基于子图交互关系的网络结构增强算法

胡　雯[1]，马　闯[2]，张海峰[1*]

(1. 安徽大学数学科学学院　合肥　230601；2. 安徽大学互联网学院　合肥　230601)

【摘要】已有研究基于子图交互关系构造子图网络来实现网络结构增强，然而其算法复杂度高。鉴于此，基于不同阶子图网络的拓扑属性分别对原始网络进行赋权，得到一阶和二阶加权网络，以权重的形式直观体现子图交互关系。同时，这两种加权网络的权重可以直接通过原始网络的拓扑结构计算得出，从而避免了子图网络的构造过程，大大降低了算法复杂度。最后，以关键点识别任务作为研究对象说明这两种加权网络在结构挖掘应用中的性能。基于加权网络定义了两种新的中心性指标，在 8 个真实网络中与 7 种经典的中心性指标进行对比，实验结果表明基于加权网络的中心性指标具有更好的性能。

关　键　词　关键点识别；　子图；　子图网络；　网络赋权

中图分类号　TP391；N94　　**文献标志码**　A　　doi:10.12178/1001-0548.2021196

Network Structure Enhancement Algorithm Based on Subgraph Interaction

HU Wen[1], MA Chuang[2], and ZHANG Haifeng[1*]

(1. School of Mathematical Science, Anhui University　Hefei　230601; 2. School of Internet, Anhui University　Hefei　230601)

Abstract　The existing studies show that the network structure can be enhanced by constructing subgraph network based on subgraph interaction relationship, but the complexity of such algorithms is high. In view of this, this paper weights the original network based on the topological attributes of different order subgraph networks, obtains the first-order and second-order weighted networks, and intuitively reflects the interaction relationships of subgraphs in the form of weights. At the same time, the weights of the two weighted networks can be calculated directly through the topology of the original network, which avoids the construction process of subgraph network and greatly reduces the complexity of the algorithm. Finally, the key nodes identification task is taken as the research object to illustrate the performance of the two weighted networks in the application of structure mining. In this paper, two new centrality indices are defined based on weighted networks, which are compared with seven classical centrality indices in eight real networks. The experimental results show that the centrality indices based on weighted network has better performance.

Key words　identification of key nodes;　subgraph;　subgraph network;　weighting network

近年来，网络科学作为一门新兴交叉学科得到越来越多的关注。现实世界中许多真实系统都可以基于网络框架进行理解，如生物网络系统[1-2]、社会网络系统[3-4]、交通网络系统[5-6] 等。挖掘网络的结构信息可以帮助我们更深入地了解这些系统并加以应用，因此如何发展有效的方法挖掘网络的结构信息有着重要的科学价值和广泛的应用前景。如识别网络上的关键点用于疾病控制和舆情扩散[7-9]、相对关键节点的识别用于罪犯挖掘和致病基因查找[10]、链路预测用于好友推荐和应对网络攻击[11]、社团划分用于潜在客户挖掘和社交网络角色检测[12] 以及动力学分析用于通信安全和疫情分析[13-14] 等。

现有的网络结构分析方法大多是基于网络的浅层结构开展的，仅考虑节点对之间是否有联系，使用的网络结构信息都是有限的，这些基于浅层结构的分析方法会造成很多深层的、隐藏的信息丢失，而结构信息的缺失会导致真实网络的内在联系无法被完整地反映。真实网络具有很多不同种类的子

收稿日期：2021－07－25；修回日期：2021－10－19
基金项目：国家自然科学基金 (61973001)；安徽省自然科学基金 (2008085QF299)
作者简介：胡雯 (1994－)，女，主要从事复杂网络方面的研究.
*通信作者：张海峰，E-mail：haifengzhang1978@gmail.com

图，子图捕捉了网络中特定的局部连接模式，这些子图的大小、类型、属性都会影响网络的结构与功能，因此很多学者开展了关于网络子图的研究。如文献 [15] 研究了大型稀疏网络中子图频率与网络结构的联系，发现利用不同子图出现的频率可挖掘网络的局部密集结构。文献 [16] 在大肠杆菌的转录调控网络中发现不同的子图结构在整个网络中对应的功能是特定的，进而研究网络不同结构的子图捕获网络局部聚集信息的差异性。文献 [17] 将三角形结构的频繁子图应用于图聚类，发现利用子图结构能更有效地实现网络的社团检测。文献 [18] 研究不确定图的数据挖掘，提出一种基于期望支持阈值来寻找不确定图的频繁子图的方法。文献 [19] 提出一种基于子图的近似图匹配技术，在生物网络中高效地搜索具有相似功能的细胞实体。

以上基于子图的数据挖掘研究仅仅考虑了子图自身的一些属性，没有考虑子图之间的交互关系。实际上，如何有效地构建子图间的交互关系来实现网络的结构增强，并用于后续的结构挖掘任务具有重要的意义。最近文献 [20] 选择网络中不同的子图结构作为节点，分别构造了不同阶数的子图网络 (subgraph network, SGN)，实现了子图之间的交互，并发现子图网络的集成可以增强后续一系列图分类算法。基于 SGN 提取的特征补充了原始网络的结构特征，可以更深入地挖掘网络结构信息，然而构建 SGN 的过程是复杂的，并且构造更高阶的 SGN 需要知道低一阶 SGN 的结构。基于此，本文提出了两种不需要构造 SGN 的赋权方法，直接从原始网络中挖掘出子图之间的交互信息，在实现结构增强的同时也大大降低了算法复杂度。主要思想是基于不同阶数 SGN 的构建过程，提出两种加权方法将 SGN 的结构信息在原始网络中以边权值的形式表现出来，一种是将一阶子图网络 (first-order subgraph network, $\mathrm{SGN}^{(1)}$) 节点的属性作为原始网络对应边的权重，构造一阶加权网络；另一种是将二阶子图网络 (second-order subgraph network, $\mathrm{SGN}^{(2)}$) 节点集合中包含原始网络同一条边的所有节点属性综合起来作为这条边的权重，构造二阶加权网络。$\mathrm{SGN}^{(1)}$ 和 $\mathrm{SGN}^{(2)}$ 考虑了两种最基本的子图结构：边和开三角结构，因此本文提出的两种赋权方法以权重的形式分别体现的是边和开三角结构的交互关系，既保留了原始网络的结构信息，又补充了特定子图的交互信息，实现了网络结构增强。为了验证本文提出的两种加权方法在后续结构挖掘任务中的有效性，本文定义了两个基于加权网络的中心性指标识别真实网络中的关键节点。实验结果表明，在关键点识别问题研究中，基于加权网络的中心性指标比经典的度中心性、接近性中心性、介数中心性、特征向量中心性、K-Shell 中心性、LeaderRank 以及显著性中心性都具有更好的性能。

1　子图网络

本文在给网络赋权的过程中涉及不同阶数 SGN 的概念及对应的构造方法，因此先介绍这些概念和构造过程。

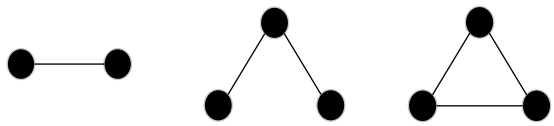
1.1　SGN 的定义

给定一个无向无权网络 $G(V,E)$，其中 $V=\{v_i|i=1,2,\cdots,N\}$表示节点集合，节点数为N，$E\subseteq(V\times V)$表示边的集合，E中元素(v_i,v_j)满足$(v_i,v_j)=(v_j,v_i)$，$i,j=1,2,\cdots,N$。定义G的子图 $g_i=(V_i,E_i)$，其中$g_i\subseteq G$，当且仅当$V_i\subseteq V$且$E_i\subseteq E$。

SGN 表示网络G到网络$G^*(V^*,E^*)$的映射 $G^*=L(G)$，其中$V^*=\left\{g_j|j=1,2,\cdots,n\right\}$，$n\leqslant N$，$E^*\subseteq(V^*\times V^*)$。$V^*$中元素即为网络$G$的子图，如果$V^*$中的两个子图$g_i$和$g_j$在原始网络$G$中包含共同的节点或包含共同的边，即$V_i\cap V_j\neq\varnothing$，那么称$G^*$中的这两个节点$g_i$和$g_j$是有连边的。同样的，$G^*$中的元素$(g_i,g_j)$满足$(g_i,g_j)=(g_j,g_i)$，$i,j=1,2,\cdots,n$，其中$n\leqslant N$。

1.2　SGN 的构造过程

网络中最基本的子图是边和三角结构，它们在大多数网络中出现的频率较高，不会造成子图网络过于稀疏的情况，并且相比于边结构，三角结构也可以补充更多网络局部结构的信息，利于更高阶子图网络的构造，因此本文选择这两种子图分别构造 $\mathrm{SGN}^{(1)}$ 和 $\mathrm{SGN}^{(2)}$。本文研究都是基于无向网络，边的情况只有一种 (如图 1a)，三角结构的情况有两种，分别为开三角结构 (如图 1b) 和闭三角结构 (如图 1c)[21]，在构造 $\mathrm{SGN}^{(2)}$ 的过程中只考虑更为简单的开三角结构，并将开三角结构中度为 2 的节点记为顶点。接下来分别介绍 $\mathrm{SGN}^{(1)}$ 和 $\mathrm{SGN}^{(2)}$ 的构造过程。



a. 连边　　b. 开三角结构　　c. 闭三角结构

图 1　子图示意图

1.2.1 $SGN^{(1)}$ 的构造过程

给定原始网络$G(V,E)$，基于边构造相应的$SGN^{(1)}$，记为$G'(V',E')$。$SGN^{(1)}$ 的节点集合为原始网络的边，如果两条边在原始网络有共同端点则相应的节点在 $SGN^{(1)}$ 中有连边。这里以图 2a 的原始网络举例说明 $SGN^{(1)}$ 的构造过程，依次提取原始网络的边作为 $SGN^{(1)}$ 的节点，如 $SGN^{(1)}$ 中标签为 (3,4) 和 (2,4) 的节点分别表示原始网络的边(v_3,v_4)和(v_2,v_4)，这两条边包含相同的端点v_4，那么 $SGN^{(1)}$ 中这两个节点有连边。按照这样的构造方法，最终得到相应的 $SGN^{(1)}$，如图 2b 所示。

1.2.2 $SGN^{(2)}$ 的构造过程

给定原始网络，基于 $SGN^{(1)}$ 构造相应的 $SGN^{(2)}$，用$G''(V'',E'')$表示。$SGN^{(2)}$ 考虑更高一阶的子图，即开三角结构作为节点集合，构造出 $SGN^{(1)}$ 之后进一步提取 $SGN^{(1)}$ 的边以获得开三角结构，如果两个开三角包含相同的边则相应的节点在 $SGN^{(2)}$ 中有连边 (需要注意的是，如果以两个开三角是否包含共同节点作为连边条件会导致子图网络过于稠密，反而不利于挖掘网络的结构信息)。图 2d 就是基于图 2b 构造的 $SGN^{(2)}$，$SGN^{(2)}$ 中标签为 (1,2,4) 的节点表示原始网络节点v_1、v_2、v_4组成的开三角，标签为 (1,2,4) 和 (1,2,3) 的两个节点包含共同的边(v_1,v_2)，那么这两个节点之间有连边。通过这样的构造方法，最终得到节点数为 6，边数为 9 的 $SGN^{(2)}$，如图 2d 所示。

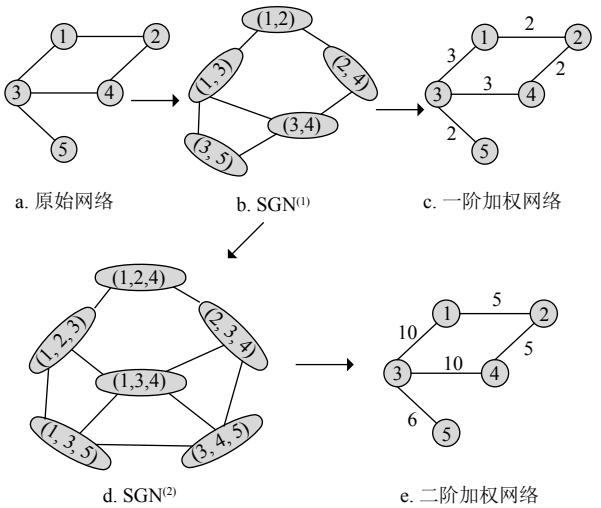


图 2 $SGN^{(1)}$ 和 $SGN^{(2)}$ 的构造过程以及对原始网络的赋权过程

2 基于 SGN 的加权网络

文献 [20] 定义了 SGN 的构造方法之后通过子图网络的结构信息以及原始网络的结构信息进行图分类，这种方法需要对新构造的 SGN 进行结构分析，有较高的复杂度，尤其在原始网络边数很多的情况下，$SGN^{(1)}$的规模会很大，并且$SGN^{(2)}$是基于$SGN^{(1)}$构造的，算法复杂度会更高。那么是否有方法既能反映不同阶子图信息又不需要添加太多额外的复杂度？基于此，本文根据$SGN^{(1)}$和$SGN^{(2)}$的拓扑结构对原始网络进行赋权，以权值的形式表现子图之间的交互关系，得到的加权网络分别记为一阶加权网络和二阶加权网络。

2.1 一阶加权网络的构造

首先介绍基于 $SGN^{(1)}$ 的赋权方法，$SGN^{(1)}$ 的每一个节点分别代表G的每一条边，那么可以选择 $SGN^{(1)}$ 节点的度作为G中相应边的权重，这个权重表示与这条边有相同端点的其他边的数目，即与这条边有交互的边数，基于此方法得到的加权网络即为一阶加权网络。

以图 2b 的 $SGN^{(1)}$ 为例介绍一阶加权网络的赋权过程，即计算出 $SGN^{(1)}$ 中每个节点的度作为对应边的权重。如标签为 (1,2) 的节点度为 2，那么给对应边(v_1,v_2)赋权重为 2，表示G中有两条边(v_1,v_3)、(v_2,v_4)与这条边包含共同的端点，给G中每条边都赋权之后，得到对应的一阶加权网络，如图 2c 所示。

根据一阶加权网络的定义规则可以直接由原始网络计算得到权重，不需要构造 $SGN^{(1)}$。记G中节点 $v_i\ (i=1,2,\cdots,N)$的度为 k_i，记 G'中节点 (v_i,v_j) $(i,j=1,2,\cdots,N)$的度为k_{ij}。将G中的边(v_i,v_j)看作一个整体，这个整体的邻居就是在节点v_i和节点v_j的所有邻居中删去这两个节点，得到k_{ij}与k_i以及k_j之间的关系式：

$$k_{ij}=k_i+k_j-2 \tag{1}$$

记一阶加权网络中边(v_i,v_j) $(i,j=1,2,\cdots,N)$的权重为$w_{ij}^{(1)}$，基于式 (1)，$w_{ij}^{(1)}$可以直接由原始网络G中节点v_i和v_j的度来表示：

$$w_{ij}^{(1)}=k_i+k_j-2 \tag{2}$$

2.2 二阶加权网络的构造

二阶加权网络基于 $SGN^{(2)}$ 对G进行赋权，$SGN^{(2)}$ 中的节点集合由G中的开三角结构构成。要给G中的某一条边赋权，先在 $SGN^{(2)}$ 的节点集合中找出包含这条边的所有开三角结构，选择这些节点的度的总和作为权重，权重也代表与这条边相关的开三

角结构有交互的开三角的数目，基于此得到的加权网络即为二阶加权网络。

以图 2d 的 $SGN^{(2)}$ 为例介绍二阶加权网络的赋权过程，给 G 的边 (v_1,v_3) 赋权，在 $SGN^{(2)}$ 中找出包含边 (v_1,v_3) 的 3 个节点，标签分别为 (1,2,3)、(1,3,4)、(1,3,5)，这 3 个节点的度分别为 3、4、3，计算它们的总和 10 即为边 (v_1,v_3) 的权重。按同样的方法给所有边赋权，最终得到对应的二阶加权网络，如图 2e 所示。

根据二阶加权网络的定义规则也可以直接由原始网络计算得到权重，省略了 $SGN^{(2)}$ 的构造过程。V'' 中由 v_i、v_j、v_l 构成的以 v_j 为顶点的开三角结构表示为 (v_i,v_j,v_l)，$i,j,l=1,2,\cdots,N$，记度为 k_{ijl}。$SGN^{(2)}$ 的节点集合是由 $SGN^{(1)}$ 的边构成的，在 $SGN^{(1)}$ 中应用式 (1)，可以得到 $SGN^{(2)}$ 中节点的度与 $SGN^{(1)}$ 中节点的度的关系式：

$$k_{ijl}=k_{ij}+k_{jl}-2 \tag{3}$$

记二阶加权网络中边 (v_i,v_j) $(i,j=1,2,\cdots,N)$ 的权重为 $w_{ij}^{(2)}$，根据二阶加权网络的赋权方法得到式 (4)：

$$w_{ij}^{(2)}=\sum_{l_1\in\Gamma(i),l_1\neq j}k_{ijl_1}+\sum_{l_2\in\Gamma(j),l_2\neq i}k_{ijl_2} \tag{4}$$

式中，$\Gamma(i)$ 表示节点 i 的邻居。

将式 (3) 代入式 (4) 并化简，$w_{ij}^{(2)}$ 可以直接由原始网络 G 中节点 v_i 和 v_j 及其邻居的度来表示：

$$w_{ij}^{(2)}=2(k_i^2+k_j^2+k_ik_j)-10(k_i+k_j)+12+\sum_{l_1\in\Gamma(i)}k_{l_1}+\sum_{l_2\in\Gamma(j)}k_{l_2} \tag{5}$$

从式 (2) 和式 (5) 可以发现，一旦知道子图网络 SGN 的定义规则以及赋权方式，那么并不需要把子图网络构造出来并加以分析，可以直接利用原始网络节点的度得到两种加权网络，显然运算复杂度会大大降低。

3　实验与结果

为了表明本文定义的赋权方法可以包含更深层次的网络结构信息，本文以网络的关键点识别作为研究对象。为此定义加权网络节点的强度作为新的中心性指标，然后与原始网络上的一些中心性指标进行比较，判断该中心性指标能否更好地刻画节点的重要性。

3.1　数据集

本文在 8 个真实网络中进行了关键点识别问题的研究，这 8 个真实网络分别是 Email[22]、TAP[23]、Yeast[24]、CA-GrQc1[25]、Rt-alwefaq[26]、Rt-obama[26]、Power[27]、Y2H[28]。表 1 为网络的基本信息，其中 N 为节点数，M 为边数，$<k>$ 为平均度，$\beta_c=<k>/<k^2>$ 为传播阈值。

表 1　网络的基本信息

Network	N	M	$<k>$	β_c
Email	1133	5451	9.622	0.054
TAP	1373	6833	9.953	0.061
Yeast	2375	11693	9.847	0.029
CA-GrQc1	4158	13422	6.456	0.056
Rt-alwefaq	4171	7059	3.385	0.008
Rt-obama	3212	3422	2.131	0.025
Power	4941	6594	2.669	0.258
Y2H	1458	1948	2.672	0.140

3.2　中心性指标

本文考虑一阶加权网络和二阶加权网络中节点的强度分别定义了两个新的中心性指标，一阶子图网络中心性 (first-order subgraph network centrality, SGN1) 和二阶子图网络中心性 (second-order subgraph network centrality, SGN2)：

$$\mathrm{SGN1}(i)=\sum_{j\in\Gamma(i)}w_{ij}^{(1)} \tag{6}$$

$$\mathrm{SGN2}(i)=\sum_{j\in\Gamma(i)}w_{ij}^{(2)} \tag{7}$$

本文选择如下几种经典的中心性指标进行比较。

度中心性 (degree centrality, DC)[29] 以节点的一阶邻居数来衡量节点的重要性，节点的邻居数越多表示节点越重要，节点 i 的度中心性定义为：

$$\mathrm{DC}(i)=\frac{k_i}{N-1} \tag{8}$$

接近性中心性 (closeness centrality, CC)[30] 体现节点与网络中其他节点的近邻程度：

$$\mathrm{CC}(i)=\frac{N-1}{\sum_{j\neq i}d_{ij}} \tag{9}$$

式中，d_{ij} 表示节点 i 与节点 j 的最短距离。

介数中心性 (betweenness centrality, BC)[31] 定义为通过该节点的最短路径在所有最短路径中的占比：

$$\mathrm{BC}(i)=\sum_{i\neq s,i\neq t,s\neq t}\frac{g_{st}^i}{g_{st}} \tag{10}$$

式中，g_{st} 表示节点 s 到节点 t 的最短路径数；g_{st}^i 表

示节点s到节点t的最短路径中包含节点i的最短路径数。

特征向量中心性 (eigenvector centrality, EC)[32] 考虑节点的邻居数以及节点邻居的重要性，节点 i 的特征向量中心性定义为：

$$\mathrm{EC}(i) = x_i = c\sum_{j=1}^{N} a_{ij}x_j \tag{11}$$

式中，x_j表示节点j的重要性；$A_{ij} = (a_{ij})_{N\times N}$是网络的邻接矩阵；$c$是比例常数。

K-Shell 中心性 (KS)[33] 是基于节点度的一种粗粒度划分，节点的核数代表了其在网络中的深度，越深层的节点重要性越高。其步骤为：1) 删去网络中度为 1 的节点，残差图中出现度为 1 的节点继续删除，直至剩余网络中没有度为 1 的节点，此时，所有删去节点记为 1-shell；2) 使用递归的方法删去网络中度为 2 的节点，记为 2-shell，依次下去直至网络中所有节点被删除。

LeaderRank(LR)[34] 是基于游走的中心性指标，LR 主要用于有向网络，对于无向网络，首先将无向网络中的无向边理解为有向网络中的双向连接，然后添加一个背景节点 g 与网络中的所有节点进行双向连接，考虑了节点邻居的重要性，节点i在t时刻的得分为：

$$\mathrm{LR}_i(t) = \sum_{j=1}^{n+1} \frac{a_{ji}}{k_j^{\mathrm{out}}}\mathrm{LR}_j(t-1) \tag{12}$$

式中，k_j^{out}表示节点j的出度。

最终，节点i的 LR 值为：

$$\mathrm{LR}(i) = \mathrm{LR}_i(t_c) + \frac{\mathrm{LR}_g(t_c)}{n} \tag{13}$$

式中，t_c表示收敛时间；$\mathrm{LR}_g(t_c)$表示在稳态下背景节点的 LR 得分。

显著性中心性 (distinctiveness centrality, DIC)[35] 考虑在某些情况下，与外围节点的连接应该更为重要：

$$\mathrm{DIC}(i) = \sum_{j=1, j\neq i}^{n} w_{ij}\lg\frac{n-1}{k_j^{\alpha}} \tag{14}$$

式中，w_{ij}表示边(i, j)的权重；惩罚因子$\alpha(\alpha \geqslant 1)$表示对大度节点进行惩罚。

3.3 评价指标

本文以 SIR(susceptible-infected-recovered) 传播模型[36] 评估网络中每个节点的重要性，得到每个节点作为源头时所感染的范围，定义感染范围的比例为节点的重要性。为了评估某一个节点的传播能力，将这个节点预先设为感染态，其他节点均为易感态进行传播，设 SIR 传播模型的感染概率为β，恢复概率为 1，直到网络中不存在感染态节点就终止 SIR 传播过程。在此过程中，节点的感染范围反映了节点的重要性，感染范围越大就代表该节点重要性越大。

用 Kendall Rank 相关系数[37]τ来评价基于中心性指标的节点重要性排名 (记为X) 与节点在 SIR 传播模型的真实传播能力排名 (记为Y) 的相关性，记X, Y中第$i(1 \leqslant \mathrm{i} \leqslant N)$个值分别为$X_i, Y_i$。如果$X, Y$中的元素满足$X_i > X_j$且$Y_i > Y_j$，或者满足$X_i < X_j$且$Y_i < Y_j$，则表明$X, Y$中这两对元素一致；如果$X_i < X_j$且$Y_i > Y_j$，或者$X_i > X_j$且$Y_i < Y_j$，则表明这两对元素不一致；如果$X_i = X_j$或$Y_i = Y_j$则表明这两对元素既不是一致的也不是不一致的，定义 Kendall Rank 相关系数τ为：

$$\tau = \frac{C-D}{N(N-1)/2} \tag{15}$$

式中，C是X, Y中拥有一致性的元素对数；D是X, Y中拥有不一致性的元素对数。

定义 M 值[38] 来量化节点重要性排名X的分辨率：

$$M(X) = \left[1 - \frac{\sum_{c\in X} N_c(N_c-1)}{N(N-1)}\right]^2 \tag{16}$$

式中，N_c表示在排名中处于同一等级c的节点数；$M(X)$取值在 0～1 之间，$M(X)$越大表示排名X的分辨率越高，当$M(X) = 1$时表示X中所有排名都处于不同等级，$M(X) = 0$表示X中所有排名都处于同一等级。

另外定义$\varepsilon(p)$[33] 量化中心性指标在识别网络中有影响力的传播者方面的性能：

$$\varepsilon(p) = 1 - \frac{L(p)}{L_{\mathrm{eff}}(p)} \tag{17}$$

式中，p是网络规模N的比例 ($p \in [0,1]$)，定义每个节点作为源头所感染范围的比例为节点的扩散效率；$L(p)$是中心性最高的pN个节点的平均扩散效率；$L_{\mathrm{eff}}(p)$是扩散效率最高的pN个节点的平均扩散效率；$\varepsilon(p)$量化了具有最高中心性的pN个节点的平均感染范围与 SIR 传播过程中最优的pN个节点的

平均感染范围的接近程度，$\varepsilon(p)$越小表示中心性指标越能准确识别网络中有影响力的节点。

3.4　结果与分析

本文在真实网络中，将新定义的两个中心性指标 SGN1、SGN2，与已有的 DC、CC、BC、EC、KS、LR、DIC 这 7 个中心性指标进行了对比实验。图 3 比较了基于中心性指标的节点排名与不同传播率下的 SIR 传播模型的真实排名的 Kendall Rank 相关系数τ，实验结果取 500 次平均。结果表明，除了在 Power 网络中，SGN1 和 SGN2 指标仅次于 LR 指标，在其他 7 个网络中，本文基于子图定义的两种新的中心性指标优于其他 7 种中心性指标，能更好地识别网络中有影响力的节点。

此外还比较了 SGN1、SGN2 和 DC、CC、BC、EC、KS、LR、DIC 的 M 值，表 2 表明 M(SGN1)、M(SGN2)、M(CC)、M(EC)、M(DIC) 的数值非常接近 1，说明这 5 种指标有很好的分辨率，并且优于 M(DC)、M(BC)、M(KS)、M(LR)，但由图 3 可知本文定义的中心性指标 SGN1、SGN2 在衡量网络的节点重要性方面效果要优于 CC、EC、DIC。

对于每一个真实网络，设置 SIR 模型的感染概率$\beta > \langle k\rangle/(\langle k^2\rangle - \langle k\rangle)$，恢复概率为 1，SIR 传播模型的传播范围取 500 次平均。定义网络规模比例 p 在 0.01～0.20 之间等距取 20 个值，在真实网络中比较了 SGN1、SGN2 和 DC、CC、BC、EC、KS、LR、DIC 这 7 种中心性指标对应的$\varepsilon(p)$。图 4 的结果表明本文的中心性指标 SGN1 和 SGN2 对应的$\varepsilon(p)$在大部分的网络中最优，且中心性指标 SGN1 和 SGN2 对应的$\varepsilon(p)$都接近 0，所以 SGN1 和 SGN2 在识别网络中有影响力的节点方面总体优于 DC、CC、BC、EC、KS、LR、DIC 指标。

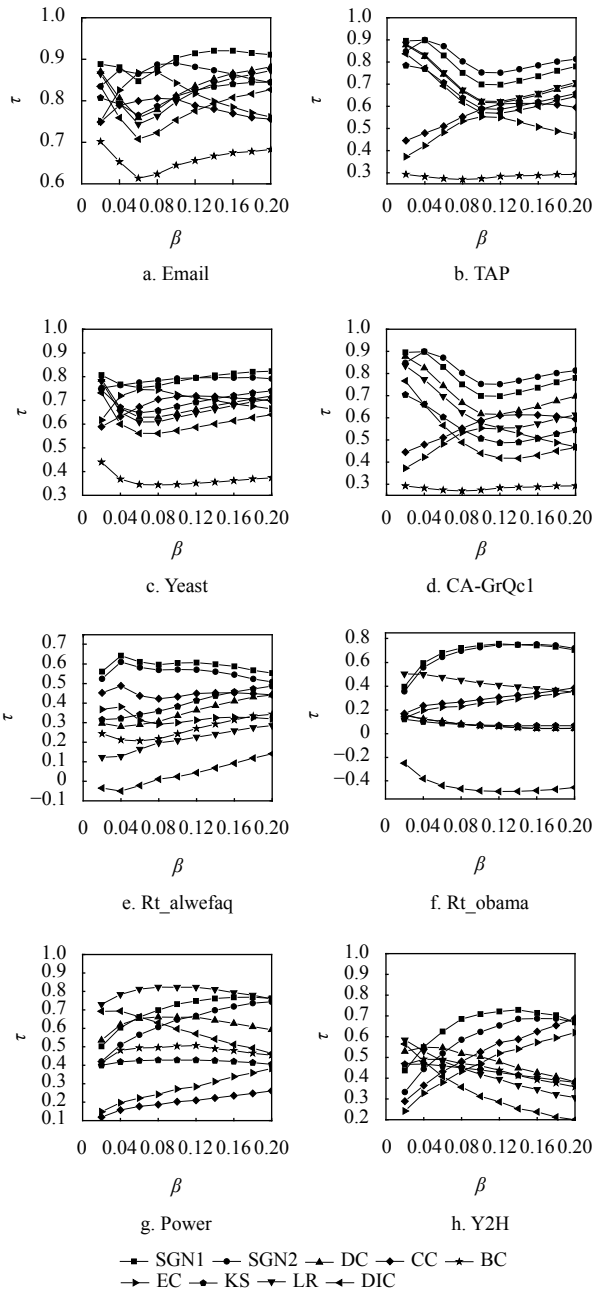


图 3　真实网络中不同中心性指标的τ值比较

表 2　不同中心性指标的 M 值比较

Network	M(SGN1)	M(SGN2)	M(DC)	M(CC)	M(BC)	M(EC)	M(KS)	M(LR)	M(DIC)
Email	0.9955	0.9998	0.8874	0.9988	0.9400	0.9999	0.8088	0.8894	0.9989
TAP	0.9941	0.9992	0.8991	0.9988	0.9238	0.9994	0.8380	0.9074	0.9985
Yeast	0.9886	0.9987	0.8314	0.9988	0.8292	0.9991	0.7737	0.8339	0.9948
CA-GrQc1	0.9829	0.9987	0.7916	0.9990	0.4849	0.9994	0.6925	0.7928	0.9952
Rt-alwefaq	0.9553	0.9610	0.3540	0.9609	0.3285	0.9611	0.3084	0.3540	0.9572
Rt-obama	0.9441	0.9613	0.1211	0.9622	0.1248	0.9624	0.0410	0.1211	0.9487
Power	0.9177	0.9896	0.5927	0.9998	0.8313	0.9999	0.2460	0.5954	0.9596
Y2H	0.9437	0.9923	0.4884	0.9957	0.5063	0.9960	0.2972	0.4884	0.9666

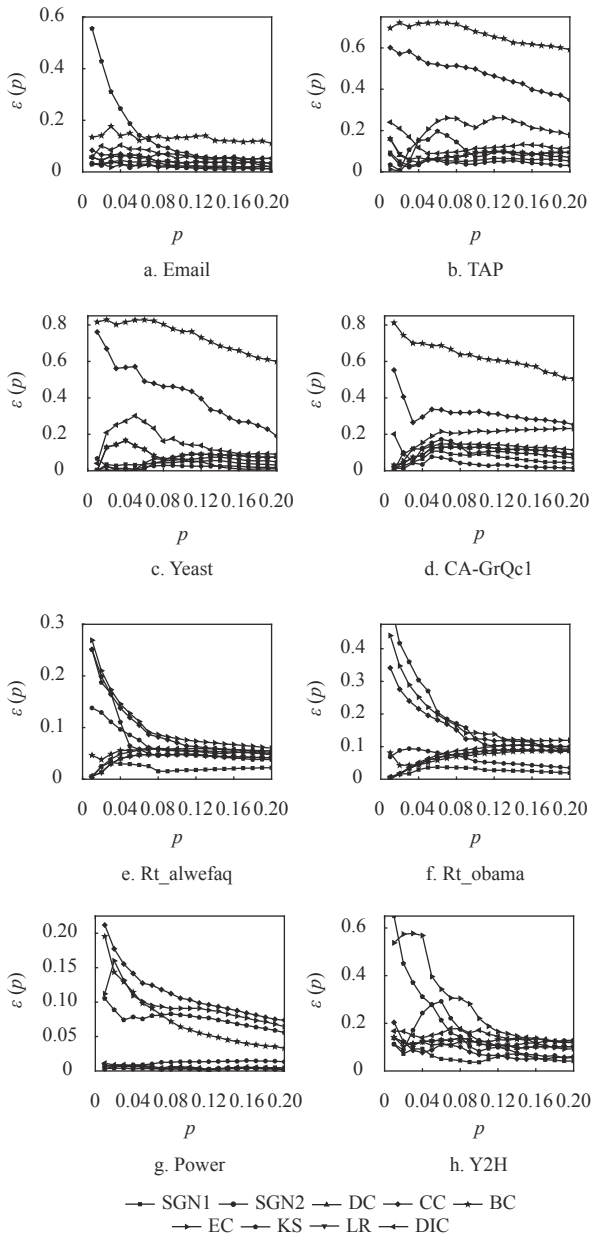


图 4 真实网络中不同中心性指标的$\varepsilon(p)$比较

4 结 束 语

综上所述，本文考虑了如何用特定子图间的交互关系来实现网络结构增强，进而可以更有效地执行结构挖掘方面的任务。基于 SGN 的构造过程将子图的结构信息以权重的形式在原始网络中表现出来，直接对原始网络进行赋权得到一阶加权网络和二阶加权网络。然后在加权网络中定义新的中心性指标 SGN1 和 SGN2，并与原始网络的 7 个中心性指标 DC、CC、BC、EC、KS、LR、DIC 进行比较。通过研究发现，这两种赋权方式可以更准确地识别网络中的关键节点。因此利用子图交互关系的赋权方法既能挖掘网络的深层次结构，又能大大降低运算复杂度。

参 考 文 献

[1] ASSENOV Y, RAMÍREZ F, SCHELHORN S E, et al. Computing topological parameters of biological networks[J]. Bioinformatics, 2008, 24(2): 282-284.

[2] CHUNG F, LU L Y, DEWEY T G, et al. Duplication models for biological networks[J]. Journal of Computational Biology, 2003, 10(5): 677-687.

[3] WELLMAN B. Computer networks as social networks[J]. Science, 2001, 293(5537): 2031-2034.

[4] GARTON L, HAYTHORNTHWAITE C, WELLMAN B. Journal of computer-mediated communication[J]. Studying Online Social Networks, 1997, 3(1): JCMC313.

[5] YANG H, HUANG H J. The multi-class, multi-criteria traffic network equilibrium and systems optimum problem[J]. Transportation Research Part B: Methodological, 2004, 38(1): 1-15.

[6] CUI Z Y, HENRICKSON K, KE R, et al. Traffic graph convolutional recurrent neural network: A deep learning framework for network-scale traffic learning and forecasting[J]. IEEE Transactions on Intelligent Transportation Systems, 2020, 21(11): 4883-4894.

[7] 张海峰, 王阳阳, 汪秉宏. 行为反应对复杂网络上传染病动力学的影响[J]. 复杂系统与复杂性科学, 2012, 9(3): 13-21.
ZHANG H F, WANG Y Y, WANG B H. The impacts of behavioral responses on the spread of infectious diseases on complex networks[J]. Complex Systems and Complexity Science, 2012, 9(3): 13-21.

[8] 赵娜, 李杰, 王剑, 等. 基于邻层传播的相对重要节点挖掘方法[J]. 电子科技大学学报, 2021, 50(1): 121-126.
ZHAO N, LI J, WANG J, et al. The impacts of behavioral responses on the spread of infectious diseases on complex networks[J]. Journal of University of Electronic Science and Technology of China, 2021, 50(1): 121-126.

[9] 曹开臣, 陈明仁, 张千明, 等. 基于网络节点中心性的新闻重要性评价研究[J]. 电子科技大学学报, 2021, 50(2): 285-293.
CAO K C, CHEN M R, ZHANG Q M, et al. Research on importance evaluation of news based on nodal centralities of complex network[J]. Journal of University of Electronic Science and Technology of China, 2021, 50(2): 285-293.

[10] 朱军芳, 陈端兵, 周涛, 等. 网络科学中相对重要节点挖掘方法综述[J]. 电子科技大学学报, 2019, 48(4): 595-603.
ZHU J F, CHEN D B, ZHOU T, et al. A survey on mining relatively important nodes in network science[J]. Journal of University of Electronic Science and Technology of China, 2019, 48(4): 595-603.

[11] WANG P, XU B W, WU Y R, et al. Link prediction in social networks: The state-of-the-art[J]. Science China Information Sciences, 2015, 58(1): 1-38.

[12] 汪小帆, 刘亚冰. 复杂网络中的社团结构算法综述[J]. 电子科技大学学报, 2009, 38(5): 537-543.
WANG X F, LIU Y B. Overview of algorithms for detecting community structure in complex networks[J]. Journal of University of Electronic Science and Technology of China, 2009, 38(5): 537-543.

[13] 张海峰, 王文旭. 复杂系统重构[J]. 物理学报, 2020, 69(8): 088906.
ZHANG H F, WANG W X. Complex system reconstruction[J]. Acta Physica Sinica, 2020, 69(8): 088906.
[14] 楼凤丹, 周银座, 庄晓丹, 等. 时效网络结构及动力学研究进展综述[J]. 电子科技大学学报, 2017, 46(1): 109-125.
LOU F D, ZHOU Y Z, ZHUANG X D, et al. Review on the research progress of the structure and dynamics of temporal networks[J]. Journal of University of Electronic Science and Technology of China, 2017, 46(1): 109-125.
[15] UGANDER J, BACKSTROM L, KLEINBERG L. Subgraph frequencies: Mapping the empirical and extremal geography of large graph collections[C]//Proceedings of the 22nd International Conference on World Wide Web. New York: Association for Computing Machinery, 2013: 1307-1318.
[16] BALAZSI G, BARABási A L, OLTVAI Z. Topological units of environmental signal processing in the transcriptional regulatory network of Escherichia coli[J]. Proceedings of the National Academy of Sciences, 2005, 102(22): 7841-7846.
[17] TSOURAKAKIS C E, PACHOCKI J, MITZENMACHER M. Scalable motif-aware graph clustering[C]//Proceedings of the 26th International Conference on World Wide Web. Republic and Canton of Geneva: International World Wide Web Conferences Steering Committee, 2017: 1451-1460.
[18] ZOU Z N, LI J Z, GAO H, et al. Mining frequent subgraph patterns from uncertain graph data[J]. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(9): 1203-1218.
[19] TIAN Y Y, MCEACHIN R C, SANTOS C, et al. SAGA: A subgraph matching tool for biological graphs[J]. Bioinformatics, 2007, 23(2): 232-239.
[20] XUAN Q, WANG J H, ZHAO M H, et al. Subgraph networks with application to structural feature space expansion[EB/OL]. (2019-03-21). https://arxiv.org/abs/1903.09022.
[21] AGARWAL S, BRANSON K, BELONGIE S. Higher order learning with graphs[C]//Proceedings of the 23rd International Conference on Machine Learning. New York: Association for Computing Machinery, 2006: 17-24.
[22] GUIMERA R, DANON L, DIAZ-GUILERA A, et al. Self-similar community structure in a network of human interactions[J]. Physical Review E, 2003, 68(6): 065103.
[23] GAVIN A C, BÖSCHE M, KRAUSE R, et al. Functional organization of the yeast proteome by systematic analysis of protein complexes[J]. Nature, 2002, 415(6868): 141-147.
[24] MERING C V, KRAUSE R, SNEL B, et al. Comparative assessment of large-scale data sets of protein-protein interactions[J]. Nature, 2002, 417(6887): 399-403.
[25] LESKOVEC J, KLEINBERG J, FALOUTSOS C. Graph evolution: Densification and shrinking diameters[EB/OL]. (2007-01-28). https://arxiv.org/abs/physics/0603229v3.
[26] RYAN A R, NESREEN K A. The network data repository with interactive graph analytics and visualization[DB/OL]. [2020-12-24]. http://networkrepository.com.
[27] WATTS D J, STROGATZ S H. Collective dynamics of ‘small-world’networks[J]. Nature, 1998, 393(6684): 440-442.
[28] BU D B, ZHAO Y, CAI L, et al. Topological structure analysis of the protein-protein interaction network in budding yeast[J]. Nucleic Acids Research, 2003, 31(9): 2443-2450.
[29] BONACICH P. Factoring and weighting approaches to status scores and clique identification[J]. Journal of Mathematical Sociology, 1972, 2(1): 113-120.
[30] SABIDUSSI G. The centrality index of a graph[J]. Psychometrika, 1966, 31(4): 581-603.
[31] FREEMAN, LINTON C. A set of measures of centrality based on betweenness[J]. Sociometry, 1977, 40: 35-41.
[32] BONACICH P. Power and centrality: A family of measures[J]. American Journal of Sociology, 1987, 92(5): 1170-1182.
[33] KITSAK M, GALLOS L K, HAVLIN S, et al. Identification of influential spreaders in complex networks[J]. Nature Physics, 2010, 6(11): 888-893.
[34] LYU L Y, ZHANG Y C, YEUNG C H, et al. Leaders in social networks, the delicious case[J]. PLoS One, 2011, 6(6): e21202.
[35] FRONZETTI C A, NALDI M. Distinctiveness centrality in social networks[J]. PLoS One, 2020, 15(5): e0233276.
[36] MORENO Y, PASTOR-SATORRAS R, VESPIGNANI A. Epidemic outbreaks in complex heterogeneous networks[J]. The European Physical Journal B-Condensed Matter and Complex Systems, 2002, 26(4): 521-529.
[37] KENDALL M G. A new measure of rank correlation[J]. Biometrika, 1938, 30(1-2): 81-93.
[38] BAE J, KIM S. Identifying and ranking influential spreaders in complex networks by neighborhood coreness[J]. Physica A: Statistical Mechanics and its Applications, 2014, 395: 549-559.

编　辑　蒋　晓