



新型双模式加权机制及其对网络计算的影响

李慧嘉 黄照词 王文璇 夏承遗

A new tunable weighting strategy for enhancing performance of network computation

Li Hui-Jia Huang Zhao-Ci Wang Wen-Xuan Xia Cheng-Yi

引用信息 Citation: *Acta Physica Sinica*, 70, 178901 (2021) DOI: 10.7498/aps.70.20210453

在线阅读 View online: <https://doi.org/10.7498/aps.70.20210453>

当期内容 View table of contents: <http://wulixb.iphy.ac.cn>

您可能感兴趣的其他文章

Articles you may be interested in

节点中心性对复杂网络传播模式的影响分析

Analysis of the effect of node centrality on diffusion mode in complex networks

物理学报. 2017, 66(12): 120201 <https://doi.org/10.7498/aps.66.120201>

基于复杂网络动力学模型的无向加权网络节点重要性评估

Evaluation methods of node importance in undirected weighted networks based on complex network dynamics models

物理学报. 2018, 67(9): 098901 <https://doi.org/10.7498/aps.67.20172295>

复杂网络电输运性能与通信序列熵之间的关联

Correlation between the electrical transport performance and the communicability sequence entropy in complex networks

物理学报. 2019, 68(11): 118901 <https://doi.org/10.7498/aps.68.20190230>

基于区域密度曲线识别网络上的多影响力节点

Identifying multiple influential nodes based on region density curve in complex networks

物理学报. 2018, 67(19): 198901 <https://doi.org/10.7498/aps.67.20181000>

一类任意 $m \times n$ 阶矩形网络的电特性

Electrical properties of an arbitrary $m \times n$ rectangular network

物理学报. 2020, 69(2): 020502 <https://doi.org/10.7498/aps.69.20191303>

基于通信序列熵的复杂网络传输容量

Quantifying complex network traffic capacity based on communicability sequence entropy

物理学报. 2021, 70(7): 078902 <https://doi.org/10.7498/aps.70.20201300>

新型双模式加权机制及其对网络计算的影响*

李慧嘉^{1)†} 黄照词¹⁾ 王文璇¹⁾ 夏承遗²⁾¹⁾ (北京邮电大学理学院, 北京 100876)²⁾ (天津理工大学计算机与通信工程学院, 天津 300384)

(2021 年 3 月 9 日收到; 2021 年 4 月 2 日收到修改稿)

高效率的网络分析方法对于分析、预测和优化现实群体行为具有重要的作用, 而加权机制作为网络重构化的重要方式, 在生物、工程和社会等各个领域都有极高的应用价值. 虽然已经得到越来越多的关注, 但是现有加权方法数量还很少, 而且在不同拓扑类型和结构特性现实网络中的效果和性能有待继续提高. 本文提出了一种新型的双模式加权机制, 该方法充分利用网络的全局和局部的重要拓扑属性 (例如节点度、介数中心性和紧密中心性), 并构建了两新型的运行模式: 一种是在原始模式中通过增加桥边的权重来提高同步能力; 另一种是在逆模式中通过弱化桥边的权重来提高聚类检测. 此外, 该加权机制仅受单一参数 α 的影响, 非常便于调控. 在人工基准网络和现实世界网络中的实验结果均验证了该模型的有效性, 可以广泛应用于大规模的现实世界网络中.

关键词: 复杂网络, 加权策略, 通信邻域图, 聚类结构**PACS:** 89.75.Hc, 89.75.Fb**DOI:** 10.7498/aps.70.20210453

1 引言

复杂网络理论^[1-3]是一种分析大规模复杂系统的有力工具, 在过去十年里已经涌现出一系列相关研究成果, 例如聚类划分技术、疾病传播和免疫策略、交通驱动模型、同步分析、级联失效、异常值搜索算法等等. 在这些成果中, 拓扑计算和结构分析是研究复杂网络内在特性的最有效途径, 并使得理解网络功能特征成为可能. 其中, 最重要的拓扑特征之一是聚类结构^[3-8]. 它在探测聚类时将节点分成若干子集, 子集之间应有清晰的边界约束, 同一组成员之间比不同组成员拥有更多的关联. 聚类信息有助于理解和分析网络的结构特征和功能特性.

既有研究已提出了许多网络聚类探测算法, 其中 Newman 等^[9,10]提出的模块度函数 Q 是最为常

用的一种探测指标, 用以评价聚类划分^[2]的好坏. 给定一个划分 $P_k = (G_1, G_2, \dots, G_K) = ((V_1, E_1), (V_2, E_2), \dots, (V_K, E_K))$, K 是聚类的数目, 模块度函数 Q ^[9] 定义如下:

$$Q = \sum_{s=1}^K \left[\frac{l(V_s, V_s)}{2L} - \left(\frac{l(V_s, V_s) + l(V_s, \bar{V}_s)}{2L} \right)^2 \right], \quad (1)$$

其中, L 是网络中边的总数目, $l(V_s, V_s)$ 和 $l(V_s, \bar{V}_s)$ 分别代表聚类 s 内部以及 s 与网络中剩余部分之间的边的数目. 通常来说, Q 值较大意味着该聚类结构较好^[11-16]. 此外, Boccaletti 等^[17]根据网络中的同步局域化理论提出了一种新的动态聚类检测方法 (opinion changing rate model, OCR). 他们将改进的 Kuramoto 同步模型应用到网络的每个节点中, 利用社团内部连接紧密而同步速度快的特征, 利用权重技术使得网络聚类之间的边权逐渐降

* 北京邮电大学提升科技创新能力行动计划 (批准号: 2020XD-A01-2) 和国家自然科学基金 (批准号: 71871233) 资助的课题.

† 通信作者. E-mail: hjli@bupt.edu.cn

低,直到网络聚类划分获得最大的模块度 Q 值. 复杂网络中另一个广泛存在的现象是同步^[18–24]. 在连通网络中,每个节点通过网络连接,在动力系统的演化下,与网络其他节点进行耦合. 在足够强的耦合作用下,节点的状态会随着时间的延长达到同步(状态一致). 在许多加权网络中,可以调整网络的结构和权重,以提升节点状态获得达到同步的能力. 事实上,已有研究表明,在许多社会和生物网络研究中,同步行为会在调节系统和组织功能中发挥重要作用.

随着数据收集技术的进步,诸如互联网、国际贸易网络等现实网络的规模极速增长,如何处理这些大规模网络已经成为巨大挑战. 传统的指数型算法,甚至是多项式时间算法,也已难以适用于大规模网络的计算与分析. 为此,研究人员可以尝试另辟蹊径,如通过改变拓扑性质来提升网络的计算效率. 但是在许多实际应用中,网络结构通常是固定的,一般不可能对边进行重连. 为了克服这一困难,可以为网络连边分配合适的权重,用来异化不同属性的边,以达到提升计算性能的目的,如弱化类间边的权重可以提高聚类探测的效率;而在信息传播研究中提高重负荷边的流通负载(通常用权重表示)可以降低网络拥塞.

当前,加权方法已经得到越来越多的关注^[21–30],但是总体来说数量还很少,而且对于不同拓扑类型和结构特性的现实网络,特定加权算法的效果和性能仍不清楚,还有待继续分析和验证. 而且现有方法(在第2节详细说明)通常仅利用一种节点中心度来定义网络边权,对于结构相对均匀的网络,效果不理想;而且对于相对稀疏的现实网络,调控参数的影响不明显. 另外,目前加权之后的动态网络中的各种隐藏特征,如聚类结构和同步能力之间的相互关系,仍未能得到很好的解释和分析.

本文提出一种新型的双模式加权机制,用来解决上述问题. 从一个简单的无权无向网络开始,通过加权算法,利用网络的全局和多种局部信息(包括节点的介数、度、紧密中心度等),将网络转化为能够提高聚类 and 同步能力的加权有向网络. 这里的加权机制包括两种模式:在原始模式下,它通过增加桥边的权重以提高网络的同步能力;而在逆模式中,则通过弱化桥边权重以提高聚类的检测效率. 该加权机制只受一个参数 α 的控制,可以很方便地进行调控. 对于聚类探测,通过在人工基准网络

(包括 GN 基准和 LFR 基准),以及若干现实网上进行测试,验证了本文提出的加权机制不仅有助于提高探测精度,而且还能缓解模块度优化的分辨率限制问题. 关于同步性,通过在具有不同结构特性的无标度网络和 Watts-Strogatz 网络上进行测试,验证了加权机制在提升网络同步性能上的高效表现. 通过比较,本文提出的加权机制在提高网络计算能力方面要优于其他以往发表的加权方法.

本文第1节和第2节分别梳理复杂网络的基本理论及国内外相关的工作;第3节引入本文使用的加权机制并提出对应的双模式加权算法;第4节对加权机制如何增强网络同步性能及其对聚类结构探测的影响进行详细分析;第5节对加权机制分别运用于人工网络和大量现实世界网络进行试验;第6节对全文进行总结,并提出一些未来研究中值得深入考虑的问题.

2 相关工作

基于现实数据重构复杂网络是了解和控制复杂系统的基本方式,而作为网络重构的重要组成部分,网络加权已经开始得到越来越多的关注. Chavez 等^[21]提出了一种通过拉普拉斯矩阵的非对角线元素 l_{ij} 来增强同步能力的加权方案,其加权后的元素 l_{ij}^w 定义如下:

$$l_{ij}^w = \sigma \frac{B_{ij}^\alpha}{\sum_u B_{iu}^\alpha}, \quad (2)$$

其中, α 为调控参数, B_{ij} 是节点 i 和 j 之间的边介数中心度, $\alpha \approx 1$ 相当于最优的同步能力. 与之相似, Wang 等^[22]提出网络可以被视为是加权无向网络和加权有向网络的叠加. 基于现实网络的特征,他们假设了一个合适的梯度场,并定义权重 l_{ij}^w 如下:

$$l_{ij}^w = \sigma \frac{k_j^\alpha}{\sum_u k_u^\alpha}, \quad (3)$$

Charvez 等和 Wang 等提出方法具有一定程度的相似性,但后者利用的不是介数中心度,而是基于节点的度. 随着 α 值的增大,网络同步化总是会随之提升. 然而,同时应当注意避免 α 值过高,否则会导致网络断开和分裂. 为了解决这个问题, Jalili 等^[23]在节点介数中心度方法的基础上提出了一种新的加权方法,其中权重 l_{ij}^w 定义如下:

$$l_{ij}^w = s \frac{B_{ij}(1 + B_j^\alpha)}{\sum_u B_{iu}(1 + B_u^\alpha)}, \quad (4)$$

其中, $B_j B$ 和 $B_{ij} B$ 分别代表节点 j 与边 e_{ij} 之间的介数中心度. 该加权机制被用来提高网络的同步性能. 但是上述加权算法仅仅利用一种节点中心度来定义网络边权, 对于结构相对均匀的网络, 效果不理想. 而且对于相对稀疏的现实网络, 调控参数的影响不明显.

最近, Khadivi 等 [31] 讨论了一种新型加权机制, 用以缓解网络聚类中的模块度缺陷, 即分辨率极限 [32] 和极端限制 [33] 等问题. 该加权机制利用了网络的局部和全局信息, 其中每条边的权重定义如下:

$$W_{ij} = \frac{B_{ij}^{-\alpha} \cdot C_{ij}^\beta}{\sum_{k \neq m} B_{km}^{-\alpha} \cdot c_{km}^\beta}, \quad (5)$$

其中, c_{ij} 定义为共同邻居比例, B_{ij} 是边介数中心度. 利用 Clauset-Newman-Moore 方法 [10] 进行试验, 验证了该加权机制能够显著地优化社团探测的速度与精度. 但是仿真分析也表明, 当网络规模较小时, 其加权效果不明显. 而且虽然该方法可以用来加强社团内部的团内边, 但是却无法明显降低团间边的权重, 从而阻碍了该加权算法的进一步应用. 另外现有的网络加权方法还包括模拟退火算法 [25]、文献计量方法 [26]、随机游走算法 [27]、级联动力学方法 [28]、变分贝叶斯方法 [29] 和概率推断方法 [30] 等. 但总体来说, 当前加权算法的数量还很少, 而且对于不同拓扑类型和结构特性的现实网络, 特定加权算法的效果和性能仍不清楚, 还有待继续分析和验证.

3 复杂网络基本理论

3.1 基本定义

假设 $G(V, E)$ 是一个无向连通图, V 为节点集, E 为边界集. 这里考虑三个最常用的衡量拓扑结构和信息流通的中心度指标.

节点度 节点度 k_i [34] 表示节点 i 相连的边的数量, 根据邻接矩阵 $A = (a_{ij})$ 将其定义为

$$k_i = \sum_{j \in V} a_{ij}, \quad (6)$$

其中, a_{ij} 是节点 i 和节点 j 之间的邻接矩阵值. 还

可以将节点度扩展到加权网络得到节点强度的概念.

节点强度 节点强度 s_i [35] 是节点 i 相关联的边的权重总和, 定义如下:

$$s_i = \sum_{j \in V} w_{ij}, \quad (7)$$

其中, w_{ij} 为节点 i 和节点 j 之间的边的权重值.

介数中心度 一个节点在信息传播中的重要性可以通过其介数中心度 [36] 来计算, 定义为

$$B_u = \sum_{i \neq u \neq j \in V} \frac{\sigma_{ij}(u)}{\sigma_{ij}}, \quad (8)$$

其中, $\sigma_{ij}(u)$ 为节点 i 和节点 j 之间的通过节点 u 的最短路径的数量, σ_{ij} 是节点 i 和节点 j 之间的最短路径的总数量. 介数中心度的概念还可以拓展到边. 边 e_{ij} 的介数中心度 B_{ij} 表示通过该边的最短路径的数量, 在信息传播研究中常用来表示流通负载.

紧密中心度 (closeness) 节点 i 的紧密中心度 [36] 定义为一个网络中其他节点到节点 i 的平均距离的倒数

$$C_i = \sum_{j \in V} \frac{n-1}{D_{ij}}, \quad (9)$$

其中, D_{ij} 是节点 i 和节点 j 之间最短路径的长度; C_i 用来测量一个节点与其他节点的 (平均) 紧密程度. 为了定义该中心性与最短路径的直接 (正比例) 关系, 还可以将其定义为 C_i 的倒数: $C_i = \frac{1}{C_i} = \frac{1}{n-1} \sum_{j \in V} D_{ij}$.

k -邻域图 k -邻域图 [37,38] 定义为将节点 i 和节点 j 连接起来的最短路径上的点的集合, 其中 i 和 j 之间的最短距离值小于等于 k . 由于邻域关系是不对称的 ($i \rightarrow j$), 因此根据 k -邻域图的定义得到的可能是一个有向子图. 在特定网络中, 节点 i 的 k -邻域图被构造为特定的相似性度量 (这里指最短路径距离) 下显示与 i 最相似的子图. 然而在大多数应用中, k 的值难以直接给定, 以致其应用领域受到限制.

3.2 双模式加权算法

1) 通信邻域图

受到 k -邻域图定义的启发, 为了增强网络信息流, 本文提出了一个新的拓扑定义——通信邻域

图 (communication neighbor graph, CNG), 并进一步提出一种新的加权算法. 通信邻域图的具体定义如下.

通讯邻域图 假设 $G(V, E)$ 是一个无自循环的无向网络, 其中节点的数量 $\|V\| = N$. 对于 G 中的节点来说, ζ_{ij} 是节点 i 与节点 j 之间的最短路径 (路径为节点序列集合). 节点 i 通过节点 j 的通信邻域图用 $\Pi_{i \rightarrow j}$ 表示, 其定义为: 给定节点 u , 如果满足通过节点 i 与 u 之间的最短路径长度大于节点 j 与 u 之间的最短路径长度, 即 $l(\zeta_{iu}) > l(\zeta_{ju})$, 那么节点 u 属于 $\Pi_{i \rightarrow j}$. 如果节点 i 和节点 u 之间存在一条以上的最短路径, 那么就认为 ζ_{iu} 是其中任意一条. 根据通信邻域图 CNG 的定义, 对于特定节点 u , 假定节点 i 与节点 u 之间有边相连, 如果 ζ_{iu} 经过节点 j , 那么节点 u 就属于 $\Pi_{i \rightarrow j}$.

注意到通信邻域图 CNG 是非对称的, 即一般情况下 $\Pi_{i \rightarrow j}$ 与 $\Pi_{j \rightarrow i}$ 是不同的. 对于任意一对相邻的节点 i 和 j 来说, 节点 j 一定属于 $\Pi_{i \rightarrow j}$. 另外, 任一节点 $u \neq i$ 属于且仅属于节点 i 的一个通讯邻域图. 图 1(a) 给出了两个相邻节点的通信邻域图的例子. 当网络是稠密且规模不大时, 该定义是非常直观的.

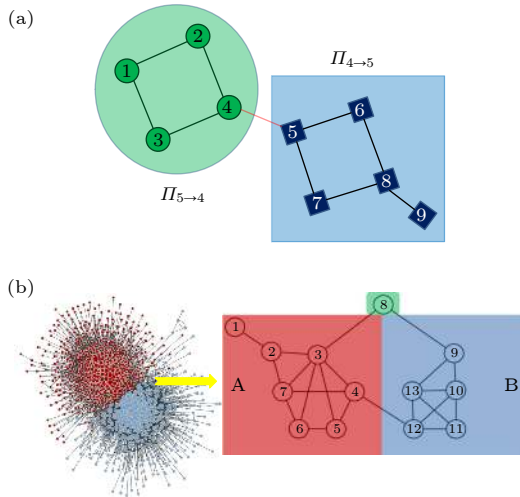


图 1 (a) 节点 4 和节点 5 之间边对应的通讯邻域图; (b) 网络中的桥边
Fig. 1. (a) Communication neighborhood graph corresponding to the edge between node 4 and node 5; (b) bridge side in the network.

2) 加权算法

如果边权与流量成正比, 那么为了增强连边上的信息流, 需要研究最有可能出现瓶颈的桥点和桥边. 因为桥点和桥边数目较少且处于两个类之间,

它们往往无法承载端点需要发送的全部信息, 如图 1(b) 所示. 如果节点具有大的度、介数中心性和紧密中心性, 那么它更有可能具有高信息吞吐量, 因此这些节点及其所连的边更可能成为瓶颈. 在本文模型中, 这种可能性被假设为与单一实数成正比, 即定义图中每个节点的 γ 值.

在给定通信邻域图 CNG 的定义后, 接下来介绍一种新型的加权机制. 令 $\gamma(u) \in R$ 表示属于 $\Pi_{i \rightarrow j}$ 中节点 u 的一个特定指标, $\Gamma(\Pi_{i \rightarrow j}) = \{\gamma(u_1), \gamma(u_2), \dots, \gamma(u_k)\}$ 且是 $\Pi_{i \rightarrow j}$ 中所有节点的 γ 值集合, 其中 $u \in \Pi_{i \rightarrow j}$. 那么, 本文的加权机制定义如下:

$$W_{ij} = \psi(\Gamma(\Pi_{i \rightarrow j})), \quad (10)$$

其中, ψ 为任意实数集中的单一实数相关的函数. 这是一种具有代表性的加权方法, 可以用来加强与网络通信相关的多种计算与应用.

本文加权算法旨在重点增强瓶颈节点的各种中心性值, 使其与加权权重成正比, 以便增强网络的整体通信能力. 定义 ψ 为一个集合所有元素的加和, 即如果 $\mathbf{X} = \{x_1, x_2, \dots, x_N\}$, 那么 $\psi(\mathbf{X}) = \sum_{i=1}^N x_i$, 同时 $\psi(\mathbf{X}) = \sum_{i=1}^N x_i$. 定义节点 u 的加权参数为

$$\gamma(u) = \left[\frac{k_u(\varepsilon + B_u)}{\tilde{C}_u} \right]^\alpha = \xi^\alpha, \quad (11)$$

其中, $\alpha \in R$, 且 k_u , B_u 和 \tilde{C}_u 分别为网络 G 中节点的度、介数中心性和紧密中心度 C_u 的倒数. ε 设定为 0.01, 以避免 γ 值为 0. 设计加权参数 γ 是为了反映瓶颈节点 u 的各项参数, 包括度 k_u , 介数中心性 B_u 和紧密中心度 C_u , 如果各项参数比较大, 则 γ 也比较大, 节点 u 成为瓶颈节点的概率也会很大. 那么, 节点 j 到节点 i 的权重为

$$\begin{aligned} W_{ij} &= \frac{\psi(\Gamma(\Pi_{i \rightarrow j}))}{\sum_m \psi(\Gamma(\Pi_{i \rightarrow m}))} = \frac{\sum_{u \in \Pi_{i \rightarrow j}} \gamma(u)}{\sum_m \sum_{u \in \Pi_{i \rightarrow m}} \gamma(u)} \\ &= \frac{\sum_{u \in \Pi_{i \rightarrow j}} \left[\frac{(\varepsilon + B_u) k_u}{\tilde{C}_u} \right]^\alpha}{\sum_m \sum_{u \in \Pi_{i \rightarrow m}} \left[\frac{(\varepsilon + B_u) k_u}{\tilde{C}_u} \right]^\alpha}, \end{aligned} \quad (12)$$

其中 $\Pi_{i \rightarrow j}$ 表示节点 i 通过节点 j 的通信邻域图. (12) 式表示的加权机制的原理为当传输线路 $i \rightarrow j$ 比较重要时, 其通信邻域图 $\Pi_{i \rightarrow j}$ 的规模也

会相应比较大. 则当 $\alpha > 0$ 时, (12) 式的分子项 $\sum_{u \in \Pi_{i \rightarrow j}} \gamma(u)$ 和相应的权重 W_{ij} 也会相应比较大, 从而实现了加权机制增强重要通信边的权重的目的. 这里, 分母项的设置是为了实现标准化, 权重矩阵的对角线元素被设置具有零行和属性, 即 $W_{ii} = -\sum_k W_{ik} = 1$.

在加权策略 ((12) 式) 中, 分母的标准化项使得入度更加均匀. 另外, γ 函数 ((11) 式) 增强了以高 γ 值节点为起点的路径, 并且进一步使得网络流量更加趋向于流向这类节点. 这个过程使节点的负载分配更加倾向于平均. 总而言之, 本文提出的加权策略使负载分布更加均匀化, 同时降低了平均路径长度. 本文加权算法具体流程如算法 1 所示. 本算法由于需要计算节点的介数、度、紧密度三种中心性度量, 时间复杂度分别为 $O(n^3)$, $O(n^2)$ 和 $O(n^3)$, 因此总的时间复杂度为 $O(n^3)$.

算法 1 网络加权算法流程

输入: 网络 G , 其节点数为 n , 边数为 m ; 加权参数 α .

输出: 网络加权矩阵 W .

1) 根据“通讯邻域图”的定义计算网络中每条边的通讯邻域图;

2) 计算 k_u , B_u 和 \tilde{C}_u , 分别为网络 G 中节点的度、介数中心性和紧密中心度 C_u 的倒数;

3) 根据 (6) 式和 (11) 式计算节点 u 的加权值 $\gamma(u)$;

4) 根据 (12) 式更新网络中边的加权值 w_{ij} .

需要说明的是, 本文提出的加权机制是使用虚拟方式改变拓扑结构, 但是仅通过加权来达到网络重连的效果是不大可能的. 例如, 在环形规则网络中, 所有节点的 γ 值是相等的, 加权机制并不能增强网络的传输容量. 因此, 本文的加权机制旨在通过降低瓶颈节点的限制, 使网络层次结构趋于平坦以提高传输容量. 这种加权机制对具有非均匀层次结构和瓶颈节点的网络 (如无标度网络) 来说很有效率. 反之, 均匀网络或无瓶颈节点的网络则不具备这种加权机制的优势. 此外, 本文的加权策略还有一个重要优点: 只受一个参数调控, 即参数 α , 可以用来控制加权权重的异质程度 (即非均匀性). 这里有三种典型的情况:

当 $\alpha > 0$ ($\alpha < 0$), 在通讯邻域图中, 具有高 ξ 值的节点对信息传输会有更多 (更少) 的贡献. 特

别地, 在同质 (均匀) 网络中, 节点呈现出均匀的负载和度分布, 那么节点的 ξ 值和随之得到的加权值也会变得均匀. 在这种情况下, 改变 α 不会导致权重和网络传输效率发生显著变化. 类似地, 当网络负载分布和度分布是异质 (非均匀) 时, 不同的 α 会导致权重和网络传输效率发生显著变化. 因此, 在异质 (同质) 网络中, α 会有更多 (更少) 的影响.

当 $\alpha = 0$ 时, γ 值将退化为对所有节点 u 都有 $\gamma(u) = 1$. 值得注意的是, 这种情况与无权网络或者标准化仅仅应用于节点度的情况不同. 在这种情况下, 仅考虑每个通信邻域图的节点数量, 它在 γ 值分布的方差较小的网络中将会相当有效, 即 γ 值的同质性.

由于受标准化因子和通讯邻域图节点数目的影响, $\alpha < 0$ 并不会产生负值, 但是会使得信息传输能力下降 (即具有高 ξ 值的节点对信息传输会有更少的贡献). 根据本文框架, 我们提出了利用转置权重矩阵以降低传输能力的逆加权方案. 为了能够更方便地转换, 权重矩阵被定义为

$$W^* = \rho W + (1 - \rho) W^T, \quad (13)$$

其中, 参数 $0 \leq \rho \leq 1$, 表示传输程度. 例如, 当 $\rho = 1$ 意味着高传输流量, 而 $\rho = 0$ 意味着低传输流量.

逆模式加权机制 在 (12) 式的加权机制中, 设定参数 $\alpha < 0$, 定义为逆模式加权机制. 根据 (12) 式, 在逆模式加权机制中, 即具有高 ξ 值的节点对信息传输会有更少的贡献, 从而会使得网络流量更加趋向于避开这类节点, 使得信息传输能力下降. 为了能够更方便地转换, 权重矩阵被定义为 (13) 式. 逆模式加权算法如算法 2 所示.

算法 2 逆模式加权算法流程

输入: 网络 G , 其节点数为 n , 边数为 m ; 加权参数 $\alpha < 0$; 传输程度参数 $0 \leq \rho \leq 1$.

输出: 网络加权矩阵 W .

1) 根据“通讯邻域图”的定义计算网络中每条边的通讯邻域图;

2) 根据 (11) 式计算节点 u 的加权值 $\gamma(u)$;

3) 根据 (12) 式更新网络中边的加权值 w_{ij}^* ;

4) 根据 (13) 式更新网络中边的加权值 w_{ij}^* .

为了测试参数 α 的影响, 将其应用到著名的 BA 网络中, 分别考虑 $\alpha = 2$ 和 $\alpha = 6$ 两种情况, 结果如图 2 所示. 可以看出, 受加权机制的影响, 节

点强度和节点度会出现明显异化. 此外, 当 $\alpha = 6$ 时, 节点强度和节点度的关系的分布相较于 $\alpha = 2$ 时更加分散. 这是因为 α 的值越大, 更多的边会得到较大的权重值. 该结果与前面加权机制的分析相符合. 因此, 通过本文的加权策略, 权重分布、节点强度以及度的关系均依赖于参数 α 的值, 该加权网络比相应的二进制无权网络提供了更多的信息.

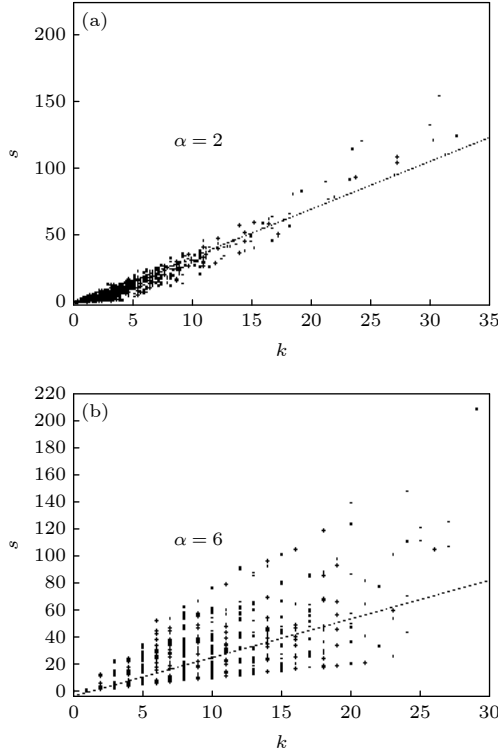


图2 参数 α 取不同值时 BA 模型中节点强度 s 和节点度 k 之间的关系 ($N = 5000$) (a) $\alpha = 2$; (b) $\alpha = 6$

Fig. 2. Relationship between node strength s and node degree k in BA model ($N = 5000$) with different values of parameter α : (a) $\alpha = 2$; (b) $\alpha = 6$.

4 复杂网络的属性分析

许多研究 [20,39,40] 发现, 网络中信息传输的难易程度可以解释为同步能力, 并且它们都被网络拓扑矩阵的特征值所限制. 由此直接引出本文提出的加权方案, 它可以试图通过改善信息流以增强同步能力. 然而, 并不是所有的应用都需要提高同步能力, 有些反而需要降低, 比如降低网络的同步性有助于辨别网络中的聚类结构 [17]. 本节将分析加权机制对网络同步能力的作用, 并进一步分析其对聚类探测及其效率的影响, 从而有效验证加权机制的高效性. 除了同步和聚类分析, 本文加权机制还可以有效应用于网络流行病传播关键路径识别、免疫策略

分析、演化博弈分析、路由策略设计、交通拥塞预防等多个方面.

4.1 同步性分析

首先说明加权机制增强网络同步性能的原因. 设 $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_N$ 为加权矩阵 \mathbf{W} 的 N 个特征值. 根据文献 [41], 矩阵 \mathbf{W} 的特征值比 λ_N/λ_2 越小, 则意味着该网络具有更大的同步能力. 如上所述, 网络通信的能力可以被解释为在网络中信息流动的难易程度. 更确切地说, 可以将动态方程写作:

$$\dot{x}_1 = F(x_i) - \frac{\sigma}{\sum_k \psi(\Gamma(G^{i \rightarrow k}))} \times \sum_{j=1}^N \psi(\Gamma(\Pi_{i \rightarrow j})) H(x_j - x_i), \quad (14)$$

通过网络加权, 总是可以将 \mathbf{W} 的对角线元素标准化为 1, 从而防止任意大或者任意小的耦合值. 权重矩阵 \mathbf{W} 是不对称的, 并且一般而言非对称矩阵的特征值非常复杂. 下面的推论证明了非对称矩阵 \mathbf{W} 具有有界的实数特征值.

推论 1 加权机制 ((12) 式) 得到的权重矩阵的所有特征值是实数, 且值介于 0—2 之间.

证明 权重矩阵 \mathbf{W} 对应的拉普拉斯矩阵 \mathbf{L}^w , 写作 $\mathbf{L}^w = \mathbf{M}\mathbf{Q}$, 其中

$$\mathbf{M} = \text{diag} \left\{ \frac{1}{\sum_k \psi(\Gamma(G^{1 \rightarrow k}))}, \dots, \frac{1}{\sum_k \psi(\Gamma(G^{N \rightarrow k}))} \right\}$$

\mathbf{Q} 是一个零行和矩阵, 具有负非对角项, 即 $Q_{ij} = -\psi(\Gamma(\Pi_{i \rightarrow j}))$. 容易看出, \mathbf{W} 的特征值 $\lambda_i (i = 1, \dots, N)$ 与 $\mathbf{M}^{1/2} \mathbf{Q} \mathbf{M}^{1/2}$ 的特征值相等, 即是实数且非负的, 且最小特征值 $\lambda_1 = 0$. 另一方面, Gerschgorin 循环定理 [42] 证明了 \mathbf{L}^w 的每个特征值都在一个半径为 1 的圆的内部, 因此 $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_N \leq 2$. 证明完毕.

接下来引入超中心节点的定义.

超中心节点 网络中具有最大 γ 值的节点被称为网络的超中心节点.

定理 1 令 $G(V, E)$ 为一个无向网络, 对于每一对节点 u 和 v , 选择它们之间的最短路径 SP_{uv} , 并根据通信邻域图的定义确定通信邻域图运用加权机制 ((12) 式). 如果该邻域图具有唯一的超中心节点, 那么当 $\alpha \rightarrow +\infty$ 时, 通过移除超中心节点的输入边, 剩余网络将会退化到一个以超中心节点为根

节点的有向生成树, 因此总有一条有向路径可以从超中心节点通向其他所有节点.

证明 根据 (12) 式, 当 $\alpha \rightarrow +\infty$ 时, 可以得到 $\psi(\Gamma(\Pi_{i \rightarrow j})) \approx \max\{\Gamma(\Pi_{i \rightarrow j})\}$, 从而有 $\sum_k \psi \times (\Gamma(\Pi_{i \rightarrow j})) \approx \max_k \{\max\{\Gamma(G^{i \rightarrow k})\}\}$. 现在, 假设节点 u^* 是唯一的超中心节点, 即网络中具有最大的 γ 值的节点. 当 $\alpha \rightarrow +\infty$ 时, 如果 $u^* \in \Pi_{i \rightarrow j}$ 并收敛于 0, 权重 W_{ij} 将收敛于 1. 换言之, 每个节点 $i \neq u^*$ 只保持一条边连接网络中具有最大 γ 值的那个节点. 由于 u^* 的输入边被删除了, 也就是所有到 u^* 的输入边的权重都等于 0, 仅保留 $N-1$ 条有向边. 因此, 从节点 u^* 到网络中所有其他节点都有一条有向路径, 且仅保留有 $N-1$ 条边, 由此得到的网络结构是一个有向生成树, 其中具有最大 γ 值的节点是根节点. 证明完毕.

推论 2 根据参考文献 [41], 定理 1 中得到的有向生成树的特征值比 $\lambda_N/\lambda_2 = 1$, 意味着该网络具有最大的同步能力. 简单来说, 由于生成树的根拥有最大的 γ 值, 因此在所得的生成树中, 大多数节点都在树的较低层次上, 也就是说大多数节点与根节点的平均距离较短. 值得一提的是, 当 $\alpha \rightarrow +\infty$ 时, 如果不移除根节点的输入边, 那么特征值比 $\lambda_N/\lambda_2 = 2$, 也就是说, $0 = \lambda_1 < \lambda_2 = \lambda_3, \dots, \lambda_{N-1} < \lambda_N = 2$. 如果超中心节点并不唯一, 当 $\alpha \rightarrow +\infty$ 时, 网络将退化为有向图, 其中所有超中心节点都属于边权重相等的团. 此外, 该团中的任意节点都能与图中的所有其他节点相连, 即对于网络中的每个节点来说, 都有一条有向路径通向每个超中心节点. 基于上述可以发现, 只有参与这些路径的边被留下来, 而其他的边都将会消失.

4.2 聚类结构分析

前文已经提到, 在一些应用中, 有时候还需要降低同步能力, 如辨别网络中的聚类结构 [17]. 本文提出的加权策略对聚类结构探测也有很强的影响, 并且可以进一步发现聚类结构与同步性之间的隐藏关系. 首先给出聚类结构的弱定义:

聚类结构的弱定义 Radicchi 等 [43] 分别从弱定义和强定义这两方面给出了衡量聚类结构的量化标准, 其中弱定义被认为是一个子图集合能够成为聚类结构的最弱标准, 即聚类是网络的子图集合, 子图内部的边的密度高于子图之间的边的密

度. 给定一个网络 $G = (V, E)$, 其中 V 是节点集, E 是边界集, $A = (a_{ij})$ 为其邻接矩阵, 令 $V_s \subset V$ 为一个特定子图, $\bar{V}_s = V \setminus V_s$ 为网络其余的节点集, 那么在弱灵敏度条件下 V_s 能够成为一个聚类的条件是:

$$L(V_s, V_s) > L(V_s, \bar{V}_s), \quad (15)$$

其中, $L(V_s, V_s) = \sum_{i \in V_s} \sum_{j \in V_s} a_{ij}$, $L(V_s, \bar{V}_s) = \sum_{i \in V_s} \sum_{j \notin V_s} a_{ij}$. 在弱定义下的网络聚类中, V_s 内部的边的个数至少应该超过其连接网络剩余部分的边个数的一半.

推论 3 如果 e_{ij} 是一个类间边, 那么邻域图 $\Pi_{i \rightarrow j}$ 和 $\Pi_{j \rightarrow i}$ 很可能是两个相邻的聚类. 此外, 在极端情况下, 如果网络是二划分且 e_{ij} 是唯一的类间边, 根据聚类结构的弱定义, $\Pi_{i \rightarrow j}$ 和 $\Pi_{j \rightarrow i}$ 将是严格的网络聚类.

证明: 根据弱定义, 如果一个网络拥有聚类结构, 由于存在高密度的类内边, 一个随机游走者将会在一个聚类中停留很多的时间. 这意味着一个节点到同一聚类内的节点的路径长度通常比到不同聚类的节点的路径长度短. 如图 2 中 $\alpha = 2$ 时, $\Pi_{i \rightarrow j}$ 中的节点到节点 j 的路径长度比到节点 i 的路径长度都要长. 相反, $\Pi_{j \rightarrow i}$ 中的节点到节点 i 的路径长度比到节点 j 的路径长度长. 因此可以得出结论: $\Pi_{i \rightarrow j}$ 和 $\Pi_{j \rightarrow i}$ 很可能是两个不同的聚类. 此外, 如果 e_{ij} 是唯一的类间界, 根据聚类结构的弱定义, 这个结论是严格的. 证明完毕.

为了阐明加权机制对聚类结构的影响, 逐步增大 α 的值并在示例网络中删除 $W_{ij} < 0.5$ 的边. 如图 3 所示, 当 α 值从 2 增加到 4, 聚类结构会显著增强.



图 3 应用加权机制后网络的动态演化示例图

Fig. 3. Example of dynamic evolution of network after applying weighting mechanism.

根据弱定义, 比率 $R = l^{\text{out}}/l^{\text{in}}$ 在解决模块度优化问题上发挥着十分重要的作用. 为了满足弱定义的标准, 这个比率的区间应该为 $[0, 2]$.

定理 2 令 l^{out} 为类间边的数量, l^{in} 为类内边的数量, 那么比率 $R = l^{\text{out}}/l^{\text{in}}$ 可以用来有效量化网络中的聚类结构.

证明 对于特定的网络 G , 其相对的超图 G^* 定义为一个有向加权的 C 团结构, 其中的每个节点对应于 G 的一个聚类. 在 G^* 中, 节点 r 和节点 s 之间的边用 $l_{\text{out}}^{rs}/l_{\text{in}}^r$ 来加权, 其中 l_{out}^{rs} 代表 G 中聚类 r 和聚类 s 之间的团间边数量, l_{in}^r 代表聚类 r 内部边的数量. G^* 对应的 $C \times C$ 拉普拉斯矩阵 $\mathbf{F} = \{F_{rs}\}$ 是不对称的, 但是可以变换为 $\mathbf{F} = \mathbf{\Delta}\mathbf{\Theta}$, 其中 $\mathbf{\Delta} = \{\Delta_{rs}\}$ 是一个对称的零行和矩阵, 其非对角元素为 $\Delta_{rs} = -l_{\text{out}}^{rs}$, 对角元素为 $\Delta_{rr} = \sum_{s \neq r} l_{\text{out}}^{rs}$, 且 $\mathbf{\Theta} = \text{diag}\{1/l_{\text{in}}^1, \dots, 1/l_{\text{in}}^C\}$. 那么可以将 \mathbf{F} 写成:

$$\mathbf{F} = \mathbf{\Theta}\mathbf{\Delta} = \begin{pmatrix} \frac{1}{l_{\text{in}}^1} & 0 & \cdots & 0 \\ 0 & \frac{1}{l_{\text{in}}^2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{l_{\text{in}}^C} \end{pmatrix} \times \begin{pmatrix} l_{\text{out}}^{11} & -l_{\text{out}}^{12} & \cdots & -l_{\text{out}}^{1C} \\ -l_{\text{out}}^{21} & l_{\text{out}}^{22} & \cdots & -l_{\text{out}}^{2C} \\ \vdots & \vdots & \ddots & \vdots \\ -l_{\text{out}}^{C1} & -l_{\text{out}}^{C2} & \cdots & l_{\text{out}}^{CC} \end{pmatrix}, \quad (16)$$

因为 \mathbf{F} 是零行和的, 所以 \mathbf{F} 的谱元素都是非负实数, 其中 \mathbf{F} 的最小特征值 $\lambda_1^* = 0$, 次小特征值 $\lambda_2^* > 0$.

根据谱方法和信息论 [42,44–46], λ_2^* 能够量化超图的连通性, 从而衡量不同聚类的性质和互相关联的程度. 应该注意到, λ_2^* 已经过标准化, 所以它应该是一维的. 这里, 可以得出 $\lambda_2^* \approx l_{\text{out}}/l_{\text{in}}$. 特殊情况下, 如果 C 个聚类都具有相等规模 $N_c = N/C$, 那么类内边数量 l_{in}^r 和类间边 l_{out}^r 数量对所有聚类来说都是一样的. 此时, 由 $l_{\text{in}}^r = l_{\text{in}}/C$ 和 $l_{\text{out}}^r = l_{\text{out}}/C$, 可以得到 $L = C(l_{\text{in}}^r + 0.5l_{\text{out}}^r) = l_{\text{in}} + 0.5l_{\text{out}}$: 由 $\mathbf{\Delta} = CI/l_{\text{in}}$, 可以得到 $\lambda_2^* = l_{\text{out}}/l_{\text{in}} = l_{\text{out}}^r/l_{\text{in}}^r$. 证明完毕.

接下来将展示本文的加权策略如何化解聚类探测中的分辨率限制问题, 以使优化方法更加有效. Fortunato 和 Barthélemy [32] 发现了聚类探测

的分辨率限制问题, 表明对于优化模块度 Q ((1) 式) 的算法, 如果 l_{in}^r 的取值区间不在 $[0, L]$ 内, 那么模块度优化算法将无法检测到这些聚类. 直观地看, 如果引入一种加权策略, 能够有效扩大这个限制区间, 那么许多优化算法就能够在较大的限制范围内探测到准确聚类. 为了分析简便, 假设权重之和边的个数 L 保持相等. 这里, 加权网络中所有的量都以上尖号为区分, 即 \hat{R} . 首先, 回顾一个关于聚类规模边界的定理 [32].

定理 3 聚类规模对模块度函数 Q 的影响有如下限制:

$$\frac{l_{\text{in}}^i}{L} < \frac{4}{(2 + R)^2}, \quad (17)$$

其中, $R = l^{\text{out}}/l^{\text{in}}$. 如果合并聚类 i 和聚类 j 不能提高模块度函数 Q 的值, 那么对于聚类 i 和聚类 j 的规模有如下约束条件 [31]:

对所有 j , 有

$$\frac{l_{\text{in}}^i}{L} \geq \frac{2R_{ij}}{(R_i + 2)(R_j + 2)};$$

对所有 i , 有

$$\frac{l_{\text{in}}^j}{L} \geq \frac{2R_{ji}}{(R_i + 2)(R_j + 2)}, \quad (18)$$

其中, 聚类 i 和聚类 j 之间的团间边权重之和为 $R_{ij}l_{\text{in}}^i = R_{ji}l_{\text{in}}^j = l_{\text{out}}^{ij}$.

设定网络加权后所有权重之和为 L . 如果 \hat{l}_{in}^i 和 \hat{l}_{in}^j 分别表示聚类 i 的类内边权重之和以及聚类 i 与其他聚类节点的类间边权重之和, 那么有 $\hat{R}_i = \hat{l}_{\text{in}}^{\text{out}}/\hat{l}_{\text{in}}^i$, $\hat{R}_{ij} = \hat{l}_{\text{in}}^{\text{out}}/\hat{l}_{\text{in}}^i$. 利用加权模块度函数 Q , 根据定理 3, 可以推导出聚类规模的上下界为

$$\frac{2\hat{R}_{ij}}{(2 + \hat{R}_i)(2 + \hat{R}_j)} \leq \frac{\hat{l}_{\text{in}}^i}{L} \leq \frac{4}{(2 + \hat{R}_i)^2}. \quad (19)$$

本文的加权策略将减少 \hat{R}_{ij} 的值, 因此能提高聚类探测规模的上限.

值得注意的是, 如果减少 \hat{R}_i , \hat{R}_j 和 \hat{R}_{ij} 的值, 聚类探测规模的下限不一定会降低. 然而, 如果假设对所有 i 都有 $\hat{R}_i \leq 2$, 将得到下界的区间:

$$\frac{1}{8}\hat{R}_{ij} \leq \frac{2\hat{R}_{ij}}{(2 + \hat{R}_i)(2 + \hat{R}_j)} \leq \frac{1}{2}\hat{R}_{ij}, \quad (20)$$

当 \hat{R}_{ij} 逐渐减少, 可以发现该区间的上下两端都将减少, 表明引入本文加权机制后, 聚类探测规模的

下界会减少.

总结来说, 如果将本文的加权策略应用于网络中, 那么能够通过优化算法探测出更大或更小的聚类, 从而可以有效地降低分辨率限制的问题.

5 实验

本节将上述加权机制运用在人工网络和大量现实世界网络中. 结果表明, 这种加权机制可以分别利用原模式和逆模式来提高网络同步能力以及增强聚类结构检测能力.

5.1 同步性

在本文提出的加权策略中, 唯一的调整参数是 α . 通过改变 α 参数, 可以控制权重值的异质性和特征值比 λ_N/λ_2 .

1) 人工基准网络

这里利用无标度网络^[2]和 Watts-Strogatz 小世界网络^[1]作为基准模型网络, 并将本文加权机制应用其中. 无标度网络的节点度分布服从幂律特性, 即该模型有两个参数, 其中节点初始边数 k_0 控制平均度, 即 $\langle k \rangle \approx 2k_0$. 另外, 参数 β 调整控制网络度分布的异质性, 当 β 增加时, 网络的异质性会减少到特定程度.

图 4 给出了在无标度网络上的实验结果, 其中数据点为 50 次实验结果的平均值. 图 4(a) 为特征值比 λ_N/λ_2 和参数 $(\alpha, \langle k \rangle)$ 的对应关系, 这里网络的平均度用 $\langle k \rangle$ 表示. 设 $\alpha = 1$, 对于所有的平均度 $\langle k \rangle$, 可以得到 $\lambda_N/\lambda_2 < 2$, 这比文献^[23]的最优情况还要好, 而又优于文献^[21, 22, 39]的研究结果. 对于较大的平均度 $\langle k \rangle$, 权重分布和度分布的同质性增加, 这表示 γ 值将更加均匀, 而 α 的影响也会较弱. 所以, 无论 α 的大小, 加权过程都将显著提高网络同步能力. 对于平均度 $\langle k \rangle$ 较低的网络, 由于网络中异质性较高, 因此较大的 α 将会得到更好的结果. 接下来研究网络的异质性 (非均匀性), 该性质可以用参数 β 来衡量. 已有研究表明, 网络度分布的异质性是影响同步性能最重要的因素之一. 无标度网络中特征值比 λ_N/λ_2 和参数 α, β 的关系如图 4(b) 所示. 对于不同的 β , 因为无标度特征导致节点度的非均匀性, 参数 α 的影响作用会更加显著. 此外可以看出, 随着 β 值增加, 网络非均匀性降低, 得到的加权网络的同步性会逐渐变差.

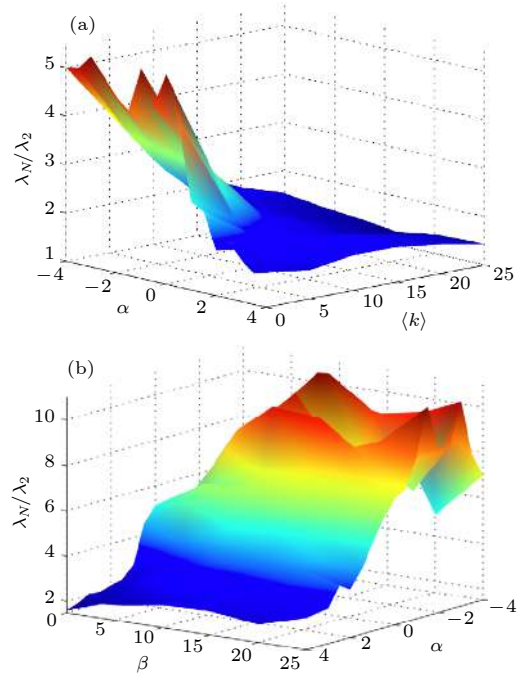


图 4 (a) 无标度网络中 λ_N/λ_2 与 $(\alpha, \langle k \rangle)$ 的对应关系 ($N = 500, \beta = 0$); (b) 无标度网络中 λ_N/λ_2 与 (α, β) 的对应关系 ($N = 500, \langle k \rangle = 4$)

Fig. 4. (a) Corresponding relationship ($N = 500, \beta = 0$) in scale-free networks between λ_N/λ_2 and $(\alpha, \langle k \rangle)$; (b) Corresponding relationship ($N = 500, \langle k \rangle = 4$) in scale-free networks between λ_N/λ_2 and (α, β) .

接下来将加权机制应用到 Watts-Strogatz 网络中, 相对于无标度网络来说, Watts-Strogatz 网络的节点度分布是均匀的. 该网络受两个参数控制, 即平均度 $\langle k \rangle = 2k_0$, 和重连概率 P . 图 5 给出了在 Watts-Strogatz 网络中应用本文加权算法的结果, 图中数据点为 50 次实验结果的平均值. 其中, 图 5(a) 为特征值比 λ_N/λ_2 随参数 α 和网络平均度 $\langle k \rangle$ 的变化情况. 和无标度网络相似, 当平均度 $\langle k \rangle$ 增加时, 网络的异构程度随之降低 α , 从而降低 α 的贡献程度. 可以看出, 当平均度 $\langle k \rangle$ 非常高时, α 对特征值比 λ_N/λ_2 几乎没有影响. 图 5(b) 为特征值比 λ_N/λ_2 随参数 α 和重连概率 P 的变化情况. 当 P 较小时, 网络度分布几乎是均匀的, 但是节点的负载分布却是极度非均匀的. 这是由于极少数的重连边可以被视作捷径, 它们承担着很高的负载, 导致负载的分布具有高度的不均匀性. 随着重连概率 P 的增加, 网络重连边的数量会逐渐增加, 负载分布会因此变得越来越均匀. 在这种情况下, 虽然度分布的方差会逐渐增加, 但仍不足以提升 γ 值的异质程度. 因此, 增加 P 会使 γ 值分布更加均匀,

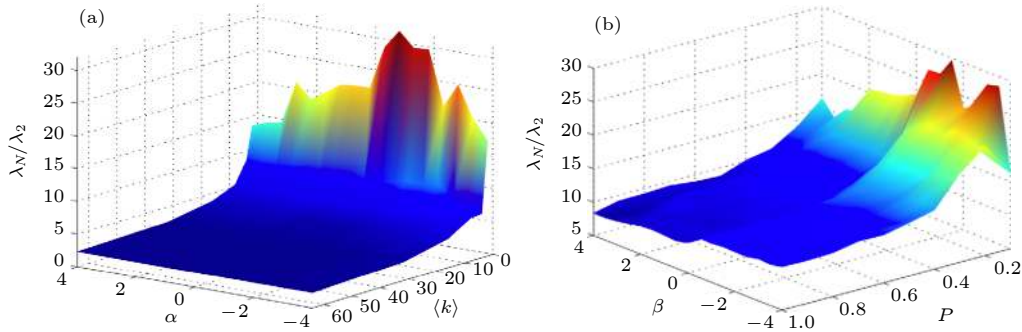


图 5 (a) Watts-Strogatz 网络中 λ_N/λ_2 与 $(\alpha, \langle k \rangle)$ 的对应关系 ($N = 500, P = 0.1$); (b) Watts-Strogatz 网络中 λ_N/λ_2 与 (α, P) 的对应关系 ($N = 500, \langle k \rangle = 4$)

Fig. 5. (a) Corresponding relationship ($N = 500, P = 0.1$) in Watts-Strogatz networks between λ_N/λ_2 and $(\alpha, \langle k \rangle)$; (b) corresponding relationship ($N = 500, \langle k \rangle = 4$) in Watts-Strogatz networks between λ_N/λ_2 and (α, P) .

表 1 在现实世界网络中使用 8 种不同加权策略的实验结果对比, 其中 RW 代表随机游走方法, SA 代表模拟退火方法, VB 代表变分贝叶斯方法, PL 代表概率推断方法

Table 1. Comparison of experimental results using eight different weighting strategies in real world networks, in which RW represents random walk method, SA stands for simulated annealing method, VB stands for variational Bayesian method, PL stands for probability inference method.

现实世界网络	N	$\langle k \rangle$	λ_N/λ_2								This work
			Chavez	Wang	Jalili	Khadiji	RW	SA	VB	PL	
蛋白质结构网络2	53	4.64	20.92	20.54	6.06	5.83	5.61	4.87	4.43	4.59	4.27
海豚网络	62	5.12	16.89	43.01	6.83	6.22	6.04	5.33	5.07	5.30	4.95
蛋白质结构网络1	95	4.48	63.1	262.2	23.5	19.82	15.71	13.65	10.59	11.88	8.45
蛋白质结构网络3	99	4.37	43.75	299.85	13.07	10.84	10.27	9.87	9.14	9.00	8.02
中国航空网络	203	18.48	13.25	5.79	3.29	2.88	2.23	2.08	1.83	1.76	1.55
电子邮件通讯	1133	9.62	8.63	5.81	5.40	4.04	3.86	3.91	3.84	3.54	3.77
酵母蛋白质交互作用	1458	2.67	52.44	269.07	25.60	17.63	15.61	13.21	10.76	12.66	9.52
蛋白质交互作用	2840	2.92	34.87	41.60	16.50	13.85	11.48	10.47	8.94	9.87	5.50
中国电力网络	865	5.20	49.77	133.8	25.43	20.90	23.44	15.19	12.04	10.56	5.04
科学家合作网络	4380	3.25	68.31	273.14	38.69	25.87	20.02	17.36	10.42	11.77	7.21
因特网AS2	7690	4.00	12.90	3.26	2.94	2.15	2.04	2.10	1.91	1.59	1.88
因特网AS5	8063	4.10	12.88	3.41	3.37	2.56	2.27	2.09	1.97	2.20	1.83

从而降低加权机制的贡献程度. 特别地, 当 $P = 1$ 时, Watts-Strogatz 网络等同于完全随机图.

2) 现实世界网络

虽然现实世界极其复杂以致无法获取全部信息, 但是为了充分地进行验证, 我们将 9 个不同的加权机制应用到大量现实世界网络中并进行对比, 包括 Chavez 方法^[21]、Wang 方法^[22]、Jalili 方法^[23]、Khadiji 方法^[31]、模拟退火 (SA) 算法^[25]、随机游走 (RW) 算法^[27]、变分贝叶斯 (VB) 算法^[29]、概率推断 (PL) 方法^[30] 和本文方法, 结果列在表 1 中, 其中设定 $\alpha = 4$. 可以看出, 本文使用的加权机制的性能超过了其他所有的加权方法. 值得一提的是, 对于所有的现实世界网络, 随着 α 的增加, 特征

值 λ_N/λ_2 比都将收敛于 2.

5.2 聚类探测

1) 人工基准网络

本文采用两个著名的人工基准网络, 即 Girvan-Newman (GN) 基准网络和 Lancichinetti-Fortunato-Radicchi (LFR) 基准网络, 用以评估加权机制对社区探测效率的影响. 其中, GN 基准测试^[9,11] 已经广泛应用于对不同聚类算法的效率评估. GN 网络共包括 128 个节点, 这些节点被分配到 4 个聚类中, 每个聚类包含 32 个节点, 每个聚类内部点之间连边和与类外节点连边的概率分别为 p_{in} 和 p_{out} . 因此, 对于每个节点, 它们在聚类内部

的边的个数和与其他类节点关联的边的期望分别为 $z_{in} = 31p_{in}$ 和 $z_{out} = 96p_{out}$, 可以看出 $z_{in} + z_{out} = 16$.

更进一步地, Lancichinetti 等^[47]提出了 LFR 基准网络, 它可以通过调整相关参数来符合现实网络中的无标度特征. 在该基准网络中, 节点的度分布和聚类规模遵循幂律分布. 此外, 还需要设定一些其他参数, 包括节点最大度、节点平均度、节点总数、聚类最大规模和最小规模以及最重要的混合参数 μ . 混合参数 μ 的取值区间为 $[0, 1]$, 决定了 LFR 基准图的聚类模糊性程度: μ 越大, 聚类就越难被正确划分. 可以看出, GN 基准网络是 LFR 基准网络的一个特例. 由于 GN 基准网络和 LFR 基准网络的社团都是预先给定的, 而不同划分算法得到的结果是不同的, 因此这里利用正确率 R (正确划分的节点和全部节点的比值) 衡量划分的正确性.

首先, 将加权策略将应用于 GN 基准网络. 在图 6(a) 的原加权网络中, α 取 2 和 4, 在逆加权网络中, α 取 -2 和 -4, 数据点为 50 次实验结果的平均值. 图 6(a) 给出了在 GN 基准网络上应用加权策略前后的平均正确率 R 的对比. 可以看出, 当 $\alpha = 4$ 和 $\alpha = 2$ 时, 加权策略的原模式都提高了平均正确率 R , 而相应的两个逆模式 ($\alpha = -4$ 和 $\alpha = -2$) 都会降低平均正确率 R . 与 $\alpha = 4$ 的情况相比, 在 $\alpha = 2$ 的情况下, 平均正确率 R 会下降约 2—3 倍. 而且在逆模式中, 当 $z_{out} \leq 6.5$ 时, $\alpha = -4$ 的情况与 $\alpha = -2$ 相比, 平均正确率 R 会降低 2 倍以上.

其次, 在 LFR 基准网络上进行分析, 其中参数设置如下: 节点数量为 1000、平均度为 20、最大节点度为 50、幂律分布指数为 2、混合参数 μ 在区间 $[0, 0.5]$ 内变化. 从图 6(b) 可以看出, 加权策略的原模式会显著提高平均正确率 R , 而相应的逆模式会降低平均正确率 R . 与 $\alpha = 4$ 的情况相比, 在 $\alpha = 2$ 的情况下, 平均正确率 R 会下降约 4—5 倍. 另外, 在逆加权模式也会得到相似的结果. 比较两种基准网络, 对于提高平均正确率 R , 加权策略在 LFR 基准网络中会取得更好的效果.

再次, 为了展示加权机制对聚类探测的影响, 首先利用具有不同 α 值的加权机制对网络赋权, 然后利用一些著名的聚类算法运行并测试其准确性. 这里使用在信息论中被广泛研究的标准化互信息 (normalized mutual information, NMI) 来衡量聚

类结果的准确性^[40]. NMI 位于区间 $[0, 1]$ 内, 其值的大小表明真实聚类结果和模型结果之间重叠的比例. 特别地, 当 NMI 的值为 1 时, 算法结果与真实聚类完全相同, 具有最高的划分效率. 真实划分 X 和模型结果 Y 之间的标准化互信息 NMI 的数学定义为

$$I(x, y) = \frac{-2 \sum_{i=1}^{c_X} \sum_{j=1}^{c_Y} N_{ij} \ln(n N_{ij} / N_i N_j)}{\sum_{i=1}^{c_X} N_i \ln\left(\frac{N_i}{n}\right) + \sum_{j=1}^{c_Y} N_j \ln\left(\frac{N_j}{n}\right)}, \quad (21)$$

其中, c_X 和 c_Y 分别表示真实划分 X 和模型结果 Y 中的聚类数量. 在 (21) 式, N_{ij} 代表真实划分中聚类 i 和实验结果中聚类 j 之间重叠的节点数量, N_i 和 N_j 分别表示矩阵 N 中第 i 行和第 j 列值的总和.

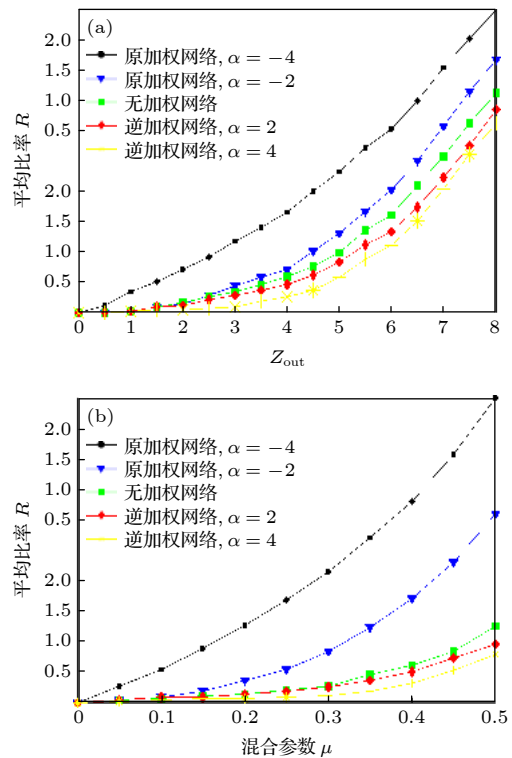


图 6 (a) GN 基准网络中使用加权机制 (逆加权机制) 前后平均比率 R 的对比; (b) LFR 基准网络中使用加权机制 (逆加权机制) 前后平均比率 R 的对比

Fig. 6. (a) Comparison of average ratio R before and after using weighting mechanism (inverse weighting mechanism) in GN benchmark network; (b) comparison of average ratio R before and after using weighting mechanism (inverse weighting mechanism) in LFR benchmark network.

图 7(a) 和图 7(b) 给出了利用两种著名算法, 即模拟退火法 (SA)^[48] 和 Duch-Arenas 法 (DA)^[49], 在运用加权机制的网络上的实验结果, 图中数据点

为 50 次实验结果的平均值. 其中, SA 算法被证明是效果最好的聚类算法之一, 然而其计算复杂性也较高. 与 SA 相比, DA 不仅在模块化优化方面具备一定可靠性, 而且计算过程也相对简单. 从图 7(a) 和图 7(b) 可以看到, 当 α 值从 -2 降低到 -4, 逆加权模式对应的 NMI 值越来越高, 要优于无权网络. 相反, 在原加权模式中, 当 α 值从 2 增加到 4, 加权网络的 NMI 相对于无权网络的 NMI 却越来越少. 这些结果均验证了本文的加权策略的有效性. 此外, 通过比较两种方法发现, 图 7(a) 中 $\alpha = 4$ ($\alpha = -4$) 和 $\alpha = 2$ ($\alpha = -2$) 两种情况的 NMI 差异区间小于图 7(b). 这可能是因为, 即使是在无权网络中, SA 都拥有更高的精确度, 难以通过加权策略提高它的表现. 但是对于 DA 来说, 通过提高 α , 加权机制能够大幅提高它的计算精度.

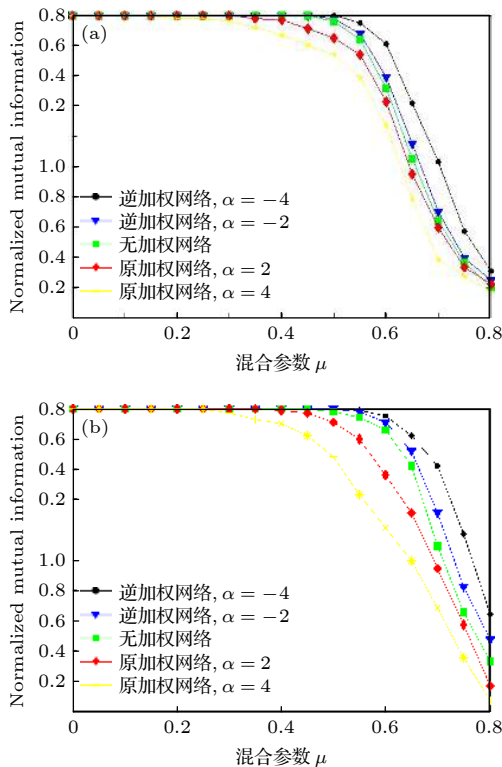


图 7 在 LFR 网络中运用加权机制 (逆加权机制), 当取不同的 α 值时, 利用 (a) SA 算法和 (b) DA 算法后 NMI 的计算值

Fig. 7. Weighted mechanism (inverse weighted mechanism) is used in LFR network. When the α value is different, the NMI value is calculated by using (a) SA algorithm and (b) DA algorithm.

此外, 本文还考虑了聚类规模的影响, 验证结果如图 8 所示, 图中数据点为 50 次实验结果的平均值. 考虑两种不同聚类规模条件下的 LFR 基准

网络, 其中, 每个聚类的节点数量在 10—50 之间为小聚类网络, 在 20—100 之间为大聚类网络. 为了增强这两种 LFR 基准网络的可比性, 规定它们的区别仅限于聚类的规模大小, 其他方面的性质则完全相同. 图 8 验证了在逆加权网络中, SA 和 DA 算法的准确性都得到提升. 此外, 通过比较不同规模 LFR 网络的聚类结果可以看出, 大聚类网络中两种算法的精确度更低, 这可能是因为小聚类网络中具有更多的类间边, 而类间边更可能是被加权的. 并且通过比较两种方法, 图 8(a) 显示的小聚类和大聚类 LFR 网络 (无论是逆加权网络还是加权网络) 之间的 NMI 区间差距小于图 8(b), 原因与图 7 相同, 完美证明了本文加权机制的有效性.

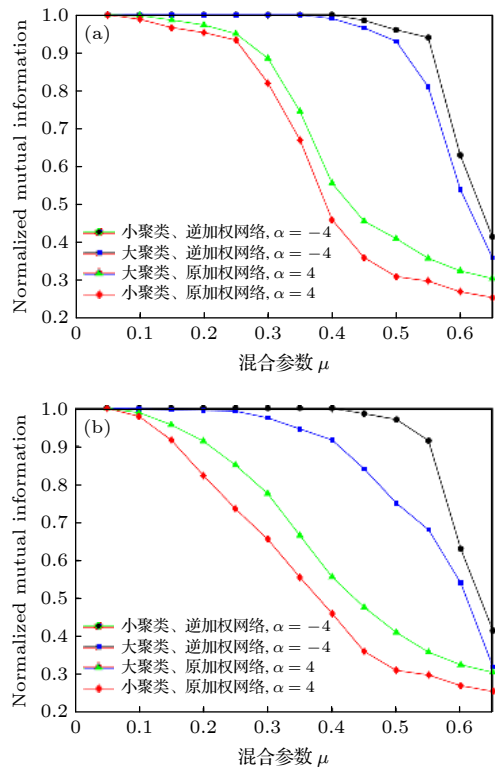


图 8 在 LFR 网络中运用加权机制 (逆加权机制), 当 $\alpha = 4$ ($\alpha = -4$) 并考虑聚类规模时, 利用 (a) SA 算法和 (b) DA 算法后 NMI 的计算值

Fig. 8. The weighted mechanism (inverse weighted mechanism) is used in LFR network. When $\alpha = 4$ ($\alpha = -4$) and considering the cluster size, the NMI calculated by (a) SA algorithm and (b) DA algorithm.

2) 现实世界网络

进一步将加权策略应用到现实世界网络中, 结果在表 2 中列出. 这里使用了 7 种广泛使用的网络, 并标识了相应的文献出处. 为了便于比较, 提供了关于 7 种网络已发表方法中的最优模块度

表 2 在不同现实网络使用加权策略得到的实验结果, 其中“/”左右表示加权后和加权前的模块度 Q 值
 Table 2. Experimental results are obtained by using weighting strategy in different real networks, where /'s left or right represents the modularity Q value after or before weighting.

网络	文献	最优 Q	SA ^[48]	DA ^[49]	CNM ^[10]
中国航空网络	[11]	—	0.644/0.525	0.589/0.428	0.577/0.483
空手道俱乐部	[50]	0.420	0.416/0.342	0.411/0.351	0.413/0.376
《悲惨世界》	[51]	0.561	0.554/0.389	0.539/0.406	0.531/0.395
海豚社会网络	[52]	0.531	0.527/0.375	0.521/0.362	0.517/0.356
电子邮件	[53]	0.579	0.568/0.462	0.543/0.436	0.538/0.444
爵士乐	[54]	0.446	0.439/0.333	0.437/0.341	0.431/0.328
PGP密钥签名	[55]	0.878	0.883/0.674	0.843/0.705	0.872/0.754

表 3 在现实世界网络中使用不同加权策略的实验结果对比, 这里网络聚类算法利用 CNM 算法, 其中 RW 代表随机游走方法, SA 代表模拟退火方法, VB 代表变分贝叶斯方法, PL 代表概率推断方法
 Table 3. Comparison of experimental results using different weighting strategies in the real world network. CNM algorithm is used as the network clustering algorithm, in which RW represents the random walk method, SA represents the simulated annealing method, VB represents the variable dB method, and PL represents the probability inference method.

网络	Chavez	Wang	Jalili	Khadivi	RW	SA	VB	PL	This work
空手道俱乐部	0.316	0.322	0.351	0.362	0.374	0.381	0.390	0.386	0.413
中国航空网络	0.449	0.478	0.423	0.432	0.506	0.543	0.564	0.578	0.603
《悲惨世界》	0.357	0.369	0.399	0.411	0.439	0.457	0.488	0.433	0.531
爵士乐	0.338	0.347	0.353	0.361	0.383	0.399	0.42	0.387	0.431
PGP密钥签名	0.583	0.678	0.676	0.715	0.744	0.786	0.839	0.820	0.872
海豚社会网络	0.357	0.381	0.371	0.374	0.406	0.444	0.483	0.500	0.517
电子邮件	0.368	0.409	0.431	0.443	0.471	0.499	0.503	0.495	0.538

Q 值^[56]. 本文计算了当 $\alpha = 4$ 时三种代表性算法, 即 SA^[48], DA^[49] 和 CNM^[10] 的加权前后模块度 Q 值, 其中“/”左右表示加权后和加权前的模块度 Q 值, 结果如表 2 所列. 这三种算法的精确度虽然并不处于最顶级的行列, 然而正如表 2 所列, 在应用加权策略后, 结果非常接近已发表的最优值, 并且计算过程更加简便.

进一步将 9 种加权策略应用到现实世界网络中, 包括 Chavez 方法^[21]、Wang 方法^[22]、Jalili 方法^[23]、Khadivi 方法^[31]、模拟退火 (SA) 算法^[25]、随机游走 (RW) 算法^[27]、变分贝叶斯 (VB) 算法^[29]、概率推断方法^[30] 和本文方法, 用以对比不同加权方法的聚类效果. 聚类算法统一使用 CNM 方法^[10], 并计算当 $\alpha = 4$ 时的模块度 Q 值. 结果列在表 3 中, 可以看出, 本文所使用的加权机制的性能超过了其他所有的加权方法, 从而验证了本文加权算法的高效性.

6 结 论

随着“大数据”技术的不断涌现和发展, 如何处理大规模网络已经成为了巨大的现实挑战^[57,58]. 同

时, 作为一个开放性问题, 提高网络同步性能或聚类探测的算法效率往往非常困难. 为此, 本文提出一个新型双模式加权机制, 通过加权改变网络拓扑结构来提高同步性能或聚类探测的算法效率. 该加权机制包括两种模式: 一是通过提高桥点的权重以增强同步能力的原始模式; 二是通过减少桥点的权重以提高聚类探测效率的逆模式. 这种加权机制仅受一个参数 α 的影响, 因此非常便于实施. 通过在人工基准网络和现实世界网络上的实验分析, 检验了模型的有效性和高效性. 特别地, 通过比较发现, 本文提出的加权策略在效率上优于以往发表的提升网络计算效率的加权方法.

在今后的研究工作中, 仍有一些问题值得深入探讨. 比如工程、信息领域应用中的群体演化博弈和平均一致性问题, 可能在无向网络中更容易分析和理解, 加权有向网络并不一定是最佳选择. 另外, 如何在拥有数百万甚至更多节点的超大规模网络中进行研究分析, 也需要进一步深入探索.

感谢吉林大学刘大有教授、中国科学院数学与系统科学研究院章祥荪研究员、中央财经大学刘志东教授对本文提出的建设性建议.

参考文献

- [1] Watts D J, Strogatz S H 1998 *Nature* **393** 440
- [2] Barabasi A L, Albert R 1999 *Science* **286** 509
- [3] Han Y, Zhu L, Cheng Z, Li J, Liu X 2020 *IEEE Trans. Cybern.* **50** 1697
- [4] Yang B, Liu D Y, Liu J M, Jin D, Ma H B 2009 *J. Software* **20** 54 (in Chinese) [杨博, 刘大有, 刘继明, 金弟, 马海宾 2009 *软件学报* **20** 54]
- [5] Ding S, Yue Z, Yang S, Niu F, Zhang Y 2020 *IEEE Trans. Ind. Inf.* **32** 2101
- [6] Liang W, Li K, Long J, Kui X, Zomaya A Y 2020 *IEEE Trans. Ind. Inf.* **16** 2063
- [7] Lu M, Zhang Z, Qu Z, Kang Y 2019 *IEEE Trans. Knowl. Data Eng.* **31** 1736
- [8] Ma X, Dong D, Wang Q 2019 *IEEE Trans. Knowl. Data Eng.* **31** 273
- [9] Newman M E J, Girvan M 2004 *Phys. Rev. E* **69** 026113
- [10] Clauset A, Newman M E J 2004 *Phys. Rev. E* **70** 066111
- [11] Du W B, Zhou X L, Lordan O, Wang Z, Zhao C, Zhu Y B 2016 *Transp. Res. Pt. E-Logist. Transp. Rev.* **89** 108
- [12] Zeng X, Wang W, Chen C, Yen G G 2020 *IEEE Trans. Cybern.* **50** 2502
- [13] Palla G, Derenyi I, Farkas I, Vicsek T 2005 *Nature* **435** 814
- [14] Li J, Wang X, Cui Y 2014 *Physica A* **415** 398
- [15] Li H J, Li H Y, Li A H 2015 *Chin. J. Comput.* **38** 301 (in Chinese) [李慧嘉, 李慧颖, 李爱华 2015 *计算机学报* **38** 301]
- [16] Hofman J M, Wiggins C H 2008 *Phys. Rev. Lett.* **100** 258701
- [17] Boccaletti S, Ivanchenko M, Latora V, Pluchino A 2007 *Phys. Rev. E* **75** 045102
- [18] Xu Y, Wu X, Li N, Liu L, Xie C, Li C 2019 *IEEE Trans. Circuits Syst. Express Brief* **67** 700
- [19] Han M, Zhang M, Qiu T, Xu M 2019 *IEEE Trans. Neural Networks Learn. Syst.* **30** 255
- [20] Hong H, Kim B J, Choi M Y, Park H 2004 *Phys. Rev. E* **69** 067105
- [21] Chavez M, Hwang D U, Amann A, Hentschel H E, Boccaletti S 2005 *Phys. Rev. Lett.* **94** 218701
- [22] Wang X, Lai Y C, Lai C H 2007 *Phys. Rev. E* **75** 056205
- [23] Jalili M, Rad A A, Hasler M 2008 *Phys. Rev. E* **78** 016105
- [24] Rad A A, Jalili M, Hasler M 2008 *Chaos* **18** 037104
- [25] Lu X, Kuzmin K, Chen M, Szymanski B K 2018 *Inf. Sci.* **424** 55
- [26] Zhang Y, Wang M, Gottwalt F, Saberi M, Chang E 2019 *J. Informetr.* **13** 616
- [27] De Meo P, Ferrara E, Fiumara G, Provetti A 2013 *J. Informetr.* **222** 648
- [28] Yang R, Wang W X, Lai Y C, Chen G 2009 *Phys. Rev. E* **79** 026112
- [29] Li H J, Daniels J J 2015 *Phys. Rev. E* **91** 012801
- [30] Meyniel F, Dehaene S 2017 *PNAS* **114** 3859
- [31] Khadivi A, Ajdari R A, Hasler M 2011 *Phys. Rev. E* **83** 046104
- [32] Fortunato S, Barthélemy M 2007 *PNAS* **104** 36
- [33] Good B H, de Montjoye Y A, Clauset A 2010 *Phys. Rev. E* **81** 046106
- [34] Newman M E J 2002 *Comput. Phys. Commun.* **147** 40
- [35] Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang D U 2006 *Phys. Rep.* **424** 175
- [36] Lu L Y, Chen D B, Ren X L, Zhang Q M, Zhang Y C, Zhou T 2016 *Phys. Rep.* **650** 1
- [37] Carmi S, Havlin S, Kirkpatrick S, Shavitt Y, Shir E 2007 *PNAS* **104** 11150
- [38] Kitsak M, Gallos L K, Havlin S, Liljeros F, Muchnik L, Stanley H E, Makse H A 2010 *Nat. Phys.* **6** 888
- [39] Motter A E, Zhou C, Kurths J 2005 *Phys. Rev. E* **71** 016116
- [40] Motter A E, Zhou C, Kurths J 2005 *EPL* **69** 334
- [41] Nishikawa T, Motter A E 2006 *Phys. Rev. E* **73** 065106
- [42] Gerschgorn S 1931 *Izv. Akad. Nauk USSR Otd. Fiz.-Mat. Nauk* **7** 749
- [43] Radicchi F, Castellano C, Cecconi F, Loreto V, Parisi D 2004 *PNAS* **101** 2658
- [44] Li H J, Yan G, Liu Z D, Li G J, Zhang X S 2017 *Sci. Sin. Math* **4** 7241 (in Chinese) [李慧嘉, 严冠, 刘志东, 李桂君, 章祥荪 2017 *中国科学: 数学* **4** 7241]
- [45] Li H J, Wang Y, Wu L Y, Zhang J, Zhang X S 2012 *Phys. Rev. E* **86** 016109
- [46] Li H J, Zhang X S 2013 *EPL* **103** 58002
- [47] Lancichinetti A, Fortunato S, Radicchi F 2008 *Phys. Rev. E* **78** 046110
- [48] Guimera R, Nunes Amaral L A 2005 *Nature* **433** 895
- [49] Duch J, Arenas A 2005 *Phys. Rev. E* **72** 027104
- [50] Zachary W W 1977 *J. Anthropol. Res.* **33** 452
- [51] Knuth D E 1994 *The Stanford Graph Base: A Platform for Combinatorial Computing* (New York: ACM Press) p592
- [52] Lusseau D, Schneider K, Boisseau O J, Haase P, Slooten E, Dawson S M 2003 *Behav. Ecol. Sociobiol.* **54** 396
- [53] Guimera R, Danon L, Diaz-Guilera A, Giralt F, Arenas A 2003 *Phys. Rev. E* **68** 065103
- [54] Gleiser P, Danon L 2003 *Adv. Complex Syst.* **6** 565
- [55] Boguna M, Pastor-Satorras R, Diaz-Guilera A, Arenas A 2004 *Phys. Rev. E* **70** 056122
- [56] Agarwal G, Kempe D 2008 *Eur. Phys. J. B* **66** 409
- [57] Xing N, Zong Q, Dou L, Tian B, Wang Q 2019 *IEEE Trans. Veh. Technol.* **68** 9963
- [58] Yang H, Yao Q, Yu A, Lee Y, Zhang J 2019 *IEEE Trans. Commun.* **67** 3457

A new tunable weighting strategy for enhancing performance of network computation^{*}

Li Hui-Jia^{1)†} Huang Zhao-Ci¹⁾ Wang Wen-Xuan¹⁾ Xia Cheng-Yi²⁾

1) (*School of Science, Beijing University of Posts and Telecommunications, Beijing 100876, China*)

2) (*School of Computer and Communication Engineering, Tianjin University of Technology, Tianjin 300384, China*)

(Received 9 March 2021; revised manuscript received 2 April 2021)

Abstract

For many real world systems ranging from biology to engineering, efficient network computation methods have attracted much attention in many applications. Generally, the performance of a network computation can be improved in two ways, i.e., rewiring and weighting. As a matter of fact, many real-world networks where an interpretation of efficient computation is relevant are weighted and directed. Thus, one can argue that nature might have assigned the optimal structure and weights to adjust the level of functionality. Indeed, in many neural and biochemical networks there is evidence that the synchronized and coordinated behavior may play important roles in the system's functionality. The importance of the network weighting is not limited to the nature. In computer networks, for example, designing appropriate weights and directions for the connection links may enhance the ability of the network to synchronize the processes, thus leading the performance of computation to improve. In this paper, we propose a new two-mode weighting strategy by employing the network topological centralities including the degree, betweenness, closeness and communication neighbor graph. The weighting strategy consists of two modes, i.e., the original mode, in which the synchronizability is enhanced by increasing the weight of bridge edges, and the inverse version, in which the performance of community detection is improved by reducing the weight of bridge edges. We control the weight strategy by simply tuning a single parameter, which can be easily performed in the real world systems. We test the effectiveness of our model in a number of artificial benchmark networks as well as real-world networks. To the best of our knowledge, the proposed weighting strategy outperforms previously published weighting methods of improving the performance of network computation.

Keywords: complex networks, weighting strategy, communication neighbor graph, community structure

PACS: 89.75.Hc, 89.75.Fb

DOI: 10.7498/aps.70.20210453

^{*} Project supported by Fundamental Research Funds for the Central Universities of China (Grant No. 2020XD-A01-2) and the National Natural Science Foundation of China (Grant No. 71871233).

[†] Corresponding author. E-mail: hjli@bupt.edu.cn