pp 35

$\boxed{\text{Lecture \# 7}}$

March 26 2013

1

GIANT COMPONENT.

(A) When $p(n) \ll \frac{\ln n}{n}$ — Edge probability

the Erdös-Rényi Graphs $G(n, p(n))$ is a.s. disconnected.

(B) In this regime [i.e. $p(n) \ll \frac{\ln n}{n}$]

$$X = \sum_{i=1}^{n} I_{i:\,iso}, \quad E[X] = n\left(1-p(n)\right)^{n-1}$$

$$E[X] = n\left(1-p(n)\right)^{\frac{1}{p(n)}\, p(n)(n-1)}$$

$$= n\, e^{-p(n)\cdot n} = n \cdot e^{-(1-\varepsilon)\ln n}$$

$$= n \cdot n^{-(1-\varepsilon)} = n^{\varepsilon} \to \infty \text{ as } n \to \infty$$

(C) Thus, when $p(n) \ll \frac{\ln n}{n}$ the graph has an arbitrarily large number of connected components.

We will study another interesting regime that gives rise to GIANT COMPONENTS.

TWO REGIMES:

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

For $\lambda < 1$, all components of the graph are "SMALL."

For $\lambda > 1$, one component of the graph is a "GIANT." (UNIQUE).

A unique giant component

A component that remains a constant fraction of individuals in the social network.

Best way to think about this structure:

BREADTH-FIRST SEARCH.

Think of two processes (related)

Assume $\lambda < 1$.

Graph Process $\rightarrow$ Z_k^G = # individuals at stage k of the graph.

Branching Process $\rightarrow$ Z_k^B = # individuals in a pure branching process.

$$Z_k^G < Z_k^B \quad \text{(e.g. triadic closure)}$$

Note

Expected # of children for a node

$$= n\, p(n) = n \cdot \frac{\lambda}{n} = \lambda$$

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

S_1 = # nodes in the Erdös - Rényi graph connected to individual 1.

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

Theorem

Let $p(n) = \frac{\lambda}{n}$ $(\lambda < 1)$

For all sufficiently large $a > 0$, we have

$$P\left(\max_{1 \le i \le n} |S_i| \ge a \ln n\right) \to 0 \quad \text{as } n \to \infty.$$

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

Proof: Omitted. □

Assume $\lambda > 1$.

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

$\boxed{\text{CLAIM}}$

$$Z_k^G \approx Z_k^B \quad \text{when } \lambda^k \le o(\sqrt{n}). \quad \square$$

Estimate the expected number of conflicts at stage $k+1$

Two of the "friends" at stage k have a common friend at stage $k+1$. (TRIADIC CLOSURE).

E[Number of conflicts at stage $k+1$]

$$= E\left[\binom{Z_k}{2}\binom{Z_k}{1}p^2\right] \le np^2\, E[Z_k^2]$$

$$= np^2\left\{E[Z_k]^2 + Var[Z_k]\right\}$$

$$\left(\because Z_k \sim \text{Poisson}(\lambda^k),\quad E[Z_k] = Var[Z_k] = \lambda^k\right)$$

$$= np^2\left(\lambda^{2k} + \lambda^k\right)$$

$$\lesssim n\frac{\lambda^2}{n^2}\cdot\lambda^{2k} = \frac{\lambda^{2(k+1)}}{n}$$

Thus conflicts become non-negligible when

$$\underline{\lambda^k \approx \sqrt{n}}.$$

THEOREM

Let $p(n) = \frac{\lambda}{n}$ $(\lambda > 1)$ in an Erdös-Rényi random graph $G(n, p(n))$. Then, there exists some $c > 0$ such that

$$\Pr[\exists \text{ a component of size} > c\sqrt{n} \text{ nodes}] \longrightarrow 1 \quad \text{as } n \to \infty$$

Thus, between any two components of size $\sqrt{n}$, the probability of having at least one link

$$\Pr[\text{there exists a link between } C_1 \,\&\, C_2]$$

$$\geq 1 - (1 - p(n))^{|C_1| \times |C_2|}$$

$$= 1 - \left(1 - \frac{\lambda}{n}\right)^{c^2 n} = 1 - \left(1 - \frac{\lambda}{n}\right)^{\frac{n}{\lambda} \cdot c^2}$$

$$= 1 - e^{-c^2 \lambda} > 0$$

A positive constant independent of n.

Thus, components of size $\leq \sqrt{n}$ connect to each other forming a connected component of $\underbrace{\text{size } qn}$ for some $0 < q < 1$.

GIANT COMPONENT.

CONTAGION AND DIFFUSION

ASSUME:

- A social network of n individuals.
- In which a randomly chosen individual is infected with a contagion virus.
- The social network is described by an Erdös-Rényi random graph with link probability $= p(n)$
- Any individual is immune to the virus with probability $= \pi$

Generate Erdös-Rényi random graph $G(n, p)$
Delete πn of the nodes at random.
Identify the component that the initially infected individual lies in ...

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

Three regimes:

(I) $p(1-\pi)\,n < 1.$

$E[\text{Size (as a fraction)}] \leq \frac{\ln n}{n} \approx 0$

(II) $1 < p(1-\pi)\,n < \log[(1-\pi)n]$

$E[\text{Size (as a fraction)}]$

$$= \frac{q\,q\,(1-\pi)\,n + (1-q)\ln n}{n} = q^2(1-\pi)$$

(III) $p(1-\pi)\,n > \log[(1-\pi)n]$

$E[\text{Size (as a fraction)}] = (1-\pi)$

q = Fraction of nodes in the giant component of the random graph with $(1-\pi)\,n$ nodes:

$$q = 1 - e^{-q(1-\pi)np}.$$

RANDOM SURFER MODEL.

Imagine a web surfer bouncing along randomly following the hyper-link graph of the web.

When the surfer arrives at a node he chooses at random, ~~the~~ hyper-links (directed edge) to a new node.

Asymptotically, the proportion of time the random surfer spends on a given node/page is a measure of
RELEVANCE (RELATIVE IMPORTANCE) of that node.

DANGLING NODES → SINKS
or PERIODICITY IN THE GRAPH
may get the surfer trapped in a limit state or a limit cycle.

→ STOCHASTIC TELEPORTATION

PAGE RANK.

Each node is <u>important</u> if it is cited by other <u>important</u> nodes.

Node $j \Rightarrow w(j)$ = Page Rank Value.

A = Adjacency Matrix
$d_{out}^{(i)}$ = Out-degree of node i

$$w(j) = \sum \frac{\omega(i)}{d_{out}(i)} A_{ij}$$

Let $P_{ij} = \dfrac{A_{ij}}{d_{out}(i)} \Rightarrow P$ = Stochastic Matrix.

$$\omega^T = \omega^T P.$$

WITH TELEPORTATION:

Jump to a random node with prob $(1-s)$, $0.8 \le s \le 0.9$

$$\omega^T = s\,\omega^T P + \frac{(1-s)}{n} e^T$$

Iterative Algorithm:

$$\boxed{W_{k+1}^T = s\,W_k^T P + \frac{(1-s)}{n} e^T}$$