

复杂网络图中心节点分布的研究^{*}

陈 琼 方 明 陈志云

(华南理工大学)

【摘 要】复杂网络图中不同的节点的作用和重要程度不同. 节点的中心性度量, 可以反映节点在图中的作用以及对其他节点的影响. 复杂网络的全局中心节点有时分布较为集中, 不能很好地反映网络图各处的重要节点. 局部中心点是中心性度量值大于等于它的所有邻居节点的中心性度量值的节点, 分散分布在网络图中, 分散分布在社区中, 是复杂网络中具有大的局部影响力的重要节点.

【关键词】复杂网络; 节点中心性; 图; 中心点

0 引言

许多现实世界的复杂系统都可以用复杂网络图来描述, 如社会网络、internet 网、蛋白质相互作用网、新陈代谢网、通信网等. 这些复杂网络图的不同节点具有不同的作用. 近年来随着大规模网络的研究, 复杂网络图的节点在网络中的作用的研究引起了国内外学者的广泛关注. 提出了许多结合实际背景的重要节点重要性度量方法, 从度中心性、介数中心性、接近中心性、特征向量中心性等方面刻画节点的重要性^[1]. 其中基于度的度中心性评估方法最为简单, 描述的是节点与邻接节点连接边的数量, 介数中心性刻画了节点或边对网络中信息或流的控制能力, 接近中心性分析节点在网络中的聚集能力, 特征向量中心性考虑与目标节点建立连接的节点的重要性并通过邻接节点的重要性来确定目标节点的地位作用. 通过节点中心性的研究可以发现复杂网络图中的重要和关键节点, 如社会网络中的领袖人物、通信网络中对信息传播具有重要控制能力的节点等.

利用节点中心性度量方法可以计算每个节

点的中心性值. 节点的中心性值较大的节点较为重要, 具有较大的影响力. 但是如果一个节点距离中心性值大的节点的距离很远, 该节点受它的某个邻近节点的影响可能比受中心性值大的节点的影响更大. 因此网络中存在一些具有较大的局部影响力的节点. 通过研究复杂网络图的中心节点的分布, 提出了节点的局部中心性分析方法. 通过局部中心性分析可以找到具有较大局部影响力的节点, 这些节点也是复杂网络图的重要节点.

1 复杂网络的节点中心性度量

研究的复杂网络是无向无权图. 设复杂网络图 $G = \langle V, E \rangle$, $V = \{v_1, v_2, \dots, v_n\}$ 是图 G 的顶点集, $E = \{e_1, e_2, \dots, e_n\}$ 为图 G 的边集. 复杂网络的节点重要性可以通过节点的中心性度量,

节点中心性度量有以下方法:

度中心性(Degree Centrality)

$$C_d(v_i) = \text{Deg}(v_i)$$

其中 $\text{Deg}(v_i)$ 是顶点 v_i 的度数. 当节点 v_i 的度中心度的值越大, 这个节点和其他节点连接的边越多, 这个节点能影响的节点越多.

接近中心性(Closeness Centrality):

收稿日期: 2011 - 11 - 01

^{*} 国家自然科学基金项目资助(61171141)

$$C_c(v_i) = \frac{n-1}{\sum_{j=1}^n gd(v_i, v_j)}$$

其中 $gd(v_i, v_j)$ 是 v_i 到 v_j 的最短距离, n 是顶点的个数. 接近中心性说明节点对网络中其他节点的影响力. 例如在社会网络中, 一个节点和另一个节点有边相连, 说明这两个人有朋友关系. 因此, 如果一个人能够把信息在最短时间内传递给其他所有人(包括自己的朋友和朋友的的朋友), 则说明这个人在这群人中的影响力最大; 接近中心性与节点 v_i 传递信息到其他所有节点的用时总和成反比, 即接近中心性越大, 用时越小, 该节点的地位更高.

介数中心性(Betweenness Centrality)

$$C_b(v_i) = \sum_{j \neq i \neq k} \frac{\sigma_{jk}(v_i)}{\sigma_{jk}}$$

其中 σ_{jk} 是从 v_j 到 v_k 的最短路径的数量, $\sigma_{jk}(v_i)$ 是从 v_j 到 v_k 的最短路径中经过 v_i 的最短路径的数量. 中间中心性说明节点对网络中信息流动的影响力. 一个节点的中间中心性值越大, 通过该节点的最短路径的数目越大. 如在一个无向星型图中, 一个中心顶点的中间中心度为 1, 每一对节点间的最短路径都通过这个中心节点; 叶结点的中间中心度为 0, 不包含在任何一条最短路径中.

此外, 度量节点重要性的方法还有: 特征向量中心性等.

2 复杂网络的中心节点

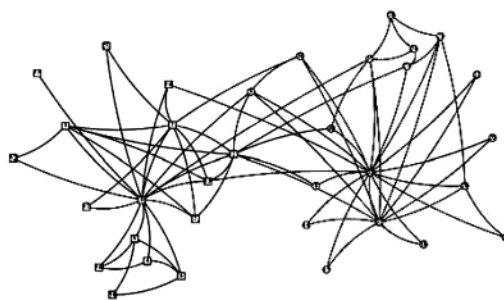
2.1 全局中心点

将所有节点按照顶点的中心性度量从大到小进行排序, 则前 k 个节点为全局中心点. 比如, 将所有节点按照度中心性从大到小进行排序, 前 k 个节点为全局度中心点.

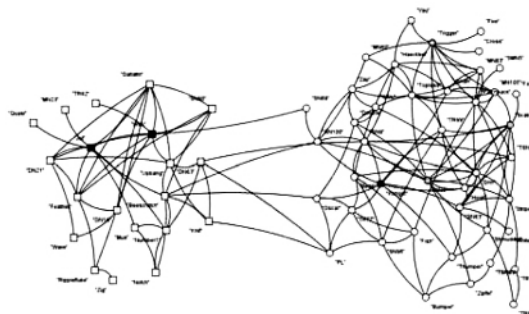
文献[2]指出在复杂网络的节点间的连接呈现两种情况: 同配性和异配性. 同配性是度数大的节点倾向于和度数大的节点连接, 异配性是度数大的节点倾向于和度数小的节点连接. 相关研究表明, 技术网络一般是异配的, 而社会网络通常是同配的. 因此, 如果图中度数大的节点大多和度数小的节点连接, 前 k 个度数大的全局中心节点在网络图上的分布较为分散; 如果图中度数大的节点大多和度数大的节点连接, 前 k 个度数大的全局中心节点在网络图上的分布较为集中; 所以, 异配性网络的全局中心点在网络图上

的分布较分散, 同配型网络的全局中心节点在网络图上的分布较为集中.

在一些真实的复杂网络图和合成复杂网络图上研究了全局中心节点的分布. 真实的复杂网络图包括: 空手道数据集、Political book 数据集、dolphins 数据集、Football 数据集. 前 k 个全局中心节点的分布见图 1. 从图 1 中可以看出, 空手道数据集的前 3 个全局中心点, 有两个是连接的, dolphins 数据集集中的前 5 个全局中心点(着绿色和黄色的节点)较为集中.



(a) 空手道网络



(b) Dolphins 网络

图 1 复杂网络图全局中心点的分布

注: 着黄色和绿色的节点是全局度中心点, 着红色和黄色的节点是局部度中心点.

进行实验的这些复杂网络具有社区结构. 复杂网络的社区结构是指整个网络是由若干个“群”构成的, 每个群内部的节点之间的连接较为紧密, 但是各个群之间的连接比较稀疏, 如图 1 所示. 不同形状的节点表示不同的社区. 从图上的全局度中心节点的分布可以看出, 在 Political book 数据集、dolphins 数据集上, 全局中心节点都集中在一个社区里, 另外一个社区没有全局中心节点. 在没有全局中心点的社区中的节点距离全局中心点较远, 它们受周围距离较近的节点的影响要比受全局中心节点的影响大.

2.2 局部中心点

在复杂网络图上,节点的中心性度量值存在最大值,也存在一些极大值,即存在一些节点,它们的中心性度量值大于等于它的所有邻居节点.这样的节点比它的所有邻居节点的都重要.如一个节点的度数比它的邻居节点的度数都大,则这个节点的影响力比它的邻居节点的影响力都大.这样的节点就是具有较大局部影响力的节点.把这种具有较大局部影响力的节点称为局部中心点.

定义 当一个节点的中心性度量值不小于它的所有邻居节点的中心性度量值时,称该节点为图的一个局部中心点.

当用度中心度来度量节点重要性时,局部中心点是局部度中心点,它的度数大于等于它的所有邻居节点的度数.局部度中心点也可称为极大度数顶点,因为和该节点邻接的节点的度数都不大于该节点的度数.图 2 中着红色和黄色的节点是局部度中心点.图 3 中着红色的节点是局部度中心点.从图 2 和图 3 可以看出,局部度中心点分散分布在复杂网络中.

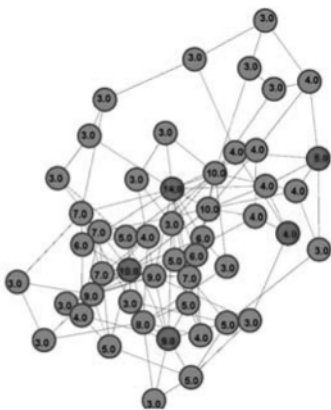


图 2 图的极大度数节点

局部中心点具有如下两个特性:

特性 1 在图 G 中,节点 v_i 和节点 v_j 是两个局部中心点, $C(v_i)$ 是节点 v_i 的中心性度量值,若 $C(v_i) \neq C(v_j)$,则节点 v_i 和节点 v_j 不相邻.

证明 用反证法证明.

假设节点 v_j 和节点 v_i 相邻.因为节点 v_i 是局部中心点, $C(v_i)$ 大于它的所有邻居节点的中心性度量值,因此 $C(v_i) \geq C(v_j)$. 因为 $C(v_i) \neq C(v_j)$,所以 $C(v_i) > C(v_j)$,根据局部中心点定义,节点 v_j 不是局部中心点.和节点 v_j 是局部中心点矛盾.

特性 2 复杂网络图的局部中心点的数目

是一确定值.

由特性 1,局部中心点分散分布在复杂网络中.如果两个局部中心点相邻,则它们的度数一定相同.按照节点中心性度量值排序得到的全局中心节点是可能集中分布,互相连接的,这时这些节点的影响力局限在这部分区域,用这些排序靠前的节点作为网络图的重要节点,实际上没有反映出网络图的其他部分的重要节点.局部中心点分散分布在复杂网络中,可以反映出复杂网络中各处的具有较大影响力的节点,所以局部中心点也是网络图中的重要节点.

对于异配性网络,度数大的节点之间很少连接,局部中心点通常也可能就是全局中心点,如空手道数据集上的两个局部度中心点也是前 2 个全局中心点.对于同配性网络,中心度值大的节点不一定是局部中心点.

根据节点中心性度量值排序,前 k 个全局中心点作为网络的重要节点,要找到分布在网络各处的重要节点, k 的合适取值较难确定.局部中心点分散分布在网络各处且数目一定,找到全部的局部中心点,就是分散在图中各处的重要节点.

同理,用接近中心性和中间中心性度量节点重要性时,就有局部接近中心点和局部介数中心点.

3 复杂网络的局部中心点的分布及和社区的关系

在 LFR 合成数据集^[9]上进行实验研究局部中心点的分布,分析局部中心点的分布和社区的关系.合成数据有 200 个节点,9 个社区.表 1 给出了局部中心点和全局中心点在 9 个社区中的分布情况,这里取和社区个数相同的全局中心点,即 $k=9$.实际上,真实网络的社区个数通常是未知的,所以 k 的取值较难确定.

根据对真实复杂网络(如图 2 所示)和合成复杂网络(如图 3 所示)的研究,可以看出,全局中心点的分布有时较为集中,如表中所示,全局中心点有些集中分布在一个社区,而局部中心节点较为分散分布在各个社区中.这里选取了 9 个全局中心点,实际上,很难预先知道多少个全局中心点就是分布在网络中各处的重要节点,如表 1 选取前 9 个全局中心点时,有 4 个社区没有全局度中心点,有 3 个社区没有介数中心点,有 5

个社区没有接近中心点. 局部介数中心点数和局部接近中心点数小于局部度中心点数, 局部度中心节点数最接近社区数. 深入研究局部中心节点和社区结构的关系对于社区结构发现将具有重要意义.

表 1 复杂网络图中心节点分布

社区 No.	度中心节点		介数中心性节点		接近中心性节点	
	全局	局部	全局	局部	全局	局部
0	3	0	3	0	3	1
1	0	1	0	1	0	0
2	1	1	1	1	0	1
3	0	1	1	0	1	0
4	1	1	1	1	0	1
5	3	2	2	1	3	0
6	0	1	0	0	0	0
7	1	0	1	0	2	1
8	0	2	0	1	0	0
总计	9	9	9	5	9	4

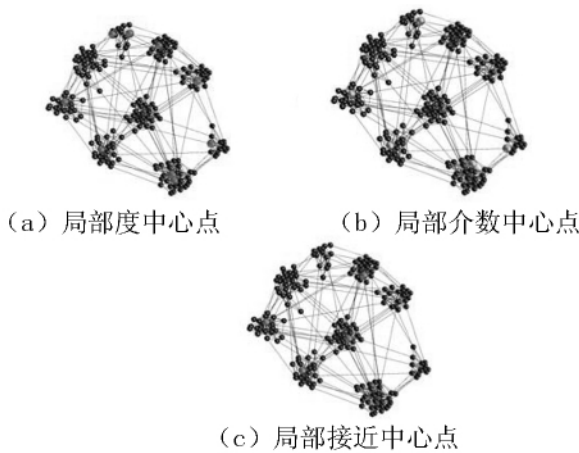


图 3 局部中心点在复杂网络中的分布

4 结束语

在复杂网络图中, 不同的节点的作用和的重要程度不同. 节点的中心性度量, 可以反映一个顶点在图中的作用以及对其他节点的影响程度. 按照节点的中心性度量值排序可以得到全局中心节点, 全局中心节点在图中可能是分散分布, 也可能是集中分布. 当前 k 个全局中心节点较为集中分布时, 它们可能不能很好地代表图的重要节点. 局部中心节点分散分布在图中, 且数目一定, 可以较好地作为图的重要节点. 在具有社区结构的复杂网络中, 局部中心点数最接近社区数, 对于研究社区结构具有重要意义.

参 考 文 献

[1] Rumi Ghosh , Kristina Lerman. Predicting Influential Users in Online Social Networks 2010 ,1005.4882v1 ,26 May .

[2] Newman M E J. Assortative mixing in networks. Phys. Rev. Ltee. ,2002 ,89(20) :208701.

[3] 安世虎,都艺兵,曲吉林. 节点集重要性测度. 中国管理科学 2006 ,14(1) .

[4] 李鹏翔,任玉晴,席西民. 网络节点(集) 重要性的一种度量指标. 系统工程 ,2004.

[5] 赫南,李德毅,涂文燕,等. 复杂网络中重要性节点发掘综述*. 计算机科学 2007 ,34(12) .

[6] Sabidussi G. T he centrality index of a graph. Psychometrika ,1966 ,31: 581 ~ 603.

[7] Freeman L. A set of measures of centrality based upon betweenness. Sociometry [J] ,1977 ,40: 35 -41.

[8] 王林、戴冠中. 复杂网络的度分布研究. 西北工业大学学报 2006 ,24(4) .

[9] Andrea Lancichinetti , Santo Fortunato , and Filippo Radicchi , Benchmark graphs for testing community detection algorithms. Physical Review E ,2008 ,78:046110.

Central Vertices Distribution in Complex Networks

Chen Qiong , Fang Ming , Chen Zhiyun
(South China University of Technology)

Abstract: The role and importance of different nodes in complex networks are different. Vertex centrality measure can reflect the role of nodes in the graph. Since the distribution of global central vertices is concentrated in some case , global central vertices cannot represent all important vertices of the entire network. Local central vertices are studied in this paper. The centrality measure of local central vertex is greater than or equal to those of all its neighboring nodes. Locating dispersedly in the networks , the local central vertices are important nodes with more influence to some vertices than the global central vertices.

Keywords: Complex Networks; Vertex centrality; Graph; Central vertex

(责任编辑: 李家云)