

一种基于互信息的复杂网络节点重要性评估方法

张翼¹ 刘玉华¹ 许凯华² 骆珍荣¹

(华中师范大学计算机科学系 武汉 430079)¹ (华中师范大学物理科学与技术学院 武汉 430079)²

摘要 在复杂网络中,如何量化节点的重要性是一个基本问题。首先阐述了现有的计算方法,在此基础上提出了一种基于互信息的节点重要性评估方法。该方法揭示了网络拓扑结构特性,准确反映了节点的相对重要程度。对该方法进行了实验论证,并与现有方法进行了分析比较,结果表明基于互信息的评估方法简单有效,特别适用于大型复杂网络节点重要性的评估。

关键词 复杂网络,节点重要性,互信息,拓扑结构,连接
中图分类号 TP393.01 文献标识码 A

Evaluation Method for Node Importance Based on Mutual Information in Complex Networks

ZHANG Yi¹ LIU Yu-hua¹ XU Kai-hua² LUO Zhen-rong¹

(Department of Computer Science, Huazhong Normal University, Wuhan 430079, China)¹

(College of Physical Science and Technology, Huazhong Normal University, Wuhan 430079, China)²

Abstract In complex networks, how to quantify the importance of the node is a basic problem. First elaborated some of the existing computational method, then proposed an evaluation method for node importance based on mutual information. This method precisely reveals topology characteristic of the network, reflecting the relative importance of nodes. The experiment on the comparison and the analysis with other methods shows that the evaluation method based on mutual information is simple and effective, it is suitable for evaluating the node importance in large-scale complex networks.

Keywords Complex networks, Node importance, Mutual information, Topology, Connection

1 引言

复杂网络,如因特网、电力网络、新陈代谢网络、科研合作网络等同我们的生活紧密相关,都可以用复杂网络理论描述^[1]。如果将真实世界中形形色色的个体用节点来表示,用边来代表个体之间的联系,许多事物都可以用网络图的理论来分析。近年来,复杂网络吸引了相关领域研究人员的广泛关注。大量研究表明,复杂网络既不是规则网络,也不是随机网络,它具有与这两者都不相同的特性:小世界特性^[2]、无标度特性^[3]、对于随机攻击的鲁棒性、对于恶意攻击的脆弱性等。随着复杂网络研究的深入,许多基础问题的探讨显得越发重要。评估节点的重要程度是复杂网络研究中的一个基本问题。在网络中发掘重要节点,对其重要性进行评估,具有很高的实用价值。比如定位恐怖组织头目从而打击恐怖组织、搜索引擎的搜索结果排序、疾病传播的控制、防止由相继故障引起的大规模停电、复杂网络社区结构中社区中心的确定等,这些都涉及到节点重要性评估计算问题。本文首先介绍了近年来复杂网络节点重要性的几种研究方法,在此基础上,提出了一种基于互信息的复杂网络节点重要性评估的方法。该方法揭示了网络拓扑结构特性,准确反映了节点的相对重要性。

实验结果表明,基于互信息的评估方法更简单有效,特别适用于大型复杂网络节点重要性的评估。

2 复杂网络中评估节点重要性的几种研究方法

评估复杂网络中节点重要性的方法很多,本质上都是基于图论的。最初的研究起源于社会学领域,随后其他学科的研究人员也开始研究这类问题。归纳起来,主要的研究方法如下^[4,9]:

(1) 社会网络分析方法:将节点的重要性等价于显著性。最著名的指标是度数和介数,节点的度数为网络中与该节点相连的边的数量,介数则反映了一个节点的影响力。在这类方法中,指标的研究不破坏网络的整体性,通过对网络的一些基本信息统计分析,进而量化节点的重要性,例如文献^[5]综合节点的全局和局部重要性评价节点重要程度。已有的指标还包括接近度、特征向量、网络直径等^[6]。这类研究方法从网络的结构特性出发研究,但方法得出的各种指标都有一定的局限性。

(2) 系统科学分析方法:移除某个节点对网络造成的破坏性等价于节点的重要性。移除节点对网络造成的破坏是多样的,例如文献^[7]提出一种基于全网平均等效最短路径数的网

到稿日期:2010-07-18 返修日期:2010-11-15 本文受华中师范大学中央高校基本业务费专项资金项目(2009043)资助。

张翼(1987-),男,硕士生,主要研究方向为复杂网络、无线传感器网络,E-mail:zhangyi19870328@sohu.com;刘玉华(1951-),女,博士,教授,博士生导师,主要研究方向为计算机网络与通信技术、无线网络、复杂网络等,E-mail:yhliu@mail.ccnu.edu.cn(通信作者)。

络抗毁度评价方法,它基于抗毁性评价节点重要性;文献[8]通过比较生成树数目的方法来评估节点重要性;文献[9]从节点移除后网络中连通节点对数目下降出发,来量化节点的重要程度。这种系统科学的分析方法并没有完全体现网络拓扑结构的差异,因此对于节点重要性的评估不是很准确。

(3)信息搜索领域分析方法:计算机科学家提出的一些算法考虑得更加全面。最著名的是 Google 的创始人 Larry Page 和 Sergey Brin 提出的网页排名算法 pagerank,它是搜索引擎 Google 的核心技术之一。pagerank 算法将文献检索中的引用理论用到 Web 中,引用网页的链接数,在一定程度上反映了该网页的重要性^[10,11]。每个到页面的链接都是对该页面的一次投票,被链接得越多,就意味着被其他网站投票越多。pagerank 算法能够在网络中准确定位节点的重要程度,并且其计算复杂度不高。

还有其它文献从不同角度来研究,如文献[12]提出在节点正常工作的情况下,收缩与该节点相连的边,收缩后得到的网络凝聚程度越高,则该节点越重要;文献[13]省略了大量迂回路由对可靠性影响的细节,只考虑迂回路由影响的效果,提出了一种跳面节点法。不同的方法从不同的角度来探讨同一问题,所以给出的指标没有好坏之分,每个指标都有自己的优缺点^[14]。

本文从网络拓扑结构出发,同时考虑节点之间连接的强弱和多少来量化节点的重要性。这种基于互信息的评估方法相对于其他一些方法更加准确,并且计算复杂度低。

3 基于互信息的节点重要性评估方法

将网络抽象为图 $G=(V,E)$,其中, V 是网络中所有点的集合, E 是网络中所有边的集合, (i,j) 代表节点 i 到节点 j 的边。香农在信息论中将信息定义为事物运动状态或存在方式的不确定性的描述,只有当信源发出消息通过信道传输给信宿后,才能消除不确定性并获得信息^[15]。对于任意一个节点,如果将它的 n 条边看作 n 条数据流,那么复杂网络可以看成是一个复杂的通信系统模型,因此,可以运用信息论的知识来评估节点的重要性。在这种基于互信息的评估方法中,每个节点所包含的信息代表了节点的重要程度,计算出这些节点的信息并进行比较,就可以得到节点之间的相对重要性。

3.1 定义

基于互信息的评估方法通过节点的信息量来评估节点的重要性,每个节点所包含的信息量由它的边决定。节点的信息量是这样计算的:先计算节点之间的互信息,即与该节点相连的每条边的信息量,然后计算该节点所包含的信息总量。

定义 1 设节点 i 到节点 j 的互信息为 $I(i,j)$,定义如下:

$$I(i,j) = \begin{cases} \log \frac{1}{p_{ij}} - \log \frac{1}{p_{ji}}, & i \text{ 与 } j \text{ 直接相连} \\ 0, & \text{其他} \end{cases} \quad (1)$$

式中, p_{ij} 为节点 i 的边 (i,j) 的概率。本文以无权无向网络为例,同一节点的每条边的概率分布视为等概率分布。所以:

$$I(i,j) = \begin{cases} \log k_i - \log k_j, & i \text{ 与 } j \text{ 直接相连} \\ 0, & \text{其他} \end{cases} \quad (2)$$

式中, k_i 为节点 i 的度。

定义 2 节点 i 的信息量是节点 i 与其他节点的互信息

之和,设为 $I(i)$:

$$I(i) = \sum_{j=0}^n I(i,j) \quad (3)$$

由以上定义可知, $I(i,j) + I(j,i) = 0$,对于整个网络来说,信息量之和 $\sum_{i=0}^n I(i)$ 为零。在计算出所有节点的信息量之后按照从大到小进行排序,信息量越大的节点重要性越强。

3.2 算法实现

根据以上定义,提出的计算方法中节点的重要性同时取决于节点之间连接的强弱和多少,因此,这种计算方法相对于其他一些方法更加简单与准确。

设网络为图 $G=(V,E)$,其中, V 是网络中所有点的集合; E 是网络中所有边的集合; n 为网络中的节点数; (i,j) 代表节点 i 到节点 j 的边。设矩阵 $A=[a_{ij}]$ 为网络 G 的邻接矩阵,其中:

$$a_{ij} = \begin{cases} 1, & \text{节点 } i \text{ 与节点 } j \text{ 有边相连} \\ 0, & \text{节点 } i \text{ 与节点 } j \text{ 无边相连} \end{cases}$$

Step1 输入网络 G 的邻接矩阵 A ,根据邻接矩阵 A 计算各节点的度数;

Step2 针对节点 i ,利用式(2)计算节点 i 与节点 j ($1 \leq j \leq n$) 的互信息 $I(i,j)$;

Step3 利用式(3)计算节点 i 的信息量;

Step4 如果 $i < n$,返回 Step2 计算节点 $i+1$ 的信息量;

Step5 根据计算出的各节点的信息量从大到小排序,即为节点重要性顺序。

从上述算法步骤看,Step2 和 Step3 是算法的主循环,因此整个节点重要性计算的时间复杂度取决于 Step2 节点间互信息的计算和 Step3 信息量的计算,得出算法的时间复杂度为 $O(n^2)$ 。下面以星形网络为例进行计算并说明以上计算过程,如图 1 所示。

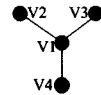


图 1 星形网络拓扑结构图

表 1 是按照以上互信息方法对星形网络进行计算的过程,表中右侧所示的信息量是由所在行的 4 个互信息相加得来。计算结果显示,星形网络的中心节点 v_1 的信息量最大,其它节点的信息量相等且都小于 v_1 ,因此 v_1 最重要,其它节点次之,并且同等重要。可见,提出的方法揭示了网络的拓扑结构特性,准确反映了节点的重要程度。在下一节中,将会针对更复杂的网络利用互信息的方法进一步分析说明,并与其的评估方法进行比较。

表 1 星形网络的计算过程

		节点间互信息 $I(i,j)$				节点的信息量 $I(i)$
$j \backslash i$	1	2	3	4		
1	0	1.099	1.099	1.099	3.296	
2	-1.099	0	0	0	-1.099	
3	-1.099	0	0	0	-1.099	
4	-1.099	0	0	0	-1.099	

4 计算实例与分析

利用 APRA 网络拓扑(如图 2 所示)来分析说明基于互

(下转第 109 页)

[4] Abbasi A A, Younis M. Movement-assisted Connectivity Restoration in Wireless Sensor and Actor Networks [J]. IEEE Transactions on Parallel and Distributed Systems, 2009, 20(9): 1366-1379

[5] Ferenets R, Lipping T. Comparison of Entropy and Complexity Measures for the Assessment of Depth of Sedation [J]. IEEE Transactions on Biomedical Engineering, 2006, 53(6): 1067-1077

[6] Rainer B, Dell'Amico M, Martello S. Assignment Problems [M]. SIAM, 2009

[7] Simulator-NS-2. TheNetwork [EB/OL]. <http://www.isi.edu/nsnam/ns/>

[8] Lopez-Nores M, Garcia-Duque J, Pazos J J. Qualitative Assessment of Approaches to Coordinate Activities of Mobile Hosts in Ad-hoc Networks [J]. IEEE Communications Magazine, 2008, 12: 108-111

(上接第 89 页)

信息的复杂网络节点重要性评估方法。APRA 网络拓扑是北美常用的干线拓扑, 由 21 个节点和 26 条边组成, 网络的平均度数为 2.48。去掉 APRA 网络中的任意一个节点, 网络仍然连通。表 2 是应用本文提出的互信息法与已有的度数法、介数法、生成树法以及 pagerank 算法计算出的 APRA 网络的节点重要程度的对照表。

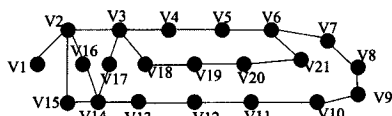


图 2 APRA 网络拓扑结构图

表 2 APRA 网络节点重要程度对照表

编号	度数法	介数法	生成树法	pagerank 法	互信息法
v ₁	2	0.000	0.6262	0.036	-1.099
v ₂	4	0.025	0.9721	0.067	1.674
v ₃	4	0.030	0.9930	0.070	2.079
v ₄	2	0.005	0.8387	0.040	-0.693
v ₅	2	0.005	0.8387	0.041	-0.405
v ₆	3	0.015	0.9836	0.061	1.216
v ₇	2	0.005	0.8797	0.043	-0.405
v ₈	2	0.005	0.8797	0.045	0.000
v ₉	2	0.005	0.8797	0.045	0.000
v ₁₀	2	0.005	0.8797	0.044	0.000
v ₁₁	2	0.005	0.8797	0.042	-0.405
v ₁₂	3	0.015	0.9780	0.057	0.811
v ₁₃	2	0.005	0.8051	0.038	-1.099
v ₁₄	4	0.030	0.9864	0.069	2.367
v ₁₅	3	0.010	0.8787	0.051	-0.170
v ₁₆	2	0.005	0.6639	0.036	-1.386
v ₁₇	2	0.005	0.6977	0.037	-1.386
v ₁₈	2	0.005	0.7701	0.038	-1.099
v ₁₉	3	0.015	0.9671	0.057	0.811
v ₂₀	2	0.005	0.8279	0.041	-0.405
v ₂₁	2	0.005	0.8279	0.042	-0.405

从表 2 可以看出, 提出的基于互信息的评估算法得出的最重要节点是 v₁₄, 而其它的几种方法得出的结果是 v₃ 或 v₁₄。从图 2 中可见, v₃, v₁₄ 两个节点局部的拓扑结构很相似, 它们都是网络中度数最大的节点。不同方法得出不同结果的原因可能是不同方法关注的角度不同。本文提出的基于互信息的评估方法与度的相关性较强, 但是与度数法相比, 本文的方法不仅考虑了局部网络的结构特性, 也考虑了全局网络的结构特性。另外在度数法、介数法、生成树法中, v₇, v₈, v₉, v₁₀, v₁₁ 这 5 个节点的重要性相同。而在本文提出的方法中, 节点 v₈, v₉, v₁₀ 相对于 v₇, v₁₁ 更为重要。观察可以发现, 移除 v₇ 节点后, v₁₂ 节点的负载加重; 而移除 v₉ 节点可能同时造成 v₆ 节点和 v₁₂ 节点的拥塞, 对网络造成的潜在破坏更大。以

上分析表明, 从本文提出的互信息方法求出的结果中, 可以明确地分辨出这些差别。

结束语 本文提出的基于互信息的复杂网络节点重要性评估方法是一种简单准确的评估方法, 而且计算速度快, 特别适用于大型复杂网络节点重要性的计算。实验结果表明, 本方法准确反映了节点重要程度的差异, 可以进一步推广到有向加权网络中。但是本方法只是从网络的结构特性出发, 并未考虑节点的异构性等其它因素, 所以需要更深一步的研究。

参考文献

[1] 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

[2] Watts D J, Strogatz S H. Collective Dynamics of Small-world Networks [J]. Nature, 1998, 440-442

[3] Albert R, Barabási A L. Emergence of Scaling in Random Networks [J]. Science, 1999, 286: 509-512

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

[5] 陈静, 孙林夫. 复杂网络中节点重要度评估 [J]. 西南交通大学学报, 2009, 44(3)

[6] Enrico N, Guido P, Peter W. Finding the Most Vital Node of a Shortest Path [J]. Theoretical Computer Science, 2003, 29(6)

[7] 饶育萍, 林竞羽, 周东方. 网络抗毁度和节点重要性评价方法 [J]. 计算机工程, 2009, 35(6)

[8] 陈勇, 胡爱群, 胡啸. 通信网中节点重要性的评价方法 [J]. 通信学报, 2004, 25(8)

[9] 李鹏翔, 任玉晴, 席西民. 网络节点(集)重要性的一种度量指标 [J]. 系统工程, 2004, 22(4)

[10] Larry P, Sergey B. The PageRank Citation Ranking: Bringing Order to the Web [C] // Stanford Digital Libraries Working Paper. 1998

[11] Zhang Yi, Liu Yu-hua, et al. Modeling of Scale-free Network Based on Pagerank Algorithm [C] // ICFC 2010. IEEE Computer Society, Wuhan, China, May 2010; V3-778-782

[12] 谭跃进, 吴俊, 邓宏钟. 复杂网络中节点重要度评估的节点收缩方法 [J]. 系统工程理论与实践, 2006, 11(11)

[13] 郭伟. 野战地域通信网可靠性的评价方法 [J]. 电子学报, 2000, 28(1)

[14] Wasserman S, Faust K. Social network analysis: methods and applications [M]. Cambridge: Cambridge University Press, 1994: 218

[15] 傅祖芸. 信息论——基础理论与应用 [M]. 北京: 电子工业出版社, 2001