

文章编号: 2095-2163(2020)01-0197-03

中图分类号: TP301.6

文献标志码: A

动态网络社区发现综述

高 阳, 张宏莉

(哈尔滨工业大学 计算机科学与技术学院, 哈尔滨 150001)

摘要: 社区发现问题是复杂网络分析中最重要的研究内容之一。给定一个复杂网络, 一个社区由网络中一定数量的节点构成, 这些节点之间的边相对较多, 这些节点与社区外部节点之间的边相对较少。目前已经提出了大量的静态网络社区发现算法。然而, 真实网络总是动态的, 网络的拓扑结构随着新节点、边的加入或现有节点和边的消失而不断变化。这些网络中的社区也因此随时间不断变化。本篇综述介绍了动态网络和动态社区的定义, 给出了动态网络社区发现算法的分类。本篇综述可以帮助不熟悉这个研究领域的科研人员了解动态网络的社区发现, 也可以帮助科研人员在有动态社区发现需求时选择最好的算法分类。

关键词: 动态网络; 社区发现; 社区变化

Community detection in dynamic networks: A survey

GAO Yang, ZHANG Hongli

(School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China)

[Abstract] The problem of community detection is one of the most important topics in the field of complex network analysis. Given a complex network, a community consists of a number of vertices that have relatively more links among them and fewer connections between them and rest of the network. A large amount of algorithms have been designed for community discovery in static networks. However, real-world networks are always dynamic, the topology structure of the networks are changing with the appearance of new vertices and edges or vanishing of existing vertices and edges. Communities in the networks evolves with time as a result. This survey introduces the definition of dynamic networks and dynamic communities, gives classification of algorithms for community detection in dynamic networks. The survey could help researchers that are unfamiliar with the area to know community detection in dynamic networks. The classification could also help researchers to find the best kind of algorithm for their needs in dynamic community detection.

[Key words] dynamic networks; community detection; community events

0 引言

真实世界网络的拓扑结构总是在不断变化的, 网络随时可能会增加、删除节点或者边, 网络的社区结构也会随之变化, 于是就产生了动态网络的社区发现问题。动态网络可以采用 2 种模型来描述^[1]:

(1) TN (Temporal Networks)^[1-4]。动态网络可表示为图 $G = (V, E, T)$, 其中 V 表示动态网络中节点的集合, V 中每个元素 (v, t_s, t_e) ($t_s \leq t_e$) 包含 3 个基本属性, v 表示网络中一个节点, $t_s, t_e \in T$ 分别表示该节点在网络中出现和消亡时间点; E 表示动态网络中边的集合, E 中每个元素 (u, v, t_s, t_e) ($t_s \leq t_e$) 包含 4 个基本属性, $u, v \in V$ 表示该边的 2 个端点, $t_s, t_e \in T$ 分别表示该边在网络中出现和消亡的时间点。

(2) SN (Network Snapshots)。动态网络表示为一个向量 $\langle G_1, G_2, \dots, G_t \rangle$, 包含动态网络在每个时

间点的拓扑结构, 其中 $G_i = (V_i, E_i)$, V_i, E_i 分别表示该时间点对应的节点和边的集合。

1 动态网络社区发现概述

给定一个动态网络, 一个动态社区 DC (Dynamic Community) 表示一些具有时间属性 (period) 节点的集合, $DC = \{(v_1, P_1), (v_2, P_2), \dots, (v_n, P_n)\}$, 其中 $P_i = ((t_{s0}, t_{e0}), (t_{s1}, t_{e1}), \dots, (t_{sN}, t_{eN}))$ ($t_s \leq t_{e_j}$) 表示节点 v_i 的 N 个存在时期。动态网络社区发现是指发现动态网络中所有的动态社区^[1]。动态网络的社区发现主要有 2 个研究目标, 对此可做阐释分述如下。

(1) 快速准确地发现动态网络在每个时间点的社区结构。

(2) 建立演化链来描述社区的生命周期及演化过程^[1]。

动态网络中的社区可以发生以下变化:

作者简介: 高 阳(1986-), 男, 博士研究生, 主要研究方向: 社区发现; 张宏莉(1973-), 女, 博士, 教授, 博士生导师, 主要研究方向: 并行计算、网络与信息安全、网络测量等。

收稿日期: 2019-03-26

(1) 出生。指一个社区第一次在网络中出现。

(2) 消亡。指一个社区从网络中消失,该社区内全部节点不再隶属于该社区。

(3) 变大。一个社区的规模由于加入了新节点而增长。

(4) 收缩。一些节点移出某个社区,该社区规模因此变小。

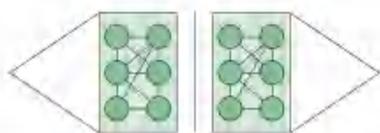
(5) 合并。2个或多个社区合并成一个社区。

(6) 分裂,由于节点或边的消失,一个社区分裂成多个部分。

(7) 不变。一个社区没有变化。

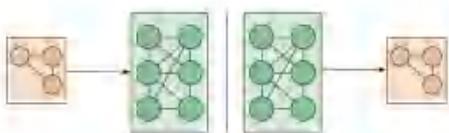
(8) 复活。一个社区消失一段时间后,再次没有任何变化地在网络中出现^[1,5-6]。

综上变化如图1所示^[1]。



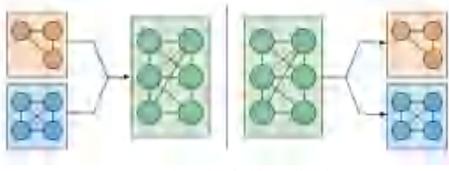
(a) 社区的出生和消亡

(a) Birth and death of a community



(b) 社区的变大和收缩

(b) Growth and contraction



(c) 社区的合并和分裂

(c) Merge and split



(d) 社区保持不变

(d) Continue



(e) 社区复活

(e) Resurgence

$t \longrightarrow t+1 \dashrightarrow t+n-1 \longrightarrow t+n$

图1 社区变化^[1]

Fig. 1 Community events^[1]

2 动态网络社区发现研究

目前获取动态网络社区结构的方法大致可分为3类,分别是:静态方法、基于演化的方法和基于增量的方法。这里拟对此展开研究论述如下。

2.1 静态方法

该类方法中,首先独立地对动态网络在每一个时间点上的网络进行社区发现,一般采用某种静态网络的社区发现算法来求解;为了发现社区的变化过程,该类方法一般会对相邻两个时间点的社区进行匹配。

Palla 等人^[6]将 CPM 算法应用到了动态网络中。算法通过采用 CPM 算法对每个时间点的网络独立地进行社区发现,再对相邻时间点的社区进行匹配,该算法把2个相邻时间点的网络组成一个网络,并在这个组合网络上用 CPM 算法进行社区发现,用组合网络上的社区帮助匹配相邻时间点的社区。Doyle 等人^[7]选择非重叠社区发现算法 Louvain 作为每个时间点网络的社区发现方法,采用二进制集合的 Jaccard 系数^[8]来计算社区之间的相似度,从而进行相邻时间点网络社区之间的匹配并分析社区的演化过程。

这类方法的主要优势是可采用现有的静态社区发现算法进行每个时间点的社区发现,具有很大的选择范围,对相邻时间点社区的匹配过程也有很多集合匹配算法作为基础;而且,不同时间点的社区发现相对独立,很容易实现算法的并行化^[1]。

2.2 基于演化的方法

Chakrabarti 等人^[9]提出了基于演化的动态网络社区发现方法(Evolutionary Clustering)。该类方法在进行社区发现时需要优化2个方面:首先,在每个时间点的网络社区结构需要与当前网络的拓扑结构尽可能地吻合;与此同时,相邻2个时间点的网络社区结构不能发生太大的变化。Chakrabarti 等人提出了2种分别基于 k-means 方法和凝聚式层次聚类的动态网络社区发现方法。Lin 等人^[10]基于同样的方法提出了 FacetNet, FacetNet 采用非负矩阵分解方法用社区演化的平滑度来帮助发现社区结构。

这类方法可以平衡某一时间点网络社区结构的质量和与前一时间点网络社区结构之间的关联,社区演化的平滑性得到了保证,但是不同时间点网络的社区结构一般需要单独计算,耗时较高,难以完成大规模动态网络的社区发现。

2.3 基于增量的方法

对于给定动态网络 $G = \{G_0, G_1, \dots\}$, 基于增量

的动态网络社区发现方法首先采用某种静态社区发现算法对第一个时间点的网络 G_0 进行完整的社区划分,根据网络拓扑结构的变化对社区进行不断更新,得到后续时间点的网络社区结构。这种方法只完成一次整体的社区划分,社区的更新一般只发生在网络结构发生变化的位置附近,因而这种方法极大地提高了动态网络社区发现的效率,同时也可保证社区演化的平滑。

Ning 等人^[11]提出了一种基于谱聚类的动态网络社区发现算法,算法将动态网络的变化分为节点之间相似度的变化和增删节点,并用关联向量和关联矩阵表示 2 种动态变化,通过增量谱聚类算法得到动态网络的社区结构。Cazabet 等人^[12]提出 iLCD 算法。iLCD 算法根据网络拓扑结构的变化(网络不断增加边和节点),对前一个时间点的社区进行更新、合并、并生成新社区来发现当前时间点的社区结构。该算法可以发现动态网络中的重叠社区结构,但是没有考虑到当网络中有节点或者边删除时社区的变化情况。

Nguyen 等人^[13]提出了基于模块化的 QCA 算法,将网络拓扑结构的变化分为增删节点和增删边,对每种网络结构的变化分别建立了相应的社区结构更新方法。该算法可快速地根据网络的变化对社区结构进行更新,但是不能发现重叠社区结构。在此基础上,Nguyen 等人^[14]又提出了 AFOCS 算法,同样将网络拓扑结构的变化分为增删节点和增删边等 4 种情况,但是允许一个节点属于多个社区,可以发现动态网络的重叠社区结构。Xin 等人^[15]提出了一种基于随机游走的静态网络重叠社区发现算法 RWS,并提出 ARWS 算法来实现对动态网络社区的更新,ARWS 算法同样将动态网络的变化分为以上四种情况,将网络中的增删节点以及其相邻节点和网络中增删边所连接的节点加入到队列中,用 RWS 算法对队列中节点进行计算,从而完成社区的更新。ARWS 算法同样具有很高的社区发现效率,但是算法具有过多的参数,不同参数值对社区发现结果影响较大。

3 结束语

针对目前存在的大量动态网络社区发现算法,本文介绍了一些文献对动态网络和动态社区的定义,分析了动态网络中社区发生的主要变化,给出了动态网络社区发现算法的分类。本文可以帮助不熟

悉这个研究领域的科研人员了解动态网络的社区发现;也可以帮助科研人员在有动态社区发现需求时选择最好的算法分类;同时也可以辅助做这方面研究的学者选择今后的研究方向。

参考文献

- [1] ROSSETTI G, CAZABET R. Community discovery in dynamic networks: a survey [J]. ACM Computing Surveys (CSUR), 2017, 51(2): 35.
- [2] HOLME P, SARAMÄKI J. Temporal networks [J]. Physics Reports, 2012, 519(3): 97-125.
- [3] ROSSETTI G, GUIDOTTI R, PENNACCHIOLI D, et al. Interaction prediction in dynamic networks exploiting community discovery [C]// Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) (ACM). Paris, France: ACM, 2015: 553-558.
- [4] VIARD T, LATAPY M, MAGNIEN C. Computing maximal cliques in link streams [J]. Theoretical Computer Science, 2016, 609: 245-252.
- [5] CAZABET R, AMBLARD F. Dynamic community detection [M]. New York: Springer, 2014.
- [6] PALLA G, BARABÁSI A L, VICSEK T. Quantifying social group evolution [J]. Nature, 2007, 446(7136): 664-667.
- [7] DOYLE D, CUNNINGHAM P, GREENE D. Tracking the evolution of communities in dynamic social networks [C]// International Conference on Advances in Social Networks Analysis and Mining (ASONAM). Odense: IEEE, 2010: 176-183.
- [8] JACCARD P. The distribution of flora in the alpine zone [J]. The New Phytologist, 1912, 11(2): 37-50.
- [9] CHAKRABARTI D, KUMAR R, TOMKINS A. Evolutionary clustering [C]// Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Philadelphia, PA, USA: ACM, 2006: 554-560.
- [10] LIN Yuru, CHI Yun, ZHU Shenghuo, et al. FacetNet: A framework for analyzing communities and their evolutions in dynamic networks [C]// International Conference on World Wide Web. Beijing, China: ACM, 2008: 685-694.
- [11] NING Huazhong, XU Wei, CHI Yun, et al. Incremental spectral clustering by efficiently updating the eigen-system [J]. Pattern Recognition, 2010, 43(1): 113-127.
- [12] CAZABET R, AMBLARD F, HANACHI C. Detection of overlapping communities in dynamical social networks [C]// IEEE Second International Conference on Social Computing. Minneapolis, MN, USA: IEEE, 2010: 309-314.
- [13] NGUYEN N P, DINH T N, XUAN Y, et al. Adaptive algorithms for detecting community structure in dynamic social networks [J]. IEEE INFOCOM, 2011, 28(6): 2282-2290.
- [14] NGUYEN N P, DINH T N, TOKALA S, et al. Overlapping communities in dynamic networks: their detection and mobile applications [C]// International Conference on Mobile Computing and Networking, MOBICOM 2011. Las Vegas, Nevada, USA: [s.n.], 2011: 85-95.
- [15] XIN Y, XIE Z Q, YANG J. An adaptive random walk sampling method on dynamic community detection [J]. Expert Systems with Applications, 2016, 58: 10-19.