动态复杂网络社团结构演变特征构建及建模研究

A study on dynamic network community detection modelling and evolution feature analysis

工程领域：软件学院
作者姓名：李天鹏
指导教师：李杰

天津大学智能与计算学部
二零一九年十一月

独创性声明

本人声明所呈交的学位论文是本人在导师指导下进行的研究工作和取得的研究成果，除了文中特别加以标注和致谢之处外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得 **天津大学** 或其他教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

学位论文作者签名:　　签字日期:　　年　　月　　日

学位论文版权使用授权书

本学位论文作者完全了解 **天津大学** 有关保留、使用学位论文的规定。特授权 **天津大学** 可以将学位论文的全部或部分内容编入有关数据库进行检索，并采用影印、缩印或扫描等复制手段保存、汇编以供查阅和借阅。同意学校向国家有关部门或机构送交论文的复印件和磁盘。

(保密的学位论文在解密后适用本授权说明)

学位论文作者签名:　　导师签名:

签字日期:　　年　　月　　日　签字日期:　　年　　月　　日

摘　　要

互联网+时代的来临给城市风险计算带来了全新的挑战，需要融合社会、物理、网络空间的多维大规模数据进行高效精准的风险感知、理解、预测。从多维城市数据中高效准确的检测出符合特定模式的社团及其演化模式是城市风险计算的基石。本课题从演化的角度，针对动态复杂网络社团检测及演化分析以及从社团到风险的挖掘模式进行了探索，具体工作如下：

首先，通过对真实复杂网络数据的分析处理，找到节点的结构属性与社团演化的规律。通过巧妙的设计，我们将社团内的节点是否在下一时刻发生转移视作二分类问题，用利用特征工程提取每个节点的结构属性作为分类特征，利用决策树对社交媒体数据、论文引用数据以及其他关系型数据的节点的社团转移进行分类，并分析其特征重要性。

其次，通过融合节点演化特征来构建社团检测模型，增强社团演化的准确性。融合节点级别的社团转移趋势以及社团级别的社团转移趋势，通过构建从社团级别转移矩阵到节点级别社团转移矩阵的层次贝叶斯结构，结合动态网络概率生成模型构建了层次贝叶斯动态随机块模型，并利用变分推断对模型进行参数估计。

最后，利用手机信令数据对以上的规律以及模型在城市风险计算中的有效性进行实证分析。利用手机信令数据验证节点的度以及节点平均邻居度对节点社团归属是否发生转移找到现实的样例进行分析。同时对层次贝叶斯动态随机块模型进行有效性分析，利用对真实网络的处理，提取出真实网络中的节点社团归属以及社团演化信息，接着通过社团信息对现实世界的事件进行提取并针对真实数据潜在风险进行分析并进行可视化。

以上工作的结果表明，节点的度以及平均邻居度是节点进行社团转移的主要结构属性，即社交网络中，用户在不同圈子间进行转移的普遍基础就是具有较多朋友，或者朋友的交友涉猎广泛。同时提出的层次贝叶斯动态随机块模型能够有效地提取出动态网络中的社团划分以及社团演化信息，并能够利用真实网络数据进行潜在风险分析。

关键词： 复杂网络分析；动态社团检测；社团演化；城市风险计算；

ABSTRACT

The advent of the Internet+ era has brought new challenges to urban risk computing. It needs to integrate multi-dimensional and large-scale data of social, physical and cyberspace for efficient and accurate risk perception, understanding and prediction. Efficient and accurate detection of associations and their evolution patterns from multi-dimensional city data is the cornerstone of urban risk calculation. From the perspective of evolution, this topic explores the detection and evolution analysis of dynamic complex network communities and the mining model from community to risk. The specific work is as follows:

Firstly, through the analysis and processing of real complex network data, the structural properties of nodes and the laws of community evolution are found. Through clever design, we regard whether the nodes in the community are transferred at the next moment as a two-category problem, using the feature engineering to extract the structural attributes of each node as the classification feature, and using the decision tree for social media data, paper reference data, and The community transfer of nodes of other relational data is classified and analyzed for its characteristic importance.

Secondly, the community detection model is constructed by integrating the evolution characteristics of nodes, and the accuracy of community evolution is enhanced. Combining the node-level community transfer trend and the community-level community transfer trend, constructing the hierarchical Bayesian dynamic random block by constructing the hierarchical Bayesian structure from the community-level transfer matrix to the node-level community transfer matrix, combined with the dynamic network probability generation model. The model is used to estimate the parameters of the model using variational inference.

Finally, the mobile phone signaling data is used to empirically analyze the above rules and the validity of the model in urban risk calculation. The mobile phone signaling data is used to verify the degree of the node and the average neighbor degree of the node to analyze whether the node community belongs to the transfer and find a reality. At the same time, the effectiveness analysis of the hierarchical Bayesian dynamic random block model is carried out. The processing of the real network is used to extract the node community membership and community evolution information in the real network, and

then the real world events are extracted and targeted by the community information. Data potential risks are analyzed and visualized.

The results of the above work show that the degree of nodes and the average neighbor degree are the main structural attributes of nodes for community transfer. In social networks, the common basis for users to transfer between different circles is to have more friends, or friends with friends. . At the same time, the hierarchical Bayesian dynamic random block model can effectively extract the community division and community evolution information in the dynamic network, and can use the real network data for potential risk analysis.

KEY WORDS： Complex network analysis, dynamic community detection,community evolution,city risk calucalation

目　录

第1章 绪论

随着互联网的蓬勃发展，复杂网络分析在实际应用特别是风险计算中的地位越来越高。尤其是动态网络的相关算法如动态网络社团检测，对于检测时序关联性数据即动态网络数据中突发事件的作用非常大。本章将会介绍动态复杂网络社团检测以及社团演化的研究背景及意义以及风险计算的研究背景、应用场景等。

1.1 研究背景及意义

随着以互联网为代表的网络信息技术的迅速发展，复杂网络对人类社会的影响越来越大。从万维网到社交网络，从病毒传播到交通流的管控分析，都能够从中提取出复杂网络结构[1]，进而通过对相应复杂网络的分析来获得现实网络的信息，进而达到改善用户体验、遏制病毒传播以及减少交通拥堵等目的。

复杂网络分析（complex network analysis）来源于图论，至今已有超过200年的历史，然而自从1998年nature上提出了复杂网络的“小世界”[2]属性以及1999年science上提出了复杂网络的“无标度”[3]属性后，复杂网络分析才进入了飞速发展的时期。

作为复杂网络分析的重要任务之一，社团检测始终吸引了大量科研人员的注意力[4]。社团最广泛的定义就是—-其作为网络中的子图，子图内部的节点连边密度多于子图与子图外其他节点的连边密度。例如，在万维网中，某个局域网内的服务器或pc之间基本是全连通的，而改局域网与外部仅仅通过一个路由器进行数据交换；再比如社交网络中的社团则代表了某个小团体，比如某兴趣团体等；而对于论文引用网络来说，社团结构则可以代表不同的研究领域等。社团检测则利用不同的理论去对网络中的节点进行聚类，以高效准确的找出不同网络中的社团结构。在现实世界中，人们往往利用动态网络对现实世界数据进行建模，并利用动态复杂网络分析的相关方法，对其进行分析。其中动态社团检测就是动态网络分析的重要的任务之一，而动态社团检测涉及两个主要任务，每个时间快照上的社团检测以及连续时间快照的社团演化分析。

动态社团检测能够帮助风险计算如城市风险事件的相关人员以及事件规模进行检测，而社团演化应用在风险计算如城市风险计算中所对应的就是风险事件的发生发展以及结束的整个过程的演化行为的计算。利用复杂网络建模，对

城市多维动态数据进行合理的融合，再利用动态社团检测方法对风险事件进行计算以及对其发生发展进行合理推理，进而可以帮助人们把控风险因素以及风险事件。

1.2 研究现状

1.2.1 复杂网络社团检测现状

目前已经有大量的研究和方法聚焦于复杂网络社团检测任务，例如基于模块度的方法[5]、基于谱聚类的方法[6]以及生成模型[7]。在生成模型中，应用最广泛且发展最迅速的模型之一就是随机块模型（Stochastic Block Model）。随机块模型提出三个假设：（1）网络中每个节点属于一个社团；（2）网络中的两个节点之间有没有边只与节点所在的社团有关。（3）网络中的社团个数固定。经过多年的改进，随机块模型已经被应用在多个领域，例如生物信息学和社会科学等[8,9]。研究者对随机块模型进行了不同方向的扩展，例如Pal[10]等人提出了混合随机块模型，为每个节点分配了一个长度为社团个数的向量来衡量每个节点属于每个社团的概率或称为隶属度，从而使每个节点可以属于多个社团；Kemp[11]等人提出了随机块模型的改进版模型，该模型允许网络可以存在无限数量的社团；而Hofman[12]等人则利用贝叶斯推断来推测随机块模型的最优解以及最优社团个数；Chen, Yudong[13]等人利用了节点的度来修正不同社团节点之间连边的概率，进而来修正随机块模型社团内部节点都是等价的这种不合理假设；类似的，Qiao, Maoying[14]等人也引入了度衰减参数来修正这种不合理假设，同时还使得社团内部的节点在该模型中服从power law。

然而上述模型以及方法均面向的是静态网络，也就是网络中的节点与边不具有时间属性，不随时间变化。而真实世界的网络往往是动态的。动态网络社团检测的研究，也是近几年研究者们所重点关注的。动态网络的数据不同于静态网络，其数据引入了时间属性。同时动态网络的数据也具有不同的格式，对于动态网络数据有两种数据格式，第一种是将连续时间的网络数据切分成等时间间隔的静态网络再进行处理；第二种是将网络数据写成三元组的形式，即源点，汇点，持续时间来表示一条边的生命周期。第一种形式便于处理，且可以有效利用静态网络的算法拓展到动态网络中，这种格式数据的缺陷是对于相邻网络切片的时间间隔不好把控；而第二种形式相比于第一种形式更加连续，更能保持动态网络的变化，然而这种格式的数据不易于处理，并且较难借鉴静态网络算法。综合考虑，我们的工作主要针对第一种数据格式，即对网络进行切片处理后进行数据分析。对于动态网络社团检测任务，主要包括两方面，第一

方面是每个时间片的社团检测；第二个方面是不同时间片的社团演化[4,15]。动态网络社团检测的方法多种多样，包括两步法、进化聚类法以及生成模型等。两步法就是对于每个时间片运行静态网络社团检测的算法，然后再在相邻时间片进行社团匹配[16]。这种方法完全割裂了社团检测和社团演化，对于网络噪声非常敏感，因此在真实网络中效果不好。进化聚类认为当前时间的网络社团结构与前一个时间片或者前几个时间片的社团有关，因此在计算当前时间片的社团结构时，进化聚类会将上一个时间片的社团检测结果输入到算法中来提高社团检测效果[17]。而生成模型则认为当前时刻的社团结构是由一个特定的与整个网络有关的分布生成的，因此生成模型会利用已知信息对网络进行建模，通过对模型的推断求解得出当前所有时间片的社团结构[18]。

Yang等人[18]在2011年将随机块模型应用到动态网络中，提出了DSBM即动态随机块模型。该模型引入了节点的社团转移矩阵，通过该矩阵来把握社团的演化行为，并通过采样对模型进行求解。该方法初步将社团检测和社团演化融合到了一起，但是对于社团演化的把握粒度较粗，同时依然需要提前确定社团个数并且不能处理重叠社团问题，并且其依然遵循社团内部节点地位等同的假设。Tang, Xuning[19]提出了DBTDP，利用狄利克雷过程进行模型选择，从而解决了社团个数必须提前确定的不合理假设。而Wu,xunxun[20]等人将节点的度衰减参数引入了动态网络，并利用变分推断对模型进行求解，使得模型的节点度符合真实网络的power law。Kao, Edward K[21]等人将混合随机块模型拓展到了动态网络，打破了动态随机块模型中每个几点只能属于一个社团的假设。 以上方

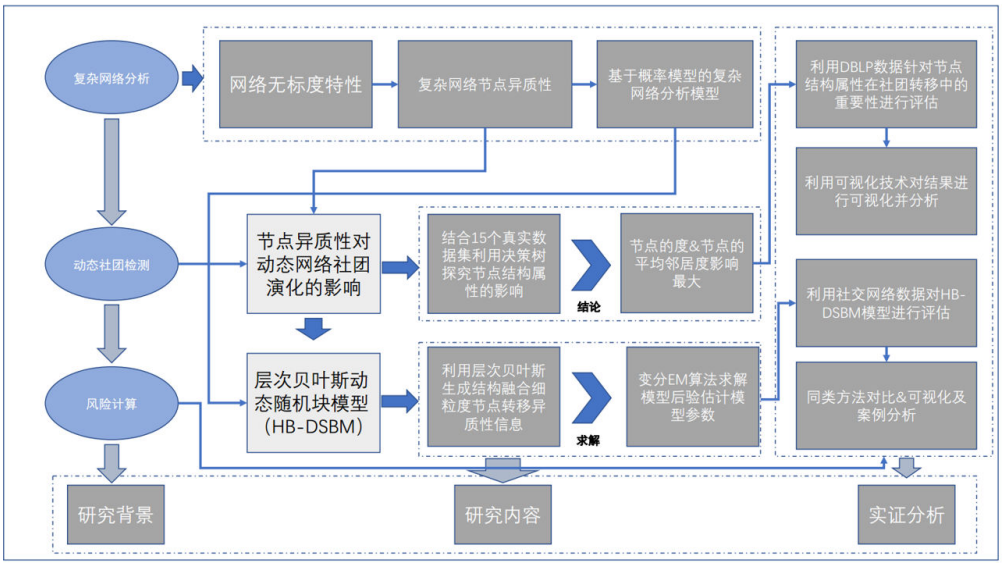


图 1-1　　研究框架图

法均重点关注动态网络社团检测的精度以及效率，而对于社团演化的关注度并不高。然而社团演化对于动态网络社团检测至关重要，同时社团演化也是一个

重要的课题。Yang, Jinfeng[22]等人利用每个社团内的关键节点来匹配相邻社团，进而来把握社团的演化。然而核心节点未必能够代表社团。Gergely Palla[23]等人提出社团的出生、消亡、分裂、合并、增长、缩小六种社团行为，被其他研究者广泛借鉴。社团的演化本质上是由节点之间的连边变化驱动的，因此社团的演化行为与社团内部节点的社团转移之间的关系对于社团演化来说至关重要，然而据我们所知目前还没有对这种关系深入的探究。同时，社团的演化行为也是由社团内的节点的变化而驱动的，目前也没有相关的文献进行详细探究。因此，如图1-1所示，以上两个问题就是本课题将要探究的目标，即探究节点异质性对动态网络社团演化的影响以及利用细粒度演化参数构建动态社团检测模型。

1.2.2 城市风险计算现状

城市风险包含的范围非常广，包括自然灾害城市风险（如地震与台风等）、公共事件类城市风险（如踩踏事件与群体恐慌等）、社会安全城市风险（如核泄漏、恐怖袭击、抢劫枪杀等）、公共卫生安全城市风险（如非典、禽流感与艾滋病等）等。因此对于城市风险来说其风险源广，风险数据量大，其综合识别与整体的管控与治理非常困难。然而针对不同的城市风险需求进行有针对化的计算与治理则难度相对较小，同时其某些特定风险计算需求与网络中的相关算法非常契合，例如网络算法中的关键节点识别就可以用来识别城市风险中的风险点，而风险事件识别则对应了动态网络社团检测。例如社团检测算法可以高效的检测出城市中的不同粒度的团体，再结合社团演化分析就可以检测出不同的事件，通过有效地设置损失函数，就可以对事件风险进行打分并排序，确定不同事件的风险大小。

城市风险计算涉及的数据范围非常广，如城市风险计算包括人类行动轨迹、车辆轨迹、人类电子足迹、社交网络数据、转账数据、城市OD流数据等等，均来自于风险监测模块的多元数据收集。收集后的数据维度广，数据量大，并不能直接进行风险计算，因而需要对数据进行融合，即信息融合。通过多维信息的多层级融合，将数据建模成复杂网络形式数据，进行风险计算。根据不同的风险需求，风险计算可以灵活选择不同的网络算法，如风险因素识别可以利用复杂网络中的关键节点识别算法进行提取，而风险事件则可以利用动态社团检测算法进行计算。

1.2.3 主要创新点

本文的主要创新点包括以下三个方面：

1. 发现了动态网络中对节点社团转移影响最大的两个结构属性；
2. 构建了节点粒度级别演化参数的动态网络社团检测生成模型并为模型求

解提出了有效的变分近似算法；

3. 成功探索了复杂网络社团检测与风险计算的紧密关联并进行了验证。

三个创新点具有紧密的内在关联性，本文通过十五个真实世界复杂网络数据集探索了动态网络中节点的社团转移与节点及社团的结构属性之间的关系，发现了影响节点社团转移的两个结构属性：节点的度以及节点的平均邻居度。这两个结构属性均为节点粒度的结构属性，而社团的属性在节点的社团转移中起到的作用不大，说明节点粒度的社团转移参数在模型中是有必要的。受上述的启发，本文提出了节点粒度的动态网络社团检测模型，并利用变分推断近似求解，使模型适用于真实世界复杂网络。而城市风险计算在城市管理中至关重要，而传统的机器学习算法无法处理具有复杂关联的城市多元数据。依赖于复杂网络及复杂网络社团检测算法，对于城市多元关联性数据，本文提出的算法可以有效的进行处理，同时基于本文算法的处理结果，利用城市风险计算的相关算法可以有效的计算不同事件的风险程度，为城市风险计算与复杂网络社团检测搭建了桥梁。

1.3 章节安排

根据本文的研究框架，文章的章节结构安排如下：

第一章，绪论，介绍复杂网络分析中，社团检测在风险计算中的重要作用以及风险计算的相关流程，研究背景及意义、研究内容框架、主要创新点等。

第二章，相关研究，介绍动态网络社团检测以及风险计算的相关研究，同时介绍本文的统一符号表示等。

第三章，节点结构属性对社团演化影响因素的探究，本文会在真实的网络数据中具体探究动态网络社团检测中，节点的哪些结构属性对社团演化影响较大。

第四章，融合节点级别社团转移参数的动态网络生成模型，本文将会针对经典的动态网络社团检测概率模型——动态随机块模型，融合更细粒度的演化参数，利用层次贝叶斯生成结构构建对社团演化掌握更准确的新模型：层次贝叶斯动态随机块模型，并在接下来介绍针对层次贝叶斯动态随机块模型的变分推断近似解法。

第五章，案例分析，本章会利用手机信令数据结合动态复杂网络社团检测算法以及风险计算的相关理论知识及算法进行实际的风险计算，验证复杂网络在风险计算中的重要作用。

第六章，总结与展望，本章将总结本课题的工作，并对未来工作进行说明与展望。

第2章 相关研究

第一章简单介绍了复杂网络社团检测的研究现状，包括静态网络社团检测、动态网络社团检测以及动态网络社团演化的研究现状，并简单说明了以上研究的区别以及关联。同时第一章还介绍了城市风险计算的研究现状，以及社团检测在城市风险计算中的重要作用。本章则从具体的方法论角度介绍动态网络社团检测以及社团演化的研究现状与趋势，包括增量聚类、进化聚类和基于模型的聚类方法。为了验证动态社团检测方法的有效性，本章进一步介绍了相关的社团检测评价指标，包含归一化互信息、均方误差和模块度的定义。接着，本章进一步详细介绍城市风险计算的相关研究以及动态社团检测在风险计算中的作用。

2.1 动态网络社团检测

动态网络社团检测本质上包含了静态复杂网络的社团检测(第一个网络快照上)、动态网络社团检测和社团演化分析三个子问题。对于第一个问题可以借用静态网络社团检测的方法完成，而第二个问题则是目前研究者关注的重点，同时第三个问题–社团演化分析则有助于理解社团的演化模式和动态复杂网络的变化规律。如图2-1所示，图中左右两张图是论文引用数据(DBLP数据)在相邻两个时间快照上的社团可视化图，左侧为前一个时间快照，右侧为后一个时间快照，而不同的颜色代表论文所属的研究领域不同，可以看到，下一时刻的社团结构与上一时刻的社团结构变化很大，只有同时融合社团演化规律与动态社团检测方法才能有效的针对动态复杂网络数据进行高精度的社团检测。

目前动态网络社团检测的主流方法将侧重点放在了第二个问题上，根据目前主流的动态复杂网络社团检测方法的本质不同，本文将这些主流方法做出了分类，如表2-1所示,增量聚类方法根据不同相邻快照之间的节点和边的变化定义不同的目标函数，增量更新社团信息；进化聚类方法结合了当前网络快照之前的一个或几个快照的信息，来进行当前网络快照的社团检测；而生成模型方法则针对网络以及社团的生成机制进行建模并进行推断，并将动态网络的社团检测问题转化为模型的参数估计问题，从而进行社团检测。

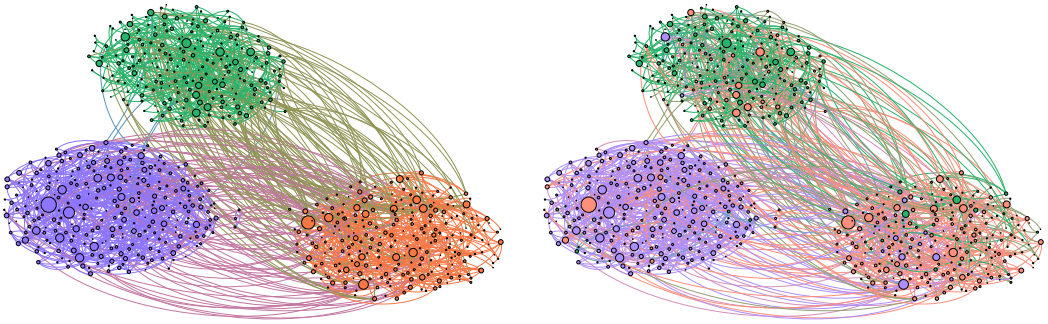


图 2-1　论文引用数据(DBLP)的社团及社团演化示意图

表 2-1　主流动态社团检测方法

动态社团检测	核心思想	方法优势	方法不足	代表性工作
增量聚类方法	将动态网络快照之间社团的变化转化为节点和边的变化，通过定义不同目标函数，增量更新节点的社团归属	复杂度低、对于局部结构变化较多的网络处理更加精准	过于依赖模型定义的目标函数，对网络中的噪声敏感	GraphScope[24] Graph-Tinker[25] TILES[26] 模块度优化[27]
进化聚类方法	融合动态网络当前快照之前的一个或几个快照社团结构信息，保持社团关系变化的平滑性	融合了历史快照信息，社团检测结果较好	进行大量的网络信息重复计算，复杂度较高	Genlouvain[28] PisCES[29] DYN-MOGA[30] 进化谱聚类[31]
生成模型方法	基于隐马尔科夫模型假设，利用生成模型对动态网络进行建模，将社团检测问题转化为参数估计问题	具有严格的理论解释和意义、模型可用于多种网络分析任务	模型优化较困难，同时概率模型的参数较多，参数域较大，因此复杂度较高	动态隐空间模型[32,33]，动态随机块模型[18,20,34]

2.1.1 增量聚类方法

增量聚类的主要思想是，首先，在动态网络的第一个快照上执行静态社团检测算法，然后根据网络在接下来的快照中的节点和边的变化调整节点的社团归属。Sun等人提出了GraphScope[24]方法，GraphScope避免了额外的参数设置(如社团个数、节点的社团转移阈值等)，利用信息论的最短描述长度来判别节点的社团变化，并通过处理流数据提升了算法的计算效率。GraphScope在二分网络上具有很好的效果。GraphTinker[25]通过定义多种不同的哈希策略来提升传统增量聚类的运行效率和精度，较传统增量聚类运行效率有两倍以上的提升。TILES[26]利用标签传播来刻画节点在网络中的局部变化，以此来减少计算节点在不同时间片上的社团变化的复杂度，同时，标签传播方法还一定程度上避免了单纯的定义目标函数造成的错误累积问题，于此同时，TILES还通过定义不同的触发机制来判定社团的分裂以及合并等演化行为。而IDCD[27]方法则通过对第一个时间片的网络对边进行重要性排序，然后通过最大化模块度来检测第一个时间片的社团结构，随后根据后续时间片节点的Jaccard相似度对相应节点的边进行排序，随后按照重要性降序对边关联的节点进行社团划分，该算法的复杂度很低，但是在真实网络中效果并不好。

总的来说，增量聚类在动态网络社团检测中具有效率高，易由静态网络社团检测方法进行迁移等优点，但是其缺点也很明显，由于增量聚类依赖于定义的目标函数，因此存在错误率累积的问题，这使得该方法对数据的噪声过于敏感，因此大部分增量聚类的方法在真实数据中的效果都不是很好。

2.1.2 进化聚类方法

进化聚类方法的主要思想是，该类方法认为动态网络在相邻的时间片时不会发生突变的，因此动态网络的社团结构在相邻时间片或某足够小的时间窗口内的变动是平滑的，所以这类方法在检测某网络快照的社团时，会融合当前快照之前的社团信息，以此来保持节点社团变化的平滑性。Chakrabarti等人首次提出了进化聚类[35]，并定义了进化聚类的框架。其本质可表示为

$$Loss = \alpha \times HL + (1 - \alpha) \times PL \tag{2-1}$$

其中Loss表示进化聚类目标函数的整体损失，HC表示当前快照的社团检测结果与上一快照社团检测结果变化的损失，而PC则为当前网络快照的社团检测目标函数的损失，α则代表平衡因子，用来控制历史快照对当前快照社团检测的影响力。

Genlouvain[28]继承自Mucha等人2010年在Science上提出的广义网络质量函

数框架[36]，模型通过定义动态网络连续时间片的*null model*，从而得到了动态网络的模块度的定义$Q_{multislice}$。利用动态网络模块度优化，进而得到动态网络的社团划分结果。Folino等人则提出了DYNMOGA[30]，DYNMOGA方法基于多目标优化将进化聚类目标函数的两个部分建模为多目标优化问题，并利用遗传算法对模型进行求解。Liu等人提出了PisCES[29]方法，通过特征向量平滑的方法约束动态网络的社团在相邻时间片变动尽可能变小，同时考虑到同社团内节点的异质性，PisCES融入了节点的度使其更能适应真实世界网络。 Huang等人[31]则结合了Fiedler特征向量与一定时间窗口的正则化拉普拉斯矩阵，进一步用谱聚类的方法进行社团检测。

总的来说，通过融合了动态网络历史社团信息，进化聚类的效果要优于增量聚类方法，但是由于进行社团计算时需要引入更多信息，其复杂度要高于增量聚类。同时，由于参数α的存在，使得模型需要进行额外的调参，不利于实际使用。

2.1.3 生成模型方法

生成模型方法认为节点的社团转移服从隐马尔科夫假设，通过对复杂网络进行概率建模，来构建网络的生成机制，进而将动态网络的社团检测转化为概率模型的参数估计问题。这部分的方法分类两个大类，分别为隐空间模型和动态随机块模型。

隐空间模型认为，节点的社团划分可以通过将节点在网络中的结构映射到多维的欧式空间中使其变得可分，进而通过利用传统机器学习聚类或者分类算法可以对其进行划分。Daniel K. Sewell等人[32]提出了动态网络有向和无向两个隐空间模型，定义了不同的距离函数，利用MCMC采样对模型进行求解。Yang等人[33]针对非负矩阵分解以及谱聚类等利用矩阵计算的隐空间模型提出了一个新的惩罚项，用以提升这些方法通过半监督学习对不完全数据进行社团检测的效果。

而**动态随机块模型**则继承自随机块模型，有Yang等人[18]于2011年通过加入社团转移矩阵A将随机块模型扩展到了动态网络。动态随机块模型认为动态网络中的任意两个节点有边的概率只与节点所在的社团有关，一个节点在一个时间片只能属于一个社团，同时在同一个社团的两个节点在社团检测与社团演化行为上是完全等价的。基于以上假设，得以构建动态随机块模型DSBM，模型通过吉布斯采样结合模拟退火算法进行参数估计，进而得到每个节点在每个时间片上的社团归属。DSBM在进行参数估计之前需要提前指定社团个数，这在真实数据进行社团检测时是不现实的，因此Tang等人[37]提出了DBTDP，不用社团转移矩阵决定节点的社团转移，而是通过狄利克雷过程刻画节点的转移，利用

中国餐馆过程对模型进行模型选择，从而实现了模型自动确定社团个数。动态随机块模型的另一个假设，即一个节点在一个时间片只能属于一个社团，在某些真实场景中也是不合理的，比如大学里每个人可以参与多个社团，或者一个人可以有多种兴趣等等。针对这种场景，Xu[38]等人扩展了著名的混合随机块模型MMSB，构建了动态混合随机块模型DMMSB。而Yu[39]等人提出了融合节点变化倾向的动态度修正随机块模型，解决了动态随机块模型第三个假设的缺陷，即同一个社团内的节点在社团转移中是完全等价的，模型融入了节点的度来修正不同节点之间连边的异质性；而对于同一社团内节点异质性的把握，Xunxun Wu等人[20]提出了DPSBM，给出了不同的思路，通过在动态随机块模型中引入了度衰减参数，并假设该参数随时间变化服从半正太分布而演化，从而刻画出了动态网络中节点度分布服从power law的无标度属性。

总体来说，生成模型的方法通过对网络的生成机制进行建模，从而探究了网络生成的本质，使其根据有可解释性，然而由于是对整个网络进行建模，因此其参数数量较增量聚类和进化聚类更大，其参数的解空间更大。同时由于其模型结构的复杂性，使得对模型的参数估计多为MCMC对解空间进行全局搜索，因此其效果更好的同时，复杂度也比前两种方法更高。

2.2 动态网络社团演化

针对动态网络社团检测的第三个子问题，即动态网络相邻时间快照的节点社团演化分析，也一直有人在关注，但是目前还处于起步阶段，对节点随着时间发生社团转移的本质依然没有统一的定论。Dakiche等人[15]就社团演化追踪进行了总结，本小结参考他们的总结并结合实际将动态网络社团演化划分为独立社团演化、非独立社团演化以及同步社团演化进行三大类，如表 2-2所示。可以看到表中的部分方法与上一小节有重叠，因为动态网络社团检测的不同方法侧重点不同，但是对社团检测以及社团演化一定都有涉及。

2.2.1 独立社团演化

独立社团演化将动态网络的社团检测与社团演化分别进行独立的计算，在每个时间片执行静态社团检测方法，随后在相邻时间片处理两个社团集合的匹配问题。社团匹配的方法大多基于相似性指标进行匹配，同时这类方法还定义了社团的演化事件，通过确定社团的演化事件来追踪社团的演化。Asur等人[40]定义了社团的五个可能事件，分别为社团的消失、形成、延续、分裂和合并。而Palla等人[23]则提出了社团的六种演化事件，分别为增长、缩小、出生、消亡、分裂和合并。在此之后，Brodka等人[41]则融合了前两种事件，提出了社

表 2-2　动态社团演化方法

动态社团演化	核心思想	方法优势	方法不足	代表性工作
独立社团演化	即熟知的两步法，在每个时间片单独进行社团检测，随后对相邻时间片进行社团匹配	容易从发展较好的静态网络社团检测算法进行扩展，可以独立定义社团演化机制	过于依赖定义的社团演化机制，对社团检测效果没有帮助	基于社团演化事件的方法[23,40,41]
非独立社团演化	假设社团在相邻时间片上不会发生突变，在进行社团检测时定义相应的惩罚项约束社团的演化	社团演化由惩罚函数进行控制，可以加强社团检测效果	进行大量的网络信息重复计算，复杂度较高，无法处理网络的突变	进化聚类[28–30] 关键节点追踪[42]
同步社团演化	基于隐马尔科夫模型假设，利用生成模型对动态网络进行建模，同时建模社团结构以及其演化规则	具有严格的理论解释和意义、模型可用于多种网络分析任务，社团检测结果与社团演化相互增强	模型优化较困难，同时概率模型的参数较多，参数域较大，因此复杂度较高	DSBM[18] 涂色优化法[43] DPSBM[20]

团增长、缩小、延续、出生、消亡、分裂和合并七种演化事件。对于这些事件的检测，以上三组团队均选择在相邻两个时间片进行检测，而Tajeuna等人[16]则选择在整个动态网络上进行社团演化事件检测。

同时对于相邻时间片的社团匹配，也有不同的社团相似性指标，如Greene等人[44]利用了Jaccard系数计算相邻时间片社团的相似性；而Brodka等人[41]则提出了inclusion度量来计算相邻时间片的社团匹配度；于此同时，Tajeuna等人[16]定义了名为互转换指标的另一种相似性度量。

独立社团演化可以自由定义社团演化追踪的方法方式，对社团演化事件的追踪也更加的细致。但是社团演化行为是建立在每个事件片上社团检测足够准确的基础上的，割裂了社团演化和社团检测，这类方法的动态社团检测结果往往在真实数据集中表现不好，因为真实世界的动态网络往往具有较高的噪声，同时这类方法也考虑不到社团演化对每个时间片社团的影响。

2.2.2　非独立社团演化

非独立社团演化将社团演化融入到相邻的时间片，以此来增强社团检测效果。这部分包括上一小节介绍的进化聚类的相关方法，如Genlouvain[28]PisCES[29]DYNMOGA[30]等。同时，非独立社团演化还包括了通过追踪关键节点的方法来追踪社团演化，如Gao等人[42]定义了每个社团的领导

节点和跟随者节点，并假设每个社团的演化行为是由领导节点的行为导致的，这使得社团的追踪演化为了对领导节点的追踪。

这类方法在相邻社团融合了社团检测和社团演化，对社团演化的追踪也有一定独到之处，但是这种方法由于其前提假设使得其无法应对网络突变，也并不能很好的把握社团演化的本质。

2.2.3 同步社团演化

同步社团演化通过对网络的建模，同时构建了动态网络社团的产生和社团的演化。如DSBM[18]通过社团转移矩阵A来刻画动态网络中节点的社团转移，通过对A的估计，即可得到节点在相邻时间片的演化倾向。然而其社团转移矩阵A并不随时间变化，因此并不能刻画节点演化倾向的变化。DPSBM[20]则利用其构建的节点度衰减参数δ来修正不同时间片不同节点社团转移的异质性，从而利用δ刻画整个网络节点的社团转移，进而把握住社团的演化。除了以上方法，同步社团演化也存在一些启发式优化方法，如Tantipathananandh等人[43]将动态网络社团检测建模为连续的图染色问题，通过求解染色问题来刻画每个时间片的社团以及社团的演化。然而该方法是NP的时间复杂度，虽然作者利用一些启发式方法对求解过程进行了优化，其复杂度依然较高。

同步社团演化结合了社团检测和社团演化，对社团检测与社团演化的效果都比较好，但是由于大部分方法建模的侧重点放在了社团检测，对于社团演化事件以及演化行为的更细节的把控程度以及社团演化的本质探究并不深入。同时概率模型的复杂性使得其求解较困难。

2.3 城市风险计算

针对城市风险的研究由来已久，早在2007年，吴竹就在政法学刊上发表了《群体性事件预警指标体系研究》[45]，文章认为，群体性事件预警是通过对社会系统中的不良因素或者负面因子的检测和评估而形成的，因此吴竹在文章中构建了涉及城市群体性事件预警指标体系，包括了六个子系统下的67个具体指标，由于这些指标涉及层面太过广，计算难度很大。而随着互联网的发展，网络空间也成为了城市风险的重要因素之一。王连强[46]在2006年提出了信息安全风险的评估方法框架，框架包括了信息安全风险评估方法ISRAM，定义了信息安全风险评估的计算框架以及平台实现；同时还包括了安全风险评估因素的分类研究等。然而其定义的信息安全风险评估计算涉及的变量及因素较多且框架并没有形成完整的系统，可适用性并不好。同时该框架仅仅涉及信息安全风险，并不与城市风险形成关联。龚俭等人[47]2017在软件学报发表了《网络安全态势感

知综述》，该综述中提出了网络安全态势感知框架，该框架的态势察觉-态势理解-态势评估模块不仅在网络安全态势感知中获得大部分人的认同，同时给物理空间安全态势感知也提供了有效的参考。文章也同时提出了利用聚类的方法将相似性的警报进行聚类同一整合，减少系统警报数量。

而随着计算机算力的指数性提升，利用计算机结合有效的算法对城市风险进行计算是目前城市风险计算的显而易见的发展方向。然而由于城市数据的维度广，不同层面的数据之间交互复杂，单纯的机器学习算法由于其对数据单元的i.i.d假设导致其并不能准确的计算城市风险。另一方面，城市风险计算方式通过传统的群体性事件预警或者评估方法需要的资源以及计算量非常大，并不能做到有效且实时。因此通过复杂网络对城市数据进行建模，进而利用复杂网络中的相关算法对城市风险进行计算评估是合理且最行之有效的。

利用复杂网络对城市系统进行建模的研究由来已久，如利用复杂网络分析台风对城市的影响[48]、利用复杂网络对电力系统进行建模并分析其脆弱性[49]、利用复杂网络建模城市道路网络，并分析其脆弱性[50]等等。由此可见，复杂网络的脆弱性计算[51]应用到城市系统中，可以在很大程度上发现城市的潜在风险。

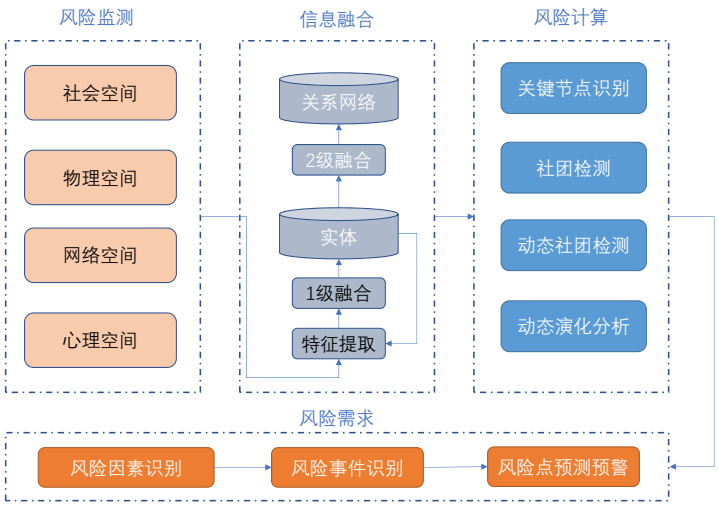


图 2-2　　风险计算框架图

本文的风险计算研究框架如图2-2所示，通过整合多渠道的城市风险计算流程总结整合而成，首先监测多元空间的数据，并通过信息融合将数据进行预处理、数据对齐等步骤将潜在风险的数据融合处理为实体，进一步经过二级融合

将数据构建成复杂网络。根据不同的风险需求，复杂网络算法会在其中起到各不相同的作用。

而动态网络社团检测则能够处理风险事件识别或风险因素识别的风险需求，通过对风险网络中的节点进行动态社团检测，并进一步进行事件提取，随后利用相关的打分算法对不同事件进行打分，随后筛选出潜在的风险事件[52]。也可以利用社团演化分析，判断团体的发展走向，结合网络脆弱性指标或者节点重要性指标，综合分析城市中某些团体的潜在威胁。

2.4 本章小结

本章首先介绍了复杂网络分析的发展历程及其主要任务，随后聚焦于复杂网络分析中的社团检测的相关研究。接着针对动态网络社团检测的目前研究现状以及存在的主要问题进行了阐释说明。最后，本章介绍了风险计算的发展现状与城市风险计算和社团检测的紧密联系进行了说明。

第3章 节点结构属性对社团演化影响因素的探究

上一章介绍了动态复杂网络社团检测的相关研究以及城市风险计算的最新进展。上一章提到了社团演化对动态复杂网络社团检测的影响非常大，而目前的大部分方法都聚焦于动态复杂网络的每个时间快照上的社团检测效果而或多或少地忽略了社团演化的重要性。更进一步，本文认为节点的结构属性特别是局部结构对节点社团演化的影响是很大的，因为社团的演化行为是由节点的社团转移行为构成的，也就是说，节点的社团转移行为才是社团演化的驱动力量。因此针对节点的结构属性对节点社团转移的影响的探究是必要且重要的。这也是本章的研究重点，即对节点的结构属性对节点社团转移的影响的探究。本章首先介绍对探究节点结构属性对节点社团转移影响的方法，然后介绍使用的真实世界复杂网络数据集，随后介绍经过第一部分介绍的方法结合第二部分真实数据集分析后得到的结论。

3.1 探究方法

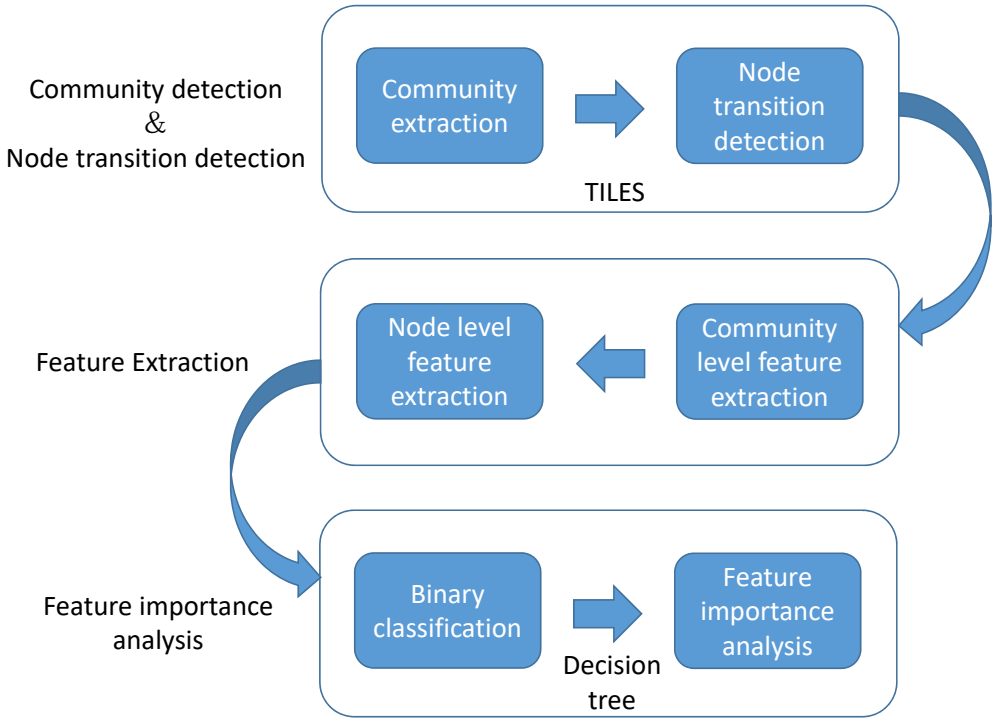


图 3-1 探究方法流程图

针对节店结构属性对节点社团转移影响的探究方法如图3-1所示，该方法共分为三步。

1. 社团检测及节点转移检测，首先利用TILES[26]对动态网络数据进行社团检测，同时TILES还会将复杂网络三元组数据进行时间片切分，即将(fi*G*fi)形式的数据处理为传统的动态网络时间片结构$W = \{W^1, W^2, ..., W^T\}$，其中$T$为动态网络时间片的个数，而$W^t \in 0,1^{N\times N}$为$t$时刻的网络邻接矩阵。我们假设所有该框架涉及到的网络均为无权图。随后我们将相邻的网络时间片组成相邻时间片组，并标注其中发生社团转移的节点与未发生社团转移的节点作为分类的节点社团转移标签。
2. 节点结构特征提取，通过特征工程方法及以往论文的经验[53]，本方法分别提取了五个社团级别结构特征及五个节点级别结构特征共十个结构特征。这些特征分别为：社团节点数、社团边数、社团内部边(即指向社团内部节点的边数)、社团外部边(即从社团内指出的边)、社团活性、社团的传导率、节点度、节点的平均邻居度、节点的接近中心性、节点的介数中心性。其具体描述见表3-1。通过相邻时间片组提取节点结构属性的方法见算法1。
3. 特征重要性分析，利用决策树对节点社团转移标签进行分类，将第二步提取的节点结构特征作为分类特征进行二分类。分类后，利用MDI计算每个特征在分类任务中的重要性，最终得到不同节点结构特征在节点发生社团转移情景中的重要性。

在该框架的第一步中，框架利用了现有的社团检测框架TILES。TILES是现有的最高水准的基于进化聚类的社团检测算法，其利用标签传播方法检测动态网络中的社团，并在该进程中划分每个动态网络时间片。TILES检测的社团结构是重叠社团，即网络中每个节点可以在一个网络快照中属于多个社团，这给社团转移检测造成了一定难度，即如何界定节点发生社团转移行为。TILES认为重叠社团类似社交网络中的账号可以属于多个圈子，针对这种思路，本框架认为当一个节点在上一时间片转移到本时间片时，其社团归属集中有新社团出现时，则该社团发生了社团转移，因为一个人的精力是有限的，那么当这个人加入了新的社团时，其在原社团中投入的精力就会减小，这也是节点的一种社团转移。

在特征提取步骤中，本文认为节点所属社团也会对节点的转移行为造成影响，例如，在社交网络中，不同的圈子人员流动性是不一样的，或者如不同的组织人员流动性是不一样的。这些组织或群体的特征会影响群体内人员的转移意愿，因此本框架引入了部分社团级别的特征作为节点的结构特征的一部分。

在特征重要性分析中，本文选择了决策树来作为节点转移二分类的分类方

法，因为决策树不同于深度神经网络，是白箱算法，因此每个特征的重要性都能通过决策树得到。本文使用Mean Decrease in Impurity(MDI)[54]计算每个特征的重要性，DMI的定义如下：

$$\Delta i(s,r) = i(r) - p_L i(r_L) - p_R i(r_R) \tag{3-1}$$

其中，$i(r)$ 是一些不纯度度量如gini index. r表示某个决策树节点，而r_L和r_R分别是 r的子节点。 $p_L = N_{r_L}/N_r$ 其中N_r是通过节点r的数据量。类似的$p_R = N_{r_R}/N_r$. 标准化后的$\Delta i(s,r)$可以给每个特征一个重要性度量，同时该计算指标非常高效，因此可以适用于大规模数据。

算法 1: Feature extraction

输入： A sequence of undirected graphs $W = W^1, ..W^T$ and the community assignment $C = C^1, ...C^T$

输出： Nodes feature set F and nodes label set L

for every graph W^t where $t \neq T$ **do**
  for every community C_l^t in C^t **do**
    Calculate community level features F_c
    for every node i in community C_l^t **do**
      Calculate node level features F_n
      Compose node i's feature sequence $F = F_c + F_n$
      if node i changes its community in W^{t+1} **then**
        node i's label $L_i = 1$
      else
        node i's label $L_i = 0$
      end if
    end for
  end for
end for

3.2 数据集

本章所用数据集包括因特网数据、Facebook数据、手机信令数据即Wiki数据等共15个多类型网络公开复杂网络数据集，对于数据集的详细描述见表3-2。如

表 3-1 notations and definitions

Symbol	Feature	Description	Definition
$f1$	Community node number	Number of nodes within the community l at time t.	n_l^t
$f2$	Community edge number	Number of edges within the community l at time t.	e_l^t
$f3$	Intra community edges	Ratio of the total number of edges between the nodes inside the community($e_l^t(in)$) to the number of nodes in the community.	$\frac{e_l^t(in)}{n_l^t}$
$f4$	Inter community edges	Ratio of the total number of edges of nodes connected outside the community($e_l^t(out)$) to the number of nodes in the community.	$\frac{e_l^t(out)}{n_l^t}$
$f5$	Community activity	Ratio of the total number of connections made in the previous snapshot by the nodes of the community(a_l^t) to the number of nodes in the community.	$\frac{a_l^t}{n_l^t}$
$f6$	Community Conductance	Ratio of the number of edges in the community to the sum of degrees of the nodes in the community.	$\frac{e_l^t}{d_l^t}$
$f7$	Node degree	Sum of links connected to node i at time t.	e_i^t
$f8$	Node average neighbor degree	Average degree of node i's neighbors, where $N(i)^t$ are the neighbors of node i at time t and e_j^t is the degree of node j which belongs to $N(i)^t$.	$\frac{1}{\lvert N(i)^t\rvert}\sum_{j\in N(i)^t} e_j^t$
$f9$	Node closeness centrality	Measuring a node i's average path length to other nodes in community, where $C_{l,-i}^t$ is a set of all nodes in community l except i at time t and $d(i, j)$ is the distance between node i and j.	$\sum_{j\in C_{l,-i}^t}\frac{C_l^t}{d(i,j)}$
$f10$	Node betweenness centrality	Measuring a node i's importance in its community connectivity, where σ_{jk} is the total number of shortest paths from node j to node k and $\sigma_{jk}(i)$ is the number of those paths that pass through i	$\sum_{j,k\in C_{l,-i}^t}\frac{\sigma_{jk}(i)}{\sigma_{jk}}$

图3-2所示，所有数据的节点度分布都服从power-law。

表 3-2 data sets description

Name	Description	$\|\mathcal{V}\|$	$\|\mathcal{E}\|$
Internet	Internet[55] topology during 04/01/2004 – 04/04/2005.	33936	104824
Facebook	Facebook New Orleans networks[56] friends links during 06/08/2008 – 21/01/2009.	62306	905565
bitcoin	Who-trusts-whom network of people who trade using Bitcoin on Bitcoin OTC[57] during 09/11/2010 – 19/01/2016.	5881	35592
Friend	Call logs of members of a young-family residential living community adjacent to a major research university in North America[58] during 10/07/2010 – 16/07/2011.	130	60518
fb-forum	The Facebook-like Forum Network[59] during 15/05/2004 – 24/10/2004.	899	33720
fb-messages	The Facebook-like Social Network[59] from an online community for students at University of California during 24/03/2004 – 22/10/2004.	1897	61734
ia-digg-reply	A reply network of the social news website Digg[59] during 29/10/2008 – 13/11/2008.	30397	87627
ia-facebook-wall-wosn-dir	The Facebook friendship graph[59] during 15/05/2004 – 24/10/2004.	44668	876993
ia-reality-call	The MIT Reality mining a small set of human call logs data[59] during 24/09/2004 – 07/01/2005.	6810	52050
ia-slashdot-reply-dir	Reply network of technology website Slashdot[59] during 01/12/2005 – 31/08/2006.	51097	140778
ia-stackexch-user-marks-post	User answering question network of Stack Overflow[59] during 03/10/2008 – 25/11/2011.	545196	1302439
ia-yahoo-messages	The message network in yahoo[59] with time presented by link sequences.	99303	3179718
soc-epinions-trust-dir	Epinion who-trusts-whom network[59] with time presented by link sequences.	131828	841373
soc-wiki-elec	Wikipedia adminship election data[59]during 14/09/2004 – 05/01/2008.	8271	107071
wiki	The Wikipedia links data[55] during 20/02/2001 – 06/12/2002.	329623	39953145

3.3 实验及验证

3.3.1 实验及结论

按照上文所述的研究框架处理上述15个数据集，并计算其特征重要性。结果如热图3-3所示，横坐标表示节点的分类特征，即分类特征；纵坐标分别表示15个数据集，颜色代表特征在不同数据集中的分类任务中所占重要性。可以看到$f7$(节点的度)与$f8$(节点的平均邻居度)在所有特征分类中均占很大比重，尤其在$fb-forum$数据中，节点的平均邻居度在分类中所占比重超过了0.5；而$f5$(社团活跃度)在$ia-slashdot-reply-dir$数据中所占比重超过其他数据，该数据为技术网站Slashdot的回复网络，从这一点来看，**科技类网站的圈子活跃度**

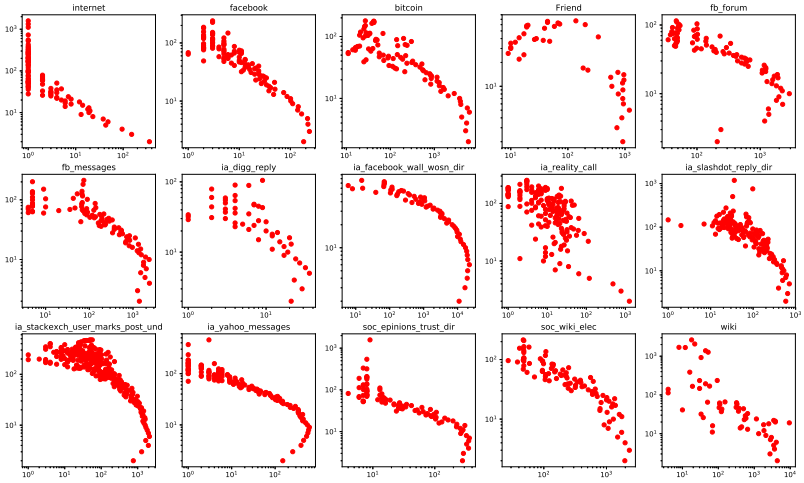


图 3-2　　15个真实数据集的度分布可视化

是影响其用户更换兴趣圈的主要因素；与此同时，在*internet*网络中，$f9$(节点的接近中心性)对其节点的社团转移影响最大，显而易见，在因特网中，连通性是其最至关重要的指标之一。虽然$f5, f9$均在某些数据集中显示出了对节点社团转移的重要影响力，但是其并不具有普遍性。反观$f7, f8$，其在所有数据集中对节点的社团转移均具有可观的影响力。因此我们得出结论，**节点的度以及节点的平均邻居度是影响节点发生社团转移的重要的结构特征**。

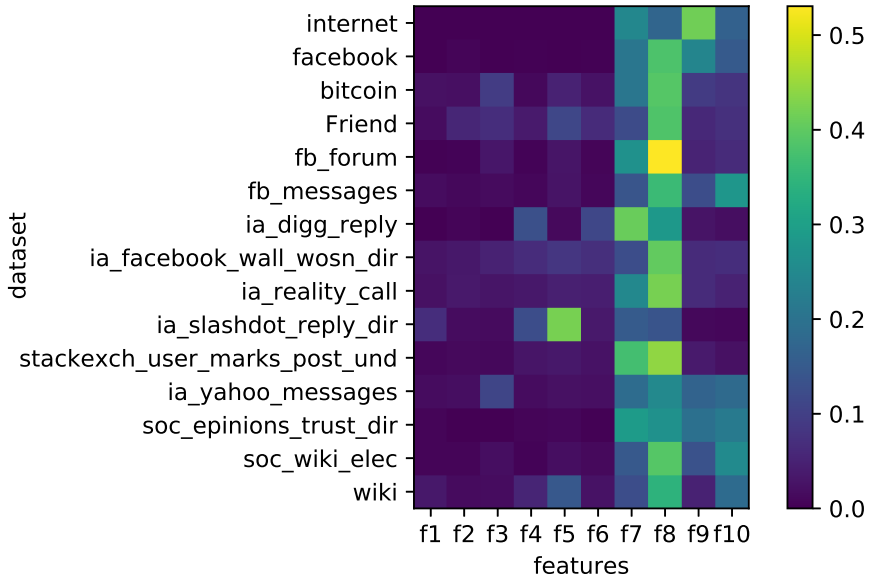


图 3-3　　15个数据集的特征重要性热图

为了探究节点的度以及节点的平均邻居度是以何种模式影响节点的社团转移的，本文分别统计了以上两个节点结构属性在15个数据集中的分布。如图3-4所示，该图展示了节点的度在十五个数据集中的分位图，每个子图中共有两列分位图，分别为发生转移的节点(横坐标为1)以及未发生转移的节点(横坐标为0)的度的分位统计图。从图中可以看到，在所有15个数据集中，发生社团转移的节点的度普遍比未发生转移的节点的度更高，即在一个网络中，节点的度越高，其越有可能转移其所在的社团。这也验证了社团检测中的大量度修正方法[60,61]的正确性。而图3-5则展示了节点的平均邻居度在十五个数据集中的分位图，每个子图中共有两列分位图，分别为发生转移的节点(横坐标为1)以及未发生转移的节点(横坐标为0)的平均邻居度的分位统计图。可以看到节点的平均邻居度在不同数据集中的发生转移以及未发生转移节点中的值并不相同，这说明不同的节点平均邻居度确实在影响节点的社团转移行为，然而其模式在所有十五个数据集中并不统一。

3.3.2 验证

为了验证结论**节点的度及节点的平均邻居度影响节点的社团转移**，本文在论文引用网络DBLP中支撑结论的案例。如图3-6所示，图中节点代表论文作者，而节点大小代表节点的度的大小。不同颜色代表节点所属社团不同，即不同颜色作者的研究兴趣不同，同时节点上的文字代表“平均邻居度-作者名”，例如“4.82-*ShuichengYan*”代表作者名为*ShuichengYan*的平均邻居度为4.82。

其中图3-6 (a)3-6 (b)展示了平均邻居度给论文作者带来的影响，图3-6 (a)与图3-6 (b)是相邻时间快照(2005年和2006年)的相同作者的论文发表可视化图。可以看到，*JunYan*,*ZhengChen*和*NingLiu* 都具有较高的平均邻居度9.33，受他们共同的具有较高节点度的邻居*ShuichengYan*的影响，在2006年，他们转移了原有的研究兴趣领域，在TKDE上与*ShuichengYan*合作发表了与*ShuichengYan*相同研究兴趣的论文[1]。

而图3-6 (c)3-6 (d)则展示了节点度对论文作者带来的影响，图3-6 (c)和图3-6 (d)也是相邻时间快照(2005年和2006年)相同作者的论文发表可视化图。如图所示，*MarcPollefey*所属的节点具有较大的直径，即其节点度较大，这代表着他的朋友较多，而由图3-6 (c)可以看出，他的绝大部分朋友都与他所在的研究兴趣领域不同。受其朋友影响，他在2006年改变了其研究兴趣领域。更详细的说，他与他的合作者*Jan − MichaelFrahm*在2006年在EDGE共同发表了一篇论

[1] Yan, Jun, et al. ”Effective and efficient dimensionality reduction for large-scale and streaming data preprocessing.” IEEE transactions on Knowledge and Data Engineering 18.3 (2006): 320-333.

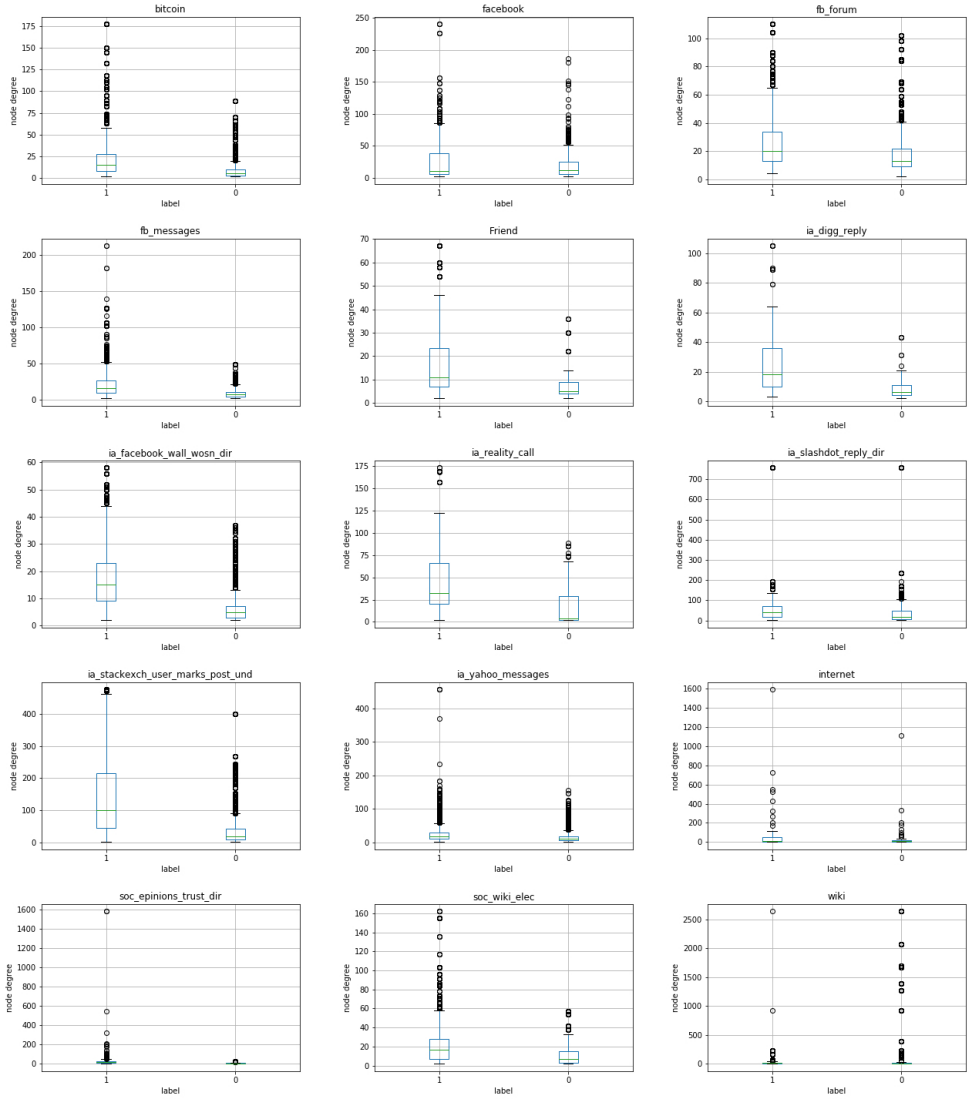


图 3-4　　15个数据集的节点度的分位图

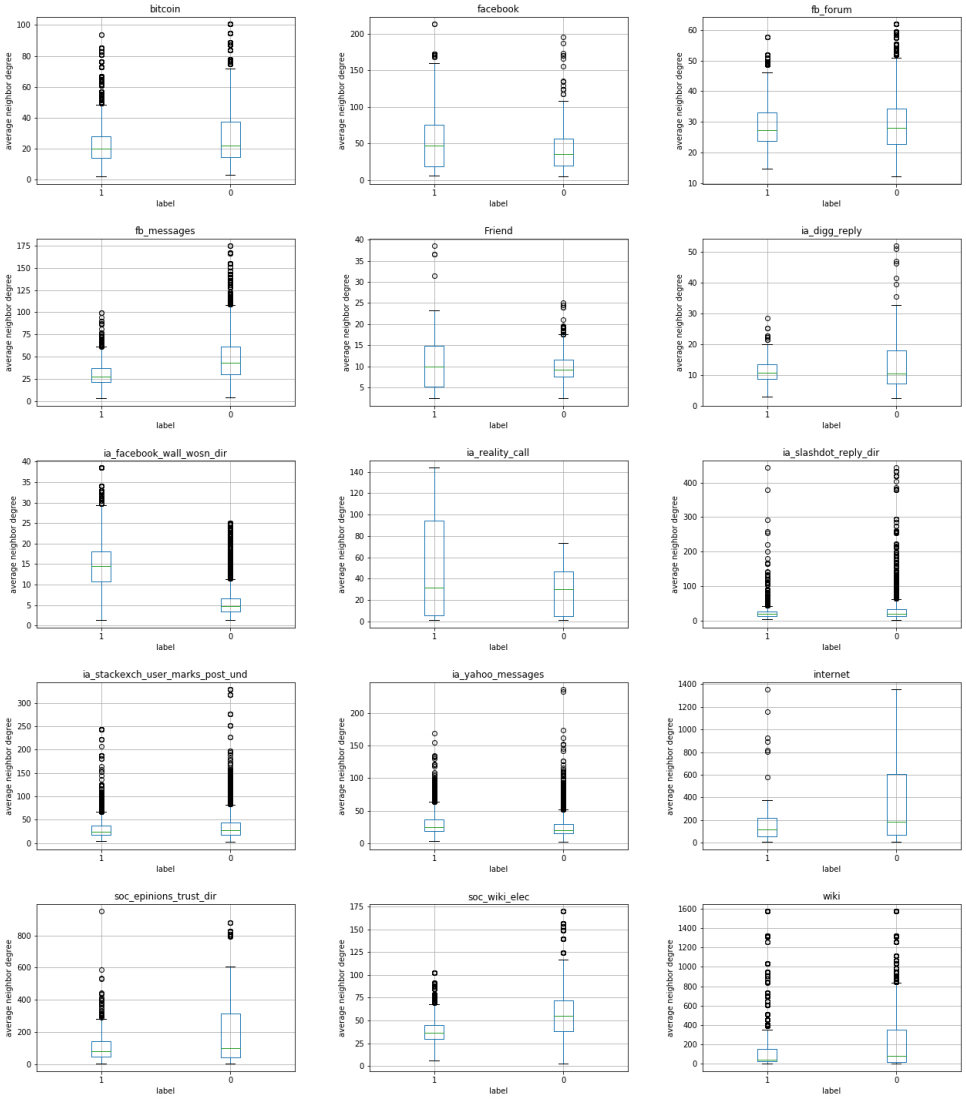


图 3-5　　15个数据集的节点的平均邻居度的分位图

文[2]。从图中的两个较小的节点*RolandMemisevic*和*ChristopherZech*也能看出平均邻居度的影响，这二人都具有较高的平均邻居度8.33，而受到两人共同的朋友*MarcPollefeys*的影响，他们也在2006年改变了研究兴趣领域。

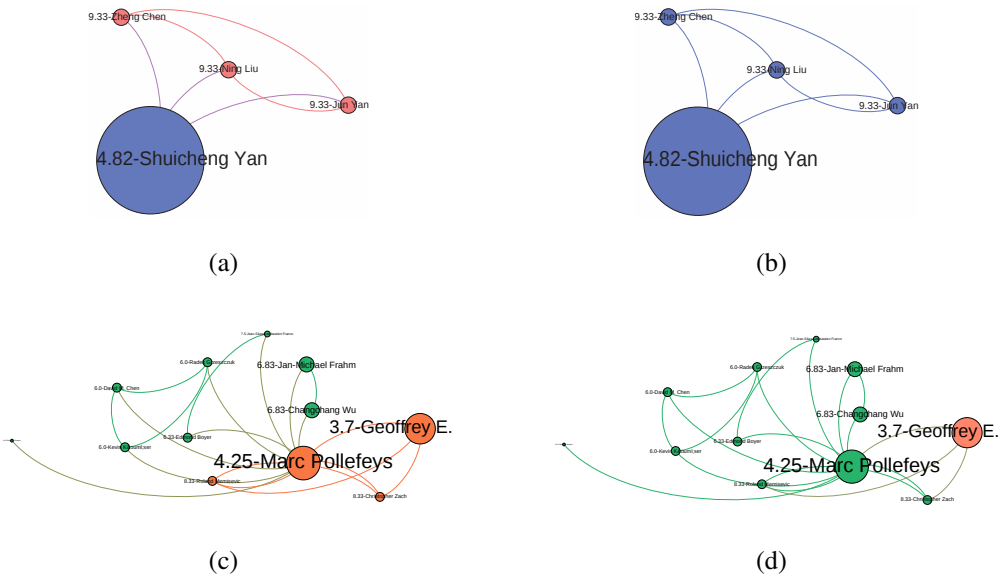


图 3-6　基于DBLP数据的验证案例

3.4　本章小结

通过将节点的结构属性视为分类特征，并将节点是否发生社团转移视为分类标签，本章利用决策树对其进行二分类，并分析了节点特征在分类中的重要性。最终得出结论：节点的度以及节点的平均邻居度对节点的社团转移影响最大且最为广泛。同时本章还在DBLP中找到了支撑结论的样例。下一章将会介绍受本章启发并改进自DSBM的社团检测生成模型HB-DSBM。

[2]Sinha, Sudipta N., et al. ”GPU-based video feature tracking and matching.” EDGE, workshop on edge computing using new commodity architectures. Vol. 278. 2006.

第4章 融合节点级别社团转移参数的动态网络社团检测生成模型

上一章介绍了利用决策树构建的框架，通过15个真实数据集的分析得到了影响节点社团转移的两个节点的结构属性：节点的度和节点的平均邻居度。受上一章启发，本章认为每个节点的社团转移由于其局部结构不同，都具有不同的转移倾向。而现有的动态网络社团检测模型大部分都将主要的关注点投入到了每个时间快照的社团检测上，而忽略了社团演化，小部分方法如DSBM[18]虽然关注了社团演化，但是其仅仅通过社团转移矩阵A来刻画社团演化，而忽略了每个节点的社团转移倾向的异质性，也就是说，在DSBM中，其假设同一个社团内的节点在下一个时刻发生转移的趋势都是一样的，且这种趋势在某个固定的社团内是一直不变的。这种假设在本文看来是不合理的。因此本章将会介绍引入更细粒度社团转移趋势的动态网络社团检测生成模型HB-DSBM。

4.1 HB-DSBM模型构建

4.1.1 符号表示

在介绍模型前，本章首先介绍模型的符号表示。如表4-1所示，本章用$W = \{W^1, W^2, ..., W^T\}$表示动态网络邻接矩阵，即$W^t$表示第$t$个网络快照的邻接矩阵，其中$t \in 1, ..., T$。这里$W^t \in \{0,1\}^{N\times N}$，其中$W_{ij}^t = 0$则代表$i$节点与$j$节点在网络快照$t$中没有边，若为1则有边。这里为了方便叙述，本章认为网络是无向无权的，这里需要说明，HB-DSBM可以有效的扩展为有向有权网络。$Z =$

表 4-1 HB-DSBM模型的符号表示

符号	描述
K,N,T	分别为社团个数、节点数和时间快照数
W^t	t时间快照对应的邻接矩阵
π_k	在时间快照1中，i节点属于社团k的概率
z_i^t	在t快照中，i节点属于哪个社团
A_k	k社团的社团级别转移倾向
C_i^t	i节点在t快照中的节点级别转移倾向, 其中$t = 2, \cdots, T$，$i = 1, \cdots, N$
B_{kl}	属于k社团的节点与属于l社团的节点在任意时间快照中的连边概率
γ ,μ	π和A_k的狄利克雷分布参数
α , β	B的Beta分布的参数

$\{Z^1, Z^2, ..., Z^T\}$代表每个网络快照中节点的社团划分，例如$z_i^t = k$则表示i节点在网络快照t中属于k社团，其中$k \in K$。

4.1.2 HB-DSBM模型

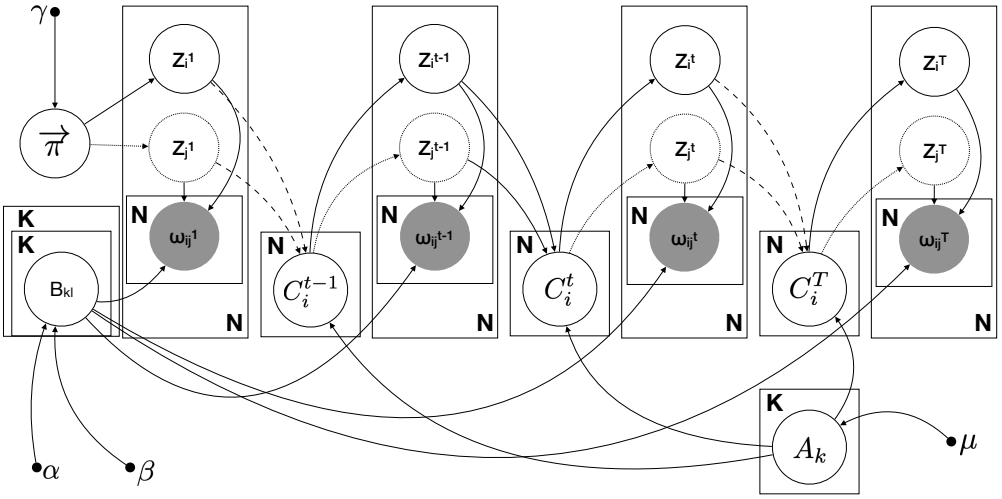


图 4-1　　HB-DSBM图模型

本节首先介绍HB-DSBM的核心生成机制来揭示它是如何通过层次贝叶斯结构同时建模社团级别和节点级别的动态演化的，随后给出其产生社团结构和动态网络的生成过程。

如图 4-1的图模型所示，HB-DSBM首先令π表示Z^1的先验分布，同时π服从参数为γ的狄利克雷分布。B是不同社团间节点的连边概率，即类随机块模型中的块矩阵，例如B_{kl}即代表属于k社团和属于l社团的节点之间连边的概率。而B服从参数为α和β的Beta分布。

本章中的$A \in [0,1]^{K\times K}$表示社团级别转移倾向矩阵，A的每一行A_k都服从参数为μ的狄利克雷分布，因此 $\sum_l A_{kl} = 1$。同时$C = \{C^2, \cdots C^T\}$表示节点级别的社团转移倾向，每个C^t都是由社团级别转移倾向矩阵A生成。对于$t > 1$的网络快照，任意节点i都有其唯一的转移向量 $C_i^t \in [0,1]^K$，该向量服从以参数为$A_{z_i^{t-1}}$的狄利克雷分布，因此$\sum_k C_{ik}^t = 1$。这一部分就是本模型的核心：层次狄利克雷生成机制。

基于以上机制，模型就可以生成社团标签以及动态网络中每个网络快照中的边。具体生成过程如下所示：

1. 生成初始社团划分概率 $\pi \sim Dir(\gamma)$
2. 生成块矩阵 $B \sim Beta(\alpha, \beta)$
3. 对于网络快照$t = 1$的每个节点i:
   (a) 生成每个节点的社团归属 $z_i^1 \sim Mult(\pi)$

(b) 生成每条边 $\omega_{ij}^1 \sim Bernoulli(\cdot|B_{z_i^1,z_j^1})$

4. 生成每个社团级别转移矩阵的社团转移向量 $A_k \sim Dir(\mu)$

5. 对网络快照$t > 1$中的每个节点i:

(a) 生成每个节点级别的社团转移向量$C_i^t \sim Dir(A_{z_i^{t-1}})$

(b) 生成每个节点的社团归属 $z_i^t \sim Mult(C_i^t)$

(c) 生成每条边 $\omega_{ij}^t \sim Bernoulli(\cdot|B_{z_i^t,z_j^t})$

由生成过程可知，当$t = 1$时，模型首先利用参数为π的多项分布生成i的社团归属z_i^1，随后以伯努利分布 $Bernoulli(\cdot|B_{z_i^1,z_j^1})$生成每对节点对$i, j$之间的连边。而当$t > 1$时，$i$节点的社团归属由以参数为$C_i^t$的多项分布决定，即$z_i^t \sim Mult(C_i^t)$。而$C_i^t$服从狄利克雷分布$Dir(A_{z_i^{t-1}})$。

根据如图 4-1的概率图模型和生成过程，可以写出HB-DSBM的联合概率分布如下：

$$\begin{aligned} &Pr(W_T, Z_T, C_T, B, A, \pi|\alpha, \beta, \gamma, \mu) \\ &= \prod_{t=1}^{T} Pr(W^t|Z^t, B) Pr(Z^1|\pi) \prod_{t=2}^{T} Pr(Z^t|C^t) \\ &\quad \prod_{t=2}^{T} Pr(C^t|A, Z^{t-1}) Pr(A|\mu) Pr(\pi|\gamma) Pr(B|\alpha, \beta) \end{aligned} \tag{4-1}$$

4.2 模型求解

本小结介绍针对本模型的高效的变分近似推断算法。

4.2.1 变分推断

由于模型参数较多且复杂，直接推出模型后验$p(Z, C, B, A, \pi|W)$是困难的, 因此基于平均场理论，本节提出了用$q(Z, C, B, A, \pi)$ 来近似$p(Z, C, B, A, \pi|W)$. 为了简介，本节用Δ 来表示参数 $\{\pi, B, A, C, Z\}$. 更详细来说，即：

$$q(\Delta) = \prod_{t=1}^{T} \prod_{i=1}^{N} q(z_i^t) \prod_{t=2}^{T} \prod_{i=1}^{N} q(c_i^t) q(B) q(A) q(\pi) \tag{4-2}$$

其中块矩阵变分参数$q(B|\widetilde{\alpha}, \widetilde{\beta}) = \prod_{k,l \geq k} Beta(\widetilde{\alpha}_{kl}, \widetilde{\beta}_{kl})$, 社团级别转移矩阵变分参数 $q(A|\widetilde{\mu}) = \prod_{k=1}^{K} \prod_{l=1}^{K} Dir(\widetilde{\mu}_{kl})$。而社团归属变分参数$q(z_i^t|\widetilde{\phi}_i^t)$ 服从以$\widetilde{\phi}_i^t$为参数的多项分布。而$q(c_i^t|\widetilde{\xi}_i^t)$和$q(\pi|\widetilde{\gamma})$都分别服从以$\widetilde{\xi}_i^t$和$\widetilde{\mu}_{kl}$为参数的狄利克雷分布。

因此有变分下界：

$$\widetilde{L}(q) = \sum_{z} \int_{\pi,B,A,C} q(\Delta) \log \frac{p(\Delta, W)}{q(\Delta)} = E_{\widetilde{\phi},\widetilde{\alpha},\widetilde{\beta}} \sum_{t=1}^{T} [\log P(W^t|Z^t, B)]$$
$$+ E_{\widetilde{\gamma},\widetilde{\phi}}[\log P(Z^1|\pi)] + E_{\widetilde{\phi},\widetilde{\xi}} \sum_{t=2}^{T} [\log P(Z^t|C^t)] + E_{\widetilde{\xi},\widetilde{\phi},\widetilde{\mu}} \sum_{t=2}^{T} [\log P(C^t|A, Z^{t-1})]$$
$$+ E_{\widetilde{\mu}}[\log P(A)] + E_{\widetilde{\gamma}}[\log P(\pi)] + E_{\widetilde{\alpha},\widetilde{\beta}}[\log P(B)] - E_{\widetilde{\gamma}}[\log q(\pi)] - E_{\widetilde{\alpha},\widetilde{\beta}}[\log q(B)] - E_{\widetilde{\mu}}[\log q(A)]$$
$$- \sum_{t=2}^{T} \sum_{i=1}^{N} E_{\widetilde{\xi}}[\log q(C_i^t)] - \sum_{t=1}^{T} \sum_{i=1}^{N} E_{\widetilde{\phi}}[\log q(z_i^t)] \tag{4-3}$$

这里 $\widetilde{\phi}, \widetilde{\xi}, \widetilde{\alpha}, \widetilde{\beta}, \widetilde{\mu}, \widetilde{\gamma}$ 即为变分参数。为了方便起见，本章省略掉了分布的条件部分。例如，本章将 $q(\Delta|\widetilde{\phi}, \widetilde{\xi}, \widetilde{\alpha}, \widetilde{\beta}, \widetilde{\mu}, \widetilde{\gamma})$ 和 $q(z_i^t|\widetilde{\phi}_i^t)$ 略写为$q(\Delta)$和$q(z_i^t)$。因此该模型的变分下界(ELBO)$\widetilde{L}(q)$被写成了如上式4-3。

随后通过最大化ELBO来获得最优的隐变量Z, π, B, A, C和模型参数$\gamma, \alpha, \beta, \mu$。通过对不同的参数$\widetilde{\phi}, \widetilde{\gamma}, \widetilde{\alpha}, \widetilde{\beta}, \widetilde{\mu}, \widetilde{\xi}$对$\widetilde{L}(q)$求偏导,并令导数为0求得每个参数的更新公式，即:

$$\nabla \widetilde{L}(q) = \{\frac{\partial \widetilde{L}}{\partial \widetilde{\gamma}}, \frac{\partial \widetilde{L}}{\partial \widetilde{\alpha}}, \frac{\partial \widetilde{L}}{\partial \widetilde{\beta}}, \frac{\partial \widetilde{L}}{\partial \widetilde{\mu}}, \frac{\partial \widetilde{L}}{\partial \widetilde{\xi}}, \frac{\partial \widetilde{L}}{\partial \widetilde{\phi}}\} = 0 \tag{4-4}$$

每个参数的结果如下所示: **对** $\widetilde{\gamma}$:

包含$\widetilde{\gamma}$的ELBO如下所示:

$$\begin{aligned} \widetilde{L}_{\widetilde{\gamma}} &= E_{\widetilde{\gamma},\widetilde{\phi}}[\log P(Z^1|\pi)] + E_{\widetilde{\gamma}}[\log P(\pi)] - E_{\widetilde{\gamma}}[\log q(\pi)] \\ &= \log \frac{\prod_{k=1}^{K} \log \Gamma(\widetilde{\gamma}_k)}{\Gamma(\sum_k \widetilde{\gamma}_k)} + \sum_{k=1}^{K} (\gamma_k - \widetilde{\gamma}_k + \sum_{i=1}^{N} \widetilde{\phi}_{ik}^1)[\psi(\widetilde{\gamma}_k) - \psi(\sum_k \widetilde{\gamma}_k)] \end{aligned} \tag{4-5}$$

这里 $\psi(x) = \frac{\Gamma'(x)}{\Gamma(x)} = \frac{d \log \Gamma(x)}{dx}$. 通过对$\widetilde{L}_{\widetilde{\gamma}}$ 针对$\widetilde{\gamma}$求偏导,并令导数等于0,得到其更新公式如下:

$$\widetilde{\gamma}_k = \gamma_k + \sum_{i=1}^{N} \widetilde{\phi}_{ik}^1, \tag{4-6}$$

对于$\widetilde{\xi}$:

ELBO中包含$\widetilde{\xi}$的式子如下所示:

$$\widetilde{L}_{\widetilde{\xi}} = E_{\widetilde{\phi},\widetilde{\xi}} \sum_{t=2}^{T} [\log P(Z^t|C^t)] + E_{\widetilde{\xi},\widetilde{\phi},\widetilde{\mu}} \sum_{t=2}^{T} [\log P(C^t|A, Z^{t-1})] - \sum_{t=2}^{T} \sum_{i=1}^{N} E_{\widetilde{\xi}}[\log q(C_i^t)] \tag{4-7}$$

利用同样方法求得更新公式如下所示:

$$\widetilde{\xi}_{ik}^t \propto \widetilde{\phi}_{ik}^t + \sum_l \widetilde{\phi}_{il}^{t-1} (\frac{\widetilde{\mu}_{kl}}{\sum_l \widetilde{\mu}_{kl}} - 1) + 1 \tag{4-8}$$

对于参数$\widetilde{\alpha}$**和**$\widetilde{\beta}$:

ElBO中包含$\widetilde{\alpha}$和$\widetilde{\beta}$的式子如下所示:

$$\begin{aligned}
\widetilde{L}_{\widetilde{\alpha},\widetilde{\beta}} &= E_{\widetilde{\phi},\widetilde{\alpha},\widetilde{\beta}}\sum_{t=1}^{T}[\log P(W^t|Z^t,B)] - E_{\widetilde{\alpha},\widetilde{\beta}}[\log q(B)] + E_{\widetilde{\alpha},\widetilde{\beta}}[\log P(B)] \\
&= \sum_{k\le l}\log\{\frac{\Gamma(\widetilde{\alpha}_{kl})\Gamma(\widetilde{\beta}_{kl})}{\Gamma(\widetilde{\alpha}_{kl}+\widetilde{\beta}_{kl})}\} \\
&+ \sum_{k=1}^{K}[\alpha_{kk}-\widetilde{\alpha}_{kk}+\sum_t\sum_{i<j}\widetilde{\phi}^t_{ik}\widetilde{\phi}^t_{jk}w^t_{ij}][\psi(\widetilde{\alpha}_{kk})-\psi(\widetilde{\alpha}_{kk}+\widetilde{\beta}_{kk})] \\
&+ \sum_{k=1}^{K}[\beta_{kk}-\widetilde{\beta}_{kk}+\sum_t\sum_{i<j}\widetilde{\phi}^t_{ik}\widetilde{\phi}^t_{jk}(1-w^t_{ij})][\psi(\widetilde{\beta}_{kk})-\psi(\widetilde{\alpha}_{kk}+\widetilde{\beta}_{kk})] \\
&+ \sum_{k<l}^{K}[\alpha_{kl}-\widetilde{\alpha}_{kl}+\sum_t\sum_{i\ne j}\widetilde{\phi}^t_{ik}\widetilde{\phi}^t_{jk}w^t_{ij}][\psi(\widetilde{\alpha}_{kl})-\psi(\widetilde{\alpha}_{kl}+\widetilde{\beta}_{kl})] \\
&+ \sum_{k<l}^{K}[\beta_{kl}-\widetilde{\beta}_{kl}+\sum_t\sum_{i\ne j}\widetilde{\phi}^t_{ik}\widetilde{\phi}^t_{jk}(1-w^t_{ij})][\psi(\widetilde{\beta}_{kl})-\psi(\widetilde{\alpha}_{kl}+\widetilde{\beta}_{kl})]
\end{aligned} \tag{4-9}$$

利用同样方法求得$\widetilde{\alpha}$和$\widetilde{\beta}$的更新公式如下:

$$\begin{aligned}
\widetilde{\alpha}_{kk} &= \alpha_{kk} + \frac{\sum_t\sum_{i<j}\widetilde{\phi}^t_{ik}\widetilde{\phi}^t_{jk}w^t_{ij}}{T} \\
\widetilde{\beta}_{kk} &= \beta_{kk} + \frac{\sum_t\sum_{i<j}\widetilde{\phi}^t_{ik}\widetilde{\phi}^t_{jk}(1-w^t_{ij})}{T} \\
\widetilde{\alpha}_{kl} &= \alpha_{kl} + \frac{\sum_t\sum_{i\ne j}\widetilde{\phi}^t_{ik}\widetilde{\phi}^t_{jl}w^t_{ij}}{T} \\
\widetilde{\beta}_{kl} &= \beta_{kl} + \frac{\sum_t\sum_{i\ne j}\widetilde{\phi}^t_{ik}\widetilde{\phi}^t_{jl}(1-w^t_{ij})}{T}
\end{aligned} \tag{4-10}$$

对于 $\widetilde{\mu}$:

ELBO中包含$\widetilde{\mu}$的式子如下所示:

$$\begin{aligned}
\widetilde{L}_{\widetilde{\mu}} &= E_{\widetilde{\xi},\widetilde{\phi},\widetilde{\mu}}\sum_{t=2}^{T}[\log P(C^t|A,Z^{t-1})] - E_{\widetilde{\mu}}[\log q(A)] + E_{\widetilde{\mu}}[\log P(A)] \\
&= \sum_{t=2}^{T}\sum_i\sum_k\sum_l\widetilde{\phi}^{t-1}_{ik}[\sum_l\psi(\widetilde{\mu}_{kl}) - \psi(\sum_l\widetilde{\mu}_{kl}) + \frac{\widetilde{\mu}_{kl}}{\sum_l\widetilde{\mu}_{kl}}(\psi(\widetilde{\xi}^t_{ik}) - \psi(\sum_l\widetilde{\xi}^t_{il}))] \\
&- \log[\Gamma(\sum_l\widetilde{\mu}_{kl})] + \sum_l\log[\Gamma(\widetilde{\mu}_{kl})] + \sum_l(\mu_l-\widetilde{\mu}_{kl})[\psi(\widetilde{\mu}_{kl}) - \psi(\sum_l\widetilde{\mu}_{kl})]
\end{aligned} \tag{4-11}$$

因此有:

$$\widetilde{\mu}_{kl} \propto \mu_l + \frac{\sum_{t=2}^{T}\sum_i\widetilde{\phi}^{t-1}_{ik}}{T-1} \tag{4-12}$$

对于 $\widetilde{\phi}$:

$\widetilde{\phi}$ 是社团划分参数Z的变分参数,其更新公式随t的不同而不同。下面分情况

讨论：

当$t=1$时:

ELBO中包含$\widetilde{\phi}$的项并考虑约束$\sum_k \widetilde{\phi}_{ik}=1$，有:

$$\begin{aligned}\widetilde{L}_{\widetilde{\phi}} &= E_{\widetilde{\phi},\widetilde{\alpha},\widetilde{\beta}}[\log P(W^1|Z^1,B)] + E_{\widetilde{\gamma},\widetilde{\phi}}[\log P(Z^1|\pi)] \\ &+ E_{\widetilde{\xi},\widetilde{\phi},\widetilde{\mu}}[\log P(C^2|A,Z^1)] - E_{\widetilde{\phi}}[\log q(Z^1)] + \rho(\sum_k \widetilde{\phi}_{ik}-1)\end{aligned} \tag{4-13}$$

因此有:

$$\begin{aligned}\widetilde{\phi}^1_{ik} \propto \exp\{&\sum_j \sum_l \widetilde{\phi}^1_{jl}[w^1_{ij}[\psi(\widetilde{\alpha}_{kl})-\psi(\widetilde{\alpha}_{kl}+\widetilde{\beta}_{kl})] + (1-w^1_{ij})[\psi(\widetilde{\beta}_{kl})-\psi(\widetilde{\alpha}_{kl}+\widetilde{\beta}_{kl})]] \\ &+ [\psi(\widetilde{\gamma}_k)-\psi(\sum_l \widetilde{\gamma}_l)] + [\sum_l \psi(\widetilde{\mu}_{kl}) - \psi(\sum_l \widetilde{\mu}_{kl}) + \sum_l (\frac{\widetilde{\mu}_{kl}}{\sum_l \widetilde{\mu}_{kl}}-1)(\psi(\widetilde{\xi}^2_{ik})-\psi(\sum_l \widetilde{\xi}^2_{il}))]\}\end{aligned} \tag{4-14}$$

当$1<t<T$:

ELBO中包含$\widetilde{\phi}$的项，同时考虑$\sum_k \widetilde{\phi}_{ik}=1$，有:

$$\begin{aligned}\widetilde{L}_{\widetilde{\phi}} &= E_{\widetilde{\phi},\widetilde{\alpha},\widetilde{\beta}}[\log P(W^t|Z^t,B)] + E_{\widetilde{\phi},\widetilde{\xi}}[\log P(Z^t|C^t)] \\ &+ E_{\widetilde{\xi},\widetilde{\phi},\widetilde{\mu}}[\log P(C^{t+1}|A,Z^t)] - \sum_{i=1}^{N} E_{\widetilde{\phi}}[\log q(z^t_i)]\end{aligned} \tag{4-15}$$

因此有:

$$\begin{aligned}\widetilde{\phi}^t_{ik} \propto \exp\{&\sum_j \sum_l \widetilde{\phi}^t_{jl}[w^t_{ij}[\psi(\widetilde{\alpha}_{kl})-\psi(\widetilde{\alpha}_{kl}+\widetilde{\beta}_{kl})] + (1-w^t_{ij})[\psi(\widetilde{\beta}_{kl})-\psi(\widetilde{\alpha}_{kl}+\widetilde{\beta}_{kl})]] \\ &+ [\psi(\widetilde{\xi}^t_{ik})-\psi(\sum_l \widetilde{\xi}^t_{il})] + [\sum_l \psi(\widetilde{\mu}_{kl}) - \psi(\sum_l \widetilde{\mu}_{kl}) + \sum_l (\frac{\widetilde{\mu}_{kl}}{\sum_l \widetilde{\mu}_{kl}}-1)(\psi(\widetilde{\xi}^{t+1}_{ik})-\psi(\sum_l \widetilde{\xi}^{t+1}_{il}))]\}\end{aligned} \tag{4-16}$$

当$t=T$:

ELBO中包含$\widetilde{\phi}$的项同时考虑约束$\sum_k \widetilde{\phi}_{ik}=1$，有:

$$\widetilde{L}_{\widetilde{\phi}} = E_{\widetilde{\phi},\widetilde{\alpha},\widetilde{\beta}}[\log P(W^T|Z^T,B)] + E_{\widetilde{\phi},\widetilde{\xi}}[\log P(Z^T|C^T)] - \sum_{i=1}^{N} E_{\widetilde{\phi}}[\log q(z^T_i)] \tag{4-17}$$

因此有:

$$\begin{aligned}\widetilde{\phi}^T_{ik} \propto \exp\{&\sum_j \sum_l \widetilde{\phi}^T_{jl}[w^T_{ij}[\psi(\widetilde{\alpha}_{kl})-\psi(\widetilde{\alpha}_{kl}+\widetilde{\beta}_{kl})]+ \\ &(1-w^T_{ij})[\psi(\widetilde{\beta}_{kl})-\psi(\widetilde{\alpha}_{kl}+\widetilde{\beta}_{kl})]] + [\psi(\widetilde{\xi}^T_{ik})-\psi(\sum_l \widetilde{\xi}^T_{il})]\}\end{aligned} \tag{4-18}$$

由更新公式可知，模型超参数μ和α 都为常量且其不影响模型结果。下面介绍得到更新公式后的迭代算法

4.2.2 迭代算法

现在已经有了所有参数的更新公式，下面给出变分EM的更新算法如算法2所示。

算法 2: HB-DSBM迭代算法

输入： 动态网络邻接矩阵 W, 对大迭代次数 n_{max} 和阈值变量ε

输出： $\widetilde{\alpha},\widetilde{\beta},\widetilde{\gamma},\widetilde{\mu},\widetilde{\xi},\widetilde{\phi}$

repeat

  给定$\widetilde{\phi}$, 根据迭代公式 4-6 4-8 4-10 4-12更新 $\widetilde{\xi},\widetilde{\gamma},\widetilde{\alpha},\widetilde{\beta},\widetilde{\mu}$

  给定 $\widetilde{\xi},\widetilde{\gamma},\widetilde{\alpha},\widetilde{\beta},\widetilde{\mu}$, 根据迭代公式4-14 4-16 4-18更新 $\widetilde{\phi}$

until $|\mathcal{L}^{new} - \mathcal{L}^{old}| < \varepsilon$ 或迭代次数 $> n_{max}$

return $\widetilde{\xi},\widetilde{\gamma},\widetilde{\alpha},\widetilde{\beta},\widetilde{\mu},\widetilde{\phi}$

算法的复杂度依赖于三个部分。更新ϕ的复杂度是$O(TN^2K^2)$, 其中 T 是网络快照的个数, N 网络中节点的数量，而K 是社团的个数。而更新 ξ的复杂度是$O(TNK^2)$。最后ELBO的计算复杂度是$O(TN^2K^2)$.因此理论上该算法的复杂度是 $O(N^2)$. 而真实世界网络的数据往往是稀疏的，因此可以通过负采样或并行计算有效提升算法的运行效率，使其适应于大规模数据集。

4.3 实验及验证

本小结将本文提出的算法HB-DSBM与四种同类动态网络社团检测算法在四个数据集(两个生成数据集与两个真实数据集)进行比较，以证明HB-DSBM的效果优于同类方法。同时在DBLP数据集中对HB-DSBM与DSBM进行社团演化分析，以证明本方法的社团演化分析更加细致且符合实际。

4.3.1 数据集

本章实验使用四个数据集对HB-DSBM及对比方法进行比较，包括两个生成数据集和两个真实数据集，所有数据集均有真相。

data1 由Lin 等人[62]基于Girvan 和Newman提出的数据的改进版数据集。该数据集具有不同的参数设置，每个参数设置组生成的数据包含10个时间快照和128个节点，这些节点分别属于4个不同的社团，每个社团均包含32个节点。参数σ表示某社团内部的节点与其他社团的节点连边的数量的均值；而nC代表每个时间快照到下一个时间快照离开其原有社团的节点的数量；aD则表示每个节点的平均度。

data2是由Greene等人[44]提出的。该数据集融合了社团级别的事件使得其更接近真实世界数据，该数据集中的每个网络包含1000个节点，15的平均度和50的最大节点度。该数据集的社团个数有20到50不等，社团间连边的概率为0.2。值得一提的是，该数据集中的节点服从网络的power-law。

KIT-email[63]数据集是email收发网络，其节点代表电子邮箱账号。本文将该数据集视作无向无权网络，并将该数据集处理成网络快照形式的数据，分别以2或3个月为一个时间间隔。其节点数以及社团个数对应于这两个时间间隔分别为138、170和23、25。

DBLP[64]数据是开源的论文数据集。本文选取了涉及计算机的三个领域的数据，包括数据挖掘(DM)、数据库(DB)和人工智能(AI)。时间跨度为9年，其中包括1163个节点和按领域划分的3个社团。本文将数据划分为9个网络快照，每个快照对应一年。

4.3.2 动态社团检测

本小结社团检测实验选取了四种动态网络社团检测方法，分别为DSBM[18]、DYNMOGA[30]、Genlouvain[28]和PisCES[29]，这四种动态社团检测方法均为不同实现方式得到的高质量的动态网络社团检测算法。以上四个算法的参数设置均遵从其原论文的设置。而HB-DSBM的超参数如前文所述，均对模型效果影响不大，因此实验部分设置参数为当$k \neq l$时 $\alpha_{kl} = 1$，而当$k \neq l$时 $\alpha_{kk} = 10$；β的设置与α相同； μ的设置和γ相同, 即 $\mu_k = \gamma_k = \frac{1}{K}$，其中$1 < k < K$。

4.3.2.1 评价指标

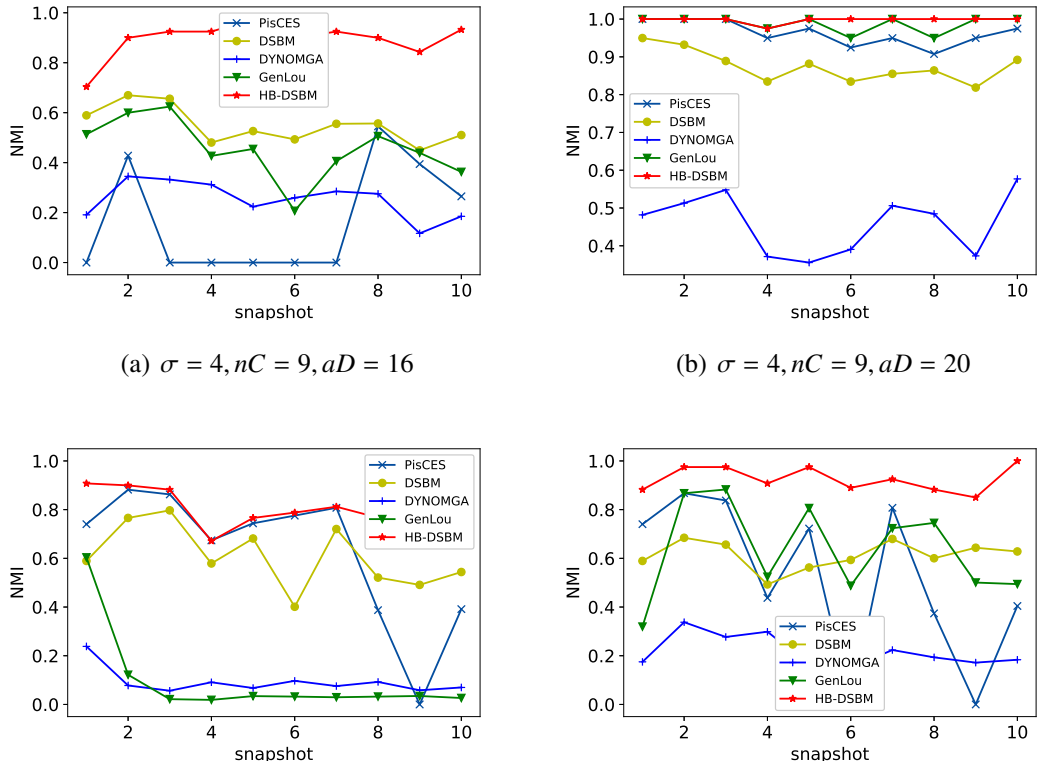
本章实验部分使用标准化互信息指标(NMI)[65]作为社团检测效果的评价指标，这也是目前大部分动态社团检测方法所使用的评价指标。NMI被设计为静态网络社团划分的评价指标，因此本章会在每个时间快照计算社团划分的NMI值。NMI被用作网络社团划分衡量指标需要知道网络社团划分的真相。令$C = \{C_1, ..., C_K\}$表示网络社团划分的真相，令 $C' = \{C'_1, ..., C'_K\}$代表待评估的社团划分，其中C_k和C'_k分别代表真相以及待评估的第k个社团的所有节点。则NMI的定义如下所示：

$$NMI(C, C') = \frac{\sum_{C,C'} p(C, C') \log \frac{p(C,C')}{p(C)p(C')}}{max(H(C)H(C'))} \tag{4-19}$$

其中，$H(C)$和$H(C')$表示社团C和C'的熵。NMI的值在0到1之间。NMI的值越高，代表待评估的社团划分与真相越接近。

4.3.2.2 实验结果

将HB-DSBM与DYNOMGA、DSBM、GenLou以及PisCES在不同数据集中进行NMI对比。在*data*1不同参数设置中的对比结果如图 4-2所示，HB-DSBM在图 4-2(a)中比效果最好的DSBM有大约明显的NMI提升，同时在图 4-2(d)中有0.1的NMI提升。虽然HB-DSBM在图 4-2(b)与图 4-2(c)中与其余方法的NMI提升差距不大，但是依然是效果最好的。HB-DSBM在五中方法中具有最好的NMI效果，证明该方法正确的把握住了节点演化的模式。HB-DSBM的曲线更加平滑，因为该模型结合了社团级别的演化以及节点级别的演化模式，成功的把握住了节点演化的异质性。DSBM的NMI曲线也比其他方法更加平滑，因为其融合了社团级别的演化模式，但是其忽略了同社团节点内演化模式的异质性。反观DYNOMGA和GenLou将社团检测与社团演化看做了两个独立的部分，因此其NMI曲线不够平滑，特别是在图 4-2(c)和图 4-2(d)中。而PisCES在不同时间快照中显示了很大的NMI反差，说明其对网络的噪声非常敏感。



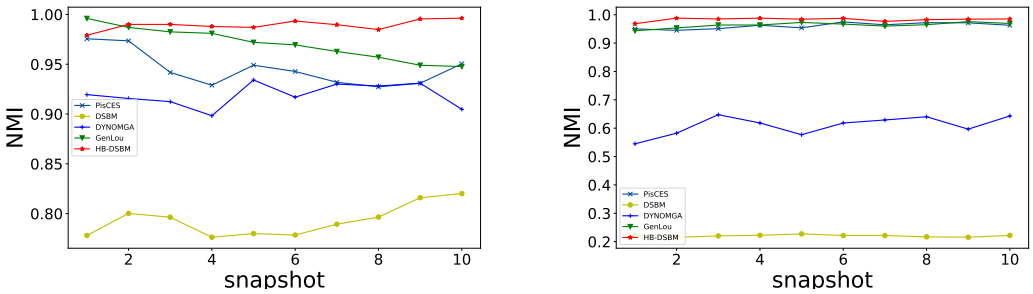
(a) $\sigma = 4, nC = 9, aD = 16$　　(b) $\sigma = 4, nC = 9, aD = 20$

(c) $\sigma = 5, nC = 3, aD = 16$　　(d) $\sigma = 5, nC = 9, aD = 20$

图 4-2　*data*1四种不同参数设置的不同方法NMI对比。

至于*data*2，NMI的结果显示如图 4-3所示，HB-DSBM依然具有最好的效果。在该数据集中，PisCES显示出比除HB-DSBM外其他方法更好的效果，

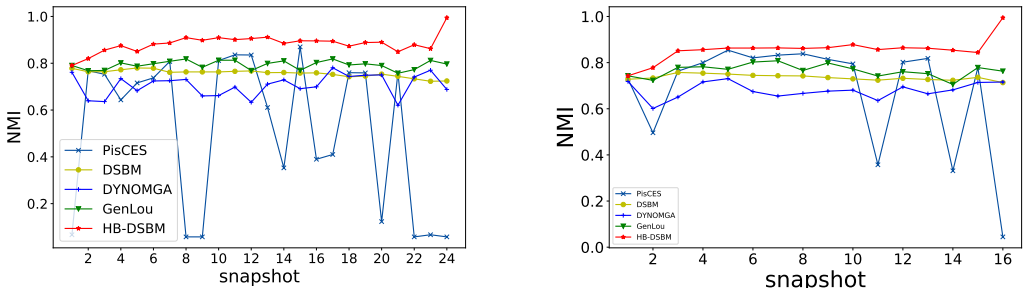
而DSBM却显示除了最差的NMI表现，因为其利用吉布斯采样对模型进行求解，使得该模型在大规模数据中的有限次迭代中很难达到局部最优，即使是在每个网络快照1000个节点的数据集中。



(a) 扩展事件网络　　(b) 社团交换事件网络

图 4-3　　*data2*两个事件网络的不同方法NMI对比。

真实世界数据集的实验如图 4-4 图4-5(a)所示，分别为KIT-email数据及DBLP数据。如图 4-4所示，HB-DSBM在KIT-email数据集中的NMI表现均高于其余方法，证明HB-DSBM不仅在生成数据集中具有较好表现，在真实世界数据集中也具有很好的社团检测效果。这里要强调的一点是，HB-DSBM在$t = 1$的时候，效果并不比其余方法具有明显提升，因为HB-DSBM的改进主要注重在动态社团检测中节点的多层次演化，因此在第一个网络快照中的效果并不明显，而随着t的增长，HB-DSBM的效果会越来越好。

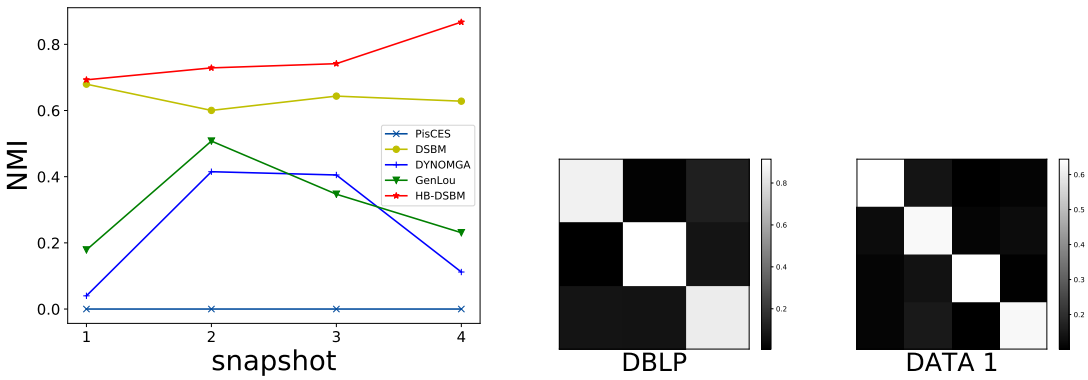


(a) 以两个月为间隔划分快照的网络　　(b) 以三个月为间隔划分快照的网络

图 4-4　　KIT-email数据不同时间间隔划分的切片网络的不同方法NMI对比

而DBLP数据的同类方法对比如图4-5(a)所示，HB-DSBM的效果明显高于其他方法，再次证明了本模型的有效性。而DSBM的效果也优于其他方法，因为DSBM为生成模型，其同时融合了不同网络快照的社团检测与社团演化。而PisCES的噪声敏感性使得其在DBLP数据中的表现非常差，将所有节点划分到了同一个社团中。该数据集的不同方法NMI对比也再次证明了本文提出的动

态网络社团演化层次贝叶斯结构的有效性。



(a) DBLP数据的NMI对比　　(b) DBLP以及*data*1的社团转移矩阵可视化

图 4-5　DBLP数据社团检测及演化效果图。

4.3.3 社团演化分析

以往的模型如DSBM将在同一个社团的节点视作等价，也就是说在同一个社团的两个节点在模型中没有任何差别，它们具有相同的社团转移倾向，并且DSBM中的社团转移倾向矩阵并不会随着时间推移而变化。也就是说，以往的模型只考虑了节点的社团级别的演化倾向，并且是不变的。在HB-DSBM中，隐变量C和A分别代表了节点级别和社团级别的演化倾向，同时通过层次生成结构将这两个参数整合到了一起。模型社团级别的转移倾向矩阵A的可视化如图4-5(b)所示，图中分别展示了DBLP和*data*1的社团级别的转移倾向。然而社团级别的转移倾向并不是一成不变的，社团级别的转移倾向会随着时间变化，如图4-6可以看到，HB-DSBM准确把握住了不同网络快照中的社团级别的节点社团转移倾向。

而同一社团的节点也具有不同的社团转移倾向，图4-7展示了DBLP数据中不同研究领域的作者的研究兴趣领域发生转移的倾向的异质性。例如，*TheoGevers*在2009年发表了一篇数据挖掘的论文。根据模型的节点转移矩阵的计算，他有很大的倾向继续在数据挖掘领域发表论文，而事实也验证了模型的推算。

通过HB-DSBM对DBLP数据的分析，本文还发现了一个有趣的现象，即大部分作者都倾向于在几年之内转换他们的研究兴趣领域。换句话说，只有很少一部分的作者会长期停留在同一个研究领域。如图4-8所示，*SvetlanaLazebnik*倾向于每两年转换一次研究兴趣领域，而这个时间间隔对*JingPeng*来说则为三年。同时*AndrewW.Fitzgibbon*从数据挖掘在$t = 7$时将研究领域转移到了人工智能领

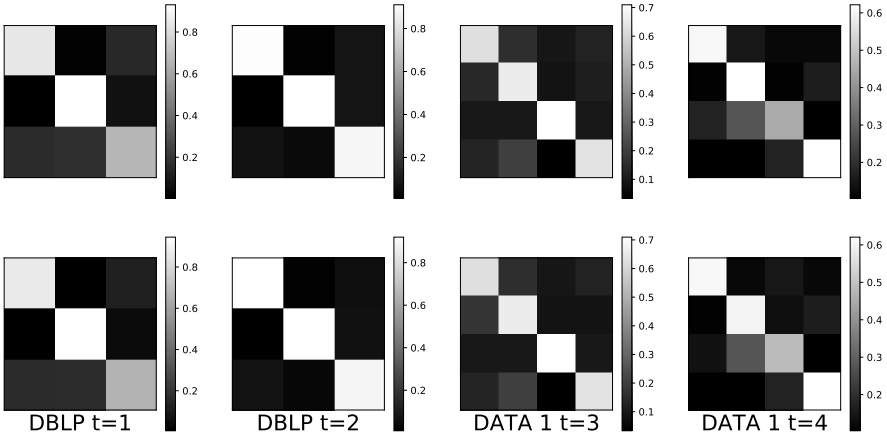


图 4-6　基于HB-DSBM模型的社团级别转移矩阵A分别在DBLP和$data1$数据的可视化(上半部分)与真实社团转移真相的可视化(下半部分)之间的对比。

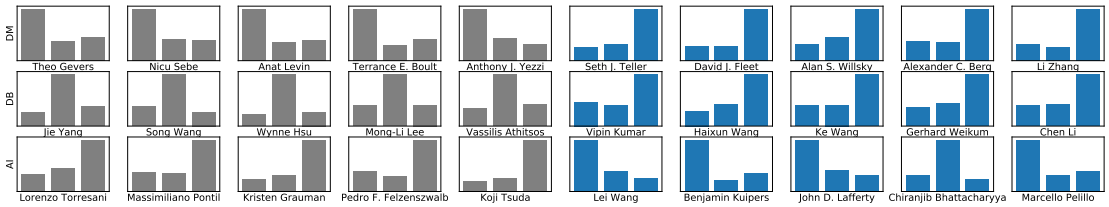


图 4-7　DBLP数据集中部分节点(2009 – 2010年)的社团转移倾向可视化(基于节点级别转移倾向参数C)。每个柱状图代表一个节点的转移倾向，从左到右分别为DM,DB,AI。所有柱状图共三行，每行代表一个领域，从上到下分别为DM,DB,AI。灰色的柱状图代表节点社团转移倾向与其所在社团转移倾向一致，而蓝色的柱状图代表节点的转移倾向与其所在社团的转移倾向不一致。

域。$AmnonShashua$则将研究兴趣领域从人工智能转移到数据挖掘进行研究，四年后又将研究兴趣领域转移回了人工智能。

4.4　本章小结

本章介绍了层次贝叶斯动态随机块模型，其层次贝叶斯生成结构能够同时融合节点级别以及社团级别的社团演化模式。同时本文利用变分推断对模型参数进行估计，利用合理的优化策略，模型可以适应大规模数据的计算。同类方法的对比也显示出HB-DSBM在动态网络社团检测任务中更加高效准确。在第五章将会介绍基于手机信令数据结合HB-DSBM的城市风险计算案例分析，以此来验证复杂网络算法在城市风险计算中的重要作用。

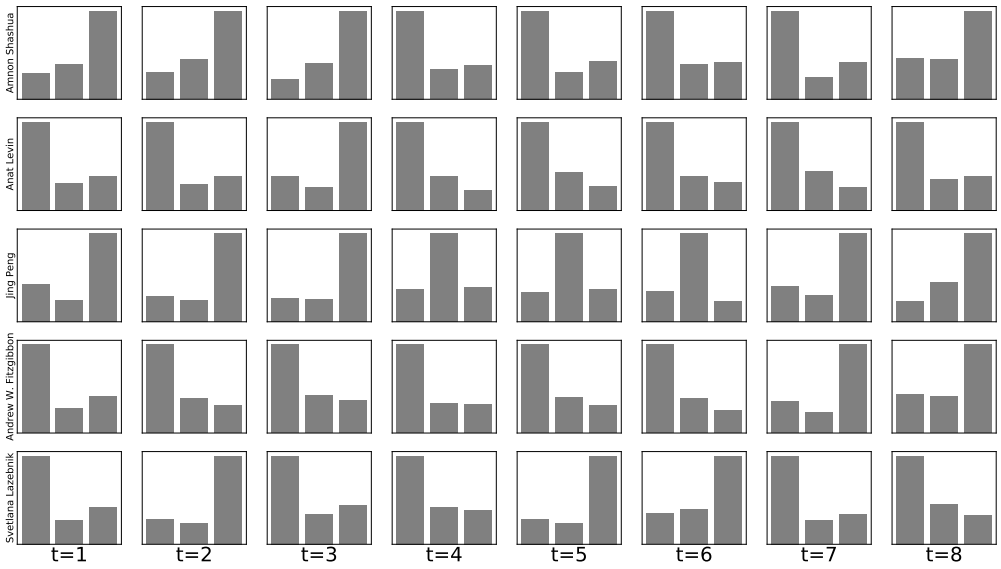


图 4-8 DBLP数据选取的五个作者($AmnonShashua$, $AnatLevin$, $JingPeng$, $AndrewW.Fitzgibbon$, $SvetlanaLazebnik$)的8年的转移倾向可视化。每个柱状图代表作者的社团转移倾向矩阵，从左到右分别为DM,DB,AI。

第5章 总结与展望

本文关注于动态复杂网络社团检测问题，依托于随机块模型框架，从模型的角度研究动态网络中的社团演化问题，同时探究了影响动态网络中节点社团转移的结构属性。随后从城市风险计算的角度对本文的方法以及探究的规律的有效性进行了实证分析。本章对本文的研究内容和研究成果进行总结分析，并指出其中模型的不足之处与未来的研究思路。

5.1 总结

社团结构作为复杂网络的重要任务之一，对于理解复杂网络结构形成、功能探索和领域应用如风险计算有重要的作用。而复杂网络的动态演化给社团检测带来了新的挑战，包括社团结构的演化、社团的产生与消亡、社团的分裂与合并等问题，与此同时，动态网络的社团检测也迫切需要我们对社团的演化行为进行建模。作为复杂网络社团检测的重要方法之一，随机块模型对网络的生成机制解释性的非常好，可以有效地建模复杂网络中社团的形成机制。而基于随机块模型构建的动态网络模型能够在有效地建模网络形成机制的同时把握网络中的社团演化机制。本文立足于动态随机块模型框架，从模型的角度对动态网络的社团结构进行建模，旨在解决动态网络中节点社团演化机制的问题。于此同时，本文还对节点的结构属性与节点的社团归属演变之间的关系通过多个真实数据集进行了探究，得到了对社团演化机制探究至关重要的结论。在应用方面，本文利用手机信令数据验证了本文提出的模型对城市风险计算的有效性。本文的主要工作和贡献点总结如下：

- 本文利用15个真实复杂网络数据集，包括社交网络数据（如twetter、facebook等）、wiki数据及社区通话数据等，通过TILES方法对数据进行动态网络社团检测，同时利用决策树将相邻时间片的节点是否发生社团转移作为二分类标签，将节点的结构属性通过特征工程组合为决策树的分类特征对节点进行二分类。通过对决策树分类后的分类模型中的特征重要性进行计算，本文得出结论：节点的度和节点的平均邻居度对节点的社团归属变化影响最大。并且在后续的案例分析中，本文找到了与之相佐证的真实情景。
- 基于动态随机块模型（DSBM），本文提出了层次贝叶斯动态随机块模型

（HB–DSBM）。该模型在相邻时间片引入了节点级别的社团转移参数，并提出了层次贝叶斯结构来生成社团级别的节点社团转移参数和节点级别的社团转移参数，并利用更细粒度的节点级别社团转移参数生成节点的在动态网络的社团归属。同时本文还对HB-DSBM提出了高效的变分推断算法，通过变分推断算法来对模型近似求解，同时提升了其算法运行效率，比之传统的DSBM模型运行效率更能适应大规模数据。通过将HB-DSBM和不同类的动态社团检测方法进行对比，并进行了社团演化分析对比，结果显示HB-DSBM在动态社团检测效果高于同类方法的同时，受益于更细粒度的社团转移参数，对于动态网络的社团演化分析更加精准。

- 通过对手机信令数据的处理，结合天津地块信息构建了多层复杂网络，并利用HB-DSBM对手机信令数据进行社团划分，并进一步通过社团标签-事件的提取方法计算出事件以及事件发生地以及事件发生时间。随后根据相关文献，通过合理的设计打分策略评判出每个事件的风险程度，并结合实际评判其合理性，以证明文章提出的HB-DSBM在风险计算中的可行性以及有效性。

5.2 展望

本文对于节点结构特征对社团演化的影响的探究以及构建的HB-DSBM模型对动态复杂网络社团演化的探究具有一定的贡献，但是随着最新的动态网络社团检测建模发展趋势以及城市风险计算发展的影响，本文的许多工作需要我们进行改进或者进一步研究。同时针对城市风险计算的需求，本文利用HB-DSBM对天津市手机信令数据进行社团检测，并利用相关文献进行事件提取以及每个事件的风险值评估，得到了风险事件的结果，然而该实证的计算依然存在一些待完善的部分。具体如下：

- 对于节点结构特征对社团演化的影响的探究中，我们得出结论：节点的度和节点的平均邻居度对节点的社团关系变化影响至关重要。其中，节点的度对节点的社团关系变化的影响是显而易见的，同时目前也有很多方法将节点的度应用于社团检测中来增强社团检测的效果，这些方法也都取得了预期的结果，这侧面证明了节点的度在动态网络社团检测中的重要作用。然而，节点的平均邻居度对节点社团转移的影响的内在规律还需要我们进一步探究，相信在不就的将来，节点的平均邻居度也可以有效的作用于动态网络社团检测。
- HB-DSBM模型在融合了节点粒度的转移参数后，确实对动态网络社团检测以及社团演化分析都起到了很大的作用，但是该模型在引入了如此细

粒度的参数后，使得其参数空间变得非常大，虽然本文提出了变分推断以及在实现中使用了随机采样等策略一定程度上降低了算法的复杂度，但是仍然达不到真正应用于生产的程度。因此对模型适当改进以降低其参数规模是HB-DSBM的下一步改进方向。

- HB-DSBM模型改进自DSBM，这使得其继承了DSBM的一大缺陷，即该模型不能适用于重叠社团的检测。在现实世界大部分情景中，同一个节点可以属于多个社团，例如大学学生可以加入多个兴趣社团等。该问题在静态网络随机块模型中得到了较好的解决，即混合随机块模型。而针对动态随机块模型的重叠社团问题，也已经有一些方法，但是均存在一些缺陷，因此对HB-DSBM进行适当改进使得其能够检测重叠社团也是其下一步重要的改进方向之一。
- HB-DSBM模型的一大假设就是需要预先知道动态网络中社团的个数，这在真实数据中是不可行的，例如在社交网络中，兴趣小组的个数不可能在不经聚类之前就提前知晓。同类方法的处理方式一般为在计算时设置社团个数$K = \log N$，其中，N为网络中的节点个数。再通过适当的方法将没有实际意义的社团舍弃以达到模型选择的目的（例如仅有一个节点的社团），而此种方法会大大增加算法的计算量，因此也不是很好的模型选择方法。HB-DSBM继承自DSBM，因此该方法的模型选择可以参考现有的针对DSBM的模型选择方法进行改进，如利用狄利克雷过程利用数据自动确定社团个数等。因此通过对HB-DSBM进行适当的改进来使该模型能够根据数据特征自动确定社团个数，即模型选择，也是其下一步改进的方向之一。
- 本文在实证部分，仅融合了手机信令数据与天津市地块数据构建了二层复杂网络对事件进行提取并对每个事件进行了风险分析。而真正的城市风险计算需要融合地块、事件、人员、人员行为等包括物理空间和网络空间的数据进行多层次融合以及计算之后才能得出准确的风险计算结果，多源信息融合也是风险计算中进行风险预测、风险决策、风险管控的基础，因此对于风险计算部分，下一步需要改进的就是完善数据收集以及信息融合的底层架构，这样才能在未来进行更深层次的风险计算。

参考文献

[1] Ben-Naim E, Frauenfelder H, Toroczkai Z. Complex networks [M]. Springer Science & Business Media, 2004.

[2] Watts D J, Strogatz S H. Collective dynamics of ‘small-world’ networks [J]. nature, 1998, 393 (6684): 440.

[3] Barabási A-L, Albert R. Emergence of scaling in random networks [J]. science, 1999, 286 (5439): 509–512.

[4] Rossetti G, Cazabet R. Community discovery in dynamic networks: a survey [J]. ACM Computing Surveys (CSUR), 2018, 51 (2): 35.

[5] Waltman L, Van Eck N J. A smart local moving algorithm for large-scale modularity-based community detection [J]. The European Physical Journal B, 2013, 86 (11): 471.

[6] Krzakala F, Moore C, Mossel E, et al. Spectral redemption in clustering sparse networks [J]. Proceedings of the National Academy of Sciences, 2013, 110 (52): 20935–20940.

[7] Holland P W, Leinhardt S. Local structure in social networks [J]. Sociological methodology, 1976, 7: 1–45.

[8] Airoldi E M, Blei D M, Fienberg S E, et al. Mixed membership stochastic block models for relational data with application to protein-protein interactions [C]. In Proceedings of the international biometrics society annual meeting, 2006.

[9] Hoff P D, Raftery A E, Handcock M S. Latent space approaches to social network analysis [J]. Journal of the american Statistical association, 2002, 97 (460): 1090–1098.

[10] Pal S, Coates M. Scalable MCMC in degree corrected stochastic block model [C]. In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2019: 5461–5465.

[11] Kemp C, Griffiths T L, Tenenbaum J B. Discovering latent classes in relational data [J], 2004.

[12] Hofman J M, Wiggins C H. Bayesian approach to network modularity [J]. Physical review letters, 2008, 100 (25): 258701.

[13] Chen Y, Li X, Xu J, et al. Convexified modularity maximization for degree-corrected stochastic block models [J]. The Annals of Statistics, 2018, 46 (4): 1573–1602.

[14] Qiao M, Yu J, Bian W, et al. Adapting Stochastic Block Models to Power-Law Degree Distributions [J]. IEEE transactions on cybernetics, 2018, 49 (2): 626–637.

[15] Dakiche N, Tayeb F B-S, Slimani Y, et al. Tracking community evolution in social networks: A survey [J]. Information Processing & Management, 2019, 56 (3): 1084–1102.

[16] Tajeuna E G, Bouguessa M, Wang S. Tracking the evolution of community structures in time-evolving social networks [C]. In 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA), 2015: 1–10.

[17] Jiang L, Shi L, Liu L, et al. An efficient evolutionary user interest community discovery model in dynamic social networks for internet of people [J]. IEEE Internet of Things Journal, 2019.

[18] Yang T, Chi Y, Zhu S, et al. Detecting communities and their evolutions in dynamic social networks—a Bayesian approach [J]. Machine learning, 2011, 82 (2): 157–189.

[19] Tang X, Yang C C. Dynamic community detection with temporal dirichlet process [C]. In 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing, 2011: 603–608.

[20] Wu X, Jiao P, Wang Y, et al. Dynamic Stochastic Block Model with Scale-Free Characteristic for Temporal Complex Networks [C]. In International Conference on Database Systems for Advanced Applications, 2019: 502–518.

[21] Kao E K, Smith S T, Airoldi E M. Hybrid mixed-membership blockmodel for inference on realistic network interactions [J]. IEEE Transactions on Network Science and Engineering, 2018.

[22] Yang J, Zhang M, Shen K N, et al. Structural correlation between communities and core-periphery structures in social networks: Evidence from Twitter data [J]. Expert Systems with Applications, 2018, 111: 91–99.

[23] Palla G, Barabási A-L, Vicsek T. Quantifying social group evolution [J]. Nature, 2007, 446 (7136): 664.

[24] Sun J, Faloutsos C, Faloutsos C, et al. Graphscope: parameter-free mining of large time-evolving graphs [C]. In Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining, 2007: 687–696.

[25] Jaiyeoba W, Skadron K. GraphTinker: A High Performance Data Structure for Dynamic Graph Processing [C]. In 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS), 2019: 1030–1041.

[26] Rossetti G, Pappalardo L, Pedreschi D, et al. Tiles: an online algorithm for community discovery in dynamic social networks [J]. Machine Learning, 2017, 106 (8): 1213–1241.

[27] 李亚芳, 贾彩燕, 于剑, et al. 一种新的社区/动态社区优化方法 [J]. 数据采集与处理, 2015, 30 (6): 1215–1224.

[28] Jutla I S, Jeub L G, Mucha P J. A generalized Louvain method for community detection implemented in MATLAB [J]. URL http://netwiki. amath. unc. edu/GenLouvain, 2011.

[29] Liu F, Choi D, Xie L, et al. Global spectral clustering in dynamic networks [J]. Proceedings of the National Academy of Sciences, 2018, 115 (5): 927–932.

[30] Folino F, Pizzuti C. An evolutionary multiobjective approach for community discovery in dynamic networks [J]. IEEE Transactions on Knowledge and Data Engineering, 2014, 26 (8): 1838–1852.

[31] Huang Q, Zhao C, Zhang X, et al. Community discovering in temporal network with spectral fusion [J]. Chaos, 2019, 29 (4): 043122.

[32] Sewell D K, Chen Y. Latent space approaches to community detection in dynamic networks [J]. Bayesian Analysis, 2017, 12 (2): 351–377.

[33] Yang L, Cao X, Jin D, et al. A Unified Semi-Supervised Community Detection Framework Using Latent Space Graph Regularization [J]. IEEE Transactions on Systems, Man, and Cybernetics, 2015, 45 (11): 2585–2598.

[34] Pensky M, Zhang T. Spectral clustering in the dynamic stochastic block model [J]. arXiv: Methodology, 2017.

[35] Chakrabarti D, Kumar R, Tomkins A. Evolutionary clustering [C]. In Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, 2006: 554–560.

[36] Mucha P J, Richardson T, Macon K T, et al. Community Structure in Time-Dependent, Multiscale, and Multiplex Networks [J]. Science, 2010, 328 (5980): 876–878.

[37] Tang X, Yang C C. Detecting social media hidden communities using dynamic stochastic blockmodel with temporal dirichlet process [J]. ACM Transactions on Intelligent Systems and Technology (TIST), 2014, 5 (2): 36.

[38] Xu K S, Hero A O. Dynamic stochastic blockmodels: statistical models for time-evolving networks [J]. international conference on social computing, 2013: 201–210.

[39] Yu L, Woodall W H, Tsui K-L. Detecting node propensity changes in the dynamic degree corrected stochastic block model [J]. Social Networks, 2018, 54: 209–227.

[40] Asur S, Parthasarathy S, Ucar D. An event-based framework for characterizing the evolutionary behavior of interaction graphs [J]. ACM Transactions on Knowledge Discovery From Data, 2009, 3 (4): 16.

[41] Bródka P, Saganowski S, Kazienko P. GED: the method for group evolution discovery in social networks [J]. Social Network Analysis and Mining, 2013, 3 (1): 1–14.

[42] Gao W, Luo W, Bu C. Evolutionary community discovery in dynamic networks based on leader nodes [J], 2016: 53–60.

[43] Tantipathananandh C, Bergerwolf T Y. Finding Communities in Dynamic Social Networks [J], 2011: 1236–1241.

[44] Greene D, Doyle D, Cunningham P. Tracking the evolution of communities in dynamic social networks [C]. In 2010 international conference on advances in social networks analysis and mining, 2010: 176–183.

[45] 吴竹. 群体性事件预警指标体系研究 [J]. 政法学刊, 2007, 24 (3): 63–67.

[46] 王连强. [S. l.]: . [s. n.], 2006.

[47] 龚俭, 臧小东, 苏琪, et al. 网络安全态势感知综述 [J]. 软件学报, 2017, 28 (4): 1010–1026.

[48] 陈长坤, 纪道溪, ChenChangkun, et al. 基于复杂网络的台风灾害演化系统风险分析与控制研究 [J]. 灾害学, 2012, 27 (1): 1–4.

[49] 基于电力系统复杂网络特征的线路脆弱性风险分析 [J]. 电力自动化设备, 2014, 34 (2): 101–107.

[50] 张树德. 基于复杂网络理论的城市道路网络脆弱性研究 [D]. [S. l.]: 哈尔滨工业大学, 2014.

[51] Gou L, Wei B, Sadiq R, et al. Topological Vulnerability Evaluation Model Based on Fractal Dimension of Complex Networks [J]. PLOS ONE, 2016, 11 (1).

[52] Moriano P, Finke J, Ahn Y. Community-Based Event Detection in Temporal Networks [J]. Scientific Reports, 2019, 9 (1): 4358.

[53] Ilhan N, Öğüdücü Ş G. Feature identification for predicting community evolution in dynamic social networks [J]. Engineering Applications of Artificial Intelligence, 2016, 55: 202–218.

[54] Menze B H, Kelm B M, Masuch R, et al. A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data [J]. BMC bioinformatics, 2009, 10 (1): 213.

[55] Mislove A. Online Social Networks: Measurement, Analysis, and Applications to Distributed Information Systems [D]. [S. l.]: Rice University, Department of Computer Science, 2009.

[56] Viswanath B, Mislove A, Cha M, et al. On the Evolution of User Interaction in Facebook [C]. In Proceedings of the 2nd ACM SIGCOMM Workshop on Social Networks (WOSN'09), August 2009.

[57] Kumar S, Spezzano F, Subrahmanian V, et al. Edge weight prediction in weighted signed networks [C]. In Data Mining (ICDM), 2016 IEEE 16th International Conference on, 2016: 221–230.

[58] Aharony N, Pan W, Ip C, et al. Social fMRI: Investigating and shaping social mechanisms in the real world [J]. Pervasive and Mobile Computing, 2011, 7 (6): 643–659.

[59] Rossi R A, Ahmed N K. The Network Data Repository with Interactive Graph Analytics and Visualization [C/OL]. In Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, 2015. http://networkrepository.com.

[60] Wilson J D, Stevens N T, Woodall W H. Modeling and detecting change in temporal networks via a dynamic degree corrected stochastic block model [J]. arXiv preprint arXiv:1605.04049, 2016.

[61] Jin D, Chen Z, He D, et al. Modeling with node degree preservation can accurately find communities [C]. In Twenty-Ninth AAAI Conference on Artificial Intelligence, 2015.

[62] Lin Y-R, Chi Y, Zhu S, et al. Facetnet: a framework for analyzing communities and their evolutions in dynamic networks [C]. In Proceedings of the 17th international conference on World Wide Web, 2008: 685–694.

[63] Görke R, Holzer M, Hopp O, et al. Dynamic network of email communication at the Department of Informatics at Karlsruhe Institute of Technology (KIT)(2011).

[64] DBLP network dataset – KONECT. April 2017. `http://konect.uni-koblenz.de/networks/dblp-cite`.

[65] Gong Y, Xu W. Machine learning for multimedia content analysis [M]. Springer Science & Business Media, 2007.

发表论文和参加科研情况说明

（一）发表的学术论文 （二）参与的科研项目

致　谢