

基于随机匹配的复杂网络最小驱动点集分析

张锡哲¹, 吕天阳², 栾 晗¹, 杨雪莹¹, 张 斌¹

(1. 东北大学 信息科学与工程学院, 沈阳 110819; 2. 审计署 计算机技术中心, 北京 100830)

摘 要: 控制复杂网络在很多领域都有着重要的应用价值. 将控制复杂网络所需的最少节点集合称为最小驱动点集. 针对网络的最小驱动点集并不唯一, 提出一种随机匹配方法来获取网络中不同的最小驱动点集, 并分析最小驱动点集集合的平均度分布以及节点在最小驱动点集集合中的出现频率. 研究发现, 多数网络的最小驱动点集分布紧密, 其节点构成与网络度分布有关; 同时, 网络中节点的控制重要性与其入度密切相关. 所得到的相关结论对于复杂网络的控制具有重要的研究意义.

关键词: 复杂网络; 结构可控性; 最小驱动点集; 拓扑分析

中图分类号: TP273

文献标志码: A

Analysis of minimum driver node set of complex network based on random matching

ZHANG Xi-zhe¹, LÜ Tian-yang², LUAN Han¹, YANG Xue-ying¹, ZHANG Bin¹

(1. College of Information Science and Technology, Northeastern University, Shenyang 110004, China; 2. Center of Computer Technology, Audit Administration, Beijing 100830, China. Correspondent: ZHANG Xi-zhe, E-mail: zhangxizhe@ise.neu.edu.cn)

Abstract: Controllability of complex networks has important application value in many areas. The minimum driver node set is defined as the minimum nodes required to the control complex network. However, the minimum driver node set of most of network is not unique. Therefore, a random matching method is proposed to obtain different minimum driver node set of a network. Then, the method analyzes the average degree distribution of the collection of minimum driver node sets, and the frequency appears in the collection of the node. It is found that the minimum driver node sets of most networks are distributed tightly. The composition of a driver node set is closely related with the degree distribution of network. The control importance of the node in a network is closely related to the in-degree of node. These conclusions have important research meaning for the controllability of the complex network.

Keywords: complex network; structural controllability; minimum driver node set; topological analysis

0 引 言

现实生活中, 许多事物及其之间的关系都可以建模为网络, 对这种日益复杂的网络系统的结构及行为分析是理解其内在规律的前提. 已有很多工作讨论了复杂网络的结构特征^[1-3]、建模规律^[4-6]和动态行为^[7-9], 但对于如何控制复杂网络的动态行为, 仍然知之甚少. 近期Liu等^[10]提出了一种分析复杂网络可控性的方法, 通过计算最大匹配给出网络的最小驱动点集(MDS); 在其基础上, Nepusz等^[11]给出了针对网络边动力学的边控制方法; Yan等^[12]探索了网络控制能

量的上下界; Wang等^[13]研究了如何通过增边减少网络的驱动节点集个数使其更容易被控制.

因为网络的最大匹配一般不是唯一的^[14], 所以复杂网络的MDS也不是唯一的. 显然, 对于不同的MDS, 其拓扑性质如平均度、节点度分布也是不同的, 因此有必要深入考察系统所有可能MDS的拓扑特征, 并分析其对系统可控性的影响. 然而, 计算网络所有的MDS是#P完全问题^[14], 这使得对于大规模复杂网络而言, 精确获得所有MDS的计算复杂度极高. 因此, 为了分析网络所有MDS的拓扑特征, 本文设计

收稿日期: 2014-02-10; 修回日期: 2014-07-10.

基金项目: 中央高校基本科研业务费项目(N120404011); 国家自然科学基金项目(60093009, 61073062, 71272216, 61100027); 国家科技支撑计划项目(2012BAH08B02).

作者简介: 张锡哲(1978—), 男, 副教授, 博士, 从事复杂网络的研究; 张斌(1964—), 男, 教授, 博士生导师, 从事服务计算等研究.

了一种MDS随机抽样方法,在匹配过程中随机选择未被匹配的节点,从而得到网络不同的控制配置.

基于所提出的算法,本文分析了若干实际网络和模型网络的MDS的结构特征.首先,给出了网络MDS的平均度分布,发现无论网络的基本拓扑如何,其MDS的平均度均呈现正态分布;然后,分析了MDS集合中的驱动节点构成及节点在MDS集合中的出现频度,探索了高频出现的核心驱动节点的拓扑特征.

1 基于随机抽样的最小驱动节点集获取方法

对于有向图,匹配边集是指没有共同起点和共同终点的边的集合.对于一个节点,若匹配边的终点指向此节点,则称其为匹配节点,否则称其为非匹配节点.含有最多边数的匹配边集称为最大匹配边集.

文献[10]提出了最小输入原理,将网络 G 记为 $G(V, E)$, V 和 E 分别为网络的节点集合和边集合,并证明了若网络 $G(A)$ 中存在一个完备匹配,则要控制网络 $G(V, E)$ 所需的最小驱动节点数 n_D 为1;否则,等于网络中最大匹配对应的非匹配节点的数量.驱动节点即为非匹配节点.

计算网络的最大匹配集常用的匹配算法是Hopcroft-Karp算法^[15].其核心就是寻找可增广路,从当前匹配 M (如果没有匹配,则取初始匹配为 $M = \phi$)出发,检查每一个未盖点,然后从它出发寻找可增广路.若找到可增广路,则沿着这条可增广路进行扩充,直到不存在可增广路为止.

注意到当求解网络最大匹配时,一般而言,有两个因素影响最终的匹配结果:一个是节点的检查顺序,算法依次检查每一个未覆盖节点,所以先检查的节点有很大可能作为匹配节点,不同的匹配次序会导

致不同的匹配结果;另一个是节点寻找增广路的顺序,选择哪一个可增广路决定了最终的匹配结果.为了分析网络驱动节点集的性质,本文设计了一个随机化匹配算法.主要思路是:在寻找增广路时,从未匹配点集合中随机抽取一个,然后从它出发寻找可增广路;对于未匹配点的多个增广路,随机选取一个进行扩充,直到不存在可增广路为止,如图1所示.算法的具体步骤如下.

Step 1: 令网络 $G(V, E)$ 的初始匹配 M 为空.

Step 2: 根据当前匹配 M ,计算未匹配节点集 V_{um} ,从 V_{um} 中随机抽取节点 n ,寻找从 n 开始的所有增广路集合 P .

Step 3: 从 P 中随机找到一个增广路 p ,并扩展该增广路,得到新匹配 M' ,令 $M = M'$.

Step 4: 重复Step 2和Step 3,直到找不到增广路为止. M 即为网络最大匹配,未匹配节点为最小驱动节点集.

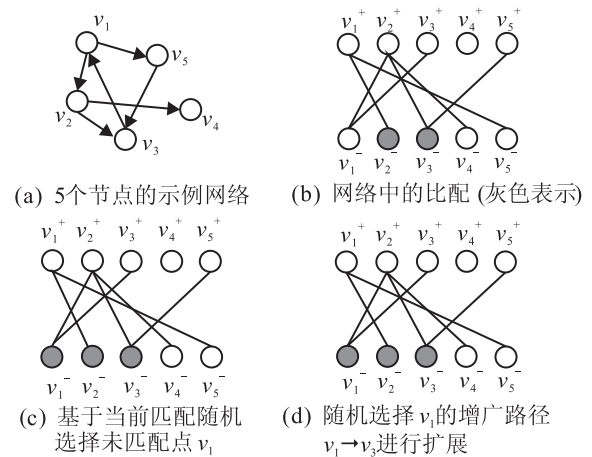


图1 随机匹配算法

2 实验及分析

为了说明问题的普遍性,本文选择了不同领域

表1 网络最小驱动节点集的统计结果

网络分类	网络名称	N	L	Γ_D	$\langle k \rangle$	$[(k_{D_{\min}}^r), (k_{D_{\max}}^r)]$
WWW	Polblogs	1 224	16 718	0.34	27.32	[10.60, 13.84]
Social- Communication	UCIonline	1 899	20 296	0.32	21.38	[5.86, 7.83]
	P2P-1	10 876	39 994	0.55	7.36	[6.87, 6.97]
Internet	P2P-2	8 846	31 839	0.58	7.20	[6.817, 6.91]
	P2P-3	8 717	31 525	0.58	7.23	[6.76, 6.86]
Power Grid	USpowerGrid	4 941	13 188	0.12	5.34	[2.63, 2.82]
Trust	Wiki-Vote	7 115	103 689	0.67	28.98	[9.66, 9.66]
Other	Florida	128	2 106	0.23	32.91	[20.77, 29.57]
Neuronal	C.elegans	306	2 345	0.19	15.79	[4.69, 6.97]
	Grassland	88	137	0.52	3.11	[2.65, 2.70]
Food Web	Littlerock	183	2 494	0.54	27.26	[1, 1]
	Seagrass	49	226	0.27	9.22	[7.08, 9.38]
	Ythan	135	601	0.51	8.90	[6.826, 7.928]
Regulatory	TRN-Yeast-1	4 441	12 873	0.96	5.80	[5.846, 5.847]

的14个真实数据集^[10]进行实验分析, 涉及信任网络、食物网络、电力网络、神经网络、引用网络、万维网、因特网、社会交际网络和社会组织网络等多个领域. 表1给出了采用随机化匹配算法对网络进行10000次的实验结果. 其中: N 表示网络的节点数, L 表示网络边数, $\langle k \rangle$ 表示网络平均度, λ_D 表示驱动节点比例, $\langle k_{D_{min}} \rangle$ 表示求出的所有MDS中平均度最小值, $\langle k_{D_{max}} \rangle$ 表示平均度最大值.

一个明显的结果是网络的MDS具有丰富的节点构成. 从驱动节点集平均度变化范围可以看出, 多数网络的MDS具有多种可能的选择; 与整个网络的平均度相比, 有些网络如TRN-Yeast-1和Seagrass的 $\langle k_{D_{max}} \rangle$ 甚至大于网络的平均度. 对于ER网络, 多数驱动节点集的平均度均显著大于网络的平均度. 这表明尽管网络中多数的驱动节点集可能主要是由低度节点构成的^[10], 但网络中仍可找到由中度和高度节点构成的最小驱动节点集, 这种驱动节点集的多样性对于网络控制具有重要的意义.

为了进一步说明MDS的拓扑特征, 本文分析了这些最小驱动点集的平均度分布. 一个有趣的结果是, 几乎所有网络的驱动节点集平均度均呈现近似正态分布, 如图2所示. 这意味着在所有可能的驱动节点所构成的高维空间中, MDS的分布极为“紧密”.

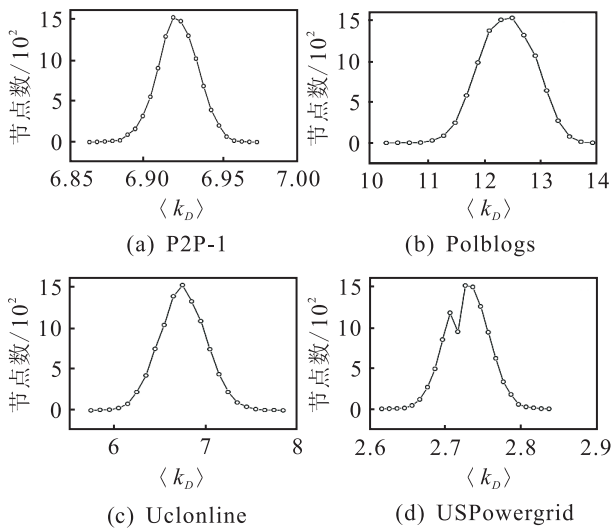


图2 网络最小驱动点集的平均度分布

为了考察节点在网络控制中的重要性, 本文分析了节点在MDS中出现频次, 如图3所示. 由图3可见, 对于BA网络和典型的实际网络, 低度节点在驱动节点集中出现频度较多, 这与文献[10]的结论基本一致. 但是, 低度节点在驱动节点集中的出现频度没有明显差异, 如P2P-1和PolBlog等数据集, 10%左右的低度节点基本覆盖了驱动节点集集合, 存在一个“稠密”区域, 在这个区域中, 网络的控制性相当灵活, 可

选择多种节点组合进行控制.

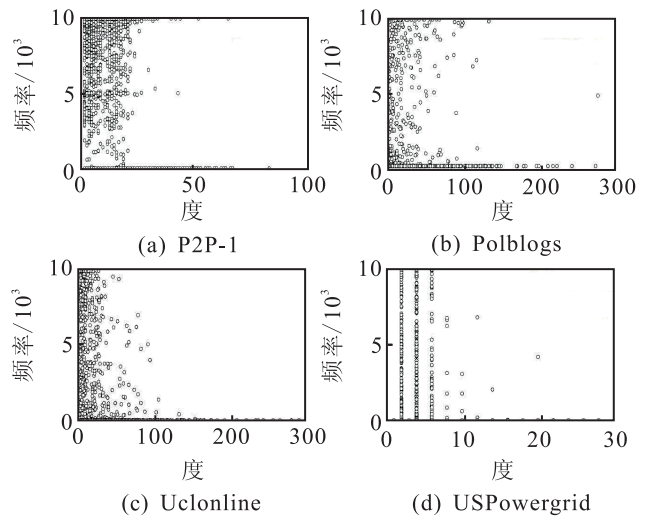


图3 网络各节点在最小驱动点集集中出现的次数

对于ER随机网络, 节点的出现频次不存在任何的倾向性, 基本呈现与度分布相似的正态分布. 高度节点和低度节点作为驱动节点的机会基本均等, 这说明在ER网络中, 既可找到主要由低度和中度节点构成的MDS, 也可能会找到主要由中度和高度节点构成的MDS, 这与文献[10]的结论明显不同.

由图3中的结果可以看出, 节点出现频次的分布与网络的节点度分布非常类似. 解释上述结果的一个可能的原因是, 节点在MDS的出现频次是由度分布决定的. BA网络 and 实际网络基本呈现幂律分布, 网络中低度节点占多数, 所以MDS主要由低度节点构成; 而ER随机网络中节点的度分布基本均衡, 低度节点和高度节点在网络中不存在多数优势, 所以MDS中的驱动节点构成也不存在任何倾向性.

为了验证这个结论, 本文统计了核心驱动节点集的平均度, 如图4所示.

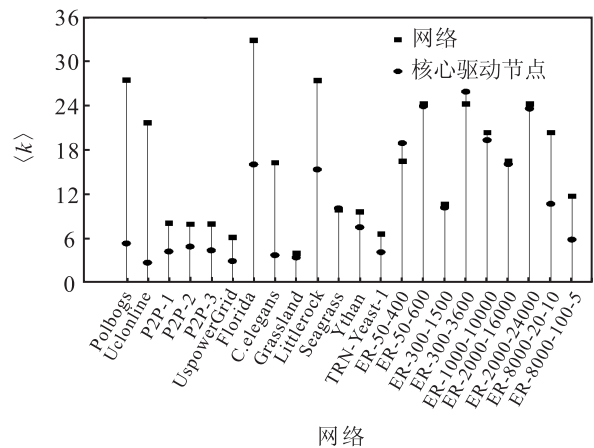


图4 核心驱动节点的平均度与网络平均度

由图4可以看出, 低度节点占优势的网络如BA、P2P、Polblog和Powergrid等, 其核心驱动节点平均度明显低于网络的平均度, 而节点度分布较均衡的网络

如 ER、Seagrass 和 Grassland 等,其核心驱动节点的平均度与网络平均度基本相等.这也说明了网络最小驱动节点集的构成与整个网络的节点构成具有密不可分的关系.

3 结 论

复杂网络的控制是具有普遍意义的问题.为了控制网络达到期望的状态,需要从部分节点输入控制信号,这些节点称为最小驱动点集.对于大多数网络,最小驱动点集不是唯一的.本文通过随机匹配的方法,分析了网络中多个最小驱动点集的拓扑性质.

本文发现的一个重要现象是多数网络都存在大量的最小驱动点集,并且这些最小驱动点集的节点构成极为相似.因此,当最小驱动点集中高频出现的核心驱动节点受到攻击或发生故障无法输入信号时,将会影响网络中多数的最小驱动点集,从而显著地降低网络的可控性.综上所述,是否能够有效地保护这些核心驱动节点是保证网络有效控制的关键.

另一方面,探索最小驱动节点集的节点构成对于网络控制也具有重要意义.本文发现,节点在最小驱动节点集中出现的频率与网络度分布具有紧密的关联,对于低度节点占多数的网络,高频出现的驱动节点一般倾向于低度节点,而在度分布较均衡的网络如 ER 随机网,高频出现的驱动节点便没有这种倾向性,这表明 MDS 的节点构成与网络的节点构成具有明显的关联.

本文的研究不仅为复杂网络控制的研究提供了新的研究视角,也提出了若干新的问题.将来的研究方向包括分析网络所有最小驱动点集的拓扑特征,寻找满足控制约束的特定的最小驱动节点集,以及研究哪种最小驱动点集具有较好的控制鲁棒性等.

参考文献(References)

- [1] Fortunato S. Community detection in graphs[J]. *Physics Reports*, 2010, 486(3): 75-174.
- [2] Ghoshal G, Barabasi A L. Ranking stability and super stable nodes in complex networks[J]. *Nature Communications*, 2011, 2(7): 394.
- [3] Karsai M, Kivela M, Pan R K, et al. Small but slow world: How network topology and burstiness slow down spreading[J]. *Physical Review E*, 2011, 83(2): 025102.
- [4] Papadopoulos F, Kitsak M, Serrano Má, et al. Popularity versus similarity in growing networks[J]. *Nature*, 2012, 489(7417): 537-540.
- [5] Barabási A L, Albert R. Emergence of scaling in random networks[J]. *Science*, 1999, 286(5439): 509-512.
- [6] Watts D J. Networks, dynamics, and the small-world phenomenon[J]. *American J of Sociology*, 1999, 105(2): 493-527.
- [7] Vespignani A. Modelling dynamical processes in complex socio-technical systems[J]. *Nature Physics*, 2012, 8(2160): 32-39.
- [8] Stanoev A, Smilkov D, Kocarev L. Identifying communities by influence dynamics in social networks[J]. *Physical Review E*, 2011, 84(4): 046102.
- [9] Palla G, Barabasi A L, Vicsek T. Quantifying social group evolution[J]. *Nature*, 2007, 446(7136): 664-667.
- [10] Liu Y Y, Slotine J J, Barabasi A L. Controllability of complex networks[J]. *Nature*, 2011, 473(7346): 167-173.
- [11] Nepusz T, Vicsek T. Controlling edge dynamics in complex networks[J]. *Nature Physics*, 2012, 8(7): 568-573.
- [12] Yan G, Ren J, Lai Y C, et al. Controlling complex networks: How much energy is needed?[J]. *Physical Review Letters*, 2012, 108(21): 218703.
- [13] Wang W X, Ni X, Lai Y C. Optimizing controllability of complex networks by minimum structural perturbations[J]. *Physical Review E*, 2012, 85(2): 026115.
- [14] Valiant L G. The complexity of computing the permanent[J]. *Theoretical Computer Science*, 1979, 8(2): 189-201.
- [15] Hopcroft J E, Karp R M. An $n^5/2$ algorithm for maximum matchings in bipartite[J]. *SIAM J on Computing*, 1973, 2(4): 225-231.

(责任编辑:滕蓉)