IJCSS International Journal of Complex Systems in Science

Int. J. Complex Systems in Science
vol. **1(2)** (2011), pp. 209–215

Node Correlations in Heterogeneous Evolving Networks

Antonio Santiago[1] and Rosa M. Benito[1,†]

[1] *Complex Systems Group, Department of Physics, ETSIA, Universidad Politécnica de Madrid (28040 Madrid, Spain)*

Abstract. In this communication we present an analytical study of the correlations between adjacent nodes in heterogeneous evolving networks. We consider a class of network models extending the Barabási-Albert model where node states induce affinities in their interactions and derive a mean-field solution of the second-order density function of degrees and states, which evidences the existence of nontrivial correlations between the degrees of adjacent nodes.

Keywords: Heterogeneous networks, preferential attachment, node correlations.
MSC 2000: 89.75.Fb, 89.75.Hc, 89.75.-k.

† **Corresponding author:** rosamaria.benito@upm.es
Received: October 19, 2011
Published: October 24, 2011

1. Introduction

A heterogeneous network is characterized by the existence of intrinsic properties of the network elements that may induce affinities in their interactions. The authors have introduced a class of heterogeneous models that extend the Barabási-Albert preferential attachment (PA) model [1, 2] to heterogeneous networks, contains previous models as particular cases and provides a framework for the systematic analysis of the influence of heterogeneity in evolving networks [3, 4]. Node correlations are an important metric that in addition to the degree distribution provides a more thorough characterization of the structure of complex networks. In this communication we present analytical results obtained by a mean-field approach concerning the correlations that develop between adjacent nodes in heterogeneous evolving networks.

2. A general class of heterogeneous network models

The PA model has been extended to heterogeneous networks by imposing an algebraic structure on the states of the nodes, while preserving the original mechanisms of growth and preferential attachment. A heterogeneous PA model is defined by three elements:
(1) R is an arbitrary space. The elements $x \in R$ are referred as node **states**.
(2) ρ is a nonnegative real function with unit measure over R referred as **node state distribution**.
(3) σ is a nonnegative real function over R^2 referred as **affinity** of the interactions.
The class of heterogeneous PA models is thus defined as the set of all triples (R, ρ, σ) whose elements comply with the previous conditions. This formalism prescribes the evolution of a network according to the following rules:
(i) The nodes v_i are characterized by their state $x_i \in R$. The states describe intrinsic properties deemed constant in the timescale of evolution. The links e_i are considered as purely topological elements.
(ii) The growth process starts with a seed composed by N_0 nodes (with arbitrary states $x_i \in R$) and L_0 links with arbitrary topology.
(iii) A new node v_a (with a constant number m of links attached) is added to the network at each iteration. The newly added node is randomly assigned a state x_a following the distribution $\rho(x)$.
(iv) The m links attached to v_a are randomly connected to the network nodes following a distribution $\{\Pi(v_i)\}$ given by an extended *attachment rule*,

$$\Pi(v_i \mid v_a) = \frac{\pi(v_i \mid v_a)}{\sum_j \pi(v_j \mid v_a)}, \ \pi(v_i \mid v_a) = k_i \sigma(x_i, x_a). \tag{1}$$

The visibility π of a node v_i in the attachment rule is given by its connectivity degree k_i times its affinity σ with the newly added node v_a, which is itself a function of the states x_i and x_a. Steps (iii) and (iv) are iterated until a desired number of nodes has been added to the network.

3. Degree distribution in heterogeneous networks

The distribution of connectivity degrees in heterogeneous network models can be analyzed by applying rate equations [5, 6] to the flows of degree densities f over network partitions. For each state x, the form of the equation will be $L_1 - L_2 = R_1 - R_2$, where:
L_1 = density of nodes with degree k at $t = N + 1$;
L_2 = density of nodes with degree k at $t = N$;
R_1 = increase in density due to nodes with degree $k - 1$ that have gained a link at $t = N$;
R_2 = decrease in density due to nodes with degree k that have gained a link at $t = N$.

Denoting by $f(k, x, N)$ the probability density function for a node having degree k and state x at $t = N$, the resulting equation for $k > m$ is

$$\begin{aligned} &(N+1)f(k,x,N+1) - Nf(k,x,N) = \\ &= m\left\langle \frac{\sigma(x,y)}{\psi(y,N)} \right\rangle_y [(k-1)f(k-1,x,N) - kf(k,x,N)], \end{aligned} \tag{2}$$

while for $k = m$ the rate equation is

$$\begin{aligned} &(N+1)f(m,x,N+1) - Nf(m,x,N) = \\ &= \rho(x) - m\left\langle \frac{\sigma(x,y)}{\psi(y,N)} \right\rangle_y m\, f(m,x,N), \end{aligned} \tag{3}$$

where the brackets mean averaging over the random variable y, and $\psi(y, N)$ is defined as the partition factor

$$\psi(y,N) = \sum_{k'} k' \int_R \sigma(x,y) f(k',x,N)\,\mathrm{d}x. \tag{4}$$

In the limit of large network size the degree density for $k < m$ tends to zero therefore the previous equations define all the possible cases in each iteration.

It can be shown by mean-field arguments that these equations yield in the thermodynamic limit $N \to \infty$ a stationary distribution $f(k, x)$ for the degree density function, well approximated for $k \geq m$ by the expression

$$f(k,x) = \frac{2\rho/\hat{w}}{m + 2/\hat{w}} \frac{B(k, 1 + 2/\hat{w})}{B(m, 1 + 2/\hat{w})}, \tag{5}$$

where Legendre's Beta function $B(y,z) = \int_0^1 t^{y-1}(1-t)^{z-1}\mathrm{d}t$ for $y, z > 0$ satisfies the functional relation $\Gamma(a)/\Gamma(a+b) = B(a,b)/\Gamma(b)$ for Euler's Gamma Γ function, and $\hat{\omega}(x) = \omega(x)/\bar{\omega}$ is a mean-field *normalized fitness* factor defined by $\omega(x) = \langle \sigma(x,y) \rangle_y$ and $\bar{\omega} = \langle \omega(x) \rangle_x$.

The previous solution is valid for any model in our class, since it does not make any assumption regarding the geometry of space R, the form of affinity σ or the distribution ρ. Notice that in the homogeneous limit the mean-field solution yields the well-known expression for the degree distribution

$$P(k) = \frac{2m(m+1)}{k(k+1)(k+2)}. \tag{6}$$

The stationary density functions follow an aymptotic behavior as $B(k, 1 + 2/\hat{w}) \sim k^{-(1+2/\hat{w})}$, so that all the degree densities have tails that are power-laws $k^{-\gamma(\hat{w})}$ with exponents $\gamma(\hat{w}) = 1 + 2/\hat{w}$. This points out to the existence of a multiscaling phenomenon in the degree densities of the heterogeneous models in the assessed class.

4. Degree correlations in heterogeneous networks

The correlations between adjacent nodes in heterogeneous network models can be analyzed by applying rate equations jointly to the flows of second-order (link) degree densities f_e as well as first-order (node) degree densities f_v over network partitions. For each pair of states (x, y), the form of the equation will be $L_1 - L_2 = R_1 - R_2 + R_3 - R_4$, where:
L_1 = density of links connecting a pair of nodes with degrees (k,l) and states (x, y) at $t = N + 1$;
L_2 = density of links connecting a pair of nodes with degrees (k,l) and states (x, y) at $t = N$;
R_1 = increase in density due to links connecting a pair of nodes with degrees $(k-1, l)$ and states (x, y), where the first node has gained a link at $t = N$;
R_2 = decrease in density due to links connecting a pair of nodes with degrees (k,l) and states (x, y), where the first node has gained a link at $t = N$;
R_3 = increase in density due to links connecting a pair of nodes with degrees $(k, l-1)$ and states (x, y), where the second node has gained a link at $t = N$;
R_4 = decrease in density due to links connecting a pair of nodes with degrees (k,l) and states (x, y), where the second node has gained a link at $t = N$.

Denoting by $f_e(k, x, l, y, N)$ the probability density function for an undirected link connecting a node with degree k and state x to a node with degree l and state y at $t = N$ the equation for link densities for the case $k > m$, $l > m$ is

$$m(N+1)f_e(k,x,l,y,N+1) - mNf_e(k,x,l,y,N) = \tag{7}$$

$$m^2\left\langle \frac{\sigma(x,z)}{\psi(z,N)} \right\rangle_z [(k-1)f_e(k-1,x,l,y,N) - kf_e(k,x,l,y,N)] +$$
$$+m^2\left\langle \frac{\sigma(y,z)}{\psi(z,N)} \right\rangle_z [(l-1)f_e(k,x,l-1,y,N) - lf_e(k,x,l,y,N)],$$

while for the case $k = m$, $l > m$ the following equation holds

$$m(N+1)f_e(m,x,l,y,N+1) - mNf_e(m,x,l,y,N) = \quad (8)$$
$$m\rho(x)\frac{\sigma(y,x)}{\psi(x,N)}(l-1)f_v(l-1,y,N) - m^2\left\langle \frac{\sigma(x,z)}{\psi(z,N)} \right\rangle_z mf_e(m,x,l,y,N) +$$
$$+m^2\left\langle \frac{\sigma(y,z)}{\psi(z,N)} \right\rangle_z [(l-1)f_e(m,x,l-1,y,N) - lf_e(m,x,l,y,N)].$$

The equation for the case $k > m$, $l = m$ is analogous to the previous case by a symmetry argument, while the equation for the case $k = l = m$ is simply

$$f_e(m,x,m,y,N) = 0. \quad (9)$$

Again in the limit of large network size the degree density for $k < m$ tends to zero therefore the previous equations define all the possible cases in each iteration. The rate equations for node densities $f_v(k,x,N)$ are independent of link densities and are derived as described in Section 3., see Eqs. 2 and 3.

It can be shown by mean-field arguments that these equations yield in the thermodynamic limit $N \to \infty$ a stationary distribution $f_e(k,x,l,y)$ for the link density function, which in the case $k = m$, $l > m$ can be well approximated by

$$f_e(k,x,m,y) = \frac{2\rho(x)\rho(y)\hat{\sigma}(x,y)}{m(2+m\hat{w}(x))B(m,1+2/\hat{w}(x))} \sum_{i=m+1}^{k} (\hat{w}(x))^{k-i} \quad (10)$$
$$\cdot \left(\prod_{j=i}^{k} \frac{(j-1)}{2+j\hat{w}(x)+m\hat{w}(y)} \right) B(i-1,1+2/\hat{w}(x)).$$

The solution for the link density function in the case $k > m$, $l = m$ is symmetrical to the previous expression with respect to k and l, while in the general case $k > m$, $l > m$ the solution becomes

$$f_e(k,x,l,y) = \frac{\Gamma(1+(k-1)\hat{w}(x))\Gamma(1+(l-1)\hat{w}(y))}{\Gamma(3+k\hat{w}(x)+l\hat{w}(y))} \cdot \quad (11)$$
$$\left(\sum_{i=m+1}^{k} \frac{\Gamma(3+(m+k+1-i)\hat{w}(x)+m\hat{w}(y))}{\Gamma(1+(m+k-i)\hat{w}(x))\Gamma(1+(m-1)\hat{w}(y))} \right.$$

$$
\cdot \frac{\Gamma(i+l-2m-1)}{\Gamma(i-m)\Gamma(l-m)} f_e(m+1+k-i,x,m,y) + \\
+ \sum_{j=m+1}^{l} \frac{\Gamma(3+m\hat{w}(x)+(m+l+1-j)\hat{w}(y))}{\Gamma(1+(m-1)\hat{w}(x))\Gamma(1+(m+l-j)\hat{w}(y))} \cdot \\
\left. \cdot \frac{\Gamma(j+k-2m-1)}{\Gamma(j-m)\Gamma(k-m)} f_e(m,x,m+1+l-j,y) \right) .
$$

This solution is likewise valid for any model in our class, and in the homogeneous limit yields the correlation function

$$
P_e(k,l) = \frac{\Gamma(k)\Gamma(l)}{\Gamma(m)\Gamma(3+k+l)} \cdot \tag{12}
$$

$$
\left(\sum_{i=m+1}^{k} \frac{\Gamma(4+2m+k-i)\Gamma(i+l-2m-1)}{\Gamma(m+k+1-i)\Gamma(i-m)\Gamma(l-m)} P_e(m+k+1-i,m) \right. \\
\left. + \sum_{j=m+1}^{l} \frac{\Gamma(4+2m+l-j)\Gamma(j+k-2m-1)}{\Gamma(m+l+1-j)\Gamma(j-m)\Gamma(k-m)} P_e(m,m+l+1-j) \right) .
$$

Despite the complicated form of the solution, the most notable feature of this expression is that it does not factorize according to $f_e(k,x,l,y) = f_v(k,x) \cdot f_v(l,y)$, which provides an evidence of the nontrivial correlations that develop between adjacent nodes as a consequence of their dynamic nature.

5. Conclusions

To sum up, we have analyzed the correlations in the degrees and states of adjacent nodes in heterogeneous evolving networks. Node correlations are an important metric that in addition to the degree distribution provides a more thorough characterization of the structure of a network and may bear an important influence on the dynamics of interactions with such topology. We have shown that the technique of rate equations can be applied to model the evolution of the density function of links connecting nodes of given degree and state, yielding a comprehensive solution to the node correlations in the models of the proposed class. In the limit of homogeneity, the analysis also yields an exact solution for the correlations between the degrees of adjacent nodes in BA networks for an arbitrary number of links. The derived solutions evidence the existence of nontrivial correlations introduced by the evolution process, which distinguish these networks from those of static nature.

Acknowledgements

Support from MICINN Spain under contracts No. MTM2009-14621 and i-MATH CSD2006-32 is gratefully acknowledged.

References

[1] A.-L. Barabási and R. Albert, *Science* **286**, 509 (1999).

[2] R. Albert and A.-L. Barabási, *Rev. Mod. Phys.* **74 No. 1**, 47 (2002).

[3] A. Santiago and R. M. Benito, *Int. J. Mod. Phys. C* **18**, 1591 (2007).

[4] A. Santiago and R. M. Benito, *Europhys. Lett.* **82**, 58004 (2008).

[5] P. L. Krapivsky, S. Redner and F. Leyvraz, *Phys. Rev. Lett.* **85**, 4629 (2000).

[6] S. N. Dorogovtsev, J. F. F. Mendes and A. N. Samukhin, *Phys. Rev. Lett.* **85**, 4633 (2000).