

适于社会网络结构分析与可视化的布局算法*

吴 鹏⁺, 李思昆

(国防科学技术大学 计算机学院, 湖南 长沙 410073)

Layout Algorithm Suitable for Structural Analysis and Visualization of Social Network

WU Peng⁺, LI Si-Kun

(College of Computer, National University of Defence Technology, Changsha 410073, China)

+ Corresponding author: E-mail: wupenny@vip.sina.com

Wu P, Li SK. Layout algorithm suitable for structural analysis and visualization of social network. *Journal of Software*, 2011, 22(10): 2467-2475. <http://www.jos.org.cn/1000-9825/3896.htm>

Abstract: In information visualization of social networks, force directed layout algorithms, which enable the creation of node-link diagrams of huge-graphs, are the most popular, however, they are not quite suitable for structural analysis and visualization which often produce images where nodes clump together in the center of the screen, making it hard to discern structural features. This paper proposes a Subgroup Analysis Layout (SAL) algorithm to solve this problem, which plots out and analyses the subgroups in social networks through the analysis of roles and key attributes. Then, the results of subgroup analysis are used to improve the force directed layout algorithm in both 2D and 3D visualization. Results with the case of terrorist organization information show that SAL algorithm can be excellent in analyzing and displaying the structure of social network.

Key words: information visualization; social network; force directed layout algorithm; structural analysis; subgroup analysis

摘 要: 力导引布局算法是社会网络信息可视化领域的主流布局算法,但却在结构分析显示方面存在缺陷.产生的图像经常是节点聚集在屏幕中央,难以分辨结构信息.因此,提出了子群分析布局 SAL(subgroup analysis layout)算法来解决这一问题.该算法通过角色分析和关键属性分析,对社会网络中的子群进行划分与分析,然后根据子群分析结果对力导引布局算法加以改进,包括 2D 和 3D 可视化.以对恐怖组织信息进行分析作为实例,也验证了 SAL 算法在社会网络结构分析与显示方面具有一定的优越性.

关键词: 信息可视化;社会网络;力导引布局算法;结构分析;子群分析

中图法分类号: TP391 **文献标识码:** A

社会网络信息可视化是当前的热门研究领域,它作为社会网络分析的重要手段,在反恐、国家安全、军事等方面具有广阔的应用前景.社会网络信息可视化的技术核心主要是布局算法,即如何在可视化空间对社会网络信息进行分布显示.目前,社会网络信息可视化领域使用的布局算法主要有:基于节点连接图表示的布局算法,如树形布局、层次布局和弹性布局等^[1];还有邻接矩阵图表示^[2,3]方法.在上述的布局算法中,弹性布局又称为力导引布局算法(force directed layout algorithm).经过近年的发展,它逐渐成为了布局算法运用的主流,能够提供

* 基金项目: 国家高技术研究发展计划(863)(2006AA01Z451, 2007AA01Z474, 2007AA010502)

收稿时间: 2010-01-14; 定稿时间: 2010-06-10

快捷的可视化实现和满意的可视化结果^[4].然而,传统的力导引布局算法把社会网络图中的节点和边看成是物理作用力系统,没有涉及到社会网络分析理论.这样固然实现简单、应用广泛,但却缺乏对社会网络信息的深层次理解.因此,它难以满足对社会网络结构进行深入分析与显示的需求;而且对具有小世界特性的社会网络来说,由于算法将相互连接的节点放置在很靠近的位置,因而产生的图像中节点经常聚集在屏幕中央,很难分辨诸如群体分布、节点角色等结构信息.

而社会网络分析的一项重要任务就是揭示社会的结构.在社会网络分析学中,社会网络是指社会行动者(actor)及其间的关系的集合^[5],凝聚子群(cohesive subgroup)则是一个行动者子集合,该集合中的行动者之间具有较强的、直接的、紧密的、经常的或者积极的关系^[6].进行子群的划分、比较,甚至进一步对行动者在子群内的角色、影响进行分析,有助于分析社会网络结构,对于发现潜在关系与目标、分析团伙行为、关注与布控关键目标等实际应用都具有重要意义.

因此,本文抓住社会网络结构分析的关键——子群,提出了子群分析布局 SAL(subgroup analysis layout)算法.SAL 算法包括两部分:子群分析过程与算法布局过程.首先,利用社会网络分析领域的思想与理论对社会网络信息进行子群分析;然后,在此基础上对传统力导引布局算法进行改进,克服其不适合于社会网络结构分析与显示的缺点.

1 力导引布局算法研究现状

力导引布局算法的基本思想是,将节点连接图看成一个顶点为钢环、边为弹簧的物理系统,当系统被赋予某个初始状态后,弹簧弹力的作用会导致钢环的移动,这种运动直到系统总能量减少到最小值时停止.这一模型易于理解和实现,因而出现了一系列的改进算法.例如,Chan 等人提出的 ODL(outdegree layout)算法^[7],采用分级思想,以节点的出度为依据将其划分为不同层次,首先挑选重要的点集进行力引导计算,然后逐步添加其余的点到图中,最终形成完整的结果.针对恐怖活动信息进行可视化的 OntoVis 系统^[8],采用改进的力导引布局算法,并运用本体图帮助进行语法提取、结构提取和重要性过滤,以便降低网络的复杂度.经过多年的发展,力导引布局及其改进算法已成为社会网络信息可视化领域应用最为广泛的算法.从各种力导引布局算法的平面可视化结果图中,例如图 1,能够看到相对聚集的节点集群,但是整个图像中节点聚集在屏幕中央,节点结构和路径难以识别,对于子群特别是子群内部的结构难以进行深入分析.因此,这类算法适合于表现网络的整体分布,但对社会网络局部结构的分析与显示支持不够,还需加以改进.

而关于力导引布局算法从二维到三维的扩展,Ware 认为^[9],从人类的空间感知力角度考虑,采用两维半的设计理念可以更好地观察和理解网络.其方法是,节点将只分布于三维空间中挑选出的若干布局平面上,而在单个平面上仍然采用经典的二维布局算法.Ahmed 等人也由此提出了一种三维可视化方法^[10],他们将各种具有不同重要级别的节点放置在平行的平面或者球面上,在各个平行面上采用改进的力导引算法进行布局.如图 2 所示,连接度被作为节点重要性的划分标准,因而适于表现无尺度网络特性,也便于从总体角度迅速掌握网络中的关键节点与连接度层次分布,但图像中节点高度聚集,节点结构和路径难以识别,也无法分析和显示诸如群体分布、节点角色等结构信息,因此在社会网络结构的分析与显示方面存在缺陷,需进一步加以改进.

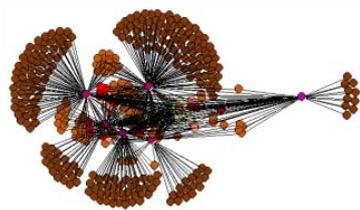


Fig.1 Layout graph of ODL algorithm^[7]
for a BGP routing topology network

图 1 BGP 路由拓扑网络 ODL 算法^[7]布局结果

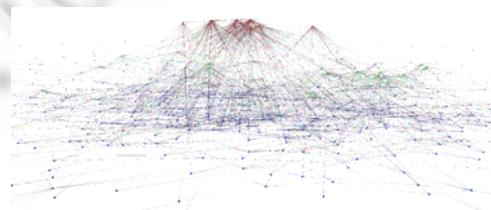


Fig.2 Parallel plane layout of citation network
produced by Ahmed's algorithm^[10]

图 2 引用与合作网络 Ahmed 算法^[10]平行平面布局结果

国内对于信息可视化算法也有一定的研究,如武汉大学的黄竞伟等学者^[11]将一般无向图的画图问题转化为函数优化问题,用遗传算法求目标函数的最优解的近似值.该方法容易实现,易于并行化.《软件学报》2008年的“软件可视化与信息可视化”专刊则集中展现了国内以及部分国外学者在软件可视化与信息可视化方面的最新研究成果,其中,任磊等学者提出了一种模型驱动的交互式信息可视化开发方法^[12],该方法采用 IIVM 建模与描述文件生成方法、系统自动生成方法,为交互式信息可视化开发的统一支撑方法问题提供了一种有效的解决方案.国防科学技术大学图形图像研究团队基于本体论对社会网络信息可视化技术进行了研究^[13],该项研究得到了国家 863 计划的支持,并在与国安局的合作项目中得到应用.总的来说,国内在信息可视化,特别是在社会网络信息可视化方面的研究较少,目前大多是综述类的文献^[4],尚无社会网络三维可视化的研究成果.

2 SAL 算法的子群分析过程

社会网络结构分析是指分析社会行动者之间实存或者潜在的关系模式^[5].与子群相关的分析,如分析网络中存在多少子群、每个子群之间是什么关系、子群内部成员之间的关系具有怎样的特点等等,是社会网络结构分析的关键任务.子群分析可以帮助理解社会网络的特性,例如,我们可以期望两个有重叠的子群,其成员间的关系紧密一些,因而冲突与矛盾可能会少一些.子群分析还有助于理解行动者的行为,例如:有的行动者在子群之间起到桥梁的作用,有的行动者与一个子群内得到所有行动者都有关系,有的却是一个相对孤立者.行动者嵌入网络的不同方式对其行为和看待问题的方式将产生重要影响.因此,子群分析对于发现潜在关系与目标、分析团伙行为、关注与布控关键目标等实际应用都具有重要意义.

SAL 算法抓住社会网络结构分析的关键——子群,来改进力导布局算法.在先期的子群分析过程中,主要利用角色分析与关键属性分析来帮助进行子群的划分与分析.

2.1 基于角色分析的子群划分

角色分析是指通过行动者的网络位置分析其在社会网络中的角色、地位^[5],角色分析也是社会网络结构分析的任务之一.角色分析通常按照对等性的思想来进行,而对等性是指,如果两个行动者具有对等的网络位置,则它们很有可能在群体中的角色相当.如果对关系网络中的行动者进行分组,使得组内的任意两个行动者都具有相同的连接集合,这样的组称为规则对等组.组内行动者是规则对等的,或者说在社会网络中具有类似的角色. SAL 算法的子群分析过程,首先按照规则对等思想对社会网络信息进行分组,这样既完成了角色分析任务,同时也对社会行动者之间的连接关系进行了梳理,方便进行下一步的子群划分.

迭代划分规则对等组时,首先将行动者分为几个大的初始组,初始组的划分可采用连接度作为依据.行动者之间的规则相关性初始取值,同组内取 1,不同组取 0.每次迭代计算后,组内规则相关性小于某一阈值的行动者形成新的分组,最终迭代结束后,每组内规则相关性都大于某一阈值.第 $t+1$ 次迭代,行动者 i 和 j 之间的规则相关性计算公式为

$$M_{ij}^{t+1} = \frac{\sum_{k=1}^g \max_{m=1}^g M_{km}^t \times \min(x_{ik}, x_{jm})}{\sum_{k=1}^g \max_{m=1}^g \max(x_{ik}, x_{jm})} \quad (1)$$

公式(1)表明,行动者 i 和 j 之间的规则相关性由它们与其他行动者的连接关系来衡量.分子在所有 g 个行动者中对每个行动者 k ,计算与行动者 i, k 的连接最为匹配的行动者 j, m 连接.由于 k 和 m 未必是完全的规则对等,所以还必须用上一次的估计值 M_{km}^t 来衡量.公式(1)计算出的是相对值,取值范围从 0~1,分母表示匹配的最大可能值.

由规则对等性的定义可知,规则对等组的计算隐含了对行动者之间邻接以及可达性关系的分析.根据子群划分的 4 个主要角度^[5]来看,下一步要得到子群划分,还需要对子群内部成员之间关系的频次、子群内部成员之间的关系相对于内外成员之间的关系密度进行考察.因此,首先假设每个规则对等组都是一个子群,然后进行各个子群之间的邻接性子群分析计算、子群内外部关系密度分析计算,对关系密切的子群进行合并,合并操作直到各个子群内部凝聚、外部关系稀疏为止,这样就通过规则对等组得到了行动者的子群分组.

2.2 基于关键属性的子群分析

基于角色分析得到的是子群的大致划分,可能还会有一部分行动者仅从关系角度分析很难准确判断其所属于子群.而社会网络分析学的研究表明,凝聚子群内部的行动者通常会具有很多共性.换句话说,具有相同关键属性的行动者常常会存在紧密联系,从而属于同一子群.因此,本文通过计算分析行动者的关键属性与子群分组的相关性,即判断两个行动者某关键属性相同时有多大概率属于同一个子群,来帮助进行子群的划分并研究子群间的密切程度.

采用社会网络分析学中的 QAP(quadratic assignment procedure)方法,可以定量地计算关键属性与子群分组的相关性.这是一种计算矩阵相关性的方法,因此,定义同子群矩阵 Q 为矩阵元素 q_{ij} 当行动者 i, j 属于同一子群时为 1, 否则为 0; 类似地,对行动者的每个关键属性定义一个同属性矩阵 A , 矩阵元素 a_{ij} 当行动者 i, j 的属性相同时为 1, 否则为 0. 采用 QAP 方法计算各个矩阵 A 和 Q 之间的相关性, 得到行动者的关键属性与子群分组的相关性, 这一相关性表明两个行动者某关键属性相同时有多大概率属于同一个子群. 因此, 设行动者的各个关键属性相对独立, 得到所有关键属性与子群分组的相关性后, 可以计算某一行动者 i 属于某一子群 j 的统计概率为

$$BS = \max_{m=1}^g q_m f_m \quad (2)$$

其中, BS 表示行动者 i 属于子群 j 的统计概率, q_m 表示第 m 个同属性矩阵与同子群矩阵的相关性, f_m 表示行动者 i 与子群 j 在第 m 个属性上的相似程度. 当公式(2)计算出的概率大于某一较高阈值时, 可以考虑将 i 合并入子群 j , 而子群间的相似度也可以根据关键属性分析计算为

$$SF = \frac{\sum_{m=1}^g q_m f_m}{\sum_{m=1}^g q_m} \quad (3)$$

SF 表示相似度, 其余变量定义如前所述.

3 SAL 算法的布局过程

SAL 算法的布局过程对力导引布局算法加以改进, 调整行动者间的适宜距离, 采用分级思想进行布局计算, 在三维可视化时, 依据相对中心度, 能够充分利用并展现子群分析的结果.

3.1 SAL算法的平面布局

在子群分析的基础上, SAL 算法对传统力导引布局算法的改进主要表现在以下几个方面:

- 1) 在规则对等组和子群已知的情况下, 为减小迭代计算的运算量, 并使最终的节点布局更加清晰、美观, 可以采用分级思想进行布局计算, 布局对象依次为子群、规则对等组和行动者;
- 2) 因为子群间的相似度可用于衡量子群间存在联系的可能性. 所以调整子群间距离, 使得相似度高的子群间距离较接近, 相似度低的子群间距离较疏远, 这样在可视化结果中用子群间距离就可以表现子群间存在联系的可能性. 同时, 这种调整还将便于子群的辨识;
- 3) 关键属性相同的行动者, 更倾向于存在联系. 因此在分布行动者时, 使得具有相同关键属性的行动者相对聚集, 具有不同关键属性的行动者相对疏离, 行动者间的距离就表现了行动者间的密切程度.

根据力导引布局算法的思想, 当行动者之间的距离偏离它们之间的适宜距离时, 它们之间的弹性力将产生势能, 系统将在总体能量最小时达到平衡. 因此, 适宜距离的确定是力导引布局算法的关键. 在传统的力导引布局算法中, 适宜距离由行动者间的路径长度决定, 定义 d_{ij} 为行动者 i 与行动者 j 在图论中的最短路径长度(值为整数), 常数 L 表示单位距离在可视化时的显示长度. 那么, 行动者 i 与 j 之间的适宜距离为

$$l_{ij} = L \times d_{ij} \quad (4)$$

为实现改进 2)、改进 3), 需根据子群分析结果对行动者间的适宜距离进行调整, 使得行动者间的适宜距离不仅由最短路径长度决定, 还与关键属性的异同和所属于子群间的相似度有关. 这样调整之后, 行动者间的适宜距离将由公式(5)进行计算, 其中, 常数 L 定义如前所述, 变量 b_{ij} 表示的距离值由最短路径长度和子群间距离来决定, 变量 s_{ij} 表示关键属性距离.

$$l_{ij}=L \times (b_{ij}+s_{ij}) \quad (5)$$

一般情况下,根据行动者 i, j 所属子群间相似度的不同,子群距离介于 0 到某个距离标称值之间,子群间相似度越高,距离越短.这时, b_{ij} 值由公式(6)中第 1 个式子计算.其中: SF_{ij} 为行动者 i, j 所属子群间的相似度; u 为 i, j 两个行动者所属子群的最大行动者数目, u 的平方根就表示该子群的大致直径.特殊情况下,即行动者 i 与 j 之间不存在路径,此时,根据行动者 i 与 j 所属子群间相似度的不同,取全图大致半径的一定比率作为行动者 i 与 j 之间的适宜距离.同时,为避免出现距离为 0 的情况,添加一个初始距离.这时, b_{ij} 值由公式(6)中第 2 个式子计算得到.其中: v 为全图中行动者的总数目, v 的平方根就表示全图的大致直径;其余变量定义如前所述.

$$b_{ij} = \begin{cases} d_{ij} + \sqrt{u} \times (1 - SF_{ij}), & d_{ij} \neq \infty \\ \sqrt{u} + \sqrt{v} \times (1 - SF_{ij}), & d_{ij} = \infty \end{cases} \quad (6)$$

关键属性距离的计算由各个属性的适宜距离叠加得到,设共有 g 个属性参与计算,依次考虑每个属性.当 i, j 两个行动者第 m 个属性相同时,也即第 m 个同属性矩阵的元素 a_{ij} 为 1 时,不增加关键属性距离;否则,增加一个关键属性距离 S_m ,计算方法见公式(7):

$$s_{ij} = \sum_{m=1}^g a_{ij} S_m \quad (7)$$

SAL 算法对适宜距离计算进行上述调整后,分级地进行布局计算.

3.2 SAL算法的三维布局

Ahmed 提出的社会网络三维可视化方法^[10]唯一的布局依据是行动者的连接度,也即社会网络分析学中的绝对点度中心度.用绝对中心度衡量一个行动者的重要性与影响力并不是一个最佳选择,这种测度仅仅关注局部信息,即与行动者存在直接连接的行动者数目.本文选择相对中心度来代替绝对中心度.此外,我们还定义了子群内的相对中心度,即行动者在子群内部的绝对中心度与子群中行动者的最大可能度数之比,通过这一数值可以对行动者在子群中的重要性 and 影响进行评估.行动者 i 在子群内部的相对中心度为

$$C'_i = (C_i - 1) / (m - 1) \quad (8)$$

其中, C_i 为行动者 i 在子群内的连接度, m 为 i 所在子群的行动者总数.子群内部的相对中心度既然是与子群相关的相对数值,那么,以此数值为依据来确定行动者的高度坐标时,直接计算行动者节点的绝对高度是不恰当的,还应考虑所处子群在平面的投影区域大小.设子群的平面范围最大直径为 R ,并设定子群的最大可能坡度为 D (小于等于 90° 的常数),则点 i 的高度 h_i 采用公式(9)计算:

$$h_i = \text{tg}(D \times C'_i) \times R / 2 \quad (9)$$

经过这样的三维扩展之后,在最终的三维显示图像中,每个子群都可以看作是一座山脉.此时,山脉的可视化属性对应于某些社会学意义:1) 山脉的节点密集程度表明子群内部行动者之间的联系密集程度.此外,节点密集程度还表示子群的关键属性相似度;2) 山脉的主峰数量代表子群中关键行动者的数目,因为主峰节点代表的行动者与子群内大多数行动者存在联系;3) 山脉的坡度表示子群内部行动者的密切程度.如果山脉整体坡度较低,表明子群内部行动者之间的联系较少.因此, SAL 算法的三维信息可视化结果能够对子群内部的结构特性进行直观的分析.

4 SAL 算法与力导引布局算法的实例对比

信息可视化实例采用恐怖组织信息.该信息由 NCTC(国家反恐中心)放置在网络上的恐怖活动知识库获得,选取了 3 类共 87 个最主要的恐怖组织信息.把恐怖组织看作社会网络中的行动者,它们之间存在的同盟、共享等关系构成同盟共享关系无向网络.

4.1 恐怖组织信息 SAL 算法子群分析过程

子群分析过程,首先基于角色分析得到 22 个规则对等组,并进一步划分出 9 个子群.然后,根据关键属性分析进行相似度计算,恐怖组织信息主要属性有:活动区域、领导者、创立时间、成员数量、组织宗旨、经济来

源等.构造同子群矩阵和同属性矩阵组,经过 QAP 方法的相关性计算,剔除部分与子群划分状况相关性较小的属性,剩余的关键属性是组织宗旨和活动区域,针对这两个属性对子群进行相似度计算.

4.2 恐怖组织信息平面可视化结果及对比

SAL 算法首先根据子群分析结果进行子群层次的布局计算,得到如图 3 所示的图像,图中子群间的距离由子群间的连接性、相似度以及关键属性的异同计算得到.另外,孤立节点也根据关键属性的不同进行分组并加入计算.经过分级计算后,最后的恐怖组织信息平面可视化结果如图 4 所示.图中央位置是基地组织(al-Qaeda)为首的恐怖组织子群,其右方是哈马斯(Hamas)组织为首的恐怖组织子群,一些小的恐怖组织如 CPN-M,IRA 和 LVF 等则分布在四周.图中颜色属性代表恐怖组织的组织宗旨属性.

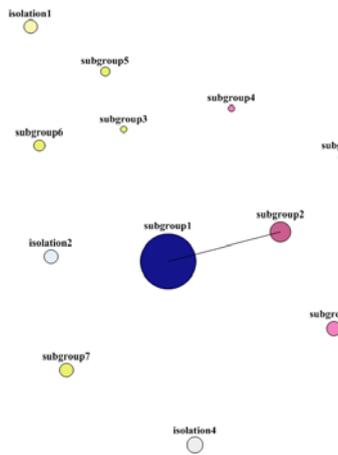


Fig.3 Subgroup layout graph of terrorist organization data produced by SAL algorithm

图 3 恐怖组织信息 SAL 算法子群布局结果

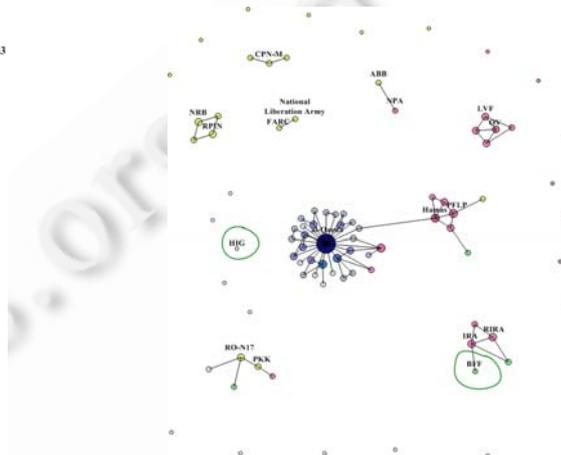


Fig.4 Layout graph of terrorist organization data produced by SAL algorithm

图 4 恐怖组织信息 SAL 算法可视化结果

OntoVis 系统^[8]使用恐怖组织信息,通过力引导布局算法得到的图像如图 5 所示.国防科学技术大学图像小组的研究基于本体论的建模与可视化方法^[13],也使用恐怖组织信息作为例子,该方法对子群、角色进行了一定的分析,通过改进的力引导布局算法得到的可视化图像如图 6 所示.SAL 算法结果与这两者的比较见表 1.

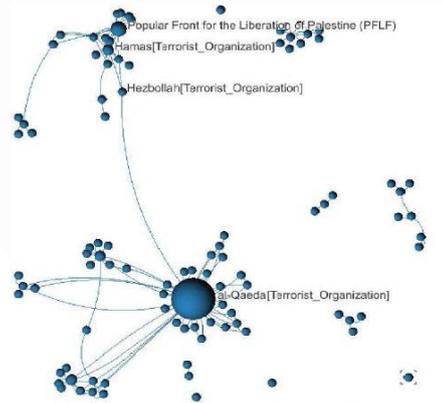


Fig.5 Layout graph of terrorist organization data in OntoVis

图 5 OntoVis 中恐怖组织信息可视化结果

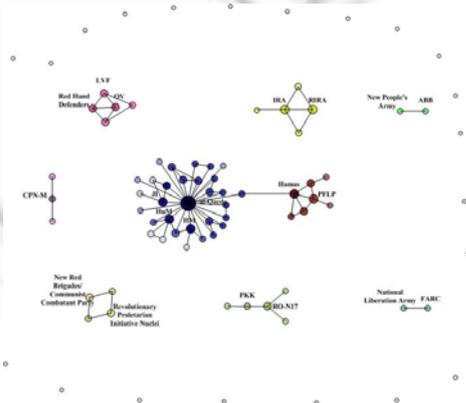


Fig.6 Layout graph of terrorist organization data produced by layout algorithm based on ontology

图 6 恐怖组织信息基于本体论的布局算法可视化结果

Table 1 Comparison of layout result for terrorist organization data

表 1 恐怖组织数据可视化结果比较

Algorithm	Layered calculation	Subgroup layout	Similarity of subgroups	Similarity of key attributes	Role of distribution
OntoVis	No	The main subgroups of al-Qaeda and Hamas can be recognized easily in figure 5.	Can't analysis and display it. The isolated nodes are placed at the brim in figure 5.	Can't analysis and display it.	In figure 5, only the node which represent the al-Qaeda is outstanding and other nodes have no difference with each other in shape and color.
Visualization algorithm based on ontology	No	The distribution of subgroups in figure 6 is more regularly and legibly than it in figure 5 for the optimization of initialization and rendering of subgroups.	Can't analysis and display it. The isolated nodes are placed at the brim in figure 6.	Can't analysis and display it.	Figure 6 describe the layout of actors well for the network roles and degree of nodes are represented by the color depths and diameters of actors.
SAL algorithm	Layered calculation is used for reducing complexity and accelerating the convergence of layout process. Moreover, it's also useful for mastering the global layout of social network and applications of visualization compress	The distribution of subgroups in figure 4 is more regularly and legibly than it in figure 5 for the computation of subgroup distance.	In figure 4, the distance between subgroups is related to the similarity of subgroups. As a result, through the similarity computation based on key attributes, the similar subgroups are near while the subgroups not similar are far off in the graph. Specially in SAL algorithm, isolated actors are placed close to subgroups which have the similar key attributes values with them. For example, the HIG terrorist organization is a religious organization at Afghanistan and Pakistan, thus it's near the subgroup of al-Qaeda because similarity between them is high. The distance between subgroups or actors could be used for estimating the possibility of relations between them. For example, the HIG terrorist organization might have unknown relations with the subgroup of al-Qaeda for they are close in the visualization result.	In figure 4, the distances between actors reflect their similarities of key attributes. Therefore, the internal cohesion of subgroups can be concluded easily from the distribution of key attributes and the distances between actors. For example, most terrorist organizations in subgroup of al-Qaeda are religious organizations and they are near in the figure 4, therefore we can get the conclusion that it is a steady team which internal cohesion is strong. There are nationalism organizations and left organizations inside the subgroup of IRA and they are a little bit far from each other in figure 4, thus it should be a loose team. The BFF is different in area and tenet with all other organizations inside its subgroup and just have link with the IRA. It might be a departing organization of the subgroup.	In figure 4, the color depths and diameters of actors represent the network roles and degree of nodes. That's means visualization results of SAL algorithm is fit for displaying the distribution of social roles. For example, the subgroup of al-Qaeda has four network roles and the deeper color indicates more important role in the subgroup.

通过上述 SAL 算法与国内外平面二维力导引布局算法的比较可以看出,SAL 算法在平面可视化布局时,能够清晰地显示子群分布,反映行动者之间的密切程度,展现行动者的关键属性分布,并直观显示行动者在子群内的角色分布与影响力,非常适合于社会网络结构化的分析与显示。

4.3 恐怖组织信息三维可视化结果及对比

Ahmed 等人提出的力导引布局三维可视化方法^[10]采用连接度作为节点重要性的划分标准,这种显示方式

适合于表现无尺度网络连接度呈幂指数分布的特性,并从总体宏观的角度上把握关键节点以及连接度层次分布.采用此方法对恐怖组织信息进行可视化,得到的结果如图 7 所示.

SAL 算法得到的恐怖组织信息三维可视化结果如图 8 所示.与图 7 相比:1) 图 8 能够清晰地显示恐怖组织子群分布和恐怖组织子群间的密切程度,因为 SAL 算法基于子群分析,并根据子群间的相似度对子群间的距离进行了相关计算处理;2) 图 8 能够对属性分布以及子群内部凝聚力进行直观的显示,因为恐怖组织间的距离与关键属性的异同相关,并且在可视化过程中采用不同颜色表示行动者的组织宗旨;3) 图 8 能够对行动者的网络角色分布进行清晰的显示,因为 SAL 算法对角色进行了分析,并采用不同颜色深度、直径大小表示角色分析结果;4) 图 7 对行动者进行平面划分的唯一依据是连接度,而 SAL 算法依据子群内的相对中心度进行三维布局.因此,图 8 更适合于显示行动者在子群内部的影响力和重要性.例如,Hum 和 Hamas 连接度都为 5,但是 Hum 所处的子群包含 31 个行动者,而 Hamas 所处的子群只有 7 个行动者,它们在所属于子群内的重要性区别很大,图 8 能够很好地表现这一区别.

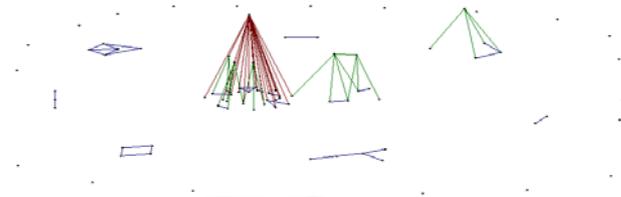


Fig.7 3D graph of terrorist organization data produced by Ahmed's algorithm

图 7 恐怖组织信息 Ahmed 布局算法三维可视化结果

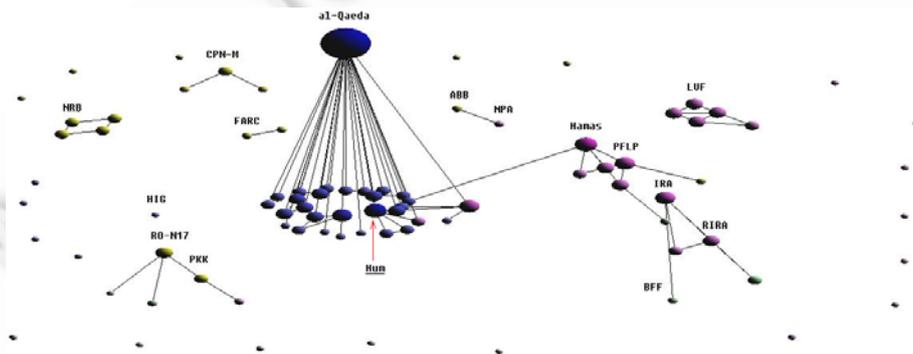


Fig.8 3D graph of terrorist organization data produced by SAL algorithm

图 8 恐怖组织信息 SAL 算法三维可视化结果

通过上述比较可以看出,在三维可视化方面,SAL 算法同样比传统力导引布局算法更适合于进行社会网络的结构化分析与显示.

5 结束语

目前,社会网络信息可视化通常采用力导引布局算法进行,该类算法易于理解、实现简单,但难以满足社会网络结构分析与可视化需求.为此,本文提出基于子群分析的 SAL 算法来满足这一需求.该算法通过角色分析与关键属性分析,划分社会网络中的子群并计算子群间的相似性.然后,根据子群分析结果对力导引布局算法加以改进,包括调整行动者之间的适宜距离和分级进行布局计算;在显示三维分布时,则以相对中心度为依据.SAL 算法克服了传统力导引布局算法在社会网络结构分析与可视化上存在的不足,其可视化结果能够清晰地显示子群分布,反映行动者间的密切程度,展现行动者关键属性分布,并直观显示行动者在子群内的角色分布与影响

力.最后,以对恐怖组织信息进行分析为实例,验证了算法在社会网络结构分析与显示方面所表现出的优越性.

下一步我们将研究如何基于子群分析对复杂社会网络进行压缩与提取,以及研究如何在社会网络演化条件下,进行节点布局变化的快速计算和可视化表现等内容.

致谢 我们向对本文工作给予支持和建议的国防科学技术大学曾亮教授以及图形图像团队的成员表示感谢.

References:

- [1] Herman I, Melancon G, Marshall MS. Graph visualization and navigation in information visualization: A survey. *IEEE Trans. on Visualization and Computer Graphics*, 2000,6(1):24–43. [doi: 10.1109/2945.841119]
- [2] Henry N, Fekete JD. MatrixExplorer: A dual-representation system to explore social networks. *IEEE Trans. on Visualization and Computer Graphics*, 2006,12(5):677–684. [doi: 10.1109/TVCG.2006.160]
- [3] Henry N, Fekete JD, McGuffin MJ. NodeTrix: A hybrid visualization of social networks. *IEEE Trans. on Visualization and Computer Graphics*, 2007,13(6):1302–1309. [doi: 10.1109/TVCG.2007.70582]
- [4] Wang B, Wu W, Xu CQ, Wu B. A survey on visualization of complex network. *Computer Science*, 2007,34(4):17–23 (in Chinese with English abstract).
- [5] Liu J. *An Introduction to Social Network Analysis*. Beijing: Social Sciences Academic Press, 2004. 4–189 (in Chinese).
- [6] Scott J, Wrote; Liu J, Trans. *Social Network Analysis: A Handbook*. 2nd ed., Chongqing: Chongqing University Press, 2007. 84–85 (in Chinese).
- [7] Chan DSM, Chua KS, Leckie C, Parhar A. Visualisation of power-law network topologies. In: Moreton N, Landfeldt B, Moors T, eds. *Proc. of the 11th IEEE Int’l Conf. on Networks*. Sydney: Bioxham & Chambers, 2003. 69–74. [doi: 10.1109/ICON.2003.1266169]
- [8] Shen ZQ, Ma KL, Eliassi-Rad T. Visual analysis of large heterogeneous social networks by semantic and structural abstraction. *IEEE Trans. on Visualization and Computer Graphics*, 2006,12(6):1427–1439. [doi: 10.1109/TVCG.2006.107]
- [9] Ware C. Designing with a 2 1/2d attitude. *Information Design Journal*, 2001,10(3):171–182.
- [10] Ahmed A, Dywer T, Hong SH, Murray C, Song L, Wu YX. Visualisation and analysis of large and complex scale-free networks. In: Brodlie KW, Duke DJ, Joy KI, eds. *Proc. of the Eurographics/IEEE VGTC Symp. on Visualization*. Leeds: Eurographics Association, 2005. 239–246.
- [11] Huang JW, Kang LS, Chen YP. A new graph drawing algorithm for undirected graphs. *Journal of Software*, 2000,11(1):138–142 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/11/138.htm>
- [12] Ren L, Wang WX, Zhou MJ, Teng DX, Ma CX, Dai GZ, Wang HA. A model driven development method for interactive information visualization. *Journal of Software*, 2008,19(8):1947–1964 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/19/1947.htm> [doi: 10.3724/SP.J.1001.2008.01947]
- [13] Wu P, Li SK. Social network information modeling and visualization based on ontology. *Journal of Computer-Aided Design & Computer Graphics*, 2009,21(4):518–525 (in Chinese with English abstract).

附中文参考文献:

- [4] 王柏,吴巍,徐超群,吴斌.复杂网络可视化研究综述. *计算机科学*,2007,34(4):17–23.
- [5] 刘军. *社会网络分析导论*.北京:社会科学文献出版社,2004.4–189.
- [6] Scott J,著;刘军,译. *社会网络分析法*.第2版,重庆:重庆大学出版社,2007.84–85.
- [11] 黄竞伟,康立山,陈毓屏.一个新的无向图画图算法. *软件学报*,2000,11(1):138–142. <http://www.jos.org.cn/1000-9825/11/138.htm>
- [12] 任磊,王威信,周明骏,腾东兴,马翠霞,戴国忠,王宏安.一种模型驱动的交互式信息可视化开发方法. *软件学报*,2008,19(8):1947–1964. <http://www.jos.org.cn/1000-9825/19/1947.htm> [doi: 10.3724/SP.J.1001.2008.01947]
- [13] 吴鹏,李思昆.社会网络信息的本体论建模与可视化. *计算机辅助设计与图形学学报*,2009,21(4):518–525.



吴鹏(1976—),男,江西南昌人,博士,助理研究员,主要研究领域为虚拟现实与可视化,分布式仿真.



李思昆(1941—),男,博士,教授,博士生导师,CCF高级会员,主要研究领域为虚拟现实与可视化,CAD,VLSI设计方法学.