Preferential Placement for Community Structure Formation

Aleksandr Dorodnykh[1], Liudmila Ostroumova Prokhorenkova[1,2(✉)], and Egor Samosvat[1,2]

[1] Moscow Institute of Physics and Technology, Moscow, Russia
[2] Yandex, Moscow, Russia
Ostroumova-la@yandex.ru

Abstract. Various models have been recently proposed to reflect and predict different properties of complex networks. However, the community structure, which is one of the most important properties, is not well studied and modeled. In this paper, we suggest a principle called "preferential placement", which allows to model a realistic community structure. We provide an extensive empirical analysis of the obtained structure as well as some theoretical heuristics.

1 Introduction

The evolution of complex networks attracted a lot of attention in recent years. Empirical studies of different real-world networks have shown that such structures have some typical properties: small diameter, power-law degree distribution, clustering structure, and others [9,15,35]. Therefore, numerous random graph models have been proposed to reflect and predict such quantitative and topological aspects of growing real-world networks [9,11,15,38,41].

The most extensively studied property of complex networks is their vertex degree distribution. For the majority of studied real-world networks, the portion of vertices of degree d was observed to decrease as $d^{-\gamma}$, usually with $2 < \gamma < 3$ [5, 18,36]. Such networks are often called scale-free. The most well-known approach to the modeling of scale-free networks is called *preferential attachment*. The main idea of this approach is that new vertices emerging in a graph connect to some already existing vertices chosen with probabilities proportional to their degrees. Preferential attachment is a natural process allowing to obtain a graph with a power-law degree distribution, and many random graph models are based on this idea, see, e.g., [10,13,23,26,45].

Another important characteristic of complex networks is their community (or clustering) structure, i.e., the presence of densely interconnected sets of vertices, which are usually called clusters or communities [19,21]. Several empirical studies have shown that community structure of different real-world networks has some typical properties. In particular, it was observed that the cumulative community size distribution obeys a power law with some parameter λ. For instance, [14] reports that $\lambda = 1$ for some networks; [3] obtains either $\lambda = 0.5$ or $\lambda = 1$; [22]

A. Bonato et al. (Eds.): WAW 2017, LNCS 10519, pp. 75–89, 2017.
DOI: 10.1007/978-3-319-67810-8_6

also observes a power law with λ close to 0.5 in some range of cluster sizes; [39] studies the overlapping communities and shows that λ is ranging between 1 and 1.6.

Community structure is an essential property of complex networks. For example, it highly affects the spreading of infectious diseases in social networks [24,29], spread of viruses over computer networks [43], promotion of products via viral marketing [25], propagation of information [42], etc. Therefore, it is crucial to be able to model realistic community structures.

Nowadays, there are a few random graph models allowing to obtain clustering structures. Probably the most well-known model was suggested in [28] as a benchmark for comparing community detection algorithms. In this model, the distributions of both degrees and community sizes follow power laws with predetermined exponents. However, there are two drawbacks of this model. First, it does not explain the power-law distribution of community sizes, these sizes are just sampled from a power-law distribution at the beginning of the process. Second, a subgraph induced by each community is very similar to the configuration model [8], which does not allow to model, e.g., hierarchical community structure often observed in real-world networks [3,14].

A weighted model which naturally generates communities was proposed in [27]. However, the community structure in this model is not analyzed in details and only the local clustering coefficient is shown. From the figures presented in [27] it seems that the community size distribution does not have a heavy tail as it is observed in real-world complex networks.

Finally, let us mention a paper [40] which analyzes a community graph, where vertices refer to communities and edges correspond to shared members between the communities. The authors show that the development of the community graph seems to be driven by preferential attachment. They also introduce a model for the dynamics of overlapping communities. Note that [40] only models the membership of vertices and does not model the underlying network.

In this paper, we propose a process which naturally generates clustering structure. Our approach is called *preferential placement* and it is based on the idea that vertices can be embedded in a multidimensional space of latent features. The vertices appear one by one and their positions are defined according to preferential placement: new vertices are more likely to fall into already dense regions. We present a detailed description of this process in Sect. 2. After n steps we obtain a set of n vertices placed in a multidimensional space. In Sect. 3 we empirically analyze the obtained structure: in particular, we show that the communities are clearly visible and their sizes are distributed according to a power law. Note that after the placement of all vertices is defined, one can easily construct an underlying network, using, e.g., the threshold model [12,31]. We discuss possible models and their properties in Sect. 4.

2 Preferential Placement

In this section, we describe the proposed approach which we call *preferential placement*. We assume that all vertices are embedded in $\mathbb{R}^d$ for some $d \geq 1$.

One can think that coordinates of this space correspond to latent features of vertices. Introducing latent features has recently become a popular approach both in predictive and generative models. These models are known by different names such as latent feature models [33,34], matrix factorization models [4,16,32], spatial models [2,6,7], or geographical models [12,31]. The basic idea behind all these models is that vertices having similar latent features are more likely to be connected by an edge.

Preferential placement is the procedure describing the embedding of vertices in the space $\mathbb{R}^d$. After that, given the coordinates of all vertices, one can construct a graph using one of many well-known approaches (see Sect. 4 for the discussion of possible variants).

Our model is parametrized by a distribution Ξ taking nonnegative values. The proper choice of Ξ is discussed further in this section.

We construct a random configuration of vertices (or points) $S_n = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, where $\mathbf{x}_i = (x_i^1, \ldots, x_i^d)$ denotes the coordinates of the i-th vertex v_i. Let $S_1 = \{\mathbf{x}_1\}$, $\mathbf{x}_1$ is the origin. Now assume that we have constructed S_t for $t \geq 1$, then we obtain S_{t+1} by adding a vertex v_{t+1} with the coordinates $\mathbf{x}_{t+1}$ chosen in the following way:

- Choose a vertex $v_{i_{t+1}}$ from $v_1, \ldots, v_t$ uniformly at random.
- Sample ξ_{t+1} from the distribution Ξ.
- Sample a direction $\mathbf{e}_{t+1}$ from a uniform distribution on a multidimensional sphere $\|\mathbf{e}_{t+1}\|_2 = 1$, where $\|\cdot\|_2$ denotes the Euclidean distance in $\mathbb{R}^d$.
- Set $\mathbf{x}_{t+1} = \mathbf{x}_{i_{t+1}} + \xi_{t+1} \cdot \mathbf{e}_{t+1}$.

We argue in this paper that in order to obtain a realistic clustering structure one should take Ξ to be a heavy tailed distribution. In this case, according to the procedure described above, new vertices will usually appear in the dense regions, close to some previously added vertices; however, due to the heavy tail of Ξ, from time to time we get outliers, which originate new clusters.

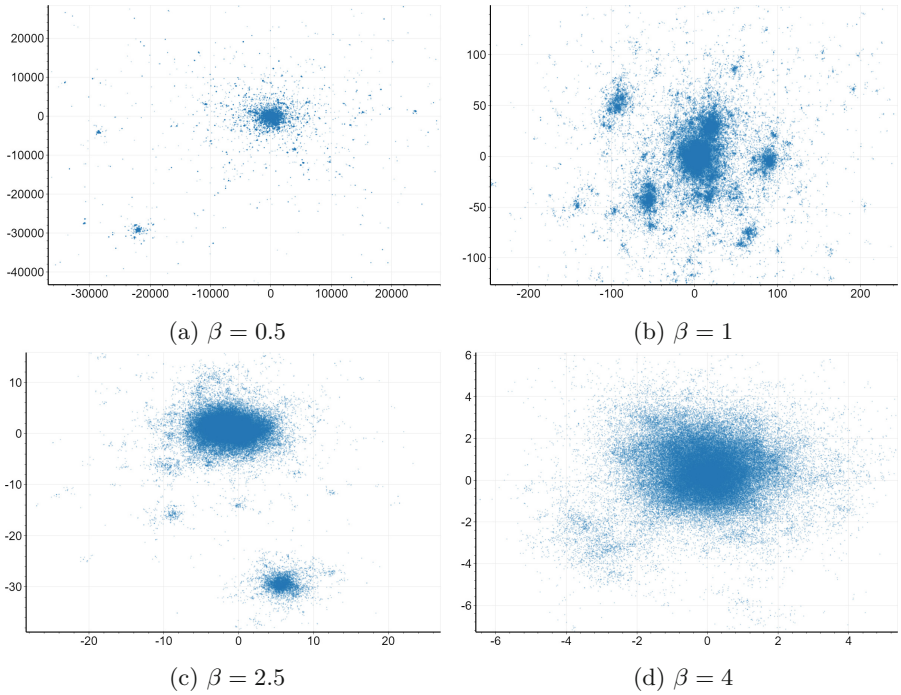
We call the described above procedure "preferential placement" due to its analogy with preferential attachment. Assume that at some step of the algorithm we have several clusters, i.e., groups of vertices located close to each other, and a new vertex appears. Then the probability that this vertex will join a cluster C is roughly proportional to its size, i.e., the number of vertices already belonging to this cluster. This is the basic intuition which we discuss further in this paper in more details.

3 Analysis of Preferential Placement

3.1 Experimental Setup

In this section, we analyze graphs obtained using the preferential placement procedure described above. We take Ξ to be a slightly modified Pareto distribution with the density function $f_\beta(x) = \frac{\beta}{(x+1)^{\beta+1}}, x \geq 0$ for fixed $\beta > 0$.

In all the experiments we take $d = 2$ since the obtained structures are easy to visualize. However, we also tried other values of $d \geq 1$ and obtained results



(a) $\beta = 0.5$ (b) $\beta = 1$

(c) $\beta = 2.5$ (d) $\beta = 4$

Fig. 1. Clustering structure depending on Ξ

similar to shown on Figs. 4 and 6. Also, if not specified otherwise, we generated structures with the number of points $n = 100\,\mathrm{K}$.

3.2 Clustering Structure Depending on Ξ

First, let us visualize the structures obtained by our algorithm. We tried several values of β, $\beta \in \{0.5, 1, 1.5, 2.5, 4\}$. The results are presented on Figs. 1 and 2. The value $\beta = 0.5$ produces the heaviest tail, in this case the distribution Ξ does not have a finite expectation. Although some clusters are clearly visible in this case, they are located far apart from each other, which seems to be not very realistic. Graphs obtained from configurations (using one of the procedures discussed in Sect. 4) are expected to have small diameter and giant connected component of size $\Theta(n)$, which does not seem to be the case for $\beta = 0.5$. Note that for too large β, e.g., for $\beta = 4$, the variance is too low and we obtain only one giant cluster with minor fluctuations, as presented on Fig. 1d. Further in this paper we discuss the case $\beta = 1.5$ presented on Fig. 2. In this case Ξ has a finite expectation but an infinite variance.

Another interesting observation is a hierarchical clustering structure produced by our algorithm. To illustrate this, we take the figure obtained for $\beta = 1.5$ and zoom it to see more details. Figure 2 shows that the largest cluster further consists of several sub-clusters.

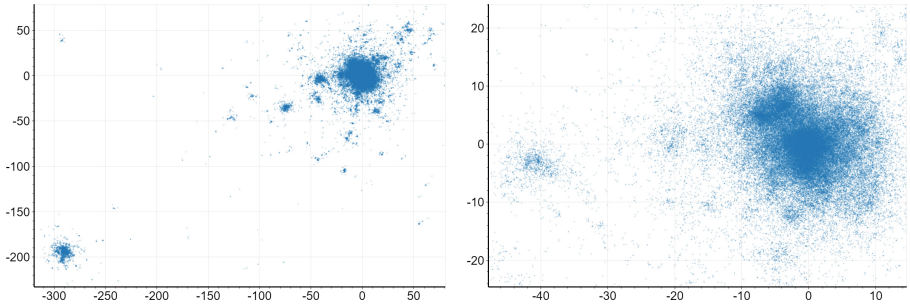


Fig. 2. $\beta = 1.5$, different scales

3.3 The Distribution of Cluster Sizes

In this section, we analyze the distribution of cluster sizes produced by preferential placement. We present both theoretical and empirical observations.

The main difficulty with the analysis of clustering structure is the fact that there are no standard definitions of clusters, both in graphs and metric spaces. For example, clusters are often defined as a result of some clustering algorithm.[1] This causes a lot of difficulties for both theoretical and empirical analysis.

Theoretical Heuristics. First, let us discuss why we expect to observe a power-law distribution of cluster sizes in our model. As we discussed above, due to the absence of a rigorous definition of a cluster, further in this section we are able to present only some heuristic theory.

Let $F_t(s)$ denote the number of clusters of size s at step t. In order to analyze $F_t(s)$ we consider its dynamics inductively. Assume that after a step t we obtain some clustering structure. At step $t+1$ we add a vertex v_{t+1} and choose its "parent" $v_{i_{t+1}}$ from $v_1, \ldots, v_t$ uniformly at random. Clearly, the probability to choose a parent from some cluster C with $|C| = s$ is equal to $\frac{s}{t}$. In this case, we call C a parent cluster for v_{t+1}. Now let us make the following assumptions:

1. All clusters can only grow, they cannot merge or split.
2. At step $t+1$ a new cluster appears with probability $p(t) = \frac{c}{t^\alpha}$, $c > 0$, $0 \leq \alpha \leq 1$.
3. Given that a vertex $t+1$ does not create a new cluster, the probability to join a cluster C with $|C| = s$ is equal to $\frac{s}{t}$.

These assumptions are quite strong and even not very realistic. For instance, it seems reasonable that two clusters can merge if many vertices appear somewhere between them. Regarding the second assumption, $p(t)$ can possibly depend

[1] *Modularity*, introduced in [37], can be used to define communities in graphs. However, this characteristic has certain drawbacks, as discussed in [20]. Moreover, modularity favors partitions with approximately equal communities, which contradicts the main idea of power-law distribution of community sizes.

on the current configuration S_t. However, these assumptions allow us to analyze the behavior of $F_t(s)$ formally. Namely, we prove the following theorem.

Theorem 1. *Under the assumptions described above the following holds.*

1. *If* $\alpha = 0$, *then*

$$\mathrm{E}F_n(s) = \frac{c\,(s-1)!\,\Gamma\left(2+\frac{1}{1-c}\right)}{(2-c)\Gamma\left(s+1+\frac{1}{1-c}\right)}\left(n + O\left(s^{\frac{1}{1-c}}\right)\right) \sim \frac{c\,\Gamma\left(2+\frac{1}{1-c}\right)}{(2-c)} \cdot \frac{n}{s^{1+\frac{1}{1-c}}}.$$

2. *If* $0 < \alpha \leq 1$, *then for any* $\epsilon > 0$

$$\mathrm{E}F_n(s) = \frac{c\,(s-1)!\,\Gamma(3-\alpha)}{(2-\alpha)\Gamma(s+2-\alpha)}\left(n^{1-\alpha} + O\left(n^{\max\{0,1-2\alpha\}} s^{1-\alpha+\epsilon}\right)\right) \sim \frac{c\,\Gamma(3-\alpha)}{2-\alpha} \cdot \frac{n^{1-\alpha}}{s^{2-\alpha}}.$$

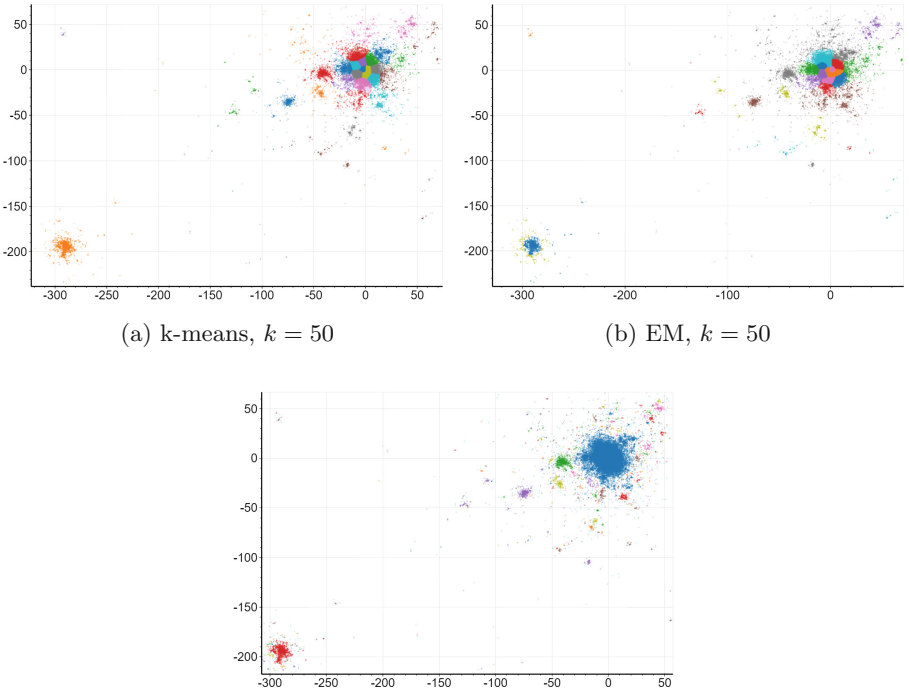
To sum up, if the probability $p(n)$ of creating a new cluster is of order $\frac{1}{n^\alpha}$ for $\alpha > 0$, then the distribution of cluster sizes follows a power law with parameter $2-\alpha$ growing with $p(n)$ from 1 to 2; if $p(n) = c$, $0 < c < 1$, then the parameter grows with c from 2 to infinity. Recall that the parameter of the cumulative distribution is one less than discussed above. The proof of Theorem 1 is technical and we place it to Appendix.

Let us also explain why we do not consider $p(n)$ decreasing faster than $\frac{c}{n}$. It is natural to assume that a new cluster appears if a new vertex chooses a parent node near the border of some cluster and then ξ_{t+1} and $\mathbf{e}_{t+1}$ are chosen such that $\mathbf{x}_{t+1} = \mathbf{x}_{i_{t+1}} + \xi_{t+1} \cdot \mathbf{e}_{t+1}$ falls quite away from the parent cluster. This probability is roughly proportional to the number of vertices located near the borders of the clusters. Extreme case, 1 vertex, provides the bound $\frac{c}{n}$.

Finally, let us mention that in practice the probability $p(n)$ of creating a new cluster can depend not only on Ξ, but also on the definition of clusters. Further in this section we demonstrate that parameters of a clustering algorithm can affect the parameter of the obtained power law.

Empirical Analysis. As we already mentioned, there is no standard definition of a clustering structure. In many cases, clusters and communities are defined just as a result of some clustering algorithm. Therefore, we first analyze the performance of several clustering algorithms, then choose the most appropriate one and analyze clusters it produces.

We compare the following algorithms: k-means [30], EM (expectation maximization), and DBSCAN (density-based spatial clustering of applications with noise) [17]. For k-means and EM one has to specify the number of clusters.



(a) k-means, $k = 50$

(b) EM, $k = 50$

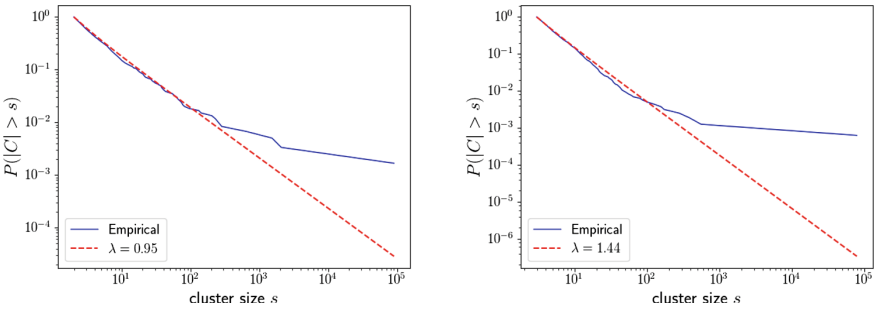
(c) DBSCAN, $L = 125$, $k = 1$

Fig. 3. The comparison of different clustering algorithms

We tried several values of k, $k \in \{10, 50, 100, 500, 1000\}$, but both algorithms turned out to be not suitable for out problem. As expected, in all cases they unnaturally split the largest cluster into several small ones (see Figs. 3a and b).

On the contrary, DBSCAN produces more realistic results. It requires two parameters: radius of neighborhood ε and the minimum number of neighbors required to form a dense region k. We consider $k \in \{1, 2, 3\}$ and ε is chosen in such a way that if we connect all vertices i, j such that $\|i - j\|_2 < \varepsilon$, then we get Ln edges, $L \in \{5, 25, 125\}$, where n is the number of vertices. For all parameters we get reasonable clustering structures. The result for $L = 125, k = 1$ is presented on Fig. 3c. For these parameters we also analyze the distribution of cluster sizes (see Fig. 4a). Note that for not too large values of s ($s < 300$) the cumulative distribution follows a power law with parameter $\lambda \approx 0.95$. In Theorem 1 this value corresponds to the case $\alpha = 0.05$, i.e., $p(n) \propto n^{-0.05}$. Based on this, we expect the number of clusters to grow as $n^{0.95}$, i.e., close to linearly. On Fig. 5 we plot the empirical number of clusters and fit it by $n^{0.95}$.

Finally, as we promised above, we show that λ can depend on the clustering algorithm. Figure 4b shows the cumulative cluster size distribution for DBSCAN with $L = 5, k = 1$. Note that $\lambda = 1.44$, so it is larger in this case. Intuitively, the reason is that $p(n)$ is larger for $L = 5$ than for $L = 125$. Smaller values of L



(a) DBSCAN with $L = 125, k = 1$ (b) DBSCAN with $L = 5, k = 1$

Fig. 4. Cluster size distribution

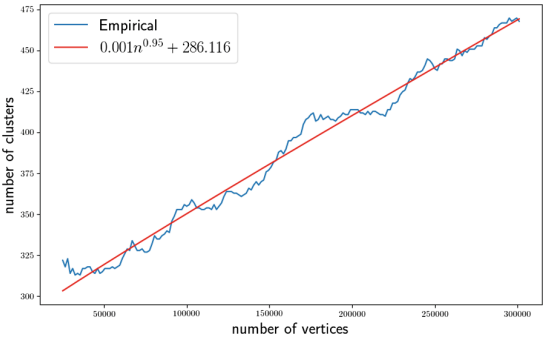


Fig. 5. Growth of the number of clusters, DBSCAN with $L = 125, k = 1$

correspond to smaller ε, which means that it is harder for a new vertex to join some existing cluster, which makes $p(n)$ larger.

4 Graph Models

4.1 Possible Definitions

In this section, we discuss how a graph can be constructed based on the vertices embedding produced by the preferential placement procedure.

The basic idea behind many known spatial models is that we want to increase the probability of connecting two vertices if they have similar latent features. Various methods can be found in the literature, which are usually combined with some other ideas like introducing weights of vertices or taking into account degrees of vertices (see, e.g., a survey of spatial models in [7]). We now briefly describe some possible approaches:

– *threshold model* [12,31]:

$$\mathrm{P}\big((v_i, v_j) \in E\big) = I\big[\|\mathbf{x}_i - \mathbf{x}_j\|_2 \leq \theta\big]\,;$$

– *p-threshold model*:

$$\mathrm{P}\big((v_i, v_j) \in E\big) = pI\big[\|\mathbf{x}_i - \mathbf{x}_j\|_2 \le \theta\big], \quad 0 < p < 1;$$

– *p-threshold model with random edges* (as in spatial small-world models [7]):

$$\mathrm{P}\big((v_i, v_j) \in E\big) = p_0 + p_1 I\big[\|\mathbf{x}_i - \mathbf{x}_j\|_2 \le \theta\big], \quad 0 < p_0, p_1, p_0 + p_1 < 1;$$

– *inverted distance model*:

$$\mathrm{P}\big((v_i, v_j) \in E\big) \propto \frac{1}{\|\mathbf{x}_i - \mathbf{x}_j\|_2};$$

– *Waxman model* [44]:

$$P\big((v_i, v_j) \in E\big) \propto e^{-\|\mathbf{x}_i - \mathbf{x}_j\|_2 / d}.$$

Here we denote by E the set of edges. We assume that all edges are mutually independent, hence to describe a random graph it is enough to define the probability of each edge. Further we focus on the threshold model, however, we expect similar results for other models.

4.2 Degree Distribution

In this section, we empirically analyze the degree distribution for the threshold model. As before, we take Ξ to be a distribution with the density function $f_\beta(x) = \frac{\beta}{(x+1)^{\beta+1}}, x \ge 0$ for $\beta = 1.5$. We choose θ such that we have $5n$ edges in our graph. The cumulative degree distribution for this case is presented on Fig. 6. Observe that the cumulative degree distribution does not follow a power law. However, it is very similar to degree distributions obtained in many real-world networks (numerous examples can be found in [1]).

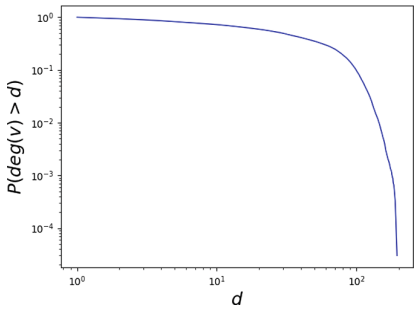


Fig. 6. Cumulative degree distribution for the threshold model

We are currently working on theoretical analysis of the degree distribution in the threshold model. We plan to add these results, together with the empirical analysis of other properties like diameter and clustering coefficient, to the extended version of this paper.

5 Conclusion and Future Work

In this paper, we introduced a principle called *preferential placement*. Our method is designed to model a realistic clustering structure. The algorithm is parametrized only by a distribution Ξ, and if Ξ is a Pareto distribution, which is the most natural choice, then we essentially have only one parameter — the exponent β. The proposed algorithm naturally models clusters and the distribution of cluster sizes follows a power law, which is a desirable property. Although preferential placement only generates the coordinates of vertices, one can easily construct a graph based on the obtained structure using one of the methods discussed in this paper. We showed that applying a threshold model to the configuration generated by preferential placement leads to a realistic degree distribution.

In this paper, we made only a first step to understanding the cluster formation in complex structures and there are many directions for future research. First of all, more formal analysis of the distribution of cluster sizes would be useful. As we discussed, the main problem here is the lack of any suitable formal definition of clusters. However, one can try, e.g., to analyze clusters produced by one of well-known clustering algorithms. Second direction is the analysis of the obtained graphs. We are currently working on theoretical analysis of the degree distribution in the threshold model. We also plan to analyze other properties, like diameter and clustering coefficient.

Acknowledgements. This work is supported by Russian President grant MK-527.2017.1.

Appendix

Proof of Theorem 1

First, recall the process of cluster formation:

- At the beginning of the process we have one vertex which forms one cluster.
- At n-th step with probability $p(n)$ a new cluster consisting of v_n is created.
- With probability $1 - p(n)$ new vertex joins already existing cluster C with probability proportional to $|C|$.

So, we can write the following equations:

$$\mathrm{E}(F_{t+1}(1)|S_t) = F_t(1)\left(1 - \frac{1-p(t)}{t}\right) + p(t)\,, \tag{1}$$

$$\mathrm{E}(F_{t+1}(s)|S_t) = F_t(s)\left(1 - \frac{s(1-p(t))}{t}\right) + F_t(s-1)\frac{(s-1)(1-p(t))}{t}\,, \quad s > 1\,. \tag{2}$$

Now we can take expectations of the both sides of the above equations and analyze the behavior of $\mathrm{E}F_t(s)$ inductively.

Consider the case $\alpha = 0$, i.e., $p(n) = c$. Let us prove that in this case

$$\mathrm{E}F_n(s) = \frac{c(s-1)!\,\Gamma\left(2+\frac{1}{1-c}\right)}{(2-c)\Gamma\left(s+1+\frac{1}{1-c}\right)}\,(n+\theta_{n,s})\ . \tag{3}$$

where $\theta_{n,s} \le C\, s^{\frac{1}{1-c}}$ for some constant $C > 0$.

We prove this result by induction on s and for each s the proof is by induction on n. Note that for $n = 1$ Eq. (3) holds for all s. Consider now the case $s = 1$. We want to prove that

$$\mathrm{E}F_n(1) = \frac{c}{2-c}\,(n+\theta_{n,1})\ .$$

For the inductive step we use Eq. (1) and get

$$\begin{aligned}\mathrm{E}(F_{t+1}(1)) = \mathrm{E}F_t(1)\left(1-\frac{1-c}{t}\right)+c &= \frac{c}{2-c}\,(t+\theta_{t,1})\left(1-\frac{1-c}{t}\right)+c\\ &= \frac{c}{2-c}\left(t+1+\theta_{t,1}\left(1-\frac{1-c}{t}\right)\right)\ .\end{aligned}$$

Since

$$C\left(1-\frac{1-c}{t}\right) \le C,$$

this finishes the proof for $\alpha = 0$ and $s = 1$.

For $s > 1$ we use Eq. (2) and get

$$\begin{aligned}\mathrm{E}(F_{t+1}(s)) &= \mathrm{E}F_t(s)\left(1-\frac{s\,(1-c)}{t}\right)+\mathrm{E}F_t(s-1)\frac{(s-1)\,(1-c)}{t}\\ &= \frac{c(s-1)!\,\Gamma\left(2+\frac{1}{1-c}\right)(t+\theta_{t,s})}{(2-c)\,\Gamma\left(s+1+\frac{1}{1-c}\right)}\left(1-\frac{s(1-c)}{t}\right)\\ &\quad+\frac{c(s-1)!\,\Gamma\left(2+\frac{1}{1-c}\right)(1-c)(t+\theta_{t,s-1})}{(2-c)\,\Gamma\left(s+\frac{1}{1-c}\right)t}\\ &= \frac{c(s-1)!\,\Gamma\left(2+\frac{1}{1-c}\right)}{(2-c)\,\Gamma\left(s+1+\frac{1}{1-c}\right)}\left(t+1+\theta_{t,s}\left(1-\frac{s(1-c)}{t}\right)+\theta_{t,s-1}\frac{s(1-c)+1}{t}\right)\ .\end{aligned}$$

To finish the proof we need to show that

$$(s-1)^{\frac{1}{1-c}}\,\frac{s(1-c)+1}{t} \le s^{\frac{1}{1-c}}\,\frac{s(1-c)}{t}\ .$$

It is easy to show that the above inequality holds.

Now we consider the case $p(n) = cn^{-\alpha}$ for $0 < \alpha \leq 1$. Let us prove that in this case

$$\mathrm{E}F_n(s) = \frac{c(s-1)!\,\Gamma(3-\alpha)}{(2-\alpha)\Gamma(s+2-\alpha)}\left(n^{1-\alpha} + \theta_{n,s}\right),$$

where $\theta_{n,s} \leq Cn^{\max\{0,1-2\alpha\}}s^{1-\alpha+\epsilon}$ for some constant $C > 0$ and for any $\epsilon > 0$.

The proof is similar to the case $\alpha = 0$. Again, for $n = 1$ the theorem holds. Consider $s = 1$. We want to prove that

$$\mathrm{E}F_n(1) = \frac{c}{2-\alpha}\left(n^{1-\alpha} + \theta_{n,1}\right).$$

Inductive step in this case becomes

$$\begin{aligned}
\mathrm{E}(F_{t+1}(1)) &= \mathrm{E}F_t(1)\left(1 - \frac{1-ct^{-\alpha}}{t}\right) + ct^{-\alpha} \\
&= \frac{c}{2-\alpha}\left(t^{1-\alpha} + \theta_{t,1}\right)\left(1 - \frac{1-ct^{-\alpha}}{t}\right) + ct^{-\alpha} \\
&= \frac{c}{2-\alpha}\left(t^{1-\alpha} - t^{-\alpha} + c\,t^{-2\alpha} + (2-\alpha)t^{-\alpha} + \theta_{t,1}\left(1 - \frac{1-ct^{-\alpha}}{t}\right)\right) \\
&= \frac{c}{2-\alpha}\left((t+1)^{1-\alpha} + O\left(t^{-\alpha-1}\right) + c\,t^{-2\alpha} + \theta_{t,1}\left(1 - \frac{1-ct^{-\alpha}}{t}\right)\right).
\end{aligned}$$

In order to finish the proof for the case $s = 1$ it is sufficient to show that

$$O\left(t^{-\alpha-1}\right) + c\,t^{-2\alpha} \leq Ct^{\max\{0,1-2\alpha\}}\frac{1-ct^{-\alpha}}{t},$$

which holds for sufficiently large C.

For $s > 1$ we have:

$$\begin{aligned}
\mathrm{E}(F_{t+1}(s)) &= \mathrm{E}F_t(s)\left(1 - \frac{s\,(1-ct^{-\alpha})}{t}\right) + \mathrm{E}F_t(s-1)\frac{(s-1)\,(1-ct^{-\alpha})}{t} \\
&= \frac{c(s-1)!\,\Gamma(3-\alpha)}{(2-\alpha)\Gamma(s+2-\alpha)}\left(t^{1-\alpha} + \theta_{t,s}\right)\left(1 - \frac{s\,(1-ct^{-\alpha})}{t}\right) \\
&\quad + \frac{c(s-2)!\,\Gamma(3-\alpha)}{(2-\alpha)\Gamma(s+1-\alpha)}\left(t^{1-\alpha} + \theta_{t,s-1}\right)\frac{(s-1)\,(1-ct^{-\alpha})}{t} \\
&= \frac{c(s-1)!\,\Gamma(3-\alpha)}{(2-\alpha)\Gamma(s+2-\alpha)}\Bigg((t+1)^{1-\alpha} + O\left(t^{-\alpha-1}\right) - c(1-\alpha)t^{-2\alpha} \\
&\qquad + \theta_{t,s}\left(1 - \frac{s\,(1-ct^{-\alpha})}{t}\right) + \theta_{t,s-1}\frac{(s+1-\alpha)\,(1-ct^{-\alpha})}{t}\Bigg).
\end{aligned}$$

In order to finish the proof, it remains to show that

$$\begin{aligned}
O\left(t^{-\alpha-1}\right) + c(1-\alpha)t^{-2\alpha} + C\,t^{\max\{0,1-2\alpha\}}(s-1)^{1-\alpha+\epsilon}\frac{(s+1-\alpha)\,(1-ct^{-\alpha})}{t} \\
\leq Ct^{\max\{0,1-2\alpha\}}s^{1-\alpha+\epsilon}\frac{s\,(1-ct^{-\alpha})}{t},
\end{aligned}$$

$$O\left(t^{-\alpha}\right) + \frac{t^{1-2\alpha}c(1-\alpha)}{1-ct^{-\alpha}} \leq Ct^{\max\{0,1-2\alpha\}}\left(s^{2-\alpha+\epsilon} - (s+1-\alpha)(s-1)^{1-\alpha+\epsilon}\right),$$

$$O\left(t^{-\alpha}\right) + \frac{t^{1-2\alpha}c(1-\alpha)}{1-ct^{-\alpha}} \leq Ct^{\max\{0,1-2\alpha\}}s^{1-\alpha+\epsilon}\epsilon,$$

which holds for sufficiently large C.

References

1. http://konect.uni-koblenz.de/plots/bidd
2. Aiello, W., Bonato, A., Cooper, C., Janssen, J., Prałat, P.: A spatial web graph model with local influence regions. Internet Math. **5**(1–2), 175–196 (2008)
3. Arenas, A., Danon, L., Diaz-Guilera, A., Gleiser, P.M., Guimera, R.: Community analysis in social networks. Eur. Phys. J. B **38**(2), 373–380 (2004)
4. Artikov, A., Dorodnykh, A., Kashinskaya, Y., Samosvat, E.: Factorization threshold models for scale-free networks generation. Comput. Soc. Netw. **3**(1), 4 (2016)
5. Barabási, A.-L., Albert, R.: Emergence of scaling in random networks. Science **286**(5439), 509–512 (1999)
6. Barthélemy, M.: Crossover from scale-free to spatial networks. EPL (Europhysics Letters) **63**(6), 915 (2003)
7. Barthélemy, M.: Spatial networks. Phys. Rep. **499**(1), 1–101 (2011)
8. Bender, E.A., Canfield, E.R.: The asymptotic number of labeled graphs with given degree sequences. J. Comb. Theory, Ser. A **24**(3), 296–307 (1978)
9. Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., Hwang, D.-U.: Complex networks: structure and dynamics. Phys. Rep. **424**(4), 175–308 (2006)
10. Bollobás, B., Riordan, O., Spencer, J., Tusnády, G., et al.: The degree sequence of a scale-free random graph process. Random Struct. Algorithms **18**(3), 279–290 (2001)
11. Bollobás, B., Riordan, O.M.: Mathematical results on scale-free random graphs. In: Bornholdt, S., Schuster, H.G. (eds.) Handbook of Graphs and Networks: From the Genome to the Internet, pp. 1–34. Wiley-VCH, Weinheim (2003)
12. Bradonjić, M., Hagberg, A., Percus, A.G.: The structure of geographical threshold graphs. Internet Math. **5**(1–2), 113–139 (2008)
13. Buckley, P.G., Osthus, D.: Popularity based random graph models leading to a scale-free degree sequence. Discrete Math. **282**(1), 53–68 (2004)
14. Clauset, A., Newman, M.E.J., Moore, C.: Finding community structure in very large networks. Phys. Rev. E **70**(6), 066111 (2004)
15. da F Costa, L., Rodrigues, F.A., Travieso, G., Villas Boas, P.R.: Characterization of complex networks: a survey of measurements. Adv. Phys. **56**(1), 167–242 (2007)
16. Dunlavy, D.M., Kolda, T.G., Acar, E.: Temporal link prediction using matrix and tensor factorizations. ACM Trans. Knowl. Discovery Data (TKDD) **5**(2), 10 (2011)
17. Ester, M., Kriegel, H.-P., Sander, J., Xu, X., et al.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: KDD, vol. 96, pp. 226–231 (1996)
18. Faloutsos, M., Faloutsos, P., Faloutsos, C.: On power-law relationships of the internet topology. In: ACM SIGCOMM Computer Communication Review, vol. 29, pp. 251–262. ACM (1999)
19. Fortunato, S.: Community detection in graphs. Phys. Rep. **486**(3), 75–174 (2010)

20. Fortunato, S., Barthelemy, M.: Resolution limit in community detection. Proc. Natl. Acad. Sci. **104**(1), 36–41 (2007)
21. Girvan, M., Newman, M.E.J.: Community structure in social and biological networks. Proc. Natl. Acad. Sci. **99**(12), 7821–7826 (2002)
22. Guimera, R., Danon, L., Diaz-Guilera, A., Giralt, F., Arenas, A.: Self-similar community structure in a network of human interactions. Phys. Rev. E **68**(6), 065103 (2003)
23. Holme, P., Kim, B.J.: Growing scale-free networks with tunable clustering. Phys. Rev. E **65**(2), 026107 (2002)
24. Hufnagel, L., Brockmann, D., Geisel, T.: Forecast and control of epidemics in a globalized world. Proc. Natl. Acad. Sci. U.S.A. **101**(42), 15124–15129 (2004)
25. Kempe, D., Kleinberg, J., Tardos, É.: Maximizing the spread of influence through a social network. In: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 137–146. ACM (2003)
26. Krot, A., Ostroumova Prokhorenkova, L.: Local clustering coefficient in generalized preferential attachment models. In: Gleich, D.F., Komjáthy, J., Litvak, N. (eds.) WAW 2015. LNCS, vol. 9479, pp. 15–28. Springer, Cham (2015). doi:10.1007/978-3-319-26784-5_2
27. Kumpula, J.M., Onnela, J.-P., Saramäki, J., Kertész, J., Kaski, K.: Model of community emergence in weighted social networks. Comput. Phys. Commun. **180**(4), 517–522 (2009)
28. Lancichinetti, A., Fortunato, S., Radicchi, F.: Benchmark graphs for testing community detection algorithms. Phys. Rev. E **78**(4), 046110 (2008)
29. Lipsitch, M., Cohen, T., Cooper, B., Robins, J.M., Ma, S., James, L., Gopalakrishna, G., Chew, S.K., Tan, C.C., Samore, M.H., et al.: Transmission dynamics and control of severe acute respiratory syndrome. Science **300**(5627), 1966–1970 (2003)
30. Lloyd, S.: Least squares quantization in PCM. IEEE Trans. Inform. Theory **28**(2), 129–137 (1982)
31. Masuda, N., Miwa, H., Konno, N.: Geographical threshold graphs with small-world and scale-free properties. Phys. Rev. E **71**(3), 036108 (2005)
32. Menon, A.K., Elkan, C.: Link prediction via matrix factorization. In: Gunopulos, D., Hofmann, T., Malerba, D., Vazirgiannis, M. (eds.) ECML PKDD 2011. LNCS, vol. 6912, pp. 437–452. Springer, Heidelberg (2011). doi:10.1007/978-3-642-23783-6_28
33. Menon, A.K., Elkan, C.: A log-linear model with latent features for dyadic prediction. In: 2010 IEEE 10th International Conference on Data Mining (ICDM), pp. 364–373. IEEE (2010)
34. Miller, K., Jordan, M.I., Griffiths, T.L.: Nonparametric latent feature models for link prediction. In: Advances in Neural Information Processing Systems, pp. 1276–1284 (2009)
35. Newman, M.E.J.: The structure and function of complex networks. SIAM Rev. **45**(2), 167–256 (2003)
36. Newman, M.E.J.: Power laws, pareto distributions and zipf's law. Contemp. Phys. **46**(5), 323–351 (2005)
37. Newman, M.E.J., Girvan, M.: Finding and evaluating community structure in networks. Phys. Rev. E **69**(2), 026113 (2004)
38. Ostroumova Prokhorenkova, L., Samosvat, E.: Recency-based preferential attachment models. J. Complex Netw. **4**(4), 475–499 (2016)

39. Palla, G., Derényi, I., Farkas, I., Vicsek, T.: Uncovering the overlapping community structure of complex networks in nature and society. Nature **435**(7043), 814–818 (2005)
40. Pollner, P., Palla, G., Vicsek, T.: Preferential attachment of communities: the same principle, but a higher level. EPL (Europhysics Letters) **73**(3), 478 (2005)
41. Raigorodskii, A.M.: Small subgraphs in preferential attachment networks. Optimization Lett. **11**(2), 249–257 (2017)
42. Romero, D.M., Meeder, B., Kleinberg, J.: Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter. In: Proceedings of the 20th International Conference on World Wide Web, pp. 695–704. ACM (2011)
43. Wang, C., Knight, J.C., Elder, M.C.: On computer viral infection and the effect of immunization. In: 16th Annual Conference on Computer Security Applications, ACSAC 2000, pp. 246–256. IEEE (2000)
44. Waxman, B.M.: Routing of multipoint connections. IEEE J. Sel. Areas Commun. **6**(9), 1617–1622 (1988)
45. Zhou, T., Yan, G., Wang, B.-H.: Maximal planar networks with large clustering coefficient and power-law degree distribution. Phys. Rev. E **71**(4), 046141 (2005)