概率方法

参考答案

1. (8 分) 用 $G(2n,p,q)$ 表示如下的随机图的分布：点集为 $V=\{1,2,\ldots,2n\}$. 选取随机的 $S^*\subseteq V$ 且 $|S^*|=n$. 这样点集被划分为了两个大小相同的块 S 和 $V\setminus S$. 对于任意两个点 (u,v)，如果 u,v 在同一个块中（$u,v\in S^*$ 或者 $u,v\notin S^*$），那么 u,v 之间以概率 p 有边相连；如果 u,v 在不同的块中，那么 u,v 之间以概率 q 有边相连.

   考虑高度稀疏的场景. 令 $p=\alpha/n$，$q=\beta/n$，其中 $\alpha,\beta\in\mathbb{R}$ 是常数. 请问是否可以依据图本身的信息对 S^* 进行一个非平凡的估计. 具体来说，是否存在一个算法 $\mathcal{M}$ 和常数 $c>0$ 使得

   $$\Pr\left[|\hat{S}|=n\wedge\frac{|\hat{S}\cap S^*|}{n}\notin(1/2-c,1/2+c)\middle|\begin{matrix}G\leftarrow G(2n,\alpha/n,\beta/n)\\ \hat{S}\leftarrow\mathcal{M}(G)\end{matrix}\right]=1-O(1/n).$$

   答案显然依赖于 (α,β) 的取值. 请找到一个尽量大的 (α,β) 的范围使得可以对 S^* 进行上述的非平凡估计.

   提示：当 $(\alpha-\beta)^2<\alpha+\beta$ 时，不存在任何算法可以非平凡地估计 S^* [Mossel-Neeman-Sly 2012].

   解　对任意 $S\subseteq V$，用 $e(S)$ 表示被被 S 割的边的数目.

   $$e(S):=|\{(u,v)\in E|u\in S,v\notin S\}|.$$

   考虑一个非常直观的算法：给定输入 G 之后，算法输出一个平衡的割 $\hat{S}\subseteq V$，使得 $e(\hat{S})$ 最大（如果 $\beta>\alpha$）或最小（如果 $\beta<\alpha$）.

   我们先考虑 $\beta<\alpha$ 的情况. 只需要证明，割边最小的平衡割大概率不是一个平凡估计.

   - 最非平凡的割，也就是 ground truth S^*，期望切割边数 $\mathbb{E}[e(S^*)]=\beta n$. 大概率 $e(S^*)$ 集中在 βn 附近. 具体来说，对于任意常数 $\delta>0$，根据 multiplicative Chernoff bound，

     $$\Pr[e(S^*)\geq(\beta+\delta)n]\leq 2^{-\Omega(n)}.$$

   - 我们选取一个常数 $c>0$. 对于任意平衡割 T，如果满足 $\frac{|T\cap S|}{n}\in(1/2-c,1/2+c)$，我们称 T 是平凡的. 对任意平凡的 T，其切割的边数的期望不低于

     $$\Big(\frac{\alpha+\beta}{2}-2c^2(\alpha-\beta)\Big)n=\Big(\frac{\alpha+\beta}{2}-\delta\Big)n.$$

     其中常数 $\delta=2c^2(\alpha-\beta)$. 通过选取合适的 c，我们可以让常数 δ 的值任意小.

     根据 multiplicative Chernoff bound，

     $$\Pr[e(T)\leq(\beta+\delta)n]\leq e^{-\left(\frac{\frac{\alpha-\beta}{2}-2\delta}{\frac{\alpha+\beta}{2}-\delta}\right)^2(\frac{\alpha+\beta}{2}-\delta)n/3}.$$

     只要常数 α,β 满足 $(\alpha-\beta)^2/(\alpha+\beta)>12\ln 2$，就可以找到足够小的常数 $\delta>0$ 使得上面的概率不超过 $2^{-(2+\delta)n}$.

     再使用 union bound，以 $1-2^{-\Omega(n)}$ 的概率，所有平凡平衡割 T 都满足 $e(T)>(\beta+\delta)n>e(S^*)$.

对称地，对于 $\alpha > \beta$ 的情况，只要常数 α, β 满足 $(\alpha-\beta)^2/(\alpha+\beta) > 8\ln 2$，以 $1-2^{-\Omega(n)}$ 的概率，所有平凡平衡割 T 都满足 $e(T) < (\beta-\delta)n < e(S^*)$.

2. (8 分) 有若干集合 $A_1, \dots, A_n, B_1, \dots, B_n$，满足 $\forall i, j \;\; A_i \cap B_j = \varnothing \iff i = j$. 证明

$$\sum_i \frac{1}{\binom{|A_i|+|B_i|}{|A_i|}} \le 1.$$

解　选取一个 $\bigcup_i (A_i \cup B_i)$ 上随机的序. 定义事件 E_i 为“A_i 中每一个元素小于 B_i 中每一个元素”. 对于每一个 i，有 $\Pr[E_i] = 1/\binom{|A_i|+|B_i|}{|A_i|}$.

用反证法. 如果

$$\sum_i \Pr[E_i] = \sum_i \frac{1}{\binom{|A_i|+|B_i|}{|A_i|}} > 1,$$

那么存在 $i \neq j$ 使得 E_i, E_j 有可能同时发生. 考虑使 E_i, E_j 同时发生的一个序. 在这个序下，$\max(A_i) < \min(B_i), \max(A_j) < \min(B_j)$. 因此必有 $\max(A_i) < \min(B_j)$ 或者 $\max(A_j) < \min(B_i)$. 这两种可能分别都会推出矛盾.

3. (5 分) 考虑如下生成社交网络过程：初始时只有一人；每当一个新人加入时，会随机关注一个人，被关注概率正比于当前被关注次数加上 β. 这里 $\beta > 0$ 是一个常数. 请问当网络中有 n 人时，第 k 个加入网络的人的被关注次数的期望约是多少.

解　令 $E_{n,k}$ 表示当网络中有 n 人时，第 k 个加入网络的人的被关注次数的期望，我们可以得到 $\sum_{k=1}^n E_{n,k} = n-1$，且有

$$\begin{aligned} E_{k,k} &= 0 \\ E_{n+1,k} &= E_{n,k} + \Pr[\text{第 } n+1 \text{ 个人关注了第 } k \text{ 个人}] \\ &= E_{n,k} + \frac{E_{n,k}+\beta}{n-1+n\beta}. \end{aligned}$$

也即有

$$E_{n+1,k} + \beta = \frac{n(\beta+1)}{n(\beta+1)-1}(E_{n,k}+\beta) = \beta \prod_{i=k}^{n} \frac{i(\beta+1)}{i(\beta+1)-1}.$$

据此，可以估算

$$\frac{E_{n,k}+\beta}{\beta} = \prod_{i=k}^{n-1} \frac{i}{i-\frac{1}{1+\beta}} = \frac{\Gamma(n)\Gamma(k-\frac{1}{1+\beta})}{\Gamma(k)\Gamma(n-\frac{1}{1+\beta})} \in \left(\left(\frac{n-1}{k-\frac{1}{1+\beta}}\right)^{\frac{1}{1+\beta}}, \left(\frac{n-\frac{1}{1+\beta}}{k-1}\right)^{\frac{1}{1+\beta}} \right)$$

所以 $E_{n,k} \approx \beta \cdot ((n/k)^{\frac{1}{1+\beta}} - 1)$.

4. (8 分) 如果集合 $S \subseteq \{0,1\}^n$ 满足

$$\forall \text{distinct } a, b, c \in S,\ \Delta(a,b) + \Delta(b,c) > \Delta(a,c),$$

我们称 S 是不共线的. 这里 Δ 表示汉明距离（Hamming distance）.

(1) 证明，对所有足够大的 n，存在不共线的 $S \subseteq \{0,1\}^n$ 满足 $|S| \geq 1.01^n$.

(2) 证明，对所有足够大的 n，任何不共线的 $S \subseteq \{0,1\}^n$ 都满足 $|S| \leq 1.99^n$.

解　对于任何三个点 $a,b,c \in \{0,1\}^n$，如果 $\Delta(a,b)+\Delta(b,c)=\Delta(a,c)$，我们称 a,b,c 共线；否则我们称 a,b,c 不共线.

(1) 如果独立随机选取 $a,b,c \in \{0,1\}^n$，那么

$$\Pr[a,b,c\text{共线}] = \Pr\Big[\forall i \in [n], (a_i,b_i,c_i) \notin \{(0,1,0),(1,0,1)\}\} = \varnothing\Big] = (3/4)^n.$$

独立随机选取 $x_1,x_2,\ldots,x_m \in \{0,1\}^n$，其中 $m = 2^{\alpha n}$. 令 $S = \{x_1,\ldots,x_m\}$. 那么

$$\Pr\big[|S| = m\big] = \Pr\big[\forall \text{distinct } i,j,\ x_i \neq x_j\big] \leq \frac{1}{2}2^{2\alpha n}\cdot 2^{-n}.$$

$$\Pr[S\text{中存在共线三元组}] \leq \frac{1}{2}2^{3\alpha n}\cdot (3/n)^n.$$

只要 $3\alpha < \log_2(4/3)$，就有非零概率 $|S| = 2^{\alpha n}$ 且 S 不共线.

(2) 对任何不共线的 S，构造一个二部图 $G=(L,R,E)$. 左右点集分别为 $L = \{0,1\}^{\lfloor n/2\rfloor}$，$R = \{0,1\}^{\lceil n/2\rceil}$. 边集与 S 一一对应：$(u,v) \in E$ 当且仅当 $u|v \in S$. 这里 $|$ 表示拼接.

图中一定没有长度为 3 的路径. 反证，假如有长度为 3 的路径 $(u,v),(u',v),(u',v')$，那么

$$\Delta(u|v,u'|v)+\Delta(u'|v,u'|v') = \Delta(u,u')+\Delta(v,v') = \Delta(u|v,u'|v),$$

说明 $u|v, u'|v, u'|v' \in S$ 这三点共线. 与 S 不共线矛盾.

换言之，图中任何两个度数大于 1 的点不相邻. 再换言之，每条边都有至少一个端点的度数等于 1. 所以存在一个从度数为 1 的顶点集到 E 的满射.

$$|S| = |E| \leq |L|+|R| = 2^{\lfloor n/2\rfloor} + 2^{\lceil n/2\rceil}.$$