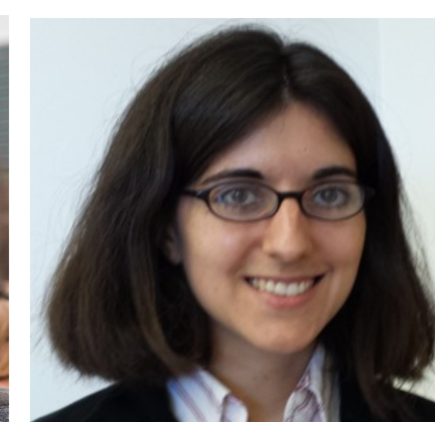
Paintboxes and probability functions for edge-exchangeable graphs

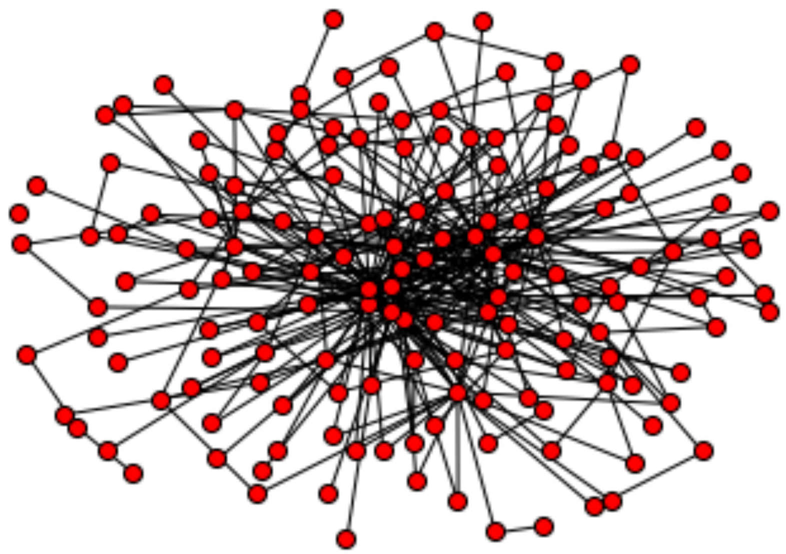
Diana Cai

Department of Statistics
University of Chicago

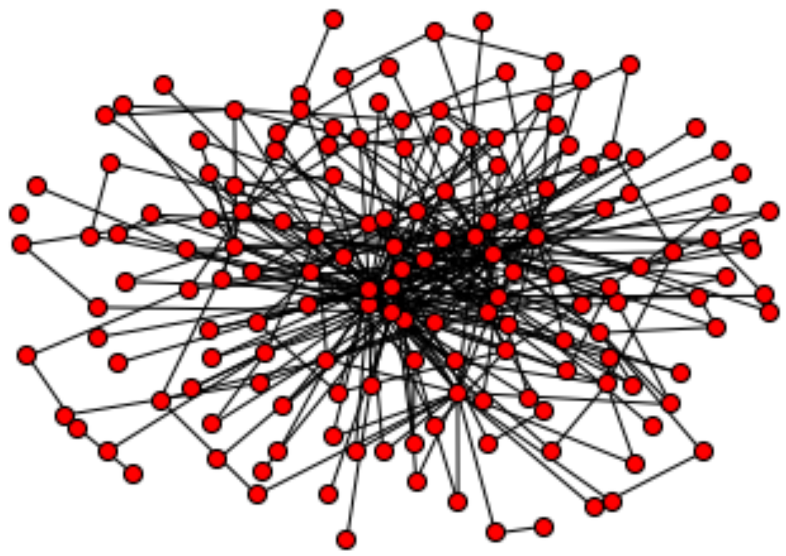
Joint work with:
Trevor Campbell and Tamara Broderick
(MIT CSAIL)



Network data (graphs): interactions between individuals

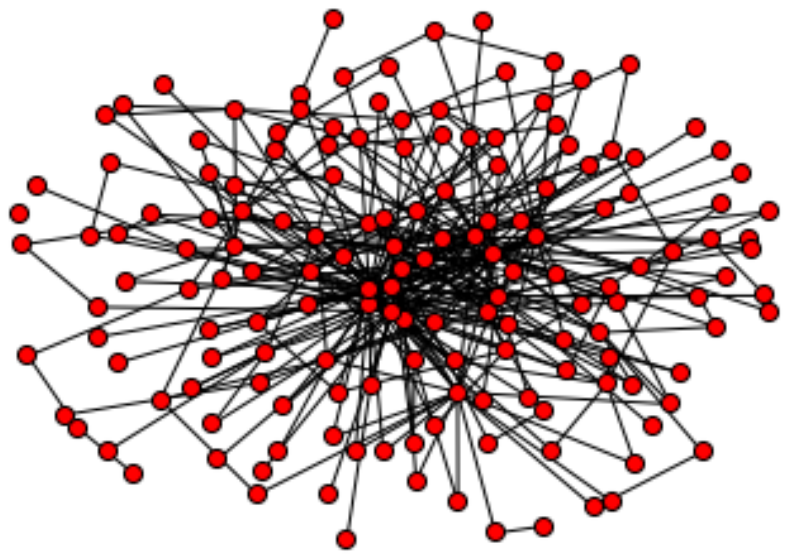


Network data (graphs): interactions between individuals



social: *Facebook, Twitter, email*

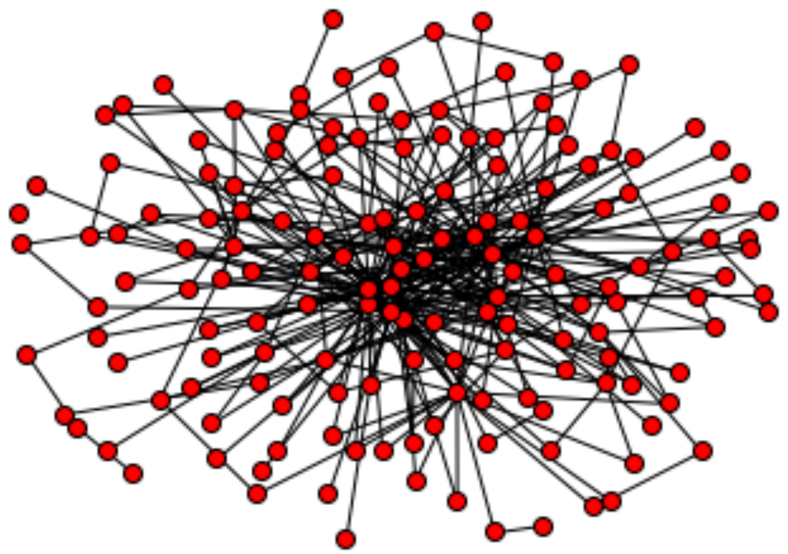
Network data (graphs): interactions between individuals



social: *Facebook, Twitter, email*

biological: *ecological, protein, gene*

Network data (graphs): interactions between individuals



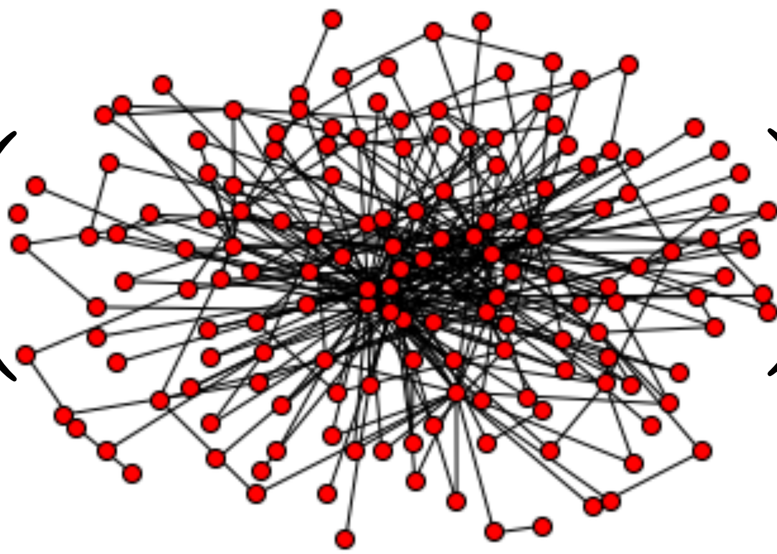
social: *Facebook, Twitter, email*

biological: *ecological, protein, gene*

transportation: *roads, railways*

Network data (graphs): interactions between individuals

Probabilistic models for graphs

$p($  $)$

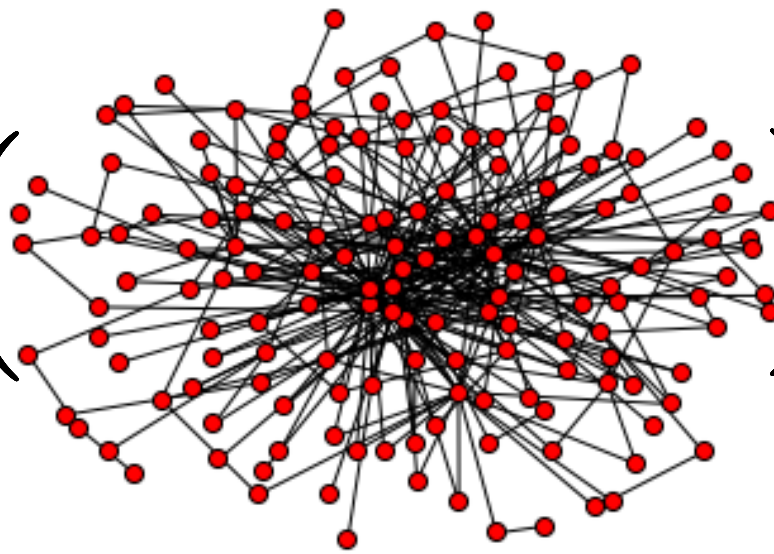
social: *Facebook, Twitter, email*

biological: *ecological, protein, gene*

transportation: *roads, railways*

Network data (graphs): interactions between individuals

Probabilistic models for graphs

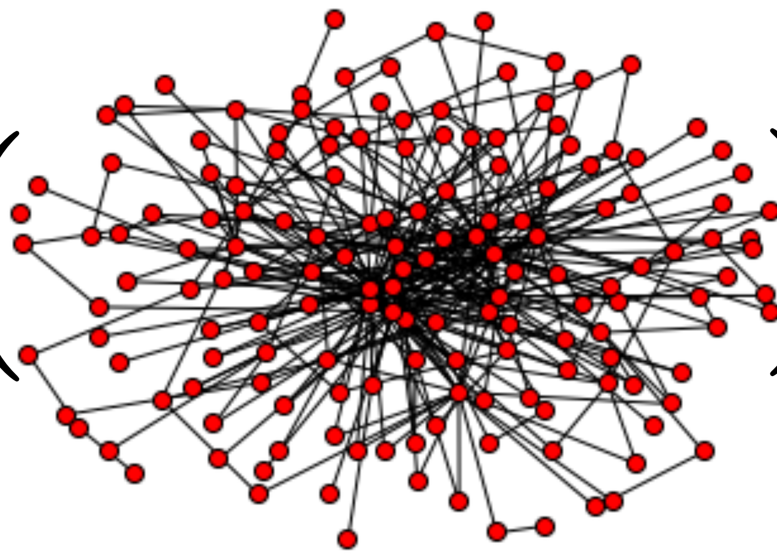
$p($  $)$

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

Many probabilistic models assume *vertex exchangeability*: dense (too many edges).

Network data (graphs): interactions between individuals

Probabilistic models for graphs

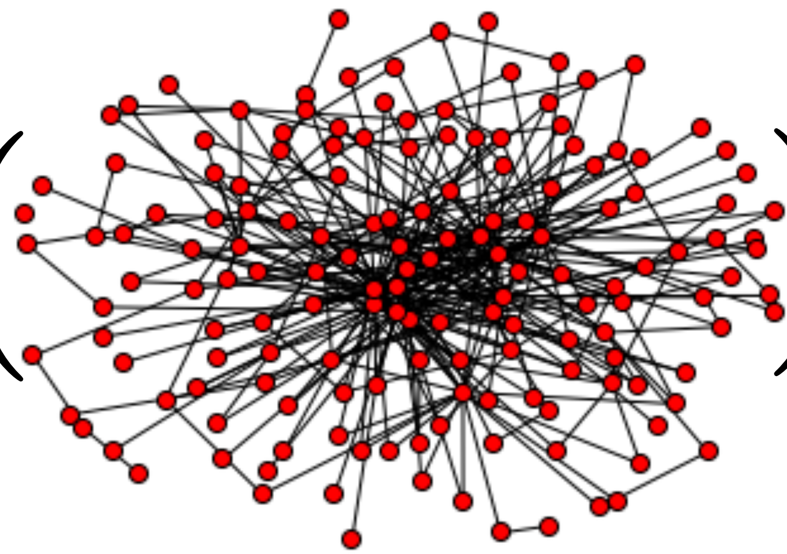
$p($  $)$

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

Many probabilistic models assume *vertex exchangeability*: dense (too many edges). Real-world graphs are *sparse*.

Network data (graphs): interactions between individuals

Probabilistic models for graphs

$p($  $)$

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

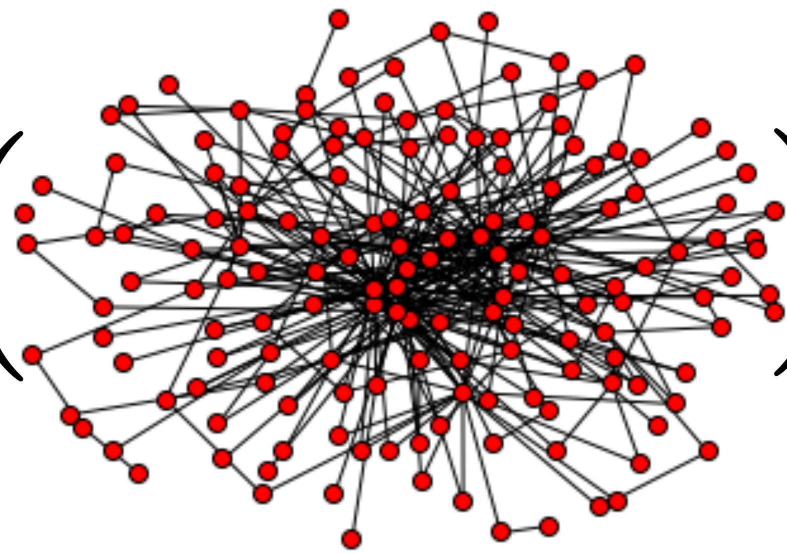
Many probabilistic models assume *vertex exchangeability*: dense (too many edges). Real-world graphs are *sparse*.

Under *edge exchangeability*, we can get sparse graphs.

[Broderick & Cai, 2015; Cai, Campbell, Broderick, 2016; Crane & Dempsey, 2015, 2016]

Network data (graphs): interactions between individuals

Probabilistic models for graphs

$p\Big($  $\Big)$

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

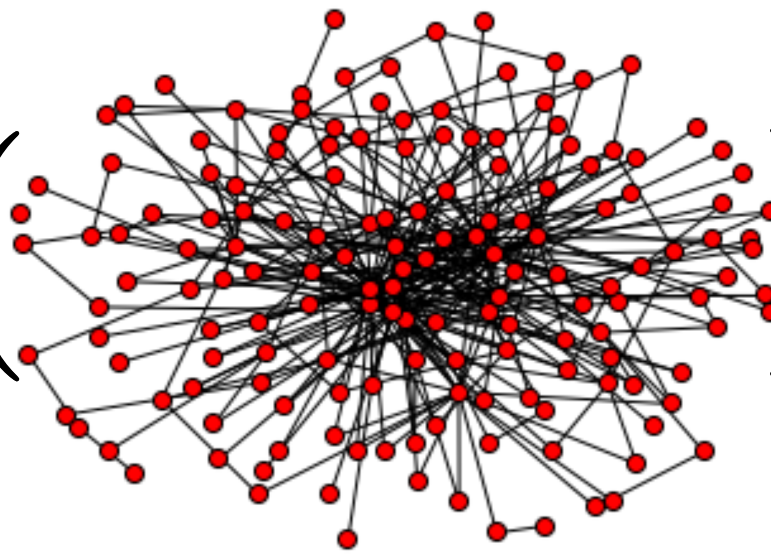
Many probabilistic models assume *vertex exchangeability*: dense (too many edges). Real-world graphs are *sparse*.

Under *edge exchangeability*, we can get sparse graphs. We want a representation theorem that characterizes all edge-exchangeable graphs.

[Broderick & Cai, 2015; Cai, Campbell, Broderick, 2016; Crane & Dempsey, 2015, 2016]

Network data (graphs): interactions between individuals

Probabilistic models for graphs

$p\Big($  $\Big)$

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

Many probabilistic models assume *vertex exchangeability*: dense (too many edges). Real-world graphs are *sparse*.

Under *edge exchangeability*, we can get sparse graphs. We want a representation theorem that characterizes all edge-exchangeable graphs.
We also want to characterize models with easy inference.

[Broderick & Cai, 2015; Cai, Campbell, Broderick, 2016; Crane & Dempsey, 2015, 2016]

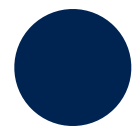
Network models should reflect **real-life scaling properties**

Network models should reflect **real-life scaling properties**

sequence of graphs

Network models should reflect **real-life scaling properties**

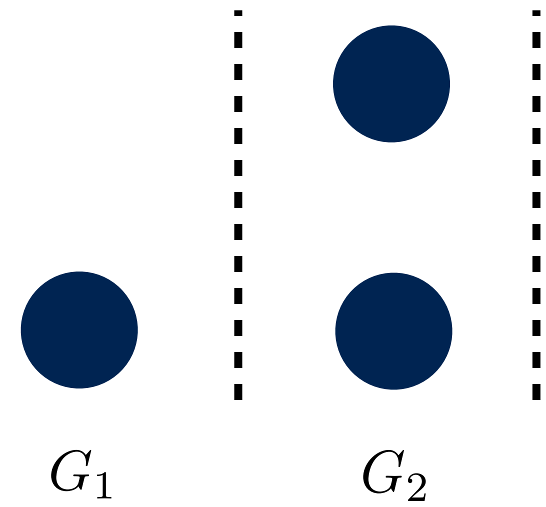
sequence of graphs



G_1

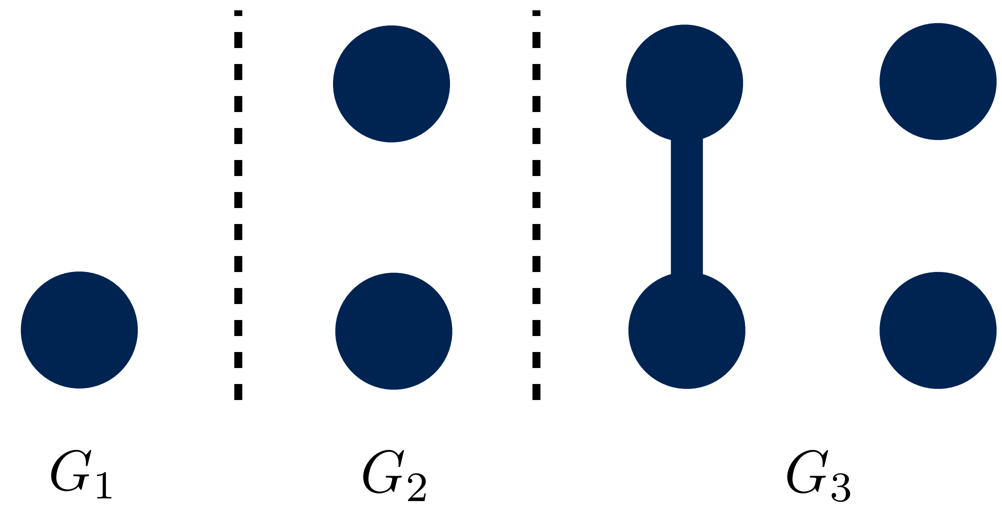
Network models should reflect **real-life scaling properties**

sequence of graphs



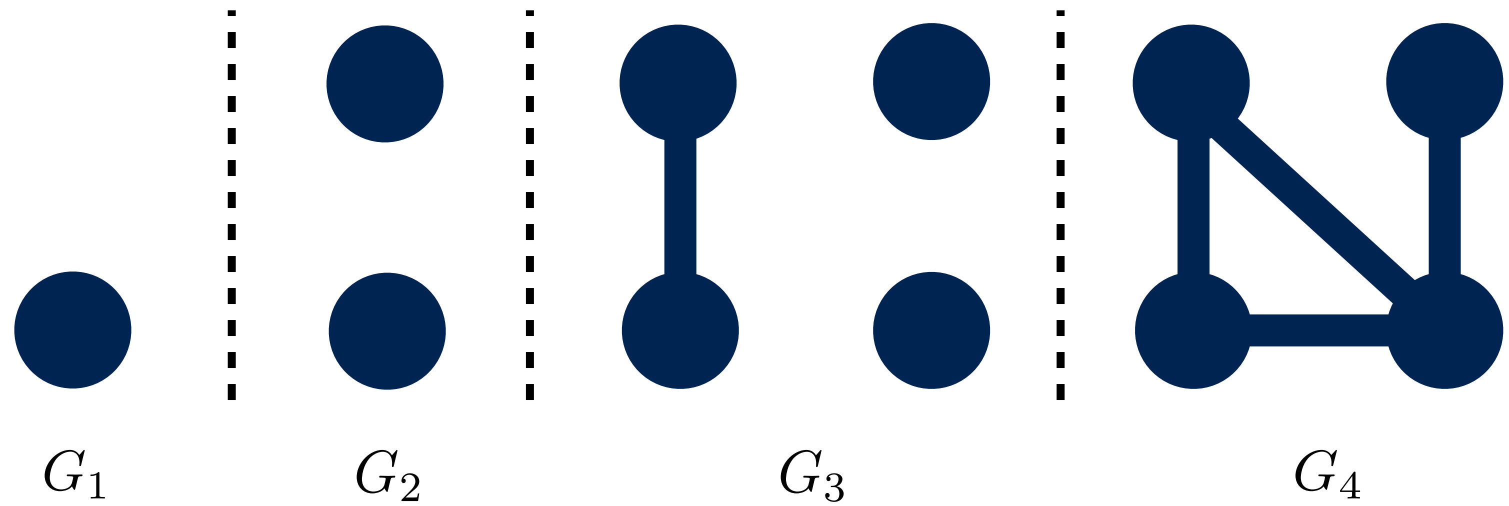
Network models should reflect **real-life scaling properties**

sequence of graphs



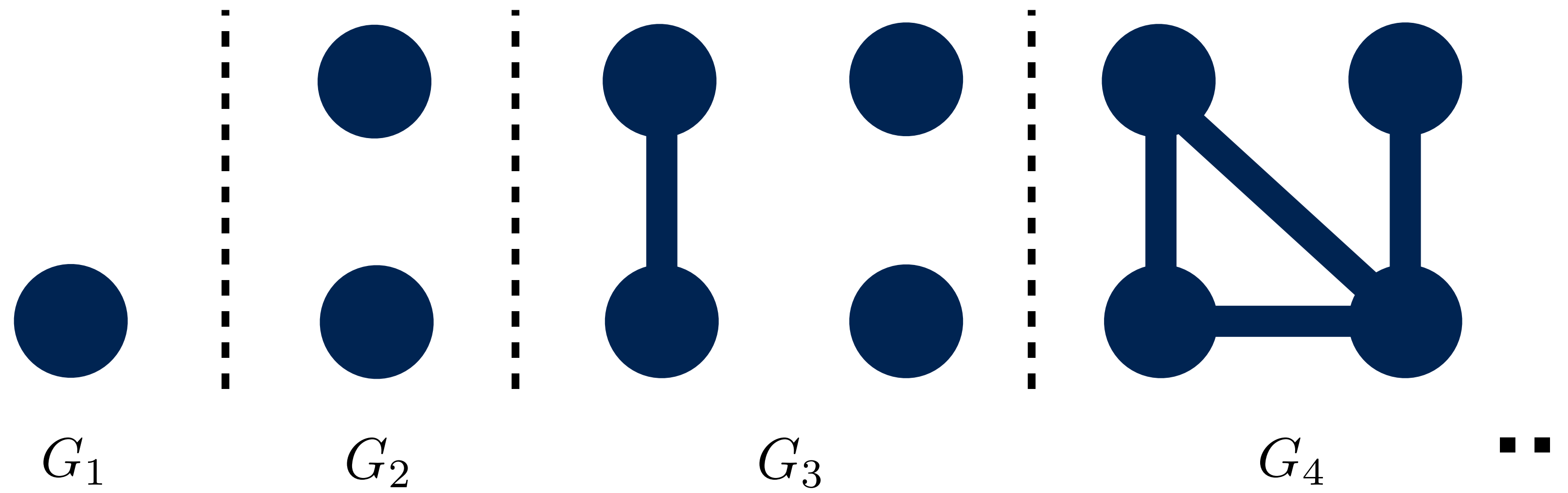
Network models should reflect **real-life scaling properties**

sequence of graphs



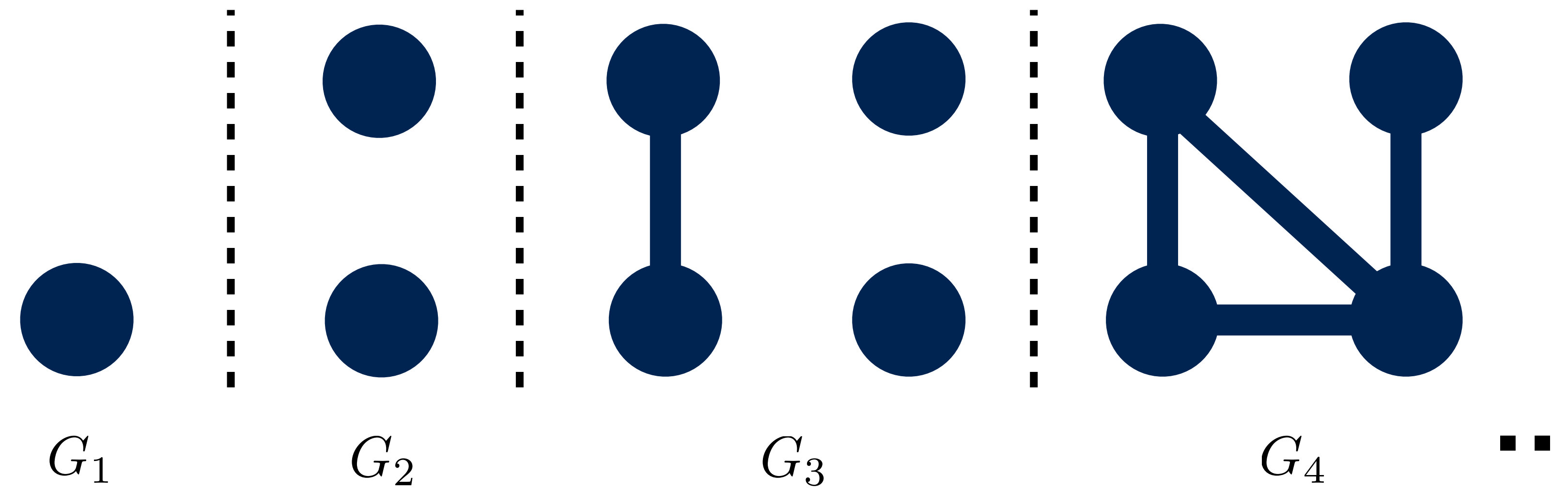
Network models should reflect **real-life scaling properties**

sequence of graphs



Network models should reflect **real-life scaling properties**

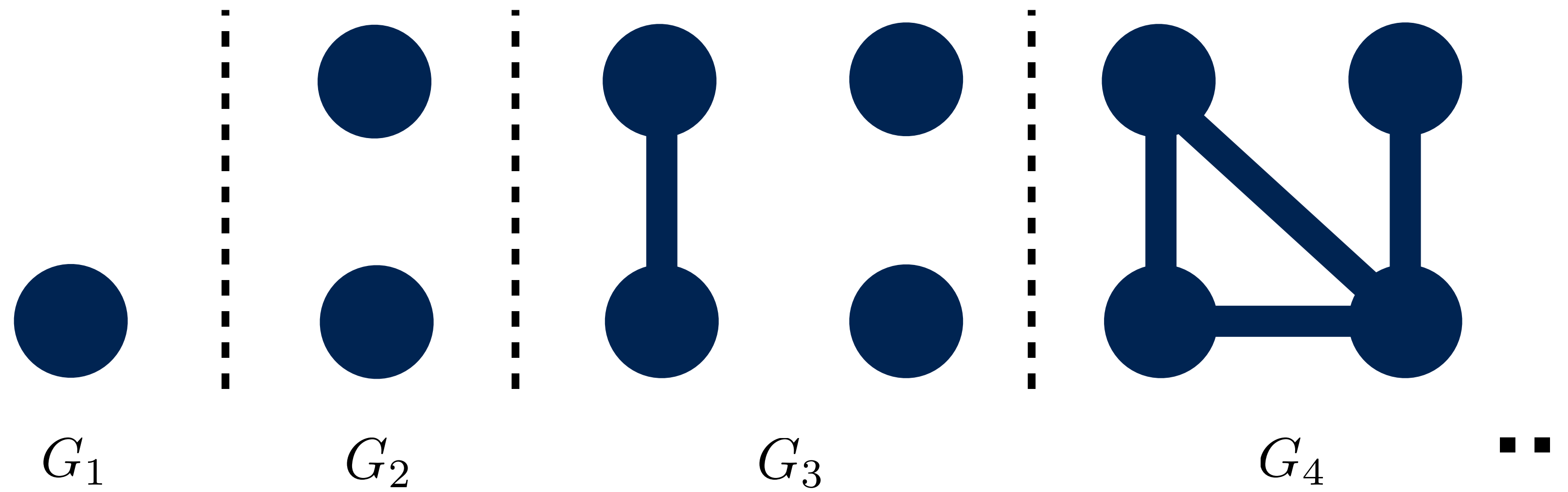
sequence of graphs



real-life scaling properties:

Network models should reflect **real-life scaling properties**

sequence of graphs

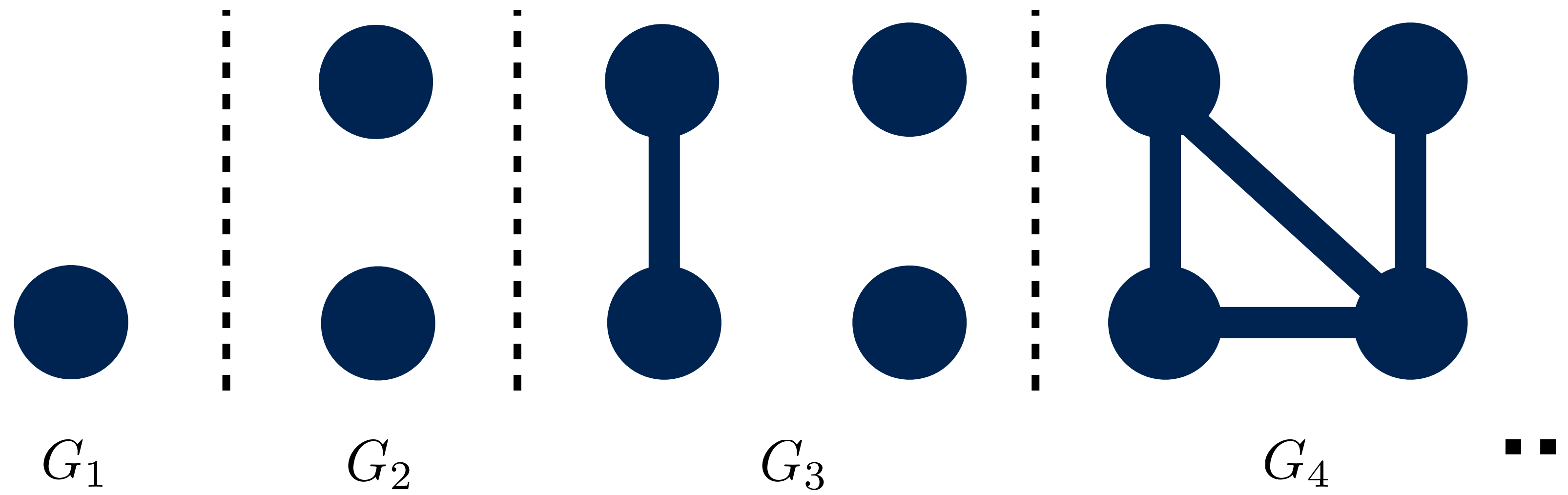


real-life scaling properties:

Sparse:

Network models should reflect **real-life scaling properties**

sequence of graphs

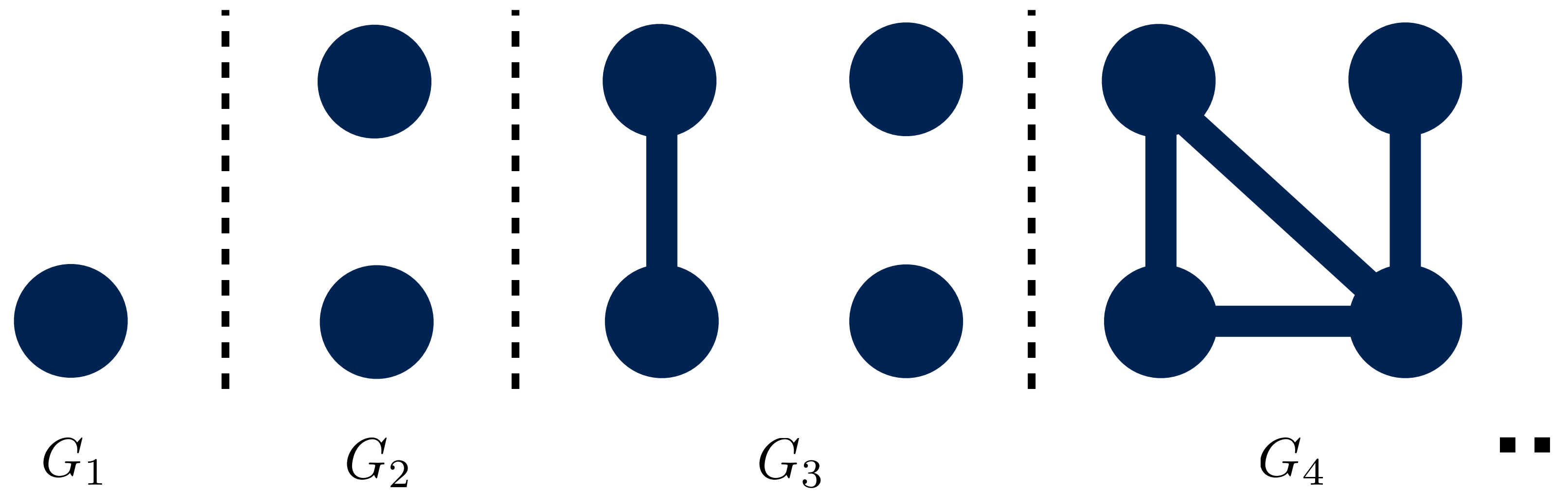


real-life scaling properties:

Sparse: $[\#\mathrm{edges}(G_n)] \in o([\#\mathrm{vertices}(G_n)]^2)$

Network models should reflect **real-life scaling properties**

sequence of graphs

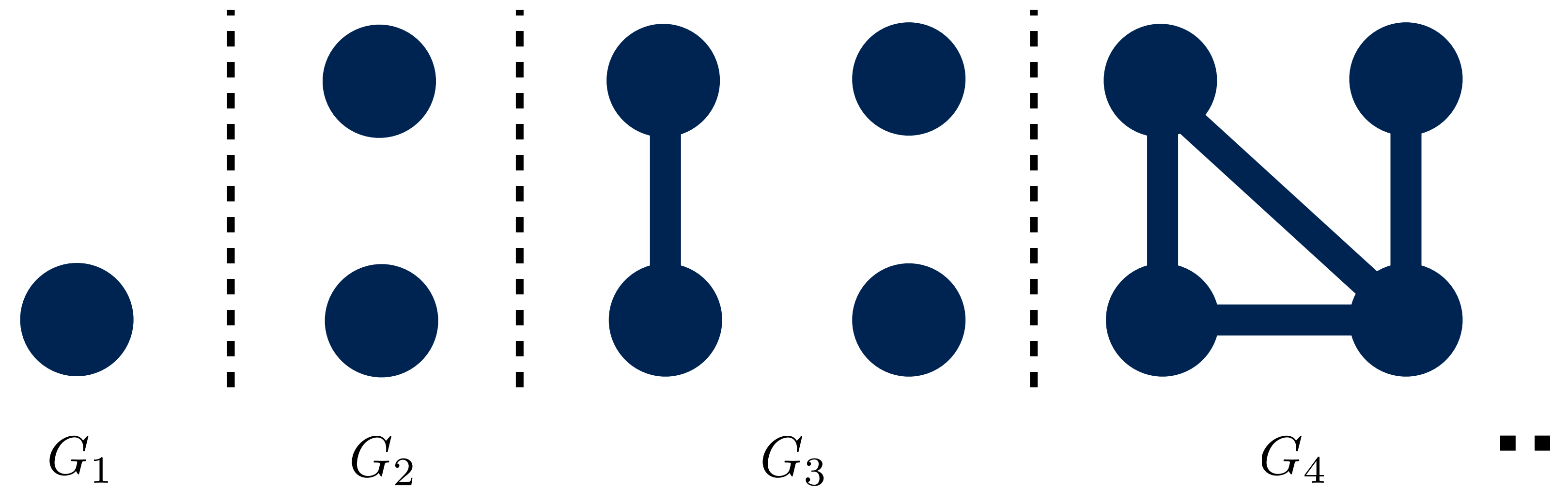


real-life scaling properties:

Sparse: $[\#\mathrm{edges}(G_n)] \in o([\#\mathrm{vertices}(G_n)]^2)$ *sub-quadratic*

Network models should reflect **real-life scaling properties**

sequence of graphs



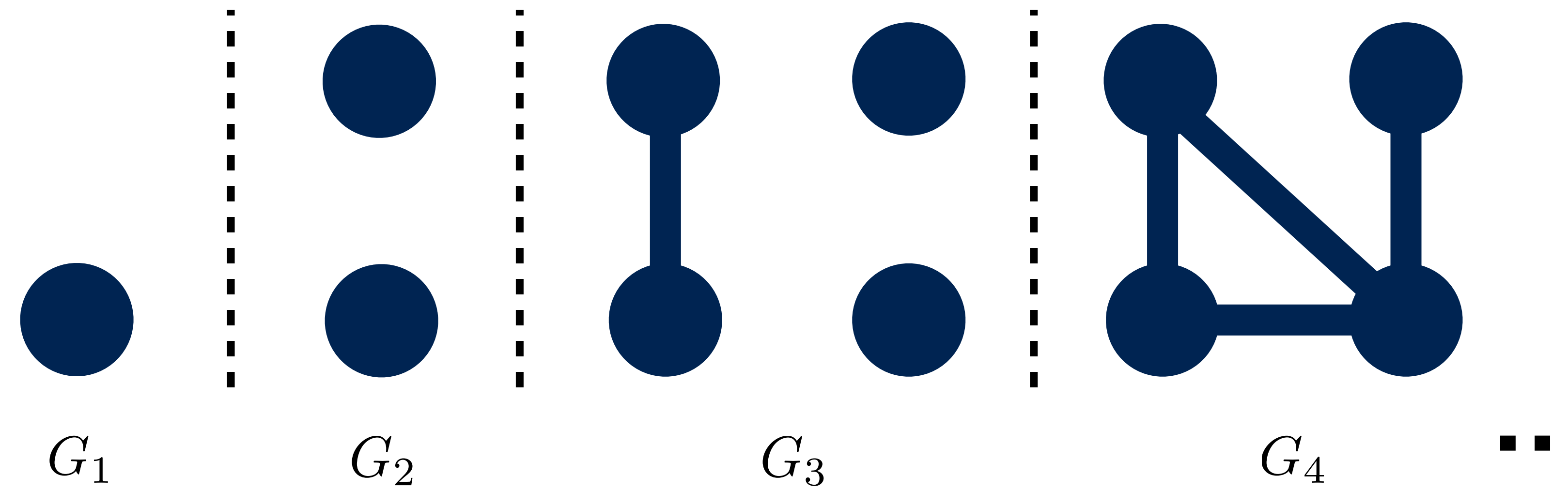
real-life scaling properties:

Sparse: $[\#\text{edges}(G_n)] \in o([\#\text{vertices}(G_n)]^2)$ *sub-quadratic*

popular models:

Network models should reflect **real-life scaling properties**

sequence of graphs



real-life scaling properties:

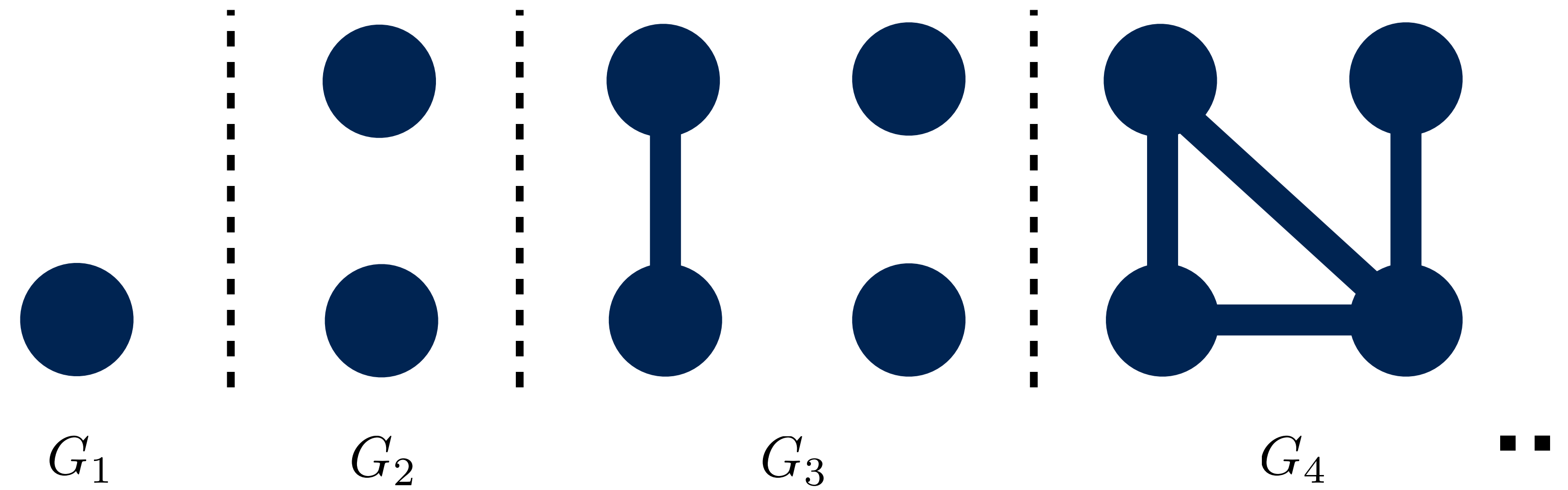
Sparse: $[\#\mathrm{edges}(G_n)] \in o([\#\mathrm{vertices}(G_n)]^2)$ *sub-quadratic*

popular models:

dense:

Network models should reflect **real-life scaling properties**

sequence of graphs



real-life scaling properties:

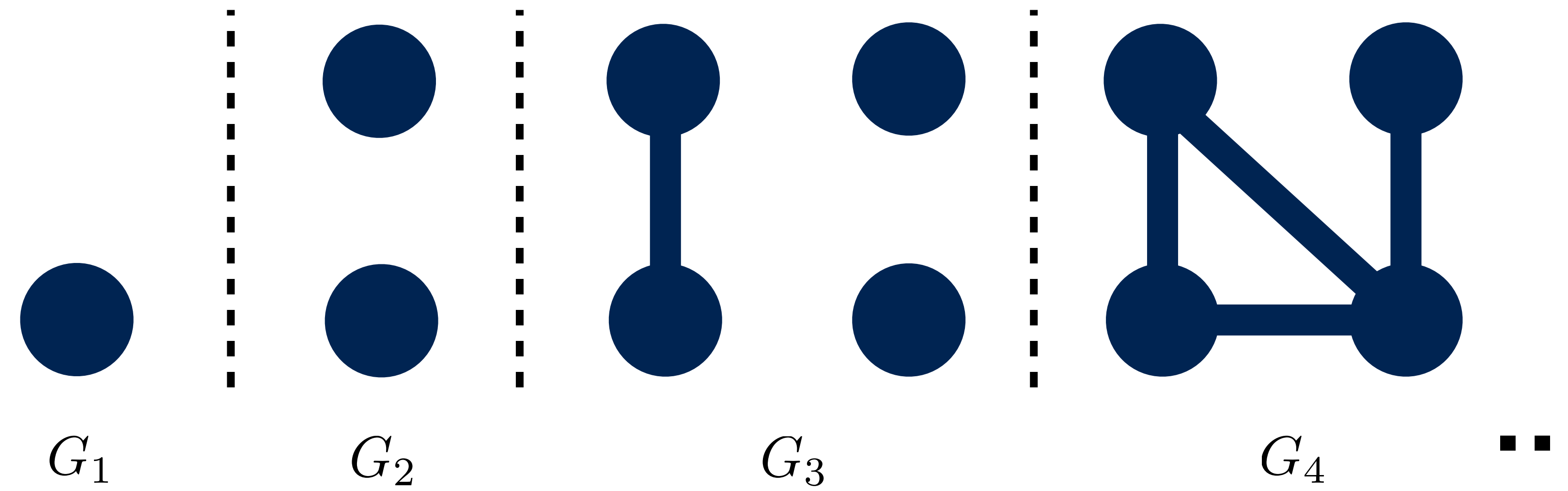
sparse: $[\#\text{edges}(G_n)] \in o([\#\text{vertices}(G_n)]^2)$ *sub-quadratic*

popular models:

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

Network models should reflect **real-life scaling properties**

sequence of graphs



real-life scaling properties:

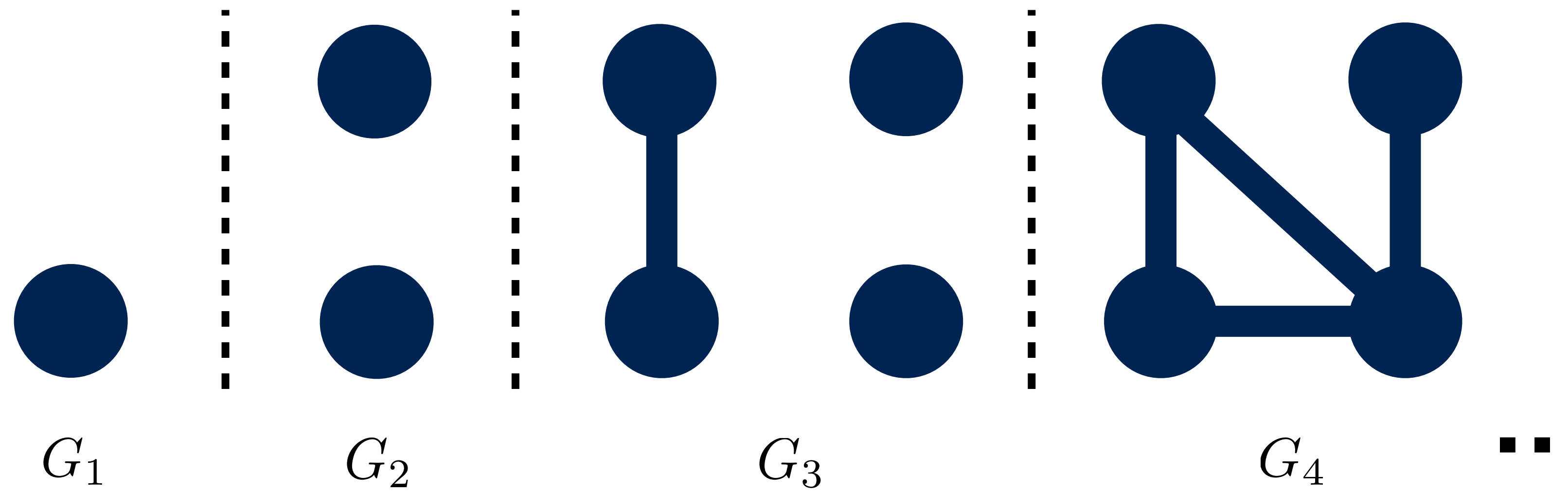
sparse: $[\#\mathrm{edges}(G_n)] \in o([\#\mathrm{vertices}(G_n)]^2)$ *sub-quadratic*

popular models:

dense: $[\#\mathrm{edges}(G_n)] \in \Omega([\#\mathrm{vertices}(G_n)]^2)$ *quadratic*

Network models should reflect **real-life scaling properties**

sequence of graphs



real-life scaling properties:

sparse: $[\#\text{edges}(G_n)] \in o([\#\text{vertices}(G_n)]^2)$ *sub-quadratic*

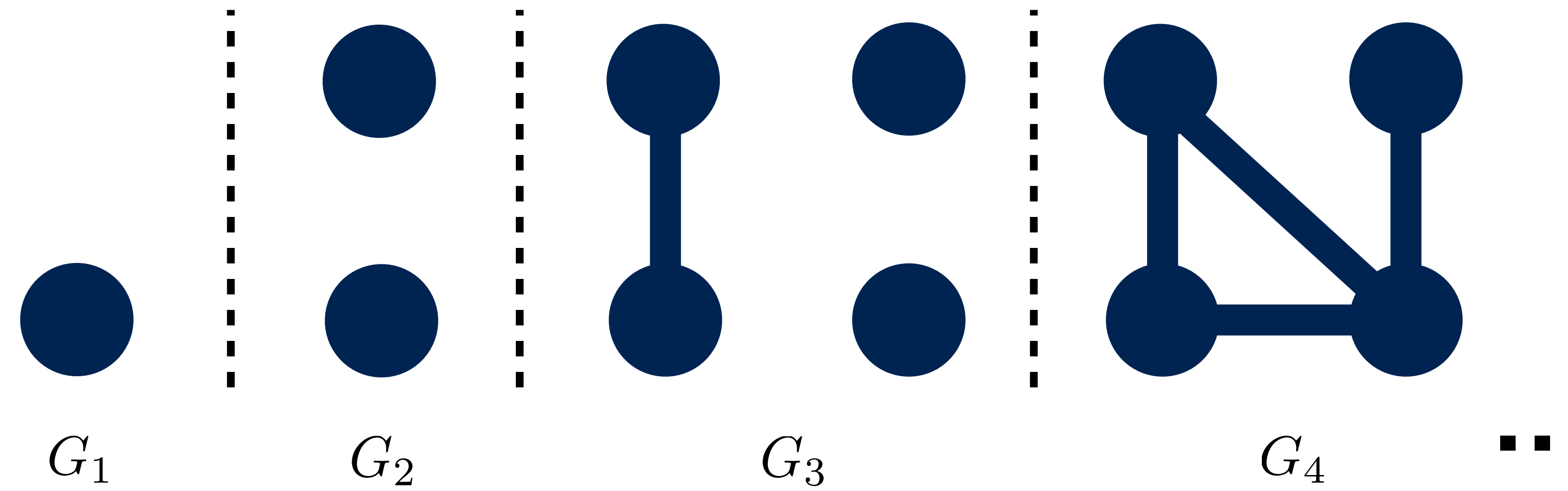
popular models:

dense: $[\#\text{edges}(G_n)] \in \Omega([\#\text{vertices}(G_n)]^2)$ *quadratic*

probabilistic models dense w.p. 1:
stochastic block model,
mixed membership stochastic block model,
infinite relational model, and many more

Network models should reflect **real-life scaling properties**

sequence of graphs



real-life scaling properties:

sparse: $[\#\text{edges}(G_n)] \in o([\#\text{vertices}(G_n)]^2)$ *sub-quadratic*

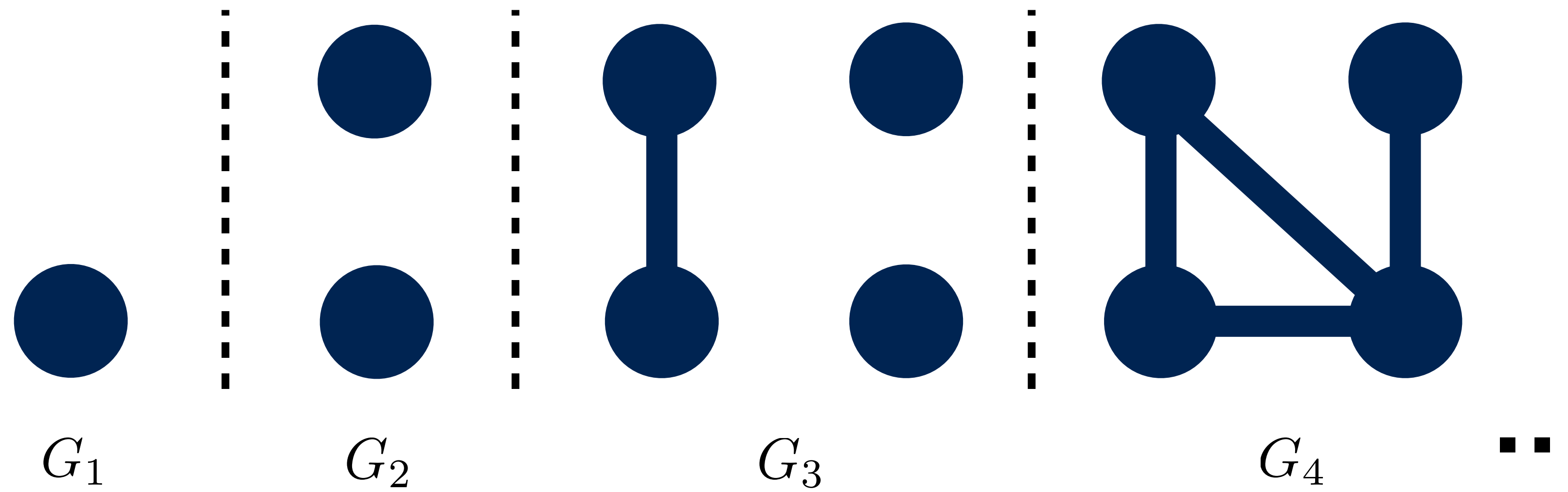
popular models: **vertex exchangeability**

dense: $[\#\text{edges}(G_n)] \in \Omega([\#\text{vertices}(G_n)]^2)$ *quadratic*

probabilistic models dense w.p. 1:
stochastic block model,
mixed membership stochastic block model,
infinite relational model, and many more

Network models should reflect **real-life scaling properties**

sequence of graphs



real-life scaling properties: **edge exchangeability**

sparse: $[\#\text{edges}(G_n)] \in o([\#\text{vertices}(G_n)]^2)$ *sub-quadratic*

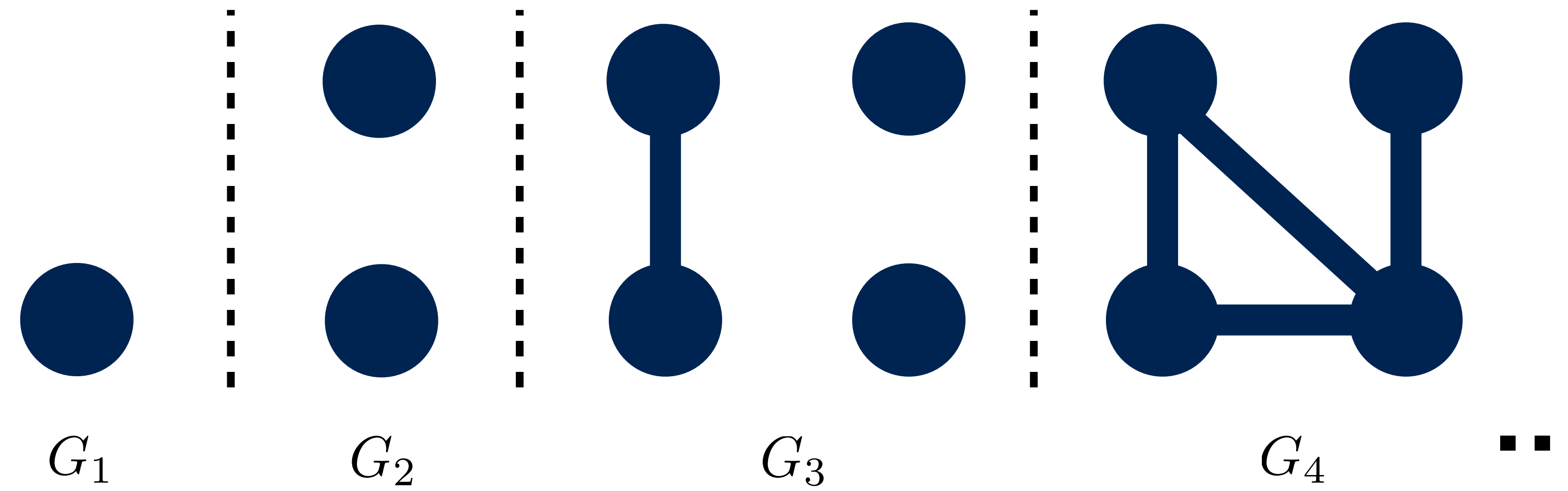
popular models: **vertex exchangeability**

dense: $[\#\text{edges}(G_n)] \in \Omega([\#\text{vertices}(G_n)]^2)$ *quadratic*

probabilistic models dense w.p. 1:
stochastic block model,
mixed membership stochastic block model,
infinite relational model, and many more

Network models should reflect **real-life scaling properties**

sequence of graphs



real-life scaling properties: **edge exchangeability**

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

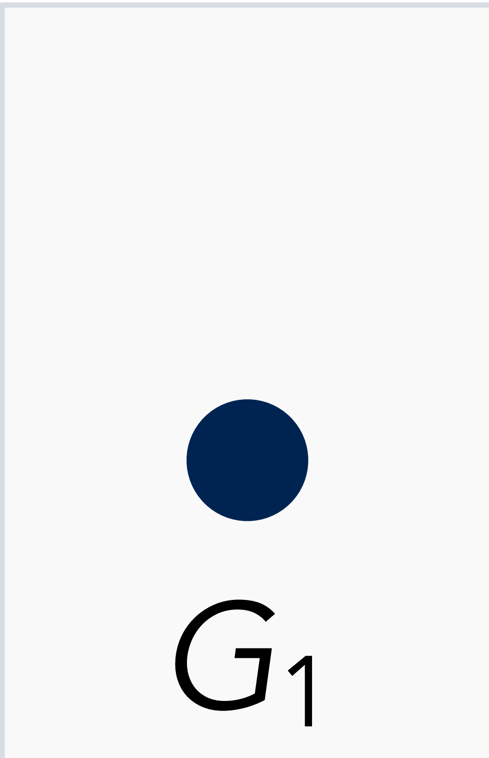
popular models: **vertex exchangeability**

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

probabilistic models dense w.p. 1:
stochastic block model,
mixed membership stochastic block model,
infinite relational model, and many more

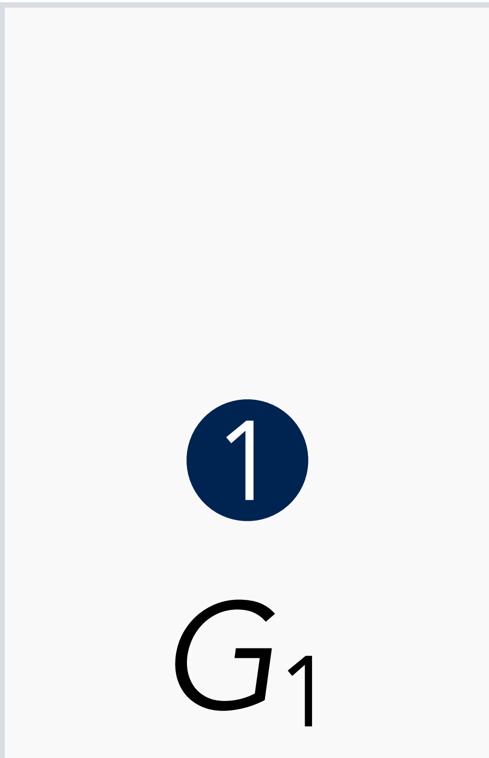
Vertex-exchangeable graph sequences

(are always dense)

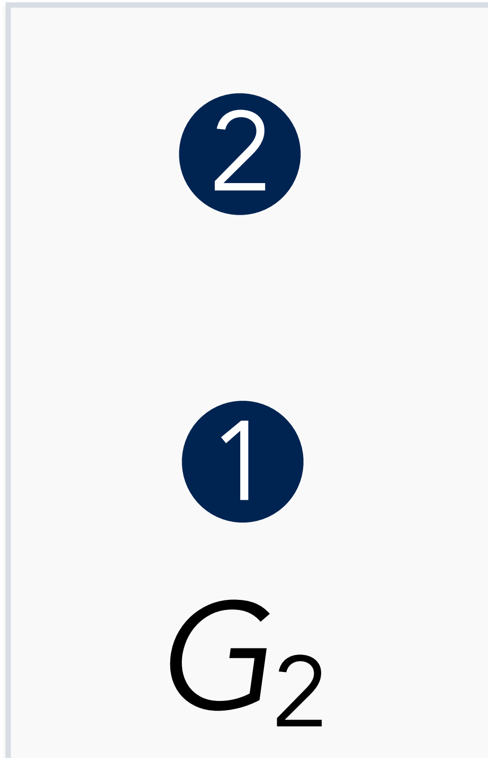

G_1

1

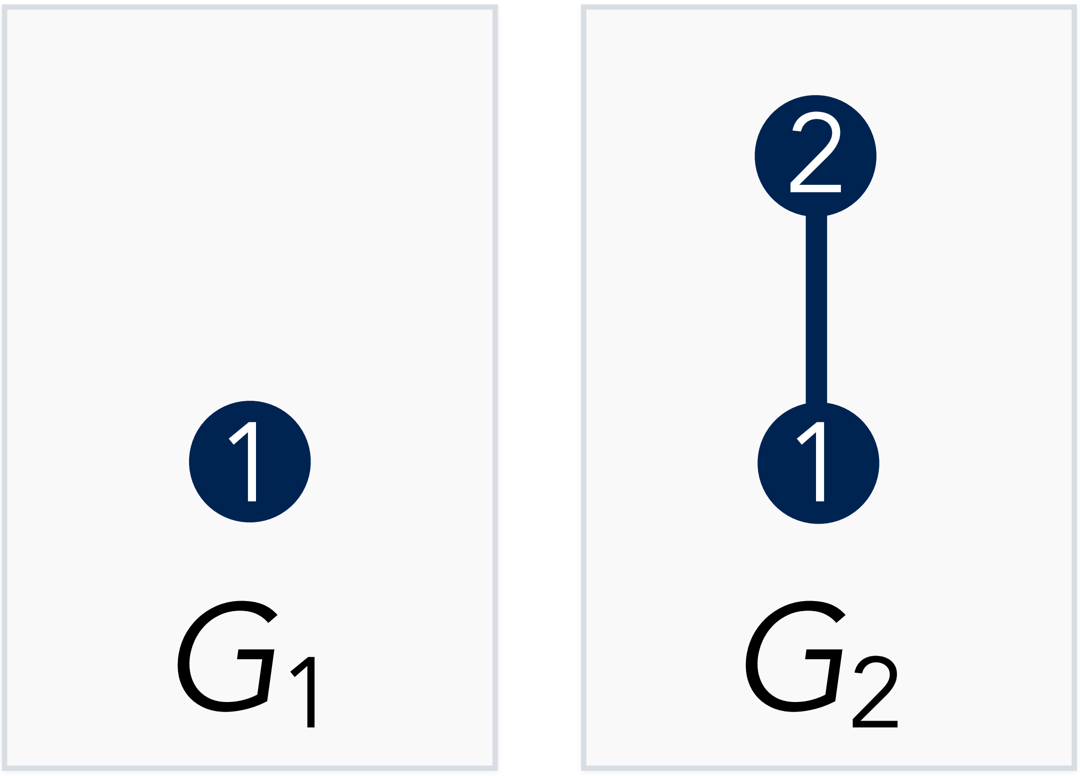
G_1

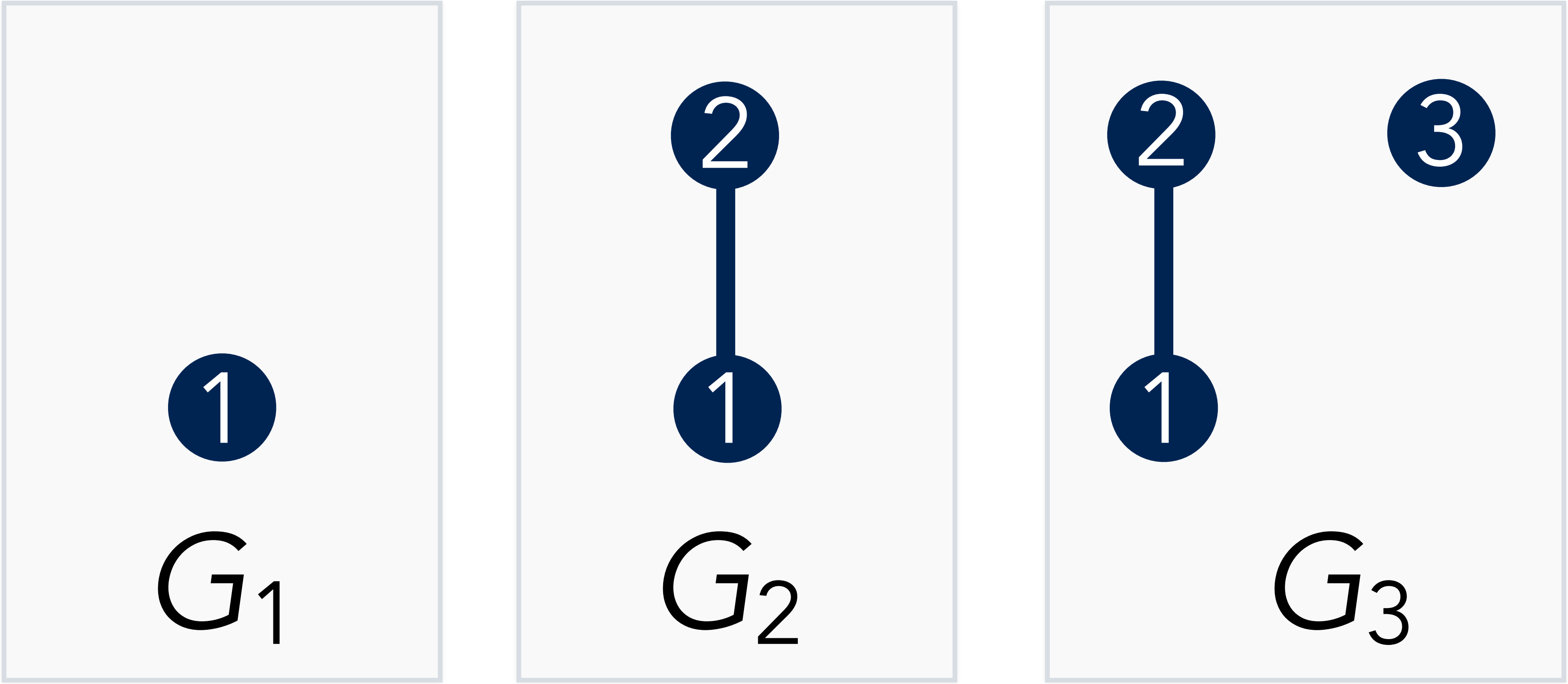


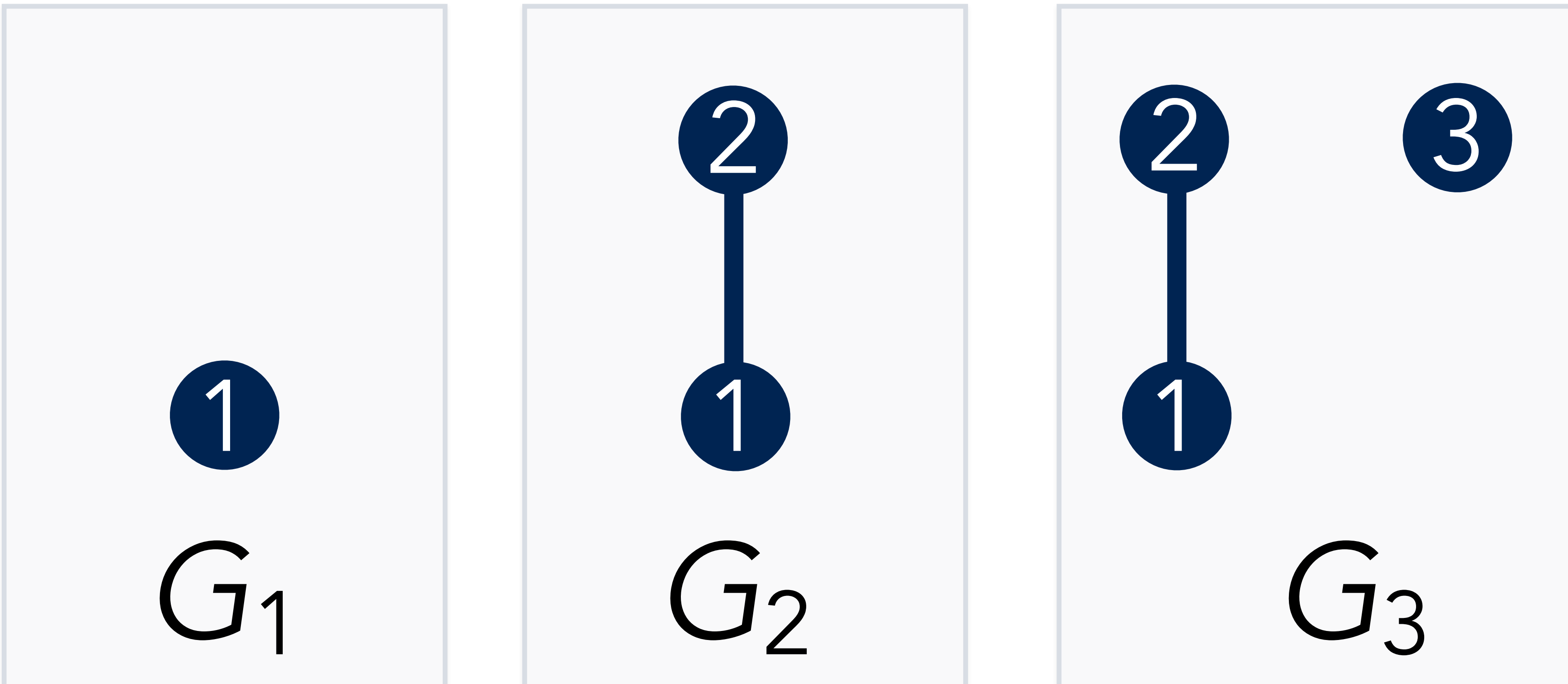
1
G_1

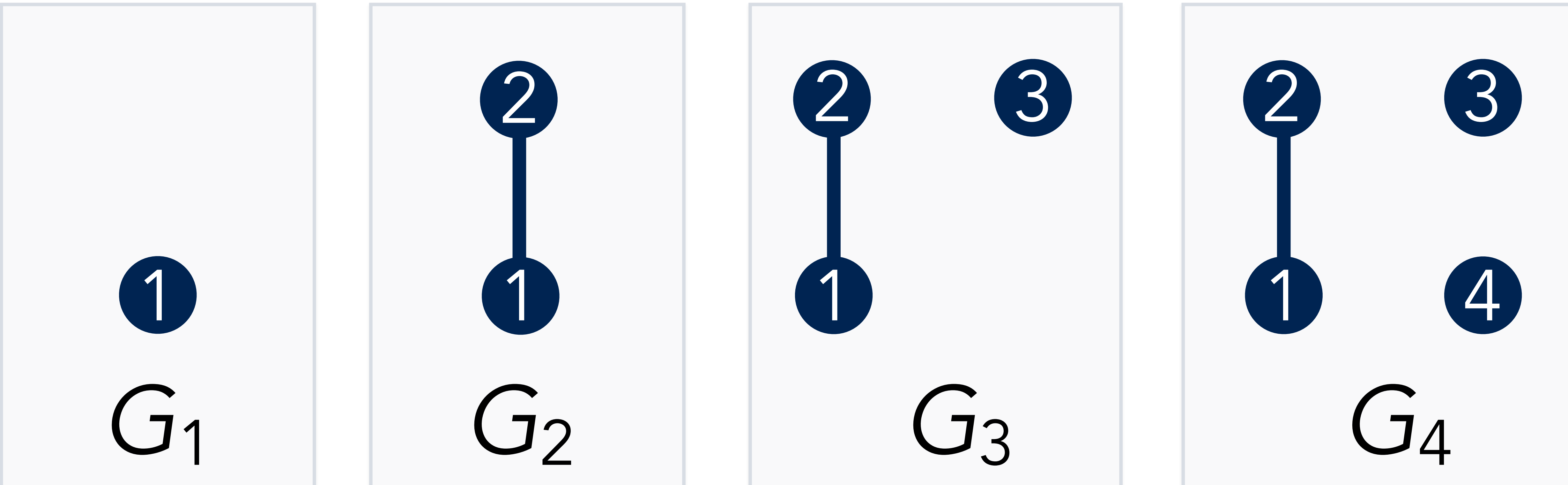


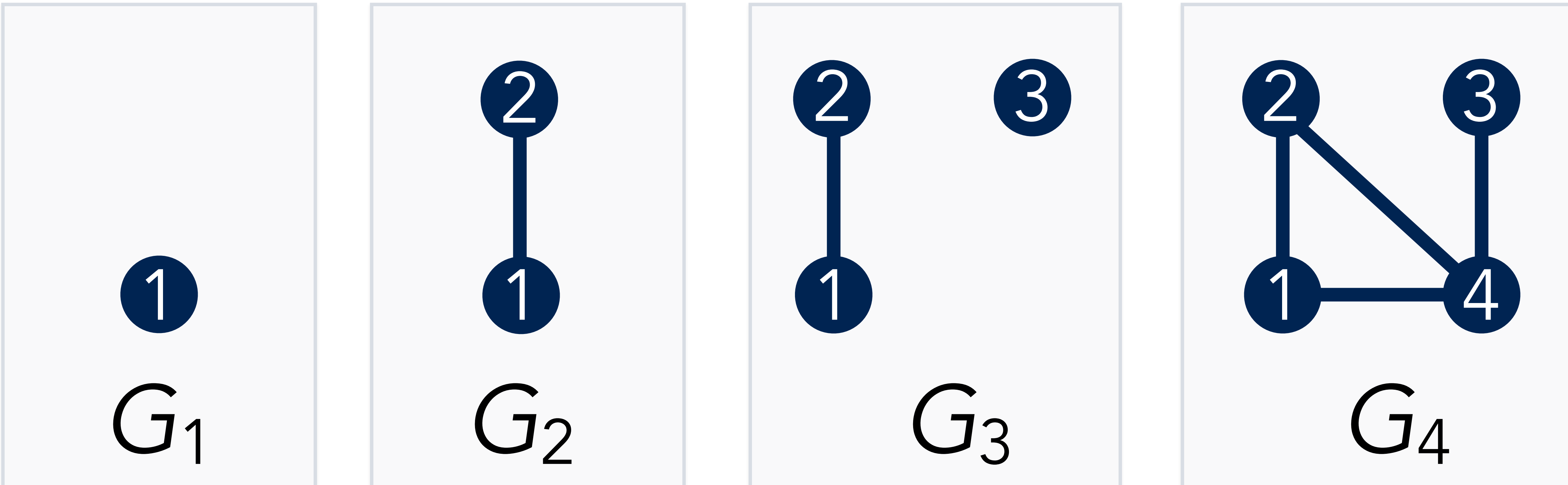
2
1
G_2

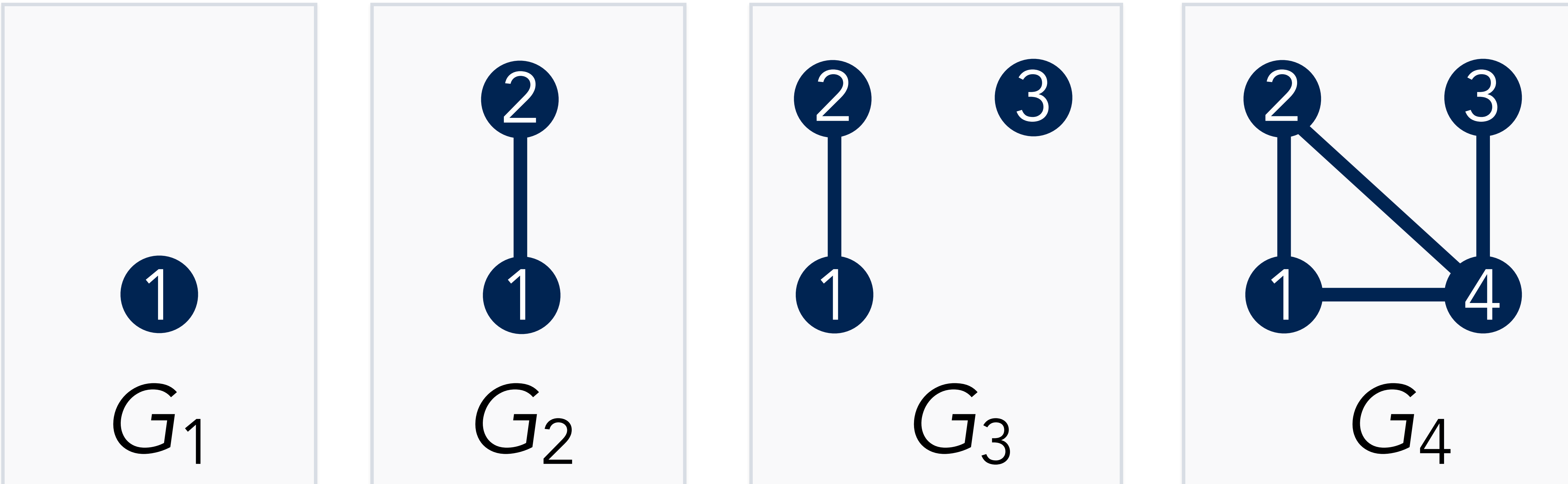

1
G_1
2
1
G_2


1
G_1
2
1
G_2
2
3
1
G_3

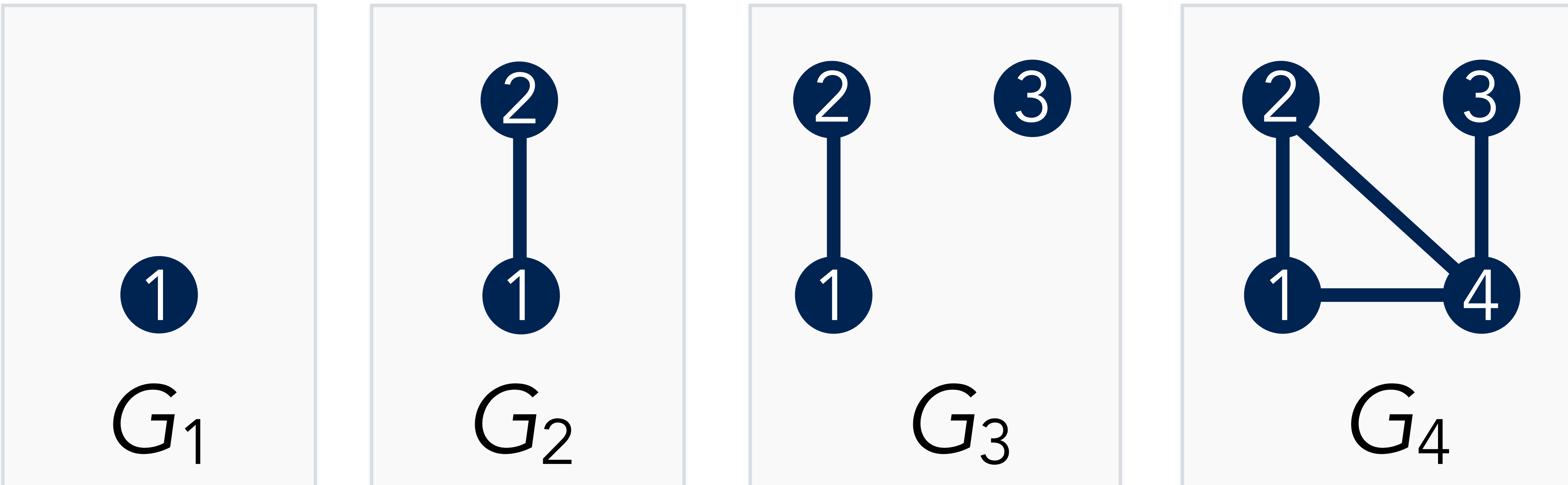

1
G_1
2
1
G_2
2
3
1
G_3

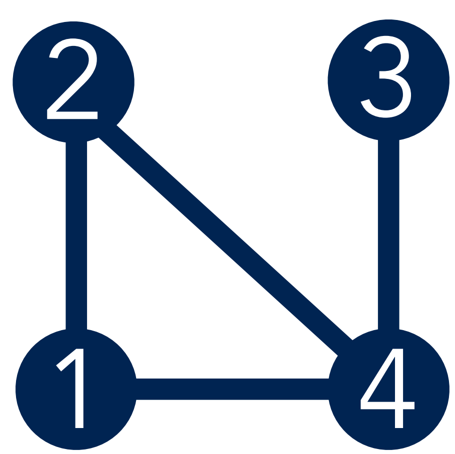

1
G_1
2
1
G_2
2
3
1
G_3
2
3
1
4
G_4

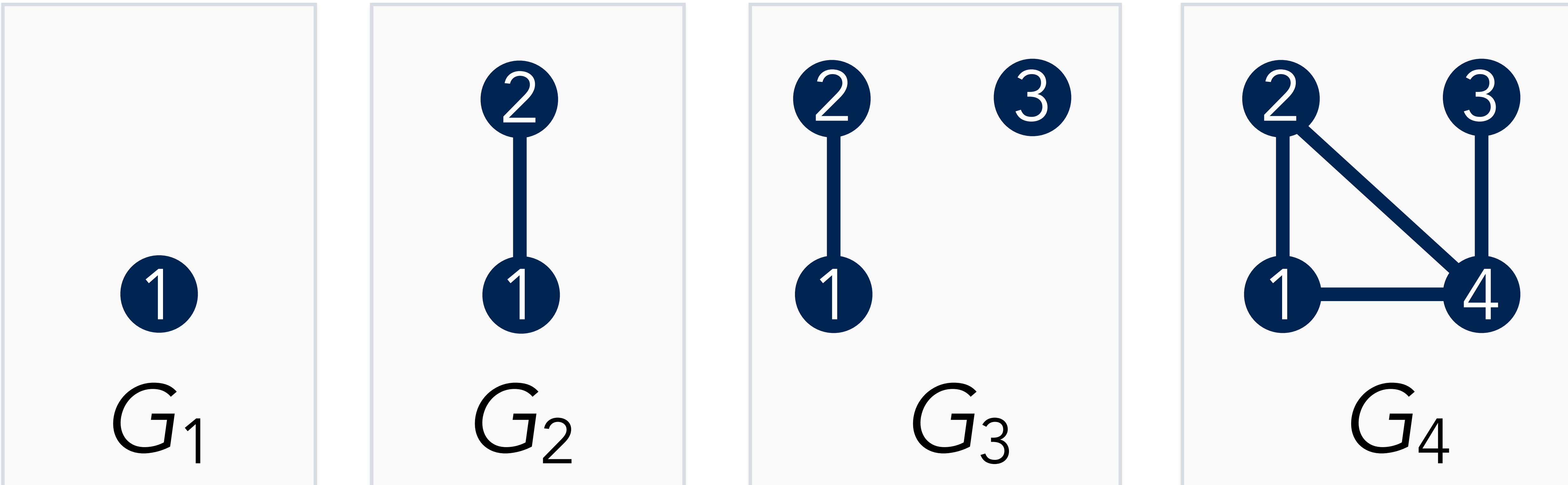

G_1
G_2
G_3
G_4

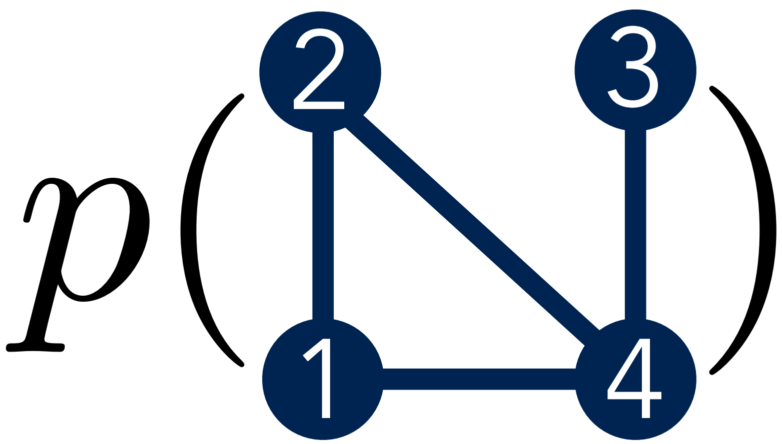


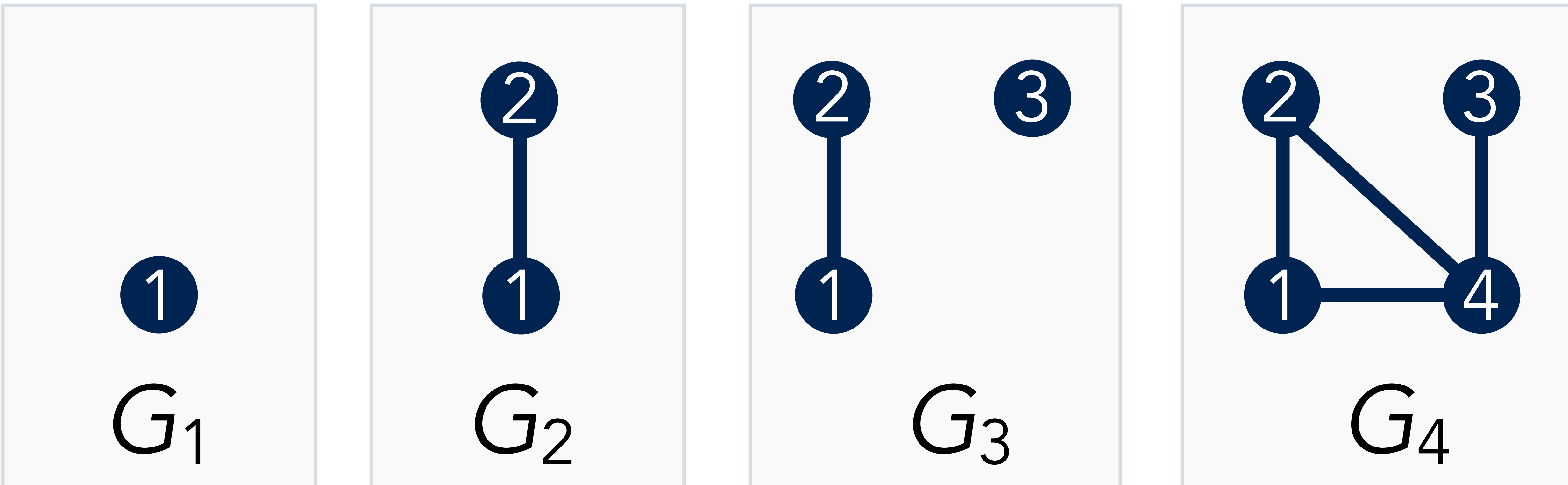
and so on …

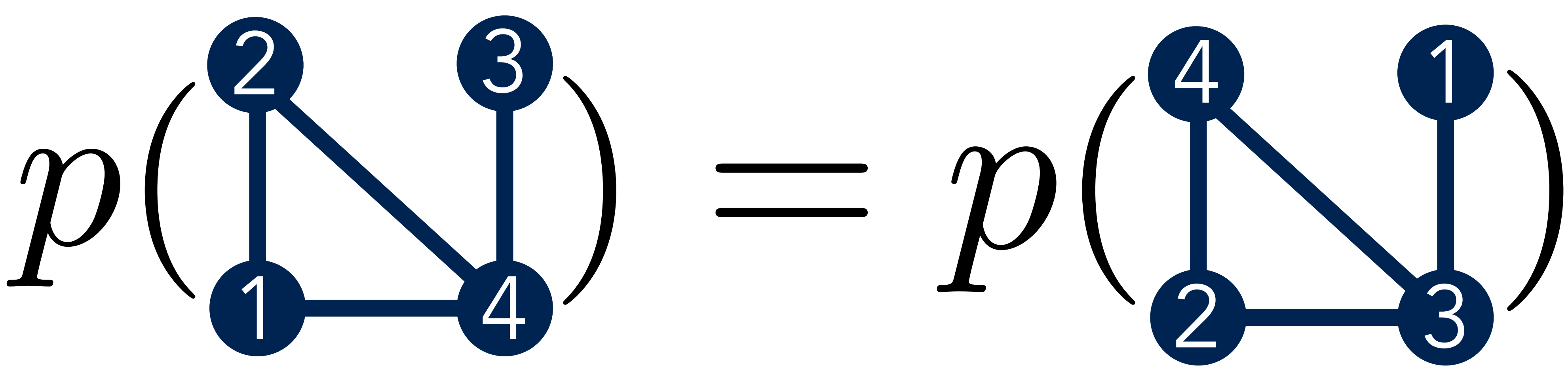

G_1
G_2
G_3
G_4

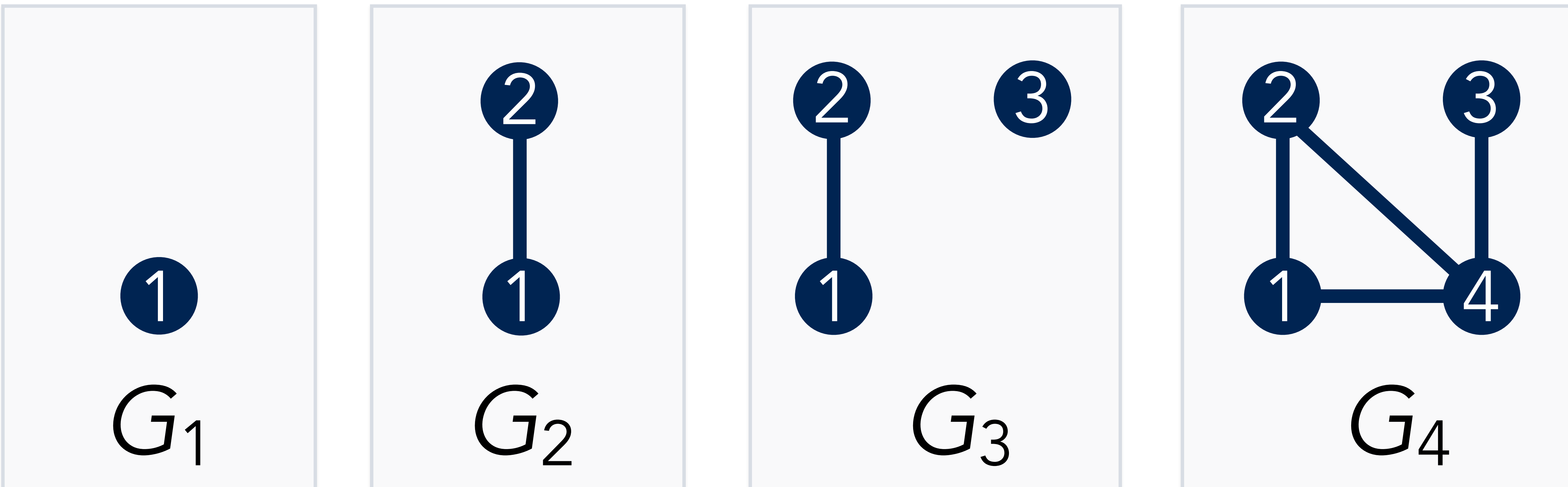


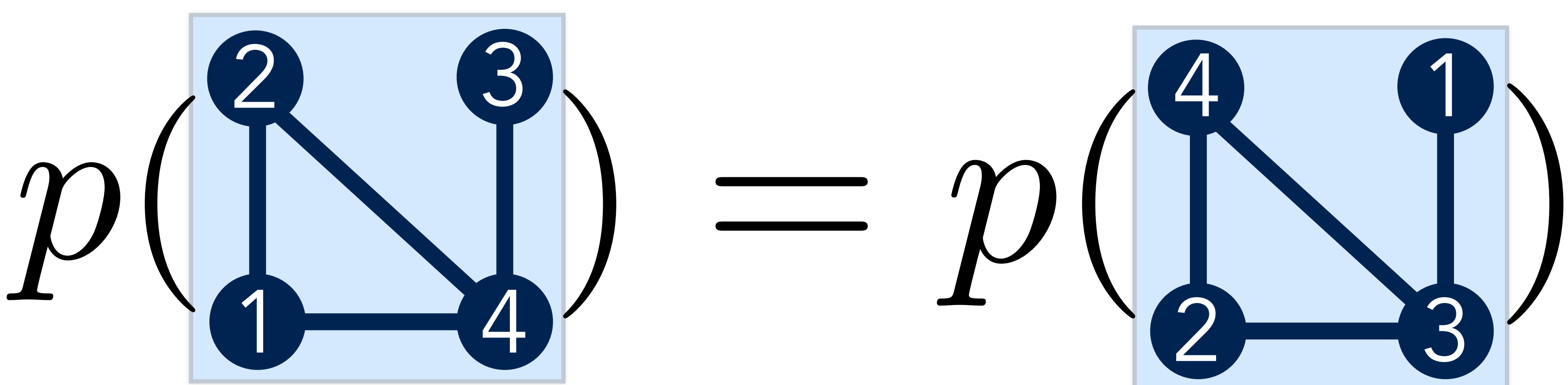

1
G_1
2
1
G_2
2
3
1
G_3
2
3
1
4
G_4


$p($
2
3
1
4
$)$


G_1
G_2
G_3
G_4

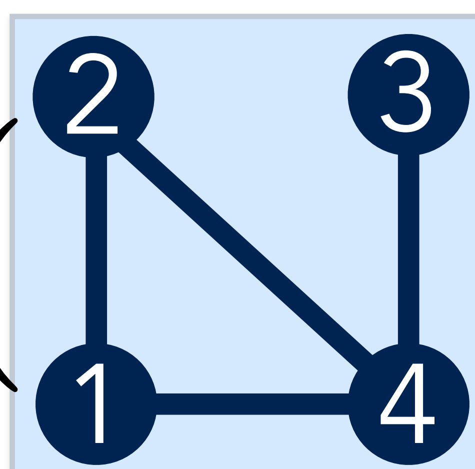

p
$=$
p

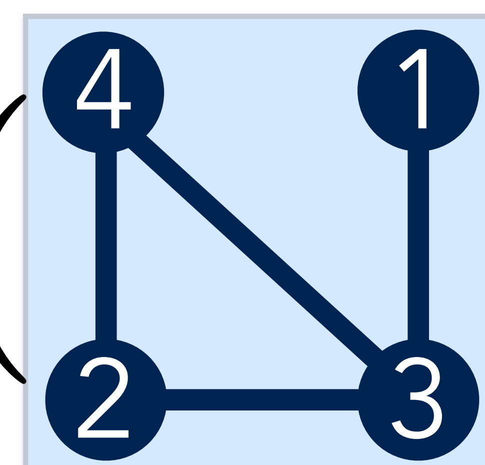

G_1
G_2
G_3
G_4



The Aldous-Hoover theorem implies that vertex-exchangeable graphs are dense or empty with probability 1.

G_1 G_2 G_3 G_4

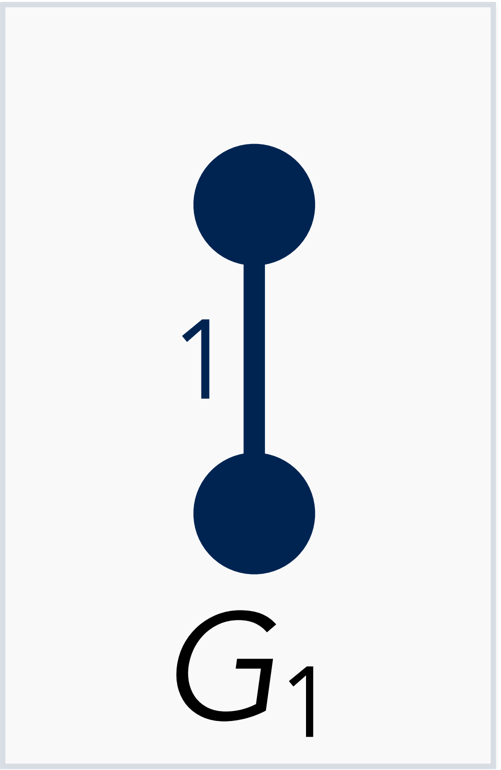
$p\Bigg($ 

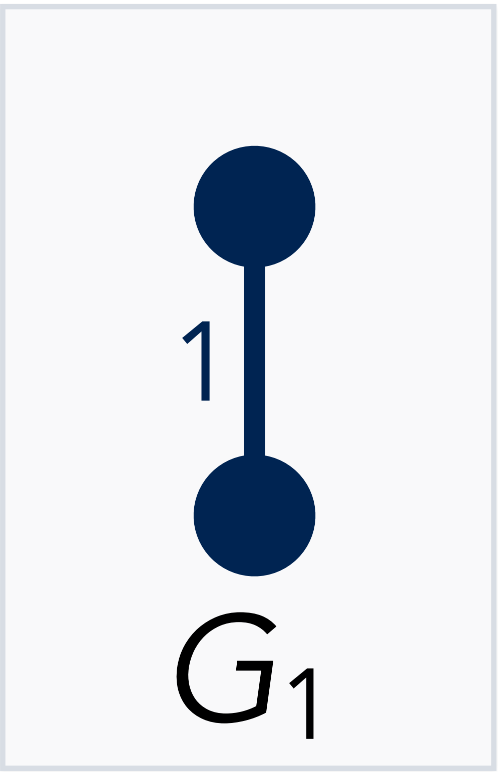
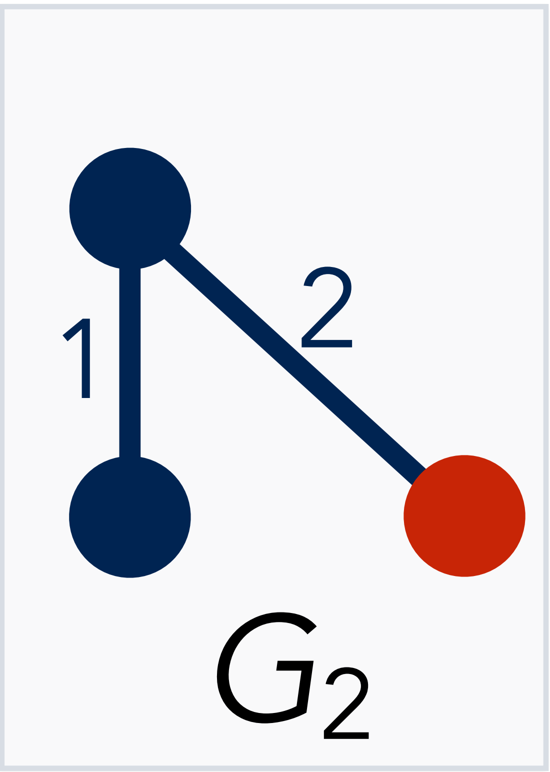
$\Bigg) = p\Bigg($ 

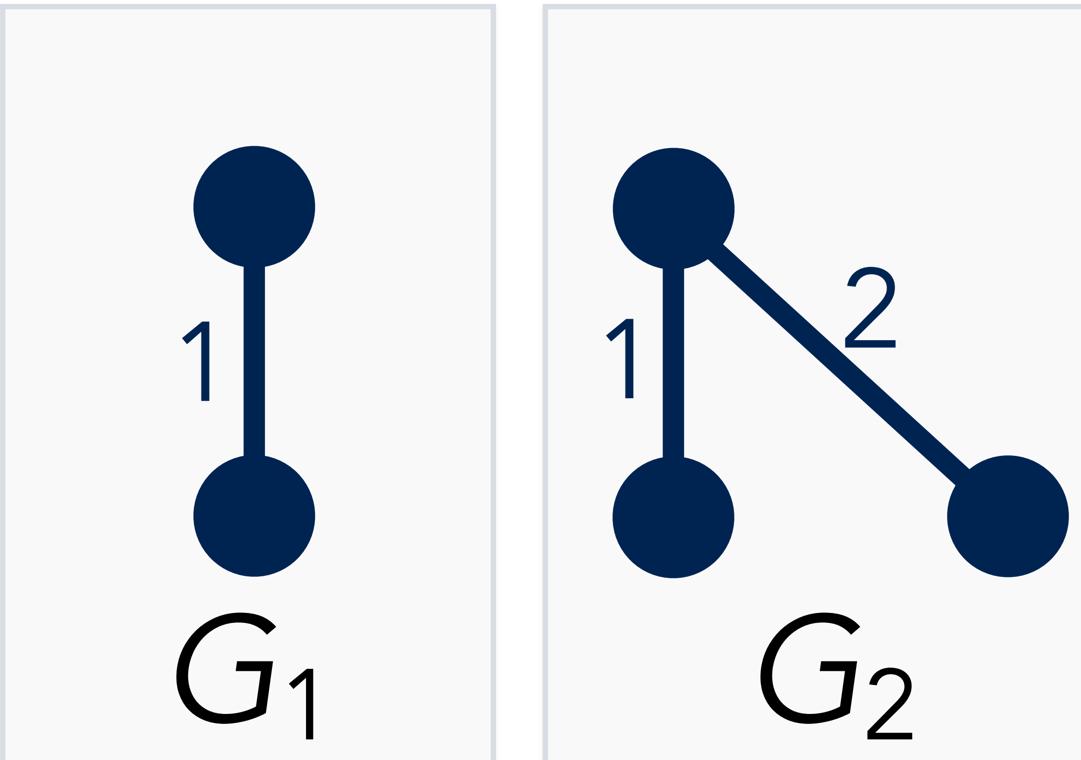
$\Bigg)$

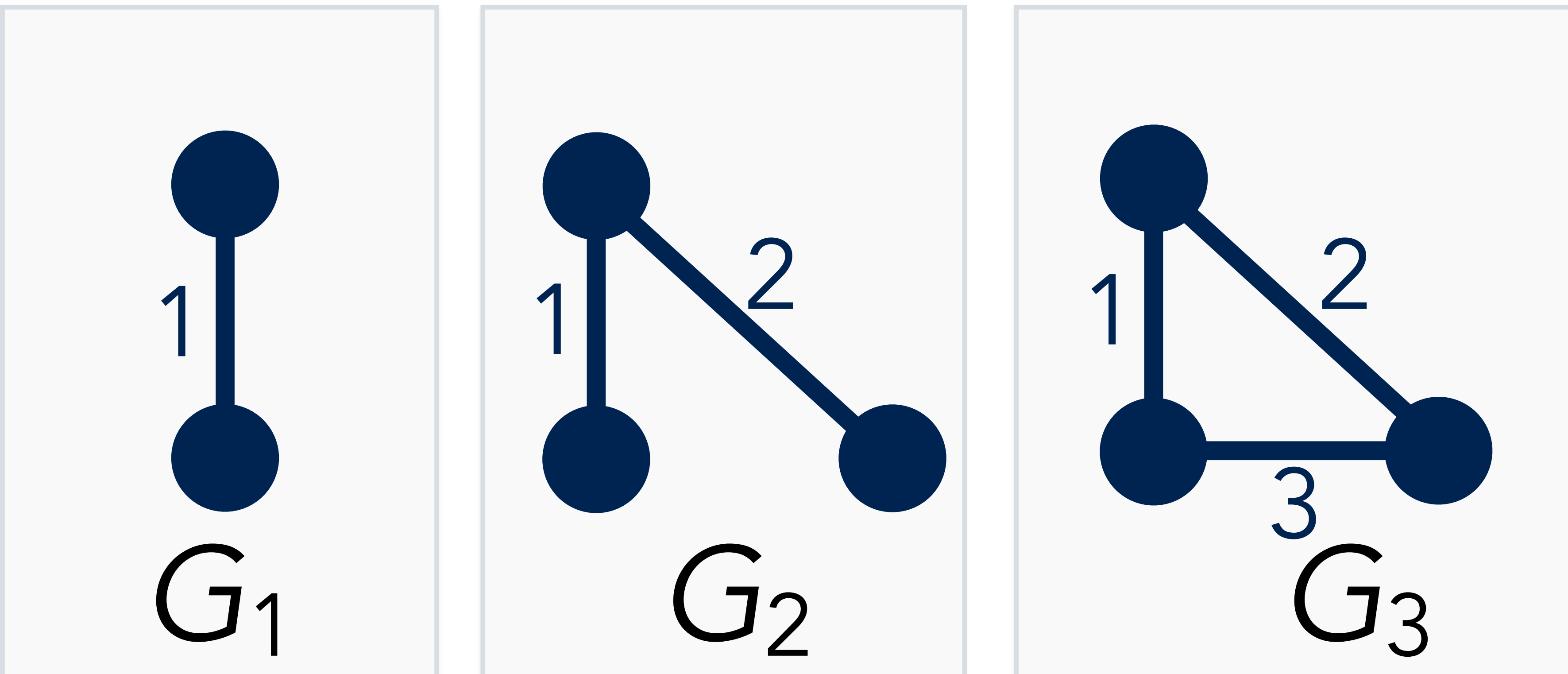
Edge-exchangeable graph sequences

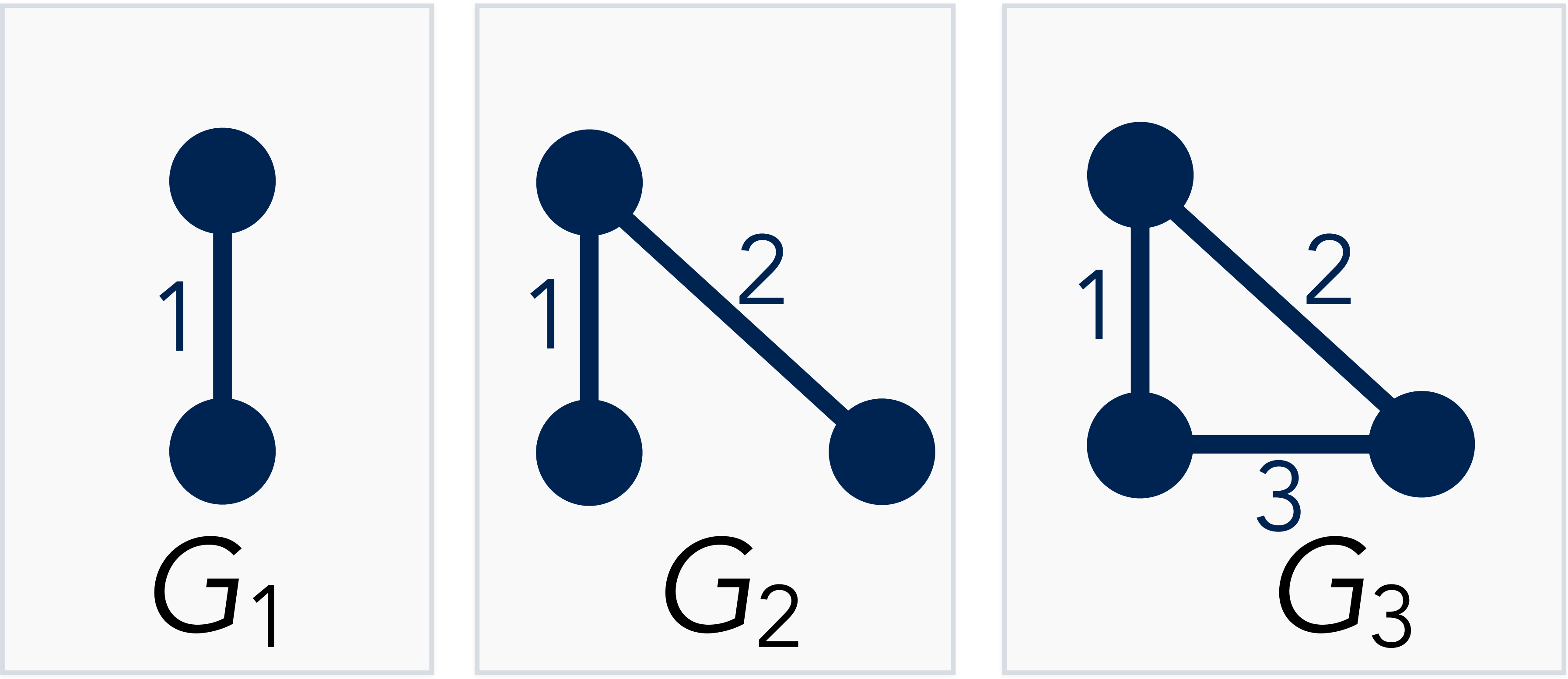
G_1

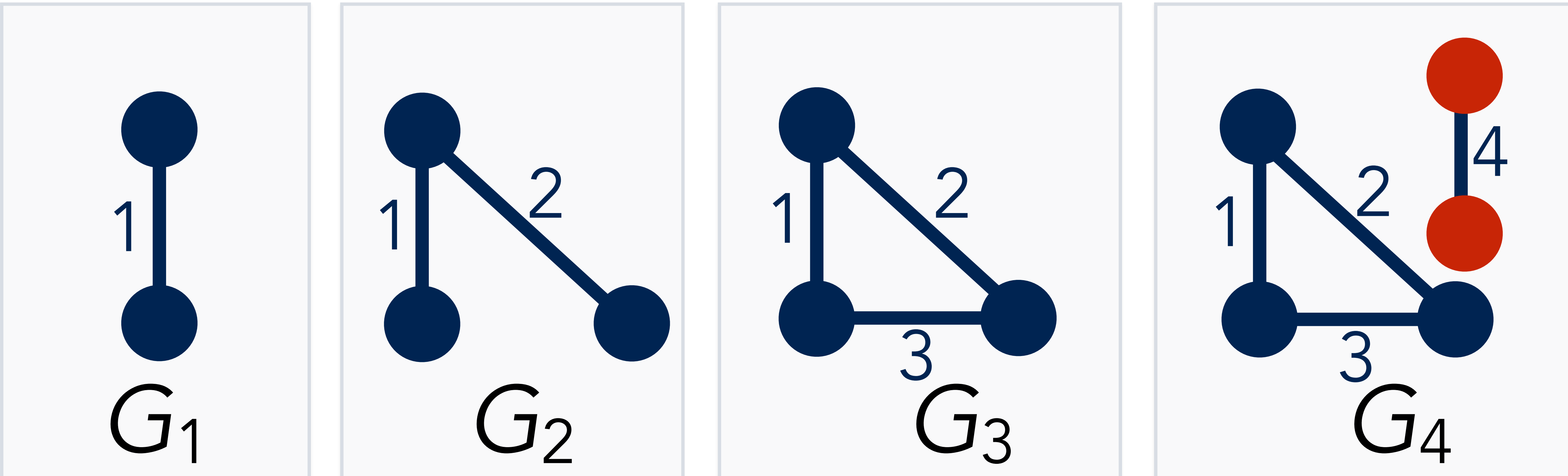

1
G_1

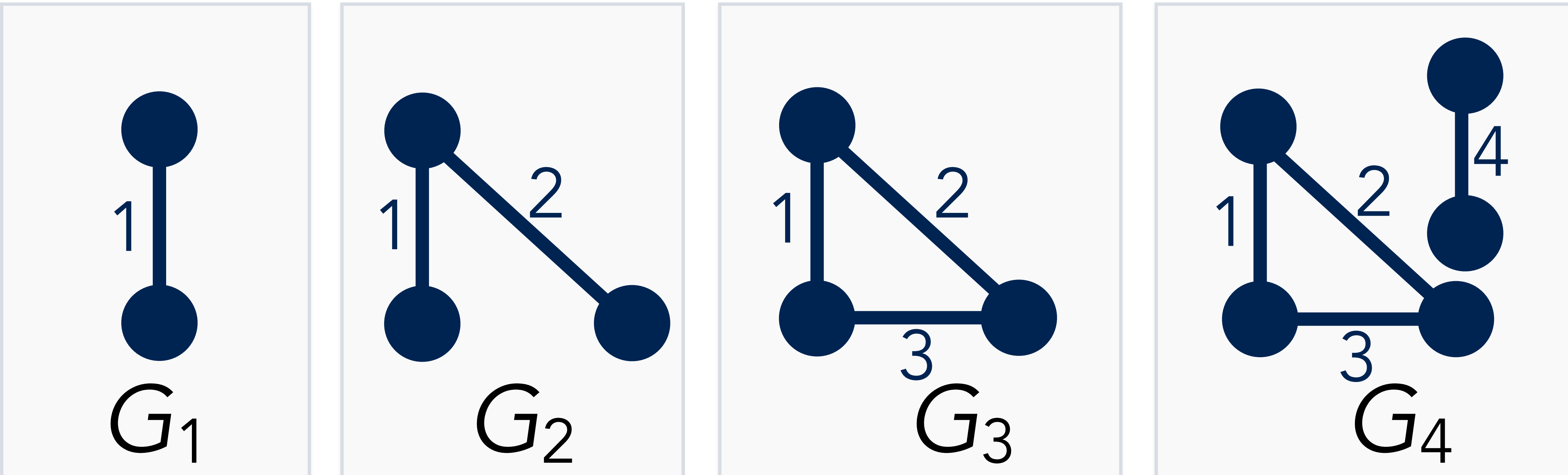

1
G_1

1
2
G_2

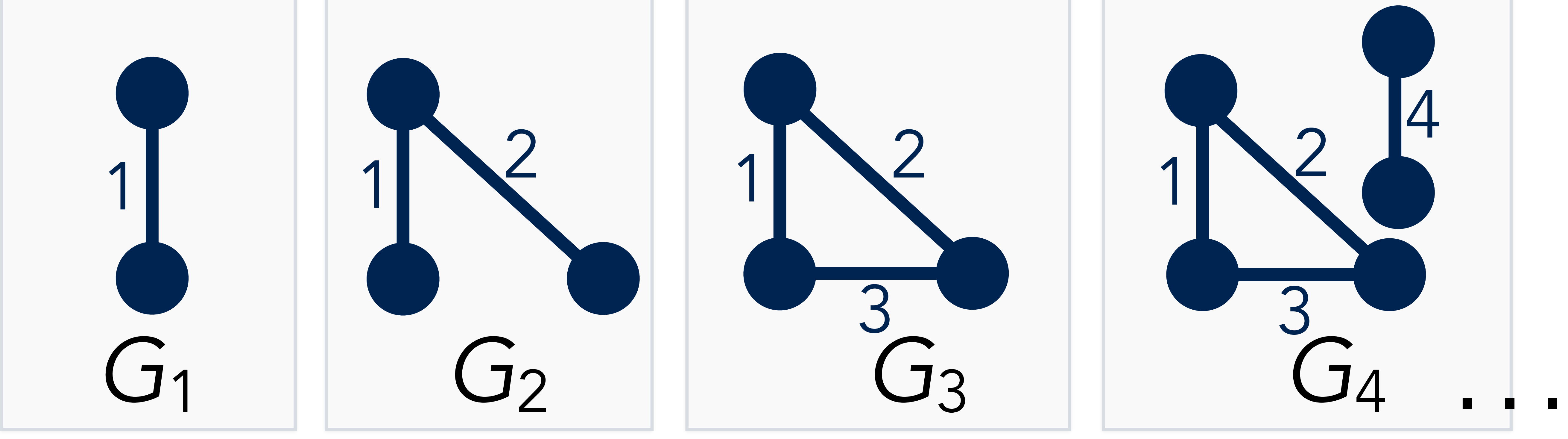

1
G_1
1
2
G_2


1
G1
1
2
G2
1
2
3
G3

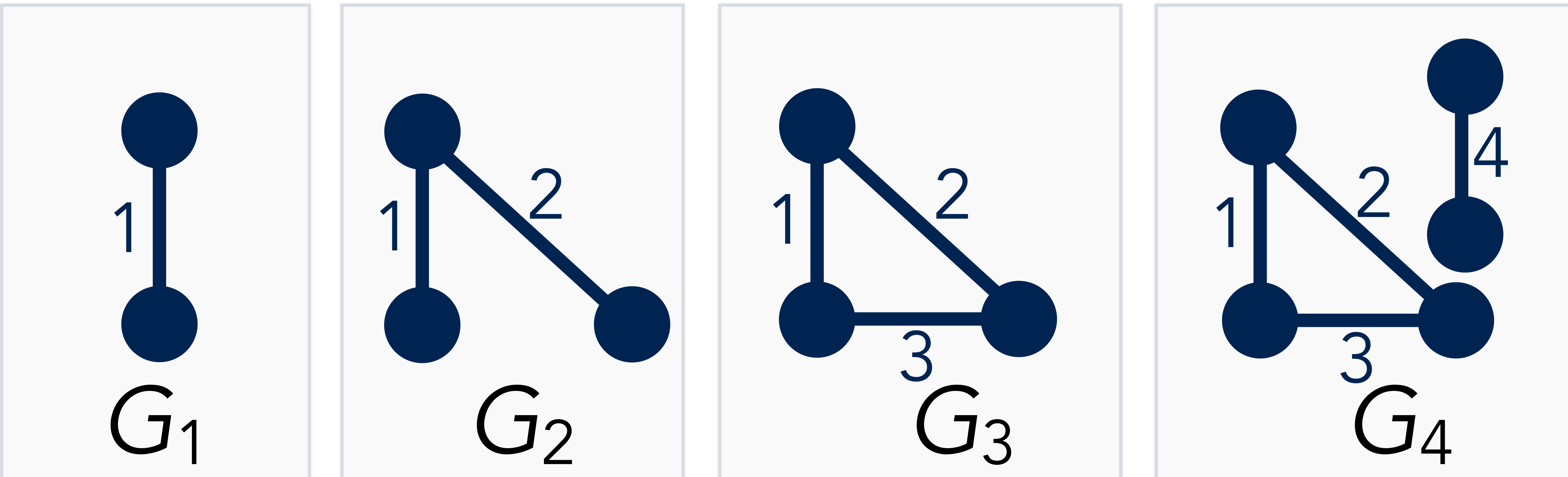

1
G_1
1
2
G_2
1
2
3
G_3

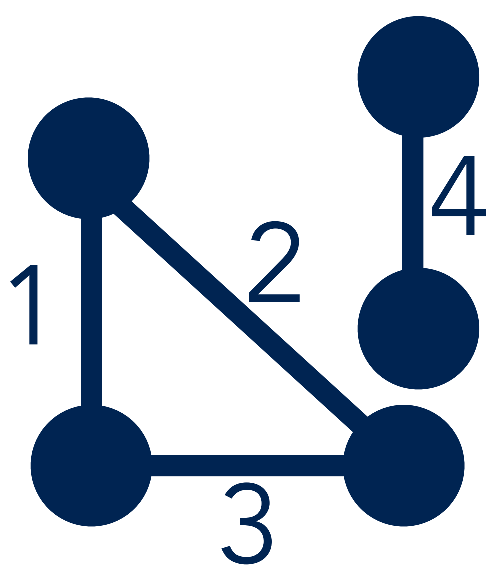

1
G_1
1
2
G_2
1
2
3
G_3
1
2
3
4
G_4

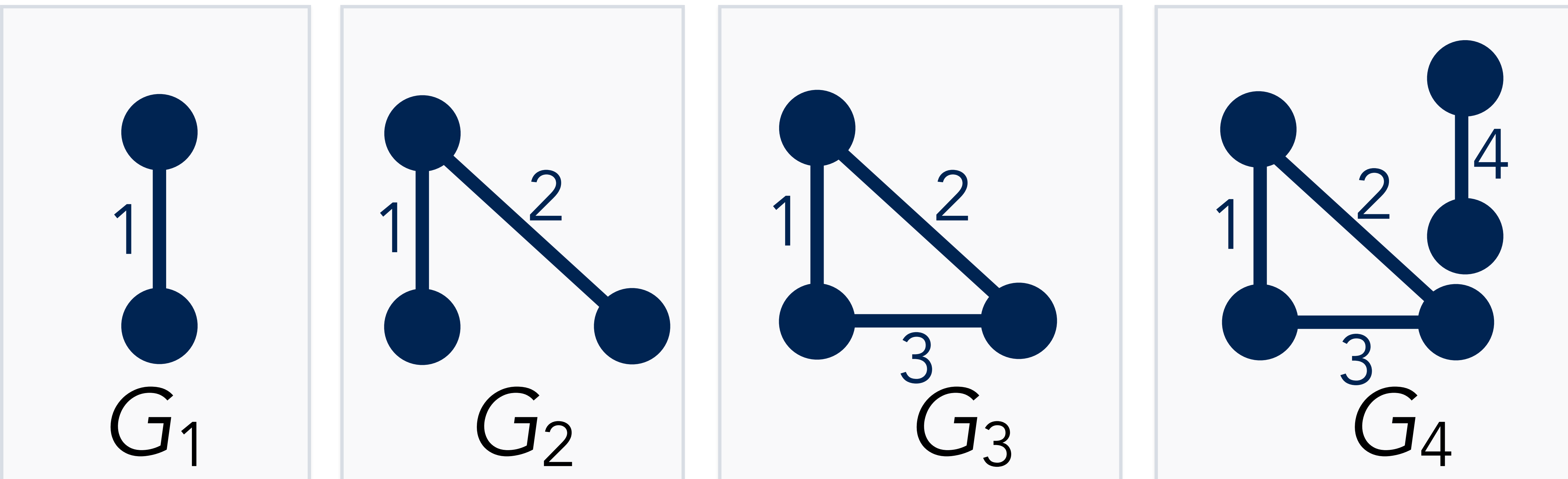

1
G_1
1
2
G_2
1
2
3
G_3
1
2
3
4
G_4

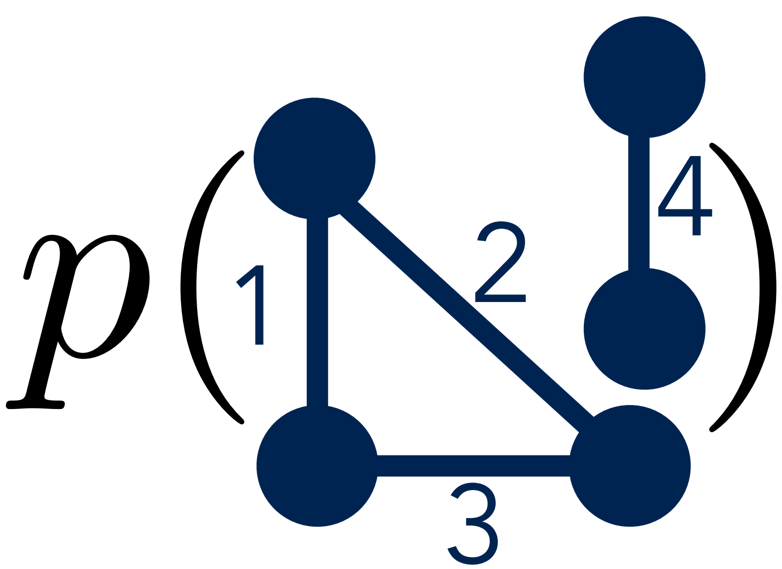


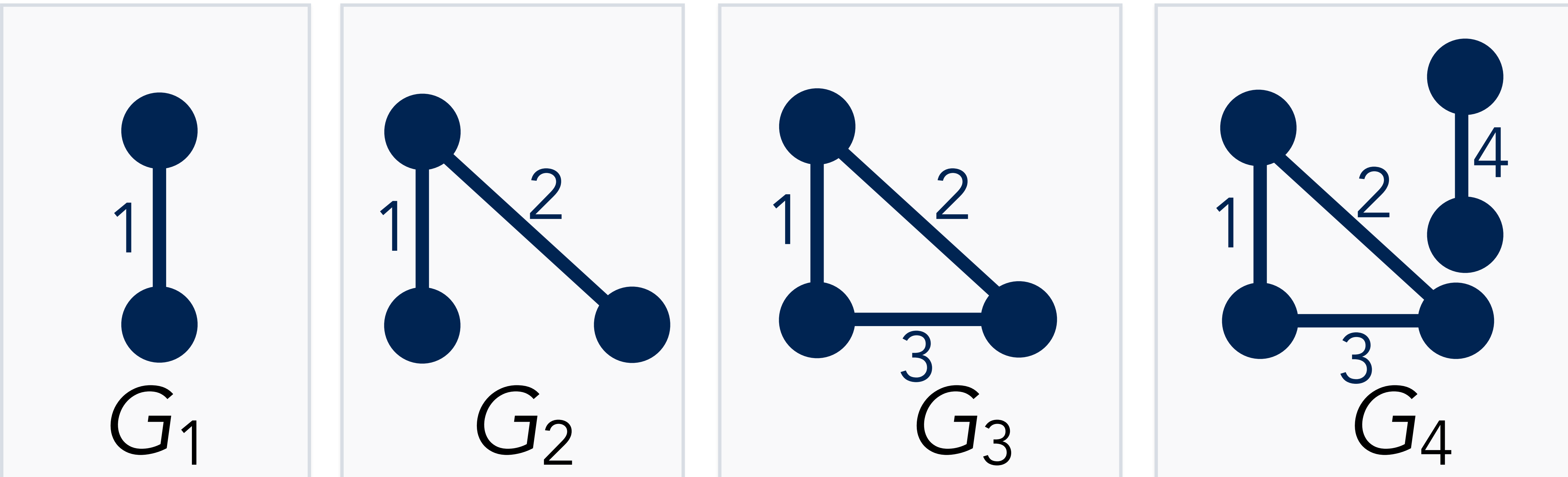
1
G_1
1
2
G_2
1
2
3
G_3
1
2
3
4
G_4
. . .

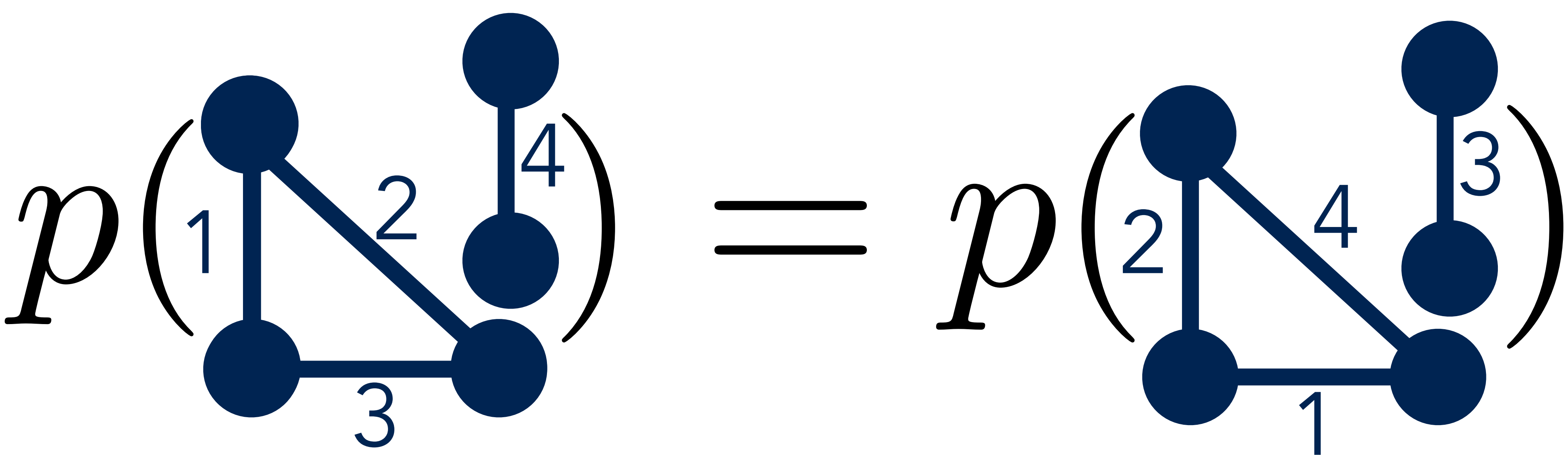

1
G1
1
2
G2
1
2
3
G3
1
2
3
4
G4


1
2
3
4

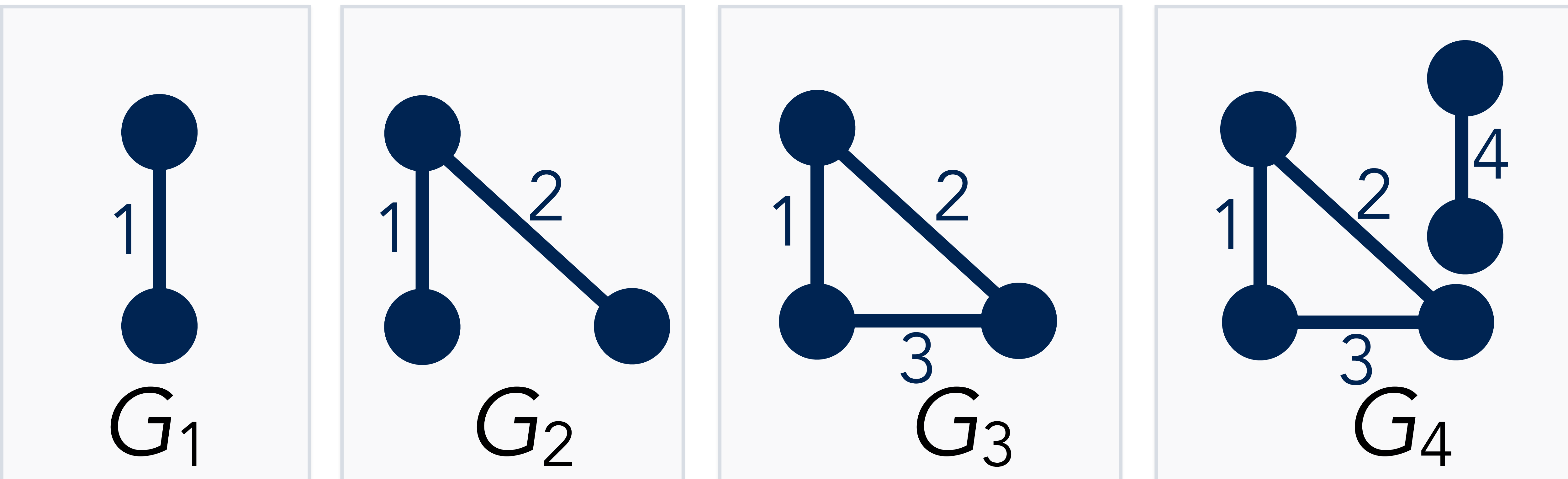

1
1
2
1
2
3
1
2
3
4
G_1
G_2
G_3
G_4

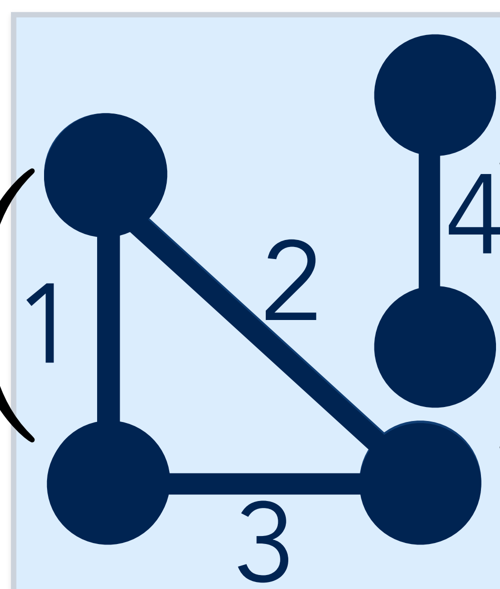
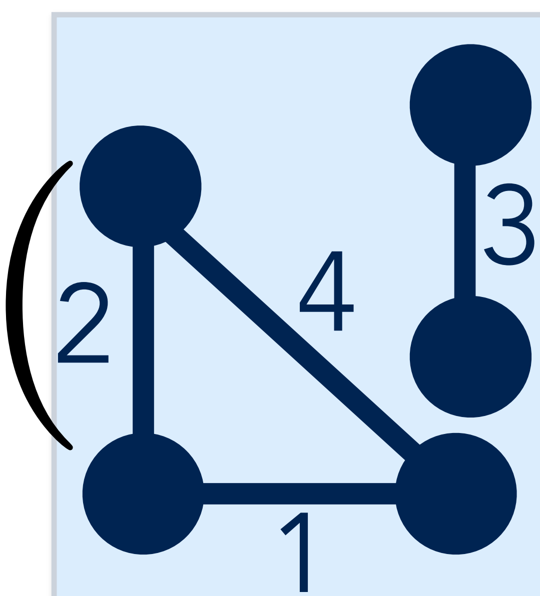

p
1
2
3
4





[Broderick & Cai, 2015; Cai, Campbell, Broderick, 2016; Crane & Dempsey, 2015, 2016]



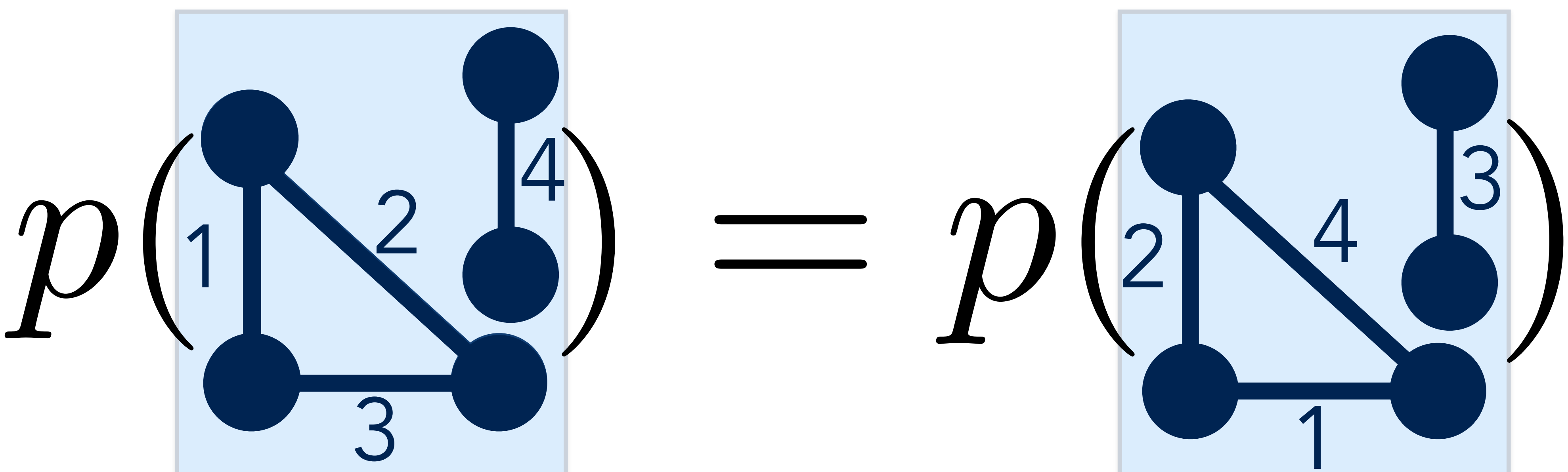
$p\Big($  $\Big) = p\Big($  $\Big)$

[Broderick & Cai, 2015; Cai, Campbell, Broderick, 2016; Crane & Dempsey, 2015, 2016]

Theorem (Cai-Campbell-Broderick, 2016). A wide class of edge-exchangeable graph models admits sparse graphs.

Want to characterize *all* sparse, edge-exchangeable graphs.

G_1 G_2 G_3 G_4

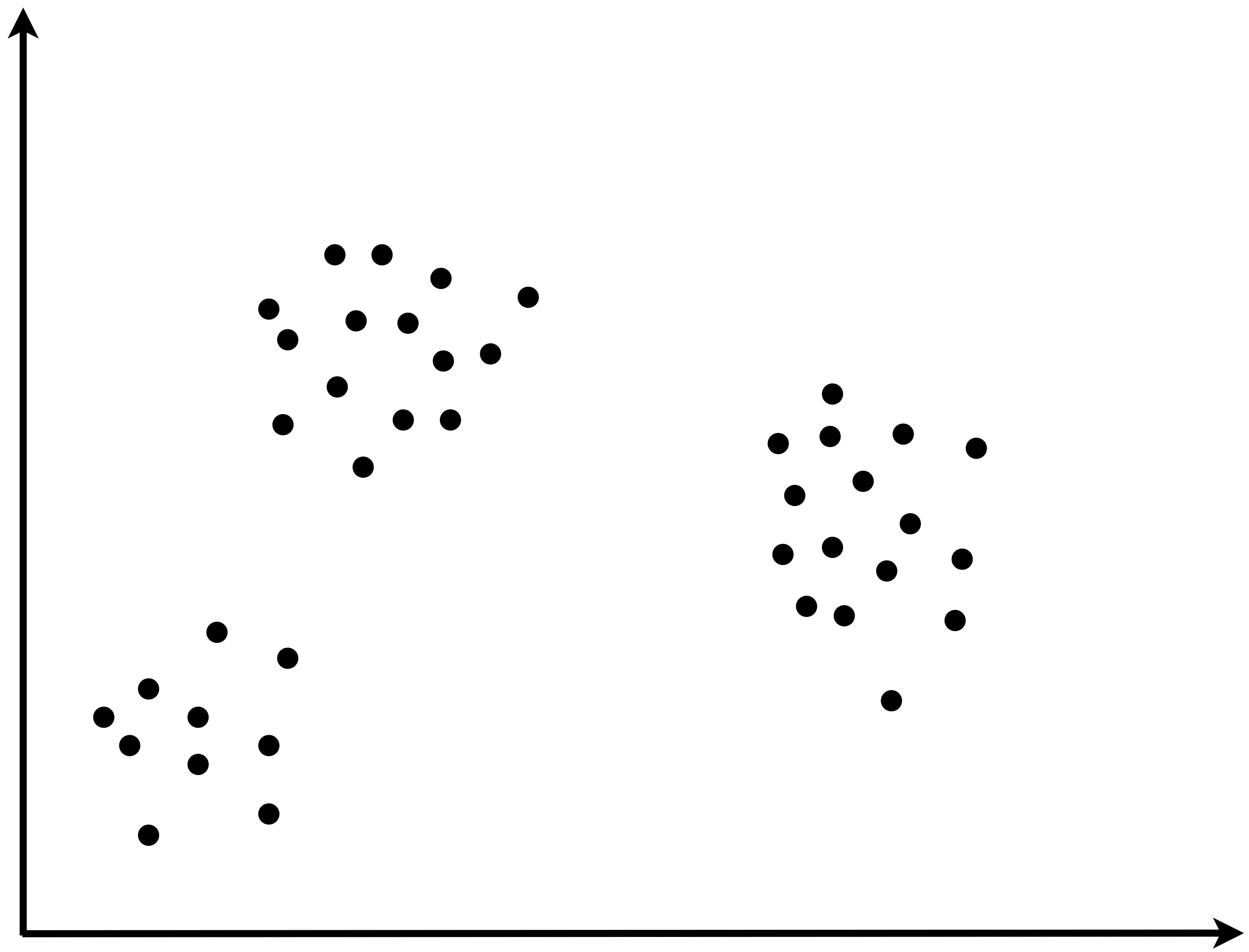


[Broderick & Cai, 2015; Cai, Campbell, Broderick, 2016; Crane & Dempsey, 2015, 2016]

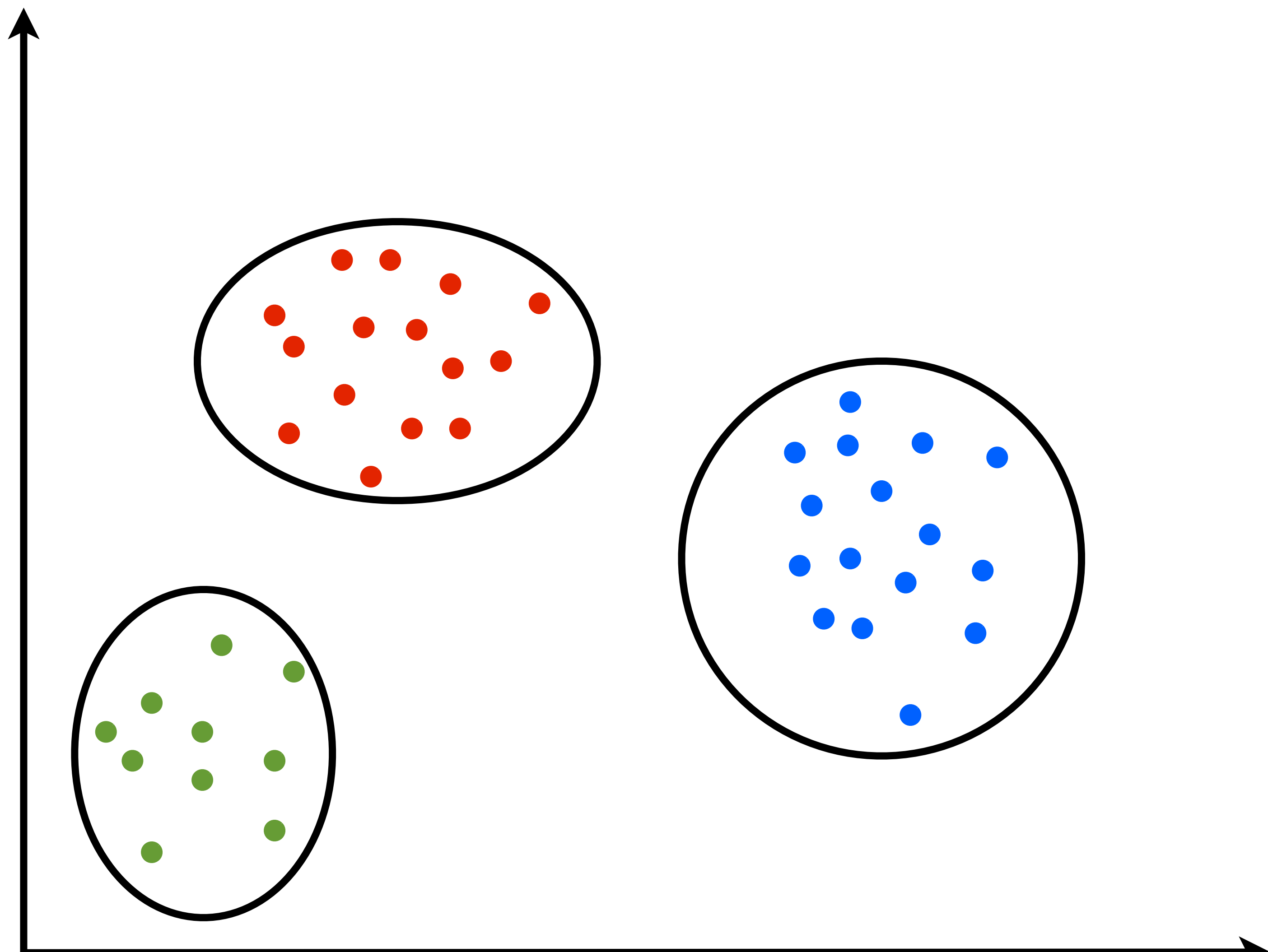
Characterizing edge-exchangeable graphs:

the **graph paintbox**

Clustering (a.k.a. partitions)

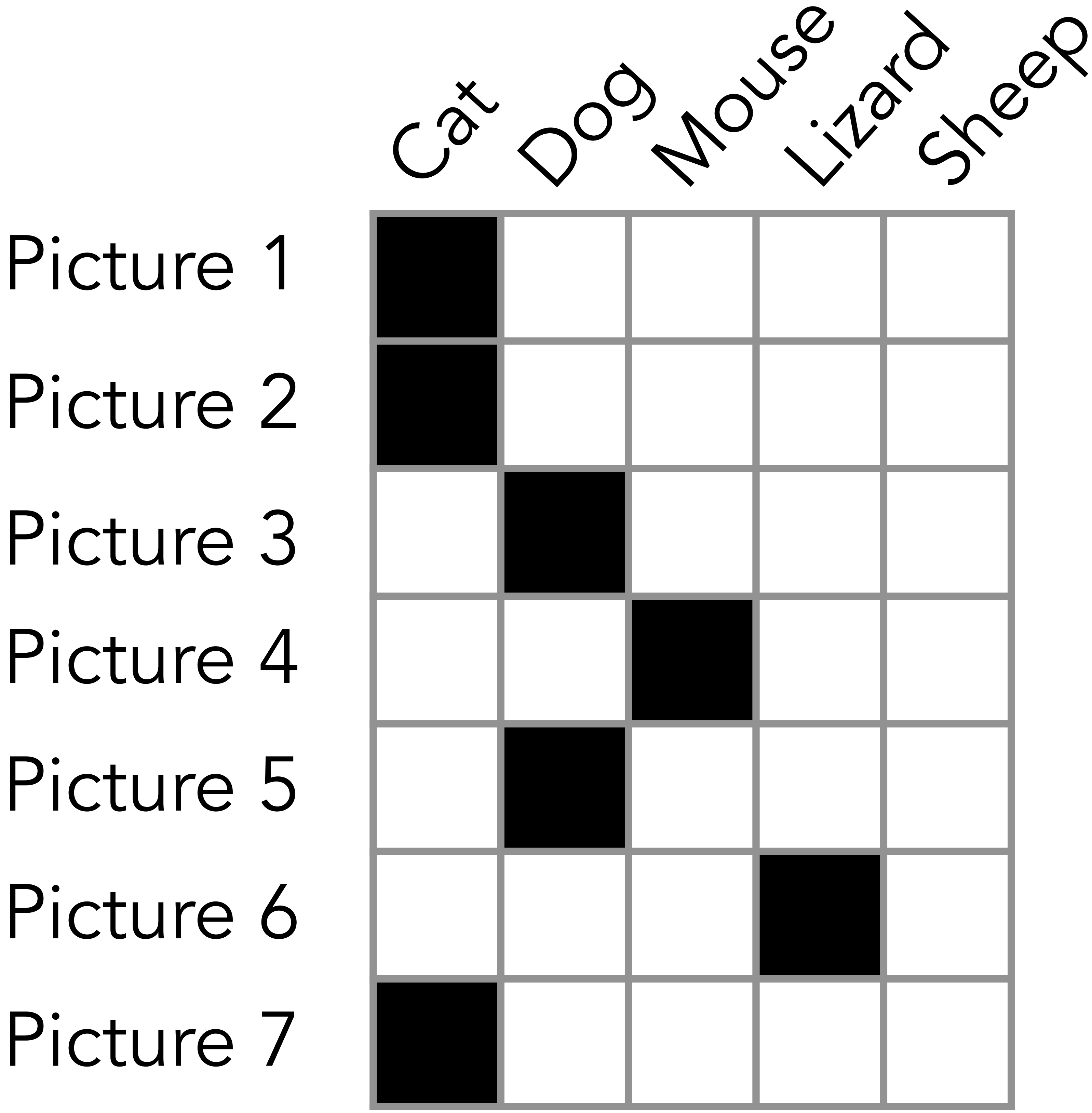


Clustering (a.k.a. partitions)



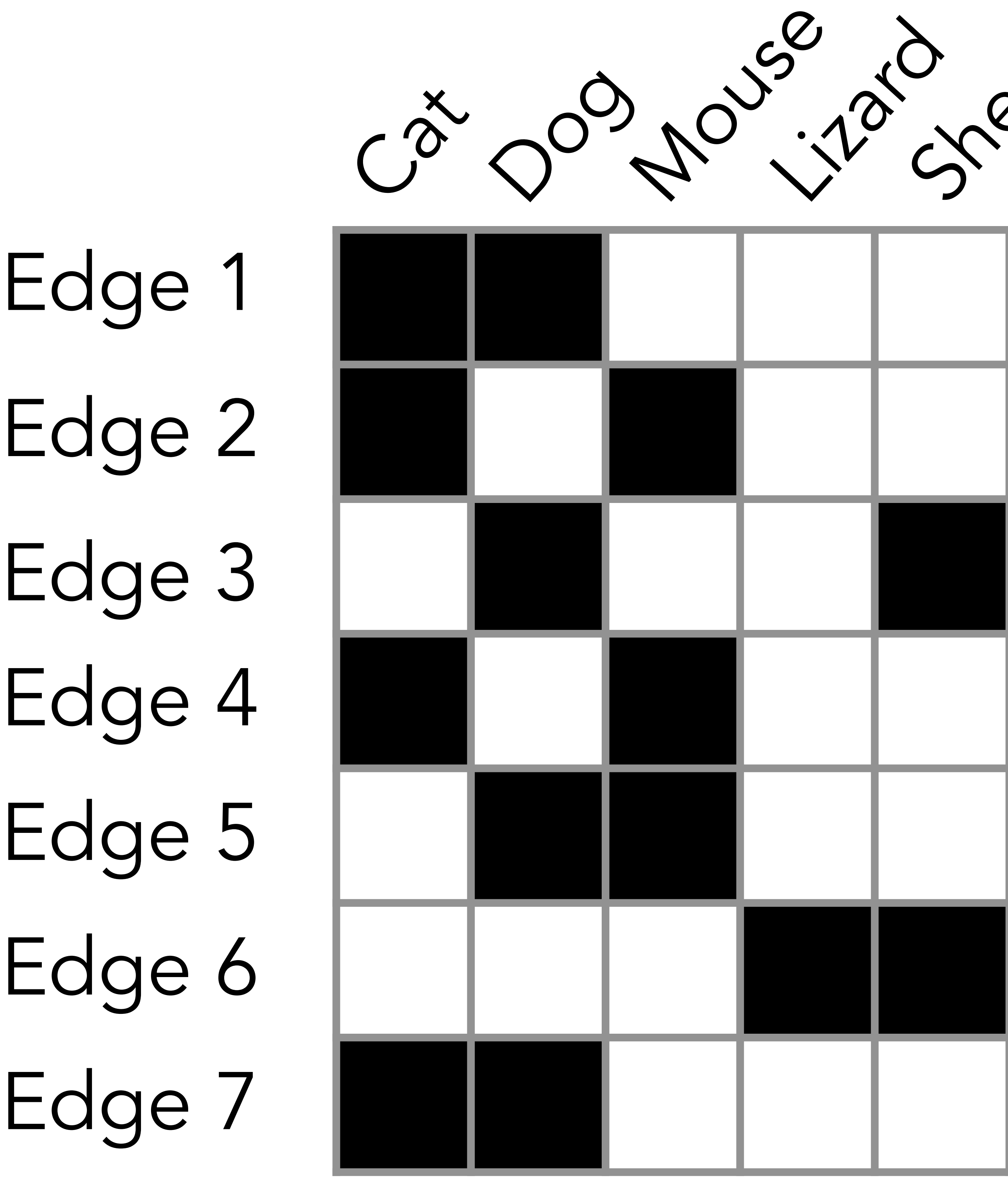
“clusters”

Clustering (a.k.a. partitions)

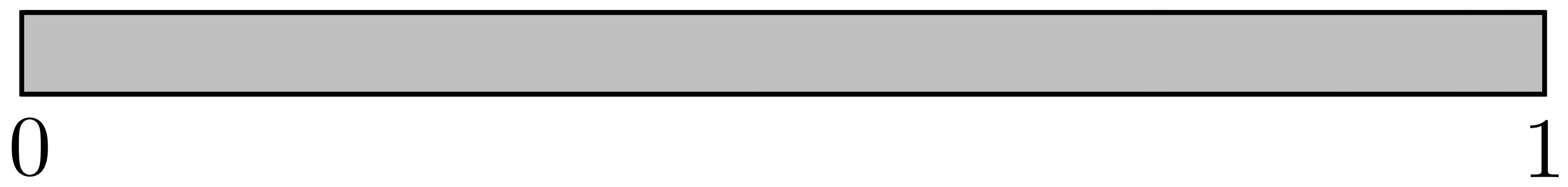


exchangeable: permuting data doesn't change distribution of the random partition

Vertex allocations

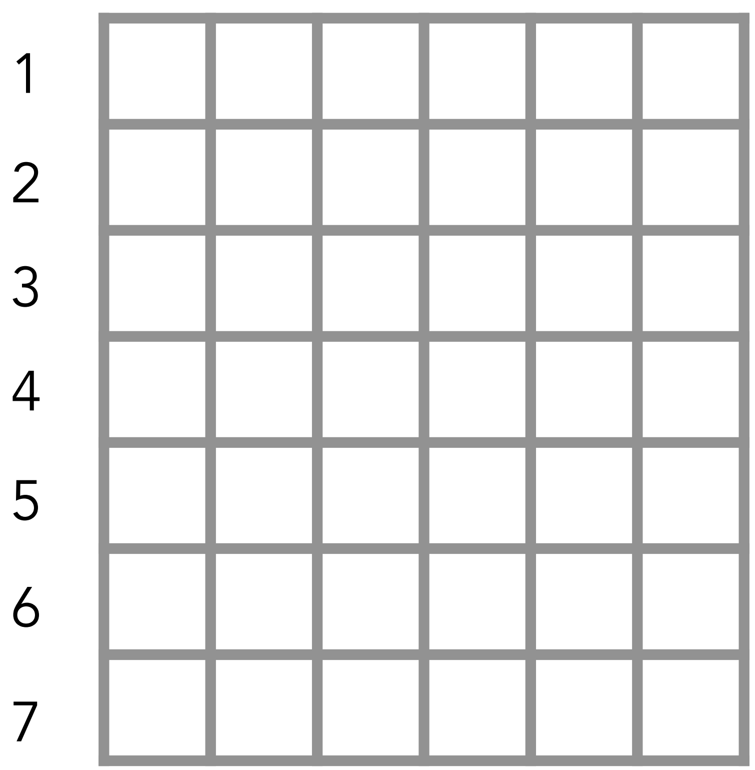


(**edge**-) **exchangeable**: permuting the edges doesn't change distribution of the random vertex allocation (graph)

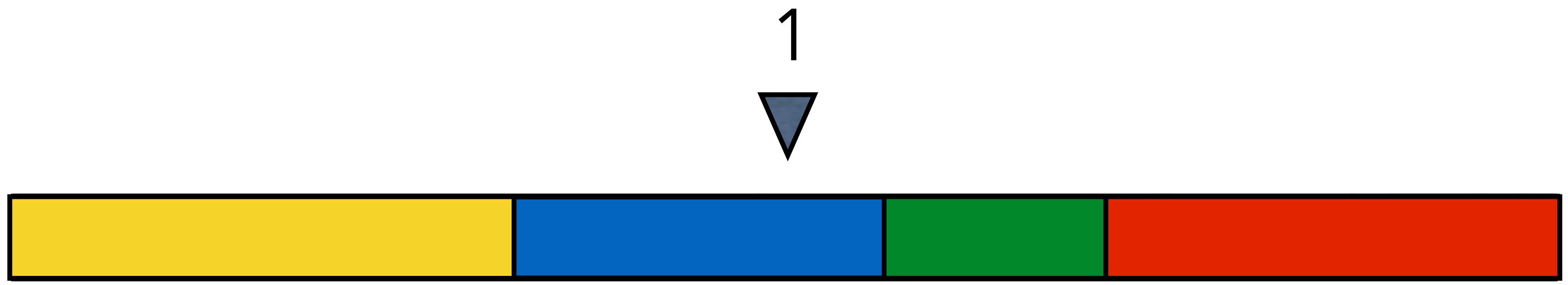


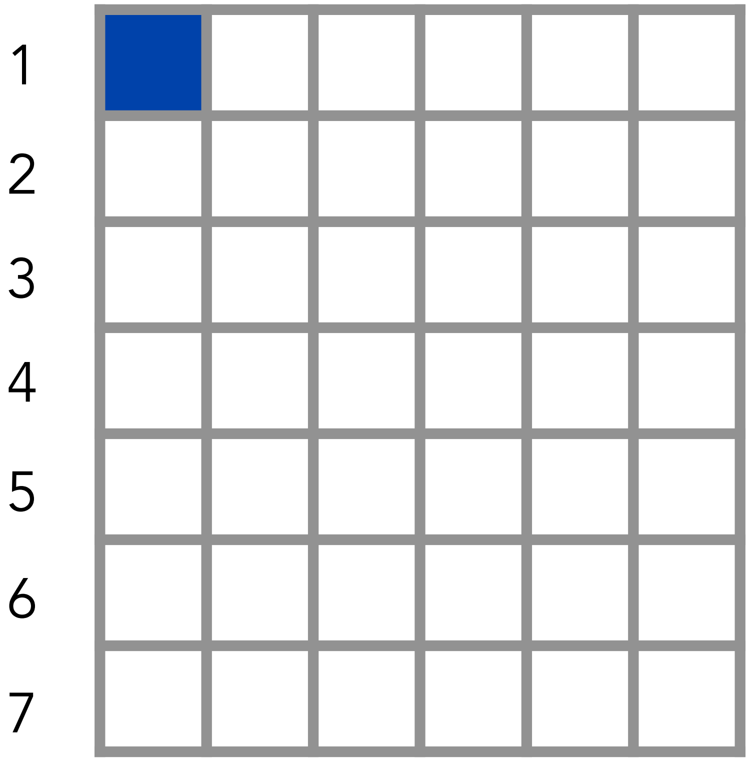
[Kingman 1978]

[Kingman 1978]

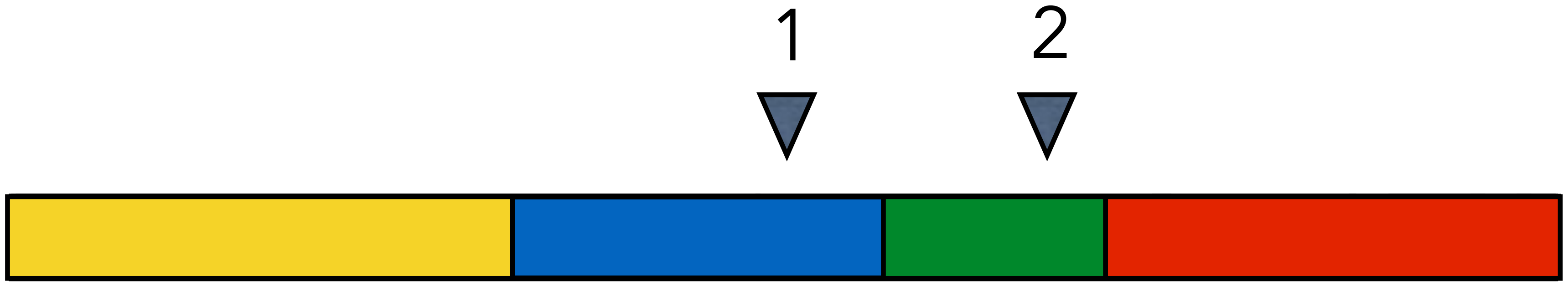


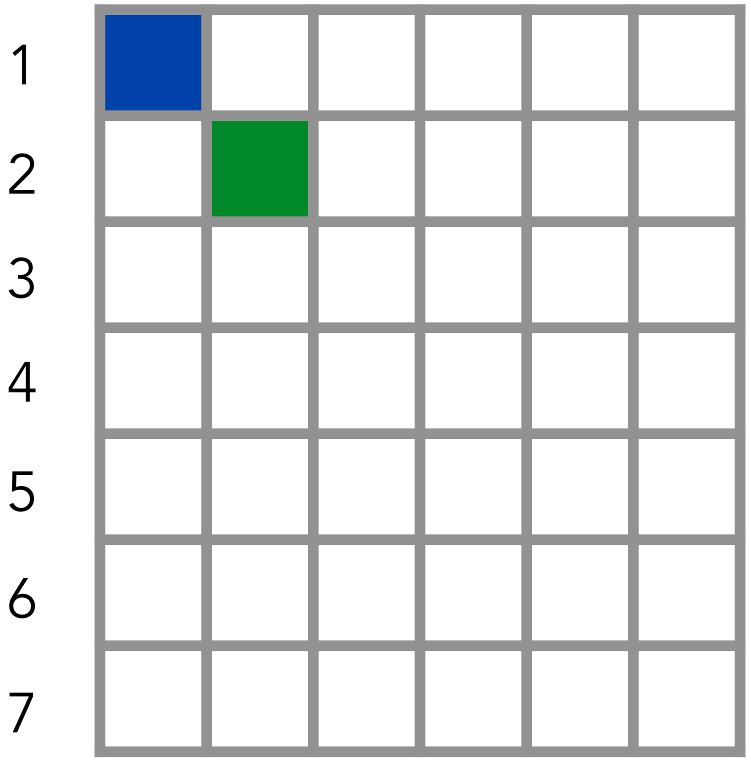
[Kingman 1978]



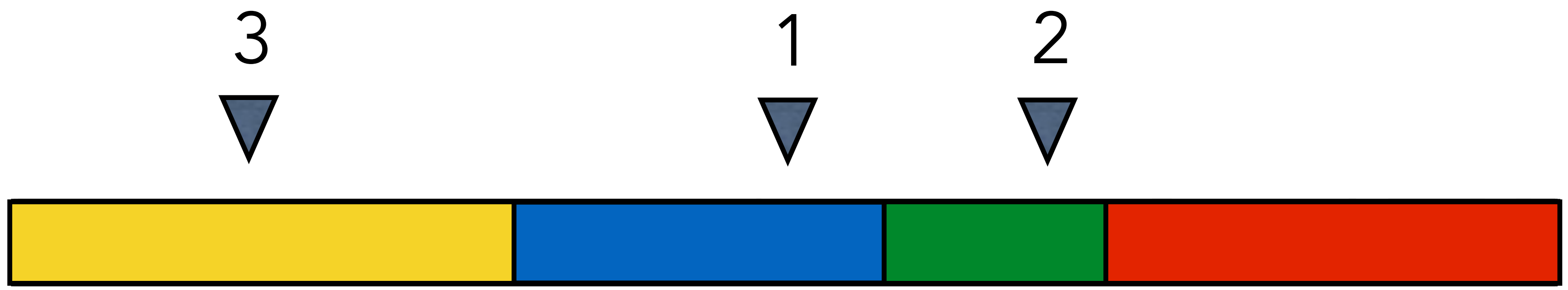


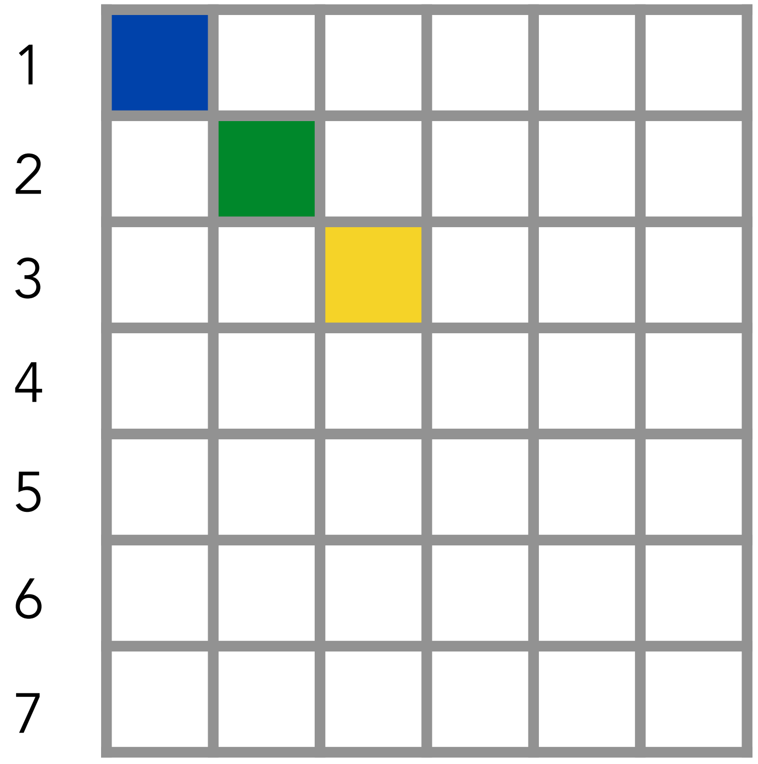
[Kingman 1978]



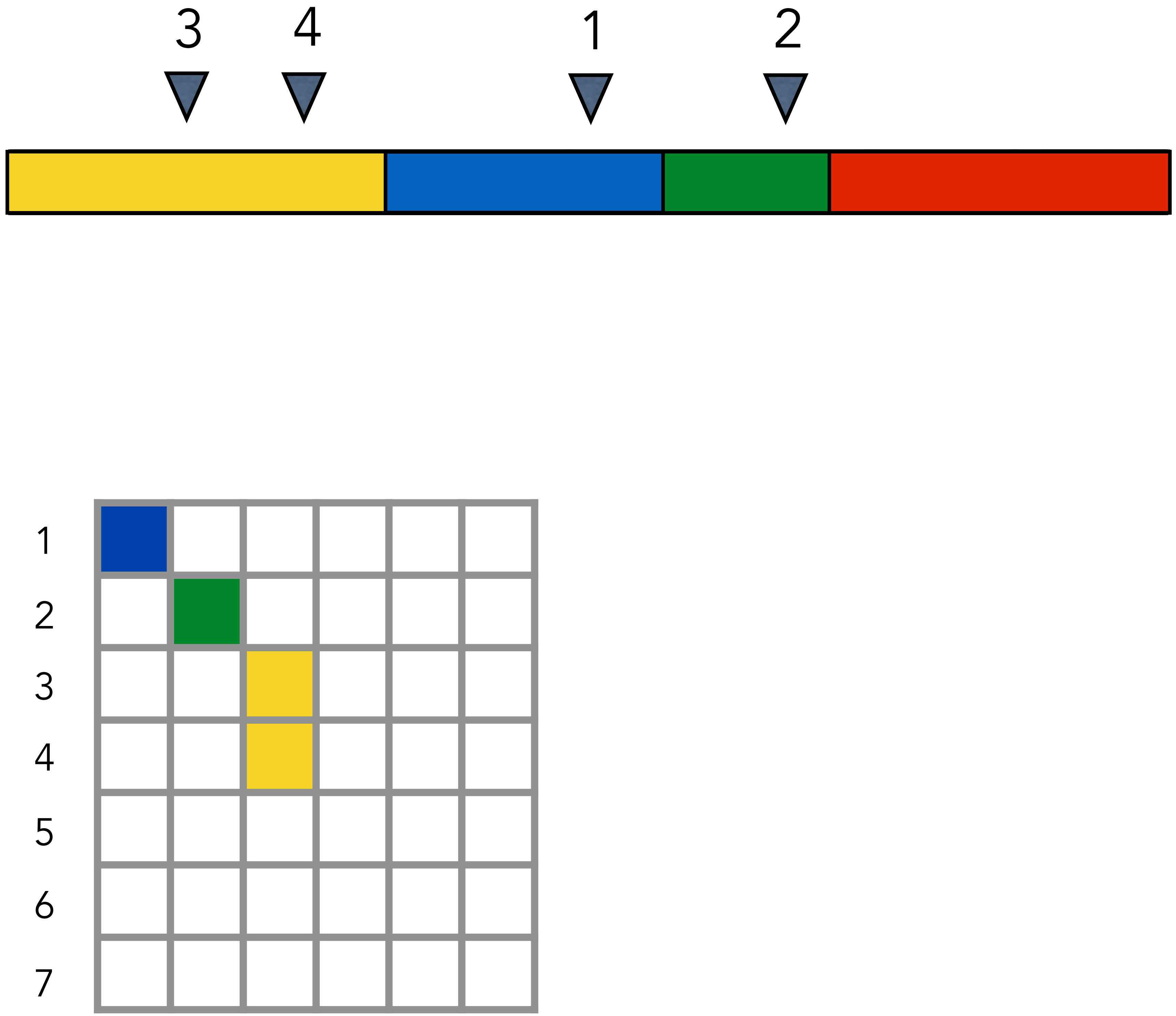


[Kingman 1978]

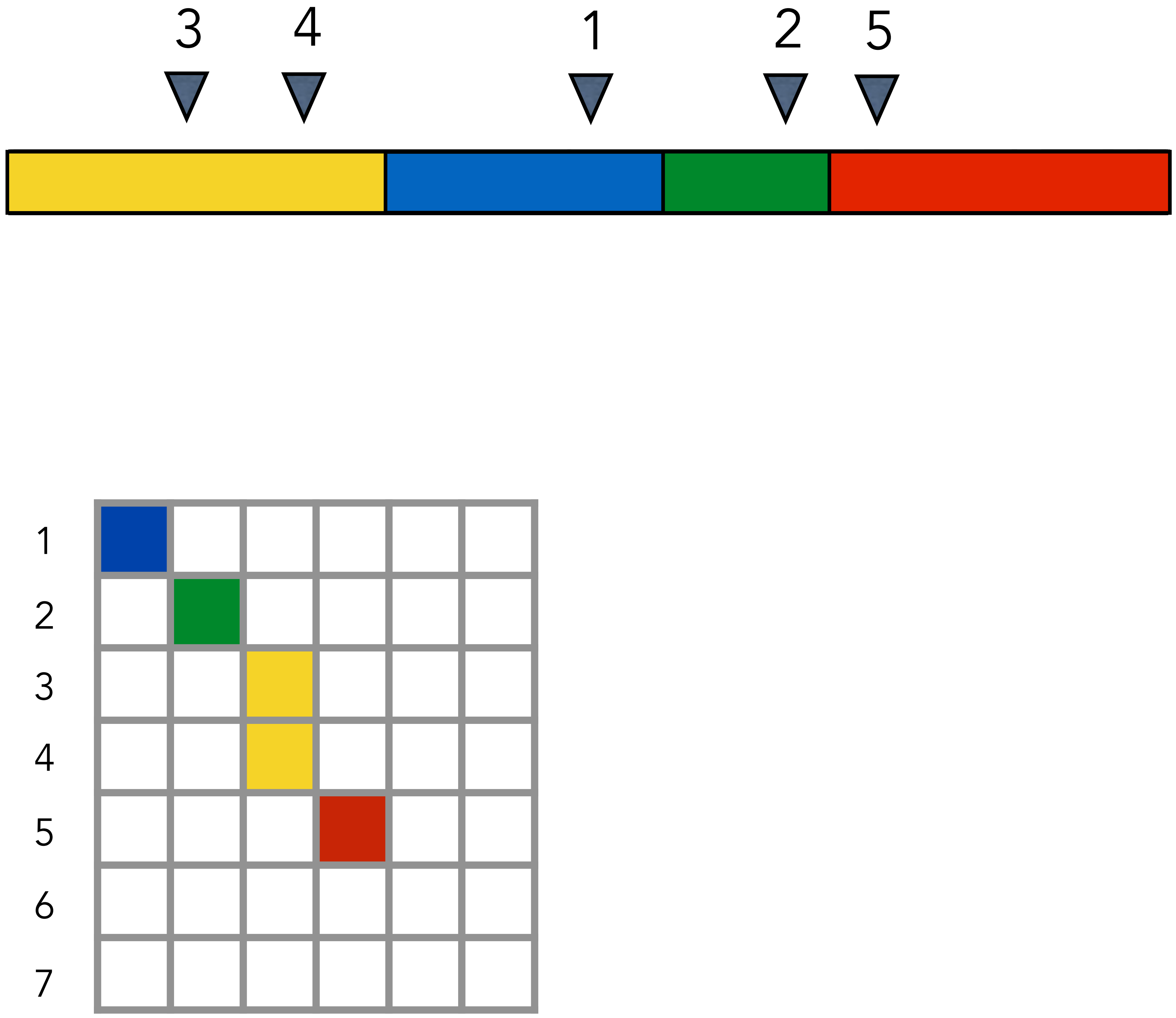




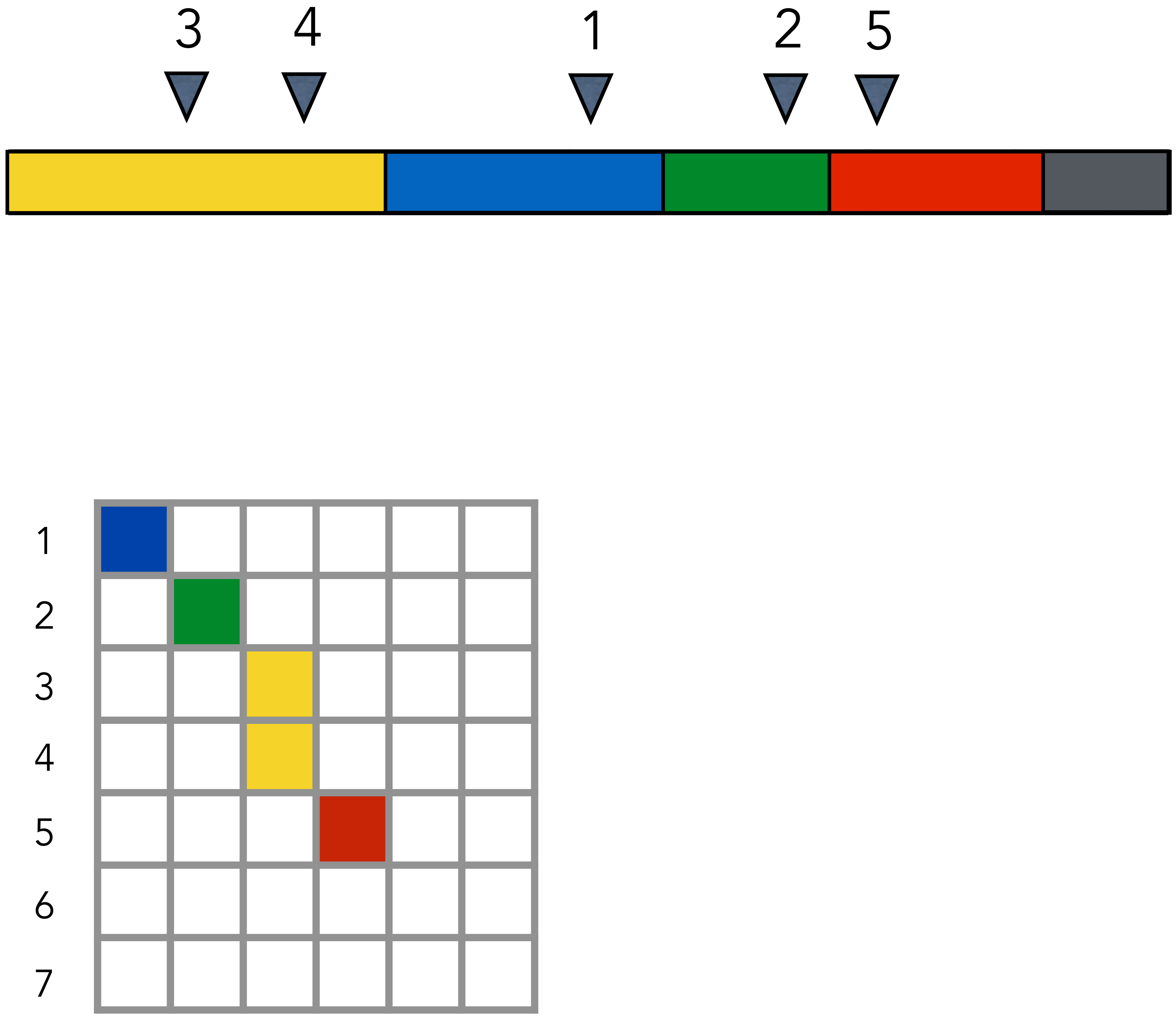
[Kingman 1978]



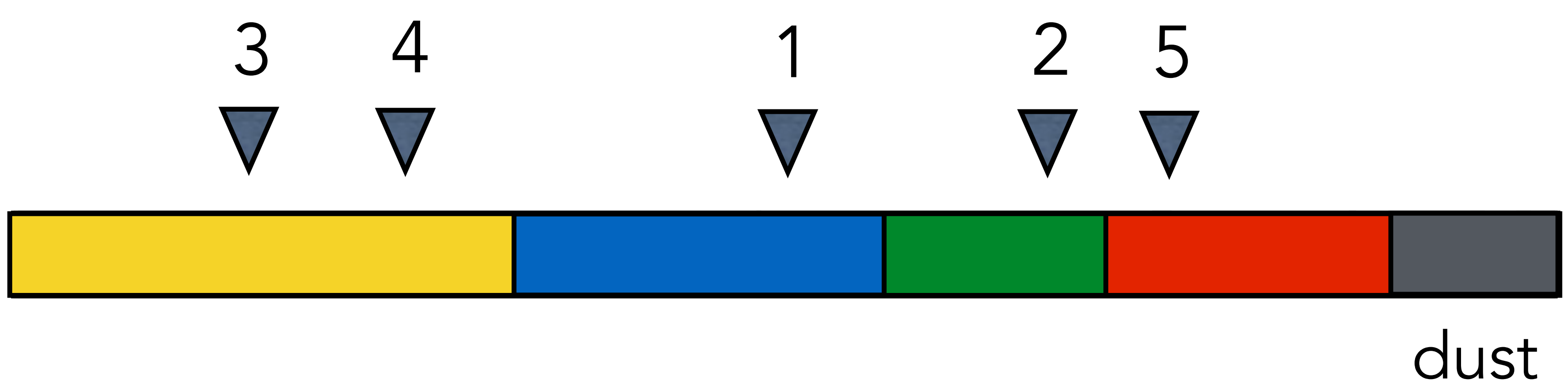
[Kingman 1978]

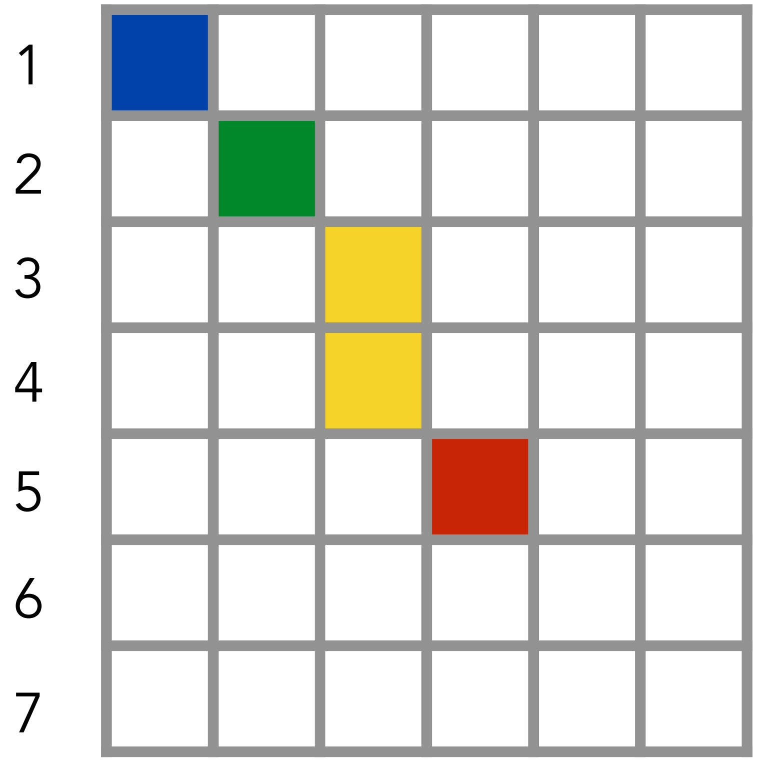


[Kingman 1978]

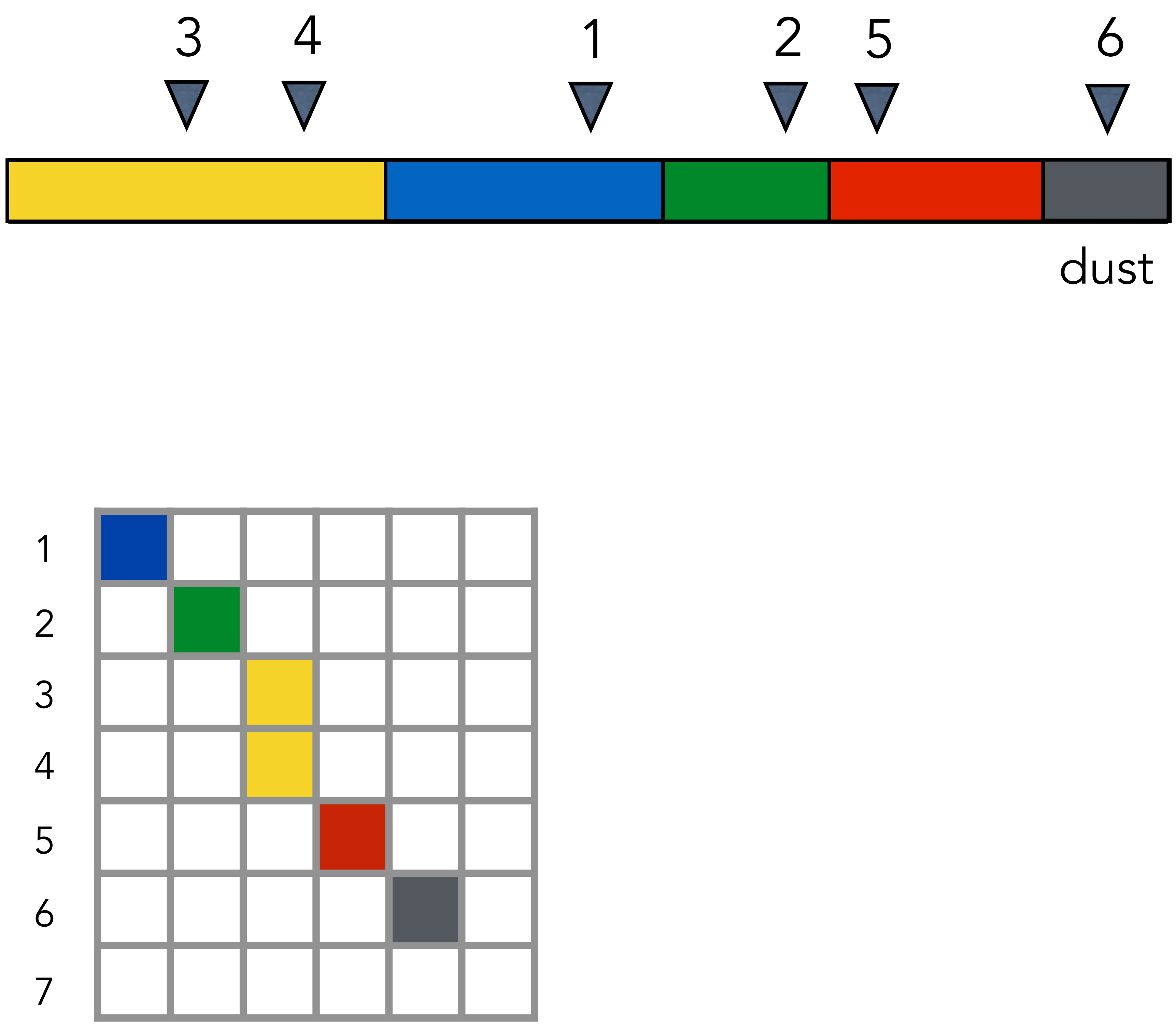


[Kingman 1978]

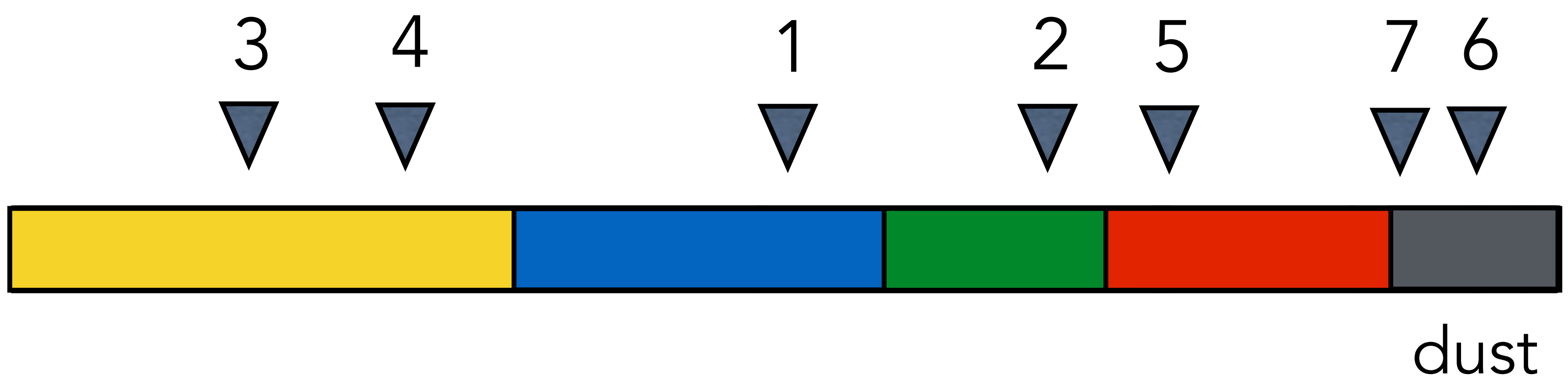


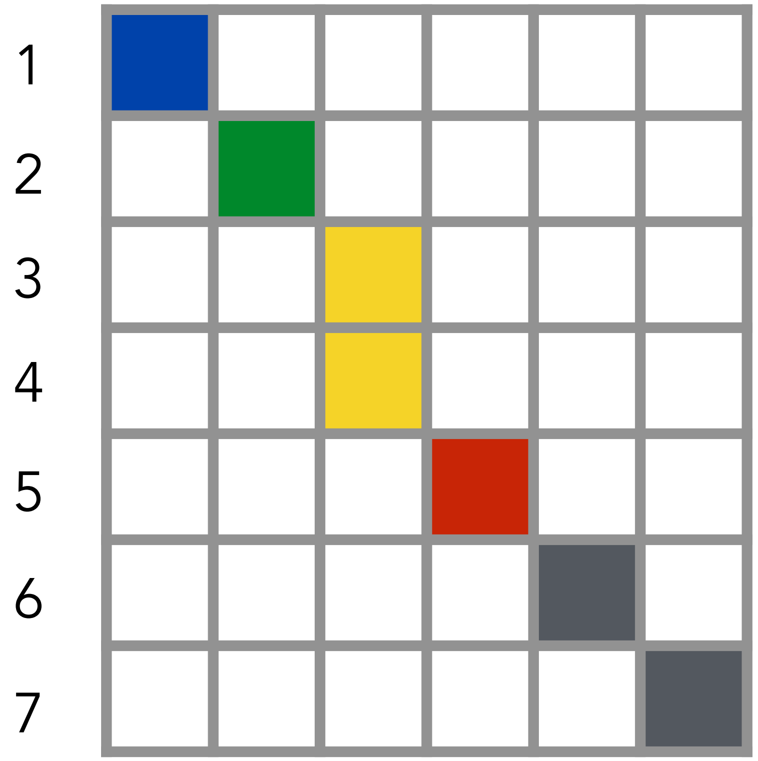


[Kingman 1978]

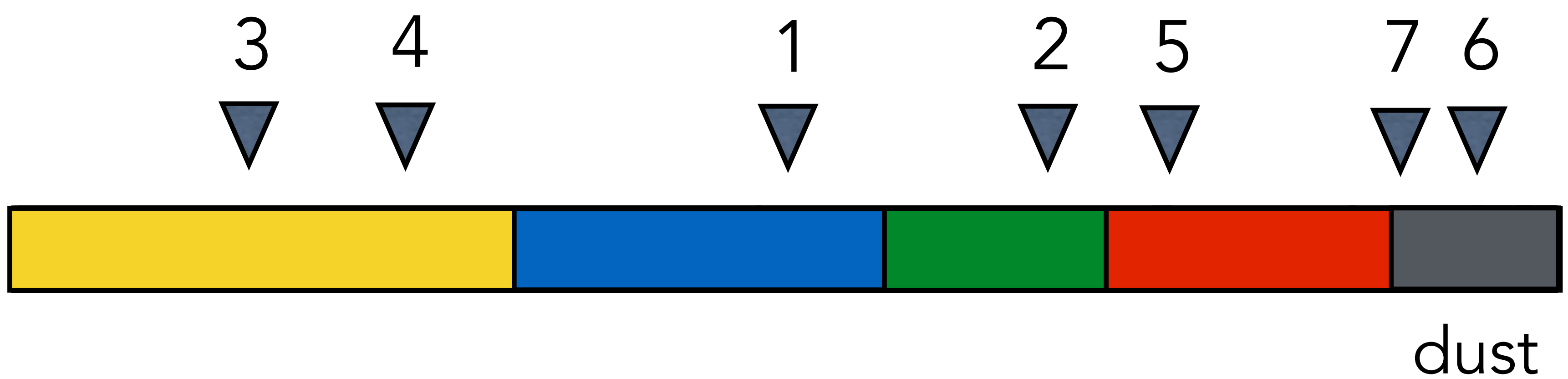


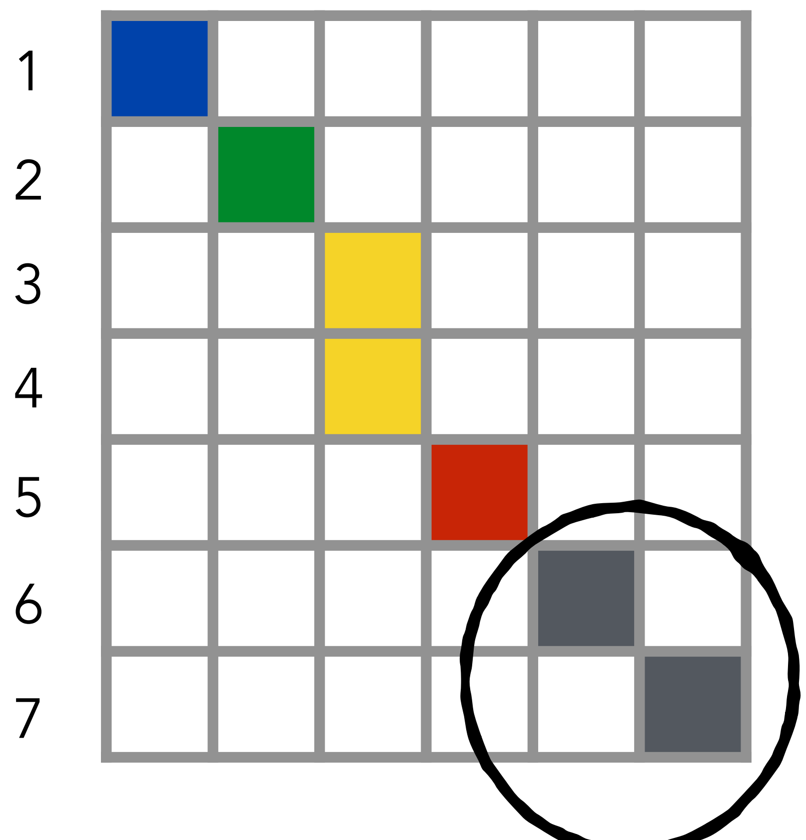
[Kingman 1978]





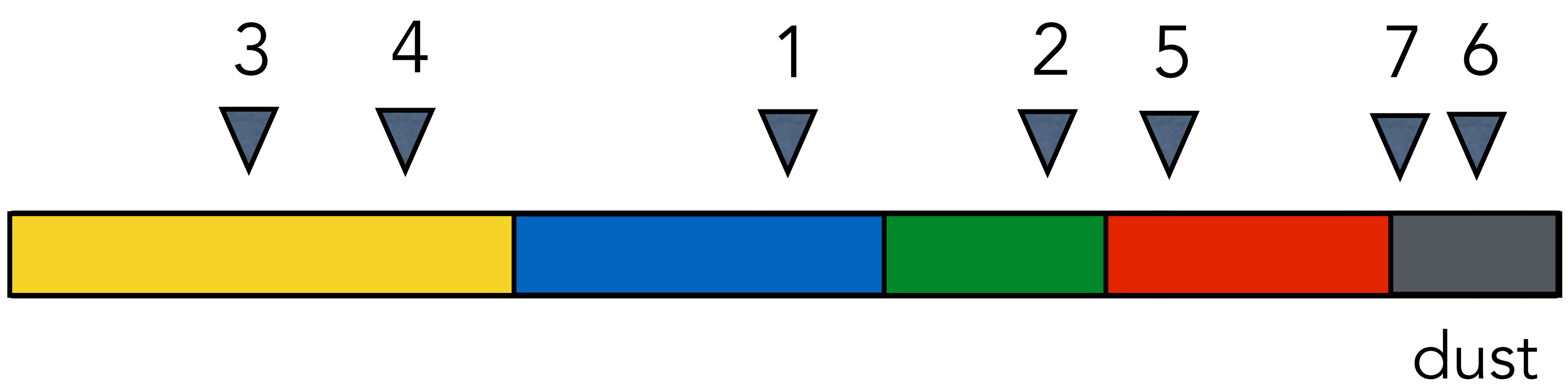
[Kingman 1978]

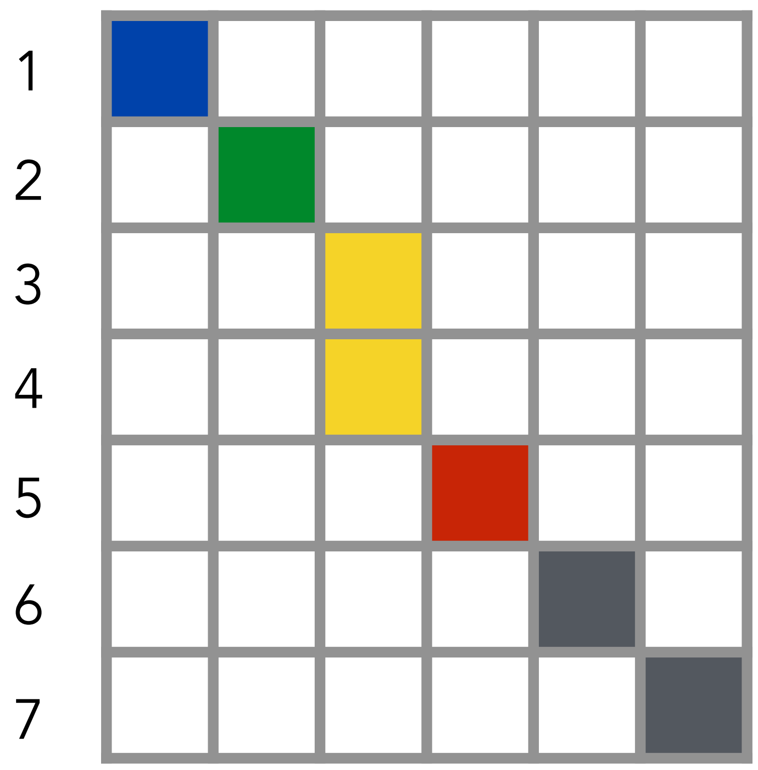




these clusters only appear in a single data point

[Kingman 1978]

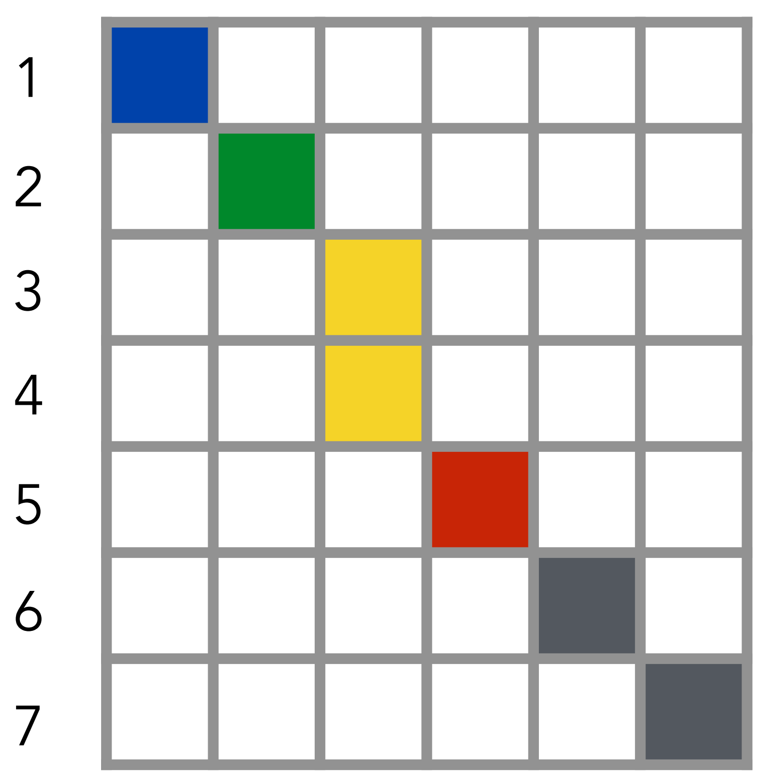




This clustering is *exchangeable*.

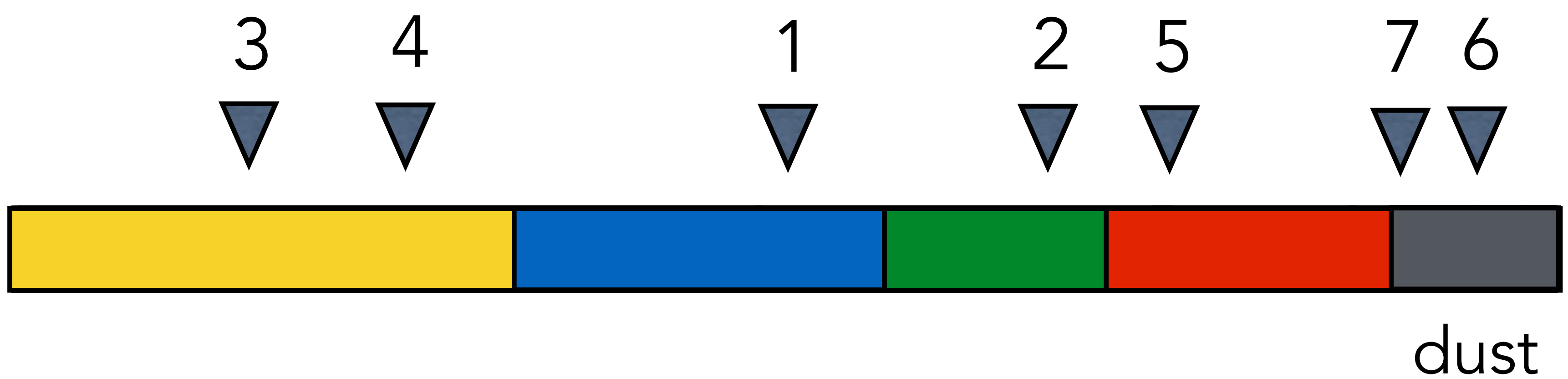
[Kingman 1978]

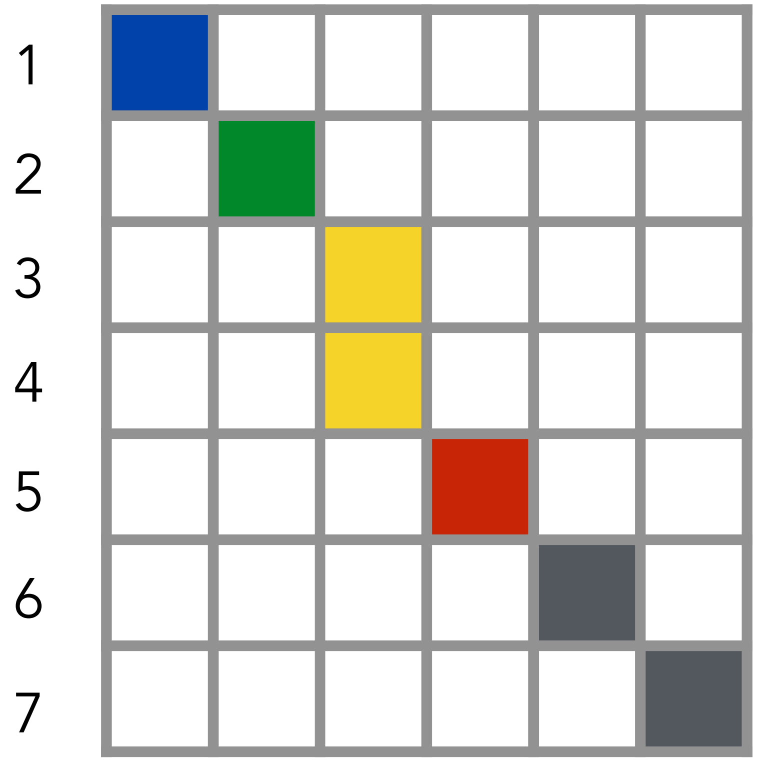
Theorem (Kingman, 1978).
A random clustering is exchangeable iff it has a Kingman paintbox representation.



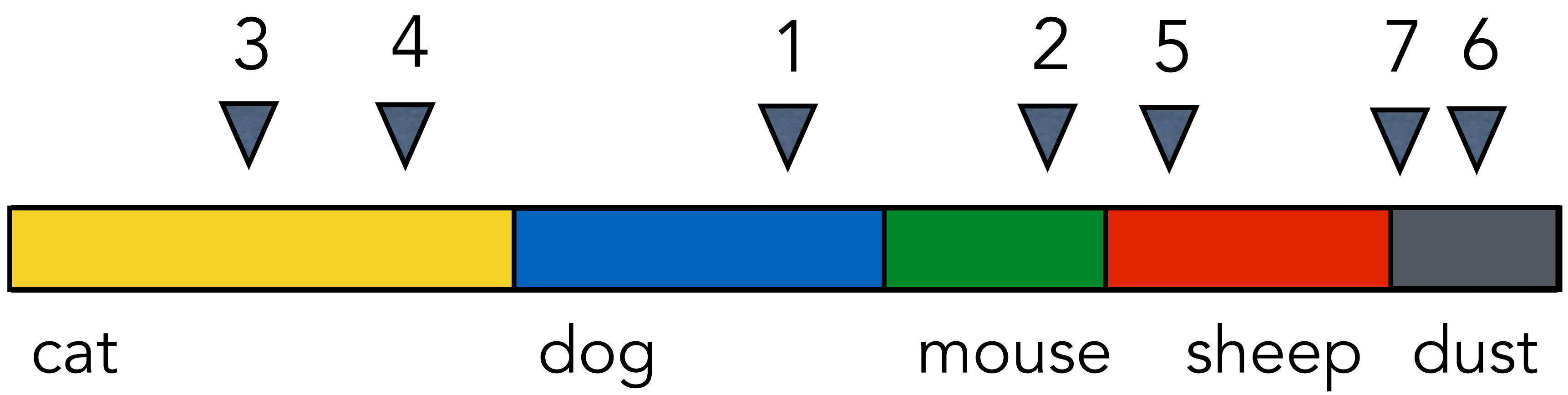
This clustering is *exchangeable*.

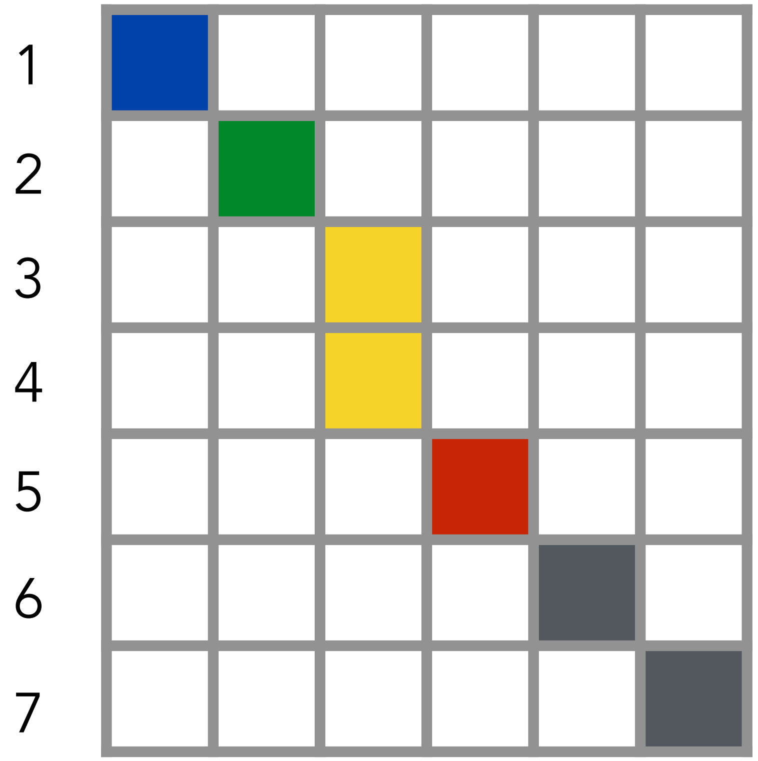
[Kingman 1978]



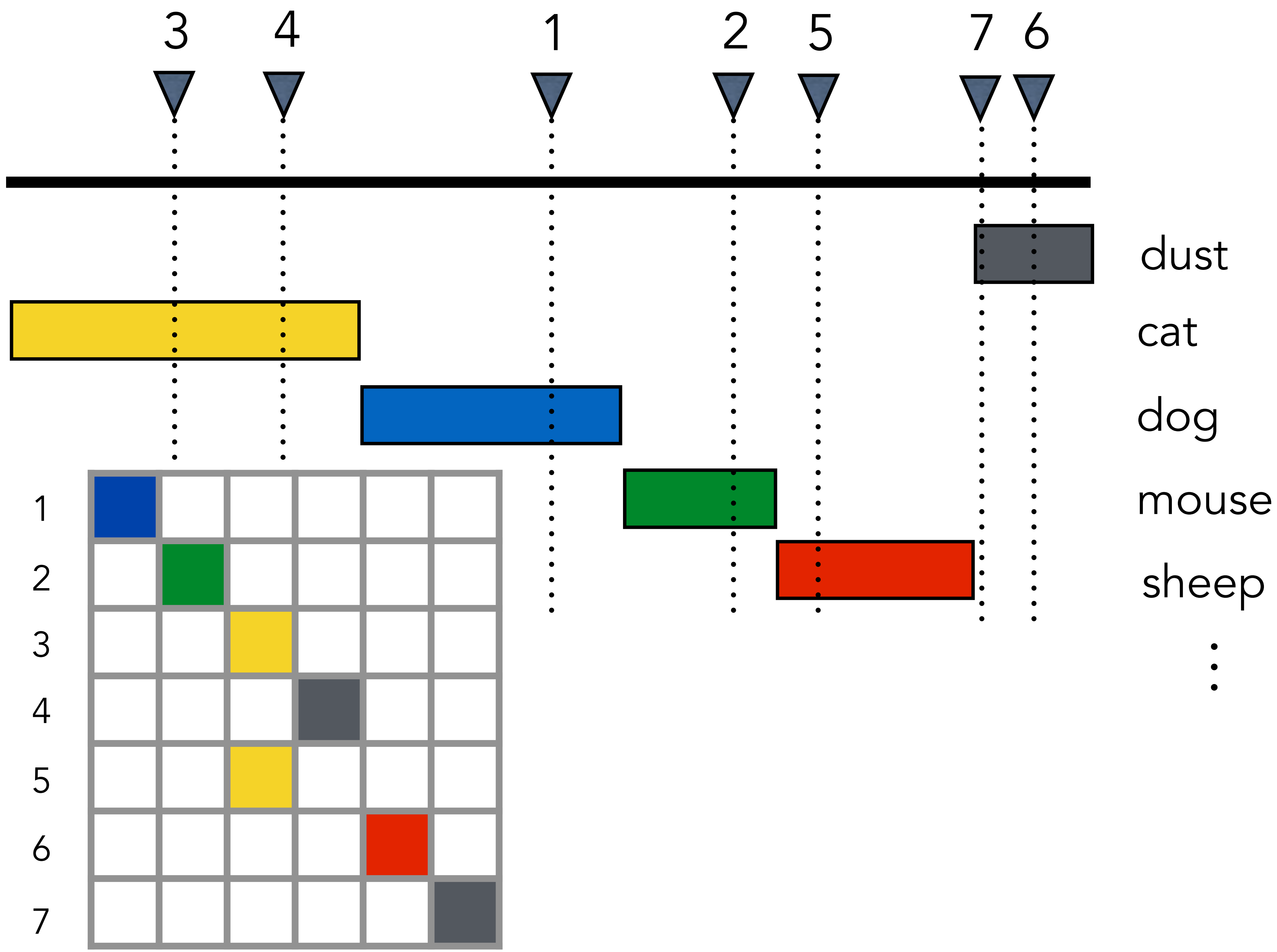


[Kingman 1978]

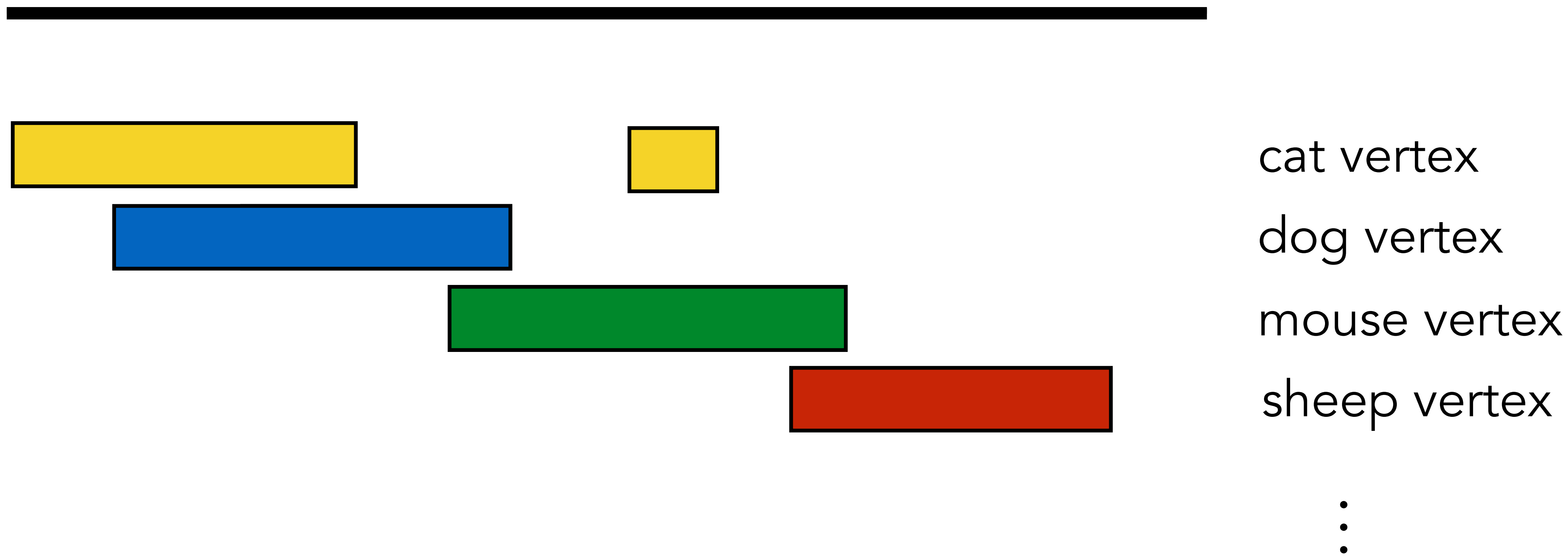




[Kingman 1978]

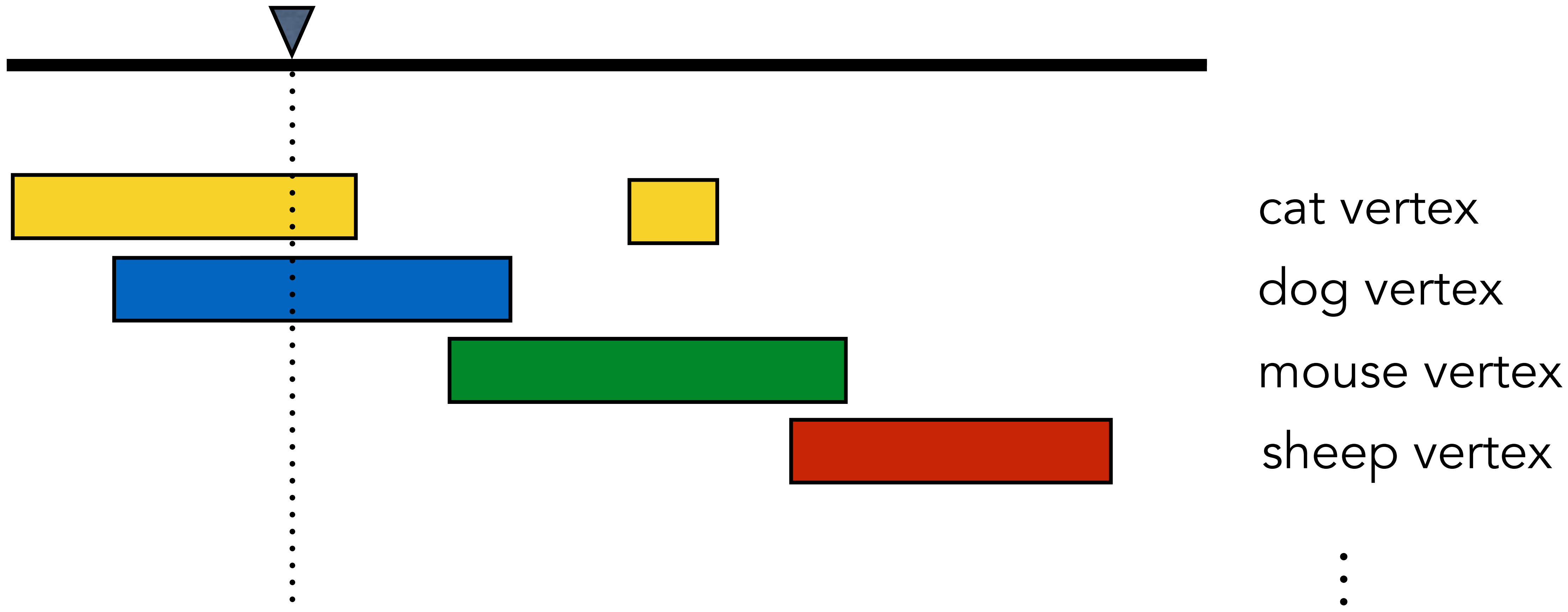


3
4
1
2
5
7
6
dust
cat
dog
mouse
sheep
1
2
3
4
5
6
7



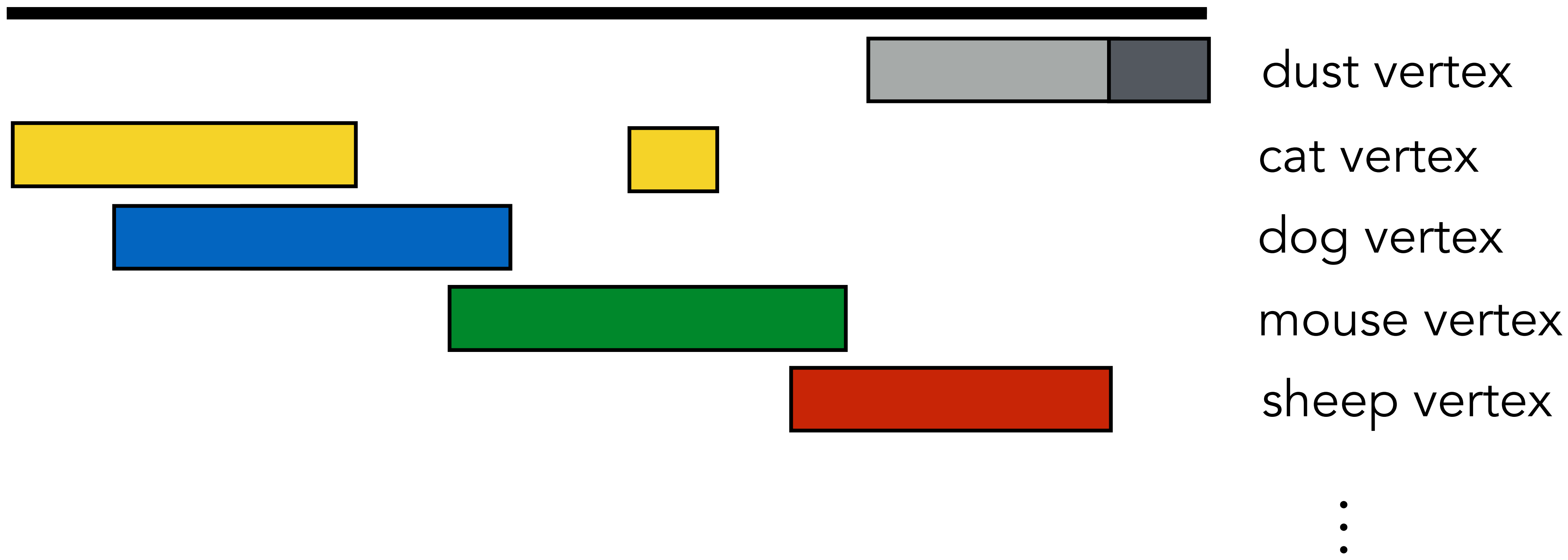
[Campbell, Cai, Broderick, 2016; Cai, Campbell, Broderick, 2016b]

any point intersects at most 2 subsets of (0,1)

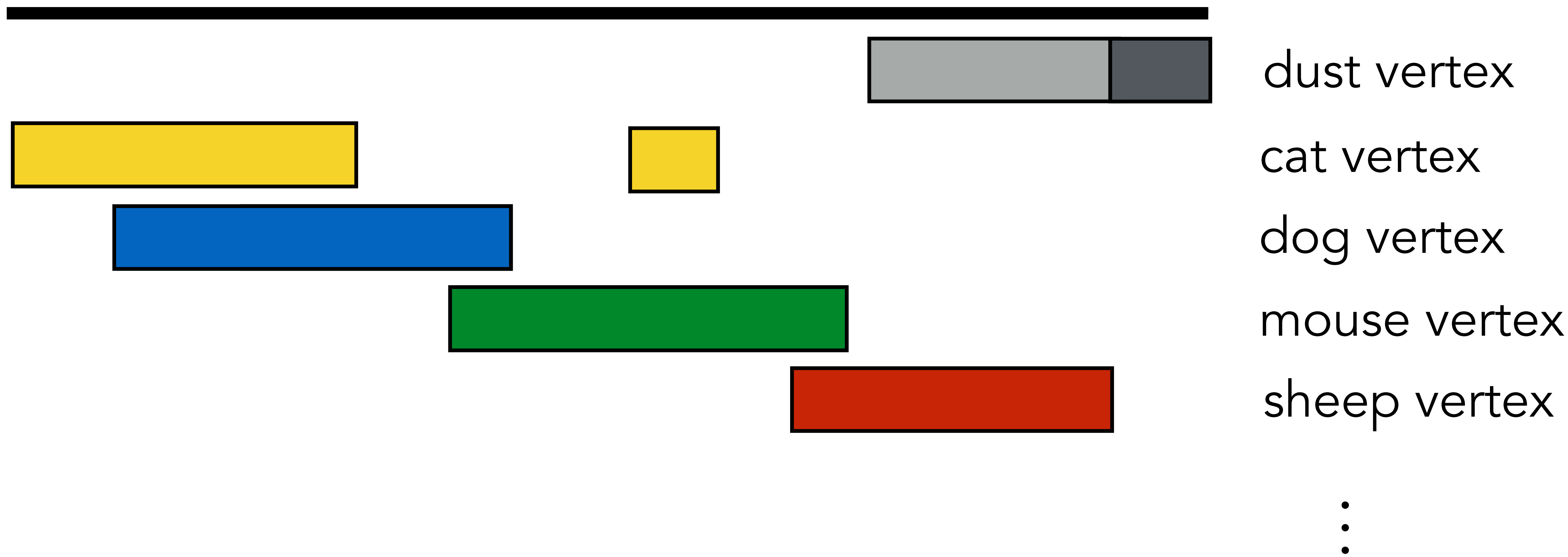


"cat and dog just interacted"

[Campbell, Cai, Broderick, 2016; Cai, Campbell, Broderick, 2016b]

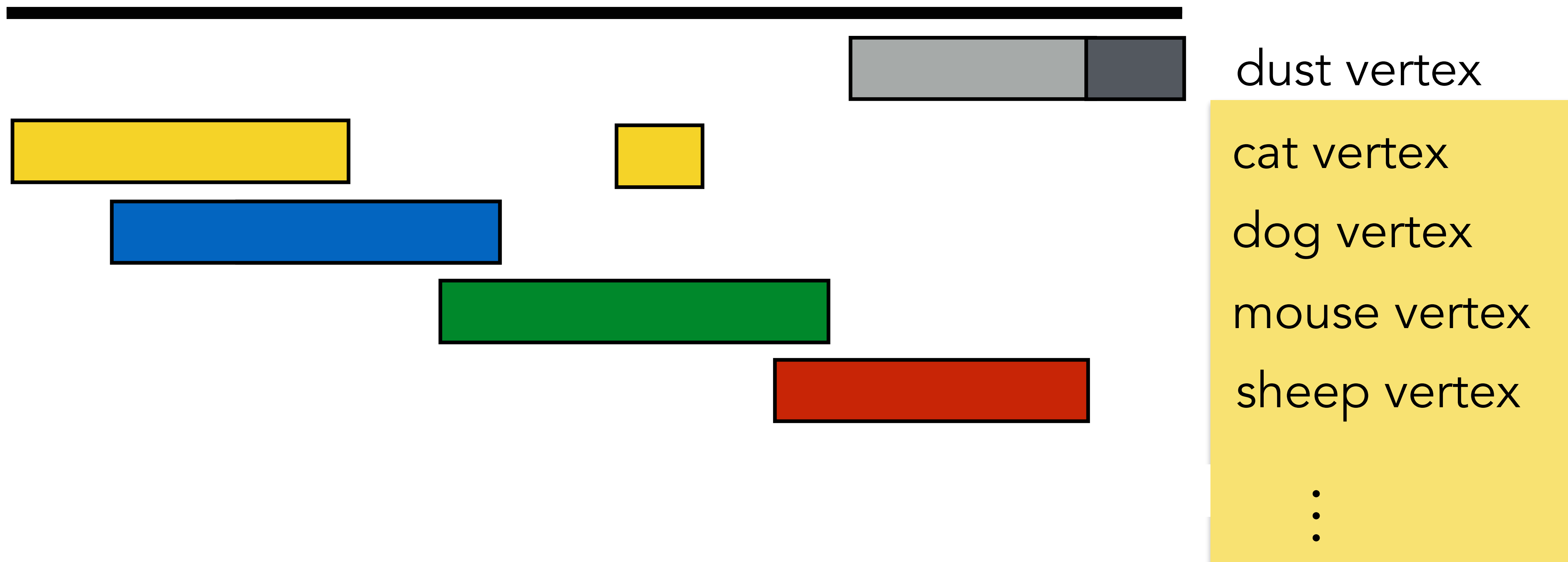


[Campbell, Cai, Broderick, 2016; Cai, Campbell, Broderick, 2016b]



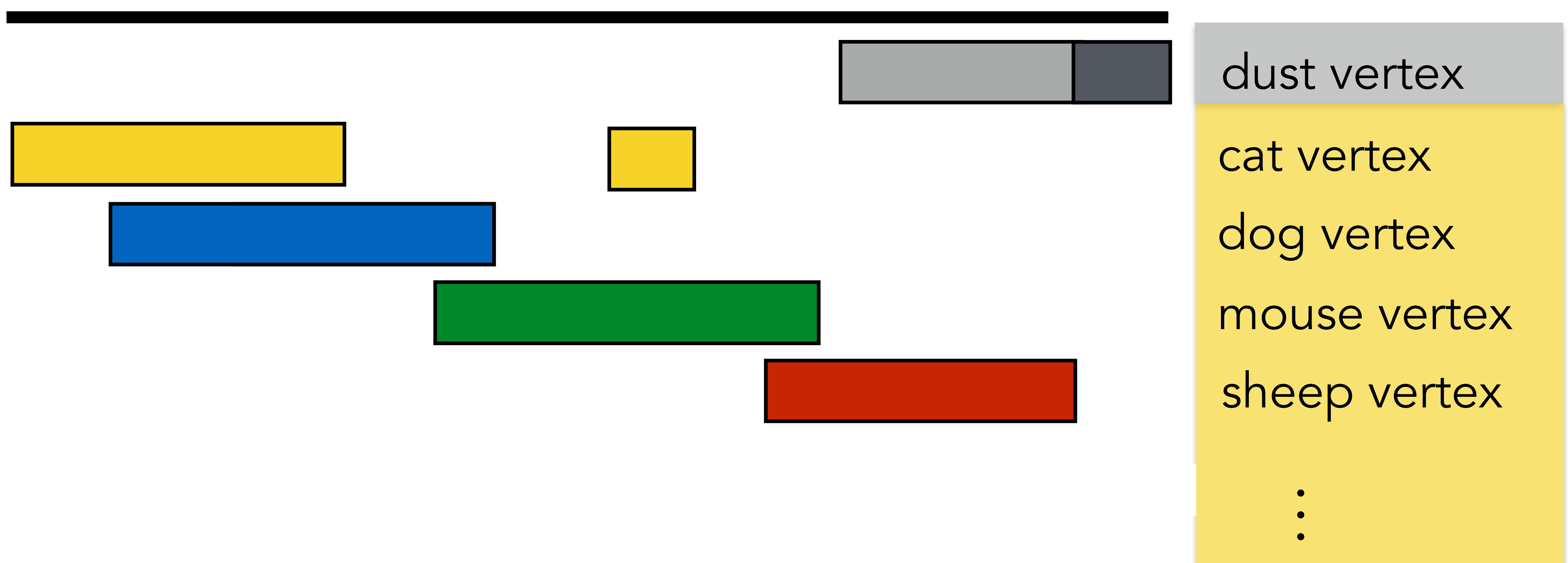
“regular” (colored) vs “dust” (gray) vertices

[Campbell, Cai, Broderick, 2016; Cai, Campbell, Broderick, 2016b]



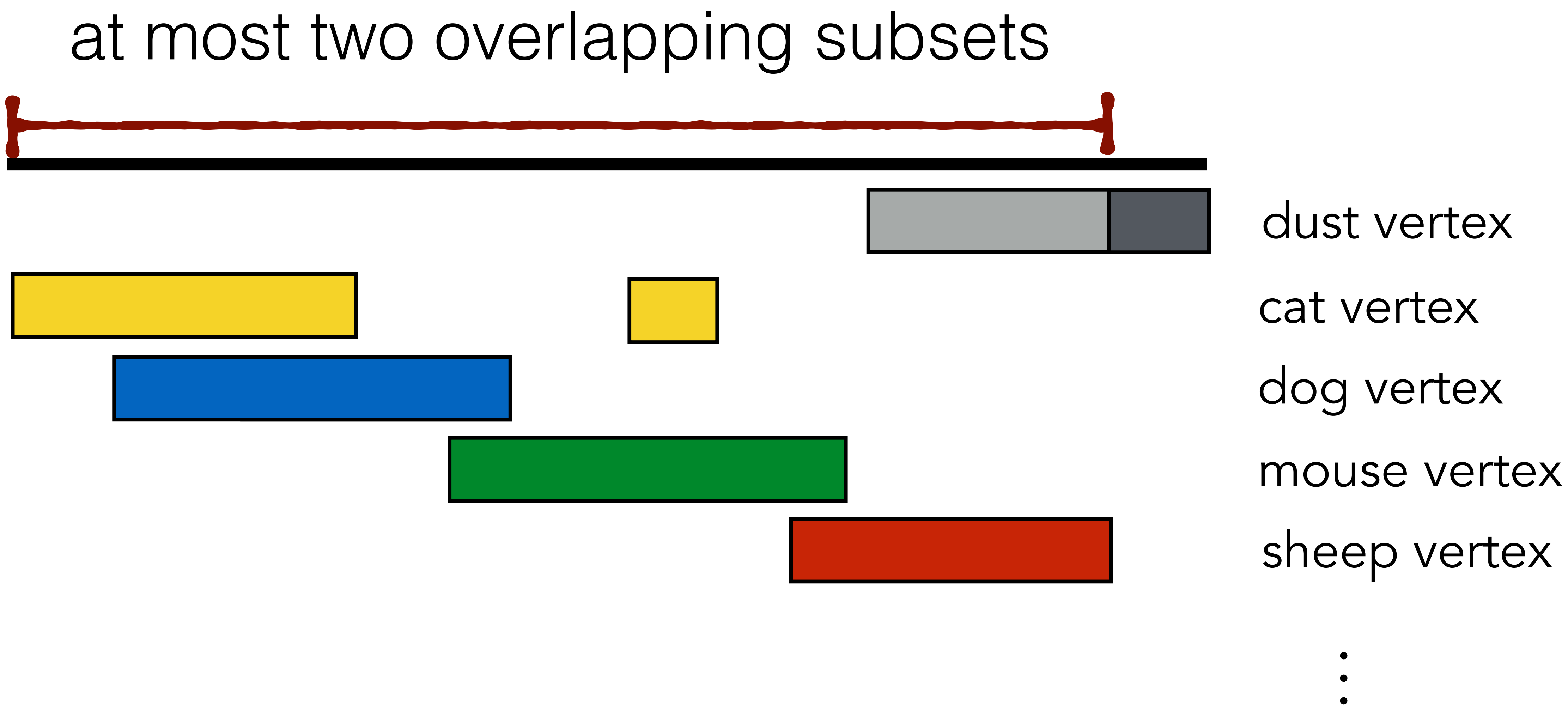
“regular” (colored) vs “dust” (gray) vertices

[Campbell, Cai, Broderick, 2016; Cai, Campbell, Broderick, 2016b]

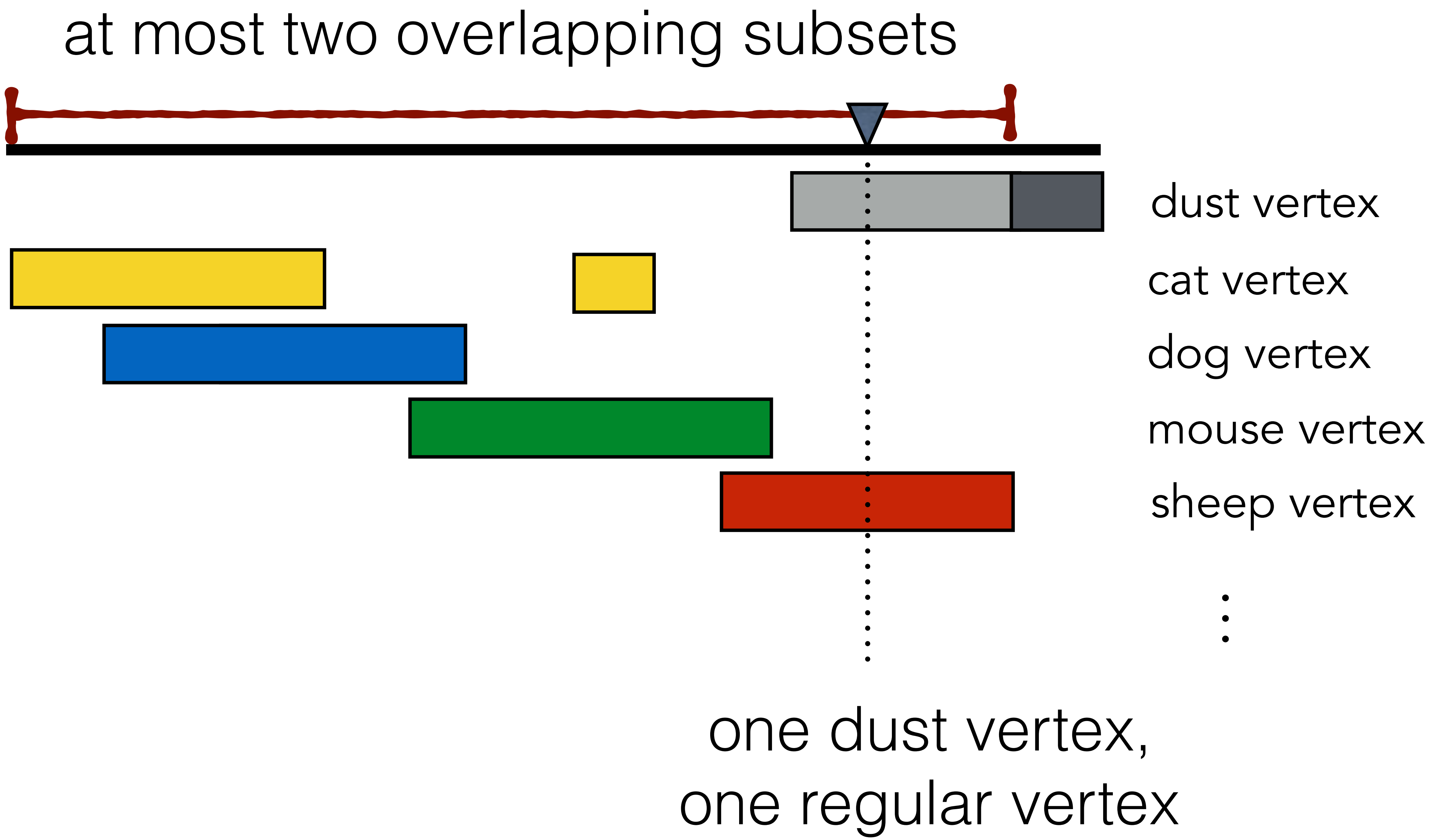


“regular” (colored) vs “dust” (gray) vertices

[Campbell, Cai, Broderick, 2016; Cai, Campbell, Broderick, 2016b]

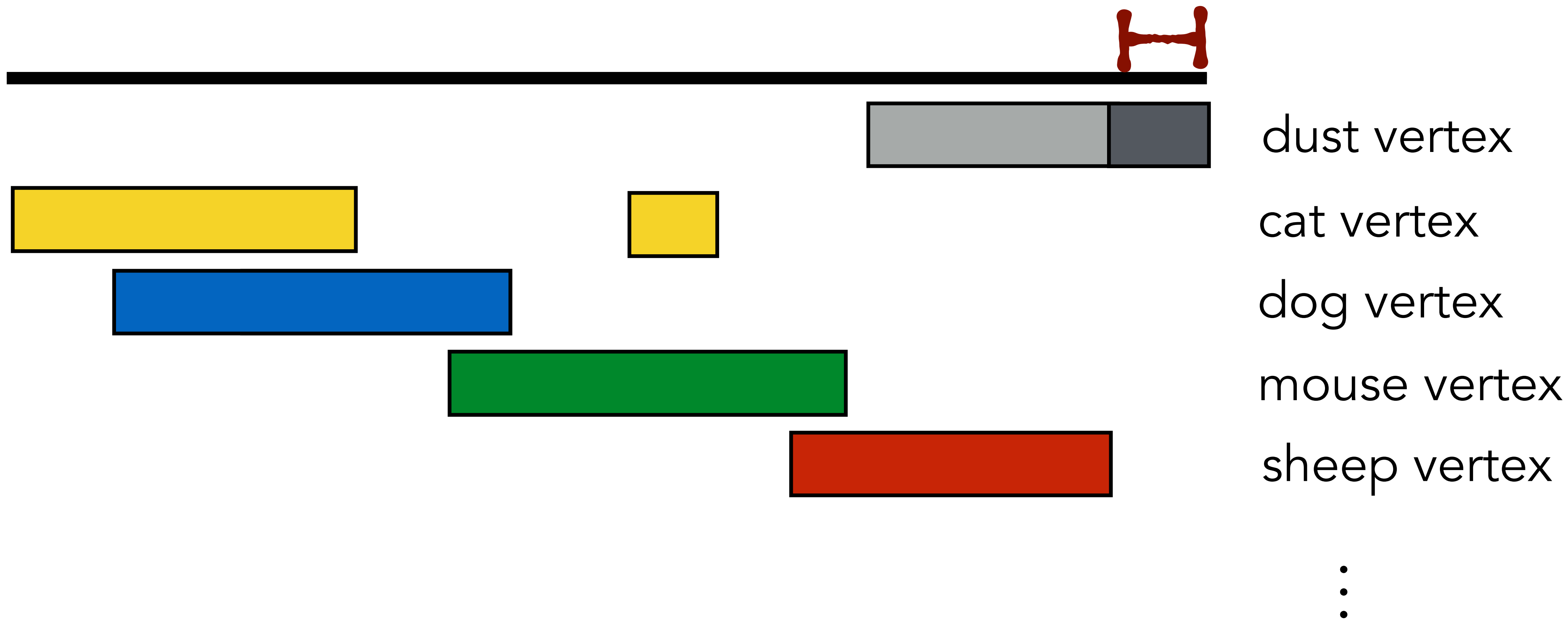


[Campbell, Cai, Broderick, 2016; Cai, Campbell, Broderick, 2016b]

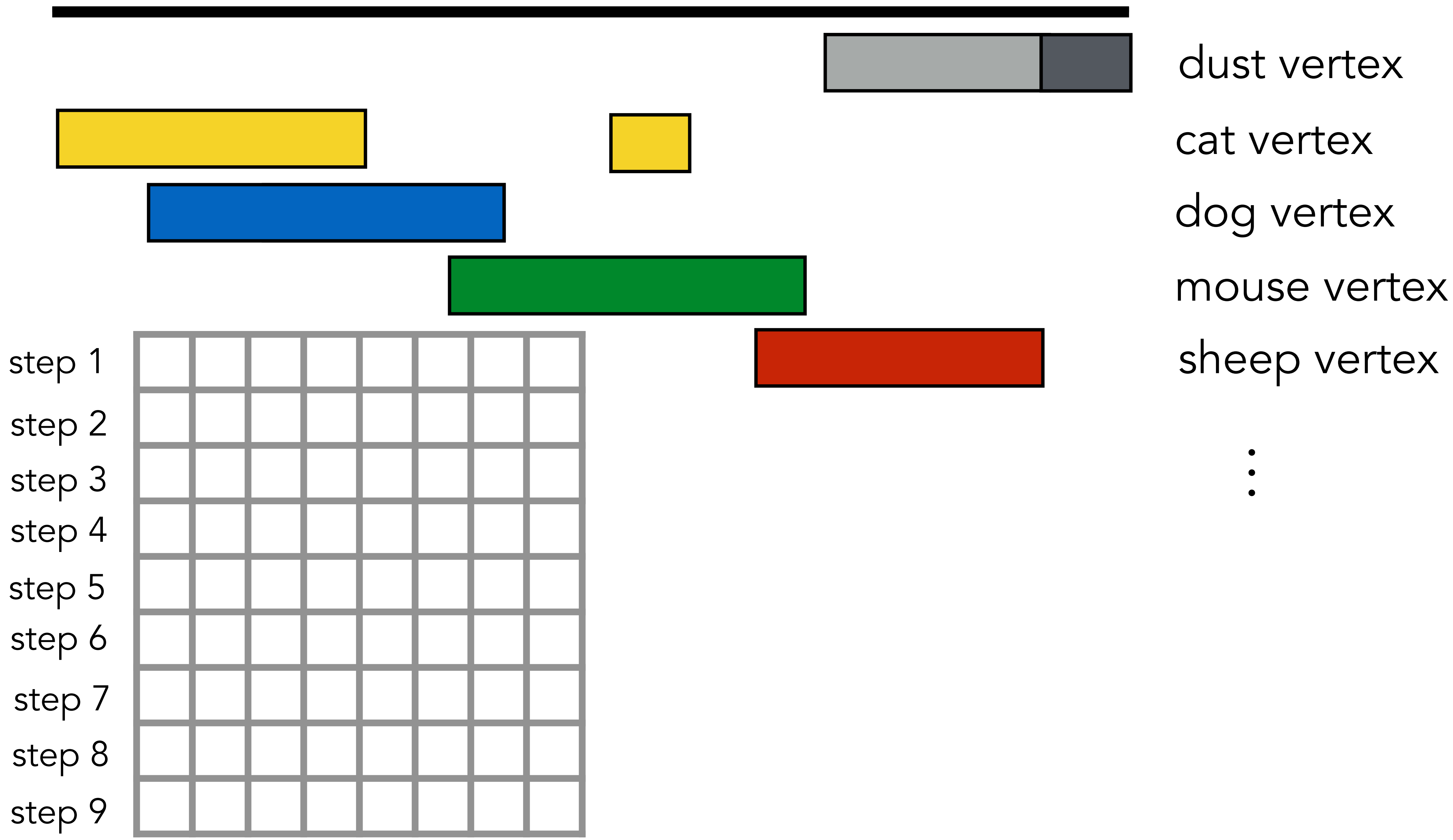


[Campbell, Cai, Broderick, 2016; Cai, Campbell, Broderick, 2016b]

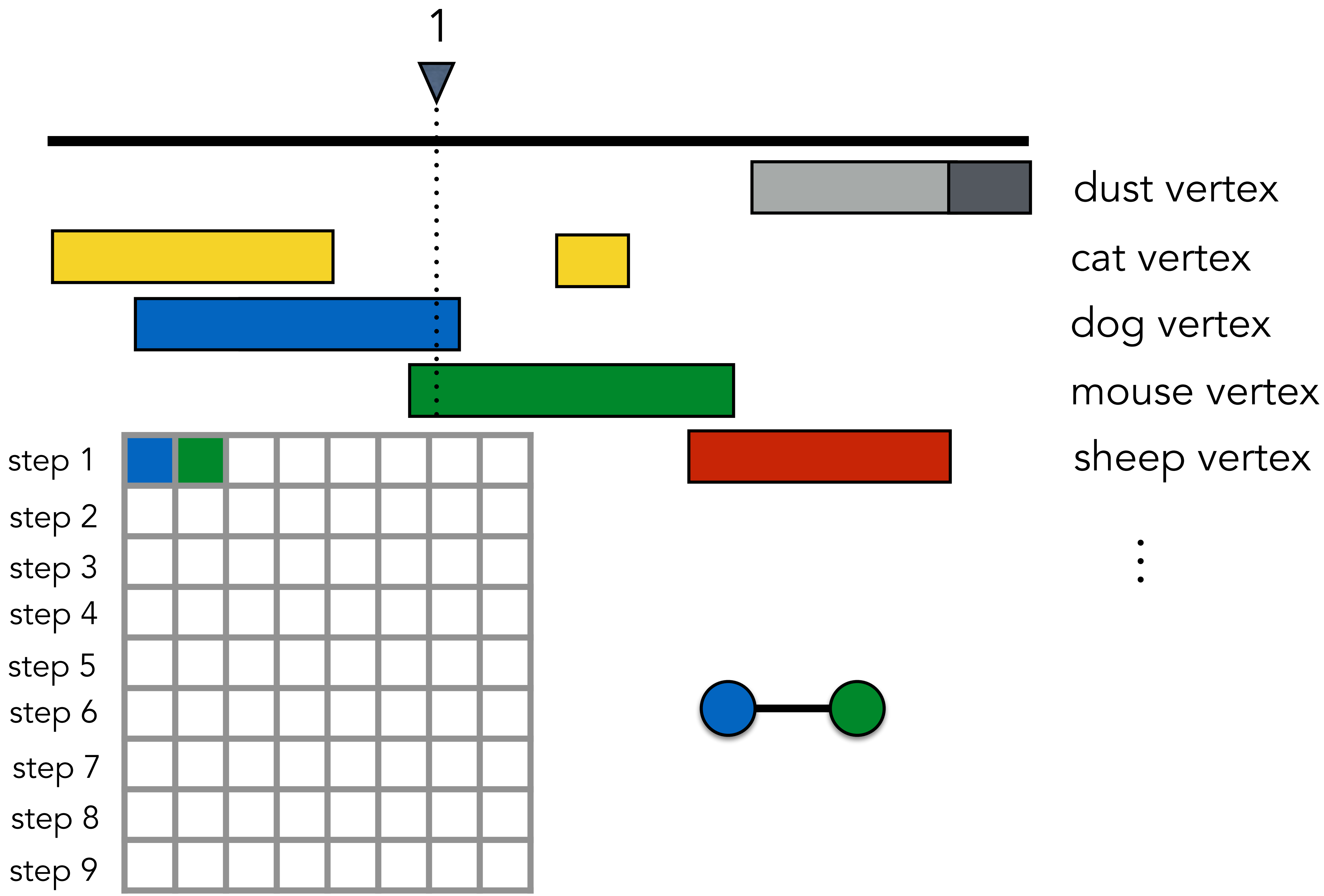
no other overlapping subsets creates 2 dust vertices



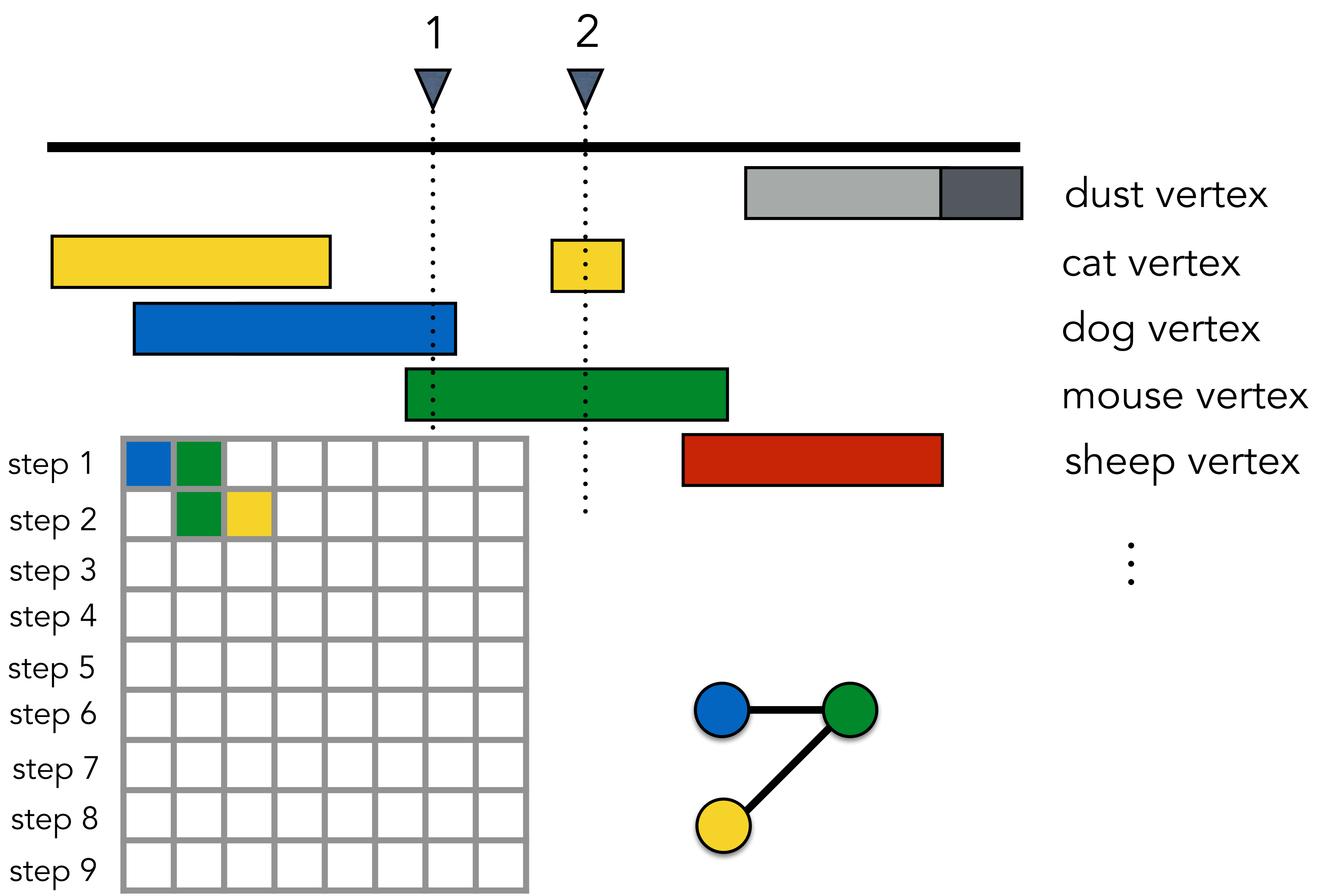
[Campbell, Cai, Broderick, 2016; Cai, Campbell, Broderick, 2016b]



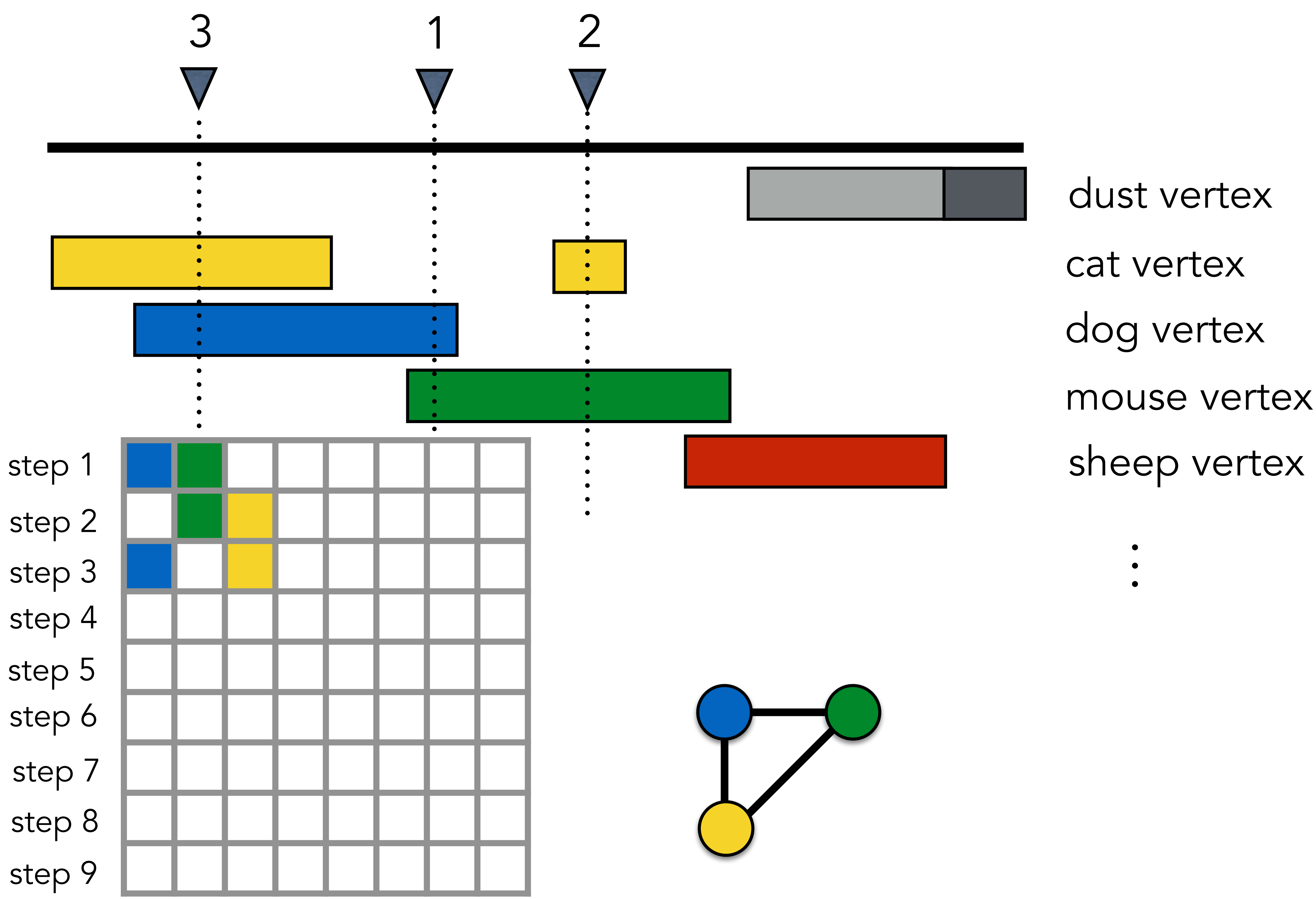
[Campbell, Cai, Broderick, 2016; Cai, Campbell, Broderick, 2016b]



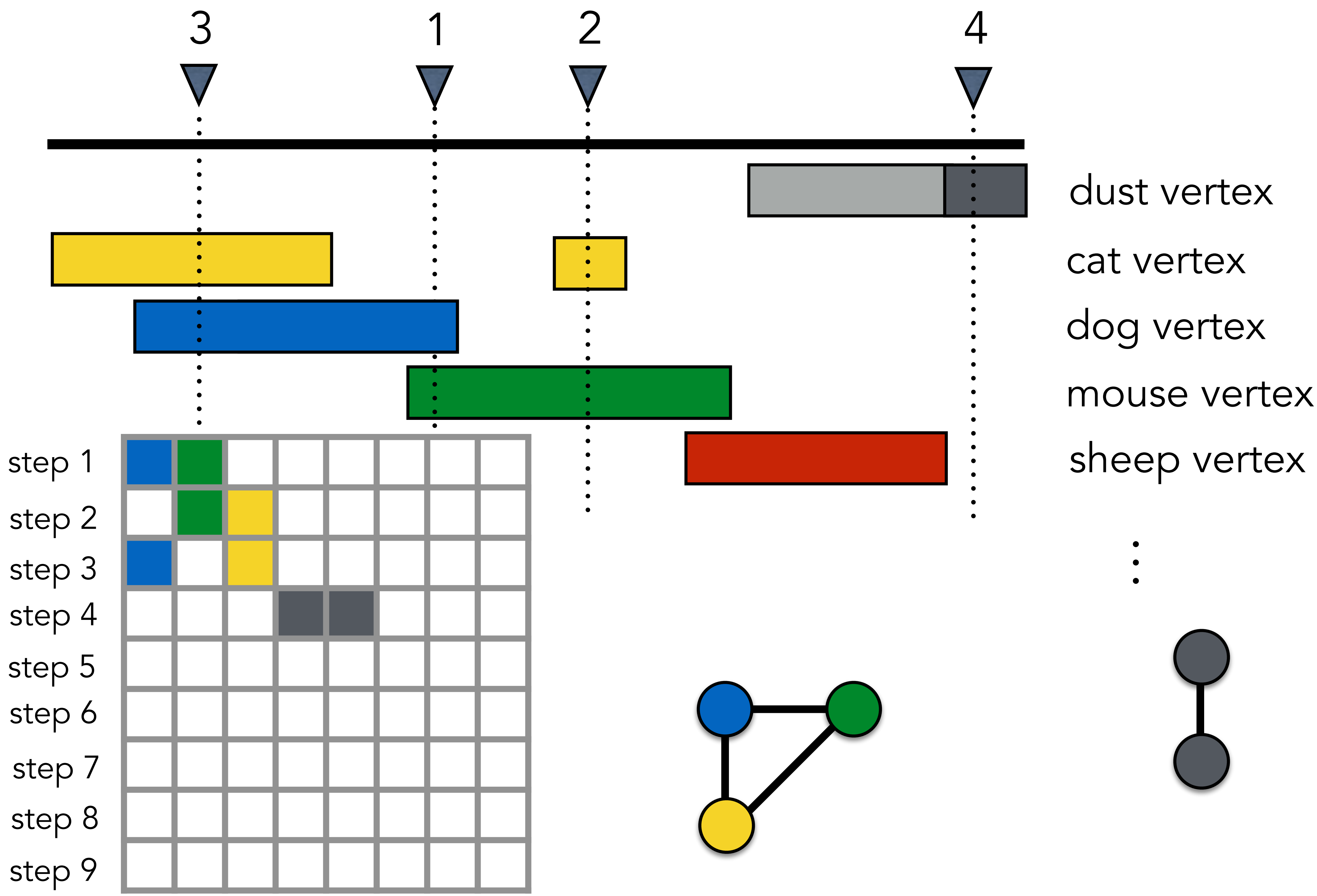
[Campbell, Cai, Broderick, 2016; Cai, Campbell, Broderick, 2016b]



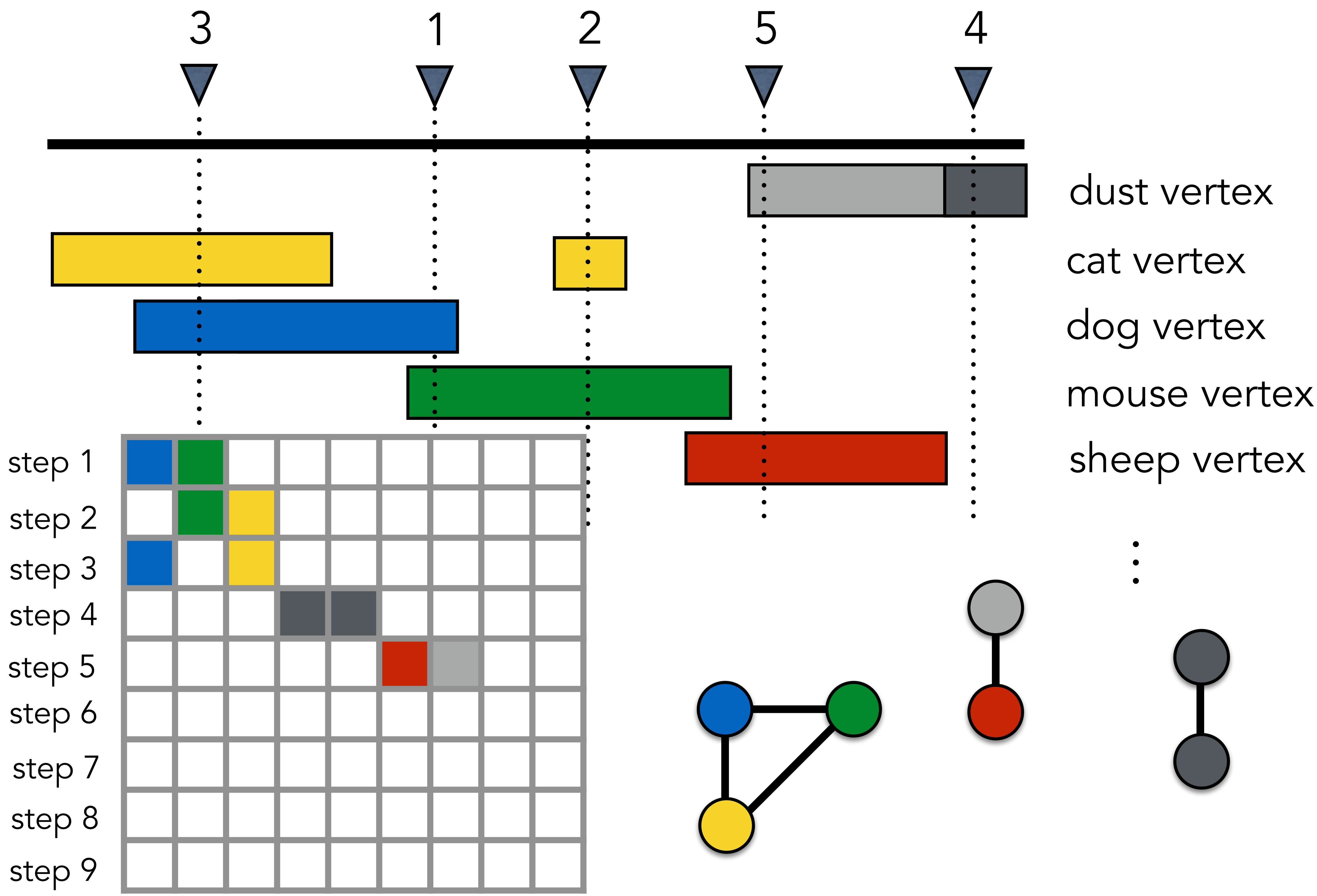
[Campbell, Cai, Broderick, 2016; Cai, Campbell, Broderick, 2016b]



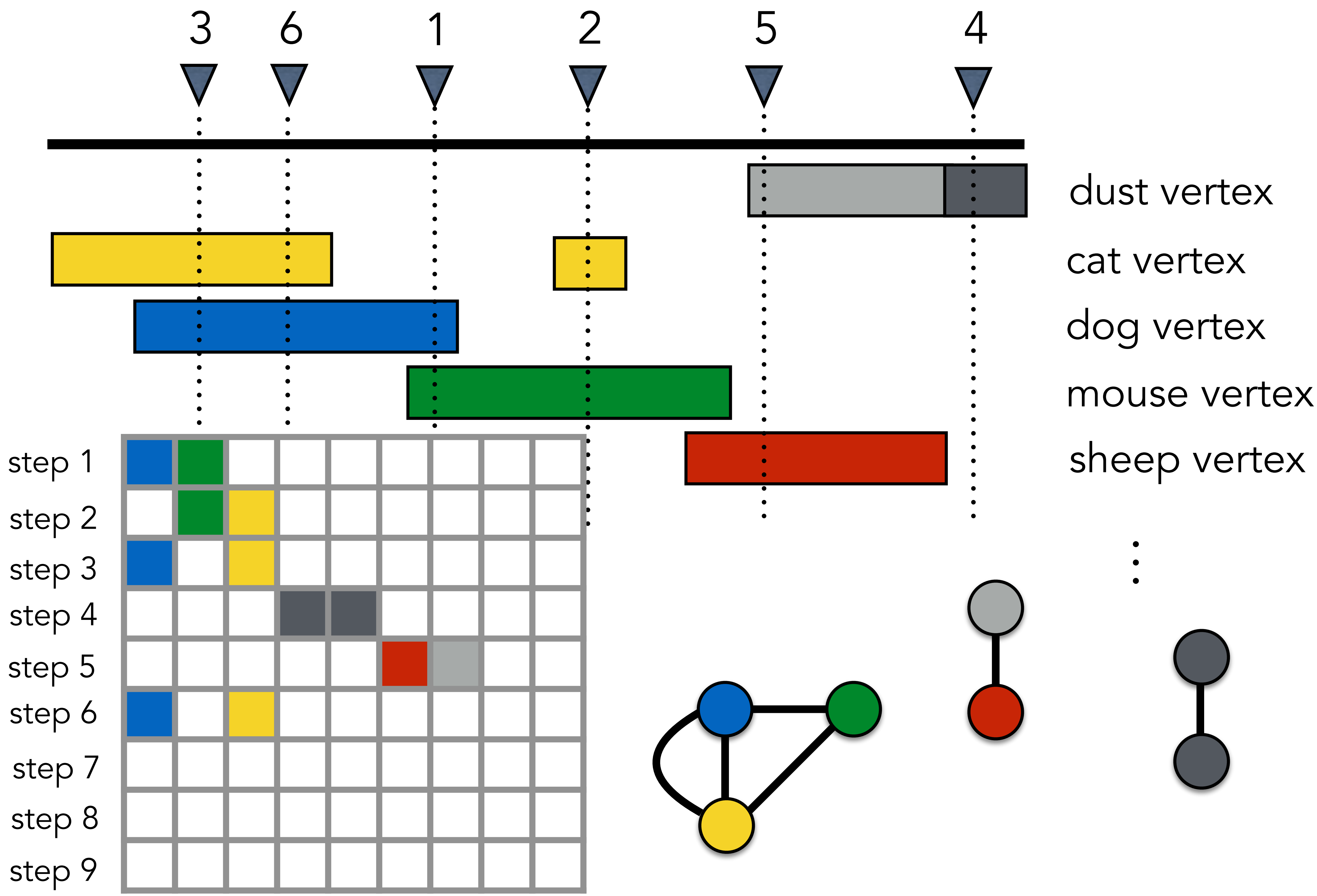
[Campbell, Cai, Broderick, 2016; Cai, Campbell, Broderick, 2016b]



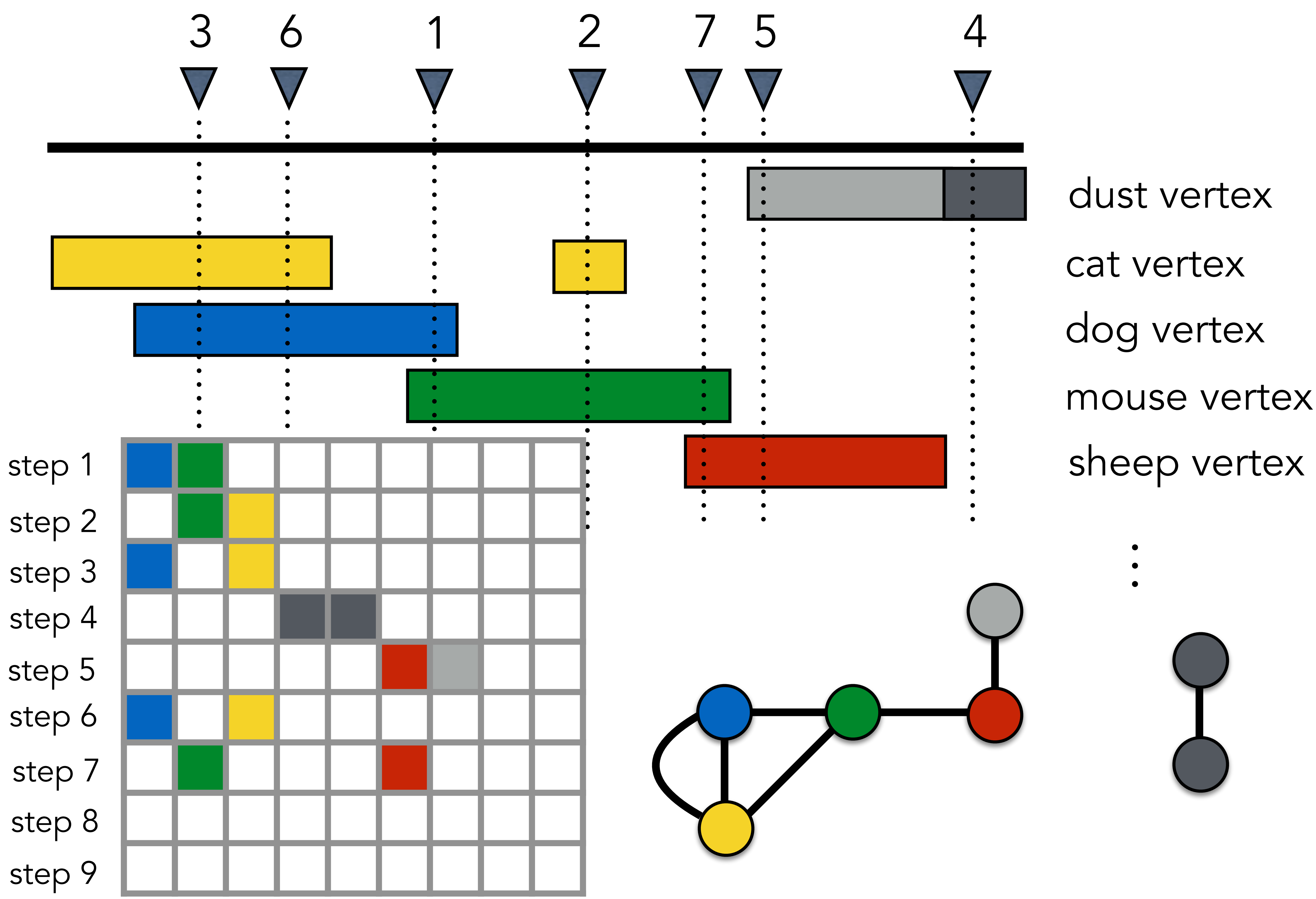
[Campbell, Cai, Broderick, 2016; Cai, Campbell, Broderick, 2016b]



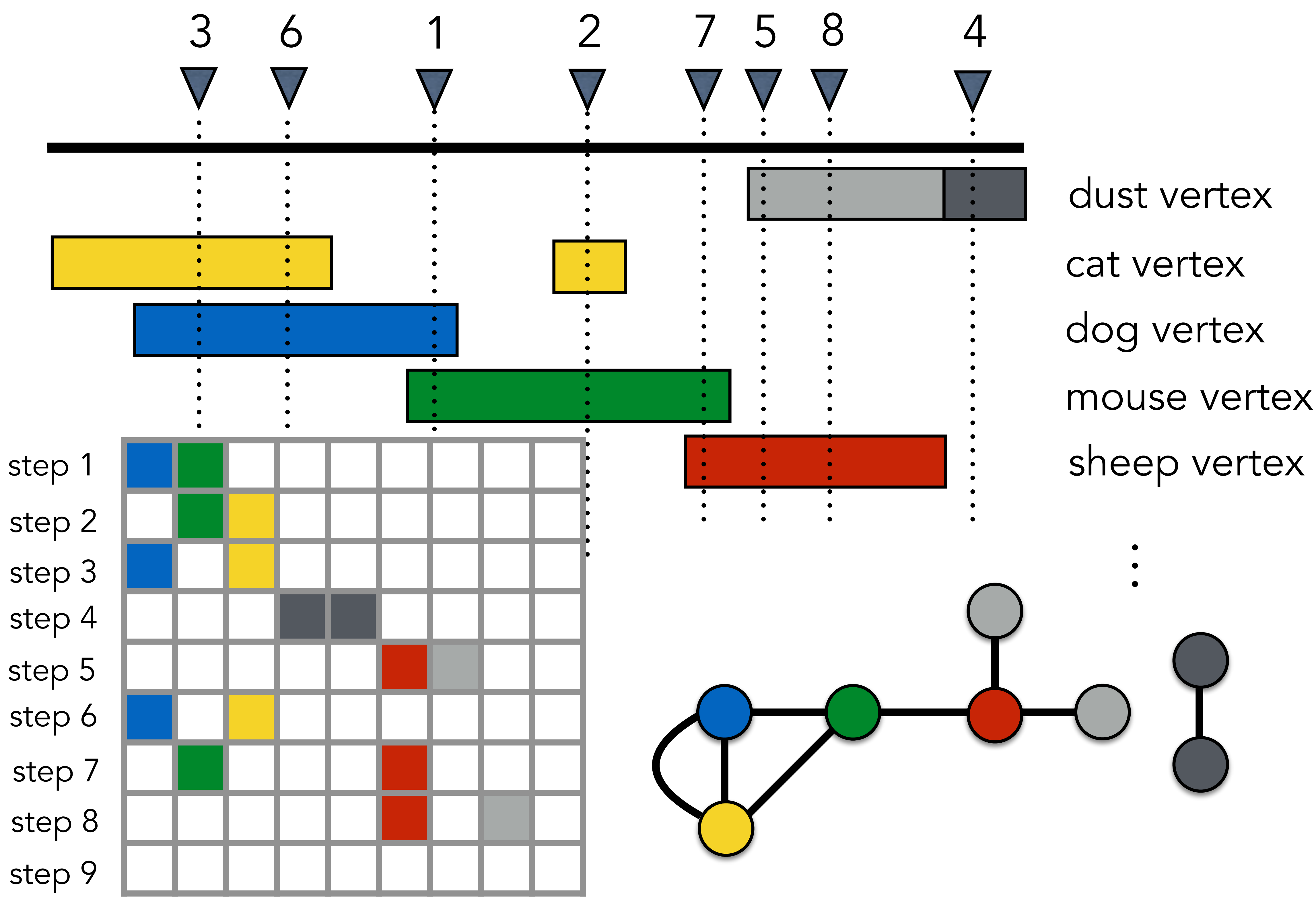
[Campbell, Cai, Broderick, 2016; Cai, Campbell, Broderick, 2016b]



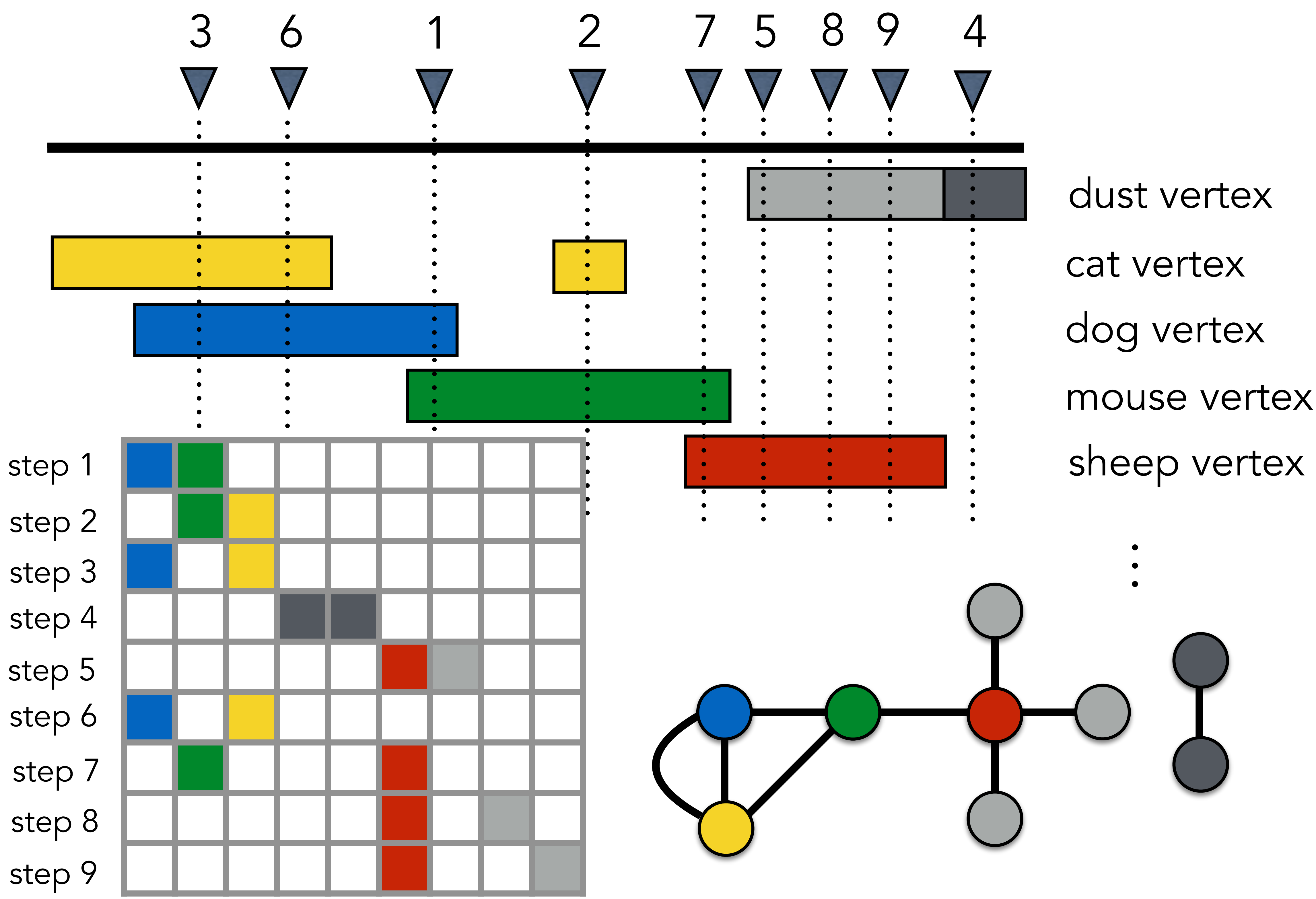
[Campbell, Cai, Broderick, 2016; Cai, Campbell, Broderick, 2016b]



[Campbell, Cai, Broderick, 2016; Cai, Campbell, Broderick, 2016b]

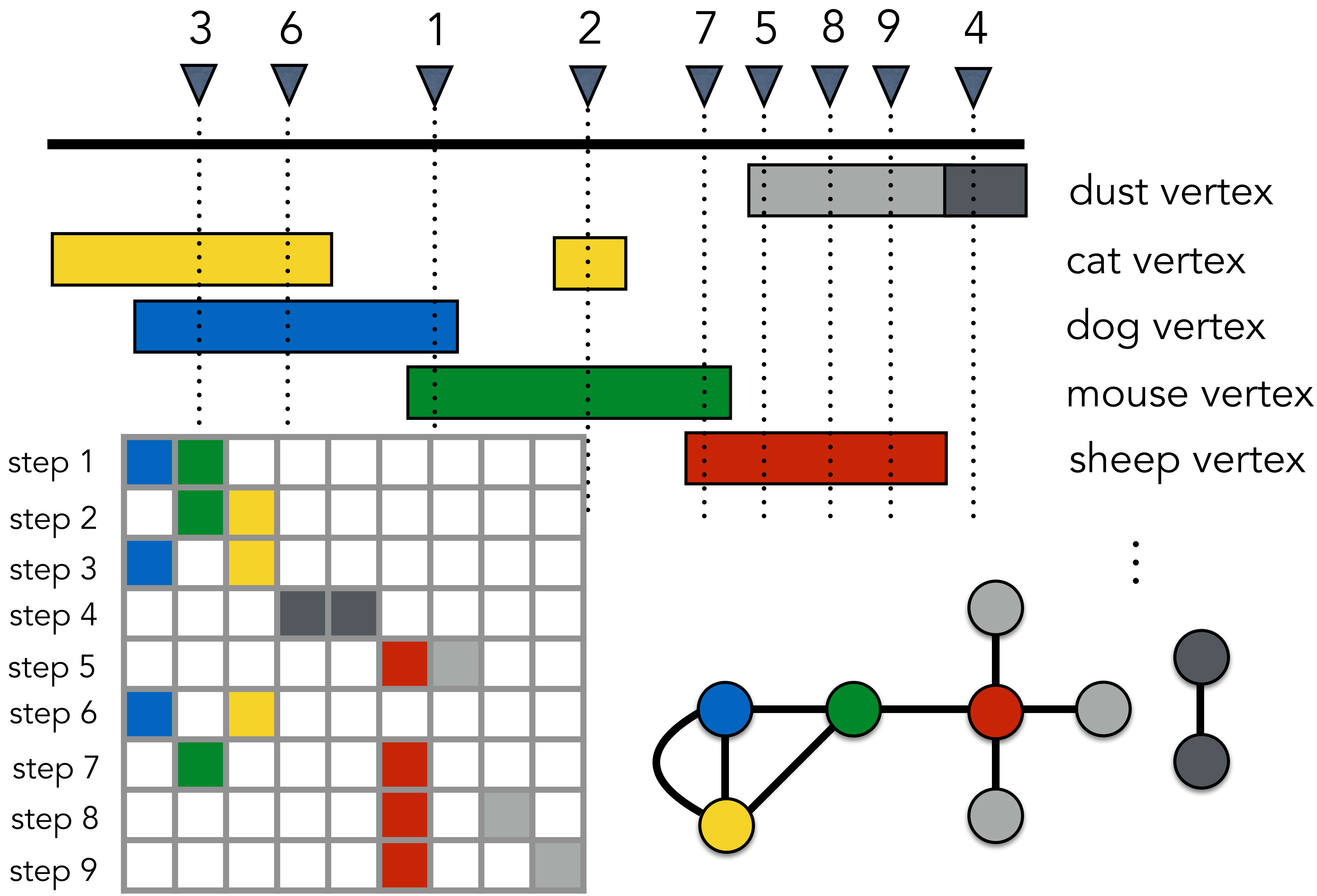


[Campbell, Cai, Broderick, 2016; Cai, Campbell, Broderick, 2016b]



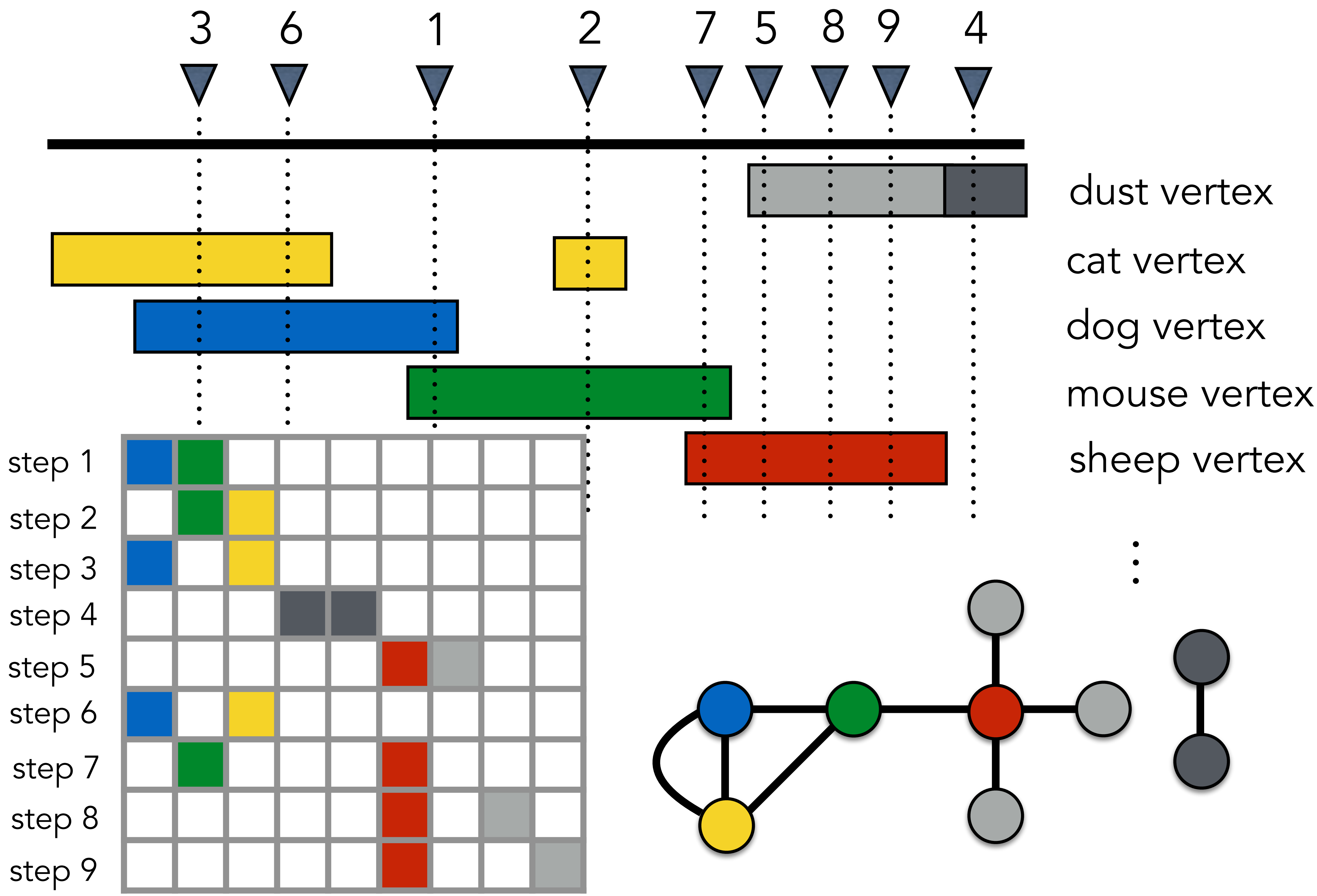
[Campbell, Cai, Broderick, 2016; Cai, Campbell, Broderick, 2016b]

This random graph is edge-exchangeable.



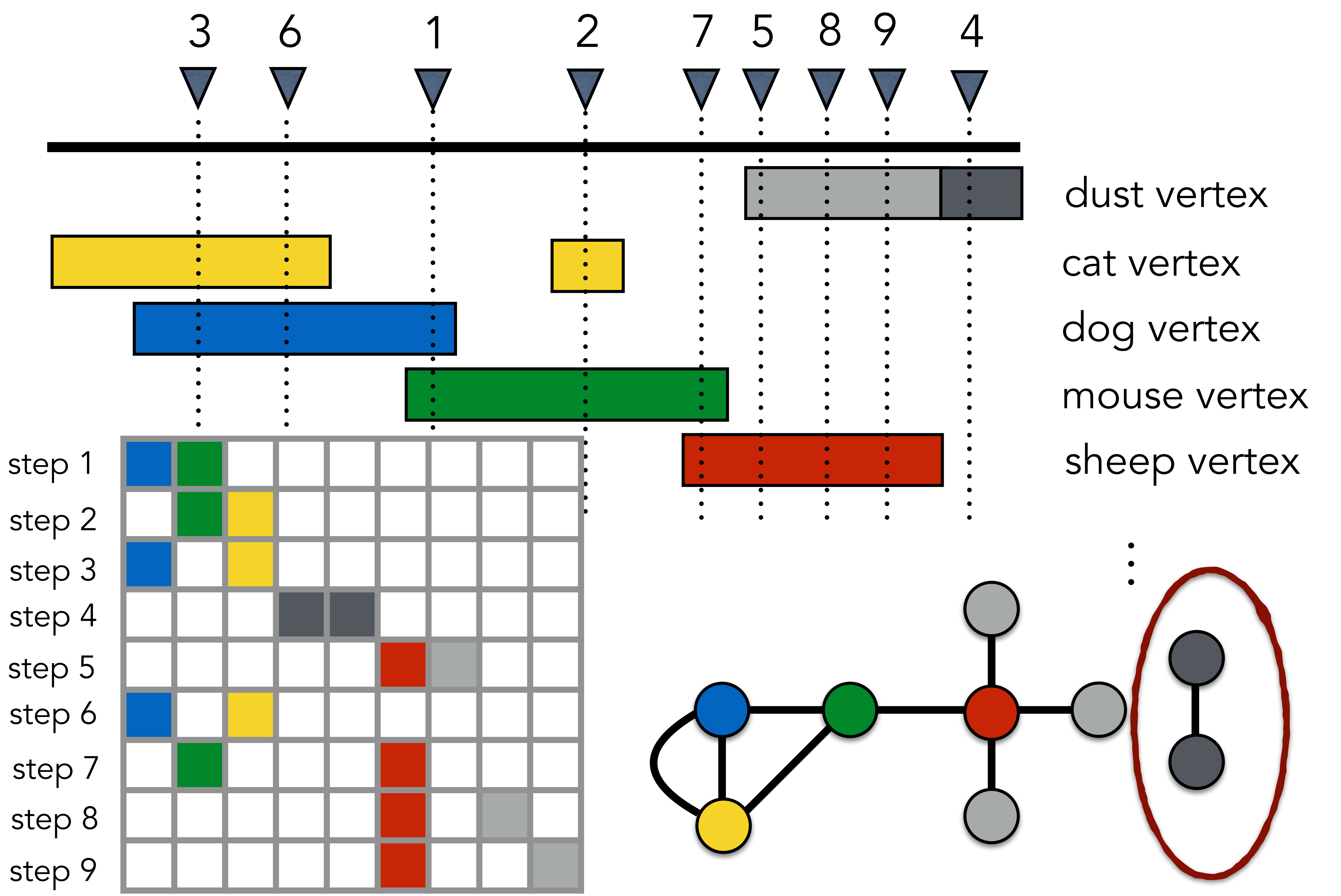
[Campbell, Cai, Broderick, 2016; Cai, Campbell, Broderick, 2016b]

The graph paintbox relates edges that connect to the same vertex, and allows us to control the topology of the graph.



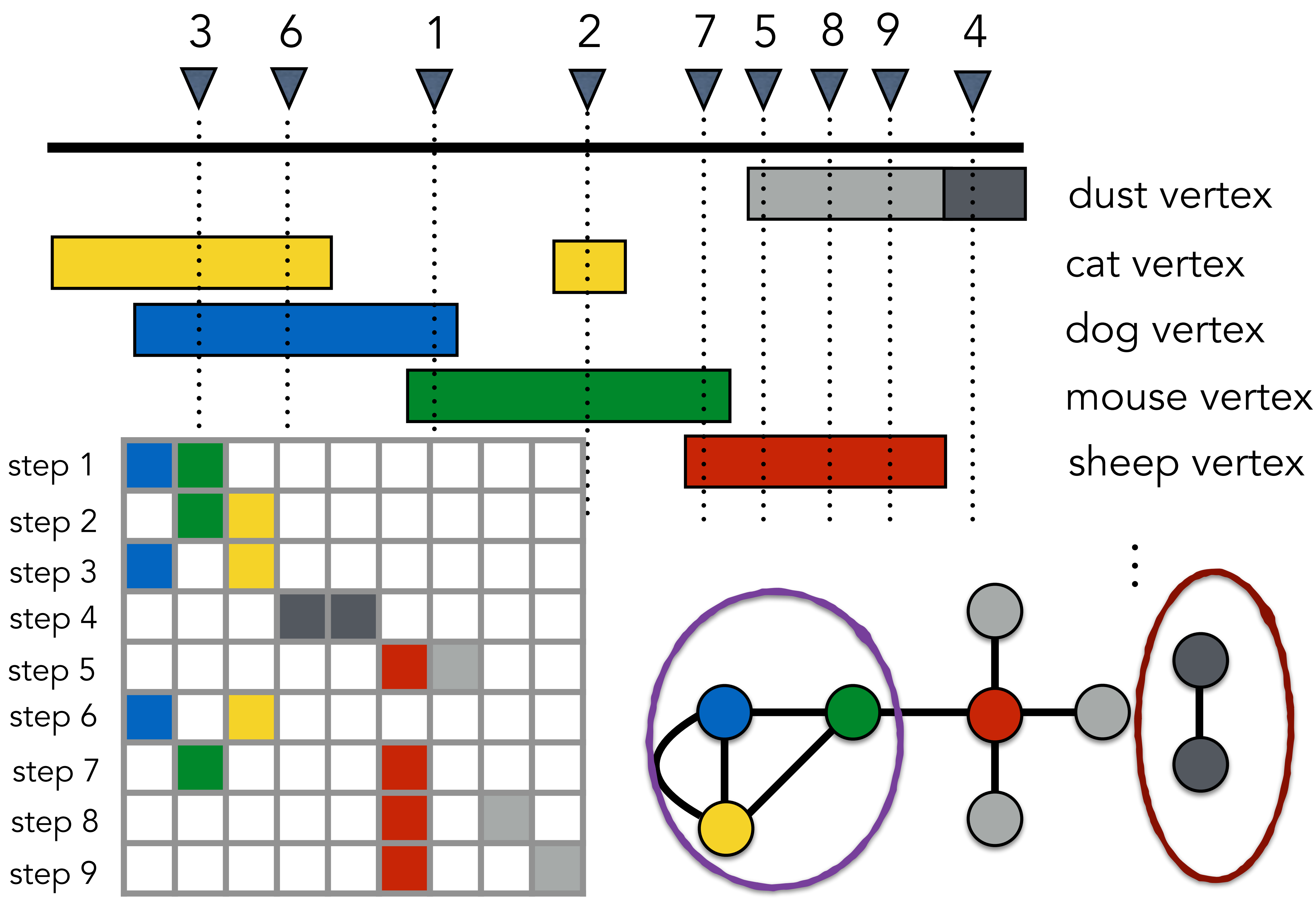
[Campbell, Cai, Broderick, 2016; Cai, Campbell, Broderick, 2016b]

The graph paintbox relates edges that connect to the same vertex, and allows us to control the topology of the graph.



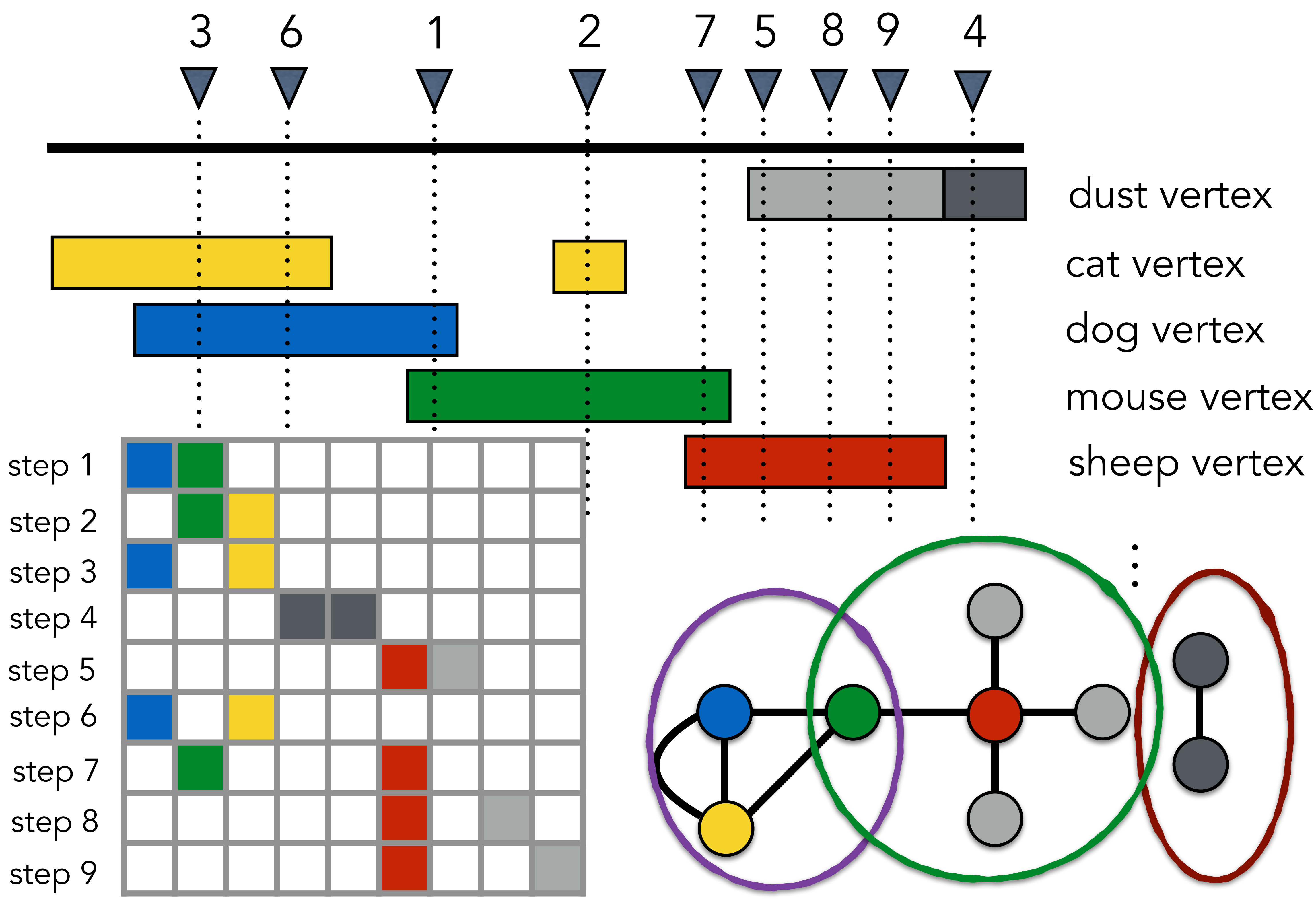
[Campbell, Cai, Broderick, 2016; Cai, Campbell, Broderick, 2016b]

The graph paintbox relates edges that connect to the same vertex, and allows us to control the topology of the graph.



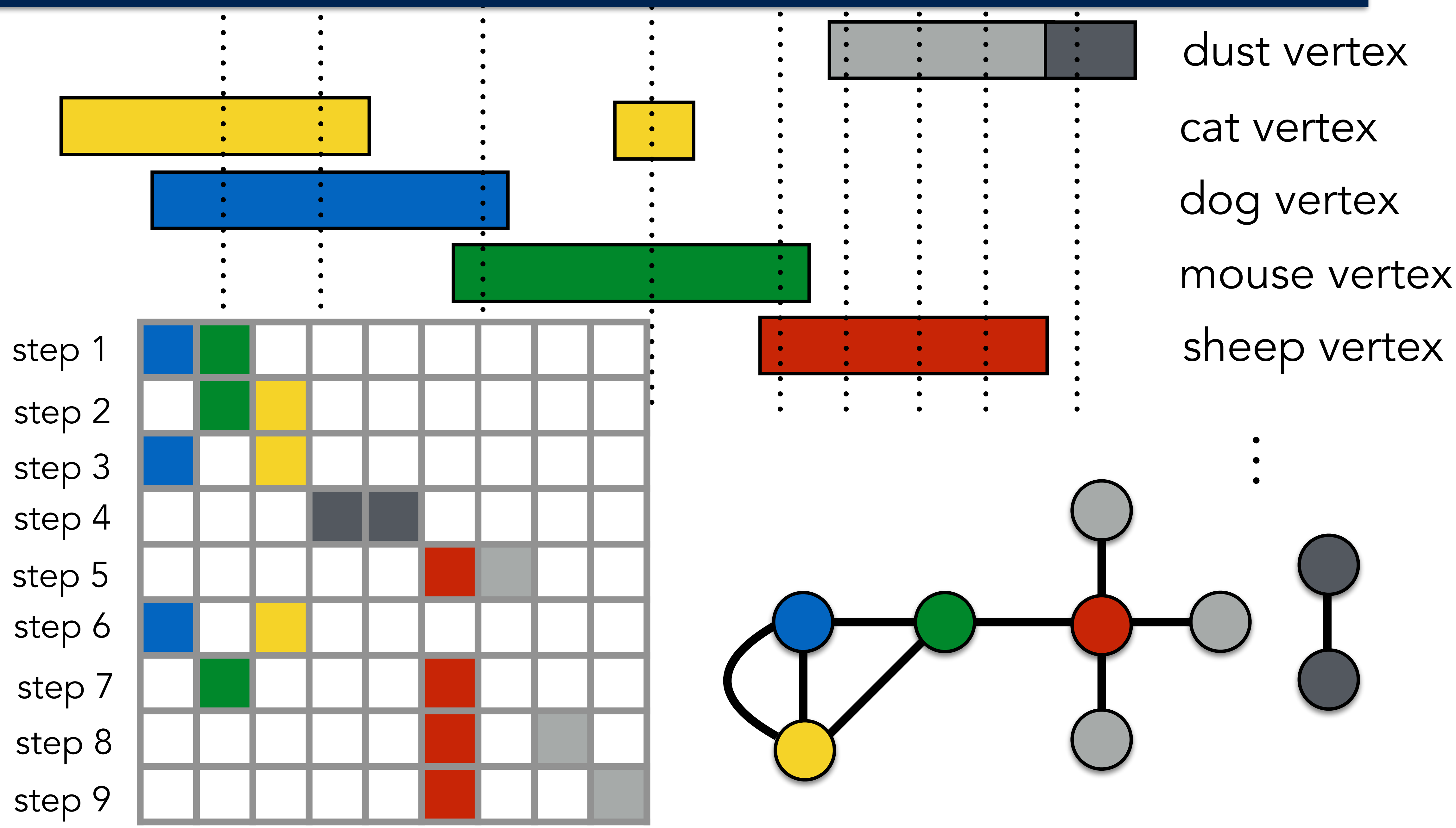
[Campbell, Cai, Broderick, 2016; Cai, Campbell, Broderick, 2016b]

The graph paintbox relates edges that connect to the same vertex, and allows us to control the topology of the graph.



[Campbell, Cai, Broderick, 2016; Cai, Campbell, Broderick, 2016b]

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



[Campbell, Cai, Broderick, 2016; Cai, Campbell, Broderick, 2016b]

Graph frequency models

=

Exchangeable vertex probability function

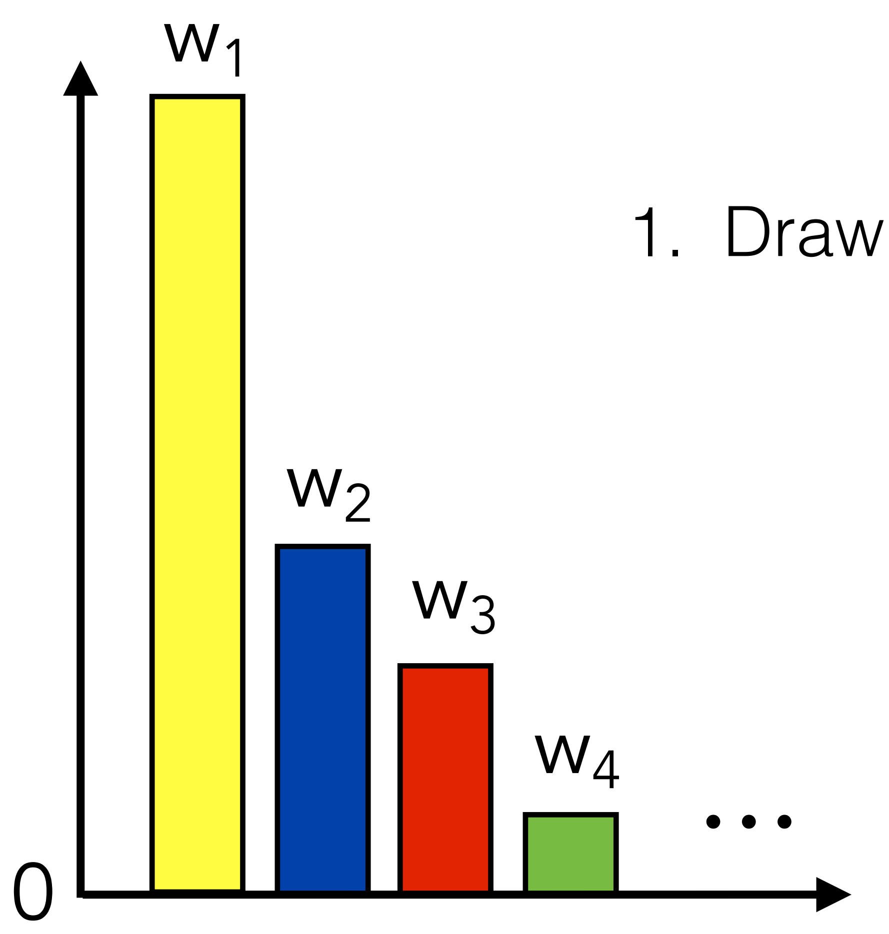
The graph paintbox is expressive but ***complex***.

The graph paintbox is expressive but ***complex***.

graph frequency model

[Cai, Campbell, Broderick, 2016a]

The graph paintbox is expressive but ***complex***.

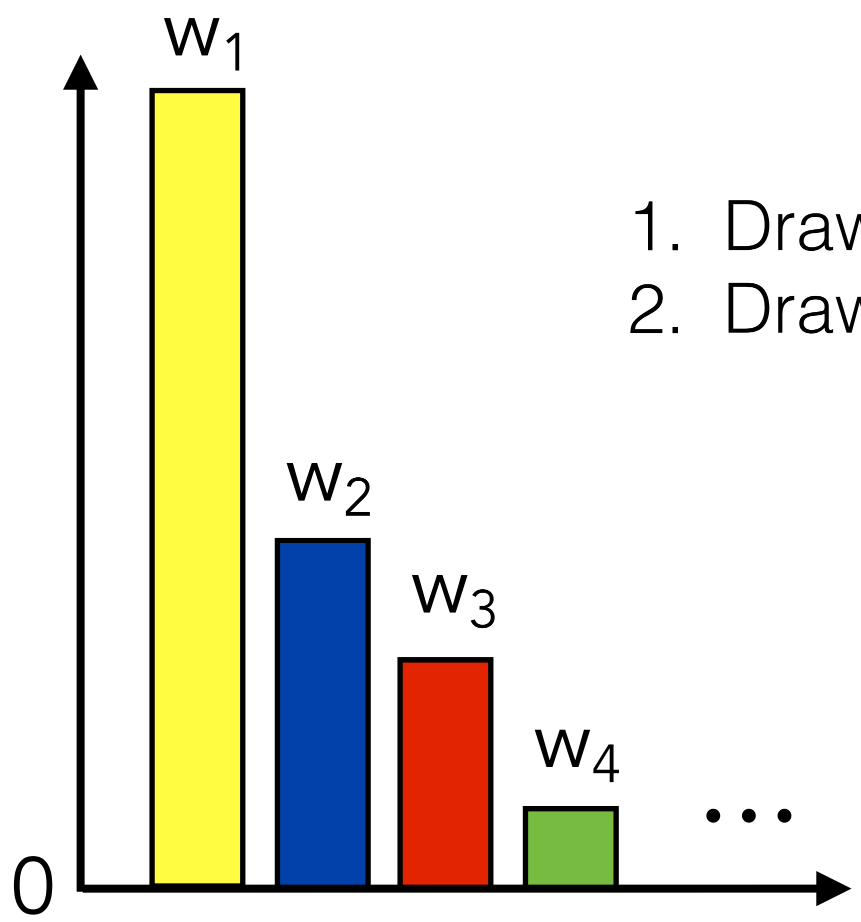


graph frequency model

1. Draw rates (w_i) from some distribution

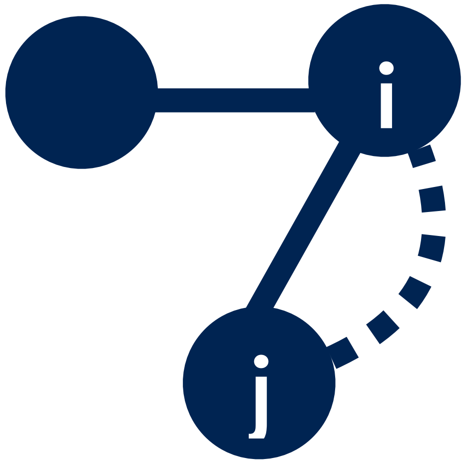
[Cai, Campbell, Broderick, 2016a]

The graph paintbox is expressive but ***complex***.



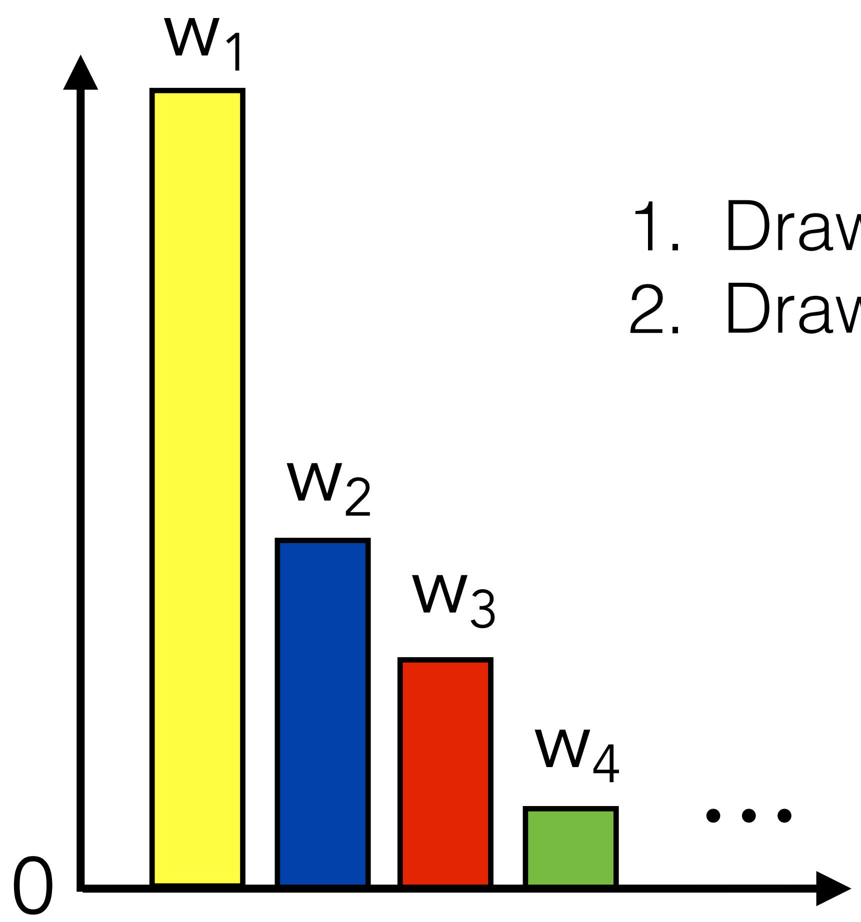
graph frequency model

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



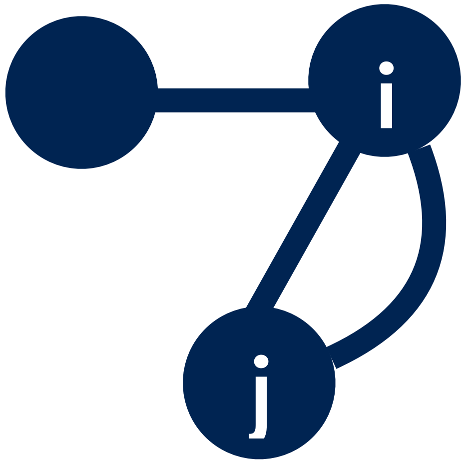
[Cai, Campbell, Broderick, 2016a]

The graph paintbox is expressive but ***complex***.



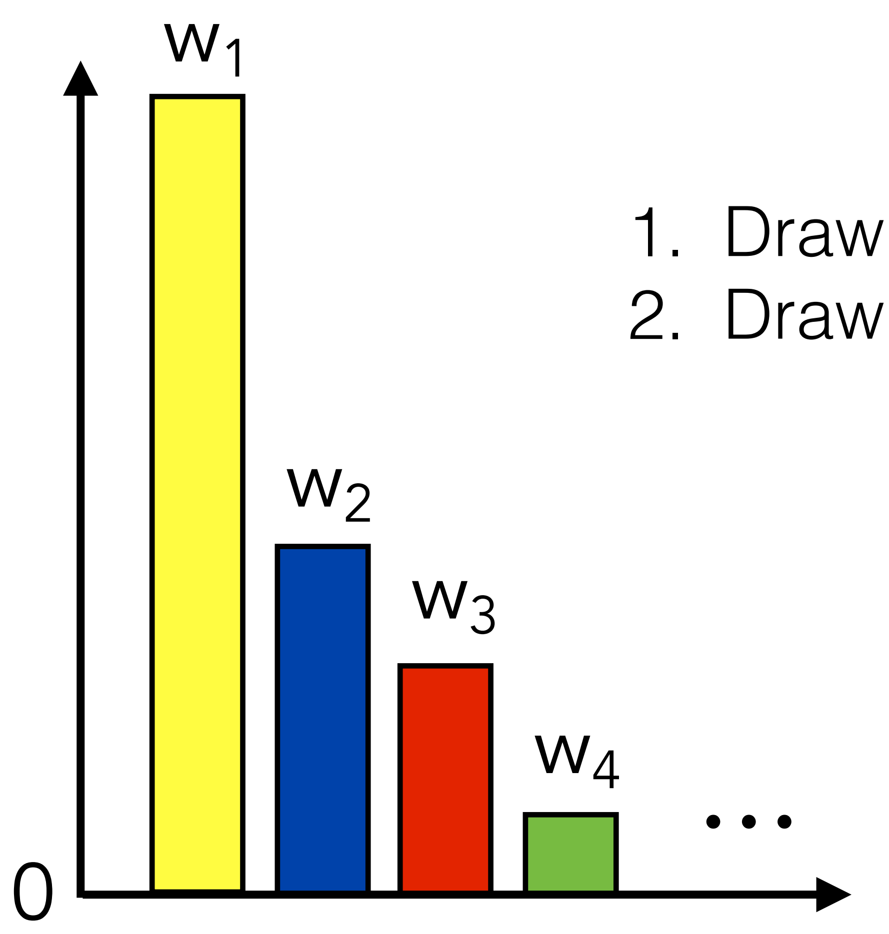
graph frequency model

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



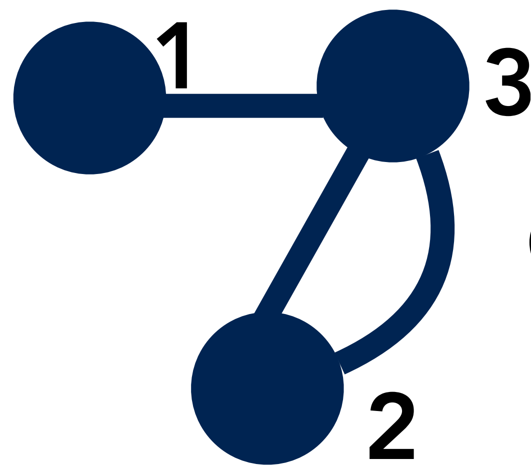
[Cai, Campbell, Broderick, 2016a]

The graph paintbox is expressive but ***complex***.



graph frequency model

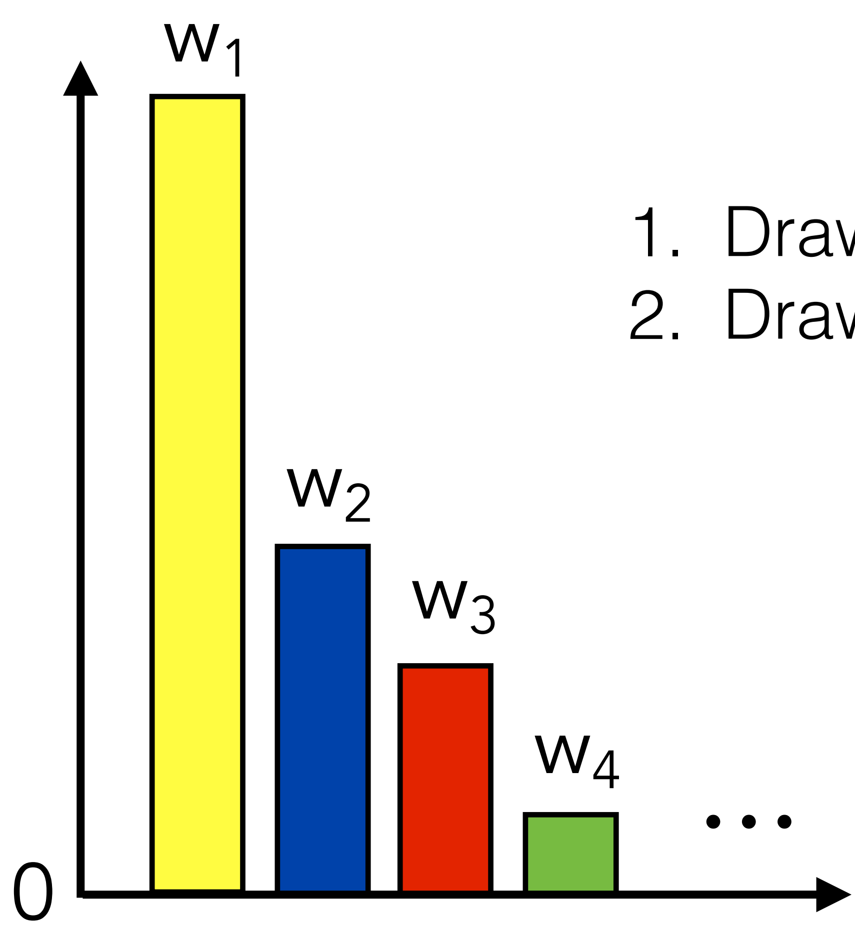
1. Draw rates (w_i) from some distribution
2. Draw edge {i,j} with probability proportional to $w_i w_j$



degrees

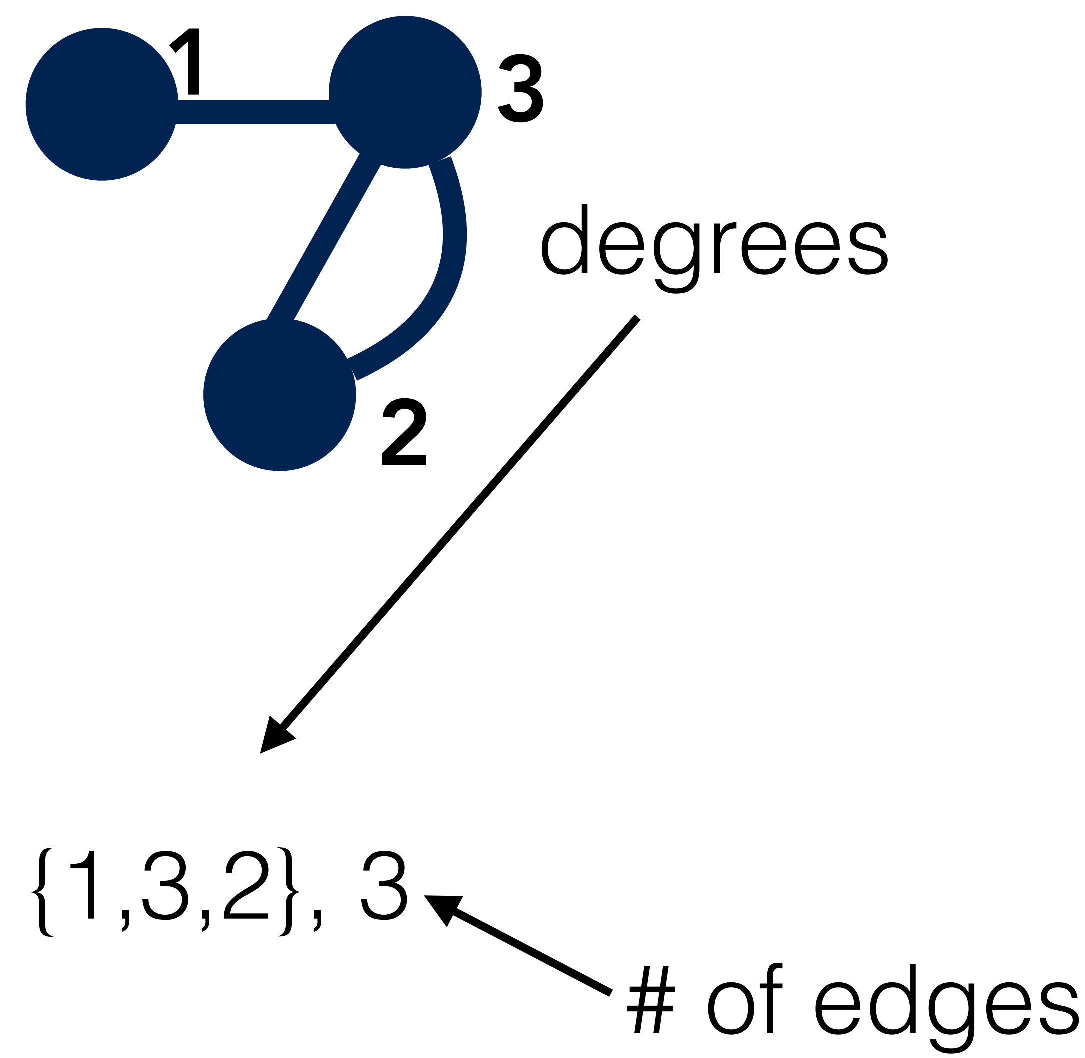
[Cai, Campbell, Broderick, 2016a]

The graph paintbox is expressive but ***complex***.



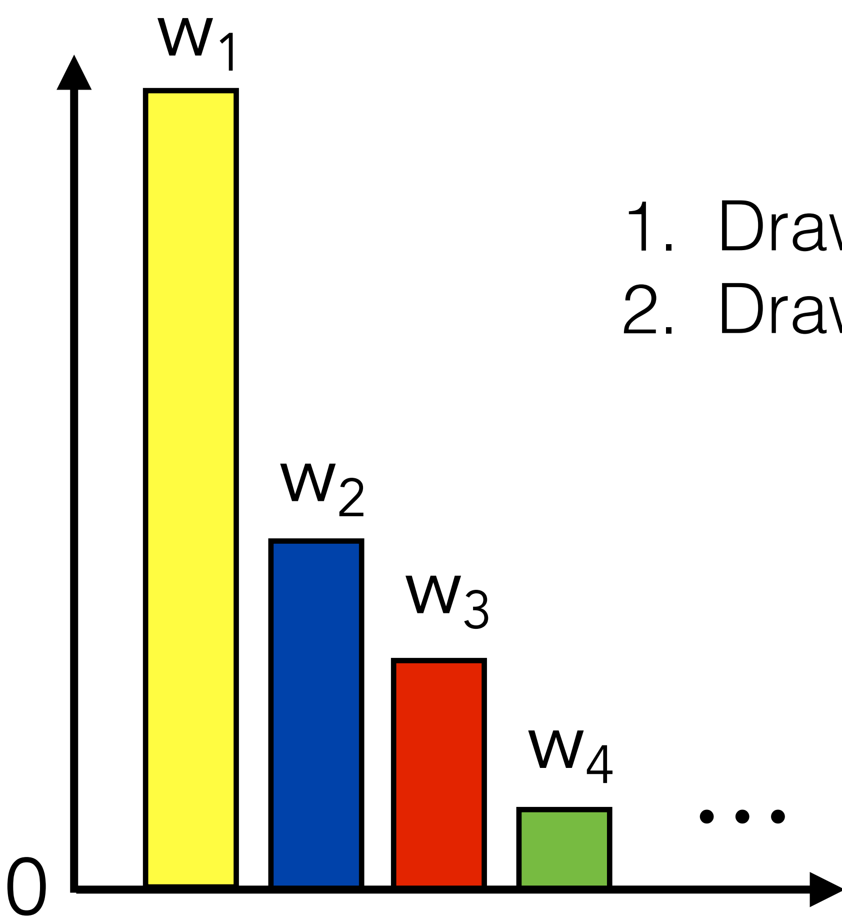
graph frequency model

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



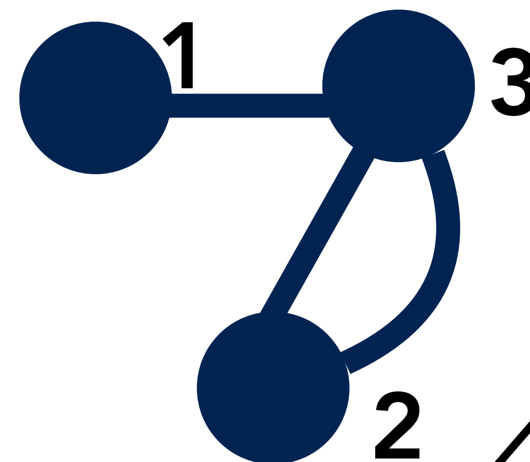
[Cai, Campbell, Broderick, 2016a]

The graph paintbox is expressive but ***complex***.



graph frequency model

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



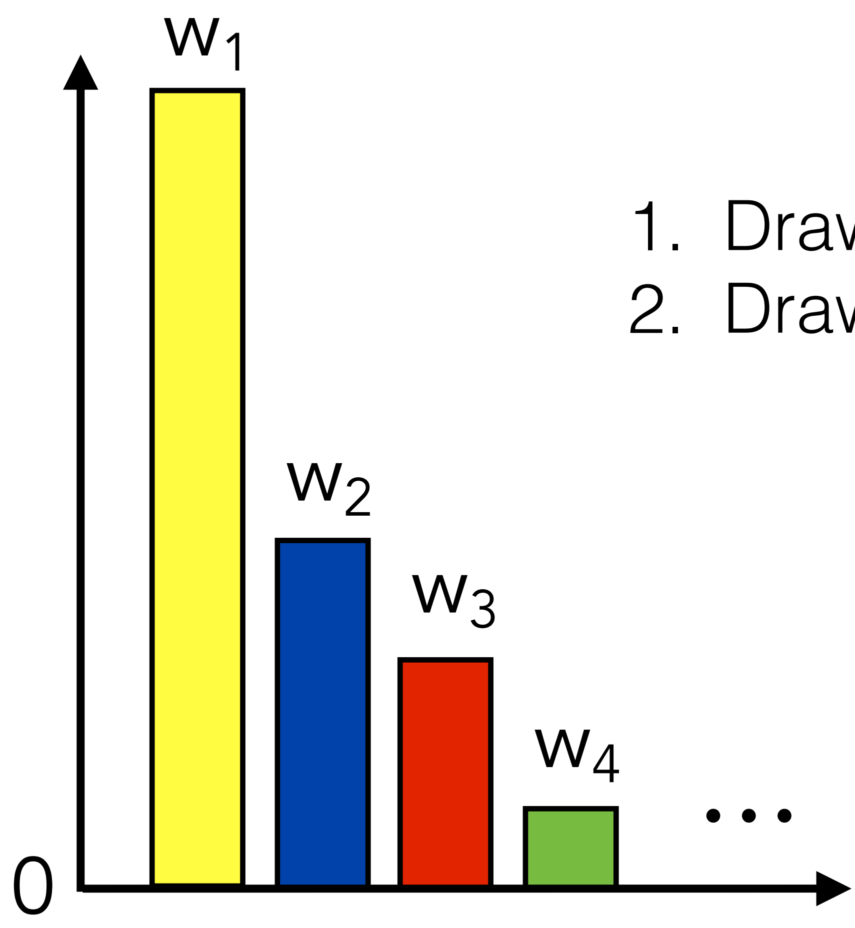
degrees

f({1,3,2}, 3)

# of edges

[Campbell, Cai, Broderick, 2016; Cai, Campbell, Broderick, 2016b]

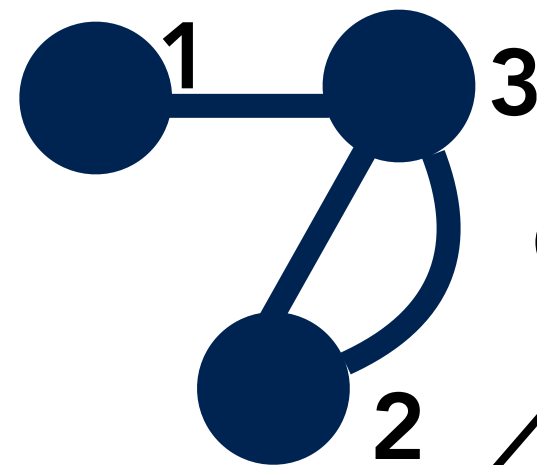
The graph paintbox is expressive but ***complex***.



Exchangeable vertex probability function (EVPF)

graph frequency model

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



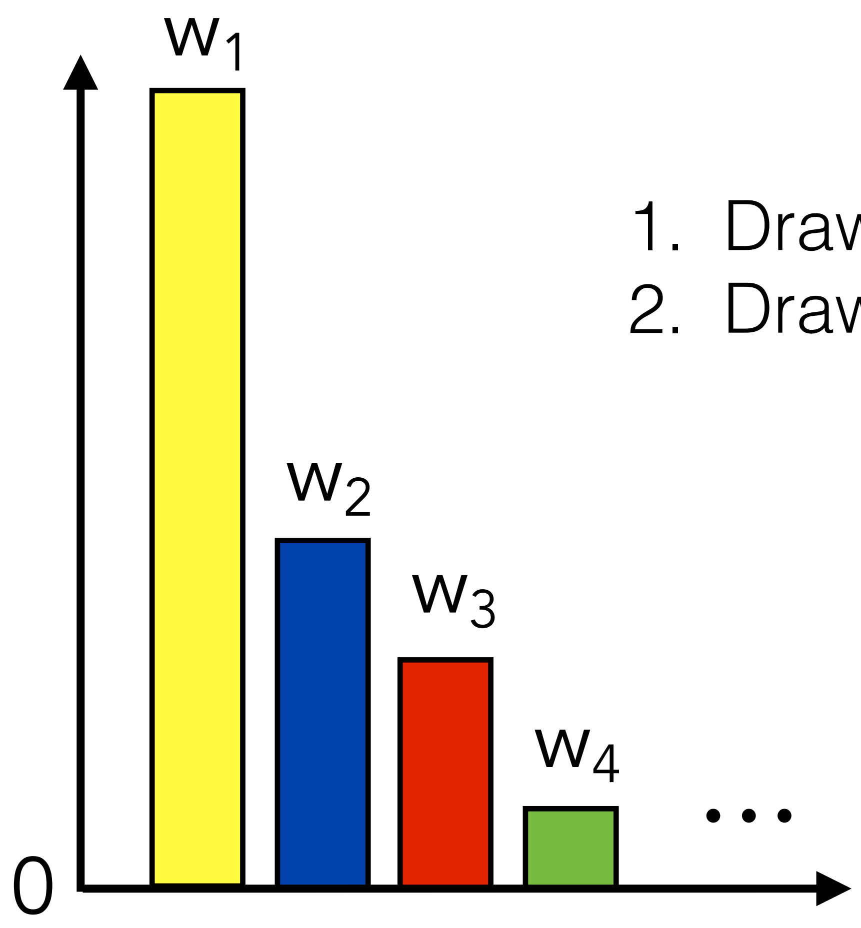
degrees

f({1,3,2}, 3)

# of edges

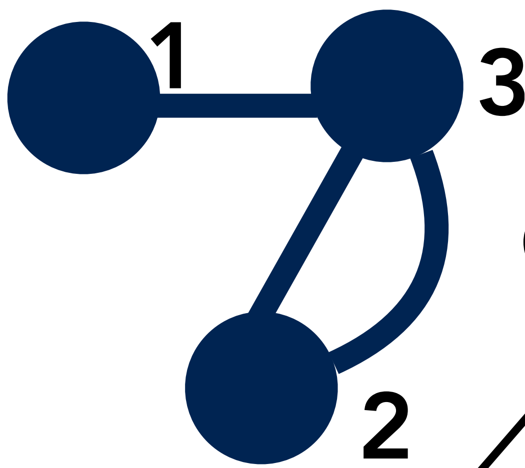
[Campbell, Cai, Broderick, 2016; Cai, Campbell, Broderick, 2016b]

The graph paintbox is expressive but ***complex***.



graph frequency model

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



Exchangeable vertex probability function (EVPF)

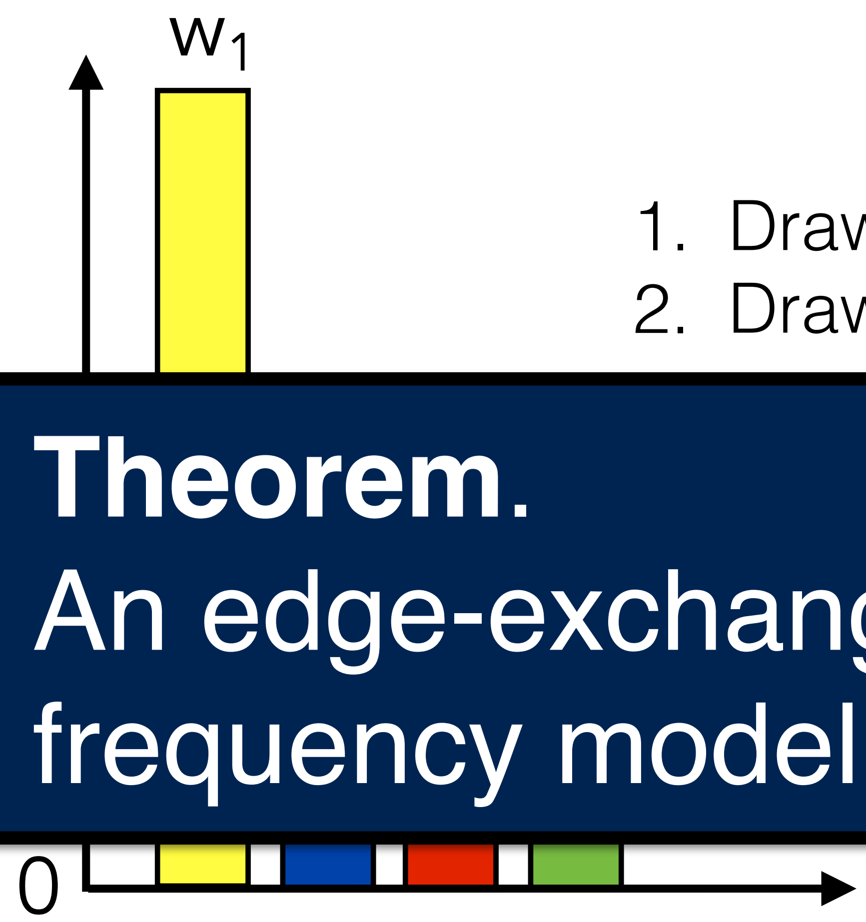
f({1,3,2}, 3)

# of edges

Reminiscent of exchangeable partition probability functions for clustering for efficient Gibbs sampling and variational inference algorithms.

[Campbell, Cai, Broderick, 2016; Cai, Campbell, Broderick, 2016b]

The graph paintbox is expressive but ***complex***.



graph frequency model

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

Theorem.
An edge-exchangeable graph has a graph frequency model iff it has an EVPF.

Exchangeable vertex probability function (EVPF)

f({1,3,2}, 3)

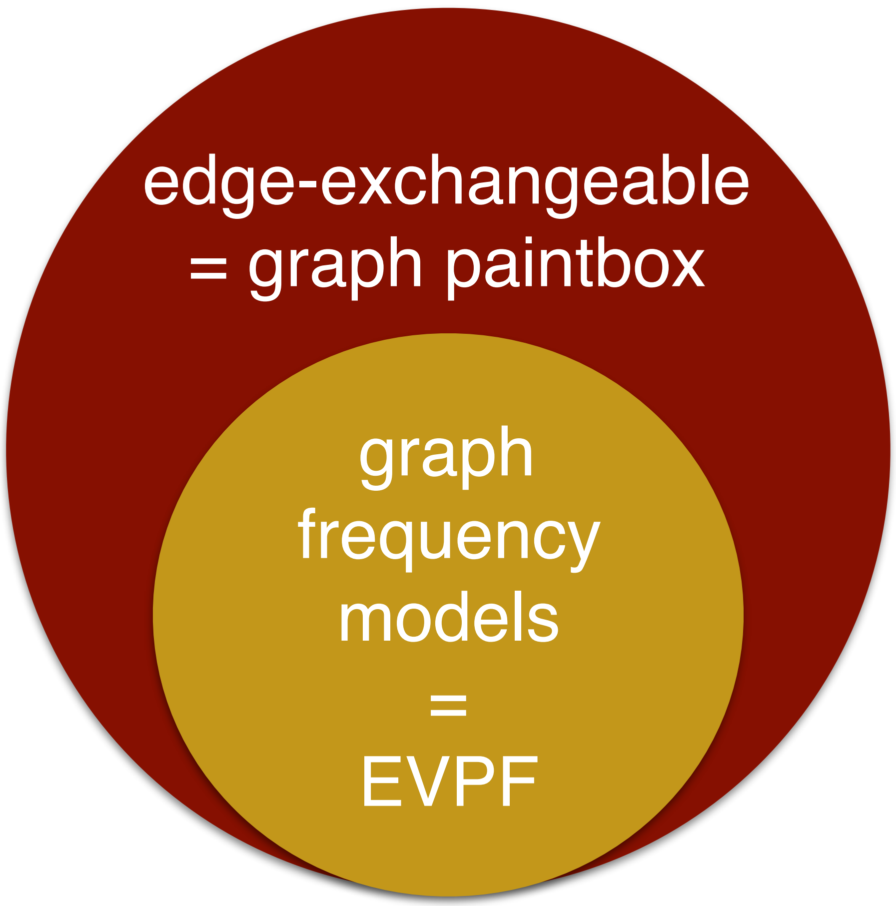
# of edges

Reminiscent of exchangeable partition probability functions for clustering for efficient Gibbs sampling and variational inference algorithms.

[Campbell, Cai, Broderick, 2016; Cai, Campbell, Broderick, 2016b]

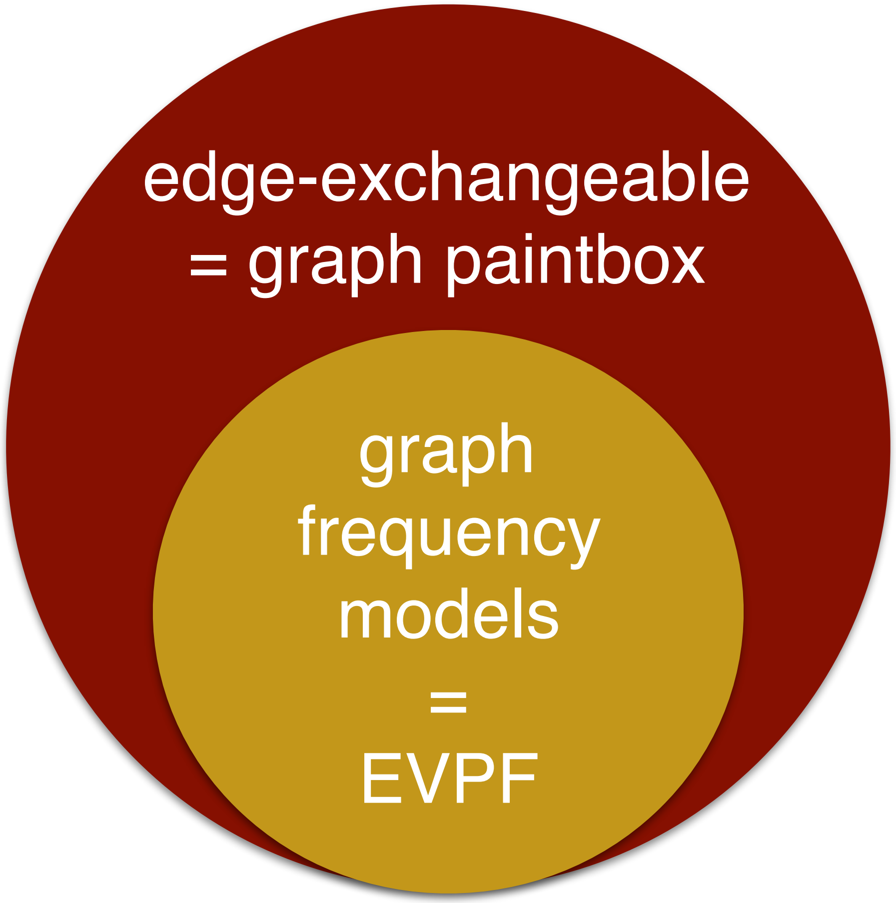
conclusions

✓characterized the class of edge-exchangeable graphs



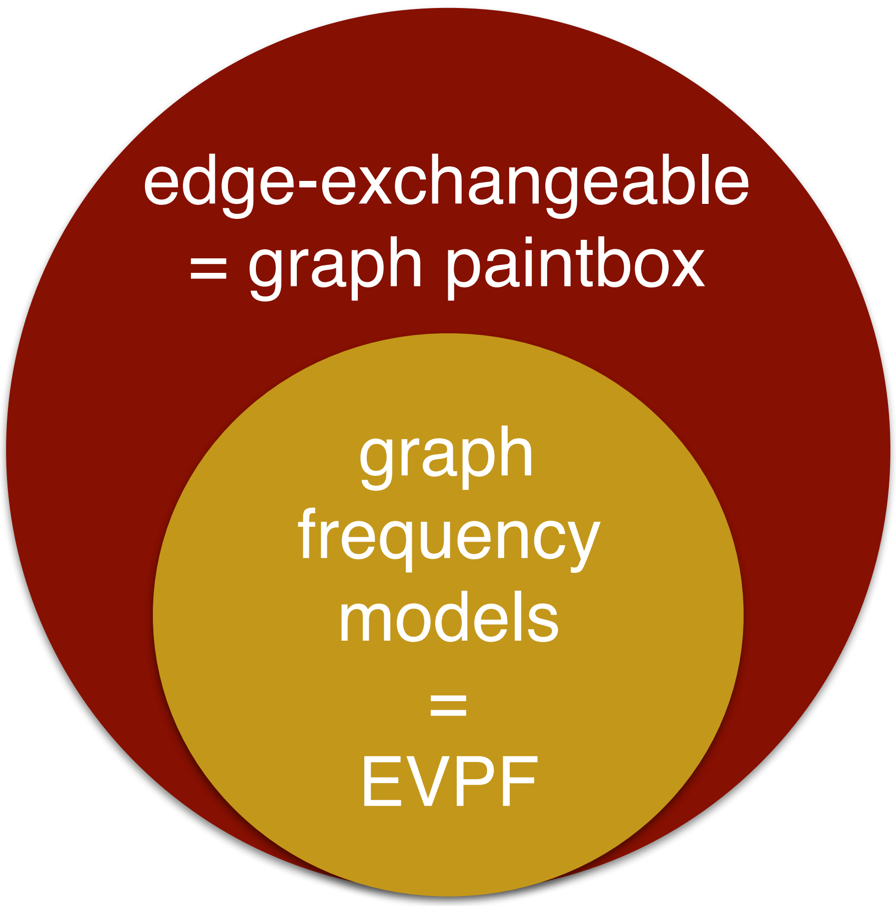
conclusions

✓characterized the class of edge-exchangeable graphs
✓characterized the class of graph frequency models



conclusions

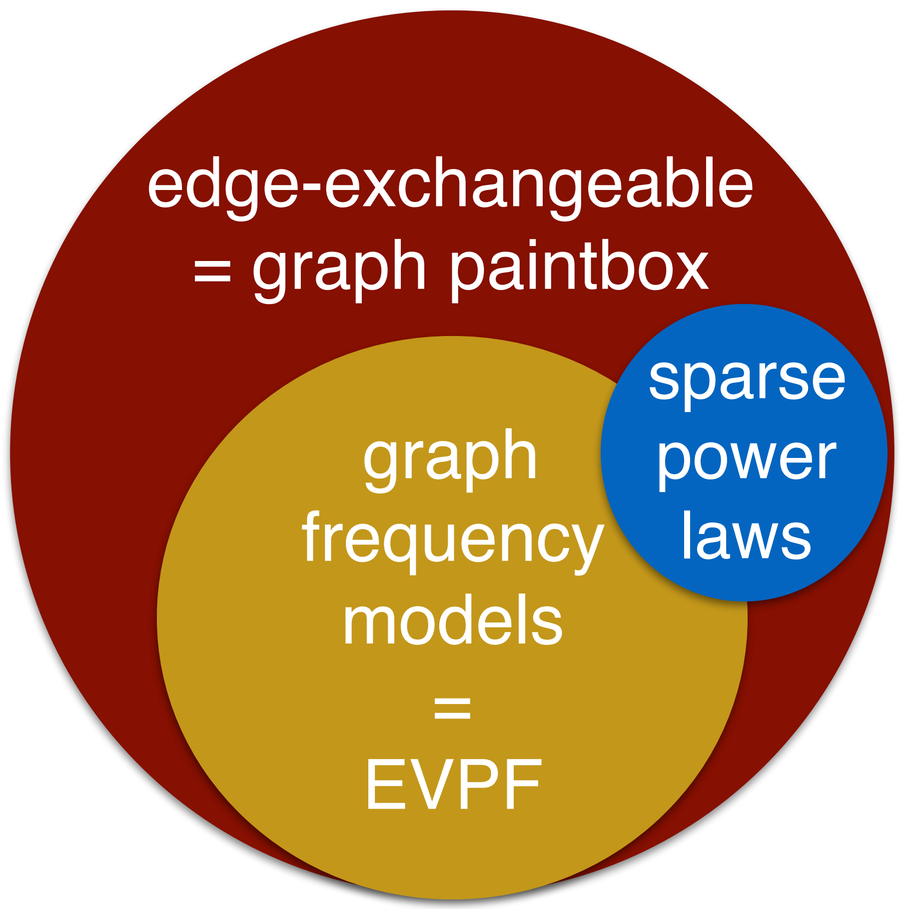
✓characterized the class of edge-exchangeable graphs
✓characterized the class of graph frequency models



future work:

conclusions

✓characterized the class of edge-exchangeable graphs
✓characterized the class of graph frequency models

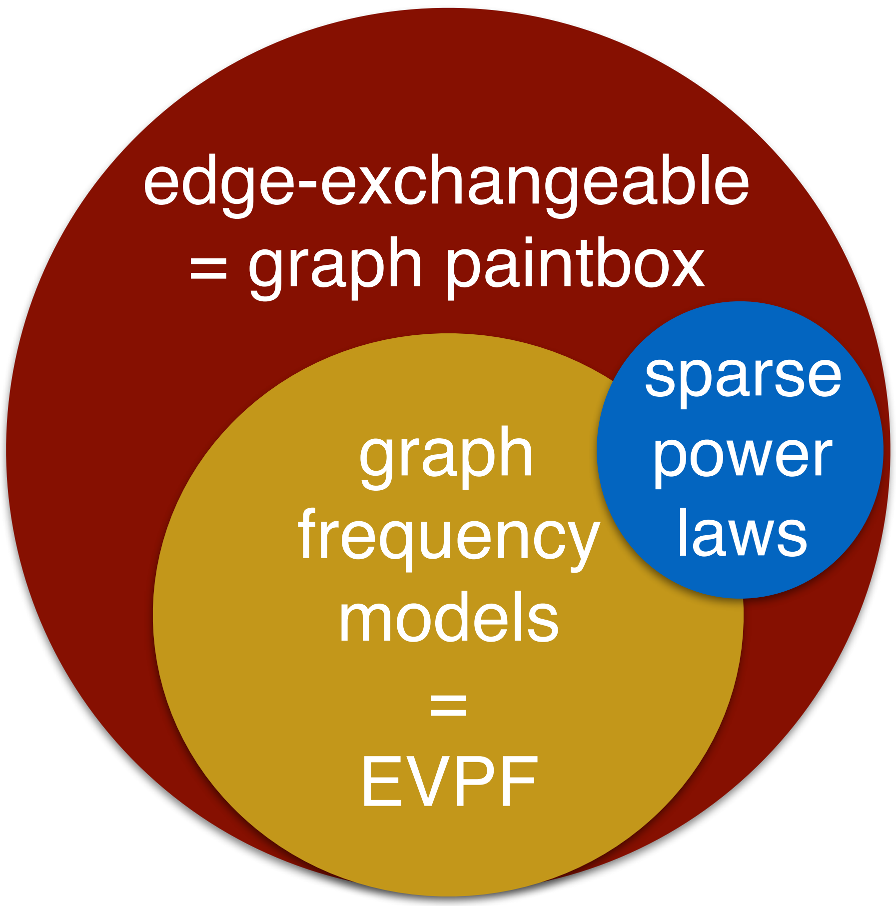


future work:

- characterize sparse, edge-exchangeable graph models

conclusions

✓characterized the class of edge-exchangeable graphs
✓characterized the class of graph frequency models

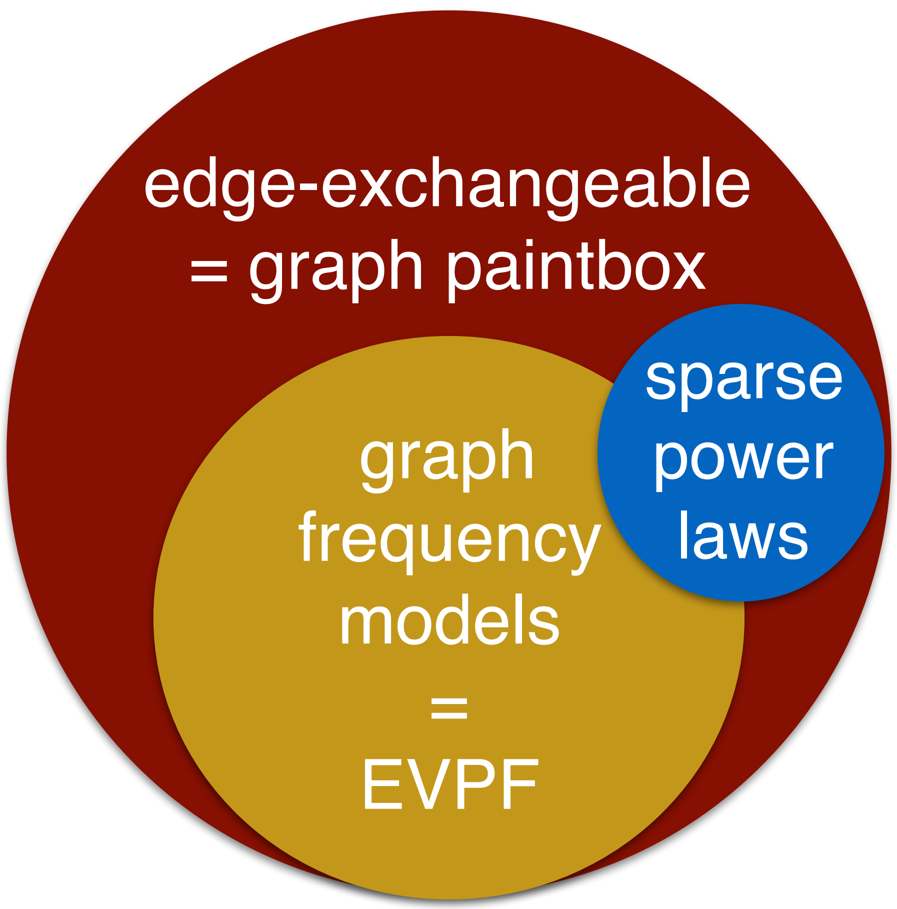


future work:

- characterize sparse, edge-exchangeable graph models
- characterize various types of sparse power laws (e.g., degrees, triangles)

conclusions

✓characterized the class of edge-exchangeable graphs
✓characterized the class of graph frequency models



future work:

- characterize sparse, edge-exchangeable graph models
- characterize various types of sparse power laws (e.g., degrees, triangles)
- truncation and practical posterior inference algorithms (frequency models and EVPF)

references

[1] **T. Campbell, D. Cai, and T. Broderick. Exchangeable trait allocations. Submitted, 2016. https://arxiv.org/abs/1609.09147**

Alternate versions in:

- ***NIPS 2016 Workshop on Practical Bayesian Nonparametrics.***
- ***NIPS 2016 Workshop on Adaptive and Scalable Nonparametric Methods in Machine learning. (2016b)***

[2] **D. Cai, T. Campbell, and T. Broderick. Edge-exchangeable graphs and sparsity. NIPS 2016. (2016a)**

Preliminary versions in:

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

[3] D. J. Aldous. Representations for partially exchangeable arrays of random variables. J. Multivariate Anal., 11(4): 581–598, 1981.

[4] T. Broderick, J. Pitman, and M. I. Jordan. Feature allocations, probability functions, and paintboxes. Bayesian Analysis. 2013.

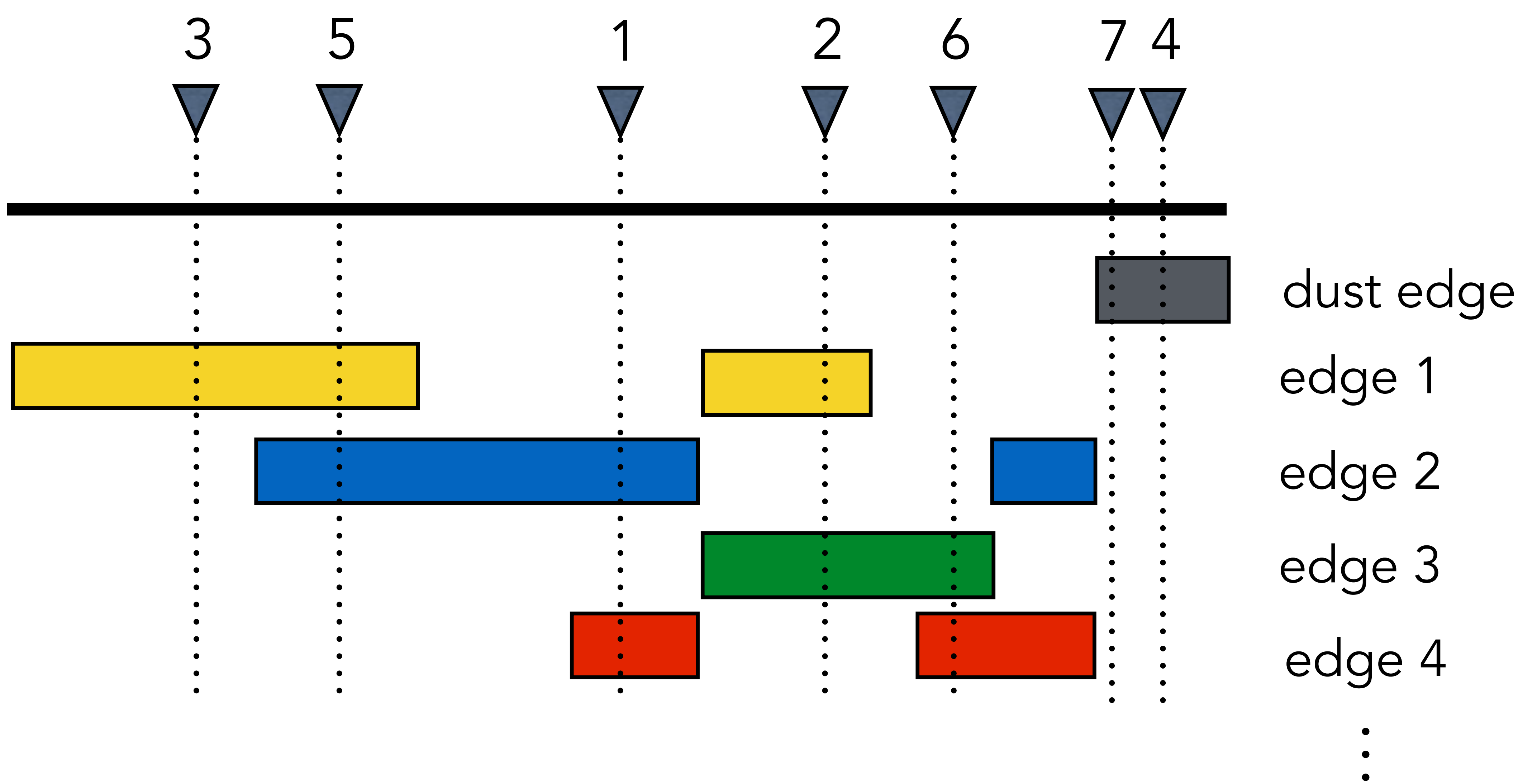
[5] T. Campbell*, J.H. Huggins*, J. How, and T. Broderick. Truncated random measures. Submitted. ArXiv: 1603.00861.

[6] H. Crane and W. Dempsey. A framework for statistical network modeling. https://arxiv.org/abs/1509.08185.

[7] H. Crane and W. Dempsey. Edge exchangeable graphs for network data. https://arxiv.org/abs/1603.04571

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

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



[Broderick and Cai, 2015; Cai, Campbell, Broderick, 2016]