

文章编号:1009-3087(2012)04-0000-05

社交网络形成和演化的特征模型研究

熊 熙¹, 曹 伟², 周 欣², 胡 勇^{1*}

(1. 四川大学 信息安全研究所, 四川 成都 610065; 2. 中国信息安全测评中心, 北京 100085)

摘 要:社交网络和微博的网络特征除具有传统无标度网络的典型特征外,还存在其他不容忽视的特征,现有模型均不能对此进行准确描述。在比较分析网络数据的基础上,提出一种混合模型对类社交网络的形成和演化特征进行描述,并建立了基于该模型的平均场方程,方程的解显示出该模型的度分布服从偏移和拉伸后的幂律。仿真结果表明,作者提出的模型能描述出社交网络的综合特征。对包括本文模型在内的几种模型进行了比较,并对各种模型间出现特征差别的原因进行了分析。

关键词:复杂网络;社交网络;群落;幂律分布

中图分类号:O550;TP393.08

文献标志码:A

Research on the Feature Model of the Formation and Evolution of Social Networks

XIONG Xi¹, CAO Wei², ZHOU Xin², HU Yong^{1*}

(1. Inst. of Info. Security, Sichuan Univ., Chengdu 610065, China;

2. China Info. Technol. Security Evaluation Center, Beijing 100085, China)

Abstract: Besides the typical characteristics of scale-free network, the indispensable features of social networking services and microblog can not be described by existing models completely and accurately. Based on comparison and analysis of network data, a hybrid model was put forward to describe the features of formation and evolution of generalized social network. Average-field equations were constructed, and its solution showed that the degree distribution follows the translated and stretched power-law. The simulation results showed that the model can describe the overall characteristics of social network. Finally, several models, including this model, were compared, and reasons were given for the differences of features among the models.

Key words: complex networks; social networking services; cluster; power-law distribution

复杂网络学,是一门研究各种具有网络结构或形态的系统的复杂性和其网络行为之间的关系的学科^[1]。1998年,2位年轻的美国物理学家 Watts 和 Strogatz 在《Nature》上发表论文^[2],提出小世界网络模型(简称 WS 模型)。该模型的典型特征为平均距离小(接近随机网络),而集群系数高(与规则网络相类似)^[3]。一年后,另外2位美国物理学家 Barabasi 和 Albert 在《Science》上发表论文^[4],提出无标度网络(Scale-free network)模型,包括2个要素:增长性和择优连接性^[5],并举例说明许多实际网络(例如互联网页面连接)都精确或近似地遵循

幂函数的度分布^[6]。用平均场解析的方法可以得到 BA 模型服从幂指数为-3的幂律分布。

WS 小世界模型和 BA 无标度模型分别从不同角度对客观世界存在的各种具有逻辑结构的网络系统的形成和演化特征进行了概括,在该研究领域产生了很大影响,并有不少后续研究提出了改进模型或应用模式^[7-12]。众所周知,BA 模型展示的无标度的幂律度分布与随机网模型^[3]展示的指数度分布是2个极端情况,自然界中大多数具有逻辑结构的系统的网络度分布却介于这两者之间。对此,文献[13]建议的一个“优选-随机模型”展示的度分布为幂律与指数函数的某种混合。由于全局信息的缺失,BA 模型建议的“全局择优连接机制”通常只能在某些“局部世界”中展现出来。这里的局部世界即所谓的群落(Cluster),也被称为簇。复杂网络学中将群落定义为:其内部连接密集、对外连接稀疏

收稿日期:2011-10-19

基金项目:国家自然科学基金资助项目(60873246);中国信息安全测评中心资助项目

作者简介:熊 熙(1983—),男,博士生.研究方向:信息安全.

* 通信联系人

的一些节点集合^[14]。2003年发表的文献[15]提出了影响很大的“局部世界模型”,指出复杂网络的度分布介于指数分布与幂律分布之间,为进一步研究打下了基础。

电子布告栏(Bulletin board system, BBS),是一种虚拟论坛。文献[16]提出了一种虚拟社区网络模型构造算法。该算法涉及增长、优先连接和加边3个过程。其核心思想与BA模型相似,度分布依然符合幂律,但是其特征参数已经偏离BA模型。

在Web2.0应用模式下^[17],网络人际关系可通过新的方式重新构建。针对博客网络,文献[18]提出一种边与节点同时增长的网络模型。该模型包括增长、随机连接及近邻互连的网络构造规则,具有可调的度分布指数及集群系数。同样,该模型具有与BA模型相似却略有差异的特征参数。在没有优先连接过程时,模型能得到幂律的度分布。这点体现了该模型与传统BA模型在构造上有本质的不同。

社交网络服务(Social networking services, SNS)和微博(如Twitter)是最新出现的2种Web2.0应用模式,都属于在线社交网络。用户的线上线下的身份比较一致。社会交际也因此产生重大变化。通过观察和分析,发现社交网络具有增长、择优连接、随机连接和聚合等多种特征,用上述任何一个模型都不能完整而准确描述。为此,作者提出一种混合模型。该模型的度分布、集群系数等各种特征参数能准确地反映社交网络形成和演化的实际情况。

鉴于SNS与微博的功能趋同,都属于在线社交网络,因此所讨论的社交网络包括SNS与微博的共同特性。

1 社交网络

社交网络源于哈佛大学的心理学教授Stanley Milgram在1967年创立的六度分割理论,即“最多通过6个人你就能认识任何一个陌生人。”按照六度分割理论,每个个体的社交圈都不断放大,最后成为一个大型网络。最初的网络条件无法满足SNS应用的需求,社交网络并没有成为主流网站。直到2004年以社交游戏和实名制交友见长的Facebook的出现,以及2006年微博网站Twitter的出现,社交网络才在网络这一虚拟社会中蓬勃兴起。从2005年开始,中国的人人网、新浪微博等模仿国外者也取得了巨大的成功。在SNS网站影响下,近3年来全球网络社区日趋SNS化。这种以互联网应用为平台从而实现人际关系网络化的社会交际网络,兼有

现实社区和虚拟社区的特点,使人际关系的发生与演变发生着深刻的变革,被认为是“正在爆发的互联网的革命”^[19]。

社交网络的多种传播方式在每一个群体内部建立起一条快速、便捷、形式多样的信息扩散通道。与目前被广泛看好的网络公共空间(BBS、博客等)相比有着更多的优势:在消息扩散方面,它更快速;在公共空间的参与性方面,它更便捷;在消息的可信度和来源方面,它更真实可靠;在监管方面,它采用纯粹的人际传播方式,最大限度地避开了第三方的监管^[20]。传播主体的能动性和去中心化使其更接近真实人际关系网络。

作者提取了国内某著名社交网站的数据,共计47 220条消息(包括原创、转发),1 186个用户信息。通过统计与分析,这一典型社交网络具有以下特性:1)度分布服从幂律。幂指数 $\gamma \in [-3, -2]$,不同于BA模型的恒定幂指数 $\gamma = -3$ 。2)用户按属性进行聚合,同属性的用户联系密切,不同属性的用户间联系较少。一个用户同时拥有多个属性,使得属性间彼此交叉。3)社交网络的传播主体是现实中的人在网络虚拟世界的映射,由于受到现实社会和虚拟世界的环境约束,必然存在相当数量的随机连接。4)以月为单位观察平均度随时间的变化,可以发现平均节点度随节点数目的增加而增长,并且平均度由快速增长变为缓慢增长,最后达到一个稳定值。这与BA网络平均度的时不变性质明显不同。

结合以上对社交网站数据的分析,可以总结出社交网络在形成和演化过程中的主要特性:1)增长性,即新用户不断加入,用户总数不断增加。2)择优连接与随机连接并存,即新用户加入网络后,首先与同自己关系密切的人建立连接,同时可随机连接一些用户,与其建立好友关系。3)聚合性,即具有相同属性的用户通常会联系更为密切,并形成一个小集群。整个网络就是由许多这样的小集群组成,各个小集群之间又相互关联。4)关系传递,即用户与好友的好友建立直接的好友关系。用户A和B、C分别有好友关系,则B、C直接相连。

由上可见:一方面,社交网络具有BA无标度网络模型的幂律分布特点,是BA模型的扩展和更新;另一方面,社交网络又具有与BA模型所不同的一些个性特征,比如聚类和相当概率的随机连接等。因此,需要重新构建模型,以正确地描述社交网络的综合特征。

2 基于社交网络的多特征混合模型

为全面准确地描述社交网络的特征,提出一种能描述其综合特征的混合模型,模型中节点表示用户,2个节点之间是否存在连接(边)则表示2个用户间是否是好友关系,此处不存在用户自环或重复连接的情况。模型所表示的演化过程为:

1) 网络初始形成时, N_0 个节点被随机地分成 m_0 个组, 每个小组里有 $n_i (i = 1, 2, \dots, m)$ 个节点,

所以 $N_0 = \sum_{i=1}^m n_i$ 。

2) 在每个演化时步, 每个群落新加一个节点。每个新加的节点发出条边到同群落中的其它节点, 并且每条新边分别以概率 p 和 $1-p$ 进行择优连接和随机连接。即其中 pm 条边以概率 $k_i / \sum_j k_j$ 择优连接到节点 i , 其余 $(1-p)m$ 条边则随机地连接。

3) 边自发增长。在每个时步, 在每个群落之中优选一个节点 i , i 以概率 r 发出1条边到其他节点。随机选定 i 的邻居 j , i 随机连向 j 的一个邻居。选邻居时, 选择同群落节点的概率是 q , 选择群落外节点的概率是 $1-q (q > 0.5)$ 。

步骤2) 中节点 i 的度 k_i 的演化方程可以表示为:

$$\frac{\partial k_i}{\partial t} = mp \frac{k_i}{\sum_j k_j} + m(1-p) \frac{1}{n_i} \quad (1)$$

式中, 等号右边第1项表示择优连接, 第2项来自随机连接。

为得到步骤3) 中 k_i 的演化方程, 假设除了节点 i 以外的所有节点度都是 k_a 。其中, k_{in} 个群落内的节点相连, 与 k_{out} 个群落外节点相连, 即 $k_a = k_{in} + k_{out}$ 。另外, 节点 i 共有 $k_i k_a$ 个间接邻点(即节点 i 的邻居的邻居)。那么, k_i 的演化方程为:

$$\frac{\partial k_i}{\partial t} = r \frac{k_i}{\sum_j k_j} + (k_i k_a) \left(r \frac{k_n}{\sum_j k_j} \frac{1}{k_n} \frac{1}{k_a} \right) \cdot (2P_{in}P_{out} + P_{out}^2 + P_{in}^2) \quad (2)$$

式中, 第2项来自其他群落。其中,

$$P_{in} = \frac{qk_{in}}{(1-q)k_{out} + qk_{in}}, P_{out} = \frac{(1-q)k_{out}}{(1-q)k_{out} + qk_{in}}。$$

那么, 式(2)就变为:

$$\frac{\partial k_i}{\partial t} = \frac{2rk_i}{\sum_j k_j} \quad (3)$$

在一个完整的演化时步(包括步骤2)和3)) 中, k_i

的演化方程为式(1)、(3) 两个方程之和。于是, 该模型的平均场方程即:

$$\frac{\partial k_i}{\partial t} = (mp + 2r) \frac{k_i}{\sum_j k_j} + \frac{m(1-p)}{n_i} \quad (4)$$

当 t 比较大时, $\sum_j k_j \approx 2mt + 2rt = 2(m+r)t, n_i = t$ 。节点 i 在 t_i 时刻加入, $k_i(t_i) = m$ 。将初始条件代入, 可以得到式(4) 的解:

$$k_i(t) = (m + \frac{b}{a}) \cdot (\frac{t}{t_i})^a - \frac{b}{a} \quad (5)$$

其中, $a = \frac{mp + 2r}{2(m+r)}, b = m(1-p)$; 一个节点的度 $k_i(t) < k$ 的概率可以表示为:

$$P[k_i(t) < k] = P[t_i > t \cdot (\frac{ak+b}{am+b})^{-\frac{1}{a}}] \approx 1 - (\frac{ak+b}{am+b})^{-\frac{1}{a}} \quad (6)$$

则可得

$$P(k) = \frac{\partial P[k_i(t) < k]}{\partial k} = \frac{1}{a} (\frac{ak+b}{am+b})^{-\frac{1}{a}-1} (\frac{a}{am+b}) = (am+b)^{\frac{1}{a}} (ak+b)^{-\frac{1}{a}-1} \quad (7)$$

显然, $P(k)$ 服从幂律分布, 该幂函数是典型幂函数 $P(k) = k^{-\frac{1}{a}-1}$ 经过平移和拉伸后得到的。该分布与 q 无关。

从式(7) 可知, 度分布服从幂律, 并且幂指数为:

$$\gamma = -(mp + 4r + 2m)/(mp + 2r) \quad (8)$$

当 $m = 1, p = 1, r = 0$ 时, γ 为 -3 , 即该模型变为 BA 模型。

3 仿真与分析

统计数据中已知的参数有 $m_0 = 131, N_0 = 1000$ 。在此条件下对数据进行幂函数曲线拟合, 可以得到其他参数: $m = 2, p = 0.8, q = 0.66, r = 0.8$ 。利用这些参数对该模型进行数值模拟, 可以观察网络性质。

仿真时, 循环进行步骤2) 和3), 直到每个群落内的节点数为 N_i 。

3.1 用户交互结构

为表示用户之间的相互关系, 用 UCINET^[21] 画出网络达到60个用户(4个群落)时的关系图, 如图1所示, 各个群落中的用户分别用不同的图形来表

示。可以明显地看出,群落内部用户间的联系比群落间更为密切。用户更倾向于与自己具有相同属性的用户成为好友。

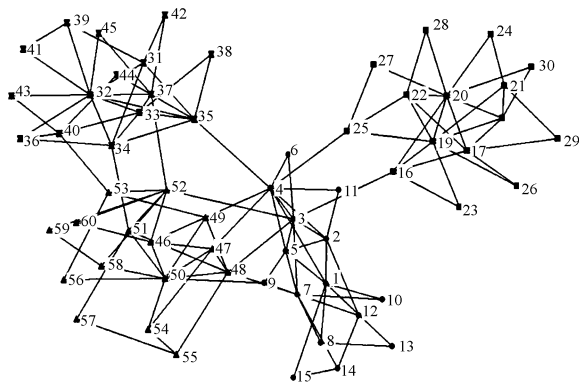


图1 用户交互结构

Fig. 1 Structure of user interaction

3.2 度分布

从图2可以看出,模拟网络的度分布服从幂律,幂指数 $\gamma = -2.723$ 。同时,可通过式(8)计算出 $\gamma = -2.75$ 。表明模型与实际数据相吻合。度分布的长尾并不明显,有一定程度的截断。

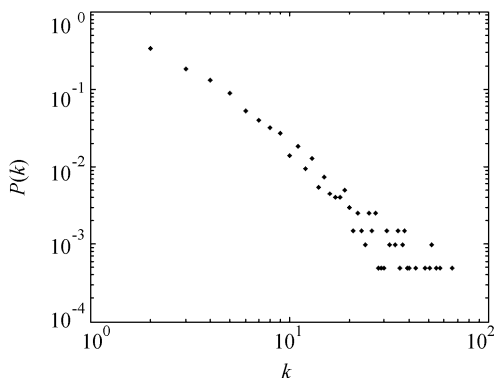


图2 多特征模型的度分布

Fig. 2 Degree distribution of the hybrid model

m_0 、 n_i 、 p 、 q 、 r 变化对网络度分布的影响不明显。从这些数据得到的 γ 略大于 BA 网络模型的幂指数 $\gamma = -3$ 。同时, SNS 可使一个群体内部联系更紧密,就意味着可能会带来的另一个效果:使群体之间联系更紧密。因为在一些群体间是存在交叉点的,比如某个人属于一个单位群体,可能还属于另一个协会群体或者一个家庭群体。在“意见领袖”的作用下一一些群体间无疑能够保持相对稳定且紧密的联系。

图3为表示网络的平均节点度与演化时间之间的关系。很明显,曲线初期增长迅速,随着时间的推移,曲线的增长趋缓,最后在一个稳定值附近波动,这基本符合实际的统计结果。

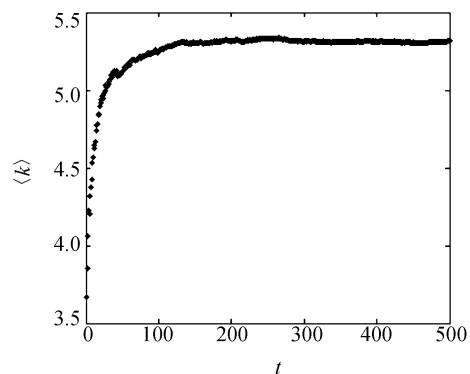


图3 多特征模型的平均度随时间的变化

Fig. 3 Change of average degree with time of the hybrid model

3.3 集群系数

集群系数衡量了网络节点聚集程度,反映了节点邻居之间的连通性。若节点 i 的度为 k_i ,则节点 i 的集群系数 c_i 为:

$$c_i = \frac{\sum_j a_{ij} \sum_l a_{jl} a_{il}}{k_i(k_i - 1)} \quad (9)$$

若节点 i 与 j 之间有连接, $a_{ij} = 1$, 否则 $a_{ij} = 0$ 。节点总数为 N 的网络平均集群系数 C 为:

$$C = \frac{1}{N} \sum_i c_i \quad (10)$$

当 $r > 0.5$ 时,本网络模型的连通性过大,度分布已偏离了幂律。图4表示节点总数为2000(4个群落), $r \leq 0.5$ 时模型的平均集群系数与 r 、 m 的关系。

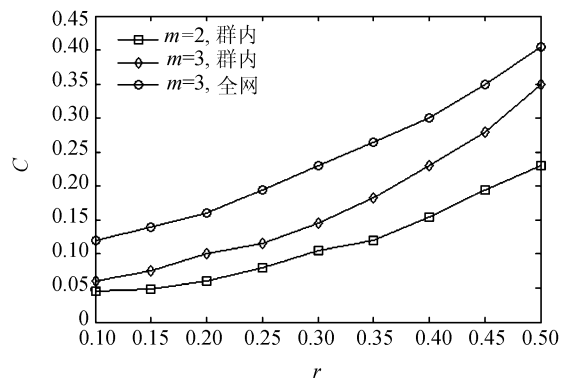


图4 集群系数随参数 m 和 r 的变化

Fig. 4 Change of cluster coefficient with m and r

与 BA 网络相比,该模型具有更大的集群系数。增大 p 或减小 n ,都能提高集群系数。同时,同一个群落内的集群系数比整个网络的集群系数大。 r 越小,边自发增长(步骤3)越缓慢。随着 r 的增大,全网的集群系数逐渐增加,且同一个群落内集群系数的增长速度大于全网集群系数的增长速度。这表明,随着边自发增长的概率 r 的增加,整个网络的结构愈加完美,且同一群落内的节点的联系比群间节点更

紧密。

3.4 几种网络模型比较

传统的 BA 网络模型与 3 种网络应用的模型 (BBS 网络、博客网络和社交网络) 之间既有共性, 又存在较大差别。这些网络模型的度分布都服从幂律, 表明网络的无标度特性是普遍存在的。4 种网络的度分布经幂函数拟合后的曲线如图 5 所示。

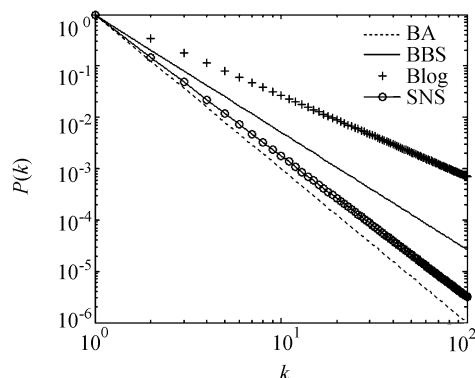


图5 社交网络与其他3种网络模型的度分布比较

Fig.5 Comparison of degree distribution of SNS model and other three network models

当 $m = 2$ 时, 本模型的幂指数 $\gamma \leq -2.5$, 图 5 中仅画出了一个典型值 $\gamma = -2.75$ 。从图 5 可以很明显地看出, BBS 网络、博客网络和社交网络的幂指数都偏离了传统 BA 模型的 -3 幂指数, 这种差别由网络模型本身的形成及演化过程决定, 比如本模型演化过程中的聚类特性。

4 结论与展望

网络社交关系的形成和演化是一个复杂的过程, 各种因素 (现实的和虚拟的) 相互影响, 呈现多种特征。社交网络将现实中的社交关系虚拟化, 因而比其它互联网应用更接近真实的人际交往。

提出的用于描述社交网络形成和演化特征的模型与传统的无标度网络模型在许多方面有所不同, 比如聚类和随机连接等个性特征。数值仿真表明该网络模型与实际社交网络数据的统计结果基本符合, 比如度分布服从幂律, 群内集群系数较大等。最后分析了多种网络特征之间存在差别的原因。

对于包括社交网络在内的复杂网络而言, 结构与传播的耦合动力学的研究正逐渐成为热点。复杂网络的研究重点也逐渐转向基于结构的传播过程。只有理解了在这种结构上的传播机理后, 才能够有根据地进行传播的控制问题研究。所提出的概念和模型将对类社交网络的结构及传播动力学研究产生直接而积极的影响。

参考文献:

- [1] 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: 147.
- [2] Watts D J, Strogatz S H. Collective dynamics of 'small-world' networks [J]. Nature, 1998, 393: 440.
- [3] Erdos P, Renyi A. On random graphs I [J]. Publ Math, 1959, 6: 290.
- [4] Barabasi A L, Albert R. Emergence of scaling in random networks [J]. Science, 1999, 286: 509.
- [5] Albert R, Barabasi A L. Statistical mechanics of complex networks [J]. Phys Rev Mod, 2002, 74: 47.
- [6] Costa L F, Rodrigues F A, Traverso G, et al. Characterization of complex networks: a survey of measurements [J]. Advances in Physics, 2007, 56 (1): 167 - 242.
- [7] Newman M E J, Watts D J. Scaling and percolation in the small-world network model [J]. Phys Rev E, 1999, 60 (6): 7332.
- [8] Mendes J F F, Dorogovtsev S N. Evolution of networks with aging of sites [J]. Phys Rev E, 2000, 62 (2): 1842.
- [9] Dorogovtsev S N, Mendes J F F. Effect of the accelerating growth of communications networks on their structure [J]. Phys Rev E, 2001, 63 (2): 25101.
- [10] Cheng X, Wang H L, Ouyang Q. Scale-free network model of node and connection diversity [J]. Phys Rev E, 2002, 65 (6): 66115.
- [11] Wu X Y, Liu Z H. How community structure influences epidemic spread in social networks [J]. Physica A, 2008, 387: 623 - 630.
- [12] Zhang Yanchao, Liu Yun, Zhang Haifeng, et al. The research of information dissemination model on online social network [J]. Acta Phys Sin, 2011, 60 (5): 50501. [张彦超, 刘云, 张海峰, 等. 基于在线社交网络的信息传播模型 [J]. 物理学报, 2011, 60 (5): 50501.]
- [13] Liu Z H, Lai Y C, Ye N, et al. Connectivity distribution and attack tolerance of general networks with both preferential and random attachments [J]. Phys Lett A, 2002, 303: 337.
- [14] Liu Z H, Hu B. Epidemic spreading in community networks [J]. Europhys Lett, 2005, 72: 315.
- [15] Li X, Chen G R. A local-world evolving network model [J]. Physica A, 2003, 328: 274.
- [16] Zhang Li, Liu Yun. Research on the evolution process of virtual community networks [J]. Acta Phys Sin, 2008, 57 (9): 5419. [张立, 刘云. 虚拟社区网络的演化过程研究 [J]. 物理学报, 2008, 57 (9): 5419.]
- [17] Shin D H. The effects of trust, security and privacy in social networking: a security-based approach to understand the pattern of adoption [J]. Interacting with Computers, 2010, 22 (5): 428.
- [18] Xiong Fei, Liu Yun, Si Xiameng, et al. Network model with synchronously increasing nodes and edges based on Web 2. 0 [J]. Acta Phys Sin, 2010, 59 (10): 6889. [熊菲, 刘云, 司夏萌, 等. 基于 Web 2. 0 的边与节点同时增长网络模型 [J]. 物理学报, 2010, 59 (10): 6889.]
- [19] 西门柳上, 马国良, 刘清华. 正在爆发的互联网革命 [M]. 北京: 机械工业出版社, 2009: 1.
- [20] Zhan Xun, Peng Tao. Effect of SNS in the course of action of interpersonal communication [J]. Press Circles, 2008, 6: 37. [詹侑, 彭涛. SNS 对人际传播中行动过程的影响 [J]. 新闻界, 2008, 6: 37.]
- [21] Borgatti S, Everett M, Freeman L. User's guide of UCINET 6 for Windows [EB/OL]. 2002 - 12 - 13. <http://www.analytictech.com/ucinet>.

(编辑 赵 婧)