Assortativity in Generalized Preferential Attachment Models

Alexander Krot[1(✉)] and Liudmila Ostroumova Prokhorenkova[1,2]

[1] Moscow Institute of Physics and Technology, Moscow, Russia
al.krot.kav@gmail.com
[2] Yandex, Moscow, Russia
ostroumova-la@yandex.ru

Abstract. In this paper, we analyze assortativity of preferential attachment models. We deal with a wide class of preferential attachment models (PA-class). It was previously shown that the degree distribution in all models of the PA-class follows a power law. Also, the global and the average local clustering coefficients were analyzed. We expand these results by analyzing the assortativity property of the PA-class of models. Namely, we analyze the behavior of $d_{nn}(d)$ which is the average degree of a neighbor of a vertex of degree d.

Keywords: Networks · Random graphs · Preferential attachment · Assortativity · Average neighbor degree

1 Introduction

Nowadays, there is a great deal of interest in structure and dynamics of real-world networks, from Internet and society networks [1,4,7] to biological networks [2]. The key problem is how to build a model which describes the properties of a given network. Such models are used in physics, information retrieval, data mining, bioinformatics, etc. [1,4,5,17].

Real-world networks have some common properties [12,19,20,22]. For example, for the majority of studied networks, the degree distribution was observed to follow the power law, which means that the portion of vertices with degree d decreases as $d^{-\gamma}$ for some $\gamma > 0$ [3,4,8,11]. Another important property of complex networks is high clustering coefficient [20] which, roughly speaking, measures how likely two neighbors of a vertex are connected.

Another key metric in complex networks analysis is the assortativity coefficient which was first introduced by Newman [18] as the Pearson's correlation coefficient for the pairs $\{(d_i, d_j)|e_{ij} \in E\}$. In assortative graphs edges tend to connect vertices of similar degrees, while in disassortative networks vertices of low degree are more likely to be connected to vertices of high degree. Assortativity coefficient lies between -1 and 1; when this coefficient equals 1, the network is said to have perfect assortative mixing patterns, when it equals 0, the network is non-assortative, while at -1 the network is completely disassortative.

A. Bonato et al. (Eds.): WAW 2016, LNCS 10088, pp. 9–21, 2016.
DOI: 10.1007/978-3-319-49787-7_2

However, as discussed in [13,15], despite Pearson's correlation coefficient is most commonly used to measure assortativity of a network, this coefficient is size-depend when the degree distribution has infinite variance. Another way to analyze assortativity is to consider the behavior of $d_{nn}(d)$ — the average degree of a neighbor of a vertex of degree d. A graph is called assortative if $d_{nn}(d)$ is an increasing function of d, whereas it is referred to as disassortative when $d_{nn}(d)$ is a decreasing function of d. We analyze $d_{nn}(d)$ instead of measuring the correlation since the obtained function of d can give a deeper insight into the network structure.

It was previously shown that in some real-world networks $d_{nn}(d)$ behaves as d^ν for some ν, which can be positive (assortative networks) or negative (disassortative networks) [4,10]. Interestingly, as we show in this paper, in a wide class of preferential attachment models $d_{nn}(d) \propto \log(d)$ as $d \to \infty$.

Assortativity has many applications, for instance, it can be used in the epidemiology. In social networks we usually observe assortative mixing, so diseases targeting high degree individuals are likely to spread to other high degree nodes. On the other hand, biological networks are usually disassortative, therefore vaccination strategies that specifically target the high degree vertices may quickly destroy the epidemic network.

In this paper, we study the behavior of $d_{nn}(d)$ in the T-subclass of the PA-class of models, which was first introduced in [21]. This class includes a lot of well-known models based on the preferential attachment principle: LCD [6], Buckley-Osthus [9], Holme-Kim [14], RAN [23], etc. Despite the fact that the T-subclass generalizes many different models, we are able to analyze $d_{nn}(d)$ in the whole class of models for $\gamma > 3$ (the case of finite variance). We prove that the expectation of $d_{nn}(d)$ asymptotically behaves as $\log(d)$ (up to a constant multiplier). However, this approximation works reasonably well only for very large values of d and for $d < 10^4$ we observe a different behavior which may look like d^ν for some $\nu > 0$.

The remainder of the paper is organized as follows. In Sect. 2, we give a formal definition of the PA-class and present some known results. Then, in Sect. 3, we state new results on the behavior of $d_{nn}(d)$. In Sect. 4, we make some simulations in order to illustrate our results for $d_{nn}(d)$. We prove all theorems in Sect. 5.

2 Generalized Preferential Attachment

2.1 Definition of the PA-Class

Let us formally define the PA-class of models which was first suggested in [21]. Let G_m^n $(n \geq n_0)$ be a graph with n vertices $\{1, \ldots, n\}$ and mn edges obtained as a result of the following process. We start at the time n_0 from an arbitrary graph $G_m^{n_0}$ with n_0 vertices and mn_0 edges. On the $(n+1)$-th step $(n \geq n_0)$, we make the graph G_m^{n+1} from G_m^n by adding a new vertex $n+1$ and m edges connecting this vertex to some m vertices from the set $\{1, \ldots, n, n+1\}$. Denote by d_v^n the degree of a vertex v in G_m^n. If for some constants A and B the following conditions are satisfied

$$\mathsf{P}\left(d_v^{n+1} = d_v^n \mid G_m^n\right) = 1 - A\frac{d_v^n}{n} - B\frac{1}{n} + O\left(\frac{(d_v^n)^2}{n^2}\right),\ 1 \le v \le n, \tag{1}$$

$$\mathsf{P}\left(d_v^{n+1} = d_v^n + 1 \mid G_m^n\right) = A\frac{d_v^n}{n} + B\frac{1}{n} + O\left(\frac{(d_v^n)^2}{n^2}\right),\ 1 \le v \le n, \tag{2}$$

$$\mathsf{P}\left(d_v^{n+1} = d_v^n + j \mid G_m^n\right) = O\left(\frac{(d_v^n)^2}{n^2}\right),\ 2 \le j \le m,\ 1 \le v \le n, \tag{3}$$

$$\mathsf{P}(d_{n+1}^{n+1} = m + j) = O\left(\frac{1}{n}\right),\ 1 \le j \le m, \tag{4}$$

then the random graph process G_m^n is a model from the PA-class. Here, as in [21], we require $2mA + B = m$ and $0 \le A \le 1$. We further omit n in d_j^n for simplicity of notation.

As it is explained in [21], even fixing values of parameters A and m does not specify a concrete procedure for constructing a network. There are a lot of models possessing very different properties and satisfying the conditions (1–4), e.g., LCD, Buckley–Osthus, Holme–Kim, and RAN models.

2.2 Power-Law Degree Distribution

Let $N_n(d)$ be the number of vertices of degree d in G_m^n. The following theorems on the expectation of $N_n(d)$ and its concentration were proved in [21].

Theorem 1. *For every model in PA-class and for every $d = d(n) \ge m$*

$$\mathsf{E}N_n(d) = c(m,d)\left(n + O\left(d^{2+\frac{1}{A}}\right)\right),$$

where

$$c(m,d) = \frac{\Gamma\left(d + \frac{B}{A}\right)\Gamma\left(m + \frac{B+1}{A}\right)}{A\,\Gamma\left(d + \frac{B+A+1}{A}\right)\Gamma\left(m + \frac{B}{A}\right)} \overset{d\to\infty}{\sim} \frac{\Gamma\left(m + \frac{B+1}{A}\right) d^{-1-\frac{1}{A}}}{A\,\Gamma\left(m + \frac{B}{A}\right)}$$

and $\Gamma(x)$ is the gamma function.

Theorem 2. *For every model from the PA-class and for every $d = d(n)$ we have*

$$\mathsf{P}\left(|N_n(d) - \mathsf{E}N_n(d)| \ge d\sqrt{n}\log n\right) = O\left(n^{-\log n}\right).$$

These two theorems mean that the degree distribution follows (asymptotically) the power law with the parameter $1 + \frac{1}{A}$.

2.3 Clustering Coefficient

A T-subclass of the PA-class was introduced in [21]. In this case, the following additional condition is required:

$$\mathsf{P}\left(d_i^{n+1} = d_i^n + 1, d_j^{n+1} = d_j^n + 1 \mid G_m^n\right) = e_{ij}\frac{D}{mn} + O\left(\frac{d_i^n d_j^n}{n^2}\right), \tag{5}$$

where $1 \leq i, j \leq n$, e_{ij} is the number of edges between the vertices i and j in G_m^n and D is a non-negative constant. Note that this property still does not define the correlation between edges completely, but it is sufficient for studying the clustering coefficients. Also, this subclass still covers all well-known models mentioned above.

There are two well-known definitions of the clustering coefficient of a graph G. The *global clustering coefficient* $C_1(G)$ is the ratio of three times the number of triangles to the number of pairs of adjacent edges in G. The *average local clustering coefficient* is defined as $C_2(G) = \frac{1}{n}\sum_{i=1}^n C(i)$, where $C(i)$ is the local clustering coefficient for a vertex i: $C(i) = \frac{T^i}{P_2^i}$, T^i is the number of edges between the neighbors of the vertex i and P_2^i is the number of pairs of neighbors.

The clustering coefficients for the T-subclass were analyzed in [16,21]. For example, in [21] it was proven that in some cases ($2A \geq 1$) the global clustering coefficient $C_1(G_m^n)$ tends to zero as the number of vertices grows for all models from the PA-class. Additionally, it was shown that the average local clustering coefficient $C_2(G_m^n)$ does not tend to zero for the T-subclass with $D > 0$. In [16] the local clustering coefficient averaged over the vertices of degree d was analyzed. It was proven that this coefficient $C(d)$ asymptotically decreases as $\frac{2D}{Am} \cdot d^{-1}$ for $A < \frac{3}{4}$.

3 Assortativity

In this paper, we analyze the assortativity property in the T-subclass. One possible way to analyze the assortativity of an undirected graph G is to consider the average degree of the neighbors of vertices with a given degree d:

$$d_{nn}(d) = \frac{1}{N_n(d) \cdot d} \sum_{i:d_i=d} \sum_{j:ij \in E(G)} d_j, \tag{6}$$

where $E(G)$ is the set of edges of the graph G. If $d_{nn}(d)$ is an increasing function of d, then the network is assortative. Vice-versa, in the disassortative case $d_{nn}(d)$ decreases.

Let $S_n(d)$ be the sum of the degrees of all neighbors of all vertices of degree d. Then $d_{nn}(d)$ can be defined as $d_{nn}(d) = \frac{S_n(d)}{dN_n(d)}$. Hence, in order to estimate $\mathsf{E}d_{nn}(d)$, we first estimate $\mathsf{E}S_n(d)$ and then use Theorems 1 and 2 on the behavior of $N_n(d)$. Namely, we prove the following theorems.

Theorem 3. *Let G_m^n belong to the T-subclass with $A < \frac{1}{2}$. Then, for any $\varepsilon > 0$ and for every $d = d(n) \geq m$*

$$\mathsf{E}S_n(d) = M(d)\left(n + O\left(n^{2A+\varepsilon} d^{2+\frac{1}{A}}\right)\right),$$

where

$$M(d) = (Ad + B + 1)\left[\frac{X}{Am + B + 1} + \sum_{i=m+1}^{d} Y(i)\right] \cdot c(m, d),$$

$$X = \frac{m}{A(m-1) + B + 1}\left[B - \frac{D}{m} + \frac{(A(m-1) + 2B + 1) \cdot (Am + B + 1)}{1 - 2A}\right],$$

$$Y(i) = \frac{1}{A(i-1) + B + 1}\left[\frac{(B - D/m)i}{Ai + B + 1} + \frac{(D/m) \cdot (i-1)}{A(i-1) + B} + m\right].$$

Asymptotically we have

$$M(d) \overset{d\to\infty}{\sim} \frac{Am + B}{A^2} \cdot \frac{\Gamma\left(m + \frac{B+1}{A}\right)}{\Gamma\left(m + \frac{B}{A}\right)} \cdot \log(d) \cdot d^{-\frac{1}{A}}.$$

Theorem 4. *Let G_m^n belong to the T-subclass of the PA-class with $A < \frac{1}{2}$. Then for any $\varepsilon > 0$ and for every $d = d(n) \geq m$*

$$\mathsf{E}d_{nn}(d) = \frac{M(d)}{d\, c(m, d)}\left(1 + O\left(\frac{n^{2A+\varepsilon} d^{2+\frac{1}{A}}}{n} + \frac{d^{2+\frac{1}{A}} \log n}{\sqrt{n}}\right)\right).$$

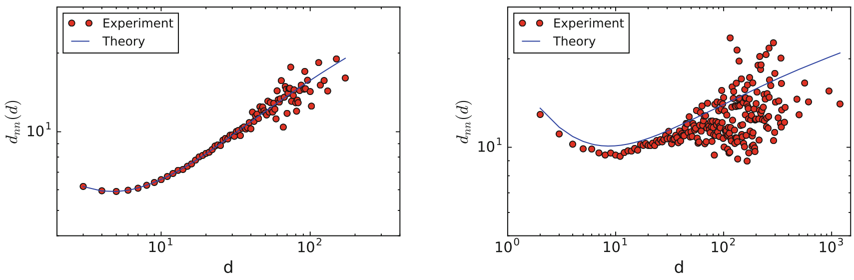
Note that $\frac{M(d)}{d \cdot c(m,d)} \overset{d\to\infty}{\sim} \frac{Am+B}{A} \cdot \log(d)$.

Note that the restriction $A < \frac{1}{2}$ is essential and for $A \geq \frac{1}{2}$ the result is expected to be completely different. Indeed, when we analyze $S_n(m)$ we have to estimate the expected sum of the degrees of the neighbors of a new vertex i. If $A \geq \frac{1}{2}$, then this sum grows faster than linearly in i and our approximations do not hold.

According to Theorem 4, all networks from the T-subclass with $A < \frac{1}{2}$ are assortative. However, $\mathsf{E}d_{nn}(d)$ increases slowly, as $\log(d)$, unlike d^ν in real-world networks. It is also worth noting that in Theorem 4 we analyze only the average value of $d_{nn}(d)$ and proving concentration is left for future research.

4 Experiments

In this section, we look at the behavior of $d_{nn}(d)$ in a three-parameter model from the family of polynomial graph models defined in [21]. This model belongs to the T-subclass and by varying the parameters we can analyze the effect of A (or, equivalently, γ) on $d_{nn}(d)$. In all the experiments we generated polynomial graphs with $n = 10^6$, $m = 2$, $D = 0.3$ and different values of A. In other words,



(a) $A = 0.2$ (b) $A = 0.4$

Fig. 1. The behavior of $d_{nn}(d)$ for different A

we fixed the probability of a triangle formation and vary the parameter of the power-law degree distribution. Detailed graph generation process is described in [21].

First, let us illustrate our main result for $\mathsf{E}d_{nn}(d)$ (see Theorem 4). We generated polynomial graphs for different A and compared the obtained values of $d_{nn}(d)$ with their theoretical approximation $\frac{M(d)}{d \cdot c(m,d)}$. We noticed that for $A < \frac{1}{3}$ the theoretical value of $\mathsf{E}d_{nn}(d)$ is extremely close to the experiment. However, if $A > \frac{1}{3}$, then $d_{nn}(d)$ turn out to be consistently smaller than their theoretical approximation. Figure 1 illustrates this observation and shows $d_{nn}(d)$ for $A = 0.2$ and $A = 0.4$. However, according to our additional experiments, the obtained for $A > \frac{1}{3}$ difference tends to zero as $n \to \infty$, as expected. The possible reason for such a slow convergence is the error term $O\left(\frac{n^{3A}}{n^2}\right)$ appearing in the proof in the case $A > \frac{1}{3}$.

We also compared the theoretical value of $\mathsf{E}d_{nn}(d)$ (for $A = 0.2$) with the asymptotic formula $\frac{Am+B}{A} \cdot \log(d)$ (see Fig. 2). Interestingly, from Fig. 1 it may seem that $d_{nn}(d)$ grows as d^{ν} for some ν (as it was observed in real-world networks). However, as d becomes large ($d > 10^4$), one can indeed observe the logarithmic growth.

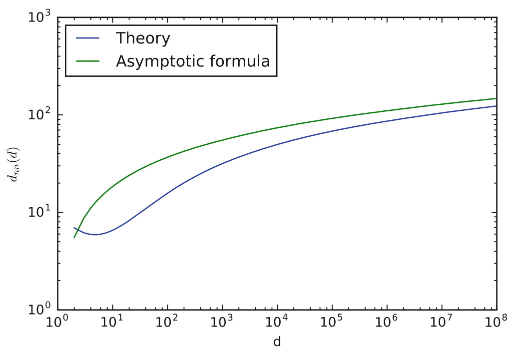


Fig. 2. Theoretical value of $d_{nn}(d)$ versus its asymptotic approximation

5 Proofs

5.1 Proof of Theorem 3

In the proof we use the notation $\theta(\cdot)$ for error terms. By $\theta(X)$ we denote an arbitrary function such that $|\theta(X)| < X$.

We need the following auxiliary theorem.

Theorem 5. *Let W_n be the sum of the squares of the degrees of all vertices in a model from the PA-class with $A < \frac{1}{2}$. Then for any $\epsilon > 0$*

$$\mathsf{E}W_n = \frac{m}{1-2A}\,(m+4B+1)\,n + O(n^{2A+\epsilon}).$$

Theorem 5 can easily be proved by induction on n. The proof is omitted due to space constraints.

Now let us prove Theorem 3. It can be shown that a.a.s. the maximum degree in G_m^n is less than $n^{A+\varphi}$ for any $\varphi > 0$. Also, $\mathsf{E}N_n(d) = c(m,d)\left(n + O\left(d^{2+\frac{1}{A}}\right)\right)$. Therefore, one can show that $\mathsf{E}S_n(d) = c(m,d)\left(n + O\left(d^{2+\frac{1}{A}}\right)\right) d\, O(n^{A+\varphi})$. As a result, for $n \le Q \cdot d^2$ (for any constant Q) we have $\mathsf{E}S_n(d) = O\left(d^2 n^{A+\varphi}\right) = M(d) \cdot O\left(n^{2A+\varepsilon} \cdot d^{2+\frac{1}{A}}\right)$ for any $\varepsilon > 0$. This concludes the proof for $n \le Qd^2$ for all d.

In order to prove Theorem 3, we use induction on d and for each d we use induction on n. Note that for each d we already have the basis for $n \le Qd^2$.

Consider the case $d = m$. At each step we add a vertex $n+1$ and m edges. We have the following possibilities.

1. At least one edge hits a vertex of degree m, then $S_n(m)$ is decreased by the sum of the degrees of the neighbors of this vertex. This happens with probability $\frac{A\,m+B}{n} + O\left(\frac{1}{n^2}\right)$. Summing over all vertices of degree m we obtain that $\mathsf{E}S_n(m)$ is decreased by $\left(\frac{Am+B}{n} + O\left(\frac{1}{n^2}\right)\right) \cdot \mathsf{E}S_n(m)$.
2. Exactly one edge hits a neighbor of a vertex of degree m and no edges hit the vertex itself, then $S_n(m)$ is increased by 1. The probability to hit a neighbor is $\frac{Ad_i+B}{n} + O\left(\frac{d_i^2}{n^2}\right)$, where d_i is the degree of this neighbor. We have to subtract the probability to hit both a vertex of degree m and its neighbor which is $\frac{D}{mn} + O\left(\frac{md_i}{n^2}\right)$. Summing over all neighbors of all vertices of degree m, we obtain that $\mathsf{E}S_n(m)$ is increased by:

$$\frac{A\mathsf{E}S_n(m)}{n} + \frac{B - D/m}{n} m\mathsf{E}N_n(m) + O\left(\frac{\mathsf{E}\sum_{\substack{i:i \text{ is a neighbor}\\ \text{of a vertex of degree } m}} d_i^2}{n^2}\right)$$
$$= \frac{A\mathsf{E}S_n(m)}{n} + \frac{B - D/m}{n} m\mathsf{E}N_n(m) + O\left(\frac{\max(n, n^{3A})}{n^2}\right).$$

Here we used the fact that:

$$\mathsf{E}\Bigg(\sum_{\substack{i:i \text{ is a neighbor}\\ \text{of a vertex of degree } m}} d_i^2\Bigg) \le \mathsf{E}\left(\sum_{i\in V(G_n^m)} d_i^3\right) = O\left(\max(n, n^{3A})\right).$$

3. If $i > 1$ edges hit a neighbor j of a vertex of degree m, which happens with probability $O\left(\frac{d_j^2}{n^2}\right)$, and no edges hit the vertex itself, then $S_n(m)$ is increased by i. Reasoning as above, we obtain that $\mathsf{E}S_n(m)$ is increased by $O\left(\frac{\max(n,n^{3A})}{n^2}\right)$.
4. The vertex $n+1$ hits some vertices, so $S_n(m)$ is increased by the sum of the degrees of these vertices. The probability to hit a vertex of degree d_i is $\frac{Ad_i+B}{n}+O\left(\frac{d_i^2}{n^2}\right)$ and after that this vertex will have a degree d_i+1. Summing over i we obtain that $\mathsf{E}S_n(m)$ is increased by:

$$\mathsf{E}\sum_{i\in V(G_n^m)}(d_i+1)\left(\frac{Ad_i+B}{n}+O\left(\frac{d_i^2}{n^2}\right)\right)$$
$$= \frac{A}{n}\mathsf{E}W_n + (2B+1)m + O\left(\frac{\max(n,n^{3A})}{n^2}\right).$$

Combining all the cases considered above, we get

$$\mathsf{E}S_{n+1}(m) = \mathsf{E}S_n(m) - \left[\frac{Am+B}{n} + O\left(\frac{1}{n^2}\right)\right]\mathsf{E}S_n(m) + \frac{A\mathsf{E}S_n(m)}{n}$$
$$+\frac{B-D/m}{n}m\mathsf{E}N_n(m) + \frac{A}{n}\mathsf{E}W_n + (2B+1)m + O\left(\frac{\max(n,n^{3A})}{n^2}\right).$$

We prove by induction on n that $\mathsf{E}S_n(m) = M(m)\left(n+\theta\left(Cn^{2A+\varepsilon}m^{2+\frac{1}{A}}\right)\right)$ for some constant $C>0$, where

$$M(m) = \frac{m\cdot c(m,m)}{A(m-1)+B+1}\left[B - \frac{D}{m} + \frac{(A(m-1)+2B+1)\cdot(Am+B+1)}{1-2A}\right]. \tag{7}$$

Assume that $\mathsf{E}S_i(m) = M(m)\left(i+\theta\left(Ci^{2A+\varepsilon}\cdot m^{2+\frac{1}{A}}\right)\right)$ for all $i < n+1$ and let us prove this result for $i = n+1$:

$$\mathsf{E}S_{n+1}(m) = \left[1 - \frac{A(m-1)+B}{n} + O\left(\frac{1}{n^2}\right)\right]\cdot M(m)\left(n+\theta\left(Cn^{2A+\varepsilon}m^{2+\frac{1}{A}}\right)\right)$$
$$+\frac{B-D/m}{n}m\cdot c(m,m)\left[n+O(1)\right] + \frac{A}{n}\cdot\frac{m}{1-2A}\left(m+4B+1\right)n$$
$$+O(n^{2A-1+\epsilon}) + (2B+1)m + O\left(\frac{\max(n,n^{3A})}{n^2}\right). \tag{8}$$

Here we use that $\mathsf{E}W_n = \frac{m}{1-2A}\left(m+4B+1\right)n + O(n^{2A+\epsilon})$ and take $\epsilon < \varepsilon$.

Next, we use (7) and the fact that $c(m,m) = 1/(Am+B+1)$:

$$\mathsf{E}S_{n+1}(m) = M(m)(n+1) + \left[1 - \frac{A(m-1)+B}{n}\right] M(m)\theta\left(Cn^{2A+\varepsilon}m^{2+\frac{1}{A}}\right) + O\left(Cn^{2A-2+\varepsilon}\right) + O(n^{2A-1+\epsilon}).$$

To complete the proof for $d = m$ we have to show that the obtained error term is not greater than $CM(m)m^{2+\frac{1}{A}}(n+1)^{2A+\varepsilon}$ for some large enough C:

$$CM(m)m^{2+\frac{1}{A}}(n+1)^{2A+\varepsilon} \geq \left[1 - \frac{A(m-1)+B}{n}\right] M(m)Cn^{2A+\varepsilon}m^{2+\frac{1}{A}} + O\left(Cn^{2A-2+\varepsilon}\right) + O(n^{2A-1+\epsilon}).$$

This inequality holds for large enough C. This completes the proof for $d = m$.

Now, consider the case $d > m$, $n > Q\,d^2$. Similarly to the previous case, once we add a vertex $n+1$ and m edges, we have the following possibilities.

1. At least one edge hits a vertex of degree d. In this case, $\mathsf{E}S_n(d)$ is decreased by $\left(\frac{Ad+B}{n} + O\left(\frac{d^2}{n^2}\right)\right) \cdot \mathsf{E}S_n(d)$.
2. One edge hits a vertex of degree $d-1$, so $S_n(d)$ is increased by the sum of the degrees of the neighbors of this vertex plus the degree of the new vertex. We get
$$\left(\frac{A(d-1)+B}{n} + O\left(\frac{d^2}{n^2}\right)\right) \cdot \left(\mathsf{E}S_n(d-1) + m \cdot \mathsf{E}N_n(d-1)\right).$$
Taking into account the case when, in addition, exactly one edge hits a neighbor of this vertex, we get that $\mathsf{E}S_n(d)$ is additionally increased by:
$$(d-1)\mathsf{E}N_n(d-1) \cdot \frac{D}{mn} + O\left(\frac{(d-1)\mathsf{E}S_n(d-1)}{n^2}\right).$$
3. Exactly one edge hits a neighbor of a vertex of degree d and no edges hit the vertex itself. In this case, $\mathsf{E}S_n(d)$ is increased by:
$$\frac{A\mathsf{E}S_n(d)}{n} + \frac{B - D/m}{n} d\mathsf{E}N_n(d) + O\left(\frac{\max(n, n^{3A})}{n^2}\right) + O\left(\frac{d \cdot \mathsf{E}S_n(d)}{n^2}\right).$$
4. All the cases with multiple edges affect $\mathsf{E}S_n(d)$ by:
$$O\left(\frac{\max(n, n^{3A})}{n^2}\right) + O\left(\frac{d^2}{n^2}\right)\mathsf{E}S_n(d) + O\left(\frac{d^3}{n^2}\right)\mathsf{E}N_n(d). \tag{9}$$

Combining all the cases considered above, we get

$$\begin{aligned}\mathsf{E}S_{n+1}(d) = \mathsf{E}S_n(d)\left[1 - \frac{A(d-1)+B}{n}\right] + \frac{A(d-1)+B}{n} \cdot \mathsf{E}S_n(d-1) \\ + \left(\frac{D(d-1)}{mn} + m\frac{A(d-1)+B}{n}\right)\mathsf{E}N_n(d-1) + \frac{(B-D/m)d}{n}\mathsf{E}N_n(d) \\ + O\left(\frac{d^2}{n^2}\right)\mathsf{E}S_n(d) + O\left(\frac{d^3}{n^2}\right)\mathsf{E}N_n(d) + O\left(\frac{\max(n, n^{3A})}{n^2}\right).\end{aligned}$$

We prove by induction on d and n that $\mathsf{E}S_n(d) = M(d)\left(n + \theta\left(Cn^{2A+\varepsilon} d^{2+\frac{1}{A}}\right)\right)$ for some constant $C > 0$. Assume that $\mathsf{E}S_i(\tilde{d}) = M(\tilde{d})\left(i + \theta\left(Ci^{2A+\varepsilon} \tilde{d}^{2+\frac{1}{A}}\right)\right)$ for $\tilde{d} < d$ and all i and for $\tilde{d} = d$ and $i < n+1$. Then

$$\begin{aligned}
\mathsf{E}S_{n+1}(d) = &\left[1 - \frac{A(d-1)+B}{n}\right] M(d)\left[n + \theta\left(Cn^{2A+\varepsilon} d^{2+\frac{1}{A}}\right)\right] \\
&+ \frac{A(d-1)+B}{n} M(d-1)\left[n + \theta\left(Cn^{2A+\varepsilon}(d-1)^{2+\frac{1}{A}}\right)\right] \\
&+ \left(\frac{D(d-1)}{mn} + m\frac{A(d-1)+B}{n}\right) c(m,d-1)\left[n + O\left(d^{2+\frac{1}{A}}\right)\right] \\
&+ \frac{(B-D/m)d}{n} c(m,d)\left[n + O\left(d^{2+\frac{1}{A}}\right)\right] + O\left(\frac{d^2}{n^2}\right) M(d)\left[n + \theta\left(Cn^{2A+\varepsilon} d^{2+\frac{1}{A}}\right)\right] \\
&+ O\left(\frac{d^3}{n^2}\right) c(m,d)\left[n + O\left(d^{2+\frac{1}{A}}\right)\right] + O\left(\frac{\max(n, n^{3A})}{n^2}\right).
\end{aligned}$$

Note that

$$\begin{aligned}
M(d) = &\frac{A(d-1)+B}{A(d-1)+B+1} M(d-1) + \frac{(B-D/m)d}{A(d-1)+B+1} c(m,d) \\
&+ \frac{\left(\frac{D}{m} + Am\right)(d-1) + Bm}{A(d-1)+B+1} c(m,d-1).
\end{aligned} \tag{10}$$

Therefore, we obtain:

$$\begin{aligned}
\mathsf{E}S_{n+1}(d) = &M(d)(n+1) + \left[1 - \frac{A(d-1)+B}{n}\right] M(d)\,\theta\left(Cn^{2A+\varepsilon} d^{2+\frac{1}{A}}\right) \\
&+ \frac{A(d-1)+B}{n} M(d-1)\,\theta\left(Cn^{2A+\varepsilon}(d-1)^{2+\frac{1}{A}}\right) \\
&+ O\left(C\frac{d^4 \log(d) \cdot n^{2A+\varepsilon}}{n^2}\right) + O\left(C\frac{d^{2-\frac{1}{A}} \log(d)}{n}\right) + O\left(\frac{\max(n, n^{3A})}{n^2}\right) + O\left(\frac{d^2}{n}\right).
\end{aligned}$$

It remains to prove that for some large enough C

$$\begin{aligned}
&CM(d) \cdot (n+1)^{2A+\varepsilon} d^{2+\frac{1}{A}} \\
&\geq CM(d) \cdot n^{2A+\varepsilon} d^{2+\frac{1}{A}} - CM(d)\left(\frac{A(d-1)+B}{n}\right) \cdot n^{2A+\varepsilon} d^{2+\frac{1}{A}} \\
&+ CM(d-1)\left(\frac{A(d-1)+B}{n}\right) \cdot n^{2A+\varepsilon}(d-1)^{2+\frac{1}{A}} \\
&+ O\left(C\frac{d^4 \log(d) \cdot n^{2A+\varepsilon}}{n^2}\right) + O\left(C\frac{d^{2-\frac{1}{A}} \log(d)}{n}\right) + O\left(\frac{\max(n, n^{3A})}{n^2}\right) + O\left(\frac{d^2}{n}\right).
\end{aligned} \tag{11}$$

First, note that

$$CM(d)\cdot(n+1)^{2A+\varepsilon}d^{2+\frac{1}{A}} - CM(d)\cdot n^{2A+\varepsilon}d^{2+\frac{1}{A}} = CM(d)\cdot n^{2A+\varepsilon}\cdot d^{2+\frac{1}{A}}\left[\frac{2A+\varepsilon}{n}+O\left(\frac{1}{n^2}\right)\right].$$

Second, one can show that

$$CM(d)\left(\frac{A(d-1)+B}{n}\right)d^{2+\frac{1}{A}} - CM(d-1)\left(\frac{A(d-1)+B}{n}\right)(d-1)^{2+\frac{1}{A}} \geq 0$$

using Eq. (10) and the inequality $(1-\frac{1}{d})^{-(2+\frac{1}{A})} \geq 1+\frac{2A+1}{Ad}$.

Therefore, Eq. (11) becomes:

$$CM(d)\cdot n^{2A+\varepsilon}\cdot d^{2+\frac{1}{A}}\left[\frac{2A+\varepsilon}{n}+O\left(\frac{1}{n^2}\right)\right] \geq O\left(C\frac{d^4\log(d)\cdot n^{2A+\varepsilon}}{n^2}\right) + O\left(C\frac{d^{2-\frac{1}{A}}\log(d)}{n}\right)+O\left(\frac{\max(n,n^{3A})}{n^2}\right)+O\left(\frac{d^2}{n}\right).$$

It is easy to see that for some large enough C and for $n \geq Q\cdot d^2$ (for some large enough Q) this inequality is satisfied. This concludes the proof of the theorem.

5.2 Proof of Theorem 4

Denote by Q the event $\{|N_n(d)-\mathsf{E}N_n(d)| < d\sqrt{n}\log(n)\}$. According to Theorem 2, $\mathsf{P}(Q) = 1-O\left(n^{-\log(n)}\right)$. Let us estimate $\mathsf{E}d_{nn}(d)$:

$$\mathsf{E}d_{nn}(d) = \mathsf{E}\left(\frac{S_n(d)}{d\,N_n(d)}\right) = \mathsf{E}\left(\frac{S_n(d)}{d\,N_n(d)}\middle|Q\right)\mathsf{P}(Q) + \mathsf{E}\left(\frac{S_n(d)}{dN_n(d)}\middle|\bar{Q}\right)\mathsf{P}(\bar{Q}).$$

Let us estimate the first term:

$$\mathsf{E}\left(\frac{S_n(d)}{d\,N_n(d)}\middle|Q\right)\mathsf{P}(Q) = \frac{\mathsf{E}\left(S_n(d)\middle|Q\right)\mathsf{P}(Q)}{d\left(\mathsf{E}N_n(d)+O\left(d\sqrt{n}\log(n)\right)\right)} = \frac{\mathsf{E}S_n(d)-\mathsf{E}(S_n(d)\middle|\bar{Q})\mathsf{P}(\bar{Q})}{d\left(\mathsf{E}N_n(d)+O\left(d\sqrt{n}\log(n)\right)\right)} = \frac{\mathsf{E}S_n(d)+O\left(n^{2-\log(n)}\right)}{d\left(\mathsf{E}N_n(d)+O\left(d\sqrt{n}\log(n)\right)\right)}.$$

Here we used that $S_n(d) = O(n^2)$. The second term can be estimated as:

$$\mathsf{E}\left(\frac{S_n(d)}{d\,N_n(d)}\middle|\bar{Q}\right)\mathsf{P}(\bar{Q}) = O\left(\frac{n^2}{d}\right)\mathsf{P}(\bar{Q}) = O\left(\frac{n^{2-\log(n)}}{d}\right).$$

Finally,

$$\mathsf{E}d_{nn}(d) = \frac{M(d)\left(n+O\left(n^{2A+\varepsilon}\cdot d^{2+\frac{1}{A}}\right)\right)+O\left(n^{2-\log(n)}\right)}{d\left(c(m,d)\left(n+O\left(d^{2+\frac{1}{A}}\right)\right)+O\left(d\sqrt{n}\log(n)\right)\right)} + O\left(\frac{n^{2-\log(n)}}{d}\right) = \frac{M(d)}{d\,c(m,d)}\left(1+O\left(\frac{n^{2A+\varepsilon}\cdot d^{2+\frac{1}{A}}}{n}+\frac{d^{2+\frac{1}{A}}\log(n)}{\sqrt{n}}\right)\right).$$

References

1. Albert, R., Barabási, A.-L.: Statistical mechanics of complex networks. Rev. Mod. Phys. **74**, 47–97 (2002)
2. Bansal, S., Khandelwal, S., Meyers, L.A.: Exploring biological network structure with clustered random networks. BMC Bioinform. **10**, 405 (2009)
3. Barabási, A.L., Albert, R.: Emergence of scaling in random networks. Science **286**, 509–512 (1999)
4. Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., Hwang, D.-U.: Complex networks: structure and dynamics. Phys. Rep. **424**(45), 175–308 (2006)
5. Bollobás, B., Riordan, O.M.: Mathematical results on scale-free random graphs. In: Handbook of Graphs, Networks: From the Genome to the Internet, pp. 1–34 (2003)
6. Bollobás, B., Riordan, O.M., Spencer, J., Tusnády, G.: The degree sequence of a scale-free random graph process. Random Struct. Algorithms **18**(3), 279–290 (2001)
7. Borgs, C., Brautbar, M., Chayes, J., Khanna, S., Lucier, B.: The power of local information in social networks. In: Goldberg, P.W. (ed.) WINE 2012. LNCS, vol. 7695, pp. 406–419. Springer, Heidelberg (2012). doi:10.1007/978-3-642-35311-6_30
8. Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., Tomkins, A., Wiener, J.: Graph structure in the web. Comput. Netw. **33**(16), 309–320 (2000)
9. Buckley, P.G., Osthus, D.: Popularity based random graph models leading to a scale-free degree sequence. Discret. Math. **282**, 53–63 (2004)
10. Echenique, P., Gómez-Gardeñes, J., Moreno, Y., Vázquez, A.: Distance-d covering problems in scale-free networks with degree correlations. Phys. Rev. E **71**, 035102(R) (2005)
11. Faloutsos, M., Faloutsos, P., Faloutsos, C.: On power-law relationships of the Internet topology. In: Proceedings of the SIGCOMM 1999 (1999)
12. Girvan, M., Newman, M.E.: Community structure in social and biological networks. Proc. National Acad. Sci. **99**(12), 7821–7826 (2002)
13. van der Hofstad, R., Litvak, N.: Degree-degree dependencies in random graphs with heavy-tailed degrees. Internet Math. **10**(3–4), 287–334 (2014)
14. Holme, P., Kim, B.J.: Growing scale-free networks with tunable clustering. Phys. Rev. E **65**(2), 026107 (2002)
15. van der Hoorn, P., Litvak, N.: Degree-degree dependencies in directed networks with heavy-tailed degrees. Internet Math. **11**(2), 155–178 (2015)
16. 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, Heidelberg (2015). doi:10.1007/978-3-319-26784-5_2
17. Leskovec, J.: Dynamics of Large Networks. ProQuest (2008)
18. Newman, M.E.J.: Assortative mixing in networks. Phys. Rev. Lett. **89**(20), 208701 (2002)
19. Newman, M.E.J.: Power laws, Pareto distributions and Zipf's law. Contemp. Phys. **46**(N5), 323–351 (2005)
20. Newman, M.E.J.: The structure and function of complex networks. SIAM Rev. **45**(2), 167–256 (2003)

21. Ostroumova, L., Ryabchenko, A., Samosvat, E.: Generalized preferential attachment: tunable power-law degree distribution and clustering coefficient. In: Bonato, A., Mitzenmacher, M., Prałat, P. (eds.) WAW 2013. LNCS, vol. 8305, pp. 185–202. Springer, Heidelberg (2013). doi:10.1007/978-3-319-03536-9_15
22. Watts, D.J., Strogatz, S.H.: Collective dynamics of 'small-world' networks. Nature **393**, 440–442 (1998)
23. 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)

Springer

http://www.springer.com/978-3-319-49786-0

Algorithms and Models for the Web Graph
13th International Workshop, WAW 2016, Montreal, QC, Canada, December 14–15, 2016, Proceedings
Bonato, A.; Graham, F.C.; Pralat, P. (Eds.)
2016, X, 165 p. 42 illus., Softcover
ISBN: 978-3-319-49786-0