第11讲 NP完全问题 (下)

罗国杰

gluo@pku.edu.cn

2025年春季学期

PKU-04833060

算法设计与分析『实验班』

复习：证明NP完全性

- 待证明：问题 Π=<D,Y> ∈ NPC

- 先证明 $\Pi \in NP$
- 再证明 $\Pi \in$ NP-hard
  - 找某个合适的已知 NPC 问题 Π'=<D',Y'>
  - 构造多项式变换 f: D'→D，使 $I' \in Y'$ 当且仅当 $f(I') \in Y$
  - 从而 $\Pi' \leq_p \Pi$

若干NP完全问题

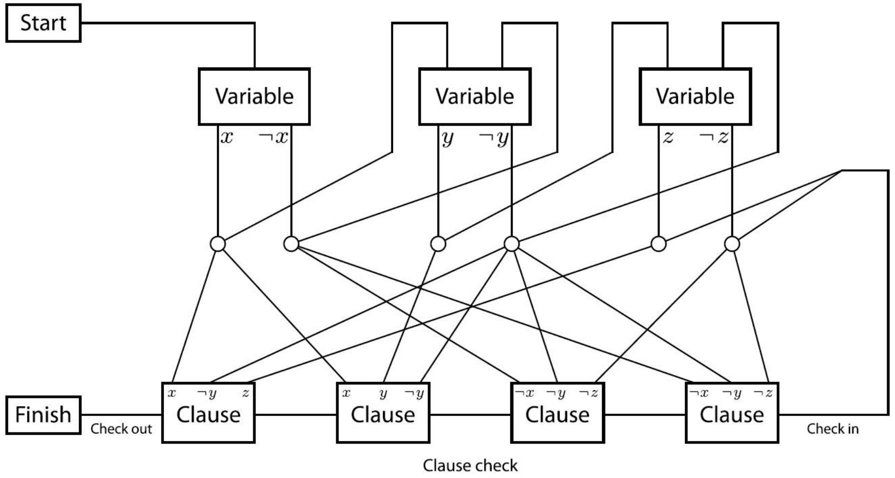
SAT

构件设计法：**SAT** — **3SAT**

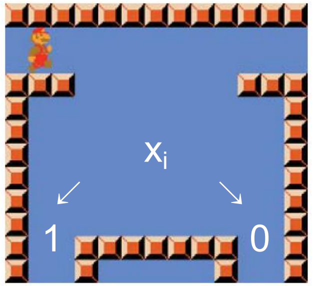
局部替换法：**SAT** — **最大可满足性**

- 3SAT实例：$F=(x_1\vee\neg x_2\vee x_3)\wedge(x_1\vee x_2\vee\neg x_3)$
  - 变元：$\{x_1,x_2,x_3\}$
  - 文字：x_1，$\neg x_2$，x_3，x_1，x_2，$\neg x_3$
  - 简单析取式：$C_1=(x_1\vee\neg x_2\vee x_3)$ 和 $C_2=(x_1\vee x_2\vee\neg x_3)$
  - 合取范式：$F=C_1\wedge C_2=(x_1\vee\neg x_2\vee x_3)\wedge(x_1\vee x_2\vee\neg x_3)$

构件（Gadget）

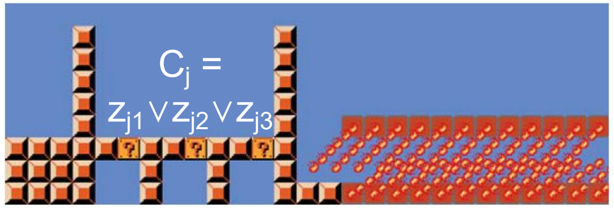


3SAT
变元构件

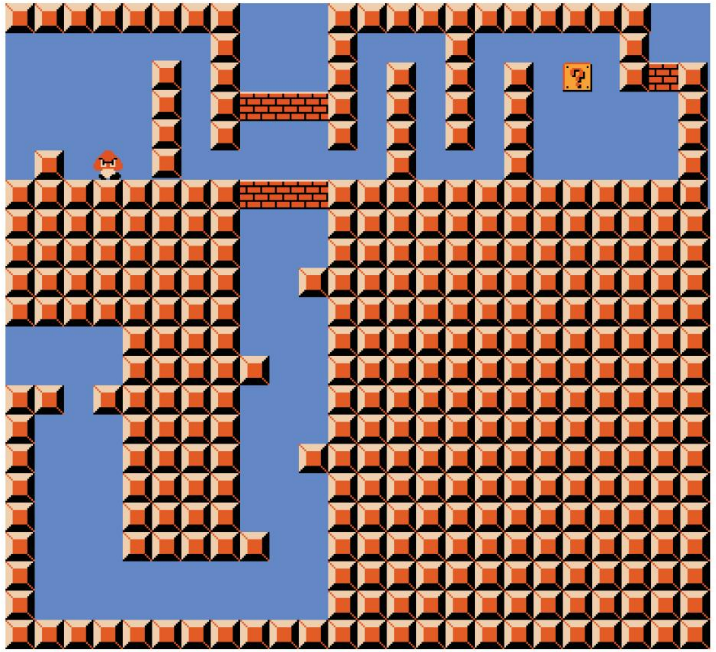


(a) Variable gadget

3SAT
子句构件



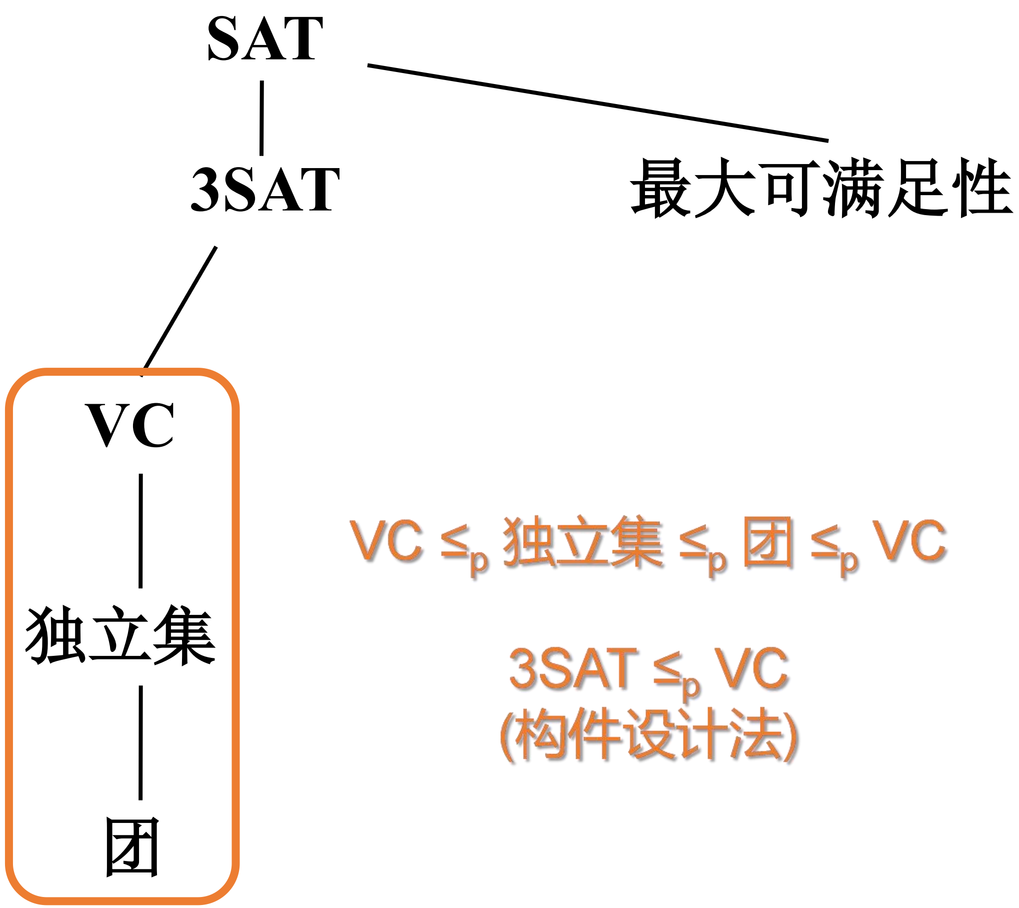
(b) Clause gadget



(c) Crossover gadget

3SAT
互连构件

若干NP完全问题



VC $\le_p$ 独立集 $\le_p$ 团 $\le_p$ VC

3SAT $\le_p$ VC
(构件设计法)

顶点覆盖、独立集、团

设无向图 $G = \langle V, E \rangle$，$V' \subseteq V$。V' 是 G 的一个

- 顶点覆盖：G 的每一条边都至少有一个顶点在 V' 中。
- 独立集：对任意的 $u, v \in V'$，都有 $(u, v) \notin E$。
- 团：对任意的 $u, v \in V'$ 且 $u \neq v$，都有 $(u, v) \in E$。

- 引理 对任意的无向图 $G = \langle V, E \rangle$ 和子集 $V' \subseteq V$，下述命题是等价的：

(1) V' 是 G 的顶点覆盖，
(2) $V - V'$ 是 G 的独立集，
(3) $V - V'$ 是补图 $G_c = \langle V, E_c \rangle$ 的团。

顶点覆盖、独立集、团的判定问题

- 顶点覆盖 (VC)：任给一个无向图 $G = \langle V, E\rangle$ 和非负整数 $K \leq |V|$，问 G 有顶点数不超过 K 的顶点覆盖吗？
- 独立集：任给一个无向图 $G = \langle V, E\rangle$ 和非负整数 $J \leq |V|$，问 G 有顶点数不小于 J 的独立集吗？
- 团：任给一个无向图 $G = \langle V, E\rangle$ 和非负整数 $J \leq |V|$，问 G 有顶点数不小于 J 的团吗？

- 根据上页引理，容易把这3个问题中的任一个问题实例多项式时间变换到另一个问题实例。

顶点覆盖

定理 顶点覆盖是NP完全的。

证：

- VC∈NP：VC 的多项式验证算法：存在多项式验证算法 $\mathcal{V}(I,W)$，在实例 I 及其顶点覆盖 V' 作为证据（W=V'）时值为1
- $3SAT\leq_p VC$：任给变元 $x_1, x_2, \ldots, x_n$ 的3元合取范式 $F = C_1 \wedge C_2 \wedge \ldots \wedge C_m$，其中$C_j = z_{j1} \vee z_{j2} \vee z_{j3}$，$z_{jk}$ 是某个 x_i 或 $\neg x_i$。
- 如下构造VC的实例 $f(F)$： $G = \langle V, E\rangle$ 和 $K = n + 2m$
- 其中 $V = V_1 \cup V_2$，$E = E_1 \cup E_2 \cup E_3$，

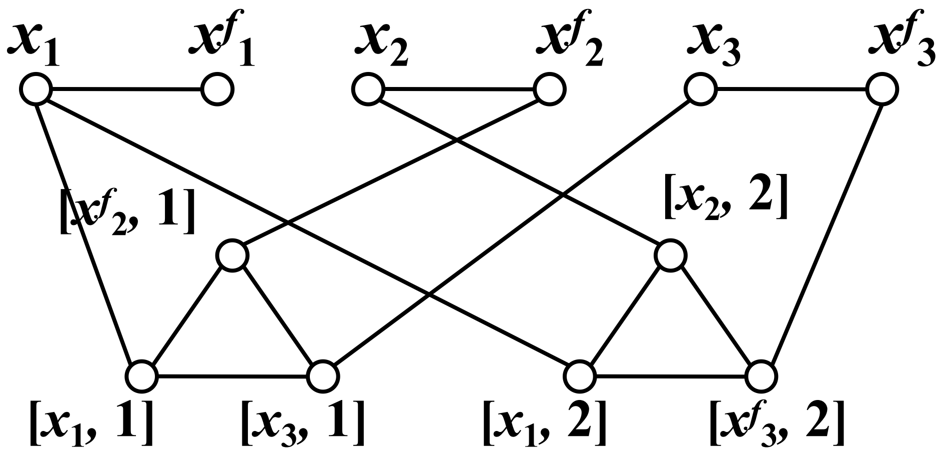
证明

$V_1 = \{x_i, x_i^f \mid 1 \le i \le n\}$，$V_2 = \{[z'_{jk}, j] \mid k = 1,2,3,\ 1 \le j \le m\}$；$E_1 = \{(x_i, x_i^f) \mid 1 \le i \le n\}$，

$E_2 = \{([z'_{j1,} j], [z'_{j2,} j]), ([z'_{j2,} j], [z'_{j3,} j]), ([z'_{j3,} j], [z'_{j1,} j]) \mid 1 \le j \le m\}$，

$E_3 = \{([z'_{jk}, j], z'_{jk}) \mid k = 1,2,3,\ 1 \le j \le m\}$。

- 这里设$C_j = z_{j1} \vee z_{j2} \vee z_{j3}$，当 $z_{jk} = x_i$ 时，$z'_{jk} = x_i$；当 $z_{jk} = \neg x_i$ 时，$z'_{jk} = x_i^f$。
- 例如，对应 $F = (x_1 \vee \neg x_2 \vee x_3) \wedge (x_1 \vee x_2 \vee \neg x_3)$ 的 $f(F)$：$K = 7$，图 G 如下



证明

要证 F 是可满足的 ⇔ G 恰好有 K 个顶点的顶点覆盖。

- 任何顶点覆盖 V' 在 x_i 和 $x^f{}_i$ 中至少取一个，在 $[z'_{j1,}j]$、$[z'_{j2,}j]$ 和 $[z'_{j3,}j]$ 中至少取 2 个，故 V' 至少有 $n+2m$ 个顶点。
- 而 $K=n+2m$，故任何顶点数不超过 K 的顶点覆盖 V' 恰好包含 K 个顶点，且在 x_i 和 $x^f{}_i$ 中取一个，这恰好对应对 x_i 的赋值，取 x_i 对应 $t(x_i)=1$，取 $x^f{}_i$ 对应 $t(x_i)=0$；每个三角形的顶点 $[z'_{j1,}j]$、$[z'_{j2,}j]$ 和 $[z'_{j3,}j]$ 中取 2 个。
- 设 t 是 F 的成真赋值，对每一个 $i\ (1\le i\le n)$，若 $t(x_i)=1$，则取 x_i；若 $t(x_i)=0$，则取 $x^f{}_i$。
- 这 n 个顶点覆盖 E_1。对每一个 $j\ (1\le j\le m)$，由于 $t(C_j)=1$，C_j 至少有一个文字 z_{jk} 的值为 1。
- 于是，从对应的三角形的顶点 $[z'_{jk},j]$ 引出的边 $([z'_{jk},j],\ z'_{jk})$ 已被覆盖。
- 取该三角形的另外2个顶点，这就覆盖了这个三角形的3条边和引出的另外2条边。
- 这样取到的 $n+2m$ 个顶点是 G 的一个顶点覆盖。

证明

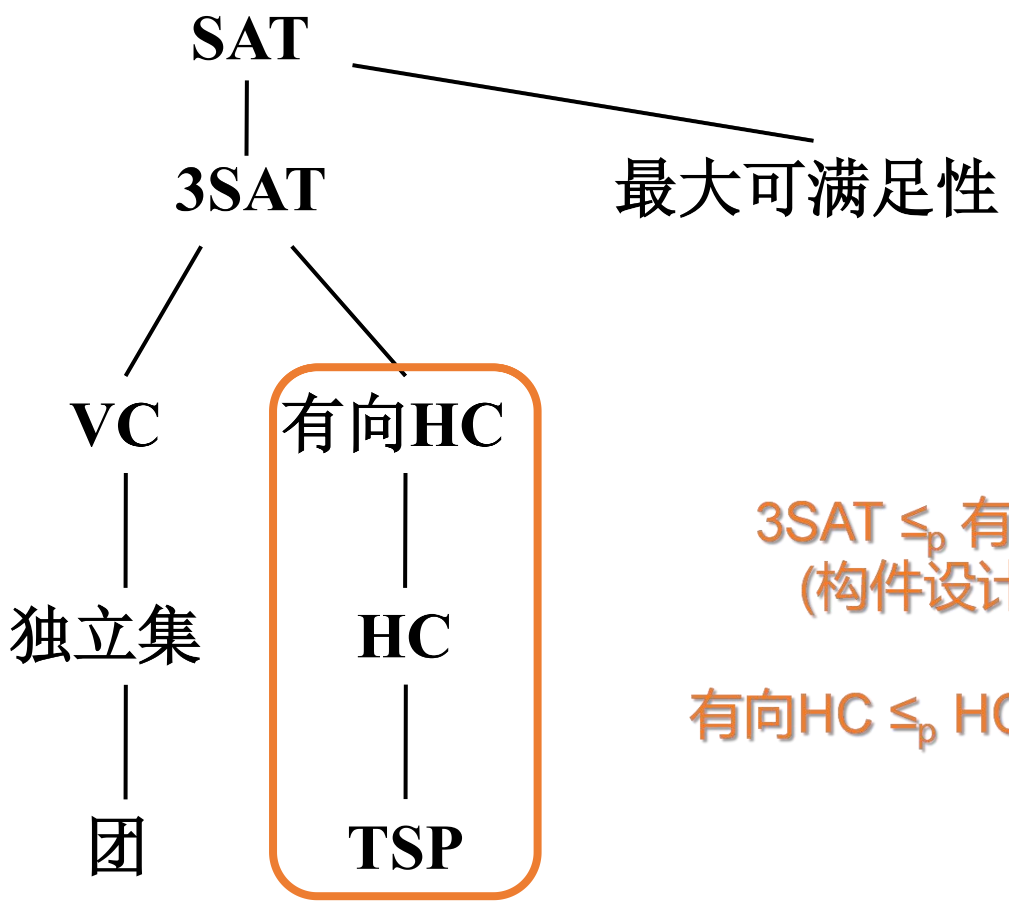
- 反之，设 $V'\subseteq V$ 是 G 的一个顶点覆盖且 $|V'|\leq K = n + 2m$。
- 根据前面的分析，每一对 x_i 和 $x^f{}_i$ 中恰好有一个属于 V'，每一个三角形恰好有2个顶点属于V'。所以 $|V'| = n + 2m$。
- 对每一个 $i\ (1 \leq i \leq n)$，若 $x_i\in V'$，则令 $t(x_i) = 1$；若 $x^f{}_i\in V'$，则令 $t(x_i) = 0$。
- 对每一个 $j\ (1 \leq j \leq m)$，设 $[z'_{jk}, j]\notin V'$，为了覆盖边 $([z'_{jk}, j], z'_{jk})$，必有 $z'_{jk}\in V'$。
- 由于 $t(z_{jk}) = 1$，从而 $t(C_j) = 1$。
- 因此，t 是 F 的成真赋值，得证 F 是可满足的。
- G 有 $2n + 3m$ 个顶点和 $n + 6m$ 条边，显然能在多项式时间内构造 G 和 K。

- 定理 独立集和团是NP完全的。

构件设计法（Gadget-based Reduction）

- 上页定理证明中设计了2种“构件”——变元构件和简单析取式构件。
- 3SAT变元构件是一对顶点 x_i，$x^f{}_i$ 及连接它们的边；
- 3SAT简单析取式构件是三角形。
- 用这些构件与构件连接构成 G，每个构件各有其功能，
- 通过这种方式到达用 VC 的实例表达 3SAT 的实例的目的。

若干NP完全问题



3SAT $\leq_p$ 有向HC
(构件设计法)

有向HC $\leq_p$ HC $\leq_p$ TSP

哈密顿回路与货郎问题

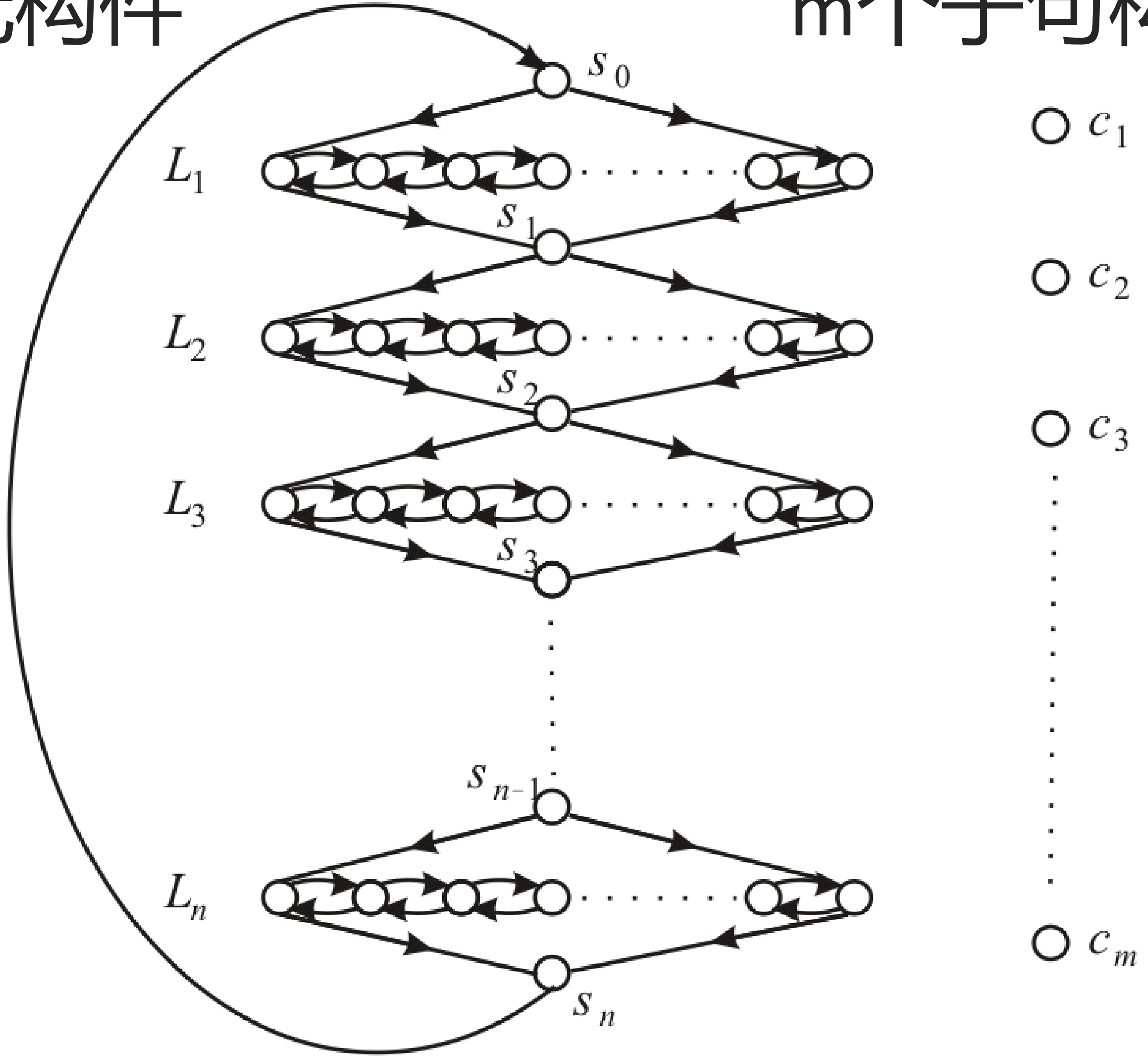
有向哈密顿回路：任给有向图 D，问：D 中有哈密顿回路吗？

定理 有向HC是NP完全的。

- 证 先证明有向HC∈NP（略），再证明 3SAT$\leq_p$有向HC。
- 任给变元 $x_1, x_2, \ldots, x_n$ 的3元合取范式 $F = C_1 \wedge C_2 \wedge \ldots \wedge C_m$，其中 $C_j = z_{j1} \vee z_{j2} \vee z_{j3}$，每个 z_{jk} 是某个 x_i 或 $\neg x_i$。
- 采用构件设计法构造有向图 D。
- 表示变元 x_i 的构件是一条由一串水平的顶点组成的链 L_i，相邻的两个顶点之间有一对方向相反的有向边。只有两种可能的方式通过 L_i 上的所有顶点——从左到右或者从右到左通过 L_i 上的所有顶点，这恰好对应 x_i 的值为1或者为0。
- 表示子句 C_j 的构件是一个顶点 c_j。
- 添加 $s_0, s_1, \ldots, s_n$，并通过它们把 $L_1, L_2, \ldots, L_n$ 连接起来。

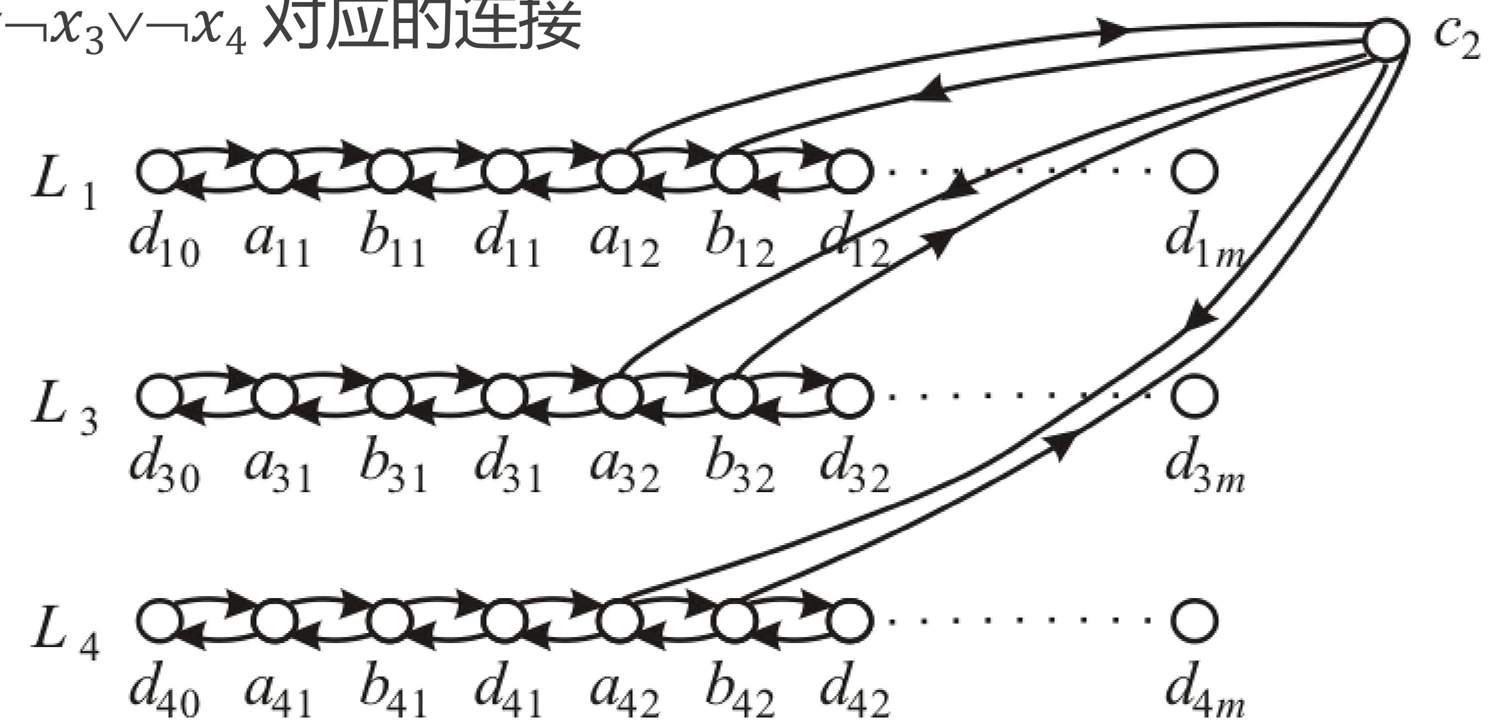
n个变元构件

m个子句构件



两种构件之间的连接

- 关键是两种构件之间的连接：链 L_i 有 $3m+1$ 的顶点，依次为 $d_{i0}, a_{i1}, b_{i1}, d_{i1}, a_{i2}, b_{i2}, d_{i2}, \ldots, a_{im}, b_{im}, d_{im}$。对每一个 $C_j = z_{j1} \vee z_{j2} \vee z_{j3}$， 如果 $z_{jk} = x_i$，则添加 $\langle a_{ij}, c_j \rangle$ 和 $\langle c_j, b_{ij} \rangle$；如果 $z_{jk} = \neg x_i$，则添加 $\langle c_j, a_{ij} \rangle$ 和 $\langle b_{ij}, c_j \rangle$。
- 例如 $C_2 = x_1 \vee \neg x_3 \vee \neg x_4$ 对应的连接



证明

- 设 F 是可满足的，t 是 F 的成真赋值。
- 要根据 t 构造一条从 s_0 到 s_n，最后回到 s_0 的哈密顿回路，先暂时不考虑所有的 c_j。
- 依次对 $i = 1,2,\ldots,n$ 进行，若 $t(x_i) = 1$，则从 s_{i-1} 到 d_{i0}，从左到右经过 L_i 的所有顶点到达 d_{im}，再到 s_i；若 $t(x_i) = 0$，则从s_{i-1} 到 d_{im}，从右到左经过 L_i 的所有顶点到达 d_{i0}，再到 s_i。
- 最后，从 s_n 回到 s_0。
- 现在要将所有 c_j 插入这条回路。
- 设 $C_j = z_{j1} \vee z_{j2} \vee z_{j3}$，由于 $t(C_j) = 1$，必有 $k\ (1 \le k \le 3)$ 使得 $t(z_{jk}) = 1$。
- 若 $z_{jk} = xi$，则通路从左到右经过 L_i，且有有向边 $\langle a_{ij}, c_j \rangle$ 和 $\langle c_j, b_{ij} \rangle$。于是，可以把 c_j 插在 a_{ij} 与 b_{ij} 之间。若 $z_{jk} = \neg x_i$，则通路从右到左经过 L_i，且有有向边 $\langle b_{ij}, c_j \rangle$ 和 $\langle c_j, a_{ij} \rangle$。于是，可以把 c_j 插在 b_{ij} 与 a_{ij} 之间。
- 这就得到 D 中的一条哈密顿回路。

证明

- 反之，设 D 有一条哈密顿回路 P，P 必须从 s_n 到 s_0。
- 不妨设 P 从 s_0 开始到 s_n，最后回到 s_0 结束。
- 我们称上面构造的那种哈密顿回路是正常的，即正常的回路从左到右或者从右到左通过每一条 L_i，每一个 c_j 插在某个 a_{ij} 和 b_{ij} 或者 b_{ij} 和 a_{ij} 之间。
- 如果 P 是正常的，容易根据 P 规定 F 的一个成真赋值 t：若 P 从左到右通过 L_i，则令 $t(x_i) = 1$；若 P 从右到左通过 L_i，则令 $t(x_i) = 0$。
- 根据 c_j 插入 L_i 的方式，不难证明必有 $t(C_j) = 1$。
- 要证 P 一定是正常的。

证明

- 要证 P 一定是正常的。假设不然，破坏正常性的唯一可能是 P 从某条链 L_s 上的顶点 u 到 c_j 后没有回到同一条链中的顶点，而是到另一条链 L_t $(s{\neq}t)$ 中的顶点。
- 若 $u = a_{sj}$，由于 b_{sj} 只与 a_{sj}、c_j 及 d_{sj} 相邻，P 已经过 a_{sj} 和 c_j，b_{sj} 只剩下一个相邻的顶点，故 P 不可能通过 b_{sj}。
- 若 $u = b_{sj}$，由于 a_{sj} 只与 b_{sj}、c_j 及 $d_{s(j-1)}$ 相邻，P 已经过 b_{sj} 和 c_j，a_{sj} 也只剩下一个相邻的顶点，故 P 不可能通过 a_{sj}。
- 都与 P 是哈密顿回路矛盾，所以 P 一定是正常的。

- 构造 D 可以在多项式时间内完成。

HC与TSP

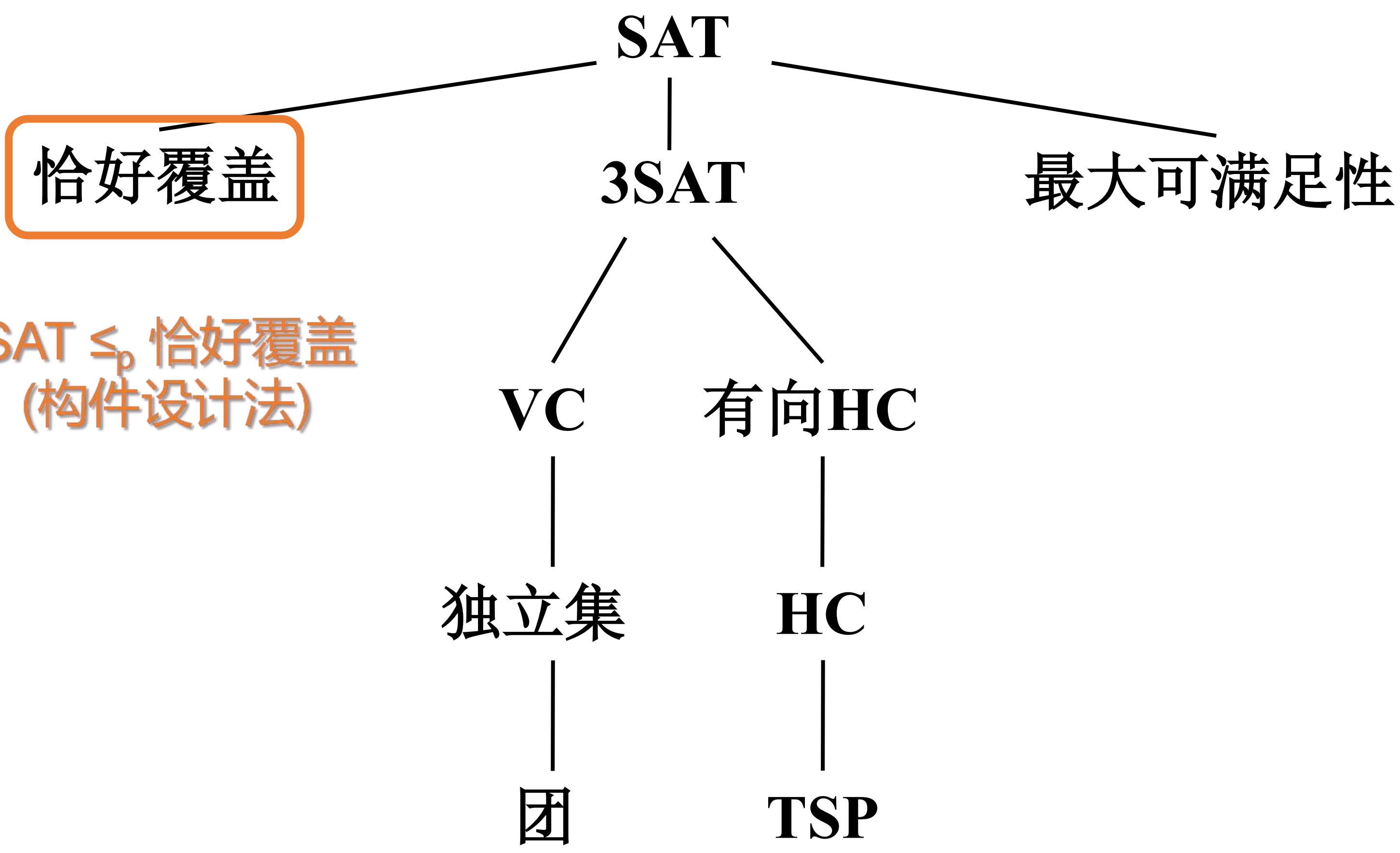
定理 HC 是 NP 完全的。

证：先证明 HC∈NP（略），再证明 有向HC $\le_p$ HC。

- 任给一个有向图 $D=\langle V,E\rangle$，要构造无向图 $G=\langle V',E'\rangle$ 使 D 有哈密顿回路当且仅当 G 有哈密顿回路。
- 把 D 的每一个顶点 v 替换成 3 个顶点 v^{in}，v^{mid} 和 v^{out}，用边连接 v^{in} 和 v^{mid}、以及 v^{mid} 和 v^{out}。
- D 的每条有向边 $\langle u,v\rangle$ 在 G 中换成 $(u^{out},\ v^{in})$。
- 即 $V'=\{v^{in},\ v^{mid},\ v^{out}\mid v\in V\}$，
- $E'=\{(u^{out},\ v^{in})\mid\langle u,v\rangle\in E\}\cup\{(v^{in},\ v^{mid}),\ (v^{mid},\ v^{out})\mid v\in V\}$。

定理 TSP 是 NP 完全的。（证明略）

若干NP完全问题



SAT $\leq_p$ 恰好覆盖
(构件设计法)

恰好覆盖

- 恰好覆盖：给定有穷集 $A = \{a_1, a_2, \ldots, a_n\}$ 和 A 的子集的集合 $W = \{S_1, S_2, \ldots, S_m\}$，问：存在子集 $U \subseteq W$ 使得 U 中的子集都是不相交的且它们的并集等于 A？称 W 这样的子集 U 是 A 的恰好覆盖。
- 例如，设 $A = \{1,2,3,4,5\}$, $S_1 = \{1,2\}$, $S_2 = \{1,3,4\}$, $S_3 = \{2,4\}$, $S_4 = \{2,5\}$，则 $\{S_2, S_4\}$ 是 A 的恰好覆盖。若把 S_4 改为 $S_4 = \{3,5\}$，则不存在 A 的恰好覆盖。

- 定理 恰好覆盖是NP完全的。

恰好覆盖：证明

证 先证明 恰好覆盖∈NP（略），再证明 可满足性$\leq_p$恰好覆盖。

- 任给变元 $x_1, x_2, \ldots, x_n$ 的合取范式 $F = C_1 \wedge C_2 \wedge \ldots \wedge C_m$，其中 $C_j = z_{j1} \vee z_{j2} \vee \ldots \vee z_{js_j}$。
- 取 $A = \{x_1, x_2, \ldots, x_n, C_1, C_2, \ldots, C_m\} \cup \{p_{jt} \mid 1 \leq t \leq s_j, 1 \leq j \leq m\}$，其中 p_{jt} 代表 C_j 中的文字 z_{jt}。
- W 包含下述4个子集

$$T^T{}_i = \{x_i, p_{jt} \mid z_{jt} = \neg x_i, 1 \leq t \leq s_j, 1 \leq j \leq m\}，1 \leq i \leq n;$$

$$T^F{}_i = \{x_i, p_{jt} \mid z_{jt} = x_i, 1 \leq t \leq s_j, 1 \leq j \leq m\}，1 \leq i \leq n;$$

$$C_{jt} = \{C_j, p_{jt}\}，1 \leq t \leq s_j, 1 \leq j \leq m；\{p_{jt}\}，1 \leq t \leq s_j, 1 \leq j \leq m。$$

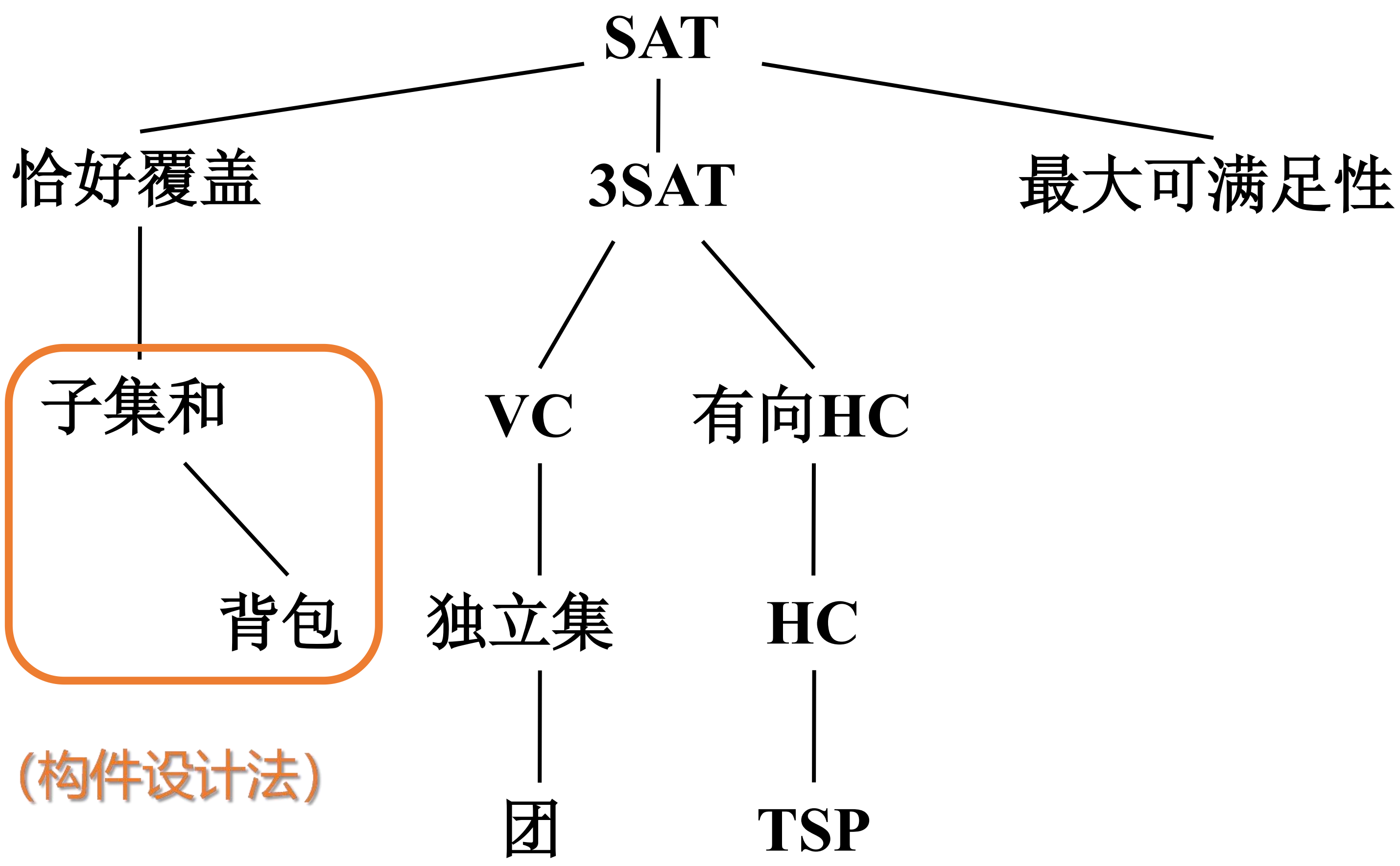
- 要证 F 是可满足的当且仅当 W 含有 A 的恰好覆盖。
- 设 $U \subseteq W$ 是 A 的恰好覆盖，对每一个 i，若 $T^T{}_i \in U$，则令 $t(x_i) = 1$；若 $T^F{}_i \in U$，则令 $t(x_i) = 0$。
- 对每一个 j，必有一个 $C_{jt} = \{C_j, p_{jt}\} \in U$。$z_{jt} = x_i$或 $\neg x_i$。
- 若 $T^T{}_i \in U$，则 $p_{jt} \notin T^T{}_i$，从而 $z_{jt} = x_i$。有 $t(x_i) = 1$，故 t 满足 C_j。
- 若 $T^F{}_i \in U$，则 $p_{jt} \notin T^F{}_i$，从而 $z_{jt} = \neg x_i$。有 $t(x_i) = 0$，故 t 也满足 C_j。得证 t 是 F 的成真赋值。

恰好覆盖：证明

- 反之，设 t 是 F 的成真赋值。
- 对每一个 i，若 $t(x_i) = 1$，则 U 包含 $T^T{}_i$；若 $t(x_i) = 0$，则 U 包含 $T^F{}_i$。
- 对每一个 j，由于 t 满足 C_j，C_j 必有一个文字 z_{jt} 使得 $t(z_{jt}) = 1$，从而 U 中现有的子集不包含 p_{jt}。
- 于是，可以把 C_{jt} 加入 U。至此，U 覆盖了所有的 x_i 和 C_j，以及部分 p_{jt}。
- 最后，把那些尚未被覆盖的 p_{jt} 构成的单元子集 $\{p_{jt}\}$ 加入 U，即可得到 A 的恰好覆盖。

- 由于 F 中的文字数不超过 mn，故 $|A| \leq n + m + mn$，W 中的子集数不超过 $2n + 2mn$，每个子集的大小不超过 $n + 1$。而且构造很简单，显然可以在多项式时间内完成。

若干NP完全问题



恰好覆盖 $\le_p$ 子集和（构件设计法）

子集和 $\le_p$ 背包（限制法）

子集和

- 子集和：给定正整数集合 $X = \{x_1, x_2, \cdots, x_n\}$ 及正整数 N，问存在 X 的子集 T，使得 T 中的元素之和等于 N 吗?

- 定理 子集和是NP完全的。
- 证：先证明 子集和∈NP（略），再证明 恰好覆盖 $\leq_p$ 子集和。
- 给定有穷集 $A = \{a_1, a_2, \cdots, a_n\}$ 和 A 的子集的集合 $W = \{S_1, S_2, \cdots, S_m\}$，对应的子集和实例包括非负整数的集合 $X = \{x_1, x_2, \cdots, x_m\}$ 及非负整数 N，每个 x_j 和 N 都可表成 kn 位的二进制数，这 kn 位分成 n 段，每段 k 位，$k = \lceil \log_2(m+1) \rceil$。

子集和：证明

- N 的每一段的第一位（最右的一位）为 1，其余的为 0。
- x_j 对应于子集 S_j。
- 当 $a_i \in S_j$ 时，从左到右 x_j 的第 i 段的第一位为 1，其余的为 0。
- 例如，$A = \{a_1, a_2, a_3, a_4\}, S_1 = \{a_1, a_2\}, S_2 = \{a_1, a_3, a_4\}, S_3 = \{a_2\}$,

  $$N = 01010101, x_1 = 01010000, x_2 = 01000101, x_3 = 00010000.$$
- 要证 W 中有 A 的恰好覆盖当且仅当存在子集 $T \subseteq X$ 使得 T 中元素之和等于 N。
- 设 $U \subseteq W$ 是 A 的恰好覆盖，令 $T = \{x_j \mid S_j \in U\}$。
- 由于 A 中的每一个元素在 U 的所有 S_j 中恰好出现一次，故对于二进制数的每一段，在 T 的所有 x_j 中恰好有一个的这一段为 $00\ldots01$，从而 T 中所有元素之和等于 N。

子集和：证明

- 反过来，设 X 的子集 T 中元素之和等于 N，令 $U = \{S_j \mid x_j \in T\}$。
- T 中至多有 m 个数，每一段有 $k = \lceil \log_2(m+1) \rceil$ 位，最大值为 $2^k - 1 \geq m$，故 T 中的数相加时不会出现段之间的进位。
- 从而，对于每一段，在 T 的所有 x_j 中恰好有一个的这一段为 $00 \ldots 01$。
- 这意味着每一个 a_i 在 U 的所有 S_j 中恰好出现一次，即 U 是 A 的恰好覆盖。
- 构造 X 和 N 显然可以在多项式时间内完成。

0-1背包与伪多项式时间算法

- 定理 0-1背包是NP完全的。
  - 0-1背包是NP的（略）
  - 0-1背包是NP难的：子集和是0-1背包的子问题——限制0-1背包的实例中所有 $w_i = v_i$ 且 $B = N$。

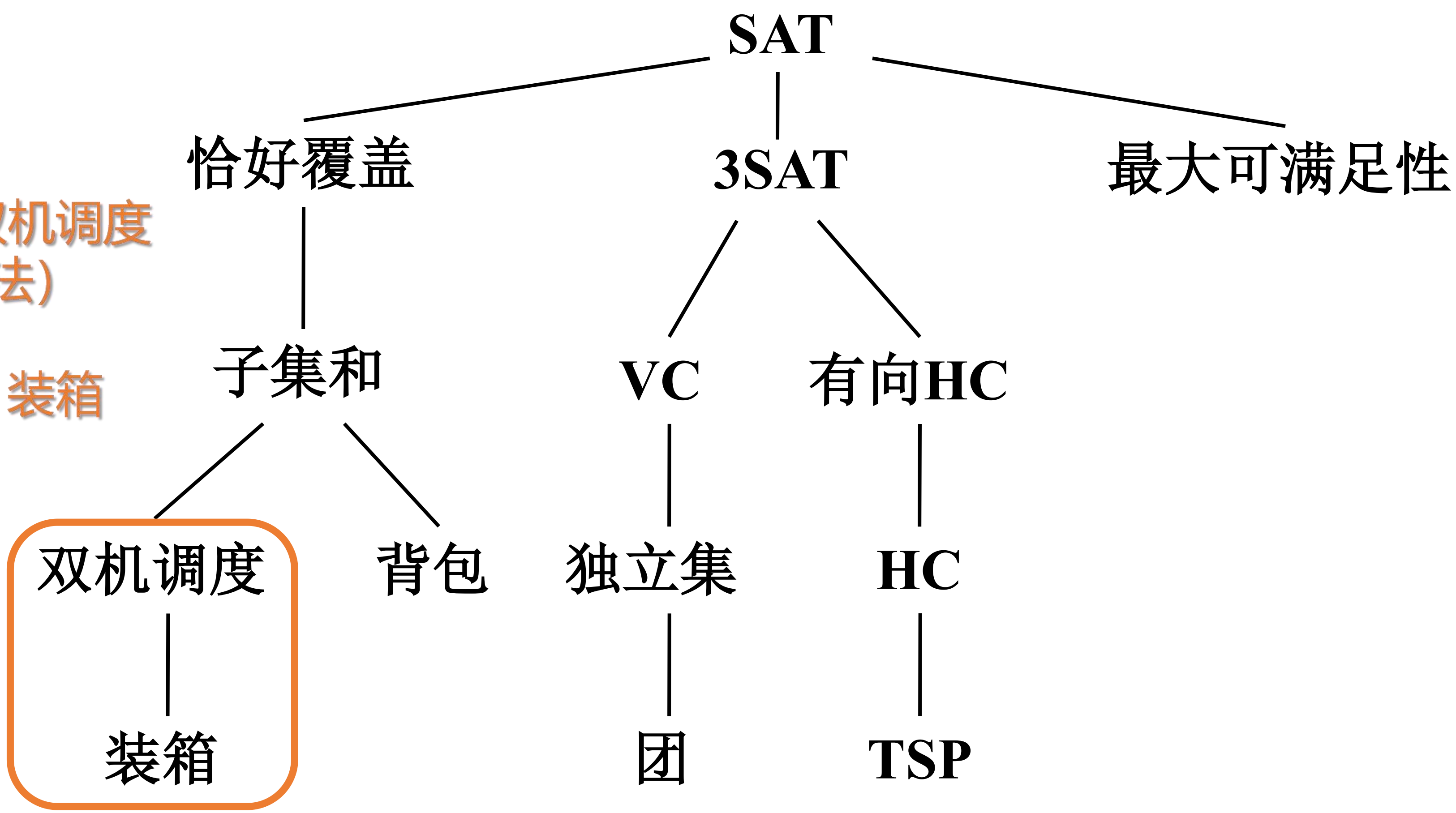
- 注意：0-1背包问题优化形式的动态规划算法, 其时间复杂度为 $O(nB)$，其中 n 是物品的个数，B 是重量限制。这不是多项式时间算法，而是指数时间算法.

- 伪多项式时间算法：算法的时间复杂度以 $|I|$ 和 $max(I)$的某个二元多项式 $p(|I|, max(I))$ 为上界，其中 $max(I)$ 是实例 I 中数的最大绝对值。

若干NP完全问题

子集和 $\le_p$ 双机调度（构件设计法）

双机调度 $\le_p$ 装箱（限制法）



双机调度与装箱

- 装箱：给定 n 件物品，物品 j 的重量为正整数 w_j, $1 \le j \le n$，以及箱子数 K。规定每只箱子装入物品的总重量不超过正整数 B，问能用 K 只箱子装入所有的物品吗?
- 双机调度：有 2 台机器和 n 项作业 $J_1, J_2, \ldots, J_n$，这2台机器完全相同，每一项作业可以在任一台机器上进行，没有先后顺序，作业 J_i 的处理时间为 t_i, $1 \le i \le n$，截止时间为 D，所有 t_i 和 D 都是正整数。问能把 n 项作业分配给这2台机器，在截止时间 D 内完成所有的作业吗?

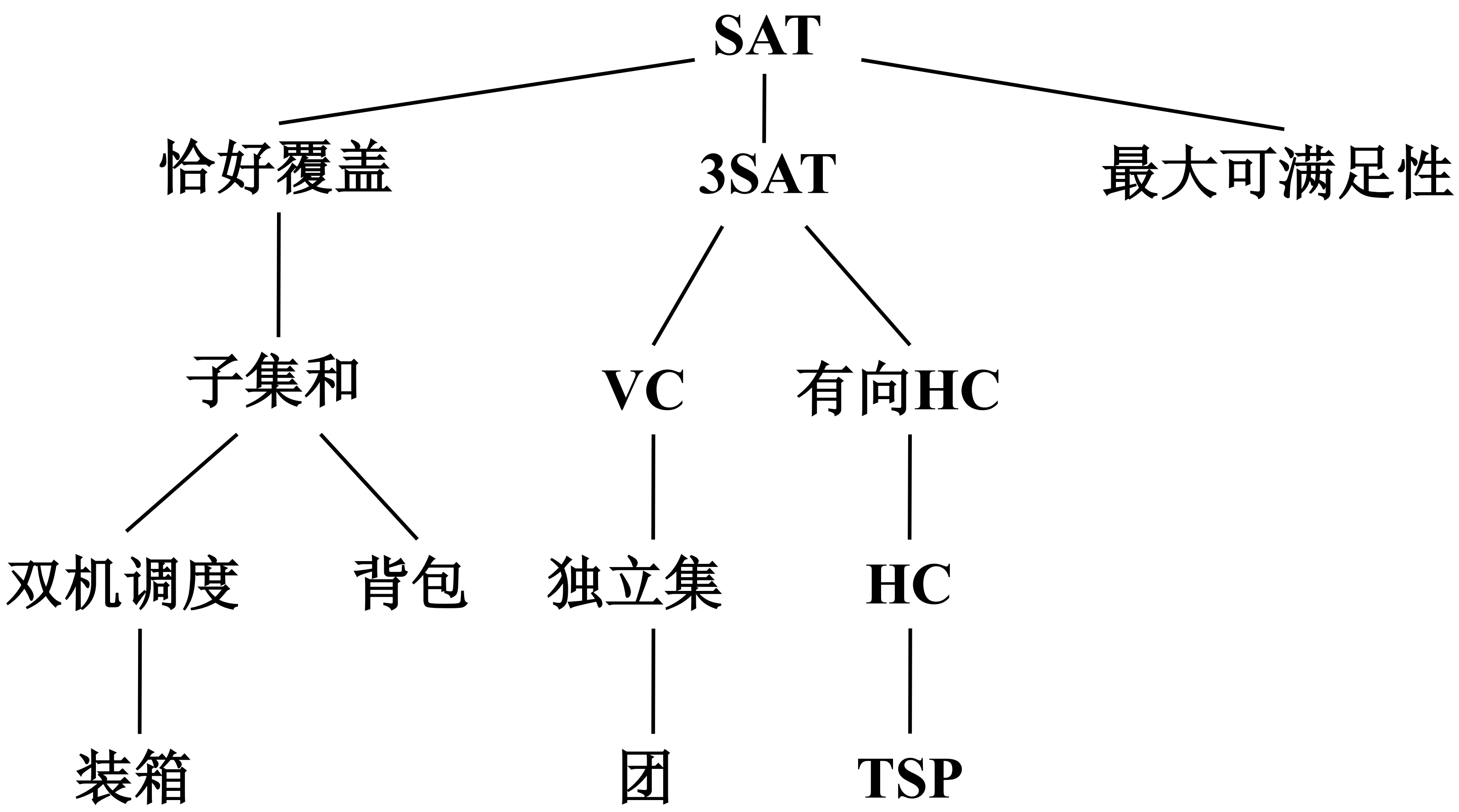
- 双机调度可以看作当箱子数 $K = 2$ 时装箱的特殊情况——把物品看作作业，物品的重量是作业的处理时间，截止时间是每只箱子允许的最大重量。

双机调度与装箱

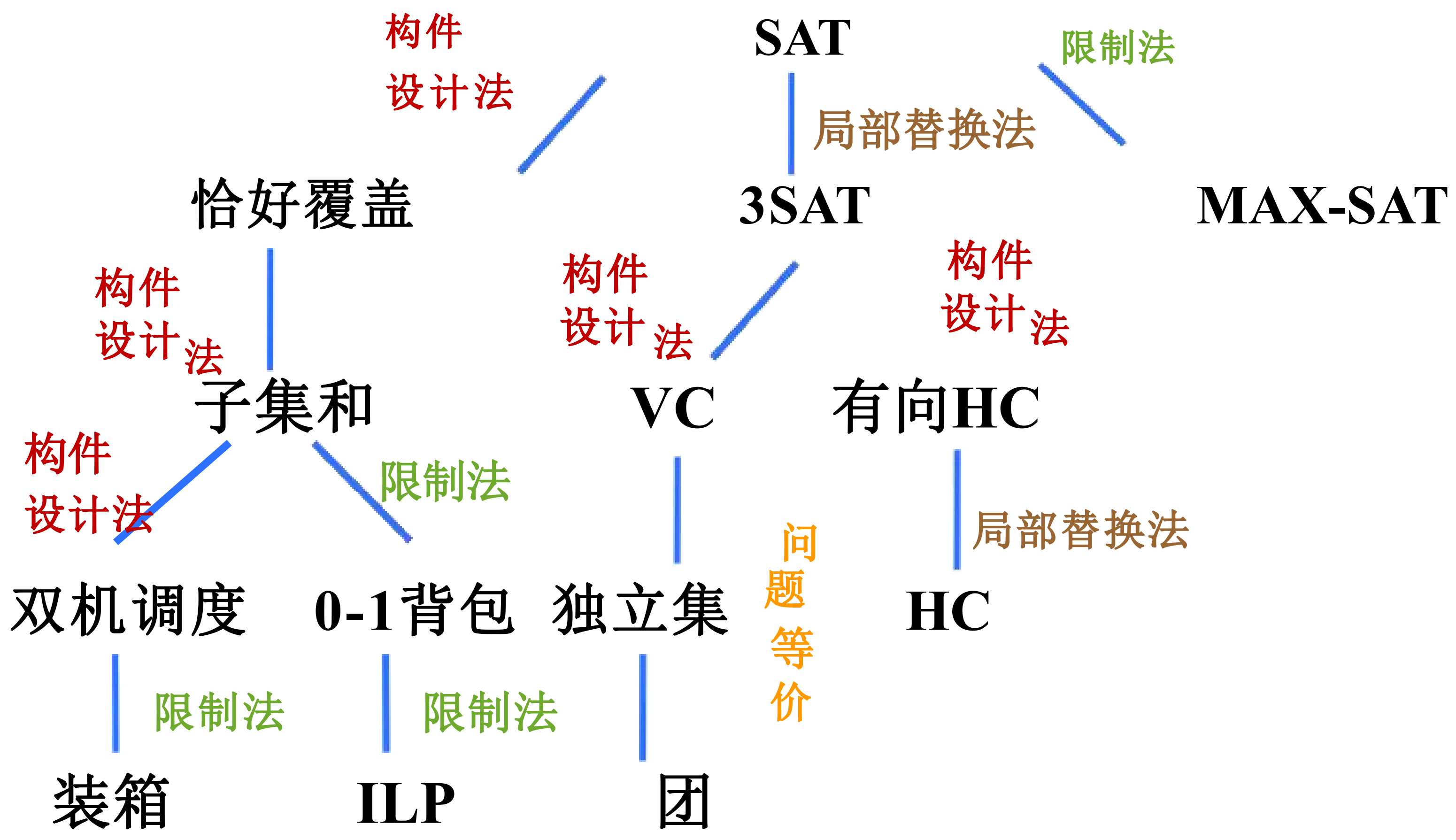
- 定理 双机调度是NP完全的。
- 证：先证明 双机调度∈NP（略），再证明 子集和$\leq_p$双机调度。
- 任给一个子集和实例：非负整数集合 $X = \{x_1, x_2, \cdots, x_n\}$ 及非负整数 N，对应的双机调度实例有 $n+2$ 项作业 $J_1, J_2, \cdots, J_{n+2}$，处理时间为 $x_1, x_2, \cdots, x_n, a, b$，截止时间为 D。
- 要求使得存在 X 的子集 T 当且仅当 $N + a = M - N + b = D$。
- 于是，$a = M - 2N + b$。
- 取 $b = M + 2N$，$a = 2M$，$D = 2M + N$，其中 $M = x_1 + x_2 + \cdots + x_n$。

- 定理 装箱是NP完全的.

若干NP完全问题



证明方法小结



NP-hard证明方法:

- 选好一个已知的NP完全问题.
- 使用限制法, 局部替换法和构件设计法.

本讲总结

- 证明问题 Π 是NP难的'
  - 选好一个已知的NP完全问题 Π'
  - 证明 $\Pi' \leq_p \Pi$ —— 常用的证明方法：限制法、局部替换法、构件设计法

- 证明问题 Π 是NP完全的
  - Π是NP的
  - Π是NP难的