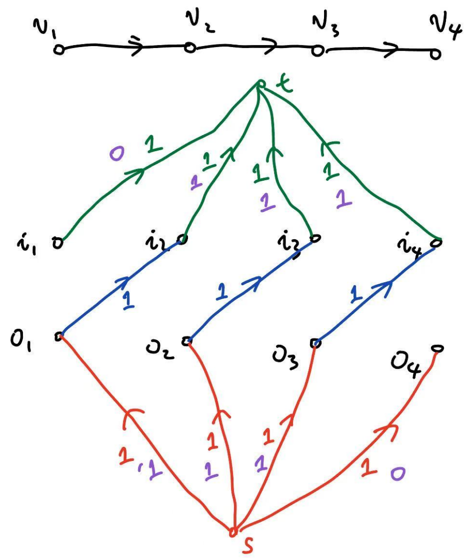
期末复习题

岳镝

例 1. 给定有向无圈图 *(DAG)* $G = (V, E)$。称 G 中若干条有向路径组成的集合 $\mathcal{P} = \{P_1, P_2, \dots, P_m\}$ 是 G 的一个（顶点不交）路径覆盖，若满足任何顶点 $v \in V$ 恰好属于 $\mathcal{P}$ 中一条路径。设计一个多项式时间算法，求 G 的最小路径覆盖。
（提示：构造一个图 G'，使得 G 的最小路径覆盖恰为 $n - \textsf{Max-Flow}(G')$）

解. 直觉上，需要限制流过每个顶点的总容量，因此考虑将每个顶点 $v \in V$ 拆成入顶点 i_v 和出顶点 o_v。对于 G 中的边 $\langle u, v\rangle$，改成 $\langle o_u, i_v\rangle$。此外增加源点 s 和汇点 t，连接 $\langle s, o_v\rangle$ 和 $\langle i_v, t\rangle$。



令 $G' = (V', E')$，其中

$$V' = \{i_v : v \in V\} \cup \{o_v : v \in V\} \cup \{s, t\},$$
$$E' = \{\langle o_u, i_v\rangle : \langle u, v\rangle \in E\} \cup \{\langle s, o_v\rangle : v \in V\} \cup \{\langle i_v, t\rangle : v \in V\}$$

令 G' 所有边的容量为 1。则 G 的最小路径覆盖大小恰为 $n - \textsf{Max-Flow}(G')$，证明留作练习。 □

例 2. 证明下列问题是 NP 完全的

(1) Π_1*:* 给定非负序列，能否划分成和相等的两个子序列？

(2) Π_2*:* 给定长度为偶数的非负序列，能否划分成和相等，长度也相等的两个子序列？

(3) Π_3*:* 给定长度为偶数的非负递增子序列，能否划分成和相等的两个子序列，且满足 a_{2i-1} 和 a_{2i} 属于不同的子序列 *(*$\forall\ 1 \le i \le n/2$*)*?

证明. 显然 $\Pi_1, \Pi_2, \Pi_3 \in \text{NP}$。下面只需证 Π_1, Π_2, Π_3 是 NP-难的。

(1) 习题 9.14.

(2) 考虑证明 $\Pi_1 \leqslant_p \Pi_2$。对于 Π_1 的实例 $A = (a_1, a_2, \ldots, a_n)$，构造 Π_2 的实例

$$B = (a_1, a_2, \ldots, a_n, \underbrace{0, 0, \ldots, 0}_{n\text{ 个}})$$

显然 B 可在多项式时间构造。只需证明 $A \in \Pi_1 \iff B \in \Pi_2$，留作练习。

(3) 考虑以下问题 Π_1'：

给定元素两两不等的非负序列，能否划分成和相等的两个子序列?

通过子集和 $\leqslant_p \Pi_1'$ 可证 Π_1' 是 NP 完全的。

以下考虑证明 $\Pi_1' \leqslant_p \Pi_3$。对于 Π_1' 的实例 $A = (a_1, a_2, \ldots, a_n)$，由以下方式构造 Π_3 的实例：首先将 A 从小到大排序，得到 $A' = (x_1, x_2, \ldots, x_n)$，其中 $x_1 < x_2 < \cdots < x_n$。随后令

$$C = (x_1, c_1, x_2, c_2, x_3, c_3, \ldots, x_n, c_n).$$

其中 $c_i = (1+\varepsilon)x_i$，ε 是一个充分小的正数，满足 $c_i = (1+\varepsilon)x_i < x_{i+1}$。

显然 C 可在多项式时间构造。只需证明 $A \in \Pi_1' \iff C \in \Pi_3$，留作练习。

□

例 3. 证明支配集问题是 NP 完全的：给定无向图 $G = (V, E)$ 和正整数 $K \le |V|$，是否存在子集 $V' \subseteq V$ 使得 $|V'| \le K$ 且 $V \setminus V'$ 中的每个顶点都至少与 V 中一个顶点相邻？

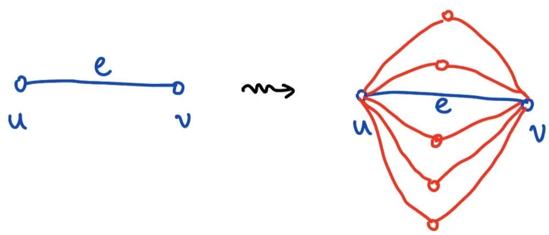
证明. 显然支配集 $\in \text{NP}$。下面证明 VC $\leqslant_p$ 支配集。

考虑 VC 的任意实例 $\langle G, K\rangle$，其中 $G = \langle V, E\rangle$。问题是如何用“覆盖顶点”来刻画“覆盖边”，一个自然的想法是把 G 中的每条边“替换”成若干顶点。

具体地，对 $e=(u,v)\in E$，令 C_e 是一个 $n+1$ 元顶点集合。考虑图 $G'=\langle V',E'\rangle$，其中

$$
\begin{aligned}
V' &= V \cup \bigcup_{e\in E} C_e,\\
E' &= E \cup \bigcup_{e=(u,v)\in E}\ \bigcup_{w\in C_e} \{(w,u),(w,v)\}
\end{aligned}
$$

注意到 $|V'|=n+(n+1)m$, $|E'|=m+2(n+1)m$，故 G' 可在多项式时间构造。下证 $\langle G,K\rangle\in$



VC $\Longleftrightarrow \langle G',K\rangle\in$ 支配集。

$\Longrightarrow$ 是显然的，证明留作练习。

$\Longleftarrow$: 假设 G' 存在大小不超过 K 的支配集 $U\subseteq V'$。我们断言任意边 $e\in E$ 必存在至少一个端点属于 U。若不然，假设边 $e=(u,v)$ 的两个端点都不属于 U。考虑顶点组 C_e，其所有顶点在 G' 中只和 u,v 两顶点相邻。为了支配 C_e，必有 $C_e\subseteq U$。故 $|U|\geq|C_e|=n+1>K$，矛盾！所以上述断言成立，这进一步说明图 G 存在一个大小不超过 K 的顶点覆盖。 □

例 4. 对 d-正则图 $G=(V,E)$，考虑支配集构造算法

1. $V'\leftarrow\emptyset$

2. *for* $i=1,2,\ldots,k$ *do*

3. 随机采样 $v\in V$

4. $V'\leftarrow V'\cup\{v\}$

(1) 证明当 $k=\frac{2n\ln n}{d+1}$ 时，V' 是支配集的概率至少 $1-1/n$

(2) 构造一个拉斯维加斯型随机算法，返回 G 的一个支配集

解. (1) 对于任意顶点 $v \in V$，v 不被 V' 支配当且仅当 v 及其所有邻居均不属于 V'。故

$$
\begin{aligned}
\Pr[v\text{ 不被 }V'\text{ 支配}] &= \Pr[v \notin V'] \cdot \prod_{(u,v)\in E} \Pr[u \notin V'] \\
&= \left(1-\frac{1}{n}\right)^{k} \cdot \left(1-\frac{1}{n}\right)^{kd} \\
&= \left(1-\frac{1}{n}\right)^{k(d+1)} \\
&\le \exp\left(-\frac{k(d+1)}{n}\right) \\
&= \frac{1}{n^2}
\end{aligned}
$$

上面用到一个熟知不等式：$1-t \le e^{-t}$。再由 Union Bound,

$$\Pr[\exists v\text{ 不被 }V'\text{ 支配}] \le \frac{1}{n}.$$

所以 V' 是支配集的概率至少 $1-1/n$。

(2) 独立重复运行上述算法，每次验证 V' 是否是支配集，直到得到一个支配集。假设算法成功概率是 p，则期望运行次数是 $1/p$。

□

例 5. 在最小支配集问题中，目标是求图 G 顶点数最少的支配集。

(1) 写出最小支配集问题的整数线性规划模型。

(2) 将约束放松为 $x_v \in [0,1]$，得到一个线性规划问题，从而可在多项式时间可求出最优解 $\{x_v^*\}_{v\in V}$。令 $V' = \{v\colon x_v^* \ge \frac{1}{\Delta+1}\}$。证明 V' 是一个 $(\Delta+1)$-近似解，其中 Δ 是图 G 的最大度。

(3) 对每个顶点 $v \in V$，构造随机变量 $M_v^1, M_v^2, \ldots, M_v^\alpha \sim_{i.i.d.} B(1, x_v^*)$。令

$$V' = \{v\colon \exists\, 1 \le i \le \alpha,\ \text{s.t. } M_v^i = 1\}$$

证明：存在整数 $\alpha = O(\log n)$，使得 V' 以概率 $1-1/n$ 是 G 的一个支配集。进而证明 V' 以常数概率构成一个 $O(\log n)$-近似解

解. (1) 用变量 $x_v \in \{0,1\}$ 表示顶点 v 是否属于支配集。对应整数线性规划 (IP)

$$
\begin{aligned}
& \text{minimize} \sum_{v\in V} x_v, \\
\text{s.t.} \quad & x_v + \sum_{u\colon (u,v)\in E} x_u \geq 1, \quad \forall v \in V \\
& x_v \in \{0,1\}
\end{aligned}
$$

(2) 放松约束后的线性规划 (LP) 是

$$
\begin{aligned}
& \text{minimize} \sum_{v\in V} x_v, \\
\text{s.t.} \quad & x_v + \sum_{u\colon (u,v)\in E} x_u \geq 1, \quad \forall v \in V \qquad (1)\\
& x_v \in [0,1]
\end{aligned}
$$

首先证明 V' 是一个支配集。考虑任意顶点 $v \in V$。注意到 v 至多有 Δ 个邻居，故根据约束条件 (1)，或者 $x_v^* \geq \frac{1}{\Delta+1}$，或者某个邻居 u 满足 $x_u^* \geq \frac{1}{\Delta+1}$。进而由 V' 的定义知 v 被 V' 支配。

近似比：对 $v \in V$，令

$$
x_v = \begin{cases} 1, & x_v^* \geq \frac{1}{\Delta+1} \\ 0, & o.w. \end{cases}
$$

则 $x_v \leq (\Delta+1)x_v^*$，且

$$
|V'| = \sum_{v\in V} x_v \leq (\Delta+1)\sum_{v\in V} x_v^* = (\Delta+1)\,\mathrm{OPT}_{LP} \leq (\Delta+1)\,\mathrm{OPT}_{IP}\,.
$$

(3) 令随机变量

$$
x_v = \begin{cases} 1, & v \in V' \\ 0, & v \notin V' \end{cases}
$$

则

$$
1 - \Pr[x_v = 1] = \Pr[x_v = 0] = \prod_{i=1}^{\alpha} \Pr[M_v^i = 0] = (1 - x_v^*)^{\alpha}.
$$

顶点 v 不被 V' 支配当且仅当 v 及其所有邻居均不属于 V'，故

$$\begin{aligned}
\Pr[v\text{ 不被 }V'\text{ 支配}] &= \Pr[x_v=0]\cdot\prod_{(u,v)\in E}\Pr[x_u=0] \\
&= (1-x_v^*)^\alpha\cdot\prod_{(u,v)\in E}(1-x_u^*)^\alpha \\
&\le e^{-\alpha x_v^*}\cdot\prod_{(u,v)\in E}e^{-\alpha x_u^*} \\
&= \exp\left(-\alpha\left(x_v+\sum_{u\colon (u,v)\in E}x_u\right)\right) \\
&\le e^{-\alpha}
\end{aligned}$$

再由 Union bound, V' 不是支配集的概率不超过 $ne^{-\alpha}$。

近似比：

$$\mathbb{E}[|V'|]=\sum_{v\in V}\Pr[x_v=1]=\sum_{v\in V}(1-(1-x_v^*)^\alpha)\le\sum_{v\in V}\alpha x_v^*=\alpha\,\mathrm{OPT}_{LP}\le\alpha\,\mathrm{OPT}_{IP}\,.$$

由马尔科夫不等式，以概率 $2/3$ 有 $|V'|\le 3\alpha\,\mathrm{OPT}_{IP}$。

综上所述，以概率 $2/3-ne^{-\alpha}$，V' 是一个 3α-近似解。取 $\alpha=O(\log n)$ 即得结论。

□

例 6. *(1)* 考虑求最小支配集的以下近似算法

每次找一对未被支配的相邻顶点，将它们加入 V'，直到所有顶点都被支配

构造实例说明该算法的近似比是 $\Omega(n)$

(2) 设计一个近似比为 $o(n)$ 的多项式时间确定性算法

解. (1) $G=\langle V,E\rangle$，其中

$$\begin{aligned}
V&=\{u,v_1,v_2,\ldots,v_n\},\\
E&=\{(u,v_i)\colon 1\le i\le n\}\cup\{(v_i,v_{i+1})\colon 1\le i\le n\}
\end{aligned}$$

(2) 算法如下：

每次将当前图中度数最大的顶点加入 V'，删去该顶点及其所有邻居，直至删去所有顶点。

设第 k 次操作后的图为 G_k，顶点数为 n_k。初始时 $G_0=G$。考虑任意正整数 k，假设 G_{k-1} 包含了最优解的 r 个顶点，则这 r 个顶点能支配 G_{k-1}，进而必有其中一个顶点的度不小于

$n_{k-1}/r - 1 \geq n_{k-1}/\operatorname{OPT} - 1$。从而

$$n_{k-1} - n_k \geq \frac{n_{k-1}}{\operatorname{OPT}}$$

解得

$$n_k \leq n\left(1 - \frac{1}{\operatorname{OPT}}\right)^k \leq ne^{-k/\operatorname{OPT}}.$$

当 $k > \ln n \cdot \operatorname{OPT}$ 时，$n_k < 1$，亦即 $n_k = 0$。这表明 $|V'| = k = O(\log n) \cdot \operatorname{OPT}$，即近似比为 $O(\log n)$。

□