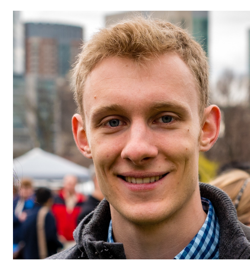
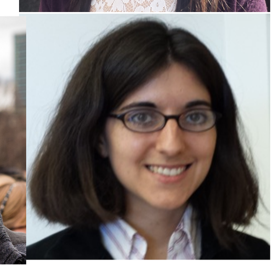
# Paintboxes and probability functions for edge-exchangeable graphs


**Diana Cai[1], Trevor Campbell[2], Tamara Broderick[2]**

[1]Statistics, U Chicago, CSAIL, [2]CSAIL, EECS, IDSS, MIT


## Motivation

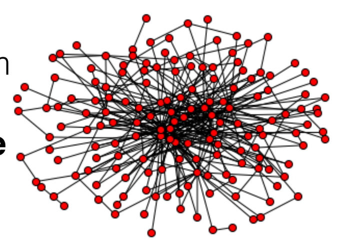
- Vertex-exchangeable graph models produce graph sequences that are dense or empty almost surely.
- But most real-world graphs are *sparse*.
- *Edge exchangeability* can produce sparse graphs.
- We introduce **the graph paintbox**, which characterizes all edge-exchangeable graphs.
- We introduce the **exchangeable vertex probability function** (EVPF), which characterizes all graph frequency models.

**Probabilistic graph models should capture real-life scaling behavior**: Given a graph sequence $G_1$, $G_2$, …

**Real-life graph sequences are sparse:**

$$[\#\text{edges}(G_n)] = o([\#\text{vertices}(G_n)]^2)$$

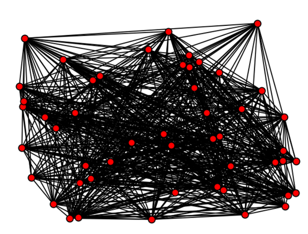
**(sub-quadratic growth)**




**Many models produce dense graphs:**

$$[\#\text{edges}(G_n)] = \Omega([\#\text{vertices}(G_n)]^2)$$

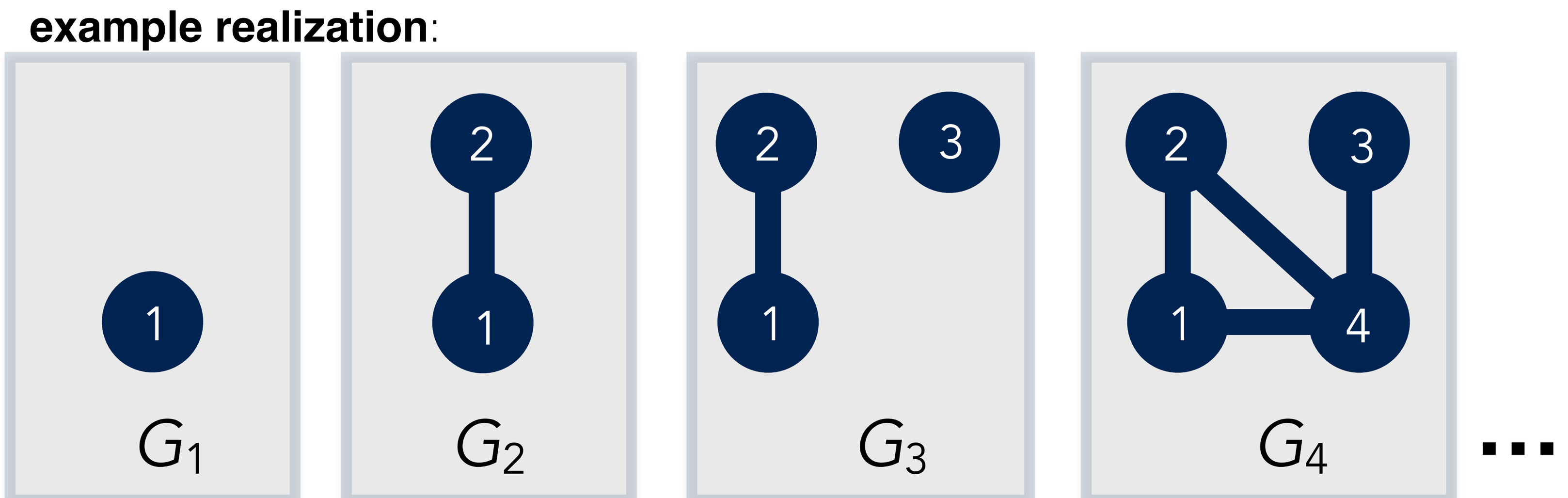
**(quadratic growth)**

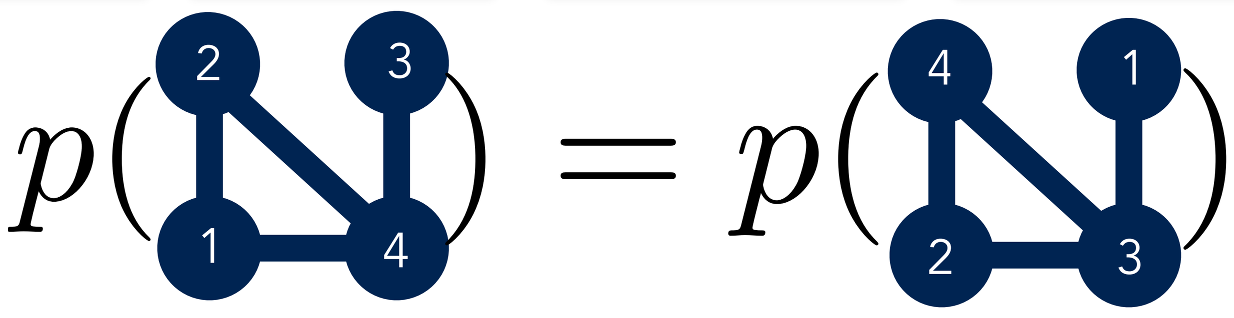



Vertex exchangeability implies a dense graph model, whereas **edge-exchangeable graph models can produce sparse graphs.**

In vertex exchangeability, a new vertex joins the graph sequence at each step, is labeled with that step number, and instantiates all edges with *existing* vertices.

**example realization**:



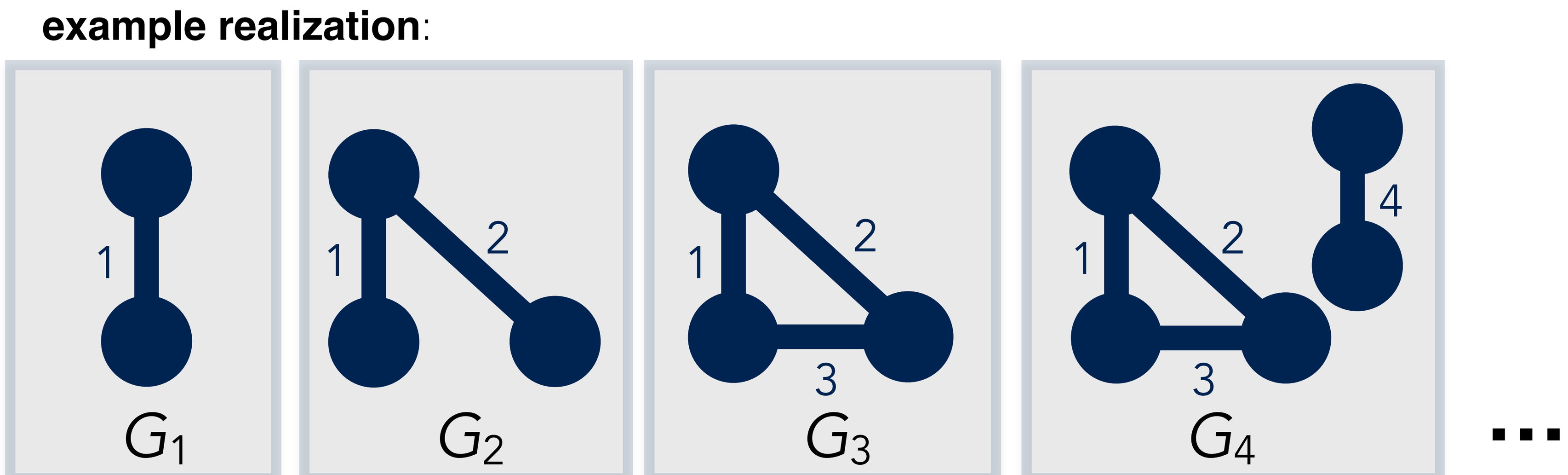



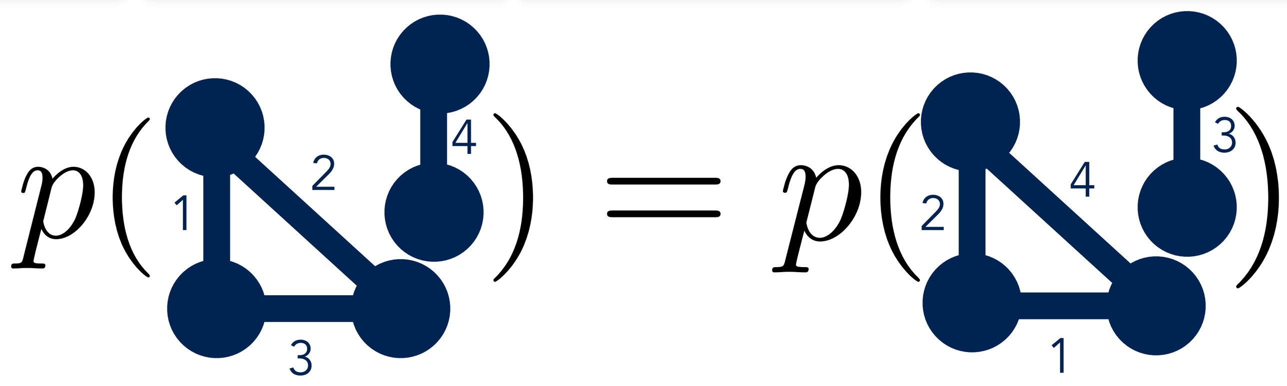
**Vertex exchangeability** means that the distribution of any step in the random graph sequence is invariant to permutations of the vertex labels. The Aldous-Hoover theorem implies that **vertex-exchangeable graphs are dense or empty with probability 1**.

We want a similar representation for edge-exchangeable graphs.

In edge exchangeability, a new *edge* joins the graph sequence at each step, is labeled with that step number, and instantiates any new vertices in that edge along the way.

**example realization**:

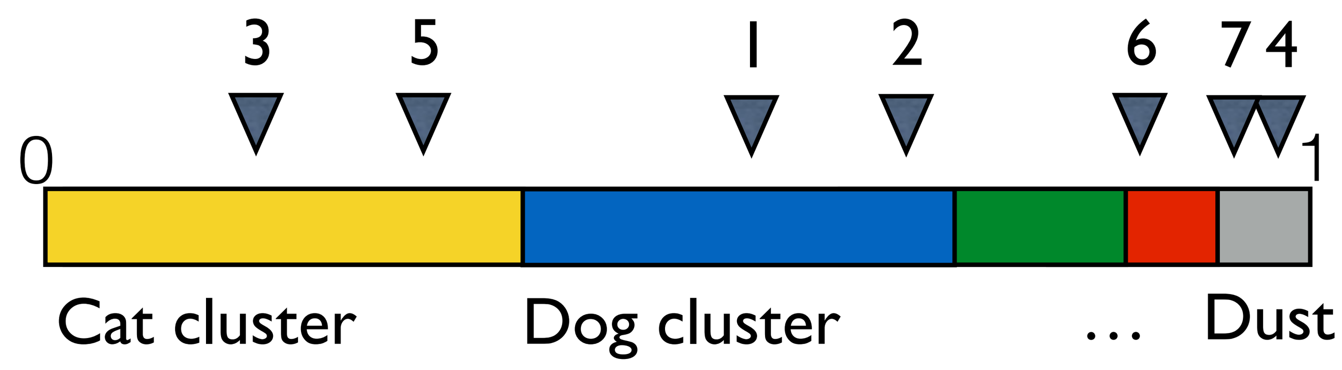


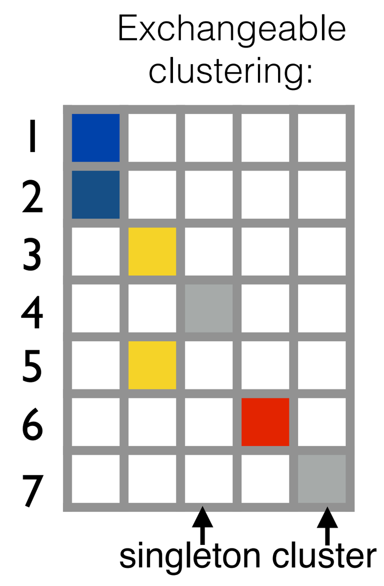



**Edge exchangeability** means that the distribution of any step in the random graph sequence is invariant to permutations of the edge labels. Now we present the graph paintbox, a representation for all edge-exchangeable graphs.

## The graph paintbox

Generalizes the Kingman paintbox for clustering:

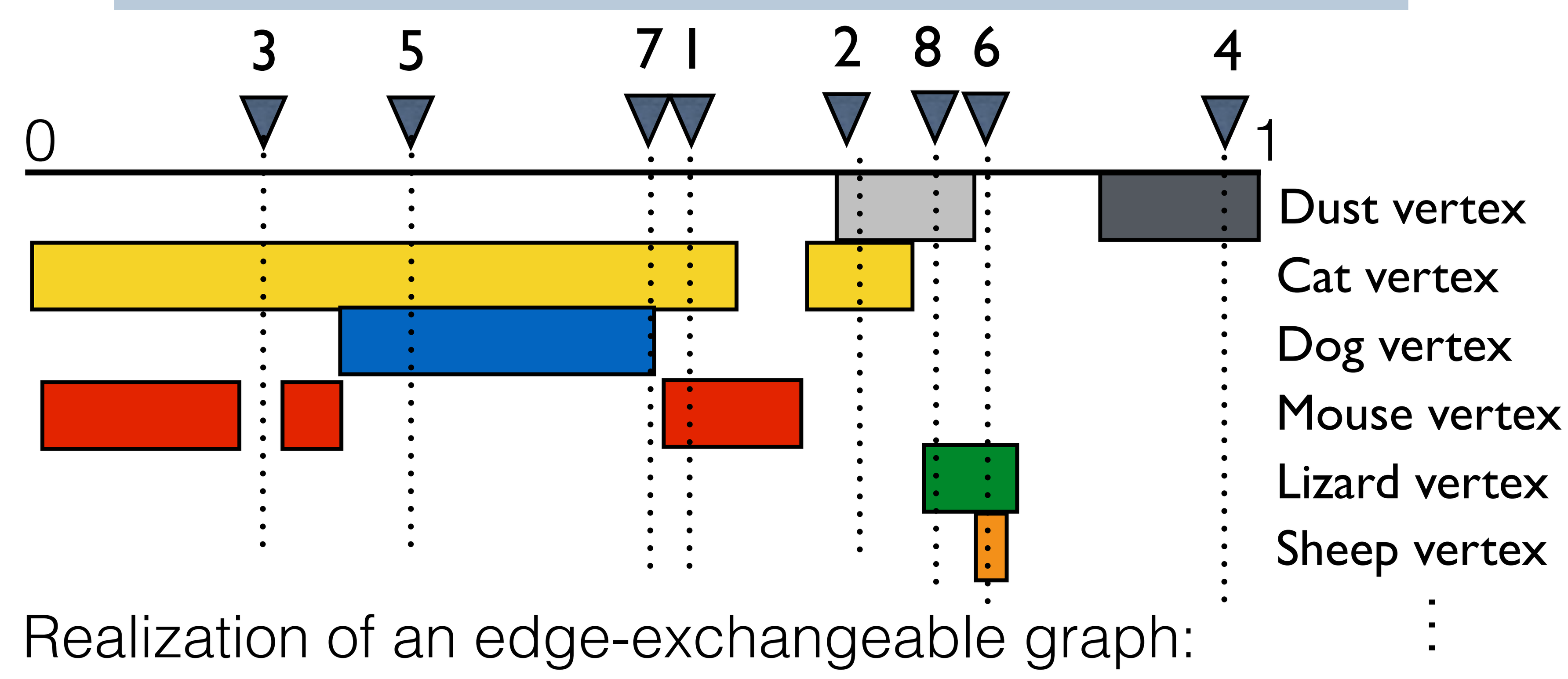


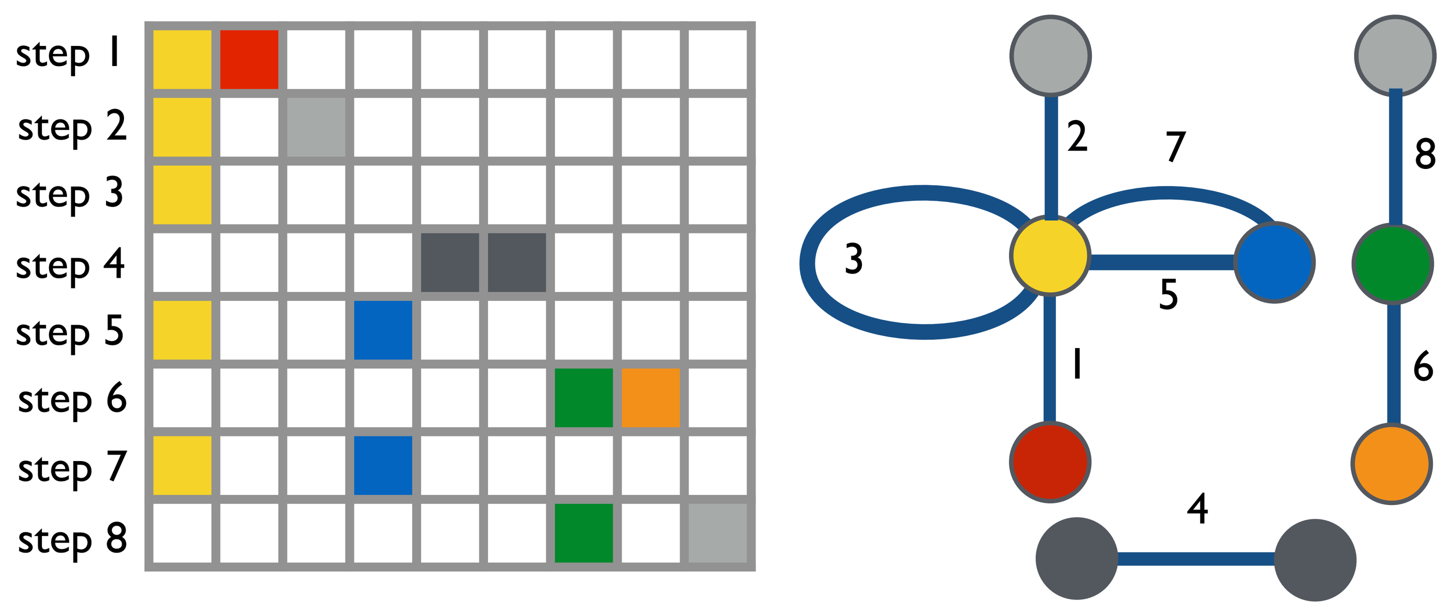

**Previous work** (Broderick & Cai, 2015; Cai, Campbell, Broderick, 2016) considered a characterization of edge-exchangeability where the paintbox subsets correspond to the *edges*.

Here we instead consider a characterization where the paintbox subsets correspond to the *vertices*, still in an edge-exchangeable framework. The earlier work most naturally treated edges with the same vertex as unrelated to each other. Our new paintbox explicitly relates edges that connect to the same vertex and allows us to control the topology of the graph.

**Theorem**: A random graph is edge-exchangeable iff it has a **graph paintbox** representation.




Realization of an edge-exchangeable graph:




## Frequency models & probability functions

The graph paintbox is expressive but complex. We want something good for efficient posterior inference algorithms:

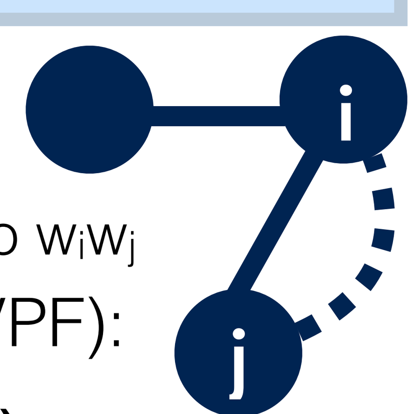
e.g., Gibbs sampling, variational Bayes, Hamiltonian Monte Carlo.

*Clustering*: exchangeable partition probability function (EPPF)
*Feature modeling*: exchangeable feature probability function (EFPF)

**Theorem**: An edge-exchangeable graph has a graph frequency model iff it has an exchangeable vertex probability function (EVPF).

Graph frequency model:

1. Draw weights ($w_i$) from some distribution
2. Draw edge $\{i,j\}$ with probability proportional to $w_i w_j$




Exchangeable vertex probability function (EVPF):

$$f(\underset{\uparrow\ \#\text{ of edges}}{8};\ \{5,1,\underbrace{2,1,1,2}_{\text{degrees of vertices}},1,1,1\})$$


**References**

1. D. J. Aldous. Exchangeability and related topics. Springer Lecture Notes in Math, 1985.
2. T. Broderick, J. Pitman, and M. I. Jordan. Feature allocations, probability functions, and paintboxes. Bayesian Analysis, 8(4):801–836, 2013.
3. T. Broderick and D. Cai. Edge-exchangeable graphs, sparsity, and power laws. NIPS Workshop on Bayesian Nonparametrics, 2015 & NIPS workshop on Networks, 2015.
4. T. Campbell, D. Cai, T. Broderick. Exchangeable trait allocations. arXiv: 1609.09147, 2016.
5. H. Crane and W. Dempsey. Edge exchangeable models for network data. arXiv: 1603.04571, 2016.
6. D. N. Hoover. Relations on probability spaces and arrays of random variables. Preprint, Institute for Advanced Study, Princeton, NJ, 1979.
7. J. F. C. Kingman. The representation of partition structures. Journal of the London Mathematical Society, 2(2):374–380, 1978.
8. P. Orbanz and D. M. Roy. Bayesian models of graphs, arrays and other