Random Graph Generator for Bipartite Networks Modeling

Szymon Chojnacki and Mieczysław Kłopotek

Institute of Computer Science,
Polish Academy of Sciences
J.K. Ordona 21, 01-237 Warsaw, Poland
{sch, klopotek}@ipipan.waw.pl

Abstract. The purpose of this article is to introduce a new iterative algorithm with properties resembling real life bipartite graphs. The algorithm enables us to generate wide range of random bigraphs, which features are determined by a set of parameters. We adapt the advances of last decade in unipartite complex networks modeling to the bigraph setting. This data structure can be observed in several situations. However, only a few datasets are freely available to test the algorithms (e.g. community detection, influential nodes identification, information retrieval) which operate on such data. Therefore, artificial datasets are needed to enhance development and testing of the algorithms. We are particularly interested in applying the generator to the analysis of recommender systems. Therefore, we focus on two characteristics that, besides simple statistics, are in our opinion responsible for the performance of neighborhood based collaborative filtering algorithms. The features are node degree distribution and local clustering coefficient.

Keywords: complex networks, random graphs, bipartite graphs, recommender systems, affiliation networks

1 Introduction

The analysis of large networks is driven by the desire to understand and model as diverse phenomena as the spread of infection, social communities creation, protein interactions or website importance assessment [1]. The interest of research community in complex networks was fueled by an empirical evidence which proved that some properties of real-life graphs are unachievable for classic random models. Moreover, the similar properties are common to networks observed in various fields. Several statistics describing networks can be measured. However, node degree distribution and mean clustering coefficient are two measures of a great importance. They are correlated for example with such macro features as an average length of a path between two nodes, the network's resilience to an attach or the pace of spread of innovations. It turns out that in diverse real-life networks:

– node degree distribution is heavy-tailed
– mean clustering coefficient is bounded away from zero

In the classic theory of random graphs developed by two Hungarian mathematicians Paul Erdős and Alfréd Rényi [2] the asymptotic node degree distribution is *Poisson*. Also the value of clustering coefficient, which measures the probability that two nodes sharing a friend are connected differs from empirical results and tends to zero as a number of nodes grows.

The seminal paper of Barabási and Albert [3] describes the driving forces which are responsible for the heavy-tailed node degree distributions. The property can be attributed to both: the growth and the preferential attachment mechanism. Moreover, none of the two results in the desired distribution on its own. Kumar and collaborators [4] proposed to substitute the preferential attachment mechanism with random selection of a neighboring node, which also leads to the heavy-tailed distribution. Liu [5] described how a mixture of preferential and random attachment enables us to generate networks with weakened heavy-tail. Vázquez [6] proposed a random graph generative procedure which results in networks with positive values of the clustering coefficient. The combined translation of the four results onto the ground of bigraphs comprises the frame of our algorithm.

Recently a few random bipartite graph generating algorithms have been introduced ([7],[8], [9], [10]). However, none of them enables to generate growing networks with varying distributions and clustering coefficient bounded away from zero.

Our contribution comprises four main results:

1. definition and formal justification of new local clustering coefficient dedicated for bigraphs - bipartite local clustering coefficient (BLCC)
2. introduction of *bouncing mechanism* responsible for the growth of BLCC
3. description and analysis of new versatile bigraph generator
4. identification of a relationship between network properties of bigraphs and the properties responsible for the complexity of recommender systems

The rest of the article is organized as follows. In Section 2 we formalize node degree distributions, local clustering coefficient and introduce BLCC. In Section 3 we outline the motivation for our research, which is based on the equivalence of bipartite graphs and user-item matrices in the recommender systems. The fourth section contains a description of our algorithm. In Section 5 we present the results of numerical simulations. The last sixth section is dedicated for the concluding remarks. Advanced mathematical transformations are described in details in two appendices.

2 Background

A graph is an ordered pair $G = (V, E)$ comprising a set of vertices V and a set of edges $E \subseteq \{V \times V\}$. A bipartite network is a graph $G = (U \cup I, E)$ which vertices can be labeled by two types U and I. The difference with a classic unipartite graph is the fact that V consists of two disjoint sets $V = \{U \cup I, U \cap I = \emptyset\}$ and edges exist only between nodes of different types $E \subseteq \{U \times I\}$. We analyze undirected graphs.

2.1 Node degree

A degree of a node stands for the number of direct (first) neighbors of the node and is equal to the number of node's edges. The probability density function (pdf) of node degree distributions in real-life datasets is usually skewed (Fig. 2). If the tail decays slowly we can observe the power-law distribution $pdf_{PL}(x) = ax^{-k}$. The tail vanishes quickly in the exponential distribution $pdf_{EX}(x) = \lambda e^{-\lambda x}$. It is convenient to visualize the two distributions on a log-log scale. From the fact that $\log(pdf_{PL}(x)) = -k\log(x) + log(a)$ follows that the power-law distribution is shaped in a straight line on a log-log chart. This distribution is called *scale-free* because $pdf_{PL}(cx) = a(cx)^{-k} = ac^{-k}pdf_{PL}(x)$. The distributions observed in real networks can not be generated by classic random graphs. The graphs studied by Erdős give the Poisson distribution. The three types of distributions are drawn in Fig. 1.

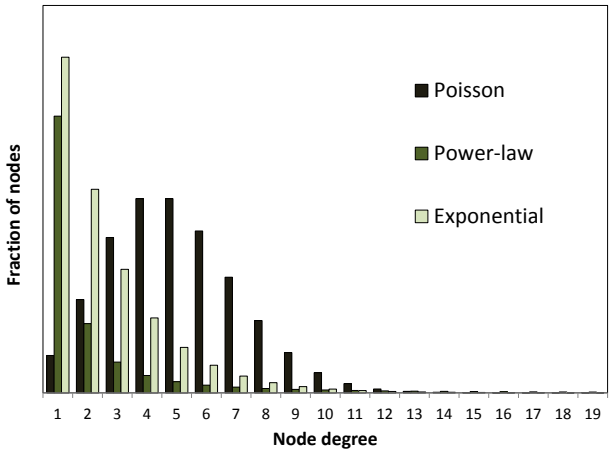

Fig. 1. Three degree distributions with the same average. The Poisson distribution is characteristic for classic random graphs. The exponential and the power-law distributions are more common in real datasets. Both of them are skewed. However, the tail of the power-law distribution decays slower.

2.2 Local clustering coefficient

Local clustering coefficient is used to measure the probability that if two nodes share a neighbor than they are also connected. It is computed for each node and an average over all nodes indicates the level of network's transitivity. Let's denote by c_j the number of connected pairs among the direct neighbors of node j and by k_j the degree of node j. The local clustering coefficient (LCC) is given by:

$$LCC_j = \frac{c_j}{k_j(k_j-1)/2}. \tag{1}$$

The value of LCC is zero for any node in a bipartite graph. Therefore, we introduce a new coefficient dedicated to measuring transitivity in bigraphs. Bipartite local clustering coefficient ($BLCC$) of node j takes values of one minus the proportion of node's second neighbors to the potential number of the second neighbors of the node. The value of $BLCC$ calculated for node j is given by:

$$BLCC_j = 1 - \frac{|N_2(j)|}{\sum_{i \in N_1(j)} (k_i - 1)}, \tag{2}$$

where $|N_2(j)|$ stands for the number of the second neighbors of node j, $N_1(j)$ is a set of the first neighbors of node j.

In order to justify the correlation between LCC and $BLCC$, we consider the values of the two coefficients in case of a unipartite graph. We denote by $f(c)$ in Eq. (3) the value of LCC calculated for a random node with c pairs of connected neighbors. We use $g(c)$ in Eq. (4) to assess the value of BLCC in case of the same node. Except of c pairs we follow the tree like structure assumption. We substitute k_i with $\frac{\langle k^2 \rangle}{\langle k \rangle}$ (i.e. the expected degree of a *neighboring node*[1] [14]) and observe that on average $|N_1(j)| = \langle k \rangle$. The logic of deriving $|N_2(j)|$ is presented in Fig. 3.

$$f(c) = \frac{2c}{\langle k \rangle (\langle k \rangle - 1)} = \frac{2c}{\langle k \rangle^2 - \langle k \rangle} \tag{3}$$

$$g(c) = 1 - \frac{\langle k \rangle \left(\frac{\langle k^2 \rangle}{\langle k \rangle} - 1 \right) - 2c}{\langle k \rangle \left(\frac{\langle k^2 \rangle}{\langle k \rangle} - 1 \right)} = \frac{2c}{\langle k^2 \rangle - \langle k \rangle} \tag{4}$$

From the fact that the variance of any distribution is nonnegative and it can be decomposed as $\sigma^2 = \langle k^2 \rangle - \langle k \rangle^2$, we assert that $g(c)/f(c)$ is constant and not larger than one.

We also considered a different definition of the number of potential second neighbors in Eq. 2. Within the local tree-like structure setting [15] it can be approximated by $\langle u \rangle \left(\frac{\langle v^2 \rangle}{\langle v \rangle} - 1 \right)$. Even though on average such definition gives positive fractions (Table 1), a value of BLCC calculated for one node can be negative and therefore we stay with the definition of BLCC as it is in Eq. 2.

[1] The formula for an average degree of a neighboring node is derivated in appendix A.

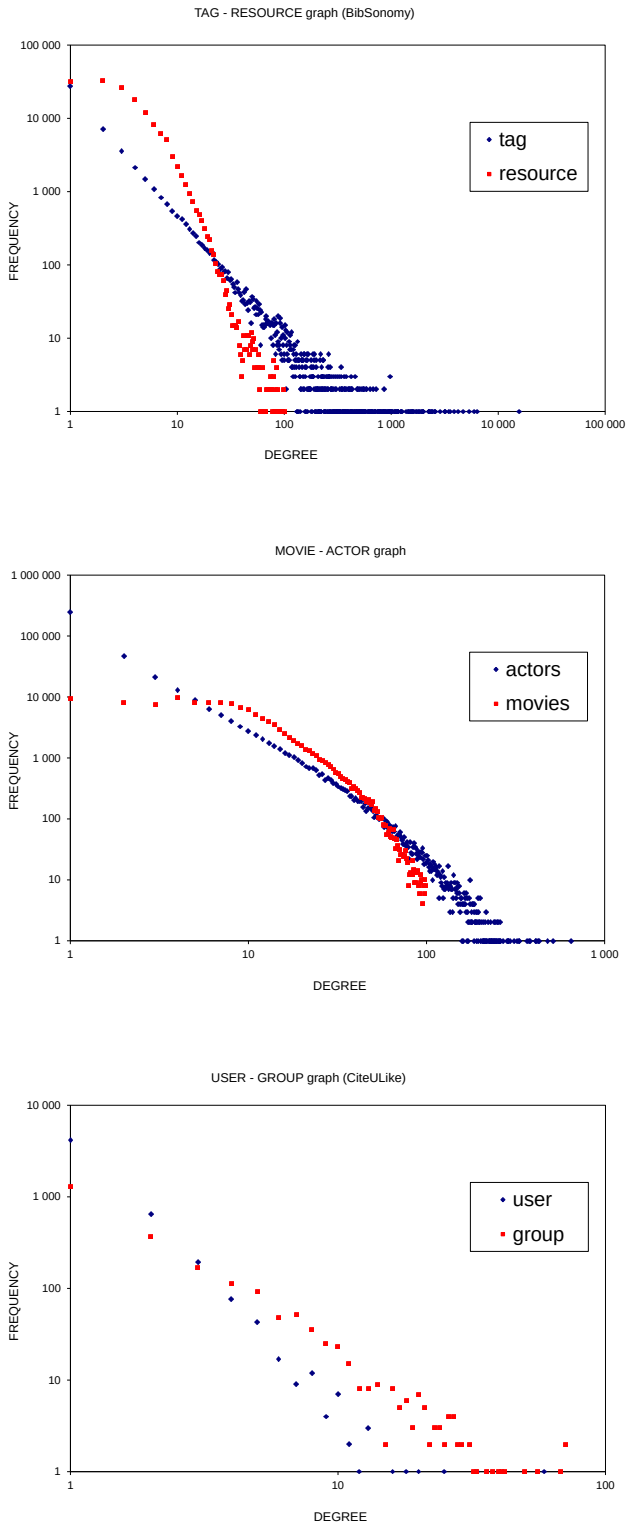

Fig. 2. The node degree distributions of three bipartite graphs. The straight line of points (on a LOG-LOG scale) in all three datasets envisions the power-law feature of the datasets. In case of BibSonomy [11] (upper chart) and IMDB [12] (middle chart) graphs, one modality tends towards exponential distribution. In case of CiteULike [13] (lower chart) dataset both modalities are shaped in a straight line.

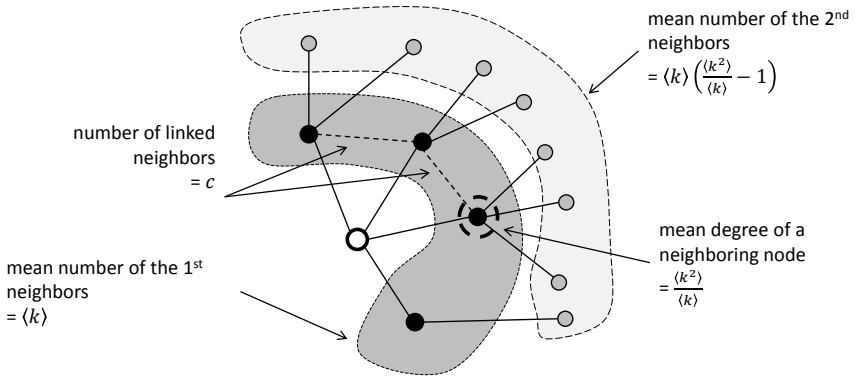

Fig. 3. In order to compute the BLCC for a unipartite graph we need to assess the potential number of the second neighbors of a given node. A random node has $\langle k \rangle$ neighbors (in the figure $\langle k \rangle = 4$). There are c connections among the neighbors on average ($c = 2$). Each neighbor has on average $\frac{\langle k^2 \rangle}{\langle k \rangle}$ edges. Each edge points to a second neighbor of the considered node or to the node ($\langle k \rangle$ edges) or to the first neighbor ($2c$ edges). We assume that there are no two different edges pointing to the same second neighbor.

	basic statistics			second neighbors		
	users	items	edges	real	theoretic	$\frac{real}{theoretic}$
CEO [16]	26	15	98	21.8	22.0	0.99
CiteULike [13]	5 208	2 336	7 196	14.2	23.9	0.59
BibSonomy [11]	3 617	93 756	253 366	500.4	6 579.2	0.08
YouTube [17]	94 238	30 087	293 360	1 269.6	2 101.3	0.60
IMDB [12]	383 640	127 823	1 470 404	78.4	211.4	0.37
Flickr [17]	395 979	103 631	8 545 307	1 217.4	52 704.9	0.02
LiveJournal [17]	3 201 203	7 489 073	112 307 385	785 194.2	1 521 273.4	0.52
Orkut [17]	2 783 196	8 730 857	327 037 487	334 863.6	2 294 114.8	0.15

Table 1. An average number of the second neighbors in eight real-life datasets is smaller than approximated by the Newman's asymptotic formula (theoretic value). The most significant shrinking is observed in the Flickr dataset. The shrinking is observed in both relatively small and very large datasets.

3 Recommender systems

Recommender systems are an important component of the Intelligent Web. The systems make information retrieval easier and push users from typing queries towards clicking at suggested links. We experience real-life recommender systems when browsing for books, movies or music. The engines are an essential part of such websites as *Amazon*, *MovieLens* or *Last.fm*. The interest of research community in the systems was fueled by the Netflix movie recommendation competition [18]. During the challenge the state-of-art systems in terms of accuracy were developed.

However, it has been shown recently during the ECML Discovery Challenge 2009 [19] that the most accurate recommender systems fail to meet real-life constraints. It is not an easy task to update trained models when new items or users enter the evaluation. The problem is usually referred to as the *Cold Start* problem. These observations constitute the motivation for our research. We believe that there exists a need for algorithms that can generate random recommendation matrices (or equivalently bipartite graphs). We are particularly interested in the neighborhood-based techniques. These methods are the best suited for the dynamically changing scenarios, but the latency of creating a recommendation depends significantly on the structure of underlying dataset (compare Fig. 4). Moreover, because of embedding iterative mechanism in our generator, it can be used to simulate the *Cold Start* cases.

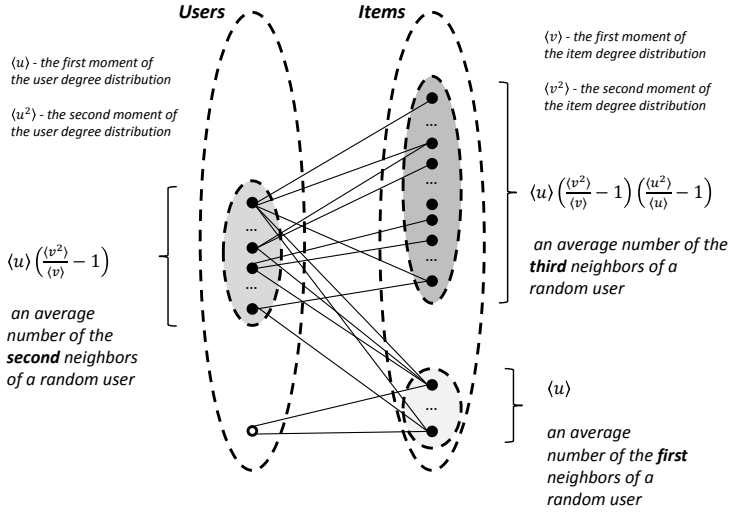

Fig. 4. In recommender systems based on the neighborhood principle the recommended items are selected from the items of the users that have rated at least one common item with an analyzed user.

4 Our algorithm

Our algorithm consists of three steps: (1) new node creation, (2) edge attachment type selection and (3) running bouncing mechanism. The procedure requires specifying eight parameters:

m - the number of initial loose edges with a user and an item at the ends
T - the number of iterations
p - the probability that a new node is a user
$(1-p)$ is the probability that a new node is an item
u - the number of edges created by each new user
v - the number of edges created by each new item
α - the probability that a new user's edge is being connected to an item with preferential attachment
β - the probability that a new item's edge is being connected to a user with preferential attachment
b - the fraction of preferentially attached edges that where created via a *bouncing mechanism*

Steps (1) and (2) are explained in Sec. 4.1 and analyzed in Sec. 4.2. In Sec. 4.3. step (3) is discussed.

4.1 Basic model

In the basic model we utilize first seven parameters. The bouncing mechanism is applied in the full model as an additional third step.

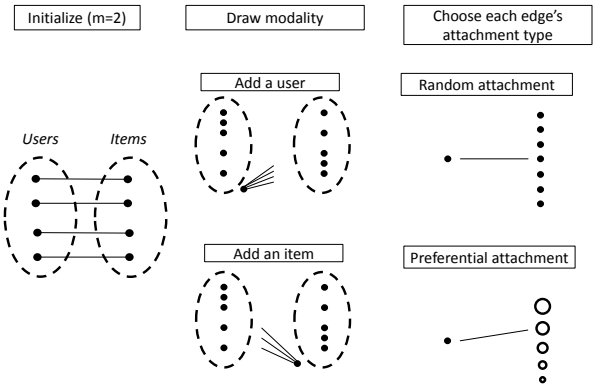

Fig. 5. The bipartite random graph generator is initialized with a set of m pairs of users and items. During each iteration two steps are performed. In the first step the type of new node is determined. In the second step a decision is made on the level of each node's edge whether to draw its ending with preferential attachment or randomly. In the preferential attachment variant the probability that a node is drawn is proportional to its degree.

The basic model is based on an iterative repetition of two steps (Fig. 5).
Step 1 If a random number is greater then p create a new user with u loose edges, otherwise create a new item with v loose edges.
Step 2 For each edge decide whether to join it to a node of the second modality randomly or with preferential attachment. The probability of selection preferential attachment is α for new user and β for new item.

4.2 Formal analysis

One can see that after t iterations the bigraph consists of $|U(t)| = 2m+pt$ users, $|I(t)| = 2m+(1-p)t$ items, and $|E(t)| = 4m+t(pu+(1-p)v)$ edges. Let's denote by η an average number of edges created during one iteration $\eta = (pu+(1-p)v)$. After relatively many iterations ($t >> m$) we can neglect m. In the presented model, an average user degree is:

$$\frac{|E(t)|}{|U(t)|} = \frac{4m + t(pu + (1-p)v)}{2m + pt} = \frac{\eta}{p},$$

analogously an average item degrees is:

$$\frac{|E(t)|}{|I(t)|} = \frac{\eta}{(1-p)},$$

the values are time invariant, but depend on both u and v.

In the following deduction we look from user modality perspective. However, the computations can be altered to the opposite item modality easily. In order to derive asymptotic node degree distribution in our model we need to specify the probability that a user node j with degree k_j gets connected to a new item. The quantity is usually represented as $\Pi(k_j)$ within the complex networks community. If nodes are selected randomly than:

$$\Pi_{random}(k_j) = \frac{1}{|U(t)|} = \frac{1}{pt}.$$

In case of random attachment $\Pi(k_j)$ does not depend on k_j. If nodes are selected with accordance to the preferential attachment rule than:

$$\Pi_{preferential}(k_j) = \frac{k_j}{|E(t)|} = \frac{k_j}{\eta t}.$$

Contrary to the random attachment scenario, the probability of node's selection is linearly proportional to its current degree. The probability of drawing a node with degree k_j is the degree divided by the number of edges. We can verify that by summing the values of Π over all user nodes we get one $\sum_j \Pi_j = 1$. In our model the decision whether to draw a user for an item with random or preferential attachment depends on β, hence the combined formula is:

$$\Pi(k_i) = \beta\frac{1}{pt} + (1-\beta)\frac{k_i}{\eta t}. \tag{5}$$

The equation (5) enables us to describe the pace of growth of nodes all with degree k_i as

$$\frac{\partial k_i}{\partial t} = (1-p)v\Pi(k_i). \tag{6}$$

We assume in the above equation that time interval between iterations is very small and that all nodes with a given degree grow in the same way. We show in the appendix that

$$P(k) \propto \left(\frac{\beta\eta + p(1-\beta)k}{\beta\eta + p(1-\beta)u}\right)^{\frac{-\eta}{(1-p)(1-\beta)v}-1}. \tag{7}$$

One can verify that for $\beta = 0$ we get power-law distribution. If $\beta \to 1$, we can utilize the fact that $\lim_{n\to\infty}\left(1+\frac{c}{n}\right)^n = e^c$ in order to obtain exponential distribution. The above result is consistent with [3]. When we put $\beta = 0$, $p = 0.5$ and $u = v$ we have power-law distribution with the scaling exponent equal to 3.

4.3 Full model

We have shown recently that node degree distributions of both modalities can be responsible for BLCC in some networks, but in others there exist additional shrinking forces responsible for high values of BLCC [20]. Therefore we introduce the *bouncing mechanism* (Fig. 6), which is based on surfing the web technique [6]. The mechanism enables us to rise BLCC, but can only by applied to the edges that are to be selected with preferential attachment. This can be attributed to the fact that the probability that a random walk is finished in a node is proportional to its degree [21]. Bouncing is performed in three micro steps: (1) a random node is drawn from the nodes that are already joined with the new node, (2) a random neighbor of the drawn node is chosen, (3) a random neighbor of the neighbor is selected for joining with the new node.

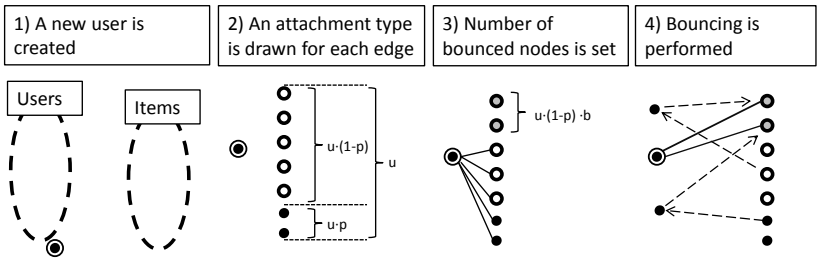

Fig. 6. For each edge of a new node, that is to be connected with an existing node with accordance to the preferential attachment mechanism, a decision is made whether to create it via a bouncing mechanism. In case of attaching new user node, u new edges are created. On average $u \cdot \alpha$ edges' endings are to be drawn preferentially and $u \cdot \alpha \cdot b$ of them are to be obtained via bouncing from the nodes that are already selected.

Algorithm 1: An iteration of the bipartite graph generator

```
if RAND() ≤ p then
    // p - the probability that a new node is a user
    for k ← 1 to u do
        // u - the number of edges created by anew user
        if RAND() ≤ α then
            // α - the probability that the new user's item is
               drawn preferentially
            if RAND() ≤ b then
                // b - the probability that new preferential node
                   was chosen by bouncing
                SelectedItem ← BounceFromRandom(TempItems) ;
            else
                SelectedItem ← DrawItemPreferentially() ;
                TempItems ← SelectedItem ;
        else
            SelectedItem ← DrawItemRandomly ;
            TempItems ← SelectedItem;
    Users ← Users ∪ NewUser;
    Edges ← Edges ∪{TempItems × NewUser} ;
else
    Process analogously with new item node
```

5 Numerical results

The results of the numerical experiments are divided into three subsections. In the first part we shortly present a Java applet developed in our Lab to play with various parameters of the generator. In the second part we show which parameters impinge on the values of node degree distributions and BLCC. In the last section we show how the number of potentially similar users and the number of their items can be determined by various levels of the generators parameters.

5.1 Graphical analysis

The applet presented in Figure 7 can be accessed online in `http://www.ipipan.eu/~sch/software/applet.html`. All parameters (except of the initial number of pairs) can be changed during graph generation. The distributions of BLCC and node degrees are being updated online for both modalities. Alse the average number of potentially similar users and their items is visualized at a chart. By an expression *similar user* we understand all users that have rated at least one item in common with the selected user.

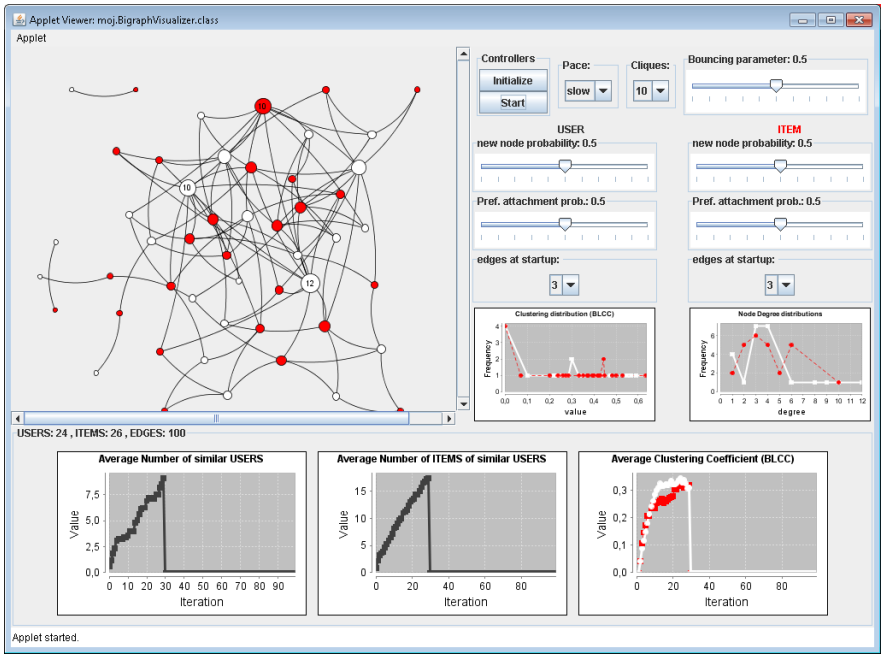

Fig. 7. A bigraph generated after $t = 30$ iterations. The values of all probabilities were set to 0.5, each new node creates three new edges $u = v = 3$, initial number of pairs $m = 10$.

5.2 Social network properties

We consider node degree distributions of both modalities and the values of BLCC as the network properties of the generated graphs. Node degree distributions are controlled by two parameters: α and β. We show in Figure 8 that if one parameter tends to one, the shape of appropriate modality becomes power-law. Low values output exponential distribution. Moreover, we do not observe any correlation between the distributions of both modalities.

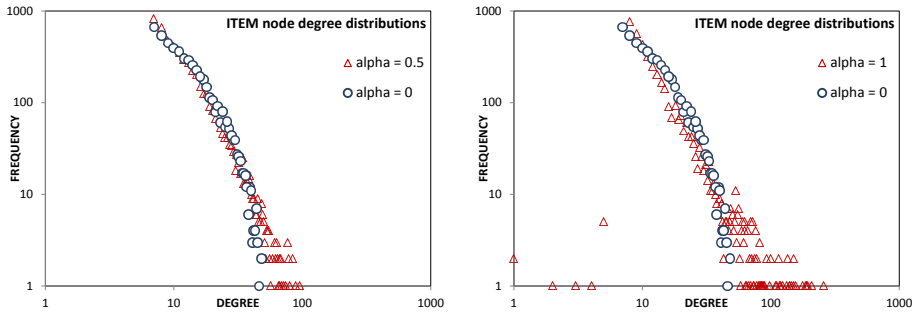

Fig. 8. Left Panel: blue circles indicate that the random attachment of users' edges (i.e. items) results in the exponential distribution of item degrees. Red triangles in both panels show that as $\alpha \to 1$ the distribution becomes power-law. Experiments run with ($m = 50$, $T = 10\ 000$, $p = 0.5$, $u = v = 7$, $\beta = 0.5$).

The values of BLCC (*bipartite local clustering coefficient*) can be controlled by the extend of the bouncing mechanism (Figure 9).

If we neglect the bouncing mechanism ($b = 0$) BLCC is controlled by node degree distributions (Figure 10).

There exist several other network properties that can be tunned by the parameters in our model. Such as an average distance between randomly selected pairs of nodes, the diameter of a bigraph, resilience to attack, spread of innovations or creation of the largest connected component. We omit the analysis of these features as they do not seem to have direct impact on the performance of the recommender systems.

5.3 Neighborhood size properties

The number of operations that a neighborhood recommender system has to perform is related to the number of similar users and the number of their items. We recommend a new item to analyzed user from the items of the users that are similar to her/him. In Figure 11 we show two intuitive results:

– the size of the neighborhood grows with the size of a graph

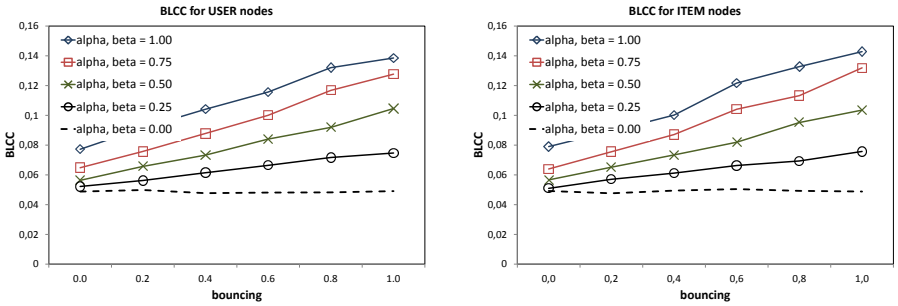

Fig. 9. The growth of the bouncing parameter b results in higher values of BLCC (bipartite local clustering coefficient). If no nodes are connected with accordance to the preferential attachment mechanism $\alpha = \beta = 0$, the values of b do not influence BLCC. Experiments run with ($m = 50$, $T = 10\ 000$, $p = 0.5$, $u = v = 7$).

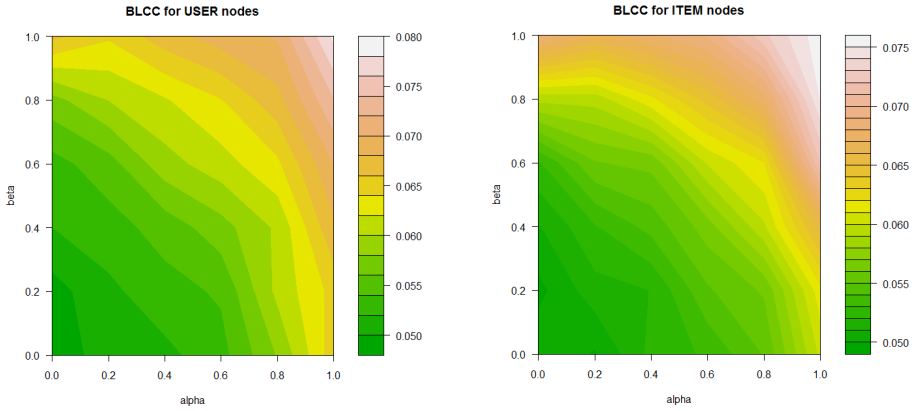

Fig. 10. BLCC growths as more edges are connected with preferential attachment mechanism. The phenomenon is observed even when the bouncing parameter is zero. Experiments run with ($m = 50$, $T = 10\ 000$, $p = 0.5$, $u = v = 7$, $b = 0$).

– the size of the neighborhood grows with the density of a graph (fixed number of nodes and growing number of edges)

The growth of the neighborhood is relatively sharper in case of the number of items. It is interesting that the number of similar users becomes stable earlier for sparser graphs (3 and 6 edges at startup) than for denser graphs (12 and 24 edges at startup).

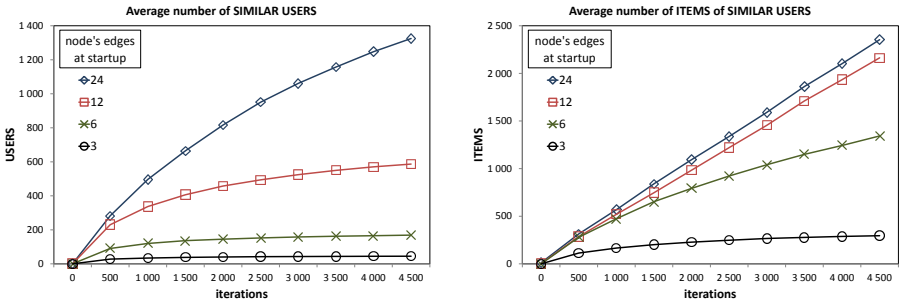

Fig. 11. An average number of similar users (having at least one common item with a considered user) follows the growth in a graph's size. The positive relation is stronger in case of the number of the items of the similar users. The density of a graph (modeled by the number of startup edges) has even stronger impact on the size of the neighborhood than the size of a graph. Experiments run with ($m = 50$, $T = 10\ 000$, $p = 0.5$, $\alpha = \beta = 0.5$, $b = 0$).

A result of potentially great importance is drawn in Figure 12. It turns out that the impact of the shapes of node degree distributions (controlled by parameters α and β) on the sizes of the neighborhoods is not monotonic. It turns out that the more exponential like than power-law like the distribution of users' degrees the smaller number of similar users is observed. In all other cases the opposite force is identified.

The result presented in Figure 13 is somewhat disappointing. The shrinking impact of the bouncing mechanism on the sizes of the neighborhoods is hardly observed. The effect of bouncing is too gentle compared to the level at which we are placed by the power-law distribution. Also random changes among various networks are stronger at the level than the shrinking forces. This drawback reflects the fact that in growing random graphs positive clustering coefficient is correlated with power-law node degree distribution and we are unable to generate graphs with both the exponential node degree distribution and high value of the clustering.

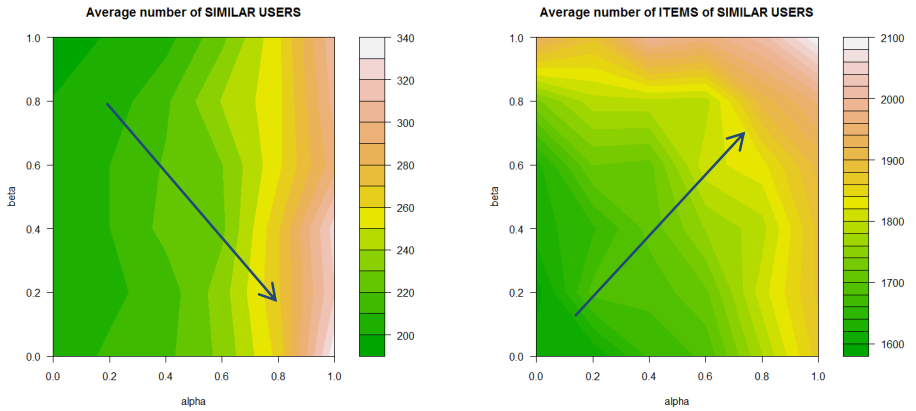

Fig. 12. The shape of node degree distributions of both modalities has opposite influence on the average number of similar users. The more power-law like item degree distribution, the more neighbors can be observed. The more heavy-tailed the distribution of user nodes the stronger shrinking of the neighborhood is obtained. The arrows indicate the direction of growth. Experiments run with ($m = 50$, $T = 10\ 000$, $p = 0.5$, $u = v = 7$, $b = 0$).

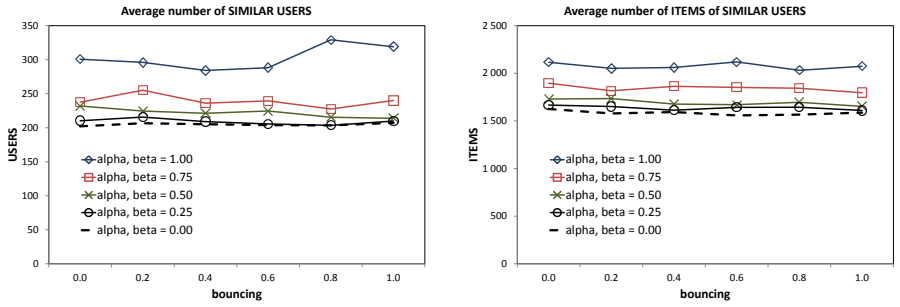

Fig. 13. The growth of the bouncing parameter b has slight negative impact of the size of both neighborhoods. However, the number of similar users and their items is determined mostly by the shapes of node degree distributions.

6 Conclusion

We have presented a new random graph generative algorithm dedicated to modeling performance of recommender systems. We have shown that the parameters of the algorithms influence not only pure network properties of created bigraphs, but also the properties related to the performance of neighborhood based collaborative filtering systems. Besides of the above features, the procedure enables us to output bigraphs of different sizes, densities and the proportions of the number of users to the number of items. We plan to compare how various features of bigraphs impinge on time and memory requirements of existing systems. Consequently, better understand the algorithms, their implementations and finally improve both of them.

Acknowledgments. This work was partially supported by Polish state budget funds for scientific research within research project *Analysis and visualization of structure and dynamics of social networks using nature inspired methods*, grant No. N516 443038.

A Degree of a neighboring node

In this appendix we derive the expected degree of a *neighboring node* in a random graph (Figure 14). Let's denote by $\langle k \rangle$ and $\langle k^2 \rangle$ the first and the second moments of the node degree distribution of graph $G = (V, E)$.

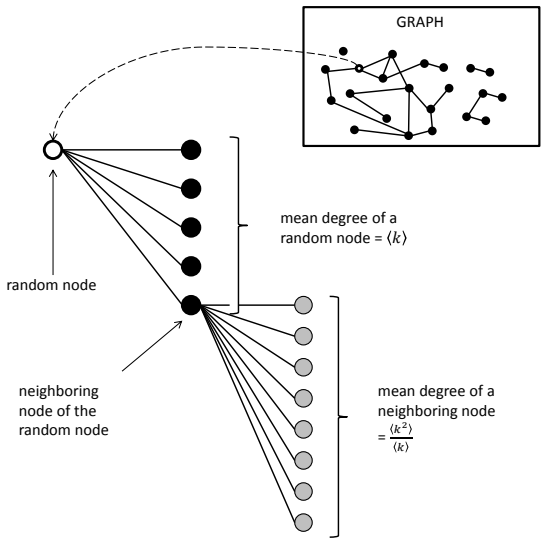

Fig. 14. The expected degree of a neighbor of randomly selected node is larger than an average node degree.

If we pick a random node from a graph then its expected number of neighbors (degree) is $\langle k \rangle$. Each of $\langle k \rangle$ edges points at a different vertex. The probability that a random edge is connected to a node is proportional to the total number of edges that are connected with the node. The probability that a random edge is connected to a node i with degree k_i is equal to $\frac{k_i}{\sum_{j \in V} k_j}$. Hence, the expected degree of a neighboring node is:

$$\sum_{i \in V} k_i \frac{k_i}{\sum_{j \in V} k_j} = \frac{\sum_{i \in V} k_i^2}{\sum_{j \in V} k_j} = \frac{\langle k^2 \rangle}{\langle k \rangle}. \tag{8}$$

The analysis is based on an assumption that there exist no correlation between the degrees of two neighboring nodes.

We can show that this value is not smaller than $\langle k \rangle$ i.e. an expected degree of a random node. Let us recall the Cauchy-Schwartz inequality:

$$\left(\sum_{i=1}^{n} x_i y_i\right)^2 \leq \left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} y_i^2\right). \tag{9}$$

By putting $x_i = 1$ for $i = 1, \ldots, n$, we get:

$$\left(\sum_{i=1}^{n} y_i\right)^2 \leq n\left(\sum_{i=1}^{n} y_i^2\right), \tag{10}$$

and

$$\frac{\sum_{i=1}^{n} y_i}{n} \leq \frac{\left(\sum_{i=1}^{n} y_i^2\right)/n}{\left(\sum_{i=1}^{n} y_i\right)/n} \Rightarrow \langle y \rangle \leq \frac{\langle y^2 \rangle}{\langle y \rangle}. \tag{11}$$

B Node degree distribution

We follow *continuum approach* [3] to derive user node degree distribution. The item node degree distribution can be obtained analogously. The calculations consist of three steps. Firstly, let's solve Eq. (6).

$$\begin{aligned}\frac{\partial k_j}{\partial t} &= (1-p)v\Pi(k_j) \\ &= (1-p)v\left(\frac{\beta}{pt} + \frac{(1-\beta)k_j}{\eta t}\right) \\ &= (1-p)v\frac{1}{t}\left(\frac{\beta\eta + p(1-\beta)k_j}{p\eta}\right),\end{aligned}$$

which yields

$$\int \frac{1}{(1-p)v} \cdot \frac{p\eta}{\beta\eta + p(1-\beta)k_j}\, dk_j = \int \frac{1}{t}\, dt. \tag{12}$$

Taking into account an initial condition $k_j(t_j) = u$, where t_j is the time of creating user j, and the fact that $\int \frac{c}{ax+b}dx = \frac{c}{a}\ln|ax+b| + C$ we obtain

$$\frac{p\eta}{(1-p)vp(1-\beta)}\left(\left[\ln\left(\beta\eta + p(1-\beta)k_j\right)\right] - \left[\ln\left(\beta\eta + p(1-\beta)u\right)\right]\right) = \left[\ln t\right] - \left[\ln t_j\right], \tag{13}$$

both sides of which can be used as exponents of e, giving

$$\left(\frac{\beta\eta + p(1-\beta)k_j}{\beta\eta + p(1-\beta)u}\right)^{\frac{\eta}{(1-p)(1-\beta)v}} = \left(\frac{t}{t_j}\right), \tag{14}$$

after reorganizing, we have

$$k_j(t) = \frac{1}{p(1-\beta)} \cdot \left((\beta\eta + p(1-\beta)u) \left(\frac{t}{t_j}\right)^{\frac{(1-p)(1-\beta)i}{\eta}} - \beta\eta \right). \tag{15}$$

The probability that k_j is smaller then a given k is:

$$\Phi\left\{k_j(t) < k\right\} = \Phi\left\{ \frac{(\beta\eta + p(1-\beta)u)\left(\frac{t}{t_j}\right)^{\frac{(1-p)(1-\beta)v}{\eta}} - \beta\eta}{p(1-\beta)} < k \right\}, \tag{16}$$

and after reorganizing

$$\Phi\left\{k_j(t) < k\right\} = \Phi\left\{ t_j > t \left(\frac{\beta\eta + p(1-\beta)k}{\beta\eta + p(1-\beta)u}\right)^{\frac{-\eta}{(1-p)(1-\beta)v}} \right\}. \tag{17}$$

We can assume that nodes are added at equal time intervals until the current iteration t. The probability the iteration of adding node j is larger than some $K \leq t$ equals $1 - \Phi(t_j \leq K) = 1 - K\frac{1}{t}$. Substituting this assumption into Eq. (17), we obtain

$$\begin{aligned}\Phi\left\{k_j(t) < k\right\} &= 1 - \Phi\left\{ t_j \leq t \left(\frac{\beta\eta + p(1-\beta)k}{\beta\eta + p(1-\beta)u}\right)^{\frac{-\eta}{(1-p)(1-\beta)v}} \right\} \\ &= 1 - \left(\frac{\beta\eta + p(1-\beta)k}{\beta\eta + p(1-\beta)u}\right)^{\frac{-\eta}{(1-p)(1-\beta)v}}.\end{aligned}$$

We can obtain probability density function of random variable k by differentiating its cumulative distribution function $P(k) = \partial\Phi\{k_j(t) < k\}/\partial k$, as a result we have

$$P(k) = \frac{\eta}{(1-p)(1-\beta)v} \cdot p(1-\beta) \cdot \left(\frac{\beta\eta + p(1-\beta)k}{\beta\eta + p(1-\beta)u}\right)^{\frac{-\eta}{(1-p)(1-\beta)v}-1}, \tag{18}$$

that is:

$$P(k) \propto \left(\frac{\beta\eta + p(1-\beta)k}{\beta\eta + p(1-\beta)u}\right)^{\frac{-\eta}{(1-p)(1-\beta)v}-1}. \tag{19}$$

References

[1] M. Newman, *Networks: An Introduction*. Oxford University Press, 2010.

[2] P. Erdős and A. Rényi, "On the evolution of random graphs," in *Publication of the Mathematical Institute of the Hungarian Academy of Sciences*, pp. 17–61, 1960.

[3] A. Barabási and R. Albert, "Emergence of scaling in random networks," *Science (New York, N.Y.)*, vol. 286, no. 5439, pp. 509–512, 1999.

[4] R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, and E. Upfal, "Stochastic models for the web graph," in *Proceedings of the 41st Annual Symposium on Foundations of Computer Science (FOCS)*, (Redondo Beach, CA, USA), pp. 57–65, IEEE CS Press, 2000.

[5] Z. Liu, Y.-C. Lai, N. Ye, and P. Dasgupta, "Connectivity distribution and attack tolerance of general networks with both preferential and random attachments," *Physics Letters A*, vol. 303, no. 5-6, pp. 337 – 344, 2002.

[6] A. Vázquez, "Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations," *Phys. Rev. E*, vol. 67, p. 056104, May 2003.

[7] E. Zheleva, H. Sharara, and L. Getoor, "Co-evolution of social and affiliation networks.," in *KDD* (J. F. E. IV, F. Fogelman-Soulié, P. A. Flach, and M. Zaki, eds.), pp. 1007–1016, ACM, 2009.

[8] J.-L. Guillaume and M. Latapy, "Bipartite structure of all complex networks.," *Inf. Process. Lett.*, vol. 90, no. 5, pp. 215–221, 2004.

[9] S. Lattanzi and D. Sivakumar, "Affiliation networks," in *STOC '09: Proceedings of the 41st annual ACM symposium on Theory of computing*, (New York, NY, USA), pp. 427–434, ACM, 2009.

[10] S. Chojnacki and M. Kłopotek, "Power-law node degree distribution in online affiliation networks," in *KKNTPD'10: III Krajowa Konferencja Naukowa Technologie Przetwarzania Danych*, pp. 71–79, WNT, 2010.

[11] F. Eisterlehner, A. Hotho, and R. Jäschke, eds., *ECML PKDD Discovery Challenge 2009 (DC09)*, vol. 497 of *CEUR-WS.org*, Sept. 2009.

[12] "Internet movie database." `http://www.imdb.com`.

[13] "Citeulike bookmarking portal." `http://www.citeulike.org`.

[14] F. Vega-Redondo, *Complex Social Networks*. Cambridge University Press, 2007.

[15] M. Newman, S. Strogatz, and D. J. Watts, "Random graphs with arbitrary degree distributions and their applications," vol. 64, July 2001.

[16] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*. Cambridge University Press, 1994.

[17] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and Analysis of Online Social Networks," in *Proceedings of the 5th ACM/USENIX Internet Measurement Conference (IMC'07)*, (San Diego, CA), October 2007.

[18] "The netflix challenge." `http://www.netflixprize.com`.

[19] "The ecml discovery challenge 2009." `http://www.kde.cs.uni-kassel.de/ws/dc09`.

[20] S. Chojnacki, K. Ciesielski, and M. Klopotek, "Node degree distribution in affiliation graphs for social network density modelling," in *Lecture Notes in Computer Science*, 2010.

[21] Z. Burda, J. Duda, J. M. Luck, and B. Waclaw, "Localization of the maximal entropy random walk," *Phys. Rev. Lett.*, vol. 102, p. 160602, Apr 2009.