Published in final edited form as:

J Am Stat Assoc. 2018 ; 113(523): 1311–1326. doi:10.1080/01621459.2017.1341413.

EDGE EXCHANGEABLE MODELS FOR INTERACTION NETWORKS

HARRY CRANE and

Department of Statistics & Biostatistics, Rutgers University, 110 Frelinghuysen Avenue,
Piscataway, NJ 08854, USA

WALTER DEMPSEY

Department of Statistics, University of Michigan, 1085 S. University Ave, Ann Arbor, MI 48109,
USA

Abstract

Many modern network datasets arise from processes of interactions in a population, such as phone calls, email exchanges, co-authorships, and professional collaborations. In such interaction networks, the edges comprise the fundamental statistical units, making a framework for edge-labeled networks more appropriate for statistical analysis. In this context we initiate the study of *edge exchangeable network models* and explore its basic statistical properties. Several theoretical and practical features make edge exchangeable models better suited to many applications in network analysis than more common vertex-centric approaches. In particular, edge exchangeable models allow for sparse structure and power law degree distributions, both of which are widely observed empirical properties that cannot be handled naturally by more conventional approaches. Our discussion culminates in the *Hollywood model*, which we identify here as the canonical family of edge exchangeable distributions. The Hollywood model is computationally tractable, admits a clear interpretation, exhibits good theoretical properties, and performs reasonably well in estimation and prediction as we demonstrate on real network datasets. As a generalization of the Hollywood model, we further identify the *vertex components model* as a nonparametric subclass of models with a convenient stick breaking construction.

Phrases

interaction data; edge-labeled network; sparse network; power law distribution; exchangeable random graph; scale-free network; edge exchangeability

1. Introduction

Interaction networks commonly arise from communications (Klimt and Yang, 2004; Opsahl and Panzarasa, 2009), collaborations (Barabási and Albert, 1999; Newman, 2001), and relationships (McAuley and Leskovec, 2012; Zachary, 1977) among individuals in a population. In this context, an unlabeled network structure, with vertices corresponding to

HARRY CRANE, URL: <http://stat.rutgers.edu/home/hcrane>

R code and links to datasets publicly available at <https://github.com/wdempsey/e2>.

elements of the population and edges representing interactions among adjacent vertices as in Figure 1(a), captures the sufficient information in the data. Some basic features of interaction networks distinguish them from the random graphs studied more broadly in the network science literature, making alternative approaches crucial to methodological development in statistical network analysis. Here we initiate the study of *edge exchangeable network models*, which we introduce below as a more appropriate way to model and analyze certain network datasets.

Edge exchangeable models are most appropriate for network datasets in which the edges are the statistical units, as they are for the interaction networks considered below. From a sampling perspective, edge exchangeability better reflects the manner in which interaction networks arise by a process of edge sampling, as opposed to the vertex sampling mechanisms implicit in vertex exchangeable models. Many other popular models, notably exponential random graph models (ERGMs), fail to admit a clear interpretation in terms of sampling. For further discussion of ERGMs we refer the reader elsewhere (Chatterjee and Diaconis, 2013; Shalizi and Rinaldo, 2013).

In a movie collaboration network, for example, each movie corresponds to an interaction among its cast of actors. If movies are sampled by drawing uniformly from, say, the Internet Movie Database (IMDB), then the sampled movies are more likely to include prolific actors. Similarly, if a co-authorship network is constructed by sampling publications randomly, then authors (i.e., vertices) with more publications are more likely to appear in the sample. In either case, vertices are not sampled directly but instead arise as a byproduct of the interaction (i.e., edge) sampling process, making higher degree vertices more likely to appear in the sample.

Edge exchangeability captures this feature of interaction networks. At an intuitive level, edge exchangeability corresponds to the assumption that the observed edges are a representative sample of the population of all edges, which in turn implies that the observed vertices are size-biased according to their prominence in interactions. Our representation theorem for edge exchangeable networks (Theorem 3.2) makes this heuristic rigorous. In practice, the observed interactions may or may not be representative of a larger population, but the edge exchangeable models presented here lay the foundation for a potentially broader framework for modeling interaction networks driven by edge growth, which is presently absent from the statistical literature.

The class of edge exchangeable models has wider implications in network science, where much of the current focus centers on modeling and inference for networks that are sparse and/or have power law degree distributions. These empirical properties reflect a certain heterogeneity in observed network structure which is not reflected in the homogeneous networks produced by the commonly studied graphon models, stochastic blockmodels, and exponential random graph models.

As the edges of many network datasets often correspond to some process of interactions, the edge exchangeable models introduced here give a natural alternative for modeling many sparse, heterogeneous networks. In particular, the heterogeneity inherent in the size-biased

sampling of vertices in edge exchangeable networks offers insight into how sparse and power law behavior can be handled in a meaningful and tractable statistical setting. Our framework of edge exchangeability, therefore, addresses the following foundational question of network modeling:

Is there a notion of probabilistic symmetry whose ergodic measures [...] describe useful statistical models for sparse graphs with network properties? (Orbanz and Roy, 2015, p. 459)

The class of edge exchangeable models provides a positive answer to this question, as we discuss further within our broader development of *edge exchangeable models* for interaction networks below.

1.1. Outline

We define the setup more precisely in Section 2 and go on to establish basic properties of edge exchangeable models throughout Sections 3-8. We formally introduce and define the concept of *edge exchangeability* in Section 3. These developments grow out of the conceptual discussion in Section 2 in which we identify edge-labeled networks as the more natural way to represent networks built from interaction datasets. In addition to modeling the data in a manner that respects the basic structure endowed by the observation process, we show that edge exchangeable models are able to account for basic structural features, namely sparsity and power law degree distribution, which appear in many modern network datasets. As part of this development, we prove a representation theorem for the class of all edge exchangeable models in Section 3.2. In Section 4 we establish fundamental properties of the newly proposed *Hollywood model*, which we introduce here as the canonical edge exchangeable model for its many elegant theoretical and computational features. We demonstrate these features of the Hollywood model throughout the closing sections. As a generalization to the Hollywood model, we identify the more general nonparametric class of *vertex components models* in Section 4.2. We note that the concept of edge exchangeability and all of the models studied here are original to this article and were first proposed here. We do, however, defer some of the technical aspects of the representation theorem for edge exchangeable graphs and further discussion of specific interaction network datasets to the supplementary material accompanying the present paper (Crane and Dempsey, 2017)

2. Interaction data

Definition 2.1 (Interaction data)

For a set $\mathcal{P}$, we write $\text{fin}(\mathcal{P})$ to denote the set of all finite (ordered) multisets of $\mathcal{P}$. An *interaction process for a population* $\mathcal{P}$ is a correspondence $\mathcal{I}: I \rightarrow \text{fin}(\mathcal{P})$ between a set I indexing interactions and finite multisets of $\mathcal{P}$.

Definition 2.1 puts no restrictions on the cardinality of the sets I and $\mathcal{P}$, but for all practical purposes it is sufficient to regard I as $[n] := \{1, \dots, n\}$ (if finite with cardinality $n - 1$) or $\mathbb{N} := \{1, 2, \dots\}$ (if countably infinite) and $\mathcal{P}$ as any finite or countable set indexing a population. In Definition 2.1, each $\mathcal{I}(i)$ is a set containing the individuals involved in interaction $i \in I$. The ordering within $\mathcal{I}(i)$ captures directional interactions present in some

applications, for example, the sending of emails from one account to a list of recipients. This setup contains the case in which interactions are undirected, which we obtain by specifying a canonical ordering of the elements in $\mathcal{P}$ and listing interactions according to that ordering. As an example, let $I = \{1, 2, 3, 4\}$ and $\mathcal{P} = \{a, b, c, d, e\}$ with $\mathcal{J}: I \rightarrow \text{fin}(\mathcal{P})$ defined by

$$\mathcal{J}(1) = (a, b), \quad \mathcal{J}(2) = (c, a), \quad \mathcal{J}(3) = (d, e), \quad \mathcal{J}(4) = (a, c). \quad (1)$$

Notice that these interactions can be represented by the directed graph with vertices labeled in $\mathcal{P}$ and edges labeled in I as in Figure 2. It is from network representations of interaction processes as in Figure 2 that we derive our theory for edge-labeled graphs below.

Definition 2.1 thus captures the structure of many interaction datasets studied in the networks literature. In the actors collaboration network (Barabási and Albert, 1999; Rossi and Ahmed, 2015), for example, I indexes a set of movies, with each movie $i \in I$ corresponding to an interaction involving the set of actors $\mathcal{J}(i)$ in its cast. In the Enron network (Klimt and Yang, 2004), interactions correspond to emails exchanged among Enron employees. In the Wikipedia network (Leskovec et al., 2010), a directed interaction (i, j) represents that individual i voted in favor of j for promotion to the administrator role on Wikipedia. In the karate club network (Zachary, 1977), each edge corresponds to a social interaction between two karate club members. See the supplementary material (Crane and Dempsey, 2017, Appendix C) for a list of other common interaction datasets.

Remark 2.2

Note that the interactions in Definition 2.1 can be of any finite size, reflecting the fact that movies generally involve more than two actors and emails can involve a list of multiple recipients. This setup bodes well for our development of edge-labeled network models, which easily incorporate the higher-order structure of these multi-way interactions by allowing edges of any finite arity. For illustrative purposes, all figures depict networks with only binary edges, but both directed and undirected networks are shown depending on context.

Remark 2.3

The interaction processes in Definition 2.1 are the primary objects of our study, for which we assume no additional information, such as covariates. Incorporating covariates is notoriously difficult in network analysis, with only a limited number of successful attempts known to date (Airoldi et al, 2011; Latouche et al, 2015; Mariadassou et al, 2010; Sweet, 2015; Tallberg, 2004; Zhang et al, 2015). Butts (2008) and Perry and Wolfe (2013) discuss elements of social theory in the context of interaction networks.

2.1. Statistical units

The convention of representing network data as a graph with labeled vertices, as in Figure 1(b), implicitly, and in most cases unintentionally, treats the vertices as the statistical units for the intended application. To be clear, the mere assignment of labels to the observed vertices presents no immediate issue. The trouble arises only when the network is often

tacitly assumed by the model to be a partial observation of a larger population process, as is often the case.

In this situation, the units are inherently tied to sampling, in that they represent the smallest elements on which observations are taken. The act of labeling vertices and treating the observed data as a graph $G = (V, E)$, with vertex set V and edge set $E \subseteq V \times V$, is consistent with the assumption that G is a partial observation of a population graph $G' = (V', E')$ with $V' \supseteq V$ and $E' \subseteq V' \times V'$, so that the data $G = G'|_V$ corresponds to the restriction of G' to vertices V , that is, $G = (V, E)$ satisfies $E = E' \cap (V \times V)$. In this way, the observed data G , even if not explicitly stated or intended, is assumed to contain information about all possible edges between sampled vertices.

For interaction data, however, the interactions, and therefore edges, are the fundamental units of observation, as is clear from the example in (1) and the datasets mentioned above. For example, we may observe the actors dataset by sampling movies without replacement from the Internet Movie Database¹ (IMDB). Or, in the Enron dataset, we may observe only those emails exchanged during a specific period of time, which constitutes a sample from the collection of all emails exchanged in the company's history. In either case, the data result from a partial observation of the interactions.

For this purpose, we assume a population process $\mathcal{J}: \mathbb{N} \rightarrow \text{fin}(\mathcal{P})$ of a potentially infinite collection of interactions indexed by $\mathbb{N}$. From $\mathcal{J}$, we observe the restriction $\mathcal{J}|_S: S \rightarrow \text{fin}(\mathcal{P})$, $s \mapsto \mathcal{J}(s)$, to some finite sample $S \subset \mathbb{N}$. The apparent role of interactions, and therefore edges, as the statistical units suggests the alternative representation as an edge-labeled network, as in Figure 1(c).

2.2. Edge-labeled networks

The foregoing discussion stresses the difference between the *population process* as a collection of interactions $\mathcal{J}: \mathbb{N} \rightarrow \text{fin}(\mathcal{P})$, as in (1), and its network representation as in Figure 2.

Let $\mathcal{J}: \mathbb{N} \rightarrow \text{fin}(\mathcal{P})$ be an interaction process. Any bijection $\rho: \mathcal{P} \rightarrow \mathcal{P}'$ induces an action $\text{fin}(\mathcal{P}) \rightarrow \text{fin}(\mathcal{P}')$ by the map

$$s = (s_1, \dots, s_r) \mapsto \rho s = (\rho(s_1), \dots, \rho(s_r)) \in \text{fin}(\mathcal{P}'). \quad (2)$$

With this, any bijection $\rho: \mathcal{P} \rightarrow \mathcal{P}'$ acts on $\mathcal{J}: \mathbb{N} \rightarrow \text{fin}(\mathcal{P})$ by composition of maps, $\mathcal{J} \mapsto \rho \mathcal{J}$, with $\rho \mathcal{J}: \mathbb{N} \rightarrow \text{fin}(\mathcal{P}')$ given by

$$(\rho \mathcal{J})(i) = \rho(\mathcal{J}(i)), \quad i \in \mathbb{N}, \quad (3)$$

¹<http://www.imdb.com/>

as defined in (2). From this, we define the *edge-labeled network induced by* $\mathcal{J}: S \rightarrow \text{fin}(\mathcal{P})$ as the equivalence class

$$\mathcal{E}_{\mathcal{J}} = \bigcup_{\mathcal{P}': \# \mathcal{P}' = \# \mathcal{P}} \{ \mathcal{J}': S \rightarrow \text{fin}(\mathcal{P}') : \rho \cdot \mathcal{J} = \mathcal{J}' \text{ for some bijection } \rho: \mathcal{P} \rightarrow \mathcal{P}' \}, \quad (4)$$

which corresponds to an edge-labeled network structure obtained by removing vertex labels from the network representation of an interaction sequence. For example, the edge labeled network on the right of Figure 3 corresponds to the equivalence class of all possible interaction processes with the same structure, of which three such examples are given on the left of Figure 3. The edge-labeled graph in Figure 3 corresponds to that of the interaction process described in (1) and shown in Figure 2. Note that $\mathcal{E}_{\mathcal{J}}$ does not depend on the specific set $\mathcal{P}$, and so we may disregard $\mathcal{P}$, or treat it implicitly as $\mathcal{P} = \mathbb{N}$, in our discussion. We write $\mathcal{E}_S$ for the set of networks with edges labeled in $S \subseteq \mathbb{N}$.

For any $S' \subseteq S$, we define the *restriction* of $\mathcal{E} \in \mathcal{E}_S$ to $\mathcal{E}_{S'}$ by $\mathcal{E}|_{S'}$, the edge-labeled network obtained by removing any edges labeled in $S \setminus S'$. If $\mathcal{E} = \mathcal{E}_{\mathcal{J}}$ for some interaction process $\mathcal{J}: S \rightarrow \text{fin}(\mathcal{P})$, then $\mathcal{E}|_{S'}$ is the edge-labeled network induced by the restricted process $\mathcal{J}|_{S'}: S' \rightarrow \text{fin}(\mathcal{P})$, $s \mapsto \mathcal{J}(s)$. Note that the restriction of any such $\mathcal{E}$ does not depend on the choice of representative $\mathcal{J}$.

2.3. Network properties

For any edge-labeled network $\mathcal{E}$, we define $\nu(\mathcal{E})$ as the number of non-isolated vertices in $\mathcal{E}$, that is, the number of vertices involved in at least one of the edges of $\mathcal{E}$. We also define $e(\mathcal{E})$ as the number of edges in $\mathcal{E}$, $M_k(\mathcal{E})$ as the number of k -ary edges in $\mathcal{E}$ for each $k \geq 1$, $N_k(\mathcal{E})$ as the number of vertices that appear exactly k times in $\mathcal{E}$, and $d(\mathcal{E}) = (d_k(\mathcal{E}))_{k \geq 1}$ as the degree distribution of $\mathcal{E}$, where $d_k(\mathcal{E}) = N_k(\mathcal{E})/\nu(\mathcal{E})$ is the proportion of vertices with degree k in $\mathcal{E}$. For example, the edge-labeled network $\mathcal{E}$ in Figure 1(c) has $\nu(\mathcal{E}) = 6$, $e(\mathcal{E}) = 6$, $M_k(\mathcal{E}) = 6$ for $k = 2$, $M_k(\mathcal{E}) = 0$ for $k \geq 3$, and $d(\mathcal{E}) = (3/6, 1/6, 1/6, 1/6, 0, \dots)$. Notice that these statistics do not depend on the edge labels assigned to $\mathcal{E}$ and they do not account for the single isolated vertex shown in Figure 1(c).

Definition 2.4 (Sparsity and power law degree distribution)—Let $(\mathcal{E}_n)_{n \geq 1}$ be a sequence of edge-labeled networks for which $e(\mathcal{E}_n) \rightarrow \infty$ as $n \rightarrow \infty$. The sequence $(\mathcal{E}_n)_{n \geq 1}$ is *sparse* if

$$\limsup_{n \rightarrow \infty} \frac{e(\mathcal{E}_n)}{m(\mathcal{E}_n)} = 0, \quad (5)$$

where $m(\mathcal{E}_n) = e(\mathcal{E}_n)^{-1} \sum_{k \geq 1} k M_k(\mathcal{E}_n)$ is the average arity of the edges in $\mathcal{E}_n$.

The sequence $(\mathcal{E}_n)_{n \geq 1}$ exhibits a *power law degree distribution* if for some $\gamma > 1$ the degree distributions $(d(\mathcal{E}_n))_{n \geq 1}$ satisfy $d_k(\mathcal{E}_n) \sim \ell(k)k^{-\gamma}$ as $n \rightarrow \infty$ for all large k for some slowly varying function $\ell(x)$, that is, $\lim_{x \rightarrow \infty} \ell(tx)/\ell(x) = 1$ for all $t > 0$, where $a_n \sim b_n$ indicates that $a_n/b_n \rightarrow 1$ as $n \rightarrow \infty$. More precisely, $(\mathcal{E}_n)_{n \geq 1}$ has power law degree distribution with index γ if

$$\lim_{k \rightarrow \infty} \lim_{n \rightarrow \infty} \frac{d_k(\mathcal{E}_n)}{\ell(k)k^{-\gamma}} = 1$$

for some slowly varying function $\ell(k)$.

Remark 2.5—The slowly varying function $\ell(k)$ makes the definition of power law distribution robust to finite sample behavior and only affects the shape of the distribution. It does not alter the tail behavior.

Our definition of sparsity in (5) refines the usual notion of sparsity for vertex-labeled graphs, as defined, for example, by Nesetril and Ossona de Mendez (2012). In the most familiar case when every interaction involves exactly 2 vertices, $m(\mathcal{E}_n) = 2$ for all $n \geq 1$ and $(\mathcal{E}_n)_{n \geq 1}$ is sparse if

$$\limsup_{n \rightarrow \infty} \frac{e(\mathcal{E}_n)}{v(\mathcal{E}_n)^2} = 0.$$

Though edge exchangeable models apply more broadly, much of the coming discussion in Section 4 focuses on sparsity and power law degree distribution because of their prevalence in network datasets, e.g., Barabási and Albert (1999), and the difficulty in modeling these properties, cf. Orbanz and Roy (2015).

3. Edge exchangeable network models

Given $\mathcal{E} \in \mathfrak{E}_S$, for $S \subseteq \mathbb{N}$, and a permutation $\sigma : S \rightarrow S$, we write $\mathcal{E}^\sigma$ to denote the edge-labeled network obtained by relabeling the edges of $\mathcal{E}$ according to σ , as demonstrated in Figure 4.

More precisely, given an edge-labeled network $\mathcal{E}$ based on interaction data $\mathcal{I} : S \rightarrow \text{fin}(\mathcal{P})$ and a permutation $\sigma : S \rightarrow S$, $\mathcal{E}^\sigma$ is the edge-labeled network induced by $\mathcal{I}^\sigma : S \rightarrow \text{fin}(\mathcal{P})$ with $\mathcal{I}^\sigma(i) = \mathcal{I}(\sigma^{-1}(i))$ for each $i \in S$. (It is clear from (4) that the definition of $\mathcal{E}^\sigma$ does not depend on which representative of the equivalence class $\mathcal{E}_{\mathcal{I}}$ is chosen when defining $\mathcal{E}^\sigma$.)

Note the distinction between the action $\rho\mathcal{I}$ defined in (3) and $\mathcal{I}^\sigma$ given here. Whereas the action $\mathcal{I} \mapsto \rho\mathcal{I}$ in (3) reassigns labels to the population through a bijection $\rho : \mathcal{P} \rightarrow \mathcal{P}'$, the operation $\mathcal{I} \mapsto \mathcal{I}^\sigma$ corresponds to relabeling interactions and $\mathcal{E} \mapsto \mathcal{E}^\sigma$ to relabeling edges.

In the following definition and in the remainder of the article, we often write $\mathcal{Y}$ to denote a random edge-labeled network. We assume that edges are labeled in a finite or countable set

S . When S is finite, so is $\mathfrak{G}_S$, and so we furnish the full power set σ -algebra. When S is countable, we implicitly assume $S = \mathbb{N}$ and we equip $\mathfrak{G}_S$ with the Borel σ -algebra associated to the product discrete topology induced by the metric

$$d(\mathcal{E}, \mathcal{E}') = 1/(1 + \sup\{n \geq 1: \mathcal{E}|_{[n]} = \mathcal{E}'|_{[n]}\}), \quad \mathcal{E}, \mathcal{E}' \in \mathfrak{G}_S,$$

with convention $1/\infty = 0$. To reassure readers uninterested in such technicalities, we stress that no complicated or obscure issues concerning measurability arise in this theory.

Definition 3.1 (Edge exchangeable network)

Let S be a finite or countable set. A random edge-labeled network $\mathcal{Y}$ in $\mathfrak{G}_S$ is *edge exchangeable* if $\mathcal{Y}^\sigma =_{\mathcal{D}} \mathcal{Y}$ for all permutations $\sigma: S \rightarrow S$, where $=_{\mathcal{D}}$ denotes *equality in distribution*.

Edge exchangeable models assign the same probability to all edge-labeled networks that are isomorphic up to relabeling, as shown in Figure 5.

3.1. Sampling interpretation of edge exchangeability

The assumption of edge exchangeability has a clear interpretation in terms of sampling in that the interactions that determine an observed edge-labeled network $\mathcal{E} \in \mathfrak{G}_S$ are assumed to be representative of a larger population of interactions $\mathcal{I}: \mathbb{N} \rightarrow \text{fin}(\mathcal{P})$. Contrast this with the assumption of exchangeability for vertex-labeled graphs, in which case sampled vertices are assumed to be a representative sample of all vertices.

For example, suppose that the interactions in (1) were obtained by sampling without replacement from a database of telephone calls exchanged among a population of phone numbers identified as a, b, c, d, e . Each interaction, say (a, b) , corresponds to a call from a to b . Under the assumed random sampling operation, the interactions $\mathcal{I}(1), \dots, \mathcal{I}(4)$ in (1) are exchangeable as a sequence in the space of pairs $\mathcal{P} \times \mathcal{P}$. Although the vertex labels carry no additional information beyond their participation in the observed interactions, they cannot simply be interchanged in any given observation. For example, the interactions in (1) and those in

$$\mathcal{I}'(1) = (c, b), \quad \mathcal{I}'(2) = (e, c), \quad \mathcal{I}'(3) = (a, d), \quad \mathcal{I}'(4) = (c, e). \quad (6)$$

both induce the same network structure, as shown in Figure 2, but the probability of observing (1) and (6) may in general differ since, for example, phone calls from a to c (represented as (a, c)) may appear in the database with higher frequency than (c, e) . Removing edge labels, as is commonly done to obtain the graphs shown in Figure 6, results in two graphs with the same structure up to relabeling vertices but which may have different probabilities as graphs. This example lays bare the folly of assuming a vertex exchangeable model in networks obtained by random edge sampling. Moreover, it makes clear why edge

exchangeability is more appropriate in this context, as the random sampling of interactions in (1) makes the sequence exchangeable and, thus, endows exchangeability to the edge-labeled network representation in Figures 3 and 5.

3.2. Characterization of edge exchangeable networks

We describe here a special case of a more general de Finetti-type representation theorem of edge exchangeable networks, which we defer to the supplementary material accompanying this article (Crane and Dempsey, 2017, Theorem A.2). The case below is most relevant to the statistical applications we consider.

To construct an edge exchangeable random network, we first choose any $f = (f_s)_{s \in \text{fin}(\mathbb{N})}$ in the $\text{fin}(\mathbb{N})$ -simplex,

$$\mathcal{F}_1 = \left\{ (f_s)_{s \in \text{fin}(\mathbb{N})} : f_s \geq 0 \text{ and } \sum_{s \in \text{fin}(\mathbb{N})} f_s = 1 \right\}.$$

Each $f \in \mathcal{F}_1$ determines a distribution on finite multisets of $\mathbb{N}$ by

$$\text{pr}(X = s | f) = f_s, \quad s \in \text{fin}(\mathbb{N}), \quad (7)$$

which in turn determines an edge exchangeable network in $\mathfrak{G}_{\mathbb{N}}$ as follows.

Let $X_1, X_2, \dots$ be independent, identically distributed (i.i.d.) random multisets drawn from (7). Given $X_1, X_2, \dots$, define the random interaction process $\mathcal{X}: \mathbb{N} \rightarrow \text{fin}(\mathbb{N})$ by $\mathcal{X}(i) = X_i$ and write $\mathcal{E}_{\mathcal{X}}$ to denote the corresponding edge-labeled network obtained through (4), whose distribution we denote by ε_f . See Figure 7(a) for an illustration.

With probability 1, each $s \in \text{fin}(\mathbb{N})$ occurs either zero or infinitely many times in the interaction process $\mathcal{X}: \mathbb{N} \rightarrow \text{fin}(\mathbb{N})$ corresponding to $X_1, X_2, \dots$ chosen i.i.d. from (7), and the resulting edge-labeled network $\mathcal{E}_{\mathcal{X}}$ is edge exchangeable. This is a special case of the more generic edge exchangeable processes, for which interactions can occur zero, one, or infinitely many times. We call an interaction that occurs only once a *blip*, and we call an interaction process and its induced edge-labeled network *blip-free* if it contains no blips. We state here the blip-free version of our general representation theorem (Crane and Dempsey, 2017, Theorem A.2).

Theorem 3.2—Let $\mathcal{Y} \in \mathfrak{G}_{\mathbb{N}}$ be an edge exchangeable network that is blip-free with probability 1. Then there exists a probability measure ϕ on $\mathcal{F}_1$ such that $\mathcal{Y} \sim \varepsilon_{\phi}$, where

$$\varepsilon_{\phi}(\cdot) = \int_{\mathcal{F}_1} \varepsilon_f(\cdot) \phi(df). \quad (8)$$

That is, every blip-free edge exchangeable network $\mathcal{Y} \in \mathfrak{G}_{\mathbb{N}}$ can be generated by first sampling $f \sim \phi$ and, given f , setting $\mathcal{Y} = \mathcal{E}_{\mathcal{X}}$ with $\mathcal{X}: \mathbb{N} \rightarrow \text{fin}(\mathbb{N})$ defined by $\mathcal{X}(i) = X_i$ for $X_1, X_2, \dots$ i.i.d. from (7).

Remark 3.3—The measure ϕ in Theorem 3.2 is not unique. Uniqueness results from a more technical treatment; see the supplementary material (Crane and Dempsey, 2017, Theorem A.2).

Corollary 3.4—The vertices in any blip-free edge exchangeable network $\mathcal{Y} \in \mathfrak{G}_{\mathbb{N}}$ arrive in size-biased order weighted by the relative frequency of their occurrence.

According to Corollary 3.4, edge exchangeability corresponds to an implicit assumption that vertices are sampled in size-biased order according to their relative frequency of occurrence in interactions, whereby sampled vertices are atypical of the population as a whole. By contrast, vertex exchangeability tacitly assumes that sampled vertices are representative of the population of all vertices. In the discussion of randomly sampling from a phone call database in Section 3.1, for example, callers who make a large number of phone calls appear in the database with high frequency and are thus more likely to be observed in a sample than callers who make few phone calls. The observed vertices, therefore, are not representative of the population as a whole.

4. The Hollywood model

We now introduce a canonical family of edge exchangeable network models with several attractive features for statistical applications. Though not limited to any specific application, we phrase the following *Hollywood model* in terms of the actors collaboration network to aid comprehension. In this description, vertices correspond to actors and edges correspond to movies, with each edge containing the set of actors involved in the corresponding movie.

Let $\nu = \{\nu_k\}_{k=1}$ be a probability distribution on the positive integers and let (α, θ) satisfy either

- $0 < \alpha < 1$ and $\theta > -\alpha$ (infinite population) or
- $\alpha < 0$ and $\theta = -k\alpha$ for some $k = 1, 2, \dots$ (finite population).

We generate a sequence of interactions $X_1, X_2, \dots$ as follows.

Given $X_1, \dots, X_{n-1}$, we choose the number of roles in the next movie independently according to $K_n \sim \nu$. Given $K_n = k$, we choose the k actors $X_{n,1}, \dots, X_{n,k}$ in order of their prominence, first filling the lead role, then the second lead role, and so on until all k roles are filled. Let $V_n(j)$ be the number of unique actors seen in all the movies through the $(j-1)$ st actor cast in movie n . (Thus, $V_n(1)$ is the number of unique actors appearing in movies $1, \dots, n-1$.) For each $i = 1, \dots, V_n(j)$, we write $D_{n,j}(i)$ to denote the number of roles for which the actor labeled i has been cast up to and including the $(j-1)$ st role of movie n . (Note that an actor may play more than one role in a given movie.) The actor $X_{n,j}$ cast in the

j th role of movie n is chosen randomly among the actors labeled $1, \dots, V_n(j)$ and a previously unseen actor, labeled $V_n(j) + 1$, according to

$$\Pr(X_{n,j} = i \mid X_1, \dots, X_{n-1}, X_{n,1}, \dots, X_{n,j-1}) \propto \begin{cases} D_{n,j}(i) - \alpha, & i = 1, \dots, V_n(j), \\ \theta + \alpha v_n(j), & i = V_n(j) + 1. \end{cases} \quad (9)$$

We update according to (9) until all k roles of movie n have been filled.

Each interaction $X_1, X_2, \dots$ records the order in which actors are selected in each movie. From $X_1, X_2, \dots$, we define $\mathcal{X}_n: [n] \rightarrow \text{fin}(\mathbb{N})$ by $\mathcal{X}_n(i) = X_i$ for $i = 1, \dots, n$, and we write for the edge-labeled network induced by $\mathcal{X}_n$ as defined in (4). We call the resulting sequence $(\mathcal{Y}_n)_{n \geq 1}$ the *Hollywood process* with parameter (α, θ, ν) , which determines a family of distributions on $\mathfrak{G}_{\mathbb{N}}$, called the *Hollywood model*.

The Hollywood model has a closed form expression for random edge-labeled networks of each finite size $n \geq 1$ given by

$$\begin{aligned} \Pr(\mathcal{Y}_n = \mathcal{E}; \alpha, \theta, \nu) &= \\ &= \left[\prod_{k \geq 1} \nu_k^{M_k(\mathcal{E})} \right] \alpha^{v(\mathcal{E})} \frac{(\theta/\alpha)^{\uparrow v(\mathcal{E})}}{\theta^{\uparrow m(\mathcal{E})}} \prod_{k=2}^{\infty} \exp\{N_k(\mathcal{E}) \log((1-\alpha)^{\uparrow (k-1)})\}, \end{aligned} \quad (10)$$

where $\mathcal{E} \in \mathfrak{G}_{[n]}$, $v(\mathcal{E})$ is the number of non-isolated vertices in $\mathcal{E}$, $(N_k(\mathcal{E}))_{k \geq 1}$ gives the number of vertices with each degree $k \geq 1$, $M_k(\mathcal{E})$ is the number of k -ary edges in $\mathcal{E}$, $m(\mathcal{E}) = \sum_{k \geq 1} k M_k(\mathcal{E})$ is the total degree of $\mathcal{E}$, and $x^{\uparrow j} = x(x+1) \cdots (x+j-1)$ is the ascending factorial function.

Though defined for edge-labeled networks with directed edges, the Hollywood model determines a distribution on undirected edge-labeled networks by ignoring the edge orientations. In this case, the probability of an undirected network $\mathcal{E}^*$ is given by the expression in (10) multiplied by a combinatorial factor $C(\mathcal{E}^*)$ which counts the number of directed networks corresponding to $\mathcal{E}^*$. While we have no closed form expression for $C(\mathcal{E}^*)$, the quantity plays no role in our inferences because it does not depend on the parameters (α, θ, ν) , as we discuss further in Section 5.1.

4.1. Interpretation of parameters and finite population model

The split parameter space of the Hollywood model enables both bounded and unbounded population sizes. The region $0 < \alpha < 1$ and $\theta > -\alpha$ gives rise to a sequence $(\mathcal{Y}_n)_{n \geq 1}$ for which $v(\mathcal{Y}_n) \rightarrow \infty$ almost surely (a.s.) as $n \rightarrow \infty$, as is reasonable to assume for the actors, Enron, and Wikipedia networks. The Karate Club dataset, on the other hand, is known to have exactly thirty-four club members but no limit on the number of interactions between

individuals. The range $\alpha < 0$ and $\theta = -k\alpha$ accommodates this case by describing an edge exchangeable sequence $(\mathcal{Y}_n)_{n \geq 1}$ for which $v(\mathcal{Y}_n) \rightarrow k$ a.s. as $n \rightarrow \infty$.

By (9), $\alpha > 0$ increases the probability of observing previously unseen vertices but decreases the probability of observing a vertex again after its initial occurrence. The range $\alpha < 0$ has the opposite effect. Thus, α values near 1 make it more likely that new edges involve previously unseen vertices, but less likely that previously seen vertices occur in future edges. On the other hand, $\alpha < 0$ corresponds to a finite population size, so that each newly observed vertex decreases the number of unseen vertices and increases the probability that future edges involve previously seen vertices. In the $0 < \alpha < 1$ regime, larger values of θ increase the probability of seeing previously unobserved vertices in new edges, but the effect of θ diminishes as $n \rightarrow \infty$. In Section 4.4, we see that $0 < \alpha < 1$ is directly related to the sparsity and power law behavior of the sequence $(\mathcal{Y}_n)_{n \geq 1}$. The interpretation of parameters

presented here parallels the interpretation of the parameters of the two parameter Ewens–Pitman process, as expounded in (Crane, 2016, Section 5.2).

4.2. Vertex components models

In Section 4.3 we discuss an alternative formulation of the Hollywood model as a special case of the more general subclass of *vertex components models*. In general, vertex components models build on the general form of edge exchangeable networks established in Theorem 3.2. To describe a vertex components model, we construct $f = (f_s)_{s \in \text{fin}(\mathbb{N})}$ by specifying a probability distribution $\nu = \{\nu_k\}_{k=1}$ on the positive integers, choosing a sequence $W = (W_j)_{j=1}$ at random from the infinite simplex

$$\Delta_1 = \left\{ (s_1, s_2, \dots) : \sum_{i \geq 1} s_i = 1 \right\},$$

and then defining

$$f(s_1, \dots, s_k) = \nu_k \prod_{j=1}^k W_{s_j}, \quad (s_1, \dots, s_k) \in \text{fin}(\mathbb{N}). \quad (11)$$

Under (11), we interpret the distribution in (7) as first choosing the size of a random multiset s according to ν and, given $\#s = k$, choosing the k elements i.i.d. from $(W_j)_{j=1}$. (Recall from Definition 2.1 that $\text{fin}(\mathbb{N})$ consists of ordered multisets. We recover the unordered case by writing each $\{s_1, \dots, s_k\}$ so that $s_1 \leq s_2 \leq \dots \leq s_k$ and multiplying (11) by the number of distinct ways to order the elements of s , which can be computed explicitly.)

Stick-breaking representations, see, for example, Sethuraman (1994) or Ishwaran and James (2001), offer a computationally tractable way to fit the vertex components model to data. In this construction, we can generate the sequence $X_1, X_2, \dots$ of interactions simultaneously with the vertex components $W = (W_j)_{j=1}$.

Let $\{\varphi_i\}_{i=1}$ be a collection of probability densities on $[0, 1]$. We first choose the size $K_n \sim \nu$ of the n th directed interaction $X_n = (X_{n,1}, \dots, X_{n,K_n})$ independently for each $n = 1, 2, \dots$. We begin by choosing $K_1 \sim \nu$, putting $X_{1,1} = 1$, and sampling $W_1 \sim \varphi_1$. We continue inductively for each $n = 1, 2, \dots$ as follows. Given K_n and $k < K_n$, we define $X_{n+1,k}$ as the set of all $X_1, \dots, X_n$ and $X_{n+1,1}, \dots, X_{n+1,k}$ up to the k th chosen element of the $(n+1)$ st interaction. Given $X_{n+1,k}$ and $W_1, \dots, W_{V_n}$, where $V_n = \max(X_{n+1,k})$ is the largest vertex label assigned among $X_{n+1,k}$, we choose $X_{n+1,k+1}$, provided $K_{n+1} > k+1$, according to

$$\text{pr}(X_{n+1,k+1} = r \mid X_{n+1, \leq k}, W_1, \dots, W_{V_n}) = \begin{cases} W_r, & r = 1, \dots, V_n, \\ 1 - \sum_{j=1}^{V_n} W_j, & r = V_n + 1. \end{cases} \quad (12)$$

If $X_{n+1,k+1} = V_n + 1$, then we choose $W_{V_n+1} \sim \varphi_{n+1}(\cdot / (1 - \sum_{j=1}^{V_n} W_j))$. (The division by $1 - \sum_{j=1}^{V_n} W_j$ represents the normalization by the remaining length of the stick in the stick-breaking construction.) We continue to draw $X_{n+1,k+2}, \dots, X_{n+1,K_{n+1}}$ as in (12).

From the sequence $X_1, X_2, \dots$ constructed above, we define $\mathcal{X}_n: [n] \rightarrow \text{fin}(\mathbb{N})$ as the interaction process constructed by putting $\mathcal{X}_n(i) = X_i$ and defining to be the edge-labeled network induced by $\mathcal{X}_n$ for each $n \in \mathbb{N}$, as in (4). We compute the joint density of $\mathcal{Y}_n$ and W by

$$\begin{aligned} \text{pr}(\mathcal{Y}_n = \mathcal{E}, (W_1, \dots, W_{v(\mathcal{E})}) \in (dw_i)_{1 \leq i \leq v(\mathcal{E})}; \{\varphi_i\}_{i \geq 1}) &= \\ &= \prod_{j=1}^{v(\mathcal{E})} \left(1 - \sum_{i=1}^{j-1} w_i \right) \varphi_j \left(\frac{w_j}{1 - \sum_{i=1}^{j-1} w_i} \right) w_j^{D_n(j)-1} dw_1 \cdots dw_{v(\mathcal{E})}, \end{aligned} \quad (13)$$

where $D_n(j)$ is the number of times the vertex corresponding to weight w_j appears in $\mathcal{E}$ and $\sum_{i=1}^0 w_i = 0$.

As we discuss below, the Hollywood model corresponds to the vertex components model in (13) with W generated from a Dirichlet process (Ferguson, 1973). This connection turns out to be powerful in establishing the basic statistical properties of the Hollywood model in Section 4.4.

Remark 4.1—We note some similarity between the above vertex components models and the class of random intersection graphs (Britton et al, 2008; Deijfen and Kets, 2009; Fill et al, 2000; Godehardt and Jaworski, 2003; Karónski et al, 1999; Singer, 1995). Random intersection graphs are able to model clustering and inhomogeneity in interactions by

assigning a random weight to each vertex in a manner similar to the vertex weights assigned in the vertex components model; see Deijfen and Kets (2009) for further discussion of this construction.

4.3. Connection between Hollywood and vertex components models

For $\alpha < 0$ and $\theta = -k\alpha$, the Hollywood model corresponds to the vertex components model with $W = (W_1, \dots, W_k)$ chosen from the symmetric Dirichlet distribution with parameter $(\alpha, \dots, \alpha)$ on the $(k-1)$ -simplex. For $0 < \alpha < 1$ and $\theta > -\alpha$, the Hollywood model is a special case of the vertex components model with $W = (W_j)_{j=1}^k$ chosen from the Griffiths–Engen–McCloskey (GEM) distribution with parameter (α, θ) on $\mathbb{R}_+^k$; see Feng (2010) for further details on the GEM distribution and its relationship to the Poisson–Dirichlet distribution. Alternatively, the Hollywood model with $0 < \alpha < 1$ and $\theta > -\alpha$ can be constructed by noting the stick-breaking construction of the GEM distribution and taking ϕ_j to be the density of the Beta distribution with parameter $(1-\alpha, \theta + j\alpha)$ for each $j \geq 1$ (Ishwaran and James, 2001). We recover (10) by marginalizing over W in (13). Crane (2016b) previously noted a connection between the Hollywood model and the Ewens–Pitman two parameter family of distributions on set partitions (Ewens, 1972; Pitman, 2005).

4.4. Statistical properties of the Hollywood model

The close connection between the Hollywood model and the Poisson–Dirichlet distribution through the vertex components interpretation of Section 4.3 addresses the fundamental question of incorporating sparse, power law behavior within a tractable statistical framework, as we discussed in Section 1. Theorems 4.3 and 4.4, in particular, follow from the corresponding power law behavior of the Ewens–Pitman process (Pitman, 2005).

Theorem 4.2—The Hollywood model with parameter (α, θ, ν) determines an edge exchangeable probability distribution on $\mathfrak{G}_{\mathbb{N}}$ for all (α, θ, ν) in the parameter space of the model.

Theorem 4.3—Let $(\mathcal{Y}_n)_{n \geq 1}$ obey the Hollywood process with parameter (α, θ, ν) . For each $n \geq 1$, let $p_n(k) = N_k(\mathcal{Y}_n)/\nu(\mathcal{Y}_n)$, $k \geq 1$, be the empirical degree distribution of $\mathcal{Y}_n$, where $N_k(\mathcal{Y}_n)$ is the number of vertices with degree $k \geq 1$ and $\nu(\mathcal{Y}_n)$ is the number of non-isolated vertices in $\mathcal{Y}_n$. Then, for every $k \geq 1$,

$$p_n(k) \sim \alpha k^{-(\alpha+1)}/\Gamma(1-\alpha) \text{ a.s. as } n \rightarrow \infty,$$

where $\Gamma(t) = \int_0^\infty x^{t-1} e^{-x} dx$ is the gamma function. That is, $(\mathcal{Y}_n)_{n \geq 1}$ has a power law degree distribution with exponent $\gamma = 1 + \alpha \in (1, 2)$.

Theorem 4.4—Let $(\mathcal{Y}_n)_{n \geq 1}$ obey the Hollywood process with parameter (α, θ, ν) for $0 < \alpha < 1$ and $\theta > -\alpha$. Then the expected number of vertices in $\mathcal{Y}_n$ satisfies

$$E(v(\mathcal{Y}_n)) \sim \frac{\Gamma(\theta + 1)}{\alpha \Gamma(\theta + \alpha)} (\mu n)^\alpha \text{ as } n \rightarrow \infty, \quad (14)$$

where $\mu = \sum_{k \geq 1} k \nu_k$ is the mean edge arity. Furthermore, if $1/\mu < \alpha < 1$, then $(\mathcal{Y}_n)_{n \geq 1}$ is almost surely sparse in the sense of (5).

For a more general class of edge exchangeable models that admits sparsity and power law, we define the α -diversity, $0 < \alpha < 1$, of an edge exchangeable process $(\mathcal{Y}_n)_{n \geq 1}$ by

$$S_\alpha = \lim_{n \rightarrow \infty} v(\mathcal{Y}_n) / e(\mathcal{Y}_n)^\alpha, \quad (15)$$

provided the limit exists and is strictly positive and finite almost surely. Following Pitman (2005, Lemma 3.11), an edge exchangeable process $(\mathcal{Y}_n)_{n \geq 1}$ from the vertex components model constructed from $\nu = \{\nu_k\}_{k \geq 1}$ and $W = (W_i)_{i \geq 1}$ has α -diversity if and only if the reordered sequence $W_{(1)} \geq W_{(2)} \geq \dots \geq 0$ satisfies $W_{(i)} \sim Zi^{-1/\alpha}$ a.s. as $i \rightarrow \infty$ for a random variable $0 < Z < \infty$. In this case, $(\mathcal{Y}_n)_{n \geq 1}$ is sparse for $1/\mu < \alpha < 1$, with $\mu = \sum_{k \geq 1} k \nu_k$ and has power law degree distribution with $\gamma = \alpha + 1$.

4.5. Projecting to a network without multiple edges

Though multiple edges are commonplace in networks constructed from interaction processes, many network models, namely the exponential random graph model, are designed to handle only simple graphs. Even some other models which could, in principle, handle multigraphs, such as graphon models and the models in Caron and Fox (2014), have been mostly studied in the context of graphs without multiple edges. The ability to handle multiple edges is an important practical feature of edge exchangeable models, as it eliminates the common practice of arbitrarily thresholding multiple edges to a single edge. The fact that many network datasets are obtained by thresholding edge multiplicities is often ignored during data analysis, and the significance of this action on inference is underexplored in the broader literature.

To demonstrate the effect of thresholding in the edge exchangeable setting, we define the (t, c) -projection of an interaction process $\mathcal{J}: \mathbb{N} \rightarrow \text{fin}(\mathcal{P})$ by $H_{\mathcal{J}}^{t,c}: \text{fin}(\mathcal{P}) \rightarrow \{0, 1\}$, where

$$H_{\mathcal{J}}^{t,c}(A) = \begin{cases} 1, & t(\{i \in \mathbb{N}: \mathcal{J}(i) = A\}) > c, \\ 0, & \text{otherwise,} \end{cases} \quad A \in \text{fin}(\mathcal{P}), \quad (16)$$

for some thresholding function t and cutoff value $c \geq 0$. We define the *standard projection* $H_{\mathcal{J}}^*$ by taking $t(A) = \#A$ and $c = 0$ in (16).

We call $(\mathcal{Y}_n)_{n \geq 1}$ a *binary Hollywood process with parameter* (α, θ) if it follows the Hollywood process with parameter (α, θ, ν) having $\nu_2 = 1$, that is, all edges involve exactly 2 vertices.

Theorem 4.5 (Sparsity of projected network data)—Let $(\mathcal{Y}_n)_{n \geq 1}$ obey the binary Hollywood process with parameter (α, θ) . For each $n \geq 1$, let $H_n^{t,c}$ be the (t, c) -projection of $\mathcal{Y}_n$ by applying (16) for any $c \geq 0$ and $t(A) = \#A$, the cardinality map. Then the sequence $(H_n^{t,c})_{n \geq 1}$ is sparse almost surely for all $0 < \alpha < 1$ and $\theta > -\alpha$.

Theorem 4.5 makes clear that projecting not only discards information but may also alter the assumed behavior of the data. By Theorems 4.4 and 4.5, $(\mathcal{Y}_n)_{n \geq 1}$ generated from the binary Hollywood model is sparse only for $1/2 < \alpha < 1$, while the projected network after applying (16) is sparse for all $0 < \alpha < 1$. The significance of this disparity ought not be overlooked.

On the other hand, Figure 8 suggests that the power law behavior of the Hollywood model, as established in Theorem 4.3, might be preserved under projection. We leave this as an open conjecture. Whether this phenomenon is real or perceived, there remains no logical justification for thresholding interaction data to a simple graph, especially when the interaction counts are available and the available class of edge exchangeable models handles multiple edge counts without issue. We also note that thresholding an edge exchangeable network according to (16) does not preserve edge exchangeability.

5. Inference from edge exchangeable models

5.1. Maximum likelihood estimation

Given edge-labeled network data $\mathcal{Y}_n$ with n edges, the log-likelihood $\ell(\alpha, \theta, \nu; \mathcal{Y}_n)$ based on (10) with parameter (α, θ, ν) satisfies

$$\begin{aligned} \ell(\alpha, \theta, \nu; \mathcal{Y}_n) = & \sum_{k=1}^{\infty} M_k(\mathcal{Y}_n) \log \nu_k + \nu(\mathcal{Y}_n) \log(\alpha) + \sum_{j=0}^{\nu(\mathcal{Y}_n)-1} \log(\theta/\alpha + j) - \\ & - \sum_{j=1}^{m(\mathcal{Y}_n)} \log(\theta + j - 1) + \sum_{k=2}^{\infty} N_k(\mathcal{Y}_n) \sum_{j=0}^{k-2} \log(1 - \alpha + j), \end{aligned} \quad (17)$$

where $M_k(\mathcal{Y}_n)$ is the number of edges in $\mathcal{Y}_n$ with exactly k vertices, $N_k(\mathcal{Y}_n)$ is the number of vertices in $\mathcal{Y}_n$ with degree k , and $m(\mathcal{Y}_n) = \sum_{k=1}^{\infty} k M_k(\mathcal{Y}_n)$ is the total degree of $\mathcal{Y}_n$.

Maximum likelihood estimation of ν returns the empirical distribution $\hat{\nu}_{MLE} = \{\hat{\nu}_k\}_{k \geq 1}$, where $\hat{\nu}_k = M_k(\mathcal{Y}_n)/n$ for each $k = 1, 2, \dots$. The partial derivatives of the log-likelihood with respect to α and θ are

$$\frac{\partial \ell(\alpha, \theta, \nu; \mathcal{Y}_n)}{\partial \alpha} = \frac{\nu(\mathcal{Y}_n)}{\alpha} + \sum_{j=0}^{\nu(\mathcal{Y}_n)-1} \frac{-\theta/\alpha^2}{\theta/\alpha + j} - \sum_{k=2}^{\infty} \sum_{j=0}^{k-2} \frac{N_k(\mathcal{Y}_n)}{1-\alpha+j} \quad \text{and} \quad (18)$$

$$\frac{\partial \ell(\alpha, \theta, \nu; \mathcal{Y}_n)}{\partial \theta} = \sum_{j=0}^{\nu(\mathcal{Y}_n)-1} \frac{1/\alpha}{\theta/\alpha + j} - \sum_{j=0}^{m(\mathcal{Y}_n)-1} \frac{1}{\theta + j}. \quad (19)$$

Given the gradient, we then estimate α and θ via a quasi-Newton method. Although the data takes values in a discrete space of edge-labeled networks, the likelihood surface of the Hollywood model is smooth when confined to either part of the parameter space, that is, α is bounded away from 0. See the supplementary material (Crane and Dempsey, 2017, Appendix B) for simulation results verifying consistency of maximum likelihood estimation. We stress that the discontinuity in the likelihood at $\alpha = 0$ does not affect our inference, since we confine estimation to one of the two regimes ($\alpha < 0$ or $0 < \alpha < 1$) based on whether the number of vertices in the population is bounded or unbounded, respectively. The boundary case $\alpha = 0$ corresponds to the one-parameter Ewens distribution, for which $\hat{\theta}_{MLE} = \nu(\mathcal{Y}_n)/\log m(\mathcal{Y}_n)$ and $\sqrt{\log(\mu n)}(\hat{\theta}_{MLE} - \theta)$ converges to the standard normal distribution, with $\mu = \sum_{k \geq 1} k\nu_k$ the average edge arity, cf. Pitman (2005, Section 3.3). Since we are mostly concerned with large scale networks that exhibit heavy tailed behavior, the $\alpha = 0$ case does not arise in the applications considered here.

The distribution in (10), and therefore the log-likelihood in (17), applies to the oriented network generated by the Hollywood model with parameter (α, θ, ν) . With $\mathcal{Y}_n^*$ denoting the undirected edge-labeled network obtained by removing the orientations from $\mathcal{Y}_n$, for each $n \geq 1$, we have

$$\text{pr}(\mathcal{Y}_n^* = \mathcal{E}^*; \alpha, \theta, \nu) = \quad (20)$$

$$C(\mathcal{E}^*) \left[\prod_{k \geq 1} \nu_k^{M_k(\mathcal{E}^*)} \right] \alpha^{\nu(\mathcal{E}^*)} \frac{(\theta/\alpha)^{\uparrow \nu(\mathcal{E}^*)}}{\theta^{\uparrow m(\mathcal{E}^*)}} \prod_{k=2}^{\infty} \exp \left\{ N_k(\mathcal{E}^*) \log((1-\alpha)^{\uparrow (k-1)}) \right\},$$

where $C(\mathcal{E}^*)$ is the combinatorial factor that counts the number distinct ways to orient the edges of $\mathcal{E}^*$ to obtain a directed network. The log-likelihood based on $\mathcal{Y}_n^*$ satisfies

$$\tilde{\ell}(\alpha, \theta, \nu; \mathcal{Y}_n^*) = \ell(\alpha, \theta, \nu; \mathcal{Y}_n) + \log C(\mathcal{Y}_n^*),$$

where $\ell(\alpha, \theta, \nu; \mathcal{Y}_n)$ is the log-likelihood from (17) with $\mathcal{Y}_n$ taken to be any oriented edge-labeled network whose edges agree with $\mathcal{Y}_n^*$. Thus, the score functions based on $\mathcal{Y}_n^*$ are the same as for any directed network corresponding to $\mathcal{Y}_n^*$, so that maximum likelihood estimation for (α, θ, ν) can be performed without issue.

Some issues arise, however, when projecting multiple edges according to the operation in (16). For definiteness, suppose H_n^* is the standard projection from the binary Hollywood model with parameter (α, θ) on $\mathfrak{G}_{[n]}$ as in (10). The log-likelihood $\ell^*(\alpha, \theta, \nu; H_n^*)$ in this case satisfies

$$\begin{aligned} \exp \left\{ \ell^*(\alpha, \theta, \nu; H_n^*) \right\} &= \\ &= \sum_{\mathcal{E} \geq H_n^*} \left[\prod_{k \geq 1} \nu_k^{M_k(\mathcal{E})} \right] \alpha^{\nu(\mathcal{E})} \frac{(\theta/\alpha)^{\uparrow \nu(\mathcal{E})}}{\theta^{\uparrow m(\mathcal{E})}} \prod_{k=2}^{\infty} \exp \left\{ N_k(\mathcal{E}) \log((1-\alpha)^{\uparrow (k-1)}) \right\} \\ &= \left[\prod_{k \geq 1} \nu_k^{M_k(H_n^*)} \right] \frac{\alpha^{\nu(H_n^*)} (\theta/\alpha)^{\uparrow \nu(H_n^*)}}{\theta^{\uparrow m(H_n^*)}} \prod_{k=2}^{\infty} \exp \left\{ N_k(H_n^*) \log(1-\alpha)^{\uparrow (k-1)} \right\} \times \\ &\quad \times \sum_{\mathcal{E} \geq H_n^*} \frac{\prod_{k \geq 1} \nu_k^{M_k(\mathcal{E}) - M_k(H_n^*)}}{\theta^{\uparrow (m(\mathcal{E}) - m(H_n^*))}} \exp \left\{ (N_k(\mathcal{E}) - N_k(H_n^*)) \log(1-\alpha)^{\uparrow (k-1)} \right\}, \end{aligned}$$

where $\mathcal{E} \geq H_n^*$ indicates that $\mathcal{E} \in \mathfrak{G}_{[n]}$ is an edge-labeled network whose standard projection is H_n^* . The log-likelihood has the form

$$\ell^*(\alpha, \theta, \nu; H_n^*) = \log C_n^*(H_n^*; \alpha, \theta, \nu) + \ell(\alpha, \theta, \nu; H_n^*), \quad (21)$$

so that without a manageable expression for $C_n^*(H_n^*; \alpha, \theta, \nu)$, maximum likelihood estimation based on projected network data H_n^* is intractable at worst and computationally burdensome at worst.

5.2. Application to Wikipedia voting and Karate Club networks

Table 1 shows the maximum likelihood estimates for the binary Hollywood model fit to the Wikipedia voting and Karate Club networks described in Crane and Dempsey (2017, Section B). Together these examples cover both regimes in the parameter space of the Hollywood model.

The Karate Club network consists of interactions among thirty-four club members, warranting the choice of $k = 34$ when fitting the model with parameters $\alpha < 0$ and $\theta = -34\alpha$. The Wikipedia network, by contrast, has no upper limit on the number of vertices, and so we fit the model under regime $0 < \alpha < 1$ and $\theta > -\alpha$.

The large standard error for maximum likelihood estimates of θ agrees with what is known about estimation of the mutation rate in Ewens's sampling formula (Ewens, 1972): although $\hat{\theta}_{MLE} \rightarrow \theta$ almost surely as the sample size grows, it converges at a rate on the order of $\log(n)$, rendering it inconsistent for practical purposes. The estimate of α is of greater interest in the applications we envision because of its relationship to the power law and sparse behavior, cf. Theorems 4.3 and 4.4. We see from Table 1 that the MLE for α has a much smaller standard error, as is supported by the simulation results of the supplementary material (Crane and Dempsey, 2017, Appendix B). When $\alpha < 0$ and the total number of vertices is taken to be known and finite $k = 1, 2, \dots$, then $\theta = -k\alpha$ and we obtain $\hat{\theta}_{MLE} = k\hat{\alpha}_{MLE}$. In this case, we have $\text{s.e.}(\hat{\theta}_{MLE}) = k \times \text{s.e.}(\hat{\alpha}_{MLE})$, with $\text{s.e.}(\cdot)$ shorthand for the standard error.

By Theorem 4.3, the maximum likelihood estimate $\hat{\alpha}_{MLE} = 0.37$ for the Wikipedia dataset implies an estimated power law of $\hat{\gamma}_{MLE} = 1.37$, which is reasonably close to the estimated power law exponent of $\hat{\gamma}_{YULE} = 1.44$ obtained by fitting the degree distribution of the network to the Yule model,

$$\text{pr}(K = k; \gamma) = (\gamma - 1)\Gamma(k)\Gamma(\gamma)/\Gamma(k + \gamma), \quad k \geq 1, \quad (22)$$

for $\gamma > 1$.

5.3. Prediction using growth dynamics

The growth dynamics of edge exchangeable models by the sequential process of edge addition facilitates predictive inferences in networks generated by a process of repeated interactions. In the setting of Section 4, we can predict the next interaction, based on an observed network $\mathcal{Y}_n$, from the update probabilities in (9) with $(\alpha, \theta, \nu) = (\hat{\alpha}_{MLE}, \hat{\theta}_{MLE}, \hat{\nu}_{MLE})$ given by the maximum likelihood estimates obtained from (17). We note that existing models, which assume vertex sampling or evolution by vertex addition, do not have a mechanism for such prediction.

For a concrete application, we consider the actors collaboration network from Barabási and Albert (1999). For (α, θ, ν) in the parameter space of the Hollywood model, we compute the predictive probability, conditional on $\mathcal{Y}_n$, that the next movie contains an actor not previously observed in the sample. A straightforward calculation based on (9) and the law of total probability gives

$$\text{pr}(\text{new actor in next movie} \mid \mathcal{Y}_n; \alpha, \theta, \nu) = 1 - \sum_{k \geq 1} \nu_k \frac{(M - \alpha N)^{\uparrow k}}{(\theta - M)^{\uparrow k}}, \quad (23)$$

where $N = \sum_{k \geq 1} N_k(\mathcal{Y}_n)$ is the total degree of $\mathcal{Y}_n$ and $M = \sum_{k \geq 1} k N_k(\mathcal{Y}_n)$ is the total number of roles in $\mathcal{Y}_n$. (To see the calculation in (23), we note that $(M - \alpha N)^{\uparrow k} (\theta - M)^{\uparrow k}$ is the conditional probability that the next movie does not feature a new actor, given that it has k actors in its cast. Summing over all $k \geq 1$ gives the total probability that the next movie does feature a new actor, yielding the rightmost term in (23). The probability that a new actor appears follows by taking the complementary probability.)

Fitting the model to the actors collaboration dataset yields $\hat{\alpha}_{MLE} = 0.66$ (s.e. 6.8×10^{-4}) and $\hat{\theta}_{MLE} = 4.21$ (s.e. 2.86), with $\hat{\nu}_{MLE}$ given in Table 2. Since $\hat{\alpha}_{MLE} = 0.66$ lies in the range $1/\mu < \alpha < 1$, with $\mu = 7.136$ computed from Table 2, Theorem 4.4 suggests that the actors collaboration network is sparse. This estimate also agrees exactly with the estimated fit $\hat{\gamma}_{YULE} = 1.66$ if the degree distribution is fit directly to the Yule distribution in (22).

The estimated predictive probability based on these maximum likelihood estimates is 0.78. We check the accuracy of this estimate by data-splitting cross validation based on samples of 2,000 movies from the total collection of about 200,000. For each iteration $j = 1, 2, \dots$, we sample 2,000 movies uniformly at random from the 200,000 to obtain $\mathcal{Y}_{2000}$. We then compute the probability in (23) based on the estimates $\hat{\alpha}_{MLE}$, $\hat{\theta}_{MLE}$, and $\hat{\nu}_{MLE}$ from $\mathcal{Y}_{2000}$ and compare this estimated probability to the empirical probability computed as the proportion of the 198,000 unsampled movies for which there appears an actor not among the 2,000 movies in the sample $\mathcal{Y}_{2000}$. The mean relative error between the fitted probability (23) and the empirical probability for 100 iterations was -0.003 with a standard deviation of 0.002.

5.4. Tests for sparsity and power law

As asymptotic network properties, sparsity and power law cannot be verified with certainty based on any finite amount of data. Statistical tests for sparsity and/or power law based on finite sample data, therefore, require that the finite sample models faithfully represent the properties exhibited by the population network. Our discussion in Section 4.5 advises caution when testing for asymptotic properties based on thresholded network data. If the parameter space Π partitions as $\Pi = \Pi_0 \cup \Pi_1$ such that $\pi \in \Pi_1$ parameterizes sparse networks and $\pi \in \Pi_0$ parameterizes non-sparse networks, then $H_0 : \pi \in \Pi_0$ versus $H_1 : \pi \in \Pi_1$ might yield a valid test for sparsity. But one ought to be aware that alternative interpretations of the partition $\Pi = \Pi_0 \cup \Pi_1$ may provide a more parsimonious conclusion. For example, under the binary Hollywood model of Section 4, the region $1/2 < \alpha < 1$ corresponds to sparse networks while $\alpha < 1/2$ parameterizes dense networks, with $\alpha < 0$ corresponding to the case where the number of vertices stays bounded and finite. By Theorem 4.5, if the network is thresholded by the standard projection, then the projected

network is sparse a.s. for all $\alpha > 0$. In this case, the partition $\Pi = \{\alpha < 0\} \cup \{0 \leq \alpha \leq 1\}$ offers the more parsimonious interpretation according to whether the population of vertices is finite ($\alpha < 0$) or infinite ($0 \leq \alpha \leq 1$) rather than sparse ($0 \leq \alpha \leq 1$) or not sparse ($\alpha < 0$).

A particular consequence of this is on display when testing for sparsity in the US airport dataset (Opsahl, 2011), which is built from the flight map between all US airports in 2010. For each pair of airports, there are as many edges as there were seats on all flights between those airports. The edge weights range over several orders of magnitude from a minimum of 1 to several hundred thousand, making the projection in (16) particularly deleterious to the data structure. For the binary Hollywood model fit to this dataset, there are two possible maximum likelihood estimates, both with log-likelihood of about -8.21×10^{-9} : if $\alpha > 0$ we get $\hat{\alpha}_{MLE,1} = 0.13$ (s.e. 0.003) and $\hat{\theta}_{MLE,1} = 0.08$ (s.e. 0.134), and if $\alpha < 0$ we get $\hat{\alpha}_{MLE,2} = -0.12$ (s.e. 0.003) by taking $k = 1574$ the number of airports in the sample. The estimate $\hat{\alpha}_{MLE,1} = 0.13$ is consistent with $\hat{\gamma}_{YULE} = 1.11$ obtained by fitting (22) to the degree distribution of the airport network. The choice of regime $\alpha < 0$ or $\alpha > 0$ is a matter of whether we believe the population of airports under consideration is fixed and finite or potentially unbounded, a choice which is closely tied to the desired hypothesis test.

To model an infinite population, we set $\alpha > 0$ and we can test for sparsity by considering $H_0 : 0 < \alpha \leq 1/2$ versus $H_1 : 1/2 < \alpha \leq 1$. From the estimated $\hat{\alpha}_{MLE,1} = 0.13$, we cannot reject the hypothesis $H_0 : \alpha \leq 1/2$. But if testing based on the projected network, then the above test is degenerate since the ranges for which the population is infinite and for which the population is sparse coincide, per Theorem 4.5. By contrast, Caron and Fox (2014, Section 7.2) conclude that the US airport network is sparse based on a 99% credible interval of $[0.10, 0.18]$ for an analogous parameter to α in our case, but that analysis appears to be based on the standard projection.²

6. Comparison to other approaches

For completeness we provide some empirical and conceptual comparisons to other common approaches.

6.1. Empirical comparison

To round out the discussion, we compare the performance of the Hollywood model to that of the exponential random graph model (ERGM) (Holland and Leinhardt, 1981), stochastic blockmodel (SBM) (Holland et al., 1983; Snijders and Nowicki, 1997), and the Erdős-Rényi (ER) model fit to the actors collaboration network. Note that in fitting the SBM we are not interested in detecting community structure; rather, we are following Bickel and Chen (2009) and others, who treat the SBM as an approximation to the graphon model.

The properties established above for the Hollywood model, and edge exchangeable models more generally, already indicate certain inevitable benefits over these other approaches. The

²We note here that Caron and Fox (2017) have since revised their conclusion of sparsity for the airport network after the original version of this article was posted.

comparison of edge exchangeable models to these other approaches is made difficult by the fact that edge exchangeability is inherently geared toward aspects of interaction networks that these other models are not. From a conceptual viewpoint, edge exchangeable models are more suitable because they capture the occurrence of multiple edges and observation by edge sampling. To illustrate these benefits, we compare the empirical performance of the Hollywood model (HW), stochastic blockmodel (SBM), Erdős-Rényi model (ER), and exponential random graph model (ERGM) on the actors collaboration dataset (Barabási and Albert, 1999), which consists of approximately 200,000 movies.

We have already fit HW to the full movie dataset in Section 5.3, but we could not compare the HW fit to these other methods on the full dataset because the best available software for SBM³ and ERGM⁴ are computationally inefficient for networks of even moderate size. Thus, for comparison we subsample 250 movies uniformly at random from the database of 200,000 movies. Since SBM, ER, and ERGM are models for graphs without multiple edges, we treat the *de facto* population network as the graph G_{250} induced by these 250 movies by including an edge for every collaboration between a pair of actors. The maximum likelihood estimate (MLE) under SBM estimates 21 blocks and log-likelihood of $-32,818$, and MLE under ER returns a log-likelihood of $-42,924$. It is also unclear how to extend the estimates for SBM beyond the network to which it has been fitted, as the number of blocks will likely increase as the network grows but the model possesses no natural way to incorporate this growth.

The ERGM with sufficient statistics given by edge and triangle density, respectively,

$$\epsilon(G) = \frac{2}{v(G)(v(G)-1)}e(G) \text{ and } \tau(G) = \frac{6}{v(G)(v(G)-1)(v(G)-2)}\Delta(G),$$

where $\Delta(G)$ is the number of triangles in G , returns a model degeneracy error when fit to G_{250} , meaning it gives no better fit than ER, which corresponds to ERGM with sufficient statistic $\epsilon(G)$. (Our choice of these sufficient statistics in the ERGM is meant to highlight the ability of ERGM for modeling non-local structure, by counting triangles. While there are many other possible non-local sufficient statistics, flexibility in this choice is limited by computational issues of ERGMs.) To avoid the model degeneracy, we also fit ERGM to the induced graph obtained from a 1,000 movie subsample. In this case, the fitted parameters are $\hat{\beta}_\epsilon = -12.26$ (s.e. 0.083) and $\hat{\beta}_\tau = 5.24$ (s.e. 0.019) with log-likelihood of $-190,630$. For comparison, we fit the Hollywood model to both the 250 and 1,000 movie subsamples and obtain a log-likelihood of -476 and $-26,320$, respectively.

Figure 9 compares the goodness of fit for the degree distribution for HW, SBM, ER, and ERGM. The fitted degree distribution for HW in (A) is based on the estimated parameters obtained from the sample of 250 movies compared to the degree distribution of the complete dataset of 200,000 movies. In (B) and (C), we compare the fitted ER and SBM to the degree distribution of subsampled network G_{250} and in (D) we compare the fitted ERGM to the

³<https://cran.r-project.org/web/packages/blockmodels/blockmodels.pdf>

⁴<https://cran.r-project.org/web/packages/ergm/ergm.pdf>

degree distribution of G_{1000} based on 1,000 movies. Note that this comparison is generous to SBM, ER, and ERGM since those models are compared only to the data to which they were fit while HW is evaluated based on how well it detects the population structure for the full dataset based on estimates from only 250 movies. This explains the difference in the empirical degree distributions across the four panels. While we would like to compare all of these models against the same dataset, we stress that this is not possible due to computational issues with SBM and ERGM. It is also not insightful as comparing the fitted ER, SBM, and ERGM to the full 200,000 movie dataset would show a very poor fit because the degree distributions from those models are known to produce degree distributions that are strongly peaked about a mean value; they do not in any way resemble the heavy tailed degree distribution of the actors network. Even with this disadvantage in comparison, the Hollywood model gives a better fit, with a clearer interpretation, than ER and SBM, while the fit under ERGM is very close to the observed degree distribution and shows strong evidence of overfitting. Finally, HW converged to the MLE without problem while SBM and ERGM suffered convergence issues and long run times.

6.2. Further discussion

We conclude the discussion by commenting on a few other approaches which have been proposed to account for sparsity and power law degree distributions in statistical network models.

6.2.1. Graphon models—Graphon models grow out of the theory of dense graph limits (Lovasz and Szegedy, 2006) and are intimately related to the Aldous-Hoover theory of partially exchangeable arrays (Aldous, 1981; Hoover, 1979). As we mentioned in Section 1, traditional exchangeable random graph models with vertex set $\mathbb{N}$ cannot reproduce sparsity.

Probabilists often address this issue by studying the graphon under a *sparse regime*, an approach brought into the statistics literature by Bickel and Chen (2009). With $\phi : [0, 1] \times [0, 1] \rightarrow [0, 1]$ symmetric in its arguments and $\{\rho_n\}_{n=1}^\infty$ satisfying $\rho_n \rightarrow \infty$ and $\rho_n^{-1} n \int_{[0,1] \times [0,1]} \phi(u, v) du dv = O(1)$ as $n \rightarrow \infty$, Bickel and Chen (2009) define the distribution of a graph $G_n = ([n], E_n)$ with vertices labeled in $[n]$ by taking $U_1, \dots, U_n$ i.i.d. Uniform $[0, 1]$ and putting

$$\text{pr}(\{i, j\} \in E_n \mid U_1, \dots, U_n) = \rho_n^{-1} \phi(U_i, U_j), \quad 1 \leq i < j \leq n, \quad (24)$$

conditionally independently for all $1 \leq i < j \leq n$. Under (24), every G_n is exchangeable, and collectively the sequence $(G_n)_{n=1,2,\dots}$ is sparse in the sense of Definition 2.4. But the formulation does not automatically correspond to a generating process for a population network since the marginal distributions in (24) are not logically related through any specified sampling mechanism.

Without a direct link between samples of different size, the meaning of the parameter ϕ varies with sample size and there is no logical way to relate parameter estimates $\hat{\phi}_n$ based on

a sample of size n to a single parameter ϕ for the population. In this case, the interpretation of estimated parameters is obscured and the significance of asymptotic statements for finite sample approximations, as derived, for example, by Borgs et al. (2015) and Wolfe and Olhede (2014), is unclear. The overall importance of accounting for the sampling scheme in network modeling has been emphasized in other recent work by Crane and Dempsey (2015).

6.2.2. Completely random measures and graphex models—Caron and Fox (2014), and later Veitch and Roy (2015) and Borgs et al (2016), studied a class of network models derived from exchangeable point processes $\mathbf{X}$ on $[0, \infty) \times [0, \infty)$. In this setting, each $(x, x') \in \mathbf{X}$ determines an edge between vertices corresponding to points x, x' . The point process $\mathbf{X}$ is assumed to be *exchangeable* in the sense that its distribution is invariant under the joint action on $[0, \infty) \times [0, \infty)$ by measure preserving transformations of $[0, \infty)$. (Note that the term *exchangeability* here is used rather loosely, as it corresponds to invariance under measure-preserving transformations of the plane as opposed to invariance with respect to the action of the symmetric group (i.e., relabeling of vertices or edges) as the term is customarily employed.)

In this context, a sequence of network data is obtained by defining a graph G_t for each $t \geq 0$, where G_t is derived from the restriction of $\mathbf{X}$ to $[0, t] \times [0, t]$ by only including in G_t those vertices that are labeled in $[0, t]$ and which are not isolated in the restriction. Veitch and Roy (2015) adapt a theorem of Kallenberg (1990) to obtain a generic *graphex representation* of the class of random graphs derived from exchangeable point processes in this way.

Graphex models allow for sparse, power law behavior, and their description in terms of the exchangeable point process $\mathbf{X}$ invokes some invariance principle in the construction. It is important to stress, however, that the notion of exchangeability applies to the generating point process $\mathbf{X}$, and not directly to the associated process of graphs $(G_t)_{t \geq 0}$. In fact, the graphs $(G_t)_{t \geq 0}$ induced from the process are not equipped with any labels, and so it is not yet clear what exchangeability of the point process implies for the induced graph sequence, or how the construction of G_t by restricting to $[0, t]$ and removing isolated vertices relates to the manner in which real world networks are observed.

7. Concluding remarks

Here we have laid out a new approach to analyzing interaction networks with edge exchangeable models. The approach put forward here better captures the manner in which certain network datasets are sampled and also presents several theoretical and practical benefits over other common network models. The observation that the interactions, and therefore edges, are the units leads naturally to the alternative representation by edge-labeled networks. The Hollywood model, in particular, is computationally tractable, performs admirably in several real data examples, and replicates the key features of sparsity and power law degree distribution of primary interest in much of the modern network science literature. Whereas the Hollywood model can be fit easily to complete datasets of hundreds of thousands or millions of edges, graphon models, stochastic blockmodels, and the exponential random graph model incur computational issues on networks only a fraction of the size, not to mention their inability to account for basic structural features such as sparsity

and power law degree distribution. The nonparametric vertex components model from Section 4.2 offers an even larger class of models with likely computational benefits due to its convenient stick-breaking construction.

The Hollywood model with parameter in the range $0 < \alpha < 1$ gives rise to networks with power law degree distribution. If α lies in the range $(1/2, 1)$, realizations from the Hollywood model are sparse. This class of models, therefore, offers an alternative to prior approaches for modeling sparse, scale-free networks, such as preferential attachment models (Barabási and Albert, 1999), sparse graphon models (Bickel and Chen, 2009; Borgs et al, 2014), and approaches using completely random measures (Caron and Fox, 2014; Veitch and Roy, 2015; Borgs et al, 2016). While these other approaches may be appropriate for certain network datasets, the edge exchangeable setting seems most suitable for interaction networks, which account for an appreciable proportion of modern network datasets.

8. Proof of main theorems

We equip $\mathfrak{G}_{\mathbb{N}}$ with the topology and Borel σ -field induced by the metric

$$d(\mathcal{E}, \mathcal{E}') = 1/(1 + \sup\{n \in \mathbb{N} : \mathcal{E}|_{[n]} \neq \mathcal{E}'|_{[n]}\}), \quad \mathcal{E}, \mathcal{E}' \in \mathfrak{G}_{\mathbb{N}},$$

with the convention $1/\infty = 0$.

8.1. Proof of Theorem 4.2

Let $(\mathcal{Y}_n)_{n \in \mathbb{N}}$ be a realization of the Hollywood process with parameter (α, θ, ν) from Section 4. Then the distribution of $\mathcal{Y}_n$ is given by (10) for each $n \in \mathbb{N}$, and these distributions are consistent with respect to the restriction operation by consequence of the sequential construction in (9). The distribution in (10) depends on $\mathcal{Y}_n$ only through $\nu(\mathcal{Y}_n)$, $N_k(\mathcal{Y}_n)$, and $M_k(\mathcal{Y}_n)$, that is, the number of vertices, number of vertices of degree $k-1$, and the number of k -ary edges for each $k-1$, all of which are invariant under relabeling edges. Edge exchangeability follows.

8.2. Proof of Theorem 4.3

Let $(\mathcal{Y}_n)_{n \geq 1}$ be a realization from the Hollywood process with parameter (α, θ, ν) . The power law behavior is apparent by the following connection between $(\mathcal{Y}_n)_{n \geq 1}$ and the two parameter Ewens-Pitman distribution. In particular, the Ewens-Pitman distribution corresponds to the special case of the Hollywood model with $0 < \alpha < 1$, $\theta > -\alpha$, and $\nu_1 = 1$. In this case, there is exactly one vertex incident to every edge and the network data corresponds to a partition of the edges. The number of blocks in the Ewens-Pitman partition corresponds to the number of vertices in the unary Hollywood process. Let $N_k(\mathcal{Y}_n)$ be the number of vertices with degree k in $\mathcal{Y}_n$ and let $\nu(\mathcal{Y}_n)$ be the total number of vertices in $\mathcal{Y}_n$. By Pitman (2005, Lemma 3.11), we know that $N_k(\mathcal{Y}_n)/\nu(\mathcal{Y}_n) \rightarrow \alpha(1-\alpha)^{\uparrow(k-1)}/k!$ a.s. for

every $k \geq 1$ as $n \rightarrow \infty$. Thus, the sequence of degree distributions

$\{d_n\}_{n \geq 1} = \{(N_k(\mathcal{Y}_n)/v(\mathcal{Y}_n))_{k \geq 1}\}_{n \geq 1}$ converges a.s. in the total variation topology to the distribution given by $p_\alpha(k) = \alpha(1-\alpha)^{\uparrow(k-1)}/k!$.

In the general Hollywood process with arbitrary distribution ν , the degree sequences of $(\mathcal{Y}_n)_{n \geq 1}$ coincide with a random subsequence of $\{d_n\}_{n \geq 1}$ indexed by $\{K_r\}_{r \geq 1}$ for

$K_r = \sum_{j=1}^r \kappa_j$, $r = 1, 2, \dots$, where $\kappa_1, \kappa_2, \dots$ are i.i.d. from ν . Thus, the degree distributions of $(\mathcal{Y}_n)_{n \geq 1}$ correspond to $\{d_{K_r}\}_{r \geq 1}$, which is a subsequence of the a.s. converging

sequence $\{d_n\}_{n \geq 1}$ and, therefore, must have the same a.s. limit. The proof is completed by noting that $\alpha(1-\alpha)^{\uparrow(k-1)}/k! \sim \alpha k^{-(\alpha-1)}/\Gamma(1-\alpha)$ as $k \rightarrow \infty$, which corresponds to the power law with exponent $\alpha + 1$.

8.3. Proof of Theorem 4.4

We once again exploit the connection to the Ewens-Pitman distribution from the proof of Theorem 4.3. Let N_n be the number of vertices in the unary Hollywood process with parameter (α, θ) satisfying $0 < \alpha < 1$ and $\theta > -\alpha$. Pitman (2005, Theorem 3.8) shows that $n^{-\alpha}N_n \rightarrow S_\alpha$ a.s., where S_α is a strictly positive and finite random variable. The sequence $\{N_n\}_{n \geq 1}$ then satisfies $N_n \sim n^\alpha S_\alpha$ a.s. as $n \rightarrow \infty$, so that

$$E(N_n) \sim \frac{\Gamma(\theta+1)}{\alpha\Gamma(\theta+\alpha)} n^\alpha \text{ as } n \rightarrow \infty,$$

by Pitman (2005, Theorem 3.8).

Let $(\mathcal{Y}_n)_{n \geq 1}$ be the Hollywood process with parameter (α, θ, ν) for arbitrary distribution ν .

Then $(v(\mathcal{Y}_n))_{n \geq 1}$ is a random subsequence of $\{N_n\}_{n \geq 1}$ given by $(N_{K_r})_{r \geq 1}$, where

$K_r = \sum_{j=1}^r \kappa_j$ for $\kappa_1, \kappa_2, \dots$ i.i.d. from ν . It follows from the above argument that

$v(\mathcal{Y}_n) = N_{K_n} \sim K_n^\alpha S_\alpha$ a.s. as $n \rightarrow \infty$. By the strong law of large numbers, $n^{-1}K_n \sim \mu$ a.s. and

$K_n^\alpha \sim (\mu n)^\alpha$ a.s. as $n \rightarrow \infty$, where $\mu = \sum_{k \geq 1} k\nu_k$; whence,

$$E(v(\mathcal{Y}_n)) \sim \frac{\Gamma(\theta+1)}{\alpha\Gamma(\theta+\alpha)} (\mu n)^\alpha \text{ as } n \rightarrow \infty$$

To establish sparsity of $(\mathcal{Y}_n)_{n \geq 1}$, we consider $(n^{-1}v(\mathcal{Y}_n)^{m_\bullet(\mathcal{Y}_n)})_{n \geq 1}$, where $m_\bullet(\mathcal{Y}_n)$ is the average total degree in $\mathcal{Y}_n$. We must identify values of $0 < \alpha < 1$ for which

$\liminf_{n \rightarrow \infty} n^{-1}v(\mathcal{Y}_n)^{m_\bullet(\mathcal{Y}_n)} = +\infty$. By the strong law of large numbers, $m_\bullet(\mathcal{Y}_n) \rightarrow \mu$ a.s. as $n \rightarrow \infty$. By the above discussion, $v(\mathcal{Y}_n) \sim (\mu n)^\alpha S_\alpha$ a.s. as $n \rightarrow \infty$ for a strictly positive and

finite random variable S_α . It follows that $n^{-1}v(\mathcal{Y}_n)^{m \bullet (\mathcal{Y}_n)} \sim n^{-1}(\mu n)^{\mu\alpha} S_\alpha$ a.s. as $n \rightarrow \infty$, which goes to infinity as long as $\mu\alpha > 1$. Thus, $(\mathcal{Y}_n)_{n \geq 1}$ is sparse with probability 1 provided $1/\mu < \alpha < 1$, for all $\theta > -\alpha$.

8.4. Proof of Theorem 4.5

Let $(\mathcal{Y}_n)_{n \geq 1}$ be the binary Hollywood process with parameter $0 < \alpha < 1$ and $\theta > -\alpha$. For each $n \geq 1$, let $H_n^{t,c}$ be the projection obtained by applying (16) to $\mathcal{Y}_n$. To establish sparsity of $(H_n^{t,c})_{n \geq 1}$, we must show

$$\limsup_{n \rightarrow \infty} \frac{e(H_n^{t,c})}{v(H_n^{t,c})^2} = 0 \text{ a.s.},$$

for which it suffices to establish sparsity under the standard projection H_n^* , since

$$e(H_n^{t,c}) \leq e(H_n^*) \text{ and } v(H_n^{t,c}) = v(H_n^*).$$

For each $k \geq 1$ and $n \geq 1$, we write $N_k(\mathcal{Y}_n)$ and $N_k(H_n^*)$ to denote the number of vertices with degree k in $\mathcal{Y}_n$ and H_n^* , respectively. We also write $e(\mathcal{Y}_n)$ and $e(H_n^*)$ to denote the number of edges in $\mathcal{Y}_n$ and H_n^* , respectively, and $v(\mathcal{Y}_n) = v(H_n^*)$ to denote the number of vertices in $\mathcal{Y}_n$ and H_n^* . Since the projection operation reduces multiple occurrences of the same edge to a single edge, the degree of each vertex in H_n^* can be no larger than $v(H_n^*)$ and

$$e(H_n^*) = \sum_{k \geq 1} k N_k(H_n^*) \leq \sum_{k \geq 1} (k \wedge v(H_n^*)) N_k(\mathcal{Y}_n). \quad (25)$$

For every $K \geq 1$, (25) implies

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{e(H_n^*)}{v(H_n^*)^2} &\leq \limsup_{n \rightarrow \infty} \sum_{k=1}^K \frac{k \wedge v(H_n^*)}{v(H_n^*)} \frac{N_k(\mathcal{Y}_n)}{v(H_n^*)} + \limsup_{n \rightarrow \infty} \sum_{k=K+1}^{\infty} \frac{N_k(\mathcal{Y}_n)}{v(H_n^*)} \\ &\leq \limsup_{n \rightarrow \infty} \sum_{k=1}^K \frac{k \wedge v(H_n^*)}{v(H_n^*)} + \limsup_{n \rightarrow \infty} \sum_{k=K+1}^{\infty} \frac{N_k(\mathcal{Y}_n)}{v(H_n^*)} \\ &\leq \sum_{k=1}^K \limsup_{n \rightarrow \infty} \frac{k \wedge v(H_n^*)}{v(H_n^*)} + \limsup_{n \rightarrow \infty} \sum_{k=K+1}^{\infty} \frac{N_k(\mathcal{Y}_n)}{v(H_n^*)}. \end{aligned} \quad (26)$$

Pitman (2005, Corollary 3.9) implies that $n^{-\alpha}v(H_n^*) \rightarrow S_\alpha$ a.s., where S_α is a strictly positive, finite random variable; thus, $v(H_n^*) \rightarrow \infty$ a.s. and $\limsup_{n \rightarrow \infty} (k \wedge v(H_n^*))/v(H_n^*) \rightarrow 0$ a.s. for every $k = 1, \dots, K$, so that (26) implies

$$\limsup_{n \rightarrow \infty} \frac{e(H_n^*)}{v(H_n^*)^2} \leq \limsup_{n \rightarrow \infty} \sum_{k=K+1}^{\infty} \frac{N_k(\mathcal{Y}_n)}{v(H_n^*)} \text{ for all } K \geq 1. \quad (27)$$

By Pitman (2005, Lemma 3.11),

$$\lim_{n \rightarrow \infty} \frac{N_k(\mathcal{Y}_n)}{v(H_n^*)} p_\alpha(k) = \alpha \Gamma(k - \alpha) / (k! \Gamma(1 - \alpha)) \text{ for every } k \geq 1 \text{ a.s.},$$

implying that for every $\varepsilon, \delta > 0$ and $k \geq 1$ there exists $R = R(\varepsilon, \delta, k)$ such that

$$\Pr\left(\left|v(H_n^*)^{-1} N_k(\mathcal{Y}_n) - p_\alpha(k)\right| < \varepsilon \text{ for all } n > R\right) \geq 1 - \delta.$$

for any $K \geq 1$, we choose $R^* = \max_{1 \leq k \leq K} R(\varepsilon/K, \delta/K, k)$ so that

$$\Pr\left(\left|v(H_n^*)^{-1} N_k(\mathcal{Y}_n) - p_\alpha(k)\right| < \varepsilon/K \text{ for all } n > R^*, \text{ for all } k = 1, \dots, K\right) \geq 1 - \delta$$

and

$$\Pr\left(\left|\sum_{k=K+1}^{\infty} v(H_n^*)^{-1} N_k(\mathcal{Y}_n) - \sum_{k=K+1}^{\infty} p_\alpha(k)\right| < \varepsilon \text{ for all } n > R^*\right) \geq 1 - \delta.$$

We combine this with (27) and the tail calculation

$$p_\alpha(> K) = \sum_{k=K+1}^{\infty} p_\alpha(k) = \Gamma(K+1)\Gamma(1-\alpha) \text{ to observe}$$

$$\begin{aligned} & \Pr(\limsup_{n \rightarrow \infty} e(H_n^*)/v(H_n^*)^2 > p_\alpha(> K) + \varepsilon) \leq \\ & \leq \Pr\left(\limsup_{n \rightarrow \infty} \sum_{k=K+1}^{\infty} N_k(\mathcal{Y}_n)/v(H_n^*) > p_\alpha(> K) + \varepsilon\right) \\ & \leq \Pr\left(\left|\sum_{k=K+1}^{\infty} v(H_n^*)^{-1} N_k(\mathcal{Y}_n) - \sum_{k=K+1}^{\infty} p_\alpha(k)\right| > \varepsilon \text{ for some } n > R^*\right) \\ & \leq \delta \end{aligned}$$

for every $K \geq 1$. Since $p_\alpha(> K) \downarrow 0$ as $K \rightarrow \infty$, it follows that

$$\Pr(\limsup_{n \rightarrow \infty} e(G_n^*)/v(H_n^*)^2 > \varepsilon) \leq \delta \text{ for all } \varepsilon, \delta > 0$$

and $(H_n^*)_{n \geq 1}$ is sparse a.s.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

H. Crane is partially supported by NSF grants CNS-1523785 and CAREER DMS-1554092.

References

- Airoldi EM, Choi DS, Wolfe PJ. Confidence sets for network structure. *Statistical Analysis and Data Mining*. 2011; 4(5):461–469.
- Aldous DJ. Representations for partially exchangeable arrays of random variables. *J Multivariate Anal*. 1981; 11(4):581–598. URL . DOI: 10.1016/0047-259X(81)90099-3 [http://dx.doi.org/10.1016/0047-259X\(81\)90099-3](http://dx.doi.org/10.1016/0047-259X(81)90099-3)
- Barabási AL, Albert R. Emergence of scaling in random networks. *Science*. 1999; 286(5439):509–512. URL . DOI: 10.1126/science.286.5439.509 <http://dx.doi.org/10.1126/science.286.5439.509> [PubMed: 10521342]
- Bickel P, Chen A. A nonparametric view of network models and Newman–Girvan and other modularities. *Proceedings of the National Academy of Sciences of the United States of America*. 2009; 106(50):21068–21073. [PubMed: 19934050]
- Borgs C, Chayes JT, Cohn H, Ganguly S. Consistent nonparametric estimation for heavy-tailed sparse graphs. *arXiv:1508.06675*. 2015
- Borgs C, Chayes J, Cohn H, Zhao Y. An L^p theory of sparse graph convergence I: limits, sparse random graph models, and power law distributions. accessed at. *arXiv:1401.2906*. 2014
- Borgs C, Chayes JT, Cohn H, Holden N. Sparse exchangeable graphs and their limits via graphon processes. *arXiv:1601.07134*. 2016
- Britton T, Deijfen M, Lagerås AN, Lindholm M. Epidemics on random graphs with tunable clustering. *J Appl Probab*. 2008; 45(3):743–756.
- Butts CT. A relational event framework for social action. *Sociol Methodol*. 2008; 38:155–200.
- Caron F, Fox EB. Sparse graphs using exchangeable random measures. Accessed at. *arXiv:1401.1137v3*. 2014
- Caron F, Fox EB. Sparse graphs using exchangeable random measures. *Journal of the Royal Statistical Society, Series B*. 2017; 79(5):1–44.
- Chatterjee S, Diaconis P. Estimating and Understanding Exponential Random Graph Models. *Annals of Statistics*. 2013; 41(5):2428–2461.
- Chung F, Lu L. Complex graphs and networks, volume 107 of CBMS Regional Conference Series in Mathematics. Published for the Conference Board of the Mathematical Sciences; Washington, DC. 2006.
- Crane H. The ubiquitous Ewens sampling formula (with discussion and a rejoinder by the author). *Statistical Science*. 2016; 31(1):1–39.
- Crane H. Rejoinder: The ubiquitous Ewens sampling formula. *Statistical Science*. 2016b; 31(1):37–39.
- Crane H, Dempsey W. A framework for statistical network modeling. *arXiv:1509.08185*. 2015
- Crane H, Dempsey W. Supplementary material for edge exchangeable models for interaction networks. 2017

- Deijfen M, Kets W. Random intersection graphs with tunable degree distribution and clustering. *Probability in the Engineering and Informational Sciences*. 2009
- Ewens WJ. The sampling theory of selectively neutral alleles. *Theoret Population Biology*. 1972; 3:87–112.
- Feng S. *Probability and its Applications*. Springer-Verlag; Berlin: 2010. The Poisson-Dirichlet Distribution and Related Topics.
- Ferguson T. A Bayesian Analysis of Some Nonparametric Problems. *The Annals of Statistics*. 1973; 1(2):209–230.
- Fill J, Scheinerman E, Singer-Cohen K. Random intersection graphs when $m = \omega(n)$: an equivalence theorem relating the evolution of the $G(n, m, p)$ and $G(n, p)$ models. *Random Structures & Algorithms*. 2000; 16:156–176.
- Gao C, Lu Y, Zhou HH. Rate optimal graphon estimation. *Annals of Statistics*. 2015; 43(6):2624–2652.
- Gao C, Ma Z, Zhang AY, Zhou HH. Community Detection in Degree-Corrected Block Models. arXiv: 1607.06993. 2016
- Godehardt E, Jaworski J. Two Models of Random Intersection Graphs for Classification. In: Schwaiger M, Opitz O, editors *Exploratory Data Analysis in Empirical Research Studies in Classification, Data Analysis, and Knowledge Organization*. Springer; Berlin, Heidelberg: 2003.
- Holland PW, Leinhardt S. An exponential family of probability distributions for directed graphs. *Journal of the American Statistical Association*. 1981:33–65.
- Holland PW, Laskey KB, Leinhardt S. Stochastic blockmodels: First steps. *Social Networks*. 1983; 5(2):109–137.
- Hoover DN. *Relations on Probability Spaces and Arrays of Random Variables*. Preprint, Institute for Advanced Studies. 1979
- Ishwaran H, James LF. Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association*. 2001; 96:161–173.
- Kallenberg O. Exchangeable random measures in the plane. *Journal of Theoretical Probability*. 1990; 3(1):81–136.
- Karónski M, Scheinerman E, Singer-Cohen K. On random intersection graphs: the subgraphs problem. *Combinatorics Probability & Computing*. 1999; 8:131–159.
- Karrer B, Newman MEJ. Stochastic blockmodels and community structure in networks. *Physical Review E*. 2011; 83:016107.
- Kivelä M, Arenas A, Barthélemy M, Gleeson JP, Moreno Y, Porter MA. Multilayer Networks. *Journal of Complex Networks*. 2014; 2(3):203–271.
- Klimt B, Yang Y. Introducing the Enron corpus. CEAS. 2004
- Klopp O, Tsybakov AB, Verzelen N. Oracle inequalities for network models and sparse graphon estimation. arXiv:1507.04118. 2016
- Kolaczyk ED. *Statistical analysis of network data*. Springer; New York: 2009.
- Latouche P, Robin S, Ouadah S. Goodness of fit of logistic models for random graphs. arXiv: 1508.00286. 2015
- Leskovec J, Huttenlocher D, Kleinberg J. Signed networks in social media. CHI. 2010
- Lovász L, Szegedy B. Limits of dense graph sequences. *J Comb Th B*. 2006; 96:933–957.
- Mariadassou M, Robin S, Vacher C. Uncovering latent structure in valued graphs: a variational approach. *Annals of Applied Statistics*. 2010:715–742.
- McAuley JJ, Leskovec J. Learning to discover social circles in ego networks. *Neural Information Processing Systems*. 2012:539–547.
- Nesetril J, Ossona de Mendez P. *Sparsity: Graphs, Structures, and Algorithms*. Springer; Heidelberg: 2012.
- Newman MEJ. The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*. 2001; 98:404–409.
- Opsahl T. Why Anchorage is not (that) important: Binary ties and Sample selection. 2011. Available at <http://wp.me/poFcYVw>

- Opsahl T, Panzarasa P. Clustering in weighted networks. *Social Networks*. 2009; 31:155–163.
- Orbanz P, Roy DM. Bayesian Models of Graphs, Arrays and Other Exchangeable Random Structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2015; 37(2):437–461. [PubMed: 26353253]
- Perry PO, Wolfe PJ. Point process modelling for directed interaction networks. *Journal of the Royal Statistical Society: Series B*. 2013; 75(5):821–849.
- Pitman J. Combinatorial stochastic processes, volume 1875 of *Lecture Notes in Mathematics*. 2005
- Rossi RA, Ahmed NK. The network data repository with interactive graph analytics and visualization. *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*. 2015. URL <http://networkrepository.com>
- Sethuraman J. A constructive definition of Dirichlet priors. *Statistica Sinica*. 1994; 4:639–650.
- Shalizi CR, Rinaldo A. Consistency under subsampling of exponential random graph models. *Annals of Statistics*. 2013; 41:508–535. [PubMed: 26166910]
- Singer K. PhD thesis. Johns Hopkins University; 1995. Random intersection graphs.
- Snijders TAB, Nowicki K. Estimation and prediction for stochastic blockmodels for graphs with latent block structure. *Journal of Classification*. 1997; 14(1):75–100.
- Sweet TM. Incorporating Covariates Into Stochastic Blockmodels. *Journal of Educational and Behavioral Statistics*. 2015 1076998615606110.
- Tallberg C. A Bayesian approach to modeling stochastic block-structures with covariates. *Journal of Mathematical Sociology*. 2004; 29(1):1–23.
- Veitch V, Roy D. The class of random graphs arising from exchangeable random measures. arXiv: 1512.03099. 2015. accessed at <http://arxiv.org/pdf/1512.03099v1.pdf>
- Wolfe PJ, Olhede SC. Nonparametric graphon estimation. Available at arXiv:1309.5936. 2014
- Zachary WW. An information flow model for conflict and fission in small groups. *Journal of Anthropological Research*. 1977; 33:452–473.
- Zhang Y, Levina E, Zhu J. Community detection in networks with node features. arXiv:1509.01173. 2015
- Zhao Y, Levina E, Zhu J. On consistency of community detection in networks. *Annals of Statistics*. 2011; 40(4):2266–2292.

Figure 1.

(a) A network structure derived from some process of interactions. Neither vertices nor edges come equipped with labels. (b) Network obtained by labeling vertices of network in Panel (a). (c) Network obtained by labeling edges of network in Panel (a).

Figure 2.
Network depiction of interaction sequence given in (1).

Figure 3.
 (Left) Network representations of interactions processes exhibiting the same interaction structure. (Right) Edge-labeled graph corresponding to the class of all interaction processes with the same structure as defined formally in (4).

Figure 4.

Relabeling and restriction operations for edge-labeled networks. Panel (a) shows a network with edges labeled $1, \dots, 6$. Panel (b) shows the network from (a) with edges relabeled according to permutation $(1354)(26)$. Panel (c) shows the restriction of the network in (a) to the edges labeled in $1, \dots, 4$.

Figure 5.

Two edge labelings of the network data from Figure 1. An edge exchangeable model assigns equal probability to both networks. Vertices are not labeled and, therefore, vertex labels play no role in the edge exchangeability condition.

Figure 6.

Graphical depiction of interaction processes from (1) (left) and (6) (right) after removing edge labels. These graphs would have equal probability under a vertex exchangeable model, but generally may have different probabilities of occurrence if obtained by random edge sampling.

Figure 7.

Illustration of the generic generating process for edge exchangeable networks described in (7) and Theorem 3.2. (a) Network representation of the interaction process $X_1 = \{2, 4\}$, $X_2 = \{1, 2\}$, $X_3 = \{1, 5\}$, $X_4 = \{6, 9\}$, $X_5 = \{2, 6\}$, $X_6 = \{2, 6\}$. (b) Edge-labeled network obtained by removing vertex labels from the network in Panel (a).

Figure 8.

Simulated degree distribution of networks and their standard projection. Plots show the degree distribution of network obtained from (A1) the binary Hollywood model with $(\alpha, \theta) = (0.67, 1)$, (B1) the binary Hollywood model with $(\alpha, \theta) = (0.25, 1)$, (A2) the standard projection of the network in (A1), and (B2) the standard projection of the network in (B1). The line with slope $-\gamma$ in (A1) and (B1) indicates the true power law based on Theorem 4.3. The line with slope $-\gamma$ in (A2) and (B2) is the conjectured power law based on visual evidence and the connection to Theorem 4.3.

Figure 9.

Comparison of fitted degree distribution (---) to empirical degree distribution (*) for (A) Hollywood model (compared to full dataset), (B) Erdős-Rényi model and (C) stochastic blockmodel (compared to 250 movie sample), and (D) ERGM (compared to 1,000 movie sample).

Table 1

Maximum likelihood estimates and standard errors for (α, θ) from binary Hollywood model fit to Wikipedia voting and Karate Club networks.

network	Maximum likelihood estimates			
	$\hat{\alpha}_{MLE}$	std. error*	$\hat{\theta}_{MLE}$	std. error*
Wikipedia	0.37	0.01	183	14.76
Karate Club	-1.80	0.47	61.3	16.04

* Standard errors estimated by the Cramér–Rao lower bound, which we have verified as a good estimate for the standard error based on simulation; see the supplementary material (Crane and Dempsey, 2017, Appendix B).

Table 2

Maximum likelihood estimates of the movie size distribution for the actors collaboration network. The mean number of actors in each movie is 7.136.

Maximum likelihood estimates			
k	$\hat{\nu}_k$	k	$\hat{\nu}_k$
1	0.081	7	0.057
2	0.071	8	0.054
3	0.065	9	0.046
4	0.072	10	0.275
5	0.062	11	0.161
6	0.059		