

# Community Discovery Algorithms Based on Node Connection Strength

Yongjun Luo<sup>1\*</sup>, Lixia Wang<sup>1</sup>, Tong Li<sup>2</sup>, Yong Yu<sup>1</sup>, Shilin Kan<sup>1</sup>, Jie Gu<sup>1</sup>

<sup>1</sup>Yunnan University, Kunming Yunnan

<sup>2</sup>Yunnan Agricultural University, Kunming Yunnan

Email: \*luoyj9511@163.com

Received: Aug. 7<sup>th</sup>, 2019; accepted: Aug. 19<sup>th</sup>, 2019; published: Aug. 26<sup>th</sup>, 2019

---

## Abstract

People are becoming more and more interested in the algorithm of community discovery in the network. In complex networks, nodes do not exist independently. From a structural point of view, if the connections of two nodes are similar, then the two nodes are likely to be in the same community. In this paper, we propose a new community discovery algorithm based on node connection strength (BNCSA), which is based on the connection strength between the number of direct connections and the number of indirect connections between nodes. Then, parameters are used to control the community structure granularity of the network, so as to avoid the disadvantage that the existing classical algorithms cannot find small communities whose granularity is less than a certain amount. Finally, we conducted some experiments to verify the partitioning results of the new algorithm. Experimental results show that the algorithm can effectively discover the community structure in real networks.

## Keywords

Complex Network, Node Connection Strength, Community Discovery

---

# 基于节点连接强度的社区发现算法

骆永军<sup>1\*</sup>, 王黎霞<sup>1</sup>, 李彤<sup>2</sup>, 郁湧<sup>1</sup>, 阚世林<sup>1</sup>, 顾捷<sup>1</sup>

<sup>1</sup>云南大学, 云南 昆明

<sup>2</sup>云南农业大学, 云南 昆明

Email: \*luoyj9511@163.com

收稿日期: 2019年8月7日; 录用日期: 2019年8月19日; 发布日期: 2019年8月26日

---

\*通讯作者。

文章引用: 骆永军, 王黎霞, 李彤, 郁湧, 阚世林, 顾捷. 基于节点连接强度的社区发现算法[J]. 计算机科学与应用, 2019, 9(8): 1604-1610. DOI: 10.12677/csa.2019.98180

## 摘要

人们对网络中社区发现的算法越来越感兴趣。在复杂网络中，节点并不是独立地存在，从结构上看，如果两个节点的连接情况很相似，那么这两个节点很可能处于同一个社区内。在本文中，我们提出了一种基于节点连接强度的社区发现算法(BNCSA)，该算法基于节点之间的直接连接数与间接连接数，度量节点之间的连接强度，然后用参数控制网络的社区结构的粒度，可以有效避免现有经典算法中无法发现粒度小于一定量的小型社区的弊端。最后，我们进行了一些实验来验证新算法的划分结果。实验结果表明，该算法能够有效地发现真实网络中的社区结构。

## 关键词

复杂网络，节点连接强度，社区发现

Copyright © 2019 by author(s) and Hans Publishers Inc.

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

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



Open Access

## 1. 引言

在现实世界中，许多系统可以被抽象成网络，例如因特网，人际关系网络，疾病传播网络和科学家合作网络。大量的研究表明，社区结构存在于复杂的网络中。节点在社区内紧密相连，而在社区之间的连接相对稀疏。社区结构在复杂网络中起着越来越重要的作用。它可以帮助我们理解复杂网络的功能，发现潜在的规律，并预测复杂网络的行为[1]。例如，社区结构代表社交网络中有相同兴趣和爱好的群体；在 Web 的网页中，网页在同一社区内有更多的链接；在文献网络中，同一社区内的文献都有相关的研究课题。

已经针对社区结构的研究开发了各种方法。有些人倾向于聚类图形，因此在聚类中存在密集的边集，并且在它们之间存在很少的边，例如，基于模块度的算法。但是，模块度不是评估的准确度量。一些人采用分层聚类的方法，并选择迭代地合并具有高相似性的顶点，或者去除连接具有低相似性的顶点的边，但是不区分顶点彼此的重要性。有些专注于桥接集群但忽略重要节点重要性的集线器。实际上，社区通常聚集在网络中的一些重要节点周围。

在本文中，我们提出了一种新的社区发现算法，称为基于节点连接强度的社区检测(BNCSA)网络算法。在结构等效的概念下，可以找到对形成社区很重要的节点。他们被认为在网络中发挥着至关重要的作用，特别是在社区内部。然后我们使用节点连接强度测量来扩展它们以形成聚类。

我们的方法具有以下特点：

- 1) 遍历每一个节点并加入邻节点之后形成社区，选取节点平均连接强度最大的社区作为待划分社区。最后用参数控制其他节点与该社区的最小连接数形成一个最终社区。
- 2) 基本思想很好地匹配真实网络中社区的演化过程，即社区逐渐从节点连接强度最大的社区进行扩展。
- 3) 高效。具有  $n$  个顶点和每个社区包含  $m$  个节点的网络上的运行时间是  $O(n^2m)$ 。它在我们的实验中胜过了 FN [2]。

## 2. 基于节点相似度的社区发现方法

首先给出一些将在算法中使用的概念。在传统的节点相似度的概念中，对于节点相似度的表述为： $G(V, E)$ 是一个无权无向图。其中， $V = \{1, 2, \dots, n\}$ 是 $G$ 中节点的集合； $E = \{(i, j) | i, j \in V\}$ 是 $G$ 中边的集合，且 $|V| = n, |E| = m$ ； $A$ 是其对应的邻接矩阵[3]，如果节点 $v_i$ 和 $v_j$ 之间有边，则其元素 $A_{ij} = 1$ ，否则， $A_{ij} = 0$ 。

### 2.1. 节点相似度度量

#### 2.1.1. 基于共同邻居数的 CN 指标

从网络的拓扑角度来看，考虑到两个节点的共同邻居数量，即 Common Neighbors (CN) [4]。两个节点拥有更多共同邻居，表明两个节点之间的连接更为紧密和相似，因此它们更有可能属于同一社区。例如，两个喜欢跑步的人，他们彼此不认识，但是两人有一些共同认识的喜欢跑步的人，如果他们相互认识的人很多，则两个人在同一地区或在同一个爱好者协会的可能性很大。网络中某个节点 $v_x$ ，其所有邻居节点的集合用 $\Gamma(x)$ 表示，则任意两个节点 $v_x$ 和 $v_y$ 的相似度可表示为：

$$S_{xy} = |\Gamma(x) \cap \Gamma(y)| \quad (1)$$

#### 2.1.2. 其他规范化指标

在 CN 指标的基础上，增加了节点度的影响，得到 6 种规范的相似性指标，包括 RA 指标[5]、HPI 指标[6]等。HPI 指标的定义是社区内两个顶点的共同邻居的数量比上他们中较小节点的度，所以大度节点与其他节点之间，更容易具有高的相似性：

$$S_{xy} = \frac{|\Gamma(x) \cap \Gamma(y)|}{\min\{k_x, k_y\}} \quad (2)$$

### 2.2. 一种新的节点相似度

在本文中我们对 HPI 指标进行改进，采用节点连接强度来评价节点之间的相似性程度。节点连接强度是节点之间的连接程度及连边在网络中重要程度的一种度量，是节点之间直接和间接连接数与节点中较小度的比值，其定义为：

$$\alpha(v_x, v_y) = \frac{e_{xy}(1 + |\Gamma(x) \cap \Gamma(y)|)}{\min(k_x, k_y)} \quad (3)$$

则一个网络的平均节点连接强度为：

$$\bar{\alpha} = \frac{\sum_{x,y=1}^n \alpha_{xy}}{2n} \quad (4)$$

节点连接强度反映了节点之间通过共同邻居加强彼此关联的程度。 $e_{xy}$ 是节点之间直接连接的数量， $k_x, k_y$ 是它们的度；显然 $0 \leq \alpha \leq 1$ 。

### 2.3. 基于节点连接强度的社区发现算法

#### 2.3.1. 算法具体步骤

基于节点连接强度的社区发现算法的具体步骤如下所示：

输入：网络的邻接矩阵  $A$ 。

输出：网络社区结构  $C$ 。

```

BEGIN
  FOR i=0 TO N DO
    FOR i=0 TO N DO
      根据式(3)计算节点及其邻居构成的社区中节点与节点的连接强度
    END FOR
    FOR i=0 TO N DO
      IF 节点与平均节点连接强度最大的社区的连边数量大于该节点度数的  $\lambda$  倍
        将该节点加入到平均节点连接强度最大的社区中
      END IF
    END FOR
    PRINT 平均节点连接强度最大的社区结构
    删除平均节点连接强度最大的社区结构
  END FOR
END

```

在 BNCSA 算法中我们设置了一个参数  $\lambda$  用来控制社区内节点与平均连接强度最大的社区的最小连接数, 实验发现  $\lambda$  取值为[0.3,0.6]时得到的社区结构更符合实际情况, 当  $\lambda$  过小时会将与社区连接度低的节点加入到社区, 导致降低社区节点的整体连接强度; 而当  $\lambda$  过大时, 会导致社区的粒度过小。

### 2.3.2. 算法时间和空间复杂度分析

BNCSA 算法中求解社区的节点平均连接强度是最为耗时的部分, 设社区有  $n$  个节点, 当网络被均分为几个社区时且每个平均节点连接强度最大的社区都不再加入新的节点, 设每个社区包含  $m$  个节点, 这时基于节点连接强度的社区发现算法的时间复杂度为  $O(n^2m)$ 。BNCSA 算法使用矩阵存储图边关系, 其空间复杂度为  $O(n^2)$ 。

## 3. 实验结果与分析

为了测试本文提出的复杂网络社区发现算法, 我们将 BNCSA 算法在四个常用的真实网络数据集上进行实验, 并将所得到的结果与 GN [7]、FN [2]划分的结果进行对比。本文算法在 win10 平台上采用 MATLAB R2012a 进行开发, 运行环境为 Intel 2.20 GHz Processor, 12 GB RAM。实验结果表明, 本文提出的社区发现算法是可行的, 能够较好的对复杂网络社区进行划分, 具有较高的精度以及较高的运行效率。

为了验证本文提出的基于节点连接强度的社区发现算法的有效性, 将算法应用在四个真实的基准网络: Karate 网络[8]、Lesmis 网络[9]、Polbook 网络[10]、Netscience 网络[11], 真实网络的信息描述如表 1 所示。

Table 1. Real network data sets

表 1. 真实网络的数据集

| 网络         | 节点数量 | 边数   | 平均度  | 网络直径 | 聚类系数  |
|------------|------|------|------|------|-------|
| Karate     | 34   | 78   | 1.77 | 7    | 0.588 |
| Lesmis     | 77   | 254  | 6.60 | 5    | 0.736 |
| Polbooks   | 105  | 441  | 8.40 | 7    | 0.488 |
| Netscience | 1589 | 2742 | 3.45 | 17   | 0.878 |

### 3.1. 实验数据集

- 1) 空手道俱乐部网络(Karate network)

空手道俱乐部网络是 Zachary 构造的美国空手道俱乐部成员关系网,网络由 34 个节点、78 条边组成,节点表示俱乐部成员,边表示成员之间存在社交互动。该俱乐部就是否抬高俱乐部收费问题产生分歧,分成了以管理者和教练为中心的两个小社区。

2) 悲惨世界人物关系网络(Lesmis network)

Lesmis 网络是 D. E. Knuth 根据 Victor Hugo 的小说《悲惨世界》章节中的人物关系所构建的真实网络。网络中的节点代表小说中的角色,边代表两个角色同时出现在同一场景中。该网络共有 77 个节点,254 条边。

3) Krebs 美国政治图书网络(Polbooks network)

美国政治图书网络来源于 21 世纪初选举美国总统时期,科学家从亚马逊网站上统计出政治图书的销售情况所构成。该网络包括 105 个节点,441 条边,每个节点代表销售的图书,边表示两本书的购买者是同一个人。该网络因为政治观点不同而形成三个群体:自由主义、保守主义以及无明显政治观点群体。

4) 网络科学家论文合作网(Netscience Network)

在复杂网络的研究过程中,许多科学家一起合作发表过文章。Newman 等人通过总结复杂网络研究领域 1589 位科学家之间的合作关系,构建了科学家合作网络,网络包括 1589 个节点,2742 条边。网络中的每个节点代表每位科学家,两个节点间的连边代表该两位科学家之间存在合作关系。

### 3.2. 评价指标

#### 模块度

模块度[12]是目前常用的一种衡量网络社区结构强度的方法,模块度在许多文献中被广泛用于评估社区的质量,最早是由 Newman 提出的。模块度的定义为:

$$Q(C) = \frac{1}{2m} \sum_{k=1}^C \sum_{v_i \in C_k, v_j \in C_k} \left( A_{ij} - \frac{k_i k_j}{2m} \right) \tag{5}$$

其中  $m$  是网络中的边数,  $A$  表示网络的邻接矩阵,邻接矩阵  $A_{ij}$  表示节点  $i$  和  $j$  之间的关系,如果  $A_{ij} = 1$ ,则  $i$  和  $j$  之间存在边,否则  $A_{ij} = 0$ 。 $k_i$  是节点  $i$  的度,  $C_i$  是节点  $i$  所属的社区。 $C$  是通过算法获得的社区数量。模块度值越大,表示网络划分的社区结构的强度越强,即划分质量越好。

### 3.3. 实验结果

我们进行了比较实验。在实验中将  $\lambda$  值设定为 0.4,结果如下所述。

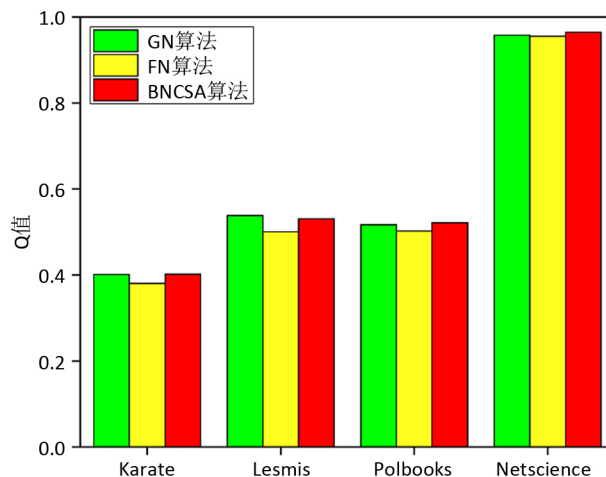


Figure 1. Comparison chart of Q values

图 1. Q 值的对比图

从图 1 中可以看出, 三种算法所对应的模块度  $Q$  值, 在 Karate 网络、Lesmis 网络、Polbooks 网络这三个经典的社区发现网络中, FN 的  $Q$  值最小, 本文提出的算法在社区划分后的  $Q$  值与 GN 算法相当。在更大的 Netscience 网络中, 本文提出的算法比 GN 算法的  $Q$  值更大, 表明本文算法在大型网络中划分出的社区结构的强度更强。

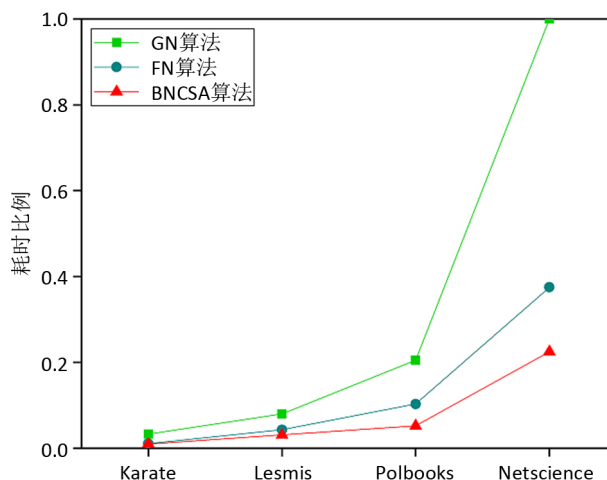


Figure 2. Time-consuming comparison chart  
图 2. 耗时对比图

为了证明 BNCSA 算法具有良好的时间复杂度, 分别在相同的计算机环境下测试 GN、FN 以及 BNCSA 算法划分 Karate 网络、Lesmis 网络、Polbooks 网络以及 Netscience 网络所需要的时间, 实验结果如图 2 所示, 纵坐标为所需时间占 GN 在 Netscience 上的时间比例。从图 2 中可以看到, 三种算法中 BNCSA 算法耗时最短, GN 耗时最长。当网络节点数较少时, FN 与 BNCSA 算法在耗时上几乎相同。然而, 随着网络节点的数量和边数的增加, BNCSA 算法在耗时方面与 FN 算法相比较, 前者有很大的优势, 所以本文算法适合大型网络的社区发现算法。

#### 4. 结论

社区发现是复杂网络中的一个重要研究方向。在本文中, 我们基于共同邻居数提出了一种新的基于节点连接强度的度量, 并基于节点连接强度提出了一种新的社区发现算法。实验结果表明, 该算法具有运算简单、实用的特点, 适用于大规模的复杂网络数据。

#### 参考文献

- [1] Coscia, M., Giannotti, F. and Pedreschi, D. (2011) A Classification for Community Discovery Methods in Complex Networks. *Statistical Analysis and Data Mining*, **4**, 512-546. <https://doi.org/10.1002/sam.10133>
- [2] Newman, M.E.J. (2004) Fast Algorithm for Detecting Community Structure in Networks. *Physical Review E*, **69**, Article ID: 066133. <https://doi.org/10.1103/PhysRevE.69.066133>
- [3] Wu, Z. and Lv, Z. (2016) A Community Detection Algorithm Based on Local Similarity. *Journal of Computer Engineering*, **42**, 196-203.
- [4] Guo, T. and Zhao, C. (2016) A New Measurement of Link Prediction Based on Common Neighbors. *Journal of China University of Metrology*, **27**, 121-124.
- [5] Zhou, T., Lv, L. and Zhang, Y.C. (2009) Predicting Missing Links via Local Information. *The European Physical Journal B*, **71**, 623-630. <https://doi.org/10.1140/epjb/e2009-00335-8>
- [6] Ravasz, E., Somera, A.L. and Mongru, D.A. (2002) Hierarchical Organization of Modularity in Metabolic Networks. *Science*, **297**, 1551-1555. <https://doi.org/10.1126/science.1073374>

- 
- [7] Girvan, M. and Newman, M.E. (2002) Community Structure in Social and Biological Networks. *Proceedings of the national Academy of Sciences*, **99**, 7821-7826. <https://doi.org/10.1073/pnas.122653799>
  - [8] Zachary, W.W. (1977) 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>
  - [9] Knuth, D.E. (1993) *The Stanford GraphBase: A Platform for Combinatorial Computing*. ACM, New York.
  - [10] Valdis Krebs. <http://www.orgnet.com>
  - [11] Newman, M.E.J. (2001) The Structure of Scientific Collaboration Networks. *Proceedings of the National Academy of Sciences*, **98**, 404-409. <https://doi.org/10.1073/pnas.98.2.404>
  - [12] Girvan, M. and Newman, M.E. (2004) Finding and Evaluating Community Structure in Networks. *Physical Review E*, **69**, Article ID: 026113. <https://doi.org/10.1103/PhysRevE.69.026113>

#### 知网检索的两种方式:

1. 打开知网首页: <http://cnki.net/>, 点击页面中“外文资源总库 CNKI SCHOLAR”, 跳转至: <http://scholar.cnki.net/new>, 搜索框内直接输入文章标题, 即可查询;  
或点击“高级检索”, 下拉列表框选择: [ISSN], 输入期刊 ISSN: 2161-8801, 即可查询。
2. 通过知网首页 <http://cnki.net/>顶部“旧版入口”进入知网旧版: <http://www.cnki.net/old/>, 左侧选择“国际文献总库”进入, 搜索框直接输入文章标题, 即可查询。

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: [csa@hanspub.org](mailto:csa@hanspub.org)