Nonparametric Bayes and Exchangeability: Part II

Tamara Broderick
Associate Professor
Electrical Engineering & Computer Science
MIT

http://www.tamarabroderick.com/tutorials.html

Roadmap

Roadmap

- Example problem: clustering

Roadmap

- Example problem: clustering
- Example NPBayes model: Dirichlet process (DP)

Roadmap

- Example problem: clustering
- Example NPBayes model: Dirichlet process (DP)

- Big questions

Roadmap

- Example problem: clustering
- Example NPBayes model: Dirichlet process (DP)

- Big questions
  - Why NPBayes?

Roadmap

- Example problem: clustering
- Example NPBayes model: Dirichlet process (DP)

- Big questions
  - Why NPBayes?
  - What does an infinite/growing number of parameters really mean (in NPBayes)?

Roadmap

- Example problem: clustering
- Example NPBayes model: Dirichlet process (DP)

- Big questions
  - Why NPBayes?
  - What does an infinite/growing number of parameters really mean (in NPBayes)?
  - Why is NPBayes challenging but practical?

Roadmap

- Example problem: clustering
- Example NPBayes model: Dirichlet process (DP)
- De Finetti for clustering: Kingman Paintbox

- Big questions
  - Why NPBayes?
  - What does an infinite/growing number of parameters really mean (in NPBayes)?
  - Why is NPBayes challenging but practical?

Roadmap

- Example problem: clustering
- Example NPBayes model: Dirichlet process (DP)
- De Finetti for clustering: Kingman Paintbox
- De Finetti for networks/graphs

- Big questions
  - Why NPBayes?
  - What does an infinite/growing number of parameters really mean (in NPBayes)?
  - Why is NPBayes challenging but practical?

Roadmap

- Example problem: clustering
- Example NPBayes model: Dirichlet process (DP)
- De Finetti for clustering: Kingman Paintbox
- De Finetti for networks/graphs

- Big questions
  - Why NPBayes?
  - What does an infinite/growing number of parameters really mean (in NPBayes)?
  - Why is NPBayes challenging but practical?
  - How does thinking about exchangeability help us use NPBayes in practice?

Dirichlet process mixture model

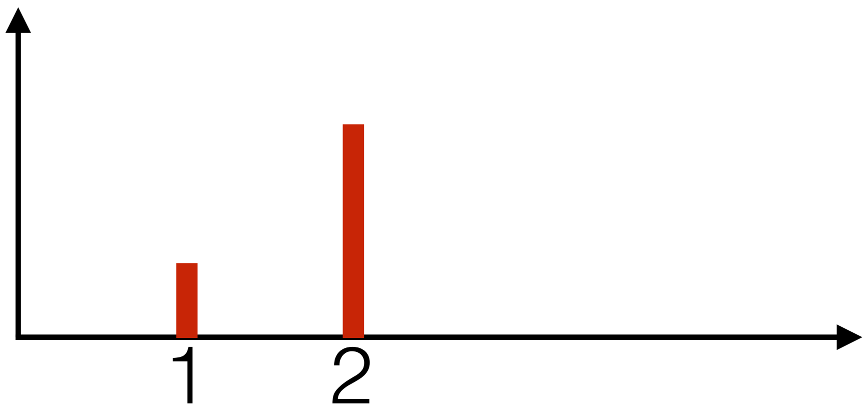
Dirichlet process mixture model

- Gaussian mixture model

Dirichlet process mixture model

- Gaussian mixture model

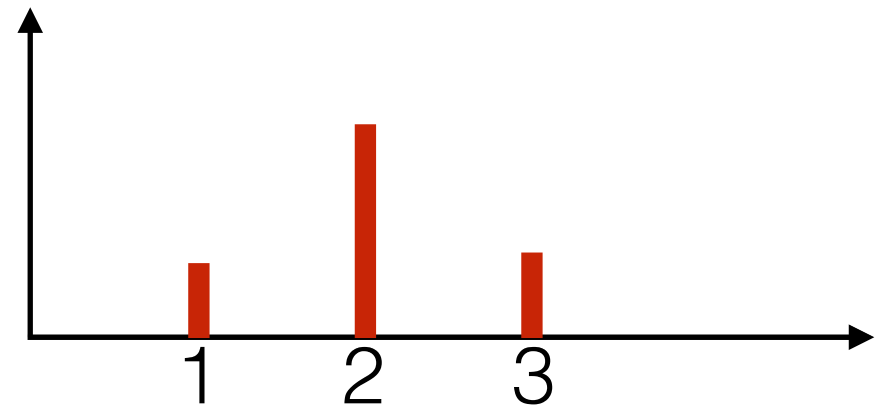
$$\rho = (\rho_1, \rho_2) \sim \mathrm{Beta}(a_1, a_2)$$



Dirichlet process mixture model

- Gaussian mixture model

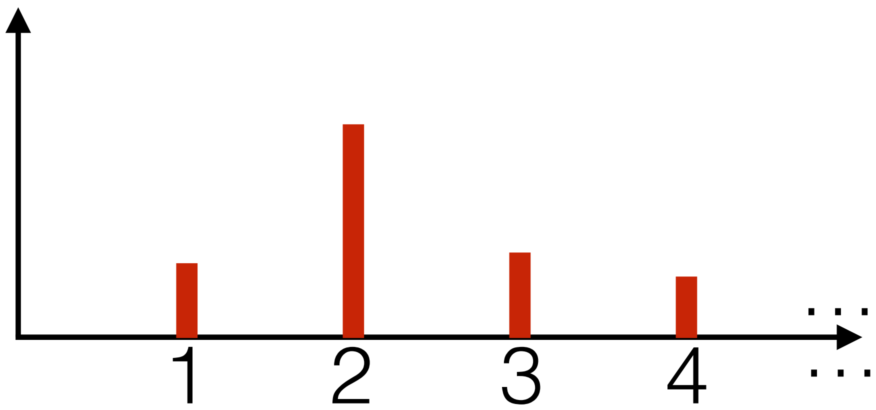
$$\rho = (\rho_1, \ldots, \rho_K) \sim \mathrm{Dir}(a_{1:K})$$



Dirichlet process mixture model

- Gaussian mixture model

$$\rho = (\rho_1, \rho_2, \ldots) \sim \mathrm{GEM}(\alpha)$$

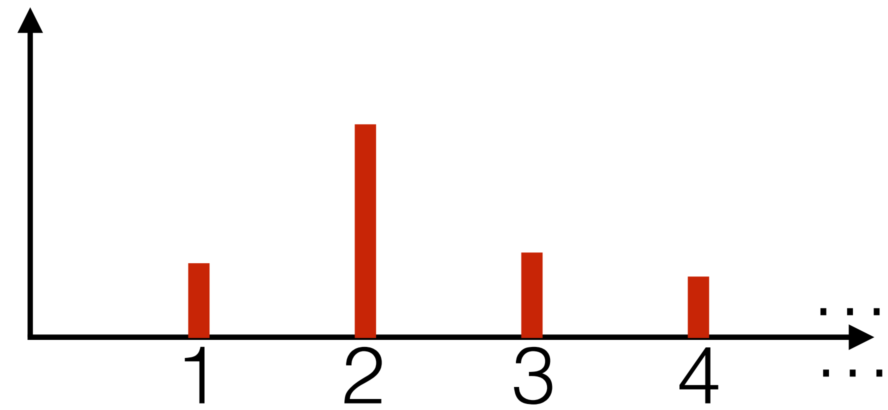


Dirichlet process mixture model

- Gaussian mixture model

$$\rho = (\rho_1, \rho_2, \ldots) \sim \mathrm{GEM}(\alpha)$$

$$\mu_k \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0), k = 1, 2, \ldots$$

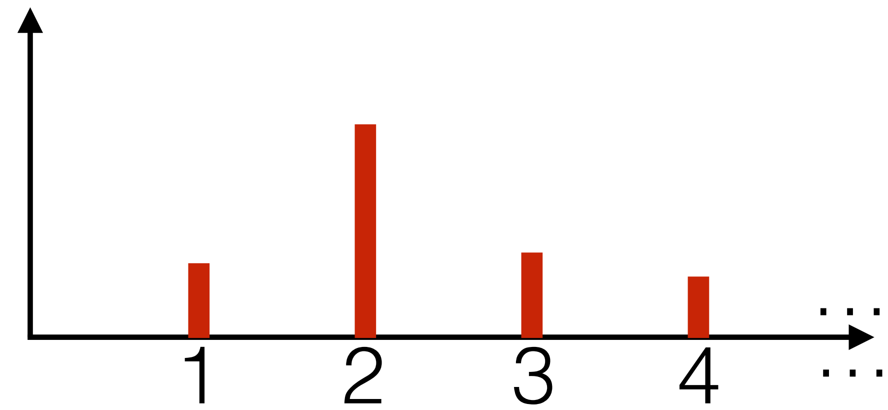


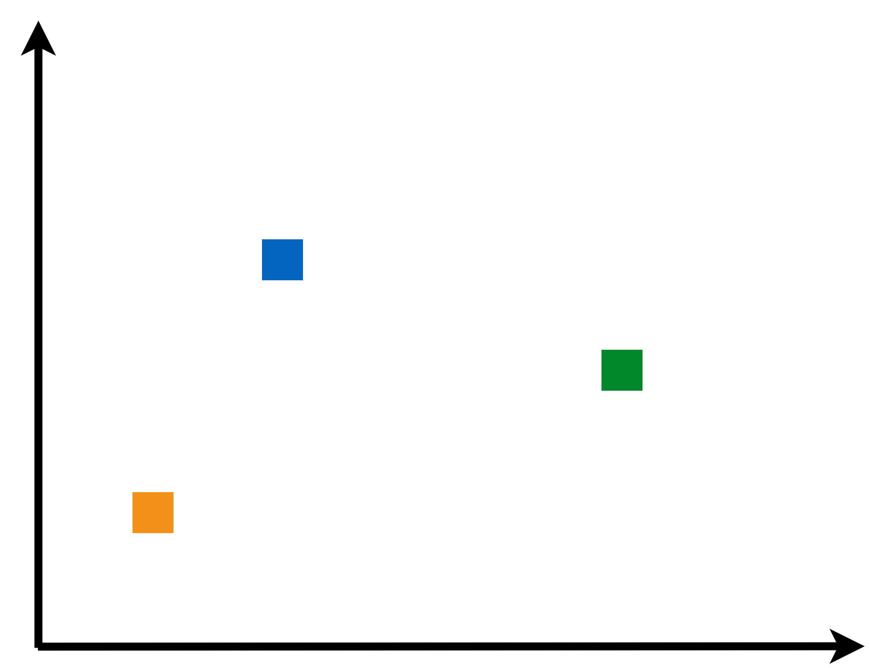
Dirichlet process mixture model

- Gaussian mixture model

$$\rho = (\rho_1, \rho_2, \ldots) \sim \text{GEM}(\alpha)$$

$$\mu_k \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0), k = 1, 2, \ldots$$



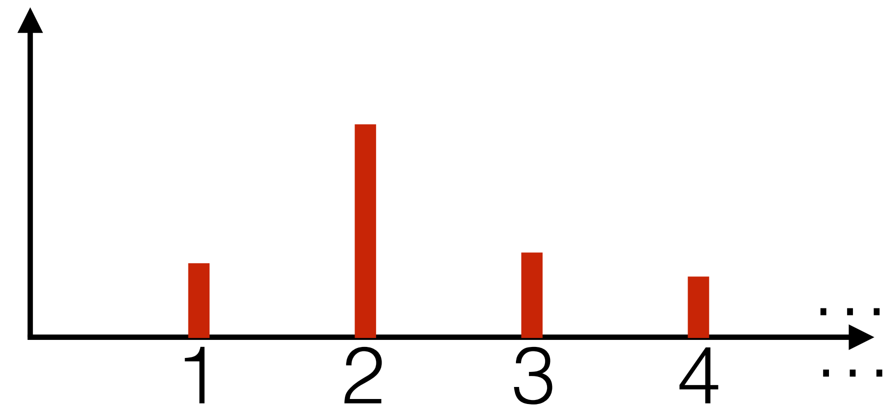


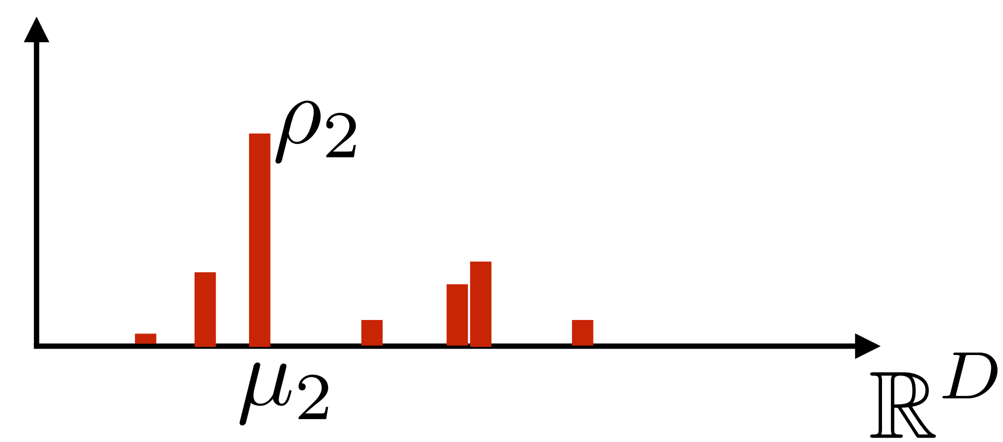
Dirichlet process mixture model

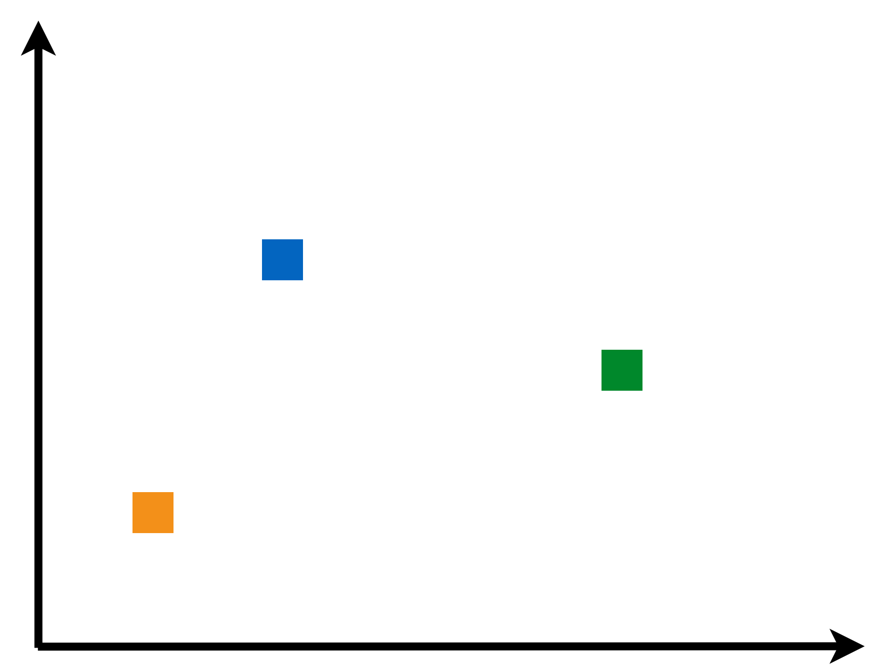
- Gaussian mixture model

$$\rho = (\rho_1, \rho_2, \ldots) \sim \mathrm{GEM}(\alpha)$$

$$\mu_k \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0), k = 1, 2, \ldots$$







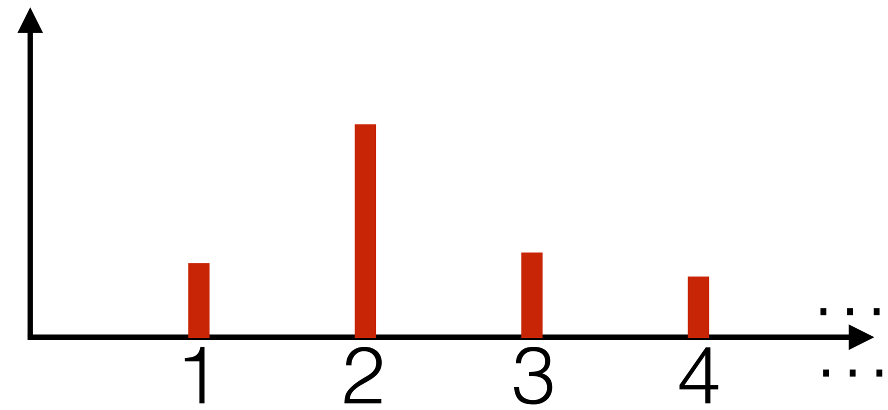
Dirichlet process mixture model

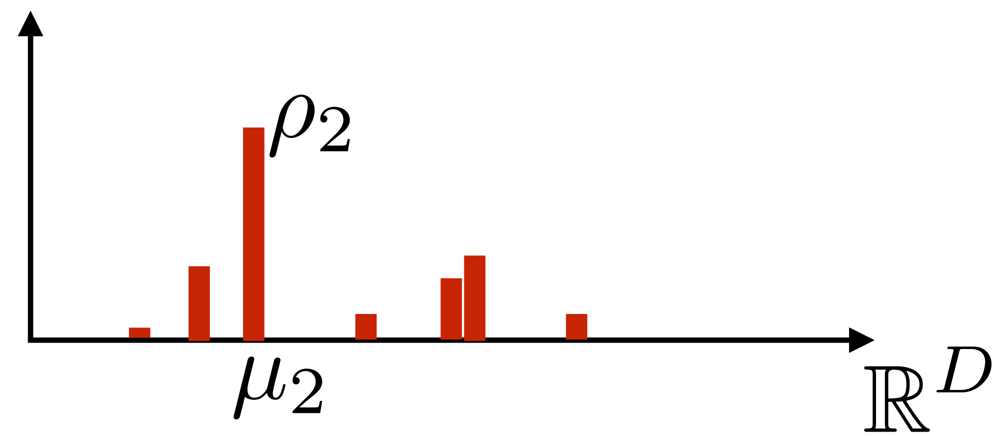
- Gaussian mixture model

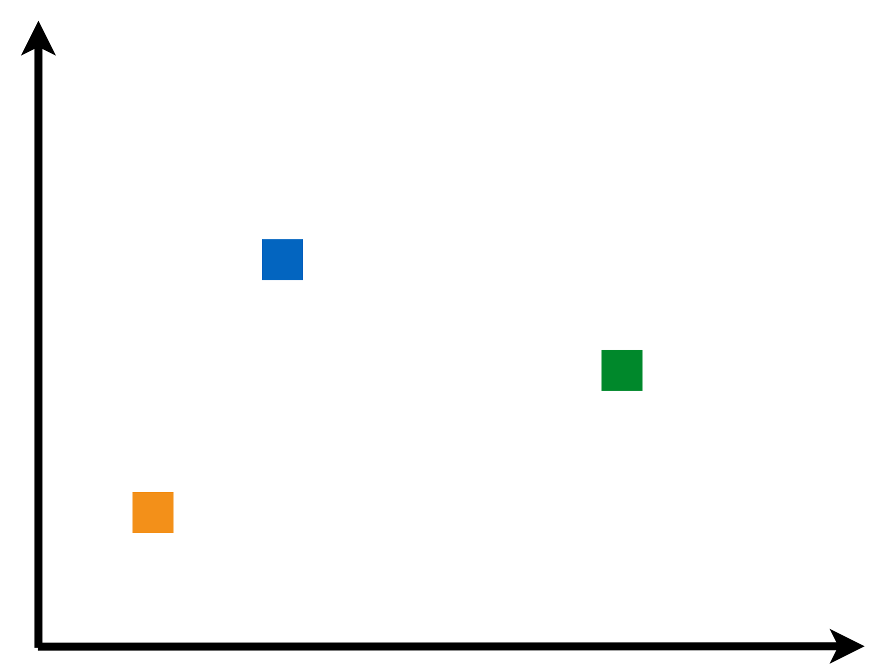
$$\rho = (\rho_1, \rho_2, \ldots) \sim \mathrm{GEM}(\alpha)$$

$$\mu_k \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0), k = 1, 2, \ldots$$

  - i.e. $G = \sum_{k=1}^{\infty} \rho_k \delta_{\mu_k}$

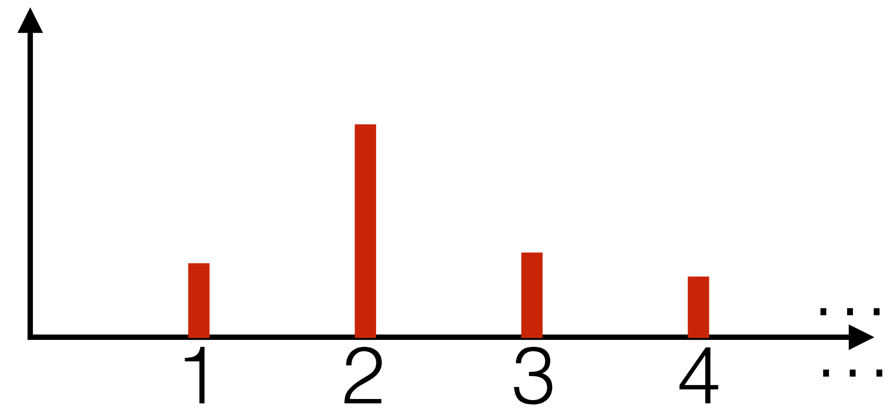






Dirichlet process mixture model

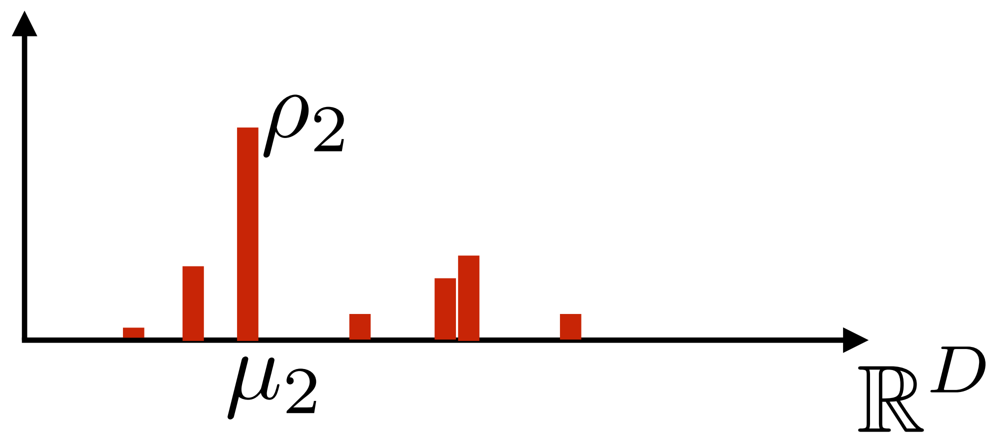
- Gaussian mixture model

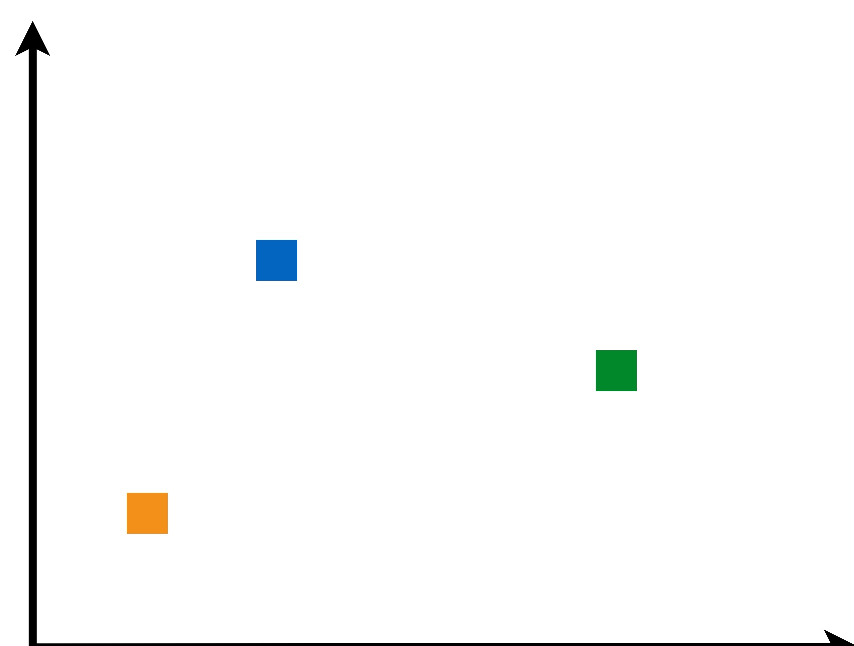


$$\rho = (\rho_1, \rho_2, \ldots) \sim \mathrm{GEM}(\alpha)$$

$$\mu_k \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0), k = 1, 2, \ldots$$

  - i.e. $G = \sum_{k=1}^{\infty} \rho_k \delta_{\mu_k} \overset{d}{=} \mathrm{DP}(\alpha, \mathcal{N}(\mu_0, \Sigma_0))$





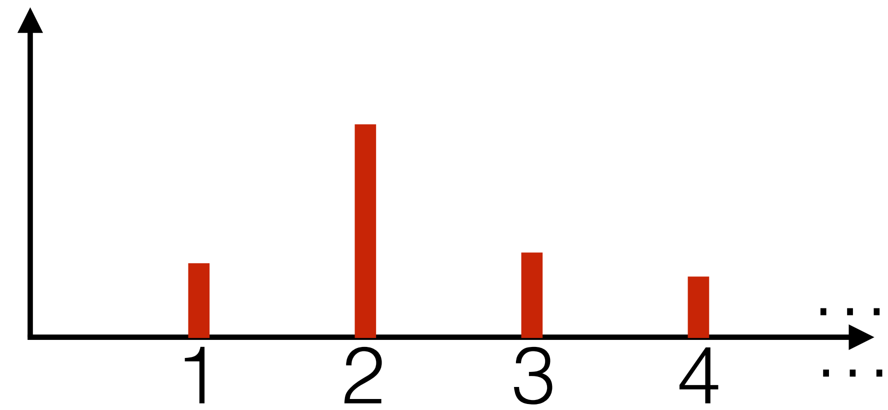
Dirichlet process mixture model

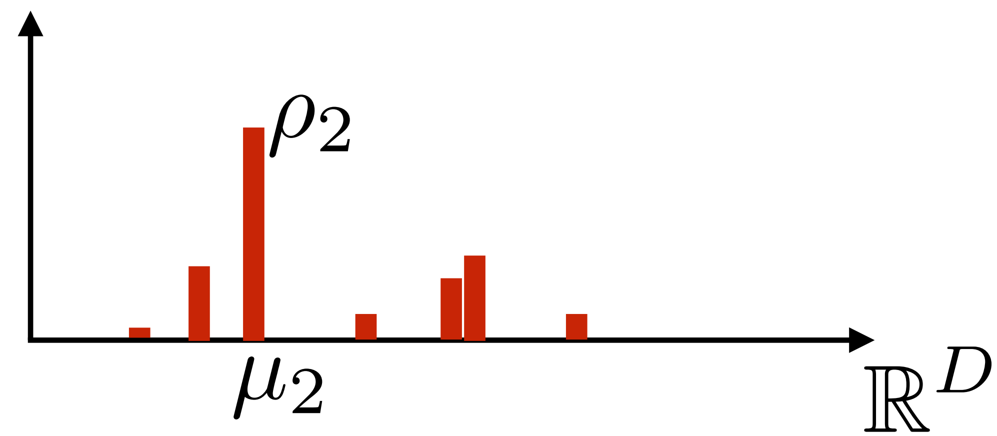
- Gaussian mixture model

$$\rho = (\rho_1, \rho_2, \ldots) \sim \mathrm{GEM}(\alpha)$$

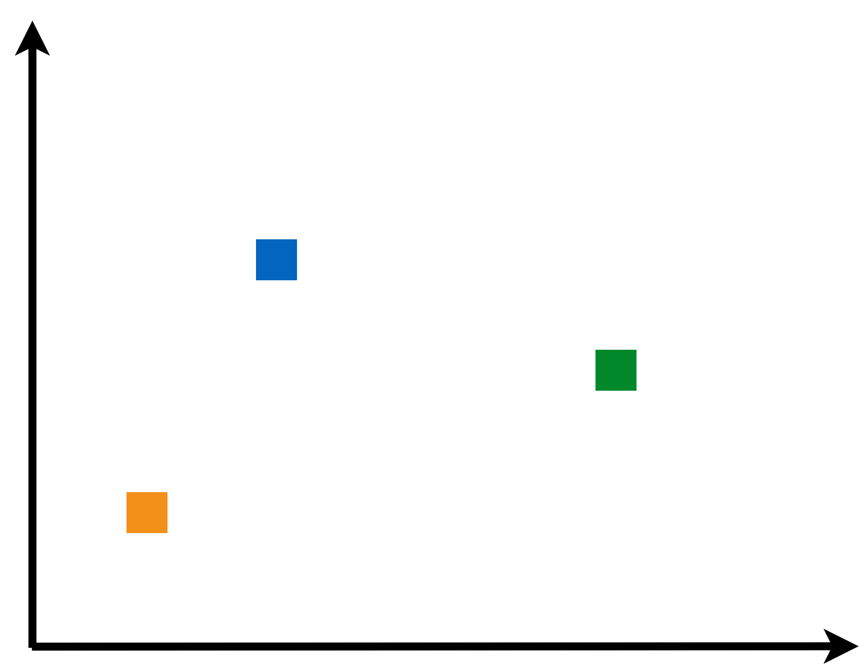
$$\mu_k \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0), k = 1, 2, \ldots$$

- i.e. $G = \sum_{k=1}^{\infty} \rho_k \delta_{\mu_k} \overset{d}{=} \mathrm{DP}(\alpha, \mathcal{N}(\mu_0, \Sigma_0))$



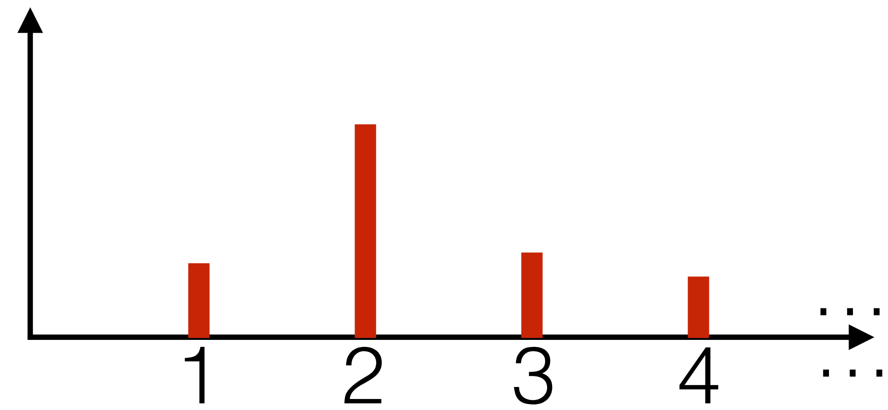


$$z_n \overset{iid}{\sim} \mathrm{Categorical}(\rho)$$



Dirichlet process mixture model

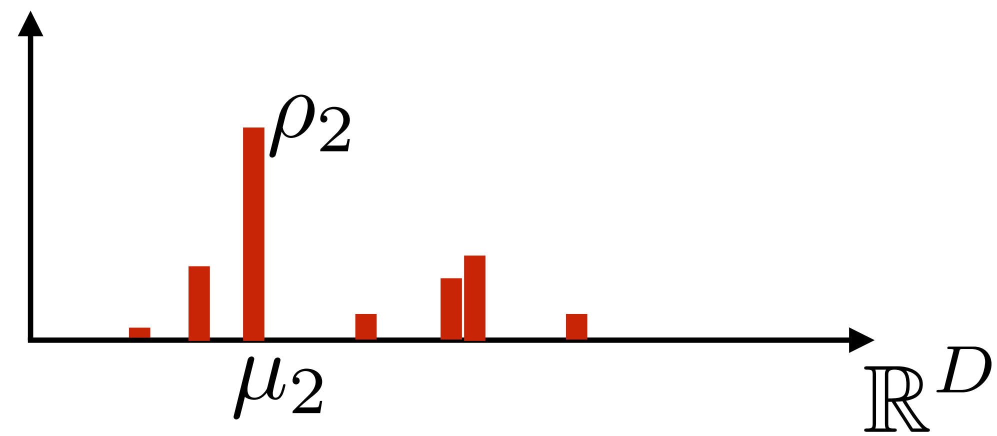
- Gaussian mixture model



$$\rho = (\rho_1, \rho_2, \ldots) \sim \mathrm{GEM}(\alpha)$$

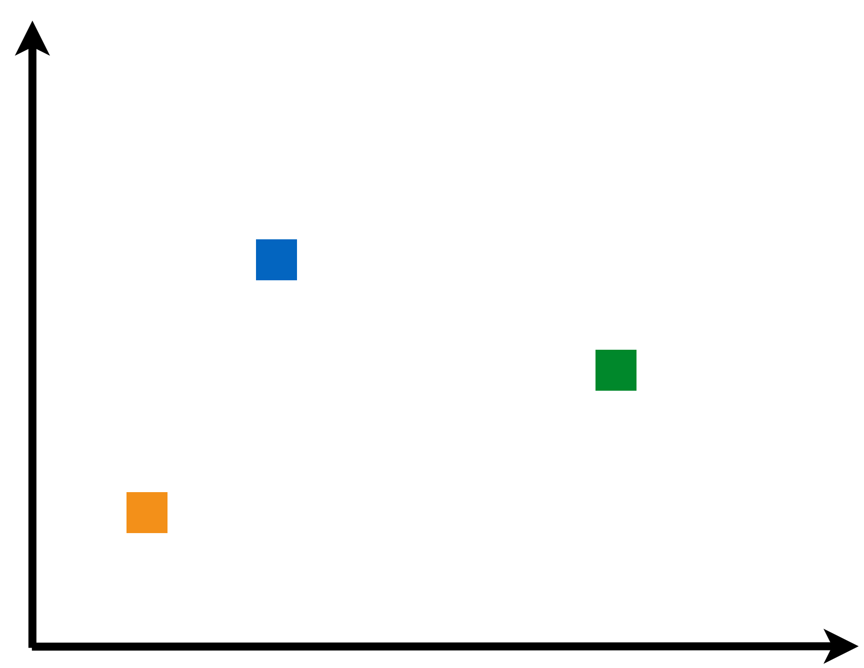
$$\mu_k \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0), k = 1, 2, \ldots$$

- i.e. $G = \sum_{k=1}^{\infty} \rho_k \delta_{\mu_k} \overset{d}{=} \mathrm{DP}(\alpha, \mathcal{N}(\mu_0, \Sigma_0))$



$$z_n \overset{iid}{\sim} \mathrm{Categorical}(\rho)$$

$$\mu_n^* = \mu_{z_n}$$

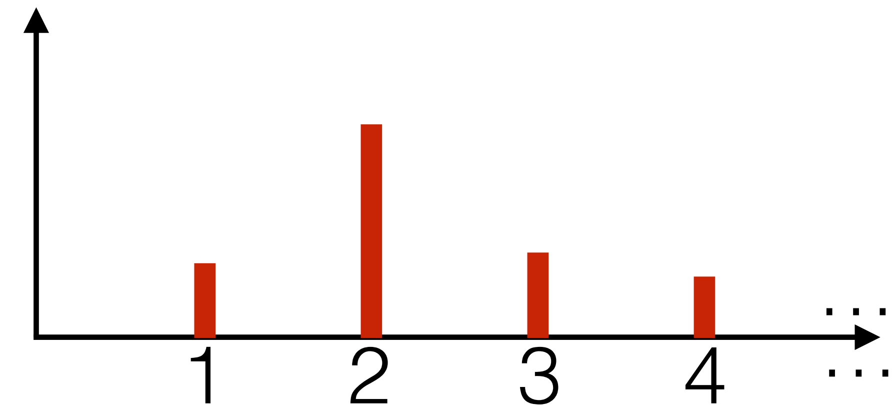


Dirichlet process mixture model

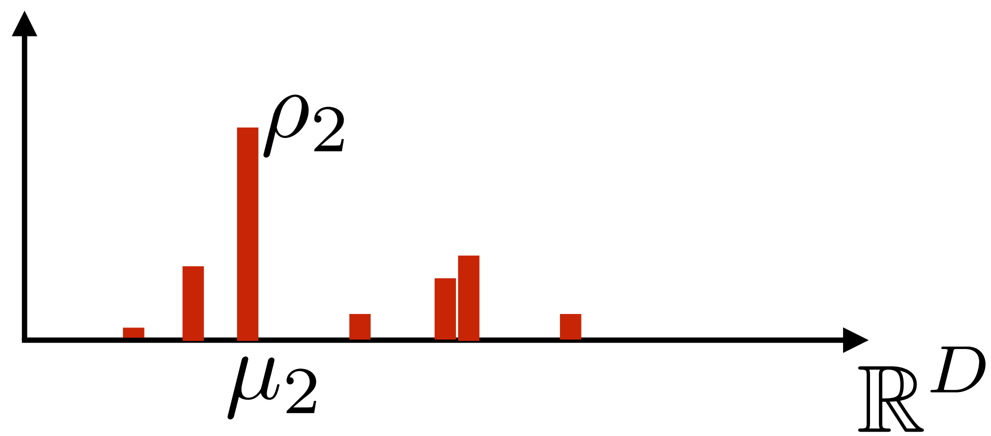
- Gaussian mixture model

$$\rho = (\rho_1, \rho_2, \ldots) \sim \mathrm{GEM}(\alpha)$$

$$\mu_k \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0), k = 1, 2, \ldots$$

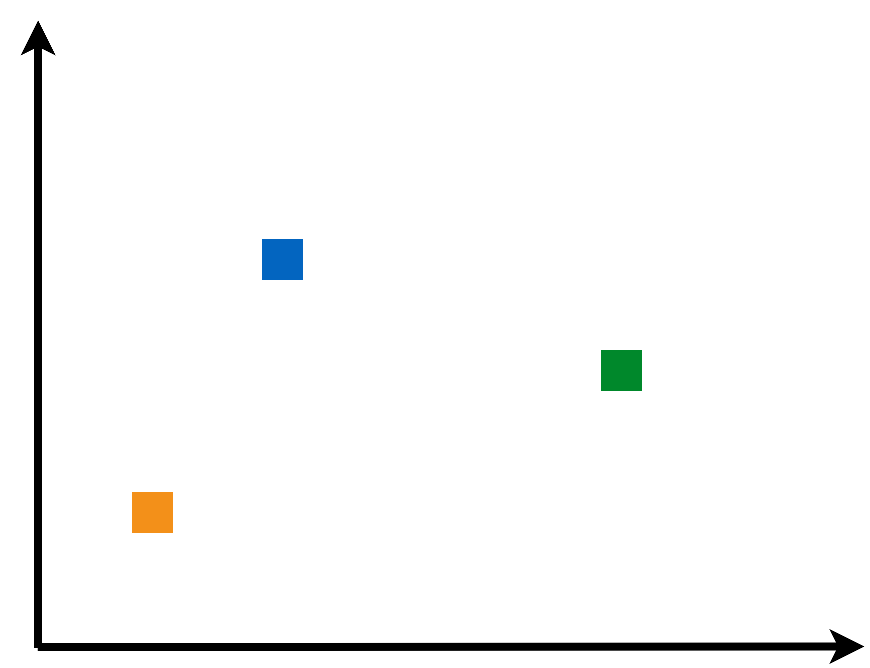


- i.e. $G = \sum_{k=1}^{\infty} \rho_k \delta_{\mu_k} \overset{d}{=} \mathrm{DP}(\alpha, \mathcal{N}(\mu_0, \Sigma_0))$

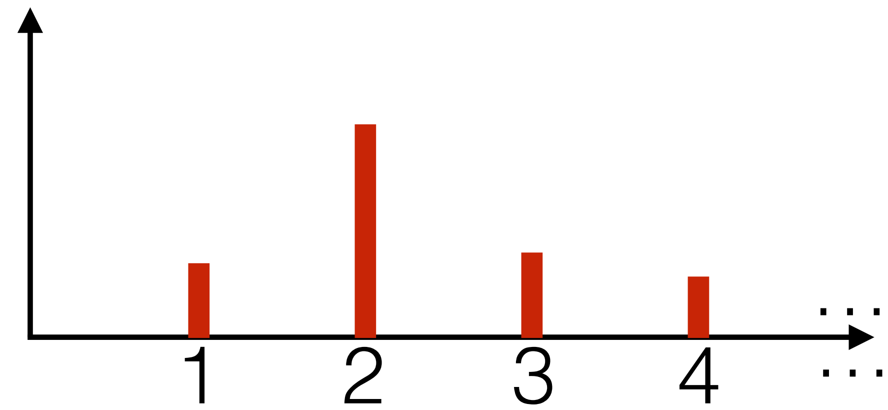


$$z_n \overset{iid}{\sim} \mathrm{Categorical}(\rho)$$

$$\mu_n^* = \mu_{z_n}$$



Dirichlet process mixture model

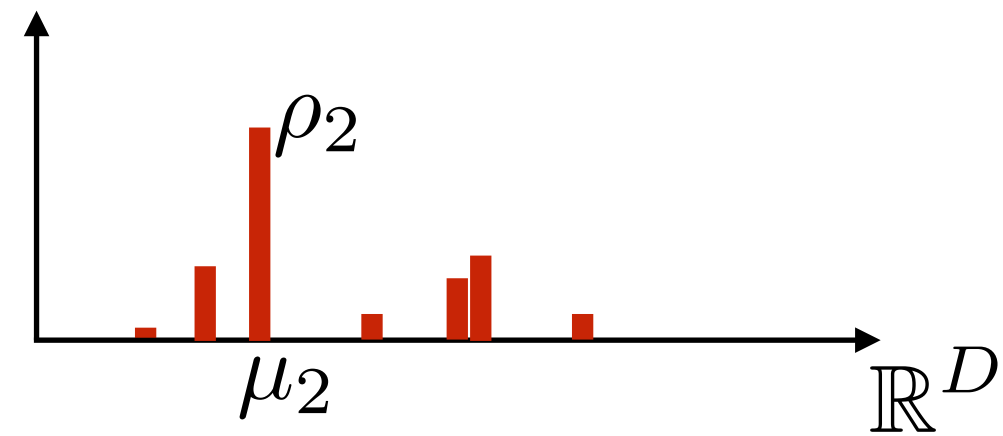


- Gaussian mixture model

$$\rho = (\rho_1, \rho_2, \ldots) \sim \mathrm{GEM}(\alpha)$$

$$\mu_k \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0), k = 1, 2, \ldots$$

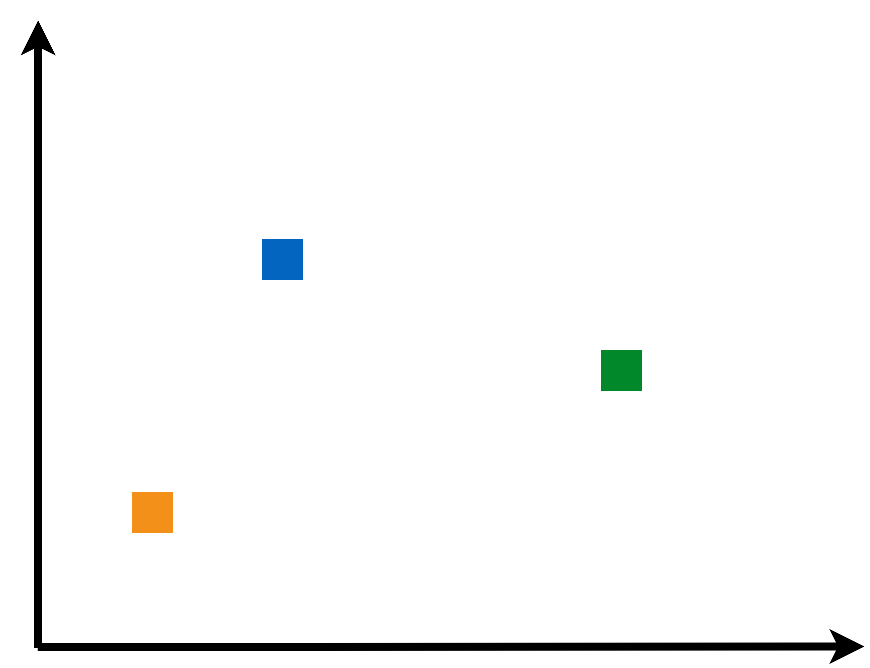
- i.e. $G = \sum_{k=1}^{\infty} \rho_k \delta_{\mu_k} \overset{d}{=} \mathrm{DP}(\alpha, \mathcal{N}(\mu_0, \Sigma_0))$



$$z_n \overset{iid}{\sim} \mathrm{Categorical}(\rho)$$

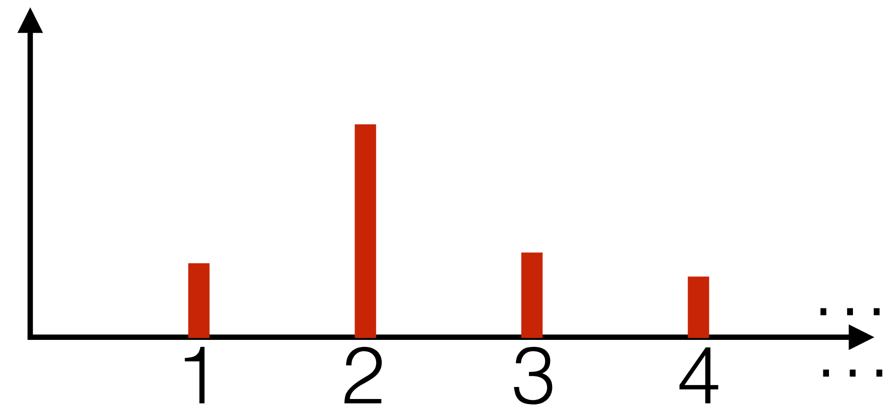
$$\mu_n^* = \mu_{z_n}$$

- i.e. $\mu_n^* \overset{iid}{\sim} G$



Dirichlet process mixture model

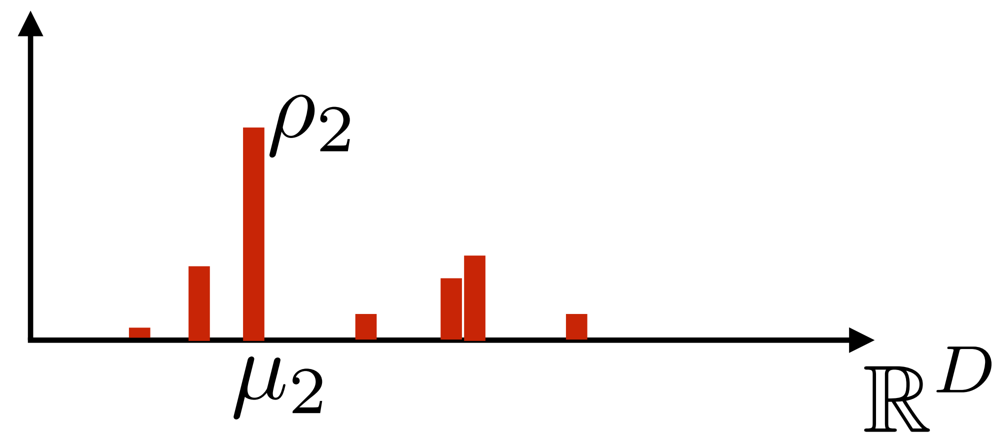
- Gaussian mixture model



$$\rho = (\rho_1, \rho_2, \ldots) \sim \mathrm{GEM}(\alpha)$$

$$\mu_k \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0), k = 1, 2, \ldots$$

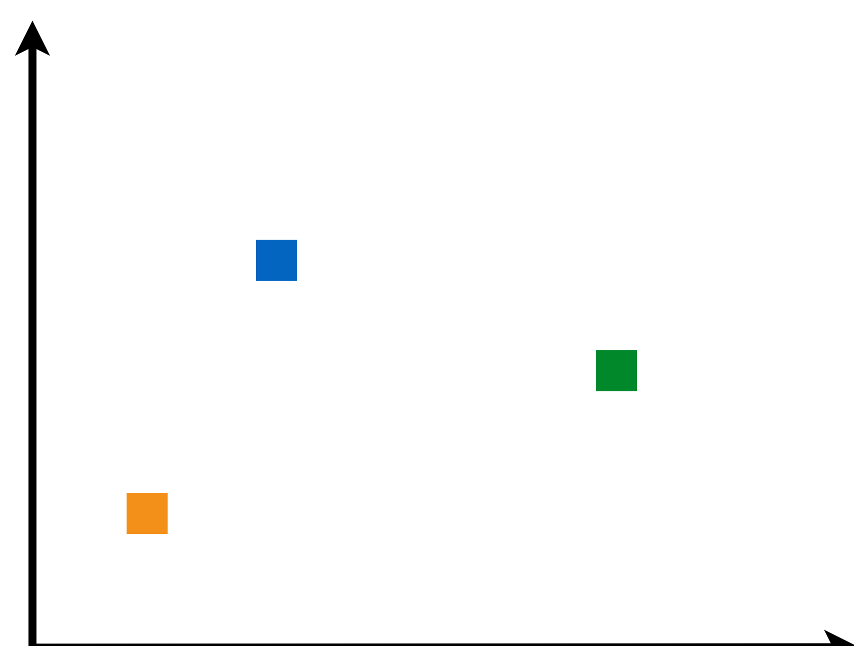
- i.e. $G = \sum_{k=1}^{\infty} \rho_k \delta_{\mu_k} \overset{d}{=} \mathrm{DP}(\alpha, \mathcal{N}(\mu_0, \Sigma_0))$



$$z_n \overset{iid}{\sim} \mathrm{Categorical}(\rho)$$

$$\mu_n^* = \mu_{z_n}$$

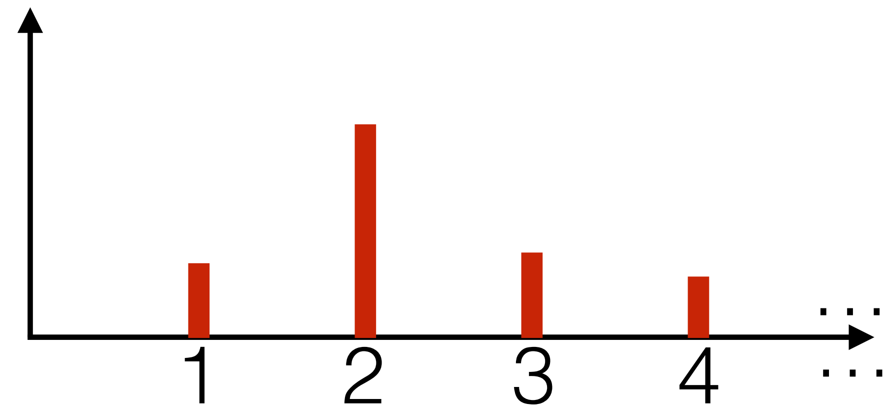
- i.e. $\mu_n^* \overset{iid}{\sim} G$



$$x_n \overset{indep}{\sim} \mathcal{N}(\mu_n^*, \Sigma)$$

Dirichlet process mixture model

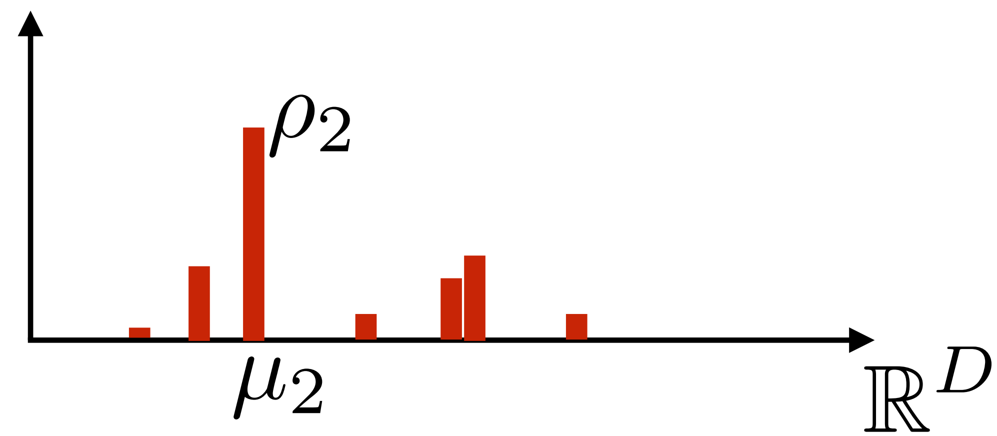
- Gaussian mixture model



$$\rho = (\rho_1, \rho_2, \ldots) \sim \text{GEM}(\alpha)$$

$$\mu_k \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0), k = 1, 2, \ldots$$

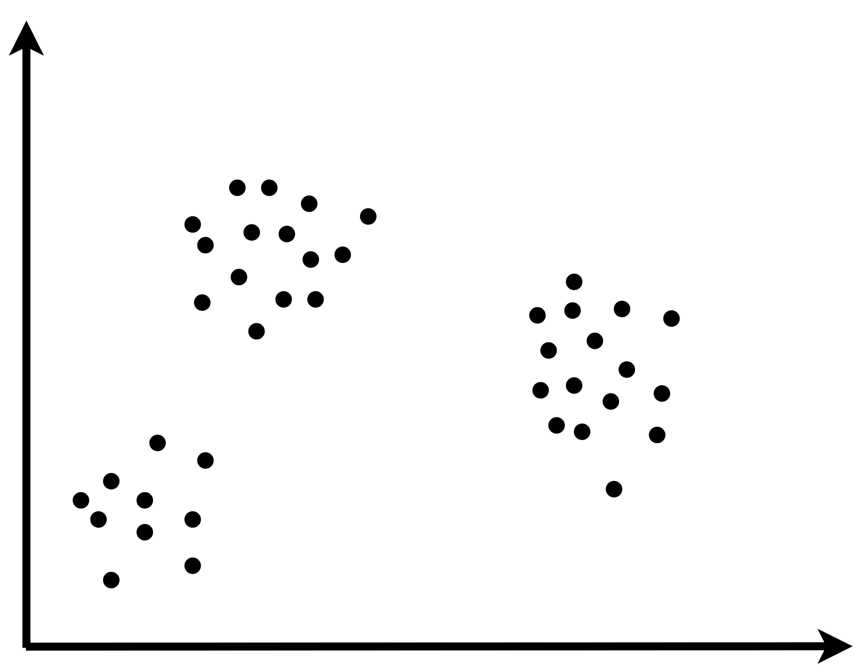
- i.e. $G = \sum_{k=1}^{\infty} \rho_k \delta_{\mu_k} \overset{d}{=} \text{DP}(\alpha, \mathcal{N}(\mu_0, \Sigma_0))$



$$z_n \overset{iid}{\sim} \text{Categorical}(\rho)$$

$$\mu_n^* = \mu_{z_n}$$

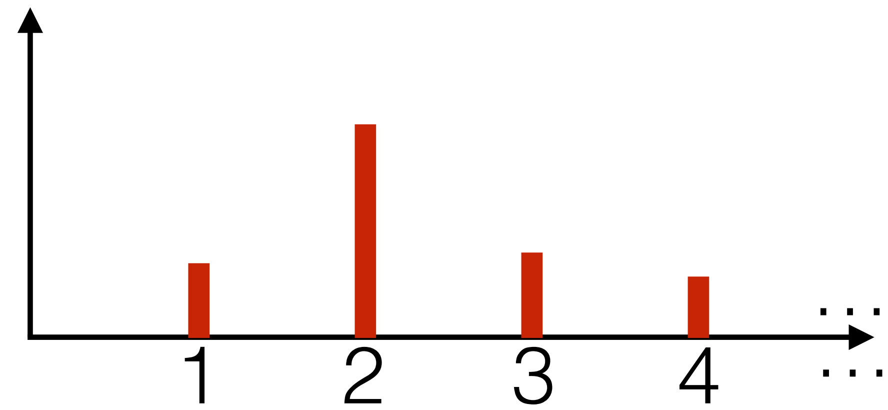
- i.e. $\mu_n^* \overset{iid}{\sim} G$



$$x_n \overset{indep}{\sim} \mathcal{N}(\mu_n^*, \Sigma)$$

Dirichlet process mixture model

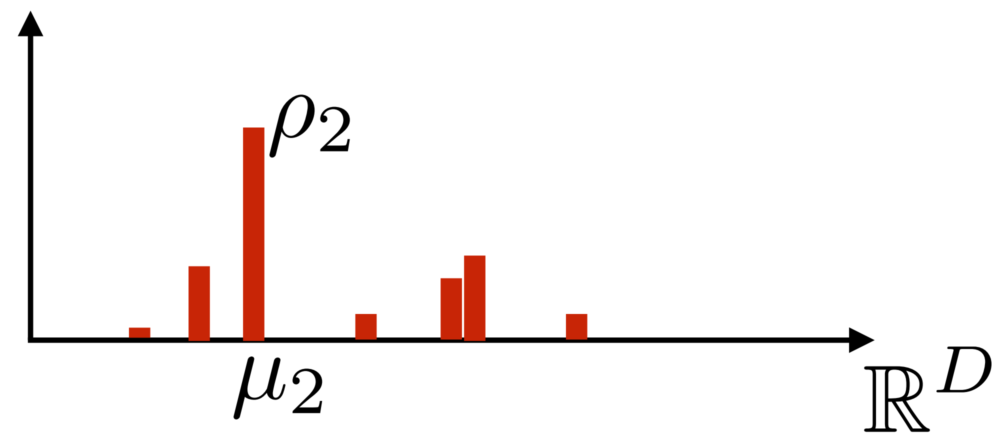
- Gaussian mixture model



$$\rho = (\rho_1, \rho_2, \ldots) \sim \mathrm{GEM}(\alpha)$$

$$\mu_k \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0), k = 1, 2, \ldots$$

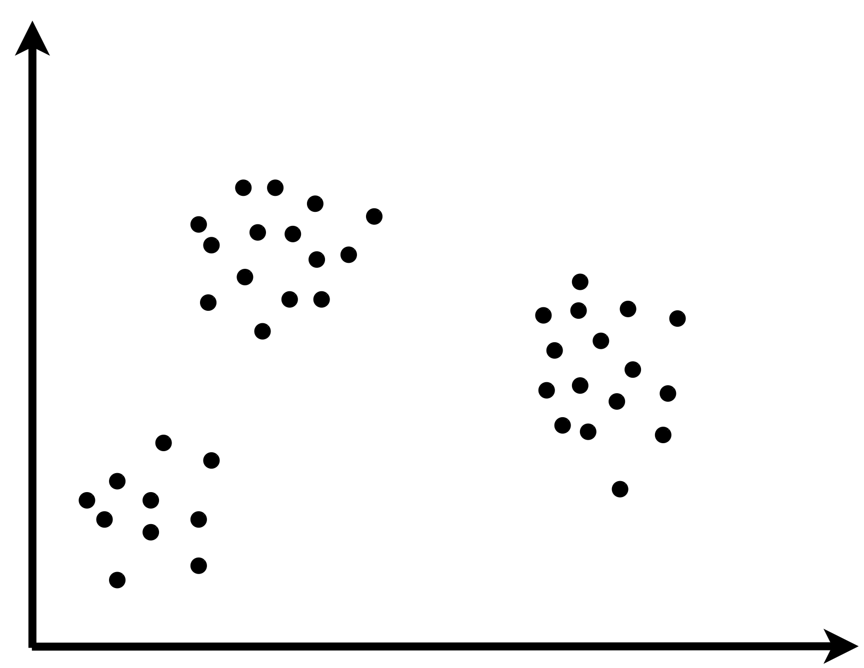
- i.e. $G = \sum_{k=1}^{\infty} \rho_k \delta_{\mu_k} \overset{d}{=} \mathrm{DP}(\alpha, \mathcal{N}(\mu_0, \Sigma_0))$



$$z_n \overset{iid}{\sim} \mathrm{Categorical}(\rho)$$

$$\mu_n^* = \mu_{z_n}$$

- i.e. $\mu_n^* \overset{iid}{\sim} G$



$$x_n \overset{indep}{\sim} \mathcal{N}(\mu_n^*, \Sigma)$$

[demo]

Roadmap

- Example problem: clustering
- Example NPBayes model: Dirichlet process (DP)
- De Finetti for clustering: Kingman Paintbox
- De Finetti for networks/graphs

- Big questions
  - Why NPBayes?
  - What does an infinite/growing number of parameters really mean (in NPBayes)?
  - Why is NPBayes challenging but practical?
  - How does thinking about exchangeability help us use NPBayes in practice?

Roadmap

- Example problem: clustering
- Example NPBayes model: Dirichlet process (DP)
- De Finetti for clustering: Kingman Paintbox
- De Finetti for networks/graphs

- Big questions
  - Why NPBayes?
  - What does an infinite/growing number of parameters really mean (in NPBayes)?
  - Why is NPBayes challenging but practical?
  - How does thinking about exchangeability help us use NPBayes in practice?

Roadmap

- Example problem: clustering
- Example NPBayes model: Dirichlet process (DP)
- De Finetti for clustering: Kingman Paintbox
- De Finetti for networks/graphs

- Big questions
  - Why NPBayes?
  - What does an infinite/growing number of parameters really mean (in NPBayes)?
  - Why is NPBayes challenging but practical?
  - How does thinking about exchangeability help us use NPBayes in practice?

Roadmap

- Example problem: clustering
- Example NPBayes model: Dirichlet process (DP)
- De Finetti for clustering: Kingman Paintbox
- De Finetti for networks/graphs

- Big questions
  - Why NPBayes? **Learn more from more data**
  - What does an infinite/growing number of parameters really mean (in NPBayes)?
  - Why is NPBayes challenging but practical?
  - How does thinking about exchangeability help us use NPBayes in practice?

Roadmap

- Example problem: clustering
- Example NPBayes model: Dirichlet process (DP)
- De Finetti for clustering: Kingman Paintbox
- De Finetti for networks/graphs

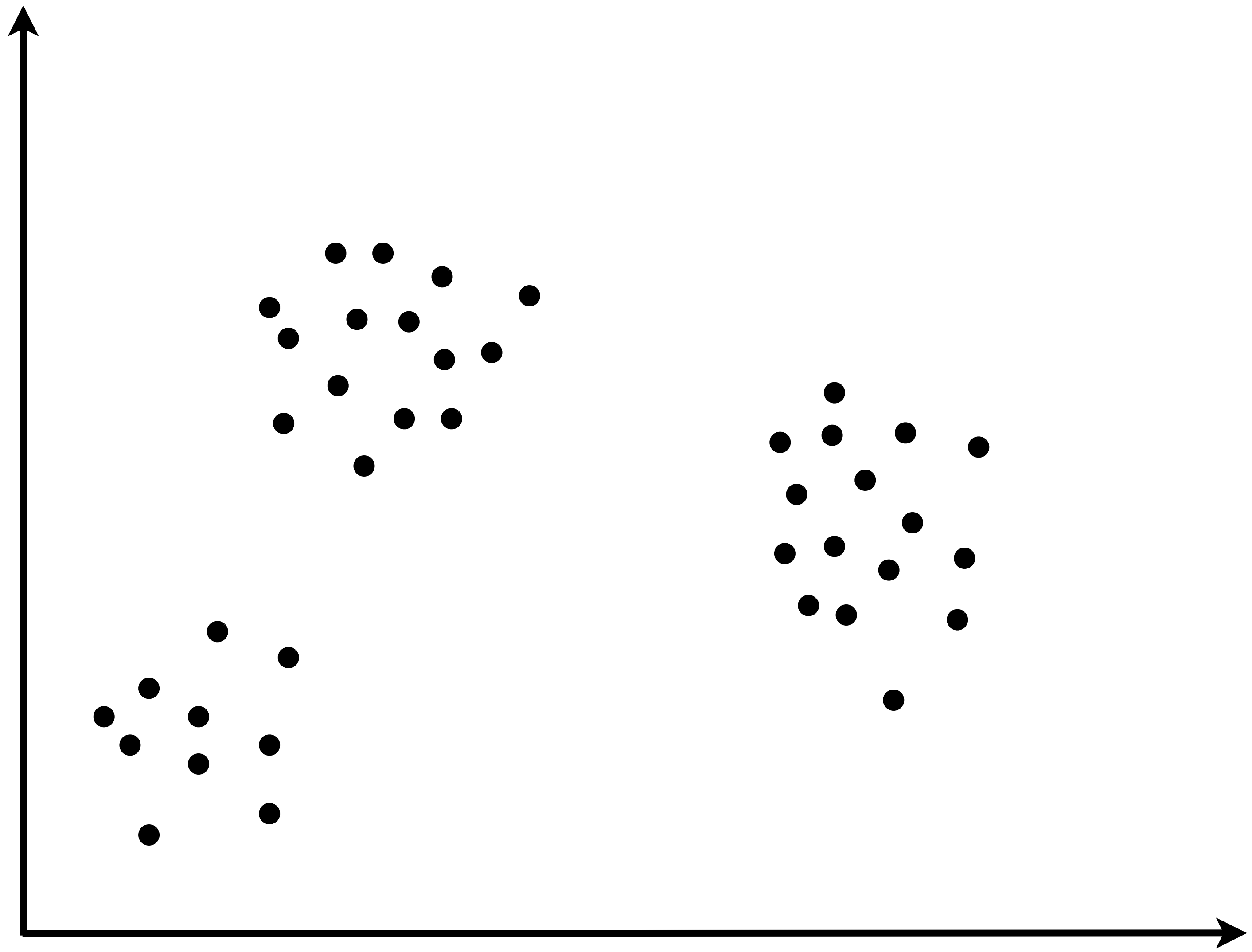
- Big questions
  - Why NPBayes? **Learn more from more data**
  - What does an infinite/growing number of parameters really mean (in NPBayes)? **Components vs clusters; latent vs realized**
  - Why is NPBayes challenging but practical?
  - How does thinking about exchangeability help us use NPBayes in practice?

Roadmap

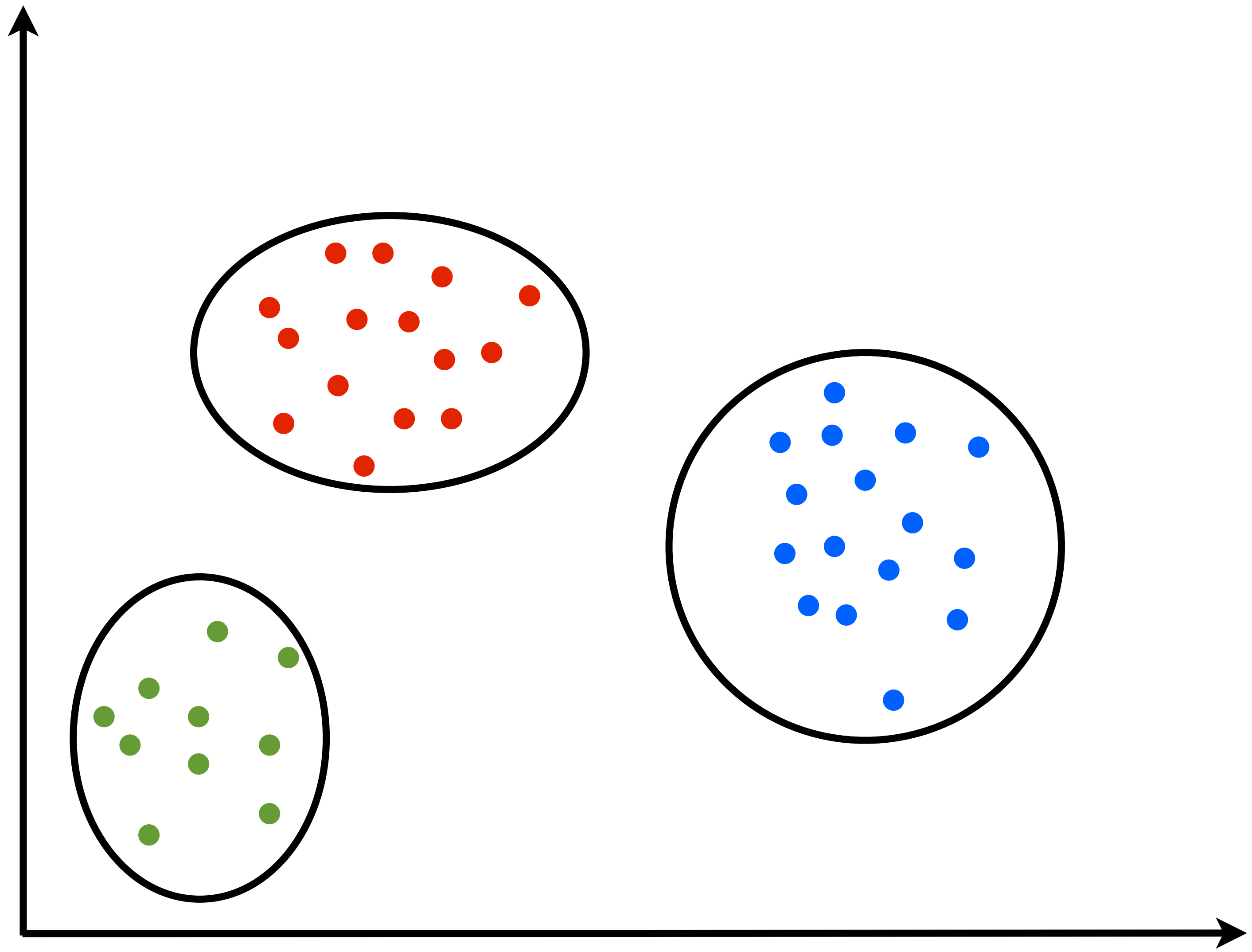
- Example problem: clustering
- Example NPBayes model: Dirichlet process (DP)
- De Finetti for clustering: Kingman Paintbox
- De Finetti for networks/graphs

- Big questions
  - Why NPBayes? **Learn more from more data**
  - What does an infinite/growing number of parameters really mean (in NPBayes)? **Components vs clusters; latent vs realized**
  - Why is NPBayes challenging but practical? **Infinite dimensional parameter but finitely many realized (in practice, e.g., can integrate out or truncate the infinity)**
  - How does thinking about exchangeability help us use NPBayes in practice?

Clustering

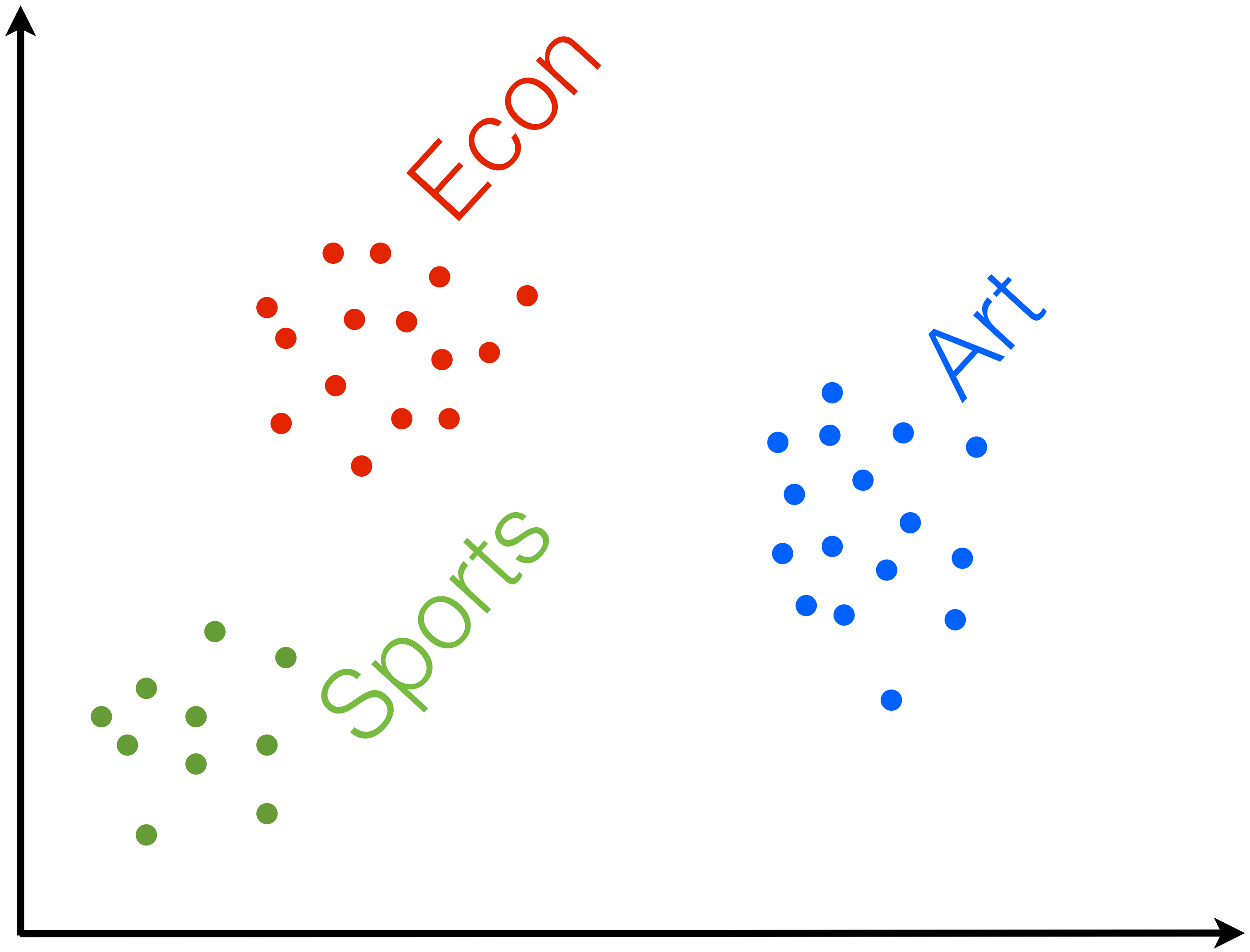


Clustering



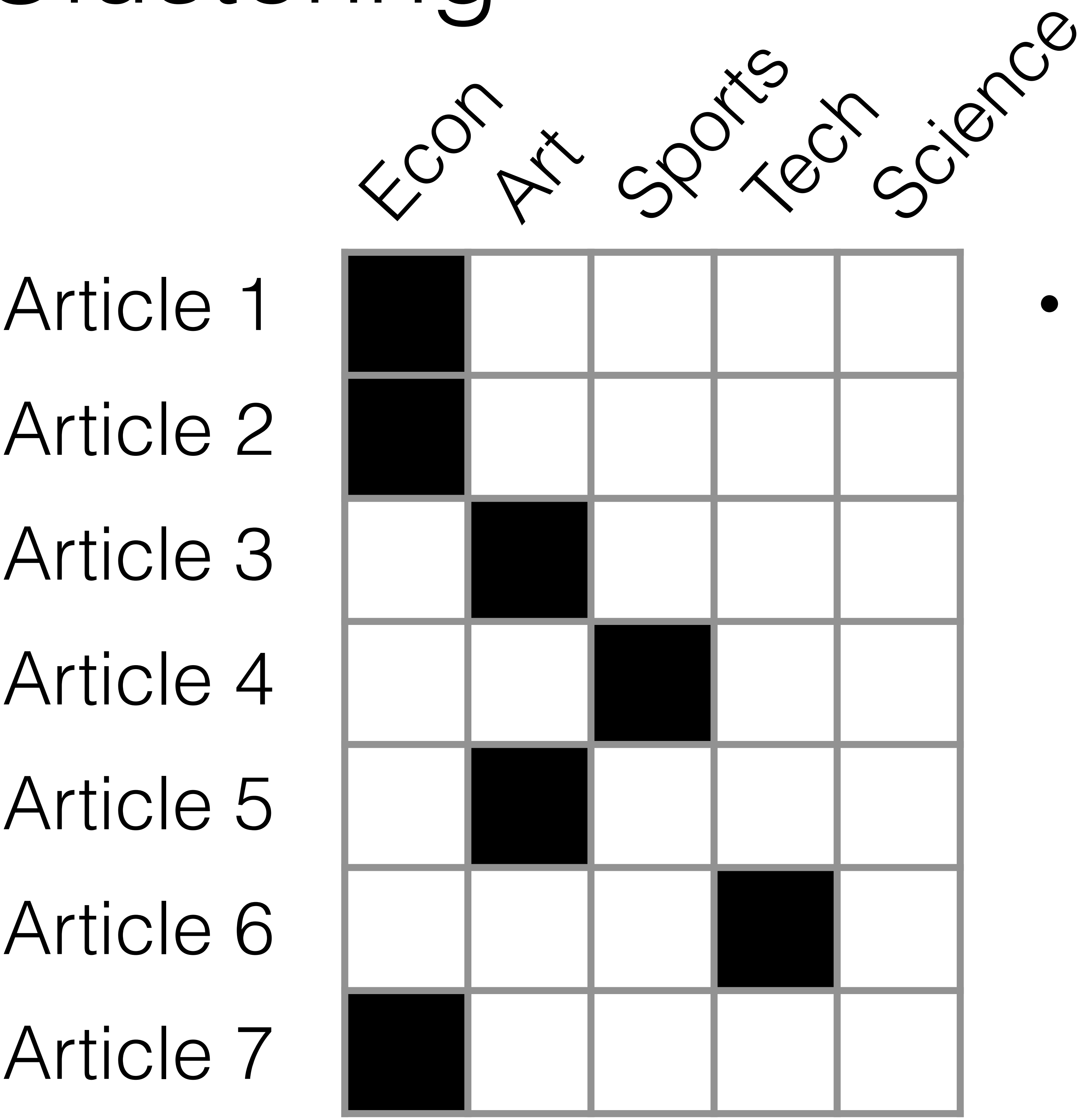
“Clusters”

Clustering



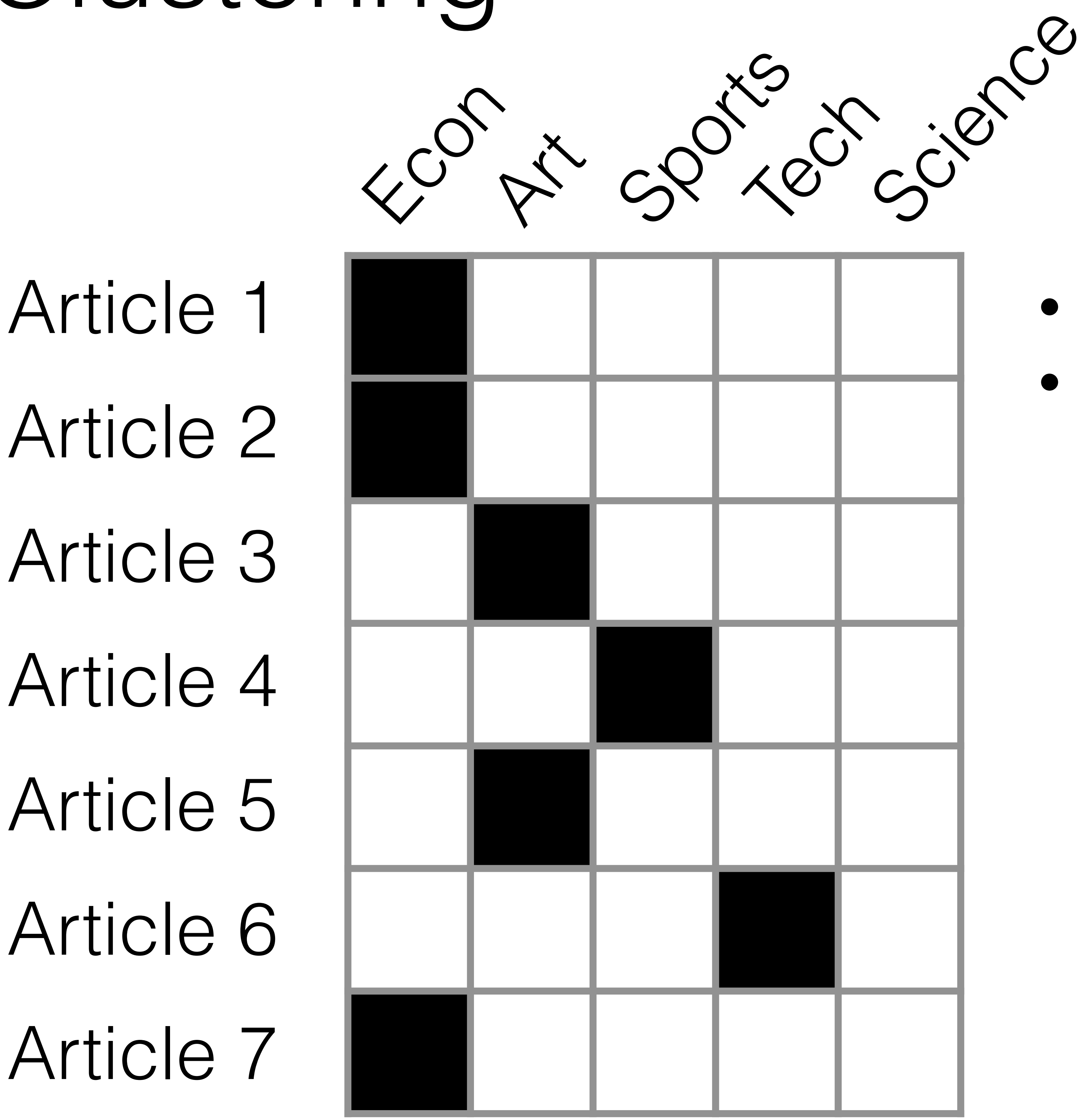
"Clusters"

Clustering



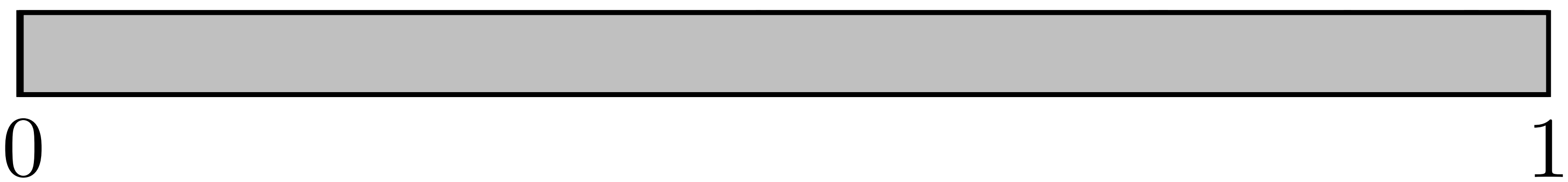
- Groups: clusters

Clustering



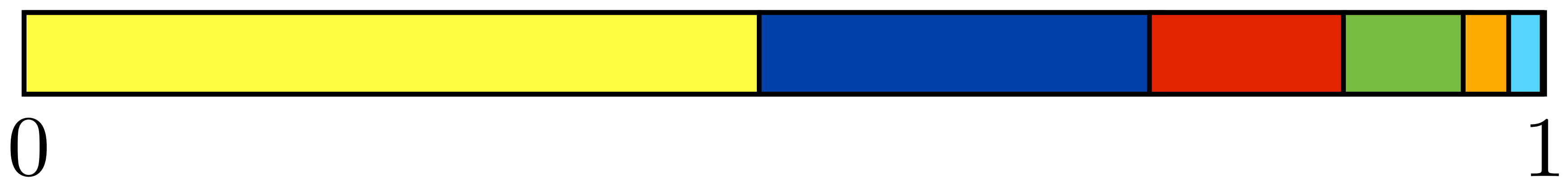
- Groups: clusters
- Exchangeable

Clustering



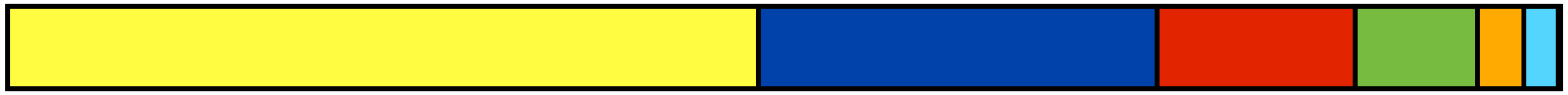
[Kingman 1978]

Clustering



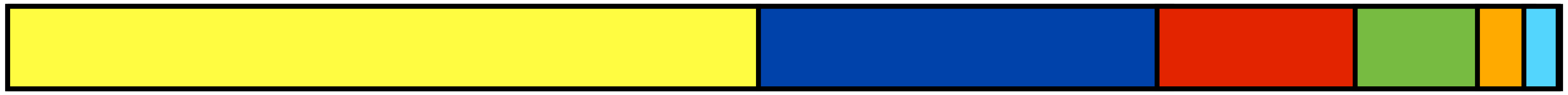
[Kingman 1978]

Clustering



[Kingman 1978]

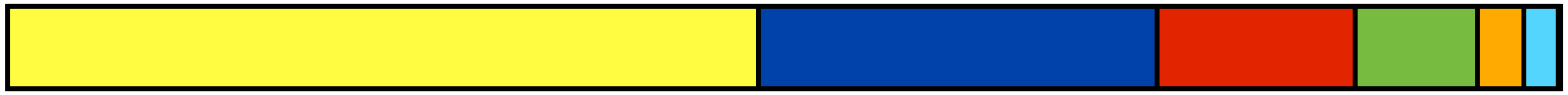
Clustering



- Finite example: Dirichlet

[Kingman 1978]

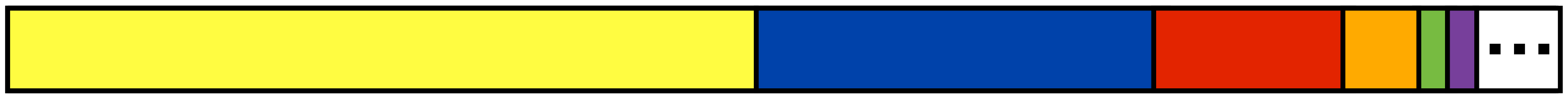
Clustering



- Finite example: Dirichlet
- Infinite example: GEM / DP

[Kingman 1978]

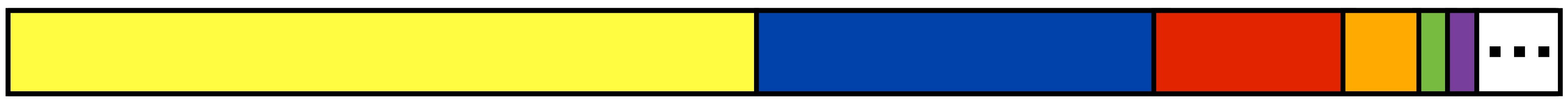
Clustering



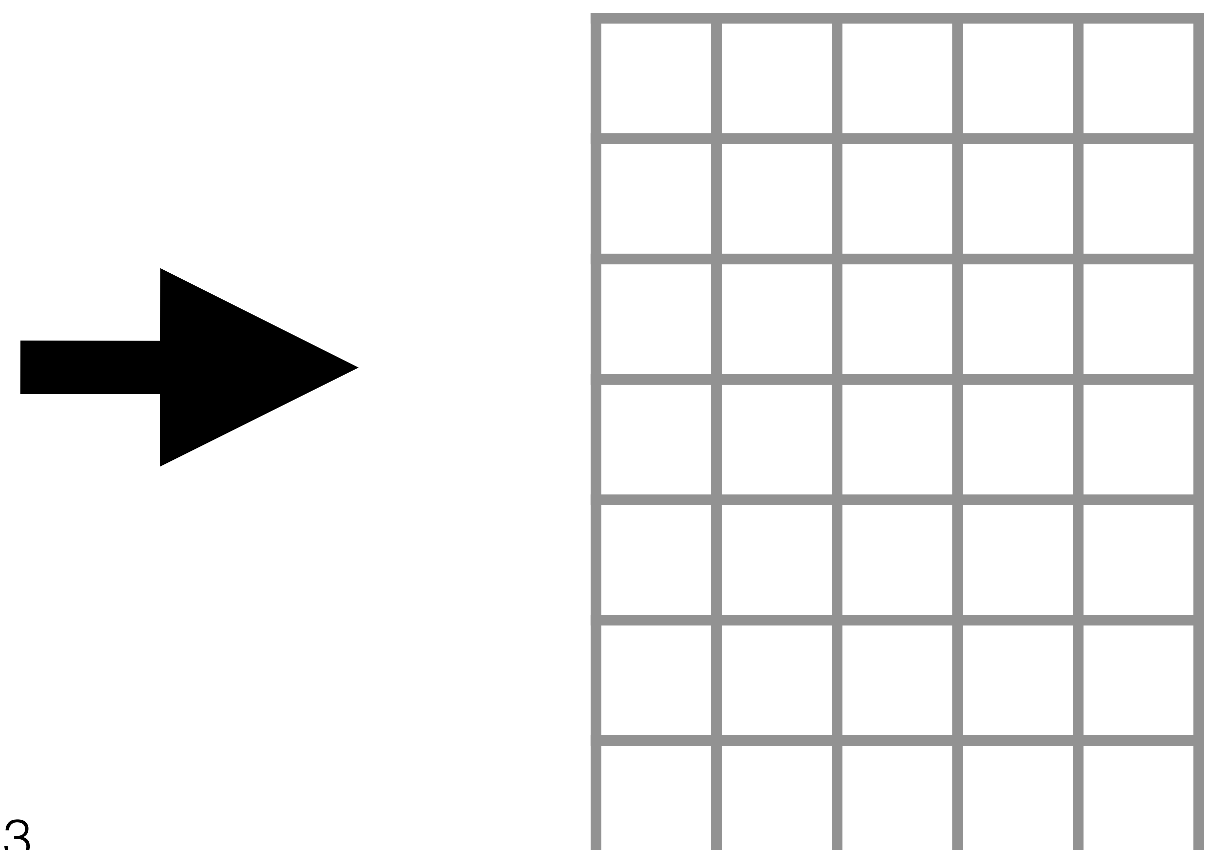
- Finite example: Dirichlet
- Infinite example: GEM / DP

[Kingman 1978]

Clustering

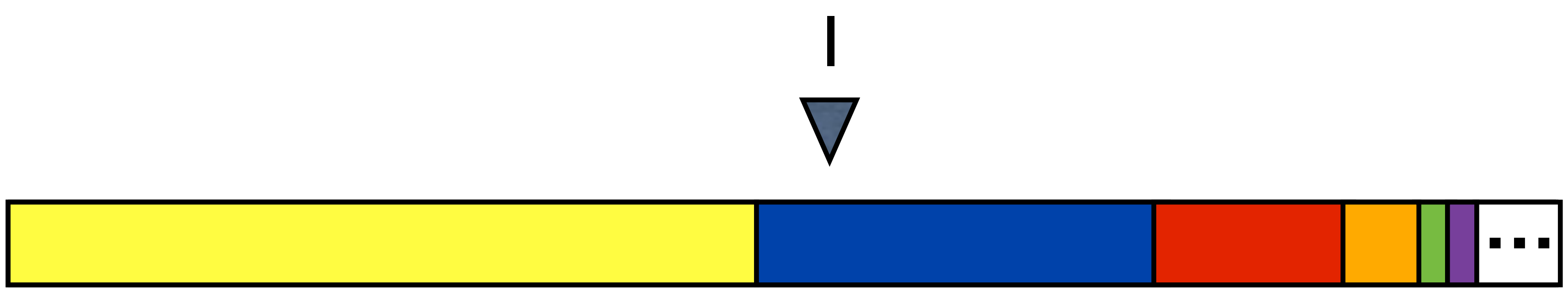


- Finite example: Dirichlet
- Infinite example: GEM / DP

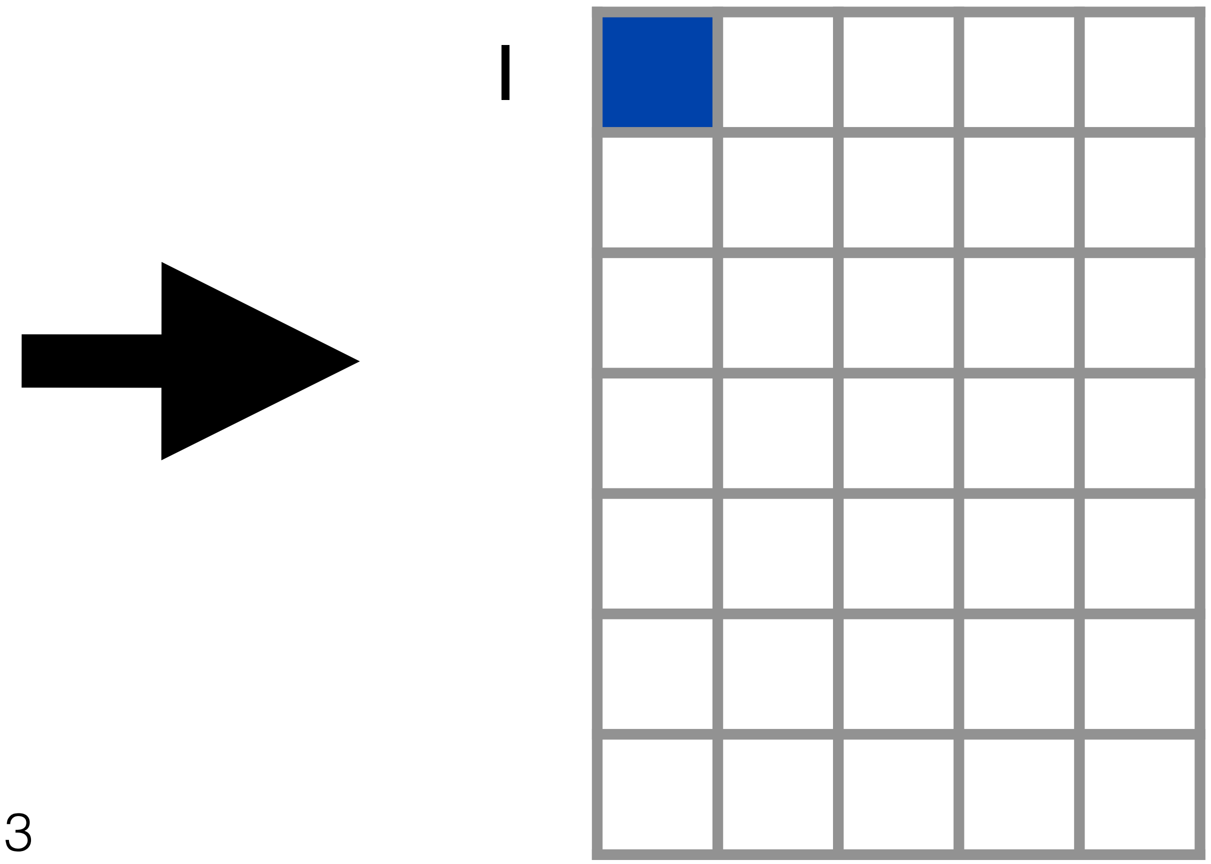


[Kingman 1978]

Clustering

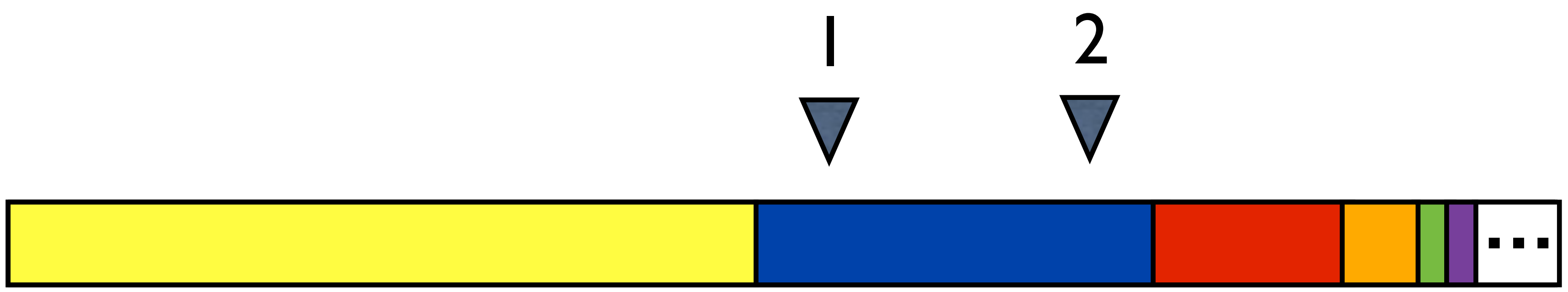


- Finite example: Dirichlet
- Infinite example: GEM / DP

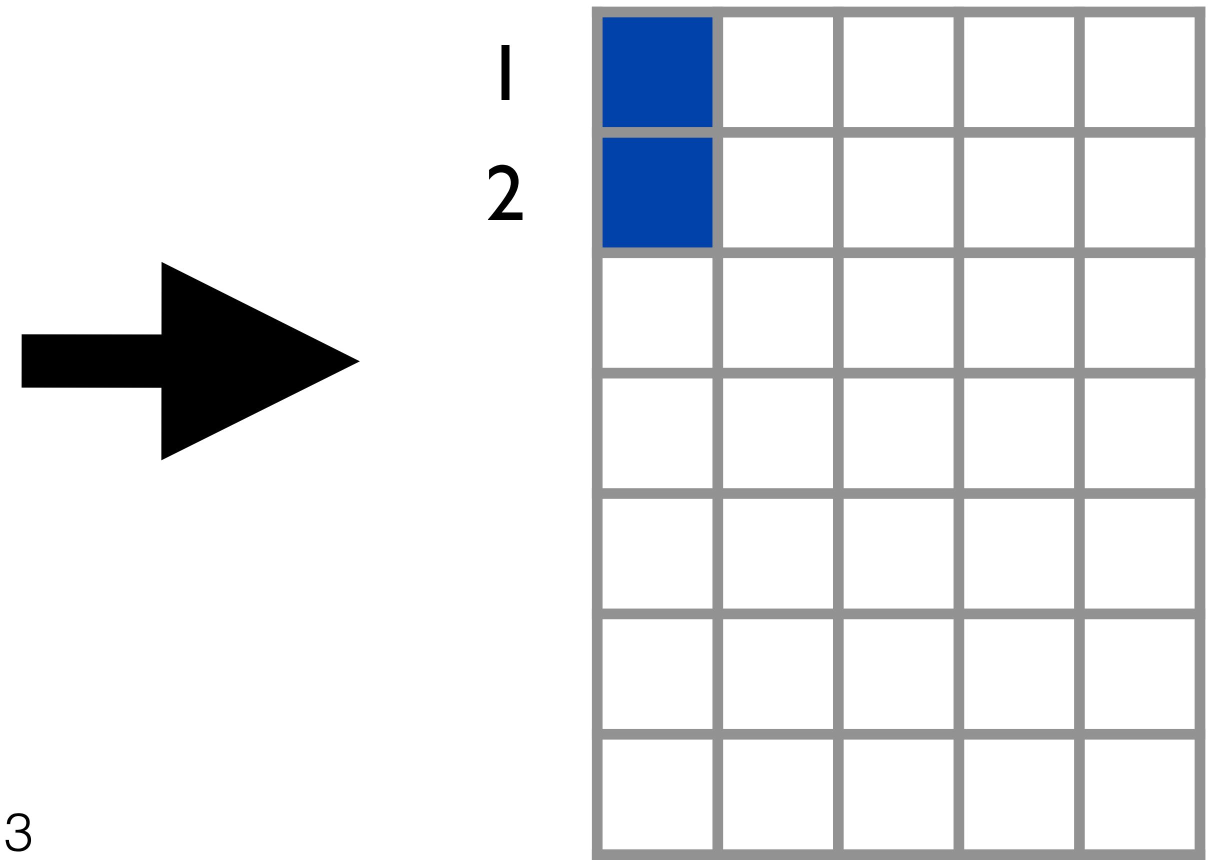


[Kingman 1978]

Clustering

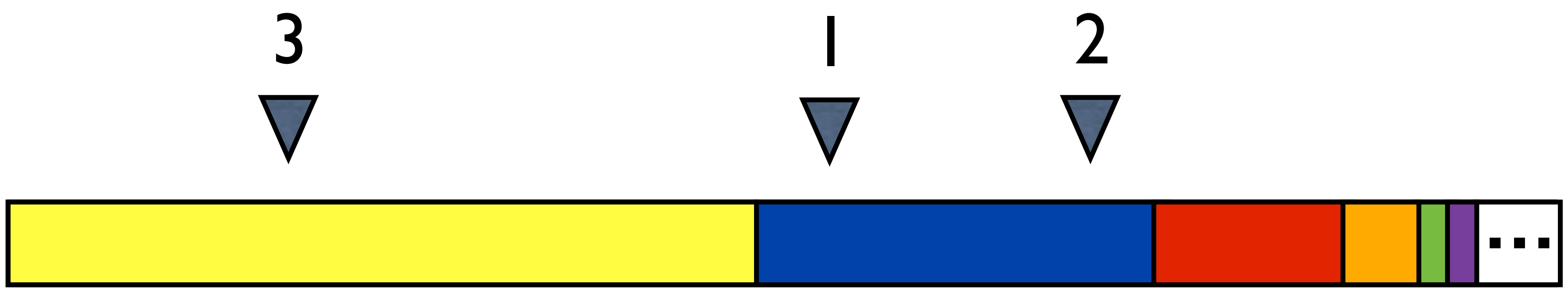


- Finite example: Dirichlet
- Infinite example: GEM / DP

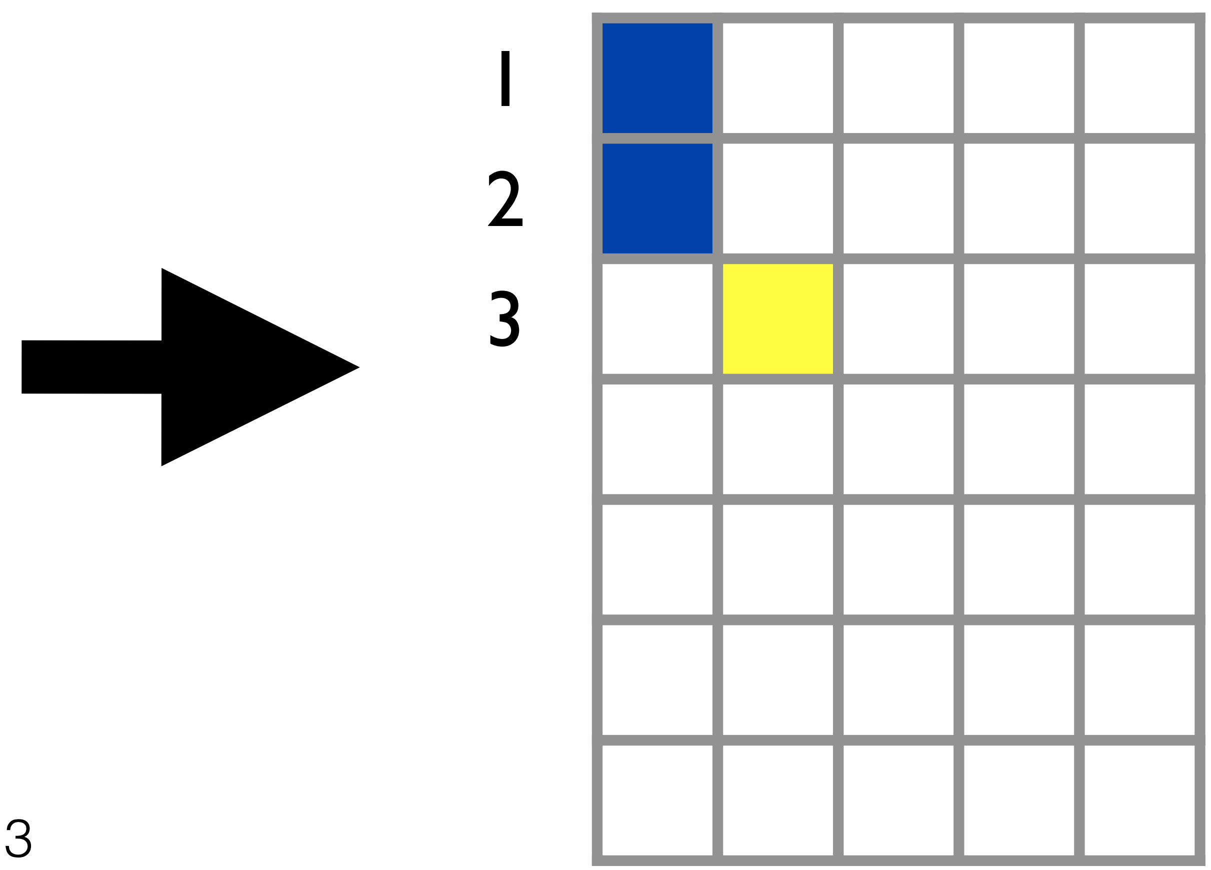


[Kingman 1978]

Clustering

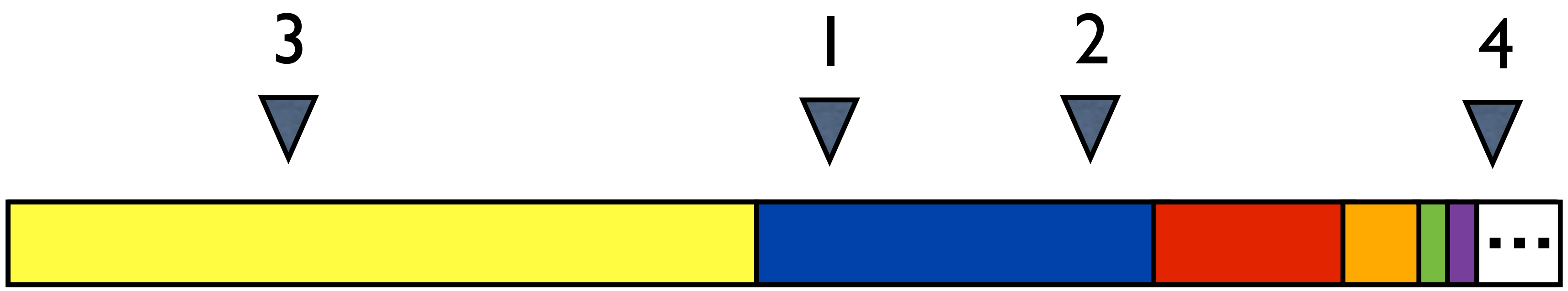


- Finite example: Dirichlet
- Infinite example: GEM / DP

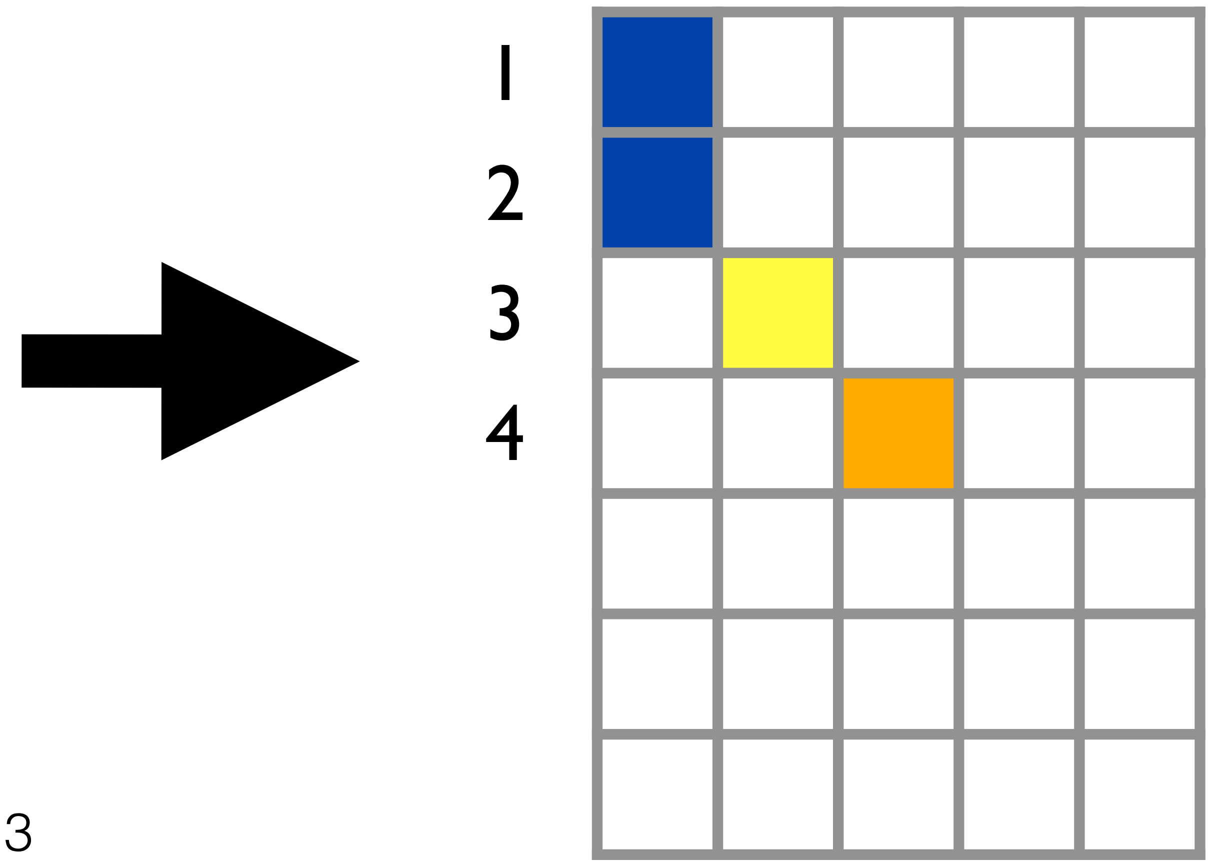


[Kingman 1978]

Clustering

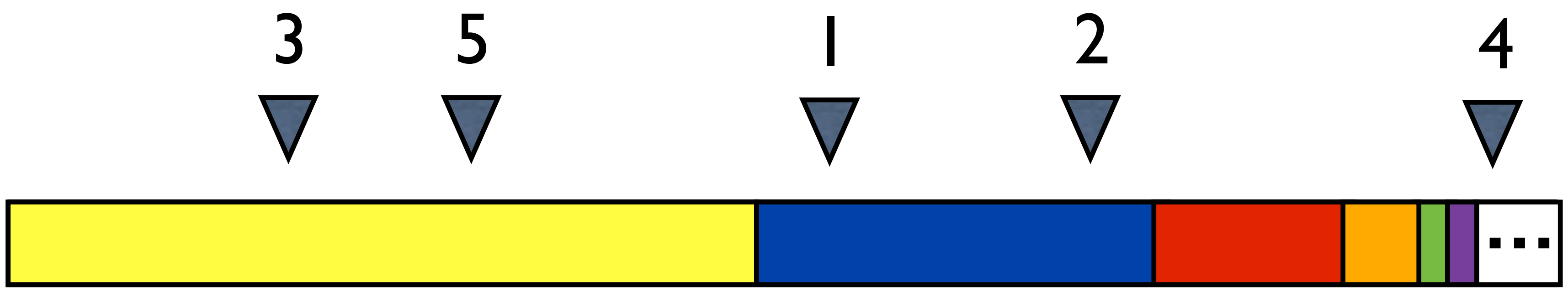


- Finite example: Dirichlet
- Infinite example: GEM / DP

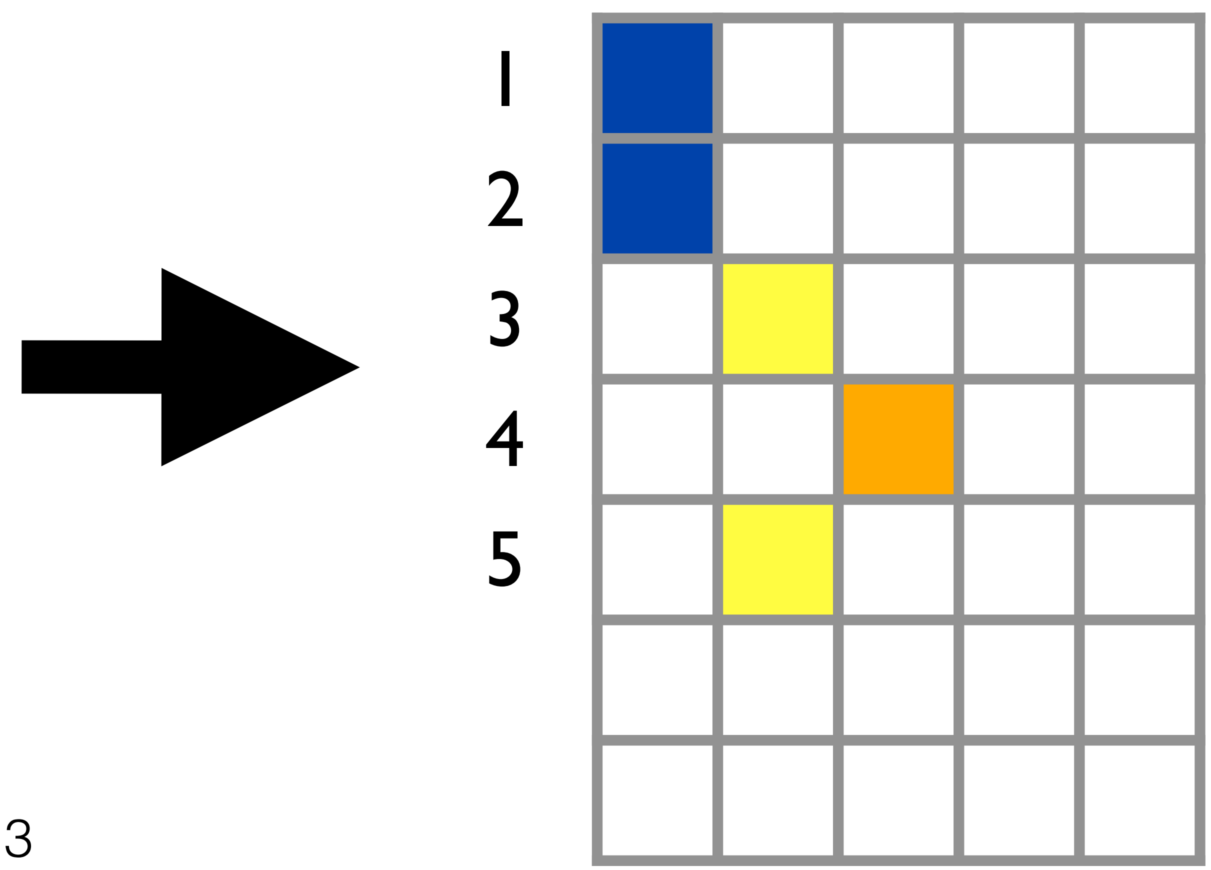


[Kingman 1978]

Clustering

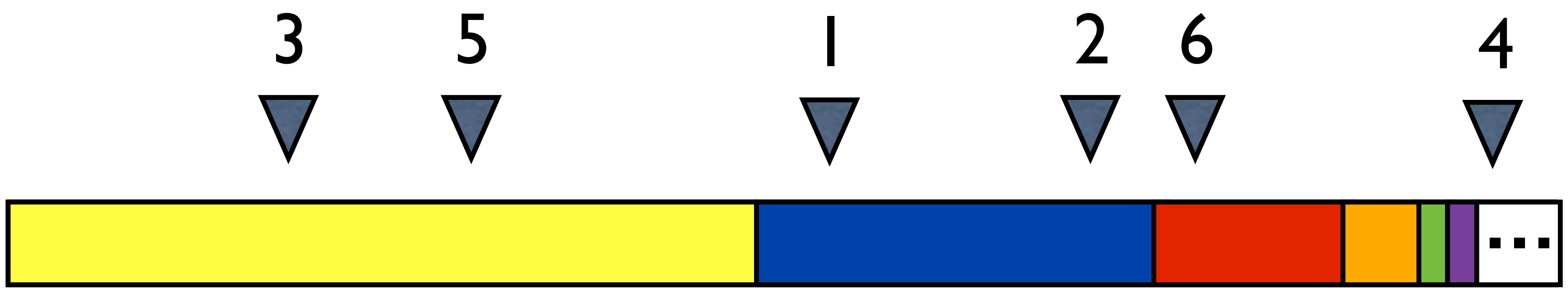


- Finite example: Dirichlet
- Infinite example: GEM / DP

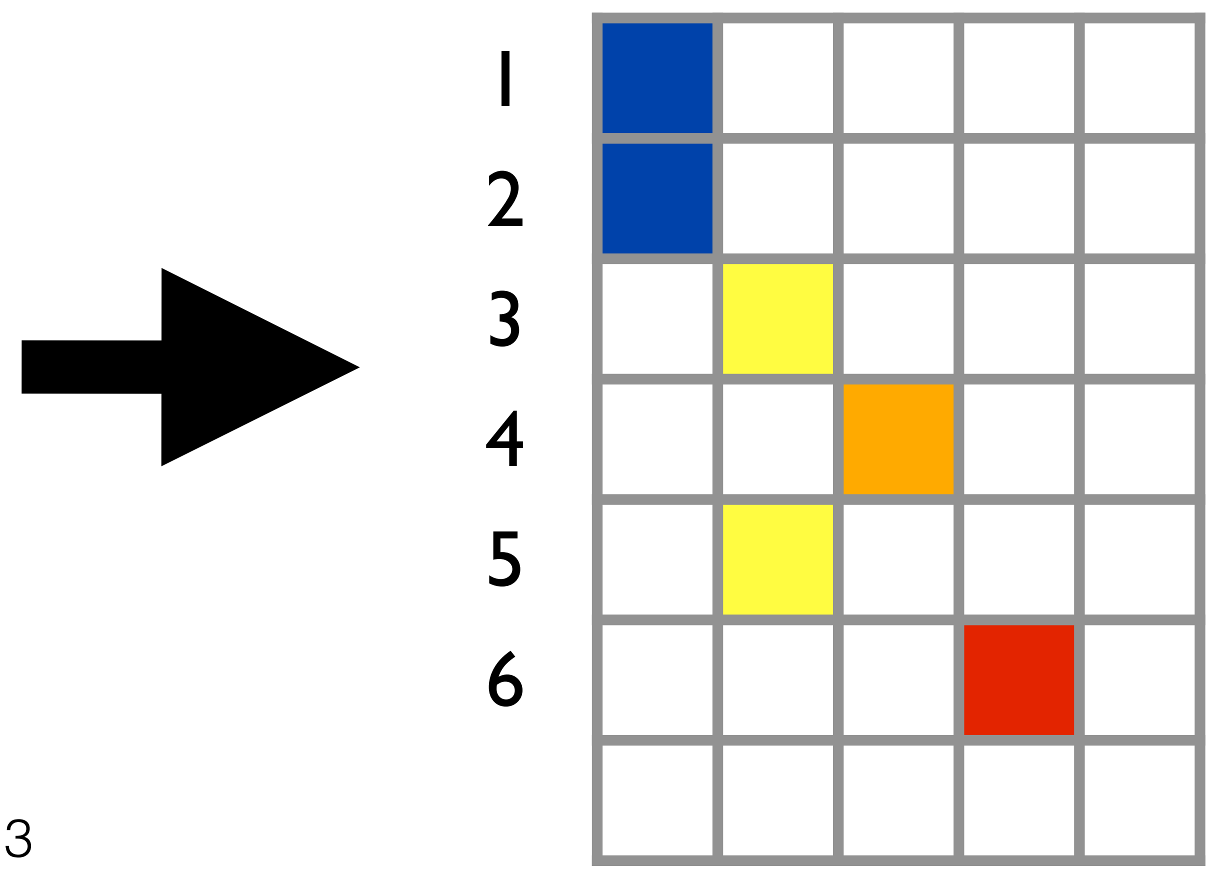


[Kingman 1978]

Clustering

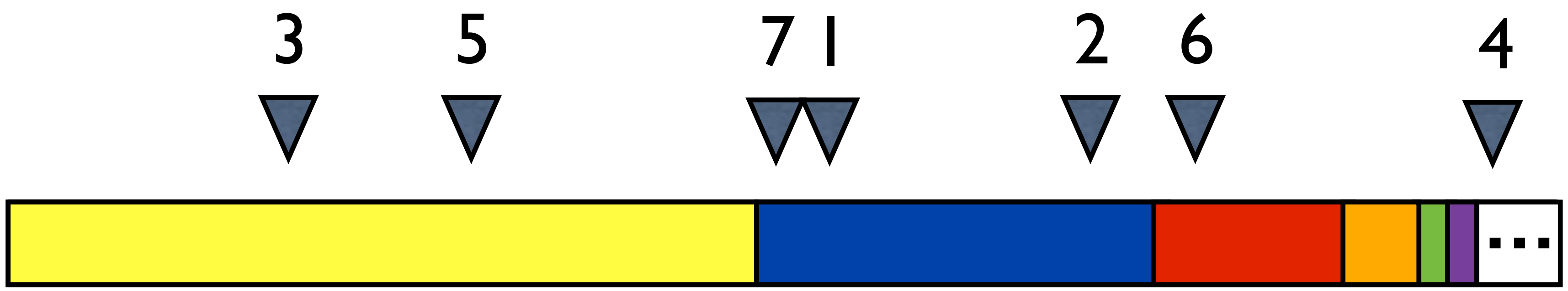


- Finite example: Dirichlet
- Infinite example: GEM / DP

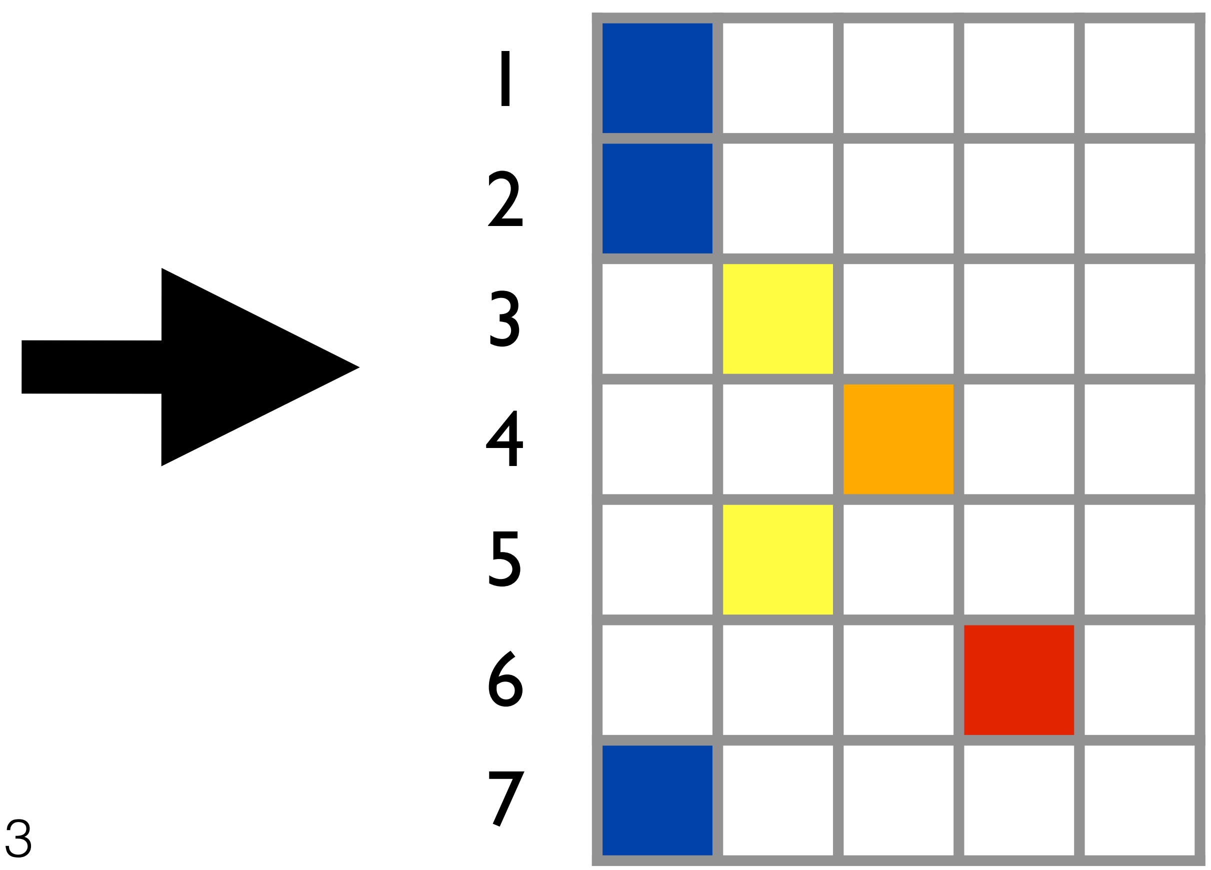


[Kingman 1978]

Clustering



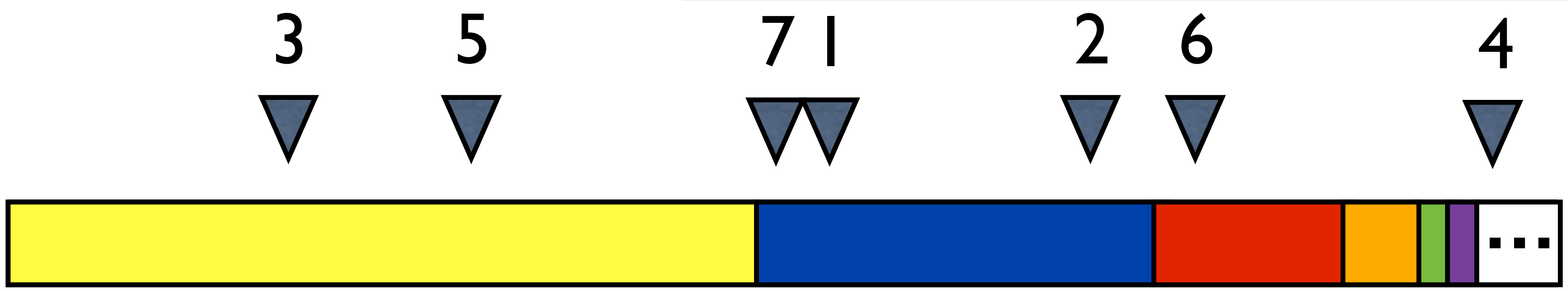
- Finite example: Dirichlet
- Infinite example: GEM / DP



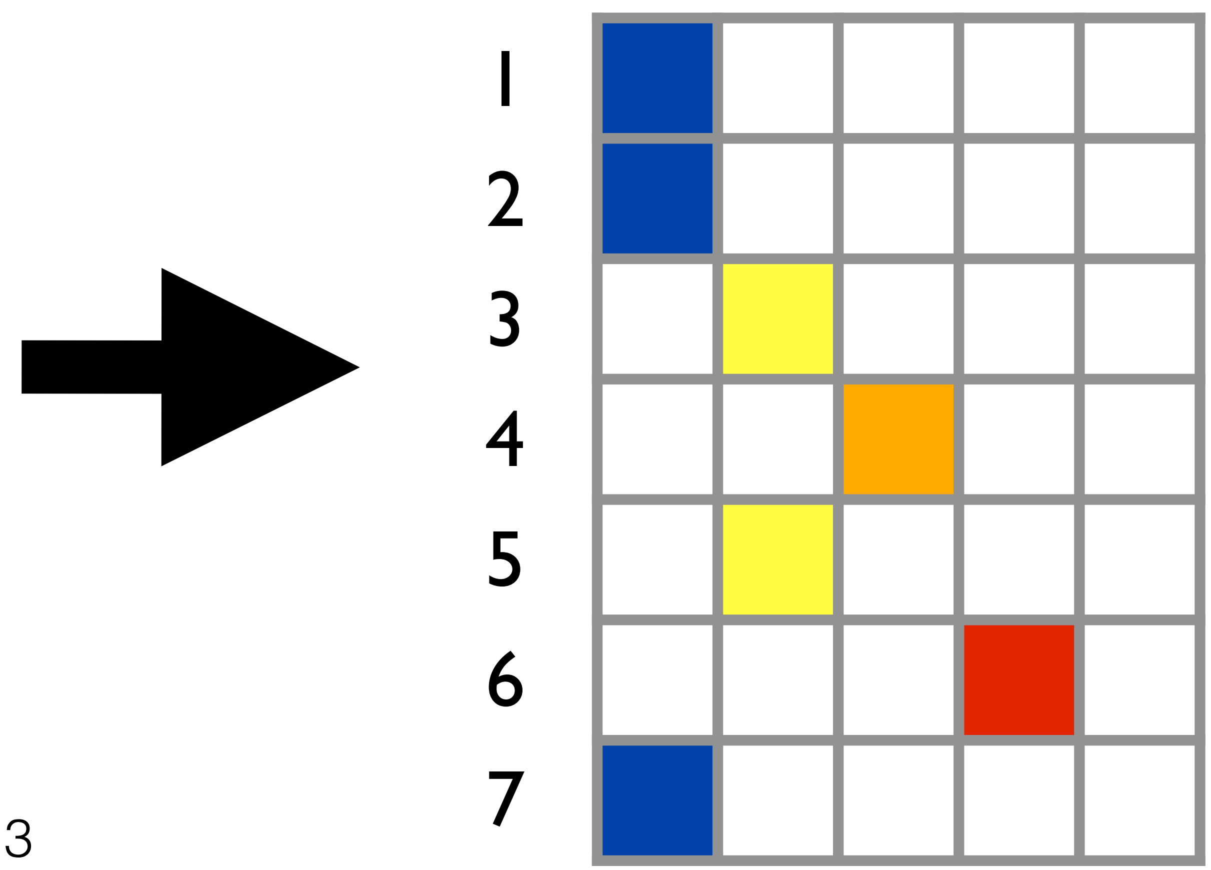
[Kingman 1978]

Clustering

Thm (Kingman). A clustering is exchangeable iff it has a Kingman paintbox representation



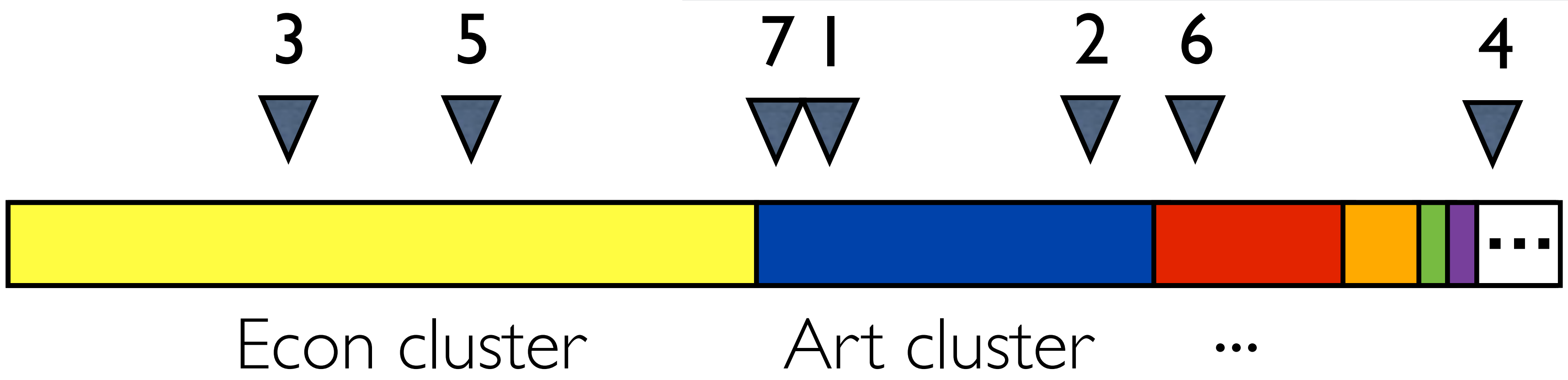
- Finite example: Dirichlet
- Infinite example: GEM / DP



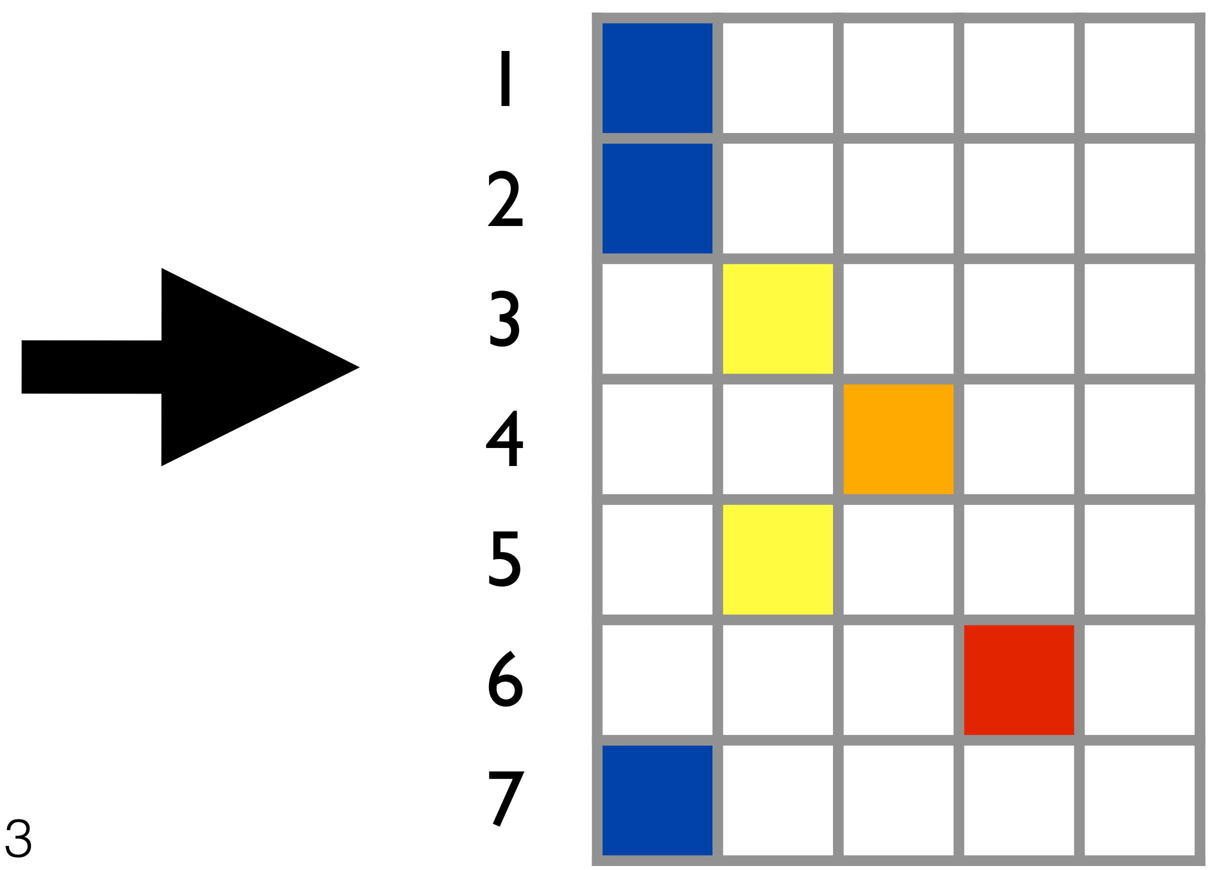
[Kingman 1978]

Clustering

Thm (Kingman). A clustering is exchangeable iff it has a Kingman paintbox representation



- Finite example: Dirichlet
- Infinite example: GEM / DP

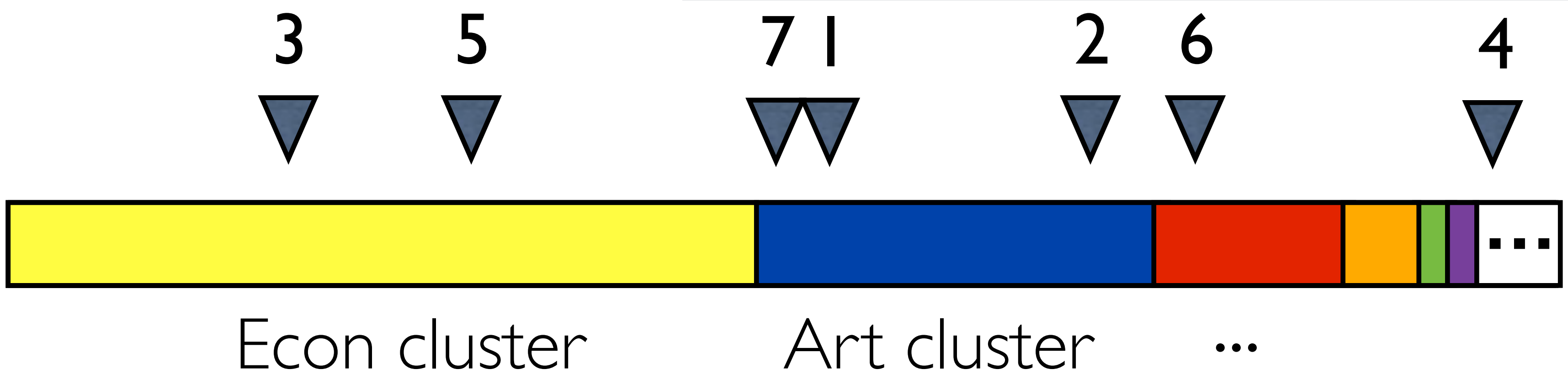


[Kingman 1978]

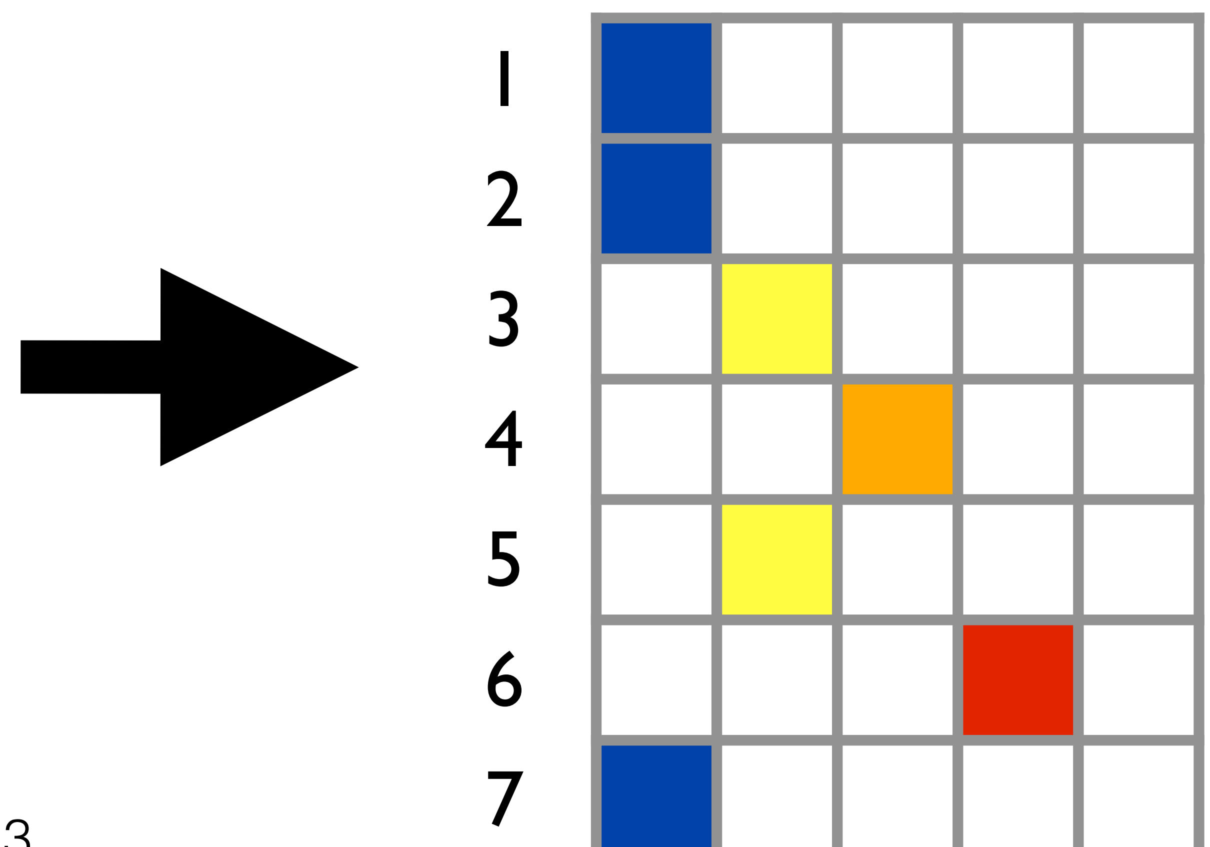
Clustering

Thm (Kingman). A clustering is exchangeable iff it has a Kingman paintbox representation

De Finetti (Bayes, NPBayes)



- Finite example: Dirichlet
- Infinite example: GEM / DP

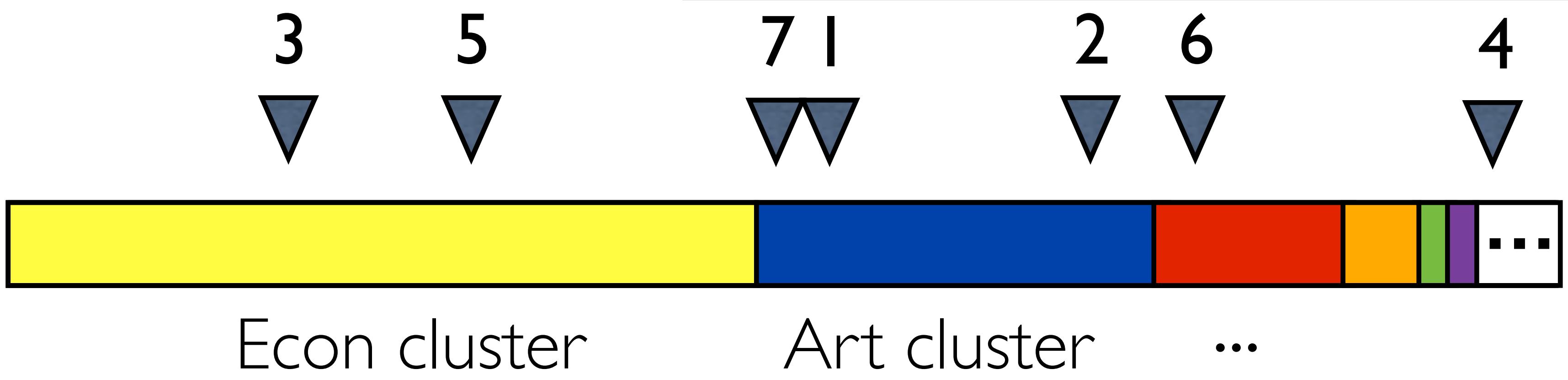


[Kingman 1978]

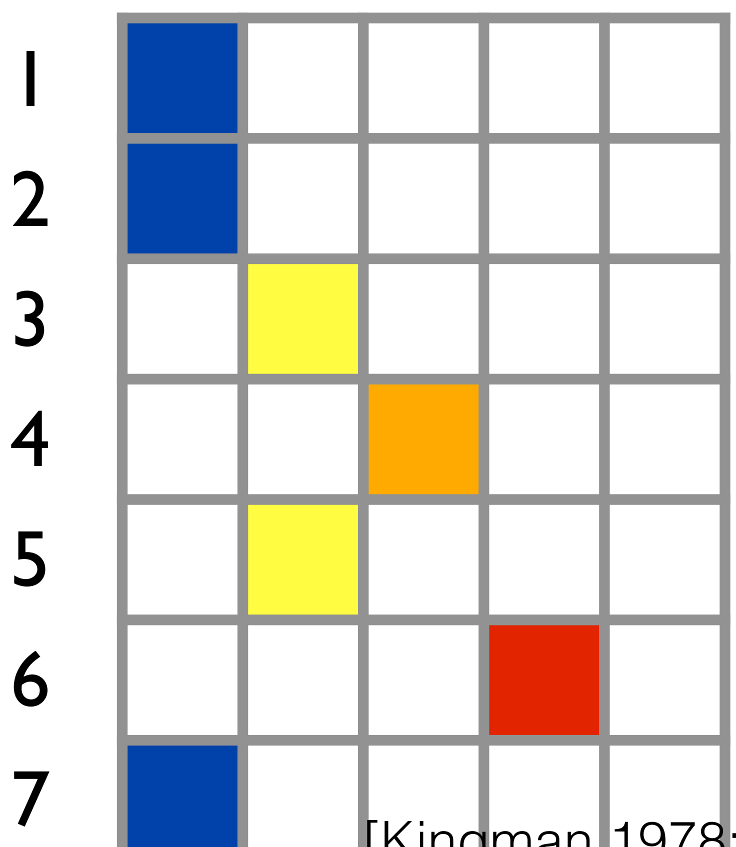
Clustering

Thm (Kingman). A clustering is exchangeable iff it has a Kingman paintbox representation

De Finetti (Bayes, NPBayes)



- Finite example: Dirichlet
- Infinite example: GEM / DP
- Implications:

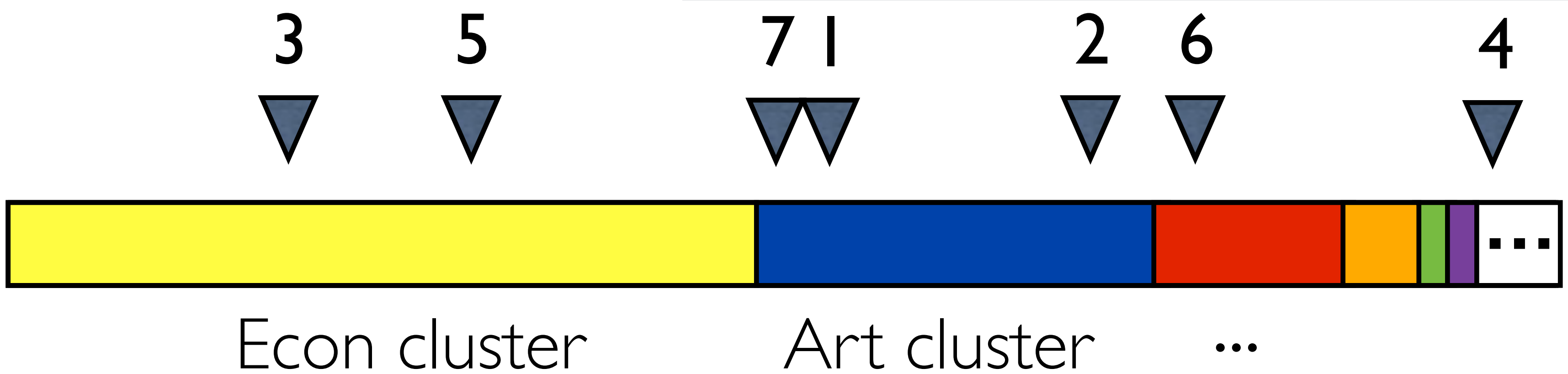


[Kingman 1978; Wallach et al 2010; Broderick, Steorts 2014; Miller et al 2015]

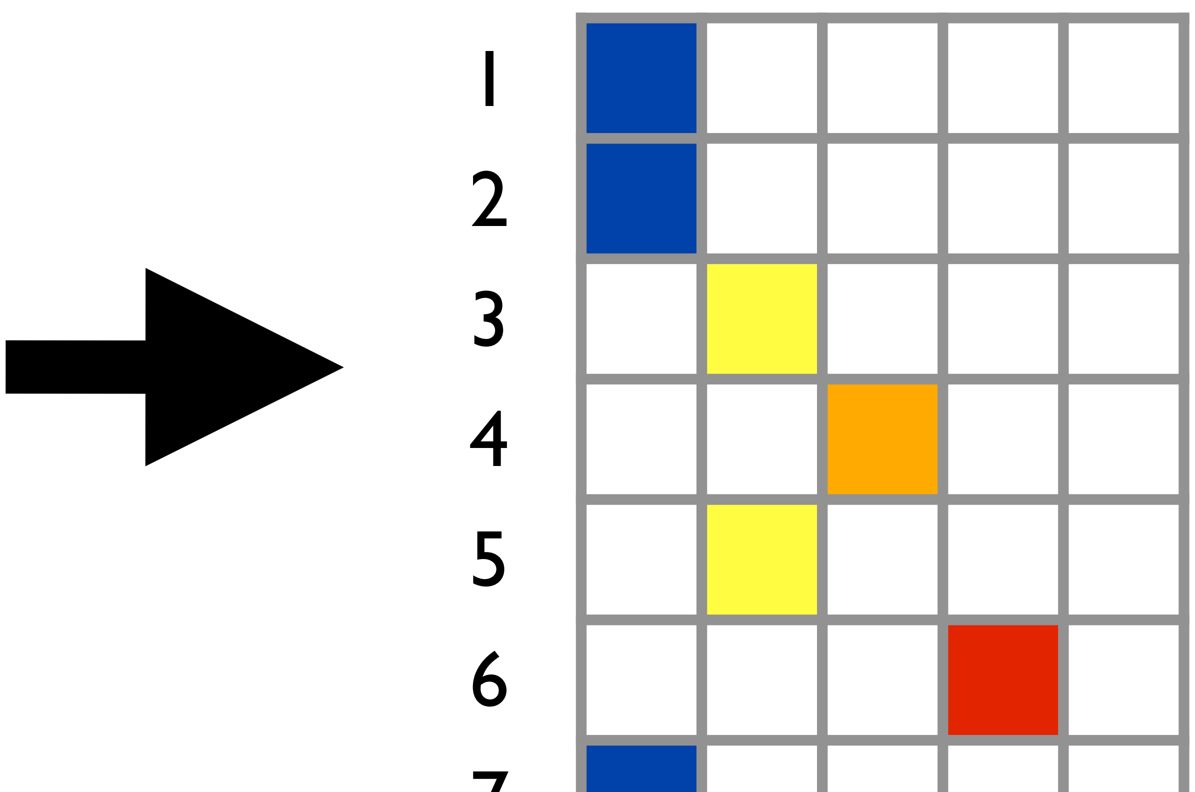
Clustering

Thm (Kingman). A clustering is exchangeable iff it has a Kingman paintbox representation

De Finetti (Bayes, NPBayes)



- Finite example: Dirichlet
- Infinite example: GEM / DP
- Implications:
  - Cluster sizes grow linearly with total # data pts

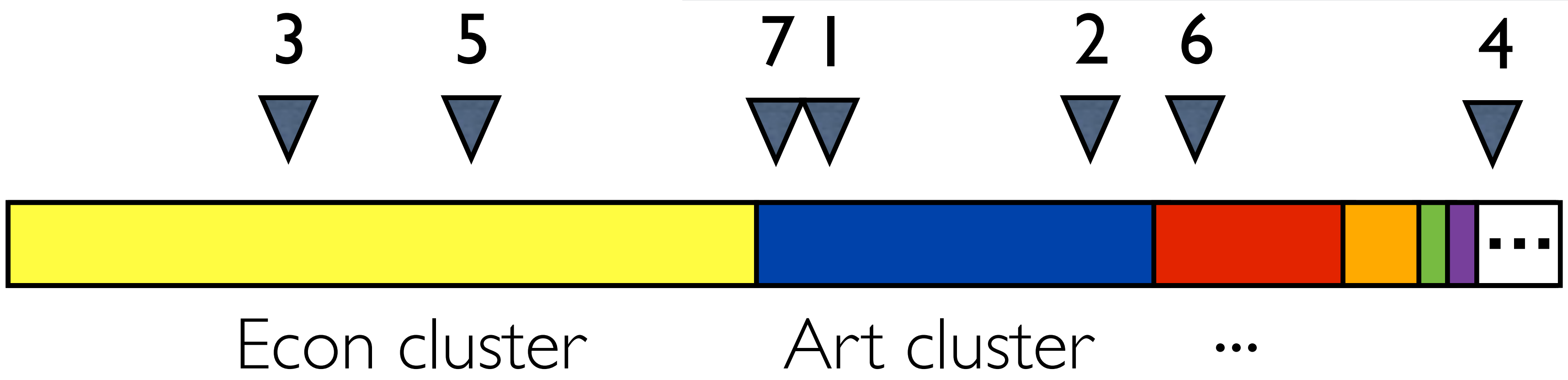


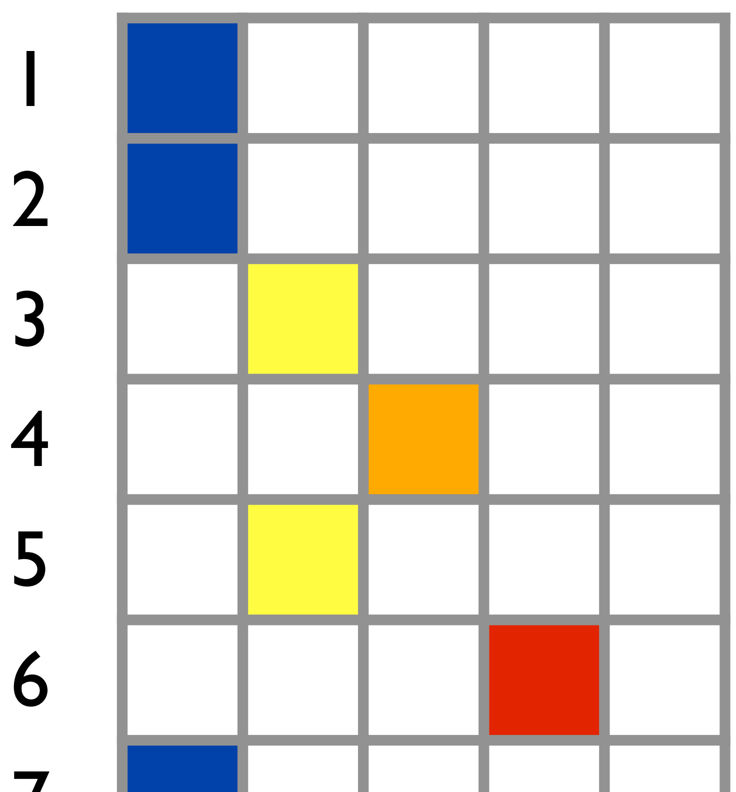
[Kingman 1978; Wallach et al 2010; Broderick, Steorts 2014; Miller et al 2015]

Clustering

Thm (Kingman). A clustering is exchangeable iff it has a Kingman paintbox representation

De Finetti (Bayes, NPBayes)





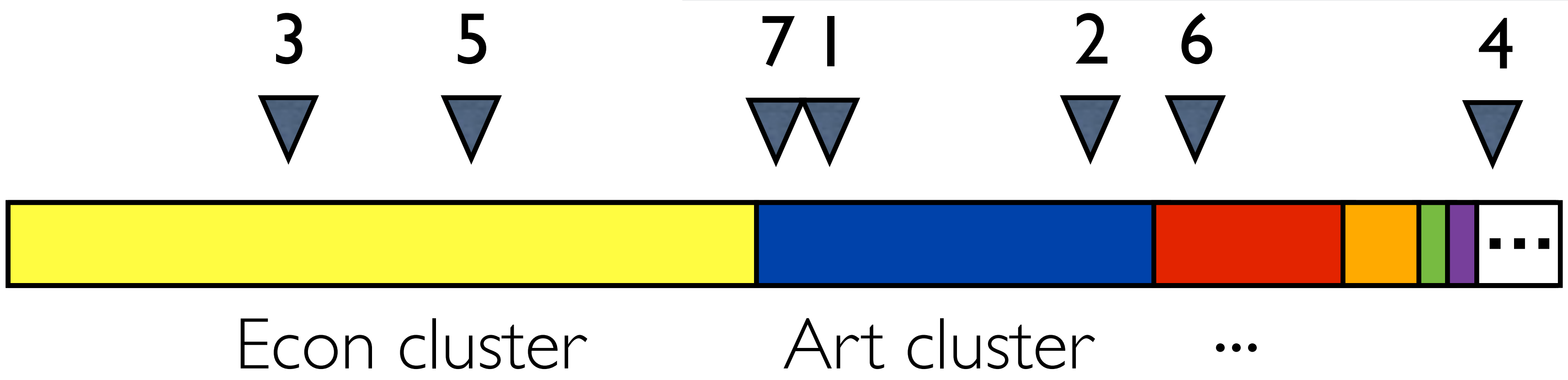
- Finite example: Dirichlet
- Infinite example: GEM / DP
- Implications:
  - Cluster sizes grow linearly with total # data pts
  - Stationary proportions

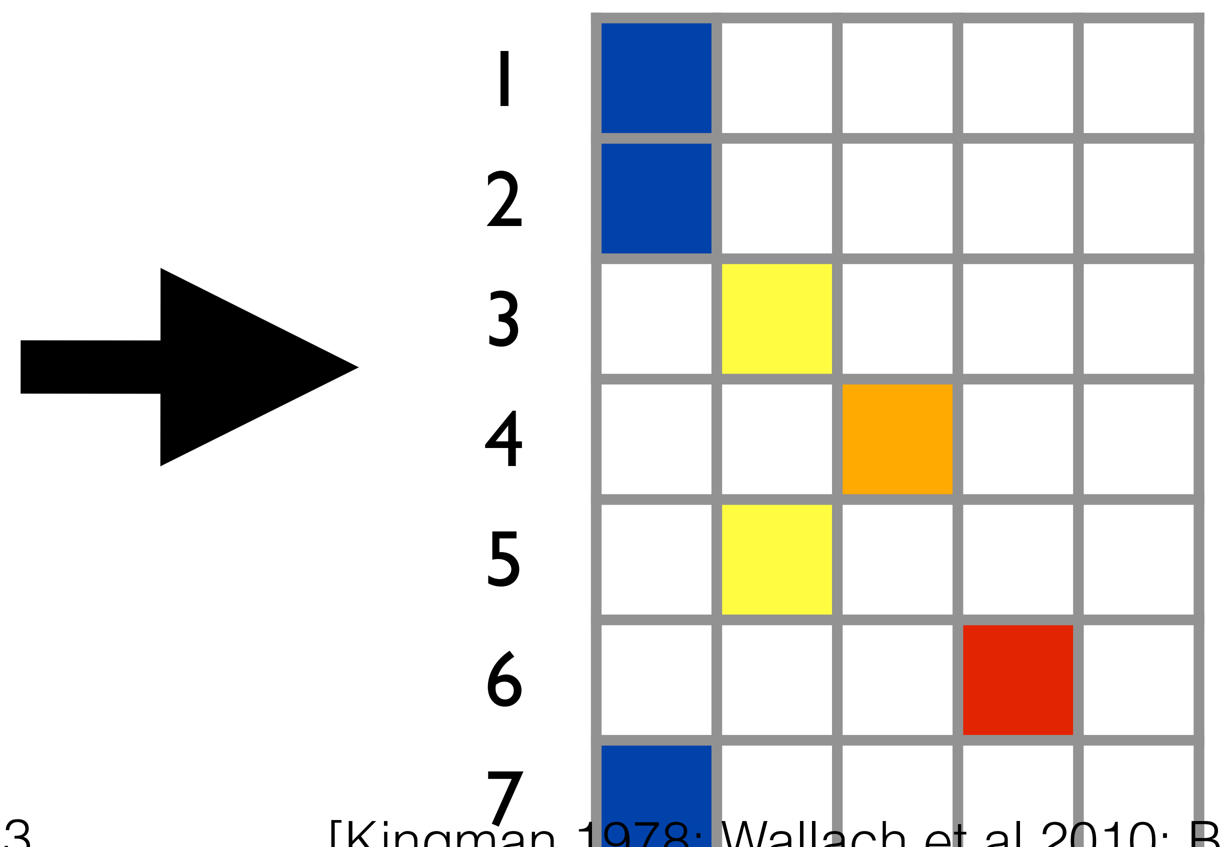
[Kingman 1978; Wallach et al 2010; Broderick, Steorts 2014; Miller et al 2015; Campbell et al 2019]

Clustering

Thm (Kingman). A clustering is exchangeable iff it has a Kingman paintbox representation

De Finetti (Bayes, NPBayes)





- Finite example: Dirichlet
- Infinite example: GEM / DP
- Implications:
  - Cluster sizes grow linearly with total # data pts
  - Stationary proportions
  - See Campbell Tues 1:45 talk “Local exchangeability”

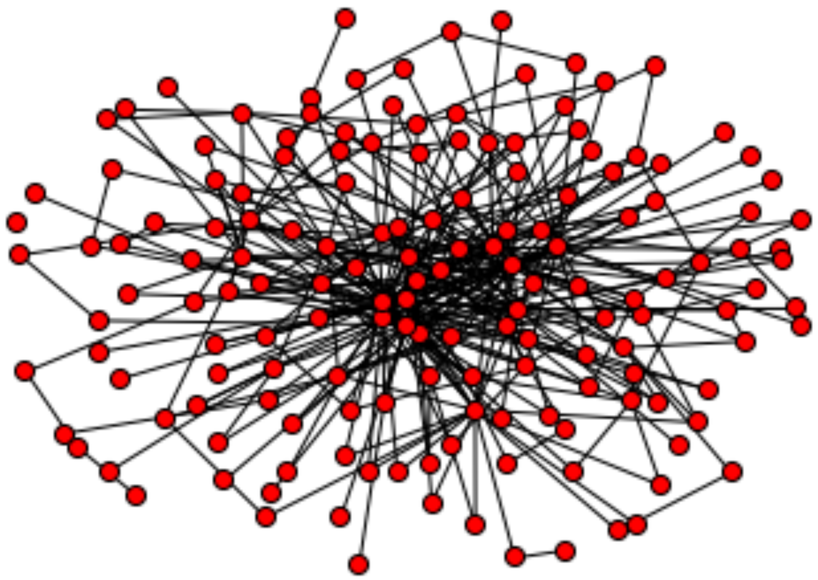
[Kingman 1978; Wallach et al 2010; Broderick, Steorts 2014; Miller et al 2015; Campbell et al 2019]

Roadmap

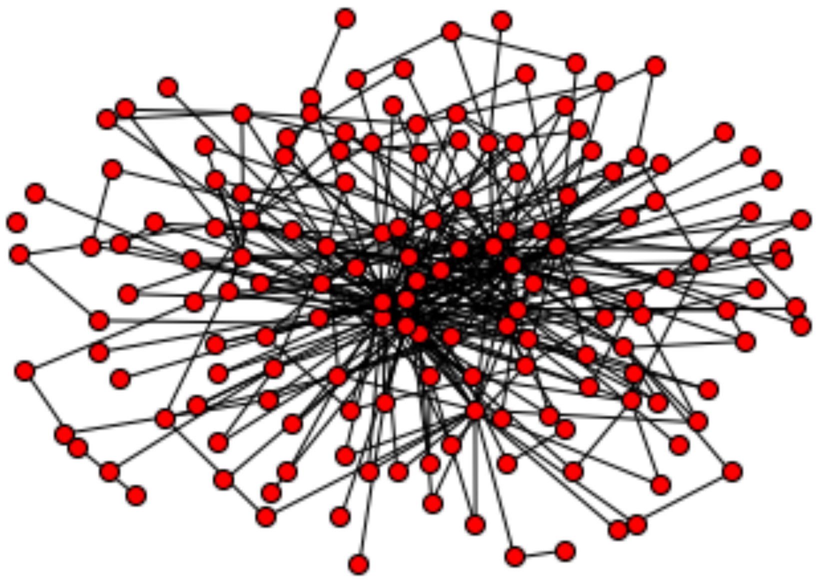
- Example problem: clustering
- Example NPBayes model: Dirichlet process (DP)
- De Finetti for clustering: Kingman Paintbox
- De Finetti for networks/graphs

Roadmap

- Example problem: clustering
- Example NPBayes model: Dirichlet process (DP)
- De Finetti for clustering: Kingman Paintbox
- De Finetti for networks/graphs



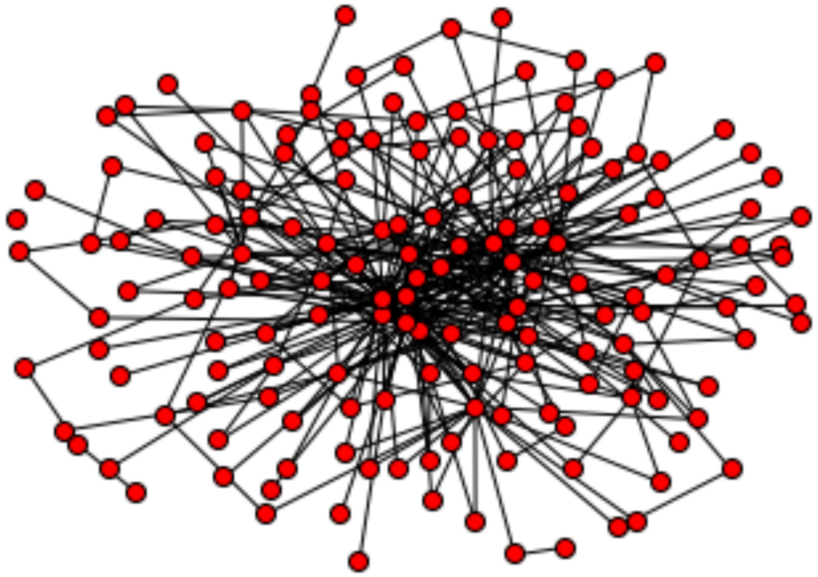
[Holland et al 1983; Kemp et al 2006; Xu et al 2007; Airoldi et al 2008; Lloyd et al 2012]



social: *Facebook, Twitter, email*
biological: *ecological, protein, gene*
transportation: *roads, railways*

[Holland et al 1983; Kemp et al 2006; Xu et al 2007; Airoldi et al 2008; Lloyd et al 2012]

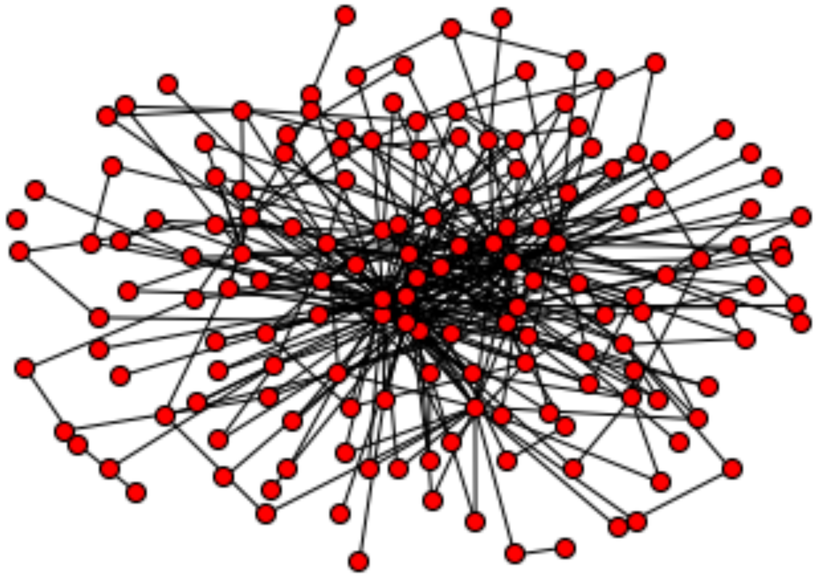
Probabilistic models for graphs

$p($  $)$

social: *Facebook, Twitter, email*
biological: *ecological, protein, gene*
transportation: *roads, railways*

[Holland et al 1983; Kemp et al 2006; Xu et al 2007; Airoldi et al 2008; Lloyd et al 2012]

Probabilistic models for graphs

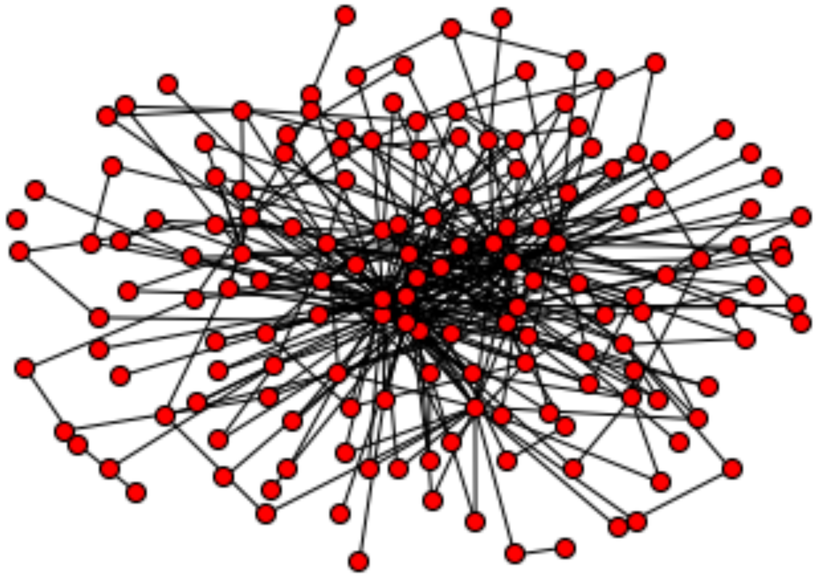
$p($  $)$

social: *Facebook, Twitter, email*
biological: *ecological, protein, gene*
transportation: *roads, railways*

- Interpretable, flexible, coherent uncertainties, expert info

[Holland et al 1983; Kemp et al 2006; Xu et al 2007; Airoldi et al 2008; Lloyd et al 2012]

Probabilistic models for graphs

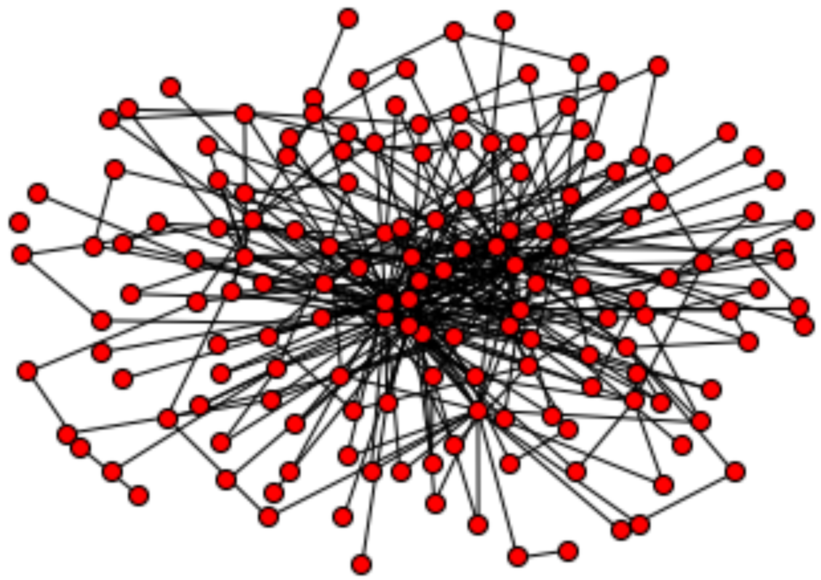
$p($  $)$

social: *Facebook, Twitter, email*
biological: *ecological, protein, gene*
transportation: *roads, railways*

- Interpretable, flexible, coherent uncertainties, expert info
- Example models:

[Holland et al 1983; Kemp et al 2006; Xu et al 2007; Airoldi et al 2008]

Probabilistic models for graphs

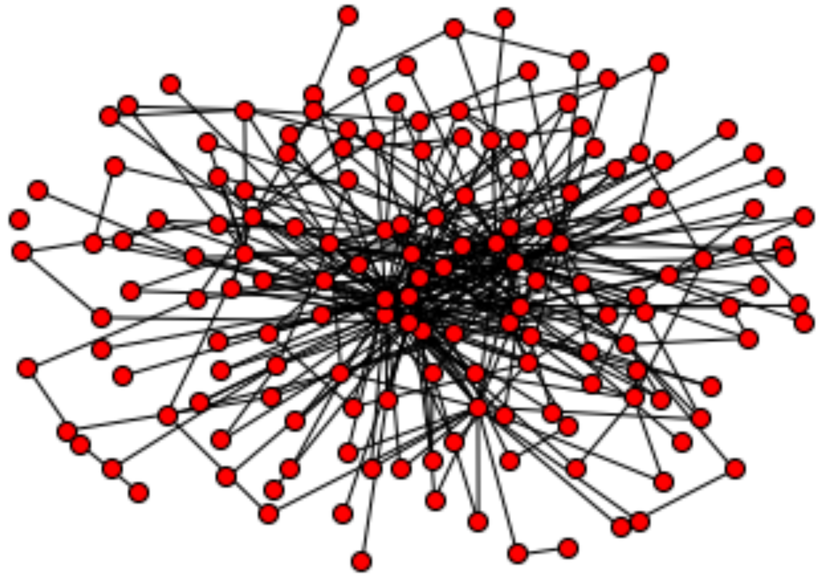
$p($  $)$

social: *Facebook, Twitter, email*
biological: *ecological, protein, gene*
transportation: *roads, railways*

- Interpretable, flexible, coherent uncertainties, expert info
- Example models:
  - Stochastic block model

[Holland et al 1983; Kemp et al 2006; Xu et al 2007; Airoldi et al 2008]

Probabilistic models for graphs

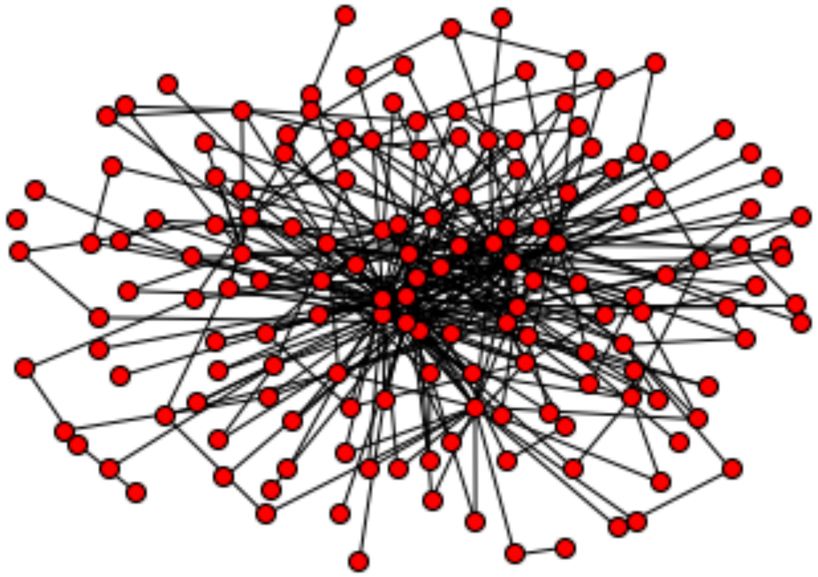
$p($  $)$

social: *Facebook, Twitter, email*
biological: *ecological, protein, gene*
transportation: *roads, railways*

- Interpretable, flexible, coherent uncertainties, expert info
- Example models:
  - Stochastic block model
  - Mixed membership stochastic block model

[Holland et al 1983; Kemp et al 2006; Xu et al 2007; Airoldi et al 2008]

Probabilistic models for graphs

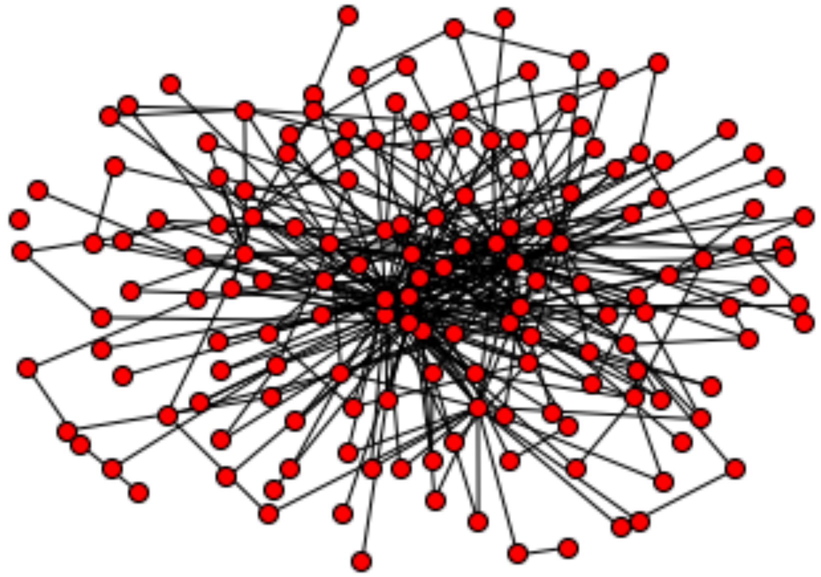
$p\Big($  $\Big)$

social: *Facebook, Twitter, email*
biological: *ecological, protein, gene*
transportation: *roads, railways*

- Interpretable, flexible, coherent uncertainties, expert info
- Example models:
  - Stochastic block model
  - Mixed membership stochastic block model
  - Infinite relational model

[Holland et al 1983; Kemp et al 2006; Xu et al 2007; Airoldi et al 2008]

Probabilistic models for graphs

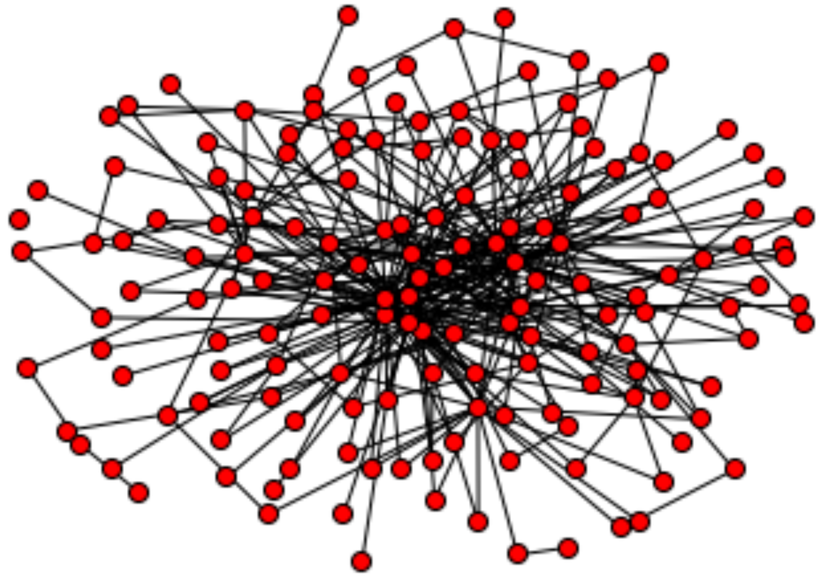
$p($  $)$

social: *Facebook, Twitter, email*
biological: *ecological, protein, gene*
transportation: *roads, railways*

- Interpretable, flexible, coherent uncertainties, expert info
- Example models:
  - Stochastic block model
  - Mixed membership stochastic block model
  - Infinite relational model
  - Latent space model
  - Eigenmodel
  - Latent feature relational model
  - Infinite latent attribute model
  - Sparse matrix-variate Gaussian process block model
  - Random function model

[Holland et al 1983; Kemp et al 2006; Xu et al 2007; Airoldi et al 2008; Lloyd et al 2012]

Probabilistic models for graphs

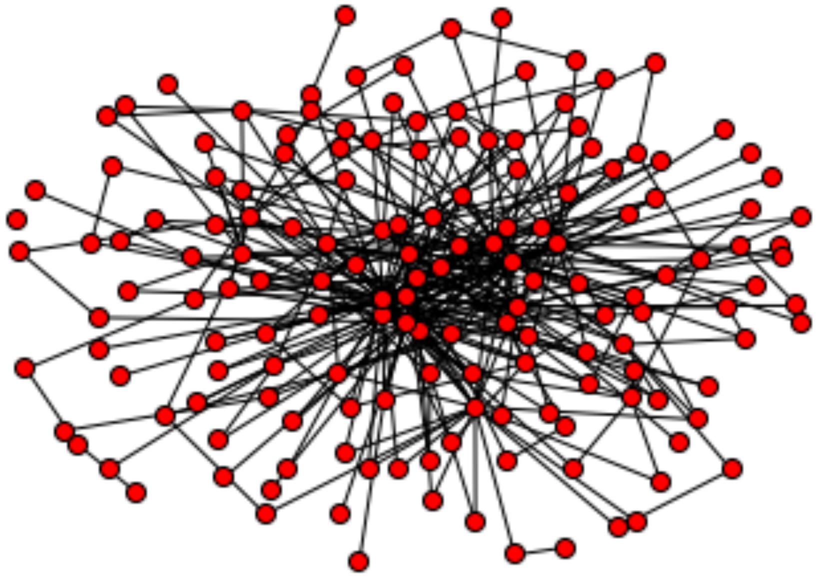
$p($  $)$

social: *Facebook, Twitter, email*
biological: *ecological, protein, gene*
transportation: *roads, railways*

- Interpretable, flexible, coherent uncertainties, expert info
- Example models: Stochastic block, mixed membership stochastic block, infinite relational, and ***many*** more

[Holland et al 1983; Kemp et al 2006; Xu et al 2007; Airoldi et al 2008; Lloyd et al 2012]

Probabilistic models for graphs

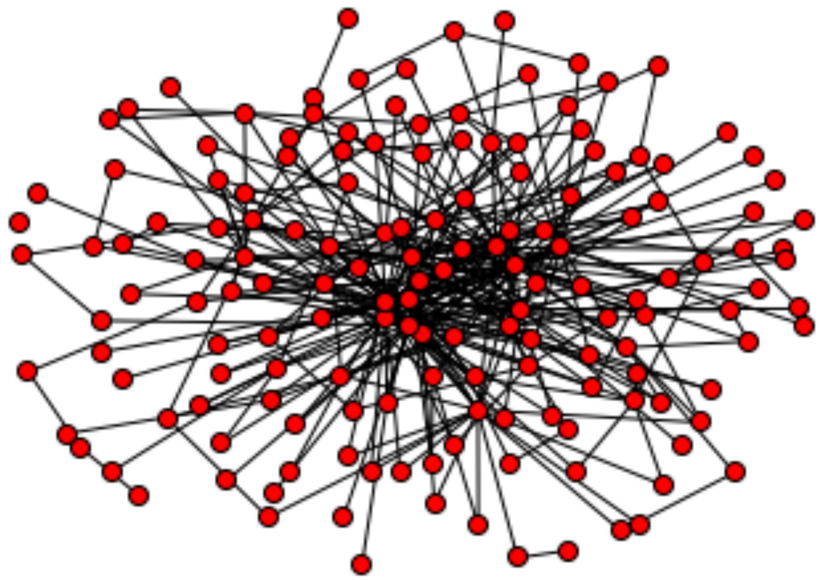
$p($  $)$

social: *Facebook, Twitter, email*
biological: *ecological, protein, gene*
transportation: *roads, railways*

- Interpretable, flexible, coherent uncertainties, expert info
- Example models: Stochastic block, mixed membership stochastic block, infinite relational, and many more
- Assume: Adding more data doesn't change distribution of earlier data (*projectivity*)

[Holland et al 1983; Kemp et al 2006; Xu et al 2007; Airoldi et al 2008; Lloyd et al 2012]

Probabilistic models for graphs

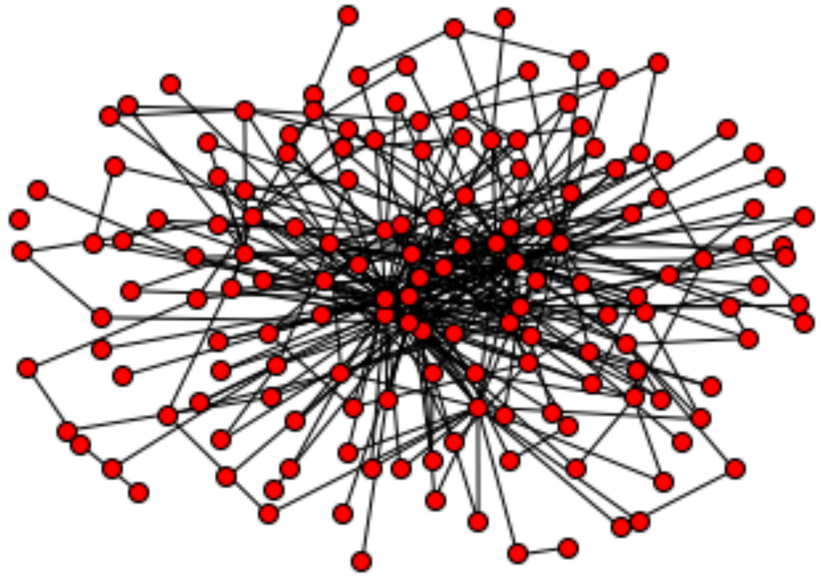
$p($  $)$

social: *Facebook, Twitter, email*
biological: *ecological, protein, gene*
transportation: *roads, railways*

- Interpretable, flexible, coherent uncertainties, expert info
- Example models: Stochastic block, mixed membership stochastic block, infinite relational, and many more
- Assume: Adding more data doesn't change distribution of earlier data (*projectivity*)
- **Problem**: model misspecification, dense graphs

[Holland et al 1983; Kemp et al 2006; Xu et al 2007; Airoldi et al 2008; Lloyd et al 2012]

Probabilistic models for graphs

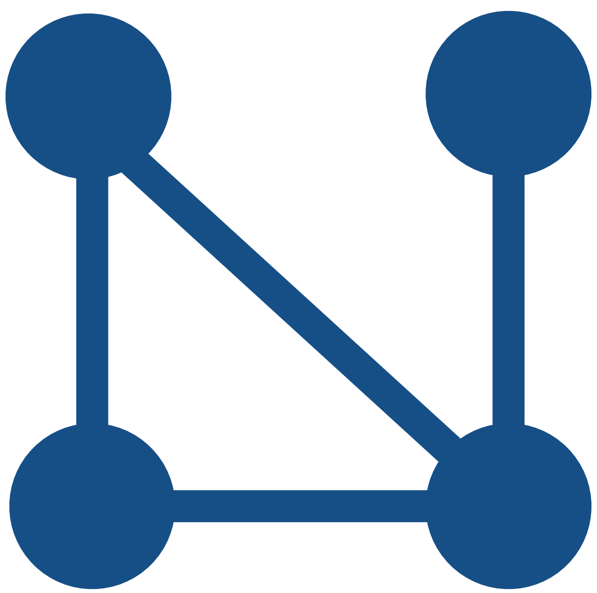
$p($  $)$

social: *Facebook, Twitter, email*
biological: *ecological, protein, gene*
transportation: *roads, railways*

- Interpretable, flexible, coherent uncertainties, expert info
- Example models: Stochastic block, mixed membership stochastic block, infinite relational, and many more
- Assume: Adding more data doesn't change distribution of earlier data (*projectivity*)
- **Problem**: model misspecification, dense graphs
- Some **solution** directions: frameworks for sparse graphs

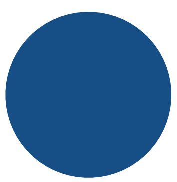
[Holland et al 1983; Kemp et al 2006; Xu et al 2007; Airoldi et al 2008; Lloyd et al 2012]

Sequence of graphs

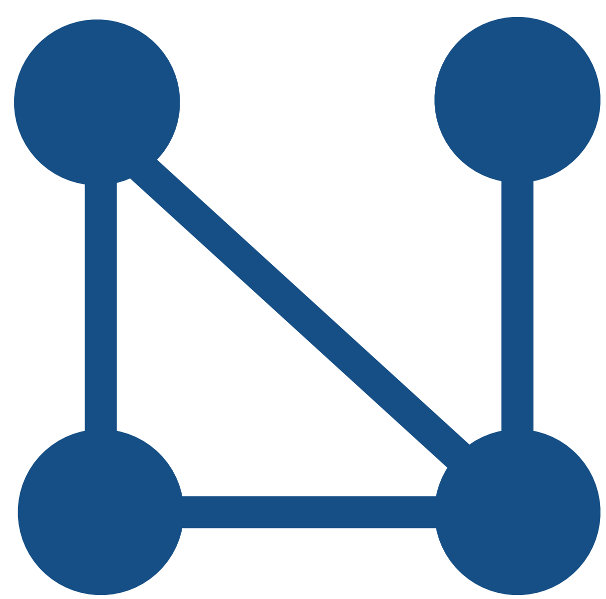


G

Sequence of graphs

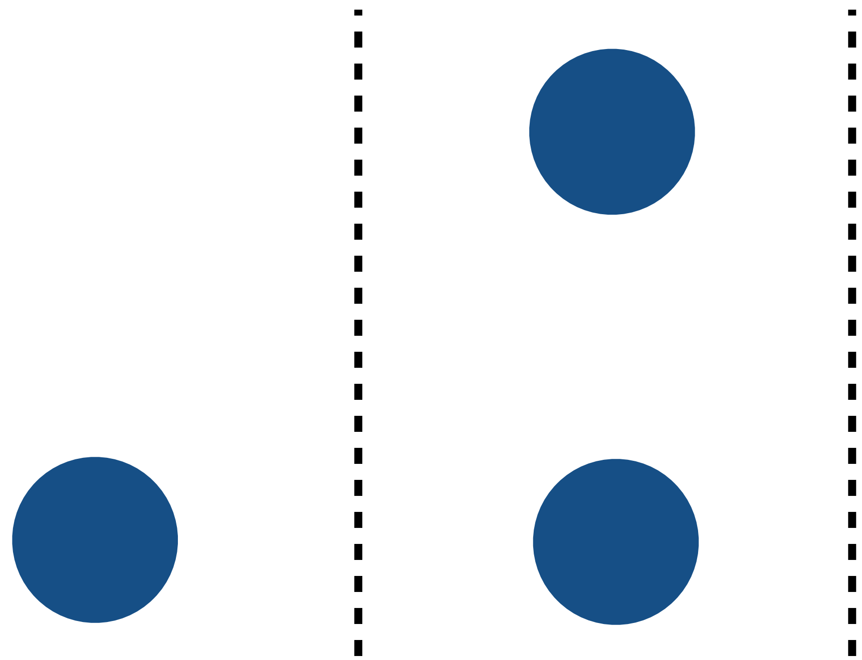


G_1

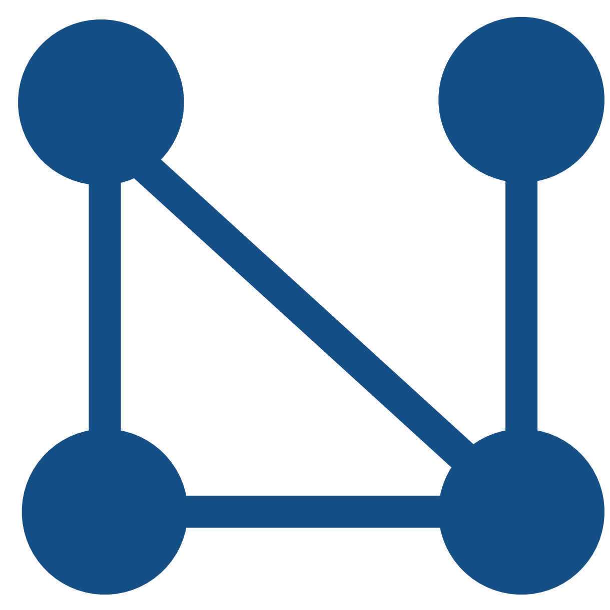


G

Sequence of graphs

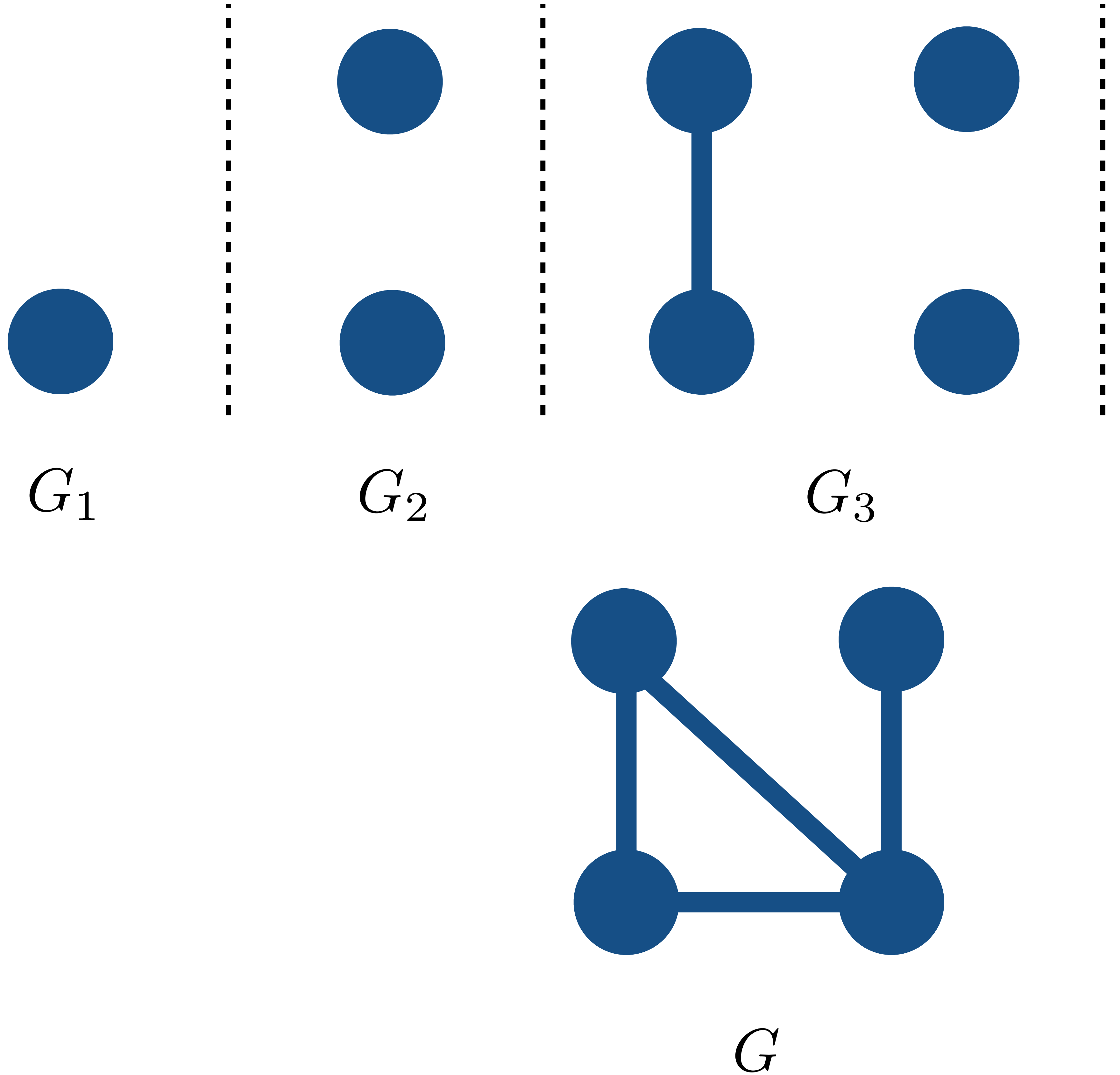


G_1 G_2

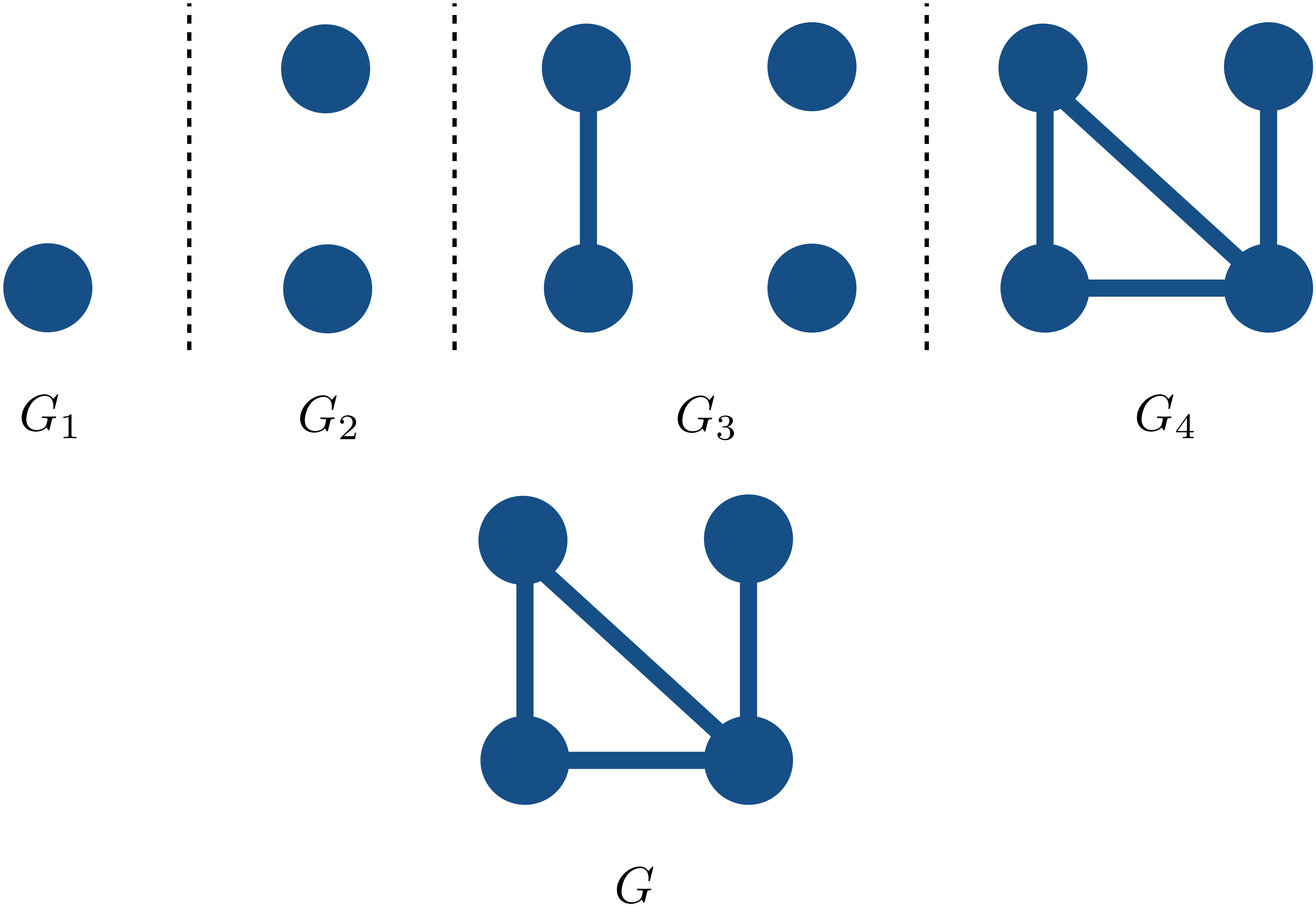


G

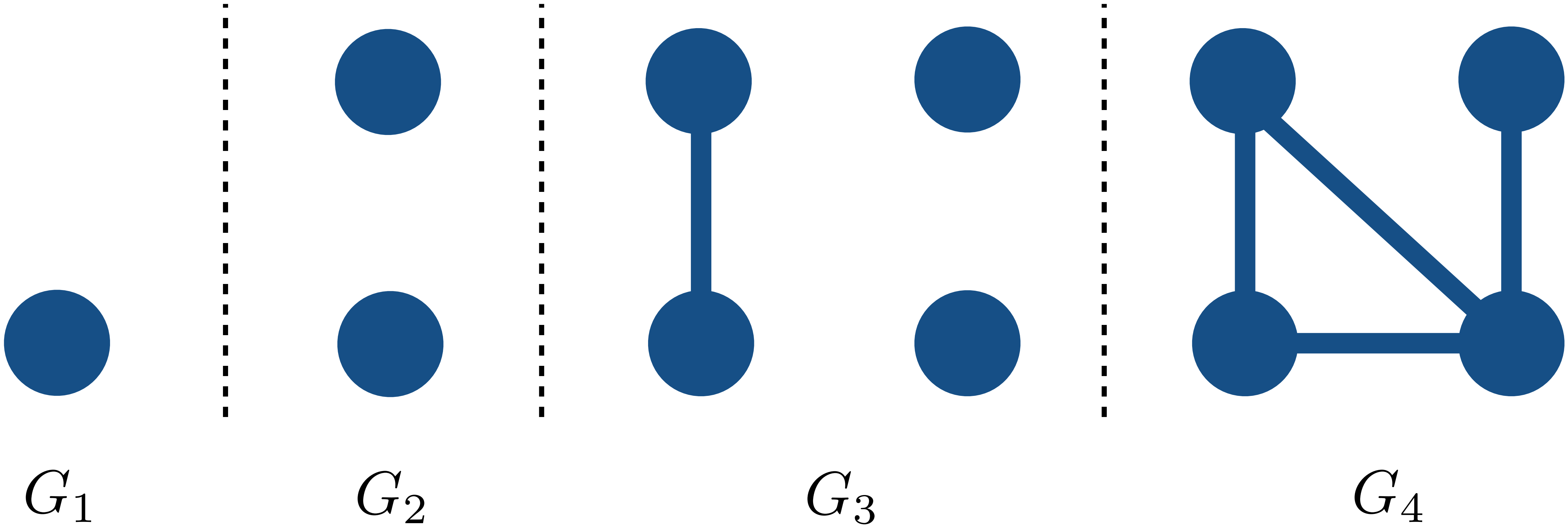
Sequence of graphs



Sequence of graphs

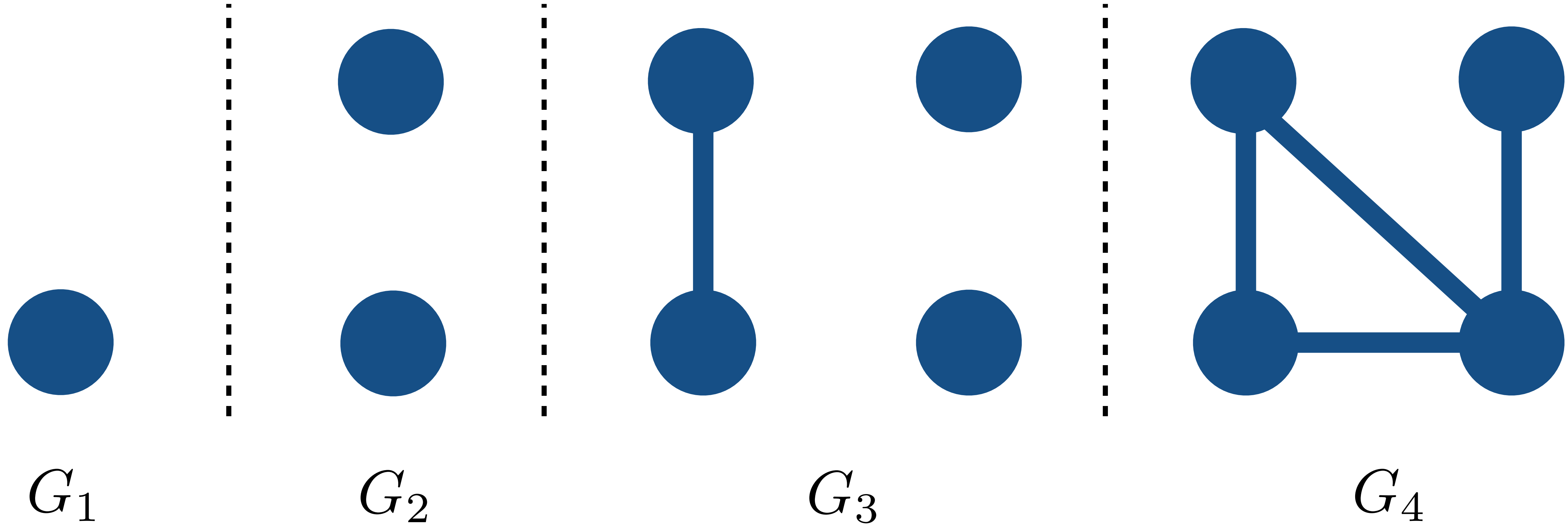


Sequence of graphs



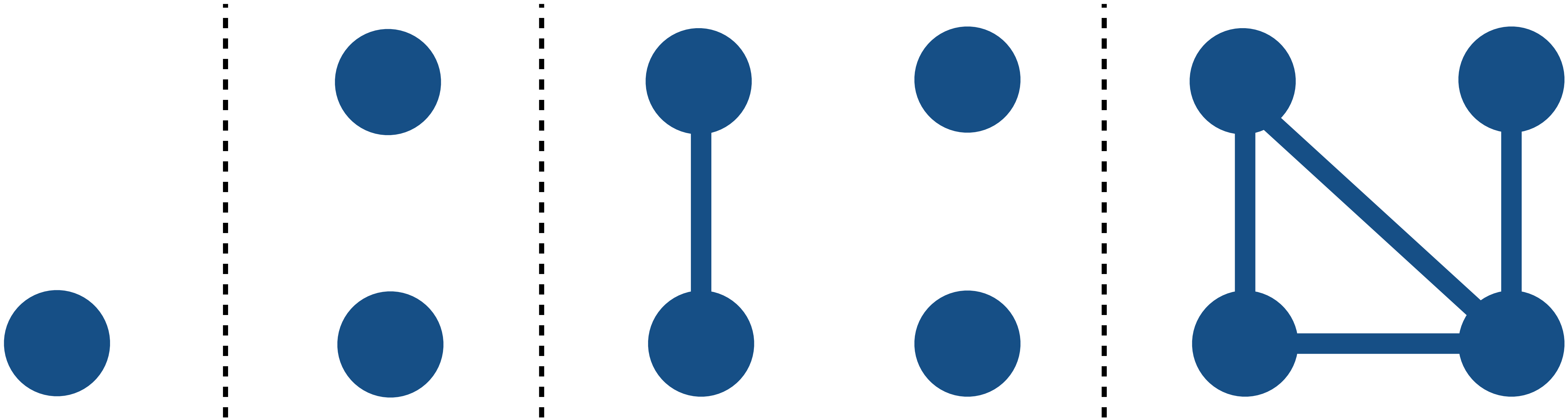
...

Sequence of graphs



If $\#\text{nodes}(G_n) \to \infty$,

Sequence of graphs

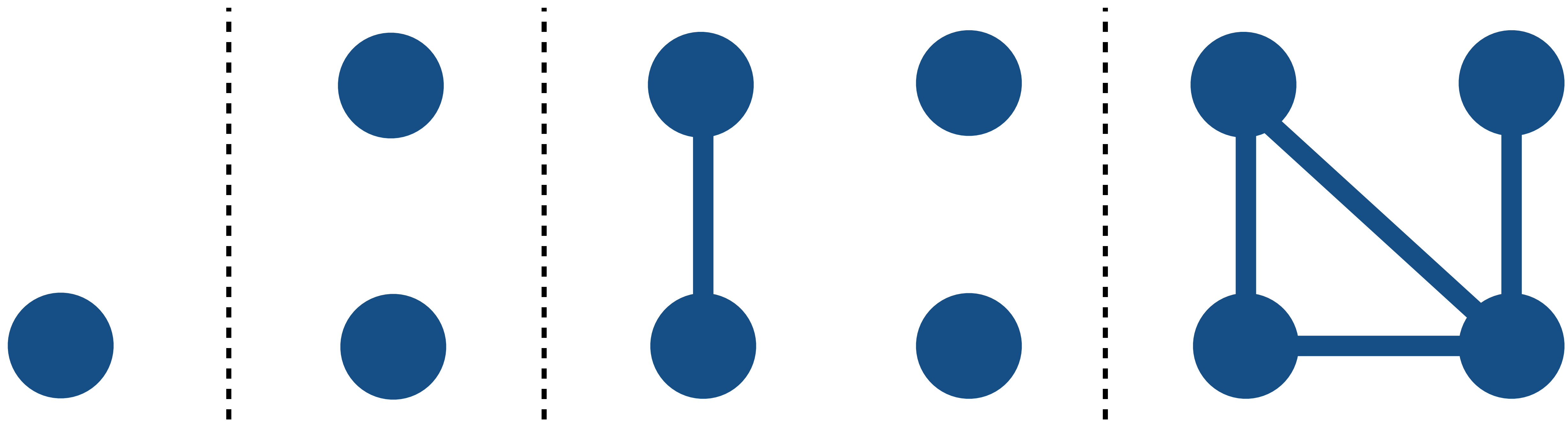


G_1 G_2 G_3 G_4

If $\#\text{nodes}(G_n) \to \infty$,

- *Dense* graph sequence $\#\text{edges}(G_n) \geq c \cdot [\#\text{nodes}(G_n)]^2$

Sequence of graphs

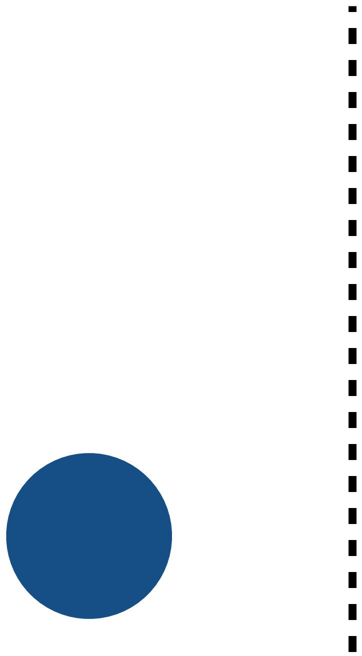


G_1 G_2 G_3 G_4

If $\#\text{nodes}(G_n) \to \infty$,

- *Dense* graph sequence $\#\text{edges}(G_n) \geq c \cdot [\#\text{nodes}(G_n)]^2$
- *Sparse* graph sequence $\#\text{edges}(G_n) \in o([\#\text{nodes}(G_n)]^2)$

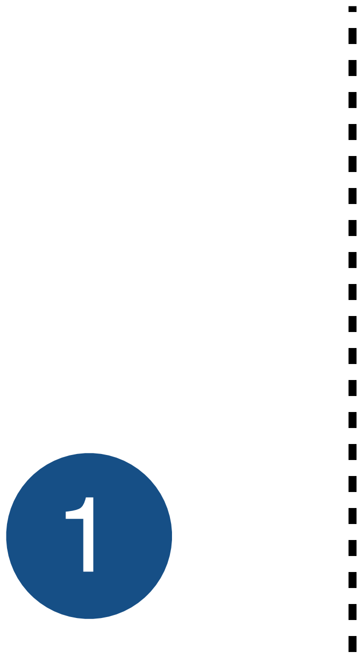
The Old Way: Nodes



G_1

[Hoover 1979, Aldous 1981]

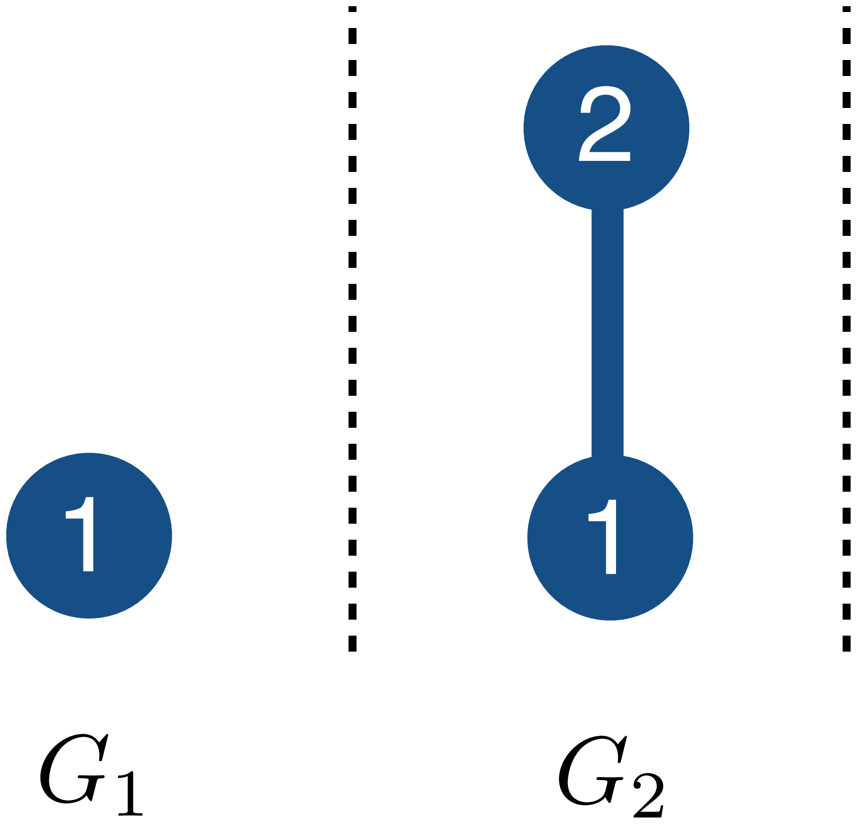
The Old Way: Nodes



G_1

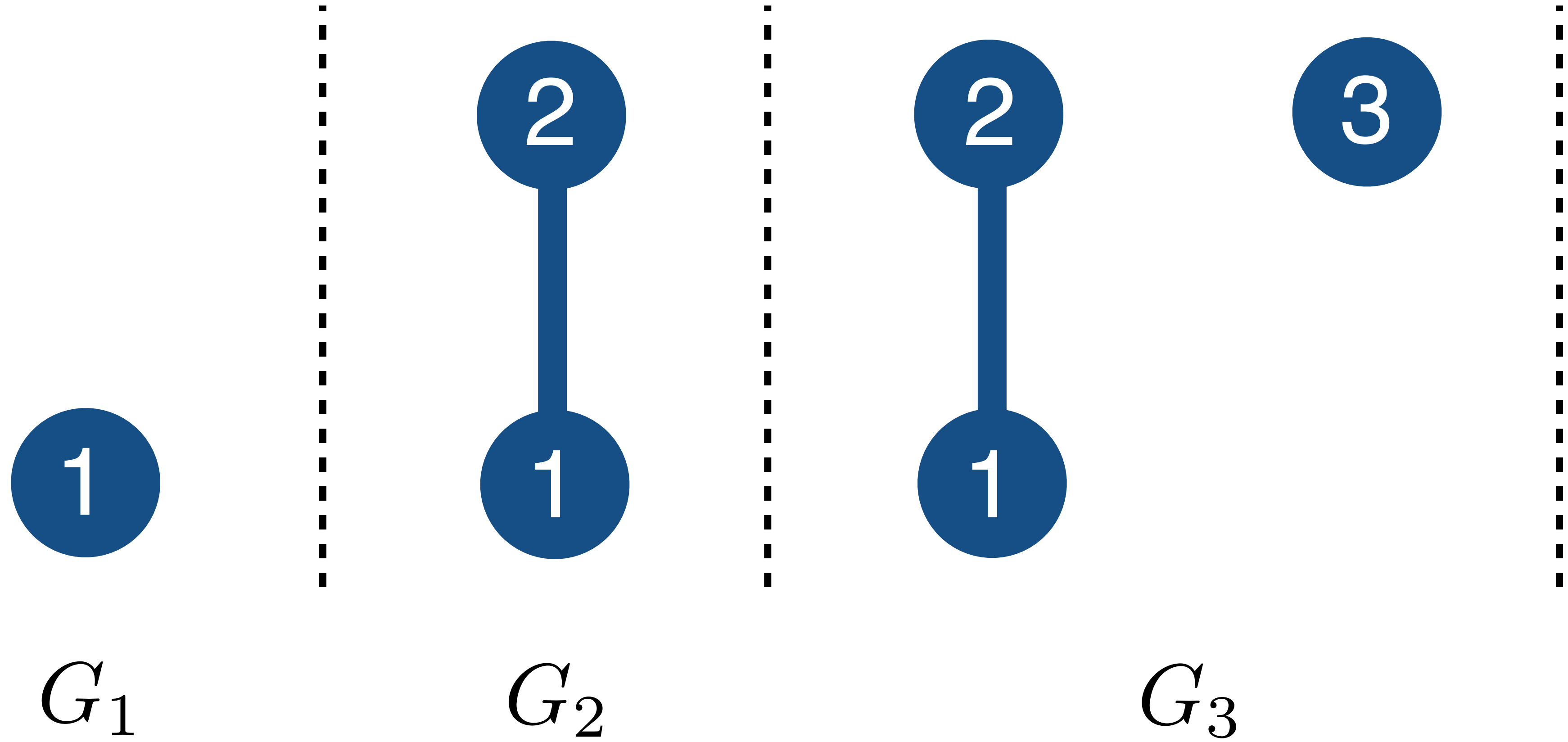
[Hoover 1979, Aldous 1981]

The Old Way: Nodes



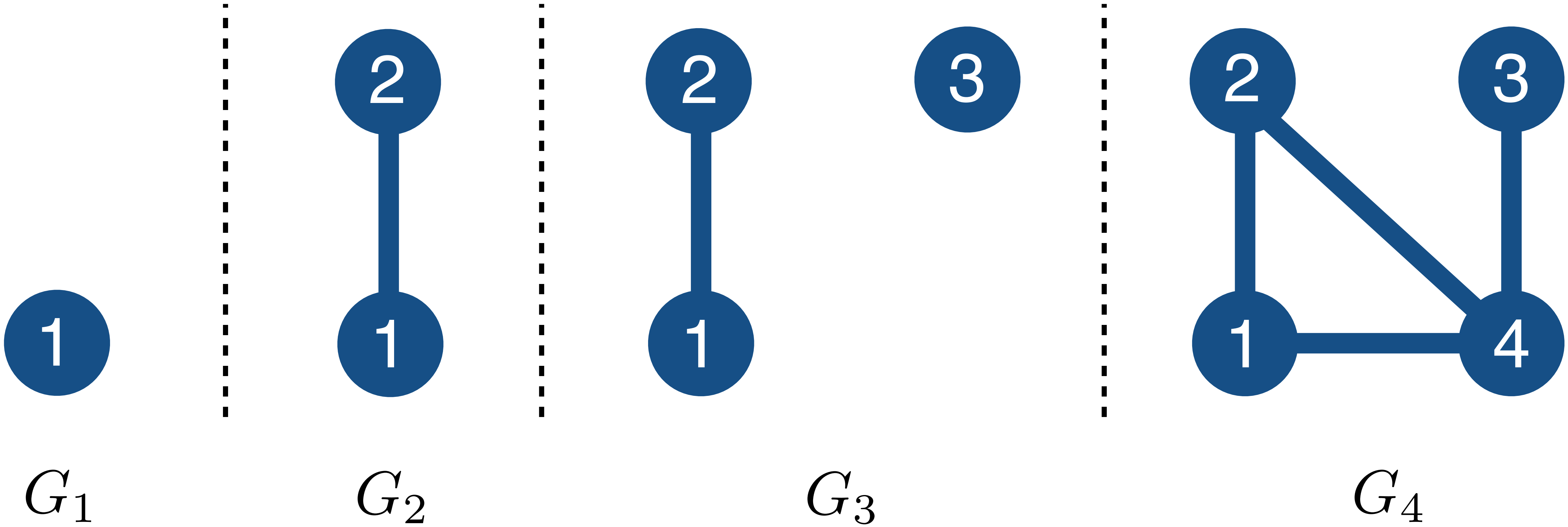
[Hoover 1979, Aldous 1981]

The Old Way: Nodes



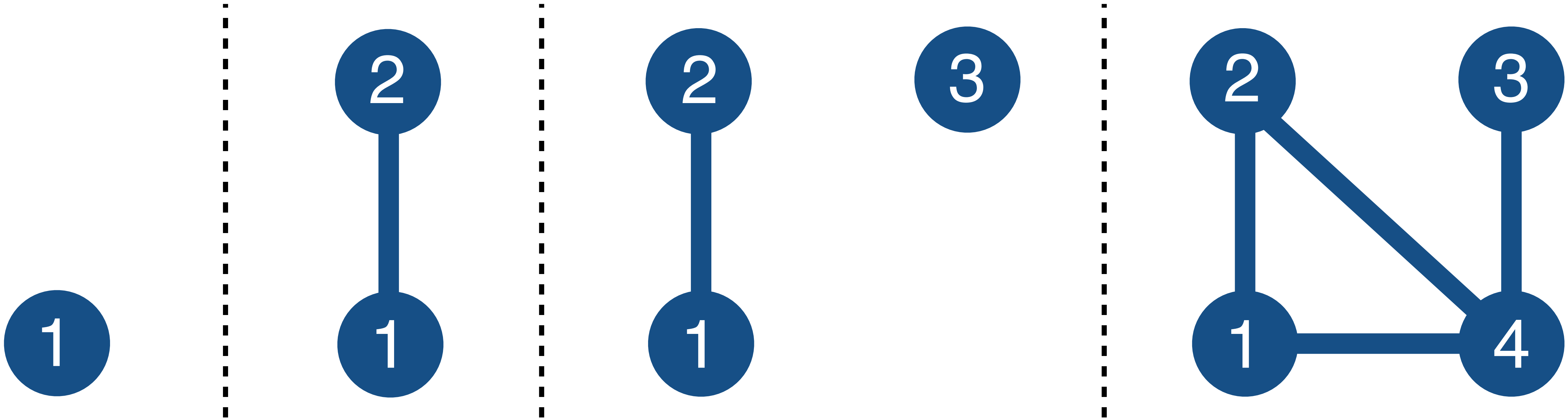
[Hoover 1979, Aldous 1981]

The Old Way: Nodes



[Hoover 1979, Aldous 1981]

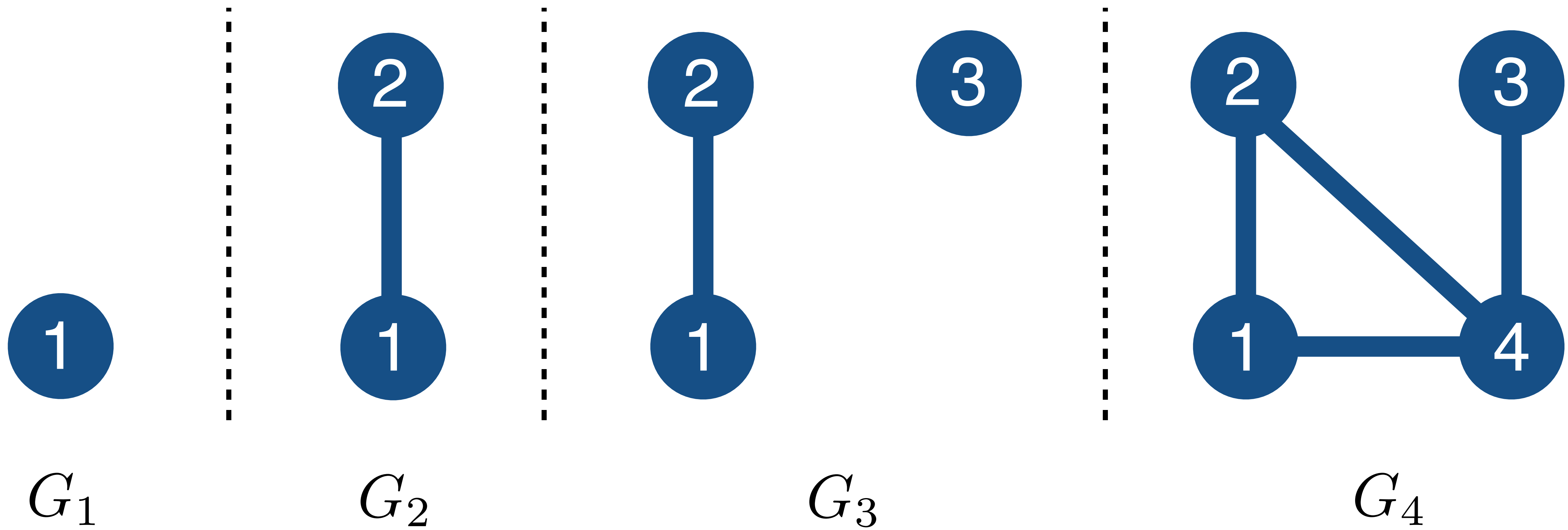
The Old Way: Exchangeability

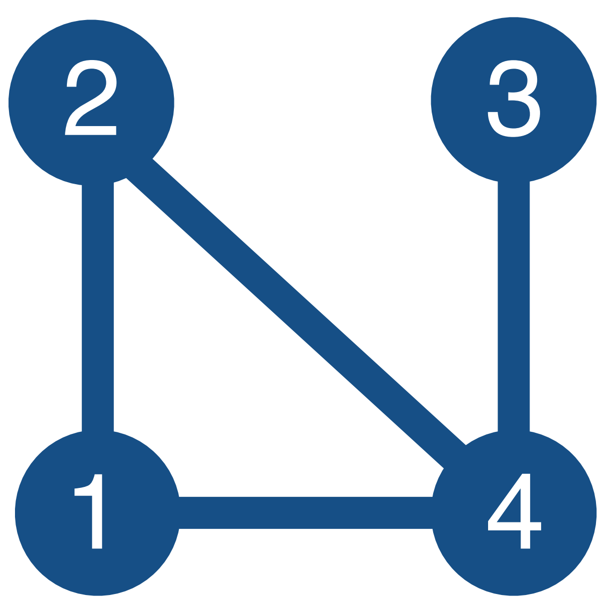


G_1 G_2 G_3 G_4

[Hoover 1979, Aldous 1981]

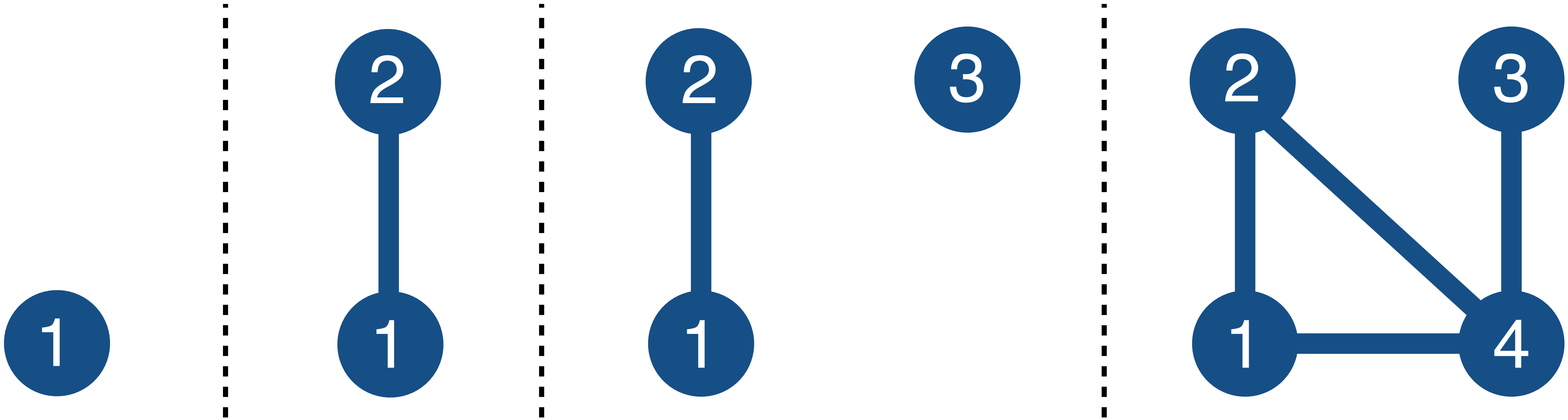
The Old Way: Exchangeability



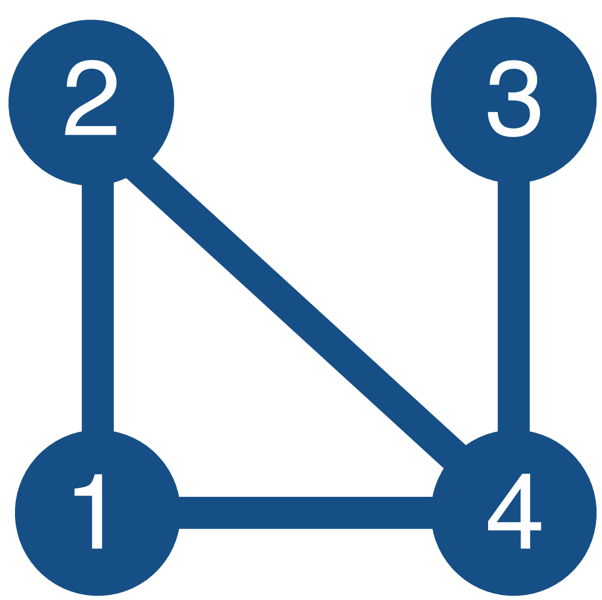


[Hoover 1979, Aldous 1981]

The Old Way: Exchangeability

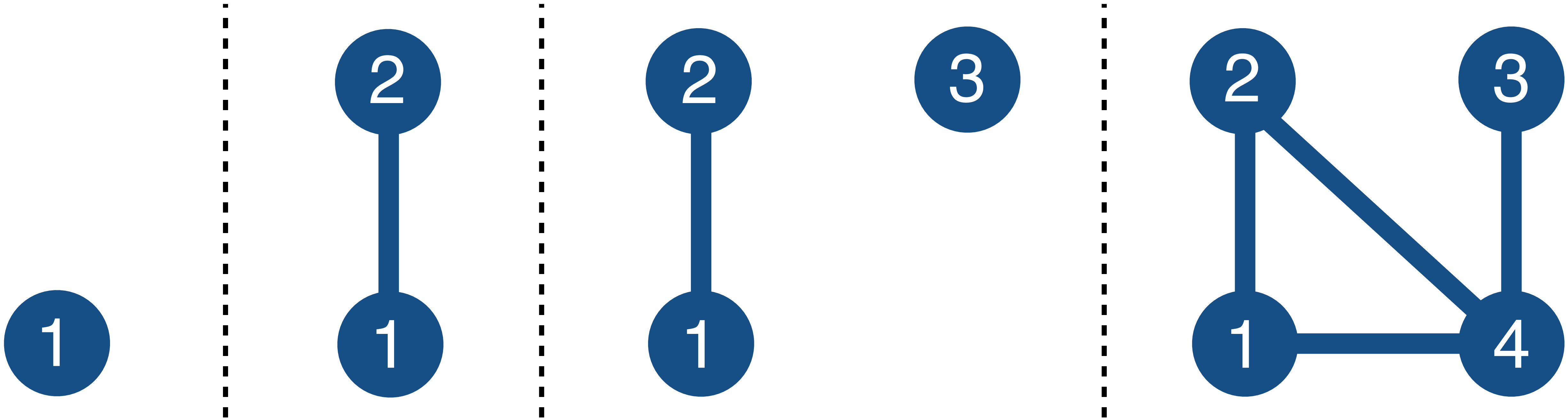


G_1 G_2 G_3 G_4

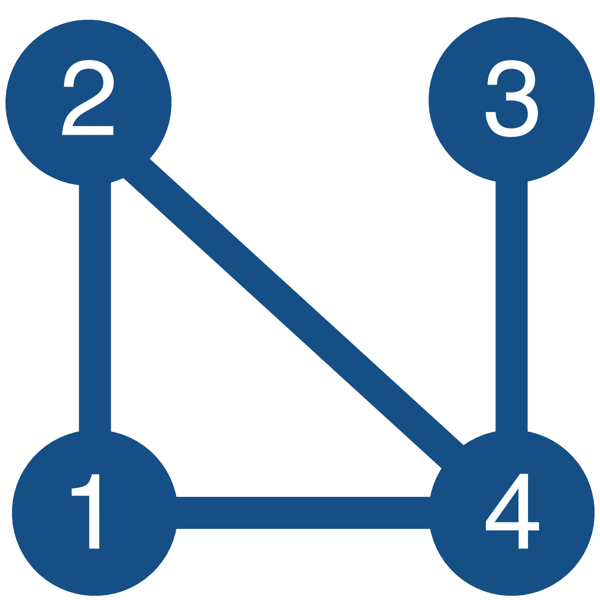
$p($  $)$

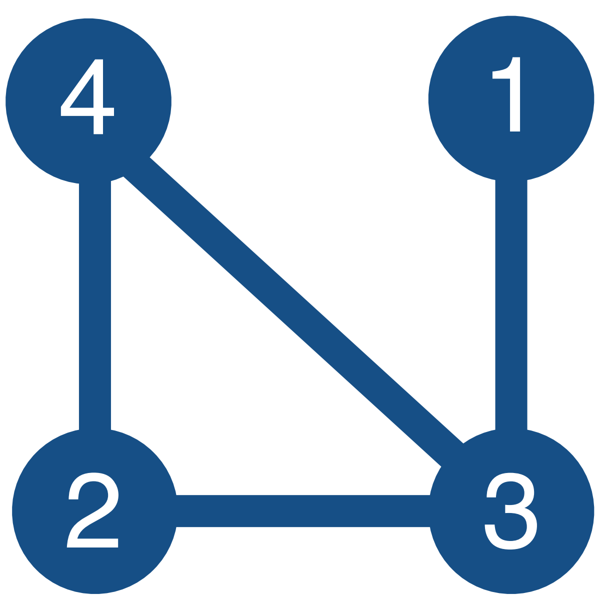
[Hoover 1979, Aldous 1981]

The Old Way: Exchangeability



G_1 G_2 G_3 G_4

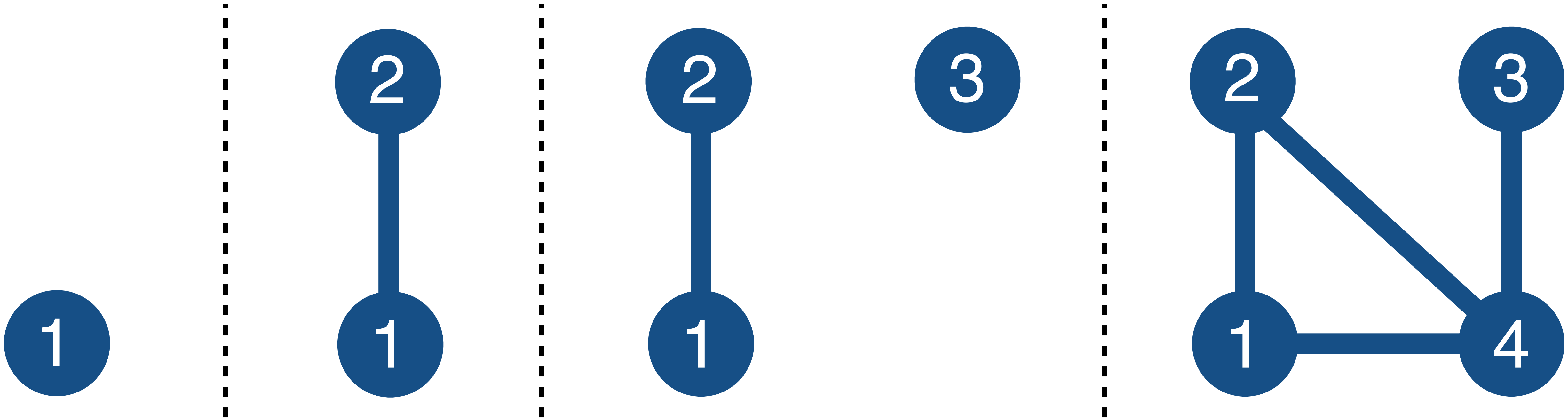
$p($ 

$) = p($ 

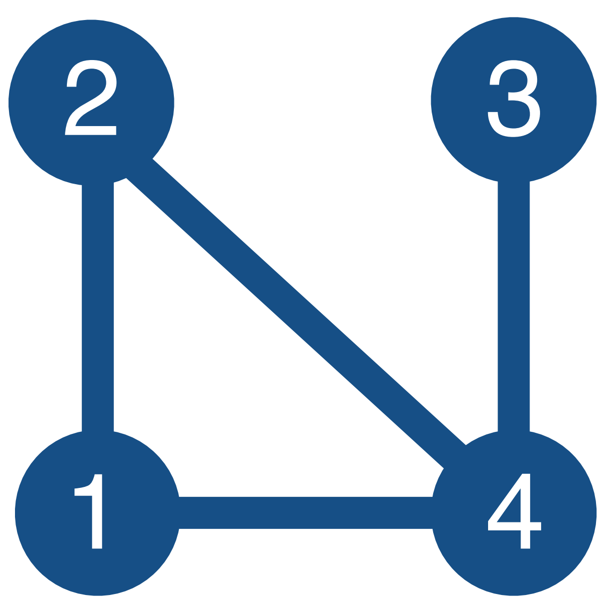
$)$

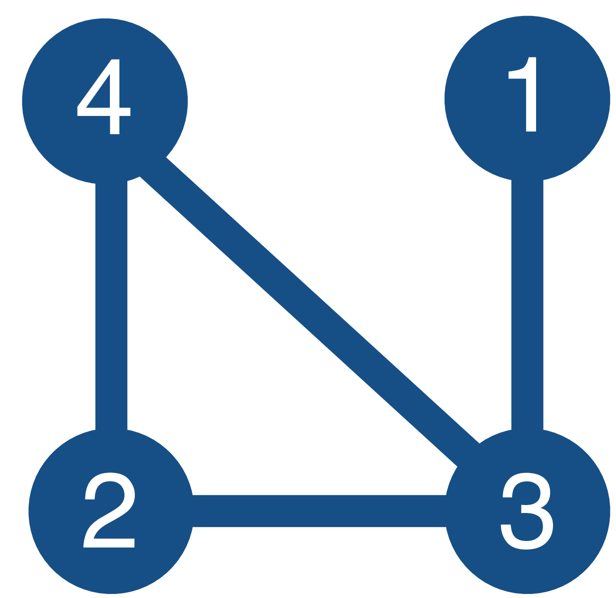
[Hoover 1979, Aldous 1981]

The Old Way: Node exchangeability



G_1 G_2 G_3 G_4

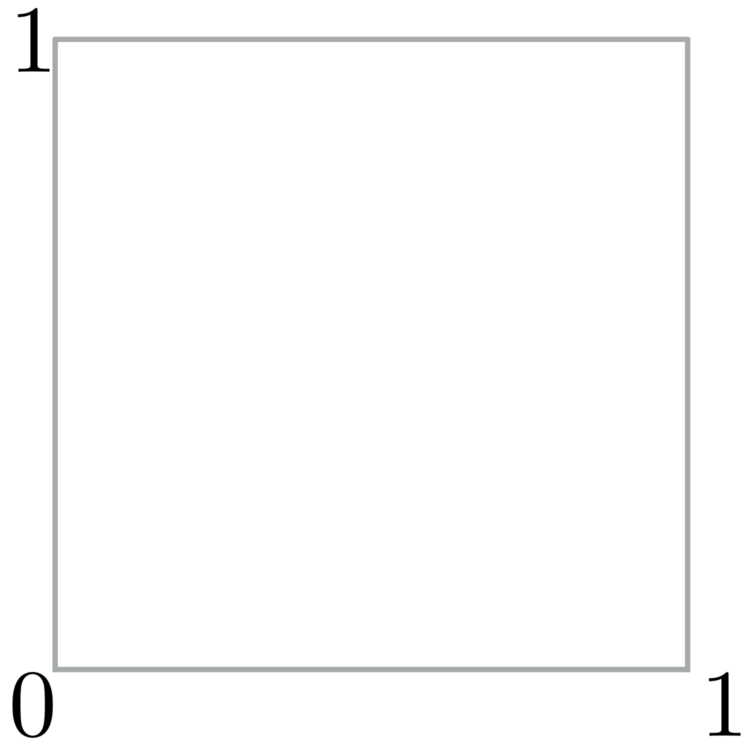
$p($ 

$) = p($ 

$)$

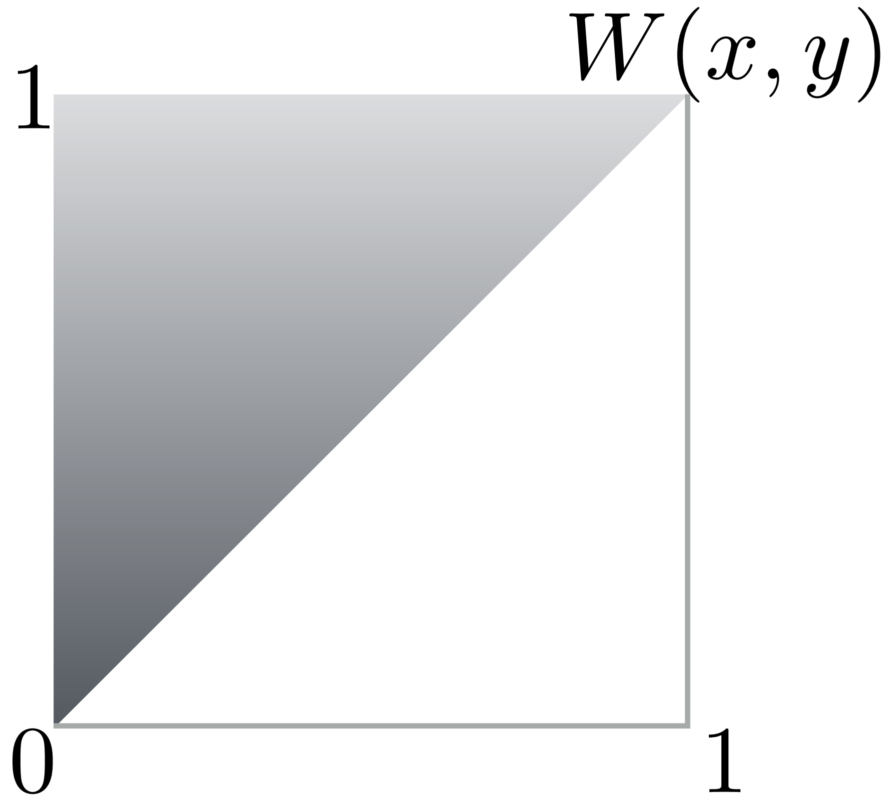
[Hoover 1979, Aldous 1981]

Aldous-Hoover



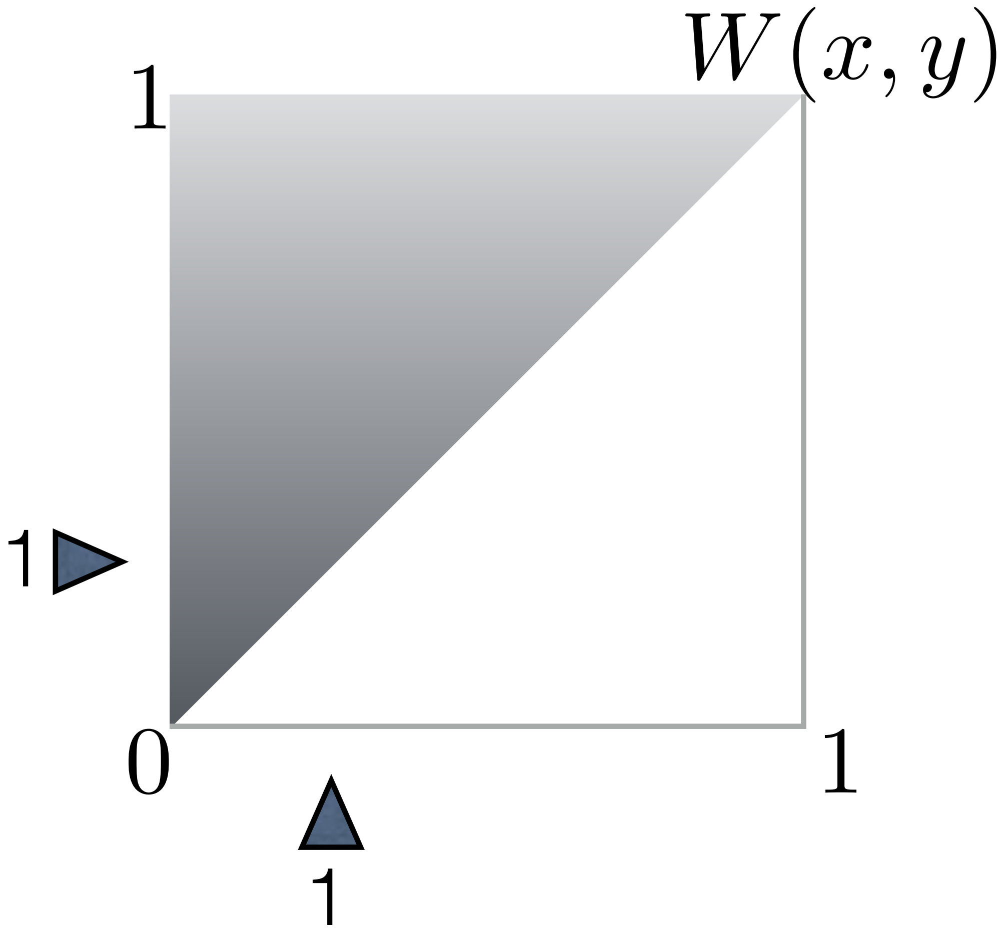
[Hoover 1979, Aldous 1981]

Aldous-Hoover



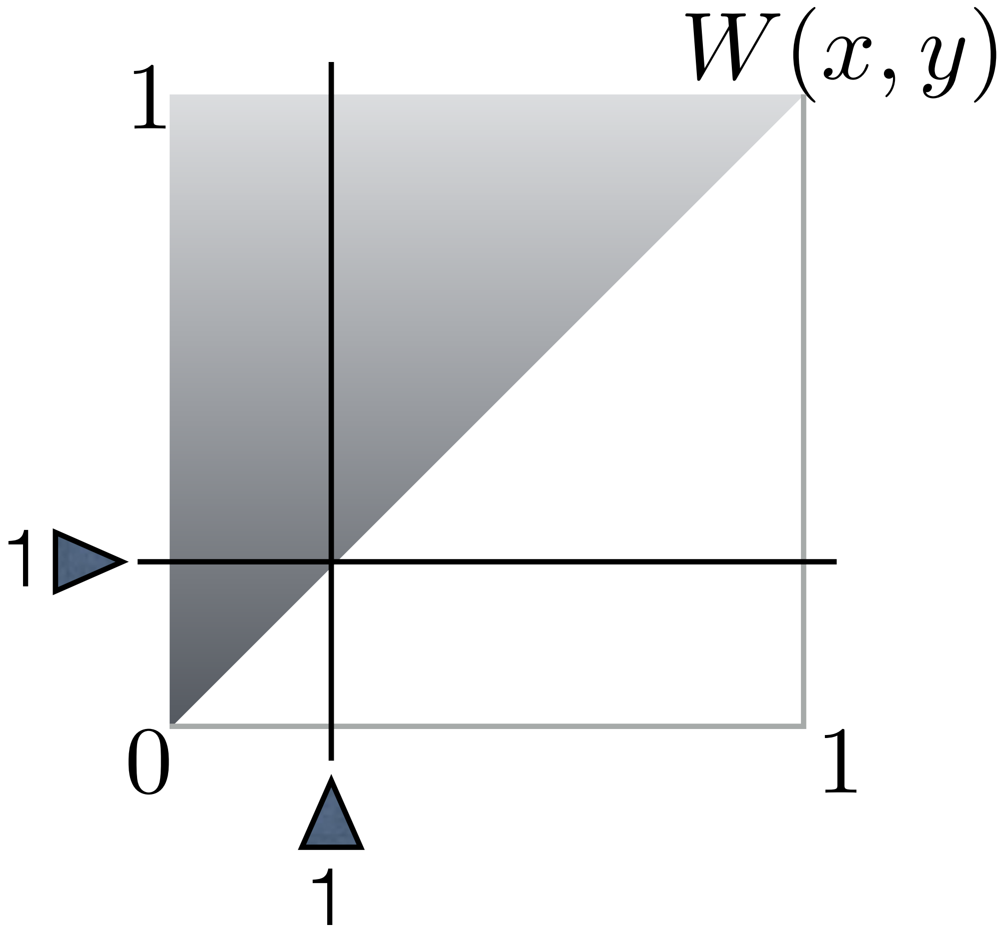
[Hoover 1979, Aldous 1981]

Aldous-Hoover



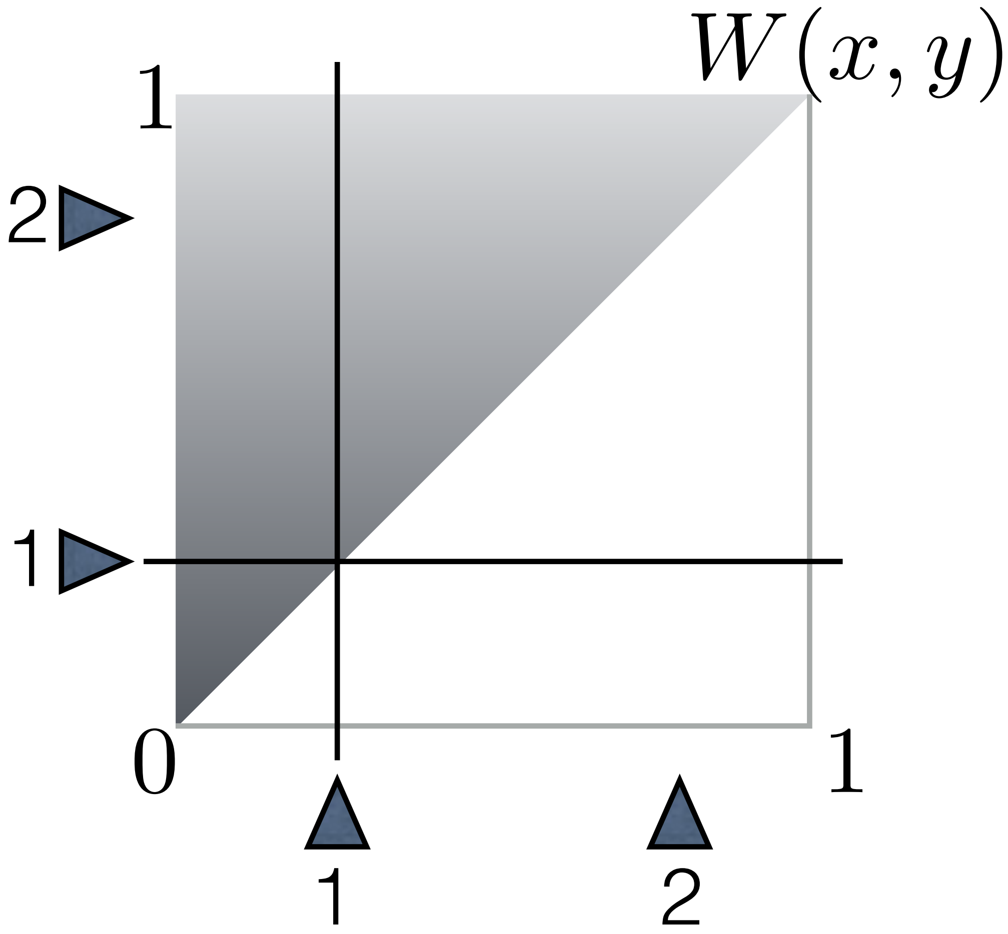
[Hoover 1979, Aldous 1981]

Aldous-Hoover



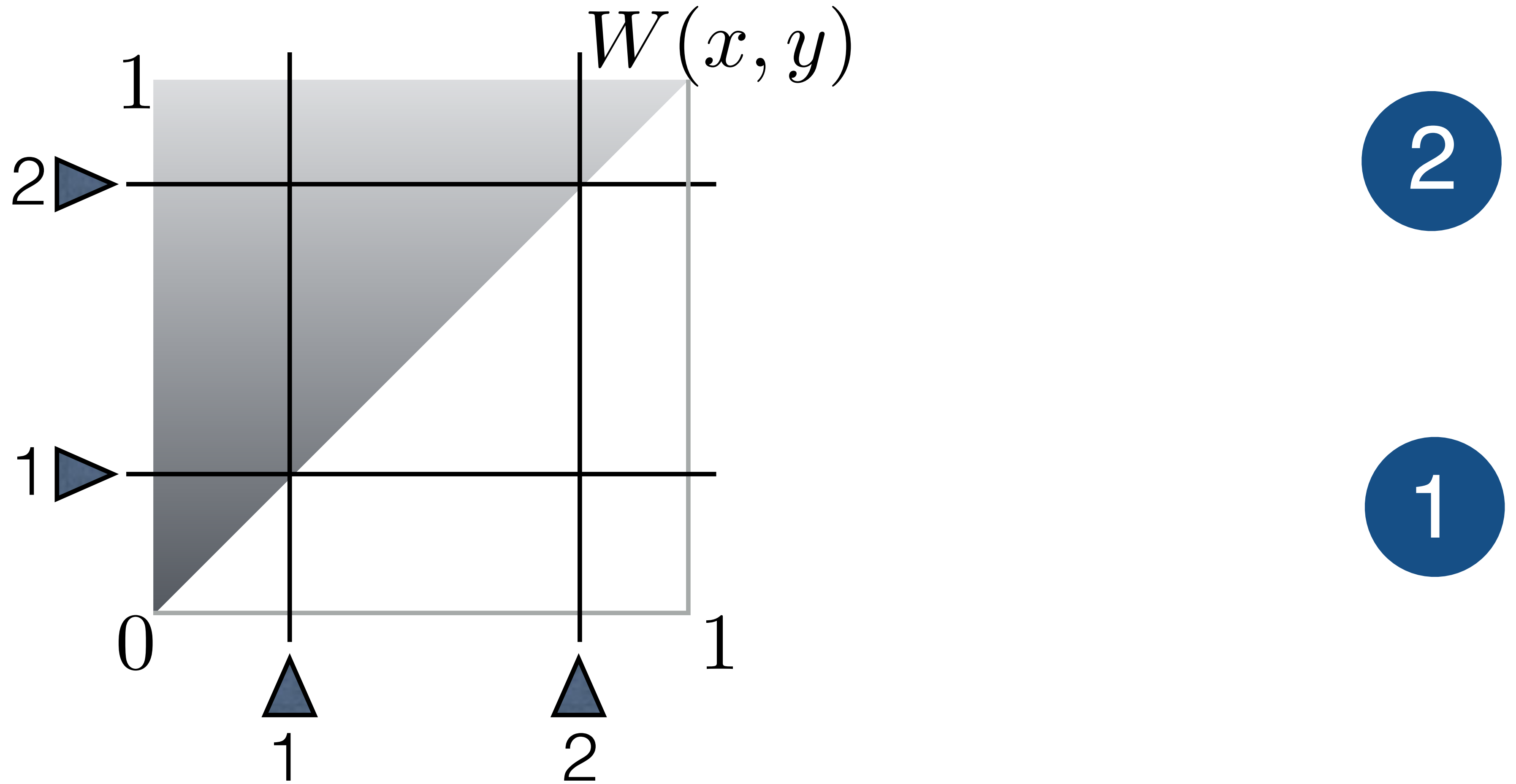
[Hoover 1979, Aldous 1981]

Aldous-Hoover



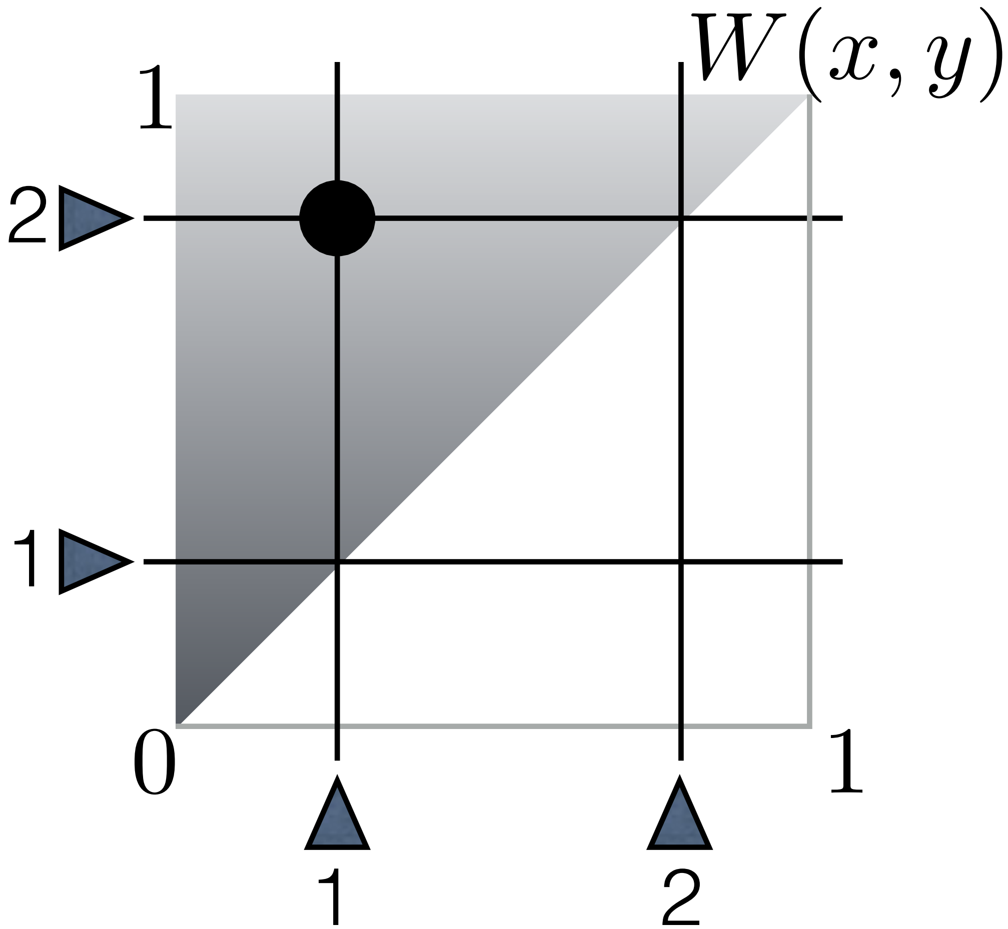
[Hoover 1979, Aldous 1981]

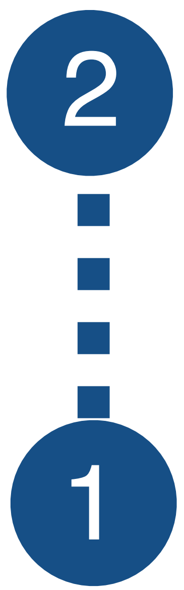
Aldous-Hoover



[Hoover 1979, Aldous 1981]

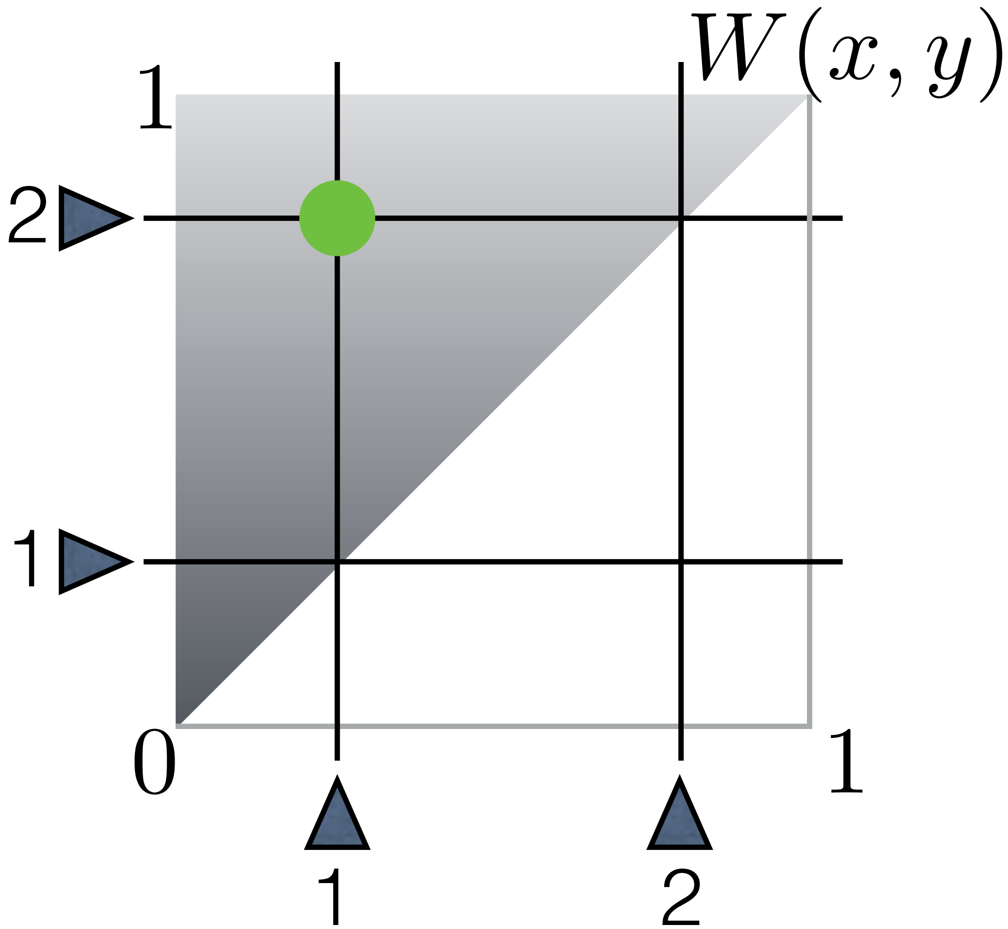
Aldous-Hoover

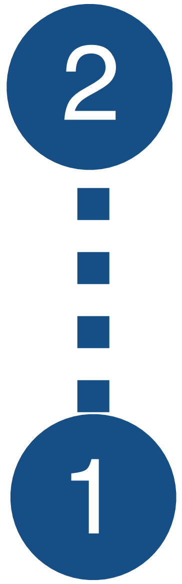




[Hoover 1979, Aldous 1981]

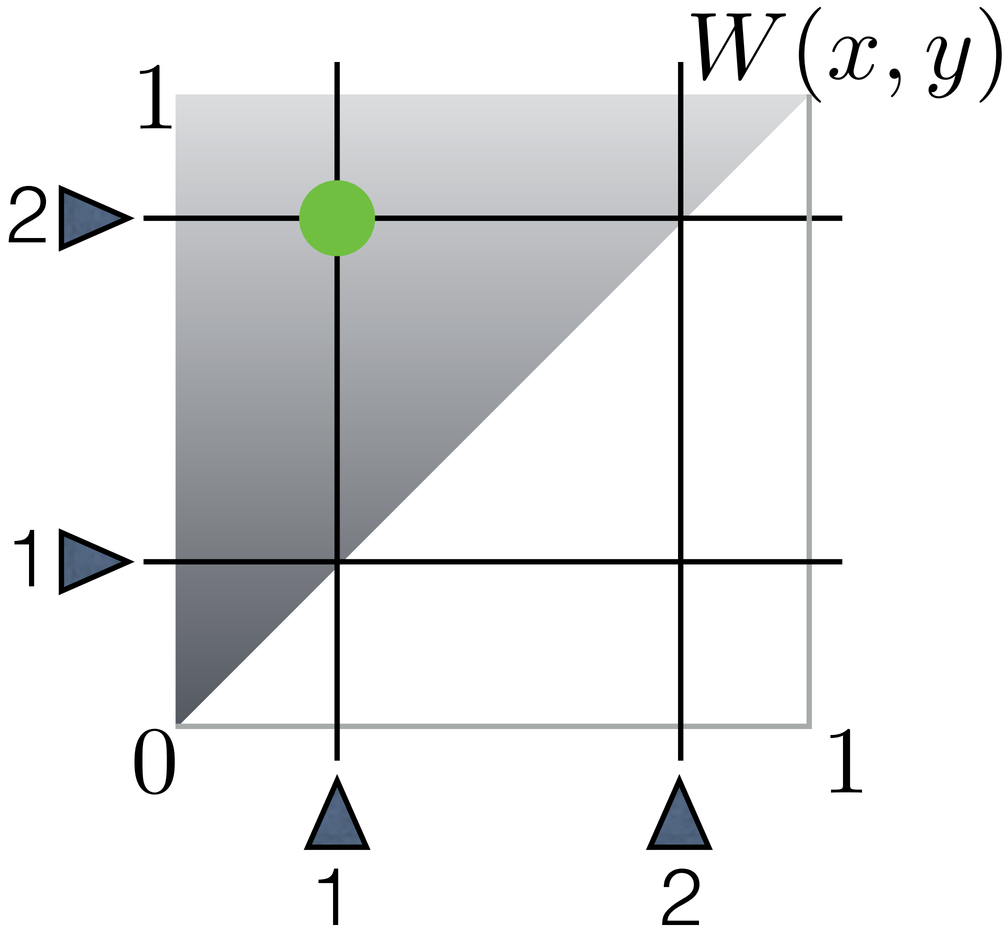
Aldous-Hoover





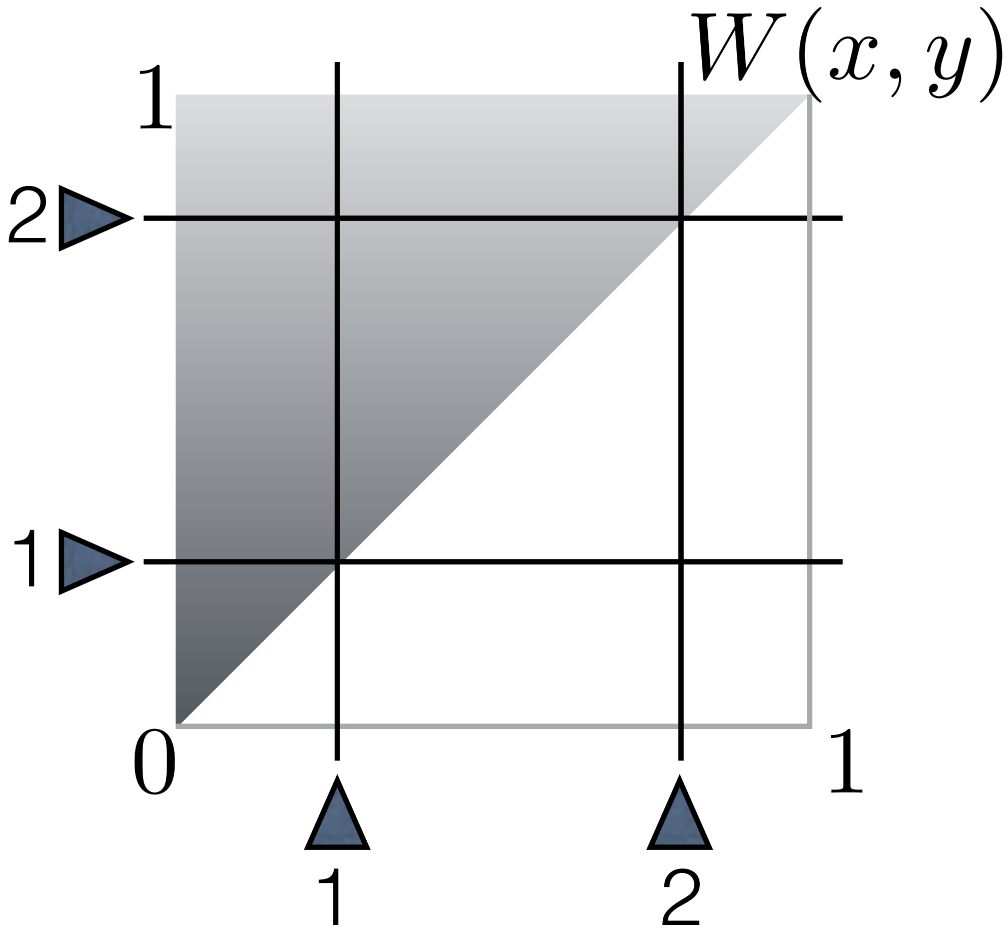
[Hoover 1979, Aldous 1981]

Aldous-Hoover



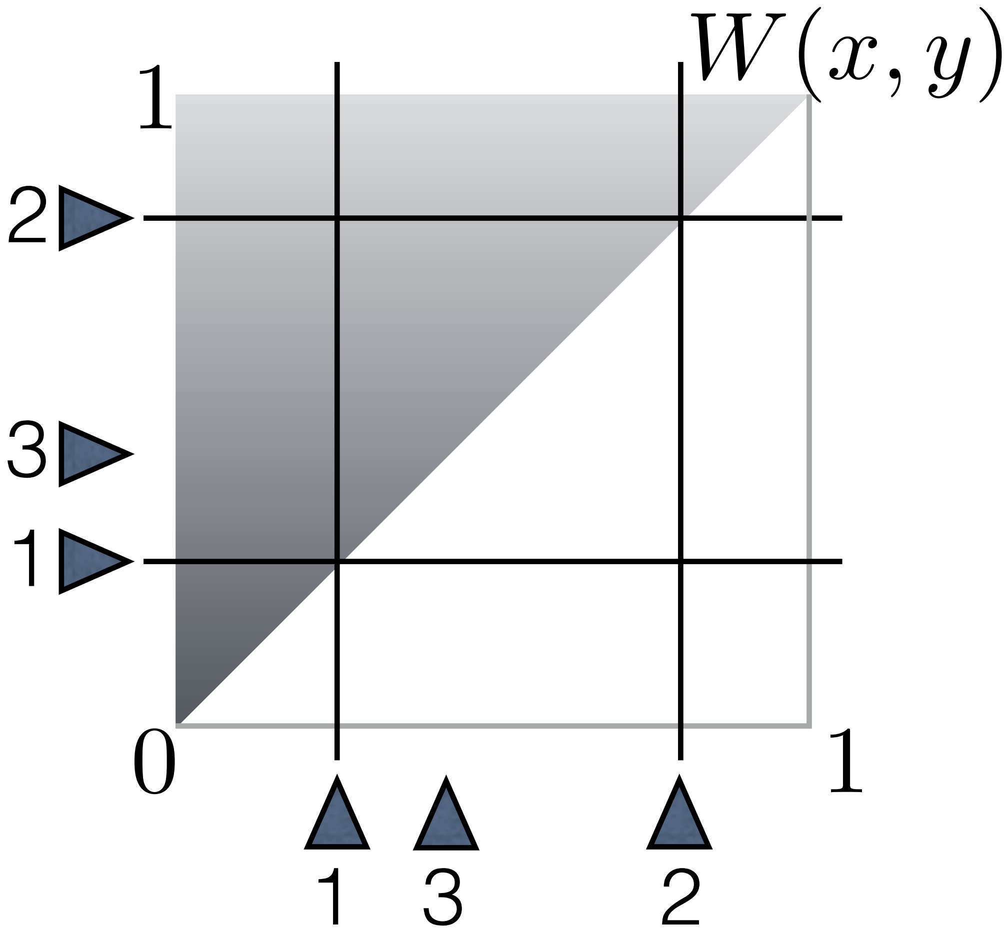
[Hoover 1979, Aldous 1981]

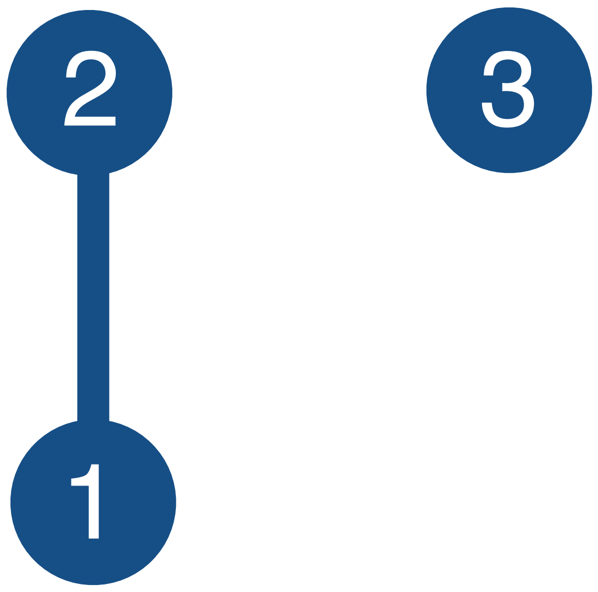
Aldous-Hoover



[Hoover 1979, Aldous 1981]

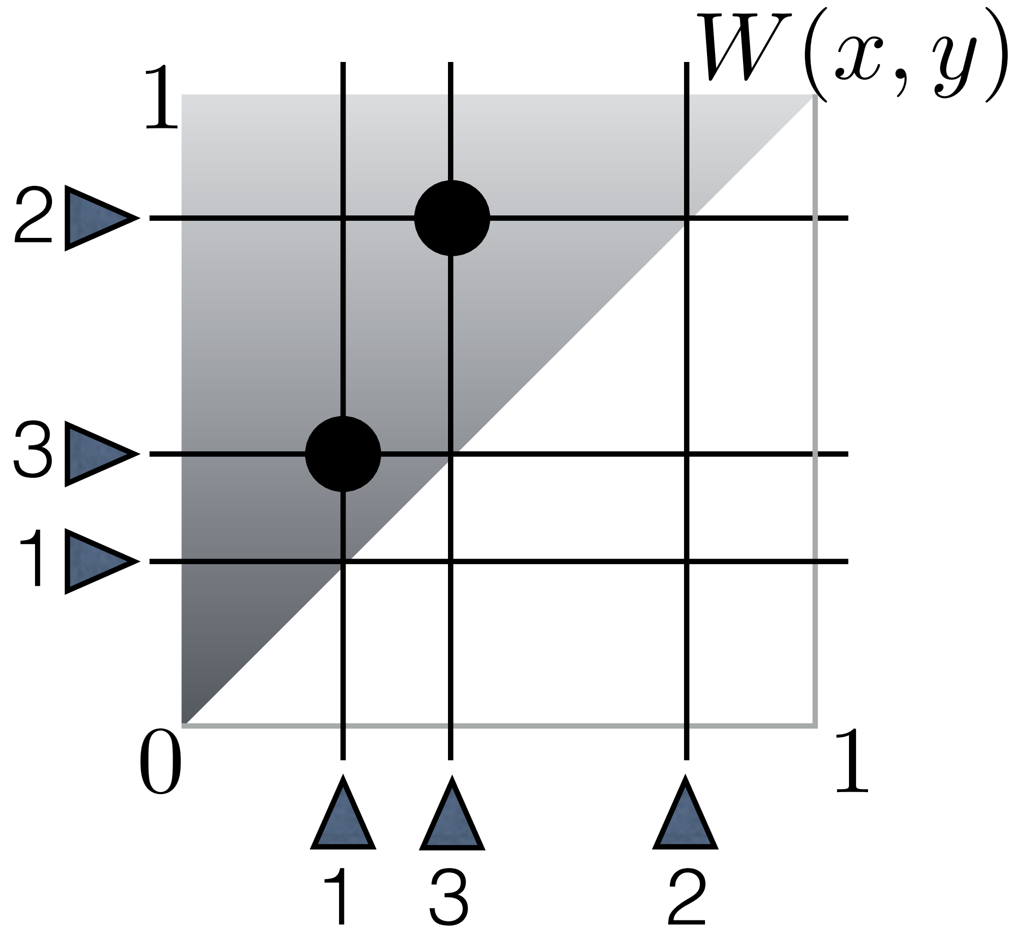
Aldous-Hoover

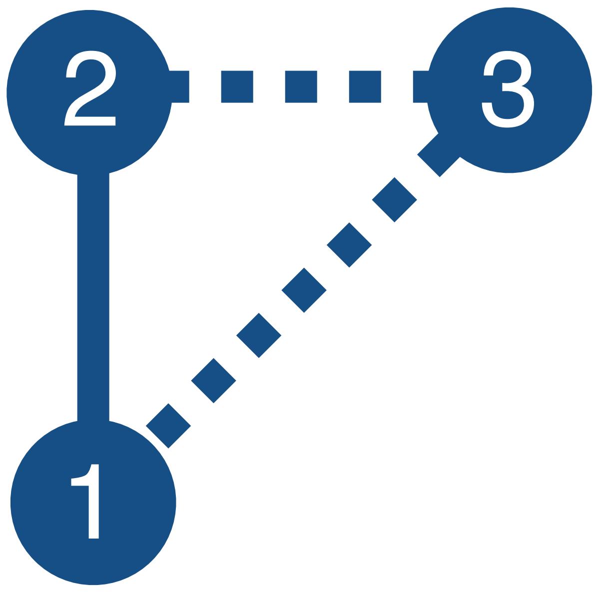




[Hoover 1979, Aldous 1981]

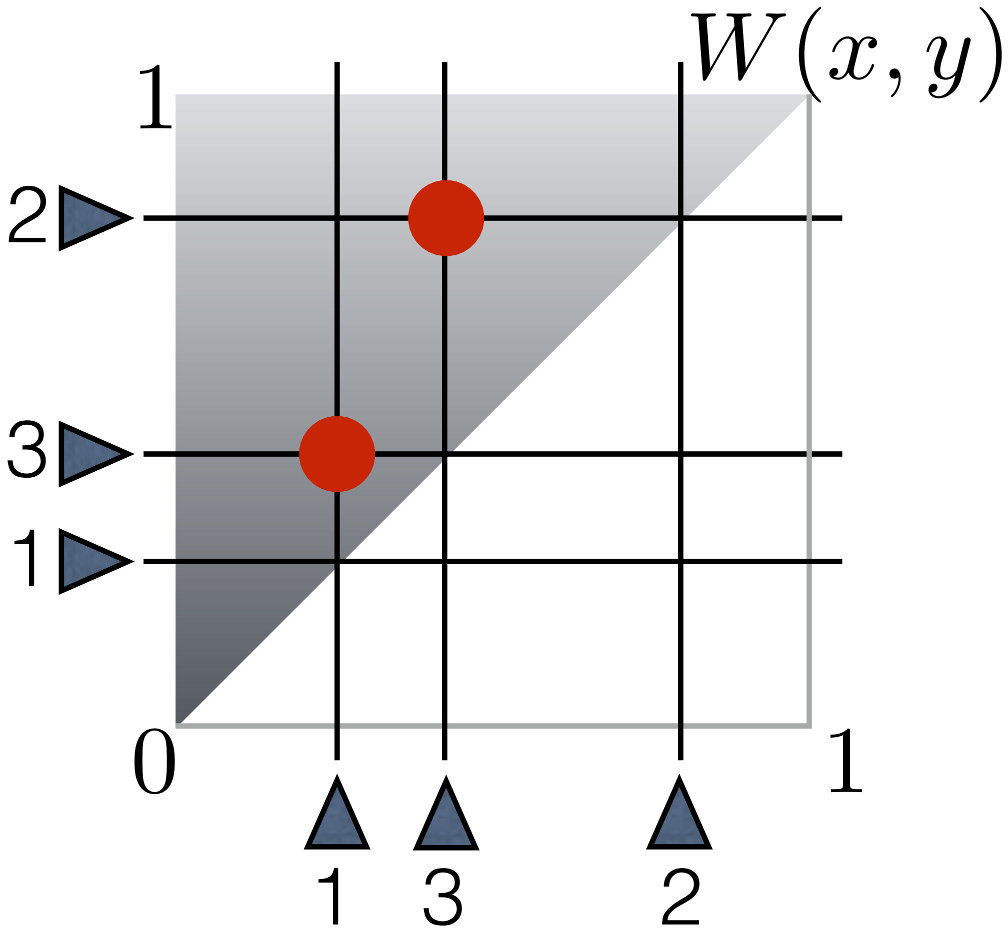
Aldous-Hoover

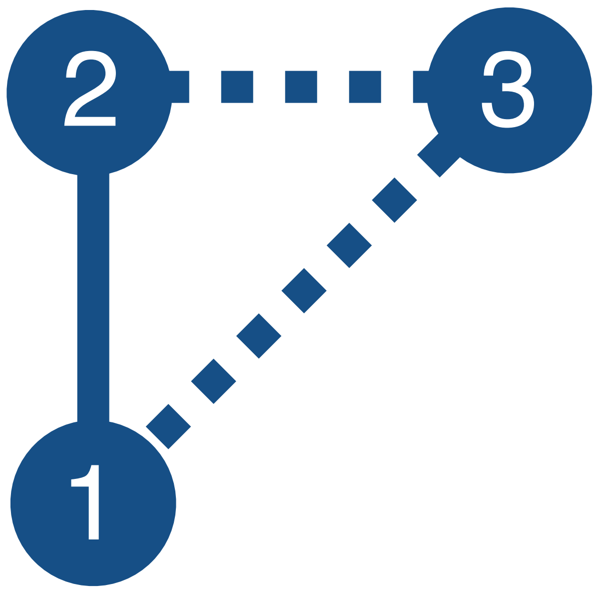




[Hoover 1979, Aldous 1981]

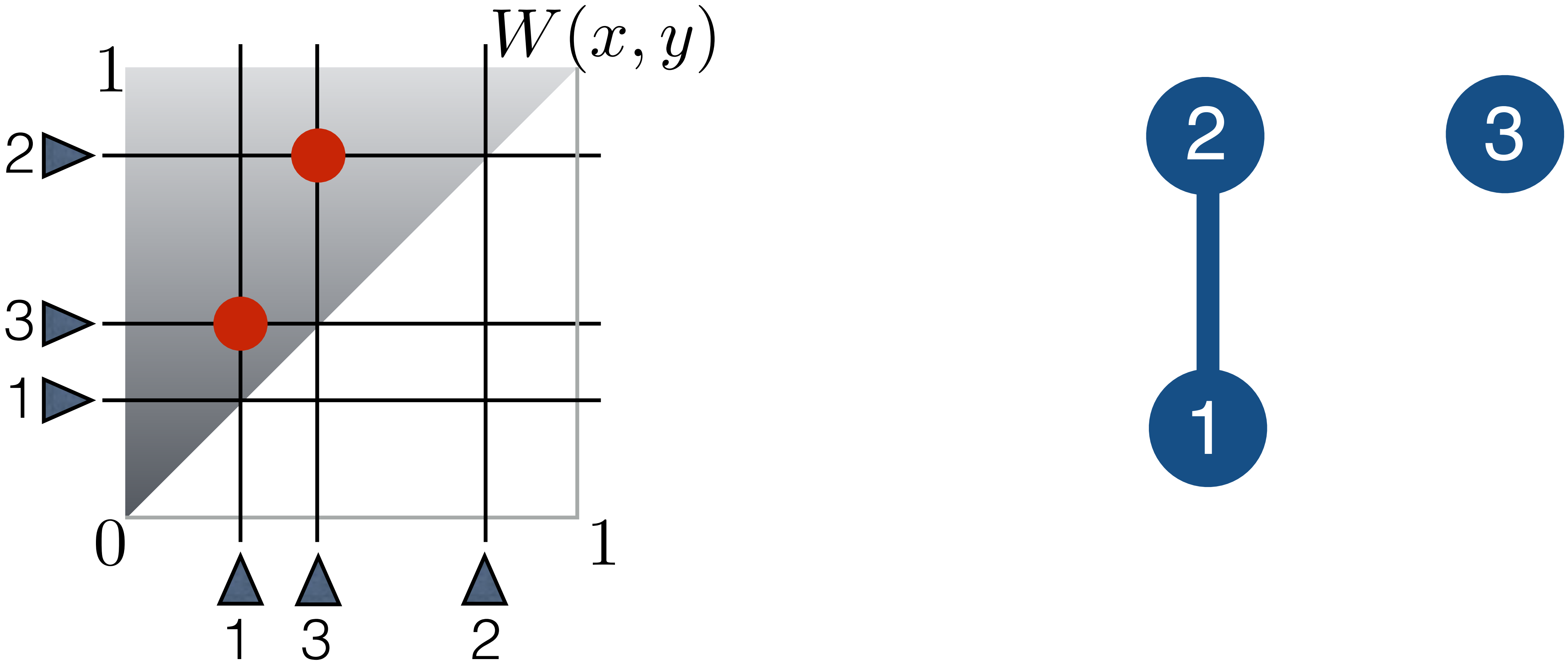
Aldous-Hoover





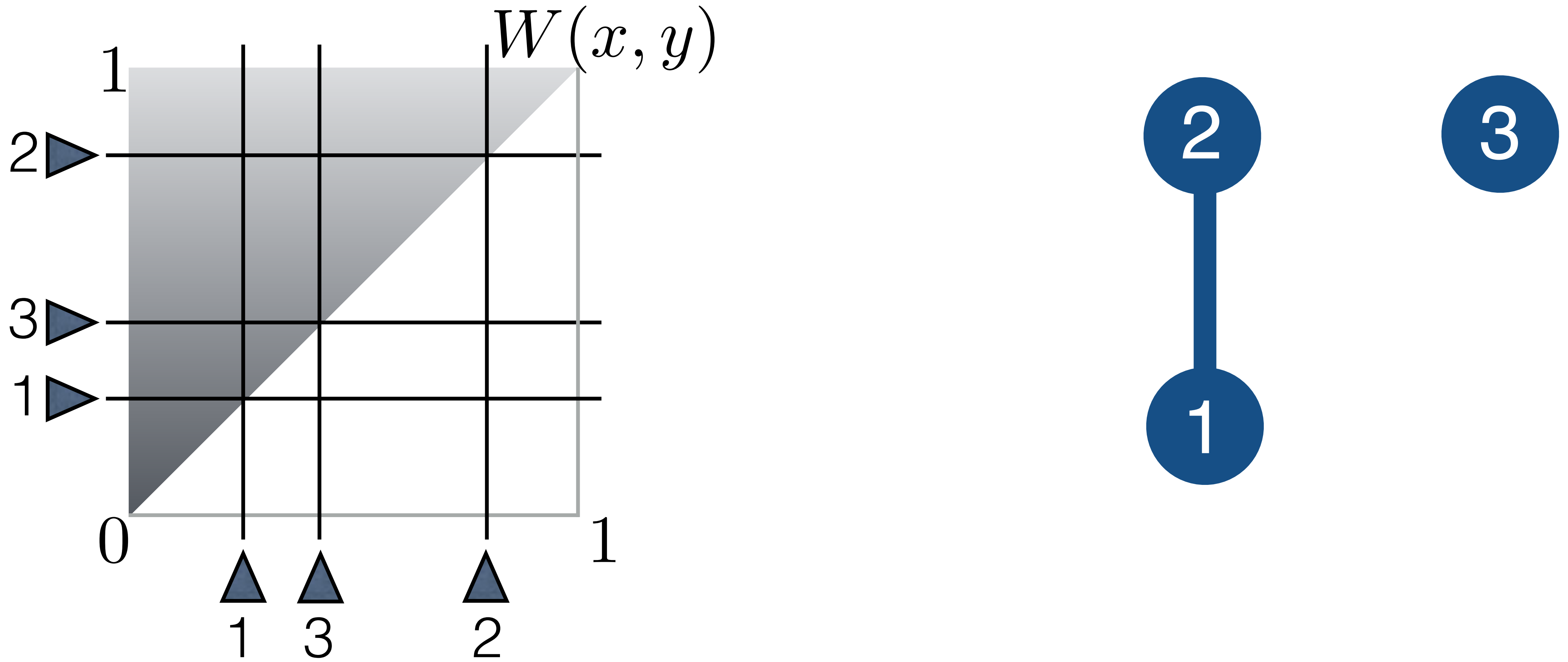
[Hoover 1979, Aldous 1981]

Aldous-Hoover



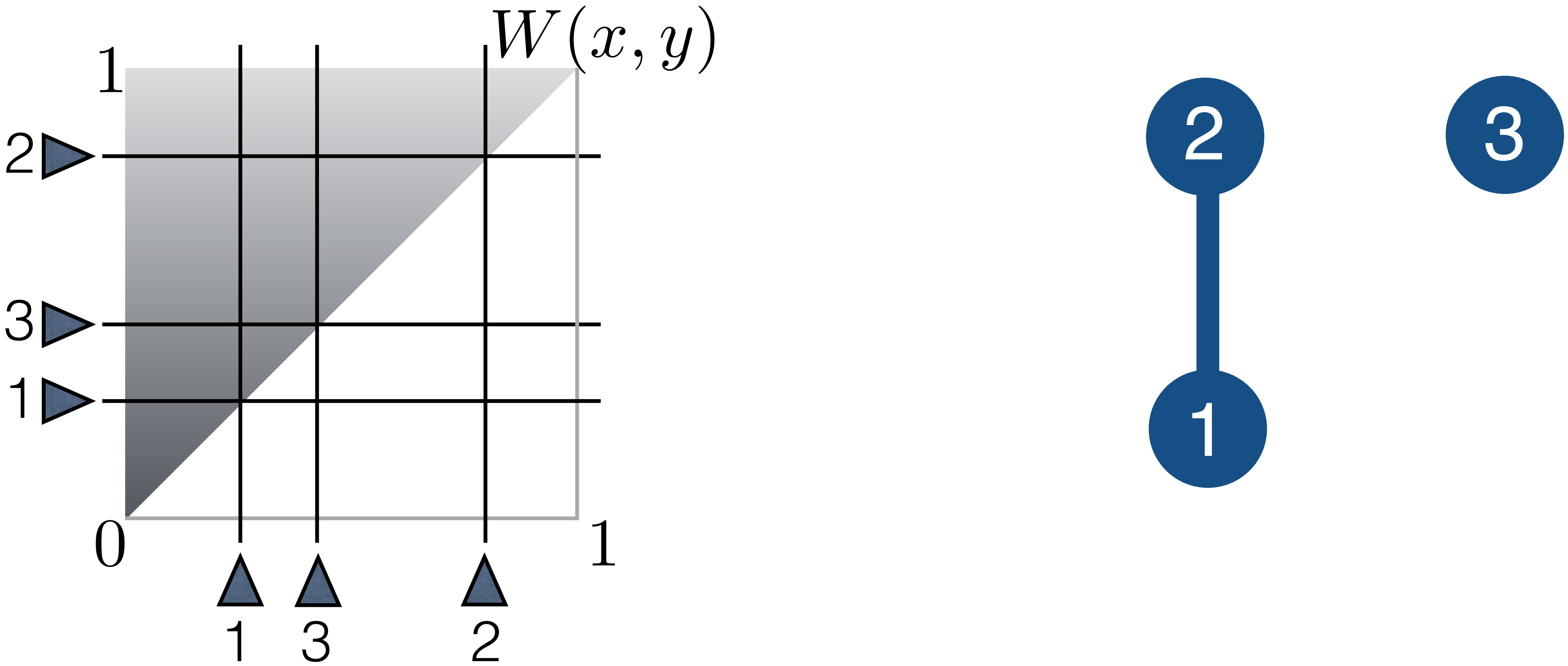
[Hoover 1979, Aldous 1981]

Aldous-Hoover



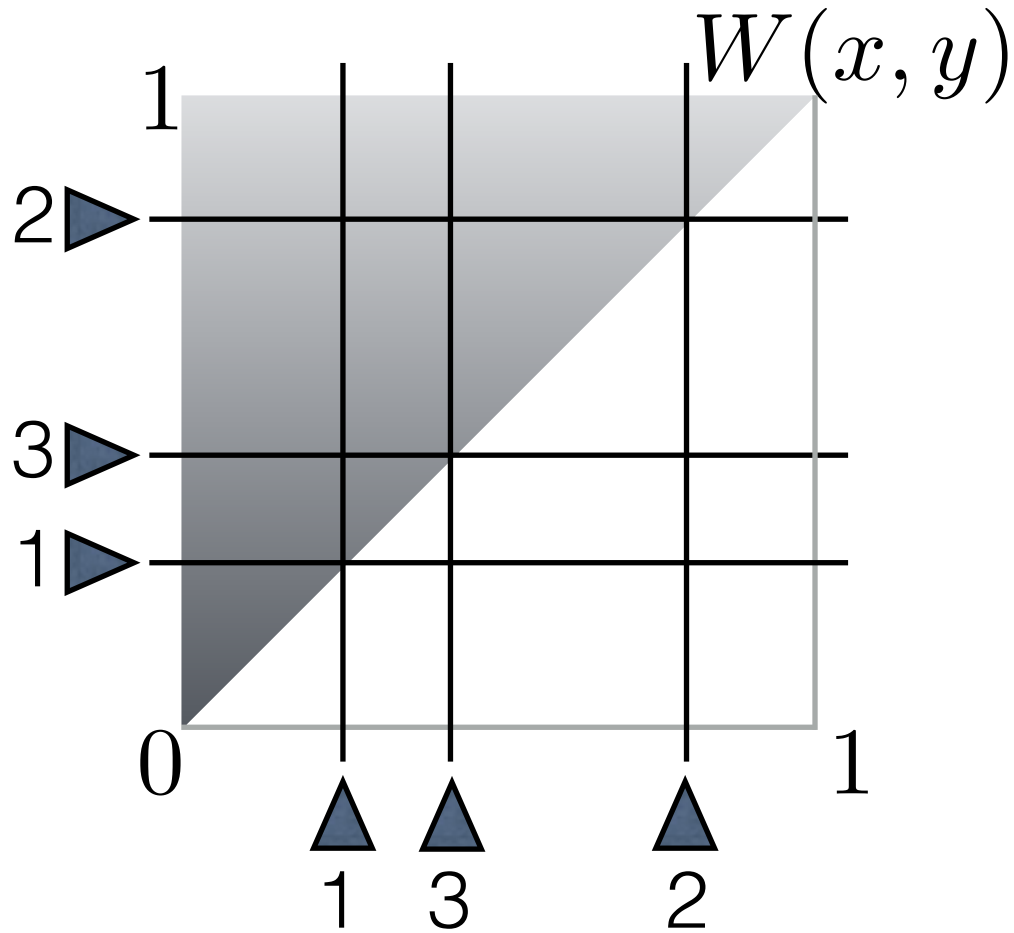
[Hoover 1979, Aldous 1981]

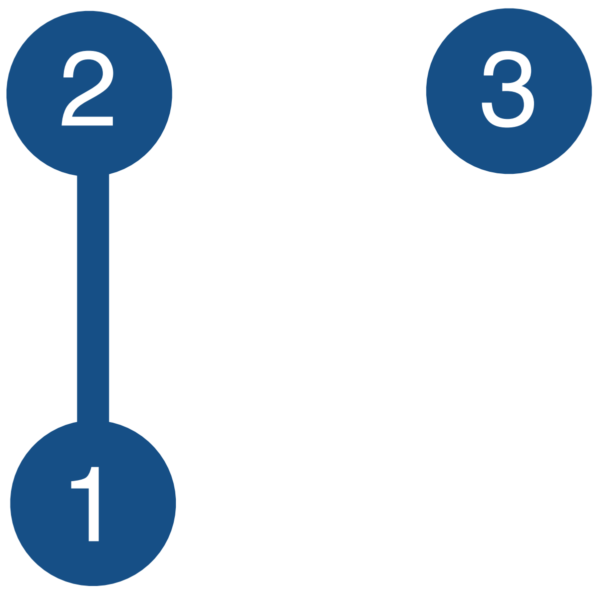
Aldous-Hoover



[Hoover 1979, Aldous 1981,
Lloyd et al 2012]

Aldous-Hoover

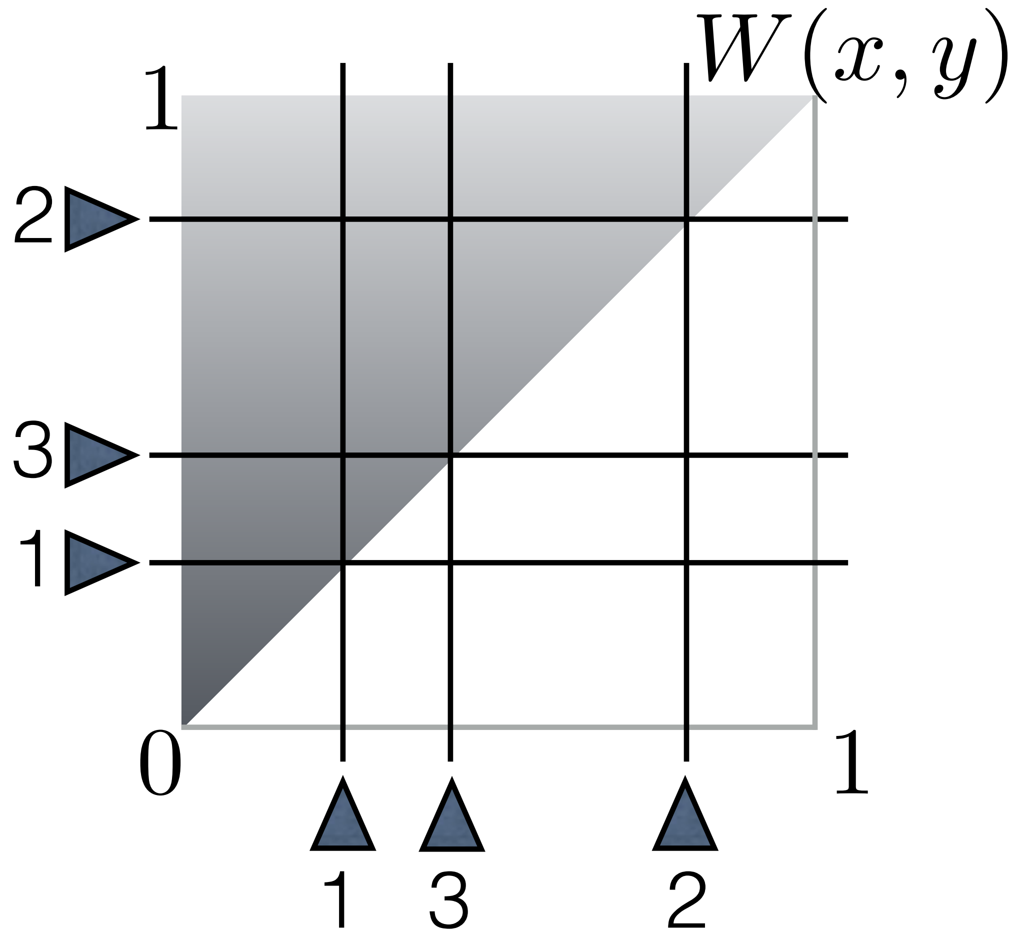


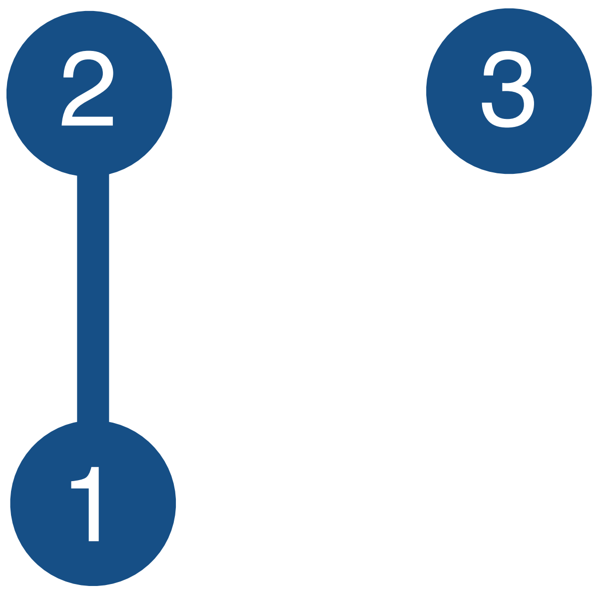


Thm (AH). Every node-exch. graph seq. has a *graphon* rep.

[Hoover 1979, Aldous 1981,
Lloyd et al 2012, Orbanz, Roy 2015]

Aldous-Hoover



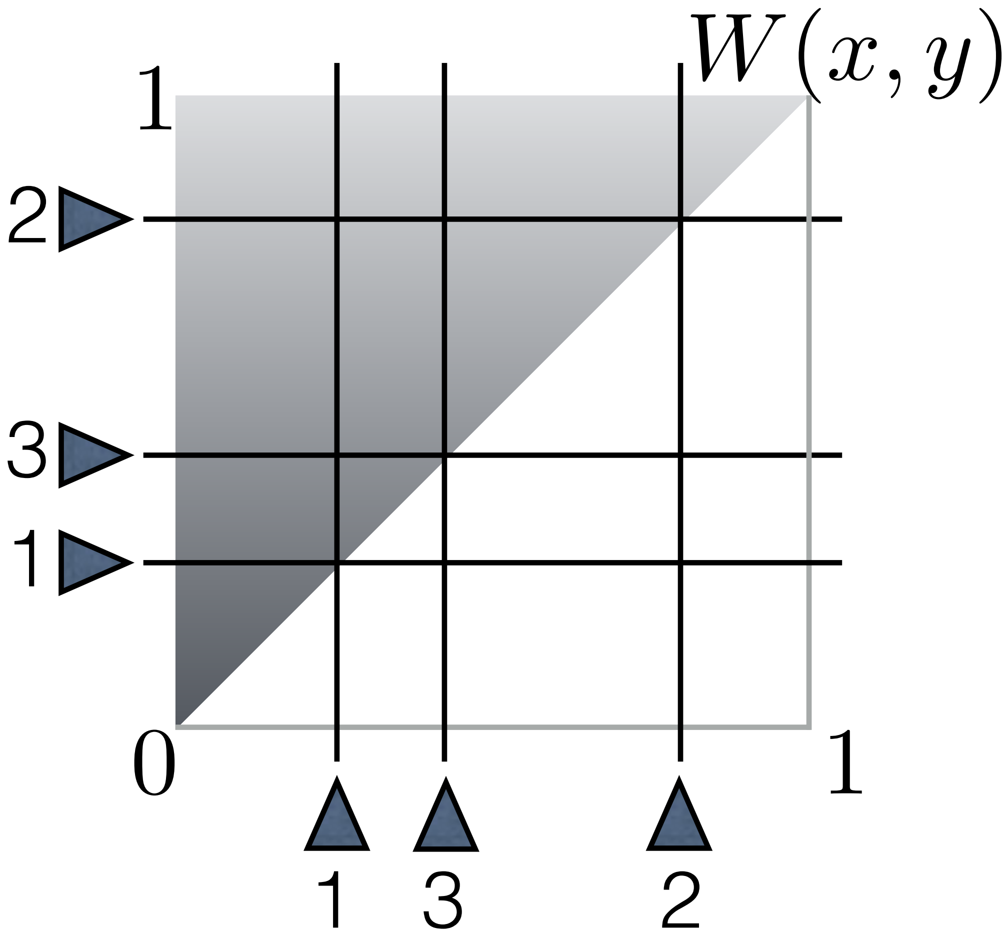


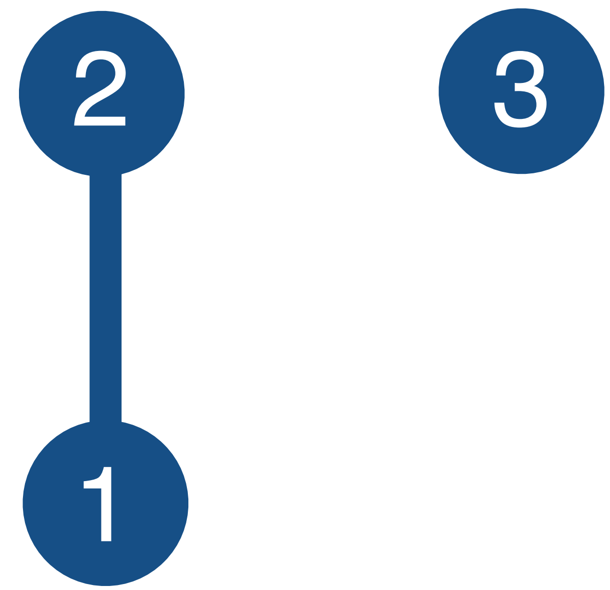
Thm (AH). Every node-exch. graph seq. has a *graphon* rep.

$$\mathbb{E}[\#\text{edges}(G_n)]$$

[Hoover 1979, Aldous 1981,
Lloyd et al 2012, Orbanz, Roy 2015]

Aldous-Hoover



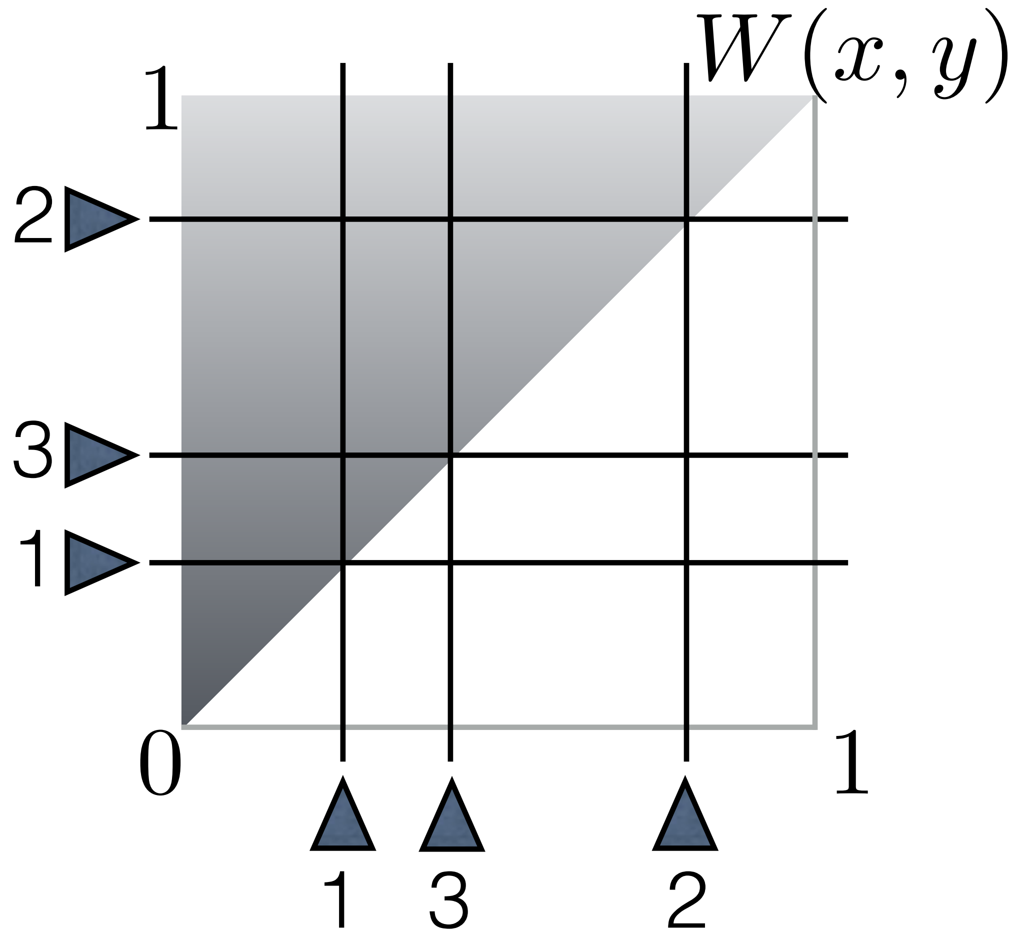


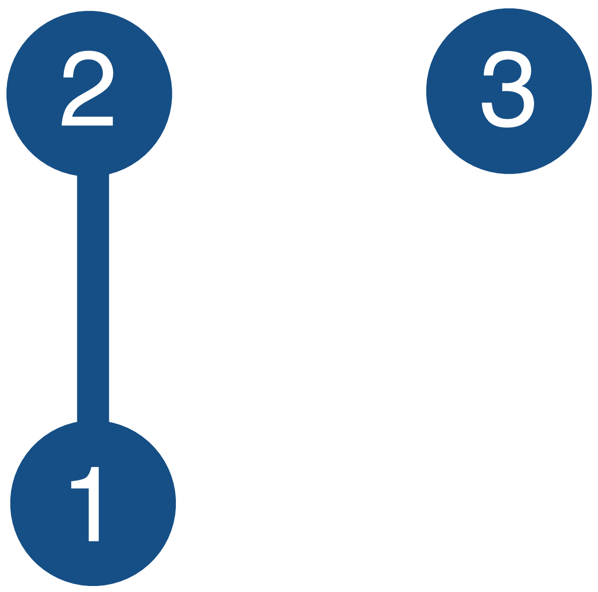
Thm (AH). Every node-exch. graph seq. has a *graphon* rep.

$$\mathbb{E}[\#\text{edges}(G_n)] = \mathbb{E}\left[\binom{n}{2}\frac{1}{2}\int_0^1\int_0^1 W(x,y)\,dx\,dy\right]$$

[Hoover 1979, Aldous 1981,
Lloyd et al 2012, Orbanz, Roy 2015]

Aldous-Hoover





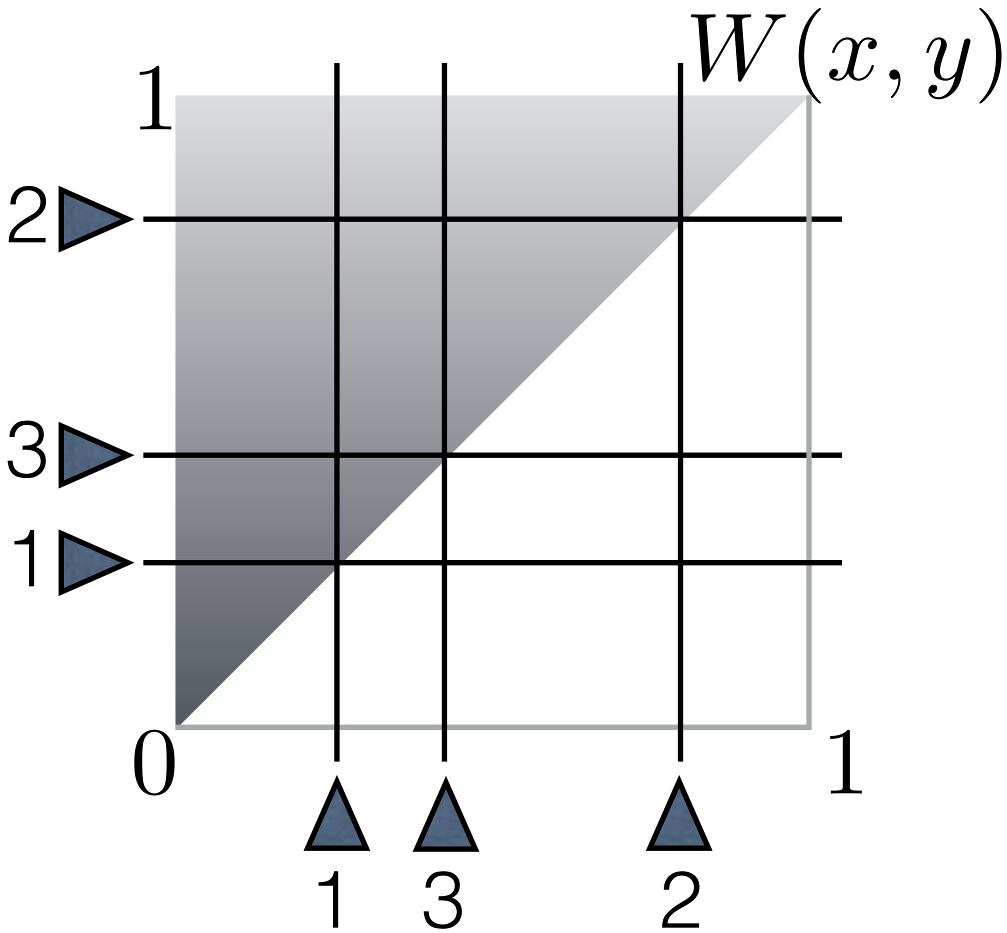
Thm (AH). Every node-exch. graph seq. has a *graphon* rep.

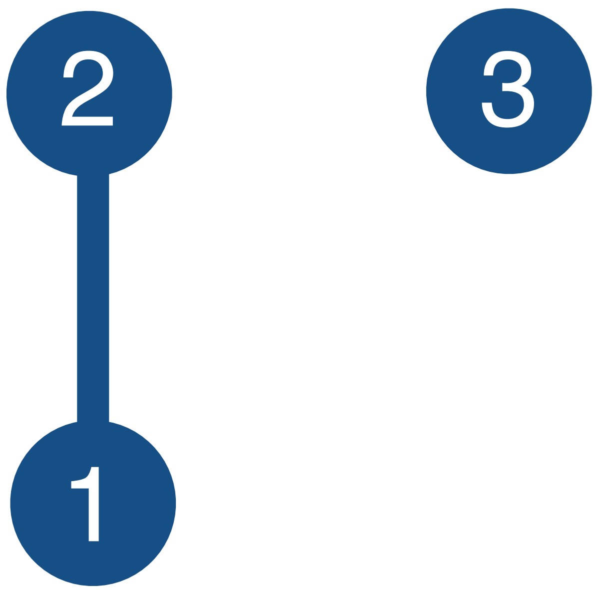
$$\mathbb{E}[\#\text{edges}(G_n)] = \mathbb{E}\left[\binom{n}{2}\frac{1}{2}\int_0^1\int_0^1 W(x,y)\,dx\,dy\right]$$

$$\sim cn^2$$

[Hoover 1979, Aldous 1981,
Lloyd et al 2012, Orbanz, Roy 2015]

Aldous-Hoover





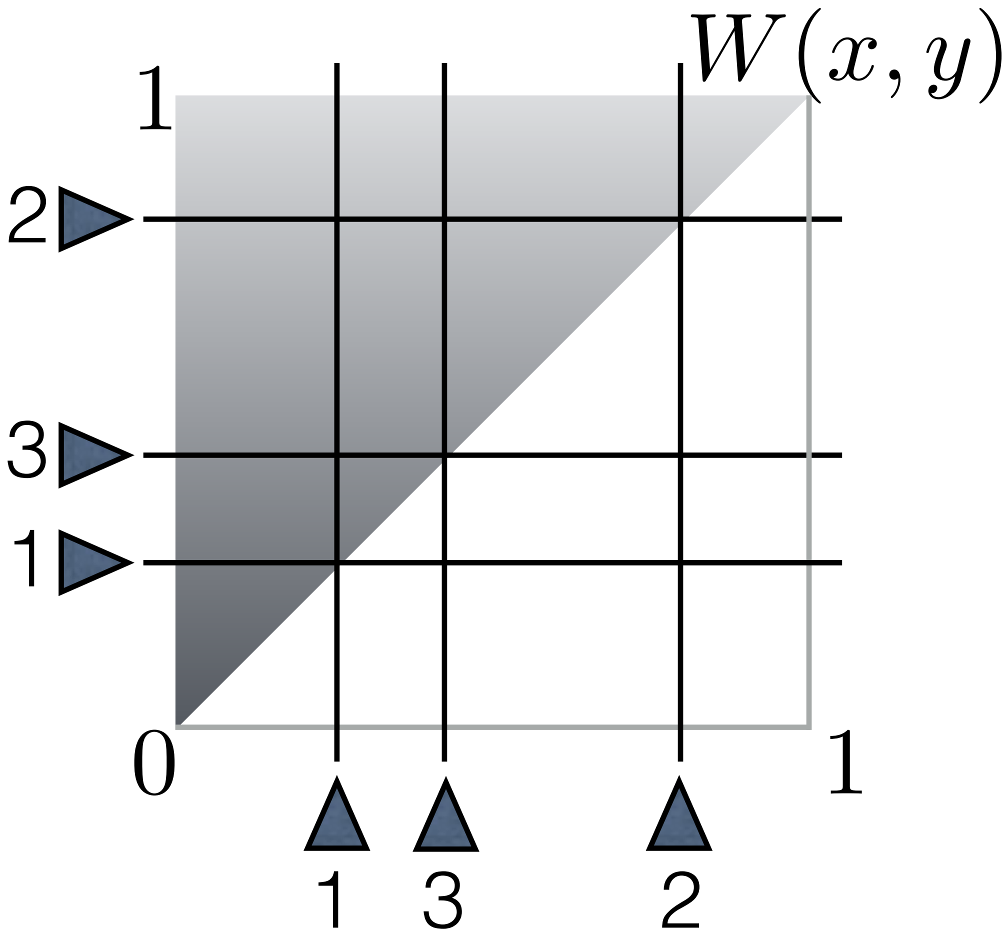
Thm (AH). Every node-exch. graph seq. has a *graphon* rep.

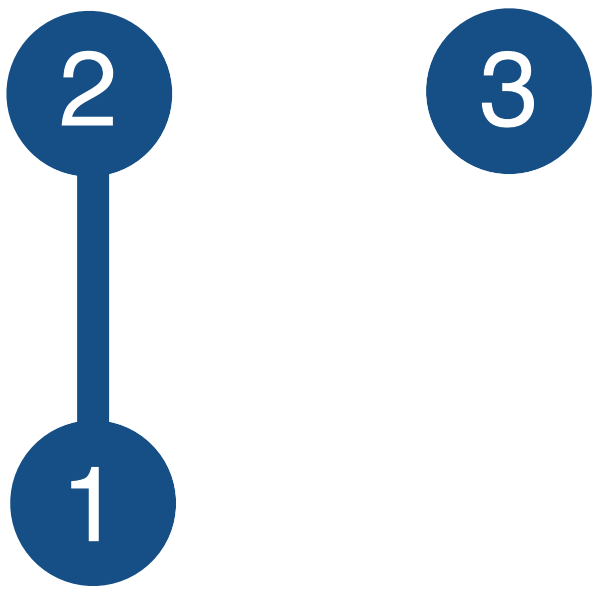
$$\mathbb{E}[\#\text{edges}(G_n)] = \mathbb{E}\left[\binom{n}{2}\frac{1}{2}\int_0^1\int_0^1 W(x,y)\,dx\,dy\right]$$

$$\sim cn^2 = c\cdot[\#\text{nodes}(G_n)]^2$$

[Hoover 1979, Aldous 1981,
Lloyd et al 2012, Orbanz, Roy 2015]

Aldous-Hoover





Thm (AH). Every node-exch. graph seq. has a *graphon* rep.

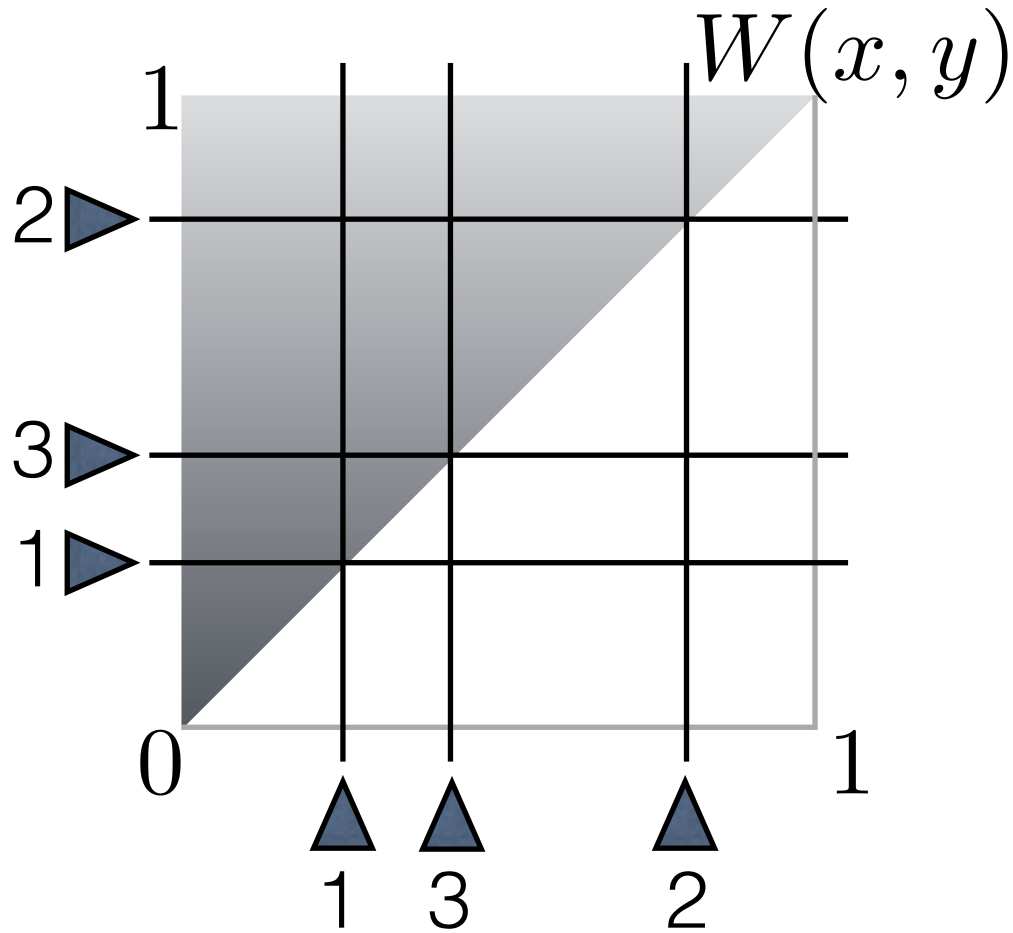
$$\mathbb{E}[\#\text{edges}(G_n)] = \mathbb{E}\left[\binom{n}{2}\frac{1}{2}\int_0^1\int_0^1 W(x,y)\,dx\,dy\right]$$

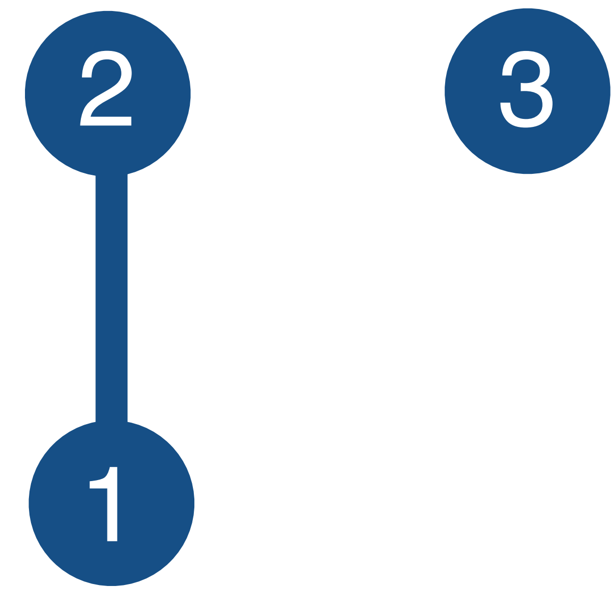
$$\sim cn^2 = c\cdot[\#\text{nodes}(G_n)]^2$$

Cor. Every node-exch graph sequence is dense (or empty) a.s.

[Hoover 1979, Aldous 1981,
Lloyd et al 2012, Orbanz, Roy 2015]

Aldous-Hoover





Thm (AH). Every node-exch. graph seq. has a *graphon* rep.

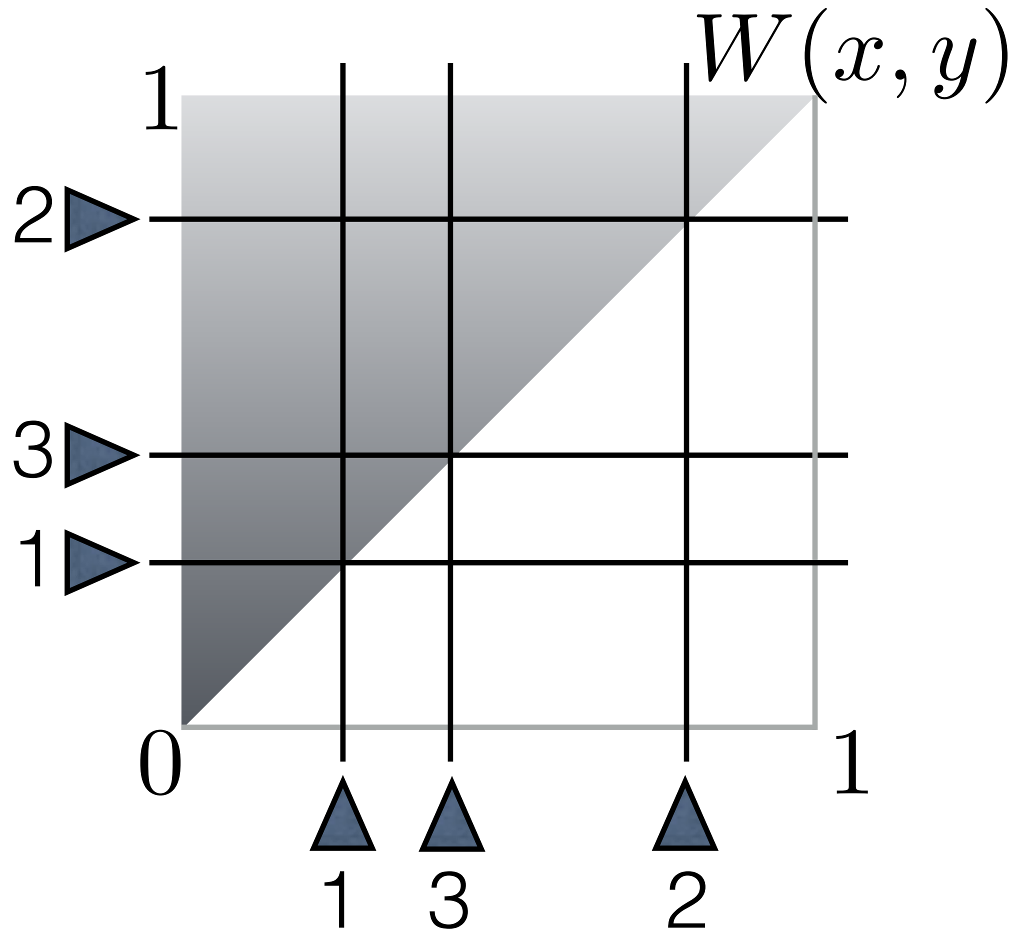
$$\mathbb{E}[\#\text{edges}(G_n)] = \mathbb{E}\left[\binom{n}{2}\frac{1}{2}\int_0^1\int_0^1 W(x,y)\,dx\,dy\right]$$
$$\sim cn^2 = c\cdot[\#\text{nodes}(G_n)]^2$$

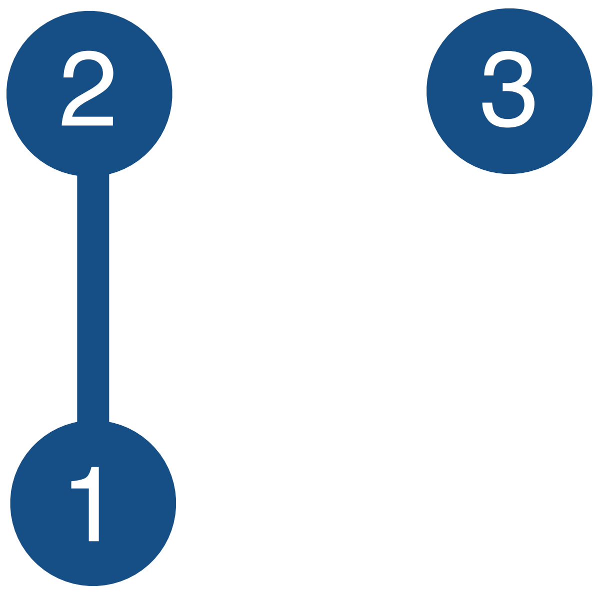
Cor. Every node-exch graph sequence is dense (or empty) a.s.

Intuition: To a given node, all other nodes look the same.

[Hoover 1979, Aldous 1981, Lloyd et al 2012, Orbanz, Roy 2015]

Aldous-Hoover





Thm (AH). Every node-exch. graph seq. has a *graphon* rep.

$$\mathbb{E}[\#\text{edges}(G_n)] = \mathbb{E}\left[\binom{n}{2}\frac{1}{2}\int_0^1\int_0^1 W(x,y)\,dx\,dy\right]$$

$$\sim cn^2 = c \cdot [\#\text{nodes}(G_n)]^2$$

Cor. Every node-exch graph sequence is dense (or empty) a.s.

Intuition: To a given node, all other nodes look the same.

Peter Orbanz Monday 11:30am

[Hoover 1979, Aldous 1981, Lloyd et al 2012, Orbanz, Roy 2015]

Q: Is there an alternative framework for sparse graphs?

Q: Is there an alternative framework for sparse graphs?

A: **Many** ideas

Q: Is there an alternative framework for sparse graphs?

A: **Many** ideas

[Caron, Fox 2017; Veitch, Roy 2015, 2016; Borgs, Chayes, Cohn, Holden 2016; Broderick, Cai 2015; Cai, Broderick 2015a,b; Crane, Dempsey 2015a,b, 2016a; Cai, Campbell, Broderick 2016, Williamson 2016]

Q: Is there an alternative framework for sparse graphs?

A: **Many** ideas

[Caron, Fox 2017; Veitch, Roy 2015, 2016; Borgs, Chayes, Cohn, Holden 2016; Broderick, Cai 2015; Cai, Broderick 2015a,b; Crane, Dempsey 2015a,b, 2016a; Cai, Campbell, Broderick 2016, Williamson 2016]

- Graphexes, at this workshop:
  - Chayes, Monday 1:45pm
  - Borgs, Monday 2:30pm

Q: Is there an alternative framework for sparse graphs?

A: **Many** ideas

[Caron, Fox 2017; Veitch, Roy 2015, 2016; Borgs, Chayes, Cohn, Holden 2016; Broderick, Cai 2015; Cai, Broderick 2015a,b; Crane, Dempsey 2015a,b, 2016a; Cai, Campbell, Broderick 2016, Williamson 2016]

- Graphexes, at this workshop:
    - Chayes, Monday 1:45pm
    - Borgs, Monday 2:30pm
- Global sparsity and local density, at this workshop:
    - Williamson, Wedn 1:45pm

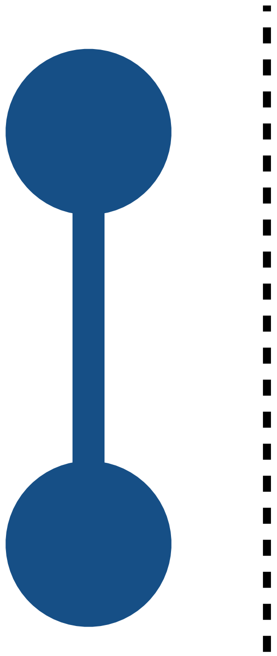
Q: Is there an alternative framework for sparse graphs?

A: **Many** ideas

[Caron, Fox 2017; Veitch, Roy 2015, 2016; Borgs, Chayes, Cohn, Holden 2016; Broderick, Cai 2015; Cai, Broderick 2015a,b; Crane, Dempsey 2015a,b, 2016a; Cai, Campbell, Broderick 2016, Williamson 2016]

- Graphexes, at this workshop:
  - Chayes, Monday 1:45pm
  - Borgs, Monday 2:30pm
- Global sparsity and local density, at this workshop:
  - Williamson, Wedn 1:45pm
- Idea: exchange the edges instead of nodes
  - Our work + Don’t miss independent graphs work by Crane & Dempsey!

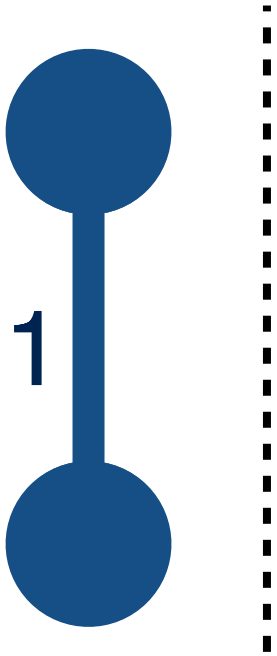
A New Way: Edges



G_1

 [Broderick, Cai 2015; Cai, Broderick 2015a,b; Crane, Dempsey 2015a,b,16a; Cai, Campbell, Broderick 2016]

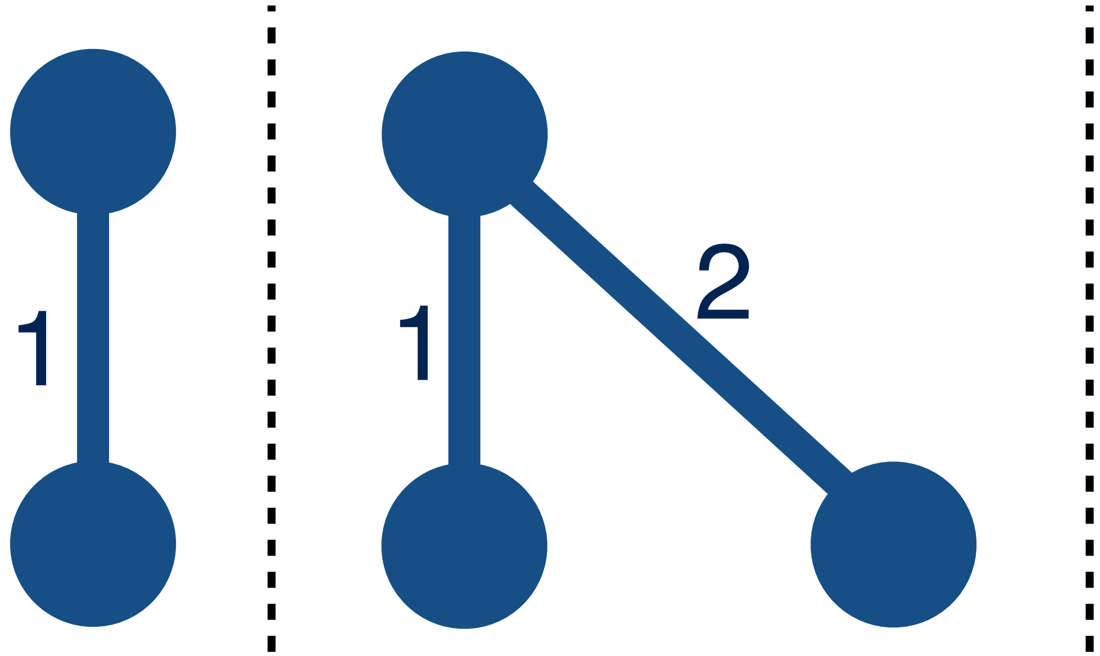
A New Way: Edges



G_1

 [Broderick, Cai 2015; Cai, Broderick 2015a,b; Crane, Dempsey 2015a,b,16a; Cai, Campbell, Broderick 2016]

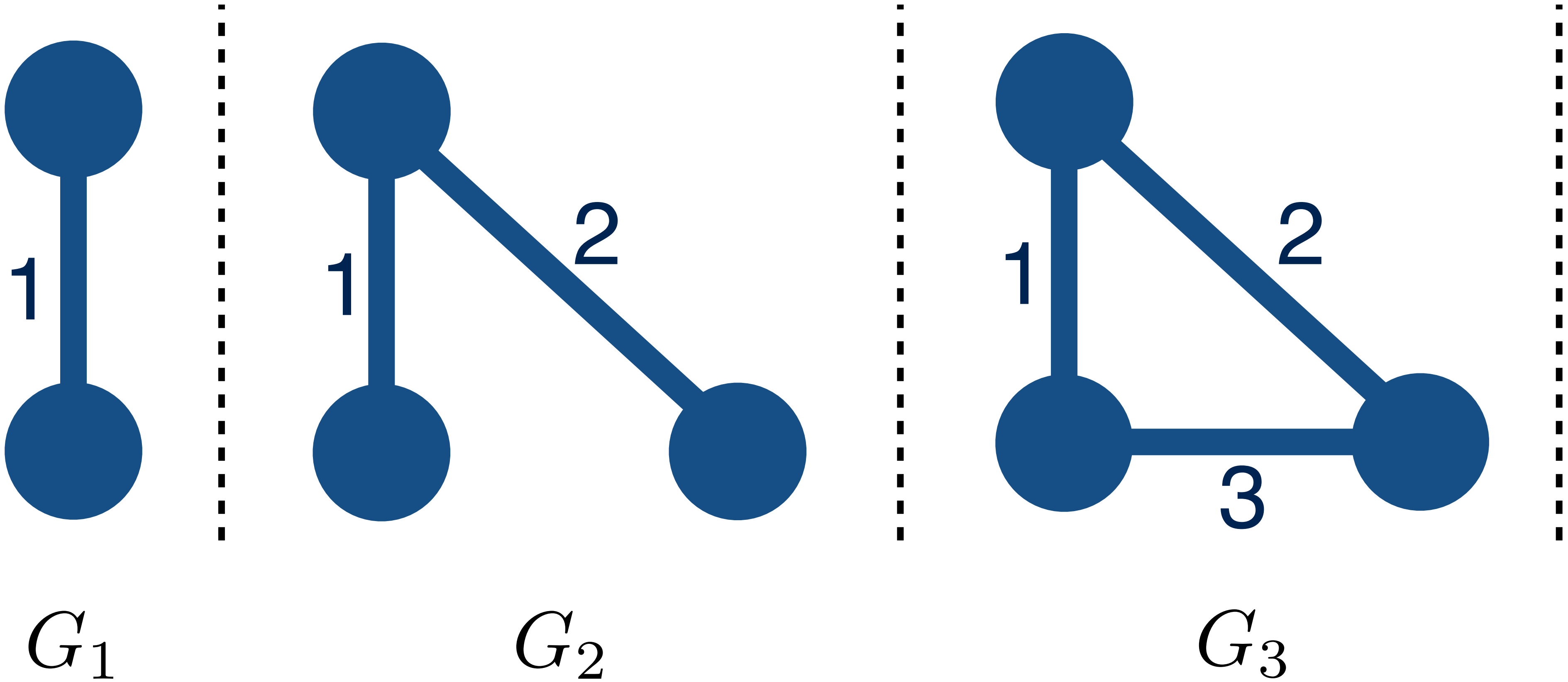
A New Way: Edges



G_1 G_2

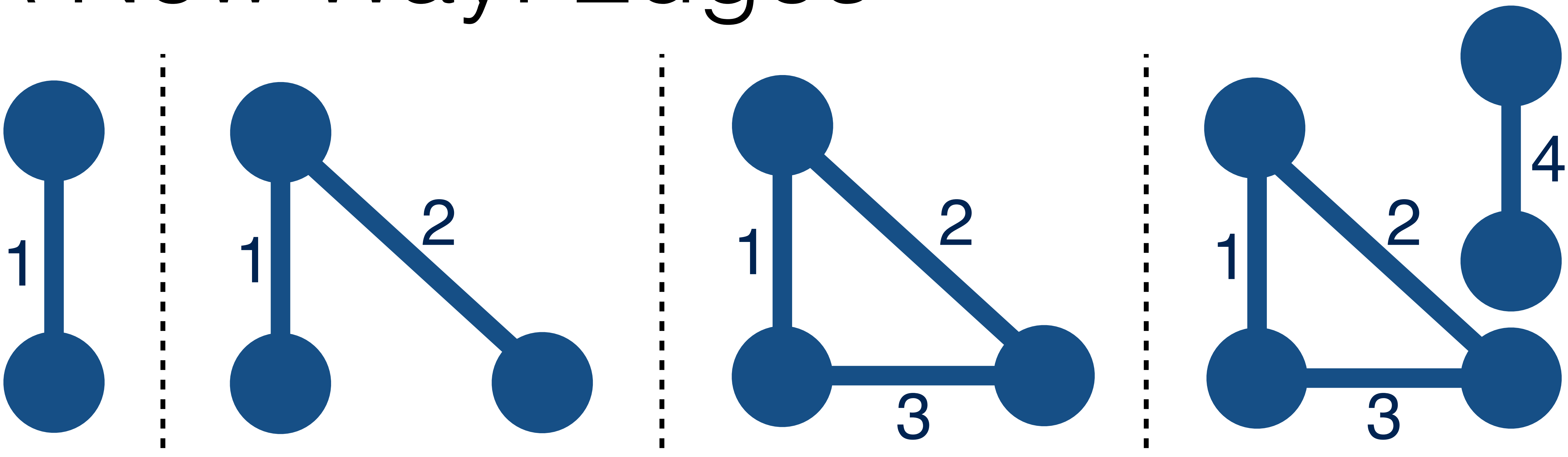
 [Broderick, Cai 2015; Cai, Broderick 2015a,b; Crane, Dempsey 2015a,b,16a; Cai, Campbell, Broderick 2016]

A New Way: Edges



 [Broderick, Cai 2015; Cai, Broderick 2015a,b; Crane, Dempsey 2015a,b,16a; Cai, Campbell, Broderick 2016]

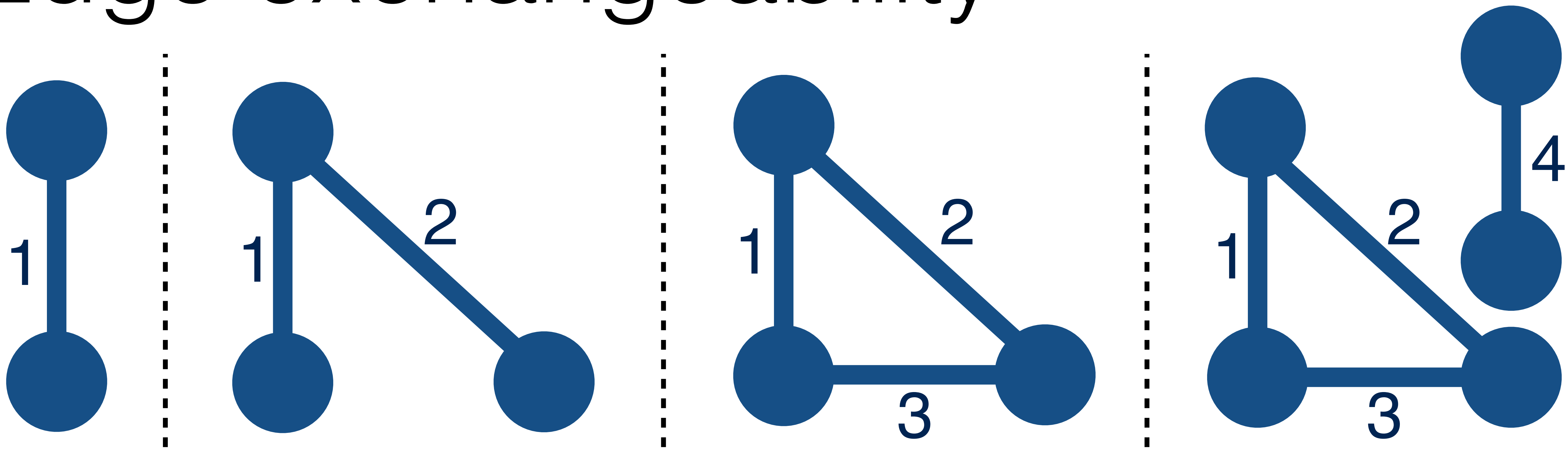
A New Way: Edges



G_1 G_2 G_3 G_4

 [Broderick, Cai 2015; Cai, Broderick 2015a,b; Crane, Dempsey 2015a,b,16a; Cai, Campbell, Broderick 2016]

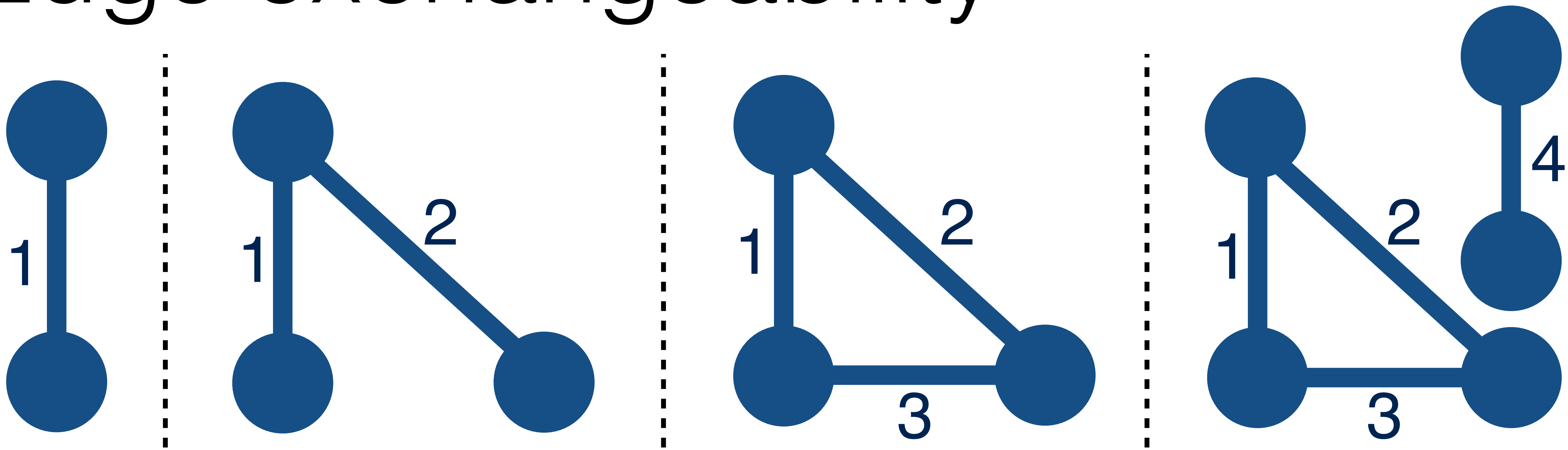
Edge exchangeability



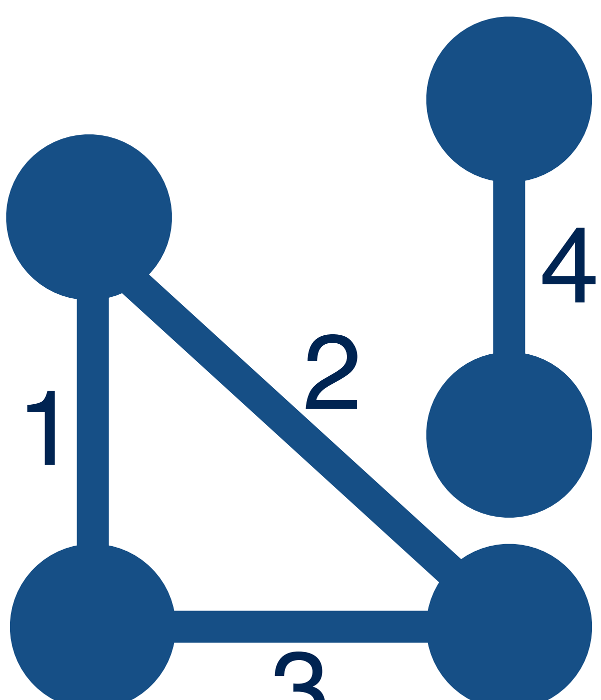
G_1 G_2 G_3 G_4

[Broderick, Cai 2015; Cai, Broderick 2015a,b; Crane, Dempsey 2015a,b,16a; Cai, Campbell, Broderick 2016]

Edge exchangeability

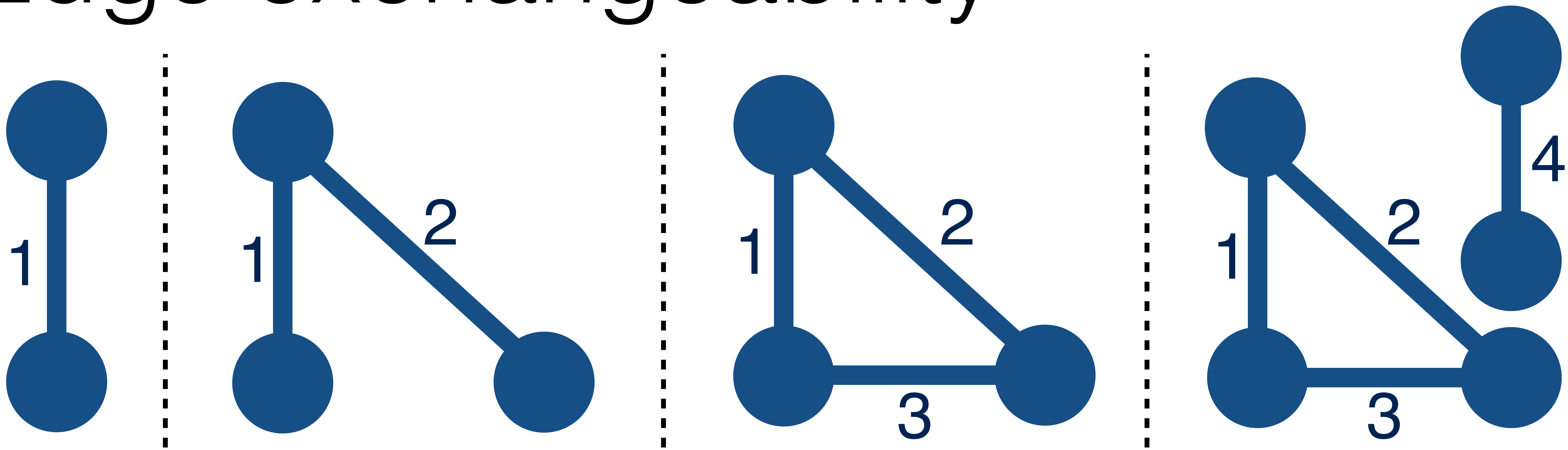


G_1 G_2 G_3 G_4

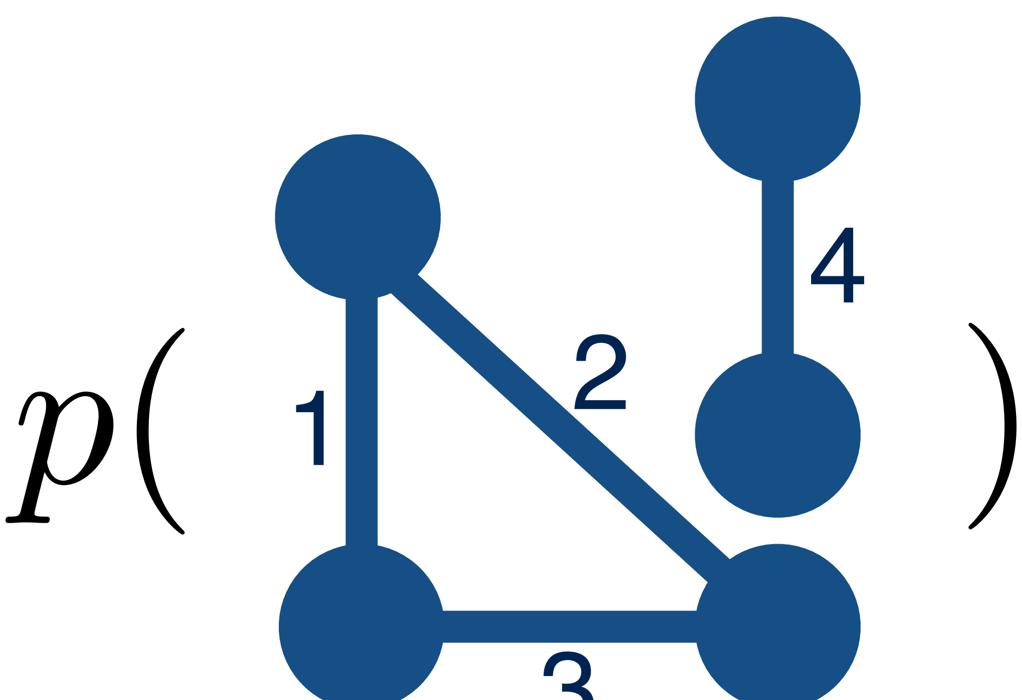


 [Broderick, Cai 2015; Cai, Broderick 2015a,b; Crane, Dempsey 2015a,b,16a; Cai, Campbell, Broderick 2016]

Edge exchangeability

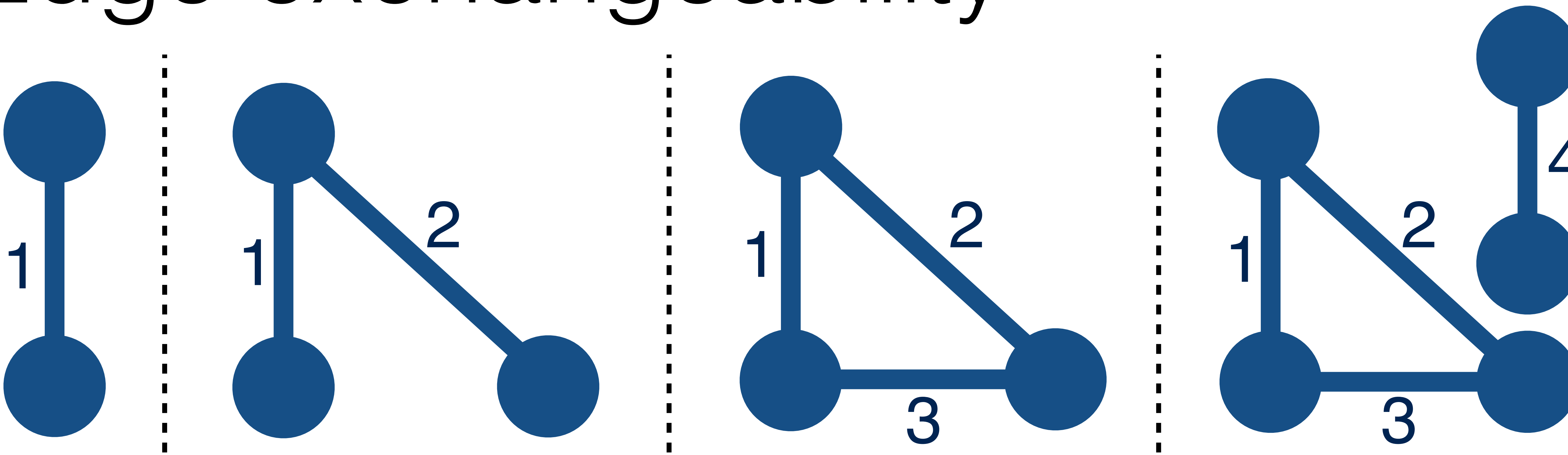


G_1 G_2 G_3 G_4

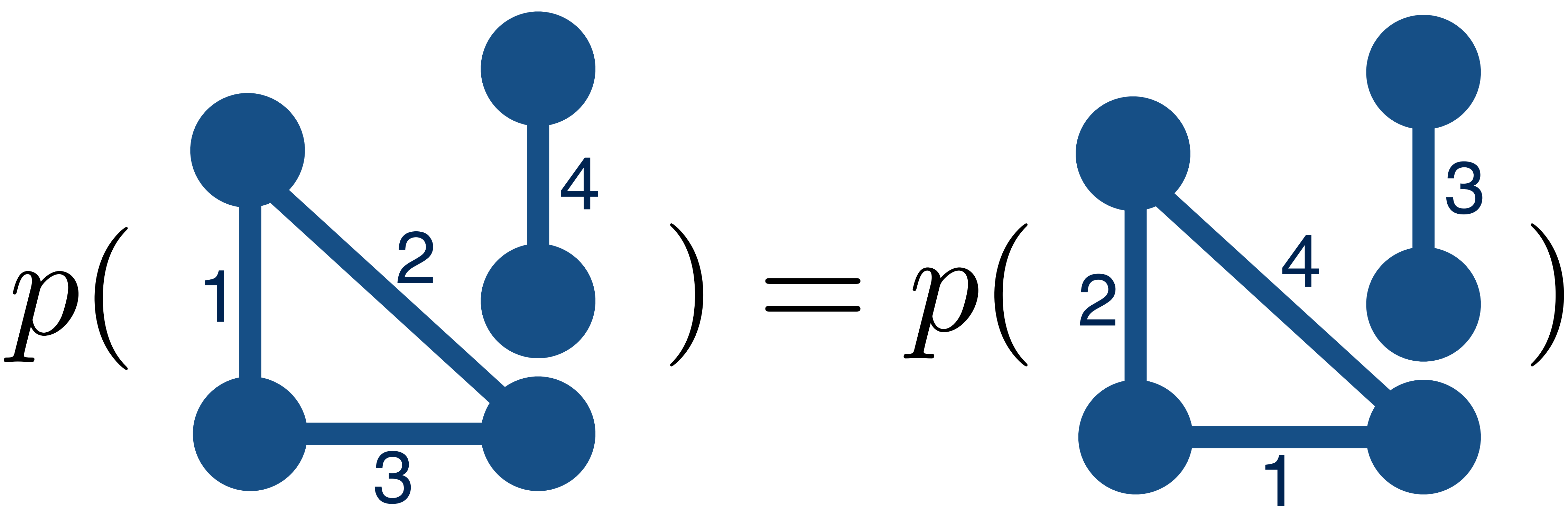


 [Broderick, Cai 2015; Cai, Broderick 2015a,b; Crane, Dempsey 2015a,b,16a; Cai, Campbell, Broderick 2016]

Edge exchangeability



G_1 G_2 G_3 G_4



 [Broderick, Cai 2015; Cai, Broderick 2015a,b; Crane, Dempsey 2015a,b,16a; Cai, Campbell, Broderick 2016]

Edge exchangeability

Thm. A wide class of edge-exchangeable graph models yields sparse graph sequences

Thm. A paintbox-style characterization for edge-exchangeable graph sequences

 [Broderick, Cai 2015; Cai, Broderick 2015a,b; Crane, Dempsey 2015a,b,16a; Cai, Campbell, Broderick 2016]

Roadmap

- Example problem: clustering
- Example NPBayes model: Dirichlet process (DP)
- De Finetti for clustering: Kingman Paintbox
- De Finetti for networks/graphs

- Big questions
  - Why NPBayes? **Learn more from more data**
  - What does an infinite/growing number of parameters really mean (in NPBayes)? **Components vs clusters; latent vs realized**
  - Why is NPBayes challenging but practical? **Infinite dimensional parameter but finitely many realized (in practice, e.g., can integrate out or truncate the infinity)**
  - How does thinking about exchangeability help us use NPBayes in practice?

Roadmap

- Example problem: clustering
- Example NPBayes model: Dirichlet process (DP)
- De Finetti for clustering: Kingman Paintbox
- De Finetti for networks/graphs

- Big questions
  - Why NPBayes? **Learn more from more data**
  - What does an infinite/growing number of parameters really mean (in NPBayes)? **Components vs clusters; latent vs realized**
  - Why is NPBayes challenging but practical? **Infinite dimensional parameter but finitely many realized (in practice, e.g., can integrate out or truncate the infinity)**
  - How does thinking about exchangeability help us use NPBayes in practice?

Roadmap

- Example problem: clustering
- Example NPBayes model: Dirichlet process (DP)
- De Finetti for clustering: Kingman Paintbox
- De Finetti for networks/graphs

- Big questions
  - Why NPBayes? **Learn more from more data**
  - What does an infinite/growing number of parameters really mean (in NPBayes)? **Components vs clusters; latent vs realized**
  - Why is NPBayes challenging but practical? **Infinite dimensional parameter but finitely many realized (in practice, e.g., can integrate out or truncate the infinity)**
  - How does thinking about exchangeability help us use NPBayes in practice? **Limitations & better models**

Roadmap [http://www.tamarabroderick.com/tutorials.html]

- Example problem: clustering
- Example NPBayes model: Dirichlet process (DP)
- De Finetti for clustering: Kingman Paintbox
- De Finetti for networks/graphs

- Big questions
  - Why NPBayes? **Learn more from more data**
  - What does an infinite/growing number of parameters really mean (in NPBayes)? **Components vs clusters; latent vs realized**
  - Why is NPBayes challenging but practical? **Infinite dimensional parameter but finitely many realized (in practice, e.g., can integrate out or truncate the infinity)**
  - How does thinking about exchangeability help us use NPBayes in practice? **Limitations & better models**

References (page 1 of 7)

T Broderick, MI Jordan, and J Pitman. Cluster and feature modeling from combinatorial stochastic processes. *Statistical Science*, 2013.

T Broderick, J Pitman, and MI Jordan. Feature allocations, probability functions, and paintboxes. *Bayesian Analysis*, 2013.

D Cai, T Campbell, and T Broderick. Edge-exchangeable graphs and sparsity. *NeurIPS* 2016. (Early versions at *NeurIPS 2015 Networks and NPBayes Workshops*.)

T Campbell, D Cai, and Broderick T. Exchangeable trait allocations. *Electronic Journal of Statistics*, 2018. (Early versions at *NeurIPS 2016 Adaptive Nonparametric and NPBayes Workshops*.)

T Campbell, S Syed, C Yang, MI Jordan, T Broderick. Local Exchangeability. ArXiv: 1906.09507.

H Crane and W Dempsey. Atypical scaling behavior persists in real world interaction networks. arXiv 1509.08184, 2015.

H Crane and W Dempsey. A framework for statistical network modeling. arXiv 1509.08185, 2015.

H Crane and W Dempsey. Edge exchangeable models for network data. arXiv 1603.04571, 2016. / H Crane and W Dempsey. Edge exchangeable models for interaction networks. *JASA*, 2018.

H Crane and W Dempsey. Relational exchangeability. arXiv 1607.06762, 2016.

S Williamson. Nonparametric Network Models for Link Prediction. *JMLR*, 2016.

References (2/7)

DJ Aldous. *Exchangeability and related topics*. Springer, 1983.

CE Antoniak. Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *The Annals of Statistics*, 1974.

E Arjas and D Gasbarra. Nonparametric Bayesian inference from right censored survival data, using the Gibbs sampler. *Statistica Sinica*, 1994.

J Bertoin. *Random Fragmentation and Coagulation Processes*. Cambridge University Press, 2006.

D Blackwell and JB MacQueen. Ferguson distributions via Pólya urn schemes. *The Annals of Statistics*, 1973.

T Broderick, MI Jordan, and J Pitman. Beta processes, stick-breaking, and power laws. *Bayesian Analysis*, 2012.

T Broderick, MI Jordan, and J Pitman. Cluster and feature modeling from combinatorial stochastic processes. *Statistical Science*, 2013.

T Broderick, J Pitman, and MI Jordan. Feature allocations, probability functions, and paintboxes. *Bayesian Analysis*, 2013.

T Broderick, AC Wilson, and MI Jordan. Posteriors, conjugacy, and exponential families for completely random measures. *Bernoulli*, 2018.

D Cai, T Campbell, and T Broderick. Edge-exchangeable graphs and sparsity. *NIPS*, 2016.

- *NeurIPS 2015 Workshop on Networks in the Social & Information Sciences*.
- *NeurIPS 2015 Workshop, Bayesian Nonparametrics: The Next Generation.*

T Campbell, D Cai, and T Broderick. Exchangeable trait allocations. *Electronic Journal of Statistics*, 2018.

- *NeurIPS 2016 Workshop on Adaptive & Scalable Nonparametric Methods in ML.*
- *NeurIPS 2016 Workshop on Practical Bayesian Nonparametrics.*

References (3/7)

T Campbell, JH Huggins, JP How, and T Broderick. Truncated Random Measures. *Bernoulli*, 2019.

T Campbell, S Syed, C Yang, MI Jordan, T Broderick. Local Exchangeability. ArXiv:1906.09507.

H Crane and W Dempsey. Atypical scaling behavior persists in real world interaction networks. arXiv 1509.08184, 2015.

H Crane and W Dempsey. A framework for statistical network modeling. arXiv 1509.08185, 2015.

H Crane and W Dempsey. Edge exchangeable models for network data. arXiv 1603.04571, 2016. / H Crane and W Dempsey. Edge exchangeable models for interaction networks. *JASA*, 2018.

H Crane and W Dempsey. Relational exchangeability. arXiv 1607.06762, 2016.

W Del Pozzo, TGF Li, and C Messenger. Cosmological inference using only gravitational wave observations of binary neutron stars. *Physical Review D*, 2017.

W Del Pozzo, CPL Berry, A Ghosh, TSF Haines, LP Singer, and A Vecchio. Dirichlet Process Gaussian-mixture model: An application to localizing coalescing binary neutron stars with gravitational-wave observations. *Monthly Notices of the Royal Astronomical Society,* 2018.

S Engen. A note on the geometric series as a species frequency model. *Biometrika*, 1975.

MD Escobar and M West. Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*, 1995.

W Ewens. The sampling theory of selectively neutral alleles. *Theoretical Population Biology*, 1972.

W Ewens. Population genetics theory -- the past and the future. *Mathematical and Statistical Developments of Evolutionary Theory*, 1987.

TS Ferguson. A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1973.

TS Ferguson. Bayesian density estimation by mixtures of normal distributions. *Recent Advances in Statistics*, 1983.

References (4/7)

EB Fox, personal website. Retrieved in 2016 from: http://www.stat.washington.edu/~ebfox/research.html --- Associated paper: EB Fox, MC Hughes, EB Sudderth, and MI Jordan. *The Annals of Applied Statistics*, 2014.

S Ghosal, JK Ghosh, and RV Ramamoorthi. Posterior consistency of Dirichlet mixtures in density estimation. *The Annals of Statistics*, 1999.

S Goldwater, TL Griffiths, and M Johnson. Interpolating between types and tokens by estimating power-law generators. *NIPS*, 2005.

A Gnedin, B Hansen, and J Pitman. Notes on the occupancy problem with infinitely many boxes: general asymptotics and power laws. *Probability Surveys*, 2007.

TL Griffiths and Z Ghahramani. Infinite latent feature models and the Indian buffet process. *NIPS*, 2005.

K Harris, TL Parsons, UZ Ijaz, L Lahti, I Holmes, and C Quince. Linking statistical and ecological theory: Hubbell's unified neutral theory of biodiversity as a hierarchical Dirichlet process. *Proceedings of the IEEE*, 2017.

DL Hartl and AG Clark. *Principles of Population Genetics, Fourth Edition*. 2003.

E Hewitt and LJ Savage. Symmetric measures on Cartesian products. *Transactions of the American Mathematical Society*, 1955.

NL Hjort. Nonparametric Bayes estimators based on beta processes in models for life history data. *The Annals of Statistics*, 1990.

DN Hoover. Relations on probability spaces and arrays of random variables, *Preprint, Institute for Advanced Study*, 1979.

FM Hoppe. Pólya-like urns and the Ewens' sampling formula. *Journal of Mathematical Biology*, 1984.

H Ishwaran and LF James. Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association*, 2001.

L James. Bayesian Poisson calculus for latent feature modeling via generalized Indian Buffet Process priors. *Annals of Statistics*, 2017.

Y Kim. Nonparametric Bayesian estimators for counting processes. *The Annals of Statistics*, 1999.

References (5/7)

JFC Kingman. The representation of partition structures. *Journal of the London Mathematical Society*, 1978.

JFC Kingman. On the genealogy of large populations. *Journal of Applied Probability*, 1982.

JFC Kingman. *Poisson processes*, 1992.

AS Lan, D Vats, AE Waters, and RG Baraniuk. Mathematical language processing: Automatic grading and feedback for open response mathematical questions. In *Proceedings of the Second (2015) ACM Conference on Learning@Scale*, 2015.

JR Lloyd, P Orbanz, Z Ghahramani, and DM Roy. Random function priors for exchangeable arrays with applications to graphs and relational data. *NeurIPS*, 2012.

SN MacEachern and P Müller. Estimating mixtures of Dirichlet process models. *Journal of Computational and Graphical Statistics*, 1998.

JW McCloskey. A model for the distribution of individuals by species in an environment. *Ph.D. thesis, Michigan State University*, 1965.

K Miller, MI Jordan, and TL Griffiths. Nonparametric latent feature models for link prediction. *NeurIPS*, 2009.

RM Neal. Density modeling and clustering using Dirichlet diffusion trees. *Bayesian Statistics*, 2003.

P Orbanz. Construction of nonparametric Bayesian models from parametric Bayes equations. *NeurIPS*, 2009.

P Orbanz. Conjugate Projective Limits. arXiv preprint arXiv:1012.0363, 2010.

P Orbanz, DM Roy. Bayesian models of graphs, arrays and other exchangeable random structures. *IEEE TPAMI*, 2015.

GP Patil and C Taillie. Diversity as a concept and its implications for random communities. *Bulletin of the International Statistical Institute*, 1977.

J Pitman and M Yor. The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *The Annals of Probability*, 1997.

References (6/7)

S Prabhakaran, E Azizi, A Carr, and D Pe'er. Dirichlet process mixture model for correcting technical variation in single-cell gene expression data. *ICML*, 2016.

A Rodríguez, DB Dunson & AE Gelfand. The nested Dirichlet process. *Journal of the American Statistical Association*, 2008.

S Saria, D Koller, and A Penn. Learning individual and population traits from clinical temporal data. *NeurIPS*, 2010.

J Sethuraman. A constructive definition of Dirichlet priors. *Statistica Sinica*, 1994.

EB Sudderth and MI Jordan. Shared segmentation of natural scenes using dependent Pitman-Yor processes. *NeurIPS*, 2009.

YW Teh. A hierarchical Bayesian language model based on Pitman-Yor processes. *ACL*, 2006.

YW Teh, C Blundell, and L Elliott. Modelling genetic variations using fragmentation-coagulation processes. *NeurIPS*, 2011.

YW Teh, MI Jordan, MJ Beal, and DM Blei. Hierarchical Dirichlet processes. *Journal of the American Statistical Association*, 2006.

R Thibaux and MI Jordan. Hierarchical beta processes and the Indian buffet process. *ICML*, 2007.

V Veitch and DM Roy. The class of random graphs arising from exchangeable random measures. ArXiv:1512.03099, 2015.

V Veitch and DM Roy. Sampling and estimation for (sparse) exchangeable graphs. ArXiv:1611.00843, 2016.

J Wakeley. *Coalescent Theory: An Introduction*, Chapter 3, 2008.

References (7/7)

M West, P Müller, and MD Escobar. Hierarchical Priors and Mixture Models, With Application in Regression and Density Estimation. *Aspects of Uncertainty*, 1994.

Y Xu, P Müller, Y Yuan, K Gulukota, and Y Ji. MAD Bayes for tumor heterogeneity—feature allocation with exponential family sampling. *Journal of the American Statistical Association*, 2015.

Image References

E Bowlby. NOAA/Olympic Coast NMS; NOAA/OAR/Office of Ocean Exploration - NOAA Photo Library. Retrieved in 2016 from: https://en.wikipedia.org/wiki/Opisthoteuthis_californiana#/media/File:Opisthoteuthis_californiana.jpg

JW Cassidy, C Caldas, and A Bruna. Maintaining tumor heterogeneity in patient-derived tumor xenografts. *Cancer research*, 2015.

ESO/L. Calçada/M. Kornmesser. 16 October 2017, 16:00:00. Obtained in 2018 from: https://commons.wikimedia.org/wiki/File:Artist%E2%80%99s_impression_of_merging_neutron_stars.jpg || Source: https://www.eso.org/public/images/eso1733a/ (Creative Commons Attribution 4.0 International License)

MIT xPro, 2017. Obtained in 2018 from: https://mitxpro.mit.edu/courses/course-v1:MITProfessionalX+DSx+2017_T2/about