http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2014.0501

网络图中边集束优化问题

姚中华[1]，吴玲达*[1,2]，宋汉辰[2]

(1. 装备学院 复杂电子系统仿真实验室，北京 101416； 2. 国防科学技术大学 信息系统工程重点实验室，长沙 410073)

摘 要：网络规模增大和复杂度提高造成的节点遮挡覆盖和边交叉阻塞等问题成为网络可视化研究的热点.针对网络中出现的视觉凌乱问题,以空间位置和群组关系为出发点,从网络中独立的边和群组两个层次,以边汇合的角度研究边集束技术,将网络中临近的边集聚成束以降低视觉复杂度,提出和改进了分段力导引算法(FDA)简化模型和群组边相容的网络图边集束模型.其中分段集束模型,提出以二次样条曲线表示网络边,通过样条控制点进行迭代汇聚的方法,实现了网络中边的集束;针对分段集束模型中部分连线过度弯曲问题,提出通过 CNM 聚类算法将网络进行群组划分,在群组结构的基础上对组内连线应用边相容原则,根据连线的匹配系数计算其集束程度的方法,网络图集束后曲线扭曲变形减少,曲线过渡更加平滑.选取国内航空网络作为案例,通过两种边集束模型进行网络图简化,分析结果表明,国内机场的群组结构具有地理属性的相近性,航空网络在整体上呈现出明显的十字脉络,东西走向和南北走向的航线分别汇聚集结成束,表现了航空网络建设在南北和东西方向的总体趋势.本集束简化算法适用性广,绘制的网络图具有良好的视觉效果和可读性.

关 键 词：网络可视化；边集束；分段力导引算法；聚类算法；集束简化算法

中图分类号：TP391

文献标识码：A **文章编号**：1001-5965(2015)05-0871-08

网络图绘制主要展现网络中节点和节点之间连线关系这两个基本元素,根据不同应用背景,产生了很多基于点布局、边布局等网络图绘制方法.力导引(Force-Directed Algorithm,FDA)布局算法由于其简洁高效易于理解,目前被广泛应用于信息可视化,由 Eades 提出[1]以及据此发展出的各种改进算法构成,包括 KK 算法[2]、DH 算法[3]、FR 算法[4]等.主要思想将图看成顶点为小球、边为弹簧组成的物理系统,系统被赋予初始状态以后,节点在弹力和节点间受力的作用下移动,直到系统总能量减少到最小值时运动停止,认为系统达到稳定状态.

随着网络规模增大带来节点和边数量的剧增,严重的边交叉和遮挡造成的视觉凌乱问题,是影响网络图可读性和信息可达性及有效性的重要原因.Ellis 等[5]从空间变形、外观、时序 3 个角度总结现有的简化策略,本文重点利用边集束从空间变形研究网络图简化技术.

边集束改变连接节点的边为平滑曲线,其核心思想是将相邻接的边收拢集束,然后在端点分散开[6],能够集束局部阻塞的边,提高网络可读性和信息可达性,使网络图更加简洁高效.Chung 等将集束应用于分子研究领域,发掘微管相关蛋白与神经系统之间的关联[7].Gansner 和 Koren 给

收稿日期：2014-04-08；**录用日期**：2014-08-11；**网络出版时间**：2014-12-01 16:01

网络出版地址：www.cnki.net/kcms/doi/10.13700/j.bh.1001-5965.2014.0501.html

基金项目：国家自然科学基金(61103081)

作者简介：姚中华(1989—)，男，安徽阜阳人，博士研究生，visworker07@163.com

*通讯作者：吴玲达(1962—)，女，上海人，教授，wld@nudt.edu.cn，研究方向为多媒体信息系统与虚拟现实技术、系统集成、网络可视化、信息可视化.

引用格式：*姚中华，吴玲达，宋汉辰. 网络图中边集束优化问题[J]. 北京航空航天大学学报，2015，41(5)：871-878.*

Yao Z H，Wu L D，Song H C. Problems of network simplification by edge bundling[J]. Journal of Beijing University of Aeronautics and Astronautics，2015，41(5)：871-878 (in Chinese).

http://bhxb.buaa.edu.cn

出改进的圆形布局算法[8],边被布局在外圆或内圆表面,通过层次化边聚类使内圆的边集束以实现空间利用率的优化. 对层次布局进行改进,Gansner等又提出适合于大型网络的多层次边集束方法[9],使用最少的像素表示网络中的边,同时对曲线的曲度进行限制. Phan 等提出基于层次聚类的流图布局技术[10],边通过层次树结构进行集束,合并通往相同根节点的分支以减少视觉凌乱. Holten 吸收这种思想,提出层次边集束法[6],集束相邻边,将层次以树形结构体现,适用于表现具有层次关系的结构数据. Hotlen 等又提出基于力导引的边集束[11],通过启发式算法插入中间点的方式将边分割为独立线段,度量相互作用的边进行集束. 该方法能够在保持网络图节点位置不变的前提下显著减少视觉杂乱,大大降低图形的视觉复杂度,适用于一般性网络图. Cui 等则通过建立控制网格[12],计算网格区域内适合聚集合并的边,根据边的方向完成合并. 这种方法依赖控制网格的优劣,适用于主体脉络清晰的网络. Lambert 等[13]根据节点位置生成网格,利用网格分割平面,计算边的最短路由,根据路径使用频率进行集束. Ozan 等提出基于中间轴的边集束[14],将网络中的边聚类分组,计算各组轮廓形状,集束组内的边,使其向中间轴靠拢. 中间轴对获取和理解网络拓扑结构有启发性意义,但其生成边的变形不易控制,会造成部分边的过度扭曲. Wong 等提出将网络图中杂乱部分移出用户关注区域的 Edge-Plucking[15] 和 EdgeLens[16] 技术. EdgePlucking 通过交互的方式拖拽阻塞区域的边,从而实现减少局部杂乱. EdgeLens 将用户关注点区域阻塞的边通过鱼眼技术“推开”. 这两种技术都能够解决局部关注区域的交叉问题,但不能解决网络图整体的视觉杂乱问题,并且依赖用户的参与. Zhou 等描述了边集束的具体应用并对其有效性等进行了评估研究[17].

1 基于群组相容的分段 FDA 集束简化

1.1 分段 FDA 的集束模型

力导引算法(Force Directed Algorithm, FDA)针对网络可视化中的节点布局,主要思想将图看成节点为小球、边为弹簧组成的物理系统,系统被赋予初始状态以后,节点在弹力和节点间受力的作用下移动,控制迭代过程,直到系统总能量减少到最小值时运动停止,认为系统达到稳定状态. FDA 算法提供了网络布局的有效方法,但布局完成后的网络图采用直线连接的方式,仍然可能因为边的重叠造成二义性问题,为网络图的理解带来困难.

因此,本文对 FDA 加以改进,将网络中的边通过插值点进行分割,然后再对插值点迭代运算,利用插值点控制生成集束曲线. 这种方法能够实现在不改变节点初始位置的同时,减少二义性的产生.

1.1.1 基于分段 FDA 的物理模型

对基本 FDA 模型加以改进,并将其应用到节点具有地理属性的网络图中. 初始输入为一般网络图的节点连接图,即有关联的节点对之间采用直线连接,节点带有坐标属性,位置固定,具有不可改变性,为实现节点对之间连线的集束,通过为线段插入插值点的方式将连线分割为多个折线段. 如图 1 所示,相互影响的两条边 P 和 Q 各自插入 4 个插值点被均分为 5 个片段,其中每条边的起始点 P_0、P_4、Q_0 和 Q_4 位置固定. P 边上插值点 p_1 受到邻接插值点 p_0 和 p_2、弹力 $\boldsymbol{F}_s$ 以及边 Q 上对应插值点 q_1 的引力 $\boldsymbol{F}_e$ 作用,其中 $\boldsymbol{F}_s$ 和 $\boldsymbol{F}_e$ 均为矢量.

同一条边上,任意相邻两个插值点(包括起始点)之间存在相互吸引的线性弹力 $\boldsymbol{F}_s$,并且弹力大小和两点之间的距离成正比,当两点之间距离为 0 时,弹力也为 0,每条边的弹性系数 k_p 由弹性常量 K 和插值点所在边的长度决定,表达式为

$$k_p = \frac{K}{|P|} \tag{1}$$

式中:$|P|$为边 P 的初始长度,$||$ 表示的是进行距离计算;K 弹性常量.

弹性常量 K 通过决定集束的强度被用来控制网络的集束数量,由于网络中的边的初始长度一般都不同,因此将其划分为折线段之后,通过上述公式计算,则任意一条边都具有不同的局部弹性系数 k_p.

根据物理模型,分割片段上的插值点均为相互吸引的带电小球,引力 $\boldsymbol{F}_e$ 被施加在相互影响的两条边上对应的插值点对之间,如图 1 所示,引力存在于 p_0 和 q_0、p_1 和 q_1、p_2 和 q_2 以及 p_3 和 q_3 之间. 引力大小反映插值点之间影响程度,由于逆平方模型相对逆线性模型能够产生更强以及更加局部化的集束,采用逆平方模型计算两插值点之间引力,使得引力和两个插值点的距离成反相关,即两插值点之间越接近,距离越短,则引力越大,距离越远,则引力越小.

每一步迭代过程中,位于边 P 上的插值点 p_i,需要计算邻接插值点 p_{i-1} 和 p_{i+1} 施加的弹力 $\boldsymbol{F}_s$ 以

http://bhxb.buaa.edu.cn

及其他所有边上对应插值点施加的引力 $\boldsymbol{F}_e$,最终插值点 p_i 收到的合力 $\boldsymbol{F}_{pi}$ 为两者矢量相加之和,即 $\boldsymbol{F}_{pi}=\boldsymbol{F}_s+\boldsymbol{F}_e$,其中

$$\boldsymbol{F}_s=k_p\cdot(\|p_{i-1}-p_i\|+\|p_i-p_{i+1}\|) \tag{2}$$

$$\boldsymbol{F}_e=\sum_{Q\in E}\frac{1}{\|p_i-q_i\|^{1/2}} \tag{3}$$

式中:k_p 为插值点所在边 P 的局部弹性系数;E 为除了插值点所在边 P 的所有边集合;‖ ‖ 表示的是进行矢量计算,结果为矢量.

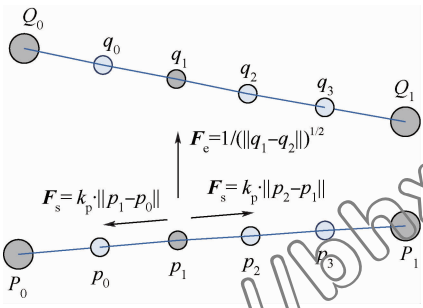


图 1　分段 FDA 物理模型

Fig. 1　Segmental FDA physical model

1.1.2　基于模拟退火算法的迭代策略

分段 FDA 模型包括物理模型以及迭代过程两部分,其中物理模型给出节点之间相互作用力模型,迭代过程则依据物理模型迭代运算节点之间的作用力,计算节点在力的作用下移动距离,更新节点位置,直到达到终止条件.

从退火算法中得到启发,进行 N 轮迭代,随着迭代次数的增加,逐次减少迭代步长,判断是否符合迭代终止条件,如果不符合则进行下一轮迭代,否则将终止迭代运算. 每一轮迭代过程中,边的插值点数量 P 逐次增加,迭代次数 I 以及迭代步长 λ 逐次减少,其中,第 i 轮迭代过程:

$$P_i=2^{i-1} \tag{4}$$

$$I_i=\frac{2\cdot I_{i-1}}{3} \tag{5}$$

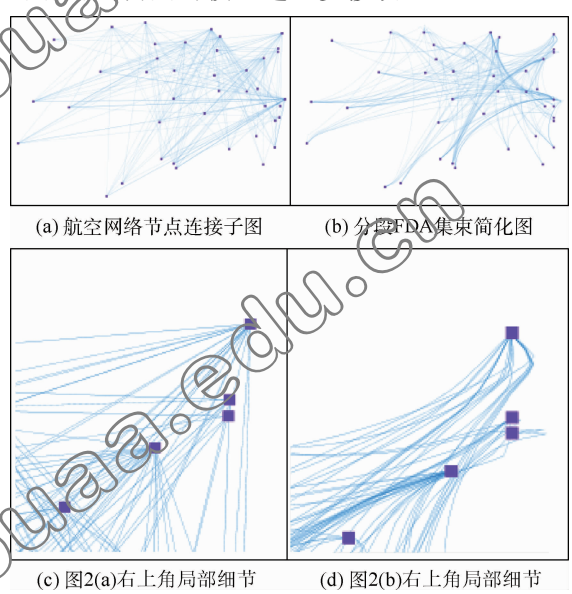
每步迭代过程,插值点的位置都将在所受合力的作用下朝合力的方向前进一定距离,步长为 λ,第 i 轮迭代的步长为 λ_i.

1.1.3　样条曲线集束模型

根据分段 FDA 模型将网络中的边分割划分,计算其相互之间作用力,经过一定次数迭代之后,系统达到稳定状态,终止迭代,即认为找到插值点的理想位置,然后将连线的起始点以及插值点作为 B 样条曲线生成的控制点,生成集束曲线. 随着曲线阶数的增大,高阶样条曲线受控制点影响过度,曲线和 B 样条特征多边形体现出高相似度,弯曲程度变得“僵硬”,难以看到曲线的明显变化,并且由于阶次的影响,高阶样条曲线可能会产生多余的摆,因此本文生成二次 B 样条曲线.

图 2(a)为国内航空网络节点连接子图,对单个节点的细节展示更加清晰,图 2(b)是应用分段 FDA 集束简化后,网络图的绘制能够实现边的汇聚,更有利于展现网络整体连接模式,生成的曲线降低了追踪连线的难度,从而减少视觉杂乱. 因此在网络相对稀疏的情况下,节点连接图能够发挥直线指向性强的特点,易于表现连线的方向以及两点之间的连接关系;而集束后的曲线方法则易于展现聚类的群组以及揭示网络的整体结构特点.

集束后的网络图仍然存在部分连线过度弯曲变形的问题,如图 2(c)、图 2(d)所示,分别为图 2(a)、图 2(b)右上角的局部细节展示. 部分节点周围曲线出现了过度弯曲,导致变形现象. 汇聚的曲线都由控制点控制生成,而控制点则由分段 FDA 算法迭代运算中的插值点生成,因此连线对应插值点之间的引力是造成集束曲线弯曲变形的主要原因,需要对模型进一步修改.



(a) 航空网络节点连接子图　(b) 分段FDA集束简化图

(c) 图2(a)右上角局部细节　(d) 图2(b)右上角局部细节

图 2　分段 FDA 集束简化图

Fig. 2　Segmental FDA bundling simplification

1.2　群组边相容集束模型

近年来,研究人员通过对城市道路交通网络拓扑结构[18]、机场航线网络[19-20]、人口迁移网络流[21]等复杂网络进行研究分析,发现了网络一般都具有一个共同性质——群组结构,即网络是由若干个“群组”或“类别”组成,群组内部节点之间的相互连接紧密,而群组和群组之间的连接则相对稀疏[22].

因此,考虑改进分段 FDA 集束模型,首先将网络中的节点通过群组发现算法进行聚类,以群组内节点关系强连接和群组间节点关系弱连接为

http://bhxb.buaa.edu.cn

约束条件,挖掘网络中的群组结构.针对群组内节点连接密度高的特点,对组内连线引入边相容原则,计算连线之间的匹配系数,利用匹配系数衡量边与边之间相似度,通过相似度度量边之间集束程度,进而调整网络中边与边之间的相互作用,最后修改 FDA 物理模型.

1.2.1 CNM 群组划分

目前使用最多的群组发现算法是基于模块度(Q)的算法,基本思路是把群组发现问题转化为网络优化问题,将搜索网络中最优群组结构作为目标约束条件. CNM 算法[23] 由 Clauset、Newman 和 Moore 等于 2004 年提出,它以 Newman 快速算法[24] 为基础,由于采用基于堆的数据结构计算和更新网络模块度的方法,算法复杂度从 $O(n^3)$ 降低到 $O((m+n)\cdot n)$,实现算法效率的大幅提高;另一方面,CNM 算法在迭代过程中合并多对群组以防止网络过早地收缩到少数较大的群组,能使网络群组均匀分布,因此采用 CNM 算法对网络中的节点进行聚类.

CNM 算法基于模块度增量矩阵,基本思路是将网络看成点集,每个节点各自单独构成群组,再合并相互间具有强连接关系的节点对,使其成为一个较大的群组,然后重复搜索并合并具有强连接性的群组.整个流程形成了一个自下而上的树状图,如图 3 所示,不同合并层次就对应着一种群组划分结构,不同颜色代表不同的群组划分.

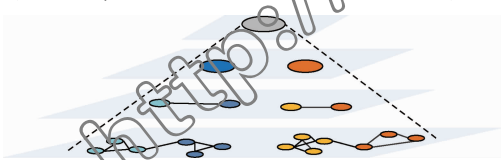


图 3 CNM 算法流程

Fig. 3 CNM algorithm flow

1.2.2 群组边相容原则

群组内部节点密度高,相互连线密集,位于同一个组内的节点之间构成强连接关系,在分段 FDA 集束模型中,不考虑任意两条边之间关系直接计算其相互作用,造成插值点偏移抖动,影响以插值点作为控制点生成的样条曲线,导致部分连线过度弯曲.因此,考虑引入一定的匹配参数,度量组内边的相似程度,从而决定其在何种程度进行集束.

针对群组内节点连接密度高的特点,对组内连线应用边相容原则,通过相容性滤除相关性较低的边.计算连线之间的匹配系数,利用匹配系数衡量边与边的相似度,通过相似度度量边之间的集束程度,进而调整网络中边与边之间的相互作用,最后对 FDA 模型进行修改.为度量任意两条边的相容性,根据 Holten 提出的匹配原则从角度、尺度、位置、视觉这 4 个方面考虑[11](见图 4).

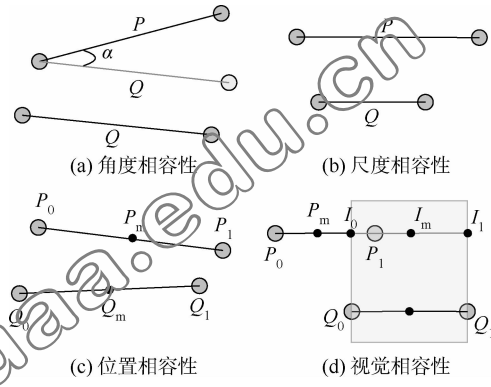


图 4 边相容原则

Fig. 4 Edge consistent principle

1) 角度相容性原则.

从边的延伸方向考虑度量两条边之间的相似性,计算他们之间的夹角,认为夹角越小,越能够体现出他们在网络中朝着相同方向延伸发展,有相似的趋势,汇聚在一起的程度越高.

几乎垂直的两条边在网络中具有截然不同的延伸方向,不应该通过集束汇聚在一起,因此,引入角度相容性系数 $C_a(P,Q)\in[0,1]$ 来度量,其中

$$C_a(P,Q)=|\cos\alpha| \tag{6}$$

$$\alpha=\arccos\left(\frac{P\cdot Q}{|P||Q|}\right) \tag{7}$$

如图 4(a)角度相容性示意图,边 P 与 Q 之间的角度相差越大,则其相容系数 $C_a(P,Q)$ 越小,当 $C_a(P,Q)$ 为 0 时,P 与 Q 垂直,当 $C_a(P,Q)$ 为 1 时,P 与 Q 相互平行,此时角度相容性系数最大,从延伸方向的角度考虑,边 P 与 Q 应具有较高的集束程度.

2) 尺度相容性原则.

从尺寸考虑边的相似性,计算两者之间的尺寸差异,这一原则主要为避免长度较小的边为适应和长度较大的边进行集束而发生过度弯曲,因此尺寸大小越相似,其汇聚程度越高.

图 5 为边扭曲变形示意图,主要由尺寸差异造成,其中两条边的原始线段以及受力弯曲后的曲线段分别用黑色和蓝色表示,两者之间存在的静电引力用橙色表示.

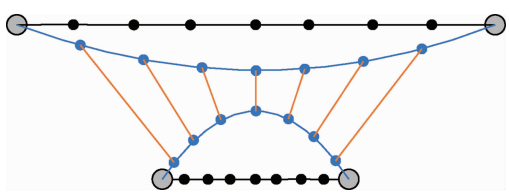


图 5 集束中边扭曲变形

Fig. 5 Curve distortion by bundling

http://bhxb.buaa.edu.cn

引入尺度相容性系数 $C_s(P,Q)\in[0,1]$,其中

$$C_s(P,Q)=\frac{2}{l_{avg}\cdot\min(|P|,|Q|)+\frac{\max(|P|,|Q|)}{l_{avg}}} \tag{8}$$

式中:l_{avg}为边 P 与 Q 长度之和的一半,即

$$l_{avg}=\frac{|P|+|Q|}{2} \tag{9}$$

如图 4(b)所示,当 P 与 Q 长度相等时,尺度相容系数 $C_s(P,Q)$为 1,两者汇聚程度最高;当较长的边和较短的边长度之比趋向无穷大时,尺度相容系数 $C_s(P,Q)$趋向于 0,两者具有较低的汇聚程度.

3) 位置相容性原则.

位置相容性从边所在位置考虑,尽量避免网络中相聚过远的边集束,减小其影响程度.以两条边中心的距离衡量两条边的距离,进而度量两者之间的位置相容性,相容性系数 $C_p(P,Q)\in[0,1]$计算公式如下:

$$C_p(P,Q)=\frac{l_{avg}}{l_{avg}+|P_m-Q_m|} \tag{10}$$

式中:P_m 和 Q_m 分别为边 P 与 Q 的中心点.

如图 4(c)所示,如果 P_m 和 Q_m 重合,则 $C_p(P,Q)$等于 1,如果 P_m 和 Q_m 趋向于无穷大,则 $C_p(P,Q)$趋近于 0.

4) 视觉相容性原则.

从视觉的角度度量两条边之间的匹配性,即使两条边同时满足以上 3 个原则,它们也可能具有较低的集束相容性.图 4(d)所示中将一条边竖直平移,则构成的相互平行的两条边,其节点起始位置差异较大,故应具有较小的相容性,视觉相容性 $C_v(P,Q)\in[0,1]$计算公式如下:

$$C_v(P,Q)=\min(V(P,Q),V(Q,P)) \tag{11}$$

$$V(P,Q)=\max\left(1-\frac{2|P_m-I_m|}{|I_0-I_1|},0\right) \tag{12}$$

式中:I_m 为 I_0 与 I_1 中点.图 4(d)为 P 至 Q 的视觉相容性系数示意图,$C(P,Q)$决定于由边 Q 延伸出的长度与边 P 的交叉部分,计算在边 P 上交叉点 I_0 与 I_1.当 P_m 与交叉部分中点 I_m 相重合(理想情况下),视觉相容性系数等于 1;当两条边的平行延伸没有重合部分时,视觉相容性系数等于 0.

根据边相容原则进行调整之后,对 FDA 模型进行适度修正,综合以上 4 种相容性系数,定义任意两条边 P 与 Q 的相容性系数 $C_e(P,Q)\in[0,1]$为

$$C_e(P,Q)=C_a(P,Q)\cdot C_s(P,Q)\cdot C_p(P,Q)\cdot C_v(P,Q) \tag{13}$$

边 P 与 Q 的整体相容性系数 $C_e(P,Q)$由 4 种相容性系数相乘求得,任意一个相容性系数较低都将对整体相容性系数造成较大影响.对应到网络图中即为,任意两条具有较高汇合程度的边必须对以上 4 条原则同时具有较高的切合度,结合上文推导的任意一点所受合力重新定义为

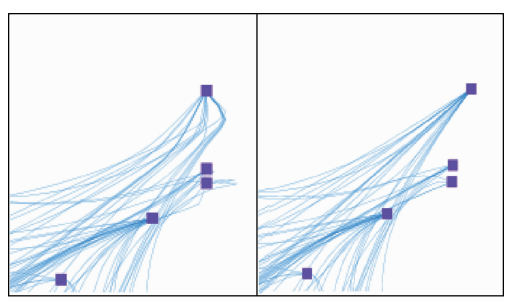
$$F_{pi}=kp\cdot(\|p_{i-1}-p_i\|+\|p_i-p_{i+1}\|)+\sum_{Q\in E}\frac{C_e(P,Q)}{\|p_i-q_i\|^{N2}} \tag{14}$$

即当 P 与 Q 两条边相容性系数为 0 时,它们之间作用力也为 0,即不考虑 P 与 Q 相互之间作用.

1.3　群组边相容集束简化

应用 CNM 聚类,建立网络的群组结构,对组内节点之间的连线,应用边相容原则,计算连线之间的相容性系数,利用相容性系数修改算法模型,迭代更新插值点作为样条曲线控制点,生成样条曲线,完成网络的绘制.

图 6 中从节点发出的连线两两之间角度差异细微,连线的延伸方向相似度高,认为它们应该具有较高的集束显著性.图 6(a)应用分段 FDA 集束模型,图 6(b)为应用群组边相容进行改进.对比两者,曲线汇聚的集束都能够表现网络中显著的连接脉络,而图 6(b)中的曲线过渡更加平滑,在节点周围以及边密集区域发生的抖动现象较少,应用群组边相容后能够改善集束产生的连线扭曲变形,达到优化网络图绘制的目的.



(a) 分段FDA集束　　(b) 群组边相容集束

图 6　群组边相容后集束对比

Fig. 6　Bundling comparison by group edge consistent

2　典型实例验证分析

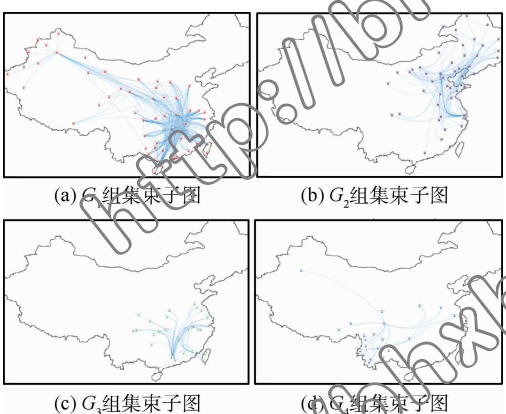
全球航班数据,采集世界各地机场间往来航班,由于航班具有往返特性,航线构成的网络图具

http://bhxb.buaa.edu.cn

有无向性. 为能清晰显示网络布局效果,便于对比分析算法效果,实验中过滤后保留国内航班数据对集束简化模型进行实验验证分析,节点数目为152,航班次数为1 167,网络直径为4,平均网络距离为2.15,网络密度为0.102.

2.1 群组结构分析

基于CNM算法将网络进行聚类,将航空网络划分为4个群组,群组内集束连接示例图7(a)~图7(d)分别对应于群组划分的 G_1 ~ G_4 组,由图可知各分组内节点能基本保持地理相近,G_1 和 G_2 组节点连接关系相对密集,G_3 和 G_4 组则相对薄弱. 群组内网络半径为3~4,即网络中任意两点最多需要经过3个中间点即可相互到达,划分的4个群组内平均网络距离都接近2,即对每一个节点平均需要经过1个中间点到达另外一个节点.



(a) G_1 组集束子图 (b) G_2 组集束子图

(c) G_3 组集束子图 (d) G_4 组集束子图

图7 群组内集束连接

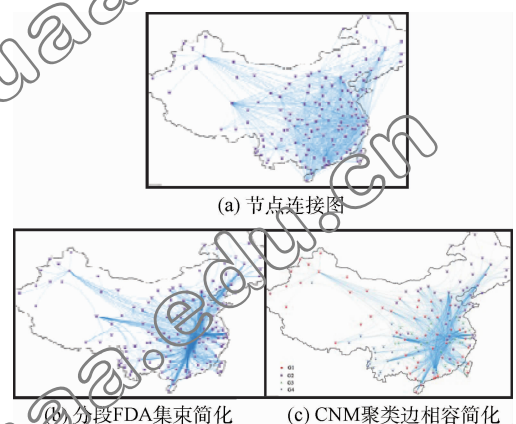
Fig. 7 Bundling within group

2.2 视觉凌乱度分析

基于边集束的网络图绘制共有的一个特征是使用曲线绘制网络中的边,Holten在层次边集束技术中使用了三次B样条曲线[6],通过调整曲线的控制点,样条曲线能够生成清晰明了、条理清楚的集束. Zhou在基于能量的网络图层次聚类中则使用了Bezier曲线和Catmull-Rom样条曲线[25]. Holten等在基于力导引的边集束[11]以及Cui等在基于几何的网络图边聚类中都使用平滑多边形的方式[12],根据多边形生成平滑曲线.

直线图于节点连线之间的指向性更加清晰准确,易于表现连线的方向以及两点之间的连接关系;而曲线则易于展现聚类的簇,以及揭示网络整体结构. 曲线的表现形式能够使其更加利于追踪,同时带来视觉上的良好体验,实验中采用二次样条曲线绘制,图8中对比节点连接图和集束简化图,可以看出集束后的网络图整体结构更加清晰,视觉凌乱现象得到减少.

比较图8(b)和图8(c),分段FDA集束简化图的集束汇聚程度过高,导致网络边集中在少数几个条带中,导致边在局部区域过度集中,为曲线追踪造成困难. 考虑分组和边相容原则后,应用CNM聚类边相容集束模型后,网络中集束相对宽松,汇聚的条带空间利用率更高. 同时整体呈现出十字型交叉脉络,边汇聚形成的东西方向集束构成十字型横边,南北方向集束构成十字型竖边,由集束后的网络图能够挖掘出网络脉络和连接模式.



(a) 节点连接图

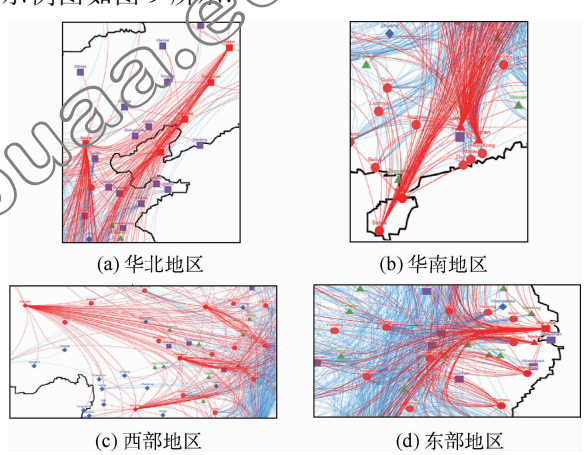
(b) 分段FDA集束简化 (c) CNM聚类边相容简化

图8 集束简化模型对比

Fig. 8 Bundling simplification model comparison

2.3 连接模式分析

对国内航空网络,按照节点地理位置所处群组,分别从群组中各自挑选出度数较高的5个节点,重点关注群组中关注点在网络中的连接情况,并将其高亮显示,获得了网络中节点的连接模式,示例图如图9所示.



(a) 华北地区 (b) 华南地区

(c) 西部地区 (d) 东部地区

图9 连接模式分析

Fig. 9 Connection mode analysis

图9为航空网络中高度数节点的分布和连接,具有两大特点:行政性和地理性. 行政性,节点大多属于省会城市,政治、经济资源集中;地理性,处于沿海或交通便利的内陆. 表明资源的集中和地理位置的便利使其与网络中的其他节点交流频

http://bhxb.buaa.edu.cn

繁,对网络有较强影响力.

分析节点连接,发现其呈现一定连接模式.图 9(a)、图 9(b)节点实际位置对应于华北地区和华南地区,对比两图可以观察到南北连线形成强烈的汇聚,集束结果表明南北跨地域间交流呈枢纽型发展.图 9(c)、图 9(d)对应于西部地区和东部地区,对比两图观察到东西部之间连线汇聚性表现不明显,表明东西跨地域间交流呈发散性发展.这种连接模式从总体上体现了国家航空网络建设的发展趋势,南北经济发达地区资源集中,交通便利,在网络中起到枢纽作用,相互交流密集,东西部地区经济相对薄弱、资源匮乏,与其他区域交流更加广泛.

3　结　论

本文针对网络中的视觉凌乱,以国内航线数据为实验对象绘制航空网络,提出了两种集束简化的改进方法,经实验验证表明:

1)算法具有一定准确性和有效性,分析结果表明本文的简化模型能够在一定程度上减少视觉凌乱,使网络的拓扑结构和连接特点更加清晰可见.

2)算法对网络进行信息挖掘,发现了航空网络的群组结构和整体上呈现十字脉络连接模式,东西和南北走向的航线分别汇聚集结成束,表明交流的密集性.

3)集束简化模型适用性广,绘制的网络图具有良好的视觉效果和可读性.

本文重点关注挖掘网络的整体连接模式,针对复杂网络分析仍需提供关于节点底层细节信息,有待进一步研究.

参考文献(References)

[1] Eades P. A heuristic for graph drawing[J]. Congressus Numerantium,1984,42(1):149-160.

[2] Kamada T,Kawai S. An algorithm for drawing general undirected graphs[J]. Information Processing Letters,1989,31(1):7-15.

[3] Davidson R,Harel D. Drawing graphs nicely using simulated annealing[J]. ACM Transactions on Graphics,1996,15(4):301-331.

[4] Fruchterman T M J,Reingold E M. Graph drawing by force-directed placement[J]. Software:Practice and Experience,1991,21(11):1129-1164.

[5] Ellis G,Dix A. A taxonomy of clutter reduction for information visualization[J]. IEEE Transactions on Visualization and Computer Graphics,2007,13(6):1216-1223.

[6] Holten D. Hierarchical edge bundles:visualization of adjacency relations in hierarchical data[J]. IEEE Transactions on Visualization and Computer Graphics,2006,12(5):805-812.

[7] Chung P J,Deek J,Feinstein H E,et al. A structure-function study of map Tau:analyzing distinct map Tau domains in mediating microtubule assembly and bundling using synchrotron SAXS[J]. Biophysical Journal,2012,102(3):700a.

[8] Gansner E R,Koren Y. Improved circular layouts[C]// Graph Drawing,14th International Symposium,GD 2006. Berlin:Springer-Verlag,2006:386-398.

[9] Gansner E R,Hu Y F,North S,et al. Multilevel agglomerative edge bundling for visualizing large graphs[C]// Proceedings of the 2011 IEEE Pacific Visualization Symposium. Piscataway,NJ:IEEE Press,2011:187-194.

[10] Phan D,Xiao L,Yeh R,et al. Flow map layout[C]// Proceedings of IEEE Information Visualization Symposium. Piscataway,NJ:IEEE Press,2005:219-224.

[11] Holten D,Van J J W. Force-directed edge bundling for graph visualization[J]. Computer Graphics Forum,2009,28(3):983-990.

[12] Cui W,Zhou H,Qu H,et al. Geometry based edge clustering for graph visualization[J]. IEEE Transactions on Visualization and Computer Graphics,2008,14(6):1277-1284.

[13] Lambert A,Bourqui R,Auber D. Winding roads:routing edges into bundles[J]. Computer Graphics Forum,2010,29(3):853-862.

[14] Ozan E,Christophe H,Fernando P,et al. Graph edge bundling by medial axes[J]. Visualization and Computer Graphics,2011,17(12):2364-2383.

[15] Wong N,Carpendale S. Using edge plucking for interactive graph exploration[C]// IEEE Information Visualization Symposium,Poster Compendium. Piscataway,NJ:IEEE Press,2005:51-52.

[16] Wong N,Carpendale S,Greenberg S. EdgeLens:an interactive method for managing edge congestion in graphs[C]// IEEE Information Visualization Symposiu. Piscataway,NJ:IEEE Press,2003:51-58.

[17] Zhou H,Xu P P,Yuan X R,et al. Edge bundling in information visualization[J]. Tsinghua Science and Technology,2013,18(2):145-156.

[18] 李树彬,吴建军,高自友,等. 基于复杂网络的交通拥堵与传播动力学分析[J]. 物理学报,2011,60(5):1-9.
Li S B,Wu J J,Gao Z Y,et al. The analysis of traffic congestion and dynamic propagation properties based on complex network[J]. Acta Physica Sinica,2011,60(5):1-9(in Chinese).

[19] Bagler G. Analysis of the airport network of India as a complex weighted network[J]. Elsevier,2008,387(12):2972-2980.

[20] Li W,Cai X. Statistical analysis of airport network of China[J]. Physical Review E,2004,69(2):1-6.

[21] Guo D. Flow mapping and multivariate visualization of large spatial interaction data[J]. IEEE,2009,15(6):1041-1048.

[22] 汪小帆,李翔,陈关荣. 复杂网络理论及其应用[M]. 北京:清华大学出版社,2006:162-163.
Wang X F,Li X,Chen G R. Complex network theory and application[M]. Beijing:Tsinghua University Press,2006:162-163(in Chinese).

[23] Clauset A,Newman M E J,Moore C. Finding community struc-

http://bhxb.buaa.edu.cn

ture in very large networks[J]. Physical Review E, 2004, 70(6):66111.

[24] Newman M E J. Fast algorithm for detecting community structure in networks[J]. Physical Review E-Statistical, Nonlinear, and Soft Matter Physics, 2004, 69(6):066133-1-5.

[25] Zhou H, Yuan X R, Cui W W, et al. Energy-based hierarchical edge clustering of graphs[C]// IEEE Pacific Visualisation Symposium 2008. Piscataway, NJ: IEEE Press, 2008:55-62.

Problems of network simplification by edge bundling

YAO Zhonghua[1], WU Lingda*[1,2], SONG Hanchen[2]

(1. Science and Technology on Complex Electronic System Simulation Laboratory, Equipment Academy, Beijing 101416, China;
2. Science and Technology on Information Systems Engineering Laboratory, National University of Defense Technology, Changsha 410073, China)

Abstract: Node occlusion and edge congestion problems, which are caused by the increment of network scale and complexity, had become a hot spot in network visualization research. To solve the visual clutter problem in network, edges close to each other in network were bundled by curving them. The bundling started from node position and group division, laying emphasis on edge bundling technique through edge convergence, edges bundled close to each other in network to reduce visual complexity. A segmental forced directed algorithm (FDA) simplification model and a group based consistent edge bundling network model were proposed and improved. In segmental FDA bundling model, quadratic spline was used to display line in network, control points of spline were produced by iteration to implement edge bundling. To solve the problem of excessive bending of some edges in segmental FDA bundling model, network was divided into different groups by CNM cluster algorithm. Lines in the same group were applied with edge consistent principle on basement of network group structure, and bundling level was calculated by the matching coefficient of edges. After edge bundling, the phenomena of curve's distortion decreased and curves became smoother. Domestic airline network data was chosen as experimental subject, and the network was simplified by two bundling models, and then the simplification result was analyzed. Experimental result shows that positions of nodes in the same group stand close to each other, and airline network shows distinct crossing skeleton. Airlines have north-south and west-east orientation bundle independently, which reveals the whole trends of airline network construction in these two directions. The bundling simplification algorithm introduced mentioned above has a wide applicability, and network visualized by this algorithm has good visual effect and readability.

Key words: network visualization; edge bundling; segmental force directed algorithm; cluster algorithm; bundling simplification algorithm

http://bhxb.buaa.edu.cn