Highly Dispersed Networks

Alan Gabel,[1] P. L. Krapivsky,[2] and S. Redner[1]

[1] *Center for Polymer Studies and Department of Physics, Boston University, Boston, MA 02215, USA*
[2] *Department of Physics, Boston University, Boston, MA 02215, USA*

We introduce a new class of networks that grow by *enhanced redirection*. Nodes are introduced sequentially, and each either attaches to a randomly chosen target node with probability $1-r$ or to the ancestor of the target with probability r, where r an increasing function of the degree of the ancestor. This mechanism leads to highly-dispersed networks with unusual properties: (i) existence of multiple macrohubs—nodes whose degree is a finite fraction of the total number of network nodes N, (ii) lack of self averaging, and (iii) anomalous scaling, in which N_k, the number of nodes of degree k scales as $N_k \sim N^{\nu-1}/k^{\nu}$, with $1<\nu<2$.

PACS numbers: 02.50.Cw, 05.40.-a, 05.50.+q, 87.18.Sn

Many of the current models for complex networks involve growth mechanisms that are based on *global* knowledge of the network. For example, in preferential attachment [1–3], a new node attaches to an existing node of the network at a rate that is a monotonically increasing function of the degree of this target node. To implement this rule faithfully, one needs to know the degree of every node in the network, and it is impractical to maintain such detailed knowledge of the network.

A counterpoint to global growth rules is provided by a class of models that require only *local* knowledge of the network, including, for example, spatial locality[4–6] and node similarity [7]. An appealing model of this genre is redirection [8–13]. Here, each newly-introduced node chooses a target node at random and attaches to this target and/or to one or more of its ancestors. If redirection occurs only to an immediate ancestor with a fixed probability, the resulting growth rule corresponds exactly to shifted linear preferential attachment [8]. Two important features of this redirection mechanism are: (i) it precisely mimics global growth rules, such as preferential attachment, and (ii) efficiency, as the addition of each node requires just two computer instructions, so that the time needed to simulate a network of N nodes scales linearly with N.

The utility of redirection as an efficient way to mimic linear preferential attachment motivates us to exploit slightly more, but still local, information around the target node. Specifically, we consider a redirection probability $r(a,b)$ that depends on the degrees of the target and ancestor nodes, a and b respectively. In *hindered redirection*, $r(a,b)$ is a decreasing function of the ancestor degree b [14], a rule that leads to sub-linear preferential attachment network growth. In this work, we investigate the complementary situation of *enhanced redirection*, for which the redirection probability r is an increasing function of the ancestor degree b with $r \to 1$ as $b \to \infty$. This seemingly-innocuous redirection rule gives rise to networks with several intriguing and practically relevant properties (Fig. 1):

- Multiple *macrohubs*—whose degrees are a finite fraction of N—arise.

- Lack of self averaging. Different network realizations are visually diverse when the growth process starts from the same initial condition, in contrast to preferential attachment [15].

- Non-extensivity. The number of nodes of degree k, N_k, scales as $N_k \sim N^{\nu-1}k^{-\nu}$, with $1<\nu<2$, again in contrast to preferential attachment, where $N_k \sim Nk^{-\nu}$ with $\nu>2$.

Several aspects of these novel attributes bear emphasis. While macrohubs also occur in superlinear preferential attachment [8, 16, 17] and in the fitness model [18, 19], these examples give a single macrohub. In contrast, enhanced redirection networks are highly disperse, with interconnected hub-and-spoke structures that are reminiscent of airline route networks [3, 20–23]. Regarding non-extensive scaling, a degree exponent in the range $1<\nu<2$ has been observed in numerous networks [24]. Taken together with extensivity so that $N_k \sim Nk^{-\nu}$, the range $1<\nu<2$ is mathematically inconsistent. Namely, for sparse networks the average degree is finite, while $\langle k\rangle = N^{-1}\sum_{k=1}^{N} kN_k$ diverges as $N^{2-\nu}$. The simplest resolution of this paradox is to posit

$$N_k \sim N^{\nu-1}k^{-\nu} \tag{1}$$

for $k \geq 2$. The number of nodes of degree 1 (leaves) N_1 must still grow linearly with N so that the sum rule $\sum_{k=1}^{N} N_k = N$ is obeyed. More precisely, the scaling with system size is

$$N - N_1 = \mathcal{O}(N^{\nu-1}), \quad N_k = \mathcal{O}(N^{\nu-1}) \;\; k \geq 2 \tag{2}$$

In our modeling, links are directed and each node has out-degree equal to 1, and thus a unique ancestor. This growth rule produces tree networks; closed loops can be generated by allowing each new node to connected to the network in multiple ways [10]. For convenience, we choose the initial condition of a single root node of degree 2 that links to itself. The root is thus both its own ancestor and its own child. Nodes are introduced one by one. Each first picks a random target node (of degree a), and then:

(a) either the new node attaches to the target with probability $1 - r(a, b)$;

(b) or the new node attaches to the ancestor (with degree b) of the target with probability $r(a, b)$.

The unexpected connection between constant redirection probability and shifted linear preferential attachment [8, 10] arises because the number of ways to redirect to an ancestor node is proportional to the number of its descendants and thus to its degree. For enhanced redirection, two natural (but by no means unique) choices for the redirection probability are:

$$r(a, b) = 1 - b^{-\lambda}, \qquad r(a, b) = \frac{a^\lambda}{a^\lambda + b^\lambda}, \qquad \lambda > 0\,. \quad (3)$$

Our results are robust with respect to the form of the redirection probability, as long as $r(a, b) \to 1$ as $b \to \infty$; we primarily focus on the first model.

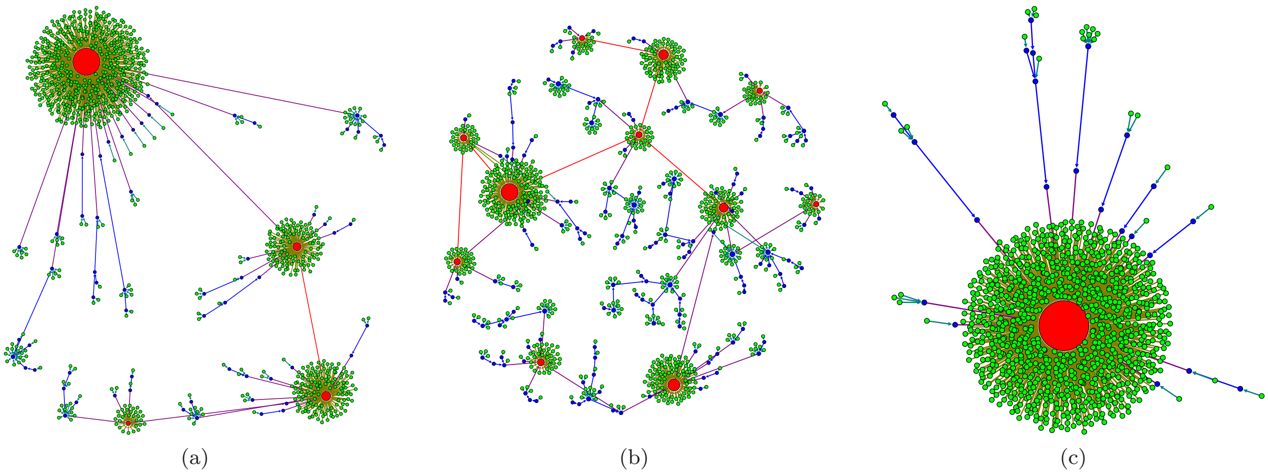


FIG. 1: Enhanced redirection networks of $N = 10^3$ nodes for $\lambda = \frac{3}{4}$ starting from the same initial state. (a) Maximum degree $k_{\max} = 548$, $\mathcal{C} = 66$ core (degree ≥ 2) nodes, and maximum diameter $D_{\max} = 10$. (b) $k_{\max} = \mathcal{C} = 154$, $D_{\max} = 12$ (the smallest $k_{\max}$ out of 10^3 realizations). (c) $k_{\max} = 963$, with $\mathcal{C} = 23$ and $D_{\max} = 6$ (the largest $k_{\max}$ out of 10^3 realizations). Green: nodes of degree 1, blue: degrees 2–20, red: degree > 20. The link color is the average of the endpoint nodes.

We now present analytical and numerical evidence for the emergence of macrohubs, the lack of self-averaging, and non-extensivity, as embodied by Eq. (2).

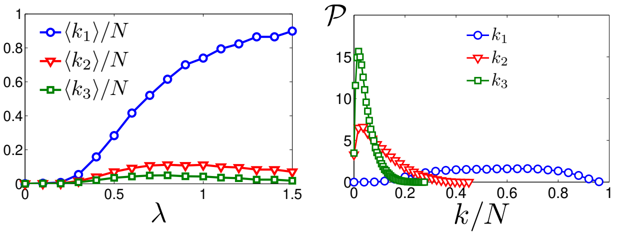
Macrohubs: Macrohubs inevitably arise in all network realizations. Figure 2(a) shows that the average largest, 2$^{\text{nd}}$-largest, and 3$^{\text{rd}}$-largest degrees are all macroscopic. These degrees, as well as the degrees of smaller hubs, are broadly distributed (Fig. 2(b)). We estimate the maximum degree $k_{\max}$ by the extremal criterion: $\int_{k_{\max}}^{\infty} N_k\, dk \sim 1$. For $N_k \sim N^{\nu-1} k^{-\nu}$ and $1 < \nu < 2$, this criterion gives $k_{\max} \sim N$. In contrast, for linear preferential attachment with $N_k \sim N k^{-3}$, $k_{\max} \sim N^{1/2}$.

The dominant role of macrohubs can be appreciated by computing the probability that the node with the highest degree attaches to every other node of the network, thereby making a star. Suppose that the network has N nodes and still remains a star. For the initial condition of a single node with a self loop, this star graph contains $N - 1$ leaves and the hub has degree $N + 1$. The probability S_N to build such a graph is

$$S_N(\lambda) = \prod_{n=1}^{N-1} \left\{ \frac{1}{n} + \frac{n-1}{n} \left[1 - (n+1)^{-\lambda} \right] \right\}. \quad (4)$$

The factor $\frac{1}{n}$ accounts for the new node attaching to the root in a network of n nodes, while the second term accounts for first choosing a leaf and then redirecting to the root. The asymptotic behavior of (4) is:

$$S_N(\lambda) \to \begin{cases} S_\infty(\lambda) & \lambda > 1 \\ A/N & \lambda = 1 \\ \exp\left(-\frac{N^{1-\lambda}}{1-\lambda}\right) & 0 < \lambda < 1 \\ \frac{1}{(N-1)!} & \lambda = 0, \end{cases} \quad (5)$$



(a) (b)

FIG. 2: (a) Average value of the three largest degrees (divided by N) as a function of λ. Each data point corresponds to 10^3 realizations. (b) Probability densities of these three largest degrees for $\lambda = \frac{3}{4}$.

where $0 < S_\infty(\lambda) < 1$, and $A = \pi^{-1}\sinh\pi \approx 3.676$. Thus a star graph occurs with positive probability when $\lambda > 1$ and $S_\infty(\lambda)$ quickly approaches 1 as λ increases (Fig. 3). This makes obvious the emergence of hubs for $\lambda > 1$. A more detailed analysis is required for $0 < \lambda \leq 1$ that also confirms the inevitability of macrohubs. We find that the probability $P(k_{\max})$ that the maximal degree is $k_{\max}$ has the scaling form [25]

$$P(k_{\max}) \to \frac{1}{N}\,\mathcal{P}(x), \quad x = \frac{k_{\max}}{N} \tag{6}$$

with $\mathcal{P}(1) = A$ for $\lambda = 1$, while for $0 < \lambda < 1$ the scaling function vanishes as $\ln\mathcal{P}(x) \sim -(1-x)^{\lambda-1}$ when $x \to 1$.

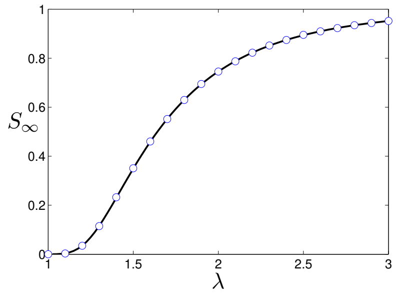
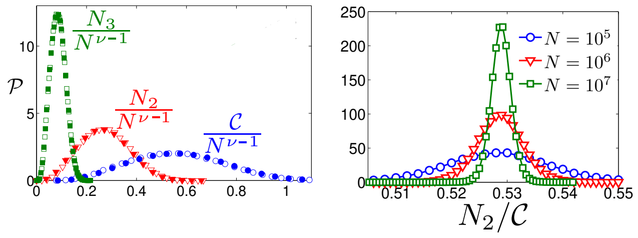


FIG. 3: Probability for a star graph, S_∞, versus λ. Data points are based on 10^4 realizations for each λ. The curve is the numerical evaluation of the product in (4).

Non Self Averaging: Enhanced redirection networks display huge sample to sample fluctuations (Fig. 1), as exemplified by (6). Another manifestation of these fluctuations is provided by the distributions for the fraction of nodes of fixed degree k, $P(N_k/N)$. For preferential attachment networks, this distribution becomes progressively sharper as N increases [19], as long as the degree is not close to its maximal value. Thus the average fractions of nodes of a given degree constitute the set of variables that characterizes the degree distribution; only the nodes with the highest degrees fail to self average [26].

In contrast, for enhanced redirection networks, essentially all geometrical features are non self-averaging. Figure 4 shows the distributions of $\mathcal{C}/N^{\nu-1}$, $N_2/N^{\nu-1}$, $N_3/N^{\nu-1}$, etc., which do not sharpen as N increases. Here $\mathcal{C} \equiv N - N_1$ is the number of non-leaf ("core") nodes. Since $\mathcal{C}$ and N_k for $k \geq 2$ all scale as $N^{\nu-1}$ (Eq. (2)), appropriately scaled distributions of these quantities would progressively sharpen as N increases if self averaging holds.

Surprisingly, the ratios $N_k/\mathcal{C}$ are self-averaging for $k \geq 2$, as the distributions $N_k/\mathcal{C}$ do sharpen as N increases (Fig. 4). The self-averaging of these ratios suggests that although the overall number of core nodes $\mathcal{C}$ varies widely between realizations, the degree distributions *given* a value of $\mathcal{C}$ are statistically the same.



(a) (b)

FIG. 4: Probability densities for enhanced redirection for: (a) $\mathcal{C}/N^{\nu-1}$, $N_2/N^{\nu-1}$, and $N_3/N^{\nu-1}$ for $N = 10^6$ (open) and $N = 10^7$ nodes (closed symbols) and (b) $N_2/\mathcal{C}$. Data are based on 10^5 realizations with $\lambda = \frac{3}{4}$ and $\nu = 1.73$.

We can understand the lack of self averaging in enhanced redirection networks in a heuristic way. Once a set of macrohubs emerges (with degrees k_1, k_2, k_3, ...) the probability of attaching to a macrohub of degree k_i asymptotically approaches k_i. This preferential attachment to macrohubs is precisely the same prescription for a multistate Pólya urn process for filling an urn with balls of several colors [27, 28], for which it is known that the long-time distribution of the number of balls of a given color is a non self-averaging quantity.

Degree Distribution: Since the degree distribution itself is non self averaging, we focus on the average over all realizations, $\langle N_k\rangle$. To avoid notational clutter we write N_k for the average $\langle N_k\rangle$. Each time a new node is introduced, the degree evolves according to

$$\frac{dN_k}{dN} = \frac{(1-f_{k-1})N_{k-1} - (1-f_k)N_k}{N} + \frac{(k-2)t_{k-1}N_{k-1} - (k-1)t_kN_k}{N} + \delta_{k,1}\,. \tag{7}$$

Here f_k and t_k are defined as the respective probabilities that an incoming link is redirected *from* a node of degree k, and *to* a node of degree k. The terms involving the factor $1 - f_j$ in (7) thus account for events where the incoming link attaches directly to the target, while the terms involving the factor t_j account for redirection to the ancestor. The term $\delta_{k,1}$ accounts for the new node of degree 1. Defining $\alpha_k = (k-1)t_k + 1 - f_k$, Eq. (7) can be written in the canonical form

$$\frac{dN_k}{dN} = \frac{\alpha_{k-1}N_{k-1} - \alpha_k N_k}{N} + \delta_{k,1}\,. \tag{8}$$

We now use the empirically-observed scaling (2), as illustrated in Fig. 5, to deduce the algebraic decay (1). We rewrite (2) more precisely as

$$N - N_1 \simeq c_1 N^{\nu-1}, \qquad N_k \simeq c_k N^{\nu-1} \quad k \geq 2\,, \tag{9}$$

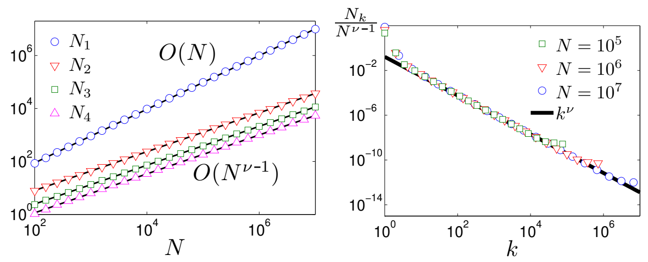
and substitute it into into the evolution equations (7). Straightforward calculation gives the product solution

$$c_k = c_1\,\frac{\nu-1}{\alpha_k}\prod_{j=2}^{k}\left(\frac{\alpha_j}{\alpha_j+\nu-1}\right). \tag{10}$$

We now need the analytic form for α_k, which requires the probabilities f_k and t_k. The latter are given by

$$f_k = \sum_{b\geq 1}\frac{r(k,b)N(k,b)}{N_k}\,, \quad t_k = \sum_{a\geq 1}\frac{r(a,k)N(a,k)}{(k-1)N_k}\,,$$

where $N(a,b)$ is the number of nodes of degree a that have an ancestor of degree b. Thus f_k is the probability of redirecting *from* a node of degree k, averaged over all such target nodes, and t_k is the probability of redirecting *to* a node of degree k, averaged over all the $(k-1)N_k$ children of nodes of degree k.



(a) (b)

FIG. 5: (a) N_k versus N and (b) $N_k/N^{\nu-1}$ versus k for enhanced redirection with $\lambda = \frac{3}{4}$ and $\nu = 1.73$ (determined numerically; Fig. 6). Data are based on 10^4 realizations, with equally-spaced bins on a logarithmic scale in (b). The lines in (a) show the prediction of Eq. (9), while the line in (b) shows the k dependence from the numerical solution of (10).

For redirection probability $r(a,b) = 1 - b^{-\lambda}$, the probabilities f_k and t_k reduce to

$$f_k = \sum_{b\geq 1}\frac{(1-b^{-\lambda})N(k,b)}{N_k} \equiv 1 - \langle b^{-\lambda}\rangle\,,$$

$$t_k = \sum_{a\geq 1}\frac{(1-k^{-\lambda})N(a,k)}{(k-1)N_k} = 1 - k^{-\lambda}\,,$$

leading to $\alpha_k = k - k^{1-\lambda} + k^{-\lambda} - f_k \to k$ in the large-k limit. Using $\alpha_k \sim k$ in the product solution (10) gives the asymptotic behavior

$$c_k \sim c_1\,\frac{\nu-1}{k}\prod_{j=2}^{k}\left(\frac{j}{j+\nu-1}\right) \sim k^{-\nu}\,. \tag{11}$$

Thus the degree distribution exhibits anomalous scaling, $N_k \sim N^{\nu-1}k^{-\nu}$, with $1 < \nu < 2$. Numerical simulations show that the exponent ν is a decreasing function of λ and that $\nu \to 1$ as $\lambda \to 2$ (Fig. 6). There is clear evidence of a transition at $\lambda = 2$; for larger λ, nodes of degree no longer appear. Thus the network consists of a collection "hairballs"—star graphs that are connected to each other by single links.

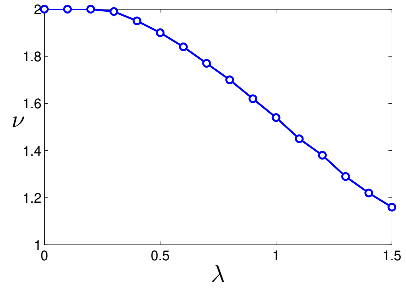


FIG. 6: Degree distribution exponent ν versus λ. Each data point is determined from fits of N_k versus N, as in Fig. 5(a).

To conclude, enhanced redirection is a simple and appealing mechanism that produces networks with several anomalous features that are observed in real networks. Among them are the existence of multiple macroscopic hubs, as arises in the airline route network [3, 20–23]. Thus our networks are highly disperse and consist of a set of loosely-connected macrohubs (Fig. 1). Also intriguing is the anomalous scaling of the degree distribution, in which the number of nodes of degree k decays more slowly than k^{-2}. Such a decay is mathematically possible in sparse networks only if the number of nodes of any degree scales sublinearly with N. Enhanced redirection may thus provide the mechanism that underlies the wide range of networks [24] whose degree distributions apparently decay more slowly than k^{-2}.

We gratefully acknowledge financial support from grant #FA9550-12-1-0391 from the U.S. Air Force Office of Scientific Research (AFOSR) and the Defense Advanced Research Projects Agency (DARPA).

[1] A. Barabási and R. Albert, Science **286**, 509 (1999).
[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, *Networks: An Introduction* (Oxford University Press, Oxford, 2010).
[4] A. Fabrikant, E. Koutsoupias, and C. H. Papadimitriou, Lecture Notes in Computer Science, **2380**, 110 (2002).
[5] V. Colizza, J. R. Banavar, A. Maritan, and A. Rinaldo, Phys. Rev. Lett. **92**, 198701 (2004).
[6] M. Barthelemy, Phys. Repts. **499** 1, (2011).
[7] F. Papadopoulos, M. Kitsak, M. Ángeles Serrano, M. Boguñá, and D. Krioukov, Nature **489**, 537 (2012).
[8] P. L. Krapivsky and S. Redner, Phys. Rev. E **63**, 066123 (2001).
[9] A. Vázquez, Phys. Rev. E **67**, 056104 (2003).
[10] H. Rozenfeld and D. ben-Avraham, Phys. Rev. E **70**, 056107 (2004).
[11] P. L. Krapivsky and S. Redner, Phys. Rev. E **71**, 036118 (2005).
[12] R. Lambiotte and M. Ausloos Europhys. Lett. **77**, 58002 (2007).
[13] E. Ben-Naim and P. L. Krapivsky, J. Stat. Mech. P06004 (2010).
[14] A. Gabel and S. Redner, J. Stat. Mech. P02043 (2013).
[15] P. L. Krapivsky and S. Redner, J. Phys. A **35**, 9517 (2002).
[16] P. L. Krapivsky, S. Redner, and F. Leyvraz, Phys. Rev. Lett. **85**, 4629 (2000); S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, Phys. Rev. Lett. **85**, 4633 (2000).
[17] P. L. Krapivsky and D. Krioukov, Phys. Rev. E **78**, 026114 (2008).
[18] G. Bianconi and A.-L. Barábasi, Phys. Rev. Lett. **86**, 5632 (2001); G. Bianconi and A.-L. Barábasi, Europhys. Lett. **54**, 436 (2001).
[19] P. L. Krapivsky and S. Redner, Comput. Netw. **39**, 261 (2002).
[20] R. F. i Cancho and R. V. Solé, *Statistical Mechanics of Complex Networks*, no. 625 in Lecture Notes in Physics, p. 114, Springer, Berlin (2003).
[21] D. L. Bryan and M. E. O'Kelly, J. Regional Science, **39**, no. 2, p. 275 (1999).
[22] J. J. Han, N. Bertain, T. Hao, D. S. Goldberg, G. F. Berriz, L. V. Zhang, D. Dupay, A. J. M. Walhout, M. E. Cusick, F. P. Roth, and M. Vidal, Nature **430**, 88 (2004).
[23] R. Guimera, S. Mossa, A. Turtschi, and L. A. N. Amaral, Proc. Natl. Acad. Sci. USA **102**, 7794 (2005).
[24] J. Kunegis, M. Blattner, and C. Moser, arXiv:1303.6271.
[25] A. Gabel, P. L. Krapivsky, and S. Redner, in preparation.
[26] P. L. Krapivsky and S. Redner, Phys. Rev. Lett. **89**, 258703 (2002).
[27] F. Eggenberger and G. Pólya, Z. Angew. Math. Mech. **3**, 279 (1923).
[28] N. Johnson and S. Kotz, *Urn Models and Their Applications: An Approach to Modern Discrete Probability Theory*, (Wiley, New York, 1977); H. M. Mahmoud, *Pólya Urn Models* (Chapman & Hall, London, 2008).