压去冗余 缩得精华

——浅谈信息学竞赛中的“压缩法”

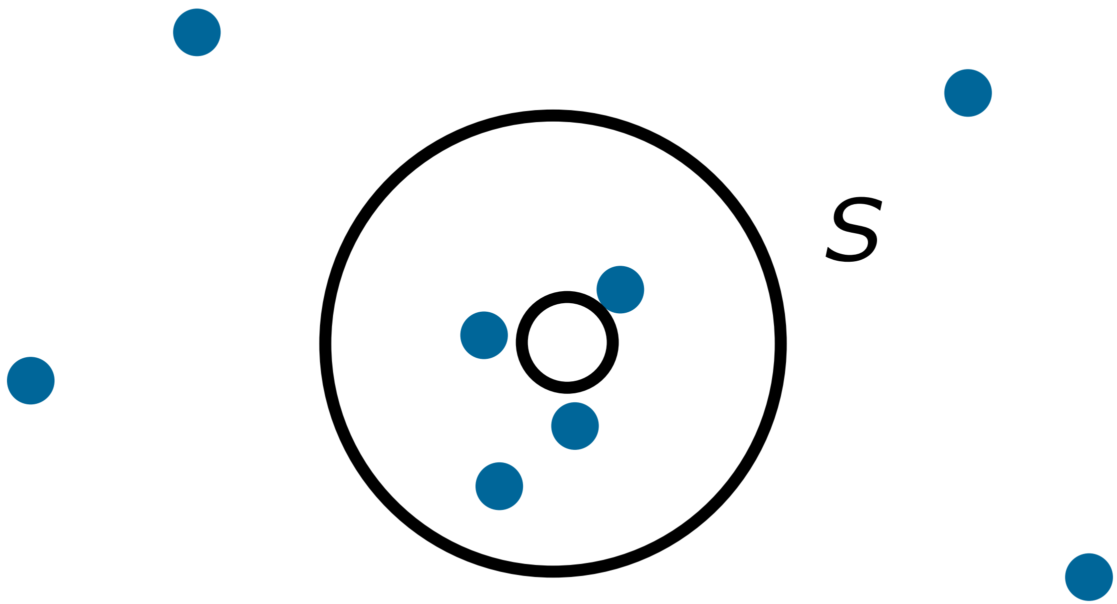
安徽 周源

引子

压缩法？？

- 压缩软件？
  - 本义：加以压力，以减小体积、大小、持续时间、密度和浓度等
- 压缩法
  - 除去冗余信息，留下精华，以减小问题的规模

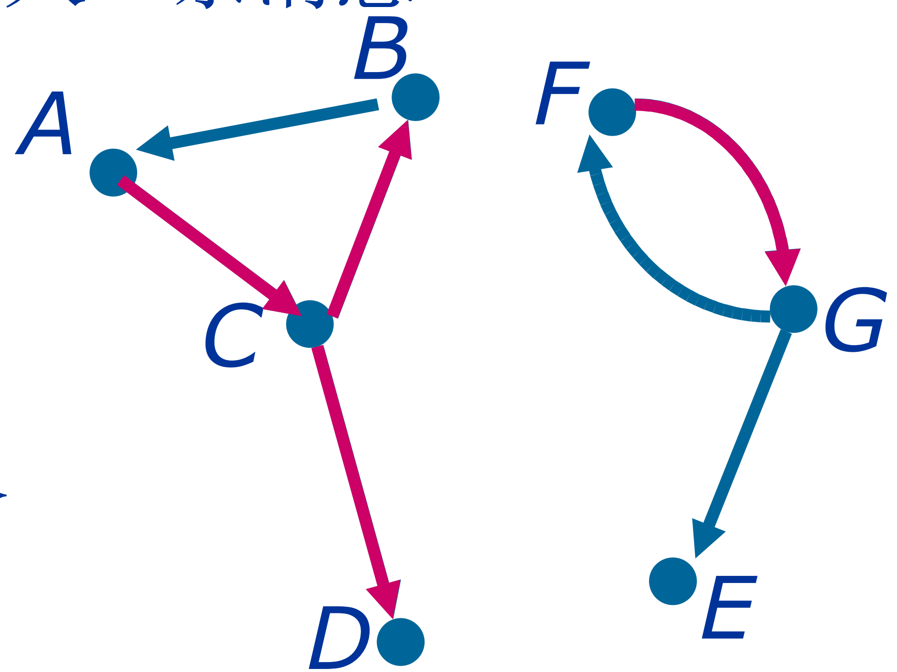
压缩法的定义



- 将集合 S 作为内外隔绝的包裹打包压缩
- 略去了包裹内部的冗余信息，从而简化问题

[例二] 球队问题（经典问题）

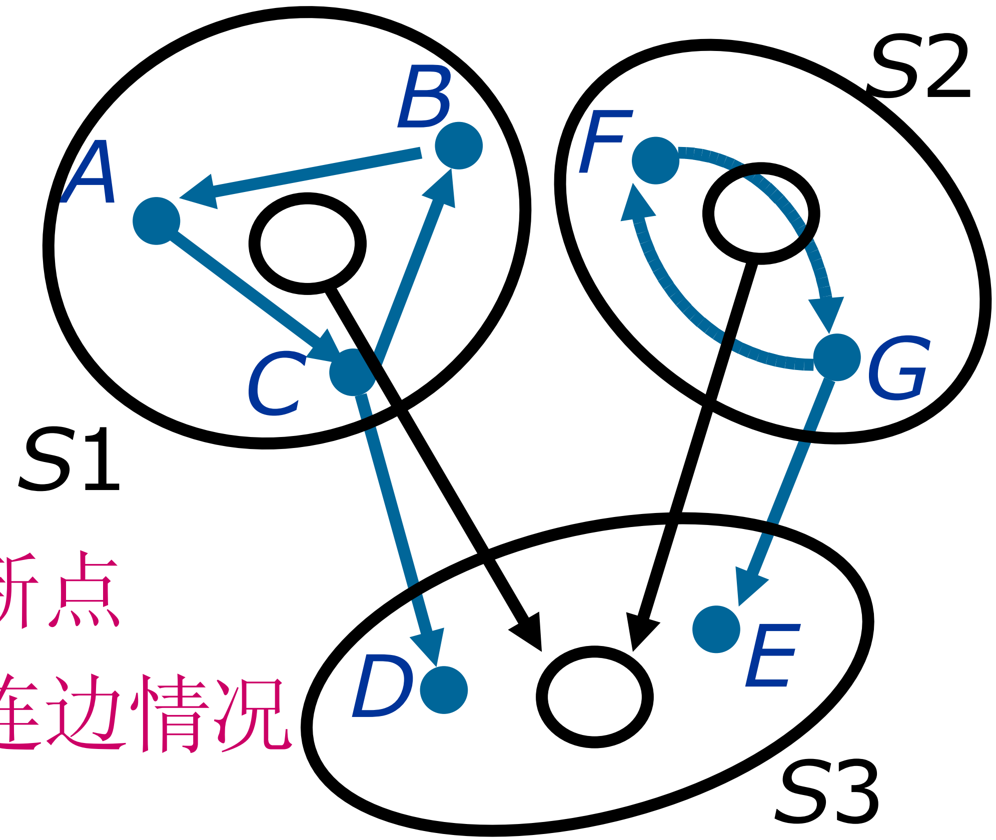
- 篮球队的球员通讯图，(B, A) 表示 B 能通知 A
- 教练需要通知所有人一条消息
- 最少需要亲自告诉几个人？



- 本例中最少为 2 人
  - 如 A, F 等等

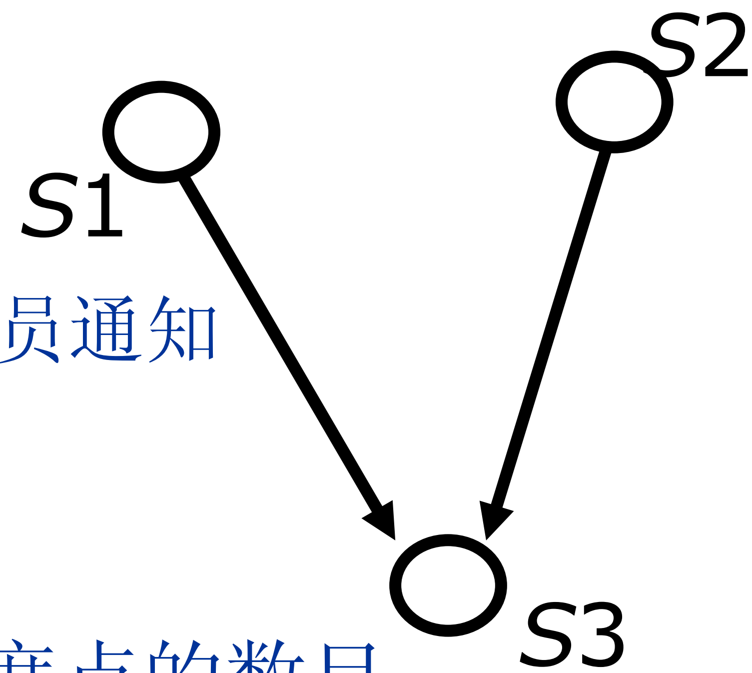
[例二] 球队问题（经典问题）

- 使用压缩法，“压去”不必要的信息
  - 求出强连通分量
    S1, S2, S3…
  - “ 压缩”
    - 每个分量　新点
    - 保留分量间连边情况



[例二]球队问题（经典问题）

- 压缩转化后的新图
  - 简洁：没有环的存在
  - 无入度的点：
    必须亲自通知
  - 有入度的点：
    必定可以由队员通知



- 结论：答案为
  压缩后图中无入度点的数目

压缩法的要点

- 1. 可压缩性
  - 集合 S 内部各个元素之间的联系不会和外部元素相互作用而影响到问题的结果即可压缩
  - [例二] 中：
    - 强连通分量内球员的通讯情况不会影响 球员是否得知消息状态的恒等性
    - 舍去分量内球员的通讯情况，压缩为新的节点

压缩法的要点

- 压：可以压去存在的冗余信息　　（可压缩性）
- 缩：保留 S 作为一个点与外部信息的联系

（替代法则）

- 2. 替代法则
  - 用什么样的形式　表现出精华部分的内容
  - [例二] 中
    - 用一个点替代一个集合
    - 用相同形式的有向边代替原来集合的内外联系

压缩法的应用

- 压缩法在很多竞赛试题中有巧妙的应用
  - [例四]欧元兑换 (BOI 2003 euro)

[例五]合并数列问题 (ZOJ p1794 改编)

  - [例五]合并数列问题 (ZOJ p1794)

[例五] 合并数列问题 (ZOJ p1794 改编)

- 有 *K* 个数列
  - $a_{1,1}, a_{1,2}, a_{1,3}\ldots a_{1,n1}$
  - $a_{2,1}, a_{2,2}, a_{2,3}\ldots a_{2,n2}$
  - …
  - $a_{k,1}, a_{k,2}, a_{k,3}\ldots a_{k,nk}$
- 如 *K*=3 时
  - 3 5 -4 5 -10 6 -3
  - 5 -10 4 -5 15 -20
  - 5 -1 3 -4 2 -6

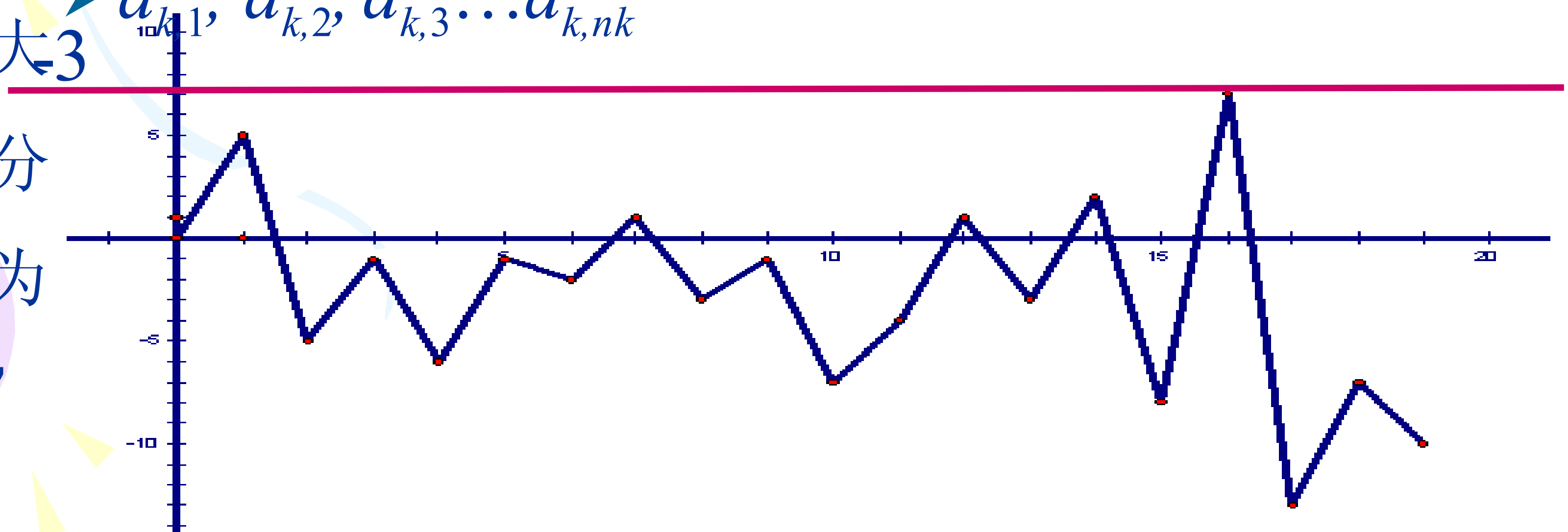
[例五] 合并数列问题 (ZOJ p1794 改编)

- 有 K 个数列
  - $a_{1,1}, a_{1,2}, a_{1,3} \ldots a_{1,n1}$
  - $a_{2,1}, a_{2,2}, a_{2,3} \ldots a_{2,n2}$
  - $a_{k,1}, a_{k,2}, a_{k,3} \ldots a_{k,nk}$
- 如 K=3 时
  - 3 5 -4 5 -10 6 -3
  - 5 -10 4 -5 15 -20
  - 5 -1 3 -4 2 -6 …
- 把这 K 个数列合并成为一个新的数列 S
- 且使得 S 中的部分和中最大的数最小
- 数据范围
  - $K \le 100000$
  - $N = \sum n_i \le 100000$

S : 5 -10 4 -5 5 -1 3 -4 2 -6 3 5 -4 5 -10 15 -20 6

最大部分和为 7

-3



初步分析

- 普通动态规划算法
  - 时间复杂度 $O(K*N^K)$
  - 显然无法承受

使用“压缩法”

观察压缩要点

- 1. 可压缩性
  - 观察：最优串中的一些子串就是原串的子串
  - 反之：若原串的子串满足性质 *P* 即为 *S* 的子串
  - 该子串可压缩，***P*** 就是可压缩性

➢3 5 -4 5 -10 6 -3

➢**5 -10 4 -5** 15 -20

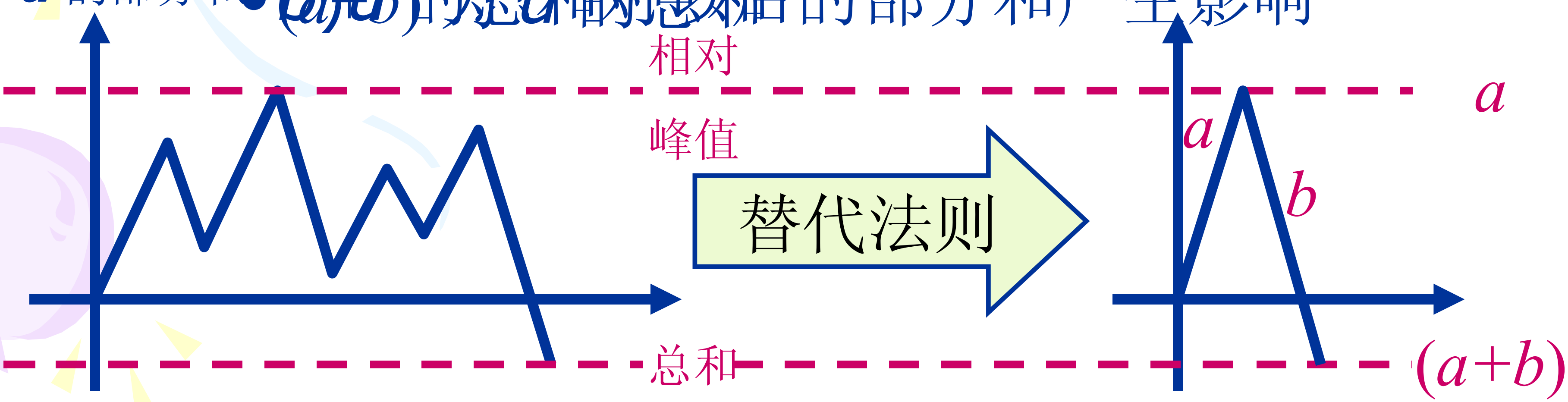
➢**5 -1 3 -4 2 -6**

最优串 *S* : 5 -10 4 -5 5 1 -1 3 -4 2 -6 …

…

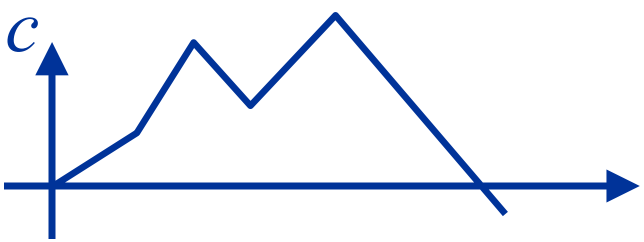
观察压缩要点

- 2. 替代法则
  - 若存在一段可压缩的子串 *u*
  - 替代法则保存 *u* 内元素与外部因素间的联系
  - 用连续的两个数 *a*, *b* 代替 *u*　一"对"(couple)
    - *a*) *u* 的局部峰值（即相对峰值）
    - *a* 为 *u* 的最大部分和，*b* 为非正值使 S 值最大部分和
    - *b*) *u* 的总和对之后的部分和产生影响
    - (*a*+*b*) 为 *u* 的总和



寻找可压缩性：第一阶段压缩

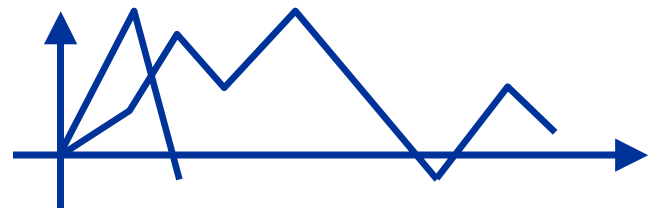
- [定理 5.1]
  - 某子串 $c_1, c_2, \ldots, c_p$ 可压缩当
  - 总和非正　其余部分和为正数



- [证明 5.1]
  - 调整法
  - 若 c 在 S 中不连续出现则可调整
  - 具体方法参见论文

寻找可压缩性：第一阶段压缩

- [定理 5.1]
  - 某子串 $c_1, c_2, \ldots, c_p$ 可压缩当
  - 总和非正　其余部分和为正数
- 压缩转化后
  - 每个数列的前一部分
    - 每一“对”的总和都是负数（或 0 ）
    - 正数后有一个绝对值更大的负数
  - 每个数列的后一部分
    - 一串任意部分和为正的子列
    - 作为对称问题考虑



3 5 -4 5 -10 6 -3

(9 -10)

5 -10 4 -5 15 -20

5 -1 3 -4 2 -6

(7 -8)

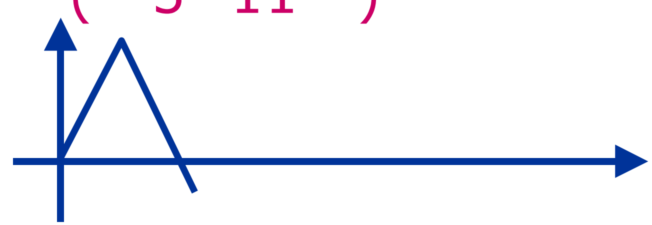
安徽 周源

寻找可压缩性：第二阶段压缩

- [定理 5.2] 和 [推论 5.2.1]
  - 若在一个数列中
  - 存在连续多个“对”
  - 当峰值在第一“对”，则可压缩
- 证明依然使用调整法
- 压缩转化后
  - 每一“对”在数列中的相对峰值递增

3 5 -4 5 -10 6 -3
(9 -10)

5 -10 4 -5 15 -20
(5 -11)



5 -1 3 -4 2 -6
(7 -8)

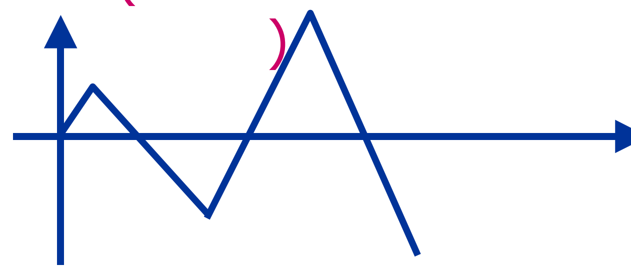
第一阶段压缩成果

最后的贪心

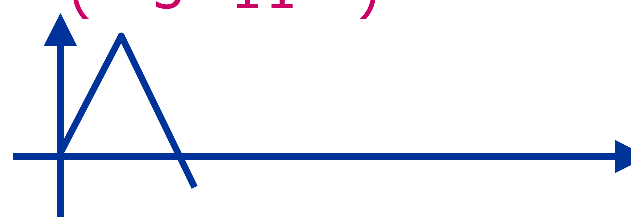
第二阶段压缩成果

- 贪心算法
  - 由于每一“对”的总和为负
  - 相对峰值大的尽量向后放
  - 将每一“对”按相对峰值归并排序即可
  - 时间复杂度：$O(N\log_2 K)$

3 5 -4 5 -10 6 -3
(9 -10)



5 -10 4 -5 15 -20
(5 -11)



5 -1 3 -4 2 -6
(7 -8)

S:

5 -11　7 -8　9 -10　15 -20　6 -3

部分和：

0 5 -6 1 -7 2 -8 **7** -13 -7 -10

[例五] 合并数列问题 (ZOJ p1794 改编)

- 小结
  - 第一阶段压缩：正数后总有绝对值更大的负数
  - 第二阶段压缩：负数后总有绝对值更大的正数
  - 得到精华信息：一个个“对”的相对峰值递增
  - 为贪心法解题创造了良好的条件

总结

化归思想

元法

[例二] 球队问题　$x+\sqrt{x-1}=7$

强连通分量 ⇒ 一个点

复杂的联系图 ⇒ 有向无环图

$\sqrt{x-1} \Rightarrow t$

复杂　简单

无理　有理

[例五] 合并数列问题　$t^2+1+t=7$

一段子串 ⇒ 一个“单峰”

K个任意数列 ⇒ 美妙的单调性

抽象　形象

总结

化归思想

法

化繁

化易

可压缩性：存在冗余

除去冗余

信息化

两个要点

合理利用信息

替代法则：除去冗余

安徽　周源

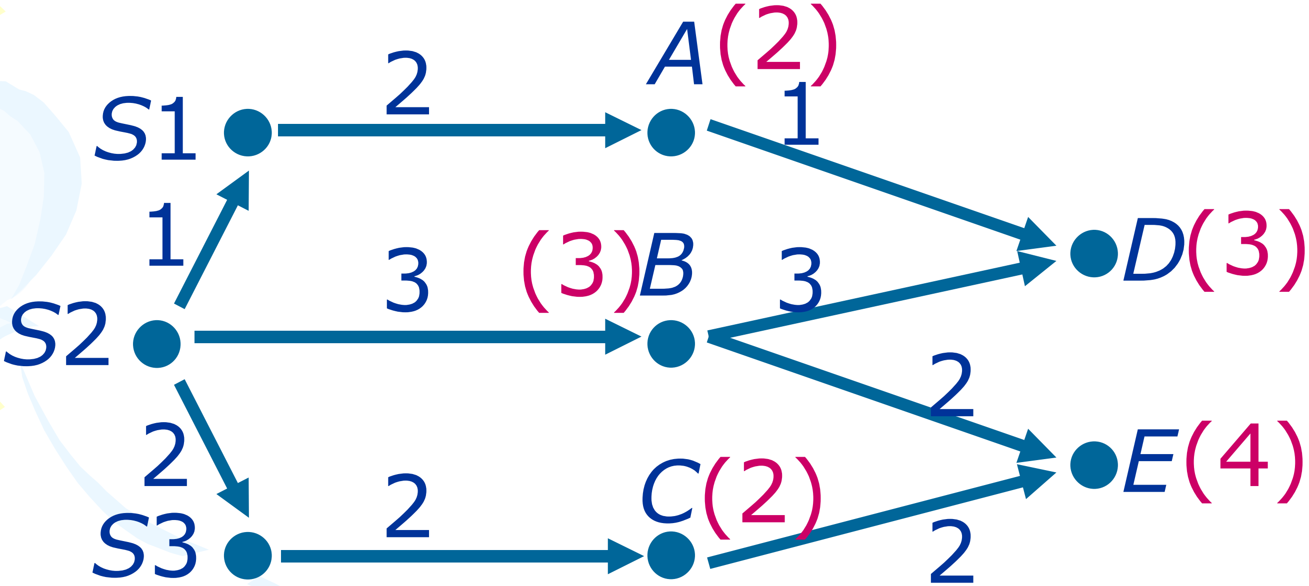
谢谢大家

当题设条件过多、信息量过大，无从下手时

压去冗余　缩得精华

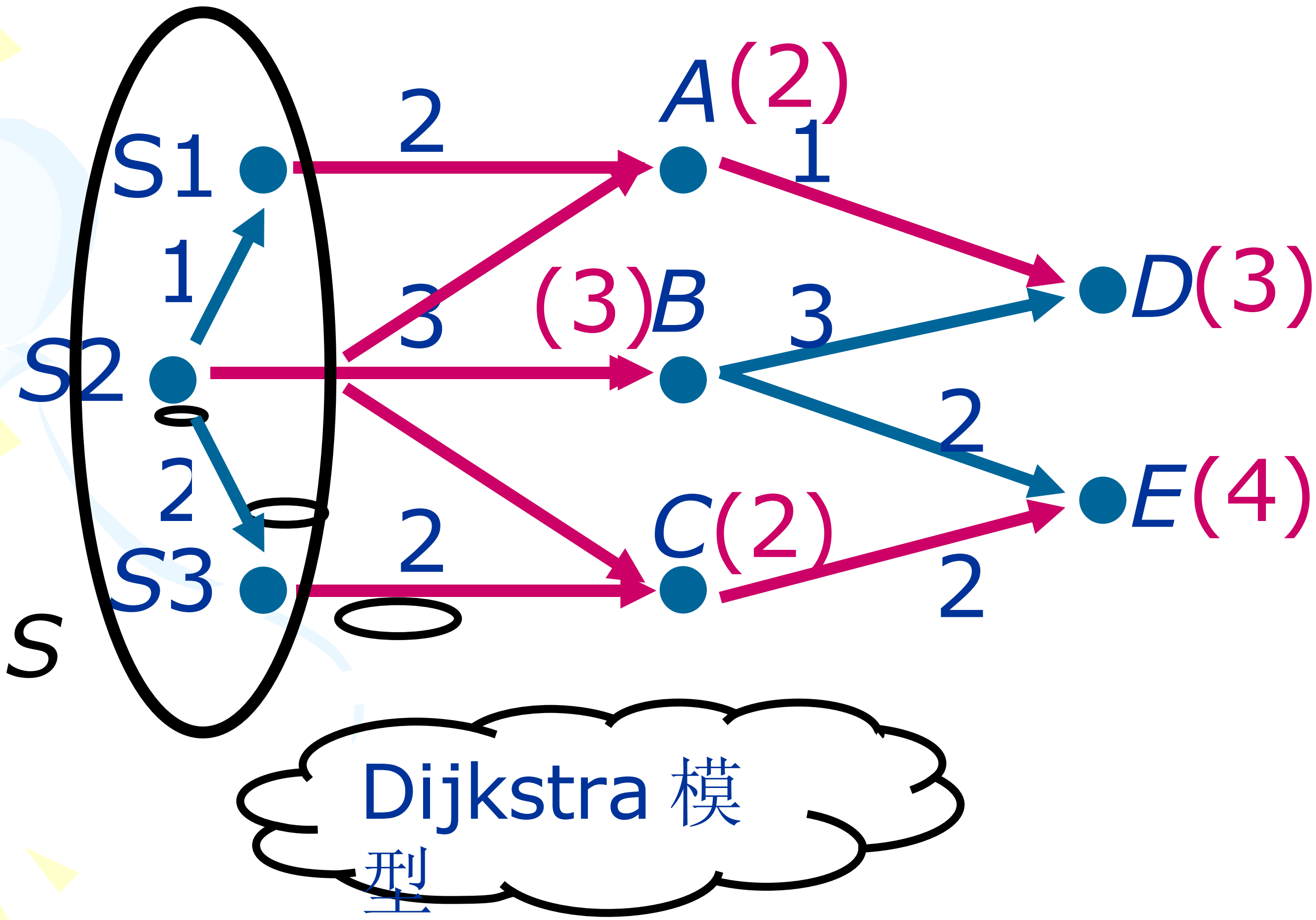
[例一] 多源最短路问题 (经典问题)

- 加权有向图中，有多个源点的最短路问题



[例一] 多源最短路问题 (经典问题)

- 舍去包裹内的连边情况，压缩为新的单源



[例三] 模方程组的替代法则（经典问题）

- 模方程组（红色字母均为未知数）
  - ➢ $a_1x \equiv c_1 \pmod{b_1}$ ➔ $a_1x + k_1b_1 = c_1$
  - – 令 $\Delta_1=(a_1, b_1)b_1$ ➔ $x = x_1 + p_1\Delta_1$ (1)
  - ➢ $a_2x \equiv c_2 \pmod{b_2}$ $\ldots_2x + k_2b_2 = c_2$
  - – 令 $\Delta_2=(a_2, b_2\ldots$ $\ldots = x_2 + p_2\Delta_2$ (2)

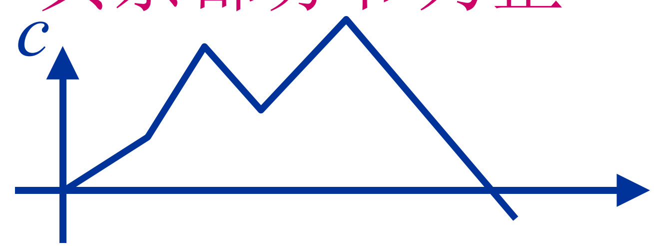
替代法则

  - – 立(1)(2)： $x_1 + p_1\ldots_1 = x_2 + p_2\Delta_2$，解得 x_0
  - – $x = x_0 + k[\Delta_1, \Delta_2]$
  - ➢ $x \equiv x_0 \pmod{[\Delta_1, \Delta_2]}$

安徽　周源

寻找可压缩性：第一阶段压缩

- [定理 5.1]
  - 某子串 $c_1, c_2, \ldots, c_p$ 可压缩当
  - 总和非正　其余部分和为正数



- [证明 5.1]
  - 调整法
  - 若 c 在 S 中不连续出现则可调整