Basics and Random Graphs continued

Social and Technological Networks

Rik Sarkar

University of Edinburgh, 2017.

- Random graphs on jupyter notebook
- Solution to exercises 1 is out
- If your BSc/MSc/PhD work is related, feel free to discuss relevant project
- Discuss with your supervisor
- Class-only students are also welcome to discuss projects

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

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

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

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

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?

Configuration model of Random graphs

- Suppose we want a graph that is random
- But has given degree for each vertex:

$$d_1, d_2, d_3, \ldots d_n$$

- At each vertex i we d_i *open-edges*
- Pair up the edges randomly
- If all degrees = d
  - Graph is called d-regular

Distances in graphs

- Paths
- Shortest paths
- BFS
- Metrics

Path

- *Length* of a path or walk is the number of edges it traverses
  - In an unweighted graph
- In a weighted graph (edges have numeric weights)
  - Length or weight of a path is the sum of weights
- In a directed graph
  - A walk or path must respect the directions

Distance

- Distance between any two nodes in a graph is the length of the shortest path between them

- Diameter of a graph:
  – Distance between the farthest pair of nodes in the graph

Metric

- A distance measure d is a metric if:
  - $d(x,y) \geq 0$
  - $d(x,y) = 0$ iff $x=y$
  - $d(x,y) = d(y,x)$
  - $d(x,z) \leq d(x,y) + d(y,z)$

The undirected graph distance

- Is a metric
- In unweighted graphs, all values are integers

Finding distance between two nodes in a graph

- Breadth first search
- Dijkstra's shortest path algorithm

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 $n^{2/3}$

Ball

- A ball of radius r at vertex v:
  - The set of all nodes within distance r from v
  - The first r layers of a BFS from v
- Usually written as
  - B(v,r) or
  - $B_r(v)$
- In a metric space:
  - The set of all points within distance r of v
- Sphere $S_r(v)$: set of points at distance exactly r from v

Asymptotic notations

- Big O: $f(n) = O(g(n))$
  - For large enough n,
  - There is a constant c such that $f(n) \leq c.g(n)$
- Big Omega: $f(n) = \Omega(g(n))$
  - For large enough n,
  - There is a constant c such that $f(n) \geq c.g(n)$
- Theta : $f(n) = \Theta(g(n))$
  - Both O and Ω

Edge Expansion

- How fast the ‘boundary’ expands relative to ‘volume’ or ‘size’ of a subset
- Boundary of S :
  - $e^{out}(S)$: edges with exactly one end-point in S
- Expansion:

$$\alpha = \min_{S \subseteq V} \frac{|e^{out}(S)|}{\min(|S|, |\overline{S}|)}$$

Expansion

$$\alpha = \min_{S \subseteq V} \frac{|e^{out}(S)|}{\min(|S|, |\overline{S}|)}$$

- Equivalently:

$$\alpha = \min_{|S| \leq n/2} \frac{|e^{out}(S)|}{|S|}$$

Expanders

- A class of graphs with expansion at least a constant

$$\alpha \geq c$$

  - For some constant c

Are the following graphs expanders?

- A chain
- A balanced binary tree
- A grid

Examples of expanders

- Random d-regular graphs for d>3
- ER graphs for large enough p

Expanders have small diameter

- A graph with degrees ≤ d and expansion ≥ α
- Has diameter

$$O(\frac{d}{\alpha} \lg n)$$

Other properties

- Expanders are well connected
- Usually sparse (number of edges much smaller than n^2)
- Diffusion processes spread fast in an expander
- Random walks mix fast (achieve steady state)
- …

Metric examples

- L_2
- L_1
- L_P
- L_∞
- Grid
- Tree

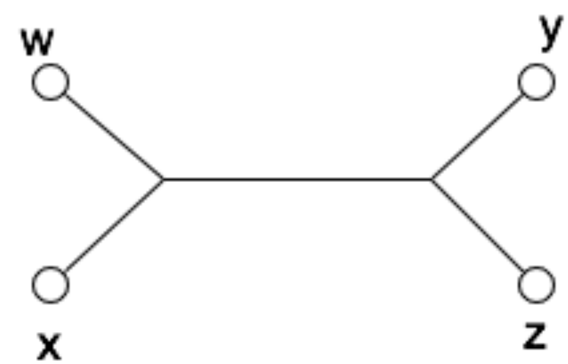
Metric growth

- Consider the number of nodes in B(v,r)
  - That is, |B(v,r)|
  - How does this grow as a function of r?

- For 2D grid?
- For 3D grid?
- For Balanced binary tree?

Metric examples

- Grid
- Tree
  - Test for tree metric
  - Any 4 points (vertices) can be to satisfy:
    - d(w,x) + d(y,z) ≤ d(w,y) + d(x,z) ≤ d(w,z) + d(x,y)
    - And d(w,y) + d(x,z) = d(w,z) + d(x,y)



Doubling dimension