Random Graphs continued

Social and Technological Networks

Rik Sarkar

University of Edinburgh, 2016.

- Random graphs on ipython notebook

Recap: clustering coefficients

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

- Large ER graphs have cc = 0
- Social graphs may have cc = 0.2 to 0.6

CC of a graph model

- If we are given a model of a graph
  - Clustering is considered significant if
  - CC is bounded from below by a constant

- Example problems:
  - What can you say about CC of a Tree?
  - A complete graph?
  - A grid with diagonals added?

Distances in graphs

- Paths
- Shortest paths
- BFS
- Metrics

Path

- A walk is a sequence of vertices v_1, v_2,…
  - Where any (v_i, v_{i+1}) is an edge in the graph
- A path is a walk where no vertex repeats
- *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

- Connected component
  - A maximal subgraph with a path between any two vertices

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

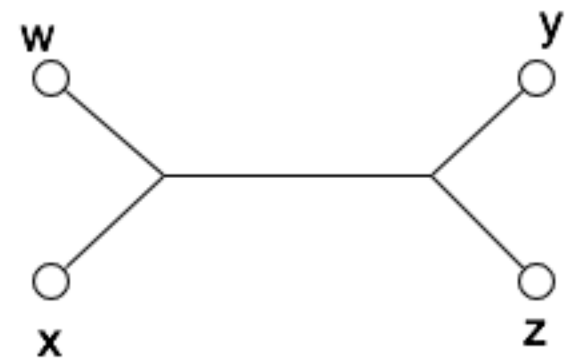
- Is a metric
- In unweighted graphs, all values are integers

Metric examples

- L2
- L1
- Grid
- Tree

Metric examples

- L2
- L1
- Grid
- Tree



  - Test for tree metric
  - Any 4 points (vertices) can be ordered as w,x,y,z to satisfy:
    - $d(w,x) + d(y,z) \le d(w,y) + d(x,z) \le d(w,z) + d(x,y)$
    - And $d(w,y) + d(x,z) = d(w,z) + d(x,y)$

Finding distance between two nodes in a graph

- Breadth first search
- Dijkstra’s shortest path algorithm

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

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

Edge Expansion

- How fast the 'boundary' expands relative to 'volume' 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

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 $\leq$ d and expansion $\geq$ 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)
- …

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

- 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 $n^{2/3}$