Random Graphs

Social and Technological Networks

Rik Sarkar

University of Edinburgh, 2016.

Graph

- V: set of nodes
- n = |V| : Number of nodes

- E: set of edges
- m=|E| : Number of edges

- If edge a-b exists, then a and b are called neighbors

Graph

- How many edges can a graph have?

Graph

- How many edges can a graph have?

$$\binom{n}{2} \quad \text{OR} \quad \frac{n(n-1)}{2}$$

- In big O?

Graph

- How many edges can a graph have?

$$\binom{n}{2} \quad \mathrm{OR} \quad \frac{n(n-1)}{2}$$

$$O(n^2)$$

Random graphs

- Most basic, most unstructured graphs
- Forms a baseline
  - What happens in absence of any influences
    - Social and technological forces
- Many real networks have a random component
  - Many things happen without clear reason

Erdos – Renyi Random graphs

Erdos – Renyi Random graphs

$$\mathcal{G}(n, p)$$

- n: number of vertices
- p: probability that any particular edge exists

  - Take V with n vertices
  - Consider each possible edge. Add it to E with probability p

Expected number of edges

- Expected total number of edges
- Expected number of edges at any vertex

Expected number of edges

- Expected total number of edges $\binom{n}{2}p$

- Expected number of edges at any vertex

$$(n-1)p$$

Expected number of edges

- For $$p = \frac{c}{n-1}$$

- The expected degree of a node is : ?

Isolated vertices

- How likely is it that the graph has isolated vertices?

Isolated vertices

- How likely is it that the graph has isolated vertices?
- What happens to the number of isolated vertices as p increases

Terminology of high probability

- Something happens with high probability if

$$p \geq \left(1 - \frac{1}{\mathrm{poly}(n)}\right)$$

- Where poly(n) means a polynomial in n
- A polynomial in n is considered reasonably 'large'

Probability of Isolated vertices

- Isolated vertices are

- Likely when: $p < \frac{\ln n}{n}$

- Unlikely when: $p > \frac{\ln n}{n}$

- Let's deduce

Useful inequalities

$$\left(1+\frac{1}{x}\right)^x \leq e$$

$$\left(1-\frac{1}{x}\right)^x \leq \frac{1}{e}$$

Union bound

- For events A, B, C …

- Pr[A or B or C ...] $\leq$ Pr[A] + Pr[B] + Pr[C] + ...

- Theorem 1:
- If $$p = (1+\epsilon)\frac{\ln n}{n-1}$$
- Then the probability that there exists an isolated vertex $$\leq \frac{1}{n^{\epsilon}}$$
- Thus for large n, w.h.p there is no isolated vertex
- Expected number of isolated vertices is miniscule

- Theorem 2
- For $p = (1-\epsilon)\dfrac{\ln n}{n-1}$
- Probability that vertex v is isolated $\geq \dfrac{1}{(2n)^{1-\epsilon}}$

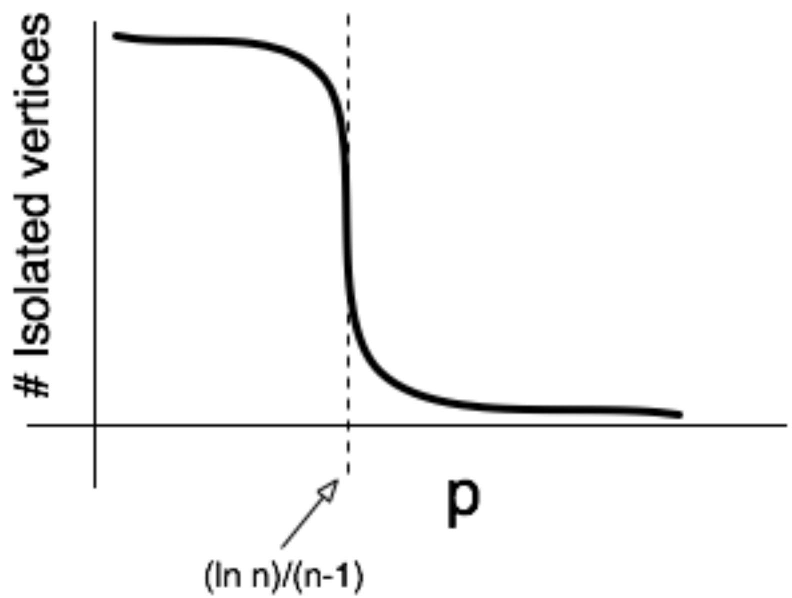
- Theorem 2
- For $p = (1-\epsilon)\dfrac{\ln n}{n-1}$
- Probability that vertex v is isolated $\geq \dfrac{1}{(2n)^{1-\epsilon}}$
- Expected number of isolated vertices:

$$\geq \frac{n}{(2n)^{1-\epsilon}} = \frac{n^{\epsilon}}{2}$$

Polynomial in n

Threshold phenomenon: Probability or number of isolated vertices

- The tipping point



Clustering in social networks

- People with mutual friends are often friends
- If A and C have a common friend B
  - Edges AB and BC exist
- Then ABC is said to form a *Triad*
  - Closed triad : Edge AC also exists
  - Open triad: Edge AC does not exist
- Exercise: Prove that any connected graph has at least n triads (considering both open and closed).

Clustering coefficient (cc)

- Measures how tight the friend neighborhoods are: frequency of closed triads
- cc(A) fractions of pairs of A's neighbors that are friends
- Average cc : average of cc of all nodes
- Global cc : ratio

$$\frac{\text{\# closed triads}}{\text{\# all triads}}$$

Global CC in ER graphs

- What happens when p is very small (almost 0)?
- What happens when p is very large (almost 1)?

Global CC in ER graphs

- What happens at the tipping point?

Theorem

- For $p = c\dfrac{\ln n}{n}$

- Global cc in ER graphs is vanishingly small

$$\lim_{n \to \infty} cc(G) = \lim_{n \to \infty} \frac{\#\ \text{closed triads}}{\#\ \text{all triads}} = 0$$

Avg CC In real networks

- Facebook (old data) ~ 0.6
  - https://snap.stanford.edu/data/egonets-Facebook.html
- Google web graph ~0.5
  - https://snap.stanford.edu/data/web-Google.html

- In general, cc of ~ 0.2 or 0.3 is considered 'high'
  - that the network has significant clustering/ community structure