

# 复杂网络中重要性节点发掘综述\*

赫南<sup>1</sup> 李德毅<sup>2</sup> 淦文燕<sup>3</sup> 朱熙<sup>1</sup>

(北京航空航天大学计算机系 北京 100083)<sup>1</sup> (电子系统工程研究所 北京 100039)<sup>2</sup>

(清华大学计算机系 北京 100084)<sup>3</sup>

**摘要** 发掘网络中重要性<sup>1</sup>节点(边)一直是图论领域的一个基本问题。随着近年来复杂网络研究热潮的兴起,特别是很多实际网络所抽象出来的复杂网络,表现出了与以往图理论不同的特性,如小世界特性、无尺度特性等。如何在复杂网络环境下,发掘重要性节点已经成为复杂网络研究的一个基本问题。本文简要介绍了复杂网络的基本概念,详细总结、分析了在复杂网络环境下几个领域中发掘重要性节点的方法,最后提出了这一领域内几个有待深入研究的问题和可能的应用方向。

**关键词** 复杂网络,节点重要性,图,相对重要性

## Mining Vital Nodes in Complex Networks

HE Nan<sup>1</sup> LI De-Yi<sup>2</sup> GAN Wen-Yan<sup>3</sup> ZHU Xi<sup>1</sup>

(Department of Computer Science, Beijing University of Aeronautics and Astronautics, Beijing 100083)<sup>1</sup>

(Institute of Electronic System Engineering, Beijing 100039)<sup>2</sup>

(Department of Computer Science & Technology, Tsinghua University, Beijing 100084)<sup>3</sup>

**Abstract** Mining vital nodes/edges has always been a basic problem in Graph Theory. As the study of Complex Networks has become more and more hot, especially that there are lots of real-world networks that can be modeled as Complex Networks, which also have many important properties different from those in the traditional Graph Theory, e. g. small-world property, scale-free property, etc. How to mine vital nodes in Complex Networks has become a new-born problem. This paper reviews some basic concepts of Complex Networks, summaries and analyses some classic measures of mining vital nodes in some fields, and predicts some further-studied questions and possible applications.

**Keywords** Complex networks, Vital nodes, Graph, Relative importance

### 1 引言:无处不在的复杂网络

自然界中存在的大量复杂系统都可以通过形形色色的网络加以描述,网络是由许多节点与连接节点之间的边组成的。其中,节点代表真实系统中不同的个体,而边则代表这些个体之间的联系。通常,当两个节点之间具有某种特定的关系则

连一条边,反之则不连边,有边相连的两个节点在网络中被看作是相邻的。数学在研究网络的时候,只关心节点之间有没有边相连,至于节点在什么位置、边是什么形状等都是不需要关心的。通常把不依赖于节点的具体位置和边的具体形态所表现出来的网络结构叫做网络拓扑<sup>[1]</sup>。

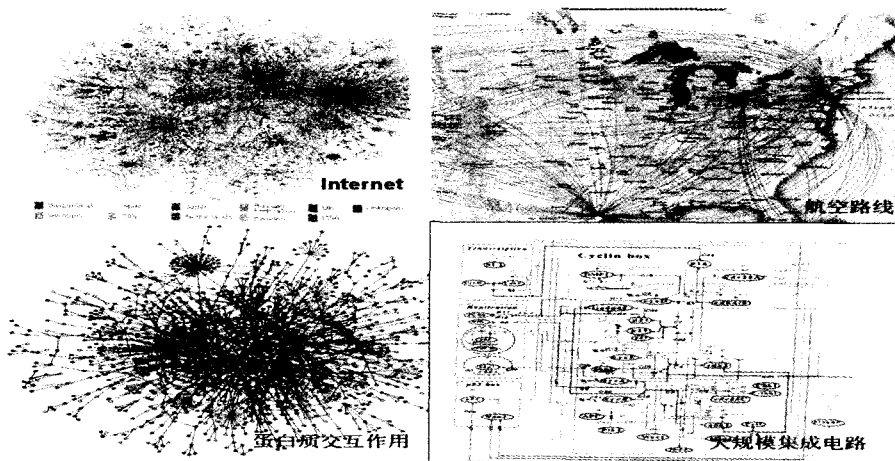


图1 复杂网络实例

\*国家自然科学基金(60496323,60375016),现代设计大型应用软件的共性基础973计划(2004CB719401)。赫南 硕士,研究方向:不确定性人工智能、复杂网络、网络化数据挖掘。

<sup>1</sup>节点的重要性根据所抽象的不同网络具有不同的含义,例如核心性、声望、吞吐量等。这里用重要性表达一种通用的定性概念。

复杂网络(Complex Networks)并没有精确严格的定义,它可以看作是大量真实复杂系统的拓扑抽象。它既不是规则网络,也不是随机网络,而是具有与两者皆不相同的统计特征的网络。复杂网络在现实世界中无处不在,如社会领域的社会关系网络<sup>[2]</sup>、科学家合作者网络<sup>[3]</sup>;信息领域的 WWW 网络<sup>[4]</sup>、论文引用网络<sup>[5]</sup>;技术领域的自治层 Internet<sup>[6]</sup>、软件类网络<sup>[7]</sup>;生物领域的食物链网络<sup>[8]</sup>、新陈代谢网络<sup>[9]</sup>、蛋白质相互作用网络<sup>[10]</sup>等。图 1 为一些复杂网络的实例。

结构决定功能是系统科学的基本观点<sup>[11]</sup>。近年来的大量研究表明,网络的拓扑结构决定了网络所拥有的特性。复杂网络表现出了与经典的随机图模型理论<sup>[12]</sup>不同的特性,其中比较重要的特性有:

#### • 小世界(Small World)

1998年,Strogatz和Watts在“Nature”上发表文章,引入了小世界模型<sup>[13]</sup>,用来描述完全规则网络到完全随机网络的转变。小世界网络既有与规则网络类似的聚集特性,又具有与随机网络类似的较小的平均路径长度。小世界模型反映了朋友关系网络的一种特性,即大部分人的朋友都是他们日常生活中经常接触的。另外,也有一些朋友离得很远,甚至异国他乡。正是由于这种不同朋友圈之间的长程连接,导致了网络具有较小的平均路径长度。著名的“六度分离(Six degrees of separation)<sup>[14]</sup>”实验,在某种程度上反映了人际关系的“小世界”特征。

#### • 无尺度(Scale-Free)

1999年Barabasi和Albert在“Science”上发表文章,指出许多现实复杂网络,包括Internet、WWW以及新陈代谢网络等,它们的度分布函数都具有幂律形式<sup>[15]</sup>。这种幂律分布的形式与分形(fractal)中的自相似(self-similarity)特点一样,没有明显的特征长度,所以该类网络称为无尺度网络。幂律分布的形成,主要是由于节点的偏附依好(preferential attachment),即新加入网络的节点倾向于与那些具有较大度的节点相连,这种现象也称为“马太效应(Matthew effect)”,主要反映了人们的从众心理。

#### • 鲁棒性和脆弱性(Robust and fragile)<sup>[16]</sup>

对网络中节点进行攻击,主要有两种策略:一是随机攻击;二是蓄意攻击某些重要节点。复杂网络由于度分布的不均匀性,导致面对不同的攻击策略,表现出不同的鲁棒性、抗毁性。面对随机攻击,即使多数节点失效、崩溃,整个网络仍然可以保持连通性。但是面对针对其中重要节点(比如,度很大的节点)的蓄意攻击,则网络就变得十分脆弱,几个“大”节点的失效就可能整个网络的崩溃。

#### • 社区结构(Community Structure)

许多现实网络中都存在这样的团体,团体内部的节点之间连接度较高,而团体之间的节点相对只有少量连接。2002年,Newman提出了复杂网络的社区结构<sup>[17]</sup>的概念,社区就是网络中节点的集合,社区中的节点之间具有紧密的连接,而社区之间则为松散的连接。它体现了复杂系统的层次和模块结构。

近年来复杂网络理论的研究越来越热,也开创了很多其他的相关分支,但是一些基本问题仍然值得我们去深入地研究。可视化技术的应用,能够定性地表示出小型网络的层次

化结构特征,可以直观地显示哪些节点是相连接的。同样,对网络的属性进行定量的刻画也是非常重要的,尤其是当网络的规模和复杂度非常大的时候,使用图形的表示来分析问题就非常困难了。在各种复杂网络中,用定量分析的方法寻找超大规模网络中哪个节点(边)最重要,或者某个节点相对于其他一个或多个节点的重要程度,是复杂网络研究中的一个基本问题。

复杂网络本质上的非同质拓扑结构<sup>[18]</sup>,决定了网络中每个节点的重要程度是不同的。在复杂网络的各种基础研究工作中,对网络中节点的重要性进行评估,发掘网络中的重要节点,具有重要的实用价值。尤其对各种各样具体的网络,更可以有针对性地分析其性质,制定正确的策略和措施。例如:

1) 在罪犯关系网络中,可以迅速定位犯罪团伙的头目,集中警力进行布控<sup>[19]</sup>;

2) 在电力网络中,对重要的断路器、发电单元等进行保护,可以有效防止由相继故障引起的大范围停电<sup>2</sup>;

3) 在搜索引擎的应用方面,可以把搜索到的正确结果根据其匹配和重要程度排序后返回给用户;

4) 在大规模计算机网络中,可以根据服务器节点的重要程度进行有针对性的备份,既能保证网络的鲁棒性,又能有效地节省资源;

5) 在传染病、病毒网络中,可以有针对性地先治疗、隔离病源,有效防止病毒的传播和扩散;

6) 在谣言传播网络<sup>[20]</sup>中,可以发掘出始作俑者,避免“蝴蝶效应”;

7) 在发掘复杂网络中的社区<sup>[21]</sup>时,也可以通过发现重要节点来确定社区的中心等。

## 2 发掘复杂网络中重要节点的几种方法

评估网络中节点重要性的方法有很多,本质上都是源于图论<sup>[22]</sup>以及基于图的数据挖掘<sup>[23]</sup>。复杂网络中节点重要性的研究起源于社会学网络分析领域<sup>[24]</sup>,Freeman等学者在早期针对社会学网络做了大量的研究,此后系统科学研究领域、信息搜索领域和文献检索领域等也分别独立地提出了类似的问题,发掘网络中重要的节点已经成为复杂网络各研究领域的一个基本问题。如下结构是,第2.1节、2.2节和2.3节分别从社会网络分析角度、系统科学分析角度和信息搜索领域的分析角度,总结并分析了近年来复杂网络在这几个领域内发掘重要性节点的方法。在第2.4节独立地讨论了网络中节点的相对重要性。最后总结了在复杂网络环境下发掘节点重要性的意义和进一步研究、应用的方向。

### 2.1 社会网络分析方法

社会网络分析的研究在20世纪40年代末就开始了,其中各种主流的方法都有这样一个假设,即节点的重要性等价于该节点与其他节点的连接而使其具有的显著性<sup>[25]</sup>,指标的研究不破坏网络的整体性(连通性)且通常不考虑节点集的重要性。这些方法的一个重要的基本思想是,网络中不同节点之间的重要性差异是通过分析网络中某种有用的信息得到的,例如节点的度、最短路径、节点和边上的权值等。通过对这些基本属性的统计、计算,能相对定量地反映出节点在网络中的位置特性,将网络节点的显著性进行“放大”来定义节点

<sup>2</sup> 2003年8月,由美国俄亥俄州克里夫兰市的3条超高压输电线路相继过载烧断,引起的北美大停电事故,使得数千万人一时陷入黑暗,经济损失估计高达数百亿美元。

的重要性。

其中,已提出的发现重要节点的指标主要分为核心性(Centrality)和声望(Prestige)两大类,度量的方法主要包括节点的度(Degree)、接近度(Closeness)、介数(Betweenness)、信息(Information)、特征向量(Eigenvector)和累计提名(Cumulated nomination)等,详见文[26~33]。现简单介绍一下其中几种常见的方法。

一个节点的度(Degree),是指网络拓扑中与此节点连接的边的数量,即

$$d_v = \sum_{i \in E} \delta_i^v$$

其中  $\delta_i^v$  取值为 1 当路径  $l$  包含节点,否则为零。

如果考虑有向图,则节点的度还可以根据边的方向分为入度(In Degree)和出度(Out Degree)。

将网络中的节点按照其度的大小进行排序,可以一定程度上反映节点的重要性。节点的度值越高,则这个节点就越重要。这种方法思想简单,反映了典型的从众心理,且计算复杂度不高( $O(L)$ ,  $L$  为网络中边的数目)。但是其缺点也同样明显,仅仅从节点度的大小并不能准确表达网络中节点的重要程度。例如,一个节点的度值虽然很高,但是连接它的其他节点并不重要,则这个节点并不一定非常重要;反之,若一个节点的度值并不是非常高,但是连接它的节点多数都非常重要,则这个节点在网络中可能是个非常重要的节点。例如有些名人的博客写得并不是很好,像记流水账,但是他们的博客在排行榜中却名列前茅,这与他们的“粉丝(fans)”数量较多有很大的关系。

接近度(Closeness)反映了节点在网络中居于中心的程度。如图 2 所示,  $P_2$  分别与节点  $P_1$ 、 $P_3$  和  $P_4$  直接相连,它到达  $P_5$  只需经过  $P_4$ 。而对于  $P_1$  来说,只有  $P_2$  与它直接相连,它要到达  $P_3$  和  $P_4$  必须经过  $P_2$ ,要到达  $P_5$  必须经过  $P_2$  和  $P_4$ 。为了到达图中的每个节点,  $P_1$  必须经过  $P_2$  三次,  $P_4$  一次。因此,  $P_2$  比  $P_1$  更趋近于网络的中心。

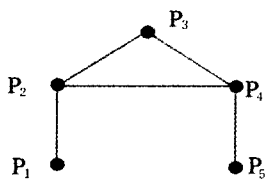


图 2 节点的接近度

Bavelas<sup>[34]</sup>指出,在网络中最中心的节点上产生的消息,将以最短的时间传播遍整个网络。网络中较短的距离意味着更少的消息传递时间和花费。

节点接近度的定义有很多形式,其中最简单、最自然的定义是由 Sabidussi 在 1966 年提出的<sup>[35]</sup>。假设  $d(P_i, P_k)$  表示以  $P_i$  为起点  $P_k$  为终点的路径所包含的边的数量,则节点  $P_k$  的接近度  $C(P_k)$  可以表示为  $C(P_k)^{-1} = \sum_{i=1}^n d(P_i, P_k)$ ,即表示节点  $P_k$  到其他所有节点距离之和的倒数。

节点的接近度越大,表明节点越居于网络的中心,它在网络中就越重要。但是,接近度对网络的拓扑结构依赖性很大,对于集中式的星形网络它可以准确地发现中心节点,但是对于民主式的正则图、ER 随机图网络则并不适合。

介数(Betweenness)最早是 Freeman 于 1977 年在研究社会网络时提出的用于衡量个体社会地位的参数<sup>[36]</sup>。节点  $u$  的介数含义为网络中所有的最短路径之中经过  $u$  的数量。记

$(i, j)$  之间最短路径的集合为  $S_{ij}$ , 则节点  $u$  归一化后的介数定义为:

$$B_u = \sum_{i,j} \frac{\sum_{l \in S_{ij}} \delta_l^u}{|S_{ij}|}$$

其中,  $\sum \delta_l^u$  表示经过节点  $u$  的最短路径的数量。

对网络中每个节点的介数进行计算、排序,也可以表达节点的某种重要性。节点的介数值越高,这个节点就越有影响力,这个节点也就越重要。使用介数来判断人际关系网络中节点的重要程度,则其表示某个人在关系网络中最短路径上出现的次数,这种次数越大,则其影响范围越大,其他人的交流渠道与此人也就越密切,因此节点也就越重要。

使用介数可以准确找到网络中某些“流量”非常大的重要节点,但其缺点是介数的计算复杂度非常高,为  $O(N^3)$ ,其中  $N$  为节点数目,即使用网络的特征简化介数计算<sup>[37]</sup>,算法的复杂度仍然很高,在无权网络和有权网络的情况下分别为  $O(MN)$  和  $O(MN + N^2 \times \log(N))$ ,其中  $N$  为网络的节点数,  $M$  为网络中边的数目。

特征向量(Eigenvector)指标,是从网络中成员的地位或名望角度考虑,将单个成员的名望看成是所有其他成员名望的线性组合,从而得到一个线性方程组,该方程组的最大特征值所对应的特征向量就是各个节点的重要性指标。Poulin 等在对 Bonacichi 求解特征向量的映射迭代方法进行改进的基础上,提出了收敛速度更快、更稳定而且适用于大网络和多分支网络的累计提名(Cumulated nomination)指标。但是这类方法是在保证网络整体性的前提下进行的,具有一定的局限性。

## 2.2 系统科学分析方法

系统科学的研究方法是利用网络的连通性来反映系统某种功能的完整性,通过度量节点删除对网络连通的破坏程度来反映网络节点(集)的重要性,即“破坏性等价于重要性”<sup>[38]</sup>。

系统科学方法的主要研究成果是系统的“核与核度”理论<sup>[39]</sup>。系统的“核”定义为那些对系统功能来讲具有重要的或支配性作用的且一旦遭到破坏会使整个系统瘫痪或造成重大损失的节点或者节点的集合,而“核度”的计算方法则采用点割集和连通分支数来定义。这种研究节点(集)重要性的思路,源于图论中点割集<sup>[40]</sup>的概念,即是通过度量节点(集)被删除后对网络连通的破坏程度来定义其重要性的。对网络连通的破坏程度越大,被删除的节点(集)越重要,因为网络连通的维持依赖于它们的存在。但是该理论的主要目的是为了解决明显连通性不同的图却有相同连通度  $K(G)$  的问题<sup>[41]</sup>。另外,对于点割集中各个节点的重要程度,也不能给出明确的等级,不同点割集中节点的重要性也无法横向比较,因为它们的删除都会使系统不再连通。

系统中节点(集)的删除除了对系统连通性可能具有破坏,还会影响到系统的一些其他指标,也可以通过计算这些指标的性能变化来度量节点的重要性。例如,Corley 和 Sha<sup>[42]</sup>提出了一种与节点间最短路径相关的度量指标。在这种方法中,如果一条最短路径上的源节点和目标节点之间的某个节点被删除,会导致源节点和目标节点之间的距离增加最大,那么这个被删除的节点就是一个重要的节点。类似地,他们也考虑了寻找最短路径上  $k$  个最重要的节点的问题,并且给出了一个指数级复杂度的寻找算法。对于  $k=1$  的情况,需要  $O(mn + n^2 \log n)$  的时间复杂度和  $O(m)$  的空间复杂度,其中  $m$

和  $n$  分别表示系统中边和节点的数量。此外,Chen 等人也提出了一种基于最小生成树的指标,即节点的重要性决定于该节点被删除后系统中最小生成树数量的变化情况<sup>[43]</sup>。去掉节点以及相关连的边后,所得到的图对应的生成树数量越少,则表明该节点越重要。

### 2.3 信息搜索领域分析方法

互联网可以被看作是一个巨大的图,其中节点代表网页,(有向)边代表网页之间的超链接。在互联网搜索领域,计算机科学家也提出了很多算法来判断网页节点的重要程度,其中近几年来两个最著名、最有代表性的算法是 Brin 和 Page 在 1998 年提出的 PageRank 算法<sup>[44]</sup> 和 Kleinberg 在同一年提出的 HITS 算法<sup>[45]</sup>。此后,互联网的搜索领域也随着 Google 的成功,越来越受到人们的关注,不断有新的算法和变种提出。Lempel 和 Moran 在 2000 年提出了 SALSA 算法<sup>[46]</sup>,它是 HITS 算法的一个变种,考虑了用户回退浏览不同页面的情况。其他一些学者也相继提出了另外的算法,如 PHITS, Bayesian, Reputation 等算法,这些算法实际上都显式或隐式地对网页节点的重要性进行了计算、排序,在实际的应用中极大地提高了检索结果的质量。

PageRank 算法是由斯坦福大学的博士研究生 Sergey Brin 和 Lawrence Page 于 1998 年提出的,并成功应用在

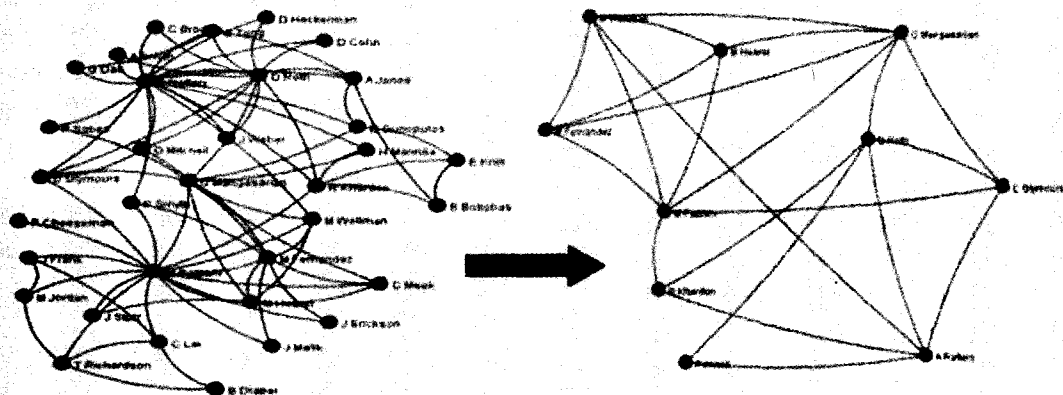


图 3 PageRank 算法发掘出的重要节点及其连接

在 PageRank 算法的基础上,许多研究者都提出对基于 PageRank 算法的改进<sup>[48]</sup>。华盛顿大学计算机科学与工程系的 Matthew Richardson 和 Pedro Domingos 提出了结合链接和内容信息的 PageRank 算法<sup>[49]</sup>。斯坦福大学计算机科学系 Taher Haveliwala 提出了主题敏感 (Topic-sensitive) PageRank 算法<sup>[50]</sup>。

HITS 算法是由美国康奈尔大学的 Jon Kleinberg 博士于 1998 年首先提出的,HITS 的英文全称为 Hypertext-Induced Topic Search。他将网页分为两类,即表达某一特定主题的 authorities 和把 authorities 串连起来的 hubs。authorities 为具有较高价值的网页,依赖于指向它的页面;而 hubs 为指向较多 authorities 的网页,依赖于它所指向的页面,每个节点也引入了两个权值:authority 权值和 hub 权值。HITS 算法的目标就是通过一定的迭代计算方法得到针对某个检索提问的最具价值的网页,即排名最高的 authority 权值。HITS 算法在学术界应用较为广泛,其计算复杂度为  $O(VI)$ ,其中  $V$  为网络中节点的数目, $I$  为算法达到收敛所需的迭代次数。

Google 搜索引擎中,Google 现在已经发展成为互联网上最好的搜索引擎之一<sup>3</sup>。其主要思想是:当网页 A 有一个链接指向网页 B 时,就认为网页 B 获得了一定的分数,该分值的多少取决于网页 A 的重要程度,即网页 A 的重要性越大,网页 B 获得的分数就越高。由于 Web 上链接相互指向的复杂程度,该分值的计算过程是一个迭代过程,最终网页将依照所得的分数进行排序并将检索结果送交用户,这个量化的分数就是 PageRank 值,其计算公式如下:

$$PR(A) = (1-d) + d(PR(T_1)/C(T_1) + \dots + PR(T_n)/C(T_n))$$

其中  $PR(A)$  是网页 A 的 PageRank 值; $d$  为界于 (0,1) 区间的衰减系数,一般取 0.85 左右; $T_1, T_2, \dots, T_n$  为指向网页 A 的其它网页, $C(T_i)$  是网页  $T_i$  向外指出的链接数目。

PageRank 算法能够在网络中准确定位节点的重要程度,即根据用户查询的匹配程度,而且计算复杂度不高,为  $O(EI)$ ,其中  $E$  为网络中边的数目, $I$  为算法达到收敛所需的迭代次数。具体效果如图 3 所示。左图是一个有 34 个节点的空手道俱乐部成员之间的关系拓扑<sup>[47]</sup>,节点旁边标注的是人名;右图是用 PageRank 算法找到的前 10 个重要成员及其之间的连接关系。

### 2.4 网络中节点的相对重要性

不管是社会网络,还是 WWW 网络,前面介绍的各种算法几乎都是站在整体的角度,或显式或隐式地对网络中所有节点的重要性进行全局的排序,很少有学者关注网络中节点的相对重要性。但是,诸如“相对网络中某个或某些节点,哪些节点是最重要的?”这类问题,提示我们有时候节点的相对重要性、局部重要性往往更加重要,尤其是当网络的规模非常大的时候。Chang, Cohn 和 McCallum 在 2000 年提出了一种个人化的 (personalized) HITS 算法变种<sup>[51]</sup>,2002 年 Haveliwala<sup>[52]</sup>, Jeh 和 Widom<sup>[53]</sup> 也分别提出了各自的个人化的 PageRank 算法变种,虽然他们的目标不是单纯地发掘网络中的重要节点,但是他们的算法的确开始更多地考虑了网络中节点的相对重要性。

2003 年 White 和 Smyth<sup>[54]</sup> 进一步总结了相关的工作,明确地提出了网络中节点的相对重要性,即网络中节点的重要性是相对于一个或多个事先指定的根节点集 (root set) 的。White 和 Smyth 通过对四种渐近性的问题描述,定义了一个

<sup>3</sup> 关于 Google 以及前 Google 时代的搜索,请参见“The search”, John Battelle, 2006。

通用的基础架构,来发掘网络中节点的相对重要性。

(1) 给定图  $G(V, E)$ 、节点  $r$  和  $t$ , 其中  $\{r, t\} \subset G$ , 根据根节点  $r$  计算节点  $t$  的重要性  $I(t|r)$ 。

(2) 给定图  $G(V, E)$ , 根节点  $r \in G$ , 根据根节点  $r$  计算集合中  $T(G)$  节点的重要性, 其中  $T \subseteq V$ 。即对集合  $T$  中的所有节点  $t$ , 根据(1)中方法分别计算其重要性  $I(t|r)$ , 计算结果大的则其相对根节点  $r$  的重要性就大。

(3) 给定图  $G(V, E)$ , 一组待计算重要性的节点集合  $T(G)$ , 一组根节点集合  $R(G)$ , 其中  $R \subseteq V$ , 根据根节点集  $R$  计算集合  $T$  中节点的重要性  $I(t|R)$ 。这里  $I(t|R)$  通常是由根集中节点确定的函数  $\{I(t|r) : r \in R\}$ , 例如可以把相对于根集  $R$  的节点  $t$  的重要性函数定为  $t$  相对  $R$  中各节点重要性的平均值:

$$I(t|R) = \frac{1}{|R|} \sum_{r \in R} I(t|r)$$

或者是取其中的最小值:

$$I(t|R) = \min\{I(t|r) : r \in R\}$$

因此, 这个函数的定义可以作为评估节点相对重要性的一个策略, 根据不同的实际问题进行定义。

(4) 给定图  $G(V, E)$ , 计算所有节点的重要性。这相当于(3)的一个特例, 即  $R=T=V$  的情况, 也就是节点全局的重要性。

基于上面的基础架构, White 和 Smyth 根据社会网络分析、图论、马尔可夫链模型和 Web 网络分析中现有的算法, 分别自定义了不同的节点重要性函数  $I(t|R)$ , 来计算网络中节点的相对重要性。通过对“911”恐怖主义网络、公司与大学间的生物科技合作网络和 CITESEER 的论文合作者网络进行评估, 实验结果表明他们提出的几种计算节点相对重要性的算法, 对这些实验网络都具有较高的一致性。

**总结** 本文介绍了在复杂网络环境下, 对复杂网络中节点重要性发掘的几个领域, 并对其中重要的方法进行了讨论和分析。主要集中在传统的社会网络、系统科学和源于互联网的信息搜索等领域中的一些算法和思想, 指出了其各自的优缺点。另外, 相对于全局地评估节点的重要性方法, 在最后也介绍了复杂网络中节点的相对重要性分析。

总的来说, 社会网络分析方法的核心思想是“重要性等价于显著性”, 对网络中重要节点的发掘不以破坏网络的整体性为基础。系统科学的分析方法则主要基于节点(集)的删除, 核心思想是“重要性等价于该节点(集)被删除后对网络的破坏性”, 对网络中重要节点的发掘, 是通过节点(集)删除前后网络连通性、性能的变化来反映的。深入应用于互联网搜索领域的以 PageRank 及其变种为代表的分析方法, 则考虑了更多因素, 节点的重要性不仅取决于其自身的连接度, 而且与周围邻居节点的重要性有关。此外, 相对于各种从全局角度排序重要性节点的方法, 网络中节点的相对重要性也日益引起学者们的重视。White 等学者深入讨论和分析了网络中节点的相对重要性, 指出了这个领域的一个新的方向。

分析网络中节点的重要性, 对于网络的安全防护等应用具有非常重要的意义。特别是现实世界中很多网络都具有复杂网络的特性, 如小世界、无尺度、幂律分布等特性, 在复杂网络的环境中, 发掘网络中节点的重要性显得更加重要。在今后的工作中仍然有很多问题值得深入研究: (1) 现有的各种度量网络中节点重要性的指标, 比如节点的度、介数等, 它们之间是否具有一定的联系? (2) 现实世界中虽然很多网络都可

以抽象为复杂网络, 但各种发掘重要性节点的方法, 是否对它们都适用? 例如, 社会关系网络<sup>[55]</sup>与恐怖主义分子网络<sup>[56]</sup>就有所不同。(3) 同一网络的不同拓扑结构, 比如有向图、无向图、带权网络、不带权网络等, 是否对发掘网络中重要性节点有影响? (4) 发掘重要性节点是否对复杂网络中的社区结构的发掘有所帮助?

因此, 发掘网络中的重要节点, 对各种复杂网络的基础研究具有非常重要的意义。

## 参考文献

- 李德毅, 杜鹃. 不确定性人工智能. 国防工业出版社, 2005. 386~391
- Wasserman S, Faust K. Social Network Analysis. Cambridge: Cambridge University Press, 1994
- Newman M E J. The structure of scientific collaboration networks. In: Proc Natl Acad Sci USA 98, 2001, 404~409
- Albert R, Jeong H, Barabasi A-L. Diameter of the world-wide web. Nature, 1999, 401: 130~131
- Redner S. How popular is your paper? An empirical study of the citation distribution. Eur J Phys, 1998, B4: 131~134
- Faloutsos M, Faloutsos P, Faloutsos C. On power-law relationships of the internet topology. Comp Comm Rev, 1999, 29: 251~262
- 韩明畅, 李德毅, 刘常显. 软件中的网络化特征及其对软件质量的贡献. 计算机工程与应用, 2006
- Williams R J, Martinez N D. Simple rules yield complex food webs. Nature, 2000, 404: 180~183
- Jeong H, Tombor B, Albert R, et al. The large-scale organization of metabolic networks. Nature, 2000, 407: 651~654
- 郑金连, 狄增加. 复杂网络研究与复杂现象. 系统辩证学学报, 2005, 13(4)
- 许国志, 等. 系统科学[M]. 上海: 上海科技教育出版社, 2000
- Erdos P, Renyi A. On the evolution of random graphs. In: Publications of the Mathematical Institute of the Hungarian Academy of Sciences, 1960(5): 17~61
- Watts D J, Strogatz S H. Collective dynamics of small world networks. Nature, 1998(393): 440~442
- Milgram S. The small world problem. Psychology Today, May 1967. 60~67
- Barabasi A L, Albert R. Emergence of scaling in random networks. Science, 1999(286): 509~512
- Grubestic T H, Murray A T. Vital Nodes, Interconnected Infrastructures, and the Geographies of Network Survivability. Annals of the Association of American Geographers, March, 2006, 96: 64
- Girvan M, Newman M E J. Community structure in social and biological networks. Proc Natl Acad, 2002
- Barabasi A L, Bonabeau E. Scale-Free networks. Sci Am, 2003
- Demo: COPLINK Criminal Network Analysis. 2005. <http://ai.eller.arizona.edu/research/coplink/crimenet.htm>
- Zanette D H. Dynamics of rumor propagation on small-world networks. Phys Rev, E, 2002, 65: 041908
- 解伟, 汪小帆. 复杂网络中的社团结构分析算法研究综述. 复杂系统与复杂性科学, 2005, 2(3)
- West D B. Introduction to Graph Theory [M]. Prentice Hall, Upper Saddle River, 2001
- Washio T, Motoda H. State of the art of graph-based data mining [J]. SIGKDD Explor Newsl, 2003, 5(1): 59~68
- Scott J. Social Network Analysis: A Handbook [M]. London: Sage Publications, 2000
- Knock D, Burt R S. Prominence [A]. In: Burt R S, Minor M J. Applied network analysis [C]. Newbury Park, CA: Sage, 1983. 195~222
- Wasserman S, Faust K. Social network analysis; methods and applications [M]. Cambridge: Cambridge University Press, 1994
- Freeman L C. Centrality in social networks; I. Conceptual clarification [J]. Social Networks, 1979, 1: 215~239
- Bonacich P. Technique for analyzing overlapping memberships [A]. In: Costner H. Sociological methodology [C]. San Francisco: Jossey Bass, 1972. 176~185
- Bonacich P. Factoring and weighting approaches to status scores and clique identification [J]. J Math Sociol, 1972, 2: 113~120
- Bonacich P. Power and centrality: a family of measures [J]. American Journal of Sociology, 1987, 92: 1170~1182
- Stephenson K, Zelen M. Rethinking centrality: methods and applications [J]. Social Networks, 1989, 11: 1~37

(下转第 17 页)

信或互相通信,对应水平通信的地址和 GLL 实体分别为  $IP_C$ 、 $IP_D$ 、 $IP_F$  和  $GLL_C$ 、 $GLL_D$ 、 $GLL_F$ 。在接入技术上,根据通信双方具备的接口不同,选择适当的技术。在此意义上,数据流从哪种无线接入技术发送取决于 GLL 本地的范围的判断,而端到端的接入方式以及路由则取决于 MRRM 层面。

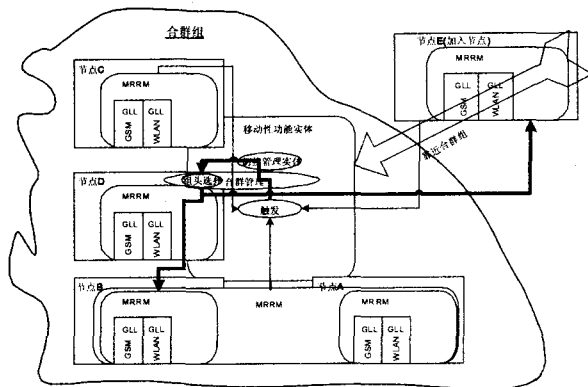


图 13 基于多无线混合通信的合群过程示意

在上述基于多无线的混合通信结构基础上,位置相近的移动台可以通过移动性功能实体中合群管理模块组织成组。在 MRRM 协调水平通信的过程中,它会将移动周围节点的信息通过网络控制接口,交给移动性功能实体中的触发模块处理。

图 13 即展示了这样一个节点加入合群组的过程。当符合触发条件时,移动台触发切换过程。这里的切换并非传统意义上的蜂窝间的切换,而是从独立状态(在某蜂窝覆盖的子网)切换到合群状态(由组头管理的自治子网),是一种网络间的更广义的垂直切换。在合群建立过程中,还会经过组头选择。

**结论** 本文提出一种支持合群管理的多无线混合网络结构模型,通过不同方式协同工作,有效利用移动通信终端物理上的合群特性,将节点组织成组管理;再利用信息融合技术,采用空间复用、冗余信息压缩、组播等方式,达到节省能量、带宽等功效。

**参考文献**

- 1 Basagni S, Conti M, et al. Mobile Ad Hoc Networking [M]. Wiley IEEE Press, 2004
- 2 Prasad R, Ruggieri M. Technology Trends in Wireless Communications [M]. Artech House Press, 2003
- 3 Ying-Dar L, Yu-Ching H. Multihop cellular: a new architecture for wireless communications [A]. In: INFOCOM 2000 Proceedings [C]. US: IEEE Press, 2000. 1273~1282
- 4 Niebert N, Schieder A, et al. Ambient Networks: An architecture for communication networks beyond 3G [J]. IEEE Wireless Comm Mag, Apr 2004
- 5 Rossi M, Badia L, Giacon P, et al. On the effectiveness of logical device aggregation in multi-radio multi-hop networks [A]. In: 2005 International Conference on Wireless Networks, Communications and Mobile Computing [C], 2005. 354~361
- 6 王芙蓉, 涂来, 黄载禄. 基于多跳蜂窝网的组位置管理策略[J]. 通信学报, 2004(07): 56~61
- 7 Wang F, Tu L, et al. Group Location Update Scheme and Performance Analysis for Location Management in Mobile Network [A]. In: IEEE VTC 05 Spring Proceedings [C]. US: IEEE Press, 2005. 2429 ~2433
- 8 IEEE 802.11b. IEEE standard for wireless LAN medium access control (MAC) and physical Layer (PHY) specification: higher-speed physical layer extension in the 2.4GHz band [S]

(上接第 5 页)

- 32 Altmann M. Reinterpreting network measures for models of disease transmission [J]. Social Networks, 1993, 15: 1~17
- 33 Poulin R, Boily M-C, Masse B R. Dynamical systems to define centrality in social networks [J]. Social Networks, 2000, 22: 187~220
- 34 Bavelas A. A mathematical model for group structures. Human Organization, 1948, 7: 16~30
- 35 Sabidussi G. The centrality index of a graph. Psychometrika, 1966, 31: 581~603
- 36 Freeman L. A set of measures of centrality based upon betweenness. Sociometry [J], 1977, 40: 35~41
- 37 Brandes U. A Faster Algorithm for Betweenness Centrality [J]. Journal of Mathematical Sociology, 2001, 25(2): 163~177
- 38 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)
- 39 许进, 席酉民, 汪应洛. 系统的核与核度 [J]. 系统科学与数学, 1993 年, (13)2: 102~110
- 40 左孝凌, 等. 离散数学. 上海科学技术文献出版社, 280~286
- 41 李鹏翔, 任玉晴, 席酉民. 网络节点(集)重要性的一种度量指标. 系统工程, 2004
- 42 Corley H W, Sha D Y. Most vital links and nodes in weighted networks. Oper. Res. Letters, 1982, 1: 157~160,
- 43 Chen Y, Hu A Q, Hu J, et al. A method for finding the most vital node in communication networks. High Technology Letters, 2004, 1: 573~575
- 44 Brin S, Page L. The Anatomy of a Large-scale Hypertextual Web Search Engine [J]. Computer Networks, 1998, 30: 107~117
- 45 Kleinberg J M. Authoritative Sources in a Hyperlinked Environment. Journal of the ACM, 1999, 46(5)
- 46 Lempel R, Moran S. The Stochastic Approach for Link-Structure

- Analysis(SALSA) and the TKC Effect. In: Proc. 9<sup>th</sup> Intl World Wide WEB Conf, 2000
- 47 Zachary W W. An information flow model for conflict and fission in small groups [J]. Journal of Anthropological Research, 1977, 33: 452~473
- 48 朱炜, 王超, 李俊, 等. Web 超链分析算法研究. 计算机科学, 2003, 30(1)
- 49 Richardson M, Domingos P. The Intelligent Surfer: Probabilistic Combination of Link and Content Information in PageRank. Cambridge, MA: MIT Press, 2002
- 50 Haveliwala T H. Topic-sensitive PageRank. In: Proc. of the Eleventh Intl World Wide WEB Conf., 2002
- 51 Chang H, McCallum A. Creating customized authority lists. In: Proceedings of the 17<sup>th</sup> International Conference on Machine Learning. San Francisco, CA: Morgan Kaufmann, 2000. 167~174
- 52 Haveliwala T. Topic-sensitive PageRank. In: Proceedings of the 11<sup>th</sup> International World Wide Web Conference, Honolulu, Hawaii, 2002. 517~526
- 53 Jeh G, Widom J. Scaling personalized Web search: [Technical Report]. Stanford University, Computer Science Department, 2002
- 54 White S, Smyth P. Algorithms for estimating relative importance in networks [J]. In: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM Press, 2003. 266~275
- 55 Hanneman R. Introduction to Social Network Methods. Department of Sociology, University of California, Riverside. (URL: http://faculty.ucr.edu/~hanneman/SOC157/Software/NET-TEXT.PDF)
- 56 Krebs V E. Uncloaking Terrorist Networks (URL: http://www.firstmonday.dk/issues/issue7\_4/krebs). FirstMonday, volume 7, number 4, April 2002