Random Walks and Optimal Stopping Strategies as a Model for Evolving Networks

IC-MSQUARE 2012: International Conference on Mathematical Modelling in Physical Sciences IOP Publishing
Journal of Physics: Conference Series **410** (2013) 012092 doi:10.1088/1742-6596/410/1/012092

Random Walks and Optimal Stopping Strategies as a Model for Evolving Networks

D.S. Vlachos, K.J. Parousis-Orthodoxou and M.M. Stamos
Department of Computer Science and Technology,
Faculty of Sciences and Technology,
University of Peloponnese,
GR-221 00 Tripolis, Greece

E-mail: dvlachos@uop.gr

Abstract. In this work, a model for evolving networks is presented based on a Brownian particle. Each time, the Brownian particle enters the network through a randomly selected node. The random walk is terminated after having created m-links with visited nodes. Two strategies have been tested: in the first one, we used a generalized algorithm for the secretary problem in order to maximize the degree of the node with which the new node is connected, while in the second strategy, the Brownian particle creates links with nodes that meets twice. In all cases, scale free, modular, dissasortative networks are created.

1. Introduction

During the last years, there is an increasing interest for the study of complex networks from researchers working in diverse areas of science [1], [2], [3], [4]. Many real world complex systems can be described by networks of nodes or agents connected with edges or interactions. These structures are characterized by both the nature of nodes and the underlying interaction that is represented by the edges. Among the various classes of models used to describe complex networks, *evolving random networks* are by far more realistic. In these models, at each time, a node is added to the network and using a strategy, one or more links are created with other nodes. Let $N(k,t)$ be the number of nodes in the network that have degree less or equal to k at time t. Then, one can easily verify that

$$\frac{\partial N(k,t)}{\partial t} = 1 - q_k^t \left(N(k,t) - N(k-1,t)\right) \tag{1}$$

where q_k^t is the probability of a node with degree k to acquire an edge at time t and it is assumed that $k > m$, where m is the number of edges that each node carries when it is entered the network. The second part of equation (1) says that at each time a new node is entered the network with degree m (thus $N(k,t)$ is increased by 1) but those nodes who have degrre k (they are exactly $N(k,t) - N(k-1,t)$) can acquire an edge and thus the number of nodes with degree $\leq k$ is decreased. Taking the stationary case $t \to \infty$ and noting that $N(k,t) = tP(k)$, where $P(k)$ is the cumulative probability distribution of node degrees, we have

$$\frac{dP(k)}{dk} = \frac{1 - P(k)}{tq_k^t} \tag{2}$$

where he have used the fact that

$$N(k,t) - N(k-1,t) \sim tdP(k)/dk \tag{3}$$

For scale free networks and for k large enough, the cumulative degree distribution must behave as $1 - Ck^{-\gamma+1}$ and replacing in equation (2) we get

$$q_k^t = \frac{k}{(\gamma-1)t} \tag{4}$$

Since $t \sim \sum k$ (because t is equal to the total network nodes and each node carries the same number of edges) we have

$$q_k^t \sim \frac{k}{\sum k} \tag{5}$$

which says that nodes with higher degree are more likely to acquire new edges than the others. This property is known as *preferential attachement* and equation (5) shows its relation to the scale free character of random evolving networks. Since preferential attachement and scale free behavior seem interdependent, how realistic is the fact that when a node is entering the network, knows the degree distribution of all the other nodes? Fortunately, there are models like the *duplication-divergence* one, in which the preferential attachement rule emerges at an effective level from local principles [5]. Following this result, it is of great interest to investigate how local principles and rules based only on limited knowledge of the network topology, can emerge other characteristic properties of real wolrd networks (for example like assortativity or disassortativity, modularity and self-similarity). This question gives the cause for reflection in the current work. The current paper examines the formation of networks under the assumption that random walkers decide to create links with network nodes applying a specific strategy. This strategy can be the optimization of a certain outcome or it can be a simple heuristic rule. The paper is organized as follows: in section 2 general results about random walks and the formulation of optimal stopping strategies are given. In section 3 experimental tests are performed and in section 4 the results from these tests are discussed.

2. Optimal Stoping Strategies in Random Walks

Since many real networks are not static, in order to build a model for them, we adopt the hypothesis that the network grows by addition of new nodes and there is a set of rules govering the way new nodes establish connections with old ones. The proposed model can be summarized as follows: at each time step, a new node is entered in the network and connects with a candidate link to a random node of the network. Then a random walk inside the network is fired, and the final connections are decided during this walk. But before we present the details of the model, we summarize some basic results about the random walks in networks and we give a general formalism for optimal stoping strategies.

2.1. Random Walk in Networks

Consider the case where a new node is entered in a given network and this node has to decide what kind of connections should establish with other nodes. In order to make this decision, the new entered node must have some information about the other nodes, their connectivity patterns etc. But what might be its perspective of the network's topology? An obvious answer is to *walk* the network for some time. For simplicity, we assume that the network is connected, undirected and all edges have the same weight. Let N be the number of nodes before the enrty of the new one and A the adjacency matrix of the network ($A_j^i = 1$ iff there is an edge between nodes i and j, which means that i, j are nearest neighbors). The number of nearest neighbors

of node i is its degree k_i. If a t a given time t a Brownian particle is located at node i, then at time $t+1$ it has the same probability to be in any of the nearest neighbors of node i. One can easily verify that after a large amount of movements (compared to the total number of edges), the probability p^i for the Brownian particle to be at node i is given:

$$p^i = \frac{k^i}{k^j \cdot e_j} \tag{6}$$

For the Brownian particle, the observed distance between two nodes (the average number of moves that the particle has to make in order to reach from one node to the other) depends not only on the direct distance (the shortest path) but also on all the possible paths that lead from one node to the other. Let d^i_j be the observed distance between nodes i and j. Then

$$d^i_j = 1 + \sum_{l,l\neq j} P^i_l d^l_j \tag{7}$$

Let now $B(j)^l_m = P^l_m$ if $m \neq j$ and $B(j)^l_j = 0$, then the observed distances are given

$$[I - B(j)] \cdot \begin{pmatrix} d^1_j \\ d^2_j \\ \vdots \\ d^N_j \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \tag{8}$$

Equation 8 can be effectively solved in the case of sparse matrices (see for example [6] and [7]). The matrix d^i_j contsins all the necessary information about how far is node i from node j (assuming Brownian motion). Note here that $d^i_j \neq d^j_i$ (for example if the nodes i, j are connected and the degree of i is $k_i = 1$ while the degree of j is $d_j > 1$ the $d^i_j = 1$ but $d^j_i > 1$). It can be easily seen that nodes with high degree have a pronounced effect on d^i_j and if the clustering coefficient of the network is high enough, a random walk in the network for the majority of the time is located on nodes with high degree.

One important feature of a random walk in a network is the probability that since the random walk is started from an arbitrary node, after m steps, the current node has a specified degree. Thus, we introduce the *modified degree distribution* $P^{(m)}(k)$ which is given:

$$P^{(m)}(k) = \sum_{k_1,k_2,\ldots,k_m} P(k_1) \cdot P(k_2|k_1) \cdot \ldots \cdot P(k|k_m) \tag{9}$$

where

- $P(k)$ is the degree distribution (the probability that a node has degree k).
- $P(k|k')$ is the conditioned degree distribution (the probability that a node with degree k is connected to a node with degree k').

Equation 9 can be rewritten using the joint degree distribution $P(k, k')$ (the probability that two connected nodes have degrees k, k'):

$$P^{(m)}(k) = \sum_{k_1,k_2,\ldots,k_m} P^{(0)}(k_1) \frac{< k >^m}{k_1 \cdot k_2 \cdot \ldots \cdot k_m} \frac{P(k_1,k_2) \cdot P(k_2,k_3) \cdot \ldots \cdot P(k_m,k)}{P(k_1) \cdot P(k_2) \cdot \ldots \cdot P(k_m)} \tag{10}$$

where we have used the identity

$$P(k'|k) = \frac{< k >}{k} \frac{P(k,k')}{P(k)} \tag{11}$$

and $< k >$ is the mean degree. Obviously $P^{(0)}(k) = P(k)$.

2.2. Optimal Stopping Strategies

Let $X_1, X_2, \ldots$ be the outcomes of random identical independent experiments and assume that all of them follow the distribution $F(x)$ ($F(x)$ is the propbability that the outcome $X \leq x$). Suppose that a *Decision Maker* (DM) observes the sequence of up to n outcomes, and at each time has two choices:either to accept or reject the current outcome. Once a decision is made (an outcome is selected), the problem terminates. If reached the n^{th} outcome must be selected, and once rejected, an outcome cannot be recalled. At each time, the DM only observes an indicator of X_t which says whether each observed oucome is the best so far. The DM's payoff for selecting the t^{th} outcome with $X_t = x_t$ is itself x_t, while in the classical secretary problem this payoff is either 1 if the best outcome is selected and 0 otherwise [8], [9].

Let

$$I_t = \begin{cases} 1 & , \quad \text{if} X_t = \max\{X_1, X_2, \ldots, X_{t-1}, X_t\} \\ 0 & , \quad \text{otherwise} \end{cases} \tag{12}$$

The probability that the outcome $X_t \leq X$ given that X_t is the maximum observed one is given:

$$\begin{aligned} P(X_t \leq X | I_t = 1) &= F(X)^{t-1} \cdot F(X) \\ &= F(X)^t \end{aligned} \tag{13}$$

and the density function of the variable X_t given that it is the maximum observed is given:

$$f(X_t | I_t = 1) = t \cdot F(X_t)^{t-1} \cdot f(X_t) \tag{14}$$

where $f(X_t)$ is the unconditional density function of X_t ($f(x) = F'(x)$). Let now E_t be the mean value of X_t given that it is the maximum observed, that is $E_t = E(X_t | I_t = 1)$. Then:

$$\begin{aligned} E_t &= \int_0^1 t \cdot x \cdot F(x)^{t-1} \cdot f(x) \cdot dx \\ &= \int_0^1 x \cdot dF(x)^t \end{aligned} \tag{15}$$

Consider now the case where the DM is called to make multiple selections in order to maximize a given function of them. Let m be the number of selections and $G(X_1, X_2, \ldots, X_m)$ is the function to be maximized. Assume that G ia a non decreasing function in any of its m-arguments. The adopted strategy is again the obvious one: the first $c-1$ outcomes are skiped and then the m maximal outcomes are selected (until the number of tries n is reached). In order to calculate averages of the given function G, one has to calculate first the joint probability density function $f_c^{n,m}(x_{t_1}, x_{t_2}, \ldots, x_{t_m}) = f(x_{t_1}, x_{t_2}, \ldots, x_{t_m} | I_{t_1} = 1, I_{t_2} = 1, \ldots, I_{t_m} = 1, t_1 < t_2 < \ldots < t_m)$ with the parameters c and n for the selected strategy. The time t_1 that the first outcome is selected can be any between c and $n-m$. The time t_2 that the second outcome is selected can be any between $t_1 + 1$ and $n - m + 1$ and finally the time t_m that the last outcome is selected can be any between $t_{m-1} + 1$ and n. Let

$$r_{t_1}^{t_2} = (1 - p_{t_1}) \cdot (1 - p_{t_1+1}) \cdot \ldots \cdot (1 - p_{t_2-1}) \cdot p_{t_2}, \; t_1 \leq t_2 \tag{16}$$

be the probability that after bypassing $t_1 - 1$ outcomes, the t_2 outcome is the first with $I_{t_2} = 1$ (p_t is the probability that $I_t = 1$ as it was defined earlier). We can argue now that:

$$\begin{aligned} f_c^{n,m}(x_{t_1}, x_{t_2}, \ldots, x_{t_m}) &= \sum_{t_1=c}^{n-m} r_c^{t_1} \cdot f(x_{t_1} | I_{t_1} = 1) \cdot f_{t_1+1}^{n,m-1}(x_{t_2}, \ldots, x_{t_m}) \\ &\quad + \frac{r_c^{n-m}}{p_{n-m}} \prod_{i=1}^{m} f(x_{t_i}) \end{aligned} \tag{17}$$

Similarly,

$$f_{t_1+1}^{n,m-1}(x_{t_2},\ldots,x_{t_m}) = \sum_{t_2=t_1+1}^{n-m+1} r_{t_1+1}^{t_2}\cdot f(x_{t_2}|I_{t_2}=1)\cdot f_{t_2+1}^{n,m-2}(x_{t_3},\ldots,x_{t_m}) + \frac{r_{t_1+1}^{n-m+1}}{p_{n-m+1}}\prod_{i=2}^{m} f(x_{t_i}) \quad (18)$$

and recursively we can calculate $f_c^{n,m}$. Then, the mean value of the function G is given

$$E_G = \int f_c^{n,m}(x_{t_1},x_{t_2},\ldots,x_{t_m})\cdot G(x_{t_1},x_{t_2},\ldots,x_{t_m})\cdot dx_{t_1}\cdot dx_{t_2}\cdot\ldots\cdot dx_{t_m} \quad (19)$$

Finally, setting $\frac{\theta E_G}{\theta c}=0$ we can obtain the optimal strategy (the value of c).

3. Experimental results

3.1. Networks using optimal stopping strategies

Two networks have been constructed using the following strategies: The first network was build using a maximum number of steps equal to 5 and a cutoff parameter equal to 2 (using the results of the previous section). The second network was build using the same strategy, but now the Brownian particle is allowed to travel for more time by using a maximum number of steps equal to 20. In both networks, the Brownian particle stops its motion after making 2 links ($m=2$). The degree distribution of these two networks is drawn in figure 1. Both networks follow a power law in degree distribution with exponents 2.4 and 2 respectively.

3.2. Networks using the m-Crossing strategy

Two networks have been constructed using the following strategy (m-Crossing strategy): a Brownian particle is entered in the network and makes a link with the first node that it meets twice. The motion can be continued until m-links have been created. The first network was created using the m-Crossing strategy with $m=2$, while in the second network we used $m=1$. Both networks follow a power law in degree distribution with exponents 2.4 and 1.9 respectively. The degree distribution of these two networks is drawn in figure 1.

4. Conclusions

We have tested a model based on a Brownian particle for evolving networks. All networks created are scale free which means that there is an emerging preferential attachment rule for the creation of links. This can be easily verified since random walks on randomly selected network edges, visit nodes with higher degree more often. In the case of the optimal stopping strategy, if the number of allowed steps is high enough, the Brownian particle can reach the most connected hubs and thus the exponent of the power law is low. On the other hand, if the number of allowed steps is low, then a percentage of links are created randomly and thus the exponent in the power law is higher [11].

Concerning the degree correlation, these networks are dissasortative. Concerning the modularity of these networks, it must be quite high, since if a Brownian particle enters in the neighborhood of a hub, and having in mind the strong dissasotativity of the network, it is difficult to escape. Moreover, the clustering coefficient of these networks must be quite high if the number of links created by newly entered node is > 1. Again this can be verified by the strong dissasortativity of the network. In the case where only one link is created, the network must have very low clustering coefficient, since the probability for the formation of a triangle is negligible.

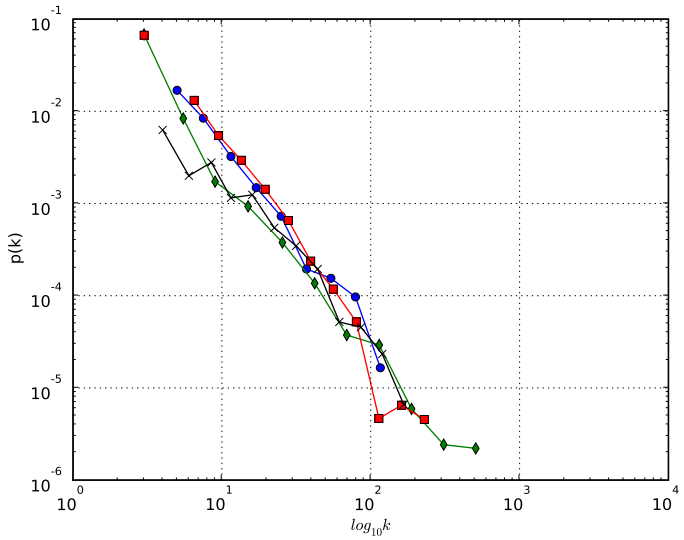

Figure 1. Degree distribution for the following networks: ∘-network using optimal stopping strategy with maximum steps 5 and cutoff paramete 2 ($\gamma \sim 2.4$), $\diamond$ -network using optimal stopping strategy with maximum steps 20 and cutoff paramete 2 ($\gamma \sim 2.$), □- network using the m-Crossing strategy with $m = 2$ ($\gamma \sim 2.4$) and ×-network using the m-Crossing strategy with $m = 1$ ($\gamma \sim 1.9$).

Acknowledgement
The project is implemented under the Operational Program "Education and Lifelong Learning" and funded by the European Union (European Social Fund) and by Greek national resources.

[1] Albert, R., Barabasi, A.-L.: Statistical Mechanics of Complex Networks. Rev. Mod. Phys. 74, 47 (2002)
[2] Dorogovtsev, S. N., Mendes, J. F. F..: Evolution of Networks: From Biological Nets to the Internet and WWW. Oxford University Press, Oxford (2003)
[3] Pastor-Satorras, R., Vespignani, A.: Evolution and Structure of the Internet: A Statistical Physics Approach. Cambridge University Press, Cambridge (2004)
[4] Cohen, R., Havlin, S.: Complex Networks: Structure, Stability and Function. Cambridge University Press, Cambridge (2008)
[5] Vázquez, A.: Disordered networks generated by recursive searches, Europhysics Letters 54, 430 (2001)
[6] Tewarson, R.P.: Sparse Matrices. Academic Press, New York (1973)
[7] Davis, T.A., Duff, I.S..: A combined unifrontal/multifrontal method for unsymmetric sparse matrices, ACM Trans. Math. Softw. 25, 1–3 (1999)
[8] Ferguson, T.S..: Who solved the secretary problem?. Statistical Science 4, 282–296 (1989)
[9] Samuels, S.M.: Secretary problems. In: Gosh, B.K., Sen, P.K. (eds.) Handbook of sequential analysis, pp. 381–405. Marcel Dekker, New York (1991)
[10] Bearden, J. Neil.: A new secretary problem with rank-based selection and cardinal payoffs. Journal of Mathematical Psychology 50, 58–59 (2006)
[11] Xenides, D., Vlachos, D.S., Simos, T.E.: Synchronization in complex systems following a decision based queuing process: rhythmic applause as a test case, Journal of Statistical Mechanics: Theory and Experiment P07017 (2008)