



上海中学

Shanghai High School



数据结构的提炼与压缩

上海市上海中学 曹钦翔

指导教师：上海市上海中学 毛黎莉





上海中学

Shanghai High School

数据结构的“化繁为简”

- 减少存储规模
- 化简存储结构



- 时空复杂度降低
- 处理方式多样





上海中学

Shanghai High School



“化繁为简”的三种手段

- 提炼：忽略无效信息，减少存储规模
- 压：调整存储方式，化简存储结构
- 缩：合并重复信息，减少存储规模





上海中学

Shanghai High School



1. 二维结构的化简

- 问题一：Ural 1568 Train car sorting
- 问题描述：对于一个序列 $\{a_n\}$ ，定义一种操作，将 $\{a_n\}$ 分成两个子列，把其中一个置于另一个前面，得到一个新的序列。现给出一个序列（这个序列是 1 到 n 的一个排列），求一种方案，通过最少的操作次数使它变成升序序列。

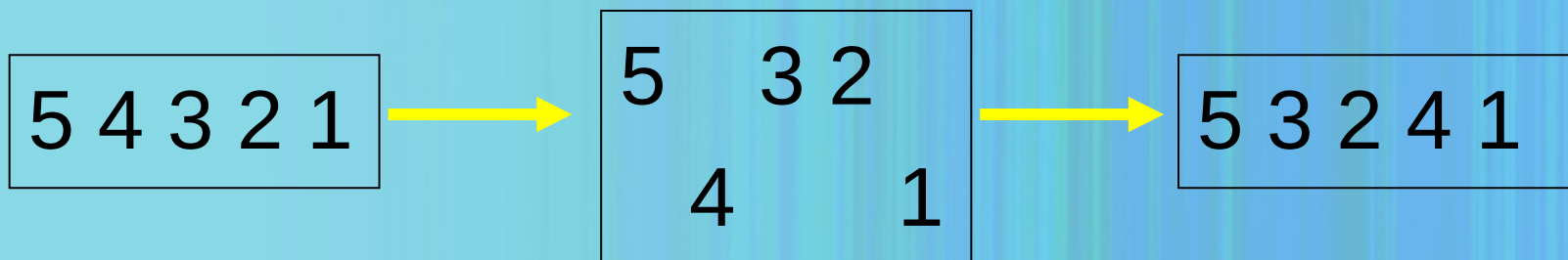




上海中学

Shanghai High School

一个操作的例子



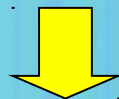


上海中学

Shanghai High School

算法

(5 , 3 , 2 , 4
 , 1)



3 , 4
 , 1

偶数行

5 0

3 4

2 0

0 并

(3 , 4 , 1 , 5
 , 2)

奇数行

5 , 2





上海中学

Shanghai High School

优化数据结构

朴素实现，单次操作复杂度 $O(n^2)$ 。需要优化



零元素过多，形成冗余



提炼：忽略零元素





上海中学

Shanghai High School

问题二：CEOI 2007 Day 2 Necklace

- 问题描述：要求编译一个库，能够对若干已知的整数串进行两个操作：
- （1）在某个已知串的左端或右端增加或减少一个元素，得到一个新的已知的串。
- （2）输出某个已知串的最左端或最右端的数。
- 在问题的一开始，只有一个已知的串：空串。





上海中学

Shanghai High School

分析

朴素方法：每个串分别储存——复杂度过高，难以承受



大量重复信息，空间严重浪费



缩：合并重复信息





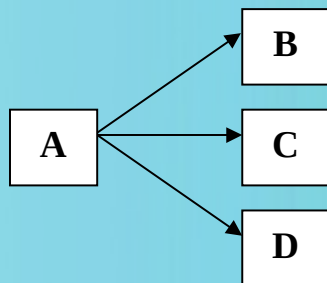
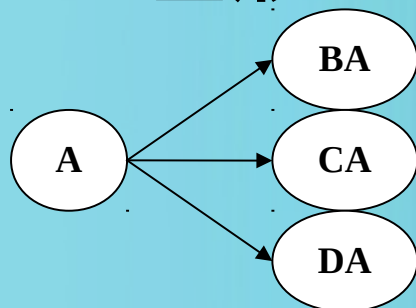
上海中学

Shanghai High School

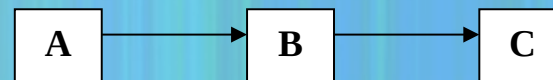
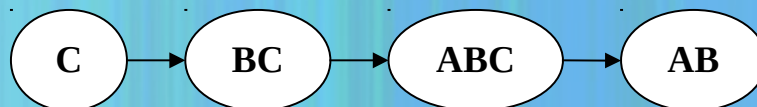
两个特例

特例
存储方式

星形



链形

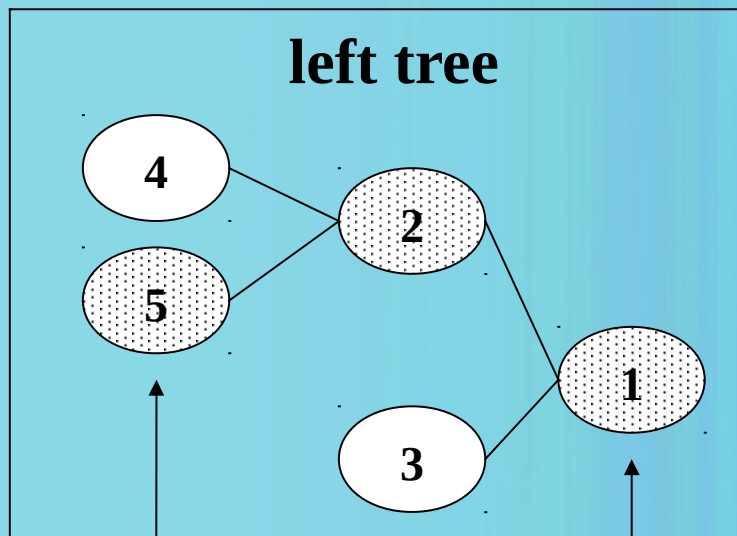




上海中学

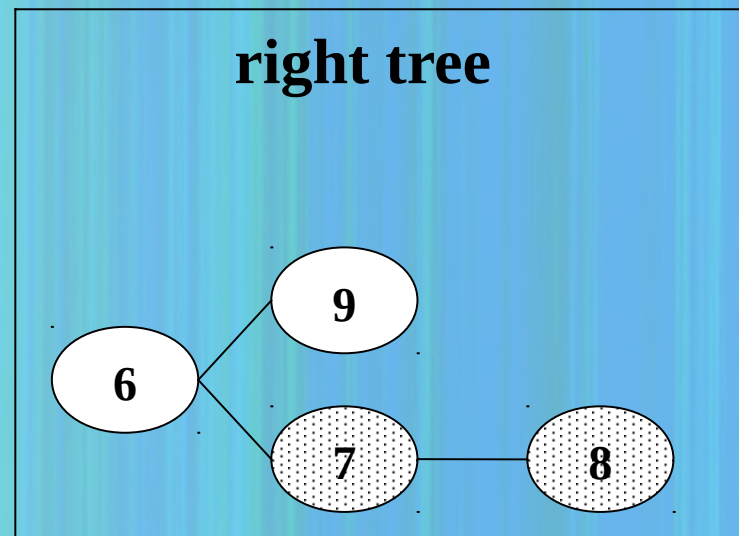
Shanghai High School

数据结构 : Left-Right Tree



left_leave

left_root



right_root

right_leave

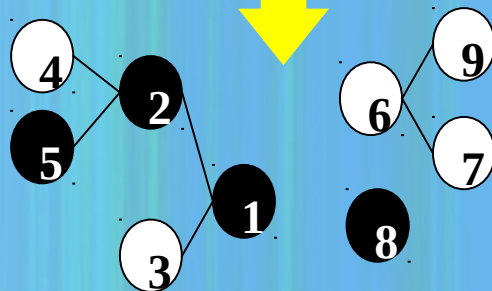
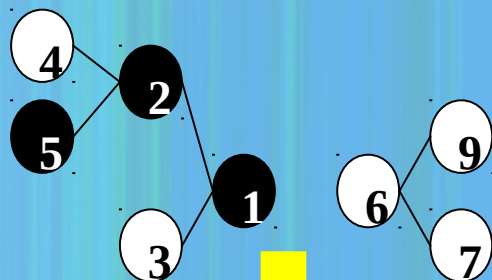
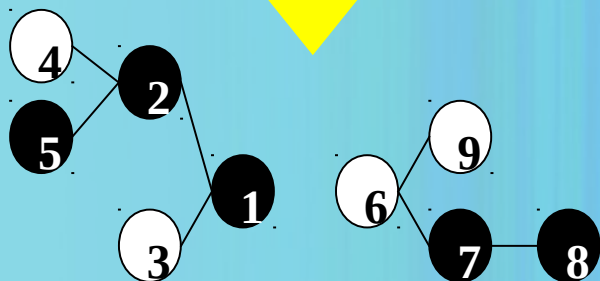
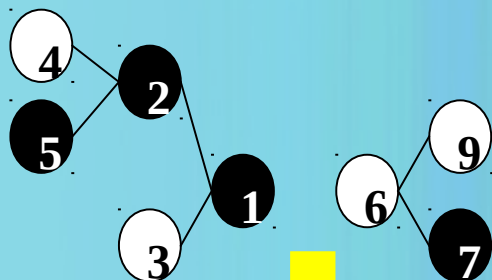




上海中学

Shanghai High School

添加新结点

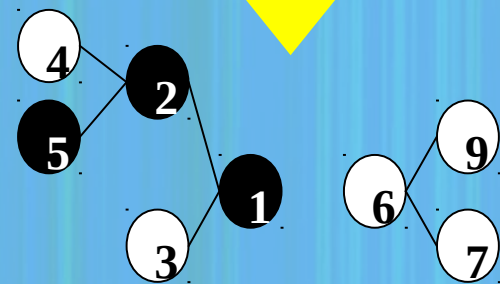
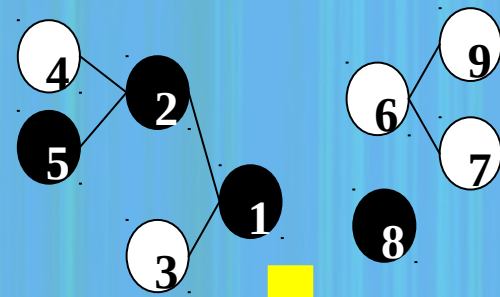
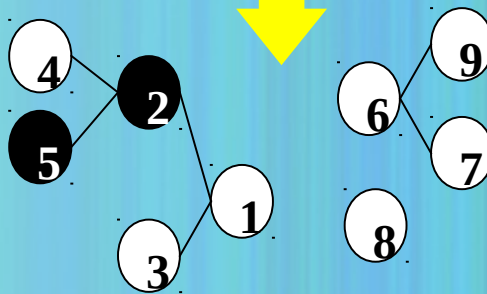
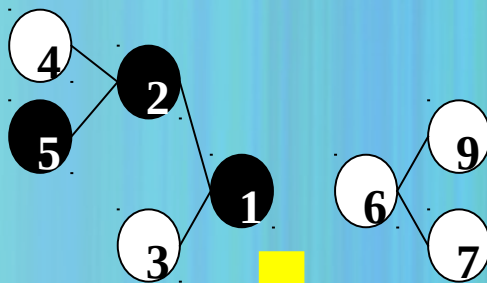
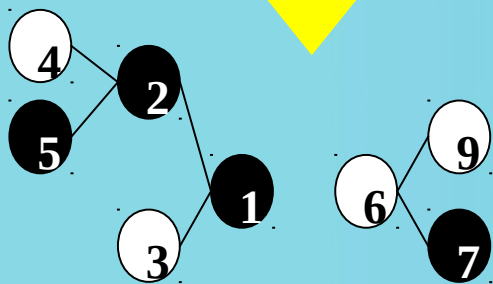
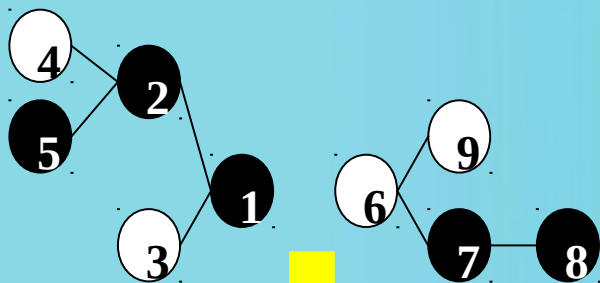




上海中学

Shanghai High School

删除结点





上海中学

Shanghai High School

转化结论

- 已知树中某链的两端点，求底部端点的父亲
- 已知树中某链的两端点，求顶部端点在链中的儿子





上海中学

Shanghai High School



2. 树形结构的化简

- 问题五：问题二的遗留问题
- 问题描述：给定一棵有根树，在线回答两种询问：
 - (1) 已知树中某链的两端点，求底部端点的父亲
 - (2) 已知树中某链的两端点，求顶部端点在链中的儿子





上海中学

Shanghai High School

分析

朴素方法：左儿子右兄弟

最坏情况下时间效率差

转化：顶结点的儿子→底结点的超级祖先

提炼：后代信息成为冗余，不再储存



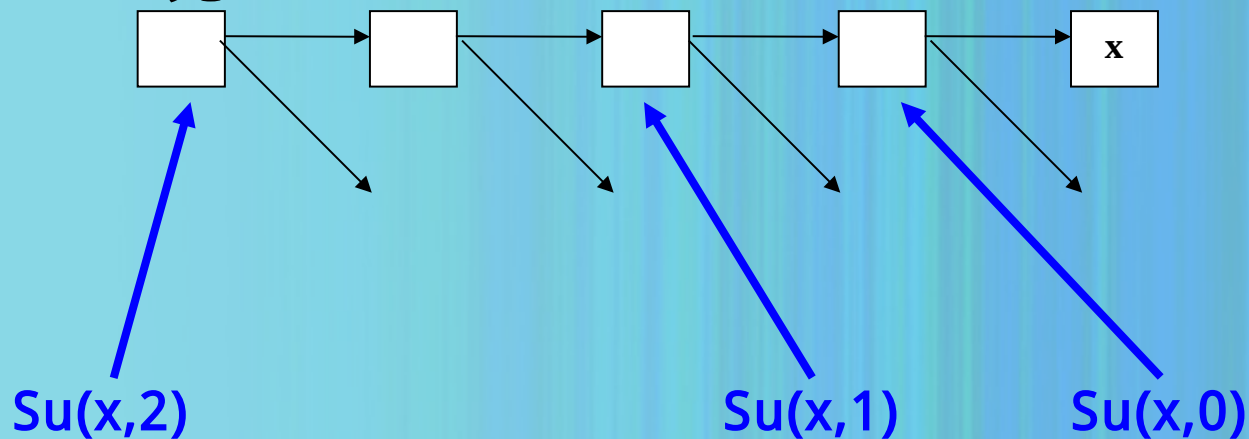


上海中学

Shanghai High School

数据结构 : Supper Father

用 $Su(x, k)$ 表示结点 x 的第 2^k 代祖先





上海中学

Shanghai High School

3. 图结构的化简

- 问题六：ural 1557 Network Attack
- 问题描述：给定一个无向连通图，若从中删去两条边能使它不连通，求所有这样的方案的总数。图点数 n 边数 m 。





上海中学

Shanghai High School

分析

核心问题：图结构复杂不易处理

关键信息：连通性



压：以图的 DFS 树为解题突破口



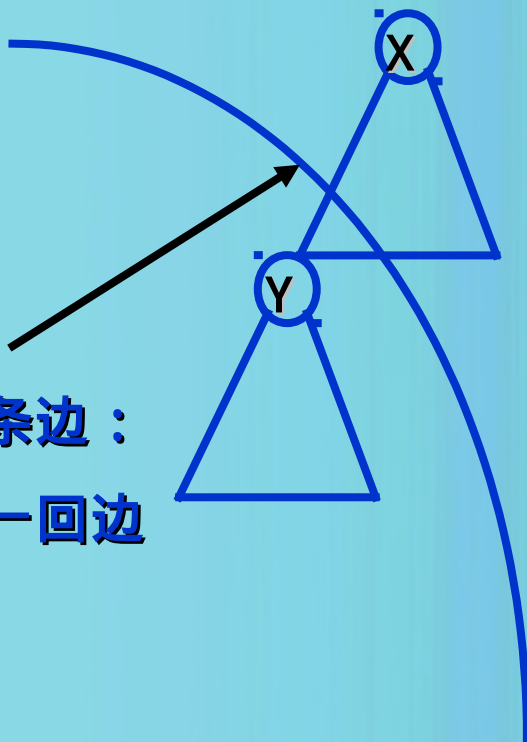


上海中学

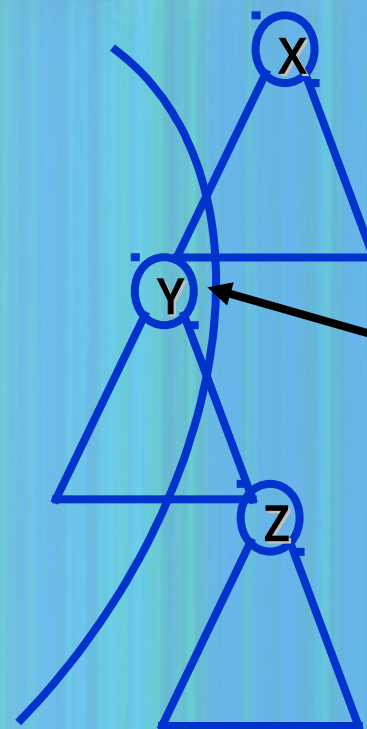
Shanghai High School

两种情况

只有两条边：
一树边一回边



只有两条边：
两条树边





上海中学

Shanghai High School

小结

因题而易，用好“三大手段”

提炼：忽略无效信息，减少存储规模

压：调整存储方式，化简存储结构

缩：合并重复信息，减少存储规模





上海中学

Shanghai High School



谢谢





上海中学

Shanghai High School



算法

- 原序列的最简母矩阵中偶数行的非零元素，形成一个子列，前置。
- 原序列的最简母矩阵中奇数行的非零元素，形成一个子列，后置。
- 形成新序列。
- 不断重复，直到的到升序列。





上海中学

Shanghai High School



2. 树形结构的化简

- 问题三：浙江 2007 年省选 捉迷藏
- 问题描述：给定一棵树，每个节点要么是黑色，要么是白色，能执行一个操作：把某一个点取反色。动态维护并返回树中距离最远的黑色点对。





上海中学

Shanghai High School

分析

问题类型：局部参数调整 + 整体属性返回



解题障碍：树形结构不易整体处理



压：线性结构存储“点对距离”

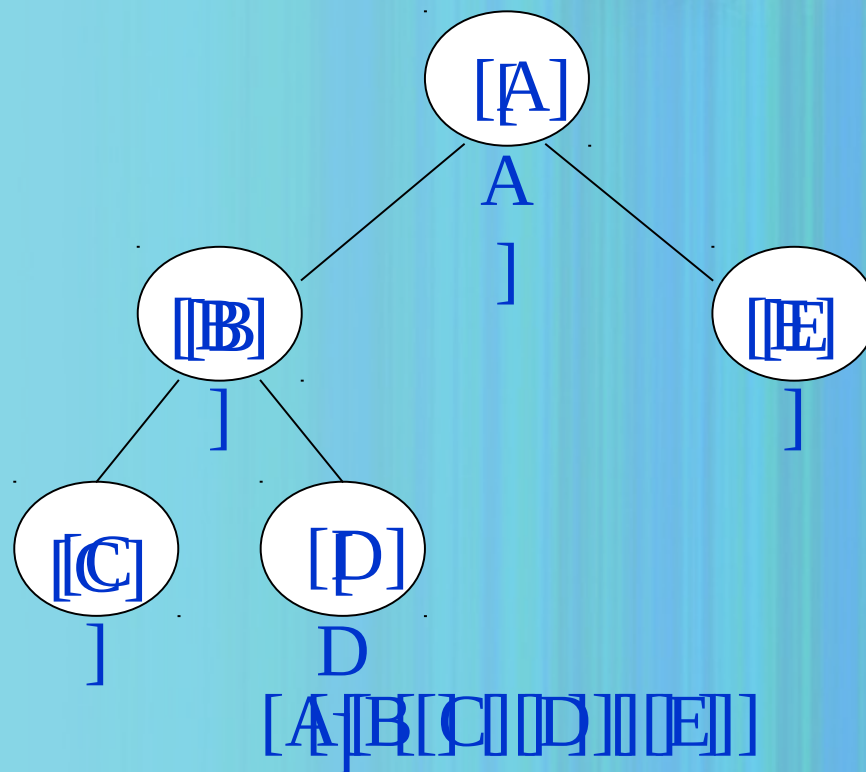




上海中学

Shanghai High School

数据结构：括号编码





上海中学

Shanghai High School

数据结构：括号编码

存储方式：]] () [— — >]][— — >
(2 , 1)

结论：对于两个点 PQ，如果介于某两点 PQ 之间编码 S 可表示为 (a,b)，PQ 之间的距离就是 $a+b$ 。





上海中学

Shanghai High School

数据结构：括号编码

合并方式： $]][[[+]]][[=]](())]] =]]][[$

$]][[[+]]][[=]](())]] =]]][[$

对于 $(a,b)=(a_1,b_1)+(a_2,b_2)$

当 $a_2 \leq b_1$ 时 $(a_1,b_1) + (a_2,b_2) = (a_1, b_1 + a_2, b_2)$

当 $a_2 > b_1$ 时 $(a_1,b_1) + (a_2,b_2) = (a_1, b_1 + b_2, a_2 + b_2)$



上海中学

Shanghai High School



数据结构：括号编码 + 线段树

- 维护以下变量

- $\text{dis}(s)$:
 $\{a+b \mid S'(a,b) \text{ 是 } S \text{ 的一个子串, 且 } S' \text{ 介于两个黑点之间} \}$
- right_plus :
 $\max\{a+b \mid S'(a,b) \text{ 是 } S \text{ 的一个后缀, 且 } S' \text{ 紧接在一个黑点之后} \}$
- right_minus :
 $\max\{a-b \mid S'(a,b) \text{ 是 } S \text{ 的一个后缀, 且 } S' \text{ 紧接在一个黑点之后} \}$
- left_plus :
 $\max\{a+b \mid S'(a,b) \text{ 是 } S \text{ 的一个前缀, 且有一个黑点紧接在 } S \text{ 之后} \}$
- left_minus :
 $\max\{b-a \mid S'(a,b) \text{ 是 } S \text{ 的一个前缀, 且有一个黑点紧接在 } S \text{ 之后} \}$





上海中学

Shanghai High School

数据结构：括号编码 + 线段树

- 利用以下公式实现合并： $S=S1+S2$ ，其中
 $S1(a,b)$ 、 $S2(c,d)$
- $dis(S)=\max\{right_plus(S1)+left_minus(S2), right_minus(S1)+left_plus(S2), dis(S1), dis(S2)\}$
- $right_plus(S)$
 $=\max\{right_plus(S1)-c+d, right_minus(S1)+c+d, right_plus(S2)\}$
- $left_plus(S)$
 $=\max\{left_plus(S2)-b+a, left_minus(S2)+b+a, left_plus(S1)\}$
- $right_minus(S)=\max\{right_minus(S1)+c-d, right_minus(S2)\}$
- $left_minus(S)=\max\{left_minus(S2)+b-a, left_minus(S1)\}$





上海中学

Shanghai High School

问题三小结

关键点 1 : 树形变线形 , 为使用线段树创造条件。

关键点 2 : 数对表编码 , 沟通整体部分关系。





上海中学

Shanghai High School

最简母矩阵

- 称一个 $p \times q$ 的矩阵 A 为序列 $\{a_n\}$ 的母矩阵，当且仅当，矩阵 A 中的所有非零元素，自上到下自左到右逐列读出得到 $\{a_n\}$ ，自左到右自下到上逐行读出得到升序序列。
- 称序列 $\{a_n\}$ 的所有母矩阵中，行数列数都最小的那个矩阵为序列 $\{a_n\}$ 的最简母矩阵。





上海中学

Shanghai High School

最简母矩阵

($\begin{pmatrix} 5 & 3 \\ 5 & 3 \\ 2 & 2 \\ 1 & 1 \end{pmatrix}$, $\begin{pmatrix} 4 \\ 2 \\ 4 \\ 1 \end{pmatrix}$) 的最简母矩阵

$\begin{pmatrix} 5 & 0 \\ 3 & 4 \\ 2 & 0 \\ 0 & 1 \end{pmatrix}$ 顺序排列





上海中学

Shanghai High School



DFS vs BFS

- 其一：性质上（结构决定性质），DFS 遍历后的可以得到 DFS 树，图中的边在 DFS 树中要么是树边要么是回边；而 BFS 遍历后往往得到层状结构，图中的边要么连接同一层中的两个点，要么连接相邻两层的两个点。
- 其二：用途上（性质决定用途），DFS 能较有效解决与连通性相关的问题（因为任意一棵子树只与它的祖先相联）；BFS 能较有效解决与点对距离相关的问题（由前面的性质，两点的距离，与所在层数密切相关）。

