Clustering of random scale-free networks

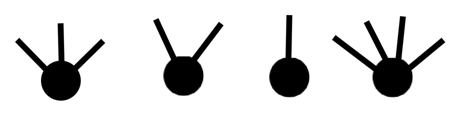
Pol Colomer-de-Simón[1] and Marián Boguñá[1]

[1] *Departament de Física Fonamental, Universitat de Barcelona, Martí i Franquès 1, 08028 Barcelona, Spain*

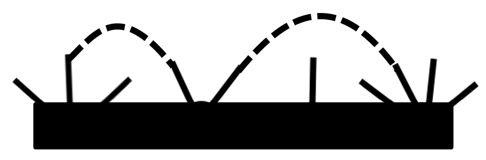
We derive the asymptotic behaviour of the clustering coefficient of scale-free random graphs generated by the configuration model with the degree distribution exponent 2 < γ < 3. Degree Heterogeneity increases the presence of triangles in many real networks even for extremely large networks. We also find that for values $\gamma \approx 2$, clustering is virtually size independent and, at the same time, becomes a *de facto* non self-averaging topological property.

Configuration model in the microcanonical ensemble

1. The degree of each node is fixed



2. The connections among nodes are realized in the most random way avoiding multiple-edges and self-edges.



In the absence of high degree nodes the Clustering coefficient of the resulting network is given by [3]:

$$C = \frac{\langle k(k-1)\rangle^2}{N\langle k\rangle^3}$$

That vanishes very fast for large system sizes $N \to \infty \Rightarrow C \to 0$

However in the case of a scale-free networks it predicts a behaviour $C \sim N^{(7-3\gamma)/(\gamma-1)}$ that diverges for γ < 7/3. This is because this derivation does not account for the structural correlations among degrees of connected nodes that appear in order to be able to close the network for degree distributions with γ < 3 [2].

Here we explain how to derive the correct scaling behaviour of clustering for scale-free random graphs with 2 < γ < 3. We also show that when $\gamma \approx 2$ clustering is very high and becomes nearly size independent.

For γ close to 2 clustering remains nearly constant. Clustering is surprisingly high

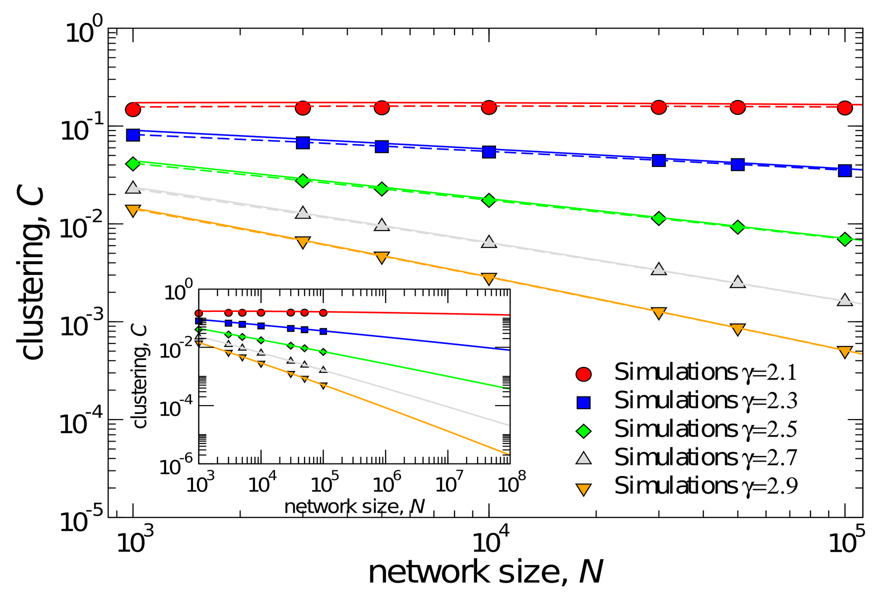


Fig 1. Clustering coefficient as measured in numerical simulations for different values of γ and size N with $\bar{\kappa}_{min} = 2$ and $\kappa_c = N^{1/(\gamma-1)}$. Each point is an average over different network realizations. Dashed lines are the exact numerical solution and the solid lines are our approximation. The inset shows an exploration up to size 10^4 using our approximation

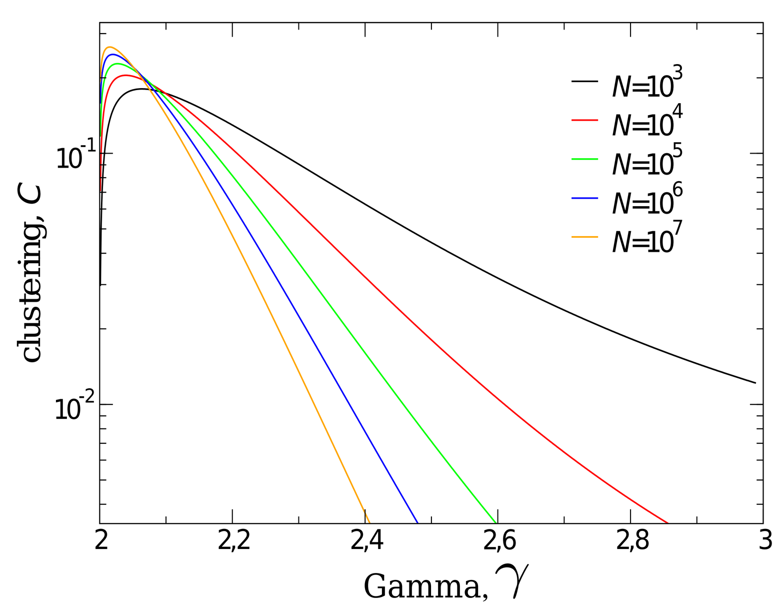


Fig 2. Clustering coefficient as a function of γ for fixed network size. Curves are evaluated from the approximation with $\bar{\kappa}_{min} = 2$ and $\kappa_c = N^{1/(\gamma-1)}$.

Conclusions:

Using the canonical ensemble of the configuration model we have been able to find both analytically and numerically the correct scaling behaviour of the clustering coefficient of the ensemble of scale-free random graphs with 2 < γ < 3. Interestingly. for (realistic) values of the exponent $\gamma \approx 2$, clustering remains nearly constant up to extremely large network sizes.

Observing the variance of the Clustering coeffcient in the numerical simulations, we found that in this case this topological property is not self-averaging.

These results are particulary important as the exponent value $\gamma \approx 2$ seems to be -for yet unknown reasons- the rule rather than the exception in real systems.

References:

[1] P. Colomer-de-Simon and M. Boguñá, ArXiv 1205.2877

[2] M. Boguñá, R.Pastor-Satorras, and A. Vespignani, Eur. Phys. J. B **38**, 205 (2004)

[3] M.E.J. Newman, *Random graphs as models of networks. In Handbook of graphs and Networks,* S.Bornholdt and H.G. Shuster (eds.) (Wiley-VCH, Berlin, 2003)

[4] D. Garlaschelli and M. I. Loffredo, Phys. Rev. E **78**, 015101 (2008)

[5] M. Boguñá and R. Pastor-Satorras, Phys. Rev. E **68**, 036112 (2003)

Configuration model in the Canonical ensemble

In this ensemble, each node is given not its actual degree but its expected degree.

1. Each node is assigned a hidden variable κ drawn from a density functions:

$$\rho(\kappa) \propto \kappa^{-\gamma} \quad \text{with} \quad 1 < \kappa < \kappa_c \quad \text{and} \quad \kappa_c \sim N^{1/(\gamma-1)}$$

2. We connect each pair of nodes with probability [4]:

$$r\left(\frac{\kappa\kappa'}{{\kappa_s}^2}\right) = \frac{\kappa\kappa'}{{\kappa_s}^2}\left(1 + \frac{\kappa\kappa'}{{\kappa_s}^2}\right)^{-1} \quad \text{where} \quad \kappa_s = \sqrt{\frac{(\gamma-1)N}{(\gamma-2)\kappa_{min}}}$$

Then the average degree of a node is:

$$\bar{k}(\kappa) = N\int_{\kappa_{min}}^{\kappa_c} d\kappa'\rho(\kappa')r\left(\frac{\kappa\kappa'}{\kappa_s^2}\right) \Rightarrow \bar{k}(\kappa) \propto \kappa$$

The Clustering coefficient of nodes of a certain hidden variable κ is calculated as [5]:

$$C(\kappa) = \left(\frac{N}{\bar{k}(\kappa)}\right)^2 \int_{\kappa_{min}}^{\kappa_c} d\kappa'\rho(\kappa') \int_{\kappa_{min}}^{\kappa_c} d\kappa''\rho(\kappa'') r\left(\frac{\kappa\kappa'}{\kappa_s^2}\right) r\left(\frac{\kappa\kappa''}{\kappa_s^2}\right) r\left(\frac{\kappa'\kappa''}{\kappa_s^2}\right)$$

If we average over all κ we get the Clustering coefficient *C:*

$$C = \int_{\kappa_{min}}^{\kappa_c} \rho(\kappa)C(\kappa)d\kappa$$

We solved analytically and we found [1]:

$$c(\kappa) \sim \frac{(\gamma-2)^2}{\kappa_s^{2(\gamma-2)}} \begin{cases} \theta(\gamma) + \Phi(-1,2,\gamma-2) & \kappa_c = \kappa_s \gg 1 \\ 2\psi(\gamma)\ln\left(\frac{\kappa_c}{\kappa_s}\right) & \kappa_c \gg \kappa_s \gg 1 \end{cases}$$

where $\psi(\gamma) = \Phi(-1,1,3-\gamma) + \Phi(-1,1,\gamma-2)$ and $\theta(\gamma) = -\pi^2 \cot\pi\gamma \csc\pi\gamma$

and Φ(z,a,b) is the transcendent Lerch function

The first line give the result for scale-free networks without structural correaltions

$C \sim N^{2-\gamma}$ when $\kappa_s \sim N^{1/2}$

The second line predicts $C \sim N^{2-\gamma}\ln N$ when $\kappa_c \sim N^{1/(\gamma-1)}$
Which corrects the incorrect scaling behavior predicted in the microcanonical ensemble.

For γ close to 2 the clustering is not a self-averaging topological property

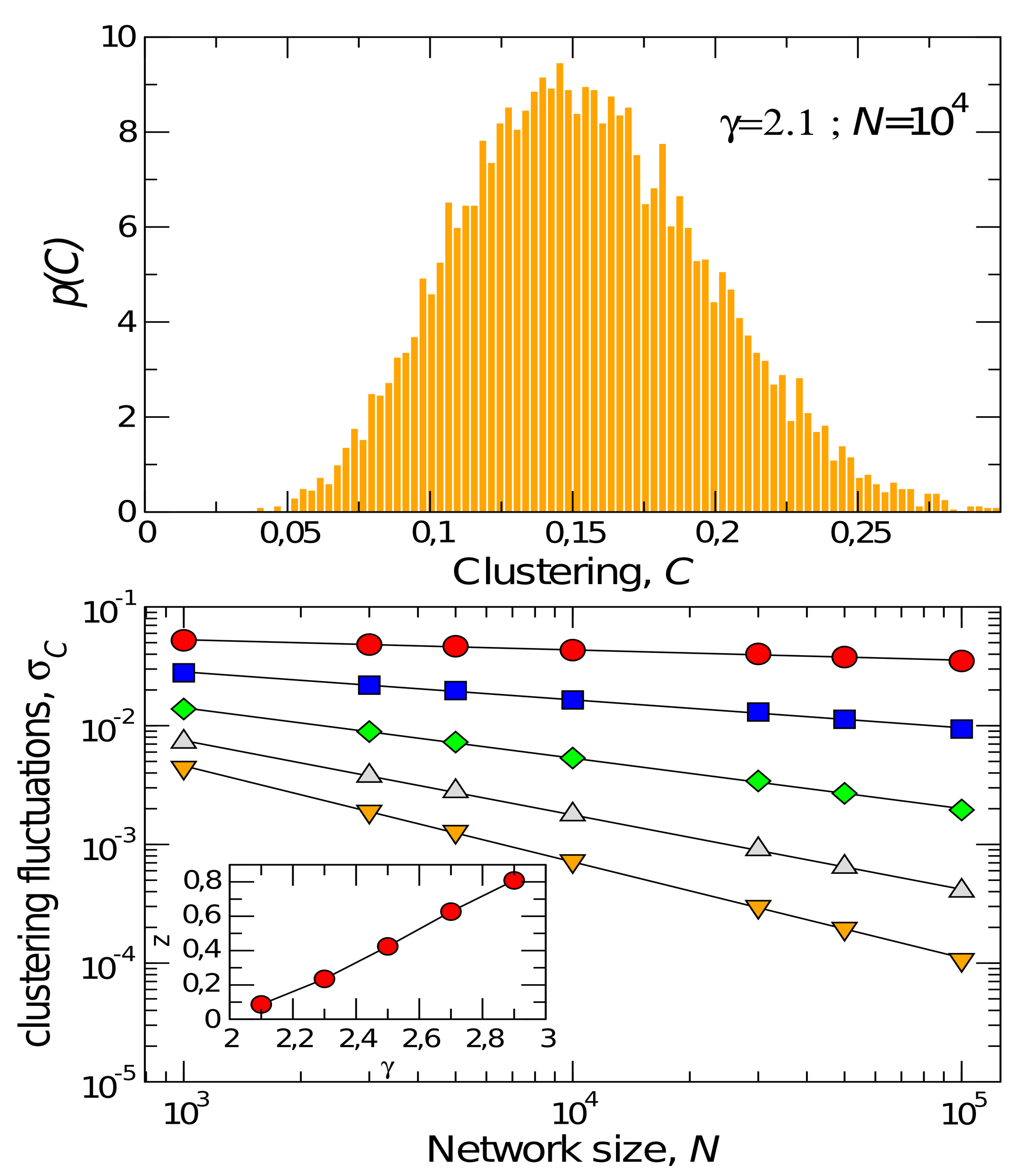


Fig 3. Sample-to-sample the fluctuations. The top plot shows the probability density function of the clustering coefficient as obtained from 10^4 network realizations for $\bar{\kappa}_{min} = 2$, $\kappa_c = N^{1/(\gamma-1)}$, γ=2 and $N = 10^4$. The bottom plot shows the standard deviation of this pdf for different values of γ as a function of the network size. Solid lines are power law fits of the form $\sigma_C \sim N^{-z}$. The exponent *z* is shown in the inset.