

RELATIONAL EXCHANGEABILITY

HARRY CRANE,* *Rutgers University*

WALTER DEMPSEY,** *Harvard University*

Abstract

A *relationally exchangeable structure* is a random combinatorial structure whose law is invariant with respect to relabeling its relations, as opposed to its elements. Historically, exchangeable random set partitions have been the best known examples of relationally exchangeable structures, but the concept now arises more broadly when modeling interaction data in modern network analysis. Aside from exchangeable random partitions, instances of relational exchangeability include edge exchangeable random graphs and hypergraphs, path exchangeable processes, and a range of other network-like structures. We motivate the general theory of relational exchangeability, with special emphasis on the alternative perspective it provides and its benefits in certain applied probability problems. We then prove a de Finetti-type structure theorem for the general class of relationally exchangeable structures.

Keywords: Edge exchangeable graph; Kingman's correspondence; paintbox process; exchangeable random partition

2010 Mathematics Subject Classification: Primary 60G09
Secondary 90B15

1. Introduction

The traditional theory of exchangeability concerns families of random variables whose distribution is invariant with respect to symmetries of a well-defined index set. This includes the classical study of exchangeable sequences $(X_i)_{i \in I}$, which satisfy $(X_{\sigma(i)})_{i \in I} \stackrel{D}{=} (X_i)_{i \in I}$ for all bijections $\sigma : I \rightarrow I$, with ' $\stackrel{D}{=}$ ' indicating *equality in distribution*. The concept naturally extends to two-dimensional arrays $(X_{ij})_{i,j \in I}$ and arbitrary d -dimensional arrays $(X_{i_1 \dots i_d})_{i_1, \dots, i_d \in I}$, for which exchangeability implies that

$$(X_{\sigma(i_1) \dots \sigma(i_d)})_{i_1, \dots, i_d \in I} \stackrel{D}{=} (X_{i_1 \dots i_d})_{i_1, \dots, i_d \in I} \quad \text{for all permutations } \sigma : I \rightarrow I.$$

An exception to this setup arises in the study of exchangeable random partitions, which were initially studied by Kingman [21] as models for classification of species in ecology, and have since proven useful throughout population genetics [17] and statistical applications for clustering [24]; see [11] for a recent survey. In this case, exchangeability is defined as symmetry with respect to relabeling of an equivalence relation.

Received 16 January 2018; revision received 3 January 2019.

* Postal address: Department of Statistics & Biostatistics, Rutgers University, 110 Frelinghuysen Avenue, Piscataway, NJ 08854, USA. Email address: hcrane@stat.rutgers.edu

Partially supported by NSF grants CNS-1523785 and CAREER DMS-1554092.

** Postal address: Department of Statistics, Harvard University, One Oxford Street, Cambridge, MA 02138, USA. Email address: wdempsey@fas.harvard.edu

As we discuss in further detail below, Kingman’s theory of exchangeable partitions arises naturally in certain genetic and ecological contexts, in which a random sample of units are partitioned according to their membership in otherwise unlabeled categories. Here we expand upon Kingman’s initial treatment by describing a general class of random structures whose distributions are exchangeable with respect to relabeling of their relations. The theory of *relational exchangeability* presented here is thus a refinement of Kingman’s theory of exchangeable partitions and the theory of edge exchangeable random networks. Before characterizing this class of structures, we first discuss several motivating examples.

1.1. Exchangeable partitions

Consider a random sequence $\mathbf{X} = (X_1, \dots, X_n)$ taking values in an at most countable set S . For concreteness, we may regard S as a set of species, so that \mathbf{X} records the species for a random sample of animals from a certain population. From \mathbf{X} , we define an equivalence relation ‘ $\sim_{\mathbf{X}}$ ’ on $[n] := \{1, \dots, n\}$ by

$$i \sim_{\mathbf{X}} j \text{ if and only if } X_i = X_j. \tag{1.1}$$

We write $\Pi(\mathbf{X}) = \{B_1, B_2, \dots\}$ to denote the set partition whose blocks B_1, B_2, \dots are the equivalence classes induced by ‘ $\sim_{\mathbf{X}}$ ’. We sometimes write ‘ \sim_{Π} ’ in place of ‘ $\sim_{\mathbf{X}}$ ’ when convenient.

Suppose now that $\mathbf{X} = (X_1, \dots, X_n)$ is *exchangeable*, meaning that

$$(X_{\sigma(1)}, \dots, X_{\sigma(n)}) \stackrel{D}{=} (X_1, \dots, X_n)$$

for all permutations $\sigma : [n] \rightarrow [n]$. Then exchangeability of \mathbf{X} induces exchangeability on the equivalence relation $\Pi := \Pi(\mathbf{X})$ as defined in (1.1) in the sense that $\Pi^{\sigma} \stackrel{D}{=} \Pi$ for all permutations $\sigma : [n] \rightarrow [n]$, with Π^{σ} defined by

$$i \sim_{\Pi^{\sigma}} j \text{ if and only if } \sigma^{-1}(i) \sim_{\Pi} \sigma^{-1}(j).$$

A key difference between exchangeable sequences and exchangeable partitions is that the equivalence classes in a set partition are unlabeled. As there are no definitive characteristics of the elements independent of how they relate to each other through the equivalence relation, the exchangeability condition is thus a distributional invariance of the associated *relation* instead of a distributional invariance with respect to the relabeling of the unobserved *labels* (i.e. species).

1.2. More general structures

The above theory of exchangeable random partitions can be refined to a wider class of structures that arise in modern data science and network analysis applications. Consider, for example, a uniform random sampling of emails from a database, e.g. the Enron email dataset [22]. If emails are sampled sequentially from the database then the resulting data structure takes the form of a sequence of binary relations among individuals in a population, i.e. the sequence of sender–receiver pairs corresponding to a given email exchange, as in Figure 1, below.

Whereas email correspondences are binary (i.e. a communication between caller and receiver), other common network structures may not be. Common examples of generic interaction networks include the network of actor collaborations formed by sampling from the Internet Movie Database (IMDb) or the network of scientific coauthorships obtained by sampling articles from arXiv. These examples give rise to edge-labeled hypergraphs, as shown in Figures 5 and 6, below.

Still other examples arise, as in the path-labeled networks of Figure 7, below, which might represent the network structure obtained by sampling paths in the Internet. Such path sampling mechanisms were critical to the early studies of Internet topology; see, e.g. [1] for some discussion of traceroute sampling and its impact on network analysis. We discuss these examples in further detail in Section 3.

The above examples are special cases of what we call *relational structures*. These structures are relational by virtue of their being observed by sampling a *relation* (e.g. species, phone call, movie collaboration, coauthorship, path between routers, etc.). The theory emerging is thus different from more traditional theories of exchangeability, for which the natural sampling unit, and thus the invariance, is defined with respect to relabeling of the underlying population.

As we observe beginning in Section 2, exchangeable relational data structures exhibit different probabilistic behaviors than more traditional exchangeable array data. A key difference lies in the assumed observation mechanism: whereas a random sequence (X_1, \dots, X_n) represents *individual measurements* taken on a sample of n distinct individuals, a random partition of $[n]$, as in (1.1), represents a *composite binary relation* on $[n]$. The difference is subtle, as can be seen through the different theories for exchangeable random sequences (de Finetti's theorem) and exchangeable random partitions (Kingman's theorem). We begin in Section 2 with a discussion of edge exchangeable random networks, the simplest nontrivial extension of Kingman's theory. We focus specifically on the distinction between the vertex- and edge-centric perspectives of network data, and why the latter is most appropriate for modeling networks constructed from sampled interactions, as in the Enron dataset. In Section 3, we extend this perspective to multiway interactions, such as networks constructed by observing scientific collaborations and paths in the Internet. In Section 4, we generalize these special cases into a framework for relationally labeled structures and prove our main theorem, which characterizes the probabilistic structure of relationally exchangeable structures uniquely in terms of a class of *interaction propensity processes*. By specializing our theory to the case of exchangeable equivalence relations, we recover Kingman's paintbox correspondence. However, the argument needed to obtain our general representation requires a careful handling of elements called 'blips', which correspond to the 'dust' in the special case of Kingman's theorem. We make concluding remarks in Section 5.

2. Edge-centric view of network structures

Consider sampling uniformly at random from a database of telephone calls exchanged among individuals in a countable set S . The sampling procedure results in an $(S \times S)$ -valued sequence $\mathbf{X} = (X_1, \dots, X_n)$, where each X_i is an ordered pair (c_i, r_i) representing the caller c_i and receiver r_i . Such a call sequence can be represented by a vertex-edge labeled network in a straightforward way. For example, the sequence of ordered pairs given by

$$X_1 = (a, b), \quad X_2 = (c, a), \quad X_3 = (d, e), \quad X_4 = (a, c) \quad (2.1)$$

can be represented by a network-like object as shown in Figure 1(a).

Assuming that the sequence \mathbf{X} is exchangeable induces a related exchangeability property on the induced vertex-edge labeled structure. By exchangeability of \mathbf{X} , the model assigns equal probability to any two structures that are isomorphic up to relabeling edges, as in Figure 1(a) and (b). If we further assume that $\mathbf{X} = (X_1, X_2, \dots)$ is countably infinite, then the one-to-one correspondence between the sequence \mathbf{X} and its vertex-edge labeled network representation implies that the class of models for such data is determined by de Finetti's theorem for exchangeable sequences.

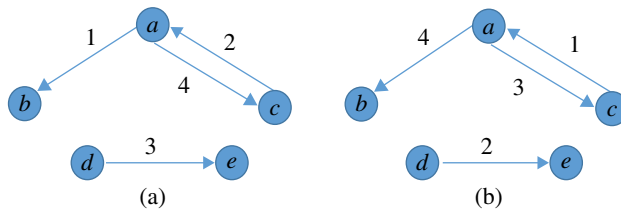


FIGURE 1: (a) Representation of the sequence of ordered pairs in (2.1). (b) Representation of the sequence of ordered pairs in $S = \{a, b, c, d, e\}$ with $X'_1 = (c, a)$, $X'_2 = (d, e)$, $X'_3 = (a, c)$, and $X'_4 = (a, b)$. The object in (b) can be obtained from that in (a) by reordering (X_1, X_2, X_3, X_4) as (X_2, X_3, X_4, X_1) . In both (a) and (b), the edges are labeled according to where they appear in the sequence.

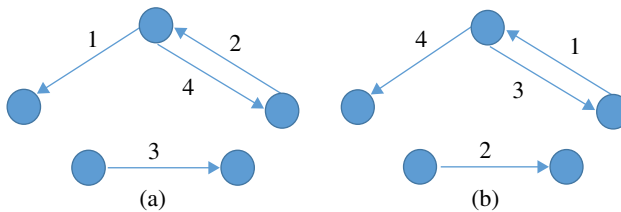


FIGURE 2: (a) Edge-labeled network induced by disregarding vertex labels in Figure 1(a). (b) Edge-labeled graph induced by disregarding vertex labels in Figure 1(b). The graphs in (a) and (b) have equal probability under an edge exchangeable model.

But in much the same way that the species ‘names’ are disregarded when representing the classification of animals by a partition whose (unlabeled) blocks represent the class of animals of a given species, the essential structure of the call sequence is determined by the network structure obtained by disregarding the vertex ‘names’ in Figure 1. A key observation, which guides our development of the general theory below, is that exchangeability of the structure in Figure 1 is induced by the invariance of X with respect to relabeling its indices. In the vertex-edge labeled network representation, the indices labeling X correspond to the edge labels, thus giving rise to the concept of *edge exchangeable random graphs*, whose distributions are invariant with respect to arbitrary relabeling of their edges, as shown in Figure 2.

The principle of edge exchangeability was described in [13] and [14] as an alternative to the more conventional property of *vertex exchangeability* exhibited by graphon models, e.g. [6], [8], and [19]. A similar nonparametric Bayesian treatment of edge exchangeability was described in [9] and [10]. Edge exchangeability has a number of statistical and practical implications, which improve upon known drawbacks of vertex exchangeable models. For example, although real-world networks are widely believed to be sparse, it is well known that vertex exchangeable models are appropriate only for modeling networks that are ‘dense’. Edge exchangeable models, on the other hand, are able to handle sparsity in a natural and computationally tractable way. (See [14, Sections 4–6] for an illustration of the computation aspects of edge exchangeable models, in particular the instantiation of the Hollywood model in [14, Section 4].) Here we build upon these observations toward a more general theory of structures behaving in an analogous way to exchangeable random partitions and edge exchangeable random graphs.

To appreciate the new perspective offered by the theory of edge and relationally exchangeable networks presented here, it is important to clarify how the edge-centric perspective arises naturally out of our motivating example of phone call sampling. In the vertex-edge labeled

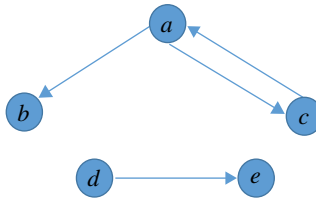


FIGURE 3: (a) Vertex-labeled network induced by disregarding edge labels in either Figure 1(a) or (b).

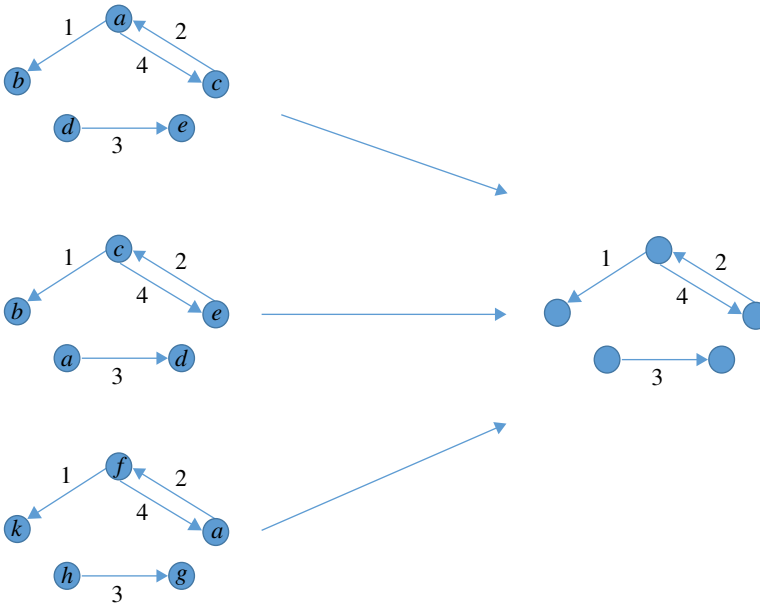


FIGURE 4: Three vertex-edge labeled networks with the same induced edge structure, as shown in the edge-labeled network on the right.

structure in Figure 1, it may seem natural to consider the vertex labels $a, b, c, d,$ and e as arbitrary ‘names’ chosen from S . Since the names are arbitrary, a standard line of reasoning may lead to the conclusion that the specific names do not matter, and thus that the model should be invariant under arbitrary reassignment of vertex names, as illustrated in Figure 3. But it is important to bear in mind that vertex exchangeability assumes not only distributional invariance with respect to relabeling of the sampled vertices, but also invariance with respect to relabeling of both sampled and unsampled vertices, thus implying that the observed vertices are *representative* of the population of all vertices (sampled and unsampled). In our running example, this is simply not the case, as uniform random sampling of telephone calls makes the observed vertices a *size-biased* sample according to their relative frequency of appearance in the database.

We conclude this section with a formal definition of the edge-labeled graphs shown in Figure 4, which serves as a precursor of our general framework of relational exchangeability below. Just as we disregarded the species names in passing from X to its induced equivalence relation ‘ \sim_X ’ in (1.1), we may also disregard the vertex labels in Figure 1 to obtain an *edge-labeled graph*, as shown in Figure 4. The edge-labeled graph is defined formally as the

equivalence class X_{\cong} of all sequences X' that yield the same structure as X after removing vertex labels:

$$X_{\cong} := \{X' : [n] \rightarrow S \times S \mid \rho X' = X \text{ for some bijection } \rho : S \rightarrow S\}. \tag{2.2}$$

Here we overload notation by allowing the bijection $\rho : S \rightarrow S$ to act on $S \times S$ by $(c, r) \mapsto (\rho(c), \rho(r))$, so that $\rho X' := \rho \circ X' : [n] \rightarrow S \times S$ is well defined by composition of functions. Exchangeability of X immediately implies *edge exchangeability* of the edge-labeled graph associated to X_{\cong} .

Though the definition in (2.2) may appear foreboding at first, we stress that it is merely the extension of (1.1) to sequences of ordered pairs. In particular, if we let $X : [n] \rightarrow S$ be a sequence of species labels, as in the discussion preceding (1.1), then the analog to (2.2) is

$$X_{\cong} = \{X' : [n] \rightarrow S \mid (\rho(X'_i))_{i \in S} = (X_i)_{i \in S} \text{ for some bijection } \rho : S \rightarrow S\},$$

which is exactly equivalent to the relation ‘ \sim_X ’ defined in (1.1). We explain this notation further in our technical treatment given in Section 4. But first we present additional examples to illustrate the breadth of the framework presented here.

3. Multiway interaction networks: extending the edge-centric perspective

In the previous section we discussed data structures constructed from sampled phone calls involving exactly two individuals, a caller and receiver. But many real-world networks are constructed from interactions involving two or more constituent elements. For example, let S represent a set of scientists, and let $X = (X_1, \dots, X_n)$ record names of coauthors on a collection of n scientific journal articles sampled uniformly from a database, such as arXiv. Then each X_i is a finite subset $\{s_{i,1}, \dots, s_{i,k_i}\} \subset S$ and the data sequence X_1, \dots, X_n gives rise to a vertex-hyperedge labeled structure.

Suppose that we observe the following sequence of ordered pairs obtained by the above procedure with $n = 4$:

$$X_1 = \{a, b, c\}, \quad X_2 = \{a, c\}, \quad X_3 = \{b, d, e\}, \quad X_4 = \{a, d, f, g\}. \tag{3.1}$$

As with the example from the previous section, the vertex labels serve only to distinguish among the observed elements. Once the relations of all observed vertices are recorded, through the induced network structure, these vertex labels can be disregarded. So while the data is observed in the form of a sequence, the essential structure can be represented by the hypergraph-labeled structure on the right-hand side of Figure 5.

The assumption that articles were observed by uniform sampling again makes X an exchangeable sequence, which in turn induces exchangeability on the associated hyperedge-labeled network constructed by disregarding vertex labels in the induced network structure in a manner analogous to the construction of edge-labeled graphs from the equivalence class (2.2). See Figure 6 for an illustration.

3.1. Examples of relational structures

Here we briefly describe additional relational structures that commonly arise in modern statistical applications and share the same essential structure as the above two examples. Each example points toward the basic structure of relationally exchangeable networks given in the next section. The α -structures of Section 4 unify these various structures into a single formal mathematical framework for studying relationally exchangeable structures.

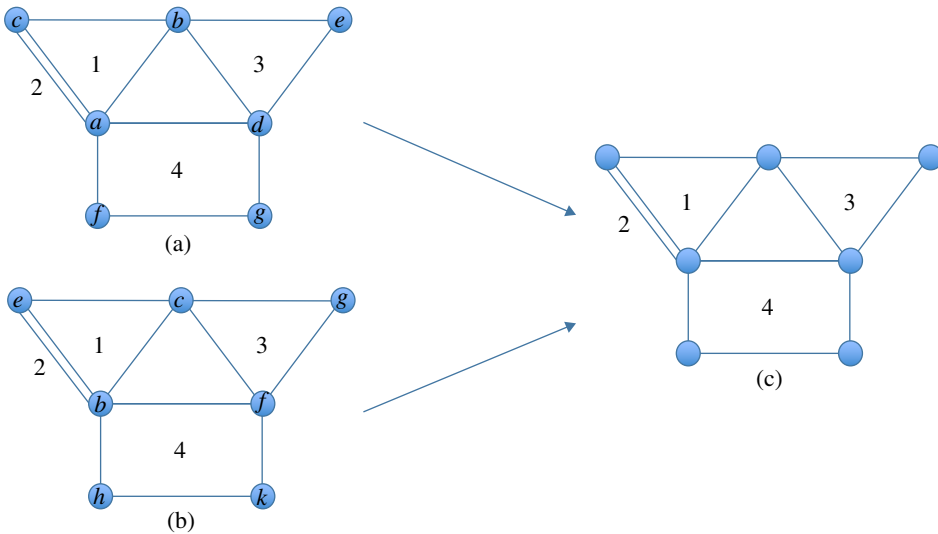


FIGURE 5: (a) Network representation of the sequence in (3.1). Note that the ‘triangle’ with vertices $a, b,$ and c is labeled 1, the ‘line’ with ends a and c is labeled 2, the ‘triangle’ with vertices $b, d,$ and e is labeled 3, and the ‘rectangle’ with vertices $a, d, f,$ and g is labeled 4. (b) Network representation of the sequence $X_1 = (b, c, e), X_2 = (b, e), X_3 = (c, f, g), X_4 = (b, f, h, k)$. (c) Edge-labeled hypergraph induced by removing vertex labels from the networks in both (a) and (b).

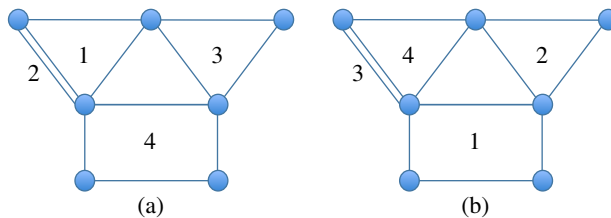


FIGURE 6: Two edge-labeled hypergraphs equivalent up to reordering of their edge labels. The graphs in (a) and (b) have equal probability under a hyperedge exchangeable model.

We start with a similar example to the previous scientific collaboration network. Let S represent a set of individuals, and let $X = (X_1, \dots, X_n)$ record names of senders and recipient lists of n emails sampled uniformly from an email database. Each X_i is an ordered pair $(s_i, (r_{i1}, \dots, r_{ik_i}))$ consisting of a single element of S (the sender) and a vector of elements of S of some finite, arbitrary length (the recipients). In an email network, we could further classify the recipients as ‘direct recipients’ (i.e. listed in the ‘to’ field of the email), or recipients who were in the ‘cc’ or ‘bcc’ field of the email. This additional classification adds structure to the data. Our more general treatment of α -structures is helpful to represent this structure.

For another example, let S be a set of Internet Protocol (IP) addresses and let each entry of $X = (X_1, \dots, X_n)$ correspond to the path traversed by a message sent between two IP addresses over the Internet. Each X_i corresponds to a path $(s_i, a_{i,1}, \dots, a_{i,k_i}, t_i)$ from source s_i to target t_i by passing through the intermediate nodes $a_{i,1}, \dots, a_{i,k_i}$, altogether indicating that the path from s to $a_{i,1}$ to $a_{i,2}$ and so on until passing from s_{i,k_1} to t_i . If X was obtained, for example, by uniform random sampling of source-target pairs (s_i, t_i) and then by applying an algorithm, such

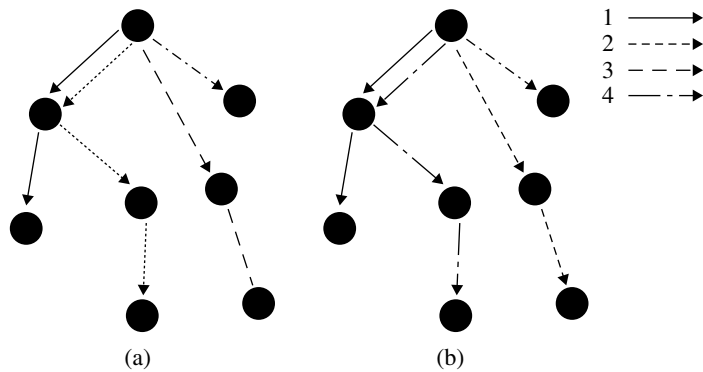


FIGURE 7: Both (a) and (b) correspond to structures induced by taking the equivalence class over paths between vertices. In this illustration, all paths have the same source vertex and each path is labeled by a line type, so that the structure in (b) can be obtained by re-typing the paths from (a) (i.e. type 2 to 4, 3 to 2, and 4 to 3).

as traceroute, to obtain a path between s_i and t_i , then the sequence of paths X is exchangeable and, therefore, so is the induced path-labeled structure shown in Figure 7. Note in Figure 7 that each path is labeled by a distinct line type, thus preserving the structure of the data in the same manner that it was observed. The more conventional approach to representing network data as a graph loses information about the dependency of the network on the path structure.

Another example in the realm of networks is to take each X_i to be an r -step ego network, obtained, e.g. by snowball sampling a neighborhood of size r from a randomly chosen vertex. The relationally labeled structure is constructed by piecing together the neighborhoods obtained from the r -neighborhoods of n randomly chosen ‘egos’ in this population. We omit the details of this example and instead move on to our general treatment.

Myriad other data structures arise according to a similar recipe as those above: let S be a set of elements, and let \mathcal{R} be a set of relations on the elements in S ; sample an exchangeable sequence X taking values in \mathcal{R} ; construct a structure from X (as in Figures 1 and 5) and obtain the corresponding relationally labeled structure by removing the vertex labels (as in Figures 4, 6, and 7). The resulting structure is exchangeable with respect to relabeling of its relations, a property which we call *relational exchangeability*. We now define this notion formally and prove a generic structure theorem for infinite relationally exchangeable structures. The formal statement is given in Theorem 4.1. Our formal definition of relational structures and proof of the associated structural theorem offer a rigorous probabilistic foundation for statistical modeling of relational structures.

4. Relational exchangeability

Let (S, \mathcal{S}) be any Borel space. A countably infinite sequence of S -valued random variables $X = (X_1, X_2, \dots)$ is *exchangeable* if $X^\sigma \stackrel{D}{=} X$ for every permutation $\sigma : \mathbb{N} \rightarrow \mathbb{N}$, where $X^\sigma := (X_{\sigma(1)}, X_{\sigma(2)}, \dots)$ is the reordering of X according to σ . By de Finetti’s theorem [16], the distribution of every such countably infinite sequence can be expressed as a mixture of independent, identically distributed (i.i.d.) sequences. In particular, with $\mathcal{P}(S)$ denoting the space of probability measures on (S, \mathcal{S}) , there is a unique probability measure ϕ on $\mathcal{P}(S)$ such that the distribution of X can be expressed as

$$\mathbb{P}(X \in \cdot) = \int_{\mathcal{P}(S)} \nu^\infty(\cdot) \phi(d\nu), \tag{4.1}$$

where ν^∞ denotes the infinite product measure induced by ν on S^∞ . We call the measure ϕ in (4.1) the *de Finetti measure* of X .

De Finetti’s theorem figures into our treatment of relationally exchangeable structures, which we now define. Note that all of the previous examples, and many more that could arise in practice, involve relational structures as they are defined in the coming section. Example 4.1 provides a concrete illustration.

4.1. Relational structures

Let $\alpha = \{R_1, R_2, \dots\}$ be a countable collection of relation symbols, and let $\alpha : \mathbb{N} \rightarrow \mathbb{N} \cup \{0\}$ be a *signature* such that $\alpha(j) = 0$ implies that $\alpha(k) = 0$ for all $k \geq j$. An α -*structure with domain* A is a collection $\mathcal{A} = (A; R_1^{\mathcal{A}}, R_2^{\mathcal{A}}, \dots)$, where A is a set and each $R_j^{\mathcal{A}} \subseteq A^{\alpha(j)}$ is a relation of arity $\alpha(j)$ on A . We adopt the convention that $A^0 := \emptyset$, so that $R_j^{\mathcal{A}} = \emptyset$ whenever $\alpha(j) = 0$. When discussing these structures below, we may sometimes leave the domain implicit, writing $\mathcal{A} = (R_1^{\mathcal{A}}, R_2^{\mathcal{A}}, \dots)$ if no confusion will result.

Remark 4.1. Although the α -structures defined above seem to be nonstandard in the probability literature, they are a standard object in the mathematical field of model theory, e.g. [18], and they provide a useful setting in which to study general exchangeable structures, as we do here. For example, the framework of α -structures has proven a powerful technique in very recent studies of exchangeable structures in theoretical and applied probability theory [2], [4], [3], [12], [15]. In our treatment below, we leverage this theory to study a general class of structures that arise naturally in applied probability and statistics.

Note that the domain A is for a *particular* α -structure. In the remainder of the text, the population of constituent elements take labels in \mathbb{Z} . Therefore, every α -structure has domain A that is a finite subset of \mathbb{Z} . An α -structure pertaining to a caller–receiver pair, for example, will have domain A of cardinality two; that is, A is the set of labels used to indicate the two individuals involved in the phone call. Of course, the actual elements of \mathbb{Z} used to identify the particular caller and receiver are immaterial. For a set of α -structures, we are simply interested in the relations, not the identities of the constituent elements. Therefore, we will investigate the equivalence class of sets of α -structures with the same inherent relational structure.

We write $\text{dom } \mathcal{A} := A$ for the *domain* of \mathcal{A} and FIN_α for the set of all α -structures \mathcal{A} with finite $\text{dom } \mathcal{A} \subset \mathbb{Z} = \{\dots, -1, 0, 1, \dots\}$ such that $\sum_{j=1}^\infty |R_j^{\mathcal{A}}| < \infty$, where $|S|$ denotes the cardinality of a set S . Writing $\text{dom } \mathcal{A} \subset_f \mathbb{Z}$ to denote that $\text{dom } \mathcal{A}$ is a finite subset of \mathbb{Z} , we have

$$\text{FIN}_\alpha := \left\{ \mathcal{A} = (A; R_1^{\mathcal{A}}, R_2^{\mathcal{A}}, \dots) : \text{dom } \mathcal{A} \subset_f \mathbb{Z} \text{ and } \sum_{j=1}^\infty |R_j^{\mathcal{A}}| < \infty \right\}.$$

The condition $\sum_{j \geq 1} |R_j^{\mathcal{A}}| < \infty$ implies that to every $\mathcal{A} \in \text{FIN}_\alpha$ there is an $r_{\mathcal{A}} := \max\{k \geq 1 : R_k^{\mathcal{A}} \neq \emptyset\} < \infty$, making it convenient to sometimes express $\mathcal{A} \in \text{FIN}_\alpha$ as $(\text{dom } \mathcal{A}; R_1^{\mathcal{A}}, \dots, R_{r_{\mathcal{A}}}^{\mathcal{A}})$ by omitting the infinite sequence of empty relations after $r_{\mathcal{A}}$. We note that the set FIN_α is countable.

For any $A^* \supseteq A$ and any injection $\rho : A^* \rightarrow A'$, there is an induced action on $\mathcal{A} = (A; R_1^{\mathcal{A}}, R_2^{\mathcal{A}}, \dots)$ by $\mathcal{A} \mapsto \rho(\mathcal{A}) = (\rho(A); R_1^{\rho(\mathcal{A})}, R_2^{\rho(\mathcal{A})}, \dots)$, with $\text{dom } \rho(\mathcal{A}) = \rho(A) := \{\rho(a) : a \in A\}$, and

$$(\rho(a_1), \dots, \rho(a_{\alpha(j)})) \in R_j^{\rho(\mathcal{A})} \text{ if and only if } (a_1, \dots, a_{\alpha(j)}) \in R_j^{\mathcal{A}} \tag{4.2}$$

for each $j = 1, 2, \dots$

For $n \geq 1$ and any subset of finite α -structures $\mathcal{R} \subseteq \text{FIN}_\alpha$ for which each $\mathcal{A} \in \mathcal{R}$ has $\text{dom } \mathcal{A} \subseteq_f \mathbb{Z}$, we consider the set $\mathcal{R}_{[n]}^{\cong}$ of equivalence classes of \mathcal{R} -valued sequences $\mathbf{x} : [n] \rightarrow \mathcal{R}$ obtained as follows. Any $\mathbf{x} : [n] \rightarrow \mathcal{R}$ determines a *relationally labeled structure* by removing the labels of the elements contained in each of the $\mathbf{x}(i)$ while maintaining the integrity of the overall structure induced by \mathbf{x} . Formally, we define the *relationally labeled structure induced by \mathbf{x}* as the equivalence class

$$\mathbf{x}_{\cong} := \{\mathbf{x}' : [n] \rightarrow \mathcal{R} \mid \rho \mathbf{x}' = \mathbf{x} \text{ for some bijection } \rho : \mathbb{Z} \rightarrow \mathbb{Z}\}, \tag{4.3}$$

where $\rho \mathbf{x} : [n] \rightarrow \mathcal{R}$ is defined by $(\rho \mathbf{x})(i) = \rho(\mathbf{x}(i))$ as in (4.2). We write $\mathcal{R}_{[n]}^{\cong}$ as the set of all \mathbf{x}_{\cong} constructed from some $\mathbf{x} : [n] \rightarrow \mathcal{R}$ in this way.

Remark 4.2. (*Notation.*) Below we reserve lowercase letters x, y, \dots for generic (nonrandom) objects and uppercase letters X, Y, \dots for random objects. We use bold letters $\mathbf{x}, \mathbf{y}, \dots$ for generic (nonrandom) functions $\mathbf{x} : [n] \rightarrow \mathcal{R}$ and $\mathbf{X}, \mathbf{Y}, \dots$ for random functions $\mathbf{X} : [n] \rightarrow \mathcal{R}$.

Example 4.1. In this example, we demonstrate how the modern data structures introduced in Section 1 fit into the framework of α -structures. For this, we write $\mathcal{A} = (A; R_1^A, R_2^A, \dots)$ as an α -structure with $\text{dom } \mathcal{A} = A$ and α varying according to the context.

- (a) For $\alpha(1) = 1$ and $\alpha(k) = 0$ for $k \geq 2$, each α -structure \mathcal{A} is determined by a subset $R_1^A \subseteq A$. If we then choose $\mathcal{R} \subseteq \text{FIN}_\alpha$ to consist of all singleton subsets of \mathbb{Z} , so that each R_1^A has the form $\{i\}$, $i \in \mathbb{Z}$, then the equivalence class of $\mathbf{x} : [n] \rightarrow \mathcal{R}$ in (4.3) corresponds to the equivalence relation induced by \mathbf{x} as in (1.1).
- (b) For $\alpha(1) = 2$ and $\alpha(k) = 0$ for $k \geq 2$, each α -structure \mathcal{A} with $\text{dom } \mathcal{A} = A$ corresponds to a binary relation $R_1^A \subseteq A^2$. Taking $\mathcal{R} \subseteq \text{FIN}_\alpha$ to consist of all structures with $R_1^A = \{(i, j)\}$ for $i \neq j$ corresponds to sampling from the phone call database in Section 1: each sampled $X_k = \{(i, j)\}$ corresponds to a caller–receiver pair (i, j) . The equivalence class \mathbf{x}_{\cong} gives an edge-labeled (directed) graph as in Figure 2. (To get an undirected graph, we would take each R_1^A to consist of symmetric pairs $\{(i, j), (j, i)\}$.)
- (c) Taking $\alpha(j) = j$ for all $j \geq 1$ means that any α -structure is determined by a collection $R_j^A \subseteq A^j$ of subsets of j -tuples for every $j \geq 1$. If we take $\mathcal{R} \subseteq \text{FIN}_\alpha$ to consist of only those α -structures of the form $R_k^A = \{(a_{\sigma(1)}, a_{\sigma(2)}, \dots, a_{\sigma(k)}) : \text{permutations } \sigma : [k] \rightarrow [k]\}$ for some $k \geq 1$ and $R_j^A = \emptyset$ for all $j \neq k$, then the elements $\mathcal{A} \in \mathcal{R}$ can be used to represent the set of coauthors in a sample of articles.
- (b) A small alteration of the previous α -structure can be used to represent the sender–receiver pair list in a sample of emails. Take $\alpha(1) = 1$ and $\alpha(j+1) = j$ for all $j \geq 2$. Then take $\mathcal{R} \subseteq \text{FIN}_\alpha$ to consist of only those α -structures of the form $R_1^A = \{(a_1)\}$ and $R_{k+1}^A = \{(a_{\sigma(1)}, a_{\sigma(2)}, \dots, a_{\sigma(k)}) : \text{permutations } \sigma : [k] \rightarrow [k]\}$ for some $k \geq 2$ and $R_j^A = \emptyset$ for all $j \neq k \geq 2$.
- (e) Also for $\alpha(j) = j$ for all $j \geq 1$, if we take $\mathcal{R} \subseteq \text{FIN}_\alpha$ to consist of only those α -structures of the form $R_k^A = \{(a_1, a_2, \dots, a_k)\}$ for some $k \geq 1$ and $R_j^A = \emptyset$ for all $j \neq k$, then each $\mathcal{A} \in \mathcal{R}$ corresponds to a path from a_1 to a_k , as discussed at the end of Section 1.

Note that the above formalism applies to structures with multiple relations, e.g. a graph with a distinguished community of vertices would have $\alpha(1) = 2$ (for the binary relation of edges), $\alpha(2) = 1$ (for the distinguished subset of vertices), and $\alpha(k) = 0$ for $k \geq 3$.

Example 4.1 demonstrates that α -structures are well-defined mathematical objects which compactly represent a wide range of relational structures prevalent in modern statistical settings. Examples of α -structures include graphs, hypergraphs, partitions, unary relations, and much more general structures. Our main theorem, Theorem 4.1, establishes a generic construction for infinitely exchangeable sequences of relational structures. This theorem will provide a necessary probabilistic foundation for research into statistical modeling of modern relational datasets.

4.2. Relational exchangeability

Our discussion below considers the case of random structures labeled by the countable set \mathbb{N} , which are those structures constructed just as in (4.3) but for a countable sequence $\mathbf{x} : \mathbb{N} \rightarrow \mathcal{R}$. It is important to note the difference between the index set \mathbb{N} and the domain of the structures in \mathcal{R} , which we take to be \mathbb{Z} . The domain labeling is ‘quotiented out’ in (4.3), while the indexing \mathbb{N} remains, serving as the ‘edge labels’ in the associated interpretation as an edge-labeled graph in Figure 2.

We write $\mathcal{R}_{\mathbb{N}}^{\cong}$ to denote the set of such structures and we equip $\mathcal{R}_{\mathbb{N}}^{\cong}$ with the Borel σ -field associated to its product-discrete topology induced by the following metric. For any $\mathbf{x}_{\cong} \in \mathcal{R}_{\mathbb{N}}^{\cong}$ and $n \geq 1$, we define the restriction $\mathbf{R}_n \mathbf{x}_{\cong}$ of \mathbf{x}_{\cong} to $\mathcal{R}_{[n]}^{\cong}$ as the structure obtained by taking any $\mathbf{x}' \in \mathbf{x}_{\cong}$ (i.e. $\mathbf{x}' : \mathbb{N} \rightarrow \mathcal{R}$ for which there exists ρ such that $\rho \mathbf{x}' = \mathbf{x}$), restricting its domain to $[n]$ by $\mathbf{x}'_{|[n]} : [n] \rightarrow \mathcal{R}, i \mapsto \mathbf{x}'(i)$, and putting $\mathbf{R}_n \mathbf{x}_{\cong} := (\mathbf{x}'_{|[n]})_{\cong}$ as in (4.3). It is clear from the definition of $\mathcal{R}_{\mathbb{N}}^{\cong}$ that this is well defined and does not depend on the specific choice of $\mathbf{x}' \in \mathbf{x}_{\cong}$. We then define the metric on $\mathcal{R}_{\mathbb{N}}^{\cong}$ by

$$d(\mathbf{x}_{\cong}, \mathbf{x}'_{\cong}) := \frac{1}{1 + \sup\{n \geq 1 : \mathbf{R}_n \mathbf{x}_{\cong} = \mathbf{R}_n \mathbf{x}'_{\cong}\}}, \quad \mathbf{x}_{\cong}, \mathbf{x}'_{\cong} \in \mathcal{R}_{\mathbb{N}}^{\cong},$$

with the convention that $1/\infty = 0$, under which $\mathcal{R}_{\mathbb{N}}^{\cong}$ is compact.

For any permutation $\sigma : \mathbb{N} \rightarrow \mathbb{N}$, we define the *relabeling* of $\mathbf{x}_{\cong} \in \mathcal{R}_{\mathbb{N}}^{\cong}$ by σ as the structure $\mathbf{x}_{\cong}^{\sigma}$ obtained by first choosing any $\mathbf{x}' \in \mathbf{x}_{\cong}$ and putting $\mathbf{x}_{\cong}^{\sigma} := (\mathbf{x}' \circ \sigma)_{\cong}$, where $\mathbf{x}' \circ \sigma : \mathbb{N} \rightarrow \mathcal{R}$ is defined by the usual composition of functions, $(\mathbf{x}' \circ \sigma)(i) := \mathbf{x}'(\sigma(i))$. It is once again clear that this does not depend on the specific choice of representative $\mathbf{x}' \in \mathbf{x}_{\cong}$ since the actions of $\rho : \mathbb{N} \rightarrow \mathbb{N}$ and $\sigma : \mathbb{N} \rightarrow \mathbb{N}$ commute for all \mathbf{x}' , i.e. $\rho(\mathbf{x}' \circ \sigma) = (\rho \mathbf{x}') \circ \sigma$.

Definition 4.1. (*Relational exchangeability.*) A random structure $\mathbf{X}_{\cong} \in \mathcal{R}_{\mathbb{N}}^{\cong}$ is *relationally exchangeable* if $\mathbf{X}_{\cong}^{\sigma} \stackrel{D}{=} \mathbf{X}_{\cong}$ for all permutations $\sigma : \mathbb{N} \rightarrow \mathbb{N}$.

4.3. Representation theorem

Our main theorem establishes a generic construction for infinite relationally exchangeable structures $\mathbf{X}_{\cong} \in \mathcal{R}_{\mathbb{N}}^{\cong}$. The construction proceeds by sampling a sequence X_1, X_2, \dots conditionally i.i.d. in a related space \mathcal{R}^* and then modifying these observations to obtain a new sequence $\mathbf{X}^{\dagger} := (X_1^{\dagger}, X_2^{\dagger}, \dots)$. We then construct the equivalence class $\mathbf{X}_{\cong}^{\dagger}$ based on the sequence $X_1^{\dagger}, X_2^{\dagger}, \dots$ as in (4.3). Since stating the theorem formally requires some new ideas and notation, we give the construction and key ideas of the proof prior to stating the result in Theorem 4.1.

As above, let $\alpha : \mathbb{N} \rightarrow \mathbb{N} \cup \{0\}$, and let $\mathcal{R} \subseteq \text{FIN}_{\alpha}$ be a set of α -structures such that $\alpha(j) = 0$ implies that $\alpha(k) = 0$ for all $k \geq j$ and each $\mathcal{A} \in \mathcal{R}$ has $\text{dom } \mathcal{A} \subseteq_f \mathbb{Z}$. Since \mathcal{R} is

at most countable, we may fix an ordering $\mathcal{R} = \{\mathcal{S}_n\}_{n \geq 1}$ of its elements. Given any $\mathcal{A} = (\text{dom } \mathcal{A}; R_1^{\mathcal{A}}, R_2^{\mathcal{A}}, \dots) \in \text{FIN}_\alpha$ and any $[0, 1]$ -valued sequence $(T_i)_{i \in \mathbb{Z}}$, we define $T\mathcal{A} := T \circ \mathcal{A}$ as the α -structure $T\mathcal{A}$ with $\text{dom}(T\mathcal{A}) := \{T_i : i \in \text{dom } \mathcal{A}\}$ and relations $R_j^{T\mathcal{A}}$ given by

$$(T_{a_1}, \dots, T_{a_{\alpha(j)}}) \in R_j^{T\mathcal{A}} \quad \text{if and only if} \quad (a_1, \dots, a_{\alpha(j)}) \in R_j^{\mathcal{A}} \tag{4.4}$$

for each $j = 1, 2, \dots$. We then define

$$\mathcal{S}_{[0,1]} := \bigcup_{n \geq 1} \{T\mathcal{S}_n : T = (T_i)_{i \in \mathbb{Z}} \in [0, 1]^{\mathbb{Z}}\}$$

as the set of α -structures obtained by associating $[0, 1]$ -valued labels to the elements of \mathcal{R} . More generally, if $(\Xi_i)_{i \in \text{dom } \mathcal{A}}$ is a collection of subsets $\Xi_i \subseteq [0, 1]$, then we define $\Xi\mathcal{A}$ by

$$\Xi\mathcal{A} := \{T\mathcal{A} : T_i \in \Xi_i \text{ for each } i \in \text{dom } \mathcal{A}\}.$$

We equip $\mathcal{S}_{[0,1]}$ with the σ -field on $\mathcal{S}_{[0,1]}$ generated by all sets of the form $\Xi\mathcal{A}$ with $\mathcal{A} \in \mathcal{R}$ and $\Xi = (\Xi_i)_{i \in \text{dom } \mathcal{A}}$ a collection of Borel subsets of $[0, 1]$.

From any $\mathbf{x} \cong \in \mathcal{R}_{\mathbb{N}}^{\cong}$ and any sequence $\xi = (\xi_i)_{i \in \mathbb{Z}}$ of i.i.d. Uniform $[0, 1]$ random variables, we write $\xi \mathbf{x} \cong$ to denote a random $\mathcal{S}_{[0,1]}$ -valued sequence $(Z_i)_{i \geq 1}$ obtained by first taking any representative $\mathbf{x}' \in \mathbf{x} \cong$ and then putting $Z_i = \xi \mathbf{x}'(i)$ for each $i \geq 1$, with $\xi \mathbf{x}'(i)$ defined as in (4.4). Since $\mathbf{x} \cong$ is fixed in this example, each $\xi \mathbf{x}'(i)$ corresponds to an assignment of Uniform $[0, 1]$ random labels to the elements in the domain of $\mathbf{x}'(i)$. Since $\xi_i \neq \xi_j$ with probability 1 for all $i \neq j$, it is immediate that the distribution of $\xi \mathbf{x} \cong$ does not depend on the manner in which representative \mathbf{x}' is chosen.

Now, for any infinite relationally exchangeable structure $\mathbf{X} \cong \in \mathcal{R}_{\mathbb{N}}^{\cong}$ and a sequence ξ of Uniform $[0, 1]$ random variables independent of $\mathbf{X} \cong$, $\xi \mathbf{X} \cong = (Z_i)_{i \geq 1}$ defines a random sequence obtained by putting $Z_i = \xi \mathbf{x}'(i)$ for a representative $\mathbf{x}' \in \mathbf{x} \cong$ on the event $\mathbf{X} \cong = \mathbf{x} \cong$. By exchangeability of $\mathbf{X} \cong$, the sequence $\mathbf{Z} = (Z_i)_{i \geq 1}$ obtained in this way is also exchangeable and de Finetti's theorem implies that \mathbf{Z} is distributed as an i.i.d. sequence from a random measure ν on $\mathcal{S}_{[0,1]}$, as described in (4.1).

Given ν , we define the *propensity of $u \in [0, 1]$ in \mathbf{Z}* by

$$\nu^*(u) := \nu(\{\mathcal{A} \in \mathcal{S}_{[0,1]} : u \in \text{dom } \mathcal{A}\}), \tag{4.5}$$

which equals the conditional probability of the event $\{u \in \text{dom } Z_j\}$ given ν for each $j \geq 1$. It is clear that, with probability 1, each $u \in [0, 1]$ appears in either 0, 1, or infinitely many of the relations $(Z_i)_{i \geq 1}$. First, if $\nu^*(u) > 0$ then the strong law of large numbers implies that u occurs in infinitely many of the relations Z_i with proportion $\nu^*(u)$. If $\nu^*(u) = 0$ and $u \in Z_i$ for some i , then the probability that u occurs in Z_j is $\nu^*(u) = 0$ independently for each $j \neq i$ and, therefore, u appears only once in \mathbf{Z} with probability 1.

Clearly, the set $\mathcal{U} = \{u : \nu^*(u) > 0\}$ is at most countable since $|\text{dom}(Z_i)| < \infty$ for each $i = 1, 2, \dots$. We can, therefore, order the elements of \mathcal{U} as u_1, u_2, \dots such that $\nu^*(u_j) \geq \nu^*(u_{j+1})$ for $j \geq 1$, breaking ties $\nu^*(v) = \nu^*(w)$ as follows. For each $\mathcal{A} \in \mathcal{R}$, we define

$$\nu^*(u; \mathcal{A}) := \nu(\{T\mathcal{A} : T = (T_i)_{i \in \mathbb{Z}} \in [0, 1]^{\mathbb{Z}} \text{ such that } u \in \text{dom}(T\mathcal{A})\}), \quad u \in [0, 1],$$

to be the measure assigned to the subset of $\mathcal{S}_{[0,1]}$ whose structure is consistent with \mathcal{A} and which contains u in its domain. Assuming that $\nu^*(v) = \nu^*(w)$, we assign the smaller label to

v if there is some $k \geq 1$ such that $v^*(v; \mathcal{S}_j) = v^*(w; \mathcal{S}_j)$ for all $j < k$ and $v^*(v; \mathcal{S}_k) > v^*(w; \mathcal{S}_k)$. If $v^*(v; \mathcal{S}_j) = v^*(w; \mathcal{S}_j)$ for all $j \geq 1$ then we label v and w in increasing order, i.e. if we are to assign labels j and $j + 1$ to v and w and if $v < w$, then we shall put $u_j = v$ and $u_{j+1} = w$; otherwise, we put $u_j = w$ and $u_{j+1} = v$. The ordering $\mathcal{U} = (u_j)_{j \geq 1}$ is thus uniquely determined by v and the fixed ordering of \mathcal{R} chosen at the outset.

Now given \mathcal{U} , for any $\mathcal{A} \in \mathcal{S}_{[0,1]}$, we define \mathcal{A}^* (suppressing the dependence on v) by replacing each occurrence of $u_j \in \mathcal{U}$ by j and replacing each occurrence of $v' \notin \mathcal{U}$ in \mathcal{A} by a unique nonpositive integer $z(v') = 0, -1, -2, \dots$ so that, for $v', v'' \notin \mathcal{U}$ and both in $\text{dom } \mathcal{A}$ with $v' \leq v''$, we have $z(v') \leq z(v'')$ and the $z(v')$ are chosen to be the largest possible nonpositive integers that satisfy this condition. (For example, if $v_1 < v_2 < v_3$ are the only elements in $\text{dom } \mathcal{A}$ that are not in \mathcal{U} , we assign $z(v_3) = 0, z(v_2) = -1$, and $z(v_1) = -2$.)

Every $\mathcal{A} \in \mathcal{S}_{[0,1]}$ thus corresponds to a unique such \mathcal{A}^* and we define \mathcal{R}^* as the set of all structures \mathcal{A}^* obtained in this way. Note that, since \mathcal{R} is at most countable, so is \mathcal{R}^* . Also, although we have constructed \mathcal{R}^* using \mathcal{U} (and therefore v), the set \mathcal{R}^* depends only on \mathcal{R} , justifying the term \mathcal{R} -simplex in our definition of $\mathcal{F}_{\mathcal{R}}$ in (4.6) below.

The elements of \mathcal{R}^* are thus α -structures \mathcal{A}^* with $\text{dom } \mathcal{A}^* \subseteq \mathbb{Z}$. We define the \mathcal{R} -simplex $\mathcal{F}_{\mathcal{R}}$ by

$$\mathcal{F}_{\mathcal{R}} := \left\{ \left\{ (f_{\mathcal{B}})_{\mathcal{B} \in \mathcal{R}^*} : f_{\mathcal{B}} \geq 0 \text{ and } \sum_{\mathcal{B} \in \mathcal{R}^*} f_{\mathcal{B}} = 1 \right\} \right\}, \tag{4.6}$$

on which we equip the metric

$$d_{\mathcal{F}_{\mathcal{R}}}(f, f') := \sum_{\mathcal{B} \in \mathcal{R}^*} |f_{\mathcal{B}} - f'_{\mathcal{B}}|, \quad f, f' \in \mathcal{F}_{\mathcal{R}},$$

and the associated Borel σ -field.

Any $f \in \mathcal{F}_{\mathcal{R}}$ determines a unique probability measure ε_f on $\mathcal{R}_{\mathbb{N}}^{\cong}$ by first drawing $\mathbf{X} = (X_1, X_2, \dots)$ i.i.d. from

$$\mathbb{P}(X_i = \mathcal{B} \mid f) = f_{\mathcal{B}}, \quad \mathcal{B} \in \mathcal{R}^*, \tag{4.7}$$

and then constructing $\mathbf{X}^{\dagger} = (X_1^{\dagger}, X_2^{\dagger}, \dots)$ from \mathbf{X} as follows. We initialize by putting $m_0 = 0$. For each $n \geq 1$, given m_{n-1} , we replace the nonpositive elements in $\text{dom } X_n$ according to the following rule.

- (i) If $\text{dom } X_n$ has no nonpositive elements then put $m_n = m_{n-1}$ and $X_n^{\dagger} := X_n$.
- (ii) If $\text{dom } X_n$ has nonpositive elements $0, \dots, -k$ for $k \geq 0$ then define X_n^{\dagger} by replacing each occurrence of $-i$ for $0 \leq i \leq k$ in X_n by $m_{n-1} - i$, then putting $m_n = m_{n-1} - k - 1$ and keeping positive elements unchanged. See Example 4.2 below for an illustration.

Note the distinction between the use of nonpositive elements in the constructions of X_i and X_i^{\dagger} , respectively. The nonpositive elements of each $X_i \in \mathcal{R}^*$ serve to denote the nonrecurring particles that appear *only within* this particular relation. The nonpositive elements of X_i^{\dagger} , on the other hand, serve to denote the nonrecurring particles across *all* relations in the sequence \mathbf{X}^{\dagger} . We define ε_f to be the distribution of $\mathbf{X}_{\mathbb{N}}^{\dagger}$ constructed by applying (4.3) to the sequence \mathbf{X}^{\dagger} .

From v , we define $f^v = (f_{\mathcal{B}}^v)_{\mathcal{B} \in \mathcal{R}^*} \in \mathcal{F}_{\mathcal{R}}$ by

$$f_{\mathcal{B}}^v := v(\{\mathcal{A} \in \mathcal{S}_{[0,1]} : \mathcal{A}^* = \mathcal{B}\}), \quad \mathcal{B} \in \mathcal{R}^*. \tag{4.8}$$

This choice of $f^\nu = (f_B^\nu)_{B \in \mathcal{R}^*}$ is uniquely determined by ν and the fixed ordering of \mathcal{R} . Conversely, given $f = (f_B)_{B \in \mathcal{R}^*} \in \mathcal{F}_{\mathcal{R}}$, we construct a measure ν_f on $\mathcal{S}_{[0,1]}$ to be the distribution of ξY from (4.4) for Y drawn from distribution (4.7) and $\xi = (\xi_i)_{i \in \mathbb{Z}}$ i.i.d. Uniform $[0, 1]$ independent of Y . Proposition 4.1 proves that the above procedure does not alter the random, relationally labeled structure.

Proposition 4.1. *Let X_{\cong} be relationally exchangeable, and let $\theta = (\theta_i)_{i \in \mathbb{Z}}$ be i.i.d. Uniform $[0, 1]$ independent of X_{\cong} . Then $((\theta X_{\cong})^*)^\dagger \cong X_{\cong}$ a.s., where $(\theta X_{\cong})^*$ denotes the application of $*$: $\mathcal{S}_{[0,1]} \rightarrow \mathcal{R}^*$ to each component of the sequence θX_{\cong} .*

Proof. Let $\theta = (\theta_i)_{i \in \mathbb{Z}}$ be i.i.d. Uniform $[0, 1]$ independent of X_{\cong} . Each event $X_{\cong} = \mathbf{x}_{\cong}$ gives rise to a probability measure ν on $\mathcal{S}_{[0,1]}$ through the de Finetti measure of $\theta \mathbf{x}_{\cong}$; see (4.1). Let $\mathcal{U} = (u_i)_{i \geq 1}$ be the ordered subset of $[0, 1]$ corresponding to the atoms of ν^* as in (4.5). Since θ is i.i.d. Uniform $[0, 1]$, we have

$$\mathbb{P}\{\theta_i \neq \theta_j \text{ for all } i \neq j\} = 1,$$

which implies that distinct $i, j \in \mathbb{Z}$ are labeled distinctly in $\theta \mathbf{x}_{\cong}$ with probability 1. In particular, the nonpositive labels in $(\theta \mathbf{x}_{\cong})^*$ account for all those elements that appear in only one entry of the sequence $\theta \mathbf{x}_{\cong}$. It follows that $((\theta \mathbf{x}_{\cong})^*)^\dagger \in \mathbf{x}_{\cong}$ with probability 1 and, thus, $((\theta \mathbf{x}_{\cong})^*)^\dagger \cong \mathbf{x}_{\cong}$ with probability 1 for all possible outcomes $X_{\cong} = \mathbf{x}_{\cong}$. \square

Example 4.2. To illustrate the above procedure, let \mathcal{R} consist of the singleton sets $\{\{i\}\}$, $i \geq 1$, just as in Example 4.1(a), written as $\{1\}, \{2\}, \dots$ for simplicity. Let $\mathcal{U} = (u_i)_{i \geq 1}$ be a countable subset of $[0, 1]$. From $\mathcal{R} = \{\{1\}, \{2\}, \dots\}$, we obtain $\mathcal{R}^* = \{\{0\}, \{1\}, \dots\}$ since, for any sequence $T = (T_i)_{i \in \mathbb{Z}}$ and any $\{i\}$, $i \geq 1$, the transformed relation $T\{i\} \equiv \{T_i\}$ has either $T_i \in \mathcal{U}$ or $T_i \notin \mathcal{U}$. If $T_i = u_k \in \mathcal{U}$ then $\{T_i\}^* = \{k\}$ for $k \geq 1$; and if $T_i = v \notin \mathcal{U}$ then $\{T_i\}^* = \{\{0\}\}$.

Given $(f_{\{0\}}, f_{\{1\}}, \dots) \in \mathcal{F}_{\mathcal{R}}$, we generate a sequence X_1, X_2, \dots of singleton sets i.i.d. as in (4.7), from which we obtain $X_1^\dagger, X_2^\dagger, \dots$ by reassigning occurrences of 0 to the greatest negative integer that has not yet appeared in the sequence. For example, if the sequence begins $\{4\}, \{0\}, \{2\}, \{0\}, \{2\}, \{2\}, \{0\}, \dots$, then we reassign labels to obtain $\{4\}, \{0\}, \{2\}, \{-1\}, \{2\}, \{2\}, \{-2\}, \dots$. Delabeling according to (4.3) (or equivalently (1.1)) yields the equivalence classes $\{1\}, \{2\}, \{3, 5, 6\}, \{4\}, \{7\}$.

For another example, let \mathcal{R} consist of ordered pairs $\{(i, j)\}$, $1 \leq i \neq j < \infty$, so that we are in the phone call example of Section 1. Let $\mathcal{U} = (u_i)_{i \geq 1}$ be a countable subset of $[0, 1]$. Then

$$\mathcal{R}^* = \{\{(i, 0)\}: i \geq 1\} \cup \{\{(0, i)\}: i \geq 1\} \cup \{\{(0, -1)\}, \{(-1, 0)\}\} \cup \mathcal{R},$$

since any $T = (T_i)_{i \in \mathbb{Z}}$ and any $\{(i, j)\} \in \mathcal{R}$ is transformed by $T\{(i, j)\} = \{(T_i, T_j)\}$, which may have $T_i = u_{i'} \in \mathcal{U}$ and $T_j = u_{j'} \in \mathcal{U}$ in which case $\{(T_i, T_j)\}^* = \{(i', j')\}$. If $T_i = u_{i'} \in \mathcal{U}$ and $T_j \notin \mathcal{U}$, then $\{(T_i, T_j)\}^* = \{(i', 0)\}$. If $T_i \notin \mathcal{U}$ and $T_j = u_{j'}$ then $\{(T_i, T_j)\}^* = \{(0, j')\}$. If $T_i, T_j \notin \mathcal{U}$ then $\{(T_i, T_j)\}^*$ will equal $\{(0, -1)\}$ or $\{(-1, 0)\}$ depending on whether $T_j < T_i$ or $T_i < T_j$, respectively.

The above construction gives the following representation theorem.

Theorem 4.1. *Let α be a signature, $\mathcal{R} \subseteq \text{FIN}_\alpha$, and fix an ordering $\mathcal{R} = (\mathcal{S}_n)_{n \geq 1}$. Let X_{\cong} be a relationally exchangeable random structure in $\mathcal{R}_{\mathbb{N}}^{\cong}$. Then there exists a probability measure ϕ on $\mathcal{F}_{\mathcal{R}}$ such that $X_{\cong} \sim \varepsilon_\phi$, where*

$$\varepsilon_\phi(\cdot) = \int_{\mathcal{F}_{\mathcal{R}}} \varepsilon_f(\cdot) \phi(df).$$

A canonical version of the measure ϕ in Theorem 4.1 can be constructed as in (4.8), and this is the measure we construct in the following proof.

Proof of Theorem 4.1. We proceed by constructing ϕ as in (4.8) and showing that $\mathbf{Y}_{\cong}^{\dagger} \sim \varepsilon_{\phi}$ satisfies $\mathbf{Y}_{\cong}^{\dagger} \stackrel{D}{=} \mathbf{X}_{\cong}$.

To see this, we first let ψ be the de Finetti measure on $\mathcal{P}(\mathcal{S}_{[0,1]})$ (i.e. the space of probability measures on $\mathcal{S}_{[0,1]}$) associated to $\xi \mathbf{X}_{\cong}$ for $\xi = (\xi_i)_{i \in \mathbb{Z}}$ i.i.d. Uniform[0, 1] independent of \mathbf{X}_{\cong} . In particular, the distribution of $\xi \mathbf{X}_{\cong}$ is conditionally i.i.d. from $\nu \sim \psi$. The measure ψ determines a measure ϕ on $\mathcal{F}_{\mathcal{R}}$ through (4.8). We need to show that $\mathbf{Y}_{\cong}^{\dagger} \sim \varepsilon_{\phi}$ satisfies $\mathbf{Y}_{\cong}^{\dagger} \stackrel{D}{=} \mathbf{X}_{\cong}$.

Given $f \sim \phi$, we construct $\mathbf{Y} = (Y_1, Y_2, \dots)$ as conditionally i.i.d. from the distribution in (4.7) and \mathbf{Y}^{\dagger} as in (i) and (ii) above. Let $\mathcal{U} = (u_i)_{i \geq 1}$ be the atoms of $\theta \mathbf{Y}^{\dagger}$ for $\theta = (\theta_i)_{i \in \mathbb{Z}}$ i.i.d. Uniform[0, 1] independent of \mathbf{Y}^{\dagger} , let $\zeta = (\zeta_k)_{k \leq 0}$ be i.i.d. Uniform[0, 1] independent of everything else, and define $\xi_f = (\xi_i)_{i \in \mathbb{Z}}$ by

$$\xi_i = \begin{cases} u_i, & i \geq 1, \\ \zeta_i, & \text{otherwise.} \end{cases}$$

First note that the conditional distribution of $\xi_f \mathbf{Y}^{\dagger}$ given f is the same as the conditional law of $\theta \mathbf{Y}^{\dagger}$ given ν_f , since the ξ_i for $i \geq 1$ were constructed from the atoms of $\theta \mathbf{Y}^{\dagger}$. Writing $\mathcal{D}(X \mid Y)$ to denote the conditional distribution of X given Y , we thus have

$$\mathcal{D}(\xi_f \mathbf{Y}_{\cong}^{\dagger} \mid f) = \mathcal{D}(\theta \mathbf{Y}^{\dagger} \mid \nu_f), \tag{4.9}$$

$$\mathcal{D}(\xi_f \mathbf{Y}_{\cong}^{\dagger} \mid f) = \mathcal{D}(\theta \mathbf{X}_{\cong} \mid \nu_f), \tag{4.10}$$

where (4.10) follows from the construction of ϕ from the de Finetti measure ψ of $\xi \mathbf{X}_{\cong}$. From (4.9) and (4.10), it follows that

$$\mathcal{D}(\xi_f \mathbf{Y}_{\cong}^{\dagger} \mid f) = \mathcal{D}(\theta \mathbf{Y}^{\dagger} \mid \nu_f),$$

$$\mathcal{D}(\xi_f \mathbf{Y}_{\cong}^{\dagger} \mid f) = \mathcal{D}(\theta \mathbf{X}_{\cong} \mid \nu_f),$$

and, thus,

$$\mathcal{D}(\xi_f \mathbf{Y}_{\cong}^{\dagger} \mid f) = \mathcal{D}(\theta \mathbf{X}_{\cong} \mid \nu_f).$$

By Proposition 4.1, we have $\mathcal{D}(\theta \mathbf{Y}^{\dagger} \mid \nu_f) = \mathbf{Y}_{\cong}^{\dagger}$ a.s. and $\mathcal{D}(\theta \mathbf{X}_{\cong} \mid \nu_f) = \mathbf{X}_{\cong}$, implying that $\mathbf{Y}_{\cong}^{\dagger} \stackrel{D}{=} \mathbf{X}_{\cong}$, as desired. □

4.4. Special cases

As discussed in Example 4.2, the case in which \mathcal{R} corresponds to the singleton sets $\{i\}$ for $i \geq 1$ gives $\mathcal{F}_{\mathcal{R}}$ equal to the ranked simplex

$$\Delta^{\downarrow} := \left\{ (f_0, f_1, \dots) : f_1 \geq f_2 \geq \dots \geq 0, f_0 \geq 0, \sum_{i \geq 0} f_i = 1 \right\}.$$

To see this, note that $\mathcal{S}_{[0,1]}$ is the set of singleton elements of $[0, 1]$, i.e.

$$\mathcal{S}_{[0,1]} := \{ \{u\} : u \in [0, 1] \}.$$

An exchangeable sequence $X = (X_1, X_2, \dots)$ in $\mathcal{S}_{[0,1]}$ gives rise to a random countable subset $\mathcal{U} \subset [0, 1]$ of elements that appear infinitely often among the X_i and its complement $[0, 1] \setminus \mathcal{U}$ consisting of all elements appearing at most once among the X_i . In the construction of X^\dagger from X described above, any occurrence $X_i = \{u\}$ for $u \notin \mathcal{U}$ gives rise to $X_i^* = \{0\}$, explaining why the definition of $\mathcal{F}_{\mathcal{R}}$ in (4.6) corresponds to the simplex of elements $(f_i)_{i \geq 0}$.

Following Example 4.1(b), the case of edge exchangeable random (directed) graphs corresponds to $\mathcal{R} \subset \text{FIN}_\alpha$ with $\alpha(1) = 2$ and $\alpha(k) = 0$ for $k \geq 2$, with each $\mathcal{A} \in \mathcal{R}$ having $R_1^{\mathcal{A}} = \{(i, j)\}$, $i \neq j$. In this case, $\mathcal{S}_{[0,1]}$ corresponds to all pairs (u, v) for $u, v \in [0, 1]$. An exchangeable sequence X in $\mathcal{S}_{[0,1]}$ determines a random subset $\mathcal{U} \subset [0, 1]$. Now, in each X_i there can be 0, 1, or 2 elements $v \notin \mathcal{U}$. Occurrences of such elements are replaced by 0 and -1 , as explained in Example 4.2.

In general, we call the occurrences of any $u \notin \mathcal{U}$ in the sequence X *blips*. These elements are merely a ‘blip’ in the overall sequence X —they occur only for an instant and then never again. Our labeling convention in defining \mathcal{R}^* is that the nonpositive labels correspond to the blips. Theorem 4.1 describes how blips arise in general relationally exchangeable structures.

5. Concluding

Though relational structures are commonplace in modern statistics, the theory of relational exchangeability presented here stands apart from most other studies of exchangeable structures by de Finetti [16], Aldous [5], Hoover [19], Kallenberg [20], and others [3], [7], [15]. The most relevant prior work is Kingman’s on exchangeable random set partitions [21], which can be reinterpreted as relationally exchangeable structures in the sense described above.

Recent work in the statistical analysis of network data underscores the significance of relationally labeled structures in applications, as many data structures which are typically represented graphically, such as social networks and networks detailing email correspondence and professional collaborations, arise from a process by which interactions or relations accumulate within a population of otherwise indistinguishable individuals. The work in [14] focused on the case of edge exchangeable random graphs, but the additional examples in Sections 1–3 involving repeated path sampling and snowball sampling are also highly relevant in network applications.

Our main representation theorem serves two immediate statistical purposes. First, the representation characterizes a general class of nonparametric statistical models of potential interest in the aforementioned applications. Second, it establishes that vertices arrive in size-biased random order in relationally exchangeable structures, explaining why the common assumption of exchangeable vertex labeling, as presented in graphon models [23], is not tenable in many applications. We reserve discussion of these practical implications for other work; see [13] and [14].

References

- [1] ACHLIOPTAS, D., CLAUSET, A., KEMPE, D. AND MOORE, C. (2005). On the bias of traceroute sampling or, power-law degree distributions in regular graphs. In *Proc. 37th Annual ACM Symposium on Theory of Computing*, Association for Computing Machinery, New York, pp. 694–703.
- [2] ACKERMAN, N. (2015). Representations of $\text{Aut}(\mathcal{M})$ -invariant measures: part 1. Preprint. Available at <https://arxiv.org/abs/1509.06170v1>.
- [3] ACKERMAN, N., FREER, C. AND PATEL, R. (2014) Invariant measures concentrated on countable structures. Preprint. Available at <https://arxiv.org/abs/1206.4011v3>.
- [4] ACKERMAN, N., FREER, C., KRUCKMAN, A. AND PATEL, R. (2017). Properly ergodic structures. Preprint. Available at <https://arxiv.org/abs/1710.09336v1>.

- [5] ALDOUS, D. J. (1981). Representations for partially exchangeable arrays of random variables. *J. Multivariate Anal.* **11**, 581–598.
- [6] ALDOUS, D. J. (1985). Exchangeability and related topics. In *École d'été de Probabilités de Saint-Flour, XIII—1983* (Lecture Notes Math. **1117**), Springer, Berlin, pp. 1–198.
- [7] AUSTIN, T. AND PANCHENKO, D. (2014). A hierarchical version of the de Finetti and Aldous-Hoover representations. *Prob. Theory Relat. Fields* **159**, 809–823.
- [8] BICKEL, P. J. AND CHEN, A. (2009). A nonparametric view of network models and Newman–Girvan and other modularities. *Proc. Nat. Acad. Sci. USA* **106**, 21068–21073.
- [9] CAI, D. AND CAMPBELL, T. (2015). Edge-exchangeable graphs and sparsity. In *NIPS 2015 Workshop on Networks in the Social and Information Sciences*.
- [10] CAI, D., CAMPBELL, T. AND BRODERICK, T. (2016). Edge-exchangeable graphs and sparsity. *Neural Information Processing Systems*.
- [11] CRANE, H. (2016). The ubiquitous Ewens sampling formula. *Statist. Sci.* **31**, 1–19.
- [12] CRANE, H. (2018). Combinatorial Lévy processes. *Ann. Appl. Prob.* **28**, 395–339.
- [13] CRANE, H. AND DEMPSEY, W. (2016). A framework for statistical network modeling. Preprint. Available at <https://arxiv.org/abs/1509.08185v4>.
- [14] CRANE, H. AND DEMPSEY, W. (2018). Edge exchangeable models for interaction networks. *J. Amer. Statist. Assoc.* **113**, 1311–1326.
- [15] CRANE, H. AND TOWNSNER, H. (2018). Relatively exchangeable structures. *J. Symbolic Logic* **83**, 416–442.
- [16] DE FINETTI, B. (1937). La prévision: ses lois logiques, ses sources subjectives. *Ann. Inst. H. Poincaré* **7**, 1–68.
- [17] EWENS, W. J. (1972). The sampling theory of selectively neutral alleles. *Theoret. Pop. Biol.* **3**, 87–112.
- [18] HODGES, W. (1993). *Model Theory* (Encyclopedia Math. Appl. **42**). Cambridge University Press.
- [19] HOOVER, D. (1979). Relations on probability spaces and arrays of random variables. Preprint. Institute for Advanced Studies, Princeton, NJ.
- [20] KALLENBERG, O. (2005). *Probabilistic Symmetries and Invariance Principles*. Springer, New York.
- [21] KINGMAN, J. F. C. (1978). Random partitions in population genetics. *Proc. R. Soc. London A* **361**, 1–20.
- [22] KLIMT, B. AND YANG, Y. (2004). Introducing the enron corpus. *CEAS*. In *Proc. CEAS 2004* (Mountain View, CA, 2004).
- [23] LOVÁSZ, L. AND SZEGEDY, B. (2006). Limits of dense graph sequences. *J. Combinatorial Theory B* **96**, 933–957.
- [24] MCCULLAGH, P. AND YANG, J. (2006). Stochastic classification models. In *International Congress of Mathematicians. Vol. III*, European Mathematical Society, Zürich, pp. 669–686.