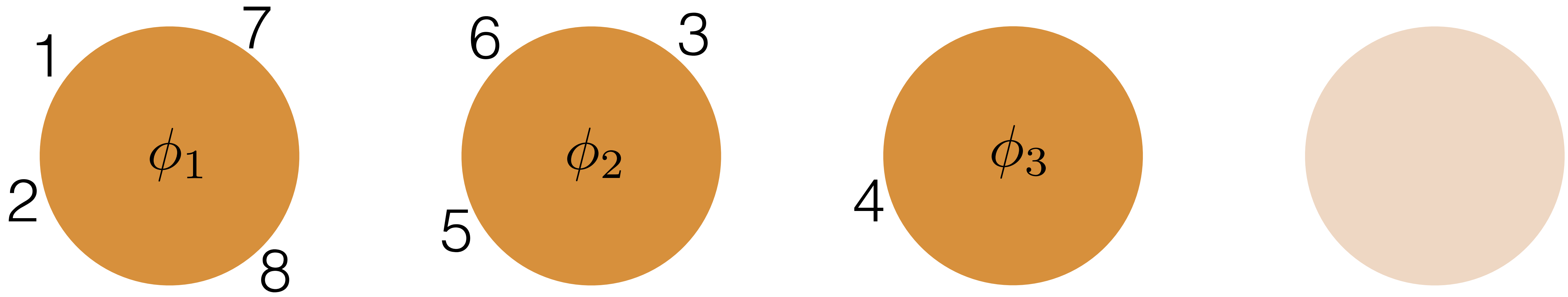
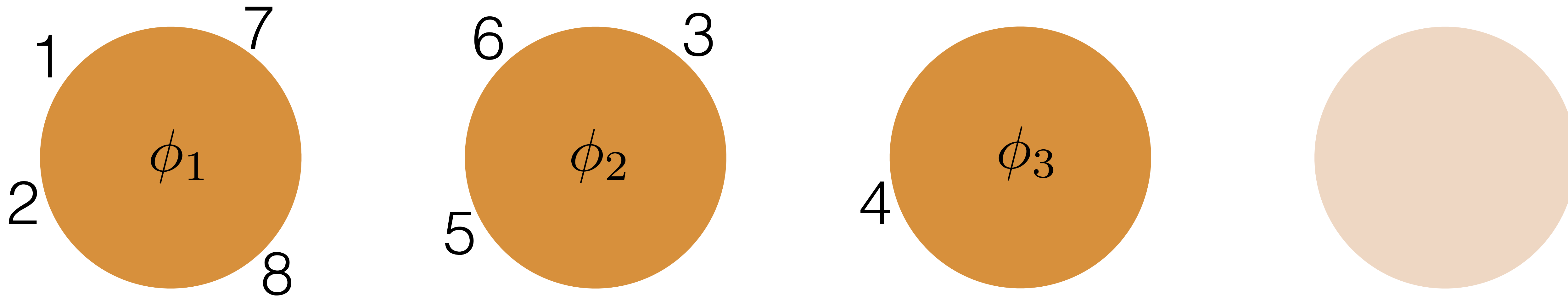
Nonparametric Bayesian Statistics: Part III

Tamara Broderick
ITT Career Development Assistant Professor
Electrical Engineering & Computer Science
MIT

Chinese restaurant process

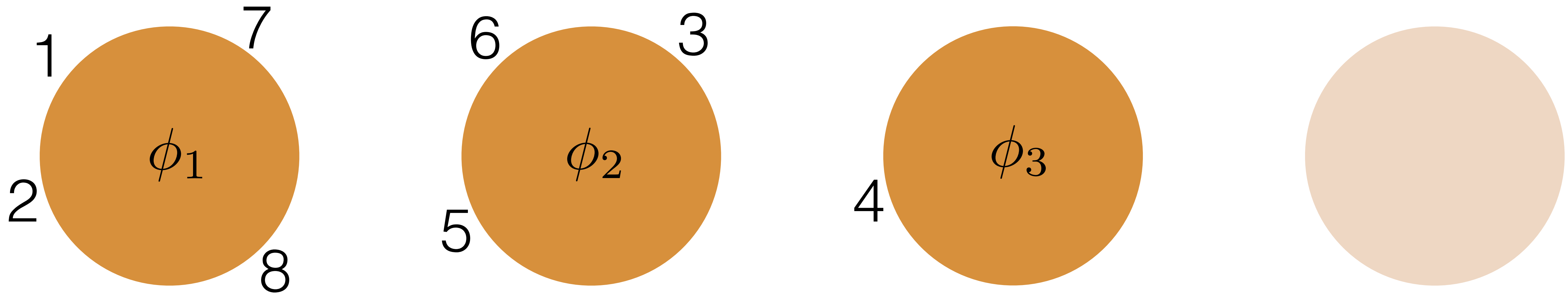


Chinese restaurant process



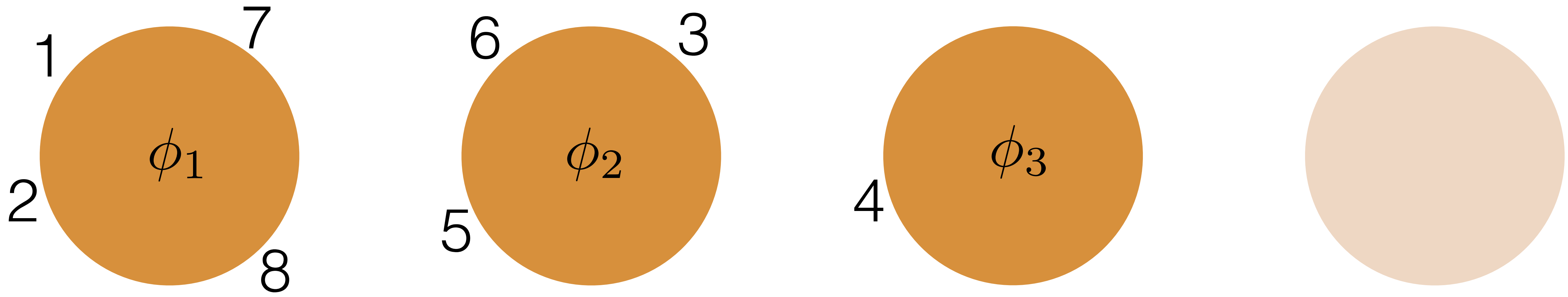
- Each customer walks into the restaurant

Chinese restaurant process



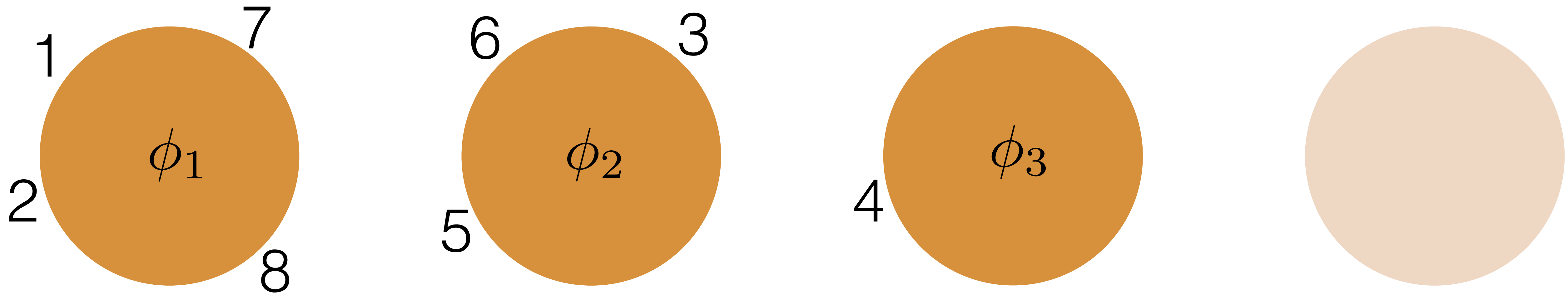
- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there

Chinese restaurant process



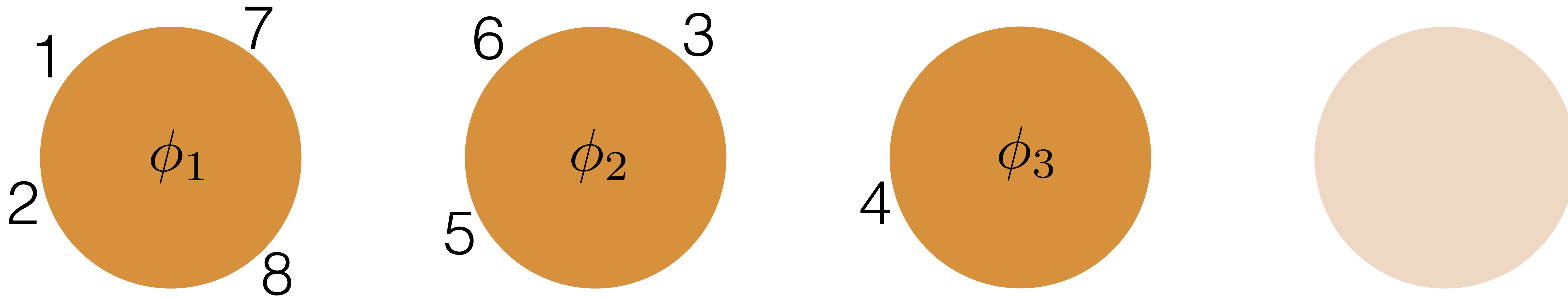
- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to α

Chinese restaurant process



- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to α
- Marginal for the Categorical likelihood with GEM prior

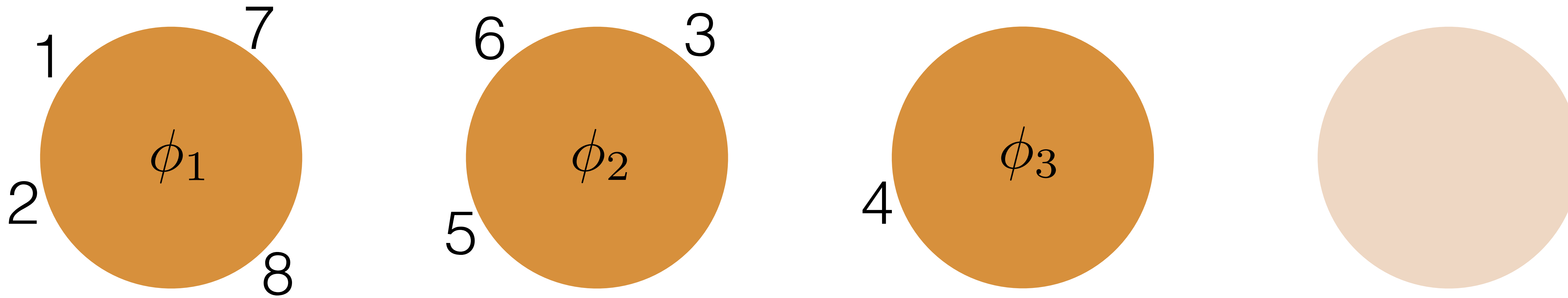
Chinese restaurant process



- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to α
- Marginal for the Categorical likelihood with GEM prior

$$z_1 = z_2 = z_7 = z_8 = 1, z_3 = z_5 = z_6 = 2, z_4 = 3$$

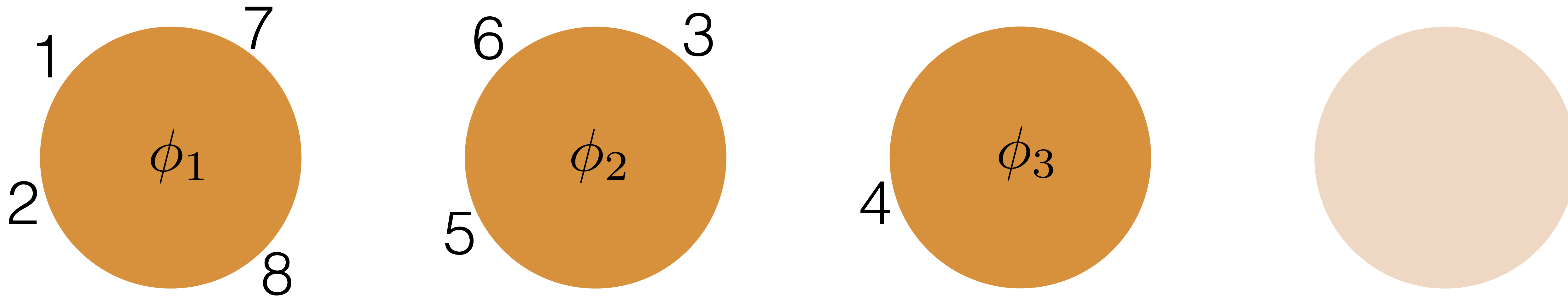
Chinese restaurant process



- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to α
- Marginal for the Categorical likelihood with GEM prior

$$z_1 = z_2 = z_7 = z_8 = 1, z_3 = z_5 = z_6 = 2, z_4 = 3$$
$$\Rightarrow \Pi_8 = \{\{1, 2, 7, 8\}, \{3, 5, 6\}, \{4\}\}$$

Chinese restaurant process

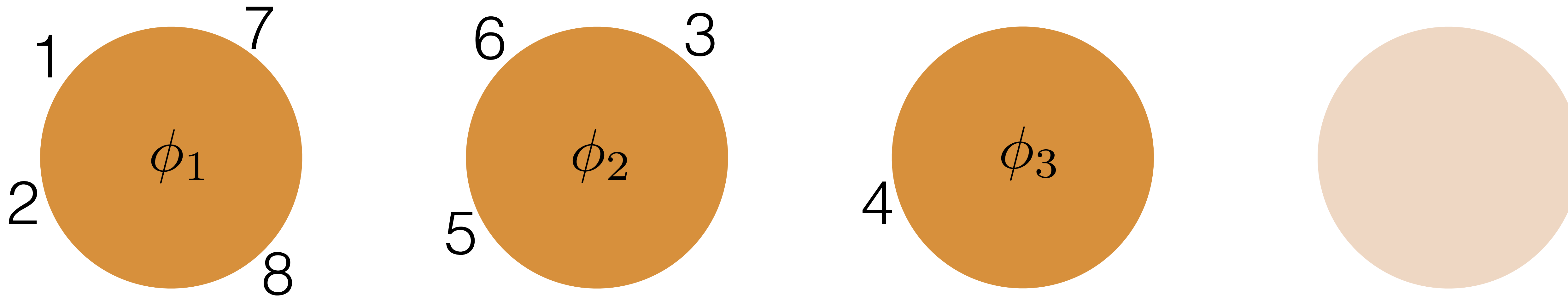


- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to α
- Marginal for the Categorical likelihood with GEM prior

$$z_1 = z_2 = z_7 = z_8 = 1, z_3 = z_5 = z_6 = 2, z_4 = 3$$
$$\Rightarrow \Pi_8 = \{\{1,2,7,8\},\{3,5,6\},\{4\}\}$$

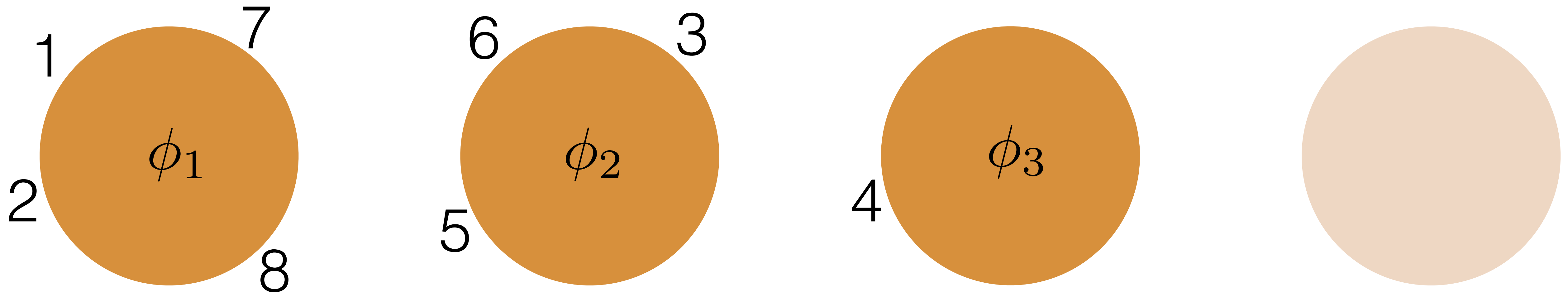
- *Partition of [8]*: set of mutually exclusive & exhaustive sets of $[8] := \{1, \ldots, 8\}$

Chinese restaurant process



- Probability of this seating:

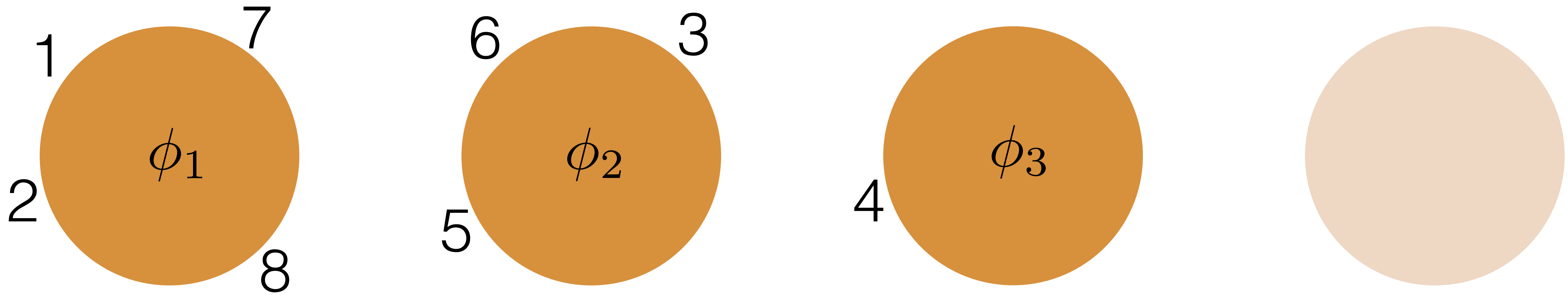
Chinese restaurant process



- Probability of this seating:

$$\frac{\alpha}{\alpha}$$

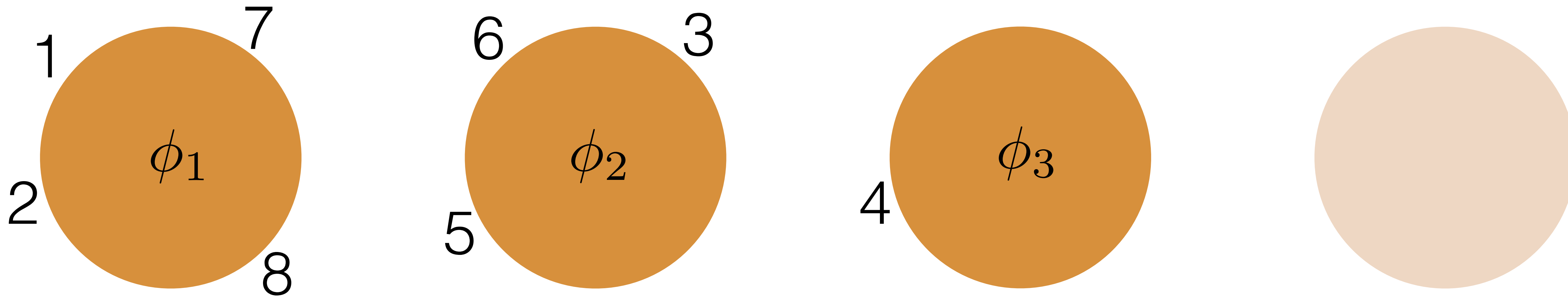
Chinese restaurant process



- Probability of this seating:

$$\frac{\alpha}{\alpha} \cdot \frac{1}{\alpha + 1}$$

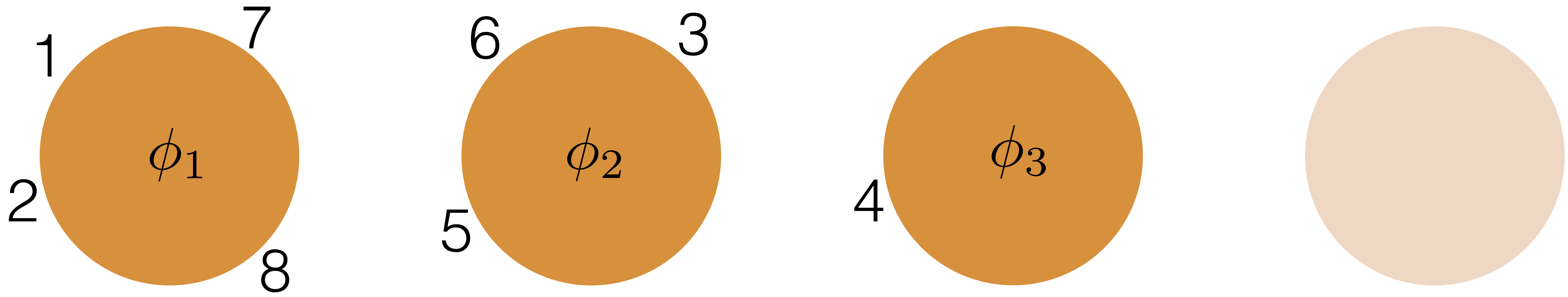
Chinese restaurant process



- Probability of this seating:

$$\frac{\alpha}{\alpha} \cdot \frac{1}{\alpha + 1} \cdot \frac{\alpha}{\alpha + 2}$$

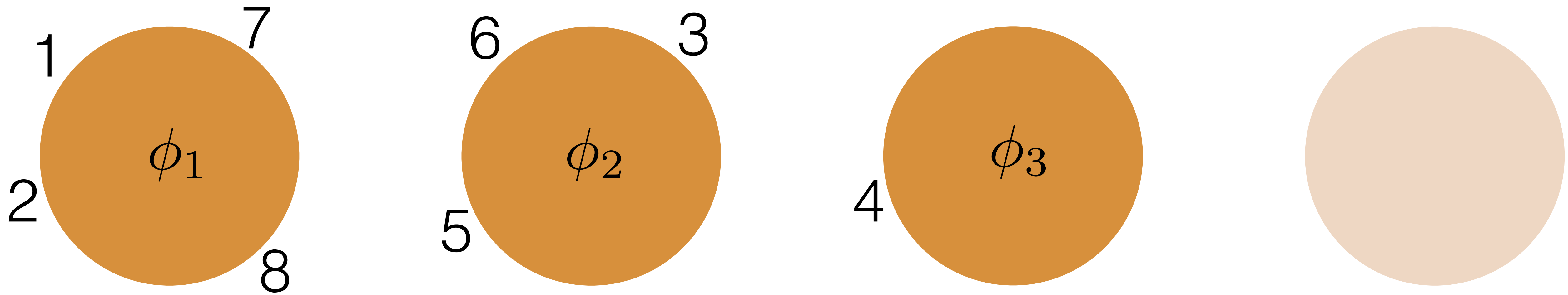
Chinese restaurant process



- Probability of this seating:

$$\frac{\alpha}{\alpha} \cdot \frac{1}{\alpha+1} \cdot \frac{\alpha}{\alpha+2} \cdot \frac{\alpha}{\alpha+3}$$

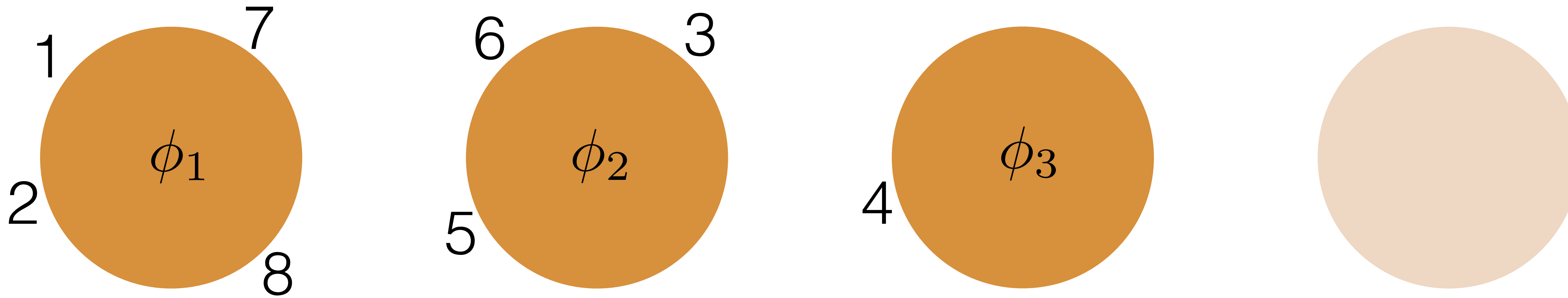
Chinese restaurant process



- Probability of this seating:

$$\frac{\alpha}{\alpha} \cdot \frac{1}{\alpha+1} \cdot \frac{\alpha}{\alpha+2} \cdot \frac{\alpha}{\alpha+3} \cdot \frac{1}{\alpha+4}$$

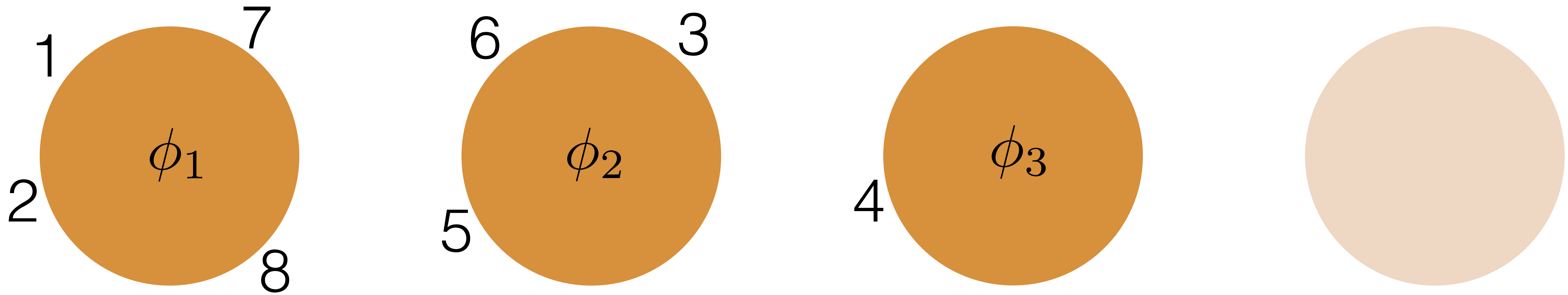
Chinese restaurant process



- Probability of this seating:

$$\frac{\alpha}{\alpha} \cdot \frac{1}{\alpha+1} \cdot \frac{\alpha}{\alpha+2} \cdot \frac{\alpha}{\alpha+3} \cdot \frac{1}{\alpha+4} \cdot \frac{2}{\alpha+5}$$

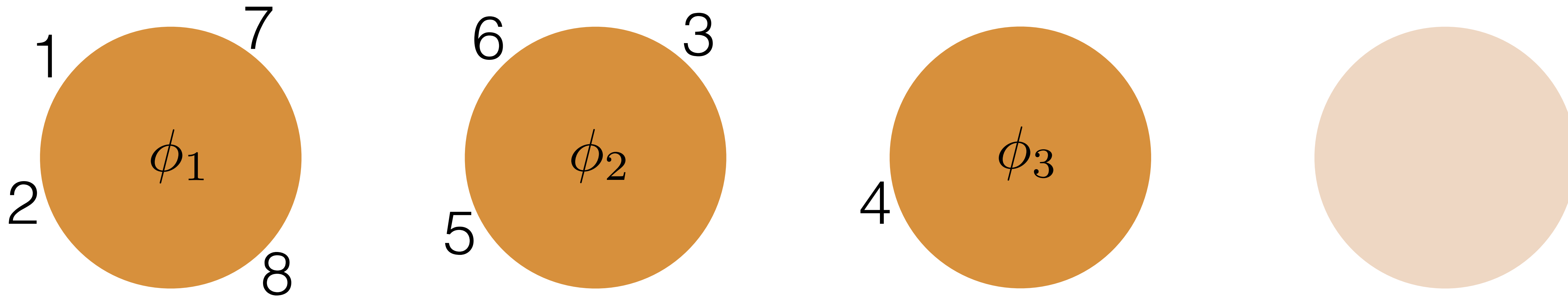
Chinese restaurant process



- Probability of this seating:

$$\frac{\alpha}{\alpha} \cdot \frac{1}{\alpha+1} \cdot \frac{\alpha}{\alpha+2} \cdot \frac{\alpha}{\alpha+3} \cdot \frac{1}{\alpha+4} \cdot \frac{2}{\alpha+5} \cdot \frac{2}{\alpha+6}$$

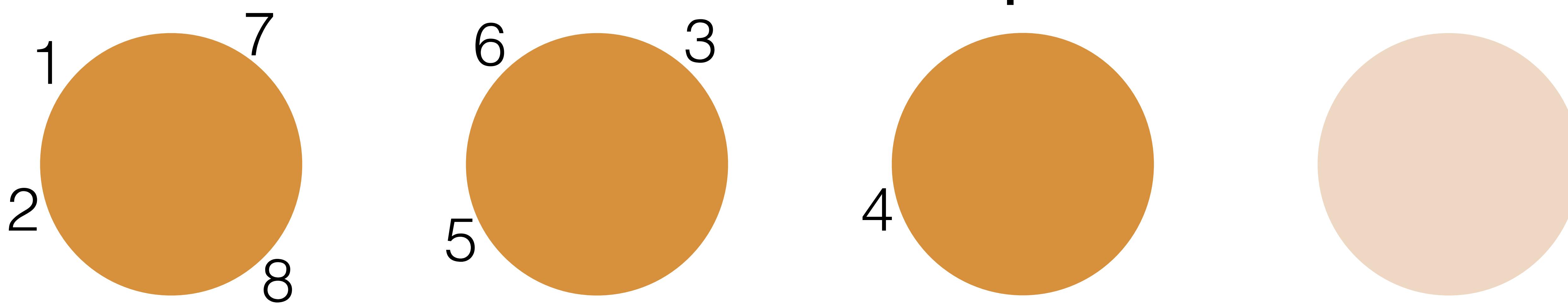
Chinese restaurant process



- Probability of this seating:

$$\frac{\alpha}{\alpha} \cdot \frac{1}{\alpha+1} \cdot \frac{\alpha}{\alpha+2} \cdot \frac{\alpha}{\alpha+3} \cdot \frac{1}{\alpha+4} \cdot \frac{2}{\alpha+5} \cdot \frac{2}{\alpha+6} \cdot \frac{3}{\alpha+7}$$

Chinese restaurant process

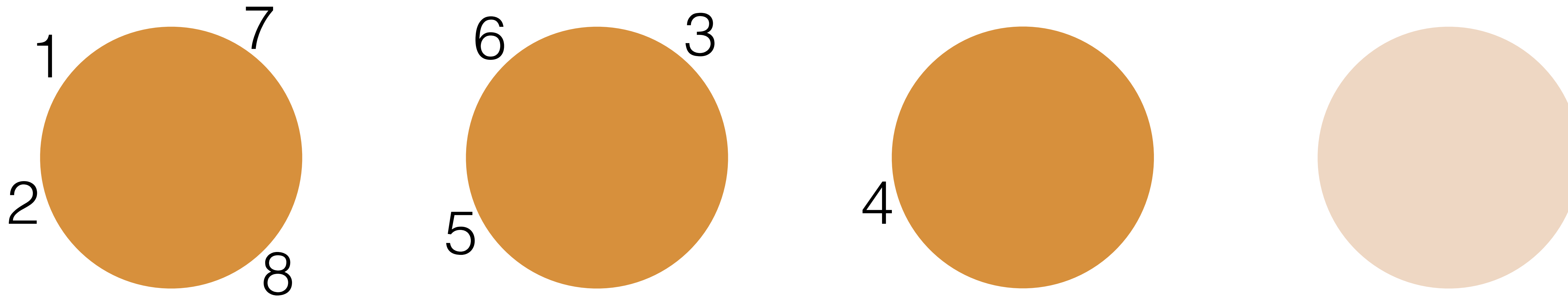


- Probability of this seating:

$$\frac{\alpha}{\alpha} \cdot \frac{1}{\alpha+1} \cdot \frac{\alpha}{\alpha+2} \cdot \frac{\alpha}{\alpha+3} \cdot \frac{1}{\alpha+4} \cdot \frac{2}{\alpha+5} \cdot \frac{2}{\alpha+6} \cdot \frac{3}{\alpha+7}$$

- Probability of *N* customers (K_N tables, n_k at table *k*):

Chinese restaurant process

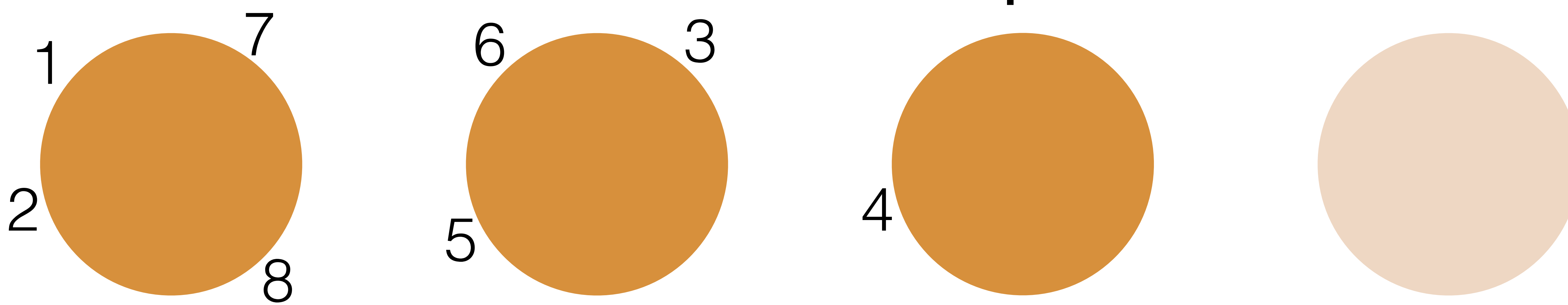


- Probability of this seating:

$$\frac{\alpha}{\alpha} \cdot \frac{1}{\alpha+1} \cdot \frac{\alpha}{\alpha+2} \cdot \frac{\alpha}{\alpha+3} \cdot \frac{1}{\alpha+4} \cdot \frac{2}{\alpha+5} \cdot \frac{2}{\alpha+6} \cdot \frac{3}{\alpha+7}$$

- Probability of *N* customers (K_N tables, n_k at table *k*):

Chinese restaurant process



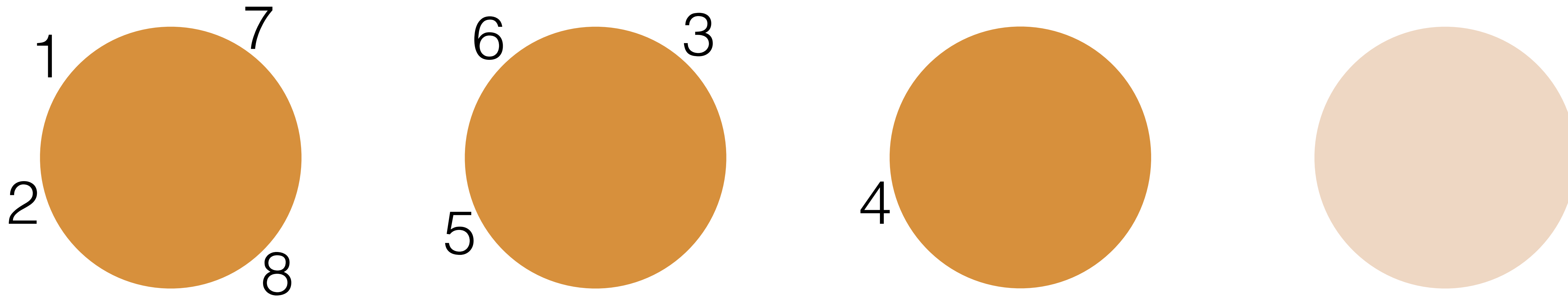
- Probability of this seating:

$$\frac{\alpha}{\alpha} \cdot \frac{1}{\alpha+1} \cdot \frac{\alpha}{\alpha+2} \cdot \frac{\alpha}{\alpha+3} \cdot \frac{1}{\alpha+4} \cdot \frac{2}{\alpha+5} \cdot \frac{2}{\alpha+6} \cdot \frac{3}{\alpha+7}$$

- Probability of N customers (K_N tables, n_k at table k):

$$\frac{}{\alpha \cdots (\alpha + N - 1)}$$

Chinese restaurant process



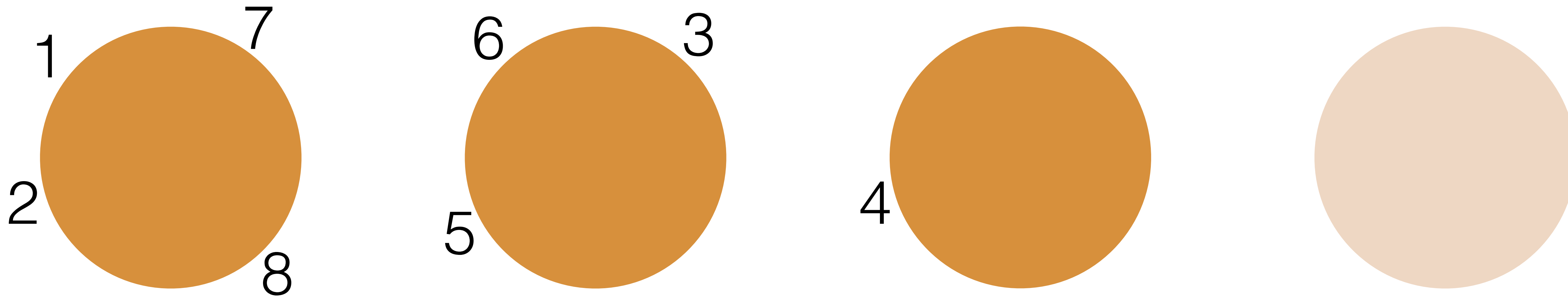
- Probability of this seating:

$$\frac{\alpha}{\alpha} \cdot \frac{1}{\alpha+1} \cdot \frac{\alpha}{\alpha+2} \cdot \frac{\alpha}{\alpha+3} \cdot \frac{1}{\alpha+4} \cdot \frac{2}{\alpha+5} \cdot \frac{2}{\alpha+6} \cdot \frac{3}{\alpha+7}$$

- Probability of *N* customers (K_N tables, n_k at table *k*):

$$\frac{\alpha^{K_N}}{\alpha \cdots (\alpha + N - 1)}$$

Chinese restaurant process



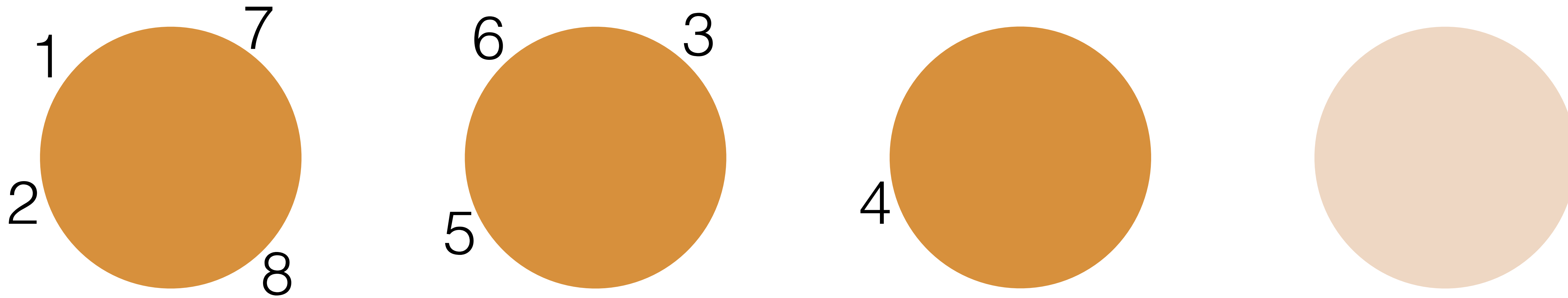
- Probability of this seating:

$$\frac{\alpha}{\alpha} \cdot \frac{1}{\alpha+1} \cdot \frac{\alpha}{\alpha+2} \cdot \frac{\alpha}{\alpha+3} \cdot \frac{1}{\alpha+4} \cdot \frac{2}{\alpha+5} \cdot \frac{2}{\alpha+6} \cdot \frac{3}{\alpha+7}$$

- Probability of *N* customers (K_N tables, n_k at table *k*):

$$\frac{\alpha^{K_N}}{\alpha \cdots (\alpha + N - 1)}$$

Chinese restaurant process



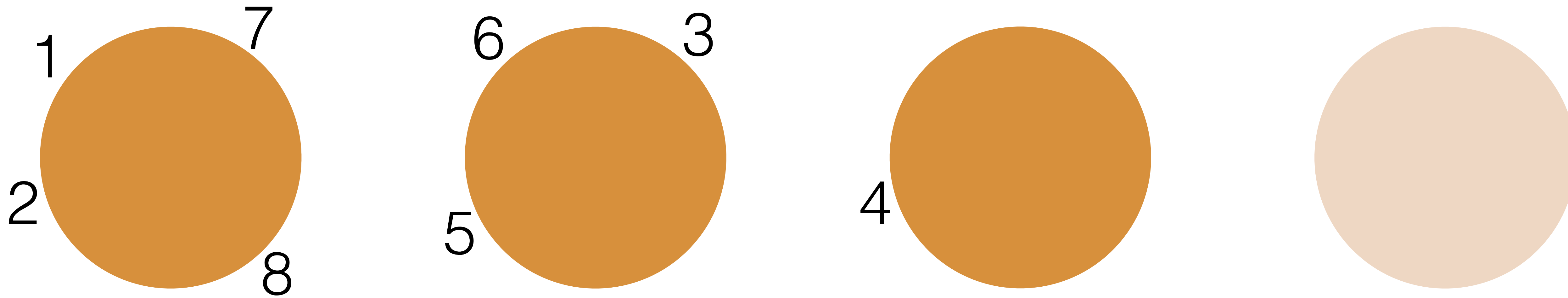
- Probability of this seating:

$$\frac{\alpha}{\alpha} \cdot \frac{1}{\alpha+1} \cdot \frac{\alpha}{\alpha+2} \cdot \frac{\alpha}{\alpha+3} \cdot \frac{1}{\alpha+4} \cdot \frac{2}{\alpha+5} \cdot \frac{2}{\alpha+6} \cdot \frac{3}{\alpha+7}$$

- Probability of *N* customers (K_N tables, n_k at table *k*):

$$\frac{\alpha^{K_N}}{\alpha \cdots (\alpha + N - 1)}$$

Chinese restaurant process



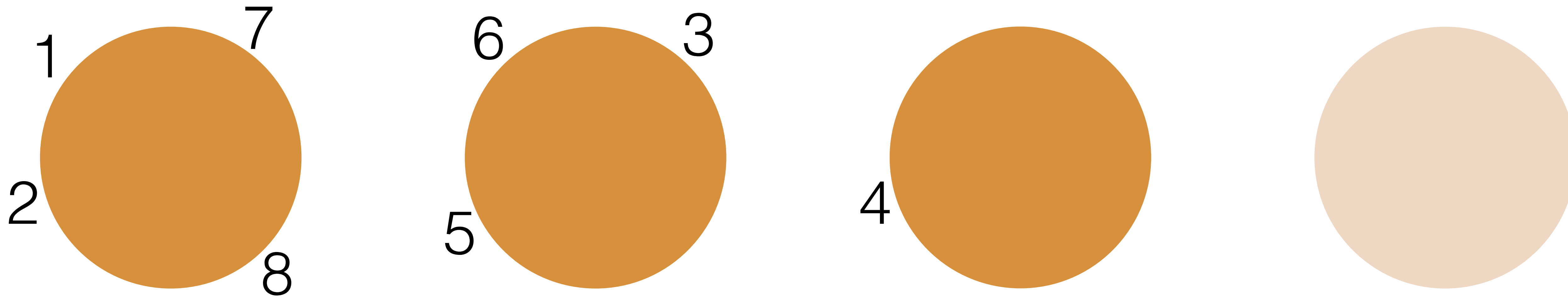
- Probability of this seating:

$$\frac{\alpha}{\alpha} \cdot \frac{1}{\alpha+1} \cdot \frac{\alpha}{\alpha+2} \cdot \frac{\alpha}{\alpha+3} \cdot \frac{1}{\alpha+4} \cdot \frac{2}{\alpha+5} \cdot \frac{2}{\alpha+6} \cdot \frac{3}{\alpha+7}$$

- Probability of *N* customers (K_N tables, n_k at table *k*):

$$\frac{\alpha^{K_N}}{\alpha \cdots (\alpha + N - 1)}$$

Chinese restaurant process



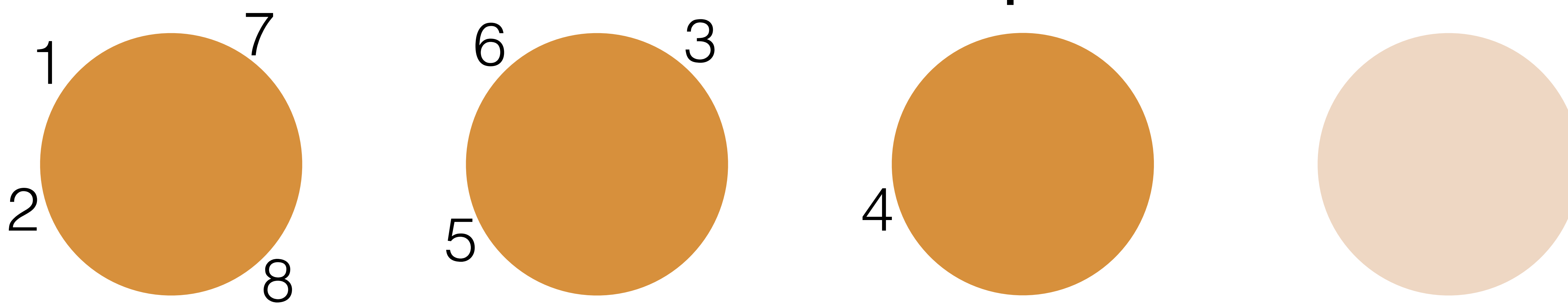
- Probability of this seating:

$$\frac{\alpha}{\alpha} \cdot \frac{1}{\alpha+1} \cdot \frac{\alpha}{\alpha+2} \cdot \frac{\alpha}{\alpha+3} \cdot \frac{1}{\alpha+4} \cdot \frac{2}{\alpha+5} \cdot \frac{2}{\alpha+6} \cdot \frac{3}{\alpha+7}$$

- Probability of *N* customers (K_N tables, n_k at table *k*):

$$\frac{\alpha^{K_N}}{\alpha \cdots (\alpha + N - 1)}$$

Chinese restaurant process



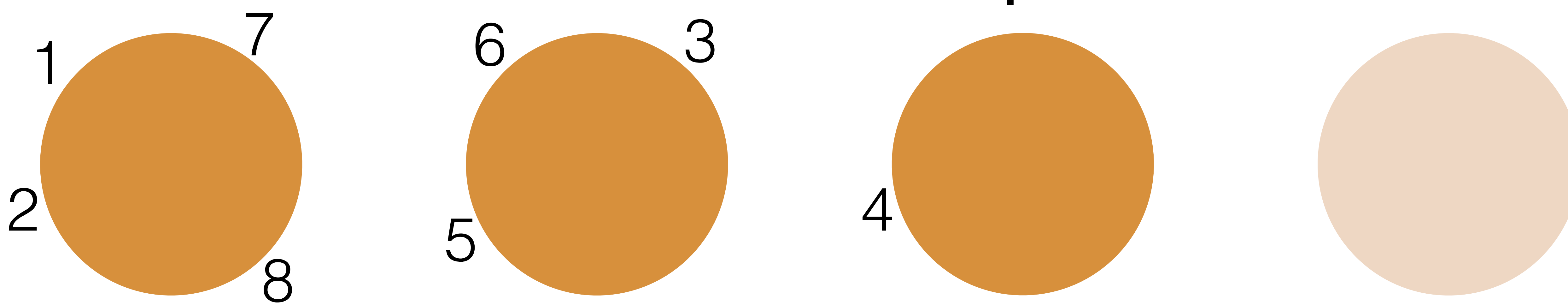
- Probability of this seating:

$$\frac{\alpha}{\alpha} \cdot \frac{1}{\alpha+1} \cdot \frac{\alpha}{\alpha+2} \cdot \frac{\alpha}{\alpha+3} \cdot \frac{1}{\alpha+4} \cdot \frac{2}{\alpha+5} \cdot \frac{2}{\alpha+6} \cdot \frac{3}{\alpha+7}$$

- Probability of N customers (K_N tables, n_k at table k):

$$\frac{\alpha^{K_N} \prod_{k=1}^{K_N} (n_k - 1)!}{\alpha \cdots (\alpha + N - 1)}$$

Chinese restaurant process



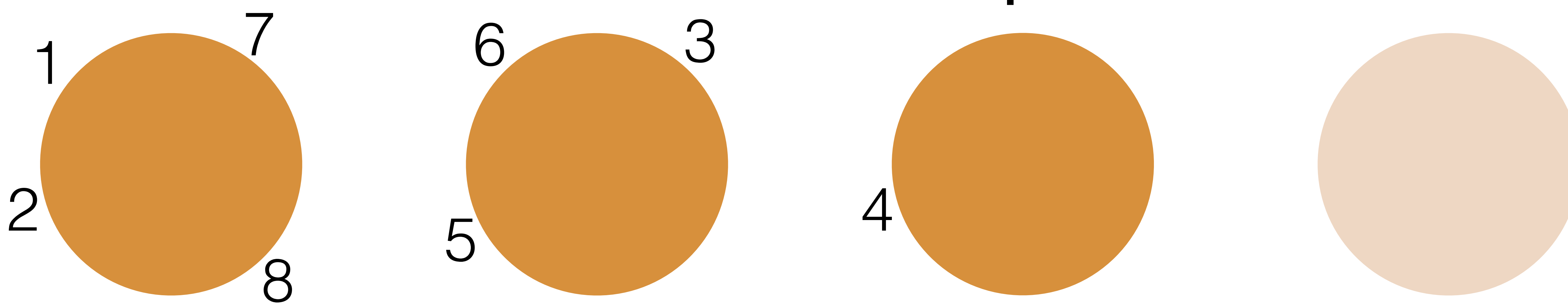
- Probability of this seating:

$$\frac{\alpha}{\alpha} \cdot \frac{1}{\alpha+1} \cdot \frac{\alpha}{\alpha+2} \cdot \frac{\alpha}{\alpha+3} \cdot \frac{1}{\alpha+4} \cdot \frac{2}{\alpha+5} \cdot \frac{2}{\alpha+6} \cdot \frac{3}{\alpha+7}$$

- Probability of *N* customers (K_N tables, #*C* at table *C*):

$$\frac{\alpha^{K_N} \prod_{C \in \Pi_N} (\#C - 1)!}{\alpha \cdots (\alpha + N - 1)}$$

Chinese restaurant process



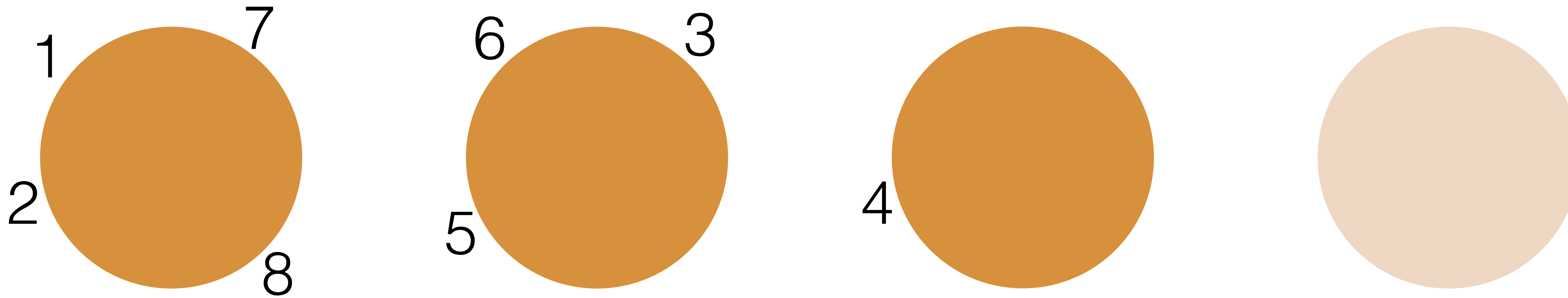
- Probability of this seating:

$$\frac{\alpha}{\alpha} \cdot \frac{1}{\alpha+1} \cdot \frac{\alpha}{\alpha+2} \cdot \frac{\alpha}{\alpha+3} \cdot \frac{1}{\alpha+4} \cdot \frac{2}{\alpha+5} \cdot \frac{2}{\alpha+6} \cdot \frac{3}{\alpha+7}$$

- Probability of N customers (K_N tables, $\#C$ at table C):

$$\frac{\alpha^{K_N} \prod_{C \in \Pi_N} (\#C - 1)!}{\alpha \cdots (\alpha + N - 1)} = \mathbb{P}(\Pi_N = \pi_N)$$

Chinese restaurant process



- Probability of this seating:

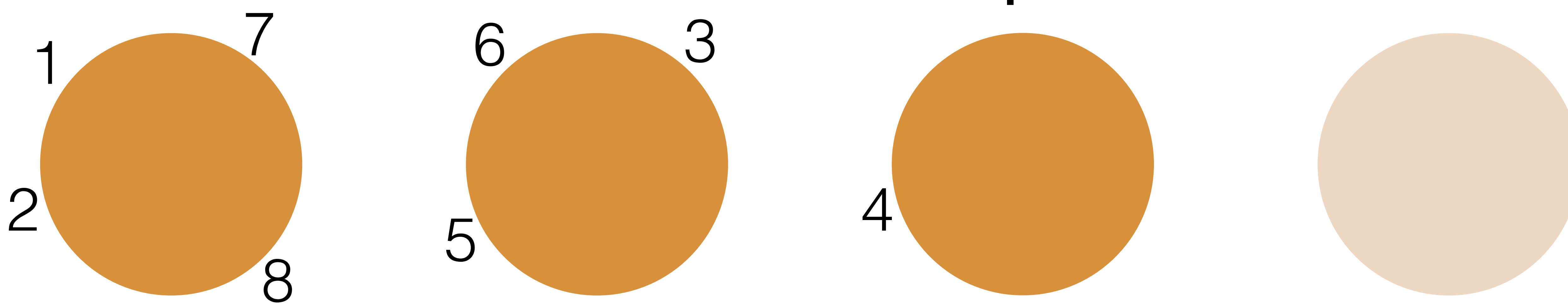
$$\frac{\alpha}{\alpha} \cdot \frac{1}{\alpha+1} \cdot \frac{\alpha}{\alpha+2} \cdot \frac{\alpha}{\alpha+3} \cdot \frac{1}{\alpha+4} \cdot \frac{2}{\alpha+5} \cdot \frac{2}{\alpha+6} \cdot \frac{3}{\alpha+7}$$

- Probability of *N* customers (K_N tables, #*C* at table *C*):

$$\frac{\alpha^{K_N} \prod_{C \in \Pi_N} (\#C - 1)!}{\alpha \cdots (\alpha + N - 1)} = \mathbb{P}(\Pi_N = \pi_N)$$

- Prob doesn't depend on customer order: *exchangeable*

Chinese restaurant process



- Probability of this seating:

$$\frac{\alpha}{\alpha} \cdot \frac{1}{\alpha+1} \cdot \frac{\alpha}{\alpha+2} \cdot \frac{\alpha}{\alpha+3} \cdot \frac{1}{\alpha+4} \cdot \frac{2}{\alpha+5} \cdot \frac{2}{\alpha+6} \cdot \frac{3}{\alpha+7}$$

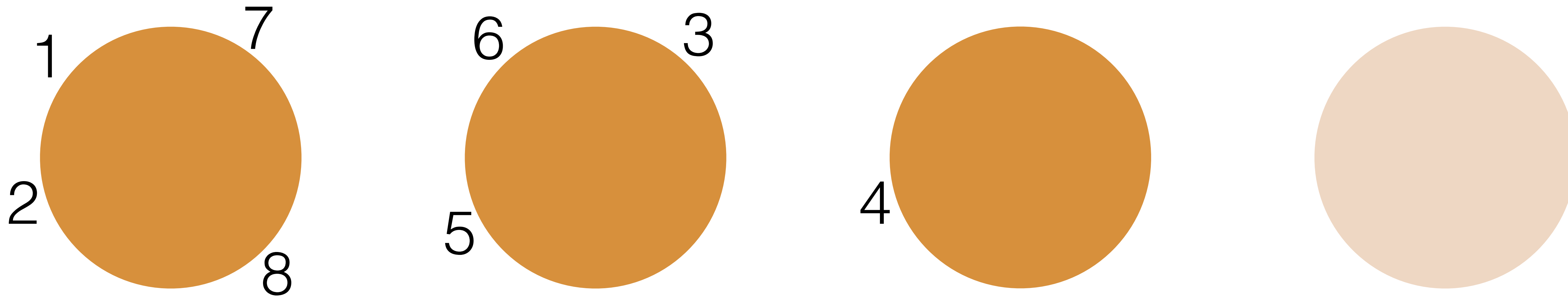
- Probability of *N* customers (K_N tables, #*C* at table *C*):

$$\frac{\alpha^{K_N} \prod_{C \in \Pi_N} (\#C - 1)!}{\alpha \cdots (\alpha + N - 1)} = \mathbb{P}(\Pi_N = \pi_N)$$

- Prob doesn't depend on customer order: *exchangeable*

$$\mathbb{P}(\Pi_8 = \{\{1,2,7,8\},\{3,5,6\},\{4\}\}) = \mathbb{P}(\Pi_8 = \{\{2,3,8,1\},\{4,6,7\},\{5\}\})$$

Chinese restaurant process



- Probability of this seating:

$$\frac{\alpha}{\alpha} \cdot \frac{1}{\alpha+1} \cdot \frac{\alpha}{\alpha+2} \cdot \frac{\alpha}{\alpha+3} \cdot \frac{1}{\alpha+4} \cdot \frac{2}{\alpha+5} \cdot \frac{2}{\alpha+6} \cdot \frac{3}{\alpha+7}$$

- Probability of N customers (K_N tables, #C at table C):

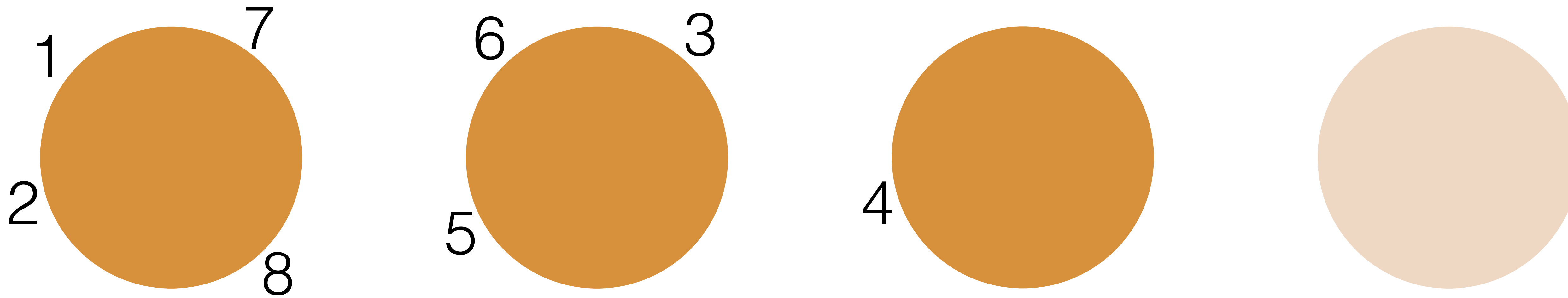
$$\frac{\alpha^{K_N} \prod_{C \in \Pi_N} (\#C - 1)!}{\alpha \cdots (\alpha + N - 1)} = \mathbb{P}(\Pi_N = \pi_N)$$

- Prob doesn't depend on customer order: *exchangeable*

$$\mathbb{P}(\Pi_8 = \{\{1,2,7,8\},\{3,5,6\},\{4\}\}) = \mathbb{P}(\Pi_8 = \{\{2,3,8,1\},\{4,6,7\},\{5\}\})$$

- Can always pretend n is the last customer and calculate $p(\Pi_N | \Pi_{N,\text{-}n})$

Chinese restaurant process



- Probability of this seating:

$$\frac{\alpha}{\alpha} \cdot \frac{1}{\alpha+1} \cdot \frac{\alpha}{\alpha+2} \cdot \frac{\alpha}{\alpha+3} \cdot \frac{1}{\alpha+4} \cdot \frac{2}{\alpha+5} \cdot \frac{2}{\alpha+6} \cdot \frac{3}{\alpha+7}$$

- Probability of *N* customers (K_N tables, #*C* at table *C*):

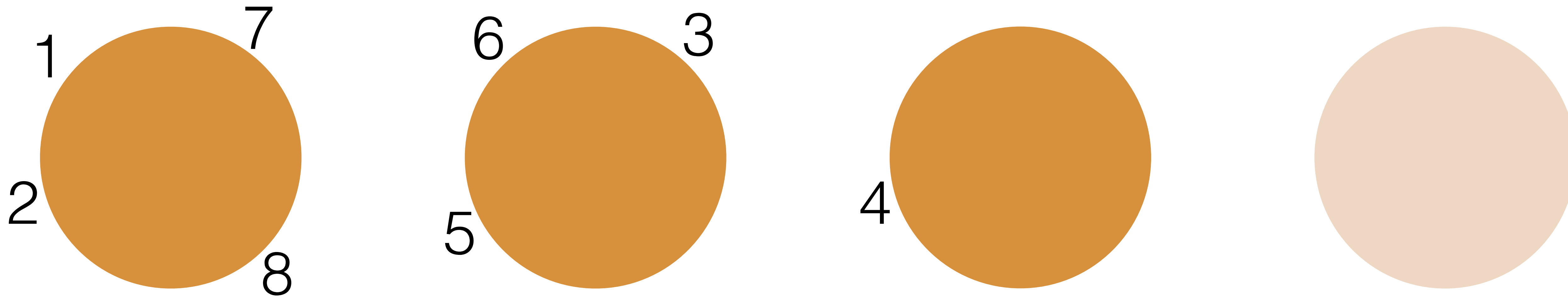
$$\frac{\alpha^{K_N} \prod_{C \in \Pi_N} (\#C - 1)!}{\alpha \cdots (\alpha + N - 1)} = \mathbb{P}(\Pi_N = \pi_N)$$

- Prob doesn't depend on customer order: *exchangeable*

$$\mathbb{P}(\Pi_8 = \{\{1,2,7,8\},\{3,5,6\},\{4\}\}) = \mathbb{P}(\Pi_8 = \{\{2,3,8,1\},\{4,6,7\},\{5\}\})$$

- Can always pretend *n* is the last customer and calculate $p(\Pi_N | \Pi_{N,\text{-}n})$
  - e.g. $\Pi_{8,\text{-}5} = \{\{1,2,7,8\},\{3,6\},\{4\}\}$

Chinese restaurant process



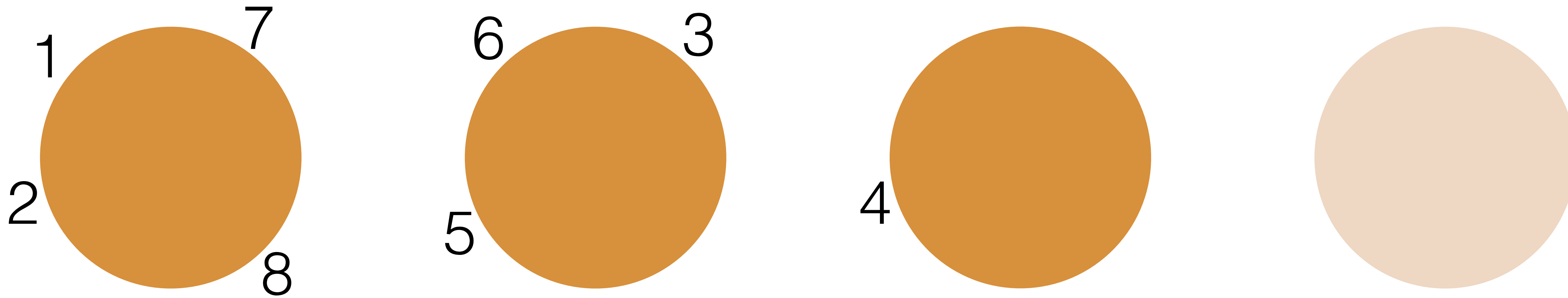
- Probability of N customers (K_N tables, $\#C$ at table C):

$$\frac{\alpha^{K_N}\prod_{C\in\Pi_N}(\#C-1)!}{\alpha\cdots(\alpha+N-1)} = \mathbb{P}(\Pi_N = \pi_N)$$

- So:

$$p(\Pi_N|\Pi_{N,-n}) =$$

Chinese restaurant process



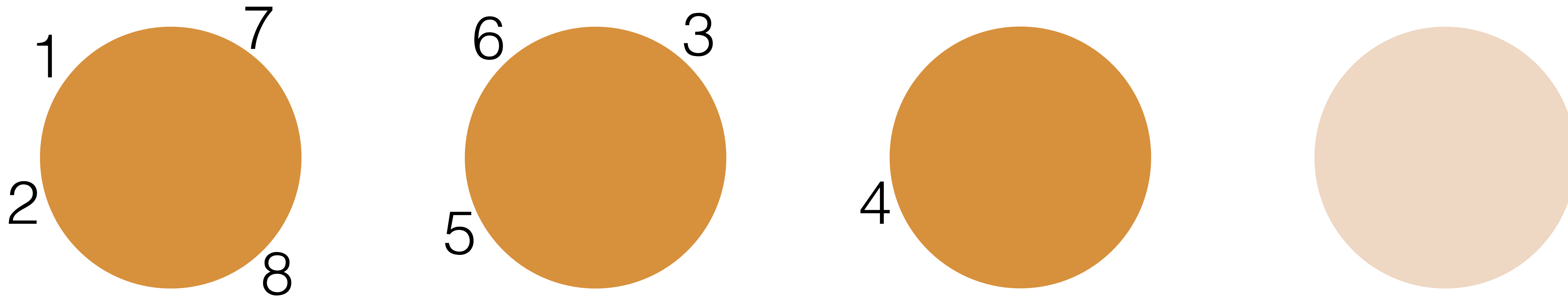
- Probability of N customers (K_N tables, $\#C$ at table C):

$$\frac{\alpha^{K_N} \prod_{C \in \Pi_N} (\#C - 1)!}{\alpha \cdots (\alpha + N - 1)} = \mathbb{P}(\Pi_N = \pi_N)$$

- So:

$$p(\Pi_N | \Pi_{N,-n}) = \left\{ \right.$$

Chinese restaurant process



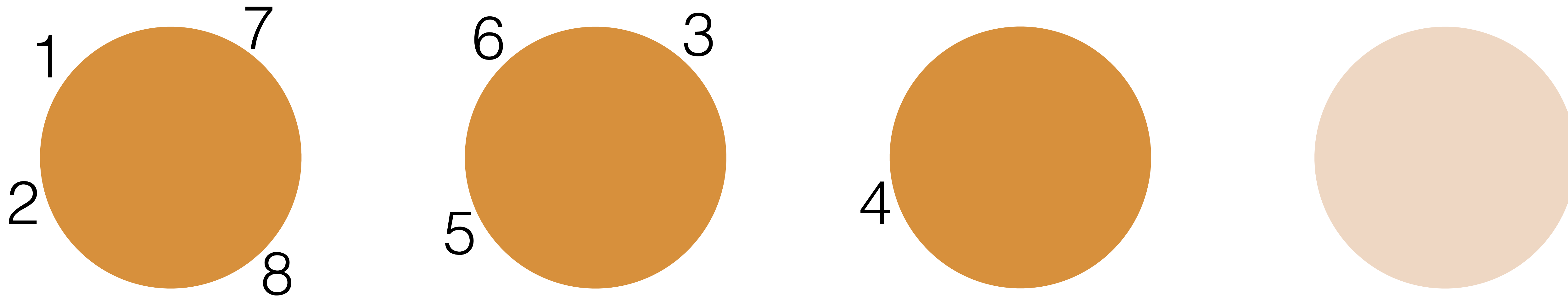
- Probability of N customers (K_N tables, $\#C$ at table C):

$$\frac{\alpha^{K_N} \prod_{C \in \Pi_N} (\#C - 1)!}{\alpha \cdots (\alpha + N - 1)} = \mathbb{P}(\Pi_N = \pi_N)$$

- So:

$$p(\Pi_N | \Pi_{N,-n}) = \begin{cases} & \text{if } n \text{ joins cluster } C \\ & \text{if } n \text{ starts a new cluster} \end{cases}$$

Chinese restaurant process



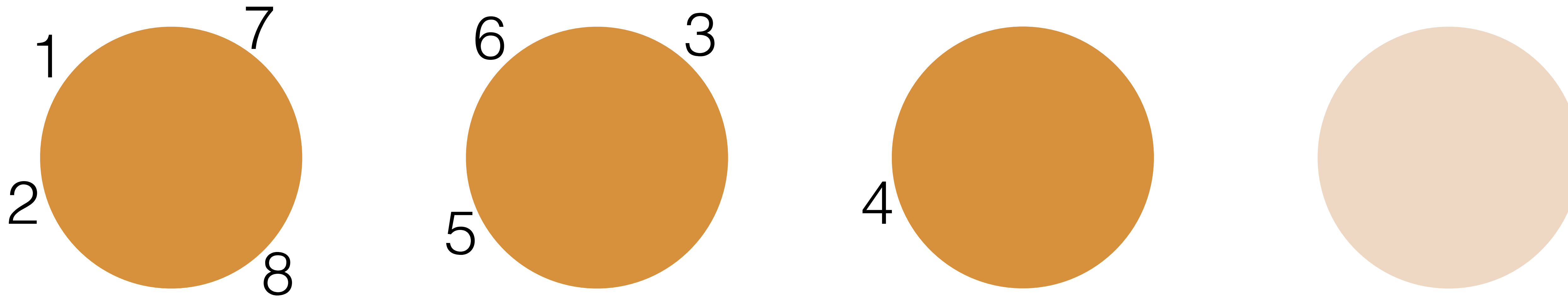
- Probability of N customers (K_N tables, #C at table C):

$$\frac{\alpha^{K_N} \prod_{C \in \Pi_N} (\#C - 1)!}{\alpha \cdots (\alpha + N - 1)} = \mathbb{P}(\Pi_N = \pi_N)$$

- So:

$$p(\Pi_N | \Pi_{N,-n}) = \begin{cases} \frac{\#C}{\alpha + N - 1} & \text{if } n \text{ joins cluster } C \\ & \text{if } n \text{ starts a new cluster} \end{cases}$$

Chinese restaurant process



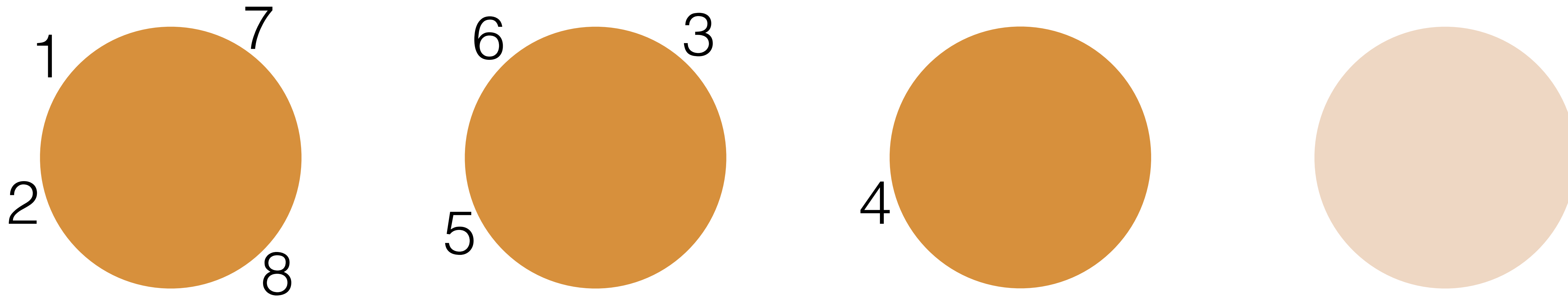
- Probability of N customers (K_N tables, #C at table C):

$$\frac{\alpha^{K_N} \prod_{C \in \Pi_N} (\#C - 1)!}{\alpha \cdots (\alpha + N - 1)} = \mathbb{P}(\Pi_N = \pi_N)$$

- So:

$$p(\Pi_N | \Pi_{N,-n}) = \begin{cases} \frac{\#C}{\alpha + N - 1} & \text{if } n \text{ joins cluster } C \\ \frac{\alpha}{\alpha + N - 1} & \text{if } n \text{ starts a new cluster} \end{cases}$$

Chinese restaurant process



- Probability of N customers (K_N tables, $\#C$ at table C):

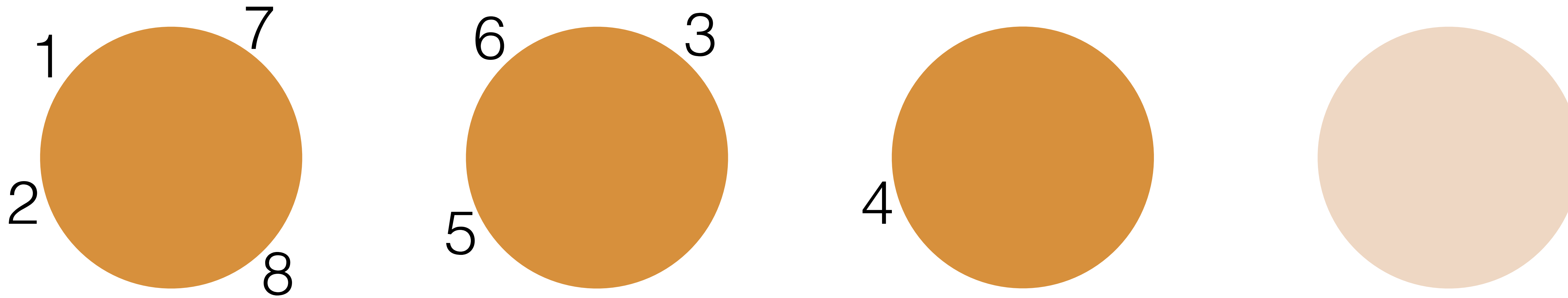
$$\frac{\alpha^{K_N} \prod_{C \in \Pi_N} (\#C - 1)!}{\alpha \cdots (\alpha + N - 1)} = \mathbb{P}(\Pi_N = \pi_N)$$

- So:

$$p(\Pi_N | \Pi_{N,-n}) = \begin{cases} \frac{\#C}{\alpha+N-1} & \text{if } n \text{ joins cluster } C \\ \frac{\alpha}{\alpha+N-1} & \text{if } n \text{ starts a new cluster} \end{cases}$$

- Gibbs sampling review:

Chinese restaurant process



- Probability of *N* customers (K_N tables, #*C* at table *C*):

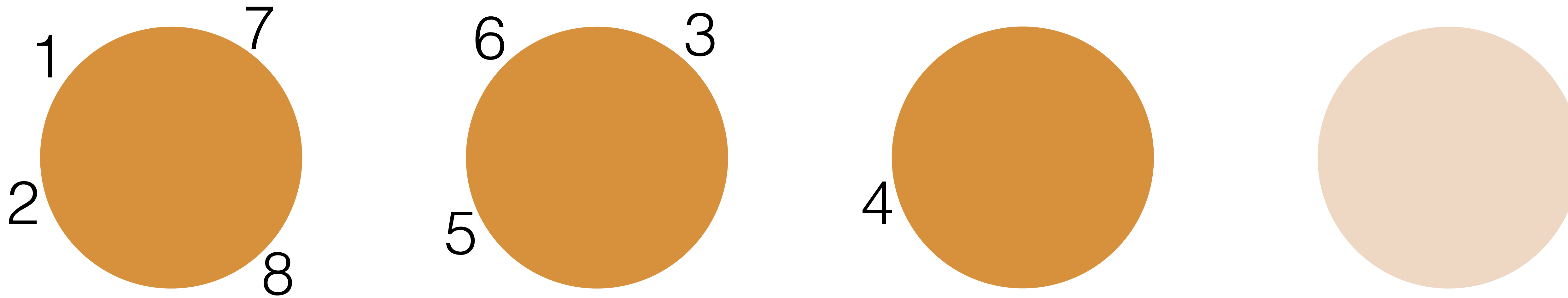
$$\frac{\alpha^{K_N} \prod_{C \in \Pi_N} (\#C - 1)!}{\alpha \cdots (\alpha + N - 1)} = \mathbb{P}(\Pi_N = \pi_N)$$

- So:

$$p(\Pi_N | \Pi_{N,-n}) = \begin{cases} \frac{\#C}{\alpha + N - 1} & \text{if } n \text{ joins cluster } C \\ \frac{\alpha}{\alpha + N - 1} & \text{if } n \text{ starts a new cluster} \end{cases}$$

- Gibbs sampling review: target distribution $p(v_1, v_2, v_3)$

Chinese restaurant process



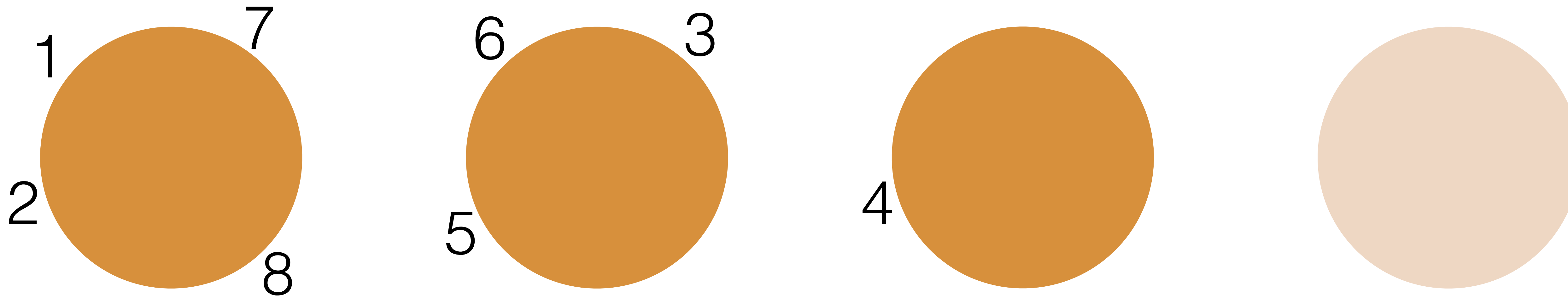
- Probability of N customers (K_N tables, #C at table C):

$$\frac{\alpha^{K_N} \prod_{C \in \Pi_N} (\#C - 1)!}{\alpha \cdots (\alpha + N - 1)} = \mathbb{P}(\Pi_N = \pi_N)$$

- So: $$p(\Pi_N | \Pi_{N,-n}) = \begin{cases} \frac{\#C}{\alpha + N - 1} & \text{if } n \text{ joins cluster } C \\ \frac{\alpha}{\alpha + N - 1} & \text{if } n \text{ starts a new cluster} \end{cases}$$

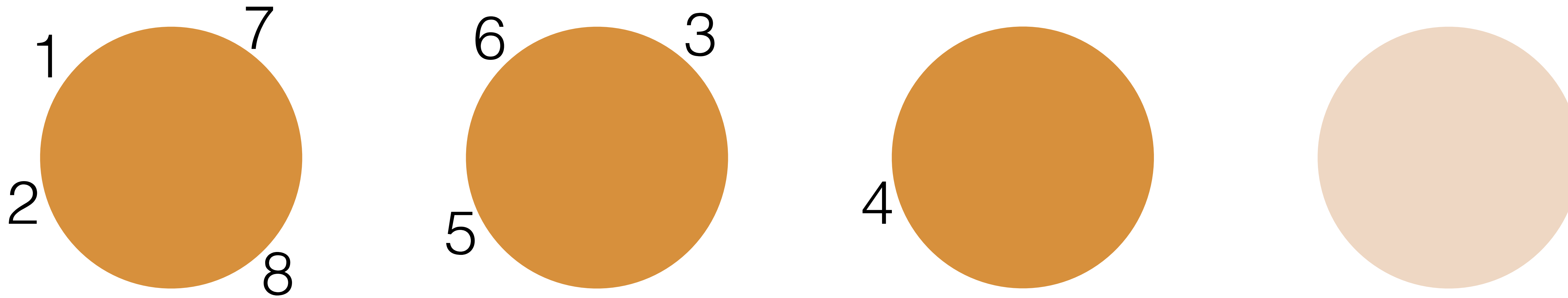
- Gibbs sampling review: target distribution $p(v_1, v_2, v_3)$
  - Start: $v_1^{(0)}, v_2^{(0)}, v_3^{(0)}$

Chinese restaurant process



- Probability of N customers (K_N tables, $\#C$ at table C):
$$\frac{\alpha^{K_N} \prod_{C \in \Pi_N} (\#C - 1)!}{\alpha \cdots (\alpha + N - 1)} = \mathbb{P}(\Pi_N = \pi_N)$$
- So: $$p(\Pi_N | \Pi_{N,-n}) = \begin{cases} \frac{\#C}{\alpha + N - 1} & \text{if } n \text{ joins cluster } C \\ \frac{\alpha}{\alpha + N - 1} & \text{if } n \text{ starts a new cluster} \end{cases}$$
- Gibbs sampling review: target distribution $p(v_1, v_2, v_3)$
  - Start: $v_1^{(0)}, v_2^{(0)}, v_3^{(0)}$
  - t th step: $v_1^{(t)} \sim p(v_1 | v_2^{(t-1)}, v_3^{(t-1)})$

Chinese restaurant process



- Probability of N customers (K_N tables, $\#C$ at table C):

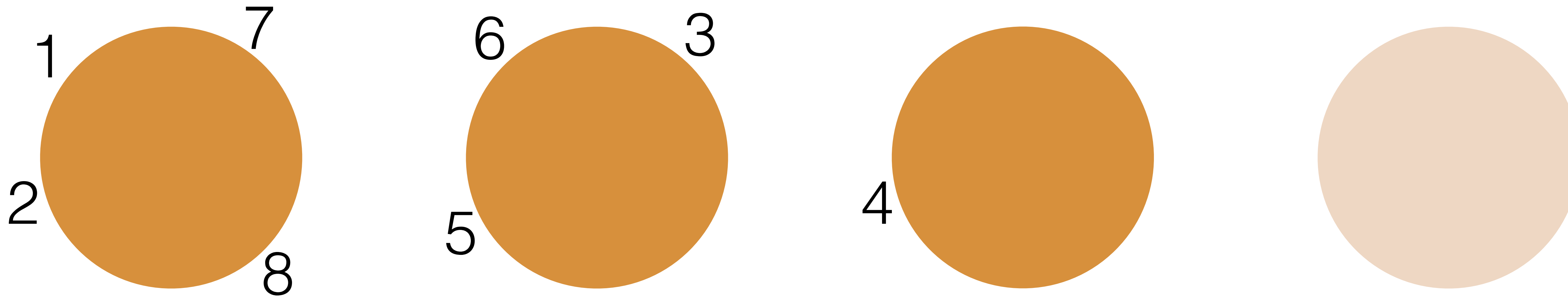
$$\frac{\alpha^{K_N} \prod_{C \in \Pi_N} (\#C - 1)!}{\alpha \cdots (\alpha + N - 1)} = \mathbb{P}(\Pi_N = \pi_N)$$

- So: $$p(\Pi_N | \Pi_{N,-n}) = \begin{cases} \frac{\#C}{\alpha+N-1} & \text{if } n \text{ joins cluster } C \\ \frac{\alpha}{\alpha+N-1} & \text{if } n \text{ starts a new cluster} \end{cases}$$

- Gibbs sampling review: target distribution $p(v_1, v_2, v_3)$
  - Start: $v_1^{(0)}, v_2^{(0)}, v_3^{(0)}$
  - t th step: $v_1^{(t)} \sim p(v_1 | v_2^{(t-1)}, v_3^{(t-1)})$

$$v_2^{(t)} \sim p(v_2 | v_1^{(t)}, v_3^{(t-1)})$$

Chinese restaurant process



- Probability of N customers (K_N tables, $\#C$ at table C):

$$\frac{\alpha^{K_N} \prod_{C \in \Pi_N} (\#C - 1)!}{\alpha \cdots (\alpha + N - 1)} = \mathbb{P}(\Pi_N = \pi_N)$$

- So: $$p(\Pi_N | \Pi_{N,-n}) = \begin{cases} \frac{\#C}{\alpha + N - 1} & \text{if } n \text{ joins cluster } C \\ \frac{\alpha}{\alpha + N - 1} & \text{if } n \text{ starts a new cluster} \end{cases}$$

- Gibbs sampling review: target distribution $p(v_1, v_2, v_3)$
  - Start: $v_1^{(0)}, v_2^{(0)}, v_3^{(0)}$
  - t th step: $v_1^{(t)} \sim p(v_1 | v_2^{(t-1)}, v_3^{(t-1)})$

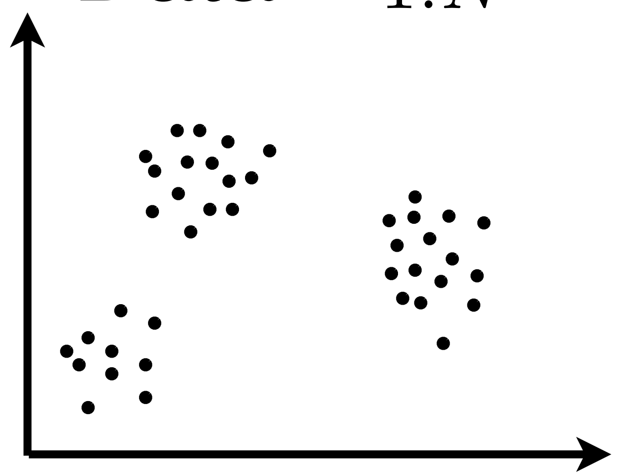
    $v_2^{(t)} \sim p(v_2 | v_1^{(t)}, v_3^{(t-1)})$

    $v_3^{(t)} \sim p(v_3 | v_1^{(t)}, v_2^{(t)})$

CRP mixture model: inference

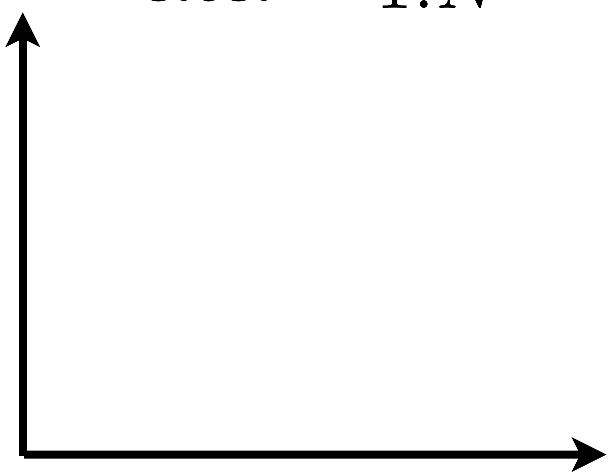
CRP mixture model: inference

- Data $x_{1:N}$



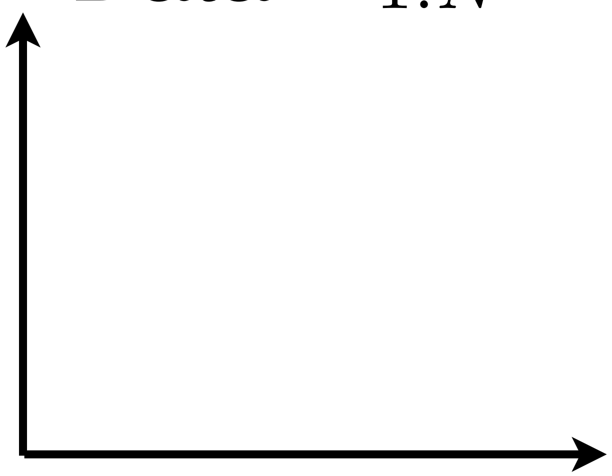
CRP mixture model: inference

- Data $x_{1:N}$



CRP mixture model: inference

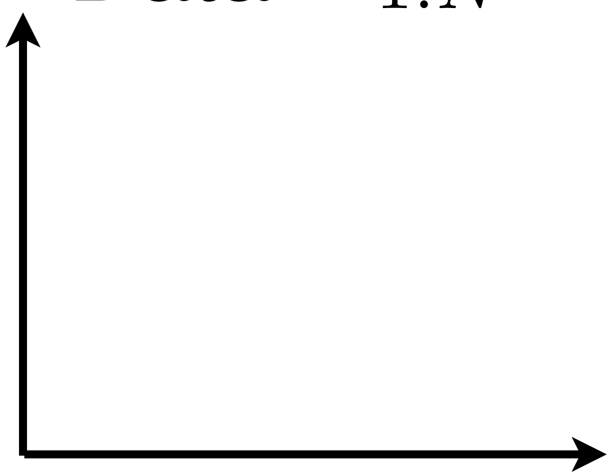
- Data $x_{1:N}$



- Generative model

CRP mixture model: inference

- Data $x_{1:N}$

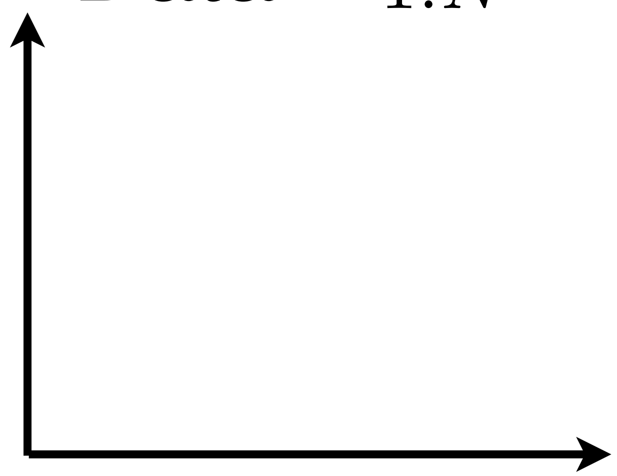


- Generative model

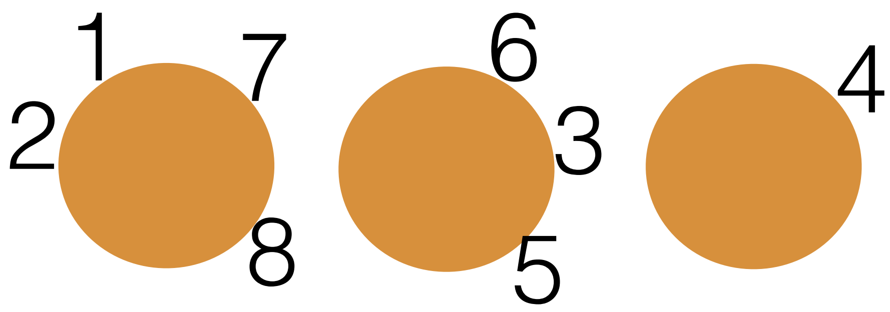
$$\Pi_N \sim \mathrm{CRP}(N, \alpha)$$

CRP mixture model: inference

- Data $x_{1:N}$

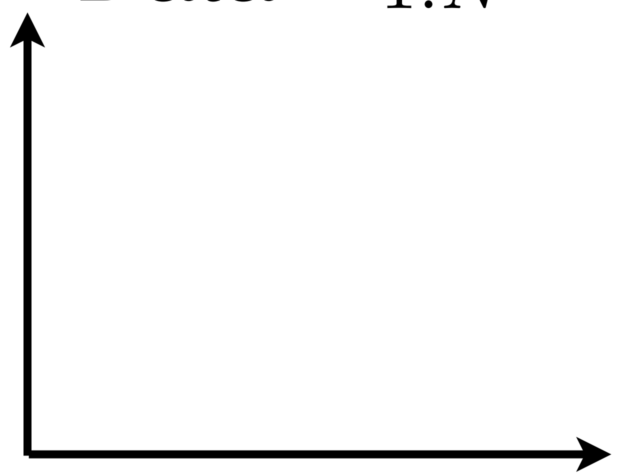


- Generative model
  $\Pi_N \sim \mathrm{CRP}(N, \alpha)$

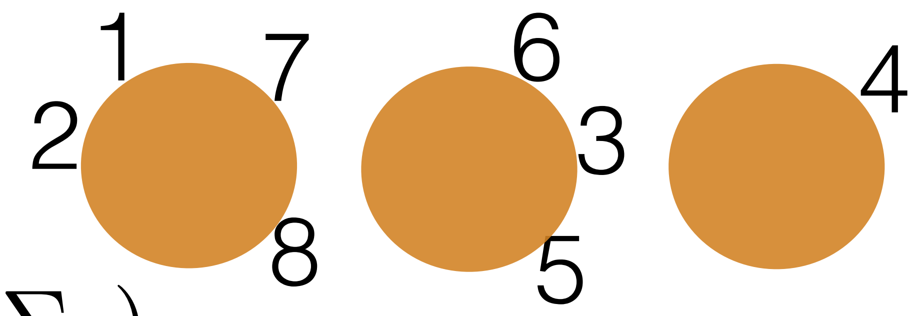


CRP mixture model: inference

- Data $x_{1:N}$

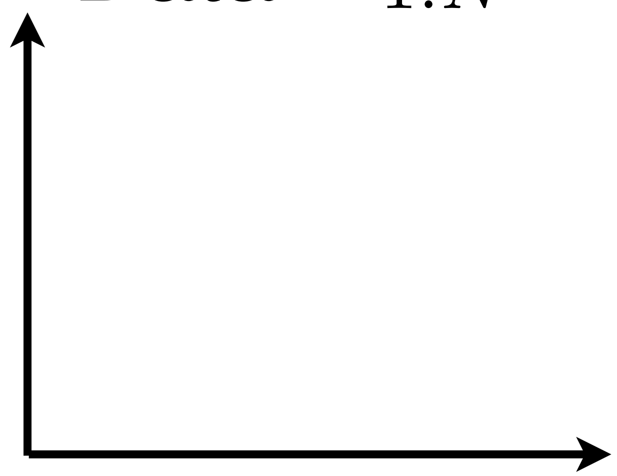


- Generative model
  $\Pi_N \sim \mathrm{CRP}(N, \alpha)$
  $\forall C \in \Pi_N, \mu_C \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0)$

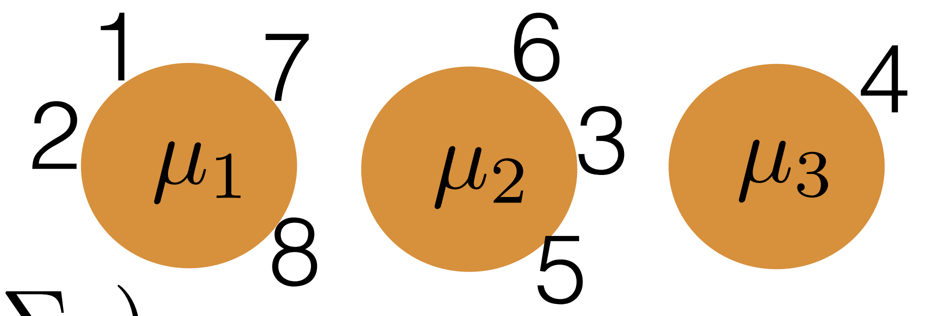


CRP mixture model: inference

- Data $x_{1:N}$

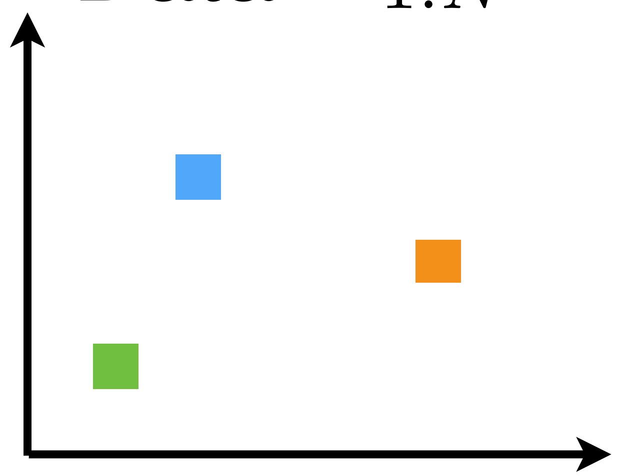


- Generative model
  $\Pi_N \sim \mathrm{CRP}(N, \alpha)$
  $\forall C \in \Pi_N, \mu_C \stackrel{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0)$



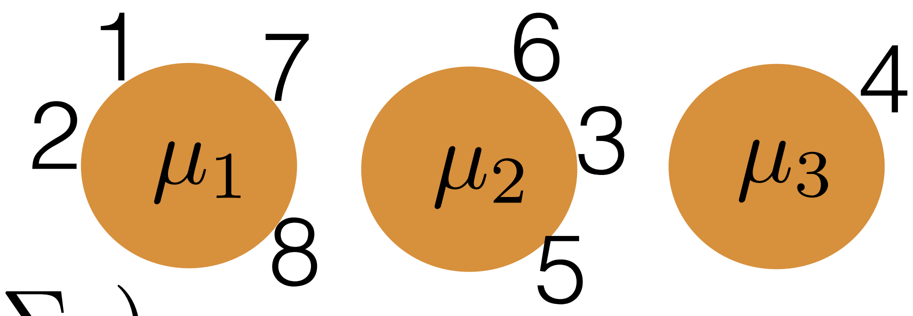
CRP mixture model: inference

- Data $x_{1:N}$



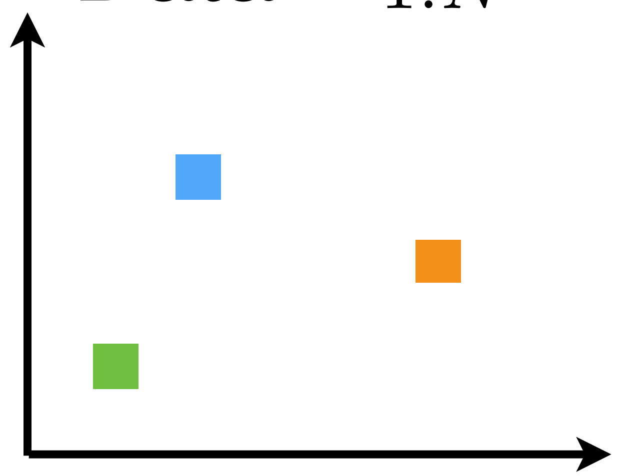
- Generative model

$$\Pi_N \sim \mathrm{CRP}(N, \alpha)$$
$$\forall C \in \Pi_N, \mu_C \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0)$$



CRP mixture model: inference

- Data $x_{1:N}$

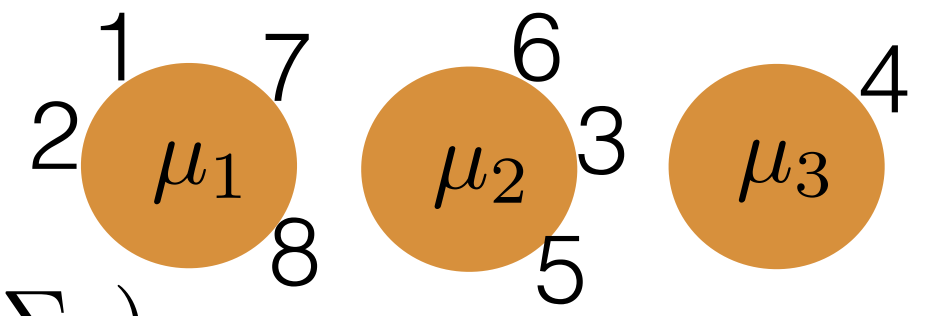


- Generative model

$$\Pi_N \sim \mathrm{CRP}(N, \alpha)$$

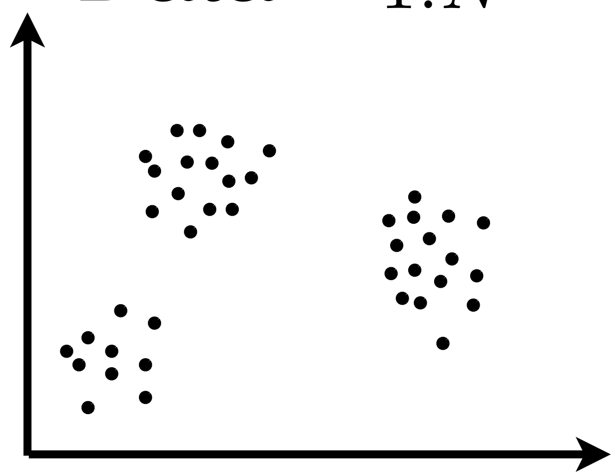
$$\forall C \in \Pi_N, \mu_C \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0)$$

$$\forall C \in \Pi_N, \forall n \in C, x_n \overset{indep}{\sim} \mathcal{N}(\mu_C, \Sigma)$$



CRP mixture model: inference

- Data $x_{1:N}$

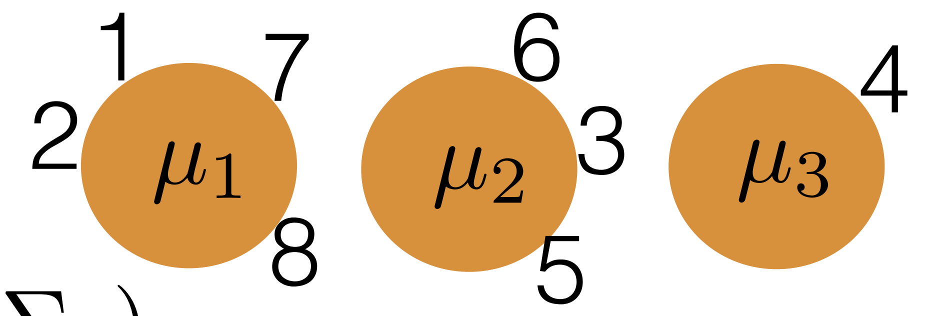


- Generative model

$$\Pi_N \sim \mathrm{CRP}(N, \alpha)$$

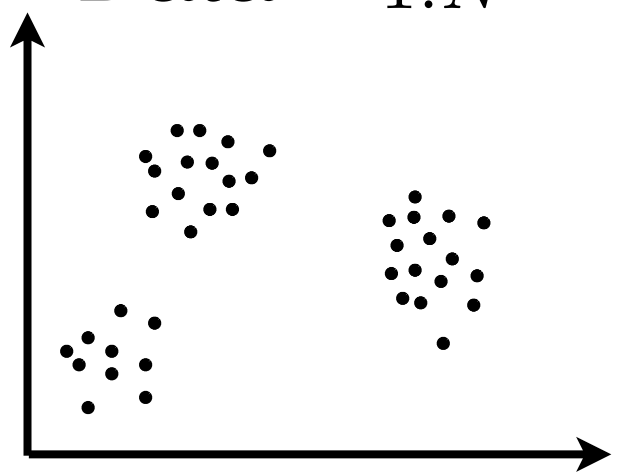
$$\forall C \in \Pi_N, \mu_C \stackrel{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0)$$

$$\forall C \in \Pi_N, \forall n \in C, x_n \stackrel{indep}{\sim} \mathcal{N}(\mu_C, \Sigma)$$



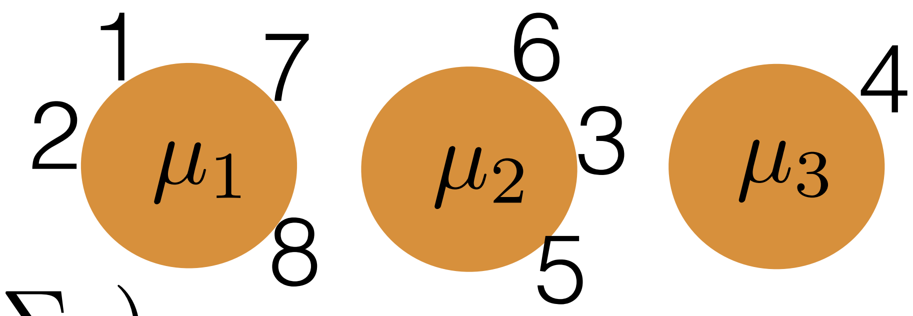
CRP mixture model: inference

- Data $x_{1:N}$



- Generative model

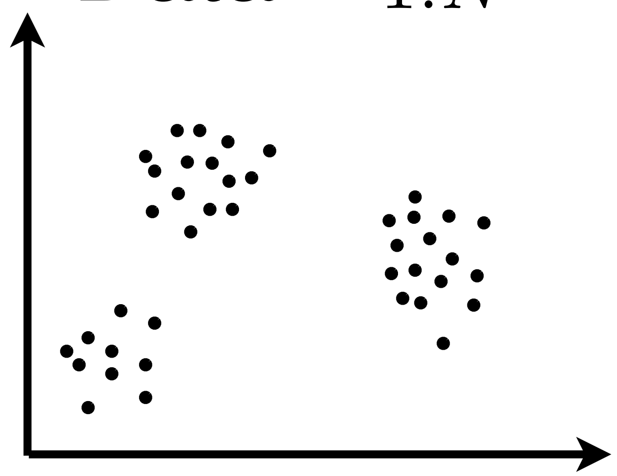
$$\Pi_N \sim \mathrm{CRP}(N, \alpha)$$
$$\forall C \in \Pi_N, \mu_C \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0)$$
$$\forall C \in \Pi_N, \forall n \in C, x_n \overset{indep}{\sim} \mathcal{N}(\mu_C, \Sigma)$$



- Want: posterior

CRP mixture model: inference

- Data $x_{1:N}$

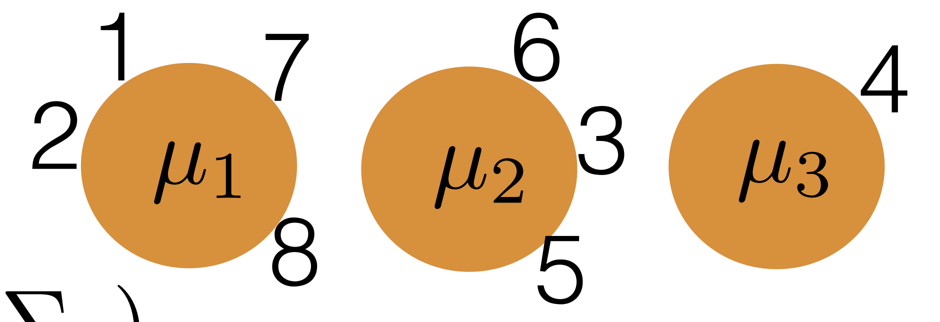


- Generative model

$$\Pi_N \sim \mathrm{CRP}(N, \alpha)$$

$$\forall C \in \Pi_N, \mu_C \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0)$$

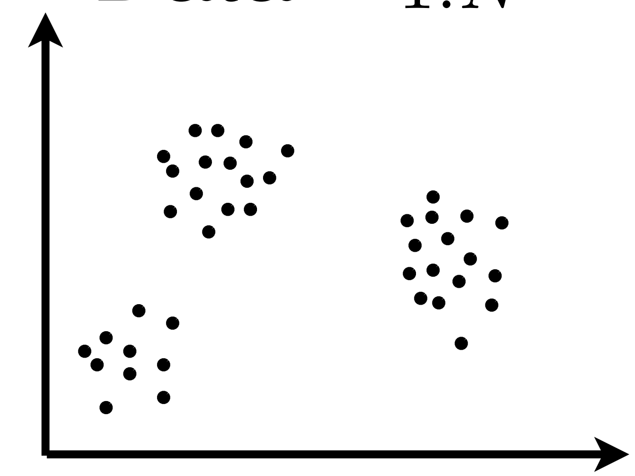
$$\forall C \in \Pi_N, \forall n \in C, x_n \overset{indep}{\sim} \mathcal{N}(\mu_C, \Sigma)$$



- Want: posterior $p(\Pi_N | x_{1:N})$

CRP mixture model: inference

- Data $x_{1:N}$

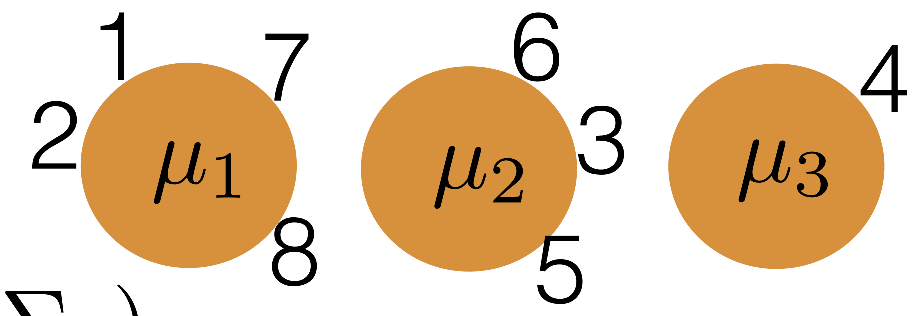


- Generative model

$$\Pi_N \sim \mathrm{CRP}(N, \alpha)$$

$$\forall C \in \Pi_N, \mu_C \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0)$$

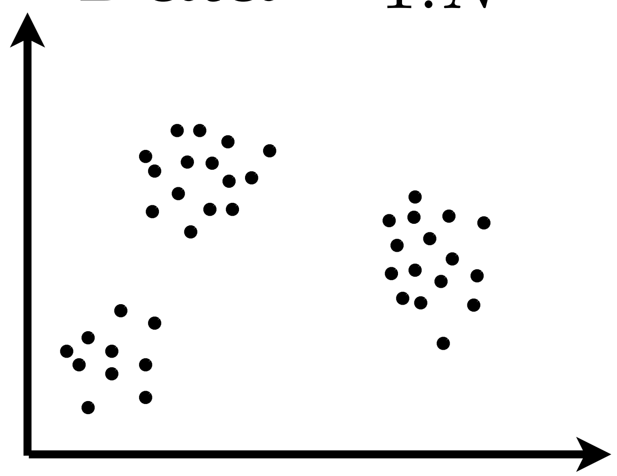
$$\forall C \in \Pi_N, \forall n \in C, x_n \overset{indep}{\sim} \mathcal{N}(\mu_C, \Sigma)$$



- Want: posterior $p(\Pi_N | x_{1:N})$
- Gibbs sampler:

CRP mixture model: inference

- Data $x_{1:N}$

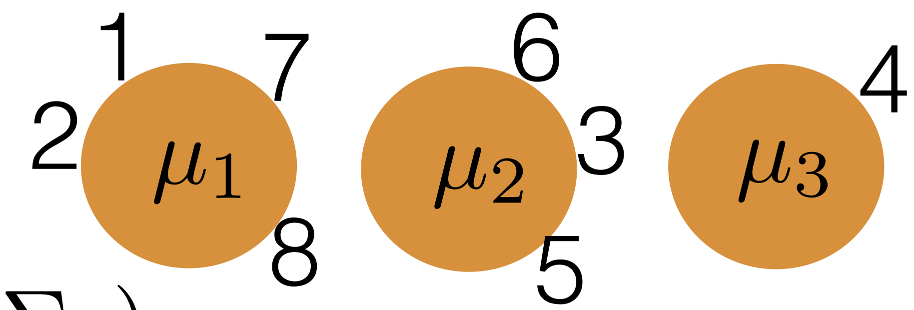


- Generative model

$$\Pi_N \sim \mathrm{CRP}(N, \alpha)$$

$$\forall C \in \Pi_N, \mu_C \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0)$$

$$\forall C \in \Pi_N, \forall n \in C, x_n \overset{indep}{\sim} \mathcal{N}(\mu_C, \Sigma)$$



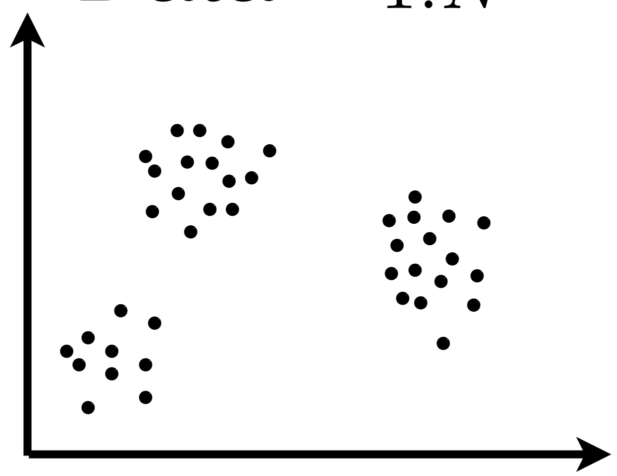
- Want: posterior $p(\Pi_N | x_{1:N})$

- Gibbs sampler:

$$p(\Pi_N | \Pi_{N,\text{-}n}, x)$$

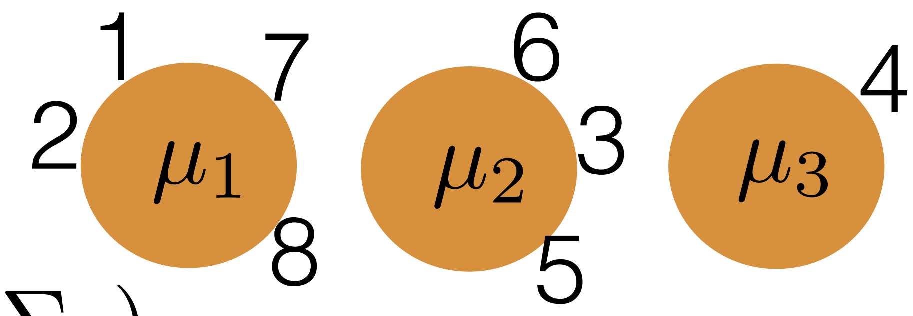
CRP mixture model: inference

- Data $x_{1:N}$



- Generative model

$$\Pi_N \sim \mathrm{CRP}(N, \alpha)$$
$$\forall C \in \Pi_N, \mu_C \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0)$$
$$\forall C \in \Pi_N, \forall n \in C, x_n \overset{indep}{\sim} \mathcal{N}(\mu_C, \Sigma)$$



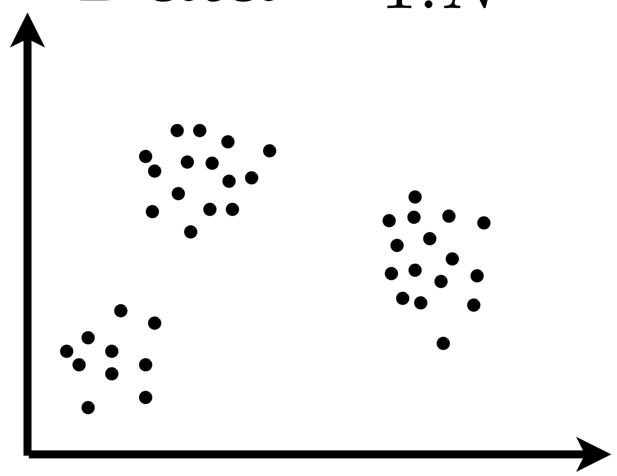
- Want: posterior $p(\Pi_N | x_{1:N})$

- Gibbs sampler:

$$p(\Pi_N | \Pi_{N,\text{-}n}, x) = \Bigg\{$$

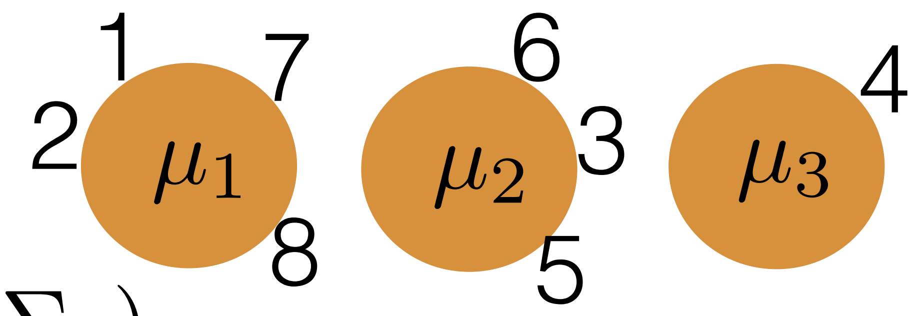
CRP mixture model: inference

- Data $x_{1:N}$



- Generative model

$$\Pi_N \sim \mathrm{CRP}(N, \alpha)$$
$$\forall C \in \Pi_N, \mu_C \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0)$$
$$\forall C \in \Pi_N, \forall n \in C, x_n \overset{indep}{\sim} \mathcal{N}(\mu_C, \Sigma)$$



- Want: posterior $p(\Pi_N | x_{1:N})$

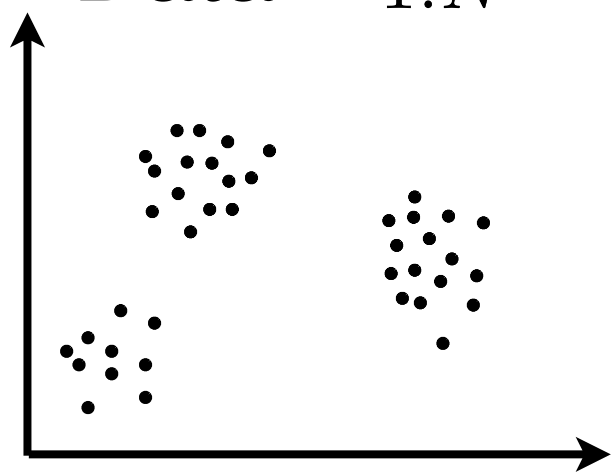
- Gibbs sampler:

$$p(\Pi_N | \Pi_{N,\text{-}n}, x) = \Bigg\{$$

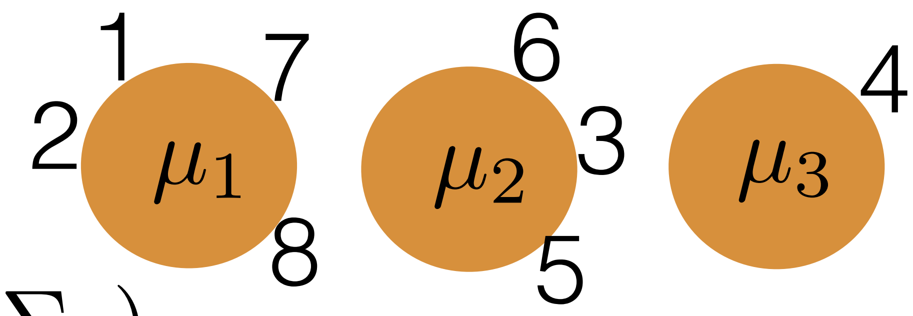
if *n* joins cluster *C*

CRP mixture model: inference

- Data $x_{1:N}$



- Generative model



$$\Pi_N \sim \mathrm{CRP}(N, \alpha)$$
$$\forall C \in \Pi_N, \mu_C \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0)$$
$$\forall C \in \Pi_N, \forall n \in C, x_n \overset{indep}{\sim} \mathcal{N}(\mu_C, \Sigma)$$

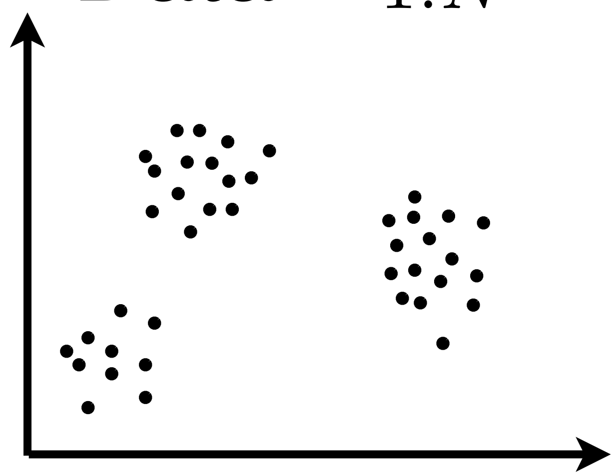
- Want: posterior $p(\Pi_N | x_{1:N})$

- Gibbs sampler:

$$p(\Pi_N | \Pi_{N,\text{-}n}, x) = \begin{cases} & \text{if } n \text{ joins cluster } C \\ & \text{if } n \text{ starts a new cluster} \end{cases}$$

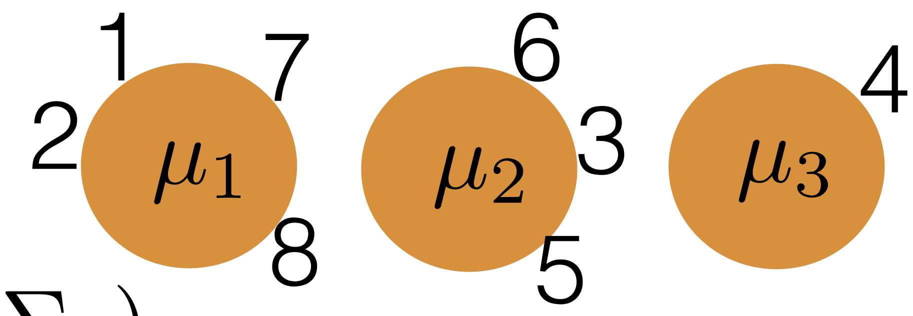
CRP mixture model: inference

- Data $x_{1:N}$



- Generative model

$$\Pi_N \sim \mathrm{CRP}(N, \alpha)$$
$$\forall C \in \Pi_N, \mu_C \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0)$$
$$\forall C \in \Pi_N, \forall n \in C, x_n \overset{indep}{\sim} \mathcal{N}(\mu_C, \Sigma)$$



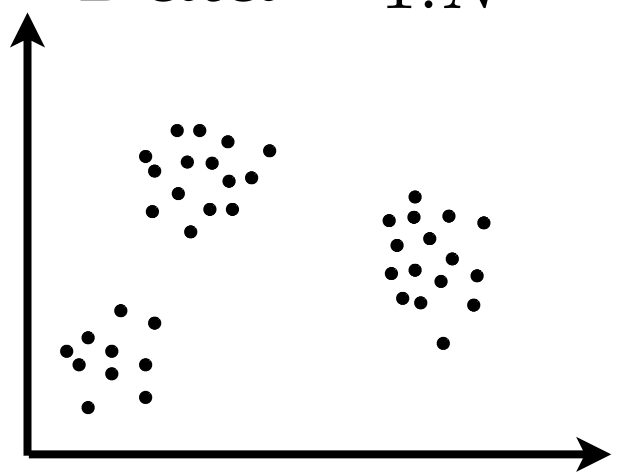
- Want: posterior $p(\Pi_N | x_{1:N})$

- Gibbs sampler:

$$p(\Pi_N | \Pi_{N,\text{-}n}, x) = \begin{cases} \frac{\#C}{\alpha+N-1} p(x_{C\cup\{n\}} | x_C) & \text{if } n \text{ joins cluster } C \\ & \text{if } n \text{ starts a new cluster} \end{cases}$$

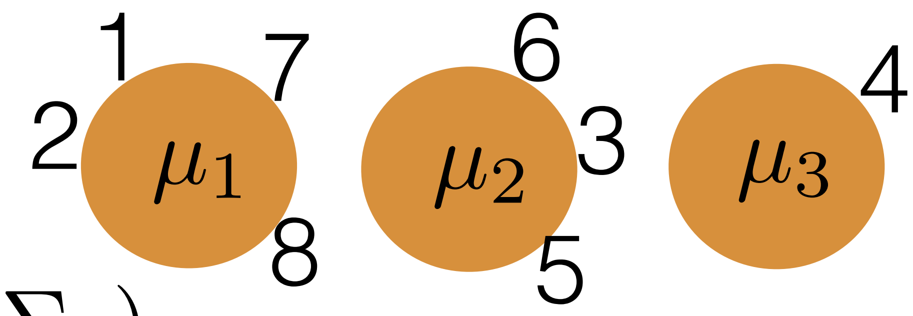
CRP mixture model: inference

- Data $x_{1:N}$



- Generative model

$$\Pi_N \sim \mathrm{CRP}(N, \alpha)$$
$$\forall C \in \Pi_N, \mu_C \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0)$$
$$\forall C \in \Pi_N, \forall n \in C, x_n \overset{indep}{\sim} \mathcal{N}(\mu_C, \Sigma)$$



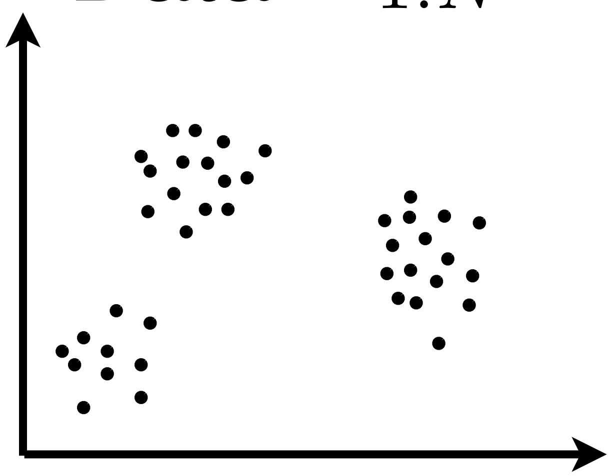
- Want: posterior $p(\Pi_N | x_{1:N})$

- Gibbs sampler:

$$p(\Pi_N | \Pi_{N,\text{-}n}, x) = \begin{cases} \frac{\#C}{\alpha+N-1} p(x_{C \cup \{n\}} | x_C) & \text{if } n \text{ joins cluster } C \\ \frac{\alpha}{\alpha+N-1} p(x_{\{n\}}) & \text{if } n \text{ starts a new cluster} \end{cases}$$

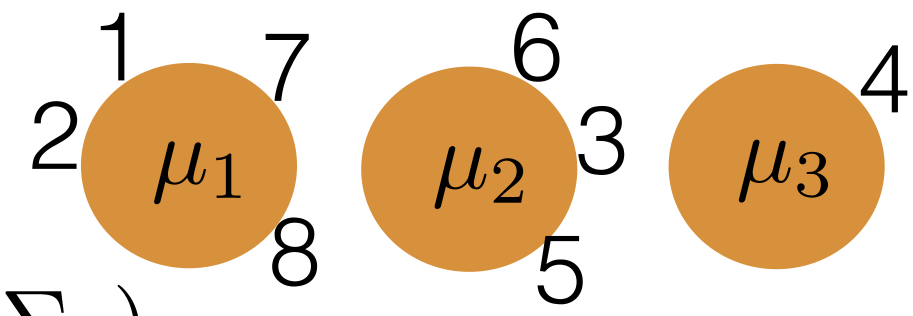
CRP mixture model: inference

- Data $x_{1:N}$



- Generative model

$$\Pi_N \sim \mathrm{CRP}(N, \alpha)$$
$$\forall C \in \Pi_N, \mu_C \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0)$$
$$\forall C \in \Pi_N, \forall n \in C, x_n \overset{indep}{\sim} \mathcal{N}(\mu_C, \Sigma)$$



- Want: posterior $p(\Pi_N | x_{1:N})$

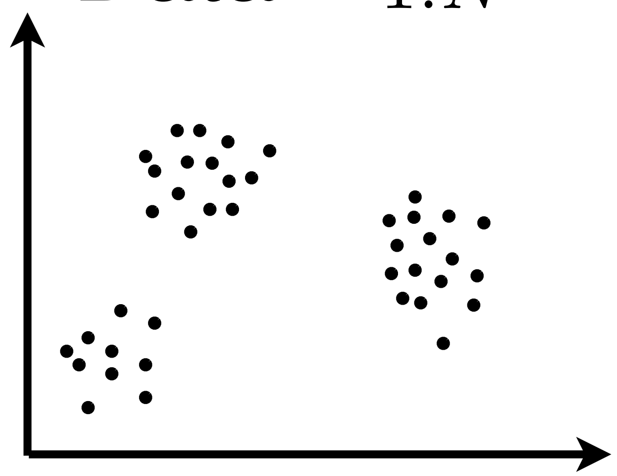
- Gibbs sampler:

$$p(\Pi_N | \Pi_{N,\text{-}n}, x) = \begin{cases} \frac{\#C}{\alpha+N-1} p(x_{C\cup\{n\}} | x_C) & \text{if } n \text{ joins cluster } C \\ \frac{\alpha}{\alpha+N-1} p(x_{\{n\}}) & \text{if } n \text{ starts a new cluster} \end{cases}$$

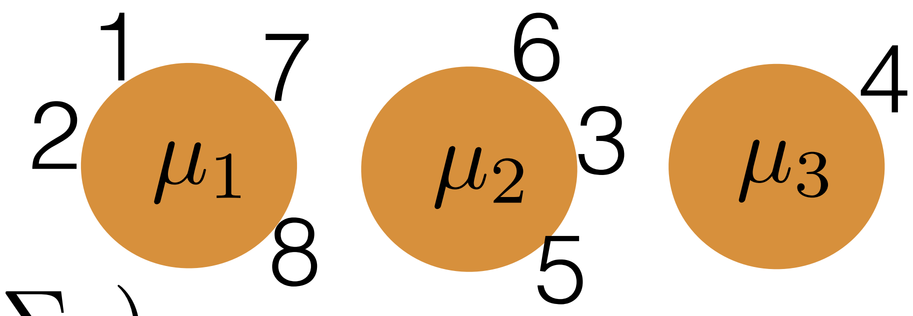
- For completeness: $p(x_{C\cup\{n\}} | x_C) =$

CRP mixture model: inference

- Data $x_{1:N}$



- Generative model



$$\Pi_N \sim \mathrm{CRP}(N, \alpha)$$
$$\forall C \in \Pi_N, \mu_C \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0)$$
$$\forall C \in \Pi_N, \forall n \in C, x_n \overset{indep}{\sim} \mathcal{N}(\mu_C, \Sigma)$$

- Want: posterior $p(\Pi_N | x_{1:N})$

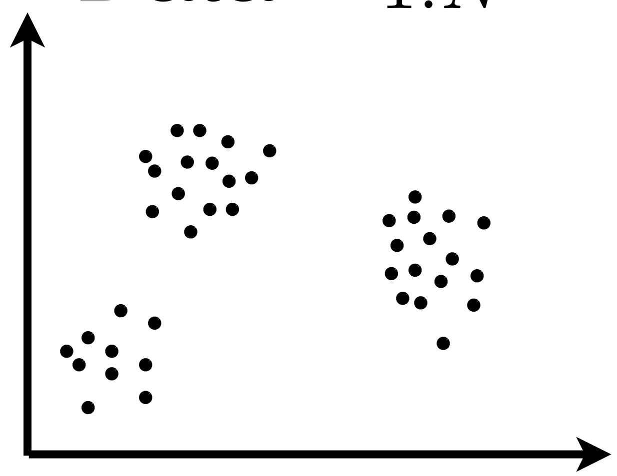
- Gibbs sampler:

$$p(\Pi_N | \Pi_{N,\text{-}n}, x) = \begin{cases} \frac{\#C}{\alpha+N-1} p(x_{C\cup\{n\}} | x_C) & \text{if } n \text{ joins cluster } C \\ \frac{\alpha}{\alpha+N-1} p(x_{\{n\}}) & \text{if } n \text{ starts a new cluster} \end{cases}$$

- For completeness: $p(x_{C\cup\{n\}} | x_C) = \mathcal{N}(\tilde{m}, \tilde{\Sigma} + \Sigma)$

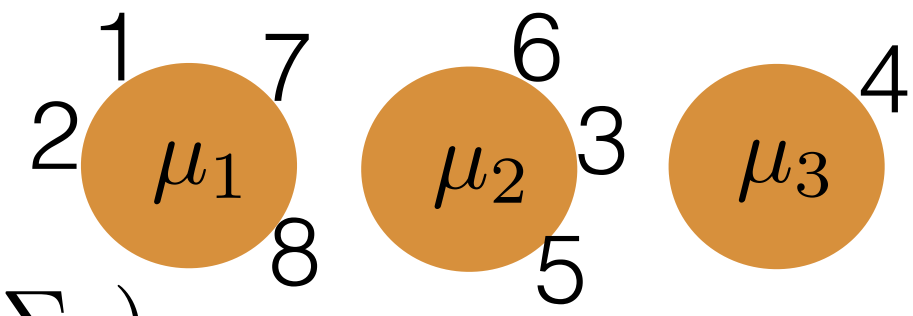
CRP mixture model: inference

- Data $x_{1:N}$



- Generative model

$$\Pi_N \sim \mathrm{CRP}(N, \alpha)$$
$$\forall C \in \Pi_N, \mu_C \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0)$$
$$\forall C \in \Pi_N, \forall n \in C, x_n \overset{indep}{\sim} \mathcal{N}(\mu_C, \Sigma)$$



- Want: posterior $p(\Pi_N | x_{1:N})$

- Gibbs sampler:

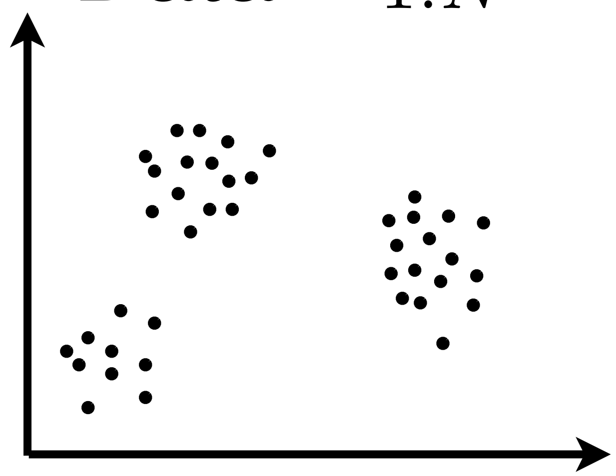
$$p(\Pi_N | \Pi_{N,\text{-}n}, x) = \begin{cases} \frac{\#C}{\alpha+N-1} p(x_{C\cup\{n\}} | x_C) & \text{if } n \text{ joins cluster } C \\ \frac{\alpha}{\alpha+N-1} p(x_{\{n\}}) & \text{if } n \text{ starts a new cluster} \end{cases}$$

- For completeness: $p(x_{C\cup\{n\}} | x_C) = \mathcal{N}(\tilde{m}, \tilde{\Sigma} + \Sigma)$

$$\tilde{\Sigma}^{-1} := \Sigma_0^{-1} + (\#C)\Sigma^{-1}$$
$$\tilde{m} := \tilde{\Sigma}\left(\Sigma^{-1} \sum_{m\in C} x_m + \Sigma_0^{-1}\mu_0\right)$$

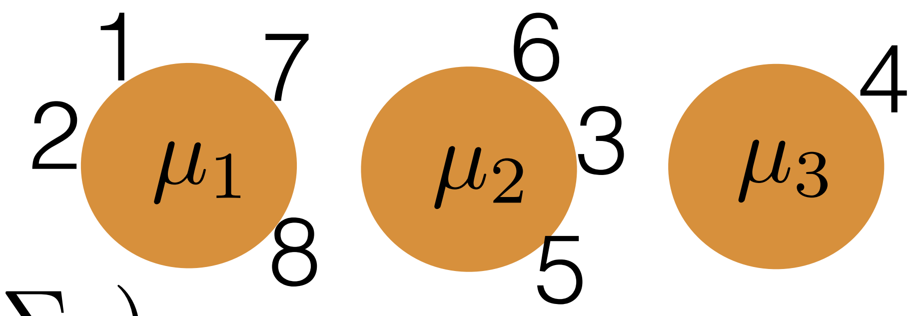
CRP mixture model: inference

- Data $x_{1:N}$



- Generative model

$$\Pi_N \sim \mathrm{CRP}(N, \alpha)$$
$$\forall C \in \Pi_N, \mu_C \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0)$$
$$\forall C \in \Pi_N, \forall n \in C, x_n \overset{indep}{\sim} \mathcal{N}(\mu_C, \Sigma)$$



- Want: posterior $p(\Pi_N | x_{1:N})$

- Gibbs sampler:

$$p(\Pi_N | \Pi_{N,\text{-}n}, x) = \begin{cases} \frac{\#C}{\alpha + N - 1} p(x_{C \cup \{n\}} | x_C) & \text{if } n \text{ joins cluster } C \\ \frac{\alpha}{\alpha + N - 1} p(x_{\{n\}}) & \text{if } n \text{ starts a new cluster} \end{cases}$$

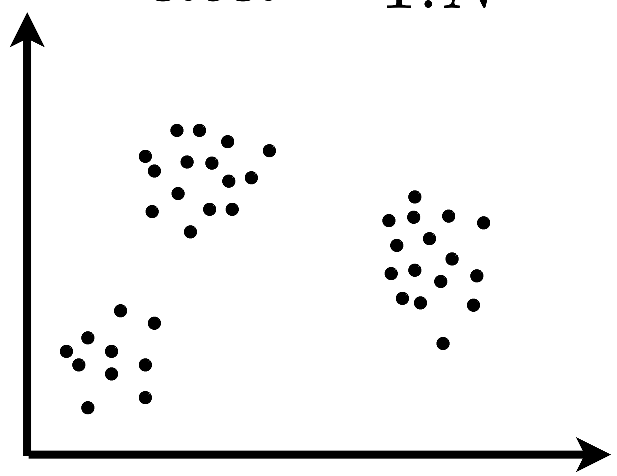
- For completeness: $p(x_{C \cup \{n\}} | x_C) = \mathcal{N}(\tilde{m}, \tilde{\Sigma} + \Sigma)$

$$\tilde{\Sigma}^{-1} := \Sigma_0^{-1} + (\#C)\Sigma^{-1}$$
$$\tilde{m} := \tilde{\Sigma} \left(\Sigma^{-1} \sum_{m \in C} x_m + \Sigma_0^{-1} \mu_0 \right)$$

[MacEachern 1994; Neal 1992; Neal 2000]

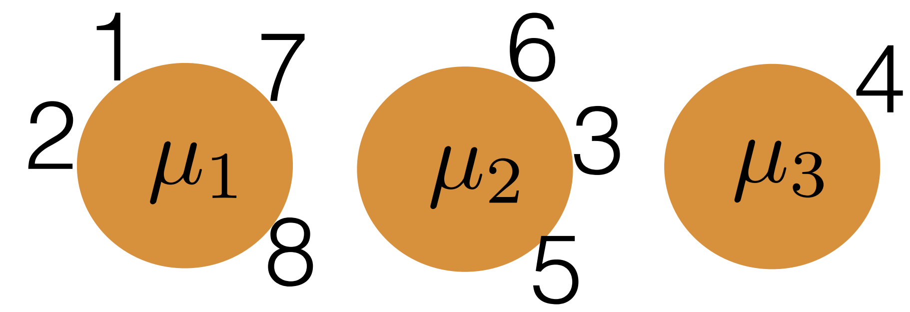
CRP mixture model: inference

- Data $x_{1:N}$



- Generative model

$$\Pi_N \sim \mathrm{CRP}(N, \alpha)$$
$$\forall C \in \Pi_N, \boxed{\phi_C \overset{iid}{\sim} G_0}$$
$$\forall C \in \Pi_N, \forall n \in C, x_n \overset{indep}{\sim} \mathcal{N}(\mu_C, \Sigma)$$



- Want: posterior $p(\Pi_N | x_{1:N})$

- Gibbs sampler:

$$p(\Pi_N | \Pi_{N,\text{-}n}, x) = \begin{cases} \frac{\#C}{\alpha + N - 1} p(x_{C \cup \{n\}} | x_C) & \text{if } n \text{ joins cluster } C \\ \frac{\alpha}{\alpha + N - 1} p(x_{\{n\}}) & \text{if } n \text{ starts a new cluster} \end{cases}$$

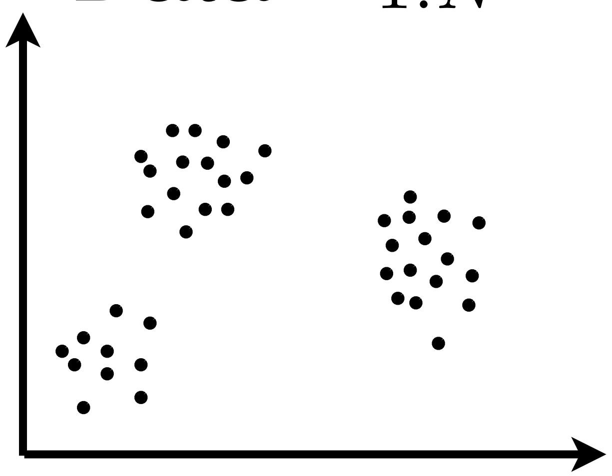
- For completeness: $p(x_{C \cup \{n\}} | x_C) = \mathcal{N}(\tilde{m}, \tilde{\Sigma} + \Sigma)$

$$\tilde{\Sigma}^{-1} := \Sigma_0^{-1} + (\#C)\Sigma^{-1}$$
$$\tilde{m} := \tilde{\Sigma} \left(\Sigma^{-1} \sum_{m \in C} x_m + \Sigma_0^{-1} \mu_0 \right)$$

[MacEachern 1994; Neal 1992; Neal 2000]

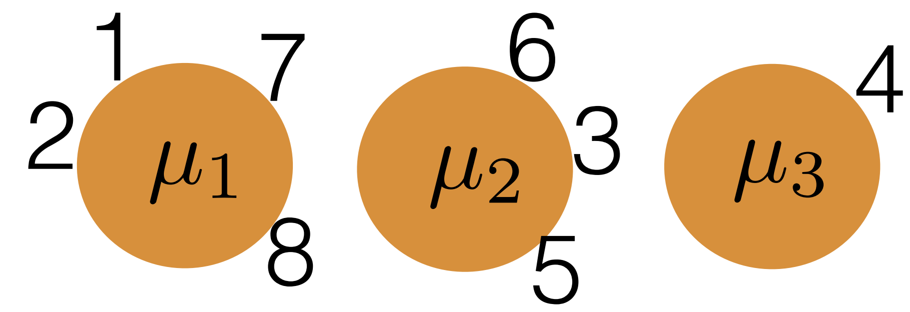
CRP mixture model: inference

- Data $x_{1:N}$



- Generative model

$$\Pi_N \sim \mathrm{CRP}(N, \alpha)$$
$$\forall C \in \Pi_N, \boxed{\phi_C \overset{iid}{\sim} G_0}$$
$$\forall C \in \Pi_N, \forall n \in C, x_n \overset{indep}{\sim} \boxed{F(\phi_C)}$$



- Want: posterior $p(\Pi_N | x_{1:N})$

- Gibbs sampler:

$$p(\Pi_N | \Pi_{N,\text{-}n}, x) = \begin{cases} \frac{\#C}{\alpha + N - 1} p(x_{C \cup \{n\}} | x_C) & \text{if } n \text{ joins cluster } C \\ \frac{\alpha}{\alpha + N - 1} p(x_{\{n\}}) & \text{if } n \text{ starts a new cluster} \end{cases}$$

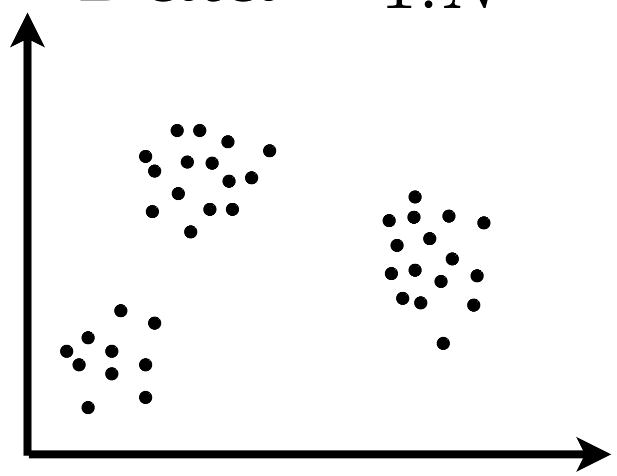
- For completeness: $p(x_{C \cup \{n\}} | x_C) = \mathcal{N}(\tilde{m}, \tilde{\Sigma} + \Sigma)$

$$\tilde{\Sigma}^{-1} := \Sigma_0^{-1} + (\#C)\Sigma^{-1}$$
$$\tilde{m} := \tilde{\Sigma}\left(\Sigma^{-1} \sum_{m \in C} x_m + \Sigma_0^{-1} \mu_0\right)$$

[MacEachern 1994; Neal 1992; Neal 2000]

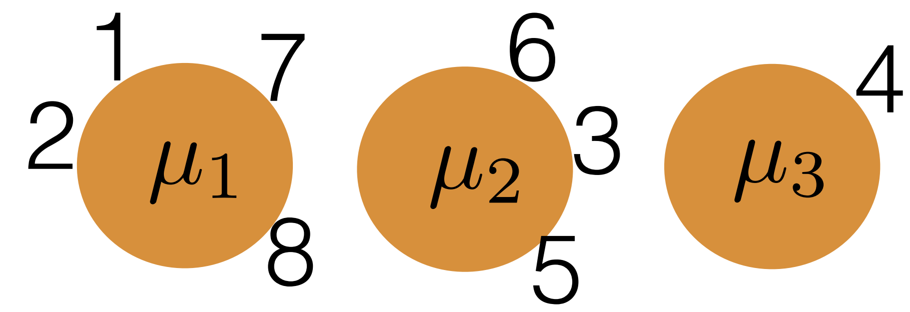
CRP mixture model: inference

- Data $x_{1:N}$



- Generative model

$$\Pi_N \sim \mathrm{CRP}(N, \alpha)$$
$$\forall C \in \Pi_N, \boxed{\phi_C \overset{iid}{\sim} G_0}$$
$$\forall C \in \Pi_N, \forall n \in C, x_n \overset{indep}{\sim} \boxed{F(\phi_C)}$$



- Want: posterior $p(\Pi_N | x_{1:N})$

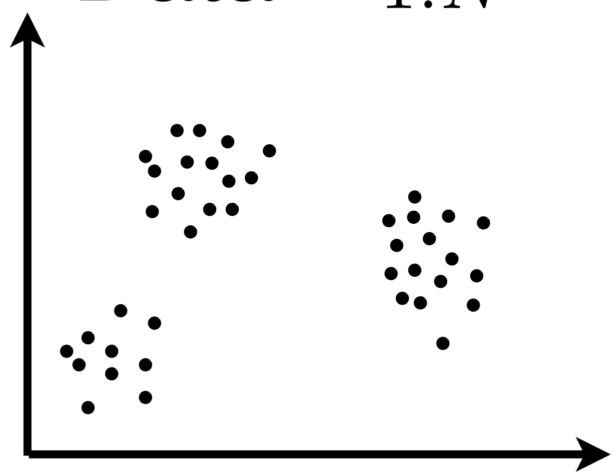
- Gibbs sampler:

$$p(\Pi_N | \Pi_{N,\text{-}n}, x) = \begin{cases} \frac{\#C}{\alpha+N-1} p(x_{C\cup\{n\}} | x_C) & \text{if } n \text{ joins cluster } C \\ \frac{\alpha}{\alpha+N-1} p(x_{\{n\}}) & \text{if } n \text{ starts a new cluster} \end{cases}$$

[MacEachern 1994; Neal 1992; Neal 2000]

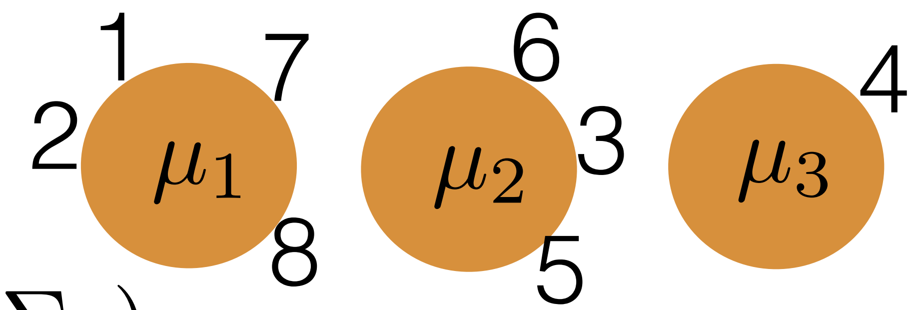
CRP mixture model: inference

- Data $x_{1:N}$



- Generative model

$$\Pi_N \sim \mathrm{CRP}(N, \alpha)$$
$$\forall C \in \Pi_N, \mu_C \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0)$$
$$\forall C \in \Pi_N, \forall n \in C, x_n \overset{indep}{\sim} \mathcal{N}(\mu_C, \Sigma)$$



- Want: posterior $p(\Pi_N | x_{1:N})$

- Gibbs sampler:

$$p(\Pi_N | \Pi_{N,\text{-}n}, x) = \begin{cases} \frac{\#C}{\alpha + N - 1} p(x_{C \cup \{n\}} | x_C) & \text{if } n \text{ joins cluster } C \\ \frac{\alpha}{\alpha + N - 1} p(x_{\{n\}}) & \text{if } n \text{ starts a new cluster} \end{cases}$$

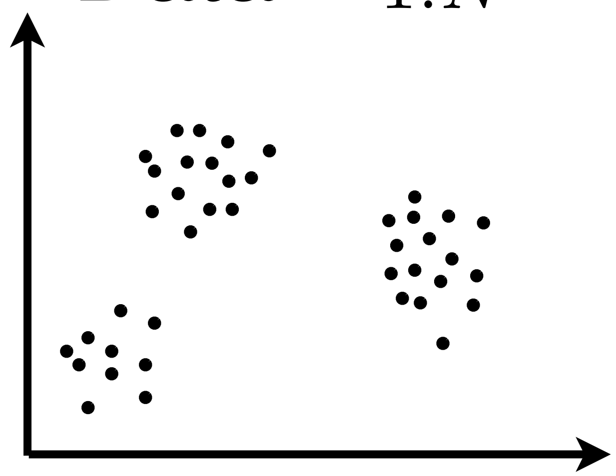
- For completeness: $p(x_{C \cup \{n\}} | x_C) = \mathcal{N}(\tilde{m}, \tilde{\Sigma} + \Sigma)$

$$\tilde{\Sigma}^{-1} := \Sigma_0^{-1} + (\#C)\Sigma^{-1}$$
$$\tilde{m} := \tilde{\Sigma}\left(\Sigma^{-1} \sum_{m \in C} x_m + \Sigma_0^{-1}\mu_0\right)$$

[MacEachern 1994; Neal 1992; Neal 2000]

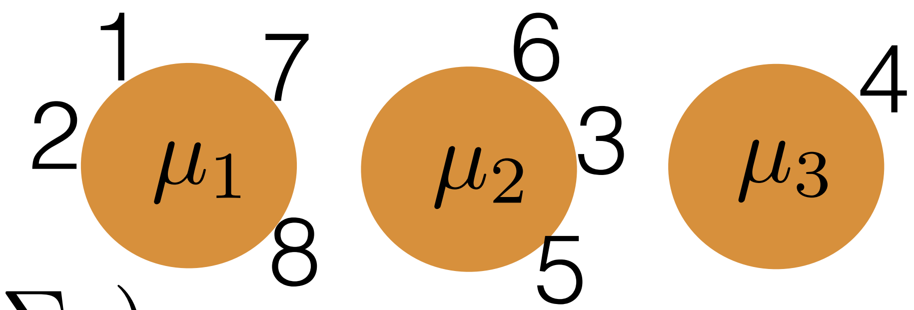
CRP mixture model: inference

- Data $x_{1:N}$



- Generative model

$$\Pi_N \sim \mathrm{CRP}(N, \alpha)$$
$$\forall C \in \Pi_N, \mu_C \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0)$$
$$\forall C \in \Pi_N, \forall n \in C, x_n \overset{indep}{\sim} \mathcal{N}(\mu_C, \Sigma)$$



- Want: posterior $p(\Pi_N | x_{1:N})$

- Gibbs sampler:

$$p(\Pi_N | \Pi_{N,\text{-}n}, x) = \begin{cases} \frac{\#C}{\alpha+N-1} p(x_{C\cup\{n\}} | x_C) & \text{if } n \text{ joins cluster } C \\ \frac{\alpha}{\alpha+N-1} p(x_{\{n\}}) & \text{if } n \text{ starts a new cluster} \end{cases}$$

- For completeness: $p(x_{C\cup\{n\}} | x_C) = \mathcal{N}(\tilde{m}, \tilde{\Sigma} + \Sigma)$

$$\tilde{\Sigma}^{-1} := \Sigma_0^{-1} + (\#C)\Sigma^{-1}$$
$$\tilde{m} := \tilde{\Sigma}\left(\Sigma^{-1} \sum_{m \in C} x_m + \Sigma_0^{-1}\mu_0\right)$$

[demo]

[MacEachern 1994; Neal 1992; Neal 2000]

More Markov Chain Monte Carlo

More Markov Chain Monte Carlo

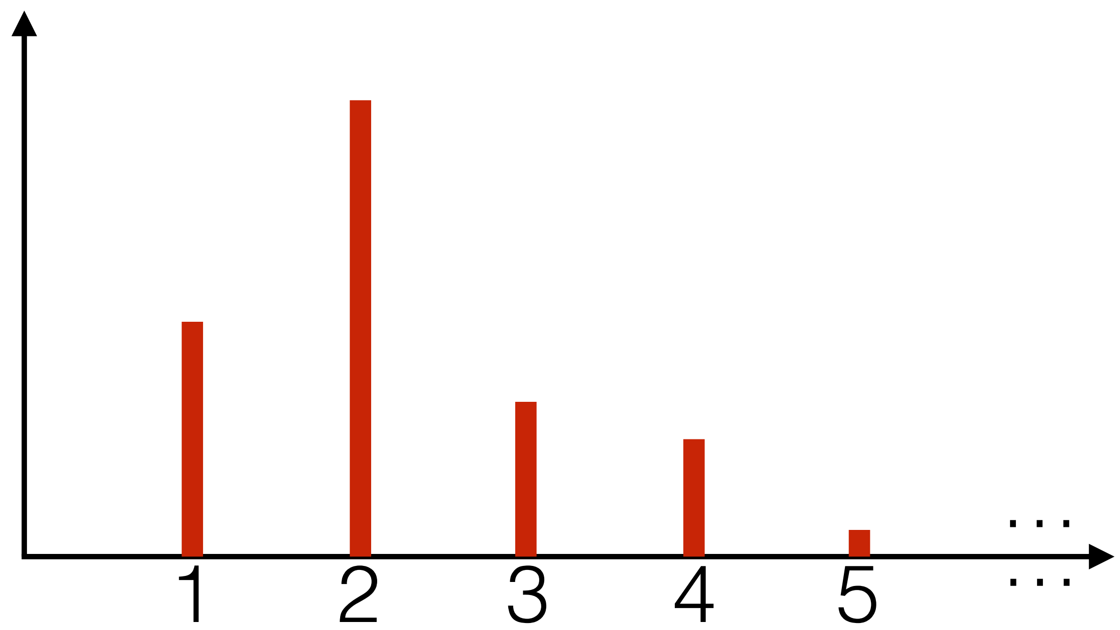
- Slice sampling

More Markov Chain Monte Carlo

- Slice sampling
  - auxiliary variable ➔ finite conditionals

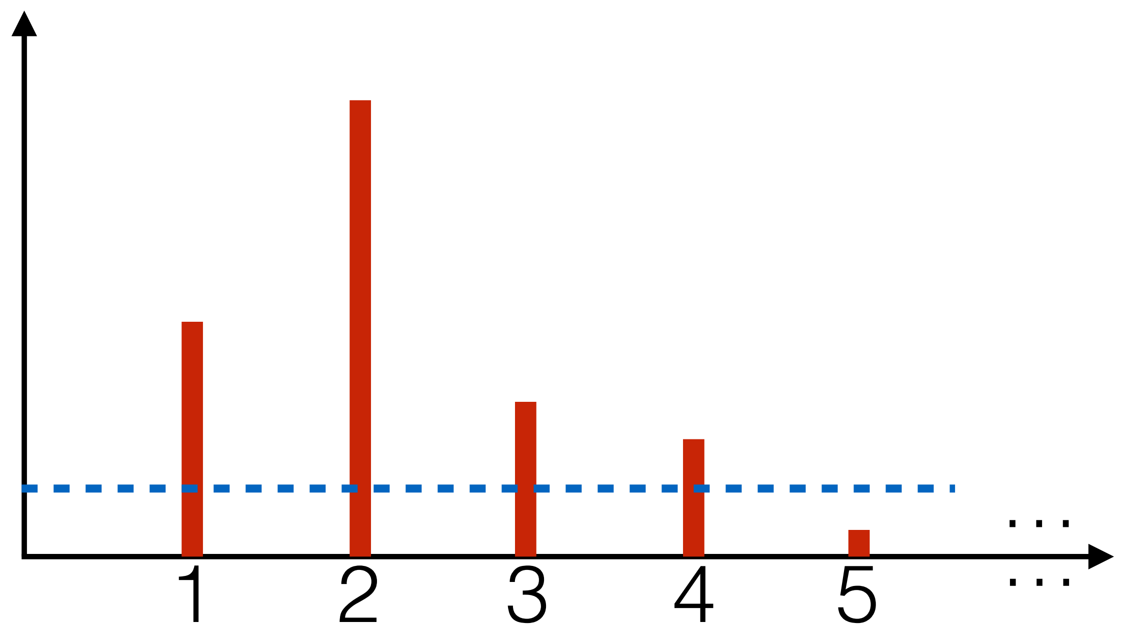
More Markov Chain Monte Carlo

- Slice sampling
  - auxiliary variable ➔ finite conditionals



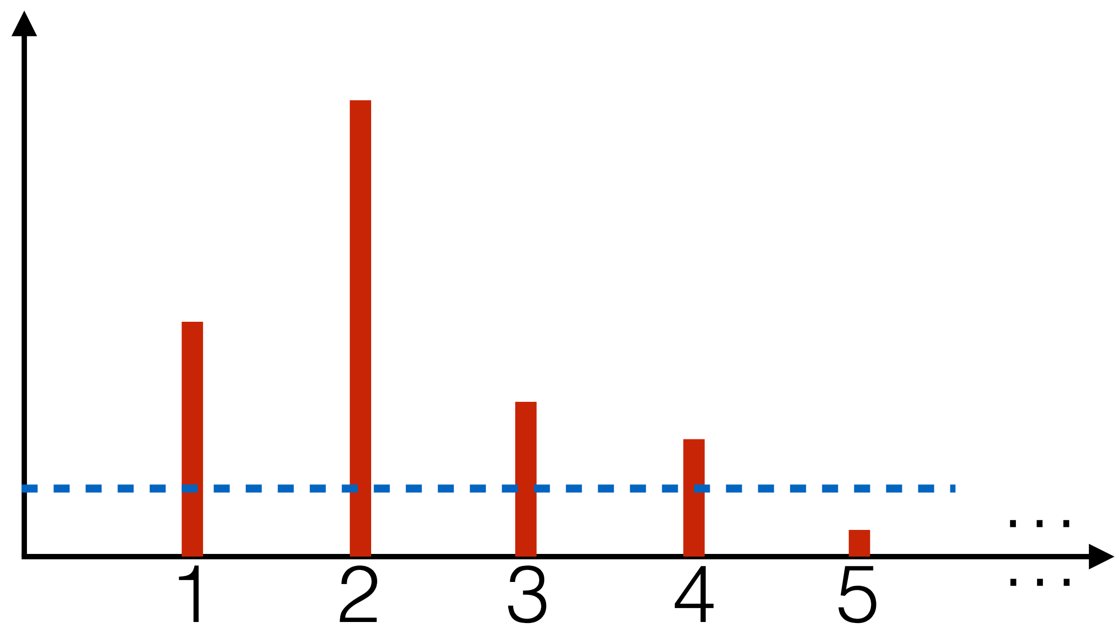
More Markov Chain Monte Carlo

- Slice sampling
  - auxiliary variable ➔ finite conditionals



More Markov Chain Monte Carlo

- Slice sampling
    - auxiliary variable ➔ finite conditionals

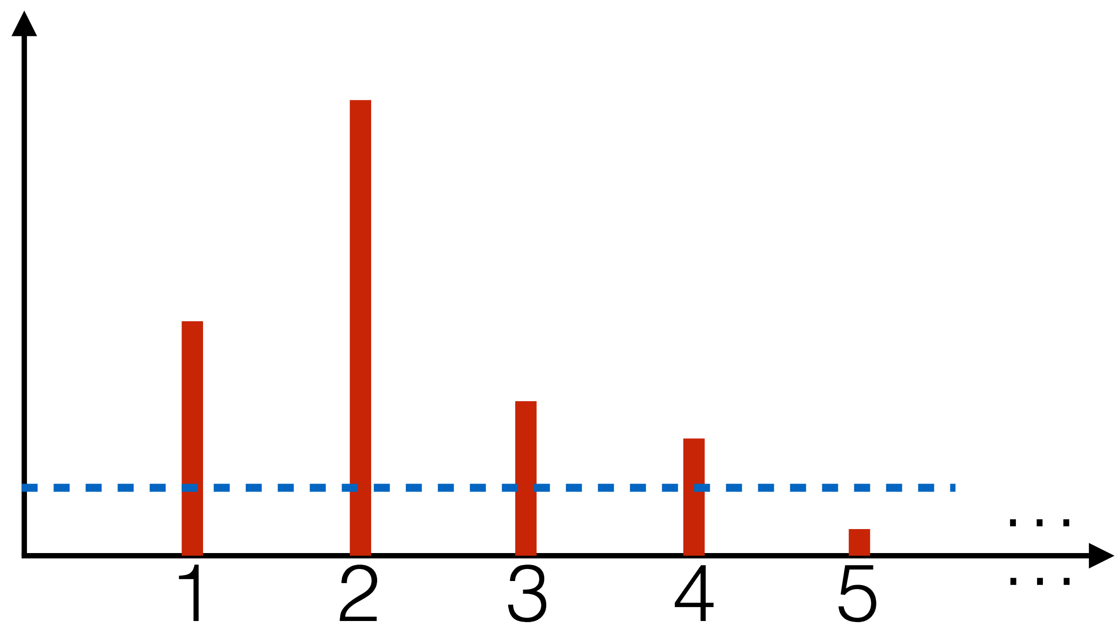


- Approximate with truncated distribution

[Ishwaran, James 2001; Campbell*, Huggins*, Broderick 2016]

More Markov Chain Monte Carlo

- Slice sampling
  - auxiliary variable ➔ finite conditionals

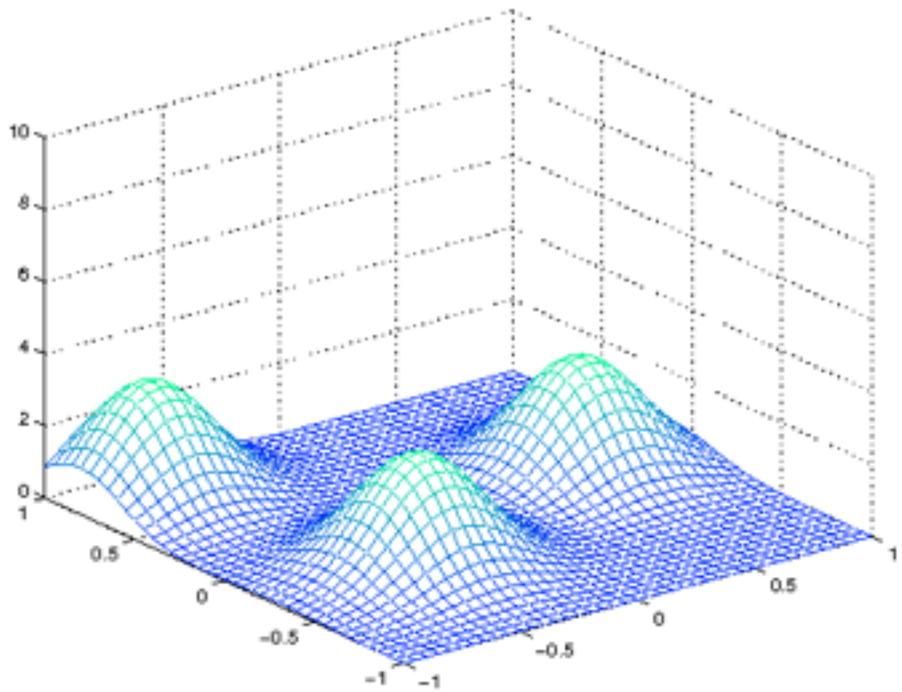


- Approximate with truncated distribution
  - E.g., Hamiltonian Monte Carlo

[Ishwaran, James 2001; Campbell*, Huggins*, Broderick 2016]

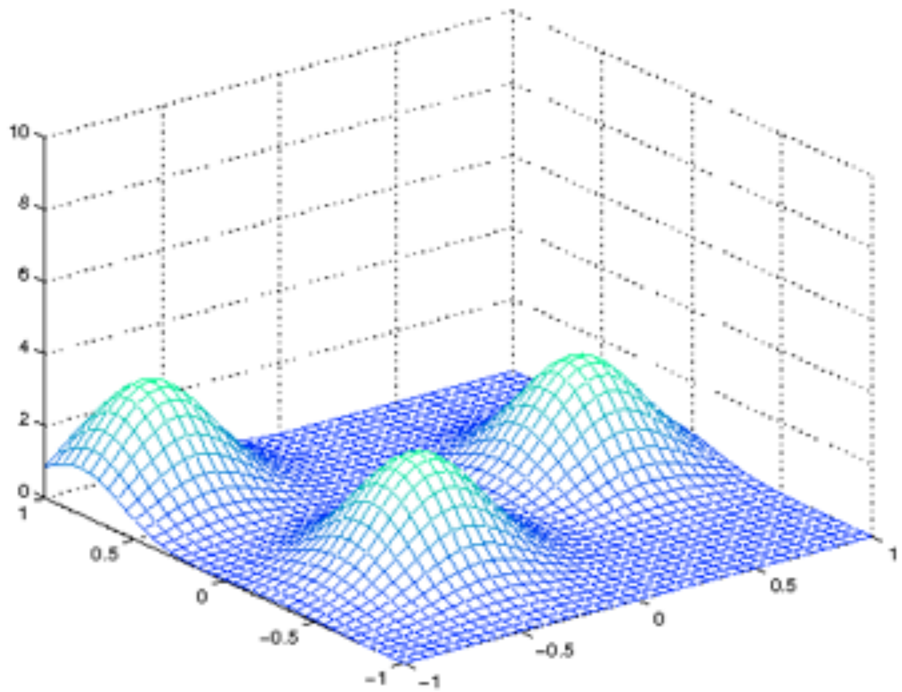
Variational Bayes

Variational Bayes



- Variational Bayes (VB)

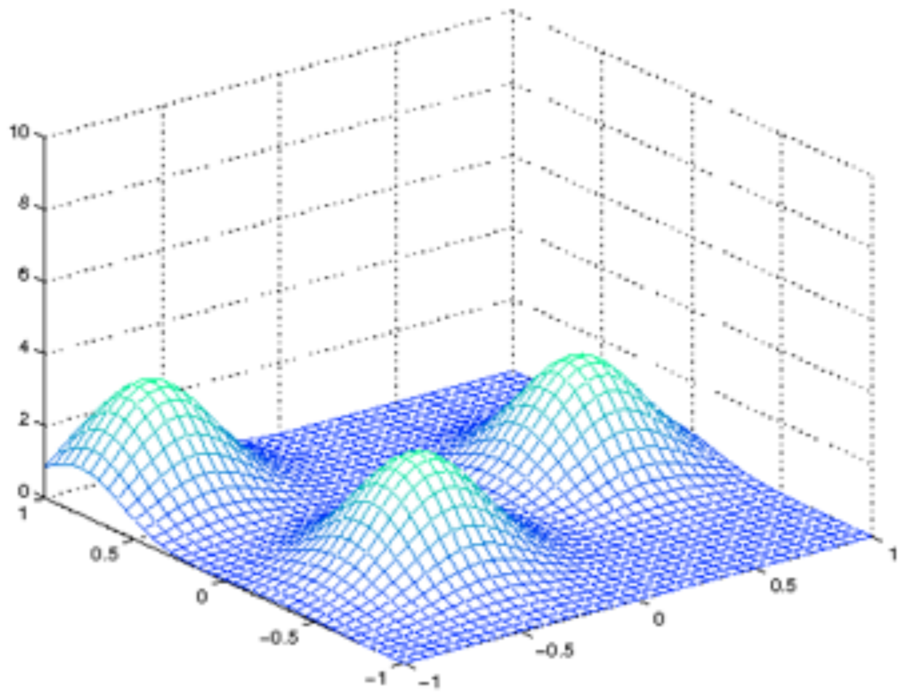
Variational Bayes

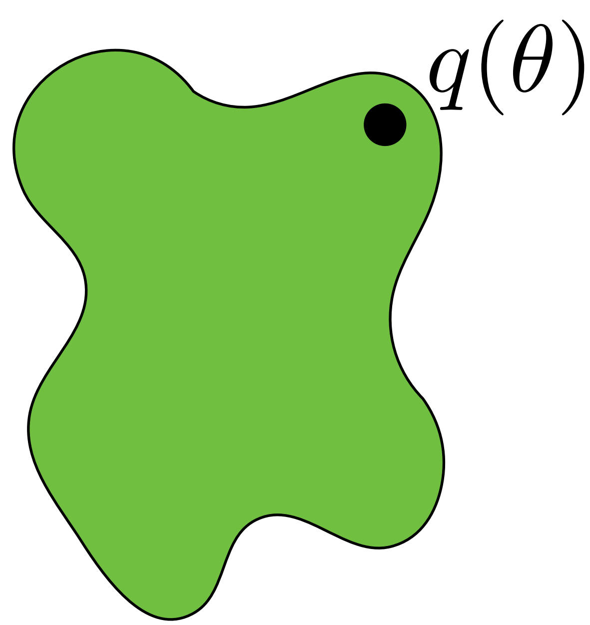


- Variational Bayes (VB)
  - Approximation $q^*(\theta)$ for posterior $p(\theta|x)$

Variational Bayes

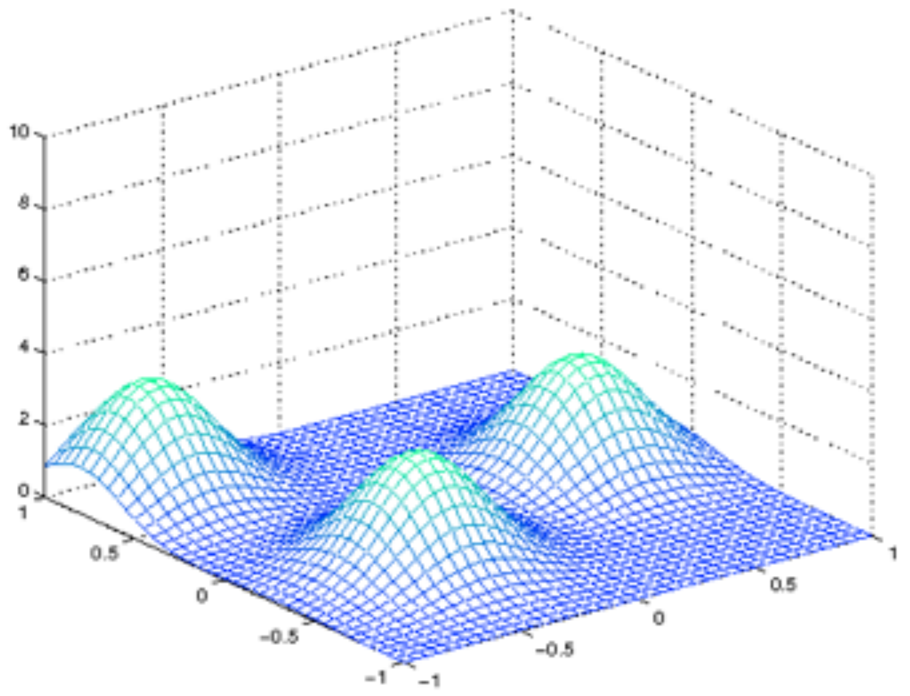
- Variational Bayes (VB)
  - Approximation $q^*(\theta)$ for posterior $p(\theta|x)$

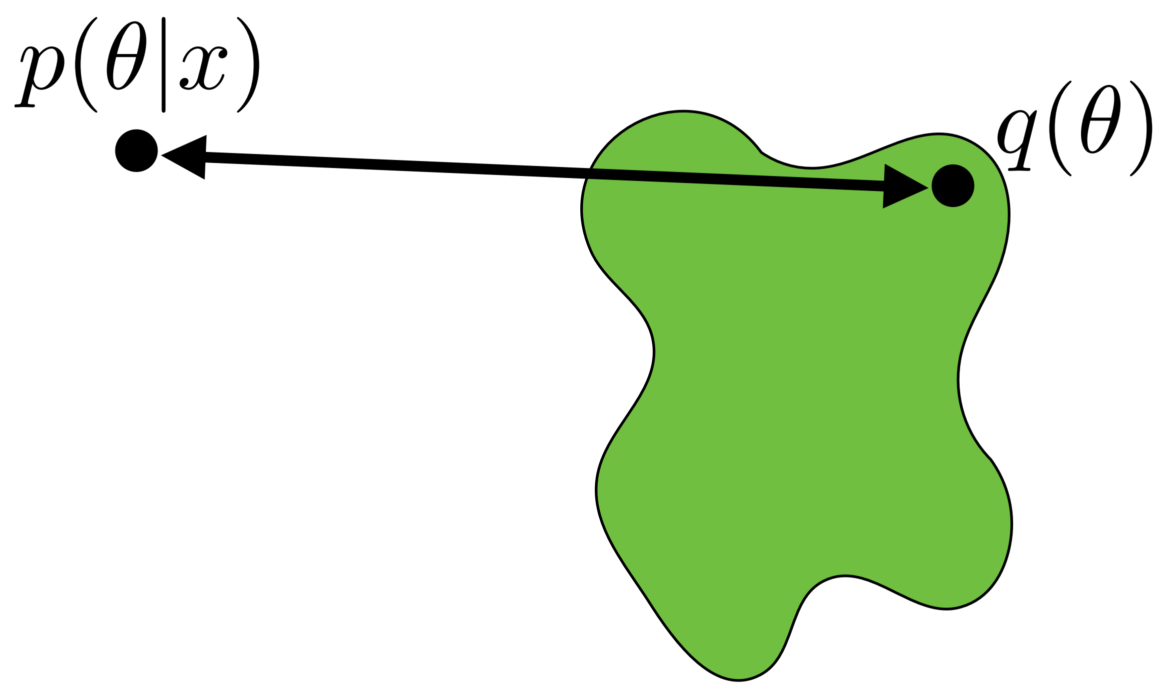




Variational Bayes

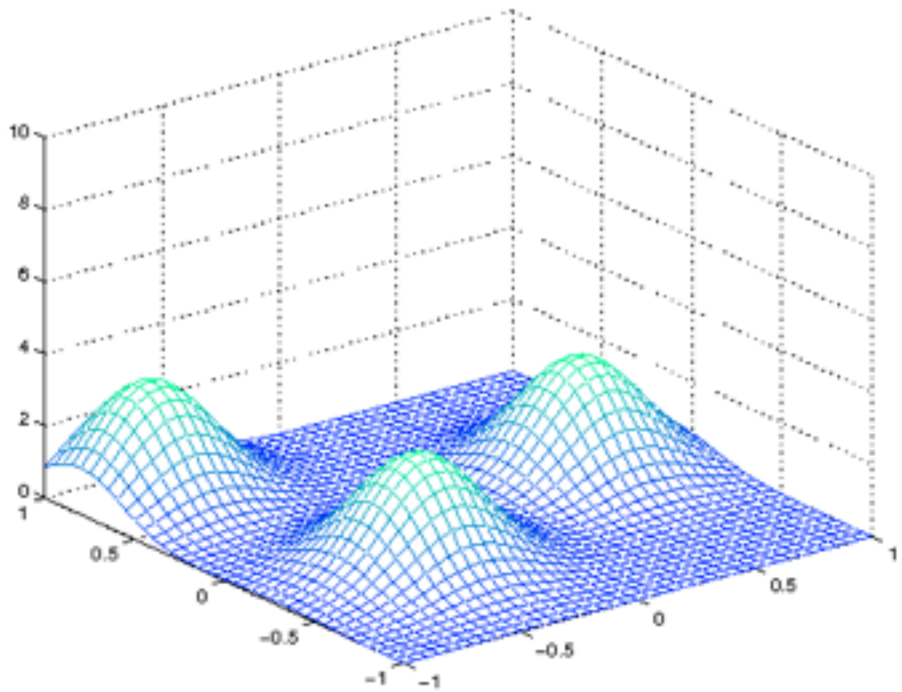
- Variational Bayes (VB)
  - Approximation $q^*(\theta)$ for posterior $p(\theta|x)$

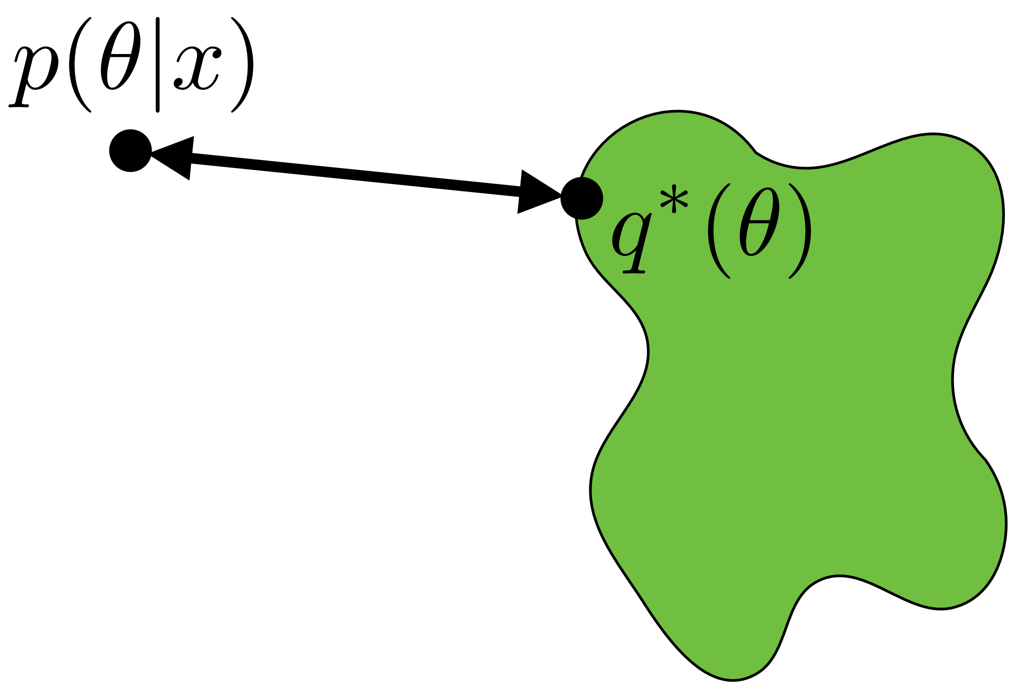




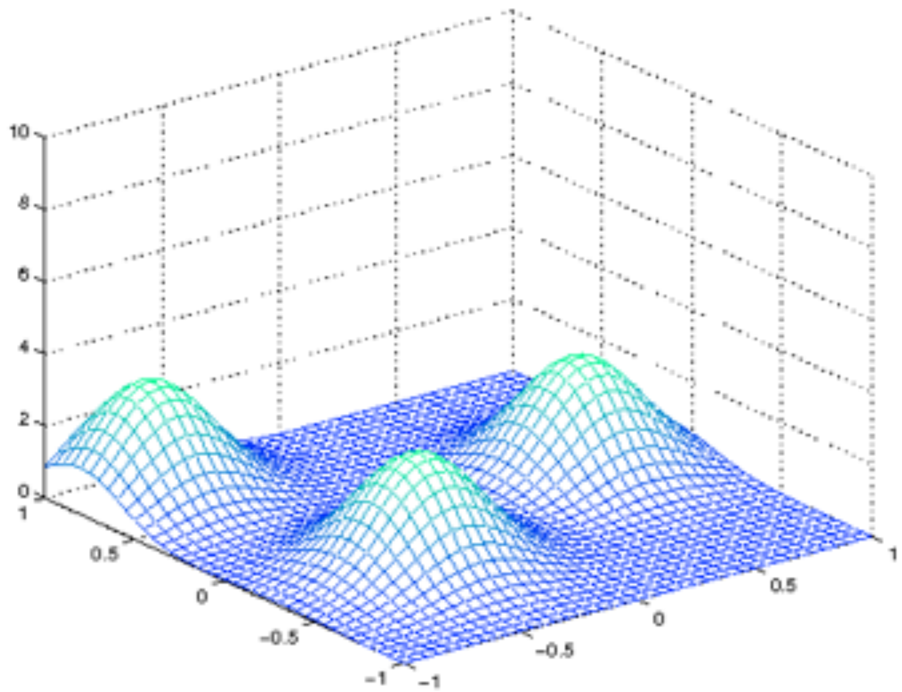
Variational Bayes

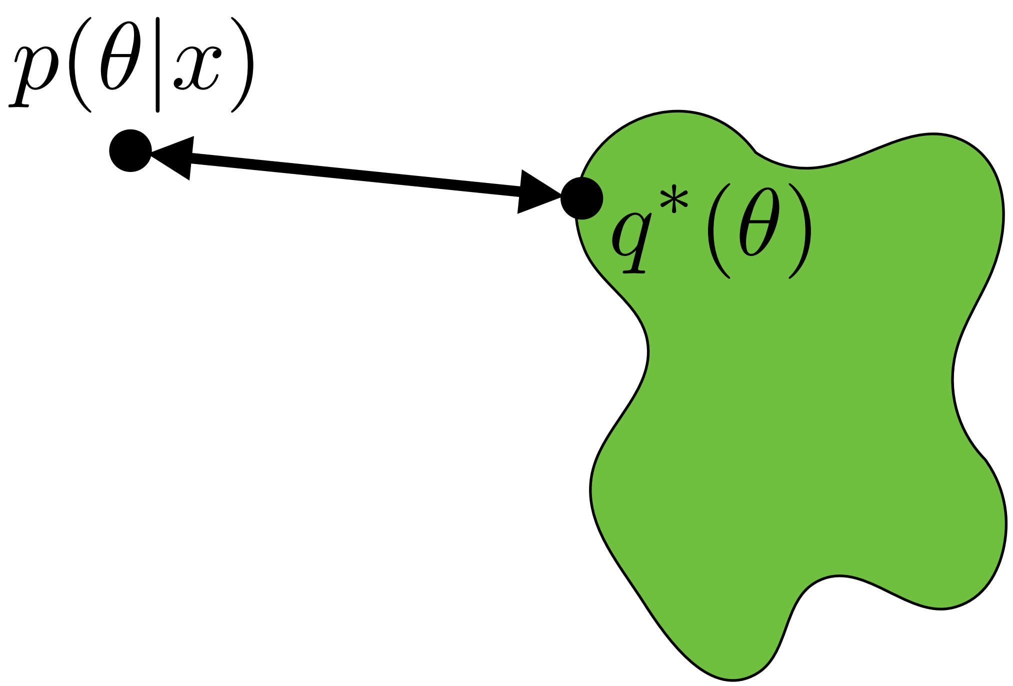
- Variational Bayes (VB)
  - Approximation $q^*(\theta)$ for posterior $p(\theta|x)$





Variational Bayes

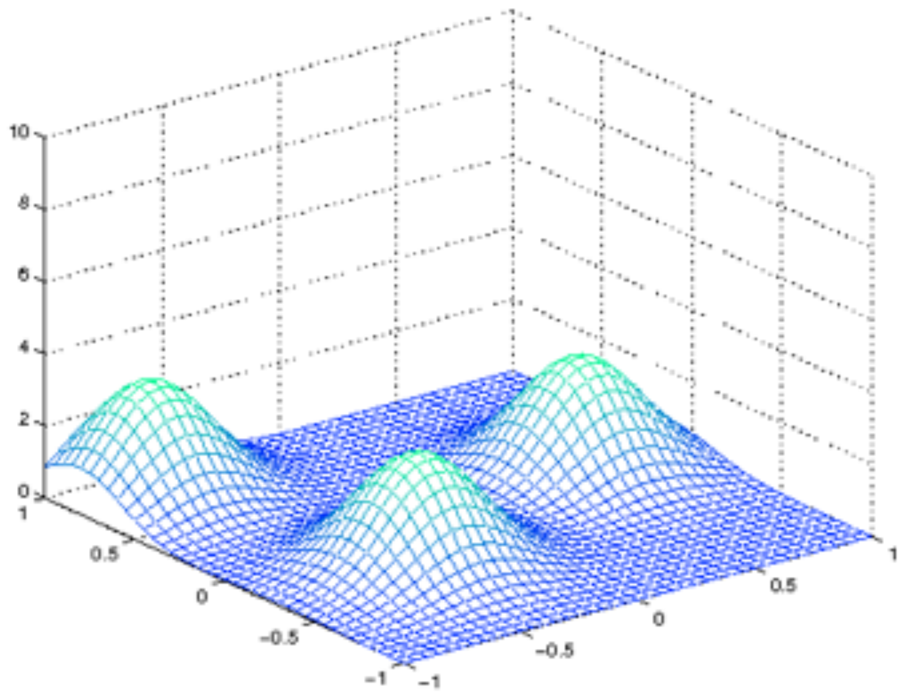




- Variational Bayes (VB)
  - Approximation $q^*(\theta)$ for posterior $p(\theta|x)$
  - "Close": Minimize Kullback-Liebler (KL) divergence:

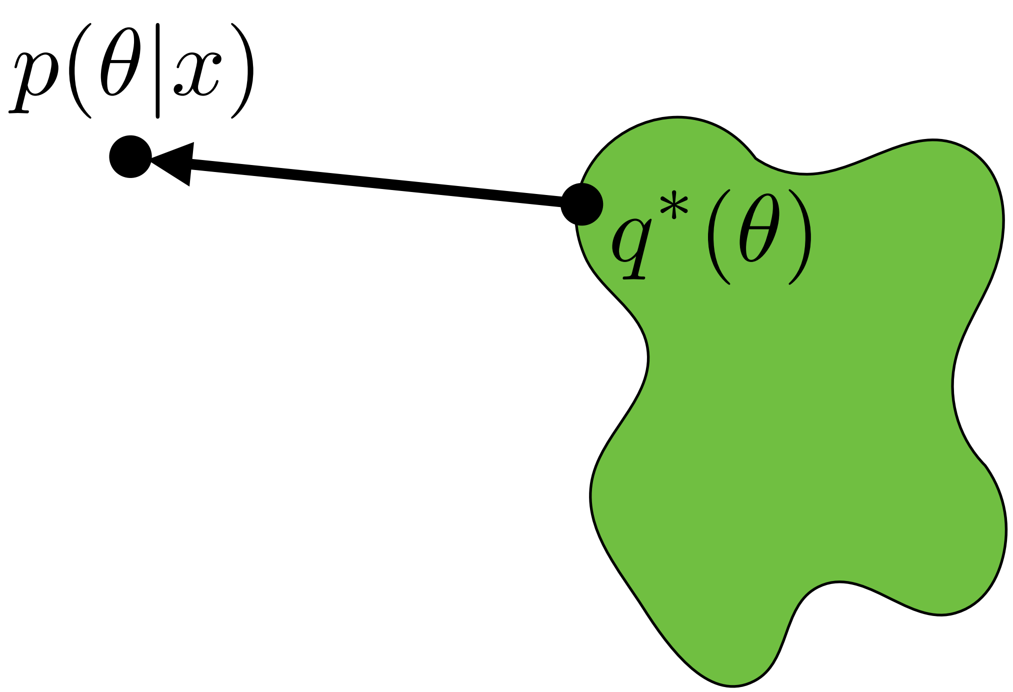
$$KL(q\|p(\cdot|x))$$

Variational Bayes

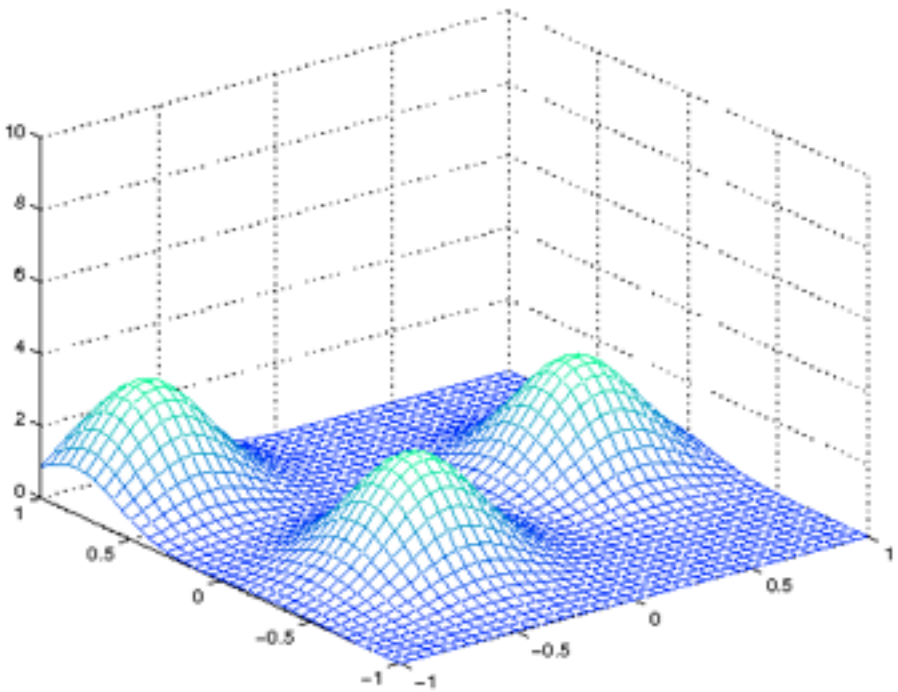


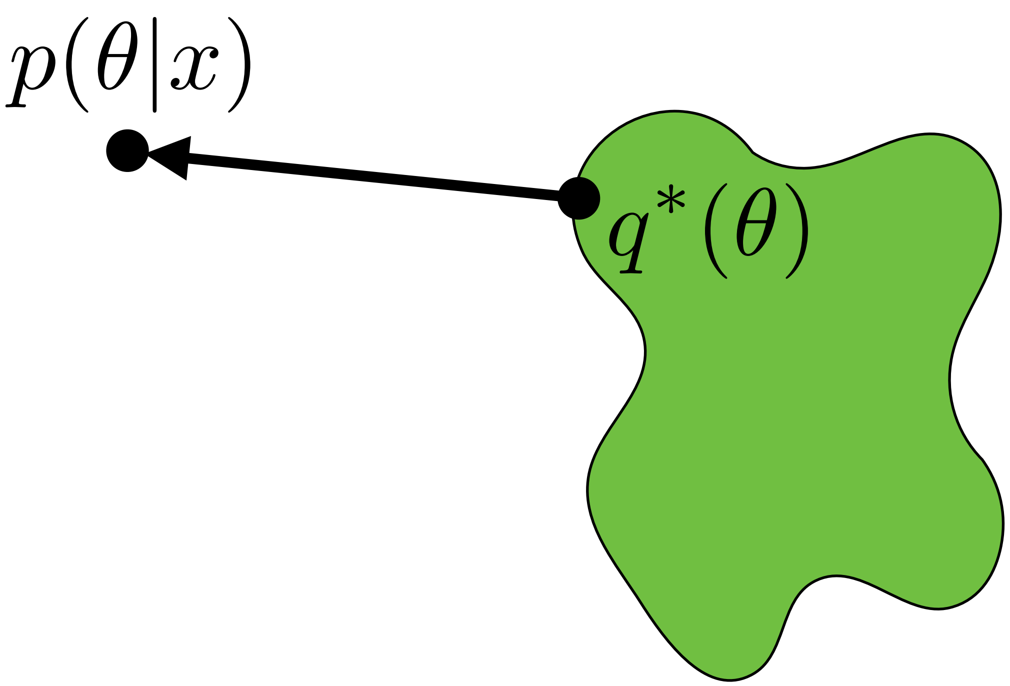
- Variational Bayes (VB)
  - Approximation $q^*(\theta)$ for posterior $p(\theta|x)$
  - "Close": Minimize Kullback-Liebler (KL) divergence:

$$KL(q\|p(\cdot|x))$$



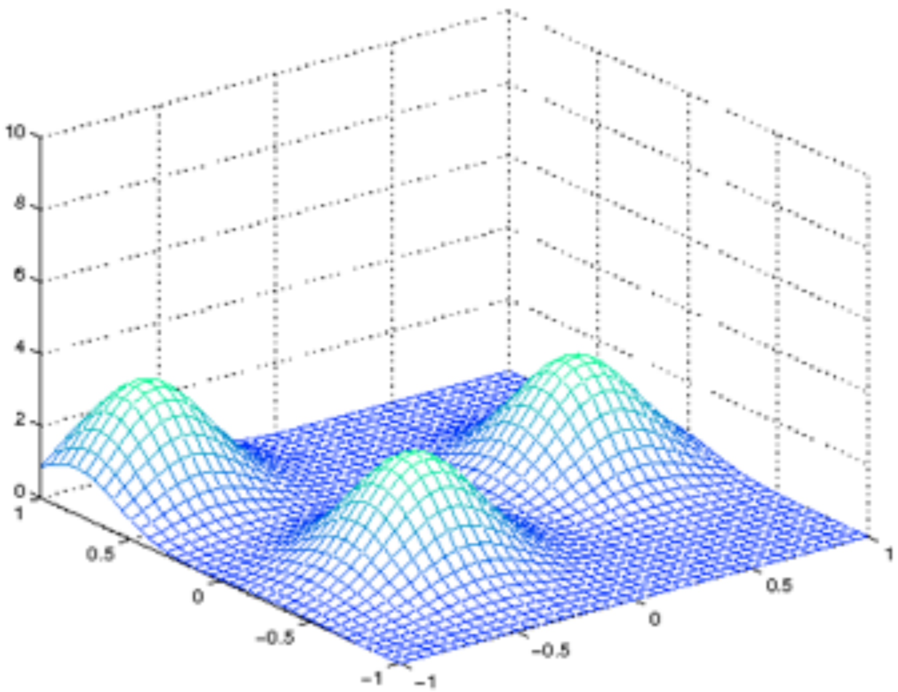
Variational Bayes

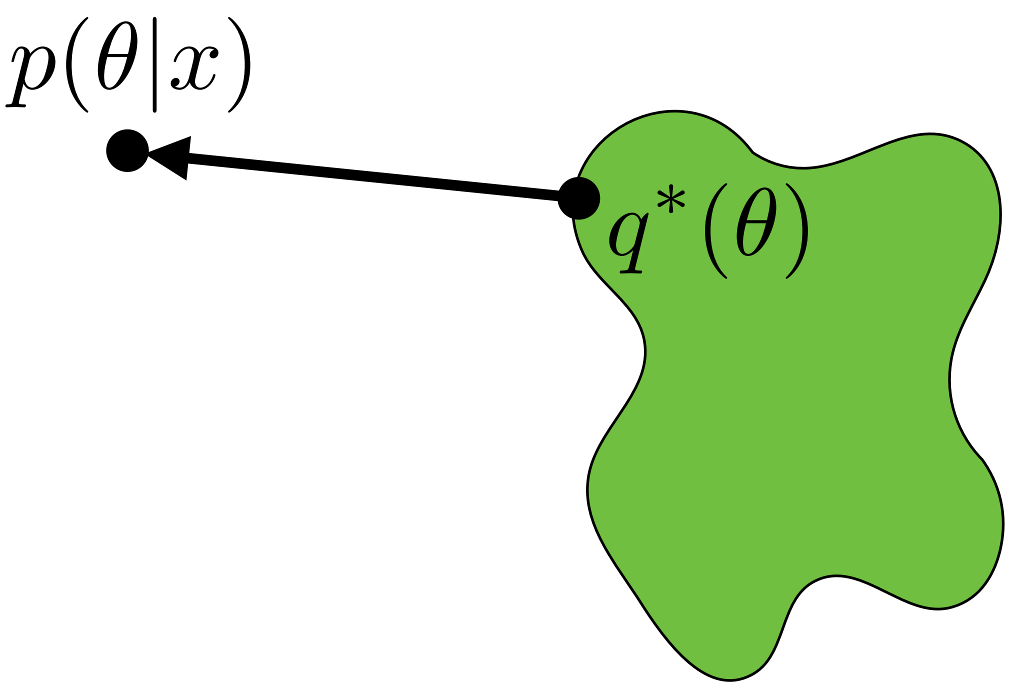




- Variational Bayes (VB)
  - Approximation $q^*(\theta)$ for posterior $p(\theta|x)$
  - "Close": Minimize Kullback-Liebler (KL) divergence:
    $$KL(q\|p(\cdot|x))$$
  - "Nice": factorizes, exponential family, truncation

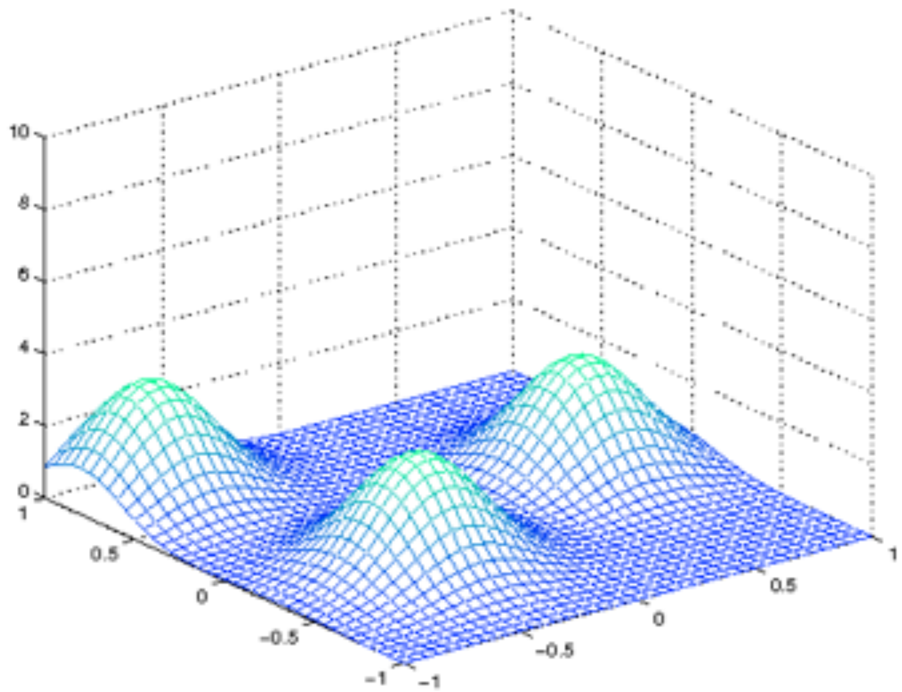
Variational Bayes

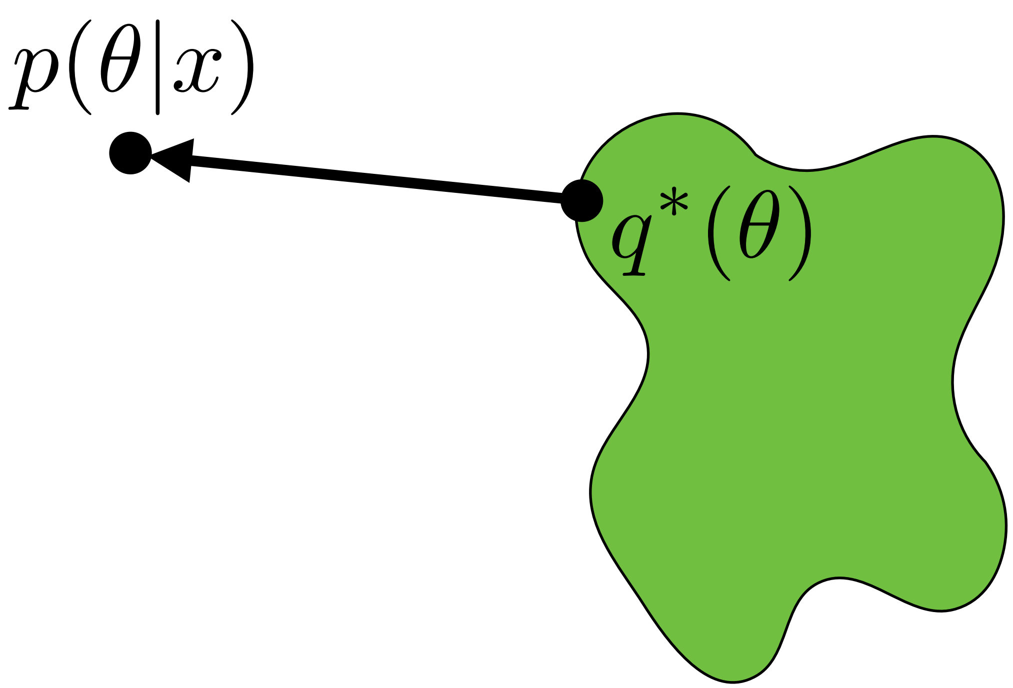




- Variational Bayes (VB)
    - Approximation $q^*(\theta)$ for posterior $p(\theta|x)$
    - "Close": Minimize Kullback-Liebler (KL) divergence:
      $$KL(q\|p(\cdot|x))$$
    - "Nice": factorizes, exponential family, truncation
- VB practical success

Variational Bayes

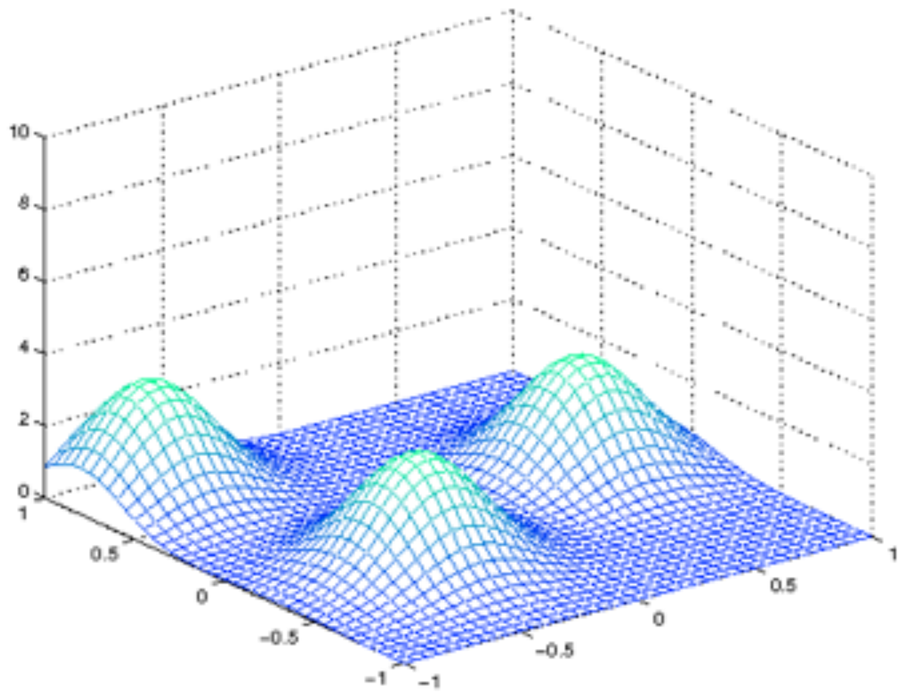


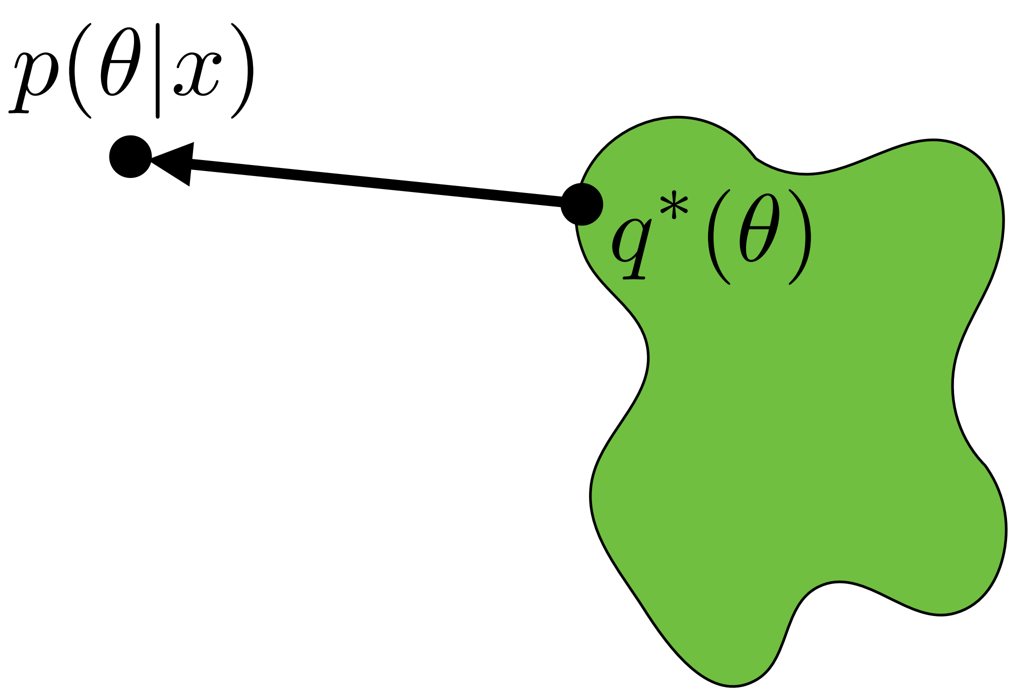


- Variational Bayes (VB)
  - Approximation $q^*(\theta)$ for posterior $p(\theta|x)$
  - "Close": Minimize Kullback-Liebler (KL) divergence:
    $$KL(q\|p(\cdot|x))$$
  - "Nice": factorizes, exponential family, truncation

- VB practical success
  - point estimates and prediction

Variational Bayes



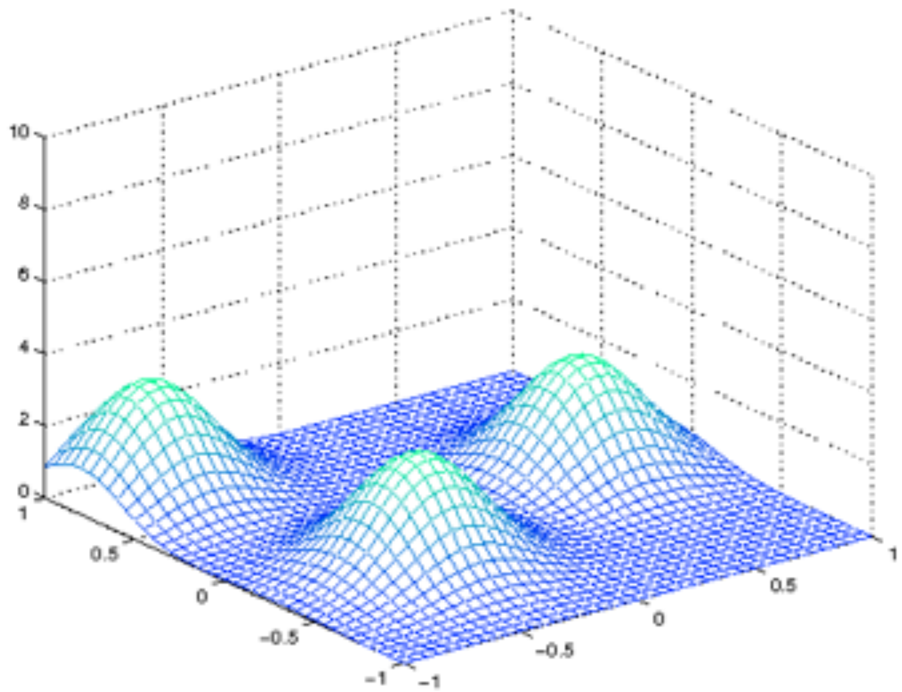


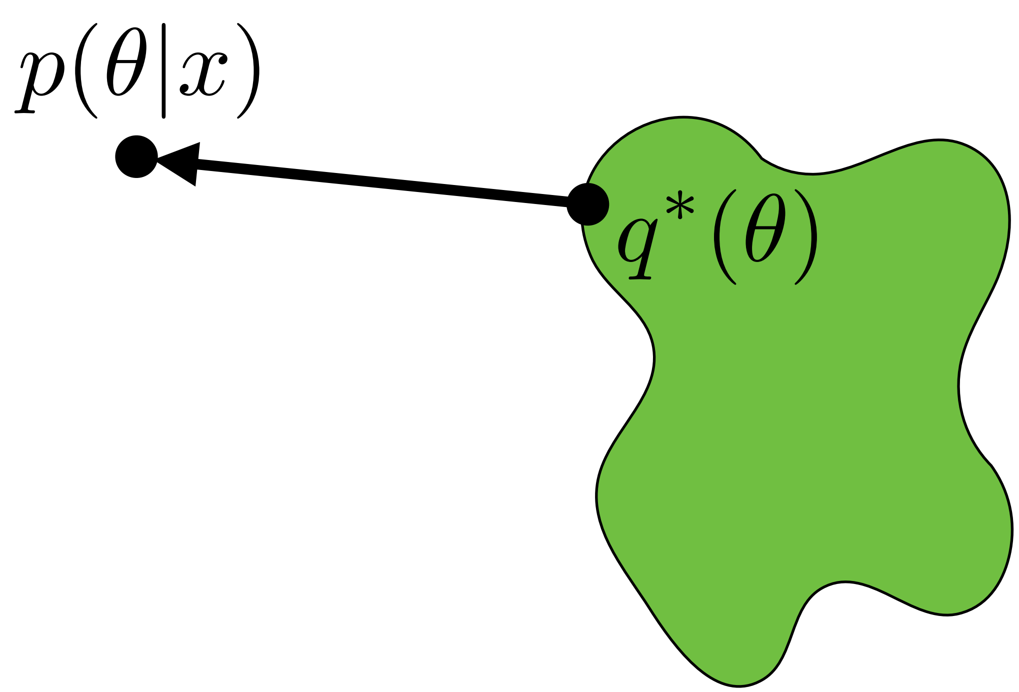
- Variational Bayes (VB)
  - Approximation $q^*(\theta)$ for posterior $p(\theta|x)$
  - "Close": Minimize Kullback-Liebler (KL) divergence:
    $$KL(q\|p(\cdot|x))$$
  - "Nice": factorizes, exponential family, truncation

- VB practical success
  - point estimates and prediction
  - fast, streaming, distributed

[Broderick, Boyd, Wibisono, Wilson, Jordan 2013]

Variational Bayes





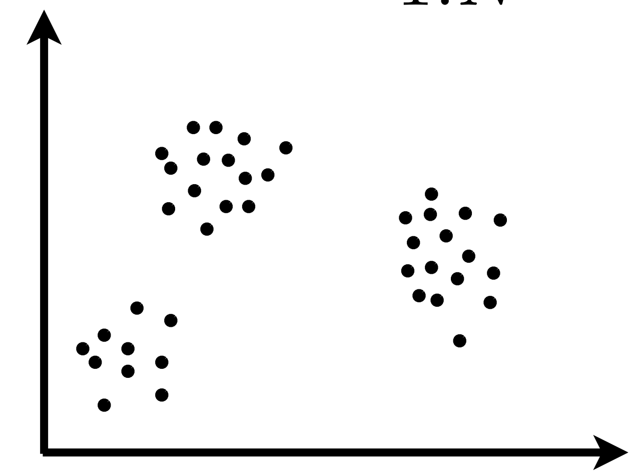
- Variational Bayes (VB)
  - Approximation $q^*(\theta)$ for posterior $p(\theta|x)$
  - "Close": Minimize Kullback-Liebler (KL) divergence:
    $$KL(q\|p(\cdot|x))$$
  - "Nice": factorizes, exponential family, truncation

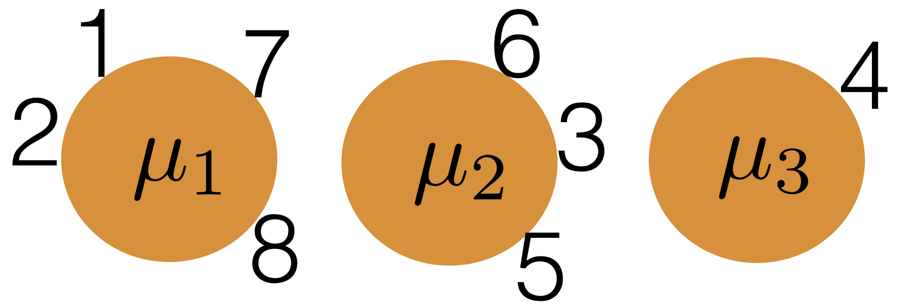
- VB practical success
  - point estimates and prediction
  - fast, streaming, distributed
  - Linear response VB (LRVB) for accurate covariance

[Broderick, Boyd, Wibisono, Wilson, Jordan 2013; Giordano, Broderick, Jordan 2015]

Exercises

- Data $x_{1:N}$

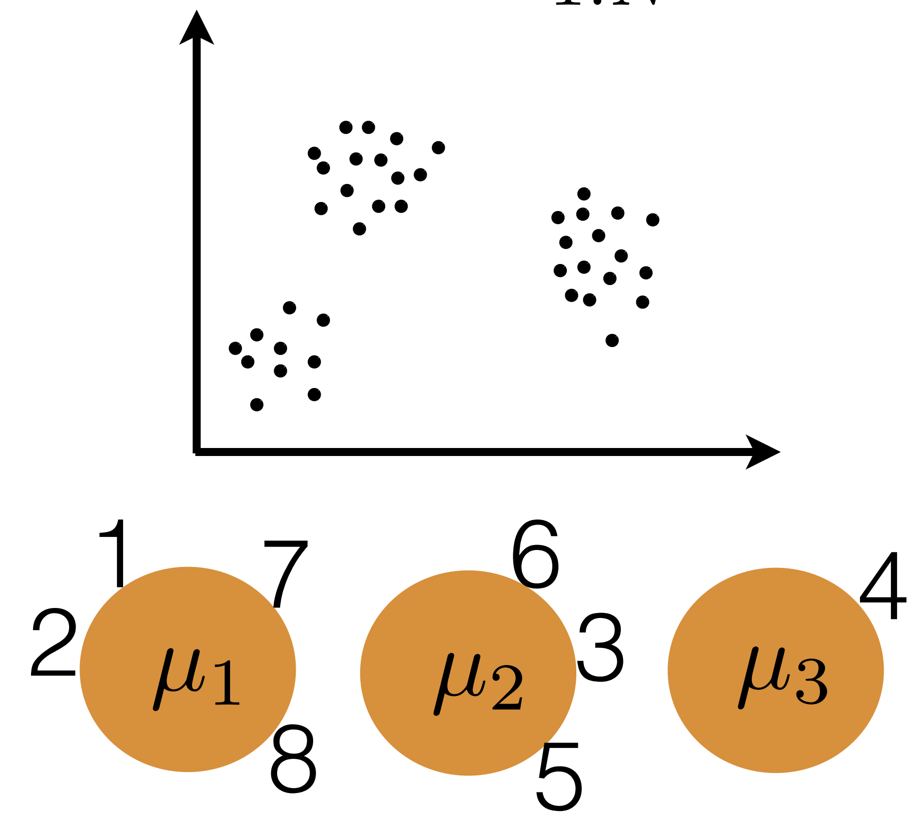




Exercises

- Code a CRP mixture model simulator

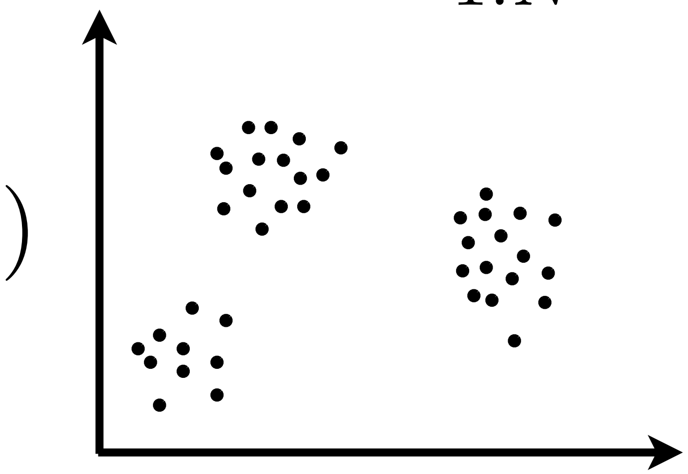
- Data $x_{1:N}$

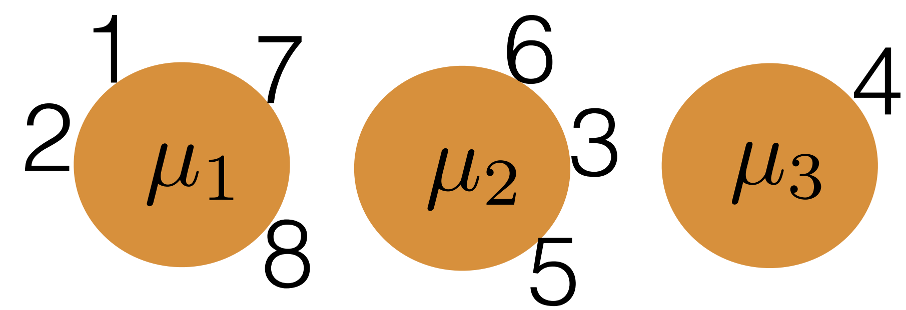


Exercises

- Code a CRP mixture model simulator
- Derive the CRP mixture model Gibbs sampler in the slides; derive $p(x_{C \cup \{n\}} | x_C)$ explicitly for a Gaussian mixture

- Data $x_{1:N}$

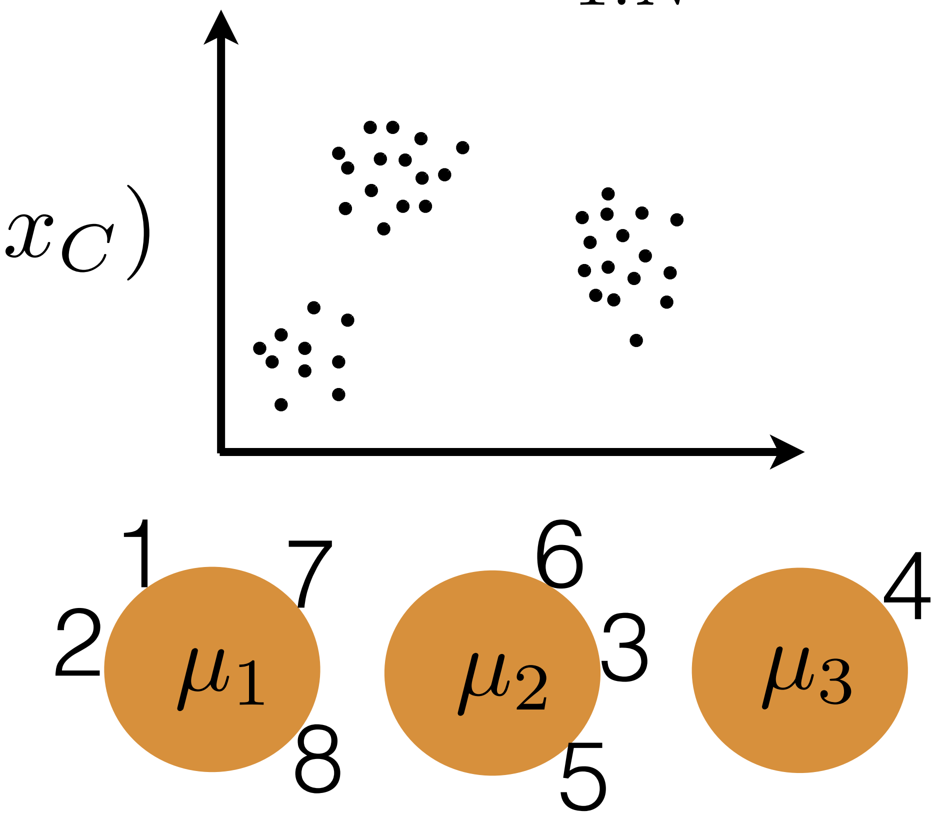




Exercises

- Code a CRP mixture model simulator
- Derive the CRP mixture model Gibbs sampler in the slides; derive $p(x_{C \cup \{n\}} | x_C)$ explicitly for a Gaussian mixture
- Extend the CRP mixture model Gibbs sampler in the slides to sample the cluster-specific parameters as well

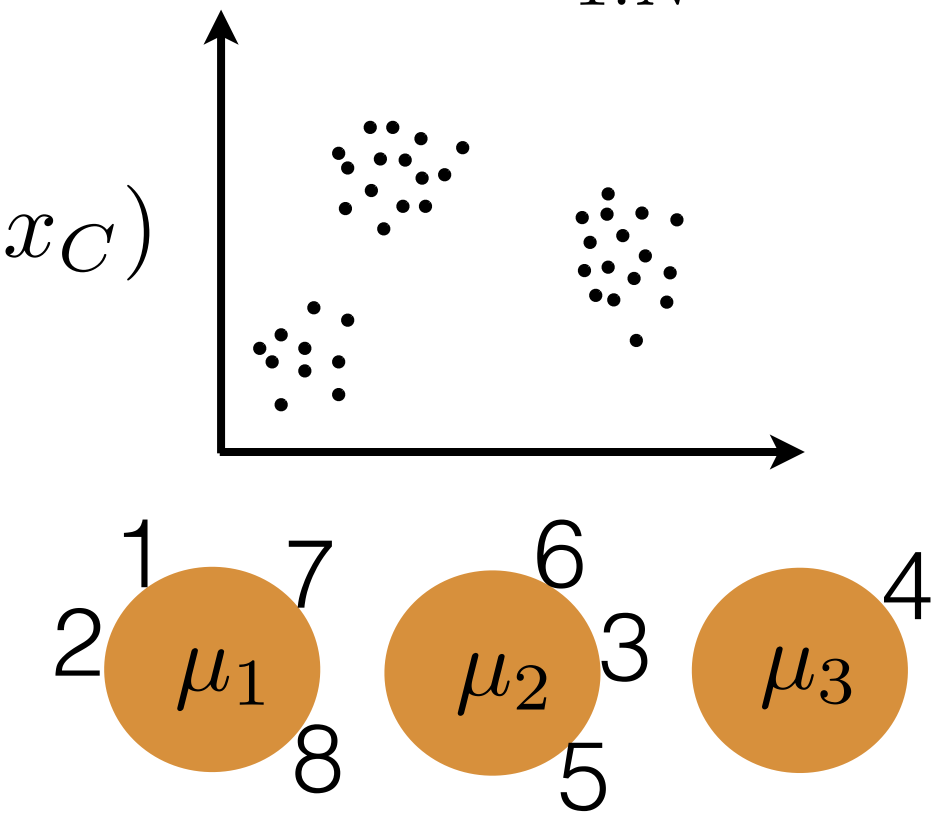
- Data $x_{1:N}$



Exercises

- Code a CRP mixture model simulator
- Derive the CRP mixture model Gibbs sampler in the slides; derive $p(x_{C \cup \{n\}} | x_C)$ explicitly for a Gaussian mixture
- Extend the CRP mixture model Gibbs sampler in the slides to sample the cluster-specific parameters as well
- Review Gibbs sampling, slice sampling [Neal 2003], variational Bayes [Bishop 2006]

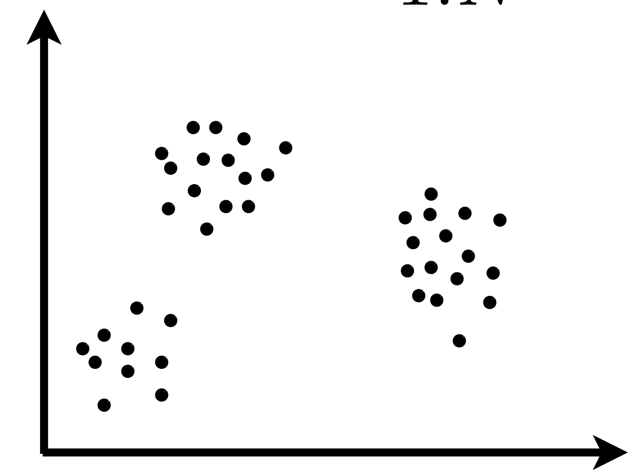
- Data $x_{1:N}$

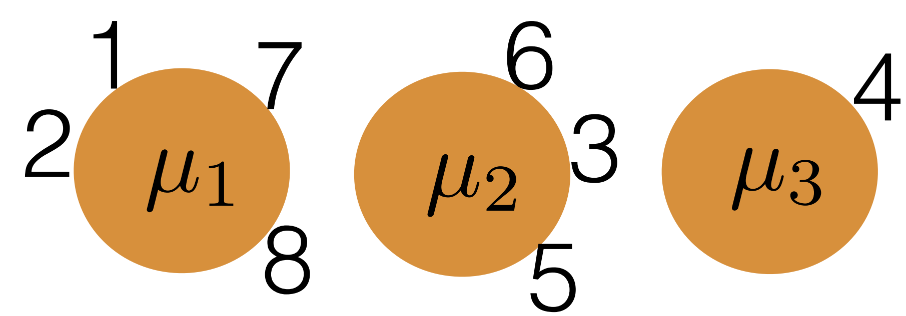


Exercises

- Code a CRP mixture model simulator
- Derive the CRP mixture model Gibbs sampler in the slides; derive $p(x_{C \cup \{n\}} | x_C)$ explicitly for a Gaussian mixture
- Extend the CRP mixture model Gibbs sampler in the slides to sample the cluster-specific parameters as well

- Data $x_{1:N}$



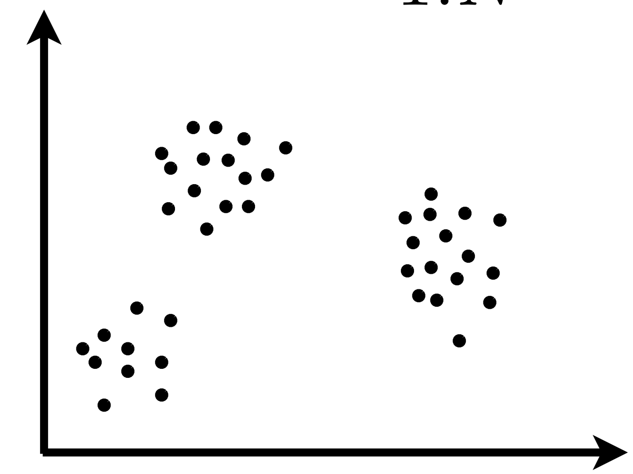


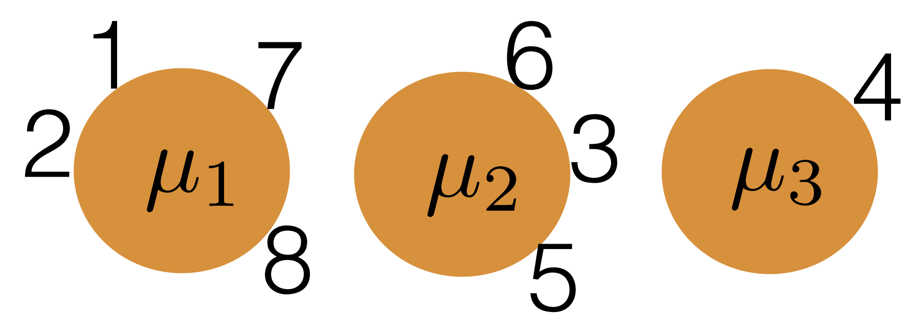
- Review Gibbs sampling, slice sampling [Neal 2003], variational Bayes [Bishop 2006]
- Read [Neal 2000] and code a DPMM Gibbs sampler

Exercises

- Code a CRP mixture model simulator
- Derive the CRP mixture model Gibbs sampler in the slides; derive $p(x_{C \cup \{n\}} | x_C)$ explicitly for a Gaussian mixture
- Extend the CRP mixture model Gibbs sampler in the slides to sample the cluster-specific parameters as well

- Data $x_{1:N}$



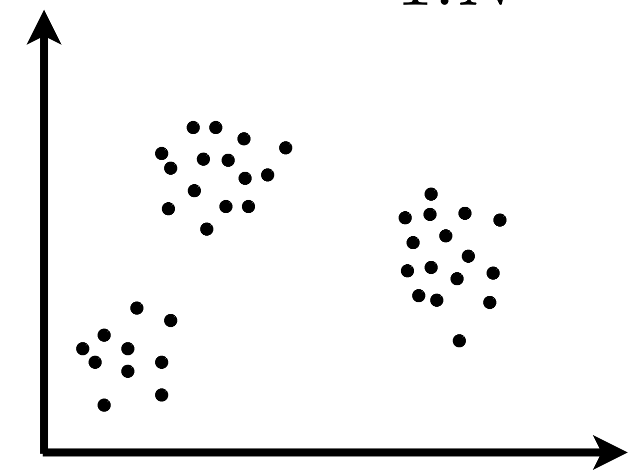


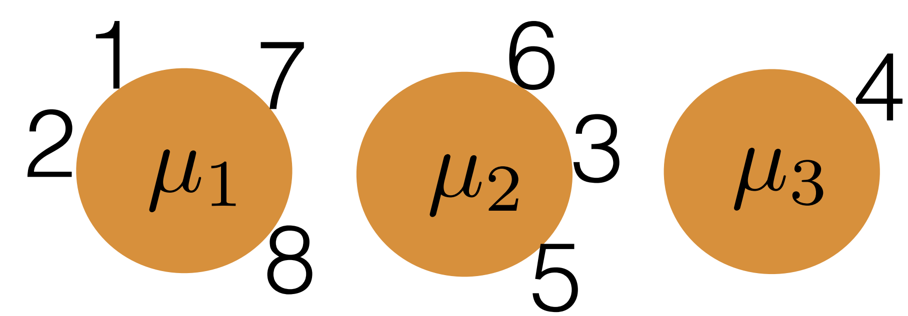
- Review Gibbs sampling, slice sampling [Neal 2003], variational Bayes [Bishop 2006]
- Read [Neal 2000] and code a DPMM Gibbs sampler
- Read [Walker 2007; Kalli, Griffin, Walker 2011] and code a DPMM slice sampler

Exercises

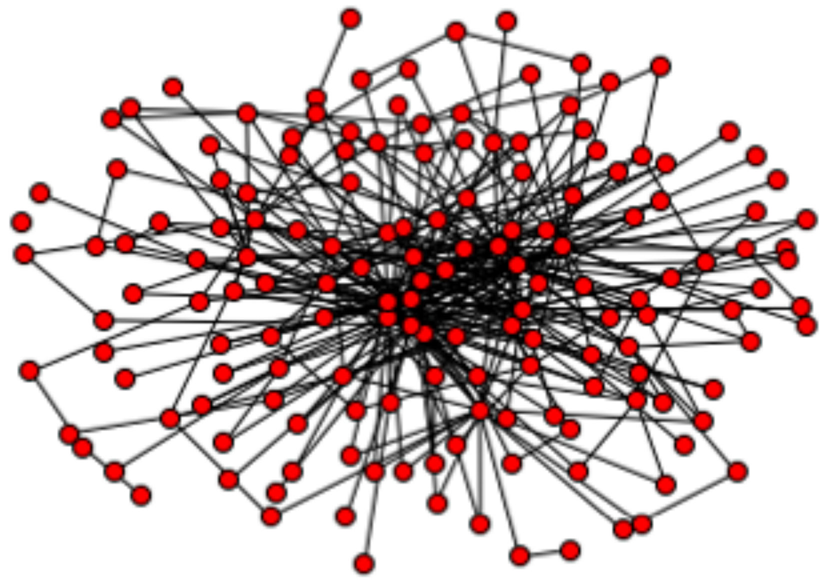
- Code a CRP mixture model simulator
- Derive the CRP mixture model Gibbs sampler in the slides; derive $p(x_{C \cup \{n\}} | x_C)$ explicitly for a Gaussian mixture
- Extend the CRP mixture model Gibbs sampler in the slides to sample the cluster-specific parameters as well
- Review Gibbs sampling, slice sampling [Neal 2003], variational Bayes [Bishop 2006]
- Read [Neal 2000] and code a DPMM Gibbs sampler
- Read [Walker 2007; Kalli, Griffin, Walker 2011] and code a DPMM slice sampler
- Read [Blei, Jordan 2006] and code variational inference for the DPMM

- Data $x_{1:N}$

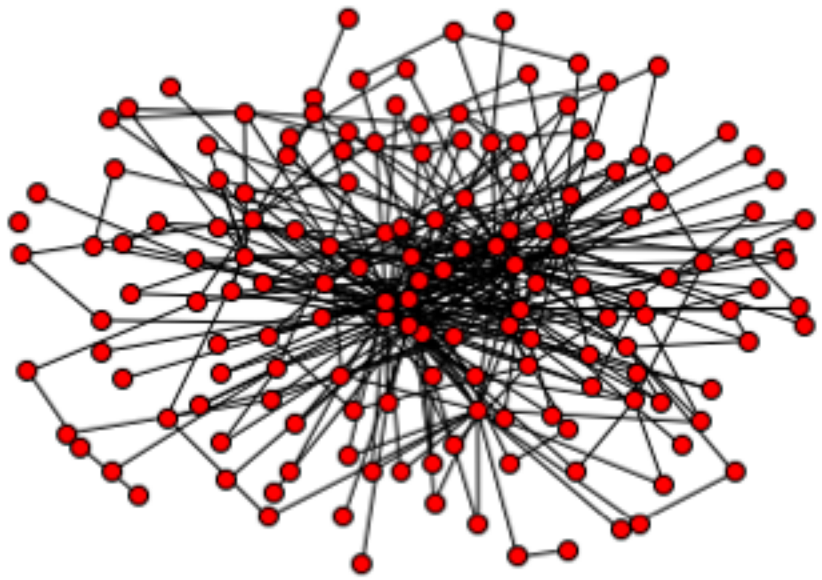




[Tiia Monto, https://commons.wikimedia.org/wiki/File:Trail_and_mountain.jpg]



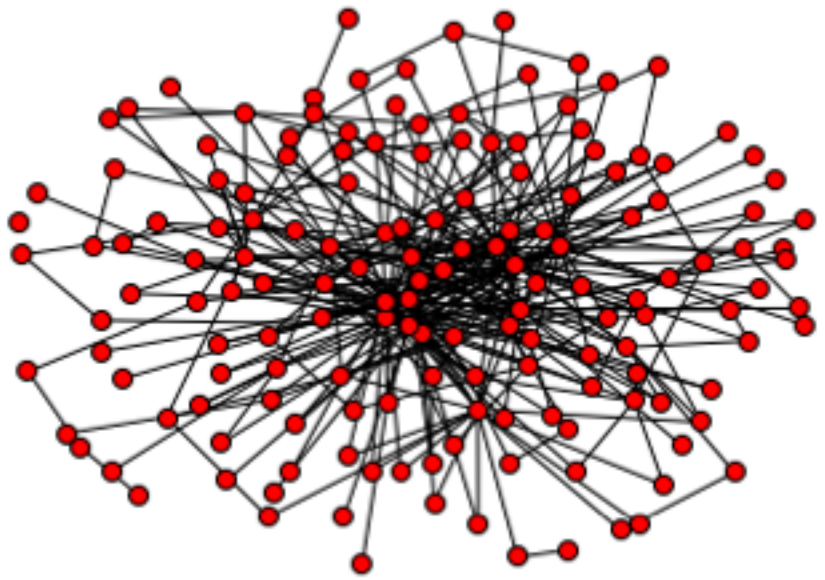
[Holland et al 1983; Kemp et al 2006; Xu et al 2007; Airoldi et al 2008; Lloyd et al 2012]



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]

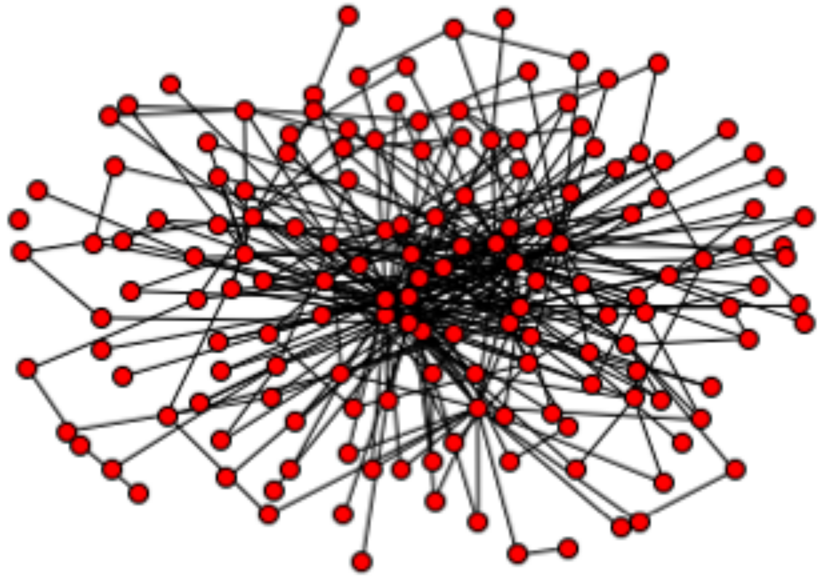
Probabilistic models for graphs

$p($  $)$

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

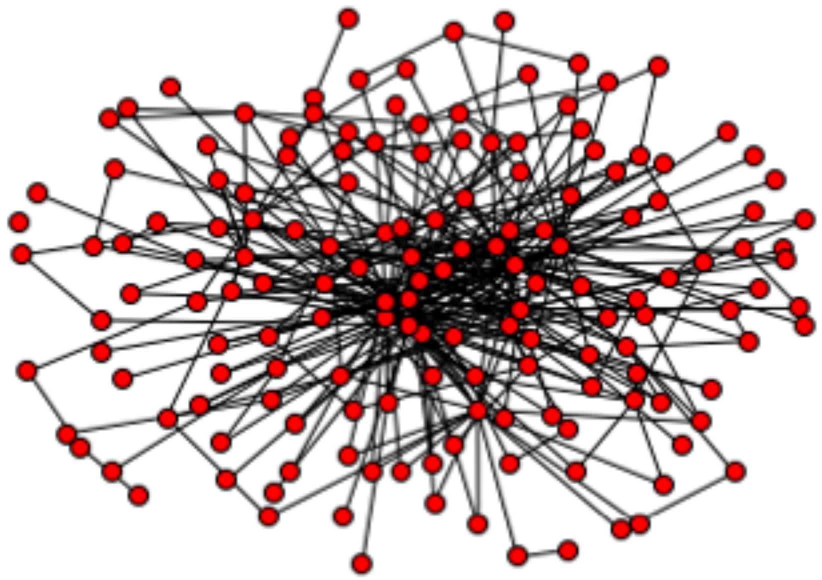
$p($  $)$

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

- Rich relationships, coherent uncertainties, prior 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

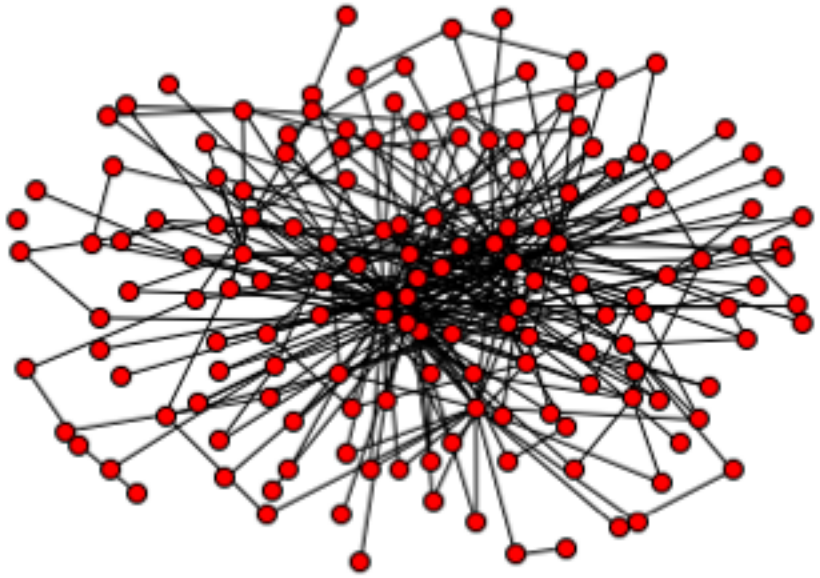
$p($  $)$

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

- Rich relationships, coherent uncertainties, prior 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]

Probabilistic models for graphs

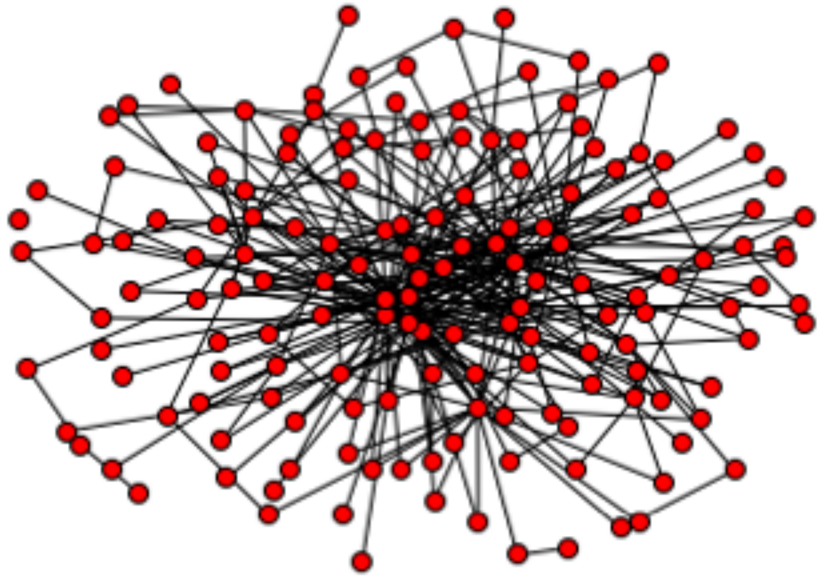
$p($  $)$

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

- Rich relationships, coherent uncertainties, prior 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

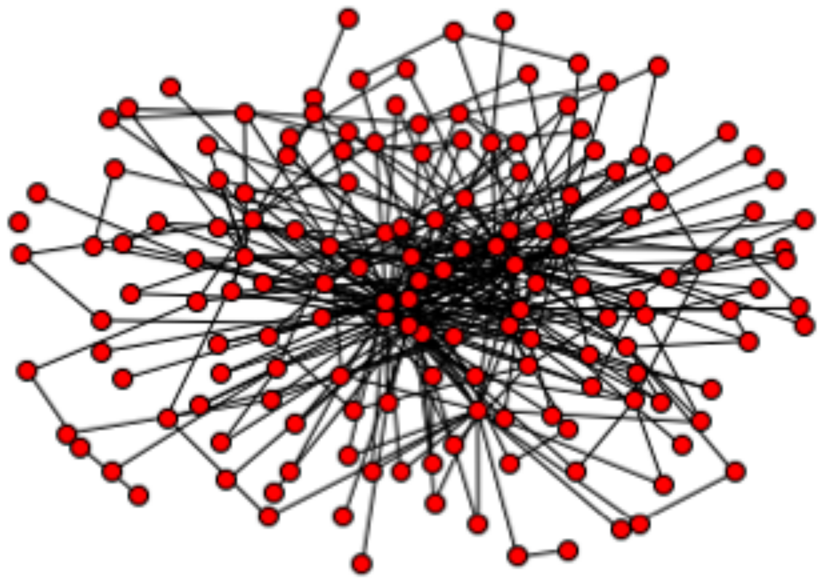
$p($  $)$

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

- Rich relationships, coherent uncertainties, prior 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

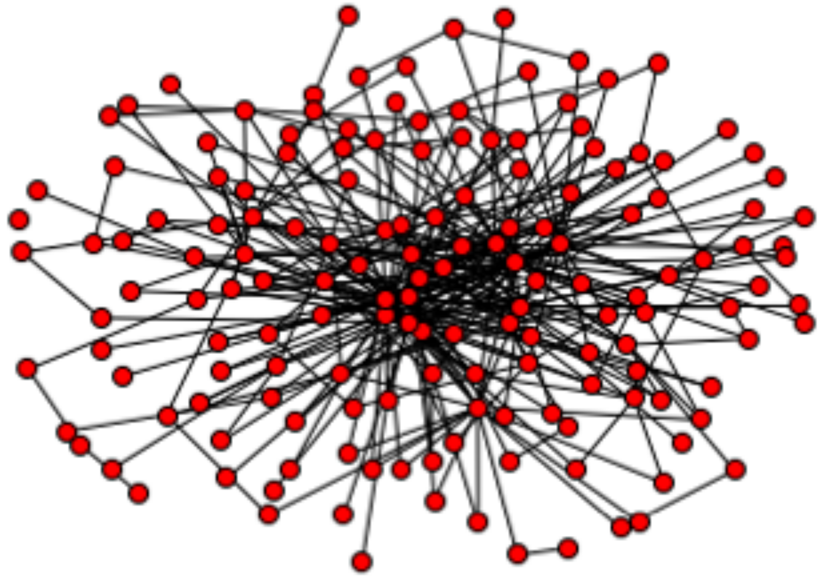
$p($  $)$

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

- Rich relationships, coherent uncertainties, prior 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

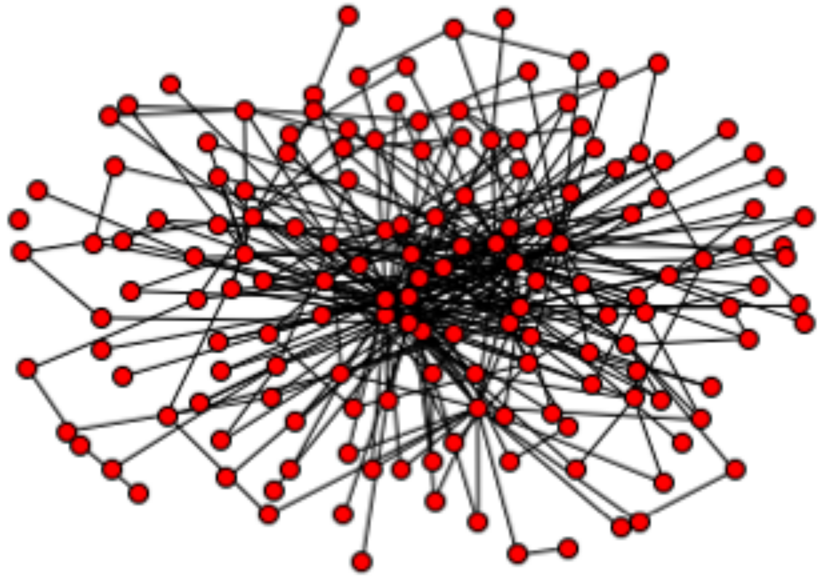
$p($  $)$

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

- Rich relationships, coherent uncertainties, prior 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
- **Our Solution**: a new framework for sparse 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

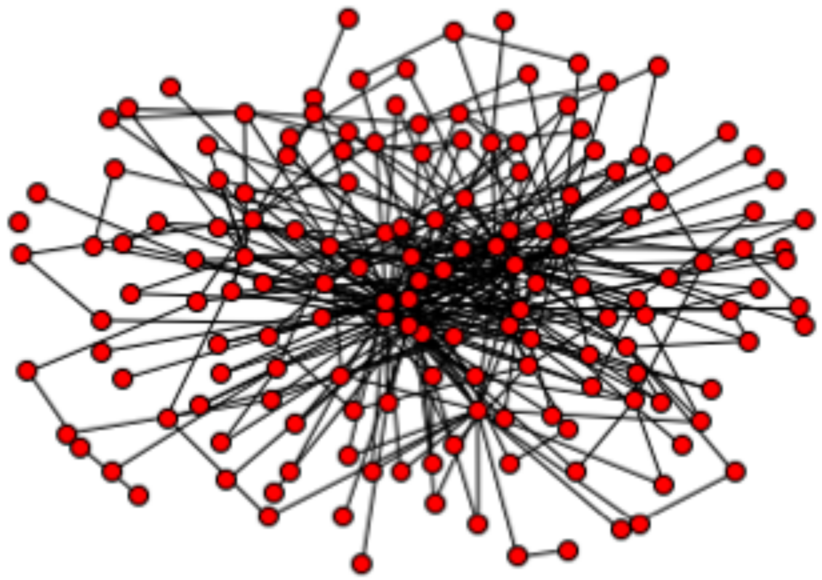
$p($  $)$

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

- Rich relationships, coherent uncertainties, prior 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**
- **Our Solution**: a new framework for sparse 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

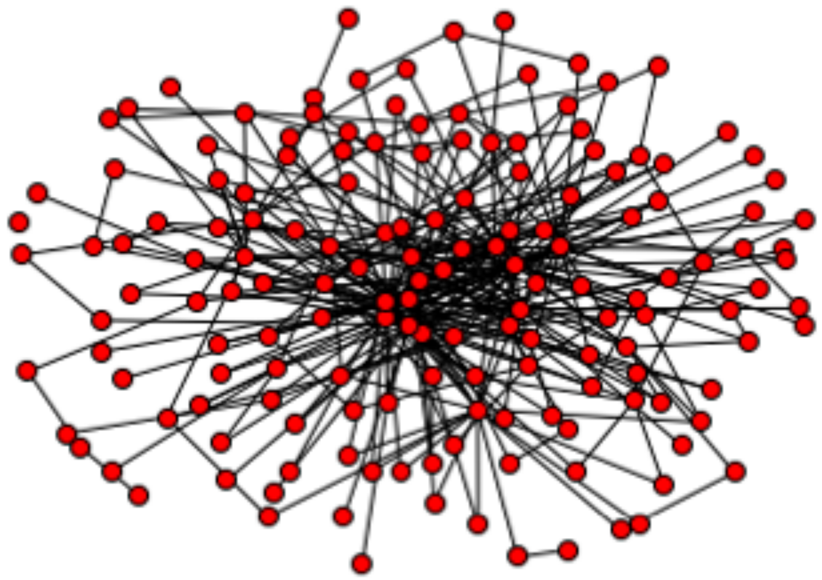
$p($  $)$

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

- Rich relationships, coherent uncertainties, prior 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
- **Our Solution**: a **new framework** for sparse 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

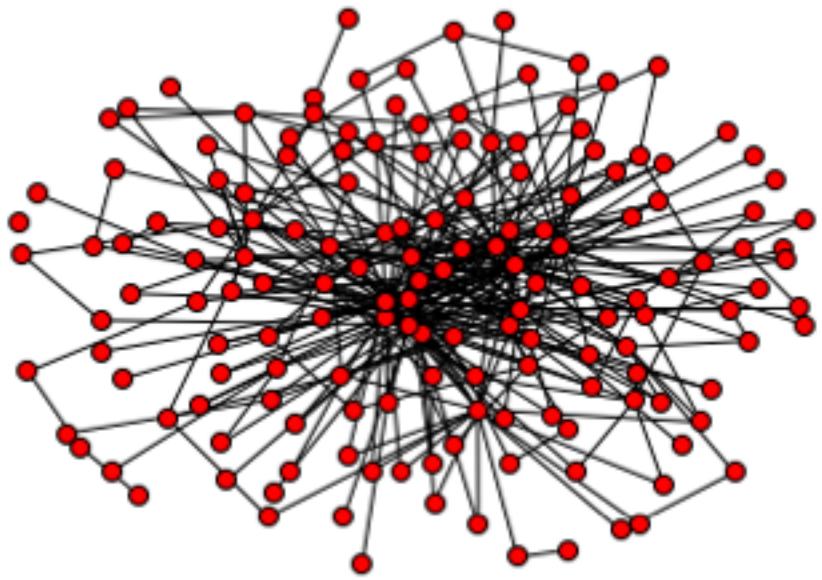
$p($  $)$

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

- Rich relationships, coherent uncertainties, prior 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
- **Our Solution**: a new framework for **sparse 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

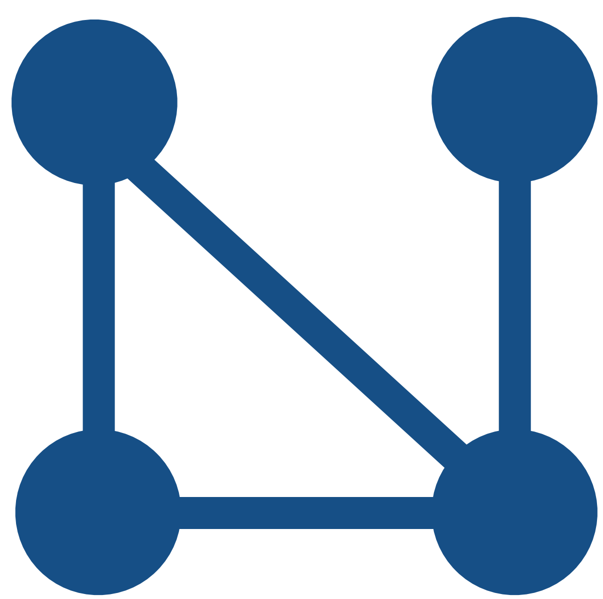
$p($  $)$

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

- Rich relationships, coherent uncertainties, prior 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
- **Our Solution**: a new framework for sparse graphs
  - Concurrent & independent graphs work by Crane & Dempsey

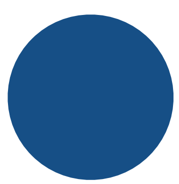
 [B, Cai 2015; Cai, B 2015a,b; Crane, Dempsey 2015a,b,16a,b; Cai, Campbell, B 2016; Campbell, Cai, B 2016]

Sequence of graphs

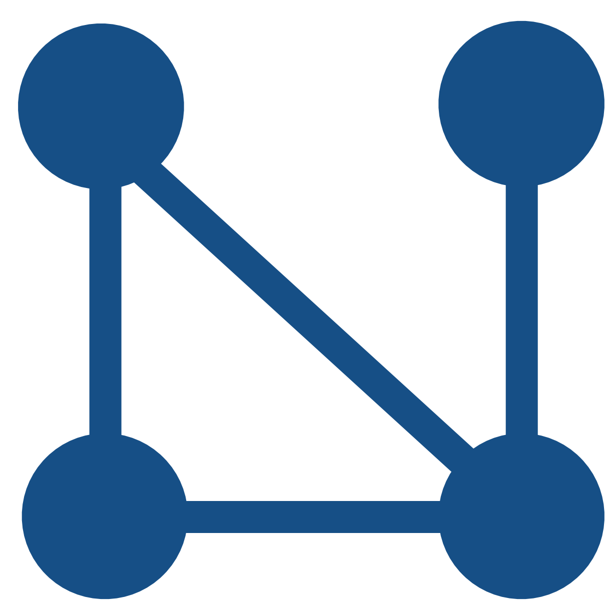


G

Sequence of graphs

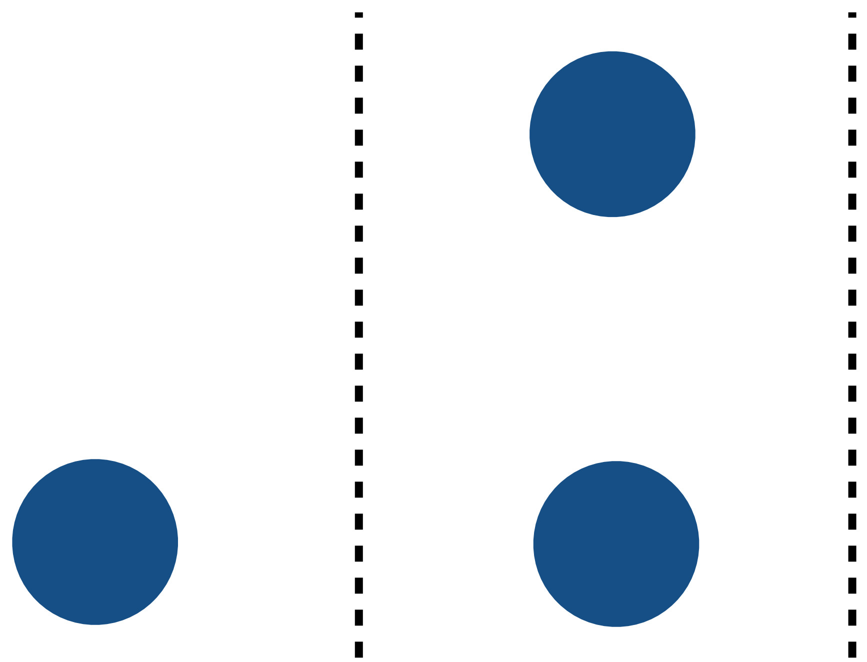


G_1

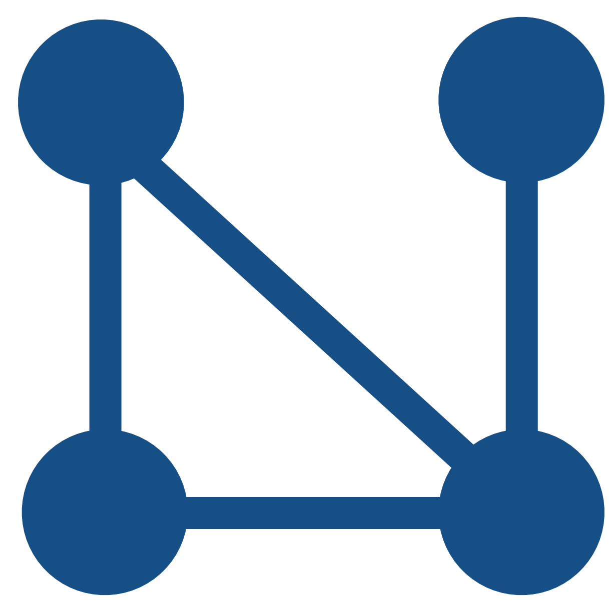


G

Sequence of graphs

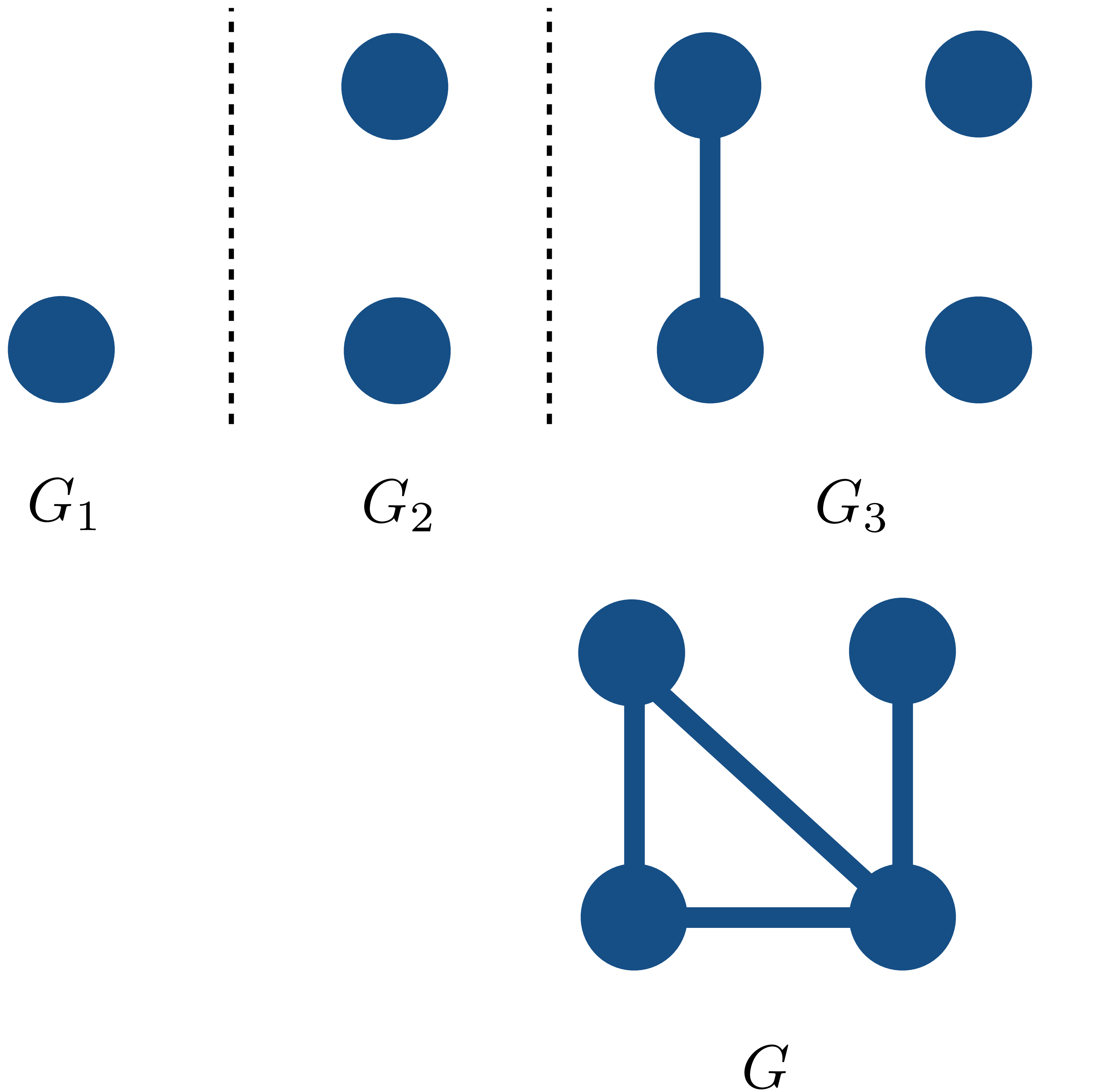


G_1 G_2

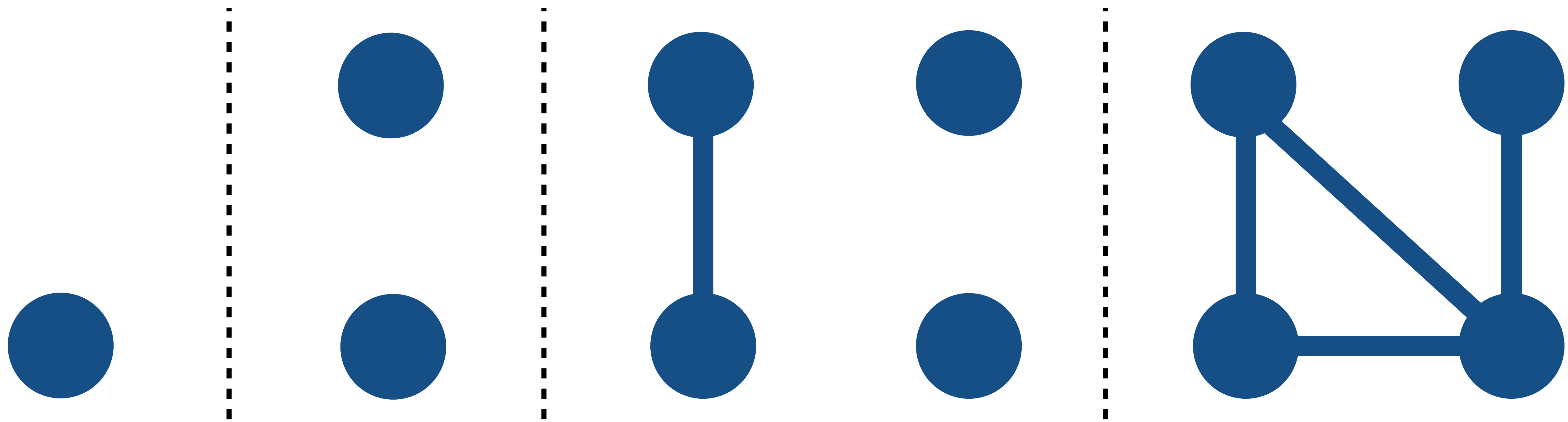


G

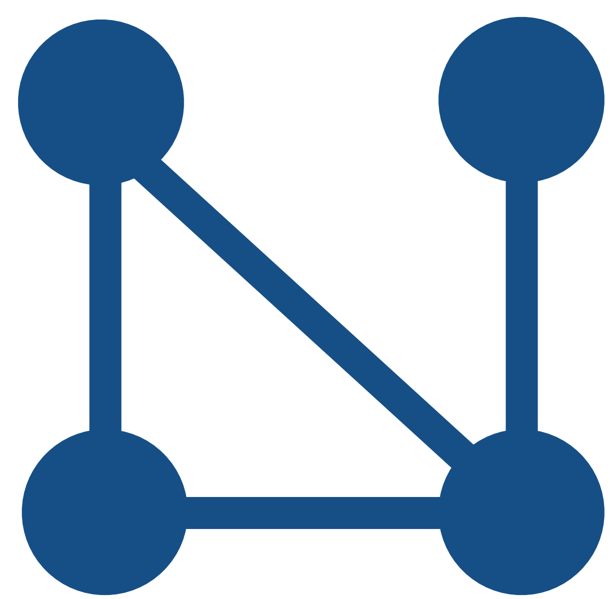
Sequence of graphs



Sequence of graphs

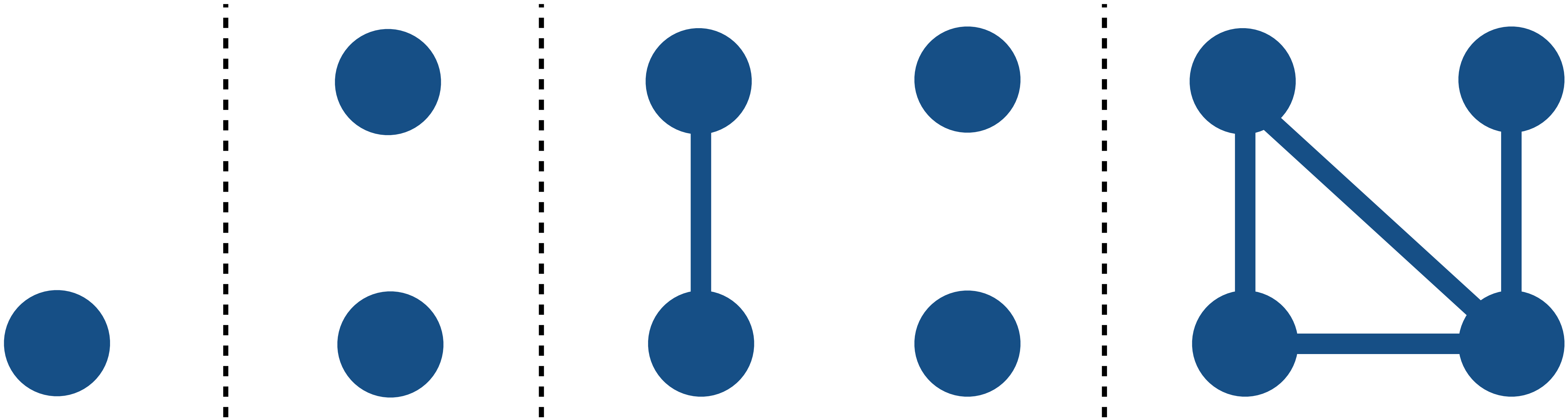


G_1 G_2 G_3 G_4



G

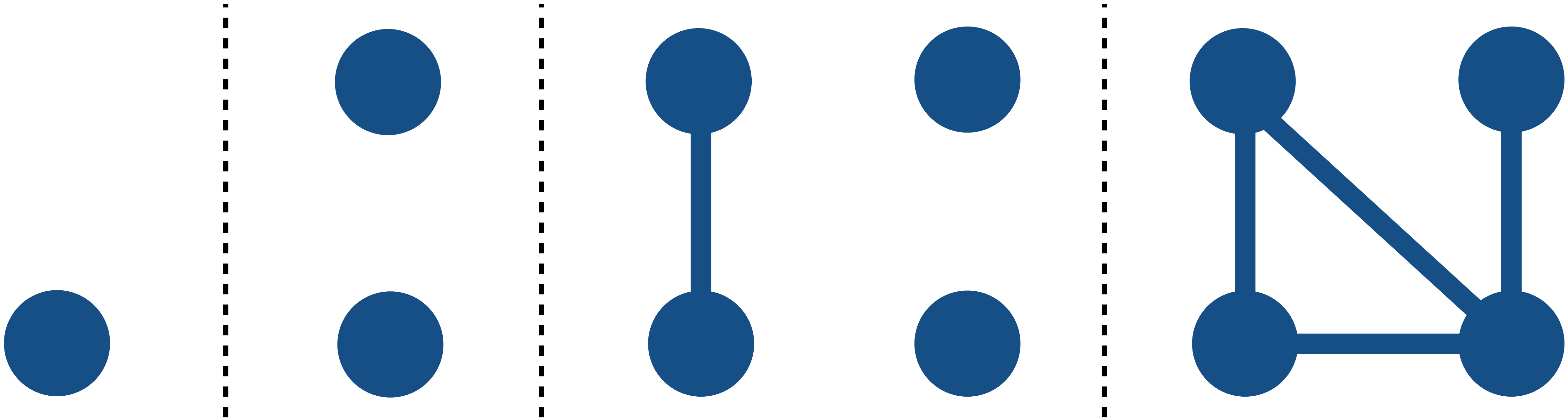
Sequence of graphs



G_1 G_2 G_3 G_4

...

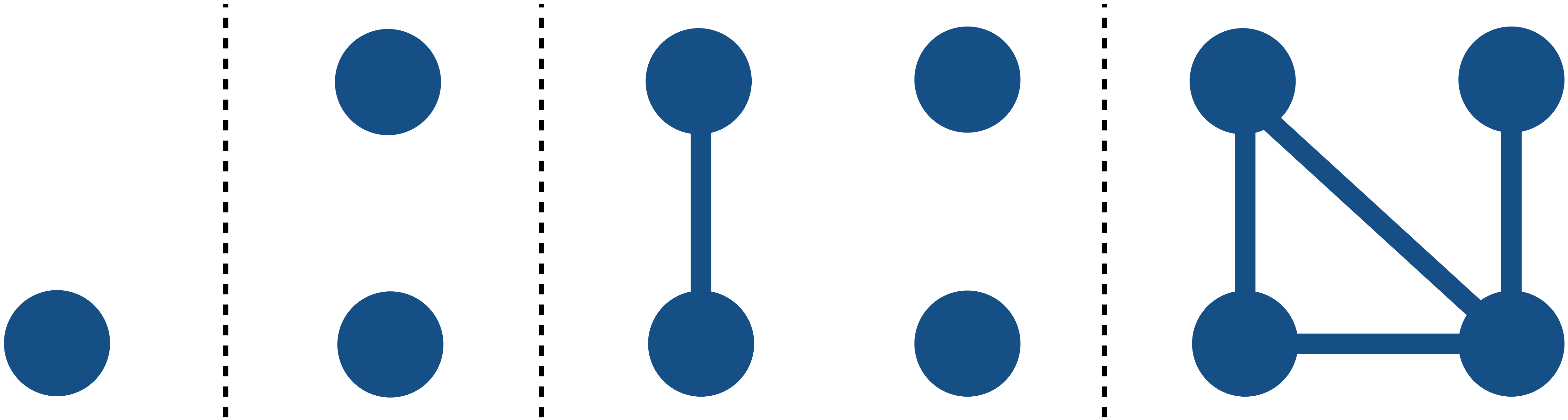
Sequence of graphs



G_1 G_2 G_3 G_4

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

Sequence of graphs

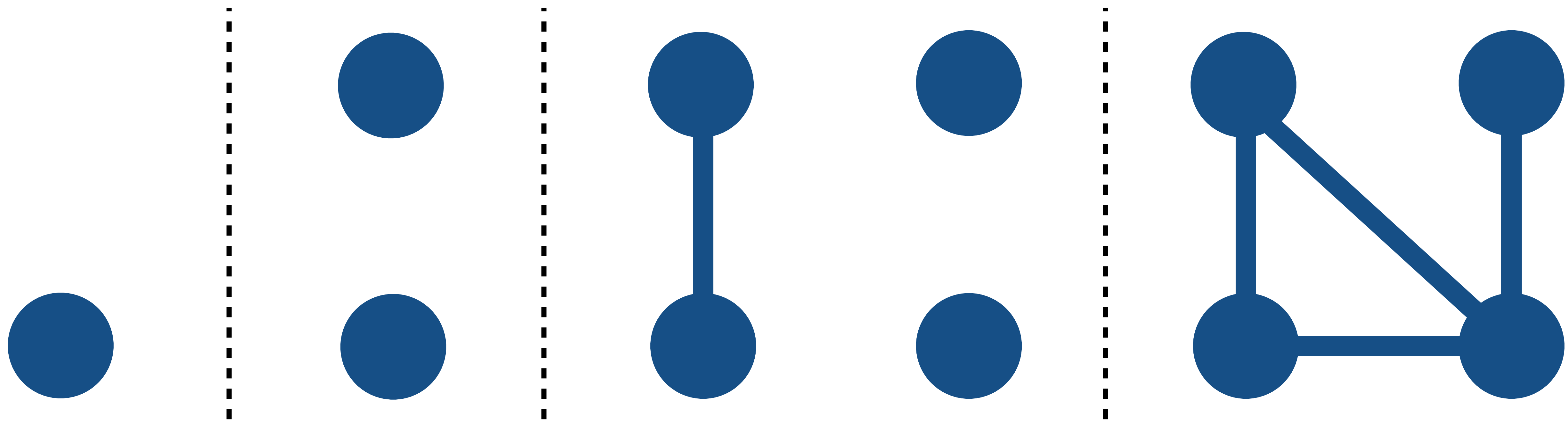


G_1 G_2 G_3 G_4

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

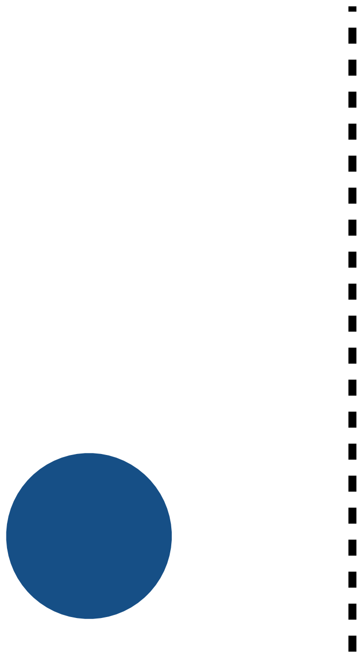


G_1 G_2 G_3 G_4

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 $\#\text{edges}(G_n) \in o([\#\text{nodes}(G_n)]^2)$

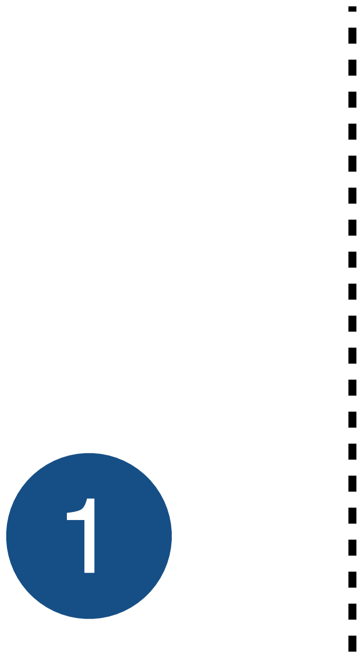
The Old Way: Nodes



G_1

[Hoover 1979, Aldous 1981]

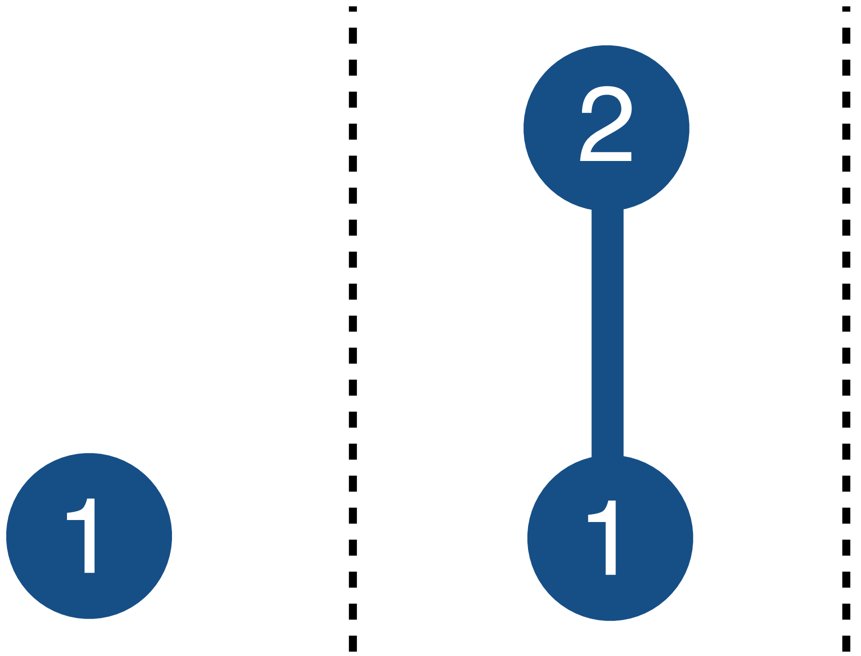
The Old Way: Nodes



G_1

[Hoover 1979, Aldous 1981]

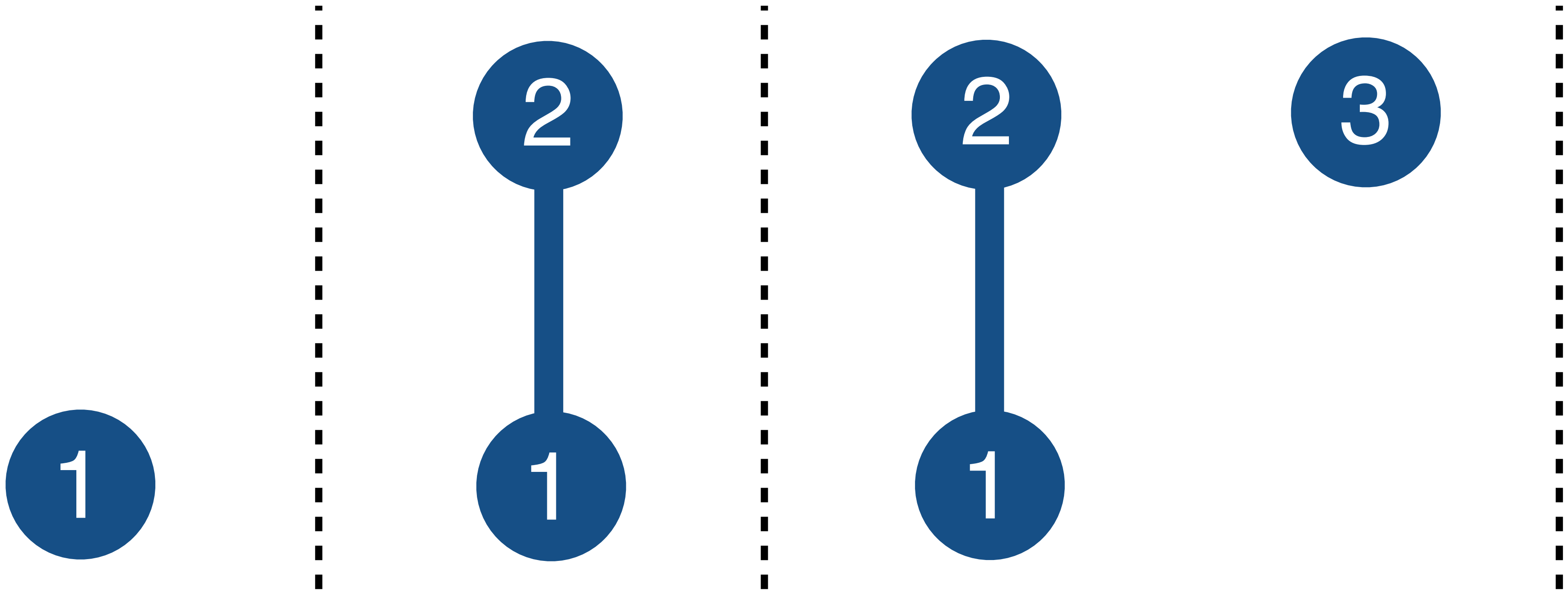
The Old Way: Nodes



G_1 G_2

[Hoover 1979, Aldous 1981]

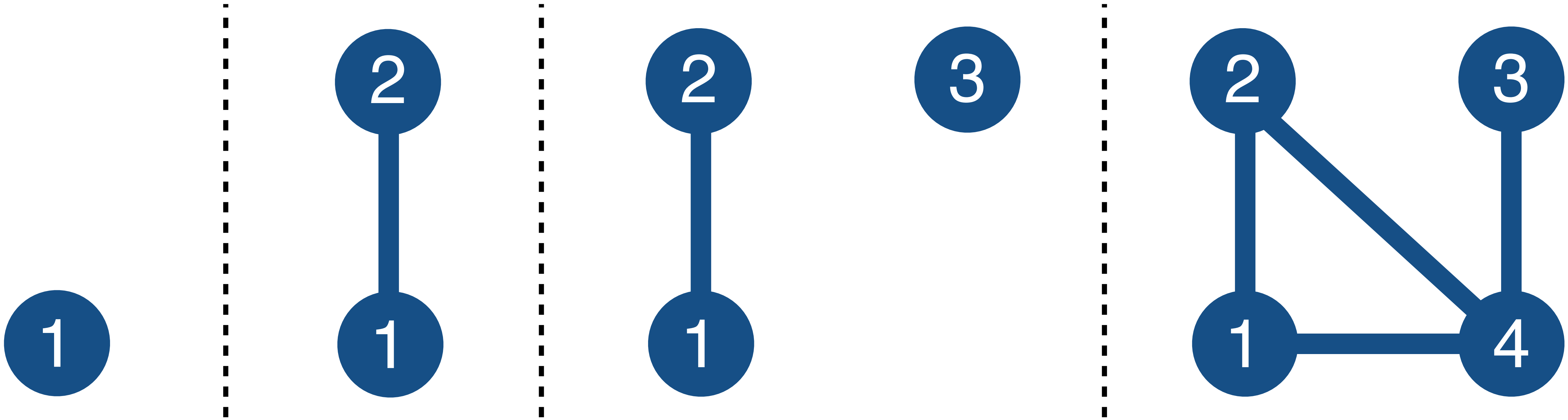
The Old Way: Nodes



G_1 G_2 G_3

[Hoover 1979, Aldous 1981]

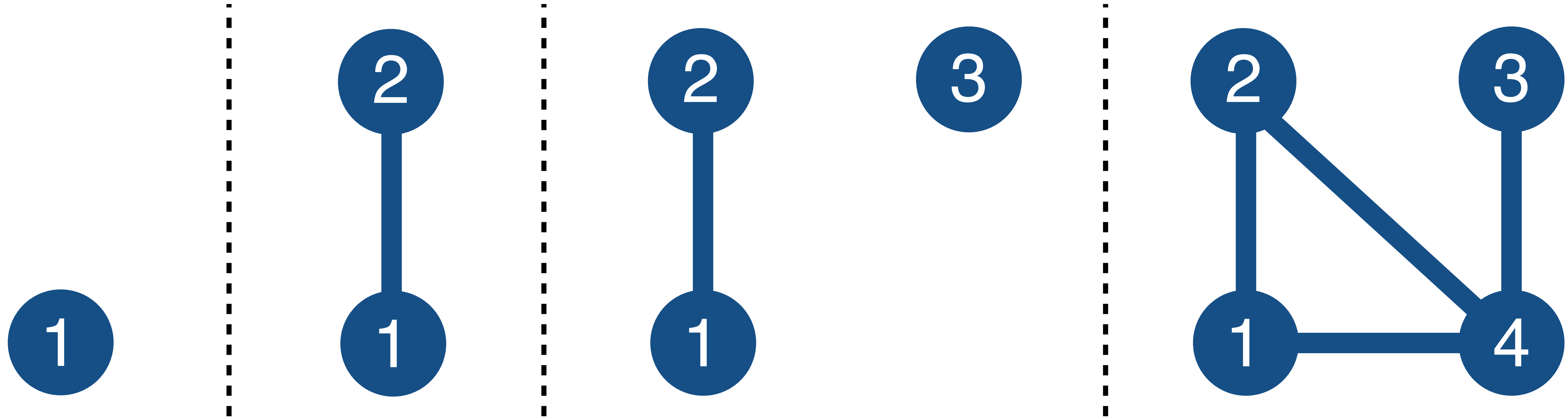
The Old Way: Nodes



G_1 G_2 G_3 G_4

[Hoover 1979, Aldous 1981]

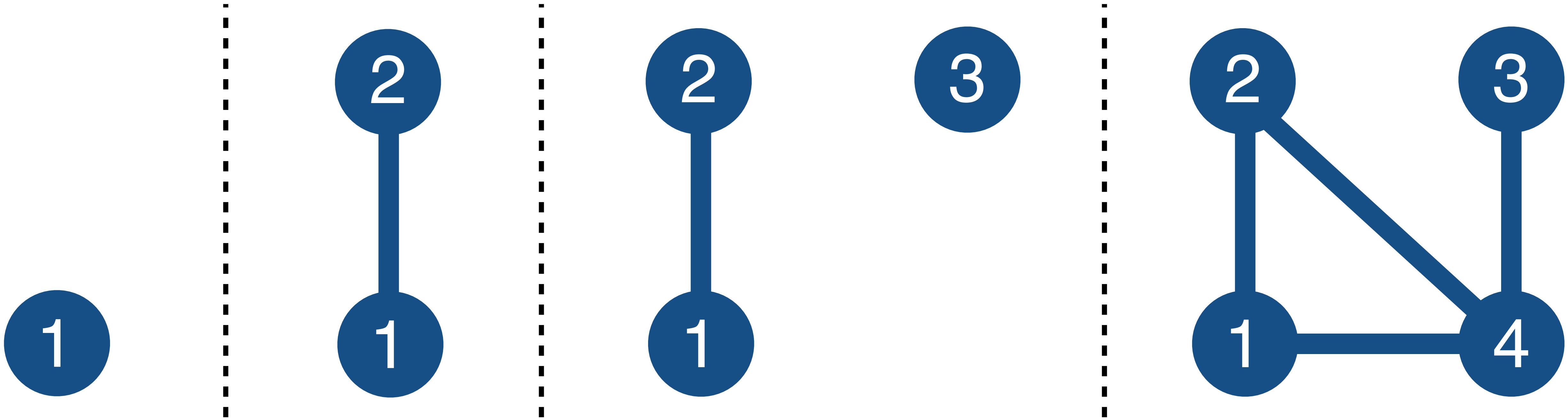
The Old Way: Exchangeability



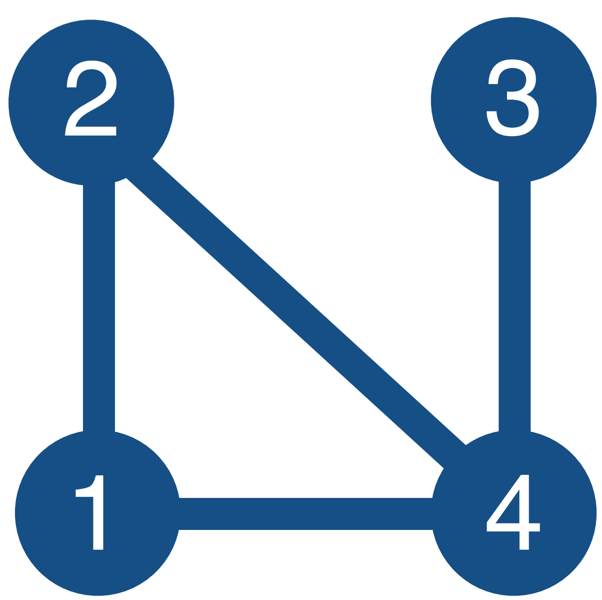
G_1 G_2 G_3 G_4

[Hoover 1979, Aldous 1981]

The Old Way: Exchangeability

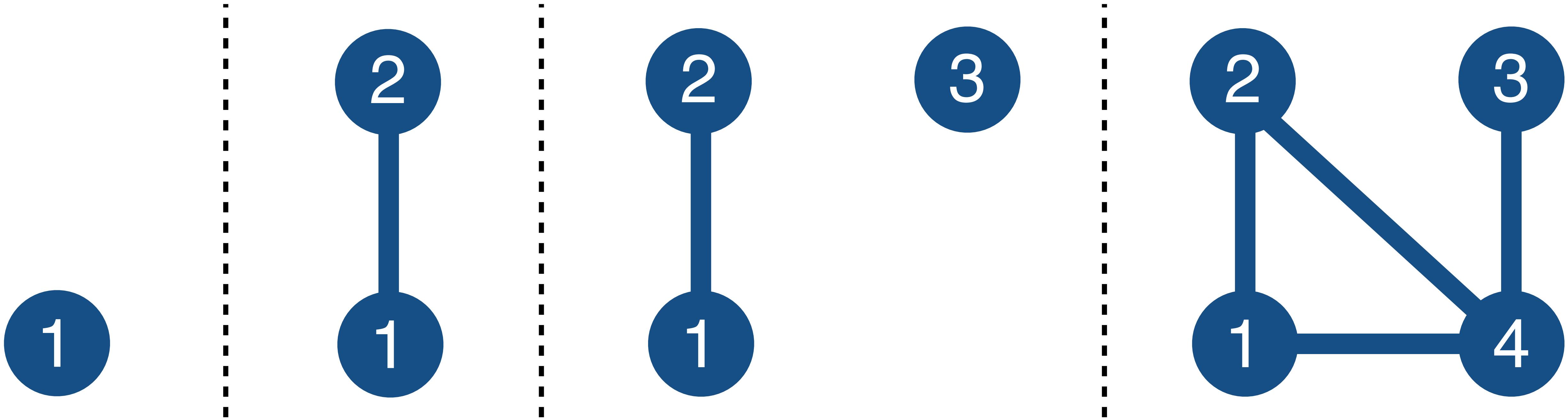


G_1 G_2 G_3 G_4

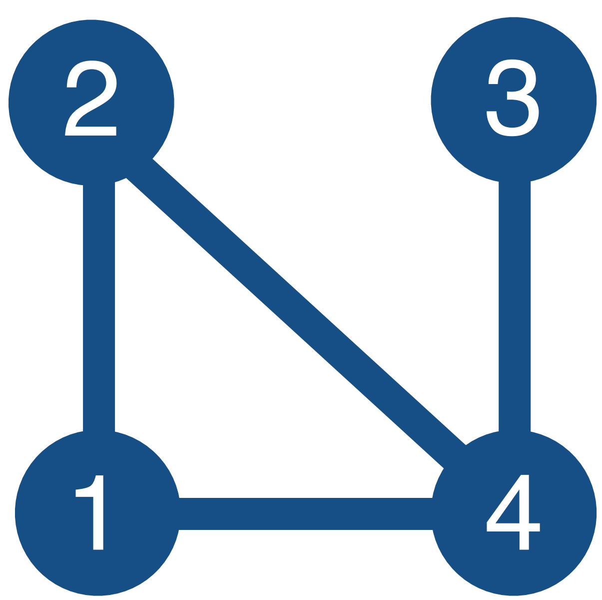


[Hoover 1979, Aldous 1981]

The Old Way: Exchangeability

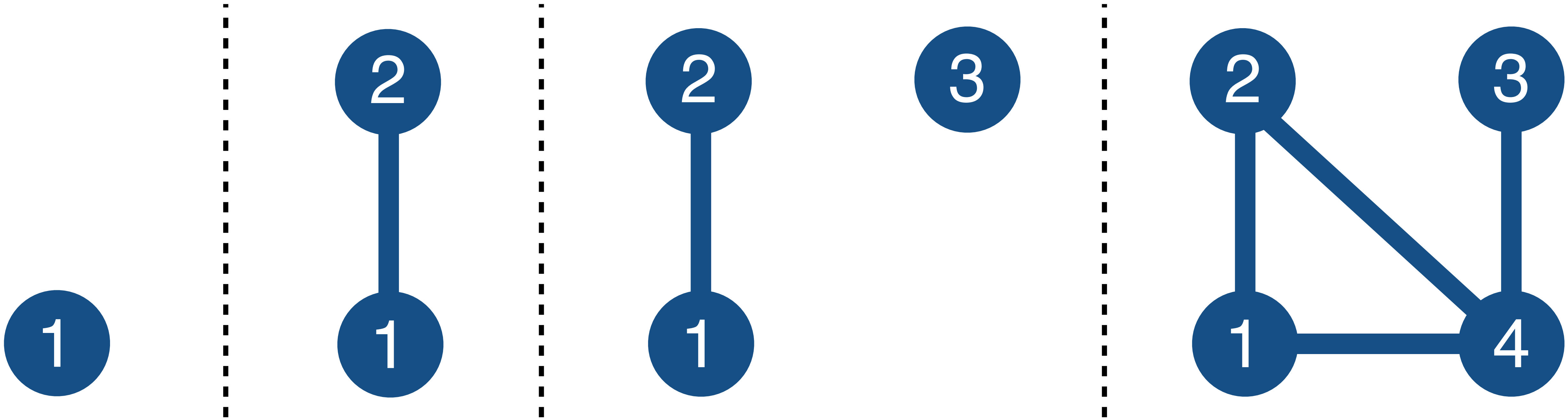


G_1 G_2 G_3 G_4

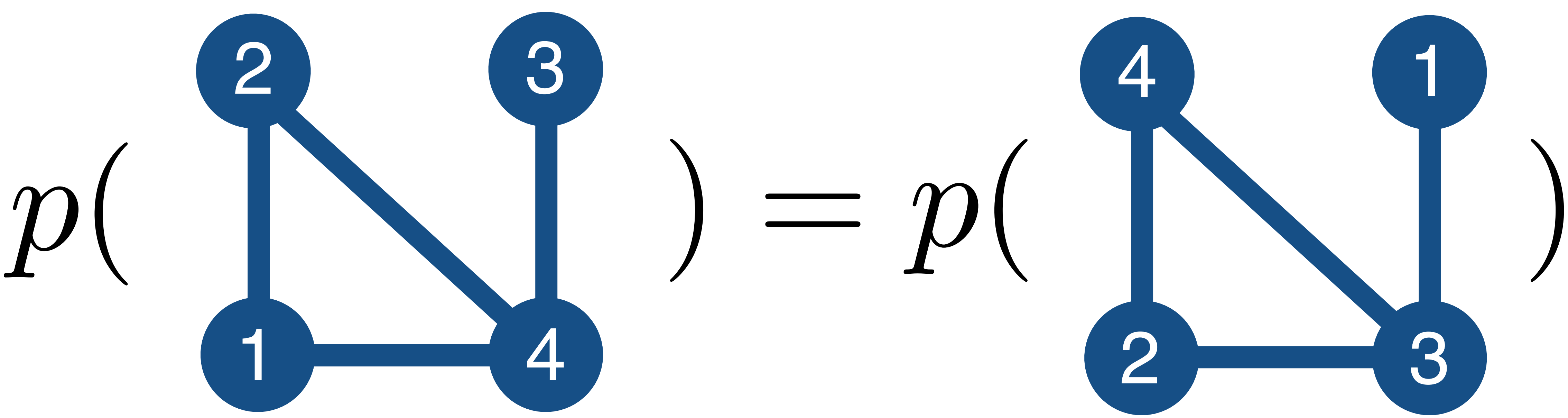
$p($  $)$

[Hoover 1979, Aldous 1981]

The Old Way: Exchangeability

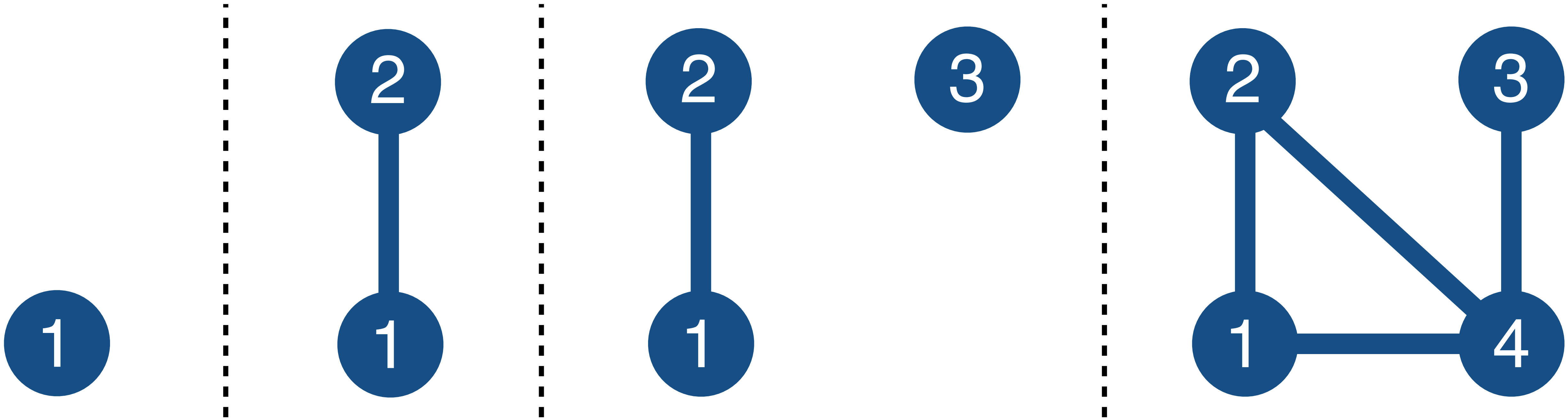


G_1 G_2 G_3 G_4

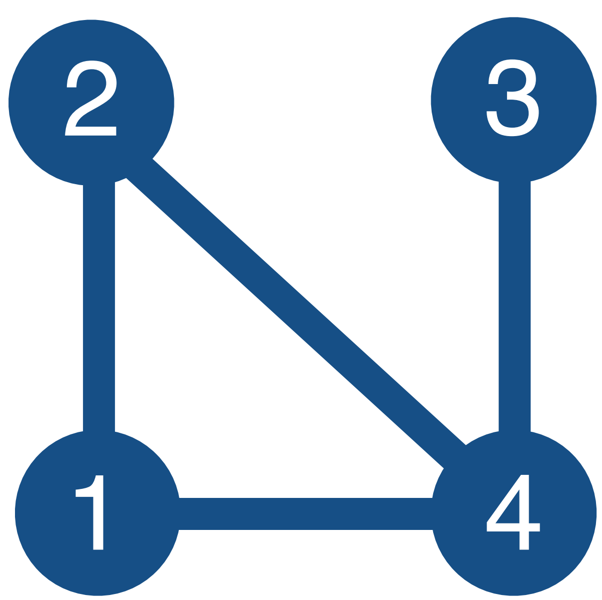


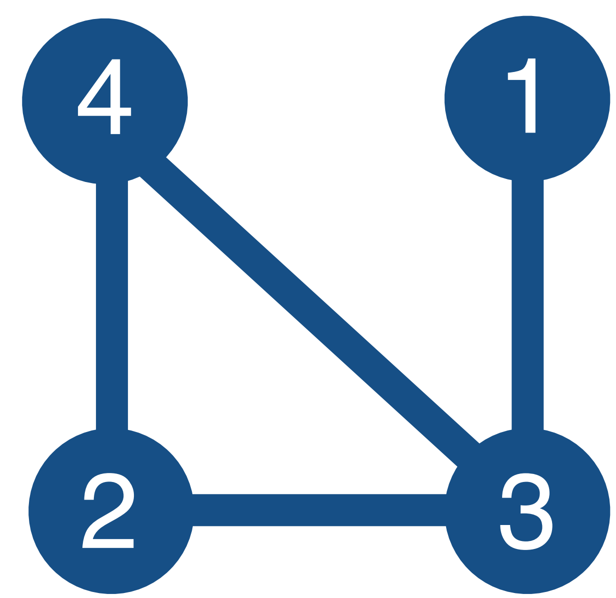
[Hoover 1979, Aldous 1981]

The Old Way: Node exchangeability



G_1 G_2 G_3 G_4

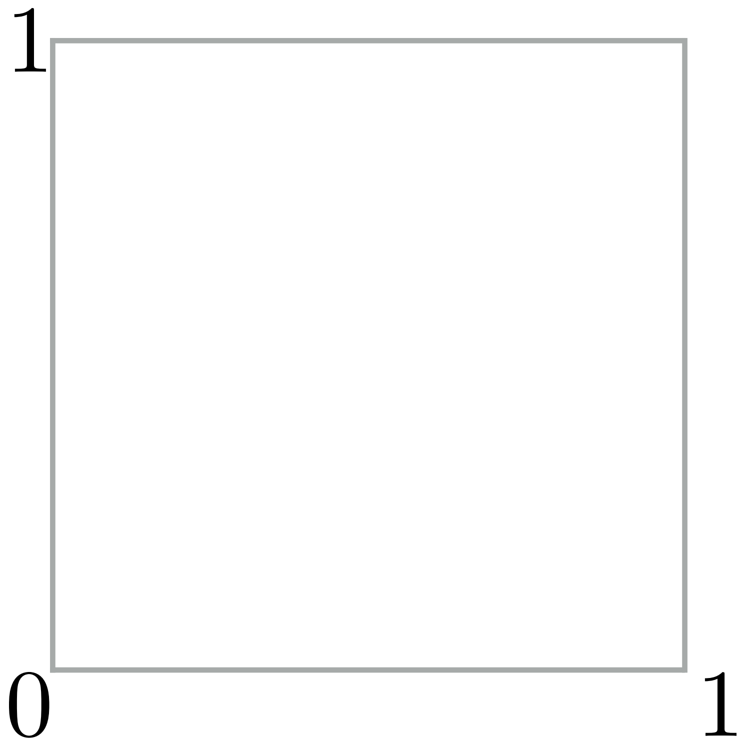
$p($ 

$) = p($ 

$)$

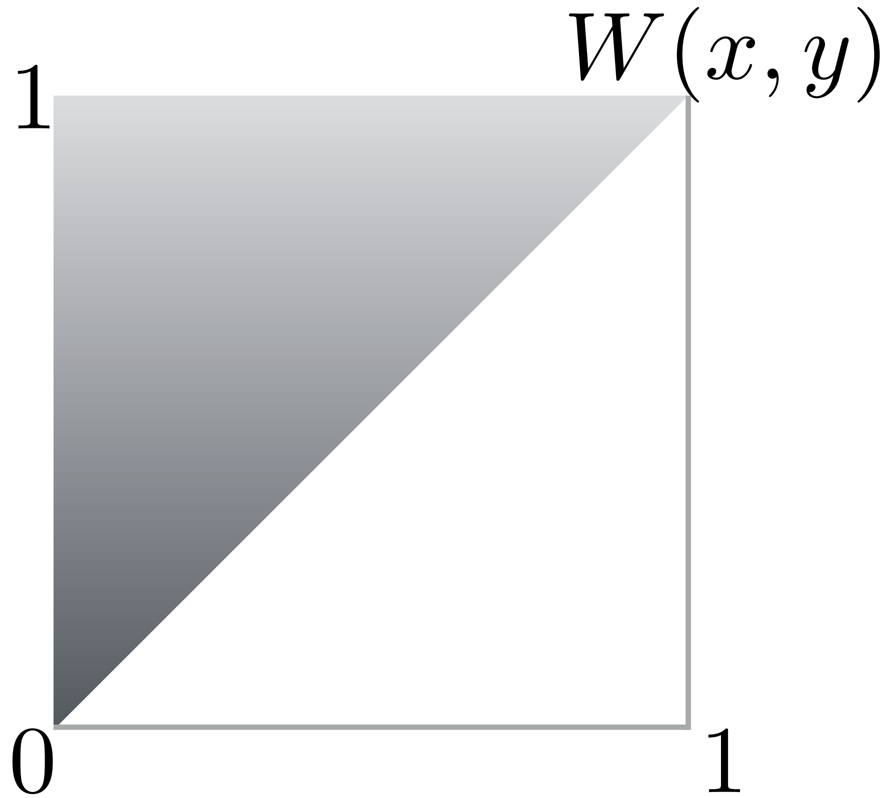
[Hoover 1979, Aldous 1981]

Aldous-Hoover



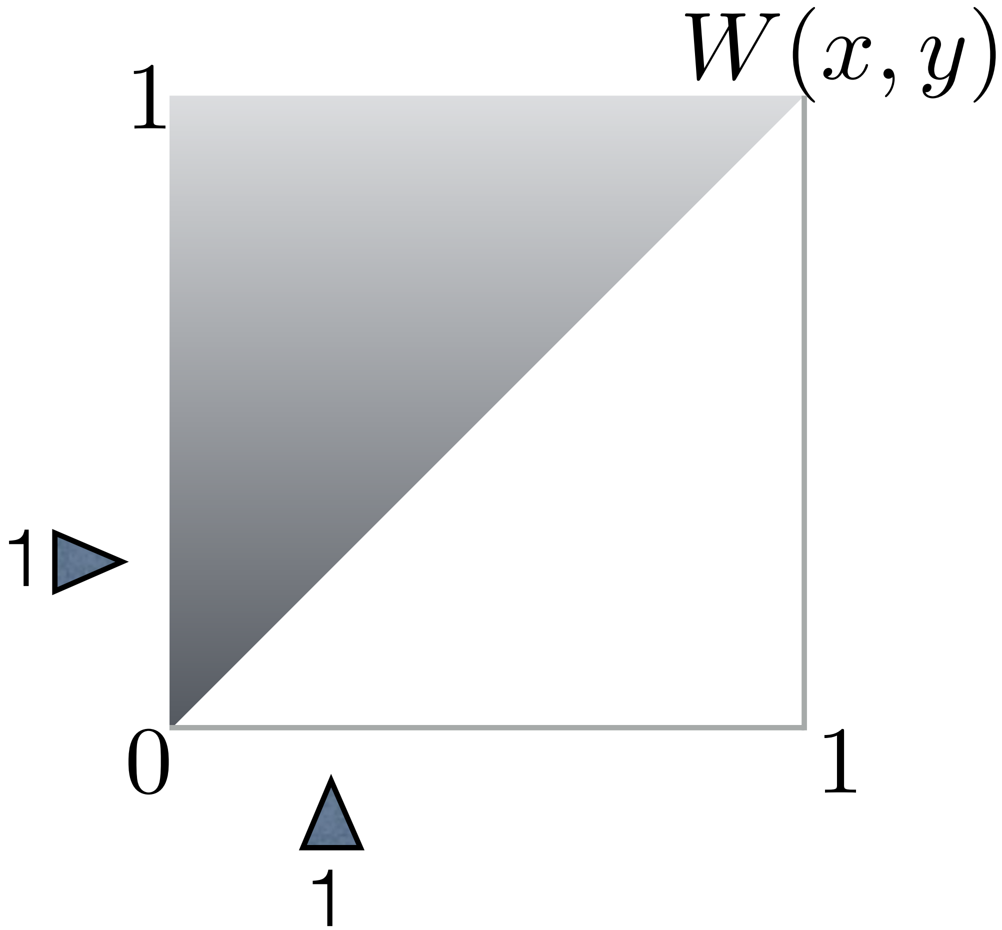
[Hoover 1979, Aldous 1981]

Aldous-Hoover



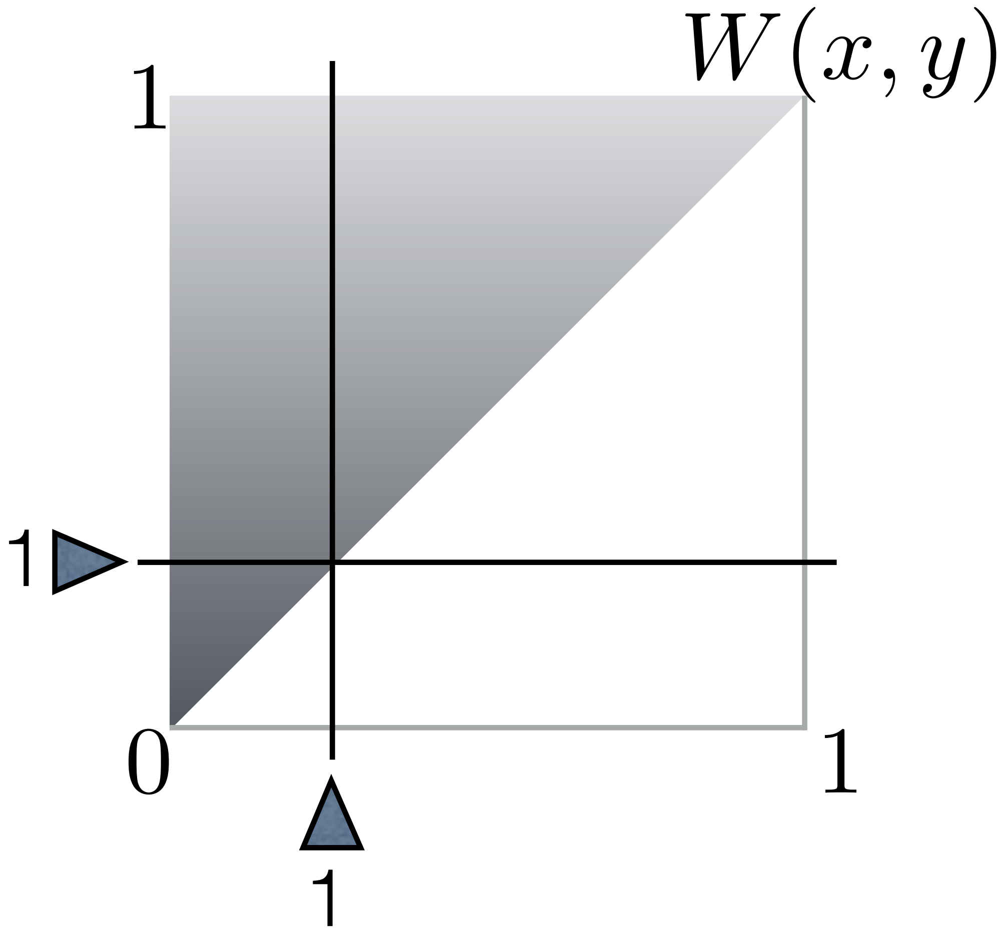
[Hoover 1979, Aldous 1981]

Aldous-Hoover



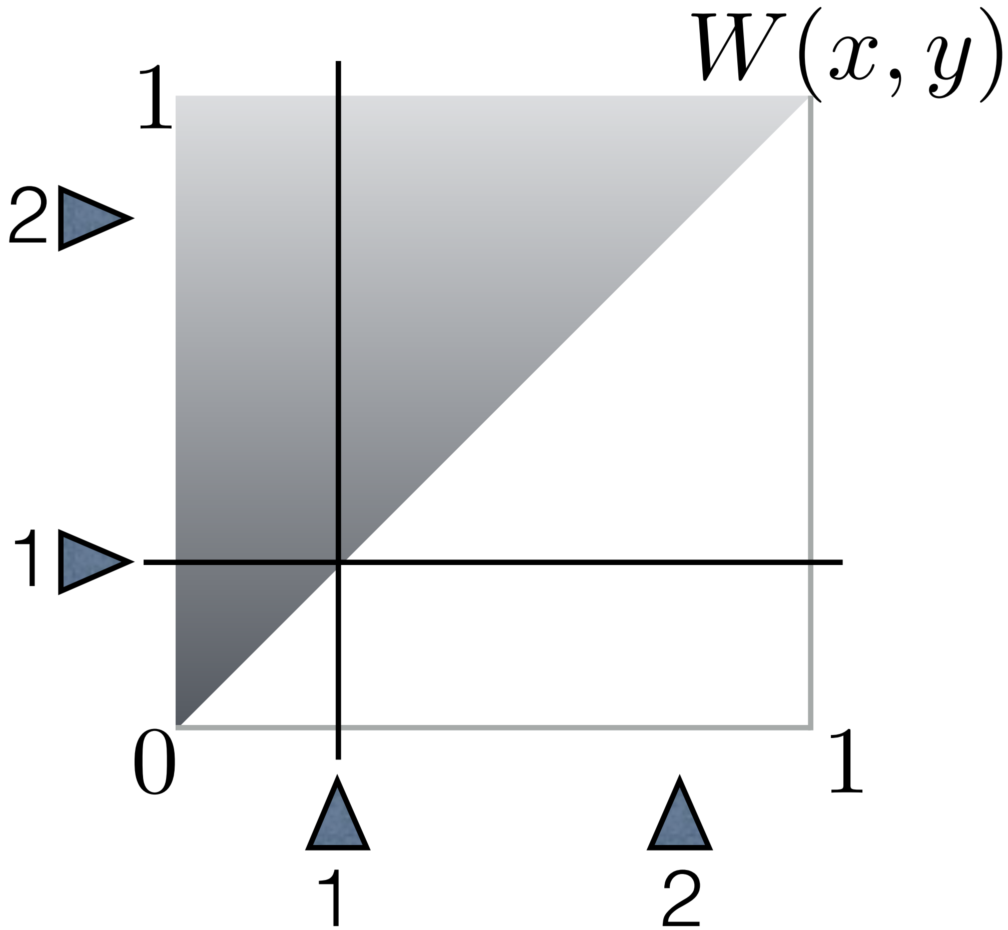
[Hoover 1979, Aldous 1981]

Aldous-Hoover



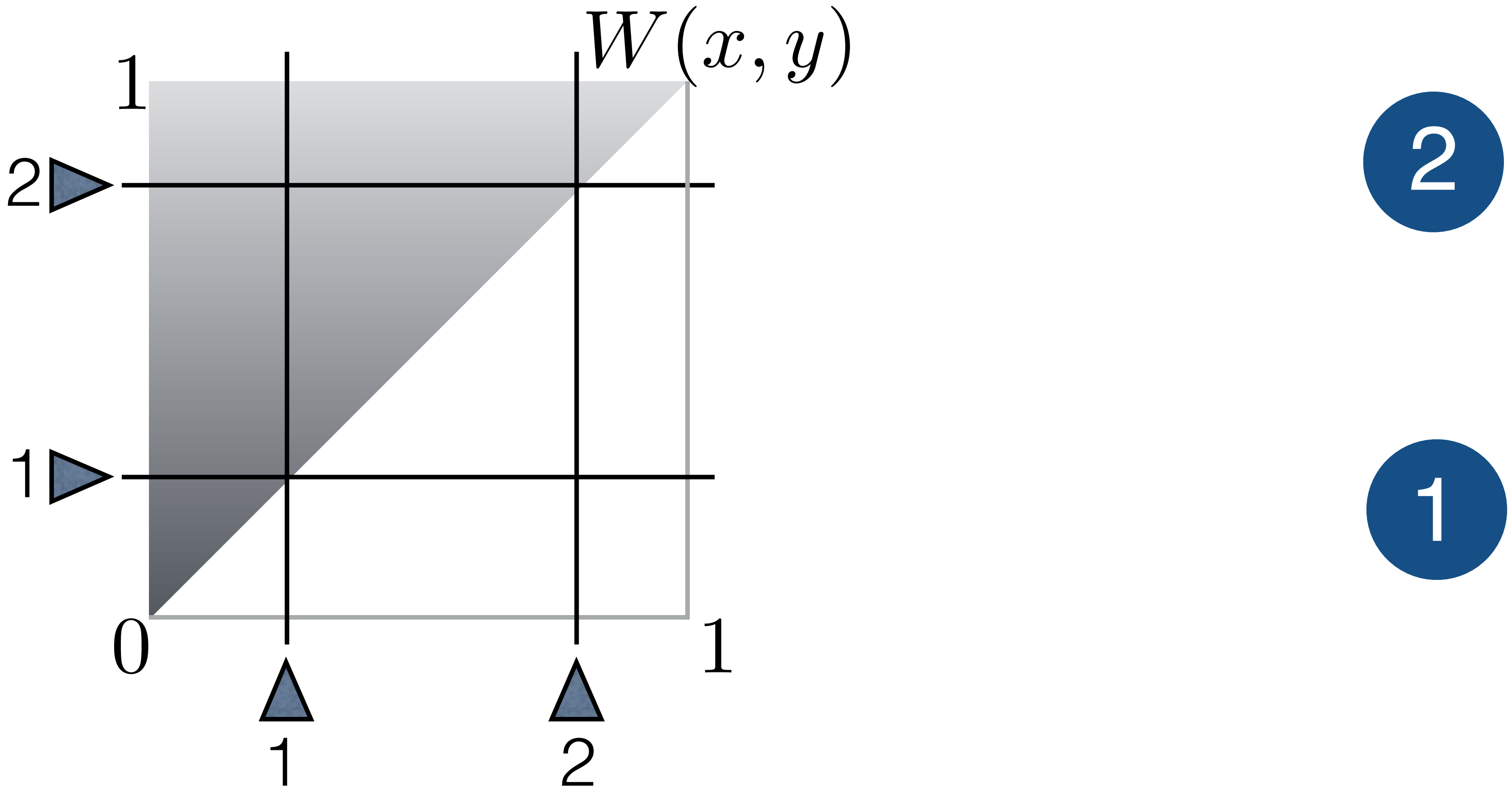
[Hoover 1979, Aldous 1981]

Aldous-Hoover



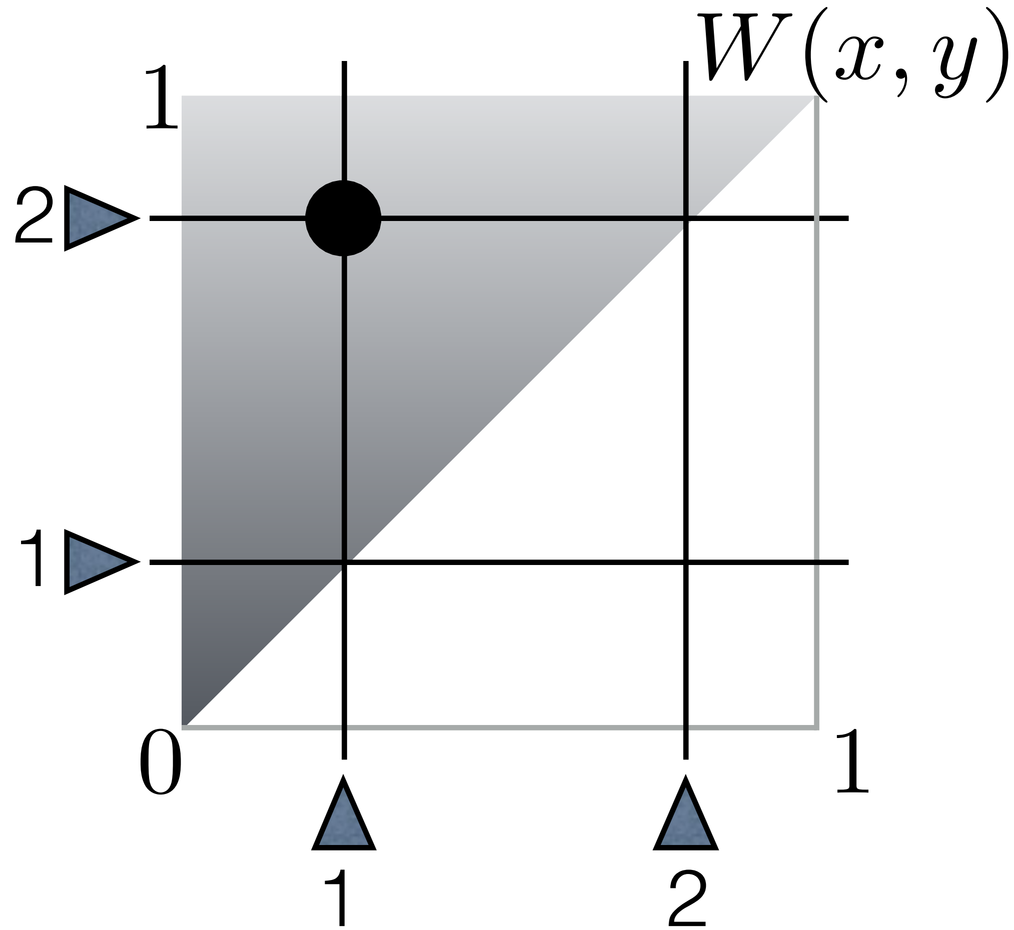
[Hoover 1979, Aldous 1981]

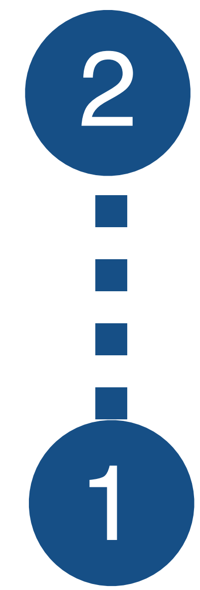
Aldous-Hoover



[Hoover 1979, Aldous 1981]

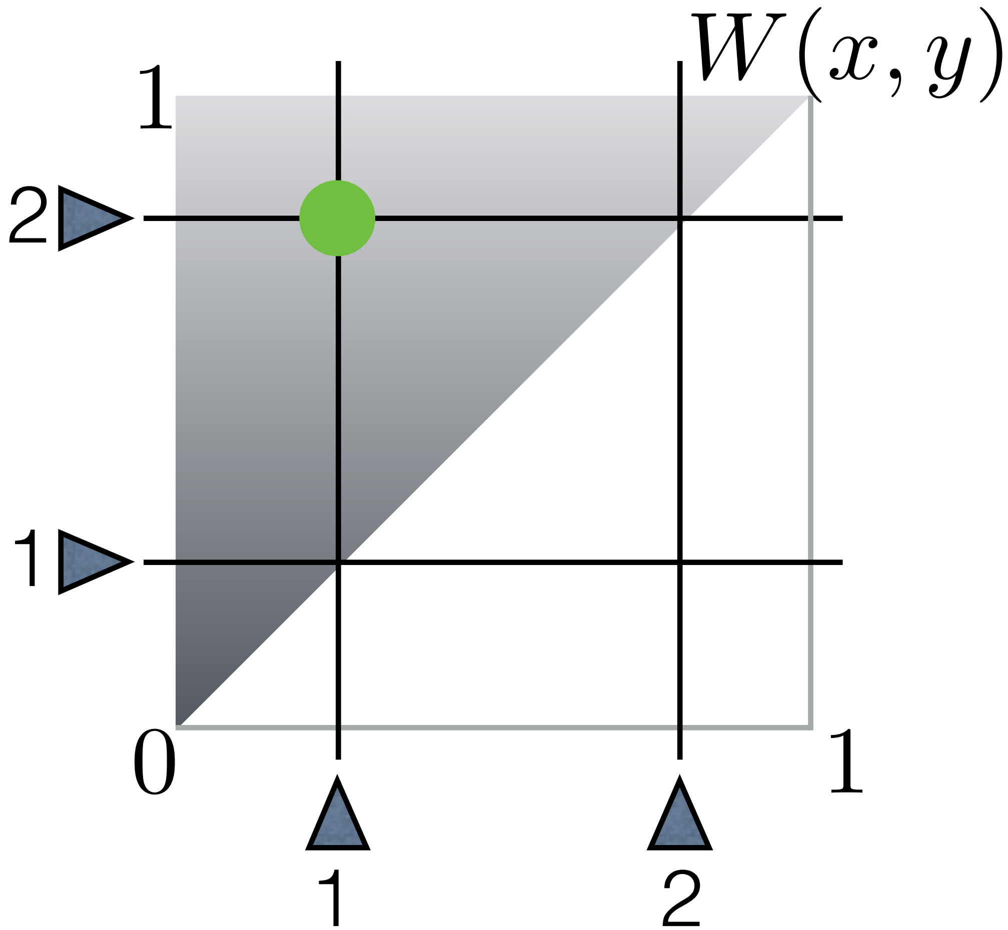
Aldous-Hoover

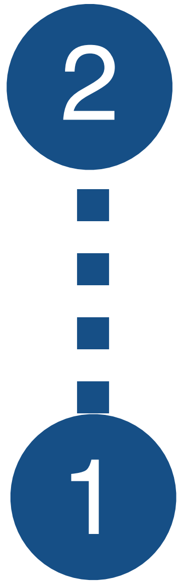




[Hoover 1979, Aldous 1981]

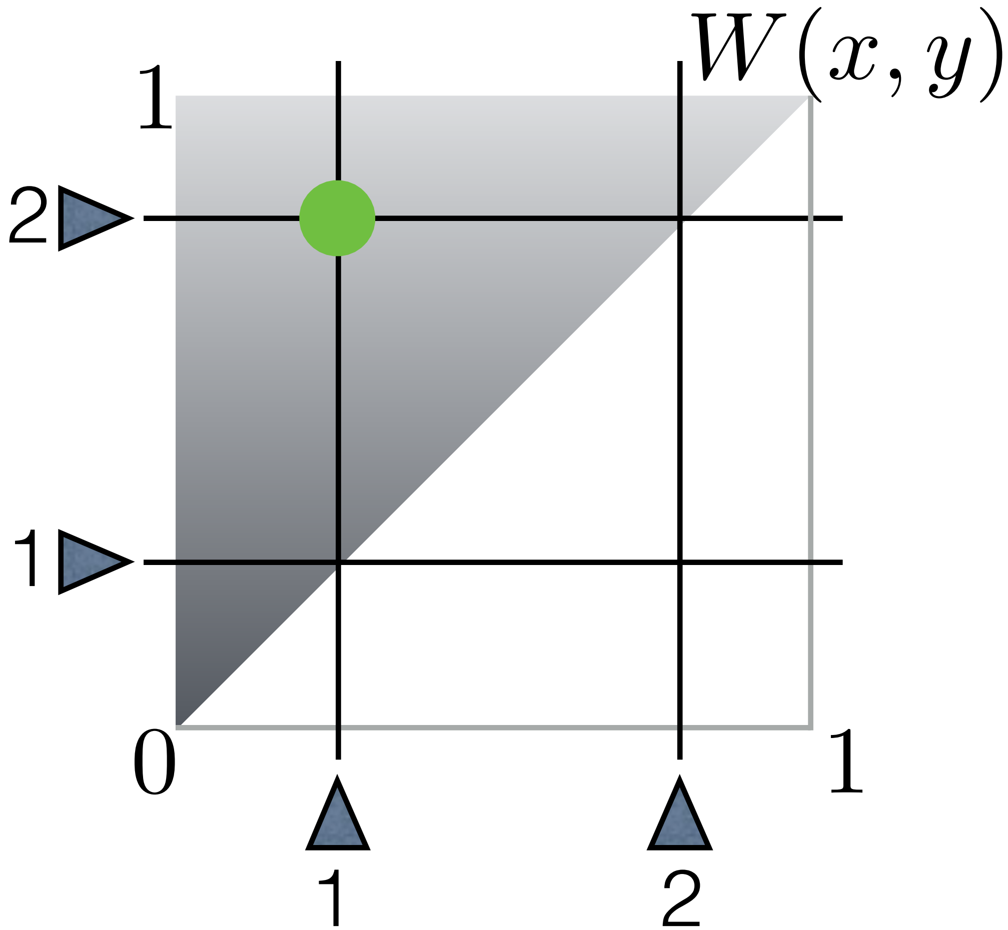
Aldous-Hoover

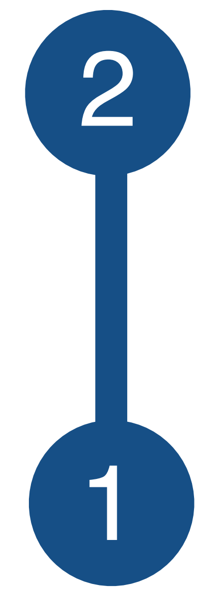




[Hoover 1979, Aldous 1981]

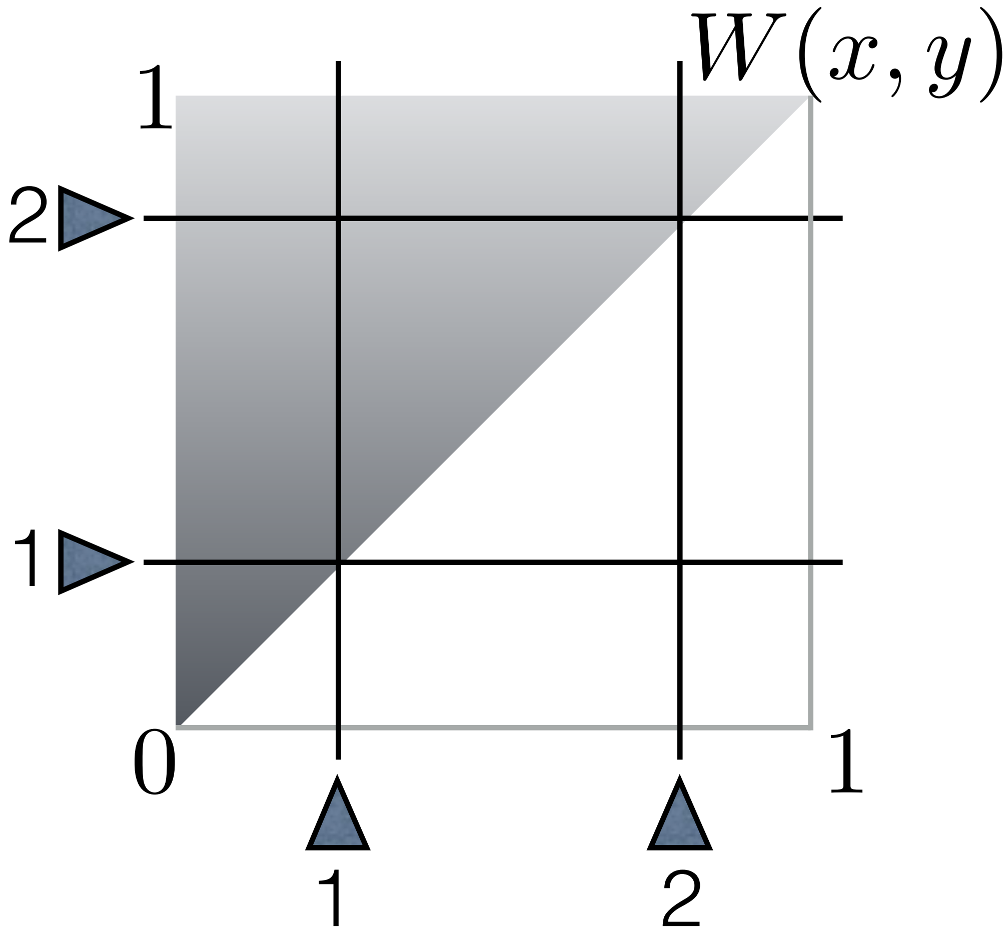
Aldous-Hoover





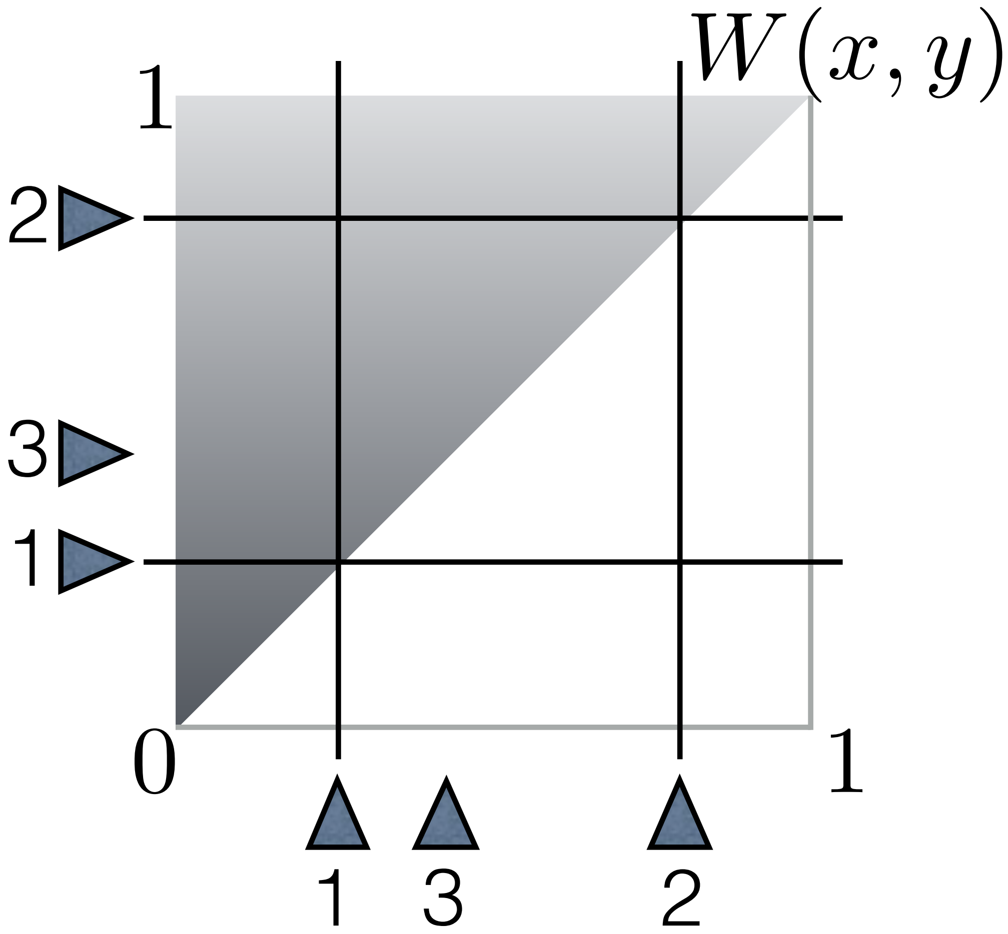
[Hoover 1979, Aldous 1981]

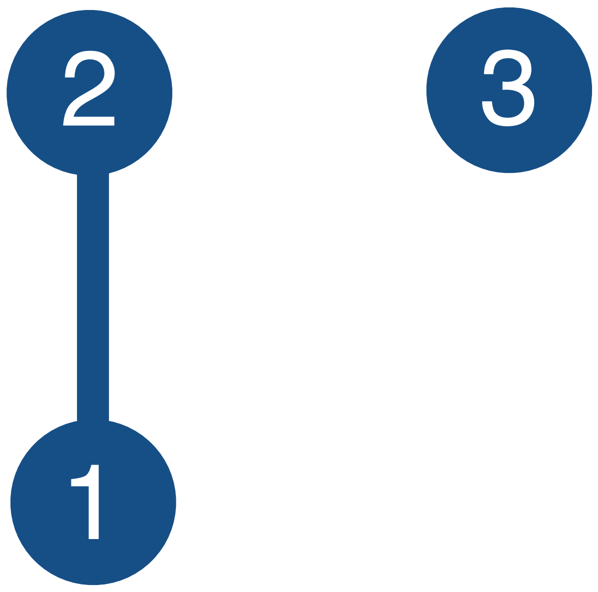
Aldous-Hoover



[Hoover 1979, Aldous 1981]

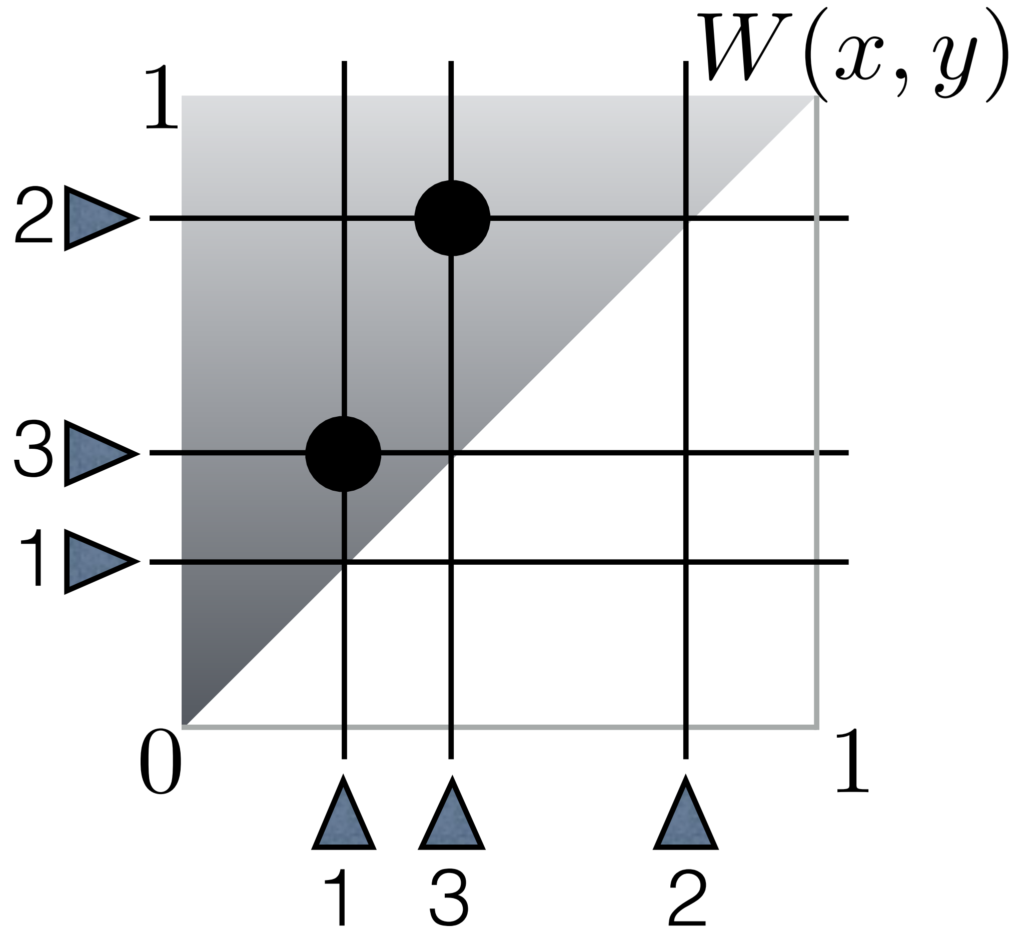
Aldous-Hoover

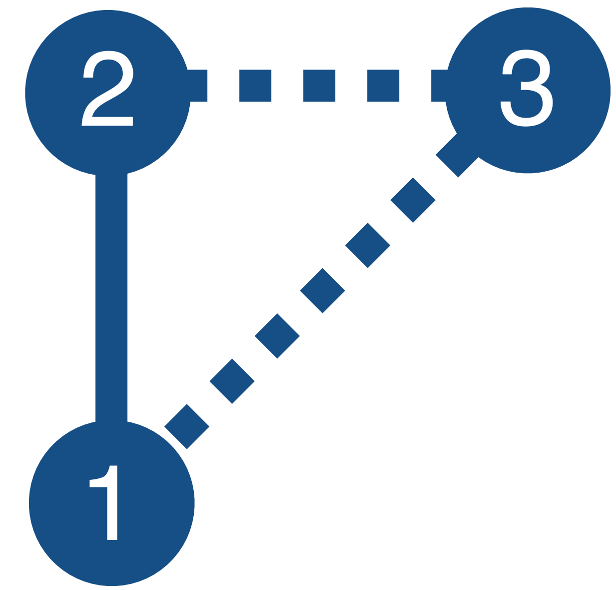




[Hoover 1979, Aldous 1981]

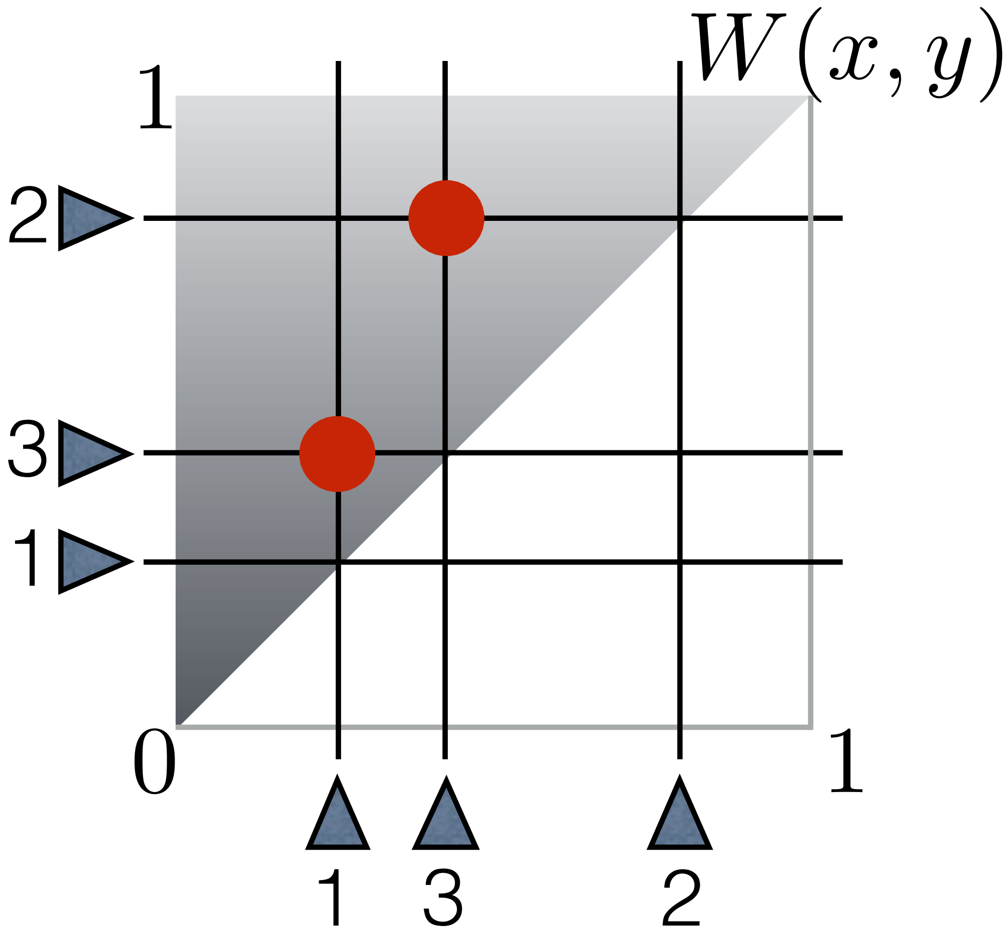
Aldous-Hoover

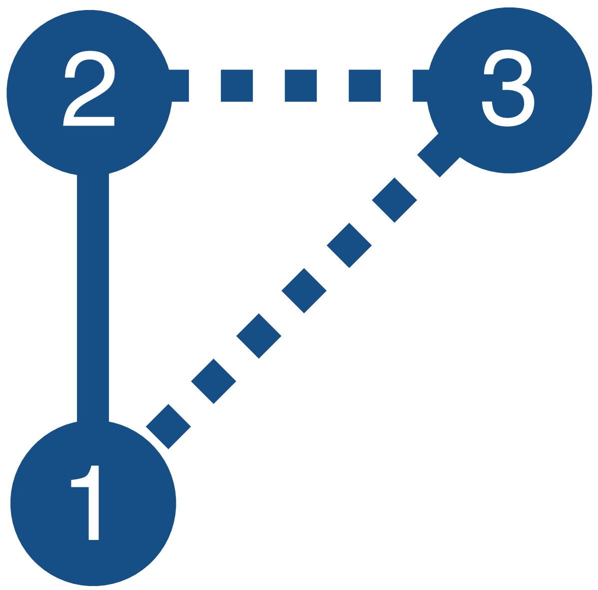




[Hoover 1979, Aldous 1981]

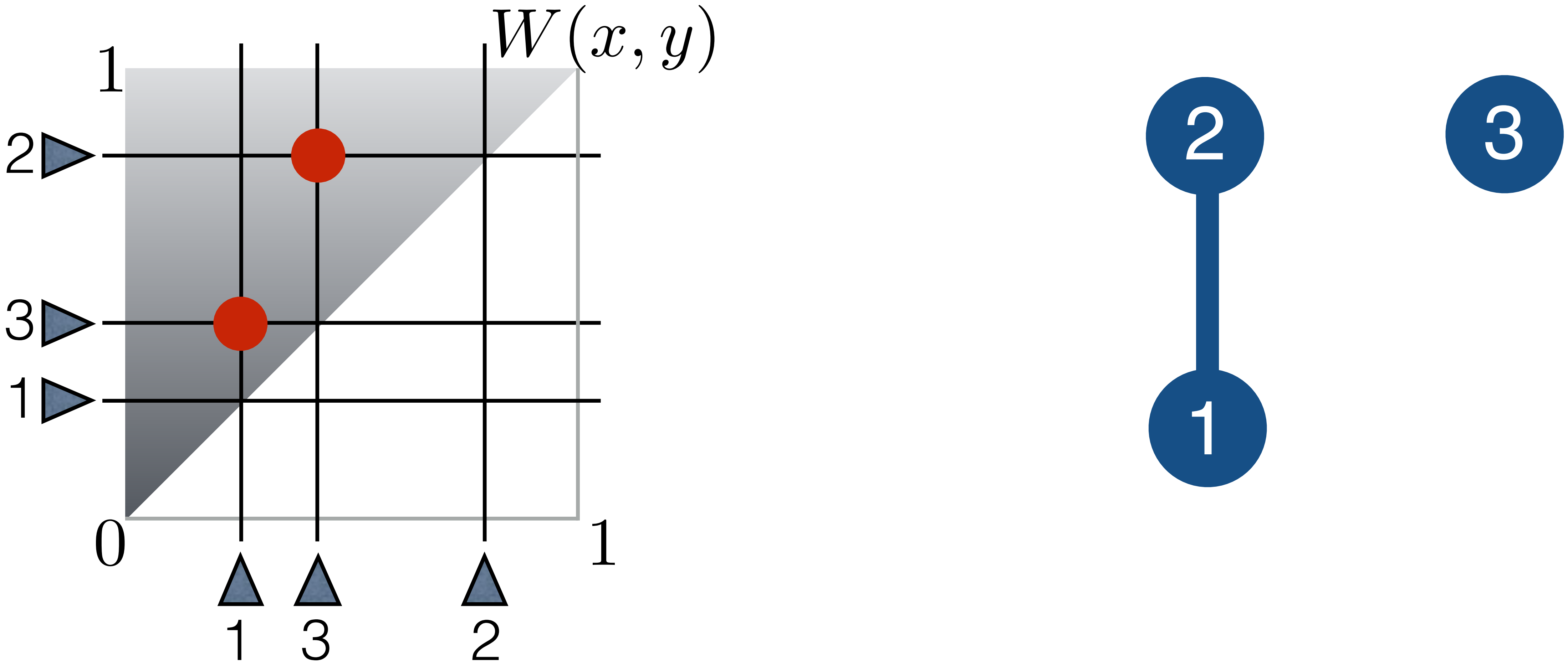
Aldous-Hoover





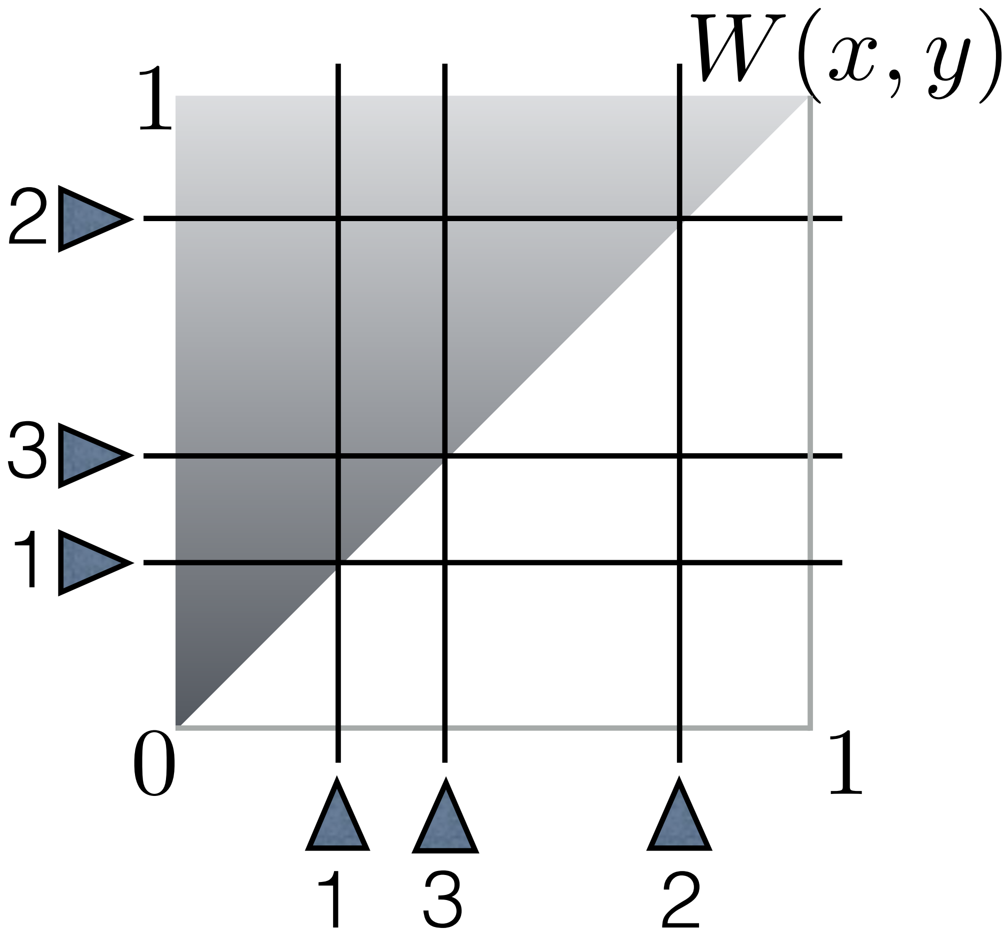
[Hoover 1979, Aldous 1981]

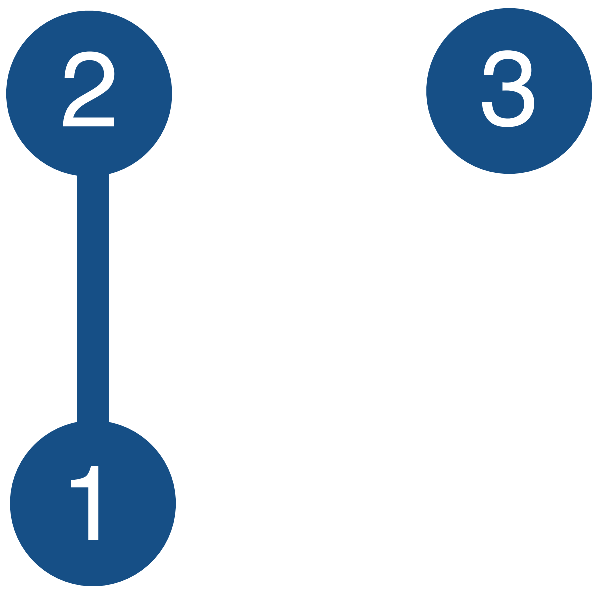
Aldous-Hoover



[Hoover 1979, Aldous 1981]

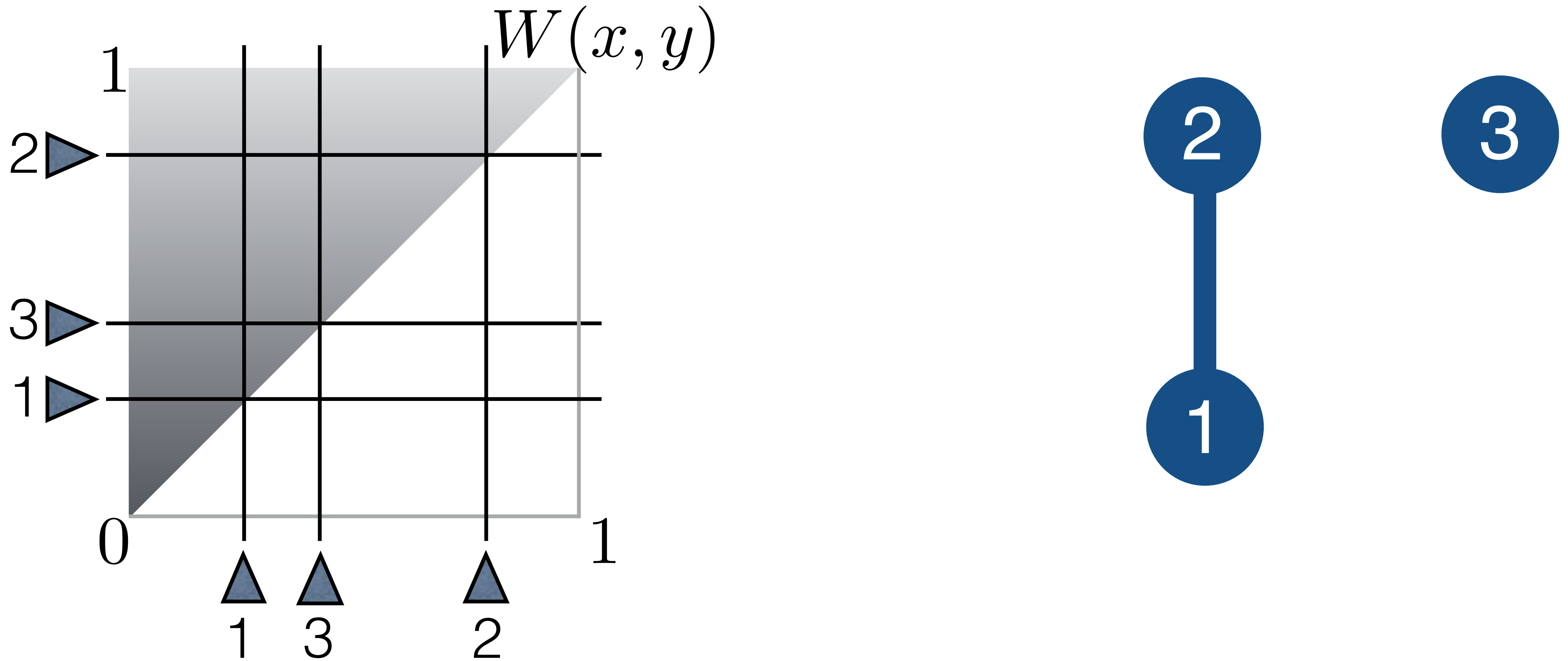
Aldous-Hoover





[Hoover 1979, Aldous 1981]

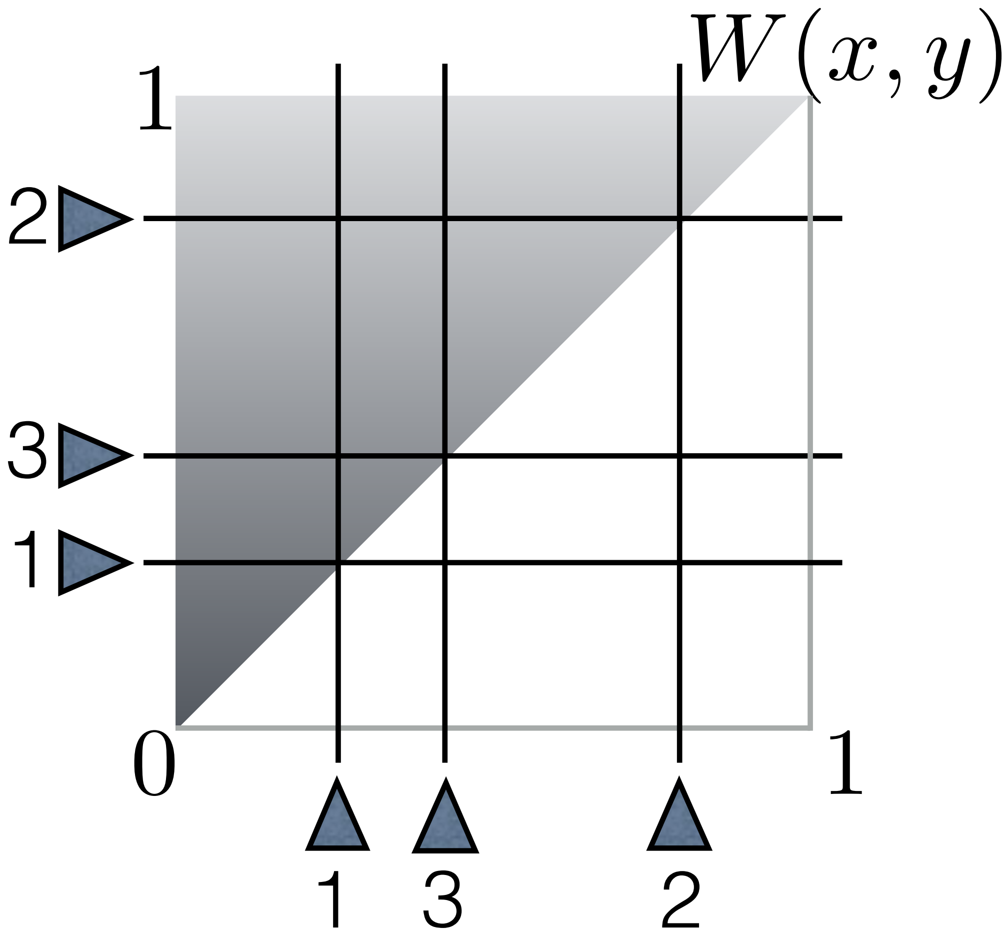
Aldous-Hoover

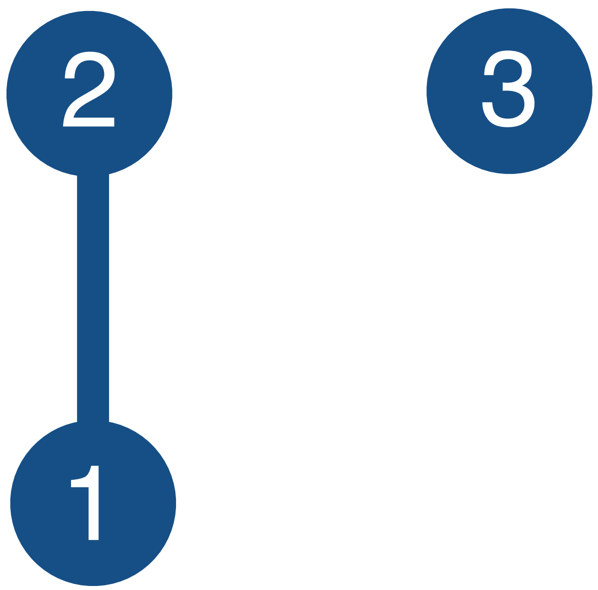


Thm (AH). Every node-exchangeable graph has a *graphon* rep

[Hoover 1979, Aldous 1981]

Aldous-Hoover



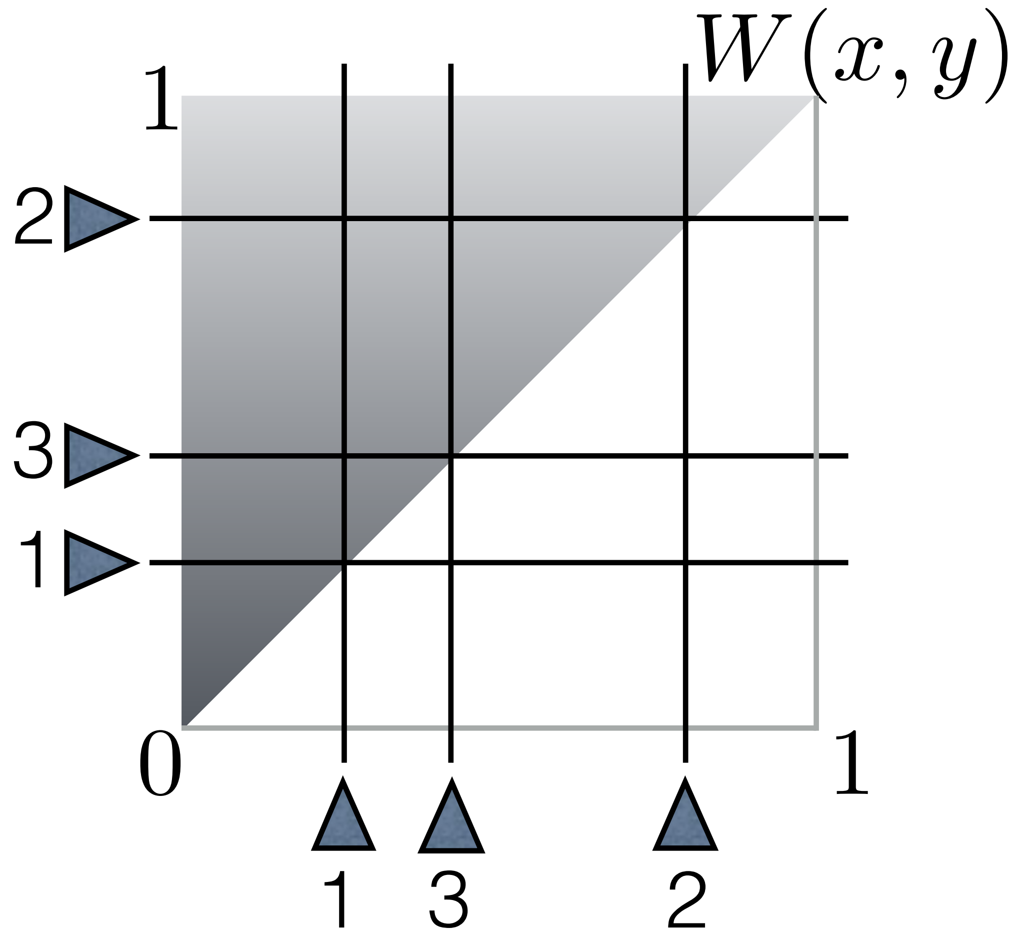


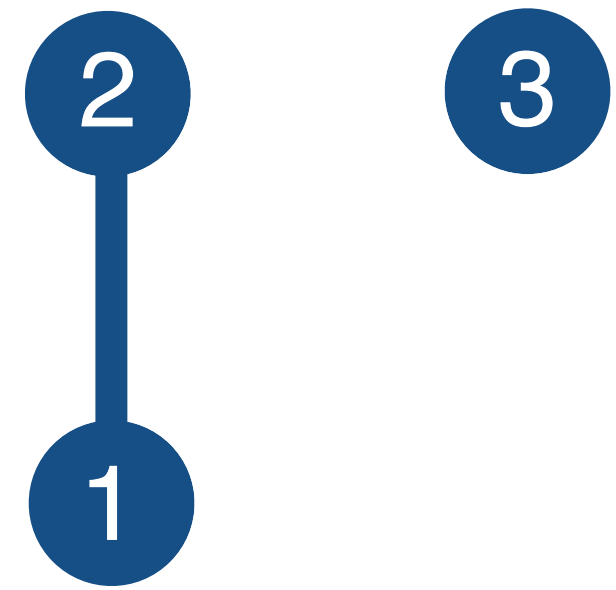
Thm (AH). Every node-exchangeable graph has a *graphon* rep

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

[Hoover 1979, Aldous 1981, Orbanz, Roy 2015]

Aldous-Hoover



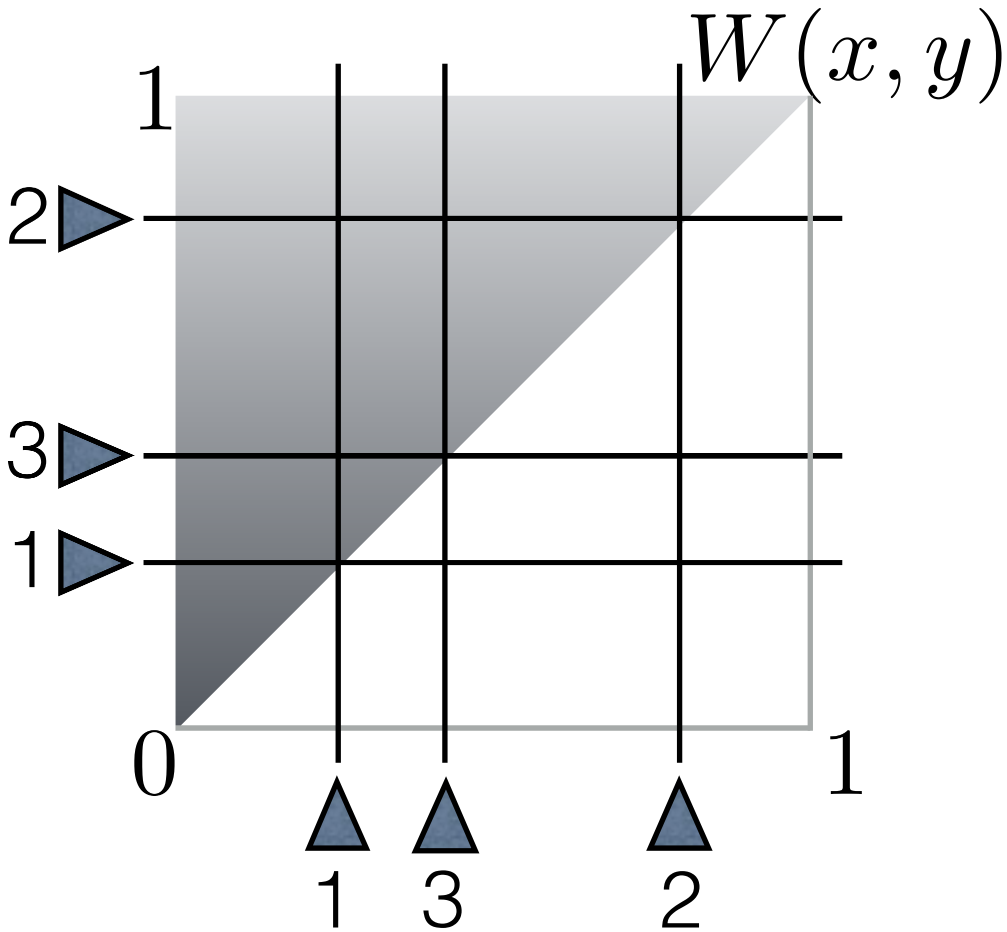


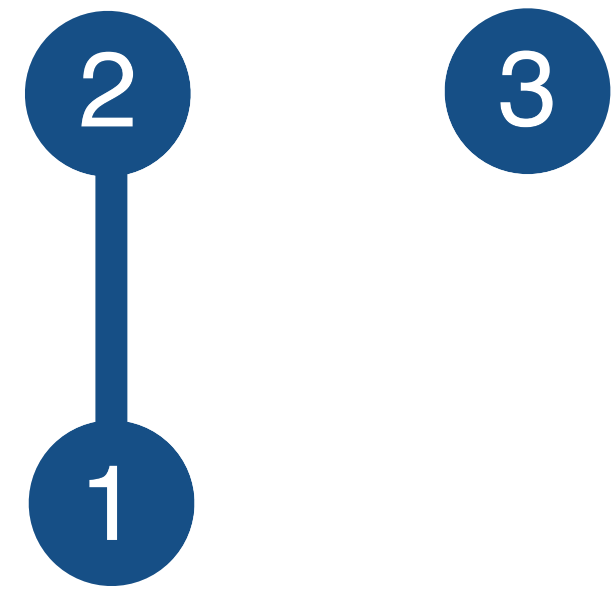
Thm (AH). Every node-exchangeable graph has a *graphon* rep

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

[Hoover 1979, Aldous 1981, Orbanz, Roy 2015]

Aldous-Hoover





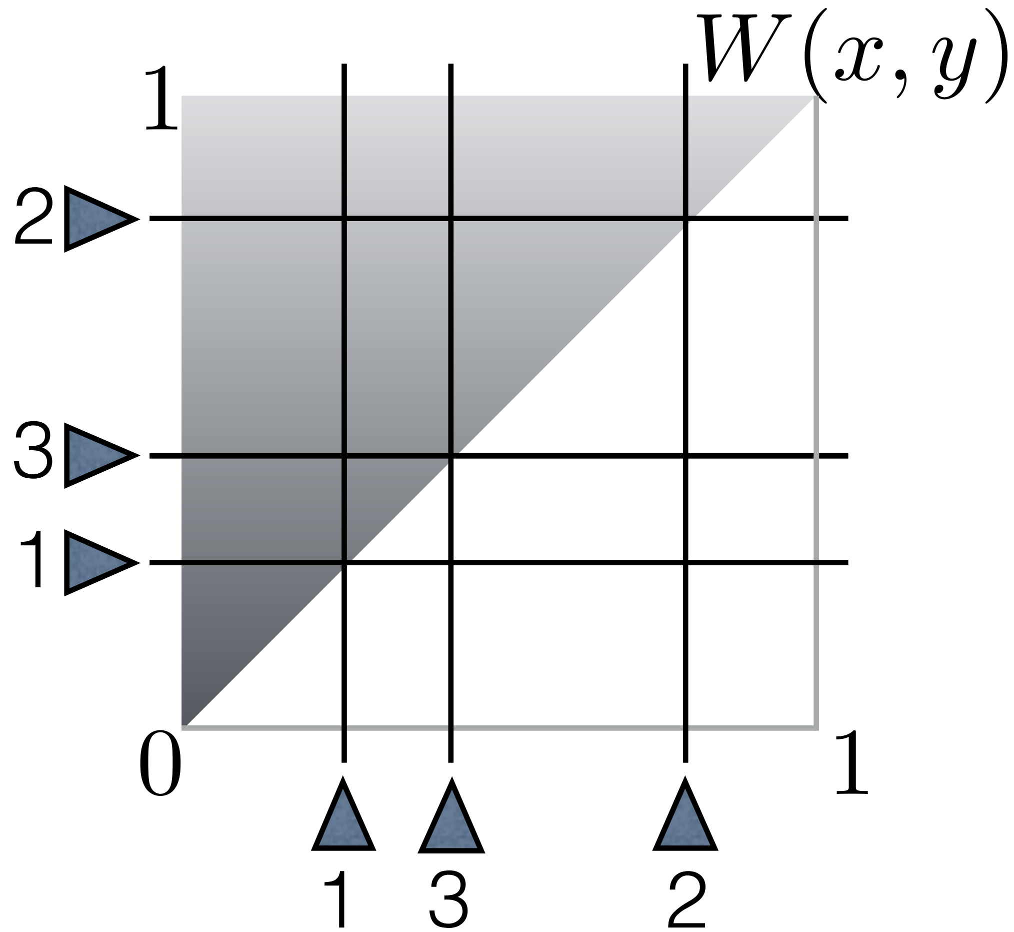
Thm (AH). Every node-exchangeable graph has a *graphon* rep

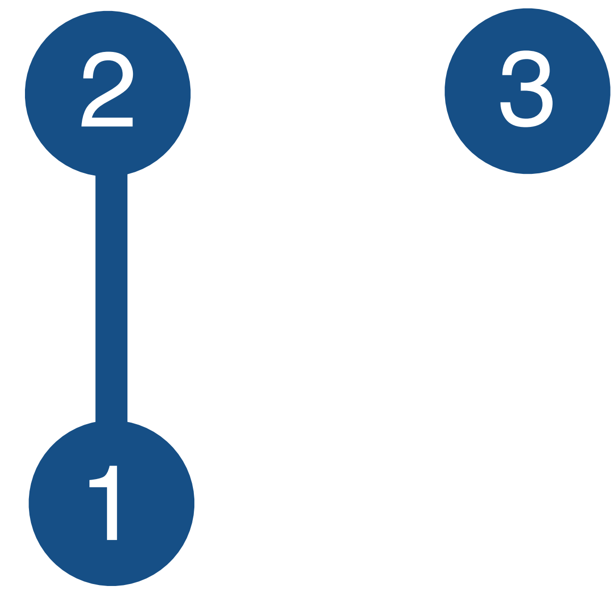
$$\mathbb{E}[\#\text{edges}(G_n)] = \mathbb{E}\left[\binom{n}{2}\frac{1}{2}\int_0^1\int_0^1 W(x,y)\,dx\,dy\right]$$

$$\sim cn^2$$

[Hoover 1979, Aldous 1981, Orbanz, Roy 2015]

Aldous-Hoover





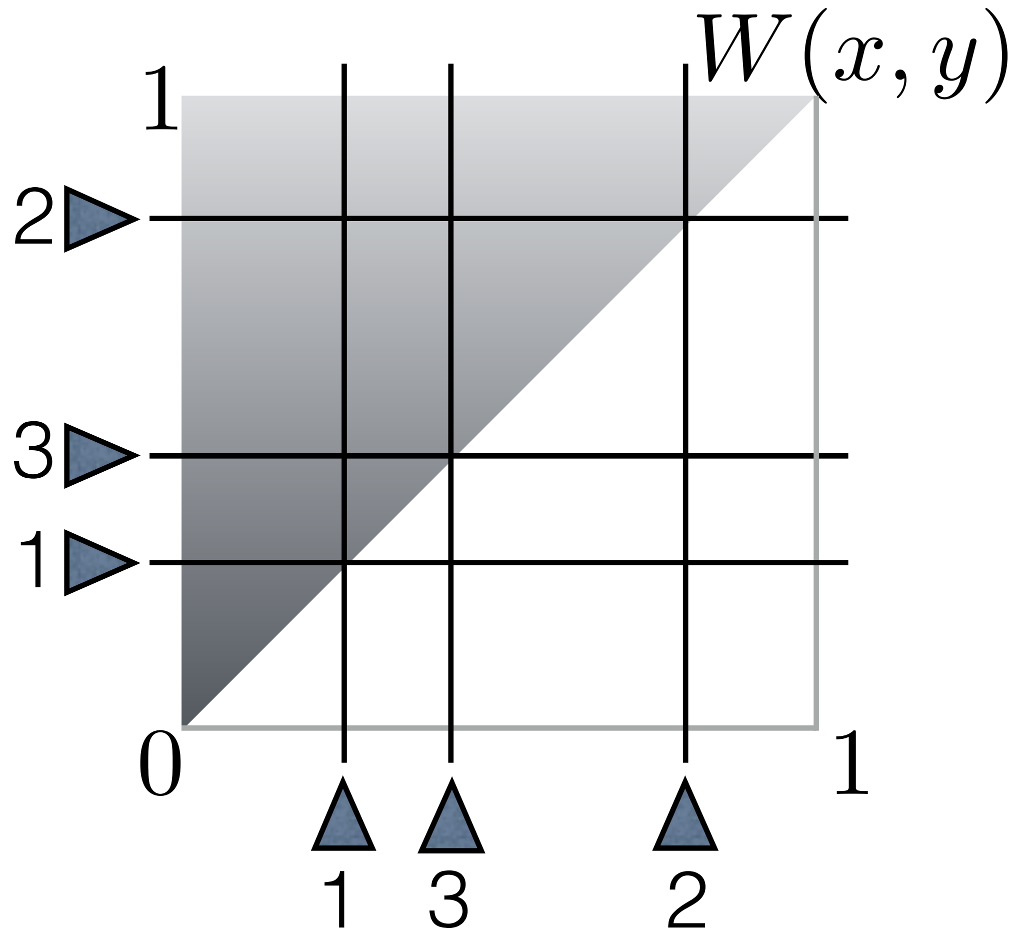
Thm (AH). Every node-exchangeable graph has a *graphon* rep

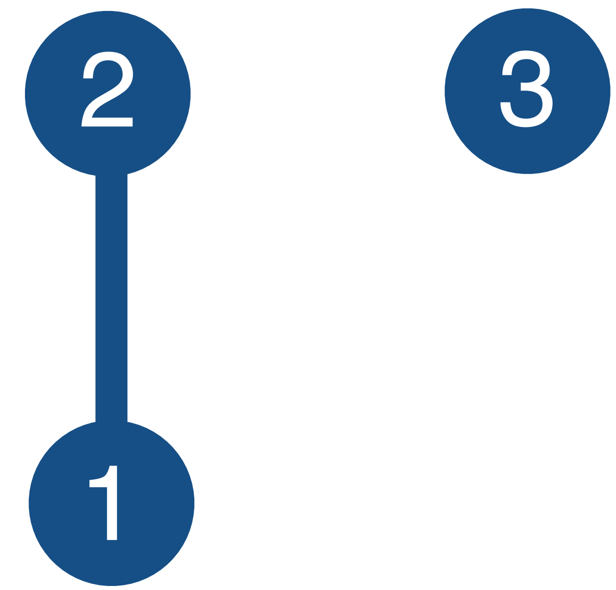
$$\mathbb{E}[\#\text{edges}(G_n)] = \mathbb{E}\left[\binom{n}{2}\frac{1}{2}\int_0^1\int_0^1 W(x,y)\,dx\,dy\right]$$

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

[Hoover 1979, Aldous 1981, Orbanz, Roy 2015]

Aldous-Hoover





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

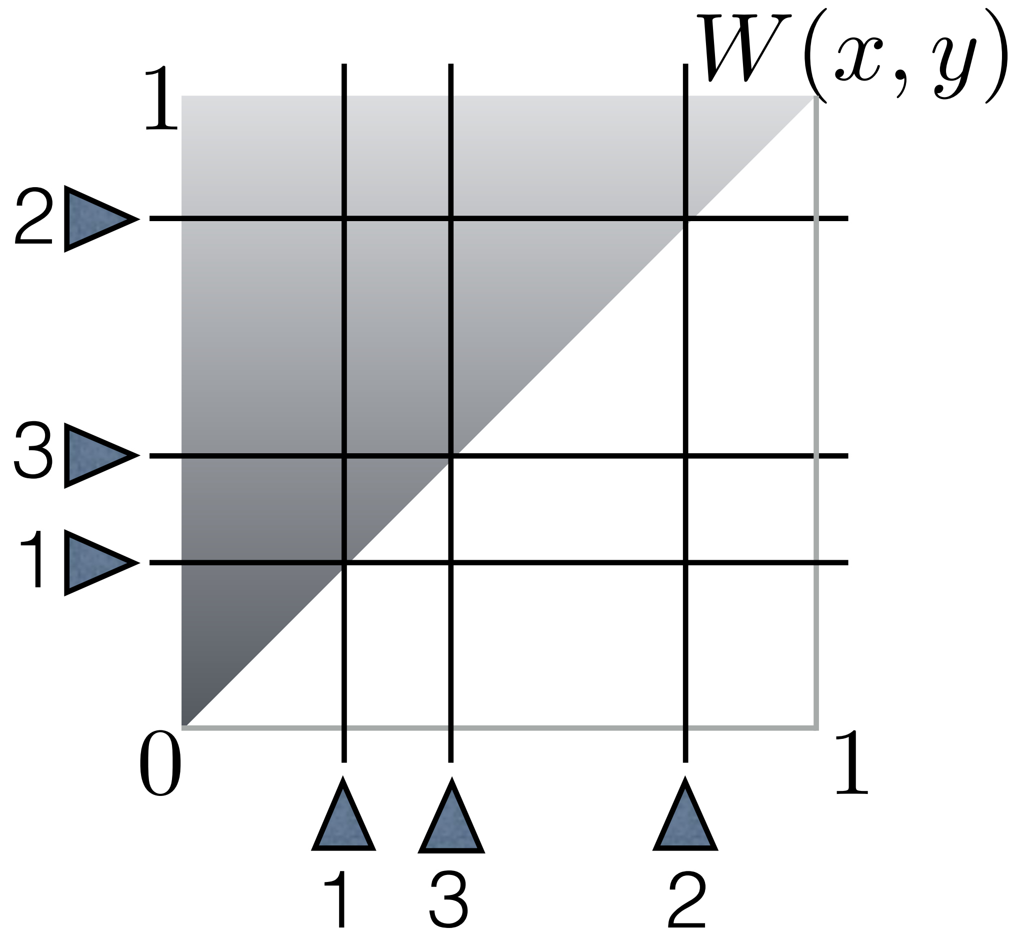
$$\mathbb{E}[\#\text{edges}(G_n)] = \mathbb{E}\left[\binom{n}{2}\frac{1}{2}\int_0^1\int_0^1 W(x,y)\,dx\,dy\right]$$

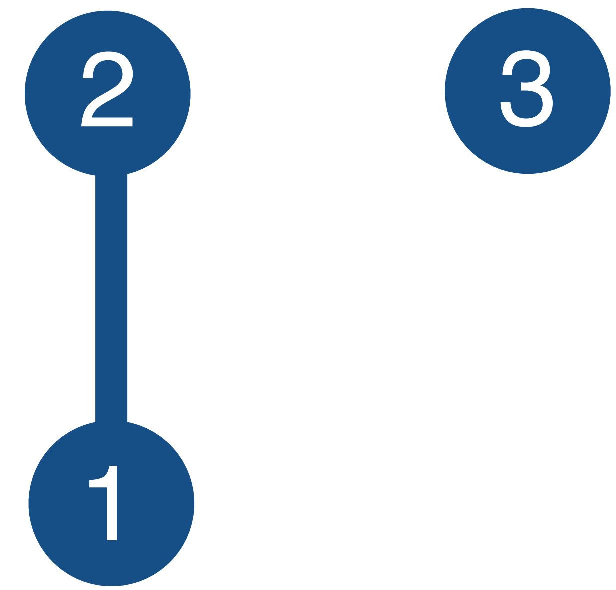
$$\sim cn^2 = c \cdot [\#\text{nodes}(G_n)]^2$$

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

[Hoover 1979, Aldous 1981, Orbanz, Roy 2015]

Aldous-Hoover





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

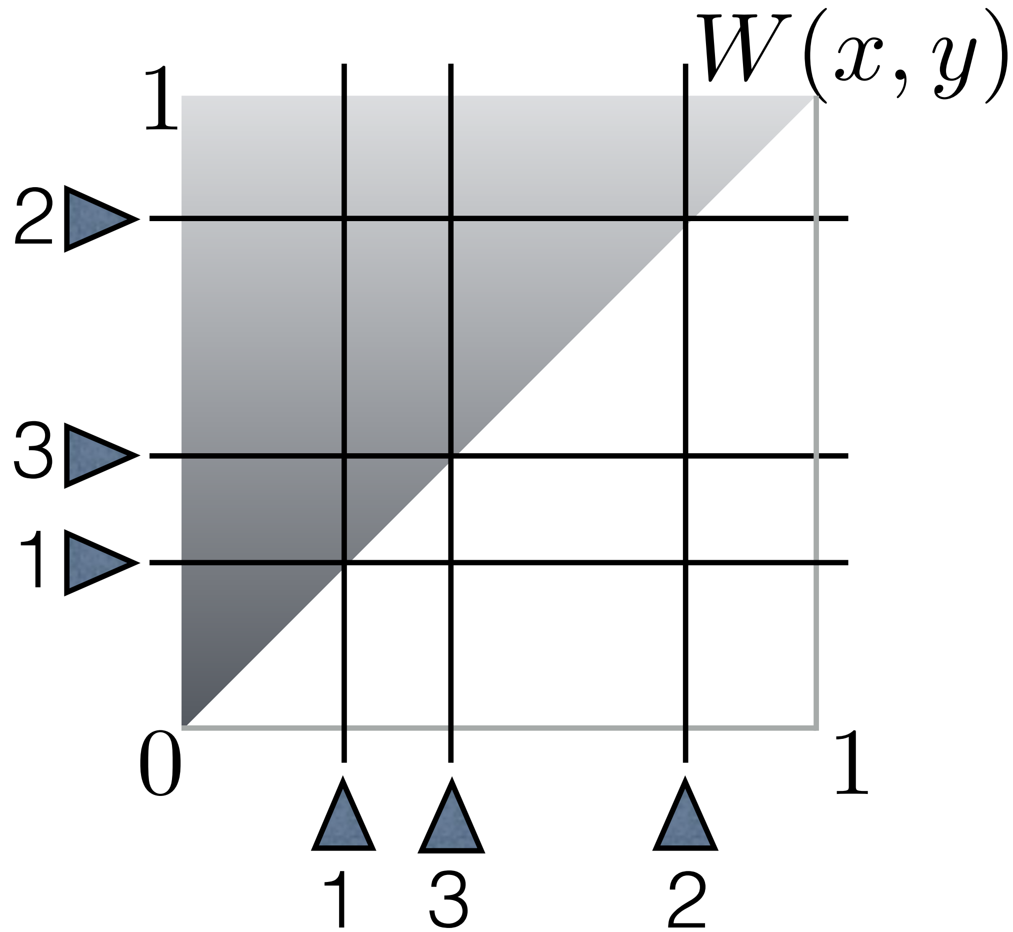
$$\mathbb{E}[\#\text{edges}(G_n)] = \mathbb{E}\left[\binom{n}{2}\frac{1}{2}\int_0^1\int_0^1 W(x,y)\,dx\,dy\right]$$
$$\sim cn^2 = c\cdot[\#\text{nodes}(G_n)]^2$$

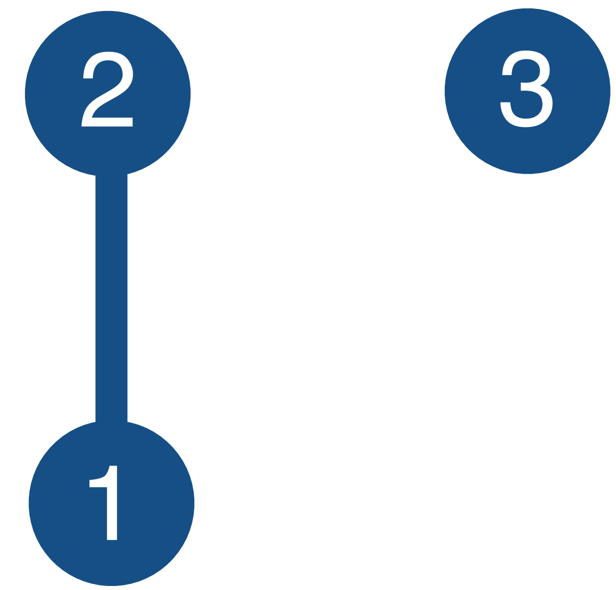
Cor. Every node-exch graph sequence is dense (or empty) a.s.

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

[Hoover 1979, Aldous 1981, Orbanz, Roy 2015]

Aldous-Hoover





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

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

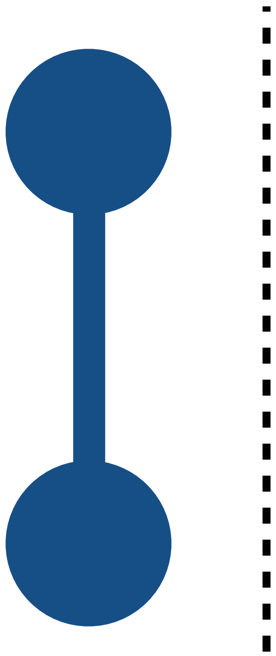
$$\sim cn^2 = c \cdot [\#\text{nodes}(G_n)]^2$$

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

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

[Caron, Fox 2014; Veitch, Roy 2015; Borgs, Chayes, Cohn, Holden 2016;
Broderick, Cai 2015; Cai, Broderick 2015a,b; Crane, Dempsey 2015a,b, 2016a; Cai, Campbell, Broderick 2016]

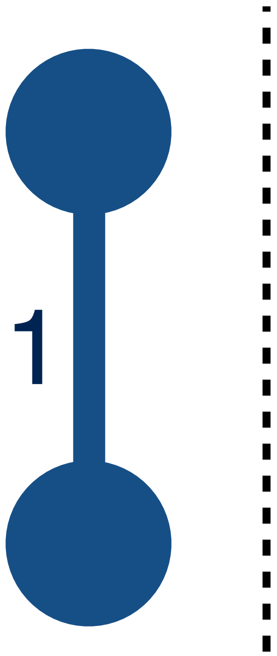
A New Way: Edges



G_1

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

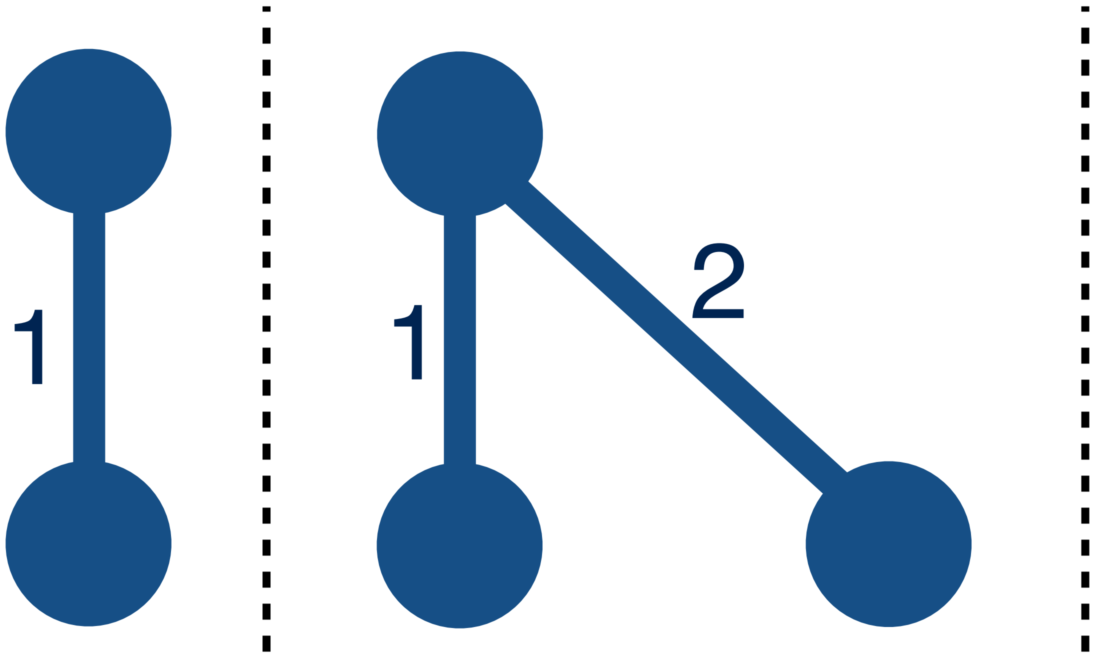
A New Way: Edges



G_1

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

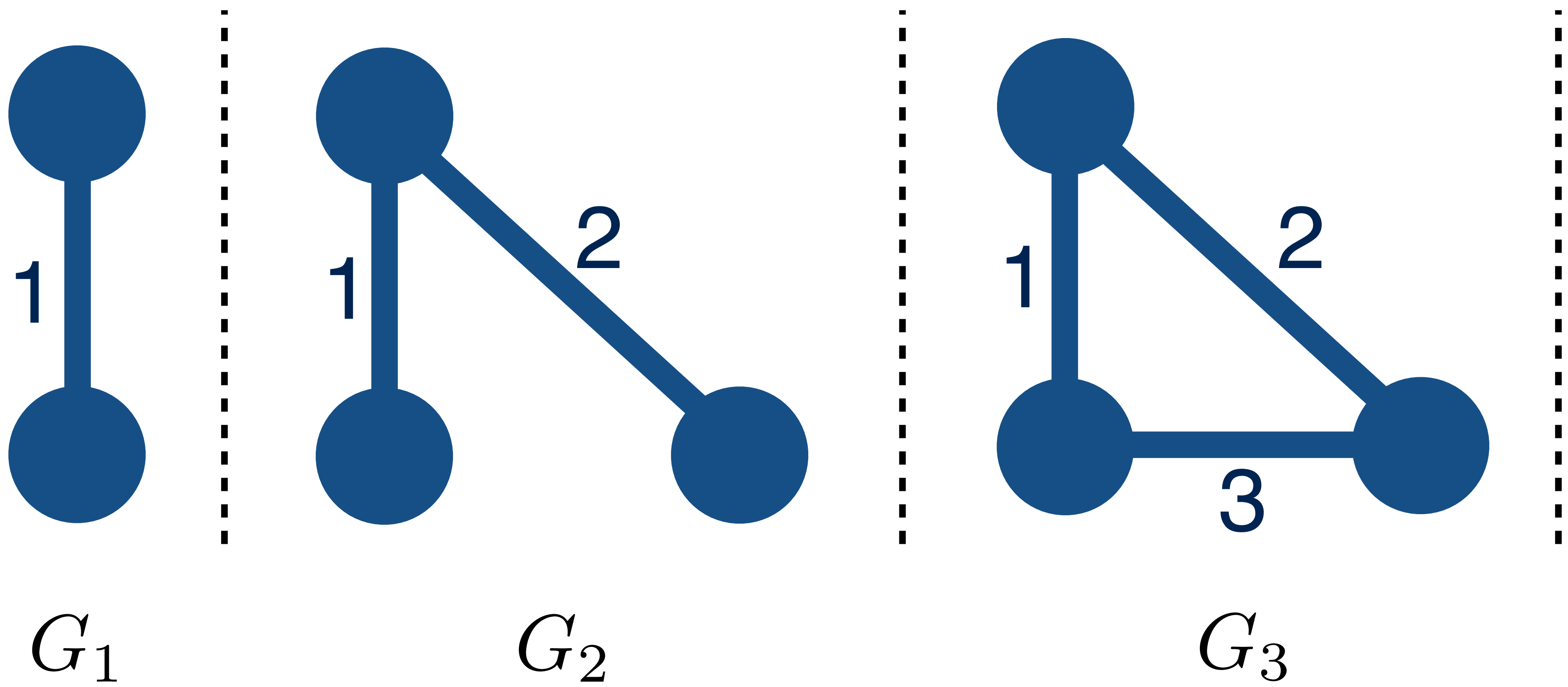
A New Way: Edges



G_1 G_2

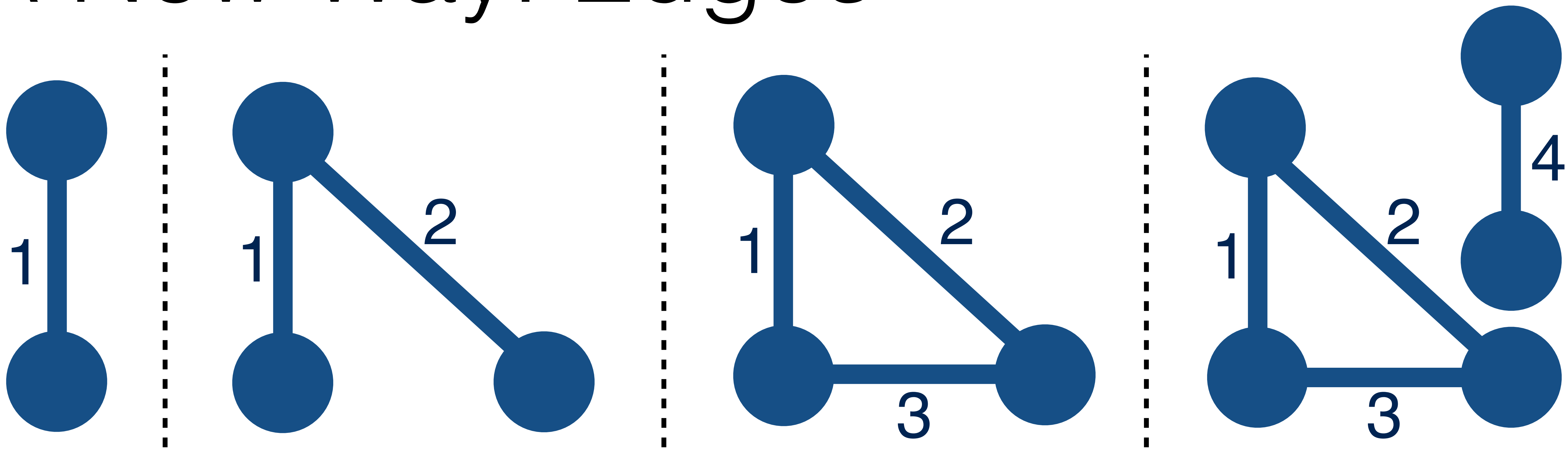
 [Broderick, Cai 2015; Cai, Broderick 2015a,b; Crane, Dempsey 2015a,b,16a; Cai, Campbell, Broderick 2016]

A New Way: Edges



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

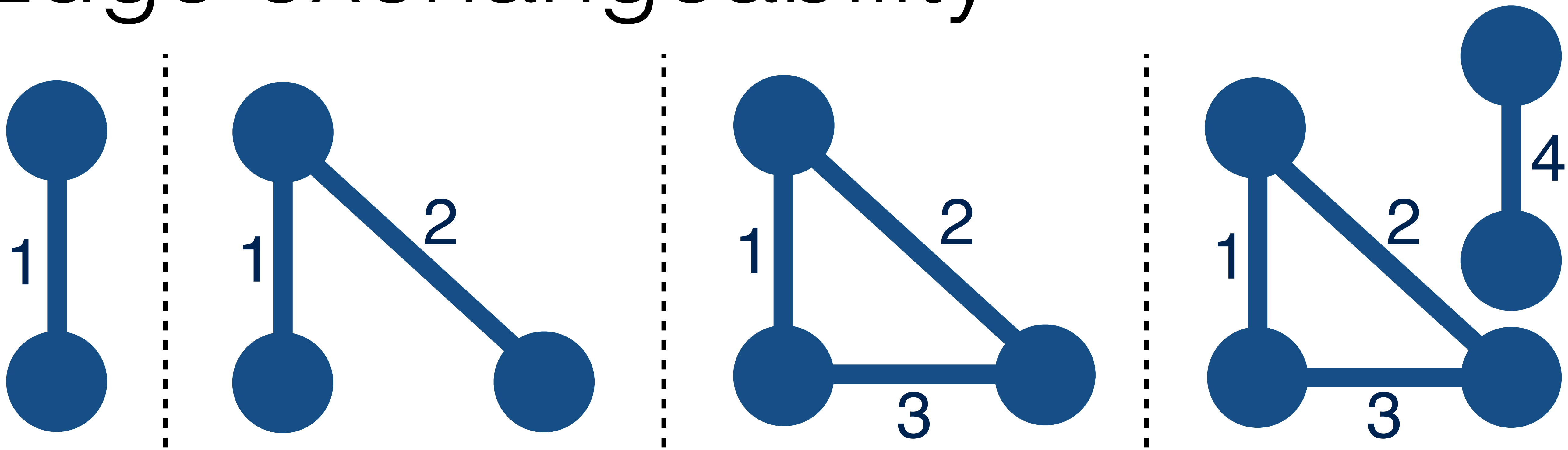
A New Way: Edges



G_1 G_2 G_3 G_4

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

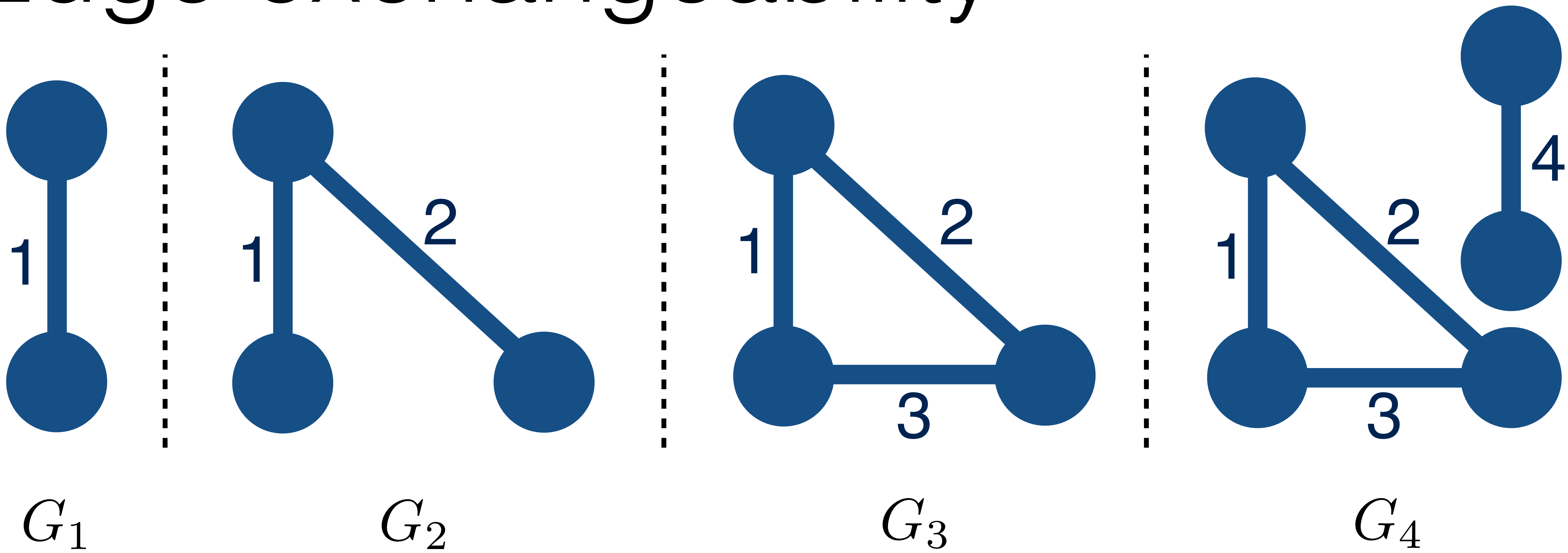
Edge exchangeability

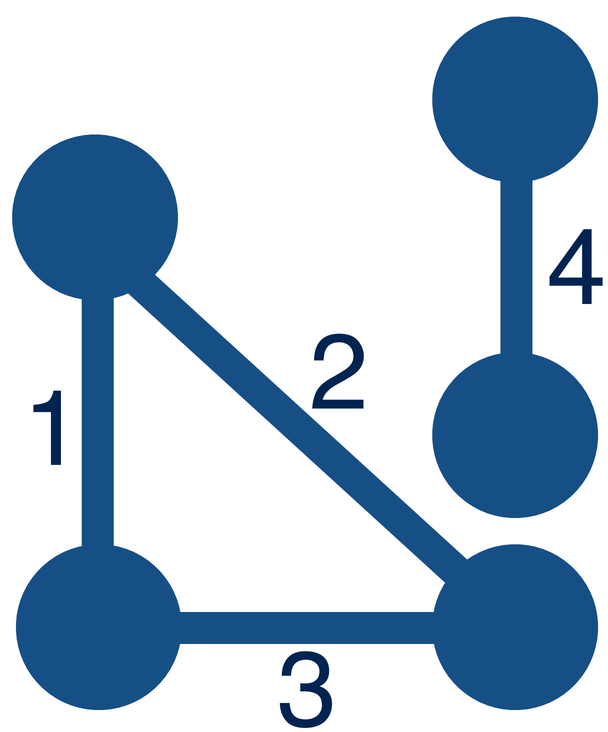


G_1 G_2 G_3 G_4

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

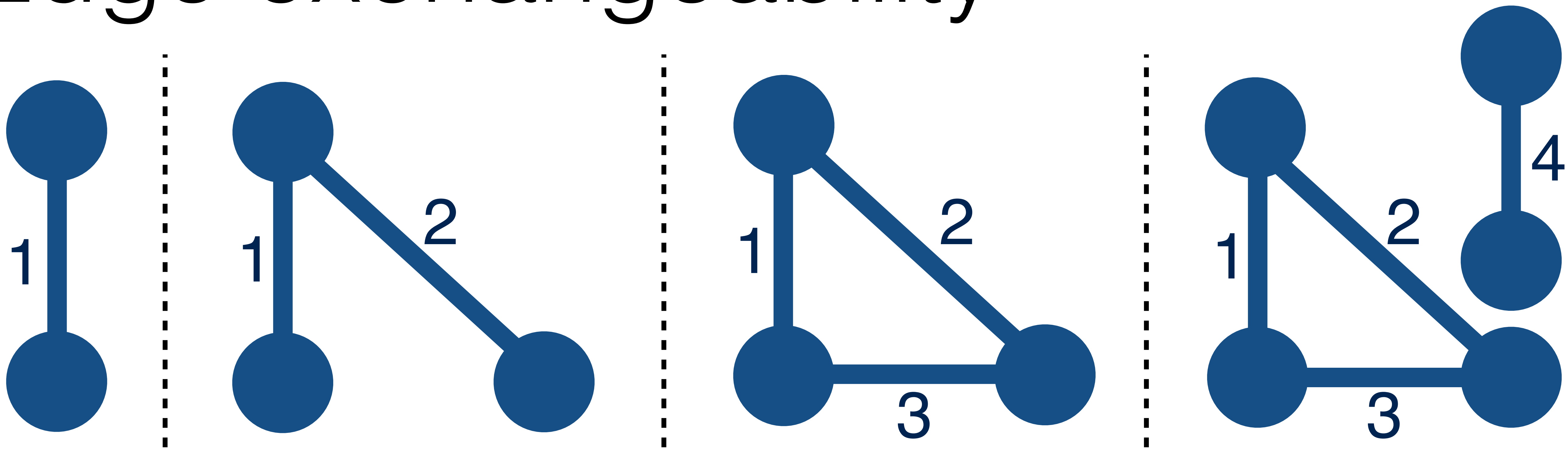
Edge exchangeability



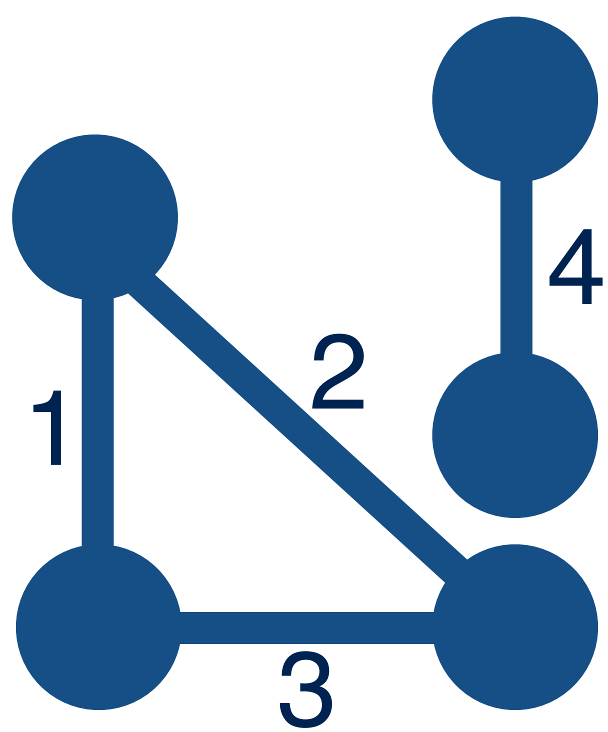


 [Broderick, Cai 2015; Cai, Broderick 2015a,b; Crane, Dempsey 2015a,b,16a; Cai, Campbell, Broderick 2016]

Edge exchangeability

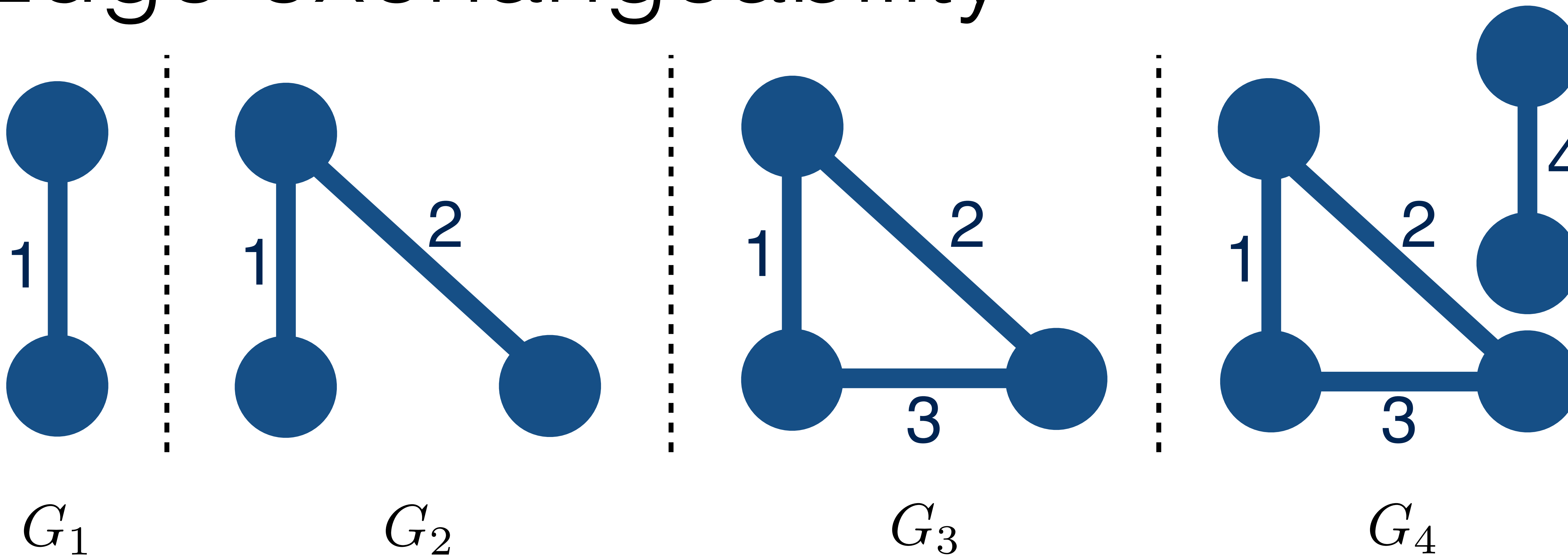


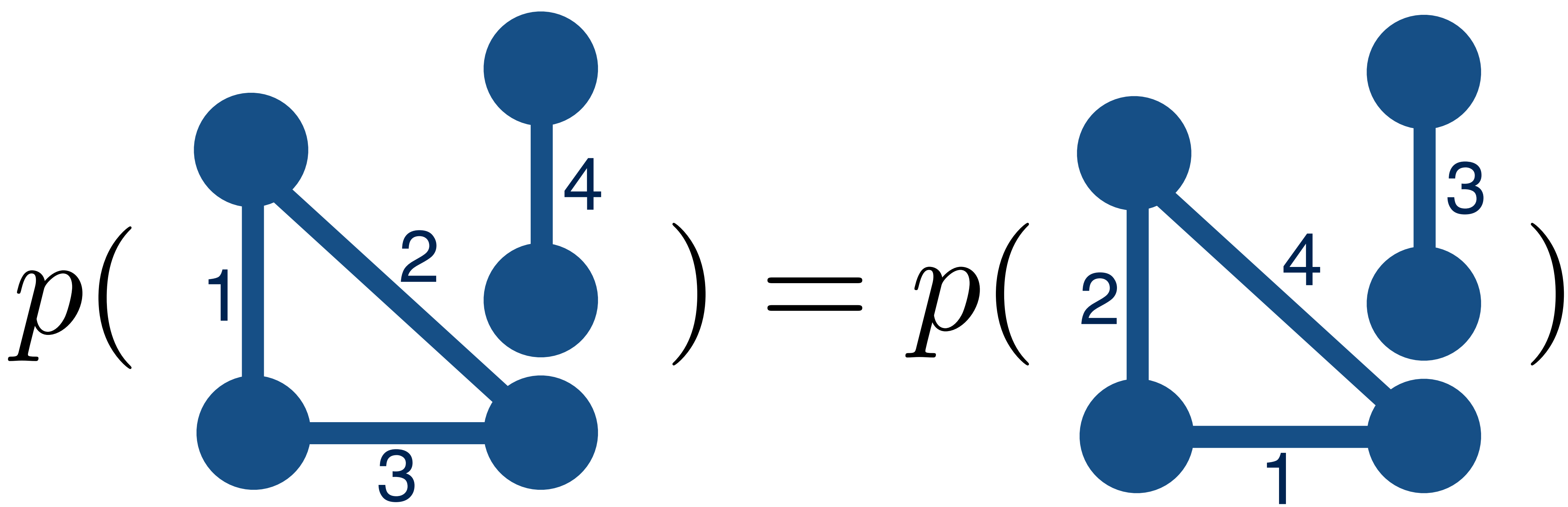
G_1 G_2 G_3 G_4

$p($  $)$

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

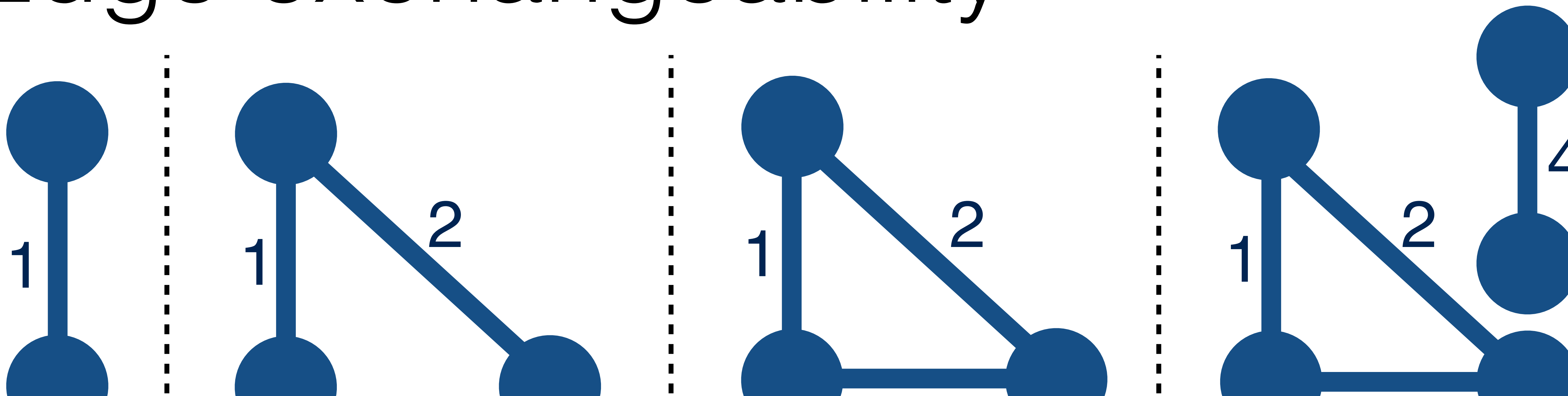
Edge exchangeability



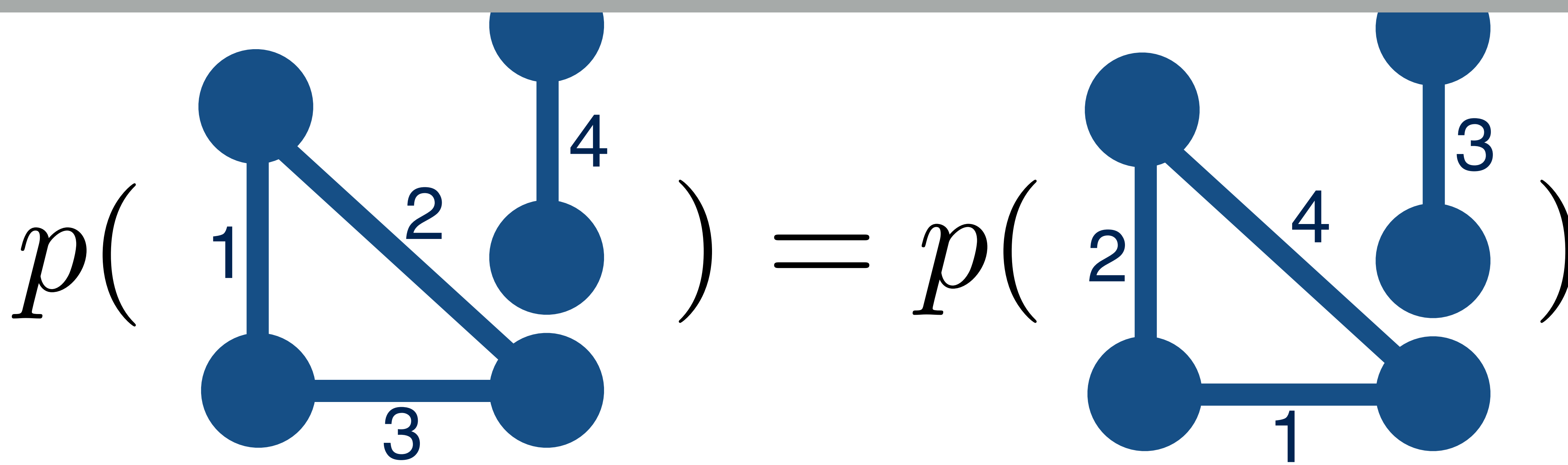


[Broderick, Cai 2015; Cai, Broderick 2015a,b; Crane, Dempsey 2015a,b,16a; Cai, Campbell, Broderick 2016]

Edge exchangeability

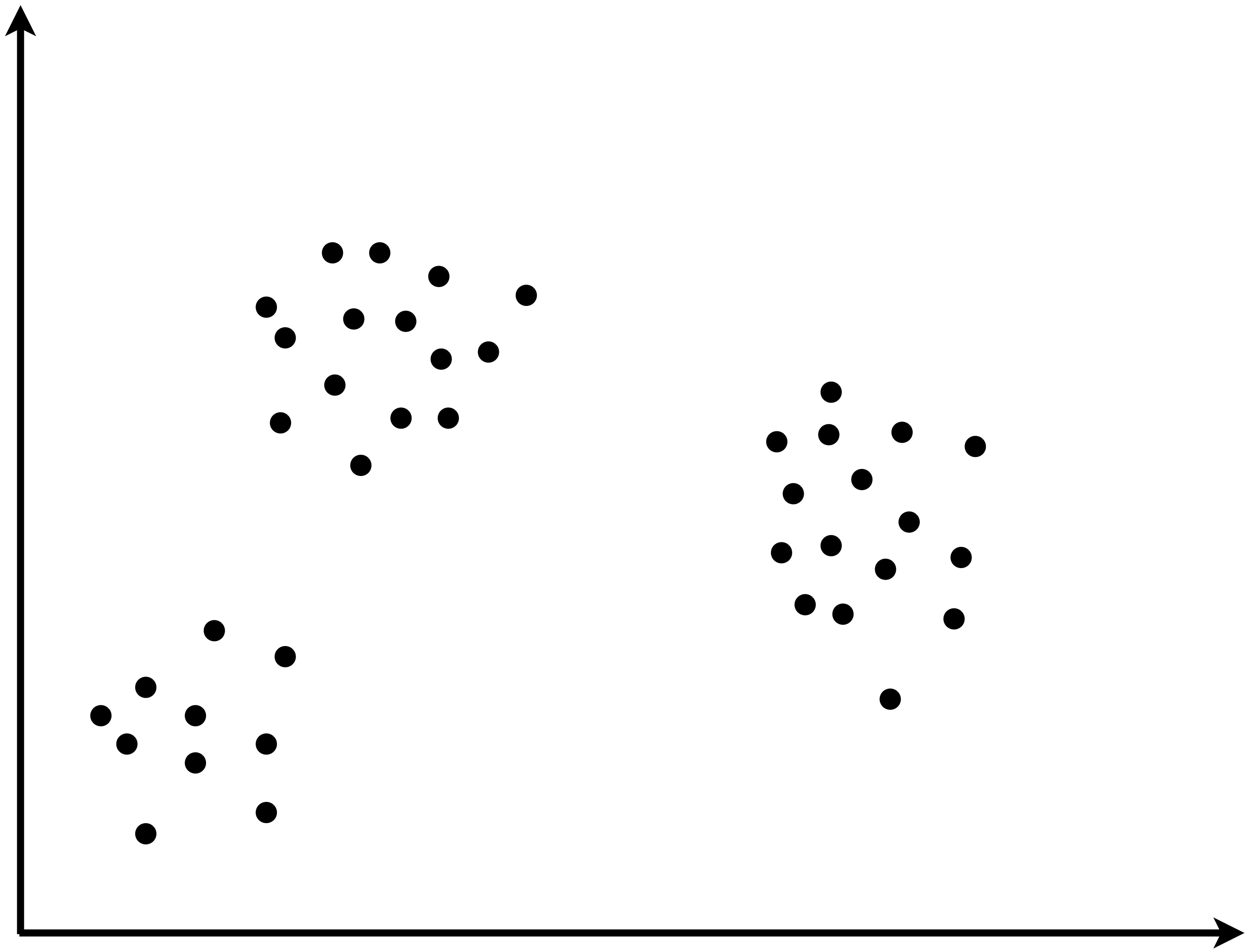


Thm (CCB). A wide class of edge-exchangeable graph models yields sparse graph sequences.

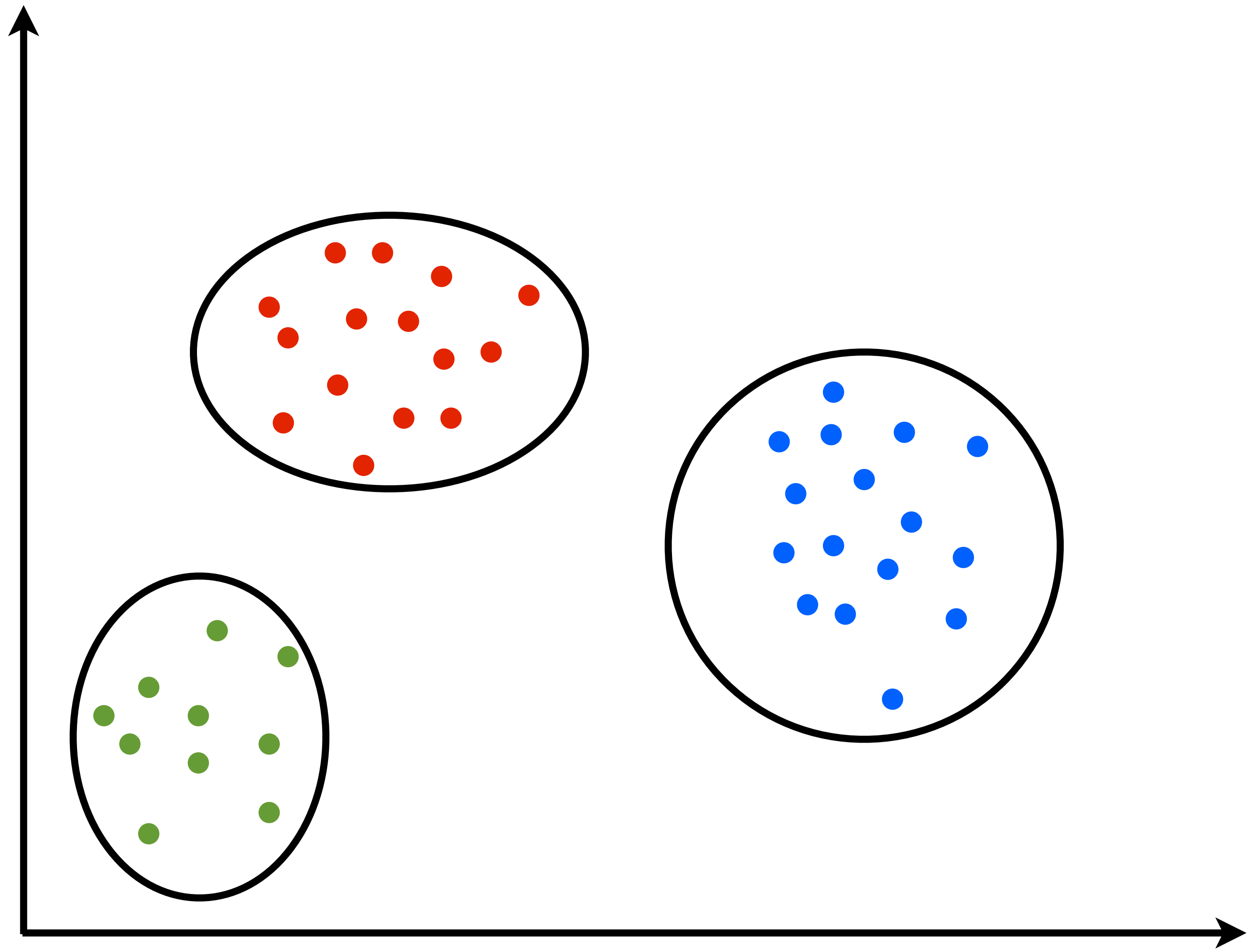


 [Broderick, Cai 2015; Cai, Broderick 2015a,b; Crane, Dempsey 2015a,b,16a; Cai, Campbell, Broderick 2016]

Clustering

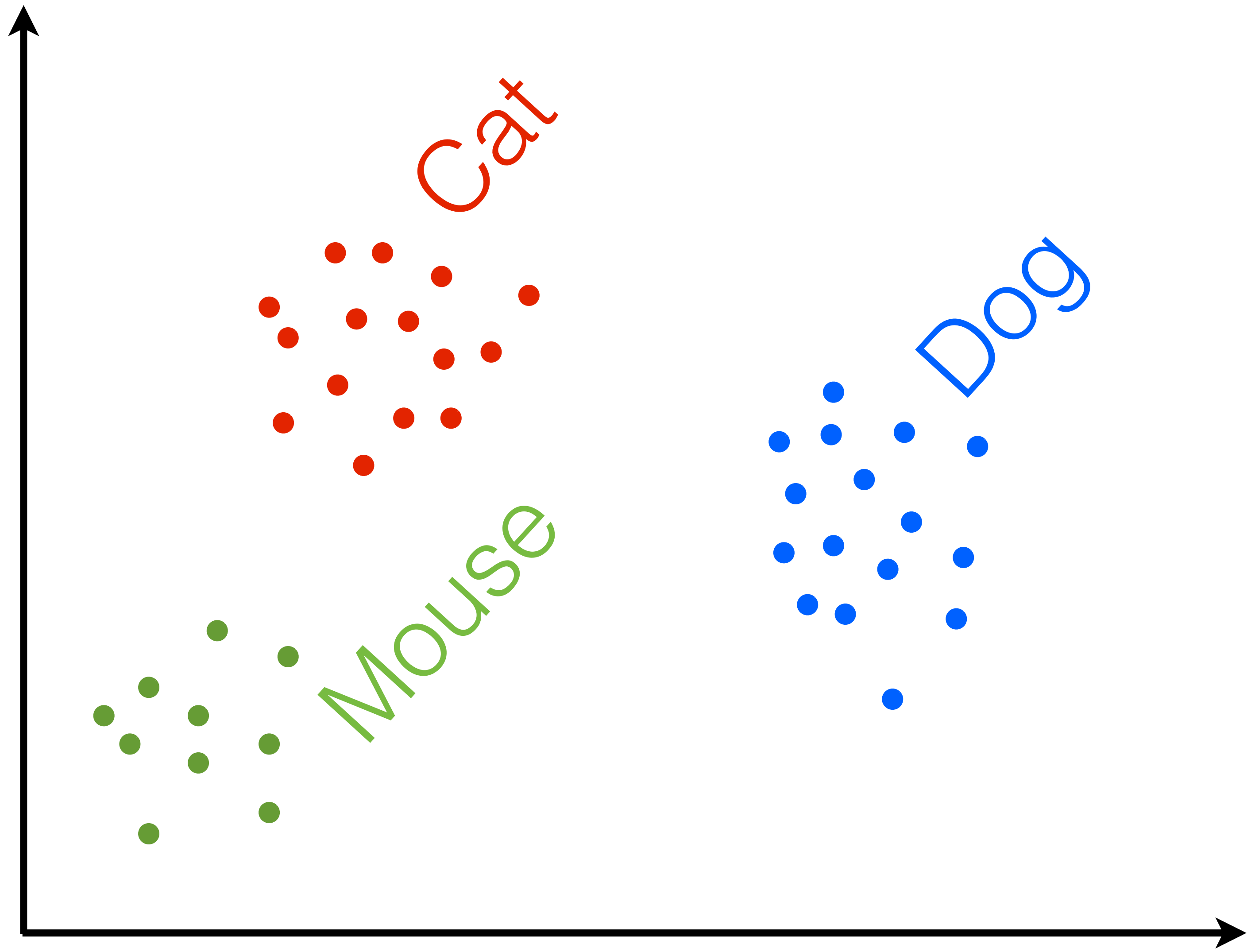


Clustering



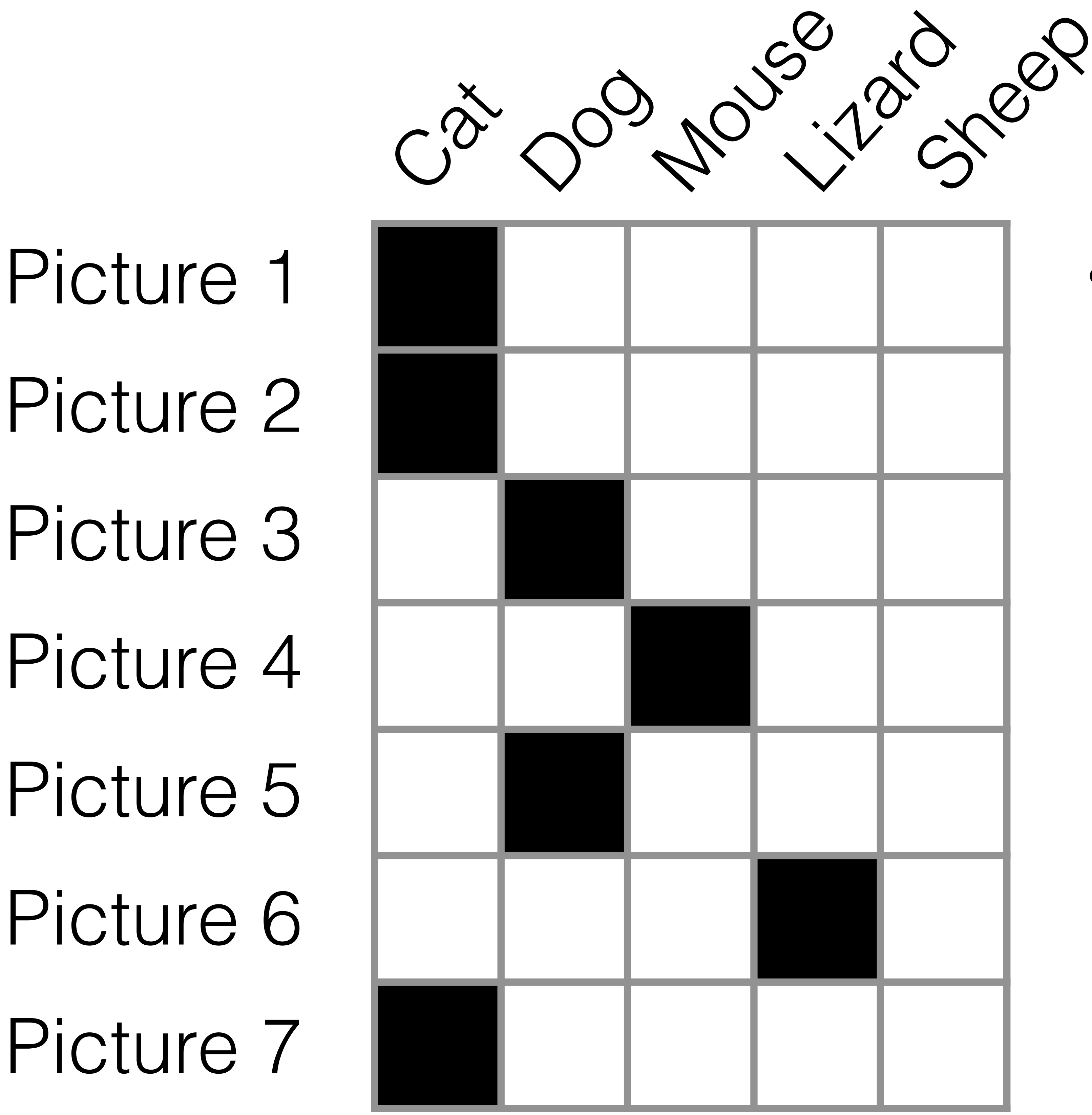
“Clusters”

Clustering



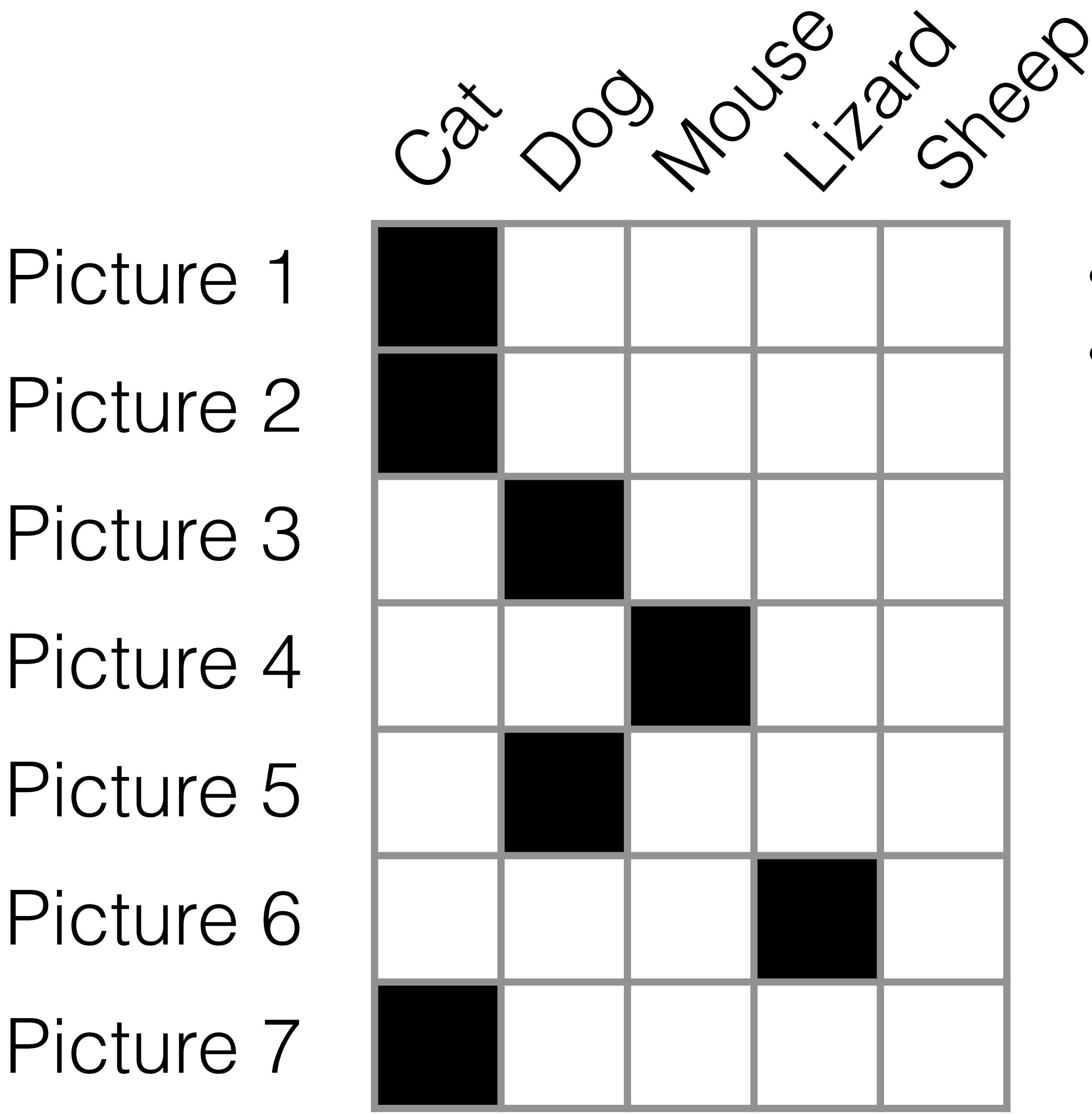
“Clusters”

Clustering



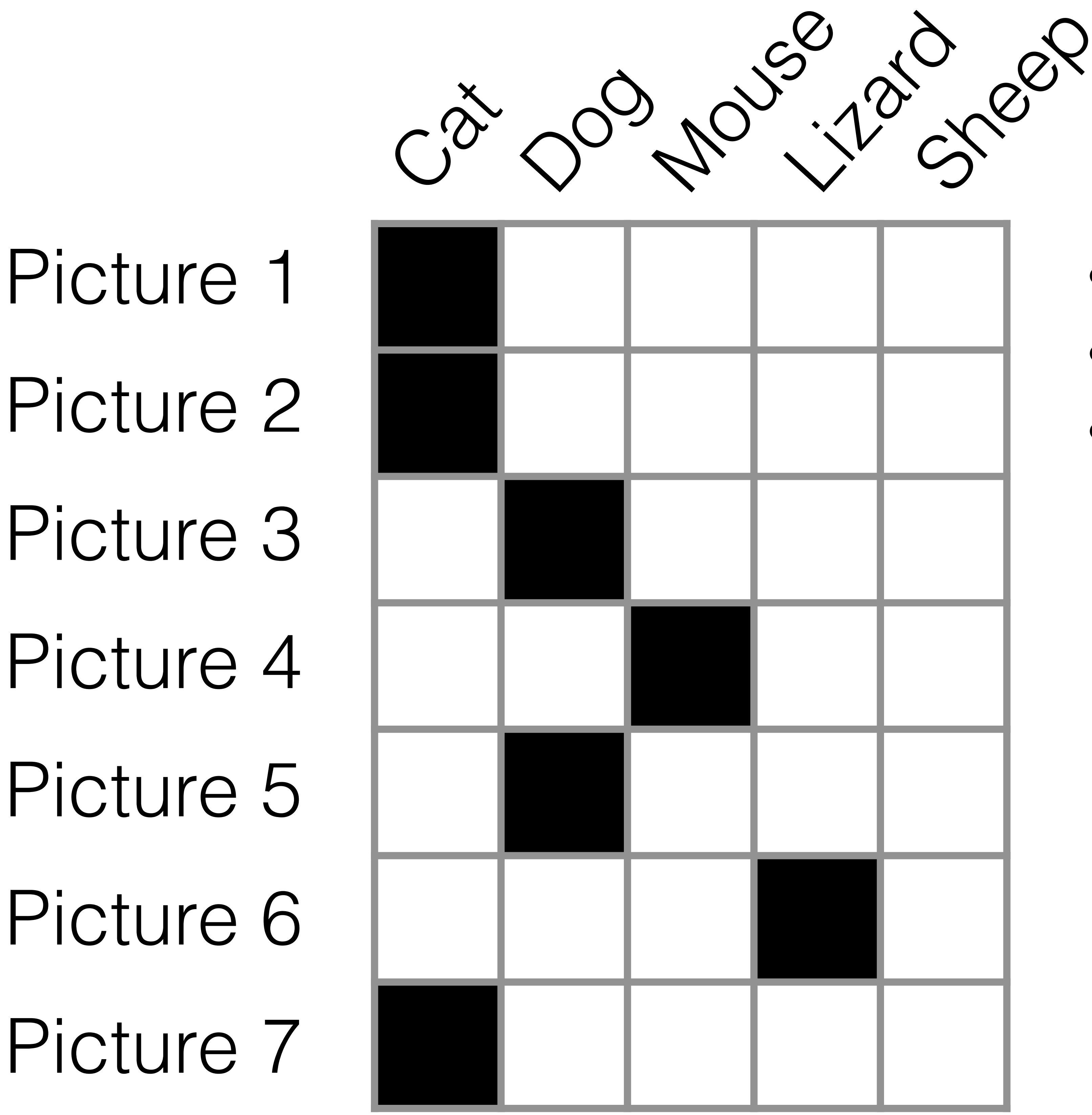
- Groups: clusters

Clustering



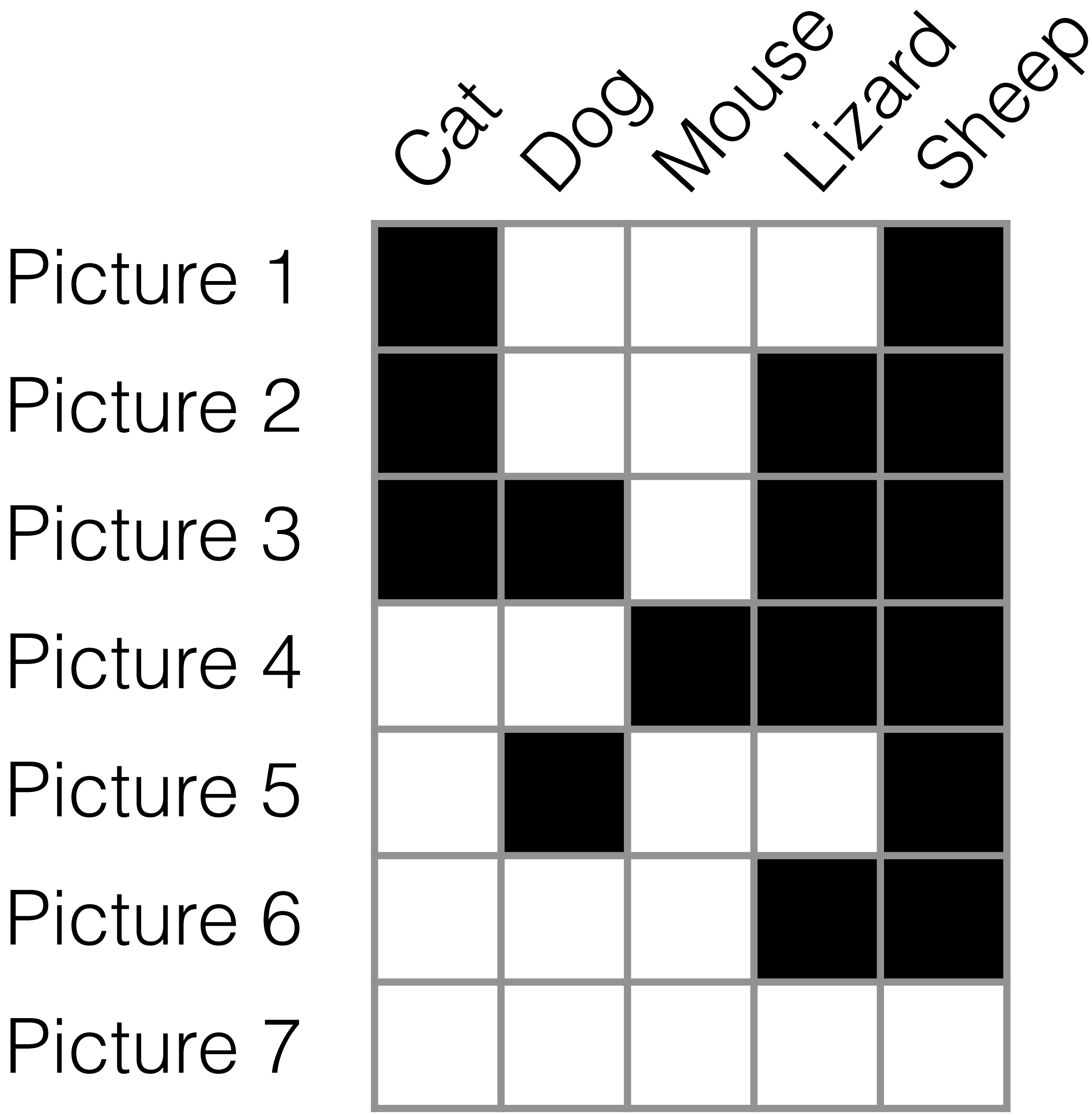
- Groups: clusters
- Exchangeable

Clustering

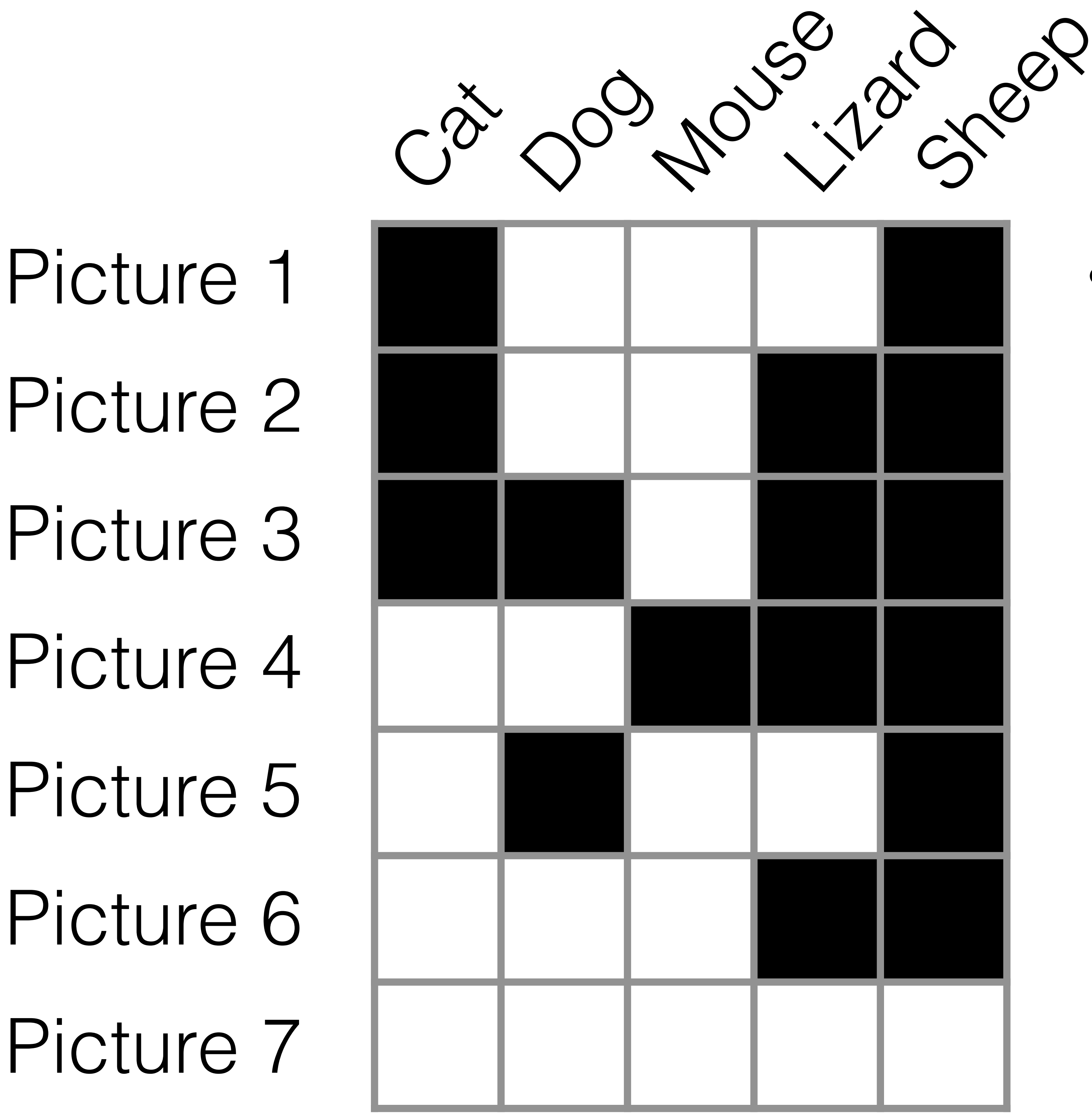


- Groups: clusters
- Exchangeable
- Chinese restaurant process/Dirichlet process

Feature allocation

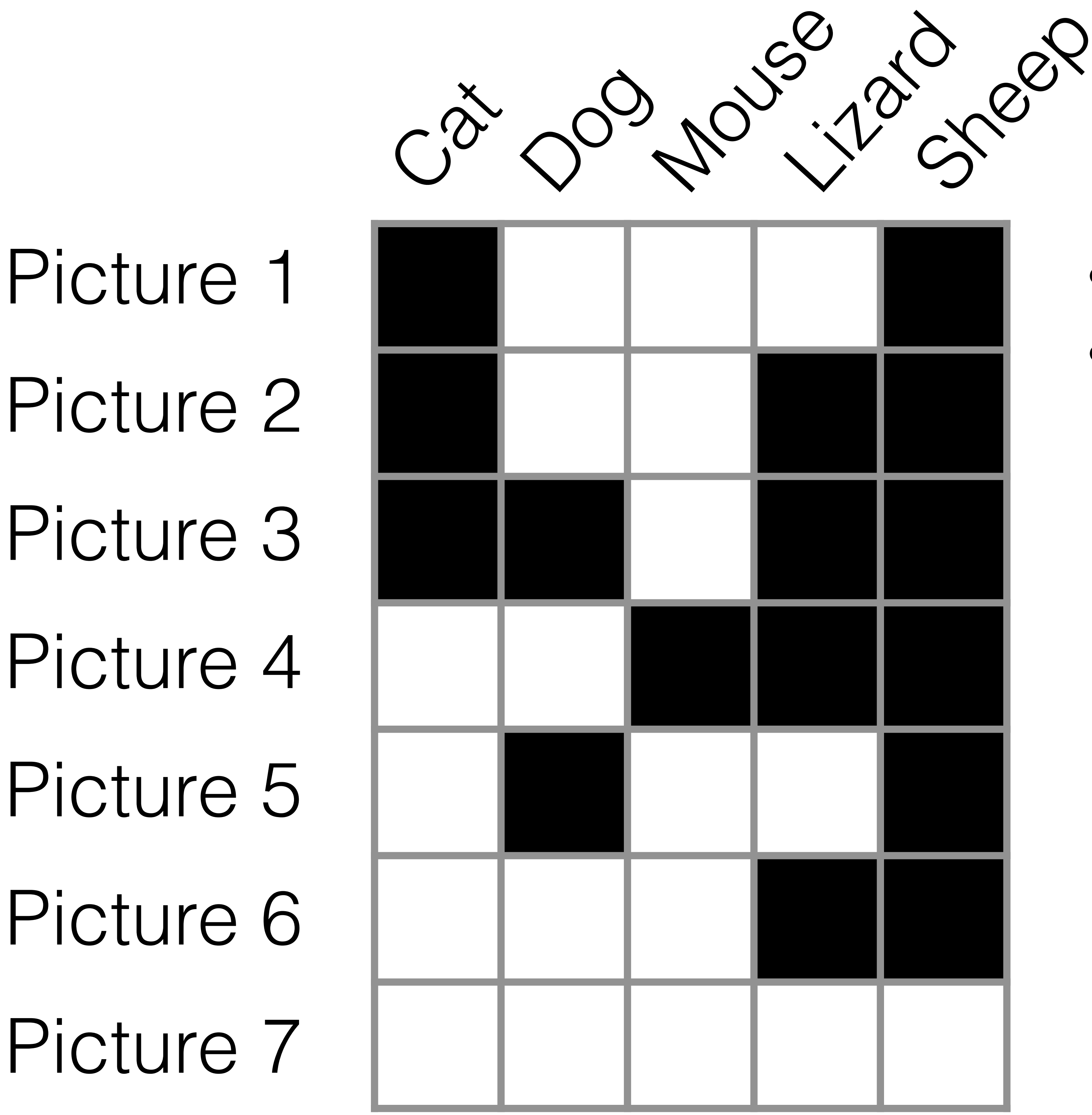


Feature allocation



- Groups: features

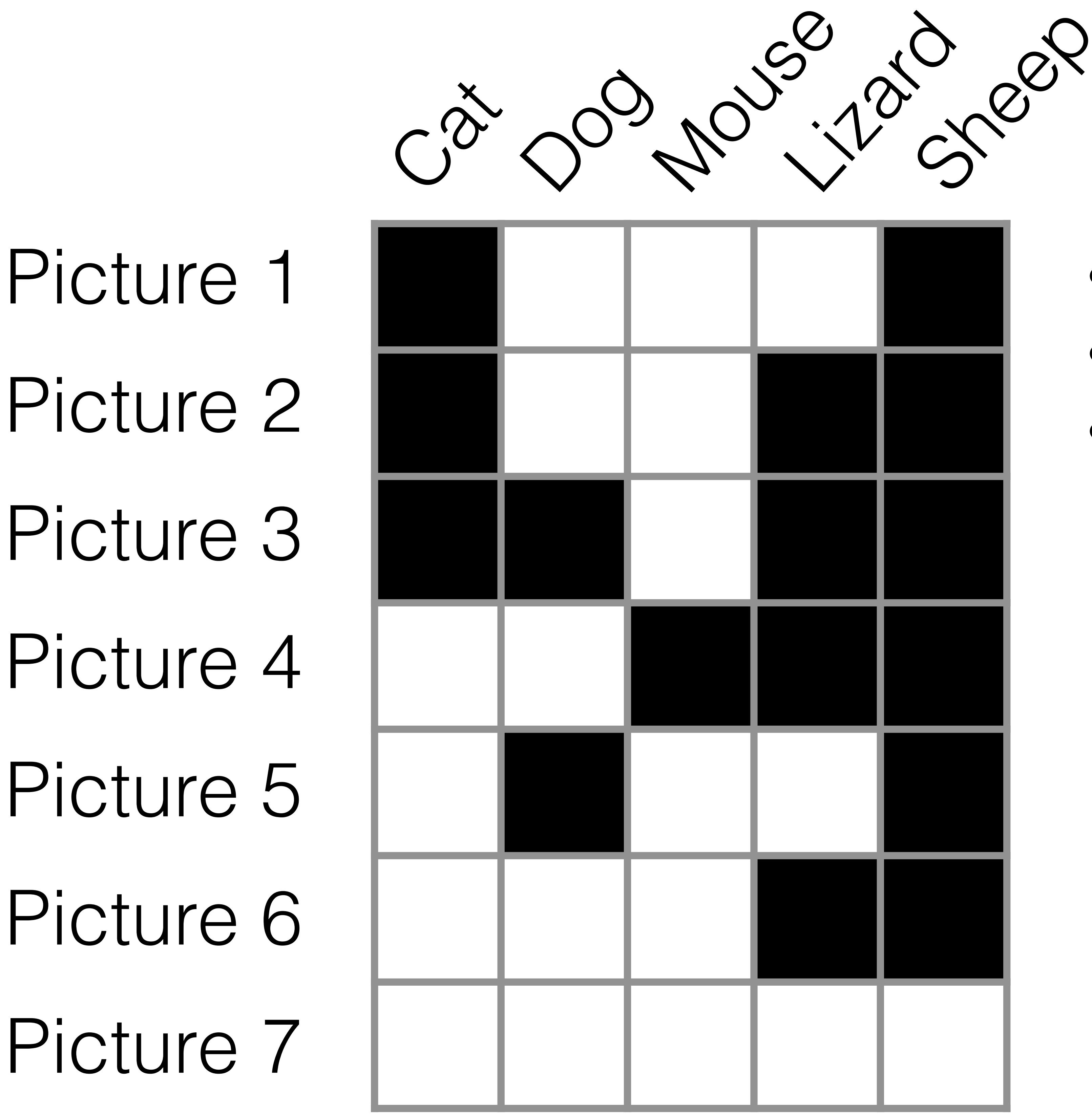
Feature allocation



- Groups: features
- Exchangeable

[Broderick, Jordan, Pitman 2013; Broderick, Pitman, Jordan 2013]

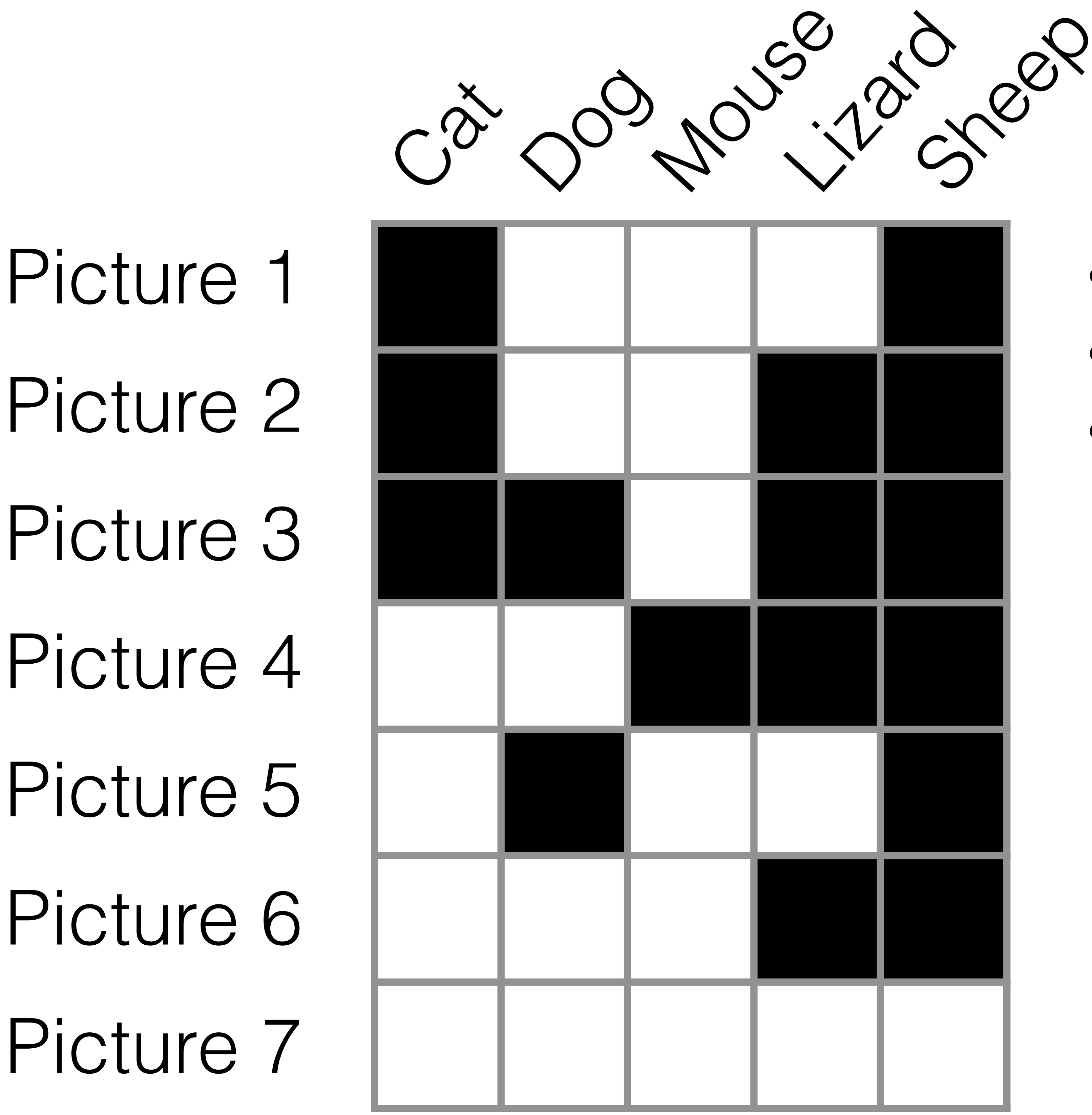
Feature allocation



- Groups: features
- Exchangeable
- Indian buffet process/beta process

[Broderick, Jordan, Pitman 2013; Broderick, Pitman, Jordan 2013]

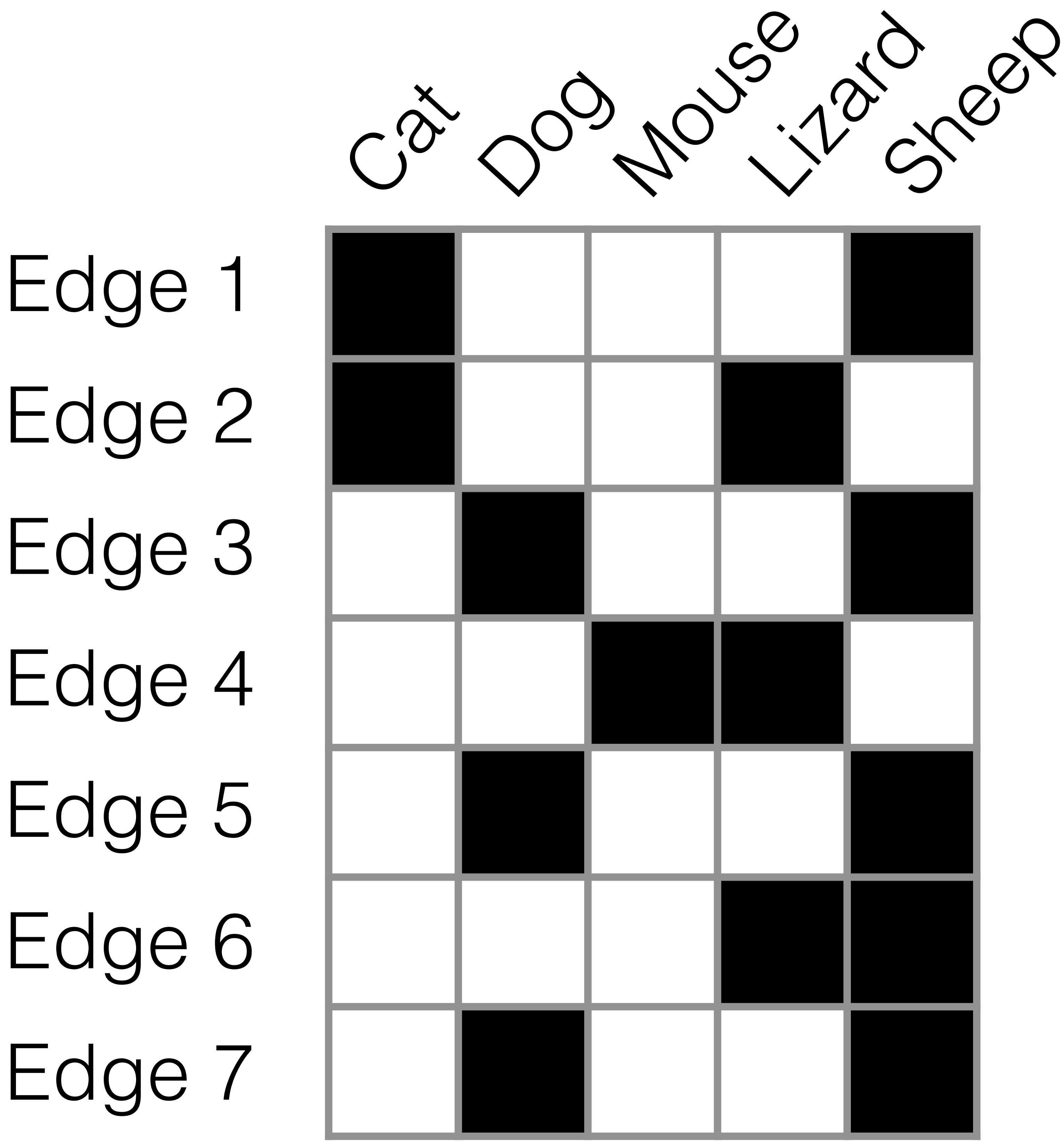
Feature allocation



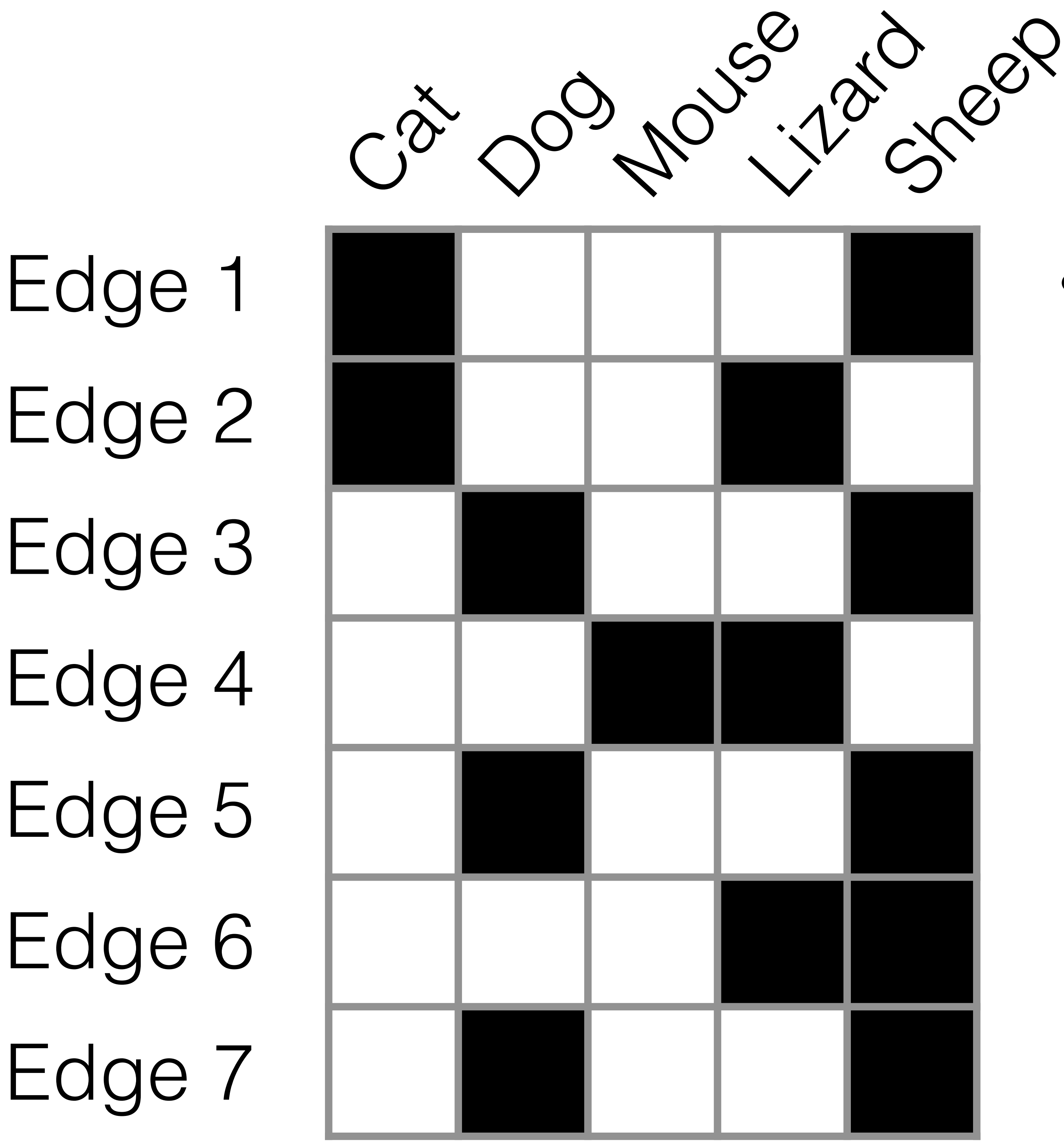
- Groups: features
- Exchangeable
- Indian buffet process/beta process

[Broderick, Jordan, Pitman 2013; Broderick, Pitman, Jordan 2013; Campbell, Cai, Broderick 2016]

Graph

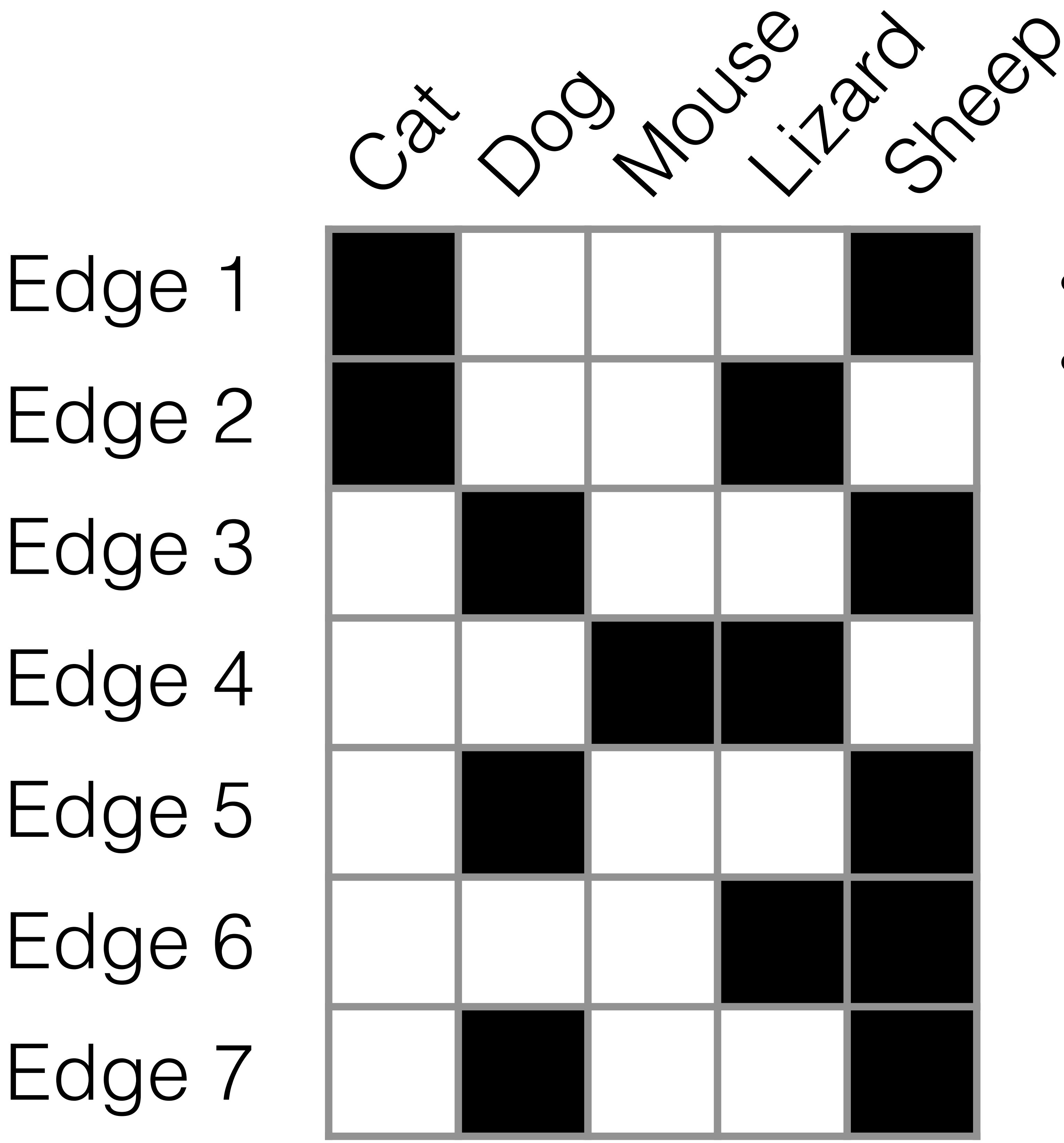

Cat
Dog
Mouse
Lizard
Sheep
Edge 1
Edge 2
Edge 3
Edge 4
Edge 5
Edge 6
Edge 7

Graph



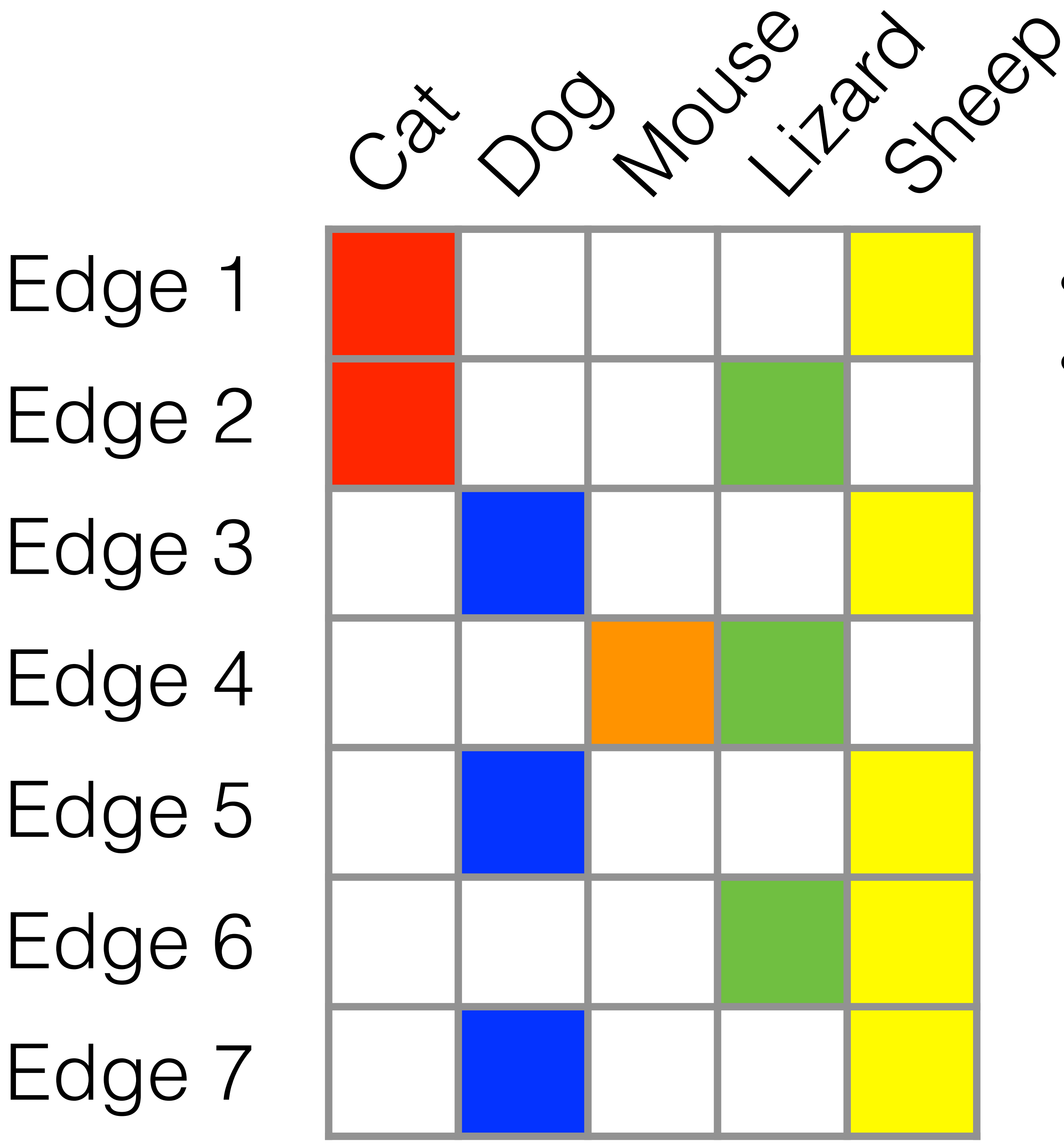
- Groups: vertices

Graph



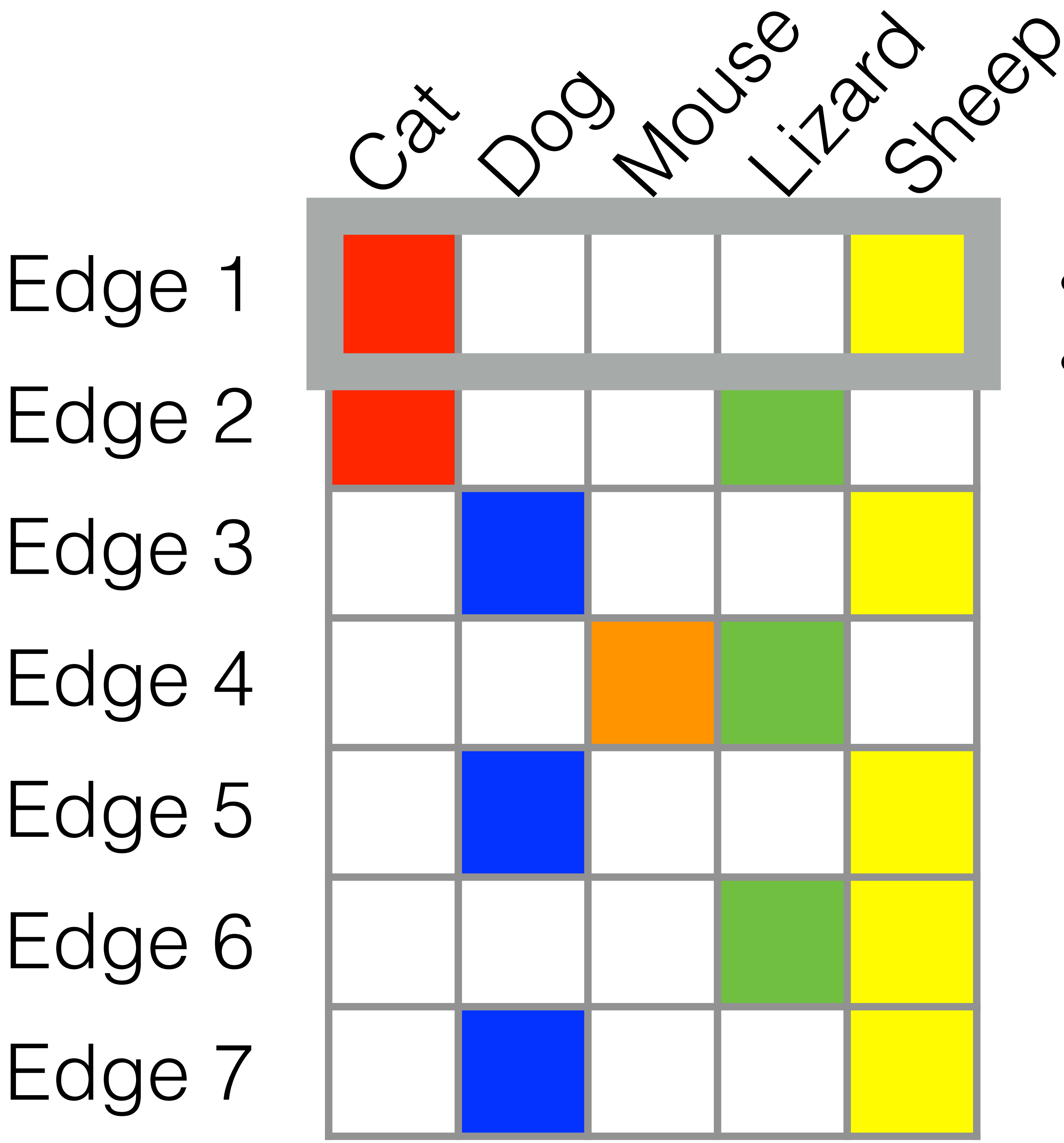
- Groups: vertices
- Edge-exchangeable

Graph



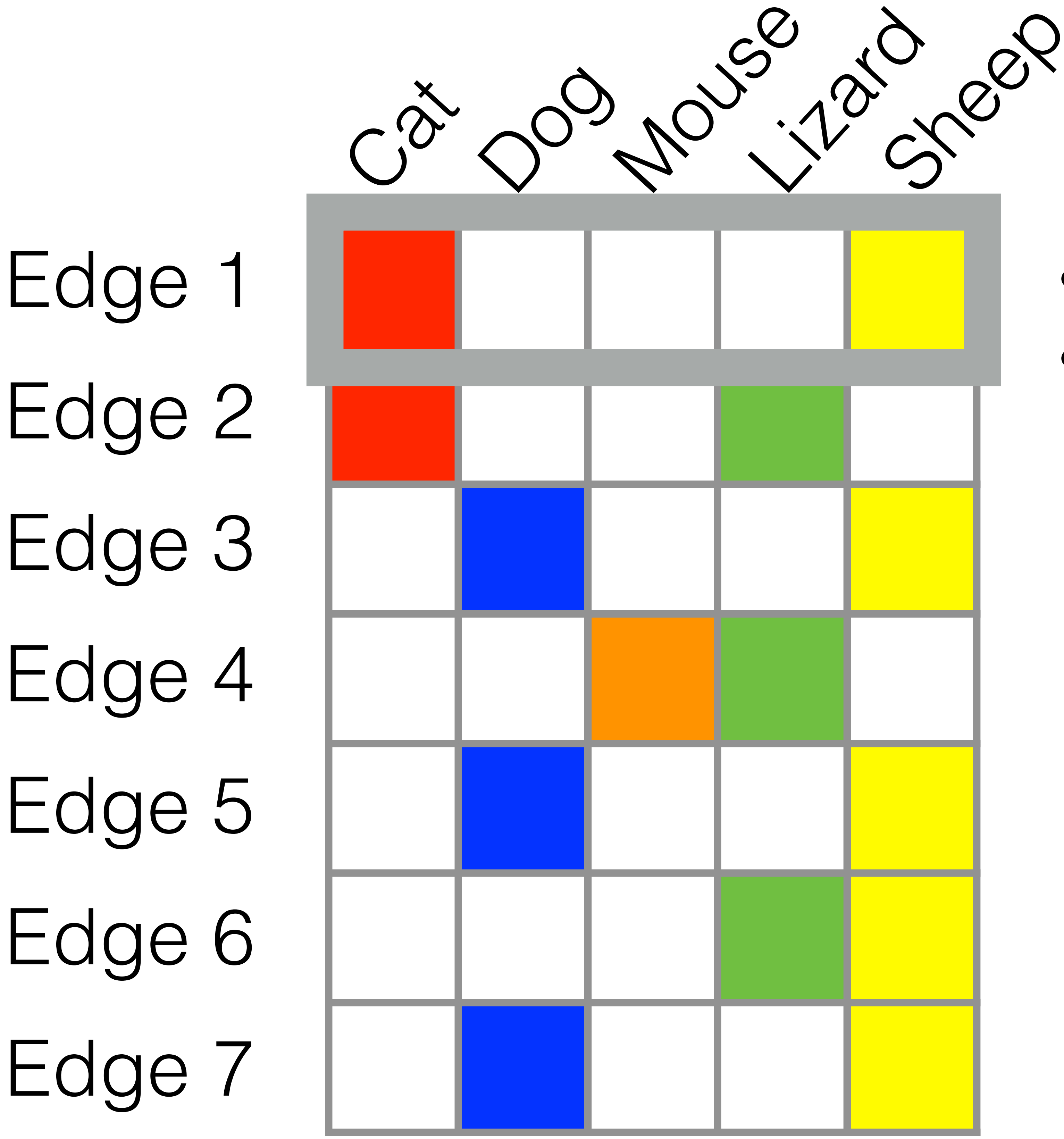
- Groups: vertices
- Edge-exchangeable

Graph

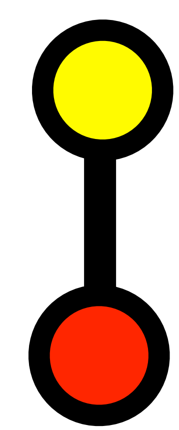


- Groups: vertices
- Edge-exchangeable

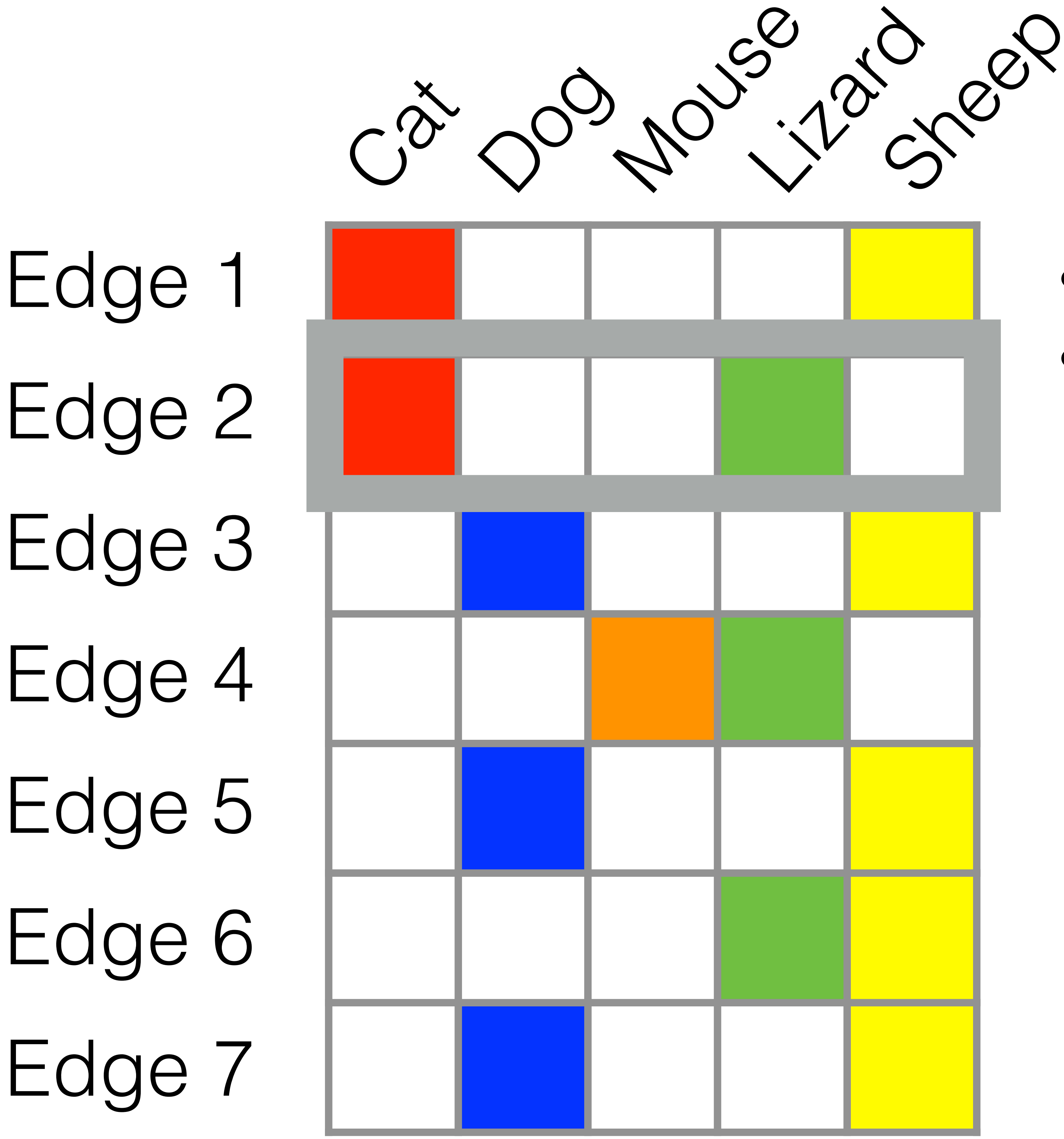
Graph



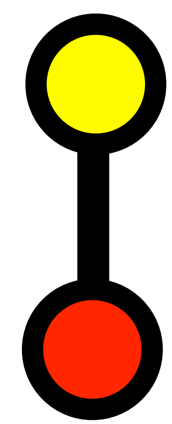
- Groups: vertices
- Edge-exchangeable



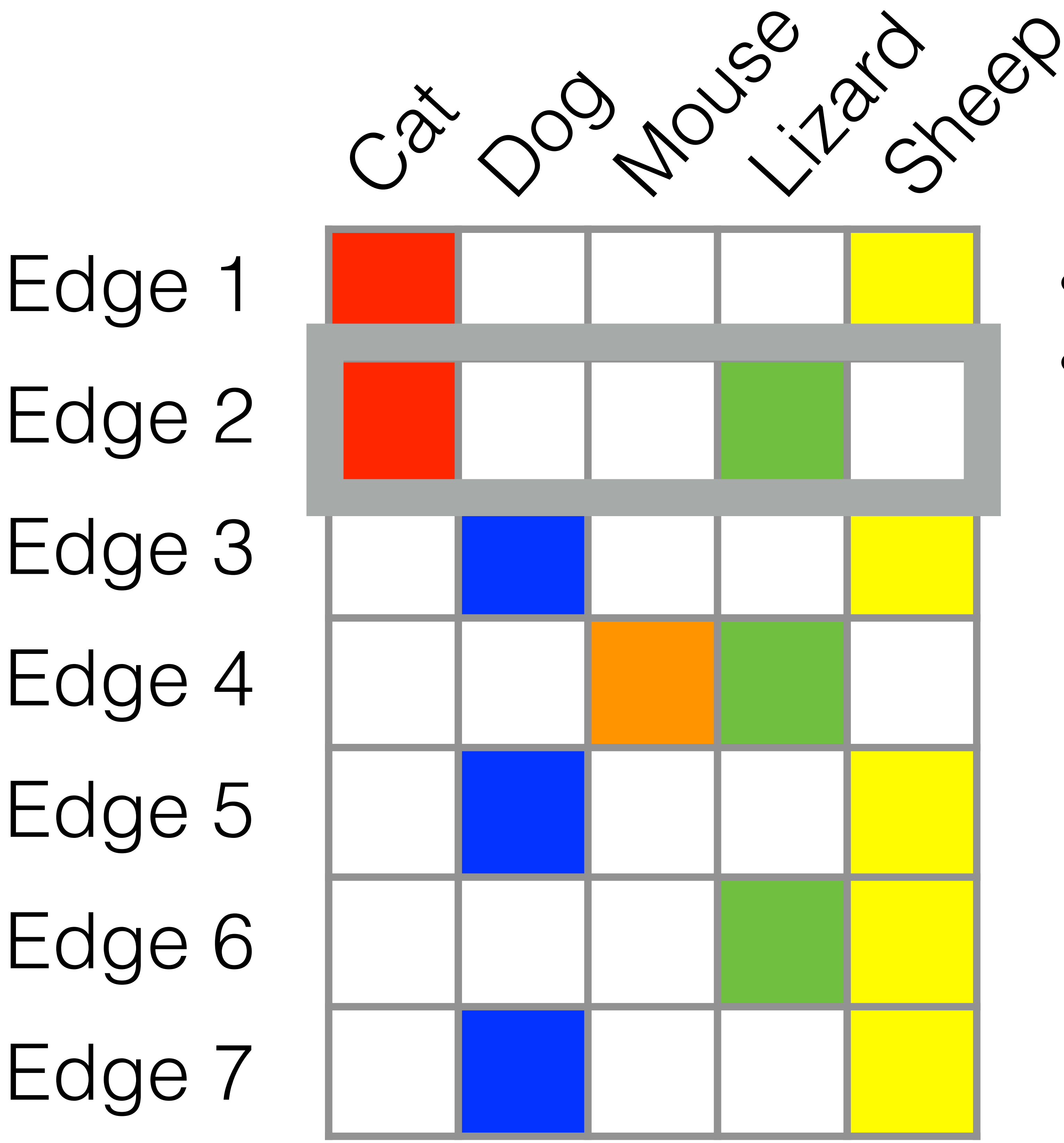
Graph



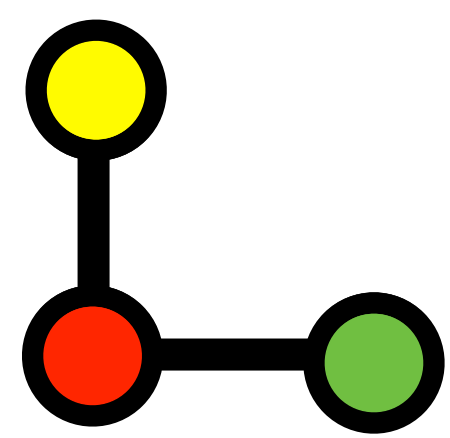
- Groups: vertices
- Edge-exchangeable



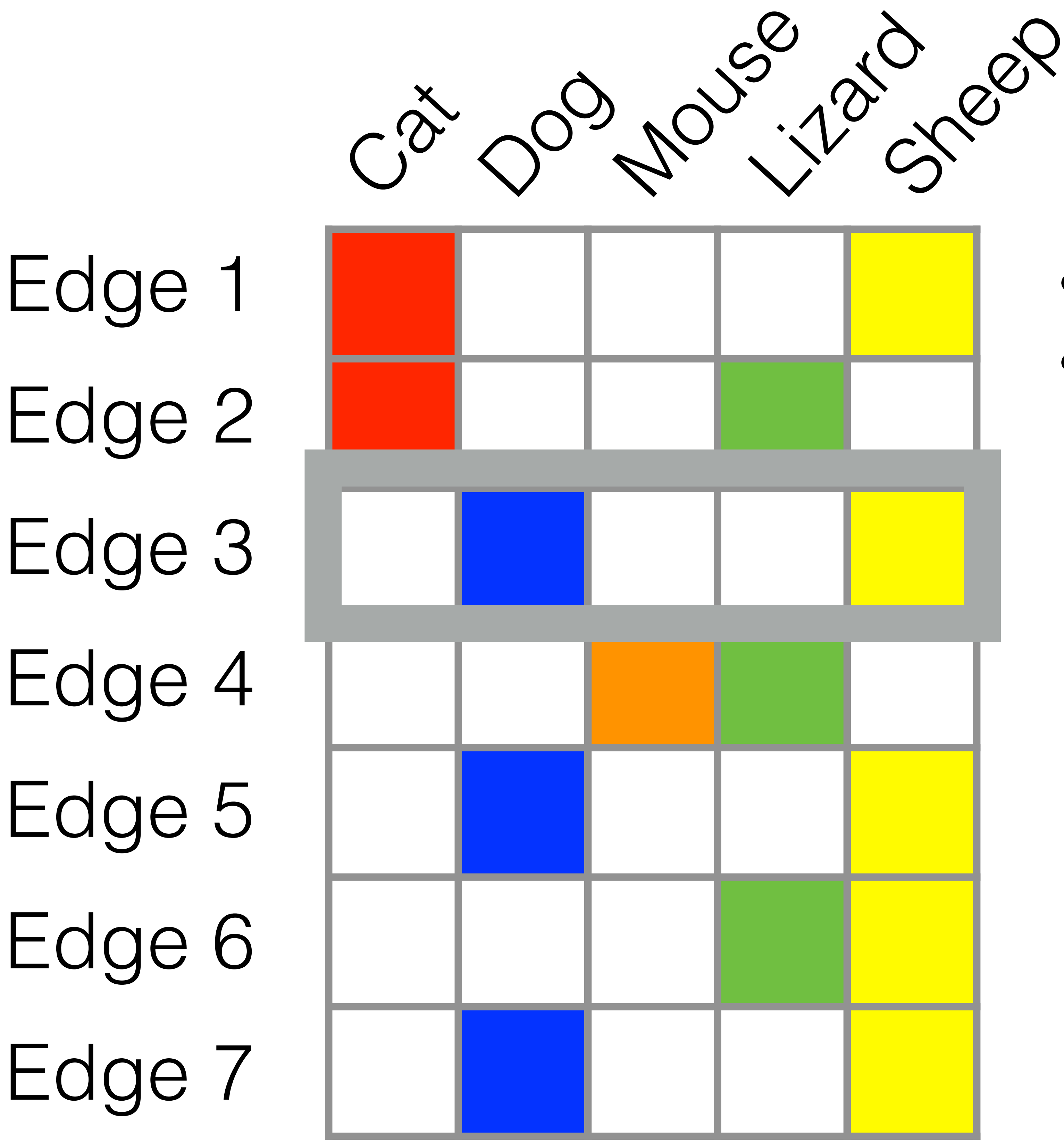
Graph



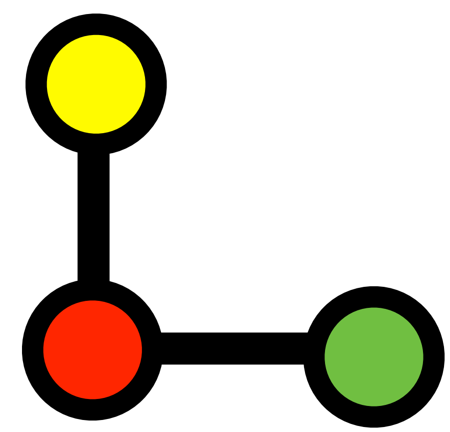
- Groups: vertices
- Edge-exchangeable



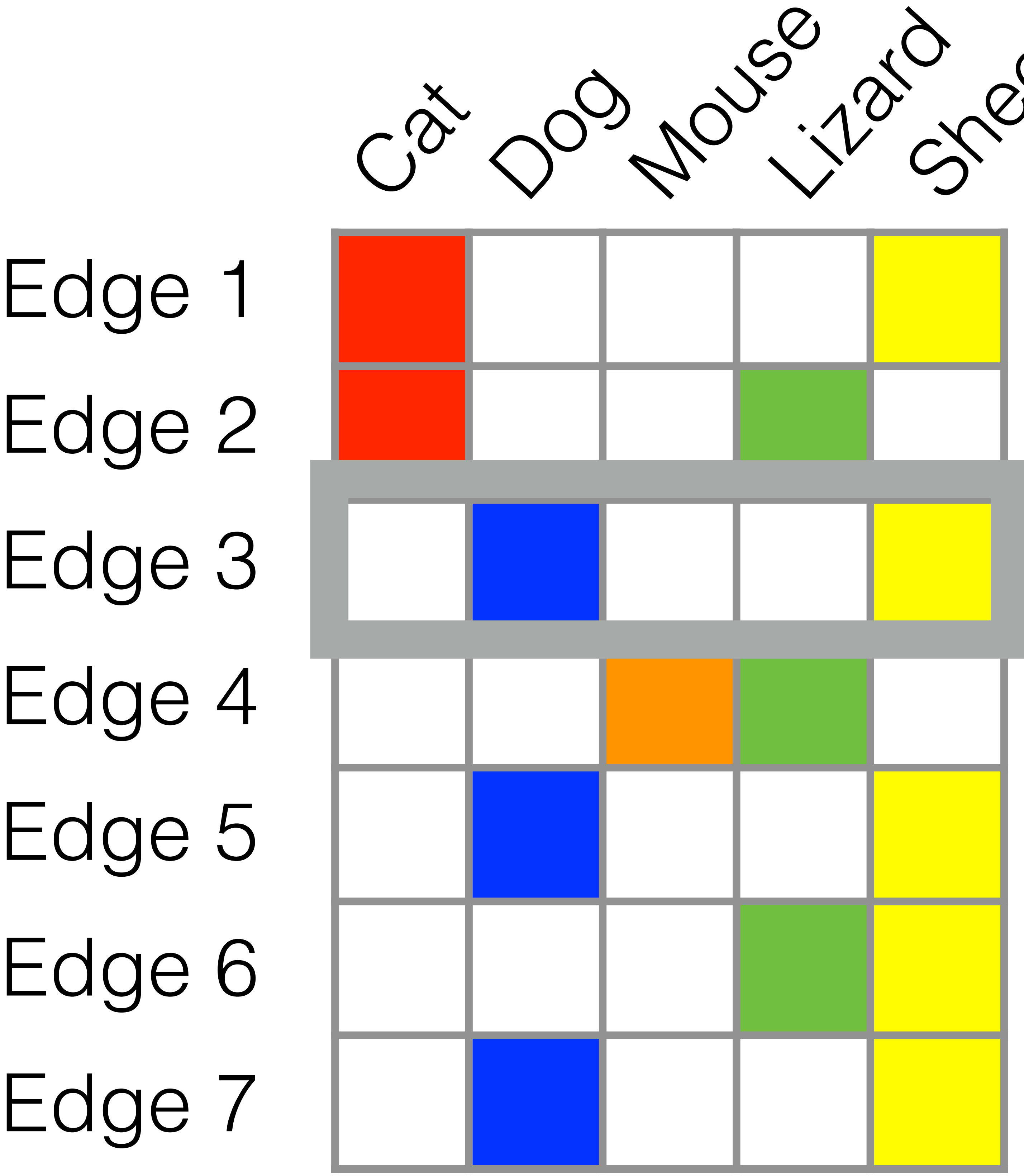
Graph



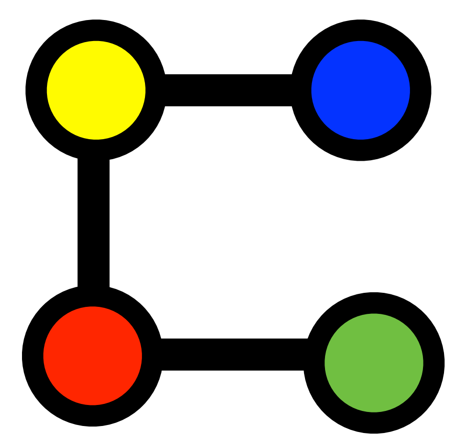
- Groups: vertices
- Edge-exchangeable



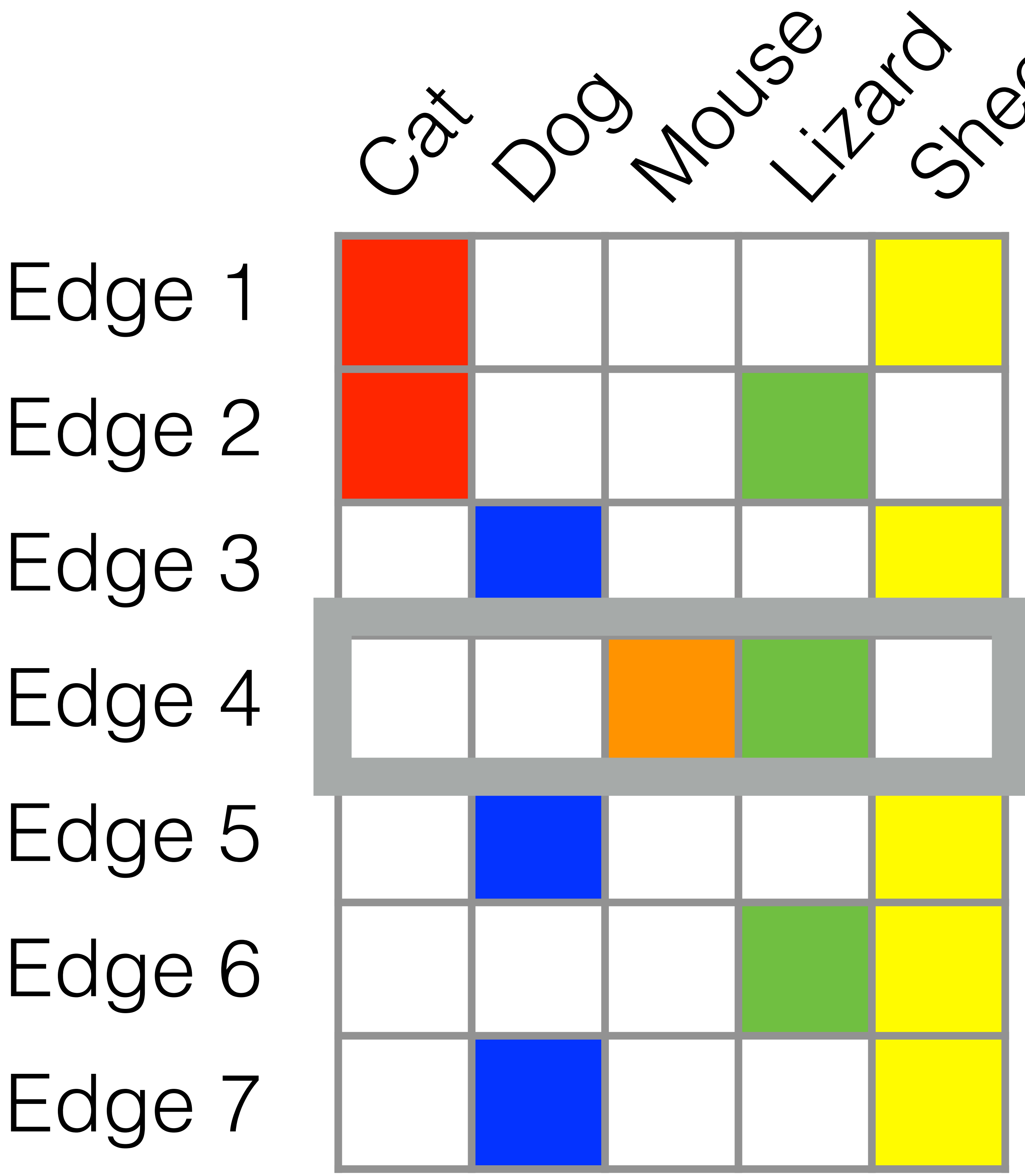
Graph



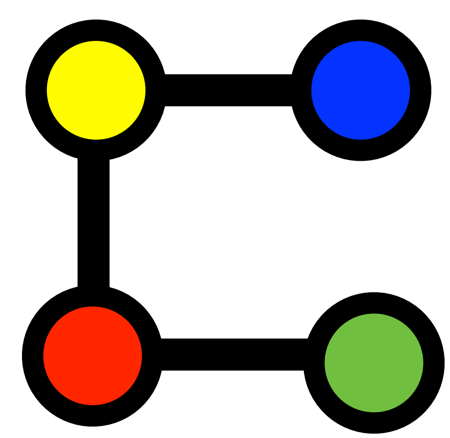
- Groups: vertices
- Edge-exchangeable



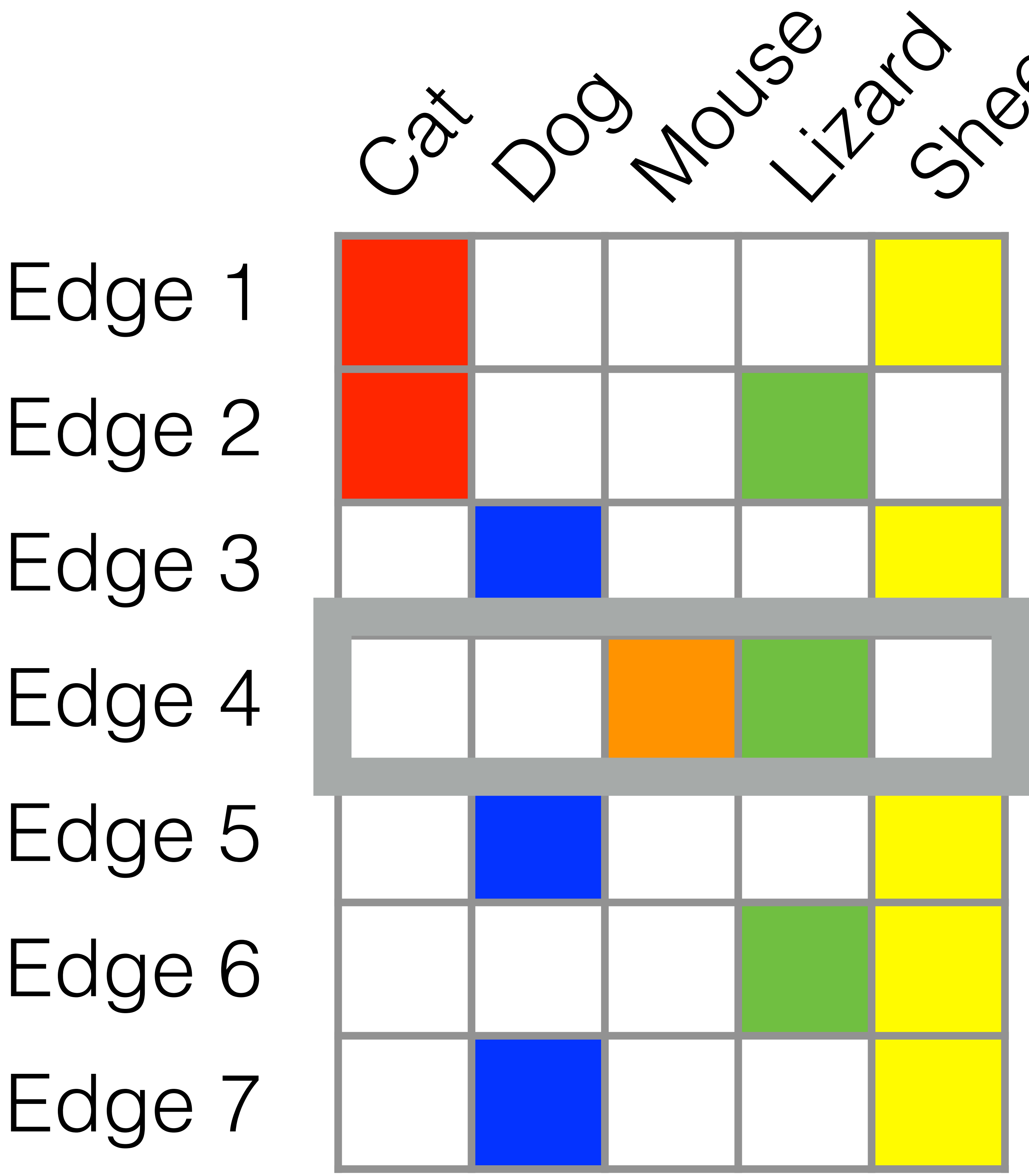
Graph



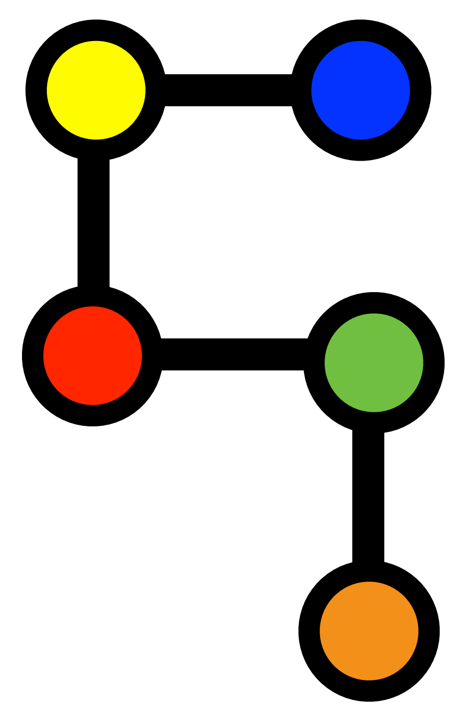
- Groups: vertices
- Edge-exchangeable



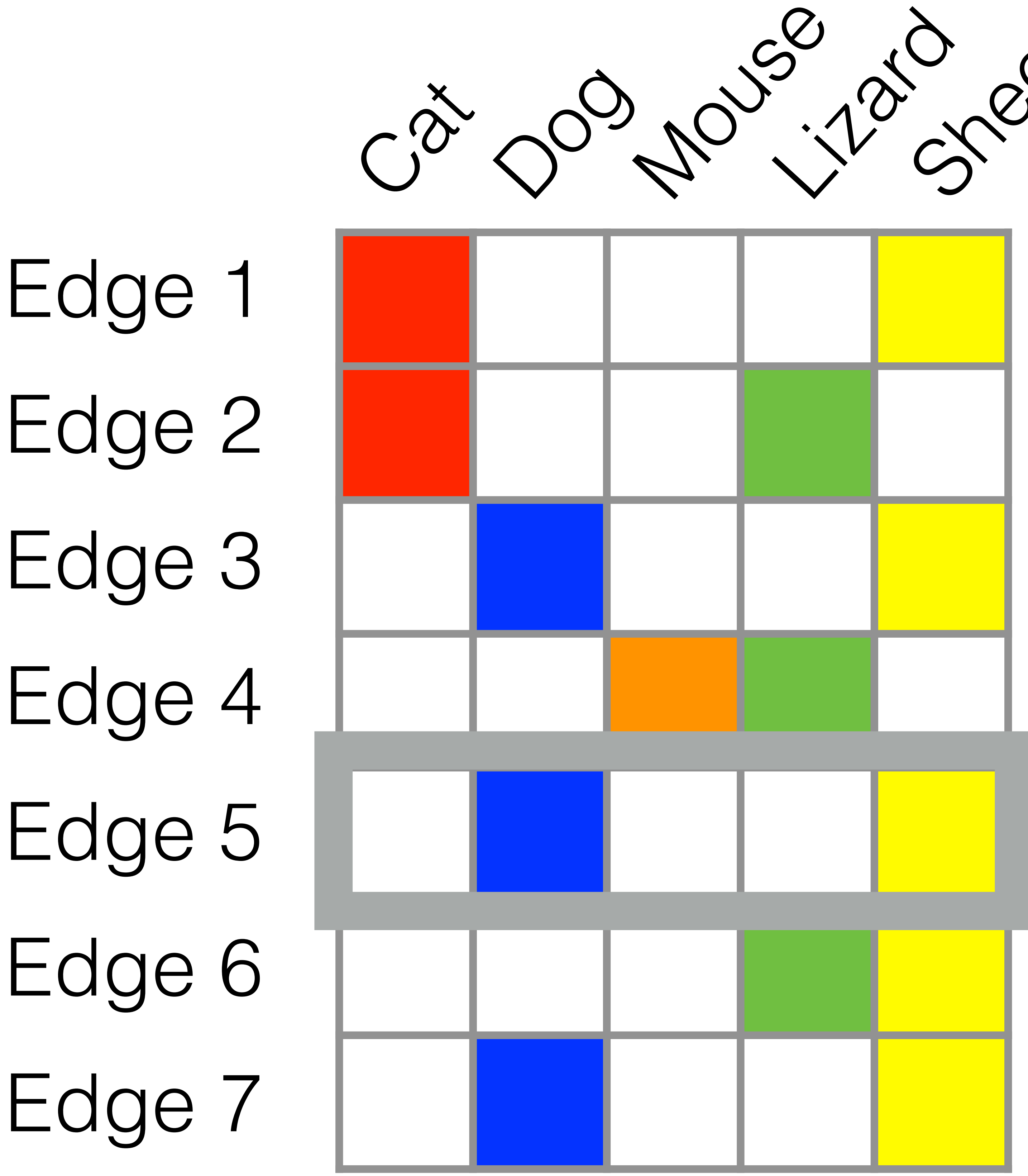
Graph



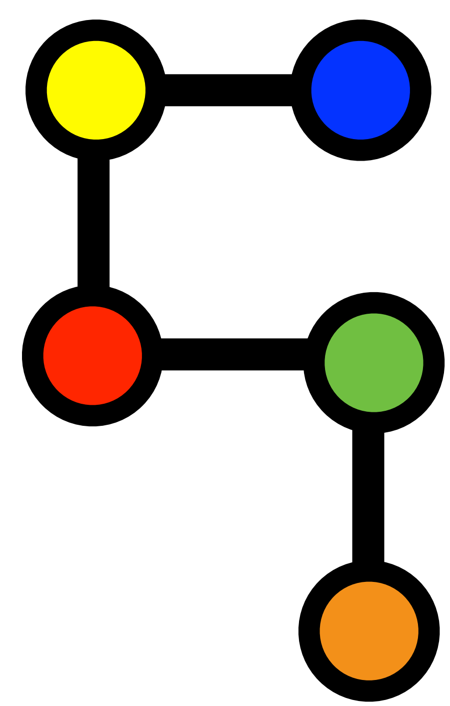
- Groups: vertices
- Edge-exchangeable



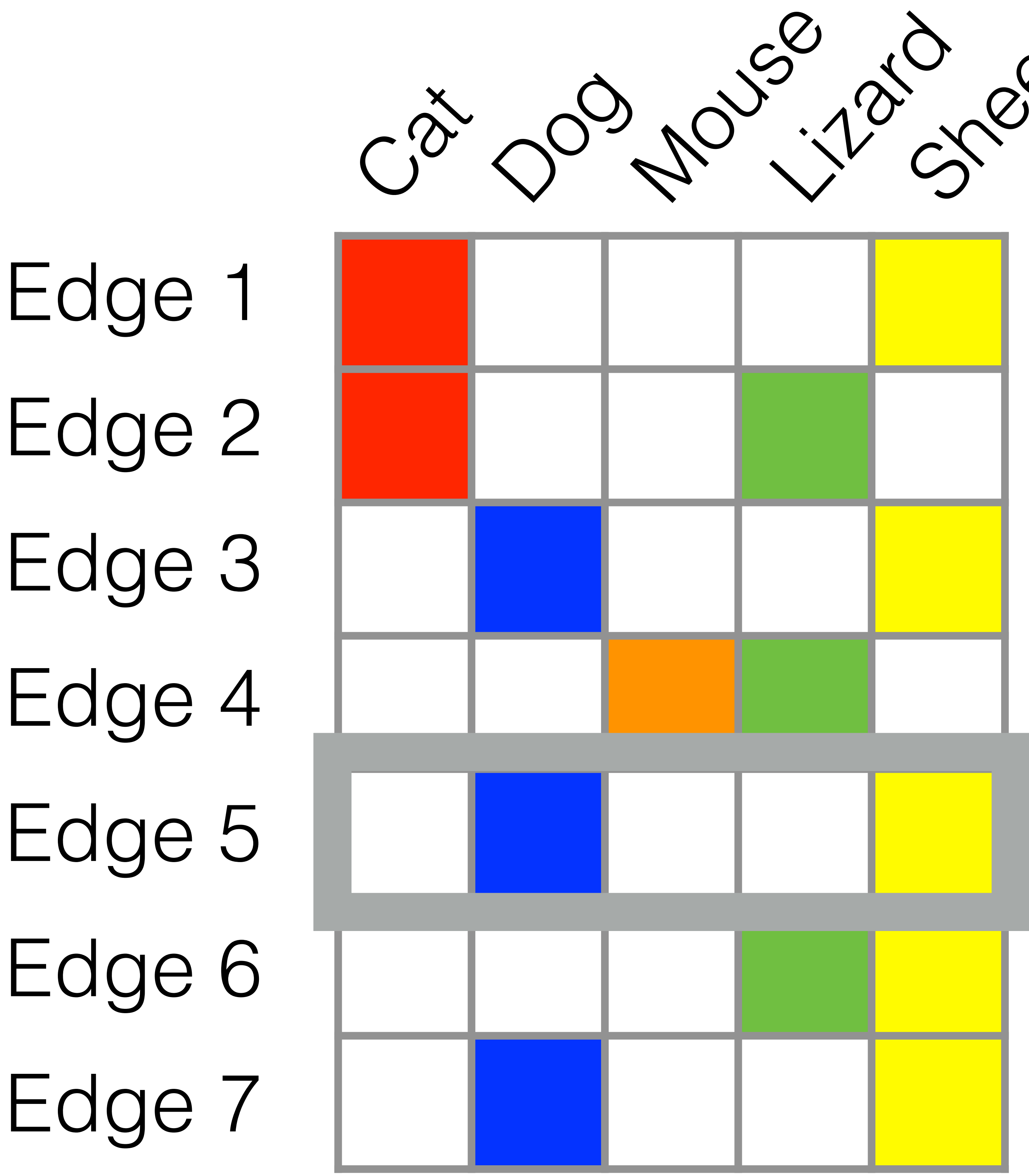
Graph



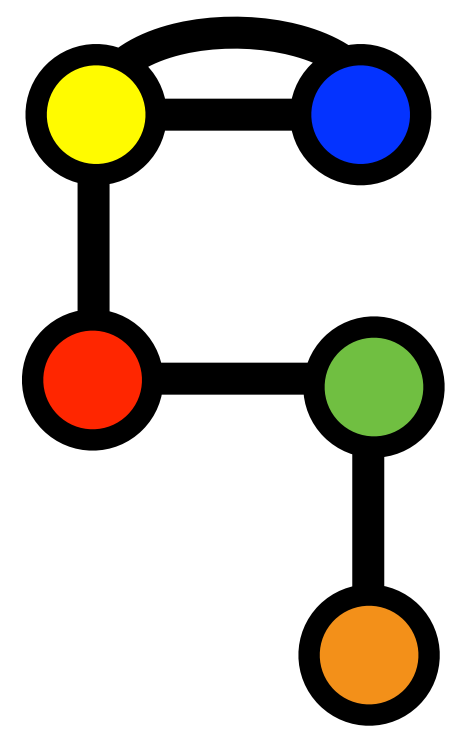
- Groups: vertices
- Edge-exchangeable



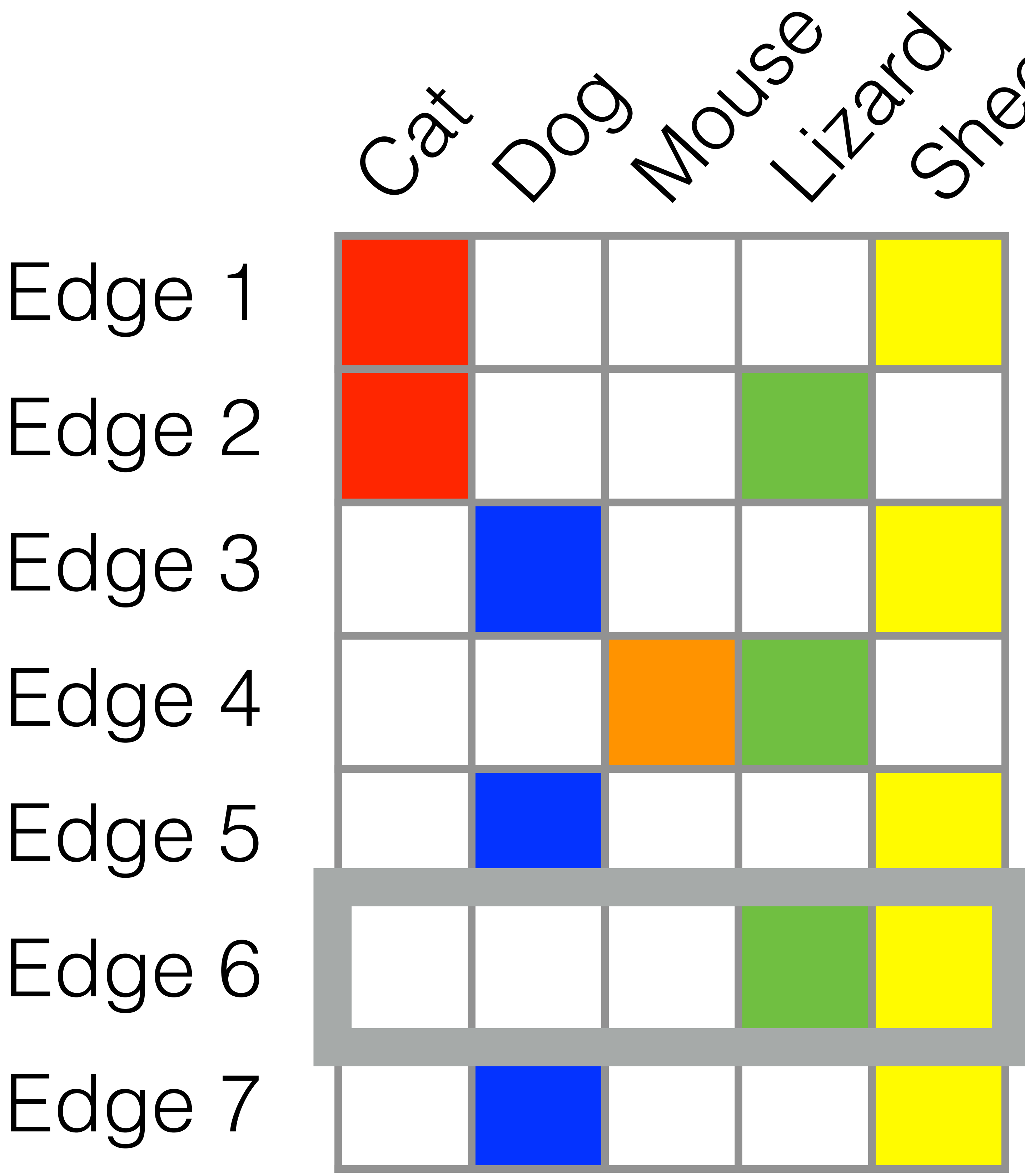
Graph



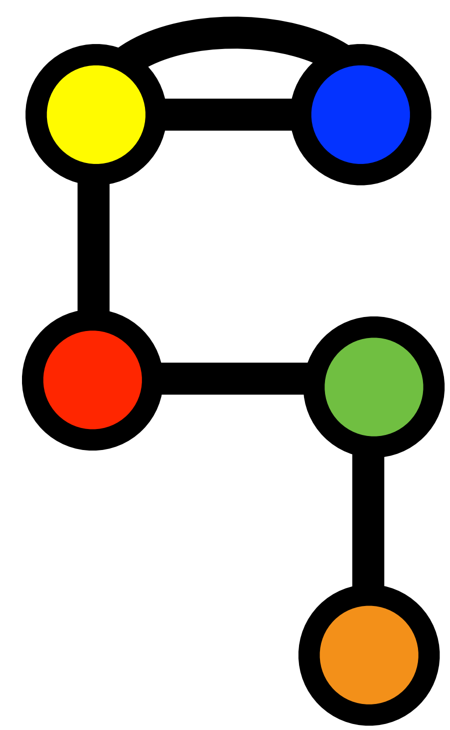
- Groups: vertices
- Edge-exchangeable



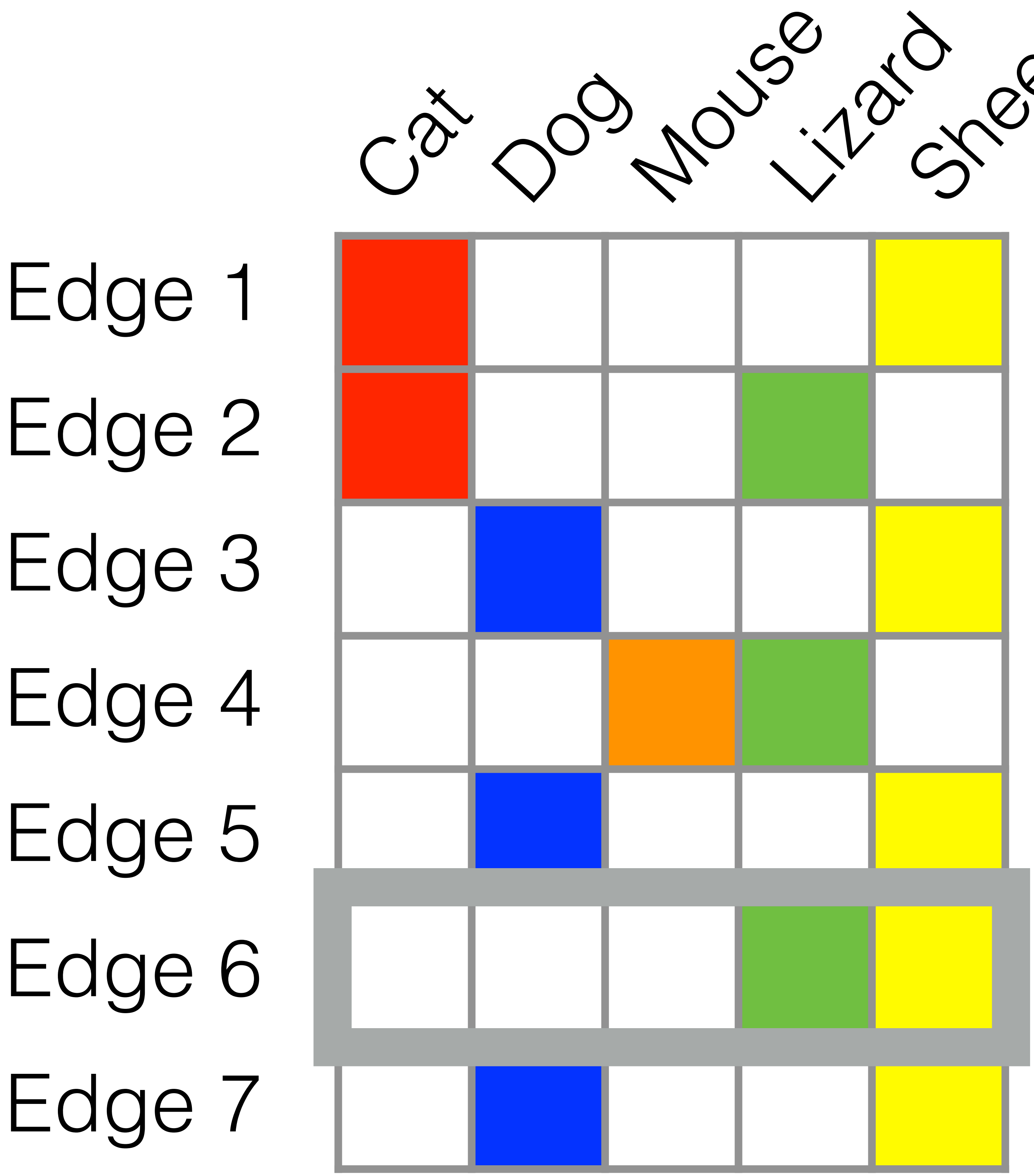
Graph



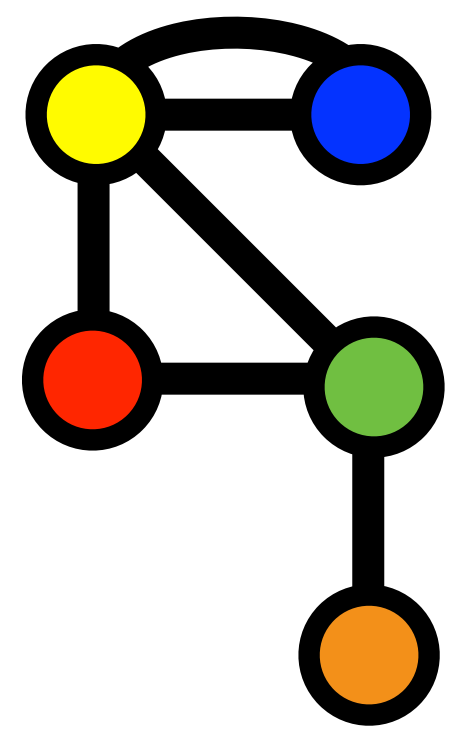
- Groups: vertices
- Edge-exchangeable



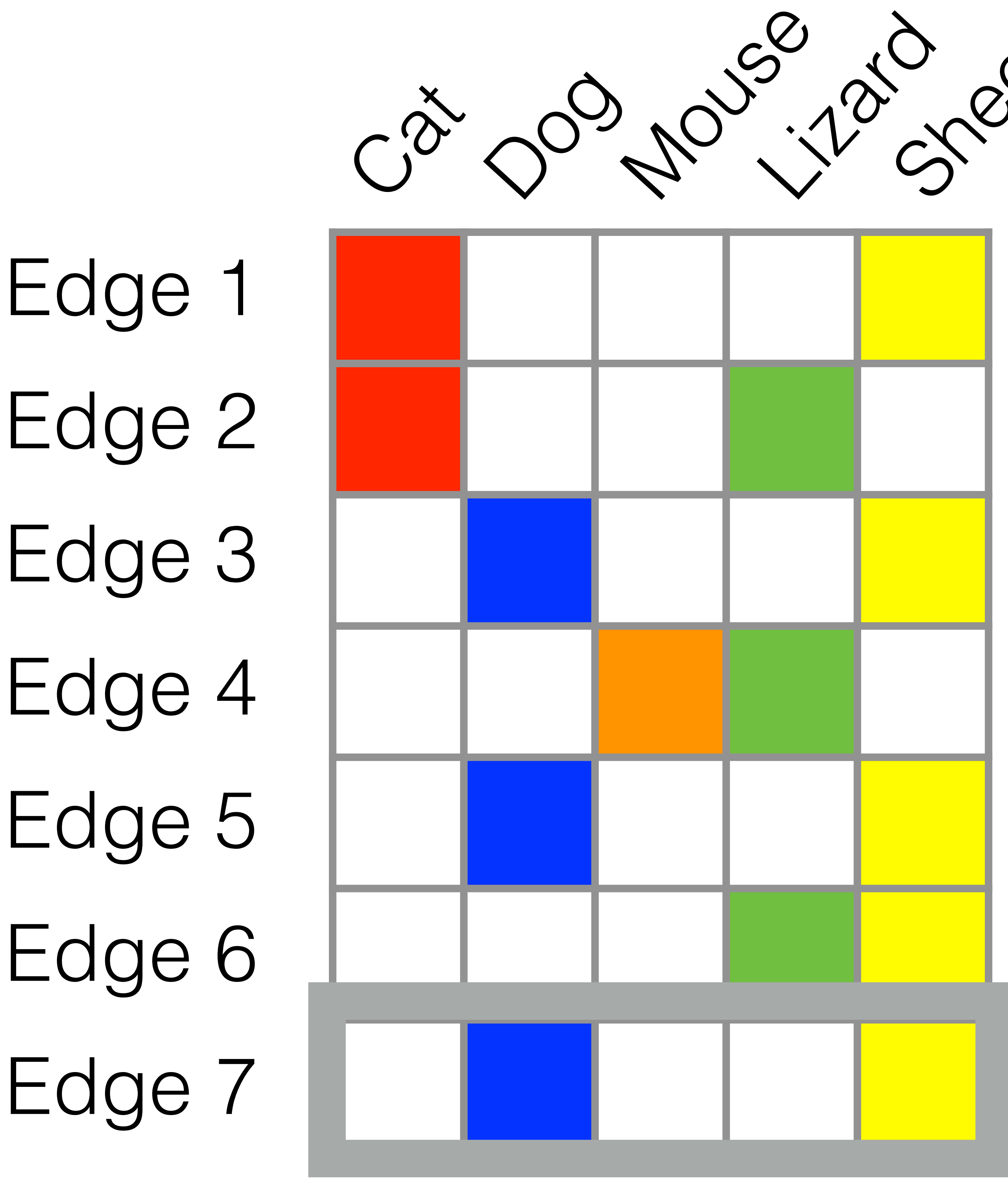
Graph



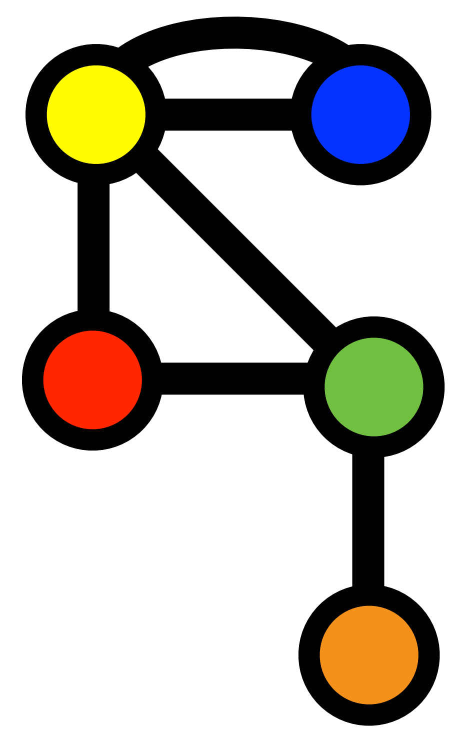
- Groups: vertices
- Edge-exchangeable



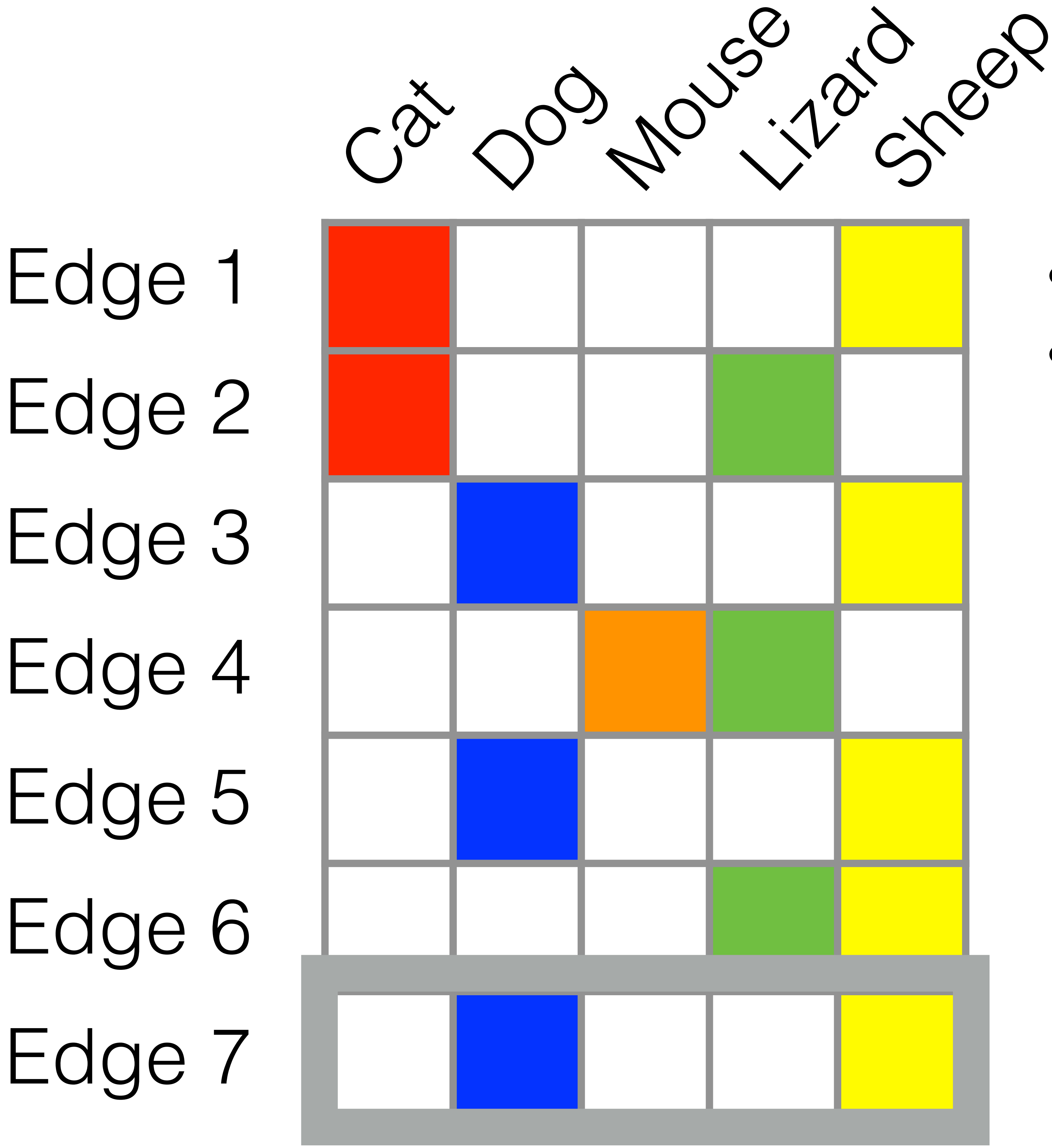
Graph



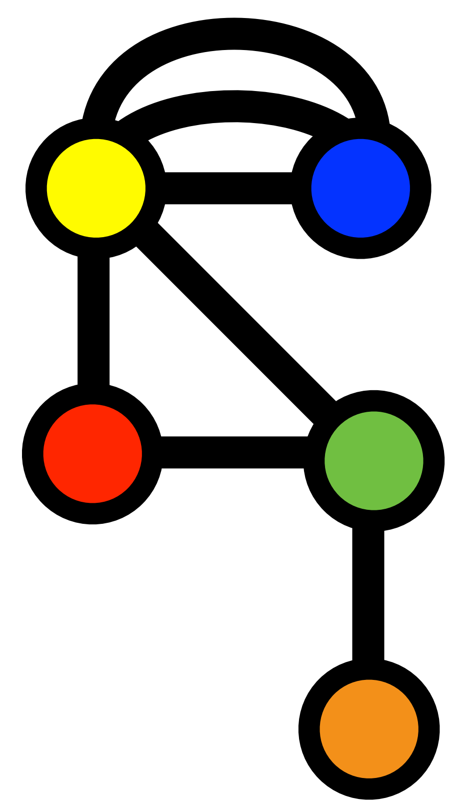
- Groups: vertices
- Edge-exchangeable



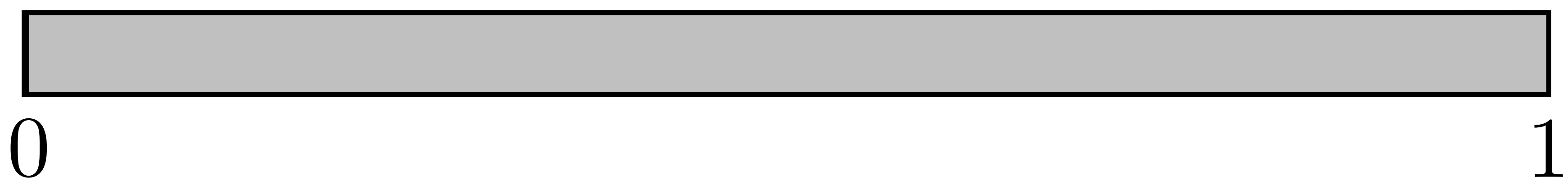
Graph



- Groups: vertices
- Edge-exchangeable

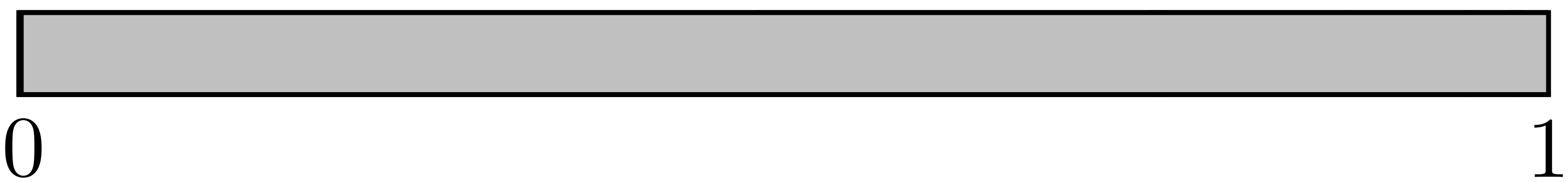


De Finetti Theorems



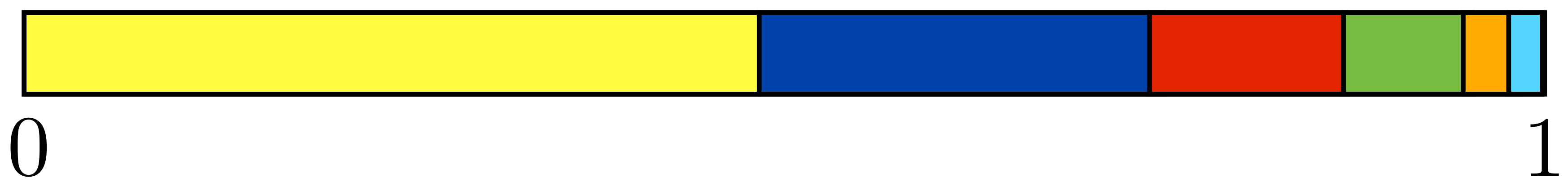
[Kingman 1978]

Clustering



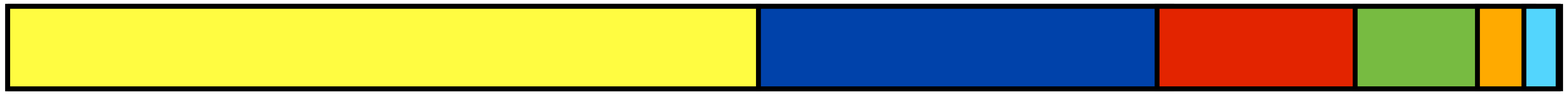
[Kingman 1978]

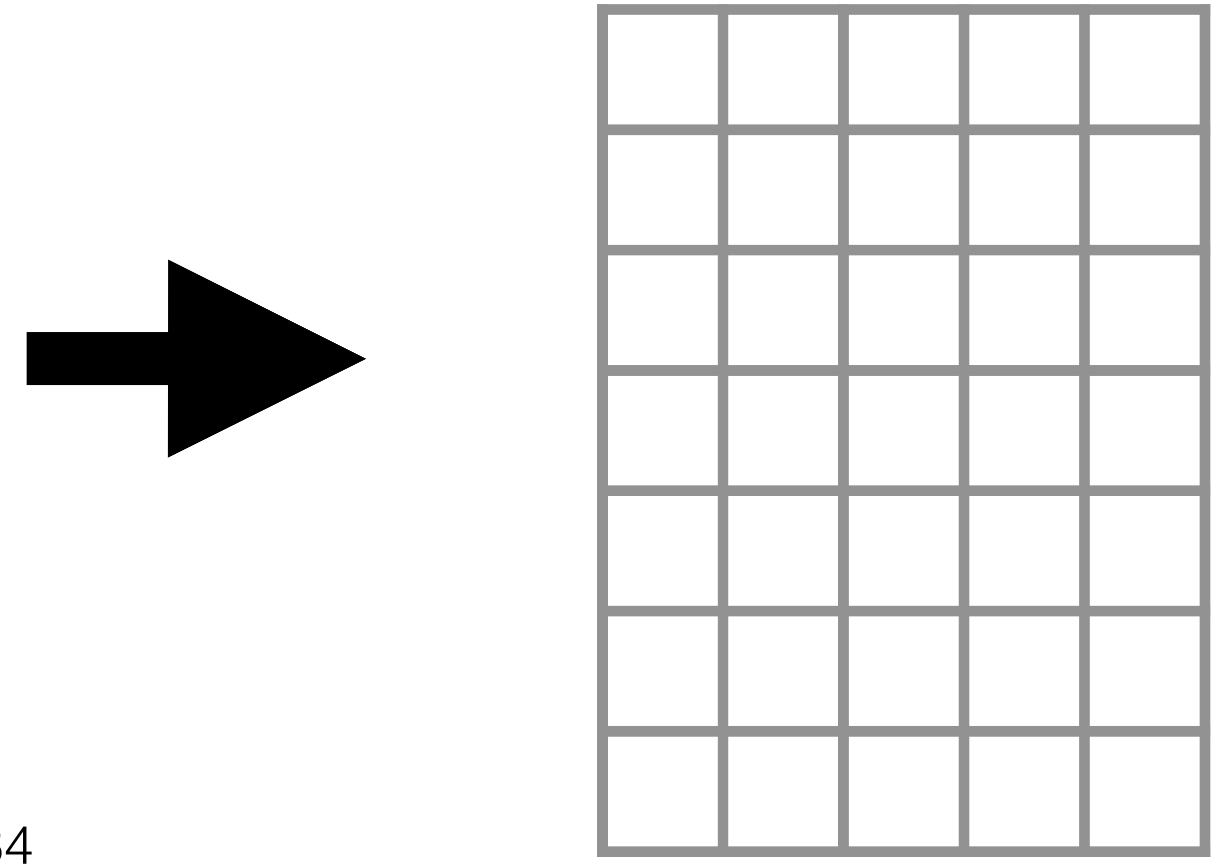
Clustering



[Kingman 1978]

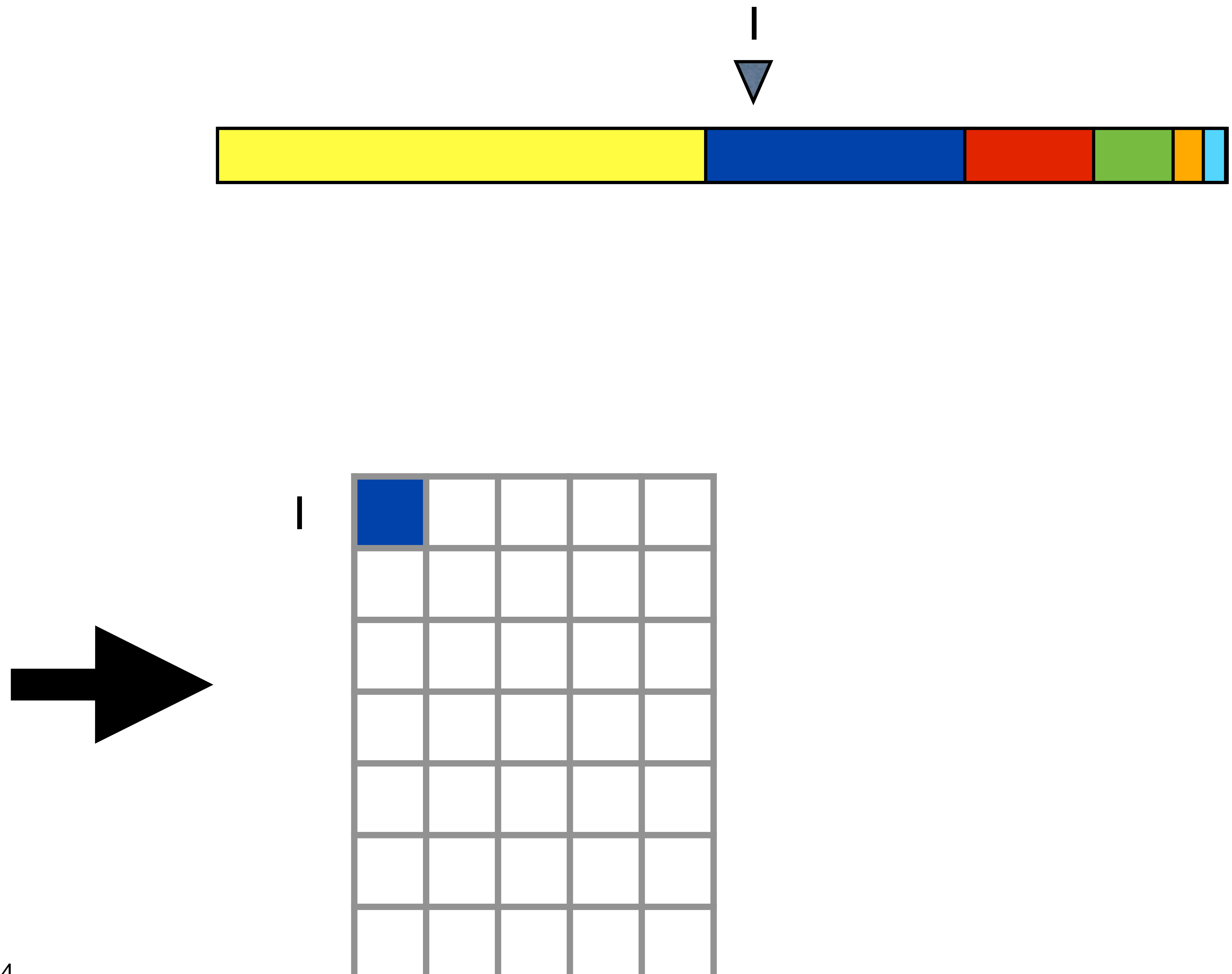
Clustering





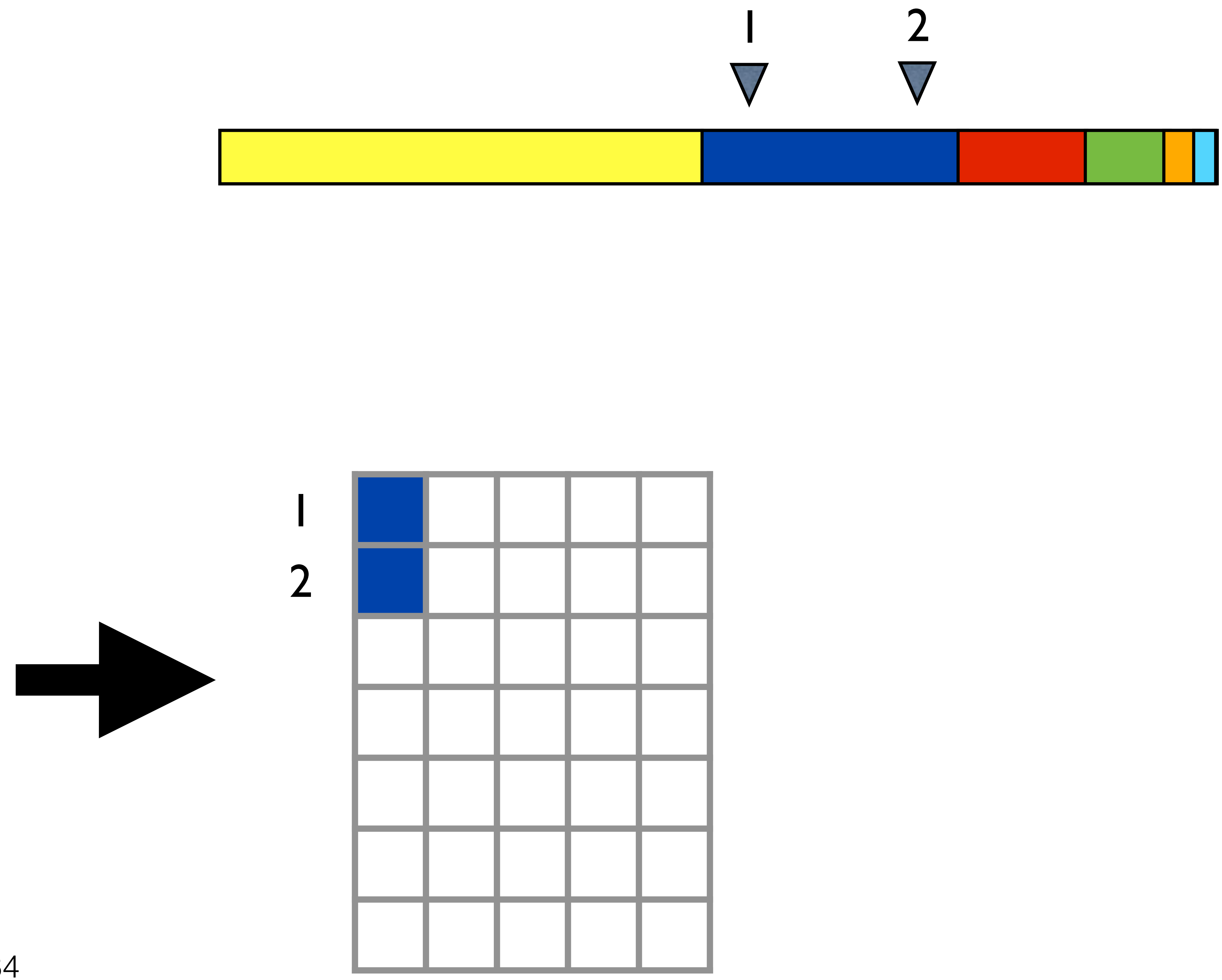
[Kingman 1978]

Clustering



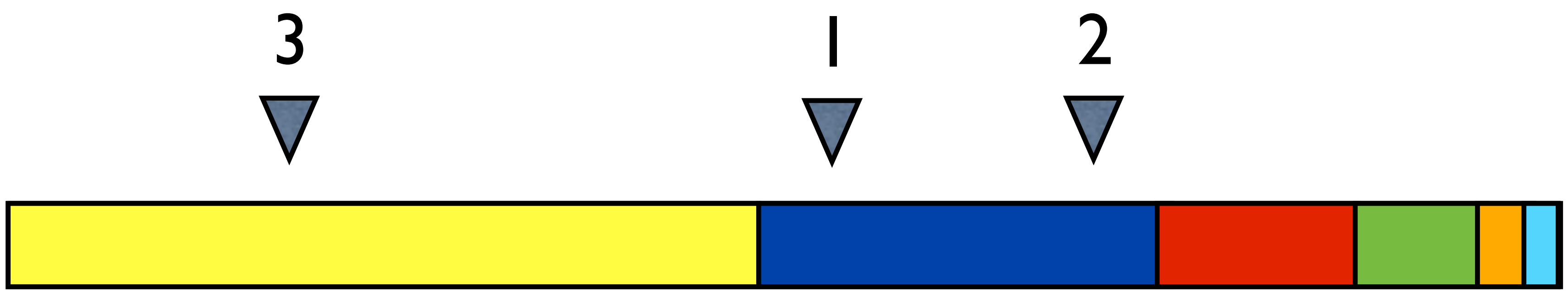
[Kingman 1978]

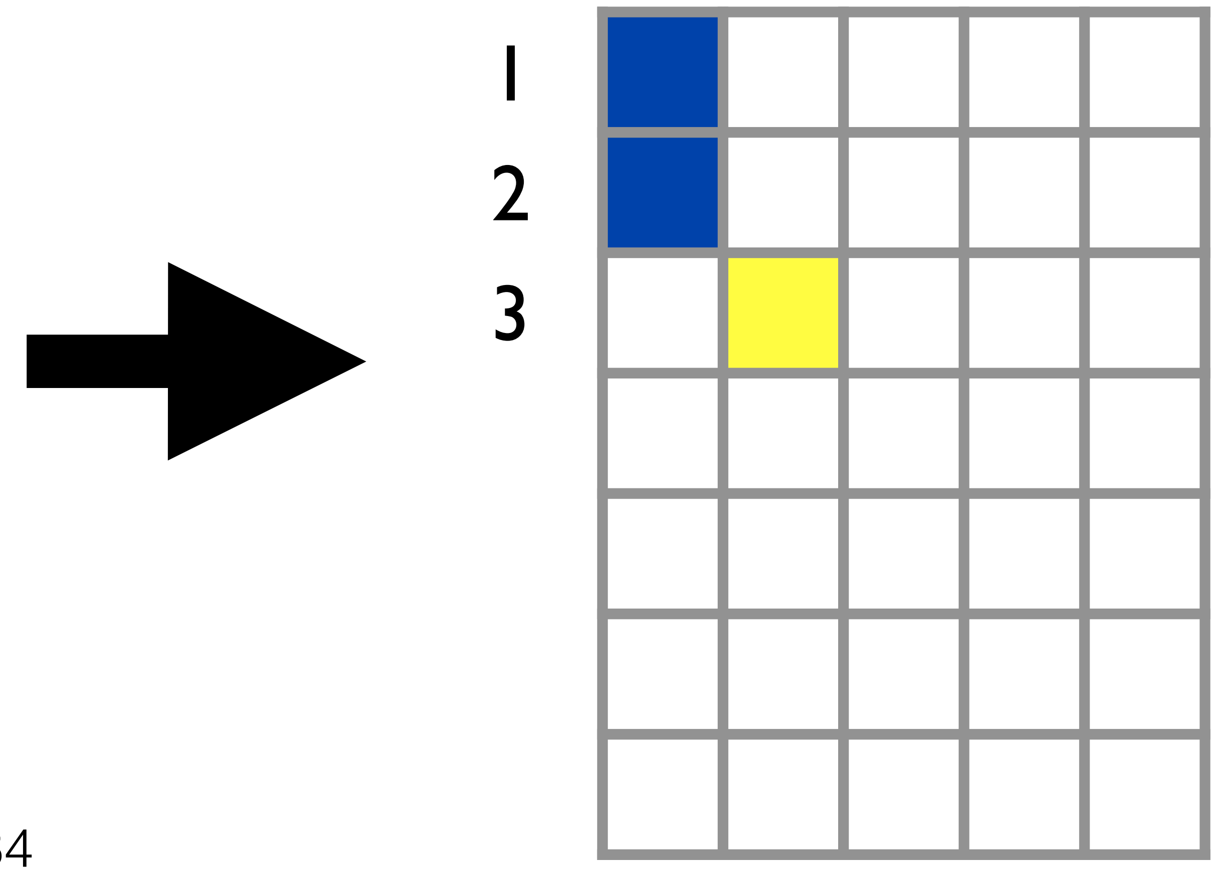
Clustering



[Kingman 1978]

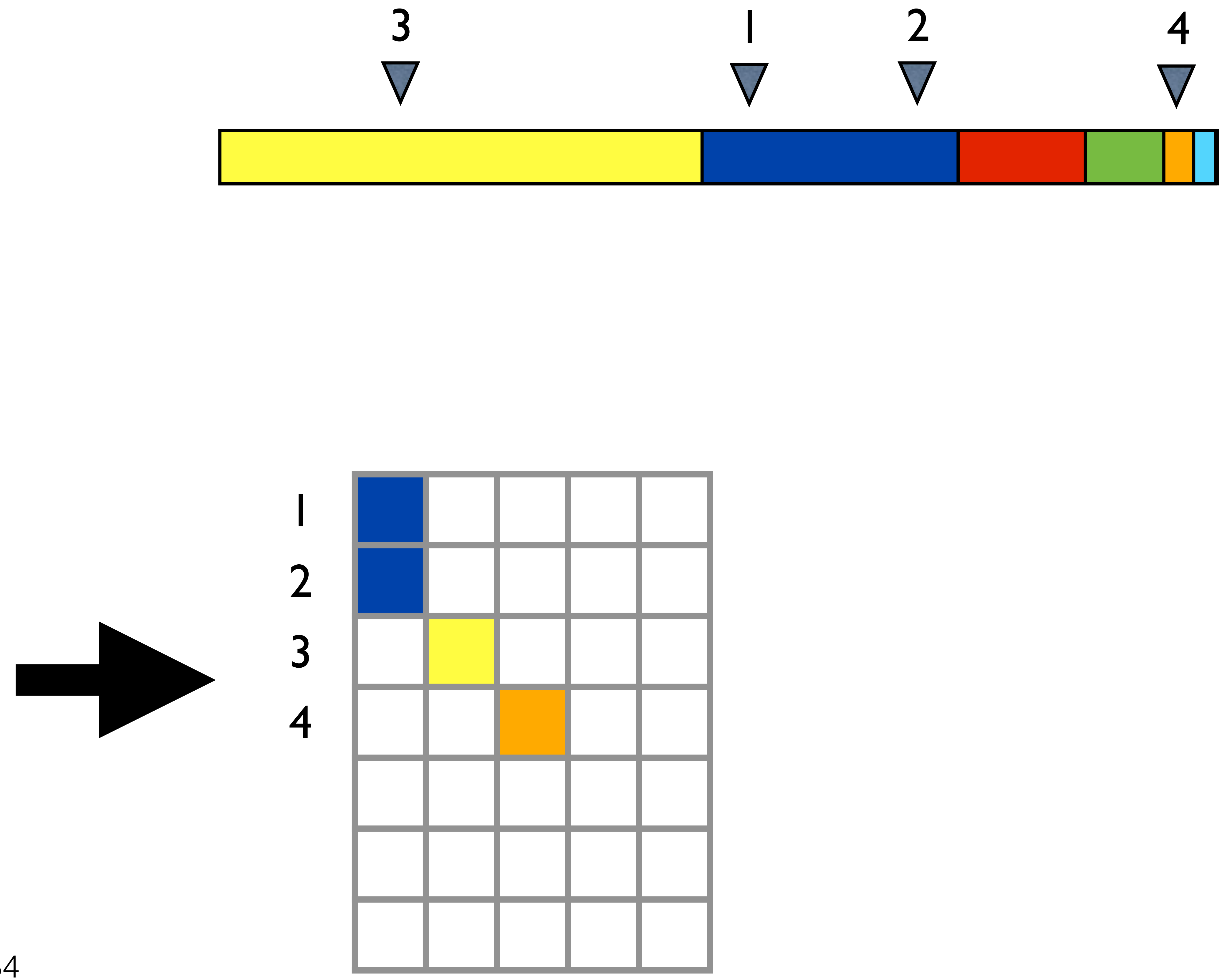
Clustering





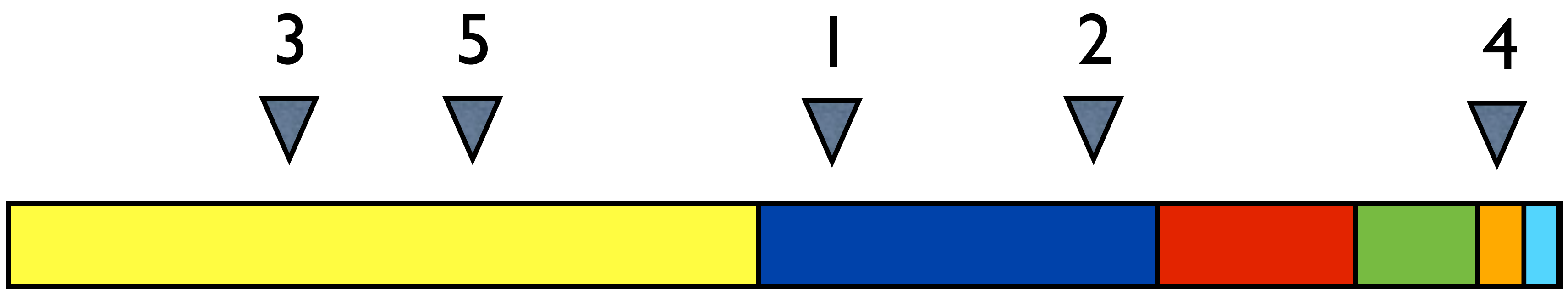
[Kingman 1978]

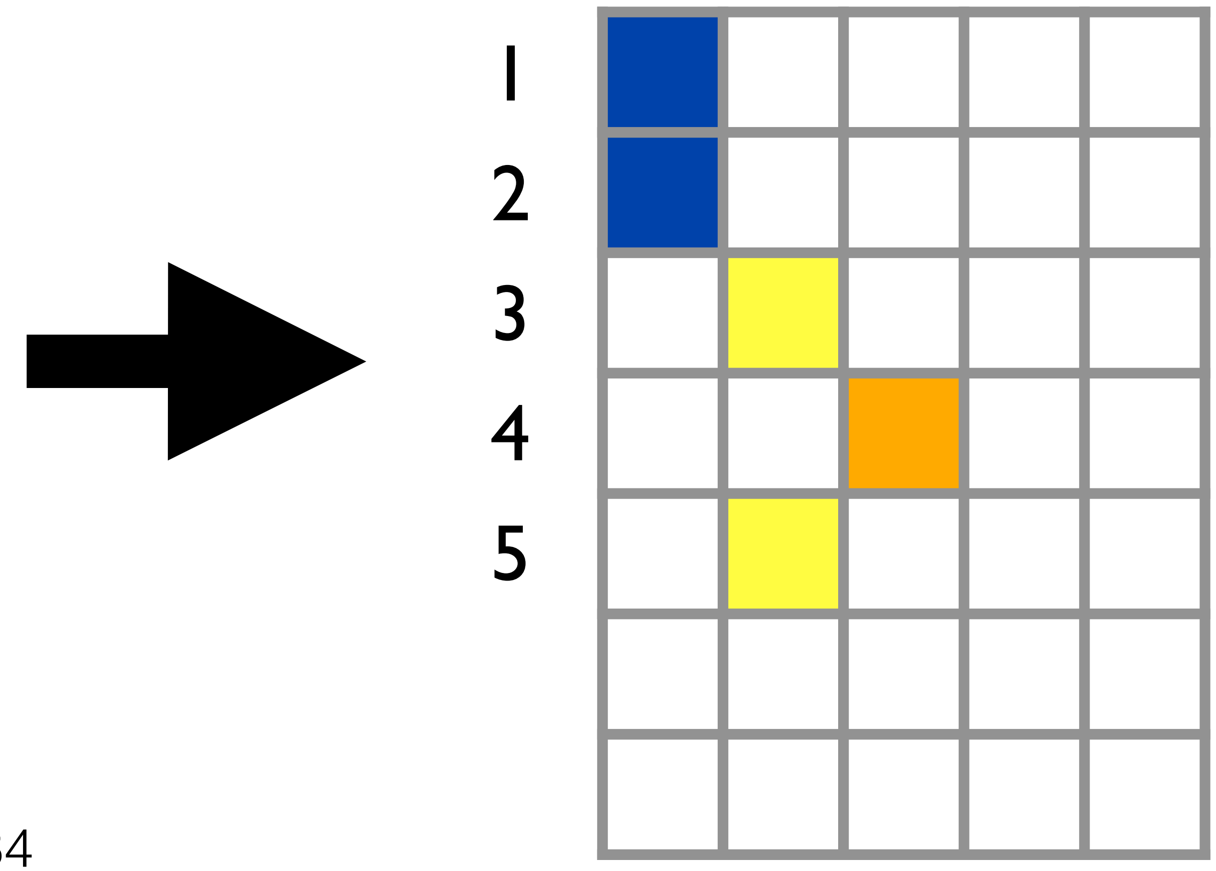
Clustering



[Kingman 1978]

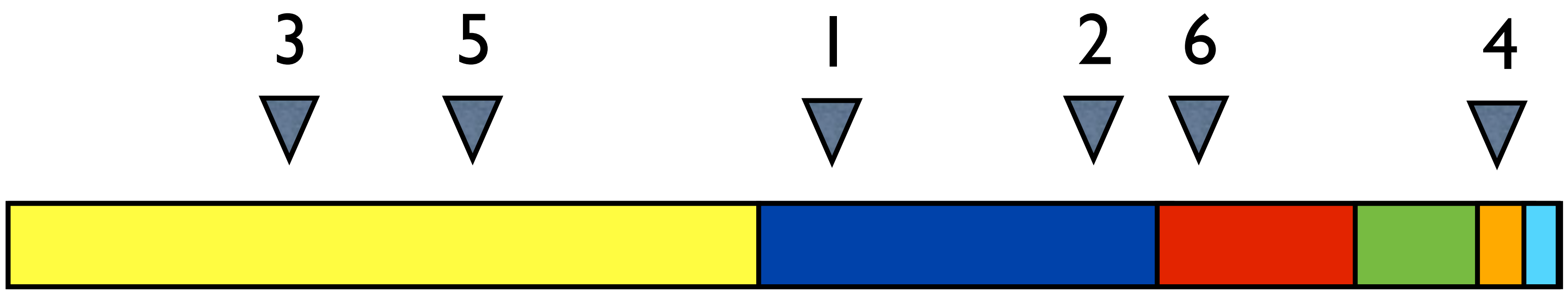
Clustering

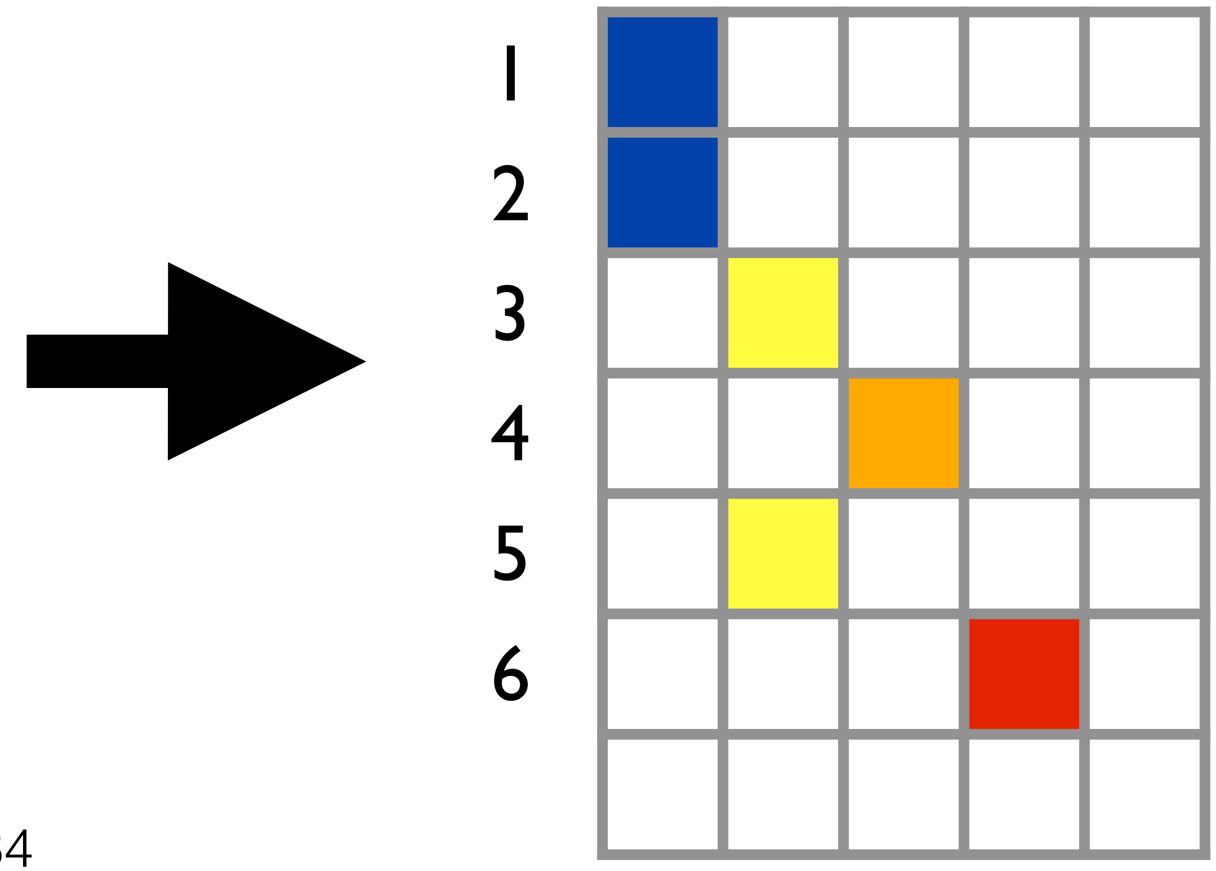




[Kingman 1978]

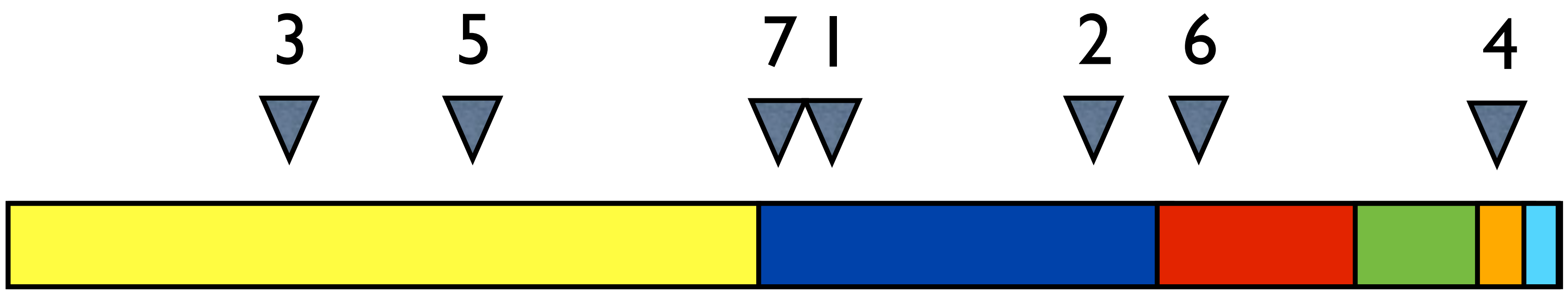
Clustering

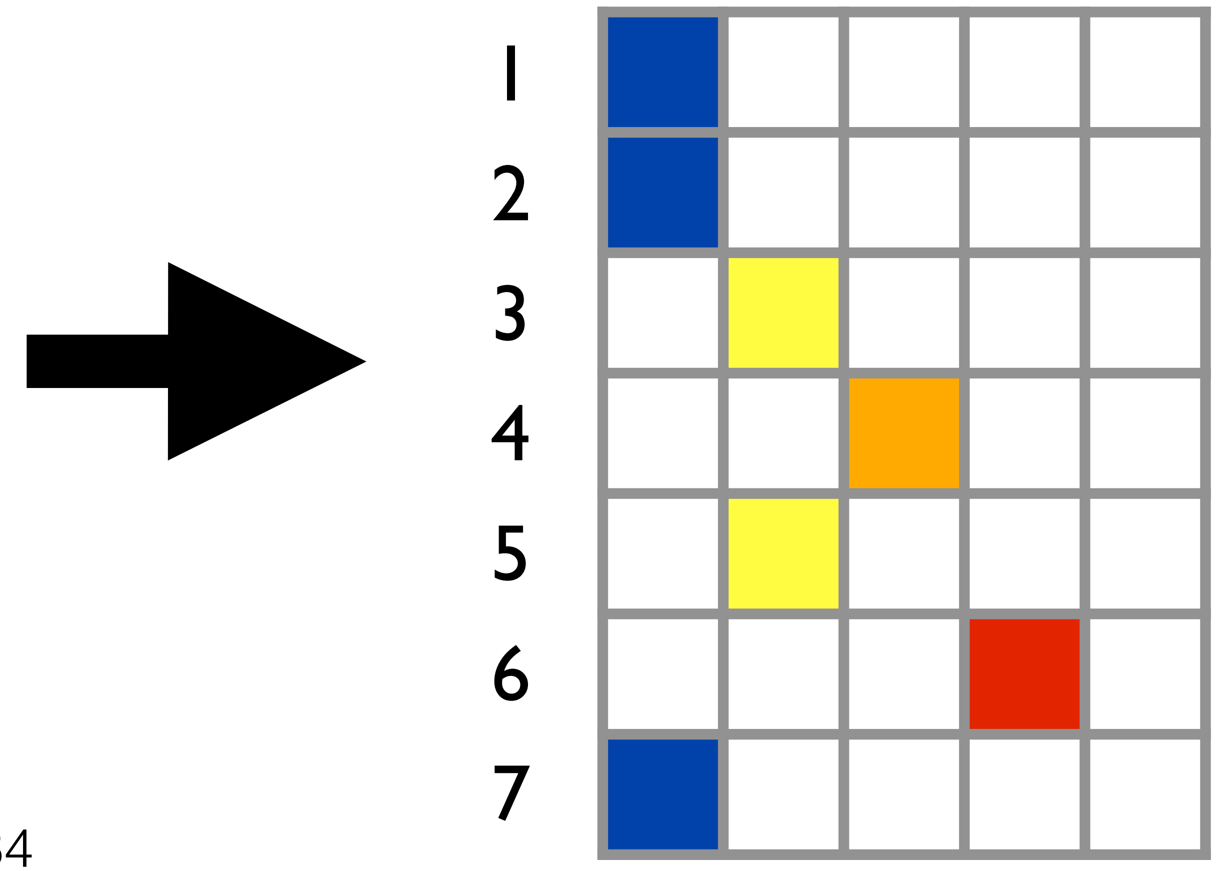




[Kingman 1978]

Clustering

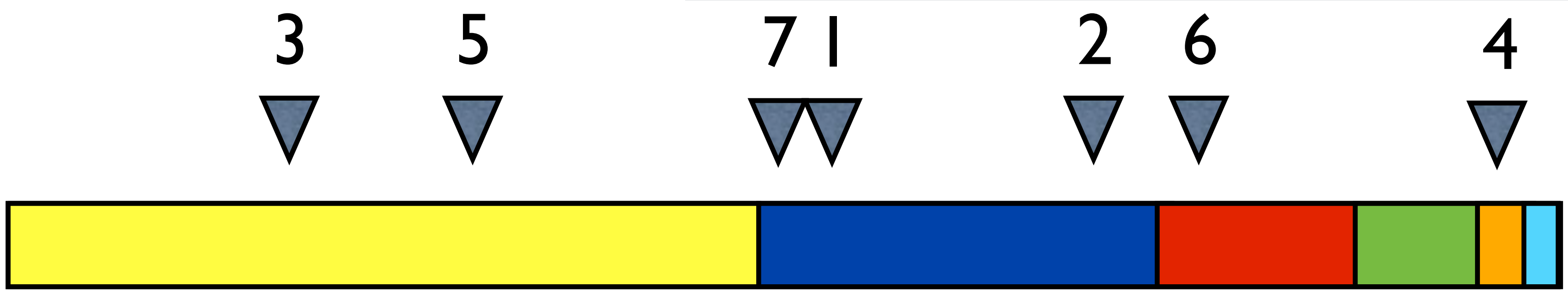


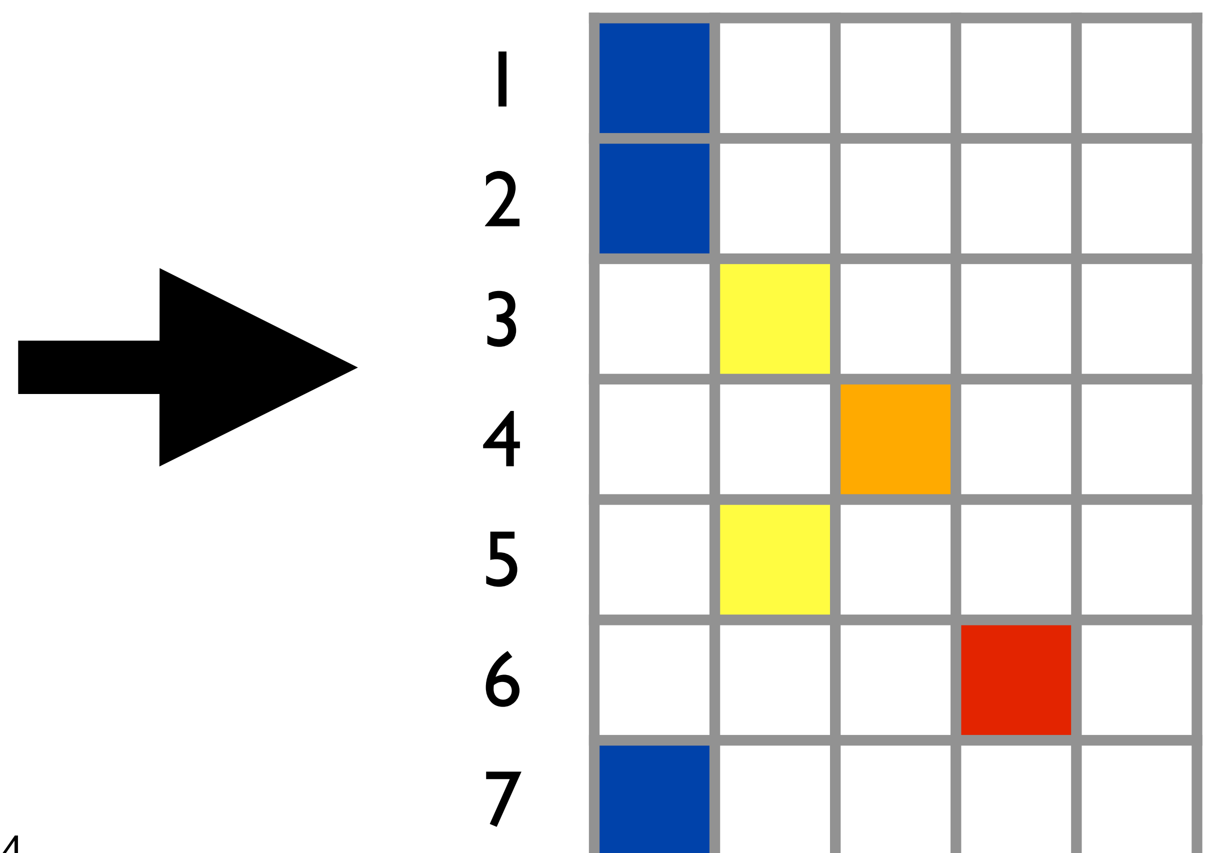


[Kingman 1978]

Clustering

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

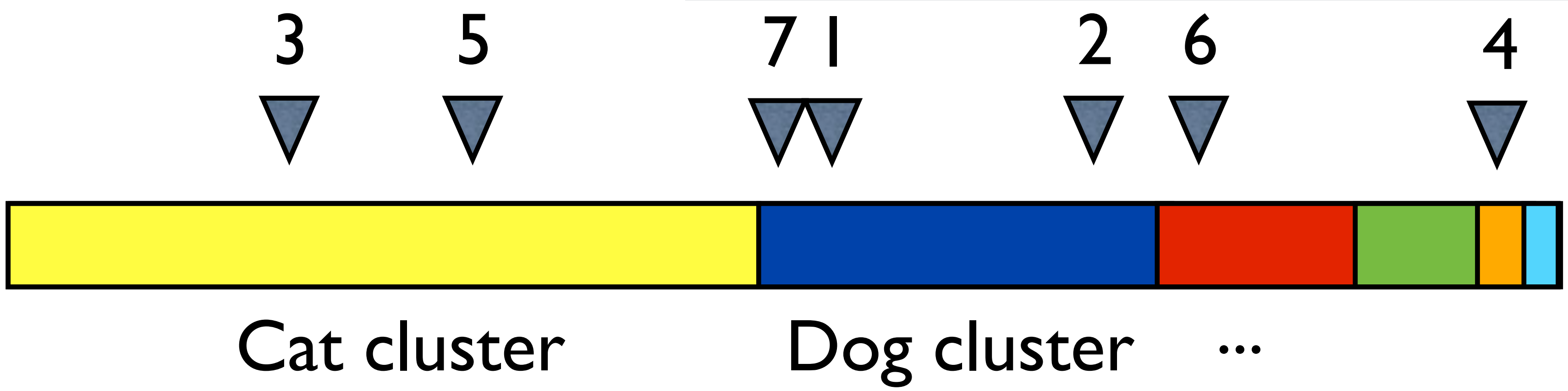


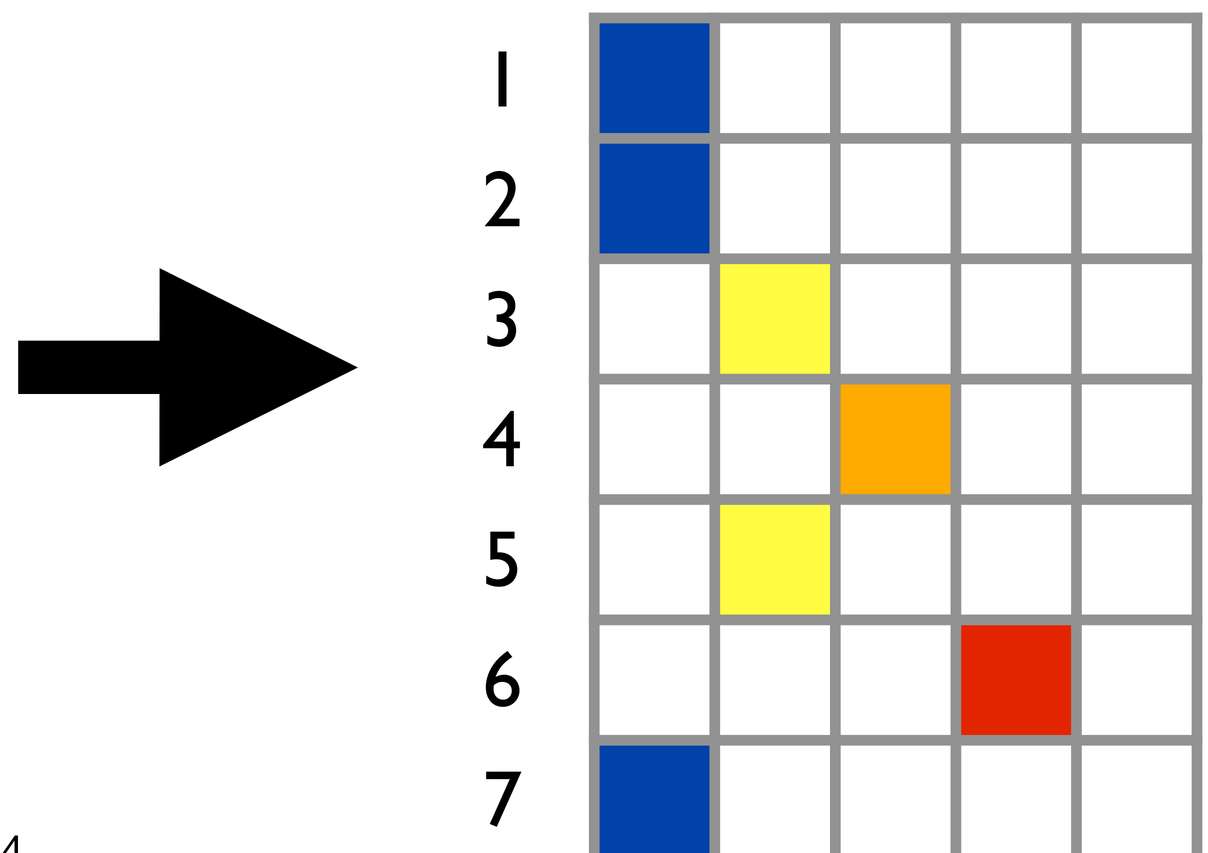


[Kingman 1978]

Clustering

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

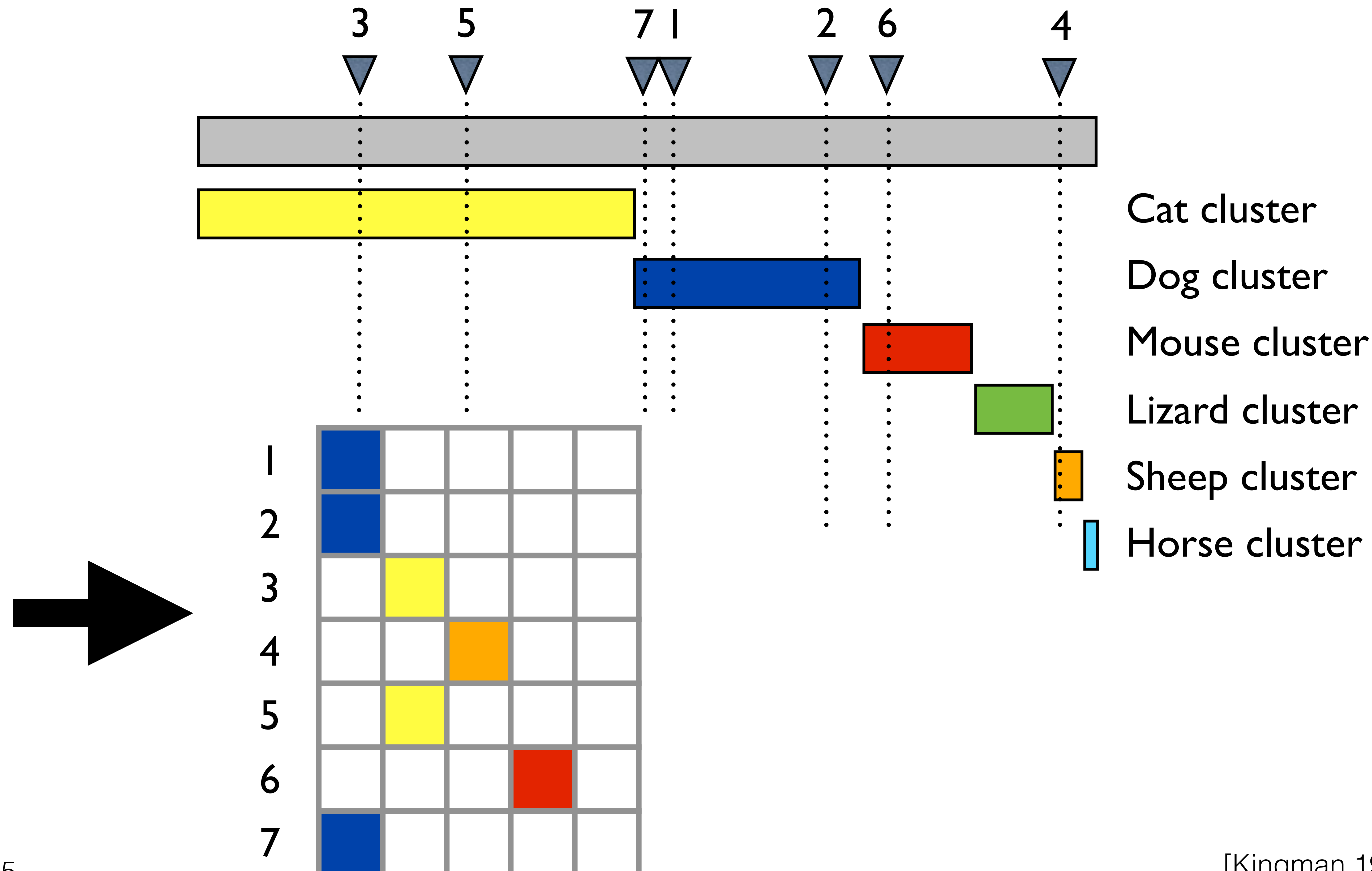




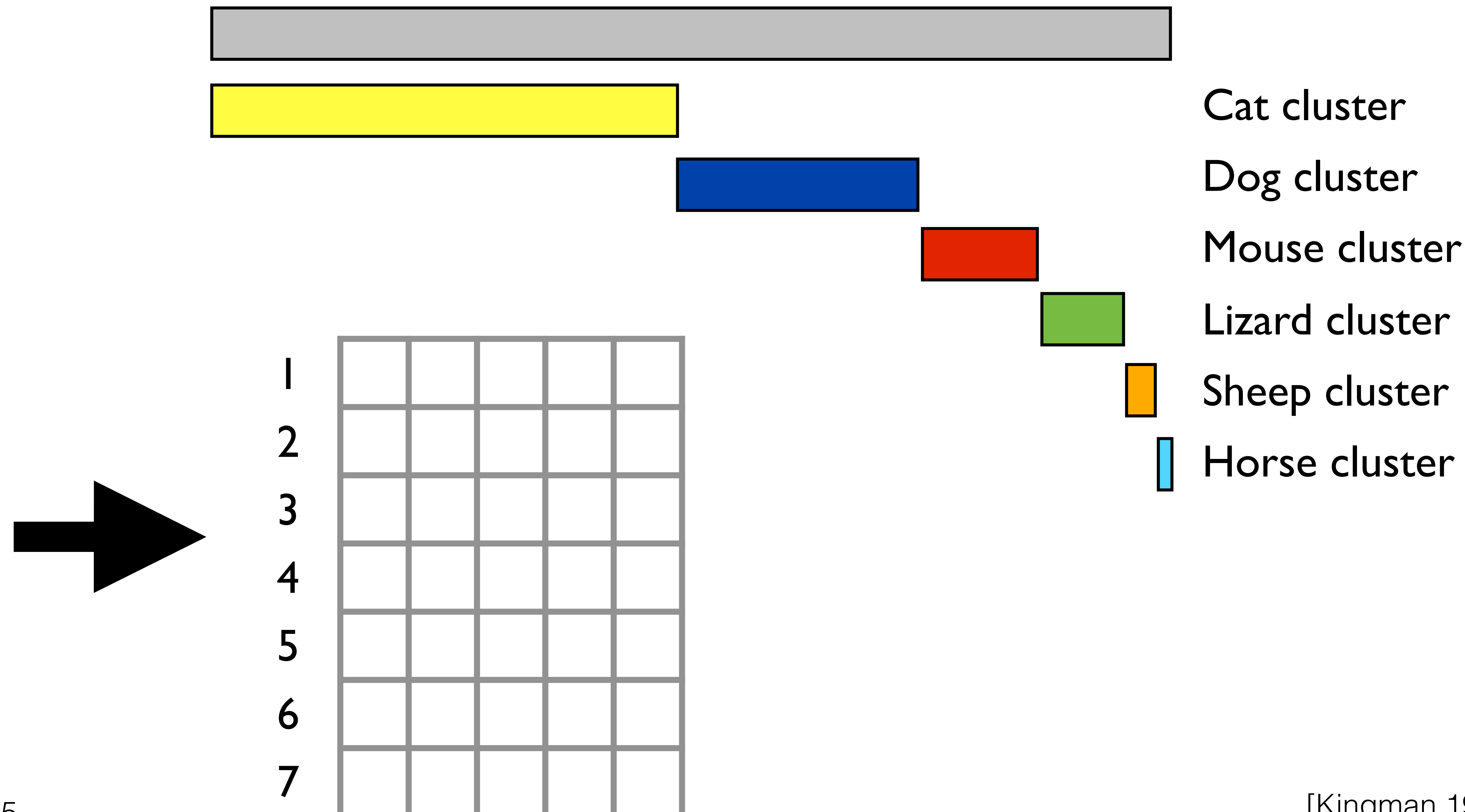
[Kingman 1978]

Clustering

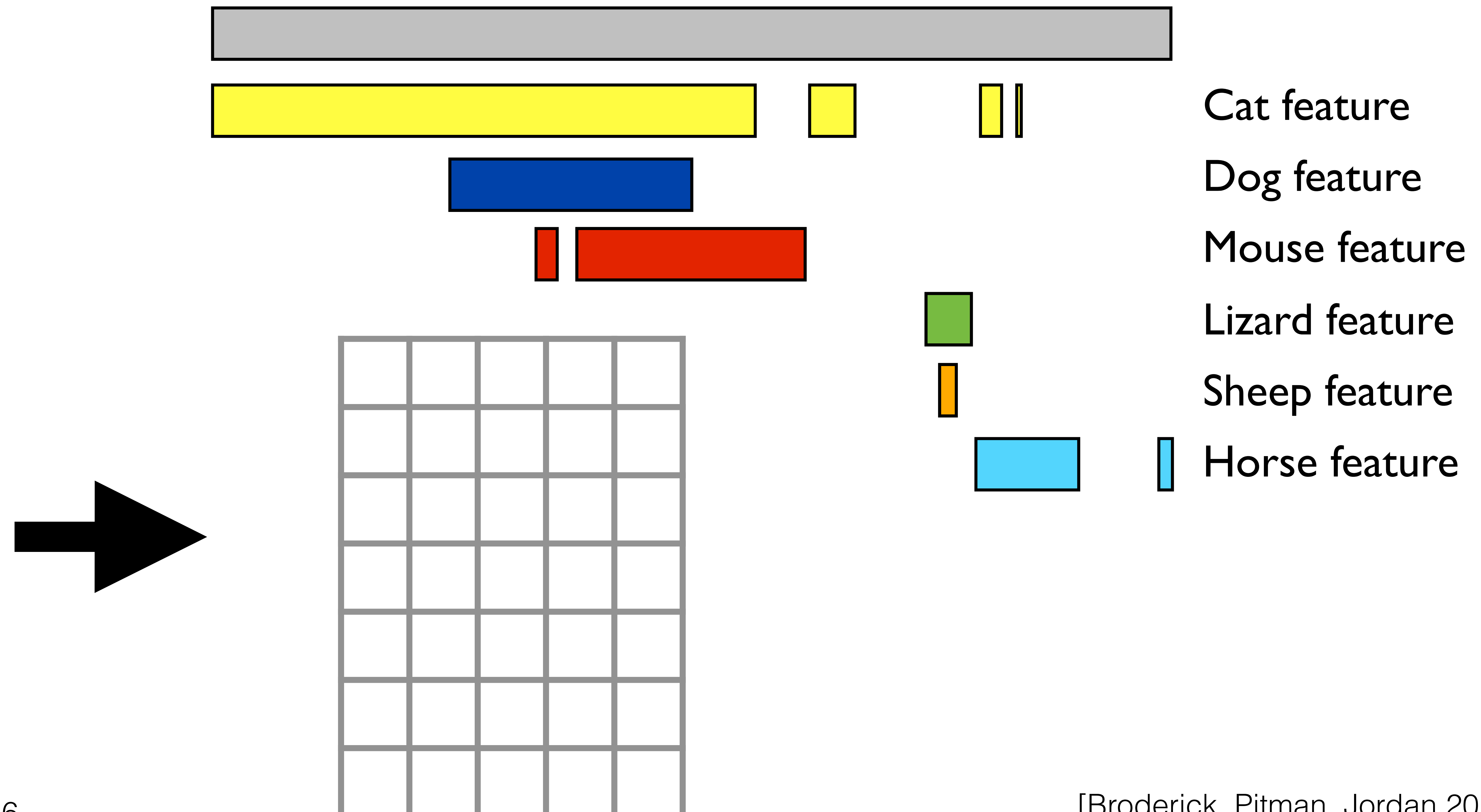
Thm (Kingman). A clustering is exchangeable iff it has a Kingman paintbox representation



[Kingman 1978]

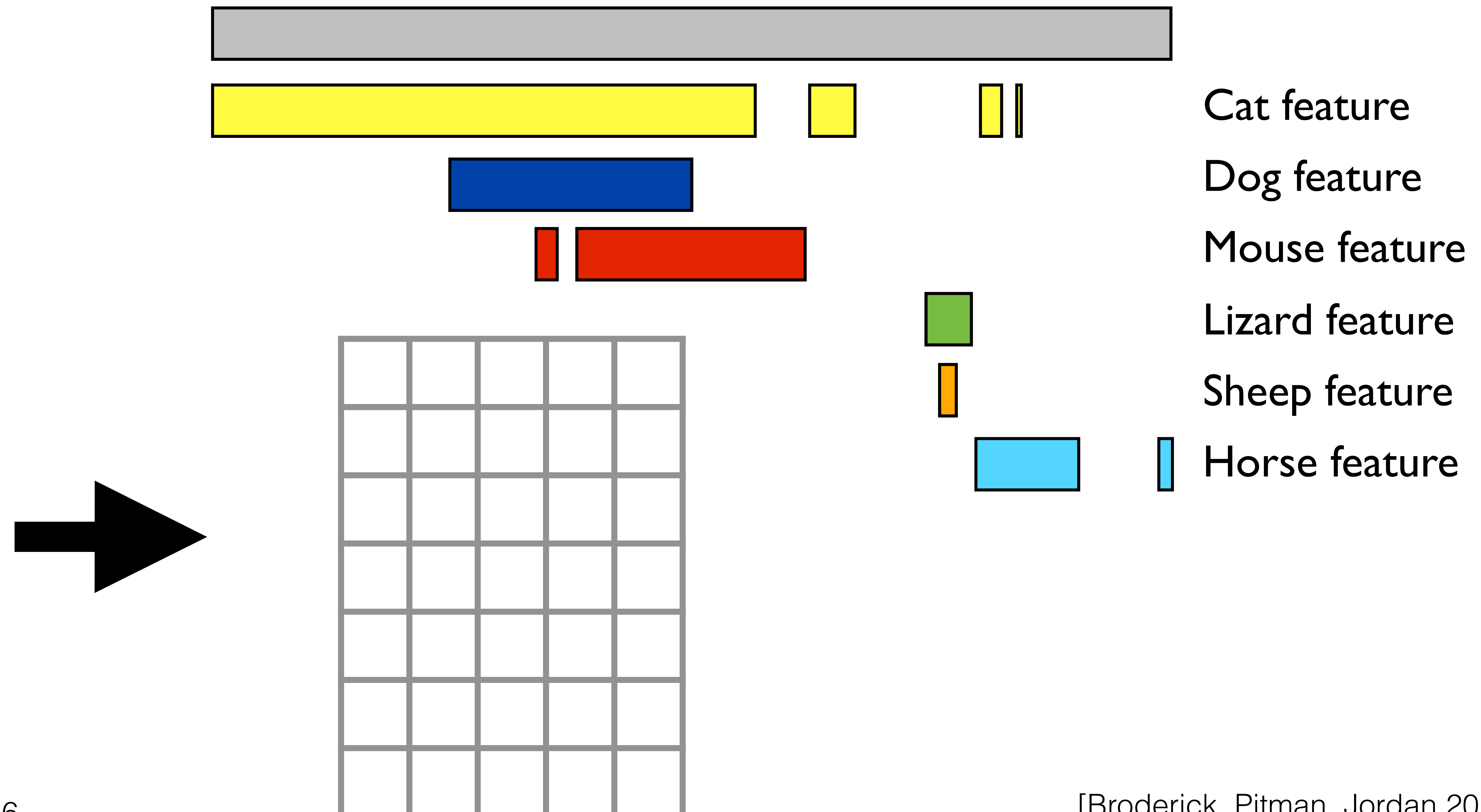


[Kingman 1978]



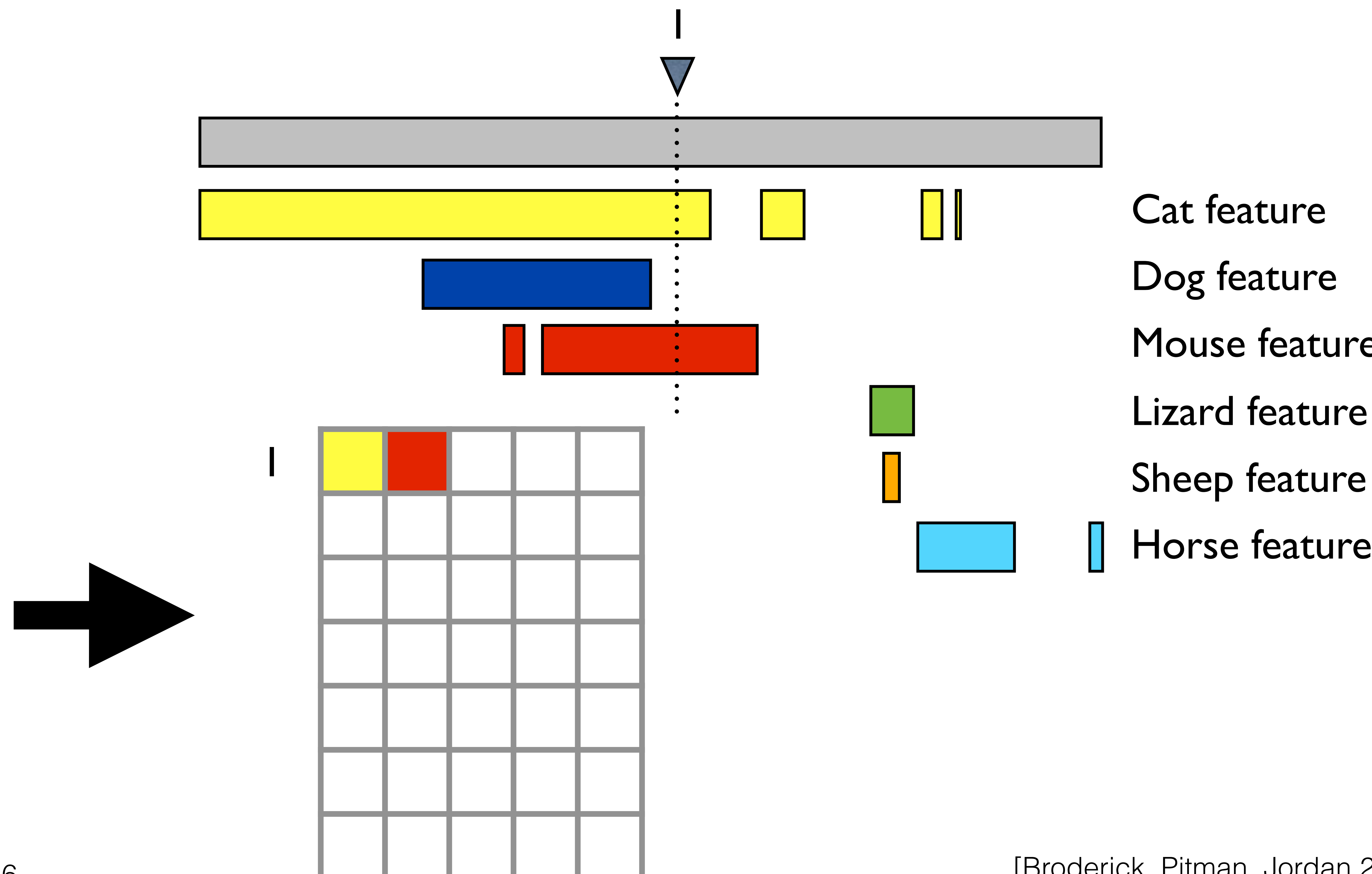
[Broderick, Pitman, Jordan 2013]

Feature allocation



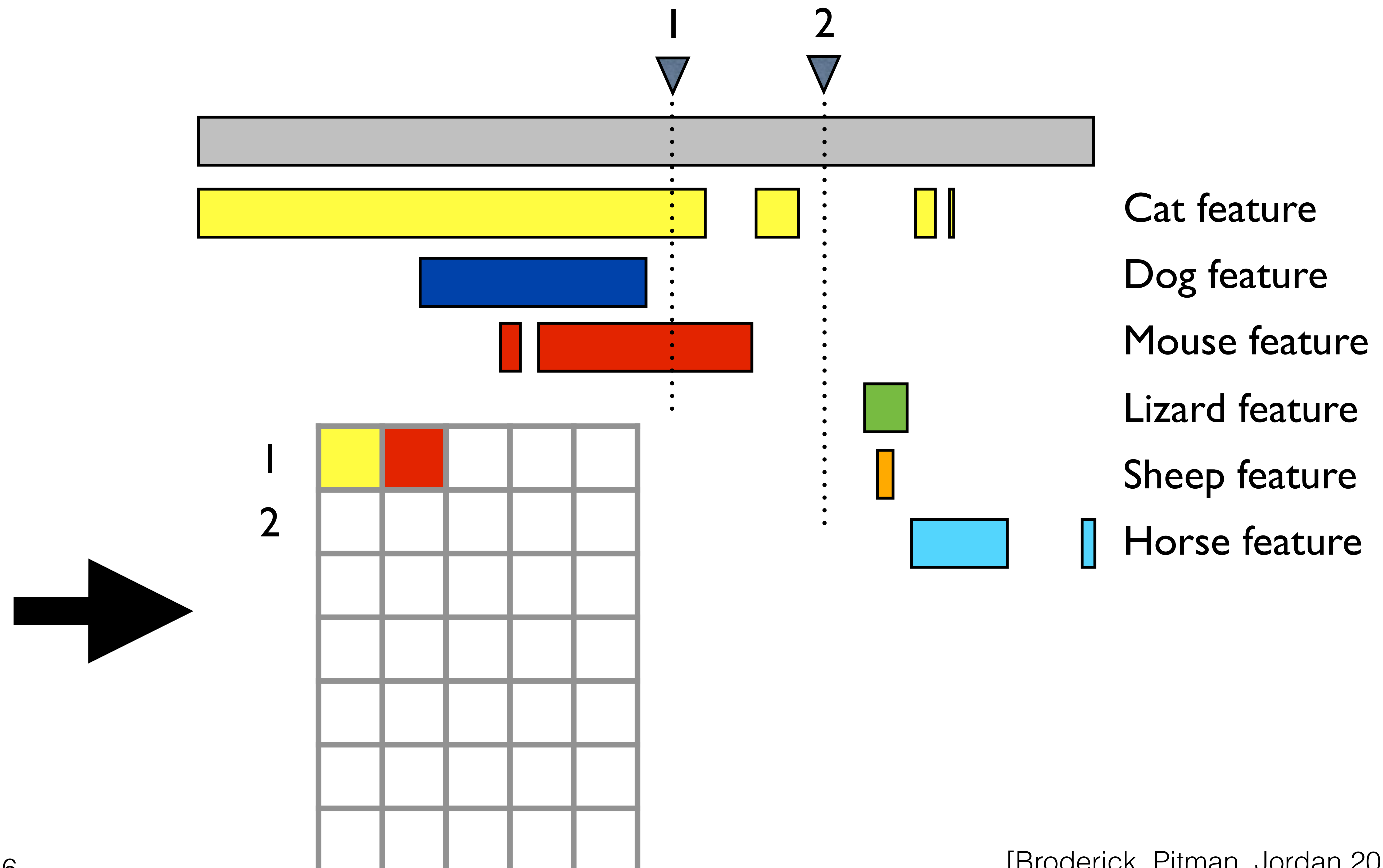
[Broderick, Pitman, Jordan 2013]

Feature allocation



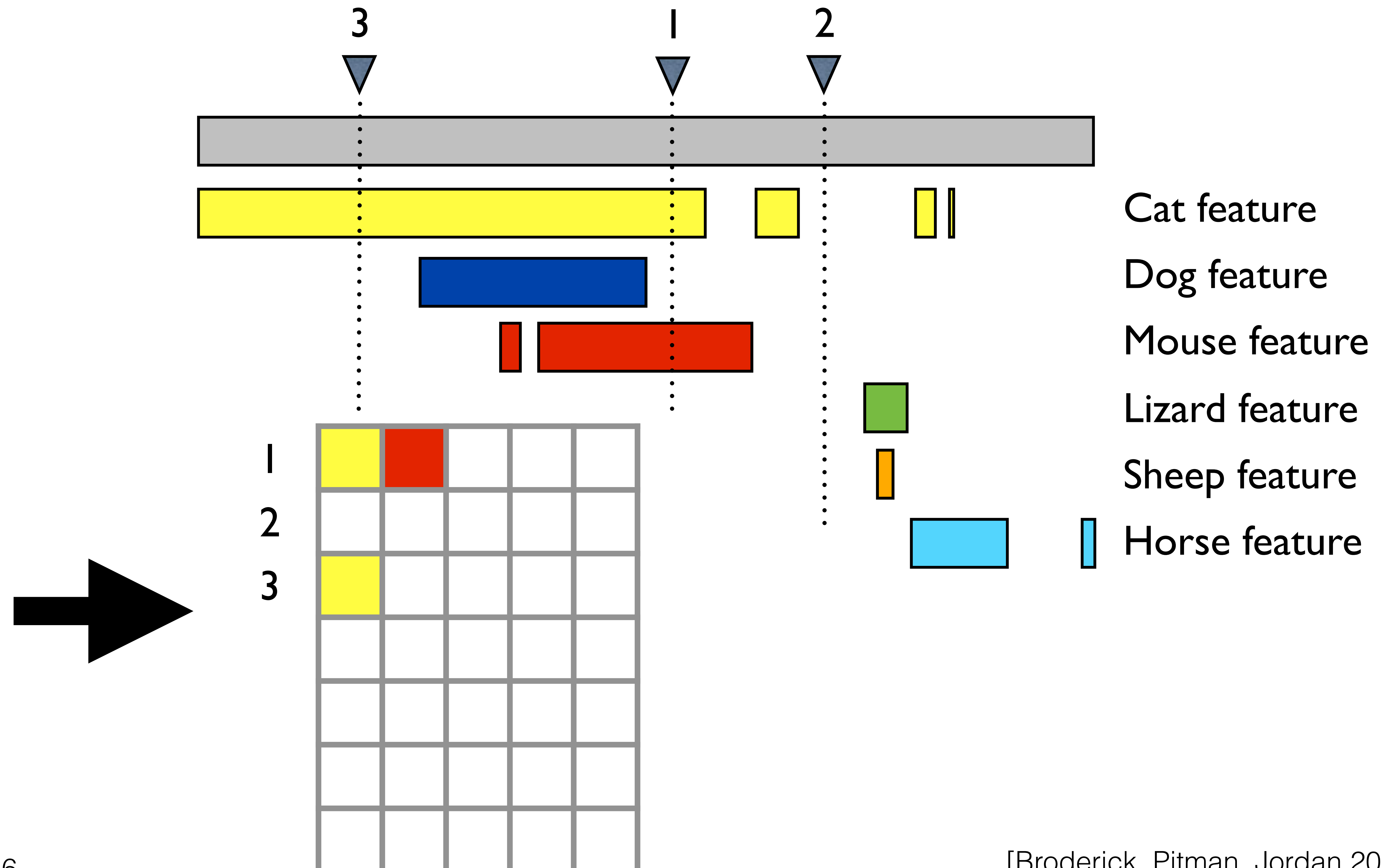
[Broderick, Pitman, Jordan 2013]

Feature allocation



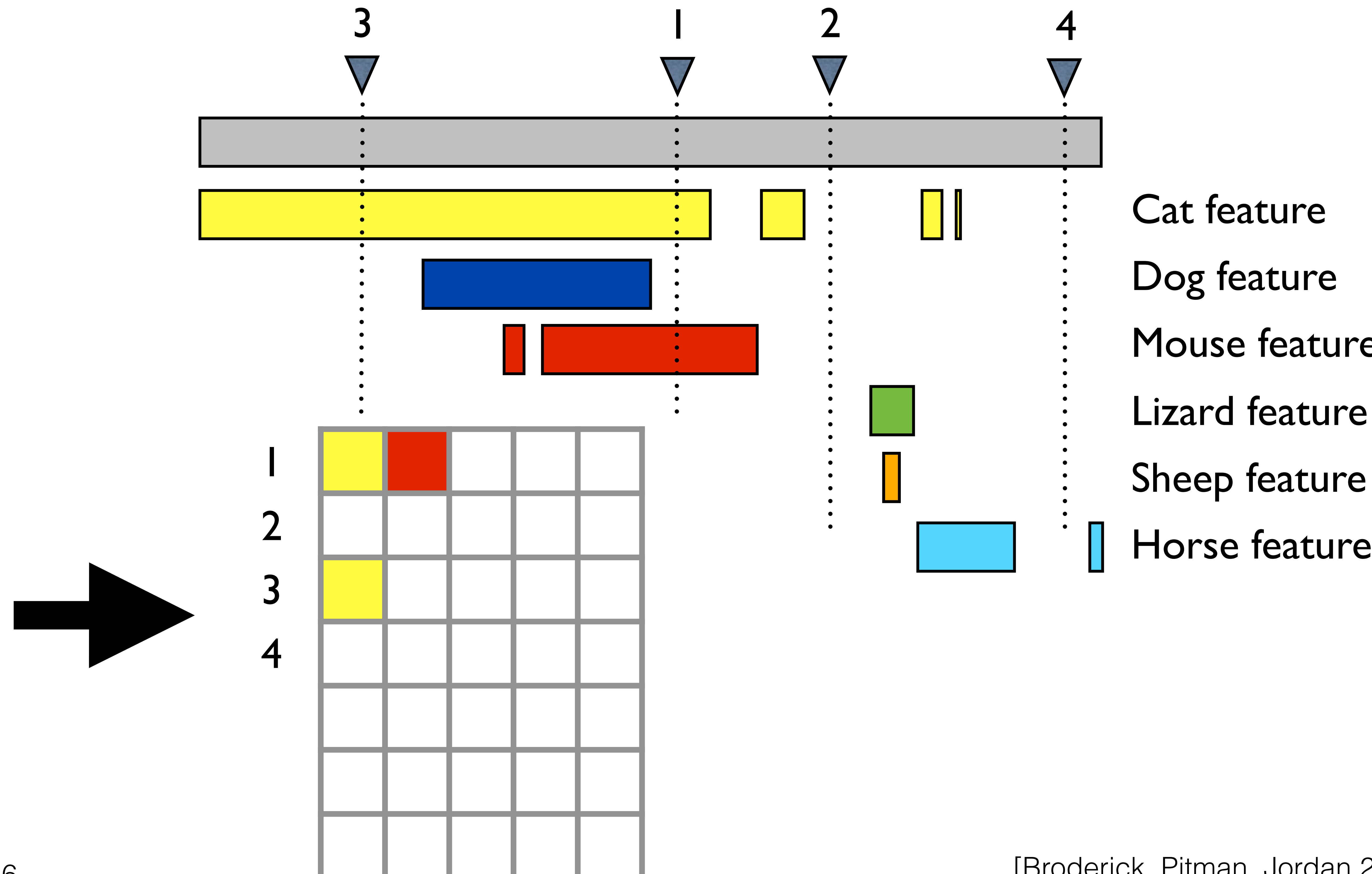
[Broderick, Pitman, Jordan 2013]

Feature allocation



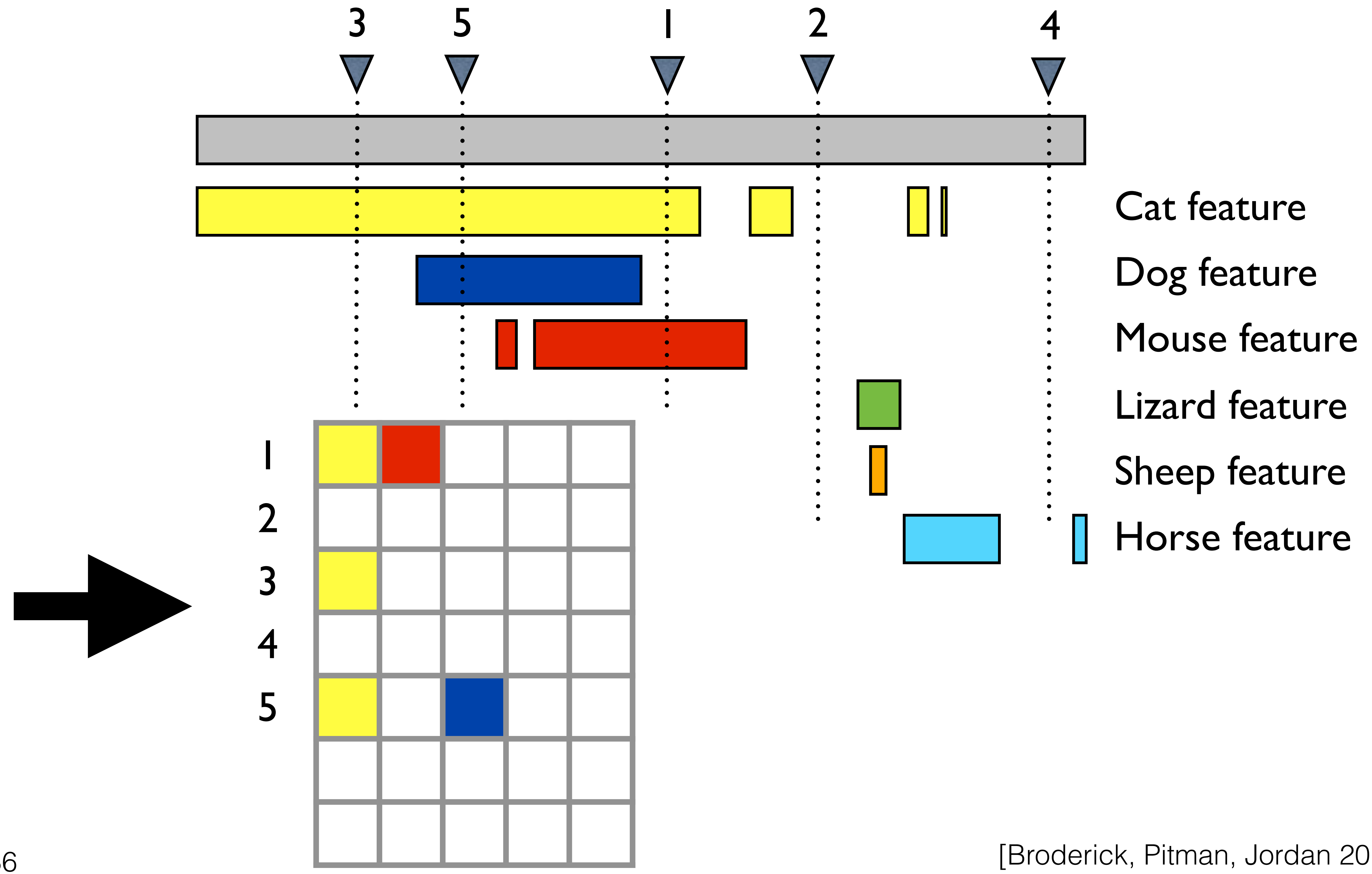
[Broderick, Pitman, Jordan 2013]

Feature allocation



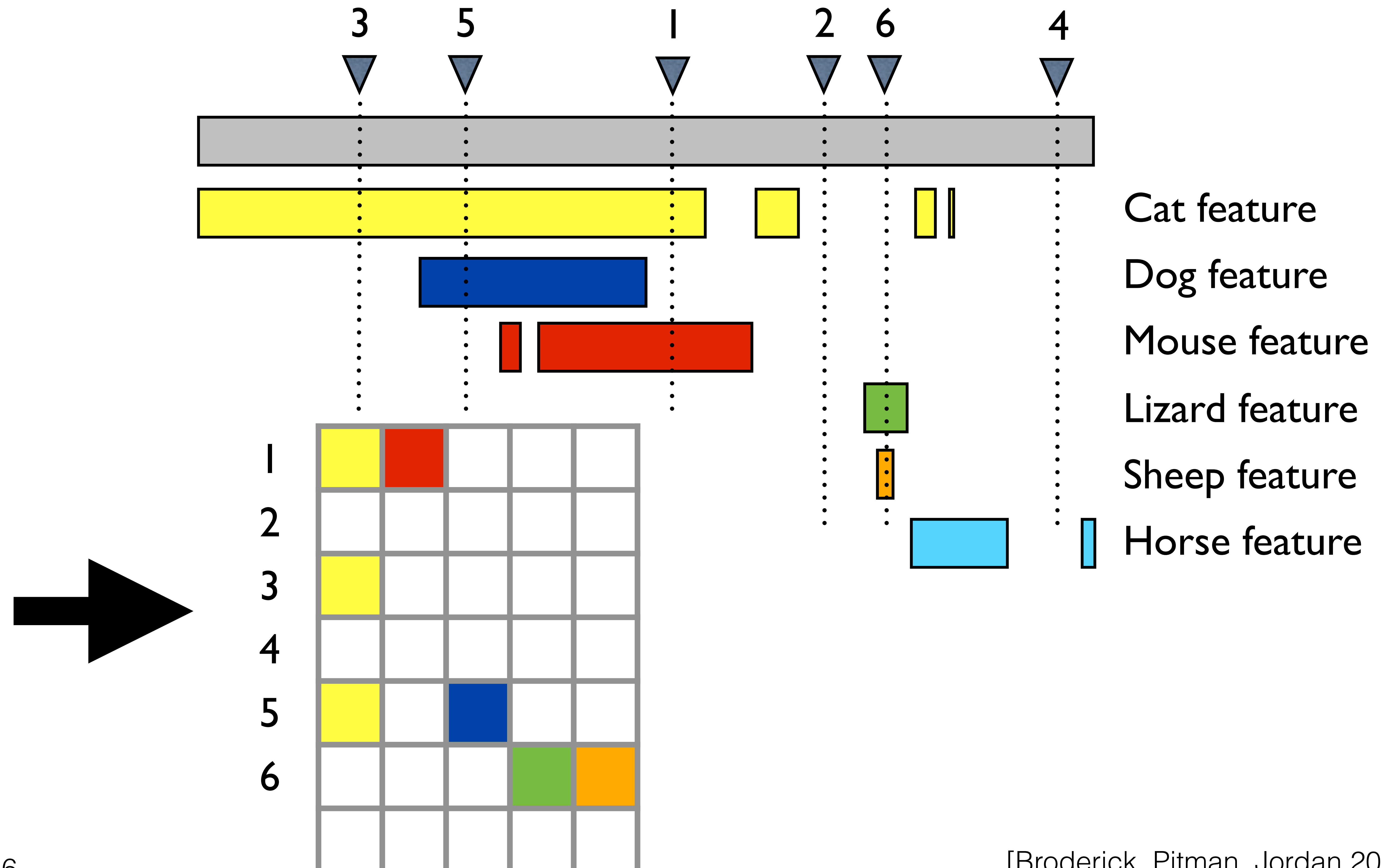
[Broderick, Pitman, Jordan 2013]

Feature allocation



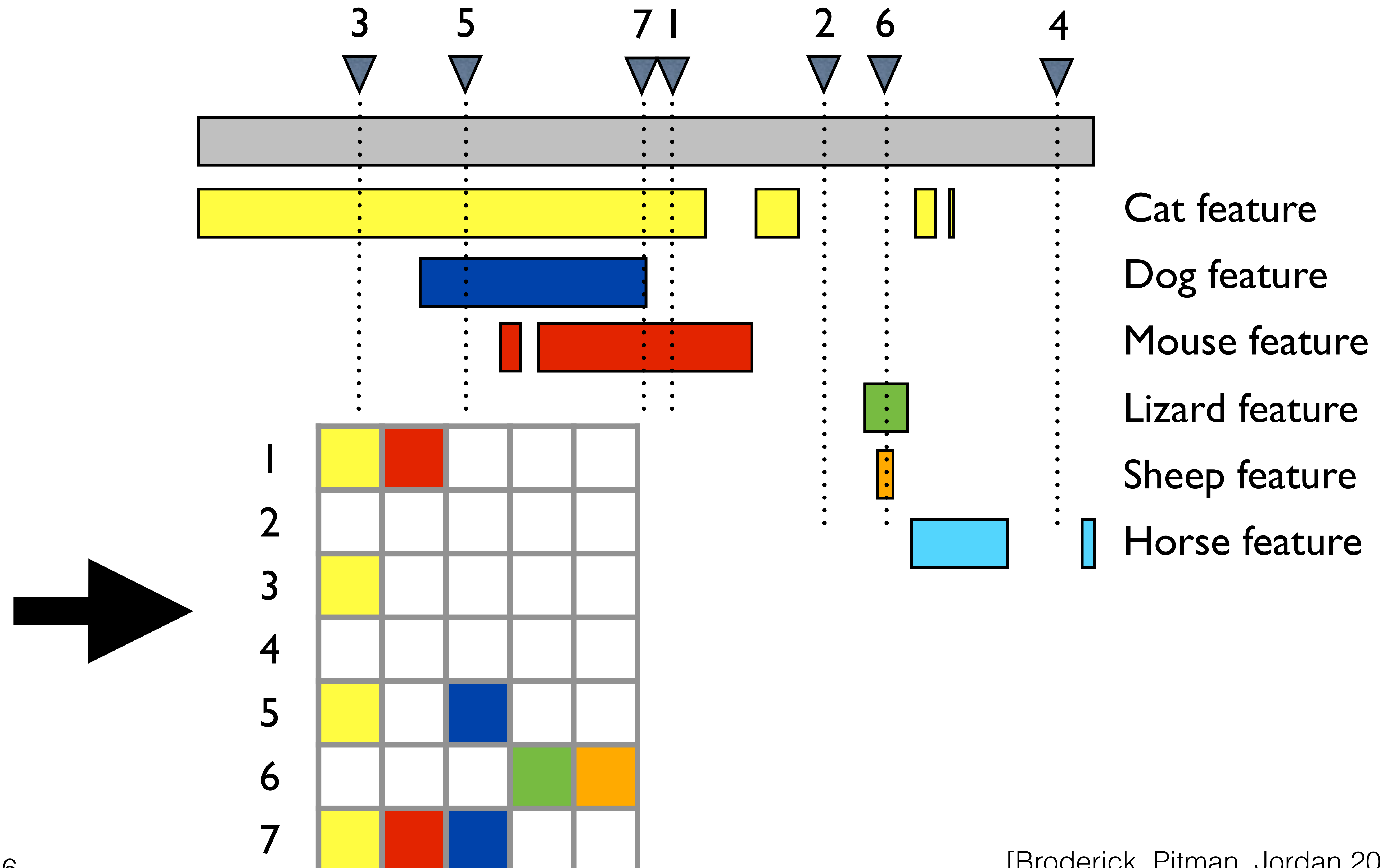
[Broderick, Pitman, Jordan 2013]

Feature allocation



[Broderick, Pitman, Jordan 2013]

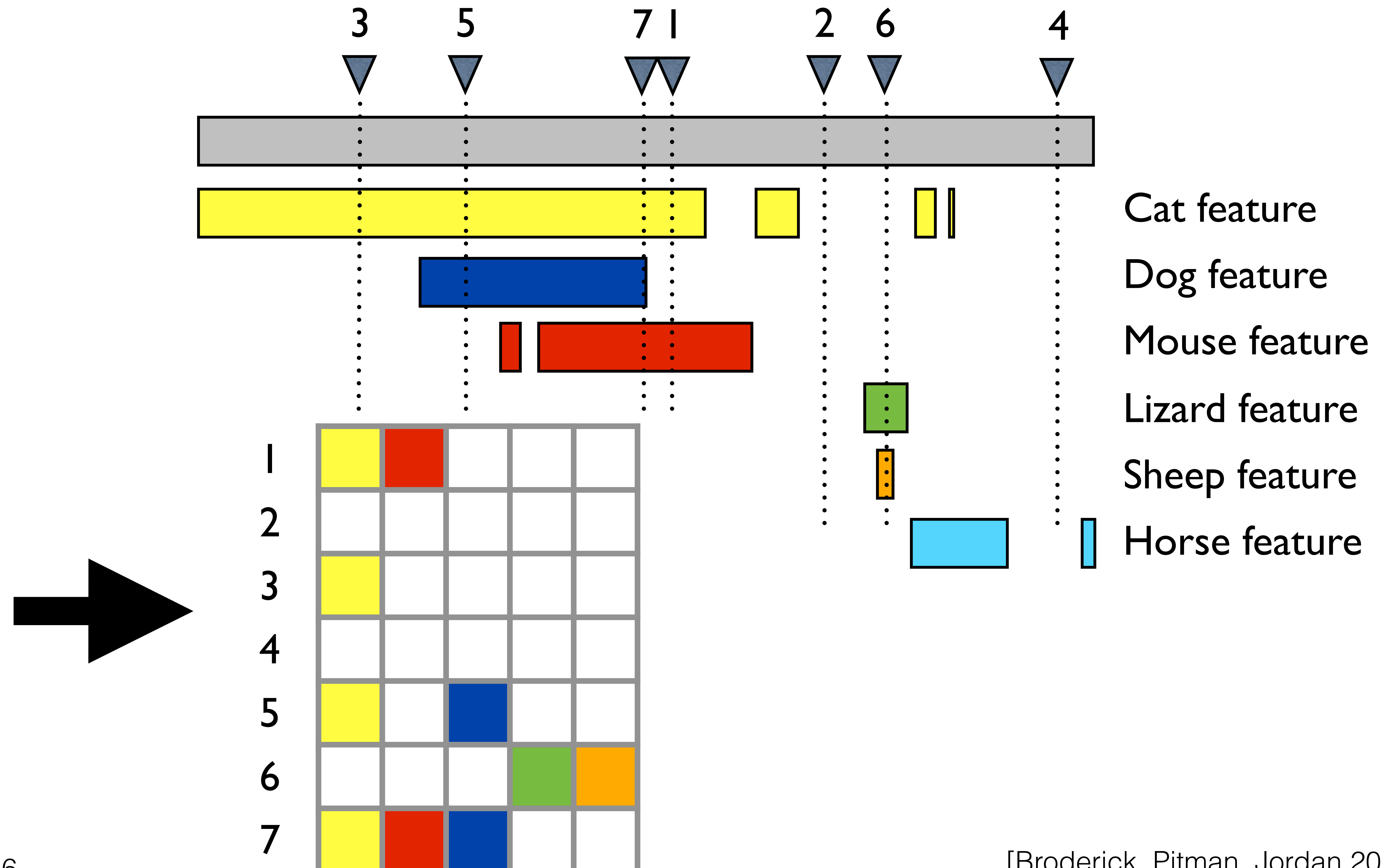
Feature allocation



[Broderick, Pitman, Jordan 2013]

Feature allocation

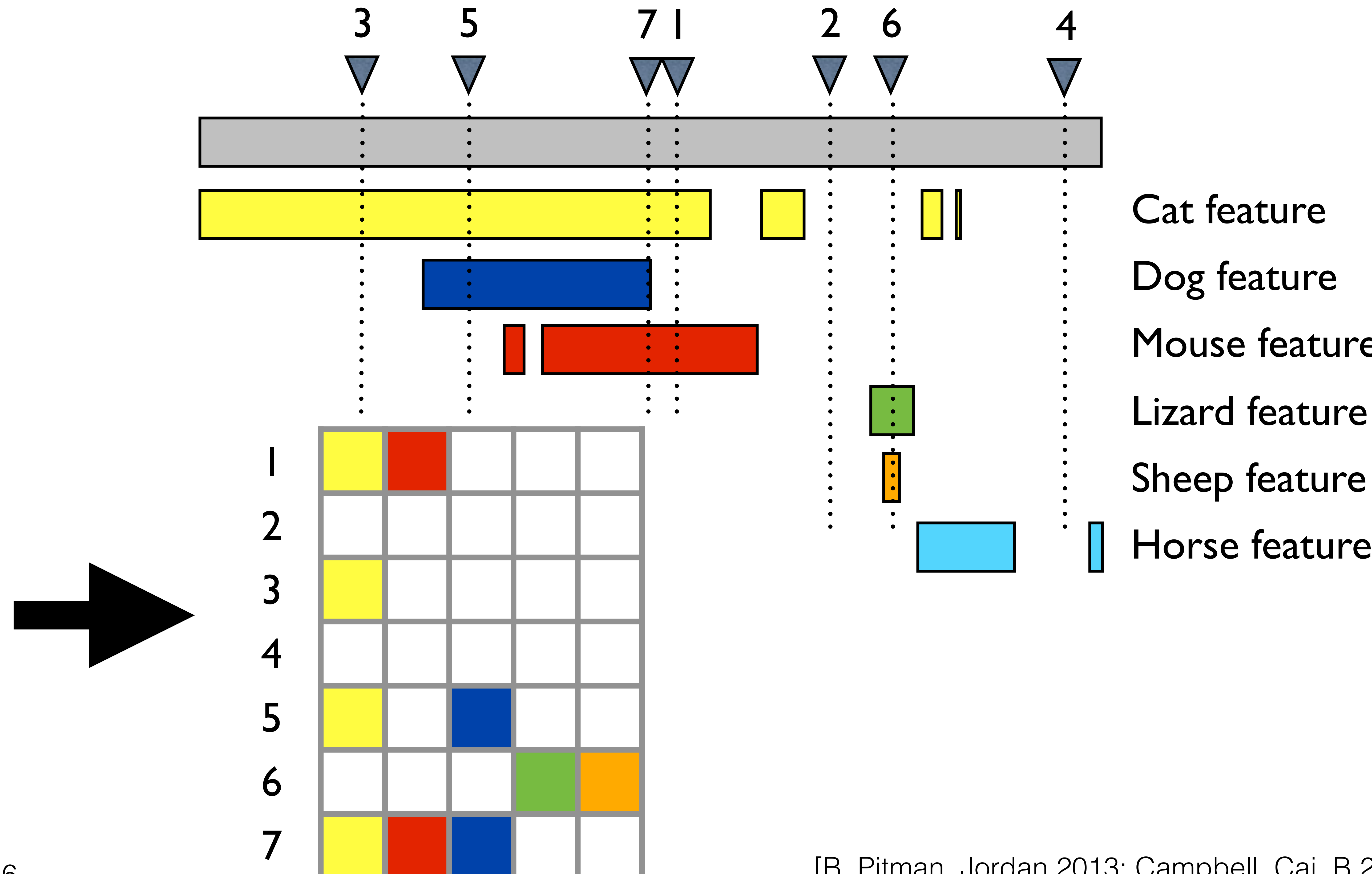
Thm (BPJ). A feature allocation is exchangeable iff it has a feature paintbox representation



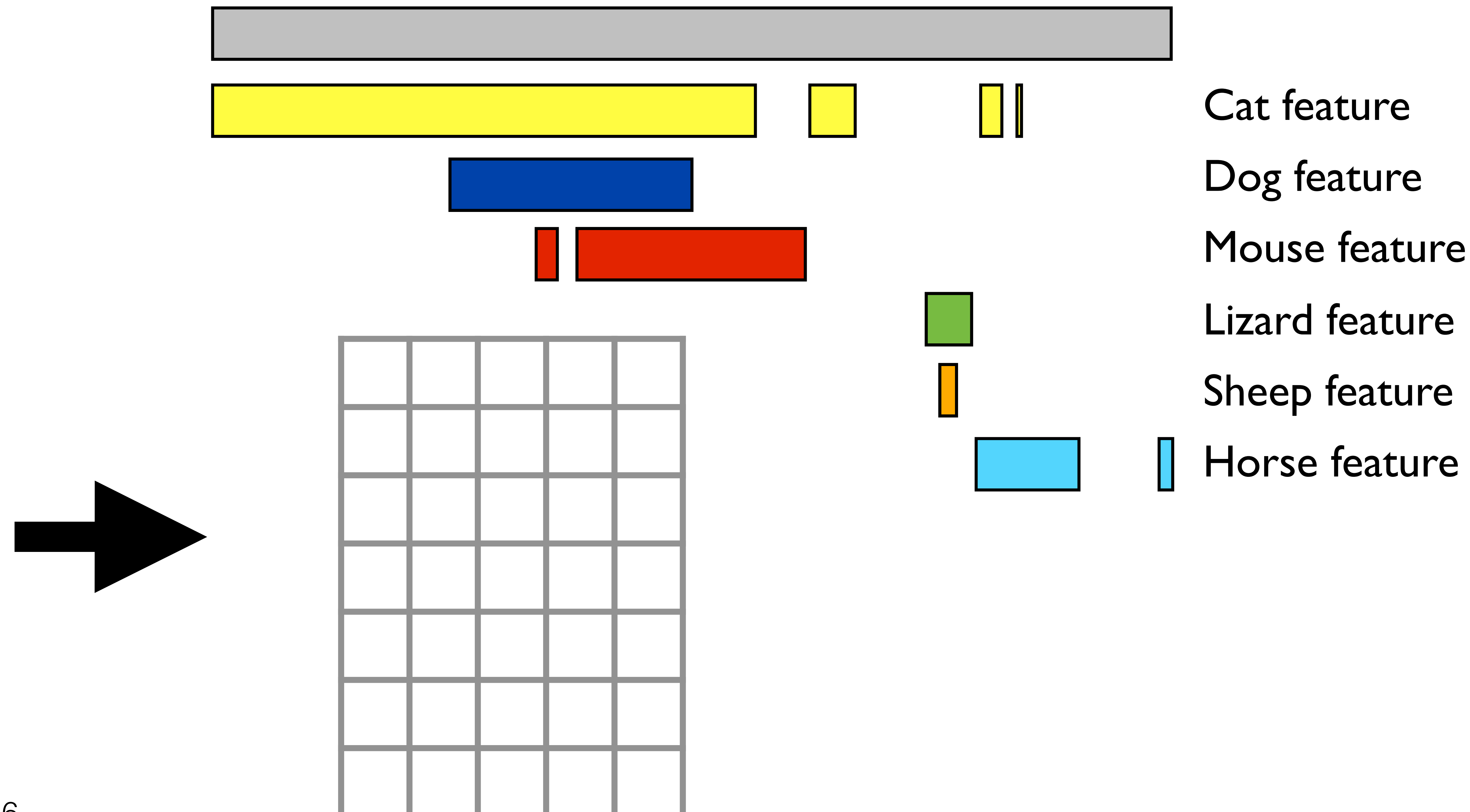
[Broderick, Pitman, Jordan 2013]

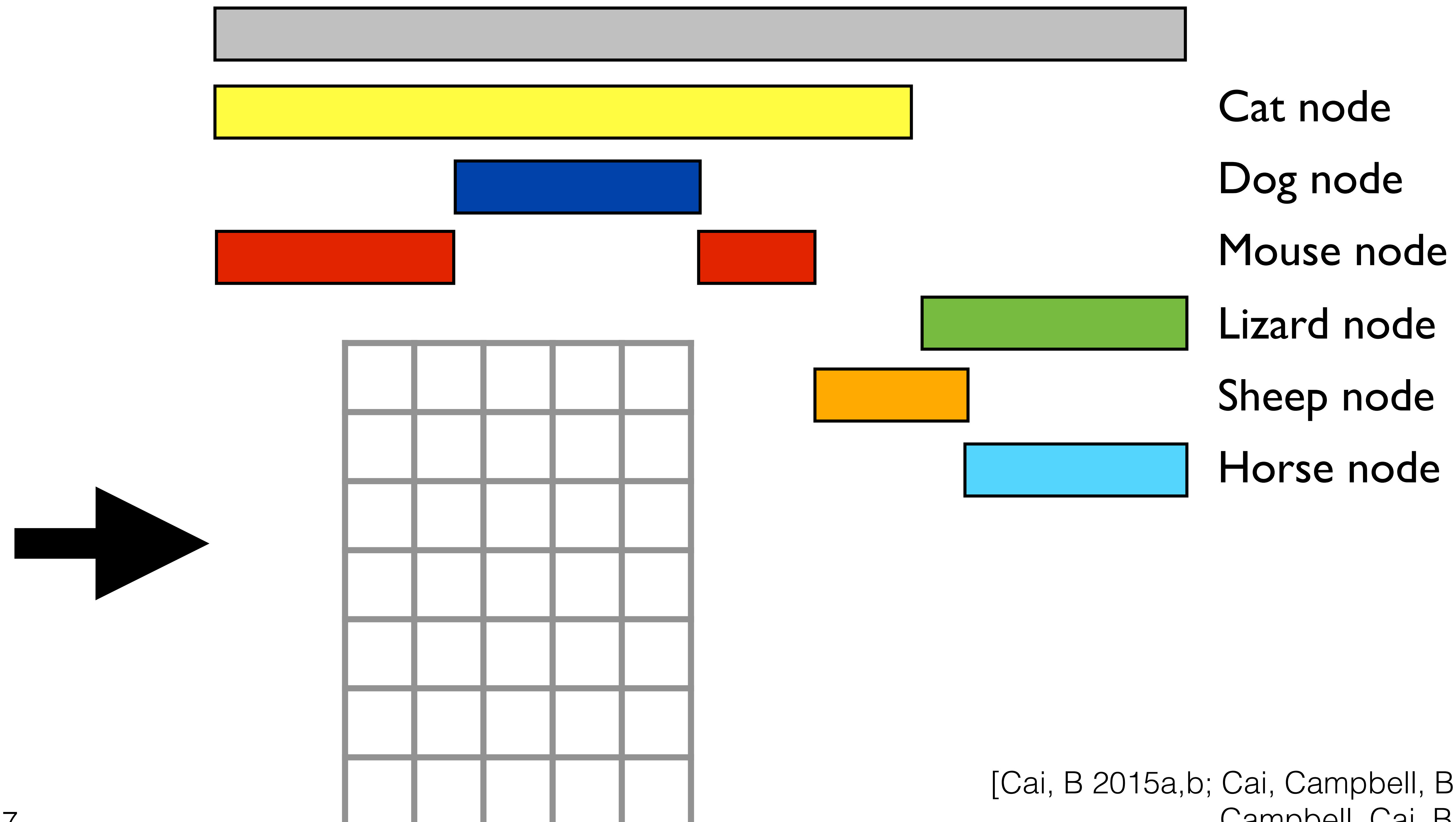
Feature allocation

Thm (BPJ). A feature allocation is exchangeable iff it has a feature paintbox representation



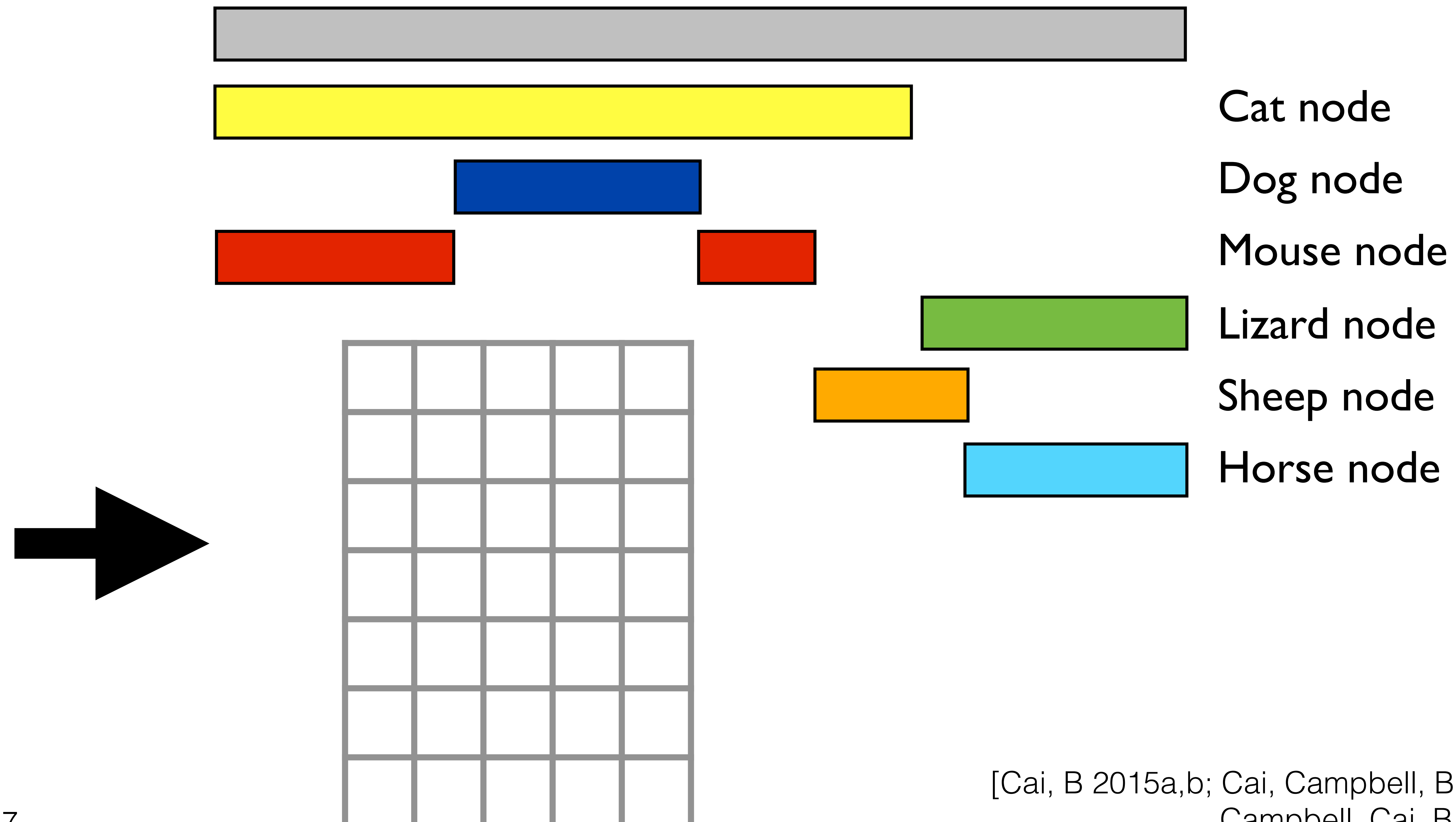
[B, Pitman, Jordan 2013; Campbell, Cai, B 2016]


Cat feature
Dog feature
Mouse feature
Lizard feature
Sheep feature
Horse feature



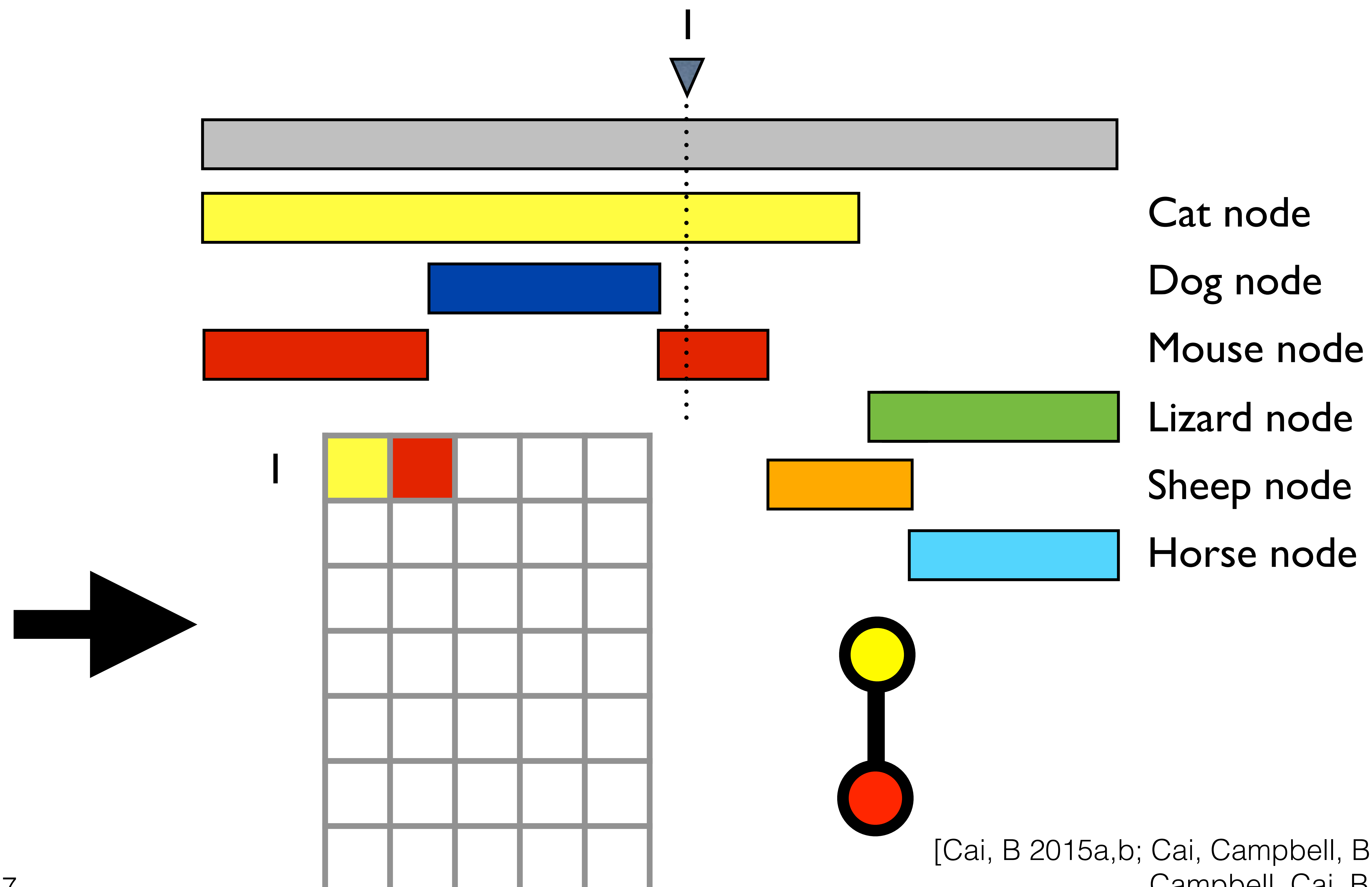
[Cai, B 2015a,b; Cai, Campbell, B, 2016; Campbell, Cai, B, 2016]

Edge-exchangeable graph



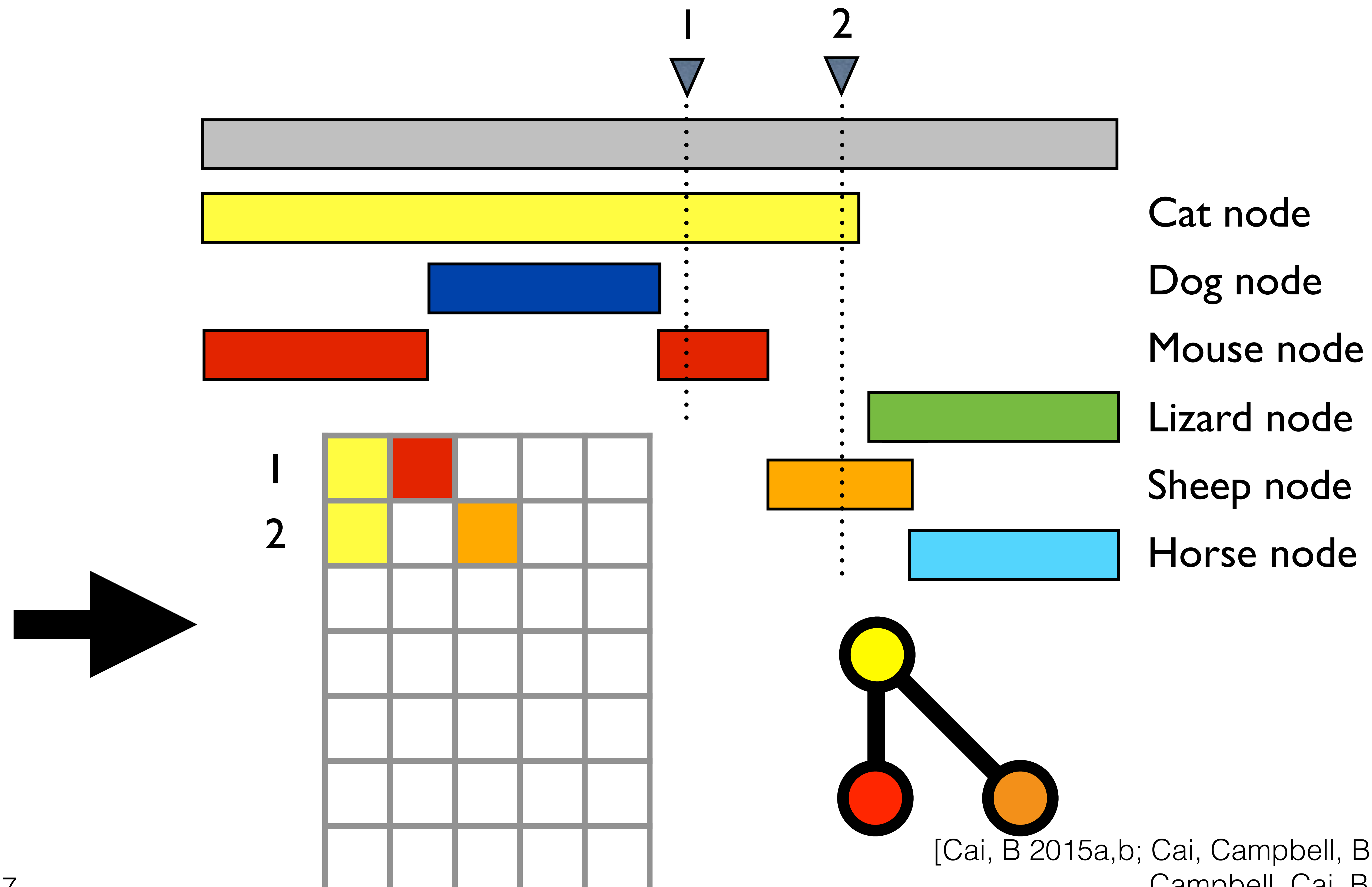
[Cai, B 2015a,b; Cai, Campbell, B, 2016;
Campbell, Cai, B, 2016]

Edge-exchangeable graph



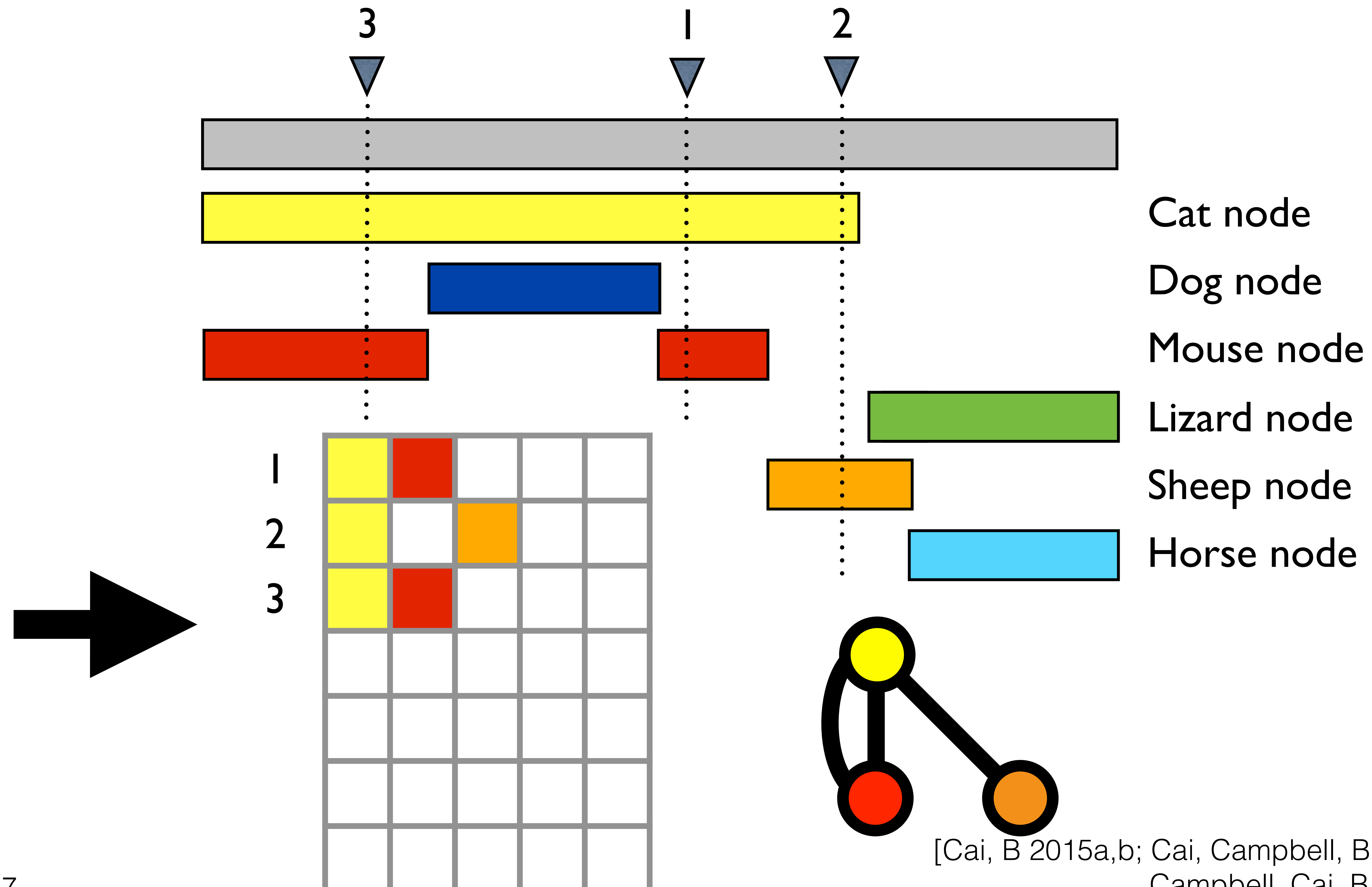
[Cai, B 2015a,b; Cai, Campbell, B, 2016;
Campbell, Cai, B, 2016]

Edge-exchangeable graph



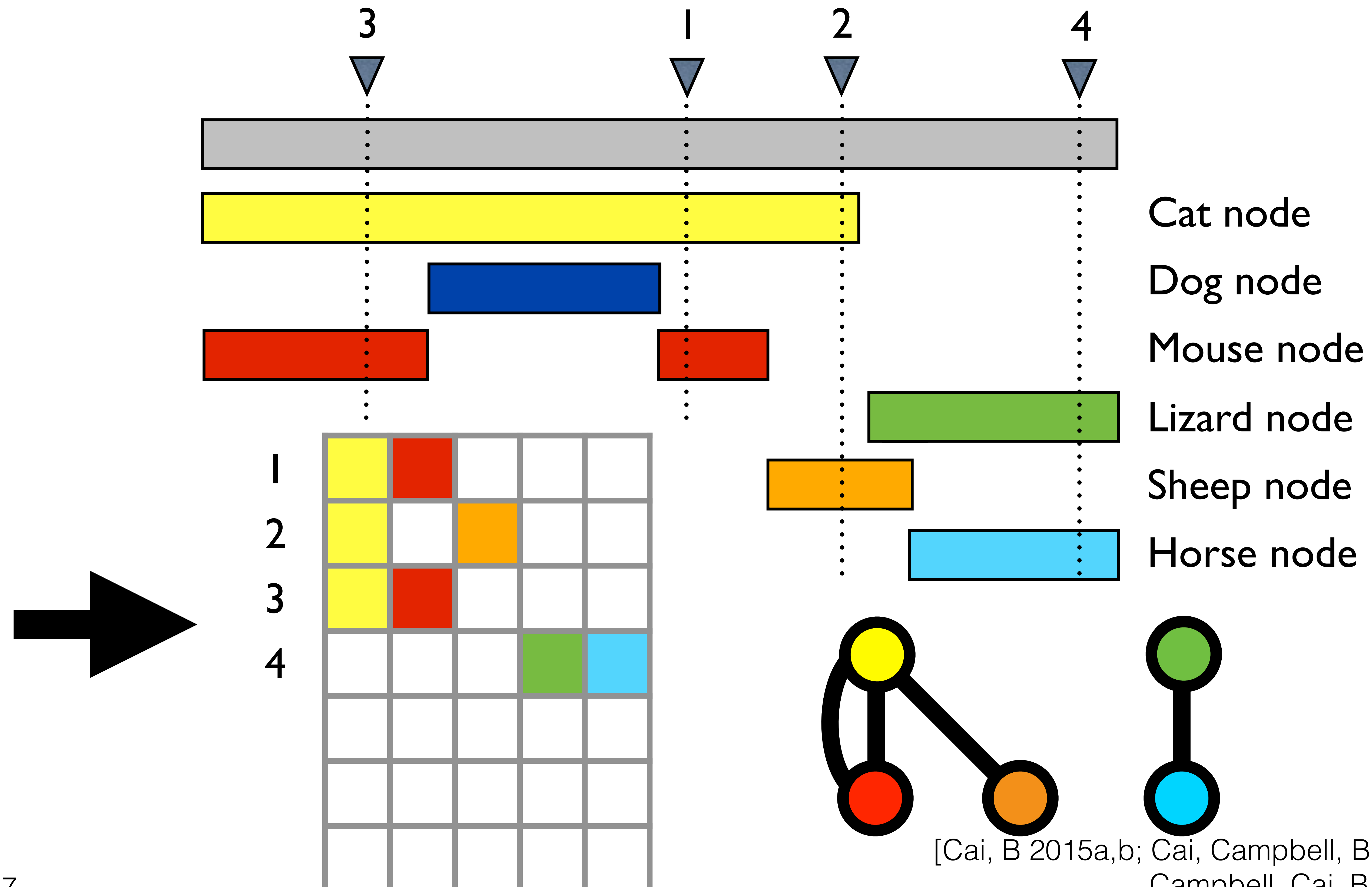
[Cai, B 2015a,b; Cai, Campbell, B, 2016; Campbell, Cai, B, 2016]

Edge-exchangeable graph



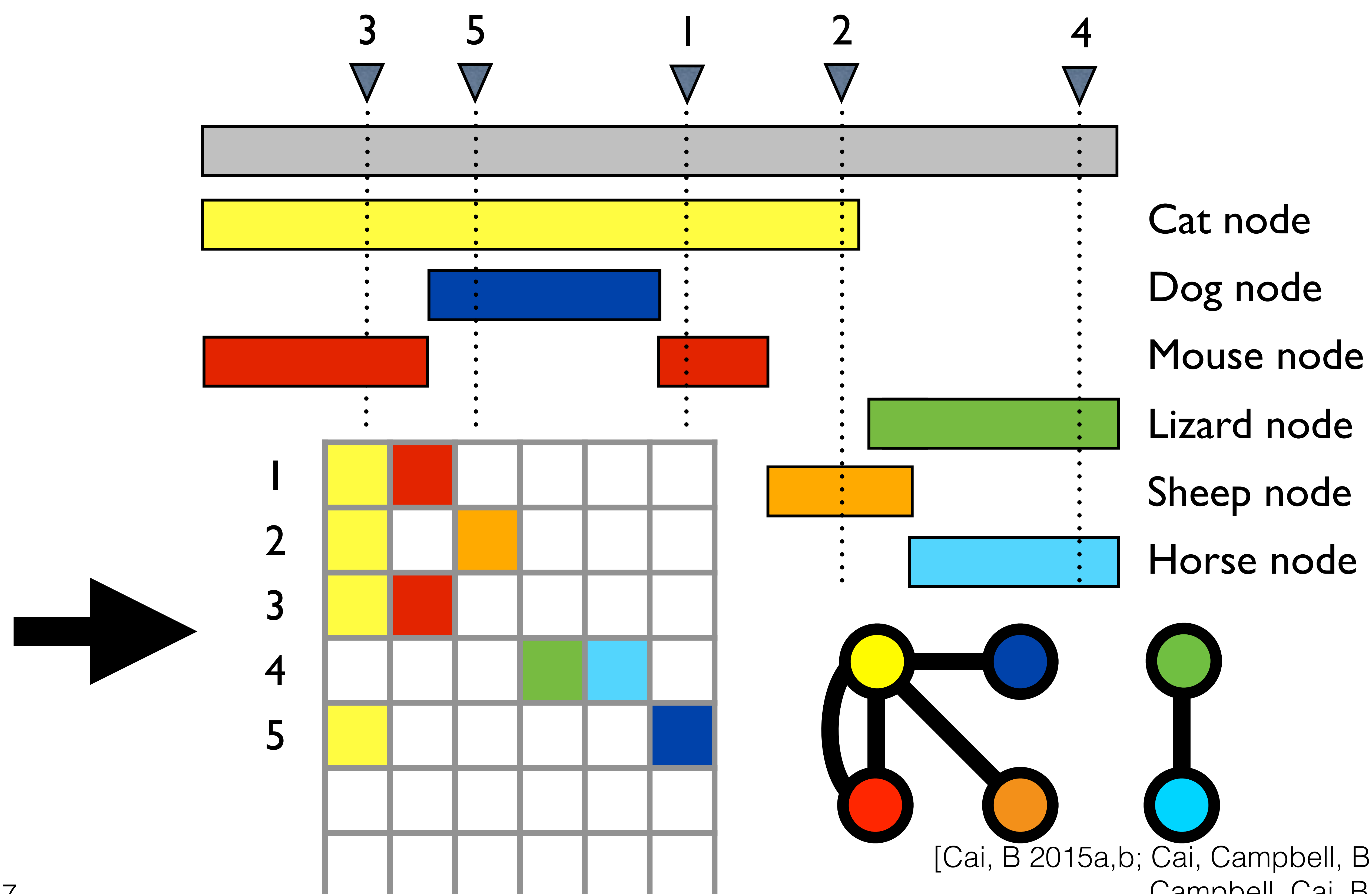
[Cai, B 2015a,b; Cai, Campbell, B, 2016; Campbell, Cai, B, 2016]

Edge-exchangeable graph



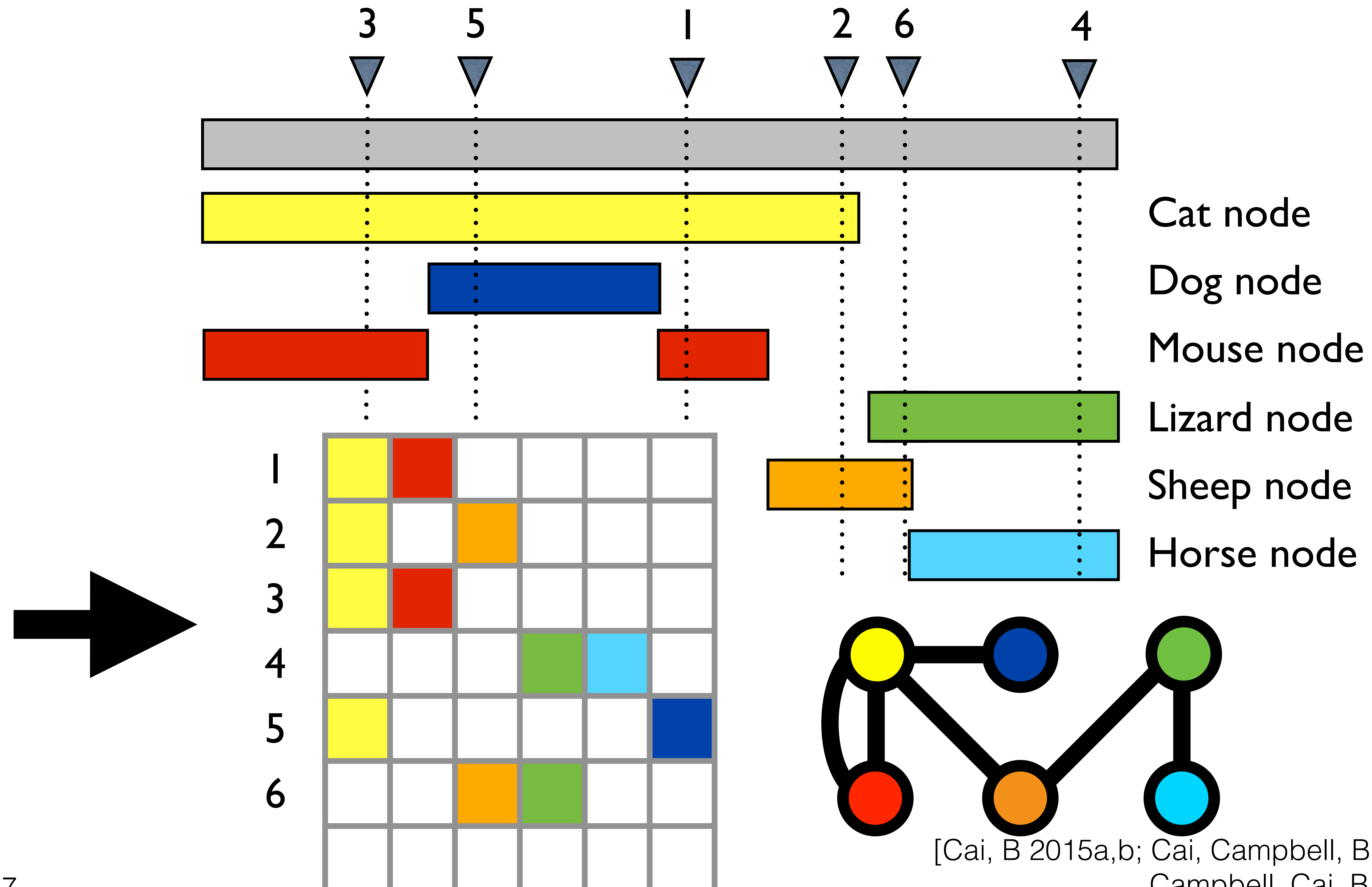
[Cai, B 2015a,b; Cai, Campbell, B, 2016; Campbell, Cai, B, 2016]

Edge-exchangeable graph



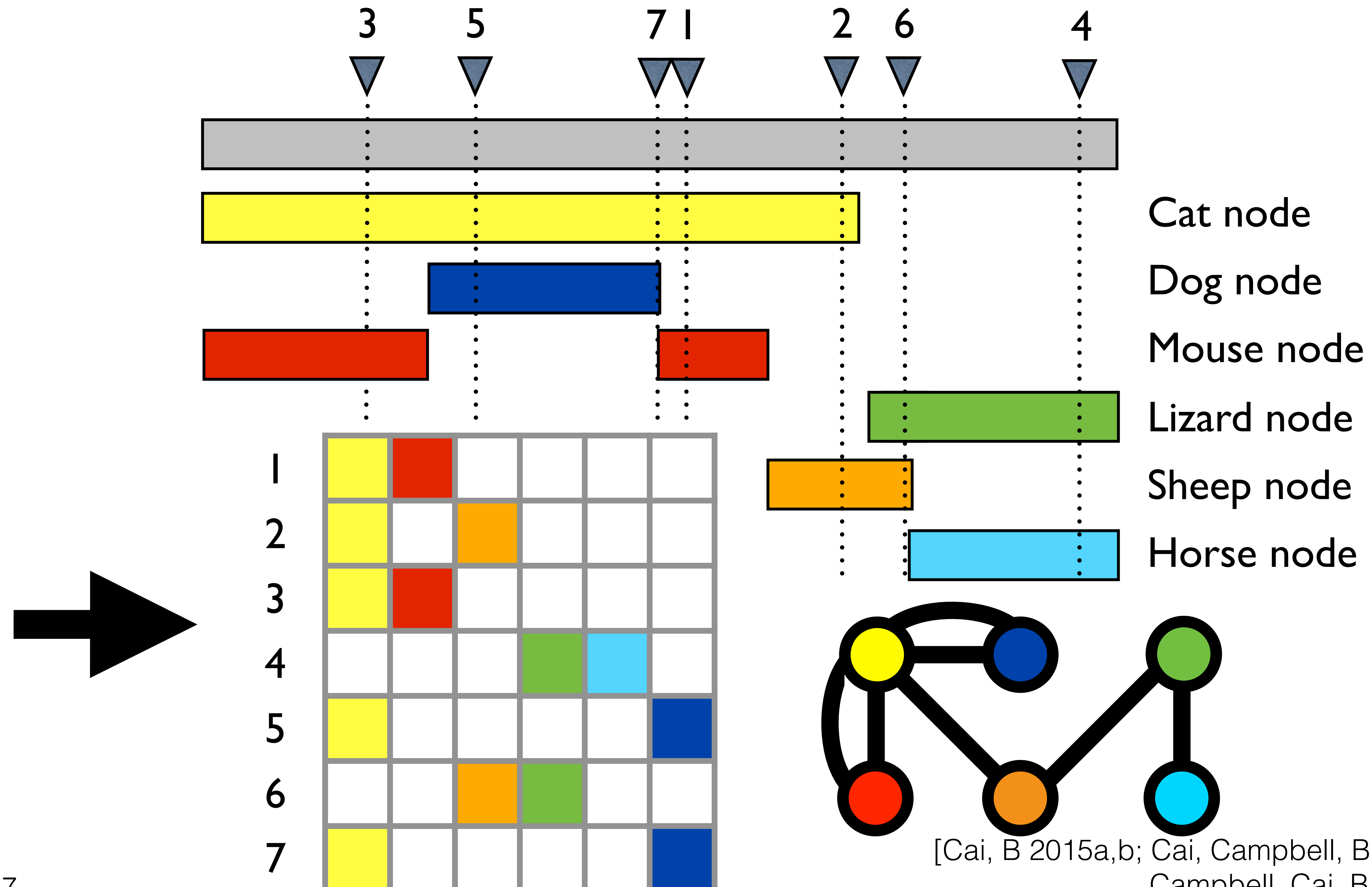
[Cai, B 2015a,b; Cai, Campbell, B, 2016;
Campbell, Cai, B, 2016]

Edge-exchangeable graph



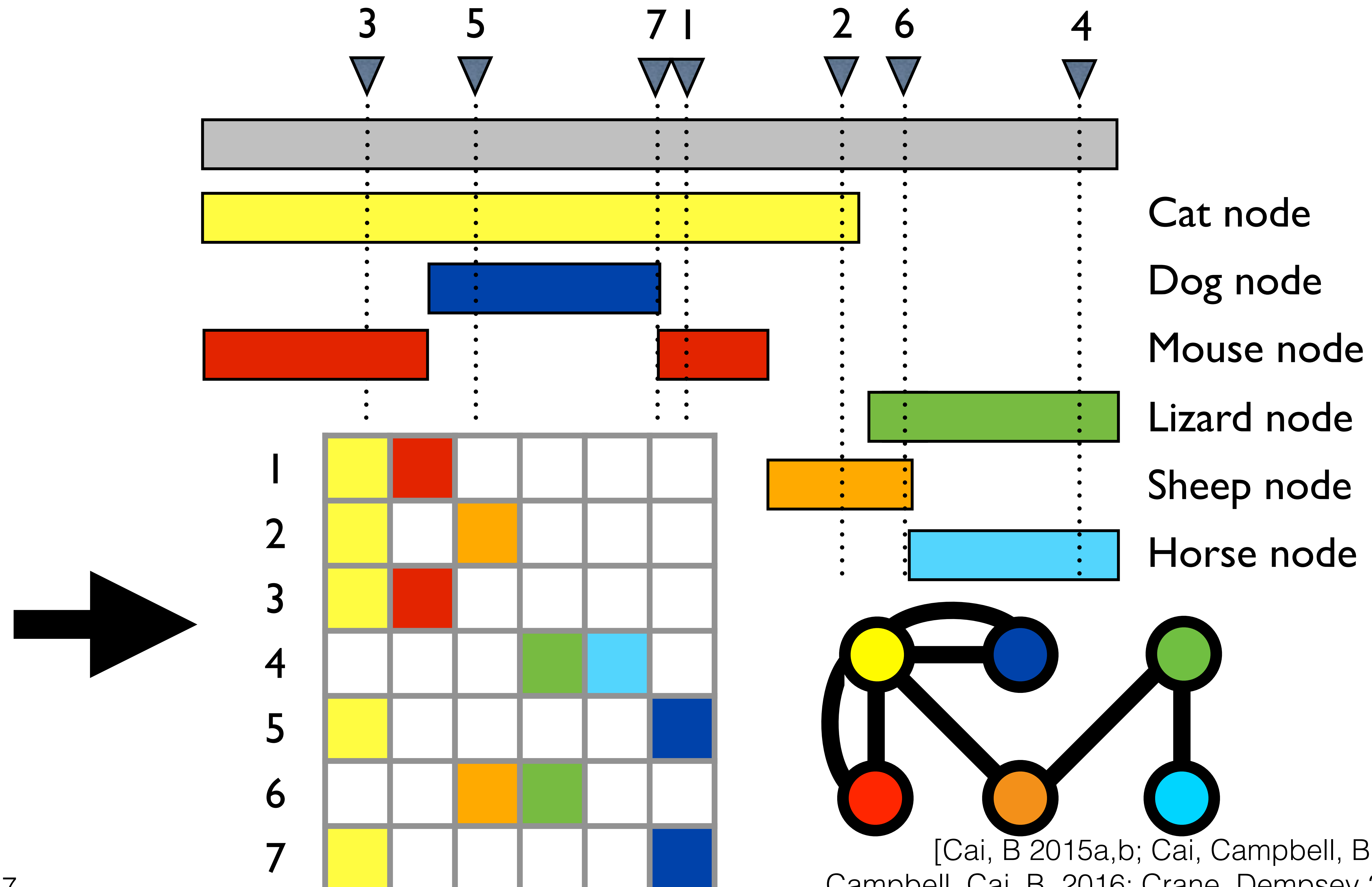
[Cai, B 2015a,b; Cai, Campbell, B, 2016; Campbell, Cai, B, 2016]

Edge-exchangeable graph



[Cai, B 2015a,b; Cai, Campbell, B, 2016;
Campbell, Cai, B, 2016]

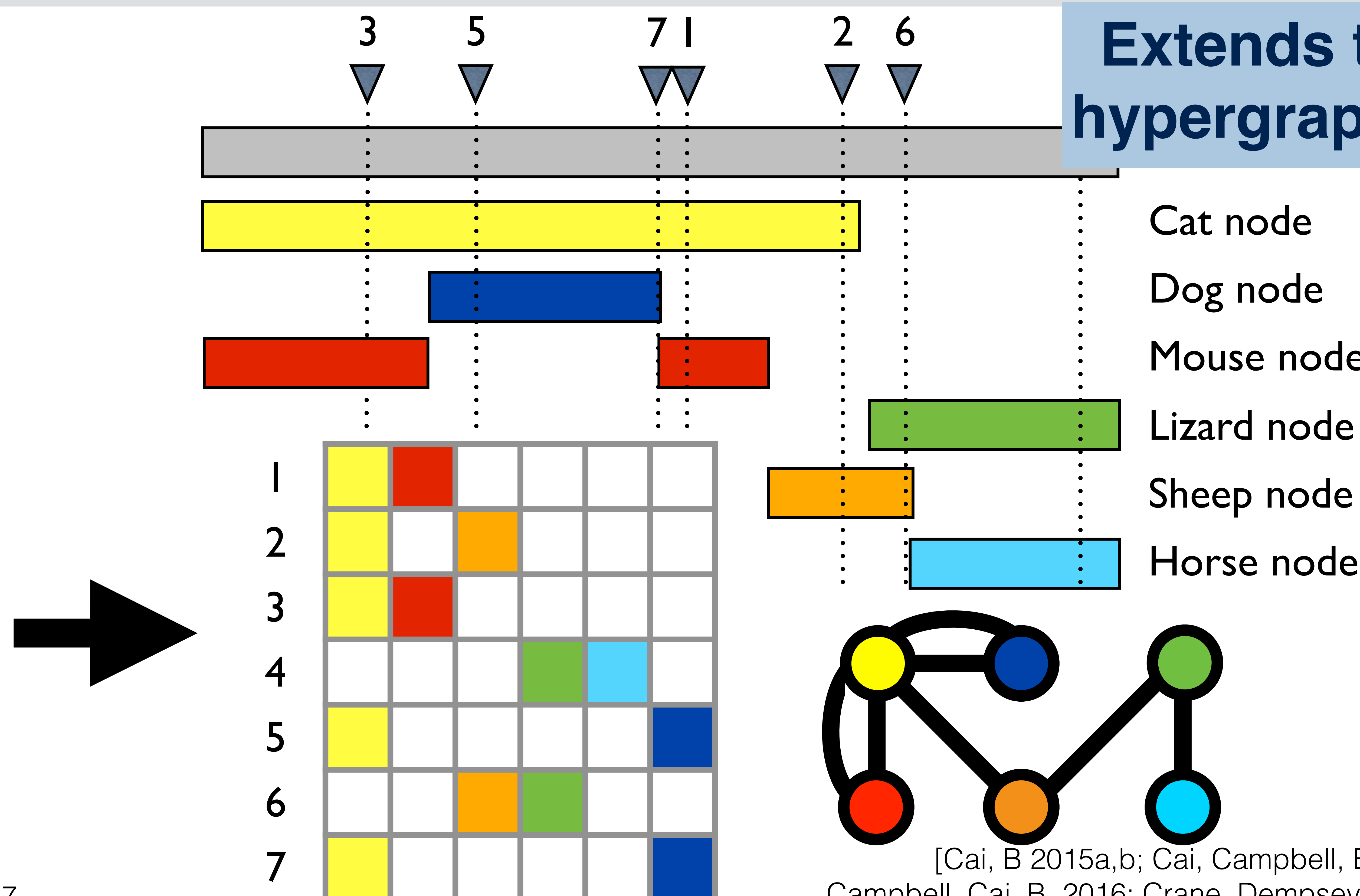
Thm (CCB). A graph sequence is edge-exchangeable iff it has a graph paintbox



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

Thm (CCB). A graph sequence is edge-exchangeable iff it has a graph paintbox

Extends to hypergraphs

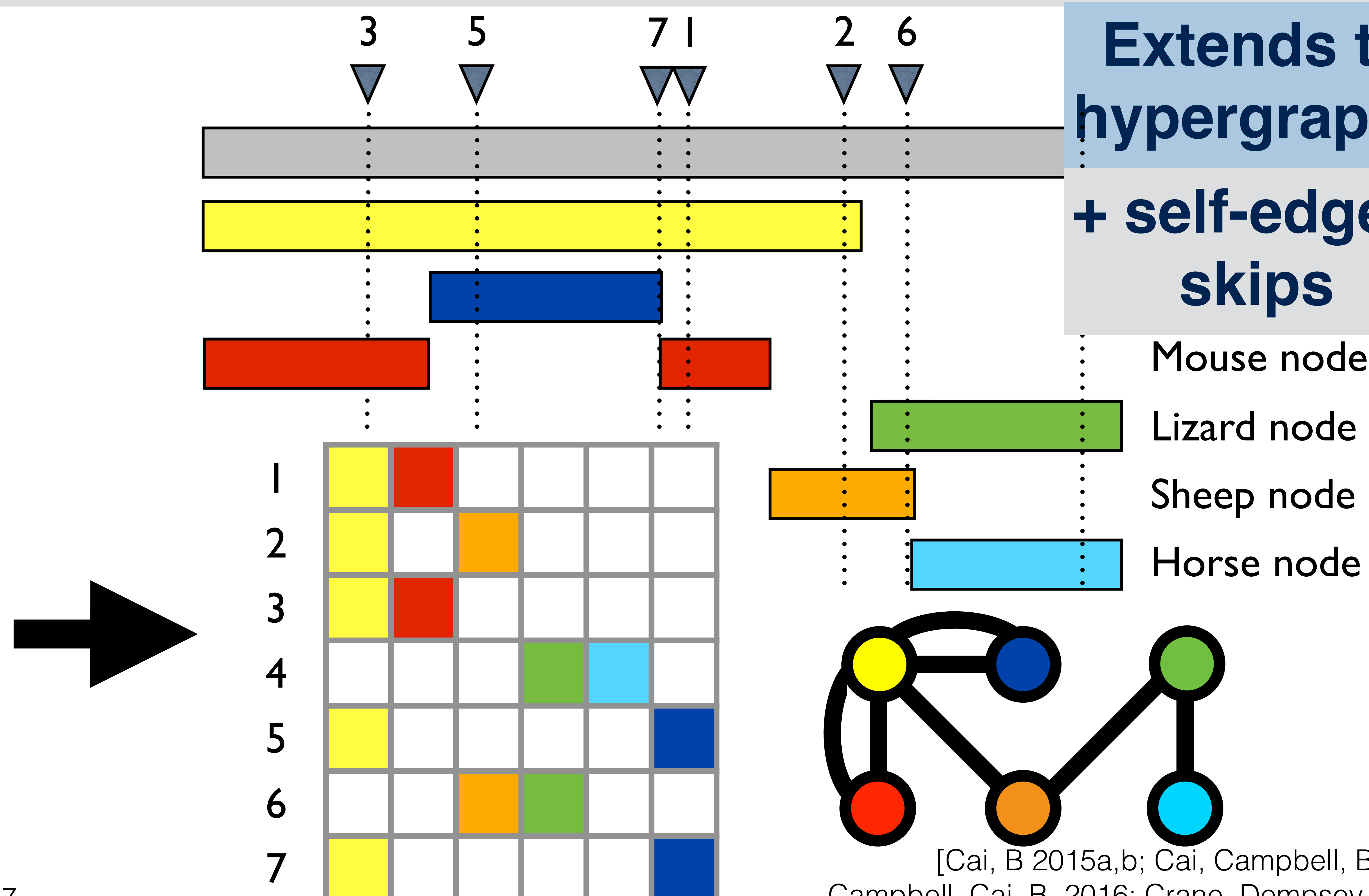


[Cai, B 2015a,b; Cai, Campbell, B, 2016; Campbell, Cai, B, 2016; Crane, Dempsey 2016b]

Thm (CCB). A graph sequence is edge-exchangeable iff it has a graph paintbox

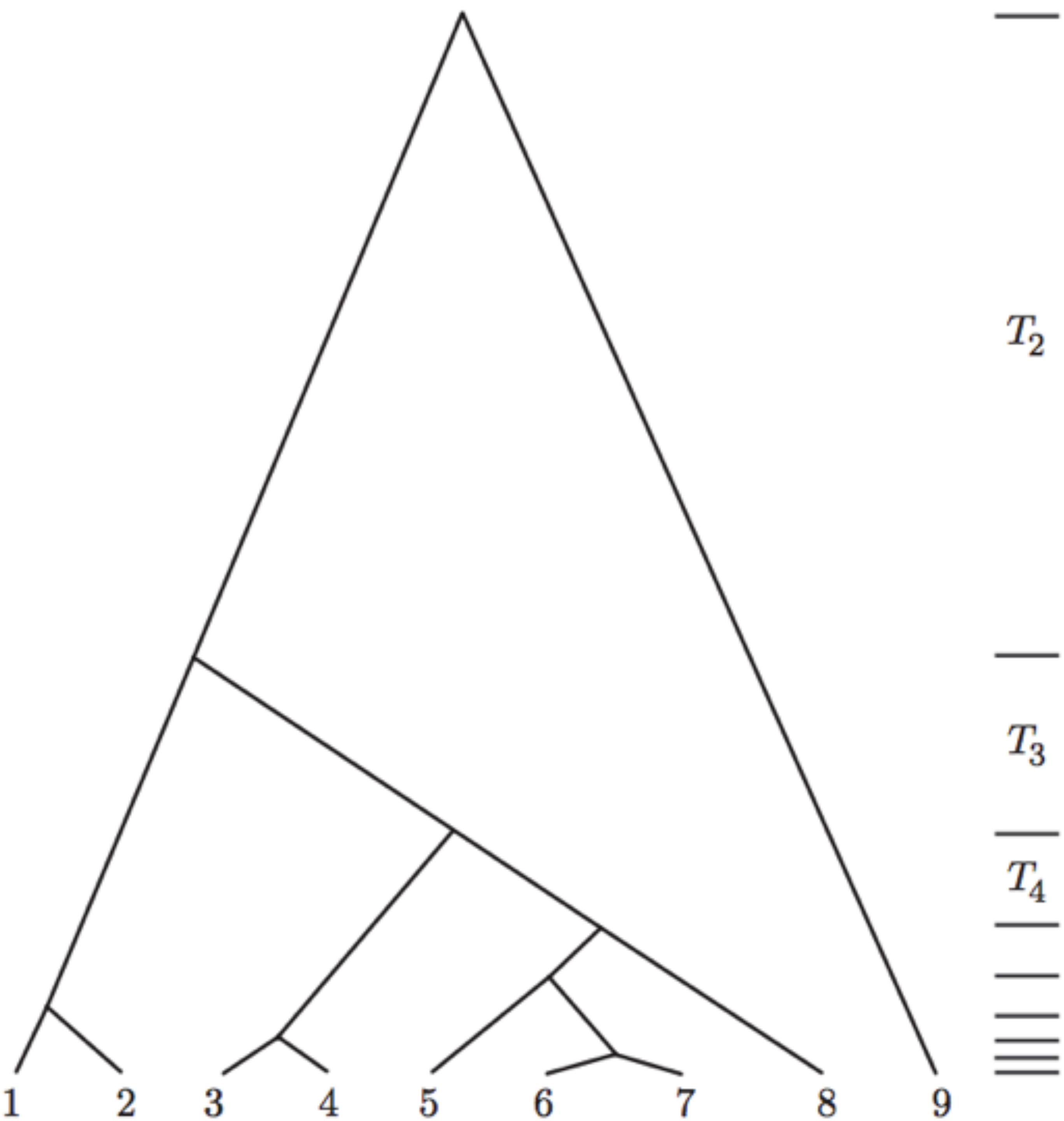
Extends to hypergraphs

+ self-edges, skips



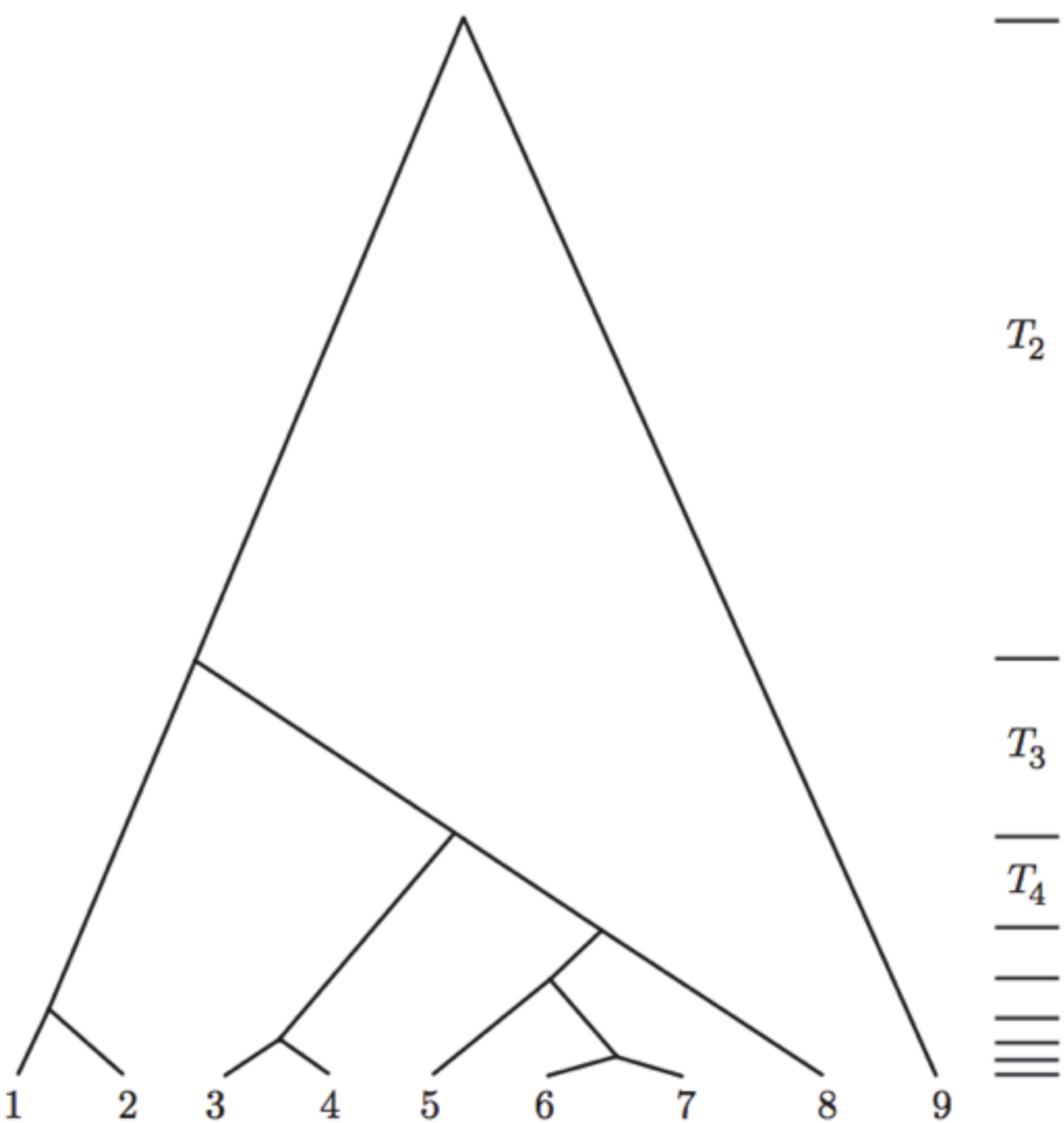
[Cai, B 2015a,b; Cai, Campbell, B, 2016; Campbell, Cai, B, 2016; Crane, Dempsey 2016b]

Genealogy, trees, beyond trees



[Wakeley 2008]

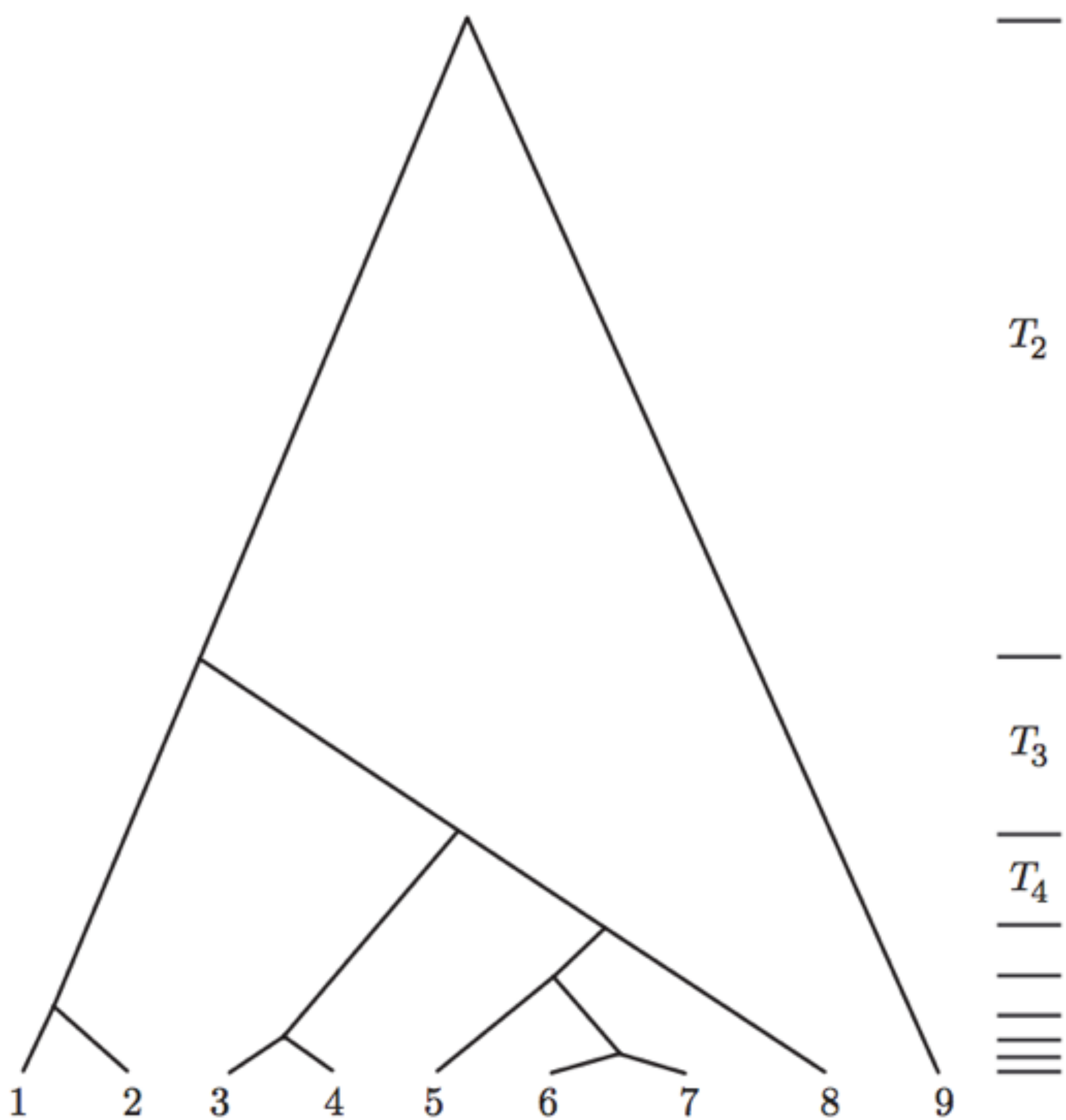
Genealogy, trees, beyond trees



[Wakeley 2008]

- Kingman coalescent

Genealogy, trees, beyond trees

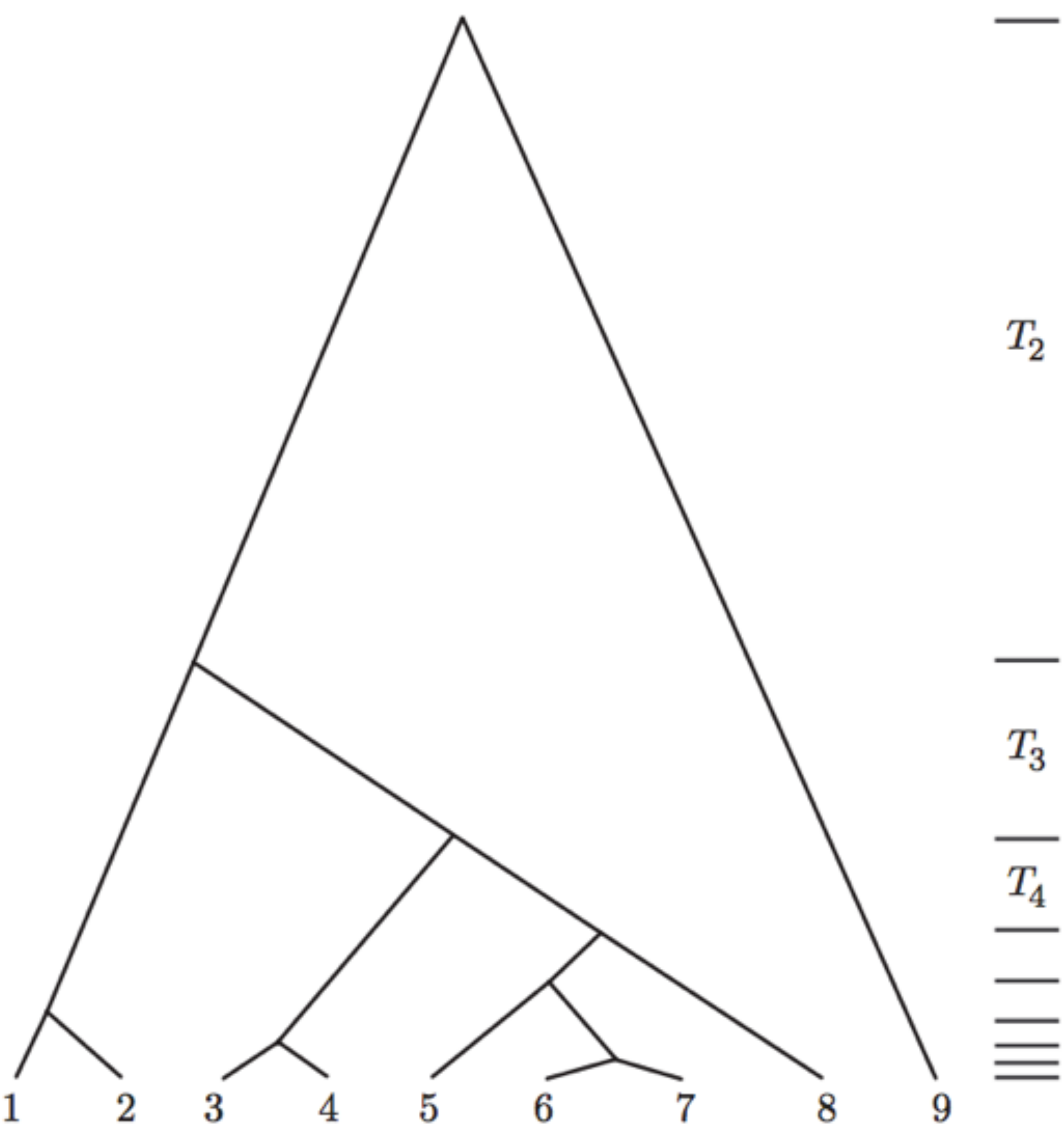


[Wakeley 2008]

- Kingman coalescent

[Kingman 1982

Genealogy, trees, beyond trees

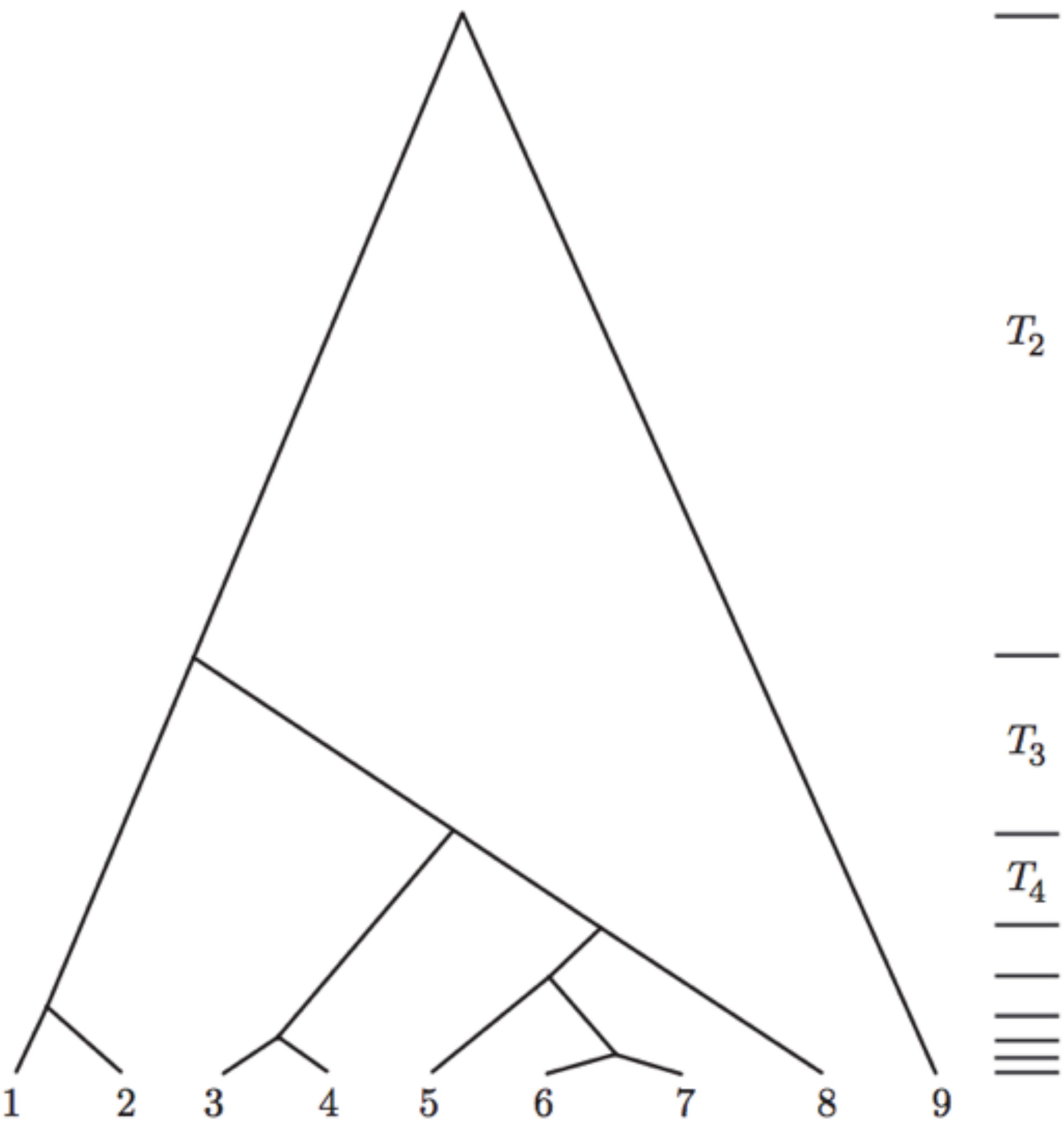


[Wakeley 2008]

- Kingman coalescent
- Fragmentation
- Coagulation

[Kingman 1982

Genealogy, trees, beyond trees

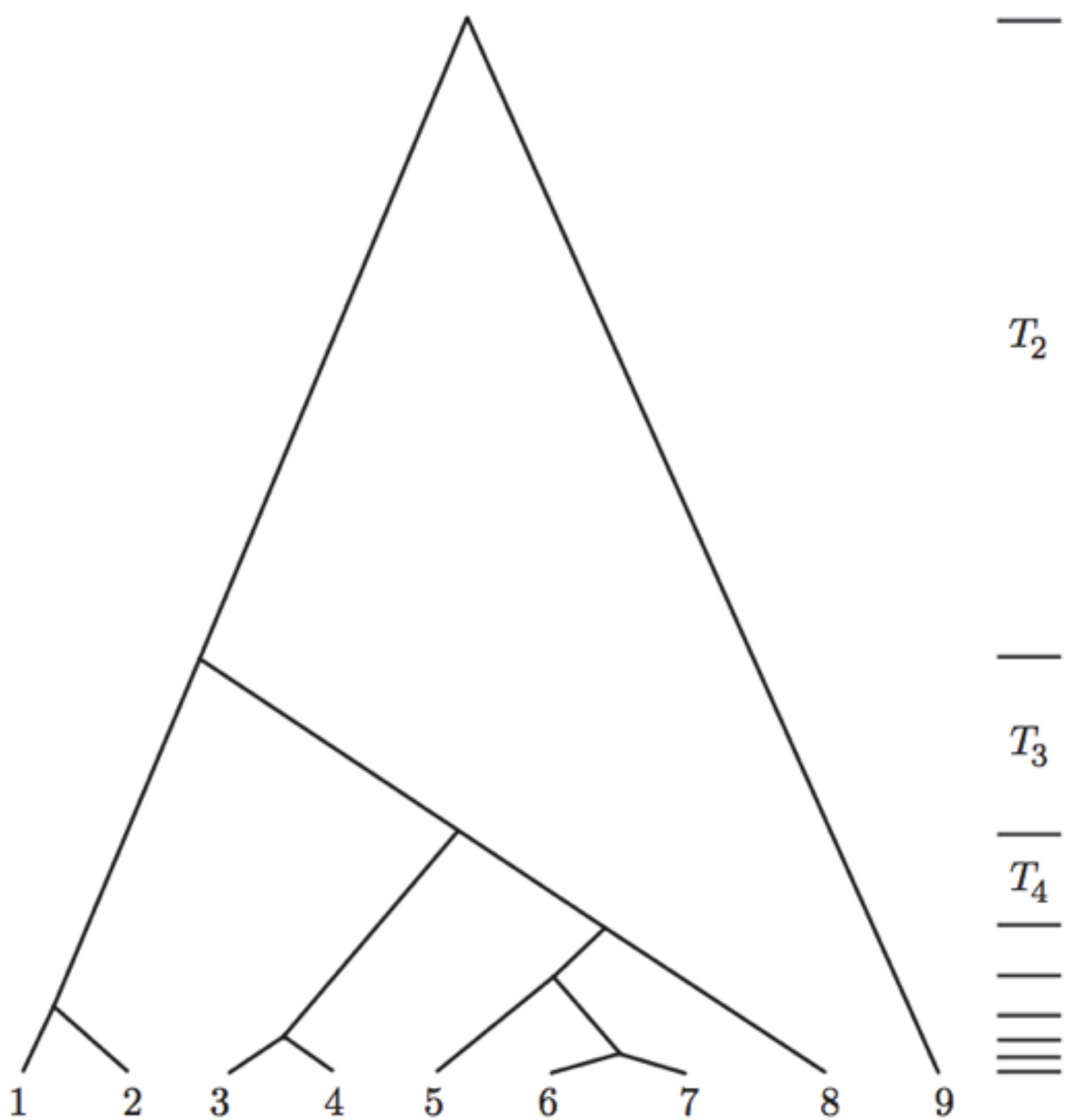


[Wakeley 2008]

- Kingman coalescent
- Fragmentation
- Coagulation

[Kingman 1982, Bertoin 2006, Teh et al 2011

Genealogy, trees, beyond trees

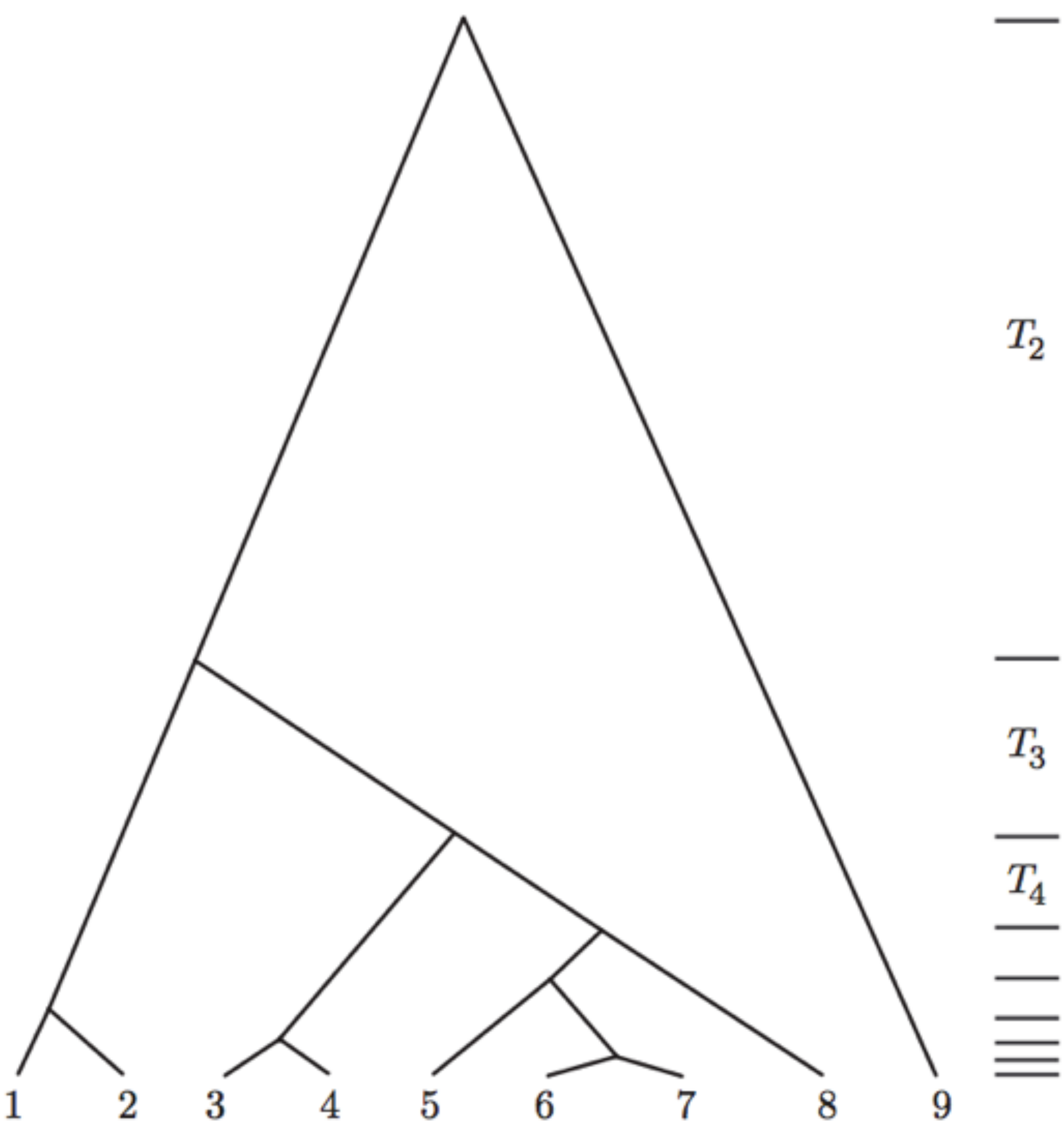


[Wakeley 2008]

- Kingman coalescent
- Fragmentation
- Coagulation
- Dirichlet diffusion tree

[Kingman 1982, Bertoin 2006, Teh et al 2011

Genealogy, trees, beyond trees

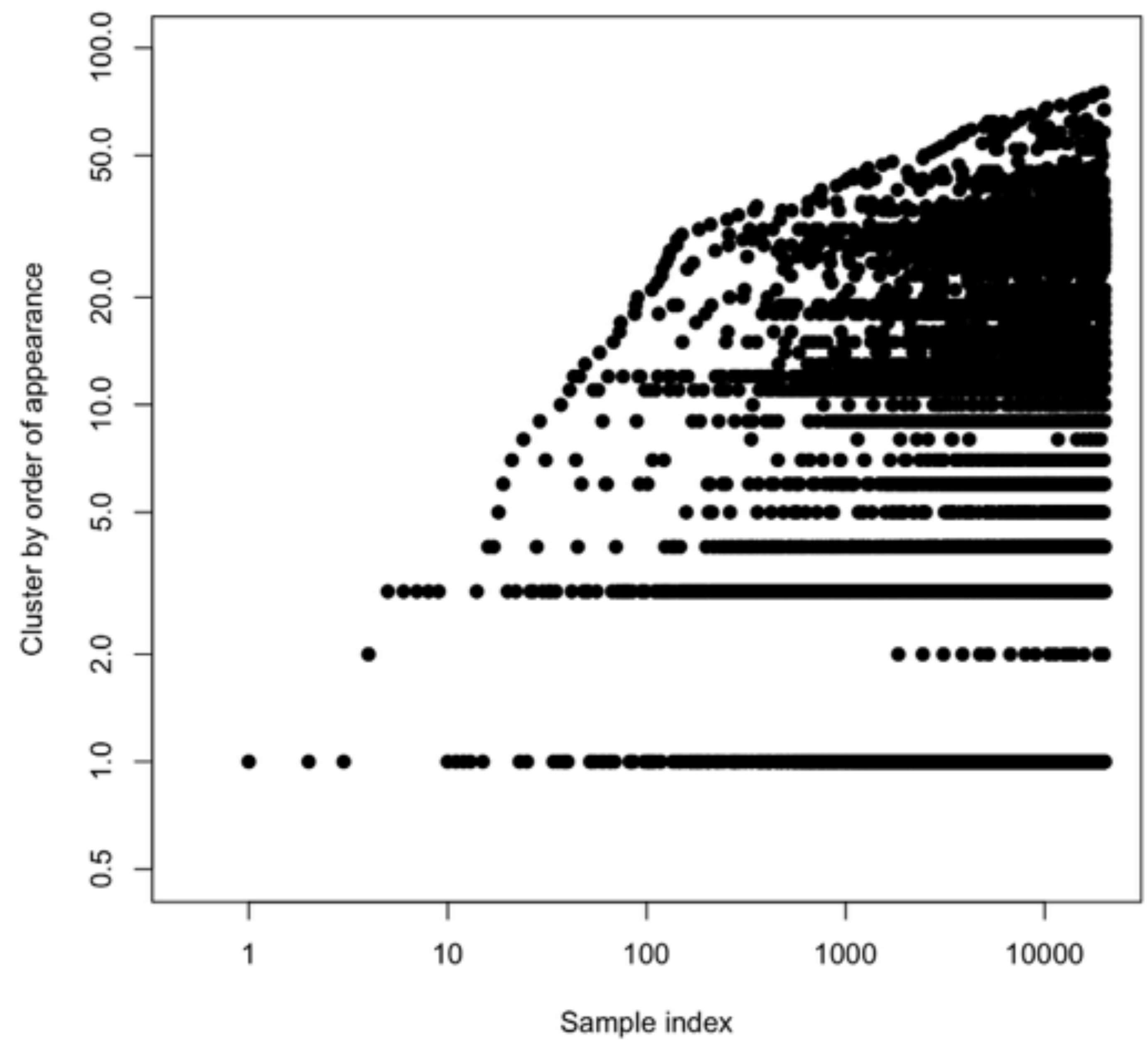


[Wakeley 2008]

- Kingman coalescent
- Fragmentation
- Coagulation
- Dirichlet diffusion tree

[Kingman 1982, Bertoin 2006, Teh et al 2011, Neal 2003]

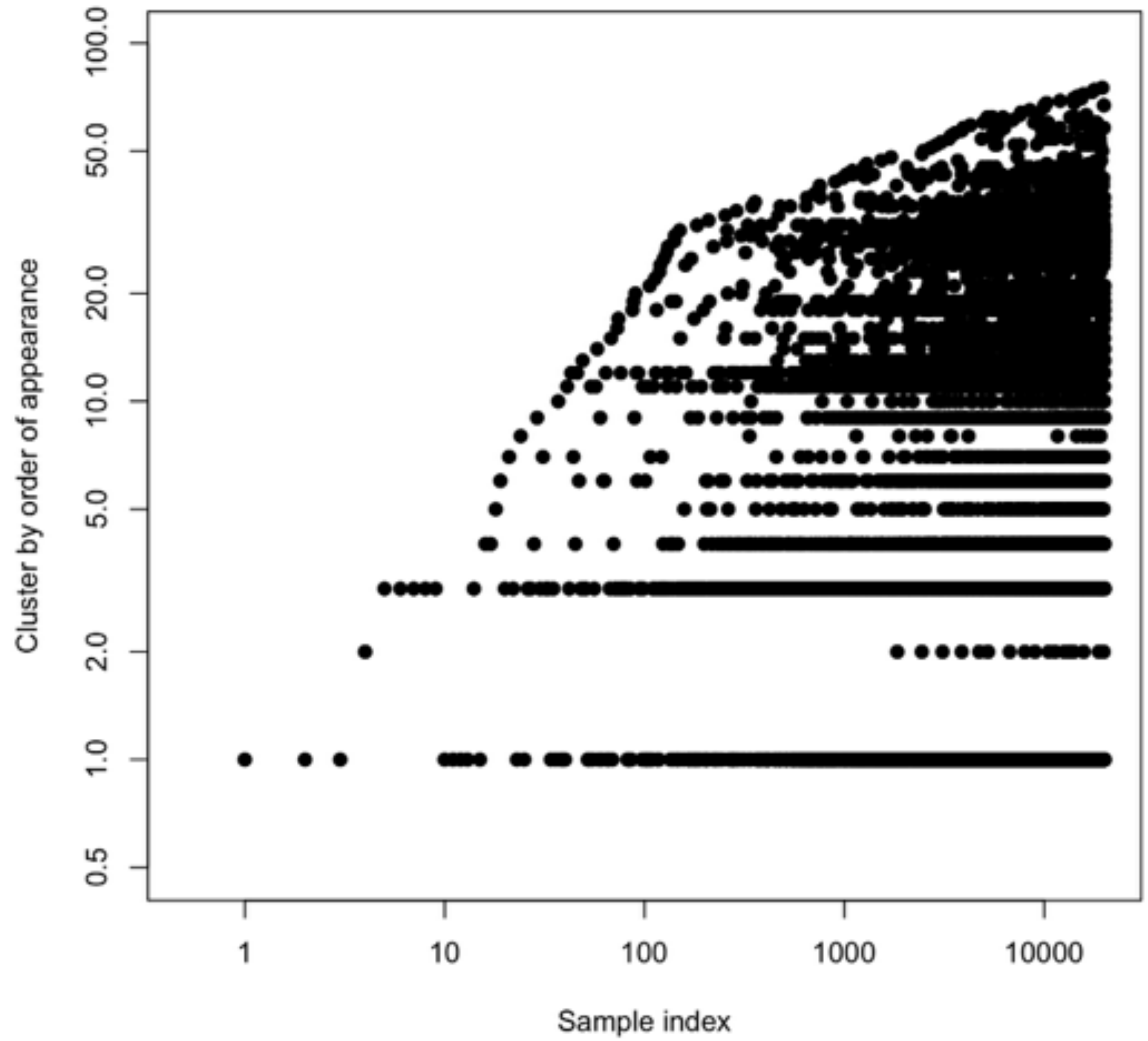
Power laws



Cluster by order of appearance
Sample index
100.0
50.0
20.0
10.0
5.0
2.0
1.0
0.5
1
10
100
1000
10000

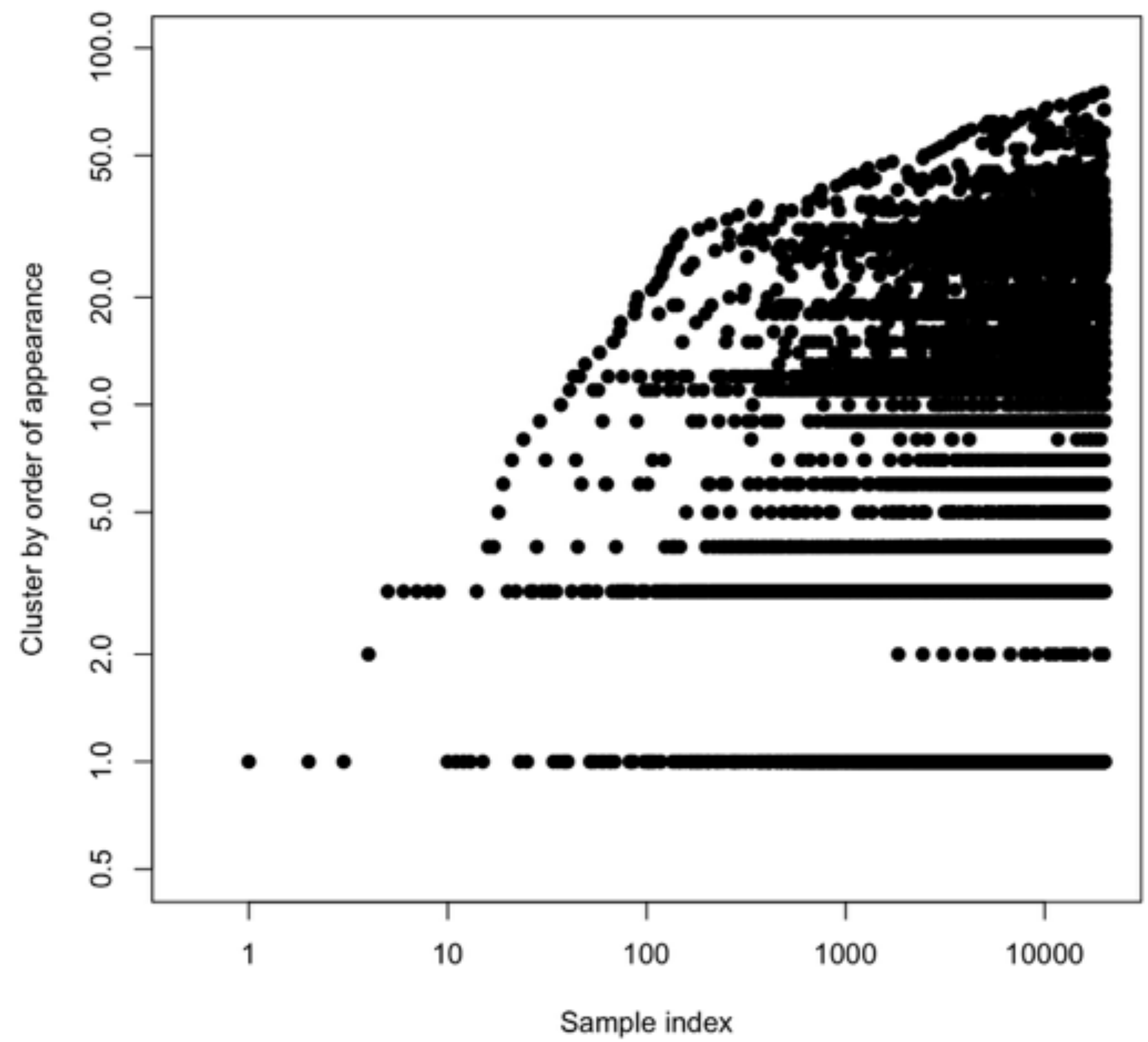
Power laws

- K_N := # clusters occupied by N data points



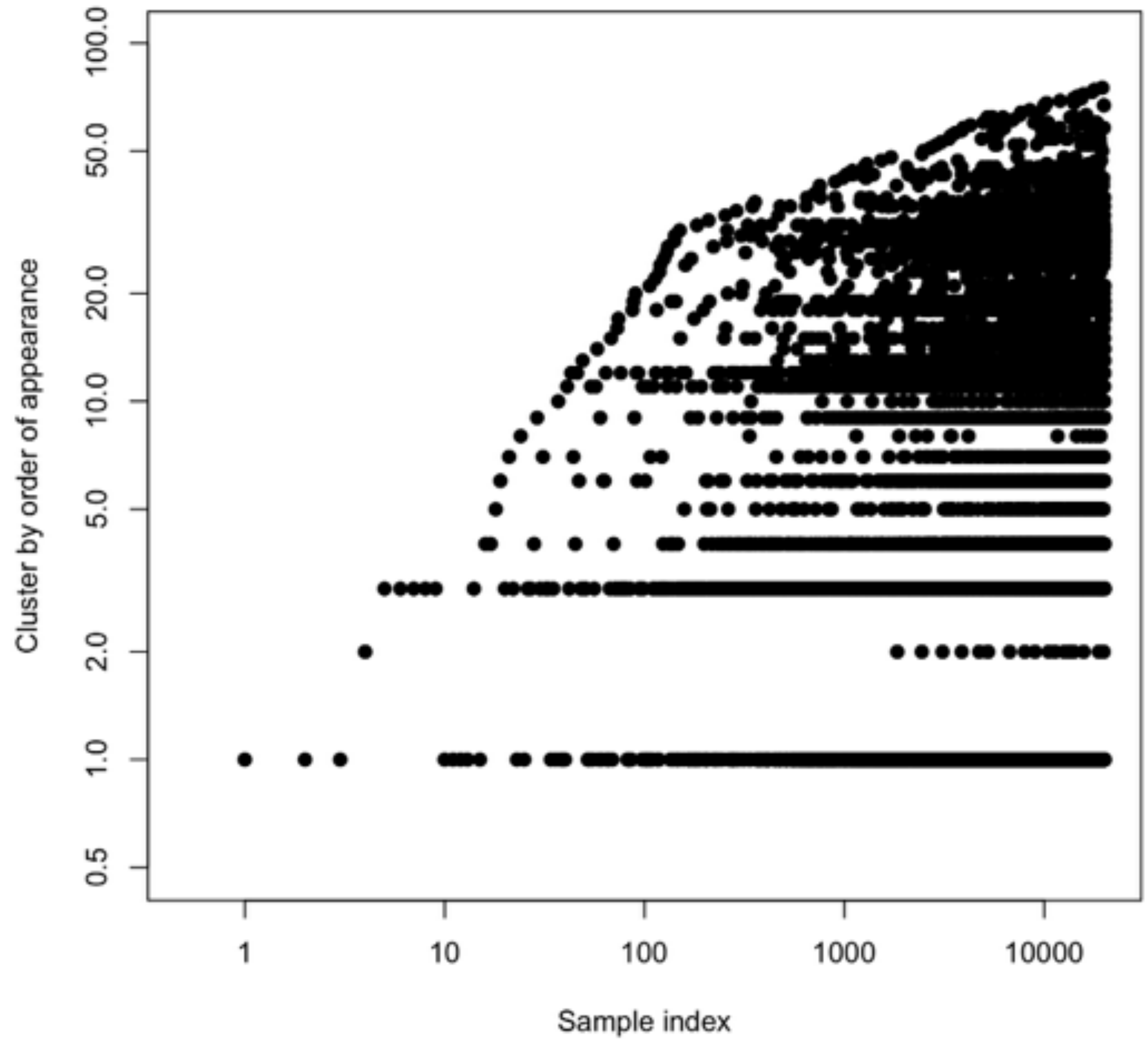
Power laws

- K_N := # clusters occupied by N data points
- CRP: $K_N \sim \alpha \log N \text{ w.p. } 1$



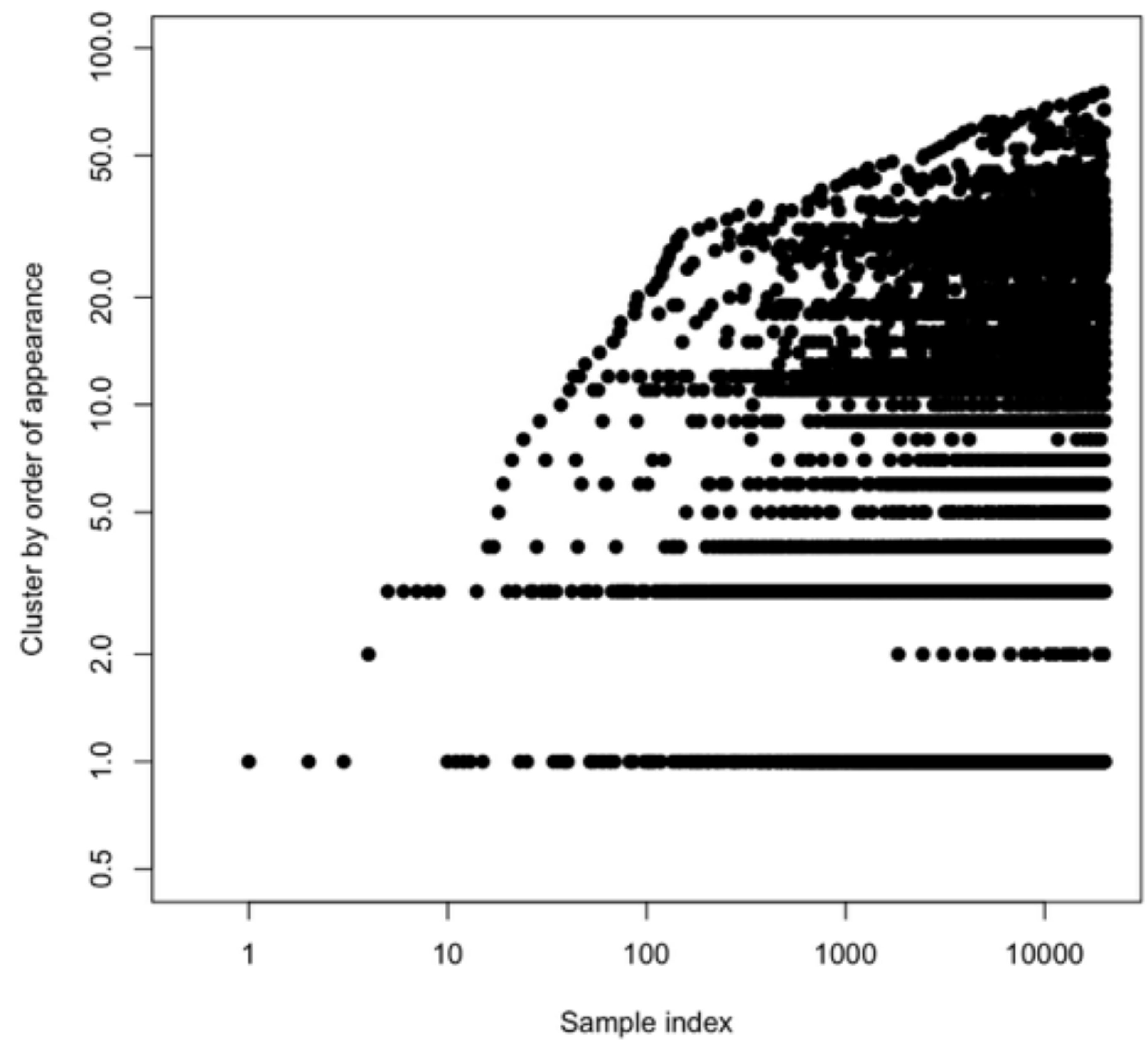
Power laws

- K_N := # clusters occupied by N data points
- CRP: $K_N \sim \alpha \log N \text{ w.p. } 1$
  - vs. Heaps' law, Herdan's law, etc



Power laws

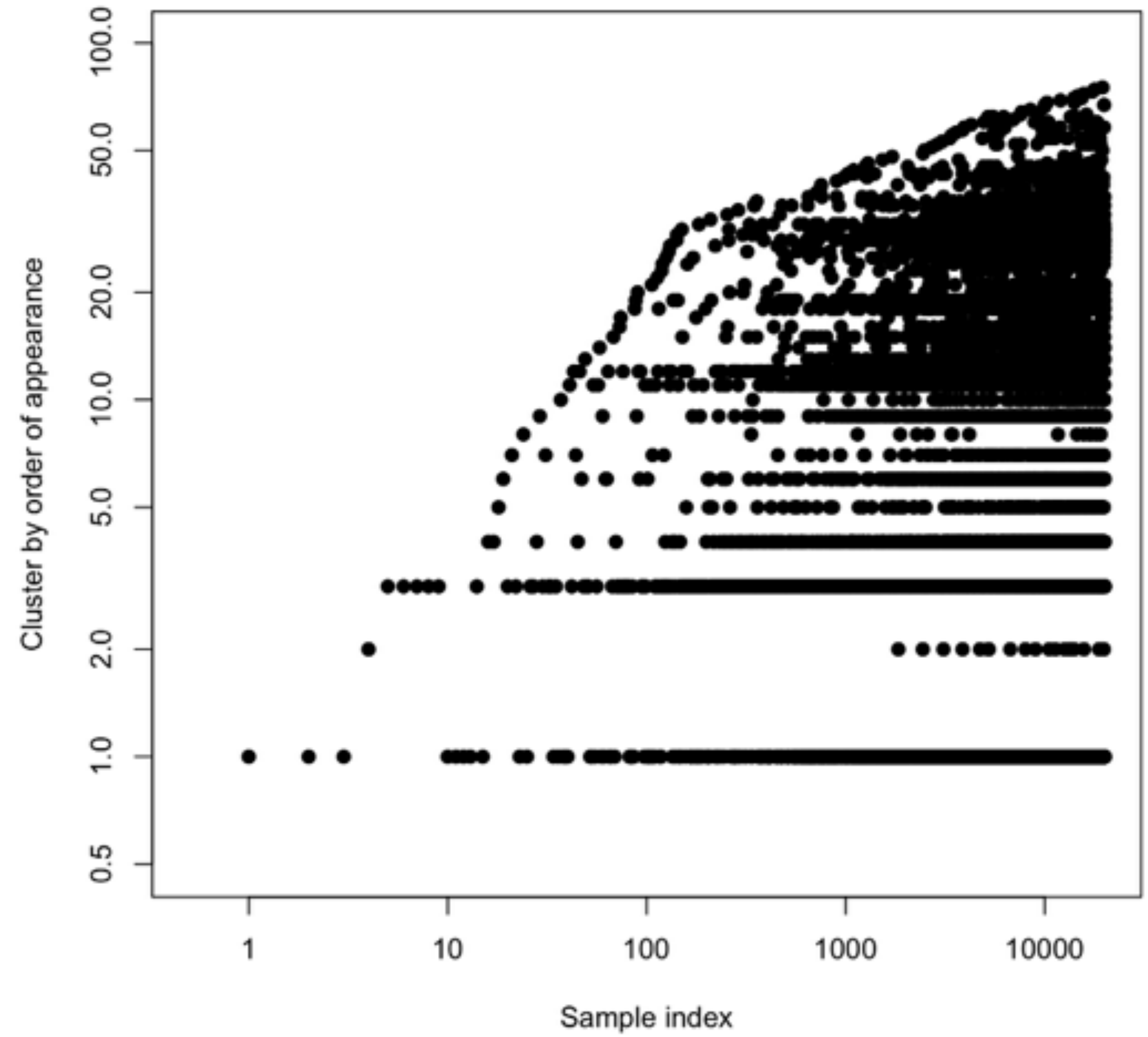
- K_N := # clusters occupied by N data points
- CRP: $K_N \sim \alpha \log N \text{ w.p. } 1$
    - vs. Heaps' law, Herdan's law, etc



[Gnedin, et al 2007

Power laws

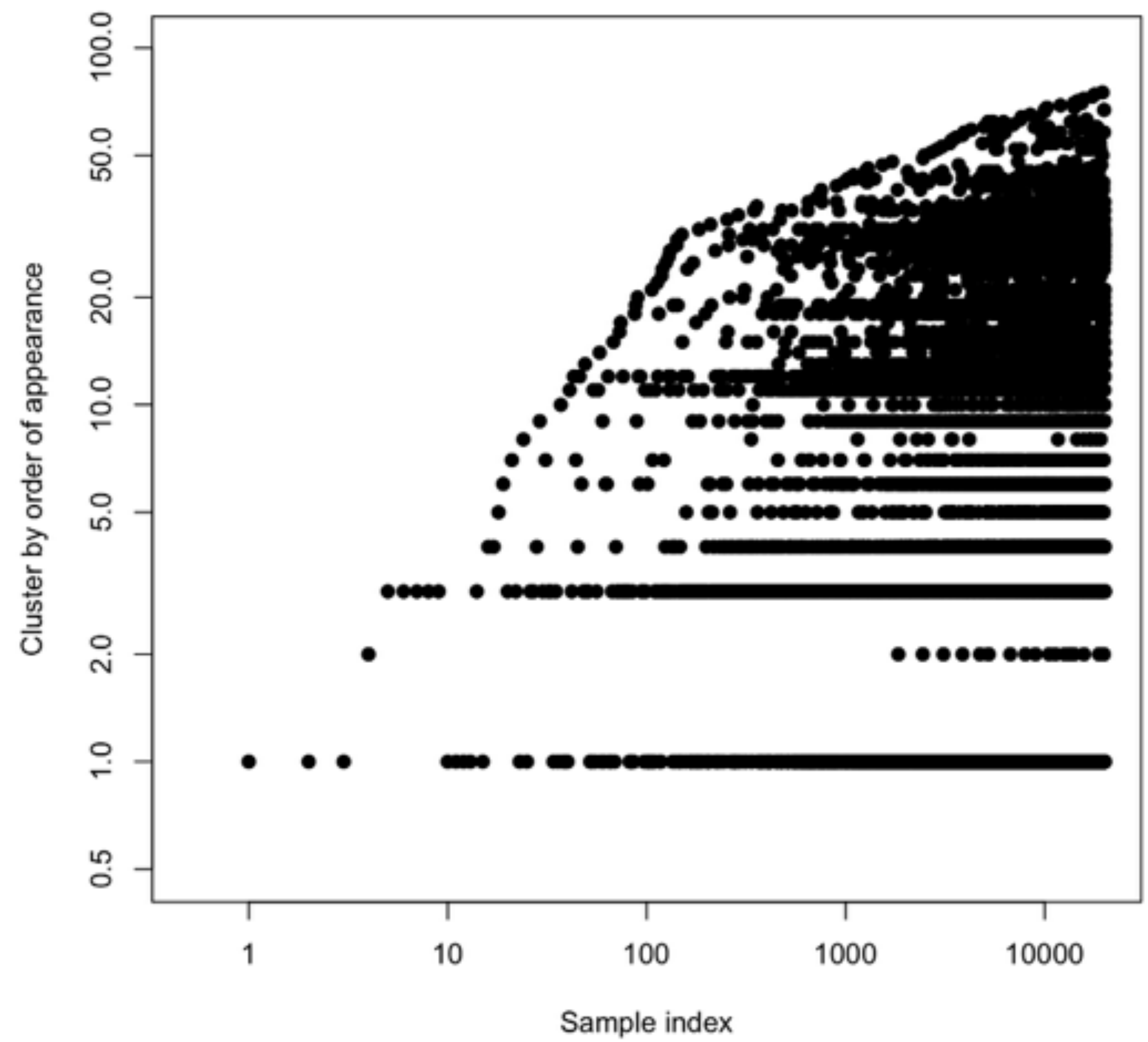
- K_N := # clusters occupied by N data points
- CRP: $K_N \sim \alpha \log N \text{ w.p. } 1$
  - vs. Heaps' law, Herdan's law, etc
- Pitman-Yor process:



[Gnedin, et al 2007

Power laws

- K_N := # clusters occupied by N data points
- CRP: $K_N \sim \alpha \log N \text{ w.p. } 1$
    - vs. Heaps' law, Herdan's law, etc
- Pitman-Yor process:

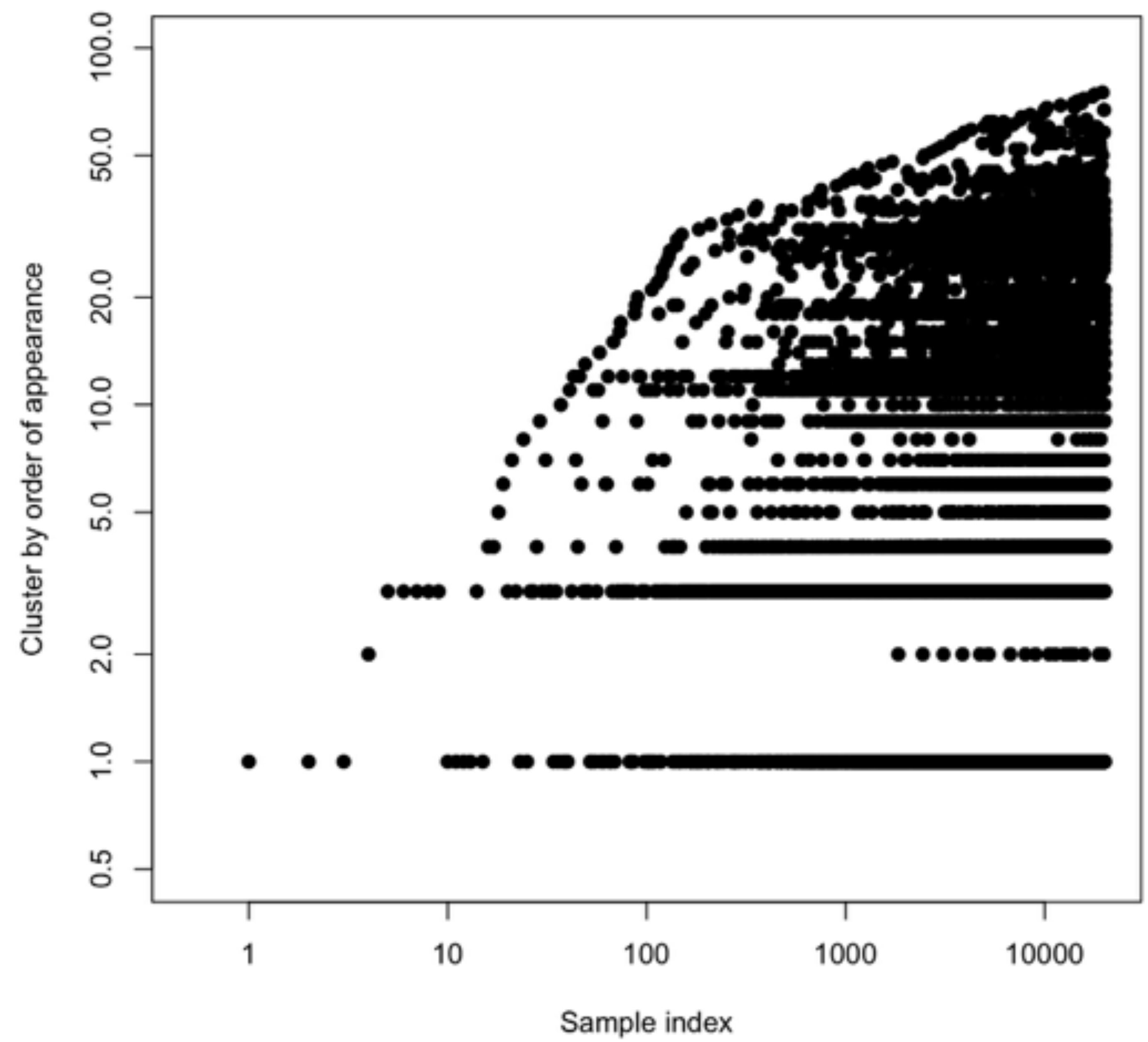


[Gnedin, et al 2007, Pitman, Yor 1997

Power laws

- K_N := # clusters occupied by N data points
- CRP: $K_N \sim \alpha \log N \text{ w.p. } 1$
  - vs. Heaps' law, Herdan's law, etc
- Pitman-Yor process:

  $K_N \sim S_\alpha N^\sigma \text{ w.p. } 1$

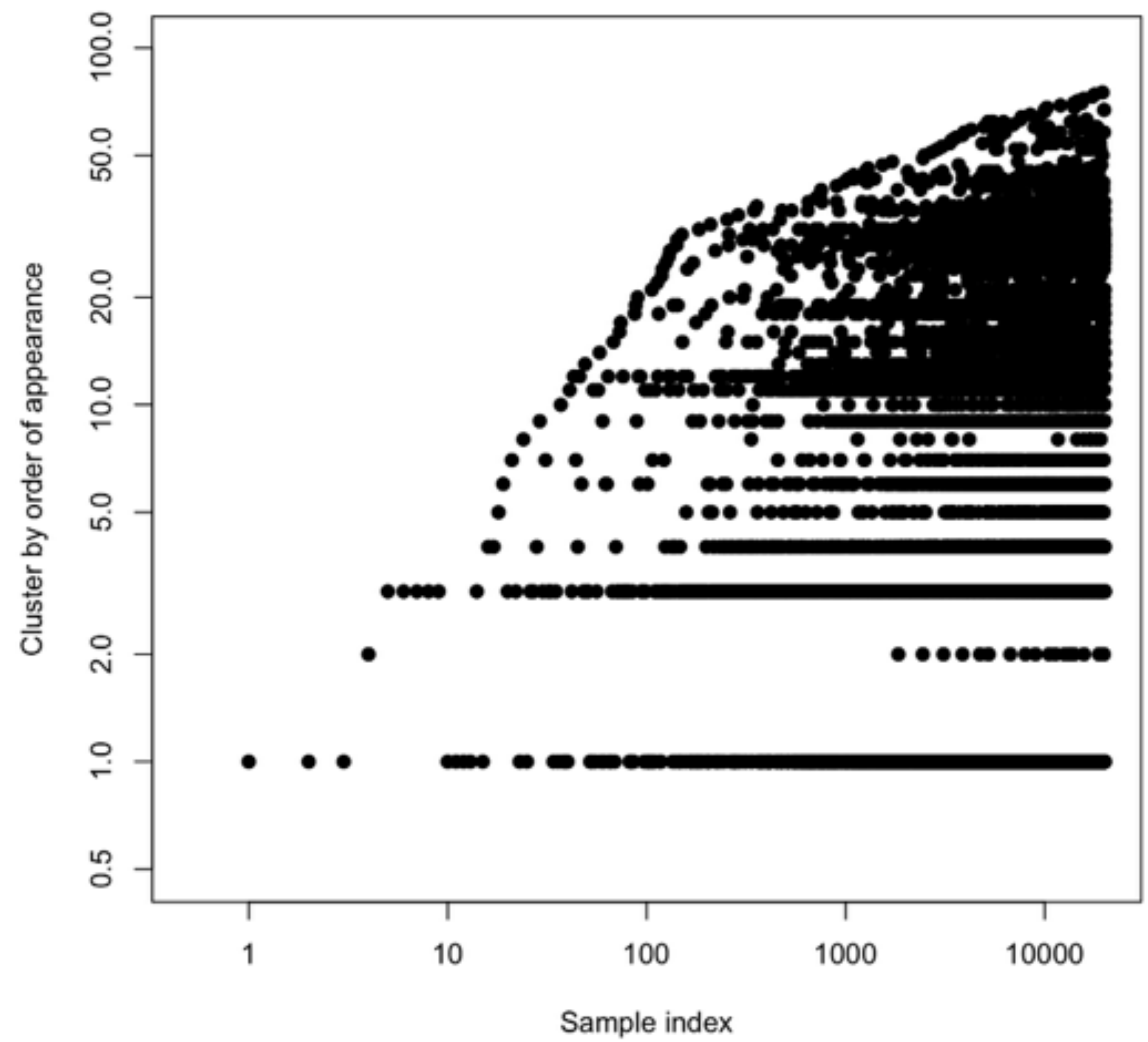


[Gnedin, et al 2007, Pitman, Yor 1997

Power laws

- K_N := # clusters occupied by N data points
- CRP: $K_N \sim \alpha \log N \text{ w.p. } 1$
  - vs. Heaps' law, Herdan's law, etc
- Pitman-Yor process:

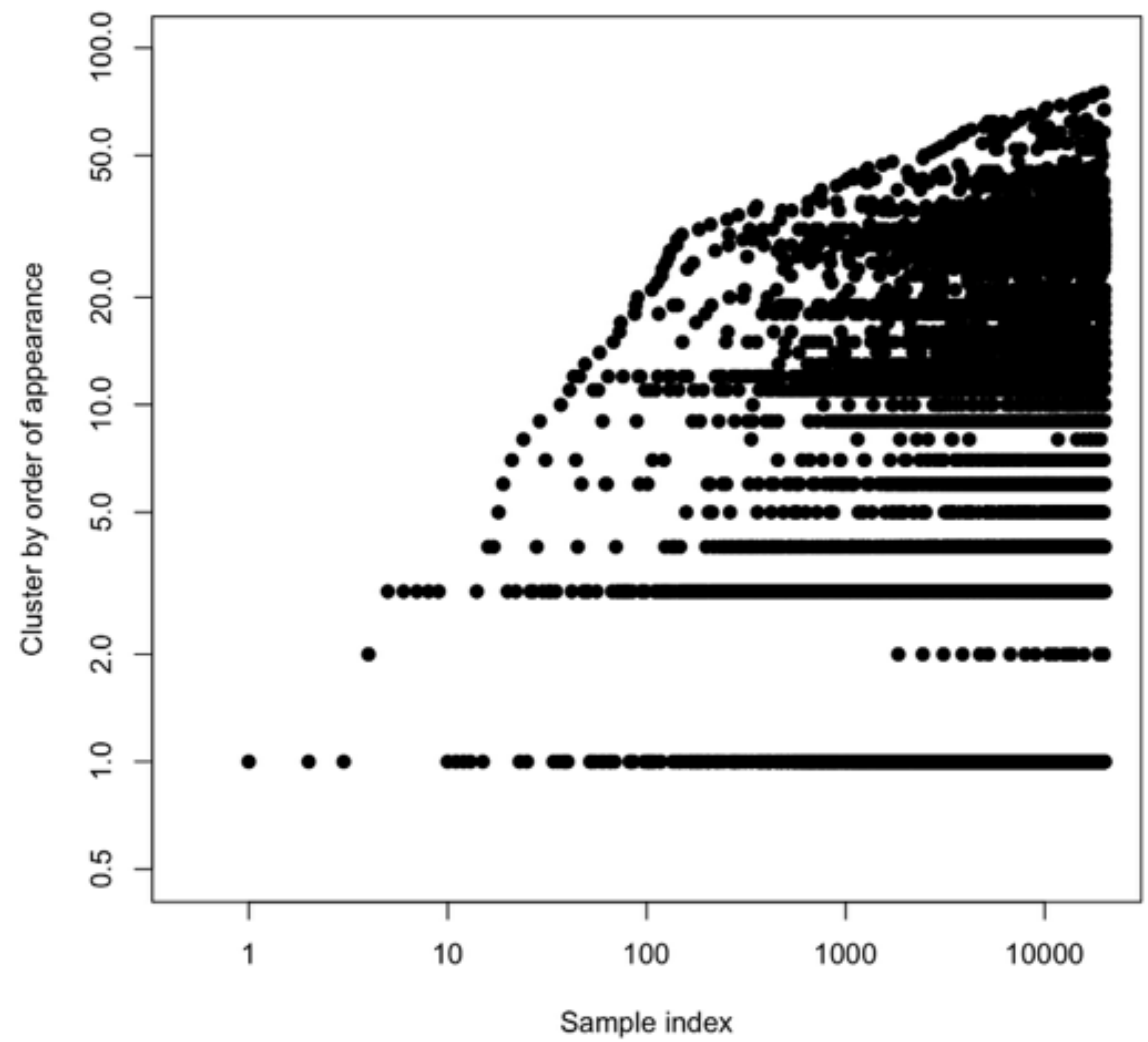
$$K_N \sim S_\alpha N^\sigma \text{ w.p. } 1$$



[Gnedin, et al 2007, Pitman, Yor 1997, Goldwater et al 2005, Teh 2006

Power laws

- K_N := # clusters occupied by N data points
- CRP: $K_N \sim \alpha \log N \text{ w.p. } 1$
  - vs. Heaps' law, Herdan's law, etc
- Pitman-Yor process:
  $K_N \sim S_\alpha N^\sigma \text{ w.p. } 1$
  - related to Zipf's law (ranked frequencies)



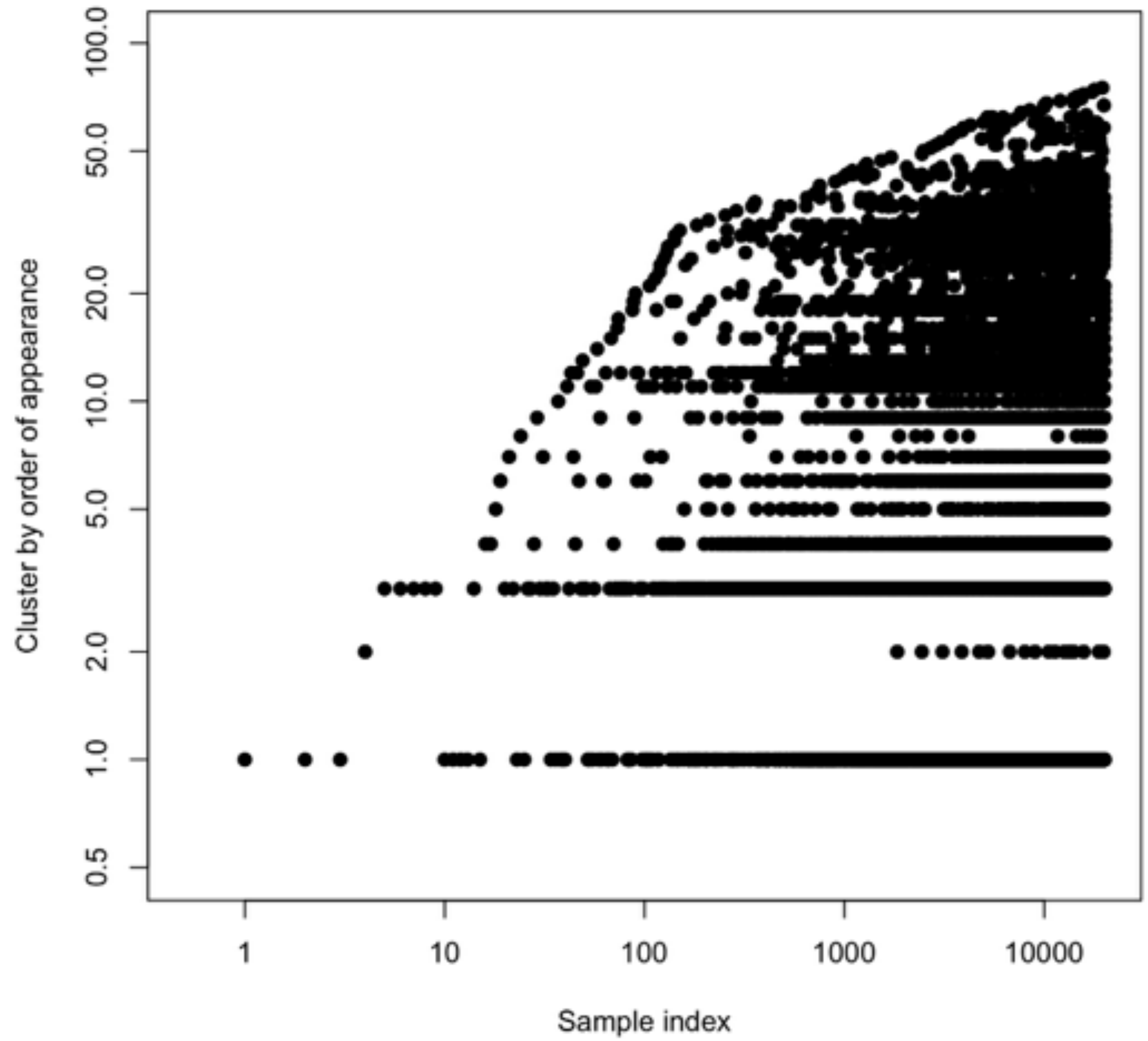
[Gnedin, et al 2007, Pitman, Yor 1997, Goldwater et al 2005, Teh 2006

Power laws

- K_N := # clusters occupied by N data points
- CRP: $K_N \sim \alpha \log N \text{ w.p. } 1$
  - vs. Heaps' law, Herdan's law, etc
- Pitman-Yor process:

  $$K_N \sim S_\alpha N^\sigma \text{ w.p. } 1$$

  - related to Zipf's law (ranked frequencies)
- Not just clusters



[Gnedin, et al 2007, Pitman, Yor 1997, Goldwater et al 2005, Teh 2006, Broderick et al 2012]

Hierarchies

Hierarchies

- Hierarchical Dirichlet process

Hierarchies

- Hierarchical Dirichlet process

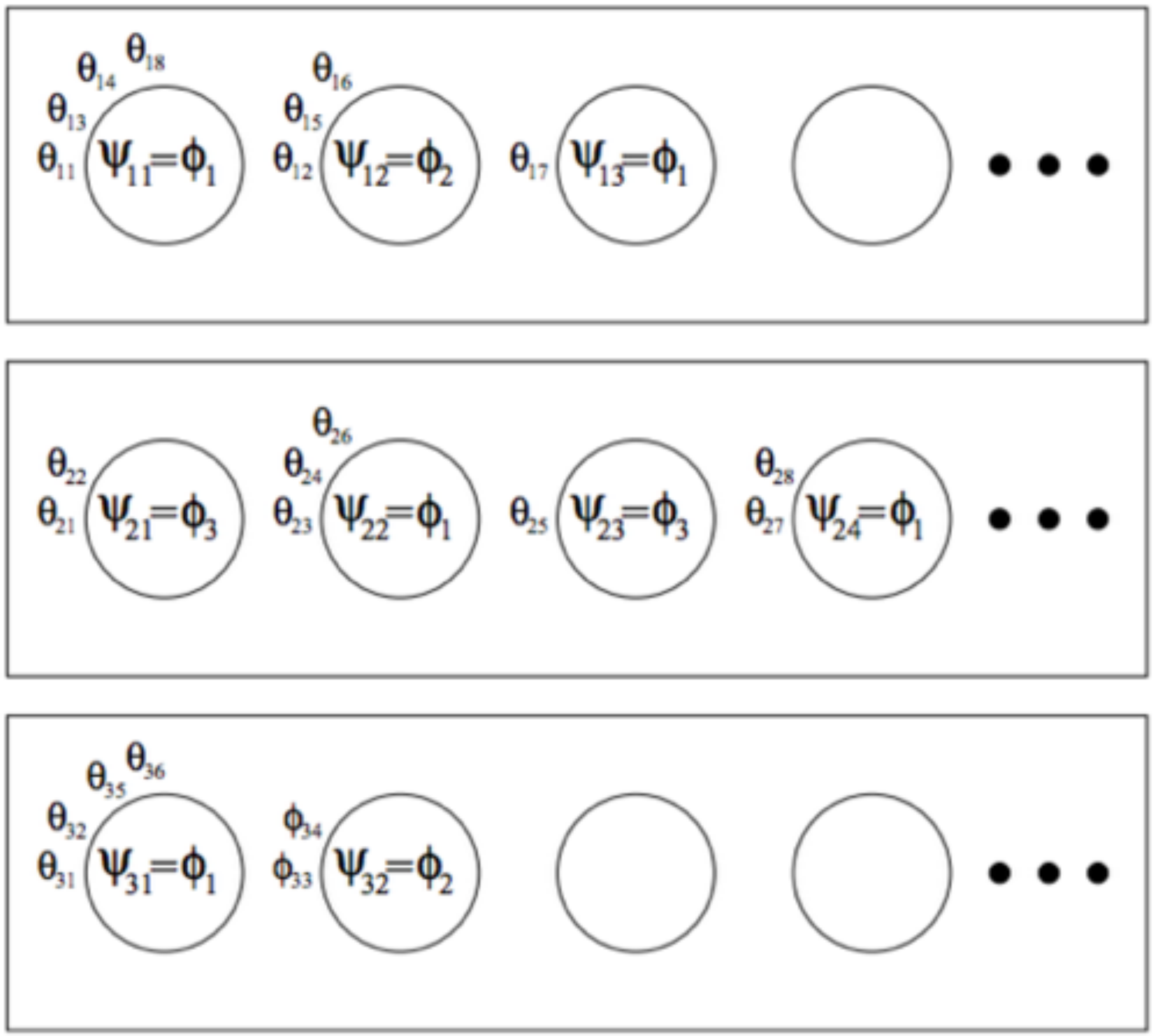
[Teh et al 2006, Rodríguez et al 2008

Hierarchies

- Hierarchical Dirichlet process
- Chinese restaurant franchise

[Teh et al 2006, Rodríguez et al 2008

Hierarchies

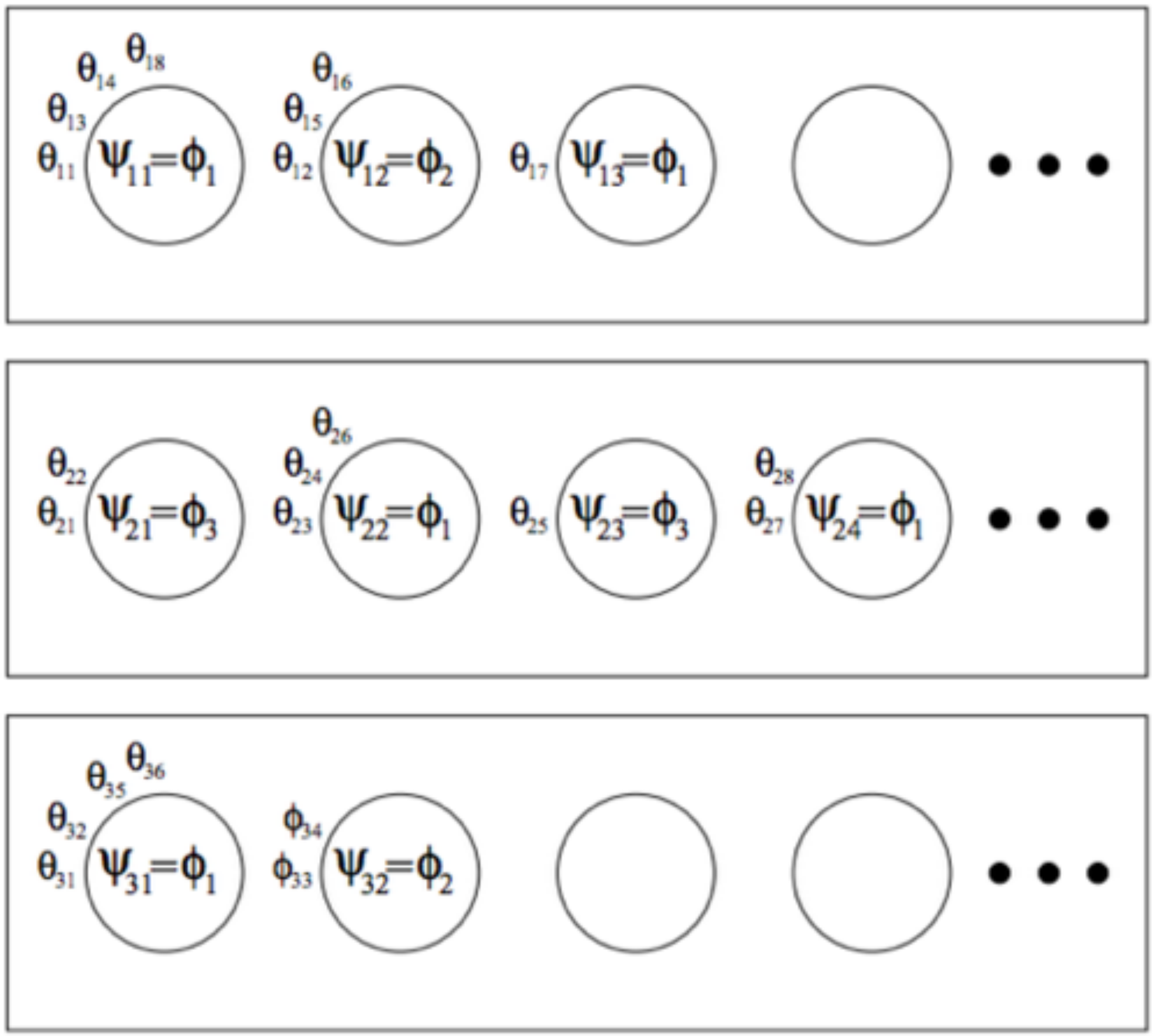


[Teh et al 2006]

- Hierarchical Dirichlet process
- Chinese restaurant franchise

[Teh et al 2006, Rodríguez et al 2008

Hierarchies

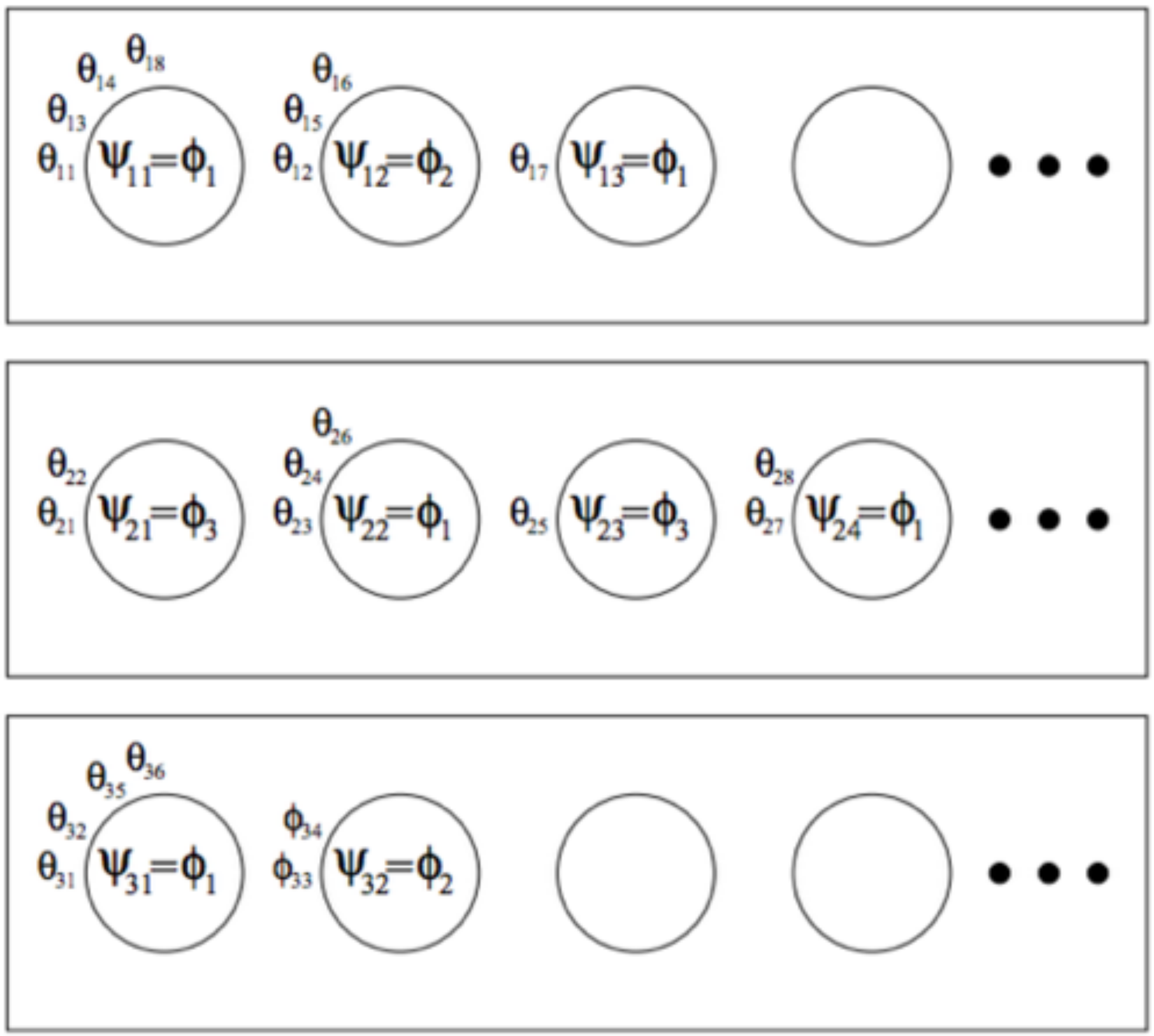


[Teh et al 2006]

- Hierarchical Dirichlet process
- Chinese restaurant franchise
- Hierarchical beta process

[Teh et al 2006, Rodríguez et al 2008

Hierarchies



[Teh et al 2006]

- Hierarchical Dirichlet process
- Chinese restaurant franchise
- Hierarchical beta process

[Teh et al 2006, Rodríguez et al 2008, Thibaux, Jordan 2007]

Nonparametric Bayes

Nonparametric Bayes

- Bayesian statistics that is not parametric

Nonparametric Bayes

- Bayesian statistics that is not parametric
- Bayesian

Nonparametric Bayes

- Bayesian statistics that is not parametric
- Bayesian

$$\mathbb{P}(\text{parameters}|\text{data}) \propto \mathbb{P}(\text{data}|\text{parameters})\mathbb{P}(\text{parameters})$$

Nonparametric Bayes

- Bayesian statistics that is not parametric
- Bayesian

$$\mathbb{P}(\mathrm{parameters}|\mathrm{data}) \propto \mathbb{P}(\mathrm{data}|\mathrm{parameters})\mathbb{P}(\mathrm{parameters})$$

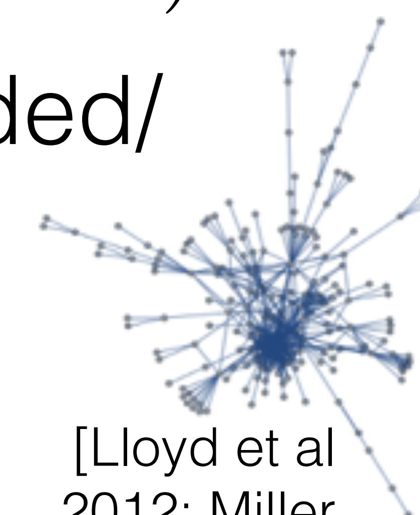
- Not parametric (i.e. not finite parameter, unbounded/ growing/infinite number of parameters)

Nonparametric Bayes

- Bayesian statistics that is not parametric
- Bayesian

$$\mathbb{P}(\text{parameters}|\text{data}) \propto \mathbb{P}(\text{data}|\text{parameters})\mathbb{P}(\text{parameters})$$

- Not parametric (i.e. not finite parameter, unbounded/ growing/infinite number of parameters)

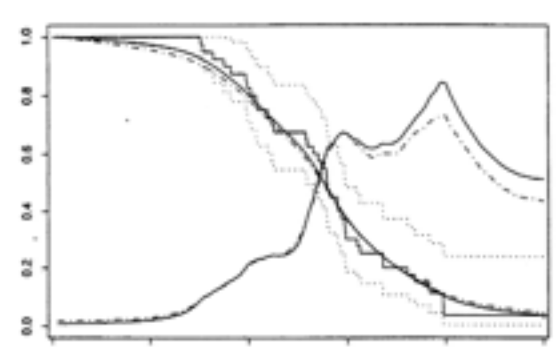


[Lloyd et al 2012; Miller et al, 2010]

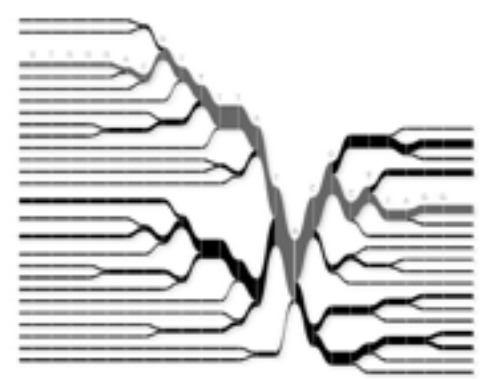
[wikipedia.org]

[Ed Bowlby, NOAA]

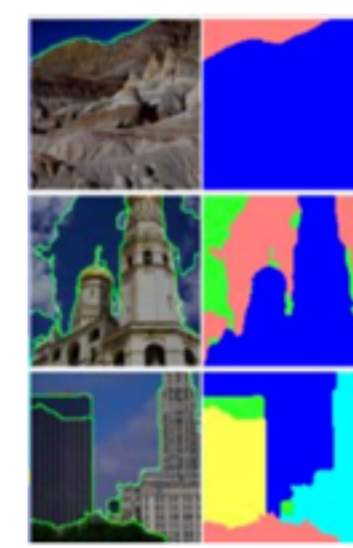
[Fox, et al 2014]



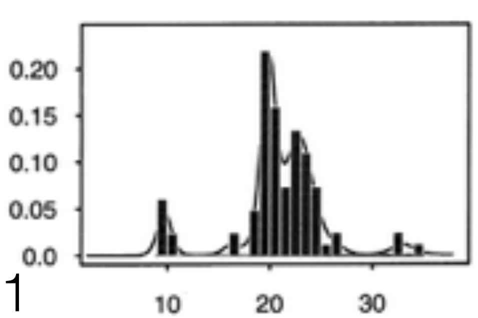
[Arjas, Gasbarra 1994]



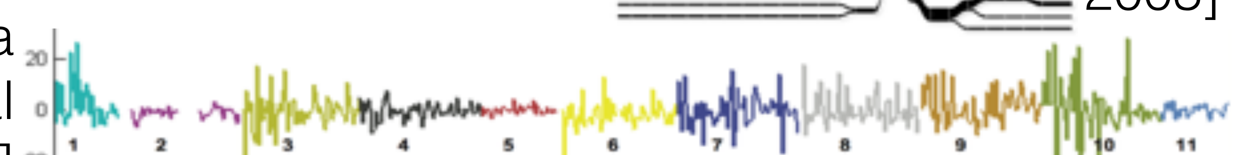
[Ewens, 1972; Hartl, Clark 2003]



[Sudderth, Jordan 2009]



[Escobar, West 1995; Ghosal, et al 1999]



[Saria et al 2010]

References (page 1 of 5)

EM Airoldi, DM Blei, SE Fienberg, & EP Xing. Mixed membership stochastic blockmodels. *Journal of Machine Learning Research*, 2008.

DJ Aldous. Representations for partially exchangeable arrays of random variables. *Journal of Multivariate Analysis*, 1981.

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.

C Borgs, J Chayes, H Cohn, and N Holden. Sparse exchangeable graphs and their limits via graphon processes. arXiv 1601.07134, 2016.

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

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. arXiv preprint arXiv:1410.6843, 2014.

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

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

T Campbell, D Cai, and Broderick T. Exchangeable trait allocations. Submitted. arXiv preprint arXiv:1609.09147.

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

References (page 2 of 5)

T Campbell*, JH Huggins*, J How, T Broderick. Truncation random measures. arXiv preprint arXiv:1603.00861, 2016.

F Caron and E Fox. Sparse graphs using exchangeable random measures. arXiv 1401.1137, 2014.

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. Relational exchangeability. arXiv 1607.06762, 2016.

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.

EB Fox, personal website. Retrieved 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.

References (page 3 of 5)

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

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

T Herlau and M Schmidt. Completely random measures for modelling block-structured sparse networks. NIPS, 2016.

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.

PW Holland, KB Laskey, and S Leinhardt. Stochastic blockmodels: first steps. *Social Networks*, 1983.

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. Poisson latent feature calculus for generalized Indian buffet processes. arXiv preprint arXiv:1411.2936, 2014.

C Kemp, JB Tenenbaum, TL Griffiths, T Yamada, and N Ueda. Learning systems of concepts with an infinite relational model. *AAAI*, 2006.

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

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.

References (page 4 of 5)

JFC Kingman. *Poisson processes*, 1992.

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

K Palla, F Caron, and YW Teh. Bayesian nonparametrics for sparse dynamic networks. arXiv 1607.01624, 2016.

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

J Pitman. Exchangeable and partially exchangeable random partitions. *Probability Theory and Related Fields*, 1995.

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

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. *NIPS*, 2010.

References (page 5 of 5)

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. *NIPS*, 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. *NIPS*, 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.

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

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

Z Xu, V Tresp, S Yu, K Yu, and H Kriegel. Fast inference in infinite hidden relational models. *Proceedings of Mining and Learning with Graphs*, 2007.