

# 树形网络的平均介数\*

霍慧鸣, 赵海兴

(青海师范大学 计算机学院, 青海 西宁 810008)

**摘要:** 主要讨论树形网络的平均节点介数, 刻画了树形网络中具有最大和最小平均节点介数的网络。

**关键词:** 介数; 无圈网; 复杂网络

中图分类号: TP391

文献标识码: A

文章编号: 1674-7720(2014)03-0055-03

## Average betweenness of tree networks

Huo Huiming, Zhao Haixing

(School of Computer Science, Qinghai Normal University, Xining 810008, China)

**Abstract:** This paper mainly discusses the average nodes betweenness in tree network. This paper characterizes the trees with the maximum and minimum average betweenness.

**Key words:** betweenness; tree network; complex network

自从 20 世纪 30 年代 Radcliffe Brown 提出社会网络分析概念以来, 社会网络分析已经逐步成为一种重要的社会定量研究方法<sup>[1]</sup>。近年来, 随着计算机技术和互联网的飞速发展, 如何利用文本挖掘、机器学习等技术面向互联网进行社会网络挖掘和分析成为了一个新的研究课题。随着数据挖掘技术的进步以及计算机处理能力的提高, 研究者们分析的社会网络的规模也急剧增大。这些分析和其他现实网络分析一样, 面临网络节点规模巨大的挑战。

现实生活中存在的大量复杂系统都可以通过不同的网络加以描述, 网络科学已成为众多学科汇聚的交叉点。复杂网络是近几年科学研究发现的一种介于规则网络和随机网络之间的一种更接近于真实网络的网络模型, 钱学森给出了复杂网络的一个较严格的定义: 具有自组织、自相似、吸引子、小世界、无标度(无尺度)中部分或全部性质的网络称为复杂网络<sup>[2]</sup>。随着复杂网络规模的不断扩大, 对于各个节点在整个图中所起作用的研究变得迫切, 各个节点的作用可以用介数来表示。介数既能刻画图中节点和边的重要性, 揭示网络层次结构, 又能用来构建基于点介数或边介数的聚类算法, 发现图

中特殊群体, 因此, 它一直是研究网络结构性质的一个重要量化手段。首先对于介数的概念要有一定的了解。在复杂网络研究中, 研究者不仅要非常客观地关注系统内个体之间的相互作用, 而且还要注视系统的整体相互作用, 表达这种整体相互作用的概念是介数。介数是一个全局的变量, 反映节点或边的作用和影响力, 可分为节点介数(Vertex Betweenness)和边介数(Edge Betweenness)两种。节点介数定义为在网络的所有最短路径中, 通过节点的最短路径条数称之为节点  $i$  的介数<sup>[3]</sup>; 边的介数定义为在网络的所有最短径经过边的介数; 网络平均节点介数定义为网络中所有节点介数的平均值; 网络平均边介数指网络中所有边介数的平均值。节点的介数在一定程度上反映的是节点在网络中的核心作用, 节点的介数越大, 表明节点的核心作用越强, 删除这样的节点会使大量节点对之间的最短路径变长, 边的介数也有同样的性质。

本文主要讨论树形网络的平均节点介数, 其中节点介数是经过该节点的所有路径, 网络平均节点介数是指所有节点介数的平均值。下面举一个具体的例子来分析节点介数和边介数。图 1 所示中各节点与边的介数可以分别表示如下。

表 1 中的平均节点介数为 4.375, 平均边介数为 7.25,

\* 基金项目: 科技部 973 专项(2010CB334708); 国家自然科学基金项目(61164005); 教育部长江学者与创新团队支持计划(IRT1068); 青海省自然科学基金项目(2012-Z-943)

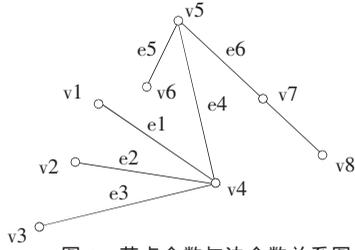


图1 节点介数与边介数关系图

依据 Newman<sup>[4]</sup>提出的介数新的计算机方法和 Brandes算法可以更快地计算出介数。更多的介绍见参考文献[5-7]。

表1 各节点介数与边介数的关系

节点	V1	V2	V3	V4	V5	V6	V7	V8
节点介数	0	0	0	15	14	0	6	0
边	e1	e2	e3	e4	e5	e6	e7	
边介数	60	6	6	16	6	12	6	

本文研究了树形网络中平均节点介数最大的图和最小的图,所有树的平均节点介数小于路的平均节点介数和其大于星形图的节点平均介数。

复杂网络树形图中的介数源于网络中对个体重要性的评估,一个节点介数表示所有节点对之间通过该节点的最短路径条数,因此它能够刻画节点在网络中的重要程度,任意节点  $v$  的介数值定义为<sup>[8]</sup>:

$$C_B(v) = \sum_{s \neq t \neq v} \delta_{st}(v), \delta_{st}(v) = \frac{\delta_{st}(v)}{\delta_{st}}$$

其中,  $\delta_{st}$  为节点  $s$  到  $t$  的最短路径数目,  $\delta_{st}(v)$  为节点  $s$  到  $t$  的最短路径中经过节点  $v$  的最短路径数目。

树形网络中节点介数就是通过该节点的路径数目。如有一个  $n$  个节点的树形网络  $N$ , 节点用  $v$  表示, 平均节点介数可表示为:

$$B(N) = \sum_v B(v) / n$$

### 1 树形网络中最大平均节点介数图

随着复杂网络规模的不断增加, 节点数目也在不断地增加, 对于一些冗余的节点删除变得更加必要, 这就需要对于节点的作用给出一个正确的评估, 需要了解哪些图形具有最大的平均节点介数, 网络结构可以尽可能地向这样的结构变化。本节主要介绍具有最大平均节点介数的网络。

引理 1 如图 2 所示,  $N_1$  是一棵树,  $N_1'$  是经过图 2 的变换得到的, 则等号成立当且仅当  $N_1$  是一条路。

证明 有  $k$  个节点的图, 如图 2 所示, 图  $N_1$  中节点  $A$  上有  $p$  个分支, 有  $n$  个节点的为分支  $P_1$ , 有  $m$  个节点的为分支  $P_2$ 。图  $N_1$  把分支  $P_1$  添加到分支  $P_2$  之后得到图的  $N_1'$  分支  $P_1+P_2$ 。则节点  $A$  上变成了  $p-1$  个分支。因为分支节点介数与分支结构和总节点个数有关, 现在树形图  $N_1$  和  $N_1'$  的平均节点介数  $B(N_1)$  和  $B(N_1')$  中, 图  $N_1$  和  $N_1'$  中除了节点  $A$ , 分支  $P_1$ 、 $P_2$  及分支  $P_1+P_2$  的节点

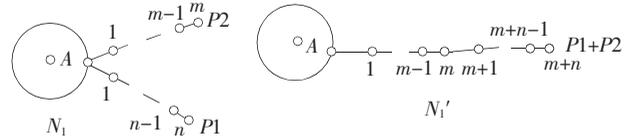


图2  $N_1$  与  $N_1'$  的关系图

介数有变化, 其他分支对应节点介数都没有改变。把其他分支节点的介数之和用  $M$  表示。

图  $N_1$  上节点介数分别表示如下:

节点  $A$  的介数为  $\sum_{i=1}^{p-1} \sum_{j=i+1}^p P_i P_j$ , 这里的  $P_i$  和  $P_j$  表示  $p$  个分支中第  $i$  和  $j$  分支的节点个数。分支  $P_1$  及  $P_2$  各节点介数如表 2、表 3 所示。

表 2 分支  $P_1$  各节点介数

1	2	...	$n-1$	$n$
0	$k-2$	...	$(n-2)(k-n-1)$	$(n-1)(k-n)$

表 3 分支  $P_2$  各节点介数

1	2	...	$m-1$	$m$
0	$k-2$	...	$(m-2)(k-m-1)$	$(m-1)(k-m)$

平均节点介数为:

$$B(N_1) = (M + A + P_1 + P_2) / n = (M + \sum_{i=1}^{p-1} \sum_{j=i+1}^p P_i P_j + \sum_{i=1}^m (i-1)(k-i) + \sum_{i=1}^n (i-1)(k-i)) / n$$

图  $N_1'$  中节点介数分别表示如下:

节点  $A$  的介数为  $\sum_{i=1}^{p-3} \sum_{j=i+1}^{p-2} P_i P_j + \sum_{i=1}^{p-2} P_i (m+n)$ 。分支  $P_1+P_2$  的介数如表 4 所示。

表 4 分支  $P_1+P_2$  的介数

1	2	...	$m+n-1$	$m+n$
0	$k-2$	...	$(m+n-2)(k-m-n-1)$	$(m+n-1)(k-m-n)$

平均节点介数可以表示为:

$$B(N_1') = (M + A + P_1 + P_2) / n = (M + \sum_{i=1}^{p-3} \sum_{j=i+1}^{p-2} P_i P_j + \sum_{i=1}^{p-2} P_i (m+n) + \sum_{i=1}^{m+n} (i-1)(k-i)) / n$$

将变化前后的两个图的平均介数进行比较, 比较结果为:

$$B(N_1') - B(N_1) = mn(k-m-n-1) \quad (1)$$

由式(1)知,  $k \geq m+n+1$ ,  $B(N_1) \leq B(N_1')$ , 等号成立当且仅当  $N_1$  是一条路时, 故可得  $B(N_1) \leq B(N_1')$ , 证毕。

路是各个节点的度不大于 2 的连通的无圈图, 两端的两个节点的度为 1, 中间各个节点的度均为 2。  $n$  个节点的路中两端的两个节点为 1 度的叶子节点, 其介数为 0。假如第  $i$  个节点的介数,  $i$  的取值范围是  $1 \sim n$ , 各个节点的介数可以表示为  $(n-i)(i-1)$ , 其平均介数为:

$$\frac{\sum_{i=1}^n (n-i)(i-1)}{n} = \frac{n^3-3n^2+2n}{6n} = \frac{n^2-3n+2}{6} \quad (2)$$

定理 1 路是具有最大平均节点介数的树形图,  $n$  个节点的平均介数为:  $\frac{n^2-3n+2}{6}$ 。

证明 用归纳法证明如下。

(1) 当  $n=4$  时, 有 4 个节点的图只有两种形式, 即路和星形图, 路的平均节点介数为 1, 星形图的平均节点介数为 0.75, 路具有较大的平均节点介数, 成立。

(2) 假如对于  $n < k$  时, 路的平均节点介数最大。当  $n=k$  时, 对于有  $k$  个节点的树形图, 选择一个节点度最大的点作为 Hub 节点, 由于各个分支只和它们自身结构与节点总个数有关, 各个分支的节点个数都是小于  $k$  的。由于  $n < k$  时, 路的平均节点介数最大成立, 因此可以以 Hub 点为中心把各个分支都变化成路的形式。整个图就变成如图 3 所示, Hub 节点的介数没有改变, 各个分支的节点介数和变成了最大, 由引理 1 可知, 把一个分支加到另一个分支上, 其总的平均介数就会变大, 如此一来, 逐渐变成路的形式平均节点介数变成最大。故可以证明, 当有  $k$  个节点时, 路的平均节点介数最大, 成立。

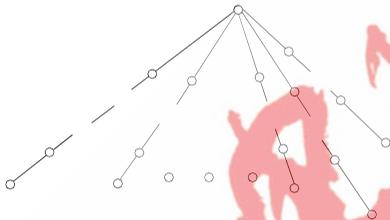


图 3  $n < k$  时以 Hub 节点为中心, 把各分支变化成路后的图

综上所述, 在树形图中, 路具有最大的平均节点介数, 证毕。

## 2 树形网络中具有最小平均节点介数的图

本节主要研究树形网络的结构中具有最小的平均节点介数。对此可以知道哪些结构的网络中节点的相对作用较小, 删除这样的节点的分支, 对于整个的网络影响相对不大。

引理 2 如图 4 所示,  $N_2$  所示是一棵树,  $N_2'$  是由  $N_2$  把图 4 所示的 1 度点分支添加到节点  $A$  得来的, 则  $B(N_2) \geq B(N_2')$ , 等式成立当且仅当  $N_2$  是一个星形图。

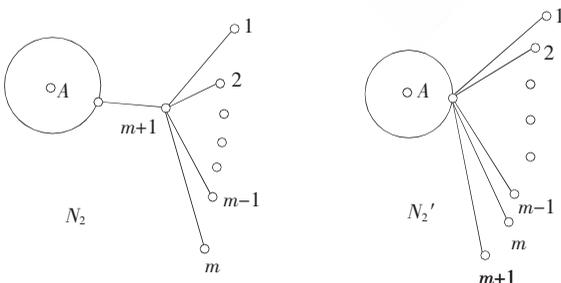


图 4 树形网络中节点分析图

证明 假如对于图 4 所示的  $N_2$  中节点  $A$  有  $p$  个分支,  $k$  个节点的图, 图中有一个  $m$  个 1 度点的分支。图  $N_2'$  把图  $N_2$  中  $m$  个 1 度的点添加到节点  $A$  上, 分支变成了 1 度点。节点  $A$  的分支个数变成了  $p+m$  个。图  $N_2$  和  $N_2'$  中的平均节点介数用  $B(N_2)$  和  $B(N_2')$  表示。  $N_2$  和  $N_2'$  中  $m$  个 1 度点的介数仍然为零, 没有改变。其中, 除了节点  $A$ 、节点  $m+1$  有变化, 其他  $p-1$  个分支对应节点介数都没有改变。未改变的分支节点介数和用  $M$  表示。

图  $N_2$  改变的节点介数和为:

$$B(N_2) = M + \sum_{i=1}^{p-2} \sum_{j=i+1}^{p-1} P_i P_j + \sum_{i=1}^{p-1} P_i (m+1) + \binom{m}{2} + m(k-m-1)$$

图  $N_2'$  的节点介数和为:

$$B(N_2') = M + \sum_{i=1}^{p-2} \sum_{j=i+1}^{p-1} P_i P_j + \sum_{i=1}^{p-1} P_i (m+1) + \binom{m+1}{2}$$

用  $B(N_2)$  减去  $B(N_2')$  等式可得:

$$B(N_2) - B(N_2') = M + \sum_{i=1}^{p-2} \sum_{j=i+1}^{p-1} P_i P_j + \sum_{i=1}^{p-1} P_i (m+1) + 2^{m+1} - M - \sum_{i=1}^{p-1} \sum_{j=i+1}^p P_i P_j - \sum_{i=1}^{p-1} P_i (m+1) - 2^m + m(k-m-1) = m(m+2-k) \quad (3)$$

由式(3)可知,  $k \geq m+2$  时, 有  $B(N_2) \geq B(N_2')$  成立, 等号成立当且仅当是一个星形图时。其他情况都是  $k > m+2$ , 故引理可证, 证毕。

星形图是只有一个节点的度不为 1, 其他节点的度都为 1 的图。星形图中, 度为 1 的节点的介数都是 0, 在星形图中, 称这个有最大度的节点为 Hub 节点, 也只有这么一个节点的介数不是 0。假如有  $n$  个节点的星形图, 其中有  $n-1$  个节点的介数为 0, Hub 节点的介数为:  $C_{n-1}^2$ 。故有  $n$  个节点的星形图的平均节点介数为:  $C_{n-1}^2/n = \frac{n^3-3n^2+2}{2n}$ 。

定理 2 星形图具有最小的节点平均节点介数,  $n$  个节点的星形图的平均节点介数为:  $\frac{n^3-3n^2+2}{2n}$ 。

证明 用归纳法证明如下。

(1) 当  $n=4$  时, 有 4 个节点的图只有两种形式, 即路和星形图, 路的节点平均介数为 1, 星形图的节点平均介数为 0.75, 星形图有最小平均节点介数, 成立。

(2) 假设当  $n < k$  时, 星形图有最小的平均节点介数成立。当  $n=k$  时, 在树形图中, 选择一个度最大的点作为 Hub 节点, 各个分支上选择与 Hub 节点直接相连的点作为分支的根节点, 各个分支的变化情况不影响 Hub 节点的介数和其他分支上节点的介数, Hub 节点的介数和分支个数与分支的节点数有关, 各分支的节点数不改变, 只要分支不增加或者减少, 则 Hub 节点的介数不变, 各个分支上的节点不受其他分支变化的影响。因为当  $n < k$

时,星形图有最小的节点平均介数,则各个分支都变为以分支根节点为中心的星形,各个分支有最小平均节点介数,Hub节点的介数不变。最终会有图5所示的图形,这种图形与原图相比,Hub节点的介数没有改变,各个分支节点介数变成了最小。根据引理2可知,把分支上的所有1度点添加到Hub节点上,其总的节点介数和减少,故逐渐把1度节点添加到Hub节点上,其节点介数和一直在减少,直到Hub节点上连接的都是1度点,总的节点介数和无法再减少。可知星形图具有最小的总节点介数和,成立。

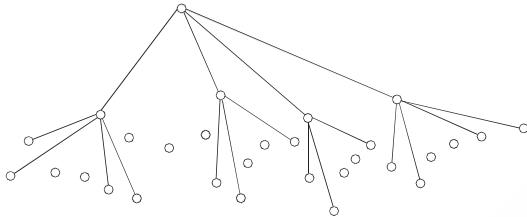


图5 树形网络具有最小平均节点介数的图

综上所述,星形图具有最小的平均节点介数,证毕。

本文主要讨论了在复杂网络中树形网络的平均节点介数最大和最小的网络,对于与之相关节点介数增加和减少给出了两个引理,节点介数是通过该节点的路径数。更多的介数应用和算法分析可以参考文献[9]~[11]。

#### 参考文献

- [1] SCOT J. Social network analysis: a handbook (2<sup>nd</sup> ed)[M]. Sage Publications, 1991.
- [2] 张嗣瀛. 复杂系统与复杂性科学[EB/OL].<http://news.qdu.edu.cn/newx?newsid=1514>, 2006-05-05.
- [3] 张晓林,袁莉,杨峰,等.基于Web的个性化信息服务机

制[J].现代图书情报技术,2001(1):25-29.

- [4] NEWMAN M E J. Scientific collaboration networks.II. shortest paths, weighted networks, and centrality[J]. Phys. Rev.E, 2001, 64(1).
- [5] LEWIS T G. 网络科学原理与应用[M].陈向阳,巨修红译.北京:机械工业出版社,2011.
- [6] 马杰良,邢雪,安莉莉.基于科研合作网络的节点枢纽特性研究[J].东北电力大学学报,2008,28(2):72-76.
- [7] Zhang Z Z, Zhou S G, Y Qi, et al. Topologies and laplacian spectra of a deterministic uniform recursive tree[J]. Eur. Phys. J.B, 2008(63):507-513.
- [8] 唐晋韬,王挺.复杂社会网络的介数性质近似计算方法研究[J].计算机工程与科学,2008,30(12):9-14.
- [9] 王小光,王锋,李森.基于介数影响矩阵的通信网络节点重要度评价方法[J].空军工程大学学报,2012,13(5):80-84.
- [10] 何宇,赵洪利,姚曜,等.介数中心性和平均最短路径长度整合近似算法[J].复杂系统与复杂性科学,2011,8(3):44-53.
- [11] 贾炜,冯登国,连一峰.基于网络中心性的计算机网络脆弱性评估方法[J].中国科学院研究生学报,2012,29(4):529-535.

(收稿日期:2013-10-10)

#### 作者简介:

霍慧鸣,女,1986年生,硕士研究生,主要研究方向:复杂网络。

赵海兴,男,1969年生,博士,教授,博士生导师,主要研究方向:复杂网络。