软件学报 ISSN 1000-9825, CODEN RUXUEW
*Journal of Software*,2013,24(9):2042−2061 [doi: 10.3724/SP.J.1001.2013.04439]
©中国科学院软件研究所版权所有.
E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

# 动态网络模式挖掘方法及其应用*

高 琳, 杨建业, 覃桂敏

(西安电子科技大学 计算机学院,陕西 西安 710071)

通讯作者: 高琳, E-mail: lgao@mail.xidian.edu.cn

**摘 要**: 静态复杂网络研究在揭示社会网络、信息网络和生物网络的形成和演化机制方面取得了重要成果,其方法和结果对系统生物学产生了重要影响.但现实世界中,很多网络是随时间发生变化的,即动态网络.以动态网络为对象,对动态网络的拓扑特性分析、动态网络相关的各种模式挖掘模型和方法进行了综述、比较和分析.特别地,将动态网络模式分析方法应用于生物网络和社会网络,分析了生物网络相关的动态功能模块和模式演化问题、科学家合作网络和社交网络的动态模式.最后指出了动态网络的模式挖掘方法及其在动态生物网络和社会网络研究中存在的问题和挑战,并对未来的研究方向进行了分析.

**关键词**: 动态网络;社会网络;生物网络;模式挖掘

**中图法分类号**: TP181 **文献标识码**: A

中文引用格式: 高琳,杨建业,覃桂敏.动态网络模式挖掘方法及其应用.软件学报,2013,24(9):2042−2061. http://www.jos.org.cn/1000-9825/4439.htm

英文引用格式: Gao L, Yang JY, Qin GM. Methods for pattern mining in dynamic networks and applications. Ruan Jian Xue Bao/Journal of Software, 2013,24(9):2042−2061 (in Chinese). http://www.jos.org.cn/1000-9825/4439.htm

## Methods for Pattern Mining in Dynamic Networks and Applications

GAO Lin, YANG Jian-Ye, QIN Gui-Min

(School of Computer Science and Technology, Xidian University, Xi’an 710071, China)

Corresponding author: GAO Lin, E-mail: lgao@mail.xidian.edu.cn

**Abstract**: Studies of static complex networks have brought significant progress in revealing the mechanism for forming and evolving of social networks, information networks, and biological networks. However, many real word networks change with time and this type of networks is the so called dynamic networks. This paper focuses on dynamic networks to study the related pattern mining method and its applications in biological and social networks. First, the study analyzes the topological properties of the dynamic networks. Then we make a comparison and analysis to the algorithms and models for variety of pattern mining in dynamic networks. Specifically, we analyze the dynamics properties of biological and social networks. Based on this property, we study the biological networks related pattern mining problems, such as dynamic function module, pattern evolution and complex diseases associated pattern, the dynamic pattern in social network. Finally, some key problems and challenges in biological and social network are highlighted, as well as the future research directions.

**Key words**: dynamic network; social network; biological network; pattern mining

近年来,动态网络受到高度关注[1],对其研究也更加深入[2−4].事实上,几乎所有复杂系统都具有某种动态特性,如信息交互网络及科学家合作网络[3]、社交网络、生物网络[5−7]、疾病分子网络[8−10]等,将这些复杂系统建模成动态网络,是一种合理有效的方式[1].

从抽象的数学模型来看,动态网络其实是一个有序的图序列,表示复杂系统在不同时刻的快照,分析每个时

* 基金项目: 国家自然科学基金(60933009, 61072103, 61100157, 61174162)

收稿时间: 2012-06-26; 修改时间: 2013-03-29; 定稿时间: 2013-06-09

© 中国科学院软件研究所 http://www.jos.org.cn

刻网络的进化情况,是动态网络研究的重点.动态网络研究目前还处于起步阶段,很多问题的提出及研究方法都源于静态网络,如动态网络的拓扑结构特征分析、动态社团挖掘、频繁模式挖掘以及某些特定条件的模式预测等,其问题及概念的定义都与静态网络类似.然而,基于动态网络的时序特点,动态模式挖掘不是各时刻快照上的简单叠加.目前,国内外很多研究人员针对动态网络的各种问题进行探索,剑桥大学计算机实验室系统研究组 John 博士致力于动态网络拓扑特征分析研究,博士期间发表了多篇关于这方面的研究成果[11−14].伊利诺伊大学芝加哥分校 Tanya 教授及其团队基于社会网络数据,对动态网络社团及特殊子结构挖掘研究较多,相关研究成果陆续出现在多个主流国际期刊及会议上[15−19].美国 NEC 实验室研究员 Yun 及其团队在动态网络社团结构分析上取得了一系列成果[20−23].动态生物网络方面的研究主要集中在生物网络的动态特性分析[24,25]、生物网络相关的功能模式[26−28]、基于系统生物学的复杂疾病相关模式发现问题的研究[29−31].社会网络分析中,目前有科学家合作网络、社交网络和推荐系统等研究[32−35].

本文针对动态网络,综述了国内外动态网络拓扑特征分析、模式挖掘相关问题和方法的最新研究成果,并对典型的模式挖掘模型和算法进行了比较和分析.对动态生物分子网络中相关模式问题,如功能模块结构、模式的演化及复杂疾病相关模式问题进行了定义与分析,进一步指出这些特定模式的挖掘可以使用动态网络模式挖掘的模型和方法.

本文第 1 节对动态网络中某些重要概念及问题进行定义.第 2 节对动态网络研究的 4 个问题——拓扑特征分析、社团模式挖掘、子图模式挖掘以及模式预测问题,从方法方面进行分析和总结.这些问题均忽略具体应用背景,关注网络拓扑结构上的分析.第 3 节分析生物网络和社会网络中相关的动态网络模式问题及应用.最后,指出动态网络模式挖掘存在的问题和未来的发展方向.

## 1 动态模式挖掘相关定义

动态网络研究中涉及到很多与静态网络类似的概念与问题,本节对这些概念与问题给出形式化定义.

**定义 1(动态网络)**. 动态网络 $G=\langle G_1,G_2,\ldots,G_T\rangle$是时间上的有序图集,其中,$G_t=(V_t,E_t)$是 $t$ 时刻网络拓扑图,$V_t$ 与 $E_t$ 分别是该时刻图的顶点集和边集.

由于应用的需要,通常将动态网络转换成其他形式的图,下面给出图的定义.

**定义 2(概要图)**. 给定一个动态网络 $G=\langle G_1,G_2,\ldots,G_T\rangle$,其对应的概要图为 $G_s=(V_s,E_s,L_e)$,其中,$V_s=V$,$V=V_1\cup V_2\cup\ldots\cup V_T$,$E_s=E_1\cup E_2\cup\ldots\cup E_T$,$L_e$ 是边集 $E_s$ 的标签集合,其元素 $l_e$ 是由 0 和 1 构成的长度为 $T$ 的字符串.当位置 $t$ 为 1 时,表示边 $e$ 出现在第 $t$ 时刻的网络 $G_t=(V_t,E_t)$中;否则,边 $e$ 不在该网络中.

根据不同的应用背景与模式挖掘方法,动态网络对应的概要图可能有其他的名称,如联合图(union graph)[36,37]、进化图(evolving graph)[38],同时对 $l_e$ 的定义也会有所差异.

**定义 3(旅程)**. 对于有序集合 $J=\{(e_1,t_1),(e_2,t_2),\ldots,(e_k,t_k)\}$,若$\{e_1,e_2,\ldots,e_k\}$在进化图 $G_s$ 的拓扑结构上是一条路径,且对$\forall i,1\leqslant i<k$,都满足 $e_i$ 出现在 $t_i$ 时刻的图中,且 $t_{i+1}\geqslant t_i$,则称 $J$ 是 $G_s$ 上的一条旅程,同时定义旅程 $J$ 的出发时间和到达时间分别为 $departure(J)=t_1$,$arrival(J)=t_k$,旅程 $J$ 的拓扑长度为$|J|=k$,时序长度为

$$\|J\|=arrival(J)-departure(J).$$

目前,已有很多文献对动态网络,特别是对进化图定义顶点间距离[38−40],也是对最短旅程的描述,由于需要将时间因素考虑进去,其定义方法与静态网络的距离定义有所不同.

**定义 4(距离)**. $J^*$表示进化图 $G_s$ 上所有旅程的集合,而 $J^*(u,v)\subseteq J^*$表示所有从顶点 $u$ 出发到达顶点 $v$ 的旅程集合,通常有以下 3 种距离定义方式:

(1) 最短距离(shortest distance):

$$d^t(u,v)=\mathrm{Min}\{|J|:J\in J^*(u,v)\cap departure(J)\geqslant t\} \tag{1}$$

(2) 最早距离(foremost distance):

$$\delta^t(u,v)=\mathrm{Min}\{arrival(J)-t:J\in J^*(u,v)\cap departure(J)\geqslant t\} \tag{2}$$

(3) 最快距离(fastest distance):

© 中国科学院软件研究所 http://www.jos.org.cn

$$\theta^t(u,v)=Min\{||J||:J\in J^*(u,v)\cap departure(J)\geqslant t\} \tag{3}$$

从以上定义可以看出,3 种距离仅考虑 $t$ 时刻后从顶点 $u$ 出发到顶点 $v$ 之间的旅程集合.最短距离度量具有最短拓扑距离的旅程,最早距离度量具有最早到达时间的旅程,而最快距离度量具有最短时序长度的旅程.

**定义 5(社团结构)**. 动态网络社团结构 $C=\langle C_1,C_2,\ldots,C_T\rangle$是时间上的有序社团结构集合,其中, $C_t=\{C_t^1,C_t^2,\ldots,C_t^k\}$ 是第 $t$ 时刻网络 $G_t$ 的社团结构,且 $C_t^i$ 是 $C_t$ 中第 $i$ 个社团.

有很多研究工作[18,41−43]关于动态网络的子图挖掘都是以静态网络为基础.这里简单回顾静态网络中的相关概念.对于图 $g_S=(V_S,E_S)$和图 $G=(V,E)$,若满足 $V_S\subseteq V$ 且 $E_S\subseteq E$,则称 $g_S$ 是 $G$ 的子图.对于图 $g_1=(V_1,E_1)$与 $g_2=(V_2,E_2)$,若满足对任意$(v_{1a},v_{1b})\in E_1$,当且仅当$(v_{2a},v_{2b})\in E_2$ 时,则称图 $g_1$ 与 $g_2$ 是同构的,其中 $v_{2a}=f(v_{1a})$且 $v_{2b}=f(v_{1b})$.如果图 $g_1$ 是图 $G$ 的一个子图,且它与图 $g_2$ 是同构的,则通常称 $g_1$ 是 $g_2$ 在 $G$ 中的一个嵌入.

**定义 6(静态频繁子图)**. 若子图 $g_{st}$ 在概要图 $G_s=(V_s,E_s,L_e)$中有多于 $t$ 个嵌入,则称 $g_{st}$ 是 $G_s$ 的一个静态频繁子图.

**定义 7(公共动态模式)**. 给定 $g_{st}$ 是概要图 $G_s$ 的一个静态频繁子图,且 $g_{st1}=(V_1,E_1)$和 $g_{st2}=(V_2,E_2)$是 $g_{st}$ 在 $G_s$ 中的两个嵌入.由于 $g_{st1}$ 和 $g_{st2}$ 是同构的,因此它们的边集 $E_1$ 与 $E_2$ 间必定存在某个双射 $f$.若所有对应边 $e_2=f(e_1)$都满足 $sub(l(e_1),i,j)=sub(l(e_2),i,j)$,其中,$1\leqslant i<j\leqslant T$,则称 $g_{st1}$ 与 $g_{st2}$ 共享一个从位置 $i$ 到 $j$ 的公共动态模式.

**定义 8(动态频繁子图)**. 对于概要图 $G_s$,$g_{st}$ 是具有 $t$ 个嵌入$\{g_1,\ldots,g_t\}$的静态频繁子图.给定一个频繁阈值$\mu$,若 $g_{st}$ 有$\mu$个以上的嵌入共享一个公共动态模式,则 $g_{st}$ 的拓扑图与此公共动态模式构成一个动态频繁子图.

从定义 6 和定义 7 可以看出,静态频繁子图仅关注概要图的拓扑结构,忽略表示边出现情况的标签 $L_e$.而动态频繁子图不但要保证在拓扑上具有频繁性,还要在边的动态性上保持一致.

**定义 9(周期子图)**. 在动态网络 $G=\langle G_1,G_2,\ldots,G_T\rangle$中,设图 $g_f=(V_f,E_f)$是 $G$ 中若干时刻网络的子图,若对于时间序列 $S_P(g_f)=\langle t:g_f\subseteq G_t\rangle$满足对$\forall i:t_{i+1}-t_i=P$ 且$|S_P(g_f)|\geqslant\sigma$,其中,$P$ 是一个常量,$\sigma$是一个最小阈值,则称 $g_f$ 是 $G$ 的一个周期子图.

**问题 1(动态社团挖掘)**. 动态社团挖掘是指对动态网络 $G=\langle G_1,G_2,\ldots,G_T\rangle$挖掘其社团结构 $C=\langle C_1,C_2,\ldots,C_T\rangle$,并进一步分析每个时刻网络社团结构的演化情况.

**问题 2(动态子图挖掘)**. 动态子图挖掘是指对动态网络 $G=\langle G_1,G_2,\ldots,G_T\rangle$中某些特殊结构子图的挖掘,如动态频繁子图和周期子图,这些子图都具有一定的时间特性.

**问题 3(时序边预测)**. 时序边预测是指在给出动态网络 $G=\langle G_1,G_2,\ldots,G_T\rangle$每个时刻的邻接矩阵$\langle A_1,A_2,\ldots,A_T\rangle$的情况下,预测 $T+1$ 时刻网络 $G_{T+1}$ 的邻接矩阵 $A_{T+1}$ 的分布情况.这里,$A_{T+1}$ 的每个元素 $a_{ij}$ 可能是 0 或 1,也可能是某个概率值.

## 2 动态模式挖掘模型与方法

本节分析了动态网络拓扑结构特征及度量指标,对动态网络不同类型模式挖掘模型与方法、同一类型问题的不同算法进行了综述、分类与分析.

### 2.1 动态网络拓扑特征分析

拓扑特征分析是复杂网络的研究基础,对于静态网络,已有的拓扑指标包括距离、直径、密度、聚集系数[44]、介数中心性[45]、模块性[46]等,涉及网络不同层面的度量,如节点、边和子模块.这些度量指标的提出,为进一步分析网络性质提供计算手段.通过距离或聚集系数,计算顶点间的相似性;通过子图密度、边介数以及模块性对网络进行模块划分.

近年来,随着动态网络研究工作的开展,很多研究者开始关注动态网络拓扑特征的分析[12,13,38,47,48].

动态网络拓扑特征最初在运输、无线以及邮件收发等网络中研究较多,这些网络都具有时序特性.Xuan 等人[38]首次在时序网络中定义了“旅程”及“距离”的概念,Casteigts 等人[39,40]对这些概念进行了更规范的定义,如定义 3 与定义 4.旅程可以看作是一种特殊的路径,构成它的边可以在多个网络中有序出现;而距离则是对最小

© 中国科学院软件研究所 http://www.jos.org.cn

旅程的描述,不同形式的距离对旅程度量的侧重点不同.

Tang 等人[11]针对动态网络,给出了最短路径、聚集系数等指标的全局及局部时序度量分析,并比较了与静态指标度量的区别.而基于旅程及不同类型的距离,Tang 等人[13]提出了针对顶点的时序介数(temporal betweenness)中心性与时序接近度(temporal closeness)中心性.通过对网络顶点时序中心性计算,与静态方式顶点中心性进行比较,发现时序中心性更能合理度量顶点在时序网络中的重要程度.另外,Tang 等人[12]研究了动态网络的小世界现象,即顶点间具有较小的时序距离,提出动态网络中的时序相关系数,见公式(4),进而提出时序网络的平均路径长度,见公式(5).

$$C=\frac{\sum_i C_i}{N}C_i=\frac{1}{T-1}\sum_{t=1}^{T-1}\frac{\sum_j a_{ij}(t)a_{ij}(t+1)}{\sqrt{\left[\sum_j a_{ij}(t)\right]\left[\sum_j a_{ij}(t+1)\right]}} \tag{4}$$

$$L=\frac{1}{N(N-1)}\sum_{ij}d_{ij} \tag{5}$$

从公式(4)容易看出:时序相关系数 $C$ 度量动态网络中相邻时刻网络相似吻合程度,当所有时刻网络都相同时,$C$=1 取最大值;而平均路径长度 $L$ 是度量动态网络中所有顶点间时序距离的平均值.为了揭示动态网络也具有小世界现象,他们将动态网络 $G$ 的网络序列 $G_1,G_2,\ldots,G_T$ 的顺序随机排列,得到一个随机的动态网络 $G^{rand}$,通过在 $G^{rand}$ 上计算 $C^{rand}$ 和 $L^{rand}$,发现原动态网络 $G$ 的时序相关系数 $C$ 要比 $C^{rand}$ 大得多,而 $L$ 与 $L^{rand}$ 却相差不大.称这个现象为时间变化系统中的小世界现象.

Grindrod 等人[47]在旅程的基础上提出动态 Katz 中心性[49−51],用一种迭代方式计算动态 Katz 中心性,对大规模稀疏网络效率很高,这种中心性考虑所有长度的路径,而不只是考虑最短路径.从定义 4 可以看出,时序距离与度量时间 $t$ 有关,也就是两点间的距离是度量时间 $t$ 的变量.显然,这种方式在某些情况下是不合理的,比如对两条一样的旅程,不能因为度量时间的原因,使它们的长度有差异.为此,Pan 等人[52]提出平均时序距离作为两点间的时序距离,并给出了计算方法.同样的,他们也提出了相应的时序接近度中心性,通过对不同实验数据的分析,进一步刻画相关静态中心性与时序中心性的差异.

随着动态网络研究的深入,提出了更多的拓扑特征,如时序模体(temporal motifs)[53]、时序视野(temporal view)[54]、网络骨干(network backbone)[54]等,其中很多都是从静态网络演化而来,结合时间因素,更好地刻画了动态网络某些方面的性质.

### 2.2 动态网络社团结构挖掘

社团结构挖掘是近些年来复杂网络研究领域最热点的问题之一,尤其是针对生物网络、社会网络和信息网络,各类社团挖掘方法更是层出不穷.而对于动态网络,社团挖掘研究还处于起步阶段.近几年来,国外学者对这个领域的研究比较活跃,针对这类问题提出了不同的模型和方法,同时对动态网络的社团演化也提出了一些新的概念.此前,文献[1−3]等从不同的角度对动态网络社团挖掘研究现状进行了综述,对问题探讨的深度及角度都有所差异.如:文献[1]是对图的社团挖掘问题进行综述,而对动态网络社团挖掘分析比较片面;文献[2,3]都是针对社会网络的社团挖掘方法的总结,其中,文献[2]更关注于方法的性能及某些方法的特定应用领域.本节从社团挖掘方法与社团演化两方面对动态网络社团挖掘相关研究进行总结与分析.

#### 2.2.1 社团挖掘方法

动态网络社团挖掘方法主要是增量聚类和进化聚类.

1. 增量聚类是一类广泛研究的聚类方法.通常可以分为两种聚类方式[4]:

(1) 对于某一个聚类对象,仅在一个时间点上进行分析;

(2) 对于某一个聚类对象,在多个时间点上进行分析.

传统意义的增量聚类又可以称为数据流聚类,属于第(1)种类型,特别是更新较快的数据流,因为海量数据的持续增加,对数据聚类时只对聚类数据顺序考察一遍,这样才能在有限的存储空间上实现对数据的有效聚类.

© 中国科学院软件研究所 http://www.jos.org.cn

第(2)种类型更接近问题的应用背景,数据建模为复杂网络.这类增量聚类目前研究比较多,Yang 等人[55]受物理学上牛顿万有引力的启发,给复杂网络中的节点间定义虚拟的排斥力与吸引力.文献[55]采用一种近似迭代方法计算节点间的吸引力与排斥力,通过多次迭代,社团内部节点之间的吸引力越来越大,这样,不同社团的顶点将凝聚成不同的斥力极性;相反的,当某条处于社团间的边,其上的排斥力经过多次迭代后超过某个阈值时,这条边将会断裂.当两个社团间的边都断裂之后,就出现了两个社团.Sun 等人[56]提出的 GraphScope 是一种基于信息论的方法,算法的目标有 3 点:

(1) 对图数据流进行分割,将它们分成 $S$ 个片段,使得每个片段内的图具有更多的相似性,而不同片段的图差异性很大;

(2) 对每个片段内若干图快照进行图的划分,这里一个片段内所有图快照具有相同的划分结构;

(3) 整个过程无需用户提供任何参数.

GraphScope 的核心思想是无损压缩,将每个片段内的所有图结构及其社团结构进行编码,目标是搜索一种图数据流分割及对应社团划分结构使得所需的编码信息达到最小.为使整个过程参数自动化,GraphScope 应用局部搜索的方法,从初始给的估计迭代搜索近似最优解.GraphScope 算法关注的是二部图数据流,图 1 给出了一个具有 3 个快照图且被分为两个片段的二部图数据流,其中,圆圈所示顶点为源点,方框所示顶点为终点.从图 1 中可以看出,这个图数据流被分为两个片段,第 1 时刻快照图 $G_1$ 和第 2 时刻快照图 $G_2$ 被分到第 1 个片段,而第 3 时刻快照图 $G_3$ 被分到片段 2.在片段 1 中,源点划分成两个模块 $I_1^{(1)}=\{1,2\}$ 和 $I_2^{(1)}=\{3,4\}$,终点划分成两个模块 $J_1^{(1)}=\{1\}$ 和 $J_2^{(1)}=\{1\}$;在片段 2 中,源点划分成 3 个模块 $I_1^{(2)}=\{1\},I_2^{(2)}=\{2,3\}$ 和 $I_3^{(2)}=\{4\}$,终点划分成 3 个模块 $J_1^{(2)}=\{1\},J_2^{(2)}=\{2\}$ 和 $J_3^{(2)}=\{3\}$.随着后续图数据的加入,GraphScope 以增量的方式,选择信息编码花费最小的方式对新增图进行处理.

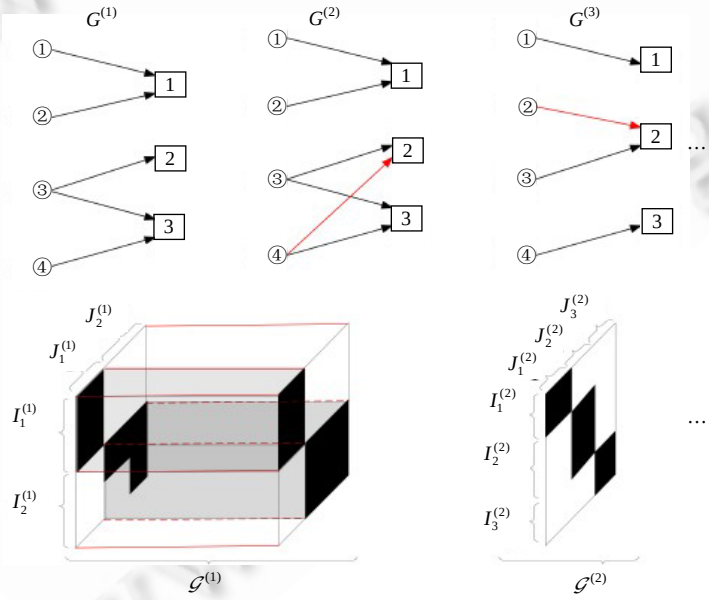



Fig.1 Bipartite graph with three snapshot[56]

图 1 一个具有 3 个图快照并被分为两个片段的二部图数据流[56]

Shan 等人[57]依据所考察网络相邻时刻拓扑结构变化不太大的特点,用网络变化增量方式考察每个时刻的社团结构信息.该方法定义了一些增量相关的顶点集合,在考察 $t$ 时刻的网络时,只需在 $t$–1 时刻网络社团结构的基础上,对增量相关节点进行分析后,就得到 $t$ 时刻的社团结构,从而避免了对整个网络节点全部重新划分,提高了算法的时间效率.但这个方法也存在不足:它没有处理网络节点新增或消失的情况,且默认所有时刻网络的社

© 中国科学院软件研究所 http://www.jos.org.cn

团数目是不变的.Ning 等人[20]提出了一种增量谱聚类的方法检测动态网络的社团结构,通过引入发生向量(incidence vector)和发生矩阵(incidence matrix)的概念,表示网络节点的增加、删除及已有节点之间相似性改变这些网络的动态特性.当网络改变时,通过增量式的迭代更新网络的谱系统,从而获取变化后的网络社团结构.

增量聚类为减少时间上的花费,一般都会对聚类质量造成一些牺牲.下面讨论另一种动态聚类概念——进化聚类,它也是近几年来动态网络社团挖掘研究较多的工作之一.

2. 进化聚类

进化聚类的概念最早由 Chakrabarti 等人[58]提出.算法框架依据动态网络变化缓慢的基本特征,在对每个时刻的网络进行聚类时,同时考虑两个相互冲突的准则,既要使聚类结果与当前时刻的网络结构尽量一致,又要使当前聚类结果与上一时刻的聚类结果差异较小.为此,文献[58]引入了快照质量(snapshot quality)和历史开销(history cost)两个概念,前者用来衡量当前聚类结果 $C_t$ 在当前网络拓扑 $G_t$ 下的聚类质量,而后者衡量当前时刻聚类结果 $C_t$ 与前一时刻聚类结果 $C_{t-1}$ 的差异性.因此,每个时刻最优的聚类结果是使快照质量最大、历史开销最小.为使二者取得平衡,引入平衡因子$\alpha$,聚类质量可以用公式(6)描述:

$$sq(C_t,G_t)-\alpha\cdot hc(C_{t-1},C_t) \tag{6}$$

在这个框架下,每个时刻的聚类算法可以灵活选择,Chakrabarti 等人[58]实现了分层聚类和 $k$-means 聚类两个版本.

最近几年,基于进化聚类的方法框架,出现了很多类似的研究.Chi 等人[21]基于谱聚类提出了保持聚类质量(PCQ)和保持聚类成员(PCM)两个框架.PCQ 是指将当前划分应用于历史数据,对历史数据的聚类质量决定公式(6)中的时序开销.例如有两种划分 $C_t$ 和 $C_t'$ 对当前数据 $G_t$ 具有同样好的划分结果,但若用 $C_t$ 和 $C_t'$ 对历史数据 $C_{t-1}$ 进行划分,$C_t$ 的效果比 $C_t'$ 更好,则认为 $C_t$ 优于 $C_t'$;而 PCM 是直接将当前聚类结果与历史聚类结果进行比较,结果的差异性决定公式(6)中的时序开销.例如,划分 $C_t$ 和 $C_t'$ 对当前数据 $G_t$ 具有同样好的划分结果,但当它们与历史划分 $C_{t-1}$ 比较时,$C_t$ 与 $C_{t-1}$ 更相似,这样我们认为 $C_t$ 与历史划分更加一致,从而认为 $C_t$ 优于 $C_t'$.为了使算法效率更高,Yun Chi 等人分别在这两个框架下实现了松弛的进化谱聚类算法.针对多特征网络,即一个网络中的节点存在多个类型,而不同类型节点之间可能会有联系,Tang 等人[59]提出一种基于谱聚类的进化聚类框架来检测这类网络的社团结构,这个框架可以处理节点及节点之间边的进化等情况.但它的不足也很明显:需要用户提供不同类型边的权重和每种类型网络的社团数目,且社团数目也不能随网络的进化而改变.鉴于动态社团变化不是稳定不变的,Kevin 等人[4]提出一种自适应进化聚类方法.这种方法基本上是在进化聚类框架下来加以实现,不同于公式(6)中平衡因子$\alpha$的确定,根据每个时刻网络的不同,$\alpha$可能会自动调整其大小,以此实现聚类结果的平滑性.

Kim 等人[60]提出一种基于微粒与密度的进化聚类方法.首先将动态网络建模成一系列由纳米社团(nano-communities)构成的网络快照,社团可以定义为纳米社团构成的稠密子图;然后提出基于密度的聚类方法,使用开销嵌入技术(cost embedding technique)实现聚类结果的时间平滑性,提高了算法的效率;同时,该技术使得这种方法不依赖聚类方法的选择及节点间相似性的定义.另外,Kim 等人[60]在相邻时刻的节点之间定义了相似性连接,通过这些连接,得到一个完整的多部图,如 $t$ 个时刻的动态网络对应一个 $t$ 部图.使用基于信息理论的映射方法,跟踪社团演化情况.该方法弥补了之前多数方法具有的两个不足:

(1) 每个时刻聚类数目不变;

(2) 以迭代方式使公式(6)达到最优而使效率降低.

为了降低噪声数据对结果的影响,提高算法稳定性,近年有一些方法应用统计学中的马尔可夫过程建模进化聚类[22,23,61].通过贝叶斯方法,Lin 等人[22]提出 FacetNet 框架,这个框架将动态社团建模成隐式马尔可夫链,最后通过 EM 算法最大化后验概率分布.FacetNet 与之前进化聚类的不同在于:不仅社团结构得到社团的进化,同时社团的进化也反过来会调整网络社团的结构.基于马尔可夫过程,Yang 等人[23]提出一种动态随机快模型(DSBM)来挖掘动态网络社团结构以及它们的进化情况.该方法采用贝叶斯方法估计参数的后验概率,为使算法更具实用性,对于大型稀疏网络,应用模拟退火算法和吉布斯抽样算法来实现对参数的后验概率估计.

© 中国科学院软件研究所 http://www.jos.org.cn

目前,在早期 Chakrabarti 提出方法的框架下,研究人员提出了很多更好的框架和方法,更加关注结果的合理性与效率的可行性.特别是与统计方法的结合,使得这些方法具有实用性.但应用统计方法时,参数估计是一个必须解决的问题,如何更有效地获取更合理的相关参数至关重要.表 1 对部分动态网络挖掘算法进行了比较,分析了不同算法的特点.

**Table 1** Comparison of algorithms for dynamic community detection

**表 1** 动态社团挖掘算法比较

| 算法类型 | 典型算法 | 算法特点 |
|---|---|---|
| 增量聚类 | IA-MCS[55] | 通过顶点间虚拟相互作用力定义顶点的相似程度,<br>算法的效率很大程度上取决于网络初始分布 |
| | GraphScope[56] | 基于信息熵定义模块划分目标函数,算法应用局部搜索策略,最优解依赖于初始划分 |
| | IC[57] | 通过改变增量相关顶点的社团归属优化目标函数,<br>不能处理顶点新增和消失的情况,社团数目不变 |
| | ISC[20] | 基于谱聚类,增量式迭代更新网络的谱系统,提高了谱聚类效率 |
| 进化聚类 | EC[58] | 基于层次聚类和$k$-means聚类,用户提供进化平衡因子$\beta$ |
| | ESC[21] | 基于谱聚类,社团数目不变,算法拓展后没有限制,用户提供进化平衡因子$\beta$ |
| | HDP-HTM[61] | 基于分层狄利克雷过程和隐式马尔可夫过程,用户提供进化平衡因子$\beta$ |
| | EMMC[59] | 基于谱聚类,用户提供边的权重及社团数目,且社团数目不变,用户提供进化平衡因子$\beta$ |
| | FacetNet[22] | 基于隐式马尔可夫链和EM算法,社团数目不变,<br>算法拓展后没有限制,用户提供进化平衡因子$\beta$ |
| | Particle-and-Density[60] | 基于密度聚类,应用开销嵌入技术实现时序平滑性,用户提供进化平衡因子$\beta$ |
| | AEC[4] | 基于相似性矩阵的自适应聚类方法,无需用户提供进化平衡因子$\beta$ |
| 模型聚类 | DSBM[23] | 基于马尔可夫过程模型,使用模拟退火和吉布斯抽样算法优化参数.<br>对参数的后验概率的估计是算法主要瓶颈 |

2.2.2 社团结构演化分析

社团演化分析是一项有意义的工作,如考察社会网络时,一个团体可能会随着时间的变化而呈现出增大或减小,甚至分裂等变化趋势,这对预测网络社团变化是很有用的,本节回顾近年来在动态网络社团演化方面的工作进展.

为了观察网络社团变化情况,Berger-Wolf 等人[15,16]提出群(group)的概念,一个群 $g$ 是网络节点集 $V$ 的子集,即 $g\subseteq V$.在动态网络某个时刻的快照图上,一个群就是一个社团.为分析社团演化情况,Berger-Wolf 等人[15]进一步提出了超群(MetaGroup)的概念.超群必须满足两个特点:

(1) 由一系列连续时刻的群组成,且任意两个群不能来自同一快照,而持续时间大于某个阈值;

(2) 相邻两个群的相似性必须大于某个阈值.

在超群的基础上,文献[15]给出了很多类似的概念,如最持久超群、最稳定超群以及最大超群等统计概念,可以通过这些统计特性,分析社团演化情况.文献[16]在文献[15]的基础上进行了更深入地分析研究,提出 5 种假设,这些假设对网络顶点的社团归属情况加以限定;同时还对其中的 3 个假设定义了事件开销函数,根据这些开销可以获取动态网络最优社团结构.另外,文献[16]还应用图的着色问题分析动态网络社团演化.

Palla 等人[32]针对社团演化,首次提出 6 种社团演化事件:增长、缩减、合并、分裂、新增和消亡,如图 2 所示.在相邻的网络快照 $G_t,G_{t+1}$ 以及它们的合并图 $G_t\cup G_{t+1}$ 上,应用经典的派系过滤算法(CPM)[62]分析相邻时刻社团之间的对应性.着重分析了社团大小与社团年龄之间的关系.通过对电话网络和科学家合作网络[32]的研究发现,大社团的年龄也比较长.

Asur 等人[63]在上述工作的基础上进行了扩展,关注的对象不仅仅是社团,还包括单个顶点.在相邻时刻,针对社团提出 5 种事件:继续、$k$-合并、$k$-分裂、形成和消亡;针对单个顶点提出 4 种事件:出现、消失、加入和离开.基于这些事件,提出了 4 个度量标准来捕获节点的行为趋势对网络演化的影响,即稳定性度量、社会性度量、吸引力度量和影响力度量.通过对这些事件与指标的分析与计算,在更广义的程度上分析了网络的进化行为,而不仅限于社团的进化.同样,在 Palla 等人[32]的基础上,Greene 等人[64]提出动态社团是由图 2 中一系列事件构成.首先,在每个静态图上应用合适的模块划分算法;之后,在相邻时刻间进行社团的匹配,得到图 2 的某个事件.

© 中国科学院软件研究所 http://www.jos.org.cn

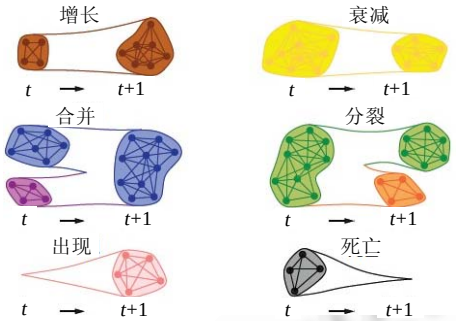



Fig.2 Possible events in community evolution[32]

图 2 社团演化过程中可能发生的事件[32]

由于在大规模网络上跟踪所有社团的演化情况是一个很困难的问题.而某些情况下,并不需要针对整个网络,只是某些特殊的顶点.如,Fenn 等人[65]提出了以顶点为中心的社团重叠系数,见公式(7).

首先,应用 Reichardt 和 Bornholdt 提出的旋转模型[66]在连续快照图上识别社团;然后,根据公式(7),得到目标顶点所在社团的演化行为.

$$a_i^t(\tau)=\frac{|C_i(t)\cap C_i(t+\tau)|}{|C_i(t)\cup C_i(t+\tau)|} \tag{7}$$

Wang 等人[67]关注的是一些核心顶点的集合,提出一种基于核心顶点的社团进化跟踪算法.例如,属于两个相邻时刻的社团,若享有同一个核心顶点,则认为后一个社团是从前一个社团进化而来.

最近,Lancichinetti 等人[68]研究了经典的一致聚类方法,并将其应用到了动态网络的社团划分上,使用 Jaccard 指数分析社团演化情况.

### 2.3 动态网络子图模式挖掘

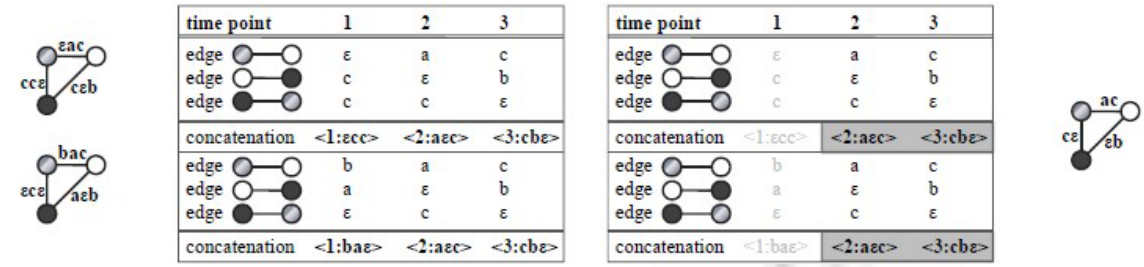
图数据中频繁模式挖掘的经典算法有 FSG[69],gSpan[70],CloseGraph[71],Gaston[72]等,这些算法大多都是针对单图或图集.由于本文的研究对象是动态网络,因此关注最近几年动态网络的频繁模式研究,发现很多工作都采用静态网络的思想.

#### 2.3.1 动态频繁子图挖掘

动态频繁子图指在具有时间特性的动态网络中频繁出现的子图模式,它与之前的频繁子图有共同之处,也有它的独特之处.Borgwardt 等人[41]首次提出动态频繁子图的概念.将一系列时序图 $G_{ts}=\langle G_1,\ldots,G_n\rangle$定义为一个动态图 $DG=(V_{DG},E_{DG},es)$,其中,$V_{DG}$ 是顶点集,$E_{DG}$ 是边集,$es$ 是边出现的模式.他们重新定义了子图,分为拓扑子图和动态子图.若 $DG_1$ 是 $DG_2$ 的拓扑子图,则要求 $DG_1$ 和 $DG_2$ 的出现时间是相同的,而拓扑上 $DG_1$ 是 $DG_2$ 的子图;若 $DG_1$ 是 $DG_2$ 的动态子图,则要求 $DG_1$ 和 $DG_2$ 的顶点与边集完全相同,只是 $DG_1$ 在出现时间上是 $DG_2$ 出现时间的某一段时间,即 $DG_1$ 的出现时间是 $DG_2$ 出现时间的连续子段.Borgwardt 等人对动态频繁子图的定义如下:如果一个动态图 $DG_2$ 具有 $t$ 个与动态子图 $DSG_1$ 相同的嵌入时,则说 $DSG_1$ 是 $DG_2$ 的动态频繁子图.这里,$t$ 是用户设定的频繁阈值.文献[41]提出的 Dynamic GREW 用来挖掘动态频繁子图,应用后缀树[73]发现所有单边的动态频繁子图,继而以迭代的方式通过合并小的动态频繁子图得到规模更大的动态频繁子图.

在 Borgwardt 工作的基础上,Wachersreuther 等人[43]提出一种高效的算法(DFSD)挖掘动态图中的动态频繁子图.采用字符串方式将静态频繁子图编码,然后应用经典的字符串公共子串发现算法得到这些静态图共有的边出现模式,从而获得动态图中的动态频繁子图.DFSD 算法的目标是在给定的动态网络 $G=\langle G_1,\ldots,G_T\rangle$中挖掘动态频繁子图.这些频繁子图不仅在拓扑结构上具有频繁性,而且共享一个公共动态模式.图 3 给出了一个动态频繁子图的挖掘过程,首先在概要图中找出频繁出现的子图,接着用矩阵形式将这些子图表示出来,然后通过公共子串的比较,找出这些子图的公共动态模式,最后还原得到动态频繁子图.

© 中国科学院软件研究所 http://www.jos.org.cn




(a) Embeddings (b) Matrix representation (c) Common substring (d) Dynamic pattern

Fig.3 An example of dynamic frequent subgraph with threshold $t$=2[43]

图 3 动态频繁子图挖掘示意图,频繁阈值 $t$=2[43]

针对变化比较缓慢且稀疏的动态网络,Inokuchi 等人[36,37,42]将动态网络表示为一系列初始网络,定义了 6 种转换规则:顶点新增、删除、重新标记,边新增、删除和重新标记.对任意两个相邻的图 $G_t$ 和 $G_{t+1}$,都可以在 $G_t$ 的基础上,通过一系列转换得到 $G_{t+1}$.文献[42]提出 GTRACE 方法,在转换序列的基础上,枚举所有的频繁子序列从而得到动态频繁子图.与 Dynamic GREW[41]不同,这个方法不仅可以处理边的新增、删除和重标记,还能处理顶点的新增、删除和重标记.由于对规模较大、时间点较多和变化较大的网络,GTRACE 大部分计算开销花费在对不相关的频繁转换子序列的挖掘上,使得效率较低.文献[37]在 GTRACE 的基础上提出一种基于逆向搜索的方法,只关注相关频繁转换子序列,提高了算法的效率.

GTRACE[42]旨在挖掘所有频繁转换子序列(FTS),这些序列代表网络频繁发生的局部变化,而与动态图上频繁子图还是有一定的差异.为此,Inokuchi 等人[36]提出了一类新的子图子序列,称为频繁的相关导出子图子序列(FRISS).根据我们的理解,FRISS 是在动态网络集上挖掘的.为挖掘所有的 FRISS,Inokuchi 等人提出了相应的有效算法.它首先由所有的动态序列图得到一个合并图集;然后在这些图集上应用 AcGM[74]算法,枚举所有频繁连通的导出子图;最后,应用一种从合并图到原有动态图的投影技术得到所有 FRISS.

为了揭示动态网络中某些子图重复出现的现象,You 等人[75,76]提出利用图的重写规则与转移规则框架来分析动态网络.首先提取图的重写规则,如图 4(b)所示,动态网络中相邻时刻的两个图 $G_i$ 与 $G_{i+1}$ 之间有公共的子图 $g_S$,$G_i$ 删除剩下部分 $R_i$ 就能得到 $g_S$,而 $G_{i+1}$ 可以在 $g_S$ 基础上增加 $A_{i+1}$ 得到.在所有相邻时刻进行这些子图提取,就得到了动态图的重写规则,如图 4(c)所示;然后,应用频繁子图挖掘方法分析网络的转移规则,如图 4(d)所示.通过这些有规律的转移规则,可以预测动态图未来的改变.

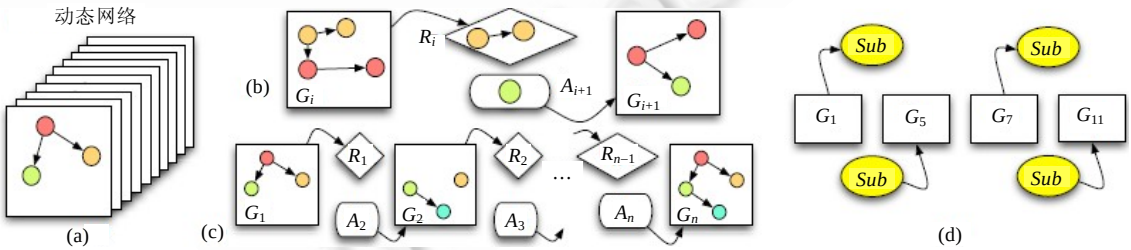



Fig.4 An Example of dynamic frequent subgraph mining[76]

图 4 动态频繁子图挖掘示例[76]

### 2.3.2 周期规律子图挖掘

由于动态网络具有时序的特点,某些子图必然会以规律性出现.近年来,越来越多的研究人员对这类问题进行研究,如文献[18,19,77–79]都是分析动态网络中周期性子图模式.Lahiri 等人[18,19]对这类问题进行了详细的探讨,证明了周期性子图模式挖掘属于 P 问题,也就是能在多项式的时间复杂度里解决.文献[18,19]应用子图哈希

© 中国科学院软件研究所 http://www.jos.org.cn

表与模式树两种数据结构,顺序遍历每个时刻的网络,得到动态网络中的周期性子图模式,PSEMiner 算法基于如下思想:顺序读取动态网络每个时刻的网络快照,并不断维护称为模式树(pattern tree)的数据列表结构,该列表记录着当前所有的周期子图嵌入与那些可能在将来某时刻具有周期性的子图.当某个周期嵌入不满足周期性时,则将它从模式树中清除,如果它的支撑集大于给定的最小阈值,则为周期规律子图.随着每个时刻网络快照 $G_t$ 的读入,不断更新这个模式树,可能涉及修改、增加和删除树节点的操作.由于这个方法不断将后续时间点有可能具有周期规律的子图保存下来,因此它的空间复杂度比较大.

Apostolico 等人[77]对文献[18,19]提出的方法进行了加速,在时刻 $t$ 对模式树进行更新时,并不是对树中所有节点进行更新检查,而是只关注周期长度 $p$ 能整除 $t$ 的节点,利用这个特点,将文献[18,19]提出的方法在时间复杂度上提高了 $T$ 倍,特别是在中等或高密度的数据集上,性能提升比较明显.为了提取动态网络周期规律的结构,与 You 等人[75,76]的研究类似,Yan 等人[78]首先将相邻时刻间网络删除的子图与增加的子图构造出来,然后在所有这些子图中挖掘最大频繁子图,进而应用随机过程模型,如马尔可夫过程,将这些频繁子图建模看成是一系列状态,若某个系列子图经过 $k$ 步转移到第 1 个子图的概率超过设定的阈值,则认为找到一个周期规律的结构.

Qin 等人[79]提出一种高效的方法挖掘动态网络中周期规律子图,与 Borgwardt 等人[41]一样,将动态网络用合并图表示,以 0 和 1 构成的字符串刻画边的出现情况.在这个基础上,首先在所有边中提取出具有周期规律的边集,然后用深度搜索方式,搜索连通且具有相同周期规律的边,这样就构成一个周期规律子图.为了进一步提高算法效率,将刻画边出现情况的 0 和 1 构成的字符串编码成整数权重,以提高判断边模式的效率.算法时间复杂度为 $O(m_{total}+m_{real}\times T^2)$,其中 $m_{total}$ 和 $m_{real}$ 分别是原动态网络总边数与合并图的总边数,$T$ 是动态网络时刻数.存在的缺陷是:某些连通但不呈现周期规律或周期长度不同的边,它们可能会构成一个具有周期规律的图,如图 5 所示,边(1,2)、边(1,4)都没有周期规律,而子图{(1,2),(1,4)}却是周期为 2 的子图.

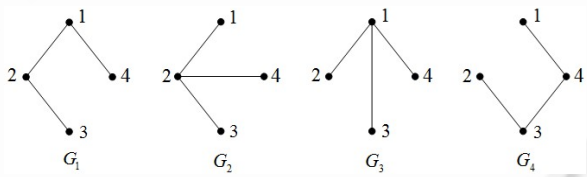



Fig.5 A dynamic network with periodic subgraphs

图 5 具有周期性子图的动态网络

2.3.3 其他子图模式挖掘

在动态网络中挖掘权重大的动态子图,也具有很重要的实际意义.Bogdanov 等人[80]提出最重动态子图(HDS)的概念,研究有权重动态网络,且边的权重为 1 或–1,顶点成员不随时间变化而改变.挖掘 HDS 可以描述为:在所有可能的时间段里,挖掘边权重之和最大的 $k$ 个不重叠子图.要解决这个问题,可以分解为两个子问题:首先,确立时间段.显然,对于 $t$ 个时刻的动态网络,可能的时间段个数为 $t(t+1)/2$,时间复杂度为 $O(t^2)$;另外,给定一个时间段,问题就简化为搜索边权重最大的子图,这是一个 NP 难问题.所以,在没有任何优化的情况下,挖掘最重动态子图需要解决 $O(t^2)$次 NP 难问题,效率太低.为此,针对第 1 个问题,Bogdanov 等人应用图的重叠率对时间段数目进行剪枝.即,具有明显重叠的子时间段,它们产生的聚合图也很相似,因此会得到相似的 HDS.通过这个方法,将原来的 $O(t^2)$搜索空间降到 $O(t\cdot\log^2(t)\cdot|E|)$.针对第 2 个问题,Bogdanov 等人应用一种启发式搜索方法 TopDown 解决经典的最重子图 HS 问题,使问题复杂度降到 $O(|E|\cdot\log(|V|))$.

HDS 是非常有实际意义的问题,如道路交通网络,一个权重大的子网络表明在某个特定的时间段交通处于很堵塞的状态.类似在生物领域,细胞网络结构不变,但分子相互之间的交互状态在改变,挖掘某段时间内交互较多的分子集合,对于识别功能模块或分析进化都是非常有意义的.

近年来,基于动态网络的子图挖掘工作研究比较多,如文献[81,82]等都是关注动态网络某些进化特征,以特定的子图模式描述.而如何以高效的手段挖掘特定网络有用的模式信息,依旧是研究者探索的重点.

© 中国科学院软件研究所 http://www.jos.org.cn

### 2.4 动态网络模式预测

预测问题是数据挖掘领域涉及最多的任务之一,如金融领域中对股票行情走势的预测,社会关系网络中对个体间联系情况的预测,生命系统中对生物分子间相互作用的预测.

#### 2.4.1 边模式预测

Liben-Nowell 等人[83,84]最早阐述了边预测(link-prediction)问题,文献[85,86]等对复杂网络中边的预测问题进行了综述,对有关边预测的方法进行了归类.Lü 等人[85]将这些方法大致分为 3 类:

(1) 基于相似性的方法[84,87−89].即通过计算网络中顶点间的某些相似性指标来预测它们之间是否有边存在.尽管这类方法非常简单,但仍然是目前的主流方法;

(2) 最大似然度方法[90,91].即通过已有的模型与参数使观察到的数据达到最大似然度来判断边的存在情况.这类方法的一个明显缺陷是非常耗时,但它能从一个更高的角度分析网络的结构;

(3) 概率模型[92−95].应用分类方法,首先建立相应的模型,然后计算当期网络不存在的边存在的概率,以此来预测边的存在情况.

Hasan 等人[86]从类似角度将边预测方法归类.然而,很多方法不只属于以上某一种方法,如在有监督的分类方法中,首先通过计算顶点间相似度,构造分类器的特征集合,然后选择合适的分类器模型进行学习.

事实上,边预测是一个广义的概念,一般包括两个方面:结构边预测(structure link prediction)和时序边预测(temporal link prediction).所谓结构边的预测指对网络中潜在的且应当存在的边的识别,也就是丢失边(missing link)的预测.这类边的预测问题一般是给出网络的部分结构,预测另外应该存在的部分.而时序边的预测如问题 3 所描述,是在给出一系列过往时刻网络拓扑信息的情况下,预测下一时刻或多个时刻的网络拓扑信息.目前,大多数文献都属于结构边的预测,而由于本文关注的对象是动态网络,本节剩余部分将重点介绍几种时序边的预测方法.

Sarkar 等人[96]提出一种基于隐藏空间模型(latent space model)的方法对动态网络的边进行预测,是对 Raftery 等人[97]用于静态网络边预测方法的拓展.该方法应用了两个假设:

- 相邻时刻网络中顶点可以在隐藏空间进行小规模的移动,而大的移动不太可能.这个假设也可以理解为,动态网络的变化比较缓慢;
- 提出了一个标准的马尔可夫假设,即顶点在 $t$+1 时刻的隐藏位置只决定于 $t$ 时刻顶点的隐藏位置,而与之前所有时刻位置都是独立的,且 $t$ 时刻网络拓扑图 $G_t$ 只决定于 $t$ 时刻顶点的隐藏位置,而与其他时刻的位置与拓扑图都独立.

文献[96]在一定程度上实现了动态网络中边的预测,但其第 2 个假设仍具片面性,很多时候动态网络的拓扑结构不只是与上一个时刻有关,还可能与之前若干时刻的网络快照有关.

Huang 等人[98]对时序边的预测问题进行了详细的讨论,分析了主流方法中以纯静态方式对将来时刻边的预测,提出了一种基于 ARIMA 模型[99]的时间系列模型.虽然这个模型能够很好应用网络时序信息,但存在的不足是它只能对以往时刻网络中出现过的边进行预测.最后,提出了一种混合时序边的预测算法,对于 $T$+1 时刻任意点对$(i,j)$,先分别计算静态预测方法得到的出现概率 $S_S(i,j)$和时间系列模型得到的出现概率 $S_T(i,j)$,最后通过公式(8)将这两个概率进行线性结合给出一个边出现的概率,其中,$m_S$ 与 $m_T$ 分别是 $S_S(i,j)$与 $S_T(i,j)$中的最小非零值.通过这个混合方式,方法的预测性能有明显提升.ARIMA 模型是通过过去点对$(i,j)$连接边发生的序列,给出出现概率 $S_T(i,j)$.模型忽视了边和边之间的相互影响,认为任意点对$(i,j)$在下个时刻是否连接,与这两个顶点以往的连接情况有关.然而,静态链路预测方法考虑了边和边之间的相互影响,混合链路预测方法结合两种方法后,不仅考虑了边的时序性而且还能考虑边与边之间的相互影响.

$$S(i,j)=(S_S(i,j)+m_S/\alpha)\times(S_T(i,j)+m_T/\alpha),\alpha>1 \tag{8}$$

Dunlavy 等人[100]对二部图形式的动态网络进行边预测,如典型的科学家合作网络,该二部图中两类顶点分别是作者与会议.分别以矩阵分解与张量分解两种思路实现边的预测.基于矩阵的方法中,应用基于矩阵奇异值分解的低秩逼近方法,使其能应用于规模较大的网络.在此基础上,还结合边预测领域中经典的 Katz 指标,在不

© 中国科学院软件研究所 http://www.jos.org.cn

增加过多计算开销的同时,使算法预测更准确.在基于张量的方法中,使用CP张量分解模型[101],通过估计时间维度上的因子,这个方法能预测一个甚至多个时刻的网络信息.

随着动态网络的发展,边预测工作[102−105]开始转向动态网络,无论从问题模型或方法选择上都会有更深更广的发展.

2.4.2 其他模式预测

本节简要回顾动态网络中边以外的其他模式预测问题.

Lahiri等人[17]用频繁闭子图对网络子结构进行预测,每个时刻的网络快照都是由一系列频繁闭子图表示而不是边,还使用了一个假设:网络中某些结构可以预测其他将来时刻出现的结构.对于每个频繁子图对$(g_X,g_Y)$,用一个随机变量$\Phi(x,y)$表示它们之间出现的时间延迟.显然,这个随机变量是非对称的.在每个时刻,对于任意频繁子图对$(x,y)$,不停更新它们之间的时间延迟分布$\Phi(x,y)$,通过建立可信的时间延迟分布之后,对下一时刻的网络进行预测.文献[17]考虑的对象是网络中的频繁子图,而不是每一条边,这在一定程度上提高了算法预测的效率.

由于顶点中心性衡量顶点在网络中的重要程度,因此在很多应用中,对将来时刻某些顶点的中心性估计是很有必要的,Kim 等人[14]对动态网络中顶点中心性进行预测,对 3 种常见中心性,通过 8 种预测函数进行顶点中心性预测.这种方法跳过对网络拓扑的预测,直接对网络顶点中心性进行预测,可以减少很多不必要的计算,实现起来也比较简单,不过对某些参数的估计是必要的.

## 3 动态模式挖掘应用

### 3.1 生物网络中的动态模式

从系统生物学角度看[106,107],生命活动不是由个别生物分子完成的,生物体内的 DNA、RNA 及蛋白质等生物大分子形成了不同的复杂生物网络,如基因调控网络、蛋白相互作用网络、代谢网络、信号传导网络等[108−110].人们已经清楚地认识到,生物功能不是只由一个或几个基因控制的,而是通过生物体内众多的分子共同构成的复杂生物网络实现的,所有的生理或病理活动都基于这种复杂分子网络的结构和动态特性[111,112].在生命活动过程中,网络中的结点和边的关系呈现出动态结构特性,而且这些节点和边形成不同结构的模式[24,25],完成细胞调控和生物体的功能实现[113].

3.1.1 动态功能模块和模式演化

生物网络是随时间动态变化的,发生作用的功能模块和模式也随时间发生变化[26].模块的动态分析,主要分为模块内的变化、模块之间的连接随着时间的变化以及模块的动态行为分析.为了研究模块的暂态变化,常用的方法结合时序基因表达数据,更进一步是研究分析模块之间的连接变化[27].

早在 2003 年,Vespignani 等人[28]在蛋白质网络的功能模块中引入了演化的概念,对网络的分析深入到生物演化过程,认为存在高度保守的拓扑结构功能模体.Komurov 等人[114]集成蛋白质相互作用和基因表达数据分析酵母菌蛋白质网络中的静态和动态模块结构.细胞中存在两类模块:由静态蛋白质构成的静态模块、供调控动态蛋白质构成的动态模块.静态模块可降低遗传扰动和蛋白质表达的噪音,动态模块主要负责细胞行为中基于特定条件的调控.但他们定义的动态模块并不是存在于随时间变化的网络中,而是针对动态蛋白质定义的.

Park 等人[115]提出动态分层的随机块模型,用于蛋白质模式的变化分析,同一模块中的蛋白质按照相互作用的模式随着时间和空间变化.并采用拟南芥 5 个组织和 3 个阶段的数据进行了实验分析,说明了蛋白质动能模式随时间发生变化.同年,Gill 等人[116]利用基因表达数据,用于分析两个不同阶段的网络,如正常情况和疾病状态下的细胞类型,涉及到模块结构的变化,使用统计方法分析了下列 3 种情况的模式变化情况:

(1) 是否不同状态的整体模块结构是不同的?

(2) 是否特定基因的连通性在不同状态下发生了变化?

(3) 是否单个基因的连通性发生了变化?

以上 3 个问题都体现了网络中不同层次的模式随着网络状态的变化在发生变化.

Ferhat 等人[117]针对基因调控网络中的功能模块结构具有的动态特性,研究了增量式模块识别方法,进而分

© 中国科学院软件研究所 http://www.jos.org.cn

析了在基因调控网络中基因在不同情况下的动态行为,及基因之间不同类型的作用关系的变化(激活或抑制). You 等人[76]研究了针对一般生物网络结构模式变化的方法,通过功能模式的演化分析,理解分析生物系统是如何随时间发生变化.随着不同类型数据的产生,目前有很多研究都是基于动态结构模式的变化[118],涉及到功能结构及疾病相关模式[119−121].

3.1.2 复杂疾病相关模式

复杂疾病是由多基因、基因与基因相互作用、基因与环境相互作用所引起的,这些相互作用形成一个多层次的复杂生物网络.正是这些复杂网络的变异,引起了疾病的发生与发展.在复杂疾病发生、发展和预后研究中,疾病不同时刻模式的动态分析尤为重要,获取恶化过程相关的基因、基因调控及其与疾病恶化过程的关系以及疾病恶化中基因调控的功能分析.主要的模式有两类:

(1) 致病基因构成的模式,因为复杂疾病是由多个基因导致的,这些基因在疾病的不同阶段产生变化;

(2) 致病基因与疾病恶化之间的关联关系模式,一旦发现了致病基因,进一步的工作是这些基因是如何相互作用影响复杂疾病的恶化过程的,即它们与疾病恶化之间到底存在怎样的关联关系.

Sol 等人[29]指出:复杂疾病需要一个新的概念框架来描述和建模,疾病被视为网络扰动的结果,必须分析生物分子网络的拓扑结构和动态性.从网络动态性方面进行分析,可以深入理解复杂疾病,分析和预测其对扰动的反应.Erler 等人[30]提到:网络结构模式的变化,能够更好地理解疾病发展过程,对于识别致病原因和干预疾病是至关重要的.通过定性分析模块之间在不同阶段或时刻的变化关系,可以深刻理解肿瘤的进展情况[27].

基因的表达形式是蛋白质.癌症会导致蛋白质的变异,从而影响蛋白质功能的改变.Taylor 等人[31]指出,蛋白质相互作用的动态模块性可用于预测乳腺癌.通过观察病人体内蛋白质性质的改变,预测癌症的后果,蛋白质性质的改变是由该蛋白质所在模块模式的变化来刻画.蛋白质相互作用的本质是实现生物体的某一种功能.之所以用动态网络来刻画,尤其是癌症的预测,主要原因是癌症是一个时间比较长的病程,其次,癌症是一个连续相关的过程,预测最终结果可能不仅仅只和一个时间状态相关.Chen 等人[8]指出:癌症的发展过程涉及到发病位置的转移及癌症转移位点随时间的变化,集成动态网络和癌症转移位置来构建癌症时序网络,来预测癌症转移模式.Hidalgo 等人[9]研究不同疾病之间的关系模式,利用病人的历史记录数据及共病关系构建表型疾病网络(phenotypic disease network).通过对这个网络的分析,可以了解疾病的进展.目前,有很多基于动态分析的方法发现复杂疾病相关模式的研究[10,122,123],并将这些研究深入到药物靶标的发现[124].

### 3.2 社会网络中的动态模式

科学家合作网、论文引用网、社交网络等属于社会网络,这些网络随着时间推移而发生变化,取决于个体角色、社会地位及研究兴趣的转变.第一个研究社会网络动态演化的是 Harrison[33]所做的桑普森修道院数据集研究.Barabasi 等人[125]以动态和进化为重点,研究了科学家合作网络,指出网络的拓扑性质由它的动态性和增长过程决定.

3.2.1 科学家合作网络

科学家合作网络是典型的动态网络,随着新作者的加入和原作者间新的合作关系的产生而不断变化.Palla 等人[32]通过分析康奈尔大学图书馆电子资源冷凝态物质学科的 30 000 多个作者 142 个月的论文合作关系,定义了社团结构(研究团队或者合作研究)的合并、分裂、增长、消减等事件,并发现了社团规模、社团稳定性以及社团的生命周期之间的关系,即:小规模社团是相对稳定的,随着时间变化成员变化较小;而大规模社团在不断地更新自身的成员,大规模社团通过不断更新成员获得更长的生命周期.研究更多的科学家合作网络是 DBLP 数据库(http://www.informatic.uni-trier.de/~ley/db),该数据库维护了计算机领域的论文、作者、会议/期刊等相关信息.整个 DBLP 合作网络数据存在一个 XML 文件中,可以通过解析该文件获取关注的论文信息.由于计算机科学包含的子领域太多,很多研究是基于该数据的子集进行的,如选择数据库、数据挖掘和人工智能这 3 个领域的 28 个会议在 1997 年~2006 年的论文作为研究对象[22,63].在建模时,每个作者作为一个节点,两个作者之间若有论文合作关系则建立一条边.一般是以年为粒度建立合作网络.然后分析研究团队的发展和变化规律,以及研究人员主导、参与或离开研究团队的情况.几个工作[22,60,63]以 DBLP 合作网络作为一个典型的动态网络

© 中国科学院软件研究所 http://www.jos.org.cn

数据,侧重于研究识别动态网络中的社团结构及其变化的方法.

研究人员合作关系中除了作者和论文关系外,还有论文关键词以及发表的会议/期刊等信息,这些对分析合作及其变化提供更丰富的内容.因此,在对合作关系建模时,Gupta 等人[126]建立了科学家合作网络的异构网络模型,用星型结构表示一篇论文,星型中心节点表示论文,外围的属性节点分别表示论文的属性,包括作者、关键词和会议/期刊等.类似地,Tang 等人[59]采用多模网络建立合作网络,包含 4 种节点类型,即论文、研究人员、关键词、会议/期刊.在该模型中,还建立了论文之间的引用关系,研究者和会议/期刊之间的编辑等关系.因此,相比于异构网络模型,该模型表示了更多的信息.这几种更复杂的网络模型表达了研究人员合作的更多的信息,但同时也为分析带来了更多的复杂性.

上述对动态的科学家合作网络分析侧重于分析研究团队及其研究人员随着时间的变化情况.实际上,对科学家合作网络分析还可以:

(1) 明确某个特定的研究领域在一定时间段的主流趋势;

(2) 明确某个特定的课题在一定时间段的主流趋势;

(3) 观察其他研究人员如何在不同领域间转换;

(4) 分析一个领域随着时间是如何变化的;

(5) 能够发现各个研究领域的热点问题、公共问题、子领域划分等;

(6) 能够介绍一个属性的相关研究人员以及一个新领域的产生.

Rosvall 等人[127]基于 Thomson-Reuter 期刊引文报告建立 1997 年~2007 年的引文网络,分析网络社团结构的变化,并以可视化的方式形象地给出一些学科在这 11 年的发展变化情况.Scott 等人[128]通过分析干细胞研究的科学家合作网络随时间的变化来回答“人类诱导性多功能干细胞(hiPSCs)的发展是否会改变干细胞研究轨迹”这样的问题.通过分析发现,hiPSCs 不会取代人类胚胎干细胞研究,相反地,这两种干细胞研究方法是互补和依赖的.

3.2.2 社交网络

与科学家合作网络一样,因特网是很重要的动态网络,节点、链接的添加与删除都很频繁,网络的演化特性也与一般的动态网络相符.近年来,随着博客、微博的兴起,信息发布呈现个性化的特点,无论是规模还是数量,社交网络都有了长足的发展.社交网络成为人们获取信息、展现自我、营销推广的窗口.随着用户的不断增加,社交网络构筑了一个庞大的网络社会.众多的门户网站、社交网站及电子商务平台提供了海量的用户和数据.研究网络的演化及如何整合有价值的信息投入应用,显得尤为重要[34].

从企业的角度而言,忠诚的用户可以转化成收益的来源,因此,保证活跃用户的数量对在线网络运营商来说是很重要的[129].以 Twitter 为例,40%的账户处于休眠状态,60%的账户在一个月内仅登陆一次[130].

Wilson 等人[131]发现,Facebook 的用户在早期十分活跃;Viswanath 等人[132]发现,用户之间的连接建立后,信息的传送数量随时间递减,也说明了用户的兴趣随时间衰减.因此,研究用户行为的演变、并使用户保持活跃状态是很有价值的研究课题.Lang 等人[129]发现,网络中度较大或聚集系数较高的节点、网络社交关系与真实社交关系重叠度大的节点存留时间较长.通过向用户推荐好友、鼓励用户建立连接等手段,可以有效促进社团增长,减缓社团衰减或死亡过程.

现代企业已经不再局限于传统的营销渠道,并把微博作为日常经营的重要内容,通过对竞争对手微博动态的跟踪分析、对用户行为的动态研究,可以有效吸引用户,提高广告效应.如把用户当作节点,行为当作边,根据边的动态变化可以研究社团的动态演化;通过应用频繁子图挖掘技术,分析网络的转移规则,可以预测动态网络未来的改变.

随着电子商务的迅速发展,推荐系统应运而生.通过分析用户的历史购买行为,为用户提供个性化服务.而用户的购买记录是有时序信息的,建立对应的动态网络模型,并通过动态模式挖掘方法来进行有效的分析,例如,可以通过对购物记录分析,获得用户感兴趣的方向,或是兴趣方向的迁移,可以对用户提供比基于静态网络模型的推荐系统更好的推荐服务,实现网络上的一对一营销.

© 中国科学院软件研究所 http://www.jos.org.cn

在现代政治中,政治领域的大数据时代已经到来.Boutet 等人[133]搜集了 2010 年英国大选阶段约 220 000 个用户的 1 150 000 条信息,发现工党最为活跃且影响力大,而保守党最有组织.Amelio 等人[35]分析了意大利议会投票中的集群效应和社团演化,通过采集一段连续时间内的网络数据,分析议员的投票记录,得到议会的内部社团结构和潜在信息.2012 年的美国大选中,奥巴马的竞选团队收集了过去两年的数据,创造出摇摆州选民的精细模型,首次尝试在 Facebook 上复制传统的上门宣传活动,分析并预测各个族群的选民在任何时刻的趋势.

## 4 总结与问题讨论

动态网络由于其广泛的应用背景,成为目前复杂网络研究领域的热点问题.本文首先分析和总结了动态网络相关模式挖掘方法的研究进展,特别地,关注了社团结构模式挖掘、子图模式挖掘以及模式预测问题.考虑了两个不同领域的应用,生物网络和社会网络.对动态生物分子网络中相关模式问题,如功能模块结构、模式的演化及复杂疾病相关模式问题进行了分析,并进一步指出这些特定模式的挖掘方法可以使用动态网络模式挖掘的模型和方法.对于社会网络,将其与社交网络以及科学家合作网络的数据分析相结合,指出了其中存在的与动态模式相关的问题.

动态网络分析方法的研究,无论是在生物信息领域还是在社会和科技领域都处于初级阶段,而动态生物分子网络分析研究处于起步阶段.构建动态生物网络数据需要建立在生物实验技术、生物医学理论、建模方法等基础上,针对具体问题建立适当的计算模型.如何基于动态生物分子网络研究复杂疾病,获得疾病发生、发展和干预的分子机制具有极大的挑战性.关于动态分子网络在疾病方面的研究可考虑:

(1) 如何构建反应生物系统的动态生物分子网络模型;

(2) 如何设计高效的算法,演化模式挖掘算法涉及到在大规模动态网络中搜索频繁子图的计算问题,使其能真正应用到人类的动态生物分子网络分析中;

(3) 如何评价动态网络模式分析方法和结果,是否可以定义与动态相关的指标进行度量.

这些问题都是未来动态网络研究涉及的关键问题.同样,在社会、政治、经济、科技等领域,对特定的问题,可以建立相关数据的动态网络模型,分析数据变化规律,预测未来发展趋势,从而进行有效地决策支持和行为调控.由于社会领域的数据规模大、变化快、内容复杂,因此,如何选择适当的时间粒度建立动态网络模型,如何集成相关的数据以进行更有效的分析,是需要考虑的重要问题.除了一些通用的模型和方法,可以结合某个 IT 应用,对应用中的具体问题建立特定的模型和设计有针对性的分析方法,从而更好地解决实际问题.

**致谢** 在此,感谢本文文献中研究人员的工作,也感谢论文评阅专家的意见和建议.

**References**:

[1] Fortunato S. Community detection in graphs. Physics Report, 2010,486(3-5):75−174. [doi: 10.1016/j.physrep.2009.11.002]

[2] Papadopoulos S, Kompatsiaris Y, Vakali A, Spyridonos P. Community detection in social media. Data Mining and Knowledge Discovery, 2012,24(3):515−554. [doi: 10.1007/s10618-011-0224-z]

[3] Spiliopoulou M. Evolution in social networks: A survey. Social Network Data Analytics, 2011. 149−175. [doi: 10.1007/978-1-4419-8462-3_6]

[4] Xu KS, Kliger M, Hero AO. Adaptive evolutionary clustering. Eprint arXiv: 1104.1990. 2011.

[5] Fraser HB. Modularity and evolutionary constraint on proteins. Nature Genetics, 2005,37(4):351−352. [doi: 10.1038/ng1530]

[6] de Lichtenberg U, Jensen LJ, Brunak S, Bork P. Dynamic complex formation during the yeast cell cycle. Science, 2005,307(5710): 724−727. [doi: 10.1126/science.1105103]

[7] Wilkins MR, Kummerfeld SK. Sticking together? Falling apart? Exploring the dynamics of the interactome. Trends in Biochemical Sciences, 2008,33(5):195−200. [doi: 10.1016/j.tibs.2008.03.001]

[8] Chen LL, Blumm N, Christakis NA, Barabási AL, Deisboeck TS. Cancer metastasis networks and the prediction of progression patterns. British Journal of Cancer, 2009,101(5):749−758. [doi: 10.1038/sj.bjc.6605214]

[9] Hidalgo CA, Blumm N, Barabási AL, Christakis NA. A dynamic network approach for the study of human phenotypes. PLoS Computational Biology, 2009,5(4):e1000353. [doi: 10.1371/journal.pcbi.1000353]

© 中国科学院软件研究所 http://www.jos.org.cn

[10] Chuang HY, Lee E, Liu YT, Lee D, Ideker T. Network-Based classification of breast cancer metastasis. Molecular Systems Biology, 2007,3(140):1−10. [doi: 10.1038/msb4100180]

[11] Tang J, Musolesi M, Mascolo C, Latora V. Temporal distance metrics for social network analysis. In: Proc. of the 2nd ACM Workshop on Online Social Networks. 2009. 31−36. [doi: 10.1145/1592665.1592674]

[12] Tang J, Scellato S, Musolesi M, Mascolo C, Latora V. Small-World behavior in time-varying graphs. Physical Review E, 2010, 81(5):055101. [doi: 10.1103/PhysRevE.81.055101]

[13] Tang J, Musolesi M, Mascolo C, Latora V, Nicosia V. Analysing information flows and key mediators through temporal centrality metrics. In: Proc. of the 3rd Workshop on Social Network Systems. 2010. 1−6.

[14] Kim H, Tang J, Anderson R, Mascolo C. Centrality prediction in dynamic human contact networks. Computer Networks, 2012, 56(3):983−996. [doi: 10.1016/j.comnet.2011.10.022]

[15] Berger-Wolf TY, Saia J. A framework for analysis of dynamic social networks. In: Proc. of the 12th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. 2006. 523−528. [doi: 10.1145/1150402.1150462]

[16] Tantipathananandh C, Berger-Wolf T, Kempe D. A framework for community identification in dynamic social networks. In: Proc. of the 13th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. 2007. 717−726. [doi: 10.1145/1281192.1281269]

[17] Lahiri M, Berger-Wolf TY. Structure prediction in temporal networks using frequent subgraphs. In: Proc. of the 2007 IEEE Symp. on Computational Intelligence and Data Mining. 2007. 35−42. [doi: 10.1109/CIDM.2007.368850]

[18] Lahiri M, Berger-Wolf TY. Mining periodic behavior in dynamic social networks. In: Proc. of the 8th IEEE Int'l Conf. on Data Mining. 2008. 373−382. [doi: 10.1109/ICDM.2008.104]

[19] Lahiri M, Berger-Wolf TY. Periodic subgraph mining in dynamic networks. Knowledge Information System, 2009,24(3):467−497. [doi: 10.1007/s10115-009-0253-8]

[20] Ning HZ, Xu W, Chi Y, Gong YH, Huang TS. Incremental spectral clustering by efficiently updating the eigen-system. Pattern Recognition, 2010,43(1):113−127. [doi: 10.1016/j.patcog.2009.06.001]

[21] Chi Y, Song XD, Zhou DY, Hino K, Tseng BL. Evolutionary spectral clustering by incorporating temporal smoothness. In: Proc. of the 13th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. 2007. 153−162. [doi: 10.1145/1281192.1281212]

[22] Lin YR, Chi Y, Zhu SH, Sundaram H, Tseng BL. FacetNet: A framework for analyzing communities and their evolutions in dynamic networks. In: Proc. of the 17th Int'l Conf. on World Wide Web. 2008. 685−694. [doi: 10.1145/1367497.1367590]

[23] Yang TB, Chi Y, Zhu SH, Gong YH, Jin R. Detecting communities and their evolutions in dynamic social networks-a Bayesian approach. Machine Learning, 2011,82(2):157−189. [doi: 10.1007/s10994-010-5214-7]

[24] Dreze M, Carvunis AR, Charloteaux B, Galli M, Pevzner SJ, Tasan M. Evidence for network evolution in an arabidopsis interactome map. Science, 2011,333(6042):601−607. [doi: 10.1126/science.1203877]

[25] Han JDJ, Bertin N, Hao T, Goldberg DS, Berriz GF, Zhang LV, Dupuy D, Walhout AJM, Cusick ME, Roth FP, Vidal M. Evidence for dynamically organized modularity in the yeast protein-protein interaction network. Nature, 2004,430(6995):88−93. [doi: 10.1038/nature02555]

[26] Alexander RP, Kim PM, Emonet T, Gerstein MB. Understanding modularity in molecular networks requires dynamics. Science Signaling, 2009,2(81):pe44. [doi: 10.1126/scisignal.281pe44]

[27] Wang XW, Dalkic E, Wu M, Chan C. Gene-Module level analysis: Identification to networks and dynamics. Current Opinion in Biotechnology, 2008,19(5):482−491. [doi: 10.1016/j.copbio.2008.07.011]

[28] Vespignani A. Evolution thinks modular. Nature Genetics, 2003,35(2):118−119. [doi: 10.1038/ng1003-118]

[29] del Sol A, Balling R, Hood L, Galas D. Diseases as network perturbations. Current Opinion in Biotechnology, 2010,21(4):566−571. [doi: 10.1016/j.copbio.2010.07.010]

[30] Erler JT, Linding R. Network-Based drugs and biomarkers. Journal of Pathology, 2010,220(2):290−296. [doi: 10.1002/path.2646]

[31] Taylor IW, Linding R, Warde-Farley D, Liu YM, Pesquita C, Faria D, Bull S, Pawson T, Morris Q, Wrana JL. Dynamic modularity in protein interaction networks predicts breast cancer outcome. Nature Biotechnology, 2009,27(2):199−204. [doi: 10.1038/nbt.1522]

[32] Palla G, Barabási AL, Vicsek T. Quantifying social group evolution. Nature, 2007,446(7136):664−667. [doi: 10.1038/nature05670]

[33] White HC. Identity and Control: A Structural Theory of Social Action. Princeton University Press, 1992.

[34] Jennifer G. The dynamics of Web-based social networks: Membership, relationships, and change. First Monday, 2007,12(11):1−17.

[35] Amelio A, Pizzuti C. Analyzing voting behavior in italian parliament: Group cohesion and evolution. In: Proc. of the Int'l Conf. on Advances in Social Networks Analysis and Mining. 2012. 140−146. [doi: 10.1109/ASONAM.2012.33]

[36] Inokuchi A, Washio T. Mining frequent graph sequence patterns induced by vertices. In: Proc. of the SIAM Int'l Conf. on Data Mining. 2010. 466−477.

[37] Inokuchi A, Ikuta H, Washio T. GTRACE-RS: Efficient graph sequence mining using reverse search. Eprint arXiv: 1110.3879. 2011.

© 中国科学院软件研究所 http://www.jos.org.cn

[38] Xuan BB, Ferreira A, Jarry A. Computing shortest, fastest, and foremost journeys in dynamic networks. Int'l Journal of Foundations of Computer Science, 2003,14(2):267−285. [doi: 10.1142/S0129054103001728]

[39] Santoro N, Quattrociocchi W, Flocchini P, Casteigts A, Amblard F. Time-Varying graphs and dynamic networks. Technical Report, University of Carleton, 2010.

[40] Santoro N, Quattrociocchi W, Flocchini P, Casteigts A, Amblard F. Time-Varying graphs and social network analysis: Temporal indicators and metrics. Eprint arXiv: 1102.0629. 2011.

[41] Borgwardt KM, Kriegel HP, Wackersreuther P. Pattern mining in frequent dynamic subgraphs. In: Proc. of the 6th IEEE Int'l Conf. on Data Mining. 2006. 818−822. [doi: 10.1109/ICDM.2006.124]

[42] Inokuchi A, Washio T. A fast method to mine frequent subsequences from graph sequence data. In: Proc. of the 8th IEEE Int'l Conf. on Data Mining. 2008. 303−312. [doi: 10.1109/ICDM.2008.106]

[43] Wackersreuther B, Wackersreuther P, Oswald A, Böhm C, Borgwardt KM. Frequent subgraph discovery in dynamic networks. In: Proc. of the 8th Workshop on Mining and Learning with Graphs. 2010. 155−162. [doi: 10.1145/1830252.1830272]

[44] Watts DJ, Strogatz SH. Collective dynamics of 'small-world' networks. Nature, 1998,393(6684):440−442. [doi: 10.1038/30918]

[45] Freeman LC. Centrality in social networks: Conceptual clarification. Social Networks, 1979,1(3):215−239. [doi: 10.1016/0378-8733(78)90021-7]

[46] Newman MEJ, Girvan M. Finding and evaluating community structure in networks. Physical Review E, 2004,69(2):026113. [doi: 10.1103/PhysRevE.69.026113]

[47] Grindrod P, Parsons MC, Higham DJ, Estrada E. Communicability across evolving networks. Physical Review E, 2011,83(4): 046120. [doi: 10.1103/PhysRevE.83.046120]

[48] Kim H, Anderson R. Temporal node centrality in complex networks. Physical Review E, 2012,85(2):026107. [doi: 10.1103/PhysRevE.85.026107]

[49] Katz L. A new status index derived from sociometric analysis. Psychometrika, 1953,18(1):39−43. [doi: 10.1007/BF02289026]

[50] Borgatti SP. Centrality and network flow. Social Networks, 2005,27(1):55−71. [doi: 10.1016/j.socnet.2004.11.008]

[51] Newman MEJ. Networks an Introduction. Oxfrod Univerity Press, 2010.

[52] Pan RK, Saramäki J. Path lengths, correlations, and centrality in temporal networks. Physical Review E, 2011,84(1):016105. [doi: 10.1103/PhysRevE.84.016105]

[53] Kovanen L, Karsai M, Kaski K, Kertész J, Saramäki J. Temporal motifs in time-dependent networks. Journal of Statistical Mechanics: Theory and Experiment, 2011,11:P11005. [doi: 10.1088/1742-5468/2011/11/P11005]

[54] Kossinets G, Kleinberg J, Watts D. The structure of information pathways in a social communication network. In: Proc. of the 14th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. 2008. 435−443. [doi: 10.1145/1401890.1401945]

[55] Yang B, Liu DY. Force-Based incremental algorithm for mining community structure in dynamic network. Journal of Computer Science and Technology, 2006,21(3):393−440. [doi: 10.1007/s11390-006-0393-1]

[56] Sun JM, Papadimitriou S, Yu PS, Faloutsos C. GraphScope: Parameter-free mining of large time-evolving graphs. In: Proc. of the 13th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. 2007. 687−696. [doi: 10.1145/1281192.1281266]

[57] Shan B, Jiang SX, Zhang S, Gao H, Li JZ. IC: Incremental algorithm for community identification in dynamic social networks. Ran Jian Xue Bao/Journal of Software, 2009,20:184−192 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/09022.htm

[58] Chakrabarti D, Kumarand R, Tomkins A. Evolutionary clustering. In: Proc. of the 12th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. 2006. 554−560. [doi: 10.1145/1150402.1150467]

[59] Tang L, Liu H, Zhang JP. Identifying evolving groups in dynamic multimode networks. IEEE Trans. on Knowledge and Data Engineering, 2012,24(1):72−85. [doi: 10.1109/TKDE.2011.159]

[60] Kim MS, Han JW. A particle-and-density based evolutionary clustering method for dynamic networks. In: Proc. of the 2009 Int'l Conf. on Very Large Data Bases. 2009. 622−633.

[61] Xu TB, Zhang ZF, Yu PS, Long B. Evolutionary clustering by hierarchical dirichlet process with hidden Markov state. In: Proc. of the 8th IEEE Int'l Conf. on Data Mining. 2008. 658−667. [doi: 10.1109/ICDM.2008.24]

[62] Palla G, Derényi I, Farkas I, Vicsek T. Uncovering the overlapping community structure of complex networks in nature and society. Nature, 2005,435(7043):814−818. [doi: 10.1038/nature03607]

[63] Asur S, Parthasarathy S, Ucar D. An event-based framework for characterizing the evolutionary behavior of interaction graphs. ACM Trans. on Knowledge Discovery from Data, 2009,3(4):16. [doi: 10.1145/1631162.1631164]

[64] Greene D, Doyle D, Cunningham P. Tracking the evolution of communities in dynamic social networks. In: Proc. of the Int'l Conf. on Advances in Social Networks Analysis and Mining. 2010. 176−183. [doi: 10.1109/ASONAM.2010.17]

[65] Fenn DJ, Porter MA, McDonald M, Williams S, Johnson NF, Jones NS. Dynamic communities in multichannel data: An application to the foreign exchange market during the 2007~2008 credit crisis. Eprint arXiv: 0811.3988. 2009.

[66] Reichardt J, Bornholdt S. Statistical mechanics of community detection. Physical Review E, 2006,74(1):016110. [doi: 10.1103/PhysRevE.74.016110]

© 中国科学院软件研究所 http://www.jos.org.cn

[67] Wang Y, Wu B, Du N. Community evolution of social network: Feature, algorithm and model. Eprint arXiv: 0804.4356. 2008.
[68] Lancichinetti A, Fortunato S. Consensus clustering in complex networks. Scientific Reports, 2012,2:336.
[69] Kuramochi M, Karypis G. Frequent subgraph discovery. In: Proc. of the 1st IEEE Int'l Conf. on Data Mining. 2001. 313−320. [doi: 10.1109/ICDM.2001.989534]
[70] Yan XF, Han JW. gSpan: Graph-based substructure pattern mining. In: Proc. of the 2nd IEEE Int'l Conf. on Data Mining. 2002. 721−724. [doi: 10.1109/ICDM.2002.1184038]
[71] Yan XF, Han JW. CloseGraph: Mining closed frequent graph patterns. In: Proc. of the 9th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. 2003. 286−295. [doi: 10.1145/956750.956784]
[72] Nijssen S, Kok JN. A quickstart in frequent structure mining can make a difference. In: Proc. of the 10th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. 2004. 647−652. [doi: 10.1145/1014052.1014134]
[73] Gusfield D. Algorithms on strings, trees and sequences. Computer Science and Computational Biology. Cambridge University Press. 1997.
[74] Inokuchi A, Washio T, Nishimura K, Motoda H. A fast algorithm for mining frequent connected subgraphs. IBM Research Report, RT0448, 2002.
[75] You CH, Holder LB, Cook DJ. Graph-Based data mining in dynamic networks: empirical comparison of compression-based and frequency-based subgraph mining. In: Proc. of the 8th IEEE Int'l Conf. on Data Mining. 2008. 929−938. [doi: 10.1109/ICDMW.2008.68]
[76] You CH, Holder LB, Cook DJ. Learning patterns in the dynamics of biological networks. In: Proc. of the 15th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. 2009. 977−986.
[77] Apostolico A, Barbares M, Pizzi C. Speedup for a periodic subgraph miner. Information Processing Letters, 2011,111(11):521−523. [doi: 10.1016/j.ipl.2011.02.016]
[78] Yan LM, Wang JW. Extracting regular behaviors from social media networks. In: Proc. of the 3rd Int'l Conf. on Multimedia Information Networking and Security. 2011. 613−617. [doi: 10.1109/MINES.2011.137]
[79] Qin GM, Gao L, Yang JY, Li JJ. Evolution pattern discovery in dynamic networks. In: Proc. of the IEEE Int'l Conf. on Signal Processing, Communications and Computing. 2011. 1−6. [doi: 10.1109/ICSPCC.2011.6061813]
[80] Bogdanov P, Mongiov M, Singh AK. Mining heavy subgraphs in time-evolving networks. In: Proc. of the 11th IEEE Int'l Conf. on Data Mining. 2011. 81−90. [doi: 10.1109/ICDM.2011.101]
[81] Berlingerio M, Bonchi F, Bringmann B, Gionis A. Mining graph evolution rules. In: Proc. of the European Conf. on Machine Learning and Knowledge Discovery in Databases. 2009. 115−130. [doi: 10.1007/978-3-642-04180-8_25]
[82] Miyoshi Y, Ozaki T, Ohkawa T. Mining interesting patterns and rules in a time-evolving graph. Lecture Notes in Engineering and Computer Science, 2011,2188(1):448−453.
[83] Liben-Nowell D, Kleinberg J. The link prediction problem for social networks. In: Proc. of the 12th Annual ACM Int'l Conf. on Information and Knowledge Management. 2003. 556−559. [doi: 10.1145/956863.956972]
[84] Liben-Nowell D, Kleinberg J. The link-prediction problem for social networks. Journal of the American Society for Information Science and Technology, 2007,58(7):1019−1031. [doi: 10.1002/asi.20591]
[85] Lü LY, Zhou T. Link prediction in complex networks: A survey. Physica A, 2011,390(6):1150−1170. [doi: 10.1016/j.physa.2010.11.027]
[86] Hasan MA, Zaki MJ. A survey of link prediction in social networks. Social Network Data Analytics, 2011. 243−275. [doi: 10.1007/978-1-4419-8462-3_9]
[87] Zhou T, Lü LY, Zhang YC. Predicting missing links via local information. The European Physical Journal B—Condensed Matter and Complex Systems, 2009,71(4):623−630. [doi: 10.1140/epjb/e2009-00335-8]
[88] Lü LY, Jin CH, Zhou T. Similarity index based on local paths for link prediction of complex networks. Physical Review E, 2009, 80(4):046122. [doi: 10.1103/PhysRevE.80.046122]
[89] Liu WP, Lü LY. Link prediction based on local random walk. Europhysics Letters, 2010,89(5):58007. [doi: 10.1209/0295-5075/89/58007]
[90] Clauset A, Moore C, Newman MEJ. Hierarchical structure and the prediction of missing links in networks. Nature, 2008,453(7191): 98−101. [doi: 10.1038/nature06830]
[91] Guimerà R, Sales-Pardo M. Missing and spurious interactions and the reconstruction of complex networks. Proc. of the National Academy of Science (PNAS), 2009,106(52):22073−22078. [doi: 10.1073/pnas.0908366106]
[92] Yu K, Chu W, Yu SP, Tresp V, Xu Z. Stochastic relational models for discriminative link prediction. In: Proc. of Neural Information Precessing Systems. Cambridge: MIT Press, 2006. 1553−1560.
[93] Taskar B, Wong MF, Abbeel P, Koller D. Link prediction in relational data. In: Proc. of the Neural Information Precessing Systems. Cambridge: MIT Press, 2004. 659−666.
[94] Kashima H, Abe N. A parameterized probabilistic model of network evolution for supervised link prediction. In: Proc. of the 6th IEEE Int'l Conf. on Data Mining. 2006. 340−349. [doi: 10.1109/ICDM.2006.8]

© 中国科学院软件研究所 http://www.jos.org.cn

[95] Wang C, Satuluri V, Parthasarathy S. Local probabilistic models for link prediction. In: Proc. of the 7th IEEE Int'l Conf. on Data Mining. 2007. 322−331. [doi: 10.1109/ICDM.2007.108]

[96] Sarkar P, Moore AW. Dynamic social network analysis using latent space models. ACM Special Interest Group on Knowledge Discovery and Data Mining, 2005,7(2):23−30. [doi: 10.1145/1117454.1117459]

[97] Hoff PD, Raftery AE, Handcock MS. Latent space approaches to social network analysis. Journal of the American Statistical Association, 2002,97(460):1090−1098. [doi: 10.1198/016214502388618906]

[98] Huang Z, Lin DKJ. The time-series link prediction problem with applications in communication surveillance. Informs Journal on Computing, 2009,21(2):286−303. [doi: 10.1287/ijoc.1080.0292]

[99] Box GEP, Jenkins GM, Reinsel GC. Time series analysis, forecasting and control. Holden-Day, 1970.

[100] Dunlavy DM, Kolda TG, Acar E. Temporal link prediction using matrix and tensor factorizations. ACM Trans. on Knowledge Discovery from Data, 2011,5(2):10. [doi: 10.1145/1921632.1921636]

[101] Carroll JD, Chang JJ. Analysis of individual differences in multidimensional scaling via an *N*-way generalization of "Eckart-Young" decomposition. Psychometrika, 1970,35(3):283−319. [doi: 10.1007/BF02310791]

[102] Hayashi K, Hirayama J, Ishii S. Dynamic exponential family matrix factorization. In: Proc. of the 13th Pacific-Asia Conf. on Advances in Knowledge Discovery and Data Mining. 2009. 452−462. [doi: 10.1007/978-3-642-01307-2_41]

[103] Raymond R, Kashima H. Fast and scalable algorithms for semi-supervised link prediction on static and dynamic graphs. In: Proc. of the European Conf. on Machine Learning and Knowledge Discovery in Databases. 2010. 131−147. [doi: 10.1007/978-3-642-15939-8_9]

[104] Oyama S, Hayashi K, Kashima H. Cross-Temporal link prediction. In: Proc. of the 11th IEEE Int'l Conf. on Data Mining. 2011. 1188−1193. [doi: 10.1109/ICDM.2011.45]

[105] Tylenda T, Angelova R, Bedathur S. Towards time-aware link prediction in evolving social networks. In: Proc. of the 3rd Workshop on Social Network Mining and Analysis. 2009. 1−10. [doi: 10.1145/1731011.1731020]

[106] Kitano H. Computational systems biology. Nature, 2002,420(6912):206−210. [doi: 10.1038/nature01254]

[107] Ideker T, Thorsson V, Ranish JA, Christmas R, Buhler J, Eng JK, Bumgarner R, Goodlett DR, Aebersold R, Hood L. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. Science, 2001,292(5518):929−934. [doi: 10.1126/science.292.5518.929]

[108] Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabási AL. Hierarchical organization of modularity in metabolic networks. Science, 2002,297(5586):1551−1555. [doi: 10.1126/science.1073374]

[109] Rives AW, Galitski T. Modular organization of cellular networks. In: Proc. of the National Academy of Science of the United States of America, 2003,100(3):1128−1133. [doi: 10.1073/pnas.0237338100]

[110] Barabási AL, Oltvai ZN. Network biology: Understanding the cell's functional organization. Nature Reviews Genetics, 2004,5(2):101−113. [doi: 10.1038/nrg1272]

[111] Prill RJ, Iglesias PA, Levchenko A. Dynamic properties of network motifs contribute to biological network organization. PLoS Biology, 2005,3(11):e343. [doi: 10.1371/journal.pbio.0030343]

[112] Qi Y, Ge H. Modularity and dynamics of cellular networks. PLoS Computational Biology, 2006,2(12):e174. [doi: 10.1371/journal.pcbi.0020174]

[113] Przytycka TM, Singh M, Slonim DK. Toward the dynamic interactome: It's about time. Brief Bioinformatics, 2010,11(1):15−29. [doi: 10.1093/bib/bbp057]

[114] Komurov K, White M. Revealing static and dynamic modular architecture of the eukaryotic protein interaction network. Molecular Systems Biology, 2007,3(100):1−11.

[115] Park YJ, Moore C, Bader JS. Dynamic networks from hierarchical Bayesian graph clustering. PLoS ONE, 2010,5(1):e8118. [doi: 10.1371/journal.pone.0008118]

[116] Gill R, Datta S, Datta S. A statistical framework for differential network analysis from microarray data. BMC Bioinformatics, 2010, 11(95):1−10. [doi: 10.1186/1471-2105-11-95]

[117] Ay F, Dinh TN, Thai MT, Kahveci T. Finding dynamic modules of biological regulatory networks. In: Proc. of the 2010 IEEE Int'l Conf. on Bioinformatics and Bioengineering. 2010. 136−143. [doi: 10.1109/BIBE.2010.31]

[118] Kovács IA, Palotai R, Szalay MS, Csermely P. Community landscapes: An integrative approach to determine overlapping network module hierarchy, identify key nodes and predict network dynamics. PLoS ONE, 2010,5(9):e12528. [doi: 10.1371/journal.pone.0012528]

[119] Tomlins SA, Mehra R, Rhodes DR, Cao XH, Wang L, Dhanasekaran SM, Kalyana-Sundaram S, Wei JT, Rubin MA, Pienta KJ, Shah RB, Chinnaiyan AM. Integrative molecular concept modeling of prostate cancer progression. Nature Genetics, 2006,39(1):41−51. [doi: 10.1038/ng1935]

[120] Edelman EJ, Guinney J, Chi JT, Febbo PG, Mukherjee S. Modeling cancer progression via pathway dependencies. PLoS Computational Biology, 2008,4(2):e28. [doi: 10.1371/journal.pcbi.0040028]

[121] Xiong H, Choe Y. Dynamical pathway analysis. BMC Systems Biology, 2008,2(9):1−17. [doi: 10.1186/1752-0509-2-9]

© 中国科学院软件研究所 http://www.jos.org.cn

[122] Lin CC, Hsiang JT, Wu CY, Oyang YJ, Juan HF, Huang HC. Dynamic functional modules in co-expressed protein interaction networks of dilated cardiomyopathy. BMC Systems Biology, 2010,4(138):1−14.
[123] Tyson JJ, Baumann WT, Chen C, Verdugo A, Tavassoly I, Wang Y, Weiner LM, Clarke R. Dynamic modelling of oestrogen signalling and cell fate in breast cancer cells. Nature Reviews Cancer, 2011,11(7):523−532. [doi: 10.1038/nrc3081]
[124] Abeel T, Helleputte T, Peer YVd, Dupont P, Saeys Y. Robust biomarker identification for cancer diagnosis with ensemble feature selection methods. Bioinformatics, 2010,26(3):392−398. [doi: 10.1093/bioinformatics/btp630]
[125] Barabási AL, Jeong H, Néda Z, Ravasz E, Schubert A, Vicsek T. Evolution of the social network of scientific collaborations. Physica A: Statistical Mechanics and its Applications, 2002,311(3):590−614. [doi: 10.1016/S0378-4371(02)00736-7]
[126] Gupta M, Aggarwal C, Han JW, Sun YZ. Evolutionary clustering and analysis of heterogeneous information networks. IBM Research Report, RC25012(W1006-064), 2010.
[127] Rosvall M, Bergstrom CT. Mapping change in large networks. PLoS ONE, 2010,5(1):e8694. [doi: 10.1371/journal.pone.0008694]
[128] Scott CT, McCormick JB, DeRouen MC, Owen-Smith J. Democracy derived? New trajectories in pluripotent stem cell research. Cell, 2011,145:820−826. [doi: 10.1016/j.cell.2011.05.032]
[129] Lang J, Wu SF. Social network user lifetime. In: Proc. of the 2011 IEEE/ACM Int'l Conf. on Social Networks Analysis and Mining. 2011. 289−296. [doi: 10.1109/ASONAM.2011.28]
[130] http://themetricsystem.rjmetrics.com/2010/01/26/new-data-on-twitters-users-and-engagement/
[131] Wilson C, Boe B, Sala A, Puttaswamy KP,Zhao BY. User interactions in social networks and their implications. In: Proc. of the 4th ACM European Conf. on Computer Systems. 2009. 205−218. [doi: 10.1145/1519065.1519089]
[132] Viswanath B, Mislove A, Cha M, andGummadi KP. On the evolution of user interaction in Facebook. In: Proc. of the 2nd ACM Workshop on Online Social Networks. 2009. 37−42. [doi: 10.1145/1592665.1592675]
[133] Boutet A, Kim H, Yoneki E. What's in Twitter: I know what parties are popular and who you are supporting now! In: Proc. of the 2012 IEEE/ACM Int'l Conf. on Advances in Social Networks Analysis and Mining. 2012. 132−139. [doi: 10.1007/s13278-013-0120-1]

**附中文参考文献:**

[57] 单波,姜守旭,张硕,高宏,李建中.IC:动态社会关系网络社区结构的增量识别算法.软件学报,2009,20:184−192. http://www.jos.org.cn/1000-9825/09022.htm

**高琳**(1964－),女,陕西乾县人,博士,教授,博士生导师,主要研究领域为数据挖掘,计算生物信息学,图论与组合优化算法及其应用.
E-mail: lgao@mail.xidian.edu.cn

**杨建业**(1987－),男,硕士生,主要研究领域为复杂网络模式挖掘.
E-mail: yangjianye0@163.com

**覃桂敏**(1977－),女,博士,副教授,主要研究领域为计算生物信息学,数据挖掘.
E-mail: gmqin@mail.xidian.edu.cn

© 中国科学院软件研究所 http://www.jos.org.cn