1 The connectivity issue

1. The question to be asked: how can connectivity be studied?

We do know the importance of trade, money, clusters of districts, cities. We know the crucial role of information and thus its flows across communities, firms etc. In particular, firms' clustering is crucial for development. More generally, interaction is an important key to explain it

2. The theory of networks does shed some important light on these issues. The following reports some notions that can be useful to apply it to questions dealing with development.

It is expedient to start with some definitions

3. Cluster: a web of edges (links) connecting a node's neighbours to each other.

4. A graph, define it as composed by a set of nodes, P, and edges, E

$$G = G(P, E)$$

5. Let us begin with notions dealing with random graphs and use them as a benchmark.

A random graph G is a graph of $|P| = N$ nodes connected by $|E| = n$ edges chosen randomly from all possible edges. The order of magnitude of the latter is

$$n_{\mathsf{max}} = C_{N.2} = \frac{N(N-1)}{2}$$

6. Another, interesting order of magnitude: since one can randomly generate a graph by connecting nodes by drawing n edges from a pool of $C_{N.2}$, there are as many as $C_{N.2}$ edges to generate such a graph. This means that one can generate as many as

$$C_{C_{N.2},n} = \frac{\frac{N(N-1)}{2}!}{n!\left[\frac{N(N-1)}{2} - n\right]!}$$

(for instance, if $N = 6$ and $n = 3$; $C_{6.2} = 15$; $C_{15,3} = 455$ graphs can be generated).

7. The interesting question: how does a random graph evolve? More specifically, why is it that certain cliques or clusters come to be when randomness prevails?

8. Let us introduce the following idea: nodes are initially entirely unconnected but then proceed to connect them with some probability p. Thus p is the probability that any two nodes be connected.

9. The expected number of connected nodes after some experiments have been made to connect them. In other words what is the expected number of edges?

$$E(\#) = \frac{N(N-1)}{2}p$$

10. What is the probability of obtaining a specific graph G_0 with n edges?

$$P(G_0) \; = \; p^n(1-p)^{\frac{N(N-1)}{2}-n}$$

(for instance a specific graph from $N = 6$; $n = 3$ and $p = .2$; say $G_0 = (adf)$, $P(adf) \simeq .00055$)

11. The average degree of random graphs: how many connections, on average, is a node likely to possess in a random graph? A node can connect to as many as $N - 1$ other nodes with probability p. The answer is $\langle k \rangle = p(N-1)$.

In an evolutionary context, if n edges have been successfully established, then we can compute the probability of connection of any two nodes: since there are $\frac{N(N-1)}{2}$ possibilities of connecting and n have been the actual ones (the favourable cases)

$$p = p(N, n) = \frac{n}{\frac{N(N-1)}{2}} = \frac{2n}{N(N-1)} \approx \frac{2n}{N^2} \quad \text{(for large } N\text{)}$$

12. What's most interesting is the emergence of some properties: the shape of a graph connection.

- Given a random graph, what is the expected number of subgraphs made up of, say, k nodes? This is an important question. In a graph of N nodes there are $C_{N,k}$ ways to generate graphs of k nodes, i.e. there are

$$C_{N,k} = \frac{N!}{k!(N-k)!}$$

possible graphs. But, from the point of view of the exact shape that a graph acquires, especially if we are considering an evolving process, each subgraph can potentially give rise to $k!$ other graphs (e.g. take a graph of $N = 6$ (a, b, c, d, e, f) nodes and consider a subgraph of $k = 3$ nodes (adf) this subgraph can also come in the shape of $(afd\ , daf, dfa, fad, fda)$. In actual and practical problems some allowance must be made for the fact that some of these subgraphs have the same relevance and thus if there are a such ones, the actual number that each graph can really generate is $\frac{k!}{a}$.

- the next question is what is the expected number of *connected* subgraphs if the available edges to connect the k nodes is l and the connection probability is p?

$$E(X) = C_{N,k}\frac{k!}{a}p^l = \frac{N!}{k!(N-k)!}\frac{k!}{a}p^l = \frac{p^l}{a}\frac{N!}{(N-k)!} \approx \frac{p^l}{a}N^k$$

(for large N and relatively small k. An example, $N = 100$; $k = 6$; $l = 3$; $a = 1; p = .2 \rightarrow E(X) = 10,000.000 * .008 = 80,000$)

13. There exists an important property of random graphs. The probability p depends on the edges that have been successfully set ($p \approx \frac{2n}{N^2}$) and also on N. It has been found that there exists a critical probability $p_c = p_c(N)$ below which (for $p = p(N) < p_c$) almost no 'property' or subgraph connections appear whilst above it (for $p = p(N) \geq p_c$) most such subgraphs connections do!

In other words, a phase transition occur. A sudden appearance of the expected number of subgraphs with l edges.

14. To see what is such a critical probability, consider $E(X) = \frac{p^l}{a} N^k$. If

$$p = p_c(N) = cN^{-\frac{k}{l}}$$

then

$$E(X) = \frac{c^l}{a}$$

15. Some cases: the critical probability at which almost every graph contains a subgraph with k

nodes and l edges

- a tree of order k $(l = k - 1)$: $p_c(N) = cN^{-\frac{k}{k-1}}$

- a cycle or order k $(l = k)$; $p_c(N) = cN^{-1}$

- a complete subgraph of order k $(l = \frac{k(k-1)}{2})$: $p_c(N) = cN^{-\frac{2}{k-1}}$

(an example, the critical probability for a graph to contain a completely connected subgraph of $k = 10$, i.e. with 45 connections and having $N = 1000$ nodes, and $c = 2$, is $p_c = 43\%$, if $N = 10000$, $p_c = 26\%$, if $N = 100000$, $p_c = 15,5\%$, thus in graphs of many nodes, cliques start appearing even for low probabilities of setting up a connection).

17. The degree distribution of random graphs. Question: what is the probability that a node named k_i has k degrees (connected to k other nodes)?

$$P(k_i = k) = P(k) = C_{N-1,k}p^k(1-p)^{(N-1)-k}$$

and the expected number of so connected nodes is

$$E(X_k) = NP(k)$$

Note that, since $\langle k \rangle \simeq pN$, and for $N \to \infty$;

$$P(k) = e^{-\langle k \rangle}\frac{\langle k \rangle^k}{k!}$$

namely a Poisson distribution.

18. Some important quantities (magnitudes) in random graphs.

- average path length, l: average distance between any pair of nodes. Assume that the average degree is $\langle k \rangle$. If on average the path length is l, then multiplying $\langle k \rangle$ a number l of times, one counts (approximately) all the nodes in the network N. Hence $\langle k \rangle^l = N$ from which

$$l_{\text{rand}} = \frac{\log N}{\log \langle k \rangle}$$

Thus, the average path length scales with the log of the network size.

- clustering coefficient of a random graph$=p$.This is so since definitionally $C_i = \frac{2E_i}{k_i(k_i-1)}$ but given that the expected number of connections is $p\frac{k_i(k_i-1)}{2}$, then $C_i = p$. Since $p = \frac{2n}{N^2}$ and $\langle k \rangle = \frac{2n}{N}$, then

$$C_i = \frac{\langle k \rangle}{N}$$

19. **The Watts and Strogatz conundrum**

-. Consider a very regular and ordered network that is highly clustered with a high path length.

Such a network can easily be represented by a ring lattice in which each node is immediately connected right and left with a given number of other nodes. Assume that this number be $K \equiv \langle k \rangle$.

-. The relevant magnitudes for a network $N \ggg K \ggg$ $\log N \ggg 1$ are :

$C(ordered) = \frac{2E}{K(K-1)} = \frac{3}{4}\frac{K-2}{K-1} \;\rightarrow \frac{3}{4}$ in the case of large networks.

$l(ordered) = \frac{N}{2K}$

Thus, highly clustered network (C close to 1) and very high average path length.

-. Consider now a random network:

$$C(\mathsf{rand}\,om) \simeq \frac{K}{N}$$

$$l(\mathsf{rand}\,om) \simeq \frac{\log N}{\log K}$$

Thus, a scarcely clustered network with a fairly short average path length.

.- Question is a short average path length always associated with a little clustered network? A long path length with a highly clustered one? The answer is NO.

.- Consider the following algorithm (procedure)

a) Begin by setting up an ordered network on a ring lattice with the mentioned magnitudes.

b) Proceed by randomly rewiring with probability p , barring duplications and self wiring. This means taking an edge at random away from one node and reconnecting it to another node at random. Since the total number of rewirable edges is N nodes times $K/2$ neighbours on either side of the ring, this procedure allows for a long range rewiring of $p\frac{NK}{2}$ edges. How does the network evolve?

c) Clearly, the network evolves according to p, if $p = 0$ the network remains as it is, i.e. ordered, if $p = 1$ the network becomes totally random. Hence, $C(p)$ and $l(p)$ are expected to vary as a function of p, from $C(ordered) \equiv C(0)$; $l(ordered) \equiv l(0)$ to $C(\text{rand}\,om) \equiv C(1)$; $l(\text{rand}\,om) \equiv l(1)$.

d) Note that $l(p)$ drops very rapidly with small increases in p while $C(p)$ varies little with p. It follows that there is a large interval in which l is short and C is high.

e) To see why this happens consider that for small p, the path length scales with the system size whilst the clustering coefficient remains roughly constant ($\simeq 3/4$). As the network becomes more and more random the path length begins to scale logarithmically (small changes) while C begins to approach the small value of K/N.

21. It is intuitively clear that the observed 'phenomenon' on the average path length, l, depends on the system size which can here be defined by KN on which the probability of rewiring operates: pKN

20. The actual mathematical form has largely been left to numerical simulations. In any case, approximations indicate that:

$$l(p, N, K) \sim \frac{N^{\frac{1}{d}}}{K} f(pKN)$$

where $f(pKN) = const$ if $pKN \lll 1$ and $f(pKN) = \ln(pKN)$ if $pKN \ggg 1$, d is the lattice dimension (the ring: $d = 1$).

22. The equivalent expression for C is more elaborate and it goes:

$$C(p) = \frac{3K(K-1)}{2K(2K-1) + 8pK^2 + 4p^2K^2}$$

23. a final word on the degree distribution: it is very similar to the random graph distribution with a peak $\langle k \rangle = K$.

24. In any case, note that even for a small p and a reasonably sized network the phenomenon emerges. Is the world a SMALL WORLD? Possibly, yes!

25. **The Barabasi's caper**

- The problem with the Watts and Strogatz' model is that it applies to networks that as noted above have approximately a Poisson distribution, that are, in a sense, 'quasi random so that the most likely number of connections for a node is just the average $\langle k \rangle$ meaning that from the point of view of connectedness they are about the same.

- Most relevant networks do not have this structure. Empirical findings have shown that, quite frequently, the distribution of nodes takes the form:

$$P(k) \ = \ ak^{-\gamma}$$

(in log-log form $p(k) = a - \gamma k$).

- This is an important finding suggesting that the distribution is not random.

- Since, $P(k) \rightarrow 0$ only for $k \rightarrow \infty$, it is a distribution that exhibits values significantly different from zero even for very large k's. In other words, in such networks there are likely to be few nodes with high k's, some with a sizable k, very many with a small k, i.e. all scales of k are likely to be present. The average $\langle k \rangle$ is not at all representative of the network scale and the ratio of the mean to the variance tends zero for $k \rightarrow \infty$. This type of networks are *scale-free* networks.

26. Evidence shows that this is the case for most transportation networks, power-line grids, city size distribution but also firm size distribution. This has great implications for a considerable number of problems. ....

The question arises: why is it that many networks have such a structure? A likely answer is because their evolution has been such that although they are the result of a stochastic process they do not feature randomness but order. Hence, the analytical task is to find an evolutionary procedure that leads to this result. The following is the devise conjectured by Albert and Barabasi.

These authors have exploited two major and historically well established ideas:

a) networks grow, i.e. their size N increases with time;

b) attachment of newly born nodes to existing ones is preferential, i.e. they attach to nodes that have already many attachments.

Proceed as follows.

a.1) start with a small number m_0 of nodes and at every time step add a new node with m edges.

b.1) the probability that the new node attaches to node i depends on k_i, hence

$\Pi(k_i) = \frac{k_i}{\sum_j k_j}$

c) the problem is now to derive the nodes' distribution. They follow a continuum theory.

27. Note that:

$$\frac{\partial k_i}{\partial t} = m\Pi(k_i) = m\frac{k_i}{\sum_j k_j}$$

knowing that by the above assumptions $\sum_j k_j = 2mt$.Thus, $\frac{\partial k_i}{\partial t} = m\frac{k_i}{2mt} = \frac{k_i}{2t}$.

Solving this differential equation by integration and by assuming that the initial condition for every node is m at some t_i, namely $k(t_i) = m$:

$$k_i(t) = m(\frac{t}{t_i})^{\frac{1}{2}}$$

all nodes basically evolve in the same way. Now, ask the question: what is the probability that $k_i(t) < k$, $P(k_i(t) < k)$?

The authors follow a very ingenious procedure. Given the above:

$$P(k_i(t) < k) = P(t_i > \frac{m^2 t}{k^2})$$

turning the question around by asking: what is the probability that node i appears with its m connections at time

$$t_i > \bar{t}_i = \frac{m^2 t}{k^2}$$

The reason to ask this question lies with the fact that it is very reasonable that the probability of appearing so connected at time t_i be the same for all $i's$ and inversely related to the number of time steps, for instance:

$$P(t_i) = \frac{1}{m_0 + t}$$

28. If this is the case,

$$P(t_i > \frac{m^2 t}{k^2}) = 1 - \frac{1}{m_0 + t}\frac{m^2 t}{k^2})$$

and such that by differentiating $\frac{\partial P(k_i(t)<k)}{\partial k} = \frac{\partial P(t_i > \frac{m^2 t}{k^2})}{\partial k}$:

$$P(k) \;\; = \;\; \frac{2m^2 t}{m_0 + t}\frac{1}{k^3}$$

asymptotically, for $t \longrightarrow \infty$

$$P(k) = 2m^2 k^{-3}$$

a power-law, scale free distribution with $\gamma = 3$. Setting $2m^2 = \alpha$

$$P(k) = \alpha k^{-\gamma}$$