快速的最大二分团枚举算法

秦才霞，廖名学，梁媛媛，郑昌文

论文题目：Efficient Algorithm for Maximal Bicliques Enumeration on Bipartite Graphs

发表的刊物/会议名称：ICNC-FSKD

联系方式：1055726610@qq.com

科研背景：图用来给数据之间的关系建模，已经被广泛运用到了数据挖掘领域，比如网页挖掘，生物信息学，顾客市场购物篮分析等。在图挖掘领域已经有了很多现有的算法用来发现数据之间的稠密关系，比如挖掘密集子图，挖掘频繁子图。最大二分团枚举是一个用来在图中发现各个顶点之间的关系的重要工具。在社交网络中，网页社区采用二分图建模，并且可以通过识别来自网络的最大二分团以找出相互作用的客户社区。二分图可以用来分析网页查询搜索的结果。

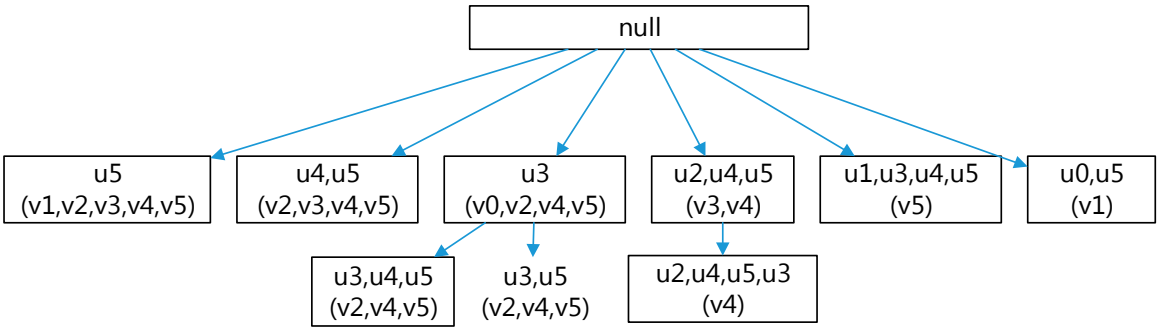

Figure 节点拓展树形图

创新点：最大二分团枚举或者闭合项目集挖掘是一个用来发现图的隐含信息的基础工具，他可以被应用在社交网络分析，购物推荐系统，生物信息学等领域。最近几年，现实世界中很多场景产生的数据集都十分巨大，因此诞生了很多并行算法用来列举最大二分团，从而加快挖掘速度。然而，每个算法都不可避免的要使用到基础的串行MBE算法来挖掘最大二分团，所以，对串行MBE算法的优化依旧是一个很重要的基础工作。本论文提出了一个最大二分团枚举的高效算法EMBE。EMBE用深度优先的搜索方式来列举最大二分团，并且在迭代过程中，不需要保存之前已经发现的最大二分团。MBE中的一个重要步骤就是检查顶点的闭合性，我们的论文目标是尽可能减少检查顶点闭合性的次数。论文提出1）在检查顶点闭合性时，剪枝算法使用了一个新的基于堆栈的实现。2）通过使用一个全局数组维护候选子节点，节省保存候选子节点的需要的时间和内存3）优先拓展邻居较多的子节点使得迭代树更加平衡。4）根据原始数据库的密集程度，设计了合并相同顶点和不合并相同顶点两种算法。为了评价衡量我们的算法，我们把我们的高效的算法在一些经典数据集和合成数据集上与一些著名的算法作对比。最后通过一些合成数据的相关实验证明，我们的算法比目前的最好算法快。