Structural Transitions in Dense 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, NM, 87501*

We introduce an evolving network model in which a new node attaches to a randomly selected target node and also to each of its neighbors with probability p. The resulting network is sparse for $p < \frac{1}{2}$ and dense (average degree increasing with number of nodes N) for $p \geq \frac{1}{2}$. In the dense regime, individual networks realizations built by this copying mechanism 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 dependences on N 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—where all nodes are connected—is non-zero as $N \to \infty$.

PACS numbers: 89.75.-k, 02.50.Le, 05.50.+q, 75.10.Hk

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 [1–6]. Much of the focus has been on *sparse* networks, where the average degree, defined as the average number of links attached to a node, is finite as the number of nodes in the network $N \to \infty$. In this letter, we introduce a minimal generative model for *dense* networks, in which the average degree grows with N.

In addition to the many new phenomena that arise in the dense regime, such networks may account for structural properties of the brain, which, for humans, has average degree 10^3 (10^{11} neurons, 10^{14} interconnections). The brain exhibits a rich spectrum of motifs—small subsets of densely interconnected nodes [7–9] that may underlie its wondrous functionality. Densification also appears to arise in many empirical networks, including, for example, the arXiv citation, patent citation, and autonomous systems graphs [10]. As we present below, our model displays many intriguing features that may mirror some of these structural properties, including a sequence of phase transitions in the densities of fixed-size cliques (complete subgraphs), non-extensivity of the degree distribution, and lack of self averaging.

Our model is based on the generic mechanism of copying (see also Ref. [10]): new nodes are introduced sequentially and each connects to a random pre-existing target node, as well as to each the neighbors of the target (friends of a friend) independently with probability p (Fig. 1). This mechanism drives the dynamics of social networks [11–13], as well as social media, such as Facebook, where people are invited to connect to a friend of a friend (see, e.g., [14, 15]). Copying is also related to triadic closure [16–21], which naturally generates highly clustered networks.

The copying and related redirection mechanisms are ubiquitous in networks; they underlie the world-wide web, citation, and other information networks [10, 22–24], the evolutionary process of gene duplication [25, 26], and protein interaction networks [27–34]. Finally, copying is local [35–37], as the creation of new links only depends on the local neighborhood of the target node, contrary to global rules such as preferential attachment [1–6]. As we will show, copying leads to highly non-trivial networks, but the simplicity of this mechanism allows for analytical solution for many network properties.

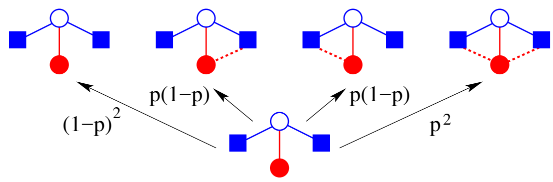

FIG. 1: The copying mechanism. A new node (filled circle) attaches to a random target (open circle) and to each of the friends of the target (squares) with probability p.

When $p = 0$, a network built by copying is a random recursive tree [38–40], 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 average number of links in a network of N nodes. Adding a new node increases the number of links, on average, by $1 + p\langle k\rangle$, where $\langle k\rangle = 2L_N/N$ is the average degree. Thus L_N grows according to

$$L_{N+1} = L_N + 1 + 2p\,\frac{L_N}{N}. \qquad (1)$$

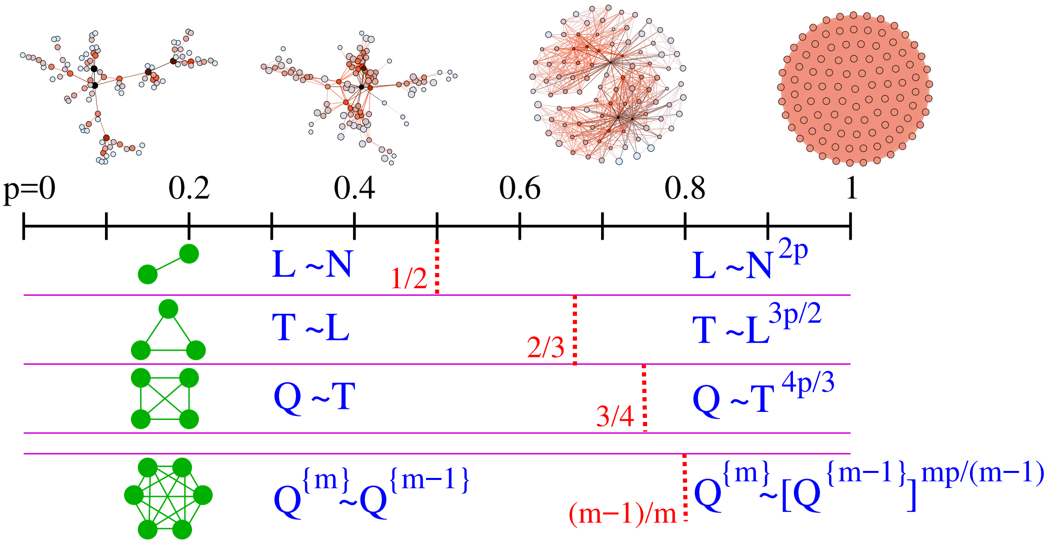

FIG. 2: Realizations of the copying model for $p = 0.1$, 0.4, 0.7, and 1 for $N = 100$, and a summary of the dense regimes.

Taking the continuum $N \gg 1$ limit and solving the resulting differential equation gives

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

with amplitude $A(p) = \big[(2p-1)\Gamma(1+2p)\big]^{-1}$ that is calculable by solving the discrete recursion (1). Indeed, the recurrence (1) admits the exact solution

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

from which the asymptotics (2) and the amplitude $A(p)$ follow [41].

We can also compute [41] the standard deviation $\Sigma_L \equiv \sqrt{\langle L_N^2\rangle - \langle L_N\rangle^2}$, which exhibits an even richer dependence on N, with transitions at $p = \frac{1}{4}$ and $p = \frac{1}{2}$:

$$\Sigma_L \sim \begin{cases} \sqrt{N} & p < \frac{1}{4}, \\ \sqrt{N\,\ln N} & p = \frac{1}{4}, \\ N^{2p} & \frac{1}{4} < p < 1, p \neq \frac{1}{2}, \\ N\,\sqrt{\ln N} & p = \frac{1}{2}. \end{cases} \tag{3}$$

The salient consequences of (2) and (3) are that L_N grows superlinearly with N and is *not* self averaging for $p > \frac{1}{2}$. This feature implies that there is a wide diversity among different network realizations, and the first few steps are crucial in shaping the evolution. Conversely, fluctuations are negligible in the sparse phase, where $\Sigma_L/\langle L_N\rangle \to 0$ as $N \to \infty$. Only for $p < \frac{1}{4}$, where Σ_L scales as $\sqrt{N}$, do we anticipate that the distribution $P(L_N)$ 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 of $k(k-1)/2$ links that completely connect k nodes within the network. 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 a 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 an induced linking, additional triangles are created whenever copying creates links to more than one neighbor of the target that were previously linked.

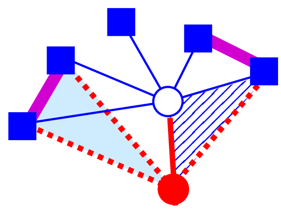

FIG. 3: Counting triangles. The target node (open circle) has five neighbors (squares), two of which are joined by 'clustering' links (heavy lines). Three copying links (dashed) create three new triangles (one is hatched for illustration) and one new triangle by induced linking (shaded).

To determine the N-dependence of average number of triangles, 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 number of triangles increases on average by

$$\Delta T = a + \frac{a(a-1)}{2}\,\frac{c}{k(k-1)/2}. \tag{4}$$

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 (4) with respect to the binomial distribution for a, we obtain, after an elementary calculation,

$$\overline{\Delta T} = pk + p^2 c. \tag{5}$$

The term p^2c 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 $\langle\overline{\Delta T_N}\rangle = 3p^2T_N/N + 2pL_N/N$ with each node addition. For $N \gg 1$, we thus obtain the rate equation

$$\frac{dT_N}{dN} = 3p^2\frac{T_N}{N} + 2p\frac{L_N}{N}\,, \tag{6}$$

whose solution is

$$T_N = \begin{cases} \frac{2p}{(1-2p)\,(1-3p^2)}\,N & p<\frac{1}{2},\\ 4N\,\ln N & p=\frac{1}{2},\\ \frac{A(p)}{1-3p/2}\,N^{2p} & \frac{1}{2}<p<\frac{2}{3},\\ \frac{4}{\Gamma(4/3)}\,N^{4/3}\ln N & p=\frac{2}{3},\\ B(p)\,N^{3p^2} & \frac{2}{3}<p\leq 1,\end{cases} \tag{7}$$

with $A(p)$ given in (2) and $B(p)$ also calculable [41] by solving the discrete recursion for T_N.

Equation (7) shows that the average number of triangles T_N undergoes a second phase transition at $p=\frac{2}{3}$ where T_N grows superlinearly in L_N (Fig. 2). Moreover, the density of triangles converges to a non-vanishing value when $0<p<\frac{1}{2}$, which mirrors the high density of triangles found in many complex networks [16–21].

The reasoning presented above can be generalized to 4-cliques (quartets) and we find that their number grows according to the rate equation [41]

$$\frac{dQ_N}{dN} = 3p^2\frac{T_N}{N} + 4p^3\frac{Q_N}{N}, \tag{8}$$

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

$$Q_N \sim \begin{cases} N & 0<p<\frac{1}{2},\\ N^{2p} & \frac{1}{2}<p<\frac{2}{3},\\ N^{3p^2} & \frac{2}{3}<p<\frac{3}{4},\\ N^{4p^3} & \frac{3}{4}<p\leq 1.\end{cases}$$

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 $Q_N^{\{m\}}$ of m-cliques evolves according to

$$\frac{dQ_N^{\{m\}}}{dN} = (m-1)p^{m-2}\frac{Q_N^{\{m-1\}}}{N} + mp^{m-1}\frac{Q_N^{\{m\}}}{N}\,. \tag{9}$$

This behavior is analogous to what occurs in duplication-divergence networks [34]. Solving (9) recursively gives

$$Q_N^{\{m\}} \sim N^{(j+1)p^j} \qquad \tfrac{j}{j+1}<p<\tfrac{j+1}{j+2}\,, \tag{10}$$

with $j=0,1,2,\ldots,m-1$. Thus the N-dependence of the average number of cliques of size m undergoes 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 [30, 36], the degree distribution evolves according to

$$\frac{dN_k}{dN} = \frac{N_{k-1}-N_k}{N} + p\,\frac{(k-1)N_{k-1}-kN_k}{N} + m_k. \tag{11a}$$

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

$$m_k \equiv \sum_{s\geq k-1} n_s\binom{s}{k-1}p^{k-1}(1-p)^{s-k+1} \tag{11b}$$

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 (11) to

$$[2+p(k+1)]n_{k+1} = [1+pk]n_k + \sum_{s\geq k} n_s\binom{s}{k}p^k(1-p)^{s-k}. \tag{12}$$

This equation is not a recurrence, but it is still possible to extract its asymptotics. First, we observe that for large k, the summand on the right is sharply peaked around $s\approx k/p$ and thus reduces to [30, 33]

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

where we used a binomial identity to compute the sum itself. Substituting this into Eq. (12) and assuming that n_k decays slower than exponentially so that differences may be replaced by derivatives, we obtain the non-local equation for the degree distribution

$$\frac{d}{dk}\,[1+pk]n_k = p^{-1}\,n_{k/p} - n_k\,. \tag{13}$$

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

$$\gamma = 1 + p^{-1} - p^{\gamma-2}\,, \tag{14}$$

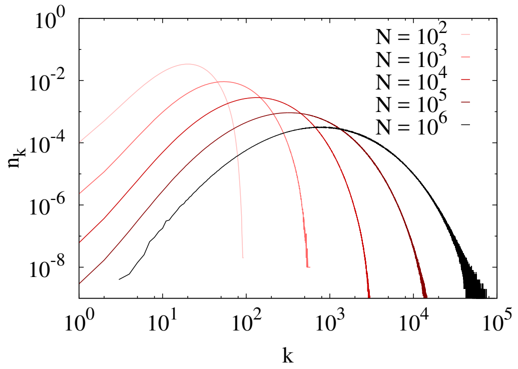

FIG. 4: Simulations of $10^{10}/N$ realizations for the degree distributions n_k for $p = 0.75$ (dense regime) and various N.

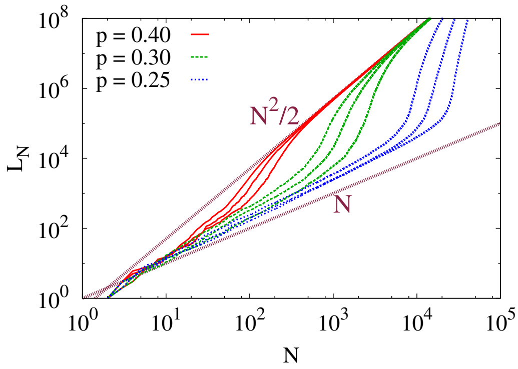

FIG. 5: The N dependence of the number of links for second-neighbor copying with $p_2 = p^2$.

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 $\gamma = \gamma(p)$ 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 analysis above no longer applies and we resort to simulations. We find that the degree distribution is no longer stationary; that is, n_k depends on N, in contrast to the stationarity in the sparse regime. Moreover, the degree distribution appears to slowly converge to a form that is close to log-normal as $N \to \infty$ (Fig. 4).

`Network Completeness.` Suppose that in addition to connecting to the neighbors of the target with probability p, a new node *also* connects to the second neighbors of the target with probability p_2. Such a mechanism naturally arises in social media, where connections to friends of a friend can extend to higher-order acquaintances. The surprising feature of second-order linking is that the network is complete with non-zero probability.

Let Π_N denote the probability that a network of N nodes always remains complete for connection probabilities p and p_2. This completeness probability is

$$\Pi_N = \prod_{r=1}^{N-1} \sum_{k=0}^{r-1} \mathrm{B}(r,k,p) \left(1-(1-p_2)^k\right)^{r-k-1}, \qquad (15)$$

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 (15) show that Π_N indeed converges to a non-zero, albeit extremely small, value [41]. 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 reasonable values of p and p_2, 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 the macroscopic differences between individual network realizations. Thus copying leads to non-self-averaging in the dense regime—unpredictable outcomes when starting from a fixed initial state. This intriguing feature also arises in empirical networks and related models [42–44], and intellectually originates with the classic Pólya urn model [45–47].

To summarize, we introduced a simple and rich generative model for dense networks based on the copying mechanism. The resulting network is dense for copying probability $p \geq \frac{1}{2}$. The dense regime partitions into distinct windows where the density of k-cliques each have unique scaling properties. Generally, different realizations of the network in the dense regime are extremely diverse. The degree distribution appears to slowly converge to a log-normal form in the dense regime. When second-neighbor connections can occur, the network asymptotically becomes complete.

Financial support for this research was also provided in part by the grants from the ARC and the Belgian Network DYSCO, funded by the IAP Programme (RL), DMR-1608211 and 1623243 from the National Science Foundation (UB and SR), the John Templeton Foundation (SR), and Grant No. 2012145 from the United States Israel Binational Science Foundation (UB).

[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, UK, 2003).

[3] M. E. J. Newman, SIAM Review **45** 167 (2003).

[4] M. E. J. Newman, A.-L. Barabási, and D. J. Watts, *The Structure and Dynamics of Networks*, (Princeton University Press, Princeton, NJ, 2006).
[5] A. Barrat, M. Barthélemy, and A. Vespignani, *Dynamical Processes on Complex Networks*, (Cambridge University Press, Cambridge, UK, 2008).
[6] M. E. J. Newman, *Networks: An Introduction*, (Oxford University Press, Oxford, 2010).
[7] V. M. Eguíluz, D. R. Chialvo, G. A. Cecchi, M. Baliki, and A. V. Apkarian, Phys. Rev. Lett. **94**, 018102 (2005).
[8] C. C. Hilgetag and M. Kaiser, Neuroinformatics **2**, 353 (2004).
[9] D. Meunier, R. Lambiotte, A. Fornito, K. D. Ersche, and E. T. Bullmore, Frontiers Neuroinformatics **3**, 37 (2009).
[10] J. Leskovec, J. Kleinberg, and C. Faloutsos, ACM Trans. Knowl. Discov. Data **1**, 1 (2007).
[11] M. Granovetter, Amer. J. Sociology **78**, 1360 (1973).
[12] R. Toivonen, L. Kovanen, M. Kivelä, J.-P. Onnela, J. Saramäki and K. Kaski, Social Networks **31**, 240 (2009).
[13] A. Ilany and E. Akçay, Nature Comm. **7**, 12084 (2016).
[14] M. O. Jackson and B. W. Rogers, Amer. Econ. Rev. **97**, 890 (2007).
[15] P. L. Krapivsky and S. Redner, Phys. Rev. E. **71**, 036118 (2005).
[16] A. Rapoport, Bull. Math. Biophys. **15**, 523 (1953).
[17] P. Holme and B. J. Kim, Phys. Rev. E **65**, 026107 (2002).
[18] J. Davidsen, H. Ebel, and S. Bornholdt, Phys. Rev. Lett. **88**, 128701 (2002).
[19] R. Toivonen, J.-P. Onnela, J. Saramaki, J. Hyvönen, and K. Kaski, Physica A **371**, 851 (2006).
[20] G. Bianconi, R. K. Darst, J. Iacovacci, and S. Fortunato, Phys. Rev. E **90**, 042806 (2014).
[21] U. Bhat, P. L. Krapivsky, and S. Redner, J. Stat. Mech. P11035 (2014).
[22] P. L. Krapivsky and S. Redner, Phys. Rev. E **71**, 036118 (2005).
[23] S. Valverde and R. V. Solé, Europhys. Lett. **72**, 858–64 (2005).
[24] T. Evans and A. D. K. Plato, Phys. Rev. E. **75**, 056101 (2007).
[25] S. Ohno, *Evolution by Gene Duplication* (Springer-Verlag, New York, 1970).
[26] J. S. Taylor and J. Raes, Annu. Rev. Genet. **9**, 615 (2004).
[27] R. V. Solé, R. Pastor-Satorras, E. D. Smith, and T. Kepler, Adv. Complex Syst. **5**, 43 (2002).
[28] F. Chung, L. Lu, G. Dewey, and D. J. Galas, J. Comput. Biology **10**, 677 (2003).
[29] A. Vázquez, A. Flammini, A. Maritan, and A. Vespignani, ComPlexUs **1**, 38 (2003).
[30] J. Kim, P. L. Krapivsky, B. Kahng, and S. Redner, Phys. Rev. E. **66**, 055101 (2002).
[31] G. C. Conant and A. Wagner, Genome Research **13**, 2052 (2003).
[32] N. V. Dokholyan, B. Shakhnovich, and E. I. Shakhnovich, Proc. Natl. Acad. Sci. U.S.A. **99**, 14132 (2002).
[33] I. Ispolatov, P. L. Krapivsky, and A. Yuryev, Phys. Rev. E **71**, 061911 (2005).
[34] I. Ispolatov, P. L. Krapivsky, I. Mazo, and A. Yuryev, New J. Phys. **7**, 145 (2005).
[35] A. Vázquez, Phys. Rev. E **67**, 056104 (2003).
[36] P. L. Krapivsky and S. Redner, Phys. Rev. E **63**, 066123 (2001).
[37] T. Evans, J. Saramäki and K. Kaski, Phys. Rev. E **72**, 26138 (2005).
[38] R. T. Smythe and H. Mahmoud, Theory Probab. Math. Statist. **51**, 1 (1995).
[39] M. Drmota and B. Gittenberger, Random Struct. Alg. **10**, 421 (1997).
[40] M. Drmota and H.-K. Hwang, Adv. Appl. Probab. **37**, 321 (2005).
[41] R. Lambiotte, P. L. Krapivsky, U. Bhat, and S. Redner, unpublished.
[42] M. J. Salganik, P. S. Dodds, and D. J. Watts, Science, **311**, 854 (2006).
[43] O. Riordan and L. Warnke, Phys. Rev. E **86**, 011129 (2012).
[44] M. Ostilli, Europhys. Lett. **105**, 28005 (2014).
[45] A. A. Markov, Izv. Acad. Sci. **11**, 177 (1917).
[46] F. Eggenberger and G. Pólya, Zeit. Angew. Math. Mech. **3**, 279 (1923).
[47] H. M. Mahmoud, *Pólya Urn Models* (CRC Press, Boca Raton FL, 2009).