

文章编号:1007-6735(2012)01-0018-09

加权网络的常用统计量

姚尊强, 尚可可, 许小可
(青岛理工大学 通信与电子工程学院, 青岛 266520)

摘要: 加权网络是复杂网络研究的一个重要分支,连边权重的异质性有助于刻画复杂系统的各种特性.但长期以来,由于加权网络上各种统计量的定义不统一、物理意义不明确,很多学者直接抛弃交互作用的强度,使用门限值将加权网络变为二值无权网络后再进行研究.本文综述了加权网络上常用的统计量,并简要介绍了这些统计量在实际复杂系统分析中的应用.本研究有助于相关研究人员明确各种加权网络统计量的物理意义,使用加权网络对复杂系统进行分析和刻画.同时,理解各种常用统计量的内在联系和应用背景是构造更有效加权网络统计量的基础.最后介绍了各种权重网络的随机化置乱方法,为分析和理解实际加权网络统计量的绝对值提供了参考和比较.

关键词: 复杂网络; 加权网络; 统计物理; 复杂系统
中图分类号: N 94 **文献标志码:** A

Fundamental Statistics of Weighted Networks

YAO Zun-qiang, SHANG Ke-ke, XU Xiao-ke
(School of Communication and Electronic Engineering, Qingdao Technological University, Qingdao 266520, China)

Abstract: At the present time, all kinds of statistics of weighted networks have not a unified definition and their physical significances are not very clear. Therefore, many researchers turn to discard link weights using the threshold method and study the unweighted version of complex systems. In the paper, fundamental statistics of weighted networks were reviewed and their applications in real-life complex systems were introduced. The study is helpful to clarify the definition of fundamental statistics of weighted networks, which can be used to analyze and characterize complex systems. Especially, understanding the intrinsic relationships and application backgrounds of these fundamental statistics is the foundation for developing more effective statistics of weighted networks. Several shuffling methods for weighted networks were presented, giving reference and comparison to study a real-life weighted complex network.

Key words: complex network; weighted network; statistical physics; complex system

自 20 世纪 90 年代末 Watts 和 Barabási 分别提出小世界网络^[1]和无标度网络^[2]后,复杂网络理论就迅速发展起来,并广泛应用于多个交叉学科.经过国内外学者不断努力,十几年来该领域取得了丰硕的研

收稿日期: 2012-01-31
项目基金: 国家自然科学基金资助项目(61004104,61104143);山东省自然科学基金资助项目(ZR2011FQ008)
作者简介: 姚尊强(1987-),男,硕士研究生.研究方向:复杂网络. E-mail: yaozunqiang@163.com
许小可(联系人),男,副教授.研究方向:复杂网络与复杂系统. E-mail: xiaokeeie@gmail.com

究成果^[3-6],但绝大多数复杂网络文章中的研究对象不考虑连接强度,只考虑拓扑结构的无权网络:节点之间的关系是二值的,要么有连边(值为1)、要么无连边(值为0).尽管无权网络简化了实际系统的复杂性,有助于我们理解和刻画各种复杂系统,并且这种简化在很多时候也确实抓住了系统的本质特征.但毫无疑问,很多系统除了很重要的拓扑性质(如度分布的无标度特性等^[2]),其连接的强度和密度等方面也表现出很强的异质性^[7-8].连边强度的异质性往往是某些复杂系统的最重要特征,例如,社会网络中的个体之间存在着强、弱两种连接^[9],而且这两种连接的功能和作用是不同的^[10].同时,学者们也在各种类型网络中发现了连接强度的差别.如科学家之间相互合作的次数分布是异质的^[11];在新陈代谢的路径中存在着不均匀的通量^[12];食物链网络中捕食者和被捕食者的相互作用是多样的^[13];神经网络中不同神经元传递电子信号能力是不同的^[1];以及航空网络中每条航线运送旅客的数量也是不同的^[14].

加权网络中的每条连边都带有一个度量连接强度的数值,为复杂网络中节点之间的关系和相互作用提供了更加细致的刻画手段.权重及其分布会对网络的性质和功能产生重要影响,有助于我们更深入地理解复杂系统的性质,因此,加权网络已经成为复杂网络研究的一个重要分支.如Barrat等提出了颇具影响力的BBV模型^[15],Thadakamalla等研究了加权网络里面的搜索策略^[16],Filippo等提出了加权网络中的信息过滤的有效算法^[17].尤其是近年来,国内学者也有很多相关的研究成果,如严钢等研究了权重网络中的疾病传播问题^[18],田柳等分析了权重分布对于网络效率的影响^[19],王文旭和汪秉宏等分析了交通流驱动下生成的加权网络模型^[20-21],Lu等分析了加权网络中的同步问题^[22],而吕琳媛和周涛研究了链路预测中弱连接的强作用^[23].

然而,相对于无权网络研究的数量,加权网络的研究成果实际上是凤毛麟角^[24].尽管我们都知道,实际网络里面的连接都有一个自然的权重,一般是不相同的,但加权网络上的各种统计量比较难计算,很难弄清其真实的物理意义,并且由于权重的引入导致了加权网络上各种统计量的定义形式不统一.因此,很多学者直接丢掉网络中节点的交互强度,使用一个门限将加权网络变为二值无权网络后再进行研究,但是,此时门限如何选取又成为一个重要问题.鉴于此,本文首先综述了加权网络上常用的统计量,同时也简要回顾了这些统计量在实际权重网络

中的应用及其物理意义,最后分析了各种权重网络的随机化置乱方法.

研究加权网络首先就要确定权重的赋予方式,目前加权网络有两种表示权重的方式:相似权和相异权.相似权表达方式中的权重表示节点之间的关系紧密程度,权重越大,关系就越亲密,两点之间的距离就越小.比如以科学家合作网为例,两个科学家合作的次数越多,相似权的权重就越大^[25-26].相异权表达方式中的权重有点像两节点之间的距离,权重越大,关系就越疏远;权重越小,关系就越亲密.本文在下面论述中所谈的权重都统一为相似权.通常在各种网络里面,相似权 w_{ij} 的范围为 $[0, \infty)$,也可以归一化到 $[0, 1]$.如果 $w_{ij} = 0$,表示节点 i 和 j 之间不连接,没有直接相互作用.此外,考虑到有向加权网络的各种统计量形式过于复杂和繁琐^[27],本文讨论的主要对象是无向加权网络,但是,很多所探讨的统计量稍加扩展后对于有向加权网络也是适用的.

1 加权网络的常用统计量

1.1 节点强度特征

加权网络最突出的特征是连边的强度值是异质的,这种异质性刻画了系统中各成分之间交互作用的差异,是系统各种非线性和自组织行为等复杂统计特征的重要原因,同时,权重和拓扑之间的非相关性也为观察这类系统的组织结构提供了互补视角.例如,细胞网络拓扑的重要性是人所共知的,但是,细胞网络是由基因、蛋白质和其它调节细胞行为的分子之间通过相互作用产生的复杂网络,最近研究发现很多重要的信息都蕴含于相互作用的强度中^[12],因此,细胞网络中节点的强度分布是很重要的.节点强度分布 $p(s)$ 度量了节点强度为 s 的概率,它和度分布 $p(k)$ 以及边权分布 $p(w)$ 一起,为我们观察加权网络提供了多个视角.比如在有些网络中,节点强度分布 $p(s)$ 以及边权分布 $p(w)$ 都呈现出很好的幂律分布特性^[28].

在权重网络中,节点 i 的度 k_i 可以很自然地推广为节点 i 的强度 s_i ,其定义为^[14-15]

$$s_i = \sum_{j \in N_i} w_{ij} \quad (1)$$

式中, N_i 为 i 的最临近节点的集合.

由式(1)可以看出,强度 s_i 整合了节点的度 k_i 和与节点相连的边权 w_{ij} 的两方面特征,是节点局域

信息的一种综合表达方式.

单位权重是在节点权重的基础上引入的,表示节点 i 所有连边的平均权重值,其定义为^[29]

$$u_i = \frac{s_i}{k_i} \quad (2)$$

单位权重描述的是节点拥有连边的平均权重情况,两个节点的单位权重相同时,并不需要这两个节点的度和点强度都相等.在单位权重相同的情况下,可能是每条连边的权重都接近于单位权重 u_i 的数值;或者相反,只有少数连边的权重占优势.对于某个度数为 k_i 、强度为 s_i 的给定节点 i ,其权重分布的差异性可以由统计量 y_i 来度量,其定义为^[12]

$$y_i = \sum_{j \in N_i} \left[\frac{w_{ij}}{s_i} \right]^2 \quad (3)$$

y_i 描述了节点 i 连边权重分布的离散程度,且间接依赖于节点 i 的度 k_i . 如果所有的权重值差异不大,则 y_i 与节点的度 k_i 的倒数成正比;相反,如果仅有一条边的权重起主导作用,则 $y_i \simeq 1$. 由于一个大网络中以上两种情况的节点可能都存在,因此,通常更关心所有度值为 k 的节点的平均值 $y(k)$ 的结果.如果经过统计发现,节点的边权重分布都很平均,相互之间都差不多,则 $y(k)$ 的值约为 $1/k$. 而如果绝大多数节点中边的权重差异性很大,比如仅有一条边的权重远远大于其它连边,则 $y(k) \simeq 1$. 换句话说,此时 $y(k)$ 独立于度值 k .

1.2 边权重和拓扑结构的关系

当网络中的连边被赋予不同权重后,很自然地研究者会期望知道权重和拓扑之间是否存在相关性.当连边权重与网络的拓扑结构无关时,度值为 k 的节点的平均强度随着 k 线性增长,近似为

$$s(k) \sim Ak^\beta \simeq \langle w \rangle k \quad (4)$$

式中, $\langle w \rangle$ 为平均权重; A, β 为待拟合系数.

式(4)表示连边权重与网络的拓扑特征(如节点的度值)之间没有关系.而当连边权重与网络的拓扑结构有相关性时,此时,要么 $\beta=1$ 但 $A \neq w$, 要么 $\beta \neq 1$.

此外,还有另外一个公式可以用来判断边权重和连边两个端点的度之间是否具有相关性^[14]

$$\langle w_{ij} \rangle \sim (k_i k_j)^\theta \quad (5)$$

式中, $\langle w_{ij} \rangle$ 为连边两端节点的度乘积 $k_i k_j$ 相同的连边的平均权重.

式(5)衡量了连边权重和连边两个端点的度乘积之间是否有相关联系,进而某种程度上衡量了边权重和网络拓扑结构的关联性.如果权重和两边端点的度乘积是相互独立的,则 $\theta \approx 0$.

1.3 匹配特性

1.3.1 匹配系数

在无权网络中,最常用的衡量节点度匹配特性的统计量就是匹配系数^[30-31],该系数定义为

$$r = \frac{M^{-1} \sum k_i k_j - [M^{-1} \sum \frac{1}{2} (k_i + k_j)]^2}{M^{-1} \sum \frac{1}{2} (k_i^2 + k_j^2) - [M^{-1} \sum \frac{1}{2} (k_i + k_j)]^2} \quad (6)$$

式中, M 为网络中连边的总条数; k_i 和 k_j 为 M 中一条连边两个端点的度数.

如果匹配系数 $r > 0$, 则称网络是同配的; 如果 $r < 0$, 则称网络是异配的. 对应于加权网络, 其加权匹配系数可以定义为^[32]

$$r_{w1} = \frac{H^{-1} \sum w_{ij} k_i k_j - [H^{-1} \sum \frac{1}{2} w_{ij} (k_i + k_j)]^2}{H^{-1} \sum \frac{1}{2} w_{ij} (k_i^2 + k_j^2) - [H^{-1} \sum \frac{1}{2} w_{ij} (k_i + k_j)]^2} \quad (7)$$

式中, H 为网络中所有连边的总权重; w_{ij} 为端点为 i 和 j 的连边的权重.

与 r 相同, r_{w1} 的值域也在 -1 和 1 之间. 当 $r_{w1} > r$ 时, 可以理解为高权重的连边倾向于存在于度差不多的节点之间; 相反, 当 $r_{w1} < r$ 时, 可理解为高权重的连边存在于度相差很大的节点之间. 尽管 r_{w1} 的定义能保证当所有的权重都相同时, $r_{w1} = r$, 但上述的定义和原始匹配系数的物理意义是有一定差别的. r 表征的是网络的度匹配特性, 那么, 加权匹配系数应该表征的是节点强度的匹配特性(倾向于连接还是不连接), 但是, r_{w1} 的定义和这个目标有很大的区别, 因此可以尝试将加权匹配系数定义为

$$r_{w2} = \frac{M^{-1} \sum s_i s_j - [M^{-1} \sum \frac{1}{2} (s_i + s_j)]^2}{M^{-1} \sum \frac{1}{2} (s_i^2 + s_j^2) - [M^{-1} \sum \frac{1}{2} (s_i + s_j)]^2} \quad (8)$$

式中, s_i 和 s_j 分别为一条连边两个端点的权重.

如果加权匹配系数 $r_{w2} > 0$, 则称加权网络是同配的, 此时点强度相似的节点倾向于相互连接; 如果 $r_{w2} < 0$, 则称加权网络是异配的, 此时点强度相差很大的节点之间倾向于存在连接关系.

1.3.2 邻居节点的平均度

在无权网络中, 邻居节点的平均度是另外一种衡量网络度匹配特性的方法^[33], 其定义为

$$k_{m,i} = \frac{1}{k_i} \sum_{j \in N_i} k_j \quad (9)$$

式中, $k_{m,i}$ 又称为节点 i 的邻居节点的平均度, 它是节点 i 所有邻居节点度的平均值.

如果先计算出每个节点的邻居节点的平均度,然后再统计一下该值和节点的度之间是否存在一定的依赖关系,就可知道网络的度匹配特性了.用 $k_{nn}(k)$ 代表度为 k 的节点的 k_{nn} 平均值,如果网络是同配的,则 $k_{nn}(k)$ 随着 k 的增加而增加;相反,如果网络是异配的, $k_{nn}(k)$ 随着 k 的增加而变小.对应于加权网络,节点 i 的最近邻节点权重的平均可以定义为^[14,32]

$$k_{w,nn,i} = \frac{1}{s_i} \sum_{j \in N_i} w_{ij} k_j \quad (10)$$

很明显,式(10)是由边的归一化权 w_{ij}/s_i 得到的最近邻度节点平均度的局域加权平均.这个统计量可以刻画加权网络的同类匹配和异类匹配特性.对于某一节点 i 而言,当大的相互作用 w_{ij} 存在于它和大度节点之间时,有 $k_{w,nn,i} > k_{nn,i}$; 否则,有 $k_{w,nn,i} < k_{nn,i}$. 这样根据实际相互作用的大小, $k_{w,nn,i}$ 度量了点 i 与度数高或低的节点相连接的情况.类似地,如果使用函数 $k_{w,nn}(k)$ 来定义所有度为 k 节点的 $k_{w,nn,i}$ 均值,那么它刻画了整个网络的加权同配性和异配性.

1.4 最短路径长度

最短路径长度是一个重要的网络统计量.一方面它是衡量网络是否是小世界网络的一个重要统计量,例如,社会网络中一个非常重要的工作就是1967年米勒实验得出人与人之间的“六度分离”现象^[34];另一方面,它也是很多网络上其它重要统计量的基础,比如中心性度量里面的介数就是依赖于网络里面节点之间的最短路径^[35].

在无权网络中,最短路径长度相对而言比较容易计算.如果两个节点之间直接相连,它们之间的最短路径就为1;如果两个节点不直接相连,而它们中间有一个中介点,那么,它们的距离就增加1,以此类推.对于权重网络而言,路径长度的计算困难出现在权重不同的时候.例如,在图1(a)中,当节点 A 和 B 之间的权重是 A 和 C 之间权重的两倍,这意味着 A 和 B 接触的频率要比 A 和 C 之间的要高.因此,相对于 C 而言,这意味着 A 可能会传播更多的信息(或疾病)给 B .实际上,如果我们看一下信息或者疾病在网络中的传播,旅行的速度或者传播的路线都会很明显地受到权重的影响.如果要计算节点 A 和 C 之间的最短路径,虽然它们之间的直接连接的权重为1,但是,通过节点 B 的非直接连接有非常大的权重.因此,信息或者疾病通过节点 B 传播可能比它们之间直接传播更加迅速.

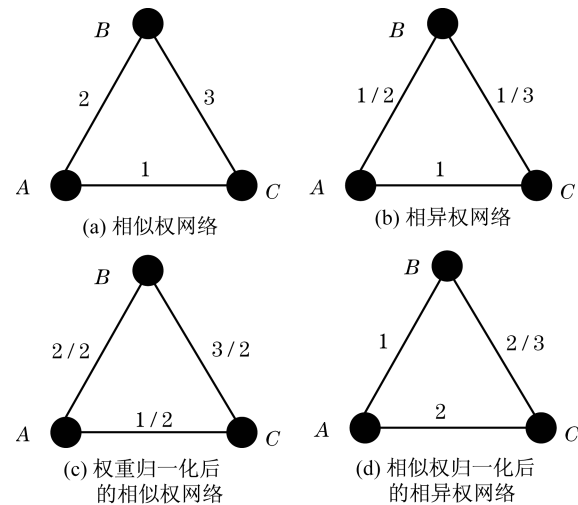


图1 拥有3个节点的加权网络

Fig.1 Weighted network with three nodes

Dijkstra 在 1959 年提出了一个算法来计算连接的费用,从而找到路径里面阻抗最小的路径^[36].例如, GPS(全球定位系统)可以使用这个算法来计算经过每一条路径时所花的时间,这样就可以找到两点间花费时间最小的路径.同样,这个算法也可以用在社会网络分析上面. Newman 在 2001 年使用该算法来研究科学家合作网中科学家之间的关系,把一篇文章的总权重设为1,然后按照合作者的数量 n 将合作者之间的权重设为 $1/n$ ^[11].对于一般网络,在得到连边的相似权重以后,可以通过取倒数的方法得到连边相异权的权重(图1(b)中的网络就是将图1(a)中相似权的权重取倒数后的网络),可将该权重当作是两个节点之间连接的费用,然后再计算最短路径长度.这种方法就意味着强连接之间的费用有可能要比弱连接之间的费用低.应用 Dijkstra 算法后,可以发现 A 和 C 之间直接的代价为1,而 A 和 C 之间经过 B 的代价为0.83 ($1/2 + 1/3$).因此,根据这个算法, A 和 C 之间通过非直接连接传递信息的速度要快于直接连接.

尽管使用 Dijkstra 算法能找到我们期望的最短路径,然而, A 和 C 之间的最短路径长度是0.83表示什么,这比较难理解,并且物理意义也很不明确.为了避免将权重信息变成费用信息以后出现的难以比较的情况,一个比较好的方式是将权重根据平均权重进行归一化^[37],这样的话,转换以后的相似权的权重如图1(c)所示.接着使用 Dijkstra 算法和图1(d)中归一化后权重的倒数,就可以得到节点之间的最短路径.此时 A 和 C 之间的最短路径长度为1.67 ($1 + 2/3$).这个值意味着这两个节点之间的距离为1.67步.以此为基础,就可以获得加权网络中

任意两点间的最短距离以及整个网络的平均最短路径长度. 而其它网络的全局统计量, 如效率^[19]和介数^[35]等, 就可以在加权最短路径的基础上进行计算. 使用归一化方法的好处是不仅使统计量的物理意义更明确, 还可以用来比较不同网络的平均最短路径长度, 即使这些网络的权重范围是不同的.

1.5 聚类系数

1.5.1 全局聚类系数

在无权重的二值网络中, 聚类系数是对拓扑性质进行度量的重要物理量, 有两种形式的定义. 一种是全局聚类系数, 一种是局域聚类系数. 全局聚类系数的定义比较简单^[4,38], 其定义为

$$C_1 = \sum \tau_{\Delta} / \sum \tau \quad (11)$$

式中, $\sum \tau_{\Delta}$ 为网络中所有三角形数目的 3 倍; $\sum \tau$ 为网络中关联三点组的数目.

全局聚类系数 C_1 反映了在整个网络中“朋友的朋友是朋友”的可能性. 在权重网络上, 加权的全局聚类系数可仿照式(11)的形式定义为^[39]

$$C_{1w} = \sum_{\tau_{\Delta}} w / \sum_{\tau} w \quad (12)$$

式中, $\sum_{\tau_{\Delta}} w$ 为网络中闭合三角形的总权重; $\sum_{\tau} w$ 为网络中关联三点组的总权重.

1.5.2 局域聚类系数

在无权网络中, Watts 和 Strogatz 提出聚类系数的另外一种定义, 该定义应用的范围更加广泛. 由于该值可用来计算每个节点的拓扑特征, 因此, 又被称为局域聚类系数. 节点的局域聚类系数反映了该节点和它直接邻居之间的集团性质, 一般而言, 近邻之间联系越紧密, 该节点的聚类系数就会越高. 某个节点 i 的局域聚类系数具体定义为^[1]

$$C_2(i) = \frac{2t_i}{k_i(k_i - 1)} \quad (13)$$

式中, t_i 为包含节点 i 的三角形数目; $k_i(k_i - 1)/2$ 为以节点 i 为中心的三点组数目.

如果直接将这个概念应用到权重网络上, 单单对拓扑结构的度量就不能反映出究竟是哪些权重的边倾向于成为三角形. 也就是说, 以前定义的聚类系数没有考虑到加权网络中有些邻居节点比其它节点更重要. 为了解决这个问题, Barrat 等定义节点 i 的加权聚类系数为^[14]

$$C_{2w}(i) = \frac{1}{s_i(k_i - 1)} \sum_{j,k} \frac{w_{ij} + w_{jk}}{2} a_{ij} a_{jk} a_{ki} \quad (14)$$

该定义考虑了三角形中 2 条边的权重 w_{ij} 和 w_{jk} , 其中, a_{ij} , a_{jk} 和 a_{ki} 表示 3 个节点之间的连接关

系, $s_i(k_i - 1)$ 为归一化因子, 以确保 $0 \leq C_{2w}(i) \leq 1$. 如果所有的权重都相等, 上式还原为节点的无权网络的聚类系数公式(13). 总体来说, 上述加权聚类系数既考虑了点 i 的邻居闭合三角形个数, 又考虑了总的相对权重. 如果 $C_{2w} > C_2$, 表示网络中闭合的关联三元组更可能由权重高的边组成; 相反地, 如果 $C_{2w} < C_2$, 则表示网络的拓扑聚集主要是由权重低的边形成的. 在后一种情况下, 由于最大部分的相互作用(通讯量或联系次数等)均不发生在关联三元组内, 很明显网络在拓扑上的聚集对网络组织只起很小的作用.

此外, Onnela 等将模体分析推广到加权网络研究中^[40], 提出另外一个局域加权聚类系数的定义

$$C_{2w}(i) = \frac{2}{k_i(k_i - 1) \sum_{j,k} \sqrt[3]{\tilde{w}_{ij} \tilde{w}_{jk} \tilde{w}_{ik}}} \quad (15)$$

该定义考虑了三角形中 3 条边的权, 而不是式(14)中的 2 条. 此外, 式(15)中的 3 个权重都应归一化后的权重, 如 $\tilde{w}_{ij} = w_{ij} / \max w_{\text{all}}$, 其中, $\max w_{\text{all}}$ 表示的是网络里面所有权重 w_{all} 的最大值. 实际上, 以上两种定义在很多情况下计算出的局域加权聚类系数是相同的, 仅仅在某些特殊情况下有区别^[40].

1.6 富人俱乐部现象

在人类社会, 有影响力的人相互之间是朋友并形成集团的现象被称为富人俱乐部. 很多实际网络中都存在着富人俱乐部现象. 比如在 Internet(互联网)中, 度大节点(富节点)之间连接往往比较多, 形成的网络子图密度也比较大^[41]. 为了描述网络这一拓扑性质, Zhou 等引入统计量 φ , 其定义为

$$\varphi(k) = \frac{2E_{>k}}{N_{>k}(N_{>k} - 1)} \quad (16)$$

式中, $N_{>k}$ 为度值大于 k 的节点的数量; $E_{>k}$ 为这些节点之间连边的数量.

最初大家认为, 如果 $\varphi(k)$ 随着 k 的增加而增加, 则可以认为网络中存在富人俱乐部现象. 然而, 进一步的研究发现, 即使是 Erdos-Renyi 网络等随机网络模型, 也可以产生 $\varphi(k)$ 随着 k 增加而增加的现象. 因此, Colizza 等提出只有当实际网络的富人俱乐部系数 $\varphi(k)$ 大于其对应的随机化网络的富人俱乐部系数 $\varphi_{\text{null}}(k)$ 时, 也就是只有当比值 $\rho(k)$ 大于 1 时, 才能认为网络中存在富人俱乐部现象^[42].

$$\rho(k) = \frac{\varphi(k)}{\varphi_{\text{null}}(k)} \quad (17)$$

图2(a)节点的大小正比于它们的强度,连边的宽度正比于它们所代表的权重.在最重要的5个节点(红色标记)之间的6条连接的权重之和为14;图2(b)最强的6条连接为黑色,它们的权重总和为21.因此,得到两者的比值为 $14/21 = 2/3$ ^[43].

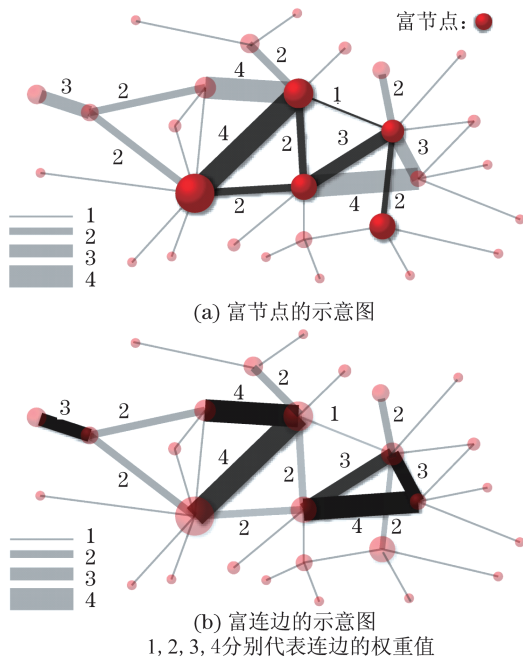


图2 权重网络的富人俱乐部系数

Fig.2 Rich-club coefficient for the weighted network

对应于加权网络,加权富人俱乐部系数主要是用来研究连接权重最大的连边是否存在于强度最大的节点之间.Opsahl 等给出了定义^[43]

$$\phi_w(k) = W_{>r} / \sum_{l=1}^{E_{>r}} W_{l,\text{rank}} \quad (18)$$

式中, $W_{l,\text{rank}}$ 为第 l 个权重最大连边的权重值.

式(18)的具体计算过程如下:首先定义节点强度大于一定值的节点为重要性节点,然后计算这些节点之间的所有连边强度之和,即上式中的 $W_{>r}$,接着计算同样数量权重最大连边的权重之和为 $\sum_{l=1}^{E_{>r}} W_{l,\text{rank}}$,最后再求出前面两项的比值.如果权重最大的连接都存在于重要节点之间,那么上面计算出的比值应为1.一般情况下,并不是所有权重最大的连接都存在于重要节点之间,因此,这个比值一般都小于1.但是,这个比值也是一个绝对量,这个量的大小是依赖于具体实际网络的.为了进行比较,很多时候都使用随机化后网络 and 实际网络进行对比,判断比值 $\rho_w(k)$ 是否大于1来作为加权网络中是否具有富人俱乐部现象的依据^[44].

$$\rho_w(k) = \frac{\phi_w(k)}{\phi_{w,\text{null}}(k)} \quad (19)$$

式中, $\phi_{w,\text{null}}(k)$ 为实际网络随机化后的加权富人俱乐部系数.

2 加权网络的随机置乱算法

当分析实际加权网络各种统计量时,很多时候很难去判定得到的结果究竟是高还是低(实际上分析无权网络的时候也是这样).例如,一个网络的聚类系数是0.2,这个值是高还是低呢?在实际应用中,往往将实际网络随机化后得到的集合当成一个比较的基准.尽管很多时候,在计算有些统计量如匹配系数的时候往往只关注其绝对值,但是,将实际网络的结果和其随机置乱网络的结果相比较其实还是有必要的,上面介绍的富人俱乐部系数就是一个很明显的例子^[45].也就是说,实际网络的统计量(如聚类系数等)是高还是低通常不应仅仅看它的绝对值,更要看它和原始网络随机置乱后的新网络对应的统计量进行比较后得到的相对结果,因此,计算网络统计量的同时很有必要了解和掌握网络的各种随机化置乱方法.

2.1 经典的随机网络

在复杂网络研究兴起以前,各种随机网络的研究就已经很成熟了,因此,有些常用的随机网络被称为经典随机网络模型.经典随机网络的概率分布一般都是均匀分布或者是泊松分布的^[1,46].尽管这些经典网络模型在很多研究中还是被作为比较的对象,而且这些网络已经有了很多非常精确的解析结果,然而,近年来的研究发现很多真实网络的度分布都是有长尾的幂律分布.幂律分布意味着网络中有一些度特别大的节点,这种特性意味着真实网络的结构特性和经典随机网络是截然不同的.因此,近年来具有代表性的满足度分布为幂律分布的Barabási-Albert模型^[2]在很多时候成为比较的对象,分析实际网络的时候往往将结果和同等规模的Barabási-Albert网络相比较.

2.2 链路置乱算法

为了弥补实际网络和经典随机网络之间的差异性,保持实际网络更多性质的随机化方法被提出来.一类方法是生成和实际网络具有相同度分布的随机网络,这种生成随机化网络的过程并不需要通过原始网络来完成,只需要已知实际网络的度分布即可^[47].对于加权网络,就需要用网络模型生成和实际网络度分布以及权重分布都相同的随机网络^[48].

另外一种方法是将原始网络通过随机化置乱连接来生成新的随机网络^[49].这种随机化的过程是在

每一步都在网络中任选4个节点,如图3(a)原始网络中节点A和B相连,C和D相连,随机化置乱以后,A和C相连,B和D相连,那么,置乱前和置乱后的网络相比,每个节点的度都是保持不变的,但是,连接关系发生了改变(图3(b)).通过不断地重复置乱过程,就能产生一个在结构上和原始网络足够随机化的副本.对于加权网络而言,权重值随着连边的置乱也失去本身的连接偏好性.置乱以后,网络的拓扑结构和权重的连接特性都发生了变化,但是,网络的度分布和权重分布都是不变的.

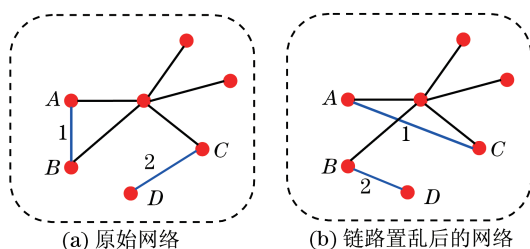


图3 原始网络和使用链路置乱算法一步置乱后的网络

Fig.3 Original network and the shuffled network using the link shuffling method

2.3 权重置乱算法

上面的随机化置乱算法本质上可以理解为对网络拓扑结构的随机化.对于权重网络,将网络随机化的时候可以不改变网络的拓扑结构,而只随机化连边的权重,这样就可以不考虑拓扑特性,单独研究网络的权重特性对于网络性质的影响^[43].具体的随机化过程如图4所示,首先是在网络中任选4个节点,节点A和B之间权重为1,C和D之间的权重为2,随机化置乱以后并不改变4个节点之间的连接关系,而是将两条连接的权重进行交换:置乱后A和B之间的权重为2,C和D之间的权重为1.

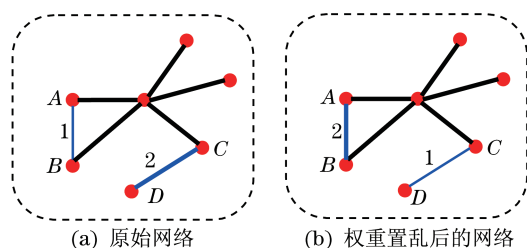


图4 原始网络和使用权重置乱算法一步置乱后的网络

Fig.4 Original network and the shuffled network using the weight shuffling method

2.4 局部权重置乱算法

按照上述方法随机置乱网络权重会出现一个问题,就是某个节点所有连边的权重之和在置乱前和置乱后发生很大变化.如果想保证每个节点的所有连边

的权重之和在置乱前后保持不变,那么对于有向权重网络可以采用如下的方法进行置乱^[43]:任选网络中某个节点A的两个输出连边,然后交换这两条连边的权重.这样交换后这两条连边的权重发生了变化,而交换前后该节点的输入权重和输出权重均没有变化.因为这一随机置乱过程都是发生在该节点上,因此是一个局部性的置乱方法.如图5所示.对于无向加权网络,在某些情况下可以先看成是双向的有向网络后,然后再使用该方法进行置乱^[14].此外,最近也有学者提出了使用马尔可夫链来随机置乱加权网络的方法^[50].

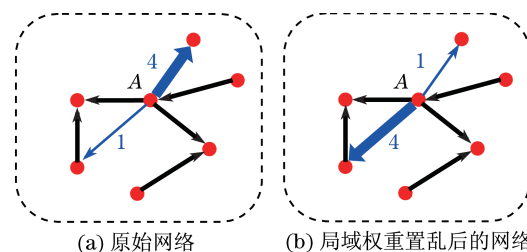


图5 原始网络和使用局部权重置乱算法一步置乱后的网络

Fig.5 Original network and the shuffled network using the local weight shuffling method

3 结语与展望

本文综述了加权网络上的常用统计量,并简要介绍了这些统计量的物理意义.同时,本文也介绍了各种加权网络的随机置乱算法,这些算法在计算实际加权网络统计量时可以生成对应的随机网络来作为参考和比较.以无权网络为代表的复杂网络研究已经兴起了十几年,尽管将复杂系统抽象成无权无向的复杂网络大大简化了系统的复杂性,但这种刻画无疑是粗糙的,因为它失去了节点之间相互作用的强度信息.通过建立加权复杂网络并分析其统计量,有利于对复杂系统进行更好的理解和刻画.近年来随着数据获取能力的提高,各种大规模社会网络中人与人交互强度的信息被提取出来^[51-52],因此,了解和掌握各种加权网络统计量是很有必要的.

限于篇幅,本文仅综述了比较常用的、比较简单的统计量,实际上加权网络中的社团结构^[53-54]和主干骨架^[55]等高级统计量和研究方法对于我们理解加权网络的复杂特性也是很有帮助的,今后的研究中有必要探索这些统计量之间的内在联系,并分析出加权网络的各种尺度特性^[56].此外,本文独立考虑了权重信息对于复杂系统的影响,实际上交互作用的方向性^[57-58]以及交互作用的时变特性^[59-60]也

是复杂系统研究中非常值得关注的两个方面,它们和交互作用的强度是有很强联系的.在今后的研究中,一个重大的挑战是如何将交互作用的强度、方向和时效特性这三者结合起来,从而更好地掌握和利用无处不在的各种复杂系统.

参考文献:

- [1] Watts D J, Strogatz S H. Collective dynamics of "small-world" networks[J]. Nature, 1998, 393 (6684): 440 - 442.
- [2] Barabási A L, Albert R. Emergence of scaling in random networks[J]. Science, 1999, 286(5439): 509 - 512.
- [3] Albert R, Barabási A L. Statistical mechanics of complex networks[J]. Review of Modern Physics, 2002, 74(1): 47 - 97.
- [4] Newman M E J. The structure and function of complex networks[J]. SIAM Review, 2003, 45(2): 167 - 256.
- [5] Boccaletti S, Latora V, Moreno Y, et al. Complex networks: structure and dynamics [J]. Physics Reports, 2006, 424(4/5): 175 - 308.
- [6] 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.
- [7] Newman M E J. Analysis of weighted networks[J]. Physical Review E, 2004, 70(5): 056131.
- [8] 朱陈平, 张永梅, 刘小廷, 等. 复杂网络稀疏性的统计物理研究综述[J]. 上海理工大学学报, 2011, 33(5): 425 - 432.
- [9] Granovetter M S. The strength of weak ties [J]. The American Journal of Sociology, 1973, 78 (6): 1360 - 1380.
- [10] Onnela J P, Saramaki J, Hyvonen J, et al. Structure and tie strengths in mobile communication networks [J]. Proceedings of the National Academy of Sciences of the United States of America, 2007, 104(18): 7332 - 7336.
- [11] Newman M E J. Scientific collaboration networks: II shortest paths, weighted networks, and centrality[J]. Physical Review E, 2001, 64(1): 016132.
- [12] Almaas E, Kovacs B, Vicsek T, et al. Global organization of metabolic fluxes in the bacterium Escherichia coli[J]. Nature, 2004, 427(6977): 839 - 843.
- [13] Nornlund C. Identifying regular blocks in valued networks: a heuristic applied to the St. Marks carbon flow data, and international trade in cereal products [J]. Social Networks, 2007, 29(1): 59 - 69.
- [14] Barrat A, Barthelemy M, Pastor-Satorras R, et al. The architecture of complex weighted networks[J]. Proceedings of the National Academy of Sciences of the United States of America, 2004, 101(11): 3747 - 3752.
- [15] Barrat A, Barthelemy M, Vespignani A. Modeling the evolution of weighted networks[J]. Physical Review E, 2004, 70(6): 1 - 13.
- [16] Thadakmalla H P, Albert R, Kumara S R T. Search in weighted complex networks[J]. Physical Review E, 2005, 72(6): 066128.
- [17] Radicchi F, Ramasco J J, Fortunato S. Information filtering in complex weighted networks[J]. Physical Review E, 2011, 83(4): 046101.
- [18] Yan G, Zhou T, Wang J, et al. Epidemic spread in weighted scale-free networks [J]. Chinese Physics Letters, 2005, 2(2): 510 - 513.
- [19] 田柳, 狄增如, 姚虹. 权重分布对加权网络效率的影响[J]. 物理学报, 2011, 60(2): 028901.
- [20] Wang W X, Wang B H, Hu B, et al. General dynamics of topology and traffic on weighted technological networks [J]. Physical Review Letters, 2005, 94 (18): 188702.
- [21] 汪秉宏, 王文旭, 周涛. 交通流驱动的含权网络[J]. 物理, 2006, 35(2): 227 - 233.
- [22] Lu X B, Wang X F, Li X, et al. Synchronization in weighted complex networks: heterogeneity and synchronizability[J]. Physica A, 2006, 370 (2): 381 - 389.
- [23] Lv L, Zhou T. Link prediction in weighted networks: the role of weak ties[J]. Europhysics Letters, 2010, 89 (1): 18001.
- [24] Costa L F, Rodrigues F A, Trivieso G, et al. Characterization of complex networks: a survey of measurements[J]. Advances in Physics, 2007, 56(1): 167 - 242.
- [25] Fan Y, Li M, Chen J, et al. Network of econophysicists: a weighted network to investigate the development of econophysics[J]. International Journal of Modern Physics B, 2004, 18 (17/18/19): 2505 - 2511.
- [26] Li M, Fan Y, Chen J, et al. Weighted networks of scientific communication: the measurement and topological role of weight[J]. Physica A, 2005, 350 (2/3/4): 643 - 656.
- [27] 吴金闪, 狄增如. 从统计物理学看复杂网络研究[J]. 物理学进展, 2004, 24(1): 18 - 46.
- [28] Li C, Chen G. A comprehensive weighted evolving network model[J]. Physica A, 2004, 343: 288 - 294.
- [29] 李艳萍. 加权复杂网络中传播问题的研究[D]. 西安: 西安理工大学, 2008.
- [30] Newman M E J. Assortative mixing in networks[J]. Physical Review Letters, 2002, 89(20): 208701.
- [31] Newman M E J. Mixing patterns in networks [J].

- Physical Review E, 2003, 67(2): 026126.
- [32] Leung C C, Chau H F. Weighted assortative and disassortative networks model[J]. Physica A, 2007, 378(2): 591 – 602.
- [33] Pastor-Satorras R, Vazquez A, Vespignani A. Dynamical and correlation properties of the Internet[J]. Physical Review Letters, 2001, 87(25): 258701.
- [34] Milgram S. The small world problem[J]. Psychology Today, 1967, 2(1): 60 – 67.
- [35] Freeman L. Centrality in social networks: conceptual clarification[J]. Social Networks, 1979, 1(3): 215 – 239.
- [36] Dijkstra E W. A note on two problems in connexion with graphs[J]. Numerische Mathematik, 1959, 1: 269 – 271.
- [37] Opsahl T, Agneessens F, Skvoretz J. Node centrality in weighted networks: generalizing degree and shortest paths[J]. Social Networks, 2010, 32(3): 245 – 251.
- [38] Wasserman S, Faust K. Social network analysis: methods and applications[M]. Cambridge: Cambridge University Press, 1994.
- [39] Opsahl T, Panzarasa P. Clustering in weighted networks[J]. Social Networks, 2009, 31(2): 155 – 163.
- [40] Onnela J P, Saramaki J, Kertesz J, et al. Intensity and coherence of motifs in weighted complex networks[J]. Physical Review E, 2005, 71(6): 065103.
- [41] Zhou S, Mondragon R J. The rich-club phenomenon in the Internet topology[J]. IEEE Communications Letters, 2004, 8(3): 180 – 182.
- [42] Colizza V, Flammini A, Serrano M A, et al. Detecting rich-club ordering in complex networks[J]. Nature Physics, 2006, 2(2): 110 – 115.
- [43] Opsahl T, Colizza V, Panzarasa P, et al. Prominence and control: the weighted rich-club effect[J]. Physical Review Letters, 2008, 101(6): 168702.
- [44] Serrano M A. Rich-club vs rich-multipolarization phenomena in weighted networks[J]. Physical Review E, 2008, 8(2): 026101.
- [45] Amaral L A N, Guimera R. Lies, damned lies and statistics[J]. Nature Physics, 2006, 2(2): 75 – 76.
- [46] Erdos P, Renyi A. On random graphs I [J]. Publicationes Mathematicae, 1959, 6: 290 – 297.
- [47] Newman M E J, Strogatz S H, Watts D J. Random graphs with arbitrary degree distributions and their applications [J]. Physical Review E, 2001, 64(2): 026118.
- [48] Snijders T A B. The statistical evaluation of social network dynamics [J]. Sociological Methodology, 2001, 31(1): 361 – 395.
- [49] Maslov S, Sneppen K. Specificity and stability in topology of protein networks[J]. Science, 2002, 296(5569): 910 – 913.
- [50] Ansmann G, Lehnertz K. Constrained randomization of weighted networks[J]. Physical Review E, 2011, 84(2): 026103.
- [51] 王科, 胡海波, 汪小帆. 中国高校电子邮件网络实证研究[J]. 复杂系统与复杂性科学, 2008, 5(4): 66 – 74.
- [52] Onnela J P, Saramaki J, Hyvonen J, et al. Analysis of a large-scale weighted network of one-to-one human communication[J]. New Journal of Physics, 2007, 9(6): 179.
- [53] Serrano M A, Boguna M, Vespignani A. Extracting the multiscale backbone of complex weighted networks[J]. Proceedings of the National Academy of Sciences of the United States of America, 2010, 106(6): 6483 – 6488.
- [54] Chen D, Shang M, Lv Z, et al. Detecting overlapping communities of weighted networks via a local algorithm[J]. Physica A, 389(19): 4177 – 4187.
- [55] Lou X, Suykens J A K. Finding communities in weighted networks through synchronization [J]. Chaos, 2011, 21(4): 043116.
- [56] Zhang J, Xu X K, Li P, et al. Node importance for dynamical process on networks: a multiscale characterization [J]. Chaos, 2011, 21(1): 016107.
- [57] Kovanen L, Saramaki J, Kaski K. Reciprocity of mobile phone calls[J]. Dynamics of Socio-Economic Systems, 2011, 2(2): 138 – 151.
- [58] Peruani F, Tabourier L. Directedness of information flow in mobile phone communication networks[J]. PLoS ONE, 2011, 6(112): e28860.
- [59] Wu Y, Zhou C, Xiao J, et al. Evidence for a bimodal distribution in human communication[J]. Proceedings of the National Academy of Sciences of the United States of America, 2010, 107(44): 18803 – 18808.
- [60] Miritello G, Moro E, Lara R. Dynamical strength of social ties in information spreading [J]. Physical Review E, 2011, 83(4): 045102.