Physica A 466 (2017) 105–110

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

# Analytic solutions for links and triangles distributions in finite Barabási–Albert networks

Ricardo M. Ferreira [a,*], Rita M.C. de Almeida [a,b], Leonardo G. Brunnet [a]

[a] Instituto de Física, Universidade Federal do Rio Grande do Sul, Av. Bento Gonçalves 9500, C.P. 15051 - 91501-970 Porto Alegre, RS, Brazil

[b] Instituto Nacional de Ciência e Tecnologia - Sistemas Complexos, Brazil

## HIGHLIGHTS

- Master equations for nodes, links and triangles for Barabási–Albert model are proposed.
- Finite time analytic solutions for nodes, links and triangles are found.
- Method may be applied to different growing network models.

## ARTICLE INFO

*Article history:*
Received 15 March 2016
Received in revised form 8 August 2016
Available online 31 August 2016

*Keywords:*
Growing networks
Exact solutions
Master equation
Finite size

## ABSTRACT

Barabási–Albert model describes many different natural networks, often yielding sensible explanations to the subjacent dynamics. However, finite size effects may prevent from discerning among different underlying physical mechanisms and from determining whether a particular finite system is driven by Barabási–Albert dynamics. Here we propose master equations for the evolution of the degrees, links and triangles distributions, solve them both analytically and by numerical iteration, and compare with numerical simulations. The analytic solutions for all these distributions predict the network evolution for systems as small as 100 nodes. The analytic method we developed is applicable for other classes of networks, representing a powerful tool to investigate the evolution of natural networks.



The pioneering works by Barabási and Albert [1,2] propose a model that builds a network adding nodes and links with a preferential attachment mechanism, yielding a degree distribution following a power-law. Here degree is defined as the number of neighbors of a given node. Barabási–Albert model was applied to a wide variety of networks such as actor collaboration, WWW, power grid data and protein–protein interaction (PPI) [3,4]. It emerges as a possible representative of a wide class of natural systems. However, it is a common controversy whether a given system may or may not be reduced to a Barabási–Albert network, with notorious examples given by metabolic pathways or PPI networks [5–7]. In fact, even networks simulated following Barabási and Albert recipe may show deviations from the typical topological statistic measures due to finite size effects. In real data such deviations may stem either from these finite size effects or from relevant differences in the underlying growth dynamics. In this context, analytic results are most welcome.

Our analytic characterization focuses on the dependence with time $t$ of three distribution functions: the number of nodes of degree $k$, $N(k,t)$; the number of links between nodes with degrees $k$ and $k'$, $L(k,k',t)$; and the number of triangles involving degrees $k$, $k'$ and $k''$, $\Delta(k,k',k'',t)$. The latter offers an insight of the neighborhood structure of nodes, whose common measure is the clustering coefficient defined as the number of triangles for which this node is a vertex, divided by

* Corresponding author.
*E-mail address:* melo.ferreira@ufrgs.br (R.M. Ferreira).

http://dx.doi.org/10.1016/j.physa.2016.08.018
0378-4371/

the number of possible triangles formed by the node itself and two of its neighbors [8]. Examples of clustering-dependent dynamics are cascading failures [9] and epidemics [10,11].

The only exact, analytic result for these three distributions when $t \to \infty$ is limited to the power law $N(k, t) \propto k^{-3}$ for Barabási–Albert and related models [1,2,12–14]. Further information on these distributions in the $t \to \infty$ limit is obtained only from simulations, as far as we know.

Nevertheless, network size is important. For example, infinite networks results may not discern between the effects due to the growth dynamics or the system finite size [3,15]. Additionally, epidemic outbreaks depend on network size [16], and social [10] and protein association networks [17–19] are typically smaller than 30 000 nodes. For finite networks some approximated results are available in the literature. Fotouhi and Rabat [20] analytically obtained a solution for $N(k, t)$, valid for $k$ less than the initial number of nodes, $N_0$, presenting a finite size correction and agreeing with previous results in the limit $t \to \infty$. Also, the average clustering coefficient of a Barabási–Albert network goes to zero for $t \to \infty$, but for finite time, the clustering coefficient for a given node was described by Fronczak et al. [21], using a mean-field approach, obtaining a dependence with $\ln(t)$. However, this result agrees with simulation data only for networks greater than 40 000 nodes [21]. Alternatively, Vazquez et al. correctly estimated the average number of triangles of a node with degree $k$ in finite biological networks based on the clustering coefficient of hierarchical scale-free networks [22]. Finally, the total number of triangles in the networks goes as $\ln(t)$, as obtained by Bianconi and Capocci [23], proving that the network clustering coefficient goes to zero for infinite networks.

In this work we analytically obtain the distributions functions $N(k, t)$, $L(k, k', t)$, and $\Delta(k, k', k'', t)$ as continuous time limit solutions of master equations and validate the results by numerically iterating these equations for discrete time and comparing to simulations of 10 000 random networks generated following the usual Barabási–Albert algorithm [1,2]. The analytic expressions are obtained using a method that is valid for network sizes smaller than 100 nodes, enough time to fade out the initial condition transient. Furthermore, as we show below, this method is straightforwardly extendable to other algorithms provided that the corresponding master equations may be decoupled.

The master equations and solutions are obtained as follows. We assume networks starting with $N_0$ nodes, each having $k_0$ neighbors. For simplicity $k_0$ is assumed to be greater than $m$, the degree of the new node added at each time step. Following Barabási–Albert algorithm, the probability $\Pi(k, t)$ that a node with degree $k$ receives a new connection is given by

$$\Pi(k, t) = \frac{mk}{\sum_{j=1}^{m_0+t-1} k_j} = \frac{mk}{2mt + N_0k_0}. \tag{1}$$

The master equation for the number of nodes with degree $k$, $N(k, t)$, can be written as

$$N(k, t+1) = N(k, t) + \delta_{k,m} + \frac{m(k-1)}{(2mt + N_0k_0)}N(k-1, t) - \frac{mk}{(2mt + N_0k_0)}N(k, t). \tag{2}$$

The second term on the right side of the equation represents the creation of nodes with degree $m$, the third (fourth) term represents the probability of a node with degree $k-1$ ($k$) to receive a link, increasing (decreasing) the number of nodes with degree $k$. Thus, this master equation describes the introduction of new nodes with degree $m$ and a flux of nodes increasing their degrees due to the attachment of new connections. This equation has been originally obtained by Dorogovtsev and Mendes [12]. The analogous master equations for the number of links $L(k, k', t)$ between nodes of degrees $k$ and $k'$ and number of triangles $\Delta(k, k', k'', t)$ formed by nodes of degrees $k$, $k'$, and $k''$ are

$$\begin{aligned} L(k, k', t+1) = {} & L(k, k', t) + \delta_{k,m}\frac{m(k'-1)}{2(2mt + N_0k_0)}N(k'-1, t) + \delta_{k',m}\frac{m(k-1)}{2(2mt + N_0k_0)}N(k-1, t) \\ & + \frac{m(k-1)}{(2mt + N_0k_0)}L(k-1, k', t) + \frac{m(k'-1)}{(2mt + N_0k_0)}L(k, k'-1, t) - \frac{m(k+k')}{(2mt + N_0k_0)}L(k, k', t) \end{aligned} \tag{3}$$

and

$$\begin{aligned} \Delta(k, k', k'', t+1) = {} & \Delta(k, k', k'', t) + \delta_{k',m}\frac{m^2(m-1)^2(k-1)(k''-1)}{6(2mt + N_0k_0)^2}L(k-1, k''-1, t) \\ & + \delta_{k,m}\frac{m^2(m-1)^2(k'-1)(k''-1)}{6(2mt + N_0k_0)^2}L(k'-1, k''-1, t) \\ & + \delta_{k'',m}\frac{m^2(m-1)^2(k-1)(k'-1)}{6(2mt + N_0k_0)^2}L(k-1, k'-1, t) \\ & + \frac{m(k-1)}{(2mt + N_0k_0)}\Delta(k-1, k', k'', t) + \frac{m(k'-1)}{(2mt + N_0k_0)}\Delta(k, k'-1, k'', t) \\ & + \frac{m(k''-1)}{(2mt + N_0k_0)}\Delta(k, k', k''-1, t) - \frac{m(k+k'+k'')}{(2mt + N_0k_0)}\Delta(k, k', k'', t). \end{aligned} \tag{4}$$

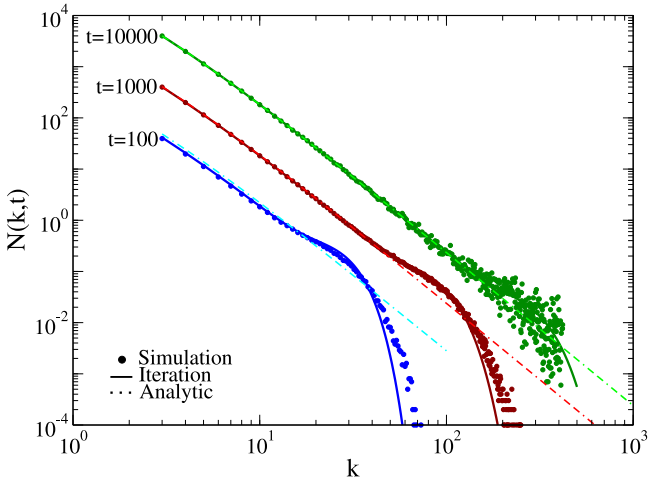


**Fig. 1.** Number of nodes with degree $k$ for different times. This figure presents a comparison of the analytic solution, the iteration results and simulation data for $N(k, t)$, for different networks sizes, averaged over 10 000 networks. The network size is given by time $t$. Points represent data obtained through simulation of Barabási–Albert algorithm, the solid lines represent iteration of master equation and dotted lines the analytic solution.

Again, the first terms in these equations, containing the $\delta$-distributions, account for the creation of new links or triangles while the other terms are the evolution of links and triangles due to the change in degree of one participant node.

For large networks continuous time may be assumed and Eq. (2) is written as a differential equation:

$$\frac{\partial N(k,t)}{\partial t} = \delta_{k,m} + \frac{m(k-1)}{(2mt + N_0k_0)}N(k-1,t) - \frac{mk}{(2mt + N_0k_0)}N(k,t), \tag{5}$$

representing a set of coupled differential equations, coupling each degree $k$ to its predecessor $k-1$. To solve these equations we begin with $k = m$. By our construction, $N(m-1, t)$ is always zero, and thus the first differential equation is decoupled:

$$\frac{\partial N(m,t)}{\partial t} = \delta_{k,m} - \frac{m^2}{(2mt + N_0k_0)}N(m,t). \tag{6}$$

This non-homogeneous differential equation, with initial condition $N(m, 0) = 0$, can be solved using Green's functions, yielding

$$N(m,t) = \frac{N_0k_0 + 2mt}{m(m+2)} - \frac{(N_0k_0)^{1+m/2}}{m(m+2)(N_0k_0 + 2mt)^{m/2}}. \tag{7}$$

In the limit $t \gg N_0k_0$ the transient term tends to zero and can be neglected resulting

$$N(m,t) = \frac{N_0k_0 + 2mt}{m(m+2)}. \tag{8}$$

This expression is now inserted in the differential equation for $N(m+1, t)$, what removes the coupling for the second equation, and again a solution may be obtained. This process may be repeated *ad infinitum*, and the expression for each degree may be written as

$$N(k,t) = \frac{(2mt + N_0k_0)(m+1)}{k(k+1)(k+2)}. \tag{9}$$

Fig. 1 presents the results from simulations, iteration of the master equation, and the analytic solution, Eq. (9), for the number of nodes with degree $k$ for three different times. Master equation iteration and the analytic solution fall one above the other, and both agree with simulations for small $k$, deviating only for high values of node degree. This deviation was first described by Dorogovtsev et al. [13] who established $k = \sqrt{t}$ as the upper limit for any analytic description of the model. To verify the limit, Fig. 2 shows the distribution $N(k, t)$ scaled by $\sqrt{t}$. We can see that the divergence between the analytic solutions and the simulations appears for values $\frac{k}{\sqrt{t}} > 1$. For clarity, the subsequent results will be presented scaled by $\sqrt{t}$, figures with no scaling can be found in Supplemental material at http://dx.doi.org/10.1016/j.physa.2016.08.018.

Eqs. (3) and (4) can be solved with the same method applied to Eq. (2). Considering continuous time, we write a set of coupled differential equations associated with both master Eqs. (3) and (4). Solving these equations iteratively, the

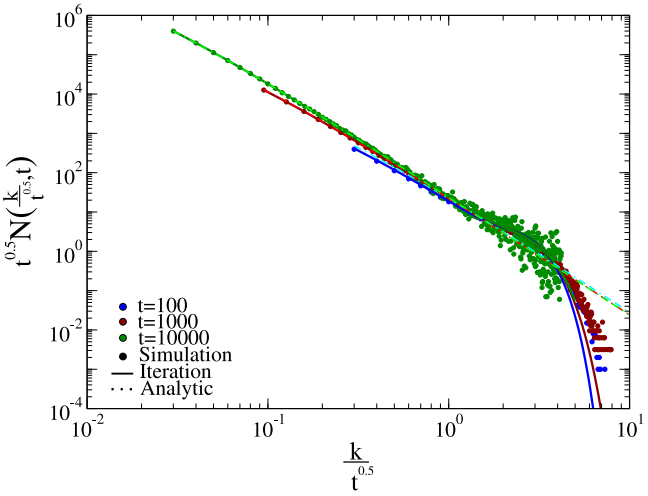


**Fig. 2.** Rescaling of $N(k, t)$ by $\sqrt{t}$. This figure presents a rescaling of the number of nodes with degree $k$, comparing the analytic solution, the iteration results and simulation data for $N(k, t)$, for different networks sizes, averaged over 10 000 networks. The network size is given by time $t$. Points represent data obtained through simulation of Barabási–Albert algorithm, the solid lines represent iteration of master equation and dotted lines the analytic solution. Different times are color coded, blue for $t = 100$, red for $t = 1000$, and green for $t = 10\,000$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

expression for any set of degrees $k$ and $k'$ is given by (see details in Supplemental material at http://dx.doi.org/10.1016/j.physa.2016.08.018,

$$L(k,k',t) = \Bigg[\sum_{i=1}^{k'-m}\binom{k-m+i-1}{i-1}\frac{(m+1)(2mt+N_0k_0)}{2(k'+1-i)(k'+2-i)}\frac{(k-1)!}{(m-1)!}\frac{(k'-1)!}{(k'-i)!}\frac{(k'+m-i+2)!}{(k+k'+2)!} + \sum_{i=1}^{k-m}\binom{k'-m+i-1}{i-1}\frac{(m+1)(2mt+N_0k_0)}{2(k+1-i)(k+2-i)}\frac{(k'-1)!}{(m-1)!}\frac{(k-1)!}{(k-i)!}\frac{(k+m-i+2)!}{(k+k'+2)!}\Bigg] \tag{10}$$

and

$$\begin{aligned}\Delta(k,k',k'',t) = \Bigg[&\sum_{i=0}^{k'-m}\sum_{j=0}^{k''-m}\frac{(k-m+i+j)!}{(k-m-2)!i!j!}\frac{(k-1)!}{(m-1)!}\frac{1}{6}m(m-1)^2 \\ &\times\frac{(k'-1)!}{(k'-i-2)!}\frac{(k''-1)!}{(k''-j-2)!}\frac{(m+k'-i+k''-j)!}{(k+k'+k'')!}\frac{L(k'-i,k''-j,t)}{2mt+N_0k_0} \\ &+\sum_{i=0}^{k-m}\sum_{j=0}^{k''-m}\frac{(k'-m+i+j)!}{(k'-m-2)!i!j!}\frac{(k'-1)!}{(m-1)!}\frac{1}{6}m(m-1)^2 \\ &\times\frac{(k-1)!}{(k-i-2)!}\frac{(k''-1)!}{(k''-j-2)!}\frac{(m+k-i+k''-j)!}{(k+k'+k'')!}\frac{L(k-i,k''-j,t)}{2mt+N_0k_0} \\ &+\sum_{i=0}^{k-m}\sum_{j=0}^{k'-m}\frac{(k''-m+i+j)!}{(k''-m-2)!i!j!}\frac{(k''-1)!}{(m-1)!}\frac{1}{6}m(m-1)^2 \\ &\times\frac{(k-1)!}{(k-i-2)!}\frac{(k'-1)!}{(k'-j-2)!}\frac{(m+k-i+k'-j)!}{(k+k'+k'')!}\frac{L(k-i,k'-j,t)}{2mt+N_0k_0}\Bigg].\end{aligned} \tag{11}$$

Observe that in the limit $t \gg N_0k_0$, Eq. (10) is also valid for $k \leq m$ and $k' \leq m$, when the above equation is null. By definition, the number of links between nodes with degrees $k$ and $k'$ is equal to the number of links between nodes with degrees $k'$ and $k$. Therefore $L(k, k', t)$ must be symmetrical regarding changes between $k$ and $k'$. This symmetry can be found in the master equation, Eq. (3), and its solution Eq. (10). Fig. 3 presents $tL(k/\sqrt{t}, k'/\sqrt{t}, t)$ for $k' = 0.5\sqrt{t}$ and $k' = \sqrt{t}$. It is clear that the analytic solution fits perfectly both simulation and iteration up to $k \sim \sqrt{t}$. Above this limit the simulation and iteration solutions start deviating from the analytic solutions.

Eq. (11) is symmetrical relative to any change between $k$, $k'$ and $k''$, as expected, given the symmetry of the model. Note that this solution is also time independent, since there $L(k, k', t)$ is divided by its time dependence. This occurs because when solving the differential equations we neglect the transient, and the remaining expression results constant in time.

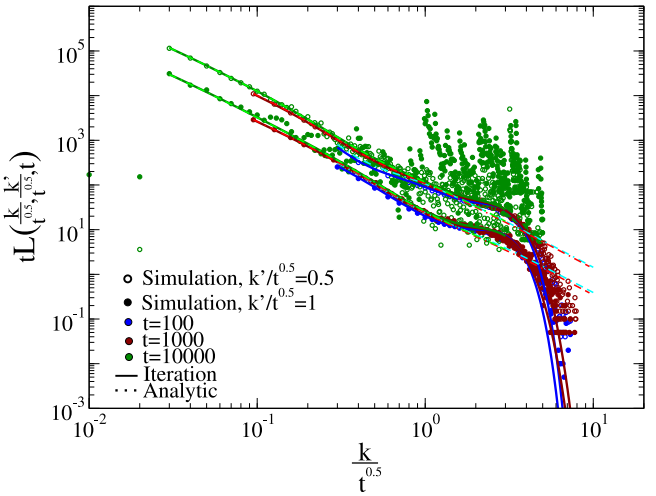


**Fig. 3.** Rescaling of number of links between degrees $k$ and $k'$. Bidimensional view of the tridimensional function $L(k, k', t)$ rescaled by $\sqrt{t}$ selecting two values of $k'/\sqrt{t}$, distinguished by solid and empty symbols, and three values of $t$, distinguished by different colors. Simulation data is represented by points, numerical iteration by continuous lines, and analytic solution by dotted lines. Blue symbols and lines represent $t = 100$, red symbol and lines represent $t = 1000$, and green symbols and lines represent $t = 10\,000$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

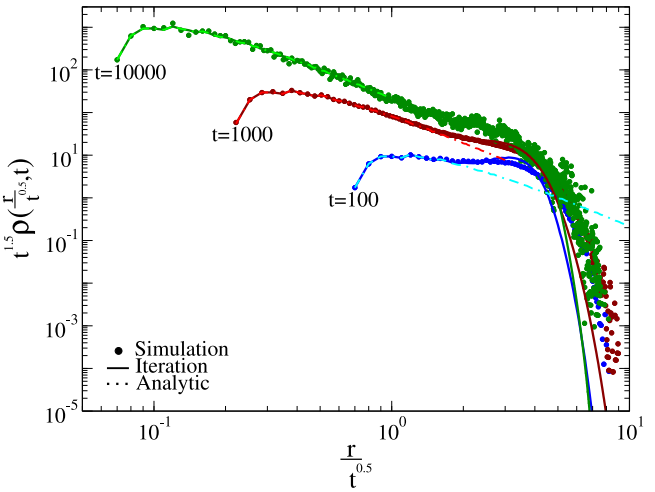


**Fig. 4.** Function $\rho(r)$ rescaled by $\sqrt{t}$. This figure presents three different network sizes and compares simulation, represented by points, iteration, by solid lines, and analytic solution, by dotted lines. To compare with different network sizes, the solution was rescaled with the corresponding times.

This result seems to be counter-intuitive since the model applies for growing networks. However the dynamics of triangle formation depends also on link distribution, hence being more complex. Therefore the behavior of $\Delta(k, k', k'', t)$ may be elusive. To verify Eq. (11) we introduce the one dimensional function $\rho(r)$ as the number of triangles between nodes with degrees $k, k'$ and $k''$ with $r = \sqrt{k^2 + k'^2 + k''^2}$, that is,

$$\rho(r) = \sum_{k} \sum_{k'} \sum_{k''} \delta_{r,\sqrt{k^2+k'^2+k''^2}} \Delta(k, k', k'', t). \tag{12}$$

Fig. 4 shows $\rho(r)$ scaled by $\sqrt{t}$. As expected, the simulations do not present the same results for different times. However, it is possible to note that the curves rapidly converge to the analytic solution for small values of $r$. This suggests that the growth dynamic acts on triangles with higher $r$. A possible explanation for this behavior lies precisely in the preferential attachment mechanism. Since a new node tends to connect to nodes with higher degrees, a triangle will be created when two high degree nodes are connected to each other. Also, our analytic solution agrees with this assumption. Since Eq. (11) represents the stationary state, and it agrees with the simulation for values of $r$ up to $\sqrt{t}$, it also indicates that the number of triangles tends rapidly to a stationary state and any difference between Eq. (11) and the simulation is given by the transient effect of the growth dynamics acting in the most connected nodes.

In summary, using this method for solving the differential equations associated with the discrete time master equations we obtain analytic solutions for the distributions of number of nodes, links and triangles of the Barabási–Albert model. These results have been verified by numerical iteration of the master equations and simulations, and proved to be valid for small networks, far from the limit of infinite time. These solutions describe accurately many real networks, failing only for $k > \sqrt{t}$, as predicted by Dorogovtsev et al. In particular, the number of triangles offers a more detailed vision of the neighborhood than average clustering coefficient. Finally, the method employed is general and may be applied to different growing network models.

## Acknowledgments

We acknowledge support from the Centro de Física Computacional, Universidade Federal do Rio Grande do Sul. This work has been partially supported by Brazilian agencies FAPERGS, CAPES, and CNPq. It is part of the project PRONEX-FAPERGS 10/0008-0.

## Appendix A. Supplementary material

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.physa.2016.08.018.

## References

[1] A. Barabási, R. Albert, Emergence of scaling in random networks, Science 286 (1999) 509.
[2] A. Barabási, R. Albert, H. Jeong, Mean-field theory for scale-free random networks, Physica A 272 (1999) 173.
[3] R. Albert, A. Barabási, Statistical mechanics of complex networks, Rev. Modern Phys. 74 (2002) 47.
[4] A. Barabási, Z. Oltvai, Network biology: understanding the cell's functional organization, Nat. Rev. Genet. 5 (2004) 101.
[5] N. Przulj, D. Corneil, I. Jurisica, Modeling interactome: scale-free or geometric? Bioinformatics 20 (2004) 3508.
[6] R. Khanin, E. Wit, How scale-free are biological networks, J. Comput. Biol. 13 (2006) 810.
[7] E. Keller, Revisiting scale-free networks, BioEssays 27 (2005) 1060.
[8] D. Watte, S. Strogatz, Collective dynamics of 'small-world' networks, Nature 393 (1998) 440.
[9] J. Ash, D. Newth, Optimizing complex networks for resilience against cascading failure, Physica A 380 (2007) 673.
[10] V. Eguíluz, K. Klemm, Epidemic threshold in structured scale-free networks, Phys. Rev. Lett. 89 (2002) 108701.
[11] M. Newman, The structure and function of complex networks, SIAM Rev. 45 (2003) 167.
[12] S.N. Dorogovtsev, J.F.F. Mendes, A.N. Samukhin, Structure of growing networks with preferential linking, Phys. Rev. Lett. 85 (2000) 4633.
[13] S.N. Dorogovtsev, J.F.F. Mendes, Evolution of networks, Adv. Phys. 51 (2002) 1079.
[14] P. Krapivsky, S. Redner, F. Leyvraz, Connectivity of growing random networks, Phys. Rev. Lett. 85 (2000) 4629.
[15] J. Dunne, R. Williams, N. Martinez, Food-web structure and network theory: The role of connectance and size, Proc. Natl. Acad. Sci. USA 99 (2002) 12917.
[16] R. Pastor-Satorras, A. Vespignani, Epidemic dynamics in finite size scale-free networks, Phys. Rev. E 65 (2002) 035108.
[17] H. Fraser, Modularity and evolutionary constraint on proteins, Nat. Genet. 37 (2005) 351.
[18] L. Li, Y. Huang, X. Xia, Z. Sun, Preferential duplication in the sparse part of yeast protein interaction network, Mol. Biol. Evol. 23 (2006) 2467.
[19] R. Ferreira, et al., Preferential duplication of intermodular hub genes: An evolutionary signature in eukaryotes genome networks, PLoS One 8 (2013) e56579.
[20] B. Fotouhi, M.G. Rabbat, Network growth with arbitrary initial conditions: Degree dynamics for uniform and preferential attachment, Phys. Rev. E 88 (2013) 062801.
[21] A. Fronczak, P. Fronczak, J. Hoyst, Mean-field theory for clustering coefficients in Barabási-Albert networks, Phys. Rev. E 68 (2003) 046126.
[22] A. Vazquez, et al., The topological relationship between the large-scale attributes and local interaction patterns of complex networks, Proc. Natl. Acad. Sci. 52 (2004) 17940.
[23] G. Bianconi, A. Capocci, Number of loops of size h in growing scale-free networks, Phys. Rev. Lett. 90 (2003) 078701.