

文章编号:1007-6735(2011)05-0452-05

基于基尼系数的网络结构洞测量

邓世果¹, 吴干华^{1,2}, 杨会杰¹

(1. 上海理工大学 管理学院, 上海 200093; 2. 华南师范大学 南海学院, 佛山 528225)

摘要: 用结构洞分析网络结构是评价网络结构引起的竞争优势的一种重要方法. 本文提出了一种新的结构洞定量描述方法, 即基于主成分分析(PCA)的结构洞的测量. 给出了类似于基尼系数的网络贡献度的概念, 在满足度分布不变的条件下利用贡献度分析对比无标度网络与对应的随机网络, 证实用贡献度评价网络有效性的优越性. 得出结论: 贡献度 S 越大, 说明网络中每个节点的重要性越平均, 结构洞程度越大; 反之, 贡献度 S 越小, 则说明网络结构信息主要集中在部分节点上, 结构洞程度越小.

关键词: 结构洞; 主成分分析; 网络贡献度; 复杂网络

中图分类号: N 941 **文献标志码:** A

Gini-coefficient-based measurement of structural holes

DENG Shi-guo¹, WU Gan-hua^{1,2}, YANG Hui-jie¹

(1. Business School, University of Shanghai for Science and Technology, Shanghai 200093, China;

2. Nanhai School, South China Normal University, Foshan 528225, China)

Abstract: Structural hole is an important concept to estimate pattern-induced competitive power for each node in a social network. Principal component analysis(PCA) was proposed to evaluate the structural holes in a global way. Gini-coefficient for social network was defined to measure quantitatively the structural holes. As examples, the Gini coefficients for some real-world networks and the extended Barabási-Albert scale-free networks were calculated. It is concluded that the bigger the contribution of S in the network, the more uniform the importance of nodes and the greater the degree of structural holes; Conversely, the smaller the contribution of S , showing the network structure information mainly concentrated in part of nodes, the smaller the degree of structural hole.

Key words: structural hole; principal component analysis; network contribution; complex network

收稿日期: 2011-10-08

基金项目: 国家自然科学基金资助项目(10975099, 10635040); 上海市重点学科建设资助项目(S30501)

作者简介: 邓世果(1985-), 女, 硕士研究生. 研究方向: 系统理论. E-mail: guobingpeng123@163.com
杨会杰(联系人), 男, 教授. 研究方向: 经济和生物复杂系统. E-mail: hjyang@usst.edu.cn

1 结构洞的概念及计算方法

随着社会关系网络研究的进展,网络结构对社会关系的影响日益受到关注.普遍存在的信息壁垒导致不完全竞争,选择往往集中在具有支配地位的玩家手里,其他的每个玩家一定程度上附属于具有支配地位的玩家.竞争场中处于某种特殊关系结构中的个体,可以透过信息过滤(information filtering)获得更多竞争优势与创新能力.关系结构对信息利益和控制利益的影响是社会竞争网络研究的核心问题^[1].

结构洞^[2]是关系结构分析方法中引起广泛关注的概念,该理论试图对个体在群体之中的关键位置进行深入的解释.结构洞指在社会关系网络中两个个体或两个群体之间不存在直接连接,且它们之间不存在间接冗余关系(redundancy),那么,两者之间的阻碍就是结构洞.如图1所示,3个菱形和三角形A均与三角形B直接相连,但3个菱形与三角形B之间没有边,要通过三角形A产生联系,这样它们之间就产生了间隙,形成结构洞,三角形A具有较好的竞争优势,更容易获得信息又没有造成冗余.三角形B的地位同于三角形A,但与三角形B直接相连的3个菱形彼此之间也有直接联系,这就造成了网络冗余,B在网络结构中的优势和控制力会明显低于A.拥有较多结构洞的个体,越趋向于财富的快速积累、地位的快速提升.竞争的社会结构不仅要有个体的资源优势比较,还应包括对拥有结构洞数量的关系优势(relations advance)比较.个体的最佳方案是尽量占据多的结构洞,然后跨越结构洞,连接原来没有关系的群体,使个体成为信息流的媒介^[3-4].

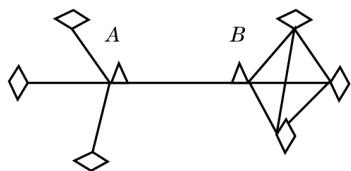


图1 结构洞示意图

Fig.1 Illustration of structural holes

网络约束系数(network constraint index)是Burt在1992年提出的计算结构洞系数的方法^[1].设节点*i*花费在*j*上的时间或者精力占其总时间或精力的比例为 P_{ij} .节点*i*的网络约束系数

$$C_i = \sum_j (P_{ij} + \sum_q P_{iq}P_{jq})^2 \quad (1)$$

该约束系数描述的是网络中某一节点与其它节点直

接或间接联系的紧密程度.网络约束系数越高,网络的闭合性越高,结构洞程度越小;反之,结构洞程度越大.此方法成功地应用于中等类型公司中信息扩散路径和行动者的心理研究^[5-6].但是,这一概念只是考察了节点最近邻和次近邻的影响,没有考虑到更加宏观的结构信息,并且没有考虑到邻居间的联系对控制力的影响.

介数中心性(betweenness centrality)广泛地用于定量描述结构洞.关系网络中某节点的介数中心度是所有节点对之间的最短路径中途经该节点的最短路径数目.其中包含两个假设:每条路径的权重系数相等,且信息从起始节点到目标节点传达的过程中总是走最短路径.因此,可测量单个节点对网络资源控制的程度.这一方法的缺点是假设节点之间通过最短路径传递信息,这并不是普遍的行为策略^[7-8],且在网络中只有聚类系数(cluster)小的节点,介数(betweenness)大时才处于结构洞位置.

网络有效结构尺度(network effective size)^[9]定量地考虑了结构洞的部分内容,简单地认为,如果间接看来节点之间不存在冗余关系,那么两者之间的空隙就是结构洞,即节点度的规模减去冗余连接的数目^[10].节点*i*的有效结构尺度

$$E_i = \sum_j (1 - \sum_q P_{iq}P_{jq}) \quad (2)$$

式中,*j*为所有与*i*相连的节点;*q*为不同于*i*和*j*的第三点; $P_{iq}P_{jq}$ 为节点*i*与节点*j*的冗余边数.

总之,这些结构洞的定量描述提供了关系网络上节点在结构上的特异性,为评价这些节点的结构上竞争性提供了重要的参考指标,但是,仍然存在一些问题需要改进,特别是没有提供一个有效的量对关系网络整体进行有效的评价^[11].实际上,结构洞的核心思想是节点之间传递信息通路的非冗余性.冗余性的存在会削弱节点的竞争力,增加维护网络结构的耗费.因此,本文将从这一思想出发,采用主成分分析方法,将基尼系数应用于结构洞特性的描述,给出网络结构的一个整体性的评价.

2 方法与材料

2.1 主成分分析测量结构洞

主成分分析(PCA)是一种从对象中提取主要信息,而忽略相对次要信息的多元统计分析方法^[12],可有效消除数据结构中的冗余信息,从而对数据进行有效的压缩.本文基于PCA的信息压缩思想,提

出了一种新的测量结构洞程度的方法.

一个社会关系网络可以采用邻接矩阵描述. 邻接矩阵的矩阵元 $A(i, j)$ 表示节点与节点之间的连接状态. 当两者直接相连时该元素为 1; 否则为 0. 现在将该矩阵的节点看作一个复杂系统的 N 个成分. 而将邻接矩阵的每一列看作对该列对应的成分的一个测量, 也就是该成分的属性. 当网络上两个节点处于对称位置时, 属性将几乎完全相同, 相关系数接近 1. 这种对称结构在网络上对应的是大量的冗余结构, 结构洞会显著减少. 一个高效率的关系网络应该是尽可能少地存在冗余结构, 也就是可压缩性尽可能达到最小, 从而使每个节点和边的存在都有其独特的性质. 当然, 冗余会带来结构的鲁棒性, 也是必不可少的, 实际的最优关系网络是达到两者的平衡.

各成分的状态表示为

$$\begin{aligned} X_i &= (x_{1,i}, x_{2,i}, \dots, x_{3,i}) = \\ &[A(1,i), A(2,i), \dots, A(N,i)], \\ &i = 1, 2, \dots, N \end{aligned} \quad (3)$$

PCA 算法过程如下:

a. 计算观测值的协方差 C

$$c_{ij} = \frac{\sum_{k=1}^n (x_i^k - \bar{x}_i)(x_j^k - \bar{x}_j)}{n-1}, \quad i, j = 1, 2, \dots, N \quad (4)$$

b. 计算 C 的 N 个特征值, 并将特征值按大小排序, 记为 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N$.

c. 以前 n 个 λ 的和, 即以可解释百分率 P 为横坐标, 以 n/N 为纵坐标做曲线. 仿照基尼系数的定义, 可定义关系网络的基尼系数, 本文中将其称为关系网络的节点贡献度 S , 如图 2 所示, 即所得曲线与 X 轴组成的面积 (图中黑色部分面积) 与直线 $X=Y$ 与 X 轴组成的面积 (图中紫色线条与 X 轴组成的面积) 的比值. 该系数越大, 说明要达到同样的可解释百分

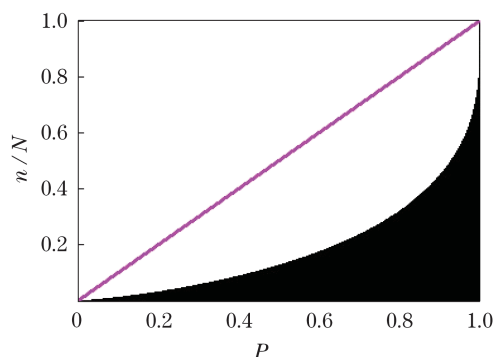


图 2 贡献度 S

Fig.2 Contribution degree S

率需要的节点数目越多, 即结构冗余程度越低, 结构洞越多.

2.2 数据

本文选用了两个真实的社会关系网络, 圣塔菲 (Santa Fe) 研究所科研合作网络和足球运动员网络. 在保持节点度不变的条件下, 随机交叉互换边, 得到这两个网络对应的随机交叉网络.

本文也考虑了扩展的无标度网络模型 (BA 网络增长模型). BA 网络增长模型是指在网络增长过程中, 新节点连接到旧节点的概率 $P(k_i)$ 正比于旧节点的度 k_i , 即 $\frac{P(k_i) \propto k_i}{\sum_j k_j}$, 也就是说新节点与旧节点相连有连接偏好. 在 BA 网络增长模型的基础上可做模型的扩展, 即 $P(k_i) \propto \frac{k_i \alpha}{\sum_j k_j}$. 模型中有两个可调参数 e 和 α , e 是每增加一个节点所增加的边数; 当 $\alpha = 1$ 时, 就是 BA 网络增长模型. 随着 α 的减小, 新节点与旧节点相连接的偏好程度变得越来越小, 各个节点对网络的贡献越来越平均. 当 α 减小到 0 时, 此模型就变为随机网络增长模型. 在 $e = 1, 2, 3, 4$ 的情况下, 分别做出了 $\alpha = 0, 0.2, 0.4, 0.6, 0.8, 1.0$ 时的增长网络, 并分别对应每个网络做出了在节点度分布不变的条件下随机交叉互换边的随机交叉网络.

3 结 果

3.1 扩展的 BA 增长模型

运用主成分分析方法计算了 28 个网络的节点贡献度 S , 结果发现, 在同一 e 的条件下, 扩展的 BA 增长模型生成的网络的节点贡献度 S_{BA} 均随着 α 的增大而减小, 如图 3 所示, 即 α 越大, 网络的结构洞程度越小. 这是因为新增节点与旧节点连接的偏好程度增大, 即新增节点偏向于与度大的节点相连, 这样就会导致网络中大部分节点都与部分中心节点相连, 因此, 网络的信息主要集中在这些中心节点上, 使其它节点的重要性降低, 网络的可压缩性增强, 造成网络冗余.

现分别对扩展的 BA 增长网络的节点贡献度 S_{BA} 和扩展的 BA 网络对应的随机交叉网络的节点贡献度 S_{BA^*} 进行比较, 结果如图 3 所示. 在 $e = 1, 2, 3, 4$ 时, BA 增长网络的节点贡献度 S_{BA} 均大于 BA 网络对应的随机交叉网络的节点贡献度 S_{BA^*} , 即 BA 网络对应的随机交叉网络的结构洞程度较低, 因为

扩展 BA 网络对应的随机交叉网络是在度不变的条件下随机交叉互换边得到的网络,这一网络在原有网络的基础上破坏了原有网络的网络结构,降低了结构洞程度,造成网络冗余.因此,扩展的 BA 网络的网络结构优于扩展的 BA 网络对应的随机交叉网络的网络结构.这说明,虽然与随机网络相比 BA 网络的贡献度 S_{BA} 较小,即结构洞程度较小,但是,BA 网络也存在一定的网络结构优势,这一网络结构要优于随机交叉网络的网络结构.

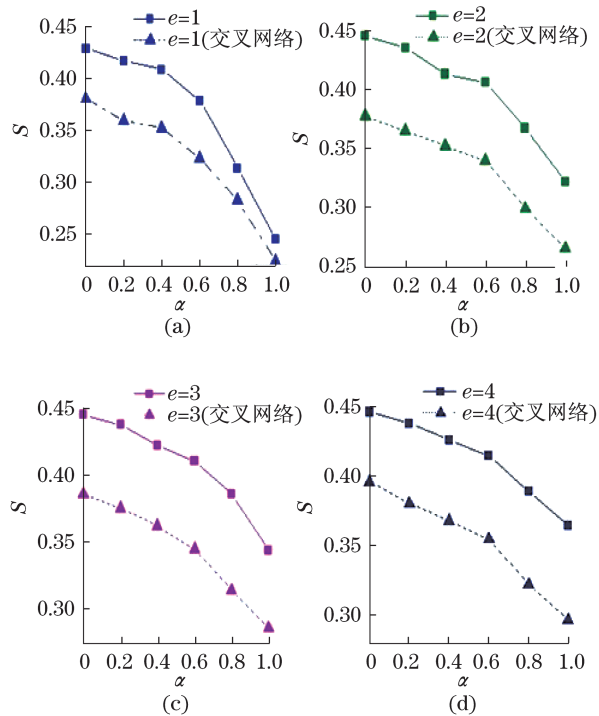


图3 e 相同时 S 与 α 的关系图

Fig.3 Relation of S versus α with same e

现比较不同 e 之间的扩展 BA 网络节点贡献度 S ,如图 4 所示.在图 4(a)中发现,随着 e 的增加, S 增大,且 S 增大的速率越来越缓慢,当 e 从 3 增大到 4 时, S 的变化已经很小,这说明 e 越大,BA 网络的网络结构性越明显,结构洞程度越高.但是, S 又不会无限制地增长,它会慢慢趋向一个稳定值.而 BA 增长网络对应的度不变随机交叉网络的 S 的变化趋势没有 BA 增长网络的明显.

3.2 真实社会网络

基于圣塔非科研合作网络和足球运动员网络的成员在俱乐部内部以及外部的社会关系,比较真实网络与其对应的随机交叉网络的贡献度,如表 1 所示.在表 1 中可以看出,真实网络的节点贡献度 S_t 大于随机交叉网络的节点贡献度 S_{te} ,这表明真实的社会网络结构优于随机交叉网络的网络结构.

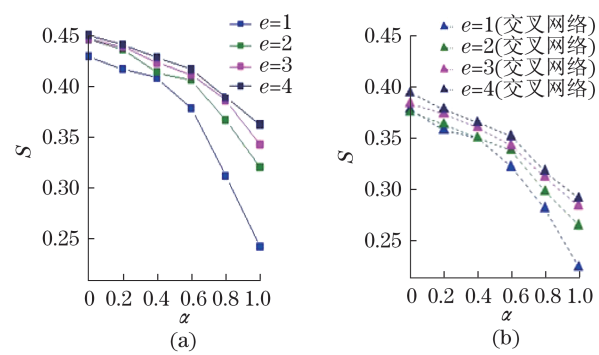


图4 e 不同时 S 与 α 的关系图

Fig.4 Relation of S versus α with different e

表1 真实网络与对应随机交叉网络的 S 比较

Tab.1 Comparison of S for real-world networks and the corresponding shuffled networks

S	圣塔非科研合作网络	足球运动员网络
S_t	0.289 9	0.221 3
S_{te}	0.229 8	0.216 5

4 结 论

提出了一种新的结构洞测量方法,即基于主成分分析(PCA)的结构洞的测量方法,并用此方法得出如上结果.通过扩展的 BA 网络在同一 e 下分别选取不同的 α 值,得出网络节点贡献度 S 随 α 的增大而减小,即 α 越大,网络结构洞程度越低,在扩展的 BA 模型对应的随机交叉网络中也得出相同的结论;通过对扩展的 BA 增长网络的节点贡献度 S_{BA} 和扩展的 BA 网络对应的随机交叉网络的节点贡献度 S_{BAr} 进行比较,得出 $S_{BAr} < S_{BA}$,即扩展的 BA 网络对应的随机交叉网络的结构洞程度低于扩展的 BA 增长网络;通过比较不同 e 之间的扩展 BA 网络节点贡献度 S ,得出 e 越大,网络的结构洞程度越高,但结构洞程度又不会无限制地增高,随着 e 的增长会趋向于稳定状态,且扩展的 BA 网络对应的随机交叉网络也有相同结论,但扩展的 BA 网络对应的随机交叉网络结构洞程度变化没有扩展的 BA 网络明显.

通过比较真实网络与其对应的随机交叉网络,得出真实的网络的贡献度高于其对应的随机交叉网络,真实的网络结构优于其对应的随机交叉网络.结合各研究领域的应用可以说明,从节点关系和网络结构的角度可以界定节点所处的位置及拥有的控制知识的力量.结构洞在网络结构研究中起导向性作

用,该理论和方法定会对商业网络分析、情报分析及计量领域等诸多社会关系网络研究起到强有力的推动作用.通过对各种结构洞算法的理解和掌握,可以促进对更大范围的网络知识的挖掘,并能够更好地理解相关网络的关系和结构,有利于对网络结构展开更深层次的研究.

参考文献:

- [1] BURT R S. The Social Structure of Competition[M]. Cambridge: Harvard University Press, 1992.
 - [2] BURT R S. The social capital of opinion leaders[J]. The Analysis of the American Academy of Political and Social Science, 1999, 556(1): 37 - 54.
 - [3] BURT R S. The contingent value of social capital[J]. Administrative Science Quarterly, 1997, 42 (2): 339 - 365.
 - [4] BURT R S. Structural holes and good ideas[J]. American Journal of Sociology, 2004, 11(2): 349 - 399.
 - [5] BURT R S. The gender of social capital[J]. Rationality and Society, 1998, 556(1): 37 - 54.
 - [6] GIUSEPPE S. Network memory. The influence of past and current networks on performance[J]. Social Network, 2004, 47(6): 893 - 906.
 - [7] FREEMAN L C. A set of measures of centrality based on betweenness[J]. Sociometry, 1977(1): 35 - 41.
 - [8] FREEMAN L C. Centrality in valued graphs: A measure of betweenness based on network flow[J]. Social Network, 1991(13): 141 - 154.
 - [9] NEWMAN M E J. A measure of betweenness centrality based on random walks[J]. Social Networks, 2005, 27(2): 39 - 54.
 - [10] MCEVILY B, ZAHEER A. Bridging ties: A source of firm heterogeneity in competitive abilities[J]. Strategic Management Journal, 1999, 20(12): 1133 - 1156.
 - [11] 胡蓉, 邓小昭. 基于结构洞理论的个人人际网络分析系统研究[J]. 情报学报, 2005, 24(4): 485 - 489.
 - [12] 郭崇慧. 基于 PCA 的复杂网络社区结构分析方法[J]. 运筹与管理, 2008, 17(6): 144 - 149.
-
- [27] CLAUSET A, SHALIZI C R, NEWMAN M E J. Power-law distributions in empirical data[J]. SIAM Review, 2009, 51(4): 661 - 703.
 - [28] 周炜星. 上证指数高频数据的多重分形错觉[J]. 管理科学学报, 2010, 13(3): 81 - 86.
 - [29] MANDELBROT B B. Long-run interdependence in price records and other economic time series[J]. Econometrica, 1970, 38(4): 122 - 123.
 - [30] HU K, IVANOV P C, CHEN Z, et al. Effect of trends on detrended fluctuation analysis[J]. Physical Review E, 2001, 64(1): 011114.
 - [31] REN F, GUO L, ZHOU W X. Statistical properties of volatility return intervals of Chinese stocks[J]. Physica A, 2009, 388(6): 881 - 890.
 - [31] ANDERSEN T G, BOLLERSLEV T, DIEBOLD F X, et al. The distribution of realized stock return volatility[J]. Journal of Financial Economics, 2001, 61(1): 43 - 76.
 - [33] REN F, GU G F, ZHOU W X. Scaling and memory in the return intervals of realized volatility[J]. Physica A, 2009, 388(22): 4787 - 4796.
 - [34] PODOBNIK B, HORVATIC D, PETERSEN A M, et al. Cross-correlations between volume change and price change[J]. Proceeding of the National Academy of Sciences of the United States of America, 2009, 106(52): 22079 - 22084.
 - [35] GU G F, ZHOU W X. Emergence of long memory in stock volatilities from a modified Mike-Farmer model[J]. Europhysics Letters, 2009, 86(4): 48002 - 48007.

(上接第 443 页)