

点群与线群的空间结构



1

点群的空间特征

点群的空间特征——基本概念

- 【点群】多个点的集合，这些点在集合上无大小的区别。地理中的许多实体可以抽象为点群，并且具有独立性和特定的含义。
- 点群的分布类型：向心分布、沿直线分布、沿线的两侧分布、均匀分布、多中心聚集、随机分布等
- 点群的空间特征：分布范围、数量、密度及其分布、分布中心、分布轴线、子群数量、子群的层次性结构等

点群的空间特征——凸壳的生成

- 分布范围特征可以通过计算点群的**凸壳**进行描述
- **【凸壳】**封闭点群中所有顶点的最小凸多边形，凸区域指要求区域内任意两点的连线仍在该区域内。
- 凸壳的算法

(1) Graham扫描算法

基本思想:设想点群内一点 O 并设为坐标原点, 这时点集 S 中所有各点相对轴 OX 有一个倾角。所有各点按倾角递增排序后, 如果某一点不是壳上顶点, 则它必然在两个壳顶点与点 O 形成的三角形内部。Graham扫描的实质是围绕已经按"倾角"排序的各顶点进行一次扫描, 在扫描过程中消去在凸壳内部的点, 留下以希望次序排列的壳顶点。

点群的空间特征——凸壳的生成

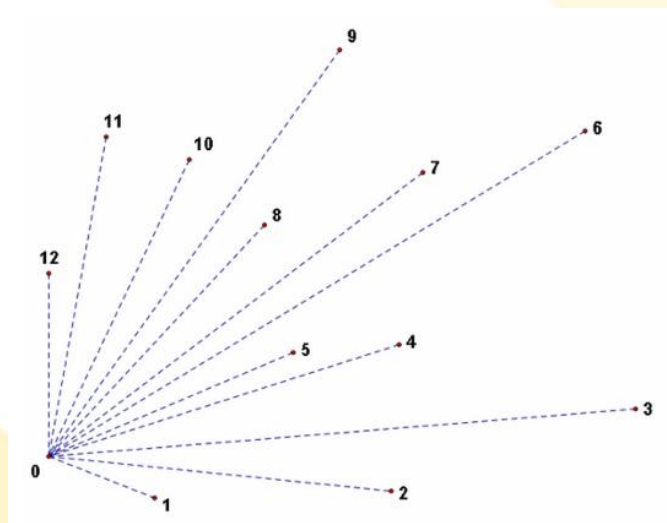
Graham扫描算法

基本步骤：

- ① **点集排序**。选出输入点集 S 中 x 坐标最小的点, 若这样的点不唯一则再由其中选出 y 坐标最小的点, 设为 O 。设想有一条从 O 向右的射线 OX , 对点集中其余每一点 P , 计算倾角 POX , 再按倾角排序, 得点序列 $Q=\{q_1, q_2, \dots\}$;



原始点群



排序结果

点群的空间特征——凸壳的生成

Graham扫描算法

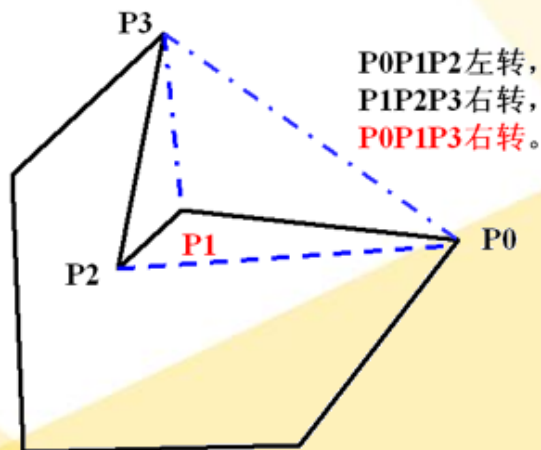
基本步骤(续):

- ② 扫描准备。将原点0, q_1 加入到凸包的点集V中, $v_1=0$, $v_2=q_1$ 。
- ③ 扫描。依次将下一顶点 q_i ($i=2, \dots, n$) 加入V中, $v_j=q_i$ ($j=3, 4, \dots$), 判断V中三个相继的点 v_{j-2}, v_{j-1}, v_j 是否形成一个左转, 若是则 $q_i \in V$; 否则, 从V中删除 q_i 。
- 左转判断方法: 对给出的三点, 设它们的坐标是 (x_1, y_1) , (x_2, y_2) , (x_3, y_3) , 计算下面的行列式

$$\Delta = \begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix}$$

若 $\Delta > 0$, 则是左转;
若 $\Delta < 0$, 则是右转;
若 $\Delta = 0$, 则三点共线

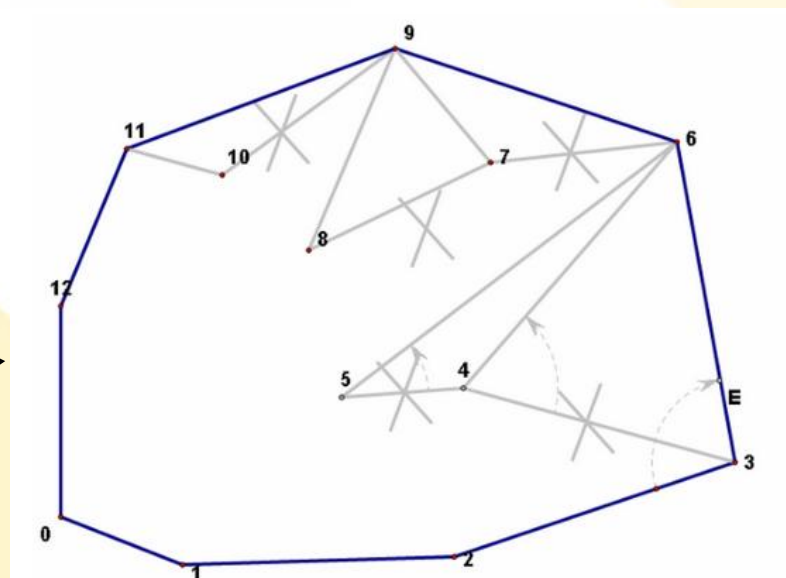
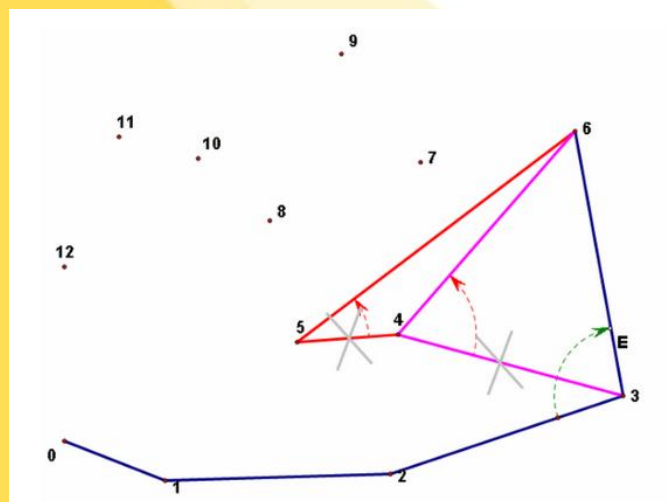
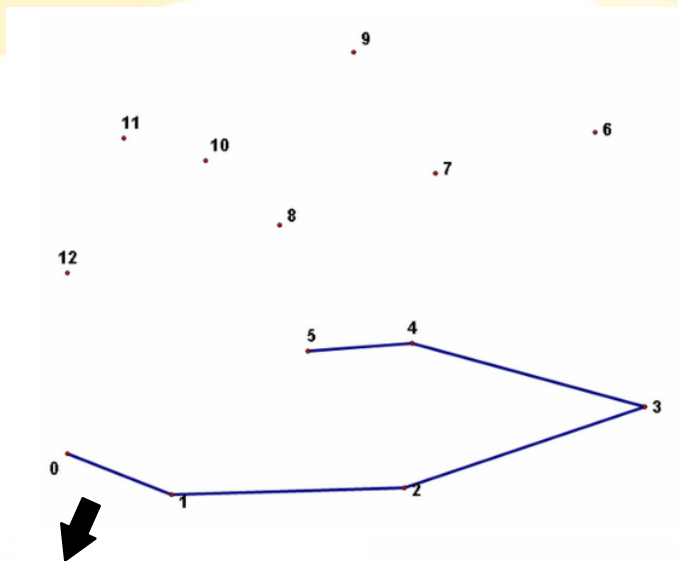
- ④ 依次连接V中的顶点, 凸壳生成结束



点群的空间特征——凸壳的生成

Graham扫描算法

基本步骤:



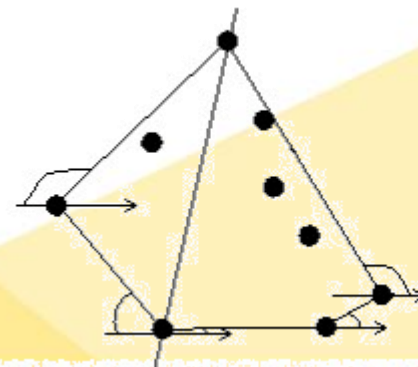
点群的空间特征——凸壳的生成

(2) Jarvis算法

- ① 找出横坐标x的最大和最小点，分别记做P和Q，过P、Q的直线将平面分成左右两个部分。
- ② 处理左边的点。重复下述步骤：
 - I. 以P为起点，向其他“左边部分”的点做射线
 - II. 找一个到X轴负方向夹角最小的射线，将这条射线的另一端点加入到凸包，将这个点记作P。直到Q被加入凸包。
- ③ 处理右边的点。重复下述步骤：
 - I. 以P为起点，向其他“右边部分”的点做射线
 - II. 找一个到X轴正方向夹角最小的射线，将这条射线的另一端点加入到凸包，将这个点记作P。直到Q被加入凸包。

其中，倾角计算方法如下：

$$A = \frac{(X_P - X_0)}{\sqrt{(X_P - X_0)^2 + (Y_P - Y_0)^2}}$$



点群的空间特征——凸壳的生成

(3) Larkin改进算法

基本步骤：

- ① 建立初始凸壳。将平面点集中各点的横坐标 x 和纵坐标 y 进行比较，分别找出 x ， y 最大和最小的点（或 $x+y$ 和 $x-y$ 的极值点），这些点一定在点集的凸壳上，将这些点连接，形成初始凸壳。
- ② 修改凸壳。任取初始凸壳的一边，计算位于此边右侧的各个点到这条边的距离，求出最大距离。如果存在这样的点，或者最大距离为0，则将这个点插入到链表中，并位于这条边的两个顶点之间，否则，不插入。
- ③ 凸壳生成。统计每一次没有插入点的操作次数，如果等于链表中顶点的总数，则终止判断，生成凸壳。

2

基于邻近图的点群聚类方法

常见邻近图的概念及其构造——基本概念

- 邻近图 (Neighbourhood Graph)

利用边将具有邻近关系的点连接起来的图为邻近图。

- 常见邻近图

根据对于“邻近范围”理解的不同，形成许多不同类型的邻近图，例如

——NNG (Nearest Neighborhood Graph, 最邻近图)

——MST (Minimum Spanning Tree, 最小支撑树)

——RNG (Relative Neighborhood Graph, 相对邻近图)

——UG (Urquhart Graph)

——GG (Gabriel Graph)

——DT (Delaunay Triangulation Graph, Delaunay三角网)

常见邻近图的概念及其构造——基本概念

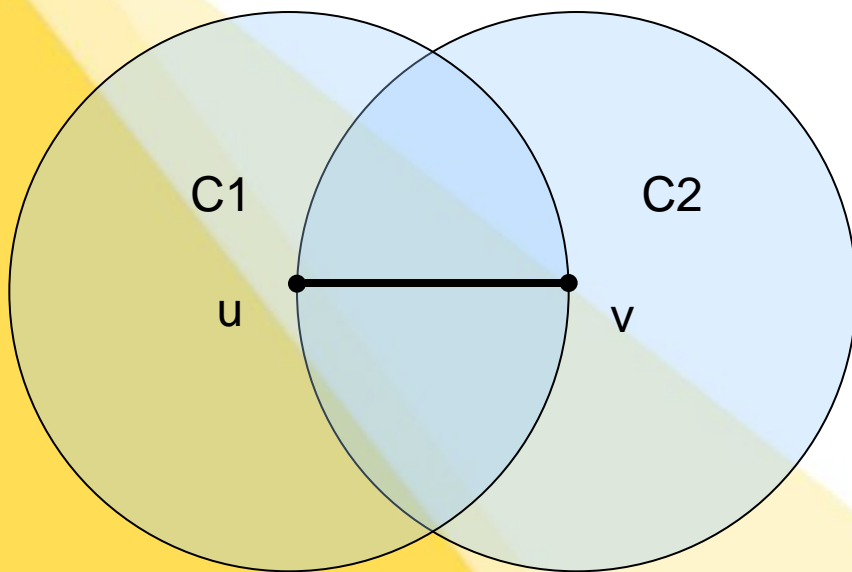
- 设在二维欧氏空间里，有邻近图 $G(V, E)$
 - 点集 $V = \{v_i \mid i=1, 2, \dots, n\}$
 - 边集 $E = \{e(u, v) \mid u \in V, v \in V, u \neq v\}$
 - G 内任意两点间距离表示为 $d(u, v)$

常见邻近图的概念及其构造——基本概念

- NNG (Nearest Neighborhood Graph)

定义：若 $e(u, v) \in \text{NNG}$ ，则 $d(u, v) \leq \min(d(w, u))$ ， $u, v, w \in V$ ， $u \neq v \neq w$

= 以 u, v 为圆心， uv 为半径的两个圆的并集内，不含其他点



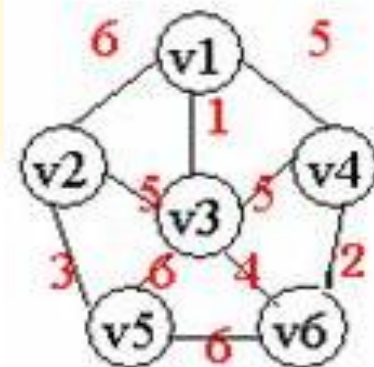
若 $e(u, v)$ 是 NNG 中的边，
则圆 C1 与 C2 内不能存在其他点

常见邻近图的概念及其构造——基本概念

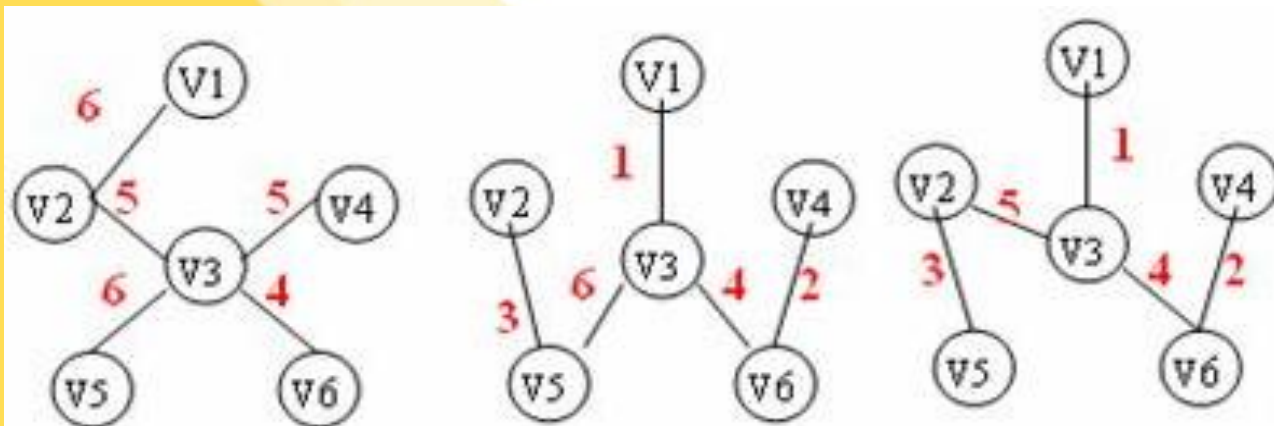
- MST(Minimum Spanning Tree)**

定义：包含原始点集 V 中所有点，且使得边合集 E 中的所有边的权值(或边长)之和最小的树状图。

MST为无回路的连通图。



所有边及对应的权值



a. 权值26

b. 权值16

c. 权值15

生成树c为最小生成树，c中的边 $E \in \text{MST}$

常见邻近图的概念及其构造——基本概念

- MST(Minimum Spanning Tree)

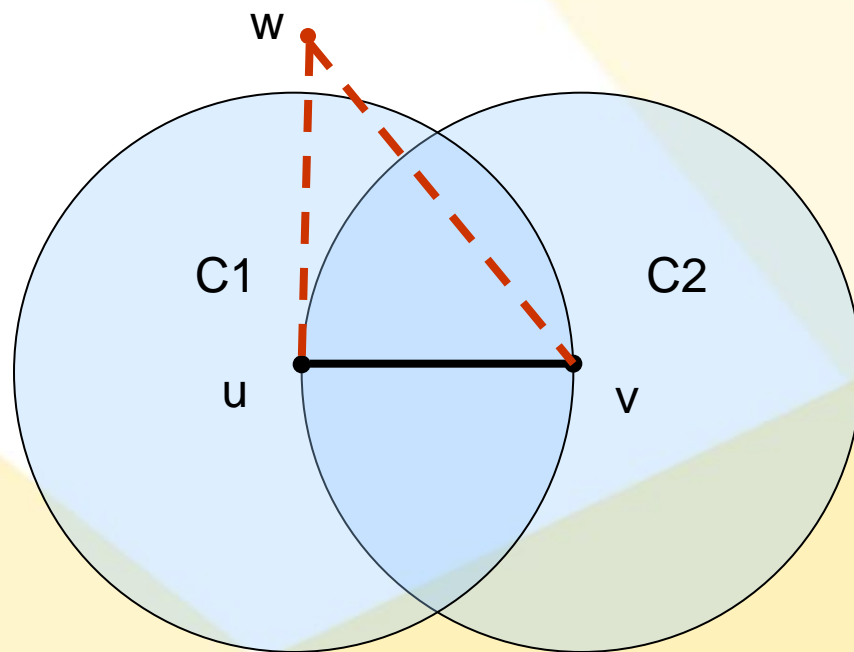
定理：若 $e(u, v)$ 是NNG的边，则其一定是MST的边，即 $NNG \subseteq MST$

- 证明：假设 $e(u, v) \in NNG$, 且 $e(u, v) \notin MST$ 成立

- ① $\because e(u, v) \in NNG \therefore$ 以 u, v 为圆心, $d(u, v)$ 为半径的圆 $C1$ 与 $C2$ 内不存在其他的任一点 $w, w \in V$

- ② 又 $\because e(u, v) \notin MST \therefore$ 假设 $w \in V$ 是在含 u 的子树内, 则存在一点 $w \in V$ 使得 $d(u, w)$ 小于 $d(u, v) \Rightarrow C1$ 内含有点 w

- ①与②矛盾, 所以假设不成立。

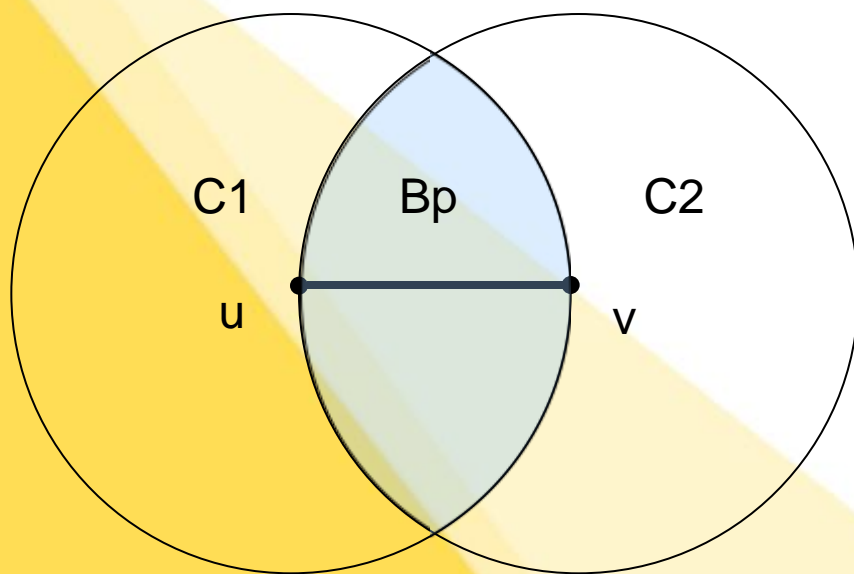


常见邻近图的概念及其构造——基本概念

- RNG(Relative Neighborhood Graph)

定义：若 $e(u, v) \in \text{RNG}$ ，则 $d(u, v) \leq \text{Max}(d(w, u))$ ， $u, v, w \in V$ ， $u \neq v \neq w$

=以 u, v 为圆心， $d(u, v)$ 为半径的两个圆的交集内，不含其他点。



若 $e(u, v)$ 是RNG中的边，
则圆C1与C2的交集 B_p 不能
存在其他点

常见邻近图的概念及其构造——基本概念

- RNG(Relative Neighborhood Graph)

定理：若 $e(u, v)$ 属于MST的边，则其一定满足RNG边的定义，即 $MST \subseteq RNG$

- 证明：假设 $e(u, v) \in MST$, 且 $e(u, v) \notin RNG$ 成立

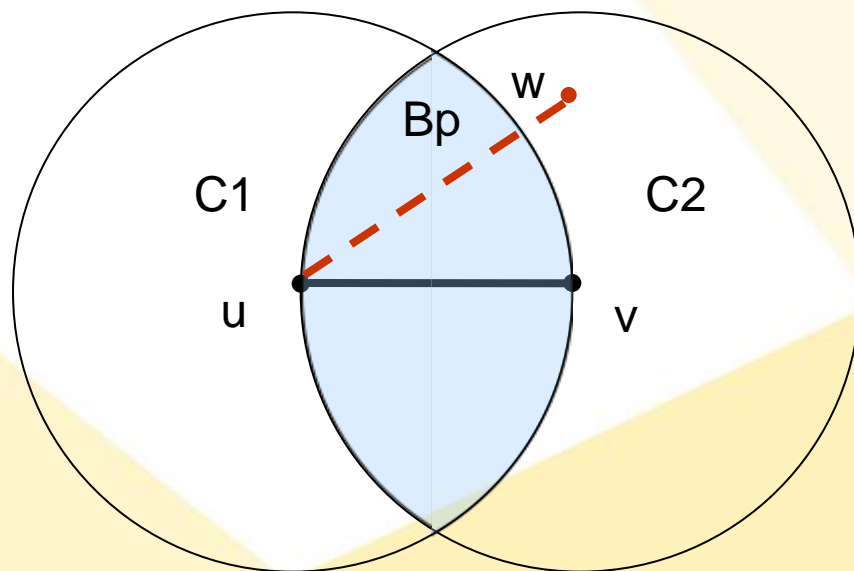
①假设 $w \in V$ 是在含 u 的子树内

$\because e(u, v) \in MST$

\therefore 不存在一点 $w \in V$ 使得 $d(u, w)$ 小于 $d(u, v) \Rightarrow C1$ 内没有点

② 又 $\because e(u, v) \notin RNG$, 则 $C1$ 与 $C2$ 的交集内存在一点 $w \Rightarrow d(u, w) < d(u, v) \Rightarrow e(u, v) \notin MST$

①与②矛盾，所以假设不成立。



常见邻近图的概念及其构造——基本概念

- UG(Urquhart Graph)

定义：若 $e(u, v) \in \text{UG}$ ，则 $d(u, v) \leq \text{Max}(d(w, u))$ ， u, v, w 为DTG中一个三角形的三个顶点。

定理： $\text{RNG} \subseteq \text{UG}$

UG与RNG相比，生产条件相似，但算法简单、速度快，且边不多于RNG的2%。

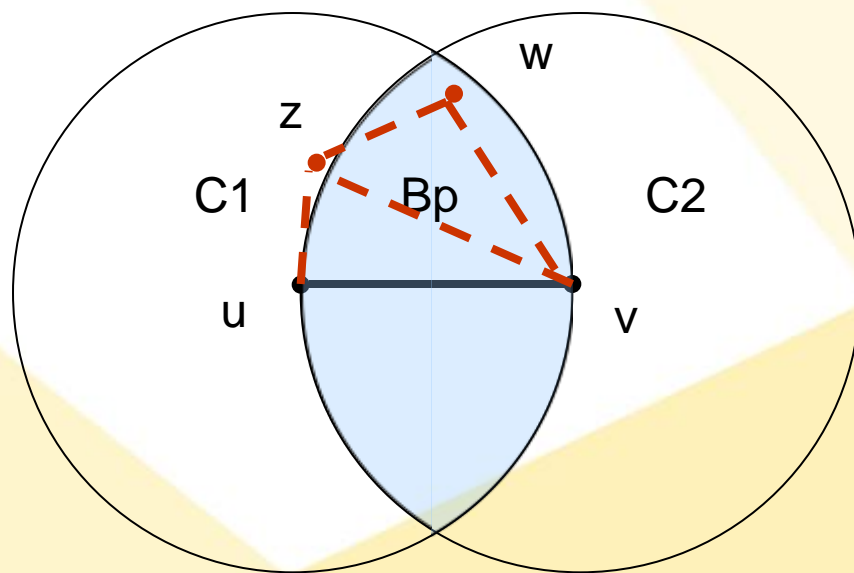
- 证明：

- $\because w$ 在 $C1, C2$ 的交集内

- \therefore 边 $e(u, v)$ 不是RNG的一条边。

- 又 $\because u, v, w$ 不构成DTG的三角形，不考察 u, v, w 间的距离关系是否满足定义

- $\therefore e(u, v)$ 是UG的一条边

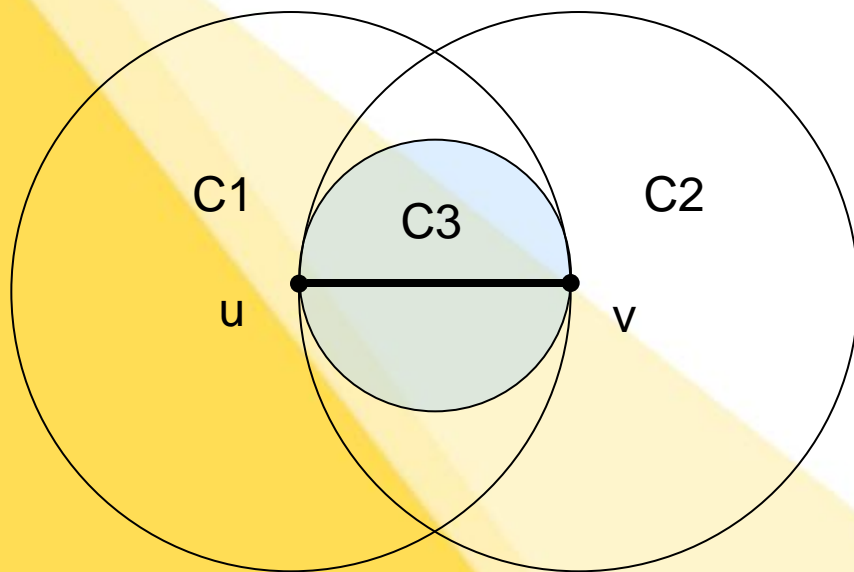


常见邻近图的概念及其构造——基本概念

- GG(Gabriel Graph)

定义：若 $e(u,v) \in GG$ ，则不存在点 $w \in V$ ，使得 $d^2(u,w) + d^2(v,w) < d^2(u,v)$ 。

=以 $d(u,v)$ 为直径的圆内及圆周不含其他点。



若 $e(u, v)$ 是RNG中的边，
则圆C3内不能存在其他点

常见邻近图的概念及其构造——基本概念

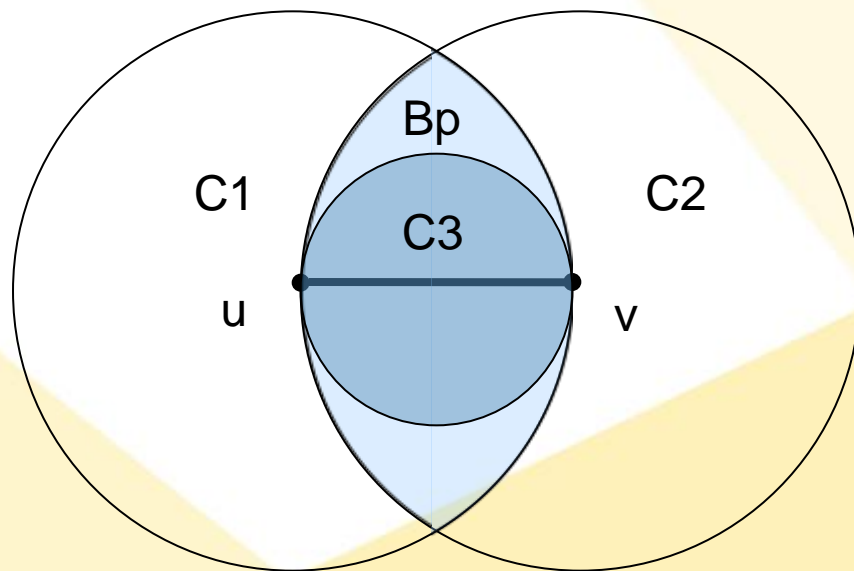
- RNG(Relative Neighborhood Graph)

定理：如果 $e(u,v) \in UG$ ，则它一定同时属于GG，即 $UG \subseteq GG$

- 证明：

$\because e(u,v) \in UG$ ，以 u, v 为圆心， $d(u, v)$ 为半径的两个圆的交集 B_p 内，不含其他点

\therefore 以 $d(u, v)$ 为直径的圆 $C3$ 内及圆周不含其他点 $\Rightarrow e(u, v) \subseteq GG$



常见邻近图的概念及其构造——基本概念

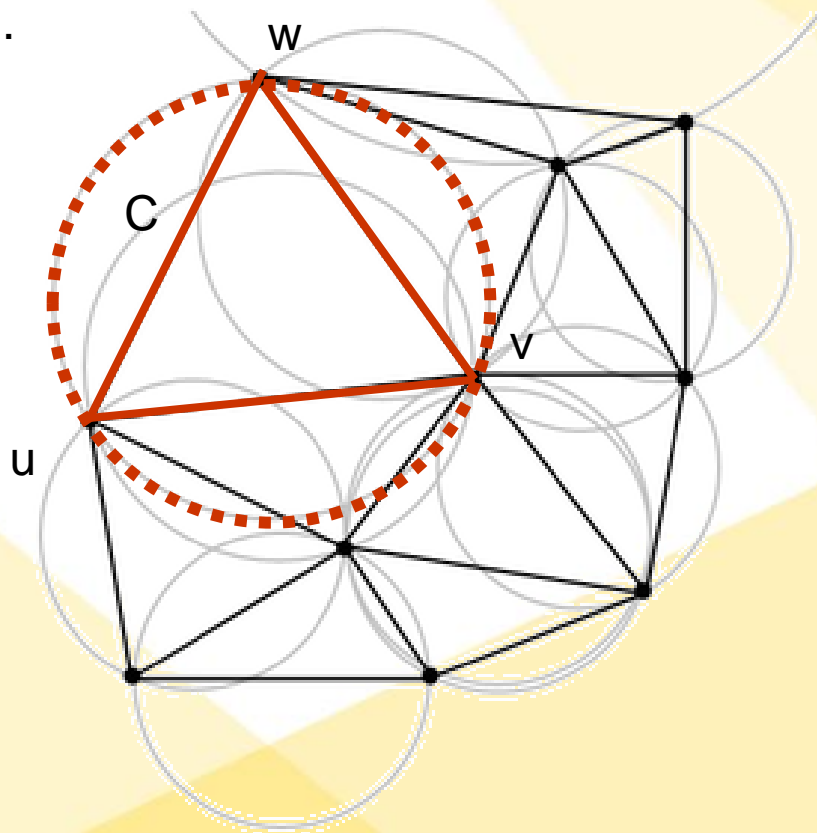
- DTG(Delaunay Triangulation Graph)

定义：若 $e(u,v) \in GG$ ，则不存在点 $w \in V$ ，使得 $d^2(u,w) + d^2(v,w) < d^2(u,v)$ 。

=以 $d(u,v)$ 为直径的圆内及圆周不含其他点。

若 $\triangle uvw$ 是DTG中的边，

则圆C内不能存在其他点



常见邻近图的概念及其构造——基本概念

- DTG(Delaunay Triangulation Graph)

定理：如果 $e(u,v) \in GG$ ，则它一定同时属于DTG，即 $GG \subseteq DTG$

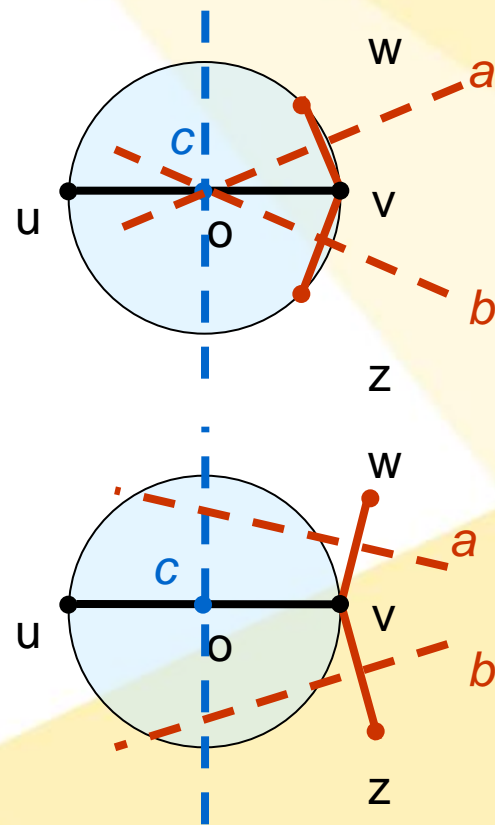
- 证明：设：a, b, c分别为w与v, v与z, u与v之间的Voronoi边。

\therefore 若w与z在以uv为直径的圆上. 则a, b一定交u, v之间的Voronoi边c于原点o \Rightarrow c的长度为0

若w与z在圆外(各自保持在原象限内)，则a, b与c的交点将分别位于o点的正上方与正下方 \Rightarrow c大于0;

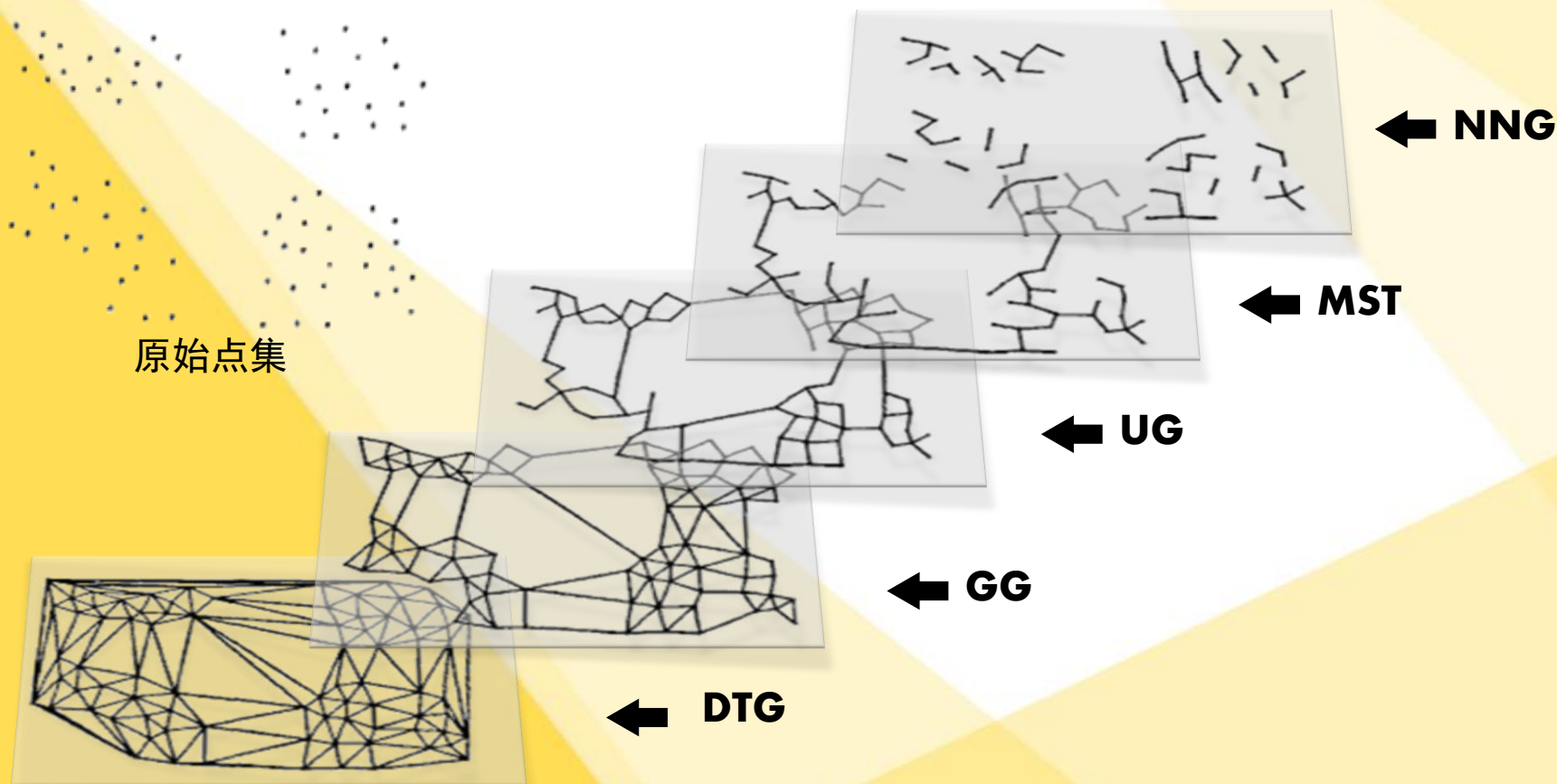
$\therefore e(u,v) \in GG \Rightarrow$ 以uv为直径的圆内和圆周上没有其他点 $\Rightarrow c \neq 0$

根据Voronoi与DTG间的对偶性, $c \neq 0 \Rightarrow e(u,v) \in DTG$



常见邻近图的概念及其构造——基本概念

综上所述，存在拓扑关系链： $NNG \subseteq MST \subseteq RNG \subseteq UG \subseteq GG \subseteq DTG$



常见邻近图的概念及其构造——构造方法

- 依据定义构造

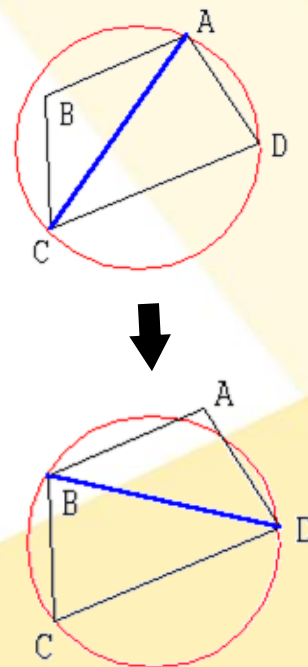
- 在DTG基础上依次构造：构造DTG→构造GG→构造RNG→构造NNG
- MST可直接从DGT构造

- 构造DTG

- (1) 逐点插入法

- ① 定义一个包含所有点的初始多边形(矩形, 三角形或凸壳等)
- ② 在多边形内选择一点a, 与初始多边形构建初始大三角网
- ③ 插入多边形内一点p, 连接p与含点p的三角形 ΔP 的顶点生产3个新的三角形; 如果p在 ΔP 的边上, 则必然在另一三角形 $\Delta P'$ 的边上, 此时连接p与 ΔP , $\Delta P'$ 的顶点, 产生4个新的三角形。

→ 通过边交换方法(LOP算法)规格化三角剖分。利用DT的空圆定义, 对加入三角网中有公共边的两个三角形的外接形进行判断, 如果外接圆包含4个顶点则交换两个三角形组成的四边形的对角线。



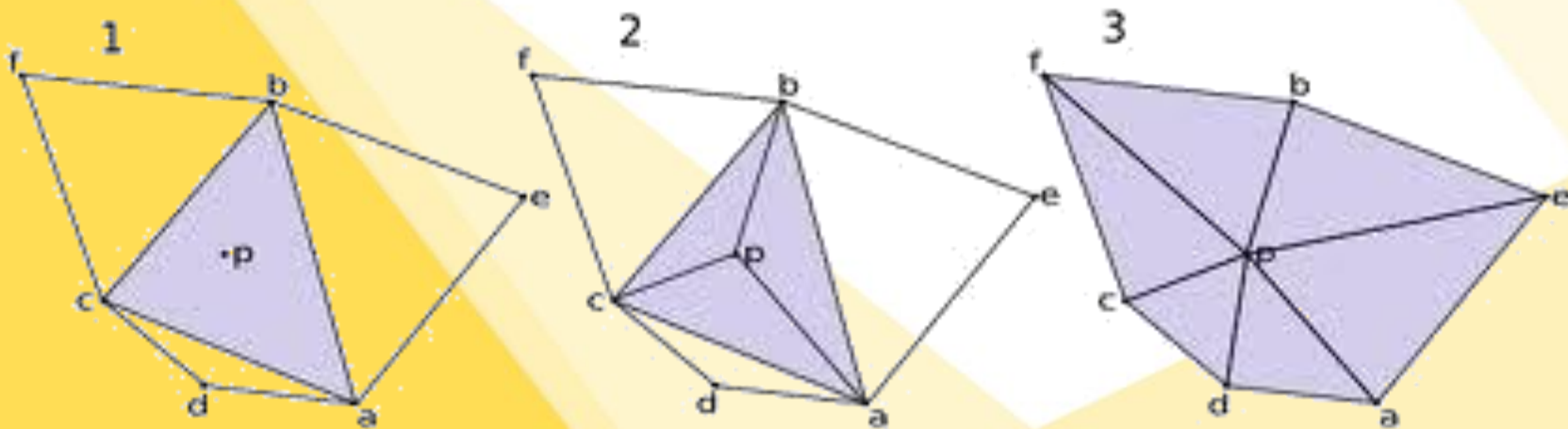
常见邻近图的概念及其构造——构造方法

- 构造DTG

- (1) 逐点插入法

- ④利用步骤③迭代处理数据，至加入所有初始数据点结束

- ⑤若初始多边形不为凸壳，中删除包含初始多边形任意顶点的三角形，三角网构建完毕。



常见邻近图的概念及其构造——构造方法

- 构造DTG

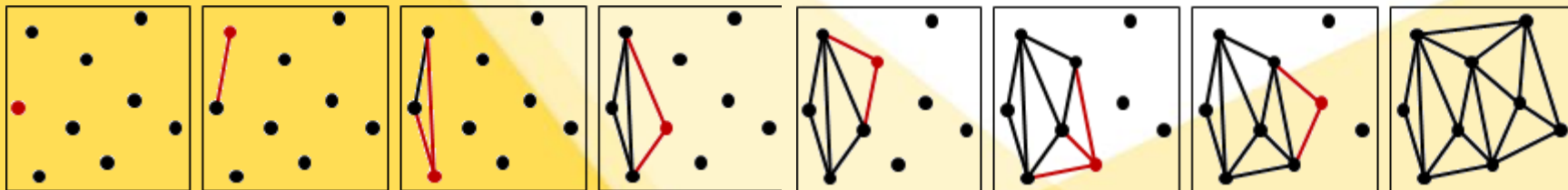
(2) 三角网生长法

① 构造初始三角形。任意选择点集P中的一点P1，并从点集P中选取距离P1，最近的点P2。两点连线成为DT的一条边；如果点P3使 $\triangle P1P2P3$ 满足空圆特性，且 $\angle P1P3P2$ 最大，则 $\triangle P1P2P3$ 为初始三角形。

② 创建初始边列表和初始三角形列表

③ 如果边列表不为空，则从边列表中取出一条边；寻找满足空圆特性的点插入三角网中；与新插入的点构建新的边和三角形，并加入到边列表和三角形列表中；

④ 重复执行步骤③，直至初始列表为空。

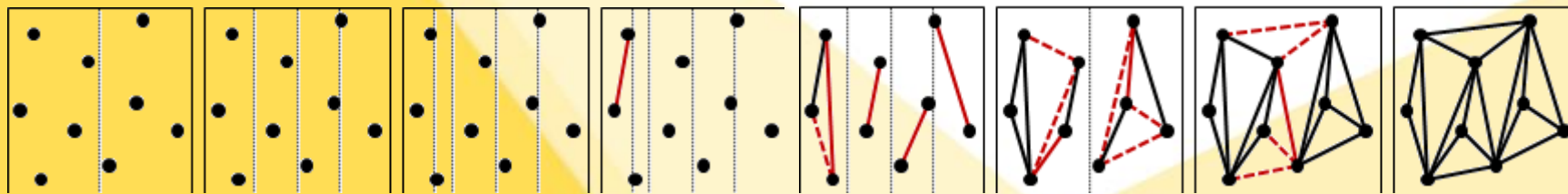


常见邻近图的概念及其构造——构造方法

- 构造DTG

(3) 分治算法

- ① 根据点坐标对点集进行排序；
- ② 将排序后的点集二分成两个点集VL，VR；
- ③ 如果点集VL或VR中点的数量大于3，对VL，VR分别重复步骤②；
- ④ 在VL，VR中生成三角网，同时用LOP算法优化，分别形成DTL与DTR。
- ⑤ 递归合并DTL与DTR，直至合成为一个DT。



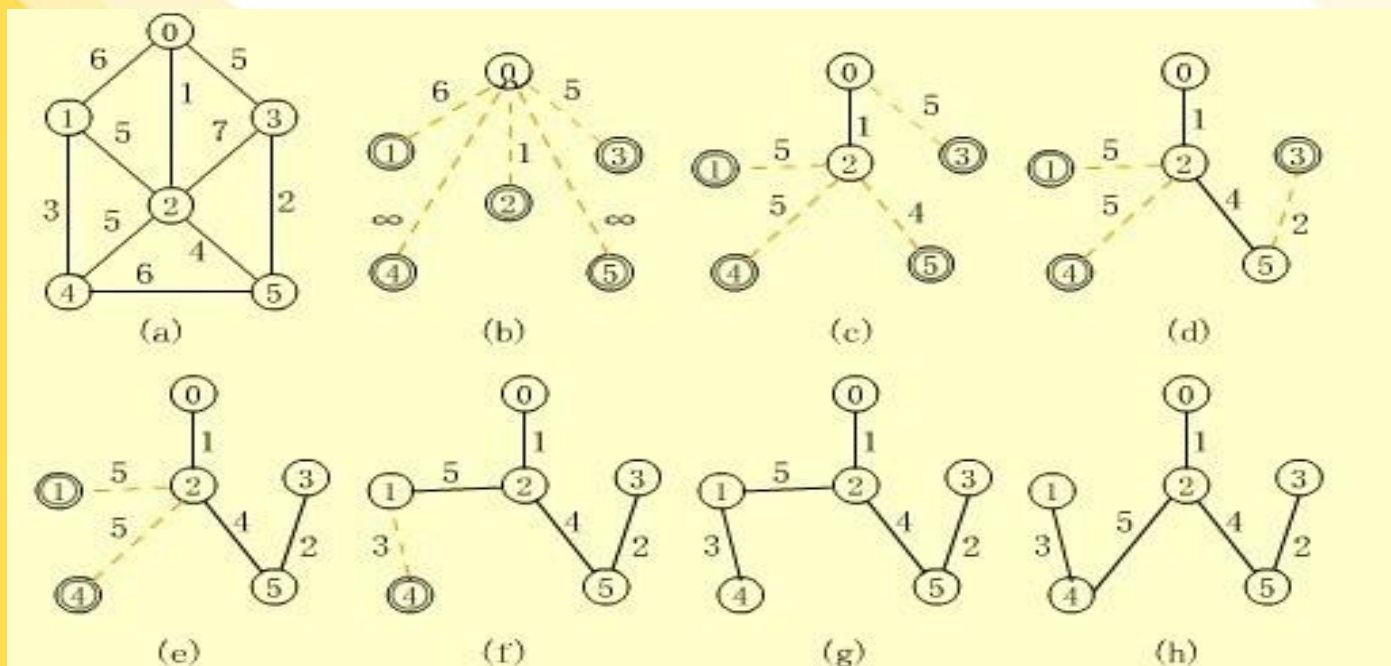
常见邻近图的概念及其构造——构造方法

• MST的构造

设 $D=(V, E)$ 为点群 V 的DGT, $T=(U, TE)$ 为 D 的最小生成树。

① 普里姆算法(Prim)

T 的初始状态为 $U=\{v_0 | v_0 \in V\}$, $E=$, 然后重复执行下述操作：在所有 $d(u, v)$ ($u \in U$, $v \in (V-U)$) 中找一条代价最小的边并入集合 TE , 同时 v 并入 U , 直至 $U=V$ 为止

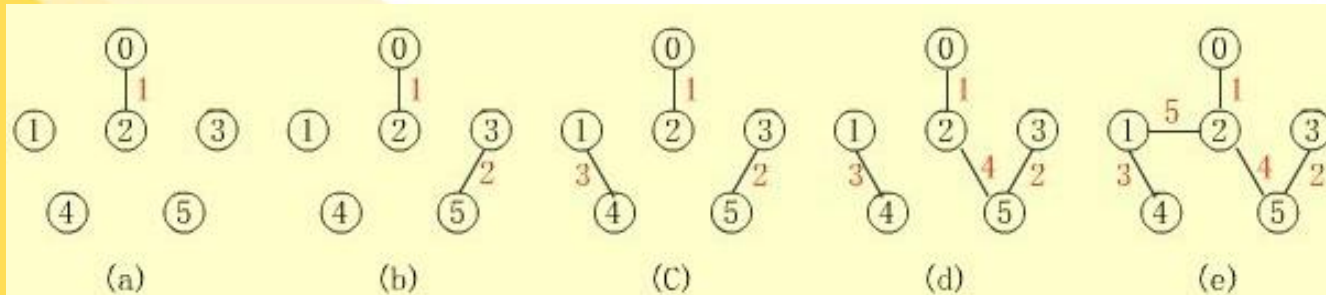
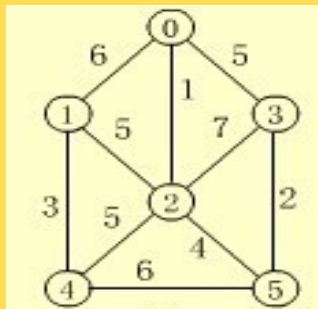


常见邻近图的概念及其构造——构造方法

• MST的构造

② 克鲁斯卡尔算法(Krusal)

T的初始状态为 $U=V$, $E=(T$ 中各顶点各自构成一个连通分量), 然后按照边的权值由小到大的顺序, 依次考察E中的各条边。若被考察的边的两个顶点属于T的两个不同的连通分量, 则将此边加入到TE中去, 同时把两个连通分量连接成一个连通分量; 若被考察边的两个结点属于同一个连通分量, 则舍去此边, 以免造成回路, 如此下去, 当T中的连通分量个数为1时, 结束考察。



基于邻近图聚类的几个重要概念

设有一类C，为一个点或点群，表示为 $V(C) = \{v_i \mid i=1, 2, \dots, n\}$

- 内边集 C^0

$C^0 = \{ e(v_i, v_j) \mid v_i \in V(C), v_j \in V(C) \}$ 。即两个点都在类C点集内的边。

- 外边集 C^1

$C^1 = \{ e(v_i, v_j) \mid (v_i \in V(C), v_j \notin V(C)) \cup (v_i \notin V(C), v_j \in V(C)) \}$ 。即仅有一个点在类C点集内的边。

※ 当类 C_1 与 C_2 合并为新的类U，则 $V(U) = V(C_1) \cup V(C_2)$ 。若设 $K(C)$ 为类C所有的内边与外边的合集，则 $K(U) = K(C_1) \cup K(C_2)$

基于邻近图聚类的几个重要概念

• 密度

① 内密度 $M(C^0)$

$M(C^0) = \text{Median}\{d(v_i, v_j) \mid e(v_i, v_j) \in C^0\}$, 即所有内边的中值。

② 外密度 $M(C^1)$

$M(C^1) = \text{Median}\{d(v_i, v_j) \mid e(v_i, v_j) \in C^1\}$, 即所有外边的中值。

• 偏差 $MM(C)$

$MM(C) = \text{Median}\{|\text{Ads}(d(v_i, v_j) - M(C^0))| \mid e(v_i, v_j) \in (C^0 \cup C^1)\}$, 即所有内边及外边与的内密度差的绝对值的中值。

• 密度置信区间 $T(C)$

$$T(C) = [M(C^0) - MM(C), M(C^0) + MM(C)]$$

• 类间距离 $d(C1, C2)$

$d(C1, C2) = \text{Median}\{d(v_i, v_j) \mid e(v_i, v_j) \in (C1 \cap C2)\}$, 即所有公共外边的中值。

基于邻近图的点群聚类方法——空间聚类约束条件

空间聚类约束条件是判断两类是否合并的依据

- 距离适应性约束

$d(C1, C2) \in T(C1)$ 且 $d(C1, C2) \in T(C2)$ ，即类间距离在两类的公共置信区间内。

- 内密度适应性约束

①当 $C10 \neq \emptyset$ 且 $C20 \neq \emptyset$ ($C1, C2$ 均不是只含有一个点)时

$M(C10) \in T(C2)$ 且 $M(C20) \in T(C1)$ ，即两类的内密度在彼此的置信区间内

②当 $C10 = \emptyset$ 或 $C20 = \emptyset$ ($C1, C2$ 中有一个类只含有一个点)时

$M(C10) \in T(C2)$ 或 $M(C20) \in T(C1)$ ，即只需其中一类的内密度在另一类的置信区间内

基于邻近图的点群聚类方法——空间聚类约束条件

空间聚类约束条件是判断两类是否合并的依据

- 偏差适应性约束

$MM(C1 \cup C2) \leq \max(MM(C1), MM(C2))$ ，即 $C1$ ， $C2$ 合并后的新类的偏差度不能同时大于 $C1$ 和 $C2$ 的偏差值。

- 外密度适应性约束(可选)

$(C1 \cup C2)_1 \neq \emptyset$ ，则 $M((C1 \cup C2)_1) \geq \max(M(C1_1), M(C2_1))$ ，即当 $C1$ ， $C2$ 合并后的新类存在外边，则新类的外密度同时大于 $C1$ 和 $C2$ 的外密度。

※在满足前3项后，第4项必定满足，判断时不必另外讨论。提出为说明应该优先考虑合并距离较近的类。

基于邻近图的点群聚类方法——基本思想

- 在确定初始类的基础上，利用邻近图根据距离、内密度和偏离适应性3个约束条件逐步合并两个类，直到没有互为有效邻近类为止结束聚类。

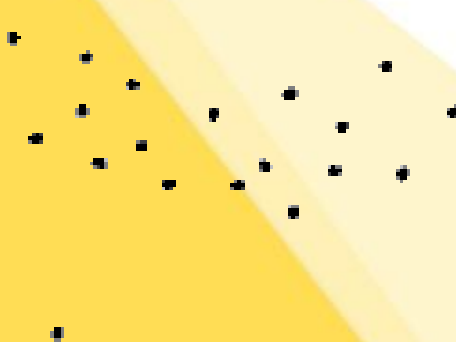
基于邻近图的点群聚类方法——具体步骤及举例

(1) 确定初始类

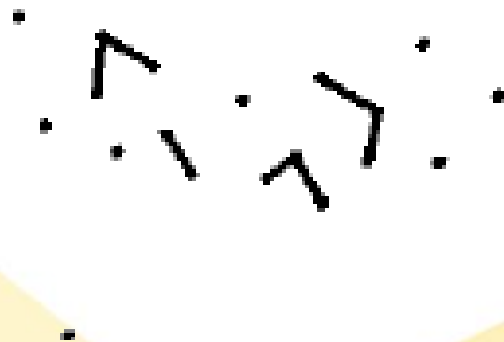
激活点集的某一种邻近图的边，并确定初始类，每个类是该邻近图上有边连接的可以连通的点集。

初始类可以是每个点，也可以是用别的邻近图聚类所得到的结果。

e. g. 将初始点群每个点看做一类，基于NNG由下至上逐步合并类，将聚类结果作为初始类（8个孤立点自成一类，非孤立点类有4个，共计12类）



初始点群



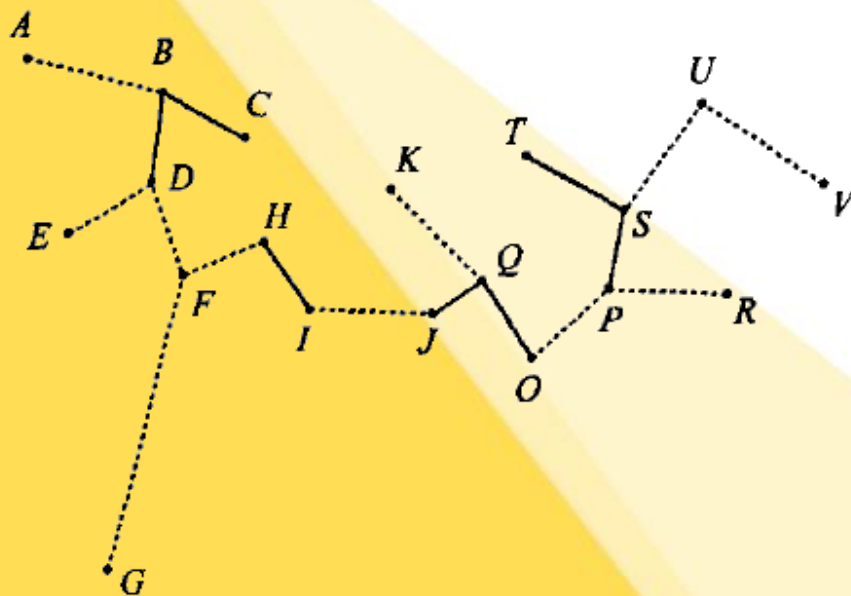
基于NNG合并类结果(as初始类)

基于邻近图的点群聚类方法——具体步骤及举例

(2) 计算边长

激活当前或更高层次的邻近图，计算邻近图中每条边的长度。

e. g. 激活MST（虚线为在NNG基础上新增的线），并计算边长。



边	长度	边	长度
$e(A,B)$	5.34	$e(J,Q)$	2.24
$e(B,C)$	3.65	$e(Q,K)$	4.9
$e(B,D)$	3.44	$e(Q,O)$	3.48
$e(D,E)$	3.78	$e(O,P)$	3.9
$e(D,F)$	3.76	$e(P,R)$	4.59
$e(F,H)$	3.34	$e(P,S)$	3.1
$e(F,G)$	11.54	$e(S,T)$	4.2
$e(H,I)$	3.07	$e(S,U)$	4.95
$e(I,J)$	4.83	$e(U,V)$	5.56

基于邻近图的点群聚类方法——具体步骤及举例

(3) 计算参数

计算各个类的内密度、偏差、置信区间以及邻近类之间的距离，并按内密度为主要关键词、偏差为次要关键词从小到大排序。

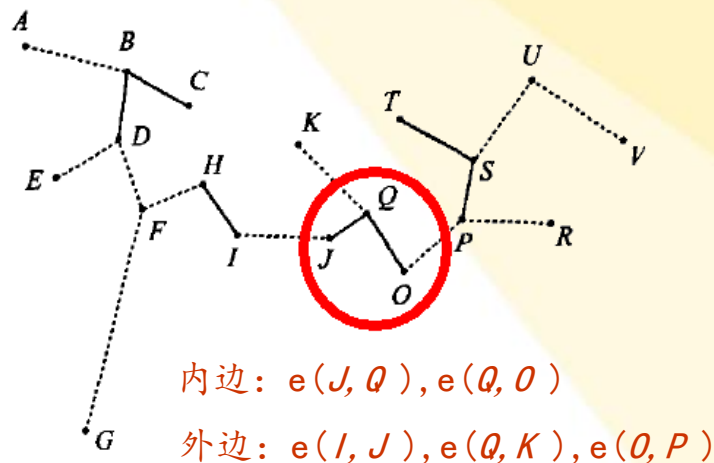
e. g. 以类 $c(J, Q, O)$ 为例

- 内密度 $M(c(J, Q, O))$

$$= \text{Median}\{d(J, Q), d(Q, O)\} = (2.24 + 3.48) \div 2 = 2.86$$

- 偏差 $MM(c(J, Q, O))$

$$\begin{aligned} &= \text{Median}\{ \text{Abs}(d(J, Q) - M(c(J, Q, O))), \text{Abs}(d(Q, O) - M(c(J, Q, O))), \\ &\quad \text{Abs}(d(I, J) - M(c(J, Q, O))), \text{Abs}(d(Q, K) - M(c(J, Q, O))), \text{Abs}(d(O, P) - M(c(J, Q, O))) \} \\ &= \text{Median}[|2.24 - 2.86| + |3.48 - 2.86| + |4.83 - 2.86| + |4.9 - 2.86| + |3.9 - 2.86|] \\ &= \text{Median}[0.62, 0.62, 1.97, 2.04, 1.04] = 1.04 \end{aligned}$$



基于邻近图的点群聚类方法——具体步骤及举例

- 置信区间 $T(c(J, Q, O))$

$$=[M(c(J, Q, O)^0) - MM(c(J, Q, O)), M(c(J, Q, O)^0) + MM(c(J, Q, O))]$$

$$=[2.86 - 1.04, 2.86 + 1.04] = [1.82, 3.9]$$

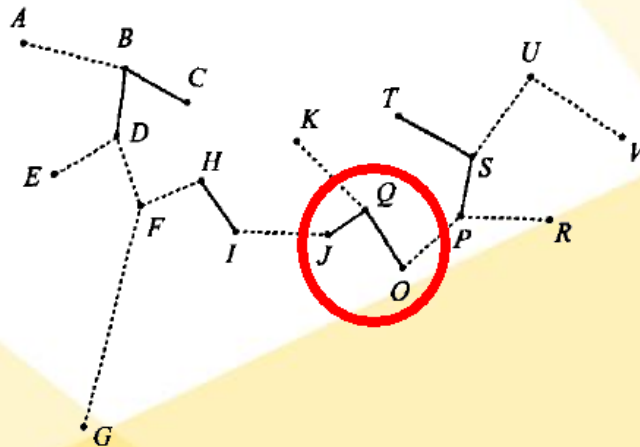
- 邻近类间距离

$c(J, Q, O)$ 与 $c(H, I)$ 、 $c(K)$ 、 $c(T, S, P)$ 有公共外边，由于MST是无回路图，所以两类间仅可能有一条公共外边，两类间距离等于公共外边长度

$$d(c(J, Q, O), c(H, I)) = d(I, J) = 4.83$$

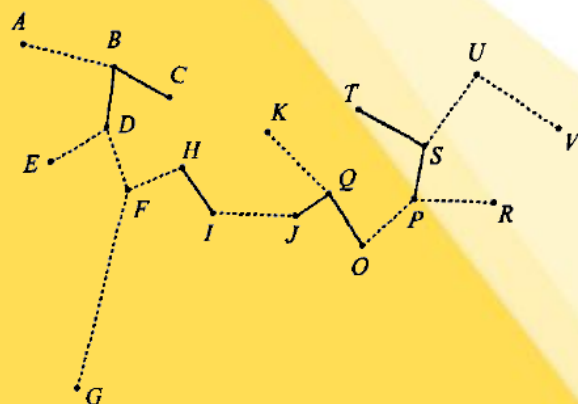
$$d(c(J, Q, O), c(K)) = d(Q, K) = 4.9$$

$$d(c(J, Q, O), c(T, S, P)) = d(O, P) = 3.9$$



基于邻近图的点群聚类方法——具体步骤及举例

- 同理计算其他类，并按内密度和偏差值从小到大排序



类	内边	外边	内密度	偏差	置信区间	邻近类	类间距离
$c(F)$	-	$e(D,F)$	0	3.76	[0, 3.76]	$c(B,C,D)$	3.76
		$e(F,G)$				$c(G)$	11.54
		$e(F,H)$				$c(H,I)$	3.34
$c(E)$	-	$e(D,E)$	0	3.78	[0, 3.78]	$c(B,C,D)$	3.78
$c(R)$	-	$e(P,R)$	0	4.59	[0, 4.59]	$c(T,S,P)$	4.59
$c(K)$	-	$e(Q,K)$	0	4.9	[0, 4.9]	$c(J,Q,O)$	4.9
$c(A)$	-	$e(A,B)$	0	5.34	[0, 5.34]	$c(B,C,D)$	5.34
$c(U)$	-	$e(S,U)$	0	5.525	[0, 5.525]	$c(T,S,P)$	4.95
		$e(U,V)$				$c(V)$	5.56
$c(V)$	-	$e(U,V)$	0	5.56	[0, 5.56]	$c(U)$	5.56
$c(G)$	-	$e(F,G)$	0	11.54	[0, 11.54]	$c(F)$	11.54
		$e(I,J)$				$c(H,I)$	4.83
		$e(Q,K)$				$c(K)$	4.9
$c(J,Q,O)$	$e(J,Q)$	$e(O,P)$	2.86	1.04	[1.82, 3.9]	$c(T,S,P)$	3.9
		$e(F,H)$				$c(F)$	3.34
		$e(I,J)$				$c(J,Q,O)$	4.83
$c(B,C,D)$	$e(B,C)$	$e(A,B)$	3.545	0.215	[3.33, 3.76]	$c(A)$	5.34
		$e(D,E)$				$c(E)$	3.78
		$e(D,F)$				$c(F)$	3.76
$c(T,S,P)$	$e(P,S)$	$e(O,P)$	3.65	0.55	[3.1, 4.2]	$c(J,Q,O)$	3.9
		$e(P,R)$				$c(R)$	4.59
		$e(S,U)$				$c(U)$	4.95

基于邻近图的点群聚类方法——具体步骤及举例

(4) 约束条件判断

从顺序表中取出最前面的一个类C，判断它与其邻近类是否满足距离，内密度与偏差适应性约束条件。

- ① 若不满足，将类C从顺序表中删除；
- ② 若满足，则合并两类(如果存在多个满足条件的有效邻近类，则与距离最近或密度最相近的有效类合并)，生成新类U。在顺序表内删除合并前两类，并添加新类U及其参数，重新排序。

基于邻近图的点群聚类方法——具体步骤及举例

(4) 约束条件判断

e. g. 判断类F的邻近类是否满足空间约束条件

$c(H, I)$ 与 $c(B, C, D)$ 均为有效邻近类, 但 $d(c(F), c(H, I)) < d(c(F), c(B, C, D))$, 所以优先合并 $c(F)$ 与 $c(H, I)$



初始类	C_1	$c(F)$		
置信区间	$T(C_1)$	$[0, 3.76]$		
内密度	$M(C_1^0)$	0		
偏差	$MM(C_1)$	3.76		
邻近类	C_2	$c(B, C, D)$	$c(G)$	$c(H, I)$
类间距离	$d(C_1, C_2)$	3.76	11.54	3.34
置信区间	$T(C_2)$	$[3.33, 3.76]$	$[0, 11.54]$	$[2.8, 3.34]$
内密度	$M(C_2^0)$	3.545	0	3.07
偏差	$MM(C_2)$	0.215	11.54	0.27
距离约束条件	① $d(C_1, C_2) \in T(C_1)$	√	×	√
	② $d(C_1, C_2) \in T(C_2)$	√	√	√
	是否满足 (①and②)	√	×	√
内密度约束条件	$M(C_1^0) \in T(C_2)$	×	√	×
	$M(C_2^0) \in T(C_1)$	√	√	√
	是否满足 (①and②)	√	√	√
偏差约束条件	① $MM(C_1 \cup C_2)$	0.21	7.78	0.555
	② $\text{Max}(MM(C_1), MM(C_2))$	3.76	11.54	3.76
	是否满足 (①≤②)	√	√	√
是否为有效邻近类		√	×	√
合并结果		合并	不合并	不合并

基于邻近图的点群聚类方法——具体步骤及举例

• 生成的新顺序表

类	内边	外边	内密度	偏差	置信区间	邻近类	类间距离
$c(E)$	-	$e(D,E)$	0	3.78	[0, 3.78]	$c(B,C,D)$	3.78
$c(R)$	-	$e(P,R)$	0	4.59	[0, 4.59]	$c(T,S,P)$	4.59
$c(K)$	-	$e(Q,K)$	0	4.9	[0, 4.9]	$c(J,Q,O)$	4.9
$c(A)$	-	$e(A,B)$	0	5.34	[0, 5.34]	$c(B,C,D)$	5.34
$c(U)$	-	$\frac{e(S,U)}{e(U,V)}$	0	5.525	[0, 5.525]	$\frac{c(T,S,P)}{c(V)}$	4.95
							5.56
$c(V)$	-	$e(U,V)$	0	5.56	[0, 5.56]	$c(U)$	5.56
$c(G)$	-	$e(F,G)$	0	11.54	[0, 11.54]	$c(F)$	11.54
$c(J,Q,O)$	$e(J,Q)$	$\frac{e(I,J)}{e(Q,K)}$	2.86	1.04	[1.82, 3.9]	$\frac{c(H,I)}{c(K)}$	4.83
		$e(O,P)$					4.9
						$c(T,S,P)$	3.9
$c(F,H,I)$	$\frac{e(F,H)}{e(H,I)}$	$\frac{e(D,F)}{e(F,G)}$	3.205	0.555	[2.65, 3.76]	$\frac{c(B,C,D)}{c(G)}$	3.76
		$e(I,J)$					11.54
						$c(J,Q,O)$	4.83
$c(B,C,D)$	$e(B,C)$	$\frac{e(A,B)}{e(D,E)}$	3.545	0.215	[3.33, 3.76]	$\frac{c(A)}{c(E)}$	5.34
		$e(D,F)$					3.78
						$c(F)$	3.76
$c(T,S,P)$	$e(P,S)$	$\frac{e(O,P)}{e(P,R)}$	3.65	0.55	[3.1, 4.2]	$\frac{c(J,Q,O)}{c(R)}$	3.9
		$e(S,U)$					4.59
						$c(U)$	4.95
$c(F,H,I)$	$\frac{e(F,H)}{e(H,I)}$	$\frac{e(D,F)}{e(F,G)}$	3.205	0.555	[2.65, 3.76]	$\frac{c(B,C,D)}{c(G)}$	3.76
		$e(I,J)$					11.54
						$c(J,Q,O)$	4.83

基于邻近图的点群聚类方法——具体步骤及举例

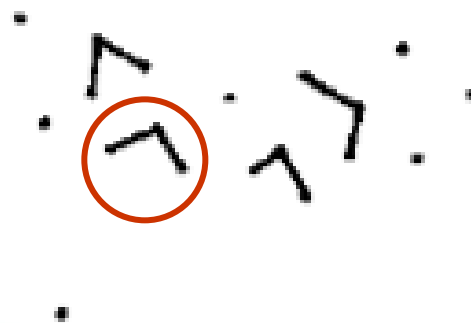
(5) 结束条件判断

判断顺序表是否为空，为空则结束聚类，否则重复步骤(4)。

e. g. 依次考察后均未出现有效邻近类，判断基于MST的聚类结束。



初始类



NNG-MST聚类结果

基于邻近图的点群聚类方法——具体步骤及举例

(6) 继续聚类

可停止聚类，或继续激发更高层次的邻近图，从步骤(1)重新开始，进一步得到新的分类结果。

e. g. 顺序激活UG, GG, DT的边继续聚类

邻近图层次越高，分类数目越小



NNG-MST聚类结果

MST-UG聚类结果

UG-GG聚类结果

GG-DTG聚类结果

基于邻近图的点群聚类方法——方法特点分析

- 特点

采用不同的邻近图组合可以产生不同的聚类结果

e. g.

①将每个点看作一类，先在NNG上聚类，然后基于该聚类结果激活MST进行聚类(上述例子)

②将初始点群生成的NNG中有相邻的点作为初始类，然后基于该聚类结果激活MST进行聚类



结果对比

基于邻近图的点群聚类方法——方法特点分析

- 优点

- 是一种无监督的空间聚类方法
- 能适应于任意形状的空间聚类，顾及群体的全局分布特性
- 且可以快速获得各种层次的高质量聚类

- 不足

- 无法结合有地理位置和属性双重特征进行聚类
- 没有考虑点群同其他地理空间要素(e. g. 河流，街道)的空间关系
- 缺乏评价聚类结果的真实性和有效性的最优手段

3

线群的空间特征

线群的空间结构——基本概念

• 【线群】：两条或多条相互关联的线的集合

• 两条线间的主要的空间特征：

- 最短距离
- 最长距离
- 平均距离
- 协调程度
- 中轴线
- 拓扑关系
- 方向关系

• 多条线间的主要的空间特征：

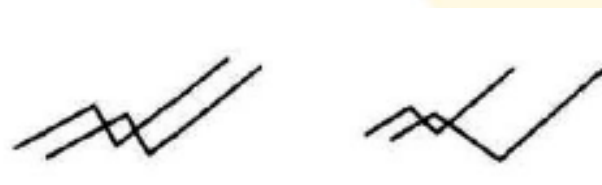
- 密度分布
- 分布范围
- 聚类特征
- 总体走向
- 排列次序
- 分形特征
- 连通情况
- 网络结构
- 拓扑关系
- 层次结构

线群的空间结构——两线间的协调程度

- 两线间的协同程度是指两线间的平行程度，但是通过平移得到的线之间在空间认知上不一定“协调”。



平行程度：小  大



平行但不协调

- 此处，介绍两种协同程度判断的方法

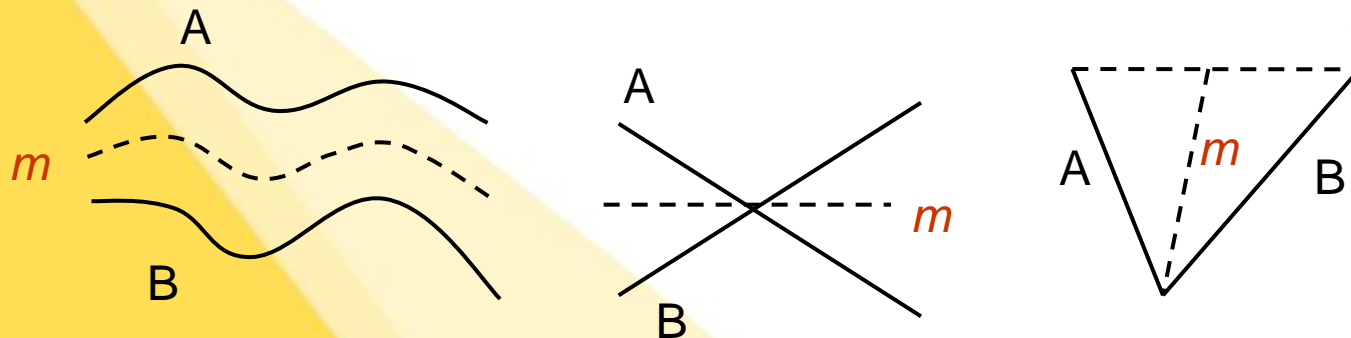
➤ 方法一 提取中心线，进行判断

➤ 方法二 以其中一条线为参考，建立一个能恰好与另一条线的两侧相切的缓冲带，以缓冲带内的中心线作为标准线，并根据一公式计算协调程度

线群的空间结构——两线间的协调程度

方法一 用中心线可以粗略的描述两条没有相交线的协调程度

- 【中心线】：位于两条线之间的分别与两条线的最短距离相等的点集，两条线之间的中心线也常用“中轴线(Medial axis)”描述



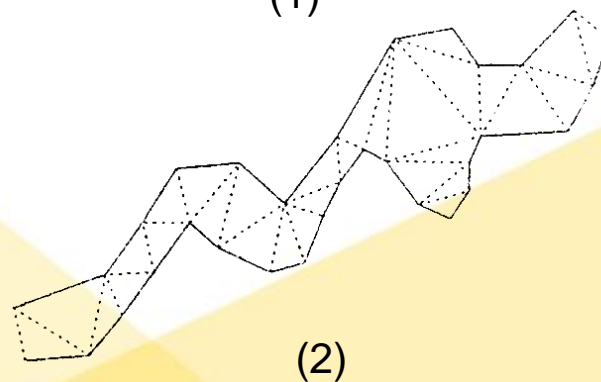
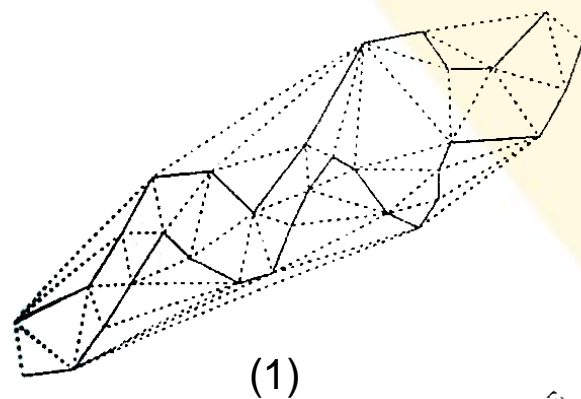
不同形态的两条线的间的中心
线

线群的空间结构——两线间的协调程度

方法一

➤ 步骤1：提取中心线。不相交的、转角不尖锐的两条线之间中心线一般情况下的近似算法步骤如下：

- (1) 建立Delaunay约束三角网, 两条线的点的密度应当通过插值的方法尽量保证接近, 同时, 三角形的内角尽量保证小于 90°
- (2) 选择顶点分属两条不同的线的三角线集合, 以及与其相接, 但没有越过约束线的三角形集合



线群的空间结构——两线间的协调程度

方法一

- (3) 按照规则建立中心线网络(即提取中轴线)

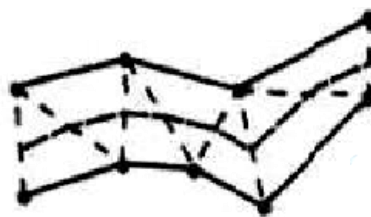
➤常用的提取规则



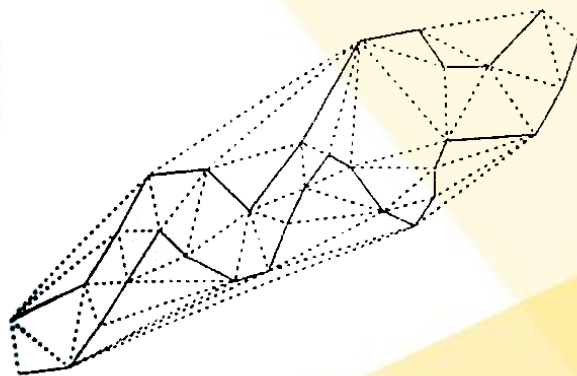
①
DT三角形的
Voronoi图的边界



②
DT三角形的
重心连线



③
DT三角形边的
中点连线

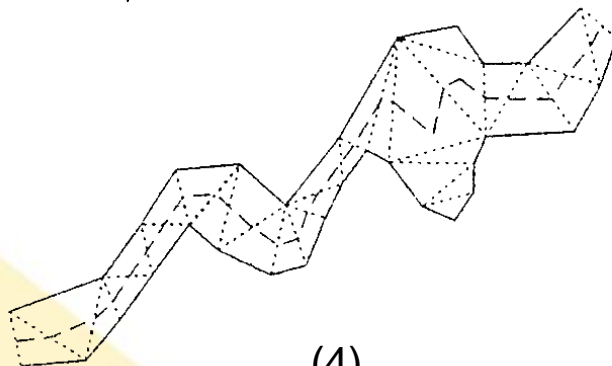


(3)

线群的空间结构——两线间的协调程度

方法一

- (4) 按照"中心线只是位于两条线之间"原则,删除中轴线的分支,并且适当把分叉点(依据提取规则的不同,可能为三角形的外心,中心,或重心)向分叉方向移位,以得到视觉上的平衡。



(4)

➤ 步骤2: 进行判断。

- 当这两条线局部的中心线相对中轴线没有偏移时,这个局部范围在视觉上就显得协调一些
- 反之,中心线相对中轴线有偏移的地方就很不协调。

线群的空间结构——两线间的协调程度

- 方法一的不足：

① 若中轴线没有分叉,即中心线没有相对中轴线有偏移时,还不能充分说明这两条线就是完全协调的,例如并不平行的两个直线段;

② 若中心线相对中轴线有偏移时,两条线肯定是不完全协调的。中轴线毛刺的长短表明了不协调的程度,但还没有确切的量化方法,不能从视觉上简单的判断;

③ 若两条线相距较远,则很难由中心线描述它们的相对协调程度。

线群的空间结构——两线间的协调程度

方法二 建立协调程度计算公式，量化评价两线间协调程度

•适用条件

- ① 被考察的两条线比较平滑，没有尖锐的转折角，因为若一条线有较尖锐的转角，即使是与这条线平行的线，在视觉上仍然是不协调的；
- ② 两条线不相交或相接、若在局部存在相交或相接，则认为这两条线在此处完全不协调；
- ③ 线与线之间应当相距较近。若两条线相离很远，则没必要单独讨论协调程度。

线群的空间结构——两线间的协调程度

方法二 建立协调程度计算公式，量化评价两线间协调程度

- 地理空间中存在大量不完全协调的两条线，这就要求对其协调程度定量计算



不完全协调的两条线

•方法适用条件

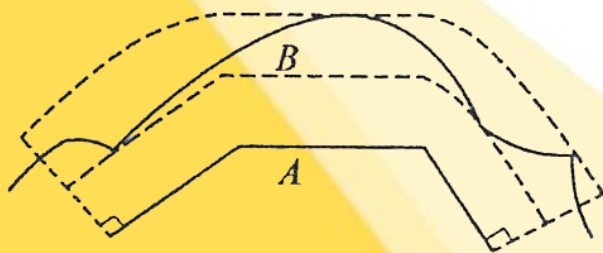
- ① 被考察的两条线比较平滑，没有尖锐的转折角，因为若一条线有较尖锐的转角，即使是与这条线平行的线，在视觉上仍然是不协调的；
- ② 两条线不相交或相接、若在局部存在相交或相接，则认为这两条线在此处完全不协调；
- ③ 线与线之间应当相距较近。若两条线相离很远，则没必要单独讨论协调程度。

线群的空间结构——两线间的协调程度

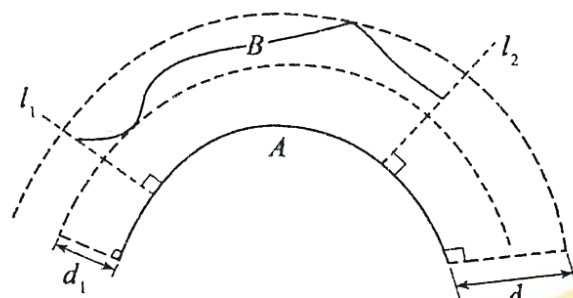
方法二 建立协调程度计算公式，量化评价两线间协调程度

• 计算公式的参数

设有两条线A, B, 以A为参考, 找到A的一个在B一侧的缓冲带, 该缓冲带刚好从两侧与B相切。缓冲带的中心线C是A的一条平行线



➤若B不完全在缓冲带内, 则认为B只是局部与A存在协调关系;



➤若B只占缓冲带的一部分, 则从两端压缩缓冲带的范围并缩短C的长度

线群的空间结构——两线间的协调程度

方法二 建立协调程度计算公式，量化评价两线间协调程度

• 计算公式的参数

① d_1, d_2

d_1 : 缓冲带的内半径

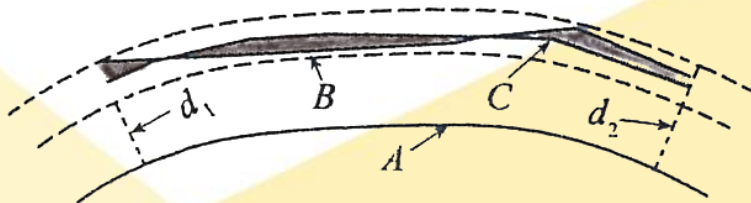
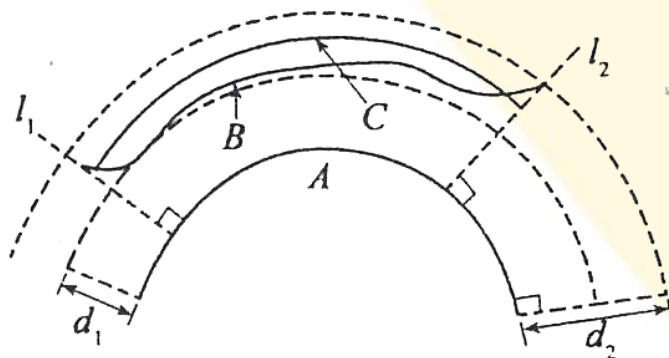
d_2 : 缓冲带的外半径

② L_C

缓冲带内标准线C的长度

③ S_i

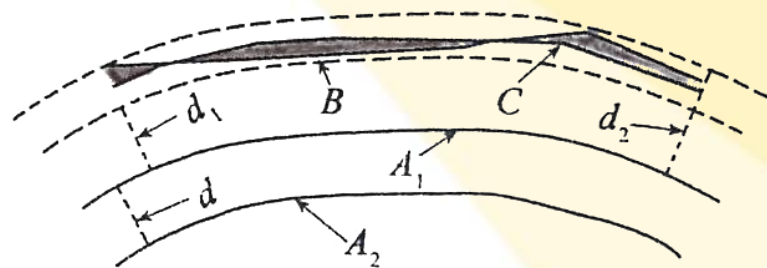
将B与C在它们的交点处分段，B由为B的第*i*段和C的第*i*段所构成的多边形面积（阴影部分），表示B与C的偏差。



线群的空间结构——两线间的协调程度

方法二 建立协调程度计算公式，量化评价两线间协调程度

- 计算公式的参数与协调程度 μ 的关系



当 $\mu=0$ 时, 两条线最协调, μ 值越大两条线之间的协调程度越差

- ① 若 $d_1=d_2$, B就是A的平行线, A与B之间完全协调。 d_2 与 d_1 的差值越大, 则B相对A的协调程度越差, 即 μ 值越大。
- ② B和C相对应的区段所构成的多边形面积之和S越大, 则B偏离C的程度越大, A与B之间的协调程度越差, 即 μ 值越大
- ③ d_1 越小, 同样的两条线在认知上的协调程度将越差, 即 μ 值越大。
- ④ 在参数相同的情况下, L_C 的值越小则两条线之间的协调程度越差, 即 μ 值越大

线群的空间结构——两线间的协调程度

方法二 建立协调程度计算公式，量化评价两线间协调程度

- μ 的计算公式

根据上述关系，可建立计算B相对A的协调程度的公式

$$\mu = \frac{d_2 - d_1}{d_1} \times \frac{\sqrt{\sum_{i=1}^m S_i}}{L_C}$$

- 方法二特点

由于协调程度是相对的，基于几个主要的相关影响因素所建的公式也只是粗略计算两条线之间的协调程度，而且以不同的线作为参考目标计算出的结果是有差异的，所以要注意不同情况下参考线的选择。

线群的空间结构——两线间的协调程度

方法二 建立协调程度计算公式，量化评价两线间协调程度

- e. g. A_2 是 A_1 一条平行线,以这两条线为参考对 B 所建的缓冲带相同,缓冲带中心线为 C 。分别求算 B 对 A_1 、 A_2 的协调程度 μ_1 , μ_2 。已知:
 - A_2 、 A_1 间距离 $d=30$, A_1 、 C 间距离 $d_1=30$, 缓冲带宽度 $d'=20$;
 - 缓冲带内中心线长度 $L_C=264$;
 - B 与 C 在交点处分成三段,相交形成面积分别为 $S_1=112$, $S_2=746$, $S_3=479$

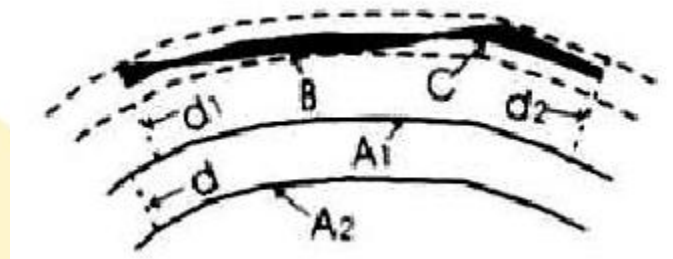
解: 偏差总面积 $S=S_1+S_2+S_3=112+746+479=1337$

对 A_1 来说, 缓冲带内半径 $d_{1内}=d_1=30$, 外半径 $d_{1外}=d_1+d'=50$

对 A_2 来说, 缓冲带内半径 $d_{2内}=d_1+d'=60$, 外半径 $d_{2外}=d_1+d'+d=80$

$$\mu_1 = \frac{50-30}{30} \times \frac{\sqrt{1337}}{264} = 0.092$$

$$\mu_2 = \frac{80-60}{60} \times \frac{\sqrt{1337}}{264} = 0.046$$



$\mu_2 < \mu_1$, B相对 A_2 比B相对 A_1 在视觉上会显得更协调些

线群的空间结构——线群的网络特征

地理学研究中常需要对可抽象为点状要素的实体（如，居民点、城市、矿产分布点、交通站点等）之间的相互联系（如，水系、构造带、交通线、通信线路等）进行分析

- 这就需要分析线群的网络特征来对上述联系进行分析
- 利用图论中图的矩阵表达可以在计算机中对线群的网络特征进行描述

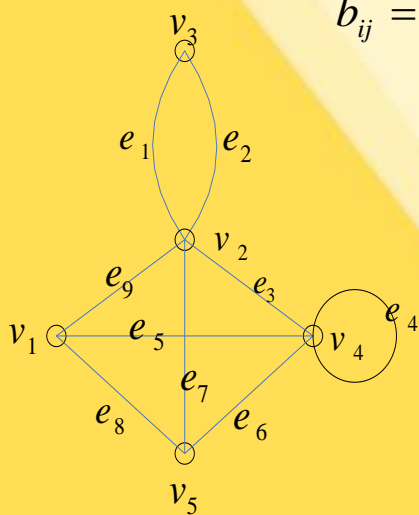
➤主要包括：关联矩阵与邻接矩阵

线群的空间结构——线群的网络特征

1、关联矩阵——（1）无向图的关联矩阵

对于图 $G=(V,E)$, $|V|=n$, $|E|=m$, 有 $m \times n$ 阶矩阵 $B=(b_{ij})_{m \times n}$, 其中, b_{ij} 表示顶点 v_i 与边 e_j 关联的次数

$$b_{ij} = \begin{cases} 2 & \text{当且仅当 } v_i \text{ 是边 } e_j \text{ 的两个端点, } e_j \text{ 是自环} \\ 1 & \text{当且仅当 } v_i \text{ 是边 } e_j \text{ 的一个端点, } e_j \text{ 不是自环} \\ 0 & v_i \text{ 与 } e_j \text{ 不关联} \end{cases}$$



对应的关联矩阵

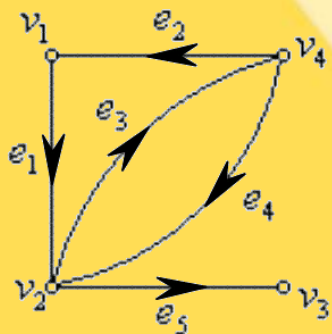
$$B(G) = \begin{matrix} & \begin{bmatrix} e_1 & e_2 & e_3 & e_4 & e_5 & e_6 & e_7 & e_8 & e_9 \end{bmatrix} \\ \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \end{bmatrix} & \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 2 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix} \end{matrix}$$

线群的空间结构——线群的网络特征

1、关联矩阵——（1）有向图的关联矩阵

对于图 $D=\langle V,E\rangle$, $|V|=n$, $|E|=m$, 有 $m\times n$ 阶矩阵 $M=(m_{ij})_{m\times n}$, 其中, m_{ij} 表示顶点 v_i 与边 e_j 关联的次数

$$m_{ij} = \begin{cases} -2 & \text{当且仅当 } v_i \text{ 是边 } e_j \text{ 的两个端点, } e_j \text{ 是自环} \\ 1 & \text{当且仅当 } v_i \text{ 是边 } e_j \text{ 的起点, } e_j \text{ 不是自环} \\ -1 & \text{当且仅当 } v_i \text{ 是边 } e_j \text{ 的终点, } e_j \text{ 不是自环} \\ 0 & v_i \text{ 与 } e_j \text{ 不关联} \end{cases}$$



对应的关联矩阵

$$M(D) = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ -1 & 0 & 1 & -1 & 1 \\ 0 & 0 & 0 & 0 & -1 \\ 0 & 1 & -1 & 1 & 0 \end{bmatrix}$$

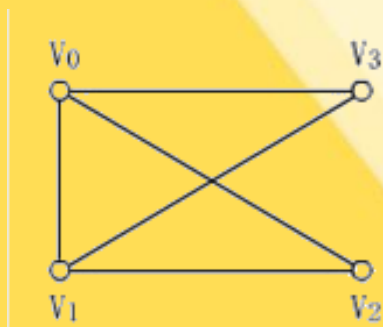
线群的空间结构——线群的网络特征

2、邻接矩阵——（1）无向图的邻接矩阵

对于图 $G=(V,E)$, $|V|=n$, $|E|=m$, 有 $m \times n$ 阶矩阵 $A=(a_{ij})_{m \times n}$, 其中, a_{ij} 表示顶点 v_i 与 v_j 之间的边数

$$a_{ij} = \begin{cases} 1, & (v_i, v_j) \in E \\ 0, & \text{其他} \end{cases}$$

A是一个对称矩阵



对应的邻接矩阵

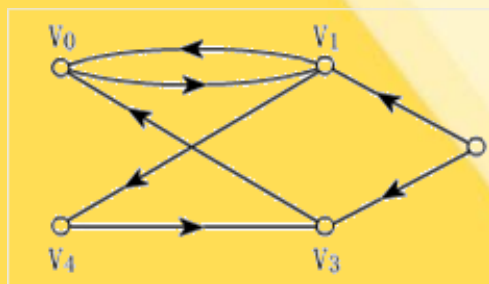
$$A(G) = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix}$$

线群的空间结构——线群的网络特征

2、邻接矩阵——（2）有向图的邻接矩阵

对于图 $D=\langle V,E\rangle$, $|V|=n$, $|E|=m$, 有 $m\times n$ 阶矩阵 $A=(a_{ij})_{m\times n}$, 其中, a_{ij} 表
顶点 v_i 与 v_j 之间的边数

$$a_{ij} = \begin{cases} 1, & (v_i, v_j) \in E \\ 0, & \text{其他} \end{cases}$$



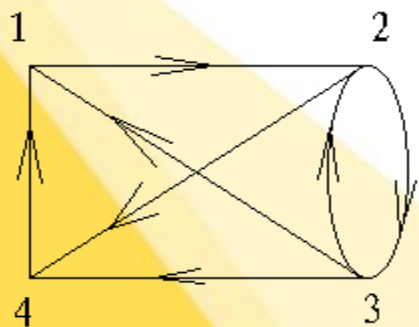
对应的关联矩阵

$$A(D) = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

线群的空间结构——线群的网络特征

2、邻接矩阵——关于邻接矩阵的幂

定理: 设 G 是具有 n 个结点集 $\{v_1, v_2, \dots, v_n\}$ 的图, 其邻接矩阵为 A , 则 A^k ($k=1, 2, \dots$) 的 (i, j) 项元素 $a^{(k)}_{ij}$ 是从 v_i 到 v_j 的长度等于 k 的路的总数。



$$A^2 = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 2 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

$$A^3 = \begin{bmatrix} 2 & 1 & 0 & 1 \\ 1 & 2 & 1 & 1 \\ 2 & 2 & 1 & 2 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

A^2 表示 i 和 j 之间具有长度为2的通路数

A^3 表示 i 和 j 之间具有长度为3的通路数

.....

线群的空间结构——线群的网络特征

3、几个网络特征测度指标

(1) 平均路径长度

任意两个节点之间的距离的平均值

$$L = \frac{\sum_{i \geq j} d_{ij}}{\frac{1}{2} N(N+1)}$$

d_{ij} 为连接*i*与*j*两点的最短路径上的边数
 N 为网络节点数

(2) 聚类系数

若网络中的一个节点*i*有 k_i 条边将它和其他 k_i 个节点相连,则在这 k_i 个节点之间最多可能有 $k_i(k_i-1)/2$ 条边。节点*i*的聚类系数为 C_i ,整个网络的聚类系数 C 就是所有节点*i*的聚类系数 C_i 的平均值。

$$C_i = \frac{2E_i}{k_i(k_i-1)}$$

E_i 为 k_i 个节点之间实际存在的边数

线群的空间结构——线群的网络特征

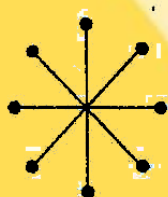
3、几个网络特征测度指标

(3) β 指数（点线率）——网络连结性指标

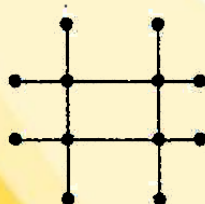
网络内每个节点的平均连线数目

$$\beta = \frac{\sum_{i=1}^N E_i}{\sum_{i=1}^N V_i}$$

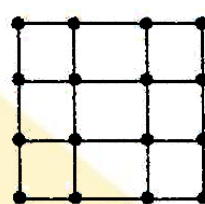
N 为网络节点数



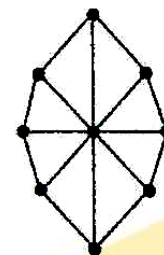
$\beta=0.89$



$\beta=1$



$\beta=1.5$



$\beta=1.78$

β 值越高，线群网络越复杂

线群的空间结构——线群的网络特征

3、几个网络特征测度指标

(4) γ 指数——网络连通性指标

网络边线的实际数目与边线的最大限度数目的比率

$$\gamma = \frac{\sum_{i=1}^N E_i}{3 \left(\sum_{i=1}^N V_i - 2N \right)} \quad N \text{ 为网络节点数}$$

$\gamma \in [0, 1]$, γ 值越高, 网络连通性越强

The background features a white central area surrounded by three large triangular shapes. Two of these triangles, located in the top-right and bottom-left corners, are composed of two overlapping layers: a light yellow layer on top and a darker yellow layer on the bottom. The third triangle, in the bottom-right corner, is a single solid darker yellow layer. The text "THE END" is centered in the white area.

THE END