# Data and Networks
# Part II

Dr. Franck Kalala Mutombo

African Institute for Mathematical Sciences (AIMS) South Africa

*franckm@aims.ac.za*

May 19, 2017

# Small World Network

# Average pathlenght

A measure of the typical separation between two nodes in the graph is given by the average shortest path length, also known as characteristic path length, defined as the mean of geodesic lengths over all pairs of nodes:

$$L = \frac{1}{n(n-1)} \sum_{i,j \in V, i \neq j} d_{ij},$$

where $d_{ij}$ is the shortest distance between nodes $i$ and $j$

# Clustering Coefficient: Global and Local

- Clustering, also known as transitivity, is a typical property of acquaintance networks,
- where two individuals with a common friend are likely to know each other,
- In terms of network topology, transitivity means the presence of a high number of triangles,
- Two versions of this measure exist:
  - the global clustering coefficient and
  - the local clustering coefficient

The global version was designed to give an overall indication of the clustering in the network, whereas the local gives an indication of the embeddedness of single nodes.

# Global Clustering Coefficient

The global clustering coefficient of a network can be seen as the relative number of transitive triples (expression borrowed from the sociology literature), i.e. the fraction of connected triples of nodes (triads) which also form triangles:

$$C = \frac{3 \times \text{ number of triangles in the network}}{\text{number of connected triplets in the network}}$$

or

$$C = \frac{\text{number of closed triplets}}{\text{number of connected triplets of vertices}}.$$

# Global Clustering Coefficient

- A triplet is three nodes that are connected by either two (open triplet) or three (closed triplet) undirected ties.
- A triangle consists of three closed triplets, one centred on each of the nodes.
- The global clustering coefficient is the number of closed triplets (or $3\times$ number of triangles) over the total number of triplets (both open and closed).
- The factor 3 in the numerator compensates for the fact that each complete triangle of three nodes contributes three connected triplets, one centred on each of the three nodes, and ensures that $0 \leq C \leq 1$, with $C = 1$ for the complete graph $K_N$.

# Local Clustering Coefficient: Watts and Strogatz

- Probability that nearest neighbours of a node are themselves nearest neighbours, i.e.
- Concretely if node $i$ has $k_i$ nearest neighbours with $e_i$ connections i.e.
- The value obtained by counting the actual number of edges in $G_i$ (the subgraph of neighbours of $i$)
- the local clustering coefficient is defined as the ratio between $e_i$ and $k_i(k_i - 1)/2$

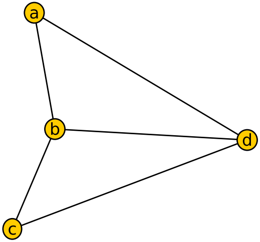
$$c_i = \frac{e_i}{k_i(k_i - 1)/2}.$$

$k_i(k_i - 1)/2$ is the maximum possible number of edges in $G_i$:

For this alternative, the clustering coefficient of a network is defined to be the average of $c_i$ over all the nodes in the network:

$$\overline{C} = \frac{1}{n} \sum_{i=1}^{n} c_i.$$

# Clustering Coefficients: Example




This network has two triangles and six connected triples (three for each triangle),

- Global clustering coefficient: $C_g = 3 \times 2/6 = 1$.
- Local clustering: The individual vertices have local clustering coefficients $c_a = c_c = 1$ and $c_b = c_d = 2/3$ and for the mean value $\overline{C} = 5/6$.

# Small world model: Stanley Milgram, 1963

- Despite their often large size, in most networks there is a relatively short path between any two nodes.
- The distance between two nodes is defined as the number of edges along the shortest path connecting them.
- The most popular manifestation of small worlds is the "six degrees of separation".
- there was a path of acquaintances with a typical length of about six between most pairs of people in the United States.

# Small world model

This feature (short path lengths) is also present in random graphs. In fact (Newman, 2003)

$$L \sim \frac{\log n}{\log \overline{k}}.$$

However, in a random graph, since the edges are distributed randomly, the clustering coefficient is considerably smaller.

$$\overline{C} = \frac{\overline{k}}{n-1}.$$

Instead, in most, if not all, real networks the clustering coefficient is typically much larger than it is in a comparable random network (i.e., the same number of nodes and edges as the real network).

Beyond Milgram's experiment, it was not until 1998 that Watts and Strogatz work stimulated the study of such phenomena.

### Watts-Strogatz model (1998)

Their main discovery was the distinctive combination of high clustering with short characteristic path length, which is typical in real-world networks (either social, biological or technological) that cannot be captured by traditional approximations such as those based on regular lattices or random graphs.

# Watts-Strogatz model

From a computational point of view, Watts and Strogatz proposed a one-parameter model that interpolates between an ordered finite dimensional lattice and a random graph. The algorithm behind the model is the following:

1. *Start with order*: Start with a ring lattice with $n$ nodes in which every node is connected to its first $k$ neighbours ($k/2$ on either side). In order to have a sparse but connected network at all times, consider $n \gg k \gg ln(n) \gg 1$.
2. *Randomise*: Randomly rewire each edge of the lattice with probability $p$ such that self connections and duplicate edges are excluded. This process introduces $pnk/2$ long-range edges which connect nodes that otherwise would be part of different neighbourhoods. By varying $p$ one can closely monitor the transition between order ($p = 0$) and randomness ($p = 1$).

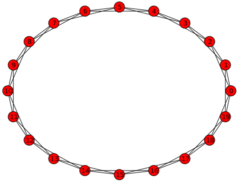
# Watts-Strogatz model: Rewiring process

- The rewiring can be considered as the process through which, with probability $p$ we replace each link $i, j$ with a link $i, k$, where $k$ is a randomly chosen node different from $i$ and $j$.
- In the case that $i, k$ is already contained in the modified network no action is considered.
- As $p \to 1$ the network tends to a completely random graph.
- The Watts-Strogatz network is often written as a three-parameters graph: $WS(n, k, p)$
- These networks have a high clustering coefficient in comparison with Erdös-Rényi random networks, i.e., if each node has a degree $k$, where $k$ is even,

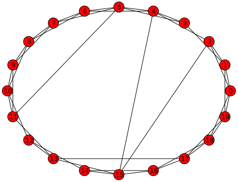
$$\overline{C} = \frac{3(k-2)}{4(k-1)},$$

which means that the clustering coefficient of these networks is independent of the network size and it tends to the value $\overline{C} = 0.75$ for large value of $k$.

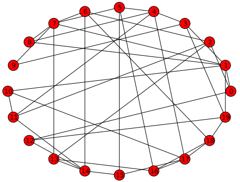
# Watts-Strogatz model



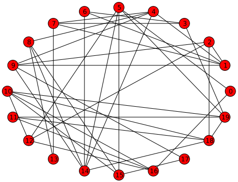
(a) $n = 20, k = 4, p = 0$



(b) $n = 20, k = 4, p = 0.110$



(c) $n = 20, k = 4, p = 0.50$



(d) $n = 20, k = 4, p = 0.85$

# Watts-Strogatz model

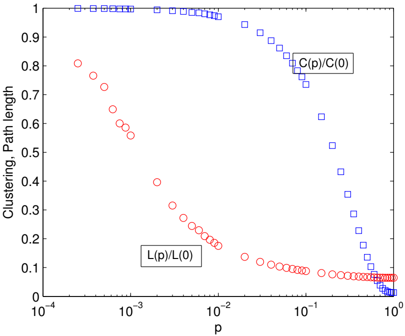



Figure : Structural evolution of the Watts-Strogatz model. Illustrations of the variation of the average path length and the clustering coefficient with the change of the rewiring probability for a network having $n = 1000$ nodes and $k = 10$. Each point is the average of 50 realisations. The values for the average path length and clustering coefficient are normalised by dividing them by the respective values obtained for $WS(1000, 10, 0)$.

# Watts-Strogatz model

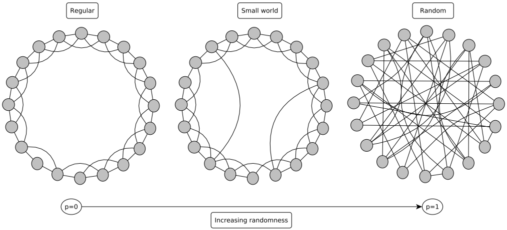



Figure : Three basic network types in the model of Watts and Strogatz. The leftmost network is a ring of 20 nodes ($n = 20$), where each vertex is connected to its four neighbours ($k = 4$). This is an ordered network which has a high clustering coefficient $C$ and a long pathlength $L$. By choosing an edge at random, and reconnecting it to a randomly chosen vertex, networks with increasingly random structure can be generated for increasing rewiring probability $p$. In the case of $p = 1$, the network becomes completely random, and has a low clustering coefficient and a short pathlength. For small values of $p$ so-called small-world networks arise, which combine the high clustering coefficient of ordered networks with the short pathlength of random networks.

# Watts-Strogatz model: Average path lenght

Although the mean degree is exactly $\overline{k} = k$, no exact expression for the degree distribution for a Watts-Strogatz small-world network is known, except when $p = 0$, in which case every node has degree $k$. An approximation for the case of $0 < p < 1$ has been calculated by Barrat and Weigt (2000). There is also no known exact expression for the average path length of this network. A scaling approximation is given by (Newman 2003):

$$L \sim \frac{n}{K} \frac{1}{\sqrt{u^2 + 4u}} \tanh^{-1}\left(\frac{u}{\sqrt{u^2 + 4u}}\right),$$

where $u = pkn$. For the case of fixed $p$ and $k$, if $k \ll n$ and $n$ is sufficiently large, then the average path length is expected to increase with network size, because the terms involving $u$ converge to a constant for large $n$ (Newman 2003).

# Watts-Strogatz model: Clustering

Although the clustering coefficient does also not have a known exact expression, it is well approximated Barrat and Weigt (2000) for large $n$ as

$$\overline{C} \sim 0.75 \left( \frac{k-2}{k-1}(1-p)^3 + O(\frac{1}{n}) \right).$$

Since $n$ is assumed to be large, for $p < 1$ the term expressed as $O(1/n)$ can be ignored.

# References

Watts D. J. 2000 (2000)
Small world: The Dynamic of Networks between Order and Randomness
*The American Mathematical Monthly*, 107, 664–668.

Watts, D. J. and Strogatz, S. H.(1998)
Collective dynamic of small-world networks
*Nature*,393,440–442.

Barabási, A.-L. and Albert, R. (1999)
Statistical mechanics of complex networks
*Science*, 286, 509–512.

GPrettejohn, B. J. and Berryman, M. J. and McDonnell, M. D. (2011)
Methods for generating complex networks with selected structural properties for simulations: A review and tutorial for neuroscientists
*Front. Comput. Neurosci.*,5

Fronczak, A. and Fronczak, P. and Holyst, J. A. (2004)
Average path length in random networks.
*Phys. Rev. E.*,70, 056110.

The End