第5章 数字矿山空间分析原理

主讲人：毕林

2024年8月13日

目录

- 5.1 空间关系分析
- 5.2 空间网络分析
- 5.3 空间统计学分析
- 5.4 探索性空间数据分析
- 5.5 空间三维分析

5.1 空间关系分析

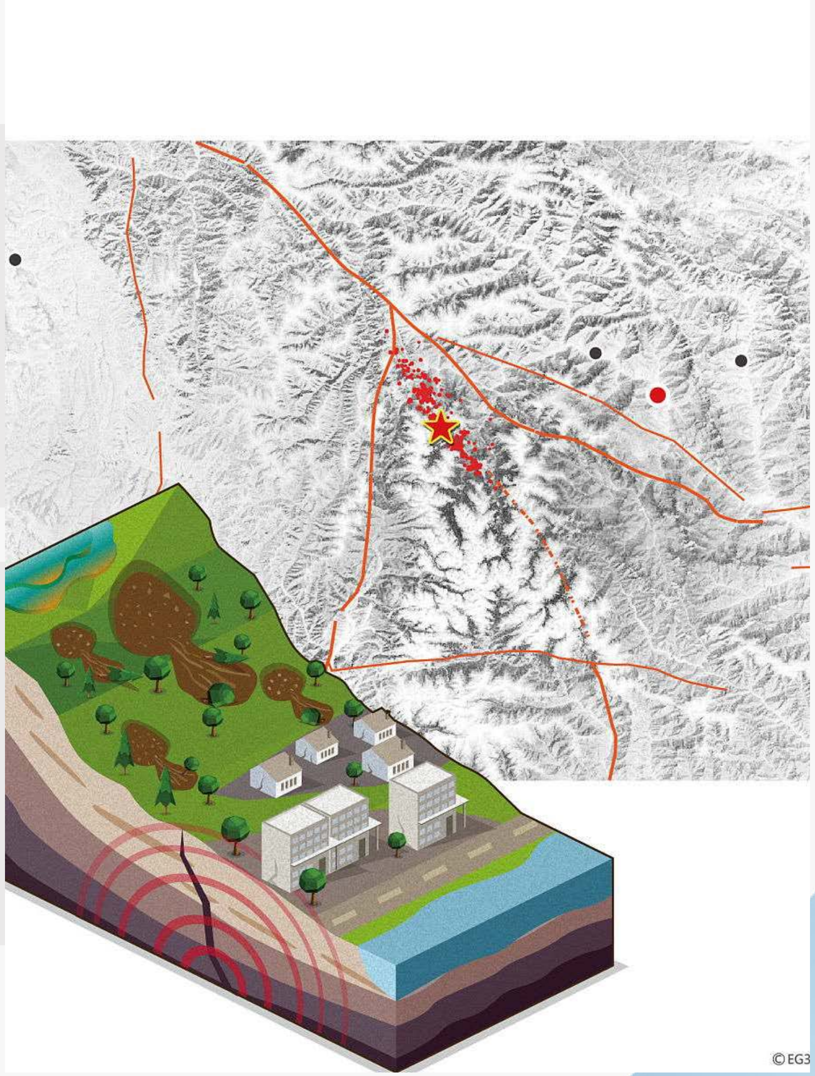
空间关系分析的重要性

空间关系分析的核心地位

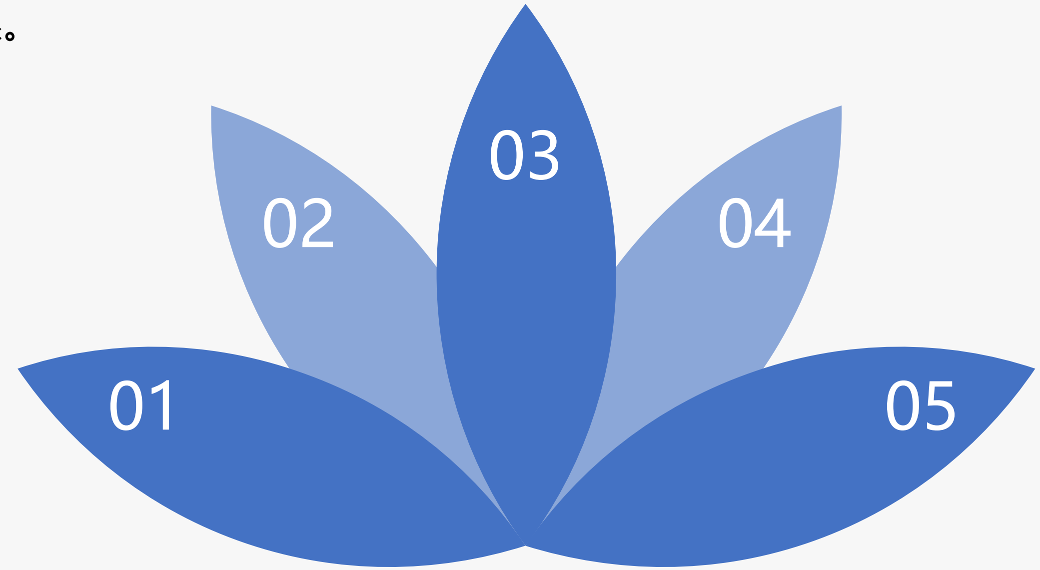
空间关系分析是GIS（地理信息系统）的核心功能之一，主要用于研究空间物体或现象之间的相互关系。

数字矿山应用系统的关键方法

数字矿山应用系统在进行开采规划、方案设计时，常采用空间关系分析方法，确保高效、科学地挖掘。



空间关系分析的类型



拓扑关系分析

拓扑关系分析是空间关系分析中最基本的一种，主要用于研究空间物体之间的连接、相邻、相离等关系
。

方向关系分析

方向关系分析主要研究空间物体之间的相对方向关系，例如两个道路交叉口的相对方向关系。

距离关系分析

距离关系分析主要研究空间物体之间的距离远近，以及不同物体之间的最短路径等问题。

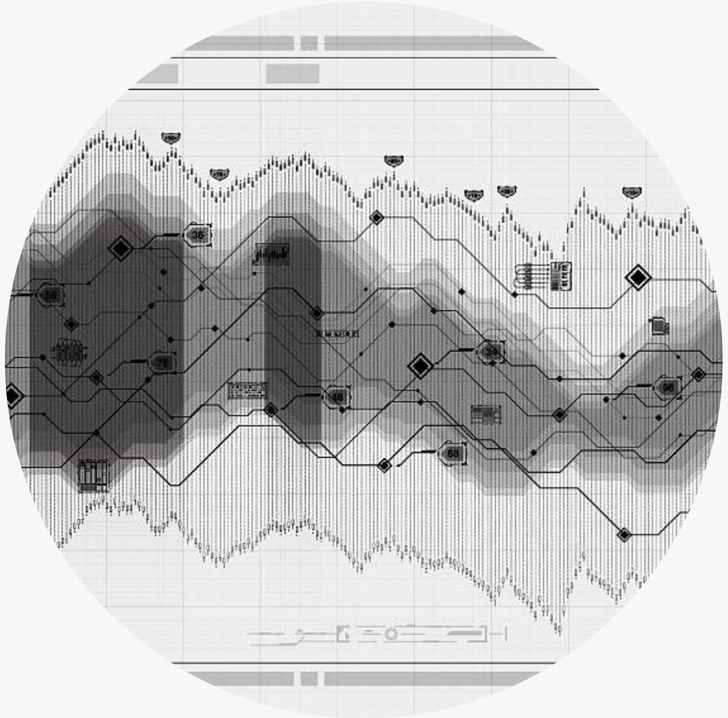
面积关系分析

面积关系分析主要研究空间物体所占面积的大小及其相互关系。

高程关系分析

高程关系分析主要研究空间物体之间的高程差异及其相互关系。

数字矿山的空间分析



数字矿山空间分析的技术

除了常见的空间关系分析类型外，还有光照关系分析、地形关系分析等多种空间关系分析方法。

数字矿山空间分析的作用

数字矿山的空间分析是从矿山地理信息目标之间的空间关系中获取派生信息和新知识的分析技术。

空间关系的复杂性

空间关系复杂多样，与地理位置、空间分布和对象属性等多方面因素有关。

空间几何关系分析

空间几何关系分析主要包括邻近度分析、叠加分析等，是数字矿山空间分析的重要手段。

5.1.1 邻近度分析

邻近度

邻近度是定性描述空间目标距离关系的重要物理量，表示地理空间中两个空间目标距离相近的程度。

距离关系分析

以距离关系为分析基础的邻近度分析是空间几何关系分析的重要手段，用于确定道路宽度和安全带。

基础设施位置选择

基础设施如工业广场、选矿厂、堆场、尾矿库、排土场等的位置选择需考虑服务范围。

邻近度分析技术发展

随着邻近度分析技术的发展以及应用范围的扩展，距离关系由欧几里得距离发展到曼哈顿距离等。

解决邻近度问题方法

解决邻近度问题的方法很多，比较成熟的分析方法有缓冲区分析、泰森多边形分析等。

1） 缓冲区分析

缓冲区分析定义

缓冲分析是指在点、线、面实体周围建立一定宽度范围的多边形，根据指定的距离划定出警戒范围。

缓冲区分析基本思想

数学角度看，缓冲区分析的基本思想是给定一个空间对象或集合，确定其邻域，邻域的大小由邻域半径R 决定。

缓冲区分析应用

为不同工作需要提供科学依据；缓冲分析还可以与其他空间分析方法结合使用。

1） 缓冲区分析

01

缓冲区数据层

缓冲区作为一个独立的数据层可以参与叠加分析，常应用到道路、河流、居民点、污染源、爆破震动等生产生活设施的空间分析，为不同工作需要提供科学依据。

02

缓冲区分析研究

结合不同的专业模型，缓冲区分析能够在景观生态、规划、军事应用等领域发挥更大的作用，例如在矿区复垦规划中利用缓冲区分析和景观结构变异系数方法进行分割研究。

03

缓冲分析其他方法

缓冲分析还可以与其他空间分析方法结合使用，以解决更复杂的实际问题，例如与网络分析结合确定最优资源分配方案，与多边形叠加分析结合提取感兴趣区域内的空间信息。

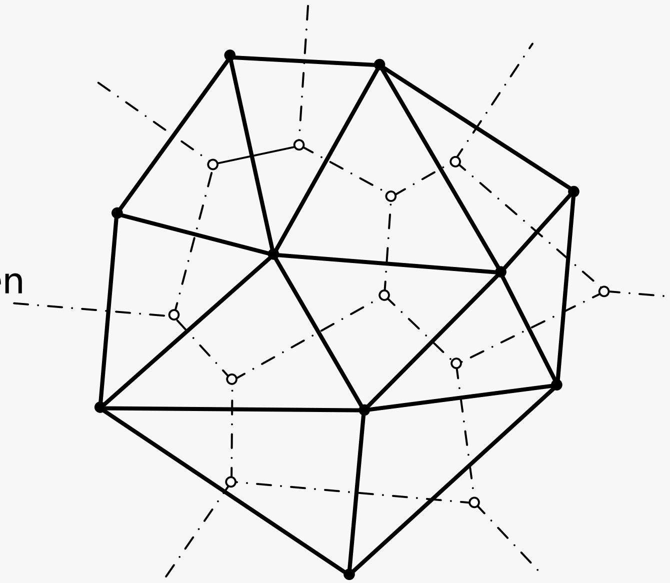
2） 泰森多边形分析

泰森多边形的定义

泰森多边形是基于一组点集的平面划分，每个点都在其周围区域内具有最小的邻近距离。

泰森多边形的几何定义

泰森多边形的几何定义为离散点集P中任意两点不同位，任意四点不共圆，则任意离散点P i 的泰森多边形的定义为。



泰森多边形的特性

泰森多边形也称为Thiessen Polygons或Thiessen Tesselations，是中虚线所围成的多边形。

泰森多边形的形成过程

任意离散点P i 的泰森多边形是一个凸多边形，且在特殊的情况下可以是一个具有无限边界的凸多边形。

泰森多边形的作用

泰森多边形是在计算几何中被广泛研究的一个问题，用于邻域分析，也称为最近邻点法。

泰森多边形的特点

泰森多边形是实现对一个平面的划分，在泰森多边形T i 中，任意一个内点到该泰森多边形的发生点P i 的距离都小于该点到其他任何发生点P j 的距离。

2）泰森多边形分析

泰森多边形内插

泰森多边形是一种空间内插方式，可以用离散点的性质来描述多边形区域的性质，并计算区域内的未知数据。

泰森多边形应用

泰森多边形适用于根据离散点影响力划分空间范围的情况，以及在缺少连续数据时作近似替代。

快速赋值方法

GIS和地理分析中经常采用泰森多边形进行快速赋值，其隐含假设是任何地点的未知数据均使用距它最近的采样点数据。

采样点数量要求

除非有足够多的采样点，否则该假设不恰当。因为降水、气压、温度等现象是连续变化的，泰森多边形插值方法得到的结果在边界上是突变的。

岩粉取样属性

露天矿炮孔岩粉取样属性代替以炮孔作为质心的泰森多边形的所以点的属性，也是常用方法。

泰森多边形不完善

泰森多边形分析的不完善之处在于，它假设多边形区域内所有点的属性与其质心点的属性相同，这可能导致结果的不准确。

5.1.2 叠加分析

叠加分析的定义

01

叠加分析是在统一空间参考系统下，通过对两个数据进行集合运算，产生新数据的过程。

叠加分析的目标

02

叠加分析的目标是分析空间对象之间的相互关系，对多种现象要素进行综合分析和评价。

叠加分析的应用

03

叠加分析可以揭示各种现象要素的内在联系及其发展规律，对地理信息进行各自的叠加处理。

叠加分析的类型

04

矢量数据模型与栅格数据模型的属性叠加处理分为代数运算与逻辑运算两大类。

矢量数据模型的叠加分析

05

矢量数据模型以点、线、面等简单几何对象来表示空间要素，这些数据直接的叠加分析。

栅格数据模型的叠加运算

06

栅格数据模型的叠加运算常被称为地图代数，应用非常广泛。

1）矢量数据叠加分析

点与多边形的叠加

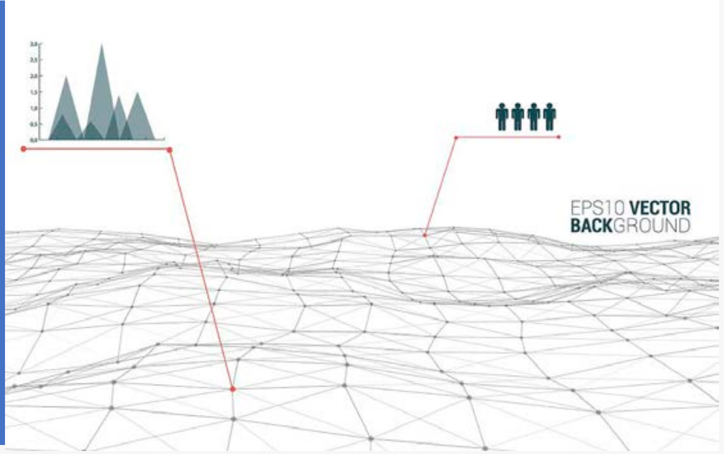
将一个点层作为输入图层叠加到一个多边形图层上，生成的新图层仍然是点层。

属性和标识

通过属性查询能够知道每个监测井是属于哪个四级区，还可查询特定的四级区内包含有哪些水质监测井等信息。

叠加分析后的图层

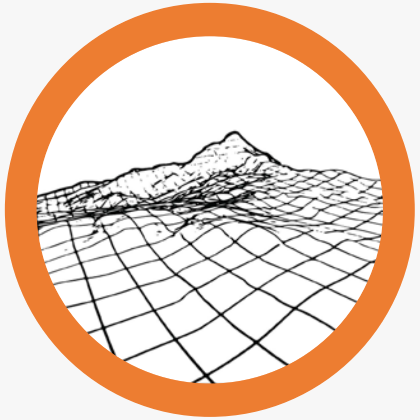
叠加分析后的图层通常会生成一个新的属性表，该属性表不仅保留了原图层的属性。



1）矢量数据叠加分析

信息关联

水资源四级区的属性表中还有属于哪个省区、面积大小等信息，监测井的属性表也可以与这些属性关联起来。



查询长度

叠加分析操作后既可确定每条线段落在哪个多边形内，也可查询指定多边形内指定线段穿过的长度。

线与多边形的叠加

将一个线图层作为输入图层叠加到一个多边形图层上，要进行线段与多边形的空间关系判别。

1）矢量数据叠加分析

交点生成

一个线目标往往跨越多个多边形，这时需要计算线与多边形的交点，只要相交就会生成一个结点。

线状数据层

叠加分析的结果产生了一个新的线状数据层，该层内的线状目标属性表发生了变化。

查询指定线段

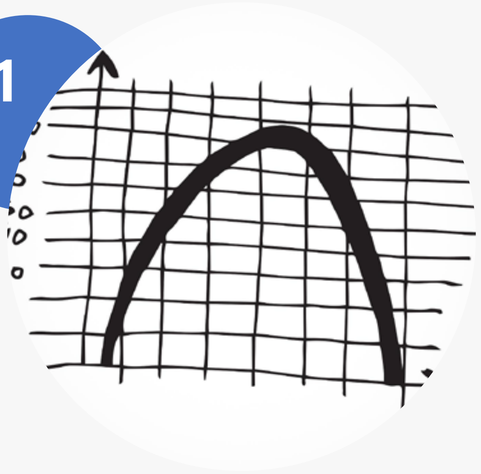
叠加分析操作后可查询任意省区内河流的长度，计算河网密度；若线层是道路层，则可计算每个多边形内的道路总长度。

多边形与多边形的叠加

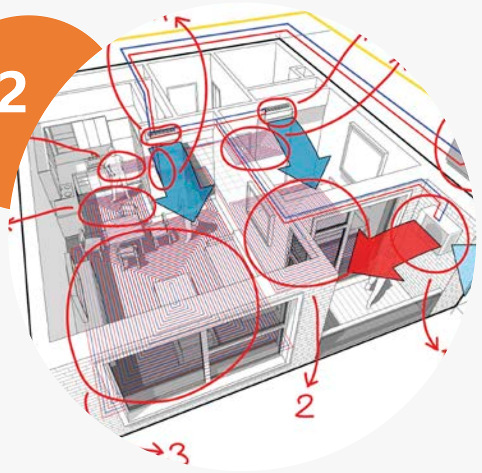
多边形与多边形的叠加要比前两种叠加复杂得多，需要进行复杂的空间关系判别和属性信息叠加。

2）栅格数据叠加分析

01



02



03

栅格数据叠加分析

栅格数据叠加分析是一种将多个栅格数据集合并计算的方法，常用于地理信息系统。

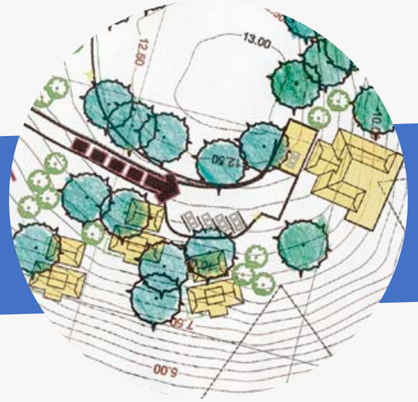
栅格数据结构

栅格数据结构简单，各种要素都可用规则格网和相应属性来表示，且精度不受多层叠加影响。

栅格叠加应用

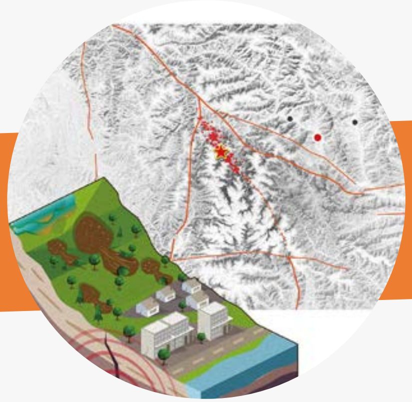
栅格叠加可用于数量统计、益本分析、类型叠加、动态变化分析以及几何提取等应用。

2）栅格数据叠加分析



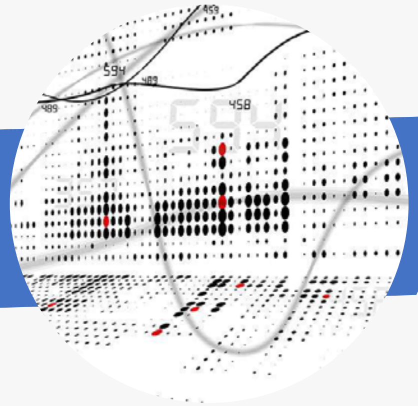
栅格数据运算

栅格数据的叠加分析操作通过栅格之间的各种运算实现，包括加、减、乘、除等数学运算。



数学关系式建立

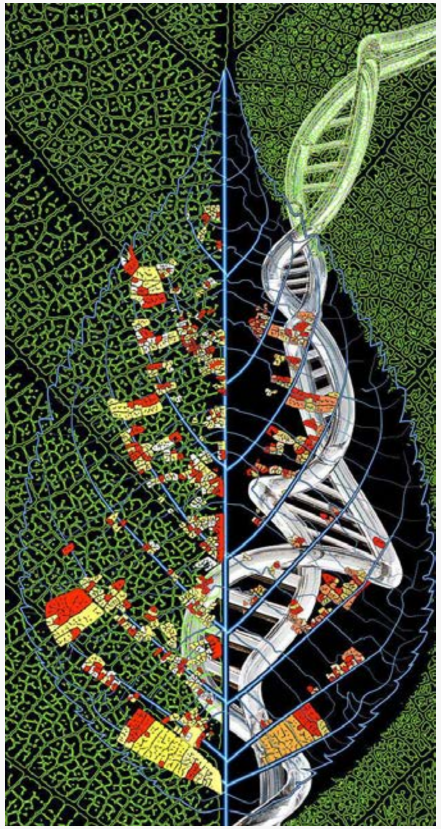
数学关系式可建立多个数据层之间的关系模型，反映各层属性与用户需求之间的联系。

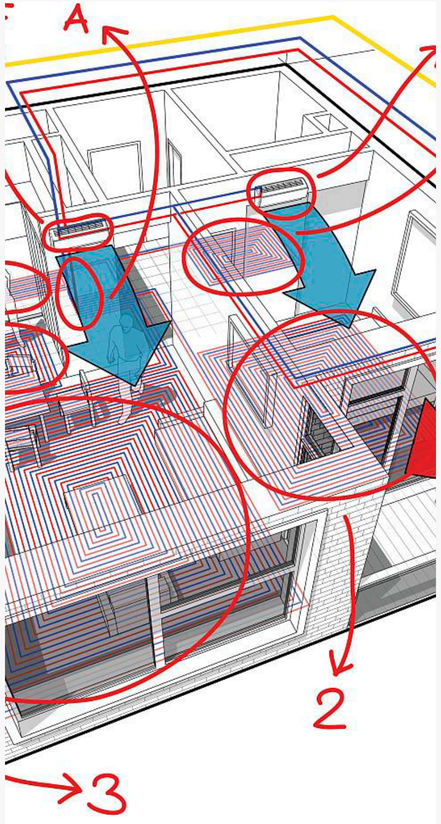


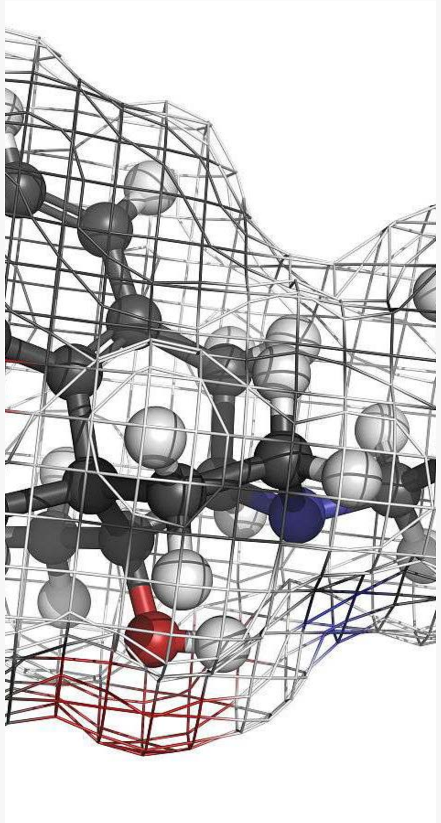
叠加后属性值

叠加操作的输出结果可能是算术运算结果，或者是各层属性数据的最大值或最小值、平均值。

2）栅格数据叠加分析







逻辑运算结果

叠加操作的结果可以通过对各层具有相同属性值的格网进行运算得到，或者通过欧几里得几何距离的运算以及滤波运算等得到。

地图代数定义

基于数学运算的数据层间叠加运算在地理信息系统中称为地图代数，其形式和概念简单，使用方便灵活。

5.2 空间网络分析

空间网络分析的定义

空间网络分析是一种基于网络模型和空间数据分析的方法，用于研究空间对象之间的连接和关系。

网络分析主要用来解决两大类问题

一类优化路径的求解、连通分量求解等问题；一类是研究资源在网络系统中的分配与流动。

5.2.1 最短路径分析

01

最短路径分析：最短路径分析是图论中的一个重要问题，旨在找到图中两个顶点之间的最短路径。

02

应用领域：最短路径分析被广泛应用于交通路线优化、通信网络设计、供应链管理等领域。

03

路径长度定义：最短路径中的“最短”不仅指一般地理意义上的距离最短，还可以是成本最少、耗费时间最短、资源流量（容量）最大、线路利用率最高等标准。

04

网络相关问题：最可靠路径问题、最大容量路径问题、易达性评价问题和各种路径分配问题均可纳入最佳路径问题的范畴之中。

05

实现方法：无论判断标准和实际问题中的约束条件如何变化，其核心实现方法都是最短路径算法。

1）Dijkstra算法

Dijkstra算法

迪杰斯特拉（Dijkstra）算法是E.W.Dijkstra于1959年提出的一种按路径长度递增的次序产生最短路径的算法。

单源点最短路径

Dijkstra算法是解决单源点间最短路径问题比较经典而且有效的算法，其基本思路是按路径长度递增的次序产生最短路径。

标号法与染色法

Dijkstra算法的基本过程包括初始化、检验未标记点、选取下一个点、找到点的前一点和标记点等步骤，又称标号法或染色法。

Dijkstra算法特点

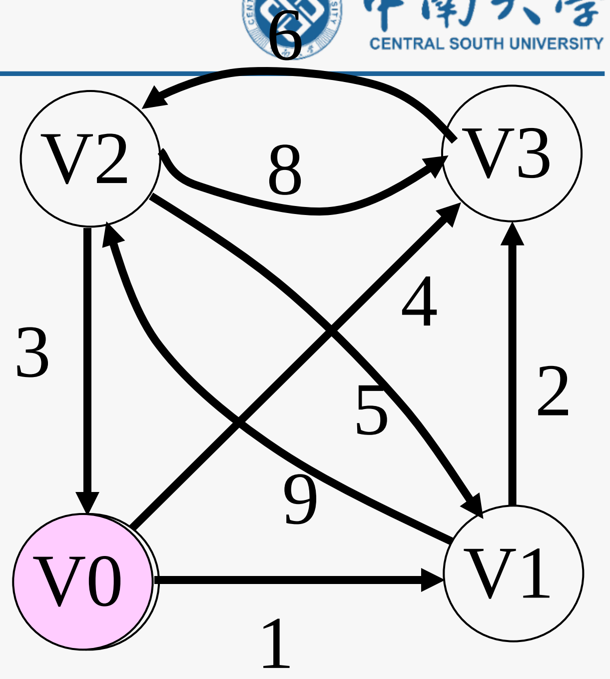
Dijkstra算法从起源点s开始，通过不断选取和标记节点，最终得到从起源点到所有其他顶点的最短路径。

2）弗洛伊德算法

Floyd算法思想：逐个顶点试探法

–求最短路径步骤

- 初始时设置一个n阶方阵，令其对角线元素为0，若存在弧<Vi,Vj>，则对应元素为权值；否则为∝
- 逐步试着在原直接路径中增加中间顶点，若加入中间点后路径变短，则修改之；否则，维持原值
- 所有顶点试探完毕，算法结束



$$D^{(0)} = \begin{array}{cc} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} & \\ \left(\begin{array}{cccc} 0 & 1 & \infty & 4 \\ \infty & 0 & 9 & 2 \\ 3 & 4 & 0 & 7 \\ \infty & \infty & 6 & 0 \end{array}\right) & \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \end{array} \end{array}$$

3）矩阵算法

矩阵算法

矩阵算法是利用矩阵来求出图的最短距离矩阵。假设是带权无向图的邻接矩阵，则。

路径长度计算

表示从结点i经过中间点1到结点j的路径长度，表示从结点i经过中间点2到结点j的路径长度。

最短路径求解

ai,j取它们中的最小值，其意义就是从结点i最多经过一个中间点到结点j的所有路径中长度最短的那条路径。

算法步骤

已知图的邻接矩阵A；求出A, A[2], A[3] ,..., A［n-2］；比较A, A[2], A[3] ,..., A［n-2］，取其中最小的一项。

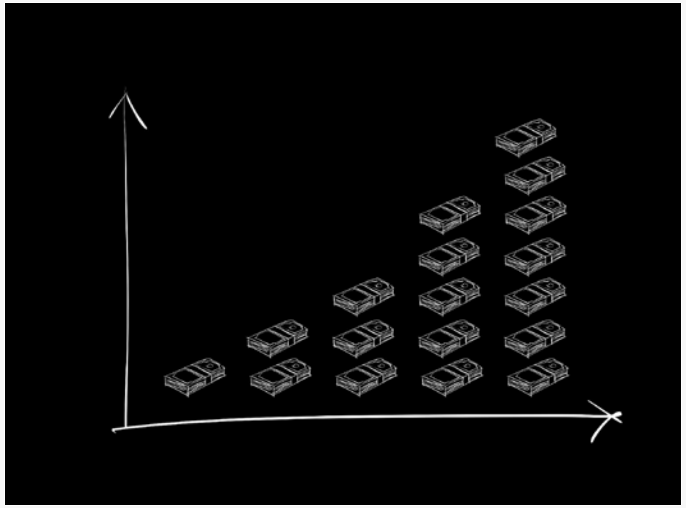
时间复杂度

最终得到的D为图的最短距离矩阵。求出矩阵中的每个值需要进行n次计算，所有元素值需要进行n2次计算。

矩阵算法限制

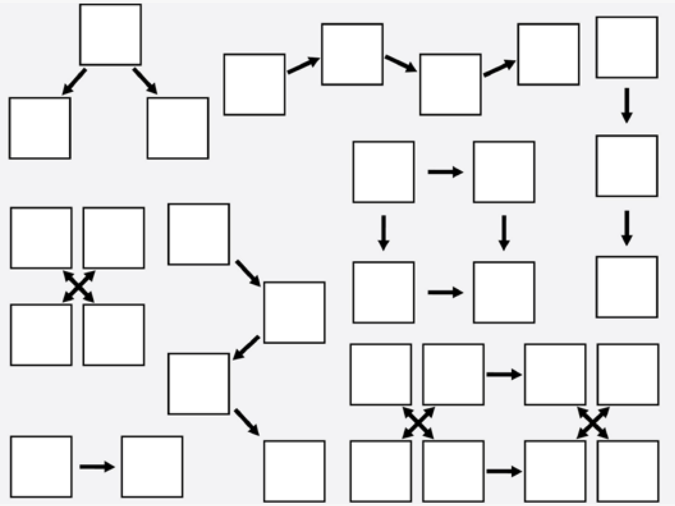
矩阵算法的时间复杂度是O(n4)，随着节点数的增加，计算量和时间复杂度呈指数级增长，因此仅适用于小规模图计算。

4）A*算法



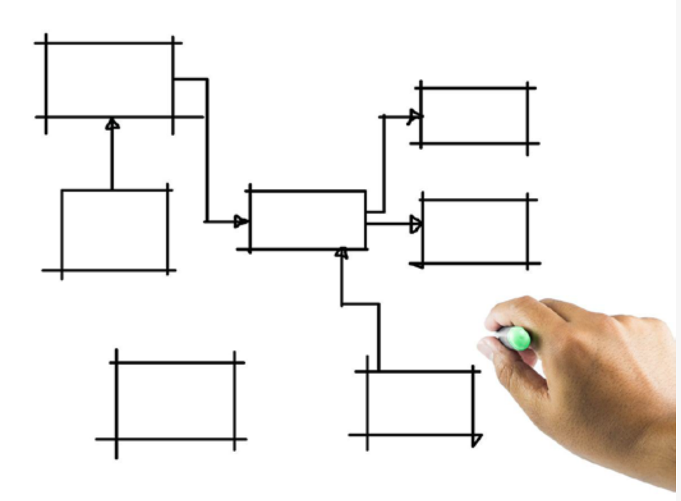
评估节点代价

A*算法使用两个函数来评估节点的代价，分别是实际代价函数g(n)和估算代价函数h(n)。



优先队列

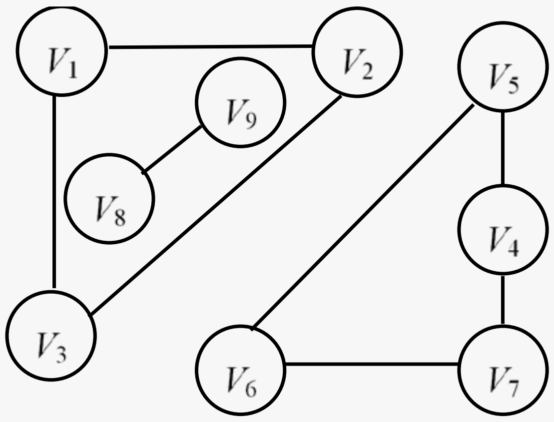
A*算法通过综合这两个代价来选择最有可能导致最优解的路径。



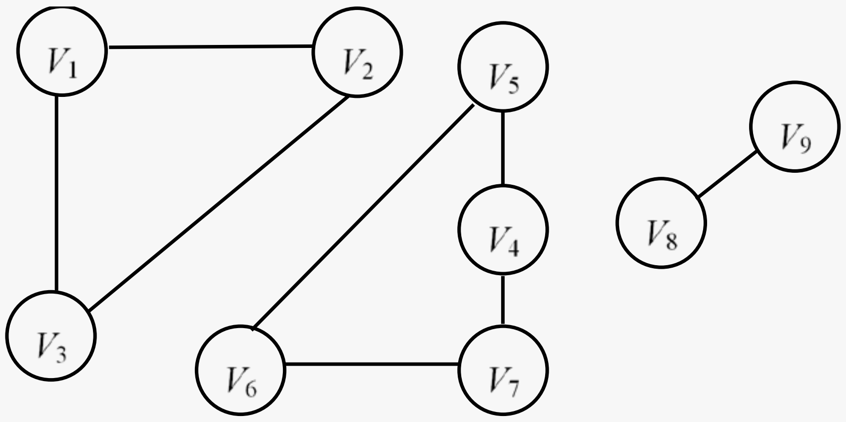
扩展搜索范围

A*算法从起点出发，逐步扩展搜索范围，同时维护一个优先队列。

5.2.2 连通分析



（a）无向图 G_2



（b）G_2 的三个连通分量

连通定义

连通图是无向图G = (V, E)中，如果从顶点Vs到顶点Vt有路径，则称Vs和Vt是连通的。

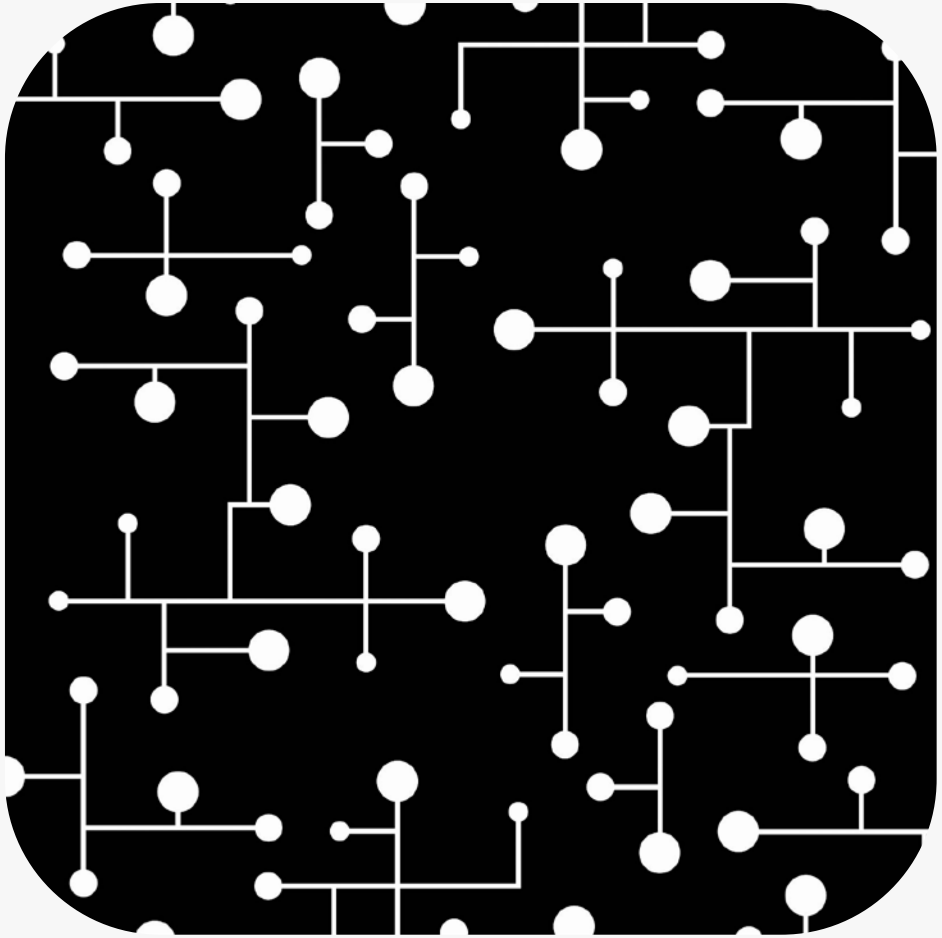
连通图定义

如果对于图G中的任意两个顶点Vi，Vj∈V，Vi和Vj都是连通的，则称G为连通图。

连通图与非连通图

G2是非连通图，G2的三个连通分量是G2的子图，且为极大连通子图。

1）连通图与生成树



连通性分析

连通性分析是根据指定的起始和终止结点，分析两点之间是否连通，或者根据多个点分析是否互通。

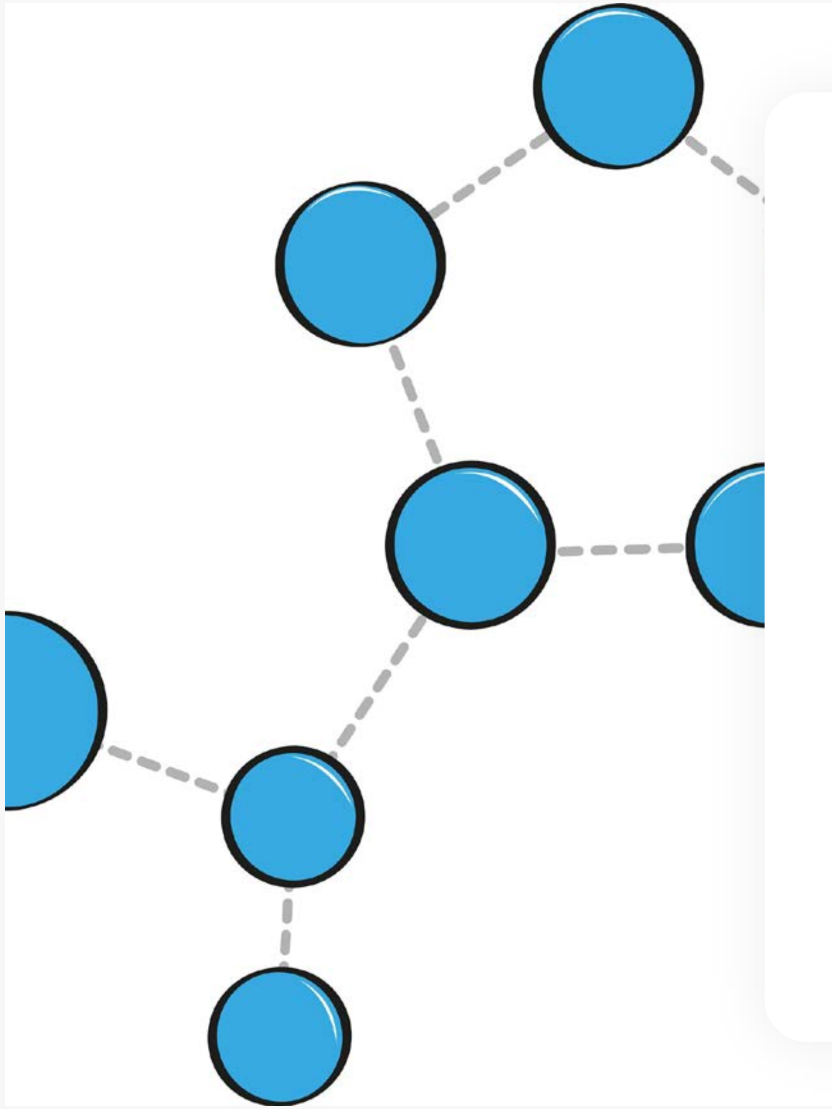
生成树的求解

连通分析的求解过程实质上是图的生成树的求解过程，其中最小生成树问题研究最多。

生成树的应用

最小生成树问题是带权连通图的一个重要应用，在解决最优（最小）代价类问题上用途非常广泛。

1）连通图与生成树



生成树与回路

若连通图G的顶点个数为n，则G的生成树的边数为n-1；树无回路，但把相邻顶点连成一条边，就会得到一个回路；树是连通的，但去掉任意一条边，就会变为不连通的。

生成树的求解方法

对于一个连通图而言，通常采用深度优先遍历或广度优先遍历来求解其生成树。

1）连通图与生成树

中南大學
CENTRAL SOUTH UNIVERSITY

图的遍历

从图中某一顶点出发访遍图中其余顶点，且使每一顶点仅被访问一次，这一过程叫作图的遍历。

深度优先搜索

深度优先搜索的基本思想是从起始节点开始，沿着一条路径尽可能深地访问，直到到达最深处，然后回溯到上一个节点，继续深入其他路径。

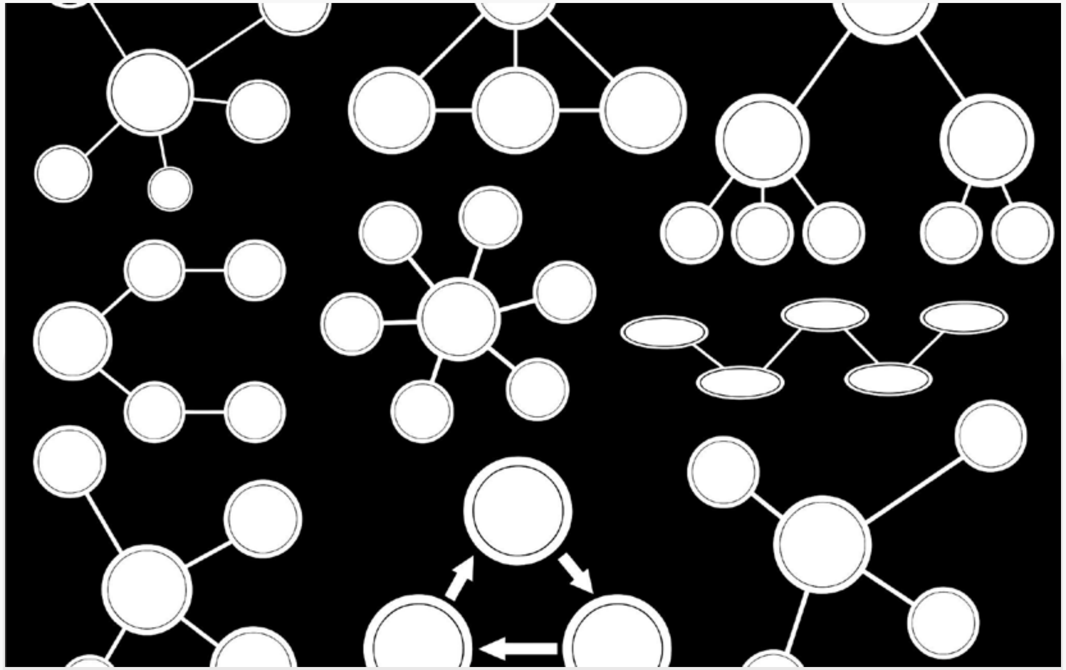
广度优先搜索

广度优先搜索从起始节点开始，先访问当前节点的所有邻居节点，然后依次访问邻居节点的邻居节点，以此类推。

搜索方法的改进

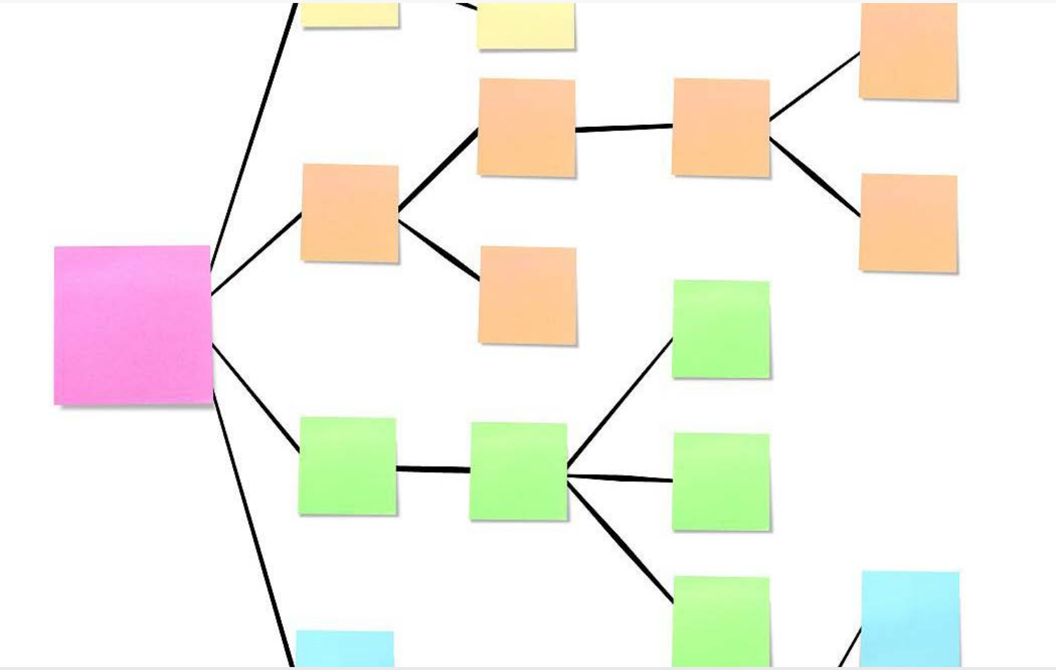
两种搜索方法是地理信息系统网络分析中比较常用的搜索方法，许多算法的提出都是基于其基本思想进行改进和优化的。

2）最小生成树算法



最小生成树算法

最小生成树算法是经典图论问题的解法，在各个领域都有着广泛的应用。



著名算法求解

已有很多算法求解此问题，其中著名的有Kruskal算法和Prim算法。

2）最小生成树算法

Prim算法

从局部最优策略开始，每次从外部选取一条边并将其加入到生成树中，直到生成树包含了所有的顶点。

初始化

将图中的所有顶点划分为独立集合，不加入任何边。

枚举边

从小到大枚举每条边，设当前枚举的边为e(u,v)，其中u属于集合A,v属于集合B。

合并集合

如果集合A和集合B的并集等于整个顶点集合，那么将e(u,v)加入到最小生成树中。

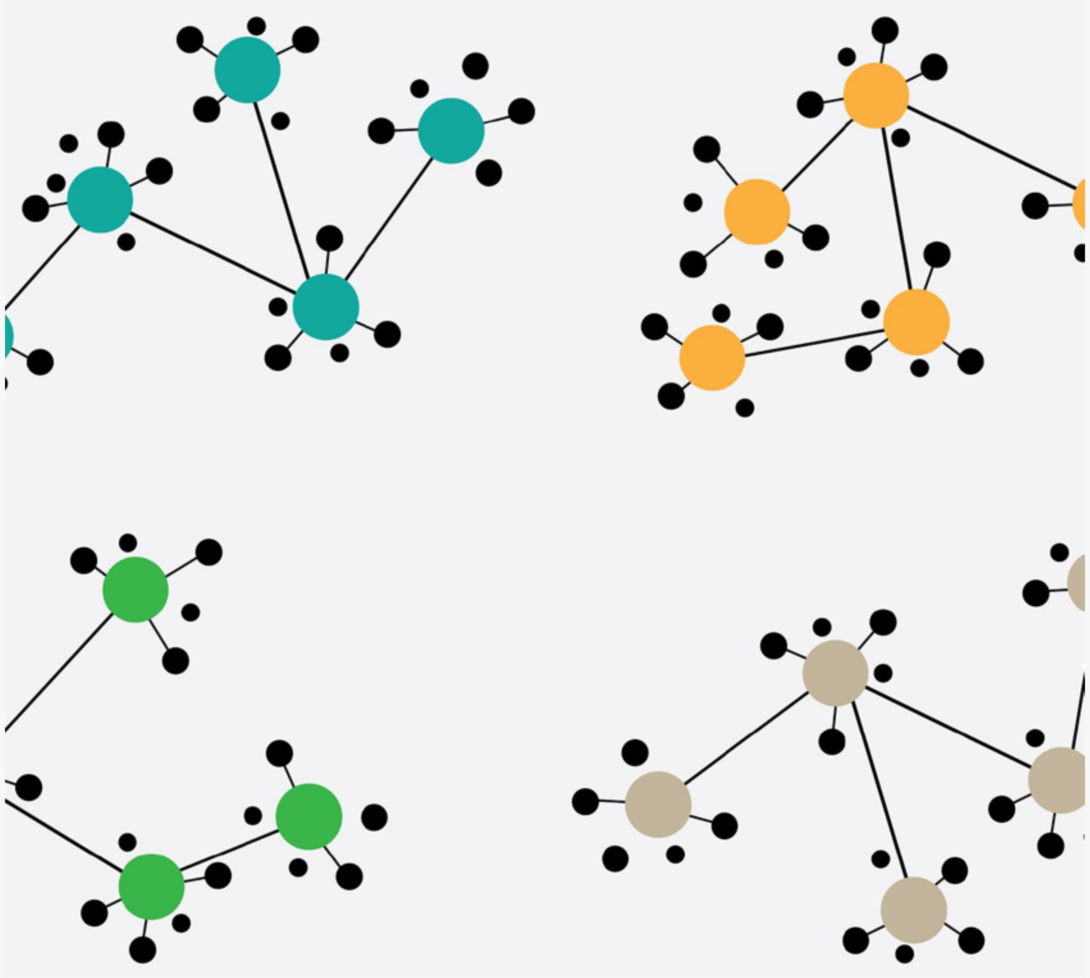
加入边

否则，将集合A和集合B的并集合并成一个集合，将e(u,v)加入到最小生成树中。

重复步骤

重复步骤②-④，直到最小生成树中包含了n-1条边，其中n为顶点数。

2）最小生成树算法



枚举边

从小到大枚举每条边，设当前枚举的边为e(u,v)，其中u和v分别属于两个不同的集合。

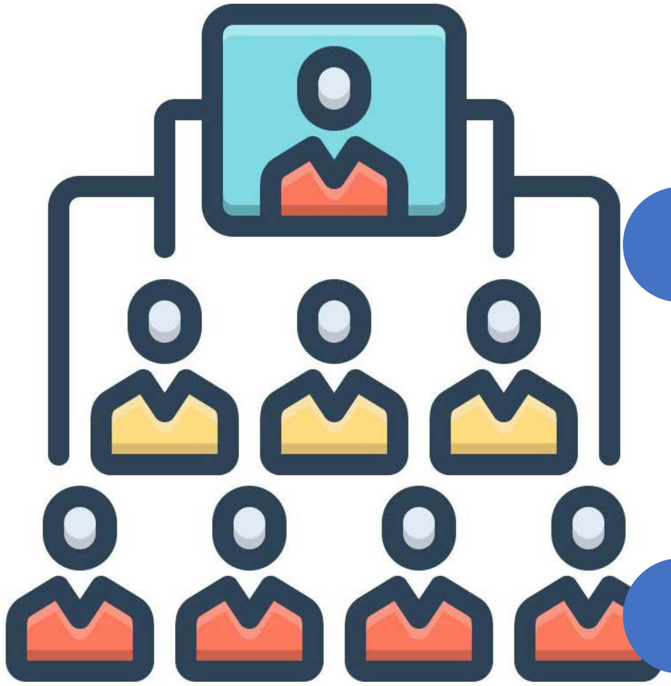
合并集合

如果u和v已经在同一个集合中，那么忽略这条边，继续枚举下一条边。

加入边与合并集合

否则，将e(u,v)加入到最小生成树中，并将u和v所在的两个集合合并成一个集合。

5.2.3 资源分配



01

资源分配的定义

资源分配也称定位与分配问题，需要解决在网络中选定几个供应中心，并将网络的各边和点分配给某一中心，使各中心所覆盖范围内每一点到中心的总的加权距离最小。

02

资源分配的问题

资源分配问题包括定位和分配两个问题。定位是指确定在哪里布设供应点最合适的问题，分配指的是已知供应点，确定其为哪些需求源提供服务的问题。

03

资源分配的工具

定位与分配是常见的定位工具，也是网络设施布局、规划所需的一个优化的分析工具，可以帮助企业合理地分配资源，提高效率和工作效果。

1）定位问题（选址问题）

中南大学
CENTRAL SOUTH UNIVERSITY

选址的问题

选址是在某一指定区域内选择服务性设施的位置，如确定市郊商店区、消防站、工厂、矿仓、堆场等的最佳位置。

选址的限制

网络分析中的选址问题一般限定设施必须位于某个结点或位于某条网线上，或限定在若干候选地点中选择位置。

选址的种类

选址问题种类繁多，实现的方法和技巧也多种多样，不同的GIS系统在这方面各有特色，主要原因是对“最佳位置”具有不同的解释。

2）分配问题

分配问题的体现

分配问题在现实生活中体现为设施的服务范围及其资源的分配范围的确定等一类问题，资源分配是为城市中的每一条街道上的学生确定最近的学校，为水库提供其供水区等。

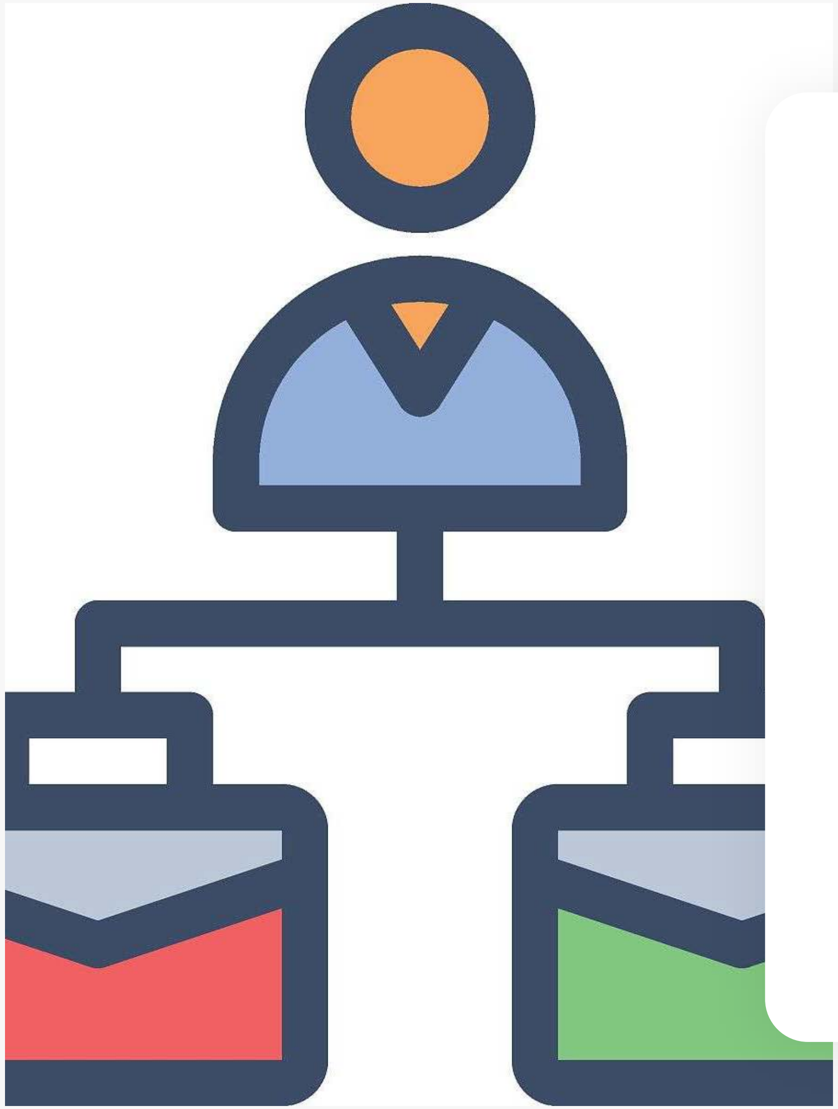
分配问题的模拟

资源分配是模拟资源如何在中心（学校、消防站、水库等）和周围的网线（街道、水路等）、结点（交叉路口、汽车中转站等）间流动的。

网络元素对资源分配的影响

在计算设施的服务范围及其资源的分配范围时，网络各元素的属性也会对资源的实际分配有很大影响，主要属性包括中心的供应量和最大阻值等。

3）中心定位与分配问题



中心定位与分配问题的介绍

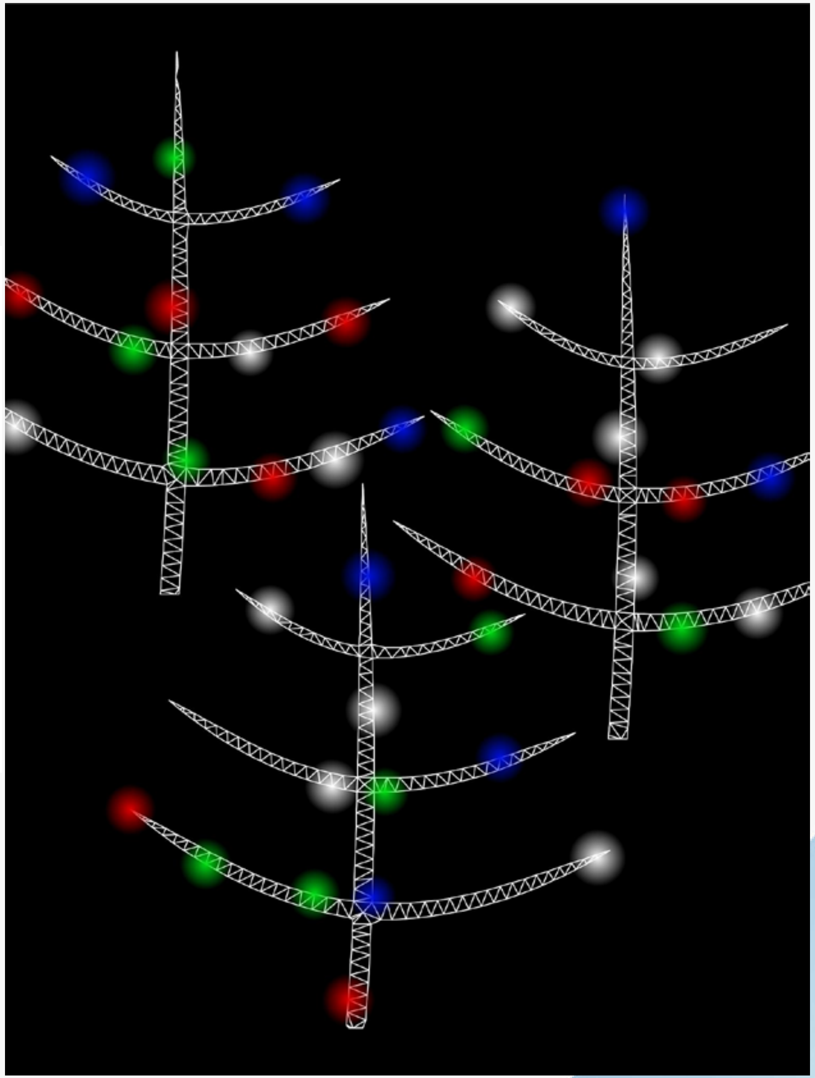
许多资源分配问题的供应点布设要求满足多种组合条件，如选择供应点时不仅要求使总的加权距离最小，还要使总服务范围最大等。

中心定位与分配问题的分解

中心定位与分配问题可以分解为多个单目标问题，利用单目标方程即最小目标值法来求解，所谓目标方程是用数学方式表达满足所有需求点到供应点的加权距离最小的条件方程。

1）网络与流

设有向图$G=(V,A,C)$，其中V表示节点的集合，v_s为V中的发点，v_t为V中的收点，其余为中间节点，A表示弧的集合。对于每一个弧$(v_i,v_j)\in A$，对应有一个弧的容量$c_{ij}\geq 0$，并称$G=(V,A,C)$为网络，网络上的流量是指定义在弧(v_i,v_j)上的函数$f_{ij}\geq 0$。



2）可行流与最大流

流量定义

网络上的流量是指定义在弧上的函数，它在实际应用中需满足两个基本条件。

可行流约束

满足容量限制约束和流量平衡约束的流称之为网络的可行流。

零流与可行流

零流是网络中所有弧上的流量均为0时的一种特殊情况，它同样是可行流。

线性规划与图论方法

网络最大流问题是一种特殊的线性规划问题，利用图论的方法是求解的基本思路。

最大流问题

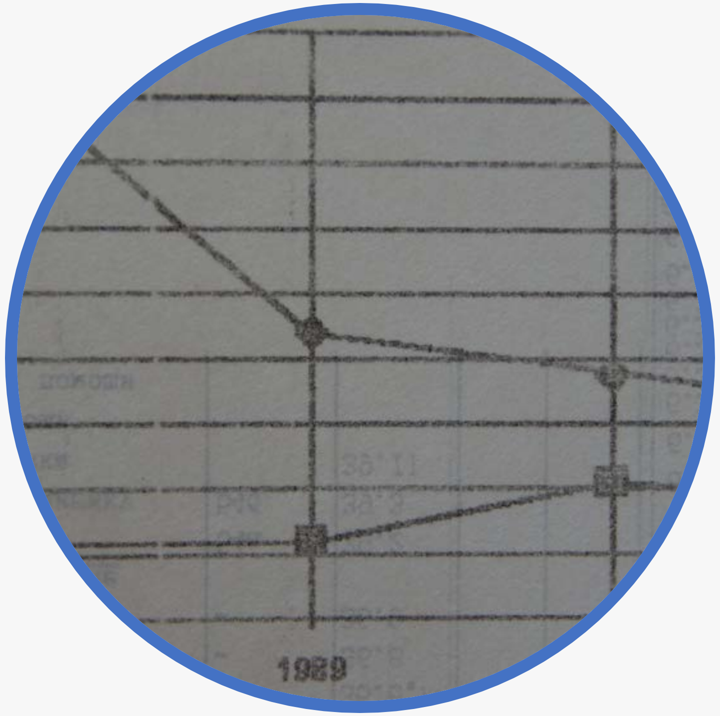
网络最大流问题是求解网络的一个可行流，使网络的总流量最大化的问题。

3）增广链

饱和弧与非饱和弧

在网络中，满足条件弧称为饱和弧，满足条件弧称为非饱和弧，满足条件弧称为零弧，满足条件弧称为非零弧

。

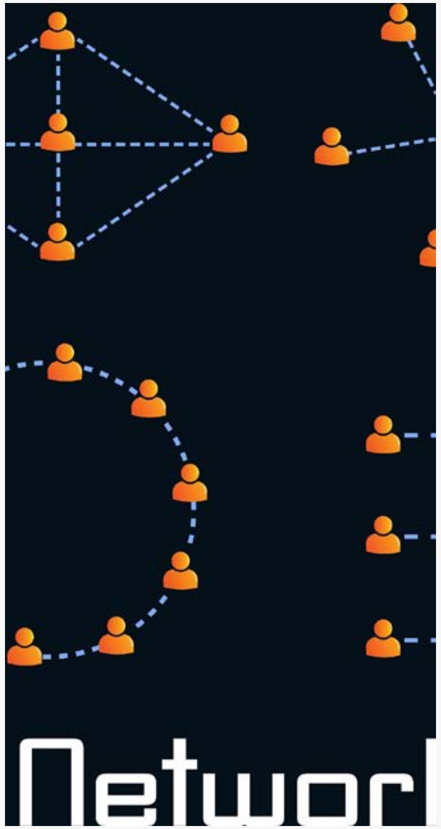


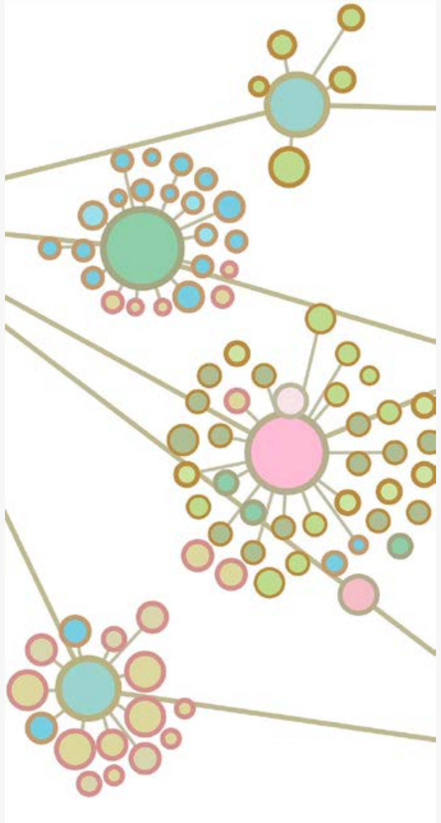
可行流与增广链

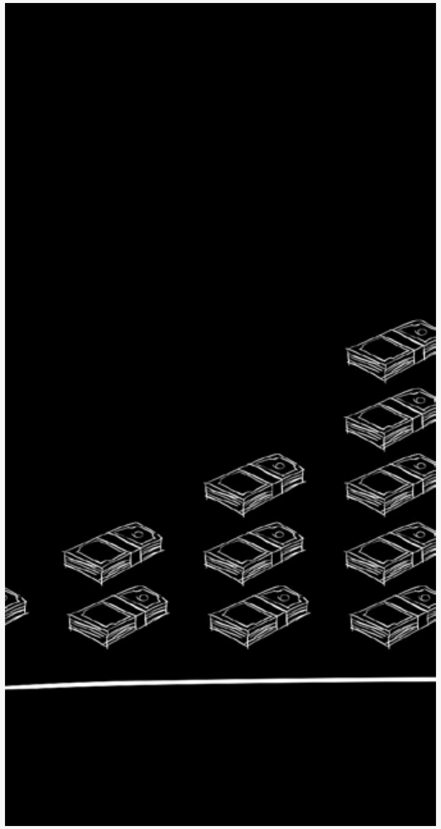
是网络的一个可行流，将网络中满足条件的弧称为饱和弧，满足条件的弧称为非饱和弧，满足条件的弧称为零弧。

增广链的方向

是网络由发点出发，经过网络中的若干中间节点，回到收点处的一条链路，其中链路的方向为→，则网络中该方向相同的称之为前向弧。







增广链的约束条件

设是网络的一个可行流，是网络由发点出发，经过网络中的若干中间节点，回到收点处的一条链路，若满足如下约束关系，则称为网络关于可行流的增广链。

增广链的作用

增广链是网络中连接发点和收点的关键路径，通过增广链，可以找到网络中的瓶颈环节，进而优化网络流量。

4）截集与截量

- 截集与截量定义：给定网络，将节点集合分割为两个非空节点集合和，其中，，则将弧集称为分割网络中发点和收点的截集。
- 截量计算与性质：截集中所有弧的容量之和称为截量，用表示。发点至收点的必定经过截集，任一可行流的流量不会超过截量。
- 最大流判断：若可行流的流量等于截量，则该可行流必为最大流。寻找最大流的一般思路是判断是否存在增广链，若存在则进行增广。
- 增广链算法：基于寻找增广链的思想，衍生出多种求解最大流问题的算法，如Ford-Fulkerson方法、AP算法和PR算法等。
- Ford-Fulkerson方法：给定有向网络和发收点，Ford-Fulkerson方法初始时将各弧的流量设置为0，然后在网络中寻找出一条增广链。
- 增广链与残量网络：增广链是残量网络中从发点至收点的路径，其中弧的剩余容量构成残量网络。沿增广链增广流量后，判断网络中是否依然存在增广链。

（1）Ford-Fulkerson方法

寻找增广链

在网络中寻找出一条增广链，增广链是残量网络中从发点至收点的路径，由剩余容量大于0的弧组成。

增广流量

沿增广链增广流量，即增加弧的流量，直到达到弧的容量上限。增广流量后，需要更新残量网络。

判断是否存在增广链

判断网络中是否依然存在增广链，若存在，继续执行②；若不存在，算法终止。

(2) SAP算法

01 初始化流量

将各弧的流量初始化为0，开始时网络中不存在增广链。

02 判断增广链

判断残量网络中是否存在增广链，若存在，执行下一步；若不存在，算法终止。

03 寻找增广链

在残量网络中寻找一条路径最短的增广链，即剩余容量最小的弧组成的路径。

04 计算剩余容量

计算出增广链中剩余容量最小的弧，对应的剩余容量为。

05 增广流量

沿增广链增广值为的流量，即增加弧的流量，直到达到弧的容量上限。

06 更新残量网络

更新残量网络，继续执行②，寻找新的增广链，直到网络中不存在增广链为止。

（3）Edmonds–Karp算法

01 Edmonds–Karp算法

Edmonds–Karp算法是指SAP算法在残量网络中使用BFS策略寻找最短路径的算法。

02 BFS寻找最短路径

算法每次用一遍BFS寻找从发点和收点的最短路径作为增广路径，然后增广流量并修改残量网络。

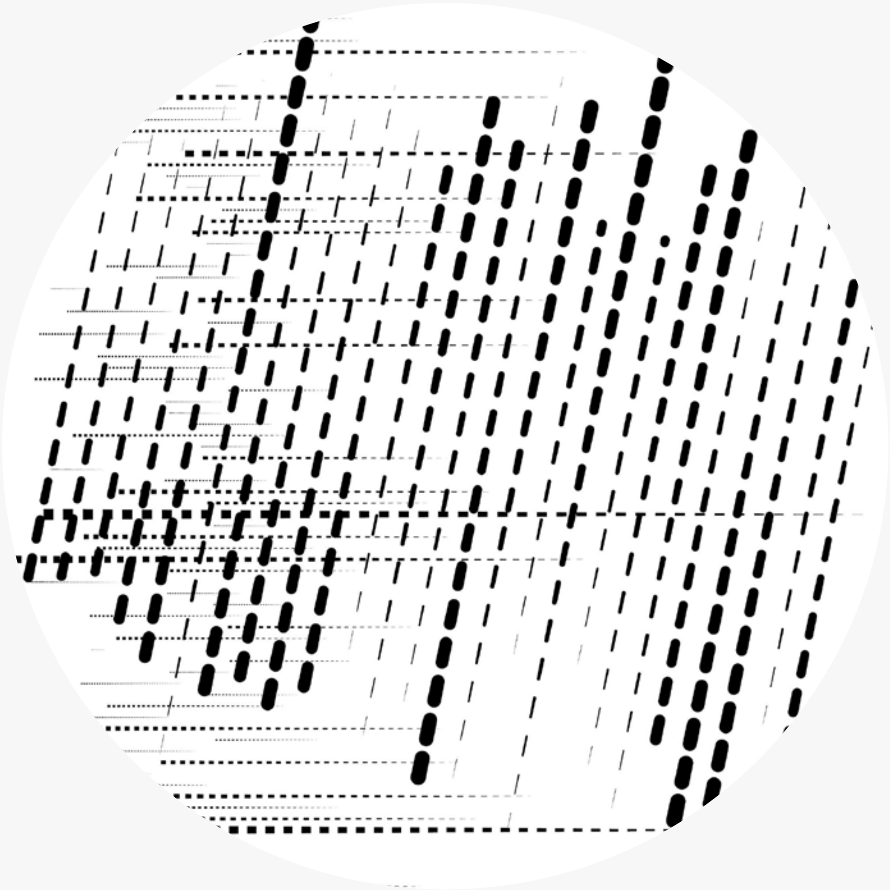
03 频繁BFS效率低

由于BFS要搜索全部小于最短距离的分支路径之后才能找到收点，因此频繁地BFS效率较低。

04 时间复杂度高

Edmonds–Karp算法的时间复杂度为，这是因为BFS需要搜索所有分支路径，导致计算量较大。

(4) Dinic算法



BFS寻找收点

BFS在寻找收点时过于缓慢，而深度优先搜索（DFS）无法保证找到最短路径。

DFS寻找增广路径

Dinic算法结合了BFS与DFS的优势，采用构造分层网络的方法可以较快找到最短增广路径。

构造分层网络

在分层网络中，只保留满足条件的边，BFS树中节点距离发点距离函数。

(4) Dinic算法

增广路径

从发点出发可直接到达的节点的距离为1，经过某一个节点可到达的节点的距离为2，依此类推。

分层网络

称所有具有相同距离的节点位于同一分层，分层网络中的任意路径就成为到达此顶点的最短路径。

算法流程

Dinic算法每次使用一遍BFS构建分层网络，然后在中使用一遍DFS找到所有到收点的增广路径。

算法结束

重新构造分层网络，若收点不在中，则算法结束。

（5）MCP算法

寻找增广路径

MCP算法每次寻找增广路径时并不是采用BFS寻找最短路径，而是采用Dijkstra寻找容量最大的路径。

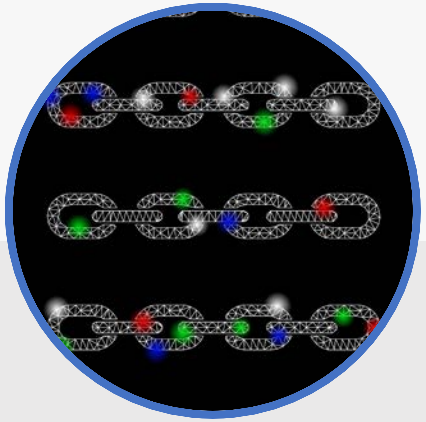
Dijkstra寻找容量

MCP算法与SAP类算法相比，可更快逼近最大流，从而降低增广操作的次数。

BFS与Dijkstra

BFS的时间复杂度为，而Dijkstra的时间复杂度为，因此MCP算法与SAP类算法相比，效率相对低下。

（6）CS算法



CS算法思想

CS算法采用二分查找的思想，寻找增广路时不必非局限于寻找最大容量，而是找到一个可接受的较大值即可。

复杂度与操作

CS算法有效降低寻找增广路时的复杂度，增广操作次数也不会增加太多。CS算法时间复杂度为。

效率对比

CS算法效率稍优于MCP算法，但与SAP类算法相比，效率依然相对低下。

5.3 空间统计学分析

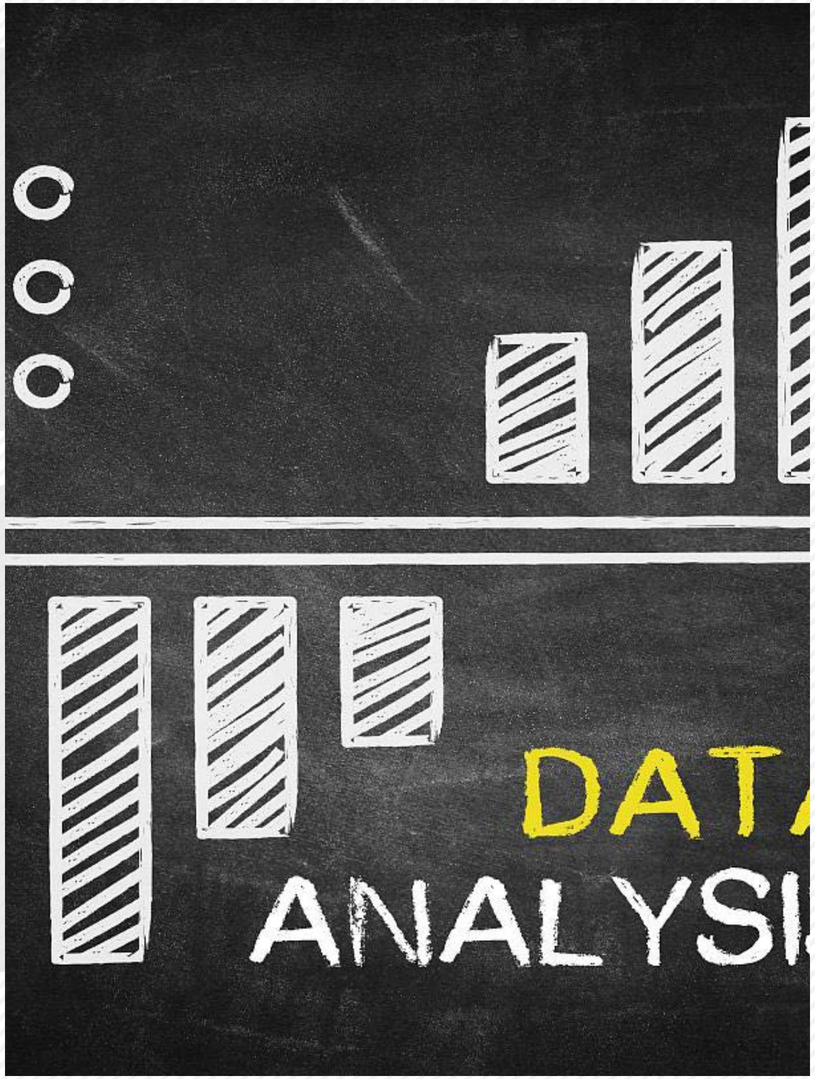
5.3 空间统计学分析

空间统计分析简介

空间统计分析是统计分析理论在空间科学的应用和拓展，是统计学与地理学交叉的学科内容。

空间统计分析的分类

空间统计分析包括“空间数据的统计分析”及“数据的空间统计分析”。空间数据的统计分析着重于空间物体和现象的非空间特性的统计分析，据的空间统计分析则是直接从空间物体的空间位置、联系等方面出发，研究既具有随机性又具有结构性，或具有空间相关性和依赖性的自然现象



空间统计学分析的应用领域

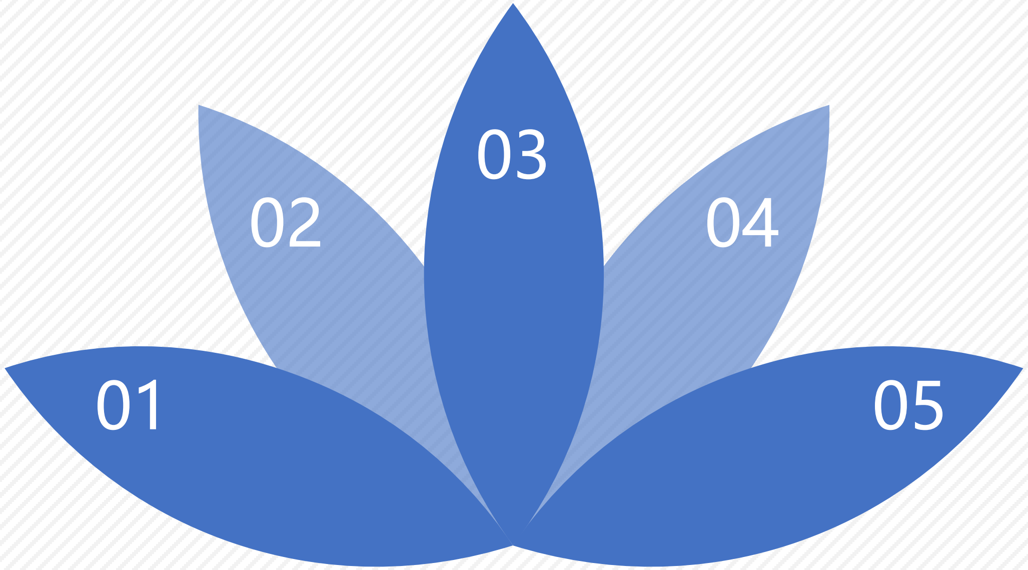
空间统计分析的应用领域：空间统计分析的方法可以应用于各种领域，包括环境科学、地理学、地质学和社会学等。

环境科学中的应用：在环境科学中，空间统计分析可用于研究环境污染、生态保护等方面的空间分布和依赖关系。

地理学中的应用：在地理学中，空间统计分析可用于研究地理现象的空间分布、变化和趋势等。

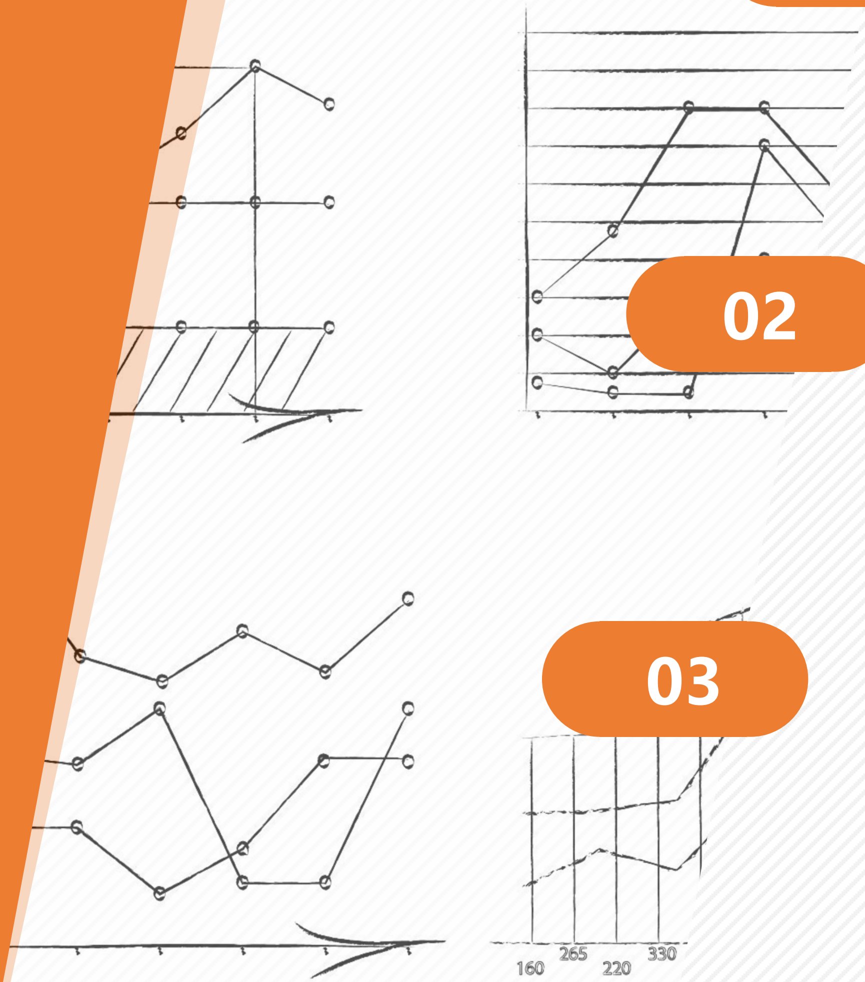
地质学中的应用：在地质学中，空间统计分析可用于研究矿产资源、地质构造等方面的空间分布和依赖关系。

社会学中的应用：在社会学中，空间统计分析可用于研究社会现象的空间分布、变化和趋势等。



5.3.1 空间统计分析方法的基本理论 1）基本概念

中南大學
CENTRAL SOUTH UNIVERSITY



01

空间统计分析方法

空间统计分析方法应运而生，以解决传统数理统计无法处理的空间样本点选取、空间估值和多组空间数据的关系等问题。

02

空间统计学

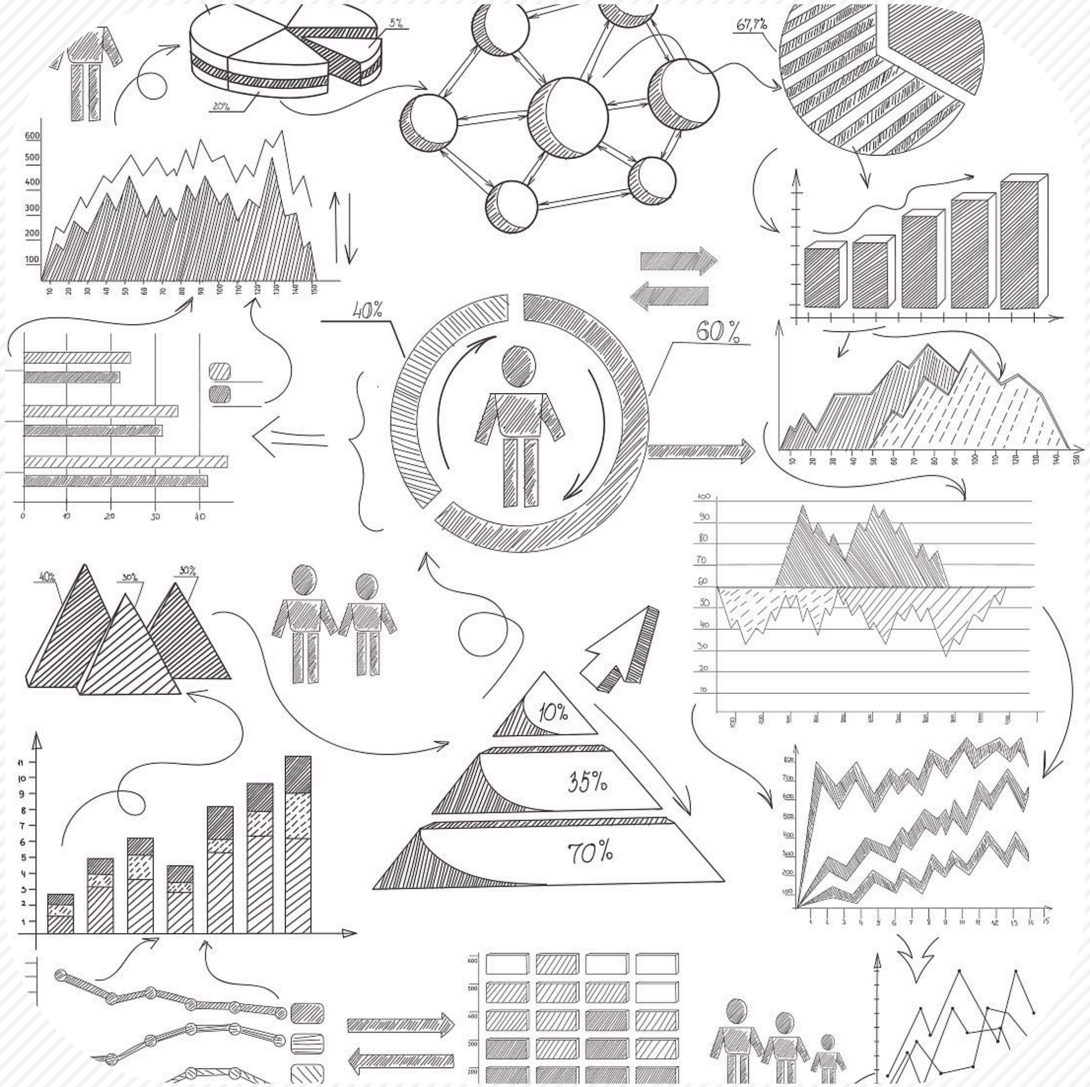
20世纪60年代，G.Matheron的理论研究基础上形成了空间统计学，以区域化变量理论为基础，研究具有地理空间信息特性的事物或现象的空间相互作用及变化规律。

03

克立格法

G.Matheron创造了“克立格法”这一名词，以表彰克立格在矿床的地质统计学评价工作中所起到的先驱作用。

5.3.1 空间统计分析方法的基本理论 1）基本概念



揭示相关规律

空间统计分析方法假设研究区中所有的值都是非独立的，相互之间存在相关性，这种相关性在空间或时间范畴内被称为自相关。

未知点预测

根据空间数据的自相关性，可以利用已知样点值对任意未知点进行预测，但事实上，在进行未知点预测之前并不知道数据间具体的相关规律。

揭示相关规律重要性

揭示空间数据的相关规律是空间统计分析的重要任务之一，而利用相关规律进行未知点预测是空间统计分析的另一个重要任务。

2）空间自相关理论

01

空间自相关

空间自相关理论是研究空间中某位置的观察值与其相邻位置的观察值之间的相互依赖性。

02

空间规律性

当变量在空间上表现出一定的规律性，即不是随机分布，则存在着空间自相关。

03

空间自相关方向

空间自相关可以分为正相关和负相关，正相关表明某单元的属性值变化与其邻近空间单元具有相同变化趋势，负相关则相反。

04

空间自相关距离

当空间自相关仅与两点间距离有关时，称为各向同性；当考虑方向的影响时，可能出现不同的自相关值。

3）空间统计分析中的理论假设 （1）区域化变量

区域化变量定义

空间统计学以区域化变量理论为基础，研究空间中呈一定结构性和随机性的自然现象。

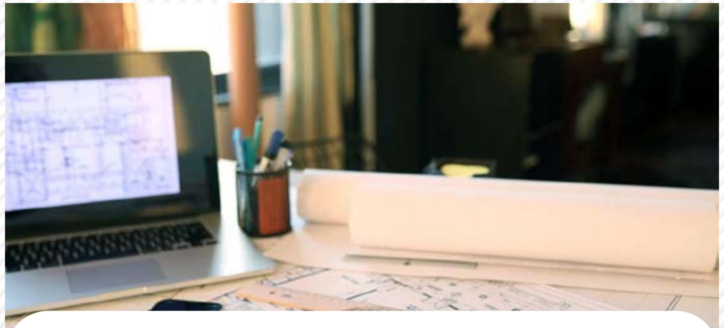
区域化变量随机场

当一个变量呈空间分布时，称之为区域化，而区域化变量是指以空间点x的三个直角坐标为自变量的随机场。

随机场与空间数据

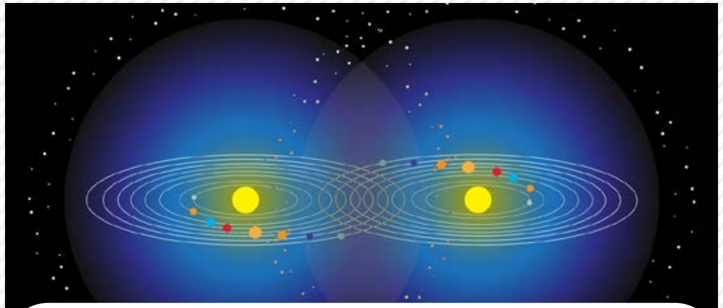
随机场反映某种空间现象的特征，通过抽样或随机观测得到现实随机场Z(x)，成为空间数据。

（1）区域化变量



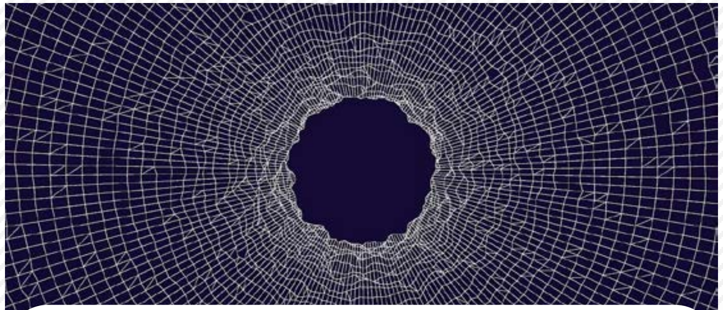
区域化变量的两重性

区域化变量的两重性表现在观测前把它看成是随机场（依赖于坐标），观测后是一个普通的空间三元函数值或一个空间点函数。



区域化变量的属性

G.Matheron定义的区域化变量是一种在空间上具有数值的实函数，它在空间的每一个点取一个确定的数值。



区域化变量的局限性

区域化变量被限制于一定空间范围，称为区域化的几何域，域内变量属性明显，域外属性不明显或为零。

（1）区域化变量

区域化变量的连续性

不同的区域化变量具有不同程度的连续性，这种连续性是通过区域化变量的半变异函数来描述的。

区域化变量的各向异性

区域化变量在各个方向上具有相同性质时称各向同性，否则称为各向异性，分析样点间的自相关程度。

区域化变量的空间相关

区域化变量在一定范围内呈一定程度的空间相关，当超出这一范围之后，相关性变弱甚至消失。

区域化变量的变异性

对于任一区域化变量而言，特殊的变异性可以叠加在一般的规律之上，表现出更加复杂的特性。

（2）协方差函数

随机函数的协方差函数

在随机函数中，当只有一个自变量x时称为随机过程，随机过程在时间t1和t2处的随机变量、的二阶混合中心矩定义为随机过程的协方差函数。

随机场的自协方差函数

当随机函数依赖于多个自变量时，称为随机场，而随机场Z(X)在空间点X和X+h处的两个随机变量Z(x)和Z(x+h)的二阶混合中心矩定义为随机场Z(X)的自协方差函数。

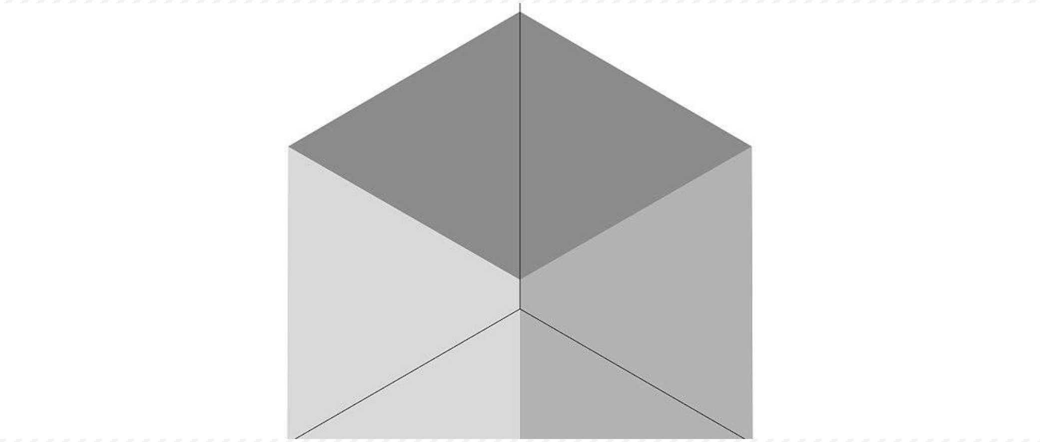
$$Cov\{Z(x),Z(x+h)\}=E[Z(x)Z(x+h)]-E[Z(x)]E[Z(x+h)]$$

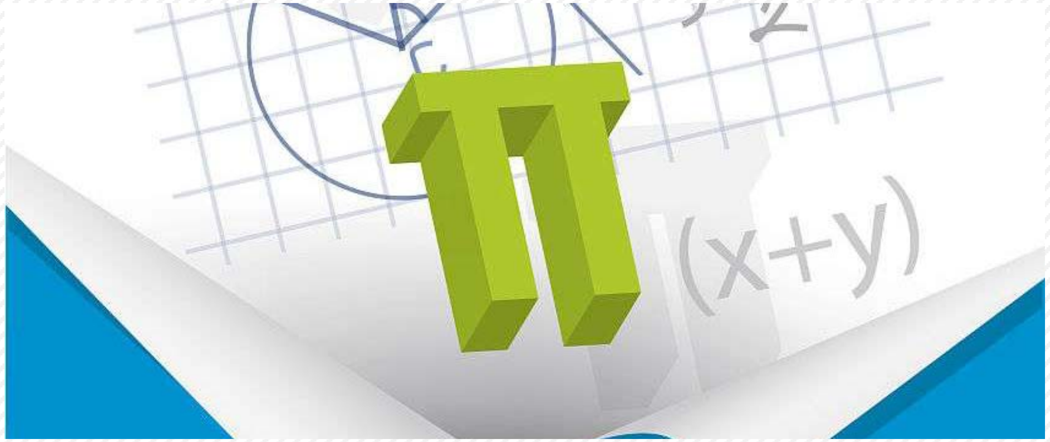
自协方差函数

随机场Z(x)的自协方差函数亦称为协方差函数，一般地，协方差函数依赖于空间点x和向量h。当h = 0时，自协方差函数变为。

$$Cov(x,x+0)=E[Z(x)]^2-\{E[Z(x)]\}^2$$

（3）变异函数





变异函数定义

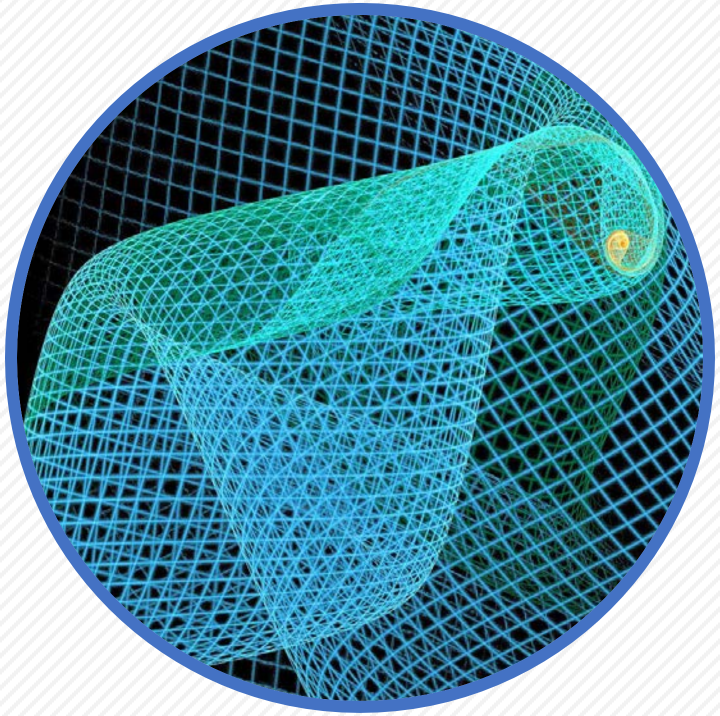
变异函数或变差函数是空间统计学的基本理论，在一维条件下，当空间点x在一维x轴上变化时，区域变量Z(x)在点x和x+h处的值Z(x)与Z(x+h)差的方差一半定义为区域变量Z(x)在x轴上的半变异函数。

$$\gamma(h)=\frac{1}{2}E[Z(x)-Z(x+h)]^2$$

（4）平稳性假设及内蕴假设

二阶平稳假设

区域化变量满足数学期望对任意x存在且等于常数，空间协方差函数对任意x和h存在且平稳，即h=0时，式(4-5)改写成。



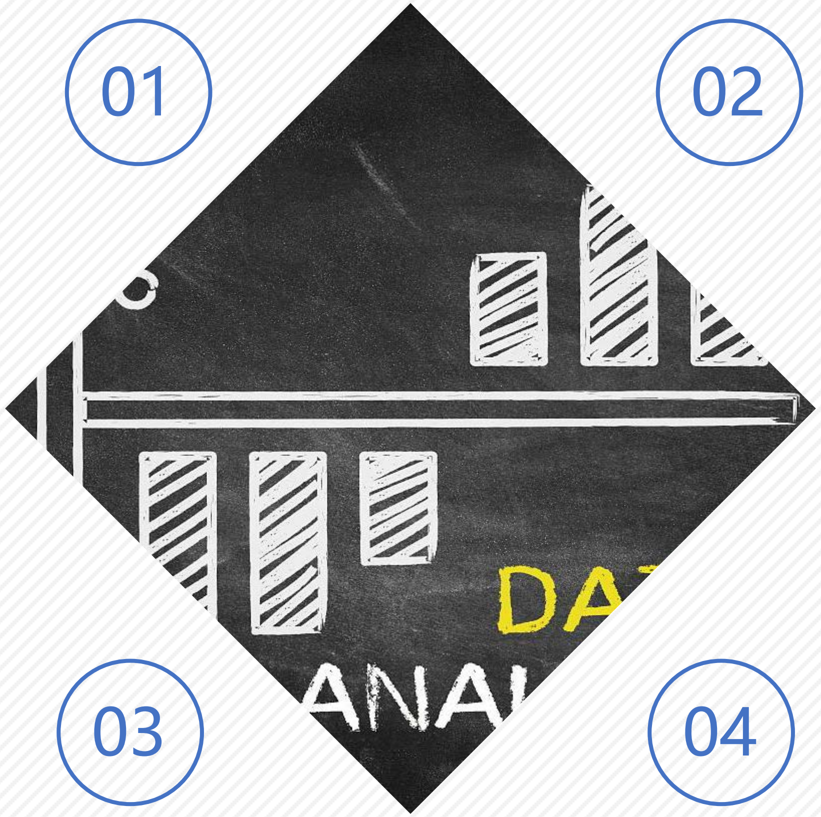
平稳性假设

设某一随机函数，其空间分布律不因平移而改变，即若对任一向量h，关系式成立时，则该随机函数为平稳性随机函数。

协方差平稳

协方差平稳意味着方差及半变异函数平稳，从而有关系式，半变异函数可用于表示矿化范围内区域化变量的空间结构性。

(4) 平稳性假设及内蕴假设



01 内蕴假设

协方差函数不存在时，区域化变量Z(x)的增量Z(x) - Z(x+h)满足条件一和条件二，称该区域化变量满足内蕴假设。

02 内蕴假设理解

随机函数Z(x)的增量Z(x) - Z(x+h)只依赖于分隔它们的向量h（模和方向）而不依赖于具体位置x。

03 半变异函数估计

被向量h分割的每一对数据可以看成是一对随机变量的一个不同现实，而半变异函数的估计量为。

04 准平稳假设

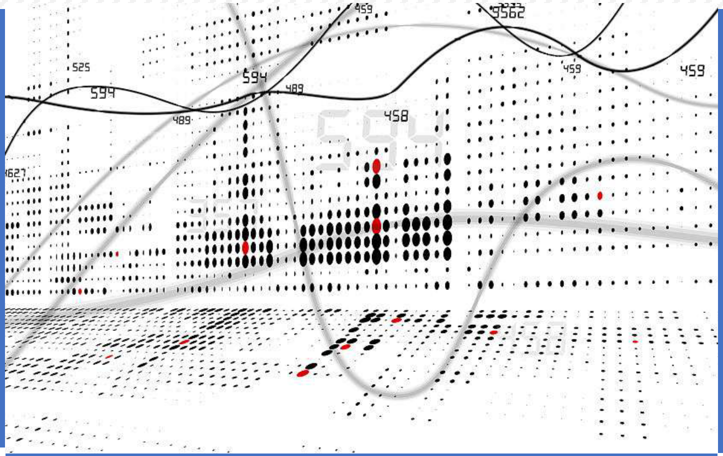
随机函数只在有限大小的邻域内是平稳的，称该随机函数服从准平稳假设，准平稳假设是一种折中方案。

$$\gamma^{*}(h)=\frac{1}{2N(h)}\sum_{i=1}^{N(h)}[Z(x_i)-Z(x_i+h)]^2$$

5.3.2 空间插值方法

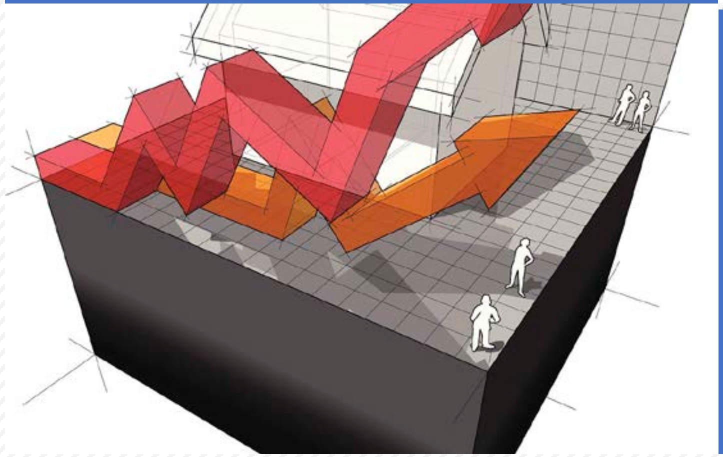
空间插值方法

是指在已知空间位置的观测数据点之间进行插值，以获得未知空间位置的观测数据点的方法。



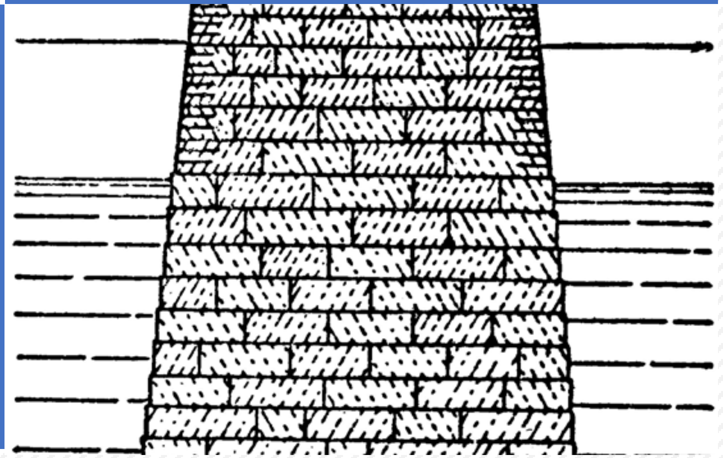
常见的空间插值方法

包括Kriging方法、距离平方倒数法、最邻近法、反距离法等，每种方法都有其特点和适用范围。



空间插值方法的作用

是地理学和环境科学中非常重要的一个研究领域，可以应用于各种实际场景中。



1）临近数据搜索

1

临近样品数据搜索

属性空间插值首要问题，包括确定搜索范围、减少数据聚类效应、加快搜索速度等。

2

八分圆搜索方法

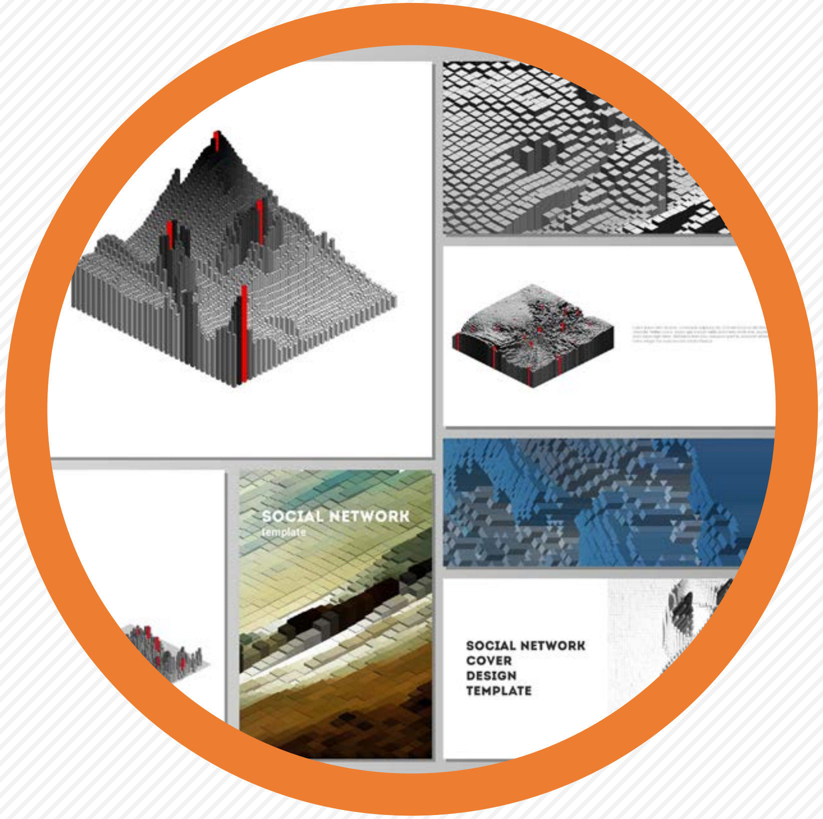
三维空间中采用八分圆的搜索方法能有效减少数据聚类效应，将搜索椭球体沿三个轴的垂直面分割，得到八个扇区。

3

最大参与插值数据个数

每个扇区内的全部数据点按照距离待估点的远近排序，从最近的开始选取，直至达到所设置的最大值。

1）临近数据搜索



八分圆与四分圆

八分圆的二维形式四分圆。

临近数据减少聚类

可通过控制临近数据中同属于一个钻孔工程的样品数目减少聚类效应，对于一个位置的插值，如果是在钻孔工程控制区域内部，临近数据中应该至少有两个钻孔工程的样品。

八分圆搜索控制工程

大多情况下使用八分圆可以在某种程度上达到这种效果，但可以设置具体的工程个数直接对临近数据的搜索加以控制。

优化搜索速度

01

建立网格

在样本点数据范围内建立一个三维网格，按其在网格中的索引位置进行排序。

02

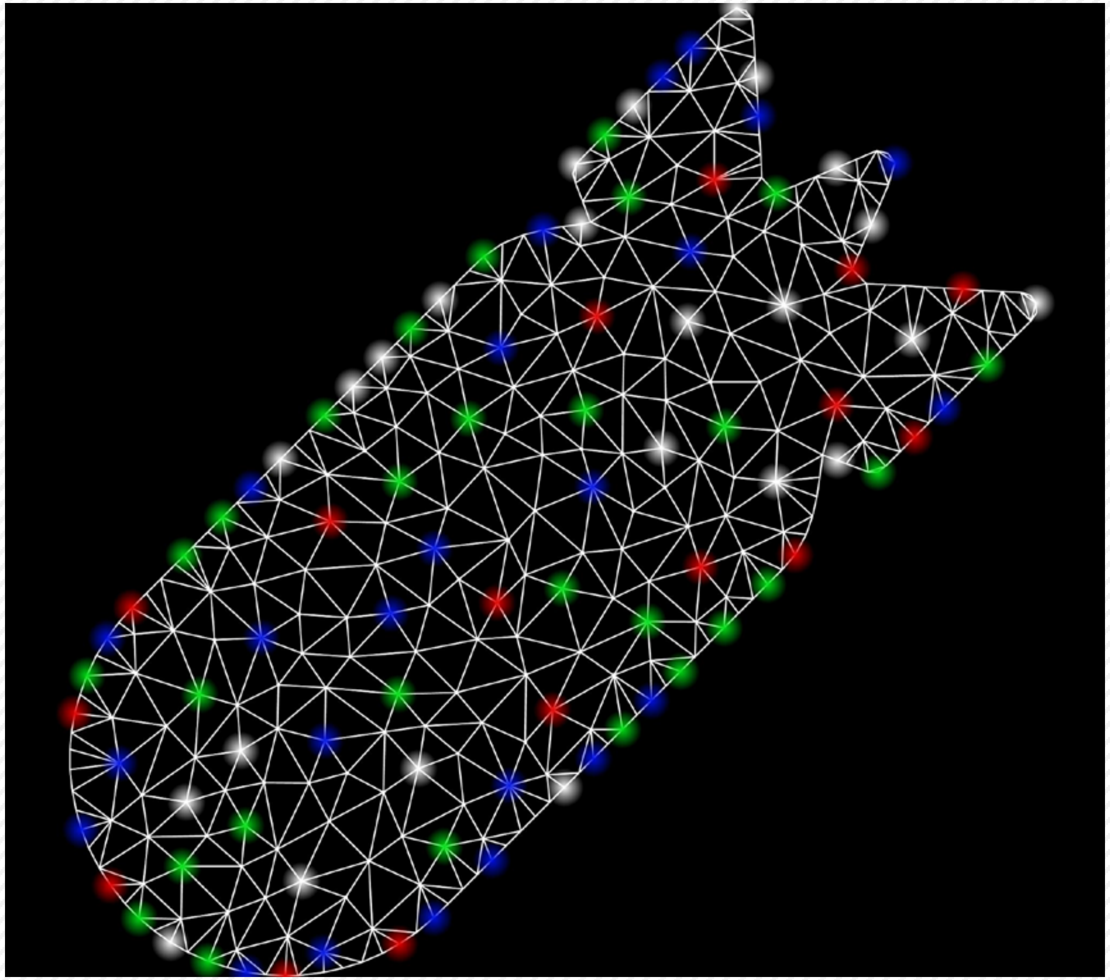
存储样本点数

建立一个与三维网格单元块数目大小一致的数组，存放网格中样本点的累积数目。

03

快速选取数据

通过数组，可以很容易地快速选取指定位置的临近数据，提高搜索效率。



空间插值

统计学中，空间插值是一种将离散点的测量数据转换为连续的数据曲面，以便与其他空间现象分布模式进行比较的方法。

空间内插算法

已知点数据推求未知点数据，通过内插目标，提高数据密度、显示数据空间分布、格网化无规则分布空间数据。

$$z^*(x_0)=\sum_{i=1}^{n}\lambda_i z(x_i)$$

空间外推算法

已知区域数据推求其他区域数据，不同插值方法所赋予的权重值不相同。

3）空间插值类型与方法

空间插值方法

根据空间插值的确定性，插值方法分为确定性方法和地质统计学方法。

确定性方法

基于已知数据点之间的数学关系，未知点的值是确定的，不考虑随机性。

地质统计学方法

基于统计原理，考虑数据之间的空间自相关性和不确定性。

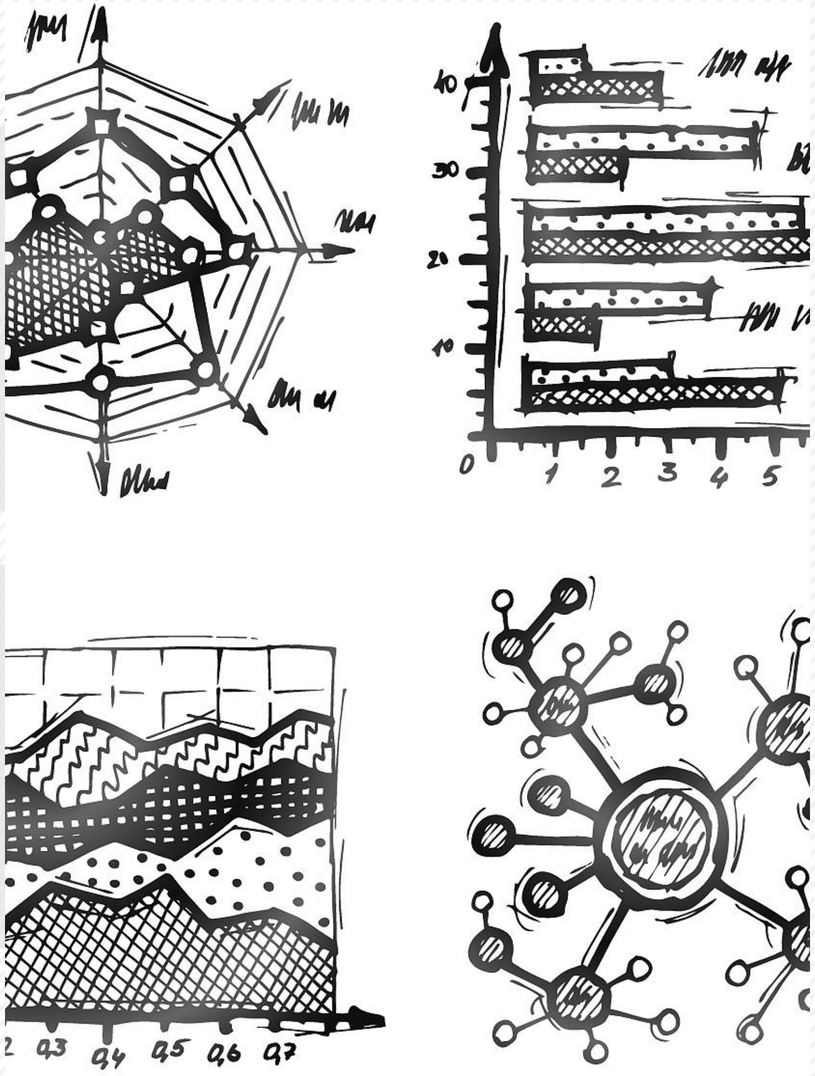
（1）几何方法

泰森多边形

泰森多边形是用于生成“领地”或控制区域的一种方法，特别适合于专题数据的内插，因为它生成专题与专题之间明显的边界。

反距离加权法

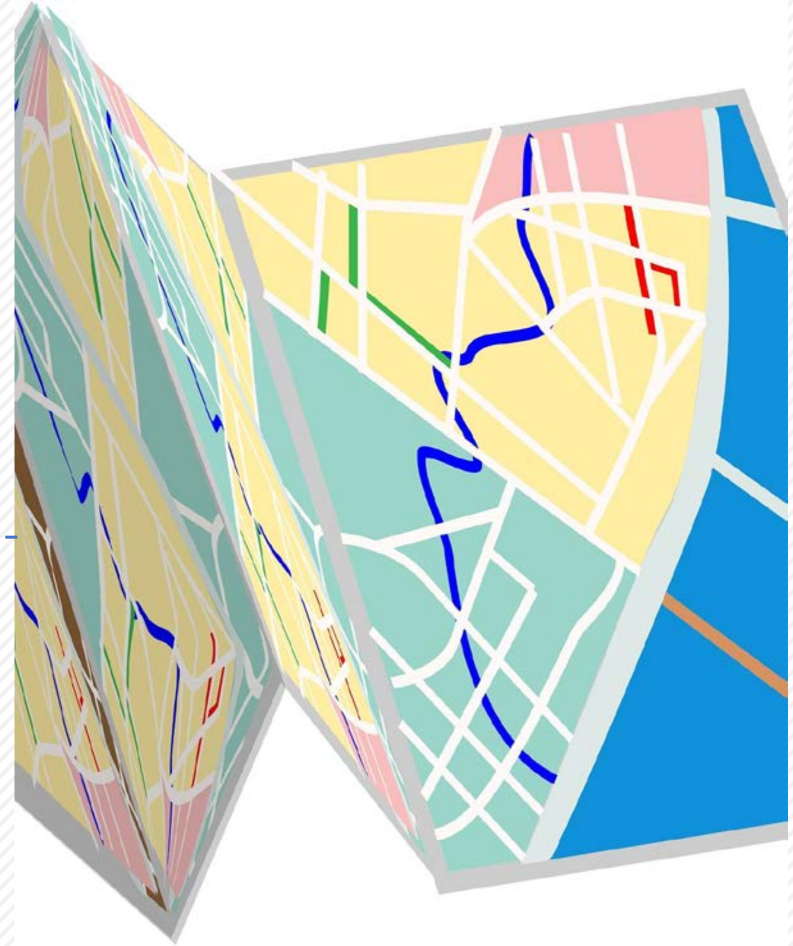
反距离加权法是最常用的空间内插方法之一，认为与未采样点距离最近的若干个点对未采样点值的贡献最大，贡献与距离幂次成反比。



(2) 统计方法

趋势面法

趋势面法可通过全局多项式插值法将由数学函数（一般为多项式）定义的平滑表面与输入采样点进行拟合。



多元回归法

多元回归法是一种统计方法，用于处理多个自变量和因变量之间的问题，不需要分布的先验知识，因此具有广泛的应用范围。

(3) 空间统计方法

Kriging插值法：Kriging插值法是空间统计分析方法的重要内容之一，它是建立在半变异函数理论分析基础上的，是对有限区域内的区域化变量取值进行无偏最优估计的一种方法。

Cokriging插值法：Cokriging插值法是Kriging插值法的一种变种，它通过探索性空间数据分析工具得到先验分布，假定空间随机变量具有二阶平稳或内蕴假设，能够克服插值中误差难以分析的问题。

空间统计学：空间统计学是一种以空间相关的先验模型为基础的方法，通过探索性空间数据分析工具得到先验分布，可以对误差做出逐点的理论估计。

变异函数：变异函数是根据经验人为选定的，它描述了空间随机变量的二阶平稳或内蕴假设，即距离较近的采样点比距离远的采样点更相似。

（4）函数方法

函数方法插值

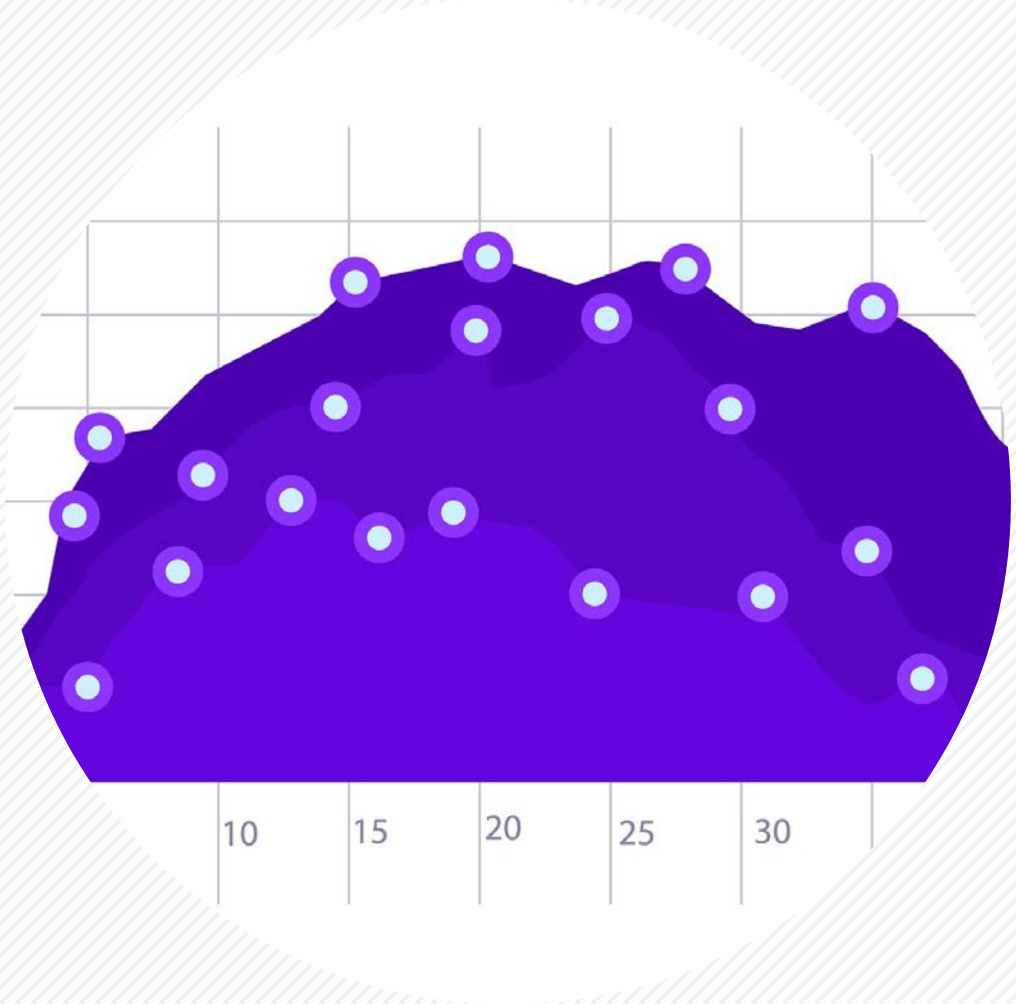
函数方法是一种数学技术，用于估计在采样点之间或采样点之外的未知数据点的值。

函数方法假设

函数方法基于一个假设，即存在一个连续函数，可以近似地描述数据点之间的关系。

函数方法目标

函数方法插值的目标是找到这个函数，以便可以在任何位置估计函数的值。



（4）函数方法

函数方法场合

函数方法在空间内插领域大多用于一些特殊场合，如利用高密度的高程数据产生等高线、为提高格网数据的空间分辨率而内插数据等。

函数方法缺点

函数方法大多不适合用于缺值预测和内插格网，因为它难以满足内插的精度，也难以估计误差。

函数方法特点

函数方法的特点是不需要对空间结构的预先估计、不需要做统计假设；缺点是难以对误差进行估计，点稀时效果不好。

函数方法常见方法

函数方法常见的方法有傅里叶级数、样条函数、径向基函数、双线性内插、立方卷积法等。

(5) 随机模拟方法

01

随机模拟方法定义：随机模拟是指把某一现实的或抽象的系统的某种特征或部分状态，用另一系统（称为模拟模型）来代替或模拟。

02

随机模拟方法概率模型：为了解决某问题，把它变成一个概率模型的求解问题，然后产生符合模型的大量随机数，对产生的随机数进行分析从而求解问题。

03

随机模拟方法基本假设：随机模拟方法的基本假设与空间统计方法不同，随机模拟认为地理空间具有非平稳性，是空间异质的。

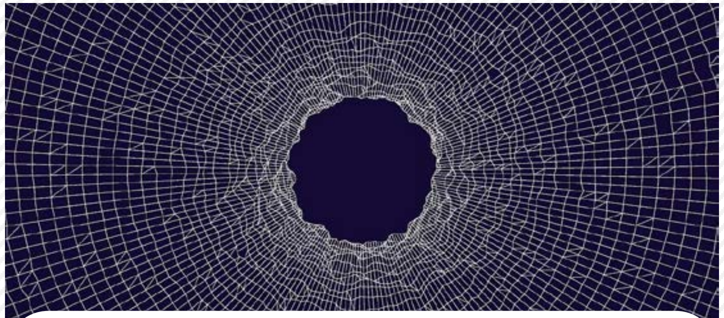
04

随机模拟方法优点：随机模拟的优点是定义了各种随机变量之间的空间相关，这类相关可以根据相邻数据把高度不确定性的先验分布更新为低不确定性的后验分布。

05

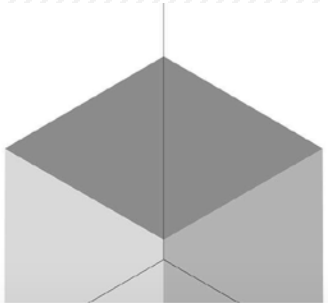
随机模拟方法缺点：随机模拟的缺点是建模困难，计算量大，常用的随机模拟方法有高斯过程、马尔科夫过程、蒙特卡罗方法、人工神经网络方法等。

（6）确定性模拟



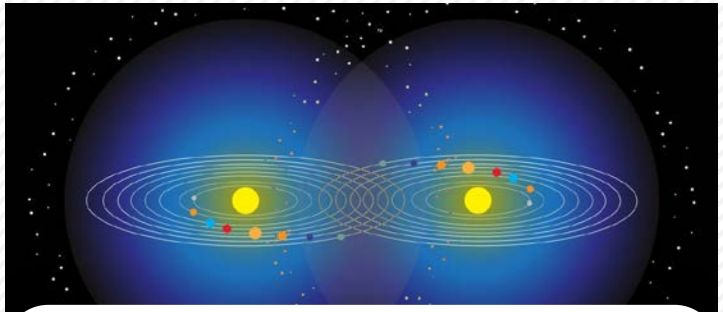
确定性模拟定义

确定性模拟是指假设变量的空间分布受物理定律控制，因此可以使用物理模型或半经验、半物理的模型模拟空间分布。



确定性模拟经验参数

对于这一类插值方法，常常是使用有限的观测值获得一些必要的经验参数，再把这些参数代入到物理模型之中。



确定性模拟优点

确定性模拟的优点是它的确定性，它不依赖或很少依赖观测样本，但空间准确性有待商榷，需要进一步验证和改进。

4）克立格法插值概述

插值方法偏差

任何一种插值方法都无法完全要求计算的平均品位估计值与实际值完全一样，偏差是不可避免的。

无偏估计

在实际中，常常要求插值方法满足所有估计块段的实际值与其估计值之间的偏差平均为0，即估计误差的期望等于0。

最小方差

块段的估计品位与实际品位之间的单个偏差应该尽可能小，即误差平方的期望值应该尽可能小。

4）克立格法插值概述

克立格法介绍

克立格（Kriging）插值法是空间统计分析方法的重要内容之一，是对有限区域内的区域化变量取值进行无偏最优估计的一种方法。

考虑距离关系

克立格法不仅考虑了待预测点与邻近样点数据的空间距离关系，还考虑了各参与预测的样点之间的位置关系。

利用分布特征

充分利用了各样点数据的空间分布结构特征，使其估计结果比传统方法更精确，更符合实际，更有效地避免了系统误差的出现。

4）克立格法插值概述

01

线性组合估计

对于任意待估计点的估计值均可以通过待估测点范围内的n个观测样本值Z(xi) 的线性组合得到。

02

权系数影响

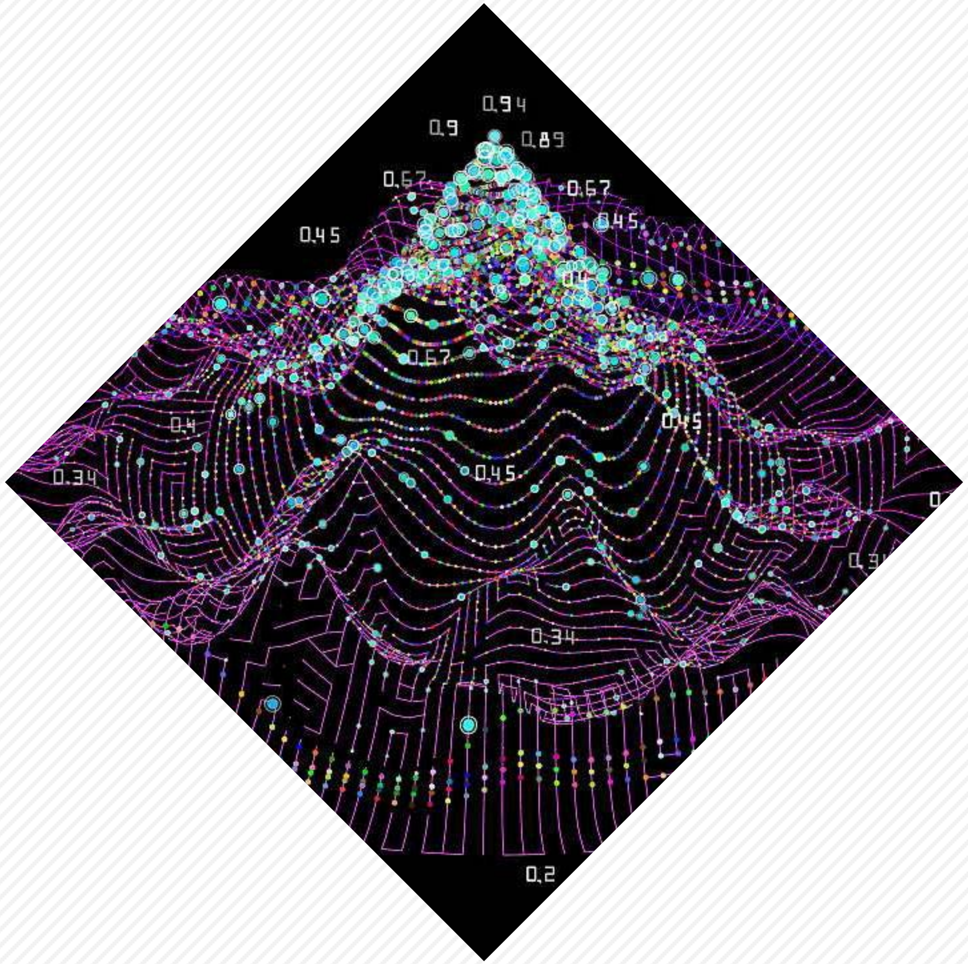
内插估计值的精度取决于权重的求解，不同的克立格方法采用不同的克立格方程组求解。

03

半变异图权系数

克立格法与确定性插值法一样，都是从预测点周围的观测值中生成权系数进行预测。

4）克立格法插值概述



不同方法权系数

克立格法中观测点的权系数更为复杂，是通过计算反映数据空间结构的半变异图得到的。

未知点预测

运用克立格法可以在研究邻域中观测值的半变异图和空间分布的基础上对研究区中未知点的值进行预测。

常见克立格方法

常见的克立格插值方法有普通克立格、简单克立格、泛克立格、概率克立格、指示克立格、析取克立格及协同克立格等。

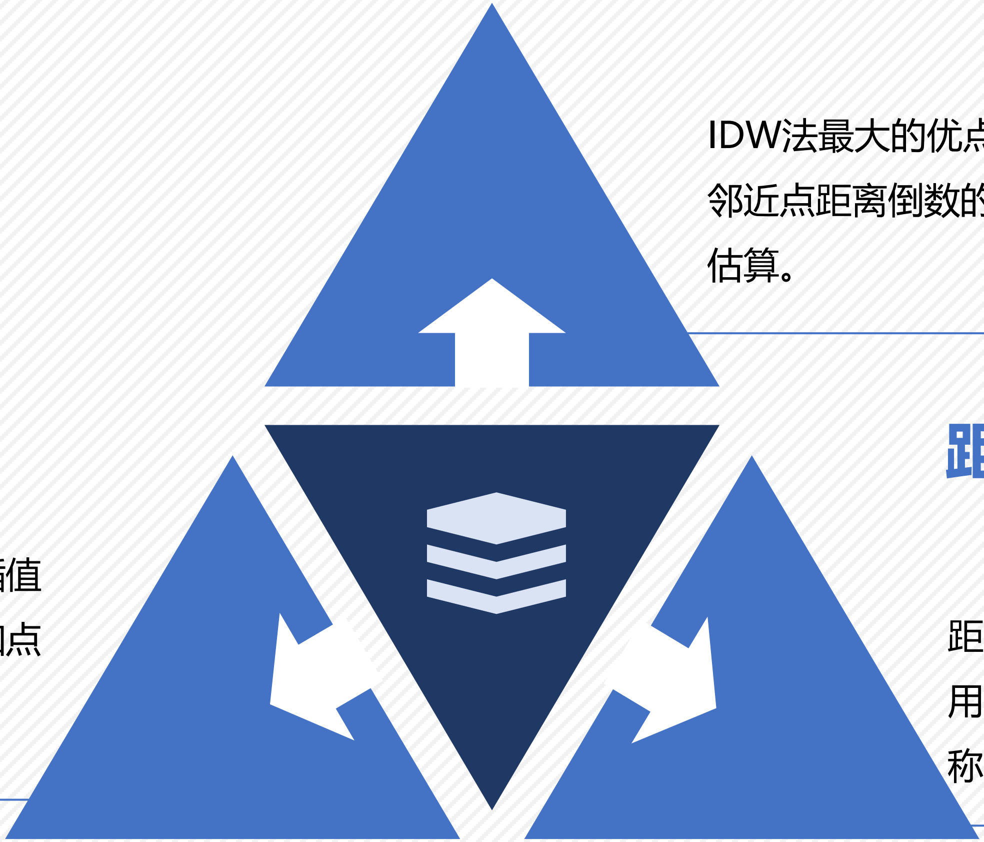
5）距离幂次反比法插值

$$z^*(x)=\frac{\sum_{i=1}^{N}[\lambda_i(x)z(x_i)]}{\sum_{i=1}^{N}\lambda_i(x)}$$

$$\lambda_i(x)=\frac{1}{d(x,x_i)^p}$$

距离幂次反比法（IDW法）

广泛应用于地球科学领域，是一种多元插值方法，通过已知空间散乱点的值计算未知点的值。



简单性

IDW法最大的优点是简单性，未知点的值只是一个邻近点距离倒数的函数，计算速度快，适合于快速估算。

距离未知点影响

距离幂次反比法的幂次不同，则有不同的适用范围和插值效果，当幂次为2时，该方法称为“距离平方反比法”。

5.3.3 克立格插值方法

克立格插值方法

克立格插值方法是一种数学方法，用于在给定数据点周围估计未知的数据。

估计未知数据

通过将已知数据点连接起来，形成一个或多个曲线，可以估计未知的数据点。

曲线拟合

克立格插值方法在科学和工程领域中广泛使用，用于估计曲线拟合、异常值处理等任务。

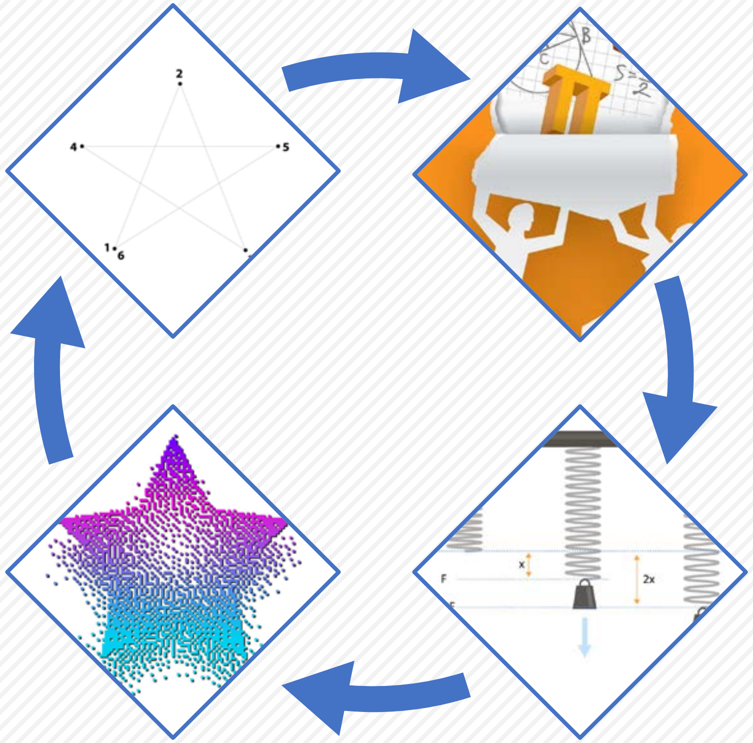
1）普通克立格（Ordinary Kringing）

普通克立格法

普通克立格法是实践中应用最广泛的克立格方法，它适用于已知分布的随机变量，且不要求期望值已知。

无偏条件

为了使估计无偏，我们需要使式(4-8)成立，这是无偏条件，避免了系统误差。



估计方差

估计方差为最小值，表示估计值的可靠性最高，因为它的波动性最小。

误差的期望值

误差的期望值为0，表示估计值与真实值之间的差距在统计上没有系统性的偏差。

1）普通克立格（Ordinary Kringing）

$$[A]=\begin{bmatrix}\gamma(x_1,x_1) & \gamma(x_1,x_2) & \cdots & \gamma(x_1,x_N) & 1\\ \gamma(x_2,x_1) & \gamma(x_2,x_2) & \cdots & \gamma(x_2,x_N) & 1\\ \vdots & \vdots & & \vdots & \vdots\\ \gamma(x_N,x_1) & \gamma(x_N,x_2) & \cdots & \gamma(x_N,x_N) & 1\\ 1 & 1 & \cdots & 1 & 0\end{bmatrix},$$

$$[\lambda]=\begin{bmatrix}\lambda_1\\ \lambda_2\\ \vdots\\ \lambda_N\\ \mu\end{bmatrix},\quad [b]=\begin{bmatrix}\gamma(x_1,x_0)\\ \gamma(x_2,x_0)\\ \vdots\\ \gamma(x_N,x_0)\\ 1\end{bmatrix}$$

权重系数和拉格朗日参数，

$$[\lambda]=\left[A^{-1}\right][b]$$

克立格方差按下式计算，

$$\sigma^{2*}(x_0)=\left[b^T\right][\lambda]$$

2）简单克立格(Simple Kriging)

变简单克立格

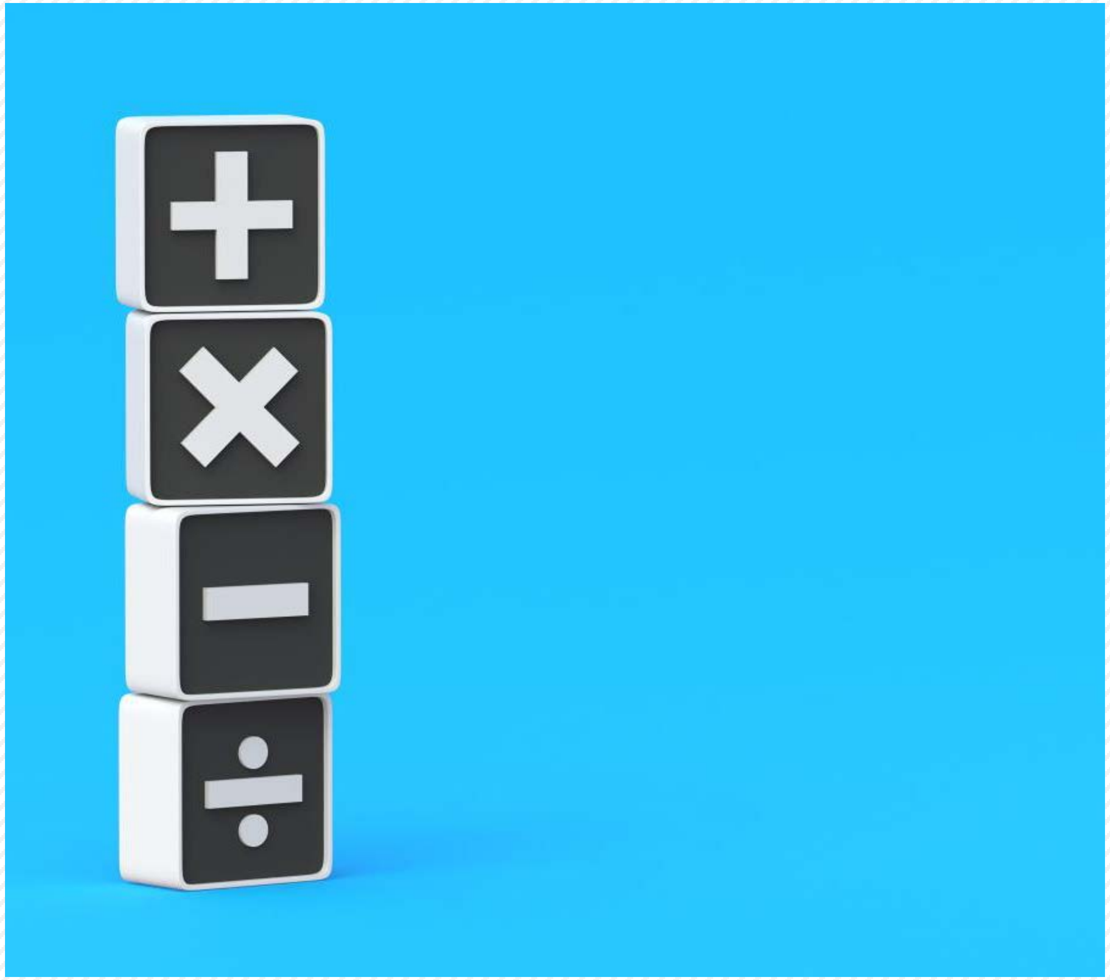
简单克立格是一种空间插值方法，它假定随机变量的均值为已知，区域化变量满足二阶平稳假设，即变异函数要有上限。

$$[A_{SK}]=[\lambda_{SK}][b_{SK}]$$

$$[A_{SK}]=\begin{bmatrix} C(x_1,x_1) & C(x_1,x_2) & \cdots & C(x_1,x_N) \\ C(x_2,x_1) & C(x_2,x_2) & \cdots & C(x_2,x_N) \\ \vdots & \vdots & & \vdots \\ C(x_N,x_1) & C(x_N,x_2) & \cdots & C(x_N,x_N) \end{bmatrix},$$

$$[\lambda_{SK}]=\begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_N \end{bmatrix}, \quad [b_{SK}]=\begin{bmatrix} C(x_1,x_0) \\ C(x_2,x_0) \\ \vdots \\ C(x_N,x_0) \end{bmatrix}$$

$$Z^*_{SK}(x_0)=\sum_{i=1}^{N}\lambda_i z(x_i)+\left\{1-\sum_{i=1}^{N}\lambda_i\right\}m$$



3）泛克立格（Universal Kriging）

中南大學 CENTRAL SOUTH UNIVERSITY

泛克立格法

泛克立格法是在漂移和非平稳随机函数的协方差或变异函数为已知的条件下，一种考虑到有漂移的无偏线性估计量的地质统计学方法。

漂移现象

在实际研究中，区域化变量Z(x)常常呈现出所谓的“漂移”现象，即自西向东或自东向西逐渐升高，但在小的局部范围内又可以认为是平稳的。

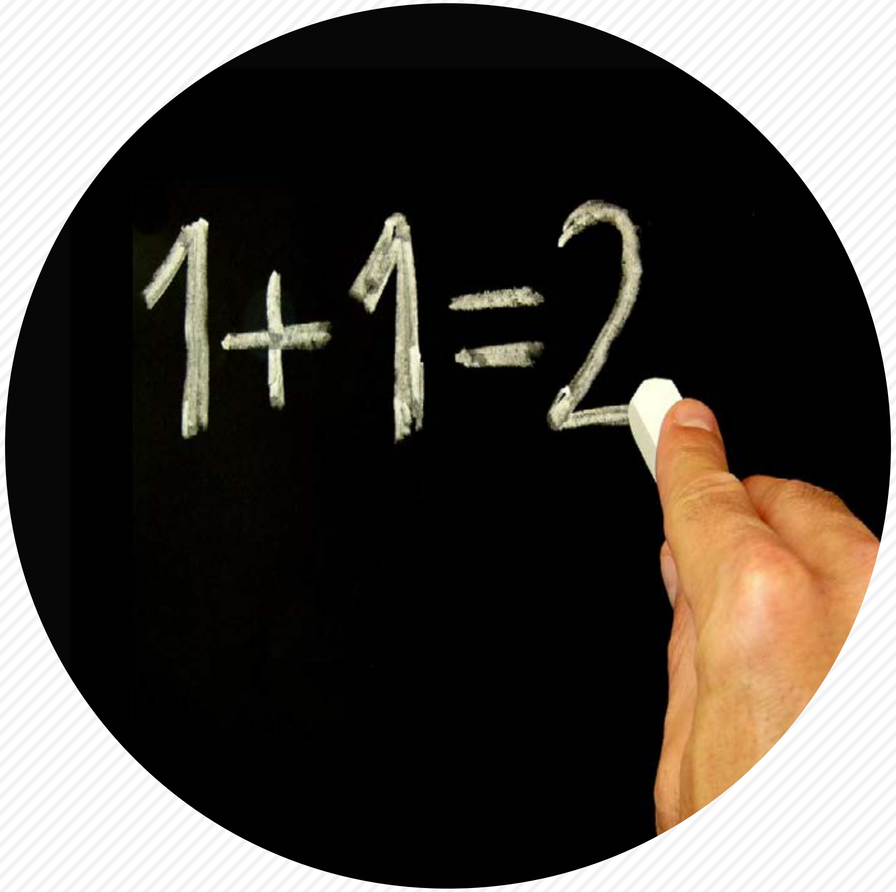
平稳区域化变量

从整体上看，区域化变量是平稳的，但某一小的局部范围却呈现漂移的现象，即具有平稳的数学期望E[Z(x)] = m(x)。

泛克立格法适用性

当估计领域内的有效数据不足以应用普通克立格法进行局部估计时，就必须考虑漂移的存在，采用泛克立格法进行估计。

3）泛克立格（Universal Kriging）



漂移和涨落

具有漂移的数据可以分解为两个部分，分别为点x处的漂移和涨落（也称为波动）。

线性漂移

漂移一般用多项式表示，当为线性漂移时，在一维条件下表示为$mx+b$，在二维条件下表示为$m1x+m2y+b$，在三维条件下表示为$m1x+m2y+m3z+b$。

漂移多项式

在采矿实践中，使用线性漂移就足够了，因为二次及以上的漂移在实践中较为罕见，且计算复杂度较高，无实际意义。

4）指示克立格（Indicator Kriging）

指示克立格法

指示克立格法是一种非参数地质统计学方法，用于处理地质、物化探数据处理及矿产储量计算中的特异值和不同矿化作用类型。

边界品位

指示克立格法根据一系列临界值（如边界品位z）对原始数据进行转换，然后求变异函数和进行克立格估值。

5）协同克立格（Co-Kriging）

在地质、采矿及其他自然现象的研究中，多个变量之间可能存在空间相关性，即一个变量的取样量不足以获得所需精度的估计量。

协同克里格法就是通过研究主变量与次级变量之间的空间相互关系，借助次级变量的样品信息以提高对主变量的估计精度。

协同克立格法 → 空间相关性 → 次级变量 → 估计精度

通过研究主变量与次级变量之间的空间相互关系，借助次级变量的样品信息以提高对主变量的估计精度。

当某一个变量的取样量不足以获得所需精度的估计量，而其他变量却有较充足的取样量时，如果前者和后者存在空间相关性。

5.4 探索性空间数据分析

5.4.1 探索性空间数据分析的基本理论

探索性空间数据分析

探索性空间数据分析（ESDA）是研究地区社会经济发展空间分布特征的基本统计方法。

空间关联测度

ESDA以空间关联测度为核心，通过对某事物或现象的空间分布的可视化分析，从而发现其空间关联性和聚集性。

空间数据探索

在地理信息系统（GIS）的帮助下，ESDA可以更有效地进行空间数据的分析和解释。

空间可视化工具

ESDA使用可视化工具，如地图和图表，来揭示空间数据的分布模式、趋势和关联。

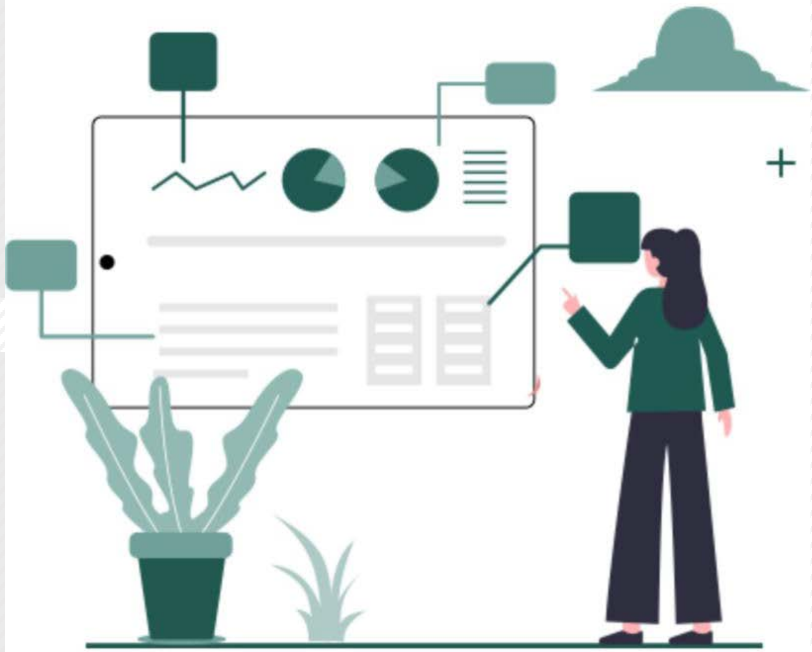
1）空间自相关

空间自相关概念

空间自相关是探索性空间数据分析的核心概念，描述了空间中相邻区域之间的数据之间的相似性或相关性。

空间自相关衡量

空间自相关可以用全局空间自相关和局部空间自相关两种方法来衡量。



2）全局空间自相关

全局空间自相关衡量

全局空间自相关是对整个研究区域中所有观测点的空间自相关的衡量。

常用指标

常用的全局空间自相关指标有莫兰指数（Moran's I）和赛斯-恩格尔-莫兰指数（Geary's C）。

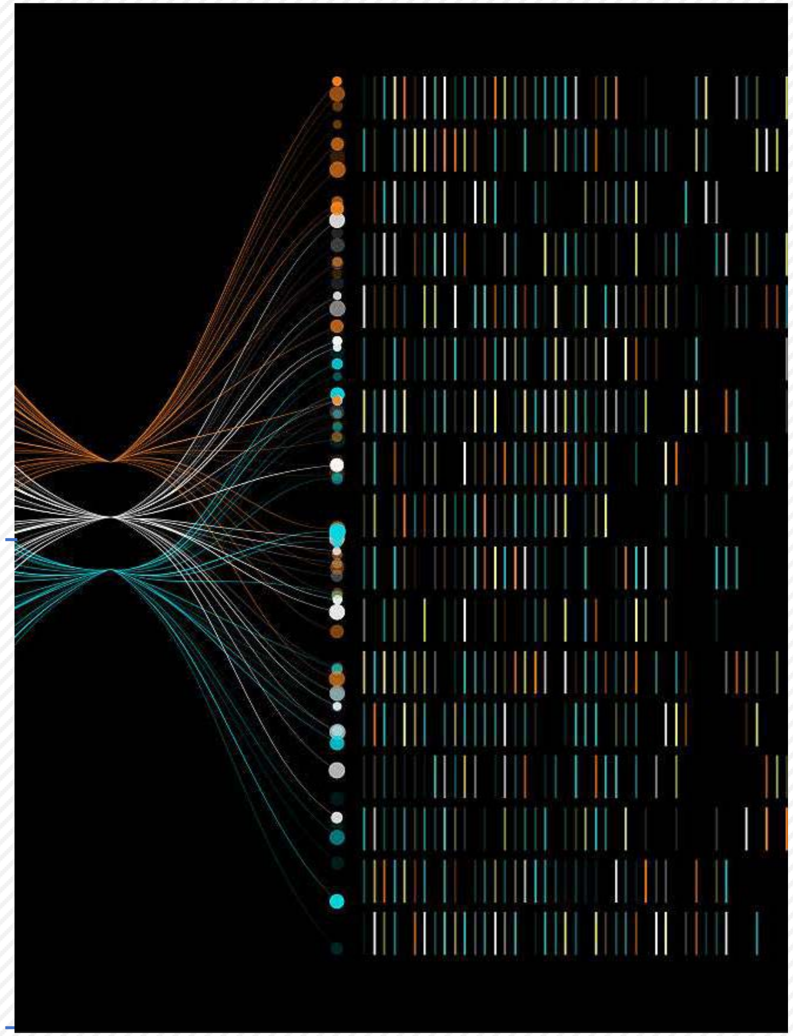
3）局部空间自相关

局部空间自相关衡量

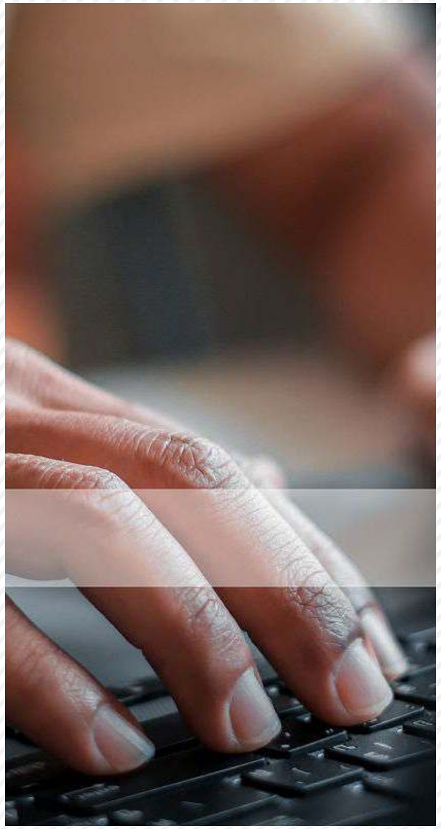
局部空间自相关是对研究区域中局部地区观测点的空间自相关的衡量。

局部空间自相关指标

常用的局部空间自相关指标有局部莫兰指数（Local Moran's I）和局部赛斯-恩格尔-莫兰指数（Local Geary's C）。



4）空间权重矩阵



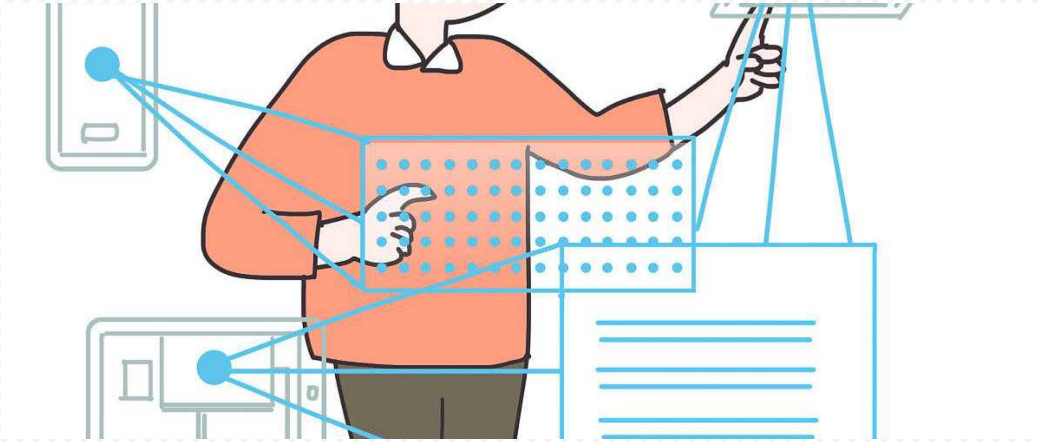
空间权重矩阵概念

空间权重矩阵是衡量空间数据点之间的相似性和距离的矩阵，用于描述空间数据点的邻近关系和相互作用。

空间权重矩阵类型

空间权重矩阵包括基于距离的权重矩阵、基于连接的权重矩阵和基于属性的权重矩阵等类型。

5）空间滞后变量



空间滞后变量概念

空间滞后变量是将一个区域的数据与相邻区域的数据进行比较，以发现空间自相关的变量。

空间滞后变量类型

空间滞后变量包括全局空间滞后变量和局部空间滞后变量，用于描述空间数据的全局和局部自相关性。

6）空间统计量

中南大學 CENTRAL SOUTH UNIVERSITY

空间统计量概念

空间统计量是用来描述空间数据分布特征的一组统计指标，包括集中趋势、离散程度、形态和相关性等。

空间统计量类型

常用的空间统计量包括平均值、中位数、方差、标准差、极差、峰度和偏度等，用于全面了解数据的特征。

空间可视化

空间可视化是将空间数据以图形或图表的形式呈现出来，以便更好地理解空间数据的分布特征和变化趋势。

空间可视化方法

常用的空间可视化方法包括地图、散点图、柱状图和热力图等，可以清晰地展示出空间数据的分布情况。

5.4.2 探索性空间数据分析的基本理论 1）直方图

直方图用途

直方图通常用于统计学、数据分析、机器学习等领域，适用于对大量样品数据进行整理加工，找出其统计规律。

直方图原理

直方图是一种通过分段计数的方式，将数据的分布情况以图形的方式呈现出来，从而帮助我们了解数据的分布情况。

直方图组成

在直方图中，数据被分段成若干个连续的区间，每个区间的中点代表该区间的平均值，区间的高度代表该区间的频率。

直方图功能

通过观察直方图，我们可以了解数据的分布情况，例如数据的集中趋势、离散程度、偏态等，从而对数据进行进一步的分析和处理。

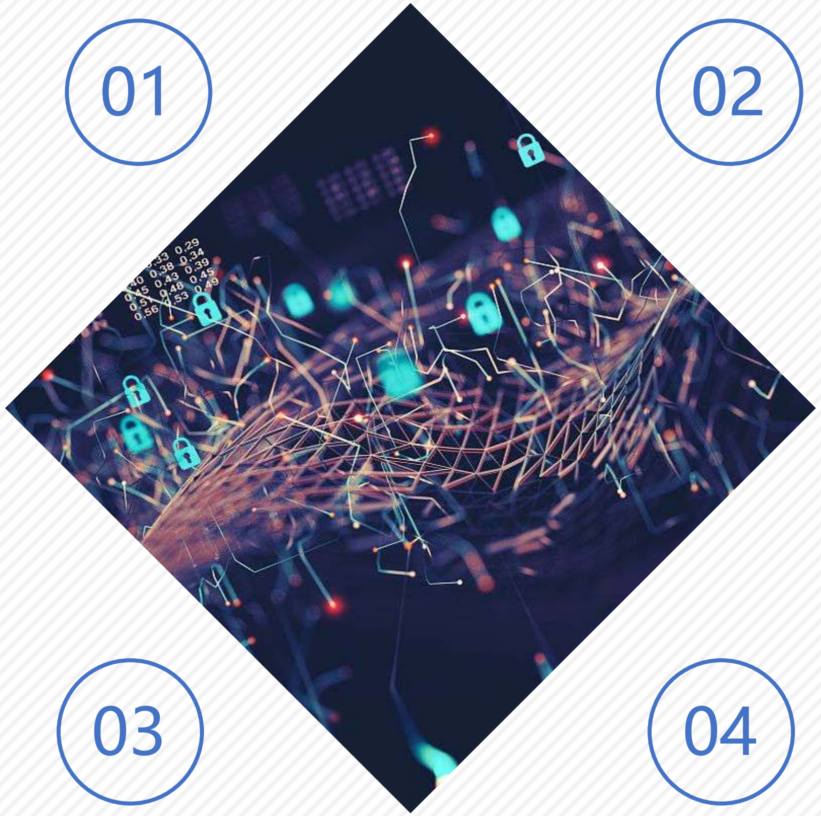
2）Q-Q概率图

01 Q-Q图定义

Q(Quantile-Quantile)图是一种用于检验数据是否符合某个理论分布的统计图。

02 Q-Q图绘制步骤

将数据按升序排列，计算每个数据点的累积分布函数（CDF）的值和对应于标准正态分布的理论分位数。



03 Q-Q图作用

绘制数据点的坐标，x轴为理论分位数，y轴为样本分位数，直观判断数据是否服从某个特定分布。

04 Q-Q图局限性

Q-Q图只能用于检验数据是否符合正态分布，不能用于检验数据是否符合其他特定的理论分布。

3）P-P图

中南大學 CENTRAL SOUTH UNIVERSITY

P-P图定义

P-P(Probability-Probability)图是一种用于检验样本分布与理论分布是否一致的统计图。

P-P图绘制步骤

将数据按升序排列，计算每个数据点的累积分布函数（CDF）的值和对应于理论分布的理论累积分布函数值。

P-P图作用

绘制数据点的坐标，x轴为理论累积分布函数值，y轴为样本累积分布函数值，判断数据是否符合指定的概率分布。

P-P图与Q-Q图区别

Q-Q图对应的是分位数，而P-P图对应的是概率值，两者在检验方法上存在差异。

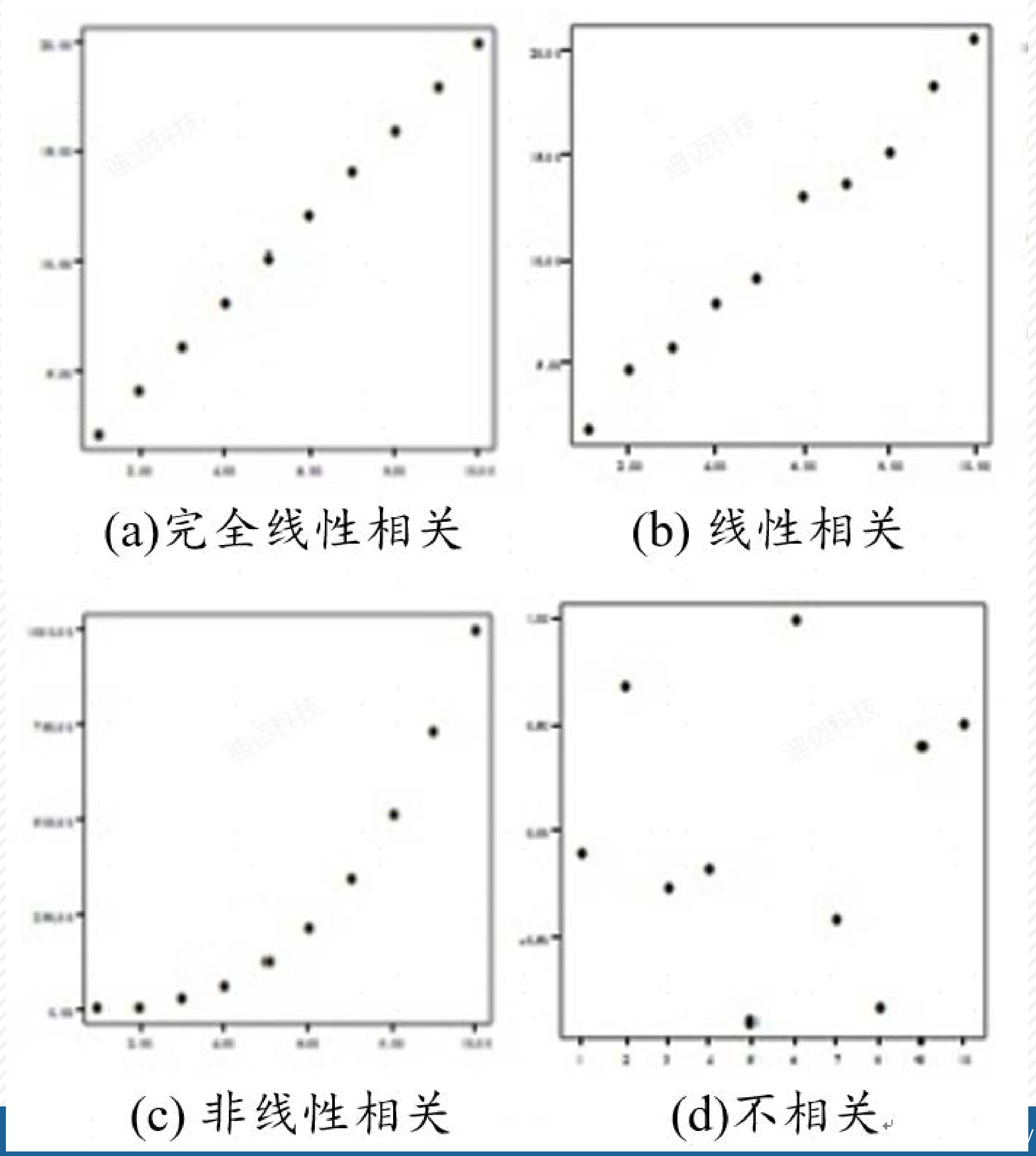
4）散点图

01

散点图定义：散点图是表示两个变量之间关系的图，又称相关图。用于分析两组数据值之间相关关系，它有直观简便的优点。

02

观察数据相关性：通过散点图对数据的相关性进行直观的观察，不但可以得到定性的结论，而且可以通过观察剔除异常数据。



(a)完全线性相关　(b) 线性相关

(c) 非线性相关　(d)不相关

5）频率分布图

数据分级与频率计算

任何一组数据都可以分为多个级别，并且可以计算每个等级内数据的个数。

频率分布组成

对于一个变量按照测量范围进行等宽度分级，统计数据落入各个级别中的个数或占总数据的百分比，这一组频率值组成频率分布。

直方图与数据分布特征

直方图可以直观地反映数据分布特征、总体规律，可以用来检验数据分布形式和寻找数据特异值。

Sfe含量直方图

某矿山Sfe含量的直方图显示，Sfe含量大致服从正态分布。

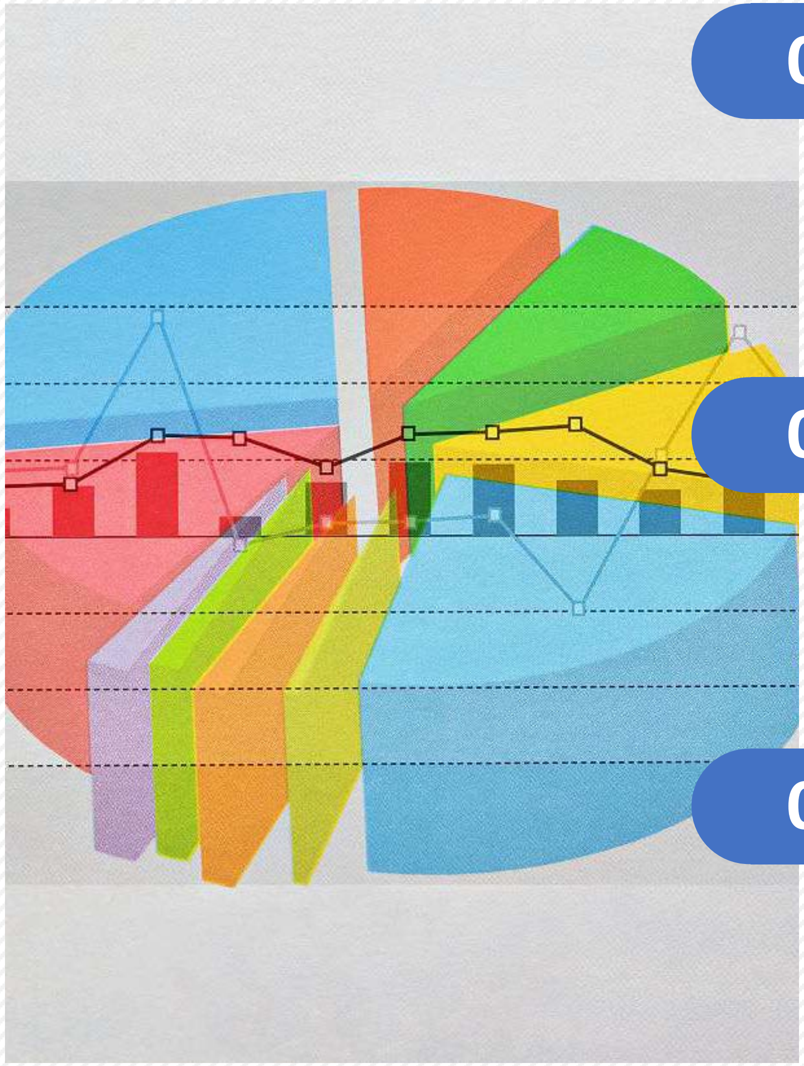
级别数量确定

直方图级别数量的要根据数据个数和数据值的范围来确定。通常情况下，数据个数越少所需级别的数量也越少，才能更好地表示数据。

区间宽度与频率成正比

相同的区间宽度确保每个条带的面积与该级别的频率成正比。

6）趋势分析图



01 样点位置与属性值表示

样点的位置可以在X、Y平面上来表示，对于感兴趣的属性值，则可通过垂直方向上的Z轴来表示，构成三维视图。

02 投影散点图

在进行趋势分析时，将Z轴数据值分别投影到X、Z平面和Y、Z平面作散点图，这也可以被看作是三维数据的侧视图。

03 多项式拟合

用多项式来拟合投影平面上的散点图，如果经过投影后的曲线是平直的，表明没有趋势；如果多项式有确定的形式，如是呈上升趋势的曲线模式，则表明数据中存在全局趋势。

5.3 空间三维分析

5.5.1 地形表面信息计算

距离计算

利用鼠标在DEM模型上选取两个不同点，通过计算连线与模型交点的三维坐标，使用公式计算出空间两点的距离。

面积计算

包括剖面积和表面积，其中剖面积需计算投影面积，表面积则是各个网格的表面积之和，或通过计算斜面面积之和得出。

5.5.2 等值线（面）生成

01

等值线：等值线是连接相邻且具有相同属性值的点的线，包括等高线、等深线、等温线等。在Grid数据和TIN数据中均可以绘制等值线。

02

（1）从Grid数据中绘制等值线：一般要经过三个步骤，包括计算等值线和网格边界交点的坐标值，找到一系列等值点，以及采用合适的样条函数，连接各等值点为一条光滑曲线。

03

（2）从TIN数据中绘制等值线：确定三角形边上存在等值点后，用内插法求得等值点的坐标，然后找出起始等值点，并追踪等值点。

04

等值面：等值面是空间中所有具有相同值的点的集合，可以表示为其中，c是常数。等值面是由许多个等值面片组成的连续曲面，文中在边界体素中生成三角面片，以三角面片拟合等值面。

05

在三维空间规则数据场中构造等值面的方法有很多种，如MC算法，MT算法，剖分立方体方法，立方体方法等，其中最具代表性的是MC方法。在MC算法中，体素是一个逻辑上的立方体，八个数据点位于体素的八个角上。

5.5.3 体积计算

DTM或DEM的体积

由四棱柱与三棱柱体积累加得到，通过抛物双曲面和斜平面拟合计算。

工程量计算

根据公式计算挖方、填方及土壤流失量，模型经过挖或填后，体积通过原始DEM体积减去新DEM体积得到。

TIN封闭模型体积

通过散度定理计算，微分向量域为多面体S的法线向量，曲面S指向外的法线向量表示为。

1）面模型体积计算

01 DTM或DEM的体积

由四棱柱和三棱柱体积累加得到，四棱柱体上表面用抛物双曲面拟合，三棱柱体上表面用斜平面拟合，下表面均为水平面。

02 TIN封闭模型的体积

基于“散度定理”计算，将多面体的体积包围起来，考虑边界曲面S的简单互连区域R，对其中的n个面Si进行积分。

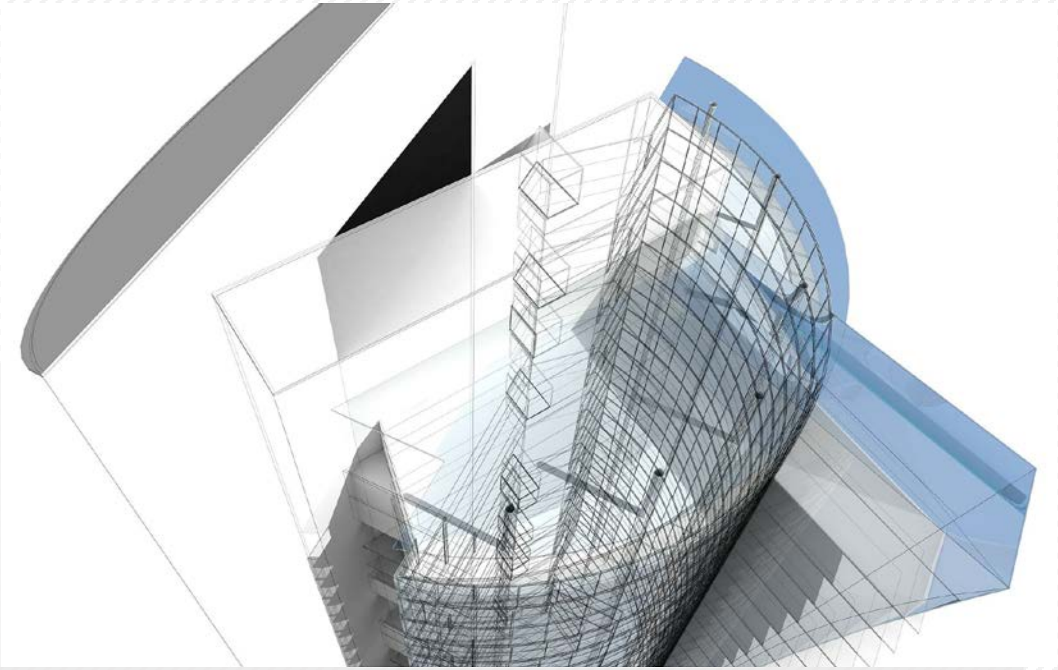
03 构建虚拟四面体计算

对于不满足计算条件的网格模型，如存在相交三角形、无效边等，可将其体素化后再进行体积计算。

04 六面体填充运算

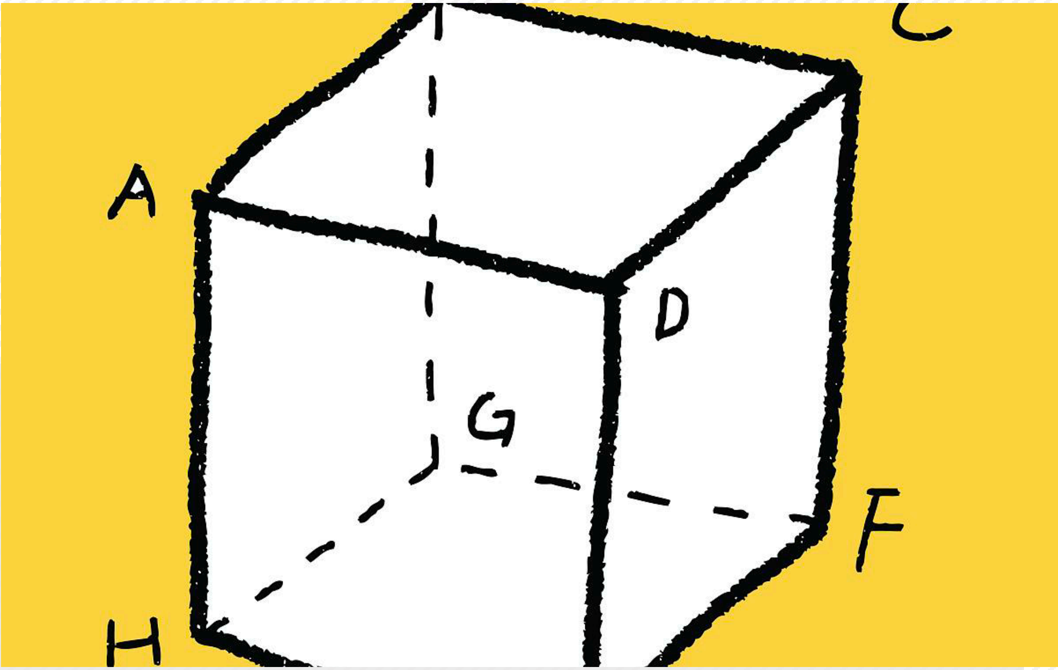
表面模型六面体化算法流程包括求表面模型体的最小外包、六面体剖分、平面切割表面模型形成轮廓线以及栅格扫描法找出处于表面模型内的六面体。

2）体模型体积计算



体模型体积计算

体模型体积计算简单，只需将所有体素的体积相加即可得到原表面模型的体积。



体素个数与体积

每个体素的体积已知，通过求和所有体素的体积，即可得到表面模型的总体积。

5.5.4 三维空间布尔分析

三维空间布尔分析

三维空间布尔分析是在三维空间中对三维对象进行布尔运算，以分析它们之间的关系、生成新的三维对象或进行空间操作的过程。

三维对象的关系

三维空间布尔分析应用于三维建模、计算机辅助设计（CAD）、虚拟现实（VR）、计算机图形学等领域，用于分析三维对象的关系。

布尔运算的过程

常见的三维空间布尔分析包括交集分析、并集分析、差集分析和对称差集分析，根据数据模型的不同可以分为栅格布尔分析和矢量布尔分析。

矢量布尔分析

矢量布尔分析基于三角形网格模型，需要确保网格模型是可定向的流形网格，不符合要求的网格模型可以通过算法进行调整。

布尔运算的步骤

实现三维网格模型的空间布尔运算的基本步骤包括相交测试、建立拓扑关系、形成多边形“结果交域”、三角化多边形和加入结果网格。

布尔运算的说明

对布尔运算实现方法进行具体阐述，并对其中应用到的原理、特殊情况处理、需要注意的问题及所采用的技巧进行详细说明。

Q&A