

NOTES

#8

B. MISHRA

MARCH 11, 2012

Giant Components in Erdős-Rényi Random Graphs

Recall from the last lecture, that a very important phase transition occurs in an Erdős-Rényi (ER) random graphs $G(n, p)$, when $p(n) = \frac{\ln n}{n}$. That is, when $p(n) \ll \frac{\ln n}{n}$, the ER graph in $G(n, p(n))$ is almost surely disconnected. But even in this regime, some very interesting phenomenon occur: for instance, the phase transitions associated with the so-called *Giant Components*.

In this regime – when $p(n) \ll \frac{\ln n}{n}$ – we will see that for the random variable $X = \sum_{i=1}^n I_{i=\text{isolated}}$, we have

$$E[X] = n(1 - p(n))^{n-1} \approx n \cdot \exp[-p(n)n] \rightarrow \infty, \quad \lim_{n \rightarrow \infty} .$$

Thus, when $p(n) \ll \frac{\ln n}{n}$, the graph has an arbitrarily large number of connected components. We will consider two regimes:

$$p(n) = \frac{\lambda}{n}, \quad \begin{cases} \lambda < 1 \\ \lambda > 1. \end{cases} \quad \text{vs.}$$

For $\lambda < 1$, all components of the graph are “small;” for $\lambda > 1$, the graph has a *unique giant component*.

A component that contains a constant fraction of individuals in the social network, with a local degree $\approx (1 + \epsilon)/n$.

Consider a breadth-first search of the graph starting with the node i , and collect all the nodes in the k th stage (i.e., at a distance k from the node i). We compare two similar processes: *Graph Process* (the one discussed above) and *Branching Process* (at each stage an individual of stage k gets some number of “new” friends – individuals in stage $k + 1$).

Z_k^G = # of individuals at stage k of the graph

Z_k^B = # of individuals at stage k in a branching process

$$Z_k^G \leq Z_k^B;$$

– If there is no triadic closure graph process keeps discovering “new” friends.

Expected number of friends for a node = $n \cdot p(n) = \lambda$.

$$E[Z_k^B] = \lambda^k.$$

Let S_1 be the number of nodes in the Erdős-Rényi graph connected to individual 1.

$$E[S_1] = \sum_{k=1}^{\Delta} E[Z_k^G] < \sum_k \lambda^k = \frac{1}{1-\lambda}.$$

Theorem: Let $p(n) = \frac{\lambda}{n}$ ($\lambda < 1$). For all sufficiently large $a > 0$, we have

$$Pr \left[\max_{1 \leq i \leq n} |S_i| \geq a \ln n \right] \rightarrow 0, \quad n \rightarrow \infty.$$

where $|S_i|$ = size of the component containing individual i .

Proof omitted. \square

Giant component with $p(n) = \frac{\lambda}{n}$ ($\lambda > 1$).

The trick is to first get a good estimate of the expected number of “conflicts” at stage $(k + 1)$. Note that when two of the “friends” at stage k have a common friend at stage $k + 1$, we count it as one “conflict.”

$$\begin{aligned} & E[\text{Number of conflicts at stage } k + 1] \\ &= E \left[\binom{Z_k}{2} \binom{Z_k}{1} p^2 \right] \\ &\leq np^2 E[Z_k^2] \\ &= np^2 \{ \text{Var}[Z_k] + E[Z_k]^2 \} \\ &= np^2 (\lambda^k + \lambda^{2k}) \quad Z_k \sim \text{Poisson}(\lambda^k) \\ &\leq n \cdot \frac{\lambda^2}{n^2} \cdot \lambda^{2k}. \end{aligned}$$

Thus the effects of “conflicts” become nonnegligible, when $\lambda^k \approx \sqrt{n}$.

Theorem: Let $p(n) = \frac{\lambda}{n}$ ($\lambda > 1$). Then there exists some $c > 0$ such that

$$Pr[\exists \text{ a component } C, |C| > c\sqrt{n}] \rightarrow 1, \quad \lim_{n \rightarrow \infty} . \quad \square$$

Between any two components of size \sqrt{n} , the probability of having a link is a positive constant (independent of n).

$$\begin{aligned} & Pr[\exists \text{ a link between components } C_1, C_2, |C_i| > \sqrt{n}] \\ &= 1 - (1 - p(n))^{|C_1||C_2|} \\ &\leq 1 - \left(1 - \frac{\lambda}{n}\right)^n \approx 1 - e^{-\lambda}. \end{aligned}$$

Claim: Components of size $\leq \sqrt{n}$ connect to each other forming a connected component of size qn for some $0 < q < 1$. Such a component is our “Giant Component.”

Contagion and Diffusion

We next examine how gossips (or an epidemics) diffuse in a social network of n individuals. We will first consider a social network that is described by an Erdős-Rényi random graph $G(n, p(n))$ with link probability $p(n)$. In this network any individual is immune to a virus (of a specific kind) with probability π .

Let us assume that in this network an individual (chosen randomly among the ones without immunity) gets infected with this contagious virus. To model this process, we can proceed as follows: First, delete πn of the nodes at random, and identify the component containing the individual who was infected first.

Such a graph can be described by another ER graph:

$$G\left((1-\pi)n, \binom{n}{2} p(n) / \binom{(1-\pi)n}{2}\right).$$

We can now estimate the size of epidemic (as a fraction of the society): S .

$E[S]$

$$= \begin{cases} O\left(\frac{\ln n}{n}\right) & \text{if } p(1-\pi)n < 1 \\ \frac{1}{n}q^2(1-\pi)n + (1-q)\ln((1-\pi)n) & \text{if } 1 < p(1-\pi)n < \ln((1-\pi)n) \\ q^2(1-\pi)n & \text{if } \ln((1-\pi)n) < p(1-\pi)n \\ (1-\pi) & \end{cases}$$

where q denotes the fraction of nodes in the giant component of the graph with $(1-\pi)n$ nodes, i.e., it is a solution to the following fixed-point equation:

$$q = 1 - \exp[-q(1-\pi)np].$$

§§§

Ranking in a Social Network

Originally, Kleinberg proposed a simple structure to think about ranking in a directed graph by giving each node a value either as an *authority* or as a *hub*.

Directed Graph: A *directed graph* or *digraph* is a pair $G = (V, E)$, $E \subseteq V \times V$ consisting of:

- a set V , whose elements are vertices (as before),
- a set E of ordered pairs of vertices, called *arcs* or *directed edges*.

It differs from an undirected graph, in that the latter is defined in terms of unordered pairs of vertices, which are usually called edges. Thus we assume that the binary relation implied by E is still irreflexive, but no longer symmetric. Thus its adjacency matrix $A \in \{0, 1\}^{n \times n}$ is also not symmetric:

$$a_{uv} = \begin{cases} 1, & \text{if } (u, v) \in E; \\ 0, & \text{otherwise.} \end{cases}$$

Authorities: Nodes pointed to by highly ranked nodes.

Hubs: Nodes pointing to highly ranked nodes.

For each node (or page in a hyperlinked web of pages), estimate its value as a potential authority ($b(p) = \text{Authority Weight}$) and as a potential hub ($h(p) = \text{Hub Weight}$).

Thus

$$\begin{aligned} \forall_i h(i) &= \sum_j A_{ij} b(j) \\ \forall_j b(j) &= \sum_i A_{ij} h(i) \end{aligned}$$

In other words:

$$h = Ab \quad \text{and} \quad b^T = h^T A \quad (\equiv b = A^T h).$$

Thus

$$b = A^T A b \quad \text{and} \quad h = A A^T h.$$

One could attempt to solve these two fixed-point equations iteratively as follows: At stage $k + 1$ compute

$$b_{k+1} = A^T A b_k \quad \text{and} \quad h_{k+1} = A A^T h_k,$$

continuing the process until they converge (i.e., $|b_{k+1} - b_k| < \epsilon$ as well as $|h_{k+1} - h_k| < \epsilon$.) We will see later these two values b and h are eigenvectors of $A^T A$ and $A A^T$, respectively.

A much simpler version of this process also occurs in the so-called PageRank:

Each node is important if it is cited by other important nodes.

For each node j , we encode its *importance* in its *PageRank* value: $w(j)$.

$$w(j) = \sum_i \frac{w(i)}{d_{out}(i)} A_{ij},$$

where A is the adjacency matrix and d_{out} is the *out-degree* of a vertex. Thus, if we write

$$P_{ij} = \frac{A_{ij}}{d_{out}(i)},$$

then $P \in R^{n \times n}$ is a *stochastic matrix*.

$$w^T = w^T P,$$

and the iterative equation is

$$w_{k+1} = P w_k,$$

carried out until convergence: $|w_{k+1} - w_k| < \epsilon$.

Page's Random Walk Model: Choose a starting node at random.

In each step, "walker" follows one of the outgoing link selected uniformly randomly from the current node – and it moves to the node pointed to by the link. Thus it goes from the node i to j with a probability $P_{ij} = \frac{A_{ij}}{d_{out}(i)}$.

PageRank of i is the *limiting probability that the random walker will end up in node i* .

$$w^T = w^T P; \quad \text{solved iteratively: } w_{k+1}^T = w_k^T P.$$

Dangling Ends, Traps and Random Teleport: In order to avoid getting trapped in an ergodic component: jump (or teleport) to a random node with a probability $(1 - s)$ (and otherwise do the random walk with a probability s). Usually $0.8 \leq s \leq 0.9$.

$$w^T = s w^T P + \frac{1}{n} (1 - s) e^T; \quad \text{solved iteratively: } w_{k+1}^T = s w_k^T P + \frac{1}{n} (1 - s) e^T.$$

SSS