Basics and Random Graphs

Social and Technological Networks

Rik Sarkar

University of Edinburgh, 2017.

Webpage

- Check it regularly
- Announcements
- Lecture slides, reading material
- Do exercises 1.

Today

- Some basics of graph theory
  - Wikipedia is a good resource for basics
- Typical types of graphs & networks
- What are random graphs?
  - How can we define "random graphs"?
- Some properties of random graphs

Graph

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

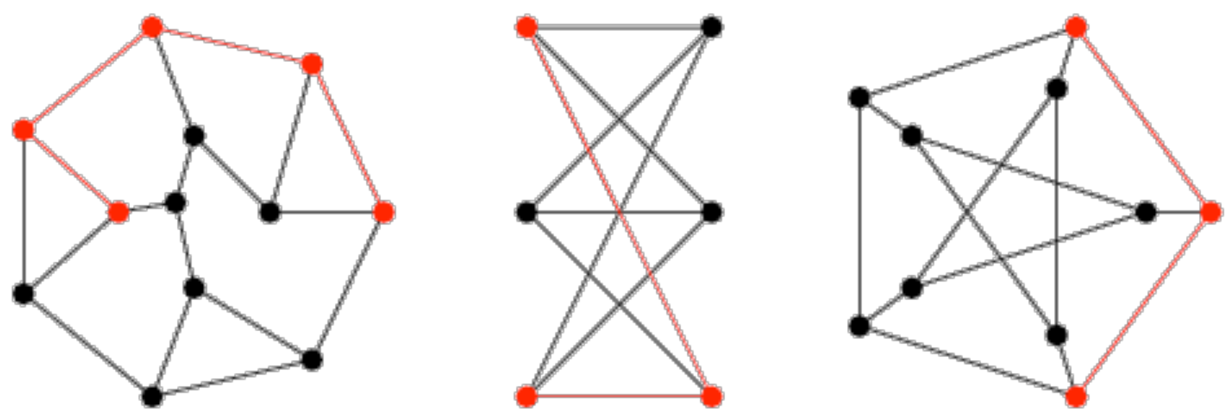
- E: set of edges
- m=|E| : Number of edges

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

Walks

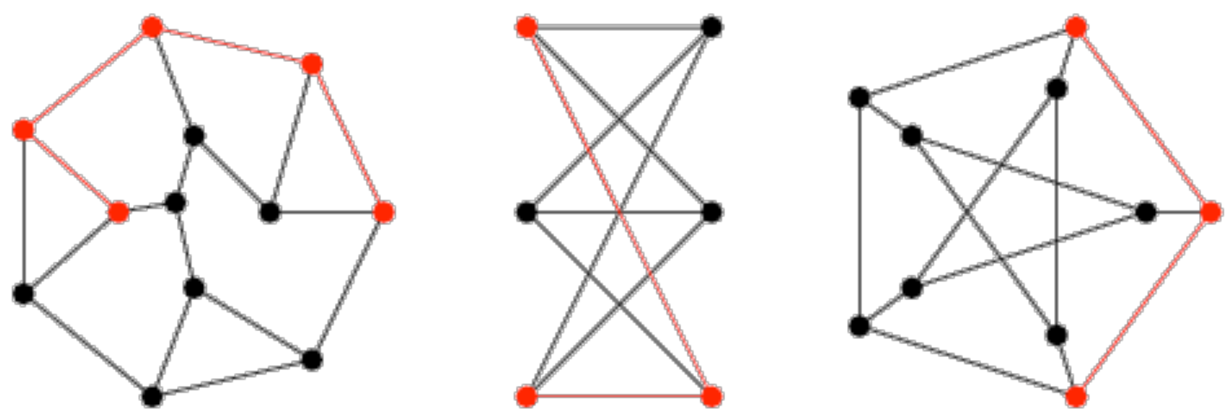
- A sequence of vertices $v_1, v_2, v_3, \ldots$
- Where successive vertices are neighbors

$$v_i, v_{i+1}, (v_i, v_{i+1}) \in E$$



Paths

- Walks without any repeated vertex



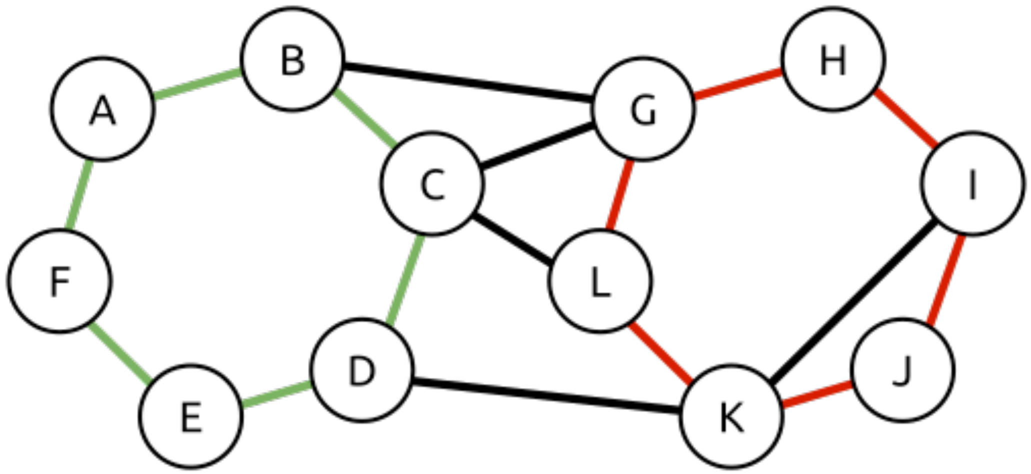
Exercises

- At most how many walks there can be on a graph?

- At most how many paths can there be on a graph?

Cycle

- A walk with the same start and end vertex



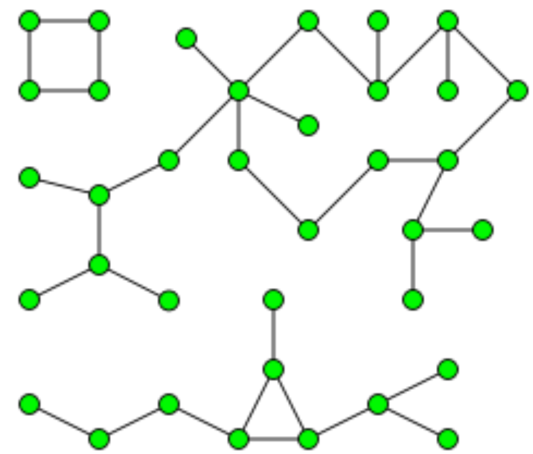
Subgraph of G

- A graph H with a subset of vertices and edges of G
  - Of course, for any edge (a,b) in H, vertices a and b must also be in H

- Subgraph induced by a subset of vertices $X \subseteq V$
  - Graph with vertices X and edges between nodes in X

Connected component

- A subgraph where
  - Any two vertices are connected by a path



- A connected graph
  - Only 1 connected component

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

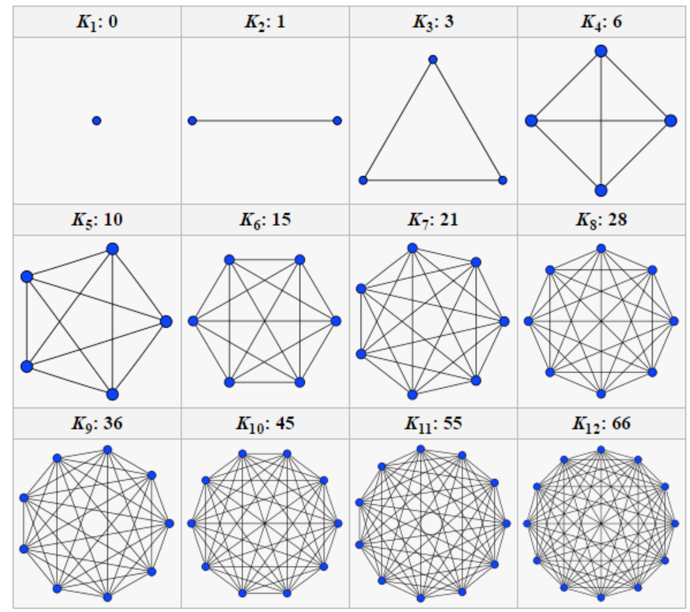
- How many edges can a graph have?

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

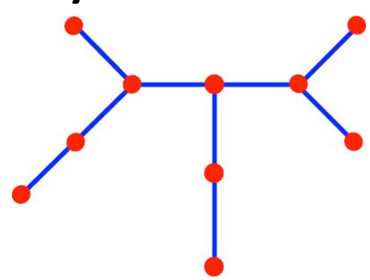
$$O(n^2)$$

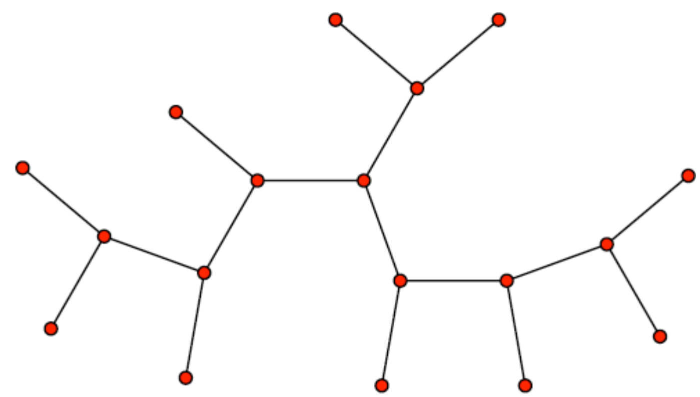
Some typical graphs

- Complete graph
  - All possible edges exist



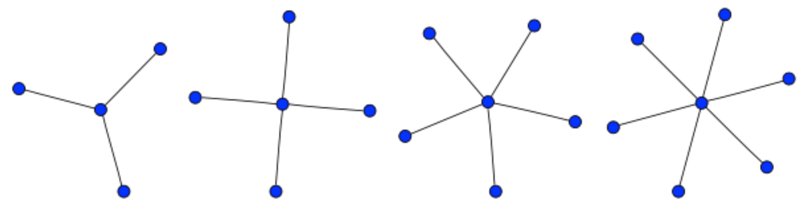
- Tree graphs
  - Connected graphs
  - Do not contain cycles



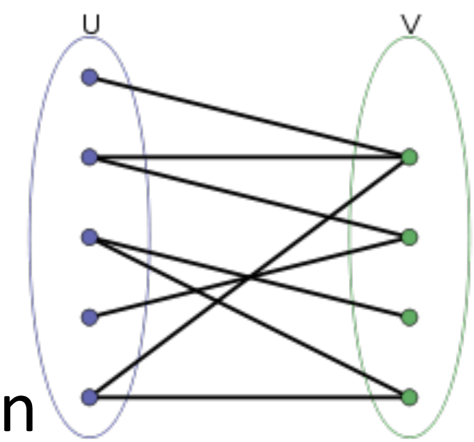


Typical graphs

- Star graphs



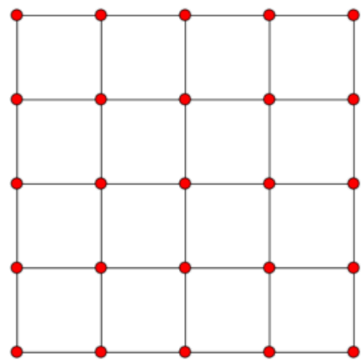
- Bipartite graphs
  - Vertices in 2 partitions
  - No edge in the same partition



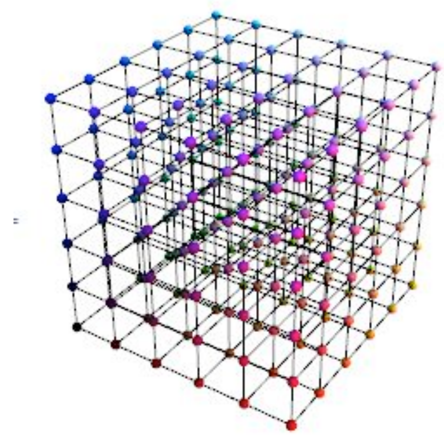
Typical graphs

- Grids (finite)

  - 1D grid (or chain, or path)



  - 2D grid



  - 3D grid

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

- 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

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

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

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

- Something happens with high probability if

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

- Where poly(n) means a polynomial in n
- A polynomial in n is considered reasonably 'large'
  - Whereas something like log n is considered 'small'

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

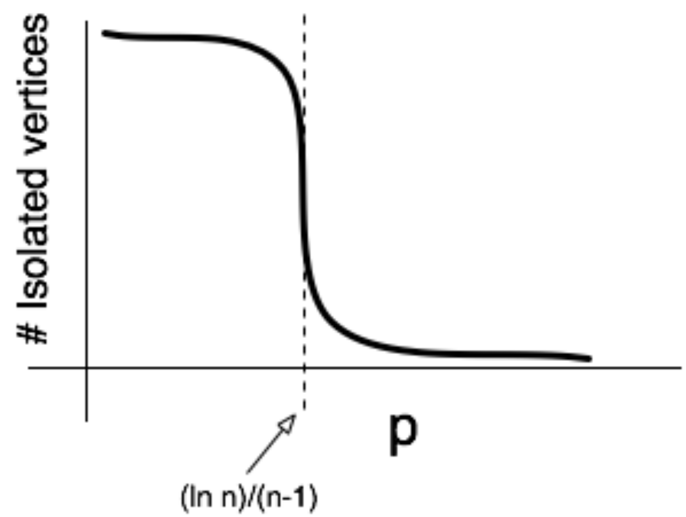
- Theorem 2
- For $p = (1 - \epsilon)\dfrac{\ln n}{n - 1}$
- Probability that vertex v is isolated $\geq \dfrac{1}{(2n)^{1-\epsilon}}$
- Expected number of isolated vertices:

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

Polynomial in n

Threshold phenomenon: Probability or number of isolated vertices

- The tipping point, 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

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

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

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

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