IJCSS International Journal of Complex Systems in Science

Int. J. Complex Systems in Science
vol. **1(2)** (2011), pp. 124–128

Matrix growth models based on centrality measures: a first analysis

Francisco Pedroche[1,†], Regino Criado[2], Esther García[2] and Miguel Romance[2]

[1] *Institut de Matemàtica Multidisciplinària, Universitat Politècnica de València. Espanya.*

[2] *Departamento de Matemática Aplicada, Universidad Rey Juan Carlos. España*

Abstract. A general growth model of random networks based on centrality measures is introduced. This formalism extends the well-known models of preferential attachment. We propose to set the preferential attachment using a linear function of some centrality measures ranging from local to global scale. The aim is to include spectral measures, such as PageRank and Bonacich, and geodesic measures, such as betweenness and closeness. In this paper we present a first analysis using degree and Personalized PageRank.

Keywords: Growing networks, PageRank, In-Degree
MSC 2000: 90B15, 15A51, 68P10

† **Corresponding author:** pedroche@imm.upv.es
Received: October 14, 2011
Published: October 24, 2011

[1]Supported by Spanish DGI grant MTM2010-18674
[2]Supported by Spanish DGI grant MTM2009-13848

1. Introduction

Most Network growth models are based on the preferential attachment model [1]. We are interested in a matrix formulation of a general class of preferential attachment. In this paper we present a theoretical framework. We include in-degree and Personalized PageRank [4], [5]. To our knowledge the first models of preferential attachment based on PageRank were [2] and [3]. Both models are based on the usual personalization vector, i.e. $\mathbf{v} = \mathbf{1}/n$. We improve the fundamentals of models that use PageRank by including a general personalization vector in our description.

2. Definitions

Let $A^{(0)} = (a_{ij}^{(0)}) \in \mathbb{R}^{n\times n}$ be the adjacency matrix of a directed network: $a_{ij} = 1$ when there is a link from node i to node j, and $a_{ij} = 0$ in other case. Let $q_i^{(0)}$ be the indegree of node i. At each time step, a new node is added to the network. This node connects with $0 < m \leq n$ different existing nodes. The adjacency matrix $A^{(k)} \in \mathbb{R}^{(n+k)\times(n+k)}$, when k nodes have been added to the network, is given by $A^{(k)} = \begin{pmatrix} A^{(k-1)} & \mathbf{r}^{(k-1)} \\ \mathbf{s}^{(k-1)T} & 0 \end{pmatrix}$ where $\mathbf{r}^{(k-1)} \in \mathbb{R}^{(n+k-1)\times 1}$, and $\mathbf{s}^{(k-1)} \in \mathbb{R}^{(n+k-1)\times 1}$, are probability distribution vectors.

We propose to set the preferential attachment by setting $\mathbf{r}^{(k-1)}$ and $\mathbf{s}^{(k-1)}$ as linear functions of some centrality measures ranging from local to global scale. We denote by $q_i^{(k)}$ the indegree of node i corresponding to the graph given by $A^{(k)}$. Let us denote $p(r_i^{(k)} = 1)$ the probability that the variable $r_i^{(k)}$ takes the value 1.

Albert-Barabasi growth model is given by:$\mathbf{r}^{(k)} = \mathbf{s}^{(k)}$, $k = 0, 1, 2, \ldots$ with $p(r_i^{(k)} = 1)_{AB} = \frac{q_i^{(k)}}{\sum_j^n q_j^{(k)}}$, $\forall i \in \mathcal{N}$, $k = 0, 1, 2, \ldots$. We denote $\mathbf{p}_{AB}^{(k)} = [p(r_1^{(k)} = 1)_{AB}, \ldots, p(r_n^{(k)} = 1)_{AB}]^T$.

Personalized PageRank model is based on computing the random vectors $\mathbf{r}^{(k)}$ and $\mathbf{s}^{(k)}$ taking: $p(r_i^{(k)} = 1)_{PR} = PR^{(k)}(i, \mathbf{v}_{out}^{(k)})$, and $p(s_i^{(k)} = 1)_{PR} = PR^{(k)}(i, \mathbf{v}_{in}^{(k)})$, $\forall i \in \mathcal{N}$, $k = 0, 1, 2, \ldots$ where $PR^{(k)}(i, \mathbf{v}^{(k)})$ is the i-th entry of the personalized PageRank vector corresponding to the graph given by $A^{(k)}$, and using the personalization vector $\mathbf{v}^{(\mathbf{k})}$. We denote: $\mathbf{p}_{PR,out}^{(k)} = [p(r_1^{(k)} = 1)_{PR}, \ldots, p(r_n^{(k)} = 1)_{PR}]^T$, and $\mathbf{p}_{PR,in}^{(k)} = [p(s_1^{(k)} = 1)_{PR}, \ldots, p(s_n^{(k)} = 1)_{PR}]^T$.

3. Results

Let us consider the toy graph in Fig. 1. We compare three different growth models: AB, PageRank with $\mathbf{v}^{(k)} = \mathbf{v_{in}} = \mathbf{v_{out}} = \mathbf{1}/(n+k)$ (that we denote

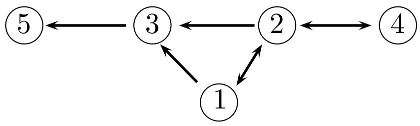

Figure 1: Toy graph

as PR) and PageRank with $\mathbf{v}^{(k)} = \mathbf{v_{in}} = \mathbf{v_{out}} = (0, 0, \ldots, 0, 1)^T$, that we denote as PPR (note that in this case we assume that the entering node has preference for the last node). In all the experiments we use 2000 steps, and $m = 3$. From Fig. 2 (left) we see that AB and PR tend to have the same probability distribution of preferential attachment. In this plot the difference decays to 2%. As a consequence, when the network is sufficiently large we have that both preferential models tend to the same growing pattern. Therefore we expect that for sufficiently large networks both models give the same in-degree distribution and PageRank distribution. This is in accordance with experiments in [3], [6], [7]. In another experiment we obtain that the quantity $\|\mathbf{p}_{AB} - \mathbf{p}_{PPR}\|_2$ oscillates around a mean value of the 20% (Fig. 2, right). Therefore, both models offer different networks. This is also shown in Fig. 3, where the adjacency matrices are shown.

In Fig. 4 we show the log-log distribution $p(k)$, i.e. the probability of having in-degree k obtained with the three models. We see that all the models follow a power-law function, which is very similar for the three models when $k \leq 100$.

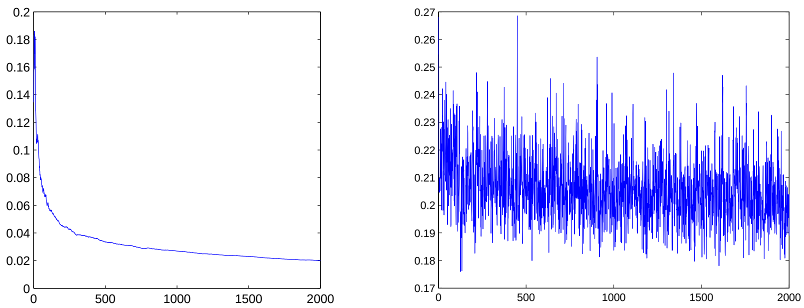

Figure 2: Evolution of $\|\mathbf{p}_{AB} - \mathbf{p}_{PR}\|_2$ (left) and $\|\mathbf{p}_{AB} - \mathbf{p}_{PPR}\|_2$. 2000 steps, $m = 3$.

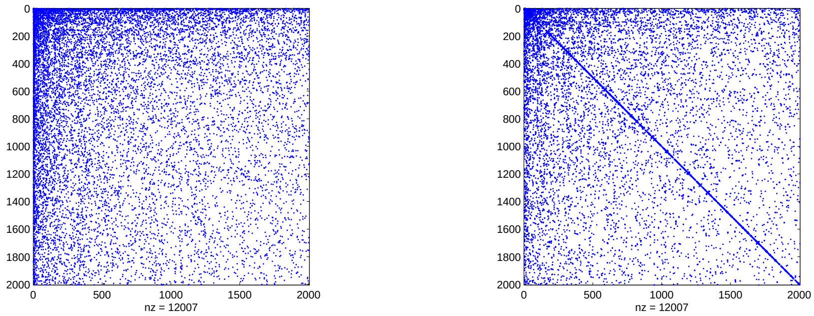

Figure 3: Adjacency matrix $A^{(2000)}$ in an execution of AB model (left) and PPR model (right), 2000 steps, $m = 3$.

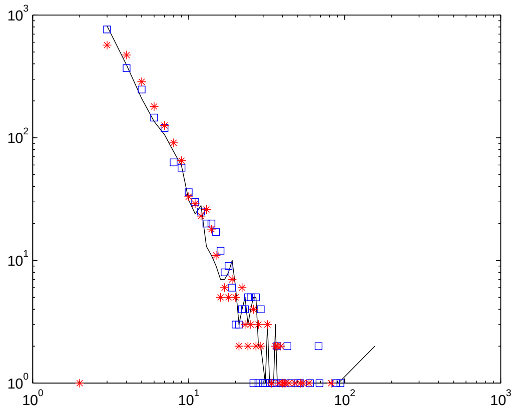

Figure 4: Log-log Distribution of $p(k)$ for the three models studied. AB (black), PR (blue) and PPR (red).

4. Conclusions

We have presented a model of network growth using a matrix formulation. This formulation allows us to include some centrality measures to guide the preferential attachment. We show three different models of growing in this framework, including a model that uses Personalized PageRank. In this first analysis we obtain the following conclusions from our experiments: 1) AB and PR lead to the same preferential attachment, 2) PPR differs from AB, and 3) PPR produces a network in which the in-degree distribution $p(k)$ follows a power-law. Future lines include to introduce new centrality measures and to stablish a general setup to quantify the differences between the models.

References

[1] R. Albert and A.L. Barabási, *Rev. Mod. Phys.* **74**, 47, (2002).

[2] E. Drinea, M. Enachescu, and M. Mitzenmacher, *Variations on Random Graph Models for the Web. Harvard TR-06-01*, (2001).

[3] G. Pandurangan, P. Raghavan, and E. Upfa, *Internet Math.*, **3**, 1, (2006).

[4] T. Haveliwala, S. Kamvar, G. Jeh., *An Analytical Comparison of Approaches to Personalizing PageRank, TR. Stanford.* (2003).

[5] A. N. Langville and C. D. Meyer, *Google's Pagerank and Beyond: The Science of Search Engine Rankings. Princeton U.P.*, (2006).

[6] N. Litvak, W.R.W. Scheinhardt, and Y.V. Volkovich, *Internet Math*, **4**, 2-3. (2007)

[7] S. Fortunato, M. Bofuñá, A. Flammini, and F. Menczer *Internet Math.*, **4**, 2-3 (2007)