模块一习题

1　为什么 Erdős-Renyi 随机图不是好的社会网络模型？

在课上我们介绍了 Erdős-Renyi 模型。对于正整数 n 和 $p \in (0,1]$, Erdős-Renyi 模型是一个 n 个点的随机无向图 $G(n,p)$, 其中对于任何一对节点 $\{u,v\}$, 都独立地以概率 p 将 $\{u,v\}$ 加入边集。我们在课上证明了当 $p = \Omega\left(\frac{\ln n}{n}\right)$ 时，大概率 $G(n,p)$ 中任何一对节点的距离是常数。这一定程度上反映了社会网络常见的“小世界”现象。

然而，我们也提到 Erdős-Renyi 并不是一个好的社会网络模型。考虑典型的强连通、有小世界性质时的 $p = \Theta\left(\frac{\ln n}{n}\right)$。**请尝试证明以下两个一般社会网络具备但是 Erdős-Renyi 不具备的性质**。

1. Erdős-Renyi 模型不具备三元闭包性质，特别地，求证聚集系数很低，是 $o(1)$ 的。回忆聚集系数的定义：对某个节点 u，设 N_u 是 u 的邻居的集合并且 $k_u = |N_u|$，则 u 的聚集系数是

$$\frac{|(v,w) \in E : v,w \in N_u|}{k_u(k_u-1)}$$

2. 求证 Erdős-Renyi 模型中节点的度的分布不是幂律。回忆幂律分布的定义是 x 出现的概率正比于 $1/x^\alpha$ $(\alpha > 0)$。这里为了方便可以假定 $\alpha = 2$。

2　幂律模型的小世界现象

课上我们介绍了如下的幂律模型：有 n 个点，按照 $i = 1, 2, \ldots, n$ 的顺序执行下面的随机过程：设当前节点是 i，以 p 的概率均匀随机选取一个 1 到 $i-1$ 的点 j 并将 i 连到 j，以剩下 $1-p$ 的概率均匀选取一个 1 到 $i-1$ 的点 j 并将 i 连到 j连接的点。这里 $p \in [0,1]$ 是一个参数。**求证：**对于任意两个节点，他们的最短路的长度的期望至多是 $O(p \ln n)$。提示：设 $f(i)$ 代表从 i 点出发到 i 之前点最远距离的期望。尝试写出 $f(i)$ 的递推式。

3　信息级联

考虑课堂上介绍的级联实验设置。令 $p = 1/2$，$q = 2/3$，用 b, r 表示信号，B, R 表示状态。（所以 $\Pr[B] = \Pr[R] = p$，$\Pr[b \mid B] = \Pr[r \mid R] = q$。）现在设想每个参与实验的人在基于已知信息算得状态概率为 0.5 的时候，以概率 $0 < c < 1$ 选择按照自己私有信号的指引对状态给出判断。**请回答下面两个问题**。

1. 论证当第 3 个人看到前面两人给出的判断为 (B, B) 时，无论她拿到的私有信号是什么，她都将给出判断 B（尽管知道第 2 个人可能拿的是 r 信号），从而级联形成。

2. 对于有 m 对 b 和 r 交替的信号序列（于是有 $2m$ 个信号），$b, r, b, r, \ldots, b, r$，给出在第 $2m+1$ 个信号上级联形成（即从第 $2m+1$ 个人开始完全放弃私有信号）的概率。并讨论其相对于不同 c 的合理性。

4 Watts-Strogatz 模型的小世界现象

考虑原始的、不能很好支持短视搜索的 Watts-Strogatz 模型：节点集合是 $n \times n$ 的二维网格，每个节点 u 都连接上下左右一个格点步内的邻居，并且按照均等概率连一条到整个 $n \times n$ 网格点的随机边。现在我们证明任何两点间的最短路径的长度都大概率是 $O(\log^2 n)$ 的。

课上我们介绍了证明的框架：从某个初始点 u 出发，定义一个两阶段的遍历过程。这里给出一个更加细化的版本。此处，我们引入了一个参数 s，它的值会在后面确定。

- 第一阶段：第一轮先用 s 步访问 u 的直接邻居，然后这些访问到的邻居通过自己的随机边来到达新的节点。一般地，从上轮找到的新的节点集出发，通过 s 步直接连接访问每个新节点的 $O(s)$ 个直接邻居，然后这些邻居通过自身的随机边到达本轮的新找到的节点。该过程持续到多于一半的节点已被访问为止，并进入第二阶段。

- 第二阶段：从第一阶段访问过的点出发，每次只直接连边完成对剩下节点的访问。

在课上，我们使用 "coupon collection" 证明了第二阶段只需 $O(\log n)$ 步即可访问所有节点。现在我们更详细地考察第一阶段。

在第一阶段中，假设某一轮中有 m 个新邻居准备发出一条随机边，那么我们希望这 m 条边都能连接到不同的未被访问的点（否则我们只会在已有的连接中绕来绕去，没有推进遍历过程）。为此，**请证明下面这个引理 1**，它描述了 m 条随机边连接的点大概率会相对均匀的分布的事实。

引理 1. *设有 N 个盒子和 m 个球（$m \le O(N)$）。现将每个球均匀随机地投入盒子中。设第一个盒子里面最后有 X 个球，则*

$$\Pr[X > \Omega(\log N)] \le 1/N^2.$$

提示：利用 Chernoff bound。

定理 (Chernoff bound)**.** *设有 k 个独立的 $[0,1]$ 上的随机变量 $Y_1, \ldots, Y_k$。设 $Y = \sum_{i=1}^{k} Y_i$。则对于任何 $t > 0$，*

$$\Pr[Y - E[Y] > t \cdot E[Y]] \le \exp(-t^2 E[Y]/(t+2))$$

使用引理 1得出结论。 引理 1不仅对第一个盒子有效，事实上对任何一个盒子都成立，因为盒子之间是对称的。利用 union bound，我们可以得到以高概率，所有 N 个盒子里面的球都不超过 $O(\log N)$ 个。

在我们的第一阶段，尚未被访问的至少 $0.5n^2$ 个点对应于 N 个盒子，如果某轮有 m 个点将要发出随机边（此时必有 $m \le O(N)$ 因为第一阶段没有结束），那么根据上述球与盒子的分析，大概率地，每一个未被访问的节点只会被 m 条中最多 $O(\log N) = O(\log n)$ 条随机边连接。因此，本轮仍能到达 $\Omega(m/\log(n))$ 个<u>不同的</u> 新节点。

我们取 $s = \Omega(\log n)$ 并且令 Ω 里面隐藏的常数足够大，就可以让下一轮有 $\Omega(m/\log n) \cdot \Omega(\log n) \ge 2m$ 个新节点。这样就可以使每轮访问到的新节点翻倍，因此在 $O(\log n)$ 轮后第一阶段可以停止。最后，因为每轮我们需要沿着 $O(s)$ 条直接连边遍历，那么第一轮总的遍历次数（大概率）是 $O(s\log n) = O(\log^2 n)$。