The Abdus Salam
International Centre for Theoretical Physics

United Nations
Educational, Scientific
and Cultural Organization

International Atomic
Energy Agency

SMR.1656 - 35

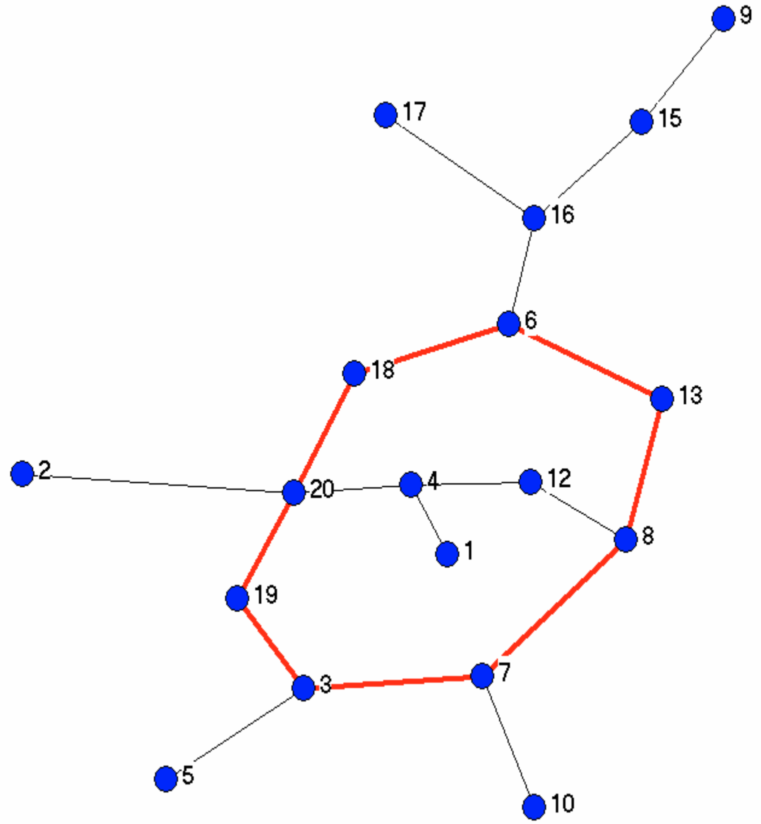
School and Workshop on
Structure and Function of Complex Networks

16 - 28 May 2005

Loops of any Size and Hamilton Cycles in Random Scale-Free Networks

Ginestra BIANCONI
the Abdus Salam ICTP
Strada Costiera 11
Trieste 34014
Italy

Loops of any size and Hamilton cycles in random scale-free networks



School and Workshop Structure and Function of Complex networks

May 2005

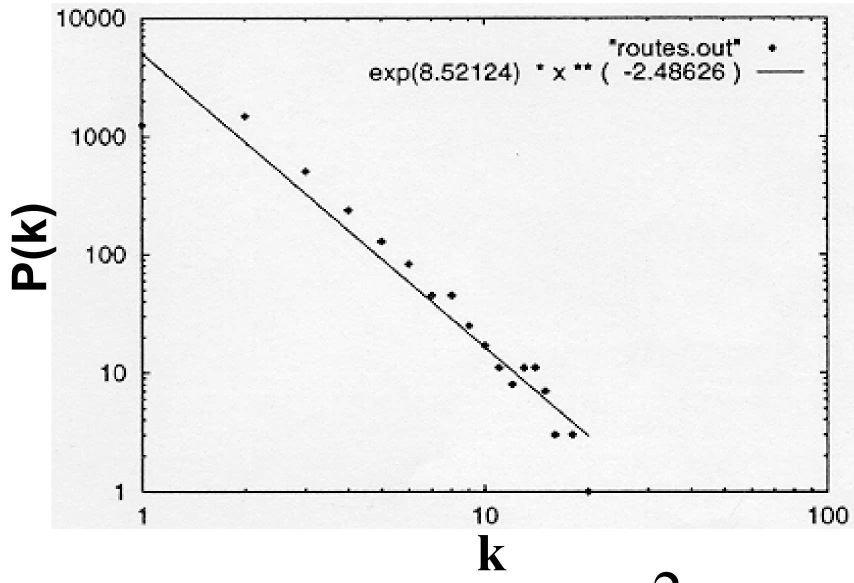
Ginestra Bianconi and Matteo Marsili

Abdus Salam Center for Theoretical Physics ICTP, Trieste Italy

Complex networks: Their degree distribution and clustering coefficient

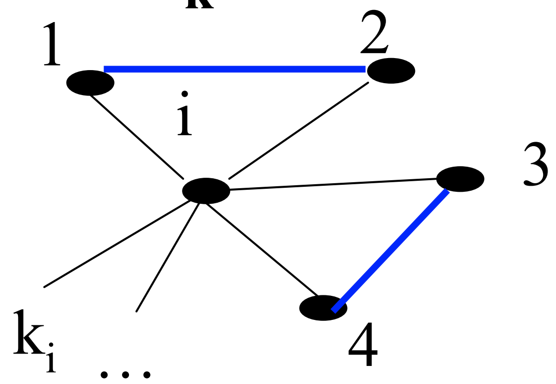
Scale-free degree distribution

$$P(k) \approx k^{-\gamma}$$



Large clustering coefficient

$$C_i = \frac{number\ \ of\ \ links\ \ between\ \ 1, 2, ..., k_i}{k_i(k_i - 1)/2}$$

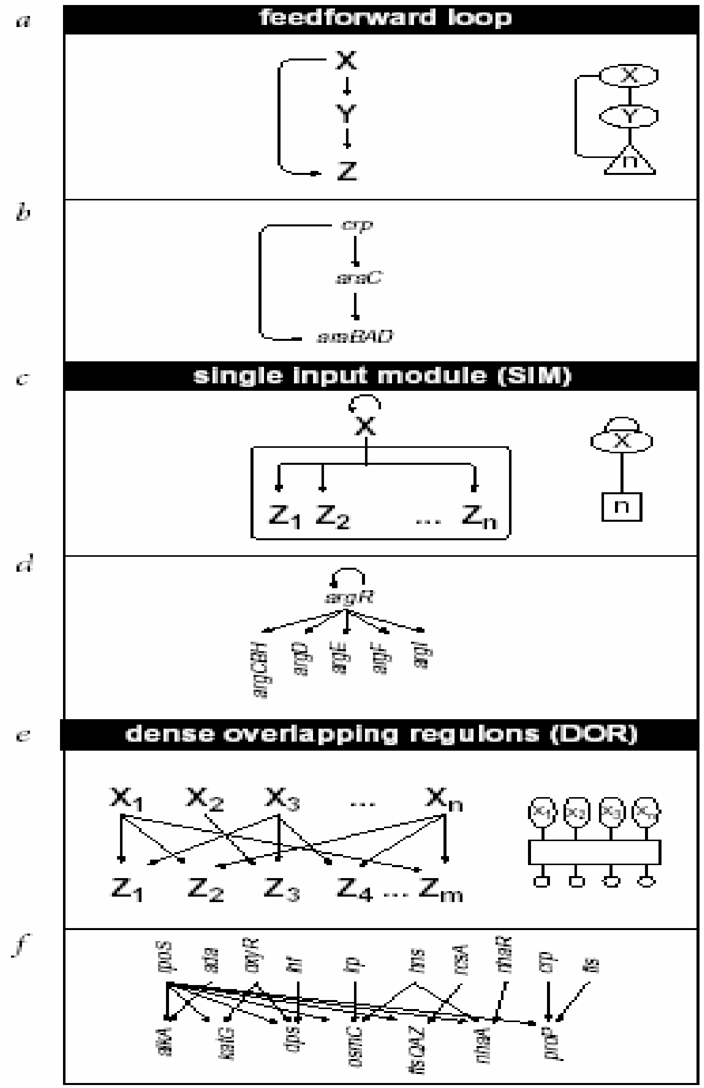


$$C_{SF} >> C_{ER}$$

Motifs → Function

Subgraphs which appear with higher frequency than in randomized networks are called **motifs** of the network.

The motifs are relevant to understand the function of the network.



Modules in the transcriptome network of e.coli.

(S.S. Shen-Orr, et al., Nature Genetics 31,64 (2002)).

To which extent

large-scale properties

effect

subgraphs frequency?

Loops

Loops are a special kind of network subgraphs

- They are responsible for the multiplicity of paths going through generic nodes of the network
- They are relevant for load distribution
- They are neglected in local tree-like approximations.

Direct counting of loops of finite size *L*

If we have access to the entire adjacency matrix $((a_{i,j}))$ with $a_{i,i} = 0$ the number of loops of finite size L is given by

$$\aleph_L = \frac{1}{2L}\left(\mathrm{Tr}\ a^L - \mathrm{corrections}\right)$$

Direct counting of loops of finite size *L*

If we have access to the entire adjacency matrix $((a_{i,j}))$ with $a_{i,i} = 0$ the number of loops of finite size L is given by

$$\aleph_3 = \frac{1}{6}\left[\sum_i (a^3)_{i,i}\right]$$

$$\aleph_4 = \frac{1}{8}\left[\sum_i (a^4)_{i,i} - 2\sum_i (a^2)_{i,i}(a^2)_{i,i} + \sum_i (a^2)_{i,i}\right]$$

$$\aleph_5 = \frac{1}{10}\left[\sum_i (a^5)_{i,i} - 5\sum_i (a^3)_{i,i}(a^2)_{i,i} + \sum_i (a^3)_{i,i}\right]$$

……

Average number of loops in regular random networks

$$P(k) = \delta(k - c)$$

The number of **small loops** with $L < \log(N)$ is finite in infinite graphs

$$< \aleph_L > \propto \frac{1}{2L}(c-1)^L$$

The number of **large loops** is exponentially large, where

$$\ell = L / N$$

$$< \aleph_L > \propto \exp(N\sigma(\ell))$$

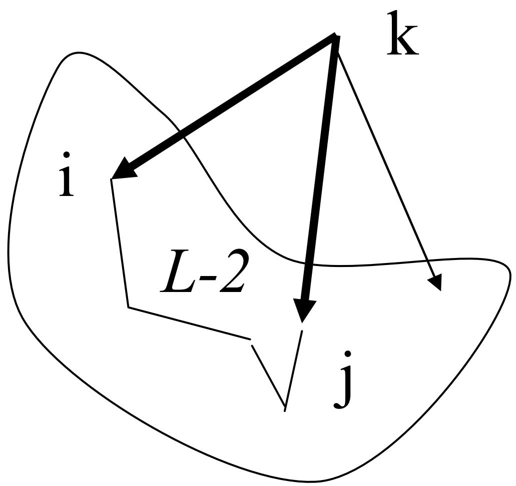
The **Hamiltonian cycles** ***(L=N)*** are present only for

$$c \geq 3$$

Number of *L*-loops in the BA network

While the BA network grows new loops are formed. These loops include necessarily the last node of the network.

For **small loops** it is possible to show



$$< \aleph_L(N) > \cong \frac{1}{2L}\left(\frac{m}{2}\log(N)\right)^L$$

Indication: scale-free graphs might have a large number of small loops

G. Bianconi and A. Capocci *Phys. Rev.Lett*. **90**, 078701 (2003).

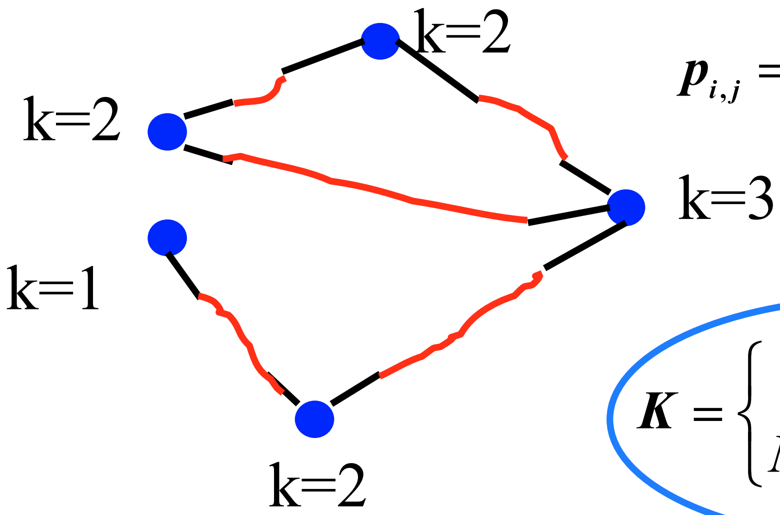
Motivation for counting loops on random scale-free networks

We would like to understand which are the consequences of pure **scale-free distribution** on **subgraphs frequency**

1. for having a reliable **null model** to compare real networks with;
2. **for having a reference point for counting loops in correlated networks.**

Ensembles of random scale-free networks

Molloy-Reed ensemble



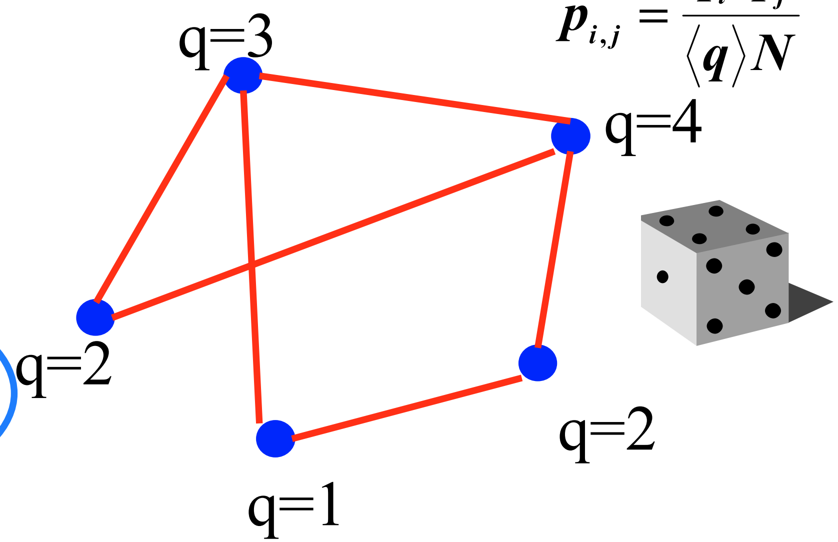
$$p_{i,j} = \frac{k_i\, k_j}{\langle k \rangle N}$$

$$P(k) \approx k^{-\gamma} \qquad k \in [m, K]$$

M. Molloy and B. Reed,(1995).

Hidden variable ensemble

$$p_{i,j} = \frac{q_i\, q_j}{\langle q \rangle N}$$



$$K = \begin{cases} N^{1/2} \text{ for } \gamma < 3 \\ N^{1/(\gamma-1)} \text{for } \gamma > 3 \end{cases}$$

$$P(q) \approx q^{-\gamma} \qquad q \in [m, K]$$

G. Caldarelliet al., *PRL* (2002)

The two ensembles: pro and contra for counting loops

- **The Molloy Reed ensemble**

  For scale-free degree distributions there are multiple link in the network.

- **The fitness ensemble**

  There is no control on the smallest value of the connectivity.

  In particular there can be some network realization with nodes of connectivity $k_i=0,1$. This is an important aspect to take into account when counting very large loops like for example Hamiltonian cycles.

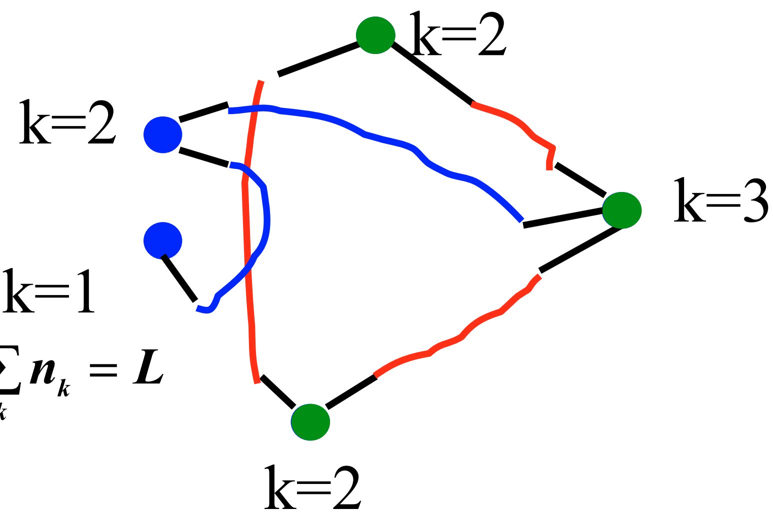
Average number of loops in the MR ensemble

To count the average number of loops of size ***L*** one have

1. to calculate in how many ways one can choose ***L*** nodes, n_k nodes with connectivity ***k*** for every allowed value of the connectivity ***k*** with

$$\sum_k n_k = L$$

2. in how many ways one can order the nodes and choose the links
3. how many are the networks in the ensemble which contain the chosen loop.



$$\langle \aleph_L \rangle = \frac{L!}{2L} \sum_{\{n_k\}}{}' \prod_k \binom{NP(k)}{n_k} \left[k(k-1)\right]^{n_k} \frac{\left(\langle k \rangle N - 2L - 1\right)!!}{\left(\langle k \rangle N - 1\right)!!}$$

Formula manipulations

Starting from

$$\langle \aleph_L \rangle = \frac{L!}{2L} \sum_{\{n_k\}}{}' \prod_k \binom{NP(k)}{n_k} \left[k(k-1)\right]^{n_k} \frac{\left(\langle k \rangle N - 2L - 1\right)!!}{\left(\langle k \rangle N - 1\right)!!}$$

1. Applying the Stirling approximations for factorials valid in the ***N>>L>>1*** limit,
2. Using the integral representation of the delta,
3. Performing the sum over n_k one gets the expression

$$\langle \aleph_L \rangle = \frac{1}{2L} \int_{-\pi}^{\pi} dx \; e^{N \left\langle \log\left[1 + k(k-1) e^{-ix} / (N \langle k \rangle)\right] \right\rangle + L\, ix + N g(L/N)}$$

Which can be evaluated with a **saddle point approximation**

Small loops

$$\langle \aleph_L \rangle \approx \frac{1}{2L}\left(\frac{\langle k(k-1)\rangle}{\langle k \rangle}\right)^L$$

$$L << \begin{cases} N^{(3-\gamma)/2} & \text{for } 2 < \gamma < 3 \\ N^{(\gamma-3)/(\gamma-1)} & \text{for } \gamma > 3 \end{cases}$$

Small loops

$$\langle \aleph_L \rangle \propto \frac{1}{2L} N^{(3-\gamma)L/2}$$

$$L << \begin{cases} N^{(3-\gamma)/2} & \textbf{for } 2 < \gamma < 3 \\ N^{(\gamma-3)/(\gamma-1)} & \textbf{for } \quad \gamma > 3 \end{cases}$$

Scale-free networks have a large number of small loops for $\gamma<3$.

Small loops

$$\langle \aleph_L \rangle \propto \frac{1}{2L} c^L$$

$$L << \begin{cases} N^{(3-\gamma)/2} & \text{for } 2 < \gamma < 3 \\ N^{(\gamma-3)/(\gamma-1)} & \text{for } \quad \gamma > 3 \end{cases}$$

For $\gamma > 3$ loops begin to be relevant for sizes L of the order of $\log(N)$.

Larger loops

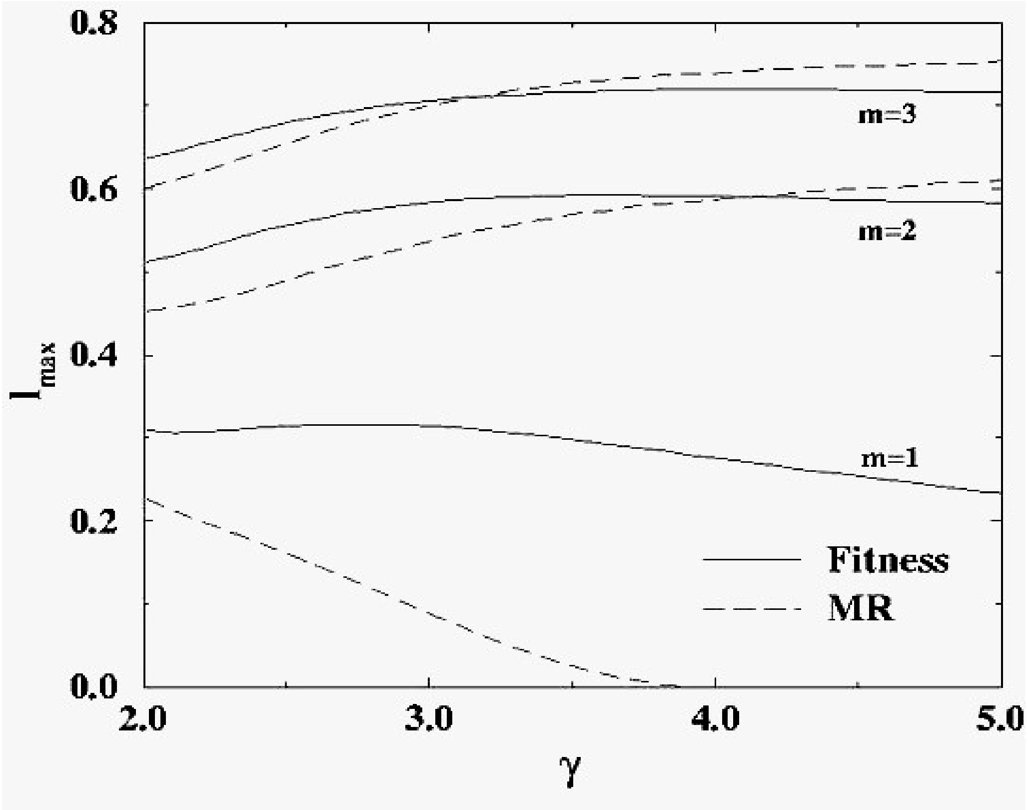
Scale-free networks have an exponential number of **large loops.**

$$< \aleph_L > \propto \exp(Nf(\ell^*))$$

$$\langle \aleph_L \rangle = \int_{-\pi}^{\pi} \frac{dx}{2\sqrt{\pi L}} e^{Nf(N\langle k \rangle e^{ix}, L/N)}$$

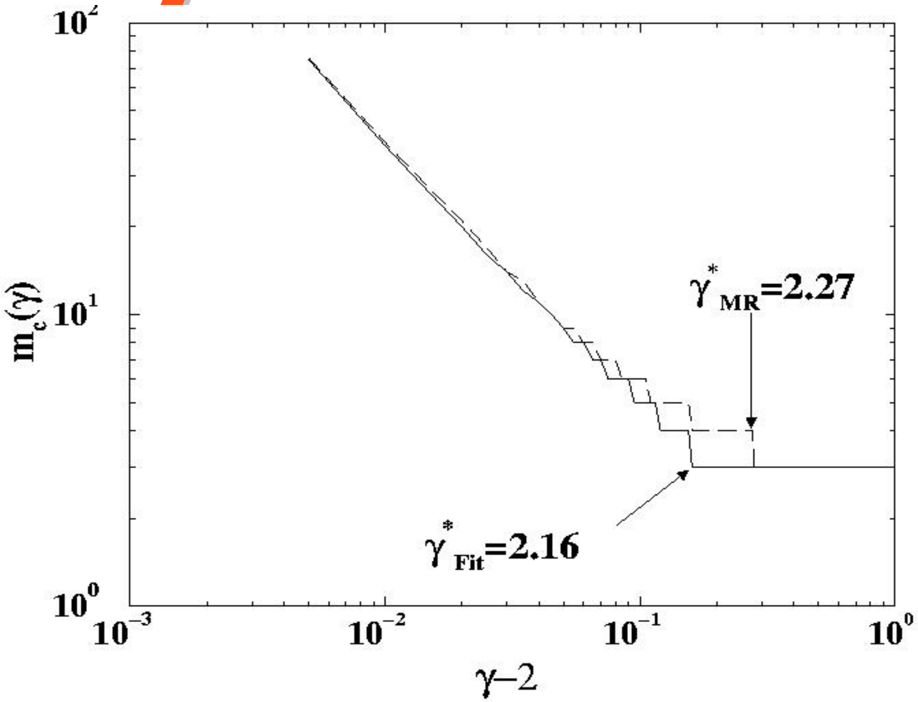
The **most frequent loop** size is of order N, and given by $L = \ell_M N$

$$\left\langle \frac{k(k-1)}{\langle k \rangle - 2\ell_M + k(k-1)\ell_M} \right\rangle = 1$$



Hamilton cycle

The random scale-free network is not Hamiltonian **also for a minimal connectivity** $m \geq 3$ as long as $\gamma < \gamma*$ in the $N \to \infty$



It can be shown that the critical *m* for having an expected number of Hamiltonian cycles in the ensemble grater than zero goes like

$$m_c(\gamma) \approx (\gamma - 2)^{-1}$$

For $\gamma < \gamma*$ it is not possible to extract a regular random graph with *c=3* ≤*m* form the SF network

Loops passing through a node of the network

The number of loops of small size L passing through a node of degree *k* is given by

$$\langle \aleph_L(k) \rangle = \frac{k(k-1)}{\langle k \rangle N} \left(\frac{\langle k(k-1) \rangle}{\langle k \rangle} \right)^{L-1}$$

The clustering coefficient of a scale-free network with $\gamma<3$ decreases with the network size as

$$C_i \approx N^{2-\gamma}$$

while for $\gamma>3$

$$C_i \approx N^{-1}$$

Loops passing through a node of the network

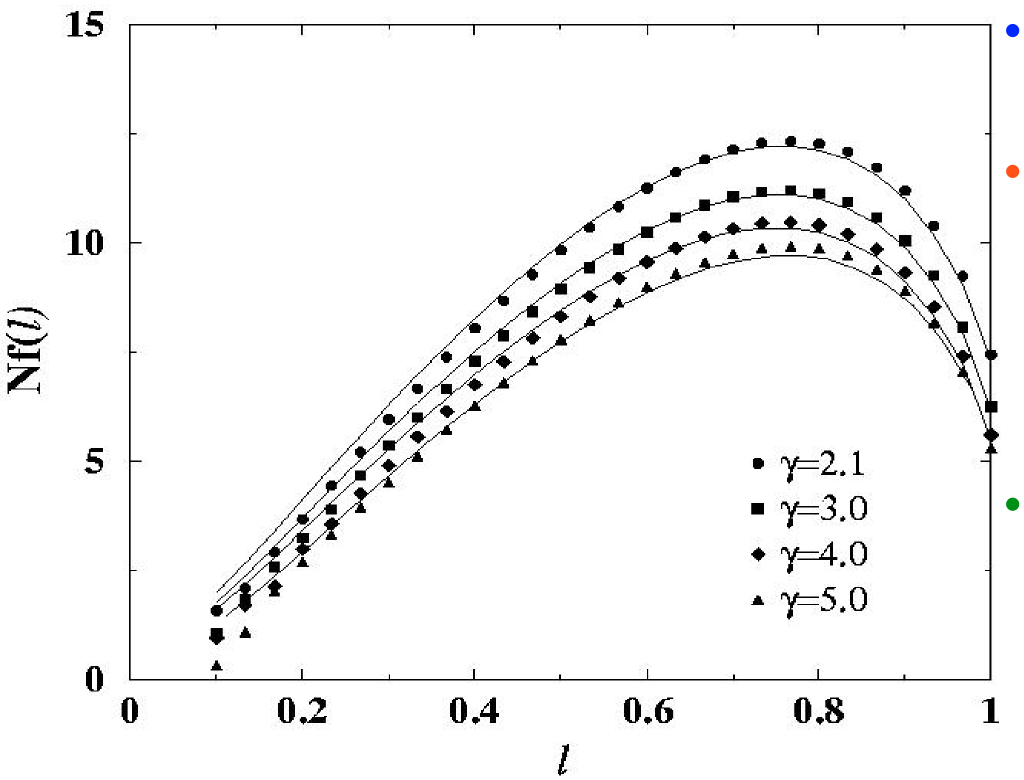
The number of loops of small size L passing through a node of degree *k* is given by

$$\langle \aleph_L(k) \rangle = \frac{k(k-1)}{\langle k \rangle \; N} \left(\frac{\langle k(k-1) \rangle}{\langle k \rangle} \right)^{L-1}$$

For a finite network with γ<3 loops of size *L* become relevant if one looks at nodes with connectivity

$$k > N^{\frac{1}{2} - \frac{3-\gamma}{4}(L-1)}$$

Comparison with direct counting results



- We simulate a Molloy Reed graph **without multiple links**.
- We use the **Johnson algorithm** to count directly each loop of a network of size N=30 and m=3 averaged over different realization.
- We use the **same degree distribution** to compare the direct counting results with the analytic results.

D. B. Johnson, SIAM J. Comput., **4** 77 (1975).

Loops in the hidden variable ensemble

The expression for the average number of loops change as in the following

$$\langle \aleph_L \rangle = \frac{L!}{2L} \sum_{\{n_q\}}{}' \prod_q \binom{NP(q)}{n_q} \frac{q^{2n_q}}{\left(\langle k \rangle N\right)^L}$$

The scaling results remain the same as in the Molloy Reed ensemble but the equation for the most frequent loop change and the value of γ^* also changes.

Conclusions

Random scale-free networks are characterized by a

1. a **large number of small loops;**
2. an exponential number of loops of length of order ***N***;
3. the **most probable loop size of order N** with the proportionality constant depending on the considered random graph ensemble.
4. Random scale-free graphs **can fail to have an Hamilton cycle** even when they have a minimal connectivity grater of equal to 3 in the large ***N*** limit provided that the power-law exponent γ is sufficiently close to two, i.e. $\gamma<\gamma*$ with $\gamma*$ depending on the ensemble.

   G.B. and M. Marsili (JSTAT 2005)

Further steps:

- Calculate the fluctuations in the number of loops and the probability that the number of loops present in the typical network is grater than zero
- Generalize the calculation to correlated scale-free networks.