Dense Power-law Networks and Simplicial Complexes

Owen T. Courtney[1] and Ginestra Bianconi[1]

[1]*School of Mathematical Sciences, Queen Mary University of London, E1 4NS, London, UK*

There is increasing evidence that dense networks occur in on-line social networks, recommendation networks and in the brain. In addition to being dense, these networks are often also scale-free, i.e. their degree distributions follow $P(k) \propto k^{-\gamma}$ with $\gamma \in (1, 2]$. Models of growing networks have been successfully employed to produce scale-free networks using preferential attachment, however these models can only produce sparse networks as the numbers of links and nodes being added at each time-step is constant. Here we present a modelling framework which produces networks that are both dense and scale-free. The mechanism by which the networks grow in this model is based on the Pitman-Yor process. Variations on the model are able to produce undirected scale-free networks with exponent $\gamma = 2$ or directed networks with power-law out-degree distribution with tunable exponent $\gamma \in (1, 2)$. We also extend the model to that of directed 2-dimensional simplicial complexes. Simplicial complexes are generalization of networks that can encode the many body interactions between the parts of a complex system and as such are becoming increasingly popular to characterize different datasets ranging from social interacting systems to the brain. Our model produces dense directed simplicial complexes with power-law distribution of the generalized out-degrees of the nodes.

I. INTRODUCTION

Networks [1–3] are a powerful framework to characterize the structure and the function of complex systems as diverse as social networks, technological networks, molecular networks and the brain. Power-law degree distributions $P(k) = Ck^{-\gamma}$ have been identified as a universal property of complex networks [1, 4] strongly affecting their dynamical behaviour [3].

The pivotal Barabási and Albert model [4] has explained the emergence of power-law distributions in complex networks by including two simple yet fundamental elements: growth and preferential attachment. This pivotal work has triggered the formulation of several other models including the Bianconi-Barabási model [5, 6], the non-linear preferential attachment model [7] and the model with initial attractiveness of the nodes [8]. However, these models usually assume that growth only occurs through the constant addition of nodes and links implying that these growing network models can only generate *sparse scale-free networks* with an average degree that does not depend on the network size. Therefore the emergent power-law network topologies are characterized by power-law exponents $\gamma \in (2, \infty)$. A model which uses preferential attachment to produce scale-free networks and which is also capable of producing dense networks is the duplication model [9–12]. In this model a new node is introduced at each time step and attaches to a randomly selected existing node of the network, as well as to each of the existing nodes' neighbors with a probability p. For $p < \frac{1}{2}$ it has been shown that the asymptotic growth in the number of links is linear and furthermore that the networks produced are sparse scale-free networks with $\gamma \in (2, \infty)$. For $p \geq \frac{1}{2}$ the networks

are indeed dense, however the degree distribution in this regime is no longer scale-free [9–12].

While the vast majority of complex networks are described by sparse networks, there is increasing evidence that networks with a diverging average degree, also called *dense networks* often occur in on-line social networks [13, 14], recommendation networks [14] and in the brain [15].

In particular the vast majority of these networks are both dense and scale-free, i.e. they have a power-law degree distribution $P(k) \simeq k^{-\gamma}$ with power-law exponent $\gamma \in (1, 2)$. Therefore it is rather relevant to develop new theoretical frameworks for modelling these networks.

Dense network models are popular among statisticians and include the graphon [16–18] in which the average degree increases linearly with the network size N, i.e. $\langle k \rangle = O(N)$ edge exchangeable models [19] and more diluted networks [20]. However in the physics community dense networks are much less popular and have been treated only by a few authors.

The trouble for physicists with dense scale-free networks has been also enhanced after the publication of a work [21] that shows that dense scale-free networks are not graphical, i.e. for the wide majority of degree distributions, if the cutoff is not chosen carefully it is not possible to construct a network without any multiedge or any tadpole (a self-connected node, also known as a loop) that displays the desired degree distribution.

This result is certainly valid, and points to the fact that realizing dense power-law networks is more challenging than realizing sparse networks. However it would be misleading to claim from these results that dense power-law networks do not exist. In fact it is sufficient to allow for some moderate level of multiedges or soft constraints

on the degree of the nodes or to impose a structural cutoff for generating dense power-law networks.

Interestingly, the fact that dense scale-free networks do exist is demonstrated by the existence of a few modelling frameworks that extend the configuration model to dense scale-free networks [13] by imposing a suitable structural cutoff, or that generate dense power-law networks with specific values of the power-law exponents (i.e. $\gamma = 1$ [22] or $\gamma = 1.5$ [13]).

Here we propose a theoretical framework that is designed to generate dense growing power-law networks with preferential attachment without imposing any ad hoc cutoff. We have based our approach on the pivotal Pitman-Yor process [23–25] also known as the Chinese Restaurant Process. This process is originally defined for generating exchangeable partitions or, in more physical terms it is defined as a ball-in-the box process. Our original aim is to generate growing dense power-law networks using a variation of the Pitman-Yor algorithm. The first model that we have proposed is a dense undirected power-law network. In order to adapt the Pitman-Yor process to our purpose we have assumed that in the growing process multiedges are allowed giving rise to an undirected weighted network.

Interestingly we have found that this model is able to generate dense power-law networks, however these networks are only marginally dense because they display a power-law exponent $\gamma = 2$ which does not change by modifying the parameter of the model. Therefore in this respect our work shows that realizing dense scale-free networks might be difficult, supporting the results found in Ref. [21].

We have therefore generalized the model by including a direction of the links showing that in this way it is possible to give rise to denser networks. However only the out-degree of these networks is scale-free while the in-degree is an homogeneous distribution.

Finally we have expanded this model to include dense simplicial complexes formed not only by nodes and links but also by triangles. Simplicial complexes [26–36] are higher order networks that can be used for analysing collaboration networks, protein interaction networks or brain function. From the network modelling perspective they provide a clear short-cut to generate networks with high clustering coefficients.

Our theoretical framework shows that dense simplicial complexes displaying dense power-law distribution of their generalized degree can be generated using a suitable modification of the Pitman-Yor process.

The paper is structured as follows: in Sec. II we characterize the structure of networks in terms of the degrees and strengths of their nodes, and show how these concepts can be extended to simplicial complexes via the generalized degrees and generalized strengths. In Sec. III we give an overview of the Yule-Simon and Pitman-Yor processes for generating power-law distributions with exponents $\gamma \in (2, \infty)$ and $\gamma \in (1, 2]$ respectively. In Sec. IV we present a modelling framework which exploits the Pitman-Yor process in order to produce dense scale-free networks and simplicial complexes. In Sec. V we derive mean-field expressions for the total number of nodes and their strengths. We use these results in Sec. VI to find the probability that a link or triangle is reinforced in any given time-step. In Sec. VII we derive mean-field equations for the degrees, and show that the degree distributions are scale-free with dense exponent $\gamma \in (1, 2)$. In Sec. VIII we explore the relation between the strengths and degrees of the nodes. In Sec. IX we examine the clustering and degree correlations produced by the model. Finally, in Sec. X we give our conclusions.

II. NETWORKS AND SIMPLICIAL COMPLEXES

In this work we will model dense networks and simplicial complexes.

Networks describe the pairwise interactions between the elements of a complex system. This type of interaction can optionally be weighted and/or directed. An undirected network has a power-law degree distribution $P(k)$ if for large values of k,

$$P(k) \simeq Ck^{-\gamma}. \tag{1}$$

Sparse networks are networks in which $\langle k \rangle$ does not diverge with the maximum degree K of the network. Therefore sparse power-law networks have power-law exponent $\gamma \in (2, \infty)$. On the contrary dense networks are networks for which the average degree diverges with K. These networks have power-law exponents $\gamma \in (1, 2]$. For a weighted network it is also interesting to consider the strength distribution $P(s)$. The strength of a node is the sum of the weights of all of its links. Interestingly many networks such as airport networks and collaboration networks [37] are characterized by a scale-free strength distribution

$$P(s) \simeq \hat{C}s^{-\delta} \tag{2}$$

with $\delta \in (1, \infty)$.

For directed networks it is possible to distinguish between the in-degree distribution and the out-degree distribution which can either be both power-law or one power-law and the other not. Additionally we can distinguish also between the in-strength and the out-strength indicating the sum of the weights of the in-coming and out-going links respectively. The in-strength and out-strength distributions characterize globally the properties of weighted directed networks.

Simplicial complexes [28–33, 35] are a generalization of networks that are able to encode the many-body interactions between more than two nodes. A d-dimensional simplex (also indicated as d-simplex) describes the interaction among a set of $d + 1$ nodes. For instance a 0-simplex is a node, a 1-simplex is a link, a 2-simplex is a triangle and a 3-simplex is a tetrahedron. A δ-face $\tilde{\alpha}$

of a d-simplex α is a δ-simplex (with $\delta < d$) constructed from a proper subset of $\delta + 1$ nodes of the simplex α. For instance the faces of a 2-simplex are its 3 links (3 1-simplices) and its 3 nodes (3 0-simplices).

A d-dimensional simplicial complex is constructed by gluing simplices of dimension smaller or equal to d along their faces. Additionally any simplicial complex satisfies the additional condition that if a simplex belongs to the simplicial complex then all its faces must also belong to it. We say a simplicial complex is 'pure' if it is formed exclusively by d-dimensional simplices and their faces (i.e. the only simplices of dimension less than d are those which are faces of a d-dimensional simplex). A network is a pure 1-dimensional simplicial complex if it does not contain isolated nodes while a pure 2-dimensional simplicial complex is constructed from triangles and the links and nodes belonging to the triangles.

We characterize the structure of simplicial complexes in terms of the generalized degrees and generalized strengths of their faces. The generalized degree $k_{d,\delta}(\alpha)$ of a δ-dimensional simplex α is the number of simplices of dimension d which α forms a subset of. If the simplices are weighted, then we define the generalized strength $s_{d,\delta}(\alpha)$ of a δ-dimensional simplex α as the total weight of simplices of dimension d which α forms a subset of.

Here we will consider pure 2-dimensional simplicial complexes which are both weighted and directional. Considering directed simplicial complexes is rather important for describing real world systems and has been proposed recently as a framework to capture the interplay between structure and dynamics of brain networks [36]. The triangles are directed in the sense that we map a differently 'directed' or 'oriented' triangle to each permutation of its three nodes, i.e. for three nodes labelled i, j and l we can create 6 distinct directed triangles: ijl, ilj, jil, jli, lij and lji. The first node in the triangle we call the 'source' node, the second we call the 'first target node' and the third we call the 'second target node'. As with an undirected simplicial complex, the triangles contain their faces of dimension 1 (links) and 0 (nodes). In the simplicial complexes we present in this paper, these links are also directed, and their directions are determined by the direction of their parent triangle. The two links coming from the source node are directed outwards from the source node towards the two target nodes. The third link between the two target nodes is directed from the first target node towards the second target node. Figure 1 is a diagram showing the relation between the direction of a triangle and the direction of its links.

Directed simplicial complexes allow us to distinguish between the simplices a node has 'gained' from distinct attachment mechanisms through its 'generalized out-degree' and 'generalized in-degree' (see Sec. IV.C). Of course the specific relation between the direction of a triangle and the directions of its links that we use in this paper is just a convention that we have chosen, and there are indeed other conventions that could be chosen instead. We have chosen ours as it produces simplices
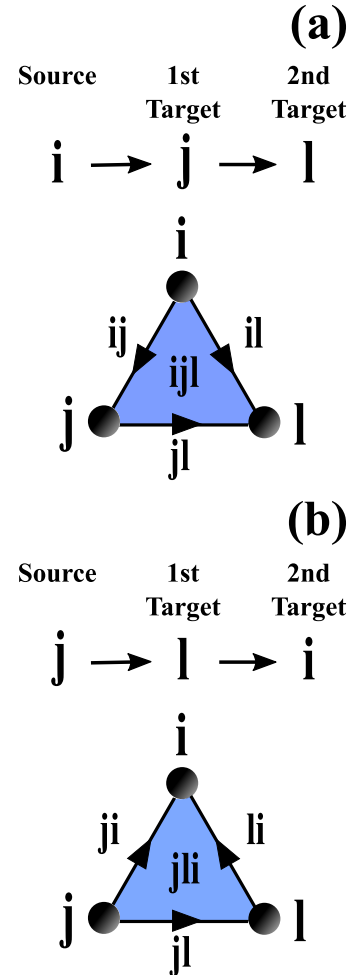

FIG. 1: (Color online). Diagram showing the relation between the direction of a triangle and the direction of its links. In panel (a) the triangle ijl has node i as its source node, node j as its first target node and node l as its second target node. There are three directed links present here: (i,j), (i,l) directed away from i towards the two target nodes and (j,l) directed from j to l. The generalized out-strength and generalized out-degree are equal to 1 for node i and 0 for nodes j and l. In panel (b) the triangle jli has node j as its source node, node l as its first target node and node i as its second target node. The three directed links here are instead (j,i), (j,l) and (l,i), while the generalized out-strength and generalized out-degree are now equal to 1 for node j and 0 for nodes i and l.

that have what could be called a 'temporal direction', where the links of the simplices produced are acyclic. Interestingly these directed simplices have been recently used [36] to analyse brain networks and coupling for topological information with the neuronal network dynamics.

III. DENSE SCALE-FREE DISTRIBUTIONS AND THE PITMAN-YOR PROCESS

The Barabási-Albert model [4] provides a fundamental mechanism for the emergence of scale-free networks with degree distribution $P(k) \sim k^{-\gamma}$ and diverging second moment $\langle k^2 \rangle$. In the Barabási-Albert model we start at time $t = 1$ from a finite network and at each time $t > 1$ we add one new node and m links connected to the new node and to a node i with degree k_i chosen with probability

$$\Pi_i^{BA} = \frac{k_i}{\sum_j k_j}. \tag{3}$$

This probability implements the preferential attachment mechanism according to which nodes which already have many links (i.e. have high degree k_i) are more likely to acquire new links. The number of nodes $N(t)$ and the number of links $L(t)$ at time t are deterministic variables growing linearly with time as we have $N(t) = t$ and $L(t) = mt$. Therefore the average degree of this network is independent on the network size and given by

$$\langle k \rangle = \frac{2L(t)}{N(t)} = 2m. \tag{4}$$

The degree distribution $P(k)$ of the Barabási-Albert model can be evaluated exactly in the large network limit $N(t) \gg 1$ and is given by [7, 8]

$$P(k) = \frac{2\Gamma(m+2)}{\Gamma(m)} \frac{\Gamma(k)}{\Gamma(k+3)} \simeq 2m(m+1)k^{-3}, \tag{5}$$

where the latter approximated expression describes the tail of the distribution where $k \gg 1$.

Other variations of the Barabási-Albert model have been shown to yield scale-free networks with tunable exponent γ. However, considering network models in which the number of nodes and the number of links both increases linearly with time can only produce networks with power-law exponent $\gamma \in (2, \infty)$. In fact power-law degree distributions $P(k) = Ck^{-\gamma}$ yield sparse network models (i.e. models in which the average degree is constant with the network size) only for $\gamma \in (2, \infty)$.

When we face the challenge of modelling dense power-law networks the question is whether it is possible to formulate alternative power-law network models in which the density of links increases in time resulting in a power-law degree distribution $P(k) \simeq Ck^{-\gamma}$ with power-law exponents $\gamma \in (1, 2]$.

In order to formulate alternative models which include growth and preferential attachment but that generate dense power-law networks we have looked closely at the mathematical literature regarding stylized ball-in-the box models.

The mathematical origin of the power-law in the Barabási-Albert model can be rooted back to a ball-in-the-box model called the Yule-Simon model [38]. This is a discrete time stochastic process analogous to placing balls in a growing number of boxes with probabilities dependent on the number of balls already in the boxes. The process starts at time $t = 1$ with a single box with one ball in it. At each subsequent time step $t > 1$ a new ball is introduced and is either placed in an existing box (with probability $\epsilon \in (0, 1)$) or a new box is created (with probability $1 - \epsilon$) and the ball is placed in it. The process may be thought of as producing a random partition of the set of balls introduced up to time t. For any given time t we indicate the total number of boxes by N, and the total number of balls by M. Additionally we indicate by s_i the number of balls in the ith box. The reinforcement dynamics called preferential attachment in network models is implemented by assuming that the probability to place a new ball in the box i grows linearly with the number of ball s_i already in the box i. Therefore in the Yule-Simon model the probability that the new ball is placed in box i is

$$\Pi_i^{YS} = \begin{cases} \epsilon s_i/t & \text{for } 1 < i \le N, \\ (1 - \epsilon) & \text{for } i = N + 1. \end{cases} \tag{6}$$

Clearly in this model the average number of boxes $\langle N(t) \rangle$ increases linearly with time, i.e. $\langle N(t) \rangle = (1-\epsilon)t$. Moreover, since at each time we add a new ball the number of balls at time t is a deterministic variable given by $M(t) = t$. It follows that the average number of balls per box is constant in time, i.e.

$$\left\langle \frac{M}{N} \right\rangle = \frac{1}{1 - \epsilon}(1 - \epsilon^t) \simeq \frac{1}{1 - \epsilon} = O(1). \tag{7}$$

This implies that if the distribution $P(s)$ of balls in the boxes decays as a power-law $P(s) \simeq s^{-\gamma}$ it must necessarily have the power-law exponent γ in the range $\gamma \in (2, \infty)$. In fact for $\gamma \in (1, 2]$ power-law distributions have a diverging average value. The exact expression of the degree distribution for the Yule-Simon process can be calculated exactly in the limit $t \to \infty$ finding

$$\begin{aligned} P(s) &= \frac{1}{\epsilon}\Gamma\left(1 + \frac{1}{\epsilon}\right)\frac{\Gamma(s)}{\Gamma(s + 1 + 1/\epsilon)} \\ &\simeq \frac{1}{\epsilon}\Gamma\left(1 + \frac{1}{\epsilon}\right)s^{-\gamma} \end{aligned} \tag{8}$$

where the last expression is derived in the limit $s \gg 1$ and where the power-law exponent γ is given by

$$\gamma = 1 + \frac{1}{\epsilon} \in (2, \infty). \tag{9}$$

Therefore the Yule-Simon model using growth and preferential attachment can generate power-law distributions $P(s) \simeq Cs^{-\gamma}$ with $\gamma > 2$ and having a finite average number of balls in the boxes. The Barabási-Albert model can be mapped to a balls-in-the-box model by assuming that each node corresponds to a box and each half-edge attached to a given node corresponds to a ball in the box. Note that for the Barabási-Albert model the number of

nodes is a deterministic variable $N(t) = t$ as is the number of half edges, which is given by twice the number of links $M(t) = 2L(t) \simeq 2mt$. However the Barabási-Albert model can be considered as being in the same *universality class* as the Yule-Simon process with $\epsilon = 1/2$.

Interestingly, a different ball-in-the box model called the Pitman-Yor process [23–25] is known to yield dense power-law distributions. This model includes growth of the number of boxes, and reinforcement dynamics (preferential attachment) but enforces that the scaling of the number of balls with the number of boxes is superlinear. In this way this model elegantly generates dense power-law distributions with exponent $\gamma \in (1, 2]$. Starting at time $t = 1$ with one ball in a single box at each time $t > 1$ a new ball is added and either placed in an existing box i or placed in a new box $i = N(t) + 1$. Specifically the probability Π_i^{PY} that the new ball goes in the box i is parametrized by the parameter $\alpha \in (0, 1)$ and given by

$$\Pi_i^{PY} = \begin{cases} \frac{s_i(t) - \alpha}{t} & \text{for } 1 < i \le N, \\ \frac{\alpha N}{t} & \text{for } i = N + 1. \end{cases} \quad (10)$$

As in the Yule-Simon process there is a growth in both the number of balls, and the number of boxes, with a preferential attachment mechanism for placing the balls. Unlike with Yule-Simon however, the probability of adding a new box is not constant, but depends on the number of boxes already added while decaying with the total number of balls already added [23, 25]. The marginal distribution of a single box in the large t limit is [23]

$$P(s) = \frac{\alpha}{\Gamma(1 - \alpha)} \frac{\Gamma(s - \alpha)}{\Gamma(s + 1)} \simeq \frac{\alpha}{\Gamma(1 - \alpha)} s^{-\gamma}, \quad (11)$$

where the latter expression is an approximation for the tail of the distribution (i.e. for $s \gg 1$). Here the power-law exponent γ is given by

$$\gamma = 1 + \alpha \in (1, 2], \quad (12)$$

which implies that the distribution has a diverging first moment. This is consistent with the fact that the number of balls increases superlinearly with the expected number of boxes $\langle N(t) \rangle$ [23] as we have

$$\begin{aligned} \langle N(t) \rangle &\simeq t^\alpha \\ M(t) &= t, \end{aligned} \quad (13)$$

and therefore

$$\left\langle \frac{M}{N} \right\rangle = O(t^{1 - \alpha}). \quad (14)$$

In this paper we explore whether the Pitman-Yor process can be exploited in order to formulate dense power-law network models, and we emphasize the challenges posed by the density of the resulting networks. In this endeavor our objective is to construct not only dense power-law networks formed by pairwise interactions but also dense simplicial complexes which allow one to go beyond the framework of pairwise interactions.

IV. EVOLUTION OF DENSE SCALE-FREE NETWORKS AND SIMPLICIAL COMPLEXES

Here we introduce a modelling framework that exploits the Pitman-Yor process in order to generate dense weighted power-law networks and simplicial complexes.

Weighted dense power-law networks evolve by the subsequent addition of nodes and the establishment of new links or reinforcement of already existing links. We consider both a version of the model where the links are directed and a version with undirected links.

This modelling framework is then extended to 2-dimensional simplicial complexes, where the growth comes from the addition of nodes, links and triangles. These simplicial complexes have densely connected skeletons and power-law distributions of the generalized strengths and generalized degrees.

Unlike in many other models of growing networks, the number of nodes in the network at a given time is not a deterministic function of time but instead depends on the stochastic growth dynamics of the network. The relative probabilities of nodes being created or selected for reinforcement are analogous to a Pitman-Yor process, with an equivalent parameter $\alpha \in (0, 1)$. Below we give the dynamics for each of the three versions of the model.

A. Undirected network growth dynamics

In the undirected network version of the model, we write the total number of nodes in the network at time t as $N(t)$. Every pair of nodes $i, j \in \{1, 2, ..., N(t)\}$ has an associated weight $w_{ij}(t)$ taking non-negative integer values.

We start at $t = 1$ with an undirected link between node 1 and node 2. At each time step $t \ge 1$ we select a node i with probability

$$\Pi_i^U = \begin{cases} \frac{s_i(t) - \alpha}{2t} & \text{for } 1 \le i \le N \\ \frac{\alpha N}{2t} & \text{for } i = N + 1. \end{cases} \quad (15)$$

We then update the value of N and select a second node j using the same algorithm. If the two selected nodes are not already linked we add a link between them, if they are already linked we reinforce the weight of the links. In other words the adjacency matrix element a_{ij} and the weight w_{ij} of the link (i, j) are updated according to

$$\begin{aligned} a_{ij}(t + 1) &= 1, \\ w_{ij}(t + 1) &= w_{ij}(t) + 1. \end{aligned} \quad (16)$$

Moreover, since the network is undirected we have

$$\begin{aligned} a_{ji}(t) &= a_{ij}(t), \\ w_{ji}(t) &= w_{ij}(t). \end{aligned} \quad (17)$$

Therefore in this model we treat half-edges as the balls of the Pitman-Yor process and we treat the nodes as the boxes of the Pitman-Yor process. Therefore it is to

be expected that the strength distribution will follow a power-law exponent with exponent $\gamma = 1 + \alpha$. However, given that the network is weighted, the degree distribution could potentially be significantly different because if a new link is placed between nodes that are already connected the strengths of these nodes will increase but their degrees will remain unchanged.

B. Directed network growth dynamics

The directed network growth dynamics assumes that links are directed and that only the source node of the links is chosen according to the Pitman-Yor reinforcement dynamics, while the target node is chosen uniformly at random among all the existing nodes of the network. In this way we expect that the density of links in the network will grow more rapidly than in the undirected case. In fact by choosing the target node with uniform probability we are more likely to add new links because we are not biasing the target node to be a node of high degree.

In the directed version, we start at $t = 1$ with a directed link from node 2 to node 1.

At each time step $t \geq 1$ a pair of nodes is selected and its weight is incremented by one. The source node is either an existing node or a new node, while the target node is chosen uniformly at random from the remaining existing nodes. The probability at time t of selecting node i as the source node is given by:

$$\Pi_i^D = \begin{cases} \frac{s_i^{out}(t) - \alpha}{t} & \text{for } 1 < i \leq N \\ \frac{\alpha(N-1)}{t} & \text{for } i = N+1. \end{cases} \quad (18)$$

The probability at time t of selecting node j as the target node is given by:

$$\hat{\Pi}_j^N = \frac{1}{N(t)}. \quad (19)$$

When both source node i and target node j have been selected we update the adjacency matrix element and weight of the link, i.e.

$$\begin{aligned} a_{ij}(t+1) &= 1, \\ w_{ij}(t+1) &= w_{ij}(t) + 1. \end{aligned} \quad (20)$$

Note that here we have chosen to select the source node of the link according to the Pitman-Yor dynamics while the target node is chosen with uniform probability. However it is also possible to consider a directed network model in which the target node is chosen according to the Pitman-Yor dynamics and the source node is chosen uniformly at random. Since the two versions of the model are simply related by the inversion of the direction of the links here we omit the explicit treatment of the latter possible definition.

C. Directed simplicial complex growth dynamics

Here we consider a directed 2-dimensional simplicial complex formed only by "directed triangles". The triangles are directed in the sense that each permutation of three nodes is associated with a different triangle. We say that the first node in the triangle is the "source node", the second node is the "first target node" and the third node is the "second target node". For the triangles in this version of the model the links are also directed, and we have chosen the convention that the two links coming from the source node are directed away from the source node towards the target nodes and the third link is directed from the first target node to the second. In this version of the model the triangles ijl also have an associated weight $w_{ijl}(t)$ taking non-negative integer values. These weights are associated specifically to the directed triangles and thus are also directed in the same sense. We define the *generalized out-strength* $\tilde{s}_i^{out}$ of a node i to be the total weight of triangles for which i is the source node, i.e. $\tilde{s}_i^{out} = \sum_{j,l=1}^{N} w_{ijl}$.

In this version of the model we start at $t = 1$ with three nodes labelled $1, 2, 3$ and the single directed triangle 123. At each time step $t \geq 1$ we select a triangle to be created or reinforced. The source node i of this triangle is selected with probability

$$\tilde{\Pi}_i^{SC} = \begin{cases} \frac{\tilde{s}_i^{out}(t) - \alpha}{t} & \text{for } 1 \leq i \leq N, \\ \frac{\alpha(N-2)}{t} & \text{for } i = N+1. \end{cases} \quad (21)$$

Once this source node i has been selected, a link (j, l) is selected uniformly at random from the set of existing links with probability

$$\hat{\Pi}_{jl}^L = \frac{1}{L(t)}, \quad (22)$$

where $L(t)$ is the number of links in the simplicial complex at time t. If this triangle already exists then its weight is reinforced according to

$$w_{ijl}(t+1) = w_{ijl}(t) + 1, \quad (23)$$

while if it doesn't exist yet then it is created with initial weight one:

$$\begin{aligned} a_{ijl}(t+1) &= 1, \\ w_{ijl}(t+1) &= 1. \end{aligned} \quad (24)$$

Figure 2 is a diagram illustrating one possible way the simplicial complex could grow in its first three time-steps. In this example we start with a single triangle at $t = 1$. At $t = 2$ a new node labelled 4 is added and forms a triangle with the link $(1, 2)$. At time $t = 3$ no new nodes are added, but instead node 4 gains an additional triangle formed with the link $(1, 3)$.
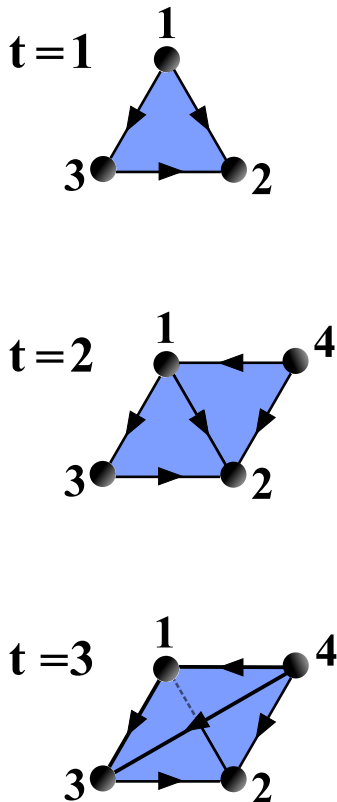
FIG. 2: (Color online). Diagram showing one possible way the simplicial complex could grow in the first three time-steps. At $t = 1$ all simplicial complexes in the model start as the single triangle 123. In this example, at $t = 2$ a new node (4) is created and randomly selects the link $(1, 2)$ to form the triangle 412. At $t = 3$ no new node is created. Instead, node 4 is selected for reinforcement and randomly selects the link $(1, 3)$ to form the triangle 413.

D. Number of links as a function of the number of nodes

In all three versions of the model, the distributions of the strengths (or out-strengths) of the nodes are generated by a Pitman-Yor process and so therefore have 'dense' power-law exponents $\delta \in (1, 2)$. However, as mentioned before, this does not guarantee that the networks themselves are dense as the links can be weighted multiple times. Therefore we have run extensive simulations of the three versions of the model to investigate whether the total number of links grows super-linearly with the number of nodes. Figure 3 shows how the number of links grows with the total number of nodes for a range of values of parameter α and for the three different versions of the model. We see that for all three versions the total number of links grows faster than the total number of nodes, indicating that the model produces dense networks and simplicial complexes. Moreover, as expected, the directed version of the model allows for the exploration of cases in which the ratio between the number of links and the number of nodes increases more rapidly
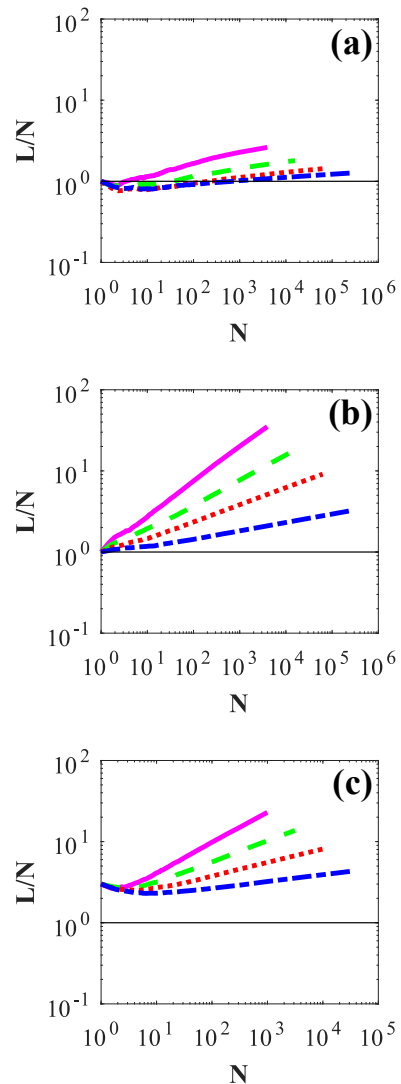

FIG. 3: (Color online). Total number of links over the total number of nodes, as a function of the total number of nodes, taken from simulation data. Panels (a), (b) and (c) show the results for the undirected network, directed network and simplicial complex respectively. For each model version and each choice of parameter α, 50 realizations of the network were generated and averaged over. The results were obtained for $t = 10^6$ and for $\alpha = 0.6$, (purple solid line), $\alpha = 0.7$ (green dashed line), $\alpha = 0.8$ (red dotted line), and $\alpha = 0.9$ (blue dot-dashed line).

than for the undirected version of the model.

V. STRENGTHS OF THE NODES

The mean-field approximation is known to give very reliable results in the context of sparse growing network models. Therefore it is natural to approach the study of dense growing networks and simplicial complexes with the same techniques. Specifically here our goal is to de-

rive the distribution of the strength s in the undirected network model, the distribution of the out-strength s^{out} in the directed network model and the distribution of the generalized out-strength $\tilde{s}^{out}$ for the simplicial complex model using the mean-field approximation.

A. Evolution of the number of nodes and of the strengths

The mean-field differential equations for the three cases differ only trivially, and therefore we will treat them using a unified set of equations that apply to all three cases. To this end we use the symbol $\hat{s}_i$ that indicates $s_i, s_i^{out}, \tilde{s}_i^{out}$ for the undirected network, directed network and directed simplicial complex versions of the model respectively. Similarly the Pitman-Yor probabilities Π_i^U, Π_i^D and Π_i^{SC} can be unified in a single expression

$$\Pi_i = \begin{cases} \frac{\hat{s}_i(t)-\alpha}{(2-a)t} & \text{for } 1 \leq i \leq N \\ \frac{\alpha(N-a-b)}{(2-a)t} & \text{for } i = N+1. \end{cases} \quad (25)$$

where we have introduced the parameters a and b taking values: $a = b = 0$ for the undirected network case; $a = 1, b = 0$ for the directed network case and $a = 1, b = 1$ for the directed simplicial complex case. Eq. (25) thus subsumes Eq.s (15), (18) and (21) for the probabilities of a node i being selected (or created) at time $t+1$ in the undirected, directed and simplicial complex versions of the model respectively.

As usual in the mean-field approximation, we will treat our stochastic variables $N(t)$, $\hat{s}_i(t)$ as deterministic continuous variables equal to their expected value over different realizations of the network or simplicial complex evolution. Since at each time the number of nodes to be chosen according to the Pitman-Yor probability is $(2-a)$ the mean-field equation determining the growth of the number of nodes in the network is given by

$$\frac{dN}{dt} = (2-a)\Pi_{N+1} = \frac{\alpha(N-a-b)}{t}, \quad (26)$$

with the initial condition

$$N(t=1) = 2 + b. \quad (27)$$

The solution is then

$$N(t) = (2-a)t^\alpha + a + b. \quad (28)$$

The differential equations for $\hat{s}_i$ of a node i born at time $t_i > 1$ is given by

$$\frac{d\hat{s}_i}{dt} = (2-a)\Pi_i = \frac{\hat{s}_i - \alpha}{t}, \quad (29)$$

with initial condition

$$\hat{s}_i(t_i) = 1. \quad (30)$$

Therefore $\hat{s}_i$ increases linearly with time, and is given by

$$\hat{s}_i = (1-\alpha)\left(\frac{t}{t_i}\right) + \alpha. \quad (31)$$

B. Strength distribution

The strength distribution can be easily derived in the mean-field approximation by using the mean-field expressions for the number of nodes $N(t)$ (Eq. 28) and the strength $\hat{s}(t)$ (Eq. (31)) as a function of time t.

To this end by using Eq. (31) we first note that the cumulative strength distribution $P(\hat{s}_i > \hat{s})$ indicating the probability that a random node i has a strength $\hat{s}_i(t) > \hat{s}$ can be written as

$$P(\hat{s}_i \geq \hat{s}) = P(t_i \leq t^\star(\hat{s})), \quad (32)$$

where $P(t_i \leq t^\star(\hat{s}))$ is the probability that a random node i arrives in the network at time $t_i \leq t^\star(\hat{s})$ and where $t^\star(\hat{s})$ satisfies

$$\hat{s} = (1-\alpha)\left(\frac{t}{t^\star}\right) + \alpha. \quad (33)$$

Moreover we observe that the probability $P(t_i \leq t^\star(\hat{s}))$ is simply given by the fraction of nodes arrived in the network before time $t^\star(\hat{s})$, i.e.

$$P(t_i \leq t^\star(\hat{s})) = \frac{N(t^\star(\hat{s})) - a - b}{N(t) - a - b}. \quad (34)$$

The strength distribution $\tilde{P}(\hat{s})$ is thus given by

$$\begin{aligned} \tilde{P}(\hat{s}) &= \frac{d}{d\hat{s}}\left[1 - P(\hat{s}_i \geq \hat{s})\right] \\ &\simeq \frac{\alpha}{1-\alpha}\left(\frac{1-\alpha}{\hat{s}}\right)^{\alpha+1}, \end{aligned} \quad (35)$$

where the last expression is valid for $\hat{s} \gg 1$. Therefore the strength distribution is power-law distributed with exponent $1 + \alpha \in (1, 2]$. Figure 4 shows the strength distributions arising from simulations of the three models. We see that in all three versions of the model, and for all values of α used, that the strength distributions follow a power-law. In the insets of each panel we see that the exponents fitted to the distributions are very close to $1 + \alpha$ as predicted by Eq. (35).

VI. REINFORCEMENT PROBABILITIES

By reinforcement probability we mean the probability that at time t we either add a new link (or a new triangle) or in the case where the link (or triangle) already exists that we increase its weight. This probability can be directly calculated in the mean-field approximation using Eqs. (28) and (31).

In the undirected case we indicate with $\pi_{ij}(t, t_i, t_j)$ the probability that at time t we reinforce or add a link between node i and node j given that node i and node j
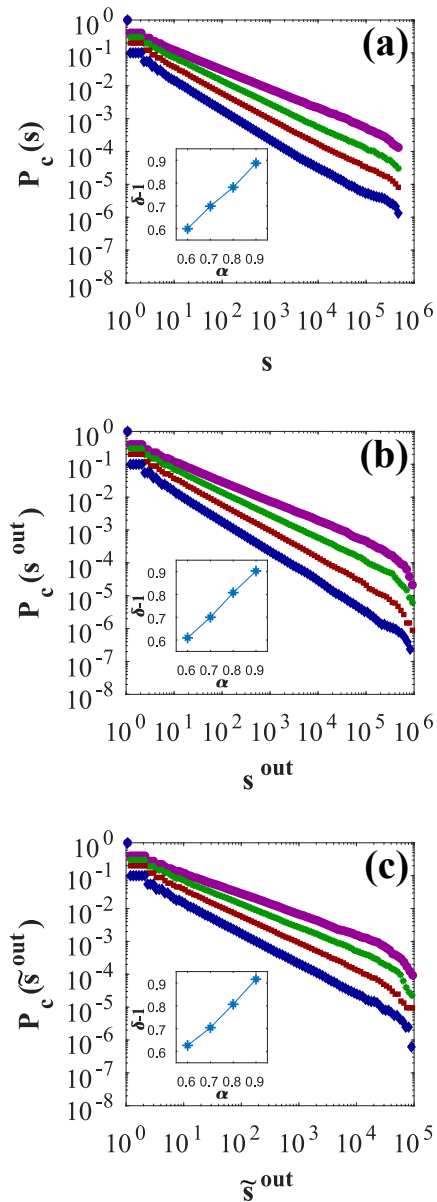
have been added to the network at time t_i and t_j respectively. The reinforcement probability is therefore given by

$$\pi_{ij}^U(t,t_i,t_j) = 2\Pi_i^U \Pi_j^U \qquad (36)$$

where Π_i^U, Π_j^U are calculated at time t. Therefore we have

$$\pi_{ij}^U(t,t_i,t_j) = \frac{(s_i(t)-\alpha)(s_j(t)-\alpha)}{2t^2}, \qquad (37)$$

which by inserting the mean-field expression for $s_i(t)$ gives

$$\pi_{ij}^U(t,t_i,t_j) = \frac{(1-\alpha)^2}{2t_it_j}. \qquad (38)$$

Similarly it can be shown that in the directed case the probability $\pi_{ij}^D(t,t_i,t_j)$ that a link from node i to node j is reinforced at time t given that nodes i iand j are arrived in the network at time t_i and t_j respectively can be expressed as

$$\pi_{ij}^D(t,t_i,t_j) = \frac{1}{N(t)}\frac{s_i^{out}(t)-\alpha}{t}, \qquad (39)$$

which using the mean-field solution of s_i^{out} and $N(t)$ gives

$$\pi_{ij}^D(t,t_i,t_j) = \frac{1-\alpha}{(t^\alpha-1)t_i}. \qquad (40)$$

For the simplicial complex we indicate with $\tilde{\pi}_{i,\ell}(t,t_i,\tau_\ell)$ the probability that a triangle with source node i and target link ℓ is reinforced, conditioned on their respective birth times t_i and τ_ℓ. We write this as

$$\tilde{\pi}_{i,\ell}(t,t_i,\tau_\ell) = \frac{\tilde{s}_i^{out}-\alpha}{t}\frac{1}{L(t)}, \qquad (41)$$

where $L(t)$ is the total number of (directed) links at time t. Using Eq. (31) we obtain

$$\tilde{\pi}_{i,\ell}(t,t_i,\tau_\ell) = \frac{1-\alpha}{t_i}\frac{1}{L(t)}. \qquad (42)$$

VII. DEGREE DISTRIBUTION

In this section we use the mean-field results of sections V and VI to derive equations for the degrees of the nodes conditioned on their birth times. We evaluate these equations numerically for a range of values of the parameter α and find power-law scalings with $\gamma = 2$ for k_i in the undirected case, and $\gamma < 2$ for k_i^{out} and $\tilde{k}_i^{out}$ in the directed and simplicial complex cases. We also compare our numerically obtained predictions with simulation results, validating our mean-field approach.

We observe that for the studied model the mean-field solution provides a very good approximation to the simulation results. Although we do not have a quantitative

FIG. 4: (Color online). Cumulative strength, out-strength and generalized out-strength distributions for the three versions of the model. For each version of the model and for each choice of parameter α, 50 realizations of the network were generated and averaged over. The simulated results were obtained for $t = 10^6$ and are represented by purple circles ($\alpha = 0.6$), green stars ($\alpha = 0.7$), red squares ($\alpha = 0.8$) and blue diamonds ($\alpha = 0.9$). The insets show the fitted exponents of the distributions for the four values of α.

theoretical argument to show the efficacy of the mean-field approximation the results is not completely surprising. In fact already for the BA model and the Simon model the mean-field approximation works very well due to the presence of the reinforcement dynamics. In the Pitman-Yor process and in the proposed dense network model the balance between the reinforcement move and the "innovation" move (placing a new ball in a new box or attaching a new link to a new node) is further shifted in favor of the reinforcement dynamics. This feature therefore could potentially justify the efficacy of the mean-field approximation.

A. Undirected network case

We write the degree of a node i born at time t_i in terms of the link probabilities $p_{ij}(t, t_i, t_j)$:

$$k_i(t, t_i) \; = \; \int_1^t dt_j \dot{N}(t_j) p_{ij}(t, t_i, t_j). \qquad (43)$$

The link probability $p_{ij}(t, t_i, t_j)$ is the probability that a link exists between nodes i and j conditioned on their birth times t_i and t_j, and may be written as

$$p_{ij}(t, t_i, t_j) \; = \; 1 - \prod_{t'=\tau}^{t} \left[1 - \pi_{ij}(t', t_i, t_j) \right], \quad (44)$$

where $\pi_{ij}(t', t_i, t_j)$ is the reinforcement probability given in (38) and $\tau = \max\{t_i, t_j\}$ is the first time that both i and j are present in the network, i.e. (44) is 1 minus the probability that the pair (i, j) is not reinforced in the time interval $[\tau, t]$. The π_{ij} reinforcement probabilities are very small for almost all pairs of nodes, so we make the approximation

$$\prod_{t'=\tau}^{t} \left[1 - \pi_{ij}(t', t_i, t_j) \right] \simeq \exp\left(-\sum_{t'=\tau}^{t} \pi_{ij}(t', t_i, t_j) \right). (45)$$

Taking t to be very large, we approximate the sum with an integral, and write (44) as

$$p_{ij}(t, t_i, t_j) \simeq 1 - \exp\left(-\int_\tau^t dt' \; \pi_{ij}(t', t_i, t_j) \right). \quad (46)$$

It is then straight-forward to obtain the following expression for the degree from Eq.s (46) and (38):

$$k_i(t, t_i) \; \simeq \; \alpha \int_1^{t_i} dt_j t_j^{\alpha-1} \left[1 - e^{-\frac{(1-\alpha)^2}{2t_i t_j}(t-t_i)} \right]$$
$$+ \alpha \int_{t_i}^{t} dt_j t_j^{\alpha-1} \left[1 - e^{-\frac{(1-\alpha)^2}{2t_i t_j}(t-t_j)} \right]. (47)$$

Eq. (47) may be written as

$$k_i(t, t_i) \; \simeq \; \alpha A^\alpha (t - t_i)^\alpha \int_{A(\frac{t}{t_i}-1)}^{A(t-t_i)} dx x^{-(1+\alpha)} \left[1 - e^{-x} \right] (48)$$

$$+ \alpha A^\alpha t^\alpha \int_0^{A(\frac{t}{t_i}-1)} dx (A+x)^{-(1+\alpha)} \left[1 - e^{-x} \right] (49)$$

where $A = \frac{(1-\alpha)^2}{2t_i}$ and in the first integral we have used the change in variable $x = A(t-t_i)t_j^{-1}$ while in the second integral we have used $x = A(\frac{t}{t_j} - 1)$. Taking $B = (1-\alpha)^2/2 \ll 1$ this equation can be approximated by the scaling function

$$k_i(t, t_i) \; \simeq \; G(t, t/t_i) \qquad (50)$$

with

$$G(t, t/t_i = y) \; = \; \alpha B^\alpha (y-1)^\alpha \int_{By(y-1)/t}^{B(y-1)} dx x^{-(1+\alpha)} \left[1 - e^{-x} \right]$$
$$+ \alpha B^\alpha y^\alpha \int_0^{By(y-1)/t} dx x^{-(1+\alpha)} \left[1 - e^{-x} \right].$$

Finally by taking the limit $t, t_i \to \infty$ with t/t_i fixed, it is possible to show that $k_i(t, t_i)$ becomes exclusively a function of t/t_i

$$k_i(t, t_i) = \tilde{G}(t/t_i) \qquad (51)$$

where for $t/t_i \gg 1$

$$\tilde{G}(t/t_i) \propto \left(\frac{t}{t_i} \right)^\alpha. \qquad (52)$$

Therefore the tail of the cumulative degree distribution $P(k_i > k)$ can be obtained within the mean-field approximation using

$$P(k_i > k) = \frac{N\left(t^\star(k)\right)}{t^\alpha} = \left(\frac{t^\star(k)}{t} \right)^\alpha, \qquad (53)$$

where $t^\star(k)$ is the birth time such that

$$k_i\left(t, t^\star(k)\right) = k. \qquad (54)$$

From Eq.s (52) and (54) we obtain the scaling

$$P(k_i > k) \propto k^{-1}, \qquad (55)$$

indicating that for large t, the degree distribution $P(k)$ has a power-law tail with exponent $\gamma = 2$. We confirm these results by comparing them to simulations of the process for a range of values of α. Figure 5 shows the average cumulative degree distributions given by the simulations. Also included are theoretical predictions obtained by evaluating Eq. (47) numerically. The inset plot shows the values of γ obtained from fitting power-laws to the tails of the simulation data. We see that for all the degrees of the nodes follow power-laws with values of γ close to the theoretical prediction of 2.

Additionally we have studied the scaling of the maximum degree (cutoff) k_{max} as a function of the network size N. This study reveals that the cutoff scales with the network size N with a proportionality constant depending on the value of α (see Figure 6).
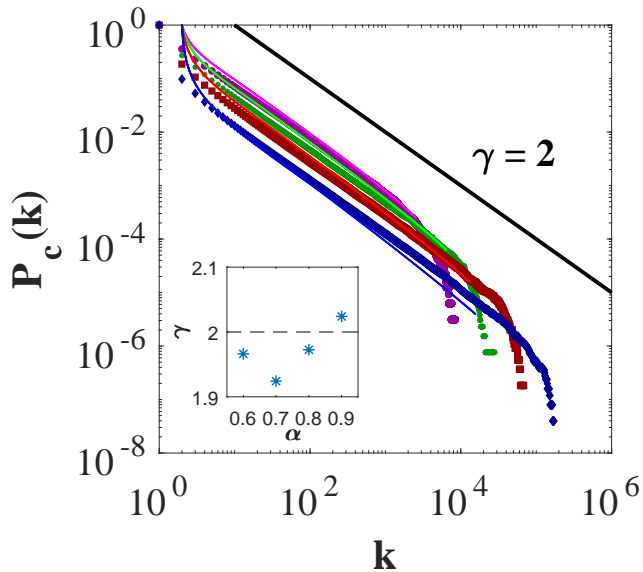
FIG. 5: (Color online). Cumulative degree distributions in the undirected case. For each choice of parameter α, 50 realizations of the network were generated and averaged over. The simulated results were obtained for $t = 10^6$ and are represented by purple circles ($\alpha = 0.6$), green stars ($\alpha = 0.7$), red squares ($\alpha = 0.8$) and blue diamonds ($\alpha = 0.9$). The numerical results are represented by the purple solid line ($\alpha = 0.6$), green dashed line ($\alpha = 0.7$), red dotted line ($\alpha = 0.8$) and blue dot-dashed line ($\alpha = 0.9$). The solid black line shows an exact power-law with exponent $\gamma = 2$ for comparison. The inset shows the fitted exponents of the simulated distributions for the four values of α.

B. Directed network case

The out-degree $k_i^{out}(t, t_i)$ at time t of node i born at time t_i in the directed case is

$$k_i^{out}(t, t_i) \simeq \int_1^t dt_j \dot{N}(t_j) p_{ij}(t, t_i, t_j), \qquad (56)$$

where $p_{ij}(t, t_i, t_j)$ is as in Eq. (46) but with $\pi_{ij}(t', t_i, t_j)$ given by Eq. (40). Let us use the direct evaluation of the integral

$$\int_\tau^t dt' \, \pi_{ij}(t', t_i, t_j) = \frac{1}{t_i}[t^{1-\alpha} - \tau^{1-\alpha}]$$

$$= t^{-\alpha}\frac{t}{t_i}\left[1 - \left(\frac{\tau}{t}\right)^{1-\alpha}\right] \quad (57)$$

to express the out-degree $k_i^{out}(t)$ of a node i with arrival time t_i at time t (56) as

$$k_i^{out}(t) = N(t_i)p(t, t_i, t_i) + \int_{t_i}^t d\tau \frac{\alpha N(\tau)}{\tau}p(t, t_i, \tau). \qquad (58)$$
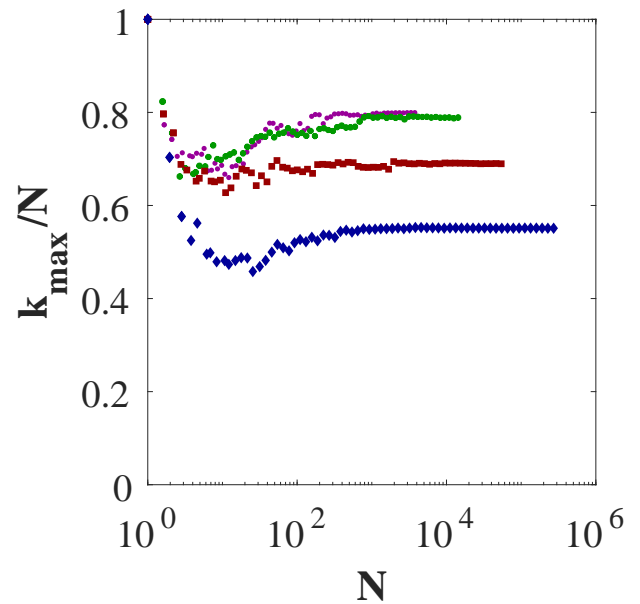
FIG. 6: (Color online). Evolution of the average normalized degree cut-off for the undirected case as a function of average total number of nodes in the network. For each choice of parameter α, 50 realizations of the network were generated and averaged over. The simulated results were obtained for $t = 10^6$ and are represented by purple circles ($\alpha = 0.6$), green stars ($\alpha = 0.7$), red squares ($\alpha = 0.8$) and blue diamonds ($\alpha = 0.9$).

Moreover integrating by parts we find

$$\frac{k_i^{out}(t)}{t^\alpha} = 1 - \left(\frac{t_i}{t}\right)^\alpha e^{-\tilde{A}(t, t_i/t)\left[1 - \left(\frac{t_i}{t}\right)^{1-\alpha}\right]}$$

$$- \alpha \int_{t_i/t}^1 dx x^{\alpha-1} e^{-\tilde{A}(t, t/t_i)[1 - x^{1-\alpha}]}$$

$$= (1-\alpha)\tilde{A}(t, t_i/t) \int_{t_i/t}^1 dx e^{-\tilde{A}(t, t/t_i)[1 - x^{1-\alpha}]} \qquad (59)$$

where

$$\tilde{A}(t, t/t_i) = t^{-\alpha}\left(\frac{t}{t_i}\right). \qquad (60)$$

Therefore we find that as $t, t_i \to \infty$ with t/t_i fixed,

$$k_i^{out}(t) \simeq (1-\alpha)\left[\frac{t}{t_i} - 1\right] \qquad (61)$$

Therefore the cumulative degree distribution may be found from

$$P(k_i(t) > k) = P(t_i < t^\star(k))) \qquad (62)$$

with

$$t^\star(k) = t\frac{(1-\alpha)}{(k + 1 - \alpha)} \qquad (63)$$

and

$$P(t_i < t^\star(k))) = \frac{N(t^\star(k))}{N(t)}. \tag{64}$$

Using the mean-field expression for the number of nodes given by Eq. (28) we find the cumulative out-degree distribution to be

$$P(k_i > k) = \left(\frac{(1-\alpha)}{(k+1)}\right)^\alpha, \tag{65}$$

which implies the out-degree distribution is

$$P(k) = \alpha\frac{(1-\alpha)^\alpha}{(k+1)^{\alpha+1}} \simeq k^{-\alpha-1}. \tag{66}$$

Therefore, we see that within our mean-field approximation the distribution of out-degrees has a power-law tail with exponent $\gamma = 1 + \alpha$. Figure 7 shows theoretical predictions for the full cumulative out-degree distribution, obtained from numerical evaluation of Eq. (56) for a selection of values of α. Also included in the figure are the results of simulations for the same values of α. The inset plot shows the values of γ obtained from fitting power-laws to the tails of the simulation data. We see that the out-degrees of the nodes follow power-laws with increasing values of γ for larger α, and with all values of γ between $\gamma = 1$ and $\gamma = 2$. From the inset plot it is clear that the exponents of tails of the distributions closely agree with the theoretical prediction of $\gamma = 1 + \alpha$.

C. Directed simplicial complex case

In the case of the simplicial complex, the generalized out-degree $\tilde{k}_i^{out}(t, t_i)$ of a node i is the number of triangles for which i is the source node. In the mean-field approximation we may write this as

$$\tilde{k}_i^{out}(t, t_i) \simeq \int_1^t d\tau_\ell \dot{L}(\tau_\ell)\hat{p}(t, t_i, \tau_\ell) \tag{67}$$

where

$$\hat{p}(t, t_i, \tau_\ell) = 1 - e^{-(1-\alpha)/t_i \int_{\max(t_i,\tau_\ell)}^t dt'[L(t')]^{-1}} \tag{68}$$

is the probability of a triangle with source node i and target link l. From figure 3 its clear that for large t, $L(t)$ grows like a power of t. We therefore assume

$$L(t) = ct^b \tag{69}$$

and obtain values c and b for each choice of α by fitting Eq. (69) to the data shown in figure 3(c). Substituting Eq. (69) in to (68) we obtain the following for the generalized out-degree of a node i born at time t_i,

$$\tilde{k}_i^{out}(t, t_i) = \\ cb \int_{t_i}^t d\tau_\ell \tau_\ell^{b-1} \left[1 - e^{-\frac{1-\alpha}{c(1-b)t_i}\left(t^{1-b} - \max(t_i,\tau_\ell)^{1-b}\right)}\right]. \tag{70}$$
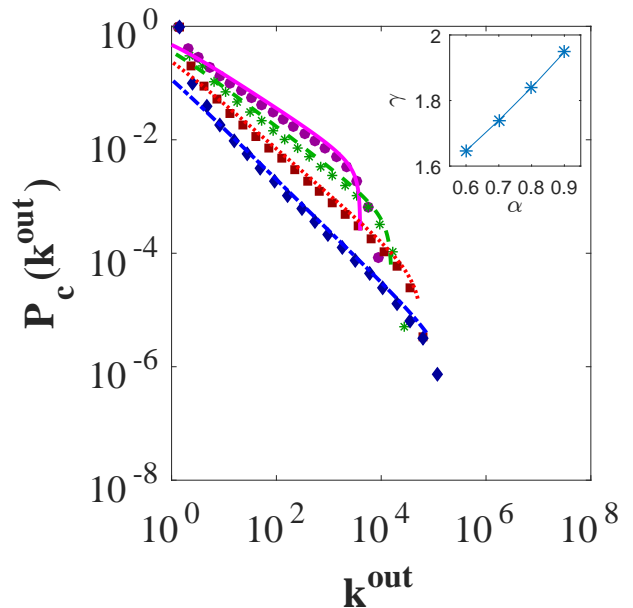

FIG. 7: (Color online). Cumulative distributions of the out-degrees. For each choice of parameter α, 50 realizations of the network were generated and averaged over. The simulated results were obtained for $t = 10^6$ and are represented by purple circles ($\alpha = 0.6$), green stars ($\alpha = 0.7$), red squares ($\alpha = 0.8$) and blue diamonds ($\alpha = 0.9$). The numerical results are represented by the purple solid line ($\alpha = 0.6$), green dashed line ($\alpha = 0.7$), red dotted line ($\alpha = 0.8$) and blue dot-dashed line ($\alpha = 0.9$). The inset shows the fitted exponents of the simulated distributions for the four values of α.

Figure 8 shows theoretical predictions for the full cumulative generalized out-degree distribution, obtained from numerical evaluation of Eq. (70) for a selection of values of α. Also included in the figure are the results of simulations for the same values of α. The inset plot shows the values of γ obtained from fitting power-laws to the tails of the simulation data. We see that the generalized out-degrees of the nodes follow power-laws with increasing values of γ for larger α, and with all values of γ between $\gamma = 1$ and $\gamma = 2$. From the inset plot we see that the exponents of tails of the distributions are quite close to the exponents of the generalized out-strengths $\delta = 1 + \alpha$.

VIII. STRENGTH VERSUS DEGREE

The models we have introduced all produce networks or simplicial complexes where the strengths of the nodes and total number of nodes follow the statistics of the Pitman-Yor model of balls in boxes. In contrast, the degree statistics do not follow the Pitman-Yor model, as links (or triangles) between nodes may be weighted multiple times without altering the degrees (or generalized degrees) of the nodes. In section VII we saw that in
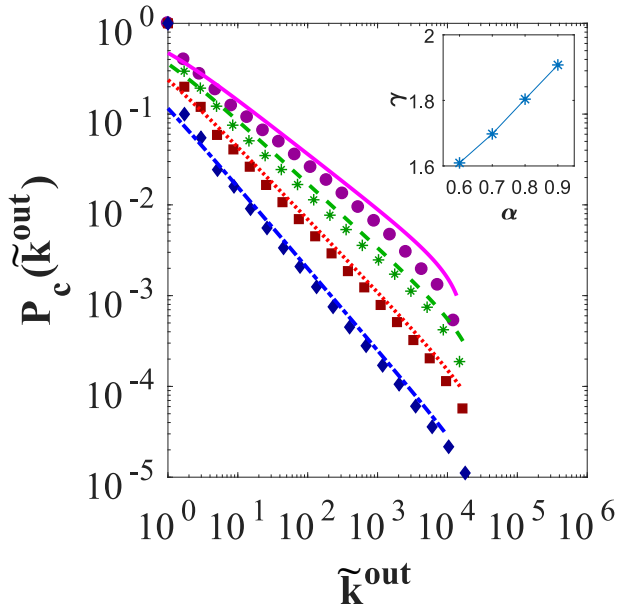
FIG. 8: (Color online). Cumulative distributions of the generalized out-degrees. For each choice of parameter α, 50 realizations of the simplicial complex were generated and averaged over. The simulated results were obtained for $t = 10^5$ and are represented by purple circles ($\alpha = 0.6$), green stars ($\alpha = 0.7$), red squares ($\alpha = 0.8$) and blue diamonds ($\alpha = 0.9$). The numerical results are represented by the purple solid line ($\alpha = 0.6$), green dashed line ($\alpha = 0.7$), red dotted line ($\alpha = 0.8$) and blue dot-dashed line ($\alpha = 0.9$). The inset shows the fitted exponents of the simulated distributions for the four values of α.

the directed network and directed simplicial complex versions, the exponents of the degree and generalized degree distributions are a fairly close match for the exponents of the strength distributions, while in the undirected version the exponent of the degree distribution is always equal to 2. An interesting question is therefore what is the relation between the degrees and generalized degrees of the nodes and their strengths and generalized strengths? In particular, in the directed network and directed simplicial complex versions we would like to know to what extent the strengths and generalized strengths can act as proxies for the degrees and generalized degrees or whether the stregth increases super-linearly with the degree of the nodes [33, 37]. To this end we have run simulations of the models for various values of α with the aim of extracting the average relations over all of the realizations. The total number of nodes in a realization is a random variable and is in fact a non-self-averaging quantity [25]. This means that over a set of realizations the total number of nodes can vary widely. This is important as the probability that a source node either gains a new link (or triangle) or has one of its existing ones reinforced depends on the ratio of the relative size of the degree of the node with respect to N. Therefore, to see the effect of this 'crowd-

ing' of the links of high degree nodes we have normalized the strength and degree data by dividing by N for each realization before averaging over all realizations. Figure 9 shows the relation between the normalized strengths and degrees of the nodes for the three versions of the model. We see from panel a) that the strengths of the nodes in the undirected version are significantly higher than their degrees, with the effect being greater for nodes with higher degree. This is expected, as the probability of a link being reinforced more than once is larger when the strengths of it's two nodes are larger. In contrast, we see from panels b) and c) that the average out-strengths and average generalized out-strengths are very close to equal to the out-degrees and generalized out-degrees of the nodes in the directed network and simplicial complex versions respectively, suggesting the strengths may indeed act as proxies for the degrees.

IX. CLUSTERING AND DEGREE CORRELATIONS

In this section we explore using simulations the clustering and degree correlations of the undirected (and unweighted) networks produced by the three versions of our model. In order to compare the results for the undirected network model with the results of the directed network and directed simplicial complex versions, we decided to discard the information about the direction of the links in the directed network and directed simplicial complex versions. Figure 10 shows the average degree $knn(k)$ of the neighbours of nodes with given degree k for the three versions. We see that for all three versions of the model, the networks produced are strongly disassortative. Interestingly in the undirected version, despite the fact that in Sec. VII we found that the degree distribution has the same power-law exponent for different values of α, the strength of the disassortativity appears to be greater for increasing values of α. A likely explanation for this trend is that for larger α there is bias away from adding links between existing high degree nodes and towards creating links between a new node and a high degree node. Figure 11 shows the average clustering of all nodes in the networks against the model parameter α for the three versions of the model. We see that the average clustering decreases with increasing values of α for all three versions.

X. CONCLUSIONS

In this paper we have presented three similar models for producing dense networks with power-law distributions of the strengths and degrees. The growth mechanisms of these models are analogous to the Pitman-Yor process, a stochastic process well-known among probability theorists for generating random partitions with power-law distributions of block sizes. Our undirected
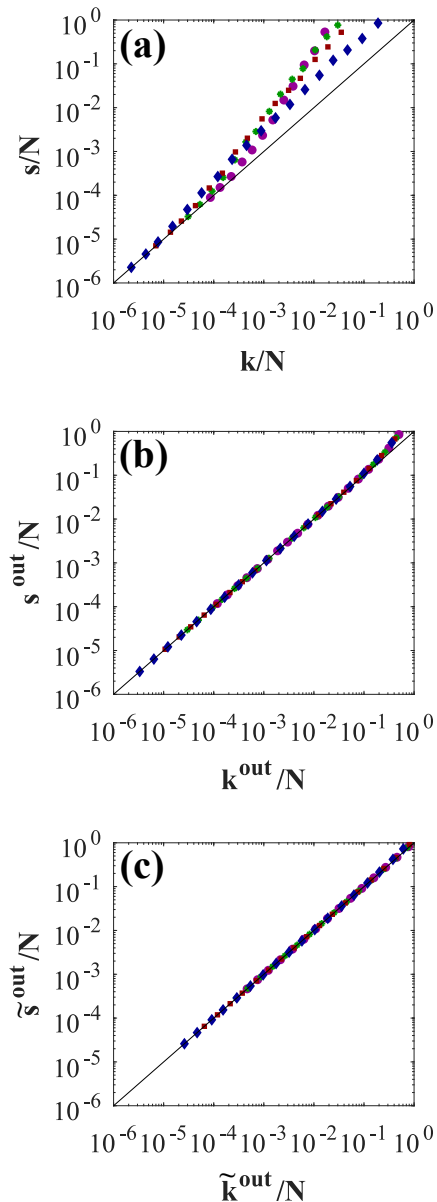
FIG. 9: (Color online). Plot of the normalized strengths versus the normalized degrees. Panels (a), (b) and (c) show the results for the undirected, directed and simplicial complex versions respectively. For each choice of parameter α, 50 realizations of the simplicial complex were generated and averaged over. The results were obtained for $t = 10^5$ and are represented by purple circles ($\alpha = 0.6$), green stars ($\alpha = 0.7$), red squares ($\alpha = 0.8$) and blue diamonds ($\alpha = 0.9$). The solid black line is the function $f(\frac{k}{N}) = \frac{k}{N}$, and is there as a guide to see how closely the strengths match the degrees.
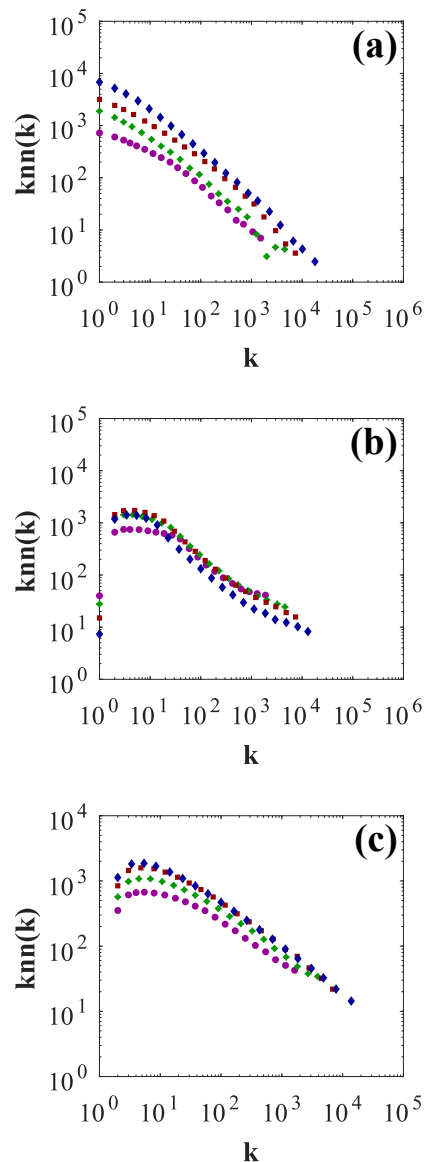

FIG. 10: (Color online). Average degree of the neighbours of nodes of degree k, taken from simulation data. Panels (a), (b) and (c) show the results for the undirected network, directed network and simplicial complex respectively. For each model version and each choice of parameter α, 50 realizations of the network were generated and averaged over. The results were obtained for $t = 10^5$ and for $\alpha = 0.6$, (purple circles), $\alpha = 0.7$ (green stars), $\alpha = 0.8$ (red squares), and $\alpha = 0.9$ (blue diamonds).

model can in one sense be thought of as a network with multiedges and a power-law degree distribution with tunable dense exponent $\gamma = 1 + \alpha \in (1, 2)$ or in a different sense as a weighted network with a power-law degree distribution with the border-line dense exponent $\gamma = 2$. Our directed network model produces dense directed networks with out-degree distributions that follow a power-law with tunable exponent $\gamma = 1 + \alpha$, and homogeneous
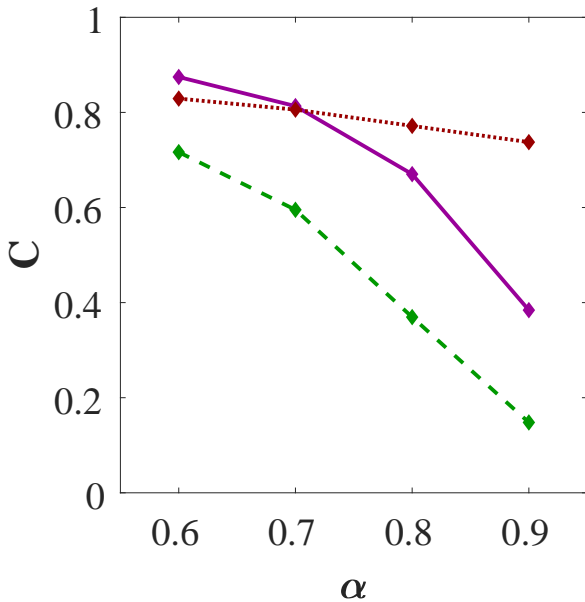
FIG. 11: (Color online). Average clustering of the nodes at different values of α, taken from simulation data with $t = 10^5$. For each model version and each choice of parameter α, 50 realizations of the network were generated and averaged over. Results from the three model versions are represented by the purple solid line (undirected), green dashed line (directed) and red dotted line (simplicial complex).

distributions of the in-degrees. Our simplicial complex version extends the concept of a scale-free network to a scale-free simplicial complex with power-law distribution of the generalized out-degrees. These models demonstrate the difficulty in producing networks that are both dense and scale-free. However, they give insight into the possible mechanisms by which real-world networks densify, and may have a use as null-models for the growth of on-line social networks, recommendation networks or the brain. We show that the models are amenable to analytical calculations through our mean-field approach, and through simulation we verify the accuracy of our mean-field calculations.

[1] A.-L. Barabási, Network science, Cambridge University Press (2016).
[2] M. E. J. Newman, Networks: An introduction, Oxford University Press (2010).
[3] S. N. Dorogovtsev and J. F. F. Mendes, Advances in physics **51**, 1079 (2002).
[4] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[5] G. Bianconi and A.L. Baraási, EPL (Europhysics Letters) **54**, 436 (2001).
[6] G. Bianconi and A.L. Barabási, Phys. Rev. Lett. **86**, 5632 (2001).
[7] P.L. Krapivsky and S. Redner, Phys. Rev. E. **63**, 066123 (2001).
[8] S. N. Dorogovtsev, J. F. F. Mendes and A.N. Samukhin, Structure of growing networks with preferential linking. Phys. Rev. Lett. **85**, 4633 (2000).
[9] J.Kim, P.L. Krapivsky, B. Kahng, and S. Redner, Phys. Rev. E, **66**, 055101 (2002).
[10] P.L. Krapivsky, and S. Redner, in *Statistical mechanics of complex networks* (pp. 3-22). Springer, Berlin, Heidelberg (2003).
[11] I. Ispolatov, P.L. Krapivsky and A. Yuryev, Phys. Rev. E. **71**, 061911 (2005).
[12] R. Lambiotte, P.L. Krapivsky, U. Bhat and S. Redner, Phys. Rev. Lett. **117**, 218301 (2016).
[13] H. Seyed-Allaei, G. Bianconi and M. Marsili, Phys. Rev. E. **73**, 046113 (2006).
[14] T. Zhou, M. Medo, G. Cimini, Z.K. Zhang and Y.C. Zhang, PLoS One **6**, e20648 (2011).
[15] P. Bonifazi, M. Goldin, M.A. Picardo, I. Jorquera, A. Cattani, G. Bianconi, A. Represa, Y. Ben-Ari and R. Cossart, Science **326**, 1419 (2009).
[16] P. Diaconis and S. Janson arXiv preprint arXiv:0712.2749 (2007).
[17] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós and K. Vesztergombi, Adv. Math., **219**, 1801 (2008).
[18] P.J. Wolfe and S.C. Olhede, arXiv preprint arXiv:1309.5936 (2013).
[19] H. Crane, and W. Dempsey, arXiv preprint arXiv:1603.04571 (2016).
[20] F. Caron and E.B. Fox, J. R. Stat. Soc. B **79**, 1295 (2017).
[21] C.I. Del Genio, T. Gross and K.E. Bassler, Phys. Rev. Lett. **107**, 178701 (2011).
[22] G. Timár, S.N. Dorogovtsev and J.F.F. Mendes, Phys. Rev. E, **94**, 022302 (2016).
[23] J. Pitman, Lecture Notes in Mathematics **1875** (2006).
[24] A. Gnedin, B. Hansen, and J. Pitman, Probability Surveys **4** 146 (2007).
[25] B. Bassetti, M. Zarei, M.C. Lagomarsino and G. Bianconi, Phys. Rev. E. **80**, 066118 (2009).
[26] C. Giusti, R. Ghrist and D.S. Bassett, Jour. Comp. Neuro. **41**, 1 (2016).
[27] G. Bianconi, EPL **111**, 56001 (2015).

[28] Z. Wu, G. Menichetti, C. Rahmede and G. Bianconi, Scientific Reports **5**, 10073 (2015).

[29] O.T. Courtney and G. Bianconi, Phys. Rev. E. **93**, 062311 (2016).

[30] K. Zuev, O. Eisenberg and D. Krioukov, Journal of Physics A: Mathematical and Theoretical **48**, 465002 (2015).

[31] G. Bianconi and C. Rahmede, Phys. Rev. E **93**, 032315 (2016).

[32] G. Bianconi and C. Rahmede, Sci. Rep. **7**, 41974 (2017).

[33] O.T. Courtney and G. Bianconi, Phys. Rev. E. **95**, 062301 (2017).

[34] D. Cohen, A. Costa, M. Farber, and T. Kappeler, Discrete & Computational Geometry, **47**, 117 (2012).

[35] J. G. Young, G. Petri, F. Vaccarino and A. Patania, Phys. Rev. E **96**, 032312 (2017).

[36] M. W. Reimann, et al. Front. Comp. Neuro., **11**, 48 (2017).

[37] A. Barrat, M. Barthélemy and A. Vespignani, Proc. Nat. Acad. Sci. U. S. A. **101** 3747 (2004).

[38] H.A. Simon, Biometrika **42**, 425 (1955).