School of Information Sciences
University of Pittsburgh

TELCOM2125: Network Science and Analysis

Konstantinos Pelechrinis

Spring 2015

Figures are taken from:
M.E.J. Newman, “Networks: An Introduction”

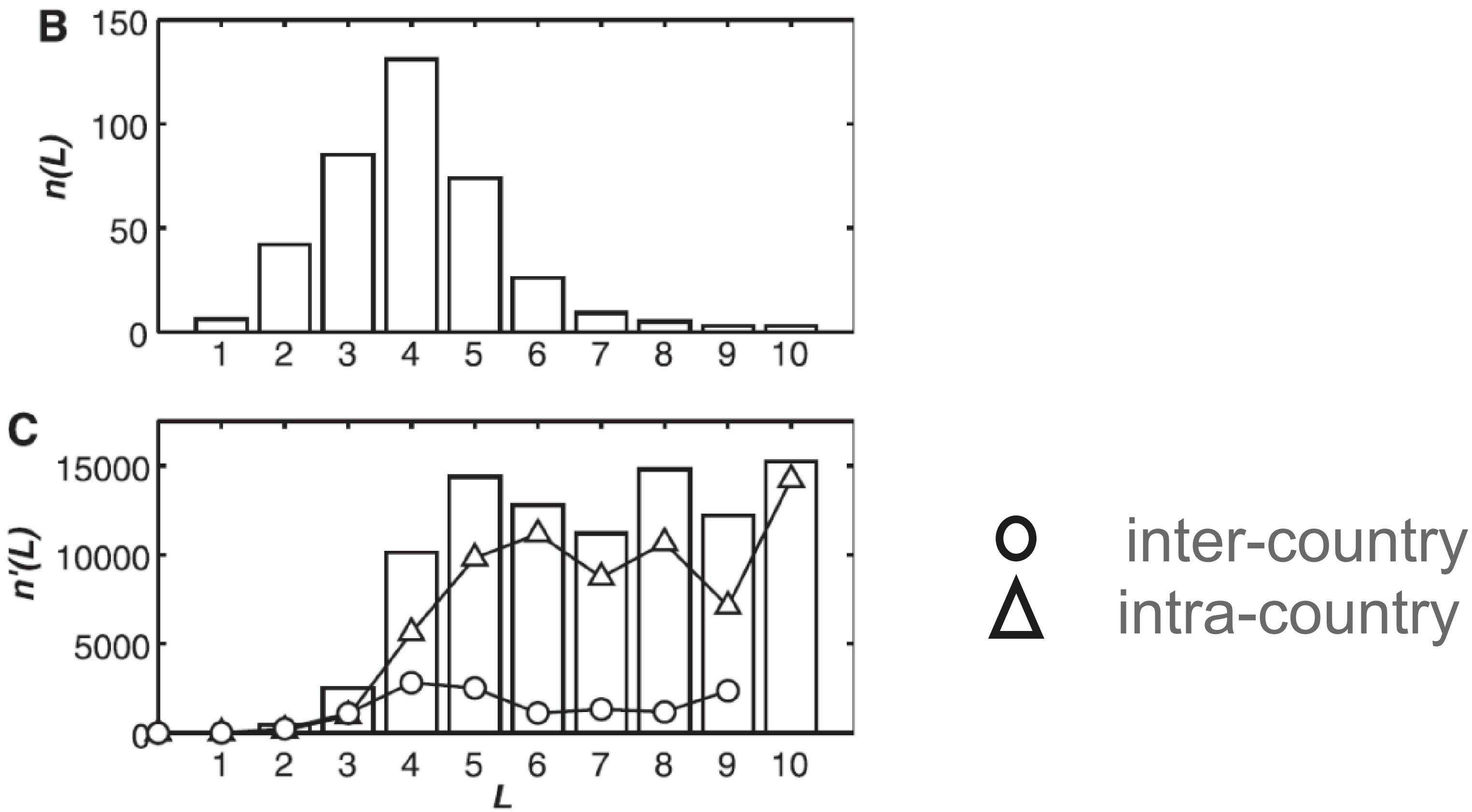
Part 8: Small-World Network Model

Small World-Phenomenon

- **Milgram's experiment**
  - Given a target individual - stockbroker in Boston - pass the message to a person you know on a first name basis who you think is *closest* to the target
- **Outcome**
  - 20% of the the chains were completed
  - Average chain length of completed trials ~ 6.5
- **Dodds, Muhamad and Watts have repeated this experiment using e-mail communications**
  - Completion rate is lower, average chain length is lower too

Small World-Phenomenon

- **Are these numbers accurate?**
  - What bias do the uncompleted chains introduce?



Source: An Experimental Study of Search in Global Social Networks: Peter Sheridan Dodds, Roby Muhamad, and Duncan J. Watts (8 August 2003); Science 301 (5634), 827.

Clustering in real networks

- **What would you expect if networks were completely cliquish?**
  - Friends of my friends are also my friends
  - What happens to small paths?

- **Real-world networks (e.g., social networks) exhibit high levels of transitivity/clustering**
  - But they also exhibit short paths too

Clustering coefficient

- **The network models that we have seen until now (random graph, configuration model and preferential attachment) do not show any significant clustering coefficient**
  - For instance the random graph model has a clustering coefficient of c/n-1, which vanishes in large networks

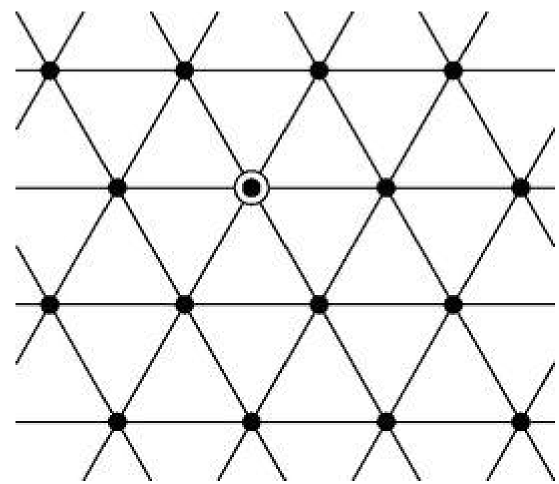
- **However, it is easy to find networks that have high clustering coefficient independent of the network size**

Triangular lattice

- For instance, consider a triangular lattice

- Due to symmetry we can consider a random vertex
  - Clustering coefficient gives the probability that two neighbors of the vertex under consideration are themselves friends

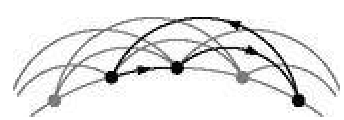
- Every vertex has six neighbors and hence there are 15 pairs of neighbors
  - From them 6 are connected
    - ✓ Hence, the clustering coefficient is 0.4
    - ✓ Independent of the network size

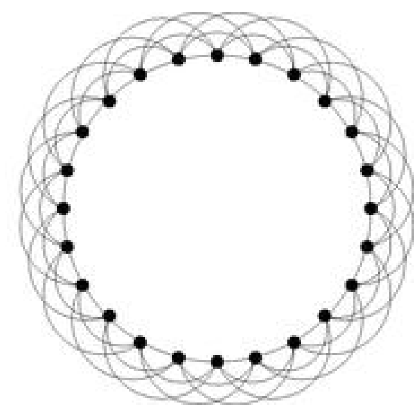


Circle model

- **In this model the vertices are arranged to a circle**
  - Each node is connected to its c nearest vertices
    - ✓ Fixed degree for all nodes
- **A triangle in this network requires two edge traversals at the same direction on the circle and one at the opposite**
  - The final/opposite step can span at most c/2 vertices
  - Hence, the number of triangles for a given node is given by the number of distinct ways of choosing the 2 *forward* target vertices from the c/2 possibilities: $\binom{c/2}{2} = \frac{1}{4}c(\frac{c}{2}-1)$
  - The number of connected triples per vertex is:

$$\binom{c}{2} = \frac{1}{2}c(c-1)$$





Circle model

- **Hence, the clustering coefficient of the circle model is:**

$$C = \frac{\frac{1}{4}nc(\frac{c}{2}-1)\times 3}{\frac{1}{2}nc(c-1)} = \frac{3(c-2)}{4(c-1)}$$

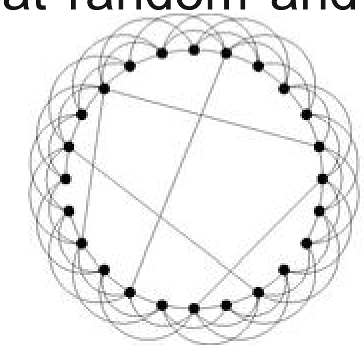
- **The clustering coefficient is not constant as in the triangular lattice but it takes values between 0 (when c=2) and 0.75 (when c→∞)**
  - However, note that C is independent of n

- **While this model exhibits large clustering coefficient it has two problems**
  - Degree distribution
  - *"Large-worlds"* → The average shortest path is not small as in real networks

Small-world models

- **Random graphs exhibit small paths but not clustering**
  - Why not try to combine these two models together?

- **The small-world model (Watts and Strogatz 1998) tries to do exactly this**
  - We start with a circle model of n vertices in which every vertex has a degree of c
  - We go through each of the edges and with some probability p we *rewire* it
    - ✓ Remove this edge and pick two vertices uniformly at random and connect them with a new edge
      - ○ Shortcut edge



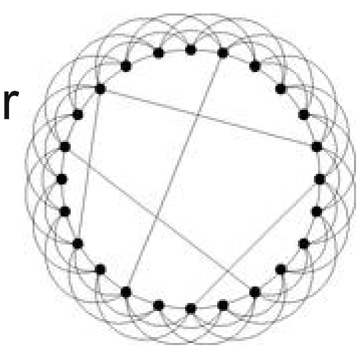
Small-world models

- **The parameter p controls the interpolation between the circle model and the random graph**
  - p=0 → ordered situation/circle model
  - p=1 → random graph
  - Intermediate values of p give networks somewhere in between

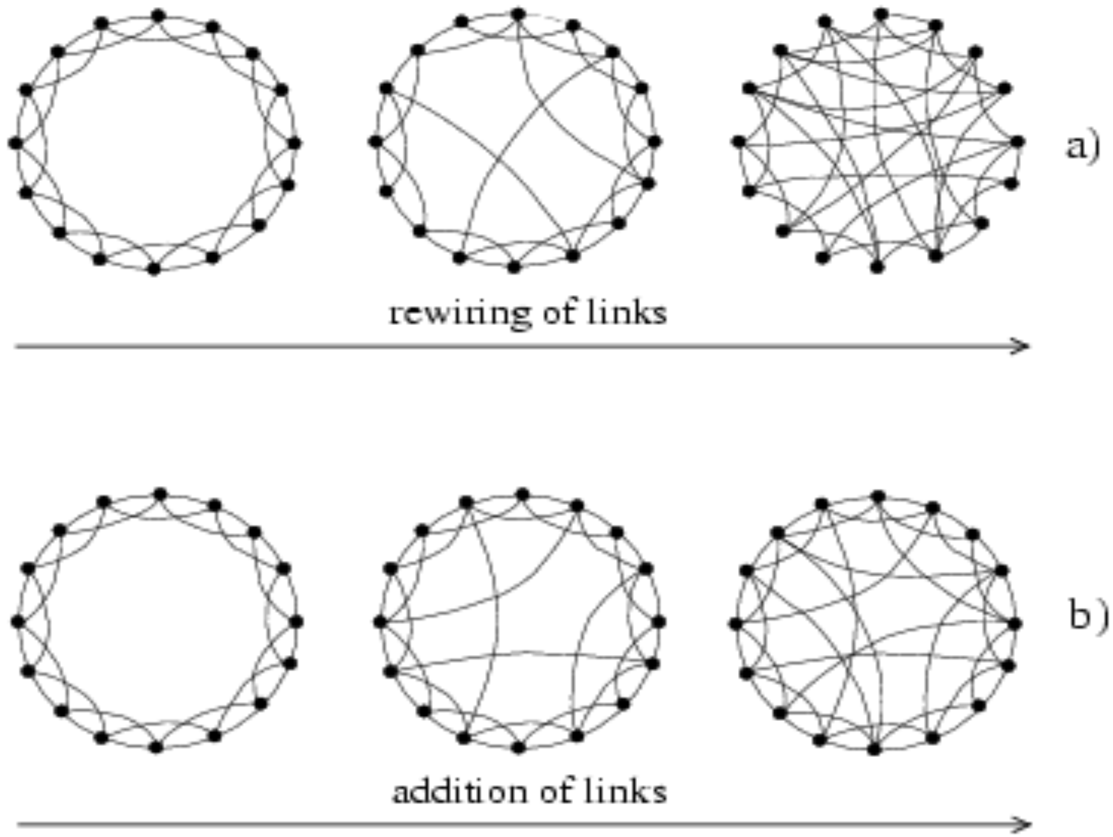
- **The crucial and interesting point is that small paths appear even for small values of p as we increase from p=0, while the high clustering remains until fairly large values of p**
  - Hence, there is a regime for values of p where both small paths as well as high clustering exists!

Small-world models

- **The above model is the original small-world model but it is rather involved to be analyzed**
- **We will use another model for our derivations**
  - Edges are added at random between two vertices in the circular lattice but no edges from the original circle are removed
  - The definition of p is remaining the same
    - ✓ For every edge at the original circle we create an additional shortcut with probability p between two randomly chosen vertices
- **When p→1 we no longer have a completely random graph**
  - This is not a big problem since we are interested in the regime where p is small
    - ✓ The only difference in this regime is that a small number of edges around the circle that would be absent in the original model are now present



Small-world models



Degree distribution

- **In the small world model that we examine every node has at least degree c**

- **The expected number of shortcut edges we add is (1/2)ncp**
  - ncp ends of shortcut edges

- **The number of shortcuts s attached to any vertex is Poisson distributed:**

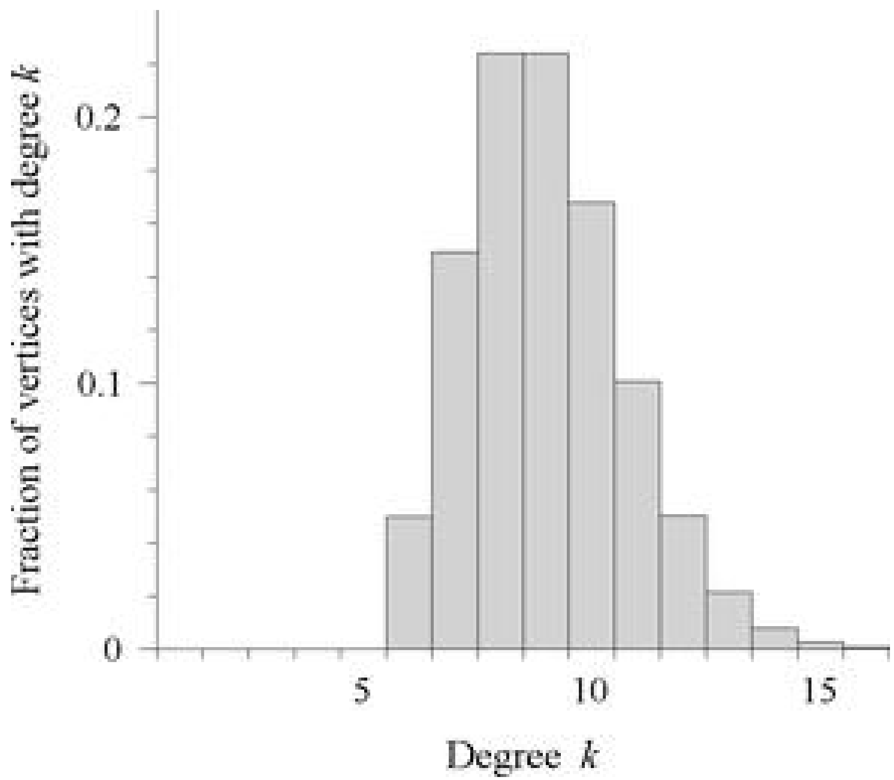
$$p_s = e^{-cp} \frac{(cp)^s}{s!}$$

- **The total vertex degree is k=c+s**

$$p_k = e^{-cp} \frac{(cp)^{k-c}}{(k-c)!}, \quad k \geq c$$

Degree distribution

- For c=6, p=0.5



Not similar to real world networks

Clustering coefficient

- **In order to calculate the clustering coefficient we need to calculate the number of triangles and connected triples after the addition of the shortcuts**
- **Number of triangles**
  - The triangles of the original circle are not changed: (1/4)nc(c-1)
  - New triangles can be created
    - ✓ In general nodes that have distance on the circle between (1/2)c+1 up to c are connected through 2-hop paths
      - ○ This number increases linear with the size of the network n
    - ✓ If a shortcut connects them then we have a new triangle
    - ✓ The probability they are connected through a shortcut is: $\frac{\frac{1}{2}ncp}{\frac{1}{2}n(n-1)} = \frac{cp}{n-1} \cong \frac{cp}{n}$
    - ✓ Hence, the number of triangles that are completed through the shortcuts is proportional to n*cp/n=cp
      - ○ At the limit of large n these triangles are negligible compared to these of the original circle

Clustering coefficient

- **Number of connected triples**
  - All connected triples of the original circle are still there: $(1/2)nc(c-1)$
  - Every shortcut creates new connected triples
    - ✓ At each end of the shortcut edge there are c edges that can form a triple
    - ✓ Hence, the total number of triples created due to a single shortcut are: $(1/2)ncp*2*c=nc^2p$
  - Pairs of shortcuts attached to a vertex can create connected triples as well
    - ✓ If a vertex has m attached shortcuts there are $(1/2)m(m-1)$ triples centered at this node
    - ✓ The number of shortcuts a node received is Poisson distributed with mean cp
      - Hence, the expected number of connected triples centered at a given vertex is $(1/2)c^2p^2$

Clustering coefficient

- **Combining all above together the clustering coefficient for the small-world network model we consider is:**

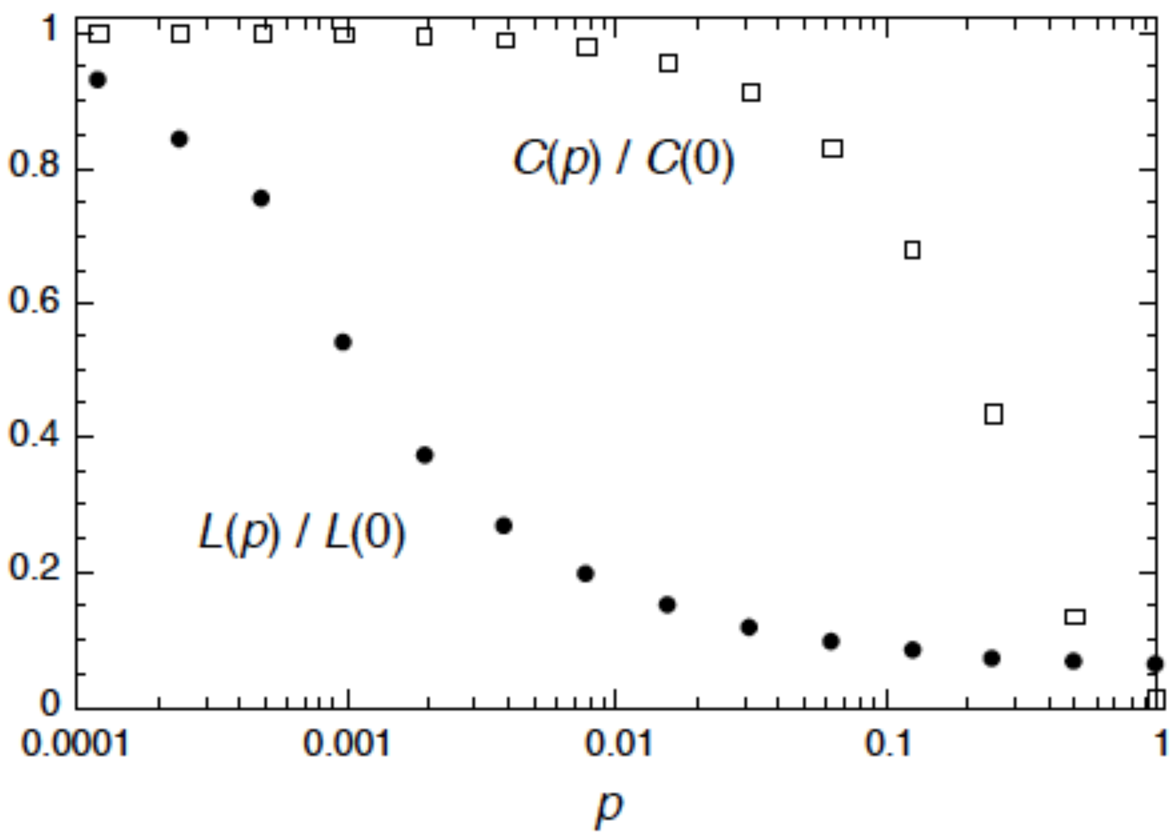
$$C = \frac{\frac{1}{4}nc(\frac{1}{2}c-1)\times 3}{\frac{1}{2}nc(c-1)+nc^2p+\frac{1}{2}nc^2p^2} = \frac{3(c-2)}{4(c-1)+8cp+4cp^2}$$

  - For p=0 we obtain the clustering coefficient of the circle model
  - As p grows the clustering coefficient reduces
    - ✓For p=1 the minimum value is $C_{\min} = \frac{3(c-2)}{4c-1}$
    - ✓This value is non zero
      - E.g., for c=6 → C_{min}=0.13

- **Note: the original small-world model from Watts and Strogatz exhibits C_{min}=0**

Clustering coefficient

- For c=6 and n=600



Average shortest path lengths

- **The analytical treatment of shortest paths in the small-world model is harder compared to degree distribution and clustering coefficient**

- **It can be argued that the average path length is given by:**

$$\ell = \frac{\ln(ncp)}{c^2 p}, \quad ncp >> 1$$

  - The average path length will increase only logarithmically with n for given c and p
    - ✓ Hence, even few shortcuts *per vertex* can produce short paths

Average shortest path lengths

- **For c=6 and n=600**