

# 基于弱社团定义的社团划分模型与算法

李文静, 高彩霞\*

内蒙古大学数学科学学院, 内蒙古 呼和浩特  
Email: lwj00828@163.com, \*smsgcx@imu.edu.cn

收稿日期: 2020年11月26日; 录用日期: 2020年12月11日; 发布日期: 2020年12月18日

## 摘要

社团结构是复杂网络的重要特征之一。为了得到更合理的网络社团结构的划分结果, 文中给出了一种基于弱社团定义的整数优化模型。由于该模型是组合优化问题, 因而较难快速求解。文中基于该模型将其转化为连续优化近似模型, 提出一个高效的块坐标下降算法。实验结果证明了基于弱社团定义的优化模型与算法的有效性。

## 关键词

复杂网络, 社团结构, 弱社团, 块坐标下降算法

# Model and Algorithm of Community Division Based on Weak Community Definition

Wenjing Li, Caixia Gao\*

School of Mathematical Sciences, Inner Mongolia University, Hohhot Inner Mongolia  
Email: lwj00828@163.com, \*smsgcx@imu.edu.cn

Received: Nov. 26<sup>th</sup>, 2020; accepted: Dec. 11<sup>th</sup>, 2020; published: Dec. 18<sup>th</sup>, 2020

## Abstract

Community structure is one of the important characteristics of complex networks. In order to obtain a more reasonable division of the network community structure, this paper considers an integer optimization model that based on the definition of weak community. This problem is often formulated as combinatorial optimization model. Consequently, to solve it efficiently is difficult.

\*通讯作者。

**Based on this model, this paper transforms it into a continuous optimization approximate model, and proposes an efficient block coordinate descent algorithm. The experimental results prove the effectiveness of the optimization model and algorithm based on the definition of weak communities.**

## Keywords

Complex Network, Community Structure, Weak Community, Block Coordinate Descent Algorithm

Copyright © 2020 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

近年来, 复杂网络的研究处于蓬勃发展阶段。随着对复杂网络的深入研究, 发现众多实际网络存在的一个共同特征, 称之为网络中的社团结构。它是指网络中的节点可以分组, 组内节点间的连接比较稠密, 组间的连接比较稀疏。社团结构在实际系统中有着非常重要的意义。例如在万维网中, 不同的社团可能代表着不同主题的网页; 在新陈代谢网络以及神经网络中, 社团反映了功能单位。相关的社团结构也存在于社会网络, 互联网以及食物网络等。研究网络的社团结构对于理解整个网络的结构与功能十分重要。

目前已经存在着大量探索复杂网络社团结构发现方法的研究[1] [2]。其中社团发现的一种较受关注的是谱分方法[3] [4] [5] [6]。它的主要思想是对通过邻接矩阵形成的 Laplace 矩阵或标准矩阵的特征值, 特征向量的分析, 探索网络中的社团结构。谱平分的思想是根据 Laplace 矩阵的第二小的特征值  $\lambda_2$  将其分为两个社团。谱平分法对于社团结构明显的网络非常有效, 但当网络的结构不明显时, 往往得不到理想的划分结果。一种被广泛使用的社团评估函数, 称之为模块函数 Q [7]。然而, 最近的研究表明, 在一些网络社团划分中模块函数 Q 存在着分辨率限制问题[8]。由李等人提出的模块密度函数 D 是一个改善社团划分的定量函数[9]。模块密度函数 D 在一定的程度上缓解了该限制问题, 但是这两种函数存在的一个共同问题是错误识别, 即划分得到的一些社团不满足弱社团的定义[10]。

文中基于弱社团定义建立了一个新的整数优化模型, 能够克服网络社团结构划分中的错误识别问题, 一般的, 整数优化模型被描述为组合优化问题, 因而较难求解, 故考虑它的连续优化近似, 并提出一个高效的块坐标下降算法。实验结果显示该模型与算法是可行的。

## 2. 模型描述

### 2.1. 复杂网络与社团结构

给出一个无权无向网络  $G(V, E)$ , 令  $V = [n] = \{1, 2, \dots, n\}$  为网络的节点集, 其中  $n$  为该网络中的节点数目。令  $E = \{e_1, e_2, \dots, e_m\}$  为边的集合, 其中  $e_l = (i, j)$  代表着节点  $i$  与  $j$  间的连边,  $m$  为网络中的连边数目, 即  $m = |E|$ 。给定一个网络, 将其划分为不重叠的社团结构, 将节点集  $[n]$  划分为  $\{C_1, C_2, \dots, C_k\}$ , 即  $C_a \cap C_b = \Phi, \forall a \neq b, \bigcup_{a=1}^k C_a = [n]$ 。

令  $A = [a_{ij}]^{n \times n}$  为网络的邻接矩阵, 其中  $a_{ii} = 0$ , 且  $\forall i \in [n]$ , 有  $i \in C_a, j \in C_b, i \neq j$ ,

$$a_{ij} = \begin{cases} 1, & (i, j) \in E \\ 0, & (i, j) \notin E \end{cases}.$$

令  $V_s$  为  $V$  的子集,  $\bar{V}_s$  为  $V$  的补集。若  $V_s$  满足弱社团的定义, 则满足

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

其中  $L(V_s, V_s) = \sum_{i \in V_s} \sum_{j \in V_s} a_{ij}$ ,  $L(V_s, \bar{V}_s) = \sum_{i \in V_s} \sum_{j \in \bar{V}_s} a_{ij}$ 。

## 2.2. 基于弱社团定义迹的最大化函数

引入一个指示向量  $t_s(i) = 1$ , 节点  $i$  属于  $V_s$ , 否则  $t_s(i) = 0$ 。因此

$$t_s = \begin{bmatrix} t_s(1) \\ t_s(2) \\ \vdots \\ t_s(n) \end{bmatrix}, s = \{1, 2, \dots, n\}.$$

定义对角矩阵  $D = \text{diag}\{d_1, d_2, \dots, d_n\}$ , 其中  $d_i = \sum_{j=1}^n a_{ij}$ 。弱社团定义

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

可具体展开为:  $t_s^T A t_s - t_s^T (D - A) t_s > 0$ ,

即

$$t_s^T (2A - D) t_s > 0.$$

定义一个分配矩阵  $T = [t_1, t_2, \dots, t_k]$ ,  $k = 1, 2, \dots, n$ 。其每一行对应着网络中相应的一个节点, 每一列对应着相应划分的社团。进一步的, 满足弱社团定义的网络社团划分情况可描述如下:

$$\sum_{s=1}^k t_s^T (2A - D) t_s = \text{tr}[T^T (2A - D) T]$$

故问题可描述为:

$$\begin{aligned} & \max_T \left\{ \text{tr}[T^T (2A - D) T] \right\} \\ & \text{s.t. } T \geq 0, \\ & T \in \{0, 1\}^{n \times k} \end{aligned} \quad (1)$$

## 2.3. 连续优化近似模型

问题(1)是离散的, 因此难以求解。故将  $T$  松弛得到  $S \in [0, 1]^{n \times k}$ 。并对  $S$  的每一行向量进行稀疏约束, 以保证结果的稀疏性。定义矩阵  $B = -(2A - D)$ 。因此连续优化近似模型可表述如下:

$$\begin{aligned} & \min_S \text{tr}\{S^T B S\} \\ & \text{s.t. } \|s_i\|^2 = 1, i = 1, 2, \dots, n, \\ & \|s_i\|_0 \leq p, \\ & S \geq 0 \end{aligned} \quad (2)$$

进一步可描述为:

$$\begin{aligned}
& \min_S \langle B, SS^T \rangle \\
& \text{s.t. } \|s_i\|^2 = 1, i = 1, 2, \dots, n, \\
& \quad \|s_i\|_0 \leq p, \\
& \quad S \geq 0
\end{aligned} \tag{3}$$

假设社团的数目为  $k$ , 且  $1 \leq p \leq k$ ,  $p$  是一个整数,  $\|s_i\|_0 \leq p$ 。当  $p \leq k$  时,  $l_0$  约束消失。

## 2.4. 社团检测框架

利用稀疏正松弛将离散问题转化为连续优化近似模型[11]。假设从(3)中得到结果  $S^*$ , 用一种直接舍入方法得出最终的社团划分矩阵  $T^*$ 。令  $t_i^T$  为  $T$  的第  $i$  行,  $e_{j_0} \in R^k$  是第  $j_0$  个元素为 1 的单位向量。

$$t_i \leftarrow e_{j_0}, j_0 = \arg \max_j (s_i)_j。$$

社团检测框架如下:

- 步骤 1: 从不同的随机点开始代入求解问题(3)得到近似结果。
- 步骤 2: 用直接舍入方法检测社团结构。
- 步骤 3: 多次重复步骤 1 和步骤 2, 直到所得的划分矩阵使所有社团均满足弱社团定义且是合格的社团结构。

## 3. 块坐标下降算法

在这部分中, 用块坐标下降算法求解问题(3)。假设网络被划分为  $k$  个社团。令  $f(S) = \langle B, SS^T \rangle$ 。对于第  $i$  个子问题, 可以求解如下子问题:

$$s_i = \arg \min_{x \in s_i} f(s_1, \dots, s_{i-1}, x, s_{i+1}, \dots, s_n) + \frac{\sigma}{2} \|x - \bar{s}_i\|^2,$$

其中  $\bar{s}_i$  是某个参考点, 且  $\sigma > 0$ 。

块坐标下降算法具体描述如下:

- 步 1: 任意给出一个  $S_0, \sigma > 0$ , 令  $t = 0, \varepsilon > 0$ 。
- 步 2: 当  $i = 1, 2, \dots, n$ , 做

$$s_i^{t+1} = \arg \min_{x \in s_i} f(s_1^{t+1}, \dots, s_{i-1}^{t+1}, x, s_{i+1}^t, \dots, s_n^t) + \frac{\sigma}{2} \|x - s_i^t\|^2, t + 1。$$

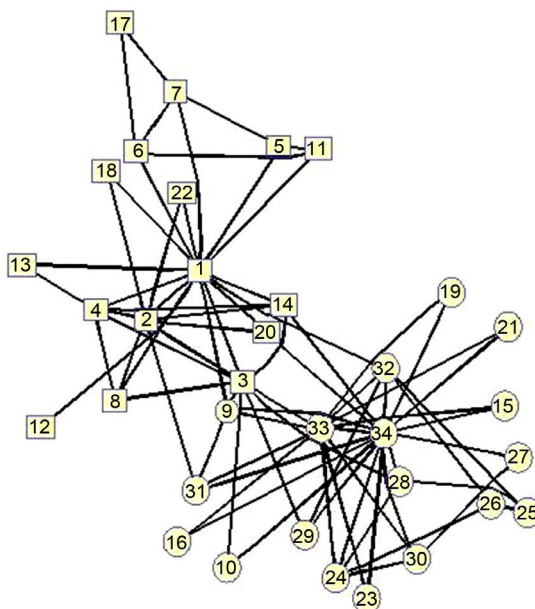
- 步 3: 当  $\|S^{t+1} - S^t\|_F < \varepsilon$ , 则得到  $S^{t+1}$ , 否则重复进行步 2。

## 4. 实验结果与分析

为了验证基于弱社团定义的优化模型与块坐标下降算法在复杂网络社团检测中的有效性与可行性, 在本节中选取 Zachary 空手道俱乐部网络[12]进行实验验证, 并与谱平分方法以及基于模块函数  $Q$  的谱平分方法对网络社团划分结果做对比和分析。

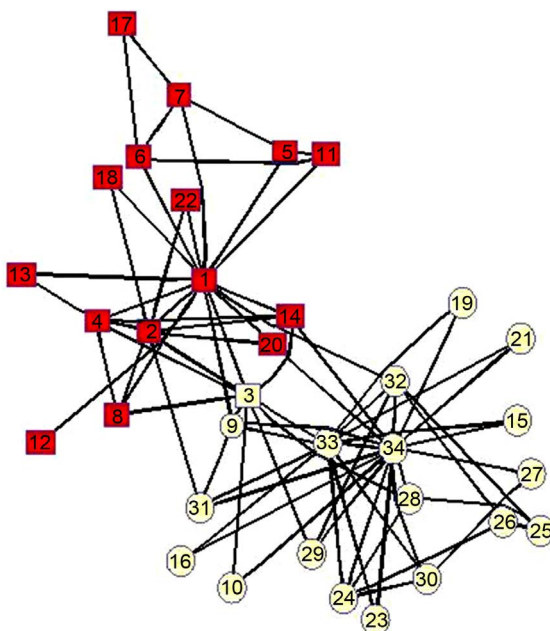
Zachary 空手道俱乐部网络是检验不同社团划分算法的一个经典实际网络。其组成有 34 个节点代表着俱乐部成员, 78 条边代表着俱乐部成员间有交往的关系。在调查后发现, 由于该俱乐部主管与校长之间因是否提高收费问题产生争执, 致使俱乐部成员分裂以主管与校长为核心的 2 个小社团。其中一个社团包含的节点为: 1, 2, 3, 4, 5, 6, 7, 8, 11, 12, 13, 14, 17, 18, 20, 22。另一组则包含余下的节点。该网络真实

的划分情况如图 1 所示。节点 1 表示这俱乐部中的主管, 节点 33 代表了俱乐部的校长。圆形与方形节点代表着分裂后小俱乐部中的各个成员。



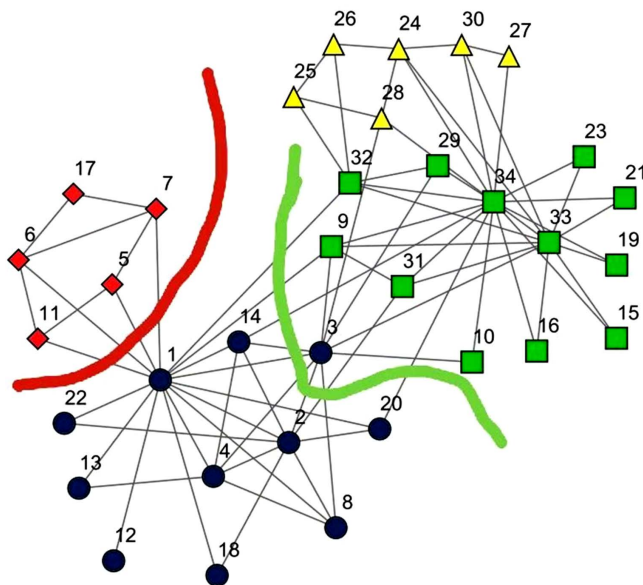
**Figure 1.** The karate club network diagram  
**图 1.** 空手道俱乐部关系图

图 2 为谱平分方法划分 Zachary 网络得到的结果。图中方形节点与圆形节点分别代表了原 Zachary 网络的实际划分结果, 而阴影和非阴影则代表了利用谱平分算法得到的结果。从图中可以直观的看出, 节点 3 被错误划分, 即谱平分算法将网络划分为两个社团时出现了错误识别的现象。



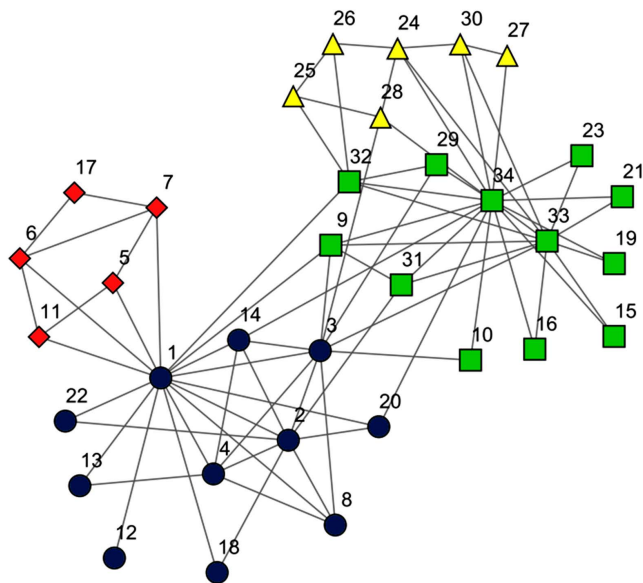
**Figure 2.** The community structure detected by spectral average method  
**图 2.** 谱平分方法划分的社团结构

在图 3 中, 基于模块函数  $Q$  的谱分方法将网络划分为 3 个社团:  $C_1 = \{1, 2, 4, 8, 12, 13, 14, 18, 20, 22\}$ ,  $C_2 = \{5, 6, 7, 11, 17\}$ ,  $C_3 = \{3, 9, 10, 15, 16, 19, 21, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34\}$ 。这表明了该算法不但可以检测出两个主要的社团, 还可以进一步的检测出网络中的小社团。由于该算法错误划分了节点 3, 故不能得到稳定合格的社团结构。



**Figure 3.** The community structure detected by spectral approach based on modularity ( $Q$ )  
**图 3.** 基于模块函数  $Q$  的谱分方法划分的社团结构

已经有学者们讨论过 Zachary 空手道俱乐部网络可以划分为更多的小社团, 多数学者们认为该网络划分为 4 个社团更为合理。图 4 是块坐标下降算法将网络划分为 4 个均满足弱社团定义的社团结构, 且不能进一步划分。如表 1 所示, 该划分结果与文献[10]中的结果一致, 且节点 3 得到了正确的划分。



**Figure 4.** The community structure detected by the algorithm in the article  
**图 4.** 文中的算法划分得到的社团结构

**Table 1.** The results of the algorithm's division of the karate club network**表 1.** 文中算法对空手道俱乐部网络的划分结果

社团	节点名称
$C_1$	1, 2, 3, 4, 8, 12, 13, 14, 18, 20, 22
$C_2$	5, 6, 7, 11, 17
$C_3$	9, 10, 15, 16, 19, 21, 23, 29, 31, 32, 33, 34
$C_4$	24, 25, 26, 27, 28, 30

基于弱社团定义的优化模型与算法克服了谱平分方法, 基于模块函数  $Q$  谱分方法在空手道俱乐部网络中错误划分个别节点的情况, 该方法不但划分得到的每一个社团都满足弱社团定义, 而且能够检测出网络中更为复杂的小社团。

## 5. 结论

文中建立的基于弱社团定义的优化模型与算法得到了正确的社团划分结果。其划分得到的每一个社团均满足弱社团定义且是合格的社团结构。该方法能够检测出网络中复杂的小社团, 将网络社团结构划分得更为精确合理。文中的方法在小型网络中取得了良好的划分效果, 进一步将这种方法应用于一些大型的网络社团划分的研究中, 这也是日后学习研究的方向。

## 基金项目

内蒙古自治区高等学校科学技术研究项目(NJZY19005), 内蒙古大学研究生科研创新基金项目(10000-15010109)。

## 参考文献

- [1] Newman, M.E.J. (2003) The Structure and Function of Complex Networks. *SIAM Review*, **45**, 167-256. <https://doi.org/10.1137/S003614450342480>
- [2] Shen, H.W. and Cheng, X.Q. (2010) Spectral Methods for the Detection of Network Community Structure: A Comparative Analysis. *Journal of Statistical Mechanics: Theory and Experiment*, **2010**, Article ID: 10020. <https://doi.org/10.1088/1742-5468/2010/10/P10020>
- [3] Luxburg, U.V. (2007) A Tutorial on Spectral Clustering. *Statistics and Computing*, **17**, 395-416. <https://doi.org/10.1007/s11222-007-9033-z>
- [4] Wang, X., Qian, B. and Davidson, I. (2014) On Constrained Spectral Clustering and Its Applications. *Data Mining and Knowledge Discovery*, **28**, 1-30. <https://doi.org/10.1007/s10618-012-0291-9>
- [5] Wang, T.S., Lin, H.T. and Wang, P. (2016) Weighted-Spectral Clustering Algorithm for Detecting Community Structures in Complex Networks. *Artificial Intelligence Review*, **47**, 1-21. <https://doi.org/10.1007/s10462-016-9488-4>
- [6] Newman, M.E.J. (2006) Finding Community Structure in Networks Using the Eigenvectors of Matrices. *Physical Review E*, **74**, Article ID: 036104. <https://doi.org/10.1103/PhysRevE.74.036104>
- [7] Newman, M.E.J. (2006) Modularity and Community Structure in Networks. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 8577-8582. <https://doi.org/10.1073/pnas.0601602103>
- [8] Fortunato, S. and Barthelemy, M. (2007) Resolution Limit in Community Detection. *Proceedings of the National Academy of Sciences*, **104**, 36-41. <https://doi.org/10.1073/pnas.0605965104>
- [9] Zhang, X., Zhang, S., Wang, R., et al. (2008) Quantitative Function for Community Detection. *Physical Review E Statistical Nonlinear and Soft Matter Physics*, **77**, Article ID: 036109.
- [10] Zhang, X.S., Li, Z., Wang, R.S., et al. (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>
- [11] Zhang, J., Liu, H., Wen, Z., et al. (2018) A Sparse Completely Positive Relaxation of the Modularity Maximization for

Community Detection. *SIAM Journal on Scientific Computing*, **40**, A3091-A3120.  
<https://doi.org/10.1137/17M1141904>

- [12] Zachary, W.W. (1976) An Information Flow Model for Conflict and Fission in Small Groups. *Journal of Anthropological Research*, **33**, 452-473. <https://doi.org/10.1086/jar.33.4.3629752>