Structural Transitions in Densifying Networks

R. Lambiotte,[1] P. L. Krapivsky,[2] U. Bhat,[2,3] and S. Redner[3,*]
[1]naXys, Namur Center for Complex Systems, University of Namur, rempart de la Vierge 8, B 5000 Namur, Belgium
[2]Department of Physics, Boston University, Boston, Massachusetts 02215, USA
[3]Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, New Mexico 87501, USA
(Received 14 July 2016; published 16 November 2016)

We introduce a minimal generative model for densifying networks in which a new node attaches to a randomly selected target node and also to each of its neighbors with probability p. The networks that emerge from this copying mechanism are sparse for $p < \frac{1}{2}$ and dense (average degree increasing with number of nodes N) for $p \geq \frac{1}{2}$. The behavior in the dense regime is especially rich; for example, individual network realizations that are built by copying are disparate and not self-averaging. Further, there is an infinite sequence of structural anomalies at $p = \frac{2}{3}, \frac{3}{4}, \frac{4}{5}$, etc., where the N dependences of the number of triangles (3-cliques), 4-cliques, undergo phase transitions. When linking to second neighbors of the target can occur, the probability that the resulting graph is complete—all nodes are connected—is nonzero as $N \to \infty$.

DOI: 10.1103/PhysRevLett.117.218301

The investigation of complex networks has blossomed into a rich discipline, with many theoretical advances and a myriad of applications to the physical and social sciences [1–3]. Network science has identified universal properties that are shared by a wide range of real-world systems, including small worldness, heterogeneous degree distributions, and network densification. The latter, observed in a variety of social, urban, and information networks [4–7], is a fundamental phenomenon where the number of edges in a network grows superlinearly with the number of nodes N. However, the vast majority of network models exclusively focus on *sparse* networks, where the average degree is finite as $N \to \infty$. The main purpose of this Letter is to introduce a minimal model for *dense* networks and to analytically determine many of its fascinating structural properties, including: (i) a densification transition, (ii) an infinite sequence of transitions in clique densities, (iii) an anomalous degree distribution for dense networks, and (iv) a completeness transition with second-neighbor copying.

Our model is based on the generic mechanism of copying (see also Refs. [4,8–12] for related modeling): new nodes are introduced sequentially and each connects to a random preexisting target node, as well as to each of the neighbors of the target (friends of a friend) independently with probability p (Fig. 1). This mechanism, related to triadic closure, is known to drive the dynamics of social networks [13–15], such as Facebook, where people are invited to connect to a friend of a friend (see, e.g., Refs. [11,12]), but also information networks [4,11,16,17] and biological networks, through the concept of duplication [10,18–21]. Copying naturally generates highly clustered, small-world networks [22–26] and has the further advantage of being a local mechanism [27–29], a feature that allows us to obtain precise results. As we will show, sufficient copying triggers instabilities in the network growth, leading to the emergence of network densification, and produces nontrivial structural properties, including an infinite sequence of phase transitions in the densities of fixed-size cliques (complete subgraphs), as well as nonextensivity and lack of self-averaging of the degree distribution. Moreover, the simplicity of the mechanism allows for analytical solution for many network properties.

When $p = 0$, a network built by copying is a random recursive tree [30,31], while for $p = 1$, a complete graph arises if the initial graph is also complete. For $p < \frac{1}{2}$, the network is sparse, while for $p \geq \frac{1}{2}$, the number of links grows superlinearly with N and the network is *dense*. In the dense regime, the network is highly clustered (Fig. 2) and undergoes an infinite series of structural transitions at $p = \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \ldots$ that signal sudden changes in the growth laws of the number of 3-cliques (triangles), 4-cliques (tetrahedra), etc.

Number of links.—We first investigate how copying affects the growth in the number of links. Let L_N denote the number of links in a given realization of a network of N nodes and let $L(N) \equiv \langle L_N \rangle$ denote the number of links averaged over many realizations. Adding a new node

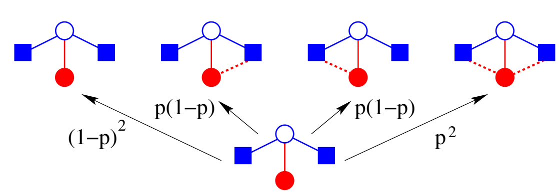

FIG. 1. The copying mechanism. A new node (filled circle) attaches to a random target (open circle) and to each of the friends of the target (squares) with probability p.

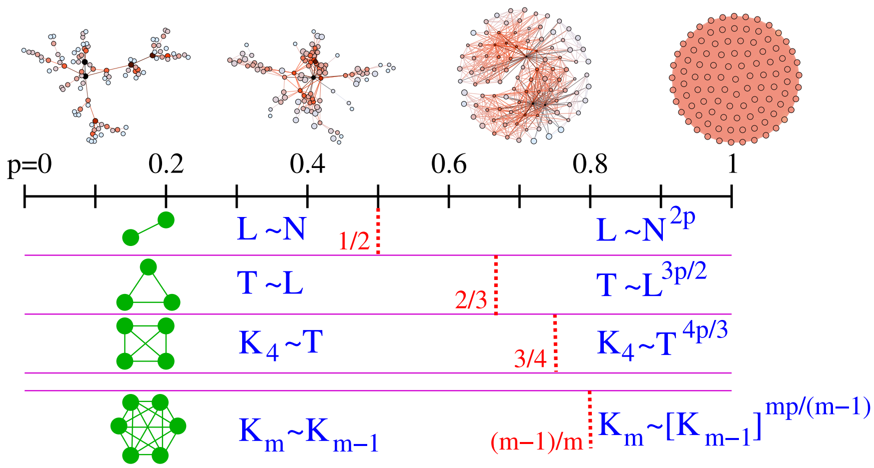

FIG. 2. Realizations of the copying model for $p = 0.1$, 0.4, 0.7, and 1 for $N = 100$, and a summary of the dense regimes.

increases $L(N)$ by $1 + p\langle k\rangle$, where $\langle k\rangle = 2L(N)/N$ is the average degree. Thus $L(N)$ grows according to

$$L(N+1) = L(N) + 1 + 2p\frac{L(N)}{N}. \tag{1}$$

The exact solution to this recurrence is [32]

$$L(N) = \frac{\Gamma(2p+N)}{\Gamma(N)}\sum_{j=2}^{N}\frac{\Gamma(j)}{\Gamma(2p+j)} \tag{2}$$

where $\Gamma(\cdot)$ is the Euler gamma function. The large-N asysmptotic behavior of this solution

$$L(N) = \begin{cases} N/(1-2p), & p < \frac{1}{2}; \\ N\ln N, & p = \frac{1}{2}; \\ A(p)N^{2p}, & \frac{1}{2} < p \le 1; \end{cases} \tag{3}$$

with $A(p) = [(2p-1)\Gamma(1+2p)]^{-1}$, illustrates the change in the N dependence at $p = \frac{1}{2}$.

The standard deviation $\Sigma_L \equiv \sqrt{\langle L_N^2\rangle - \langle L_N\rangle^2}$ exhibits an even richer dependence on N, with transitions at $p = \frac{1}{4}$ and $p = \frac{1}{2}$ [32]:

$$\Sigma_L \sim \begin{cases} \sqrt{N}, & p < \frac{1}{4}; \\ \sqrt{N\ln N}, & p = \frac{1}{4}; \\ N^{2p}, & \frac{1}{4} < p < 1, p \neq \frac{1}{2}; \\ N\sqrt{\ln N}, & p = \frac{1}{2}. \end{cases} \tag{4}$$

The salient consequences of Eqs. (3) and (4) are that $L(N)$ grows superlinearly with N and is *not* self-averaging for $p > \frac{1}{2}$. That is, there is a wide diversity between different network realizations starting from the same initial condition—the first few steps are crucial in shaping the evolution. Conversely, fluctuations are negligible only in the sparse phase, where $\Sigma_L/L(N) \to 0$ as $N \to \infty$; for $p < \frac{1}{4}$, where Σ_L scales as $\sqrt{N}$, we further anticipate that the distribution in the number of links for a network of N nodes, $P(L, N)$, is asymptotically Gaussian.

Triangles and larger cliques.—A related set of transitions occurs in the densities of larger-size cliques. A k-clique is a complete subgraph of k nodes that are connected by $k(k-1)/2$ links. We first investigate the number of 3-cliques (triangles). There are two mechanisms that increase the number of triangles as a result of a copying event—direct and induced linking. In direct linking, a triangle is created in each copying event that consists of the new node, the target node, and the neighbor of the target that receives a "copying" link (Fig. 3). In induced linking, additional triangles are created whenever copying creates links to more than one neighbor of the target that were previously linked to each other.

To determine the N dependence of the average number of triangles $T(N)$, suppose that the target node has degree k and that its neighbors are connected via c "clustering" links (Fig. 3). If a copying links are made, the increase in the number of triangles, ΔT, is

$$\Delta T = a + \frac{a(a-1)}{2}\frac{c}{k(k-1)/2}. \tag{5}$$

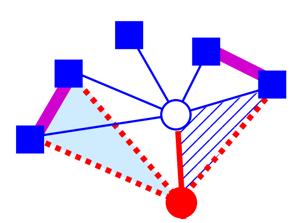

FIG. 3. Counting triangles. The target node (open circle) has five neighbors (squares), two of which were previously joined by "clustering" links (heavy lines). Three copying links (dashed) create three new triangles by direct linking (one is hatched for illustration) and one new triangle by induced linking (shaded).

The first term on the right accounts for direct linking and the second for induced linking. For the latter, we need to count how many of $a(a-1)/2$ possible links between a neighbors of the target, which also connect to the new node, are actually present. Averaging Eq. (5) with respect to the binomial distribution for a, we obtain, after an elementary calculation,

$$\overline{\Delta T} = pk + p^2 c. \tag{6}$$

The term $p^2 c$ arises because two connected neighbors of the target also connect to the new node with probability p^2, since linking to each node occurs independently.

We now express the average number of clustering links $\langle c \rangle$ in terms of the number of triangles $T(N)$. To this end, we note that c equals the number of triangles that contain the target node, $\langle c \rangle = 3T(N)/N$. Using these relations, the average number of triangles increases by $\overline{\Delta T(N)} = 3p^2 T(N)/N + 2pL(N)/N$ with each node addition. While this recursion can be solved exactly [32], the detailed solution is cumbersome and not illuminating. It is simpler to specialize to the $N \gg 1$ limit, where the above recursion reduces to the rate equation

$$\frac{dT(N)}{dN} = 3p^2 \frac{T(N)}{N} + 2p \frac{L(N)}{N}, \tag{7}$$

whose solution is

$$T(N) = \begin{cases} \frac{2p}{(1-2p)(1-3p^2)} N, & p < \frac{1}{2}; \\ 4N \ln N, & p = \frac{1}{2}; \\ \frac{A(p)}{1-3p/2} N^{2p}, & \frac{1}{2} < p < \frac{2}{3}; \\ \frac{4}{\Gamma(4/3)} N^{4/3} \ln N, & p = \frac{2}{3}; \\ B(p) N^{3p^2}, & \frac{2}{3} < p \le 1; \end{cases} \tag{8}$$

with $A(p)$ given in Eq. (3) and $B(p)$ also calculable [32] by solving the discrete recursion for $T(N)$.

Thus the average number of triangles $T(N)$ undergoes *two* transitions, with the second at $p = \frac{2}{3}$, where $T(N)$ grows superlinearly in $L(N)$ (Fig. 2). Moreover, the density of triangles converges to a nonvanishing value even in the sparse regime of $p < \frac{1}{2}$, which mirrors the high density of triangles found in many complex networks [22–26].

The reasoning presented above can be generalized to 4-cliques (quartets) and we find that their number grows according to the rate equation [32]

$$\frac{dK_4(N)}{dN} = 3p^2 \frac{T(N)}{N} + 4p^3 \frac{K_4(N)}{N}, \tag{9}$$

from which the average number of quartets grows as (with all prefactors omitted)

$$K_4(N) \sim \begin{cases} N, & 0 < p < \frac{1}{2}; \\ N^{2p}, & \frac{1}{2} < p < \frac{2}{3}; \\ N^{3p^2}, & \frac{2}{3} < p < \frac{3}{4}; \\ N^{4p^3}, & \frac{3}{4} < p \le 1. \end{cases}$$

At the transition points $p = \frac{1}{2}, \frac{2}{3}$, and $\frac{3}{4}$, the algebraic factor is multiplied by $\ln N$. Generally, the average number $K_m(N)$ of m-cliques evolves according to

$$\frac{dK_m(N)}{dN} = (m-1)p^{m-2} \frac{K_{m-1}(N)}{N} + mp^{m-1} \frac{K_m(N)}{N}. \tag{10}$$

Solving Eq. (10) recursively gives

$$K_m(N) \sim N^{(j+1)p^j} \qquad \frac{j}{j+1} < p < \frac{j+1}{j+2}, \tag{11}$$

with $j = 0, 1, 2, \ldots, m-1$. Thus the N dependence of the average number of cliques of size m undergoes $m-1$ transitions at $p = 1 - \frac{1}{n}$ with $n = 2, \ldots, m$.

Degree distribution.—Let N_k be the number of nodes of degree k. Following standard reasoning [28,33], the degree distribution evolves according to

$$\frac{dN_k}{dN} = \frac{N_{k-1} - N_k}{N} + p \frac{(k-1)N_{k-1} - kN_k}{N} + m_k. \tag{12a}$$

The first term on the right is the contribution due to attachment to the target node, the second term accounts for attachments to the neighbors of the target node, and the third term

$$m_k \equiv \sum_{s \ge k-1} n_s \binom{s}{k-1} p^{k-1} (1-p)^{s-k+1} \tag{12b}$$

is the probability that the new node acquires a degree k because it attaches to a target of degree s and to $k-1$ neighbors of this target. Here $n_s = N_s/N$ denotes the fraction of nodes of degree s.

When the network is sparse and large, we assume that the fractions n_k do not depend on N to recast Eq. (12) as

$$[2 + p(k+1)] n_{k+1} = [1 + pk] n_k + \sum_{s \ge k} n_s \binom{s}{k} p^k (1-p)^{s-k}. \tag{13}$$

This equation is not a recurrence, but it is still possible to extract the asymptotic behavior of n_k. First, we observe that for large k, the summand on the right is sharply peaked around $s \approx k/p$ and thus reduces to [20,33]

$$n_{k/p}\sum_{s\geq k}\binom{s}{k}p^k(1-p)^{s-k} = p^{-1}n_{k/p},$$

where we use a binomial identity to compute the sum itself. Substituting this into Eq. (13) and assuming that n_k decays slower than exponentially so that differences may be replaced by derivatives, we obtain the nonlocal equation for the degree distribution

$$\frac{d}{dk}[1+pk]n_k = p^{-1}n_{k/p} - n_k. \tag{14}$$

The algebraic form $n_k \sim k^{-\gamma}$ solves this equation and also gives the transcendental relation for the exponent,

$$\gamma = 1 + p^{-1} - p^{\gamma-2}, \tag{15}$$

which admits two solutions. One, $\gamma = 1$, is unphysical because the corresponding degree distribution is not normalizable. The other applies when $0 \leq p < \frac{1}{2}$, where $\gamma \equiv \gamma(p)$ decreases monotonically with p, with $\gamma(0) = \infty$ and $\gamma(\frac{1}{2}) = 2$. Because $\gamma > 2$ for $0 \leq p < \frac{1}{2}$ the average degree $\langle k\rangle = \sum_{k\geq 1} k n_k$ is finite so that the network is indeed sparse for $0 \leq p < \frac{1}{2}$.

In the dense regime, the scaling ansatz $n_k = N_k/N$ fails (see Fig. 4) and many features of the degree distribution become anomalous. For example, the distribution does not self-average, nodes of small finite degree are absent in sufficiently large networks, and the distribution appears to slowly converge to a form that is visually close to, but distinct from, a log-normal as $N \to \infty$ (Fig. 4). The resolution of the degree distribution in this regime represents an intriguing challenge.

Network completeness.—Finally, suppose that a new node connects to the neighbors of the target with probability p *and* to the second neighbors of the target with probability q. Such a mechanism naturally arises in social media, where connections to friends of a friend can extend to higher-order acquaintances. The unexpected feature of second-order linking is that the network is complete with nonzero probability.

Let $\mathcal{C}(N)$ denote the probability that a network of N nodes always remains complete for connection probabilities p and q. This completeness probability is

$$\mathcal{C}(N) = \prod_{r=1}^{N-1}\sum_{k=0}^{r-1} \mathrm{B}(r,k,p)[1-(1-q)^k]^{r-k-1}, \tag{16}$$

where $\mathrm{B}(r,k,p) = \binom{r-1}{k}p^k(1-p)^{r-k-1}$ is the binomial probability that copying leads to k links to the neighbors of the target. The second factor is the probability that *all* of the remaining $r-k-1$ neighbors of the target are connected by second-order links.

Asymptotic analysis and numerical evaluation of Eq. (16) show that $\mathcal{C}(N)$ indeed converges to a nonzero, albeit extremely small, value [32]. A more relevant criterion is not defect-free completeness, but whether the number of links eventually scales as $N^2/2$, as in the complete graph. Simulations show that for representative values of p and q, $L(N)$ initially grows linearly with N but then crosses over to growing as $N^2/2$ (Fig. 5). Thus second-order copying generically leads to networks that are effectively complete—eventually each individual knows almost everybody. Moreover, Fig. 5 illustrates that individual network realizations are macroscopically disparate. This intriguing feature also arises in empirical networks and related models [34–36], and intellectually originates with the classic Pólya urn model [37–39].

To summarize, we introduced a simple generative model for network densification based on the copying mechanism that leads to rich structural properties. A dense network arises for copying probability $p \geq \frac{1}{2}$. This regime further partitions into disjoint windows where the densities of k-cliques each have distinct scaling properties. Different network realizations starting from the same initial state are extremely diverse and all features of the resulting degree distribution are unconventional. When second-neighbor connections are made, the network asymptotically becomes

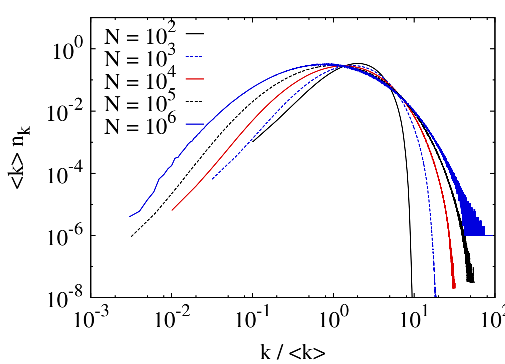

FIG. 4. Simulations of $10^{10}/N$ realizations for the degree distributions n_k for $p = 0.75$ (dense regime) and various N.

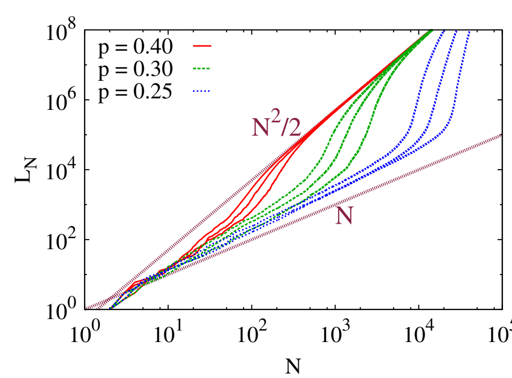

FIG. 5. N dependence of the number of links, with three realizations for each value of p, for second-neighbor copying with $q = p^2$.

complete. These theoretical findings provide a simple mechanism for the emergence of network densification in real-world networks, and calls for future empirical analyses of the scaling of elemental network motifs with network size.

Financial support for this research was provided in part by the ARC (Action de Recherche Concerté) on Mining and Optimization of Big Data Models, funded by the Wallonia-Brussels Federation, and the Belgian Network DYSCO (Dynamical Systems, Control and Optimisation), funded by the Interuniversity Attraction Poles Programme initiated by the Belgian State Science Policy Office (R. L.)., DMR-1608211 and 1623243 from the National Science Foundation (U. B. and S. R.), the John Templeton Foundation (S. R.), and Grant No. 2012145 from the United States Israel Binational Science Foundation (U. B.).

[*]Corresponding author.
redner@santafe.edu

[1] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[2] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford University Press, Oxford, England, 2003).
[3] M. E. J. Newman, *Networks: An Introduction* (Oxford University Press, Oxford, England, 2010).
[4] J. Leskovec, J. Kleinberg, and C. Faloutsos, ACM Trans. Knowl. Discovery Data **1**, 2 (2007).
[5] Y.-Y. Ahn, S. Han, H. Kwak, S. Moon, and H. Jeong, *Proceedings of the 16th International Ionference on World Wide Web* (ACM, New York, 2007), p. 835.
[6] L. M. A. Bettencourt, D. I. Kaiser, and J. Kaur, J. Inform. **3**, 210 (2009).
[7] M. Schläpfer, L. M. A. Bettencourt, S. Grauwin, M. Raschke, R. Claxton, Z. Smoreda, G. B. West, and C. Ratti, J. R. Soc. Interface **11**, 20130789 (2014).
[8] J. M. Kleinberg, R. Kumar, P. Raghavan, S. Rajagopalan, and A. S. Tomkins, *Proceedings of the International Conference on Combinatorics and Computing* (Springer, Berlin, Heidelberg, 1999), pp. 1–18.
[9] R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, and E. Upfal, *Proceedings of the 41st Annual Symposium on Foundations of Computer Science* (IEEE, 2000), pp. 57–65.
[10] A. Vázquez, A. Flammini, A. Maritan, and A. Vespignani, Complexus **1**, 38 (2003).
[11] P. L. Krapivsky and S. Redner, Phys. Rev. E **71**, 036118 (2005).
[12] M. O. Jackson and B. W. Rogers, Am. Econ. Rev. **97**, 890 (2007).
[13] M. Granovetter, Am. J. Sociology **78**, 1360 (1973).
[14] R. Toivonen, L. Kovanen, M. Kivelä, J.-P. Onnela, J. Saramäki, and K. Kaski, Soc. Networks **31**, 240 (2009).
[15] A. Ilany and E. Akçay, Nat. Commun. **7**, 12084 (2016).
[16] S. Valverde and R. V. Solé, Europhys. Lett. **72**, 858 (2005).
[17] T. S. Evans and A. D. K. Plato, Phys. Rev. E **75**, 056101 (2007).
[18] S. Ohno, *Evolution by Gene Duplication* (Springer-Verlag, Berlin, 1970).
[19] J. S. Taylor and J. Raes, Annu. Rev. Genet. **38**, 615 (2004).
[20] I. Ispolatov, P. L. Krapivsky, and A. Yuryev, Phys. Rev. E **71**, 061911 (2005).
[21] I. Ispolatov, P. L. Krapivsky, I. Mazo, and A. Yuryev, New J. Phys. **7**, 145 (2005).
[22] J. Davidsen, H. Ebel, and S. Bornholdt, Phys. Rev. Lett. **88**, 128701 (2002).
[23] P. Holme and B. J. Kim, Phys. Rev. E **65**, 026107 (2002).
[24] R. Toivonen, J.-P. Onnela, J. Saramaki, J. Hyvönen, and K. Kaski, Physica A **371A**, 851 (2006).
[25] G. Bianconi, R. K. Darst, J. Iacovacci, and S. Fortunato, Phys. Rev. E **90**, 042806 (2014).
[26] U. Bhat, P. L. Krapivsky, and S. Redner, J. Stat. Mech. (2014) P11035.
[27] A. Vázquez, Phys. Rev. E **67**, 056104 (2003).
[28] P. L. Krapivsky and S. Redner, Phys. Rev. E **63**, 066123 (2001).
[29] T. S. Evans and J. P. Saramäki, Phys. Rev. E **72**, 026138 (2005).
[30] R. T. Smythe and H. Mahmoud, Theory Probab. Math. Stat. **51**, 1 (1995).
[31] M. Drmota and H.-K. Hwang, Adv. Appl. Probab. **37**, 321 (2005).
[32] U. Bhat, P. L. Krapivsky, R. Lambiotte, and S. Redner, arXiv:1610.01662.
[33] J. Kim, P. L. Krapivsky, B. Kahng, and S. Redner, Phys. Rev. E **66**, 055101 (2002).
[34] M. J. Salganik, P. S. Dodds, and D. J. Watts, Science **311**, 854 (2006).
[35] O. Riordan and L. Warnke, Phys. Rev. E **86**, 011129 (2012).
[36] M. Ostilli, Europhys. Lett. **105**, 28005 (2014).
[37] A. A. Markov, Izv. Acad. Sci. **11**, 177 (1917).
[38] F. Eggenberger and G. Pólya, Z. Angew. Math. Mech. **3**, 279 (1923).
[39] H. M. Mahmoud, *Pólya Urn Models* (CRC Press, Boca Raton FL, 2009).