J. Appl. Prob. **53,** 434–447 (2016)
*doi:*10.1017/jpr.2016.11
© *Applied Probability Trust* 2016

DEGREES IN RANDOM SELF-SIMILAR BIPOLAR NETWORKS

CHEN CHEN * AND
HOSAM MAHMOUD,* ** *The George Washington University*

Abstract

We investigate several aspects of a self-similar evolutionary process that builds a random bipolar network from building blocks that are themselves small bipolar networks. We characterize admissible outdegrees in the history of the evolution. We obtain the limit distribution of the polar degrees (when suitably scaled) characterized by its sequence of moments. We also obtain the asymptotic joint multivariate normal distribution of the number of nodes of small admissible outdegrees. Five possible substructures arise, and each has its own parameters (mean vector and covariance matrix) in the multivariate distribution. Several results are obtained by mapping bipolar networks into Pólya urns.

Keywords: Random structure; network; self-similarity; random graph; degree; stochastic recurrence; Pólya urn; multivariate normal distribution

2010 Mathematics Subject Classification: Primary 05082; 90B15
Secondary 60C05; 60F05

1. Introduction

The classical assumption about the evolution of networks (particularly trees) is that edges (or nodes) are added gradually, one at a time, to an existing random structure. However, in many real applications the additions come in self-similar 'clusters' such as small building blocks, which are themselves graphs or small networks. The graph rewriting concepts of formal languages have numerous applications in biology, software construction and verification, and picture generation. For instance, when multicellular organisms reproduce, they undergo complex development controlled by the genetic information in their cells, replacing existing components with self-similar structures. Such a complex reproductive process is not completely understood yet, however it is believed to follow a strict graph rewriting system of rules that regulate the replacements [9].

This study trend is picking up in the very recent research. For instance, trees with 'thickened nodes' were considered in [3], where a random tree is endowed with the property that each of its nodes bursts into a thicker cluster (graph with certain required connectivity properties), whereas [5] considered random trees built from building blocks, which are themselves trees. Gopaladesikan *et al.* [5] cites numerous potential applications. These applications include the spreading of religion as described in [12], which purports that tree growth from blocks of trees can be used to study the growth of a religion through conversion. An entire family (tree) converts, when the head of the family decides to follow a new religion. Subsequently, all the members of the family try to convert other families (trees).

Received 30 September 2014; revision received 13 May 2015.
* Postal address: Department of Statistics, The George Washington University, 801 22nd St. NW, Washington, DC 20052, USA.
** Email address: hosam@gwu.edu

https://doi.org/10.1017/jpr.2016.11 Published online by Cambridge University Press

In this paper we study the growth of large bipolar networks (graphs) from smaller building graphs that are themselves bipolar graphs. We look into the admissible outdegrees in such graphs and study the polar degrees and the asymptotic joint distribution of the number of nodes of small admissible outdegrees.

These themes are organized in sections. Section 2 is an overview of the structure and applications of bipolar networks. In Section 3 we characterize the admissible outdegrees that can appear in the network. In Section 4 we say a few words, by way of introducing notation, on Pólya urn models, a tool in this study. In Section 5 we develop the asymptotic distribution of the polar degrees. In Section 6 we discuss the five structural cases that can arise (Section 6.1) and derive the multivariate central limit theorem for the joint distribution (after suitable normalization) of the number of nodes of a small admissible outdegree (Section 6.2). We present the limit bivariate distribution of the number of nodes of the smallest and second smallest admissible outdegrees, which in principle can be extended to an arbitrary fixed number of admissible outdegrees. In Section 7 we discuss the limiting covariance computation. The section is divided into two units: in Section 7.1 we discuss the computations for a large class of bipolar networks, and in Section 7.2 we illustrate how in principle the computations can be done exactly and asymptotically, even if the case does not follow the conditions needed in Section 7.1. In Section 8 we offer a few concluding remarks.

2. Bipolar networks

Let $G = (V, E)$ be a given directed graph with vertex (node) set V and a set of directed edges E and without cycles. We call such a graph *bipolar* when it has a node designated as the North Pole (N), which is the only node with no edges ending on it, and another designated as the South Pole (S), and there is a directed path connecting any node $v \in V - \{\mathrm{S}\}$ to S, which is the only node with no edges out of it. The number of edges emanating out of a node $v \in V$ is the *outdegree* of v, to be denoted outdeg(v). Let Δ be the outdegree of N, and assume that there are $k \geq 1$ distinct outdegrees in the graph, which are namely (in increasing order) $d_1 < d_2 < \cdots < d_k$. Necessarily, we must have $d_1 \leq \Delta \in \{d_1, d_2, \ldots, d_k\}$.

A *self-similar bipolar network* (henceforth, we shall refer to it simply as a bipolar network) grows in the following manner. At the beginning of discrete time $n = 0$, we have $G_0 = G$, a bipolar *seed* graph (henceforth simply called the seed). Thereupon, a series of graphs $G_0, G_1, G_2, \ldots$ evolves, where for $n \geq 1$, G_n is obtained from G_{n-1} by choosing an edge of G_{n-1} at random (all edges being equally likely), and replacing it with a copy of G as follows. If the edge chosen for replacement is xy, directed from vertex x to vertex y in G_{n-1}, we remove the edge (but not its end vertices), obtain an additional copy of G, and identify its North Pole with x and its South Pole with y. In Figure 1 we illustrate the growth of a bipolar network in two time steps; the seed is the leftmost graph. The edge chosen for replacement at a step is flagged with an asterisk.

Among many other potential applications, random bipolar networks can be models for road maps between trading centers, with N being a source of shipping merchandise, and S being a market. The routes connecting the poles evolve into towns and other road configurations. Etymologically, boroughs (and many similar derivatives like boro, borough, berg, borg, burg, and bourg) are configurations that historically appeared as sporadic spots on trade routes providing services to travelers such has lodging, and they themselves grew in time to become cities in their own right. Salzburg in Austria is an example of a town that evolved on the trading routes of the salt industry. The spreading of towns and boroughs on the map is a manifestation of bipolar networks.

https://doi.org/10.1017/jpr.2016.11 Published online by Cambridge University Press

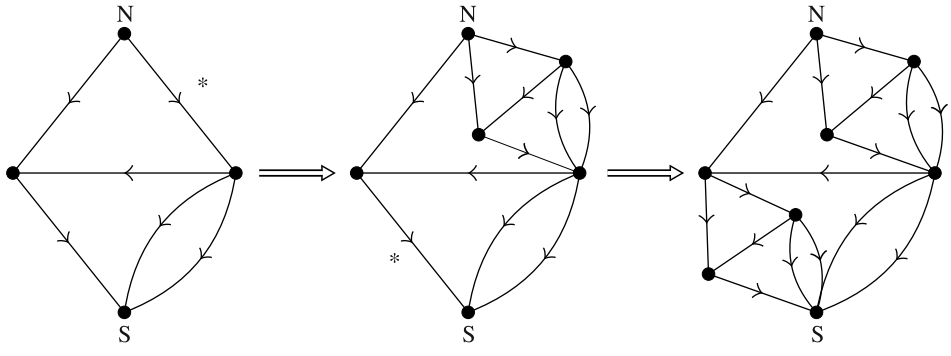


FIGURE 1: The evolution of a bipolar network in two steps. At each step, the starred edge is the one chosen for replacement.

3. Admissible outdegrees

As the bipolar network grows, copies of the seed G are added, and there can be many nodes of outdegree d_1, for example. Other outdegrees not in the set $\{d_1, d_2, \ldots, d_k\}$ can also appear later. Let $d_1^* < d_2^* < d_3^* \ldots$ be the *admissible outdegrees* throughout the history of evolution, i.e. the outdegrees that can appear in any of $G_0, G_1, \ldots$; note that $d_1^* = d_1$, but d_2^* is not necessarily equal to d_2, and the seed may not contain nodes of outdegree d_2^*. In the next theorem we describe the outdegrees that can appear in a bipolar network.

Theorem 1. *Suppose that a bipolar network grows out of a bipolar seed with outdegrees given by* $\{d_1, d_2, \ldots, d_k\}$. *Then,* d^* *is an admissible outdegree if and only if it has the form* $d^* = d_i + (\Delta - 1)r$ *for* $i = 1, \ldots, k$ *and some integer* $r \geq 0$.

Proof. For $i = 1, \ldots, k$, each time we pick an edge coming out of a node of outdegree d_i, we replace it with a copy of G, increasing that outdegree by $\Delta - 1$ (the -1 accounts for the removal of the edge upon replacement). If we start at a vertex with outdegree d_i, for $i = 1, \ldots, k$, and we keep choosing one of the edges emanating out of it (and their number keeps changing) $r \geq 0$ times, we obtain a node of outdegree $d_i + (\Delta - 1)r$. Hence, the outdegree $d_i + (\Delta - 1)r$ can appear in the network (i.e. is admissible) for $i = 1, \ldots, k$, and any $r \geq 0$.

We next argue that the given form is the only possible form of outdegrees in the growing network. We prove this claim by induction. Assume (as an induction hypothesis) that the outdegree profile after n replacements is $d_i + (\Delta - 1)r$ for $i = 1, \ldots, k$ and nonnegative integers $0 \leq r \leq n$. At step $n + 1$, we choose an edge from the existing network (and all its outdegrees have the stated form by hypothesis) and replace it with a copy of G. The edge is coming out of a node with outdegree $d_i + (\Delta - 1)r$ for some $i = 1, \ldots, k$ and some $r \in \{0, \ldots, n\}$, which gets altered to $d_i + (\Delta - 1)r + \Delta - 1 = d_i + (\Delta - 1)(r + 1)$, and no other outdegree is changed. The added vertices of G are also of the correct form. So, after $n+1$ steps, the outdegrees in the network are $d_i + (\Delta - 1)r$ for $0 \leq r \leq n + 1$ and all $i = 1, \ldots, k$. The induction is complete. □

Corollary 1. *If* $d_1 = \Delta = 1$ *and* $k = 1$, *the only admissible outdegree is 1.*

Corollary 1 indicates the possibility of having a finite sequence of admissible outdegrees. Indeed, this case is very trivial—after stage n, the bipolar network in this case is a path of length

https://doi.org/10.1017/jpr.2016.11 Published online by Cambridge University Press

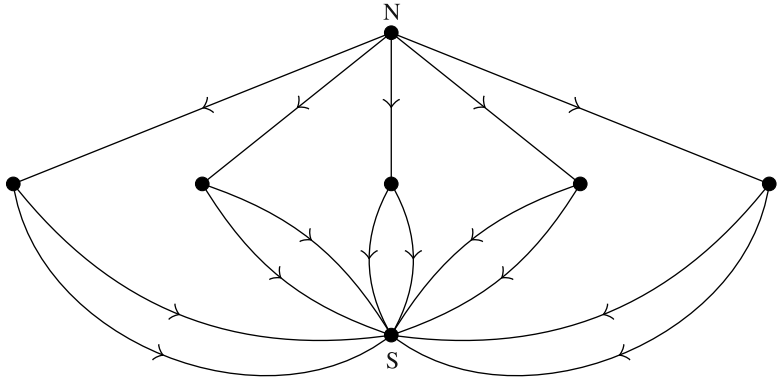


FIGURE 2: A seed with outdegrees 2 and 5.

$(|E| - 1)n + |E|$ edges and the outdegree of all the nodes is 1. There is no randomness to reckon with in the outdegrees of such a graph. The case will be excluded from the forthcoming study.

Corollary 2. *If $d_1 = 1$ and $\Delta = 2$, all natural numbers are admissible outdegrees.*

Corollary 2 indicates that when the seed graph G has nodes of outdegree 1, the network can be somewhat degenerate. We shall observe that indeed, the analysis of such a case is simpler than the other cases. For instance, in the networks of Figure 1, the seed has nodes in it of outdegree 1. As we continue inserting copies of the seed, nodes of outdegree j for any $j \in \mathbb{N}$ can appear. This is to be opposed to a network starting from a graph with outdegrees 2, 5, such as the one shown in Figure 2. In the latter network, the only admissible outdegrees are from the sequence

$$2, 5, 6, 9, 10, 13, 14, 17, \ldots.$$

4. Pólya urns

Pólya urn schemes will be a systematic tool in this study. Here we say a brief introductory word about these schemes and establish some notation. A k-color Pólya *urn scheme* is an urn containing balls of up to k different colors, subject to rules of evolution in discrete time. After each unit of time, a ball is drawn uniformly at random (all balls being equally likely) from the urn, and the color of the ball is observed. The ball is then placed back in the urn, and depending on its color, a certain number of balls of each color are added. If the ball drawn is of color i, we add $A_{i,j}$ balls of color j for $j = 1, \ldots, k$. (Interpret 'add' to mean take out balls if $A_{i,j}$ is negative.) The dynamics of the evolution of the urn can thus be represented by the *replacement matrix*

$$\boldsymbol{A} = \begin{pmatrix} A_{1,1} & A_{1,2} & \ldots & A_{1,k} \\ A_{2,1} & A_{2,2} & \ldots & A_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ A_{k,1} & A_{k,2} & \ldots & A_{k,k} \end{pmatrix}.$$

Entry (i, j) in this matrix is the number of balls of color j that we add upon drawing a ball of color i.

https://doi.org/10.1017/jpr.2016.11 Published online by Cambridge University Press

5. The outdegree of the North Pole

As a model for trading routes, the outdegree of the North Pole is of particular significance. It is a constraint on the maximal capacity of the network, i.e. the maximal amount of goods that can be shipped out of the source.

Suppose we colored the edges coming out of the North Pole with color 1 (say white (W)), and all the other edges with color 2 (say blue (B)). We think of the edges in the bipolar network as balls in a Pólya urn. Let W_n be the number of white edges in the graph after n steps of evolution (number of white balls in the urn after n draws). As we start from $G_0 = G$, we have $\mathrm{W}_0 = \Delta$ and $\mathrm{B}_0 = |E| - \Delta$, where $|E|$ is the cardinality of E. At each stage, we pick an edge at random (a ball from the urn at random). If the edge is white, we are replacing an edge connected to the North Pole with a copy of G. Thus, $\Delta - 1$ white edges and $|E| - \Delta$ blue edges are added to the graph ($\Delta - 1$ white balls and $|E| - \Delta$ blue balls are added to the urn). If, instead, we sample a blue edge (ball), we add $|E| - 1$ blue edges to the graph (blue balls to the urn) because this new added copy of G does not affect the outdegree of the North Pole. The replacement matrix associated with this urn is

$$\begin{pmatrix} \Delta - 1 & |E| - \Delta \\ 0 & |E| - 1 \end{pmatrix}.$$

We appeal to the recent results in [8] and [14] to quickly establish the properties of the outdegree of the North Pole. These results are in terms of Pochhammer's symbols for the rising factorials and Stirling's numbers of the second kind. Pochhammer's symbol for the rising factorial is

$$\langle x \rangle_s = x(x+1)\cdots(x+s-1) \quad \text{for any } x \in \mathbb{R}$$

and any integer $s \geq 0$, with the interpretation that $\langle x \rangle_0 = 1$. The numbers $\left\{ {r \atop j} \right\}$ are Stirling's numbers of the second kind. These are the number of ways to partition a set of r distinct objects into j nonempty parts. For properties, we refer the reader to textbooks such as [2] and [6].

Proposition 1. *After n steps of evolution, let* W_n *be the outdegree of the North Pole of a random bipolar network grown from a bipolar seed with* $|E|$ *edges and North Pole outdegree* Δ*. The moments of this random variable are*

$$\mathbb{E}[\mathrm{W}_n^r] = \frac{(\Delta-1)^r}{\langle |E|/(|E|-1) \rangle_n} \sum_{j=1}^{r} (-1)^{r-j} \left\{ {r \atop j} \right\} \left\langle \frac{\Delta}{\Delta - 1} \right\rangle_j \left\langle \frac{|E| + j(\Delta-1)}{|E|-1} \right\rangle_n.$$

With the proper normalization, a limit distribution follows from results in [4], [8], and [14], represented in terms of the standard Riemann gamma function.

Proposition 2. *Let* W_n *be the outdegree of the North Pole after n steps of evolution in a random bipolar network grown from a bipolar seed with* $|E|$ *edges and North Pole outdegree* Δ*. As* $n \to \infty$*, the random variable* $\mathrm{W}_n / n^{(\Delta-1)/(|E|-1)}$ *converges in distribution to a random variable* $\mathrm{W}_\star$*, which is uniquely characterized by its moments; these moments are*

$$\mathbb{E}[\mathrm{W}_\star^r] = (\Delta-1)^r \Gamma\left(\frac{|E|}{|E|-1}\right) \Gamma\left(r + \frac{\Delta}{\Delta-1}\right) \left(\Gamma\left(\frac{\Delta}{\Delta-1}\right) \Gamma\left(\frac{|E| + r(\Delta-1)}{|E|-1}\right)\right)^{-1}.$$

https://doi.org/10.1017/jpr.2016.11 Published online by Cambridge University Press

Corollary 3. *After n steps of evolution, the outdegree of the North Pole has mean*

$$\begin{aligned}\mathbb{E}[\mathrm{W}_n] &= \frac{\Delta}{\langle |E|/(|E|-1)\rangle_n}\left\langle \frac{|E|+\Delta-1}{|E|-1}\right\rangle_n \\ &\sim (\Delta-1)\left\{\Gamma\left(\frac{|E|}{|E|-1}\right)\Gamma\left(\frac{2\Delta-1}{\Delta-1}\right)\right. \\ &\qquad \left.\times\left(\Gamma\left(\frac{\Delta}{\Delta-1}\right)\Gamma\left(\frac{|E|+\Delta-1}{|E|-1}\right)\right)^{-1}\right\} n^{(\Delta-1)/(|E|-1)},\end{aligned}$$

and variance

$$\begin{aligned}\mathrm{var}[\mathrm{W}_n] &= \frac{(\Delta-1)^2}{\langle |E|/(|E|-1)\rangle_n}\left(\left\langle \frac{\Delta}{\Delta-1}\right\rangle_2\left\langle \frac{|E|+2(\Delta-1)}{|E|-1}\right\rangle_n\right. \\ &\quad \left.-\left(\frac{\Delta}{\Delta-1}+\frac{\Delta^2}{(\Delta-1)^2\langle (|E|+\Delta-1)/(|E|-1)\rangle_n}\right.\right. \\ &\qquad \left.\left.\times\left\langle \frac{|E|+\Delta-1}{|E|-1}\right\rangle_n\right)\left\langle \frac{|E|+\Delta-1}{|E|-1}\right\rangle_n\right) \\ &\sim (\Delta-1)^2\left(\Gamma\left(\frac{|E|}{|E|-1}\right)\Gamma\left(2+\frac{\Delta}{\Delta-1}\right)\left\{\Gamma\left(\frac{\Delta}{\Delta-1}\right)\Gamma\left(\frac{|E|+2(\Delta-1)}{|E|-1}\right)\right\}^{-1}\right. \\ &\qquad -\left(\Gamma\left(\frac{|E|}{|E|-1}\right)\Gamma\left(1+\frac{\Delta}{\Delta-1}\right)\right. \\ &\qquad \left.\left.\left.\times\left\{\Gamma\left(\frac{\Delta}{\Delta-1}\right)\Gamma\left(\frac{|E|+\Delta-1)}{|E|-1}\right)\right\}^{-1}\right)^2\right) n^{2(\Delta-1)/(|E|-1)}.\end{aligned}$$

6. Counting nodes of small outdegrees in the network

For $i \geq 1$, let $D_n^{(i)}$ be the number of nodes of the admissible outdegree d_i^* in the bipolar network, after n steps of evolution. Here we show how to obtain the asymptotic joint distribution of $D_n^{(1)}$ and $D_n^{(2)}$. The method used is applicable in principle to obtaining the asymptotic joint distribution of $D_n^{(1)}, D_n^{(2)}, \ldots, D_n^{(m)}$ for fixed m.

6.1. A color code

Toward asymptotic distributions, we use a color code. Let us color the edges coming out of nodes of outdegree d_1^* with color 1, edges coming out of nodes of outdegree d_2^* with color 2, and all other edges (coming out of nodes of higher outdegrees) with color 3. Think of the colored edges as balls in an urn. Depending on the interplay between the outdegrees Δ, d_1^*, and d_2^*, there are five Pólya urn schemes that can arise. We summarize these cases in the next lemma. The arguments in the proof give the rationale for each case.

The number of edges out of nodes of outdegree d_i^* in the seed will be called n_i for $i = 1, 2$, and n_3 is the number of edges out of nodes of outdegrees higher than d_2^*. For instance, the seed graph in Figure 2 has the two outdegrees $d_1^* = d_1 = 2$ and $d_2^* = d_2 = 5$, and their frequency of appearance is $n_1 = 10$ and $n_2 = 5$, respectively. As this seed has no edges out of nodes of outdegrees higher than 5, we take $n_3 = 0$.

Lemma 1. *Let G_n be a bipolar network grown from a given bipolar seed G, which is not a directed path and in which there are n_i edges emanating out of nodes of outdegree d_i^* for $i = 1, 2$, and n_3 edges coming out of nodes of higher outdegrees. Color the edges as discussed*

https://doi.org/10.1017/jpr.2016.11 Published online by Cambridge University Press

just before this lemma. Consider the evolution of the bipolar network as a Pólya urn scheme. We note that the case $d_1 = \Delta$ and $d_2^* > 2\Delta - 1$ cannot arise. The corresponding ball replacement matrix falls into one of the following cases:

(i) *if $d_1 = \Delta$ and $d_2^* < 2\Delta - 1$, the replacement matrix is*

$$\begin{pmatrix} n_1 - 2d_1 & n_2 & n_3 + 2d_1 - 1 \\ n_1 - d_1 & n_2 - d_2^* & n_3 + d_1 + d_2^* - 1 \\ n_1 - d_1 & n_2 & n_3 + d_1 - 1 \end{pmatrix};$$

(ii) *if $d_1 = \Delta$ and $d_2^* = 2\Delta - 1$, the replacement matrix is*

$$\begin{pmatrix} n_1 - 2d_1 & n_2 + 2d_1 - 1 & n_3 \\ n_1 - d_1 & n_2 - d_2^* & n_3 + d_1 + d_2^* - 1 \\ n_1 - d_1 & n_2 & n_3 + d_1 - 1 \end{pmatrix};$$

(iii) *if $1 < d_1 < d_2^* = \Delta$, the replacement matrix is*

$$\begin{pmatrix} n_1 - d_1 & n_2 - d_2^* & n_3 + d_1 + d_2^* - 1 \\ n_1 & n_2 - 2d_2^* & n_3 + 2d_2^* - 1 \\ n_1 & n_2 - d_2^* & n_3 + d_2^* - 1 \end{pmatrix};$$

(iv) *if $1 = d_1 < d_2^* = \Delta$, the replacement matrix is*

$$\begin{pmatrix} n_1 - 1 & n_2 & n_3 \\ n_1 & n_2 - 2d_2^* & n_3 + 2d_2^* - 1 \\ n_1 & n_2 - d_2^* & n_3 + d_2^* - 1 \end{pmatrix};$$

(v) *if $d_1 < d_2^* < \Delta$, the replacement matrix is*

$$\begin{pmatrix} n_1 - d_1 & n_2 & n_3 + d_1 - 1 \\ n_1 & n_2 - d_2^* & n_3 + d_2^* - 1 \\ n_1 & n_2 & n_3 - 1 \end{pmatrix}.$$

Proof. Suppose that the edge picked from the bipolar network $\mathcal{G}$ is xy, directed from x to y, and this edge is replaced by a copy of G, resulting in a new network $\mathcal{G}'$. We argue case (i) in detail and only sketch the remaining cases, noting differences.

(i) When we pick edge xy of color 1 in $\mathcal{G}$ (we draw a ball of color 1 from the urn), we replace the edge with a copy of G. That copy comes with d_1 edges emanating out of its North Pole. We identify that North Pole with x and the South Pole with y. Prior to the insertion, the outdegree of x in $\mathcal{G}$ is $\text{outdeg}(x) = d_1 = \Delta$. After insertion it is increased to $\text{outdeg}(x) + \Delta - 1 = 2d_1 - 1$ in $\mathcal{G}'$, and that is an outdegree higher than d_2^* by the condition of the case. All the edges out of x should be of color 3; we add $n_3 + 2d_1 - 1$ edges of color 3 to $\mathcal{G}$ to obtain $\mathcal{G}'$ ($n_3 + 2d_1 - 1$ balls of color 3 to the urn). The added copy of G has n_1 edges out of nodes of outdegree d_1; these include d_1 edges at the North Pole of G. When G is hooked to $\mathcal{G}$ to construct $\mathcal{G}'$, those d_1 edges receive color 3 as mentioned: the incoming graph adds only $n_1 - d_1$ edges of color 1, and $\mathcal{G}$ loses an additional d_1 edges of color 1 (which are recolored with color 3). While executing these dynamics, the inserted copy of G adds n_2 edges of color 2.

https://doi.org/10.1017/jpr.2016.11 Published online by Cambridge University Press

If instead, the edge xy we picked in $\mathcal{G}$ is of color 2, the added graph contributes $n_1 - d_1$ new edges of color 1; and only $n_2 - d_2^*$ edges of color 2 (with G adding n_2 and $\mathcal{G}$ losing d_2^* at x). At x the outdegree is modified from d_2^* in $\mathcal{G}$ to $d_2^* - 1 + \Delta > d_2^*$ (the case imposes the condition $\Delta \geq 3$), and thus the edges out of x in the new network receive color 3, so that new network has $n_3 + d_2^* - 1 + \Delta$ additional edges of color 3.
The remaining alternative is xy is of color 3. In this case, we add $n_1 - d_1$ edges of color 1, G adds n_2 edges of color 2. The outdegree of x in $\mathcal{G}$ is larger than d_2^*. These edges are of color 3 in $\mathcal{G}$, and remain of color 3 in the new network as the degree of x is increased. By adding G, the new network receives n_3 edges of color 3, plus the edges at the North Pole of G, while losing the edge of color 3 that got replaced by G.

(ii) The argument is similar to that in case (i). The difference is only in the top row. Upon choosing an edge of color 1, we modify the outdegree of x: Incoming $d_1 - 1$ edges at the North Pole of the attached copy of G receive color 2, and G adds only $n_1 - d_1$ edges of color 1.

(iii) The argument is similar to that in cases (i) and (ii).

(iv) The argument is similar to that in cases (i) and (ii), and the condition introduces degeneracy.

(v) The argument is similar to that in cases (i) and (ii), and the condition does not have the degeneracy of case (iv).

This completes the proof. □

6.2. Probabilistic analysis

The results below hold in view of known urn theory [1], [7], [13]; see also an account of these results in textbook style in [10]. According to that theory, the number of balls of color j, $j = 1, \ldots, c$ in a c-color urn with certain properties, when appropriately normalized, converges in distribution to a multivariate normal random variable. The multivariate theorem needs several conditions to ensure tenability of the urn, the existence of a dominant color, and a certain eigenvalue structure of the ball replacement matrix. Tenability means that the ball drawing never stops, no matter which stochastic path is followed. Our urn is naturally tenable, as it corresponds to a graph associated with a physical process that can grow indefinitely. In the presence of negative entries on the diagonal, color c is dominant, if every other color can appear, upon starting with only balls of color c. Our urn is of this type, and color 1 can be seen to be dominant.

The rest of the conditions concern eigenvalues. Let the ball replacement matrix underlying the growth of the network be called $\boldsymbol{A}$, and let $\boldsymbol{A}^\top$ be its transpose. Suppose that we arranged the eigenvalues of $\boldsymbol{A}^\top$ according to their descending values of the real part, i.e. they are of the form

$$\operatorname{Re}\lambda_1 \geq \operatorname{Re}\lambda_2 \geq \operatorname{Re}\lambda_3.$$

Hence, λ_1 is the *principal eigenvalue*, the one with the largest real part. We call the corresponding eigenvector the *principal eigenvector*. The urn theory [1], [7], [13] is applicable, delivering strong laws and central limit theorems, if λ_1 is simple and is an eigenvalue of the submatrix of the dominant colors (all three in our case, and the submatrix alluded to is the entire ball replacement matrix). The three eigenvalues in all our cases are

$$\lambda_1 = |E| - 1, \qquad \lambda_2 = -d_1, \qquad \lambda_3 = -d_2^*,$$

https://doi.org/10.1017/jpr.2016.11 Published online by Cambridge University Press

where $|E| = n_1 + n_2 + n_3$. However, each case has different eigenvectors. In the sequel, we use $\xrightarrow{\text{a.s.}}$ to indicate almost sure convergence. For random vectors $\boldsymbol{Y}_n$ and $\boldsymbol{Y}$, the notation $\boldsymbol{Y}_n \xrightarrow{\text{a.s.}} \boldsymbol{Y}$ is for component-wise almost sure convergence.

Proposition 3. *Let $D_n^{(i)}$ be the number of nodes of admissible outdegree d_i^* for $i = 1, 2$ in a bipolar network after n steps of evolution. Let $\boldsymbol{D}_n$ be the vector with these two components. We then have*

$$\frac{1}{n}\boldsymbol{D}_n \xrightarrow{\text{a.s.}} (|E| - 1)\begin{pmatrix} \dfrac{v_{1,\text{case}}}{d_1} \\ \dfrac{v_{2,\text{case}}}{d_2^*} \end{pmatrix} =: (|E| - 1)\boldsymbol{v}_{\text{case}}^*,$$

where $v_{i,\text{case}}$ is the ith component of $\boldsymbol{v}_{\text{case}}$, $i = 1, 2$, and $\boldsymbol{v}_{\text{case}}$ is the (normalized to length 1) principal eigenvector of the transpose of the replacement matrix, depending on the case.

Proof. Consider the Pólya urn underlying the growth of the network. Let $X_n^{(i)}$ be the number of balls of color i in the urn, for $i = 1, 2, 3$, after n draws, and let

$$\boldsymbol{X}_n = \begin{pmatrix} X_n^{(1)} \\ X_n^{(2)} \\ X_n^{(3)} \end{pmatrix}.$$

We appeal to a classic urn result by Athreya and Karlin [1] (see also [7] and [13]) that relates the proportion of balls of various colors to the principal eigenvalue and eigenvector. In our case, this result translates into the following:

$$\frac{1}{n}\boldsymbol{X}_n \xrightarrow{\text{a.s.}} \lambda_1 \boldsymbol{v}_{\text{case}}.$$

We have $\lambda_1 = |E| - 1$, and $\boldsymbol{v}_{\text{case}}$ is the corresponding (normalized) principal eigenvector of $\boldsymbol{A}^\top$ (of three components) depending on the case. A calculation yields $\boldsymbol{v}_{\text{case}}$ for each of the five cases as stated in the theorem.

The d_i^* edges out of a node of outdegree d_i^* for $i = 1, 2$ account for one node in the network with this outdegree. That is, $D_n^{(1)} = X_n^{(1)}/d_1$ and $D_n^{(2)} = X_n^{(2)}/d_2^*$, which completes the proof. □

In the five cases we have, the eigenvalues are the same, but the (normalized) principal eigenvectors (of $\boldsymbol{A}^\top$) are different for each case and somewhat complicated. In a local technical report the authors obtain the eigenvectors in all cases. These computations are available upon request. We only present the form of the simplest case (case (iv)) to give the reader an idea about the type of the calculations involved:

$$\boldsymbol{v}_{(\text{iv})} = \begin{pmatrix} \dfrac{n_1}{|E|} \\ \dfrac{(n_2 - d_2^*)|E| + n_1 d_2^*}{|E|(|E| - 1 + d_2^*)} \\ 1 - \dfrac{n_1}{|E|} - \dfrac{(n_2 - d_2^*)|E| + n_1 d_2^*}{|E|(|E| - 1 + d_2^*)} \end{pmatrix};$$

https://doi.org/10.1017/jpr.2016.11 Published online by Cambridge University Press

hence, we have

$$\boldsymbol{v}^*_{(\mathrm{iv})} = \begin{pmatrix} \dfrac{n_1}{|E|} \\ \dfrac{(n_2 - d_2^*)|E| + n_1 d_2^*}{|E|(|E| - 1 + d_2^*)d_2^*} \end{pmatrix}.$$

And so, in case (iv), we have

$$\frac{1}{n}\begin{pmatrix} D_n^{(1)} \\ D_n^{(2)} \end{pmatrix} \xrightarrow{\text{a.s.}} (|E| - 1)\begin{pmatrix} \dfrac{n_1}{|E|} \\ \dfrac{(n_2 - d_2^*)|E| + n_1 d_2^*}{|E|(|E| - 1 + d_2^*)d_2^*} \end{pmatrix}.$$

In the next theorem, we shall use the notation $\mathcal{N}_i(\mathbf{0}, \boldsymbol{\Sigma})$ to denote a centered multivariate normal vector with mean $\mathbf{0}$ (of i components) and an $i \times i$ covariance matrix $\boldsymbol{\Sigma}$ (of dimensions $i \times i$). We use the symbol $\xrightarrow{\text{D}}$ to indicate convergence in law (in distribution).

Theorem 2. *Let $D_n^{(1)}$ and $D_n^{(2)}$ be respectively the number of nodes of outdegree $d_1^* = d_1$ and d_2^* in the bipolar network, after n steps of evolution. Let $\boldsymbol{D}_n$ be the vector with these two components. We then have, as $n \to \infty$,*

$$\frac{\boldsymbol{D}_n - (|E| - 1)\boldsymbol{v}^*_{\text{case}} n}{\sqrt{n}} \xrightarrow{\text{D}} \mathcal{N}_2(\mathbf{0}, \boldsymbol{\Sigma}^*_{\text{case}})$$

for an effectively computable vector $\boldsymbol{v}^*_{\text{case}}$ and covariance matrix $\boldsymbol{\Sigma}^*_{\text{case}}$ that depend on the case of the bipolar network.

Remark 1. The proof is based on [13], which ascertains the existence of $\boldsymbol{\Sigma}^*_{\text{case}}$, but does not give it constructively. Certain urns in the class of extended Pólya urns in Smythe [13] fall into a class of urn schemes discussed in Janson [7]. The latter reference provides explicit formulæ for the covariance matrix. We shall see that the urns induced by all bipolar networks fall into Smythe's extended Pólya urn schemes, but only some are in Janson's class.

Proof. Consider the Pólya urn underlying the growth of the network. Let $X_n^{(i)}$ be the number of balls of color i in the urn for $i = 1, 2, 3$ after n draws, and let

$$\boldsymbol{X}_n = \begin{pmatrix} X_n^{(1)} \\ X_n^{(2)} \\ X_n^{(3)} \end{pmatrix}.$$

According to [13], asymptotic normality holds if tenability and the existence of a dominant color are satisfied, as well as a special condition concerning the eigenvalues. Namely, the condition is $\operatorname{Re}\lambda_2 < \frac{1}{2}\operatorname{Re}\lambda_1$. In all five cases of bipolar networks, our urns are tenable, color 1 is dominant, and we have $-d_1 = \operatorname{Re}\lambda_2 < \frac{1}{2}\operatorname{Re}\lambda_1 = |E| - 1$ (recall that $|E| \geq 2$). According to [13], we have the multivariate central limit theorem

$$\frac{\boldsymbol{X}_n - (|E| - 1)\boldsymbol{v}_{\text{case}} n}{\sqrt{n}} \xrightarrow{\text{D}} \mathcal{N}_3(\mathbf{0}, \boldsymbol{\Sigma}_{\text{case}}),$$

where $\boldsymbol{v}_{\text{case}}$ (depends on the case) is the (normalized) principal eigenvector of $\boldsymbol{A}^\top$.

https://doi.org/10.1017/jpr.2016.11 Published online by Cambridge University Press

This can be computed for each of the five cases of bipolar networks. Since $D_1^{(n)} = X_n^{(1)}/d_1$ and $D_2^{(n)} = X_n^{(2)}/d_2^*$, it follows that

$$\boldsymbol{\Sigma}^*_{\text{case}} = \begin{pmatrix} \dfrac{c_{1,1}}{d_1^2} & \dfrac{c_{1,2}}{d_1 d_2^*} \\ \dfrac{c_{2,1}}{d_1 d_2^*} & \dfrac{c_{2,2}}{(d_2^*)^2} \end{pmatrix},$$

where $c_{i,j}$ for $1 \le i, j \le 2$ are the corresponding entries of $\boldsymbol{\Sigma}_{\text{case}}$. □

7. Covariance computation

Finding the covariance matrix in a growing combinatorial construct is a notoriously difficult problem. Entire papers have been dedicated to the computation of one variance; see, for example, [11]. The difficulty usually is a manifestation of the phenomenon called the combinatorial explosion in the computation of moments, where each higher moment is significantly more complicated and given by longer formulae than the one before. Smythe [13] gives the asymptotic averages, but does not give an explicit construction of the covariance matrix. It was not until some years later that Janson [7] gave a method to compute covariances. However, the class of urn schemes discussed in [7] is slightly different from that in [13], but has a fairly large intersection with it.

7.1. Bipolar networks fitting Janson's urn scheme

For a large class of bipolar networks, the work of Janson [7] gives a program for the computation of $\boldsymbol{\Sigma}_{\text{case}}$. That class is restricted to having all the diagonal elements of the ball replacement matrix being at least -1. Such would be the structure in any of the five cases, for sufficiently large values of n_1, n_2, and n_3.

For bipolar networks with such underlying Pólya urn structures, we outline the steps we need to extract from Janson's program to complete our calculation. Following [7], let $\boldsymbol{\xi}_i$ be the ith column of $\boldsymbol{A}^\top$, i.e.

$$\boldsymbol{\xi}_i = \begin{pmatrix} A_{i,1} \\ A_{i,2} \\ A_{i,3} \end{pmatrix} \quad \text{for } i = 1, 2, 3.$$

Let $\boldsymbol{u}_1$ and $\boldsymbol{v}_1 = \boldsymbol{v}_{\text{case}}$ denote left and right principal eigenvectors of $\boldsymbol{A}^\top$ corresponding to the principal eigenvalue λ_1, i.e. vectors satisfying

$$\boldsymbol{u}_1^\top \boldsymbol{A}^\top = \lambda_1 \boldsymbol{u}_1^\top, \qquad \boldsymbol{A}^\top \boldsymbol{v}_1 = \lambda_1 \boldsymbol{v}_1.$$

We assume that $\boldsymbol{u}_1$ and $\boldsymbol{v}_1$ are normalized such that

$$(1 \quad 1 \quad 1)\boldsymbol{v}_1 = \boldsymbol{v}_1^\top \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = 1, \qquad \boldsymbol{u}_1^\top \boldsymbol{v}_1 = \boldsymbol{v}_1^\top \boldsymbol{u}_1 = 1.$$

Similarly, let $\boldsymbol{u}_i$ and $\boldsymbol{v}_i$ be normalized left and right eigenvectors of $\boldsymbol{A}^\top$ corresponding to the eigenvalue λ_i, $i = 2, 3$. We next define the matrices

$$\boldsymbol{B}_i := \boldsymbol{\xi}_i \boldsymbol{\xi}_i^\top, \quad i = 1, 2, 3, \qquad \boldsymbol{B} := \sum_{i=1}^{3} v_{1,i} \boldsymbol{B}_i,$$

https://doi.org/10.1017/jpr.2016.11 Published online by Cambridge University Press

where $v_{1,i}$ is the ith component of $\boldsymbol{v}_1$. Now, the covariance matrix $\boldsymbol{\Sigma}_{\text{case}}$ is given by

$$\boldsymbol{\Sigma}_{\text{case}} = \sum_{j,k=2}^{3} \frac{\boldsymbol{u}_j^\top \boldsymbol{B}\boldsymbol{u}_k}{\lambda_1 - \lambda_j - \lambda_k} \boldsymbol{v}_j \boldsymbol{v}_k^\top.$$

Certainly the covariance matrix in any of the five cases is too lengthy for the page constraints of a journal publication, and would not fit in matrix form on a page. To give the reader an idea of what is involved, we only give one component of $\boldsymbol{\Sigma}_{\text{(iv)}} := [c_{j,j}]_{i,j=1,2}$, which is the simplest of all five cases. In case (iv), where $d_1 = 1$, when $n_2 > 2d^*$, we have

$$c_{1,1} = \frac{n_1(n_2 + n_3)}{|E|^2(|E| + 1)}.$$

Therefore, as $n \to \infty$, we have

$$\text{var}[D_n^{(1)}] \sim \frac{n_1(n_2 + n_3)}{|E|^2(|E| + 1)} n.$$

The other entries of the covariance matrix are much more complicated.

7.2. Cases not in Janson's urn scheme

All five cases of bipolar networks fit in Smythe's scheme. So, we always obtain a bivariate central limit theorem. However, numerous cases do not fit schemes covered by the covariance computations in [7]. For these we can obtain the covariance matrix on a case-by-case basis. In fact, in many cases we can develop it exactly, as well as asymptotically. We illustrate this on a bipolar network grown from the diamond seed shown Figure 3. This diamond has two types of outdegrees ($k = 2$), which are $d_1 = 1$, and $d_2 = 3$, with corresponding counts $n_1 = 2$ and $n_2 = 3$. As there are no nodes of outdegrees higher than 3, we take $n_3 = 0$. We also have $\Delta = 3$. By Theorem 1 the admissible outdegree sequence is $1, 3, 5, 7, \ldots$; we thus have $d_2^* = d_2 = 3$.

This bipolar network falls into case (iv), with ball replacement matrix

$$\begin{pmatrix} 1 & 3 & 0 \\ 2 & -3 & 5 \\ 2 & 0 & 2 \end{pmatrix}.$$

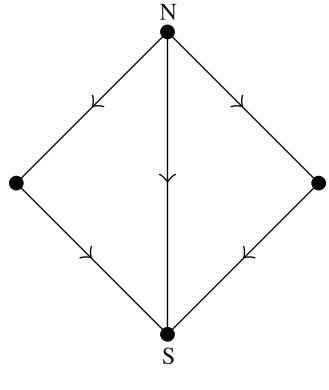


FIGURE 3: A diamond seed.

https://doi.org/10.1017/jpr.2016.11 Published online by Cambridge University Press

We can set up a stochastic recurrence for $X_n^{(1)}$. Let $\mathbf{1}_n^{(1)}$ be the indicator of the event of picking a ball of color 1 in the nth drawing. Note that, if we do not pick that color, we add two balls of color 1 to the urn, regardless of whether the ball picked is of color 2 or 3. We write the recurrence

$$X_n^{(1)} = X_{n-1}^{(1)} + \mathbf{1}_n^{(1)} + 2(1 - \mathbf{1}_n^{(1)}) = X_{n-1}^{(1)} - \mathbf{1}_n^{(1)} + 2.$$

Taking expectations, we obtain

$$\mathbb{E}[X_n^{(1)}] = \mathbb{E}[X_{n-1}^{(1)}] - \mathbb{E}[\mathbf{1}_n^{(1)}] + 2.$$

Let N_n be the network at time n; such a network has $4n+5$ edges in it. We have the conditional expectation

$$\mathbb{E}[\mathbf{1}_n^{(1)} \mid N_{n-1}] = \frac{X_{n-1}^{(1)}}{4n+1}.$$

We can take expectations and substitute the result into the recurrence for the expected number of balls of color 1, and obtain

$$\mathbb{E}[X_n^{(1)}] = \frac{4n}{4n+1}\mathbb{E}[X_{n-1}^{(1)}] + 2,$$

a recurrence to be solved under the initial condition $\mathbb{E}[X_0^{(1)}] = 2$. The solution is

$$\mathbb{E}[X_n^{(1)}] = \tfrac{8}{5}n + 2 \quad \text{for } n \geq 0.$$

The second moment can be obtained in similar steps, starting with squaring the stochastic recurrence for $X_n^{(1)}$, then taking expectations and solving the ensuing recurrence. When we subtract the square of the mean, we obtain

$$\operatorname{var}[X_n^{(1)}] = \frac{4}{25}n + \frac{1}{5} - \frac{\pi\sqrt{2}\Gamma(n+3/4)}{20\Gamma(n+5/4)\Gamma^2(3/4)} = \frac{4}{25}n + \frac{1}{5} + O\left(\frac{1}{n}\right);$$

we obtained the error term by Stirling's approximation to the gamma function.

In a similar manner we can write a stochastic recurrence for $X_n^{(2)}$. The exact average for balls of color 2 follows from

$$\mathbb{E}[X_n^{(2)}] = \frac{24}{35}n + \frac{6}{7} + \frac{(15/28)\sqrt{2\pi}\Gamma(n+1/2)}{\Gamma(3/4)\Gamma(n+5/4)}.$$

Multiplying the stochastic recurrence for $X_n^{(1)}$ by that for $X_n^{(2)}$, we can obtain a stochastic recurrence for $X_n^{(1)}X_n^{(2)}$, from which by averaging we obtain a recurrence for $\mathbb{E}[X_n^{(1)}X_n^{(2)}]$. An exact covariance is obtained, from which we get the asymptotic equivalent:

$$\operatorname{cov}[X_n^{(1)}X_n^{(2)}] \sim -\tfrac{141}{350}n.$$

By squaring the stochastic recurrence for $X_n^{(2)}$, then averaging, we obtain a recurrence for $\mathbb{E}[(X_n^{(2)})^2]$. That recurrence involves $\mathbb{E}[X_n^{(1)}X_n^{(2)}]$, which we now have. Performing all these computations, we obtain

$$\operatorname{var}[X_n^{(2)}] \sim \tfrac{3987}{2450}n.$$

Recall that

$$\begin{pmatrix} D_n^{(1)} \\ D_n^{(2)} \end{pmatrix} = \begin{pmatrix} X_n^{(1)} \\ \frac{1}{3}X_n^{(2)} \end{pmatrix},$$

https://doi.org/10.1017/jpr.2016.11 Published online by Cambridge University Press

and so, for a bipolar network grown from a diamond seed, the central limit theorem is

$$\left\{\begin{pmatrix} D_n^{(1)} \\ D_n^{(2)} \end{pmatrix} - \begin{pmatrix} \frac{8}{5} \\ \frac{8}{35} \end{pmatrix} n\right\} n^{-1/2} \xrightarrow{\mathrm{D}} \mathcal{N}_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{4}{25} & -\frac{47}{350} \\ -\frac{47}{350} & \frac{443}{2450} \end{pmatrix}\right).$$

8. Concluding remarks

We studied node outdegrees in self-similar directed bipolar networks. We looked at the outdegree of the North Pole. The indegree of nodes can be obtained by very similar methods. For instance, the indegree of the South Pole can be obtained just like that of the North Pole, if we let Δ be the indegree of the South Pole in the seed, and replace the outdegrees of the seed by its indegrees. Likewise the number of nodes of small indegrees mirrors that of small outdegrees in both methods and results. So, only very minor modifications are needed to study the degrees in self-similar undirected networks, since the degree of a node is the sum of the indegree and outdegree in a directed counterpart.

Acknowledgement

The authors would like to thank Panpan Zhang for valuable help in producing the graphics and preparing the final document.

References

[1] Athreya, K. B. and Karlin, S. (1968). Embedding of urn schemes into continuous time Markov branching processes and related limit theorems. *Ann. Math. Statist.* **39,** 1801–1817.
[2] David, F. N. and Barton, D. E. (1962). *Combinatorial Chance*. Hafner, New York.
[3] Drmota, M., Gittenberger, B. and Panholzer, A. (2008). The degree distribution of thickened trees. In *Proceedings of the Fifth Colloquium on Mathematics and Computer Science,* Association of Discrete Mathematics and Theoretical Computer Science, Nancy, pp. 149–162.
[4] Flajolet, P., Dumas, P. and Puyhaubert, V. (2006). Some exactly solvable models of urn process theory. In *Fourth Colloquium on Mathematics and Computer Science Algorithms, Trees, Combinatorics and Probabilities,* Association of Discrete Mathematics and Theoretical Computer Science, Nancy, pp. 59–118.
[5] Gopaladesikan, M., Mahmoud, H. and Ward, M. D. (2014). Building random trees from blocks. *Prob. Eng. Inf. Sci.* **28,** 67–81.
[6] Graham, R. L., Knuth, D. E. and Patashnik, O. (1994). *Concrete Mathematics*, 2nd edn. Addison-Wesley, Reading, MA.
[7] Janson, S. (2006). Limit theorems for triangular urn schemes. *Prob. Theory Relat. Fields* **134,** 417–452.
[8] Kuba, M. and Mahmoud, H. (2016) Two-color balanced affine urn models with multiple drawings. II. large index and triangular urns. Submitted. Available at http://arxiv.org/abs/1509.09053.
[9] Lobo, D., Vico, F. J. and Dassow, J. (2011). Graph grammars with string-regulated rewriting. *Theoret. Comput. Sci.* **412,** 6101–6111.
[10] Mahmoud, H. M. (2009). *Pólya Urn Models*. CRC, Boca Raton, FL.
[11] Matthews, P. C. and Rosenberger W. F. (1997). Variance in randomized play-the-winner clinical trials. *Statist. Prob. Lett.* **35,** 233–240.
[12] Rambo, L. R. (2010). Conversion studies, pastoral counseling, and cultural studies: engaging and embracing a new paradigm. *Pastoral Psychology* **59,** 433–445.
[13] Smythe, R. T. (1996). Central limit theorems for urn models. *Stoch. Process. Appl.* **65,** 115–137.
[14] Zhang, P., Chen, C. and Mahmoud, H. (2015). Explicit characterization of moments of balanced triangular Pólya urns by an elementary approach. *Statist. Prob. Lett.* **96,** 149–153.

https://doi.org/10.1017/jpr.2016.11 Published online by Cambridge University Press