

高阶生成机理导向的大规模动态网络社团检测及演化分析研究

Higher-Order Generative Mechanism Guided large-scale dynamic Community Detection and Evolution Analysis

一级学科：计算机科学与技术
学科专业：计算机科学与技术
作者姓名：李天鹏
指导教师：王文俊 教授

智能与计算学部
二〇二五年四月

摘要

动态网络社团检测及其演化分析可以有效地挖掘复杂网络的结构和功能及其演化模式，广泛应用于城市规划、推荐系统、突发事件检测等领域。动态网络社团检测任务的核心是对动态网络演化模式的有效建模，现有方法对动态网络演化模式的刻画忽略了对网络高阶生成机理的融合，无法有效支撑社团演化分析。针对上述问题，本文在动态随机块模型的基础上，从实证出发，研究基于随机块模型的动态网络社团检测与演化分析问题。主要研究内容包括：

第一，针对现有动态随机块模型的高阶生成机理启发式建模的问题，首先提出了动态网络节点与社团演化的关联挖掘框架，通过将动态网络中节点的社团演化问题转化为相邻网络快照中节点是否发生社团归属转移的二分类问题，验证了节点演化异质性的存在。并基于该实证提出了基于层次狄利克雷生成结构的层次动态随机块模型**HB-DSBM**，通过引入社团-节点的层次社团转移高阶生成结构建模了同一社团内节点的演化异质性。

第二，针对动态随机块模型的演化分析能力提升，提出了动态网络层次演化异常检测方法，引入节点流行度参数捕获节点在拓扑与演化层面的异质性，提出动态网络层次演化异常检测指标以识别动态网络节点、社团、网络快照各层次的演化异常，扩展模型对动态网络演化分析的维度，在真实世界网络中进行了验证与演化异常挖掘。

第三，提出了基于深度生成模型的动态网络社团检测算法。针对高精度随机块模型参数过多、参数后验推断复杂而造成模型对大规模数据处理能力差的核心问题，通过引入深度生成模型的变分自编码器架构扩展了动态随机块模型的大规模数据应用能力，同时保留参数的可解释性。

本文从动态网络的高阶生成机理出发，以动态网络高阶演化异质性实证挖掘为基础，基于随机块模型设计了动态网络社团检测与演化分析方法，扩展了随机块模型的演化分析深度，依托深度生成模型为动态随机块模型的大规模数据应用提供了方法框架，实验与实证分析验证了所提出方法的有效性与实用性。

关键词： 复杂网络，社团检测，动态网络，社团演化分析

目 录

摘 要	I
ABSTRACT	III
目 录	V
第一章 绪论	1
1.1 研究背景与意义	1
1.2 研究现状	2
1.2.1 静态网络社团检测研究现状	3
1.2.2 动态网络社团检测研究现状	4
1.3 研究内容与创新点	12
1.4 论文组织结构	15
第二章 相关研究基础	17
2.1 动态网络社团检测相关研究	17
2.1.1 基于启发式优化的方法	17
2.1.2 基于矩阵分解的方法	20
2.1.3 基于谱优化的方法	21
2.1.4 基于多目标优化的方法	22
2.1.5 基于动力学的方法	22
2.1.6 基于随机块模型的方法	23
2.1.7 基于深度学习的方法	25

2.2	深度生成模型相关研究基础	26
2.3	动态网络社团检测评价指标	28
2.4	本章小结	29
第三章	动态网络高阶生成机理的随机块模型建模	31
3.1	引言	31
3.2	网络高阶生成机理挖掘	32
3.2.1	用于网络高阶生成机理挖掘的数据描述	36
3.2.2	节点演化异质性归纳	36
3.3	融合节点社团转移异质性的动态随机块模型	40
3.3.1	HB-DSBM的准备工作	40
3.3.2	HB-DSBM的模型设计	40
3.4	HB-DSBM模型求解	42
3.4.1	HB-DSBM参数的近似后验推断	42
3.4.2	HB-DSBM算法迭代	44
3.5	HB-DSBM的实验及验证	45
3.5.1	数据集	45
3.5.2	社团检测实验	46
3.5.3	社团演化分析	49
3.5.4	节点的演化异常分析	49
3.6	本章小结	52
第四章	融合节点流行度的动态网络建模及层次异常检测	53
4.1	引言	53
4.2	GEABS模型描述	56
4.2.1	GEABS的准备工作	57
4.2.2	GEABS模型的生成过程	57

4.2.3	GEABS模型的参数近似后验推断	60
4.2.4	动态网络层次演化异常检测定义	62
4.3	模型推断及参数求解	63
4.3.1	隐变量更新公式推断	64
4.3.2	参数更新公式推断	66
4.3.3	算法流程	67
4.4	实验评测	68
4.4.1	数据集	68
4.4.2	社团检测	69
4.4.3	快照级别异常检测	70
4.4.4	社团级别异常检测	73
4.4.5	节点级别异常检测	75
4.4.6	GEABS数据挖掘实证	75
4.5	本章小结	78
第五章	基于变分自编码器的动态网络深度社团检测方法	79
5.1	引言	79
5.2	VGRGMM模型描述	81
5.2.1	符号与问题定义	82
5.2.2	VGRGMM模型框架	83
5.2.3	优化过程	86
5.3	VGRGMM模型实验	87
5.3.1	数据集	89
5.3.2	实验设置	89
5.3.3	评价指标	91
5.3.4	动态社团检测实验结果	92

5.3.5 可视化结果	96
5.3.6 参数敏感性分析	97
5.3.7 实证分析	97
5.4 本章小结	98
第六章 总结与展望	101
6.1 研究总结	101
6.2 研究展望	103
参考文献	104
发表论文和参加科研情况	123
附录	124
致谢	140

第1章 绪论

本章首先介绍动态复杂网络社团检测及其演化分析的研究背景以及其在实际应用中的重要意义，然后详细介绍该领域当前国内外研究现状，并基于此提出本文研究内容与创新点，最后总结了本文的章节结构。

1.1 研究背景与意义

自1998年小世界网络^[1]与无标度网络^[2]提出以来，复杂网络及复杂性科学即迎来了长足的发展。复杂网络作为一种数据结构能够更大程度地保留复杂系统各模块间的交互，因此对复杂网络的研究与探索能够帮助人们理解现实世界的运作模式，如人工系统的运作机理如电网^[3]、水网^[4]；人类社会的发展模式，如邮件网络传播模式^[5]、科研合作模式^[6]；自然界的演化规律，如食物链的鲁棒性^[7]、城市的韧性^[8]等。针对复杂网络的研究也逐渐发展出多个不同分支，例如链接预测^[9]、异常检测^[10]、社团检测等^[11]。

社团指的是网络中紧密连接的连通子图，是复杂网络分析中的重要统计特征之一。社团检测在现实世界复杂系统的交互模式挖掘中具有重要意义，联通了网络微观节点与宏观网络结构，能够帮助研究者理解网络的结构和功能，在链接预测、网络动力学分析、网络生成、网络控制、网络同步等复杂网络建模的下游任务中均起到了简化网络结构，提升任务效果的作用^[11,12]。

现实世界的网络往往是动态变化的，因此动态网络社团检测能够更好地帮助理解和挖掘现实世界的演化规律。动态网络社团检测指的是在动态网络中考虑时序演化信息对每个网络快照进行社团结构识别，根据任务侧重不同可以分为动态社团检测与社团演化分析。动态社团检测侧重于识别动态网络每个快照中节点的社团归属，社团演化分析侧重于社团在时序演化过程中的规律挖掘。社团演化分析往往通过预定义的各种社团事件如社团合并、分裂、扩张、缩小等行为^[13]来对连续快照的社团进行匹配与识别，一定意义上已经脱离了社团检测的范畴而更倾向于数据挖掘，其方法也大多属于启发式地提升社团匹配效果或对社团演化规律

模型考虑了社团结构在网络演化中的作用，并总结出社团边缘的节点变动对网络鲁棒性影响较大。基于结构相似性的方法以定义网络演化相似性度量实现对动态网络与动态社团演化的建模与分析，这种方式依赖于指标设计与社团检测效果，且其模型往往缺乏可解释性，不具备指标外的泛化能力。

基于预定义的方法则将社团演化行为视为网络中的各种事件，通过识别网络中的预定义事件的模式来分析动态网络及动态社团的演化。Greene等人^[85]提出了一种可扩展的社团演化事件识别方法，通过定义社团诞生、死亡、合并、分裂、扩展、收缩事件，并设计一系列阈值进行匹配来实现对社团演化事件的识别。Bródka等人^[86]则进一步将社团演化事件扩展成8种，并定义了社团归属指标 I 来衡量两个社团的相似性，基于该指标进一步定义了一系列的阈值进行社团演化事件的识别。Mazza等人^[87]则提出了无阈值框架，通过将所识别的社团结构视为节点，社团间的相似性视为边构建了社团相似性网络，并在该网络中进行模块度优化聚类，用聚类结果实现动态网络社团演化分析。Zhao等人^[88]则在考虑人类移动的空间平滑性更新了上述社团演化事件的定义，并将其应用在城市人类移动分析中。基于预定义的方法往往以预定义的事件及事件特征为基准，通过设定阈值识别网络中的事件，其对动态网络及动态社团的演化分析比较直观，但且缺乏统一标准，需要人工信息指导。

总的来说，动态网络社团演化分析方法往往依赖于社团检测准确性，难以完全脱离动态社团检测任务而独立存在。

1.3 研究内容与创新点

本文以动态复杂网络为研究对象，从动态网络社团检测与演化分析任务出发，重点关注动态网络的生成机制建模与分析，从随机块模型的角度构建动态网络的高阶生成机理，从实证分析出发，解决生成模型在动态网络演化机制中的社团与节点的演化异质性缺失问题，以提升随机块模型对动态网络演化模式的建模能力与演化分析能力，并结合图神经网络与深度序列模型探索了动态社团检测的深度生成建模方法，为动态社团检测的大规模数据可解释性建模提供基础。从研究目标上，以随机块模型为理论基础，针对真实世界动态网络数据挖掘动态网络的高阶生成机理，以支撑动态网络演化模式改进；针对动态随机块模型对节点演化异质性生成建模缺失的问题进行改进，提出动态网络的微观节点-介观社团-宏观网

络快照的层次狄利克雷生成框架，扩展动态随机块模型的建模深度；从动态随机块模型的动态网络演化分析出发，提出动态网络层次演化异常识别指标，扩展动态随机块模型的动态网络演化分析深度；针对动态随机块模型参数量过高，难以针对大规模数据进行应用的问题，提出动态网络深度生成模型社团检测建模，解决深度学习模型的黑盒问题同时，为动态网络生成模型的大规模应用提供方法范式。其研究思路及主要研究内容关系如图 1-2所示：

第一，提出动态网络高阶生成机理挖掘框架，探索动态网络节点变化与社团演化之间的本质关联。考虑到动态网络的演化由节点和边的变化驱动，而社团的演化则由节点的社团归属变化而驱动，因此需要明确动态网络中节点演化与社团演化的相互作用关系。在其指导下，研究融合动态网络高阶生成模式的动态随机块模型。针对经典随机块模型对动态网络生成机理建模能力提升的问题，设计生成机制，进一步地提升动态随机块模型的建模能力，并研究有效的动态随机块模型参数推断方法，进而研究该模型在社团检测与社团演化分析的有效性。

第二，研究动态随机块模型社团演化分析的深入挖掘方法。在研究内容二的基础上，改进并简化动态随机块模型的建模方法，提升模型参数的可解释性与易理解性，研究基于动态随机块模型的动态网络各层次演化异常定义以提升动态随机块模型的社团演化分析能力，并在真实数据中验证其有效性。

第三，针对动态随机块模型难以大规模应用问题，研究基于深度学习的动态社团深度生成建模方法。以生成模型为出发点，基于深度生成模型的框架研究生成模型参数的变分自编码器求解方法，并验证模型的动态社团检测能力及其参数的动态社团演化分析能力与大规模数据应用能力。

针对以上研究内容，本文从动态网络演化建模的角度，深入分析并提出了解决方法，取得了一定研究成果，并通过实证分析验证了提出方法的贡献。其主要贡献和创新点包括一下几个方面：

第一，本文采用特征工程的方法，将动态网络节点的社团转移行为定义成相邻快照节点是否转移的二分类问题，并提取动态网络各快照中节点与社团的属性信息，基于决策树分类实现对节点在动态网络演化过程中的高阶生成机制挖掘，通过收集大量真实世界动态网络数据进行横向对比，应用不同的经典社团检测方法剔除社团检测方法的纵向影响，最终得出结论，即动态网络的生成过程中，节点的异质性起到最主要作用。这为后续模型设计提供了实证依据。

第二，在研究内容一的支撑下，本文提出建模动态网络节点-社团演化机制的层次狄利克雷生成结构，基于动态随机块模型的基本假设，进一步引入节点级别

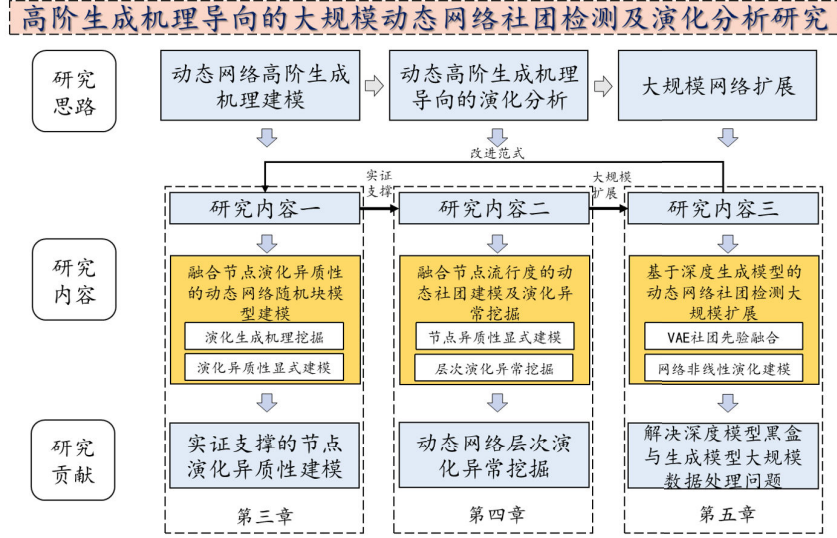


图 1-2 本文研究思路及关联

的社团转移参数，并定义了社团转移参数与节点级别的社团转移参数的生成关系。通过变分推断方法对所提出模型进行参数后验推断，以估计模型参数拟合结果。通过在生成数据集与真实数据集的社团检测、社团演化与节点演化实验，验证了所提出方法对网络高阶生成机理建模的有效性和可行性，并挖掘了论文引用网络中节点与社团演化的规律。

第三，考虑到动态随机块模型对网络生成机制的显式建模能力，本文提出了融合节点流行度的动态随机块模型以同时建模节点拓扑异质性与演化异质性，根据该模型参数和隐变量的实际意义提出了面向微观节点、介观社团以及宏观网络快照的演化异常识别指标，拓展模型的动态网络演化分析能力，通过对真实世界数据的实验验证了本模型对动态网络及动态社团的建模能力与所提出的动态网络层次演化异常指标的有效性。

第四，针对生成模型对动态网络生成及演化的可解释性建模能力强但算法效率低、模型求解方法复杂，以及动态网络深度模型建模相对简单且运行效率高（易于优化）但黑盒特性限制其对真实世界动态网络演化分析能力的问题，提出结合二者优势的深度生成模型，根据变分自编码器与生成模型变分推断求解的理论关联，提出融合变分自编码器与混合高斯的动态网络深度社团检测算法，并利用深度序列模型建模动态网络演化。针对生成数据集与真实数据集的实验验证了所提出方法的有效性以及基于深度生成模型的动态网络演化模式挖掘的可行性，为后续动态网络社团检测算法的推广与应用提供了有效的建模框架。

1.4 论文组织结构

本文对动态社团生成机制及其深度建模问题展开研究并提出了三个工作，全文的章节结构如下所示：

第一章，主要介绍了动态网络社团检测的研究背景和意义，指出了本文的研究问题，分析了当前国内外的研究现状和挑战，随后对本文研究内容进行阐述。

第二章，介绍了本文的相关研究基础，包括动态网络社团检测相关研究技术与前沿方法，包括其研究思路以及优缺点，进一步介绍了基于深度学习的动态网络社团检测方法的相关技术路线以及深度生成模型的相关基础。

第三章，针对动态网络社团演化与节点演化的关联性设计研究框架，利用大量真实世界数据进行动态网络高阶生成机理的规律挖掘，并在此实证基础上提出融合节点演化异质性的动态随机块模型，随后介绍了模型的变分推断估计方法与实验结果及实证分析。

第四章，针对动态随机块模型的演化分析能力提升，首先提出融合节点流行度的动态随机块模型以提升模型的可解释性，随后介绍基于模型参数的动态网络、社团、节点的各层次演化异常指标，最后介绍实验以验证模型与所定义指标的有效性并进行实证分析。

第五章，针对生成模型可解释性强但模型推断复杂难以应用于大规模数据的问题与基于深度模型的方法可解释性差但方法设计简单且运行效率高的问题，提出了基于深度生成模型的动态网络深度社团检测算法，随后介绍模型实验以验证所提出方法能够融合深度模型与生成模型的优势，验证了所提出方法在动态网络演化分析的应用潜力。

第六章，总结与展望，对本文的工作和贡献进行了总结，并展望未来的研究思路。

第2章 相关研究基础

本章从方法论的角度首先将当前主流的动态网络社团检测相关方法划分为基于启发式优化的方法、基于模块度优化的方法、基于矩阵分解的方法、基于谱优化的方法、基于多目标优化的方法、基于随机块模型的方法和基于深度学习的方法。并介绍与本文相关的深度生成模型研究基础。为了验证动态社团检测方法的有效性，本章还介绍了用于衡量动态社团检测效果的常用评价指标。

2.1 动态网络社团检测相关研究

动态网络的社团检测的核心任务是识别网络中每个快照的社团结构，为了更好地对动态社团进行识别，则需要建模动态网络及动态社团的演化，大量的研究人员围绕着动态网络的演化建模以及动态社团检测从不同技术路线进行创新^[89]，其中最主要的方法可以划分为基于启发式优化的方法、基于模块度优化的方法、基于矩阵分解的方法、基于谱优化的方法、基于多目标优化的方法、基于随机块模型的方法和基于深度学习的方法。各方法的优劣势及原理对比见表 2-1。需要注意的是，各类方法的设计思路往往并不单独属于动态网络建模的某一个思路如即时社团检测等。

2.1.1 基于启发式优化的方法

动态网络社团检测基于启发式优化的方法思路较朴素，即根据动态网络社团的基本定义，通过设计优化目标，再利用贪婪搜索或极值优化的方法对优化目标的解空间进行求解。

Tant等人^[110]首先根据社团的基本定义与动态网络的特性提出动态社团的启发式优化方法，该方法首先定义动态社团演化的假设，即动态网络中的节点更倾向于不频繁更换其社团归属，且大多数时间都倾向于与同一社团内的其他节点进行交互，随后证明了动态社团检测属于NP-hard完全，最后提出将动态社团检测分解

刻属于同一社团时, $\delta_{i,j}^{(t)} = 1$, 否则为 -1 。

需要注意的是, NMI 、 AC 和 ARI 仅在数据中存在社团标签时适用, 而 Q 可以在没有真实社团标签的情况下进行测量。较大的 NMI 、 ARI 和 Q 值表示社团划分结果更好, 而 AC 越小则表示社团划分结果更好。

2.4 本章小结

本章首先从方法论的角度对动态网络社团检测的主流方法进行了划分, 详细介绍了各方法的基本概念、主要研究趋势以及优劣势。总的来说, 以随机块模型为代表的生成模型方法在动态网络建模、分析以及除社团外的其他应用均具有一定优势, 但存在模型参数求解困难的问题。进一步, 本章又介绍了存在解决随机块模型缺陷的深度生成模型的主要方法对比与变分自编码器模型的详细概念。随后, 介绍了用于动态网络社团检测效果评估的主要评价指标。需要注意的是, 在本文的工作中, 其他主流方法的优秀工作也将作为对比方法在实验中进行体现。

第3章 动态网络高阶生成机理的随机块模型建模

本章主要研究动态网络演化中微观节点、介观社团、宏观网络快照的关联关系，在动态随机块模型的框架下进行建模，以实现动态社团检测与动态社团演化的同时建模。动态随机块模型是概率图模型针对快照网络建模的重要框架，随机块模型通过建模动态网络连续快照 t 到 $t+1$ 的生成关系，将节点所在社团与节点对之间的连边概率进行关联，并通过社团转移矩阵控制节点的时序演化。通过上述设置，动态随机块模型可以实现节点、社团与网络快照的演化建模。本章针对经典动态随机块模型高阶生成模式建模问题，首先探究了现实世界中节点属性与其社团演化的关联，挖掘节点演化异质性，依此设计了层次贝叶斯动态随机块模型（Hierarchical Bayesian Dynamic Stochastic Block Model, HB-DSBM）建模动态网络中节点在时序演化过程中的介观演化异质性，并通过生成数据集与真实数据集进行实验，证明了HB-DSBM在动态网络建模中的有效性与模型设计的合理性。

3.1 引言

动态复杂网络分析能够帮助人们有效地理解复杂系统的运行规律与结构功能^[154]，而动态社团检测作为动态复杂网络分析中的重要任务之一，能够帮助研究者更好地理解复杂网络的生成和演化机制^[155]。社团的一般定义是网络中紧密连接的连通子图，在现实世界中的不同网络上具有不同的现实意义^[20]。基于随机模型的复杂网络建模方法也广泛应用在社团检测任务中，如基于非负矩阵分解方法、指数随机图模型等。随机块模型由于其参数具有明确的现实或物理意义和有效的推断方法收到部分学者的青睐，并且其生成式的建模不仅可以支撑社团检测，也可以支撑如链接预测^[11]、网络生成^[156]等多种下游任务。

上述方法主要应用于静态网络建模，缺乏对于动态网络及社团的时序演化模式的刻画，无法直接应用于动态网络建模。动态社团检测主要涉及两个任务，即检测动态网络中每个快照的社团结构与建模动态网络中社团演化机制。二者相辅相成，准确刻画动态社团的演化机制可以帮助模型更精准地识别快照中的社团，

3.6 本章小结

本章首先从实际出发,通过创新性地将在动态网络的社团转移倾向异质性挖掘问题转化为节点的社团转移与否的二分类问题,对15个各类真实世界动态网络数据进行了实证分析,进而得出结论,动态网络中,节点的社团转移存在异质性,且与节点的度具有较高关联性。根据上述结论,本章进一步基于动态随机块模型设计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在刻画动态社团演化模式时同时建模节点的社团转移异质性。另外针对所提出的HB-DSBM模型,利用变分推断对模型参数的更新进行了详细的推断,利用合理的优化策略,模型具备适应大规模数据计算的潜力。同类方法的对比也显示出HB-DSBM在动态网络社团检测以及社团演化分析任务中更加高效准确。另外,本文提出的基于节点转移异质性参数提出的节点演化异常指数也在实际分析中得以验证其有效性。

然而从模型设计的角度来看,HB-DSBM仍然存在一定缺陷,首先,其对节点的社团转移异质性建模主要从生成角度而进行的设计,即假设节点的社团转移异质性与社团级别的节点转移倾向有关。具体来说,节点的社团转移异质性服从以其社团级别社团转移倾向为参数的狄利克雷分布。该设计虽然合理,但该参数并未与动态网络中的节点属性直接关联,因此在实际的下游任务中无法直接用于解释不通节点的社团转移倾向;其次,生成模型的优势之一是针对多种下游任务的适配,而所提出的节点演化异常指数还有进一步挖掘其能力的可能。因此,后续章节将针对HB-DSBM中存在的上述问题进行改进。

第4章 融合节点流行度的动态网络建模及层次异常检测

本章将针对动态网络高阶生成机理显式建模问题进一步完善基于动态随机块模型的动态网络建模方法，使其能够刻画节点拓扑及演化异质性，并从演化分析能力提升层面提出针对动态网络的演化异常检测任务提出涵盖动态网络快照-社团-节点的异常检测方法，在真实数据集中进行了应用，以验证本章提出的方法在动态网络建模与应用的有效性。

4.1 引言

生成模型通过设计动态网络的生成机制，刻画了复杂系统从宏观到微观的生成和演化模式，而社团作为动态网络生成机制中的介观结构，承载了对动态网络生成机制建模的重要角色^[175]，在对动态网络生成机制建模过程中起到承上启下作用，融入社团结构的生成模型往往能够刻画动态网络更加深层次的演化模式。例如，动态随机块模型在对网络进行建模后，可以通过社团归属参数对动态网络社团演化进行分析，挖掘所检测社团的扩张、收缩等时序变化^[48]。这使得该模型成为经典的复杂网络生成模型之一而获得了部分研究者的青睐。

近年来，部分研究者针对随机块模型在静态网络中存在的问题进行了改进。例如，针对经典随机块模型对网络中节点异质性建模缺失问题所提出的度修正随机块模型DCSBM^[176]通过引入节点的度对网络中节点的连边概率进行修正；针对动态随机块模型的重叠社团建模缺失问题所提出的混合随机块模型MMSB^[177]通过提出混合参数来刻画节点属于多个社团的场景。也有研究者考虑随机块模型在不同类型网络中的应用对其进行扩展，例如对异质网络进行适配的CVC算法^[178]、对属性网络进行适配的NEMBP算法^[179]、对超图进行适配的DCHSBM算法^[180]等进行适应性的改进与推理等。

上述方法并不适配动态网络，未考虑针对动态网络时序变化进行建模。Yang等人针对动态网络首先提出了动态随机块模型DSBM^[48]，其改进自经典随机块模型SBM，并引入社团转移矩阵来建模动态网络节点的社团划分随时间的变化；

4.5 本章小结

本章从动态随机块模型对动态网络建模的能力与演化分析能力出发提出了GEABS生成模型。首先提出了引入节点流行度参数来刻画动态网络中的节点异质性以更好地提升生成模型对动态网络的建模能力；其次提出了基于GEABS不同维度参数的网络-社团-节点不同层次的演化异常检测指标以扩展生成模型在动态网络中的演化分析能力；最后提出了基于变分推断的模型求解方法，利用变分EM方法结合拉格朗日法实现了对模型参数和隐变量的交替更新求解。实验表明，GEABS相较对比方法具有更优的动态网络建模能力，且提出的网络层次演化异常识别参数能够有效地识别动态网络节点、社团与网络快照的演化异常。通过对世界贸易网络World Trade数据集的案例研究也证明了本章所提出的层次演化异常识别指标能够在横向、纵向对比中挖掘动态网络节点、社团、网络快照之间更加复杂的交互依赖关系，能够为动态复杂系统规律挖掘提供有效支撑。

而从模型复杂度角度来看，GEABS具有生成模型的通病，即随着模型设计的精细化，其参数求解效率逐渐下降。虽然可以通过引入负采样、并行计算等方法提升模型的计算效率，但依然难以适用于大规模网络。而能够适用于大规模网络的深度学习方法由于其黑盒特性，难以像生成模型一样有效地支撑对于真实世界的规律挖掘与分析。因此在下一章，本文将提出结合生成模型与深度模型的动态网络建模方法，通过深度生成模型实现网络表示学习与生成模型的有效结合。

第5章 基于变分自编码器的动态网络深度社团检测方法

随着网络表示学习的快速发展，生成模型对网络拓扑的解耦以及对网络语义信息的聚合大大降低了大规模网络的计算复杂度，逐渐替代了现有的矩阵分解、概率图模型等方法对网络进行直接建模的方法。然而其基于图神经网络的深度学习对网络的建模是黑盒建模，即其模型中的参数并不具备现实或物理意义，而其强大的拟合能力能够较好地刻画节点的拓扑与属性信息，间接支撑复杂网络的下游任务效果，但难以为动态网络演化模式分析、网络功能结构挖掘等提供帮助。因此，本章提出基于深度生成模型的动态网络建模方法（VGRGMM），通过变分自编码器架构，结合深度序列模型建模动态网络，引入混合高斯先验来同时计算动态网络的社团结果，即提升了模型的运行效率，也为后续网络演化分析提供了可解释的中间参数用以支撑对动态网络的后续挖掘。在真实世界数据和生成数据的大量实验表明，VGRGMM模型在社团检测及社团演化分析任务中都表现出优于对比方法的效果，证明了VGRGMM模型可以在结合深度学习模型框架与生成模型的结合中融合二者的优点。

5.1 引言

动态网络由于其能够建模许多现实世界的复杂系统，长期以来一直受到研究者的关注，其能够建模包括社交网络中的人员关系^[113]、学术论文之间的引用关系与合作关系^[213,214]以及生物网络如蛋白质-蛋白质相互作用关系^[215,216]等。动态社团检测旨在发现隐藏在动态网络中的有意义的群体或簇^[40,217]，这对信息传播^[218]、链接预测^[219-221]等多种应用及动态网络规律^[222]挖掘都具有重要意义。另外，作为动态社团检测的一个重要组成部分，社团演化分析则侧重于分析社团及网络在时序演化过程中的模式，还可以揭示网络的演化特征，量化连续快照之间社团的过渡关系^[72]。因此，动态网络社团检测通过显式或隐式地建模动态网络的演化可以更好地模拟真实的复杂系统，并且相比于静态网络中的社团检测具有更多的挑战^[40]。

有5个社团的子图的变化。在快照7中，VGRGMM几乎完美地识别了快照中的社团结构，而VGRNN则错误地划分了一些紫色节点的社团结果。从快照 $t = 7$ 到 $t = 8$ 的横向对比可以看出，红色社团发生了明显变化，一些节点加入了橙色社团（从图5-10 (a) 到 (d)）。受这种节点社团演化的影响，VGRNN无法准确识别快照8的社团结果，例如节点8和88，反之，VGRGMM仍然实现了较准确的社团划分结果。同时，从社团演化的角度（即纵向比较VGRGMM和VGRNN），VGRNN将太多属于红色社团的节点转移到其他社团，这与真实值不一致。

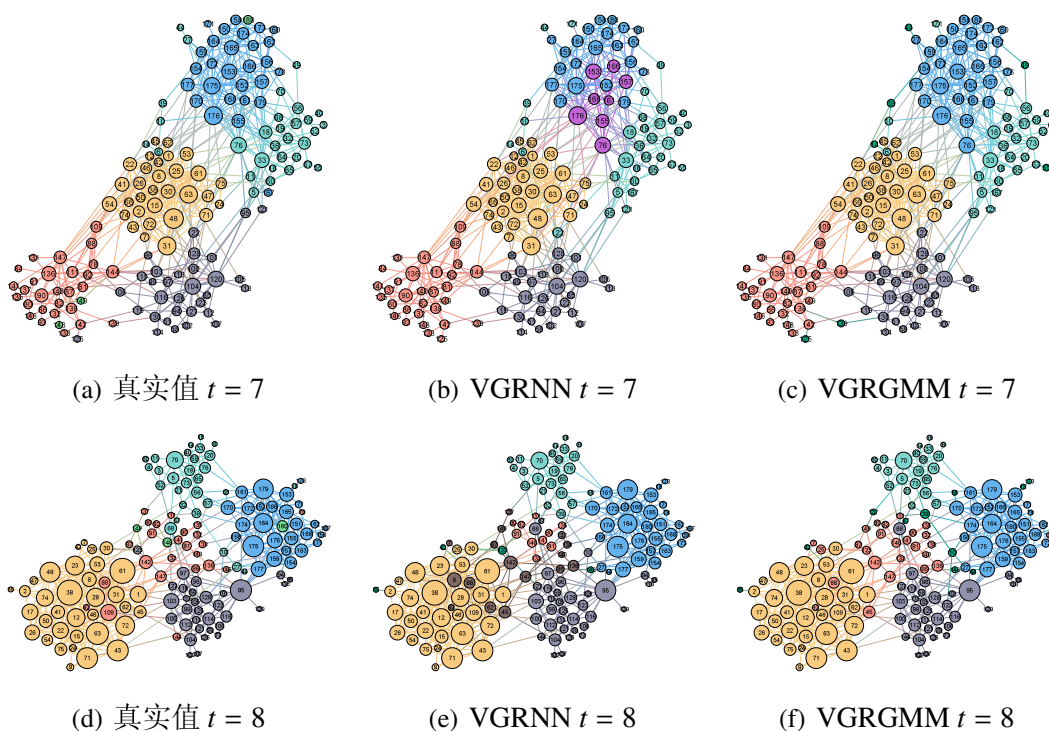


图 5-10 HS12网络在快照 $t = 7$ 和 $t = 8$ 中子图的网络可视化，不同颜色代表不同的社团。

5.4 本章小结

本章针对现有动态网络检测方法缺乏可解释性参数或无法建模大规模动态网络的问题，提出了基于变分图自编码器架构的深度动态网络社团检测算法VGRGMM。该模型首先利用变分图自编码器架构的基本设计，为深度学习模型引入了可解释可学习的参数，解决了深度模型的黑盒问题；其次，通过引入混合高斯先验，将动态网络的表示向量引入了社团结构信息，实现了对动态网络社团

检测的神经网络化，提升了社团检测方法的运行效率；最后，利用改进的GRU模型刻画了动态网络的演化模式，并建模了动态网络的非线性变化，提升了模型的动态社团检测效果。实验证明了VGRGMM相较于传统社团检测算法与动态网络表示学习方法的“两步法”社团结构识别具有更好的动态社团检测效果，对各模型的运行效率实验也证明了本方法能够适应较大规模的动态网络数据。

但该模型仍然存在一定缺陷，首先，虽然VGRGMM成功地实现了对动态网络社团检测的深度建模，验证了深度生成模型在动态网络社团检测任务的有效性，但对于超大规模网络的计算仍然存在一定困难，因此需要进一步引入优化策略来实现模型对于超大模型的建模；其次，混合高斯先验的引入实现了动态网络社团检测与动态网络表示学习的结合，但前述章节的模型证明了动态网络的演化存在层次性，且其演化更加复杂，因此需要进一步将动态网络生成模型与动态网络表示学习进行结合以更好地建模动态网络的演化模式；最后，虽然可以利用该模型参数间接的实现动态网络的演化分析，但模型本身对动态网络的非线性演化是通过GRU进行实现的，因此仍然存在模型黑盒难题，即无法可解释性地建模动态网络的演化。

第6章 总结与展望

本文研究了融合用户和商品多维度特征的个性化评论摘要生成方法，提出了一系列研究工作。本章总结了本文的相关研究成果，并对评论摘要生成未来研究方向进行了展望。

6.1 研究总结

动态网络作为建模现实世界复杂系统的重要工具，受到越来越多的学者的研究与国内外企业的应用。动态网络社团检测作为分析和挖掘复杂系统生成与演化模式的重要任务，对人们理解人类社会运行规律、挖掘各实体及其交互关系具有重要作用，基于随机块模型的动态网络社团检测方法由于其模型能够可解释地建模动态网络的生成和演化模型，因此能够有效地支撑上述任务。本文围绕“面向网络高阶生成机理的大规模动态网络社团检测及演化分析研究”问题，立足于基于随机块模型的动态网络社团检测方法，针对动态网络社团检测模型的高阶生成机理建模与网络演化分析的问题展开研究，主要工作总结如下：

- 首先，本文通过对15个不同类型真实世界网络的演化分析挖掘节点与社团的演化关系，确定动态网络的高阶生成机理。本文通过特征工程的方法，将节点的社团演化行为转化成节点在相邻快照是否发生社团转移的二分类问题，并将节点、社团级别的多种属性信息作为分类特征，基于决策树分类实现对节点在动态网络与社团演化过程中的社团标签变化的本质因素挖掘。通过各类别的真实世界动态网络数据进行实验与挖掘，最终发现动态网络演化过程中，社团对节点的影响相对较小（社团级别属性权重较低），而节点的度与平均邻居度对其社团转移存在较大影响。这证明了节点在社团演化以及网络演化过程中存在异质性且并不仅依赖节点的度。
- 其次，本文根据所挖掘出的真实世界网络的生成机理，基于动态随机块模型提出了能够建模动态网络演化层面的节点异质性的层次贝叶斯随机块模型HB-DSBM，该模型通过引入节点级别的社团转移参数，定义了社团级别

与节点级别的转移参数的生成关系。本文还提出有效的变分推断方法对该模型参数后验进行估计，并设计了参数估计的算法。通过在生成数据集和真实数据集的实验表明，该模型有效地建模了动态网络的高阶生成模式，在动态社团检测、社团演化分析与节点演化分析任务中均表现出了显著的效果。另外，本文还利用所提出的算法，针对论文引用网络的演化模式进行了案例分析。

- 随后，针对动态随机块模型的网络演化分析能力提升问题，提出了基于动态随机块模型的动态网络从此演化异常检测方法。随机块模型能够适用于各种下游任务，但其模型不具备对动态网络节点、社团及网络快照的演化异常的识别能力。针对该问题，本文首先针对前述HB-DSBM模型在对节点异质性建模过程中存在的模型设计复杂导致推断难度高、节点异质性参数建模不直观的问题引入了节点的流行度参数，通过节点的流行度参数来建模节点的异质性。随后，根据模型参数所代表的现实意义，提出了动态网络变更点（网络级别演化异常）、社团级别演化异常以及节点级别演化异常的定义与识别指标。通过对真实世界数据的实验验证了本模型对动态网络演化的建模能力与所提出的动态网络层次演化异常指标的有效性，并进一步针对世界贸易网络的各层次演化异常进行了联合挖掘，分析了从国际事件到世界贸易的级联影响。
- 最后，针对高精度随机块模型随参数规模大、参数后验推断复杂的问题，提出基于深度生成模型中的变分自编码器架构的动态网络深度社团检测算法。该算法同时融合了深度模型的数据处理能力及易于改进优化的特点与生成模型的可解释性，利用深度序列模型GRU建模动态网络的非线性演化模式，并将变分自编码器参数先验改进为混合高斯先验以检测动态网络的社团结构。在针对生成数据与真实数据的社团检测实验验证了该模型的建模能力与算法效果，并通过对真实动态网络的演化实证验证了该模型的动态网络演化分析潜力。为后续动态网社团检测算法的推广与应用提供了有效的建模框架。

6.2 研究展望

本文从生成模型的角度刻画了动态网络社团检测模型的高阶生成机理，并提出了生成模型的深度生成改进框架，主要探究了动态网络社团检测算法在建模精度、演化分析能力以及大规模数据应用的问题与挑战，提出了改进方法及实验验证。然而，结合本文工作与最新的动态网络建模技术，仍有许多工作需要进一步研究与探索：

- 近年来，有方法证明了随机块模型与谱聚类的理论关联^[41]，并提出了通过谱聚类对随机块模型参数进行求解的方法。而谱GNN^[258]的相关方法也相继被提出，根据不同卷积核设计，谱GNN的计算效率相较于GNN在空域的卷积提升巨大，另外基于谱GNN的卷积模型具有额外的理论保证与可解释能力，故结合谱GNN与随机块模型理论的动态网络社团检测算法将克服大规模数据能力限制，且能够有效支撑动态网络及动态社团的演化分析。因此，如何利用谱GNN实现对动态随机块模型的求解具有较高的研究意义。
- 本文基于深度生成模型实现了对动态网络社团检测的可解释性建模，并拓展了动态网络社团演化分析的方法边界，但其依然存在模型运行效率问题，虽然可以通过batch降低模型的空间复杂度，但会降低模型的社团检测效果。因此，如何通过蒸馏或其他模型优化方法对动态网络深度生成模型进行优化也是一项具有实际意义的研究。
- 另外，利用变分自编码器对生成模型参数进行拟合的本质是基于平均场假设对模型参数依赖进行解耦。平均场假设属于强假设，会较多地损失模型的精度，因此，如何改进变分自编码器以克服平均场假设对模型精度的影响也是一项具有挑战的研究。
- 随着大语言模型的提出，生成式人工智能对各领域均存在颠覆性的挑战。当前已经有静态网络的预训练模型通过生成式人工智能进行建模，使其具有较高的多任务泛化能力与生成能力，但针对动态网络的预训练模型仍然处于起步阶段，因此如何建模动态网络的大模型以实现对真实世界动态网络的多任务泛化、分布外泛化以及规律涌现与生成式未来亟待解决的重要研究热点之一。