本科毕业论文（设计）开题报告

动态社会网络团体发现与挖掘算法研究

An Approach to Discover and Analyze Communities in Dynamic Social Networks

1. 课题背景

社会中人与人之间的错综复杂的关系，可以抽象成一个非常复杂的网络图，每个人就是这个图中的一个节点，而人与人之间的关系就是这个图的边。人和人的关系有陌生有紧密，紧密的关系使相关人们形成一个团体。研究社会网络中的团体的发现方法和进化规律对于社会学研究和相关应用有非常重要的意义。例如，对犯罪分子的犯罪网络团体的挖掘和研究能够帮助公安机关深层次地挖掘埋藏在人际关系中的重要线索，辅助公安刑侦人员的案件侦破，对公安刑侦具有非常重要的现实意义；研究互联网社区(如微薄、论坛等)团体的进化，有助于相关组织者把握最新的流行趋势，帮助参与者，也能对商业决策提供有力的支持。

而随着时间的变化，社会网络中的团体构成随时都在发生着变化。相关的团体可能发生如分裂、合并、生长、消亡等变化。研究动态的社会网络中的团体变化和其变化规律，可以帮助社会学家研究社会发展的相关规律，帮助商家和受众掌握最新的流行趋势，可以帮助政府部门根据社会的发展制定最优决策……

社会中人与人之间的联系表现为人们的电话通信记录、邮件记录、共同参加一个活动的记录等等。使用一定的算法和数学模型对这些数据进行建模，可以把社会网络抽象为一个网络图。研究基于这个网络图上的团体发现和分析的相关算法，并开发有关的计算机程序，可以自动化地高效地分析数据，挖掘出有用的信息和规律。

1. 目标和任务

对于静态的网络进行团体发现和挖掘，已经有相当多的研究和算法，如层次聚类、k-means聚类、betweenness切片划分算法等等。这些算法能够根据不同的聚类相似度依据对静态网络进行分析并挖掘出其中的团伙。对于静态网络，也有很多的研究和方法来分析团伙内成员、团伙间的联系。关于静态网络的数据挖掘现在已经有非常成熟的知识体系和方法。但是，现有的静态网络的挖掘方法对实际分析的支持还并不充分，因为实际的社会网络，总是处在时刻变化中。如果有一套行之有效的对于动态网络的分析挖掘方法，将为相关领域提供非常大的帮助。

与静态网络不同，在动态网络中，由于节点（人）之间的联系随着时间轴的推进而不断变化，而网络结构也时刻在发生着改变。使用传统的静态分析挖掘方法就难以提供有效地对动态网络进行分析和挖掘。本次研究的目的，就是研究出一套较有效的挖掘方法和算法对动态网络进行分析，挖掘出动态网络中的团伙信息，并应用这套方法分析和研究在动态网络中的团伙进化现象及进化规律。概括之，其主要研究点包括以下几点：

1. 研究并实现一套有效的对动态网络进行挖掘和分析的方法和算法
2. 设计软件实现相关的算法，并使用数据集进行验证和分析
3. 针对数据集探讨动态网络中团伙的演化现象和演化规律

1. 研究方案和关键技术考虑

对于基于时间线的动态社会网络，

《《《研究思路》》》

《《《betweenness算法》》》

《《《fast betweenness计算》》》

《《《module度量》》》

《《《二部图的最优匹配算法》》》

《《《k团伙行为界定》》》

《《《动态的core挖掘方法》》》

1. 预期研究结果
2. 研究给出一套有效的对动态社会网络进行挖掘和分析的方法和算法
3. 开发实现出相关的算法，计划使用java为核心语言平台辅助以xml、python、Shell程序及mysql数据库实现相关的算法和数据结构
4. 开发出一套具有可视化界面的程序，把算法程序整合到可视化程序中，便于形象生动的表现动态网络的情况和算法的效果。Java swing实现
5. 使用多类型的实际数据集对算法进行验证实验。在实验中分析团伙的演化现象和规律，得出一些有意义的结论
6. 进度计划

目前开题报告的工作基本接近尾声。接下来的两个月的时间里，主要的工作就是对算法的实现和改进，以及使用实际的数据集进行验证和研究

**3.15-4.1:**

对整个软件进行设计，实现图的描述和基本的设计数据结构主要是动态图的描述。实现betweenness边算法，实现以modules聚类质量度量为依据的自动切边聚类。实现静态部分的可视化软件开发

**4.1-4.15:**

实现二部图的匹配算法，建立模型分析时间切片间的图演化。并使用数据集验证，改进分析的算法。实现动态图分析的软件可视化部分

**4.15-4.30：**

使用数据集进行验证和研究在相关数据集中团伙的演化行为。继续优化算法和软件可视化部分。可扩展研究其他的如core动态分析方法的应用

**5.1-5.20：**

1. 完成毕业论文，并且给出根据实验结果而得到的解释，以及一些算法复杂度分析参考文献：