PRAMANA — journal of physics
© Indian Academy of Sciences
Vol. 71, No. 2
August 2008
pp. 391–401

Correlations and clustering in a scale-free network in Euclidean space

A K NANDI[1], G MUKHERJEE[1,2] and S S MANNA[1,*]
[1]Satyendra Nath Bose National Centre for Basic Sciences, Block-JD, Sector-III, Salt Lake City, Kolkata 700 098, India
[2]Bidhan Chandra College, Asansol 713304, Dt. Burdwan, India
[*]Corresponding author. E-mail: manna@bose.res.in

Abstract. Empirical study shows that many real networks in nature and society share two generic properties: they are scale-free and they display a high degree of clustering. Quite often they are modular in nature also, implying occurrences of several small tightly linked groups which are connected in a hierarchical manner among themselves. Recently, we have introduced a model of spatial scale-free network where nodes pop-up at randomly located positions in the Euclidean space and are connected to one end of the nearest link of the existing network. It has been already argued that the large scale behaviour of this network is like the Barabási–Albert model. In the present paper we briefly review these results as well as present additional results on the study of non-trivial correlations present in this model which are found to have similar behaviours as in the real-world networks. Moreover, this model naturally possesses the hierarchical characteristics lacked by most of the models of the scale-free networks.

Keywords. Scale-free network; Euclidean network; clustering.

PACS Nos 89.75.Hc; 89.20.Hh; 89.75.Fb; 05.60.-k

1. Introduction

In recent years empirical observations on a large number of real-world networks have revealed two common features in their structural properties. Namely, these networks have highly heterogeneous structures and it is apparent in their nodal degree distributions having power-law tails. For this reason they are called the scale-free networks (SFN). In addition, a significant amount of local correlations are observed in these networks, which are characterized by the clustering coefficients and are measured by the three-point correlation functions among the local nodes of the network [1–5].

A number of different models have been proposed for these networks over the last several years which mainly tried to reproduce these features. Among them one of the most well-known networks is the one by Barabási and Albert (BA) [1]. Though this model is very successful to generate an SFN with a non-trivial power-law degree distribution, no significant local correlation as present in real-world networks has been observed in this model [1,2].

One of the most important scale-free networks is Internet [6]. The difference the Internet has with other networks is that a large part of Internet is embedded in Euclidean space by the cable and wire connections [7–9]. Therefore, it is indeed necessary to frame a suitable model of an SFN whose nodes as well as links both lie on the Euclidean space and whose growth procedure is determined by considerations related to Euclidean measures. In a recent work [10] we have proposed a Euclidean scale-free network whose growth is controlled by a local selection rule. In the limiting case of very large network sizes this model is seen to follow the same degree distribution as in the BA model. In the present paper we argue using the numerical evidence that the same model network indeed has a hierarchical structure associated with non-trivial clustering coefficient as observed in real-world networks.

The number of links k meeting at a node of the network is called the 'degree' of that node. The probability that a randomly selected node has degree k is denoted by $P(k)$. For the large SFNs the degree distribution varies as a power law: $P(k) \sim k^{-\gamma}$, where γ is regarded as a critical exponent characterizing the degree distribution [2]. Secondly, the clustering coefficient for an arbitrary node i with degree k_i is defined as $\mathcal{C}_i = 2n_i/[k_i(k_i-1)]$, where n_i is the number of links among the k_i neighbours of i. Empirical measurements on the clustering properties of the real-world networks have revealed the following common features: (i) The clustering coefficient $\langle \mathcal{C} \rangle$ averaged over all nodes is significantly larger compared to the random networks of similar size [11], (ii) the clustering coefficients of these networks do not depend on their sizes N [2] and (iii) the clustering coefficient averaged over all nodes of same degree k shows the $\langle \mathcal{C}(k) \rangle \sim k^{-1}$ scaling, which is regarded as the signature of a hierarchical structure in many real-world networks [12].

Apart from this, many networks are found to be modular: small groups of nodes that are highly interconnected between themselves, but intergroup links are comparatively few in numbers. In social networks such modules represent groups of friends or coworkers. These numerous groups combine with each other in a hierarchical manner, generating a hierarchical network.

In the above description the networks are considered to have binary structures, i.e., a pair of nodes are either connected by a link or disconnected. However, in reality, even the linked pairs have a large variation of their strengths. Quite often, these strengths of ties between nodes are not similar and may even be highly heterogeneous. This strength of a link is called the weight of the link and such a network is called the weighted network. A number of examples from widely different fields can be cited for such networks, e.g., the number of passengers between a pair of airports [13], strength of pair-interaction between two species in ecological system [14], number of co-authored papers of two authors [15], magnitude of data traffic in a link of the Internet [6] or volume of trade between two countries are few well-known examples of weighted networks [16,17].

In this context one can define a nodal strength s_i in a weighted network which supports the total weight of all links meeting at the node i: $s_i = \sum_j w_{ij}$. For the world-wide airport network (WAN) it has been observed that the nodal strength averaged over all nodes of degree k varies as $\langle s(k) \rangle \propto k^{\beta}$ with $\beta > 1$ where the link weight is defined either by the number of passenger seats or by the Euclidean

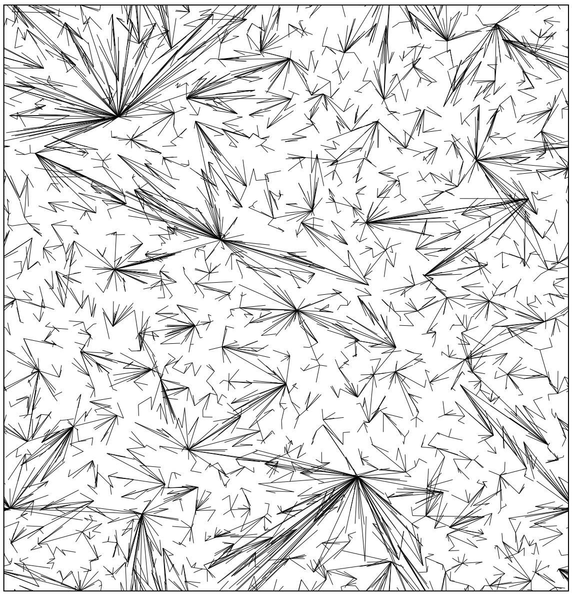

Figure 1. The picture of a network of $N = 2^{12}$ nodes, each node is a randomly positioned point on the unit square and is connected randomly to one end of the nearest link.

distance between successive stops [18]. Average link weight again scales with the product of the degree values of the two end-nodes: $\langle w_{ij}\rangle \propto (k_i k_j)^{\theta}$ with $\theta \approx 0.5$ for WAN [18]. For the metabolic networks where link weights represent the optimal metabolic fluxes, β is observed to be approximately 0.8 for *E. Coli*. A number of models have been proposed to reproduce similar non-linear behaviours [19–21].

A detailed look into the different unweighted networks shows that most models have difficulty in capturing simultaneously these two features, the high clustering and scale-free degree distribution.

Let us discuss the BA model in some more detail [1,2]. Starting from a small connected network, here the network grows incrementally by adding nodes one by one. The new nodes get connections to the already existing nodes of the network in a probabilistic fashion guided by a linear attachment probability rule ('rich get richer' principle) which guarantees that large degree nodes grow at higher rates. In BA model the degree exponent γ is known to be exactly 3.

On the other hand, the BA networks cannot reproduce the non-trivial correlations observed in many real-world networks. The BA networks produce a larger clustering coefficient than the corresponding random networks of same size. The average clustering coefficient decreases with the system size as $\mathcal{C}(N) \sim N^{-0.75}$ for BA network and $\sim (\ln N)^2/N$ for the random networks [22]. However for several real-world systems $\mathcal{C}(N)$ is practiclly independent of N. Moreover, in BA model $\mathcal{C}(k)$ is independent of k contrary to the empirical observation that $\mathcal{C}(k) \sim k^{-a}$.

In a recent paper [10], we questioned the general requirement of the 'linear attachment' rule of the BA model as a necessary criterion to achieve an SFN. We

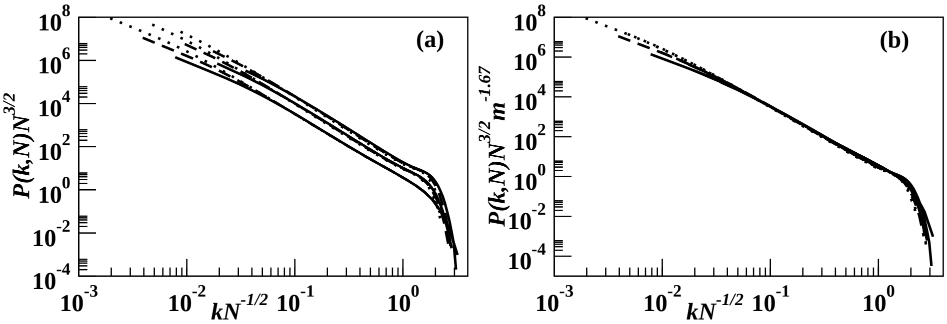

Figure 2. (**a**) Finite-size scaling analysis of the degree distribution for three system sizes, $N = 2^{14}$ (solid line), 2^{16} (dashed line), 2^{18} (dotted line) and for three values of the number of outgoing links $m = 1$, 2 and 3 (from left to right). (**b**) Same plot as in (a) but scaled with $m^{-1.67}$.

argued that at least for networks embedded in the Euclidean space like the Internet and the airport networks, existence of such a rule for the expanding network seems to be highly inappropriate. Qualitatively one can say that whenever a new computer is connected to the Internet and become a member of the world-wide Internet network, it always gets a connection to the existing local nodes of the Internet network. In fact one would hardly give any extra importance to large ISP hubs in Tokyo, Stuttgart or Chicago rather than small providers in his locality. Considering the whole world-wide Internet network, it is apparent that the new nodes pop up randomly in space and time without any spatial correlations. Similar arguments can be put forward for the airport network as well. A new airport in some remote city in some country is quite likely to be connected to the neighbouring airport first by direct non-stop flights and very unlikely to have cross-country inter-continental flights to begin with.

In ref. [10] we studied a growing network on Euclidean space where new nodes are added one by one and are connected to the neighbouring nodes of the growing network. We showed that even such a network has scale-free degree distribution. The degree exponent γ of this network is numerically found to be nearly 3 and we argued that in the limit $N \to \infty$ it is exactly the same as the BA network. In addition, these networks have non-trivial dependence of average strength on the nodal degrees.

In the present work we show that this model has additional realistic features. There exists a strong correlation between link weights and the topological properties. The assortative behaviour of the system is enhanced when the weighted definition of average nearest-neighbour degree is taken into account. Moreover, the presence of hierarchical architecture is also identified.

2. The model

In this section we describe our method of generating the scale-free network on the two-dimensional Euclidean space. The network is grown within a unit square on

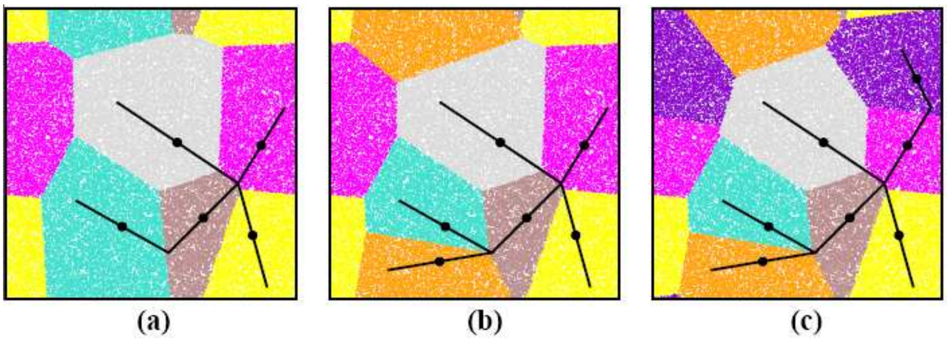

Figure 3. The associated Voronoi cells for the growing network: (**a**) Five links, (**b**) six links and (**c**) seven links. The whole space is partitioned into different cells around each link centre.

the x–y plane. We drop points one by one at randomly selected locations within this area. Their coordinates are denoted by $\{x_i, y_i\}, i = 1, N$ for N such points. These points are the nodes of the spatial network to be grown. Initially when two points are dropped we join them by a straight line which is the first link of the network. Then the third point is connected either to the point 1 or to the point 2 by the second link and this process is repeated one by one to make all N nodes connected to the growing network.

How a new node is connected to the existing network can be understood in the following way. In general every new node connects to one end of the nearest link. Suppose at a certain stage the network has t links connecting $t+1$ nodes. For connecting the $(t+2)$th node to the network, the nearest link centre is selected. One of the two end-nodes of the nearest link is then chosen with probability 1/2 and is connected to the new node to create the $(t+1)$th link (figure 1). These connections can be done very efficiently by keeping the local information into memory. The unit square area is divided into a lattice of size $\sqrt{N} \times \sqrt{N}$. As the network grows the serial numbers of the links whose centres are positioned within a lattice cell are stored at the associated lattice site. To find out the nearest link centre one starts from the cell of the $(t + 2)$th node and searches the lattice cells shell by shell till the nearest link centre is found out. When $t \sim N$, only the nearest shell needs to be searched.

After the network has grown to N nodes, the degree distribution $P(k, N)$ is calculated. From a direct measurement of the slope of the $\log P(k, N)$ vs. $\log k$ plot the exponent γ_k is estimated to be 3.00 ± 0.05, which is close to BA value. Moreover, a scaling of $P(k, N)$ is also studied (figure 2a):

$$P(k, N) \propto N^{-\eta}\mathcal{G}(k/N^{\zeta}), \tag{1}$$

where $\eta = 3/2$ and $\zeta = 1/2$ are used to obtain the best data collapse giving $\gamma_k = \eta/\zeta = 3$. The network so generated has a tree structure. However, networks having more general structures with multiple loops can be generated with m links coming out from every incoming node and are attached in a similar way to the first m neighbouring links in increasing order. In figure 2b we show that degree distributions for $m = 2$ and 3 can be scaled as $P(k, N)m^{-1.67} \propto N^{-\eta}\mathcal{G}(k/N^{\zeta})$.

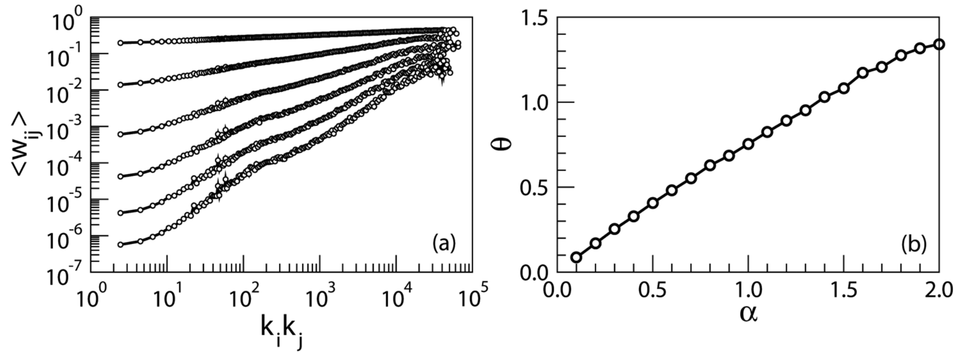

Figure 4. (**a**) Plot of the average link weight vs. the product of degree values of the two end-nodes of the link for a network of $N = 2^{14}$ nodes. Tuning parameter α as defined in eq. (2) has values 0.1, 0.4, 0.8, 1.2, 1.6 and 2 from top to bottom. Plot shows that slope θ of the curves increases with α. (**b**) Plot of the link weight-degree exponent $\theta(\alpha)$ vs. the tuning parameter α shows almost linear increase.

It is interesting to observe that when a new point is dropped its very position determines to which node this point would be connected. This is because around every link centre there is an associated surrounding region. All points of this region have this link centre situated at the shortest distance. This implies that given a connected network of t links, the whole space is partitioned into t non-overlapping regions similar to the well-known Voronoi cells [23]. When a new link is added into the system, the number of such cell increases and the area of each cell is rearranged (figure 3). The probability that a randomly selected point is within a particular cell is equal to the area of the cell. Since different Voronoi cells have different areas, the probability of selecting a link centre is in general non-uniform. It is known that for a two-dimensional Poisson–Voronoi tessellation, the cell sizes follow a gamma distribution whose width scales as $1/t$ [24]. Therefore, though for finite t the cell sizes are non-uniform, for $t \to \infty$ the cell size distribution is similar to a delta function implying that all cells have uniform size. In this limit the probability that a particular node of degree k will be linked is $k/2$ times the cell size – which gives rise to the linear preferential attachment as in BA model. Therefore, for a very large network ($N \to \infty$) in our model the node selection probability is different from that in BA model at early times (t small) whereas it asymptotically converges to the BA preferential rule in the limit $t \to \infty$. We conclude that the leading behaviour of our spatial network model is like the BA model. We have already seen that even for finite N the degree exponent $\gamma_k \approx 3$ as in the BA model.

This result can be interpreted as follows. Let us assume that the area of every Voronoi cell in the network with t links is uniform and is equal to $1/t$. This implies that the ith node with degree $k_i(t)$ has probability $k_i(t)/(2t)$ to be linked with the new $(t+2)$th node. The factor 2 comes from the fact that one of the two end-nodes of every link is selected with equal probability. Therefore, $\mathrm{d}k_i(t)/\mathrm{d}t = k_i(t)/(2t)$ and $k_i(t) = (t/i)^{1/2}$, which is exactly similar to the BA model resulting $P(k) \propto k^{-3}$.

Like weighted airport network [18] the strength s_i of a node i is measured as the sum of the Euclidean lengths raised to the power α for all links meeting at i:

$$s_i = \sum_j w_{ij} = \sum_j \ell_{ij}^{\alpha} \tag{2}$$

ℓ_{ij} being the length of the link between the nodes i and j and α is a continuously tunable parameter; the sum j runs over the nearest neighbours of i. Using the above definition of link weights and nodal strengths, one can analytically show that the strength distribution $P(s)$ follows a power-law with exponent $\gamma_s(\alpha) = 1+4/(2+\alpha)$ and using the general relation $\gamma_s = \gamma_k/\beta + 1 - 1/\beta$ [25] and using $\gamma_k = 3$ we get $\beta(\alpha) = 1 + \alpha/2$. The exponent of the weight distribution $P(w)$ varies as $\gamma_w(\alpha) = 1 + 2/\alpha$. A number of checks have been done to verify these results [10].

2.1 *Weight–topology correlation*

A strong correlation between weight and the topological properties is observed for this model. Like world-wide airport network (WAN), the dependence of the link weight w_{ij} on the degrees of the end-point nodes k_i and k_j can be well approximated by the relation $\langle w_{ij}\rangle \propto (k_i k_j)^{\theta}$. For a large network with $N = 2^{14}$, the average weights are calculated for different values of α.

In figure 4a, normalized link weight and product of the degree values of two end-nodes is plotted for six different α values. In the double-logarithmic plot it is found that the variation is a power-law over decades and the slope of the exponent θ systematically increases with α.

In figure 4b, the link weight-degree exponent θ is plotted with α. Exponent θ found to grow with α. It is expected here as with α link weights increases but the static structure of the network and the degree distribution remain unchanged. Hence the degree of any two end-nodes of a given link remains unaltered but the weight of the link increases. Plot shows that θ increases almost linearly with α.

2.2 *Degree–degree correlation*

In figure 5, average nearest-neighbour degree $\langle k_{\rm nn}\rangle$ of any arbitrary node of degree k is plotted with k. It gives an idea of the degree–degree correlation of the neighbouring nodes. The exact conditional probability $P(k_1|k)$ that a node of degree k is connected to a node of degree k_1 is in general difficult to find out and interpret. The average nearest-neighbour degree of a node i of degree k_i is defined as [25]:

$$k_{{\rm nn},i} = \frac{1}{k_i}\sum_j k_j, \tag{3}$$

where the sum j runs over the nearest neighbours of i.

Now, when averaged over the nodes according to degree k we find $k_{\rm nn}(k)$ which is related to $P(k_1|k)$ as

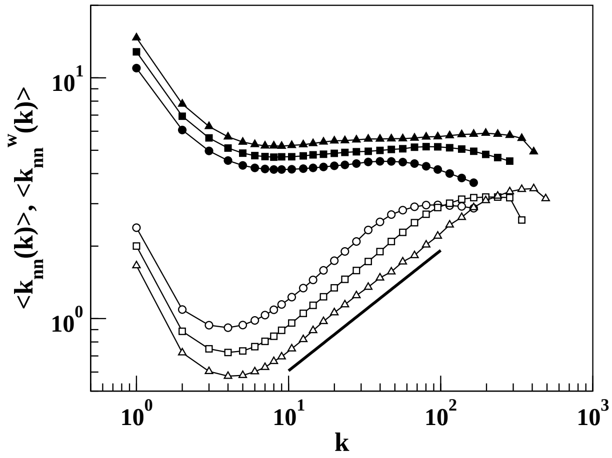

Figure 5. The average degree $\langle k_{\rm nn}(k)\rangle$ and the average weighted degree $\langle k^{\rm w}_{\rm nn}(k)\rangle$ of a neighbour of a node of degree k are plotted with k with solid and empty symbols for networks with $m = 1$. Plots for three different system sizes $N = 2^{12}, 2^{14}$ and 2^{16} are denoted by circles, squares and triangles respectively. The solid line is guide to the eye having a slope 1/2.

$$k_{\rm nn}(k) = \frac{1}{N_k}\sum_{k_i=k} k_{{\rm nn},i} = \sum_{k_1} k_1 P(k_1|k). \tag{4}$$

If there is no degree–degree correlation, $P(k_1|k)$ is a function of k_1 only and $k_{\rm nn}(k)$ is a constant. But if $k_{\rm nn}(k)$ increases with k (assortative) it tells that large degree nodes prefer to join with the large degree nodes and disassortative if $k_{\rm nn}(k)$ decreases with k.

In the case of weighted networks the appropriate characterization of the assortative behaviour is obtained by the weighted average nearest neighbours degree, defined as

$$k^{\rm w}_{{\rm nn},i} = \frac{1}{s_i}\sum_j w_{ij}k_j, \tag{5}$$

where the sum j runs over the nearest neighbours of i.

Here the behaviour of the function $k^{\rm w}_{\rm nn}(k)$ (defined as the average of $k^{\rm w}_{{\rm nn},i}$ over all vertices with degree k), marks the weighted assortative or disassortative properties considering the actual interactions among the system's elements [25].

Plot shows that the unweighted network has no significant degree–degree correlation, similar to BA model. But the weighted network on the contrary show clear assortative property which is common in most of the social networks.

2.3 *Clustering and hierarchy*

In figure 6a, we plotted the average clustering coefficient of the network for different system sizes N. Unlike BA model we find that the clustering coefficient is quite high

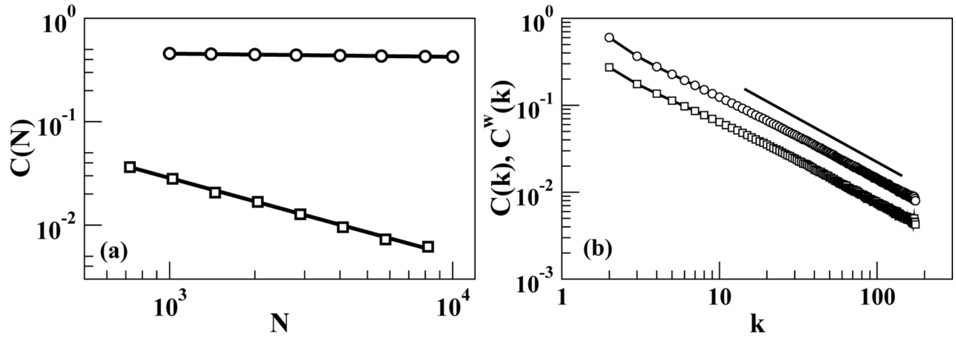

Figure 6. (**a**) Comparison of the variation of overall clustering coefficient $\mathcal{C}(N)$ with the network size N ($m = 2$) for the present model (circles) and the BA model (squares). The slope is -0.03 in our model, which is much larger than -0.75 of the BA model. (**b**) Plot of the average clustering coefficient $C(k)$ and the weighted clustering coefficient $C^{\mathrm{w}}(k)$ of nodes of degree k with k of a network of size $N = 10{,}000$ nodes ($m = 2$). For both curves the slopes are ≈ -1 (as indicated by the solid line).

and its variation with system size is almost negligible. In the log–log plot $C(N) \sim N^{-\mu}$, where $\mu \sim 0.03$. This shows a marked departure from the BA behaviour. The search for a new connection as far as possible in the local neighbourhood gives rise to high local clustering and as the nodes pop-up randomly in the unit square this feature is common in every portion of the square. Hence the average clustering coefficient is high and is practically independent of N.

This topological clustering does not take into account the fact that all neighbours are not equal. Weighted clustering coefficient is introduced by Barrat *et al* [25] which combines the topological information with the link weight distribution of the network. If node j and node h are neighbours of node i, then the weighted clustering coefficient can be defined as

$$C^{\mathrm{w}}(i) = \frac{1}{s_i(k_i - 1)} \sum_{j,h} \frac{(w_{ij} + w_{ih})}{2} a_{ij} a_{ih} a_{jh}, \tag{6}$$

where a_{ij} stands for the adjacency matrix of the network, which is 1 if there is a link between node i and node j, otherwise 0.

This quantity $C^{\mathrm{w}}(i)$ counts for each triplet formed in the neighbourhood of vertex i, the weights of the two participating edges of i. C^{w} and $C^{\mathrm{w}}(k)$ are defined as the weighted clustering coefficients averaged over all vertices of the network and over all vertices with degree k, respectively.

Hierarchy in a network is characterized in a quantitative manner following the finding in [26], that in deterministic scale-free networks the clustering coefficient of a node with k links follows the scaling law

$$C(k) \sim k^{-1}. \tag{7}$$

Barabási *et al* argued that this scaling is applicable for other scale-free networks also [12].

In figure 6b we find that $C(k) \sim k^{-\nu}$, where ν is almost 1 and the weighted and unweighted plots have similar slopes. This is a clear signature of the presence of hierarchical organization in the network which is again different from the BA behaviour.

To summarize, we argue that in many real-world networks it is more likely that the new nodes get their connections in the local neighbourhood and that produces the scale-free degree distribution. Indeed, a spatial scale-free network is grown using the criterion of the local selection rule. This network shows a non-linear dependence of the nodal strength on the degree. The clustering behaviour is very different from the BA model, showing almost negligible variation with the system size. Moreover, the system shows a strong hierarchical scaling found to be present in many real networks.

Acknowledgement

GM and SSM thankfully acknowledge hospitality at the Indian Institute of Technology Guwahati during the conference on 'Statistical Physics Approaches to Multidisciplinary Problems'.

References

[1] A L Barabási and R Albert, *Science* **286**, 509 (1999)
[2] R Albert and A-L Barabási, *Rev. Mod. Phys.* **74**, 47 (2002)
[3] M E J Newman, *SIAM Rev.* **45**, 167 (2003)
[4] I Bose, in: *Frontiers in biophysics* edited by T P Singh and C K Dasgupta (Allied Publishers, New Delhi, 2005) p. 87
[5] S N Dorogovtsev and J F F Mendes, *Evolution of networks* (Oxford University Press, 2003)
[6] R Pastor-Satorras and A Vespigniani, *Evolution and structure of internet: A statistical physics approach* (Cambridge University Press, Cambridge, 2004)
[7] M Faloutsos, P Faloutsos and C Faloutsos, *Proc. ACM SIGCOMM, Comput. Commun. Rev.* **29**, 251 (1999)
[8] R Pastor-Satorras, A Vazquez and A Vespignani, *Phys. Rev. Lett.* **87**, 258701 (2001)
[9] S H Yook, H Jeong and A-L Barabási, *Proc. Natl Acad. Sci. (USA)* **99**, 13382 (2002)
[10] G Mukherjee and S S Manna, *Phys. Rev.* **E74**, 036111 (2006)
[11] D J Watts and S H Strogatz, *Nature (London)* **393**, 440 (1998)
[12] E Ravasz and A L Barabási, *Phys. Rev.* **E67**, 026112 (2003)
[13] L A N Amaral, A Scala, M Barthélémy and H E Stanley, *Proc. Natl Acad. Sci. (USA)* **97**, 11149 (2000)
[14] A E Krause, K A Frank, D M Mason, R E Ulanowicz and W W Taylor, *Nature (London)* **426**, 282 (2003)
[15] M E J Newman, *Phys. Rev.* **E64**, 016131 (2001)
[16] M A Serrano and M Boguna, *Phys. Rev.* **E68**, 015101 (R) (2003)
[17] K Bhattacharya, G Mukherjee, J Saramki, K Kaski and S S Manna, *J. Stat. Mech.* P02002 (2008)
[18] A Barrat, M Barthélémy and A Vespignani, *J. Stat. Mech.* P05003 (2005)
[19] W X Wang, B H Wang, B Hu, G Yan and Q Ou, *Phys. Rev. Lett.* **94**, 188702 (2005)

[20] G Bianconi, *Europhys. Lett.* **71**, 1029 (2005)
[21] K-I Goh, B Kahng and D Kim, *Phys. Rev.* **E72**, 017103 (2005)
[22] K Klemm and V M Eguluz, *Phys. Rev.* **E65**, 057102 (2002)
[23] G Voronoi and J Reine, *Angew. Math.* **134**, 198 (1908)
[24] F Járai-Szabó and Z Néda, *Physica* **A385**, 518 (2007)
[25] A Barrat, M Barthélémy, R Pastor-Satorras and A Vespignani, *Proc. Natl Acad. Sci. (USA)* **101**, 3747 (2004)
[26] S N Dorogovtsev, A V Goltsev and J F F Mendes, *Phys. Rev.* **E65**, 066122 (2002)