

基于社交网络结构的节点影响力度量方法

李泽鹏^{1,2}, 左 杨^{1,2}, 王宏宇^{1,2}

(1. 北京大学高可信软件技术教育部重点实验室, 北京 100871; 2. 北京大学信息科学技术学院, 北京 100871)

摘 要: 度量社交网络节点影响力是社交网络结构分析的关键问题之一. 目前研究社交网络节点影响力的方法主要有两大类: 中心度方法和节点删除方法. 前者主要通过度或最短路径等因素来判断节点的影响力, 不考虑网络的连通性; 后者通过节点删除后对网络结构的破坏程度来判断, 计算复杂性很高, 不适用于较大规模的社交网络. 通过结合社交网络的局部连通度及节点间的最短路径, 提出了连通中心度来度量社交网络中节点的影响力, 并给出了连通中心度的计算方法和一些特殊网络中节点的连通中心度的值. 最后, 通过实验说明该指标能很好地度量社交网络中节点的影响力.

关键词: 社交网络; 节点影响力; 中心度方法; 连通中心度; 最短路

中图分类号: TP399 **文献标识码:** A **文章编号:** 0372-2112 (2016)12-2967-08

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2016.12.022

An Influence Measure of Nodes Based on Structures of Social Networks

LI Ze-peng^{1,2}, ZUO Yang^{1,2}, WANG Hong-yu^{1,2}

(Key Laboratory of High Confidence Software Technologies (Peking University), Ministry of Education, Beijing 100871, China;

2. School of Electronics Engineering and Computer Science, Peking University, Beijing 100871, China)

Abstract: Identifying the influence of nodes is one of the major research topics in the structural analysis of social networks. The current measures of researching the node influence can be divided into two categories: centrality measure and node removal measure. The former mainly identifies the influence of the node by degree or shortest path, without considering the connectivity of social networks; while the latter by the damage of the structure of a social network when some nodes are removed. The node removal measure is incapable to be applied in large-scale since the computational complexity. We propose a new parameter, connectedness centrality, to identify the influence of nodes in networks by combining the local connectivity and shortest paths. We give a method to compute the connectedness centrality of the node and obtain the precise values in some specific networks. Finally, an experiment using a real-word network shows that our method can well identify the influence of the node in social networks.

Key words: social networks; node influence; centrality measure; connectedness centrality; shortest path

1 引言

社交网络的拓扑结构是网络各项功能实现的基础, 对社交网络的拓扑结构、功能及其之间的关系目前已有大量研究^[1-3]. 分析社交网络的拓扑结构, 一个主要的方法是挖掘社交网络中有重要影响力的节点, 这些节点对社交网络的信息传播、网络结构的演化起着关键的作用^[4,5]. 因此, 如何度量节点的影响力, 如何在社交网络中发现这些重要节点是研究社交网络结构及

功能的关键.

目前刻画社交网络节点影响力的方法主要有两大类: 一类是中心度方法, 如度中心度^[6]、紧密中心度^[7]、介数中心度^[8]、特征向量中心度^[9,10]等; 另一类是节点删除方法, 即通过度量节点删除后对社交网络的破坏程度来确定节点影响力的方法.

中心度方法主要是通过社交网络中节点的度或节点间最短路径等因素来判断社交网络中节点的影响力, 计算复杂程度较低, 因此这类方法被广泛应用于社

收稿日期: 2015-01-30; 修回日期: 2016-01-20; 责任编辑: 梅志强

基金项目: 国家自然科学基金 (No. 61672050, No. 61372191, No. 61572492, No. 61572046, No. 61502012); 国家 973 重点基础研究发展计划 (No. 2013CB329600); 中国博士后科学基金 (No. 2016M591013); 江西省教育厅项目 (No. GJJ150686)

交网络的结构分析技术中^[11].

度中心度的方法是通过节点在社交网络中的大小来确定节点的影响力,即通过每个节点的邻居节点数来判断,其优势是计算代价小,可应用于大规模网络中.但是,节点度的大小只能体现社交网络结构的一个局部性质,所反映的是节点与其邻居节点之间的关系,无法判断节点在整个社交网络中的影响力.为此,Chen 等人^[12]在度中心度的基础上,进一步提出了局部中心度的方法来度量节点在社交网络中的影响力.节点 v 的局部中心度 $C_{\text{LOC}}(v)$ 定义为:

$$Q(u) = \sum_{w \in N_1(u)} |N_2(w)|, C_{\text{LOC}}(v) = \sum_{u \in N_1(v)} Q(u),$$

其中 $N_i(u)$ 表示到节点 u 的距离不超过 i 的所有节点构成的集合, $i=1, 2$. 局部中心度实际上是度中心度的推广,节点 v 的局部中心度受到与 v 的距离不超过 4 的所有节点的影响,并且与 v 的距离越小对其影响越大.

紧密中心度和介数中心度方法主要是通过社交网络中最短路径来刻画节点的影响力.节点 v 的紧密中心度 $C_{\text{CLO}}(v)$ 为:

$$C_{\text{CLO}}(v) = \frac{1}{\sum_{w \in V} d(v, w)},$$

其中 $d(v, w)$ 表示节点 v 到节点 w 的距离,即它们之间的最短路径的长度;节点 v 的介数中心度 $C_{\text{BET}}(v)$ 定义为:

$$C_{\text{BET}}(v) = \sum_{u \neq v \neq w \in V} \frac{\sigma_{uw}(v)}{\sigma_{uw}},$$

其中 σ_{uw} 表示节点 u 与节点 w 之间最短路的条数, $\sigma_{uw}(v)$ 表示这些最短路中经过节点 v 的条数.可以看出,紧密中心度所反映的是节点在社交网络中的居中程度,节点的紧密中心度越大,说明此节点到其他节点的平均距离越小,即从此节点到其他节点的信息传播越快,也就间接地表示该节点在社交网络中的影响力越大.而介数中心度反映的是节点在社交网络拓扑结构中位置的重要性,是对其他节点之间信息传播影响的一种度量指标.节点的介数中心度越大,说明社交网络中信息传播时经过该节点的信息量越大,即表示该节点在社交网络中的影响力越大;反之也就越小.

特征向量中心度的方法主要是利用随机游走的特征来判断节点的重要程度.节点的特征向量中心度定义如下:

$$\lambda x_i = \sum_{j=1}^n a_{i,j} x_j, i=1, 2, \dots, n,$$

其中 λ 表示邻接矩阵的最大特征值.特征向量中心度所反映的是节点的影响力与其邻接节点的影响力相关,特别是与游走路径上其他节点的影响力相关,其他节点影响力越大,当前节点的影响力也就越大,反之就

越小.

另一类度量社交网络中节点影响力的方法是节点删除方法,主要通过刻画节点删除后对社交网络结构的破坏程度来判断节点的影响力,其考虑到社交网络的整体连通程度,这类方法更多的是应用在系统科学领域的研究中,主要理论有网络的坚韧集与坚韧度以及核与核度等.

对一个社交网络图 $G=(V, E)$,若节点子集 $S \subset V$ 满足 $G-S$ 不连通,则称 S 为 G 的一个节点割集.记 $C(G)$ 为 G 中所有节点割集构成的集合;对一个节点割集 S ,记 $\omega(G-S)$ 为 $G-S$ 的连通分支的数目.网络图 G 的坚韧度 $\tau(G)$ ^[13] 定义如下:

$$\tau(G) = \min \left\{ \frac{|V'|}{\omega(G-V')} : V' \in C(G) \right\} \quad (1)$$

且将满足式(1)的节点割集 S^* 称为 G 的坚韧集.网络图 G 的核度 $h(G)$ ^[14] 定义如下:

$$h(G) = \max \left\{ \omega(G-V') - |V'| : V' \in C(G) \right\} \quad (2)$$

且将满足式(2)的节点割集 S^* 称为 G 的核.

图的核度最初被称为离散数^[15].Jung^[15] 提出图的离散度是用于研究图的 Hamilton 性问题的.此后,许进等^[14] 将图的离散数引入到系统以及网络的研究中,并形象地称其为系统的核度.

从上述两个定义可以看出,社交网络图的坚韧度与核度都是根据节点割集的阶数和删除这个节点割集后的网络的连通分支数来定义的,均是刻画社交网络整体连通程度的指标.不同的是,根据图的坚韧度来判断时,坚韧度越大,说明连通性越好;但根据核度来判断时,核度越小其连通性越好.相应地,坚韧集与核均是刻画对网络具有重要的或支配性的节点(或节点集)的指标,反映的是社交网络的整体性质,无法对同一个网络中不同节点之间的影响力作出判断.而且,在应用坚韧度或核度来判断社交网络连通性及稳定性时,计算量非常大,不适用于较大规模的社交网络.对社交网络中节点影响力,特别是在线社交网络中节点影响力度量方法和模型的详细介绍可参见文献[16,17].

最近,仍有许多学者研究社交网络的节点影响力.Ma, Ren 和 Ye 等^[18] 提出了基于资源分配动态和 k -层分解的节点影响力度量方法. Saito, Kimura 和 Ohara 等^[19] 提出了基于信息传播的节点影响力度量方法,称为超级传递者. Hu 和 Liu^[20] 结合特征向量中心度、介数中心度、紧密中心度、度中心度以及相互信息等因素,提出了基于线性判别分析的多指标算法. Kim 和 Anderson^[21], Kas, Wachs 和 Carley 等^[22], Gao, Shi 和 Chen^[23] 研究了动态社交网络下的节点影响力度量方法.另外,关于社交网络影响力最大化的研究,可参见文献[24~29].

本文在中心度方法与节点删除方法的基础上,提

出了一种新的度量社交网络中节点影响力的参数:连通中心度. 这个参数结合了节点在社交网络结构中位置的重要性以及对其他节点之间的局部连通度的影响力.

2 连通中心度

在本文中,用图 $G=(V,E)$ 表示社交网络的拓扑结构,其中 V 表示节点集, E 表示边(连接)集. 若无特别说明,均指无向图. 记 $n=|V|$ 为 G 的节点数, $m=|E|$ 为 G 的边数.

对图 $G=(V,E)$ 中的两个不相邻的节点 v_i 与 v_j ,若节点子集 $S \subset V$ 满足在 $G-S$ 中节点 v_i 与 v_j 不连通,则称 S 为 v_i 与 v_j 的一个分离集. 用 $c(v_i, v_j)$ 表示 G 中节点 v_i 与 v_j 的最小分离集中节点的数目,称 $c(v_i, v_j)$ 为节点 v_i 与 v_j 的局部连通度. 记

$C(v_i, v_j) = \{v \in S: S \text{ 是 } v_i, v_j \text{ 的分离集且 } |S| = c(v_i, v_j)\}$.

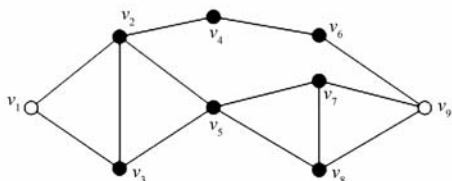


图1 一个9节点的网络例子

例如,在图1中, $S = \{v_4, v_5\}$ 是节点 v_1 与 v_9 的一个分离集; $S = \{v_6, v_7, v_8\}$ 也是节点 v_1 与 v_9 的一个分离集;并且,节点 v_1 与 v_9 的局部连通度为2, $C(v_1, v_9) = \{v_2, v_3, v_4, v_5, v_6\}$.

用 $p(v_i, v_j)$ 表示 G 中节点 v_i 与 v_j 之间内部互不相交的路的最大条数. Menger^[30] 证明了 $c(v_i, v_j)$ 与 $p(v_i, v_j)$ 有如下关系:

定理1^[30] 设 $G=(V,E)$ 是一个图, v_i 与 v_j 是 G 中不相邻的两个节点,则有 $c(v_i, v_j) = p(v_i, v_j)$.

2.1 连通中心度的定义

定义1 连通中心度. 对一个图 $G=(V,E)$, v_i, v_j 是 G 中的两个不相邻的节点,任意节点 $v_t \in V \setminus \{v_i, v_j\}$ 对 v_i, v_j 的影响力定义为

$$\varphi_{ij}(v_t) = \begin{cases} \max\left\{\frac{1}{p(v_i, v_j)}, \frac{\sigma_{ij}(v_t)}{\sigma_{ij}}\right\}, & \text{若 } v_t \in C(v_i, v_j); \\ \frac{\sigma_{ij}(v_t)}{\sigma_{ij}}, & \text{否则.} \end{cases}$$

其中 σ_{ij} 表示节点 v_i 与 v_j 之间最短路的条数, $\sigma_{ij}(v_t)$ 表示这些最短路中经过节点 v_t 的条数. 当 v_i, v_j 在 G 中相邻时,定义 $\varphi_{ij}(v_t) = 0$.

节点 v_t 的连通中心度定义为 v_t 对 G 中所有的节点对 v_i 与 v_j 的影响力之和,即

$$C_{\text{CON}}(v_t) = \sum_{i \neq j} \varphi_{ij}(v_t).$$

特别,如果 $G-v_t$ 是一个完全图,即 $G-v_t$ 中任意两个节点是相邻的,则定义 $C_{\text{CON}}(v_t) = 0$.

由定义1可知,对 G 中的任意两个不相邻的节点 v_i, v_j 以及 $v_t \in V \setminus \{v_i, v_j\}$, 有

$$0 \leq \varphi_{ij}(v_t) \leq 1.$$

若 $\varphi_{ij}(v_t) = 0$, 则 v_t 不在 v_i 与 v_j 的任何一个分离集中,也不在它们的最短路上,即说明在网络 G 中,节点 v_t 对 v_i, v_j 没有影响;若 $\varphi_{ij}(v_t) = 1$, 则节点 v_t 分离 v_i 与 v_j 或节点 v_i 与 v_j 的任意最短路必经过节点 v_t , 说明在网络 G 中,节点 v_t 对 v_i, v_j 的影响很大.

连通中心度既考虑了经过节点的最短路径的数目,也考虑了此节点对其他节点对的局部连通性,是一个从网络整体出发度量节点影响力的指标. 节点的连通中心度越大,说明此节点对整个网络的影响越大,重要性越高. 反之,连通中心度越小,对网络的影响越小,重要性越低.

例如,在图1中,节点 v_1 与节点 v_9 之间的最短路共有5条,其中经过 v_4 的只有1条,经过 v_8 的有2条;在 G 中将节点 v_4 删除后, v_1 与 v_9 之间内部不相交的路只剩1条,而将 v_8 删除后, v_1 与 v_9 之间内部不相交的路仍有2条. 因此,节点 v_4 对 v_1, v_9 的影响力为 $\varphi_{1,9}(v_4) = \frac{1}{2}$;

节点 v_8 对 v_1, v_9 的影响力为 $\varphi_{1,9}(v_8) = \frac{2}{5}$, 故对节点对 v_1 与 v_9 来说,节点 v_4 的影响力更大一些. 如果采用介数中心度来衡量节点影响力,可得到节点 v_4 比节点 v_8 对 v_1, v_9 的影响力小.

2.2 连通中心度的性质

在一个社交网络中,节点的连通中心度反映的是当前节点对其他节点相互交往过程中的影响力. 连通中心度的大小受到社交网络结构的限制. 下面根据网络的结构,分析连通中心度的性质.

根据定义1,如果节点 v 是 G 的叶子节点,那么 v_t 对其余节点的影响力为0,即有:

性质1 设节点 v 是网络图 G 的叶子节点,则 $C_{\text{CON}}(v) = 0$.

性质2 设 G 是包含 n 个节点的完全图,则根据定义1可知, G 中任意节点 v 的连通中心度为 $C_{\text{CON}}(v) = 0$.

通过性质1和性质2可以看出,一个节点在社交网络中的影响力与其度的大小没有必然的联系,而与网络的整体结构或局部结构关系紧密. 在性质1与性质2的基础上,可得到更一般的结论.

性质3 设 v 是网络图 G 的一个节点, v 在 G 中的所有邻居节点构成的集合记为 $N(v)$, 若 $N(v)$ 在 G 中的

导出子图是一个完全图,则 $C_{\text{CON}}(v) = 0$.

通常,在一个规模较大的网络中存在很多节点具有相同的影响力,下面给出判断两个节点具有相同连通中心度的一个充分条件.

性质 4 设 v, w 是网络图 G 中的两个节点,若 $G - v$ 与 $G - w$ 是同构的,则节点 v 与节点 w 的连通中心度相同,即有 $C_{\text{CON}}(v) = C_{\text{CON}}(w)$.

2.3 连通中心度的计算

对一个图 $G = (V, E)$, 分别用 $n = |V|$ 和 $m = |E|$ 表示 G 的节点数和边数. 对 G 中的任意节点 v_i , 计算 v_i 的连通中心度最直观的方式就是对每一对不相邻的节点 v_i, v_j 计算出 $\varphi_{ij}(v_i)$, 然后将所有的非相邻节点对的值进行求和计算出 $C_{\text{CON}}(v_i)$, 具体的计算步骤如下:

(1) 需要计算 G 中每对不相邻节点 v_i, v_j 间的内部互不相交的路的条数, 即计算 v_i, v_j 的局部连通度, 以及 v_i, v_j 之间最小节点割集的并构成的集合 $C(v_i, v_j)$. Ford 和 Fulkerson^[31,32] 给出了一种利用最大流-最小割的方法来计算不相邻节点的局部连通度. 利用这种方法, Edmonds 和 Karp^[33] 给出一个时间复杂度为 $O(n \cdot m^2)$ 的算法; 对边稠密的网络, 可利用 Goldberg 和 Tarjan^[34] 提出的 relabel-to-front 算法, 其时间复杂度为 $O(n^3)$.

(2) 需要计算每对不相邻节点间最短路径的条数. Floyd^[35] 给出了在一般图中复杂度为 $O(n^3)$ 的算法, 而对边稀疏的网络, Johnson^[36] 给出了一个时间复杂度为 $O(n^2 \log n + nm)$ 的算法.

(3) 根据定义 1, 计算节点 $v_i \in V \setminus \{v_i, v_j\}$ 对节点对 v_i, v_j 的影响力.

(4) 计算出节点 v_i 的连通中心度 $C_{\text{CON}}(v_i)$.

综上所述, 通过上述算法计算图 G 中任意节点 v_i 的连通中心度 $C_{\text{CON}}(v_i)$ 的时间复杂度为 $O(n^3)$, 其中 n 为 G 的节点数.

以图 1 所示的 9 个节点的网络 G 为例, 下面计算每个节点 v_i 的连通中心度. G 中任意不相邻的节点对 v_i, v_j 的局部连通度、最短路的条数以及最小节点割集的并构成的集合见表 1.

通过表 1 可得到 G 中节点连通中心度依次为:

$$\begin{aligned} C_{\text{CON}}(v_1) &= 0; C_{\text{CON}}(v_2) = 12.1; C_{\text{CON}}(v_3) = 3; \\ C_{\text{CON}}(v_4) &= 9; C_{\text{CON}}(v_5) = 13.47; C_{\text{CON}}(v_6) = 8.33; \\ C_{\text{CON}}(v_7) &= C_{\text{CON}}(v_8) = 2.07; C_{\text{CON}}(v_9) = 8.67. \end{aligned}$$

可以看出, 影响力最大的节点是 v_5 , 最小的是 v_1 .

进一步, 每个节点的连通中心度与其在网络拓扑结构中的具体位置有着重要的关系. 例如, 在网络图 G 中, 除节点 v_1 外, v_7, v_8 的连通中心度是最小的, 这是因为其余节点之间的经过 v_7 或 v_8 的互不相交的路以及最短路都很少. 从而说明在进行信息传播时, 节点 v_7 或 v_8 起到的作用很小, 对整个网络中的影响也就很小. 另外,

节点 v_4 的度数小于 v_9 的度数, 但 v_4 的连通中心度大于 v_9 的连通中心度. 这说明节点的连通中心度不依赖于其度数的大小.

表 1 图 1 所示网络中任意节点对的局部连通度、最短路条数和最小节点割集的并集的计算

节点对	$c(v_i, v_j)$	σ_{ij}	$C(v_i, v_j)$
v_1, v_4	2	1	$\{v_2, v_3, v_5, v_6, v_9\}$
v_1, v_5	2	2	$\{v_2, v_3\}$
v_1, v_6	2	1	$\{v_2, v_3, v_4, v_5, v_9\}$
v_1, v_7	2	2	$\{v_2, v_3, v_4, v_5, v_6, v_9\}$
v_1, v_8	2	2	$\{v_2, v_3, v_4, v_5, v_6, v_9\}$
v_1, v_9	2	5	$\{v_2, v_3, v_4, v_5, v_6\}$
v_2, v_6	2	1	$\{v_4, v_5, v_9\}$
v_2, v_7	2	1	$\{v_4, v_5, v_6, v_9\}$
v_2, v_8	2	1	$\{v_4, v_5, v_6, v_9\}$
v_2, v_9	2	3	$\{v_4, v_5, v_6\}$
v_3, v_4	2	1	$\{v_2, v_5, v_6, v_9\}$
v_3, v_6	2	1	$\{v_2, v_4, v_5, v_9\}$
v_3, v_7	2	1	$\{v_2, v_4, v_5, v_6, v_9\}$
v_3, v_8	2	1	$\{v_2, v_4, v_5, v_6, v_9\}$
v_3, v_9	2	2	$\{v_2, v_4, v_5, v_6\}$
v_4, v_5	2	1	$\{v_2, v_6, v_9\}$
v_4, v_7	2	2	$\{v_2, v_5, v_6, v_9\}$
v_4, v_8	2	2	$\{v_2, v_5, v_6, v_9\}$
v_4, v_9	2	1	$\{v_2, v_5, v_6\}$
v_5, v_6	2	3	$\{v_2, v_4, v_9\}$
v_5, v_9	3	2	$\{v_2, v_4, v_6, v_7, v_8\}$
v_6, v_7	2	1	$\{v_2, v_4, v_5, v_9\}$
v_6, v_8	2	1	$\{v_2, v_4, v_5, v_9\}$

3 特殊网络中节点的连通中心度

树状网络是最简单的一种连通网络, 下面通过一种简单的方法, 给出树中任意非叶子节点的连通中心度.

定理 2 设 $T = (V, E)$ 是树, $v \in V$ 是 T 的非叶子节点, 则 v 的连通中心度为

$$C_{\text{CON}}(v) = \sum_{1 \leq i < j \leq r} n_i n_j,$$

其中 r 表示 $T - v$ 的连通分支数, n_k 表示 $T - v$ 中第 k 个连通分支的节点数, $k = 1, 2, \dots, r$.

证明 由于 T 是树, 对 T 中任意一对不相邻的节点 v_i, v_j , 存在唯一的一条路连接节点 v_i 与 v_j , 因此有

$$\varphi_{ij}(v) = \begin{cases} 1, & \text{若 } v \in C(v_i, v_j); \\ 0, & \text{否则.} \end{cases}$$

所以, 节点 v 只对 $T - v$ 中不连通分支上的节点对有影响, 从而

$$C_{\text{CON}}(v) = \sum_{v \in C(v, v_j)} \varphi_{ij}(v) = \sum_{v \in C(v, v_j)} 1 = \sum_{1 \leq i < j \leq r} n_i n_j.$$

证毕.

下面考虑圈中节点的连通中心度. 对含 4 个节点的圈, 易验证每个节点 v 的连通中心度为 $C_{\text{CON}}(v) = \frac{1}{2}$. 当圈中节点数至少为 5 时, 有下面的结论.

定理 3 设 $C = (V, E)$ 是包含 n 个节点的圈, $n \geq 5$, 则 C 中任意节点 v 的连通中心度为

$$C_{\text{CON}}(v) = \begin{cases} \frac{1}{4}n(n-3) - \frac{1}{2}, & \text{若 } n \text{ 是奇数;} \\ \frac{1}{4}n(n-3) - 1, & \text{若 } n \text{ 是偶数.} \end{cases}$$

证明 对 C 中任意节点 v , $C-v$ 中不相邻的节点对的数目为

$$\begin{aligned} & \binom{n-1}{2} - (n-2) \\ &= \frac{1}{2}(n-1)(n-2) - (n-2) \\ &= \frac{1}{2}(n-2)(n-3). \end{aligned}$$

另一方面, 可以看出, 对 $C-v$ 中任意一对不相邻的节点 v_i, v_j , 都有 $v \in C(v_i, v_j)$ 成立. 进一步, 当 n 是奇数时, C 中任意一对不相邻的节点 v_i, v_j 之间的最短路径只有 1 条, 且恰有 $n-4$ 对不相邻的节点, 满足其最短路径经过节点 v . 所以, v 的连通中心度为

$$\begin{aligned} C_{\text{CON}}(v) &= 1 \cdot (n-4) + \frac{1}{2} \cdot \left\{ \frac{1}{2}(n-2)(n-3) - (n-4) \right\} \\ &= \frac{1}{4}n(n-3) - \frac{1}{2}. \end{aligned}$$

当 n 是偶数时, 由于 $n \geq 6$, 可以验证 C 中恰有 $n-5$ 对不相邻的节点, 满足其最短路径经过节点 v . 所以, v 的连通中心度为

$$\begin{aligned} C_{\text{CON}}(v) &= 1 \cdot (n-5) + \frac{1}{2} \cdot \left\{ \frac{1}{2}(n-2)(n-3) - (n-5) \right\} \\ &= \frac{1}{4}n(n-3) - 1. \end{aligned}$$

证毕.

定理 4 对完全二部图 $K_{s,t}$, $s, t \geq 2$, 令 $K_{s,t}$ 的二部划分为 V_1, V_2 , 其中 $V_1 = \{v_i: 0 \leq i \leq s-1\}$, $V_2 = \{u_j: 0 \leq j \leq t-1\}$, 有

(1) 若 $s \neq t$, 则任意节点 v_i 的连通中心度为

$$C_{\text{CON}}(v_i) = \frac{t(t-1)}{2s};$$

任意节点 u_j 的连通中心度为

$$C_{\text{CON}}(u_j) = \frac{s(s-1)}{2t},$$

其中 $i=0, 1, \dots, s-1; j=0, 1, \dots, t-1$.

(2) 若 $s=t$, 则 $K_{s,t}$ 中任意节点 v 的连通中心度为

$$C_{\text{CON}}(v) = \frac{1}{2}(s-1).$$

证明

(1) 因为 V_1, V_2 是 $K_{s,t}$ 的二部划分, 所以 V_1 中的节点互不相邻, V_2 中的节点也互不相邻. 故对 V_1 中任意不相邻节点对 v_{i_1}, v_{i_2} , 它们之间存在 $|V_2|$ 条最短路, 且对每个 V_2 中的节点 u_j , 有一条最短路经过 u_j , 因此,

$$C_{\text{CON}}(u_j) = \frac{1}{|V_2|} \cdot \binom{|V_1|}{2} = \frac{|V_1|(|V_1|-1)}{2|V_2|} = \frac{s(s-1)}{2t},$$

其中 $j=0, 1, \dots, t-1$.

同理可得:

$$C_{\text{CON}}(u_j) = \frac{1}{|V_1|} \cdot \binom{|V_2|}{2} = \frac{|V_2|(|V_2|-1)}{2|V_1|} = \frac{t(t-1)}{2s},$$

其中 $i=0, 1, \dots, s-1$.

(2) 若 $s=t$, 则由上述分析可知, 对 $K_{s,t}$ 中任意节点 v , 有

$$C_{\text{CON}}(v) = \frac{1}{2}(s-1).$$

证毕.

定理 5 对完全 k -部图 K_{n_1, n_2, \dots, n_k} , 令它的 k -部划分为 V_1, V_2, \dots, V_k , 则对任意的节点 $v_i \in V_i$, v_i 的连通中心度为

$$C_{\text{CON}}(v_i) = \sum_{\substack{r=1 \\ r \neq i}}^k \frac{n_r(n_r-1)}{2(n-n_r)},$$

其中 $i=1, 2, \dots, k, n_1+n_2+\dots+n_k=n$.

证明 由于 K_{n_1, n_2, \dots, n_k} 的不同部分之间的节点都是相邻的, 故每对不相邻的节点属于同一部分. 对任意的节点对 $v_{i_1}, v_{i_2} \in V_r$, v_{i_1}, v_{i_2} 之间内部不相交的路的条数为 $n-n_r$, 因此 V_i 中的任意节点 v_i 对 v_{i_1}, v_{i_2} 的影响力为:

$$\varphi_{i_1, i_2}(v_i) = \frac{1}{n-n_r},$$

其中 $i \in \{1, 2, \dots, k\}$, 且 $i \neq r$. 进一步, V_r 中互不相邻的节点对的数目为:

$$\binom{|V_r|}{2} = \binom{n_r}{2} = \frac{n_r(n_r-1)}{2}.$$

因此, 节点 v_i 对 V_r 中所有节点对的影响之和为:

$$\binom{|V_r|}{2} \cdot \varphi_{i_1, i_2}(v_i) = \frac{n_r(n_r-1)}{2(n-n_r)}$$

所以, 节点 v_i 的连通中心度为:

$$C_{\text{CON}}(v_i) = \sum_{\substack{r=1 \\ r \neq i}}^k \frac{n_r(n_r-1)}{2(n-n_r)}.$$

证毕.

4 数据分析

本节以 Karate 俱乐部成员网络^[37]为例, 如图 2 所

示,通过连通中心度计算每个成员的影响力,并将此指标与度中心度、紧密中心度和介数中心度等指标进行了比较,其中节点的标号与文献[37]中是一致的。

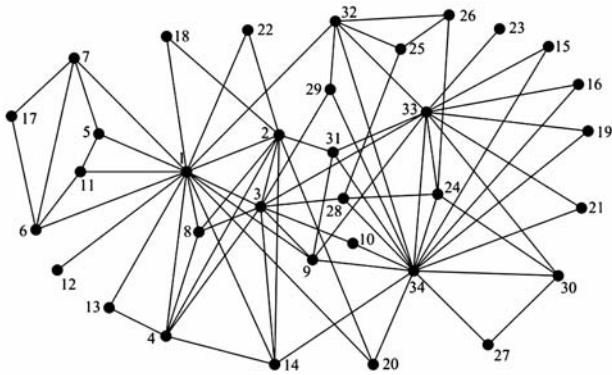


图2 Karate俱乐部成员关系网络

表2 Karate俱乐部网络节点影响力4种度量参数的比较

节点 v	$C_{\text{DEG}}(v)$	$C_{\text{CLO}}(v)$	$C_{\text{BET}}(v)$	$C_{\text{CON}}(v)$
1	16	0.01724	231.08	235.83
2	9	0.01471	28.45	47.99
3	10	0.01695	80.40	88.91
4	6	0.01408	6.32	16.27
5	3	0.01149	0.33	0.33
6	4	0.01163	15.83	16.33
7	4	0.01163	15.83	16.33
8	4	0.01333	0	0
9	5	0.01563	29.96	31.28
10	2	0.01299	0.45	0.45
11	3	0.01149	0.33	0.33
12	1	0.01111	0	0
13	2	0.01124	0	0
14	5	0.01538	21.96	28.53
15	2	0.01124	0	0
16	2	0.01124	0	0
17	2	0.00862	0	0
18	2	0.01136	0	0
19	2	0.01124	0	0
20	3	0.01493	15.50	17.26
21	2	0.01124	0	0
22	2	0.01136	0	0
23	1	0.01042	0	0
24	5	0.01190	9.63	13.65
25	3	0.01136	1.17	5.17
26	3	0.01136	2.03	8.16
27	2	0.01087	0	0
28	4	0.01370	11.46	15.70
29	3	0.01351	0.95	6.21
30	4	0.01163	2.04	14.46
31	4	0.01389	7.78	8.23
32	6	0.01640	74.10	83.36
33	12	0.01563	96.18	117.38
34	16	0.01639	145.23	155.71

从表2可以看出,通过连通中心度来判断 Karate 俱

乐部网络中各节点的影响力时,影响力最大的5个节点依次为:节点1,节点34,节点33,节点3,节点32.因此,通过连通中心度能够准确地选择影响力较大的节点。

连通中心度与介数中心度比较相似,介数中心度为0的节点的连通中心度也为0,但是,介数中心度不为0的节点,其连通中心度总是大于等于介数中心度,这是因为连通中心度在计算的过程中考虑了每对不相邻节点间的内部互不相交的路的条数,即节点对的局部连通度。

一个节点 v 的连通中心度表示的是 v 对其他节点影响的大小,其值与度中心度、紧密中心度以及介数中心度没有必然的联系.例如,对节点30与节点31, $C_{\text{DEG}}(30) = C_{\text{DEG}}(31)$, $C_{\text{CLO}}(30) < C_{\text{CLO}}(31)$, $C_{\text{BET}}(30) < C_{\text{BET}}(31)$,但是, $C_{\text{CON}}(30) > C_{\text{CON}}(31)$. 因为节点27到其余很多节点的最短路或内部互不相交的路经过节点30,说明节点30对节点27与其他节点间的信息传播起着重要的作用,从而使得节点30的连通中心度较大.而经过节点31的最短路或内部互不相交的路较少,因此节点31的连通中心度较小.这说明,在 Karate 俱乐部成员网络中,节点30比节点31的影响力大。

节点对之间的内部互不相交的路是信息并行传递的基础,因此,利用连通中心度能更好地度量社交网络中节点的影响力。

5 结论

本文在目前两类度量网络节点影响力方法的基础上,即结合了中心度方法和节点删除方法,提出了一种新的刻画网络中节点重要程度的方法.该方法既考虑了节点在网络结构中位置的重要性,也考虑了节点对网络局部连通性的影响,更全面地刻画了节点的重要程度.在信息并行传播的社交网络中,其传播路径为两个节点间的内部互不相交的路,因此对具有这种特征的网络,利用连通中心度来度量节点的重要性更准确一些。

不同性质的网络具有不同的结构特征,我们所考虑的只是无向的连通网络,对有向网络和赋值网络还需进一步的研究.另外,针对规模很大的网络,计算其节点的连通中心度将耗费大量的计算资源,精确计算每个节点连通中心度不现实.所以,将连通中心度应用于大规模网络中时,需要对计算过程做进一步的简化,但要求对节点的连通中心度的影响不太大,这也是一个值得研究的问题。

参考文献

- [1] Newman M E. The structure and function of complex networks[J]. SIAM Review, 2003, 45(2): 167-256.

- [2] Boccaletti S, Latora V, Moreno Y, et al. Complex networks: Structure and dynamics[J]. *Physics Reports*, 2006, 424 (4): 175 – 308.
- [3] Zhang J, Zhou C, Xu X, et al. Mapping from structure to dynamics: A unified view of dynamical processes on networks[J]. *Physical Review E*, 2010, 82 (2): 026116.
- [4] Zhou T, Wang B H. Catastrophes in scale-free networks [J]. *Chinese Physics Letters*, 2005, 22 (5): 1072 – 1075.
- [5] Pastor-Satorras R, Vespignani A. Immunization of complex networks[J]. *Physical Review E*, 2002, 65 (3): 036104.
- [6] Freeman L C. Centrality in social networks conceptual clarification[J]. *Social Networks*, 1979, 1(3): 215 – 239.
- [7] Sabidussi G. The centrality index of a graph [J]. *Psychometrika*, 1966, 31(4): 581 – 603.
- [8] Freeman L C. A set of measures of centrality based on betweenness[J]. *Sociometry*, 1977, 40(1): 35 – 41.
- [9] Bonacich P. Factor and weighting approaches to status scores and clique identification[J]. *Journal of Mathematical Sociology*, 1972, 2(1): 113 – 120.
- [10] Bonacich P. Some unique properties of eigenvector centrality[J]. *Social Networks*, 2007, 29(4): 555 – 564.
- [11] Borgatti S P, Everett M G. A graph-theoretic perspective on centrality [J]. *Social Networks*, 2006, 28 (4): 466 – 484.
- [12] Chen D B, Lü L Y, Shang M S, et al. Identifying influential nodes in complex networks[J]. *Physica A*, 2012, 391 (4): 1777 – 1787.
- [13] Chvátal V. Tough graphs and Hamiltonian circuits[J]. *Discrete Mathematics*, 1973, 5 (3): 215 – 228.
- [14] 许进, 席酉民, 汪应洛. 系统的核与核度(I) [J]. *系统科学与数学*, 1993, 13(2): 102 – 110.
Xu J, Xi Y M, Wang Y L. On system core and coritivity (1) [J]. *Journal of System Science and Mathematical Science*, 1993, 13(2): 102 – 110. (in Chinese)
- [15] Jung H A. On a class of posets and the corresponding comparability graphs[J]. *Journal of Combinatorial Theory Series B*, 1978, 24(2): 125 – 133.
- [16] Sun J, Tang J. A survey of models and algorithms for social influence analysis[A]. *Social Network Data Analytics [C]*. New York: Springer, 2011. 177 – 214.
- [17] 吴信东, 李毅, 李磊. 在线社交网络影响力分析[J]. *计算机学报*, 2014, 37(4): 735 – 752.
Wu X D, Li Y, Li L. Influence analysis of online social networks [J]. *Chinese Journal of Computers*, 2014, 37 (4): 735 – 752. (in Chinese)
- [18] Ma S J, Ren Z M, Ye C M, et al. Node influence identification via resource allocation dynamics[J]. *International Journal of Modern Physics C*, 2014, 25(11): 1450065.
- [19] Saito K, Kimura M, Ohara K, et al. Super mediator—A new centrality measure of node importance for information diffusion over social network[J]. *Information Sciences*, 2016, 329: 985 – 1000.
- [20] Hu F, Liu Y H. Multi-index algorithm of identifying important nodes in complex networks based on linear discriminant analysis[J]. *Modern Physics Letters B*, 2015, 29 (03): 1450268.
- [21] Kim H, Anderson R. Temporal node centrality in complex networks [J]. *Physical Review E*, 2012, 85 (2): 605 – 624.
- [22] Kas M, Wachs M, Carley K M, et al. Incremental algorithm for updating betweenness centrality in dynamically growing networks[A]. *Advances in Social Networks Analysis and Mining (ASONAM)*, 2013 IEEE/ACM International Conference on IEEE[C]. Niagara Falls, Canada: IEEE/ACM, 2013. 33 – 40.
- [23] Gao Z X, Shi Y, Chen S Z. Measures of node centrality in mobile social networks[J]. *International Journal of Modern Physics C*, 2015, 26(09): 1550107.
- [24] Zhang X, Zhu J, Wang Q, et al. Identifying influential nodes in complex networks with community structure [J]. *Knowledge-Based Systems*, 2013, 42(2): 74 – 84.
- [25] Shakarian P, Salmento J, Pulleyblank W, et al. Reducing gang violence through network influence based targeting of social programs [A]. *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining [C]*. New York, USA: ACM, 2014. 1829 – 1836.
- [26] He X, Kempe D. Stability of influence maximization[A]. *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining [C]*. New York, USA: ACM, 2014. 1256 – 1265.
- [27] Zhou C, Zhang P, Zang W, et al. Maximizing the long-term integral influence in social networks under the voter model[A]. *Proceedings of the 23rd International Conference on World Wide Web Conferences Steering Committee [C]*. Seoul, Korea: ACM, 2014. 423 – 424.
- [28] Morone F, Makse H A. Influence maximization in complex networks through optimal percolation [J]. *Nature*, 2015, 524(7563): 65 – 68.
- [29] Lamba H, Narayanam R. A model-independent approach for efficient influence maximization in social networks [J]. *Social Network Analysis & Mining*, 2015, 5(1): 1 – 11.
- [30] Menger K. Zur allgemeinen Kurventheorie[J]. *Fundamenta Mathematicae*, 1927, 10: 96 – 115.
- [31] Ford L P, Fulkerson D R. Maximal flow through a network [J]. *Canadian Journal of Mathematics*, 1956, 8(3): 399 – 404.

- [32] Ford L P, Fulkerson D R. Flows in Networks[M]. Princeton NJ: Princeton University Press, 1962.
- [33] Edmonds J, Karp R M. Theoretical improvements in algorithmic efficiency for network flow problems[J]. Journal of the ACM, 1972, 19(2): 248 - 264.
- [34] Goldberg V A, Tarjan E R. A new approach to the maximum flow problem[J]. Journal of the ACM, 1988, 35(4): 921 - 940.
- [35] Floyd R W. Algorithm 97: Shortest path[J]. Communications of the ACM, 1962, 5(6): 345.
- [36] Johnson D B. Efficient algorithms for shortest paths in sparse networks[J]. Journal of the ACM, 1977, 24(1): 1 - 13.
- [37] Zachary W W. An information flow model for conflict and fission in small groups[J]. Journal of Anthropological Research, 1977, 33(4): 452 - 473.

作者简介



李泽鹏(通讯作者) 男, 1987年出生于甘肃定西. 北京大学信息科学技术学院博士后, 主要研究方向为图论与组合优化, 社交网络拓扑结构.

E-mail: lizepeng@pku.edu.cn



左 杨 男, 1990年出生于安徽安庆. 北京大学信息科学技术学院硕士研究生, 主要研究方向为社交网络结构分析.

王宏宇 女, 1988年出生于黑龙江牡丹江. 北京大学信息科学技术学院博士研究生, 主要研究方向为图论与组合优化, 社交网络结构.