

基于信号自适应传递的社团发现算法

谭春妮¹, 张玉梅², 张嘉桐³, 吴晓军^{1,2*}

(1. 陕西师范大学 物理学与信息技术学院, 西安 710119; 2. 陕西师范大学 计算机科学学院, 西安 710119;

3. 西北大学 文化遗产学院, 西安 710127)

(* 通信作者电子邮箱 xjwu@snnu.edu.cn)

摘要:为了准确地检测出复杂网络的社团结构,提出一种基于信号自适应传递的社团发现方法。首先使信号在复杂网络上自适应地传递,从而获取网络中各节点对整个网络的影响向量,然后把网络中节点的拓扑结构转化成代数向量空间上的几何关系,最后结合聚类特性发现网络中的社团结构。为获取更加合理的空间向量,提出最佳传递次数,缩小搜索空间,增强算法寻优能力。该算法在计算机生成网络、Zachary 网络和美国大学生足球赛网络上进行实验测试,并与 GN 算法、谱聚类算法、极值优化算法和信号传递算法进行实验对比,社团划分的准确性和精确性均有所提高,证明该算法具有有效性和可行性。

关键词:复杂网络;社团结构;自适应;传递次数;社团发现算法

中图分类号: TP393.02 **文献标志码:** A

Community detection algorithm based on signal adaptive transmission

TAN Chunni¹, ZHANG Yumei², ZHANG Jiatong³, WU Xiaojun^{1,2*}

(1. School of Physics and Information Technology, Shaanxi Normal University, Xi'an Shaanxi 710119, China;

2. School of Computer Science, Shaanxi Normal University, Xi'an Shaanxi 710119, China;

3. School of Cultural Heritage, Northwest University, Xi'an Shaanxi 710127, China)

Abstract: In order to accurately detect the community structure of complex networks, a community detection algorithm based on signal adaptive transmission was proposed. First, the signal was adaptively passed on complex networks, thereby getting the vector affecting on the entire network of each node, then the topological structure of each node was translated into geometrical relationships of algebra vector space. Thus, according to the nature of the clustering, the community structure of the network was detected. In order to get the feasible spatial vectors, the optimum transfer number was determined, which reduced the searching space, and effectively strengthened the search capability of community detection. The proposed algorithm was tested on computer-generated network, Zachary network and American college football network. Compared with Girvan-Newman (GN) algorithm, spectral clustering algorithm, extremal optimization algorithm and signal transmission algorithm, the results show that the accuracy and precision of the proposed community division algorithm is feasible and effective.

Key words: complex network; community structure; adaptability; transfer number; community detection algorithm

0 引言

关于复杂网络的研究已成为当今科学界的前沿和热点,如今已遍及学术界的各个领域,如人际关系网、万维网、新陈代谢网等。人们通过大量的研究表明,复杂网络除了有小世界性、无标度性之外,社团结构也越来越被广泛关注。社团结构指的是整个网络由若干个子集组成,各子集之间的连接相对稀疏、而子集内部的连接相对稠密^[1]。社团结构的发现对于人们深入地理解网络的动态性以及结构功能等十分重要。

近年来,复杂网络中社团结构的研究受到了很多学者的关注,许多社团发现算法^[2-3]被相继提出。算法大致分为三大类:启发式方法、基于优化的方法和基于相似性的方法。启发式方法主要是把复杂网络中的社团发现问题转变成预定义的启发式规则的设计问题,如 GN(Girvan-Newman)算法^[4]、极

值优化算法^[5]、Wu-Huberman 算法^[6]等;基于优化的方法是将复杂网络聚类问题转化为优化问题,通过对目标函数的最优化预定义而得到复杂网络的社团结构,如谱聚类算法^[7]、信号传递算法^[8]等;基于相似性的方法是根据网络拓扑结构来定义节点间的相似度,如基于结构全等的相关系数^[9]、节点聚类中心度^[10]等。现如今的社团发现算法如何能较好、较快地划分社团,即所有社团的发现算法中复杂度和查找的准确性依然是最需要解决的两个问题。

基于信号传递的社团发现算法在降低算法的复杂度和提高划分社团结构的准确率方面具有重要的意义,但是对于直径较大的网络,当信号仅传递 3 或 4 次时,边缘节点是无法获取信息的,因而导致社团划分的结果较差。针对上述问题,本文提出一种基于信号自适应传递的社团发现算法,在信号传递的过程中,传递次数不再是静态固定值,而是自适应的动态

收稿日期:2015-01-07;修回日期:2015-04-07。

基金项目:陕西自然科学基金资助项目(2014JZ021);陕西省重点科技创新团队项目(2014KTZ-18);榆林市产学研合作项目(2012cxy3-6)。

作者简介:谭春妮(1989-),女,陕西宝鸡人,硕士研究生,主要研究方向:智能信息处理; 张玉梅(1977-),女,陕西榆林人,副教授,博士,CCF 会员,主要研究方向:非线性时间序列建模及预测; 张嘉桐(1995-),女,黑龙江汤原人,主要研究方向:传感器网络; 吴晓军(1970-),男,陕西凤翔人,教授,博士,主要研究方向:复杂系统。

改变值,从而使所有节点都能够获取信息而又不会出现信息溢出,增强了算法优化的能力。

1 信号传递算法

Hu 等^[8]提出了通过在复杂网络中进行传递信号来发现社团的方法。该方法的思路是将每个节点看作是一个可以受启发的系统,该系统能够发送、接收并记录信号。首先,随机在网络中选择一个节点作为信息源并赋予初始的单位信息,而其他节点没有信息;然后信息源把信息平均传递给相邻节点和自己;随后,获得信息的节点再传给相邻节点和它自己,同时获取由其他节点发送的信息,经过传递3或4次后,大部分节点都获取一定量信息源的信息。因此,信号在节点上的传递可以看作是源节点对于整个网络的影响。对于有 n 个节点的网络,通过这个传递过程可以把网络节点的拓扑结构转化为 n 维几何空间向量的形式。图1说明了在5个节点构成的网络上的信号传输的过程。

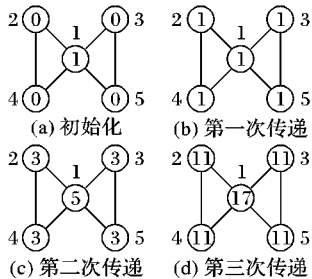


图1 以节点1为源节点进行信号传输的模拟过程

可以将上述的传递过程用公式 $X_i = (I + A)^T$ 来表示,其中 A 是网络的邻接矩阵, I 是单位矩阵, T 是网络中传递信号的次数, X_i 是网络中各节点对源节点的影响向量。

经过上述的信号传递过程,然后把网络的拓扑结构映射为向量,之后选取合适的相似度计算节点间的相似程度,最后运用聚类算法划分社团结构。

为了验证得出的社团结构是否合理或最佳,该算法用 Newman 等^[11]提出的一种模块度的评判标准,模块度的定义如下:

$$Q = \sum_{c=1}^{n_c} \left[\frac{l_c}{m} - \left(\frac{d_c}{2m} \right)^2 \right]$$

其中: m 是网络总边数; n_c 是社团划分的个数; l_c 是社团内部连接边的数目; d_c 是社团内部节点的总度数。 Q 函数的值越趋向于1,则相应的社团结构越好。在实际应用中,一般 Q 的值在0.3 ~ 0.7^[12]。

2 基于信号自适应传递的社团发现算法

2.1 传递次数自适应动态调整

信号传递过程中的传递次数 T 是影响网络空间向量的重要参数,它决定了划分社团结构效果的优越性。基于信号传递的社团发现算法中,传递次数 T 始终是静态固定值,无法动态地调整算法在不同网络中的信息传递性能。若 T 值过大,可能导致信息重复传递以至于其他节点所包含信息量过多而无法与源节点区分;若 T 值过小,边缘节点可能无法获取到源节点的信息。针对上述问题,本文提出自适应动态地调整传递次数的策略,使该算法在传递过程中不再是静态固定值,而是自适应的动态改变值,从而更好地提升算法在不同网络中的信息传递性能。

自适应动态调整传递次数的策略:当信号传递算法中传

递次数为 $T = i (i = 0, 1, \dots, n)$ 时,遍历整个网络,看是否每个节点都有信息量,若还有节点没有信息量,则更新传递次数 T ;否则,停止信号传递,并将此时的传递次数定义为最佳的传递次数 T_{best} 。这样,根据遍历的结果来动态更新本次迭代的传递次数,具有更好的自适应性。因此,实现传递次数的自适应动态调整,算法的寻优精度将会有较大的提高。为验证此结论的合理性,下面通过实验来证明。

用计算机随机生成一个节点 $N = 128$ 的网络,将它分为4个社团,并且每个社团内部均包含32个节点。每个社团内部节点的平均度为 $\langle k_{in} \rangle$,而社团之间的平均度为 $\langle k_{out} \rangle$,每个节点度的期望 $\langle k_{in} \rangle + \langle k_{out} \rangle = 16$ 。随着 $\langle k_{out} \rangle$ 由0越来越大,社团结构越来越模糊,划分社团越来越困难^[1]。图2显示了随着社团之间平均度 $\langle k_{out} \rangle$ 的增大社团划分准确度的变化情况比较,其中传递次数分别为2,4,8, T_{best} ;图中每个值是在同种情况下运行100次取平均值。

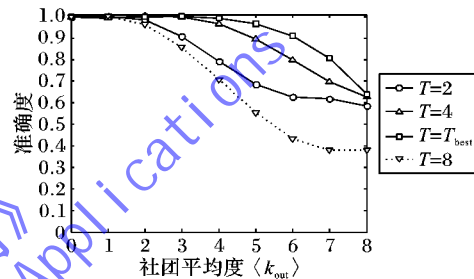


图2 不同传递次数 T 与准确度关系

由图2可知,传递次数取 T_{best} 时获取的社团准确度最高,就是说传递次数取 T_{best} 时是可行且最佳的。

2.2 算法描述

在无向网络 G 中任意选择一个节点作为源节点 S 并赋予一个单位信息,而其他节点没有信息。其中, A 为网络 G 的邻接矩阵; T 为信号传递次数; X_s 表示节点 S 对网络的影响,即节点 S 映射到 n 维空间向量。算法具体步骤如下:

输入 网络的邻接矩阵 A ;

输出 网络的社团结构。

步骤1 初始化 $X_s^0 = (0, 0, \dots, 1, \dots, 0)$, $T = 0$ 。

步骤2 由公式 $X_s = (I + A)^T$, 计算出每个节点对网络的影响向量 $X_s (S = 1, 2, \dots, n)$ 。

步骤3 遍历整个网络,若网络的影响向量 X_s 中有0向量,则更新 X_s , 令 $T = T + 1$, 并返回步骤2;否则,执行步骤4。

步骤4 把 X_1, X_2, \dots, X_n 视为到 n 维单位球面上的点,采用余弦相似度和 Ward 凝聚聚类技术获取网络社团结构。

步骤5 结合社团的定义以及模块度、准确度获取最佳社团结构,并画出此时所获取的社团结构关系图。

上述算法中遍历的过程相当于对整个网络进行一次广度优先搜索,该搜索过程的时间复杂度为 $O(n^2)$,因此在最坏的情况下此算法的时间复杂度为 $O(n^2 + n)$,一般情况下,算法的复杂度远远小于该复杂度。

3 实验与分析

在计算机生成网络和真实网络上分别进行了实验,测试结果证明了本文算法的有效性。

3.1 计算机生成网络

在获取社团结构之后,一般情况下并不知道这个社团结构能否反映网络的本质,并且不知道社团个数是否为最合适。

所以还需要通过计算社团的准确度和精确度来确定社团划分的效果,准确度指运用某种算法获得网络的社团结构和网络实际社团结构的接近程度,而精确度指某种网络运用同样的算法划分社团结构若干次而获得的结果之间的差异性。

为了进一步验证本文算法的有效性和准确性,仍然选择 2.1 节中的计算机生成网络,分别与 GN 算法^[4]、极值优化算法^[5]、谱聚类算法^[7]、信号传递算法^[8]作比较,比较结果如图 3 所示。

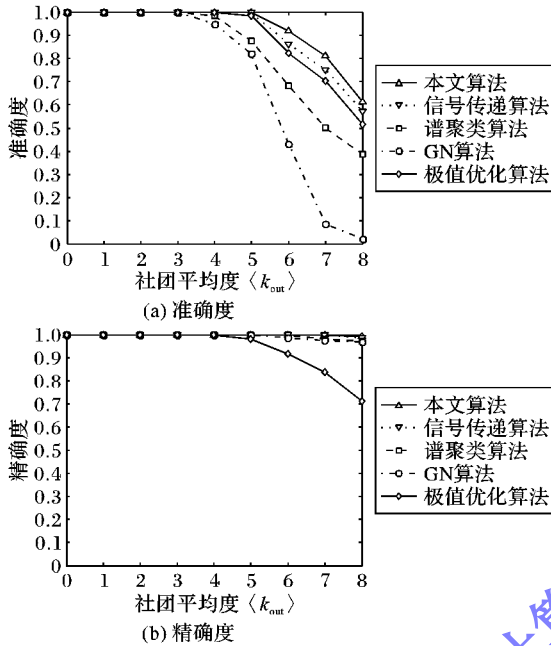


图 3 不同算法准确度和精确度的变化情况

图 3 是随着计算机生成网络中社团之间 $\langle k_{out} \rangle$ 变化准确度和精确度的变化情况。从图 3 可知,本文算法准确度比 GN 算法、谱聚类算法要好,与极值优化算法信号传递算法相近,精确度与谱聚类算法、信号传递算法和 GN 算法类似,但比极值优化算法高很多。由上述可知,该算法具有较高的准确度和精确度。

3.2 真实网络

3.2.1 Zachary 空手道俱乐部网络

“Zachary 空手道俱乐部网络”^[13]是用来判断社团划分效果的典型的实验网络。在 20 世纪 70 年代初,Zachary^[14]花费两年时间观察美国一所大学中空手道俱乐部成员间的人际关系并因此建立一个网络——Zachary 空手道俱乐部网络。该网络共有 34 个节点,每个节点代表俱乐部的成员;所有节点之间共有 78 条边,每条边代表成员之间的人际关系。由于俱乐部管理者与校长之间因为是否要提高俱乐部的收费而产生了分歧,致使俱乐部划分成两部分。图 4 是 Zachary 俱乐部网络的社团划分图,不同颜色的节点代表不同的社团,实验结果表明本文的社团发现算法划分的结果与网络的实际划分结果一致。其中网络中的节点 1 代表了该俱乐部的管理者,节点 33 代表了俱乐部的校长^[14]。

3.2.2 美国大学足球赛网络

2000 年, Girvan 等^[4]发现美国大学生足球赛俱乐部是一个非常合适验证社团算法的有效网络。其所产生的网络具有众所周知的社团结构,该俱乐部共有 115 个球队,按照地理位置分成了 12 个赛区,图 5 是大学生足球赛网络的实际结构图,不同形状

颜色的节点分别代表不同社团,边代表球队之间的比赛。

表 1 是运用本文算法结合网络的实际情况划分 12 个社团得到的结果,其中夹角余弦为节点相似度量,并采用 Ward 凝聚聚类技术。

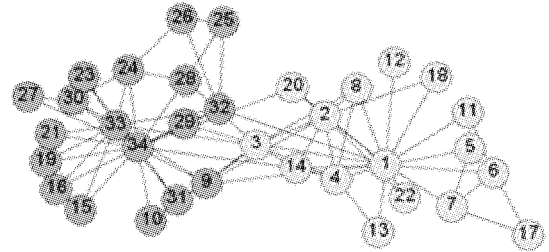


图 4 Zachary 网络的社团划分

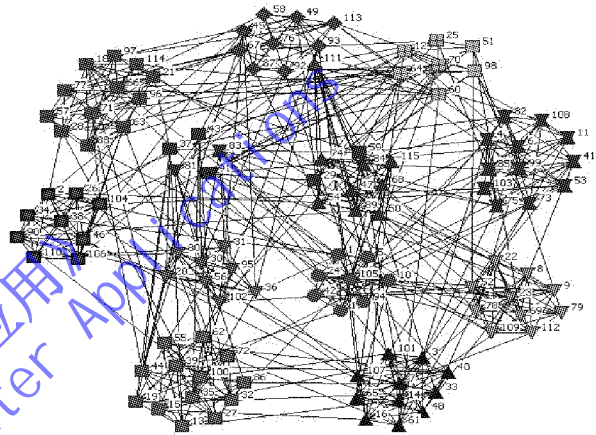


图 5 美国大学生足球赛网络的实际结构

表 1 美国大学生足球赛网络社团划分结果。

社团编号	社团内节点
1	1, 5, 10, 17, 24, 42, 94, 105
2	2, 26, 34, 38, 46, 90, 104, 106, 110
3	3, 7, 14, 16, 33, 40, 48, 61, 65, 101, 107
4	4, 6, 11, 41, 53, 73, 75, 82, 85, 99, 103, 108
5	8, 9, 22, 23, 52, 69, 78, 79, 109, 112
6	12, 25, 29, 51, 70, 91
7	13, 15, 19, 27, 32, 35, 37, 39, 43, 44, 55, 72, 86, 100
8	18, 21, 28, 57, 63, 66, 71, 77, 88, 96, 97, 114
9	20, 30, 31, 36, 56, 80, 81, 83, 95, 102
10	47, 50, 54, 68, 74, 84, 89, 111
11	59, 60, 64, 98, 115
12	45, 49, 58, 67, 76, 87, 92, 93, 113

由表 1 可知,与网络实际情况对比运用本文算法节点正确划分的个数为 106,即准确度为 92.17%,优于 Hu 等^[8]的信号传递算法的准确度 90.43%,更优于 GN 算法的 67.13%。此刻网络模块度 Q 为 0.587,优于信号传递算法的 0.536,以及谱聚类算法的 0.579。

4 结语

本文基于自适应的思想对现有信号传递算法进行了改进,提出一种有效社团发现的算法。经实例验证,当信号传递次数根据网络的规模而自适应动态地调整时,获取的空间

(下转第 1594 页)

- 东, 韩建军, 王天江. 一种基于 VFD 多核系统的硬实时任务节能调度算法[J]. 计算机研究与发展, 2012, 49(5): 1018-1027.)
- [2] KAVOUSIANOS X, CHAKRABARTY K, JAIN A, *et al.* Test scheduling for multicore SoCs with dynamic voltage scaling and multiple voltage islands [C]// Proceedings of 2011 20th Asian Test Symposium. Piscataway: IEEE, 2011: 33-39.
 - [3] TOKARNIA A M, PEPE P C F, PAGOTTO L D. Path-based dynamic voltage and frequency scaling algorithms for multiprocessor embedded applications with soft delay deadlines [C]// Proceedings of 2011 14th Euromicro Conference on Digital System Design. Piscataway: IEEE, 2011: 109-116.
 - [4] ZHANG D, WU F, CHEN F, *et al.* An overhead-aware optimal energy-efficient real-time scheduling algorithm on multiprocessors [J]. Chinese Journal of Computers, 2012, 35(6): 1297-1312. (张冬松, 吴飞, 陈芳园, 等. 开销敏感的多处理器最优节能实时调度算法[J]. 计算机学报, 2012, 35(6): 1297-1312.)
 - [5] ZHU X, HE C, WANG J, *et al.* An elastic energy-aware scheduling strategy for heterogeneous computing systems [J]. Chinese Journal of Computers, 2012, 35(6): 1313-1326. (朱晓敏, 贺川, 王建江, 等. 异构计算系统中弹性节能调度策略研究[J]. 计算机学报, 2012, 35(6): 1313-1326.)
 - [6] KIM J K, SIEGEL H J, MACIEJEWSKI A A, *et al.* Dynamic resource management in energy constrained heterogeneous computing systems using voltage scaling [J]. IEEE Transactions on Parallel and Distributed Systems, 2008, 19(11): 1445-1457.
 - [7] LEE W Y. Energy-efficient scheduling of periodic real-time tasks on lightly loaded multicore processors [J]. IEEE Transactions on Parallel and Distributed Systems, 2012, 23(3): 530-537.
 - [8] ZHU D, AYDIN H. Reliability-aware energy management for periodic real-time tasks [J]. IEEE Transactions on Computers, 2009, 58(10): 1382-1397.
 - [9] ZHU D, MELHEM R, MOSSE D. The effects of energy management on reliability in real-time embedded systems [C]// Proceedings of IEEE/ACM International Conference on Computer Aided Design. Piscataway: IEEE, 2004: 35-40.
 - [10] HAQUE M A, AYDIN H, ZHU D. Energy-aware task replication to manage reliability for periodic real-time applications on multicore platforms [C]// Proceedings of the 2013 International Green Computing Conference. Piscataway: IEEE, 2013: 1-11.
 - [11] LIN M, PAN Y, YANG L T, *et al.* Scheduling co-design for reliability and energy in cyber-physical systems [J]. IEEE Transactions on Emerging Topics in Computing, 2013, 1(2): 353-365.
 - [12] LI Z, WANG L, LI S, *et al.* Reliability guaranteed energy-aware frame-based task set execution strategy for hard real-time systems [J]. The Journal of Systems and Software, 2013, 86(12): 3060-3070.
 - [13] LI J, SHU L C, CHEN J, *et al.* Energy-efficient scheduling in nonpreemptive systems with real-time constraints [J]. IEEE Transactions on Systems, 2013, 43(2): 332-344.
 - [14] ZHANG Y, CHAKRABARTY K. Energy-aware adaptive checkpointing in embedded real-time systems [C]// DATE'03: Proceedings of the Design, Automation and Test in Europe Conference and Exhibition. Piscataway: IEEE, 2003: 918-923.

(上接第 1554 页)

向量最佳,因为此时检测网络的社团效果比传统方法更为准确。本文算法克服了原有信号传递算法仅对小规模网络正确划分的缺陷,该算法对较大网络的划分同样具有较高准确度。但本文的研究仅限于不重叠的无权网络,对于加权重叠网络的研究将是下一步的研究方向。此外,信息孤岛问题也将是以后需要考虑的问题。

参考文献:

- [1] WANG X, LI X, CHEN G. Complex network theory and its applications [M]. Beijing: Tsinghua University Press, 2006: 162-165. (汪小帆, 李翔, 陈关荣. 复杂网络理论及其应用[M]. 北京: 清华大学出版社, 2006: 162-165.)
- [2] SHI C, CAI Y N, FU D, *et al.* A link clustering based overlapping community detection algorithm [J]. Data & Knowledge Engineering, 2013, 87: 394-404.
- [3] LI K, PANG Y. A unified community detection algorithm in complex network [J]. Neurocomputing, 2014, 130: 36-43.
- [4] GIRVAN M, NEWMAN M E J. Community structure in social and biological networks [J]. Proceedings of National Academy of Sciences, 2001, 99(12): 7821-7826.
- [5] DUCH J, ARENAS. A community detection in complex networks using extremal optimization [J]. Physical Review E, 2005, 72(2): 027104.
- [6] WU F, HUBERMAN B A. Finding communities in linear time: a physics approach [J]. The European Physical Journal B, 2004, 38(2): 331-338.
- [7] DONETTI L, MUNOZ M A. Detecting network communities: a new systematic and efficient algorithm [J]. Journal of Statistical Mechanics: Theory and Experiment, 2004, 37(10): 10012.
- [8] HU Y, LI M, ZHANG P, *et al.* Community detection by signaling on complex networks [J]. Physical Review E: Statistical, Nonlinear and Soft Matter Physics, 2008, 78: 016115.
- [9] WASSERMAN S, FAUST K. Social network analysis: methods and applications [M]. Cambridge: Cambridge University Press, 1994: 249-290.
- [10] YANG B, LIU J. Discovering global network communities based on local centralities [J]. ACM Transactions on the Web, 2008, 2(1): article 9.
- [11] NEWMAN M E J, GIRAN M. Finding and evaluating community structure in networks [J]. Physical Review E, 2004, 69(2): 293-313.
- [12] ZHANG C, SHEN H. Modularity function for community structure based on natural density of networks [J]. Journal of University of Electronic Science and Technology of China, 2012, 41(2): 185-191. (张聪, 沈惠璋. 网络自然密度社团结构模块度函数[J]. 电子科技大学学报, 2012, 41(2): 185-191.)
- [13] ZACHARY W W. An information flow model for conflict and fission in small groups [J]. Journal of Anthropology Research, 1977, 33(4): 452-473.
- [14] XIE Z, WANG X. Review of the algorithm analysis research for community structure in complex networks [C]. Complex Systems and Complexity Science, 2005, 2(3): 1-12. (解岱, 汪小帆. 复杂网络中的社团结构分析算法研究综述[J]. 复杂系统与复杂性科学, 2005, 2(3): 1-12.)