

加权复杂网络社团结构的组合模型

马领娟, 尹小彩, 苏日娜, 丁爱婷, 高彩霞*

内蒙古大学数学科学学院, 内蒙古 呼和浩特

Email: 1486201568@qq.com, xyin188@126.com, surina0920@qq.com, 1842181650@qq.com,
*gaocx0471@163.com

收稿日期: 2021年4月21日; 录用日期: 2021年5月7日; 发布日期: 2021年5月25日

摘要

社团结构是复杂网络的一个重要特性, 通过对社团结构的检测可以发现网络中藏有的规律。本文利用弱社团定义去识别加权网络的社团结构, 提出了优化模型和均值切割算法。数值实验表明对网络进行加权处理更具有现实意义。

关键词

加权复杂网络, 弱社团, 均值切割算法

Combination Model of Weighted Complex Network Community Structure

Lingjuan Ma, Xiaocai Yin, Rina Su, Aiting Ding, Caixia Gao*

School of Mathematical Sciences, Inner Mongolia University, Hohhot Inner Mongolia

Email: 1486201568@qq.com, xyin188@126.com, surina0920@qq.com, 1842181650@qq.com,
*gaocx0471@163.com

Received: Apr. 21st, 2021; accepted: May 7th, 2021; published: May 25th, 2021

Abstract

Community structure is an important characteristic of complex networks, and the hidden rules in networks can be found by detecting the community structure. In this paper, the community structure of weighted networks is identified by using the weak community definition, and an optimization model and a mean cut algorithm are proposed. Numerical experiments show that it is more

*通讯作者。

practical to weight networks.

Keywords

Weighted Complex Network, Weak Community, Mean Cutting Algorithm

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

复杂网络具有诸多拓扑特性, 如小世界网络, 无尺度网络等[1], 其中社团结构是其尤为重要的一个特性[2]。近年来, 有关社团结构的检测已有很多较为完善的模型。通过对社团结构检测的研究发现已有的模型和算法都存在着不同的局限性。

模块化函数 Q 是由 Newman 和 Girvan (2004)定义的, 用于检测社团结构的度量函数。Li 等人(2008)提出了另一种度量函数 D , 用来优化社团结构的检测。然而, 用模块函数 Q 和 D 检测社团结构可能产生分辨率极限, 甚至出现错误识别现象[3] [4] [5]。为了更加量化地解决社团结构的检测问题, Radicchi 等人(2004)提出了弱社团意义下的社团定义[4]。

由于在现实背景下, 复杂网络中节点间的关系是错综复杂的, 普通网络已无法展现节点间联系的强弱程度。通过对网络进行加权处理, 以便更客观的刻画节点间的重要性, 本文将两个节点间的权重 n 模拟成节点之间有 n 条边相连, 提出了优化的弱社团定义, 即候选子网络中的权重之和应不小于该子网络其余部分的权重之和。通过对网络进行加权处理, 建立了一个满足弱社团定义的优化模型, 并设计了相应的均值切割算法求解该模型。进一步模拟日本空手道俱乐部的数值算例, 验证了所提出的模型和算法具有实用性及有效性。

2. 加权网络社团识别的模型

给定一个加权的无向图 $G=(V, E)$ 及其邻接矩阵 $A=[a_{ij}]$, G 的顶点集 $V=\{v_1, v_2, \dots, v_n\}$, a_{ij} 表示节点 v_i 和 v_j 之间的联系程度, 也就是权重。

本文将 V_t 表示为 V 的子集, $\bar{V}_t=V/V_t$ 为 V_t 的补集, V_t 是定义在弱社团下的, 即

$$M(V_t, V_t) \geq M(V_t, \bar{V}_t) \quad (1)$$

其中

$$M(V_t, V_t) = \sum_{i \in V_t} \sum_{j \in V_t} a_{ij}, \quad M(V_t, \bar{V}_t) = \sum_{i \in V_t} \sum_{j \in \bar{V}_t} a_{ij}$$

在网络 G 中, 给定一个划分 $P_K=(G_1, G_2, \dots, G_K)=((V_1, E_1), \dots, (V_K, E_K))$, K 是满足弱社团定义的社团数量。在不考虑权重下提出的模块化函数 Q [5] [6]

$$Q = \sum_{S=1}^K \left[\frac{L(V_S, V_S)}{2L} - \left(\frac{L(V_S, V_S) + L(V_S, \bar{V}_S)}{2L} \right)^2 \right] \equiv \sum_{S=1}^K Q_S$$

和 Li 等人[7]提出的另一种模块化函数 D :

$$D = \sum_{s=1}^k \frac{L(V_s, V_s) - L(V_s, \bar{V}_s)}{|V_s|} \equiv \sum_{S=1}^K D_S$$

其中 $L = L(V, V)/2$ 是网络中的连接总数。这里， Q 和 D 可能会产生错误识别现象及分辨率极限问题。Zhang 等人[8] (2010)提出的组合优化模型

$$\begin{aligned} & \max \sum_{k=1}^n y_k \\ & \sum_{k=1}^n x_{ik} = 1 \\ & z_{lk} \leq x_{ik} \\ & z_{lk} \leq x_{jk} \\ & x_{ik} + x_{jk} - 1 \leq z_{lk} \\ s.t. & \sum_{i=1}^n x_{ik} \geq y_k \\ & \sum_{i=1}^n x_{ik} \leq n y_k \\ & 2 \sum_{l=1}^L z_{lk} \geq \sum_{j=1}^n \sum_{i=1}^n x_{ik} a_{ij} - 2 \sum_{l=1}^L z_{lk} + y_k \\ & x_{ik} \in \{0, 1\}, y_k \in \{0, 1\}, z_{lk} \in \{0, 1\}, i = 1, 2, \dots, n, k = 1, 2, \dots, n, l = 1, 2, \dots, L \end{aligned}$$

其中， n 是网络中的节点数， L 是网络中的边数， z_{lk} 表示边 e_l 是否属于 k 社团， $e_l = (v_i, v_j)$ 表示连接节点 v_i 和 v_j 的边， x_{ik} 是指节点 v_i 是否属于第 k 个社团[8]。该模型一定程度上克服了 Q 和 D 所暴露出的错误识别现象及分辨率极限问题，但没有考虑到节点之间的权重影响。

本文在该组合优化模型基础上进行了加权处理，能更进一步表述出社团与社团之间的联系。给定一个加权网络，社团识别问题是将该网络划分成尽可能多的不重叠子网络，使每个子网络满足给定的弱社团定义。

约定 n 是一个正整数，定义下面几个集合

$$S^n = \{S \in R^{n \times n} \mid S \text{ 是对称矩阵}\}$$

$$S_0^n = \{S \in R^{n \times n} \mid S \text{ 是对称矩阵, } S \geq 0, \text{ 且 } \text{diag}(S) = 0\}$$

其中， $S \geq 0$ 代表矩阵 S 的所有元素都是非负的， $\text{diag}(S)$ 代表由矩阵 S 的对角线上的元素组成的向量[9]。其中， a_{ij} 表示节点 v_i 和 v_j 之间的权重，对每一个矩阵 $M \in R^{n \times n}$ ，定义 $L(M)$ 是一个 $n \times n$ 维的矩阵，它的分量有如下表达式

$$L(M)_{ij} = \begin{cases} -M_{ij}, & i \neq j \\ \sum_{k \neq i} M_{ik}, & i = j \end{cases}$$

L 构成 $R^{n \times n}$ 上的线性算子。对于每一个矩阵 $A \in S_0^n$ ，构造一个带权重的无向图 $G(V, E)$ ，其中 V 是图的顶点集，矩阵 A 是该图邻接矩阵[9]。

设 y_k 表示第 k 个社团是否为空， $y_k = 0$ 当且仅当第 k 个社团没有节点[8]。即，

$$y_k = \begin{cases} 1, & \text{第 } k \text{ 个社团非空} \\ 0, & \text{其他} \end{cases}$$

令 A^k 表示第 k 个社团的邻接矩阵, $A^k = [a_{ij}^k]$ 即,

$$a_{ij}^k = \begin{cases} a_{ij}, & v_i \text{ 和 } v_j \text{ 都属于第 } k \text{ 个社团} \\ 0, & \text{其他} \end{cases}$$

弱社团定义可以表示为

$$\text{tr}(L(A^k)) \geq 2 \sum_{i=\text{rank}(A^k)+1}^n \sum_{j=1}^{\text{rank}(A^k)} a_{ij}$$

即删除候选网络的一些边使得删去的边的总权重比网络其余部分的权重小。其中目标函数

$$\sum_{k=1}^n y_k \text{tr}(L(A^k))$$

表示出了删去的边的权重尽可能小, 且得到的社团越多越好。因此建立了如下社团识别模型:

$$\begin{aligned} & \max \sum_{k=1}^n y_k \text{tr}(L(A^k)) \\ & \text{s.t. } \text{tr}(L(A^k)) \geq 2 \sum_{i=\text{rank}(A^k)+1}^n \sum_{j=1}^{\text{rank}(A^k)} a_{ij}. \end{aligned}$$

3. 均值切割算法

若 $v_i \in G'$ 且 $v_j \in G'$ 则 e_{ij} 为 G' 内边, s_{ij} 表示 e_{ij} 的权值。如果, 仅 $v_i \notin G'$ 或 $v_j \notin G'$ 成立, 则 e_{ij} 为割边, 记为 e'_{ij} , s'_{ij} 表示 e'_{ij} 的权值。 G' 所有割边组成的集合记为割集 L' 。

步骤 1: 输入图 $G=(V, E)$, $k=1$, $\tilde{P}=\phi$, 随机选取 $G' \subset G$, 若满足 $\sum(L(s)) \geq \sum(L'(s'))$, 则 G' 合格。否则, 转到步骤 3。

步骤 2: 判断 $E(L(s)) \geq E(L'(s'))$ 是否成立, 若不等式成立, 则 G' 为一个合格社团切割, $k=k+1$, $\tilde{P}=\tilde{P} \cup \{G'\}$, 否则, 返回步骤 1。

步骤 3: 设 $G=G \setminus G'$, 返回步骤 1。

步骤 4: 重复上述步骤, 直到 $G=\phi$, 完成所有切割, 输出 k , \tilde{P} 是 G 的最优划分。

4. 数值实验

Zachary (1977)通过两年的观察分析了著名空手道俱乐部网络, 该网络是由 34 个节点和 78 条边组成的, 其中节点代表俱乐部的成员, 边代表俱乐部成员之间的友谊。由于俱乐部的管理者和俱乐部的教练之间产生了分歧, 俱乐部被分成了几个子俱乐部。Zhang 等人[8] (2010)在不加权的条件下, 根据弱社团定义将网络划分为以下四个社团(图 1):

$$\begin{aligned} C_1 &= \{5, 6, 7, 11, 17\}, C_2 = \{24, 25, 26, 27, 28, 30\}, C_3 = \{1, 2, 3, 4, 8, 12, 13, 14, 18, 20, 22\}, \\ C_4 &= \{9, 10, 15, 16, 19, 21, 23, 29, 31, 32, 33, 34\} \end{aligned}$$

然而, 在上述划分中俱乐部成员间关系的重要性没有提现。在本文中, 通过对边进行加权处理, 以权重表示成员间关系的强弱, 对该算例进行随机数值模拟。在优化后的弱社团定义下该网络被划分为以下三个社团(图 2):

$$\begin{aligned} G_1 &= \{1, 4, 5, 6, 7, 11, 12, 13, 17, 18, 22\}, G_2 = \{2, 3, 8, 9, 10, 14, 16, 20, 31, 32\} \\ G_3 &= \{15, 19, 20, 21, 23, 24, 25, 27, 28, 29, 30, 33, 34\} \end{aligned}$$

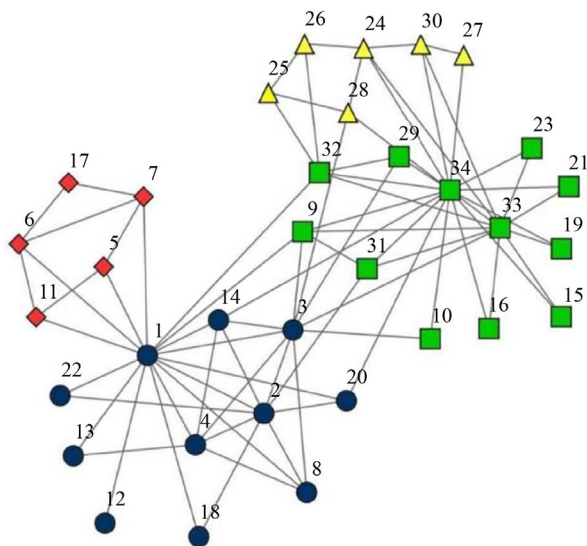


Figure 1. Unweighted
图 1. 不加权

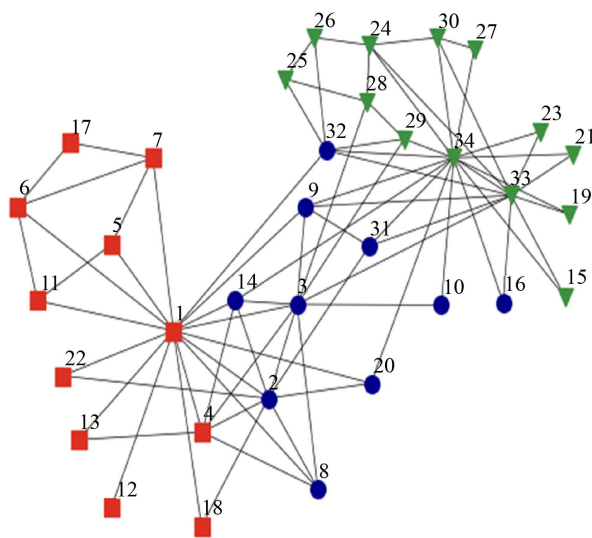


Figure 2. Weighted
图 2. 加权

通过对该现实算例进行加权处理, 结合现实背景表明加权后的划分更加直观的反映出社团间的关系。

5. 结论

本文基于 Radicchi 等人(2004)的弱社团意义下的社团定义, 提出了一个优化的弱社团定义。根据该定义, 建立了加权网络社团结构检测的模型, 该模型更进一步阐述了社团与社团之间的联系。为求解该模型, 给出均值切割算法, 数值实验验证提出的模型及算法对实际问题更具有应用性, 但存在使用模型和算法检测到的加权网络社团结构与假设的实验结果有偏差的问题。

基金项目

本文受到内蒙古大学校级大学生创新创业训练计划项目(项目编号: 202011223)和内蒙古自治区高等

学校科学技术研究项目(项目编号: NJZY19005)的资助。

参考文献

- [1] Albert, R. and Barabási, A.-L. (2002) Statistical Mechanics of Complex Networks. *Reviews of Modern Physics*, **74**, 47-97. <https://doi.org/10.1103/RevModPhys.74.47>
- [2] Girvan, M. and Newman, M.E.J. (2002) Community Structure in Social and Biological Networks. *Proceedings of the National Academy of Sciences of the USA*, **99**, 7821-7826. <https://doi.org/10.1073/pnas.122653799>
- [3] Ravasz, E., Somera, A.L., Mongru, D.A., Oltvai, Z.N. and Barabasi, A.-L. (2002) Hierarchical Organization of Modularity in Metabolic Networks. *Science*, **297**, 1551-1555. <https://doi.org/10.1126/science.1073374>
- [4] Radicchi, F., Castellano, C., Cecconi, F., Loreto, V. and Parisi, D. (2004) Defining and Identifying Communities in Networks. *Proceedings of the National Academy of Sciences of the USA*, **101**, 2658-2663. <https://doi.org/10.1073/pnas.0400054101>
- [5] Newman, M.E.J. (2006) Modularity and Community Structure in Networks. *Proceedings of the National Academy of Sciences of the USA*, **103**, 8577-8582. <https://doi.org/10.1073/pnas.0601602103>
- [6] Newman, M.E.J. and Girvan, M. (2004) Finding and Evaluating Community Structure in Networks. *Physical Review E*, **69**, Article ID: 026113. <https://doi.org/10.1103/PhysRevE.69.026113>
- [7] Li, Z.P., Zhang, S.H., Wang, R.-S., Zhang, X.-S. and Chen, L.N. (2008) Quantitative Function for Community Detection. *Physical Review E*, **77**, Article ID: 036109. <https://doi.org/10.1103/PhysRevE.77.036109>
- [8] Zhang, X.-S., Li, Z.P., Wang, R.-S. and Wang, Y. (2012) A Combinatorial Model and Algorithm for Globally Searching Community Structure in Complex Networks. *Journal of Combinatorial Optimization*, **23**, 425-442. <https://doi.org/10.1007/s10878-010-9356-0>
- [9] 刘歆, 吴国宝, 张瑞, 张在坤. 一种连续的谱聚类优化模型[J]. 计算数学, 2018, 40(4): 354-366.