Probability in the Engineering and Informational Sciences, **27**, 2013, 297–307.
doi:10.1017/S026996481300003X

SOME NODE DEGREE PROPERTIES OF SERIES–PARALLEL GRAPHS EVOLVING UNDER A STOCHASTIC GROWTH MODEL

Hosam M. Mahmoud
Department of Statistics
The George Washington University
Washington, D.C. 20052, U.S.A.
E-mail: hosam@gwu.edu

We introduce a natural growth model for directed series-parallel (SP) graphs and look at some of the graph properties under this stochastic model. Specifically, we look at the degrees of certain types of nodes in the random SP graph. We examine the degree of a pole and will find its exact distribution, given by a probability formula with alternating signs. We also prove that, for a fixed value s, the number of nodes of outdegree $1, \ldots, s$ asymptotically has a joint multivariate normal distribution. Pólya urns will systematically provide a working tool.

1. INTRODUCTION

Series–parallel (SP) graphs are network models that can represent the flow of, for example, commercial goods from a source to a market. To the best of our knowledge, these networks have been studied under two models of randomness: the uniform model, where all SP networks of a certain size are equally likely [2,4], and the hierarchical lattice model [5].

Yet another natural stochastic model can be considered, and may cover a wide variety of additional realistic applications. We introduce this model and study some of its properties (specifically, the degree distribution of certain types of nodes in the network under this growth model).

We use, as usual, the notation K_n for the complete graphs on n vertices. There are a few definitions of families of SP graphs. One popular variant of the definitions views SP graphs in an algorithmic constructive way. In this variant, the smallest SP graph is K_2. The two vertices are called *poles*. When drawn in a vertical position on a page, the top pole is called the *North Pole* (N), and the bottom one is called the *South Pole* (S). Larger SP graphs are obtained from smaller ones by one of two compositions: a *series composition*, which identifies the South Pole of a graph with the North Pole of the other, or a *parallel composition*, which identifies the two North Poles together, and the two South Poles together. For the network flow application we have in mind, we think of SP graphs as directed, with orientation assigned to the edges to allow the flow to move from the North Pole to the South Pole. Figure 1 shows two directed SP graphs, and two directed SP graphs that can be obtained from them by a series and a parallel composition.

 0269-9648/13 $25.00

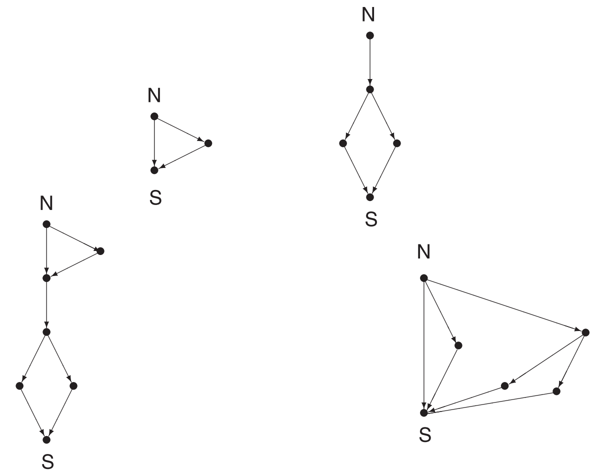

FIGURE 1. Two directed series-parallel graphs on top, and a graph obtained by a series composition (bottom left), and another graph obtained by their parallel composition (bottom right).

We have already mentioned two models of randomness: the uniform model and the hierarchical lattice model. We propose a rather natural growth model, which is different from these two. Starting with a directed K_2 (with its sole edge directed from the North to the South Pole), we grow an SP graph in steps. At each step, we choose an edge from the existing graph at random (all edges being equally likely). We subject that edge to either a series extension with probability p, or a parallel doubling with probability $q := 1 - p$. A *series extension* is performed as follows. If the edge chosen has u and v as end vertices (and points from u to v), we create a new vertex, say x, remove the directed edge uv and replace it with the two new directed edges ux (directed from u to x) and xv (directed from x to v). A *parallel doubling* of an edge is performed as follows. If the edge chosen has u and v as end vertices (and points from u to v), we just add another edge with these same endpoints and directed in the same sense. Henceforth "random" will always mean the model we have just introduced. Note that S_n, the size (number of nodes in the graph) after n edge additions to K_2, is 2 plus a binomial random variable on n trials, with rate of success p, and thus $(S_n - pn)/\sqrt{pqn}$ converges in distribution to the standard normal random variate.

In this orientation scheme, all the edges of the random SP graph are directed with arrows pointing away from the North Pole toward the South Pole. We qualify the entire orientation scheme with the phrase *away from the North Pole*. All the graphs in Figure 1 have an orientation away from the North Pole.

In this article we look at some properties of a random SP graph grown under the randomness model just discussed. Specifically, we look at the degrees of certain types of nodes in the random SP graph. We shall examine the degree of a pole. We find its exact distribution, given by a probability formula with alternating signs. We also prove that, for a fixed number s, the number of nodes of outdegree $1, \ldots, s$ asymptotically has a joint multivariate normal distribution. Pólya urns will systematically provide a working tool.

2. THE DEGREE OF A POLE

The number of edges coming out of the North Pole, or the number of trading routes emanating out of the source of the market, is a measure of the volume of trading and and the amount of goods that can be shipped out of the source. This number is the North Pole's outdegree (it is also its degree).

Suppose we colored the edges coming out of the North Pole with white (W), and all the other edges with blue (B). We think of the edges as balls in a Pólya urn. Let W_n be the number of white balls in the urn after n edge additions to K_2. As we start from a directed K_2, we have $W_0 = 1$, and $B_0 = 0$. At each stage we pick an edge at random (a ball from the urn at random). We also generate $\mathrm{Ber}(p)$, a Bernoulli random variable with success probability p. If $\mathrm{Ber}(p) = 1$ (success), we extend the chosen edge, otherwise we double the edge. If the edge is white, and the Bernoulli random variable indicates success, we are extending an edge connected to the North Pole; this adds one blue directed edge to the graph (one blue ball to the urn). If the edge is white, and the Bernoulli random variable indicates failure, we only double the edge, adding another white edge to the graph (a white ball to the urn). If, instead, we sample a blue edge (ball), we add a blue edge (ball), because no matter what operation is performed on the edge, it does not change the outdegree of the North Pole.

The dynamics of a two-color Pólya urn scheme are often represented with a replacement matrix, the rows and columns of which are indexed with the two colors, and the entries are the number of balls added. The replacement matrix associated with our urn is

$$\begin{pmatrix} 1-\mathrm{Ber}(p) & \mathrm{Ber}(p) \\ 0 & 1 \end{pmatrix};$$

the entry at position (C_1, C_2) represents the number of balls of color C_2 that we add upon withdrawing a ball of color C_1 from the urn, for $C_1, C_2 \in \{\mathrm{W}, \mathrm{B}\}$; the rows are indexed with W and B from top to bottom, and the columns are indexed with W and B from left to right.

It is shown in [6] how to get an exact distribution by solving a certain parametric pair of differential equations underlying an urn in $x(t)$ and $y(t)$ for an urn of this type. If $X(t, x(0))$ and $Y(t, y(0))$ are the solution, then $X^{W0}(t, (x(0))Y^{B_0}(t, y(0))$ is a history generating function. Specialized to our case, the differential equations are

$$\begin{aligned} x'(t) &= px(t)y(t) + qx^2(t), \\ y'(t) &= y^2(t). \end{aligned}$$

We solve this system under the initial condition $x(0) = u$, and $y(0) = v$, and get

$$\begin{aligned} x(t) &= \frac{uv}{u - uvt - (u-v)(1-vt)^p}, \\ y(t) &= \frac{v}{1-vt}. \end{aligned}$$

Following [6], these solutions give rise to the history generating function

$$\sum_{0 \le w,b,n < \infty} \mathbf{Prob}(W_n = w, B_n = b) u^w v^b z^n = \left(\frac{uv}{u - uvz - (u-v)(1-vz)^p}\right)^{W_0} \times \left(\frac{v}{1-vt}\right)^{B_0}.$$

Recall that $W_0 = 1$, and $B_0 = 0$. By setting $v = 1$, we get

$$\sum_{n=0}^{\infty} \sum_{w=0}^{\infty} \mathbf{Prob}(W_n = w) u^w z^n = \frac{u}{u - uz - (u-1)(1-z)^p}. \quad (1)$$

https://doi.org/10.1017/S026996481300003X Published online by Cambridge University Press

PROPOSITION 1: *Let W_n be the outdegree of the North Pole*[1] *in a random series-parallel graph. We then have*

$$\mathbf{E}[W_n] = \frac{(n+q)(n-1+q)\dots(1+q)}{n!} \sim \frac{1}{\Gamma(q+1)}\, n^q.$$

PROOF: Differentiate (1) once with respect to u, and set $u = 1$ to get a generating function of averages

$$\sum_{n=0}^{\infty} \mathbf{E}[W_n]\, z^n = (1-z)^{p-2}.$$

Extracting the nth coefficient, we get the exact average as stated.

Note that the average can be written in terms of Gamma functions:

$$\mathbf{E}[W_n] = \frac{\Gamma(n+q+1)}{n!\,\Gamma(q+1)}.$$

The asymptotic equivalent follows from Stirling's approximation of the Gamma function. ■

THEOREM 1: *Let W_n be the outdegree (indegree) of the North (South) Pole in a random series-parallel graph. Then, it has the exact probability distribution*

$$\mathbf{Prob}(W_n = w) = \sum_{k=0}^{w-1} (-1)^{n+k} \binom{qk-p}{n}\binom{w-1}{k}.$$

PROOF: Let $[x^i y^j]$ be the the operator that extracts the coefficient of $x^i y^j$ from a bivariate function of x and y. Extracting coefficients from (1), we have (for $|u| < 9^{p-1}/(8^{p-1}+9^{p-1})$, and $|z| < \frac{1}{8}$)

$$\begin{aligned}
\mathbf{Prob}(W_n = w) &= [u^w z^n]\left(\frac{u}{u - uz - (u-1)(1-z)^p}\right)\\
&= -[u^w z^n]\left(\frac{u}{(u-1)(1-z)^p} \times \frac{1}{1 - \dfrac{u}{(u-1)(1-z)^{p-1}}}\right)\\
&= -[u^w z^n]\frac{u}{(u-1)(1-z)^p}\sum_{k=0}^{\infty}\left(\frac{u}{(u-1)(1-z)^{p-1}}\right)^k\\
&= -[u^w z^n]\sum_{k=0}^{\infty} u^{k+1} \sum_{m=0}^{\infty}(-1)^{k+1+m}\binom{-k-1}{m}u^m\\
&\quad \times \sum_{n=0}^{\infty}(-1)^n\binom{-kp+k-p}{n}z^n\\
&= \sum_{k=0}^{w-1}(-1)^{n+w+1}\binom{-k-1}{w-k-1}\binom{qk-p}{n}.
\end{aligned}$$

[1] By symmetry, the South Pole has the same indegree distribution.

https://doi.org/10.1017/S026996481300003X Published online by Cambridge University Press

Noting that

$$\binom{-k-1}{w-k-1} = (-1)^{w+k+1}\binom{w-1}{k},$$

the result follows. ■

Remark: Probability formulas with alternating signs are remarkable and not always intuitive. There are quite a few of them that appear in similar contexts, such as the classic occupancy problem (see [3], a classic work of de Moivre). After all, probability is nonnegative, and somehow cancellations in the formula with alternating signs always occur in a way to produce a nonnegative answer.

3. NODES OF SMALL OUTDEGREE

The outdegree and indegree of a node in a trading network are indications of the local importance of a trading center to its neighbors. They determine how many neighbors will be affected, if the node becomes dysfunctional. The indegrees are symmetrical to the outdegrees, for we can imagine the polarity of the graph reversed (the sense of edge orientation leads away from the South Pole), and the indegrees with the old polarity will become outdegrees in the reversed graph. Therefore, it is sufficient to study the outdegrees of the SP graph under the original orientation.

In this section, we examine the distribution of the number of nodes of outdegree up to some fixed number, say s. Let us utilize $s+1$ colors to code the outdegrees. We color each edge out of a node of outdegree i with color $i = 1, \dots, s$; color $s+1$ is *special:* we color all the other edges with color $s+1$; these edges are pointing away from nodes of outdegree $s+1$ or higher. Again, think of the edges as balls in a Pólya urn. This urn evolves in the following way. If at stage n we pick an edge of a nonspecial color i (pointing away from a node of outdegree i), we either extend it (with probability p) into a path of two edges directed away from the North Pole, or double it (with probability q), and a new edge pointing out of the Northern end node is added. In the case of extending the chosen edge, we do not change the outdegree of the Northern end of the edge being extended; we only add a new node of outdegree 1 (a new edge of color 1). In the case of doubling, we change the degree of the Northern end of the edge being doubled—it is increased by 1. Thus, we remove i edges of color i, and add $i+1$ edges of color $i+1$. When we pick a special edge, we either increase the outdegree of its northern end, or keep it the same. If the operation is an extension, the number of special edges does not change, but we add one node of outdegree 1 (we add an edge of color 1). If the operation is the doubling of the special edge, the outdegree of the node at the Northern end of the edge goes up by 1 (we add an edge with the special color).

Let us represent the dynamics of this $(s+1)$-Pólya urn scheme with a ball replacement matrix, the rows and columns of which are indexed with the $s+1$ colors, and the entries are the number of balls added. Let B be a Bernoulli random variable with success probability p. The replacement matrix associated with our urn is

$$\mathbf{A} = \begin{pmatrix} 2B-1 & 2(1-B) & 0 & \cdots & 0 & 0 \\ B & -2(1-B) & 3(1-B) & \cdots & 0 & 0 \\ B & 0 & -3(1-B) & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ B & 0 & 0 & \cdots & -s(1-B) & (s+1)(1-B) \\ B & 0 & 0 & \cdots & 0 & 1-B \end{pmatrix};$$

the rows (from top to bottom) and columns (from right to left) of this matrix are indexed with the numbers (colors) $1, \dots, s+1$. The entry in row i and column j represents the number of balls of color j that we add upon withdrawing a ball of color i from the urn, for $i, j = 1, 2, \dots, s+1$. Note that the sum across any row of the replacement matrix is 1. Pólya urn schemes satisfying such a condition are called *balanced*. They enjoy the property that, regardless of the stochastic path followed, τ_n, the total number of balls in the urn after n draws is deterministic; in our case it is

$$\tau_n = n + 1.$$

Let $X_n^{(r)}$ be the number of edges in the SP graph of color r after the random insertion of n edges, and let $\mathbf{X}_n$ be the vector with the $s+1$ components $X_n^{(1)}, X_n^{(2)}, \dots, X_n^{(s+1)}$. Strong limit laws and asymptotic distributions are known for this type of balanced urn (where all the rows add up to the same constant, which is 1 in our case).

Assume the eigenvalues of $\mathbf{E}[\mathbf{A}]$ are numbered according to the decreasing order of their real parts:

$$\Re\lambda_1 \ge \Re\lambda_2 \ge \cdots \ge \Re\lambda_{s+1}.$$

The eigenvalue with largest real part, λ_1, is called the *principal eigenvalue*, and the corresponding eigenvector is a *principal eigenvector*. It is shown in [1] that for urns of this type

$$\frac{X_n^{(r)}}{n} \xrightarrow{a.s.} \lambda_1 v_r,$$

where λ_1 is the principal eigenvalue of the average of the replacement matrix, and $\mathbf{v} = (v_1, v_2, \dots, v_{s+1})$ is the corresponding principal left eigenvector of $\mathbf{E}[\mathbf{A}]$. Also, under the condition that λ_2, the eigenvalue with second largest real part satisfies $\Re\lambda_2 < \frac{1}{2}\lambda_1$, it is shown in [7] that

$$\frac{\mathbf{X}_n - \lambda_1 \mathbf{v}}{\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathbf{\Sigma}),$$

for some covariance matrix $\mathbf{\Sigma}$. Smythe [7] states that $\mathbf{\Sigma}$ is generally hard to compute. We shall prove a multivariate central limit theorem of this type. In fact, we shall obtain an exact form for the covariance matrix of $\mathbf{X}_n$, for $s = 2$, and show how in principle we can extend this to higher values of s.

To deal with the exact mean and covariances, we derive the recurrence equations from the dynamics of the construction. Let $\mathcal{F}_n$ be the sigma field generated by the the first n edge insertions. Let $I_n^{(r)}$ be the indicator of the event that an edge of color r is picked at the nth draw. For color 1, we write the conditional recurrence

$$\begin{aligned}\mathbf{E}[X_n^{(1)} \,|\, \mathcal{F}_{n-1}] = X_{n-1}^{(1)} &+ \mathbf{E}[(2B-1) I_n^{(1)} \,|\, \mathcal{F}_{n-1}] \\ &+ \mathbf{E}[B I_n^{(2)} \,|\, \mathcal{F}_{n-1}] \\ &\ \ \vdots \\ &+ \mathbf{E}[B I_n^{(s+1)} \,|\, \mathcal{F}_{n-1}].\end{aligned}$$

Noting the independence of B and $\mathcal{F}_{n-1}$, we write the latter equation as

$$\begin{aligned}\mathbf{E}[X_n^{(1)} \,|\, \mathcal{F}_{n-1}] = X_{n-1}^{(1)} &+ (2p-1)\,\mathbf{E}[I_n^{(1)} \,|\, \mathcal{F}_{n-1}] \\ &+ p\,\mathbf{E}[I_n^{(2)} \,|\, \mathcal{F}_{n-1}] + \cdots + p\,\mathbf{E}[I_n^{(s+1)} \,|\, \mathcal{F}_{n-1}].\end{aligned}$$

https://doi.org/10.1017/S026996481300003X Published online by Cambridge University Press

The indicator $I_n^{(r)}$ is a Bernoulli random variable $\mathrm{Ber}(X_{n-1}/\tau_{n-1})$ that conditionally (given $\mathcal{F}_{n-1}$) has the expectation $X_{n-1}^{(r)}/\tau_{n-1}$. The conditional expectation for the first color then takes the form

$$\mathbf{E}[X_n^{(1)} \,|\, \mathcal{F}_{n-1}] = X_{n-1}^{(1)} + (2p-1)\,\frac{X_{n-1}^{(1)}}{n} + p\,\frac{X_{n-1}^{(2)}}{n} + \cdots + p\,\frac{X_{n-1}^{(s+1)}}{n}.$$

Note that the coefficients of the random variables come down spanning the entries of the average of the first column of the replacement matrix.

Writing a similar equation for each color, and putting them in matrix form, we get

$$\mathbf{E}[\mathbf{X}_n \,|\, \mathcal{F}_{n-1}] = \left(\mathbf{I} + \frac{1}{n}\,\mathbf{E}[\mathbf{A}^T]\right)\mathbf{X}_{n-1},$$

where $\mathbf{I}$ is the $(s+1)\times(s+1)$ identity matrix, and $\mathbf{A}^T$ is the transpose of $\mathbf{A}$. We can take expectations and write

$$\mathbf{E}[\mathbf{X}_n] = \left(\mathbf{I} + \frac{1}{n}\,\mathbf{E}[\mathbf{A}^T]\right)\mathbf{E}[\mathbf{X}_{n-1}] := \mathbf{R}_n\,\mathbf{E}[\mathbf{X}_{n-1}].$$

This form can be iterated, and we get

$$\mathbf{E}[\mathbf{X}_n] = \mathbf{R}_n\mathbf{R}_{n-1}\ldots\mathbf{R}_1\mathbf{E}[\mathbf{X}_0]. \tag{2}$$

Observe that the eigenvalues of $\mathbf{E}[\mathbf{A}]$ are

$$\lambda_1 = 1, \quad \text{and} \quad \lambda_r = -(r-1)q, \quad \text{for } r = 2,\ldots,s+1.$$

The eigenvalues are real and distinct, with $\lambda_2 = -q < \frac{1}{2} = \frac{1}{2}\lambda_1$. As the eigenvalues are distinct, they give rise to simple Jordan normal forms—the matrix $\mathbf{R}_j$ can be written as

$$\mathbf{M}\,\mathbf{D}_j\,\mathbf{M}^{-1} = \mathbf{M}\begin{pmatrix} 1+\frac{1}{j} & 0 & 0 & \ldots & 0 & 0\\ 0 & 1-\frac{q}{j} & 0 & \ldots & 0 & 0\\ \vdots & \vdots & & \ddots & \vdots & \vdots\\ 0 & 0 & 0 & \ldots & 0 & 1-\frac{sq}{j}\end{pmatrix}\mathbf{M}^{-1},$$

where $\mathbf{M}$ is the modal matrix[2] of $\mathbf{E}[\mathbf{A}^T]$, which is invertible, because the eigenvalues are distinct. Eq. (2) can now be simplified to

$$\begin{aligned}\mathbf{E}[\mathbf{X}_n] &= (\mathbf{M}\,\mathbf{D}_n\,\mathbf{M}^{-1})(\mathbf{M}\,\mathbf{D}_{n-1}\,\mathbf{M}^{-1})\cdots(\mathbf{M}\,\mathbf{D}_1\,\mathbf{M}^{-1})\mathbf{E}[\mathbf{X}_0]\\ &= \mathbf{M}\,\mathbf{D}_n\mathbf{D}_{n-1}\cdots\mathbf{D}_1\mathbf{M}^{-1}\begin{pmatrix}1\\0\\0\\\vdots\\0\end{pmatrix}.\end{aligned} \tag{3}$$

[2] The modal matrix of a given matrix with distinct eigenvalues is the matrix formed by placing the jth eigenvector as the jth column.

We thus have the exact vector of means:

$$\mathbf{E}[\mathbf{X}_n] = \frac{1}{n!}\,\mathbf{M}\begin{pmatrix} \Gamma(n+1) & 0 & 0 & \dots & 0 & 0 \\ 0 & \frac{\Gamma(n+1-q)}{\Gamma(1-q)} & 0 & \dots & 0 & 0 \\ \vdots & \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & \frac{\Gamma(n+1-sq)}{\Gamma(1-sq)} \end{pmatrix}\mathbf{M}^{-1}\begin{pmatrix}1\\0\\0\\ \vdots\\0\end{pmatrix}.$$

We illustrate this program with the small instance $s = 2$.

THEOREM 2: *Let $Y_n^{(r)}$ be the number of nodes of outdegree $r \in \{1, 2\}$ in a random directed series-parallel graph, and let $\mathbf{Y}_n$ be the vector with these two components. We have*

$$\mathbf{E}[Y_n^{(1)}] = \frac{p(n+1)}{q+1} + \frac{2q\,\Gamma(n+p)}{(q+1)\,\Gamma(p)\,\Gamma(n+1)},$$

$$\mathbf{E}[Y_n^{(2)}] = \frac{pq(n+1)}{(2q+1)(q+1)} + \frac{4^p q\,\Gamma(p-\frac{1}{2})\,\Gamma(n+p)}{2\sqrt{\pi}(q+1)\,\Gamma(-1+2p)\Gamma(n+1)} - \frac{3q\,\Gamma(n-1+2p)}{(2q+1)\,\Gamma(-1+2p)\,\Gamma(n+1)}.$$

Also, $\mathbf{Y}_n$ converges in distribution to a bivariate normal vector:

$$\frac{\mathbf{Y}_n - \begin{pmatrix}\frac{p}{q+1}\\ \frac{pq}{(2q+1)(q+1)}\end{pmatrix} n}{\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} \frac{2pq(3-p)}{(2-p)^2(3-2p)} & -\frac{2p^2q}{(4-3p)(3-2p)(2-p)^2} \\ -\frac{2p^2q}{(4-3p)(3-2p)(2-p)^2} & \frac{pq(24p^4-157p^3+356p^2-342p+120)}{(5-4p)(4-3p)(3-2p)^2(2-p)^2}\end{pmatrix}\right).$$

PROOF: Note that $Y_n^{(1)} = X_n^{(1)}$, and $Y_n^{(2)} = \frac{1}{2}X_n^{(2)}$. Therefor it suffices to get the results for $X_n^{(1)}$ and $X_n^{(2)}$.

Consider the 3×3 replacement matrix corresponding to $s = 2$. For this we have

$$\mathbf{M} = \begin{pmatrix} \frac{p(3-2p)}{6q^2} & -\frac{1}{3} & 0 \\ \frac{p}{3q} & -\frac{2}{3} & -1 \\ 1 & 1 & 1. \end{pmatrix}$$

Multiplying out, as required in (3), the exact averages for the number of edges of colors 1 and 2 follow, after some lengthy simplification.

The exact second moments are given by rather lengthy expressions. Here, we only set up the recurrence equations and indicate how to solve them. We shall show the derivation in a bit of detail for$X_n^{(1)}$, and will skip most of the details for the second moment of $X_n^{(2)}$ and for the covariance between the counts of the two colors.

https://doi.org/10.1017/S026996481300003X Published online by Cambridge University Press

For exact second moment of $X_n^{(1)}$, we start with a recurrence obtained from the 3×3 replacement matrix:

$$X_n^{(1)} = X_{n-1}^{(1)} + B - (1 - B)\mathrm{Ber}\left(\frac{X_{n-1}^{(1)}}{n}\right). \qquad \textbf{(4)}$$

Squaring both sides, we get

$$(X_n^{(1)})^2 = (X_{n-1}^{(1)})^2 + B + (1 - B)\mathrm{Ber}\left(\frac{X_{n-1}^{(1)}}{n}\right) + 2BX_{n-1}^{(1)} - 2(1 - B)X_n^{(1)}\mathrm{Ber}\left(\frac{X_{n-1}^{(1)}}{n}\right).$$

So, the conditional second moment for this color is

$$\mathbf{E}[(X_n^{(1)})^2 \,|\, \mathcal{F}_{n-1}] = (X_{n-1}^{(1)})^2 + p + q\frac{X_{n-1}^{(1)}}{n} + 2pX_{n-1}^{(1)} - 2qX_{n-1}^{(1)}\frac{X_{n-1}^{(1)}}{n}.$$

This gives a recurrence for the (unconditional) second moment:

$$\mathbf{E}[(X_n^{(1)})^2] = \left(1 - \frac{2q}{n}\right)\mathbf{E}\big[\big(X_{n-1}^{(1)}\big)^2\big] + \left(2p + \frac{q}{n}\right)\mathbf{E}[X_{n-1}^{(1)}] + p.$$

Plug in $\mathbf{E}[X_n^{(1)}]$, which we have developed. This recurrence, and several other in the sequel, are of the general form

$$a_n = \left(1 - \frac{b}{n}\right)a_{n-1} + h(n), \qquad \textbf{(5)}$$

for constant b and known asymptotically linear function $h(n)$, with an asymptotically quadratic solution. The solution to the recurrence for $\mathbf{E}[(X_n^{(1)})^2]$ is

$$\begin{aligned}\mathbf{E}[(X_n^{(1)})^2] = {} & \frac{p(q^2 + 3q + qp + 2qpn + 2 + pn - p)}{(2 - p)(1 + 2q)(1 + q)}(n + 1) \\ & + (-p^4 + 2p^2 + 7p^2q^2 - 2qp - 4q^3p - 5q^2p - 4q^2 + 4q^4 \\ & + 2q^3 - p - 2q + 3p^2q)\,\Gamma(n + 1 - 2q) \\ & \times ((1 + q)(1 + 2q)(p + 2q)(-1 + p + 2q) \\ & \times\ \Gamma(n + 1)\,\Gamma(1 - 2q))^{-1} \\ & + (2(2p^2n - 2pn + 4qpn + 5qp + 2q^2))\,\Gamma(n + p) \\ & \times ((-2 + p)(p + 2q)(-1 + p + 2q)\Gamma(n + 1)\,\Gamma(-1 + p))^{-1}.\end{aligned}$$

After subtracting off the square of the mean, and computing asymptotics, a linear asymptotic variance ensues (as $n \to \infty$):

$$\mathbf{Var}[X_n^{(1)}] \sim \frac{2pq(3 - p)}{(2 - p)^2(3 - 2p)}\,n.$$

For the second moment of $X_n^{(2)}$ and the covariance, we only sketch the key steps. We start from a stochastic recurrence (again obtained from the dynamics of the construction):

$$X_n^{(2)} = X_{n-1}^{(2)} + 2(1 - B)\left[\mathrm{Ber}\left(\frac{X_{n-1}^{(1)}}{n}\right) - \mathrm{Ber}\left(\frac{X_{n-1}^{(2)}}{n}\right)\right]. \qquad \textbf{(6)}$$

https://doi.org/10.1017/S026996481300003X Published online by Cambridge University Press

Multiply (4) and (6), and take expectation (handling the Bernoulli random variables via a double expectation). Take into consideration that $\mathrm{Ber}(\frac{X_{n-1}^{(1)}}{n})$ and $\mathrm{Ber}(\frac{X_{n-1}^{(2)}}{n})$ are mutually exclusive (if one of them is 1, the other must be 0). This gives an exact recurrence for the mixed moment $\mathbf{E}[X_n^{(1)} X_n^{(2)}]$. This recurrence involves $\mathbf{E}[(X_n^{(1)})^2]$, which we already have. Thus, the recurrence is in the form of (5). We solve the recurrence and obtain the exact mixed moment $\mathbf{E}[X_n^{(1)} X_n^{(2)}]$. Extracting leading asymptotics, we get a linear covariance equivalence, as ($n \to \infty$):

$$\mathbf{Cov}[X_n^{(1)}, X_n^{(2)}] \sim -\frac{4p^2 q}{(4-3p)(3-2p)(2-p)^2}\, n.$$

Finally, square (6), and take expectations. The resulting recurrence has the expectations of $X_n^{(1)}$ and $X_n^{(2)}$, as well as the expectation of their product. We already have all these ingredients in exact form. We plug in the results we have and solve the recurrence (also in the form of (5)) to get $\mathbf{E}[(X_n^{(2)})^2]$. Subtracting off the square of $\mathbf{E}[X_n^{(2)}]$, we get an exact variance. The formula is so huge to be listed, and we only give its linear asymptotic equivalent:

$$\mathbf{Var}[\mathbf{E}[X_n^{(2)}] = \frac{4pq(24p^4 - 157p^3 + 356p^2 - 342p + 120)}{(5-4p)(4-3p)(3-2p)^2(2-p)^2}\, n.$$

■

It is evident that we can extend this computation to higher s. Considering the rth column of the replacement matrix, for colors $r = 2, \ldots, s$, we have the recurrence

$$X_n^{(r)} = X_{n-1}^{(r)} + r(1-B)\left[\mathrm{Ber}\left(\frac{X_{n-1}^{(r-1)}}{n}\right) - \mathrm{Ber}\left(\frac{X_{n-1}^{(r)}}{n}\right)\right]. \qquad (7)$$

with $\mathrm{Ber}\left(\frac{X_{n-1}^{(r-1)}}{n}\right)$ and $\mathrm{Ber}\left(\frac{X_{n-1}^{(r)}}{n}\right)$ being mutually exclusive (if one of them is 1, the other must be 0). The averages of $X_n^{(r)}$, for $r = 1, \ldots, s$, can be obtained by a bootstrapped program: we obtain inductively the average number of edges of one color and plug it in the recurrence for the average number of edges of the next color, with $\mathbf{E}[X_n^{(1)}]$ serving as basis for the induction. Working with asymptotics, as $n \to \infty$, drastically simplifies the appearance of the average counts to

$$\mathbf{E}[X_n^{(r)}] \sim \frac{r!\, pq^{r-1}}{(rq+1)((r-1)q+1)\ldots(q+1)}.$$

The average number of nodes of outdegree r is then $\frac{1}{r}\mathbf{E}[X_n^{(r)}]$.

The variances and covariances are significantly more computationally intensive. Nevertheless, the steps are clear. It is also a bootstrapped program in the fashion of dynamic programming: obtain all the results up to color $r-1$ (in addition to all first moments, obtain all the mixed moments $\mathbf{E}[X^{(i)} X^{(j)}]$, for $i, j = 1, \ldots, r-1$). Now, write a recurrence for $\mathbf{E}[X_n^{(1)} X_n^{(r)}]$, which we obtain by taking the product of (4) and (7), then averaging. This mixed moment will involve some first moments and the mixed moment $\mathbf{E}[X_n^{(1)} X_n^{(r-1)}]$, so it is in the form of (5). We then move on to a recurrence for $\mathbf{E}[X_n^{(2)} X_n^{(r)}]$, which beside the recursive term will involve only moments computed so far, and so the recurrence is in the form of (5). We can then proceed in a similar fashion via recurrences for $\mathbf{E}[X_n^{(r')} X_n^{(r)}]$, for

$r' \le r$, and it will involve, beside the recursive term, only already computed moments, so these recurrences are in the form of (5). Ultimately, for $\mathbf{X}_n$ being the vector with components $X_n^{(1)}, \ldots, X_n^{(s)}$, we have the strong law

$$\frac{X_n^{(r)}}{n} \xrightarrow{a.s.} \frac{r!\,pq^{r-1}}{(rq+1)(r-1)q+1)\ldots(q+1)},$$

and with Smythe's condition $-q = \Re\lambda_2 < \frac{1}{2}\lambda_1 = 1$ being satisfied, we have the multivariate central limit theorem

$$\frac{\mathbf{X}_n - \begin{pmatrix} \frac{p}{q+1} \\ \frac{2pq}{(2q+1)(q+1)} \\ \vdots \\ \frac{r!\,pq^{r-1}}{(rq+1)((r-1)q+1)\ldots(q+1)} \end{pmatrix} n}{\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathbf{\Sigma}),$$

for some effectively computable covariance matrix $\mathbf{\Sigma}$.

References

1. Athreya, K. & Karlin, S. (1968). Embedding of urn schemes into continuous time Markov branching process and related limit theorems. *The Annals of Mathematical Statistics* **39**: 1801–1817.
2. Bernasconi, N., Panagiotou, K., & Steger, A. (2008). On the degree sequences of random outerplanar and series-parallel graphs. *Lecture Notes in Computer Science* **5171**: 303–316.
3. De Moivre, A. (1712). *The Doctrine of Chances*. London: Millar. Reprinted in 1967 by Chelsea, New York.
4. Drmota, M., Giménez, O., & Noy, M. (2010). Vertices of given degree in series-parallel graphs. *Random Structures & Algorithms* **36**, 273–314.
5. Hambly, B. & Jordan, J. (2004). A random hierarchical lattice: the series-parallel graph and its properties. *Advances in Applied Probability* **36**: 824–838.
6. Morcrette, B. & Mahmoud, H. (2012). Exactly solvable balanced tenable urns with random entries via the analytic methodology. *Discrete Mathematics and Theoretical Computer Science* **proc.**, 219–232.
7. Smythe, R. (1996). Central limit theorems for urn models. *Stochastic Processes and Their Applications* **65**: 115–137.