

# SUPPLEMENTARY MATERIAL: EDGE EXCHANGEABLE MODELS FOR INTERACTION NETWORKS

HARRY CRANE AND WALTER DEMPSEY

ABSTRACT. This is the supplement to Crane and Dempsey (2017). We characterize edge exchangeable network models, analyze the performance of maximum likelihood estimators under the Hollywood model, and describe additional examples of interaction datasets.

## APPENDIX A. CHARACTERIZATION OF EDGE EXCHANGEABLE NETWORKS

Our representation in Crane and Dempsey (2017, Theorem 4.2) applies only to the special case of edge exchangeable networks that are *blip-free*. We prove here a more general, complete characterization for all edge exchangeable networks in the binary case for which each edge is undirected and involves exactly two vertices. The argument for directed networks with edges of any finite arity is analogous but more technical. Recall the notation for edge-labeled networks introduced in Crane and Dempsey (2017).

Let  $\text{fin}_2(\mathcal{P}) \subset \text{fin}(\mathcal{P})$  be the set of all size 2 multisets of  $\mathcal{P}$ . Given a binary edge-labeled network  $\mathcal{E}$  with  $e(\mathcal{E}) = n$ , we call  $S : [n] \rightarrow \text{fin}_2(\mathbb{N})$  a *selection function* for  $\mathcal{E}$  if  $\mathcal{E}_S = \mathcal{E}$ , where  $\mathcal{E}_S$  is as defined by Crane and Dempsey (2017, Equation (5)); that is,  $S$  is an interaction process whose induced edge-labeled network agrees with  $\mathcal{E}$ . We think of the selection function as a way of labeling the vertices of the edge-labeled network  $\mathcal{E}$ .

Selection functions  $S, S' : [n] \rightarrow \text{fin}_2(\mathbb{N})$  are equivalent, written  $S \equiv S'$ , if they correspond to the same edge-labeled network, that is,  $\mathcal{E}_S = \mathcal{E}_{S'}$ . To every edge-labeled network  $\mathcal{E}$  with  $n$  edges we associate a *canonical selection function*  $S_{\mathcal{E}} : [n] \rightarrow \text{fin}_2(\mathbb{N})$  defined by labeling the vertices *in order of appearance*, as follows.

We initialize by putting  $S_{\mathcal{E}}(1) = \{1, 1\}$  if the edge labeled 1 is a self loop and otherwise we put  $S_{\mathcal{E}}(1) = \{1, 2\}$  to indicate an edge between two distinct vertices. Given  $S_{\mathcal{E}}(1), \dots, S_{\mathcal{E}}(i-1)$ , we define

---

*Date:* May 8, 2017.

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

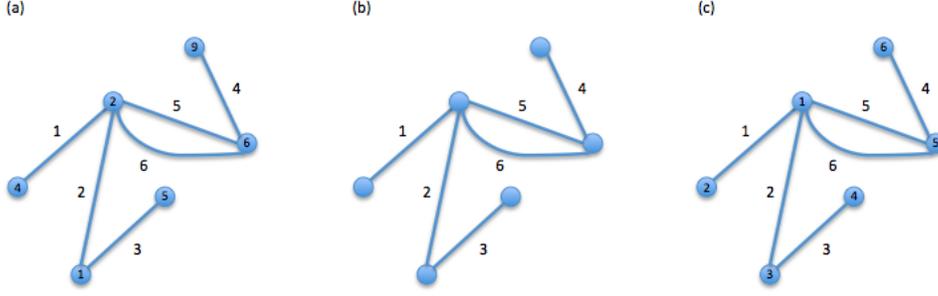


FIGURE 1. Illustration of the generic process of generating edge exchangeable networks in Section A. (a) Network representation of interaction process  $X_1 = \{2, 4\}$ ,  $X_2 = \{1, 2\}$ ,  $X_3 = \{1, 5\}$ ,  $X_4 = \{6, 9\}$ ,  $X_5 = \{2, 6\}$ ,  $X_6 = \{2, 6\}$ . (b) Edge-labeled network  $\mathcal{E}_X$  obtained by removing vertex labels from the vertex-edge-labeled network in Panel (a). (c) Vertex labeling of the network in Panel (b) according to its canonical selection function.

$S_{\mathcal{E}}(i) = \{v_1(i), v_2(i)\}$  by choosing  $v_1(i) \leq v_2(i)$  to be the smallest vertex labels consistent with the structure of  $\mathcal{E}|_{[i]}$ . In this way,  $v_j(i) = r$  coincides with a previously observed vertex label if one of the vertices involved in the  $i$ th interaction corresponds to the vertex labeled  $r$  in the previous interactions  $S_{\mathcal{E}}(1), \dots, S_{\mathcal{E}}(i-1)$ . See Figures 1(b) and 1(c) for an illustration.

The  $\text{fin}_2(\mathbb{N})$ -simplex consists of all  $(f_{i,j})_{j \geq i \geq -1}$  such that  $f_{i,j} \geq 0$  for all  $j \geq i \geq -1$ ,  $f_{-1,i} = 0$  for all  $i \neq 0$ , and  $\sum_{j \geq i \geq -1} f_{i,j} = 1$ . (The labels  $-1$  and  $0$  have a special status in the construction, as we shall see.) For any  $f = (f_{i,j})_{j \geq i \geq -1}$  in the  $\text{fin}_2(\mathbb{N})$ -simplex and  $i \in \mathbb{N}$ , we define

$$f_{\bullet}^{(i)} = \sum_{j=0}^{\infty} f_{i,j}$$

as the sum of masses involving element  $i$ .

Every  $f = (f_{i,j})_{j \geq i \geq -1}$  in the  $\text{fin}_2(\mathbb{N})$ -simplex determines a probability distribution on edge-labeled networks, denoted  $\epsilon_f$ , as follows. Let  $X_1, X_2, \dots$  be i.i.d. random pairs  $\{i, j\}$  with

$$(1) \quad \text{pr}(X_k = \{i, j\} \mid f) = f_{i,j}, \quad j \geq i \geq -1.$$

Given  $X_1, X_2, \dots$ , we define the selection function  $\mathcal{X} : \mathbb{N} \rightarrow \text{fin}_2(\mathbb{Z})$ , where  $\mathbb{Z} = \{\dots, -1, 0, 1, \dots\}$ , as follows. We initialize with  $m_0 = 0$ . For  $n \geq 1$ , suppose  $m_{n-1} = z \leq 0$ . If  $X_n$  contains no 0s, then we define  $\mathcal{X}(n) = X_n$  and update  $m_n = m_{n-1}$ . If  $X_n = \{0, j\}$  for some  $j \geq 1$ , then we put  $\mathcal{X}(n) = \{z-1, j\}$  and update  $m_n = z-1$ . Otherwise, if  $X_n = \{0, 0\}$ , then we put  $\mathcal{X}(n) = \{z-1, z-1\}$  and update

$m_n = z - 1$ ; and if  $X_n = \{-1, 0\}$ , then we put  $\mathcal{X}(n) = \{z - 1, z - 2\}$  and update  $m_n = z - 2$ . (Thus, events with  $-1$  or  $0$  involve vertices that appear once and never again. These are the “blips” that we previously ruled out in Crane and Dempsey (2017, Theorem 4.2).) We define  $\mathcal{Y} = \mathcal{E}_{\mathcal{X}} \sim \epsilon_f$  to be the edge-labeled network induced by  $\mathcal{X}$ . See Figure 1 for an illustration of this procedure.

**Proposition A.1.** *The edge-labeled network  $\mathcal{Y} = \mathcal{E}_{\mathcal{X}}$  corresponding to  $X_1, X_2, \dots$  i.i.d. from (1) is edge exchangeable for all  $f$  in the  $\text{fin}_2(\mathbb{N})$ -simplex.*

For identifiability, we define the *ranked reordering* of  $f$  by  $f^\downarrow = (f_{\{i,j\}}^\downarrow)_{j \geq i \geq -1}$ , the element of the  $\text{fin}_2(\mathbb{N})$ -simplex obtained by putting  $f_{\{-1,0\}}^\downarrow = f_{\{-1,0\}}$ ,  $f_{\{0,0\}}^\downarrow = f_{\{0,0\}}$ , and reordering elements  $1, 2, \dots$  so that  $f_{\bullet}^{(i)} \geq f_{\bullet}^{(i+1)}$  for all  $i \geq 1$  and then breaking ties  $f_{\bullet}^{(i)} = f_{\bullet}^{(i+1)}$  by declaring that  $(f_{\{i,i\}}, f_{\{i,i+2\}}, \dots)$  comes before  $(f_{\{i+1,i+1\}}, f_{\{i+1,i+2\}}, \dots)$  in the lexicographic ordering. We write  $\mathcal{F}^\downarrow$  to denote the space of rank reordered elements of the  $\text{fin}_2(\mathbb{N})$ -simplex.

As the vertex labels other than  $-1$  and  $0$  are inconsequential, it is clear that  $\epsilon_f$  and  $\epsilon_{f'}$  determine the same distribution for any  $f, f'$  for which  $f^\downarrow = f'^\downarrow$ . For any edge-labeled network  $\mathcal{E}$ , we write  $|\mathcal{E}|^\downarrow \in \mathcal{F}^\downarrow$  to denote its *signature*, if it exists, as follows. Let  $S_{\mathcal{E}} : \mathbb{N} \rightarrow \text{fin}_2(\mathbb{N})$  be the canonical selection function for  $\mathcal{E}$ . For every  $\{i, j\} \in \text{fin}_2(\mathbb{N})$ ,  $j \geq i \geq 1$ , we define

$$f_{\{i,j\}}(\mathcal{E}) = \lim_{n \rightarrow \infty} n^{-1} \sum_{k=1}^n \mathbf{1}\{S_{\mathcal{E}}(k) = \{i, j\}\} \quad \text{and}$$

$$f_{\bullet}^{(i)}(\mathcal{E}) = \lim_{n \rightarrow \infty} n^{-1} \sum_{k=1}^n \mathbf{1}\{i \in S_{\mathcal{E}}(k)\},$$

if the limits exist. We also define

$$f_{\{0,i\}}(\mathcal{E}) = f_{\bullet}^{(i)} - \sum_{j=1}^{\infty} f_{\{i,j\}}(\mathcal{E}), \quad i \geq 1,$$

$$f_{\{0,0\}}(\mathcal{E}) = \lim_{n \rightarrow \infty} n^{-1} \sum_{k=1}^n \left( \sum_{\ell \geq 1} \mathbf{1}\{S_{\mathcal{E}}(k) = \{\ell, \ell\}\} \right) - \sum_{i=1}^{\infty} f_{\{i,i\}}(\mathcal{E}), \quad \text{and}$$

$$f_{\{-1,0\}}(\mathcal{E}) = \lim_{n \rightarrow \infty} n^{-1} \sum_{k=1}^n \left( \sum_{\ell, r \geq 1: \ell \neq r} \mathbf{1}\{S_{\mathcal{E}}(k) = \{\ell, r\}\} \right) - \sum_{j > i \geq 1} f_{\{i,j\}}(\mathcal{E}).$$

Provided each of the above limiting frequencies exists, we define  $|\mathcal{E}| = (f_{\{i,j\}}(\mathcal{E}))_{j \geq i \geq -1}$  and  $|\mathcal{E}|^\downarrow = (f_{\{i,j\}}^\downarrow(\mathcal{E}))_{j \geq i \geq -1}$ .

(Note the role of  $f_{\{0,0\}}(\mathcal{E})$  and  $f_{\{-1,0\}}(\mathcal{E})$  for recording the residual proportion of loops and edges, respectively, that do not contribute to the limiting frequencies  $f_{\{i,j\}}(\mathcal{E})$  for a given  $j \geq i \geq 1$ . For example, the interaction process  $\mathcal{I} : \mathbb{N} \rightarrow \text{fin}_2(\mathbb{N})$  given by  $\mathcal{I}(i) = \{i, i\}$  for each  $i \geq 1$  corresponds to the edge-labeled network  $\mathcal{E} = \mathcal{E}_{\mathcal{I}}$  for which every edge is a loop at a distinct vertex. For any given  $i \geq 1$ ,  $f_{\{i,i\}}(\mathcal{E}) = 0$  and we have  $f_{\{0,0\}}(\mathcal{E}) = 1$ . Conversely, for  $f = (f_{\{i,j\}})_{j \geq i \geq -1}$  with  $f_{\{0,0\}} = 1$ , the sequence  $X_1, X_2, \dots$  i.i.d. from (1) yields  $\{0,0\}, \{0,0\}, \dots$  with probability 1, whose associated interaction process has  $\mathcal{X}(1) = \{-1, -1\}$ ,  $\mathcal{X}(2) = \{-2, -2\}$ , and  $\mathcal{X}(n) = \{-n, -n\}$ , for each  $n \geq 1$ , with associated edge-labeled network having all edges corresponding to a loop at a distinct vertex.)

**Theorem A.2.** *Let  $\mathcal{Y} \in \mathfrak{G}_{\mathbb{N}}$  be an edge exchangeable network. Then there exists a unique probability measure  $\phi$  on  $\mathcal{F}^{\downarrow}$  such that  $\mathcal{Y} \sim \epsilon_{\phi}$ , where*

$$(2) \quad \epsilon_{\phi}(\cdot) = \int_{\mathcal{F}^{\downarrow}} \epsilon_f(\cdot) \phi(df).$$

*That is, every edge exchangeable network  $\mathcal{Y}$  can be generated by first sampling  $f \sim \phi$  and, given  $f$ , putting  $\mathcal{Y} = \mathcal{E}_{\mathcal{X}}$  for  $\mathcal{X} : \mathbb{N} \rightarrow \text{fin}_2(\mathbb{Z})$  constructed from  $X_1, X_2, \dots$  i.i.d. according to (1).*

Theorem 4.2 of Crane and Dempsey (2017) follows as a corollary to Theorem A.2 by ruling out blips, that is, by confining  $\phi$  to the subset of  $f \in \text{fin}_2(\mathbb{N})$  for which  $f_{\{i,0\}} = 0$  for all  $i \geq -1$ .

**A.1. Proof of Theorem A.2.** We equip  $\mathfrak{G}_{\mathbb{N}}$  with the product-discrete topology induced by the metric

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

with convention  $1/\infty = 0$ , and  $\mathcal{F}^{\downarrow}$  with the topology induced by

$$d_{\mathcal{F}^{\downarrow}}(f, f') = \sum_{j \geq i \geq -1} |f_{\{i,j\}} - f'_{\{i,j\}}|, \quad f, f' \in \mathcal{F}^{\downarrow}.$$

We then work with the respective Borel  $\sigma$ -fields induced by these topologies.

Let  $\mathcal{Y} \in \mathfrak{G}_{\mathbb{N}}$  be an edge exchangeable random network, let  $S_{\mathcal{Y}} : \mathbb{N} \rightarrow \text{fin}_2(\mathbb{N})$  be its canonical selection function, and let  $\xi_1, \xi_2, \dots$  be an i.i.d. sequence of Uniform $[0, 1]$  random variables which are independent of  $\mathcal{Y}$ . Given  $\mathcal{Y}$  and  $(\xi_i)_{i \geq 1}$ , we define  $\mathcal{Z} : \mathbb{N} \rightarrow \text{fin}_2([0, 1])$  by  $\mathcal{Z}(n) = \{\xi_i, \xi_j\}$  on the event  $S_{\mathcal{Y}}(n) = \{i, j\}$ , for  $n \geq 1$ .

By independence of  $\mathcal{Y}$  and  $(\xi_i)_{i \geq 1}$  and edge exchangeability of  $\mathcal{Y}$ ,  $(\mathcal{Z}(n))_{n \geq 1}$  is an exchangeable sequence taking values in the Polish space  $\text{fin}_2([0, 1])$ . By de Finetti's theorem, see, for example, Aldous (1985), there exists a unique measure  $\mu$  on the space  $\mathcal{P}(\text{fin}_2([0, 1]))$  of probability measures on  $\text{fin}_2([0, 1])$  such that  $\mathcal{Z} =_{\mathcal{D}} \mathcal{Z}^* = (\mathcal{Z}^*(n))_{n \geq 1}$  with

$$\text{pr}(\mathcal{Z}^* \in \cdot) = \int_{\mathcal{P}(\text{fin}_2([0, 1]))} m^\infty(\cdot) \mu(dm),$$

where  $m^\infty$  denotes the infinite product measure of  $m$ . In particular, there exists a random measure  $\nu$  on  $\text{fin}_2([0, 1])$  such that

$$\text{pr}(\mathcal{Z} \in \cdot \mid \nu) = \nu^\infty \quad \text{a.s.}$$

Given  $\nu$ , we define

$$f_{\{i, j\}} = \nu(\{\{\xi_i, \xi_j\}\}), \quad i, j \geq 1,$$

$$f_{\bullet}^{(i)} = \nu(\{\{x, y\} \in \text{fin}_2([0, 1]) : \xi_i \in \{x, y\}\}), \quad i \geq 1,$$

$$f_{\{0, i\}} = f_{\bullet}^{(i)} - \sum_{j=1}^{\infty} f_{\{i, j\}}, \quad i \geq 1,$$

$$f_{\{0, 0\}} = \nu(\{\{u, u\} \in \text{fin}_2([0, 1])\}) - \sum_{i=1}^{\infty} f_{\{i, i\}}, \quad \text{and}$$

$$f_{\{-1, 0\}} = \nu(\{\{u, v\} \in \text{fin}_2([0, 1]) : v \neq u\}) - \sum_{j > i \geq 1} f_{\{i, j\}}.$$

By construction  $(f_{\{i, j\}})_{j \geq i \geq -1}$  is in the  $\text{fin}_2(\mathbb{N})$ -simplex and, therefore,  $f^\downarrow \in \mathcal{F}^\downarrow$ . (Note that  $\mathcal{F}^\downarrow$  is a subset of the  $\text{fin}_2(\mathbb{N})$ -simplex and  $f \mapsto f^\downarrow$  is measurable with respect to the Borel  $\sigma$ -field induced by the metric  $d_{\mathcal{F}^\downarrow}(\cdot, \cdot)$  given above.)

Given  $\nu$ , we let  $(\mathcal{Z}', S')$  be an i.i.d. copy of  $(\mathcal{Z}, S_{\mathcal{Y}})$  and let  $\mathcal{Y}' = \mathcal{E}_{S'}$  be the edge-labeled network induced by  $S'$ . We complete the proof by showing  $\text{pr}(\mathcal{Y}' \in \cdot \mid \nu) = \epsilon_{f^\downarrow}$ , for  $f^\downarrow$  as defined above from  $\nu$ .

First, let  $A = \{i \in \mathbb{N} : f_{\bullet}^{(i)} > 0\}$  and  $\xi_A = \{\xi_i : i \in A\}$ . It follows that

$$\text{pr}(\mathcal{Z}'(1) \cap \xi_A = \emptyset \mid \nu) = f_{\{0, 0\}} + f_{\{-1, 0\}} \quad \text{and} \quad \text{pr}(\mathcal{Z}'(1) \cap \xi_A = \{\xi_i\} \mid \nu) = f_{\{0, i\}}.$$

## Maximum likelihood estimates

(A)	MLE for $\alpha$		
	$\alpha = 0.2$	$\alpha = 0.4$	$\alpha = 0.6$
$\theta = 5$	0.18 (0.064, 0.068)	0.39 (0.041, 0.042)	0.59 (0.022, 0.025)
$\theta = 10$	0.19 (0.060, 0.058)	0.39 (0.038, 0.037)	0.60 (0.021, 0.022)
$\theta = 20$	0.19 (0.050, 0.050)	0.40 (0.030, 0.033)	0.60 (0.021, 0.021)
(B)	MLE for $\theta$		
	$\alpha = 0.2$	$\alpha = 0.4$	$\alpha = 0.6$
$\theta = 5$	6.29 (2.81, 2.56)	5.94 (2.66, 2.79)	5.99 (3.71, 3.42)
$\theta = 10$	10.64 (3.37, 3.57)	11.71 (4.18, 4.19)	12.11 (5.23, 5.21)
$\theta = 20$	21.94 (6.00, 5.58)	21.92 (5.87, 6.20)	21.67 (6.79, 7.41)

TABLE 1. Performance of maximum likelihood estimation (MLE) for  $(\alpha, \theta)$  from the Hollywood model based on network data with 2,000 edges and 100 iterations. Each cell in Panel (A) contains the mean  $\hat{\alpha}_{MLE}$  in 100 iterations of MLE from the model for the specified combination of  $(\alpha, \theta)$ . The values  $(\hat{s}_{CRLB}(\hat{\alpha}), \hat{s}_e(\hat{\alpha}))$  are the empirical Cramér–Rao lower bound based on inverting the information matrix and the simulation-based estimate of the standard error, respectively. Panel (B) contains the analogous information for the MLE of  $\theta$ .

By exchangeability,  $i \notin A$  implies  $\xi_i$  appears at most once in  $\mathcal{Z}$  with probability 1. We further have that

$$\begin{aligned} \text{pr}(\mathcal{Z}'_1 \cap \xi_A = \emptyset \text{ and } \mathcal{Z}'(1) = \{u, u\} \text{ for some } u \in [0, 1] \mid \nu) &= f_{\{0,0\}} \quad \text{and} \\ \text{pr}(\mathcal{Z}'(1) \cap \xi_A = \emptyset \text{ and } \mathcal{Z}'(1) = \{u, v\} \text{ for } u \neq v \mid \nu) &= f_{\{-1,0\}}. \end{aligned}$$

Now, define  $\mathcal{X}' : \mathbb{N} \rightarrow \text{fin}_2(\mathbb{N} \cup \{-1, 0\})$  and the random selection function  $S_{\mathcal{X}'} : \mathbb{N} \rightarrow \text{fin}_2(\mathbb{Z})$  as follows. Let  $m_0 = 0$ . For  $n \geq 1$ , suppose  $m_{n-1} = z \leq 0$ . If  $\mathcal{Z}'(n) \cap \xi_A = \{\xi_i, \xi_j\}$  for some  $i, j \in \mathbb{N}$ , then put  $\mathcal{X}(n) = S_{\mathcal{X}}(n) = \{i, j\}$ . If  $\mathcal{Z}'(n) \cap \xi_A = \{\xi_i\}$  for some  $i \geq 1$ , then put  $\mathcal{X}(n) = \{0, i\}$ ,  $S_{\mathcal{X}}(n) = \{z - 1, i\}$ , and  $m_n = z - 1$ . If  $\mathcal{Z}'(n) \cap \xi_A = \emptyset$  and  $\mathcal{Z}'(n) = \{u, u\}$  for some  $u \in [0, 1]$ , put  $\mathcal{X}(n) = \{0, 0\}$ ,  $S_{\mathcal{X}}(n) = \{z - 1, z - 1\}$ , and  $m_n = z - 1$ . And if  $\mathcal{Z}'(n) \cap \xi_A = \emptyset$  and  $\mathcal{Z}'(n) = \{u, v\}$  for  $u \neq v$ , put  $\mathcal{X}(n) = \{-1, 0\}$ ,  $S_{\mathcal{X}}(n) = \{z - 1, z - 2\}$ , and  $m_n = z - 2$ .

By construction, we have  $S_{\mathcal{X}'} \equiv S'$  a.s. and, given  $f$ ,  $\mathcal{X}'$  is conditionally i.i.d. from distribution (1). The integral representation in (2) follows by de Finetti's theorem, completing the proof.

## APPENDIX B. SIMULATION STUDY OF MAXIMUM LIKELIHOOD UNDER HOLLYWOOD MODEL

For estimating the parameters  $(\alpha, \theta)$  from the Hollywood model, we use a quasi-Newton method based on the gradients in Equations (16) and (17) of Crane and Dempsey (2017). Since the likelihood surface is smooth (i.e., gradient is a smooth function of  $\alpha$  and  $\theta$ ) we expect maximum likelihood estimates to converge to their true parameter values. Our empirical investigation supports this conclusion. Table 1 confirms asymptotic consistency of the MLE for  $(\alpha, \theta)$  via simulation. From this we see that the MLE for  $\alpha$  is accurate even for the relatively small sample size of 2,000 edges. The MLE for  $\theta$  is less accurate, as we expect based on what is known of estimation of the mutation rate in Ewens's sampling formula (Ewens, 1972). In practice, however, accurate estimation of  $\alpha$  is of greater interest, as it is directly related to the power law exponent in the degree distribution. We also notice that inverting the empirical information matrix gives a reasonable approximation to the standard errors for both  $\hat{\alpha}_{MLE}$  and  $\hat{\theta}_{MLE}$ .

## APPENDIX C. DESCRIPTION OF NETWORK DATASETS

Below we describe and provide references for some interaction datasets discussed in Crane and Dempsey (2017).

- Actors collaboration (Barabási and Albert, 1999): Network built from collaborations among actors in a given sample of movies. Each edge connects the actors who played a role in the corresponding movie.
- Enron email corpus (Klimt and Yang, 2004): Network built from a corpus of about 500,000 emails. Vertices are employees in the Enron Corporation with a directed edge from vertex  $i$  to  $j$  for each email sent from  $i$  to  $j$ . Some versions of this dataset project edge multiplicities and ignore edge direction so that for each pair of vertices  $i$  and  $j$  there is an undirected edge between  $i$  and  $j$  if at least one email was exchanged between the two.
- Karate Club (Zachary, 1977): Network built from social interactions among 34 members of a karate club. Vertices are the members of the club and an edge between  $i$  and  $j$  corresponds to a social interaction between the two. The network exhibits no vertex sampling or growth since it is assumed all club members have been observed.

- Wikipedia voting (Leskovec et al., 2010): The Wikipedia voting network represents voting behavior for elections to the administrator role in Wikipedia. Vertices are Wikipedia users and a directed edge points from  $i$  to  $j$  if user  $i$  voted for user  $j$ .
- US Airport (Colizza et al., 2007; Opsahl, 2011): The network from Opsahl (2011) is built from the flight map between all US airports in 2010. A directed edge from  $i$  to  $j$  indicates that a flight was scheduled from airport  $i$  to airport  $j$  in 2010. Edges are weighted by the number of seats on the scheduled flights. The network grows as a consequence of additional flights between airports.
- Co-authorship (Newman, 2001): Network built from co-authorship of preprints on the Condensed Matter section of arXiv between 1995 and 1999. Vertices are of two types, authors and papers, and edges only exist between vertices of a different type. An edge between  $i$  (author) and  $j$  (paper) indicates that  $i$  is an author on paper  $j$ . The data is more succinctly represented by an edge-labeled network as in Crane and Dempsey (2017) by associating each article to an interaction involving all of its authors.
- UC Irvine (Opsahl and Panzarasa, 2009): Network built from UC Irvine online community. Vertices are active members of the community and a directed edge from  $i$  to  $j$  indicates that a message was sent from user  $i$  to user  $j$ .
- Political blogs (Adamic and Glance, 2005): Network built from hyperlinks between political blogs. Vertices are websites (blogs) with a directed edge from  $i$  to  $j$  for every hyperlink from website  $i$  to website  $j$ . Sampling is similar to the Facebook network.

#### REFERENCES

- [1] L. A. Adamic and N. Glance. The Political Blogosphere and the 2004 US Election: Divided They Blog. In *Proceedings of the 3rd international workshop on Link discovery*, pages 36–43, 2005.
- [2] D. J. Aldous. Exchangeability and related topics. In *École d'été de probabilités de Saint-Flour, XIII—1983*, volume 1117 of *Lecture Notes in Math.*, pages 1–198. Springer, Berlin, 1985.
- [3] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999. ISSN 0036-8075. doi: 10.1126/science.286.5439.509. URL <http://dx.doi.org/10.1126/science.286.5439.509>.

- [4] V. Colizza, R. Pastor-Satorras, and A. Vespignani. Reaction diffusion processes and metapopulation models in heterogeneous networks. *Nature Physics*, 3:276–282, 2007.
- [5] H. Crane and W. Dempsey. Edge exchangeable models for interaction networks. 2017.
- [6] W. J. Ewens. The sampling theory of selectively neutral alleles. *Theoret. Population Biology*, 3:87–112, 1972. ISSN 0040-5809.
- [7] B. Klimt and Y. Yang. Introducing the Enron corpus. *CEAS*, 2004.
- [8] J. Leskovec, D. Huttenlocher, and J. Kleinberg. Signed networks in social media. *CHI*, 2010.
- [9] J.J. McAuley and Jure Leskovec. Learning to discover social circles in ego networks. In *Neural Information Processing Systems*, pages 539–547, 2012.
- [10] M.E.J. Newman. The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, 98:404–409, 2001.
- [11] T. Opsahl. Why Anchorage is not (that) important: Binary ties and Sample selection. Available at <http://top.me/poFcY-Vw>, 2011.
- [12] T. Opsahl and P. Panzarasa. Clustering in weighted networks. *Social Networks*, 31:155–163, 2009.
- [13] D. Watts and S. Strogatz. Collective dynamics of ‘small-world’ networks. *Nature*, 393:440–442, 1998.
- [14] W.W. Zachary. An information flow model for conflict and fission in small groups. *Journal of Anthropological Research*, 33:452–473, 1977.

DEPARTMENT OF STATISTICS & BIostatISTICS, RUTGERS UNIVERSITY, 110 FRELINGHUYSEN AVENUE, PISCATAWAY, NJ 08854, USA

*E-mail address:* hcrane@stat.rutgers.edu

*URL:* <http://stat.rutgers.edu/home/hcrane>

DEPARTMENT OF STATISTICS, UNIVERSITY OF MICHIGAN, 1085 S. UNIVERSITY AVE, ANN ARBOR, MI 48109, USA

*E-mail address:* wdem@umich.edu