度。全部节点的度的算术平均值称之为网络的平均度, 用 $\langle k \rangle$ 来表示, 即

$$\langle k \rangle = \left(\sum_i k_i \right) / N \quad (1)$$

式中, N 表示网络中的节点数目; k_i 表示第 i 个节点的度值。在朋友网络中, k 表示一个人的朋友数, $\langle k \rangle$ 则表示网络中平均每人有多少朋友。经过计算, 得到4个网络的平均度分别为27.806, 15.903, 6.322 6和1.806 5。

可见, 随着朋友关系的亲密度逐渐增加, 网络的平均度也在逐渐下降, 从平均每个人大约有28个普通朋友到平均每个人只有不到2个最好的朋友。数据的差异也说明了图1中网络从复杂变的愈发简单的根本原因。

2.2 无向网络的平均路径长度

最短路径长度指的是从节点 i 到 j 要经历的边的最小数目, 记作 $d(i, j)$ 。平均路径长度(average path length) $\langle d \rangle$ 指的是网络中所有节点对之间最短路径长度的算术平均值, 即:

$$\langle d \rangle = \frac{1}{N(N-1)} \sum_{i \neq j} d(i, j) \quad (2)$$

式中, N 是节点数目。

若按有向图研究, 则网络2、网络3和网络4这3个都不是强连通图, 平均路径长度也就无从谈起; 若按无向图研究, 则只有网络4是不连通的。若按无向图考虑, 则网络2、网络3亦可计算平均路径长度, 但网络4由于有4个点与其他点无相连(这从上面给出的网络拓扑图中可以清晰的看出来), 故无法求其平均路径长度。将网络1、网络2和网络3按无向图进行处理, 可得到其平均路径长度分别为1.008 6, 1.311 8、1.311 8。

数据的变化不难理解, 朋友少了, 两个人之间的“距离”也就拉大了。数据的变化也暗合了中国的俗语“多一个朋友多一条路”。

2.3 无向网络的集聚系数

集聚系数(clustering coefficient)也称群聚系数、集群系数或聚类系数, 用来描述节点的邻点之间也互为邻点的比例。节点 i 的集聚系数 C_i 等于所有与它相连的节点相互之间所连的边的数目除以这些节点之间可以连出的边数的最大值, 即:

$$C_i = 2e_i / k_i(k_i - 1) \quad (3)$$

式中, k_i 表示节点 i 的度; e_i 表示节点 i 的所有邻接点之间实际存在的边数。显然 C_i 是一个介于0与1之间的数。 C_i 越接近1, 表示这个节点附近的点越有“抱团”的趋势。整个网络的集聚系数 C 为所有点集聚系数的算术平均值, 即:

$$C = \frac{1}{N} \sum_{i=1}^N C_i \quad (4)$$

式中, N 为网络中的节点数目。通过计算, 得到4个网络的集聚系数 C 分别为0.928 6、0.590 14、0.325 65和0.156 45。

从数据中可以看出, 网络1的集聚系数高达0.928 6, 在上文中还得到网络1的平均路径长度为1.008 6。网络1的集聚系数与平均路径长度的值都接近1, 这表明其接近完全耦合网络。文献[1]揭示, 小世界网络模型的统计特征为高集聚系数和低平均路径长度, 由于网络1具有上述特征, 故网络1(即普通朋友网络)显然是一个小世界网络。随着朋友关系亲密度的增加, 网络的集聚系数逐渐变小, 逐渐丧失了小世界特性, 最终网络变得不连通。

2.4 无向网络的介数

介数^[16,20](betweenness)亦称边介、中心等, 通常分为边介数和节点介数两种, 反映了节点或边的作用和影响力。本文中所探讨的是点介数, 节点 i 的介数 $B(i)$ 具体算法^[20]如下:

$$B(i) = \frac{2 \sum_{j < k} g_{jk}(i) / g_{jk}}{(n-1)(n-2)} \quad (4)$$

式中, g_{jk} 表示 j 与 k (两者都不等于 i) 之间的路径数目; $g_{jk}(i)$ 表示 j 与 k 之间经过 i 的路径总数, 除以 $(n-1)(n-2)/2$ 表示归一化。

通过计算, 得到网络中各节点介数如图2所示。从图2中可以观察到各网络介数值最大的节点, 如表1所示。

表1 各网络中介数最大点

	网络1	网络2	网络3	网络4
介数最大点	第17个、21个	第21个	第3个	第26个

介数值最大的节点对于整个网络的拓扑结构有着重要的影响。网络1介数值最大的为第21个和第17个, 网络2则对应第21个, 网络3变成了第3个, 而网络4则变成了第26个。由下文中的图3可以看出, 它们并不一定是度最大的节点(有向网络里节点的出度与入度中较大的那个值等于对应无向化网络对应节点的度)。这个研究结果表明, 介数值最大的节点不一定是度最大的节点, 反之亦然。此外, 这也从一个侧面表明, 网络的拓扑结构有重大变化。随着朋友关系亲密程度的增加, 网络中最“亲密的人”也发生改变。

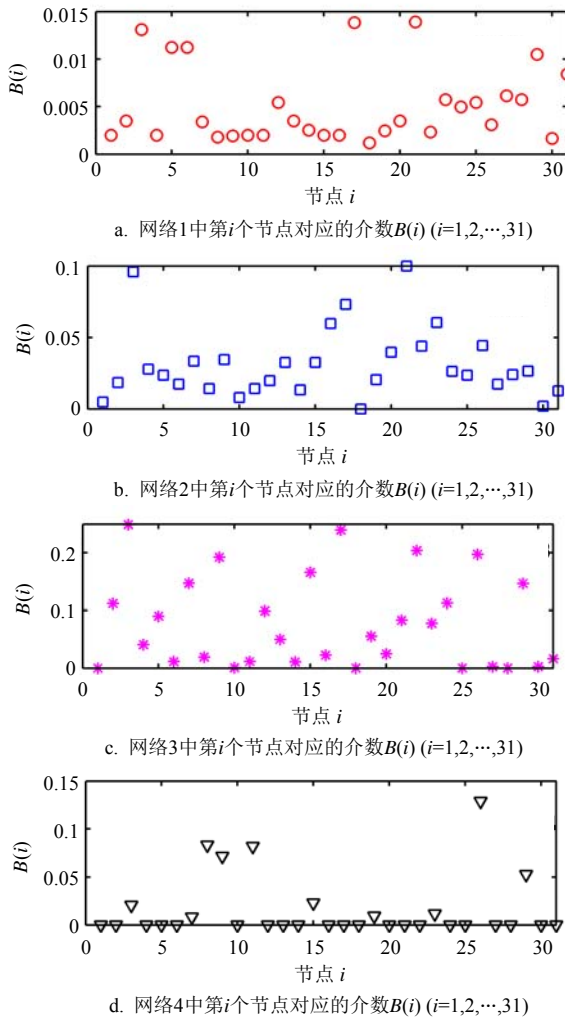


图2 网络*j*中的各节点点介数(*j*=1,2,3,4)

2.5 无向网络的度相关性

度相关性(degree correlations)^[14]描述的是网络中不同节点之间的连接关系,用*r*表示。如果度大的节点倾向于连接度大的节点,则称网络是正相关的;反之,如果度大的节点倾向于和度小的节点连接,则称网络是负相关的。文献[21-22]给出了一种相关性的计算方法,即计算网络节点度的Pearson相关系数*r*:

$$r = \frac{M^{-1} \sum_i j_i k_i - \left[M^{-1} \sum_i \frac{1}{2} (j_i + k_i) \right]^2}{M^{-1} \sum_i \frac{1}{2} (j_i^2 + k_i^2) - \left[M^{-1} \sum_i \frac{1}{2} (j_i + k_i) \right]^2} \quad (5)$$

式中, j_i, k_i 分别表示连接第*i*条边的两个顶点*j*和*k*的度; *M*表示网络的总边数; *r*的取值范围为 $-1 \leq r \leq 1$, 当 $r > 0$ 时, 网络是正相关的; 当 $r < 0$ 时, 网络是负相关的; 当 $r = 0$ 时, 网络是不相关的。文献[21-22]还指出网络对应的度相关性与其类型有关, 如演员合作网、中学生的朋友关系网等社会网

络往往是正相关的, 而电力网、因特网、海洋食物链网等技术和生物网络往往是负相关的^[22]。其后的研究表明, 并非所有社会网络的度相关性都为正^[16]。

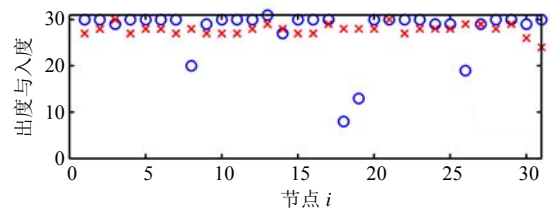
本文中, $r > 0$ 表示朋友多的同学更容易和朋友多的同学结成朋友关系, 而 $r < 0$ 表示朋友多的同学更容易和朋友少的同学结成朋友关系。经过计算得到4个网络的相关系数分别为-0.079 249、-0.105 34、-0.159 57和0.116 12。

通过对比数据发现: 在前3个网络中, 其*r*值均为负, 与通常所说的社会网络往往正相关相悖。事实上, 文中所研究的朋友关系网络是一个节点数只有31的小网络, 与演员网络、科研合著网络等动辄几百上千个节点的实证网络是无法比拟的, 更无法与在线网络千万级的节点数作比较。因此, 自身的网络规模决定了所研究的网络与文献中已有内容的不同。且上文也提及, 不一定所有的社会网络其*r*参数都取正, 负相关的结论也可为之提供些参考价值。另一方面, 朋友关系网络是一个复杂的实证网络, 朋友的结成并不如在线网络那样纯粹而直接, 同乡、同宿舍、同爱好等诸多原因影响了朋友关系的建立, 各种错综复杂的因素结合在一起, 使网络呈现出与其他网络截然不同的统计特征。第4个网络的*r*值又变回正值, 也从一个侧面反映了网络的复杂性。

2.6 有向网络的出度与入度

上述讨论都基于所有网络是无向的这一假设。但在实际中, 两人之间的朋友关系并不是相互的, 即网络是有向的。朋友网络中一个节点的出度表示本节点对应的人所认可的朋友数, 入度表示把该节点对应的人作为朋友的人数。分析所获得的有向网络模型, 统计并绘制各网络节点的出度与入度分布如图3所示。(“○”表示出度, “×”表示入度)

从图3中可以明显看出, 大多数节点的出度与入度值相差很小, 但也有一些节点, 两者的值差距很大。特别是第22个节点, 出度分别为30、30、30、6, 而入度却分别只有27、12、3、0, 其在各个网络中都是对应出度最大的节点之一, 而出度与入度的差值是最大的。尤其在很好朋友网络和好朋友网络中, 入度更是减少到了3和0。



a. 网络1中第*i*个节点对应的出度与入度(*i*=1,2,...,31)

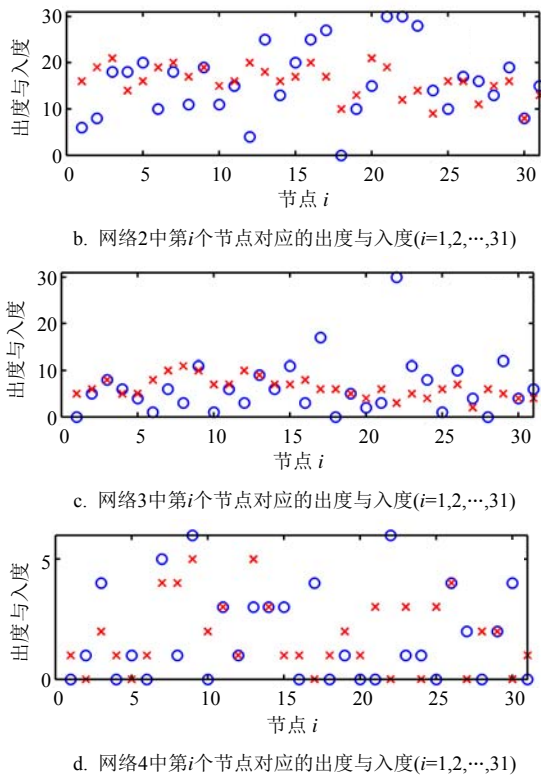


图3 各网络中节点的出度与入度

3 结论

本文基于复杂网络理论,对某班级同学之间的朋友关系网络进行了研究和探讨。通过对4个不同网络的分析比较,发现个人的朋友选择行为(个体的动力学行为)对于整个朋友网的构成(网络的拓扑结构)有着重要的影响。

直观看无向化后得到的拓扑图,网络结构从复杂变得简单,由连通图转化为不连通图;从统计性质看,网络的平均度和集聚系数愈变愈小;平均路径长度越来越大,直至出现网络的不连通性;度相关性亦由负相关变为正相关。网络的节点没有发生变化,而网络从具有小世界特性到逐渐消失该特性。普通朋友网络是小世界的,而最要好朋友的网络是不连通的。若考虑有向图,则有些节点虽具有较大的出度,但入度很小。

虽然本文只选取了一个仅有36人容量的班级作为研究对象,且最后得到的有效数据只有31个,但可以为研究更大规模朋友网络的拓扑性质提供参考。

虽然采集数据量较少,希望以后可以有条件和条件研究具有更多节点的网络。本文只是对朋友关系网络进行的一个初步实证研究,还有更多的网络性质变化情况在本文中并未涉及,这也是将来继续要做的工作。除此之外,对于该网络上的各种动态

行为,如同学之间流言的传播能力、兴趣由个人到群体的扩散过程、各种影响朋友关系网络的形成等尚待进一步深入研究。

参 考 文 献

- [1] WATTS D J, STROGATZ S H. Collective dynamics of small-world Networks[J]. Nature, 1998, 393(4): 440-442.
- [2] BARABÁSI A L, ALBERT R. Emergence of scaling in random networks[J]. Science, 1999, 286(5439): 509-512.
- [3] 周涛,汪秉宏,韩筱璞,等. 社会网络分析及其在舆情和疫情防控中的应用[J]. 系统工程学报, 2010, 25(6): 742-754.
ZHOU Tao, WANG Bing-hong, HAN Xiao-pu, et al. Social network analysis and its application in the prevention and control of propagation for public opinion and the epidemic[J]. Journal of Systems Engineering, 2010, 25(6): 742-754.
- [4] BORGATTI S P, MEHRA A J, BRASS D J, et al. Network analysis in the social sciences[J]. Science, 2009, 323(5916): 892-895.
- [5] BALL B, NEWMAN M E J. Friendship networks and social status[DB/OL].[2012-05-30]. <http://arxiv.org/abs/1205.6822>.
- [6] YIN Hai-ping, RONG Zhi-hai, YAN Gang. Development of friendship network among young scientists in an international summer school[J]. Physica A, 2009, 388(17): 3636-3642.
- [7] EAGLE N, PENTLAND A, LAZER D. Inferring friendship network structure by using mobile phone data[J]. Proceedings of the National Academy of Sciences, 2009, 106(36): 15274-15278.
- [8] GOLDR S A, WILKINSON D, HUBERMAN B A. Rhythms of social interaction: messaging within a massive online network[C]//Proceedings of the Third Communities and Technologies Conference. London: Springer, 2007: 41-66.
- [9] KWAK H, LEE C, PARK H, et al. What is twitter, a social network or a news media?[C]//Proceedings of the 19th International Conference on World Wide Web. New York: ACM, 2010: 591-600.
- [10] 杨洪勇,李凯旋,林娜. 基于复杂网络的学生交流网络模型[J]. 控制工程, 2008, 15(4): 437-439.
YANG Hong-yong, LI Kai-xuan, LIN Na. Student intercommunication model based on complex networks[J]. Control Engineering of China, 2008, 15(4): 437-439.
- [11] 尹书华. 基于复杂网络的微博用户关系网络特性研究[J]. 西南师范大学学报(自然科学版), 2011, 36(6): 57-61.
YIN Shu-hua. A research of user relations properties based on a complex network of microblog[J]. Journal of Southwest China Normal University(Natural Science Edition), 2011, 36(6): 57-61.
- [12] FU Feng, CHEN Xiao-jie, LIU Liang-huan, et al. Social dilemmas in an online social network: the structure and evolution of cooperation[J]. Physics Letters A, 2007, 371(1-2): 58-64.
- [13] 刘晶,李永先,薛伟莲. 国内外在线社会网络研究综述[J]. 农业网络信息, 2011(3): 66-68, 86.

- LIU Jing, LI Yong-xian, XUE Wei-lian. Review of online social network research at home and abroad[J]. Agriculture Network Information, 2011(3): 66-68, 86.
- [14] 胡海波. 在线社会网络的结构、演化及动力学研究[D]. 上海: 上海交通大学, 2010.
HU Hai-bo. Research on the structure, evolution and dynamics of online social networks[D]. Shanghai: Shanghai Jiao Tong University, 2010.
- [15] HU Hai-bo, WANG Xiao-fan. Evolution of a large on line social network[J]. Physical Letters A, 2009, 373(12/13): 1105-1110.
- [16] 何大韧, 刘宗华, 汪秉宏. 复杂系统与复杂网络[M]. 北京: 高等教育出版社, 2009.
HE Da-ren, LIU Zong-hua, WANG Bing-hong. Complex systems and complex networks[M]. Beijing: Higher Education Press, 2009.
- [17] 汪小帆, 李翔, 陈关荣. 复杂网络理论及其应用[M]. 北京: 清华大学出版社, 2006.
WANG Xiao-fan, LI Xiang, CHEN Guan-rong. Complex network theory and its applications[M]. Beijing: Tsinghua University Press, 2006.
- [18] 周涛, 柏文洁, 汪秉宏, 等. 复杂网络研究概述[J]. 物理, 2005, 34(1): 31-36.
ZHOU Tao, BAI Wen-jie, WANG Bing-hong, et al. A brief review of complex networks[J]. Physics, 2005, 34(1): 31-36.
- [19] COSTA L D F, RODRIGUES F A, TRAVIESO G, et al. Characterization of complex networks: a survey of measurements[J]. Advances in Physics, 2007, 56(1): 167-242.
- [20] BARTHÉLEMY M. Betweenness centrality in large complex networks[J]. The European Physical Journal B - Condensed Matter and Complex Systems, 2004, 38(2): 163-168.
- [21] NEWMAN M E J. Assortative mixing in networks[J]. Physical Review Letters, 2002, 89(20): 208071.
- [22] NEWMAN M E J. Mixing patterns in networks[J]. Physical Review E, 2003, 67(8): 026126.2010.

编辑 蒋晓