The Topology of Preferential Attachment

The Asymptotics of the Expected Betti Numbers of Preferential Attachment Clique Complexes

Chunyin Siu
Cornell University
cs2323@cornell.edu

So, preferential attachment…

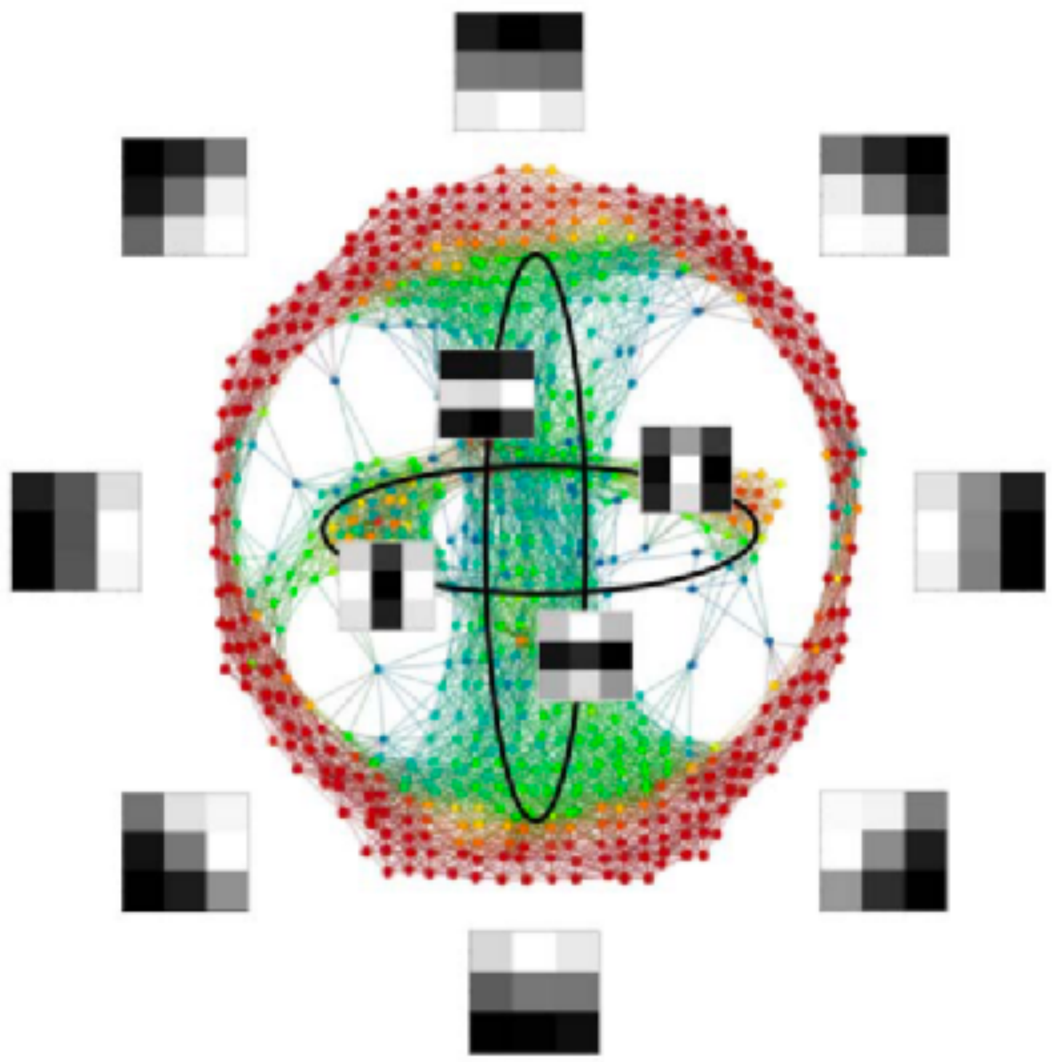
(Stephen Coast
https://www.fractalus.com/steve/stuff/ipmap/)

So, preferential attachment…

- Just a bouquet of circles?

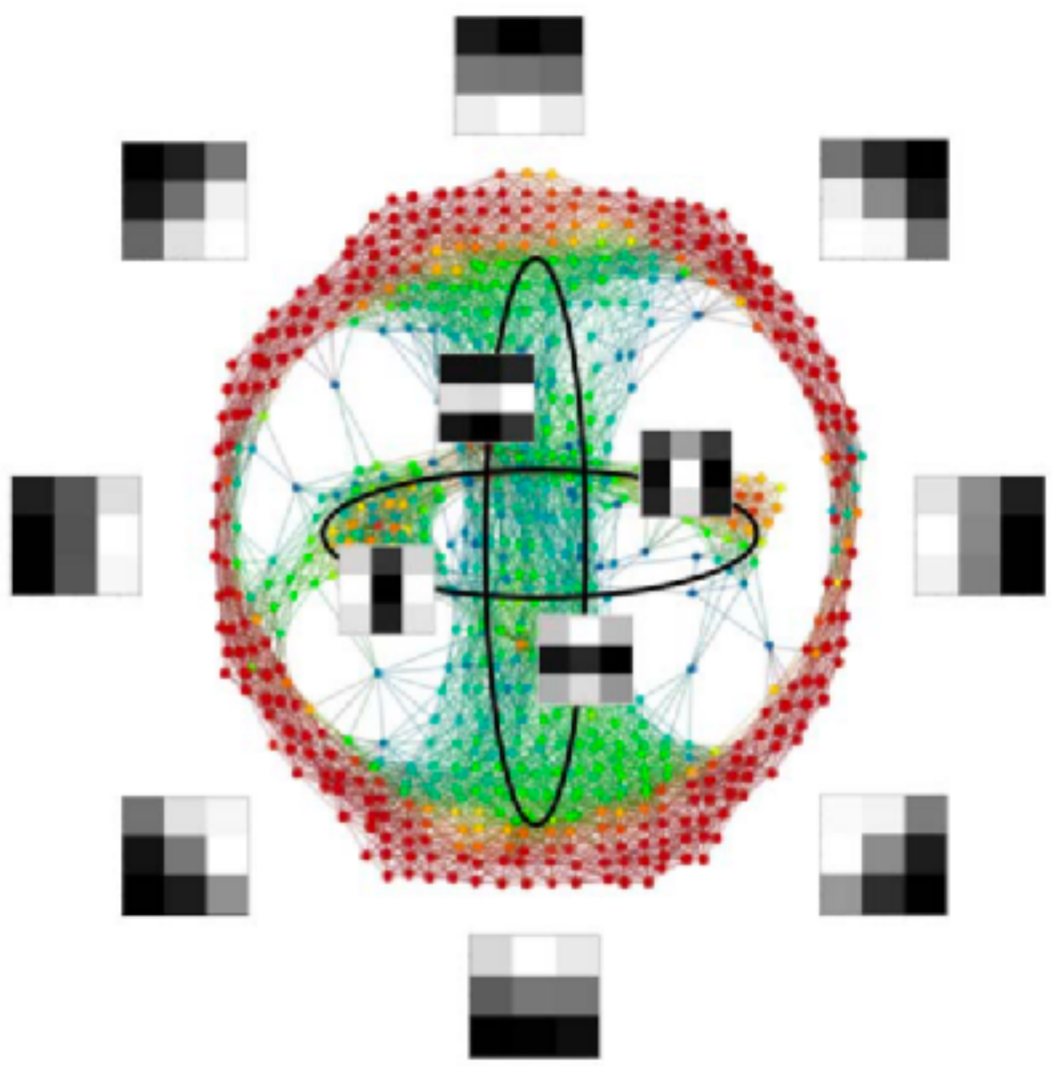
(Stephen Coast
https://www.fractalus.com/steve/stuff/ipmap/)

Agenda



topological data analysis

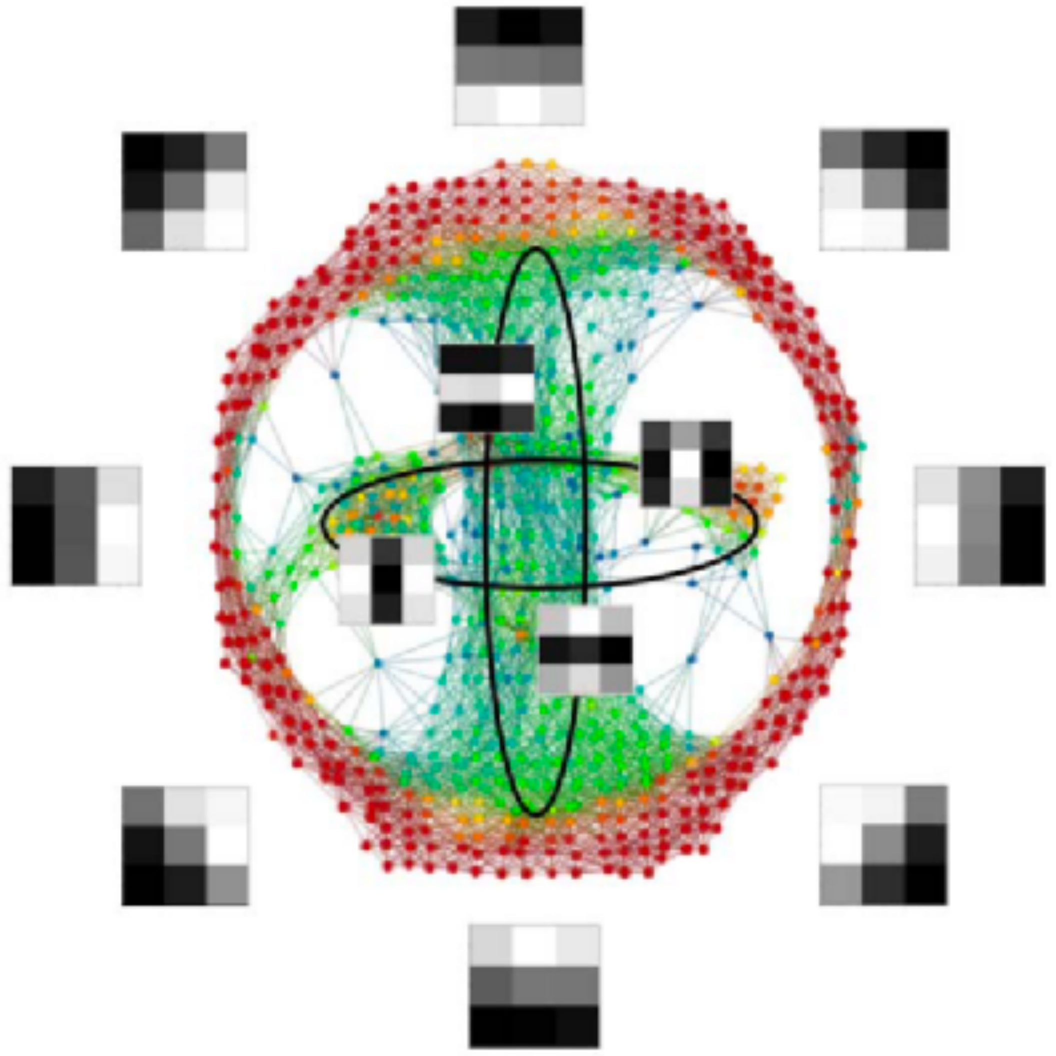
Agenda



topological data analysis

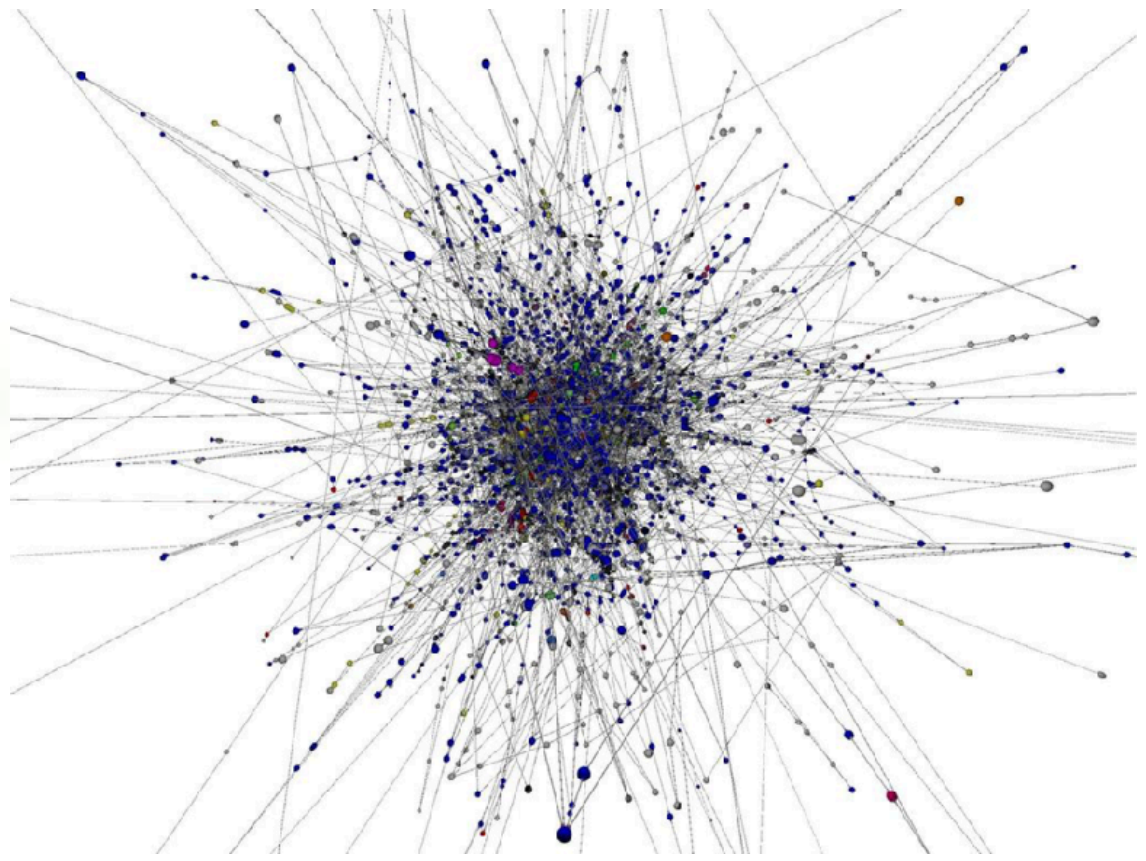
stochastic topology

Agenda



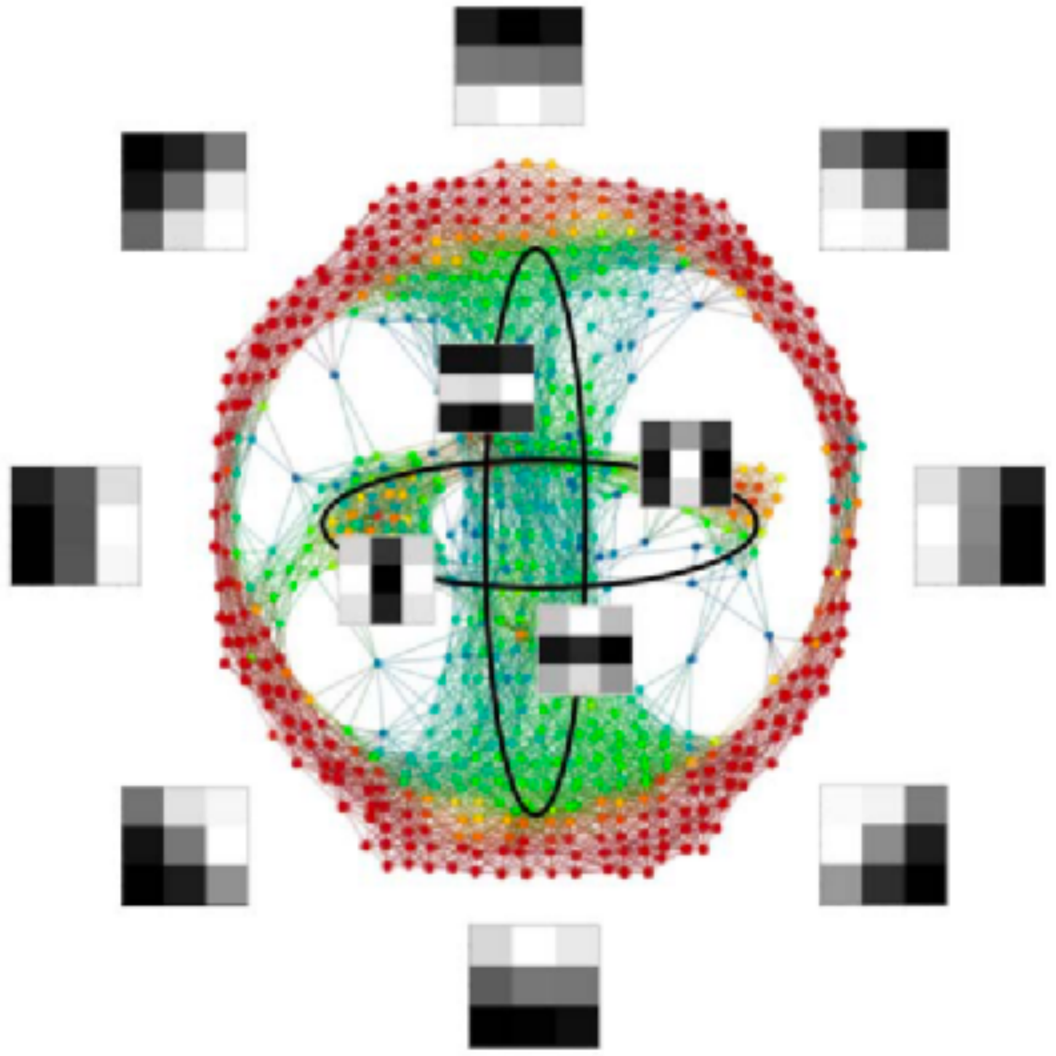
topological data analysis

stochastic topology



preferential attachment

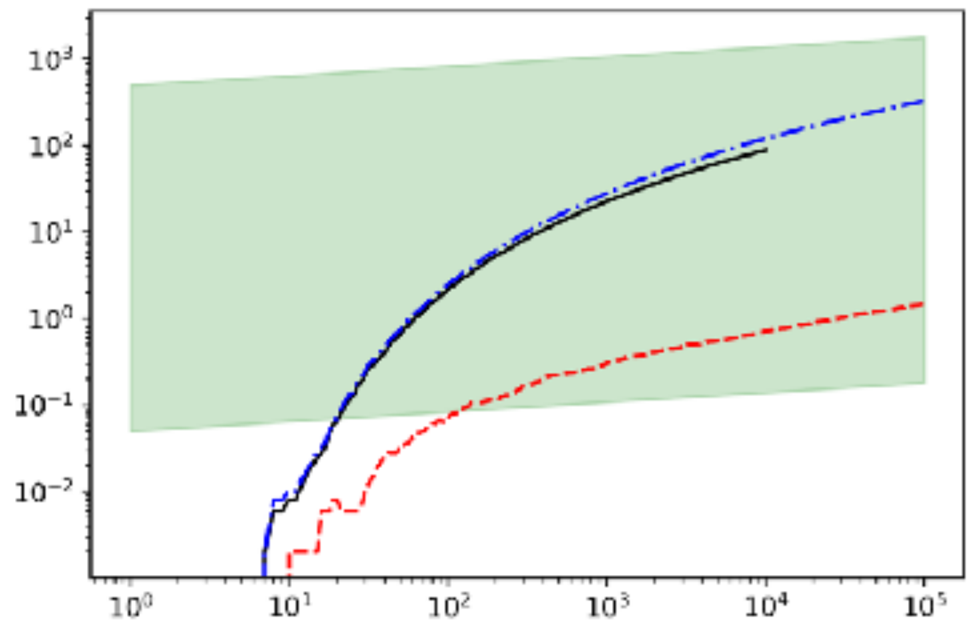
Agenda



topological data analysis

stochastic topology

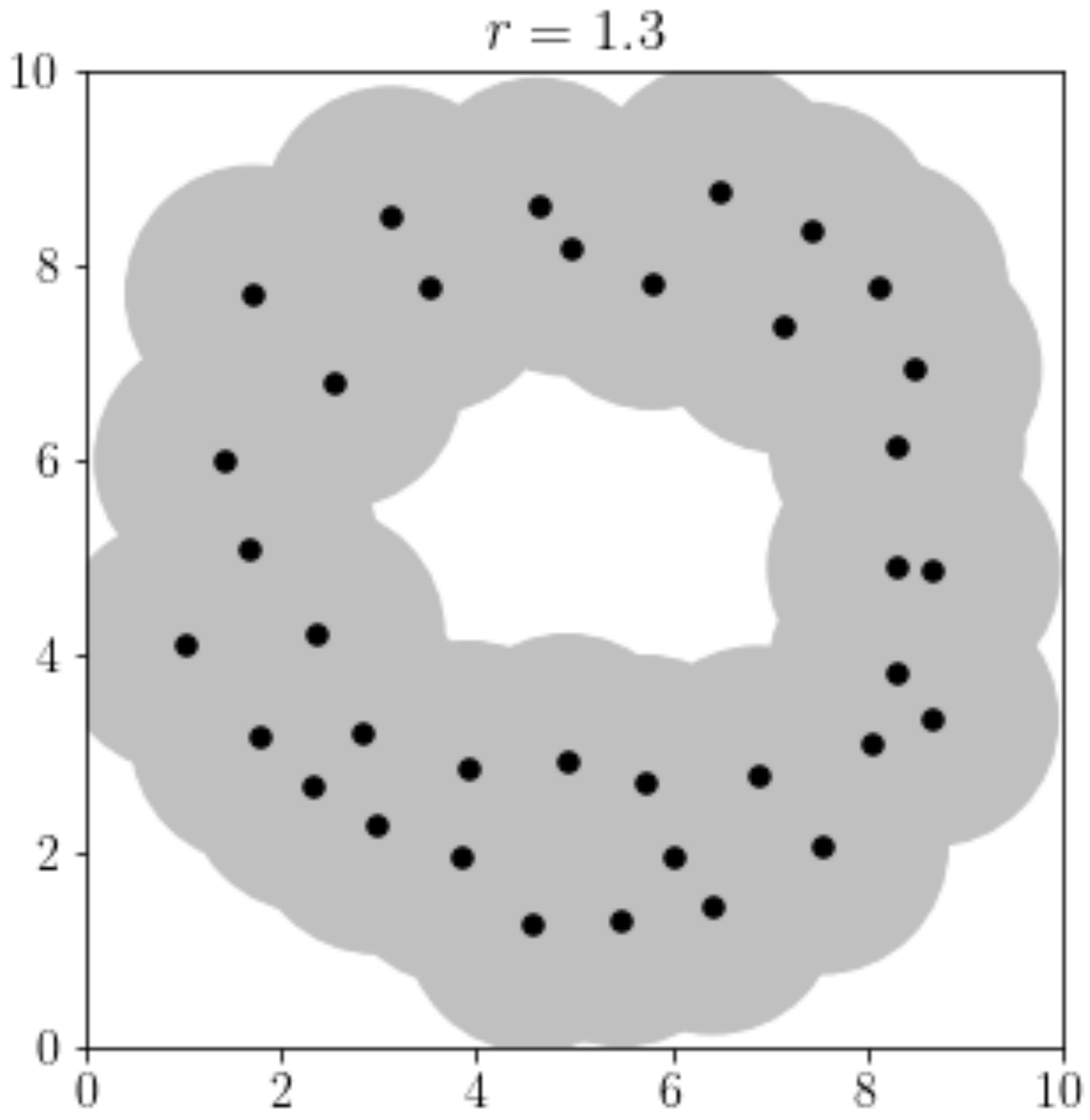
preferential attachment



our result

I. Topological Data Analysis

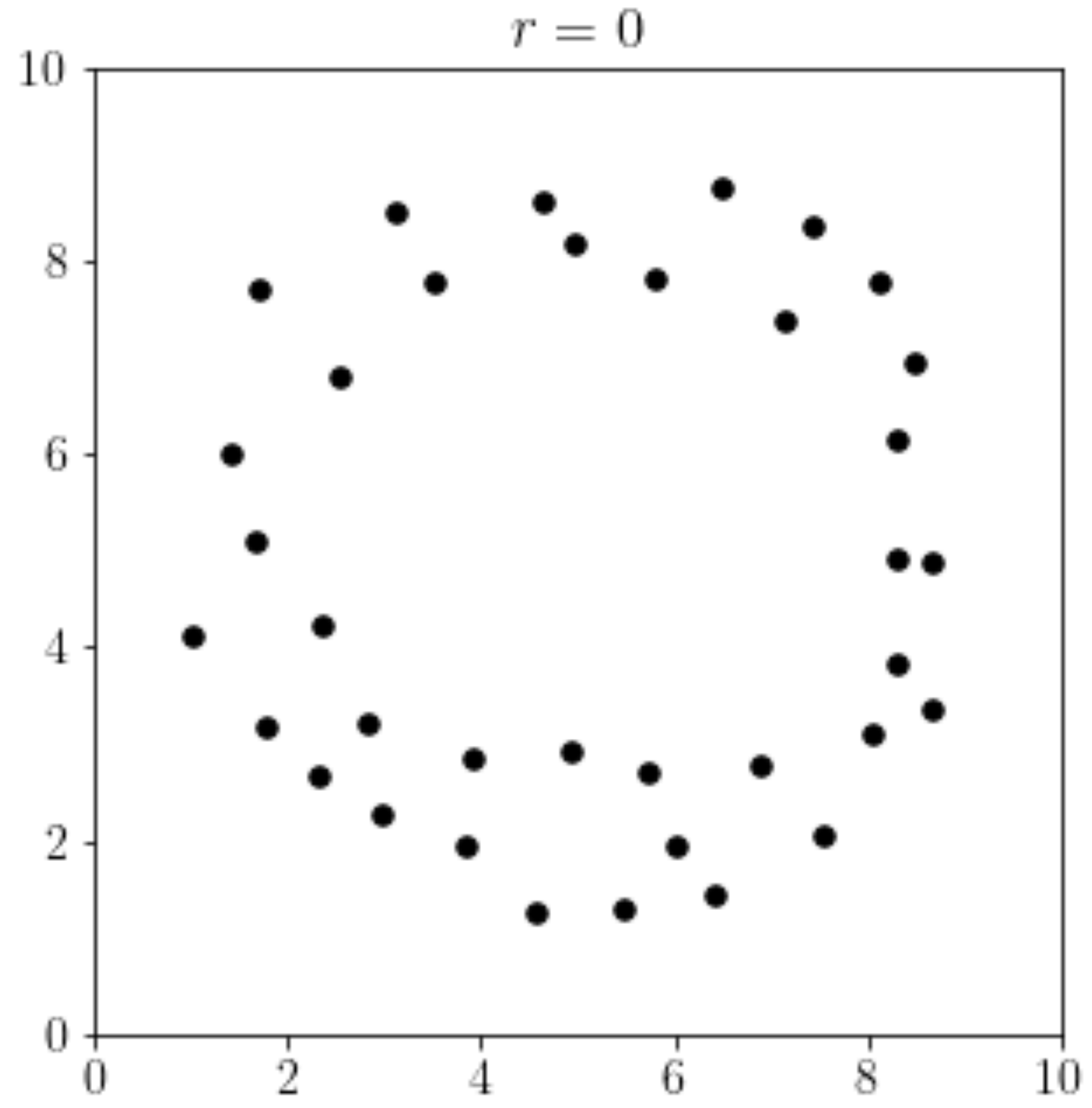
Two Approaches



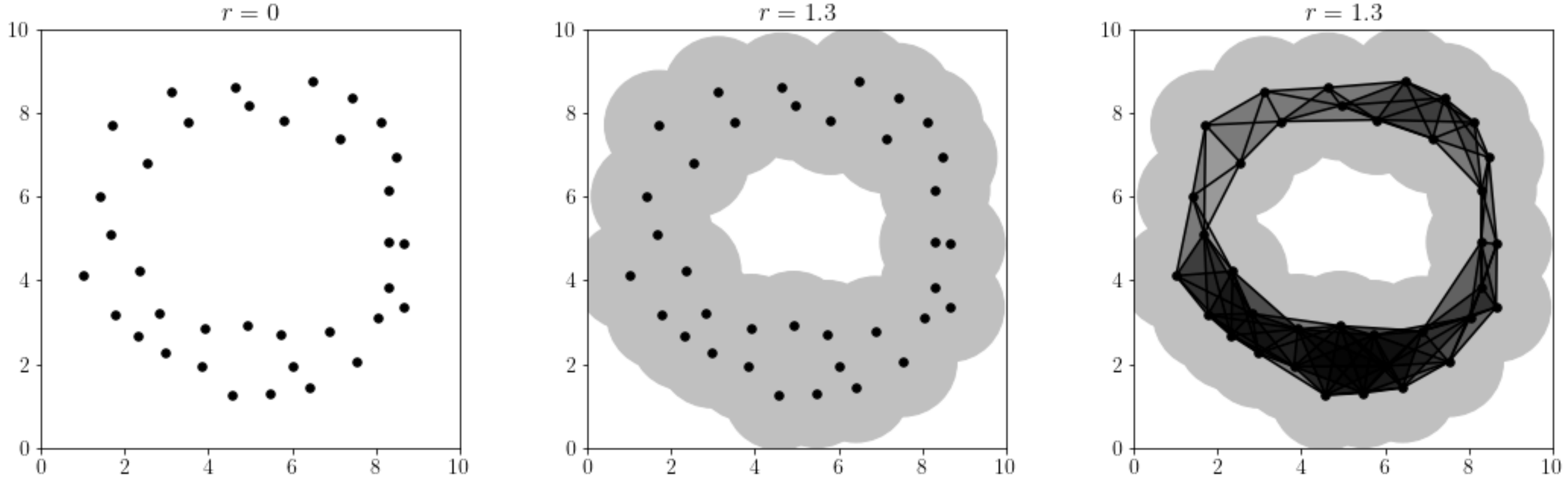
points

networks and complexes

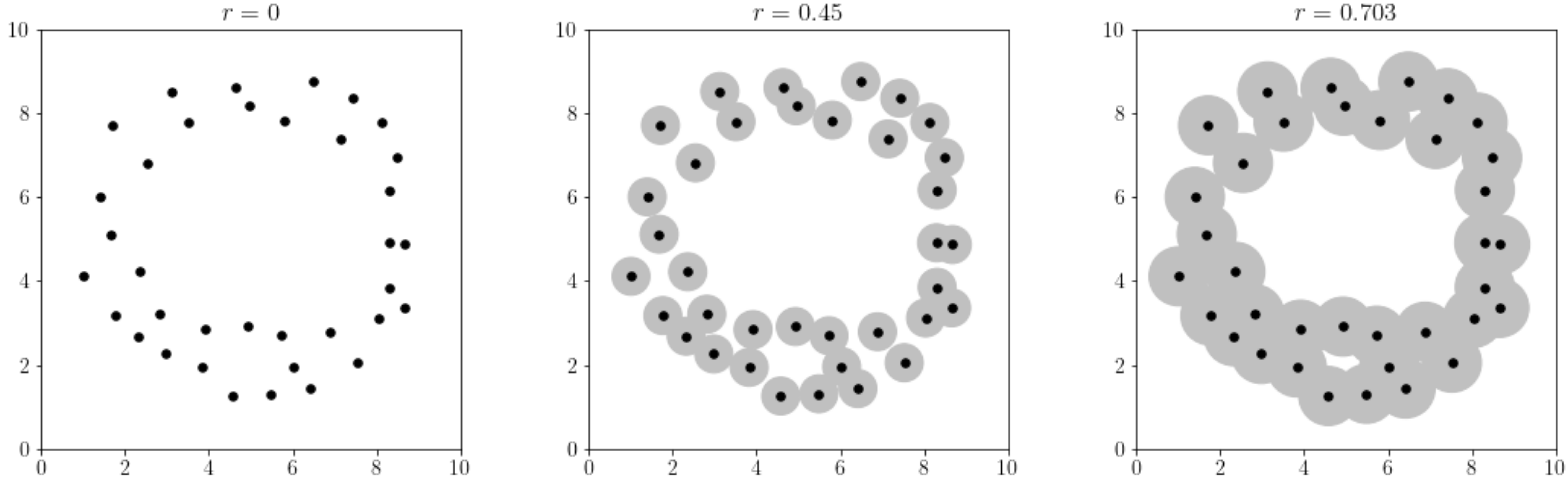
Points



Points



Points



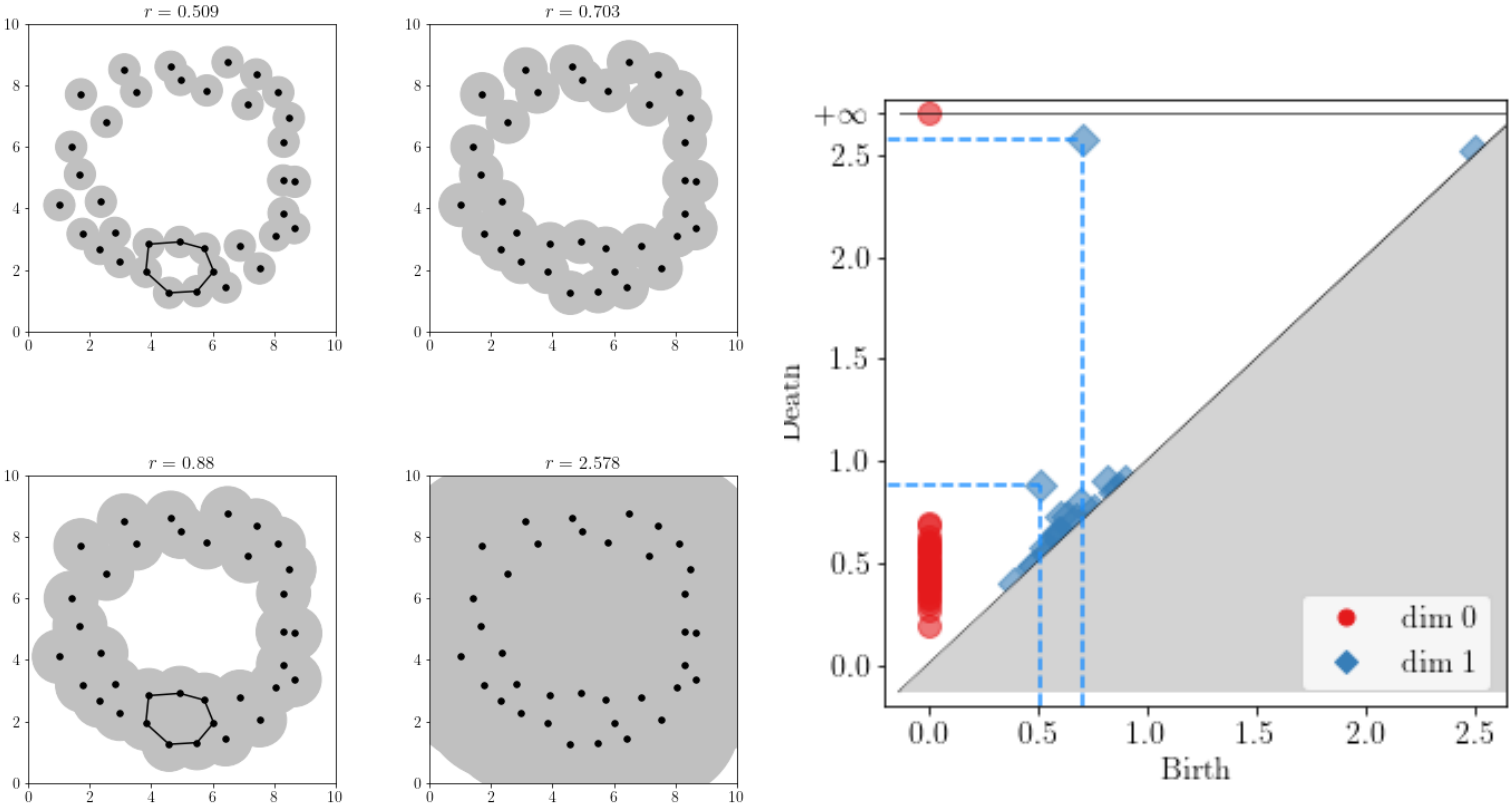
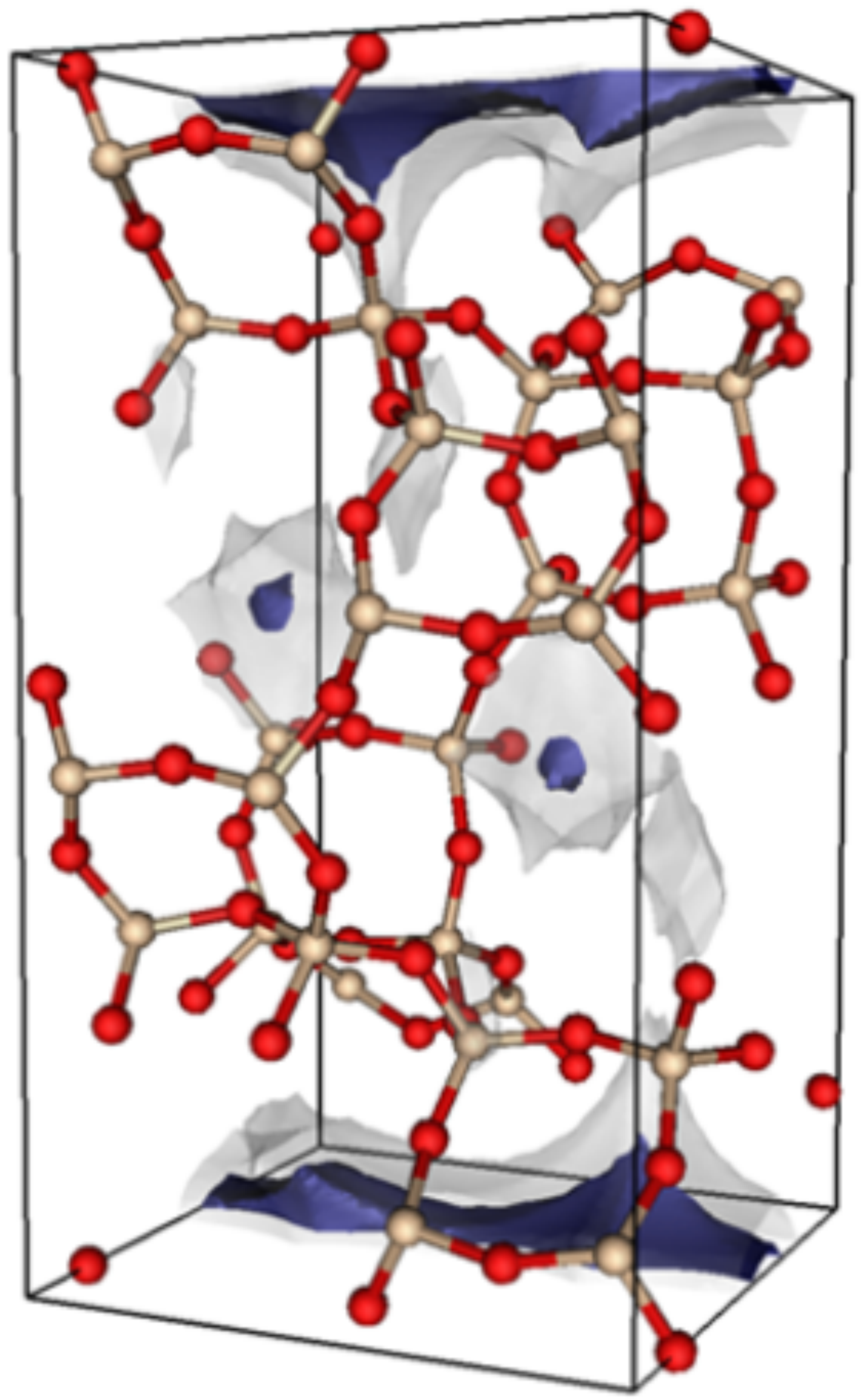


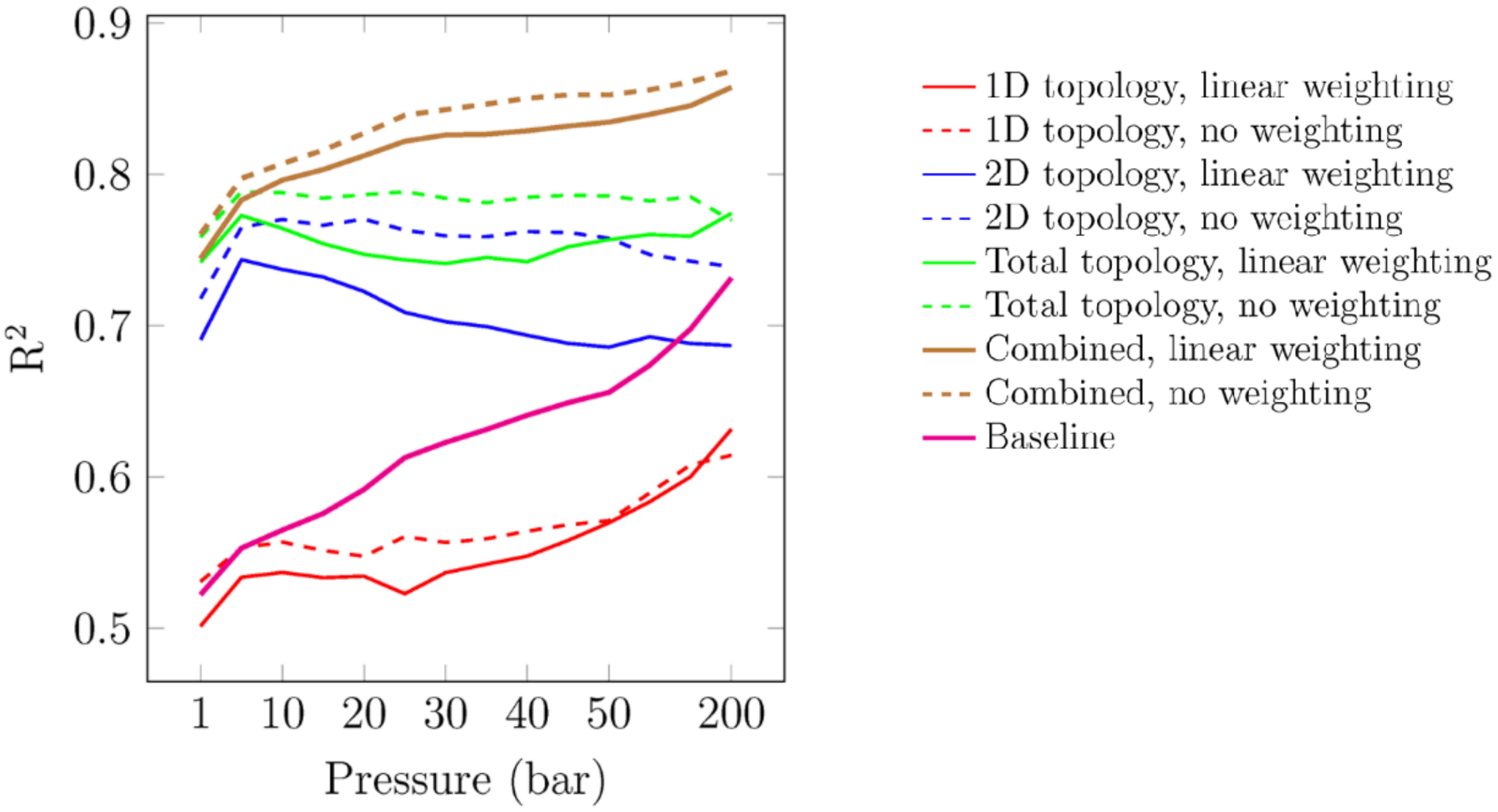
diagram credit: Andrey Yao

Zeolite crystals

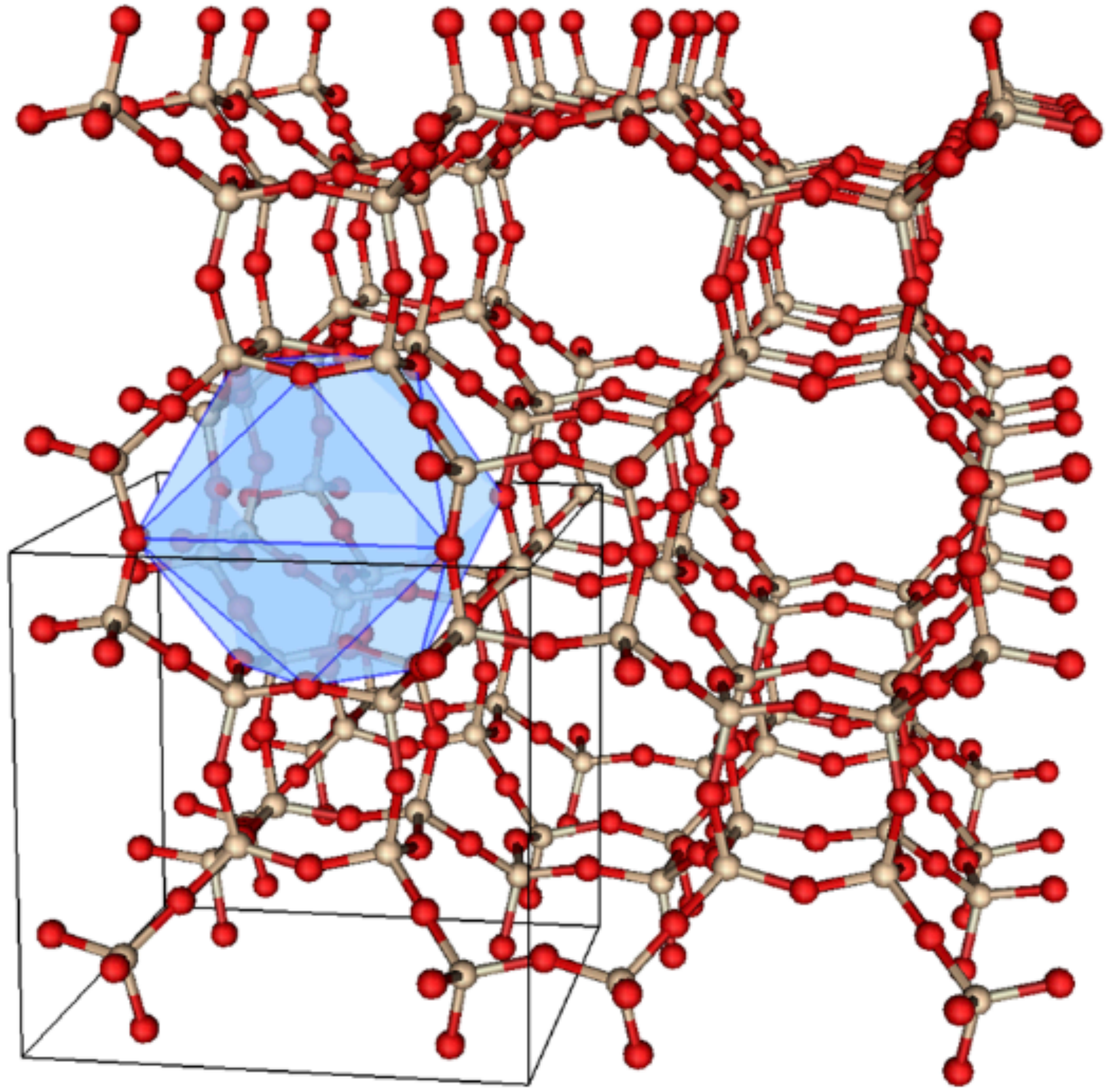
[Krishnapriyan et al, 2020]



Zeolite crystals
[Krishnapriyan et al, 2020]



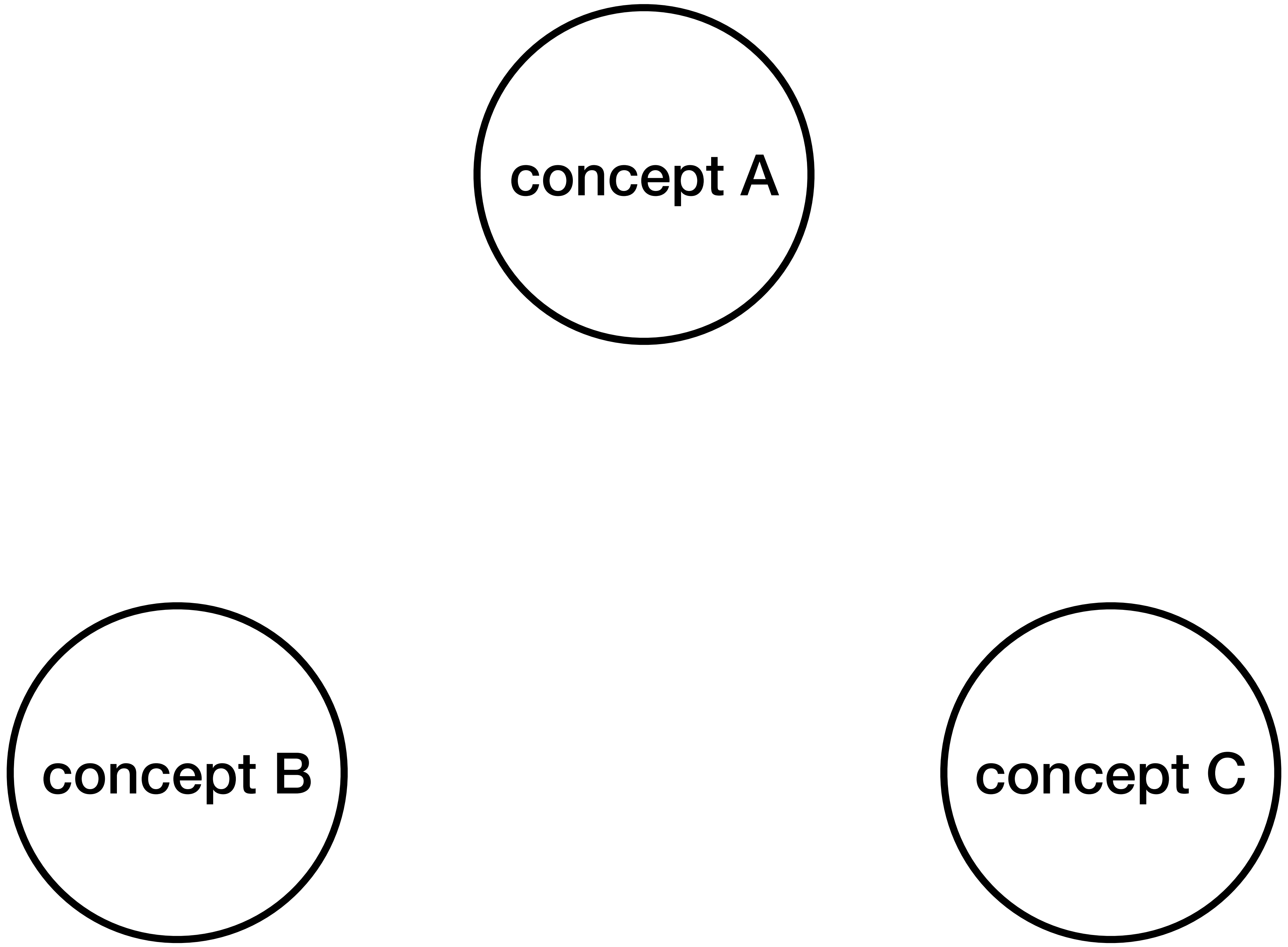
Zeolite crystals
[Krishnapriyan et al, 2020]

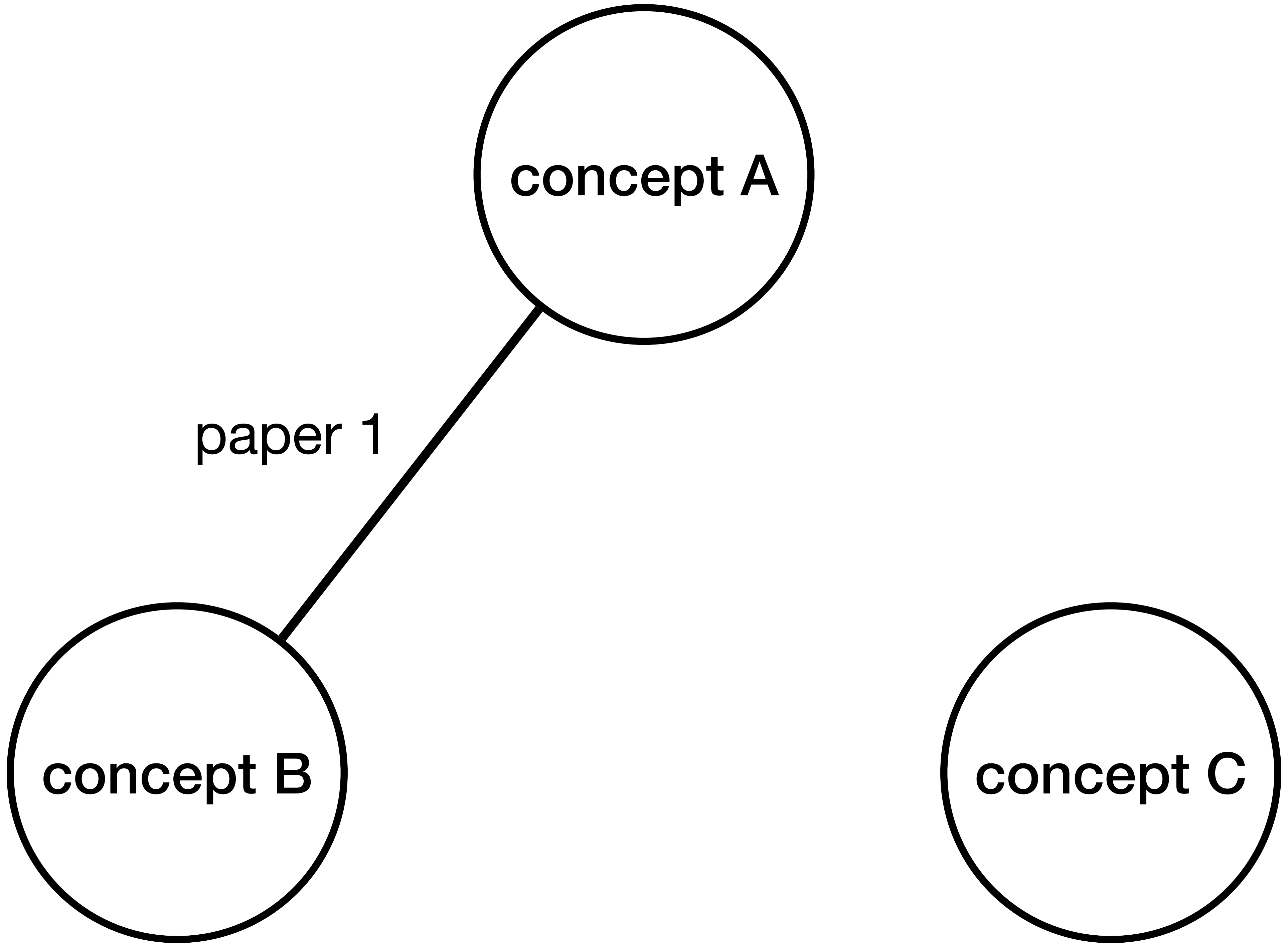


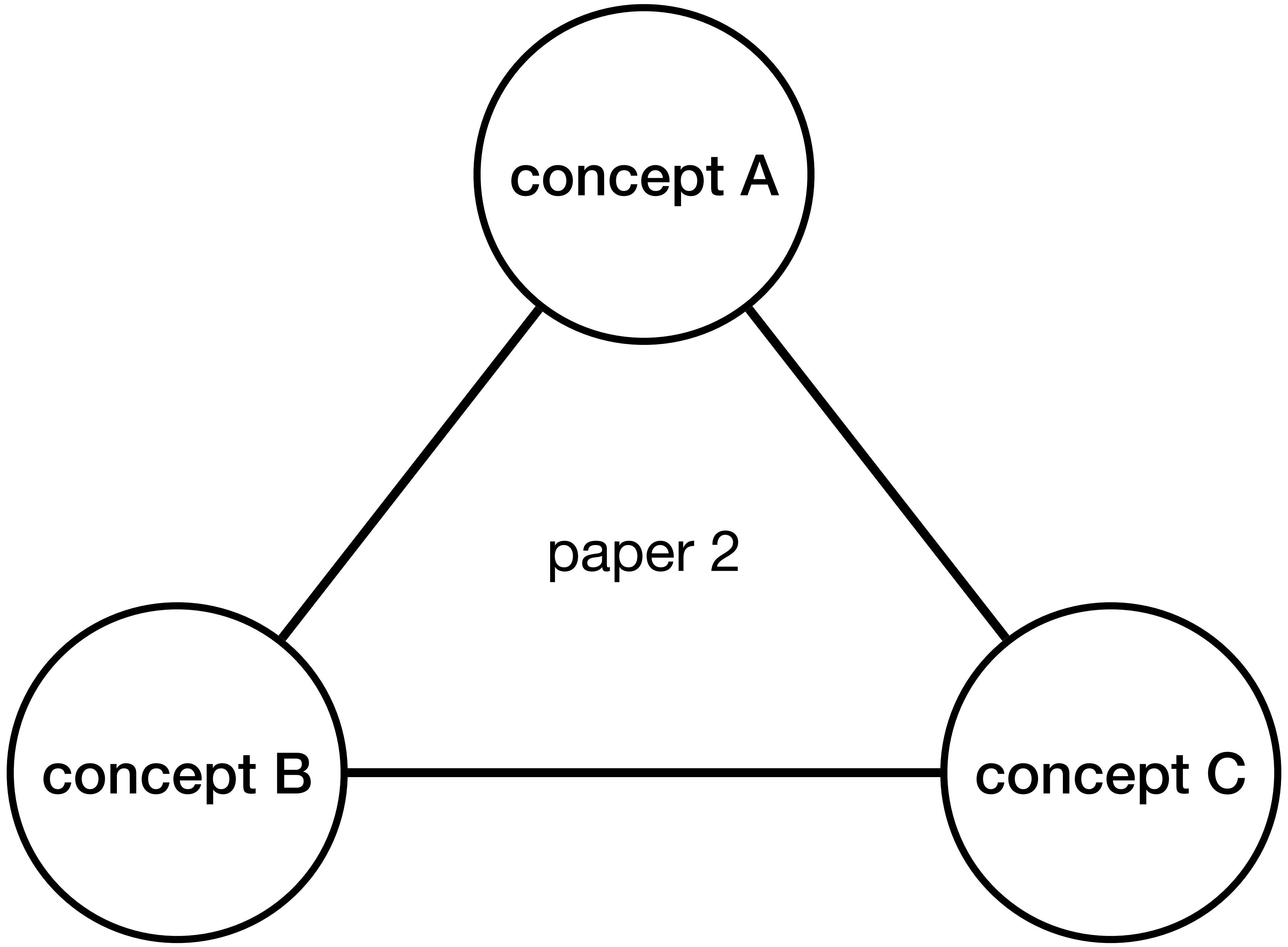
Networks and Complexes

Networks and Complexes

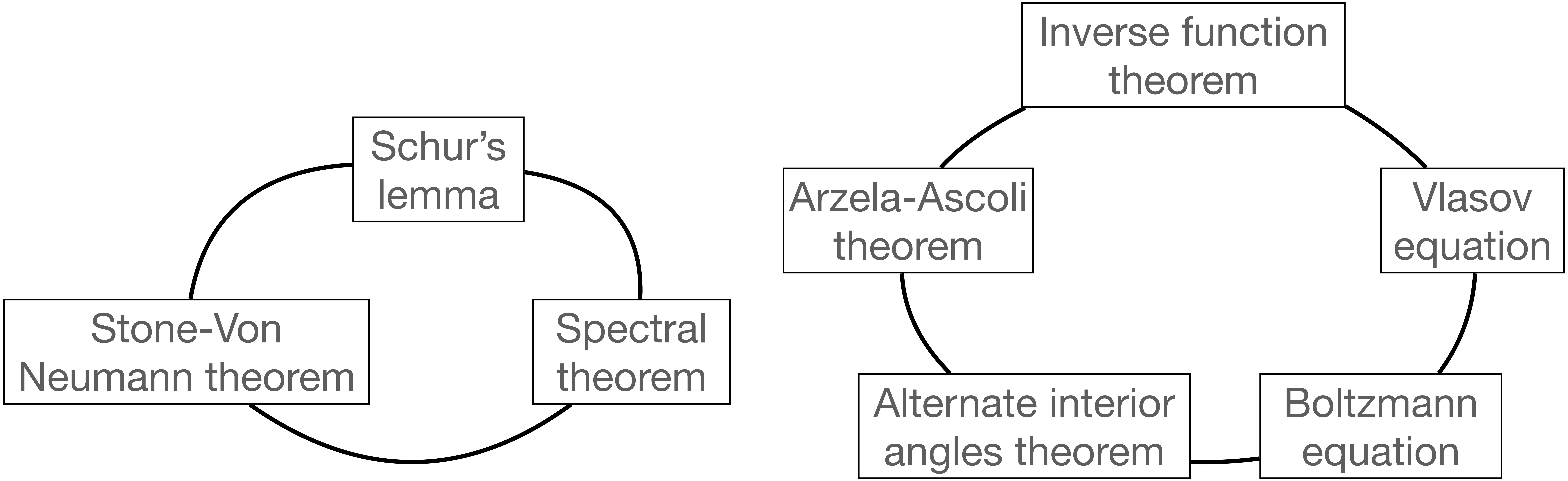
- Co-occurence complex in Math research paper [Salikov et al, 2018]


concept A
concept B
concept C


concept A
paper 1
concept B
concept C


concept A
paper 2
concept B
concept C

Gap in Understanding

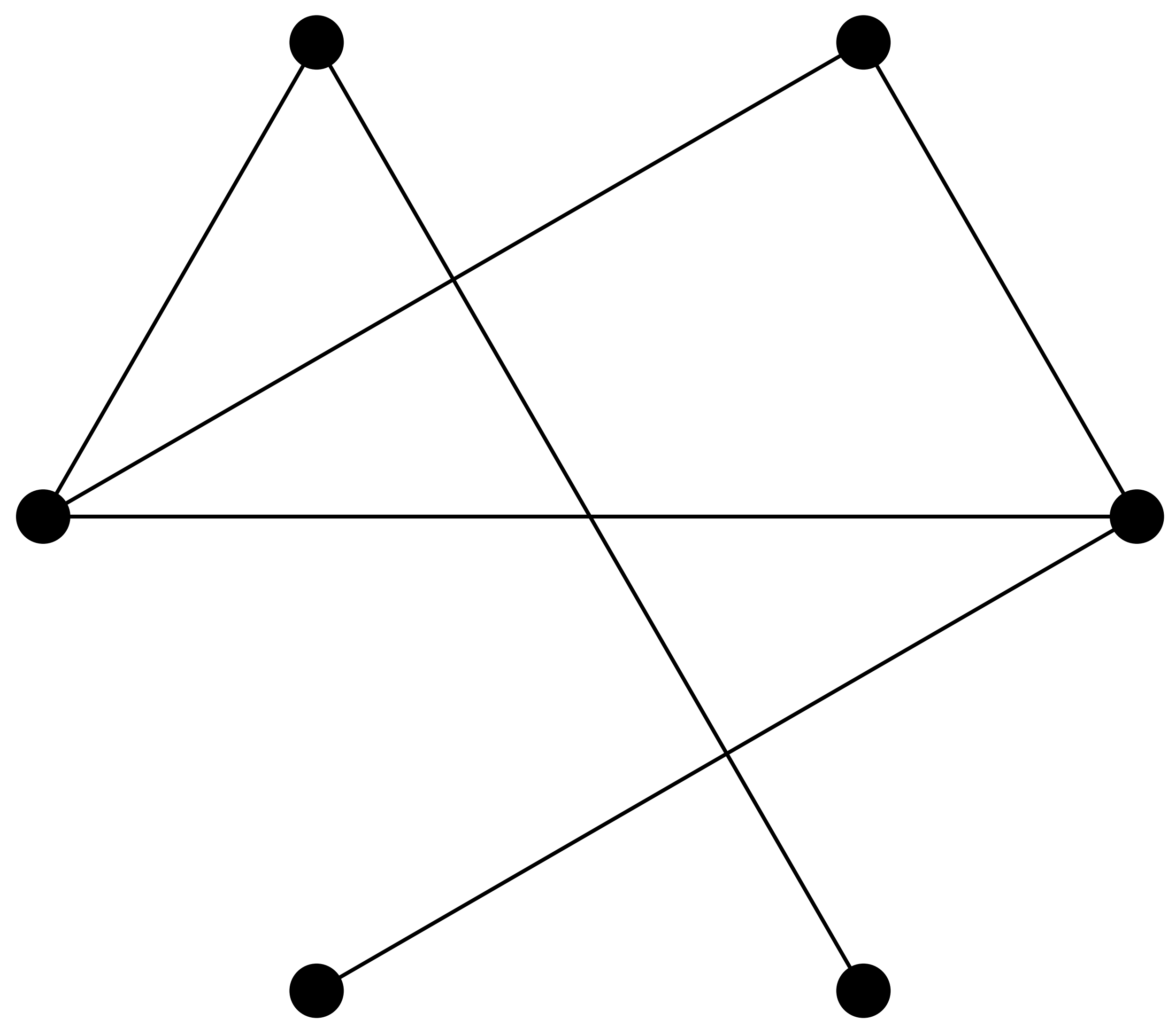


Benchmark of Comparison?

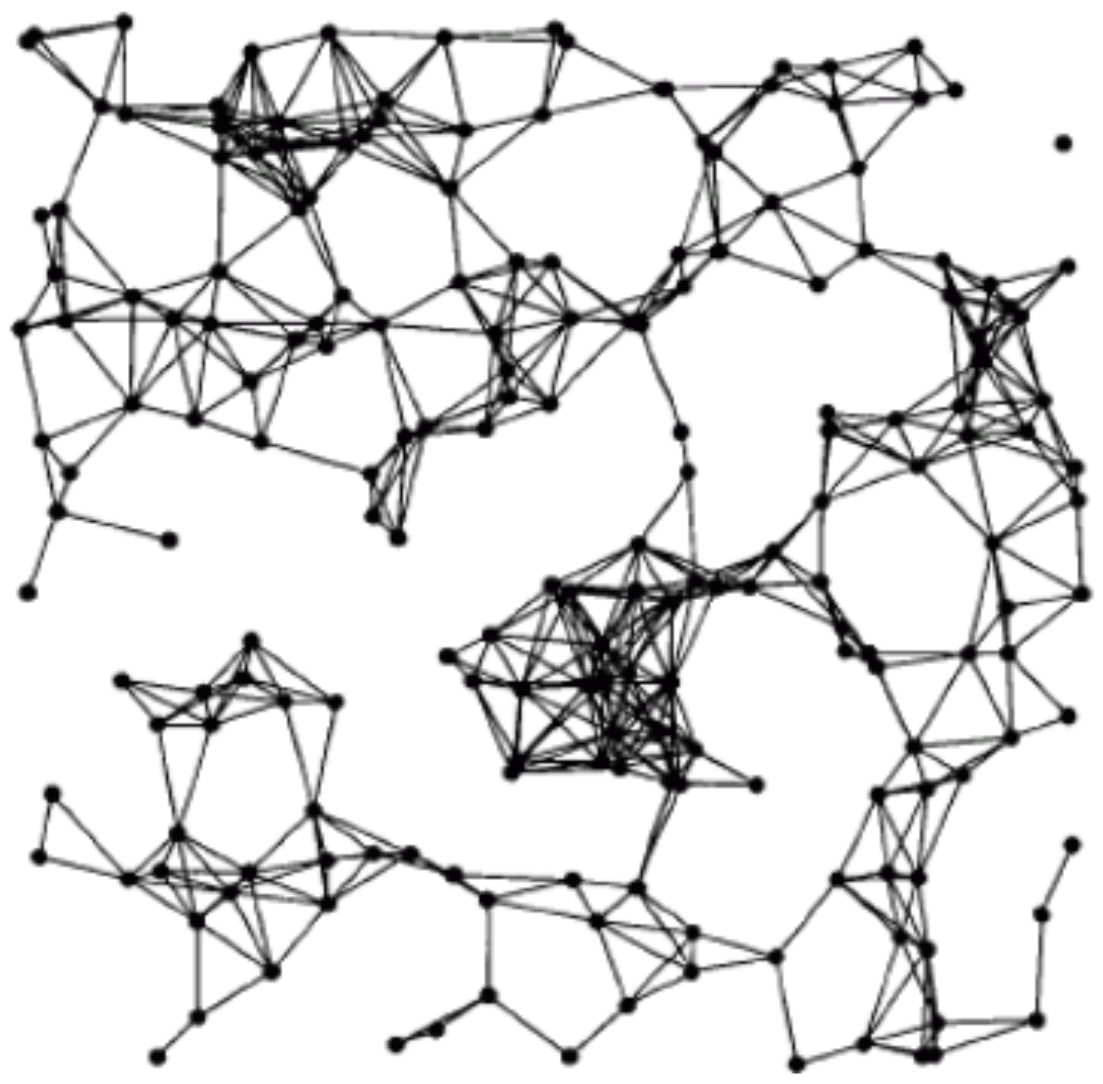
II. Stochastic Topology

Mug doesn’t play dice?

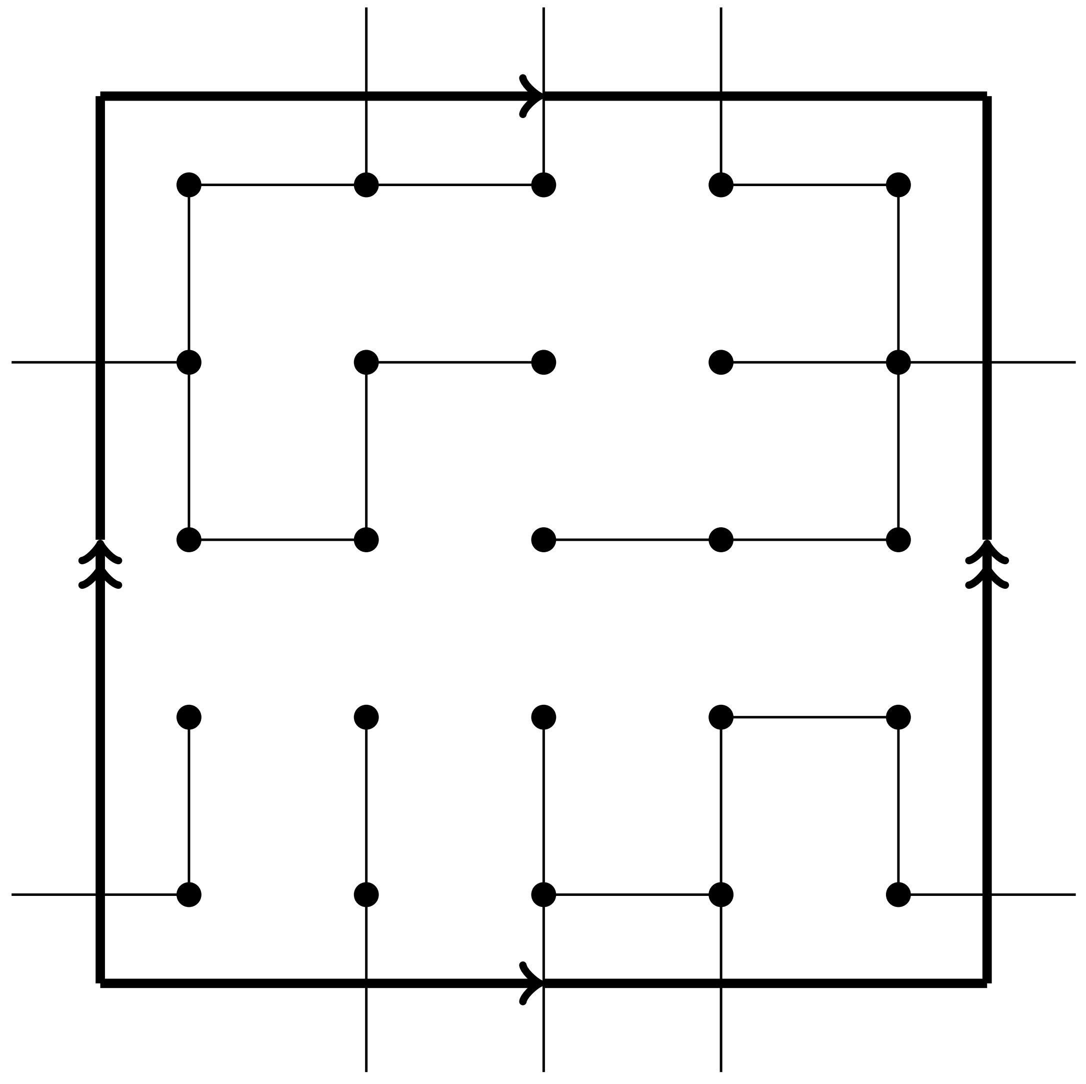
Tapas of Random Topology



Erdo-Renyi Complexes

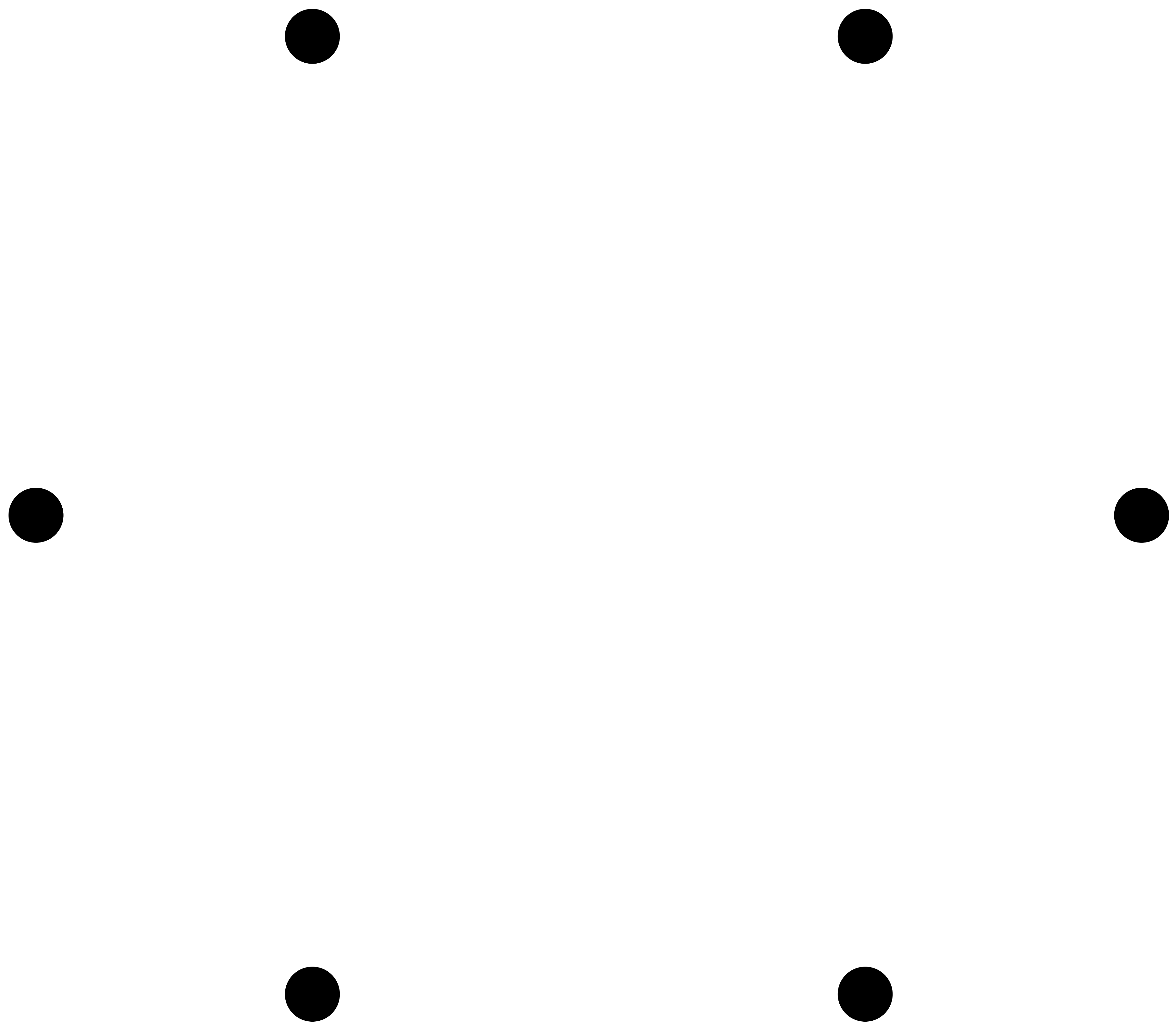


Geometric Complexes

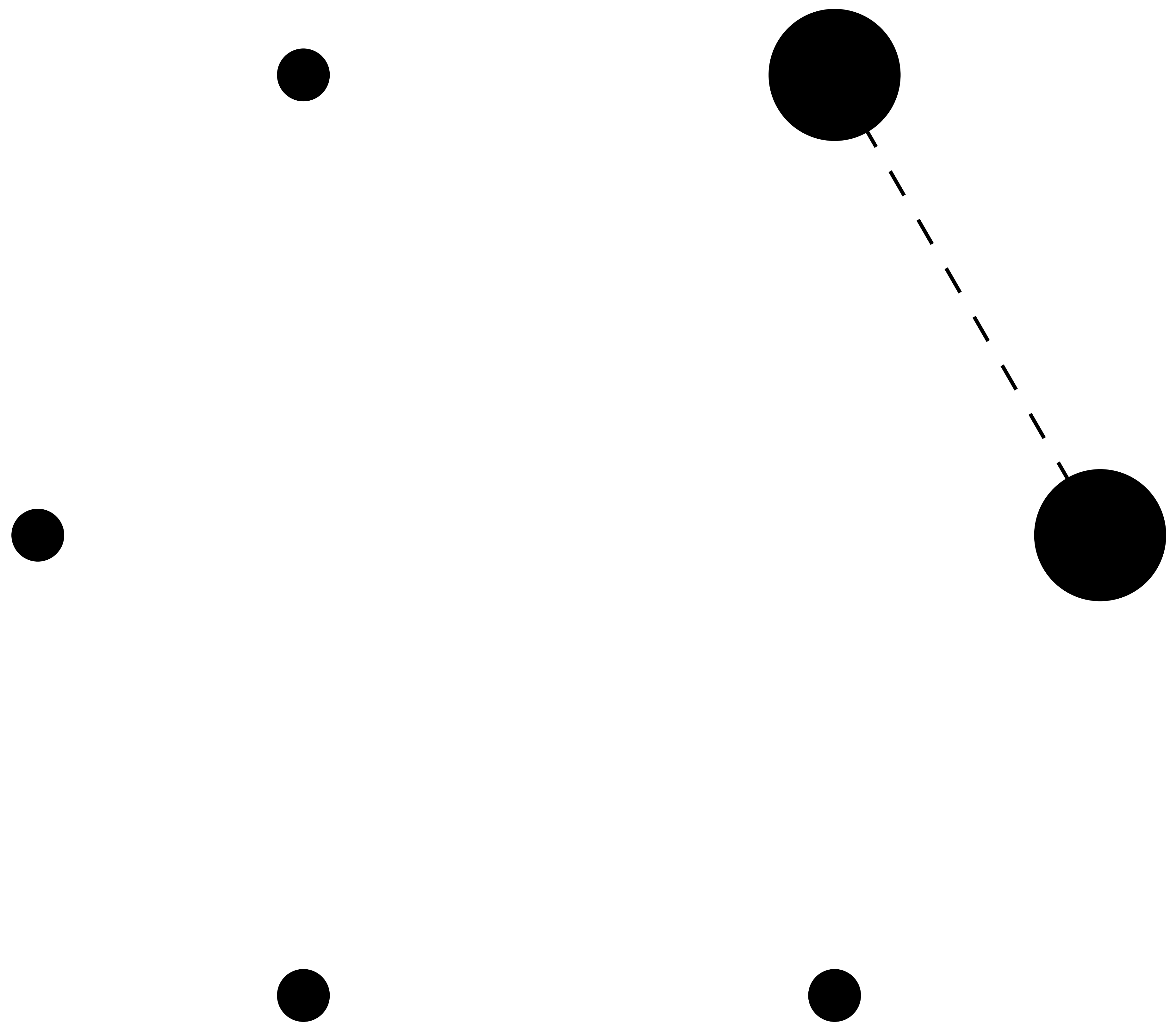


Topological Percolation

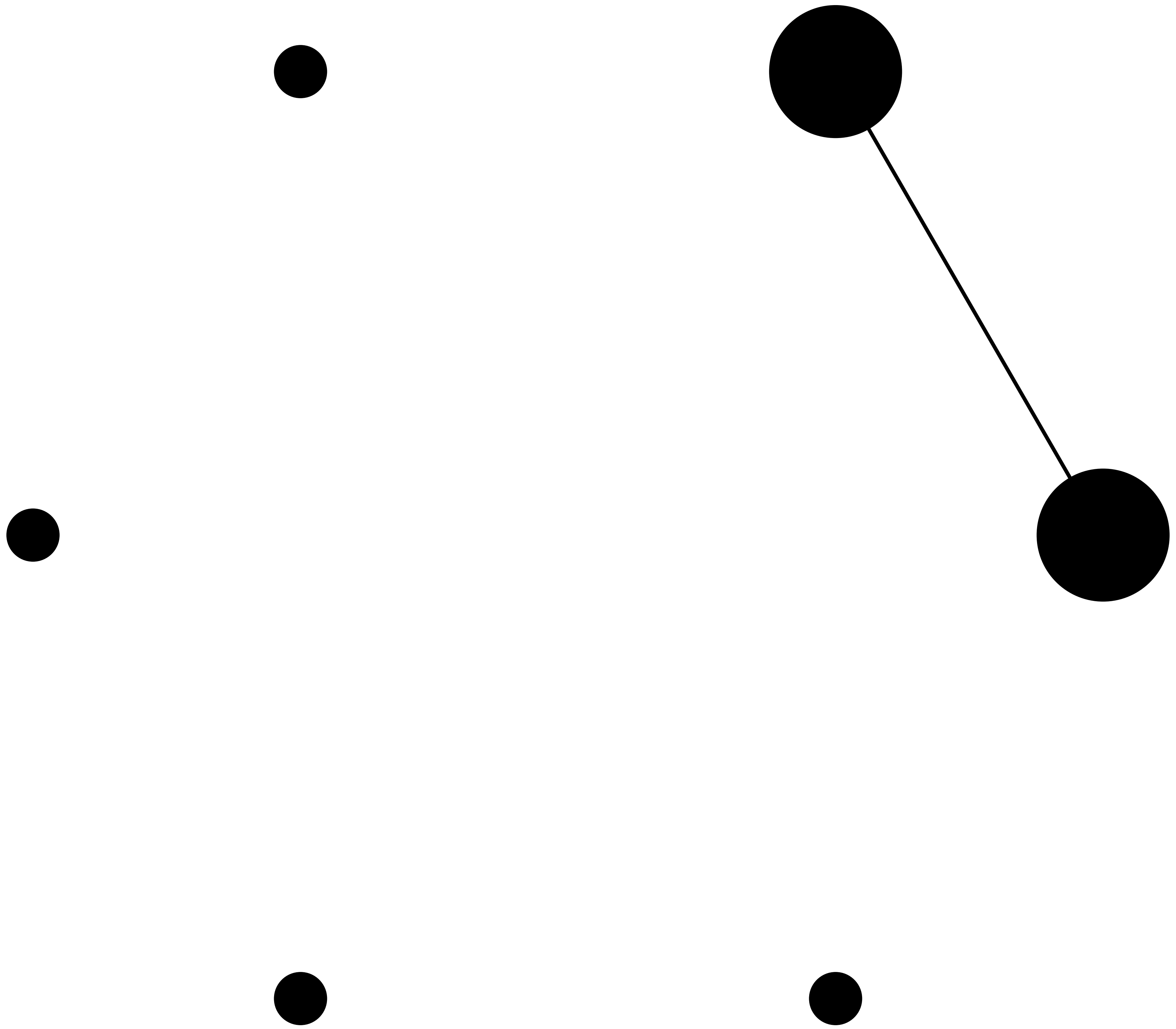
Erdos-Renyi graphs



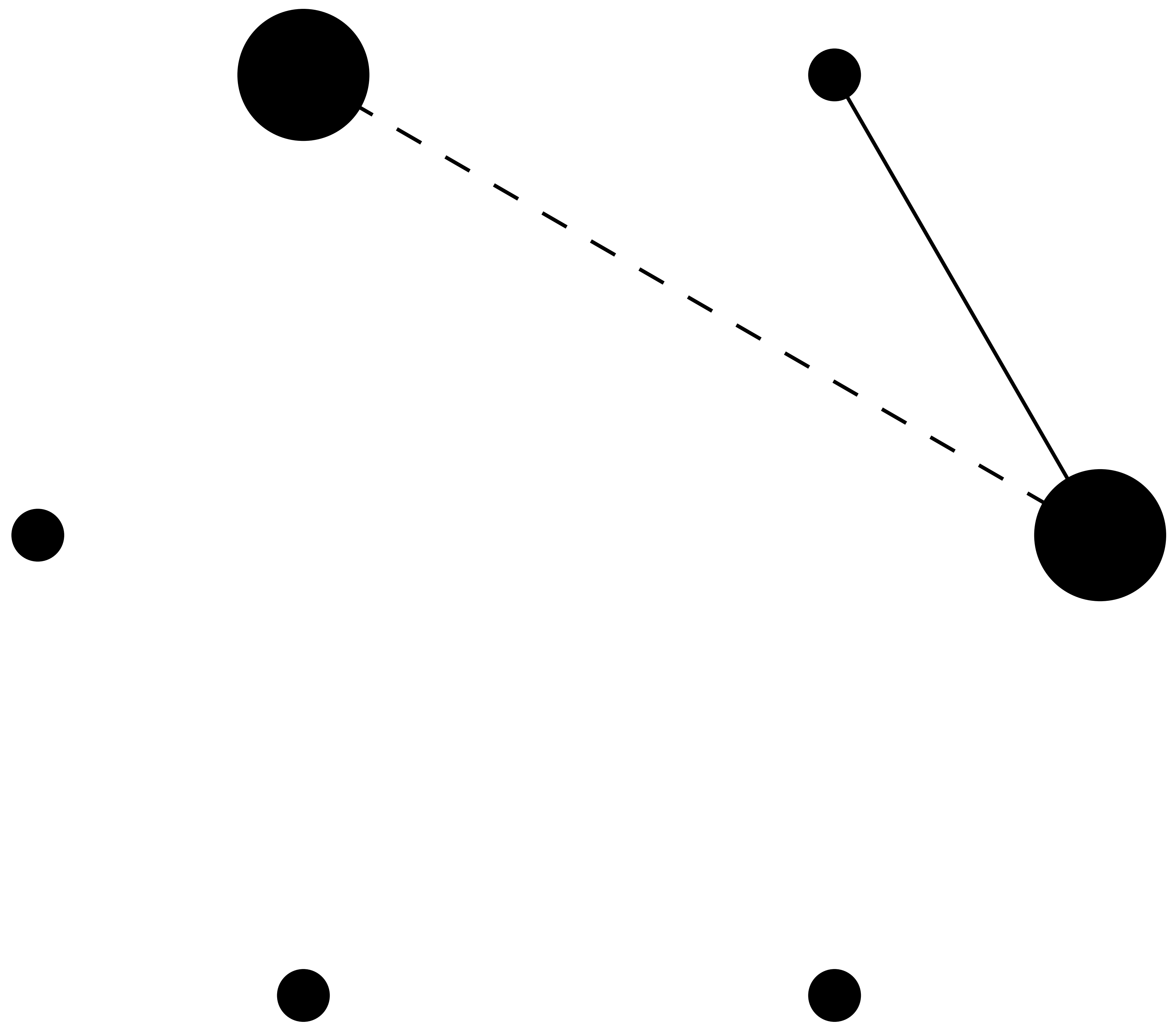
Erdos-Renyi graphs



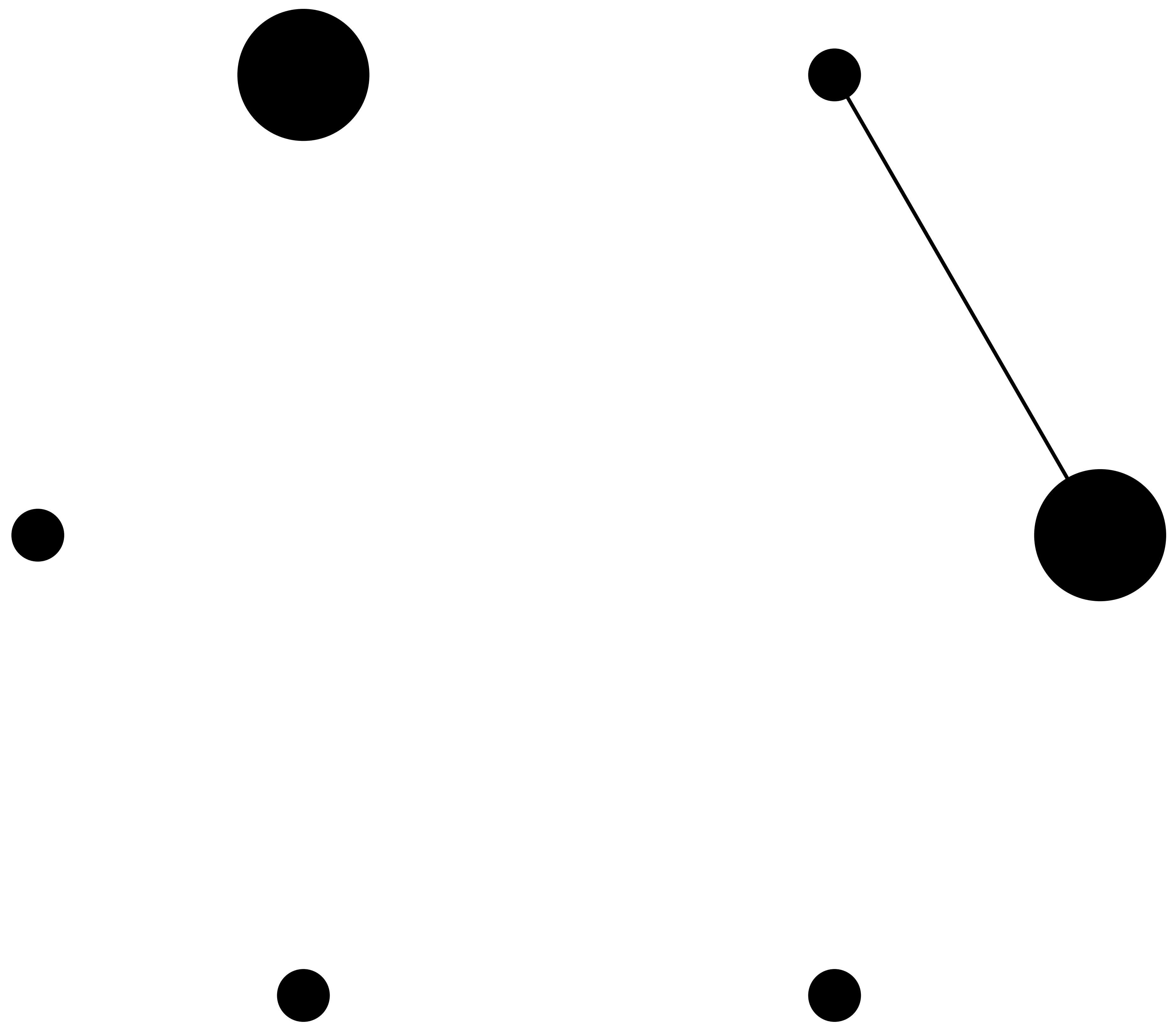
Erdos-Renyi graphs



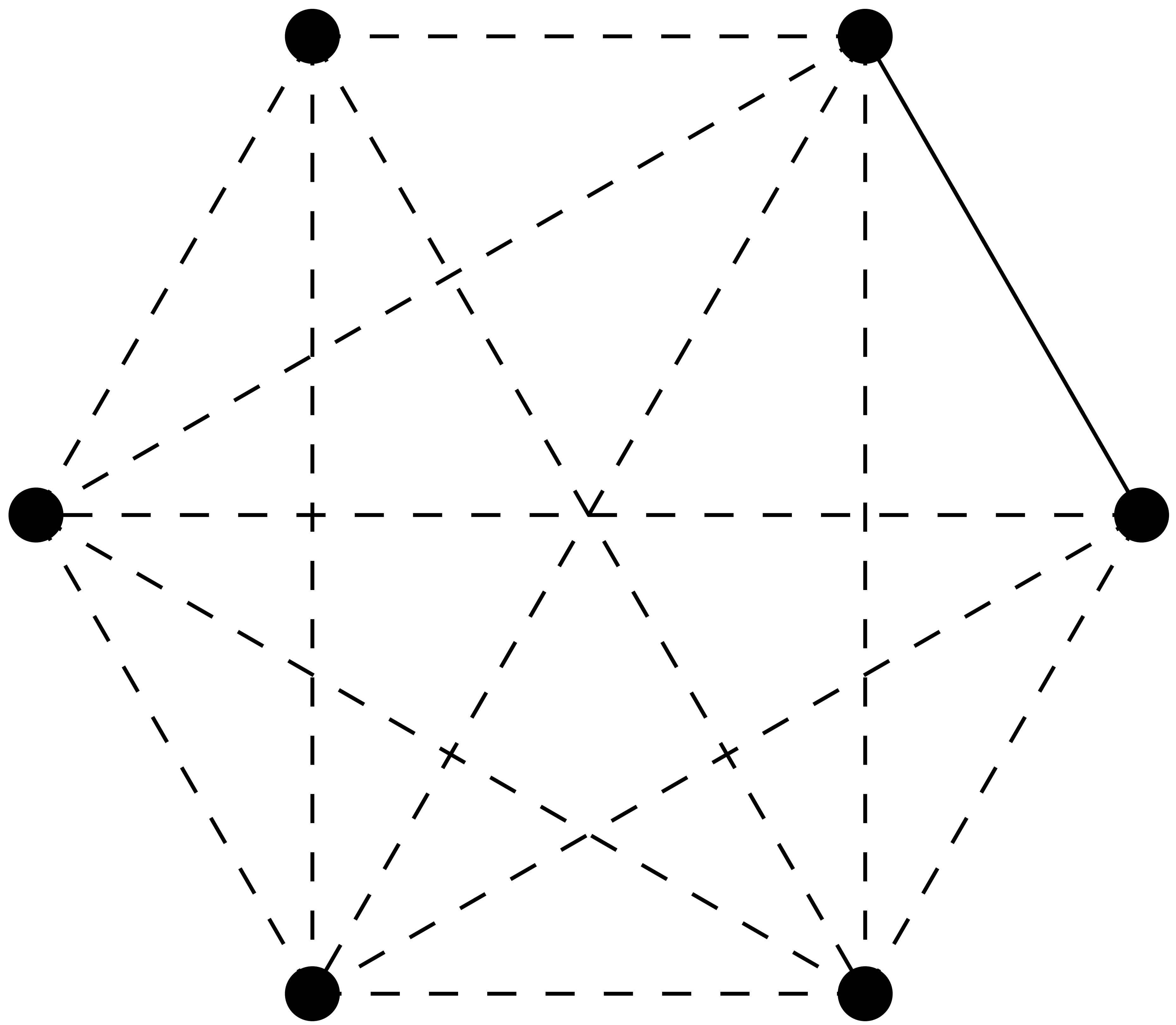
Erdos-Renyi graphs



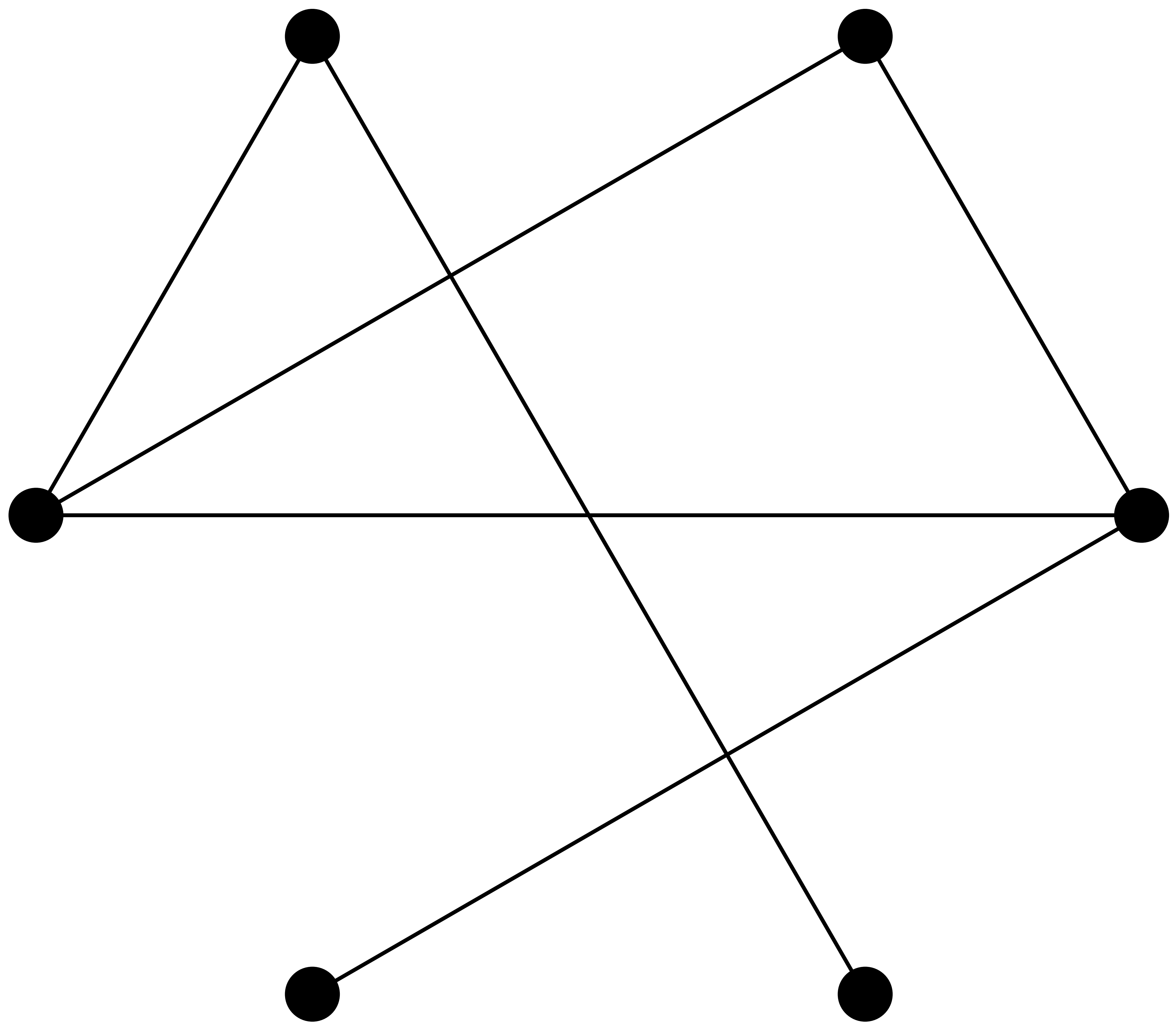
Erdos-Renyi graphs



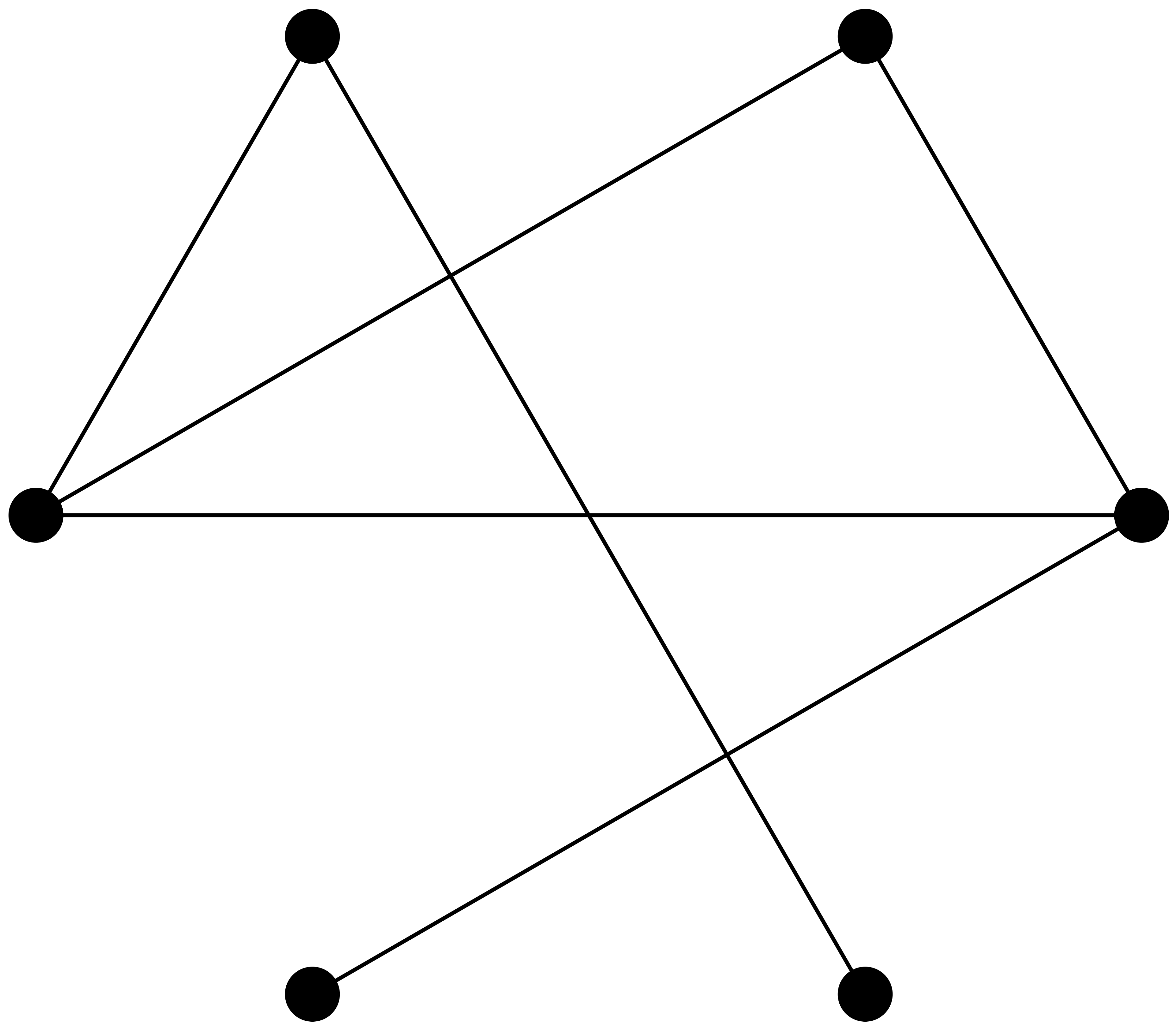
Erdos-Renyi graphs



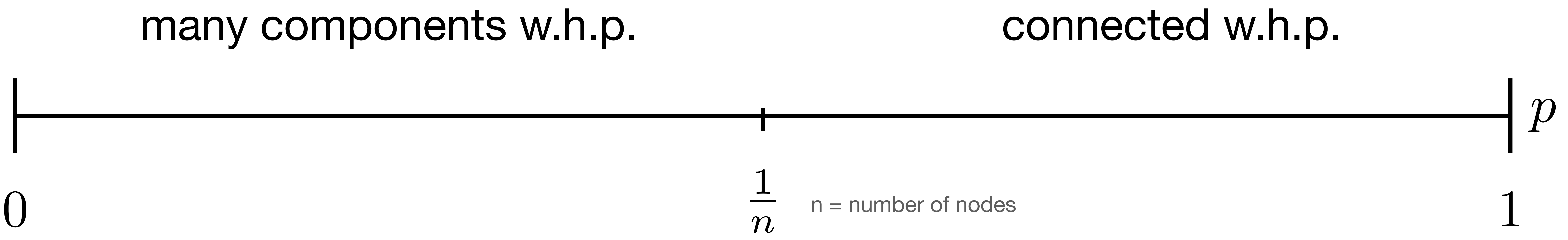
Erdos-Renyi graphs



Erdos-Renyi graphs

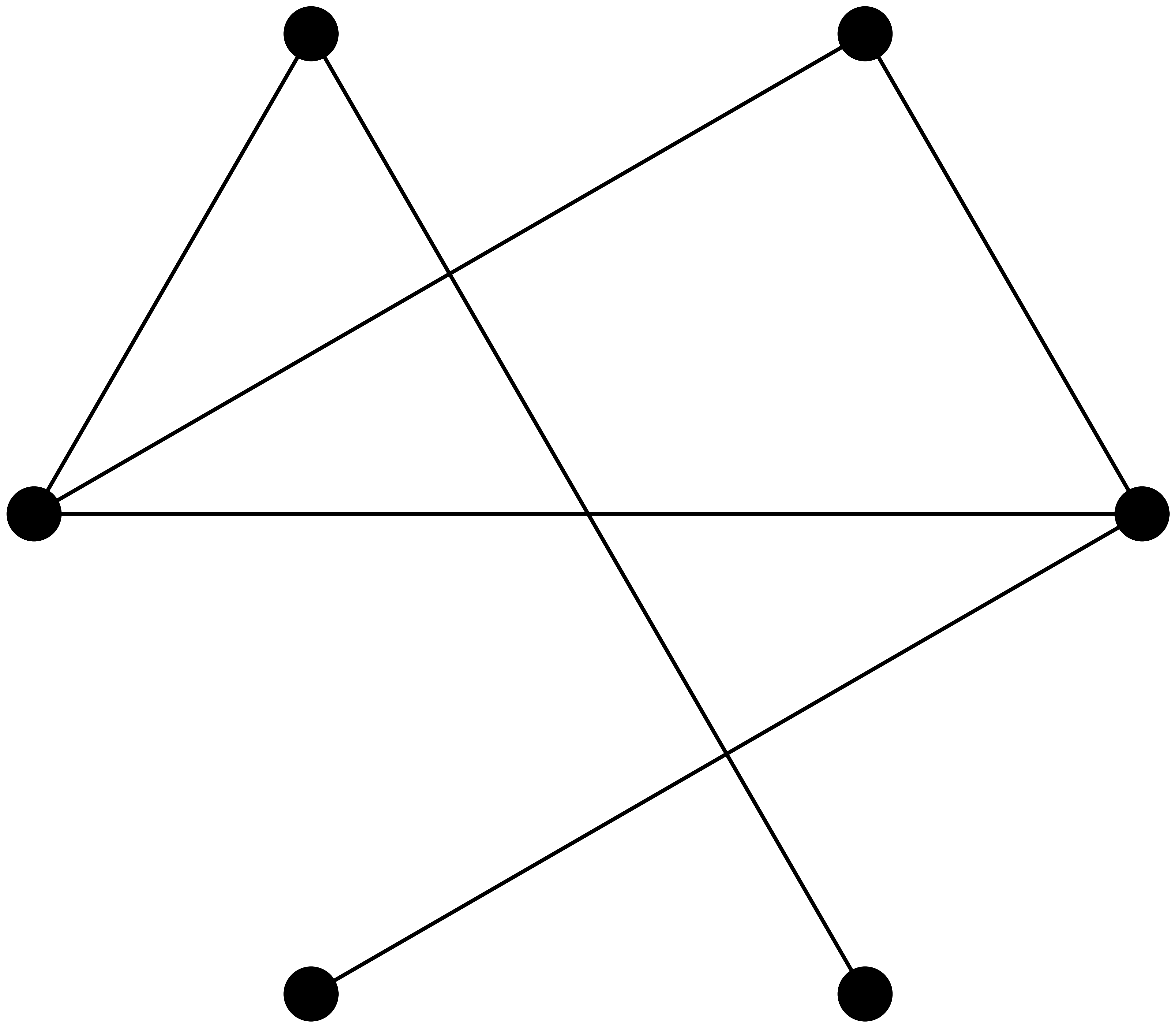


Phase Transition
[Erdos-Renyi 1960]

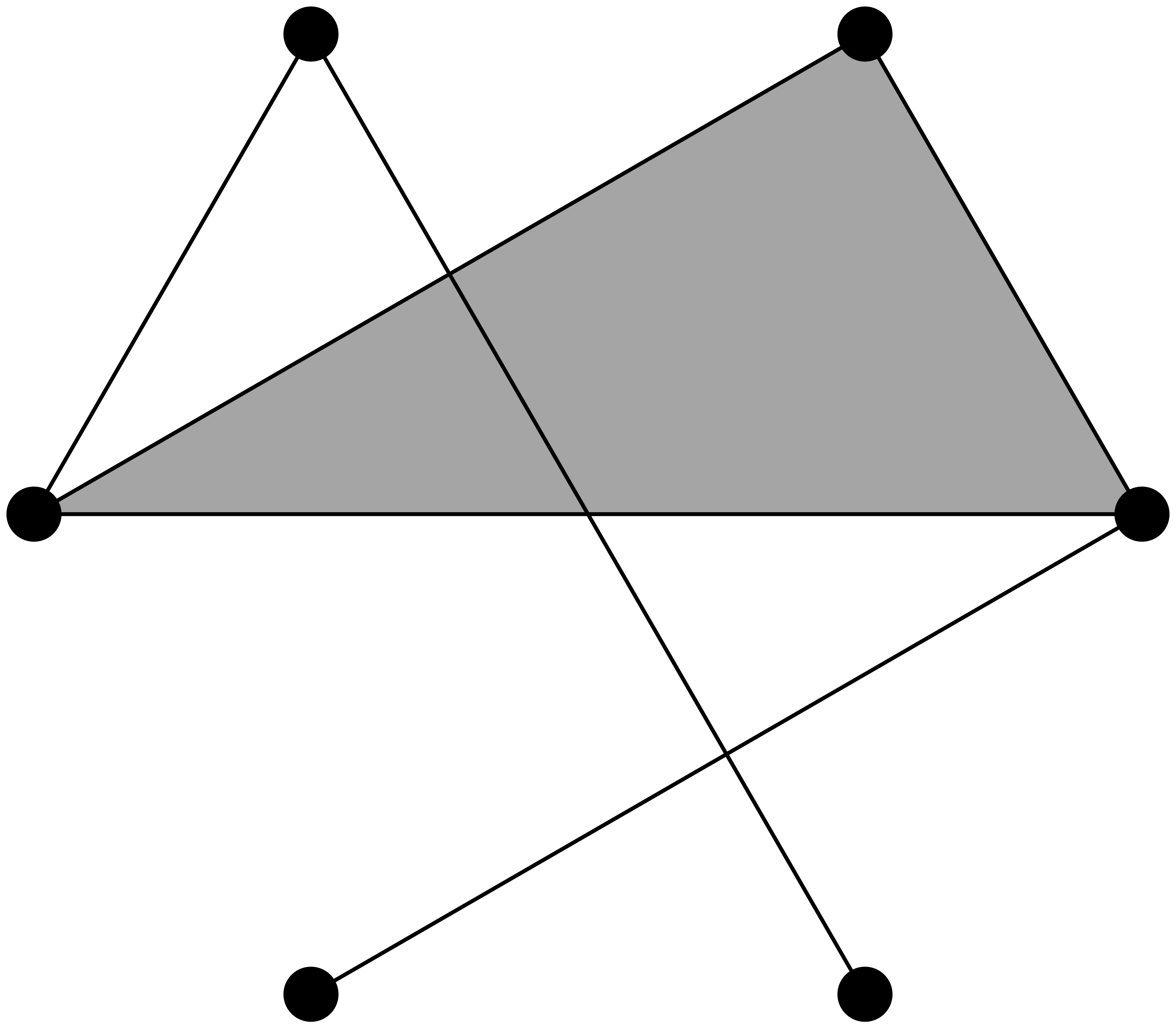


all log terms and constants forgone

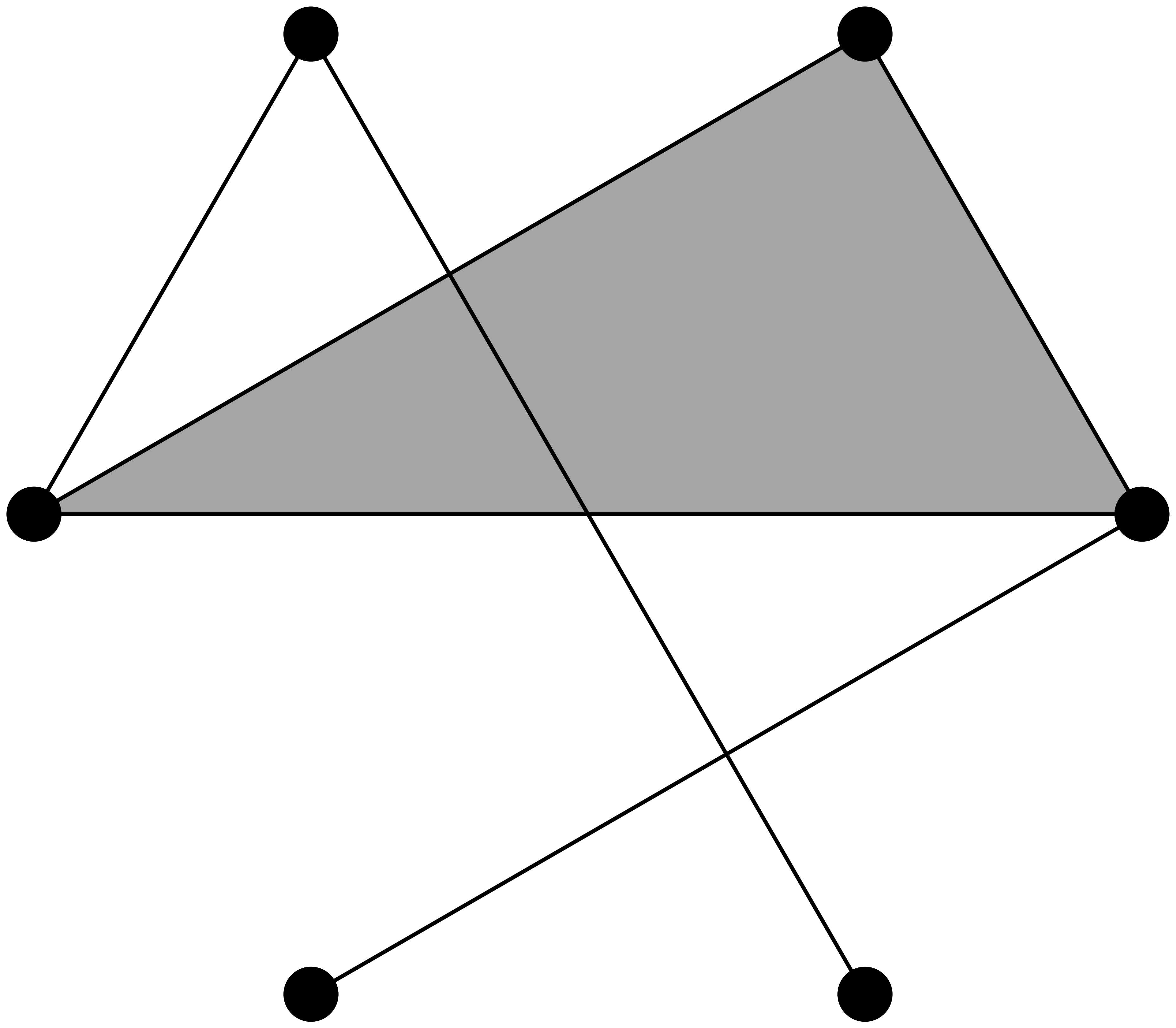
Erdos-Renyi Clique Complex

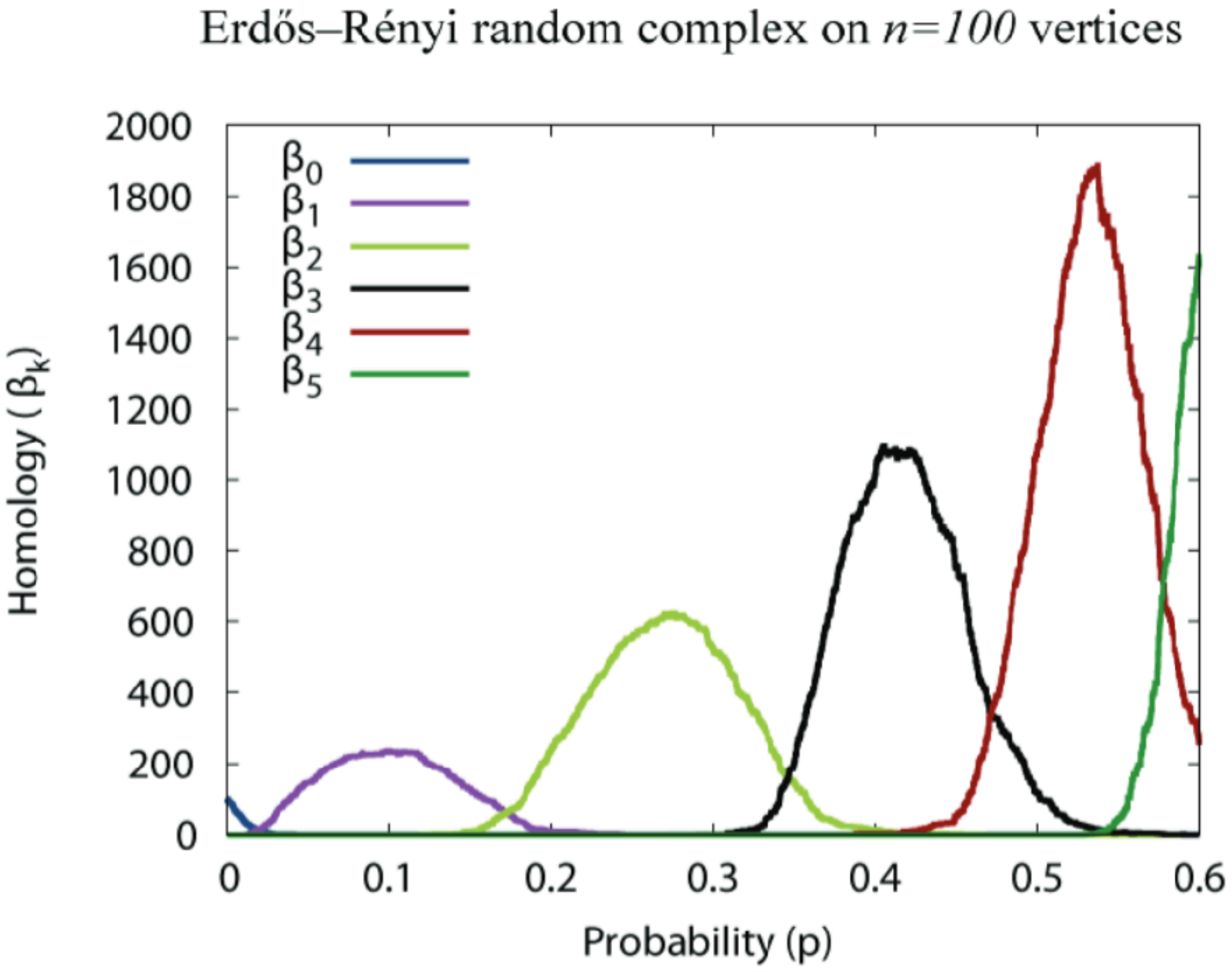


Erdos-Renyi Clique Complex



Betti Numbers

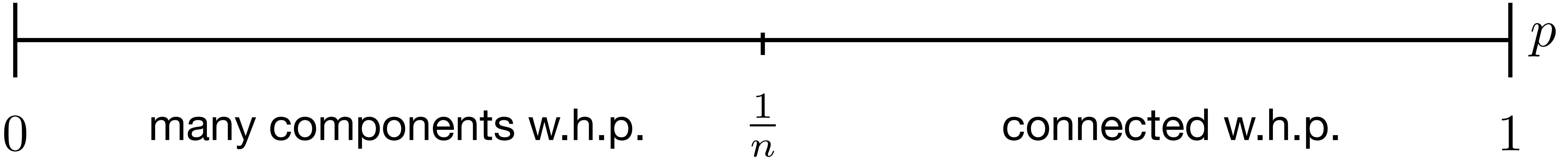




computation and plotting done by Zomorodian

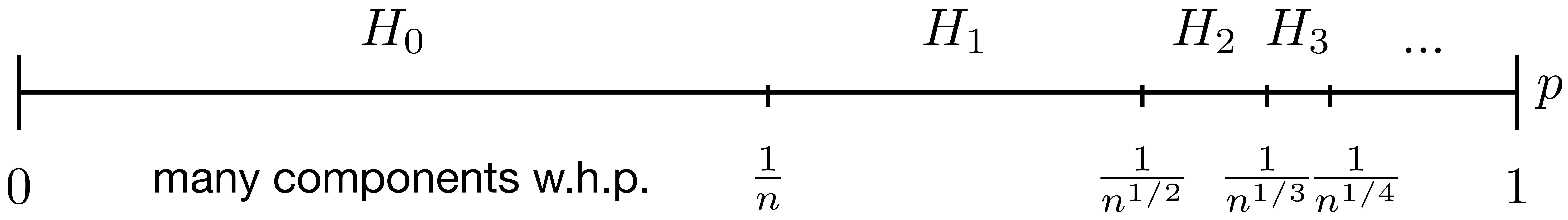
Phase Transition

[Erdos-Renyi 1960]



n = number of nodes
all log terms and constants forgone

Phase Transition
[Kahle 2009, 2014]

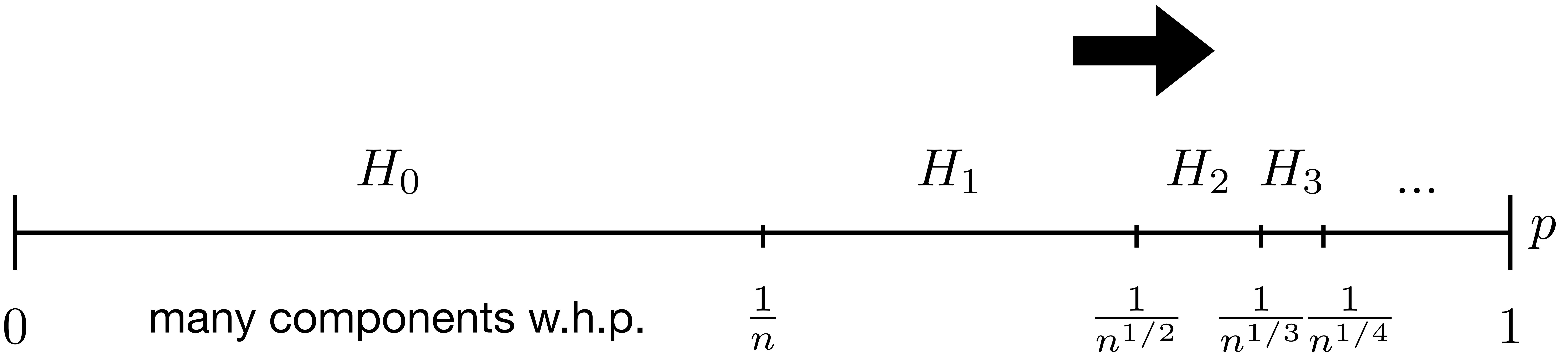


n = number of nodes
all log terms and constants forgone

Phase Transition

[Kahle 2009, 2014]

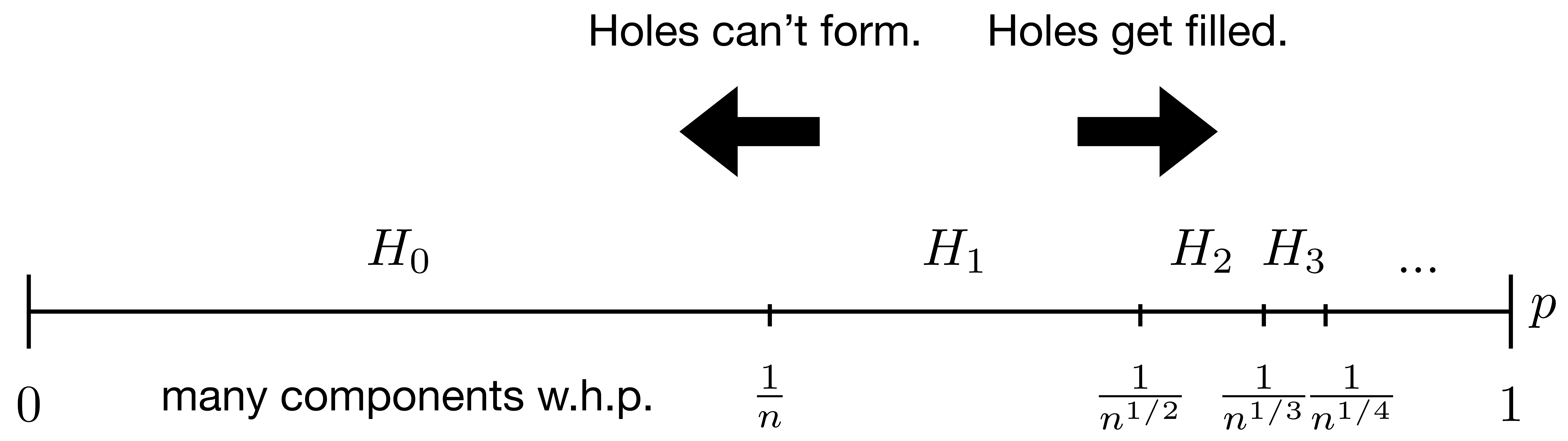
Holes get filled.



n = number of nodes
all log terms and constants forgone

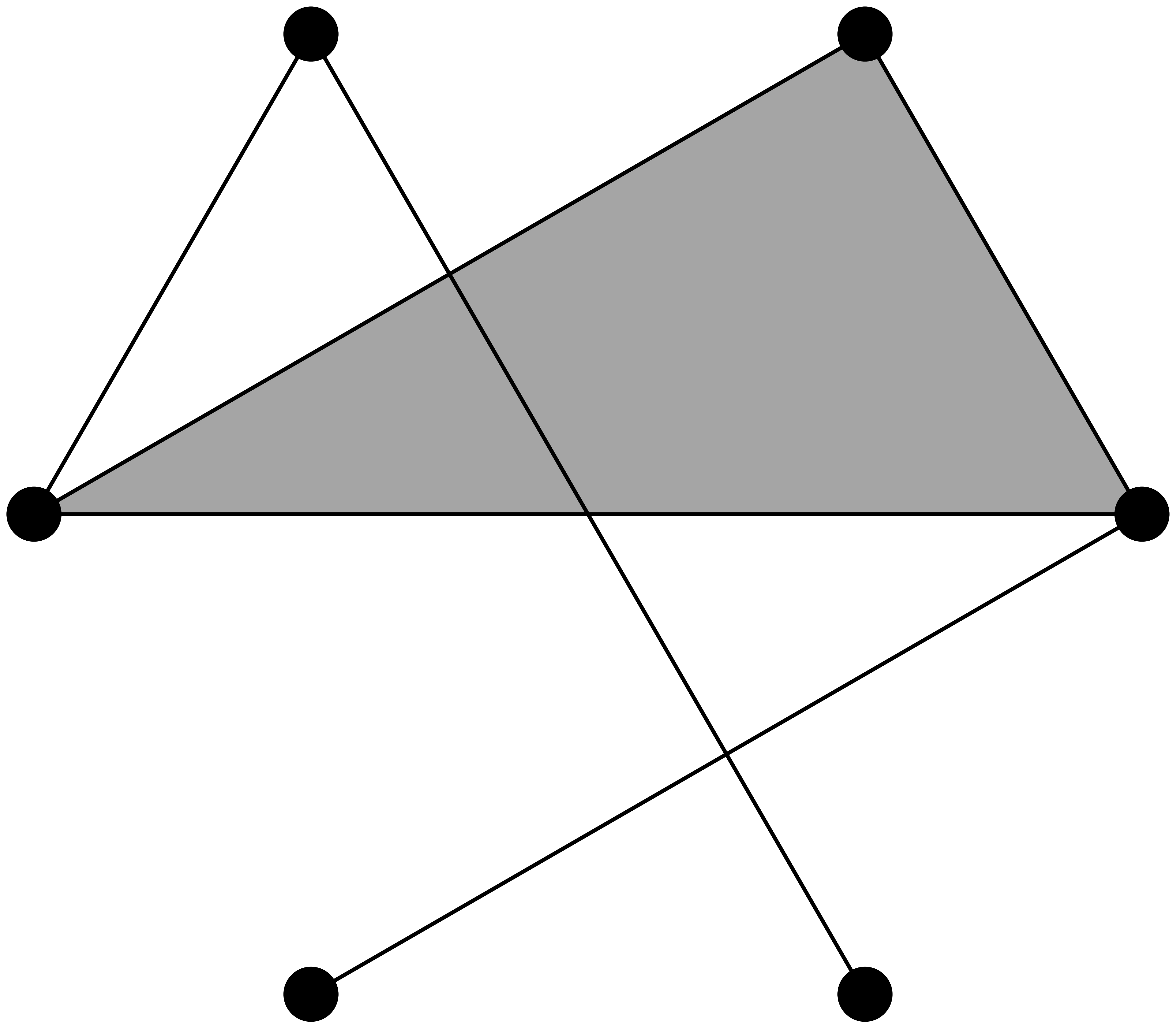
Phase Transition

[Kahle 2009, 2014]



n = number of nodes
all log terms and constants forgone

Erdos-Renyi Clique Complex



Geometric Complexes

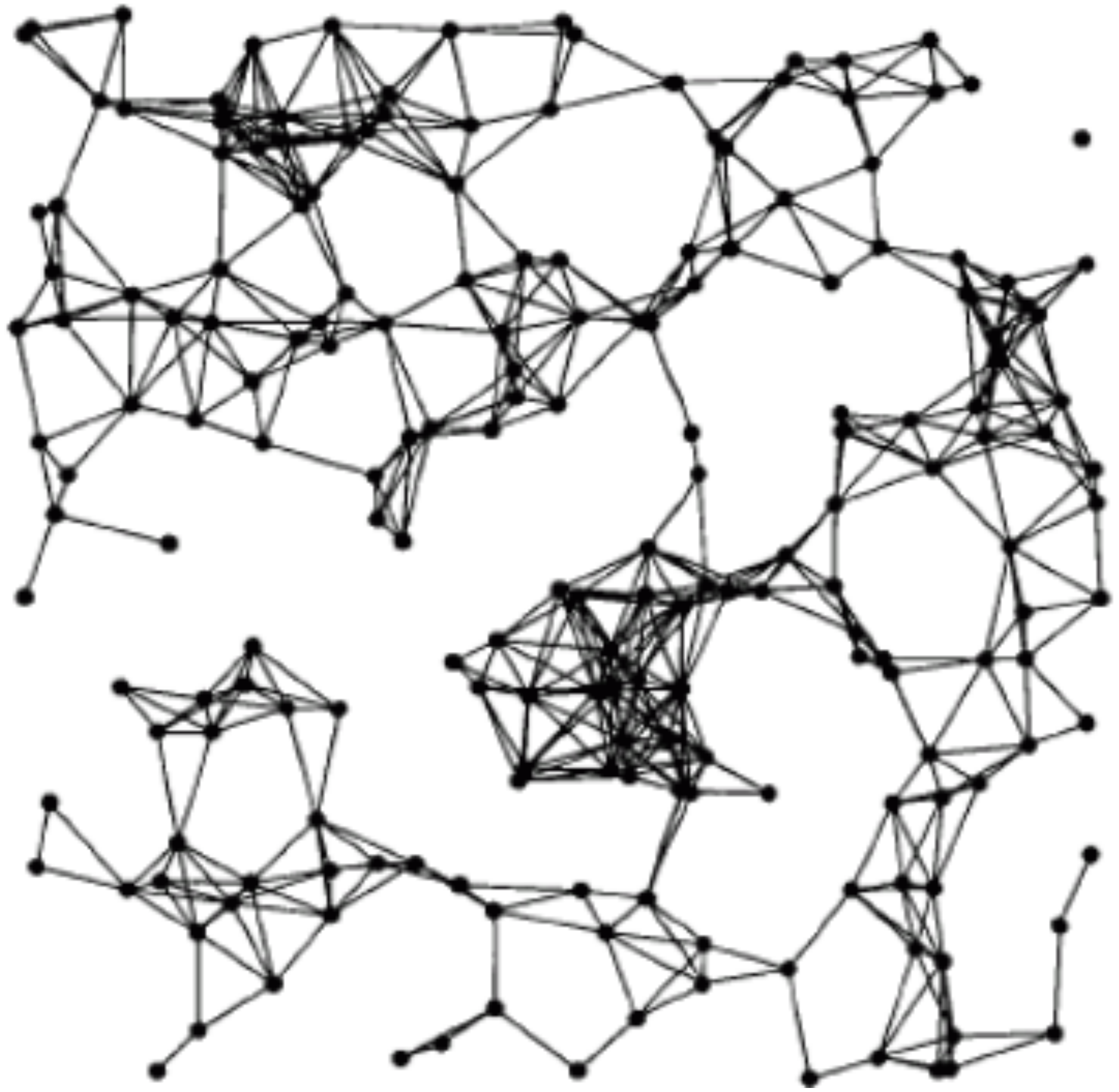


image credit: Penrose

Geometric Complexes

- Rips
- Cech

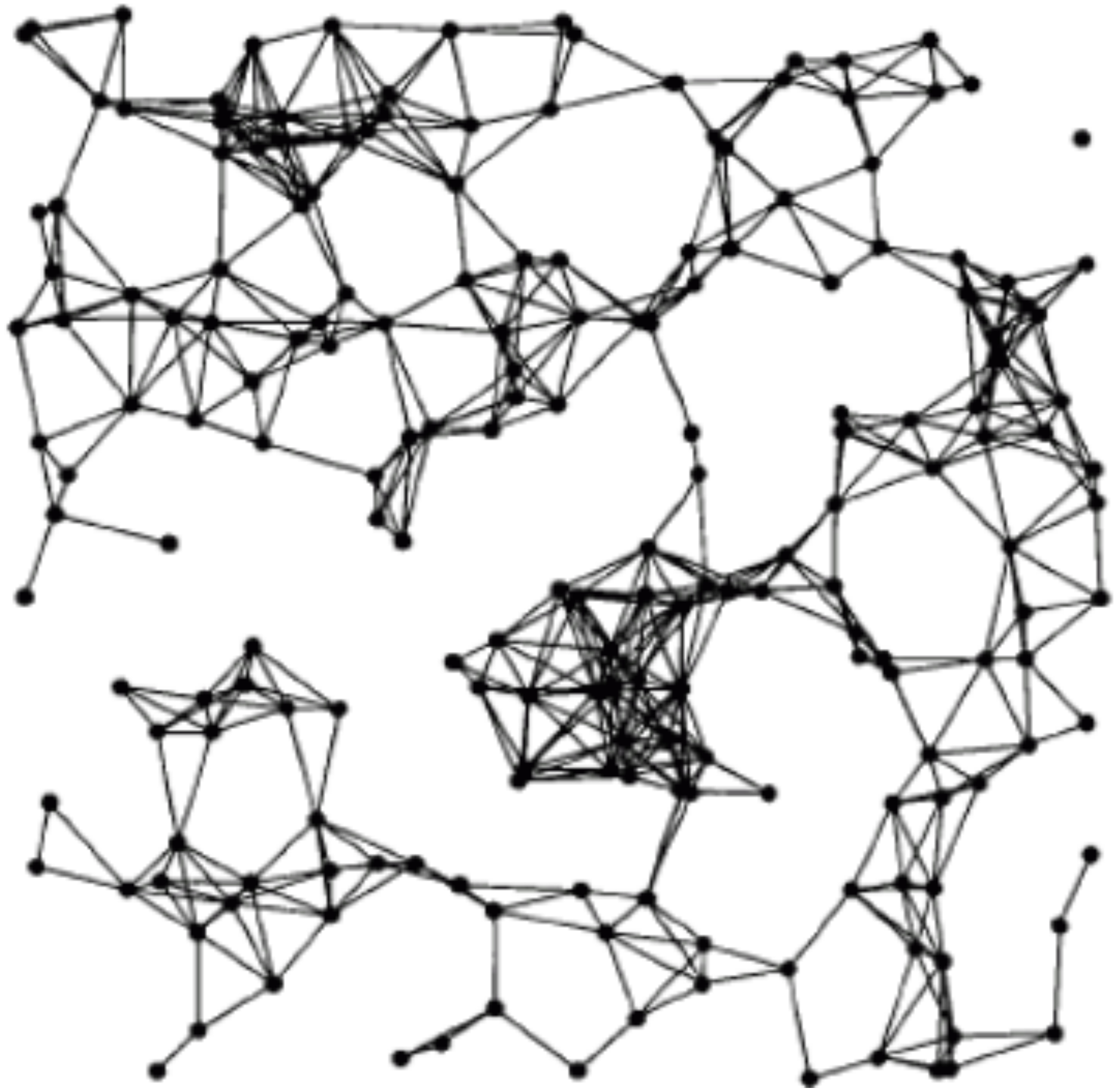


image credit: Penrose

Geometric Complexes

- Rips (clique)
- Cech

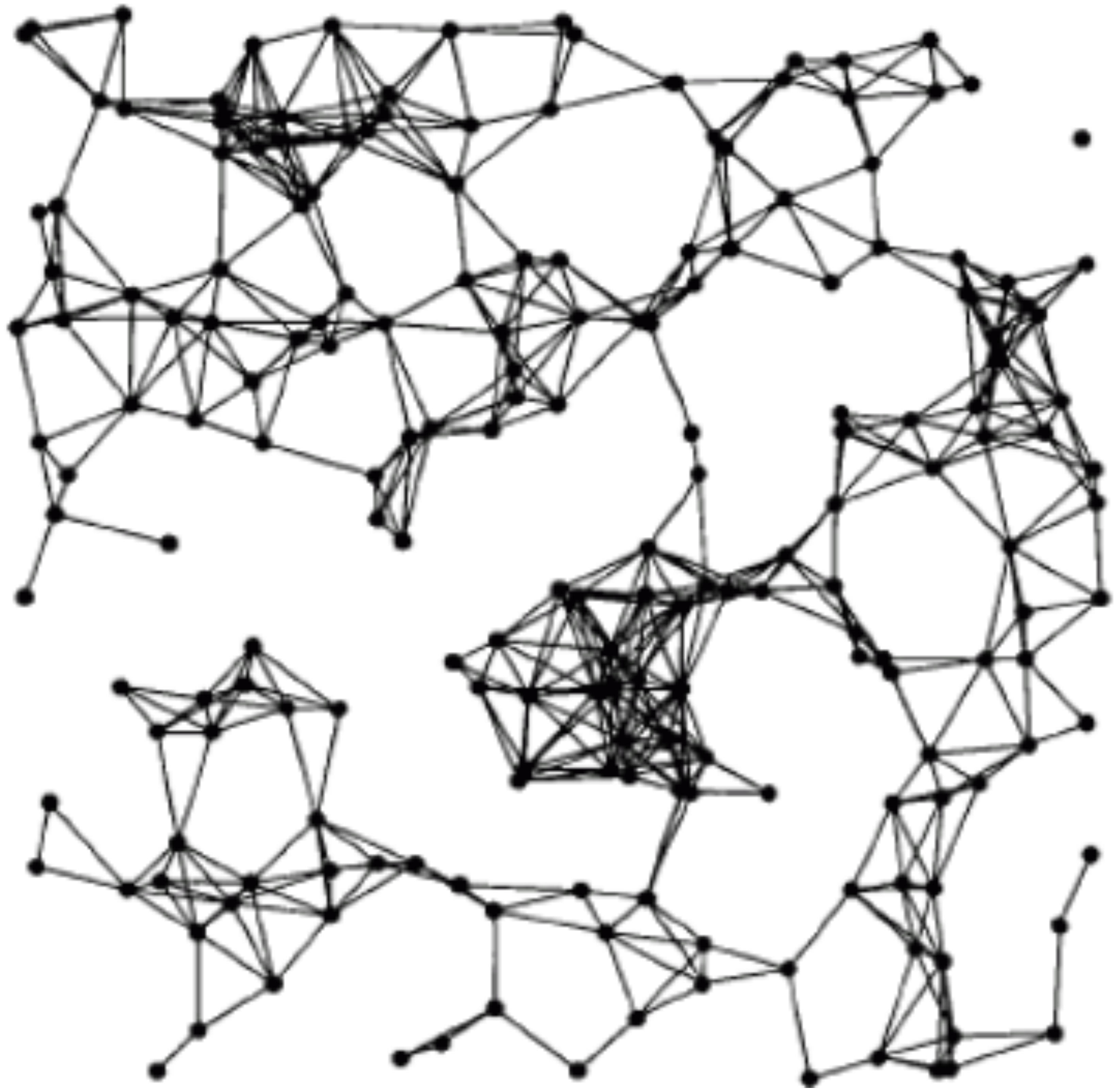
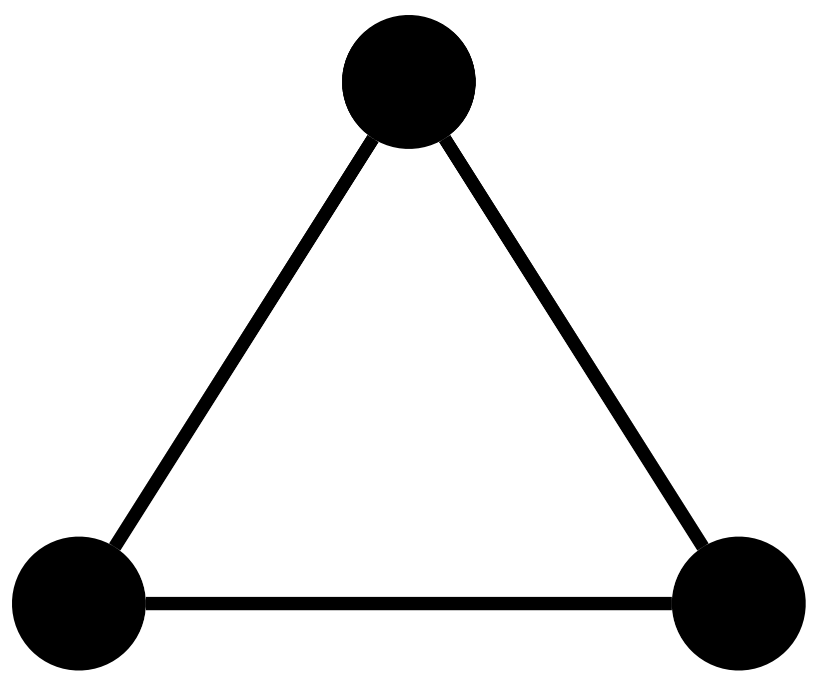


image credit: Penrose

Geometric Complexes

- Rips (clique)
- Cech



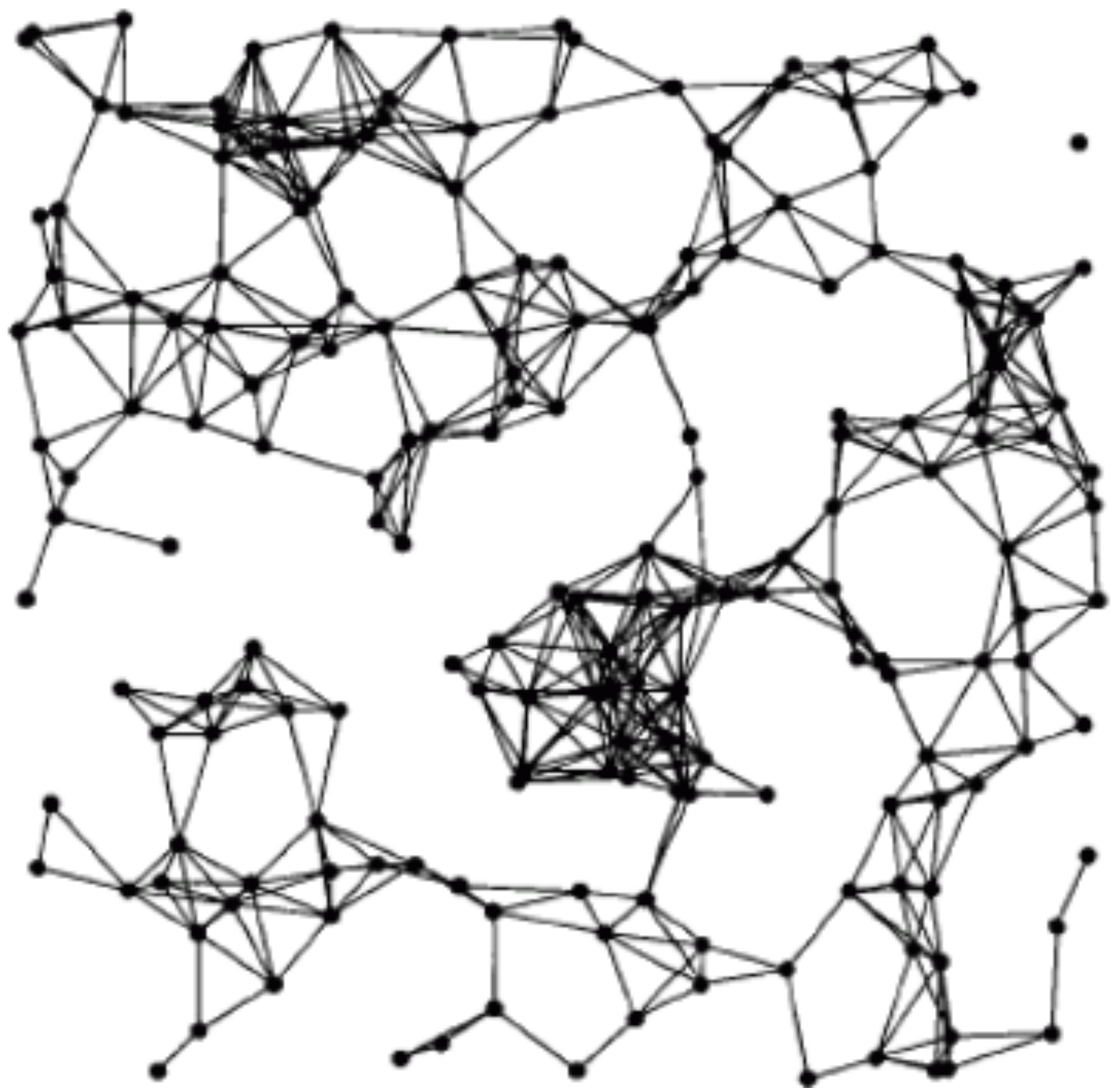
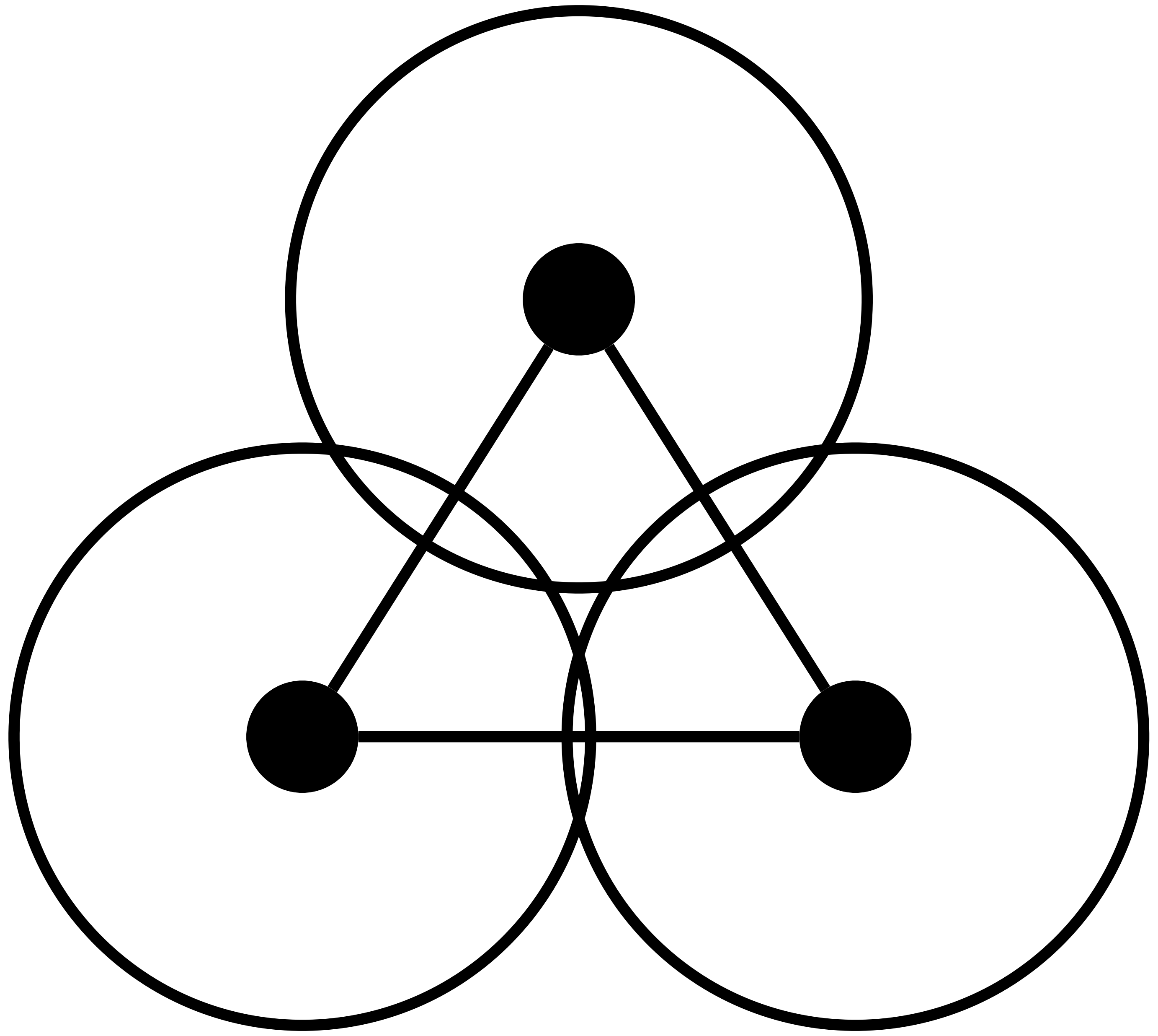


image credit: Penrose

Geometric Complexes

- Rips (clique)
- Cech



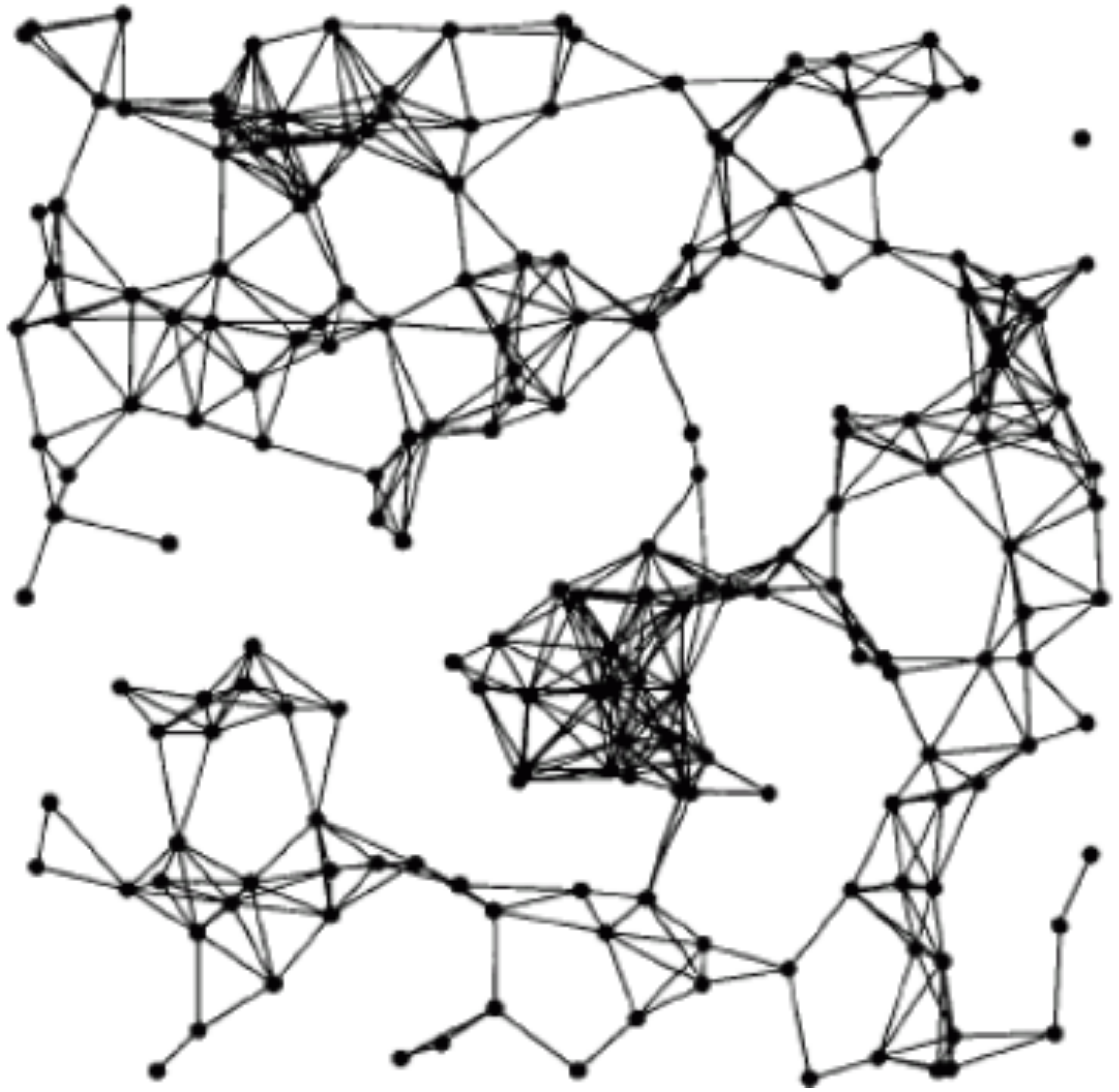


image credit: Penrose

Expected Betti numbers at dimension k

[Kahle 2011]

Expected Betti numbers at dimension k

[Kahle 2011]

- n, the number of points

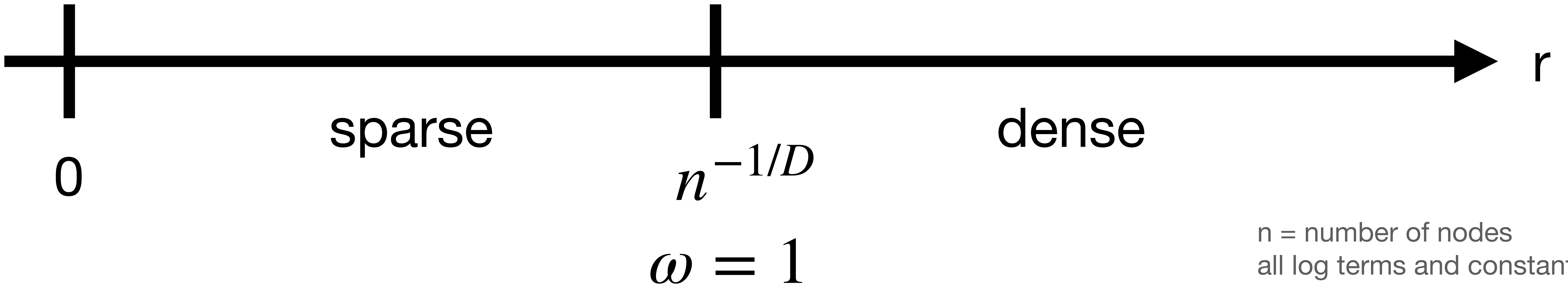
Expected Betti numbers at dimension k

[Kahle 2011]

- n, the number of points
- $\omega = nr^D$, where D is the ambient dimension

Expected Betti numbers at dimension k
[Kahle 2011]

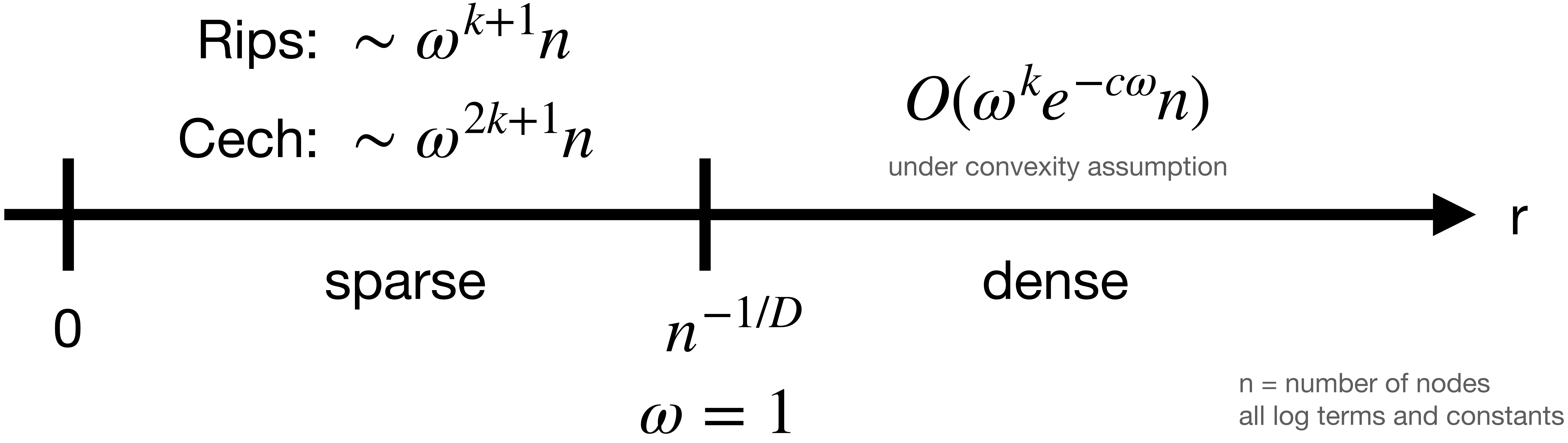
- n, the number of points
- $\omega = nr^D$, where D is the ambient dimension



n = number of nodes
all log terms and constants forgone

Expected Betti numbers at dimension k
[Kahle 2011]

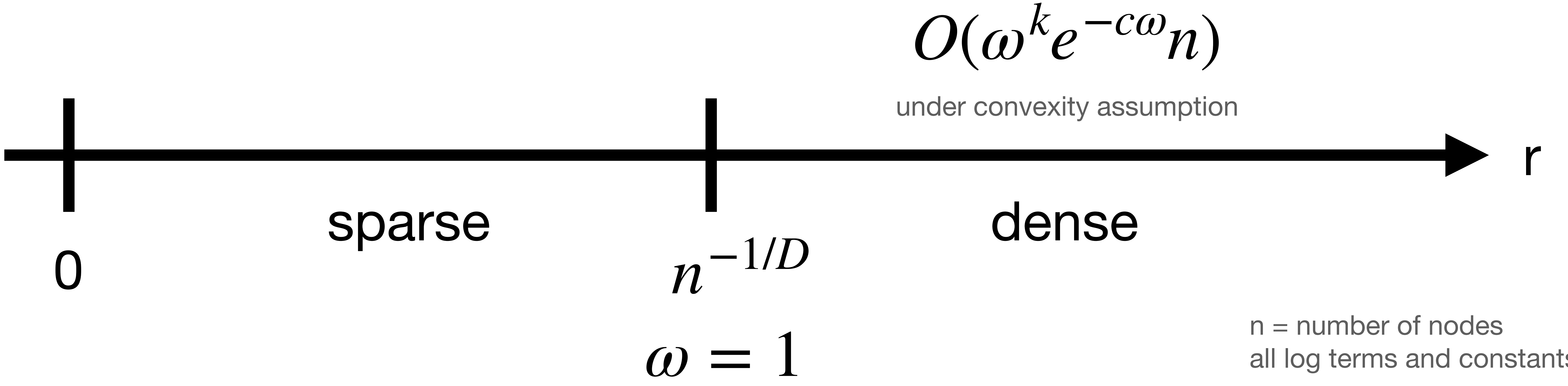
- n, the number of points
- $\omega = nr^D$, where D is the ambient dimension



n = number of nodes
all log terms and constants forgone

Expected Betti numbers at dimension k
[Kahle 2011]

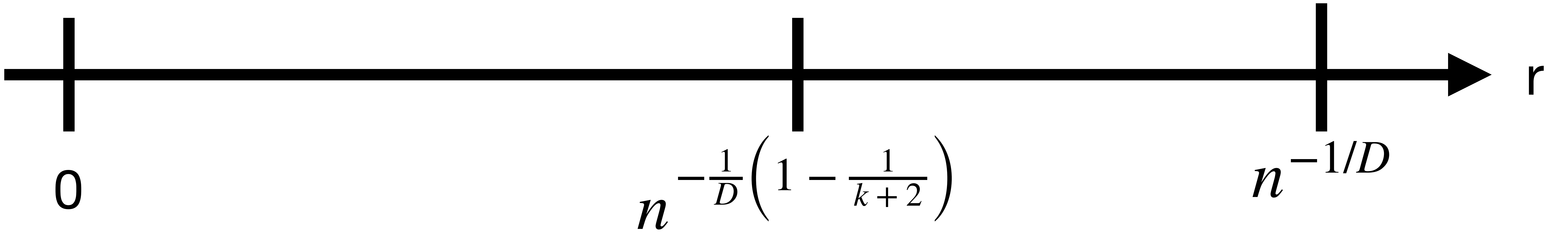
- n, the number of points
- $\omega = nr^D$, where D is the ambient dimension
- $E\beta_k(\text{Cech}) \sim \omega^{2k+1}n$



n = number of nodes
all log terms and constants forgone

Expected Betti numbers at dimension k
[Kahle 2011]

- n, the number of points
- $\omega = nr^D$, where D is the ambient dimension
- $E\beta_k(\text{Cech}) \sim \omega^{2k+1}n$

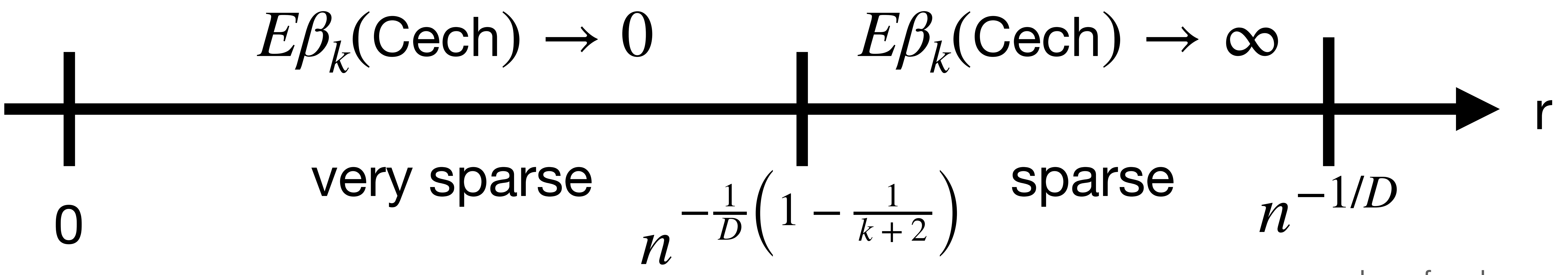


n = number of nodes
all log terms and constants forgone

Expected Betti numbers at dimension k

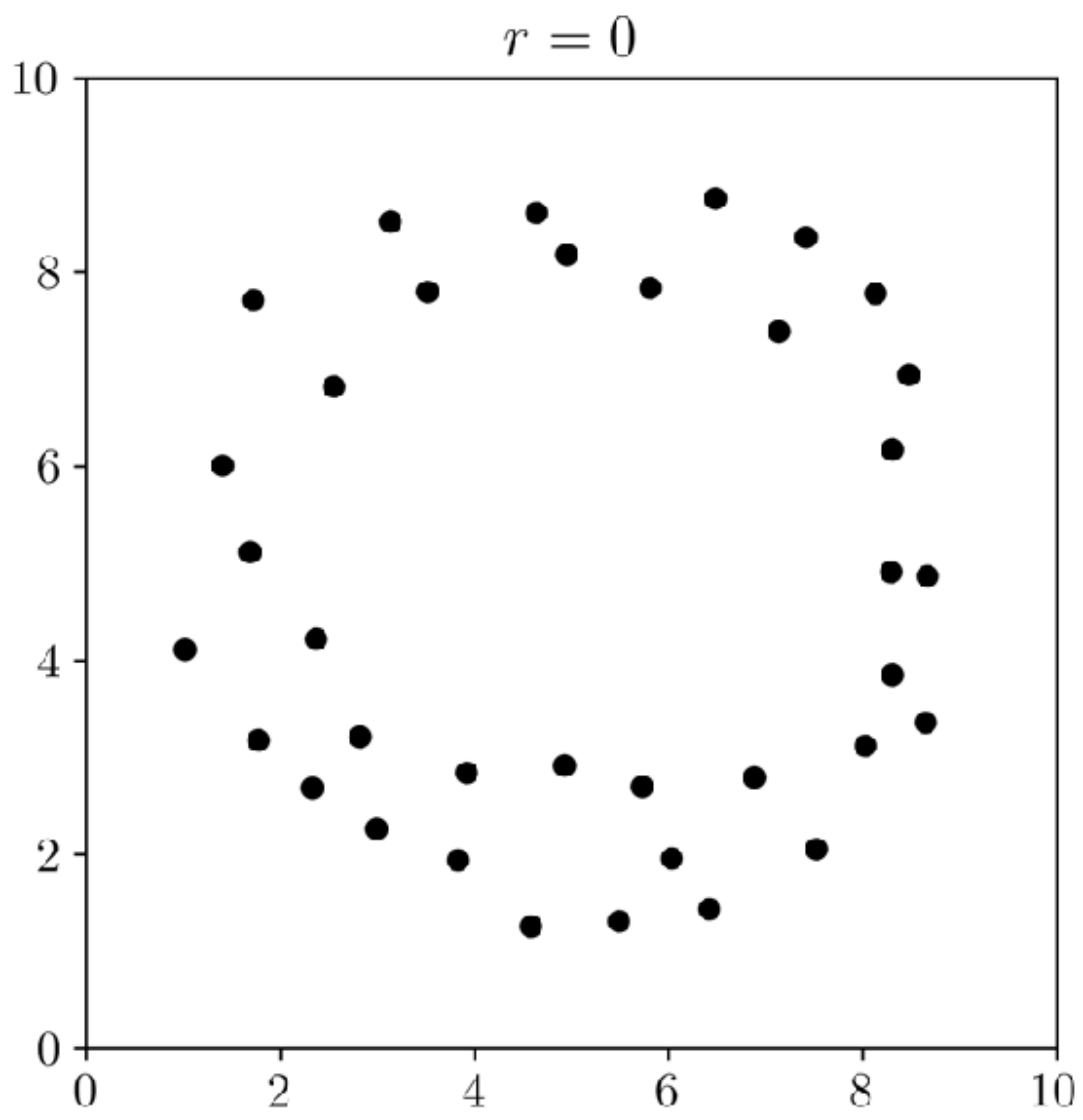
[Kahle 2011]

- n, the number of points
- $\omega = nr^D$, where D is the ambient dimension
- $E\beta_k(\text{Cech}) \sim \omega^{2k+1}n$



n = number of nodes
all log terms and constants forgone

Maximally Persistent Cycles



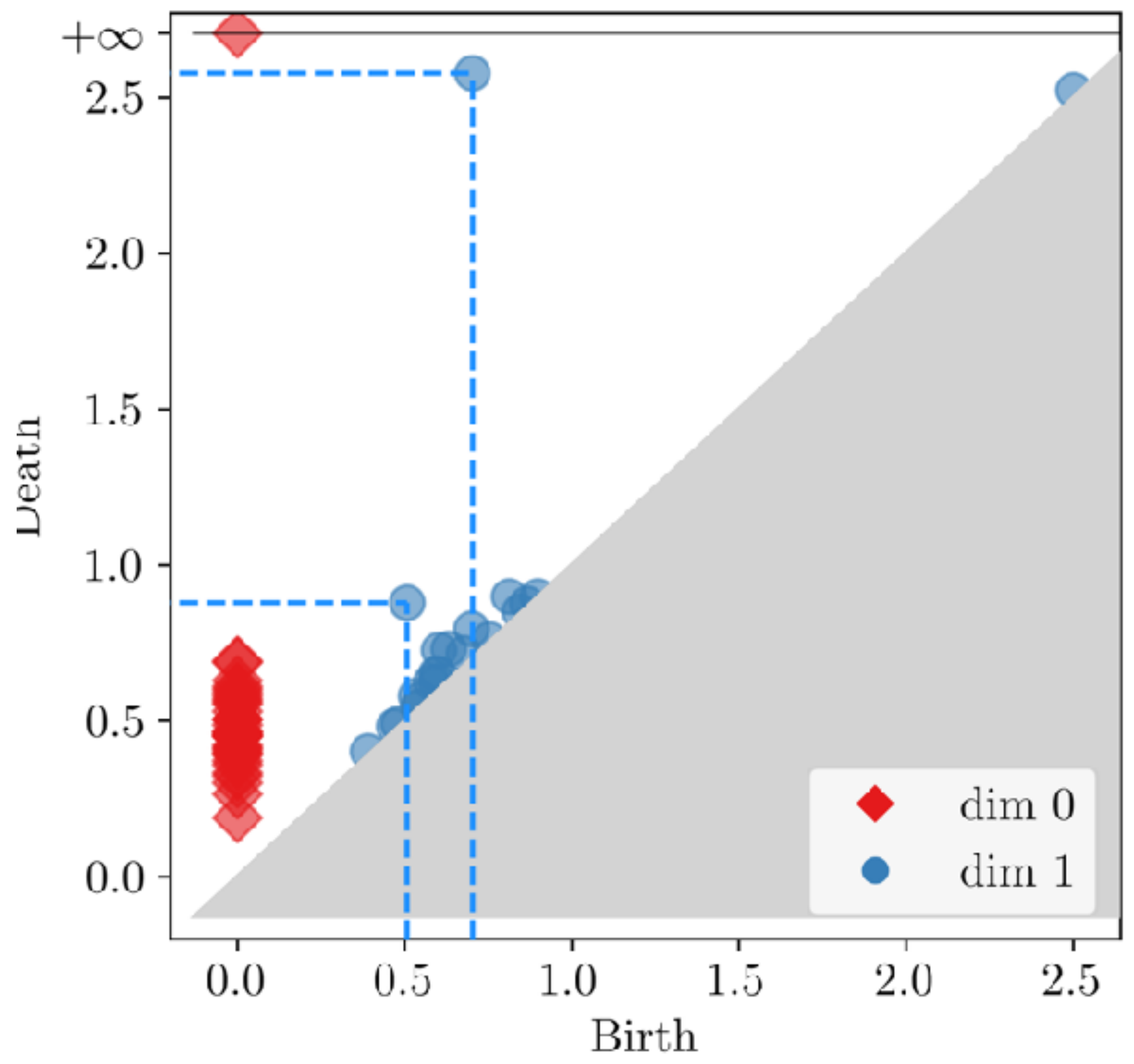


image credit: Andrey Yao

Maximally Persistent Cycles

n points in expectation

k-cycle

Maximally Persistent Cycles

[Bobrowski-Kahle-Skraba 2017]

n points in expectation

k-cycle

$$c\left(\frac{\log n}{\log\log n}\right)^{1/k} \leq \text{max persistence} \leq C\left(\frac{\log n}{\log\log n}\right)^{1/k}$$

a.a.s.

Geometric Complexes

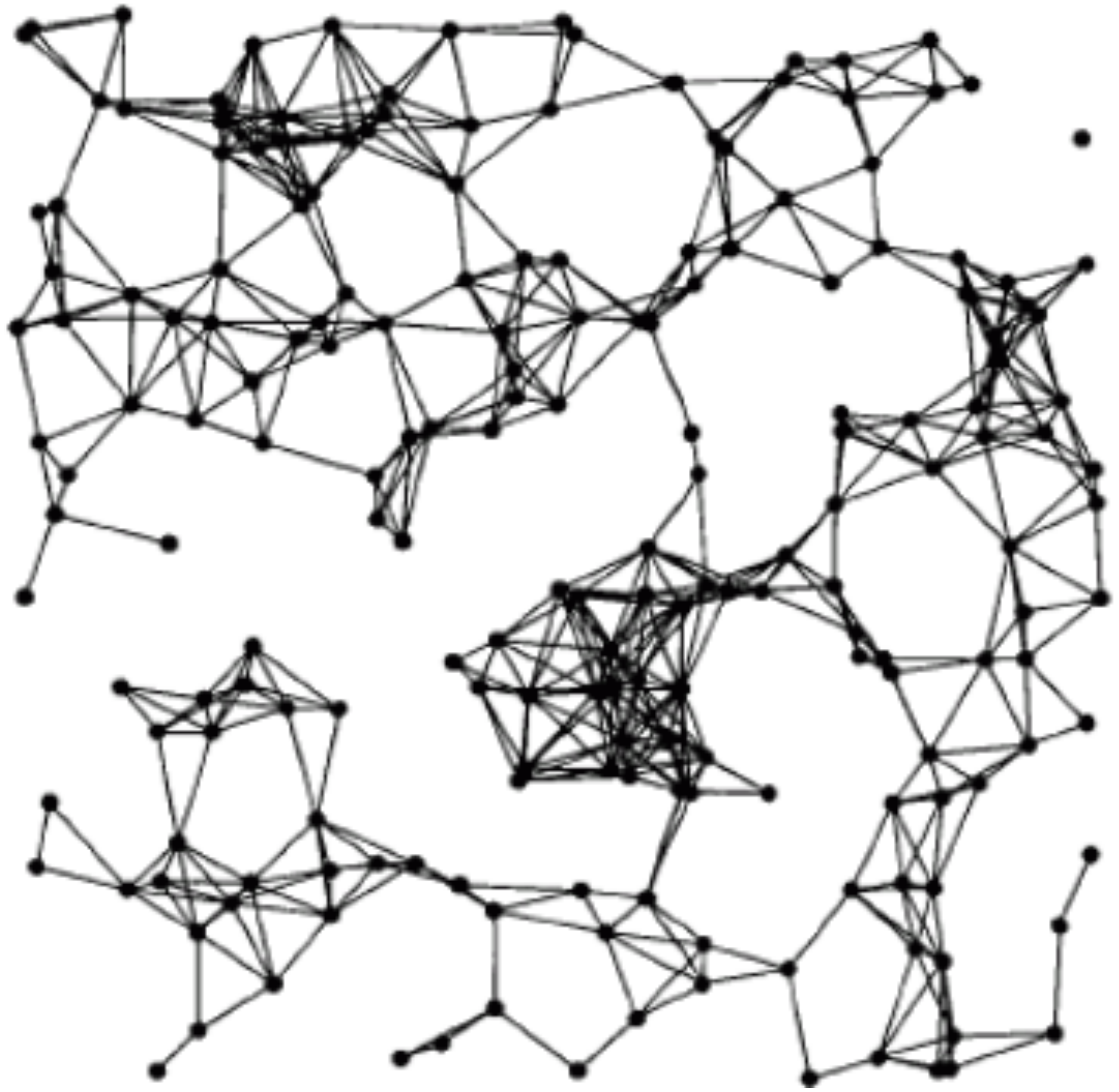
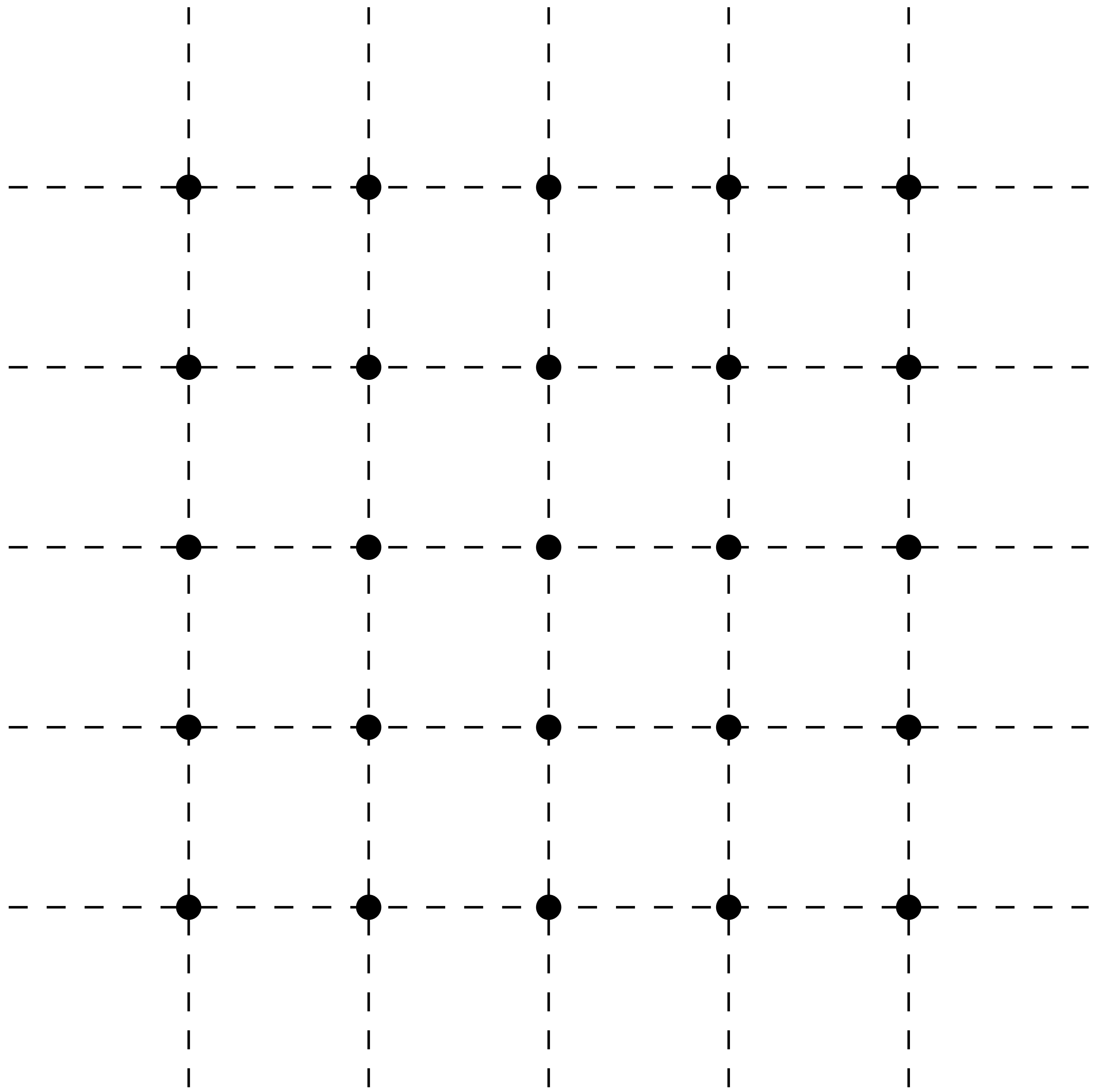
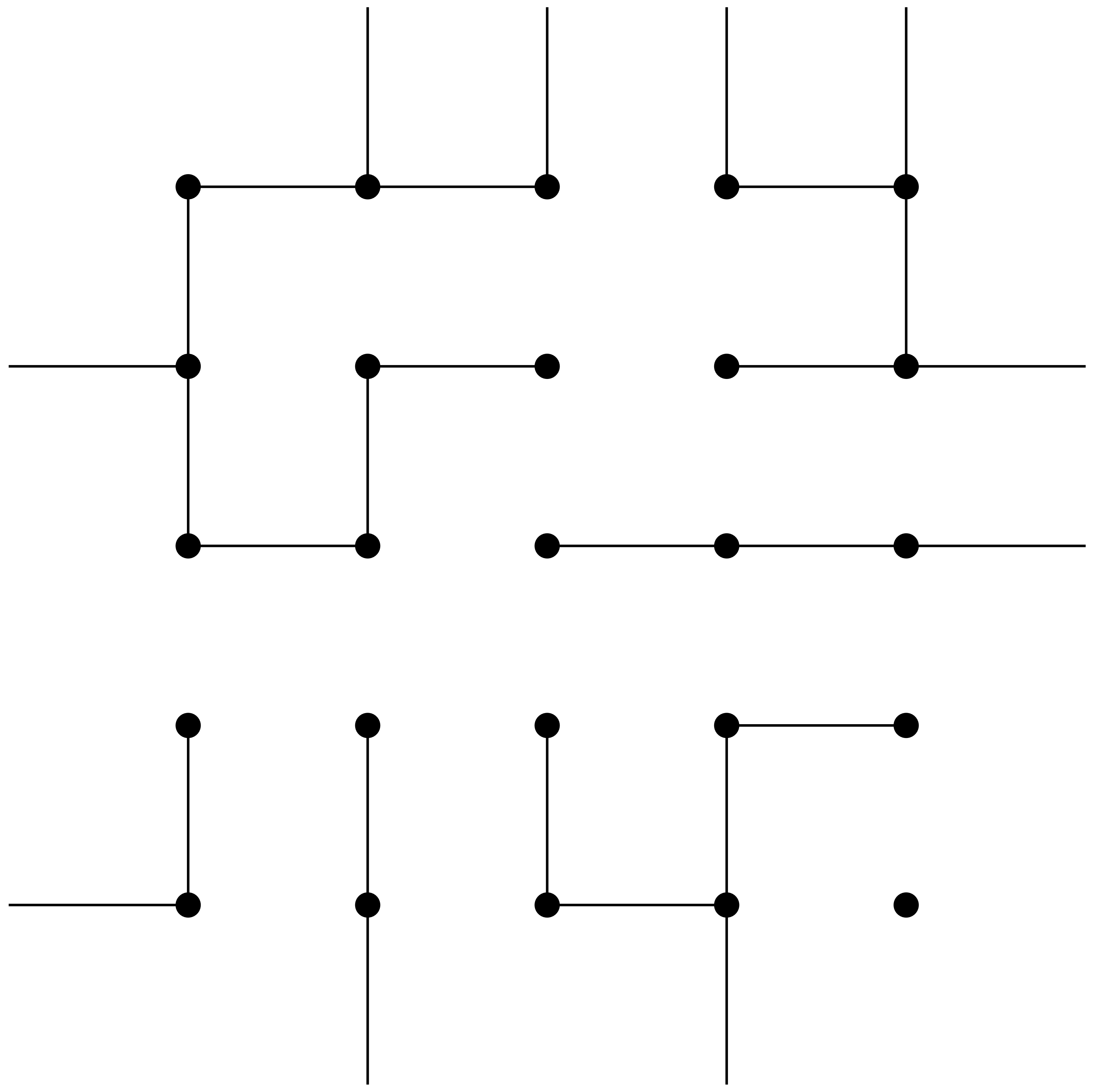


image credit: Penrose

Bernoulli Bond Percolation

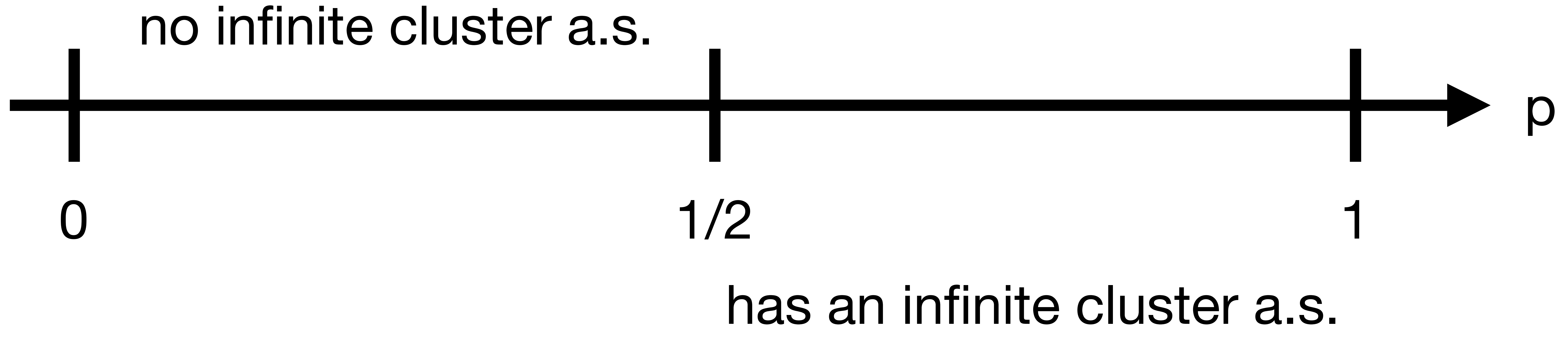


Bernoulli Bond Percolation



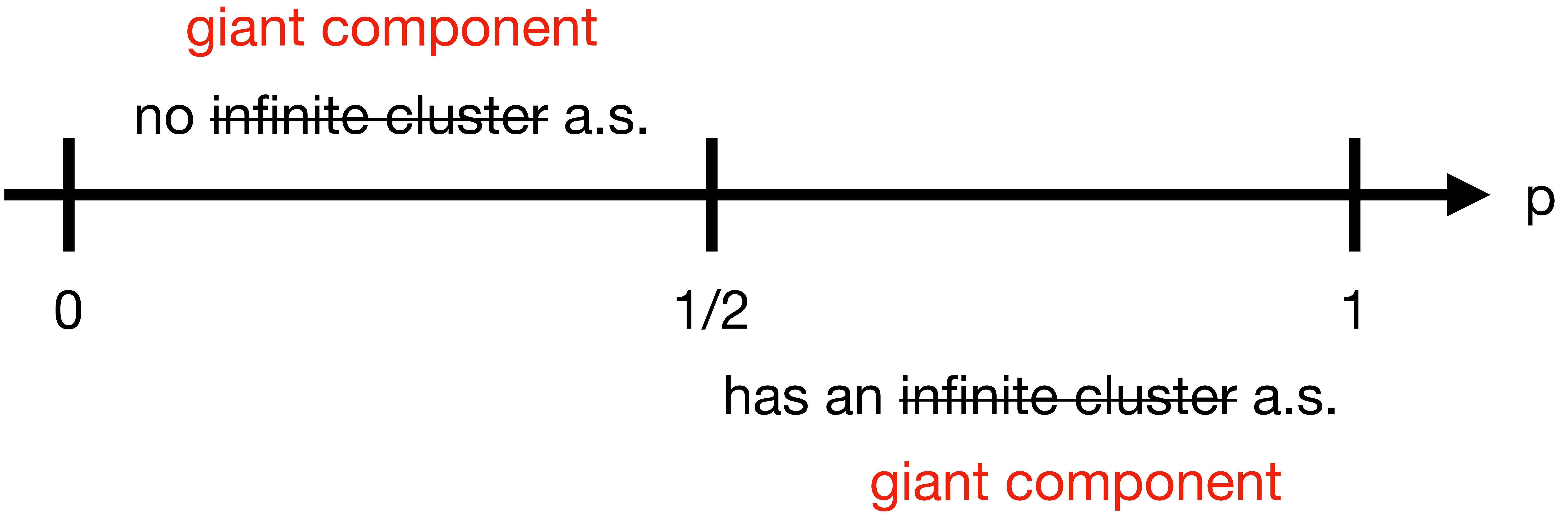
Phase Transition

[Harris 1960, Kesten 1980]



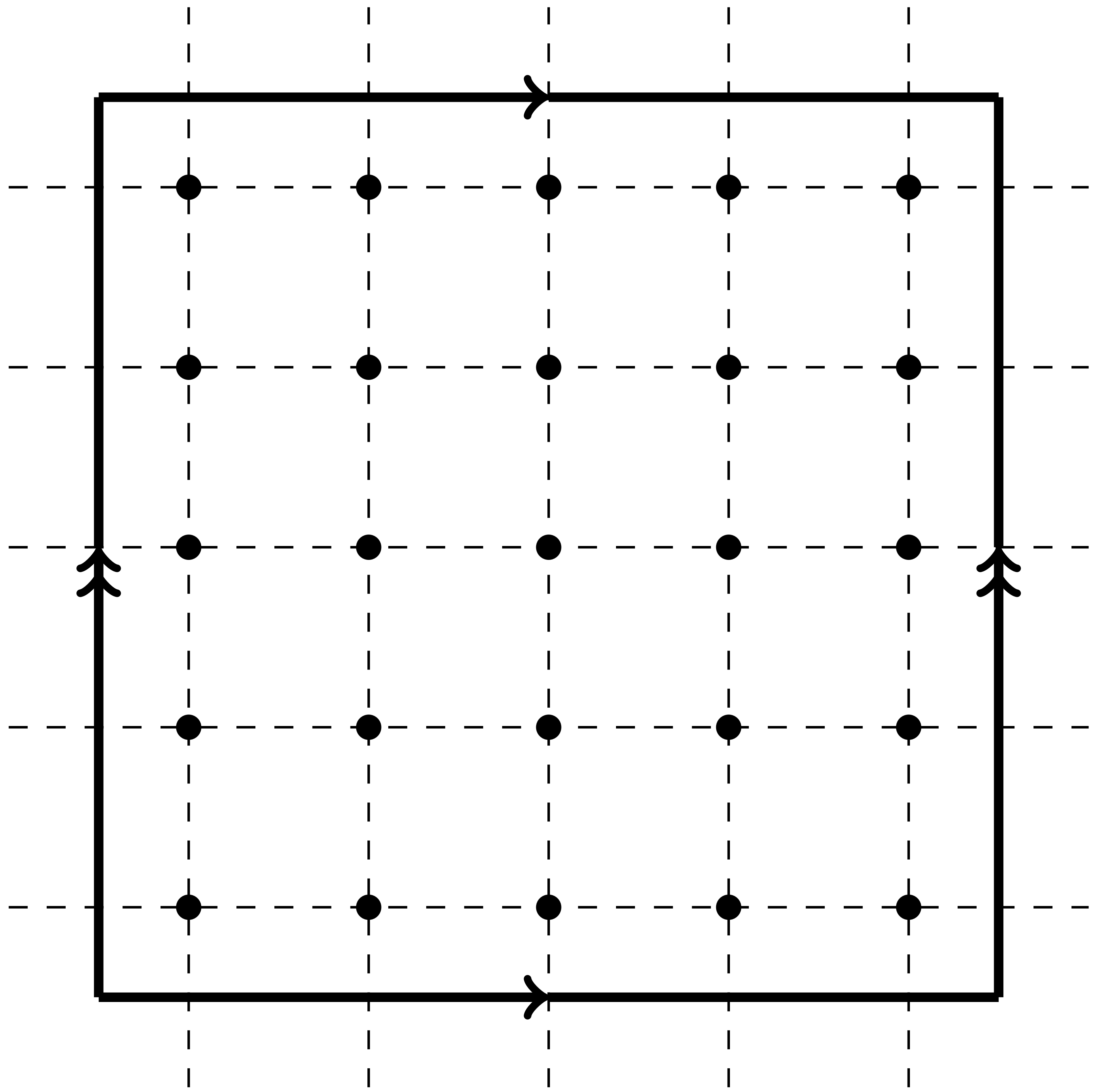
Phase Transition

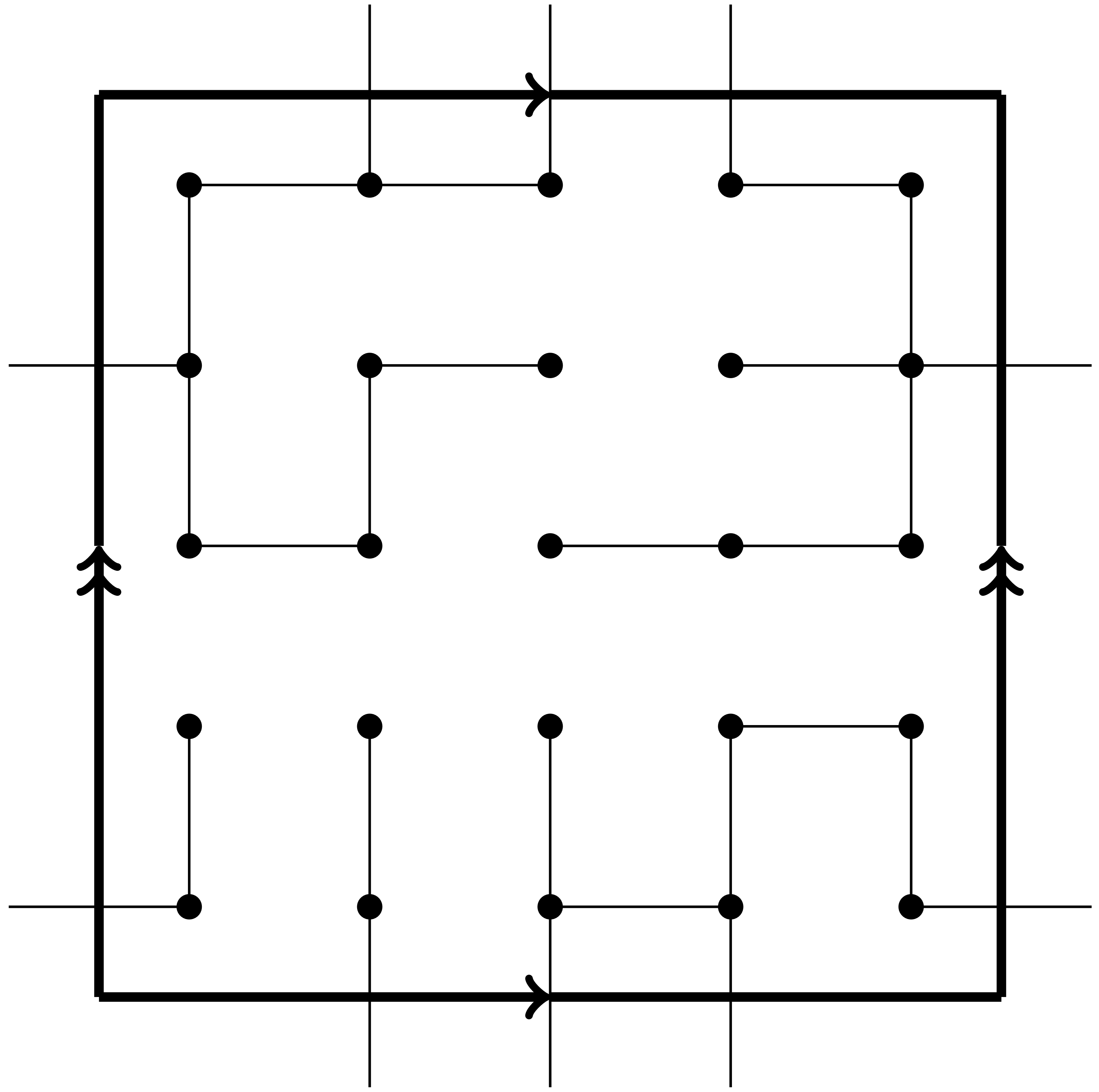
[Harris 1960, Kesten 1980]



Giant Cycles?

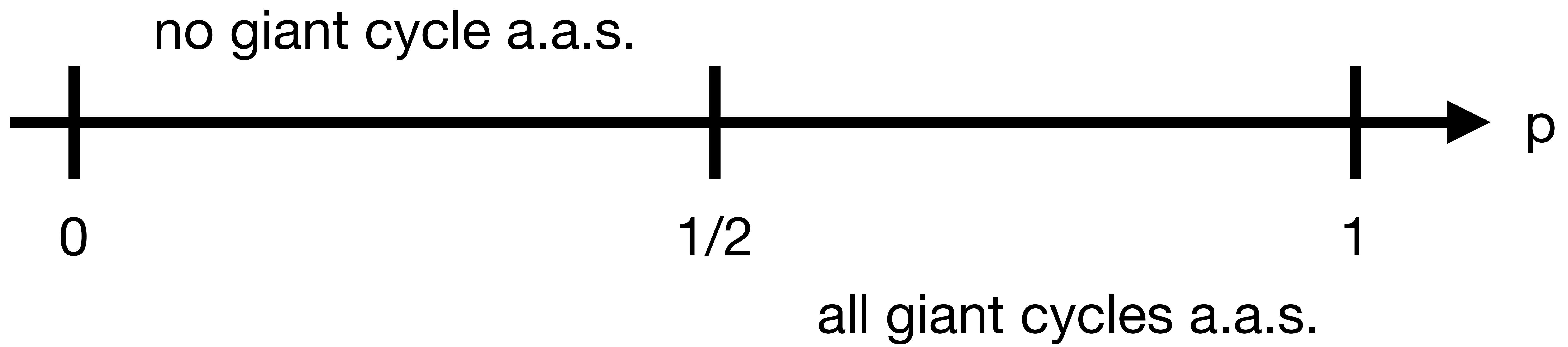
Bernoulli Bond Percolation



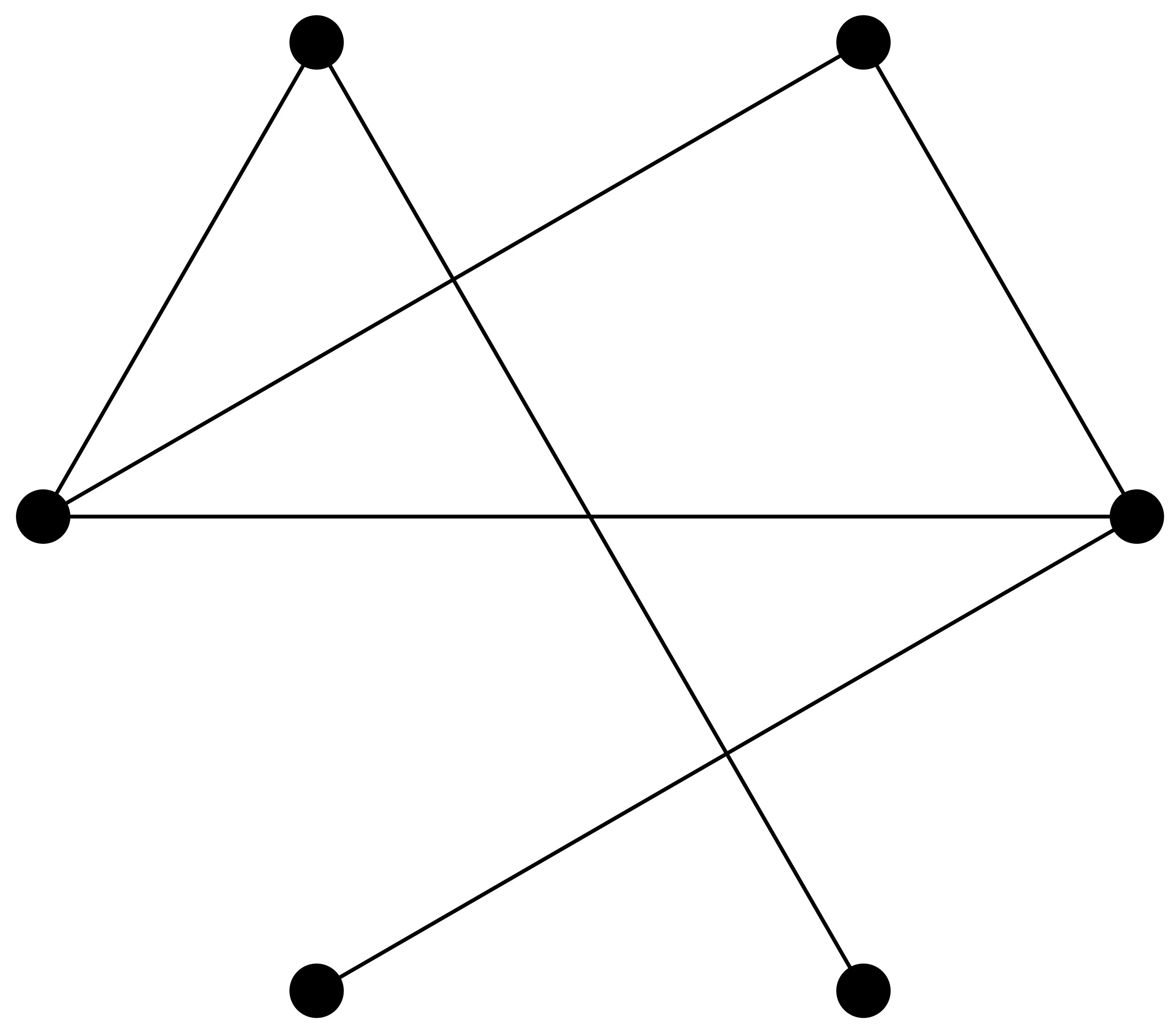


Phase Transition

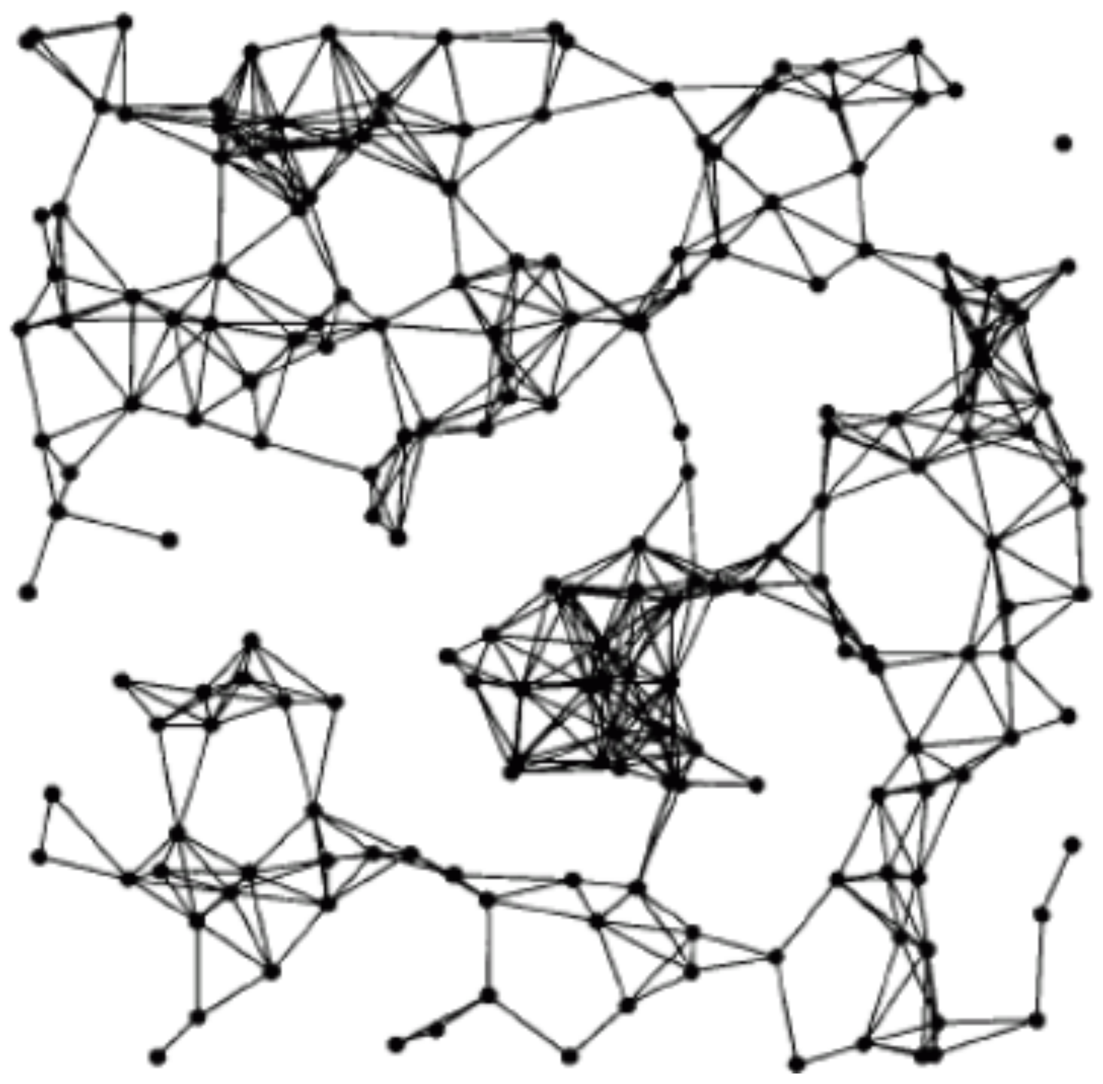
[Duncan-Kahle-Schweinhart, 2021]



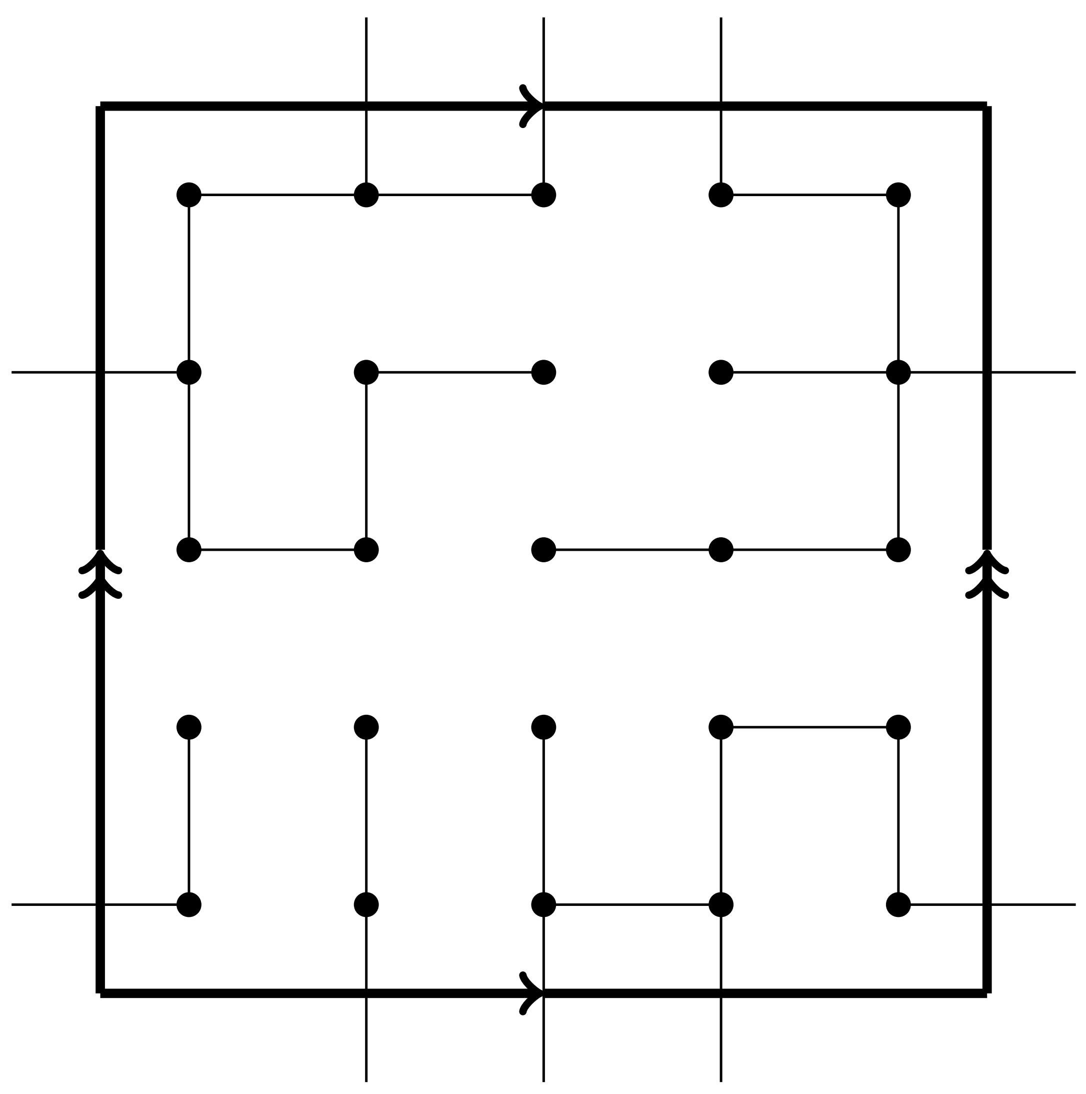
Tapas at Random Topology



Erdo-Renyi Complexes



Geometric Complexes



Topological Percolation

III. Preferential Attachment

A Non-Homogeneous Model

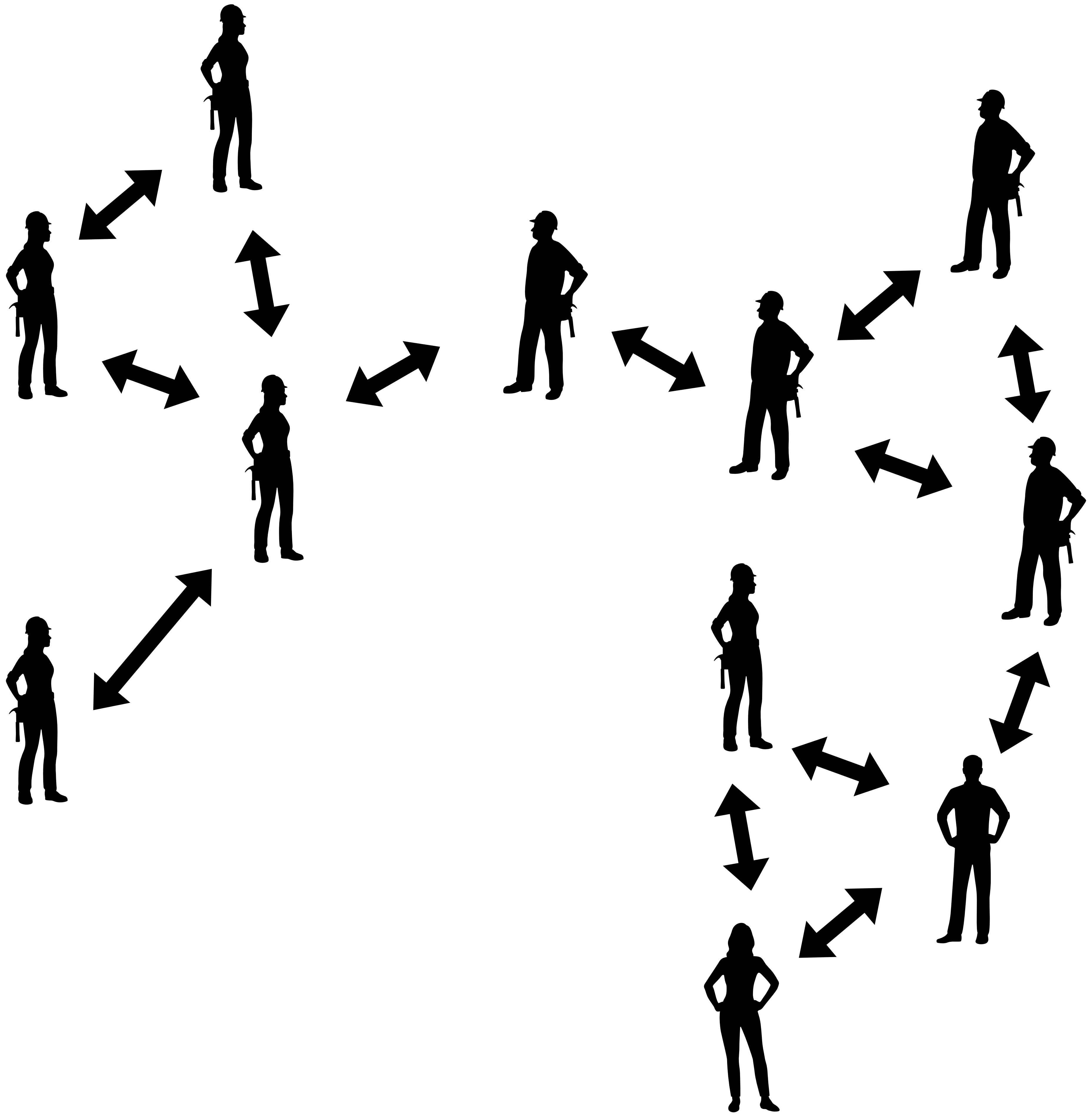
Preferential Attachment

[Albert and Barabasi 1999]

(Stephen Coast
https://www.fractalus.com/steve/stuff/ipmap/)

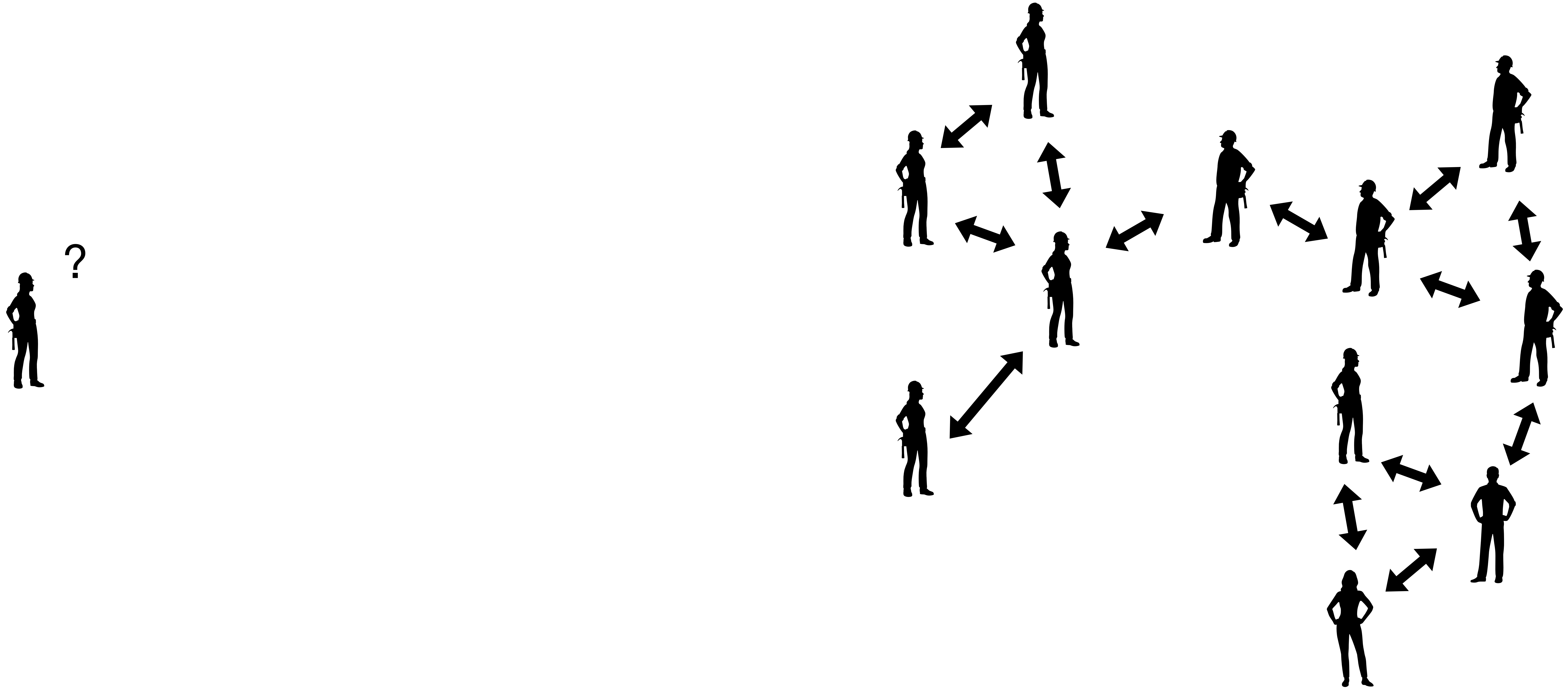
Preferential Attachment

[Albert and Barabasi 1999]



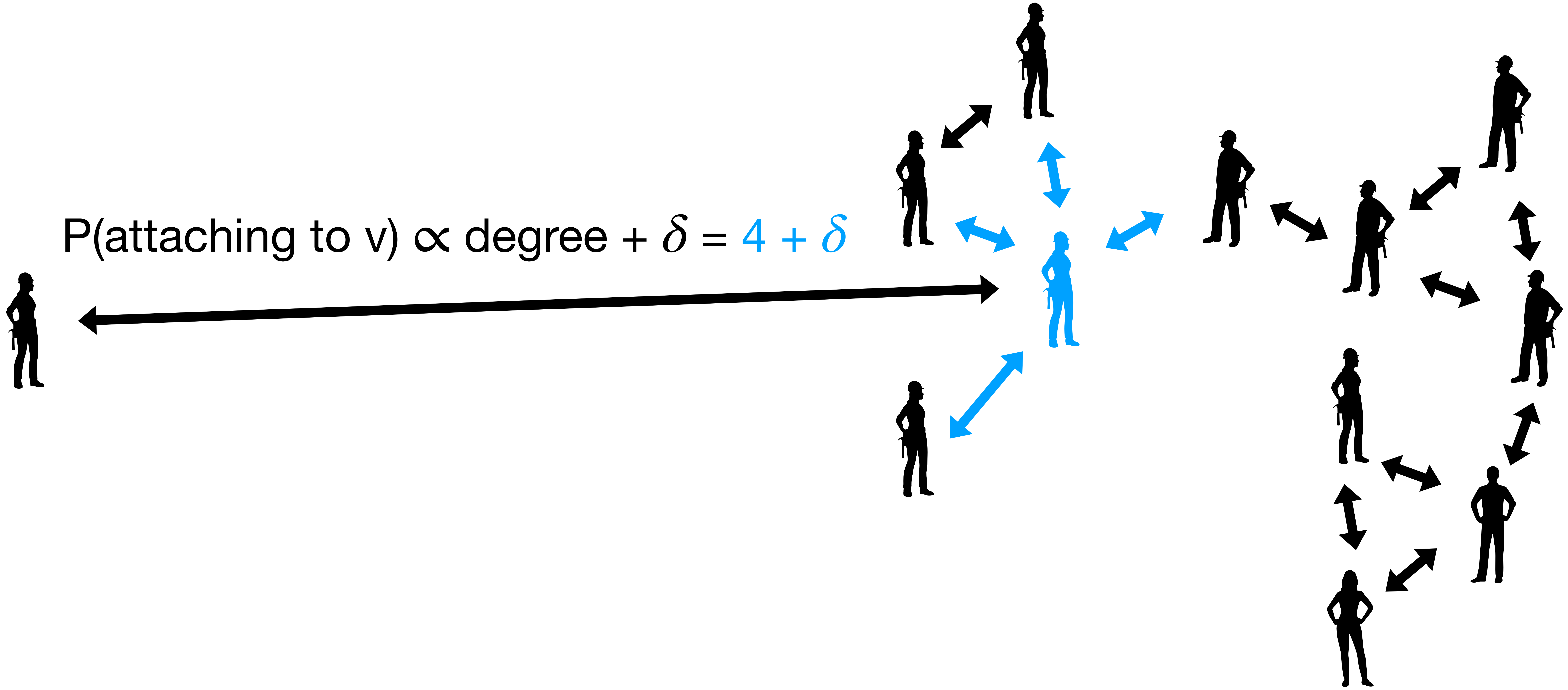
Preferential Attachment

[Albert and Barabasi 1999]



Preferential Attachment

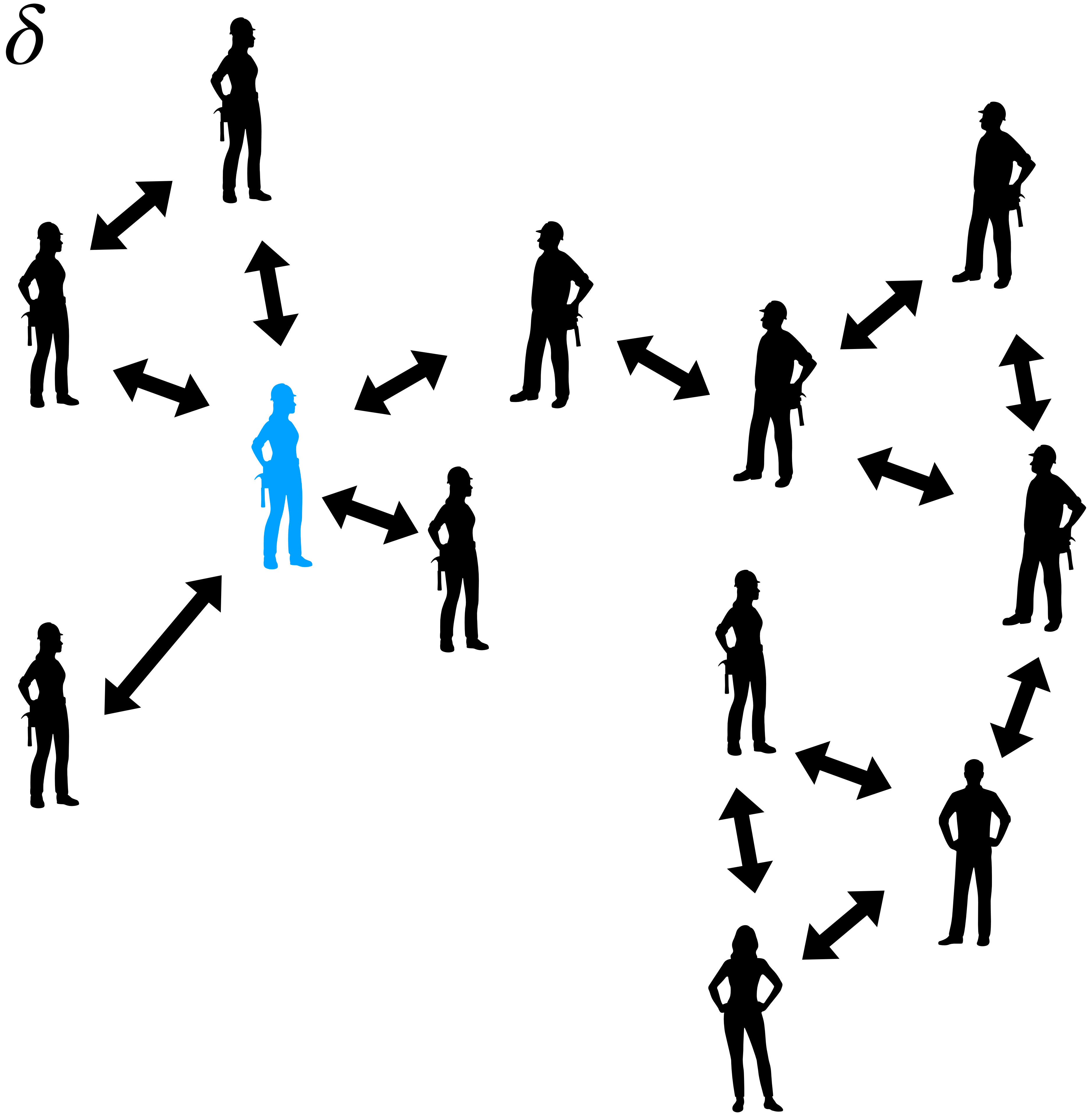
[Albert and Barabasi 1999]



Preferential Attachment

[Albert and Barabasi 1999]

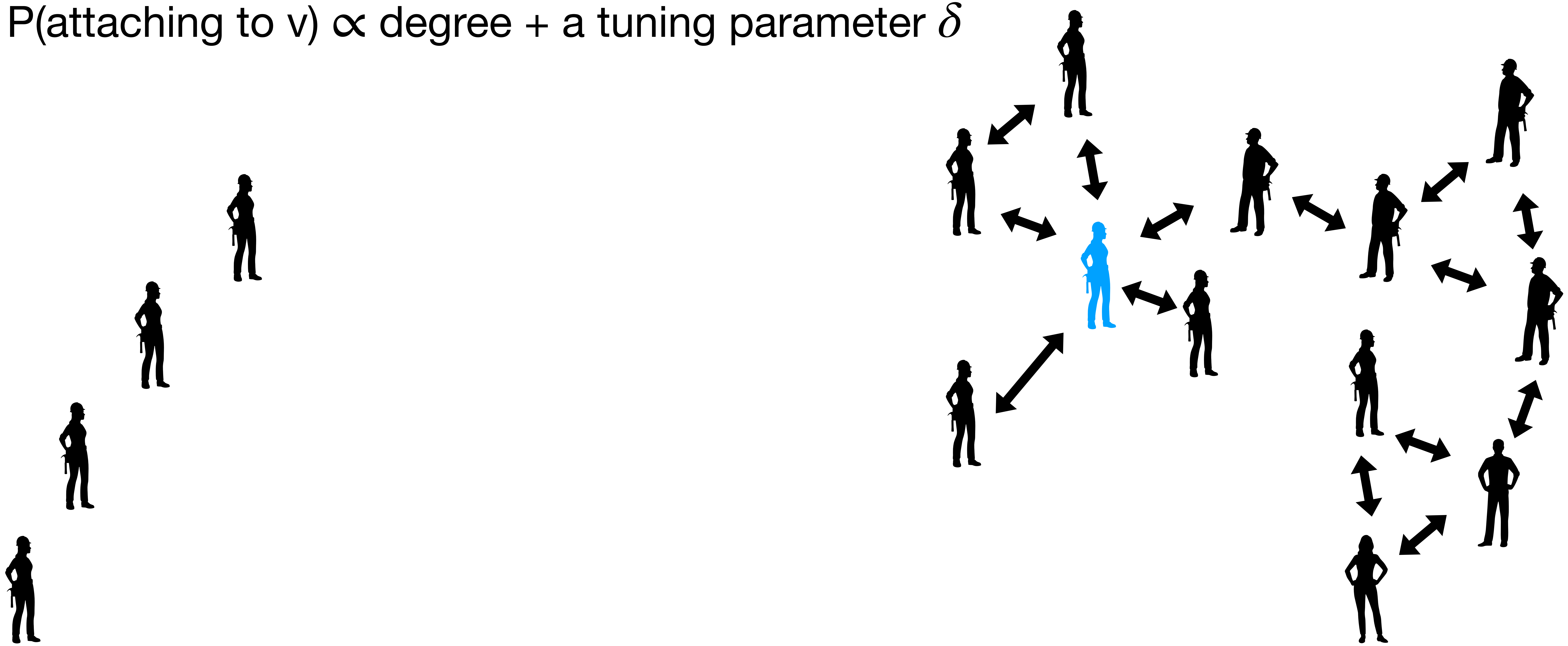
P(attaching to v) $\propto$ degree + a tuning parameter δ



Preferential Attachment

[Albert and Barabasi 1999]

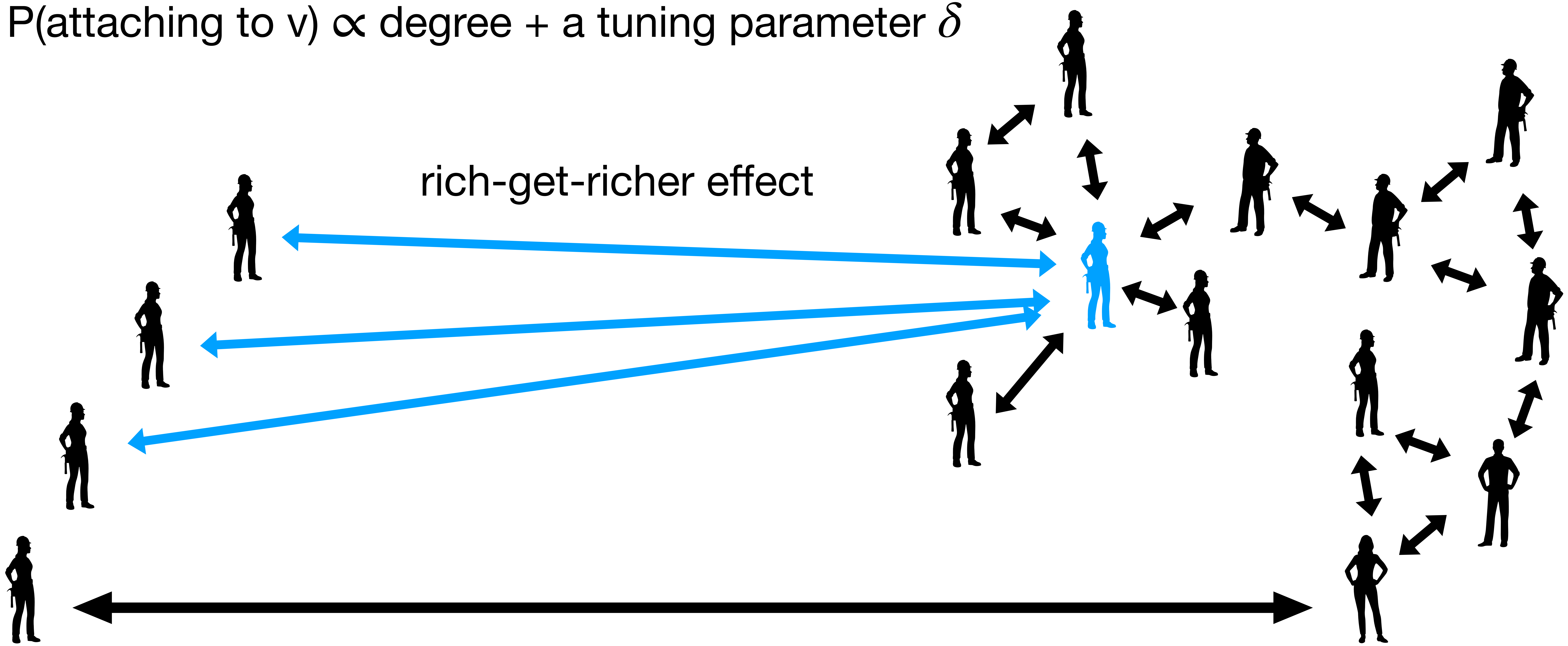
P(attaching to v) $\propto$ degree + a tuning parameter δ



Preferential Attachment

[Albert and Barabasi 1999]

P(attaching to v) $\propto$ degree + a tuning parameter δ



What do we know?

What do we know?

- triangle counts and clustering coefficient [Bollobas and Ridden 2002, Prokhorenkova et al 2013]

What do we know?

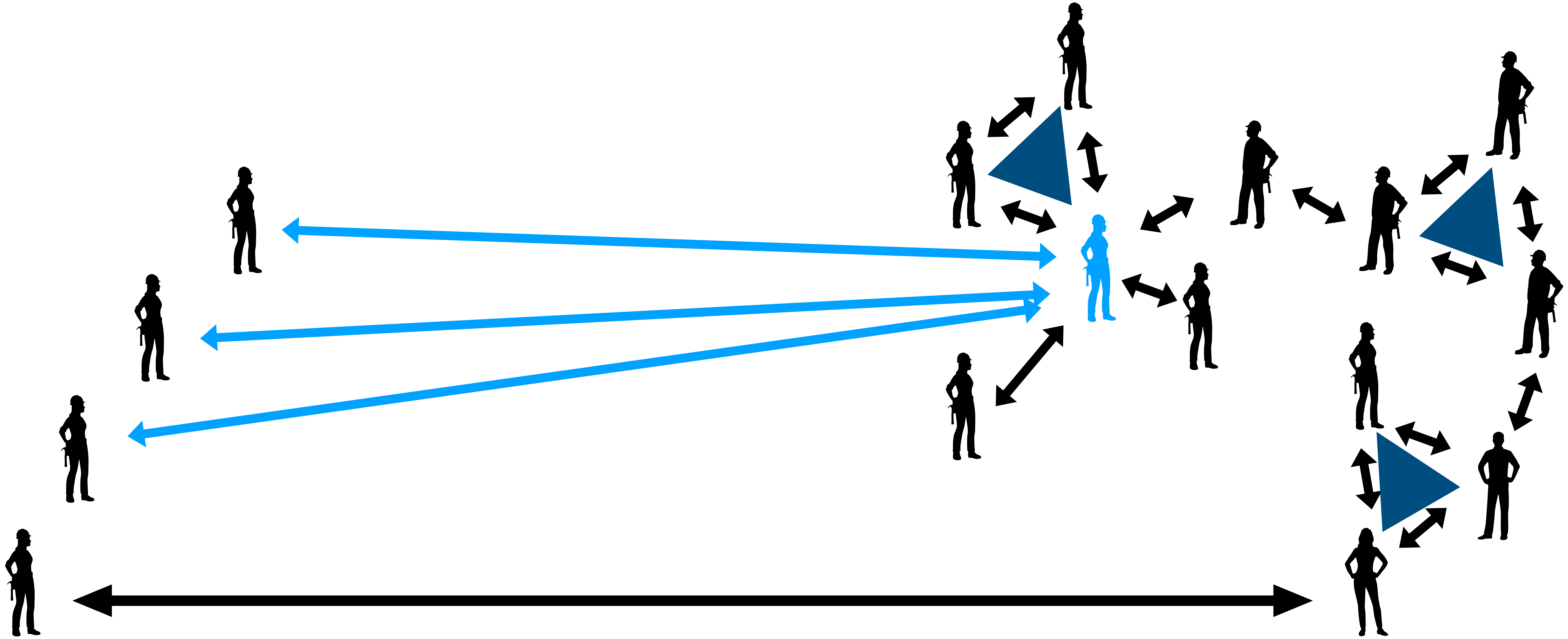
- triangle counts and clustering coefficient [Bollobas and Ridden 2002, Prokhorenkova et al 2013]
- subgraph counts [Garavaglia and Steghuis 2019]

What do we know?

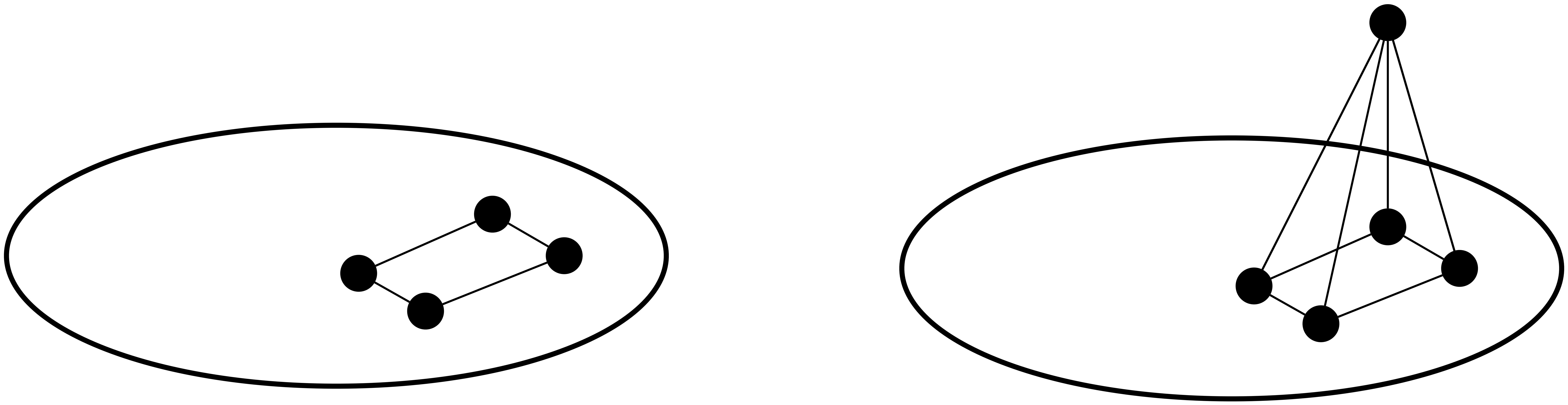
- triangle counts and clustering coefficient [Bollobas and Ridden 2002, Prokhorenkova et al 2013]
- subgraph counts [Garavaglia and Steghuis 2019]
- and more…

Clique Complex

aka Flag Complex

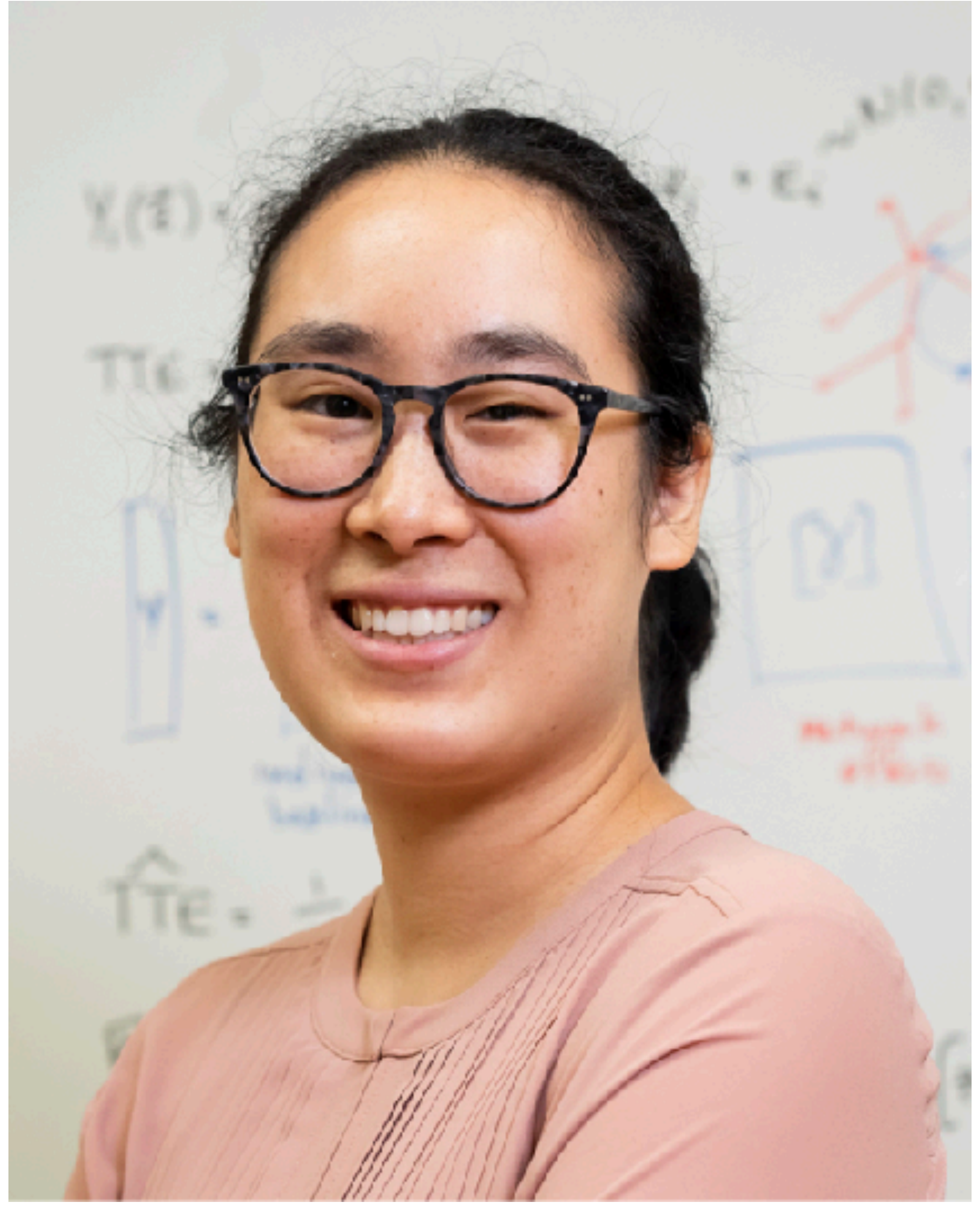


Clique Complex = Mapping Cone



III Topology of Preferential Attachment

My Lovely Collaborators



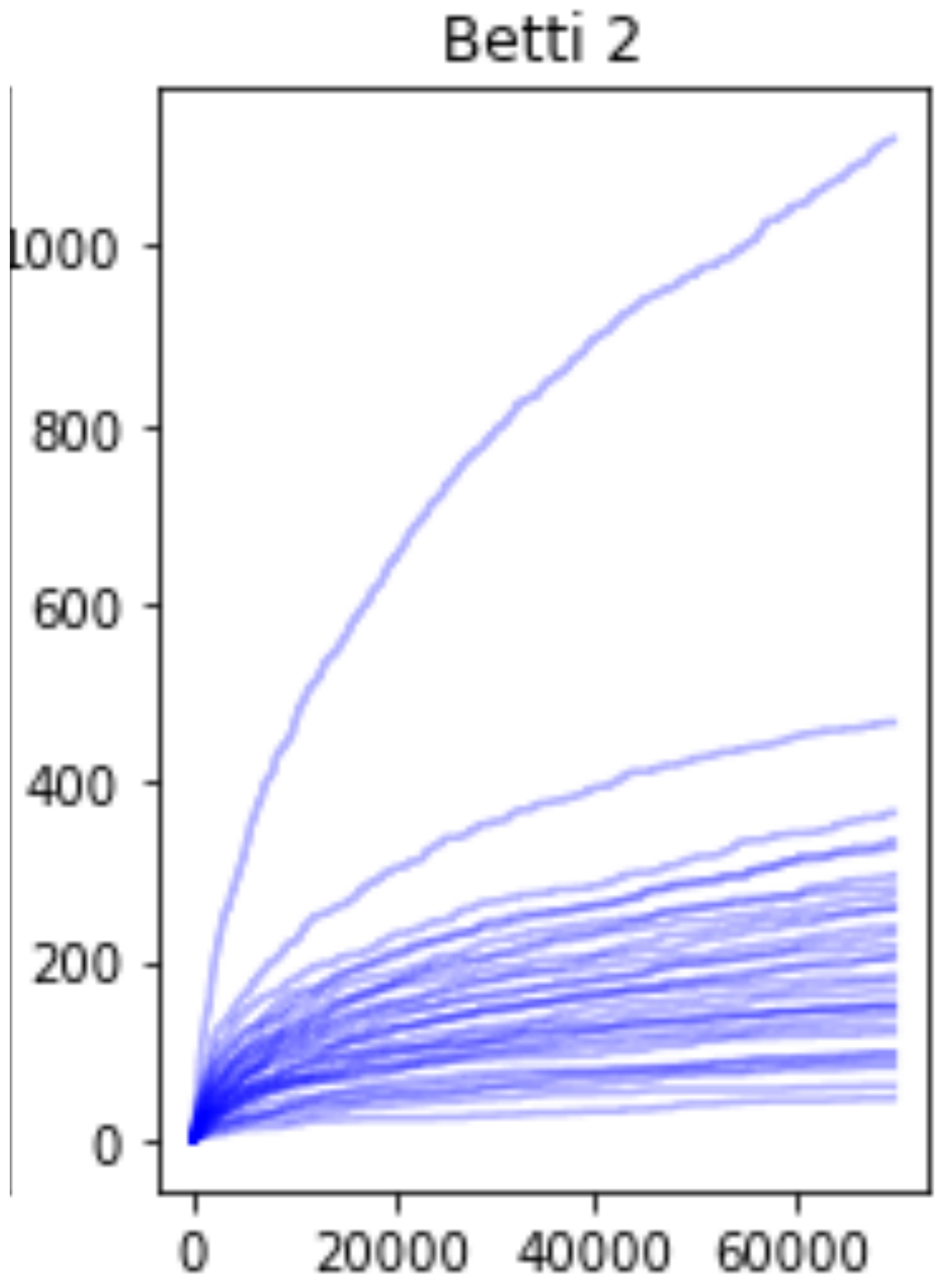
Christina Lee Yu

Gennady Samorodnitsky

Rongyi He (Caroline)

Expected Betti Number $E[\beta_q]$

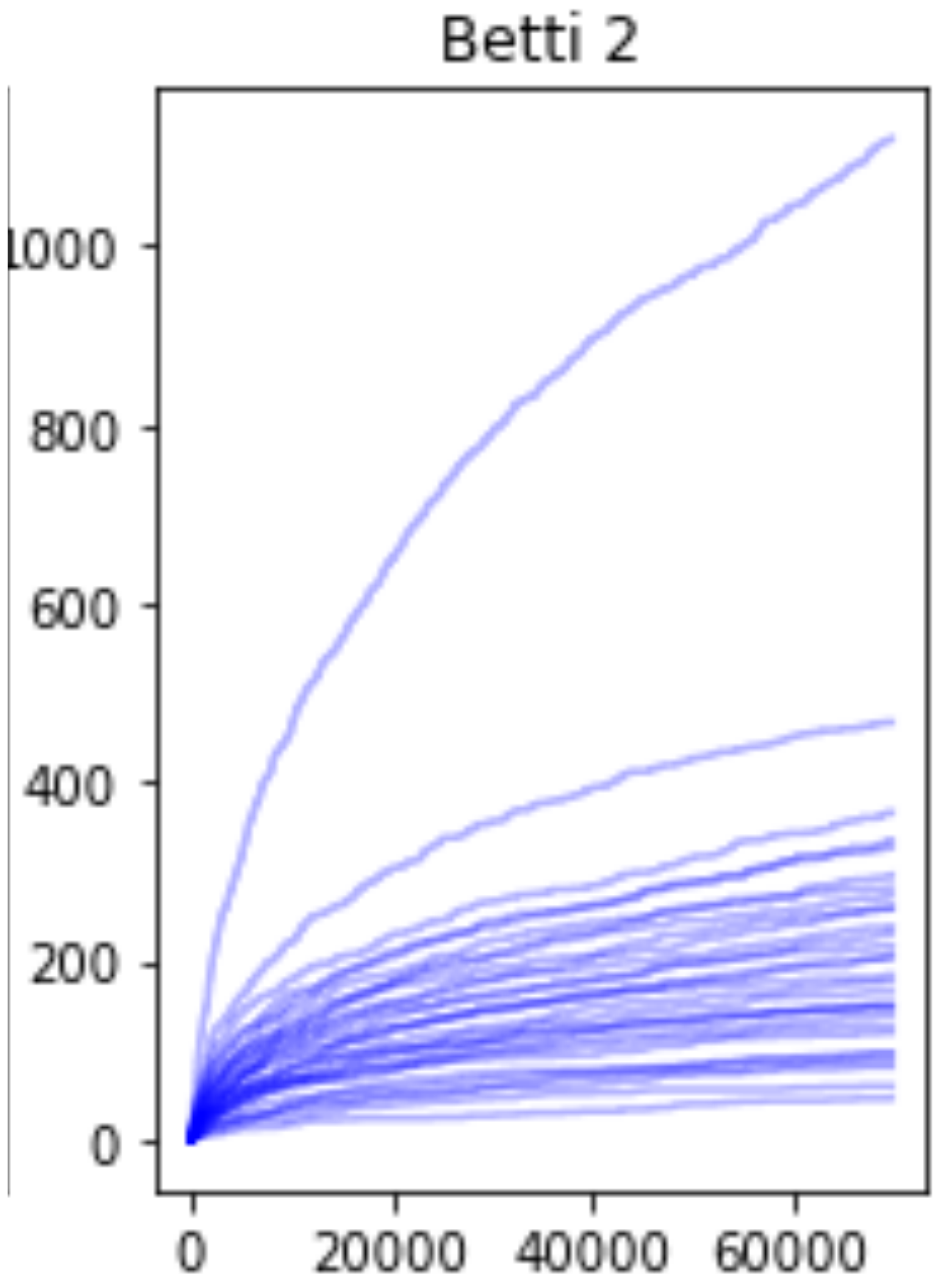
Expected Betti Number $E[\beta_q]$



Different curves, different random seeds.
All curves have the same model parameters.

Expected Betti Number $E[\beta_q]$

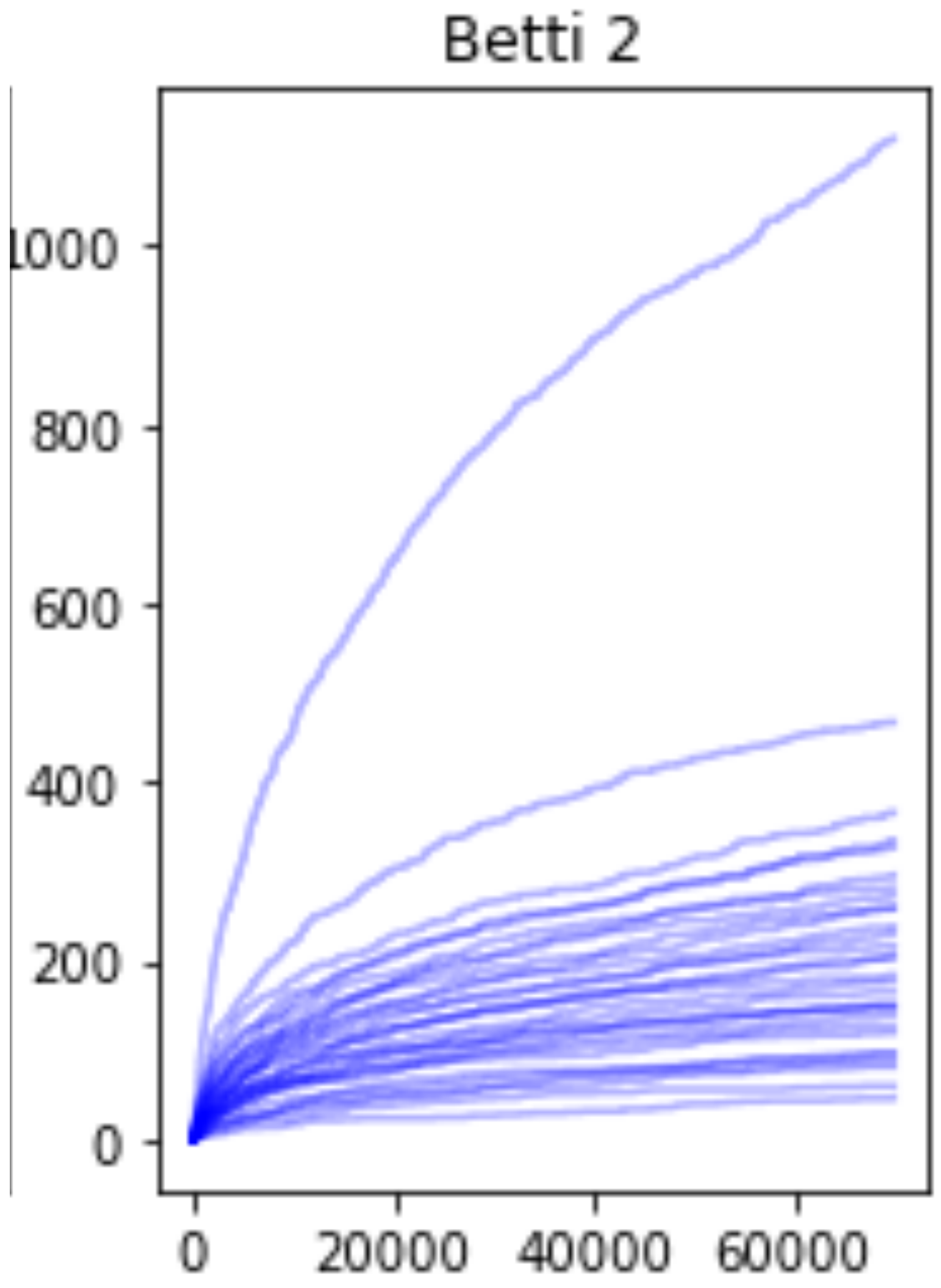
- increasing trend



Different curves, different random seeds.
All curves have the same model parameters.

Expected Betti Number $E[\beta_q]$

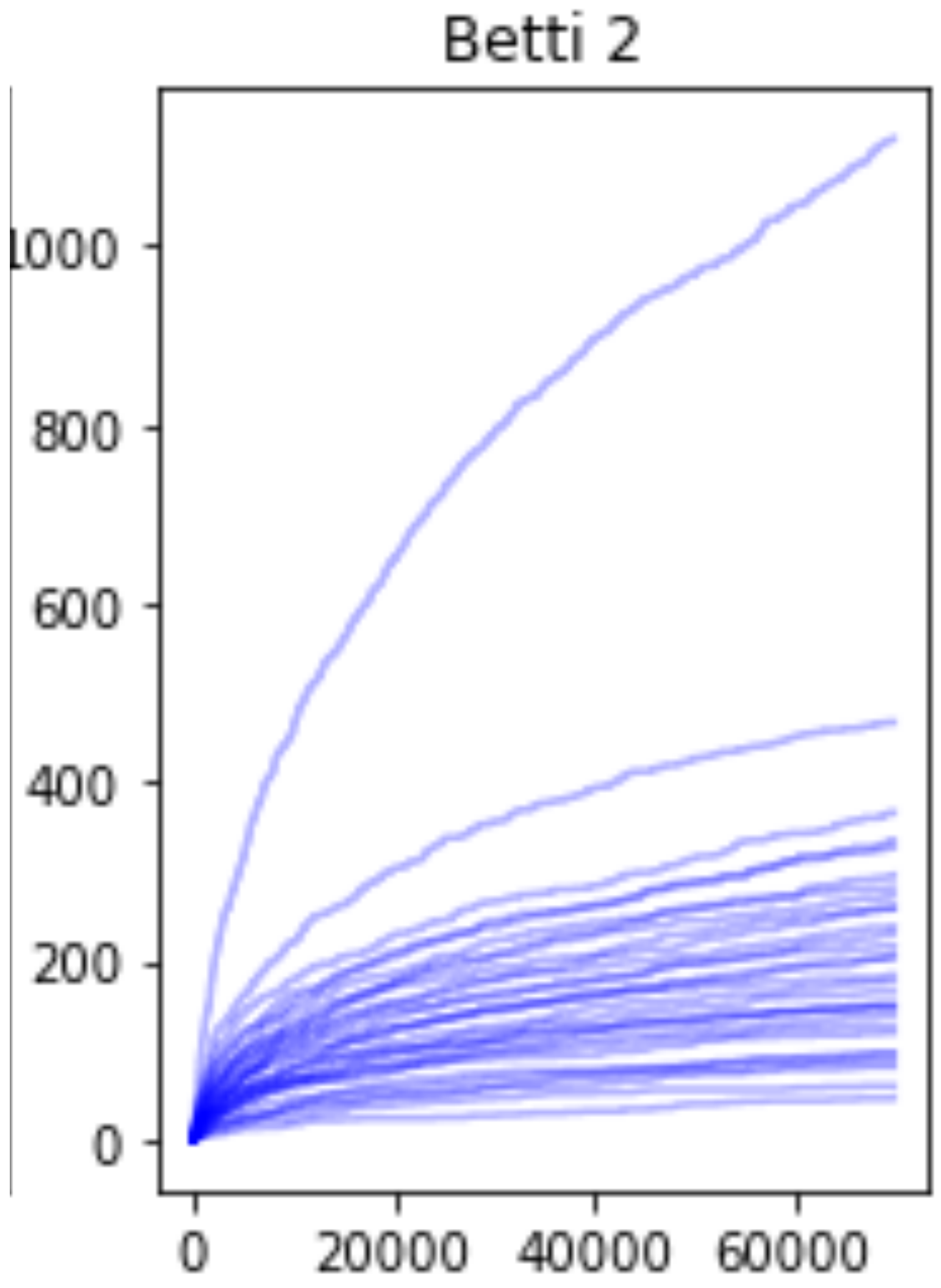
- increasing trend
- concave growth



Different curves, different random seeds.
All curves have the same model parameters.

Expected Betti Number $E[\beta_q]$

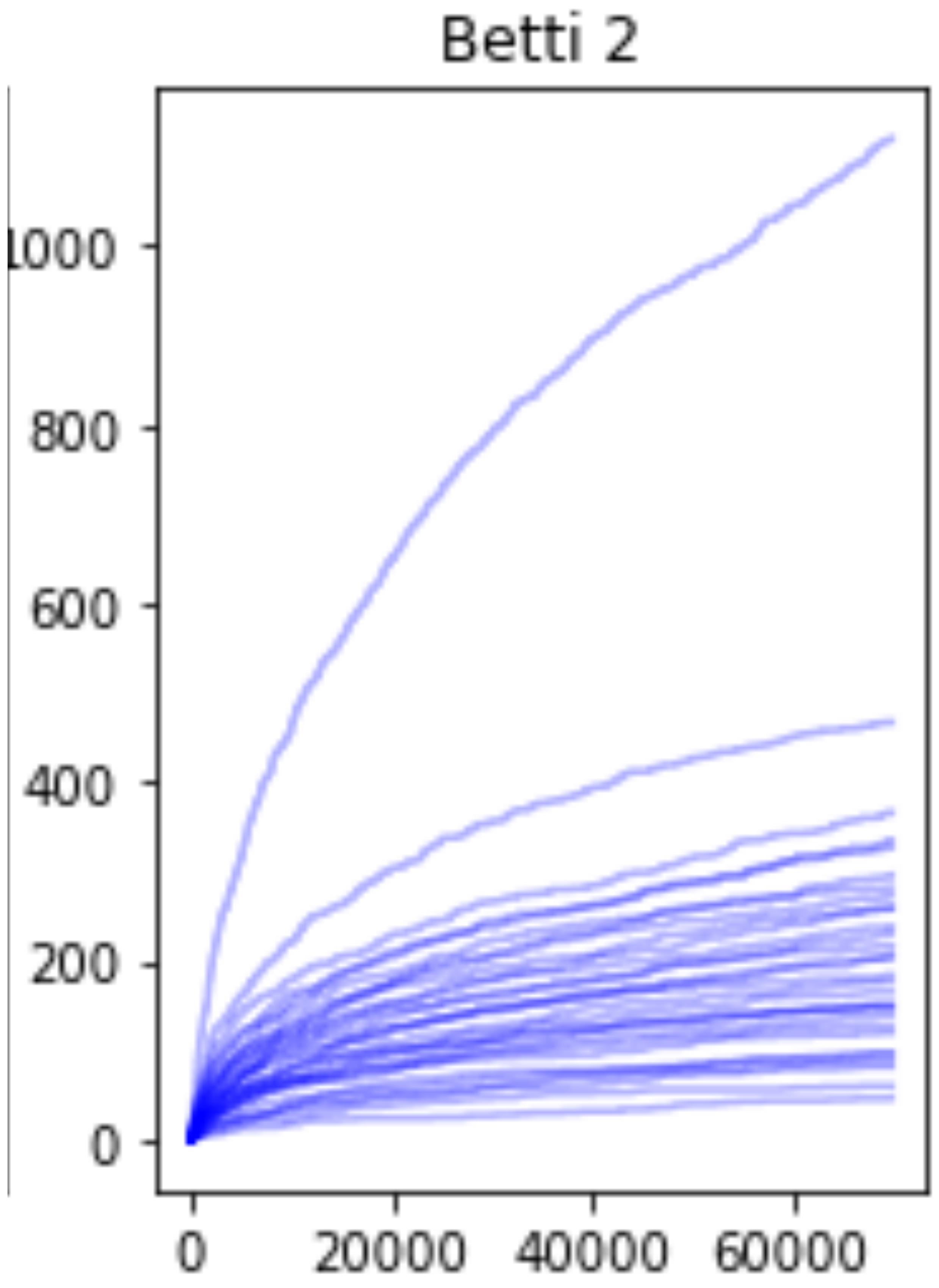
- increasing trend
- concave growth
- outlier



Different curves, different random seeds.
All curves have the same model parameters.

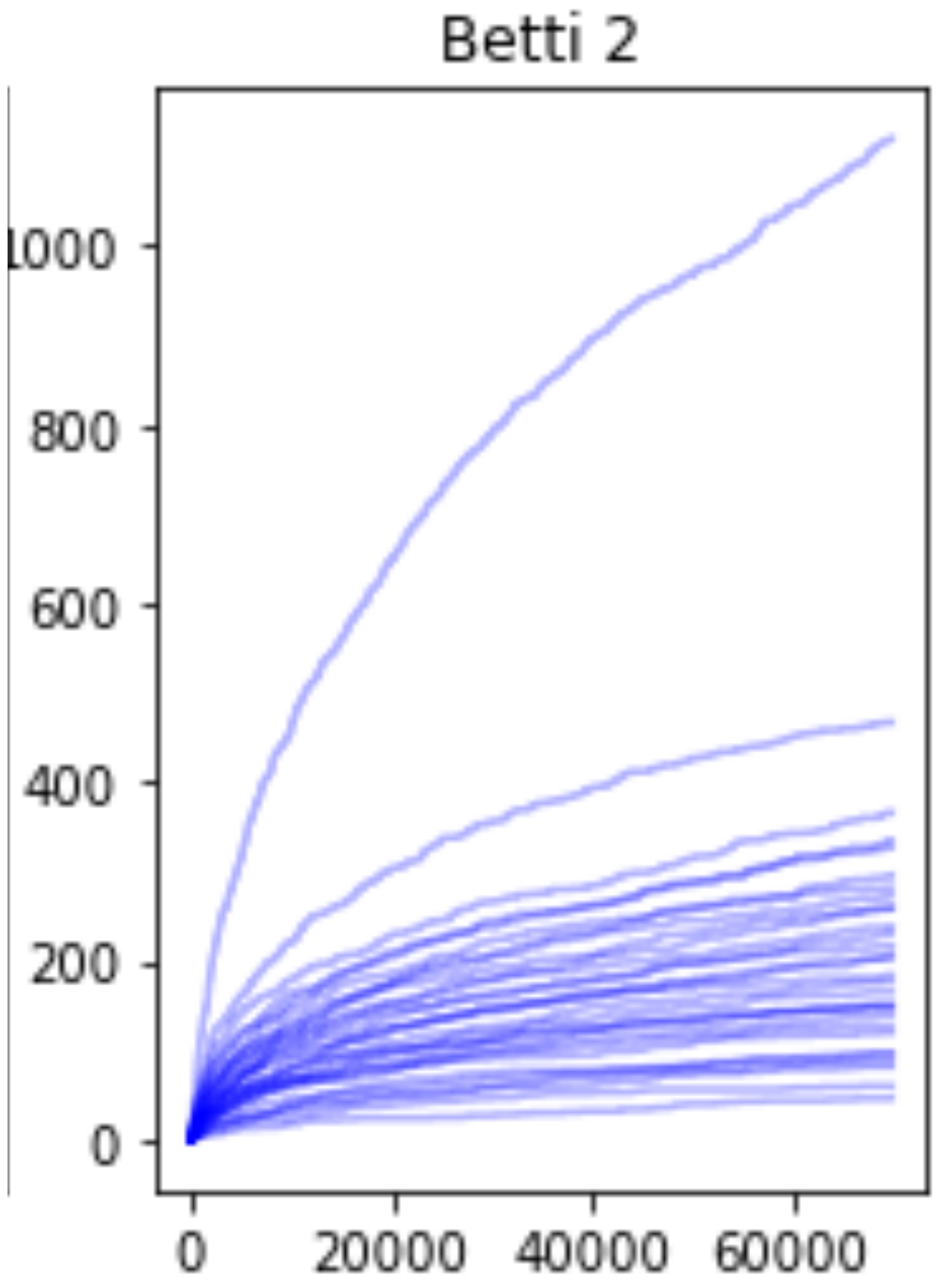
Expected Betti Number $E[\beta_q]$

- $c(\text{num of nodes}^{1-4x}) \leq E[\beta_2] \leq C(\text{num of nodes}^{1-4x})$
  under mild assumptions
  - $x \in (0,1/2)$ depends on the preferential attachment strength.



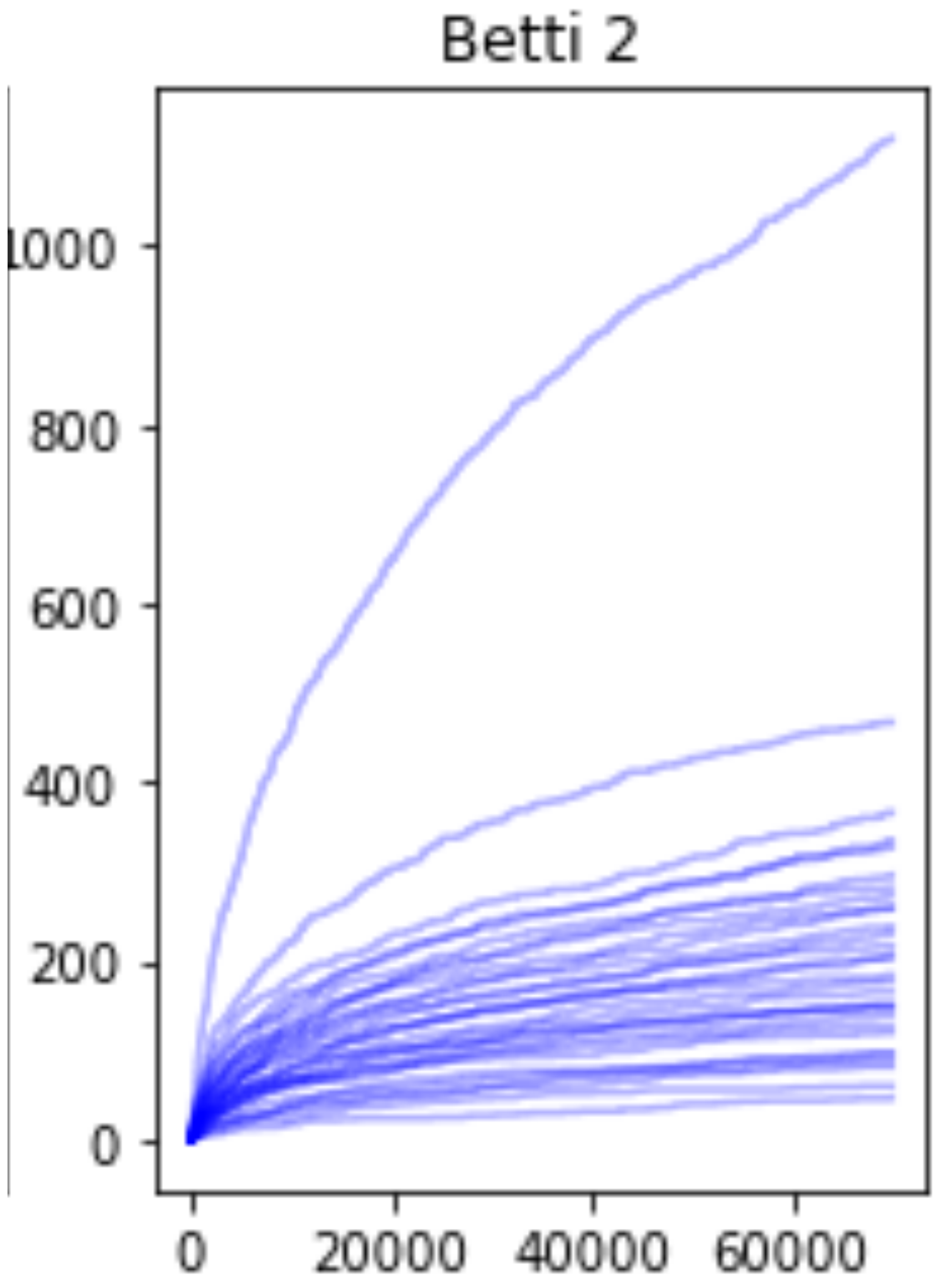
Expected Betti Number $E[\beta_q]$

- $c(\text{num of nodes}^{1-4x}) \leq E[\beta_2] \leq C(\text{num of nodes}^{1-4x})$
  under mild assumptions
  - $x \in (0,1/2)$ depends on the preferential attachment strength.
  - If $1 - 4x < 0$, then $E[\beta_2] \leq C$.



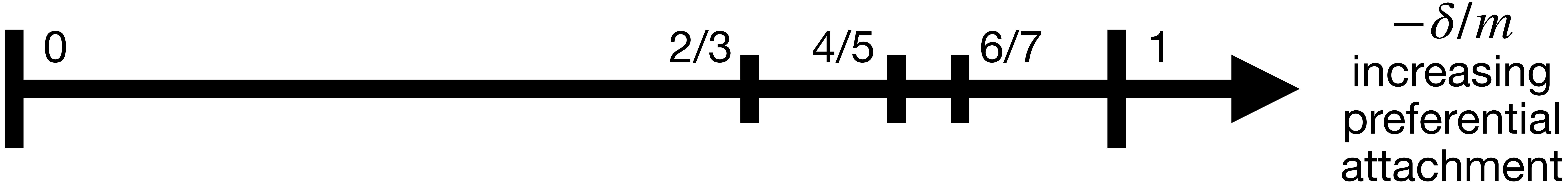
Expected Betti Number $E[\beta_q]$

- c(num of nodes^{1-4x}) $\leq E[\beta_2] \leq C$(num of nodes^{1-4x})
  under mild assumptions
  - $x \in (0,1/2)$ depends on the preferential attachment strength
  - If $1 - 4x < 0$, then $E[\beta_2] \leq C$.
- c(num of nodes^{1-2qx}) $\leq E[\beta_q] \leq C$(num of nodes^{1-2qx}) for $q \geq 2$.



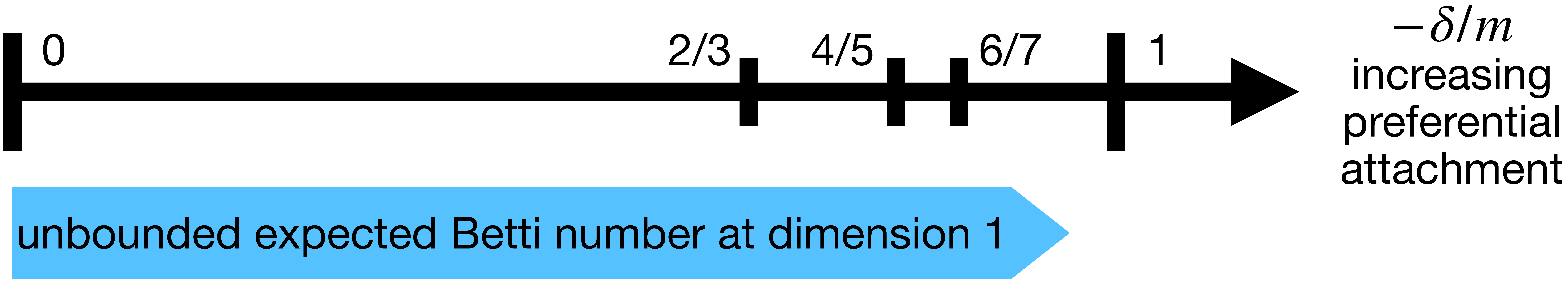
Phase transition

Recall
P(attaching to v) $\propto$ degree + δ
m = number of edges per new node



Phase transition

Recall
P(attaching to v) ∝ degree + δ
m = number of edges per new node

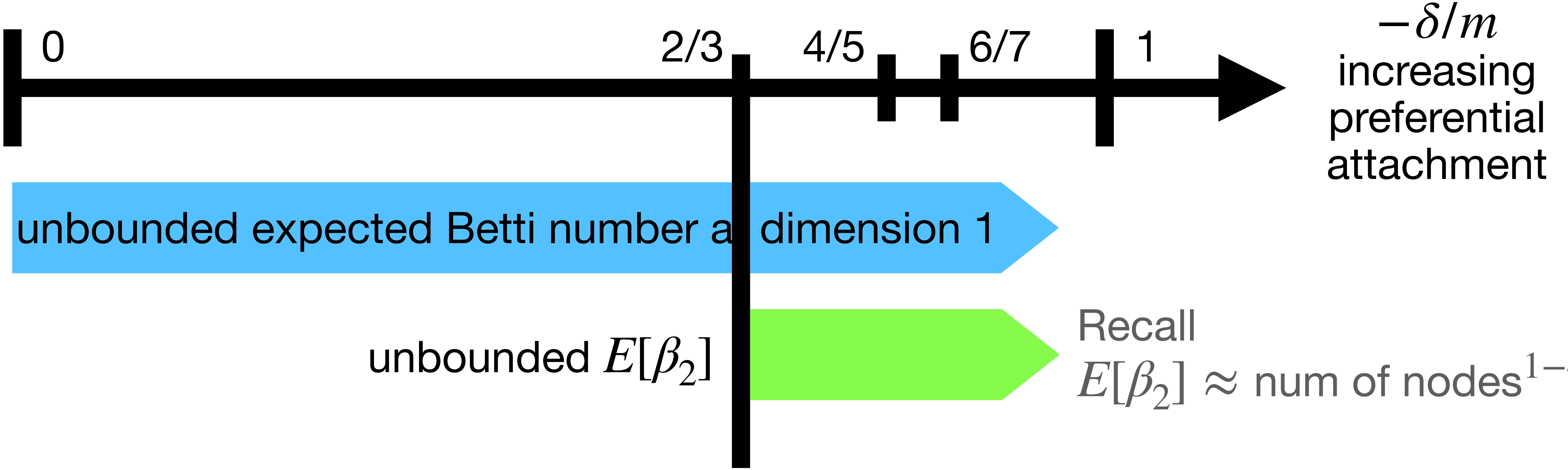


Phase transition

Recall

P(attaching to v) ∝ degree + δ

m = number of edges per new node



Recall

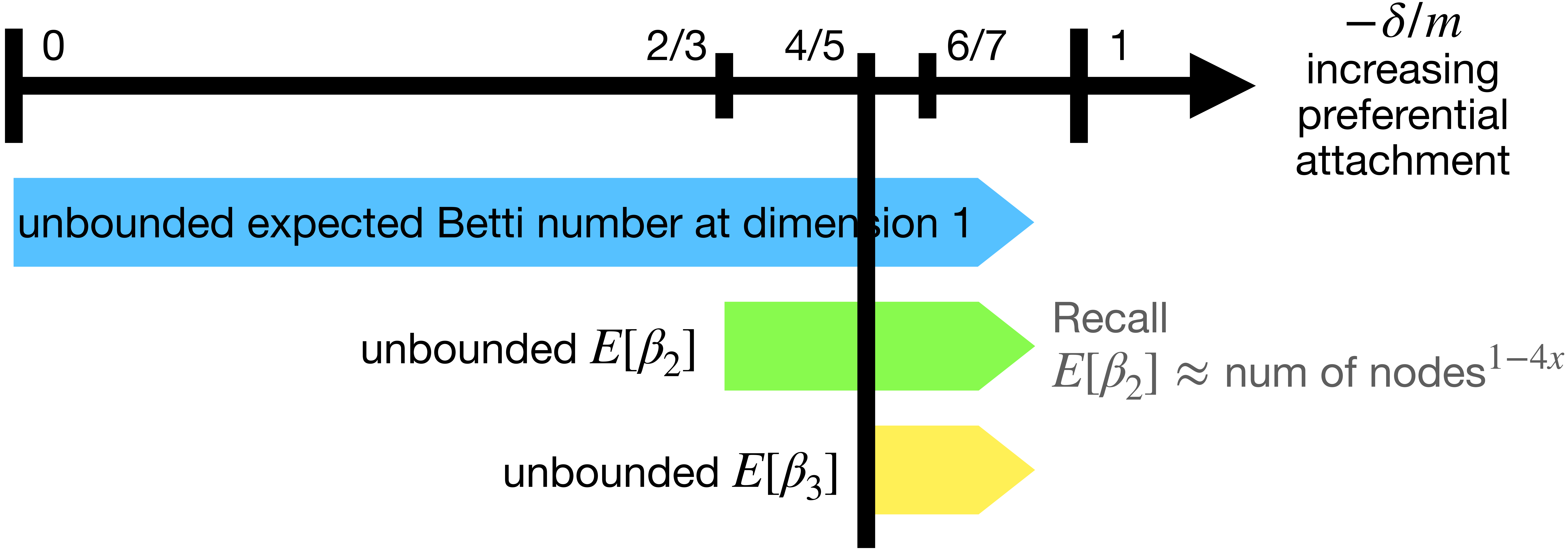
$E[\beta_2] \approx$ num of nodes^{1-4x}

Phase transition

Recall

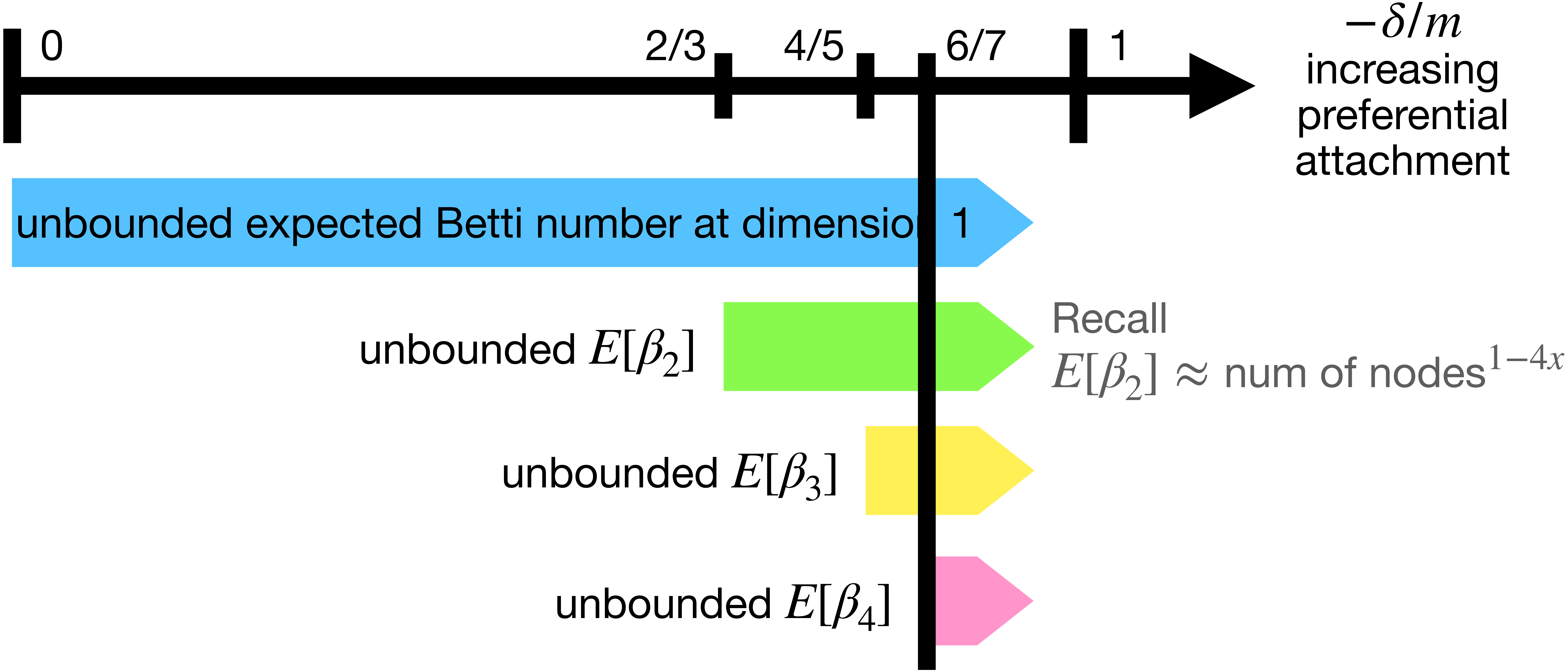
P(attaching to v) $\propto$ degree + δ

m = number of edges per new node



Phase transition

Recall
P(attaching to v) ∝ degree + δ
m = number of edges per new node

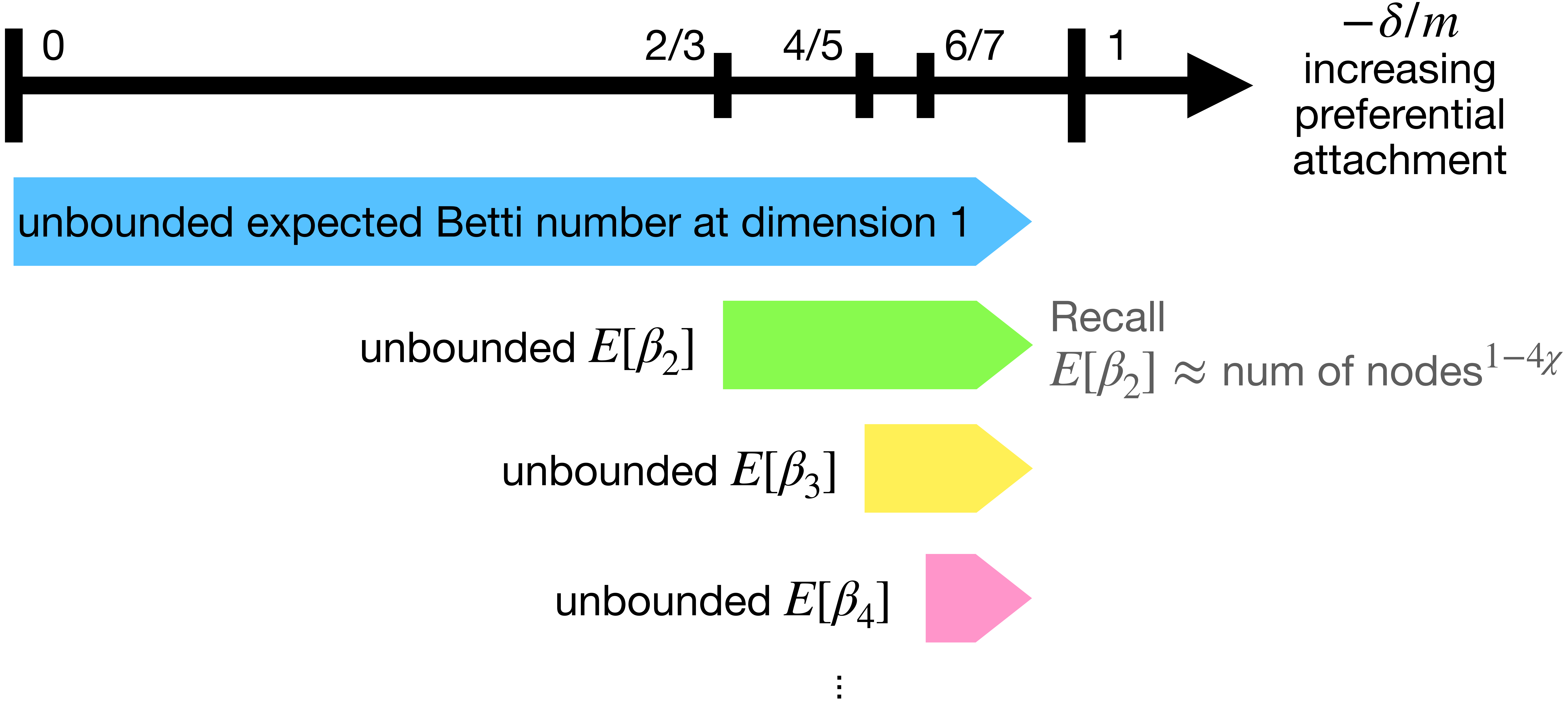


Phase transition

Recall

P(attaching to v) $\propto$ degree + δ

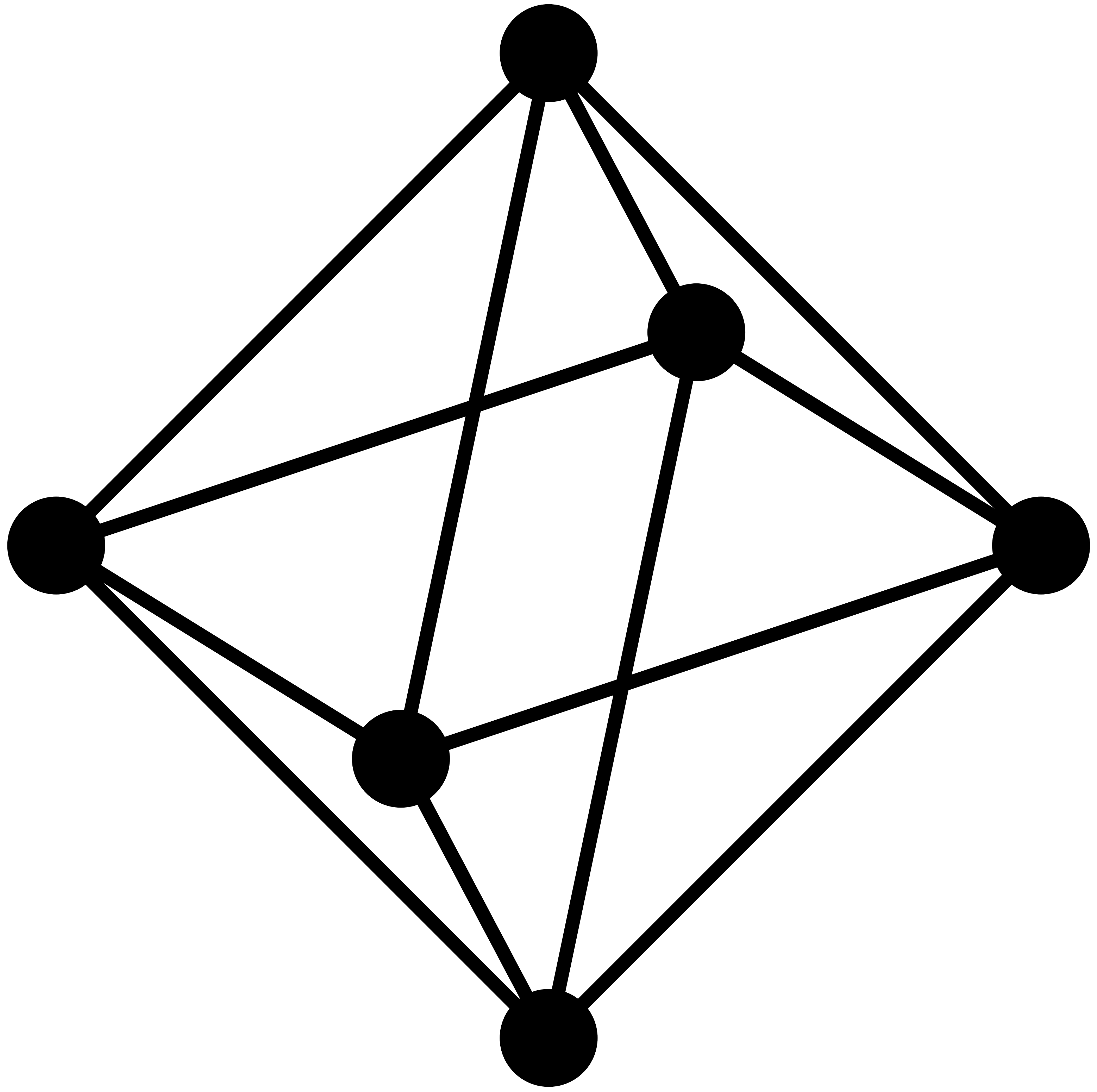
m = number of edges per new node



Theorem: $E[\beta_2] \approx$ num of nodes^{1-4x}

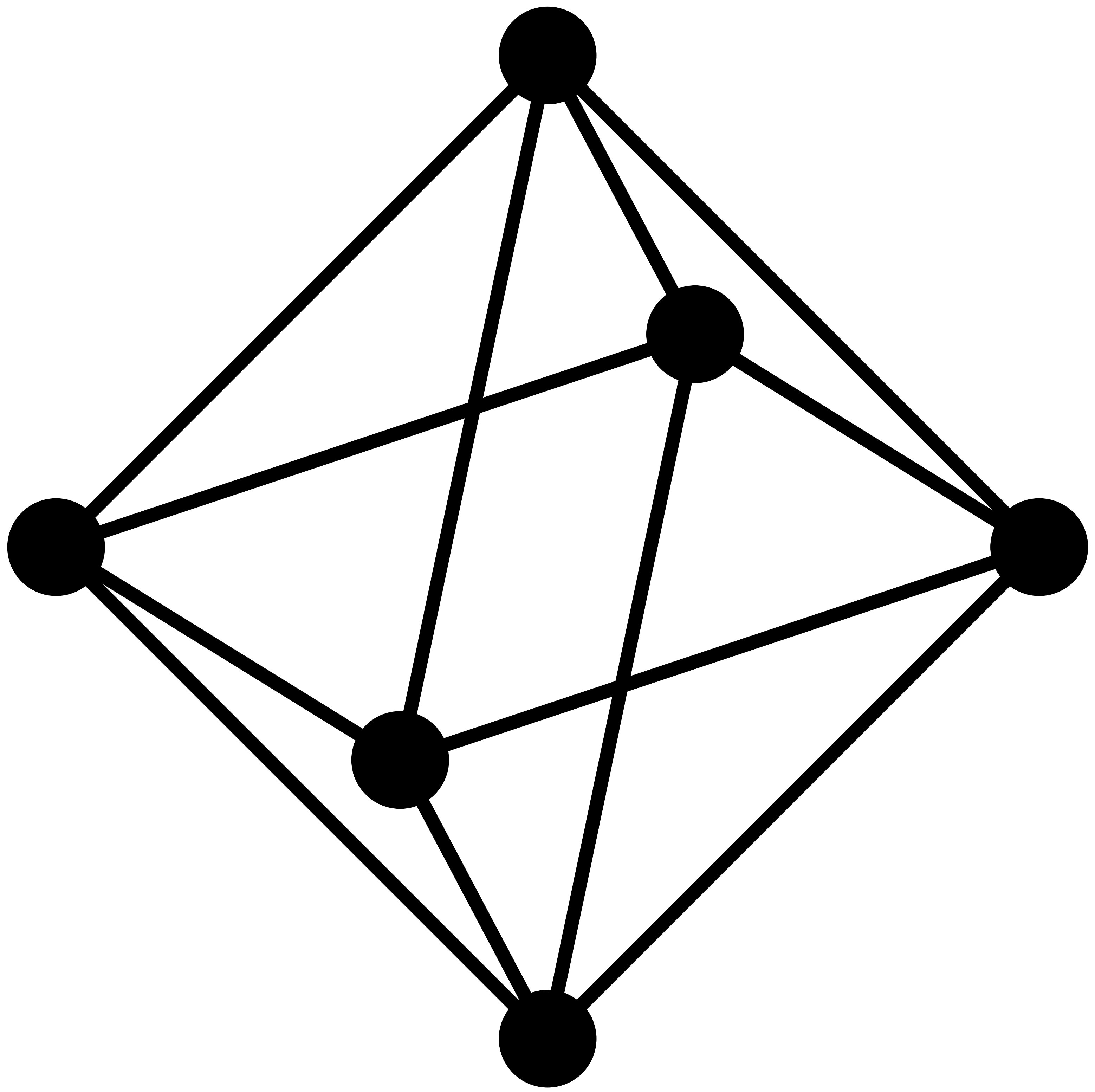
Proof?

Proof of $E[\beta_2] \approx$ num of nodes^{1-4x}

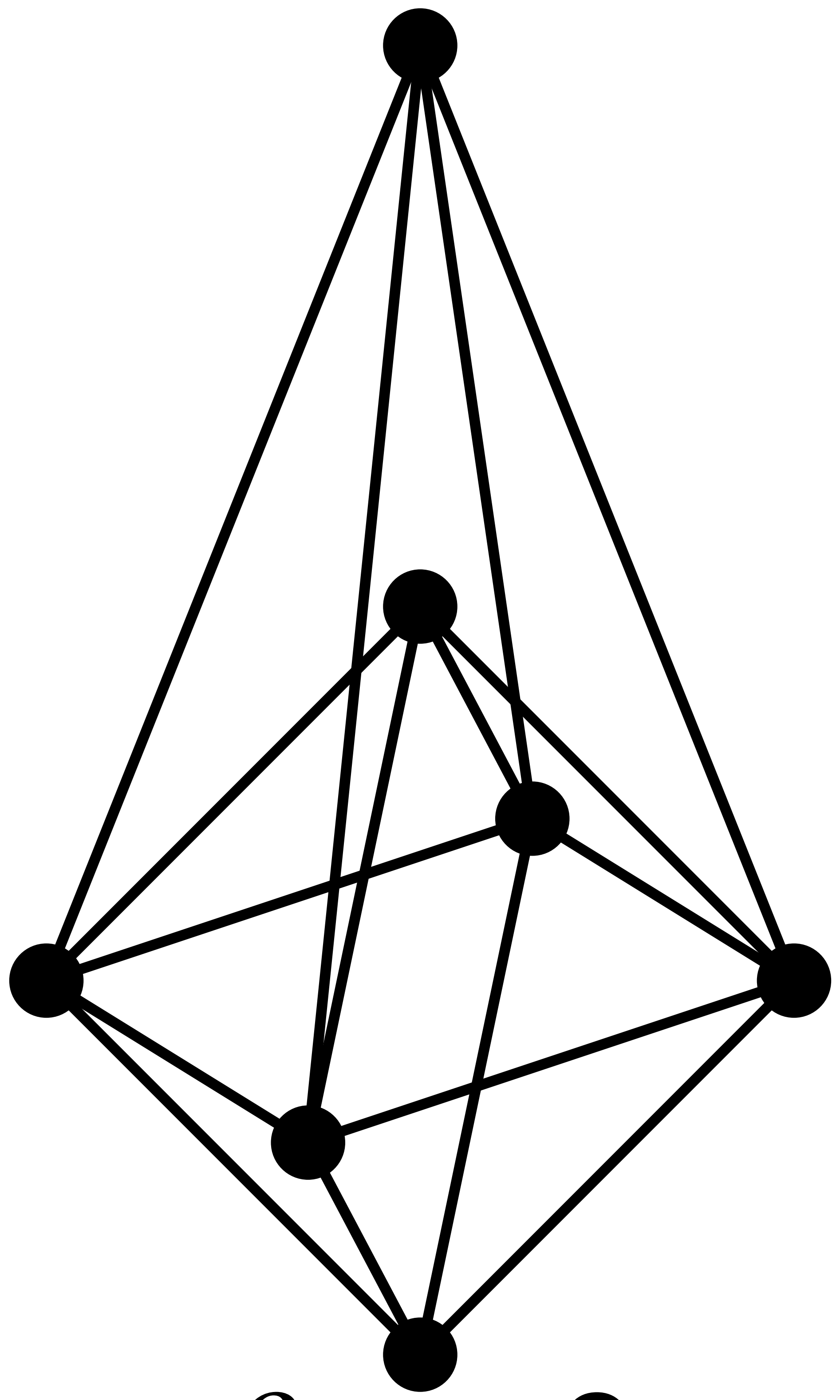


$\beta_2 = 1$

Proof of $E[\beta_2] \approx$ num of nodes^{1-4x}

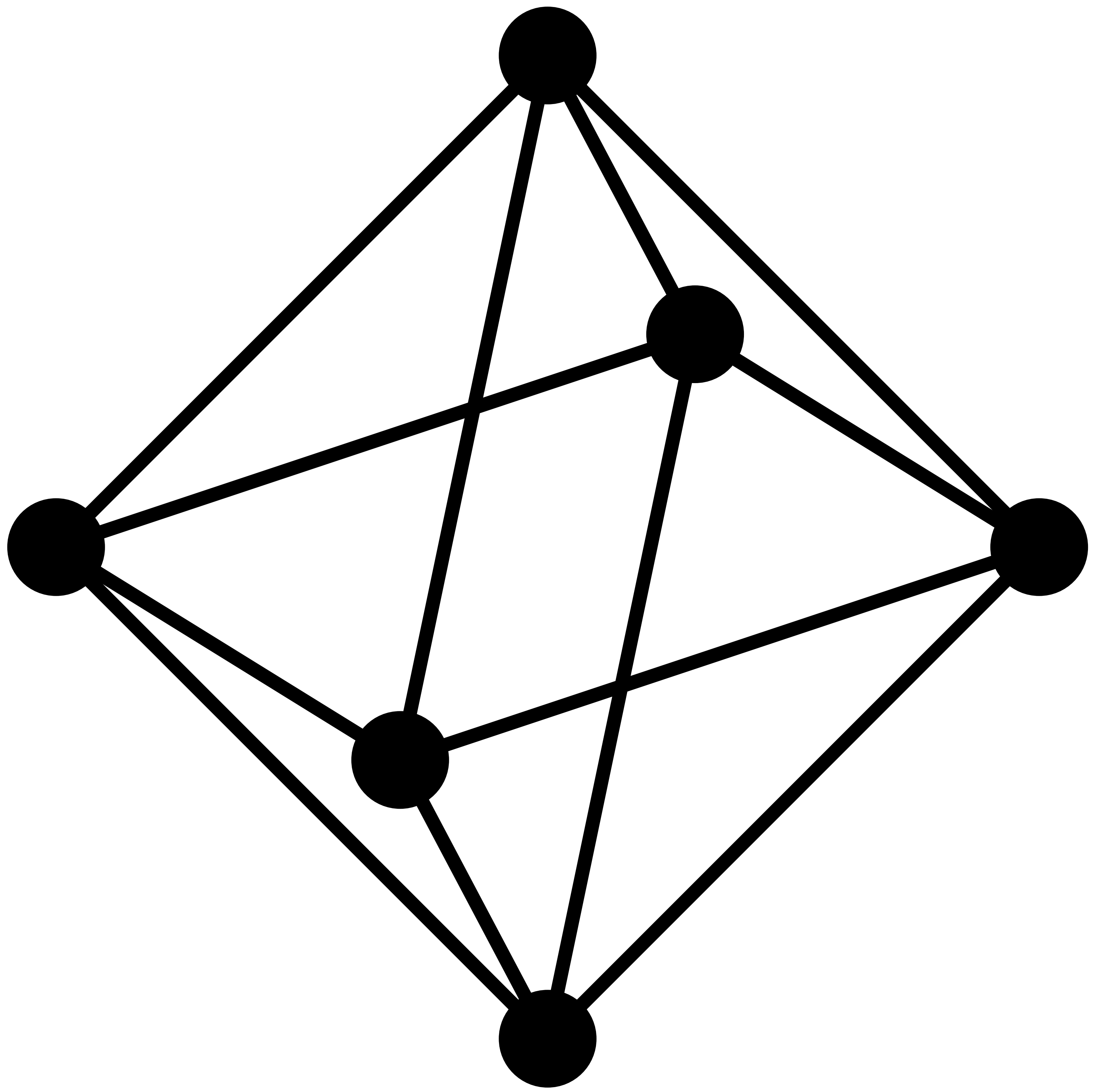


$\beta_2 = 1$

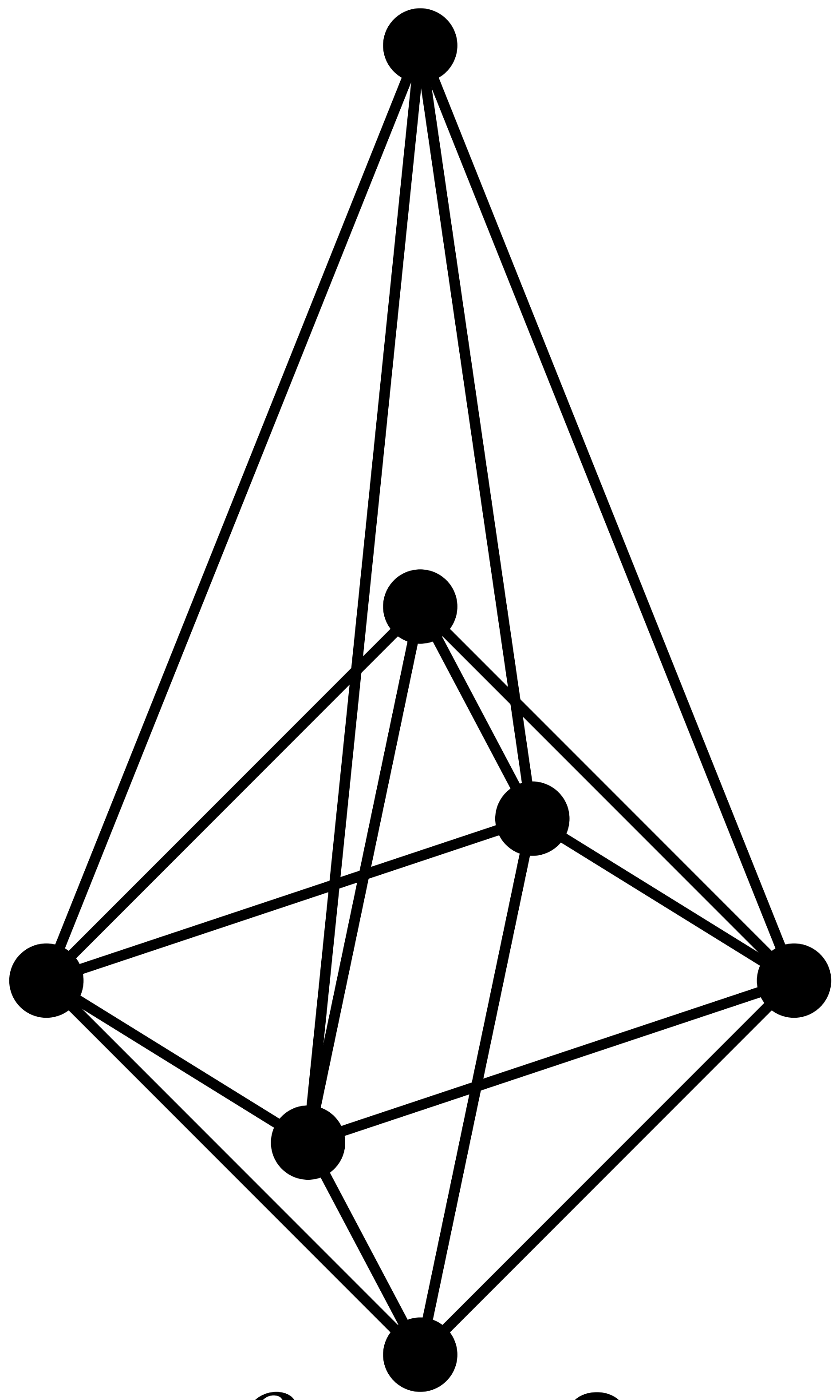


$\beta_2 = 2$

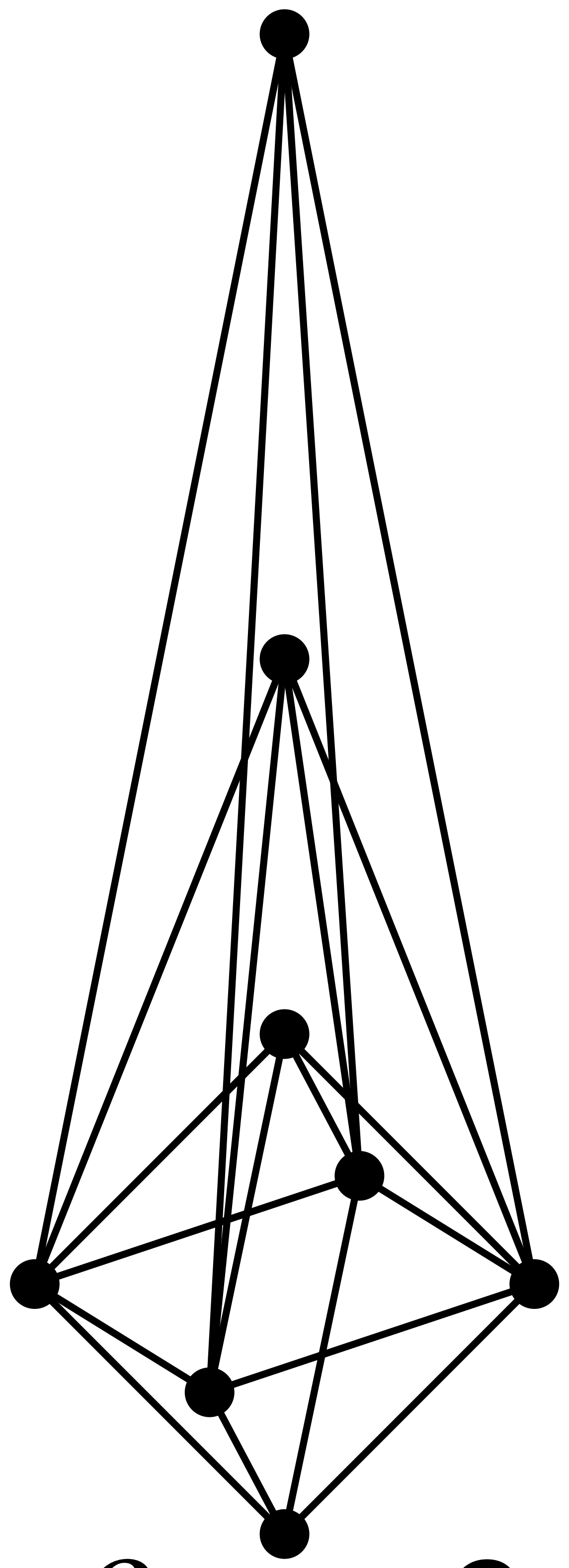
Proof of $E[\beta_2] \approx$ num of nodes^{1-4x}



$\beta_2 = 1$



$\beta_2 = 2$



$\beta_2 = 3$

Subtleties

- Need homological algebra to relate Betti numbers with counts

Subtleties

- Need homological algebra to relate Betti numbers with counts
- Identify the “square count” as the main term with minimal cycle results in [Gal 2005] and [Kahle 2009]

Subtleties

- Need homological algebra to relate Betti numbers with counts
- Identify the “square count” as the main term with minimal cycle results in [Gal 2005] and [Kahle 2009]
- Generalize minimal cycle results in the language of homological algebra

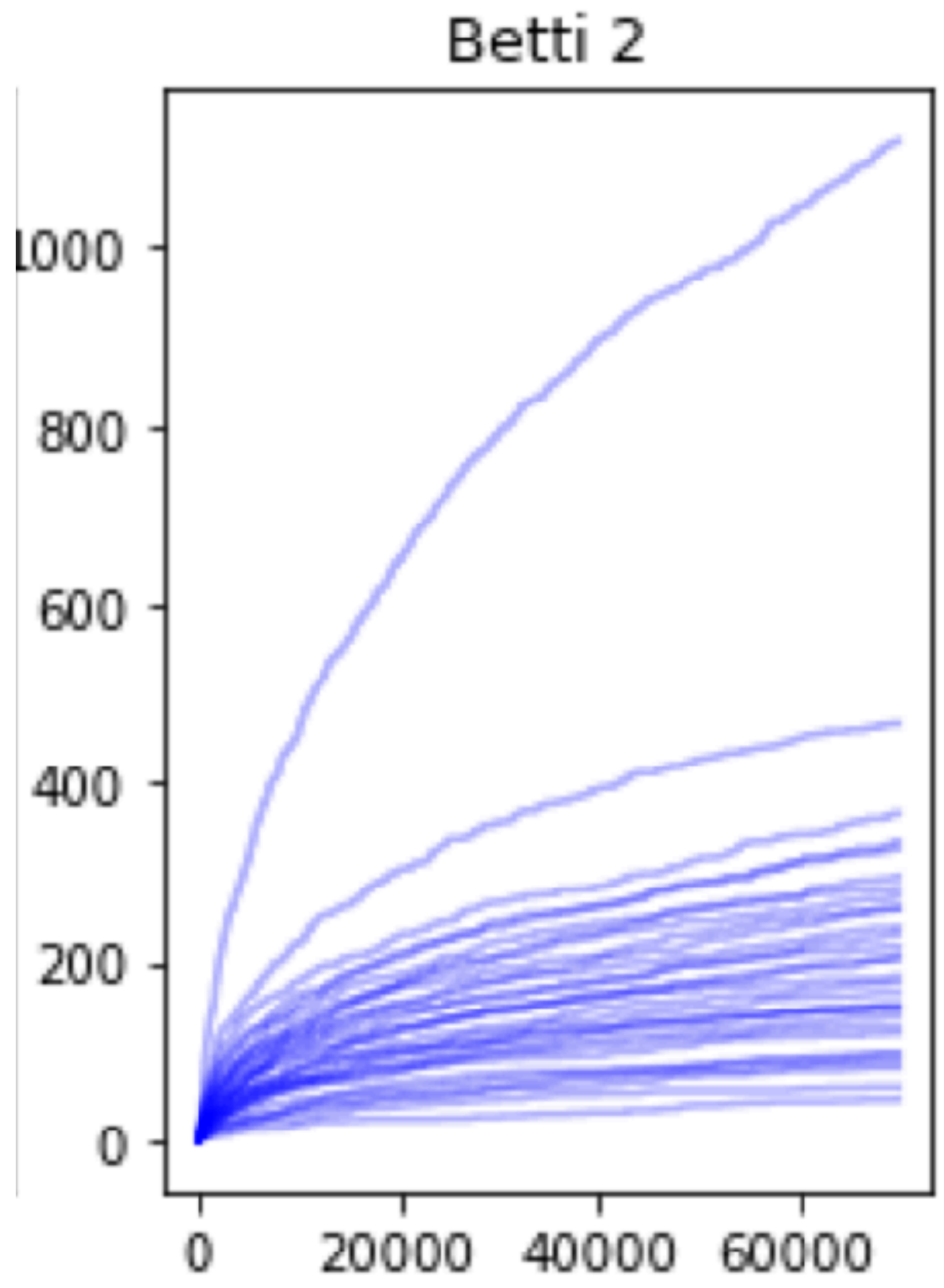
Subtleties

- Need homological algebra to relate Betti numbers with counts
- Identify the “square count” as the main term with minimal cycle results in [Gal 2005] and [Kahle 2009]
- Generalize minimal cycle results in the language of homological algebra
- Apply graph counting result in [Garavaglia and Stegehuis 2019] on a large class of subgraphs

Theorem: $E[\beta_2] \approx$ num of nodes^{1-4x}

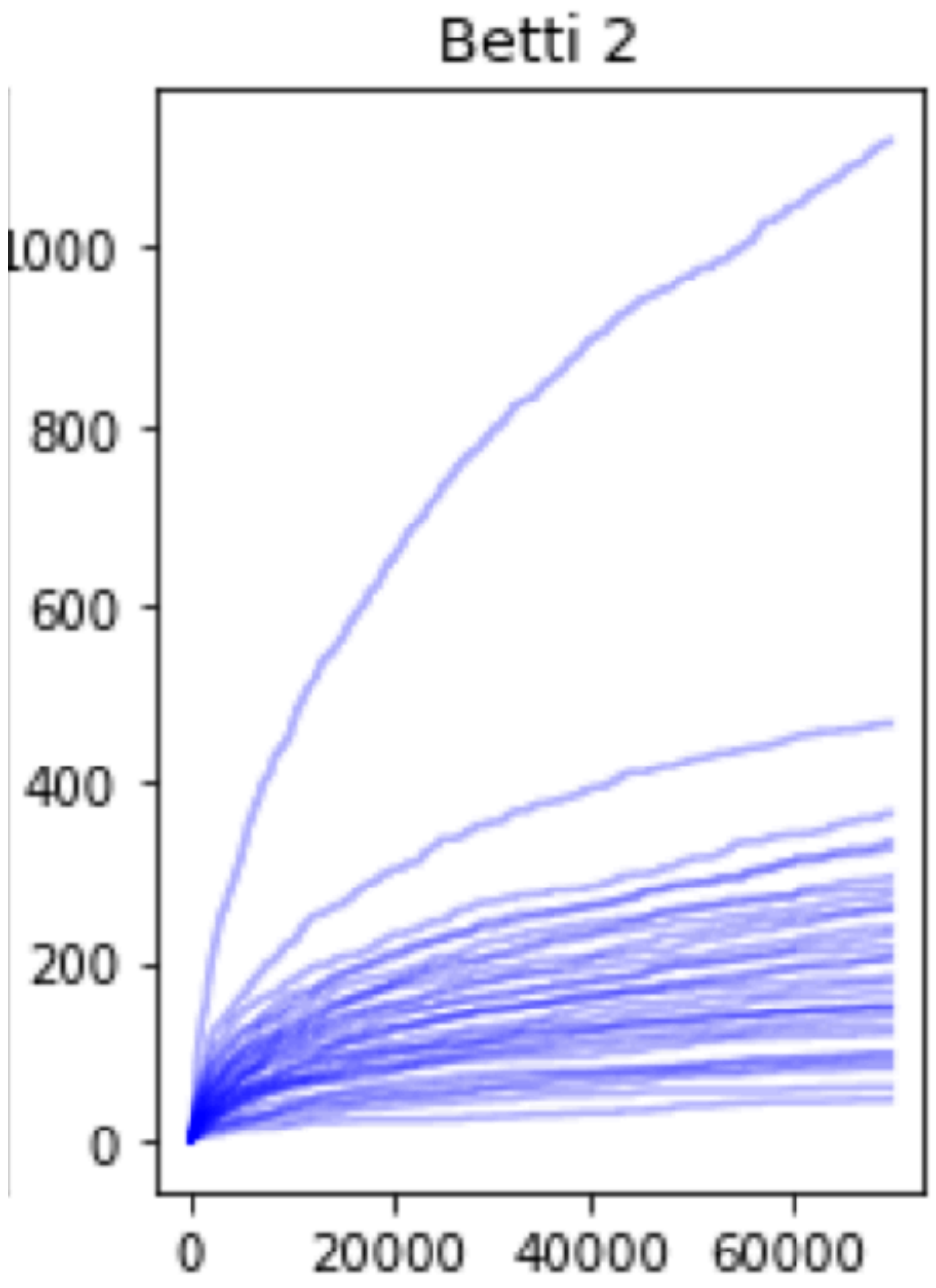
In practice???

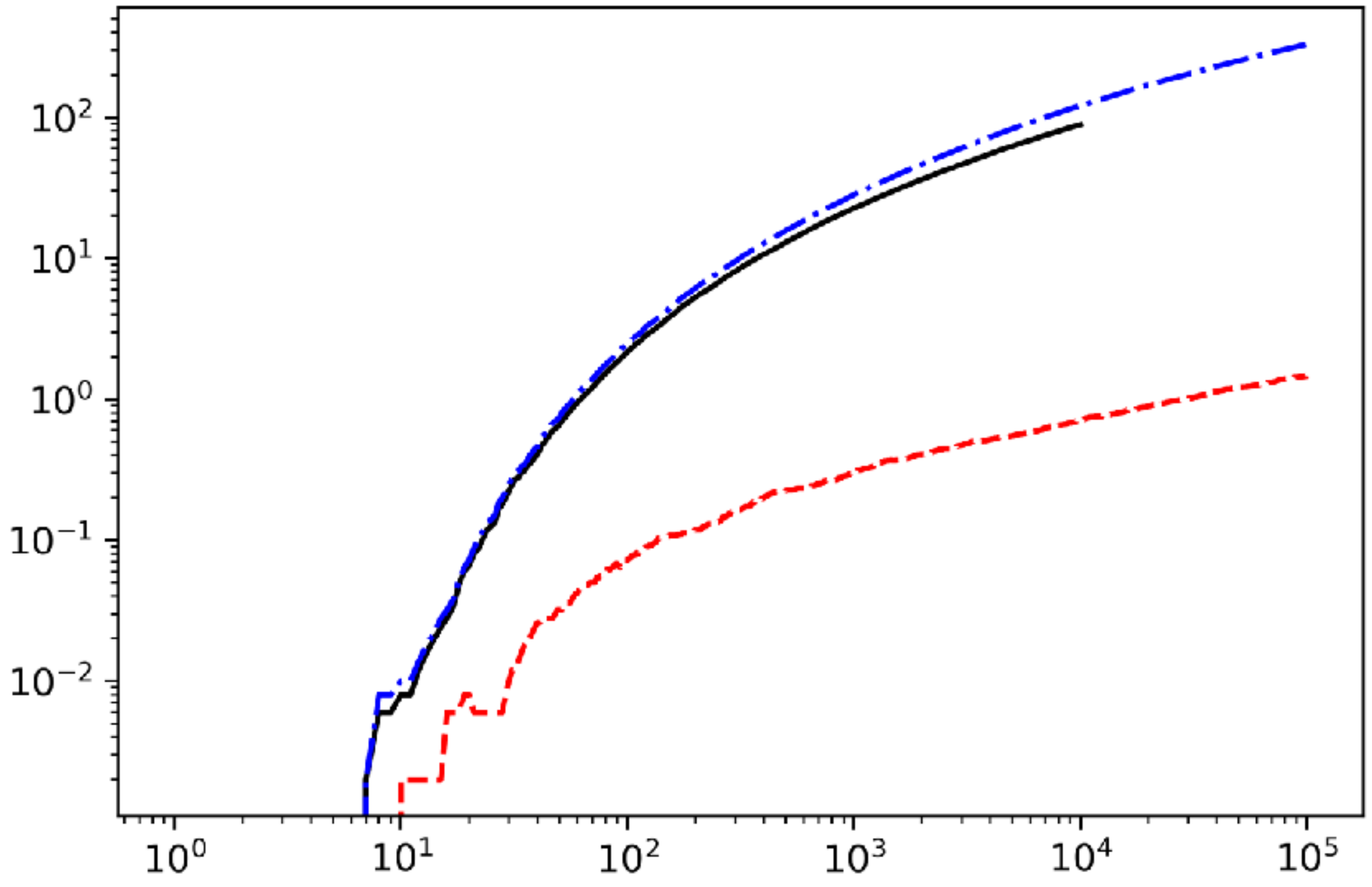
$E[\beta_2] \approx$ **num of nodes**$^{1-4x}$



$$E[\beta_2] \approx \textbf{num of nodes}^{1-4x}$$

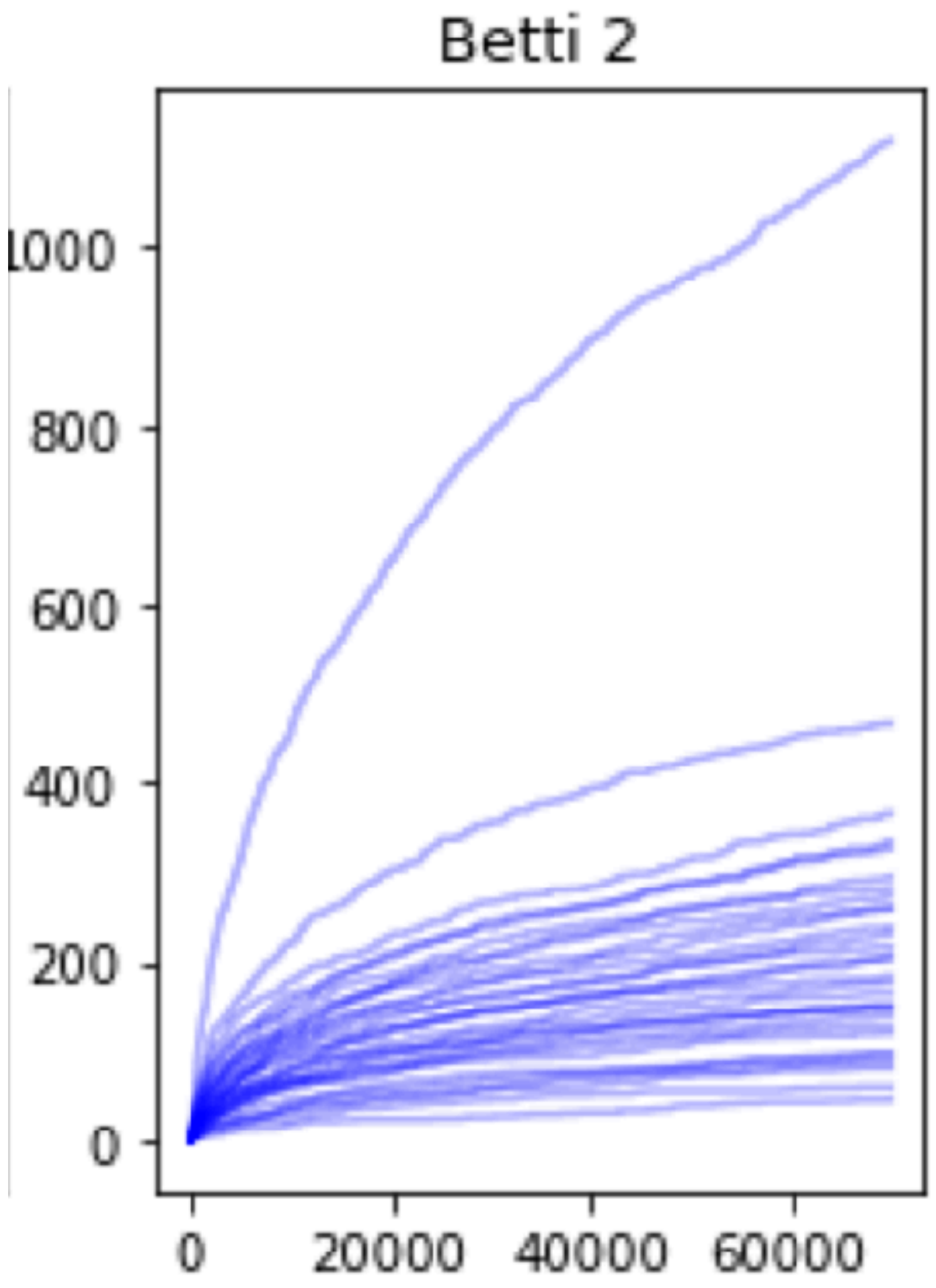
$$\log E[\beta_2] \approx (1-4x)\log(\textbf{num of nodes})$$

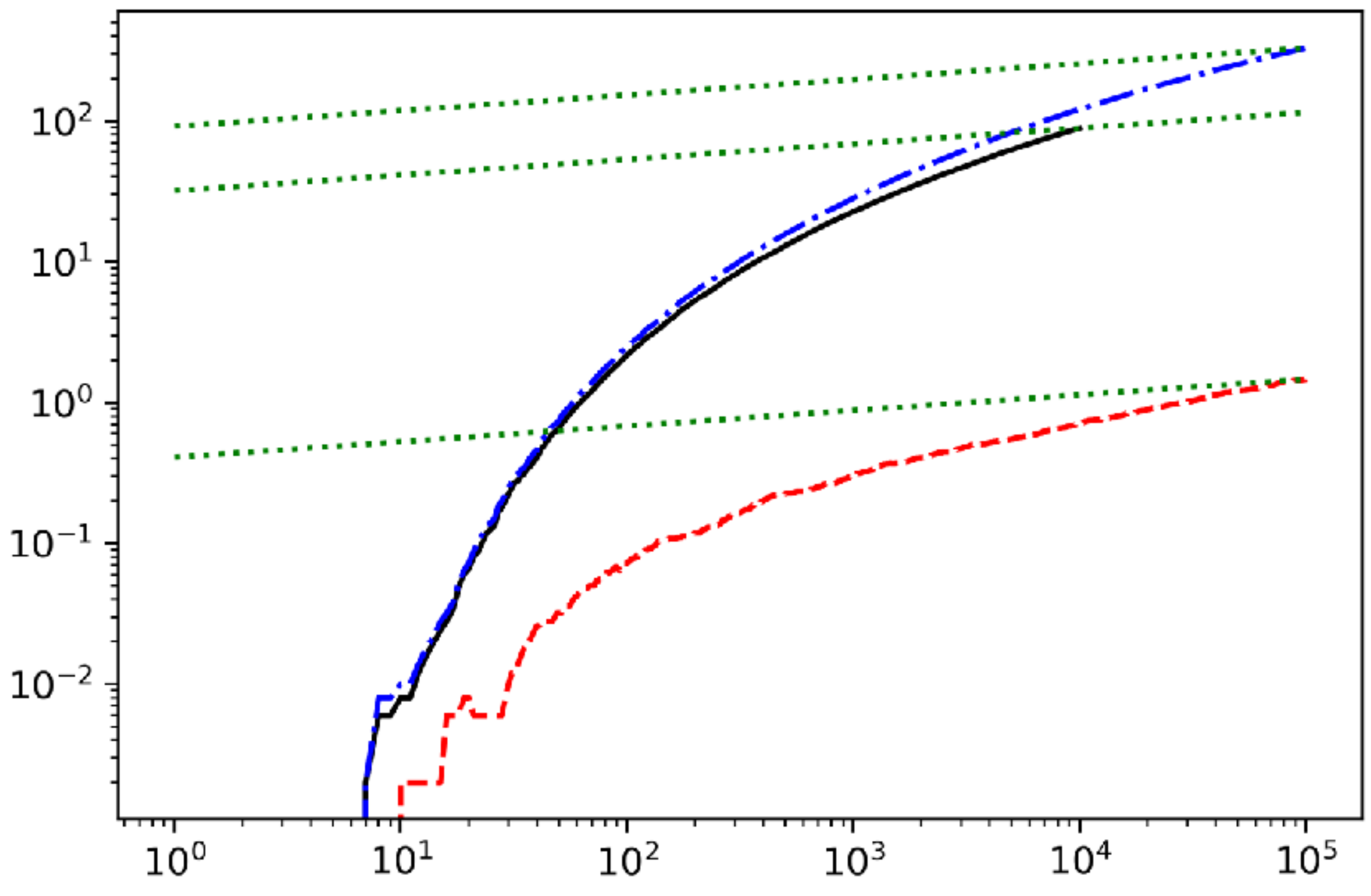




$$E[\beta_2] \approx \textbf{num of nodes}^{1-4x}$$

$$\log E[\beta_2] \approx (1-4x)\log(\textbf{num of nodes})$$





V. What lies ahead

order of magnitude of expected Betti numbers

homotopy connectedness of the infinite complex?

order of magnitude of expected Betti numbers

parameter estimation?

homotopy connectedness of the infinite complex?

order of magnitude of expected Betti numbers

parameter estimation?

homotopy connectedness of the infinite complex?

order of magnitude of expected Betti numbers

simplicial preferential attachment?

parameter estimation?

homotopy connectedness of the infinite complex?

order of magnitude of expected Betti numbers

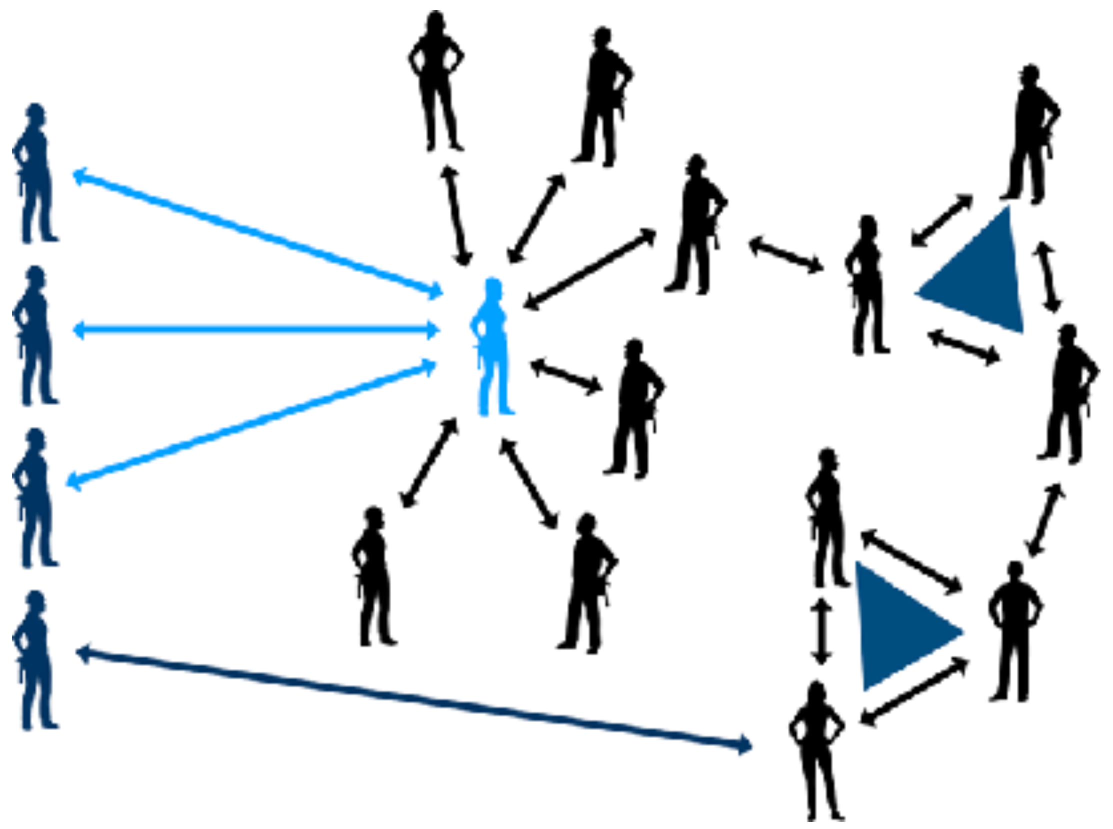
simplicial preferential attachment?

other non-homogeneous complexes?

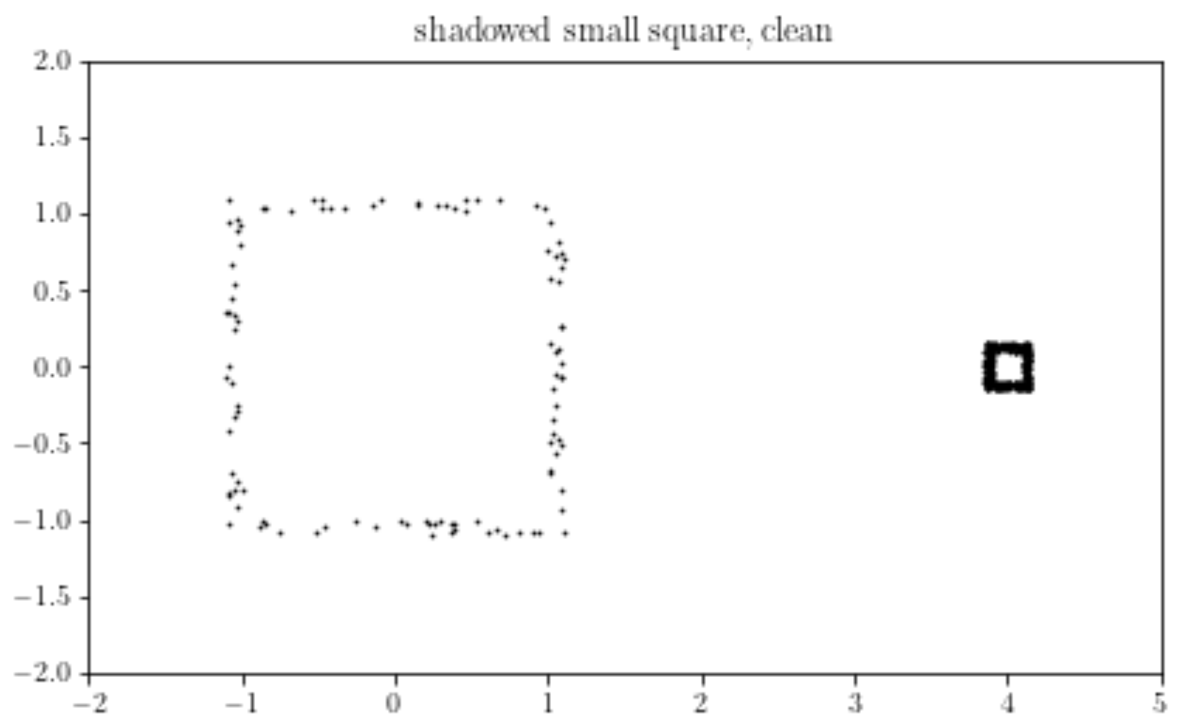
What did we learn today?

- Random topology is cool.
- Preferential attachment graph has interesting topology.
- More interesting things are waiting to be discovered.

Chunyin Siu cs2323@cornell.edu
Cornell University



arxiv paper

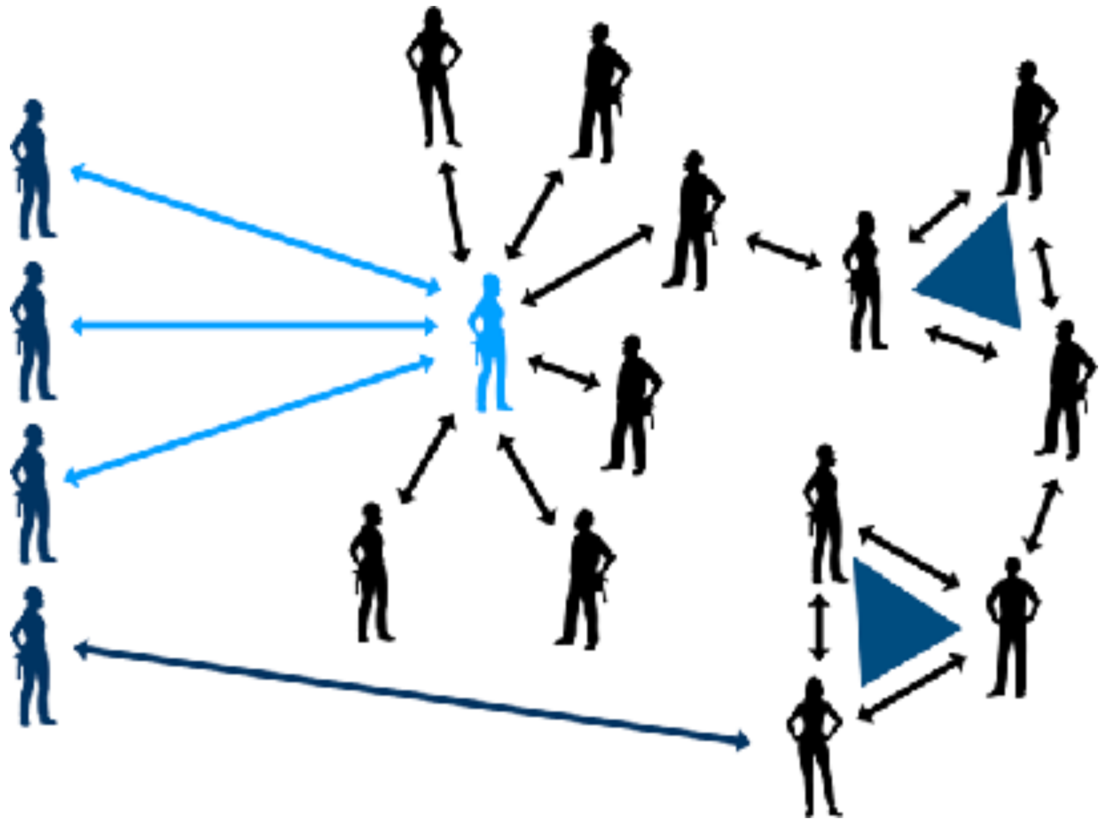


my video about small holes

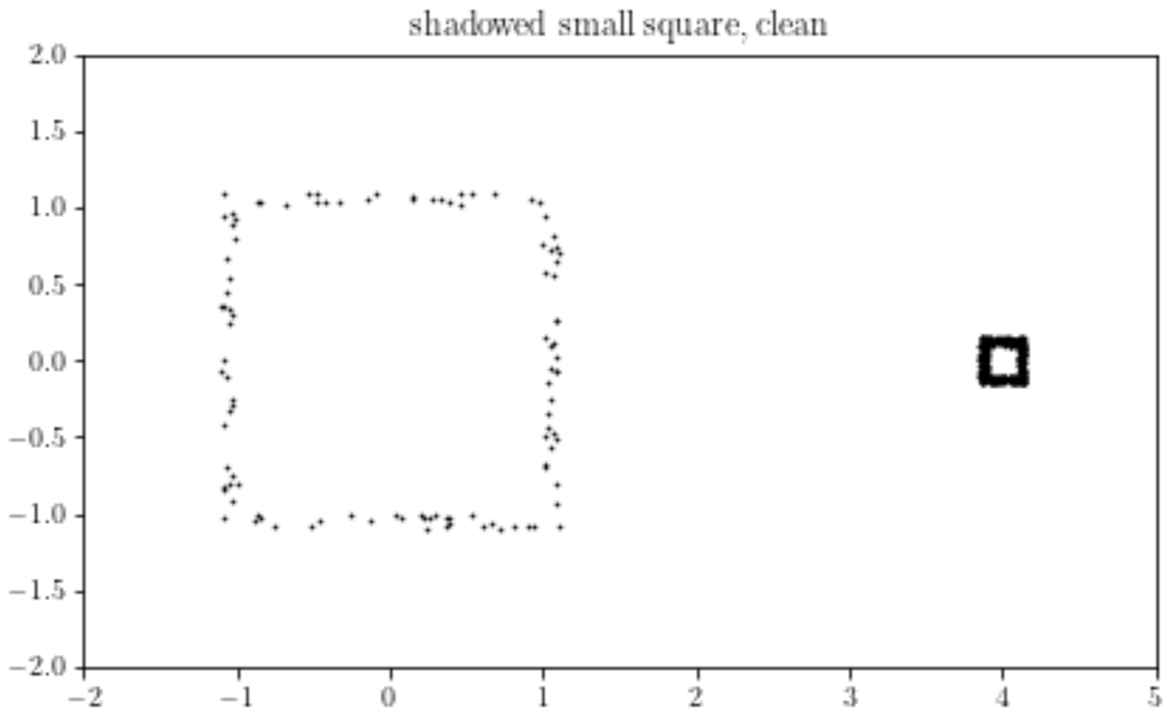
Thank you!

Chunyin Siu
Cornell University

c-siu.github.io
cs2323@cornell.edu



arxiv paper



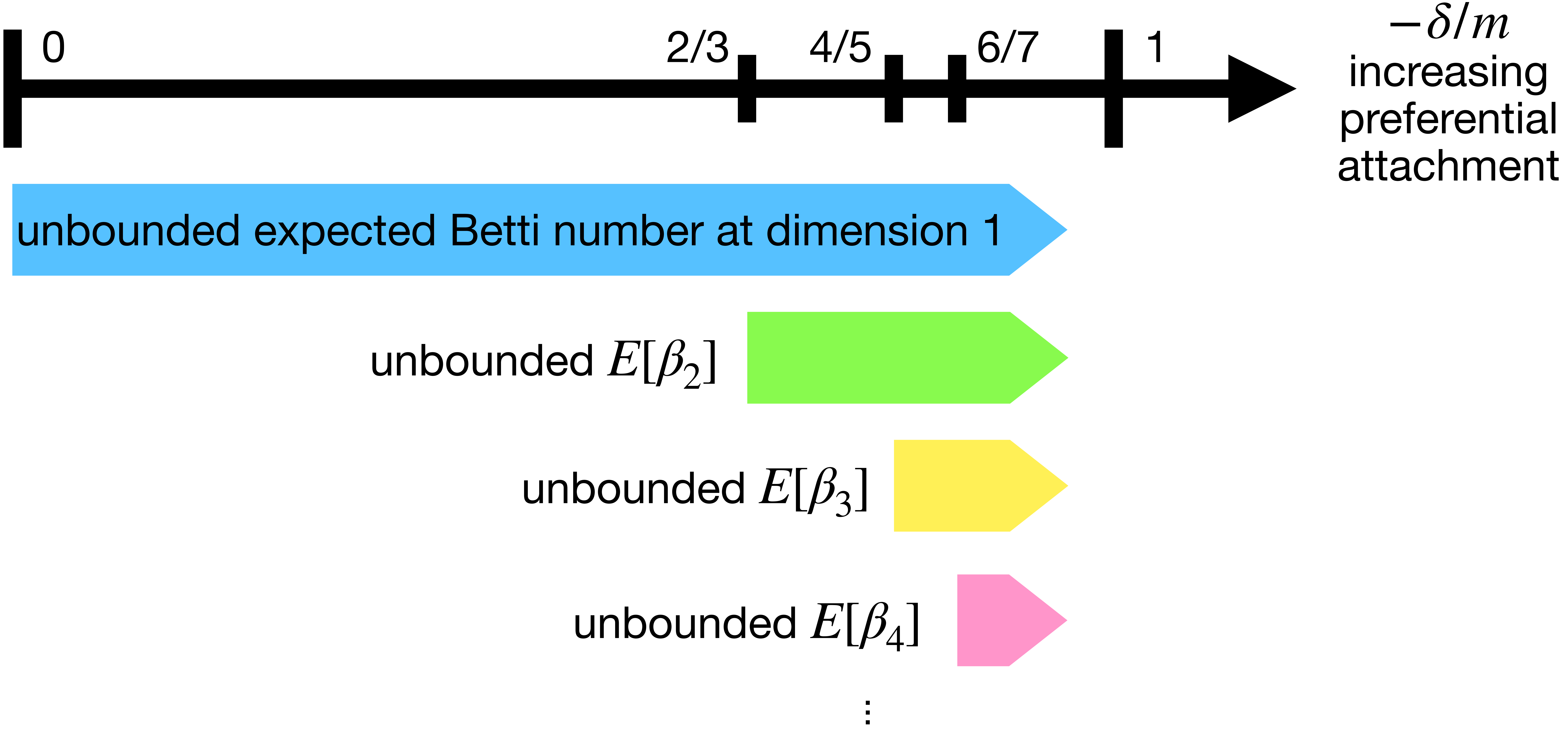
my video about small holes

Phase transition

Recall

P(attaching to v) ∝ degree + δ

m = number of edges per new node

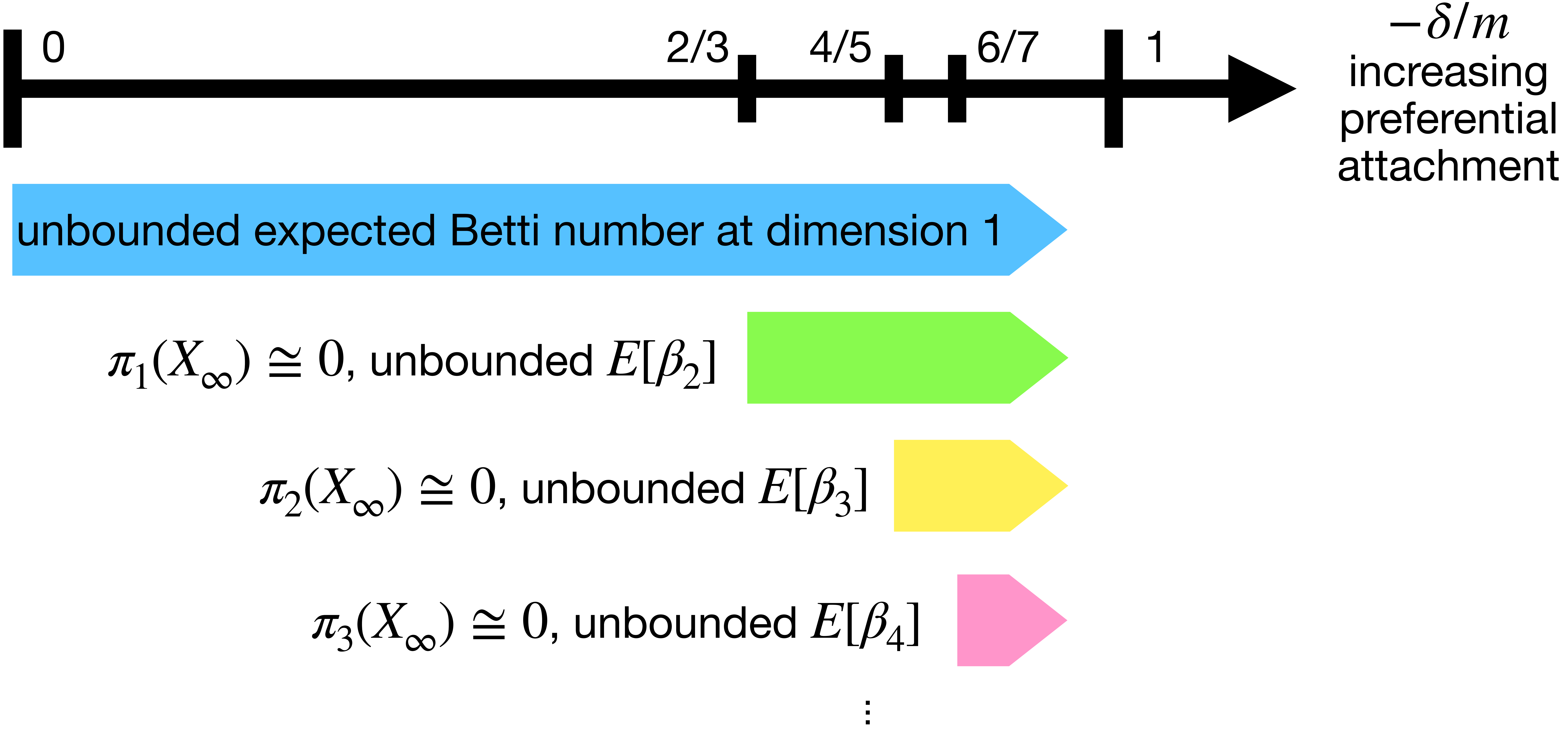


Phase transition

Recall

P(attaching to v) $\propto$ degree + δ

m = number of edges per new node



Subtleties

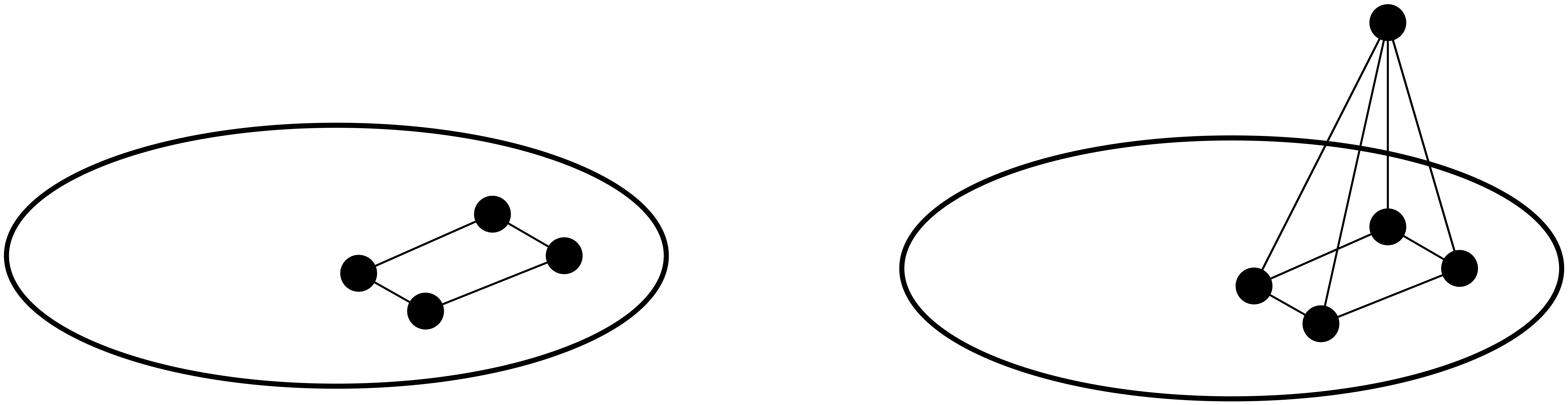
- Need homological algebra to relate Betti numbers with counts

Subtleties

- Need homological algebra to relate Betti numbers with counts
  - adding a vertex = construct mapping cone

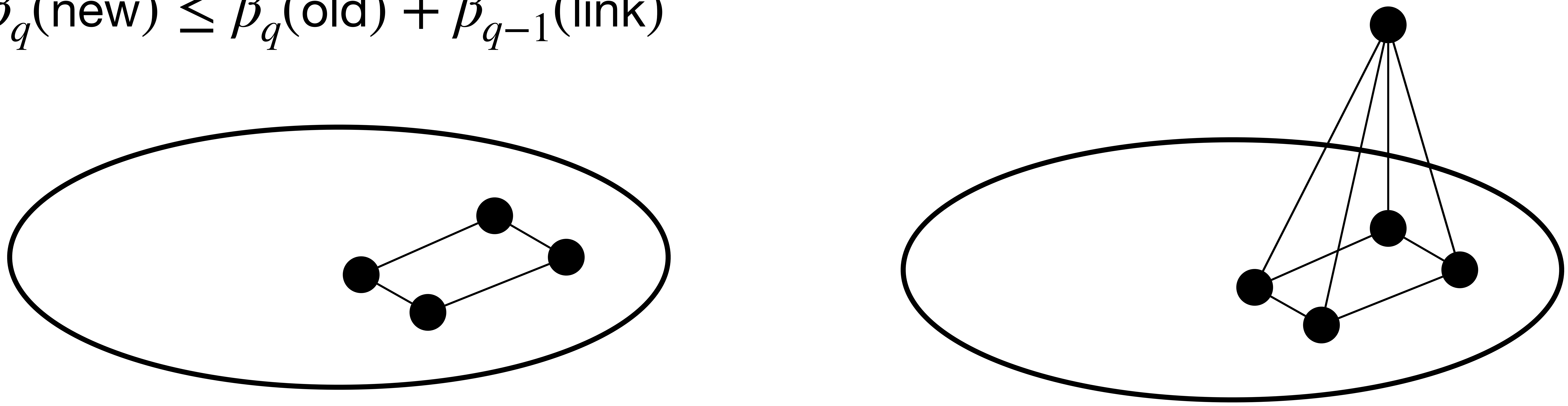
Subtleties

- Need homological algebra to relate Betti numbers with counts
  - adding a vertex = construct mapping cone



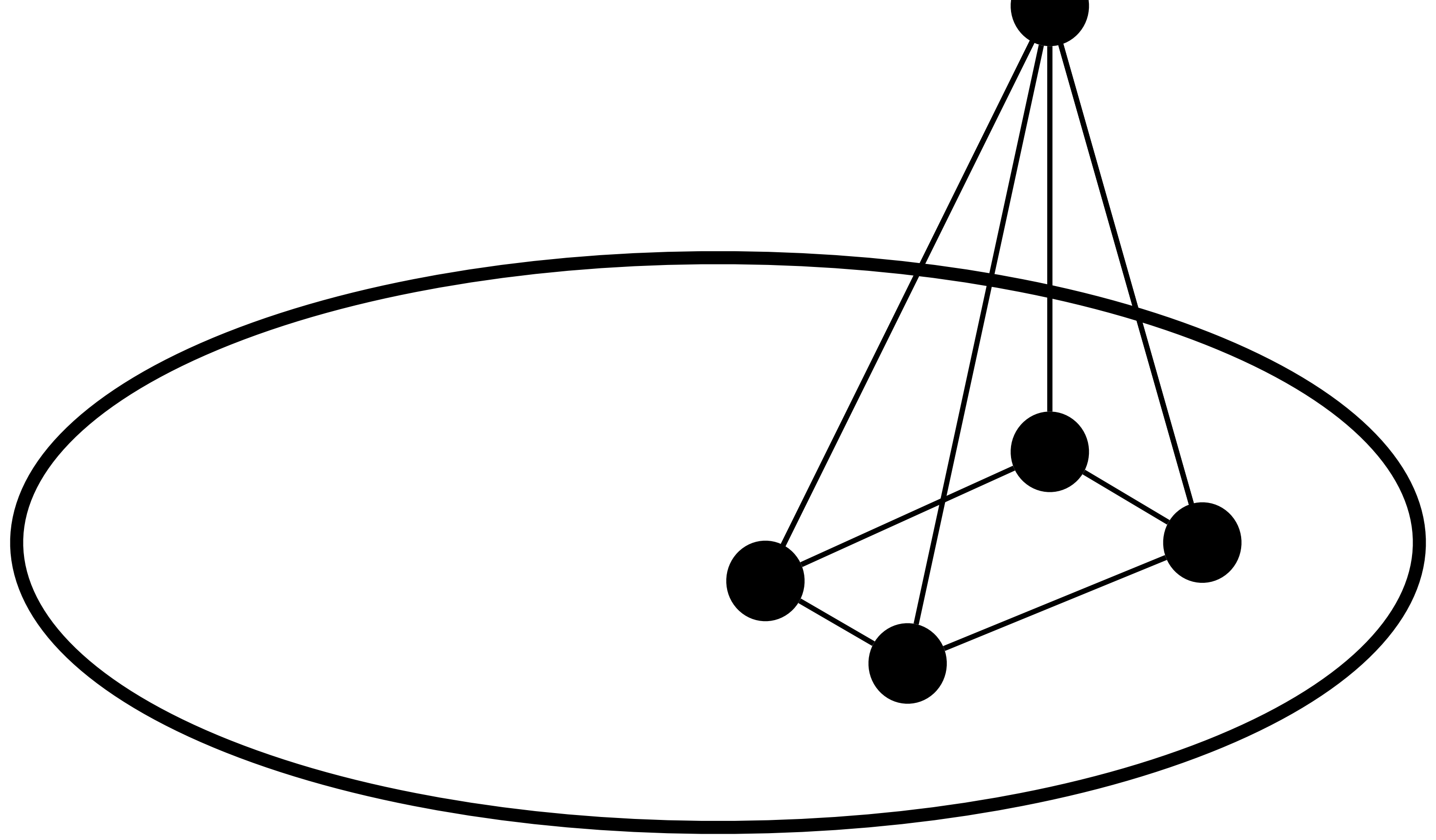
Subtleties

- Need homological algebra to relate Betti numbers with counts
  - adding a vertex = construct mapping cone
  - $\beta_q(\text{new}) \leq \beta_q(\text{old}) + \beta_{q-1}(\text{link})$



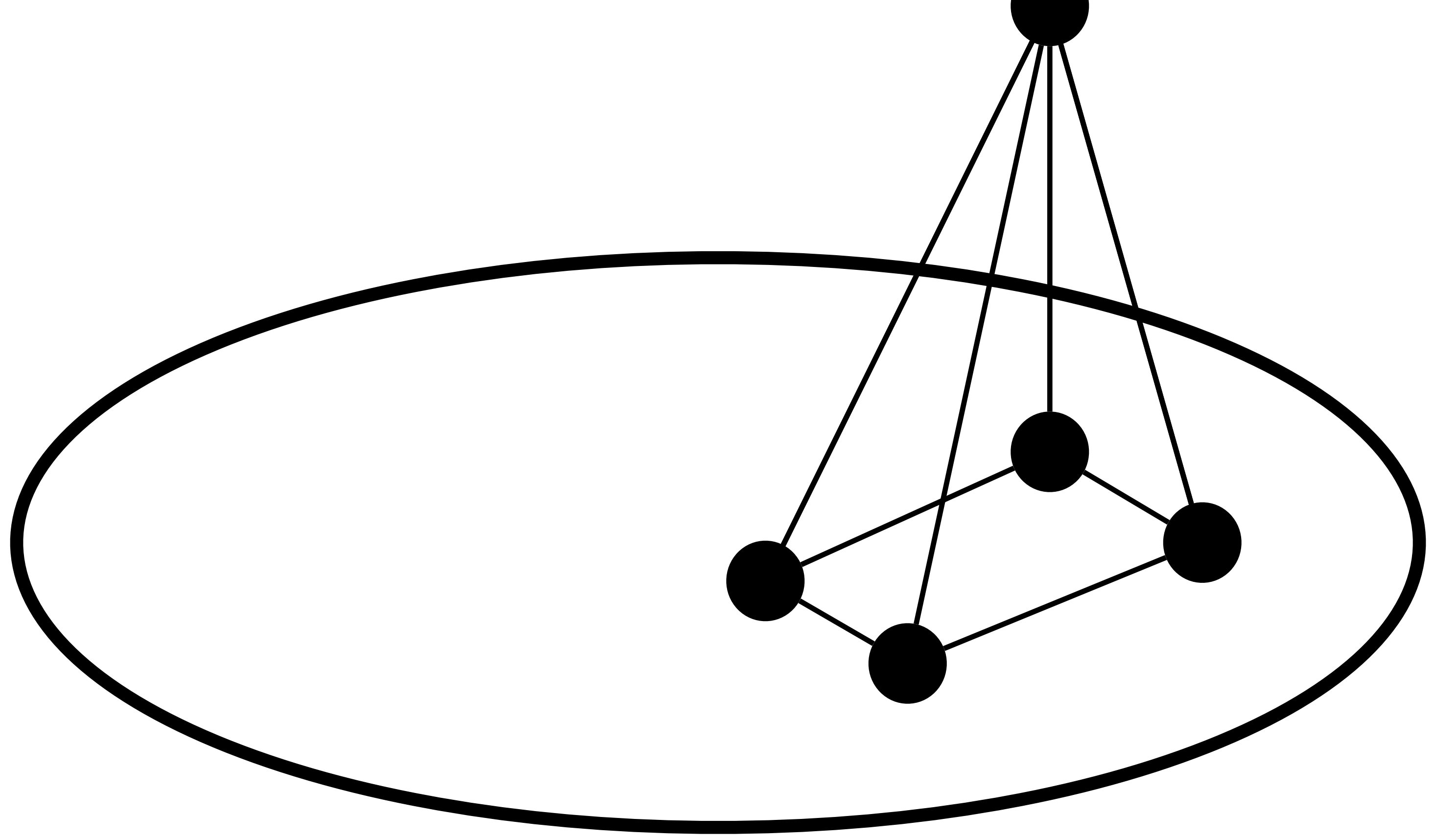
Subtleties

- Need homological algebra to relate Betti numbers with counts
  - $\beta_q(\text{new}) - \beta_q(\text{old}) \leq \beta_{q-1}(\text{link})$



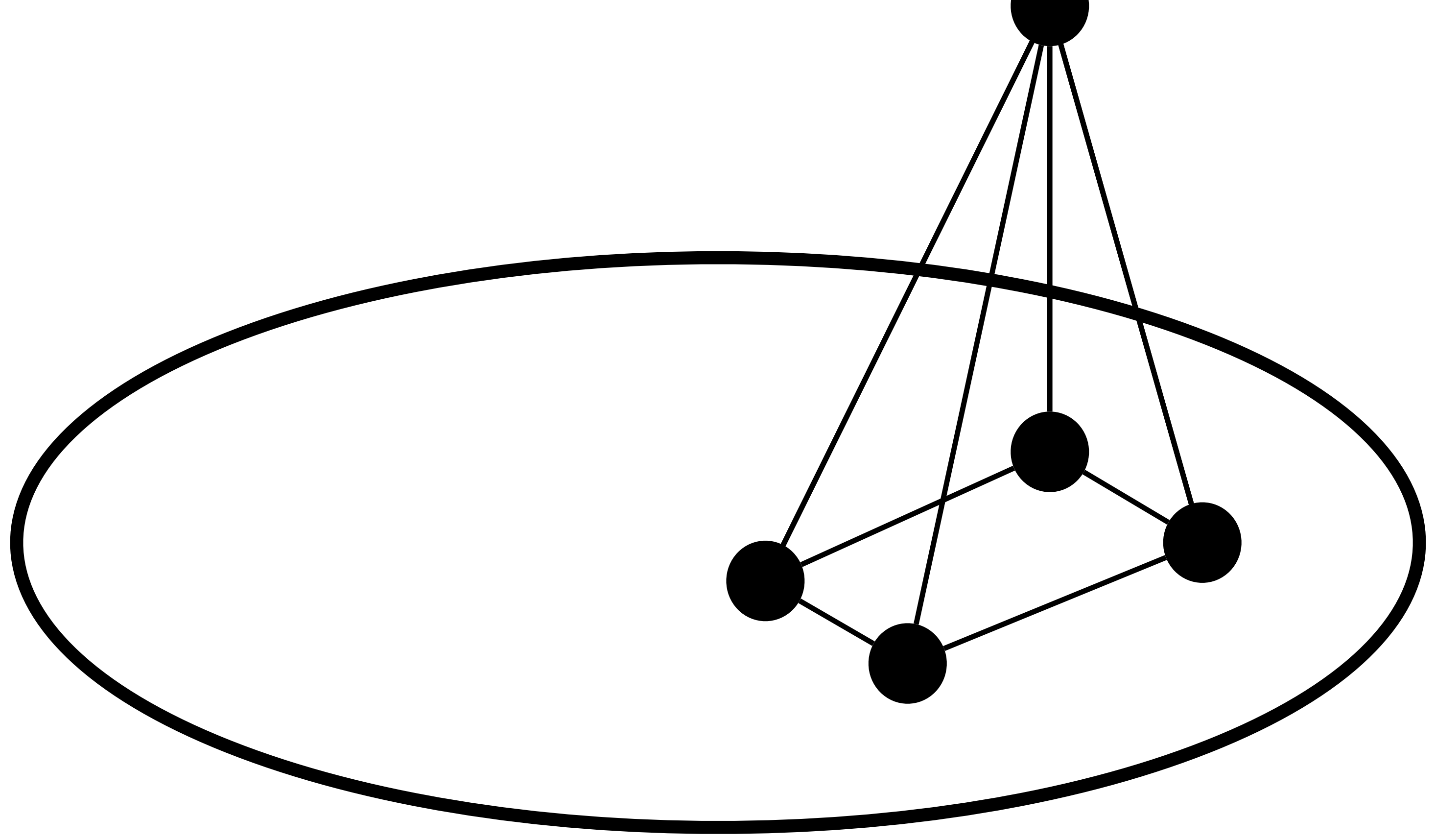
Subtleties

- Need homological algebra to relate Betti numbers with counts
    - $\beta_q(\text{new}) - \beta_q(\text{old}) \leq \beta_{q-1}(\text{link})$
- Identify the "square count" as the main term with minimal cycle results in [Gal 2005] and [Kahle 2009]

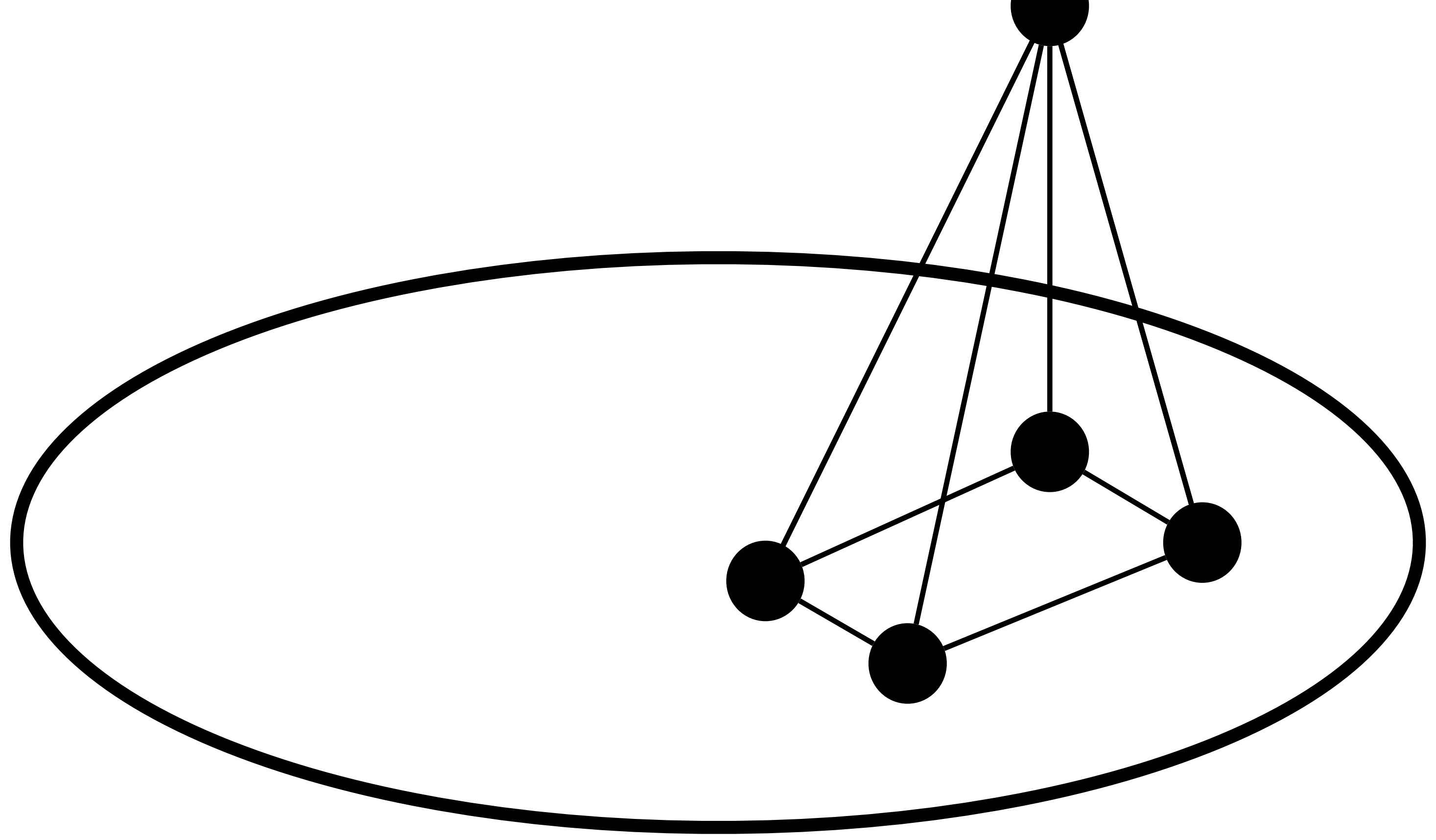


Subtleties

- Need homological algebra to relate Betti numbers with counts
  - $\beta_q(\text{new}) - \beta_q(\text{old}) \leq \beta_{q-1}(\text{link})$
- Identify the "square count" as the main term with minimal cycle results in [Gal 2005] and [Kahle 2009]
- Generalize minimal cycle results with homological algebra



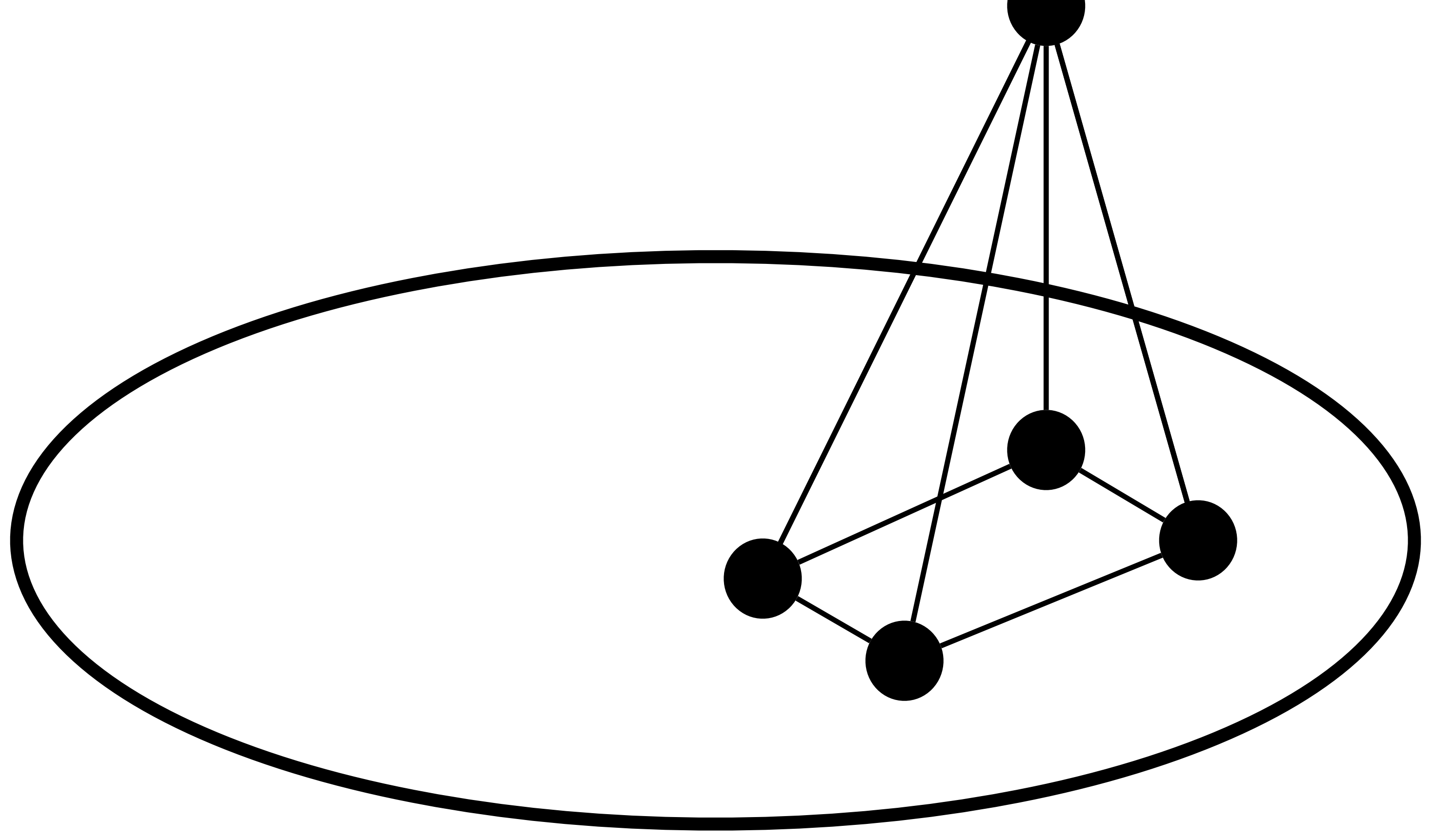
Subtleties

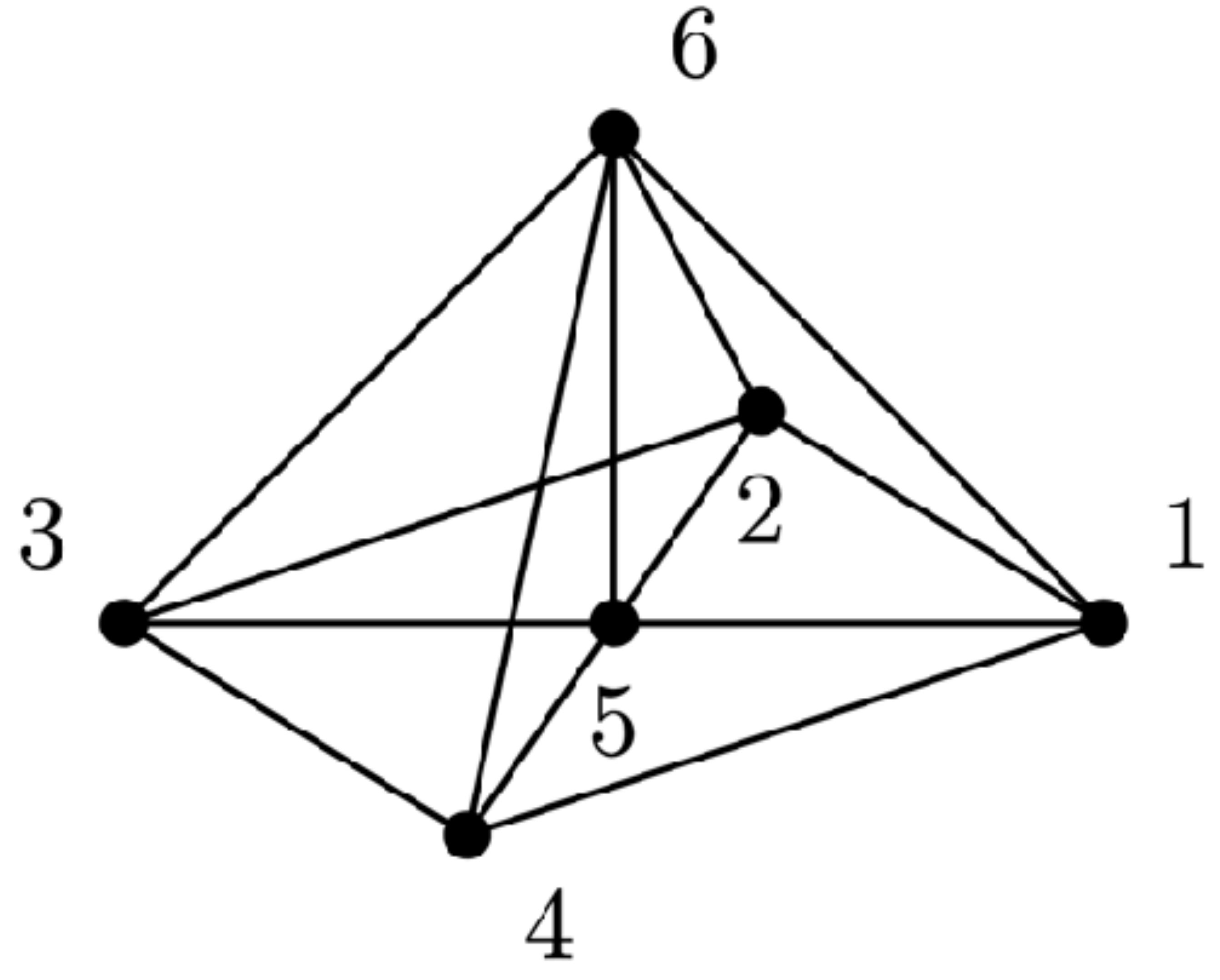


- Need homological algebra to relate Betti numbers with counts
  - $\beta_q(\text{new}) - \beta_q(\text{old}) \leq \beta_{q-1}(\text{link})$
- Identify the "square count" as the main term with minimal cycle results in [Gal 2005] and [Kahle 2009]
- Generalize minimal cycle results with homological algebra
  - $1 - \beta_q(\text{link}, S^{q-1}) - \beta_q(\text{link}) \leq \beta_q(\text{new}) - \beta_q(\text{old}) \leq \beta_{q-1}(\text{link})$

Subtleties

- Need homological algebra to relate Betti numbers with counts
  - $\beta_q(\text{new}) - \beta_q(\text{old}) \leq \beta_{q-1}(\text{link})$
- Identify the "square count" as the main term with minimal cycle results in [Gal 2005] and [Kahle 2009]
- Generalize minimal cycle results with homological algebra
  - $1 - \beta_q(\text{link}, S^{q-1}) - \beta_q(\text{link}) \leq \beta_q(\text{new}) - \beta_q(\text{old}) \leq \beta_{q-1}(\text{link})$





Subtleties

- Need homological algebra to relate Betti numbers with counts
  - $\beta_q(\text{new}) - \beta_q(\text{old}) \leq \beta_{q-1}(\text{link})$
- Identify the “square count” as the main term with minimal cycle results in [Gal 2005] and [Kahle 2009]
- Generalize minimal cycle results with homological algebra
  - $1 - \beta_q(\text{link}, S^{q-1}) - \beta_q(\text{link}) \leq \beta_q(\text{new}) - \beta_q(\text{old}) \leq \beta_{q-1}(\text{link})$
- Apply graph counting result in [Garavaglia and Stegehuis 2019] on a large class of subgraphs