Physica A 509 (2018) 588–598

Contents lists available at ScienceDirect

Physica A

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

Master equation for the degree distribution of a Duplication and Divergence network

Vítor Sudbrack, Leonardo G. Brunnet, Rita M.C. de Almeida, Ricardo M. Ferreira, Daniel Gamermann*

Instituto de Física - Universidade Federal do Rio Grande do Sul (UFRGS), Av. Bento Gonçalves, 9500, Brazil

HIGHLIGHTS

- Modeling of network evolution dynamics through Markov chains.
- Determination of the model's parameter range for which the network reaches a steady state.
- Evaluation of the asymptotic network degree distribution.
- Original and copied nodes mutation rates are treated independently.

ARTICLE INFO

Article history:
Received 7 December 2017
Received in revised form 10 April 2018
Available online xxxx

Keywords:
Duplication and divergence
Markov process
Network evolution

ABSTRACT

Network growth as described by the Duplication–Divergence model proposes a simple general idea for the evolution dynamics of natural networks. In particular it is an alternative to the well known Barabási–Albert model when applied to protein–protein interaction networks. In this work we derive a master equation for the node degree distribution of networks growing via Duplication and Divergence and we obtain an expression for the total number of links and for the degree distribution as a function of the number of nodes. Using algebra tools we investigate the degree distribution asymptotic behavior. Analytic results show that the network nodes average degree converges if the total mutation rate is greater than 0.5 and diverges otherwise. Treating original and duplicated node mutation rates as independent parameters has no effect on this result. However, difference in these parameters results in a slower rate of convergence and in different degree distributions. The more different these parameters are, the denser the tail of the distribution. We compare the solutions obtained with simulated networks. These results are in good agreement with the expected values from the derived expressions. The method developed is a robust tool to investigate other models for network growing dynamics.

1. Introduction

Physicists are required to build models that extract the essence of observable phenomena seen in nature in order to understand and describe them. As physics endeavors in studying complex phenomena in distinct fields such as biology or social sciences, a commonly applied paradigm is the use of graph theory. In this approach the system under study is described as a network consistent of a set of nodes and a set of links among them.

Examples of systems studied within this approach are social networks [1], author citations [2], flights connections [3], metabolic models [4,5], protein–protein interactions [6,7], electrical grids [8,9], among others.

* Corresponding author.
E-mail address: danielg@if.ufrgs.br (D. Gamermann).

https://doi.org/10.1016/j.physa.2018.06.066
0378-4371/

In the present work we are interested in *Network dynamics*, i.e. the study of the dynamics behind a growing graph. The motivation lies in the dynamics of biological networks, given that many such systems find a natural description within those models. Understanding the evolutionary selection rules resulting in networks with similar topological characteristics as the real observed ones, may give insights about the underlying biological processes (natural selection) behind these structures. For example, the protein association networks can help understand the evolution of species' genomes [10–13].

In the present work, the network dynamics is described as a Markovian process. Within this approach, the network state at a given time depends only on its configuration on the previous moment. Given a set of rules that describe how the network changes in each time step, we construct the corresponding master equation representing the evolution of the system's configurations. A similar methodology was used in the work by Ferreira et al. [14]. With this approach, they presented analytical results and simulations of networks growing according to the Barábasi–Albert rule. Here we focus on a complementary approach to model protein–protein network dynamics [15] and explore the evolution of an adapted version of the Duplication Divergence model [16]. This model is usually applied to study the evolution of proteins related networks [17–20]. The importance in describing the average behavior of stochastic processes in this manner is to know the network behavior for different values of the parameters without the need of long, time consuming, numerical simulations to obtain statistically relevant information.

This article is organized as follows. In the next section we focus in explaining the Duplication and Divergence model. Following the model explanation, we derive an expression for the total number of links as a function of the number of nodes, which gives us a straight forward way to obtain the mean degree of the graph. Then we study the graph growth as a Markovian chain, in which the next degree distribution of the network is a function of the current degree distribution, pondered by the probabilities of all possible occurrences in each time step. Finally, we study the asymptotic limit of the degree distribution.

2. The model

Given an initial small network (three nodes connected to each other forming a triangle[1]) we study the Markov process where, in each time step a node of the network is randomly chosen to be copied i.e. a new node is created with exactly the same neighbors as the chosen one. In what follows we refer to the copied node as original and its copy as duplicated. After duplication, original and duplicated nodes may diverge, meaning that each link of the original node is lost with probability m_o and each link in the duplicated node is lost with probability m_d. Also, a link between original and the duplicated nodes is always added.

This model is an adaptation of the Duplication and Divergence model, originally developed by Vázquez et al. [16,21]. In the model proposed by Vázquez a new node is also added by copying an existing node and all its links. New node and ancestor are linked with a probability p. Also, either the link between new node and a third neighbor or the link between ancestor and this neighbor is lost with probability q [21]. Our model treats independently the loss of a link by original and duplicated nodes and sets $p = 1$.

In Vázquez's model the network growth is based on local rules, that is, rules that require only information on one node instead of rules that require information over all the network. This model of network dynamics applied to protein–protein interaction networks allows all proteins to evolve from a common ancestor through gene copies (represented by duplications) and mutations (divergence). Therefore, it would mimic the entire history of a genome evolution [21].

3. Mean degree and total number of links

First, let us evaluate the behavior of the network average node degree $\bar{k}$ as a function of the number of nodes in the network, t. Given that a node with degree k is chosen to be duplicated, the number of links in the next step changes. The mean change in the number of links is given by[2]:

$$\begin{aligned} L_{t+1} &= L_t + \overline{(k - m_o k - m_d k + 1)} \\ &= L_t + (1 - m_o - m_d)\bar{k} + 1 \end{aligned} \tag{1}$$

Terms in the right-hand side of Eq. (1) represent, from left to right: existing links, links added due to the duplication of a k-degree node, mean number of lost links of the original node, mean number of lost links for the duplicated node and creation of the original–duplicated link.

Let us define the total mutation parameter M as the sum of the independent mutation parameters m_o and m_d: $M = m_o + m_d$. Now, given the relation between node's degree distribution and number of links in a network, known as the

1 As will be discussed latter on, the asymptotic behavior of the graph's properties is independent of this choice of seed (initial) graph.

2 The bar over a term in the equation indicates that the average of the process will be considered.

handshake lemma [22],

$$2L = \sum_{i=1}^{N} k_i = N\bar{k}, \tag{2}$$

and, considering the time as the number of nodes (since we add a node in each time step), the following map for the network average degree is written

$$\frac{(t+1)\bar{k}_{t+1}}{2} = \frac{t\bar{k}_t}{2} + (1-M)\bar{k}_t + 1\ . \tag{3}$$

This equation can be rearranged to explicit the $\bar{k}_{t+1}$ term:

$$\bar{k}_{t+1} = \frac{(t+2-2M)\bar{k}_t + 2}{t+1}$$

For long times (big values of t), this can be approximated as a continuous differential equation:

$$\frac{\partial \bar{k}}{\partial t} = \frac{1-2M}{t+1}\bar{k} + \frac{2}{t+1}$$

For $0 \leq M \leq 2$, the solution of this ODE is:

$$\bar{k}(t) = \begin{cases} c(t+1)^{1-2M} + \dfrac{2}{2M-1} & \text{for } M \neq 0.5; \\ 2\log(t+1) + c & \text{for } M = 0.5\ , \end{cases} \tag{4}$$

where the constant c is related to network initial conditions.

For instance, consider the case of a process without divergence ($M = 0$). Starting with a triangle ($\bar{k}(3) = 2$), the duplication will make every node connected with every other, i.e., a complete graph. In this particular case Eq. (4) results in

$$\bar{k}(t) = (t+1) - 2 = t - 1.$$

In the limit $t \to \infty$, the possible asymptotic behaviors for Eq. (4) are:

For M > 0.5: The mean degree converges to $\frac{2}{2M-1}$.

For M < 0.5: The mean degree diverges as t^{1-2M}.

For M = 0.5: The mean degree diverges logarithmically.

We represent these possible situations in Fig. 1, where each point refers to a network evolved through the model dynamics with a different value for the total mutation parameter M. In the vertical axis one has the correspondent mean degree calculated by Eq. (4) (in the limit $t \to \infty$) and in the horizontal axis it is presented the results for the mean degree evaluated from simulated networks growing according to the model rules. The correspondence of the points colors with the parameter M can be read in the color scale. The figure clearly shows that, for high values of the total mutation M (values equal or bigger than 0.8), a network with 20 thousand nodes has already reached its stationary behavior (the term t^{1-2M} can be neglected). The closer the total mutation gets to $M = 0.5$, the slower one observes the convergence to the stationary limit. For values of total mutation under 0.5 it is expected that the mean degree diverges and so are the points in this figure departing from the main diagonal.

Using again the handshake lemma in Eq. (4), we obtain the total number of links as a function of time

$$L(t) = \begin{cases} ct(t+1)^{(1-2M)} + \dfrac{t}{2M-1} & \text{for } M \neq 0.5; \\ t\log(t+1) + ct & \text{for } M = 0.5 \end{cases} \tag{5}$$

In Fig. 2 we show results comparing the mean value for the number of links in 5 thousand networks as a function of time with the expected value found through from Eq. (5).

4. Degree distribution for finite times

Considering the model dynamics, it is possible to establish relations among the number of nodes with degree k at time t, $N(k, t)$, and at time $t + 1$:

$$N(k, t+1) = \qquad\qquad N(k, t)$$

$$- \qquad\qquad \frac{N(k, t)}{t}$$

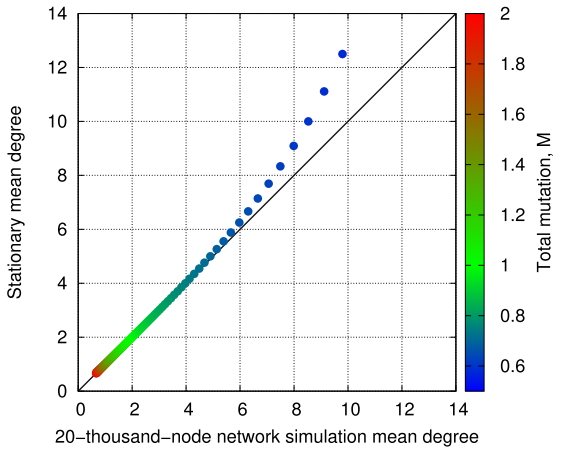

Fig. 1. [color on-line] Scatter plot of the average degree for 20 thousand node networks obtained numerically compared with the result given by Eq. (4) as a function of the total mutation $M = m_o + m_d$. Each point corresponds to a single network evolving according to the model dynamics with its total mutation rate M given by the color scale. Note that the greater the total mutation M, the faster the stationary limit is reached.

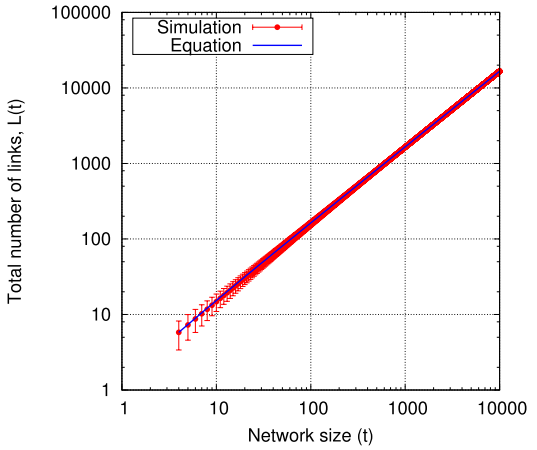

Fig. 2. [color on-line] A comparative between the results of the average (over 5 thousand networks) total number of links as a function of the network size and the result given by Eq. (5) for $M = m_o + m_d = 0.8$.

$$
\begin{aligned}
&+ \sum_{i=k-1}^{t-1} P_o(i \to k)\frac{N(i,t)}{t} \\
&+ \sum_{i=k-1}^{t-1} P_d(i \to k)\frac{N(i,t)}{t} \\
&+ \frac{(k-1)N(k-1,t)}{t}(1-m_o)(1-m_d) \\
&+ \frac{(k+1)N(k+1,t)}{t}m_o m_d \\
&- \frac{kN(k,t)}{t}[(1-m_o)(1-m_d)+m_o m_d],
\end{aligned} \tag{6}
$$

the terms in this equation represent, from top to bottom: the number of existing nodes of degree k; the probability of a k-degree node to be chosen to duplicate, leaving this degree; the first sum is the total probability that a node of degree different from k is duplicated and after the divergence process becomes a k-degree node; the second sum is the same total probability for the duplicated node. Finally, the last three terms are the probabilities that the neighbors of the node chosen to be duplicated arrive, from a different degree (either $k+1$ or $k-1$) to the degree k and the probability that a k-degree neighbor goes to $k+1$ or $k-1$.

Since each link, during the divergence process, is lost independently with probability m_o or m_d, the probabilities inside the sums are binomial distributions representing the probabilities that either the original or duplicated node goes from degree i to degree $k-1$ after the divergence, and finally received the original–duplicated link:

$$P_o(i \to k) = C^i_{k-1}(1-m_o)^{k-1}(m_o)^{i-k+1} \tag{7a}$$

$$P_d(i \to k) = C^i_{k-1}(1-m_d)^{k-1}(m_d)^{i-k+1} \tag{7b}$$

It is important to note the domain of the above functions (7) and (7b), $i \geq k-1$, that is, the node after passing through duplication and divergence can increase its degree by one unit (which means keeping all its links and adding the copied-duplicated link), keep the same degree (losing one neighbor and adding the copied-duplicated link) or else it will have its degree decreased due to the loss of more than one neighbor.

One can verify the addition of a single node in each time step by summing the master equation for $N(k,t)$, Eq. (6), over all possible degrees

$$\sum_{k=0}^{\infty}(N(k,t+1)-N(k,t)) = 1. \tag{8}$$

As shown in Fig. 3, the numerical solution of the map in Eq. (6) produces results in excellent agreement with the mean number of nodes with degree k at the time t evaluated from thousands of simulated networks growing through the stochastic process.

As an analytic example, using $m_o = 0$ e $m_d = 0$, no links are lost by neither node, and Eq. (6) becomes,

$$\begin{aligned} N(k,t+1) = & \quad N(k,t) - \frac{N(k,t)}{t} + 2\frac{N(k-1,t)}{t} \\ & + \frac{(k-1)N(k-1,t)}{t} - k\frac{N(k,t)}{t} \end{aligned} \tag{9}$$

Eq. (9) can be written as

$$N(k,t+1) - N(k,t) = \frac{k+1}{t}\Big(N(k-1,t) - N(k,t)\Big) \tag{10}$$

Note that, in the case where the initial network is totally connected, that is, $k = t-1$, there is only the flux of all nodes having its degree increased by one in a totally connected network, $N(k,t) = t\delta_{k-1,t}$.

5. Asymptotic degree distribution

Eq. (6) can be conveniently written in its matrix form, defining the column vector $\vec{N}(t)$ whose components are $N_k(t) = N(k,t)$:

$$\frac{d}{dt}\vec{N} = \frac{1}{t}A\vec{N}, \tag{11}$$

where the matrix A shows the following elements, constant with respect to time:

$$\begin{aligned} a_{i,k} = & \quad -(k[(1-m_o)(1-m_d) + m_o m_d] + 1)\delta_{k,i} \\ & + (k-1)(1-m_o)(1-m_d)\delta_{k-1,i} \\ & + (k+1)m_o m_d \delta_{k+1,i} \\ & + P_o(i \to k) + P_d(i \to k) \end{aligned} \tag{12}$$

The coefficients of Eq. (12) explicit the couplings in Eq. (11). It is possible to solve Eq. (11) through a matrix decomposition of the matrix $A_{n\times n}$ in its eigenvalues and eigenvectors: Writing $A = X^{-1}DX$, putting this decomposition in Eq. (11) and multiplying from the left by the matrix X, one decouples the equations and is able to solve it for each component of the vector $\vec{N}$ in the space where the matrix A is diagonal. Returning to the original space one has:

$$\vec{N} = \sum_{i=1}^{n} c_i \vec{X}_i t^{\lambda_i} \tag{13}$$

Where the λ_i are the eigenvalues of matrix A, $\vec{X}_i$ the respective eigenvectors (the columns of the matrix X, properly normalized, are the vectors $\vec{X}_i$) and the constants c_i depend on the initial conditions (seed graph).

Normalizing the vector $\vec{N}(t)$ one works with the fraction of nodes with degree k:

$$\vec{P}(t) = \frac{1}{t}\vec{N(t)}, \tag{14}$$

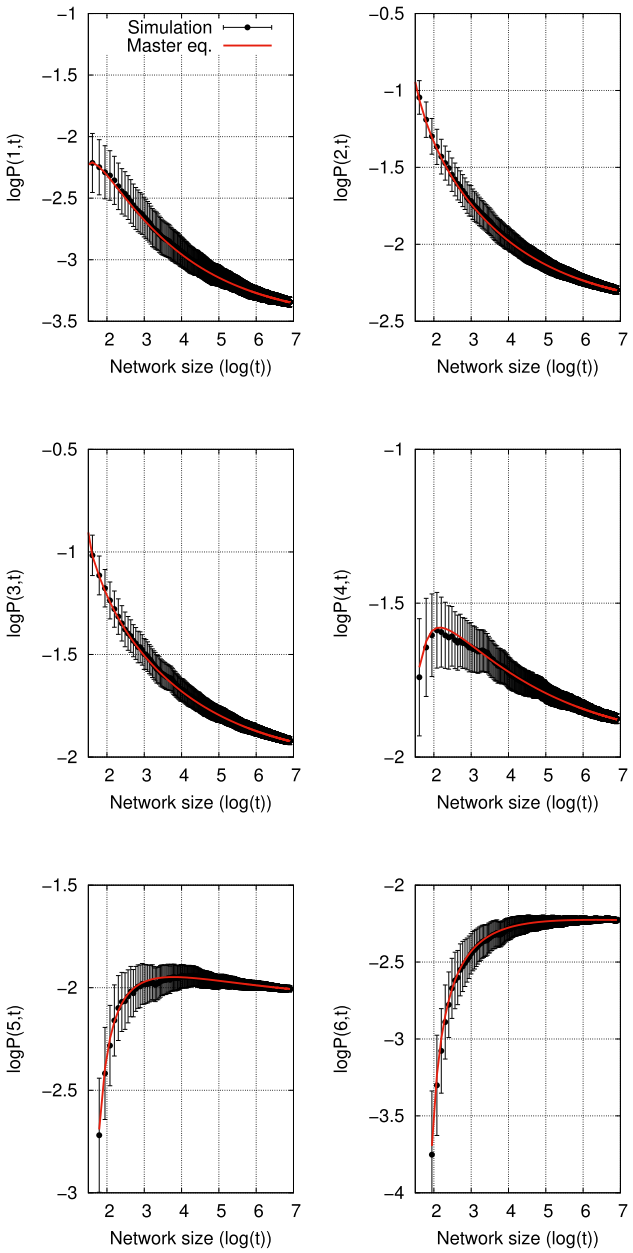

Fig. 3. [color on-line] Comparison of the results for the average fraction ($P(k,t) = N(k,t)/t$) of nodes of degree k in 5-thousand networks simulated from 3 nodes until 10^3 with the results predicted by the numerical integration of Eq. (6) for $m_o = 0.25$ and $m_d = 0.40$ (resulting in $M = 0.65$).

note that $P_k(t)$ is the fraction of nodes of degree k at time t. The solution (13) can be written in terms of the new vector as follows:

$$\vec{P} = \sum_{i=1}^{t} c_i \vec{X}_i t^{\lambda_i - 1} \tag{15}$$

In the stationary state, condition (8) implies that $\lambda = 1$ is an eigenvalue of matrix A (whose left-eigenvector is $(1, 1, 1, 1..., 1)$)[23], and, to conserve probability, this must be the greatest eigenvalue, and the normalized right-eigenvector which corresponds to the unitary eigenvalue (called main eigenvector) will be the asymptotic solution of the degree distribution of a network whose growth is governed by the Duplication and Divergence model with mutation rates m_o and m_d. Note that this asymptotic solution does not depend on the initial (seed) graph used, only the c_i parameters, which may play a role only in the transient state, depend on the initial graph.

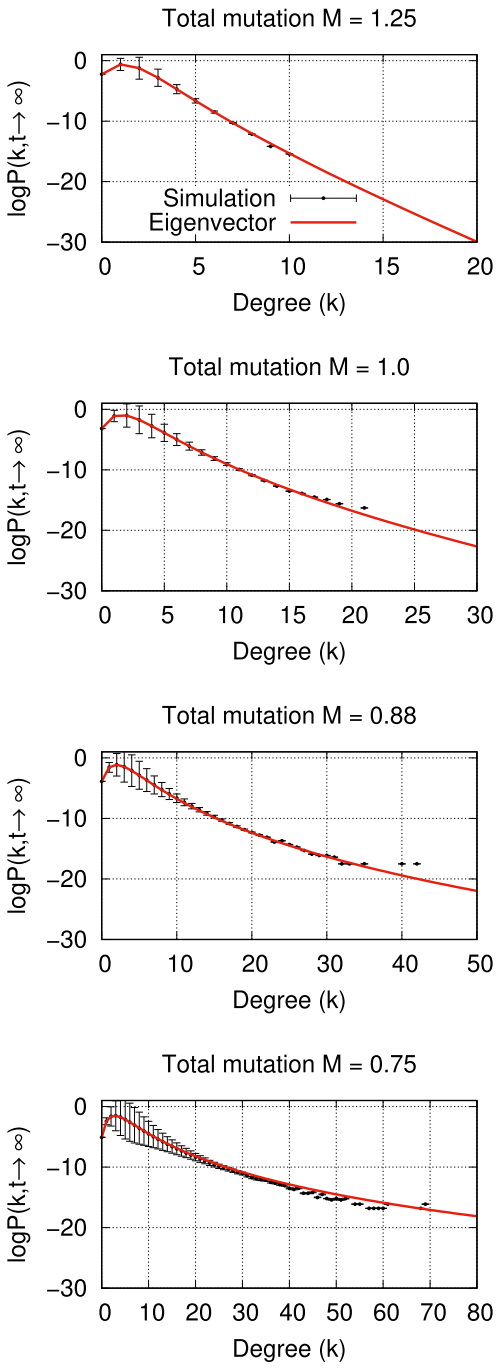

Fig. 4. [color on-line] Results of the degree distribution for $t \to \infty$ obtained through the eigenvector decomposition compared to results obtained from simulated networks up to 20 thousand nodes for values of total mutation rates 1.25, 1.00, 0.88 and 0.75 ($m_o = m_d = M/2$). It is possible to observe that 20-thousand-node networks are large enough to generate the asymptotic degree distribution.

Fig. 4 compares the main eigenvector numerically obtained with the simulation of networks up to 80 thousand nodes.

Note that the matrix $A_{n\times n}$ has to be truncated for taking the limit $t \to \infty$ because its actual size is $t \times t$ in Eq. (11), which is the greatest degree possible for $(t+1)$-node networks. For $M > 0.5$, this truncation still results in a matrix with an eigenvalue whose value is 1, which is related, as mentioned, to the stationary distribution. When $M < 0.5$, any truncation results in a matrix A which has all eigenvalues less than 1. So when the mean degree diverges, there is obviously no stationary degree distribution.

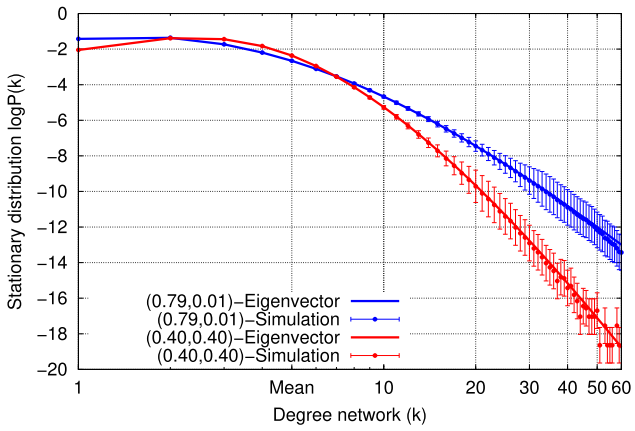

Fig. 5. [color on-line] Stationary degree distribution, $P(k, t \to \infty)$, for networks with the same total mutation rate $M = 0.80$, but different mutation rates m_o and m_d.

6. Parameters m_o and m_d are independent

As seen in Eq. (4), the mean degree (stationary or time-dependent) is a function that depends only on the total mutation rate, $M = m_o + m_d$. However, Eq. (6) cannot be expressed considering only the sum of the individual mutations, each must be kept as an independent parameter. Therefore, different partitions of the same total mutation lead to different distributions with the same average. The parameters are interchangeable in both equations which reflects the fact that, after the duplication, the original and copied nodes are indistinguishable.

Finding the stationary distribution numerically for $M = 0.80$ in two possible scenarios, shown in Fig. 5, gives us insights about the network behavior in both cases. First, we should note that the distribution as a whole does not appear to follow neither an exponential nor a power-law function. Only the distribution's far right-tail should be adjusted to any of these monotonically decreasing distributions, but only if one considers degrees bigger than around 10, region where the distributions seem to follow a straight line in Fig. 5, which has its vertical axis in logarithm scale. Though it is a common claim that natural networks have scale-free (power-law) degree distributions, many of these claim fail to survive a thorough statistical analysis [24] and many authors dispute this general claim [25–27].

From Fig. 5, one observes that, when the mutation parameters are equal ($m_o = m_d = 0.4$), the distribution is denser near the distribution's average and when the mutation parameters are very dissimilar ($m_o = 0.79$; $m_d = 0.01$), the resulting distribution has a denser tail, indicating more nodes with lower and higher degrees than the average value. This is a very reasonable result considering the divergence process. In Fig. 6, we show this difference. It is possible to see the time evolution of the degree distribution $P(k, t)$ for degrees in these three cases: lower, near and greater than the stationary average degree (that is $\bar{k} = 10/3$, for $M = 0.80$) obtained numerically by integrating Eq. (6) and simulations.

One can also observe the real part of the eigenvalues of matrix A in both cases, shown in Fig. 7. The imaginary part, which is not shown in this figure, is symmetric, which only reflects the fact that the elements of the matrix are real, and therefore the complex eigenvalues come in pairs as complex conjugates. From the real part of the eigenvalues one can infer the convergence rate. Therefore, one can conclude that for different values of m_o and m_d the convergence will take longer, since more eigenvalues are greater that in the other situation: in the symmetrical case, there are smaller eigenvalues than in the asymmetrical case.

The rate of convergence to the stationary solution can also be inferred by the second highest eigenvalue of matrix A. In Fig. 8 dots represent the real part of the highest eigenvalues of A. It is possible to see that the second highest eigenvalue of matrix A for total mutation rate higher than 0.55 will be given by $2 - 2M$, and therefore $\bar{P}(t) \propto t^{1-2M}$, which is the same exponent of the convergence of the mean degree given by Eq. (4).

7. Conclusions

In the present work, we derived the master equation for the degree distribution of networks evolving through the duplication and divergence model considering the mutation rates of original and duplicated nodes as independent parameters. Numerical integration of the resulting maps agree well with the average values obtained from simulated networks.

The maps obtained for the network nodes average degree and number of links can be approximated as ordinary differential equations and solved analytically. The asymptotic solution for these ODEs, agrees well with simulated data for values of total mutation greater than 0.5, limit for which the network converges to a stationary degree distribution.

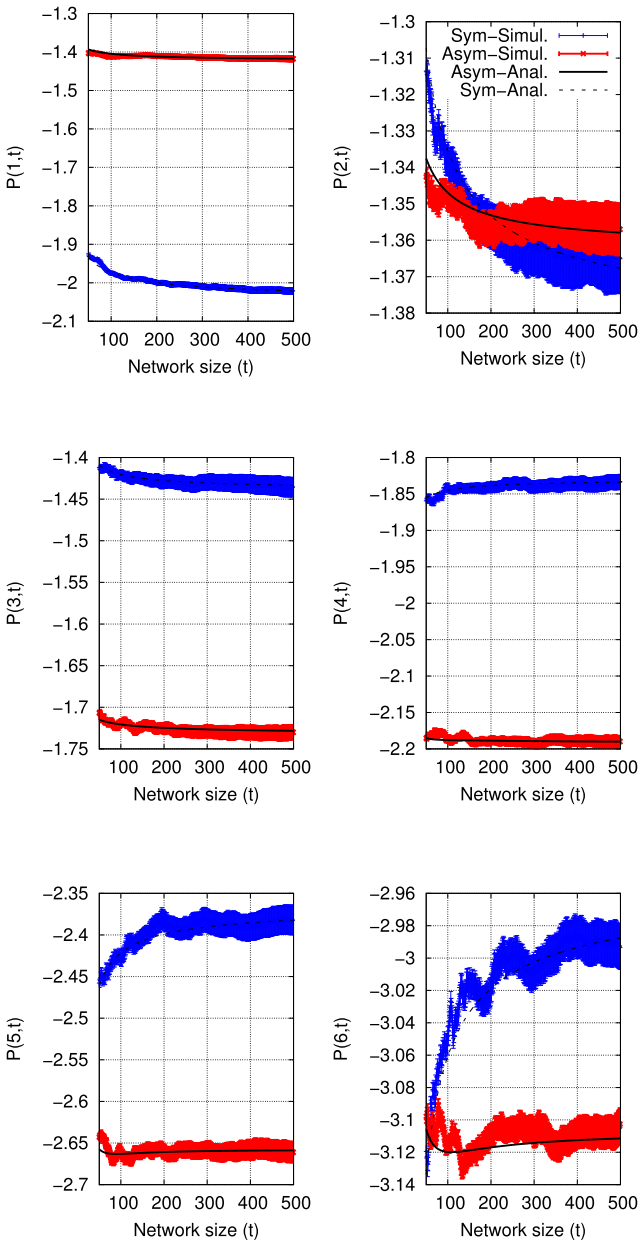

Fig. 6. [color on-line] Degree distribution $P(k, t)$ as a function of time, for networks with the same total mutation $M = 0.8$, but different mutation rates m_o and m_d. The top graphs are for low degree nodes, the middle ones are near the stationary average and the bottom ones are high degree nodes.

Though the average node degree of a network evolving through this process only depends on the total mutation rate parameter, its degree distribution and rate of convergence will depend on the mutation parameter difference between original and duplicated nodes. The behavior of this distribution can be studied from the spectrum of the matrix A, which contains the transition probabilities between nodes of different degrees in each time step of the Markovian process. The more similar the two parameters are, the faster the distribution converges because, in this case, one has less eigenvalues close to 1, and the less dense the distribution tail will be.

Using Markovian processes to describe network evolution allows one to obtain iterative maps for the graphs evolution in a wide variety of models. In the asymptotic limit, maps can be approximated as ODEs, and therefore solved such that one is able to study qualitatively the dynamics dependence on the model parameters without the need to run computationally intensive numerical simulations. Finally, different network growing models may be treated using the present method. In particular, a work considering a hybrid Barabasi-Duplication/Divergence model is under development.

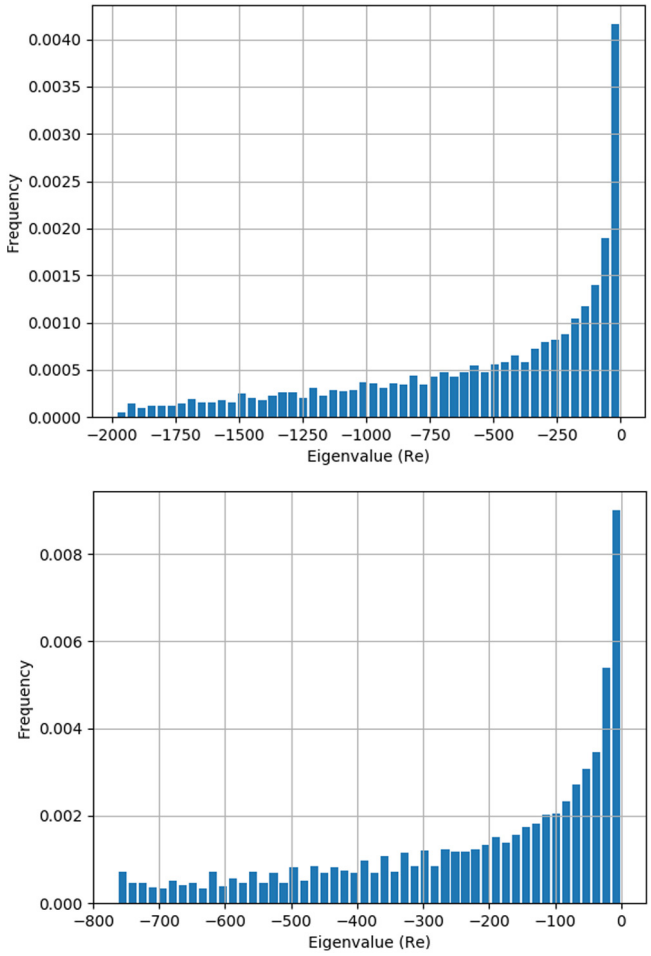

Fig. 7. Spectrum of the 2000 × 2000 matrix A for the dynamics with the same total mutation, and $m_o = m_d = 0.4$ on top and $m_o = 0.79$, $m_d = 0.01$ on bottom. We can conclude that the eigenvalues in the equal parameter case are lower than the asymmetrical case.

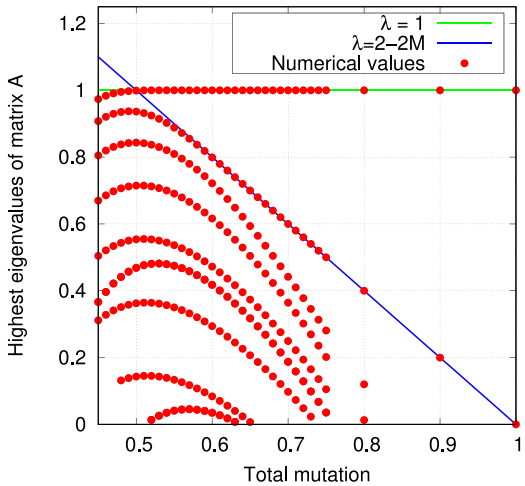

Fig. 8. [color on-line] Highest eigenvalues of a size 2000 × 2000 matrix A calculated numerically with for $m_o = m_d = M/2$. The points in each vertical line correspond to the firsts eigenvalues for the matrix with a total mutation rate given by the horizontal axis. The blue (oblique) line corresponds to the exponent of convergence of $\bar{k}$.

Acknowledgments

We thank the contributions of collaborators from MTC-IF-UFRGS. We also thank CNPq, Brazil for the financial support.

References

[1] S.P. Borgatti, A. Mehra, D.J. Brass, G. Labianca, Network analysis in the social sciences, Science 323 (5916) (2009) 892–895.
[2] D.J. de Solla Price, Networks of scientific papers, Science 149 (3683) (1965) 510–515.
[3] Chi Li-Ping, Wang Ru, Su Hang, Xu Xin-Ping, Zhao Jin-Song, Li Wei, Cai Xu, Structural properties of US flight network, Chin. Phys. Lett. 20 (8) (2003) 1393.
[4] J. Forster, Genome-scale reconstruction of the saccharomyces cerevisiae metabolic network, Genome Res. 13 (2) (2003) 244–253.
[5] J. Triana, A. Montagud, M. Siurana, D. Fuente, A. Urchueguia, D. Gamermann, J. Torres, J. Tena, P.F. de Cordoba, J.F. Urchueguia, Generation and evaluation of a genome-scale metabolic network model of synechococcus elongatus PCC7942, Metabolites 4 (3) (2014) 680–698.
[6] N. Przulj, D.A. Wigle, I. Jurisica, Functional topology in a network of protein interactions, Bioinformatics 20 (3) (2004) 340–348.
[7] Fereydoun Hormozdiari, Petra Berenbrink, Nataša Przulj, S. Cenk Sahinalp, Not all scale-free networks are born equal: the role of the seed graph in ppi Network Evolution, PLoS Computational Biology 3 (7) (2007) e118.
[8] Réka Albert, István Albert, Gary L. Nakarado, Structural vulnerability of the north american power grid, Phys. Rev. E 69 (2004) 025103.
[9] Giuliano Andrea Pagani, Marco Aiello, The Power Grid as a complex network: A survey, Physica A 392 (11) (2013) 2688–2700.
[10] Susumu Ohno, Evolution by Gene Duplication, Springer, 2013.
[11] Wen-Hsiung Li, Molecular Evolution, Sinauer Associates, 1997.
[12] I. Ispolatov, P.L. Krapivsky, A. Yuryev, Duplication-Divergence model of protein interaction network, Phys. Rev. E 71 (2005) 061911.
[13] John S. Taylor, Jeroen Raes, Duplication and divergence: the evolution of new genes and old ideas, Annu. Rev. Genet. 38 (1) (2004) 615–643, PMID: 15568988.
[14] Ricardo M. Ferreira, Rita M.C. de Almeida, Leonardo G. Brunnet, Analytic solutions for links and triangles distributions in finite barabási–albert networks, Physica A 466 (2017) 105–110.
[15] Ricardo M. Ferreira, José Luiz Rybarczyk-Filho, Rodrigo J.S. Dalmolin, Mauro A.A. Castro, José C.F. Moreira, Leonardo G. Brunnet, Rita M.C. de Almeida, Preferential duplication of intermodular hub genes: an evolutionary signature in eukaryotes genome networks, PLOS ONE 8 (2) (2013) 1–11.
[16] Alexei Vázquez, Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations, Phys. Rev. E 67 (2003) 056104.
[17] I. Ispolatov, P.L. Krapivsky, A. Yuryev, Duplication-Divergence model of protein interaction network, Phys. Rev. E 71 (6) (2005).
[18] Pei Wang, Jinhu Lu, Xinghuo Yu, Zengrong Liu, Duplication and divergence effect on network motifs in undirected bio-molecular networks, IEEE Transactions on Biomedical Circuits and Systems 9 (3) (2015) 312–320.
[19] Ibrokhim Y. Abdurakhmonov, Zabardast T. Buriev, Carla Jo Logan-Young, Abdusattor Abdukarimov, Alan E. Pepper, Duplication, divergence and persistence in the Phytochrome photoreceptor gene family of cottons (Gossypium spp.), BMC Plant Biology 10 (1) (2010) 119.
[20] Shuiming Cai, Zengrong Liu, H.C. Lee, Mean field theory for biology inspired duplication-divergence network model, Chaos 25 (8) (2015) 083106.
[21] Alexei Vázquez, Alessandro Flammini, Amos Maritan, Alessandro Vespignani, Modeling of protein interaction networks, Complexus 1 (1) (2003) 38–44.
[22] Leonhard Euler, Solutio Problematis ad Geometriam Situs Pertinentis, Commentarii Academiae Scientiarum Imperialis Petropolitanae 8 (1736) 128–140.
[23] C.R. MacCluer, The many proofs and applications of Perron's theorem, SIAM Review 42 (3) (2000) 487–498.
[24] Aaron Clauset, Cosma Rohilla Shalizi, M.E.J. Newman, Power-law distributions in empirical data, SIAM Rev. 51 (4) (2009) 661–703.
[25] Raya Khanin, Ernst Wit, How scale-free are biological networks, J. Comput. Biol. 13 (3) (2006) 810–818.
[26] Gipsi Lima-Mendez, Jacques van Helden, The powerful law of the power law and other myths in network biology, Mol. BioSyst. 5 (2009) 1482–1493.
[27] D. Gamermann, J. Triana, R. Jaime, A comprehensive statistical study of metabolic and protein-protein interaction network properties, 2017, arXiv: 1712.07683.