Edge-exchangeable graphs and sparsity

Tamara Broderick
Department of EECS
Massachusetts Institute of Technology
tbroderick@csail.mit.edu

Diana Cai
Department of Statistics
University of Chicago
dcai@uchicago.edu

Abstract

A known failing of many popular random graph models is that the Aldous–Hoover Theorem guarantees these graphs are dense with probability one; that is, the number of edges grows quadratically with the number of nodes. This behavior is considered unrealistic in observed graphs. We define a notion of edge exchangeability for random graphs in contrast to the established notion of infinite exchangeability for random graphs—which has traditionally relied on exchangeability of nodes (rather than edges) in a graph. We show that, unlike node exchangeability, edge exchangeability encompasses models that are known to provide a projective sequence of random graphs that circumvent the Aldous–Hoover Theorem and exhibit sparsity, i.e., sub-quadratic growth of the number of edges with the number of nodes. We show how edge-exchangeability of graphs relates naturally to existing notions of exchangeability from clustering (a.k.a. partitions) and other familiar combinatorial structures.

1 Introduction

As graph, or network, data become more ubiquitous and large-scale—in the form of social networks, collaboration networks, networks expressing biological interactions, etc.—a number of probabilistic models for graphs have been proposed. However, it has become apparent that many of the most popular probabilistic models for graphs are fundamentally misspecified in a way that worsens as the graphs scale to larger numbers of nodes [14]. At the heart of the problem is that generative probabilistic modeling relies on what seems at first to be a very weak assumption, that of *exchangeability*. Exchangeability is essentially the idea that seeing our data in a different order does not change its distribution and is a much weaker assumption than the popular "independent and identical distribution (iid)" assumption. In graphs, this assumption has historically taken the following form: we assume that if we relabeled our nodes, it would not change the probability of the graph [1, 9, 14]. Exchangeability assumptions are fundamentally tied with probabilistic modeling since they imply the existence of parameters, likelihoods, and priors by *de Finetti's Theorem*. The particular version of de Finetti's Theorem for graphs with this form of *node-exchangeability* is known as the *Aldous–Hoover Theorem* [1, 9]. Notably, the Aldous-Hoover Theorem for exchangeable graphs implies that, if we assume our graph is node-exchangeable, it must be that our graph is *dense* [14]. That is, the number of edges in a dense graph grows quadratically with the number of nodes.

But most real-world graphs have been observed to be *sparse* [13]; that is, the number of edges grows sub-quadratically in the number of nodes. While there are ad hoc solutions to this mismatch between model and data, they have other undesirable properties, such as lacking *projectivity*—a property that facilitates handling streaming data, performing distributed data analysis, and consistent hierarchical modeling. Caron and Fox [7] have recently suggested an example model that has some desirable sparse scaling properties. They consider an alternative form of exchangeability in the sense of independent increments of subordinators. By contrast, we here consider a new form of exchangeability for graphs where we consider permuting the *edges* rather than the nodes. In the remainder of the current work, we both describe node-exchangeability in more detail and introduce our new concept of *edge-exchangeability*. We show how edge-exchangeability of graphs relates naturally to existing notions of exchangeability from clustering (a.k.a. partitions), feature allocations [3, 4], and other combinatorial structures [6, 5]. We describe how the Caron and Fox [7] model fits into our framework. We outline remaining connections and characterizations to be made in future work.

2 Edge-exchangeability in graphs

An undirected graph is defined by a set of nodes (i.e., vertices) called V and a set of edges E. In particular, each element e of E is an unordered set $\{u, v\}$ of two nodes $u, v \in V$ and represents a link between u and v. We consider V to be a set where each element is unique, but we allow E to be a multiset; that is, we allow edges to potentially occur multiple times in E. We define *active nodes* to be those nodes that appear in some edge, where an edge between a vertex and itself counts for this purpose. In what follows, we take the approach that we are only interested in active nodes. In this case, an undirected graph may be characterized by specifying only its edge set E. That is, we can consider a graph as an unordered set of tuples, which represent the edges of the graph. In this case, we can obtain the (active) node set by representing all the unique elements of E: $V = \bigcup_{e \in E} e$.

Example 2.1. Consider the graph (containing only active nodes) defined by $E = \{\{2,3\},\{1,4\},\{3,6\},\{6,6\},\{3,6\}\}$. Then the active nodes are $V = \{1,2,3,4,6\}$. It is not important that "5" does not appear in V since the elements of V are treated as arbitrary labels. ■

To consider exchangeability in graphs, it has been traditional to think of a sequence of (random) graphs $E_1, E_2, \ldots$ such that $E_m \subseteq E_n$ for any $m < n$ and such that E_n represents the edges between the vertices with labels in $[n] := \{1, \ldots, n\}$. One typically thinks of the step of going from E_n to E_{n+1} as adding in a new vertex labeled $n + 1$ (and all its edges to existing vertices and itself). Then we say that the graphs are *infinitely exchangeable* if, for any positive integer n, we have that permuting the vertex labels in $[n]$ does not change the probability of these graphs. This form of exchangeability, which we will refer to as *node exchangeability* implies, via the Aldous–Hoover Theorem [9, 1], that our graphs must be dense [14]. Thus, if we wish to maintain some notion of exchangeability while modeling sparse graphs, it behooves us to consider alternative forms of exchangeability.

Example 2.2. An example realization of a sequence of random graphs when considering infinite exchangeability in the traditional sense is given by

$$E_1 = \emptyset, E_2 = \{\{1,2\}\}, E_3 = \{\{1,2\}\}, E_4 = \{\{1,2\},\{1,4\},\{2,4\},\{3,4\}\}, \ldots.$$

Note that it is not necessary to specify the node set V_1, V_2, V_3, ... here since it is understood that $V_n = [n]$. Further note that V_n here represents all the nodes, not only the active ones. ■

In the present work, we instead introduce a notion of exchangeability of the *edges* rather than the *vertices*. In particular, consider a new sequence of graphs $E_1, E_2, \ldots$, where $E_m \subseteq E_n$ for any $m < n$. Now we think of E_{n+1} as adding some new edges relative to E_n, but these new edges need not be connected to any particular vertex. We can make the step on which we add in edge e explicit by augmenting the edge set E. In particular, define a *step-augmented graph* E' as a collection of tuples, where the first element is the edge and the second element is the step on which the edge is added.

Example 2.3. There is exactly one sequence of step-augmented graphs that corresponds to the graph sequence in Example 2.2. It is

$$\begin{aligned}
E'_1 &= \emptyset, \\
E'_2 &= \{(\{1,2\},2)\}, \\
E'_3 &= \{(\{1,2\},2)\}, \\
E'_4 &= \{(\{1,2\},2),(\{1,4\},4),(\{2,4\},4),(\{3,4\},4)\}.
\end{aligned}$$

In the traditional exchangeability setup of Example 2.2, the step of an edge is always the maximum node value in that edge. ■

Example 2.4. In our new setup, the step need not be the maximum node value. For instance, consider a step-augmentation of Example 2.1:

$$E'_4 = \{(\{2,3\},1),(\{1,4\},4),(\{3,6\},1),(\{6,6\},3),(\{3,6\},3)\}.$$

This augmentation would be equivalent to the graph sequence

$$\begin{aligned}E_1 &= \{\{2,3\},\{3,6\}\},\\ E_2 &= \{\{2,3\},\{3,6\}\},\\ E_3 &= \{\{2,3\},\{3,6\},\{6,6\}\},\{3,6\}\},\\ E_4 &= \{\{2,3\},\{1,4\},\{3,6\},\{6,6\},\{3,6\}\}.\end{aligned}$$

■

Now suppose we treat the *edges* (rather than the nodes) in this sequence of edge sets as arbitrary labels; if two edges are the same, they have the same labels, and otherwise they have different labels. This representation is technically a superclass of the objects defined above since there is additional structure in the edge labels of graphs; graph edge labels may agree in one element but not the other.

Example 2.5. Using an order of appearance scheme [4] to index the labels, E'_4 in Example 2.4 becomes $\{(\phi_1,1),(\phi_2,1),(\phi_3,3),(\phi_1,3),(\phi_4,4)\}$. ■

We note that if we consider only the uniqueness of the labels in E'_n and not their actual values, the information in this structure can be expressed as a set of subsets of the steps $[n]$. That is, for each unique label ϕ that occurs in any tuple in E'_n, we let A_ϕ be the collection of steps that co-occur in a tuple with ϕ. In Example 2.5, $A_{\phi_1} = \{1,3\}$. Then let C_n be the collection of A_ϕ for all values of ϕ. Thus, C_n is a set of subsets of $[n]$ since E'_n contains only edges added on step n or earlier. We call the sequence (C_n) the *step collection sequence* of a sequence of graphs.

Example 2.6. The C_4 corresponding to E'_4 in Example 2.5 is $\{\{1,3\},\{1\},\{3\},\{4\}\}$. ■

Finally, now, we recognize the step collection C_n as a familiar combinatorial object: either a partition (a.k.a. clustering) [15], feature allocation [4], or a trait allocation (defined below and reminiscent of [5]); we explain each of these connections below and show how they give a natural notion of exchangeability in the *edges* of a graph.

2.1 Partition connection

First consider the connection to partitions. In this case, suppose that each index in $[n]$ appears exactly once across all of the subsets of C_n. This assumption on C_n is equivalent to assuming that in the original graph sequence $E_1, E_2, \ldots$, we have that E_{n+1} always has exactly one more edge than E_n. In this case, C_n is exactly a *partition* of $[n]$; that is, C_n is a set of mutually exclusive and exhaustive subsets of $[n]$. If the edge sequence (E_n) is random, then (C_n) is random as well.

We say that a partition sequence $C_1, C_2, \ldots$, where C_n is a (random) partition of $[n]$ and $C_m \subseteq C_n$ for all $m \le n$, is infinitely exchangeable if, for all n, permuting the indices of n does not change the distribution of the (random) partitions [15]. Permuting the indices $[n]$ in the partition sequence (C_m) corresponds to permuting the order in which edges are added in our graph sequence (E_m). Contrast this with the traditional form of exchangeability in graphs (node exchangeability) as described above.

Recall further that the *Kingman paintbox theorem* [11] tells us that we have an infinitely exchangeable partition sequence if and only if we can find a sequence of (potentially random) probabilities $p_1, p_2, \ldots$ such that $p_k \in (0,1)$ and $\sum_{k=1}^{\infty} p_k = 1$ and such that drawing partition elements according to these probabilities yields the same partition distribution as our original random partition. The sequence $(p_k)_{k=1}^{\infty}$ is called the *Kingman paintbox*. In the graph domain, we can interpret the Kingman paintbox probability p_k as the probability of a particular edge in the graph.

Example 2.7. We consider a generative model proposed by Caron and Fox [7]. Let $W = \sum_{k=1}^{\infty} w_k \delta_{\phi_k}$ be a random measure such that the pairs $\{(w_k,\phi_k)\}_{k=1}^{\infty}$ are generated from a Poisson point process with rate measure $\nu(dw, d\phi) = \nu(dw)G(d\phi)$ for some proper distribution G. We assume that ν is a positive measure with support on $\mathbb{R}_+$. In this case, W is a *completely random measure*. For $n = 1, 2, \ldots$, we draw whether the graph acquires edge $\{i,j\}$ at step n according to the distribution $(p_{\{i,j\}})_{i,j}$ where

$$p_{\{i,j\}} \propto \begin{cases} 2w_i w_j & i \neq j \\ w_i^2 & i = j \end{cases} .$$

Defining this distribution requires that the sum over the $w_i w_j$ be finite, but this finiteness condition will hold for a wide range of models, including the generalized gamma process [10, 12] with $\phi \in [0, \alpha]$ for some fixed $\alpha > 0$, as considered by [7]. In fact, this construction returns exactly the generative model of Caron and Fox [7] if the total number of edges N is chosen to be Poisson with appropriate rate parameter.[1] ■

2.2 Feature allocation connection

Next we notice that it need not be the case that exactly one edge is added at each step of the graph sequence, e.g. between E_n and E_{n+1}. If we allow multiple unique edges at any step, then the step collection C_n is just a set of subsets of $[n]$, where each subset has at most one of each index in $[n]$. Suppose that any m belongs to only finitely many subsets in C_n for any n. That is, we suppose that only finitely many edges are added to the graph at any step. Then C_n is an example of a *feature allocation* [4]. Again, if (E_n) is random, then (C_n) is random as well.

We say that a (random) feature allocation sequence (C_m) is infinitely exchangeable if, for any n, permuting the indices of $[n]$ does not change the distribution of the (random) feature allocations [3, 4]. In this feature allocation case, permuting the indices $[n]$ in the sequence (C_m) corresponds to permuting the steps when edges are added in the edge sequence (E_m).

Similarly to the partition case in Section 2.1, we can apply known results from feature allocations to characterize edge-exchangeable graph models of this form. For instance, we know that the *feature paintbox* [4] characterizes distributions over feature allocations (and therefore this sequence of edge-exchangeable graphs) just as the Kingman paintbox characterizes distributions over partitions (and therefore the edge-exchangeable graphs with exactly one new edge per step).

We may also consider feature paintbox distributions with extra structure. For instance, we may say that an edge-exchangeable graph sequence (with multiple unique edges per step) has an *edge-exchangeable graph probability function* (EGPF) if the probability of the graph can be expressed as a function only of the total number of steps N and the edge multiplicities (and where the probability is symmetric in the edge multiplicities). This definition directly corresponds to the notion of an exchangeable feature probability function (EFPF) [4, 3] on the feature allocations (C_n).[2]

Also, we may define a *graph frequency model* as built around a random measure $B = \sum_{k=1}^{\infty} V_k \delta_{\phi_k}$. In particular, we draw a random graph conditional on B as follows. For each step index n, independently make a Bernoulli draw with success probability V_k. If the draw is a success, the edge indexed by k appears at time n. Otherwise it does not.

Then Theorem 17 ("Equivalence of EFPFs and feature frequency models") from [4] translates into a theorem about edge-exchangeable graphs as follows.

Theorem 2.8. *Let λ be a non-negative random variable (which may have some arbitrary joint law with the frequencies in a graph frequency model). We can obtain an edge-exchangeable graph by generating an edge-exchangeable graph from a graph frequency model and then, for each time n, including an independent* $\mathrm{Poisson}(\lambda)$*-distributed number of unique edges (which are different from those previously generated and which will never appear again). A graph of this type has an EGPF. Conversely, every graph with an EGPF has the same distribution as one generated by this construction for some joint distribution of λ and the edge frequencies.*

Example 2.9. We consider a graph frequency model related to the model from Example 2.7. In particular, let $W = \sum_{k=1}^{\infty} w_k \delta_{\phi_k}$ be a completely random measure with rate measure $\nu(dw, d\phi) = \nu(dw)G(d\phi)$ for some proper distribution G. We now assume that ν is a positive measure with support on $[0, 1]$. For $n = 1, 2, \ldots$, we draw whether the graph has an edge $\{i, j\}$ at time step n as

$$\mathrm{Bern}(q_{\{i,j\}}) \quad \text{where} \quad q_{\{i,j\}} := \begin{cases} 2w_i w_j & i \neq j \\ w_i^2 & i = j \end{cases}.$$

Since the graph frequency representation is given explicitly, the EGPF existence follows by Theorem 2.8. ■

[1] This observation makes use of the fact that a gamma process prior paired with a Poisson likelihood process yields the same distribution as a Dirichlet process paired with a multinomial likelihood and a Poisson-distributed number of data points.

[2] Both functions are related to exchangeable partition probability functions (EPPFs) [15] for partitions.

2.3 Trait allocation connection

Finally, we may consider the case where at every step, any non-negative (finite) number of edges may be added *and* those edges may have non-trivial (finite) multiplicity; that is, the multiplicity of any edge at any step can be any non-negative integer. By contrast, in Section 2.2, each unique edge occurred at most once at each step. In this case, the step collection C_n is a set of subsets of $[n]$. The subsets need not be unique or exclusive since we assume any number of edges may be added at any step. And the subsets themselves are multi-sets since an edge may be added with some multiplicity at step n. We say that C_n is a *trait allocation*,[3] which we define as a generalization of a feature allocation where the subsets of C_n are multisets. As above, if (E_n) is random, (C_n) is as well.

We say that a (random) trait allocation sequence (C_m) is infinitely exchangeable if, for any n, permuting the indices of $[n]$ does not change the distribution of the (random) trait allocation. Here, permuting the indices of $[n]$ corresponds to permuting the steps when edges are added in the edge sequence (E_m).

Example 2.10. Consider again the construction from Example 2.7. Now suppose that at each step, we add $n_{\{i,j\}}$ instances of edge $\{i, j\}$, where $n_{\{i,j\}}$ is drawn from some distribution $h(\cdot|\theta_{\{i,j\}})$, where $\theta_{\{i,j\}}$ equals $2w_i w_j$ for $i \neq j$ and w_i^2 for $i = j$. The case where $h(\cdot|\theta)$ is Poisson with mean parameter θ and we take exactly one step can be seen as another interpretation of the model in Caron and Fox [7]. ■

3 Conclusion

In this work, we have defined a notion of edge exchangeability for random graphs in contrast to the traditional notion of node exchangeability for random graphs. While the Aldous-Hoover Theorem guarantees that node-exchangeable random graphs must be dense with probability one, we have seen in Example 2.7 that edge-exchangeability encompasses models that are known to provide a projective sequence of random graphs that circumvent this theorem and exhibit sparsity. It remains to more fully characterize the asymptotic properties of edge-exchangeable random graphs. For one, we have considered how certain types of structure (e.g., the EGPF in Section 2.2) affect edge-exchangeable random graphs (Theorem 2.8). But perhaps an even more natural type of structure for graphs would be the idea that the probability of a graph depends only on the number of times a node occurs. We believe that this will yield a probability function structure like the EGPF. Moreover, it remains to consider power laws and other asymptotic behaviors in graphs in the style of [8] for partitions and [2] for feature allocations.

References

[1] D. J. Aldous. *Exchangeability and related topics*. Springer Lecture Notes in Math, 1985.

[2] T. Broderick, M. I. Jordan, and J. Pitman. Beta processes, stick-breaking, and power laws. *Bayesian Analysis*, 7(2):439–476, 2012.

[3] T. Broderick, M. I. Jordan, and J. Pitman. Cluster and feature modeling from combinatorial stochastic processes. *Statistical Science*, 2013.

[4] T. Broderick, J. Pitman, and M. I. Jordan. Feature allocations, probability functions, and paintboxes. *Bayesian Analysis*, 8(4):801–836, 2013.

[5] T. Broderick, A. C. Wilson, and M. I. Jordan. Posteriors, conjugacy, and exponential families for completely random measures. *arXiv preprint arXiv:1410.6843*, 2014.

[6] T. Broderick, L. Mackey, J. Paisley, and M. I. Jordan. Combinatorial clustering and the beta negative binomial process. *IEEE TPAMI*, 2015.

[7] F. Caron and E. B. Fox. Bayesian nonparametric models of sparse and exchangeable random graphs. *arXiv preprint arXiv:1401.1137v3*, 2015.

[8] A. Gnedin, B. Hansen, and J. Pitman. Notes on the occupancy problem with infinitely many boxes: general asymptotics and power laws. *Probab. Surv.*, 4:146–171, 2007. ISSN 1549-5787.

[3] We introduce this terminology following the discussion in [5].

[9] D. N. Hoover. Relations on probability spaces and arrays of random variables. Preprint, Institute for Advanced Study, Princeton, NJ, 1979.

[10] P. Hougaard. Survival models for heterogeneous populations derived from stable distributions. *Biometrika*, 73(2):387–396, 1986.

[11] J. F. C. Kingman. The representation of partition structures. *Journal of the London Mathematical Society*, 2(2):374–380, 1978.

[12] J. P. Klein and M. L. Moeschberger. *Survival analysis: techniques for censored and truncated data*. Springer Science & Business Media, 2003.

[13] M. E. J. Newman. Power laws, Pareto distributions and Zipf's law. *Contemporary physics*, 46 (5):323–351, 2005.

[14] P. Orbanz and D. M. Roy. Bayesian models of graphs, arrays and other exchangeable random structures. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 37(2):437–461, 2015.

[15] J. Pitman. Exchangeable and partially exchangeable random partitions. *Probability theory and related fields*, 102(2):145–158, 1995.