CS 416A: Small World Networks

Raj Venkat

1 Clustering Coefficient of $G(n,p)$ ER Random Graph

Consider node i. Clustering coefficient of i can be calculated as:

$$CC_i = \frac{\text{No. of edges between i's neighbors}}{\text{No. of possible edges between i's neighbors}}$$

The expected degree of node i (in other words, the expected number of neighbors i has) is

$$k_i = (n-1)p$$

The expected number of edges between k_i nodes in $G(n,p)$ can be calculated as

$$\binom{k_i}{2}p$$

and the number of possible edges between k_i nodes in $G(n,p)$ is simply

$$\binom{k_i}{2}$$

Therefore, the clustering coefficient for node i is

$$CC_i = \frac{\binom{k_i}{2}p}{\binom{k_i}{2}} = p$$

This is independent of i and n, the number of nodes. The global clustering coefficient therefore is

$$CC(G) = CC_i = p$$

2 Average Shortest Path Length for ER Random Graph

2.1 Existence of Giant Component for $\bar{k} \geq 1$

Lemma. *For average node degree $\bar{k} \geq 1$, a node i is in the giant component with high probability.*

Heuristic Argument. Let $G(n,p)$ be an ER random graph with n vertices. Let $i \in G$ and $j \in G$ be two vertices in the graph. Let u be the probability that a randomly chosen node in G is not part of the giant component (GC). Then, if i is not in the giant component, one of the following must hold true:

1. i and j do not share an edge (with probability $(1-p)$)
2. i and j share and edge AND j is not in the GC (with probability pu)

For each $i \notin GC$, there are $(n-1)$ options for j. Hence, we can write $u = (1 - p + pu)^{n-1}$. Now, average degree in an ER graph, $\bar{k} \approx pn$, therefore $p = \bar{k}/n$. Substituting above,

$$u = \left(1 - \frac{\bar{k}}{n}(1-u)\right)^{n-1}$$

Taking log on both sides,

$$\log u = (n-1)\log\left(1 - \frac{\bar{k}}{n}(1-u)\right)$$

For small x, $\log 1 + x \approx x$. Thus,

$$\log u \approx -\frac{\bar{k}(n-1)(1-u)}{n}$$

For large n, $n - 1 \approx n$.

$$\log u \approx -\bar{k}(1-u) \implies u = \mathrm{e}^{-\bar{k}(1-u)}$$

Let $v = 1 - u$ be the probability that a node is in the GC.

$$v = 1 - \mathrm{e}^{-\bar{k}v}$$

Plot $f(v) = v$ and $g(v) = \mathrm{e}^{-\bar{k}v}$ (left figure). For $\bar{k} > 1$, there is one positive non-trivial solution.

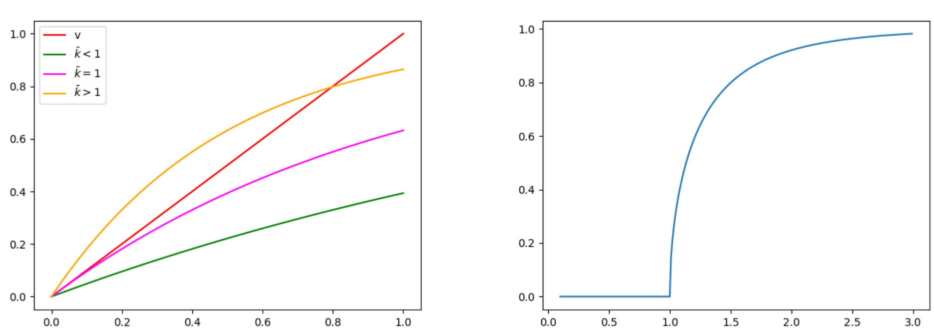

Figure 1: Graphing solutions to $v = 1 - \mathrm{e}^{-\bar{k}v}$

Plot the positive real-valued solutions for v, at various $\bar{k}$ (right figure). Probability v that a node is in the giant component goes to 1 as $\bar{k} > 1$. □

2.2 Average Shortest Path Length

Having shown that for $\bar{k} > 1$, most nodes are connected with high probability, we can calculate the average shortest path length as follows. We know that an ER graph exhibits low clustering, therefore it is safe to assume that in general, given a node i, its neighbors' neighbors are not i's neighbors.

Therefore, from a given node, the number of nodes reachable at a distance of 1 is $\bar{k}$, at a distance of 2 is $\bar{k}^2$ and so on, until for some distance s, all nodes are reached. Thus we calculate the diameter s as follows:

$$\bar{k}^s = n$$

$$s \ln \bar{k} = \ln n$$

$$s = \frac{\ln n}{\ln \bar{k}}$$

As the clustering coefficient is very low, the length of most shortest paths will be approximately the diameter.

3 Clustering Coefficient of Watts-Strogatz Network

We first calculate the clustering coefficient for a $WS(n, k, 0)$ graph (i.e. a WS-graph with no rewiring), and then generalize it to the graph family $WS(n, k, p)$ where each edge is rewired with a probability p. We first calculate the clustering coefficient for a single vertex in $WS(n, k, 0)$, and as all vertices are structurally identical, $CC(G) = cc(v \in V(G))$.

Let $G = WS(n, k, 0)$. Let $v \in G$ be a random node from G, and let $N(v)$ denote the set of v's neighbors. We say that the subgraph H formed by nodes in $N(v)$ and the edges between them is the subgraph *induced by* $N(v)$.

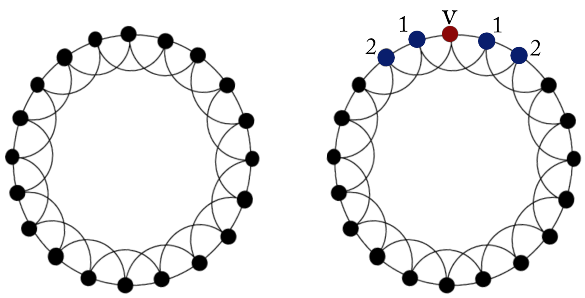

Figure 2: WS(n,2,0) and neighborhood $N(v)$

Now recall that $N(v)$ consists of v's $k/2$ left hand neighbors and $k/2$ right hand neighbors (Fig. 2). Let vertices on the right side of v be labeled from 1 to $k/2$, then we see that the i^{th} vertex, has $k/2 - i$ right hand neighbors in H, and $k/2 - 1$ left hand neighbors in H. Similarly, a vertex on the left of v will have $k/2 - i$ left hand neighbors in H, and $k/2 - 1$ right hand neighbors in H.

The degree of the i^{th} vertex on either side of v is $(k-i-1)$. Let the degree of a node u be denoted k_u. Then, total number of edges in H is given by

$$|E(H)| = \frac{1}{2}\sum_{u\in H} k_u$$

$$= \sum_{i=1}^{k/2}(k-i-1)$$

$$= \frac{3}{8}k(k-2).$$

As the number of nodes in H, $|V(H)| = k$, clustering coefficient for vertex v is

$$cc(v) = \frac{|E(H)|}{\binom{|V(H)|}{2}} = \frac{\frac{3}{8}k(k-2)}{\frac{1}{2}k(k-1)} = \frac{3(k-2)}{4(k-1)}.$$

Finally, in the general case of $WS(n,k,p)$, for each pair of neighbors in $N(v)$, there must be two edges - one from each - terminating in v and one edge between the two neighbors, thus forming a triangle, in order to count towards the global CC, and each of these edges exists with probability $(1-p)$. Thus,

$$CC(WS(n,k,p)) = \frac{3(k-2)}{4(k-1)}(1-p)^3$$

4 Web Demos

Giant Component and Phase Transition - `http://www.netlogoweb.org/launch#http://ccl.northwestern.edu/netlogo/models/models/Sample%20Models/Networks/Giant%20Component.nlogo`

Watts Strogatz Networks - `http://www.netlogoweb.org/launch#http://www.netlogoweb.org/assets/modelslib/Sample%20Models/Networks/Small%20Worlds.nlogo`