Basics and Random Graphs

Social and Technological Networks

Rik Sarkar

University of Edinburgh, 2018.

Webpage

- Check it regularly
- Announcements
- Lecture slides, reading material
- Do exercises. (exercise 0 up now)
- Preliminary reading for next week will be up.
- Set up ipython notebook and visualize some graphs
  - Setup instructions on web page

Today

- What are random graphs?
  - How can we define "random graphs"?
- Some properties of random graphs

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 in an ER graph

- 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 $p = \dfrac{c}{n-1}$

- 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?

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 …] ≤ 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 $$\leq \frac{1}{n^{\epsilon}}$$

Terminology of high probability

- Poly(n) means a polynomial in n
- A polynomial in n is considered reasonably ‘large’
  - Whereas something like constant, or log n is considered ‘small’

- Something happens with high probability if

$$\Pr[event] \geq \left(1 - \frac{1}{\text{poly}(n)}\right)$$

- Thus for large n, w.h.p there is no isolated vertex
- Expected number of isolated vertices is miniscule

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

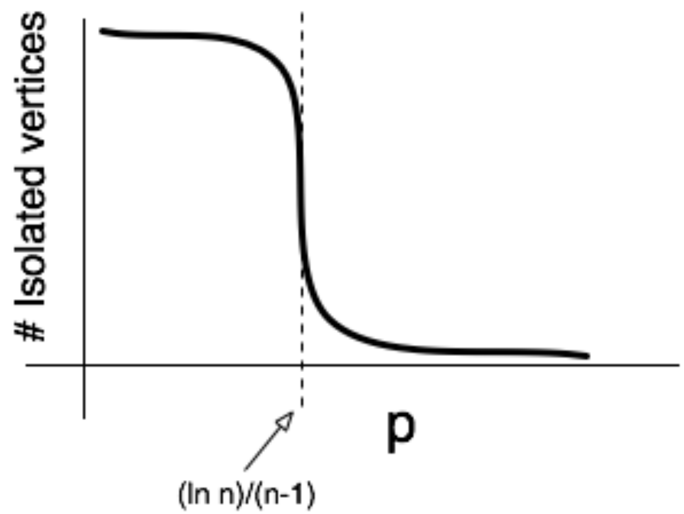
- Theorem 2
- For $p = (1-\epsilon)\frac{\ln n}{n-1}$
- Probability that vertex v is isolated $\geq \frac{1}{(2n)^{1-\epsilon}}$
- Expected number of isolated vertices:

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

Polynomial in n

Threshold phenomenon: Probability or number of isolated vertices

- The tipping point, phase transition



- Common in many real systems

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

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

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

CC of a graph model

- If we are given a model of graphs
  - Clustering is considered significant if
  - CC is bounded from below by a constant
    - E.g. $cc(G) > 0.1$
    - Note that $cc(G) > 1/n$ does not help, since this can be very small

- Example problems:
  - What can you say about CC of Trees?
  - Complete graphs?
  - Grids?
  - Grids with diagonals added?

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 $p = c\frac{\ln n}{n}$

- Global cc in ER graphs is vanishingly small

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

- In other words, there is no constant c
  - Such that cc(ER-graph) > c
  - At the tipping point

Random graphs: Emergence of giant component

- Suppose N_G is the size of the largest connected component in an ER graph
- How does N_G/N change with p?
- When is N_G/N at least a constant?
  - (giant component: at least a constant fraction of nodes)

Giant component

- When $p = (1-\varepsilon)/n$
  - W.h.p no GC, components of size $O(\log n)$
- When $p = (1+\varepsilon)/n$
  - W.h.p GC exists, where $N_G/N \sim \varepsilon$
- When $p = 1/n$
  - W.h.p Largest component has size $n^{2/3}$