

一个大规模股市图交互式可视化分析系统

涂志兵^{1,2} 成杰峰^{1,2} 冯圣中^{1,2}

¹ (深圳市高性能数据挖掘重点实验室 深圳 518055)

² (中国科学院深圳先进技术研究院 深圳 518055)

摘 要 随着物联网、移动互联网、云计算以及各种数据自动采集技术的迅猛发展,许多领域迅速积累了大量具有图结构的可用数据。其中一个重要的图应用是股市图。如何分析股市图达到合理充分的投资决策支持一直是一个重要的课题。其中极大团(Maximal Clique)分析是分析股市图的一个重要方法。股市图的规模庞大,传统的极大团枚举算法仅仅罗列图中所有的极大团。但一个图中可以有指数级数量的极大团,而一支股票对应的点可以参与到任意多的极大团中。因此,传统的极大团枚举算法不能直接有效支持股市图分析。本文提出一个支持快速选择、自动分组及导航浏览三种股市图交互式可视化操作的大规模股市图分析系统。根据用户感兴趣的股市图节点,这三种股市图交互式可视化操作从股市图中快速枚举出与这些特定股票相关的极大团、查看这些特定股票之间的组合关系以及显示与这些特定股票相关的其他股票,是有效支持股市图分析的必要手段。同时基于对某些特定顶点或边相关的极大团枚举的需求,本文提出了从图中枚举出与特定顶点或边相关的极大团算法。我们使用真实数据验证了本文提出的算法的优越性。

关键词 股市图;极大团;图算法;图可视化;大规模图

On Interactive Visualization for Large-Scale Stock Market Graphs

TU Zhi-bing^{1,2} CHENG Jie-feng^{1,2} FENG Sheng-zhong^{1,2}

¹(Shenzhen Key Laboratory of High Performance Data Mining, Shenzhen 518055, China)

²(Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China)

Abstract Maximal clique analysis is an important method of stock market graph analysis. Traditional maximal clique enumeration algorithms enumerate all maximal cliques in the graph, which cannot support efficient stock market graph analysis. In this paper, we propose interactive visualization methods for large-scale stock market graphs. According to user's interested stocks, we provide functions to enumerate all maximal cliques related to those stocks quickly, and to view their combination relations as well as other related stocks. Our interactive visualization methods are very useful to stock market graph analysis. Moreover, traditional maximal clique enumeration algorithms cannot be applied to support those functions. Due to the need of enumerating all maximal cliques related to specific nodes or edges, we propose a new maximal clique enumeration algorithm containing specific nodes or edges. We use real a dataset to verify the superior performance of our algorithm.

Keywords market graph; maximal clique; graph algorithm; graph visualization; large-scale graph

1 引 言

随着物联网、移动互联网、云计算以及各种数据自动采集技术的迅猛发展,许多领域迅速积累了大量具有图结构的可用数据。其中一个重要的图应用是股

市图^[1]。设 $\langle R_i \rangle$ 为股票 i 在预定的考虑时间窗内的一系列单位时间点的价格,比如一周内每天的收盘价,我们可以通过如下公式量化两支股票的价格相关性。这种相关性定量地描述了股票之间在这一时间段内价格波动一致的程度。

$$C(i, j) = \frac{\langle R_i R_j \rangle - \langle R_i \rangle \langle R_j \rangle}{\sqrt{\langle R_i^2 - \langle R_i \rangle^2 \rangle \langle R_j^2 - \langle R_j \rangle^2 \rangle}}$$

作者简介: 涂志兵, 硕士研究生, 主要研究方向为高性能计算、大规模图挖掘, E-mail: zb.tu@siat.ac.cn; 成杰峰, 博士, 副研究员, 主要研究方向为大规模图挖掘; 冯圣中, 博士, 研究员, 主要研究方向为高性能计算、云计算、生物信息学。

当 $C(i, j)$ 大于某一个给定的阈值时, 那么股票 i 和股票 j 之间在统计上具有紧密的联系。如果我们将每支股票看做股市图中的一个顶点且两支股票之间的紧密程度超过某一阈值时则表示股票图中对应的两个顶点之间存在一条边, 我们就可以得到一个表示整个股市价格波动的股市图。

如何分析股市图达到合理充分的投资决策支持一直是一个重要的课题。其中极大团分析是分析股市图的一个重要方法^[1]。极大团 (Maximal Clique) 是指任意两个顶点均相连的完全子图。它在图理论、图算法以及许多现实应用中具有举足轻重的地位。一方面, 许多基本的图问题都与它有非常紧密的联系, 比如最大独立子集^[2]、图着色^[3]、最大公共子图^[4]以及最大团^[5]等。另一方面, 在现实世界中也具有广泛的应用, 比如从社交网络中发现社区, 从蛋白质交互网络中发现蛋白质复合体等。从股市图中枚举极大团就是从股市图中寻找一组相互之间的价格波动具有紧密关系的股票, 它可以直接用于设计具有分散风险的股票投资组合。

股市图的规模庞大, 传统的极大团枚举算法仅仅罗列图中所有的极大团。但一个图中可以有指数级数量的极大团, 且一支股票对应的点可以参与到任意多的极大团中。因此, 传统的极大团枚举算法不能直接有效支持股市图分析。并且, 通常用户感兴趣的只是某一些股票, 如现有投资组合中存在的股票, 一种能支持快速选择、自动分组和导航浏览的股市图交互式可视化分析方法和系统有很大的意义。其中, 快速选择是指系统能根据用户感兴趣的一组股市图节点, 快速枚举出股市图中与这些特定股票相关的极大团; 自动分组是指系统能根据用户感兴趣的一组股市图节点查看这些特定股票之间的组合关系, 同时能组合出对应的股市图中与这些特定股票相关的极大团, 并按照这种影射通过可视化的手段加以表示; 导航浏览是指根据用户感兴趣的这组股市图节点之间关联的组合, 进一步显示相关的极大团中的其他新的股市图节点。这三种股市图交互式可视化操作是有效支持股市图分析的必要手段。

传统的极大团枚举算法并不能直接应用到这三种股市图交互式可视化操作去有效支持股市图分析, 这要求对极大团枚举算法进一步扩展。极大团枚举算法是一个经典的NP问题。极大团枚举内存算法在最坏情况下的时间复杂度被证明为 $O(3^{n/3})$ (n 为图中最大连通分量的大小)^[6]。基于对某些特定顶点或者边相关

的极大团枚举, 而不必要枚举图中所有的极大团的需求, 本文提出了从图中枚举出与特定顶点或者边相关的极大团算法。该算法具有广泛的应用, 比如在社交网络分析中, 具有相同爱好或者特征的用户之间往往具有非常紧密的联系。如果我们把每一个用户看做社交网络图中的一个顶点且它们之间存在一条边且仅当两个用户具有相同的某种爱好或者特征, 那么从社交网络图中枚举极大团就是从社交网络中寻找一组具有共同爱好或者特征的用户。寻找一组喜欢围棋的用户或者一些类似Jimmy和Bob那样在周五看NBA且周末打篮球的用户就是从社交网络图中枚举与特定用户相关的极大团。

本文第二部分介绍大规模股市图交互式可视化分析系统设计; 第三部分对从图中枚举出与特定顶点或边相关的极大团算法的问题进行描述; 第四部分介绍BK算法, 并提出与特定顶点或边相关的极大团枚举算法; 第五部分介绍可视化分析和极大团枚举算法的相关工作; 第六部分通过实验验证算法的时间性能; 第七部分是本文的总结。

2 大规模股市图交互式可视化分析系统

为了有效地支持股市图分析, 便于投资者查看各个股票之间的价格波动关系, 我们基于开源软件GraphStream^[7]开发了一个大规模股市图交互式可视化分析系统。系统中采用真实的数据源S&P 500数据集 (简称SP500), 数据集的详细信息详见表1。图1是对基于SP500数据集构建的股市图可视化后的效果图。

该系统提供了快速选择、自动分组和导航浏览这三种主要功能, 能够根据用户感兴趣的股市图节点,

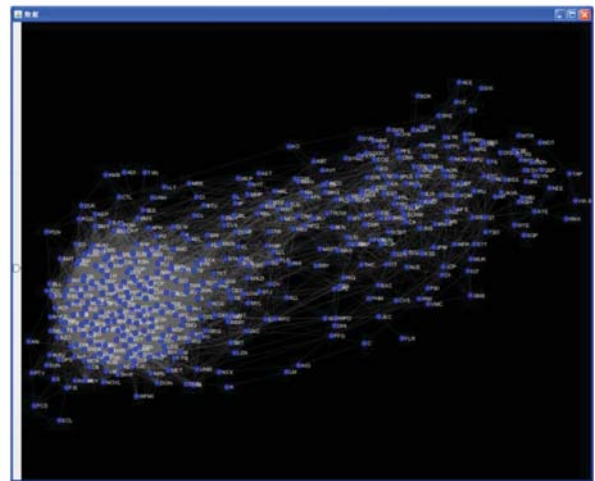


图1 SP500股市图的可视化

表1 数据集

数据集	SP500	LJ
V	523	4.8M
E	44785	25.6M

从股市图中快速枚举与这些特定股票相关的极大团，查看这些特定股票之间的组合关系，并根据这些组合关系进一步查看股市图中其他相关的股票等。

2.1 快速选择

快速选择功能能够根据用户感兴趣的一组股市图节点，快速枚举并在股市图中显示与这些特定股票相关的极大团。

为了便于用户从股市图中选择股市图节点，我们

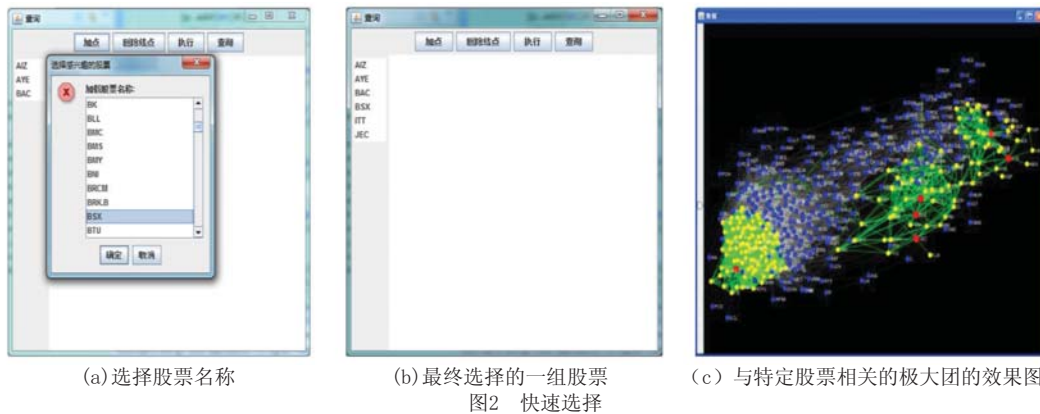


图2 快速选择

2.2 自动分组

自动分组功能能够根据用户感兴趣的一组股市图节点，查看它们之间的组合关系，并能够从股市图中组合出与这些特定股票相关的极大团。

图3中的中间图形显示部分显示了用户感兴趣的股市图节点之间的组合关系，即BSX与AYE，BAC同时与JEC和ITT具有紧密关系。

图4基于图3中6支股票的组合关系，在股市图中显示出与组合中这些特定股票相关的极大团。其中图4 (a) 和图4 (b) 表示系统根据BAC, JEC, ITT这一组合关系，在股市图中显示出与它们相关的极大团；

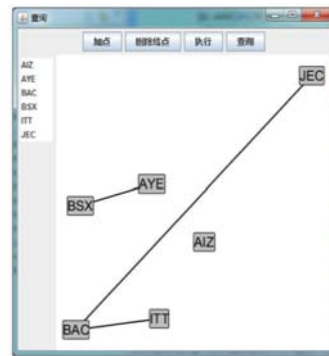


图3 特定股票节点之间的组合关系

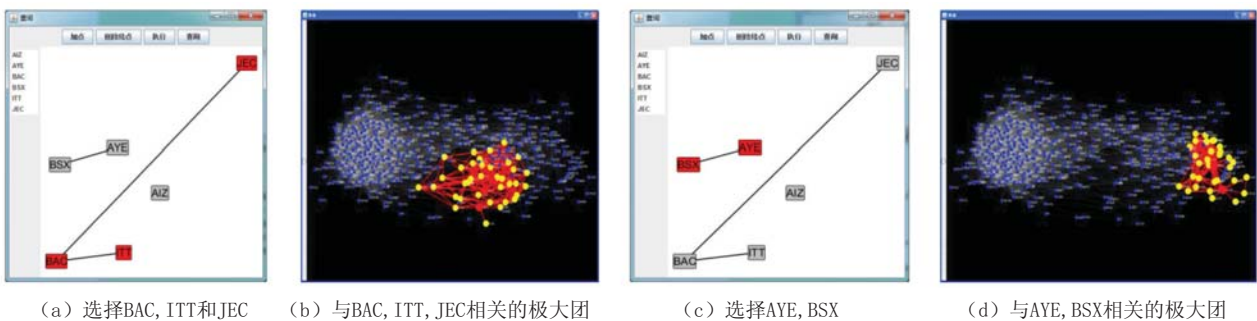


图4 自动分组与导航浏览

设计了便于用户输入感兴趣的股市图节点的界面。图2 (a) 显示了可供用户选择的股票节点，用户只需从可供选择的股票列表中选择感兴趣的股票即可。图2 (b) 左上角部分列举了用户最终选择的一组感兴趣的股票。

当选定完感兴趣的几支股票后，点击运行，系统能够从股市图中快速枚举出与这些特定股票相关的极大团。图2 (c) 在股市图中显示了与这些特定股票相关的极大团，其中标记为红色的顶点为用户感兴趣的股市图节点，标记为黄色的顶点为与这些特定股票相关且与它们能够形成极大团的股票，而标记为蓝色的顶点则表示其他不相关的股票，即不能够与这些特定股票形成极大团的股票。同时与这些特定股票相关的极大团的所有边均标记为绿色。

图4 (c) 和图4 (d) 表示系统根据BSX与AYE这一组合关系, 在股市图中显示出与它们相关的极大团。

2.3 导航浏览

导航浏览功能能够根据用户感兴趣的这组股市图节点之间的组合关系, 进一步显示相关极大团中其他的新的股市图节点。

图4 (b) 中标记为黄色的顶点即为与BAC, ITT和JEC这一组合相关的极大团中其他新的股市图顶点, 而图4 (d) 标记为黄色的顶点即为与AYE和BSX这一组合相关的极大团中其他新的股市图顶点。

3 面向分析的极大团问题描述

设 $G=(V,E)$ 是无向图, 其中 V 是图中的顶点集, E 是图中的边集。定义 $n=|V|$ 和 $m=|E|$ 分别表示图中的顶点数和边数。假设图以邻接表形式存储在内存或磁盘上, 每个顶点赋予唯一的ID, 且顶点按照其ID排好序。给定一个顶点集 S , 定义 $G_s=(V_s, E_s)$ 为由 S 中顶点所构建的导出图, 其中 $V_s=S$, $E_s=\{(u,v) \mid u,v \in S \text{ and } (u,v) \in E\}$ 。定义 $\text{adj}(v)=\{u \mid (u,v) \in E\}$ 为顶

点 v 的邻居顶点集, 顶点 v 在图 G 中的度数定义为 $\text{deg}G(v)=|\text{adj}(v)|$ 。

图 G 中的团 (clique) 是指一组顶点集 C 且 C 中任意两个顶点在图 G 中均相连。一个团被称为极大团 (Maximal Clique, 简称MC) 当且仅当不存在向当前团 C 中添加新的顶点可以形成更大的团 C' , 即不存在 $C' \supset C$ 。

与某些特定顶点或边相关的极大团枚举问题定义如下: 给定一个无向图 G 和一些由应用或用户所指定的特定顶点或边, 从图 G 中枚举出与这些特定顶点或边相关的极大团。

4 与特定顶点或边相关的极大团枚举算法

由于当前大多数极大团枚举算法几乎都是基于BK算法, 因此在介绍我们的算法之前, 我们将简要介绍BK算法。BK算法的伪代码如算法1和算法2所示。

BK算法采用回溯技术从图 $G=(V, E)$ 中枚举出所有的极大团。算法执行过程中, 为了记录已经被处理过或待处理的顶点, 同时避免枚举重复的极大团, BK算法动态地维护了三个数据结构: compsub , cand 和 not 。其中 compsub 记录了已经被处理过的且两两相互相连的顶点集, cand 记录了可能与 compsub 中顶点形成极大团的候选顶点集, not 则记录了已经处理过的顶点集且这些顶点与 compsub 中顶点形成的极大团已

算法1 BK 主算法

Input: Graph $G=(V,E)$

Output: all maximal cliques in G

$\text{compsub} \leftarrow \phi;$

$\text{cand} \leftarrow V;$

$\text{not} \leftarrow \phi;$

$\text{BK}(\text{compsub}, \text{cand}, \text{not});$

算法2 BK 递归算法

Input: $G=(V,E)$, compsub , cand , not

Output: all maximal cliques in G

if $\text{cand} == \phi$ and $\text{not} == \phi$

$\text{output } \text{compsub}$ as maximal clique.

else

$\text{pivot} \leftarrow$ vertex has maximum degree between cand and $\text{not};$

for $v \in \text{cand}$ **do**

if $(v, \text{pivot}) \notin E$

$\text{cand} \leftarrow \text{cand} \setminus \{v\};$

$\text{BK}(\text{compsub} \cup \{v\}, \text{cand} \cap \text{adj}(v), \text{not} \cap \text{adj}(v));$

$\text{not} \leftarrow \text{not} \cup \{v\};$

算法3 基本算法(baseline)

Input: $G=(V,E)$, compsub , cand , not , specverset ,

specedgeset

Output: all maximal cliques in G

if $\text{cand} == \phi$ and $\text{not} == \phi$

if $\text{compsub} \cap \text{specverset} \neq \phi$

or $\text{compsub} \cap \text{specedgeset} \neq \phi$

$\text{output } \text{compsub}$ as maximal clique.

else

$\text{pivot} \leftarrow$ vertex has maximum degree between cand and $\text{not};$

for $v \in \text{cand}$ **do**

if $(v, \text{pivot}) \notin E$

$\text{cand} \leftarrow \text{cand} \setminus \{v\};$

$\text{BK}(\text{compsub} \cup \{v\}, \text{cand} \cap \text{adj}(v), \text{not} \cap \text{adj}(v));$

$\text{not} \leftarrow \text{not} \cup \{v\};$

经被枚举出。

最初, $compsub$ 和 not 均为空, $cand$ 包含了图 G 中所有的顶点, 也就是, $compsub=not=\phi$, $cand=V$ 。接着从 $cand$ 和 not 中选择一个与其他顶点具有最大度数的顶点作为枢纽顶点。算法将基于此枢纽顶点和 $cand$ 集合中与此枢纽顶点不相连的顶点继续调用自身。而对于上述每一个顶点, 算法在调用之前将根据 $compsub$, $cand$ 和 not 这三种数据结构的定义产生新的 $compsub$, $cand$ 和 not , 并以它们作为参数调用自身。该算法将递归调用自身直到 $cand$ 为空, 即没有候选顶点。如果此时 not 也为空, 则 $compsub$ 中的顶点就构成了一个极大团。否则, $compsub$ 中的顶点是已经枚举出的极大团中的部分顶点, 算法立即停止。

根据之前的分析, 大多数应用并不需要枚举出图 G 中所有的极大团, 而是与某些特定顶点相关的极大团。因此基于BK算法, 我们提出了与特定顶点或者边

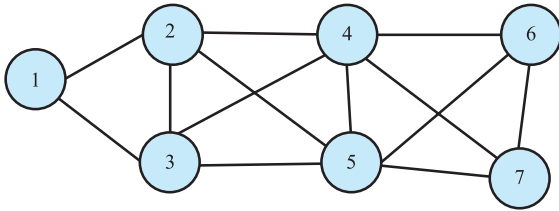


图5 节省极大团枚举代价的示例

算法4 与特定顶点集相关的极大团枚举算法 modBK4

Input: $G=(V,E)$, $compsub$, $cand$, not , $specverset$,

Output: all maximal cliques in G

if $cand == \phi$ and $not == \phi$

output $compsub$ as maximal clique.

else

if $|compsub| == 0$

$candidates \leftarrow cand \cap specverset$;

else

$candidates \leftarrow cand$;

$pivot \leftarrow$ vertex has maximum degree between

$cand$ and not ;

for $v \in candidates$ **do**

if $|compsub| > 0$ && $(v, pivot) \in E$

continue;

$cand \leftarrow cand \setminus \{v\}$;

$BK(compsub \cup \{v\}, cand \cap adj(v), not \cap adj(v))$;

$not \leftarrow not \cup \{v\}$;

相关的极大团枚举算法。

直觉上来说, 为了获取与这些特定顶点或者边相关的极大团, 仅需要检查所有已经枚举出的极大团并找出至少包含这些特定顶点中一个顶点或者这些特定边中一条边的极大团即可。因此基本算法是首先从图 G 中枚举出所有的极大团, 然后依次检查每一个极大团是否包含特定顶点集中至少一个顶点或特定边集中至少一条边。如果该极大团至少包含某一特定顶点或边, 则此极大团就是用户所感兴趣的极大团。算法3中描述了该算法, 其中 $specverset$ 和 $specedgeset$ 分别指应用或用户所指定的一组特定顶点集和边集。

然而, 我们的目标是得到所有与某些特定顶点或边相关的极大团而非所有的极大团。因此没有必要枚举出不相关的极大团, 即不包含任何特定顶点或边的极大团, 且耗费大量的时间。根据极大团的定义, 我们可以在算法执行之前从图 G 中过滤满足以下条件的顶点: 不与特定顶点或边集中任何一个顶点相连。换句话说, 我们仅需要从特定顶点或边以及它们邻居顶点所构成的子图中枚举极大团即可。

但是, 仅仅满足上述条件还不够。我们仍可以得

算法5 与特定边集相关的极大团枚举算法 modBK5

Input: $G=(V,E)$, $compsub$, $cand$, not , $specedgeset$,

Output: all maximal cliques in G

if $cand == \phi$ and $not == \phi$

output $compsub$ as maximal clique.

else

if $|compsub| == 0$

$candidates \leftarrow cand \cap specedgeset$;

else if $|compsub| == 1$

$candidates = \{v | (u, v) \in specedgeset$

and $u = compsub[0]\}$;

else

$candidates \leftarrow cand$;

$pivot \leftarrow$ vertex has maximum degree between $cand$

and not ;

for $v \in candidates$ **do**

if $|compsub| > 1$ && $(v, pivot) \in E$

continue;

$cand \leftarrow cand \setminus \{v\}$;

$BK(compsub \cup \{v\}, cand \cap adj(v), not \cap adj(v))$;

$not \leftarrow not \cup \{v\}$;

到不包含任何特定顶点的极大团, 因为这个极大团完全是由这些特定顶点的邻居顶点所构成。例如, 在图5中, 如果仅需要得到包含顶点1或者6的极大团, 从由顶点1和6以及它们的邻居顶点所构成的子图中枚举出的极大团 {2, 3, 4, 5} 却不包含顶点1或者6。

为了避免枚举出上述无关的极大团, 仅需要从特定顶点集中的顶点或特定边集中的边开始枚举极大团即可, 即确保第一个用于扩展成极大团的候选顶点必须是应用或用户所指定的顶点或边。随着算法的进行, 将得到所有与这些顶点或边相关的极大团。枚举出与特定顶点集相关的极大团算法如算法4所示。

枚举与特定边集相关的极大团枚举算法如算法5所示。与算法4不同的是, 算法5不仅需要确保第一个枢纽顶点选自特定的边集中某一条边的顶点, 而且还需要确保第二个顶点也必须选自特定边集中的某一条边的顶点同时前后两个顶点所形成的边必须来自于特定的边集。

在许多真实应用中, 处理的图规模太大以至于难以完全将整个图装入机器内存。同时大多数大规模图满足幂律特性: 少数顶点与图中大多数顶点均相连, 即少数顶点的度数非常高。根据此特性, 如果指定的顶点就是这些度数非常高的顶点, 那么由这些度数高的顶点及其邻居顶点所构建的子图将与原图具有同等大小。正如前面所说, 当算法执行过程中所需访问的图不能完全装入机器内存时将会频繁访问磁盘从而导致很高的I/O开销。因此为了能够处理这种大规模图, 必须首先从原图中抽取由部分特定顶点集或边集中顶点或边及其邻居顶点所构成的子图且确保该子图能够刚好装入机器内存, 接着基于此子图寻找与这些特定顶点或边相关的极大团。下一步则继续从原图中抽取与剩余特定顶点或边及其邻近顶点所构成的子图, 并重复上述过程, 直至处理完特定顶点集中所有顶点或特定边集中所有边。

5 相关工作

图数据可视化技术 (Graph Visualization) 在众多数据分析工具中具有重要的地位, 尤其是面临海量数据时, 能够帮助人们理解这些大规模图数据并进一步挖掘相关数据所隐含的重要特性。

图数据可视化的必要手段是对图结构和交换关系网络的可视化技术, 通常基于所谓的力驱动模型去对图布局。即考虑图中任意两点间都存在排斥力的情况

下, 去计算图中所有节点在空间中的稳定位置坐标让系统势能最小。Tutte^[8]和Eades^[9]分别在1963年和1984年, 奠基了力驱动模型的布局方法。本文股市图的可视化表示采用的是力驱动模型的布局方法。

图可视化方法还与具体的数据挖掘算法有关, 需要针对数据挖掘算法的特征设计可视化方法和工具。目前典型的工具和方法有慕尼黑大学的PBC (Perception Based Classification) 交互式决策树分类器^[10]、Prefuse^[11]、Xmdvtool^[12]等。Huang等人在2000年提出交互式聚类算法和分类算法^[13, 14]。

在交互式可视化网络数据分析系统方面, 浙江大学汪加才等人构建了一个交互式可视化数据挖掘原型系统VisMiner^[15], 用户可从SOM (Self-organization Maps) 标记图或距离图中交互选择感兴趣区域进行分析, 但该系统不支持大规模、复杂的网络数据分析。

极大团枚举算法一直是图理论和图算法的研究重心。近些年来, 随着社交网络的兴起, 研究大规模图上图问题吸引越来越多人的注意, 而极大团在图算法以及图理论中无可替代的地位也使它成为研究的热点。

当前最有效的极大团枚举算法被证明均与BK算法^[16]有关。BK算法是由Bron和Kerbosch于1973年提出, 该算法采用回溯技术限制搜索空间以提高算法效率。但是在算法执行过程中它需要访问图中任意顶点或者边, 当图的规模大到机器内存不足以完全加载时将频繁地访问存放在磁盘上的图信息, 从而导致很高的I/O开销。

为了解决大规模图上极大团枚举算法具有很高的I/O开销的问题, Cheng等人提出了基于H-graph^[17]或B-graph^[18]的极大团枚举算法。H-graph或B-graph是从原图中抽取的一个子图且刚好能够装入机器内存, 这样算法执行过程中就不必去访问磁盘中的图信息, 避免了过多的I/O开销。同时从H-graph或B-graph中枚举出的极大团就是原图中的极大团。

为了能够处理更大规模的图并进一步减少算法的执行时间, 人们开始提出了许多极大团枚举的并行算法。Schmidt等人^[19]通过引入candidate path这种新的数据结构将BK算法执行过程中所构建的回溯树分解, 这样增加了并行度, 从而将BK算法并行化。而Cheng等人^[20]进一步借鉴H-graph和B-graph的思想且根据传统极大团枚举算法中频繁执行集合交集操作这一特性, 提出了可扩展种子子图和抽取更小子图的代价模型从而分别降低了I/O开销和CPU开销。

6 实验分析

本节中我们将评估算法的性能。将算法(modBK4、modBK5)与基本算法(baseline)做比较。算法用C++语言实现,实验的运行环境如下:8核Intel Xeon E5506 2.13GHz CPU, 16GB RAM的机器,Ubuntu操作系统, g++ 4.6.3编译器。

为了比较算法的时间性能,我们使用以下两个数据集:S&P 500股票数据和LiveJournal数据集(简称LJ)。SP500是收集了S&P 500股市中523支股票在2009年8月至2010年8月内247天的收盘价,并基于此数据生成了对应的股市图。LJ数据集^[21]则是由社交网

络所构建的图,其中图中每个顶点代表一个社交网络成员,而边则表示他们之间是朋友关系。表1列出了这两种数据集的详细信息。

从LJ数据集中随机选取4组大小分别为10, 20, 50, 100的顶点集和边集来比较算法modBK4, modBK5与基本算法(baseline)的运行时间。表2中基于LJ数据集比较了枚举与特定顶点集相关的极大团算法的运行时间。表3中也是基于LJ数据集比较了枚举与特定边集相关的极大团算法的运行时间。从上述两个表中可以看出,算法modBK4和modBK5均比基本算法节省了大量的运行时间。

表2 枚举与特定顶点集相关极大团运行时间(seconds)

specvset	10	20	50	100
baseline	4845.261	4936.650	4978.466	5041.975
modBK4	2.760	2.758	634.711	645.321

表3 枚举与特定边集相关极大团运行时间(seconds)

specedgeset	10	20	50	100
baseline	4913.454	4939.360	5537.617	6675.338
modBK5	4.074	4.175	4.279	4.595

7 总结

本文构建了一个支持快速选择、自动分组以及导航浏览功能的大规模股市图交互式可视化分析系统,它能有效地支持股市图分析。并基于对某些特定顶点或边相关的极大团枚举的需求,提出了从图中枚举出与特定顶点或边相关的极大团算法。经实验证明,该算法能够极大地减少运行时间。

参考文献

- [1] Boginski V, Butenko S, Pardalos P M. On structural properties of the market graph [J]. Innovation in Financial and Economic Networks, 2003: 29-45.
- [2] Tsukiyama S, Ide M, Shirakawa I. A New algorithm for generating all the maximal independent sets [J]. Sociaety for Inustrial and Applied Mathematics Journal on Computing, 1977, 6(3): 505-517.
- [3] Byskov J M. Algorithms for k-colouring and finding maximal independent sets [J]. Symposium on Discrete Algorithms, 2003: 456-457.
- [4] Koch I. Enumerating all connected maximal common subgraphs into two graphs [J]. Theoretical Computer Science, 2001, 250(1-2): 1-30.
- [5] Tarjan R E, Trojanowski A E. Finding a maximum independent set [J]. Sociaety for Inustrial and Applied Mathematics Journal on Computing, 1977, 6(3): 537-546.
- [6] Moon J W, Moser L. On cliques in graphs [J]. Israel Journal of Mathematics, 1965, 28(1): 23-28.
- [7] Dutot A, Guinand F, Olivier D, et al. GraphStream:a tool for bridging the gap between complex systems and dynamic graphs [J]. In Emergent Properties in Natural and Artificial Complex Systems, 2007.
- [8] Tutte T W. How to draw a graph [J]. Proceedings in London Mathematical Society, 1963, 13(52): 743-768.
- [9] Eades P. A heuristic for graph drawing [J]. Congressus Numerantium, 1984, 41: 149-160.
- [10] Ankerst M, Ester M, H P K. Towards an effective cooperation of the user and the computer for classification [J]. Knowledge Discovery and Data Mining, 2000: 179-188.
- [11] the prefuse visualization toolkit [EB/OL]. <http://prefuse.org/>.
- [12] Xmdv Tool [EB/OL]. <http://davis.wpi.edu/xmdv/>.
- [13] Huang Z, Lin T. A visual method of cluster validation with fastmap [J]. Proceedings of Pacific Asia Knowledge Discovery and Data Mining, 2000.
- [14] Z. Huang, M. NG, T. Lin and D C. An interactive approach to building classification models by clustering and cluster

- validation [J]. Proceedings of the Second International Conference on Intelligent Data Engineering and Automated Learning, Data Mining, Financial Engineering, and Intelligent Agents, 2000: 23-28.
- [15] 汪加才, 陈奇, 赵杰煜, 等. VISMiner: 一个交互式可视化数据挖掘原型系统 [J]. 计算机工程, 2009, 29: 17-19.
- [16] Bron C, Kerbosch J. Algorithm 457: finding all cliques of an undirected graph [J]. Communications of the ACM, 1973, 16(9): 575-577.
- [17] Cheng J, Ke Y, et al. Finding maximal cliques in massive networks by h*-graph [C] // In Special Interest Group on Management Of Data Conference, 2010: 447-458.
- [18] Cheng J, Ke Y, et al. Finding maximal cliques in massive networks [J]. ACM Transactions on Database Systems, 2011, 36(4): 21.
- [19] Schmidt M C, Samatova N F, Thomas K, et al. A scalable, parallel algorithm for maximal clique enumeration [J]. Journal of Parallel and Distributed Computing, 2009, 69(4): 417-428.
- [20] Cheng J, Zhu L, Chu S. Fast algorithms for maximal clique enumeration with limited memory [J]. In Knowledge Discovery and Data Mining, 2012: 1240-1248.
- [21] Snap [EB/OL]. stanford.edu.cn.