


RESEARCH ARTICLE

Generating preferential attachment graphs via a Pólya urn with expanding colors

Somya Singh¹, Fady Alajaji² and Bahman Ghahesifard³

¹ICTEAM Institute, UCL, Louvain-la-Neuve, Belgium, ²Department of Mathematics and Statistics, Queen's University, Kingston, Canada, and ³Electrical and Computer Engineering Department, UCLA, Los Angeles, CA, USA

Corresponding author: Somya Singh; Email: somya.singh@uclouvain.be

Abstract

We introduce a novel preferential attachment model using the draw variables of a modified Pólya urn with an expanding number of colors, notably capable of modeling influential opinions (in terms of vertices of high degree) as the graph evolves. Similar to the Barabási-Albert model, the generated graph grows in size by one vertex at each time instance; in contrast however, each vertex of the graph is uniquely characterized by a color, which is represented by a ball color in the Pólya urn. More specifically at each time step, we draw a ball from the urn and return it to the urn along with a number of reinforcing balls of the same color; we also add another ball of a new color to the urn. We then construct an edge between the new vertex (corresponding to the new color) and the existing vertex whose color ball is drawn. Using color-coded vertices in conjunction with the time-varying reinforcing parameter allows for vertices added (born) later in the process to potentially attain a high degree in a way that is not captured in the Barabási-Albert model. We study the degree count of the vertices by analyzing the draw vectors of the underlying stochastic process. In particular, we establish the probability distribution of the random variable counting the number of draws of a given color which determines the degree of the vertex corresponding to that color in the graph. We further provide simulation results presenting a comparison between our model and the Barabási-Albert network.

Keywords: Pólya urn; preferential attachment; reinforcement stochastic processes; degree distribution of random graphs

1. Introduction

Preferential attachment graphs are an important class of randomly generated graphs which are often used to capture the “rich gets richer” phenomenon. This class of random graphs has been widely studied within the areas of statistical mechanics (Hartmann and Weigt, 2001; Tsallis and Oliveira, 2022), network science (Csányi and Szendrői, 2004), probability theory (König, 2005; Rácz and Sridhar, 2022), and game theory (Santos and Pacheco, 2005). One of the most popular model of a preferential attachment graph is the so-called Barabási-Albert model (Barabási and Albert, 1999), which has since been modified in a variety of ways (Zhongdong et al., 2012; Lin et al. 2006; Alves et al., 2021). Various other models have been devised thereafter to generate preferential attachment graphs, for example in Berger et al. (2005b) the growth of the random graph is competition-based. Given a graph at a certain time step, the new vertex attaches itself to an existing vertex which ends up minimizing a certain cost function. For a vertex, this cost function depends on its centrality and distance from the root ensuring that the vertices with higher degrees have lower cost functions. In Krapivsky et al. (2000), a continuous-time equation governing the number of vertices with degree k is formulated to study citation networks. In Flaxman et al. (2006), a randomly growing graph algorithm that combines the features of a geometric random graph and a preferential attachment graph is analyzed. In Capocci et al. (2006), properties

of Wikipedia are studied by representing topics as vertices and hyperlinks between them as edges. Several preferential attachment hypergraphs (i.e., graphs in which an edge can join any number of vertices) generating models have also been devised in the literature (Avin et al., 2019; Giroire et al., 2022; Inoue et al., 2022).

Our main objective in this paper is to introduce a new preferential attachment graph-generating algorithm using a modified Pólya urn model. In the classical two-color Pólya urn model, at each time instant t , a ball is drawn from the urn and returned to the urn along with one ball of the color drawn (Pólya, 1930; Mahmoud, 2008; Pemantle, 2007). In this sense, the Pólya urn model is a suitable reinforcement process for modeling preferential phenomena. In particular, in the context of randomly growing graphs, the reinforcements favor a high-degree vertex in getting an even higher degree as the network grows. Indeed, Pólya urn models have been widely used to model reinforcements, for instance, in communication channels (Alajaji and Fuja, 1994), image segmentation (Banerjee et al., 1999), social and epidemic networks (Avin et al., 2020; Kim and Jo, 2010; Toivonen et al., 2006; Hayhoe et al., 2018; Singh et al., 2022c, a, b; Jadbabaie et al., 2023; He et al., 2023), citation networks (Jeong et al., 2003; Milojević, 2010; Newman, 2001), actor collaborations (Albert and Barabási, 2000; Jeong et al., 2003), mechanics (Albert and Barabási, 2002), and polymer formations (Bhutani et al., 2022). Naturally, Pólya urns have been used to model preferential attachment graphs in the literature. For example, in Collevocchio et al. (2013) a generalized Pólya urn process is used to devise a preferential attachment graph-generating algorithm (refer to Chung et al., 2003; Oliveira, 2009 for a detailed description of this generalized Pólya urn process.) In Berger et al. (2014), a preferential attachment-type multi-graph (i.e., a graph that can have more than one edge between a pair of vertices) is constructed using different variations of the Pólya urn process which was used to study the spread of viruses on the internet in Berger et al. (2005a). In Marcaccioli and Livan (2019), the probability distribution of weights on the edges of a fixed network is established via the draws of the classical two-color Pólya urn. This setup ensures that the more two vertices have interacted in the past, the more likely they are to interact in the future. More elaboration on the similarities between Pólya urns and preferential attachment graphs is given in the survey in Pemantle (2007).

The Barabási-Albert model is well known to exhibit a power law distribution as the number of vertices becomes sufficiently large, given by $p(k) \sim k^{-3}$, where $p(k)$ is the probability of randomly selecting a vertex with degree k in the network. Despite the fact that this power law can be used to study various properties of the Barabási-Albert model such as Hirsch index distribution and the clustering coefficient, see Barabási (2009), Bollobás and Riordan (2004), Albert and Barabási (2000) for definitions, the likelihood of vertices gaining new edges is solely determined by their degree. This is not realistic, when modeling scenarios where newly added individuals are accompanied with *impactful ideas that can lead to rapid or disruptive influence, regardless of their initially low degree*.

Motivated by the above-mentioned shortcoming of the Barabási-Albert model, in this paper, we construct randomly growing undirected graphs using the draw variables of a single modified Pólya urn with an expanding number of colors. The Pólya process is modified in the sense that at each time instant, not only is a ball drawn and returned to the urn along with additional reinforcing balls of the same color but another ball of a new color is also added to the urn. This new color corresponds to a new vertex which is added to the graph at this time instant. More specifically, the network is generated by associating each incoming vertex to the new color ball added after each draw and by attaching it to the existing vertex represented by the drawn color. The number of colors in the urn grows without bound with the number of draws, and the generated network has a preferential attachment property as the vertices corresponding to dominant colors (i.e., colors in the urn with a large number of balls) are more likely to attract newly formed vertices as their neighbors. The resulting preferential attachment growing graph is thus constructed via a Pólya urn; this enables us to track and characterize the degree count of individual vertices in the network through the draw variables of their corresponding colors giving each vertex a unique

identity which is absent in the Barabási-Albert model. Moreover, using an expanding color Pólya urn with each vertex uniquely corresponding to a ball color sets our model apart from other models in the literature (Collecchio et al., 2013; Marcaccioli and Livan, 2019; Berger et al., 2005a) which use either the classical two-color Pólya urns or a generalized version of Pólya urns (with finitely many colors) for generating preferential attachment graphs. Indeed, the draw variables of the Pólya urn capture the entire structure of the generated graph and hence it is enough to study the behavior of these draw variables to understand the properties of the graph. Moreover, we use an extra time-varying parameter to set the number of balls (not necessarily an integer) added to the Pólya urn to reinforce the color of the drawn ball at each time instant. The time-varying nature of this reinforcement parameter allows us for any given vertex to tweak the likelihood of amplifying or dampening its degree of growth depending on the time at which it was introduced in the network (i.e., its birth time). This feature can be used to regulate the dominance (in terms of gaining edges) of high-degree vertices over low-degree ones in the generated preferential attachment graph. Therefore, unlike the Barabási-Albert algorithm and the aforementioned models, our model can be used to generate random networks with a variety of degree distributions catering to a wide range of real-world growing networks other than power law distributions, including the generation of networks where recently formed vertices can play a disruptive role. Furthermore, we observe that in the special case of using a reinforcement parameter equal to 1, our model is actually equivalent to the Barabási-Albert algorithm (except for the initialization step).

This paper is organized as follows. In Section 2, we describe our model for constructing preferential attachment-type graphs using a modified Pólya urn with expanding colors. We determine the Pólya urn's composition in terms of its draw random variables and use it to establish the conditional distribution of the urn's draw vector at a given time instant given all past draw vectors. In Section 3, we define a counting random variable which tracks the degree of vertices in the graph as it evolves and analytically derives its probability distribution. We further verify that the resulting distribution expression of the counting random variable defines a legitimate probability mass function over its support set. In Section 4, we present via simulations a detailed comparison of our model with the Barabási-Albert model by plotting the degree distribution and vertices' average birth time for various choices of reinforcement parameters. We discuss the advantages of generating preferential attachment networks through our model over the Barabási-Albert model. Finally, conclusions are stated in Section 5.

2. The model

We construct a sequence of undirected graphs \mathcal{G}_t , where $t \geq 0$ denotes the time index, using a Pólya reinforcement process. We start with $\mathcal{G}_0 = (V_0, \mathcal{E}_0)$, where the initial vertex and the edge set are respectively, $V_0 = \{c_1\}$ and $\mathcal{E}_0 = \{(1, 1)\}$, i.e., a self-loop on vertex 1. At each time step $t \geq 1$, a new vertex enters the graph and forms an edge with one of the existing vertices. The latter vertex is selected according to the draw variable of a Pólya urn with an expanding number of colors as follows:

- At time $t = 0$, the Pólya urn consists of a single ball of color c_1 .
- At each time instant $t \geq 1$, we draw a ball and return it to the urn along with $\Delta_t > 0$ additional (reinforcing) balls of the same color. We also add a ball of a new color c_{t+1} . We then introduce a new vertex to the graph \mathcal{G}_{t-1} (corresponding to the color c_{t+1}) and connect it with the vertex whose color ball is drawn at time t . This results in the newly formed graph \mathcal{G}_t . Note that at time $t = 0$, the urn consists of only one c_1 color ball. Hence, the draw variable at time $t = 1$ is deterministic and corresponds to drawing a c_1 color ball.

At any given time instant t , we define the draw random vector

$$\mathbf{Z}_t := (Z_{1,t}, Z_{2,t}, \dots, Z_{t,t})$$

of length t , where

$$Z_{j,t} = \begin{cases} 1 & \text{if a } c_j \text{ color ball is drawn at time } t \\ 0 & \text{otherwise} \end{cases} \quad \text{for } 1 \leq j \leq t. \tag{1}$$

The vector \mathbf{Z}_t is a standard unit vector for all time instances $t \geq 1$, and since at time $t = 1$ there is only c_1 color ball present in the urn, $\mathbf{Z}_1 = Z_{1,1} = 1$. We denote the ‘‘composition’’ of the Pólya urn at any given time instant t by the random vector

$$\mathbf{U}_t := (U_{1,t}, U_{2,t}, \dots, U_{t+1,t}),$$

where

$$U_{j,t} = \frac{\text{Number of balls of color } c_j \text{ in the urn at time } t}{\text{Total number of balls in the urn at time } t}, \tag{2}$$

for $1 \leq j \leq t + 1$. In the following lemma, we express the vector \mathbf{U}_t in terms of the draw variables.

Lemma 1. *Given $t \geq 0$, \mathbf{U}_t is given by*

$$\mathbf{U}_t = \frac{1}{1 + t + \sum_{k=1}^t \Delta_k} \left(1 + \sum_{n=1}^t \Delta_n Z_{1,n}, 1 + \sum_{n=2}^t \Delta_n Z_{2,n}, \dots, 1 + \sum_{n=t-1}^t \Delta_n Z_{t-1,n}, 1 + \Delta_t Z_{t,t}, 1 \right) \tag{3}$$

*almost surely.*¹

Proof. To compute the ratio in (2), recall that at time $n = 0$, we have one ball in the urn (this ball is of color c_1) and for each time instant $n \geq 1$, we add $\Delta_n + 1$ balls to the urn (Δ_n of the color drawn and 1 of the new color c_{n+1}). Hence the total number of balls in the urn at time t is given by $1 + \sum_{n=1}^t (\Delta_n + 1)$.

To determine the number of balls of color c_j in the urn after the t th draw, we note that the first c_j color ball is added to the urn at time $j - 1$. After that, at every time instant n (where $j \leq n \leq t$) at which a c_j color ball is drawn, we add Δ_n balls of c_j color to the urn. Hence, the number of balls of color c_j in the urn at time t is equal to $1 + \sum_{n=j}^t \Delta_n Z_{j,n}$. Therefore, the ratio of color c_j balls in the urn at time t in (2) is given by:

$$U_{j,t} = \frac{1 + \sum_{n=j}^t \Delta_n Z_{j,n}}{1 + t + \sum_{k=1}^t \Delta_k} \quad \text{for } 1 \leq j \leq t + 1, \tag{4}$$

which yields (3). □

Remark 1. As expected, the sum of the components of \mathbf{U}_t in (3) is one for all $t \geq 0$. To see this we first note that, for $t = 0$, $\mathbf{U}_0 = U_{1,0} = 1$. For any time $t \geq 1$, we have the following from (3):

$$\begin{aligned} \sum_{j=1}^{t+1} U_{j,t} &= \frac{1}{1 + t + \sum_{k=1}^t \Delta_k} \left((t + 1) + \sum_{n=1}^t \Delta_n Z_{1,n} + \sum_{n=2}^t \Delta_n Z_{2,n} + \dots + \sum_{n=t-1}^t \Delta_n Z_{t-1,n} + \Delta_t Z_{t,t} \right) \\ &= \frac{1}{1 + t + \sum_{k=1}^t \Delta_k} \left((t + 1) + \sum_{i=1}^t \sum_{n=i}^t \Delta_n Z_{i,t} \right) \\ &= \frac{1}{1 + t + \sum_{k=1}^t \Delta_k} \left((t + 1) + \sum_{n=1}^t \Delta_n \sum_{i=n}^t Z_{i,t} \right) \end{aligned} \tag{5}$$

but since \mathbf{Z}_t is a standard unit vector for all $t \geq 1$, the right-hand side of (5) simplifies as follows:

$$\sum_{j=1}^{t+1} U_{j,t} = \frac{1}{1 + t + \sum_{k=1}^t \Delta_k} \left((t + 1) + \sum_{n=1}^t \Delta_n \right) = 1.$$

An illustration of our model is given in Fig. 1 where we show a sample path of the random vectors \mathbf{U}_t and \mathbf{Z}_t , for $t \leq 3$ and with $\Delta_t = 2$.

We further write the conditional probabilities of the draw variables given in the past. More specifically, for $1 \leq j \leq t$, using (4), we have that

$$\begin{aligned} P(\mathbf{Z}_t = \mathbf{e}_{j,t} \mid \mathbf{Z}_{t-1}, \mathbf{Z}_{t-2}, \dots, \mathbf{Z}_1) &= P(Z_{j,t} = 1 \mid \mathbf{Z}_{t-1}, \mathbf{Z}_{t-2}, \dots, \mathbf{Z}_1) \\ &= P(a_j \text{ color ball is drawn at time } t \mid \mathbf{Z}_{t-1}, \mathbf{Z}_{t-2}, \dots, \mathbf{Z}_1) \\ &= U_{j,t-1} = \frac{1 + \sum_{n=j}^{t-1} \Delta_n Z_{j,n}}{1 + (t - 1) + \sum_{k=1}^{t-1} \Delta_k}, \end{aligned} \tag{6}$$

where $\mathbf{e}_{j,t}$ represents a standard unit vector of length t whose j th component is 1. Considering the case where $j = t$ in (6) and the convention that $\sum_{n=t}^{t-1} \Delta_n Z_{t,n} = 0$, we obtain

$$U_{t,t-1} = \frac{1 + \sum_{n=t}^{t-1} \Delta_n Z_{t,n}}{t + \sum_{k=1}^{t-1} \Delta_k} = \frac{1}{t + \sum_{k=1}^{t-1} \Delta_k} = P(Z_{t,t} = 1)$$

and hence

$$P(Z_{t,t} = 1 \mid \mathbf{Z}_{t-1}, \mathbf{Z}_{t-2}, \dots, \mathbf{Z}_1) = P(Z_{t,t} = 1) = \frac{1}{t + \sum_{k=1}^{t-1} \Delta_k}, \tag{7}$$

i.e., the conditional probability of drawing a ball of color c_t at time t equals the marginal probability of drawing a ball of color c_t at time t . Similarly, we have that

$$P(Z_{t,t} = 0 \mid \mathbf{Z}_{t-1}, \mathbf{Z}_{t-2}, \dots, \mathbf{Z}_1) = P(Z_{t,t} = 0) = \frac{t - 1 + \sum_{n=1}^{t-1} \Delta_n}{t + \sum_{k=1}^{t-1} \Delta_k}, \tag{8}$$

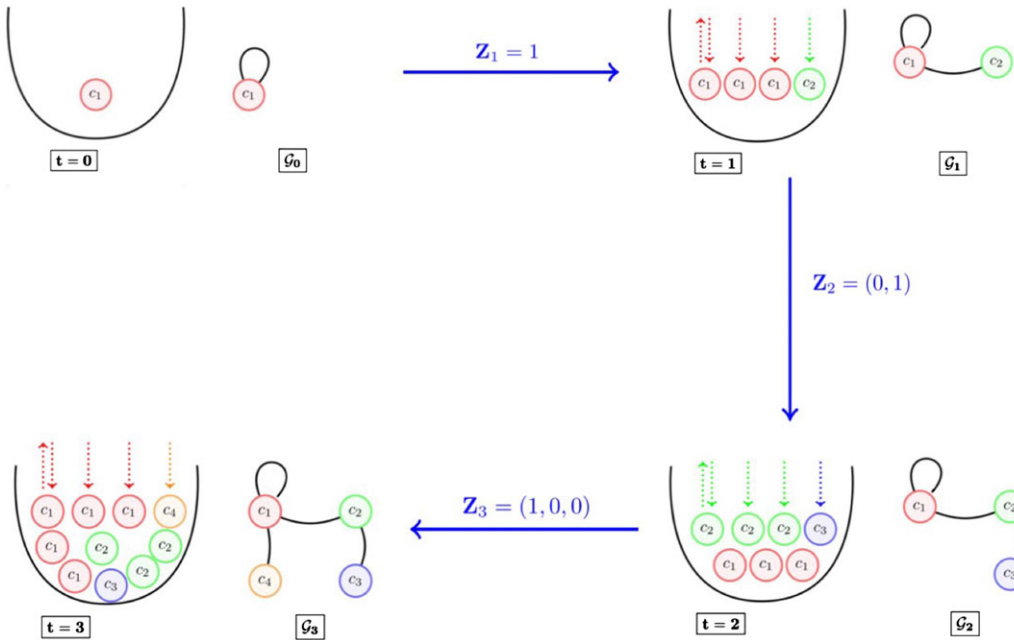


Figure 1. We illustrate a sample path for constructing a preferential attachment graph using an expanding color Pólya urn with $\Delta_t = 2$. For $t = 0$, the urn has only one ball of color c_1 . This urn corresponds to G_0 and $\mathbf{U}_0 = U_{1,0} = 1$. For $t = 1$, the c_1 color ball is drawn from and returned to the urn (i.e., $\mathbf{Z}_1 = Z_{1,1} = 1$). Two additional c_1 color balls are added to the urn along with a new c_2 color ball and so $\mathbf{U}_1 = (3/4, 1/4)$. For $t = 2$, a c_2 color ball is drawn from and returned to the urn (i.e., $\mathbf{Z}_2 = (0, 1)$). Two additional c_2 color balls are added to the urn along with a new c_3 color ball; hence $\mathbf{U}_2 = (3/7, 3/7, 1/7)$. For $t = 3$, a c_1 color ball is drawn from and returned to the urn (i.e., $\mathbf{Z}_3 = (1, 0, 0)$). Two additional c_1 color balls are added along with a new c_4 color ball; thus $\mathbf{U}_3 = (5/10, 3/10, 1/10, 1/10)$.

implying that $Z_{t,t}$ is independent of the random vectors $\{\mathbf{Z}_{t-1}, \mathbf{Z}_{t-2}, \dots, \mathbf{Z}_1\}$. More generally, we obtain the marginal probability for the random variable $Z_{j,t}$ for any $1 \leq j \leq t$ by taking expectation on both sides in (6) with respect to the random vectors $\mathbf{Z}_{t-1}, \mathbf{Z}_{t-2}, \dots, \mathbf{Z}_1$ as follows:

$$P(Z_{j,t} = 1) = E(U_{j,t-1}) = \frac{1 + \sum_{n=j}^{t-1} \Delta_n P(Z_{j,n} = 1)}{t + \sum_{k=1}^{t-1} \Delta_k} = 1 - P(Z_{j,t} = 0) \quad \text{for } 1 \leq j \leq t. \quad (9)$$

For $j = t$, the formula in (9) for $P(Z_{t,t} = 1)$ reduces to (7), but for $j < t$, the formula for $P(Z_{j,t} = 1)$ is a recursive function of the marginal probabilities of past draw variables: $P(Z_{j,1} = 1), \dots, P(Z_{j,t-1} = 1)$.

We further note that, for graph G_t , the edge between the new vertex to one of the existing vertices in G_{t-1} is made using the realization of the draw vector \mathbf{Z}_t . Using (6), we observe that the conditional probability $P(\mathbf{Z}_t = \mathbf{e}_{j,t} | \mathbf{Z}_{t-1}, \dots, \mathbf{Z}_1)$ can be written in terms of the draw variables $Z_{j,j}, \dots, Z_{j,t-1}$. Hence, all the spatial information of the graph G_t is encoded in the sequence of random draw vectors $\{\mathbf{Z}_1, \dots, \mathbf{Z}_{t-1}, \mathbf{Z}_t\}$. We illustrate this property in the following example, where we retrieve the graph G_4 using $\{\mathbf{Z}_1, \mathbf{Z}_2, \mathbf{Z}_3, \mathbf{Z}_4\}$.

Example 1. Consider the following realizations for the random draw vectors $\mathbf{Z}_1, \mathbf{Z}_2, \mathbf{Z}_3$ and \mathbf{Z}_4 for all $t \geq 1$:

$$\begin{aligned} \mathbf{Z}_1 &= 1, & \mathbf{Z}_2 &= (1, 0), \\ \mathbf{Z}_3 &= (0, 1, 0), & \mathbf{Z}_4 &= (0, 1, 0, 0). \end{aligned}$$

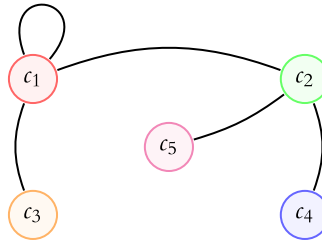


Figure 2. An illustration of how the sequence of draw vectors $\{\mathbf{Z}_4 = (0, 1, 0, 0), \mathbf{Z}_3 = (0, 1, 0), \mathbf{Z}_2 = (1, 0), \mathbf{Z}_1 = 1\}$ determines G_4 .

By construction, the graph G_0 consists of only one vertex c_1 with a self-loop. Since we start with only one ball of color c_1 in the urn, the random variable $\mathbf{Z}_1 = 1$ is deterministic and results in an edge drawn between the c_1 vertex and the new incoming vertex c_2 . For $t = 2$, since $\mathbf{Z}_2 = (1, 0)$, the new incoming vertex c_3 is connected to c_1 . Similarly for $t = 3$, the new vertex c_4 is connected to c_2 and finally, for $t = 4$, using the value of \mathbf{Z}_4 , we connect c_5 to c_2 . Hence, the graph G_4 is as shown in Fig. 2.

3. Analyzing the degree count of the vertices in G_t

The goal of this section is to establish a formula for the probability distribution of degree count of a fixed vertex in the preferential attachment graph constructed via our modified Pólya process until time t (including time t). We obtain this formula by writing the degree of a fixed vertex at time t in terms of the total number of balls of color corresponding to this vertex drawn until time t . To this end, let random variable $N_{j,t}$ count the number of draws of color c_j from the urn until time t (including time t). Since by construction, for a fixed color c_j , at any time $t \geq j - 1$ the degree of vertex c_j at time t , denoted by $d_{j,t}$, is one more than the number of times a c_j color ball is drawn from the urn until time t ; the additional one here is due to the fact that at each time instant, the new vertex which is added has degree one. For instance in Fig. 1, the color c_2 (green) is drawn once until time 3 and hence the degree of vertex corresponding to color c_2 in G_3 is two. Therefore,

$$d_{j,t} = 1 + N_{j,t} \quad \text{for all } 1 \leq j \leq t + 1,$$

where $N_{t+1,t} = 0$. Also, note that the first time a color c_j ball can be drawn is at time j . In the following theorem, we establish an analytical expression for the (marginal) probability mass function of random variable $N_{j,t}$.

Theorem 1. Fix $t \geq 1$. For a color c_j , $1 \leq j \leq t$, we have that

$$P(N_{j,t} = k) = \begin{cases} \sum_{(i_1, i_2, \dots, i_k) \in \mathcal{A}_{j,k}^{(t)}} \frac{\prod_{a=1}^k \left(1 + \sum_{b=1}^{a-1} \Delta_{i_b}\right) \prod_{p=j, p \notin \{i_1, \dots, i_k\}}^t \left((p-1) + \sum_{l=1, l \notin \{i_1, \dots, i_k\}}^{p-1} \Delta_l\right)}{\prod_{n=j-1}^{t-1} \left((n+1) + \sum_{m=1}^n \Delta_m\right)} & \text{for } 1 \leq k \leq t - j + 1 \\ \frac{\prod_{p=j}^t \left((p-1) + \sum_{l=1}^{p-1} \Delta_l\right)}{\prod_{n=j-1}^{t-1} \left((n+1) + \sum_{m=1}^n \Delta_m\right)} & \text{for } k = 0, \end{cases} \tag{10}$$

where

$$\mathcal{A}_{j,k}^{(t)} = \begin{cases} \{(i_1, i_2, \dots, i_k) \mid 1 = i_1 < i_2 < \dots < i_k \leq t\} & \text{for } j = 1 \\ \{(i_1, i_2, \dots, i_k) \mid j \leq i_1 < i_2 < \dots < i_k \leq t\} & \text{for } 1 < j \leq t. \end{cases} \tag{11}$$

Proof. Note that the set $\mathcal{A}_{j,k}^{(t)}$ defined in (11) gives all possible ways in which k elements can be chosen from a set of $t - j + 1$ consecutive integers. In the context of our model, this set represents all possible k length tuples of time instants such that a color c_j ball is drawn at each of these time instants. For $j = 1$, the first draw at $t = 1$ is deterministic and hence $i_1 = 1$ for $j = 1$ as given in (11). For $1 \leq k \leq t - j + 1$, we have

$$\begin{aligned} P(N_{j,t} = k) &= \sum_{(i_1, \dots, i_k) \in \mathcal{A}_{j,k}^{(t)}} P(Z_{j,j} = 0, \dots, Z_{j,i_1-1} = 0, Z_{j,i_1} = 1, Z_{j,i_1+1} = 0, \dots, Z_{j,i_2-1} = 0, Z_{j,i_2} \\ &= 1, Z_{j,i_2+1} = 0, \dots, Z_{j,i_3-1} = 0, Z_{j,i_3} = 1, \dots, Z_{j,i_k-1} = 0, Z_{j,i_k} = 1, Z_{j,i_k+1} \\ &= 0, \dots, Z_{j,t} = 0) \\ &= \sum_{(i_1, \dots, i_k) \in \mathcal{A}_{j,k}^{(t)}} [P(Z_{j,t} = 0 \mid Z_{j,j} = 0, \dots, Z_{j,i_1-1} = 0, Z_{j,i_1} = 1, Z_{j,i_1+1} = 0, \dots, Z_{j,i_2-1} \\ &= 0, Z_{j,i_2} = 1, Z_{j,i_2+1} = 0, \dots, Z_{j,i_3-1} = 0, Z_{j,i_3} = 1, \dots, Z_{j,i_k-1} = 0, Z_{j,i_k} = 1, Z_{j,i_k+1} \\ &= 0, \dots, Z_{j,t-1} = 0) \times P(Z_{j,j} = 0, \dots, Z_{j,i_1-1} = 0, Z_{j,i_1} = 1, Z_{j,i_1+1} = 0, \dots, Z_{j,i_2-1} \\ &= 0, Z_{j,i_2} = 1, Z_{j,i_2+1} = 0, \dots, Z_{j,i_3-1} = 0, Z_{j,i_3} = 1, \dots, Z_{j,i_k-1} = 0, Z_{j,i_k} = 1, Z_{j,i_k+1} \\ &= 0, \dots, Z_{j,t-1} = 0)]. \end{aligned} \tag{12}$$

By substituting (6) in the conditional probability expressions in (12), we obtain the following:

$$\begin{aligned} P(Z_{j,t} = 0 \mid Z_{j,j} = 0, \dots, Z_{j,i_1-1} = 0, Z_{j,i_1} = 1, Z_{j,i_1+1} = 0, \dots, Z_{j,i_2-1} = 0, Z_{j,i_2} = 1, Z_{j,i_2+1} \\ = 0, \dots, Z_{j,i_3-1} = 0, Z_{j,i_3} = 1, \dots, Z_{j,i_k-1} = 0, Z_{j,i_k} = 1, Z_{j,i_k+1} = 0, \dots, Z_{j,t-1} = 0) \\ = 1 - \frac{1 + \sum_{n=j}^{t-1} \Delta_n Z_{j,n}}{t + \sum_{m=1}^{t-1} \Delta_m} \\ = 1 - \frac{1 + \sum_{l=1}^k \Delta_{i_l}}{t + \sum_{m=1}^{t-1} \Delta_m} \\ = \frac{(t - 1) + \sum_{l=1, l \notin \{i_1, \dots, i_k\}}^{t-1} \Delta_l}{t + \sum_{m=1}^{t-1} \Delta_m}. \end{aligned} \tag{13}$$

Now, substituting the conditional probability expression obtained in (13) in (12) yields

$$\begin{aligned} P(N_{j,t} = k) &= \sum_{(i_1, \dots, i_k) \in \mathcal{A}_{j,k}^{(t)}} \frac{\left((t - 1) + \sum_{l=1, l \notin \{i_1, \dots, i_k\}}^{t-1} \Delta_l \right)}{t + \sum_{m=1}^{t-1} \Delta_m} P(Z_{j,j} = 0, \dots, Z_{j,i_1-1} = 0, Z_{j,i_1} \\ &= 1, Z_{j,i_1+1} = 0, \dots, Z_{j,i_2-1} = 0, Z_{j,i_2} = 1, Z_{j,i_2+1} = 0, \dots, Z_{j,i_3-1} = 0, Z_{j,i_3} \\ &= 1, \dots, Z_{j,i_k-1} = 0, Z_{j,i_k} = 1, Z_{j,i_k+1} = 0, \dots, Z_{j,t-1} = 0). \end{aligned} \tag{14}$$

Similar to (12) and (14), we continue to recursively write the joint probability as a product of conditional and marginal probabilities and substitute the expressions for the conditional probability using (13) as follows:

$$\begin{aligned}
 P(N_{j,t} = k) &= \sum_{(i_1, \dots, i_k) \in \mathcal{A}_{j,k}^{(t)}} \left(\frac{(t-1) + \sum_{l=1, l \notin \{i_1, \dots, i_k\}}^{t-1} \Delta_l}{t + \sum_{m=1}^{t-1} \Delta_m} \right) \left(\frac{(t-2) + \sum_{l=1, l \notin \{i_1, \dots, i_k\}}^{t-2} \Delta_l}{t-1 + \sum_{m=1}^{t-2} \Delta_m} \right) \\
 &P(Z_{j,j} = 0, \dots, Z_{j,i_1-1} = 0, Z_{j,i_1} = 1, Z_{j,i_1+1} = 0, \dots, Z_{j,i_2-1} = 0, Z_{j,i_2} = 1, Z_{j,i_2+1} \\
 &= 0, \dots, Z_{j,i_3-1} = 0, Z_{j,i_3} = 1, \dots, Z_{j,i_k-1} = 0, Z_{j,i_k} = 1, Z_{j,i_k+1} = 0, \dots, Z_{j,t-2} = 0) \\
 &\vdots \\
 &= \sum_{(i_1, \dots, i_k) \in \mathcal{A}_{j,k}^{(t)}} \left[\left(\frac{(t-1) + \sum_{l=1, l \notin \{i_1, \dots, i_k\}}^{t-1} \Delta_l}{t + \sum_{m=1}^{t-1} \Delta_m} \right) \left(\frac{(t-2) + \sum_{l=1, l \notin \{i_1, \dots, i_k\}}^{t-2} \Delta_l}{t-1 + \sum_{m=1}^{t-2} \Delta_m} \right) \dots \right. \\
 &\left(\frac{i_k + \sum_{l=1, l \notin \{i_1, \dots, i_k\}}^{i_k} \Delta_l}{i_k + 1 + \sum_{m=1}^{i_k} \Delta_m} \right) \left(\frac{1 + \sum_{l=1}^{k-1} \Delta_{i_l}}{i_k + \sum_{m=1}^{i_k-1} \Delta_m} \right) \left(\frac{i_k - 2 + \sum_{l=1, l \notin \{i_1, \dots, i_k\}}^{i_k-2} \Delta_l}{i_k - 1 + \sum_{m=1}^{i_k-1} \Delta_m} \right) \dots \\
 &\left(\frac{i_{k-1} + \sum_{l=1, l \notin \{i_1, \dots, i_k\}}^{i_{k-1}} \Delta_l}{i_{k-1} + 1 + \sum_{m=1}^{i_{k-1}} \Delta_m} \right) \left(\frac{1 + \sum_{l=1}^{k-2} \Delta_{i_l}}{i_{k-1} + \sum_{m=1}^{i_{k-1}-1} \Delta_m} \right) \left(\frac{i_{k-1} - 2 + \sum_{l=1, l \notin \{i_1, \dots, i_k\}}^{i_{k-1}-2} \Delta_l}{i_{k-1} - 1 + \sum_{m=1}^{i_{k-1}-1} \Delta_m} \right) \\
 &\dots \left(\frac{1}{i_1 + \sum_{m=1}^{i_1-1} \Delta_m} \right) \left(\frac{i_1 - 2 + \sum_{l=1, l \notin \{i_1, \dots, i_k\}}^{i_1-2} \Delta_l}{i_1 - 1 + \sum_{m=1}^{i_1-1} \Delta_m} \right) \dots \left. \left(\frac{j-1 + \sum_{l=1, l \notin \{i_1, \dots, i_k\}}^{j-1} \Delta_l}{j + \sum_{m=1}^{j-1} \Delta_m} \right) \right] \\
 &= \sum_{(i_1, \dots, i_k) \in \mathcal{A}_{j,k}^{(t)}} \left[\frac{\prod_{a=1}^k (1 + \sum_{b=1}^{a-1} \Delta_{i_b})}{\prod_{n=j-1}^{t-1} ((n+1) + \sum_{m=1}^n \Delta_m)} \left(\prod_{l_1=j-1}^{i_1-2} (l_1 + \sum_{l=1, l \notin \{i_1, \dots, i_k\}}^{l_1} \Delta_l) \right) \right. \\
 &\left(\prod_{l_2=i_1}^{i_2-2} (l_2 + \sum_{l=1, l \notin \{i_1, \dots, i_k\}}^{l_2} \Delta_l) \right) \left(\prod_{l_3=i_2}^{i_3-2} (l_3 + \sum_{l=1, l \notin \{i_1, \dots, i_k\}}^{l_3} \Delta_l) \right) \dots \\
 &\dots \left(\prod_{l_k=i_{k-1}}^{i_k-2} (l_k + \sum_{l=1, l \notin \{i_1, \dots, i_k\}}^{l_k} \Delta_l) \right) \left(\prod_{l_{k+1}=i_k}^{t-1} (l_{k+1} + \sum_{l=1, l \notin \{i_1, \dots, i_k\}}^{l_{k+1}} \Delta_l) \right) \left. \right] \\
 &= \sum_{(i_1, i_2, \dots, i_k) \in \mathcal{A}_{j,k}^{(t)}} \frac{\prod_{a=1}^k (1 + \sum_{b=1}^{a-1} \Delta_{i_b}) \prod_{p=j, p \notin \{i_1, \dots, i_k\}}^t ((p-1) + \sum_{l=1, l \notin \{i_1, \dots, i_k\}}^{p-1} \Delta_l)}{\prod_{n=j-1}^{t-1} ((n+1) + \sum_{m=1}^n \Delta_m)}.
 \end{aligned}$$

Therefore, (10) holds for $1 \leq k \leq t - j + 1$. We determine $P(N_{j,t} = 0)$ as follows:

$$\begin{aligned}
 P(N_{j,t} = 0) &= P(Z_{j,j} = 0, Z_{j,j+1} = 0, \dots, Z_{j,t-1} = 0, Z_{j,t} = 0) \\
 &= P(Z_{j,j} = 0) \prod_{n=j+1}^t P(Z_{j,n} = 0 \mid Z_{j,j} = 0, Z_{j,j+1} = 0, \dots, Z_{j,n-1} = 0). \tag{15}
 \end{aligned}$$

Now, using (6) for the conditional probabilities in (15) we obtain

$$\begin{aligned}
 P(N_{j,t} = 0) &= \left(1 - \frac{1}{j + \sum_{m=1}^{j-1} \Delta_m}\right) \left(1 - \frac{1}{j+1 + \sum_{m=1}^j \Delta_m}\right) \cdots \left(1 - \frac{1}{t + \sum_{m=1}^{t-1} \Delta_m}\right) \\
 &= \frac{\prod_{p=j}^t \left((p-1) + \sum_{l=1}^{p-1} \Delta_l\right)}{\prod_{n=j-1}^{t-1} \left((n+1) + \sum_{m=1}^n \Delta_m\right)}.
 \end{aligned}$$

Hence, (10) holds for $k = 0$. □

The analytic formula obtained in (10) is quite involved when the reinforcement parameter Δ_t is time-varying. For the special case of $\Delta_t = \Delta$ for all $t \geq 1$, Theorem 1 reduces to the following corollary.

Corollary 1. Fix $t \geq 1$. For a color c_j , $1 \leq j \leq t$ and $\Delta_n = \Delta$ for all $n \geq 1$, the marginal probability for $N_{j,t}$ is given by:

$$P(N_{j,t} = k) = \begin{cases} \frac{\sum_{(i_1, i_2, \dots, i_k) \in \mathcal{A}_{j,k}^{(t)}} \prod_{a=1}^k (1+(a-1)\Delta) \prod_{p=j, p \notin \{i_1, \dots, i_k\}}^t \left((p-1)(\Delta+1) - \Delta \sum_{l=1}^k \mathbb{1}(i_l \leq p-1) \right)}{\prod_{n=j-1}^{t-1} ((\Delta+1)n+1)} & \text{for } 1 \leq k \leq t-j+1 \\ \frac{\prod_{p=j}^t (p-1)(\Delta+1)}{\prod_{n=j-1}^{t-1} ((\Delta+1)n+1)} & \text{for } k = 0, \end{cases} \tag{16}$$

where the set $\mathcal{A}_{j,k}^{(t)}$ is defined in (11) and $\mathbb{1}(\mathcal{E})$ is the indicator function of the event \mathcal{E} .

The analytical expression in (16) can be further simplified for the case of $\Delta_t = 1$, $t \geq 1$, as follows.

Corollary 2. Fix $t \geq 1$. For a color c_j , $1 \leq j \leq t$ and $\Delta_n = 1$ for all $n \geq 1$, the marginal probability for $N_{j,t}$ is given by:

$$P(N_{j,t} = k) = \begin{cases} \frac{\sum_{(i_1, i_2, \dots, i_k) \in \mathcal{A}_{j,k}^{(t)}} \frac{\Gamma(k+1) \prod_{p=j}^t \left(2(p-1) - \sum_{l=1}^k \mathbb{1}(i_l \leq p-1) \right)}{\prod_{n=j-1}^{t-1} (2n+1)}}{\prod_{n=j-1}^{t-1} (2n+1)} & \text{for } 1 \leq k \leq t-j+1 \\ \frac{2\Gamma(t+1)}{\Gamma(j-1) \prod_{n=j-1}^{t-1} (2n+1)} & \text{for } k = 0, \end{cases} \tag{17}$$

where $\Gamma(\cdot)$ is the gamma function.

As we will see in Section 4, our model with $\Delta_t = 1$ for all $t \geq 0$ has exactly the same mechanism as the Barabási-Albert model (except for the initialization). Therefore, (17) can be used to predict the degree count of vertices in the Barabási-Albert model for sufficiently large t . To illustrate (17), we present a simulation in Fig. 3 for the probability mass function of the counting random variable $N_{2,12}$ for the case of $\Delta_t = 1$, $1 \leq t \leq 12$.

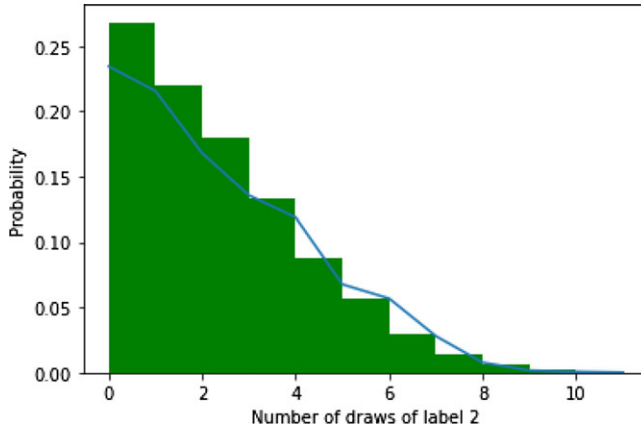


Figure 3. A simulation of the probability distribution given by (17) in Corollary 2 for the case of $\Delta_t = 1$ with $1 \leq t \leq 12$. A normalized histogram of the counting random variable $N_{2,12}$ from our model is plotted (by averaging over 1000 simulations) and is shown to concord with the curve of (17) (in blue).

Since the first time instant at which a c_j color ball can be drawn from the Pólya urn is at time j , the total number of draws of a c_j color ball till time t can be at most $t - j + 1$. Therefore,

$$\sum_{k=0}^{t-j+1} P(d_{j,t} = k + 1) = \sum_{k=0}^{t-j+1} P(N_{j,t} = k) = 1,$$

which implies that $P(N_{j,t} = k)$ is a probability mass function on the support set $\{0, 1, \dots, t - j + 1\}$. We next only verify that $P(N_{j,t} = k)$ obtained in (16) does indeed sum up to one (over k ranging from zero to $t - j + 1$) and is hence a legitimate probability mass function. For simplicity, we focus on the case with $\Delta_t = \Delta$; the proof for the general case follows along similar lines. To this end, we write the set $\mathcal{A}_{j,k}^{(t)}$ as the following disjoint union:

$$\begin{aligned} \mathcal{A}_{j,k}^{(t)} &= \{(i_1, i_2, \dots, i_k) \mid j \leq i_1 < i_2 < \dots < i_k \leq t\} \\ &= \{(i_1, i_2, \dots, i_k) \mid j \leq i_1 < i_2 < \dots < i_k \leq t - 1\} \sqcup \mathcal{B}_{j,k}^{(t)} = \mathcal{A}_{j,k}^{(t-1)} \sqcup \mathcal{B}_{j,k}^{(t)}, \end{aligned} \tag{18}$$

where $\mathcal{B}_{j,k}^{(t)} := \{(i_1, \dots, i_{k-1}, t) \mid j \leq i_1 < \dots < i_{k-1} \leq t - 1\}$. Note that

$$\mathcal{B}_{j,k}^{(t)} = \{(i_1, \dots, i_{k-1}, t) \mid (i_1, \dots, i_{k-1}) \in \mathcal{A}_{j,k-1}^{(t-1)}\}. \tag{19}$$

Theorem 2. Fix $t \geq 1$. For a c_j , $1 \leq j \leq t$ and $\Delta_t = \Delta$ for all $t \geq 1$, we have that:

$$\sum_{k=0}^{t-j+1} P(N_{j,t} = k) = 1. \tag{20}$$

Proof. We write the left-hand side of (20) using (16):

$$\begin{aligned} & \frac{\prod_{p=j}^t (p - 1)(\Delta + 1)}{\prod_{n=j-1}^{t-1} ((\Delta + 1)n + 1)} \\ & + \sum_{k=1}^{t-j+1} \sum_{(i_1, \dots, i_k) \in \mathcal{A}_{j,k}^{(t)}} \frac{\prod_{a=1}^k (1 + (a - 1)\Delta) \prod_{p=j, p \notin \{i_1, \dots, i_k\}}^t ((p - 1)(\Delta + 1) - \Delta \sum_{l=1}^k \mathbf{1}(i_l \leq p - 1))}{\prod_{n=j-1}^{t-1} ((\Delta + 1)n + 1)}. \end{aligned} \tag{21}$$

Therefore, showing that (20) holds is equivalent to showing that:

$$\begin{aligned} & \prod_{p=j}^t (p-1)(\Delta+1) \\ & + \sum_{k=1}^{t-j+1} \sum_{(i_1, \dots, i_k) \in \mathcal{A}_{jk}^{(t)}} \left(\prod_{a=1}^k (1+(a-1)\Delta) \prod_{p=j, p \notin \{i_1, \dots, i_k\}}^t \left((p-1)(\Delta+1) - \Delta \sum_{l=1}^k \mathbb{1}(i_l \leq p-1) \right) \right) \\ & = \prod_{n=j-1}^{t-1} ((\Delta+1)n+1). \end{aligned} \tag{22}$$

We prove (22) by induction on $t-j+1 \geq 1$.

Base Case: $t-j+1 = 1$ or $t=j$. For this case, the left-hand side of (22) is the following:

$$\begin{aligned} & \prod_{p=j}^j (p-1)(\Delta+1) \\ & + \sum_{\mathcal{A}_{j,1}^{(j)}} \left(\prod_{a=1}^1 (1+(a-1)\Delta) \prod_{p=j, p \neq j}^j \left((p-1)(\Delta+1) - \Delta \sum_{l=1}^1 \mathbb{1}(i_l \leq p-1) \right) \right) \\ & = (j-1)(\Delta+1) + 1 \end{aligned}$$

which, upon simplification and noting that the set $\mathcal{A}_{j,1}^{(j)} = \{j\}$, equals the right-hand side of (22) for $t-j+1 = 1$.

Induction Step: We now show the induction step: assuming that (22) is true for $t-j+1 = s$, we show that it holds for $t-j+1 = s+1$. We thus assume that the following holds:

$$\begin{aligned} & \prod_{p=j}^{j+s-1} ((p-1)(\Delta+1)) + \sum_{k=1}^s \sum_{\mathcal{A}_{jk}^{(j+s-1)}} \left(\prod_{a=1}^k (1+(a-1)\Delta) \prod_{p=j, p \notin \{i_1, \dots, i_k\}}^{j+s-1} \left((p-1)(\Delta+1) \right. \right. \\ & \qquad \qquad \qquad \left. \left. - \Delta \sum_{l=1}^k \mathbb{1}(i_l \leq p-1) \right) \right) \\ & = \prod_{n=j-1}^{j+s-2} ((\Delta+1)n+1). \end{aligned} \tag{23}$$

We next show the induction step using (23), by starting from the right-hand side:

$$\begin{aligned} & \prod_{n=j-1}^{j+s-1} ((\Delta+1)n+1) = ((\Delta+1)(s+j-1)+1) \prod_{n=j-1}^{j+s-2} ((\Delta+1)n+1) \\ & \stackrel{(a)}{=} ((\Delta+1)(s+j-1)+1) \prod_{p=j}^{j+s-1} (p-1)(\Delta+1) \end{aligned}$$

$$\begin{aligned}
 & + \sum_{k=1}^s ((\Delta + 1)(s + j - 1) + 1) \sum_{\mathcal{A}_{j,k}^{(j+s-1)}} \left(\prod_{a=1}^k (1 + (a - 1)\Delta) \prod_{p=j, p \notin \{i_1, \dots, i_k\}}^{j+s-1} ((p - 1)(\Delta + 1) \right. \\
 & \qquad \qquad \qquad \left. - \Delta \sum_{l=1}^k \mathbb{1}(i_l \leq p - 1) \right) \\
 \stackrel{(b)}{=} & \prod_{p=j}^{j+s} (p - 1)(\Delta + 1) + \prod_{p=j}^{j+s-1} (p - 1)(\Delta + 1) + \sum_{k=1}^s ((\Delta + 1)(s + j - 1) \\
 & - \Delta k + \Delta k + 1) \sum_{\mathcal{A}_{j,k}^{(j+s-1)}} \left(\prod_{a=1}^k (1 + (a - 1)\Delta) \prod_{p=j, p \notin \{i_1, \dots, i_k\}}^{j+s-1} \left((p - 1)(\Delta + 1) - \Delta \sum_{l=1}^k \mathbb{1}(i_l \leq p - 1) \right) \right) \\
 \stackrel{(c)}{=} & \prod_{p=j}^{j+s} (p - 1)(\Delta + 1) + \prod_{p=j}^{j+s-1} (p - 1)(\Delta + 1) \\
 & + \sum_{k=1}^s ((\Delta + 1)(s + j - 1) - \Delta k) \sum_{\mathcal{A}_{j,k}^{(j+s-1)}} \left(\prod_{a=1}^k (1 + (a - 1)\Delta) \prod_{p=j, p \notin \{i_1, \dots, i_k\}}^{j+s-1} ((p - 1)(\Delta + 1) \right. \\
 & \qquad \qquad \qquad \left. - \Delta \sum_{l=1}^k \mathbb{1}(i_l \leq p - 1) \right) \\
 & + \sum_{k=1}^s (\Delta k + 1) \sum_{\mathcal{A}_{j,k}^{(j+s-1)}} \left(\prod_{a=1}^k (1 + (a - 1)\Delta) \prod_{p=j, p \notin \{i_1, \dots, i_k\}}^{j+s-1} \left((p - 1)(\Delta + 1) - \Delta \sum_{l=1}^k \mathbb{1}(i_l \leq p - 1) \right) \right) \\
 \stackrel{(d)}{=} & \prod_{p=j}^{j+s} (p - 1)(\Delta + 1) + \prod_{p=j}^{j+s-1} (p - 1)(\Delta + 1) \\
 & + \sum_{k=1}^s \sum_{\mathcal{A}_{j,k}^{(j+s-1)}} \left(\prod_{a=1}^k (1 + (a - 1)\Delta) \prod_{p=j, p \notin \{i_1, \dots, i_k\}}^{j+s} \left((p - 1)(\Delta + 1) - \Delta \sum_{l=1}^k \mathbb{1}(i_l \leq p - 1) \right) \right) \\
 & + \sum_{k=1}^s \sum_{\mathcal{A}_{j,k}^{(j+s-1)}} \left(\prod_{a=1}^{k+1} (1 + (a - 1)\Delta) \prod_{p=j, p \notin \{i_1, \dots, i_k\}}^{j+s-1} \left((p - 1)(\Delta + 1) - \Delta \sum_{l=1}^k \mathbb{1}(i_l \leq p - 1) \right) \right) \\
 \stackrel{(e)}{=} & \prod_{p=j}^{j+s} (p - 1)(\Delta + 1) + \prod_{p=j}^{j+s-1} (p - 1)(\Delta + 1) \\
 & + \sum_{k=1}^s \sum_{\mathcal{A}_{j,k}^{(j+s-1)}} \left(\prod_{a=1}^k (1 + (a - 1)\Delta) \prod_{p=j, p \notin \{i_1, \dots, i_k\}}^{j+s} \left((p - 1)(\Delta + 1) - \Delta \sum_{l=1}^k \mathbb{1}(i_l \leq p - 1) \right) \right) \\
 & + \sum_{k=2}^{s+1} \sum_{\mathcal{A}_{j,k-1}^{(j+s-1)}} \left(\prod_{a=1}^k (1 + (a - 1)\Delta) \prod_{p=j, p \notin \{i_1, \dots, i_{k-1}\}}^{j+s-1} \left((p - 1)(\Delta + 1) - \Delta \sum_{l=1}^{k-1} \mathbb{1}(i_l \leq p - 1) \right) \right)
 \end{aligned}$$

$$\begin{aligned}
 & \stackrel{(f)}{=} \prod_{p=j}^{j+s} (p-1)(\Delta+1) + \prod_{p=j}^{j+s-1} (p-1)(\Delta+1) \\
 & + \sum_{k=1}^s \sum_{\mathcal{A}_{j,k}^{(j+s-1)}} \left(\prod_{a=1}^k (1+(a-1)\Delta) \prod_{p=j, p \notin \{i_1, \dots, i_k\}}^{j+s} \left((p-1)(\Delta+1) - \Delta \sum_{l=1}^k \mathbb{1}(i_l \leq p-1) \right) \right) \\
 & + \sum_{k=2}^{s+1} \sum_{\mathcal{B}_{j,k}^{(j+s)}} \left(\prod_{a=1}^k (1+(a-1)\Delta) \prod_{p=j, p \notin \{i_1, \dots, i_{k-1}, j+s\}}^{j+s} \left((p-1)(\Delta+1) - \Delta \sum_{l=1}^{k-1} \mathbb{1}(i_l \leq p-1) \right) \right) \\
 & \stackrel{(g)}{=} \prod_{p=j}^{j+s} (p-1)(\Delta+1) + \prod_{p=j}^{j+s-1} (p-1)(\Delta+1) \\
 & + \sum_{\mathcal{A}_{j,1}^{(j+s-1)}} \left(\prod_{a=1}^1 (1+(a-1)\Delta) \prod_{p=j, p \notin \{i_1\}}^{j+s} \left((p-1)(\Delta+1) - \Delta \sum_{l=1}^1 \mathbb{1}(i_l \leq p-1) \right) \right) \\
 & + \sum_{k=2}^s \sum_{\mathcal{A}_{j,k}^{(j+s)}} \left(\prod_{a=1}^k (1+(a-1)\Delta) \prod_{p=j, p \notin \{i_1, \dots, i_k\}}^{j+s} \left((p-1)(\Delta+1) - \Delta \sum_{l=1}^k \mathbb{1}(i_l \leq p-1) \right) \right) \\
 & \quad + \prod_{a=1}^{s+1} (1+(a-1)\Delta) \\
 & \stackrel{(h)}{=} \prod_{p=j}^{j+s} (p-1)(\Delta+1) \\
 & \quad + \sum_{k=1}^{s+1} \sum_{\mathcal{A}_{j,k}^{(j+s)}} \left(\prod_{a=1}^k (1+(a-1)\Delta) \prod_{p=j, p \notin \{i_1, \dots, i_k\}}^{j+s} \left((p-1)(\Delta+1) - \Delta \sum_{l=1}^k \mathbb{1}(i_l \leq p-1) \right) \right).
 \end{aligned}$$

In the above set of equations, we obtain (a) by substituting (23) in the left-hand side of (a). In (b), we add and subtract Δk to the term $((\Delta+1)(s+j-1)+1)$ and split the summation across the terms $((\Delta+1)(s+j-1)-\Delta k)$ and $(\Delta k+1)$ in (c). In (d), we absorb the terms $((\Delta k+1)(s+j-1)-\Delta k)$ and $(\Delta k+1)$ into the product. We replace k by $k-1$ in the fourth term on the left-hand side of (e). We obtain the fourth term on the right-hand side of (f) using (19). On the right-hand side of (g), the second term can be written as follows:

$$\begin{aligned}
 & \prod_{p=j}^{j+s-1} (p-1)(\Delta+1) \\
 & = \sum_{\mathcal{B}_{j,1}^{(j+s)}} \left(\prod_{a=1}^1 (1+(a-1)\Delta) \prod_{p=j, p \notin \{j+s\}}^{j+s} \left((p-1)(\Delta+1) - \Delta \sum_{l=1}^0 \mathbb{1}(i_l \leq p-1) \right) \right)
 \end{aligned}$$

which is merged with the third term on the right-hand side of (g) to obtain the $k=1$ term on the right-hand side of (h). Similarly, for the terms for $k=2$ to $k=s$, we merge both of the terms on the right-hand side of (f) using (18) to obtain the fourth term on the right-hand side of (g). The last term on the right-hand side of (g) is the evaluation of the fourth term on the right-hand side

of (f) at $k = s + 1$. Finally (h) is obtained by writing all the terms under one summation. Hence the proof follows from induction on $t - j + 1$.

4. Simulation results

In this section, we present a comparative study² between our model and the Barabási-Albert model in terms of following three features:

- Structural differences in small-sized graphs;
- Degree distributions of the graphs obtained;
- Expected birth time of vertices with a fixed degree.

We first illustrate the structural differences (in terms of vertex color allocation and degree) between the graphs generated by our model and the Barabási-Albert model. In the next set of simulations, we compare the degree distribution of both models by plotting the probability of randomly choosing a k degree vertex versus k (on a log – log scale) for a graph generated until a fixed time instant. We give the degree distribution of graphs generated for 5000 time steps (averaged over 250 simulations) via the standard Barabási-Albert model and our model with different choices of the reinforcement parameter Δ_t and discuss the similarities and differences obtained in the degree distributions. In the third set of simulations, we compare both models in terms of vertices expected birth time versus degree which we define as follows.

Definition 1. Given a random network/graph generated until time t , we define the vertices expected birth time for a fixed degree k , where $1 \leq k \leq t$, as the expected value of all the times when the vertices which have degree k at termination time t were introduced. It is denoted by $\bar{b}_t(k)$ and is given by the following expression:

$$\bar{b}_t(k) = \sum_{j=1}^t (j - 1) p_k^{(j)}(t) \tag{24}$$

where $p_k^{(j)}(t)$ is the probability that vertex j has degree k at termination time t .

Note that we write $(j - 1)$ in (24) because vertex j is introduced in the network/graph at time $(j - 1)$. In the experiments, we determine the empirical version of (24), which we call the average birth time and denote it by $b_t(k)$ for a degree k and termination time t . It is given by

$$b_t(k) = \sum_{j=1}^t (j - 1) \frac{\mathbb{1} \left(\sum_{n=j}^t Z_{j,n} = k - 1 \right)}{\sum_{j'=1}^t \mathbb{1} \left(\sum_{n=j'}^t Z_{j',n} = k - 1 \right)}. \tag{25}$$

Analyzing the expected birth time for a random graph provides insights on the average “age” of vertices gaining a certain degree in the generated network. A high expected birth time for a high-degree k indicates that newer incoming vertices have (on average) accumulated more connections than older vertices. In the context of social networks, this scenario corresponds to *tardy influential* vertices that quickly gained more popularity over already existing ones despite being introduced later in the network. We will compute the empirical expected birth time (average birth time) for different choices of Δ_t and compare them with the Barabási-Albert model later in this section.

In Fig. 4, two 15-vertex networks are depicted, one generated by our model (on the left-hand side) and the other by the Barabási-Albert model (on the right-hand side). We make two observations: First, in contrast with the Barabási-Albert model, in our model all vertices are labeled by distinct colors. This one-to-one correspondence between vertices and colors encodes all the

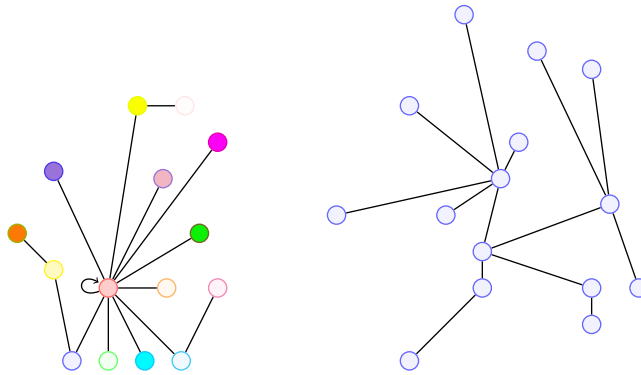


Figure 4. On the left-hand side is a 15-vertex network generated via the draws from a Pólya urn with expanding colors and $\Delta_t = 5$ for all $t \geq 1$ and on the right-hand side is a network with 15 vertices generated via Barabási-Albert model. For our model, unlike the Barabási-Albert model, each vertex is represented by a distinct color which corresponds to a color type of balls in the Pólya urn at that time instant. Furthermore, the extra reinforcement parameter Δ_t in our model provides versatility in the level of preferential attachment. The parameter $\Delta_t = 5$ in our model enables the central vertex of the graph on the left-hand side to obtain a higher degree (11 in this case) as compared to the right-hand side Barabási-Albert network in which the highest degree achieved is 6.

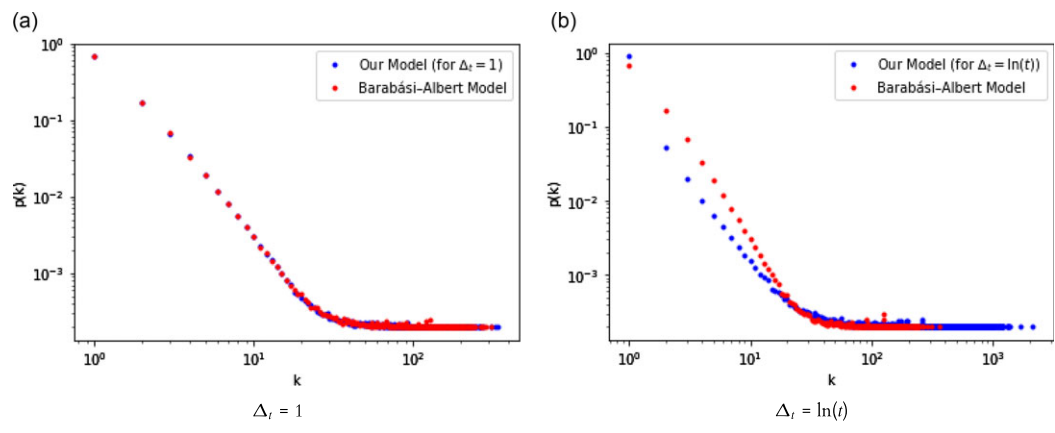


Figure 5. Degree distributions of networks generated until time 5000 (averaged over 250 simulations) for the Barabási-Albert model and our model with $\Delta_t = 1$ and $\Delta_t = \ln(t)$. In **(a)**, the degree distributions of both models are nearly identical. While in **(b)** the degree distributions are quite different.

information of the generated graph in the draw vectors of the underlying Pólya urn. Second, the maximum degree achieved is higher in the left-hand side network generated via our model (which achieves a maximum degree of 11 compared to 6 in the right-hand side Barabási-Albert network). This happens along one sample path due to the choice of reinforcement parameter (here $\Delta_t = 5$ for all $t \geq 1$) in our model which allows for an already selected vertex to be chosen with a higher probability than in the case of the Barabási-Albert model where the vertices are chosen proportional to their degree. In fact, this property is observed to persist in our simulations.

In Figs. 5 and 6, we plot the degree distribution of networks generated for four different choices of Δ_t : 1, $\ln(t)$, $f(t)$, $g(t)$, where the functions $f(t)$ and $g(t)$ are defined in (26). We observe the deviation of the degree distribution of the graphs generated via our model for the above-mentioned

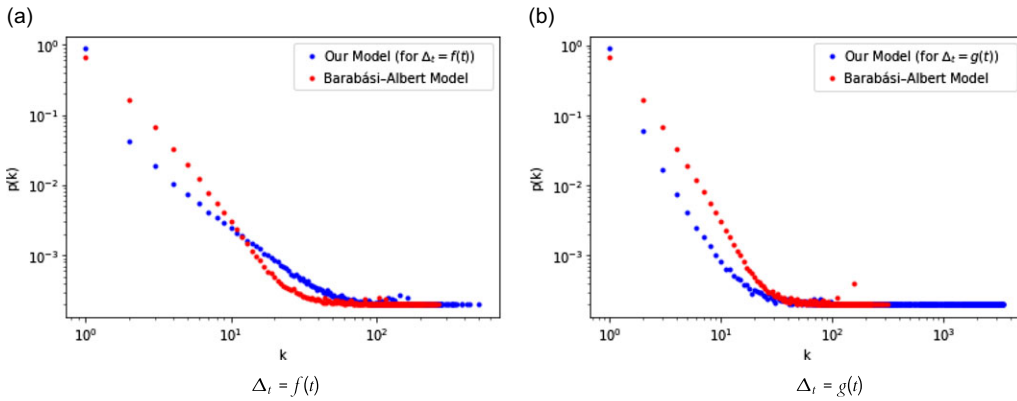


Figure 6. Degree distribution of the Barabási-Albert model and our model generated for two different choices of Δ_t , **(a)** $\Delta_t = f(t)$ and **(b)** $\Delta_t = g(t)$, where $f(t)$ and $g(t)$ are defined in (26). Both plots are averaged over 250 simulations, where each simulation is a generation of a 5000-vertex graph.

choices of Δ_t from the degree distribution of the Barabási-Albert network which follows the relation $p(k) \sim k^{-3}$, where $p(k)$ is the probability of randomly choosing a vertex of degree k in the network. The similarity between the Barabási-Albert algorithm and our model with $\Delta_t = 1$ (as observed in Fig. 5(a)) can be represented in the following way:

$$\begin{aligned}
 &P(\text{incoming vertex at time } t \text{ connects to vertex corresponding to color } c_j) \\
 &= \text{ratio of } c_j \text{ color balls in the expanding color Pólya urn at time } t - 1 \\
 &= \frac{\text{degree of vertex corresponding to color } c_j \text{ in graph } \mathcal{G}_{t-1}}{\text{sum of degrees in graph } \mathcal{G}_{t-1}} \\
 &= P(\text{incoming vertex connects to the vertex added at time } j - 1 \\
 &\quad \text{in a standard Barabási-Albert network}).
 \end{aligned}$$

Hence in the case where $\Delta_t = 1$, the mechanisms of both models for iteratively constructing new vertices and edges are equivalent. However, the initialization of our model is different from the Barabási-Albert model. In our model, the initial graph has only one vertex with a self-loop, whereas in the Barabási-Albert model, the initial graph can potentially have more than one vertex equipped with an edge set and no self-loops. Even though the initialization of both models is different, the equivalent procedures for adding new vertices and edges between our model with $\Delta_t = 1$ and the standard Barabási-Albert model ensure that the generated graphs via both models will show similar properties for sufficiently large t . While it is an intricate task to analytically solve for the asymptotic degree distribution and other properties of our model due to the fact that its reinforcement dynamics is much more involved than that of the Barabási-Albert model, such an investigation is a worthwhile future direction.

In Fig. 5(b), we observe that the degree distribution of our model with $\Delta_t = \ln(t)$ significantly differs from the degree distribution of Barabási-Albert model. The former has a lower probability of obtaining lower degree vertices (degree range $10^0 - 10^1$) as compared to the latter but has a slightly higher probability of gaining moderate degree vertices (degree range $50 - 150$). Additionally, the maximum degree attained in the case of $\Delta_t = \ln(t)$ for our model in Fig. 5 is much higher ($\sim 10^3$ as compared to only 200 in Barabási-Albert network).

In Fig. 6, we present two more cases in which the degree distributions differ substantially between our model and the Barabási-Albert model. More specifically, we generate our

network for Δ_t being an increasing step function $f(t)$ and a decreasing continuous function $g(t)$ given by:

$$f(t) = \begin{cases} 1 & \text{for } 0 \leq t < 1000 \\ 10 & \text{for } 1000 \leq t < 2500 \\ 100 & \text{for } 2500 \leq t \leq 5000, \end{cases} \quad g(t) = \begin{cases} 10 & \text{for } 0 \leq t \leq 1000 \\ \frac{10^4}{t} & \text{for } 1000 \leq t \leq 2000 \\ 5 & \text{for } 2000 \leq t \leq 3000 \\ \frac{15 \times 10^3}{t} & \text{for } 3000 \leq t \leq 4000 \\ 3.75 & \text{for } 4000 \leq t \leq 5000. \end{cases} \quad (26)$$

We remark from Fig. 6 that the maximum degree attained in both figures (generated via $\Delta_t = f(t)$ and $\Delta_t = g(t)$) is higher than in the Barabási-Albert model. Under the function $g(t)$, the constant value of $\Delta_t = 10$ up until time 1000 allows for the ball colors corresponding to older vertices (birth time ≤ 1000) to get accumulated in large numbers. Thereafter, the value of Δ_t continuously decreases, and therefore, the ball colors corresponding to younger vertices (birth time ≥ 1000) do not grow in large numbers. This gap between number of balls corresponding to younger and older vertices is significant enough for most of the older vertices to achieve high degrees with the younger vertices not acquiring much connections. Hence, most of the vertices either get a very high degree or a very low degree resulting in less vertices with moderate degrees ($10^1 - 10^2$) compared to the Barabási-Albert network. On the contrary, the increasing step function f allows for a wider range of ball colors (corresponding to vertices born between time 1000 to 2500) to grow their proportion, thus producing more vertices in the moderate degree range ($10^1 - 10^2$) compared to the Barabási-Albert network.

In the next set of simulations, we compare (25) for our network generated with $\Delta_t = 1, \ln(t), f(t)$ and $g(t)$ and the Barabási-Albert network. Fig. 7(a) demonstrates that in both the Barabási-Albert model and our model for $\Delta_t = 1$ vertices of the same degree are born at similar times. The stark similarities between the Barabási-Albert model and our model for $\Delta_t = 1$ in Figs. 5 and 7 strongly suggest that both networks have very similar structures; however, a rigorous analytic study is required to confirm if our model with $\Delta_t = 1$ is stochastically equivalent to the standard Barabási-Albert model. In Fig. 7(b), we observe that for our model with $\Delta_t = \ln(t)$, the network shows slightly more connectivity in the vertices which are born at similar times as compared to the Barabási-Albert network. This effect of same-age vertices showing more connectivity when compared to Barabási-Albert networks is much more amplified when our model uses $\Delta(t) = f(t)$ as shown in Fig. 7(c). Both cases provide a much richer algorithm for generating real-life networks in which the “rich gets richer” phenomenon needs to be dampened as it allows the more recently born vertices to get more connectivity. In contrast, Fig. 7(d) shows an amplification of the “rich gets richer” phenomenon when compared to the Barabási-Albert network as the first two richest vertices achieve a significantly higher degree (around 3500) compared to all other vertices. The rest of the vertices have very similar connectivity as that of the Barabási-Albert network. The choice $\Delta_t = g(t)$ of the reinforcement parameter in our model provides an algorithm to generate graphs which are spatially similar to the Barabási-Albert network but demonstrate a higher effect of preferential attachment.

5. Conclusion

We constructed a preferential attachment-type graph using the draw vectors of a Pólya urn with growing colors and a tunable time-varying reinforcement parameter Δ_t . The network obtained is essentially equivalent to the Barabási-Albert network for the case $\Delta_t = 1$ and gains a significant

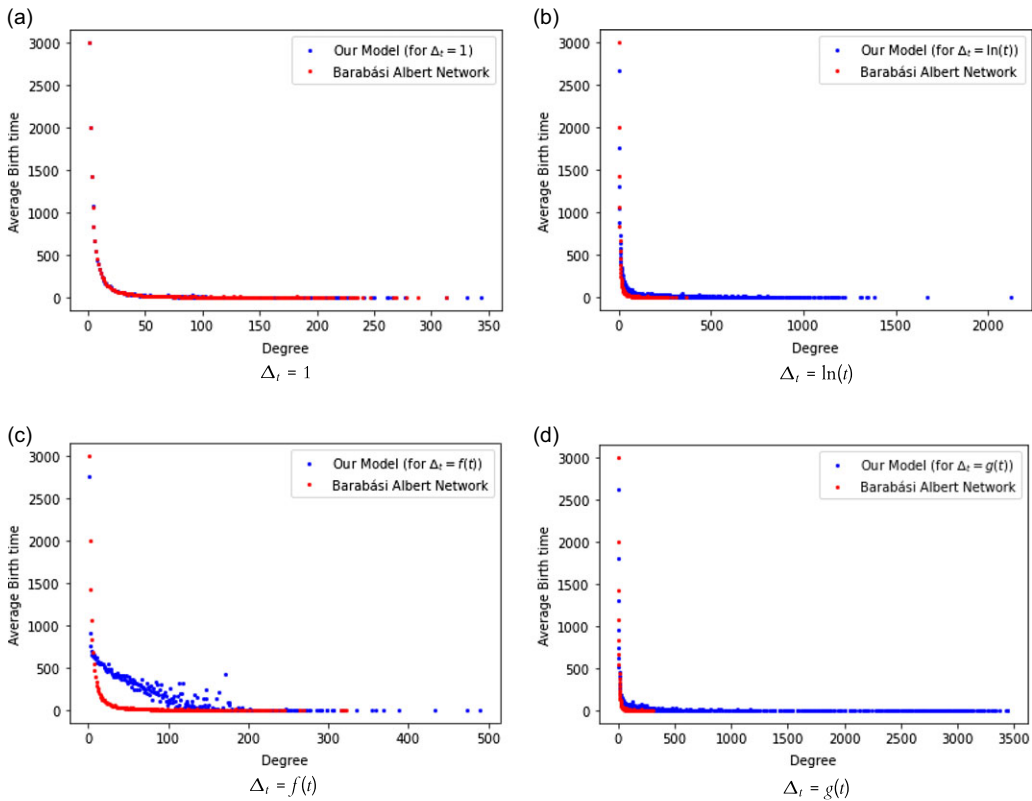


Figure 7. Vertices average birth time versus degree for our model using **(a)** $\Delta_t = 1$; **(b)** $\Delta_t = \ln(t)$; **(c)** $\Delta_t = f(t)$ and **(d)** $\Delta_t = g(t)$ (where the functions $f(t)$ and $g(t)$ are given in (26)) and for the Barabási-Albert network. All networks are generated for 5000 time steps and the average of 250 such networks is plotted.

amount of versatility when Δ_t is a time-varying function. We analyzed the draw vectors of the underlying stochastic process and derived the probability distribution of a random variable counting the draws of a particular color of this Pólya process. This random variable can be written in terms of the degree of the vertex in the constructed preferential attachment network corresponding to this color. We provided simulation evidences for the structural similarities between our model and the Barabási-Albert model for $\Delta_t = 1$ and also justified the richness and versatility of our model for general Δ_t . Future directions include devising a preferential attachment graph-generating algorithm using a Pólya urn with finitely many colors, formulating strategies for choosing the best possible Δ_t for a randomly growing graph, incorporating removal of edges in the graph through removal of balls from the Pólya urn and setting an upper limit on the maximum degree a vertex can achieve.

Acknowledgements. This work was funded in part by the Natural Sciences and Engineering Research Council of Canada.

Competing interests. None.

Notes

- 1 All identities involving random variables or vectors are (implicitly) understood to hold almost surely.
- 2 For details about the simulations, refer to the following link: https://drive.google.com/drive/folders/1uOmz4B6RQ0hRmuu-CTfb02jJ03B1SyEl?usp=share_link

References

- Alajaji F and Fuja T (1994) A communication channel modeled on contagion. *IEEE Transactions on Information Theory* 40(6), 2035–2041.
- Albert R and Barabási AL (2000) Topology of evolving networks: local events and universality. *Physical Review Letters* 85(24), 5234–5237.
- Albert R and Barabási AL (2002) Statistical mechanics of complex networks. *Reviews of Modern Physics* 74(1), 47–97.
- Alves TFA, Alves GA, Macedo-Filho A, Ferreira RS and Lima FWS (2021) The diffusive epidemic process on Barabási–Albert networks. *Journal of Statistical Mechanics: Theory and Experiment* 2021(4), 043203.
- Avin C, Daltrophe H, Keller B, Lotker Z, Mathieu C, Peleg D and Pignolet YA (2020) Mixed preferential attachment model: homophily and minorities in social networks. *Physica A: Statistical Mechanics and its Applications* 555, 124723.
- Avin C, Lotker Z, Nahum Y and Peleg D (2019) Random preferential attachment hypergraph. In *Proceedings IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 398–405.
- Banerjee A, Burlina P and Alajaji F (1999) Image segmentation and labeling using the Pólya urn model. *IEEE Transactions on Image Processing* 8(9), 1243–1253.
- Barabási AL (2009) Scale-free networks: a decade and beyond. *Science* 325(5939), 412–413.
- Barabási AL and Albert R (1999) Emergence of scaling in random networks science. *Science* 286(5439), 509–512.
- Berger N, Borgs C, Chayes J and Saberi A (2005) On the spread of viruses on the internet. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms*, 301–310.
- Berger N, Borgs C, Chayes JT, D’Souza RM and Kleinberg RD (2005) Degree distribution of competition-induced preferential attachment graphs. *Combinatorics, Probability & Computing* 14(6), 697–721.
- Berger N, Borgs C, Chayes JT and Saberi A (2014) Asymptotic behavior and distributional limits of preferential attachment graphs. *Annals of Probability* 42(1), 1–40.
- Bhutani KR, Kalpathy R and Mahmoud H (2022) Random networks grown by fusing edges via urns. *Network Science* 10(4), 347–360.
- Bollobás B and Riordan O (2004) The diameter of a scale-free random graph. *Combinatorica* 24(1), 5–34.
- Capocci A, Servedio VDP, Colaiori F, Buriol LS, Donato D, Leonardi S and Caldarelli G (2006) Preferential attachment in the growth of social networks: the internet encyclopedia wikipedia. *Physical Review E* 74(3), 036116.
- Chung F, Handjani S and Jungreis D (2003) Generalization of Pólya’s urn problem. *Annals of Combinatorics* 7(2), 141–153.
- Collecchio A, Cotar C and LiCalzi M (2013) On a preferential attachment and generalized pólya’s urn model. *Annals of Applied Probability* 23(3), 1219–1253.
- Csányi G and Szendrői B (2004) Structure of a large social network. *Physical Review E* 69(3), 036131.
- Flaxman AD, Frieze AM and Vera J (2006) A geometric preferential attachment model of networks. *Internet Mathematics* 3(2), 187–205.
- Giroire F, Nisse N, Trolliet T and Sulkowska M (2022) Preferential attachment hypergraph with high modularity. *Network Science* 10(4), 400–429.
- Hartmann AK and Weigt M (2001) Statistical mechanics perspective on the phase transition in vertex covering of finite-connectivity random graphs. *Theoretical Computer Science* 265(1), 199–225.
- Hayhoe M, Alajaji F and Gharesifard B (2018) A Pólya contagion model for networks. *IEEE Transactions on Control of Network Systems* 5(4), 1998–2010.
- He X, Lu J, Du H and Jin X (2023) Social influence in signed networks. *IEEE Transactions on Computational Social Systems*, 11(1), 1–10.
- Inoue M, Pham T and Shimodaira H (2022) A hypergraph approach for estimating growth mechanisms of complex networks. *IEEE Access* 10, 35012–35025.
- Jadbabaie A, Makur A, Mossel E and Salhab R (2023) Inference in opinion dynamics under social pressure. *IEEE Transactions on Automatic Control* 68(6), 3377–3392.
- Jeong H, Néda Z and Barabási AL (2003) Measuring preferential attachment in evolving networks. *Europhysics Letters* 61(1), 56–62.
- Kim JY and Jo HH (2010) A signalling explanation for preferential attachment in the evolution of social networks. *Journal of Evolutionary Economics* 20(3), 375–393.
- König W (2005) Orthogonal polynomial ensembles in probability theory. *Probability Surveys* 2, 385–447.
- Krapivsky PL, Redner S and Leyvraz F (2000) Connectivity of growing random networks. *Physical Review Letters* 21, 4629–4632.
- Lin M, Wang G and Chen T (2006) A modified earthquake model based on generalized Barabási–Albert scale-free networks. *Communications in Theoretical Physics* 46(6), 1011–1016.
- Mahmoud H (2008) *Pólya Urn Models*. CRC Press.
- Marcaccioli R and Livan G (2019) A Pólya urn approach to information filtering in complex networks. *Nature Communications* 10(1), 745–745.
- Milojević S (2010) Modes of collaboration in modern science: beyond power laws and preferential attachment. *Journal of the American Society for Information Science and Technology* 61(7), 1410–1423.

- Newman MEJ (2001) Clustering and preferential attachment in growing networks. *Physical Review E* 64(4), 025102.
- Oliveira RI (2009) The onset of dominance in balls-in-bins processes with feedback. *Random Structures & Algorithms* 34(4), 454–477.
- Pemantle R (2007) A survey of random processes with reinforcement. *Probability Surveys* 4, 1–79.
- Pólya G (1930) Sur quelques points de la théorie des probabilités. *Annales de l'institut Henri Poincaré* 1(2), 117–161.
- Rácz MZ and Sridhar A (2022) Correlated randomly growing graphs. *The Annals of Applied Probability* 32(2), 1058–1111.
- Santos FC and Pacheco JM (2005) Scale-free networks provide a unifying framework for the emergence of cooperation. *Physical Review Letters* 95(9), 098104.
- Singh S, Alajaji F and Ghahesifard B (2022a) A finite memory interacting Pólya contagion network and its approximating dynamical systems. *SIAM Journal on Control and Optimization* 60(2), 347–369.
- Singh S, Alajaji F and Ghahesifard B (2022b) Consensus using a network of finite memory Pólya urns. *IEEE Control Systems Letters* 6, 2780–2785.
- Singh S, Alajaji F and Ghahesifard B (2022c) Modeling network contagion via interacting finite memory Pólya urns. In *Proceedings IEEE International Symposium on Information Theory (ISIT)*, 348–353.
- Toivonen R, Onnela JP, Saramäki J, Hyvönen J and Kaski K (2006) A model for social networks. *Physica A: Statistical Mechanics and its Applications* 371(2), 851–860.
- Tsallis C and Oliveira R (2022) Complex network growth model: possible isomorphism between nonextensive statistical mechanics and random geometry. *Chaos* 32(5), 053126–053126.
- Zhongdong X, Guanghui Z and Wang B (2012) Using modified Barabási and Albert model to study the complex logistic network in eco-industrial systems. *International Journal of Production Economics* 140(1), 295–304.