The Topology of Preferential Attachment

Higher-Order Connectivity of Random Interactions

Chunyin Siu
Cornell University
cs2323@cornell.edu

Probabilists

Statisticians

Network Scientists

Topologist

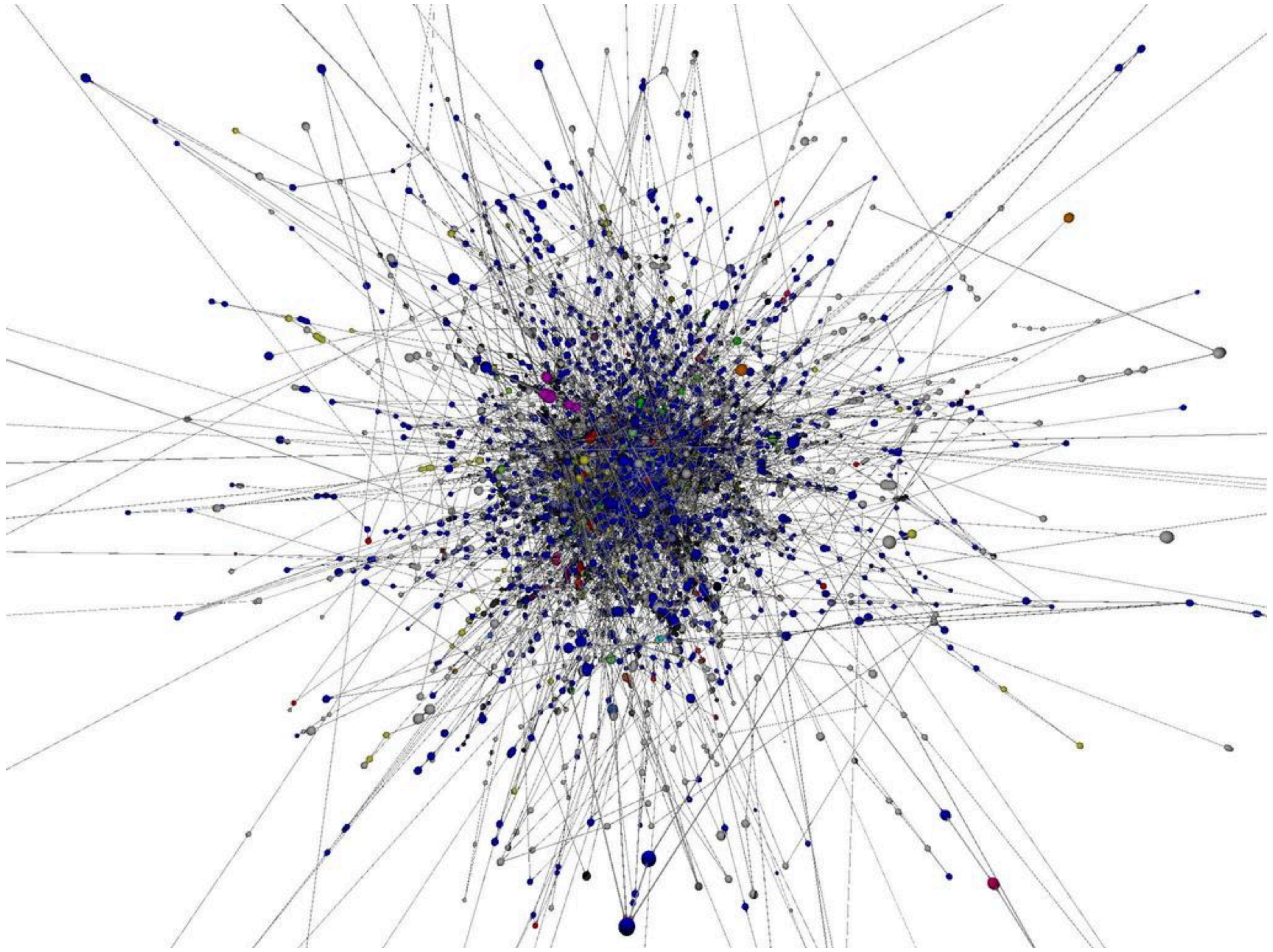
The Topology of Preferential Attachment

Higher-Order Connectivity of Random Interactions

Chunyin Siu
Cornell University
cs2323@cornell.edu

So, preferential attachment…

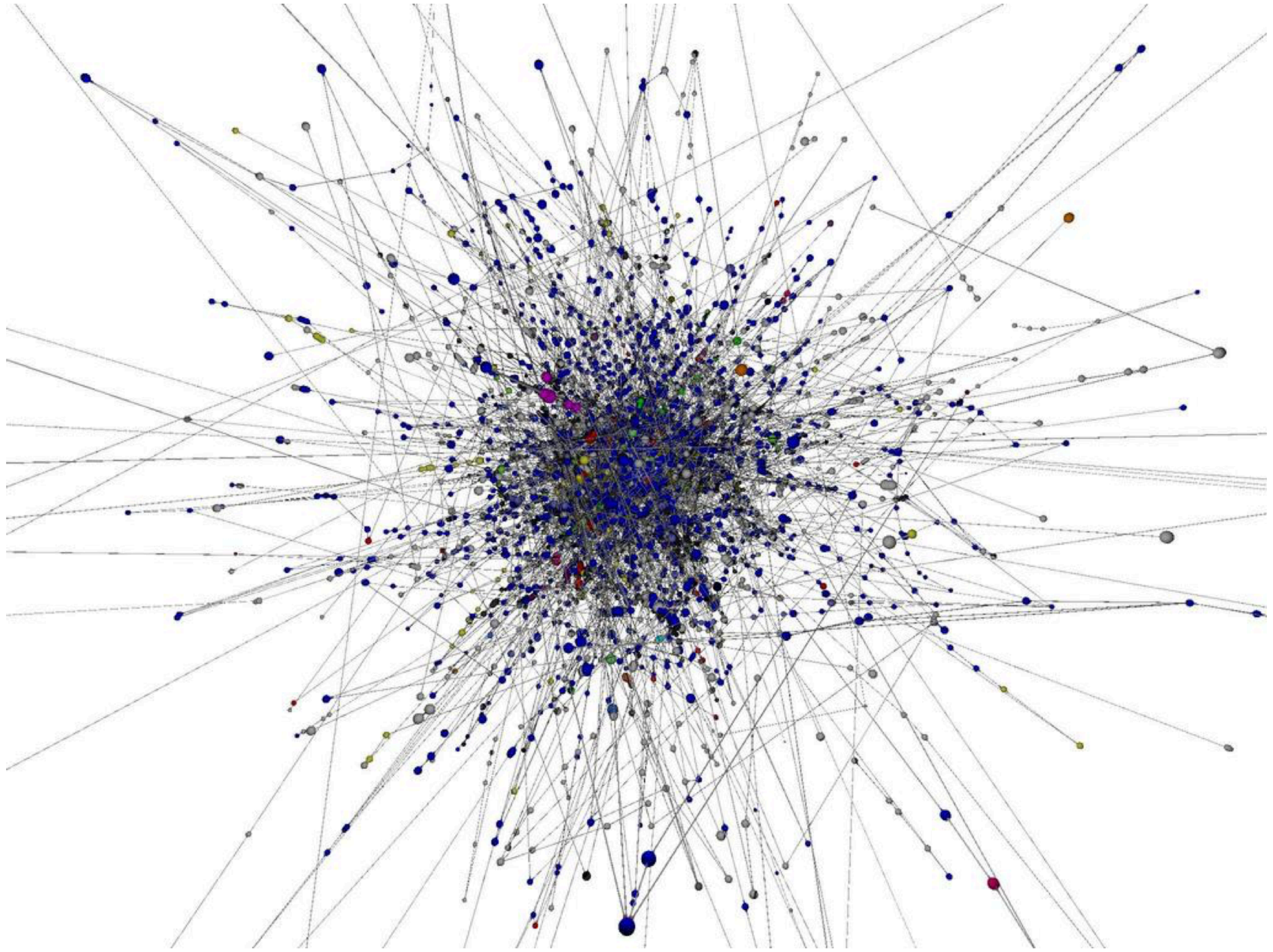
- Highly connected hubs



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

So, preferential attachment…

- Highly connected hubs
- Dense core of hubs?



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

So, preferential attachment…

- Highly connected hubs
- Dense core of hubs?
- Beyond pairwise connections?

- —> topological properties

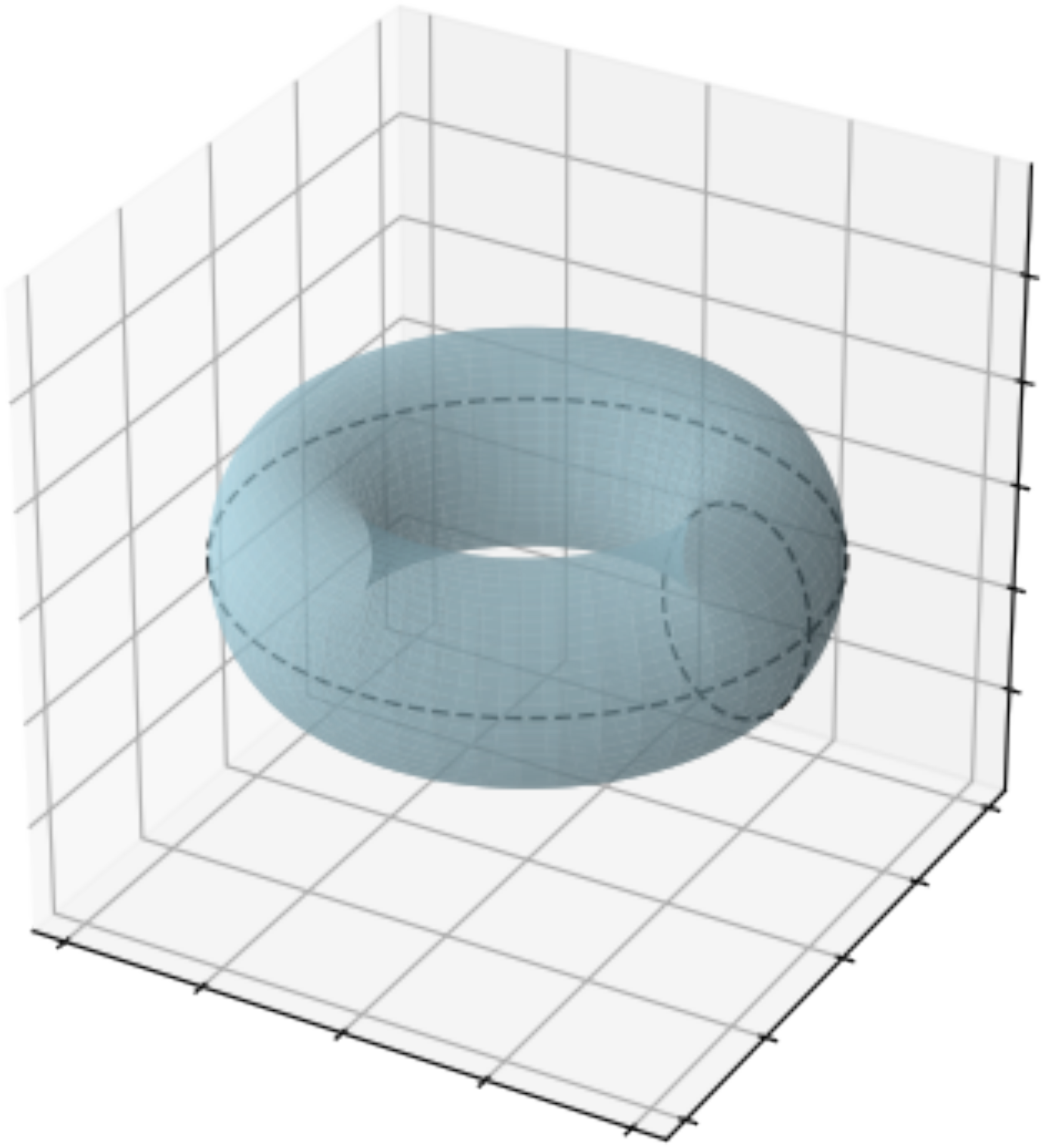
(Stephen Coast
https://www.fractalus.com/steve/stuff/ipmap/)

Agenda

preferential attachment

Agenda

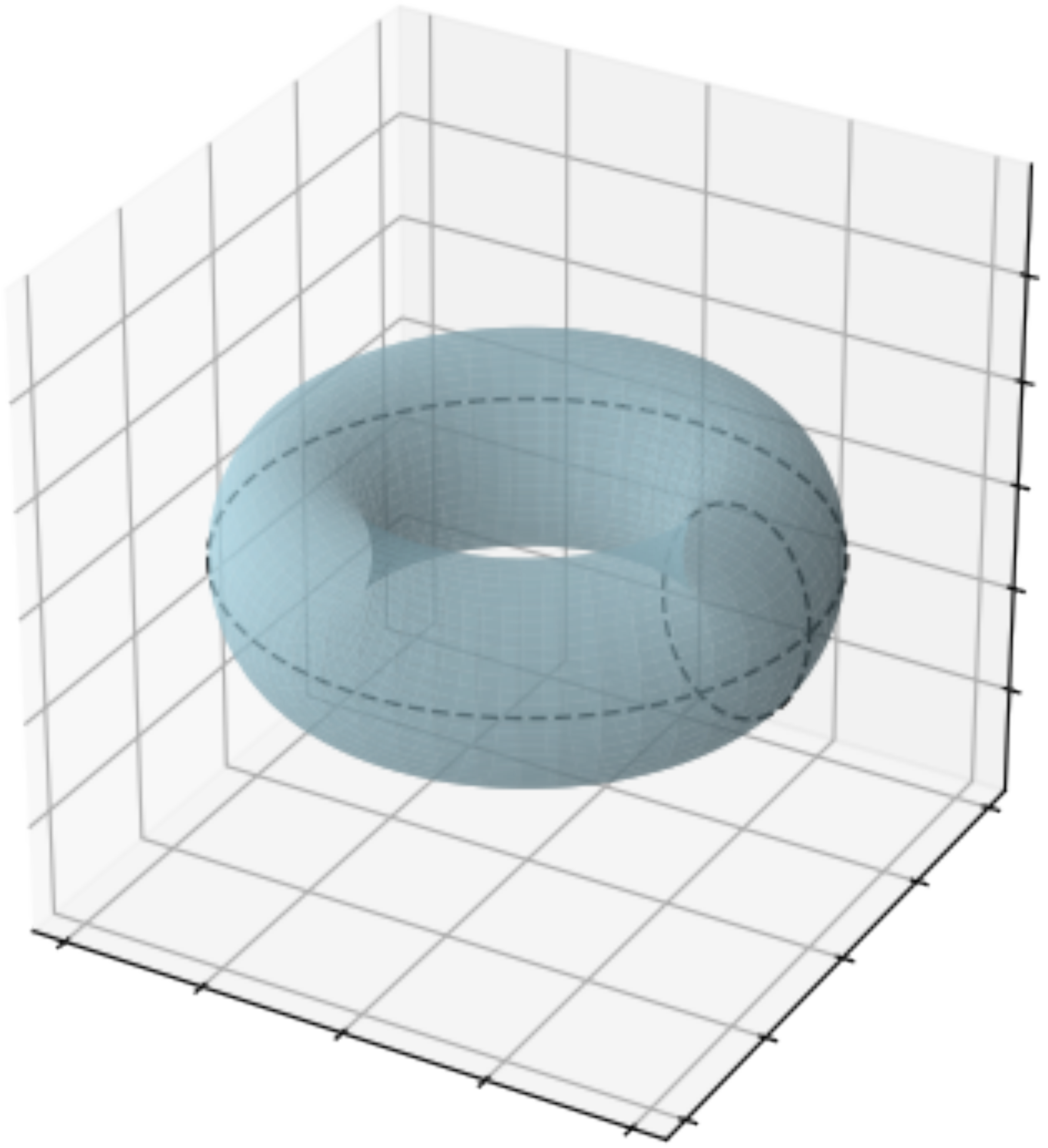
preferential attachment



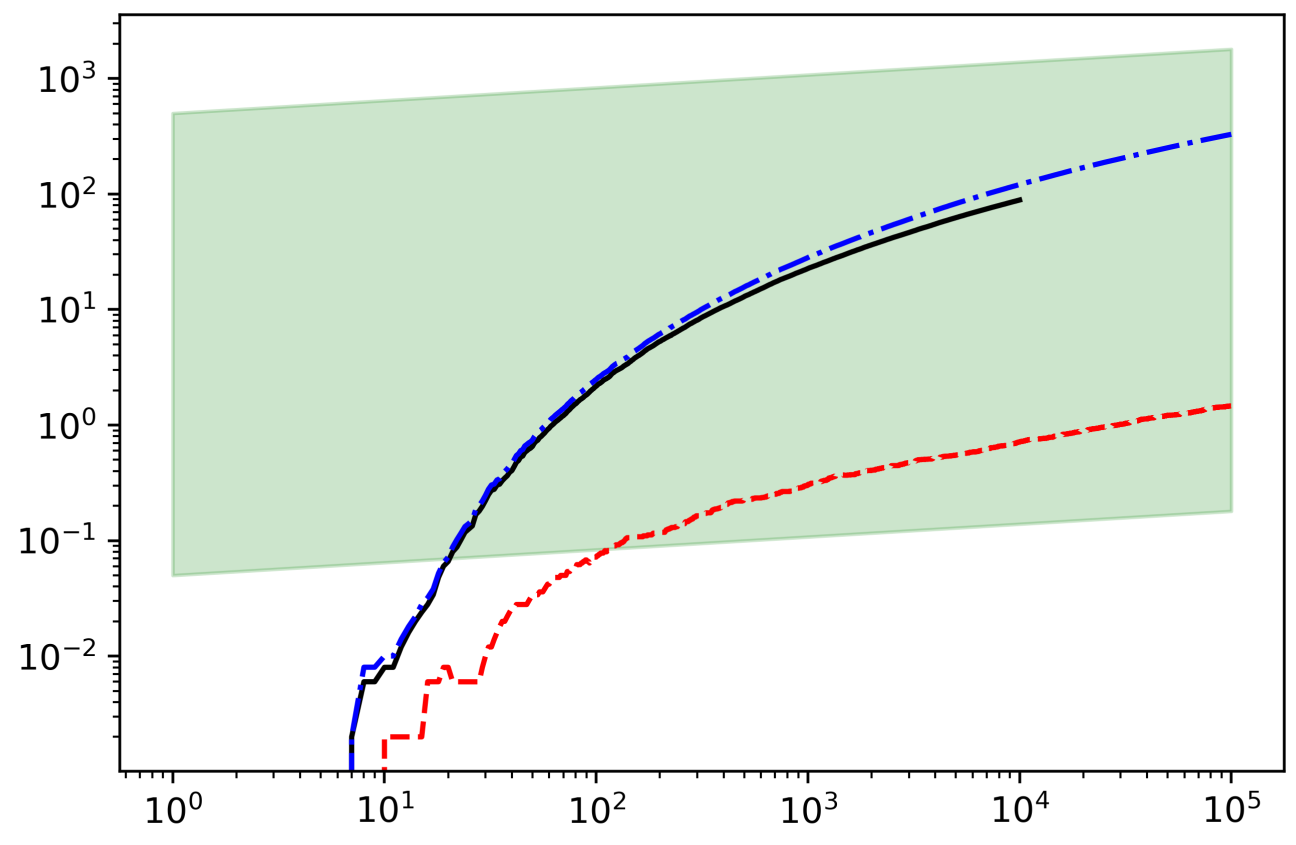
topology

Agenda

preferential attachment



topology

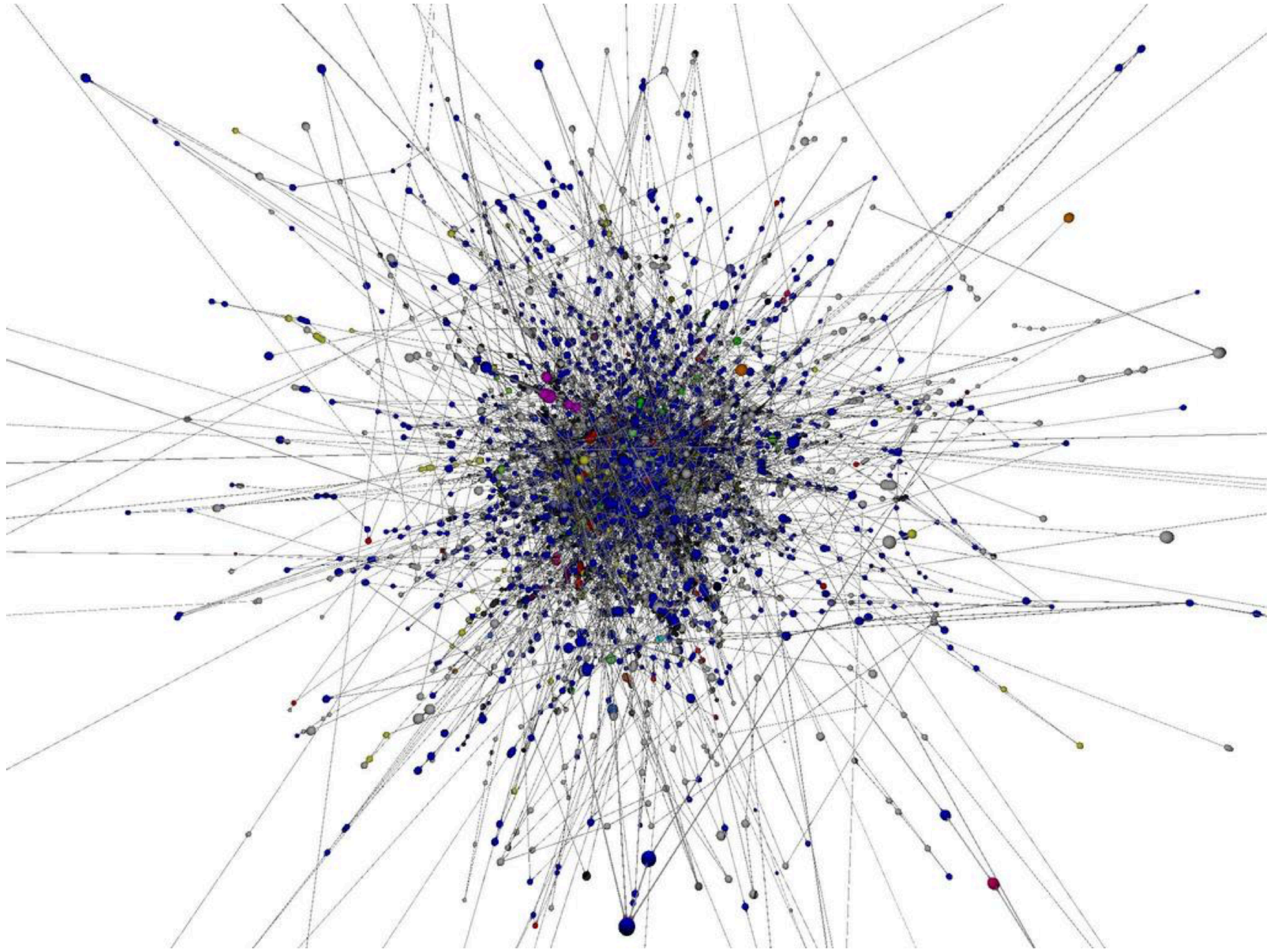


our result

I. Preferential Attachment

Preferential Attachment

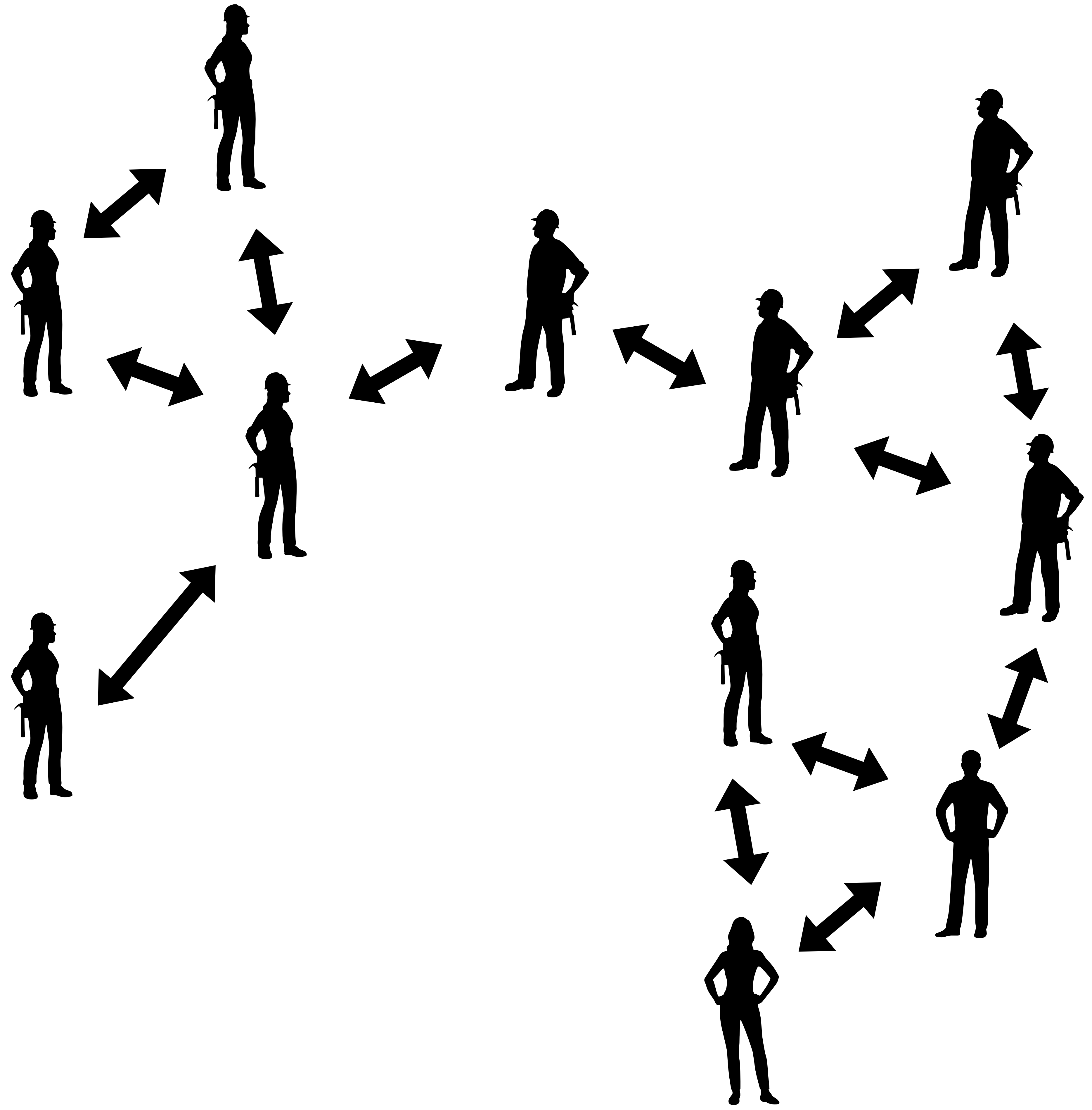
[Albert and Barabasi 1999]



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

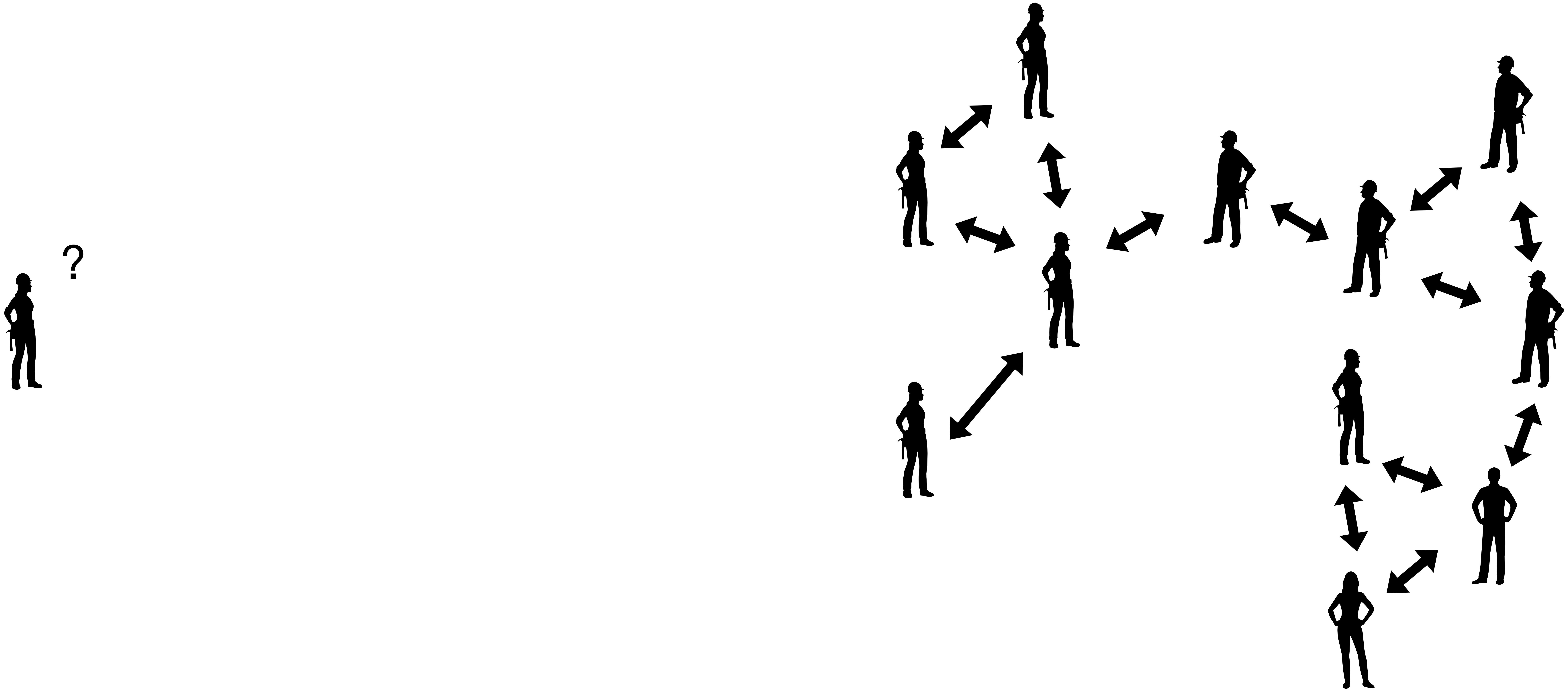
Preferential Attachment

[Albert and Barabasi 1999]



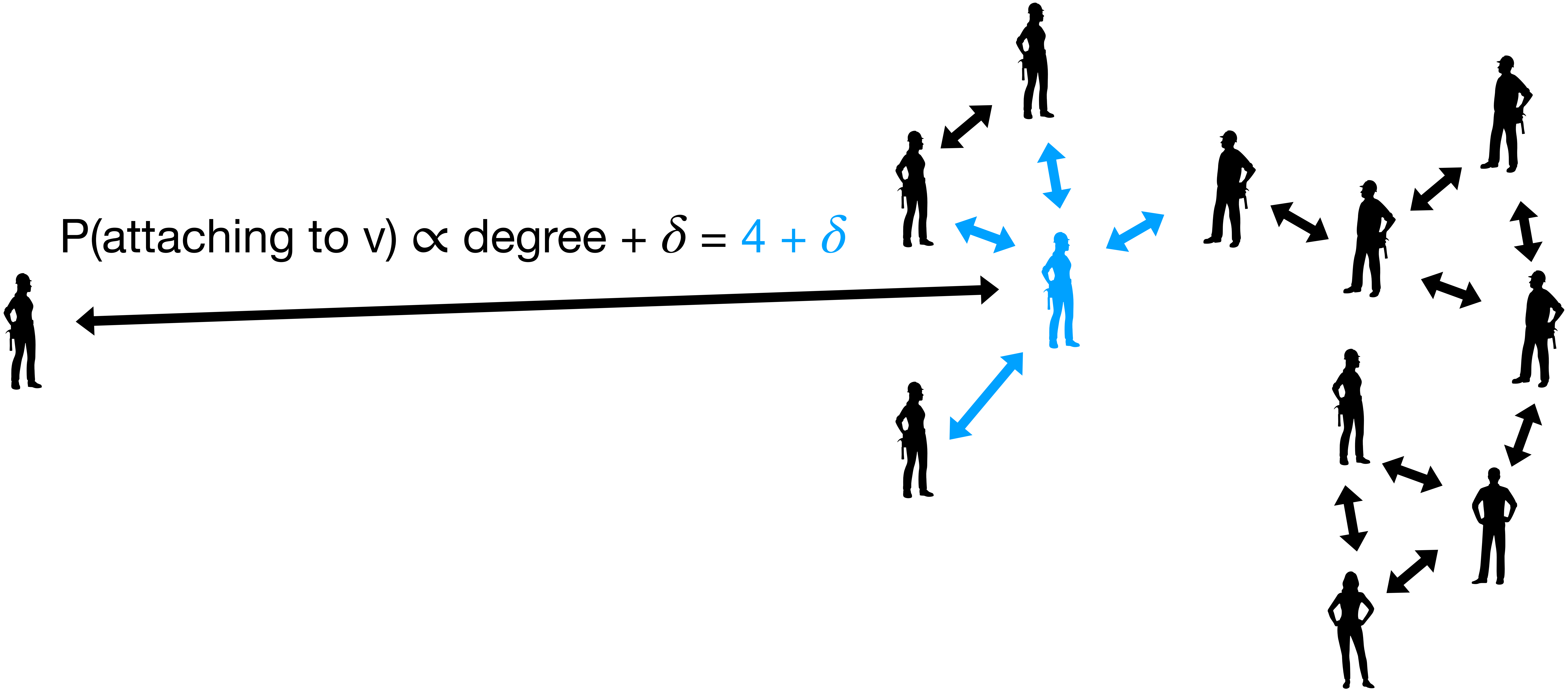
Preferential Attachment

[Albert and Barabasi 1999]



Preferential Attachment

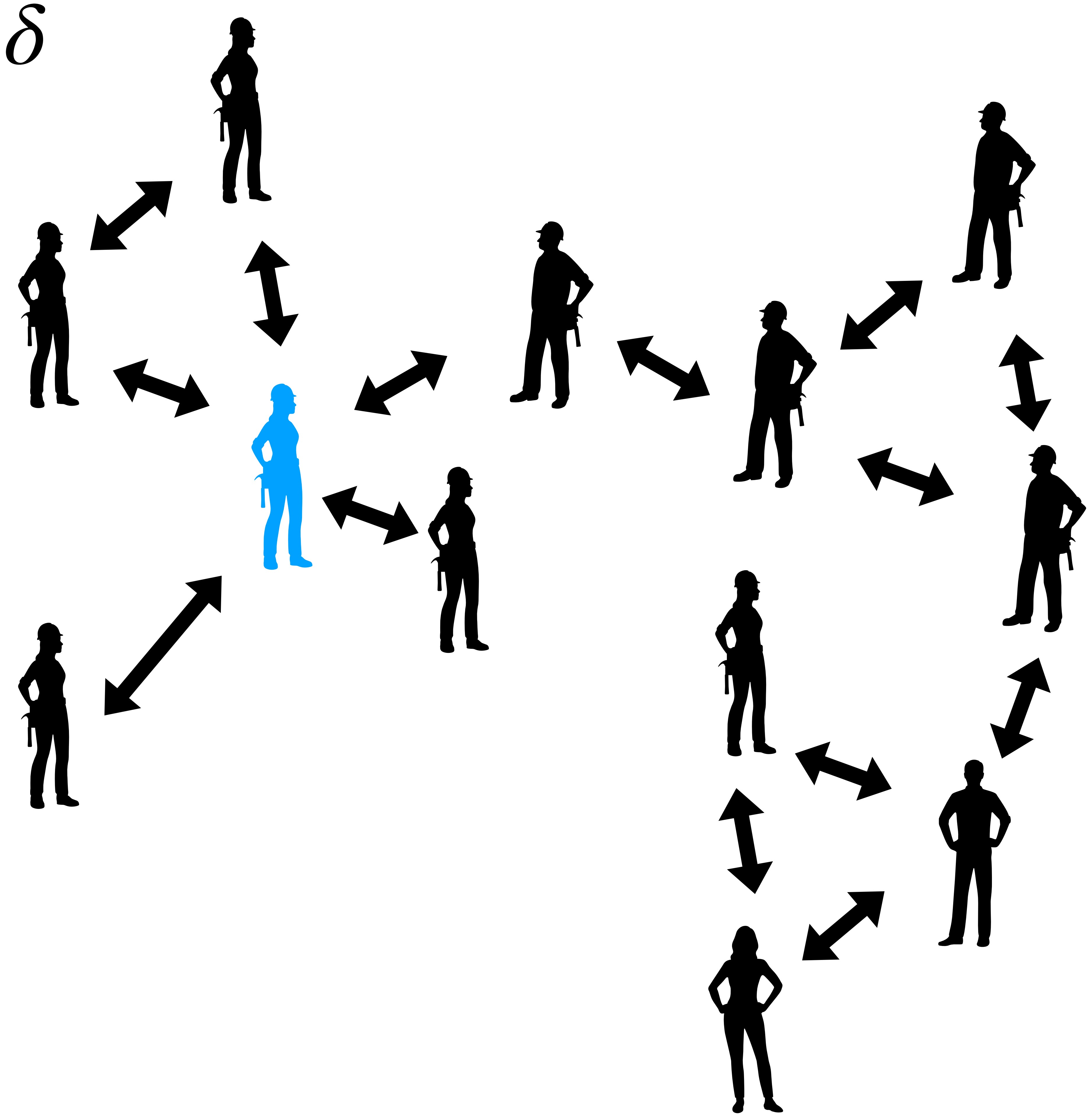
[Albert and Barabasi 1999]



Preferential Attachment

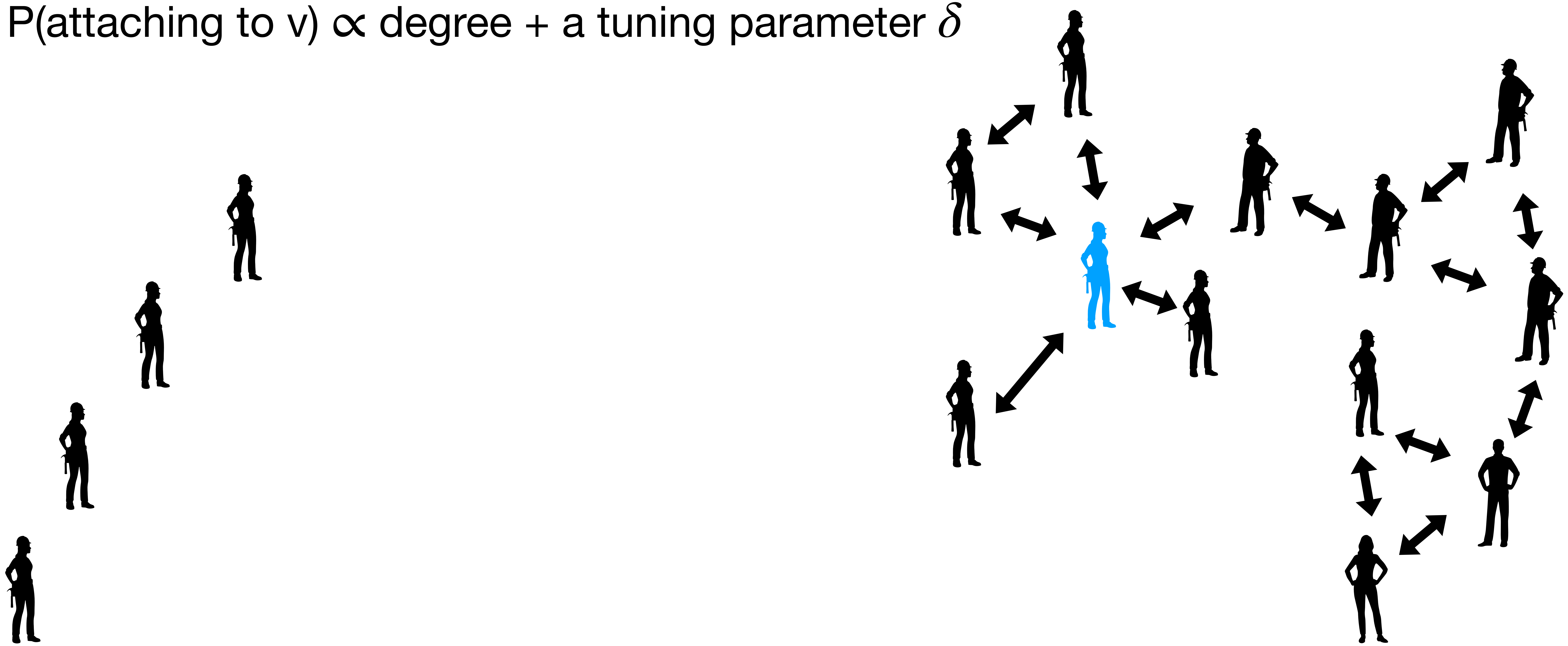
[Albert and Barabasi 1999]

P(attaching to v) ∝ degree + a tuning parameter δ



Preferential Attachment
[Albert and Barabasi 1999]

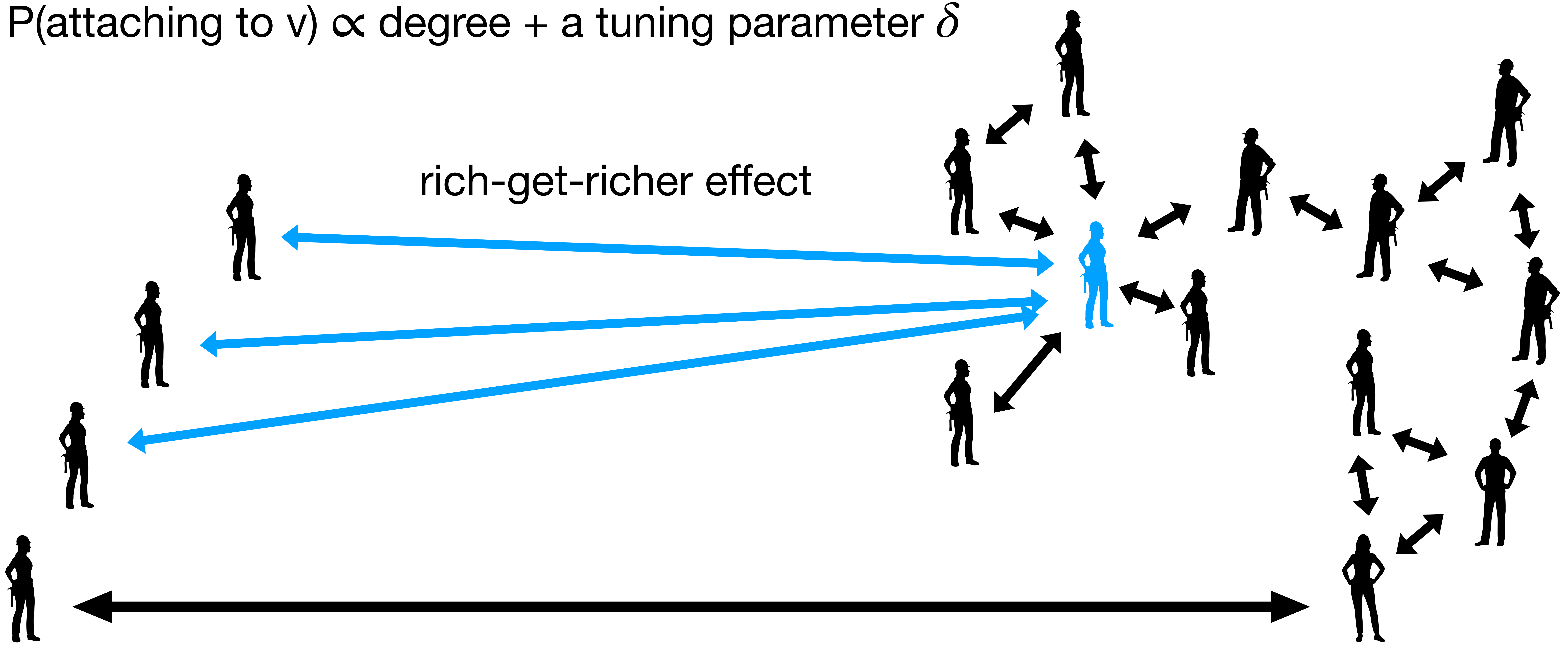
P(attaching to v) $\propto$ degree + a tuning parameter δ



Preferential Attachment

[Albert and Barabasi 1999]

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



What do we know?

What do we know?

- Scale-freeness and Degree distribution
  [Barabasi and Albert 1999; Dorogovtsev, Mendes and Samukhin 2000; Krapivsky, Redner and Leyvraz 2000]

pmf of degree distribution

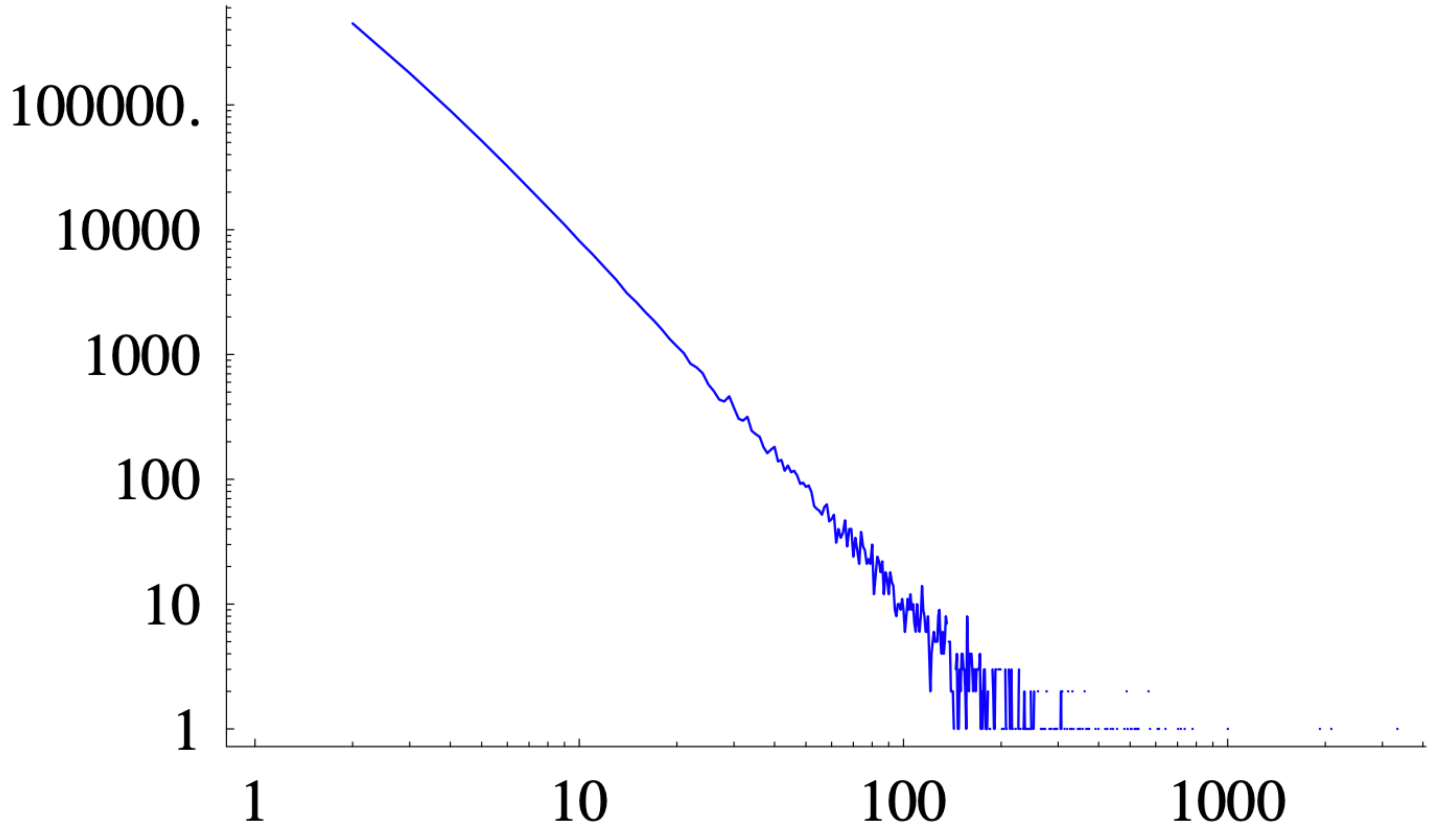


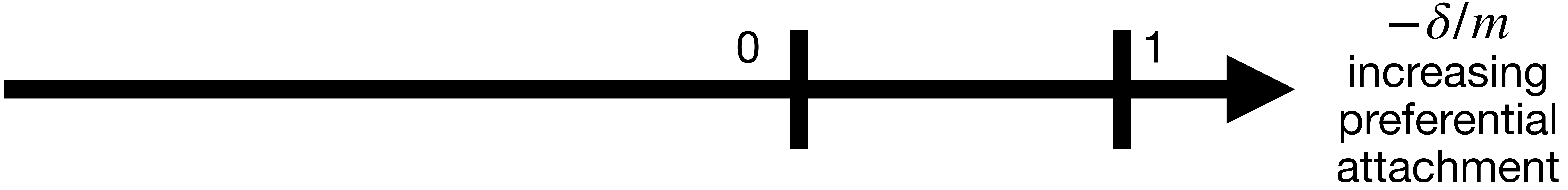
Fig 8.3 of R. Hofstad (2013).
Random Graphs and Complex Networks.
https://doi.org/10.1017/9781316779422

Phase transition

Recall

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

m = number of edges per new node

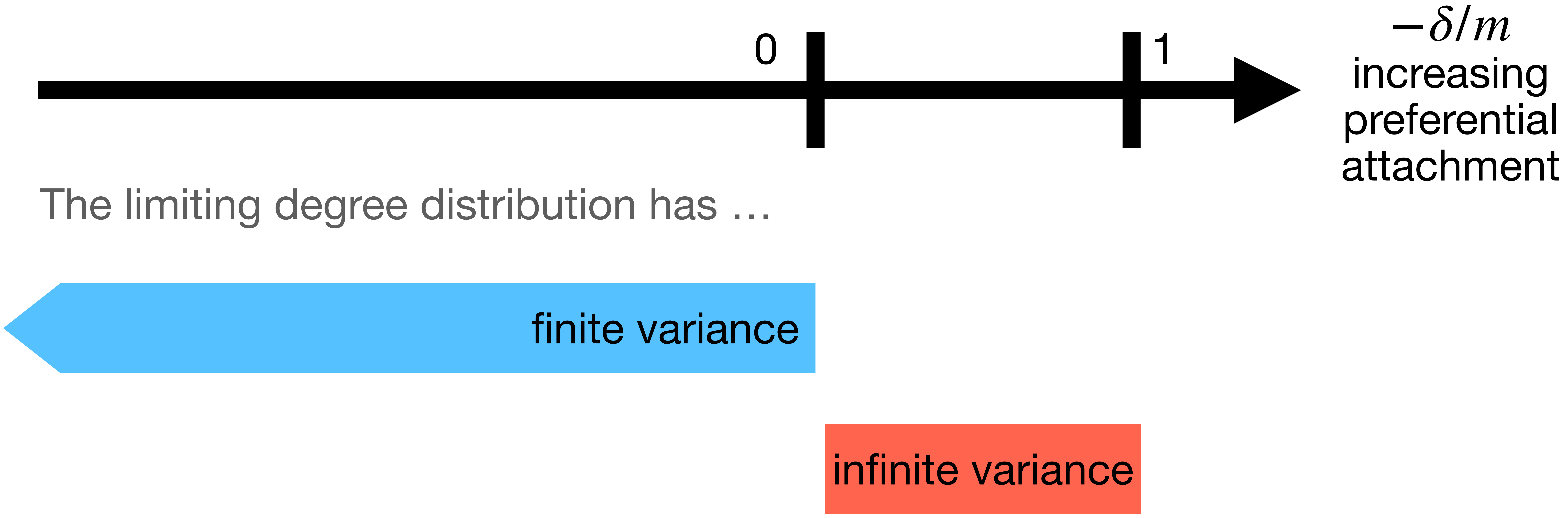


Phase transition

Recall

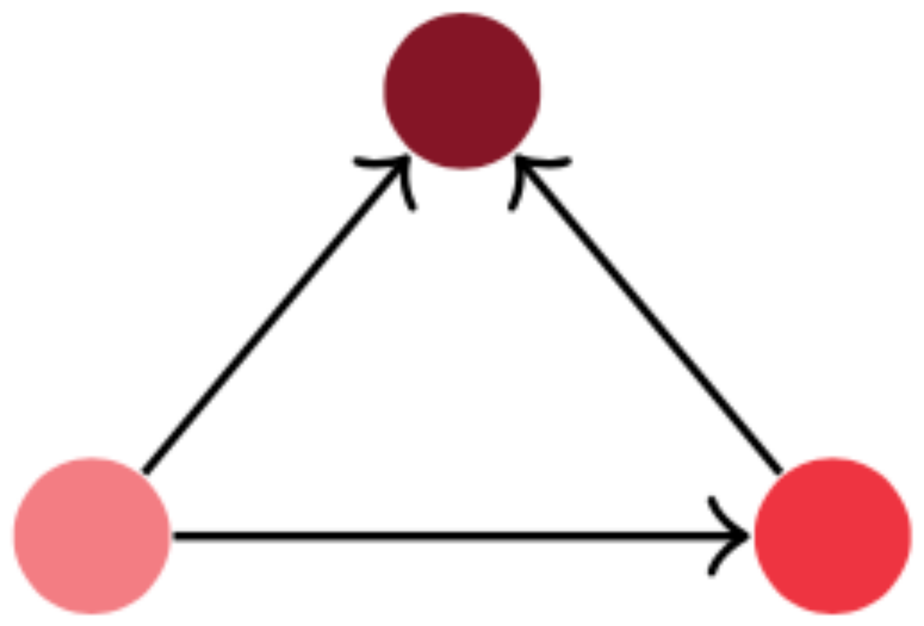
P(attaching to v) $\propto$ degree + δ

m = number of edges per new node



What do we know?

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



(a) $t^{(3-\tau)/(\tau-1)} \log(t)$

Fig 2 of A. Garavaglia and C. Stegehuis (2019).
Subgraphs in Preferential Attachment Models.
https://doi.org/10.1017/apr.2019.36

What do we know?

- subgraph counts [Garavaglia and Stegehuis 2019]

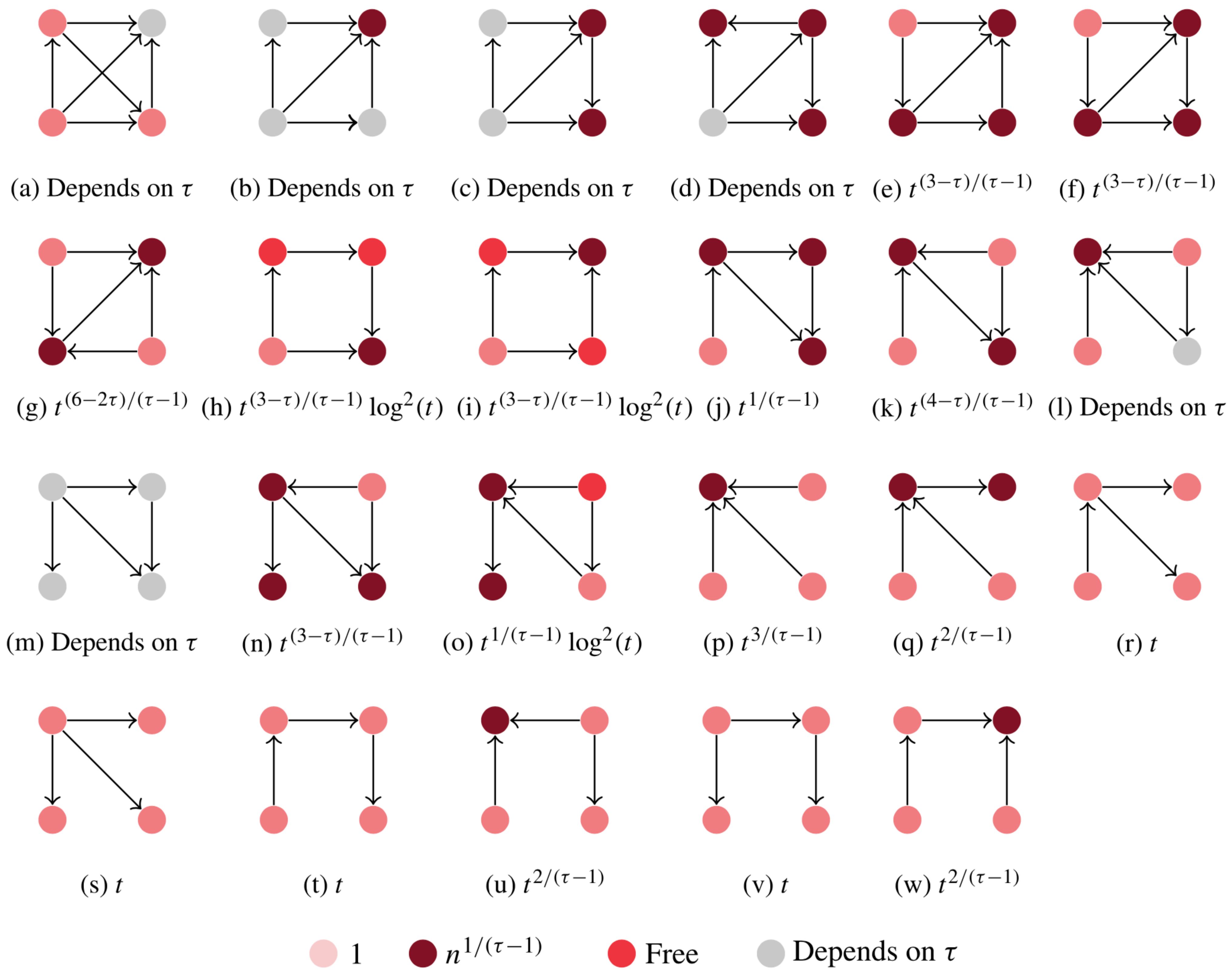
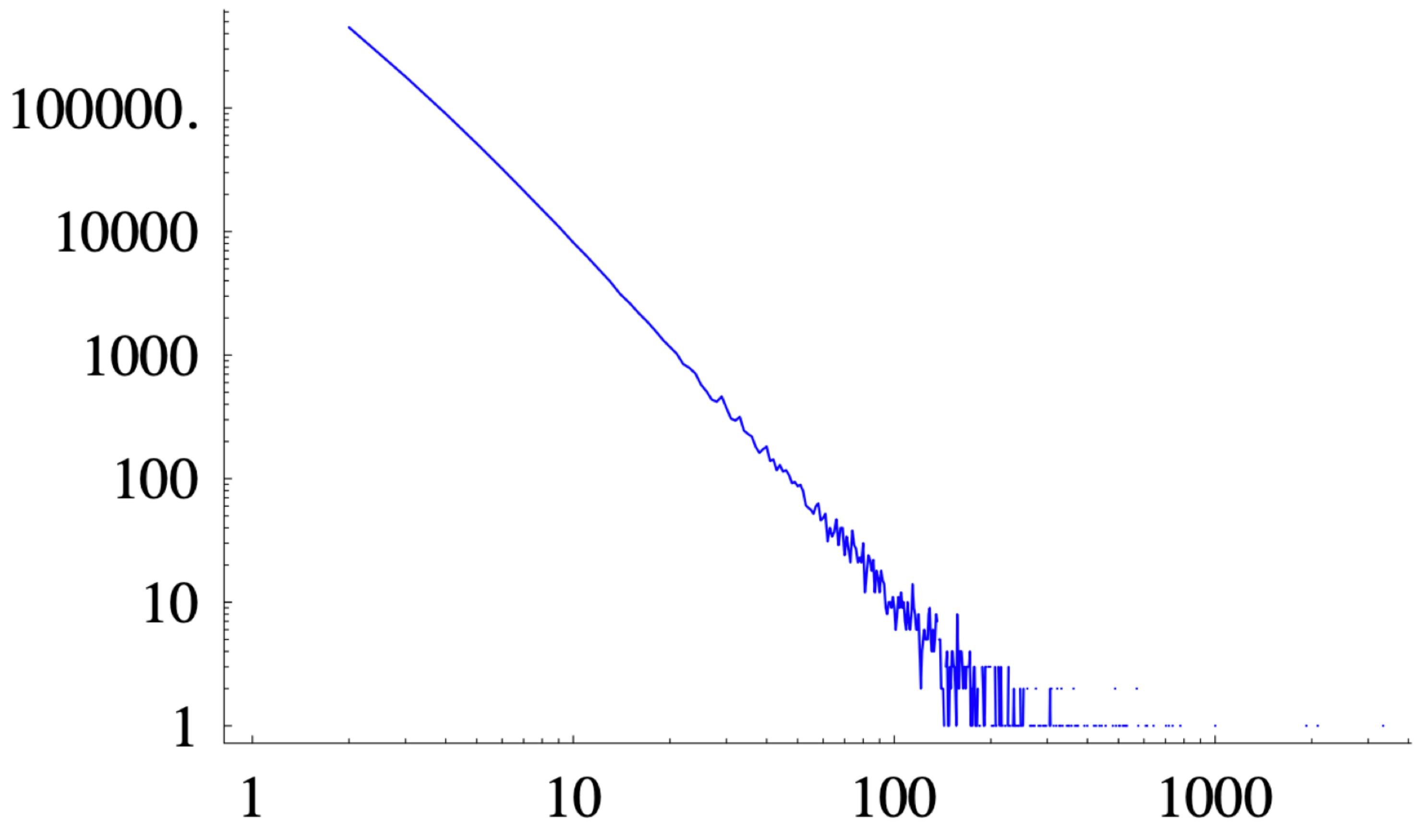
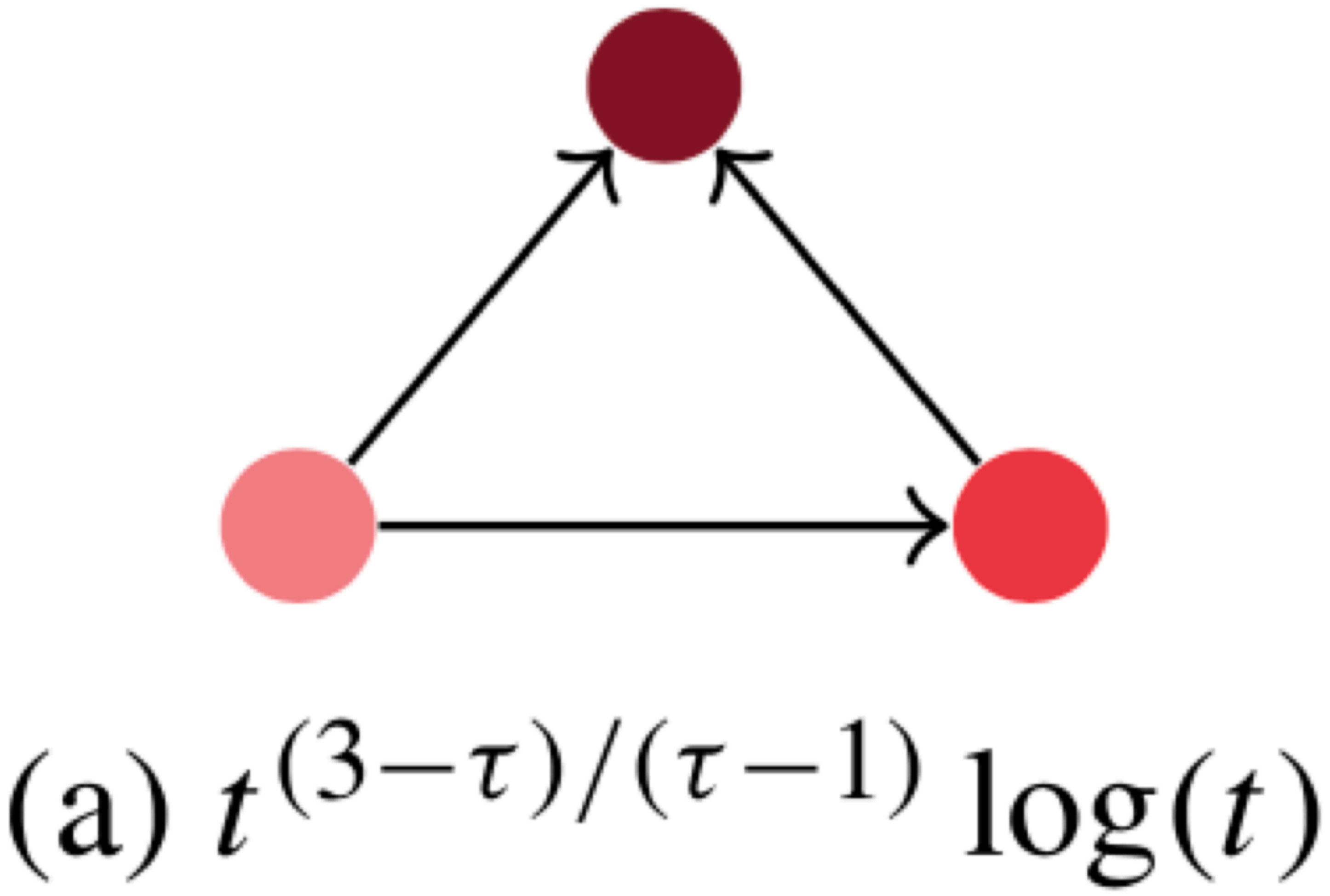


Fig 3 of A. Garavaglia and C. Stegehuis (2019). Subgraphs in Preferential Attachment Models. https://doi.org/10.1017/apr.2019.36

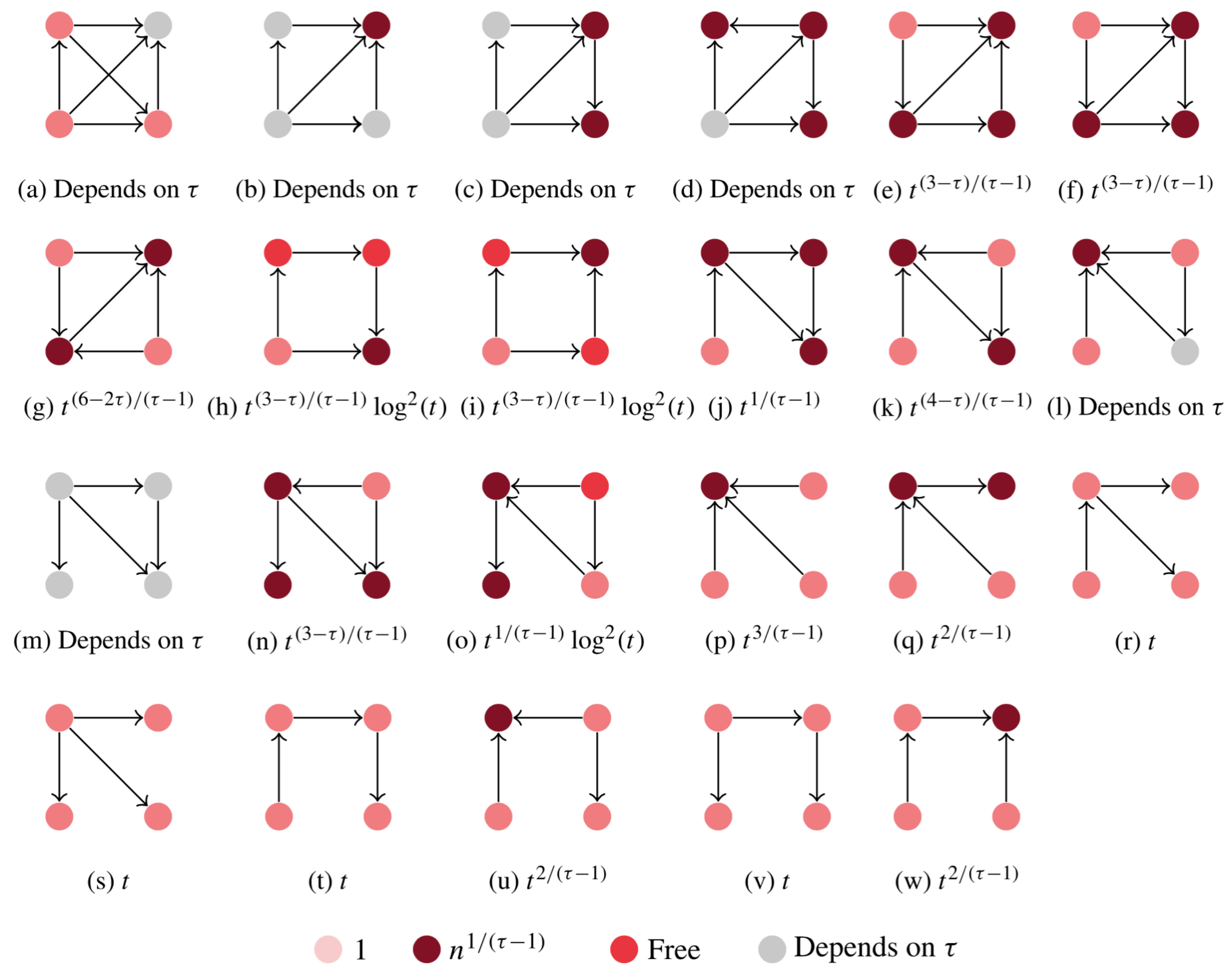
What do we know?



degree distribution

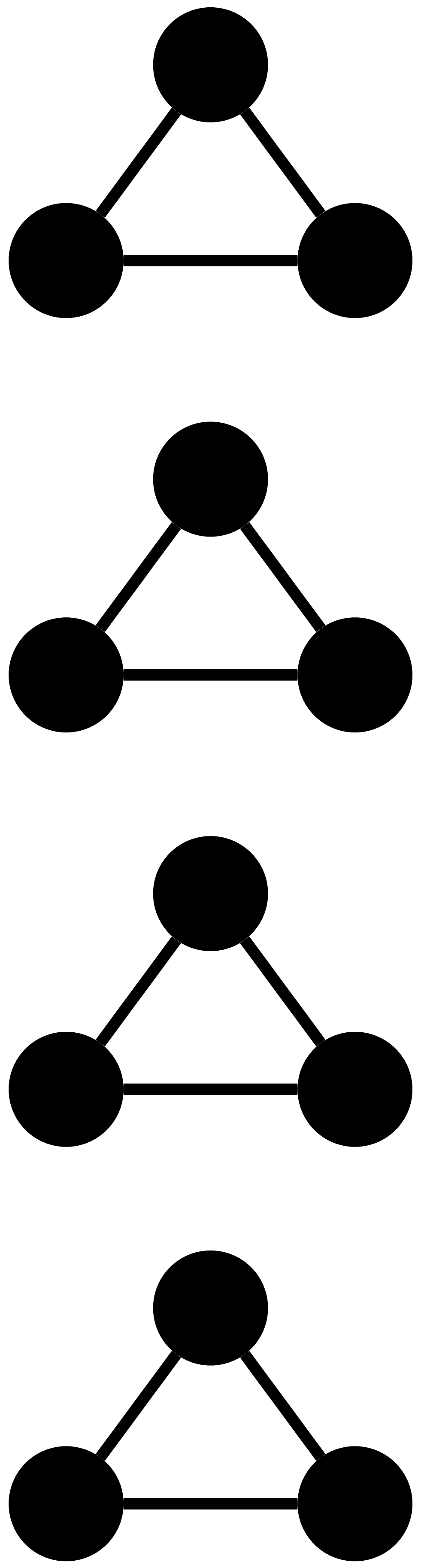


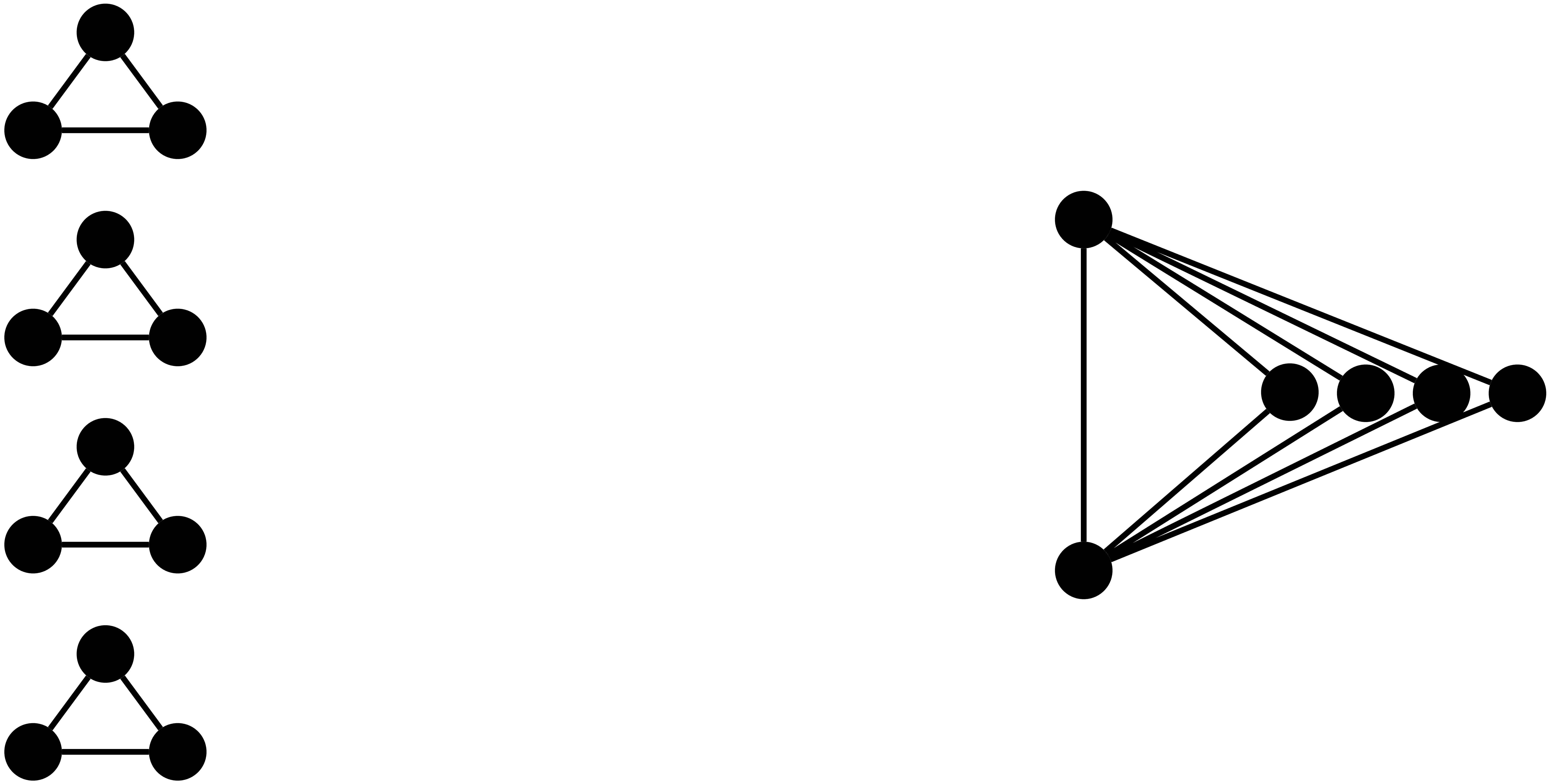
triangle counts



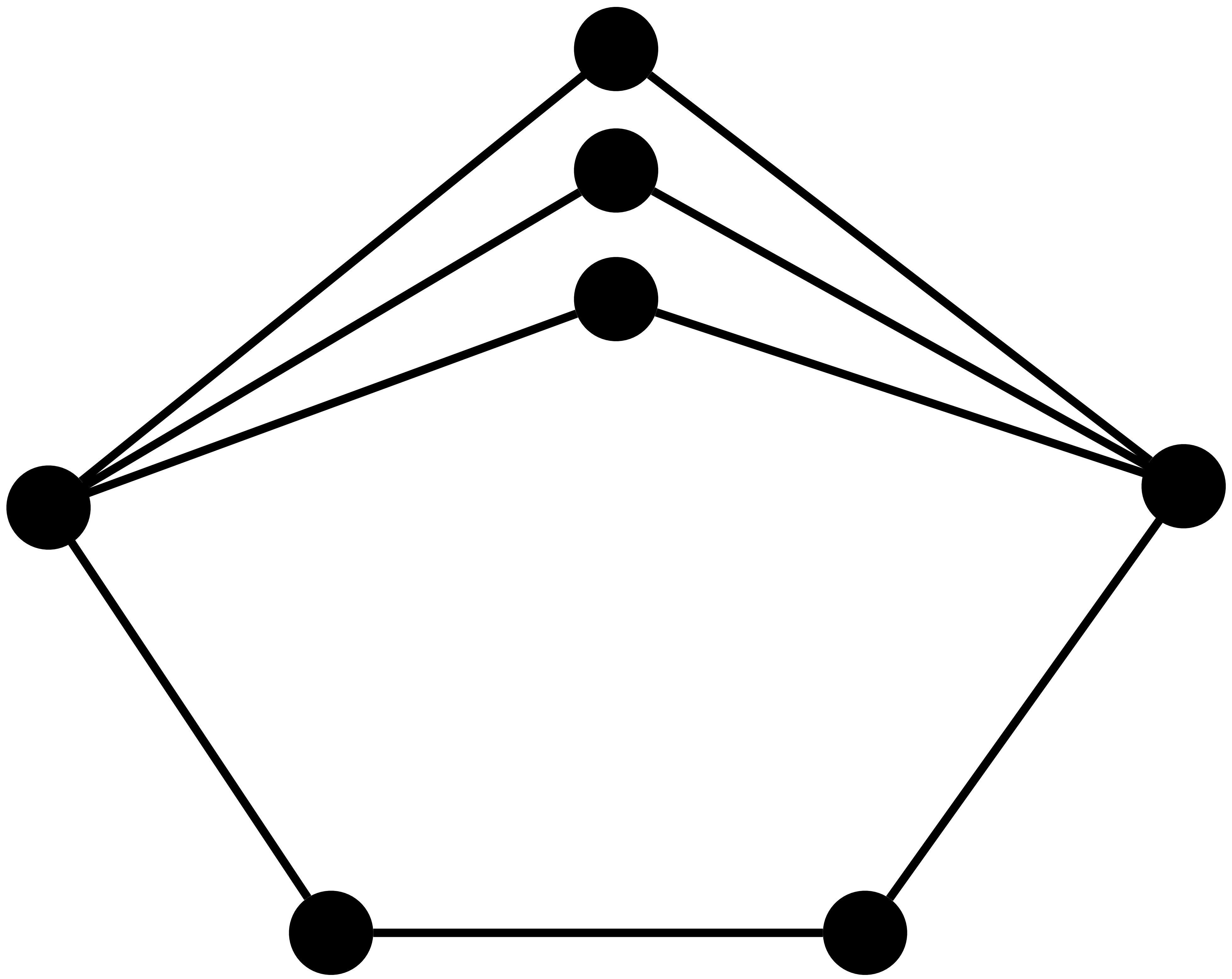
subgraph counts

What should we count?
And how?

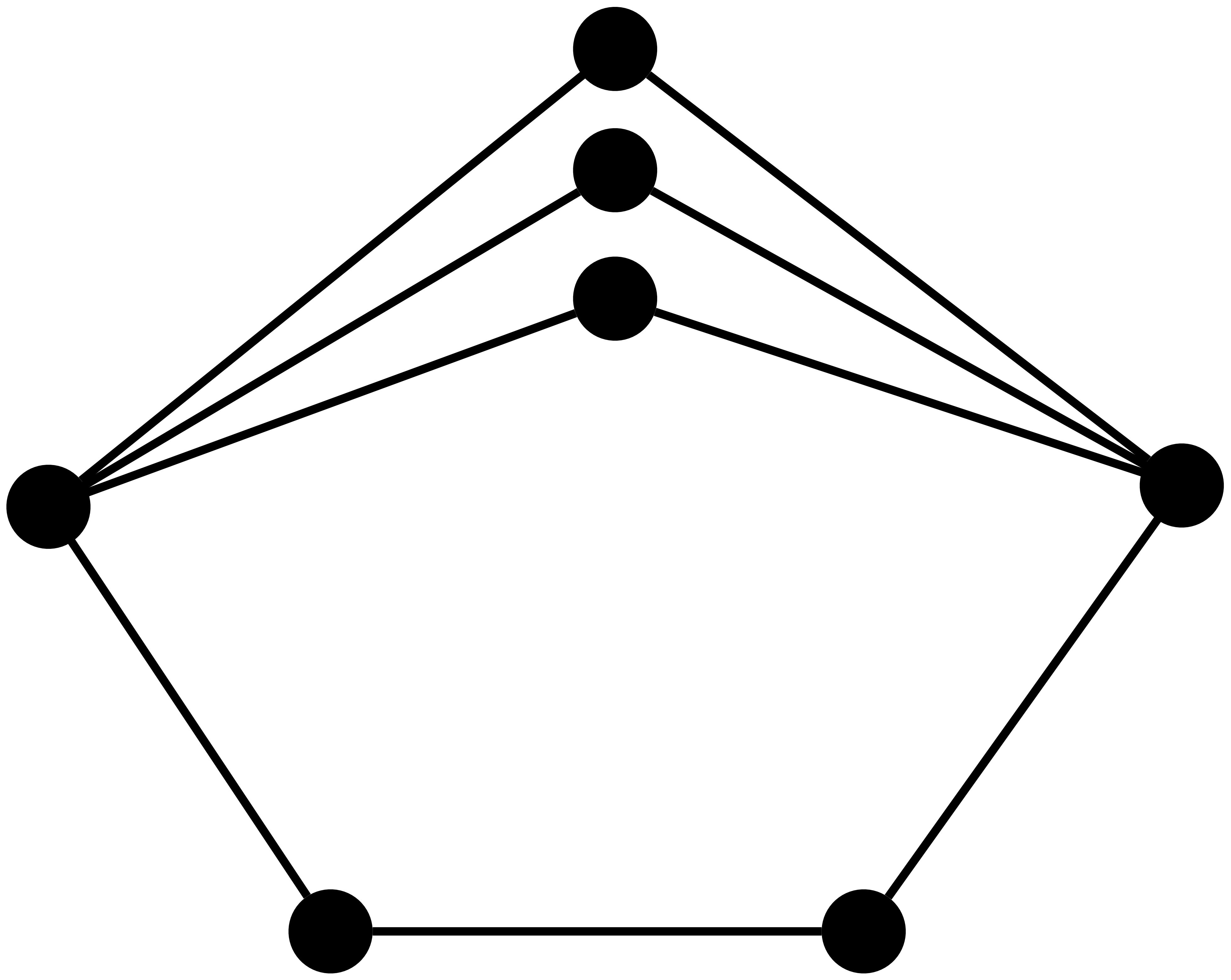




Paths from left to right?

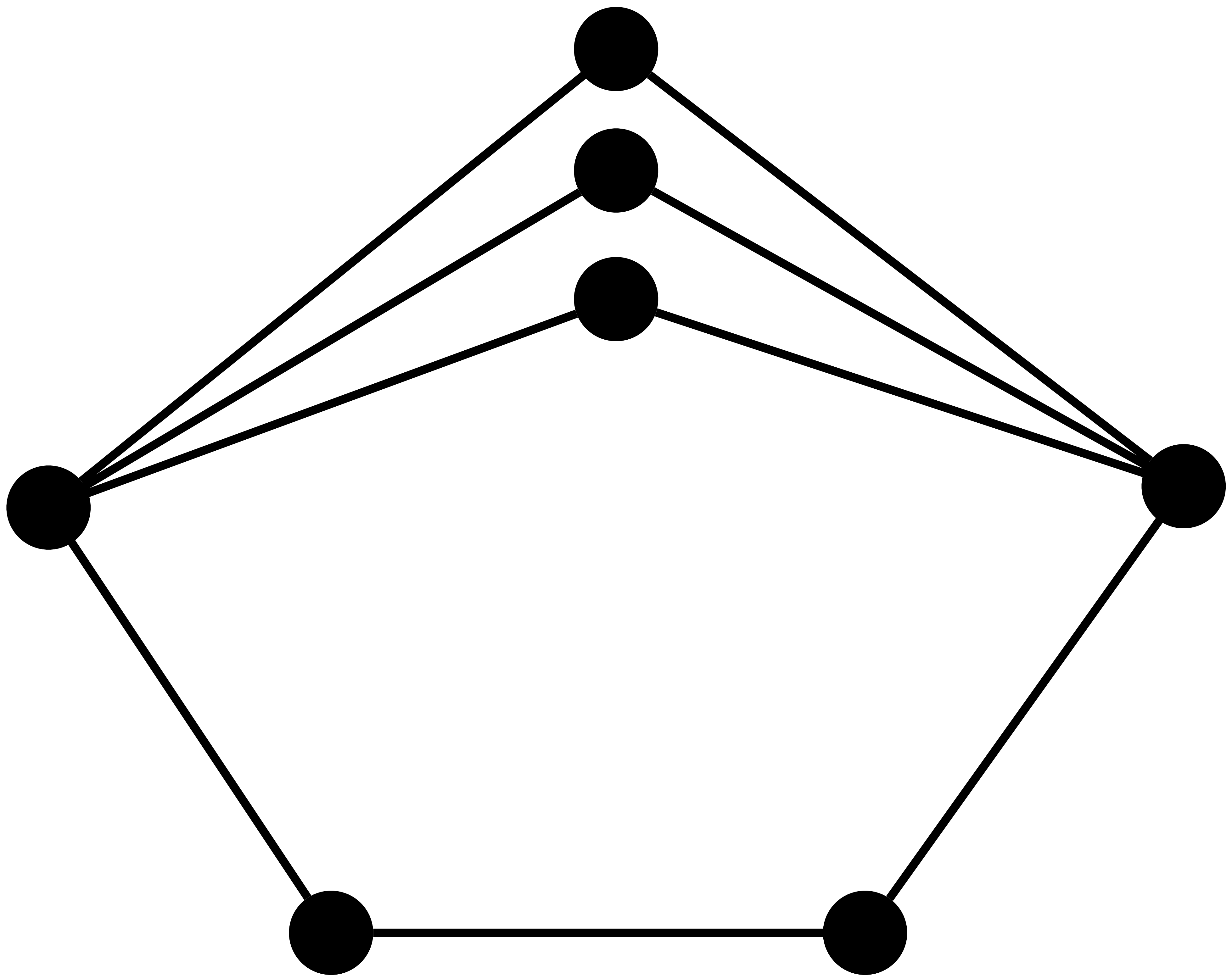


Paths from left to right?



- backtracking?

Paths from left to right?



- backtracking?
- concatenating with loops?

II. Into Topology

Counting everything in every dimension all at once

Betti numbers count repeated connections “in all dimensions”.

Betti numbers count repeated connections “in all dimensions”.

GOOD ✓

“correct” way to count things

Betti numbers count repeated connections “in all dimensions”.

GOOD ✓

“correct” way to count things

homological algebra

Betti numbers count repeated connections "in all dimensions".

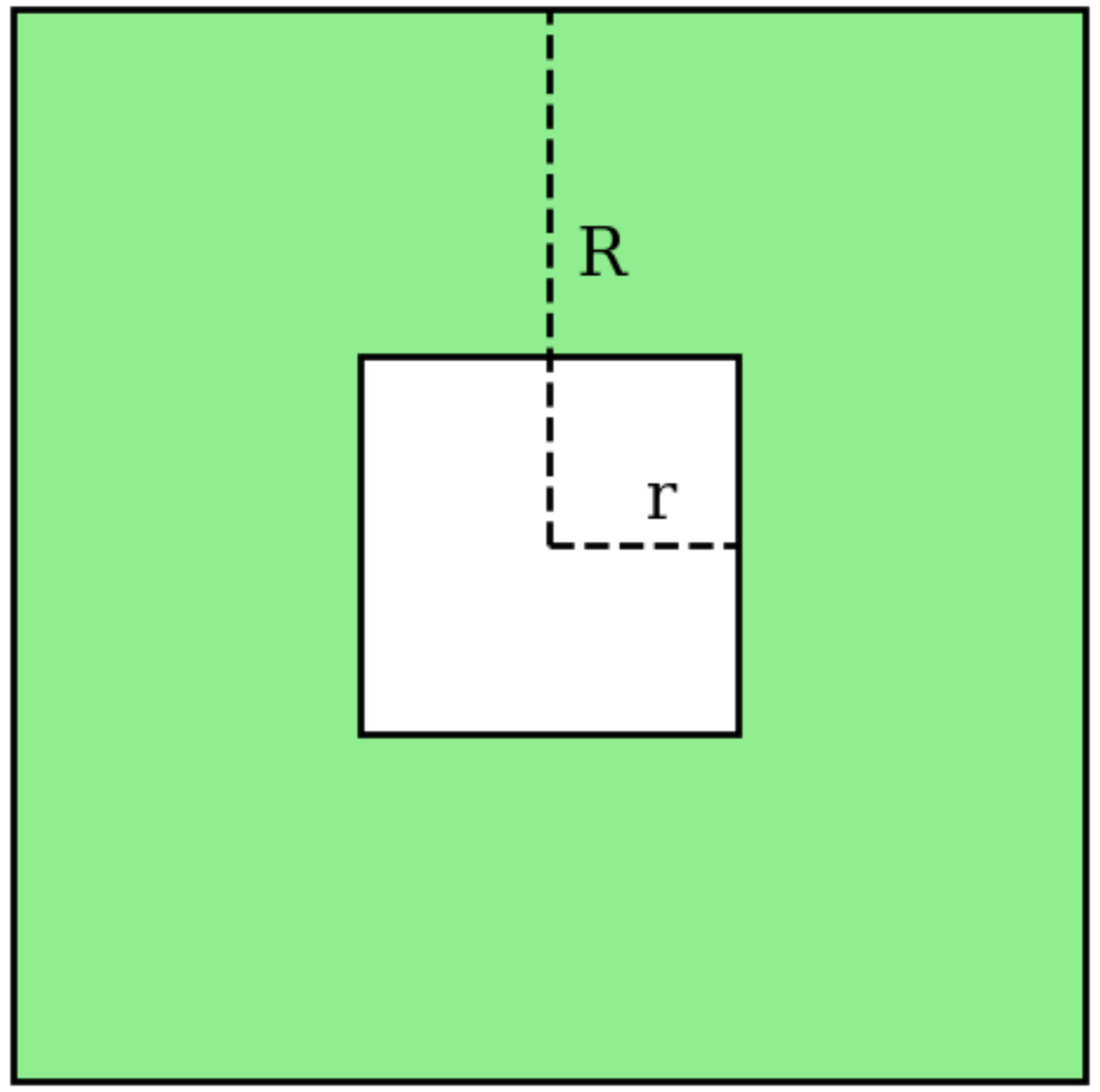
GOOD ✓	**BAD** ✗
"correct" way to count things	hard to write down
homological algebra	hard to do

Betti numbers β_k

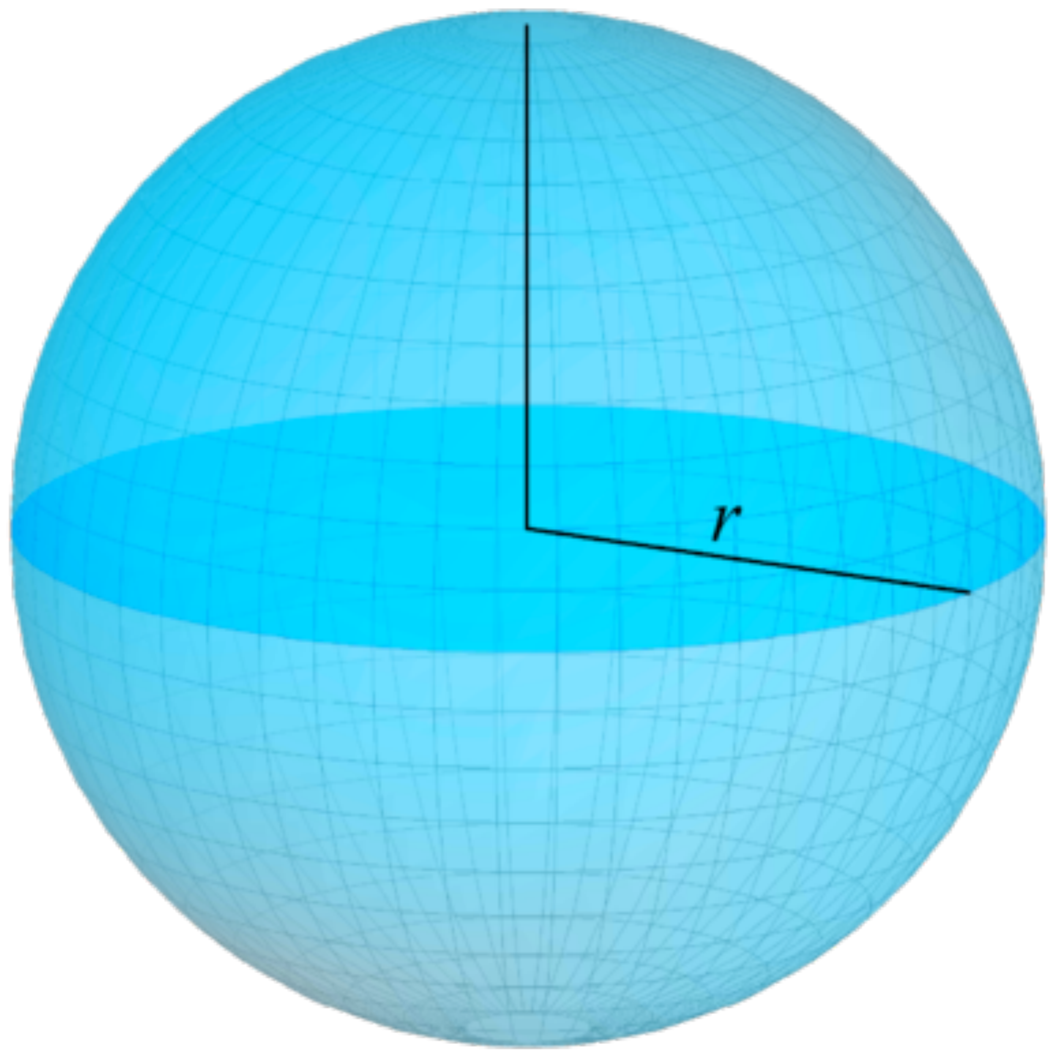
- Repeated connections?
- Holes?

Betti numbers β_k

Count of Holes



$\beta_1 = 1$: 1 loop

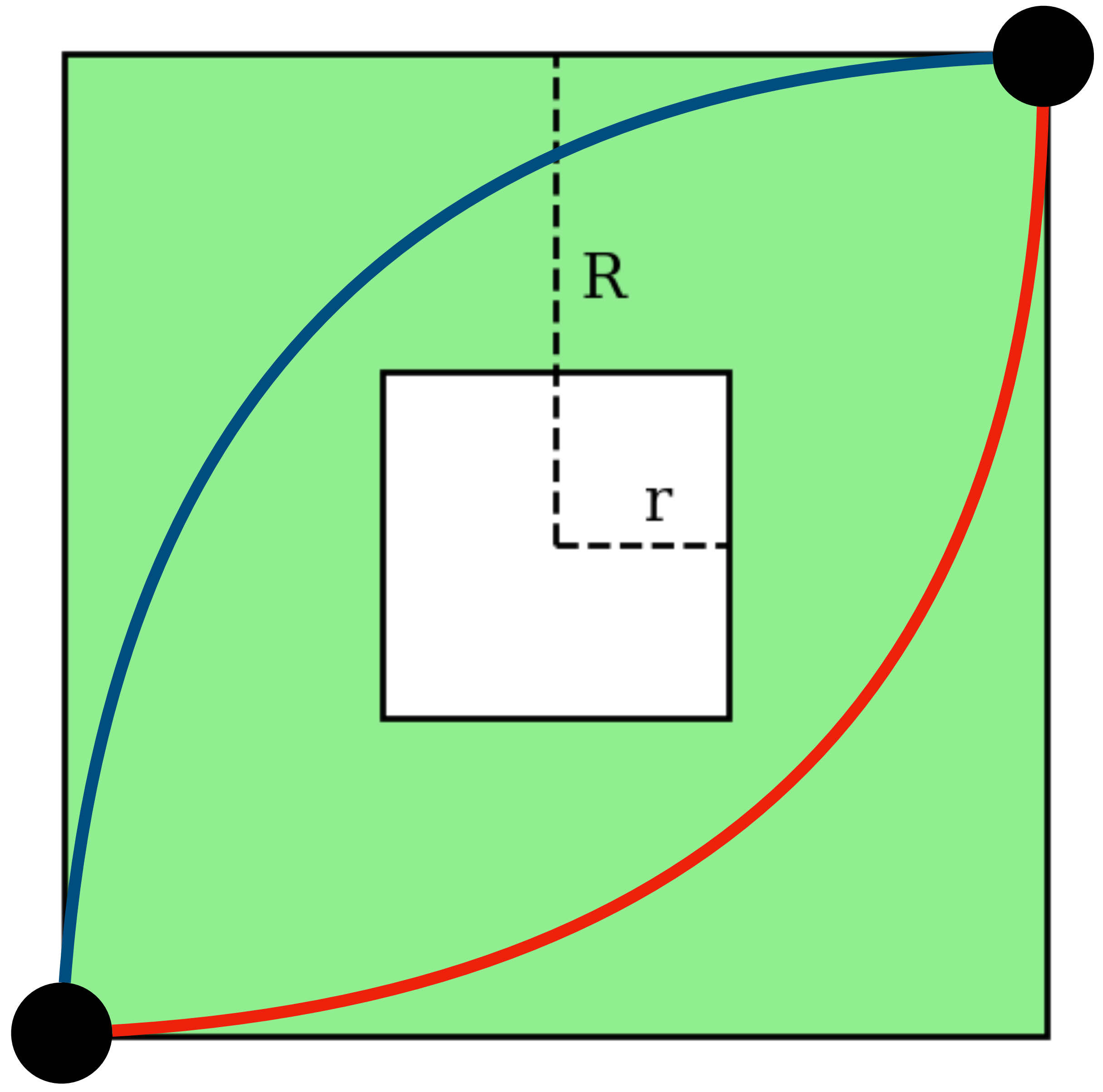


$\beta_1 = 0$: 0 loop

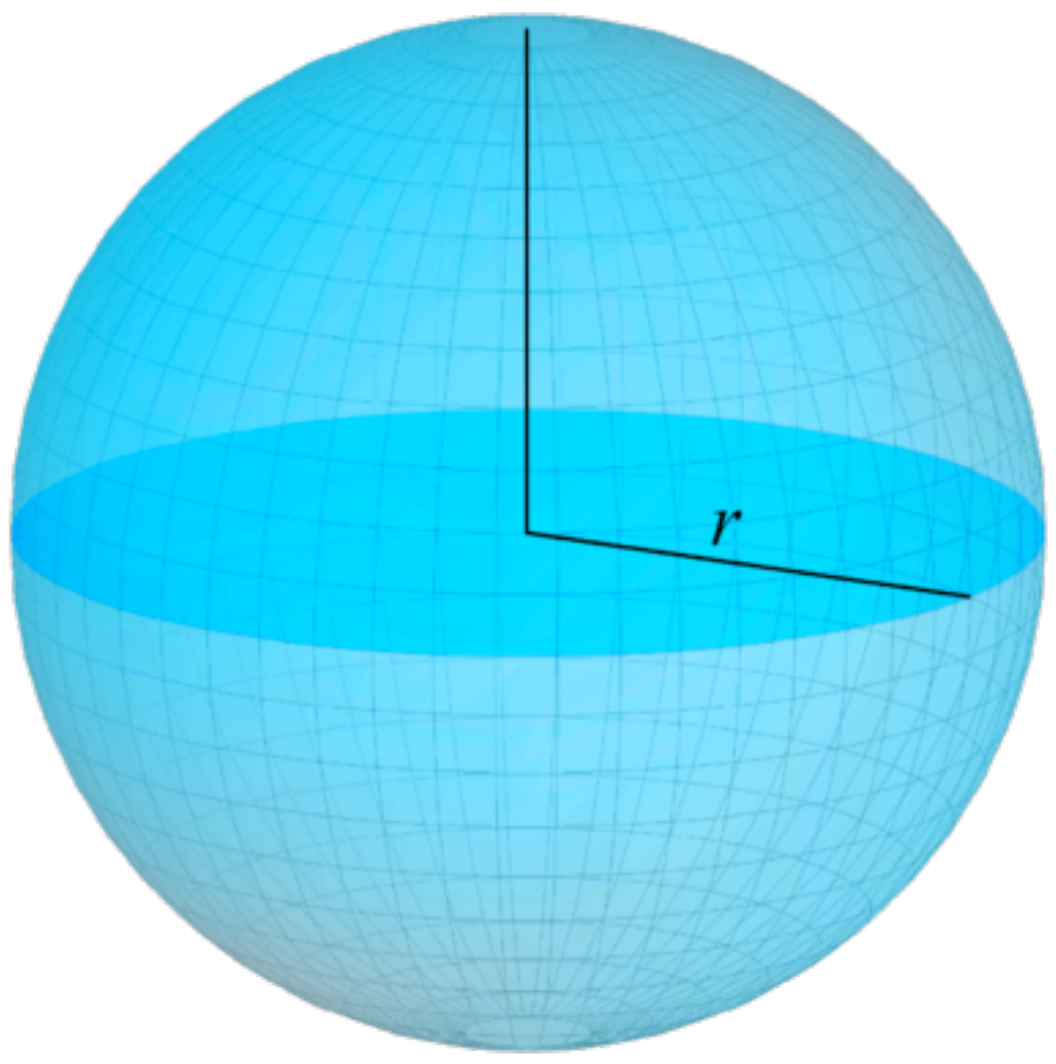
$\beta_2 = 1$: 1 cavity

Betti numbers

Count of Repeated Connections



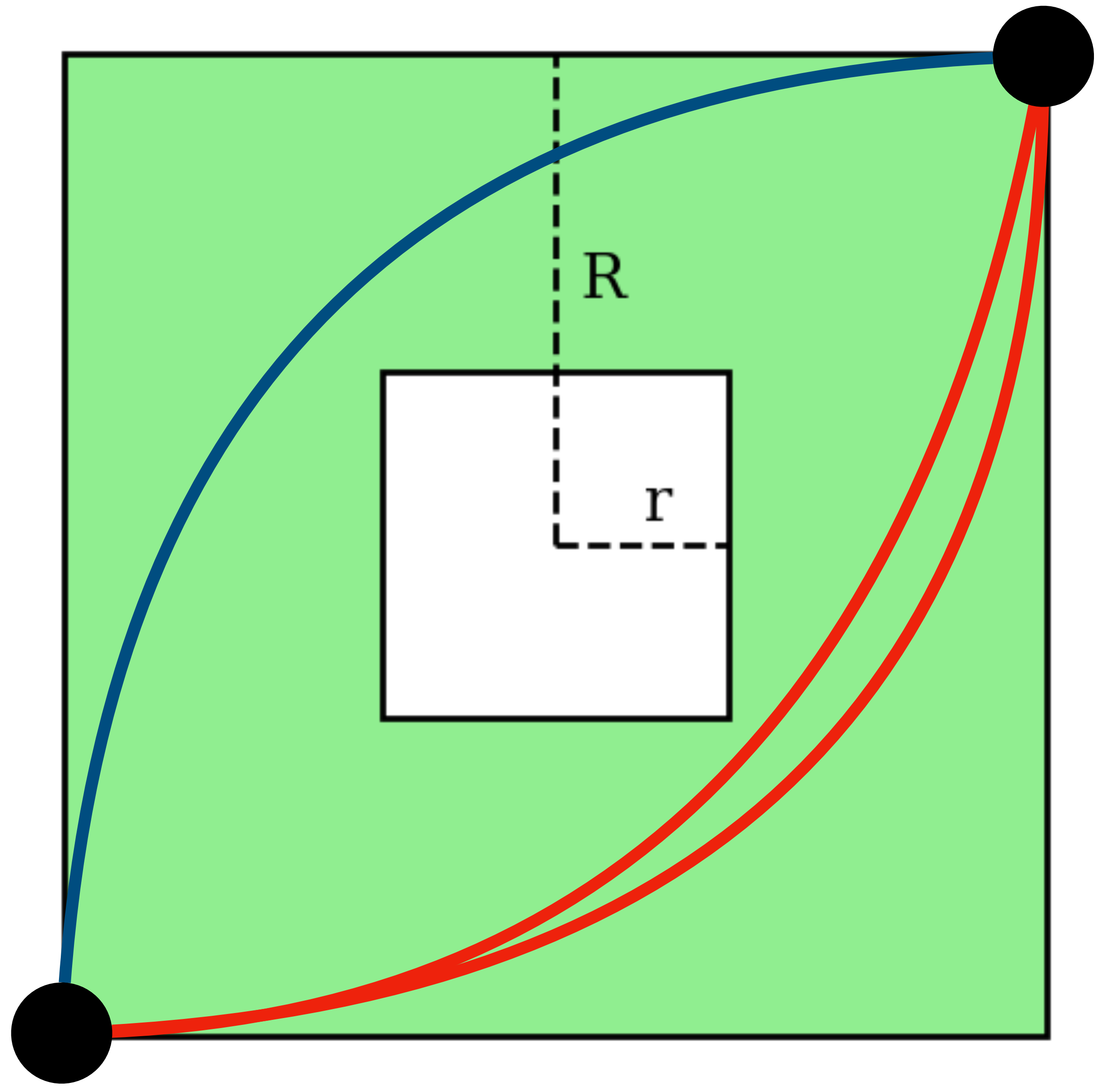
1 alternative path



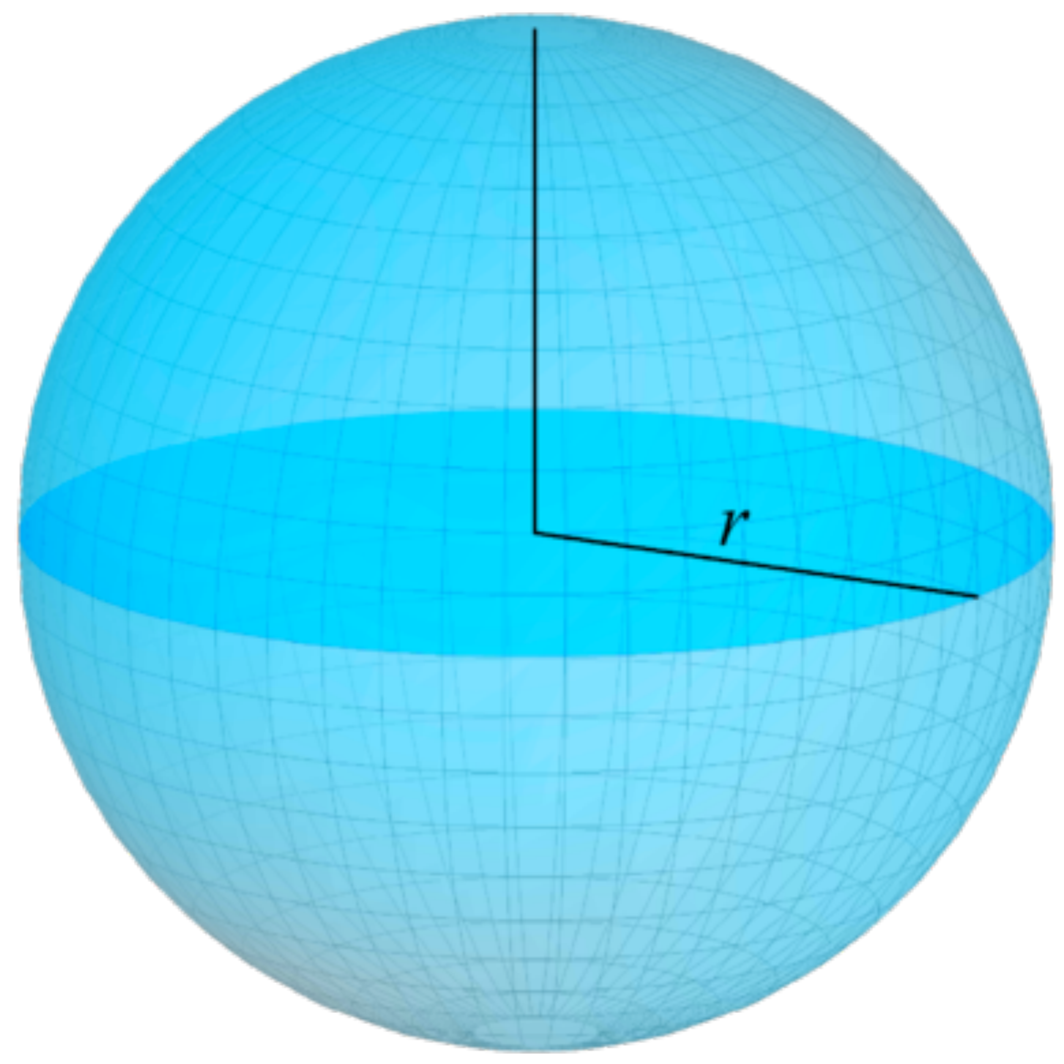
0 loop
1 cavity

Betti numbers

Count of (Independent) Repeated Connections



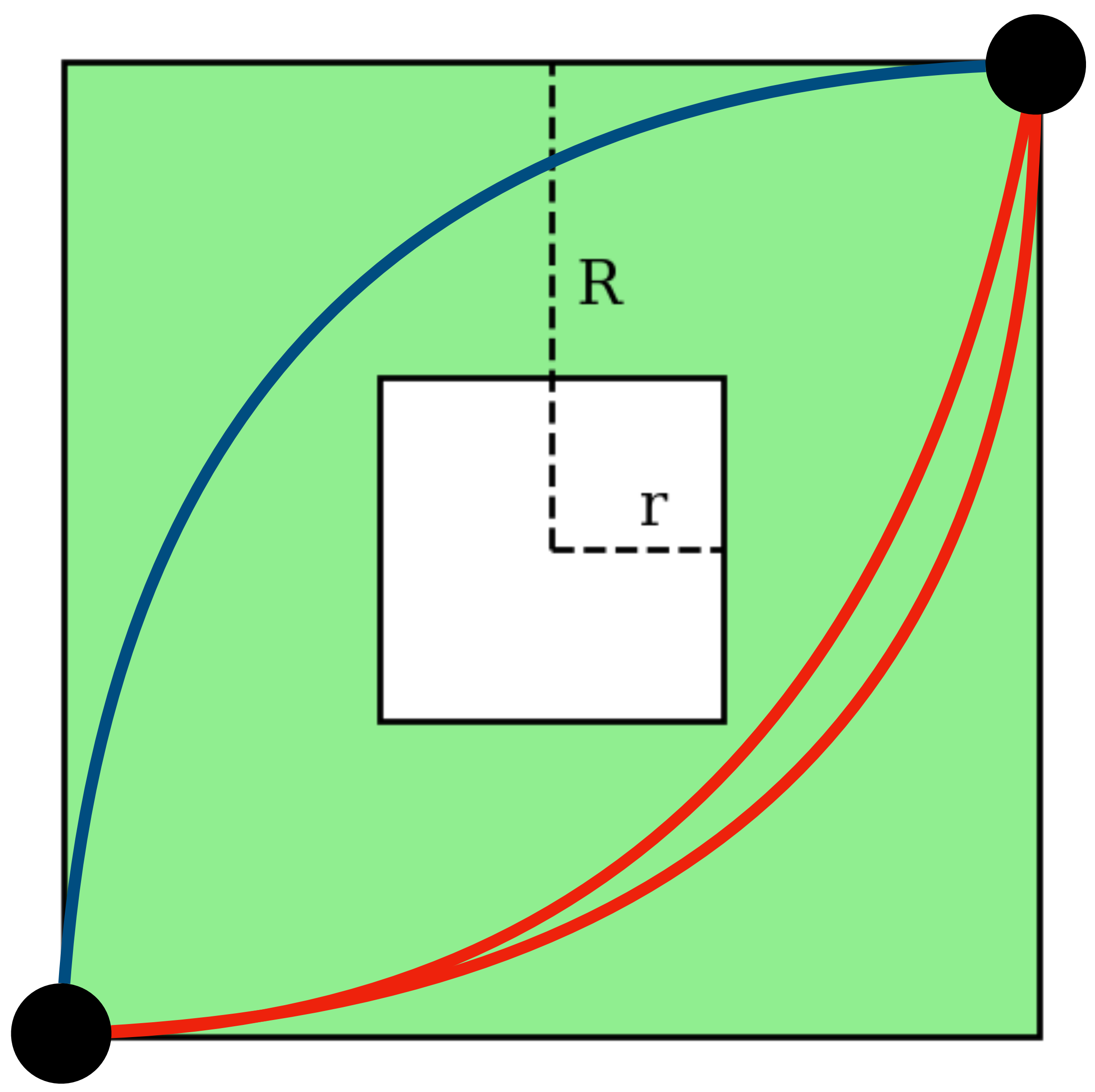
1 alternative path



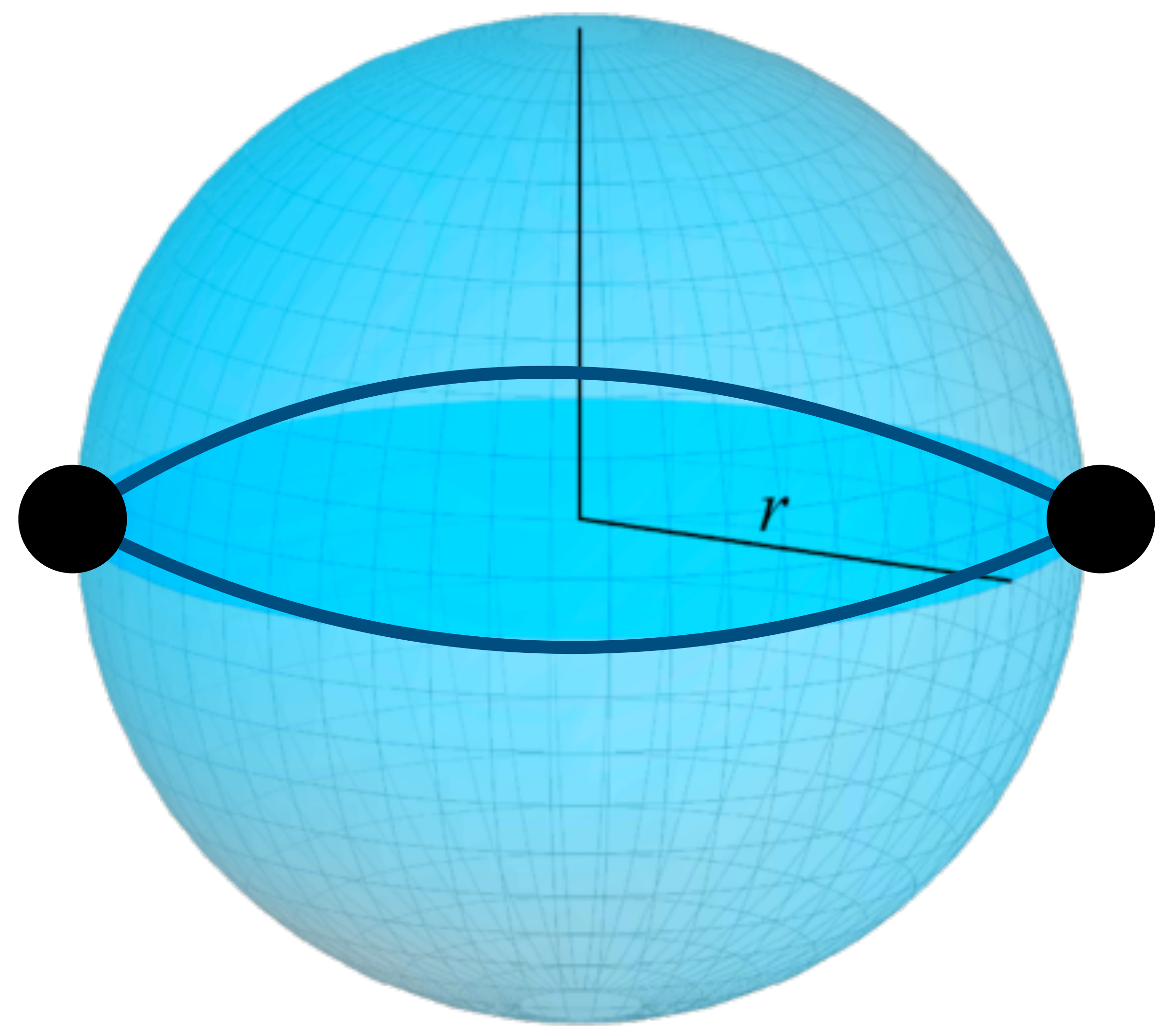
0 loop
1 cavity

Betti numbers

Count of (Independent) Repeated Connections



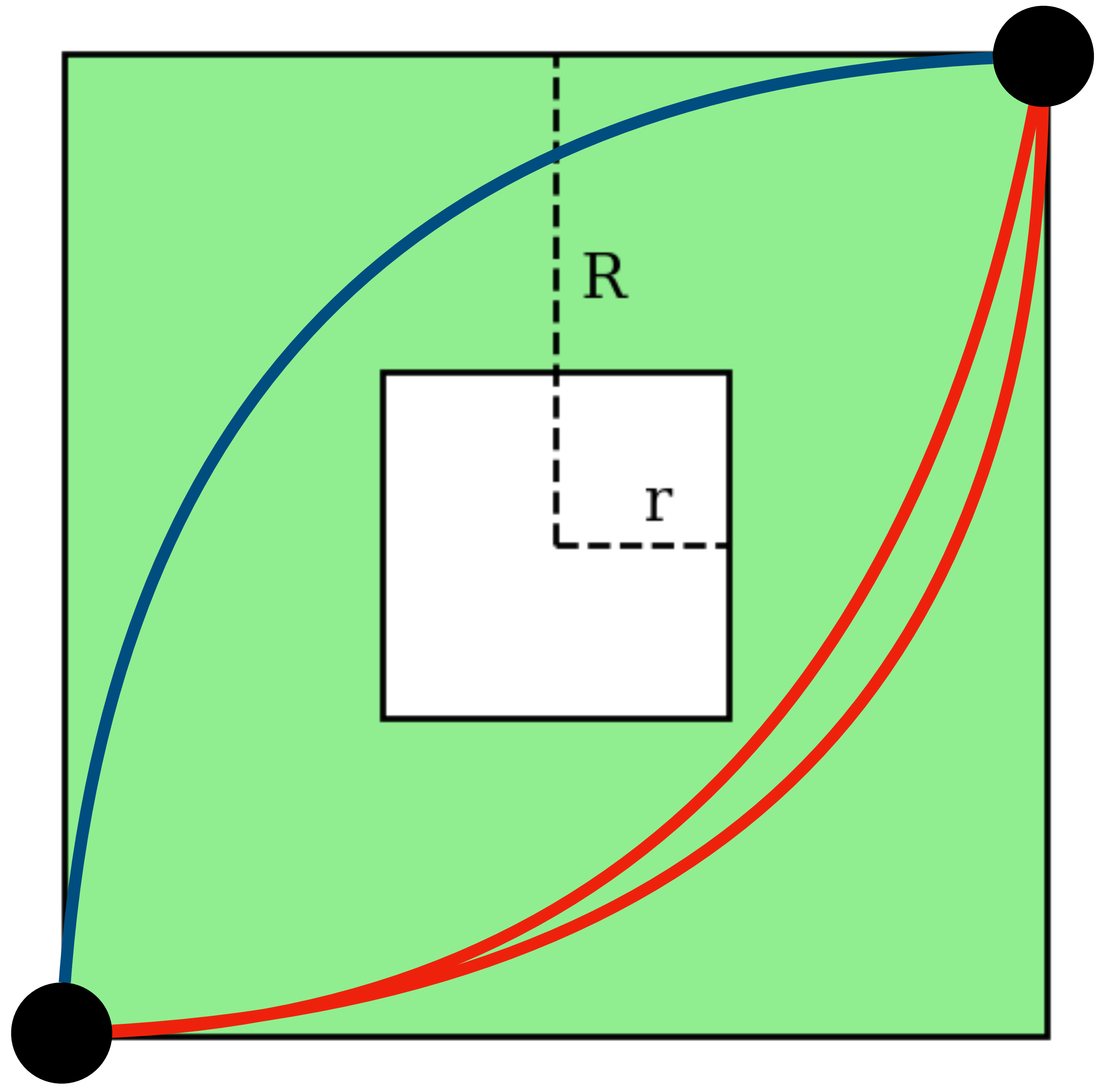
1 alternative path



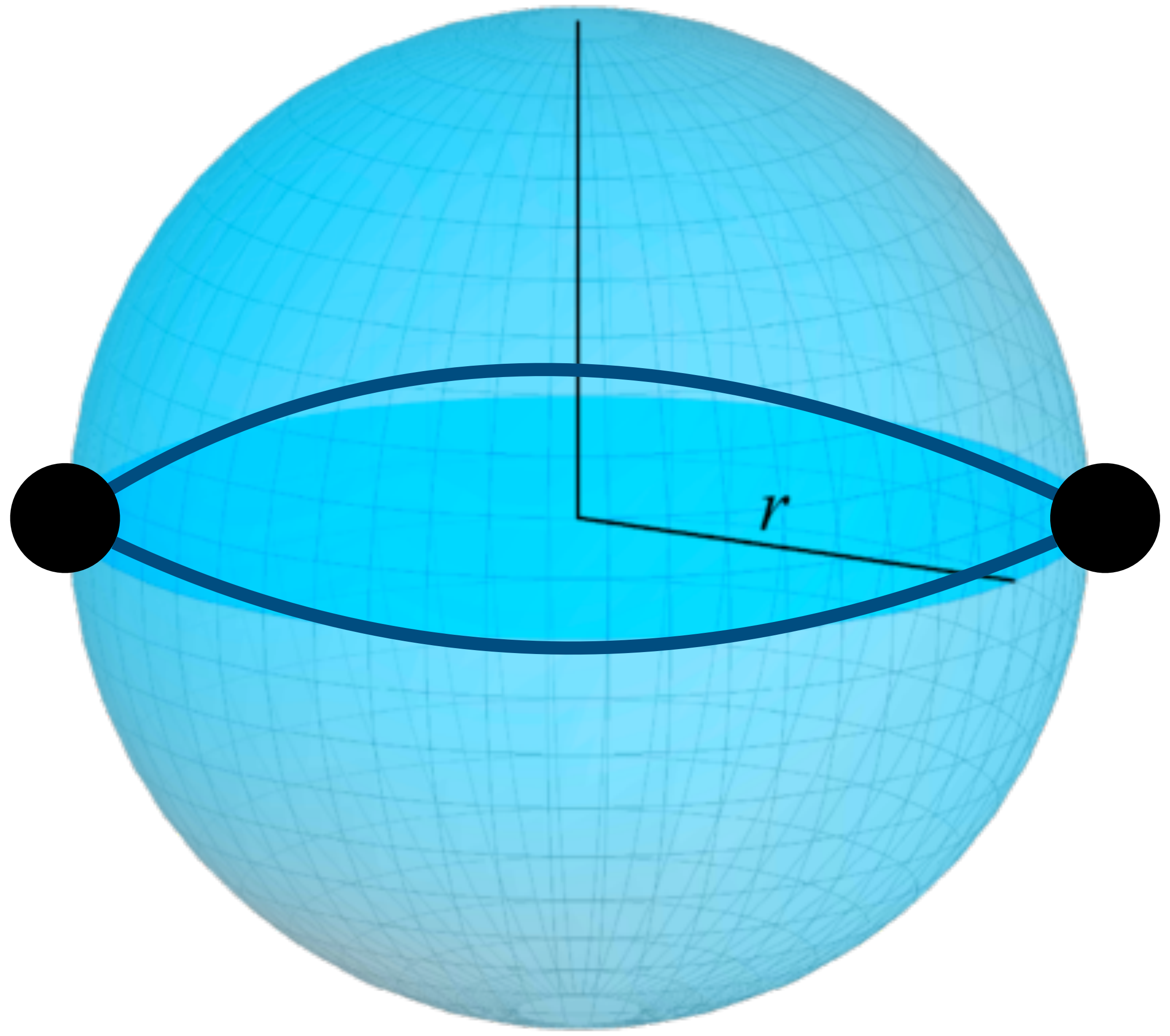
0 alternative path (slide through upper hemisphere)
1 cavity

Betti numbers

Count of (Independent) Repeated Connections



1 alternative path



0 alternative path (slide through upper hemisphere)
1 alternative way to slide a path

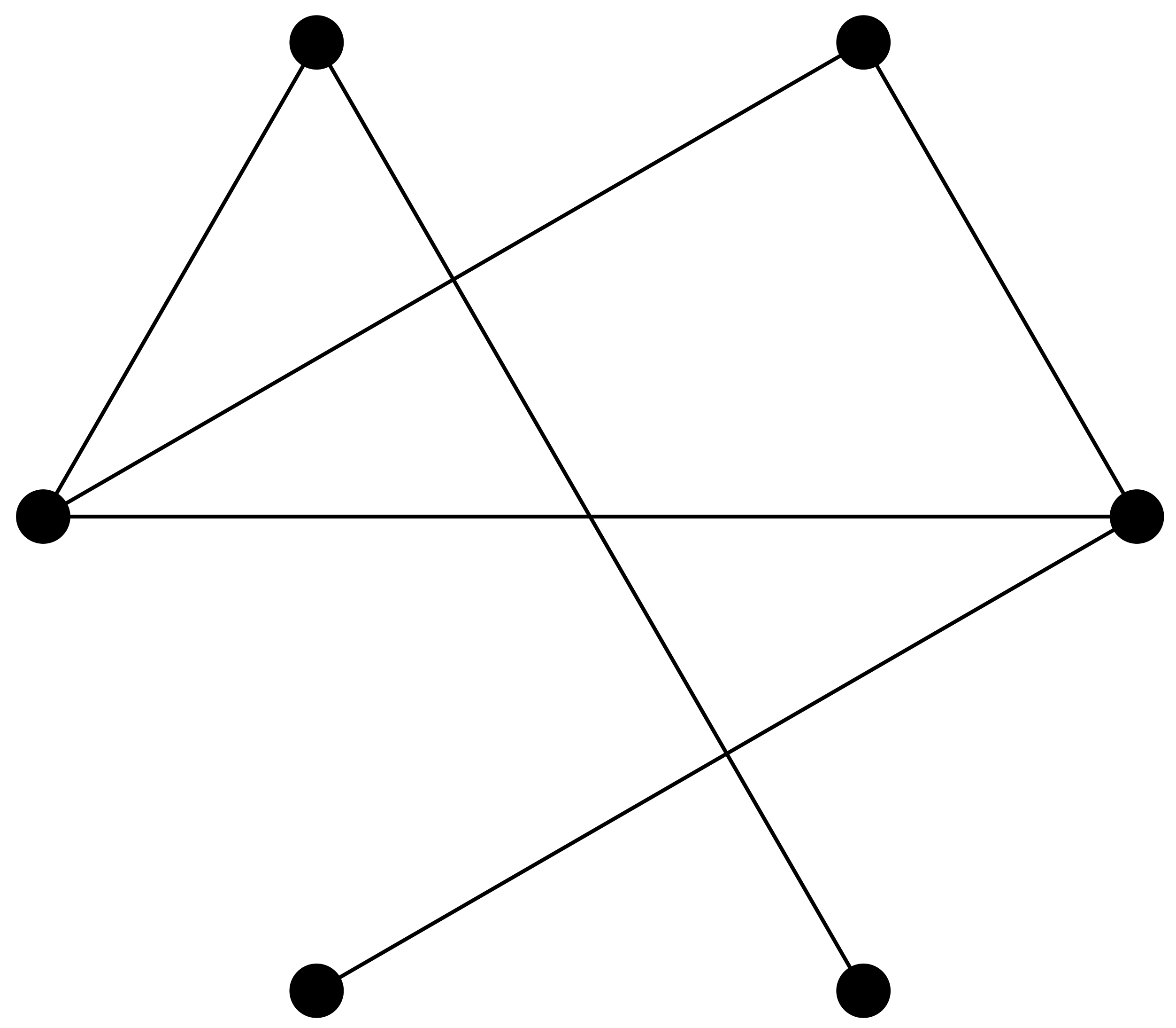
Betti numbers count repeated connections “in all dimensions”.

Interlude:

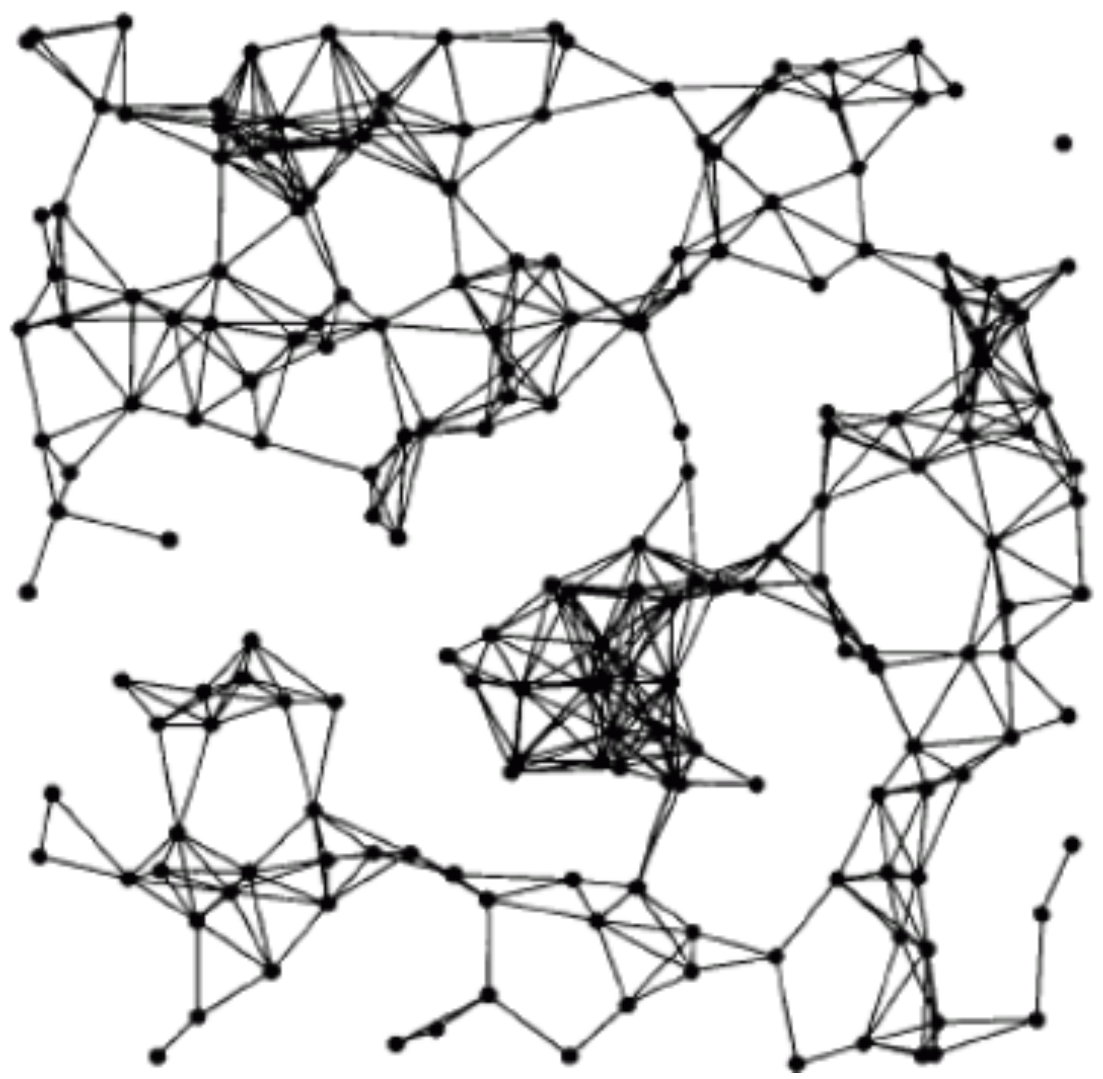
Random Walk in the Literature

What Random Topologists Already Know

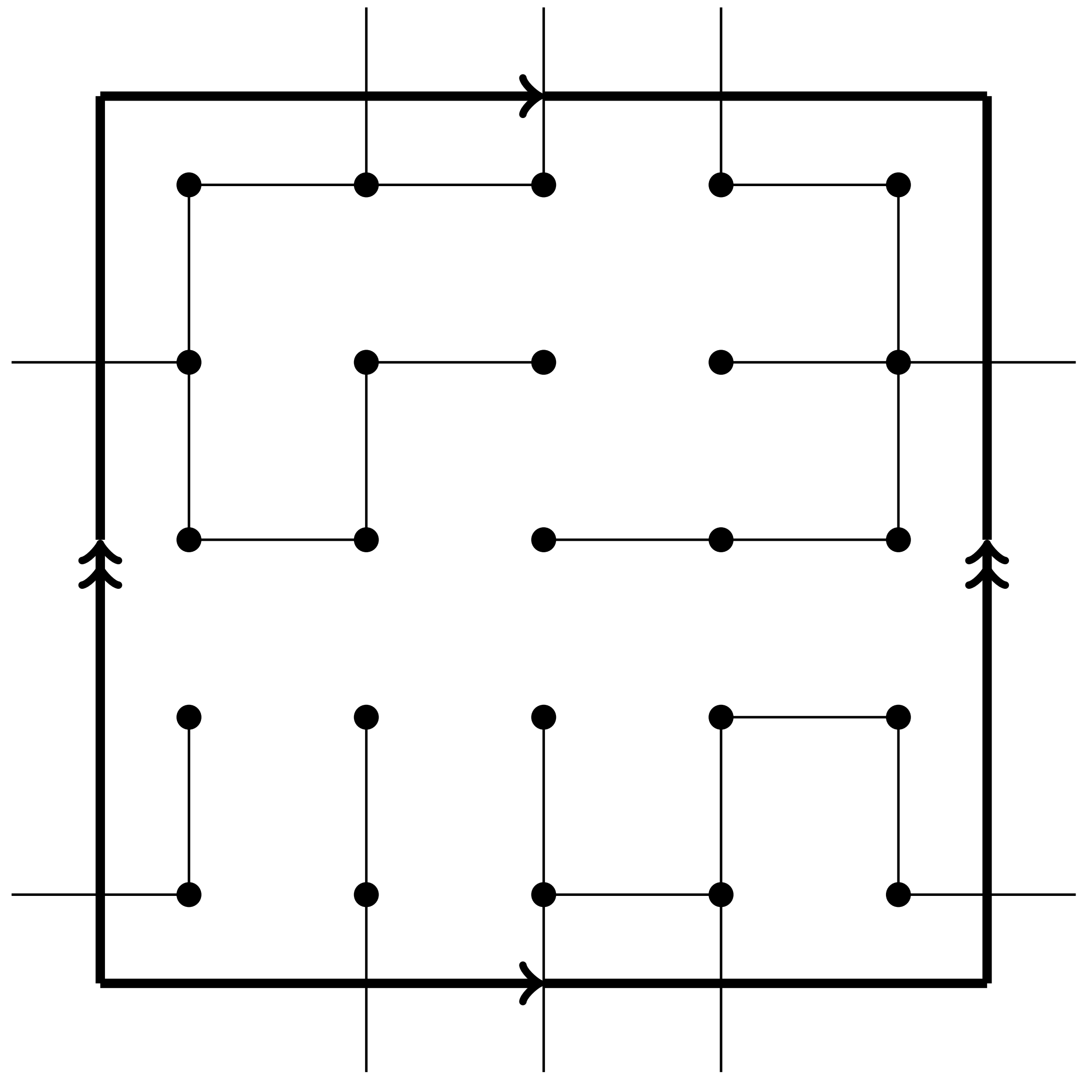
Tapas of Random Topology



Erdo-Renyi Complexes

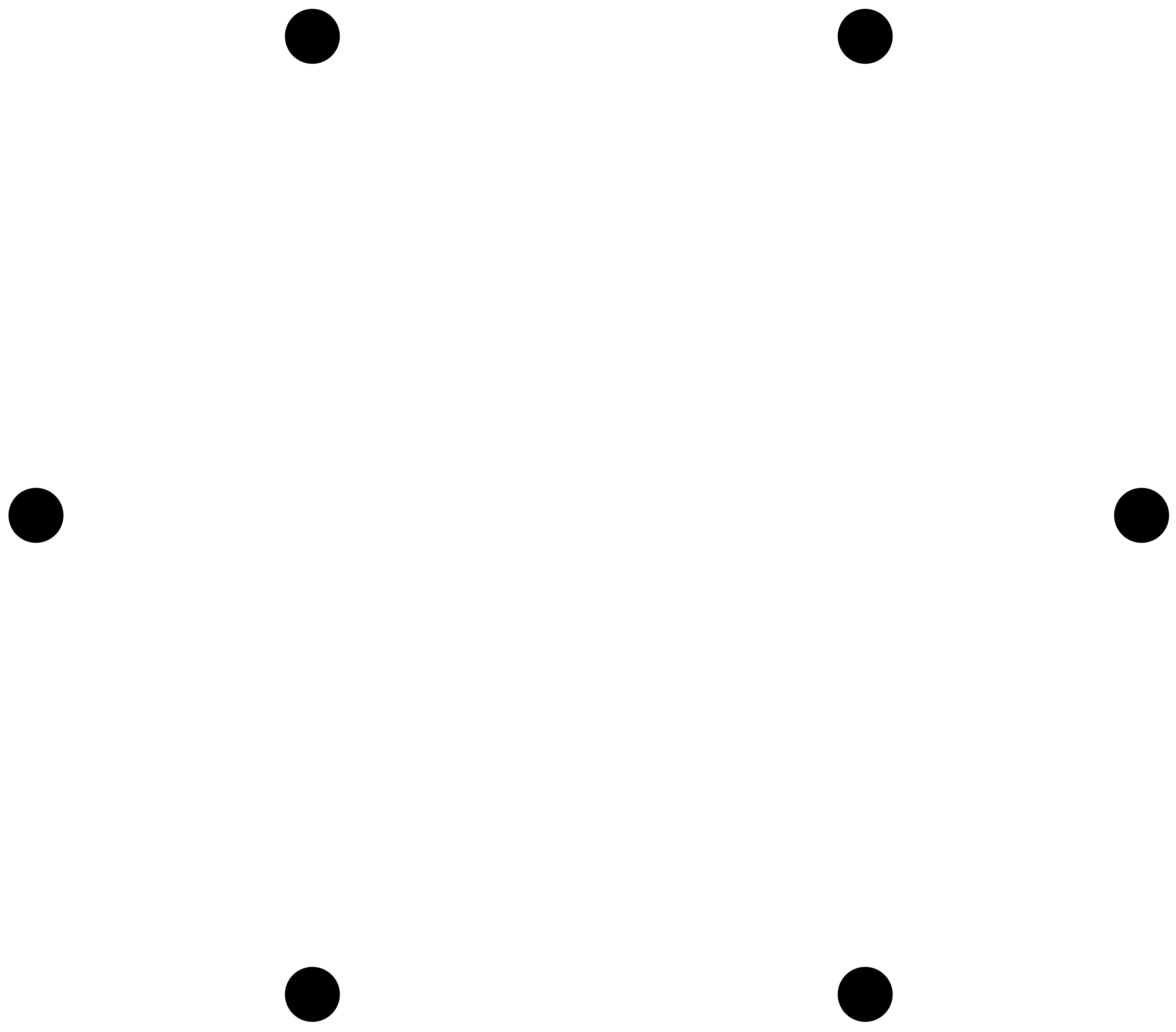


Geometric Complexes

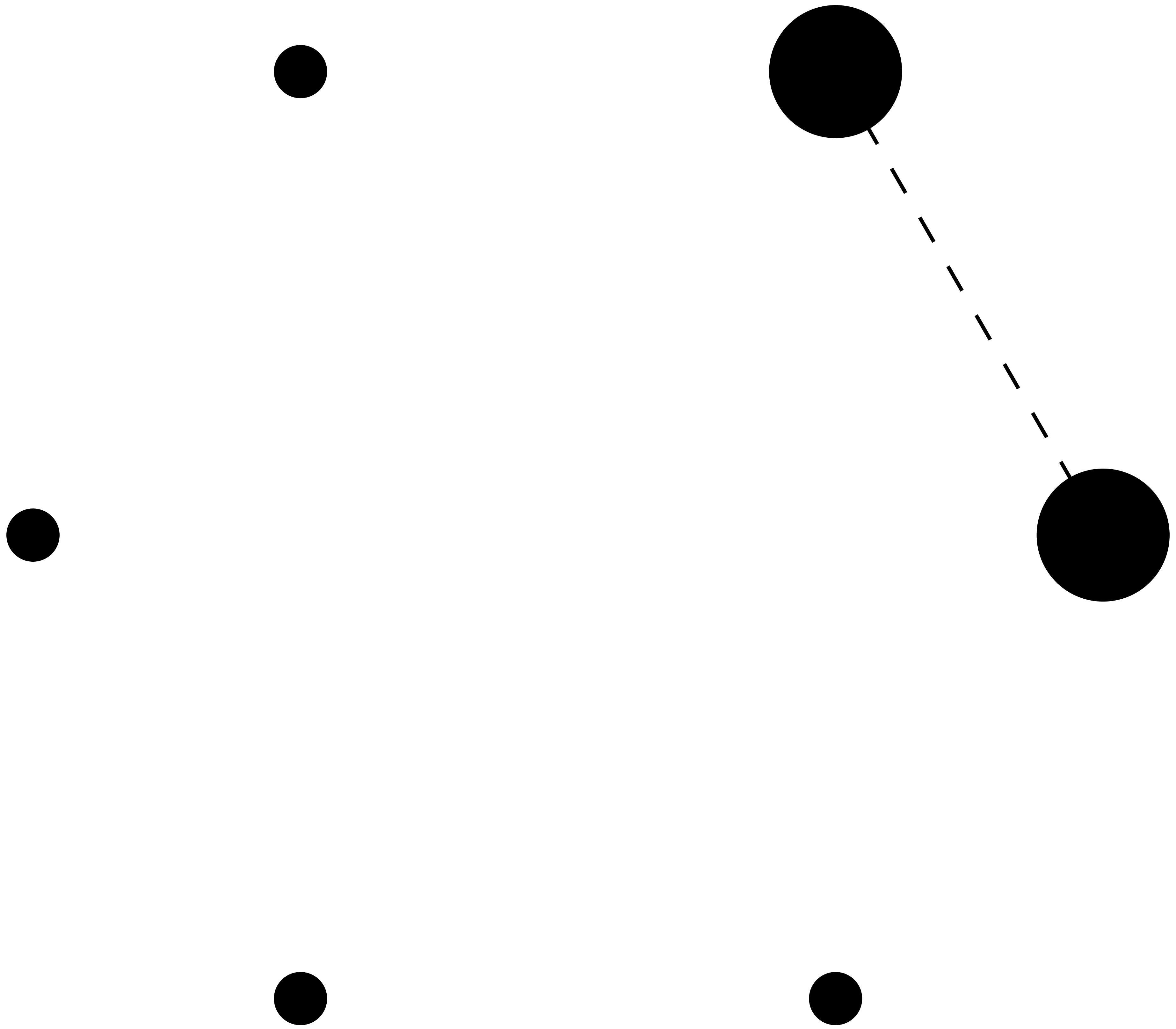


Topological Percolation

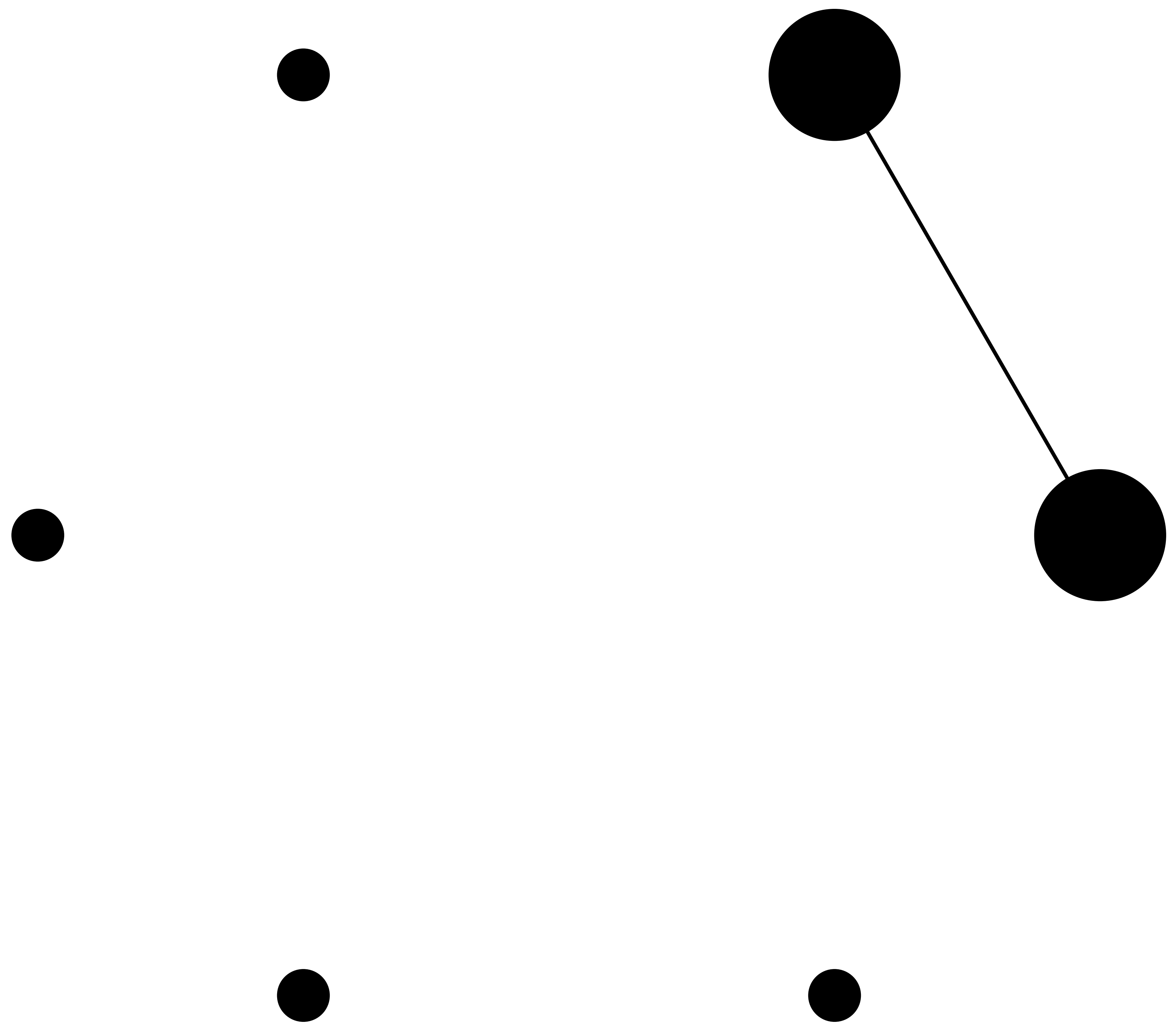
Erdos-Renyi graphs



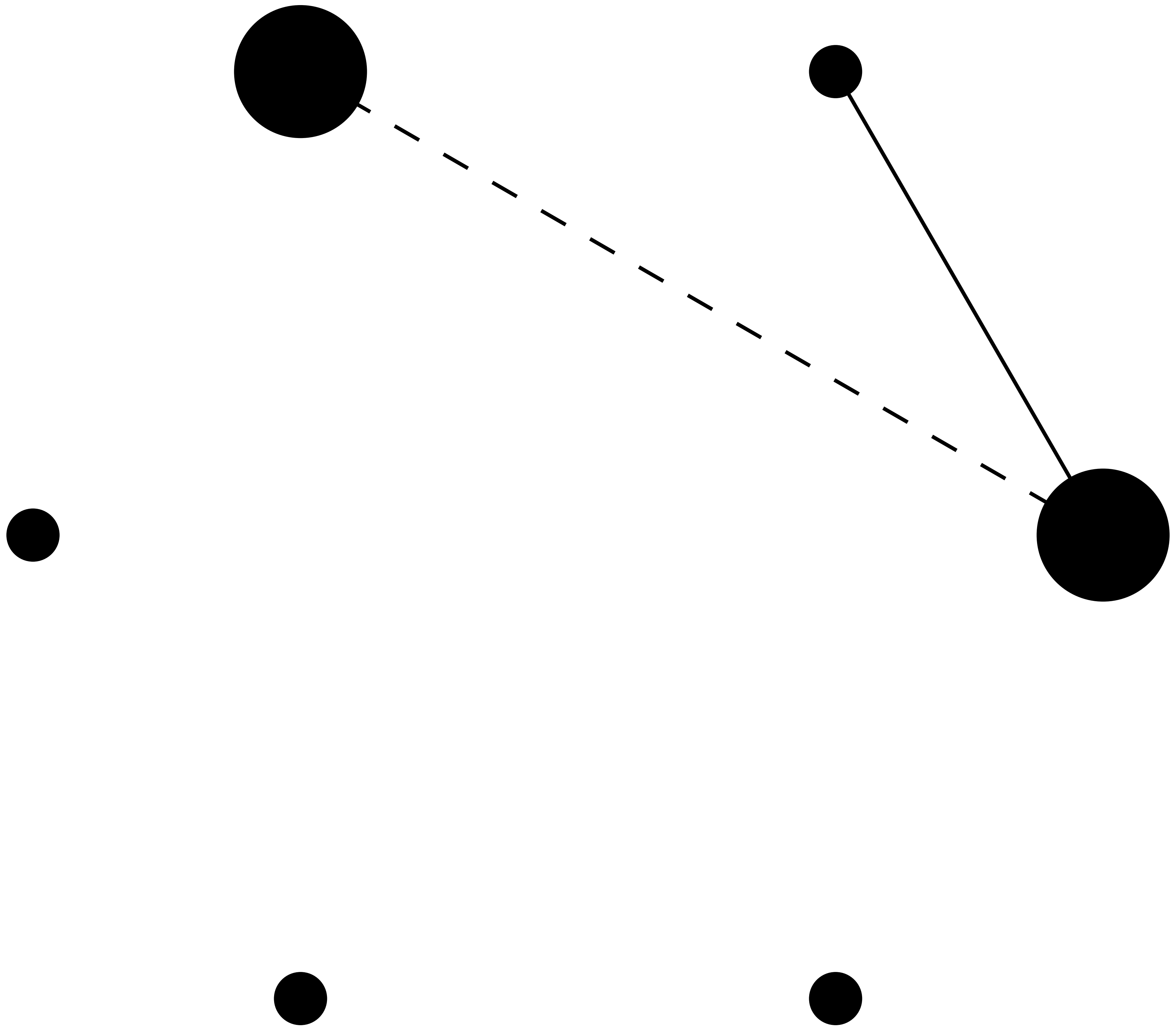
Erdos-Renyi graphs



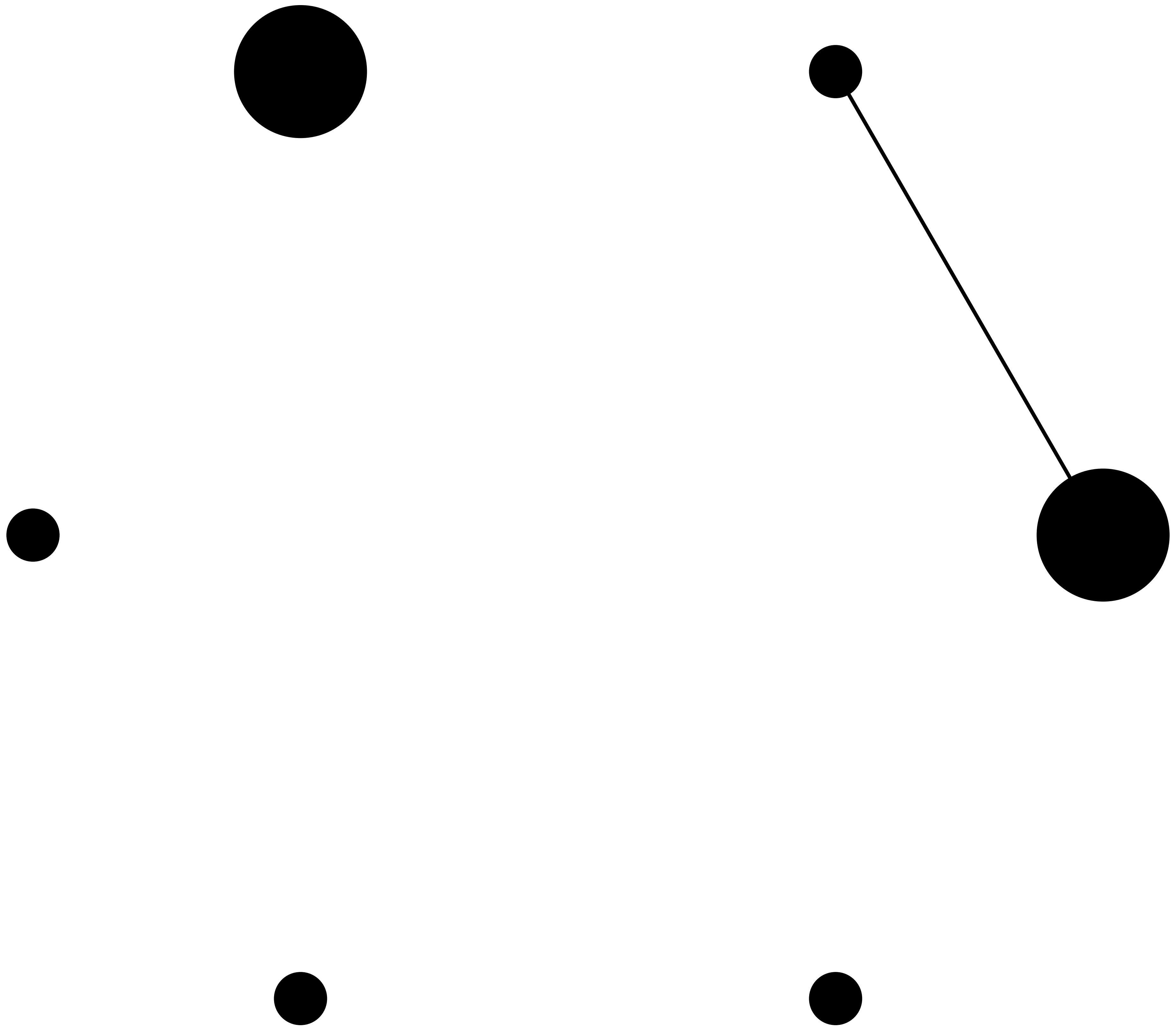
Erdos-Renyi graphs



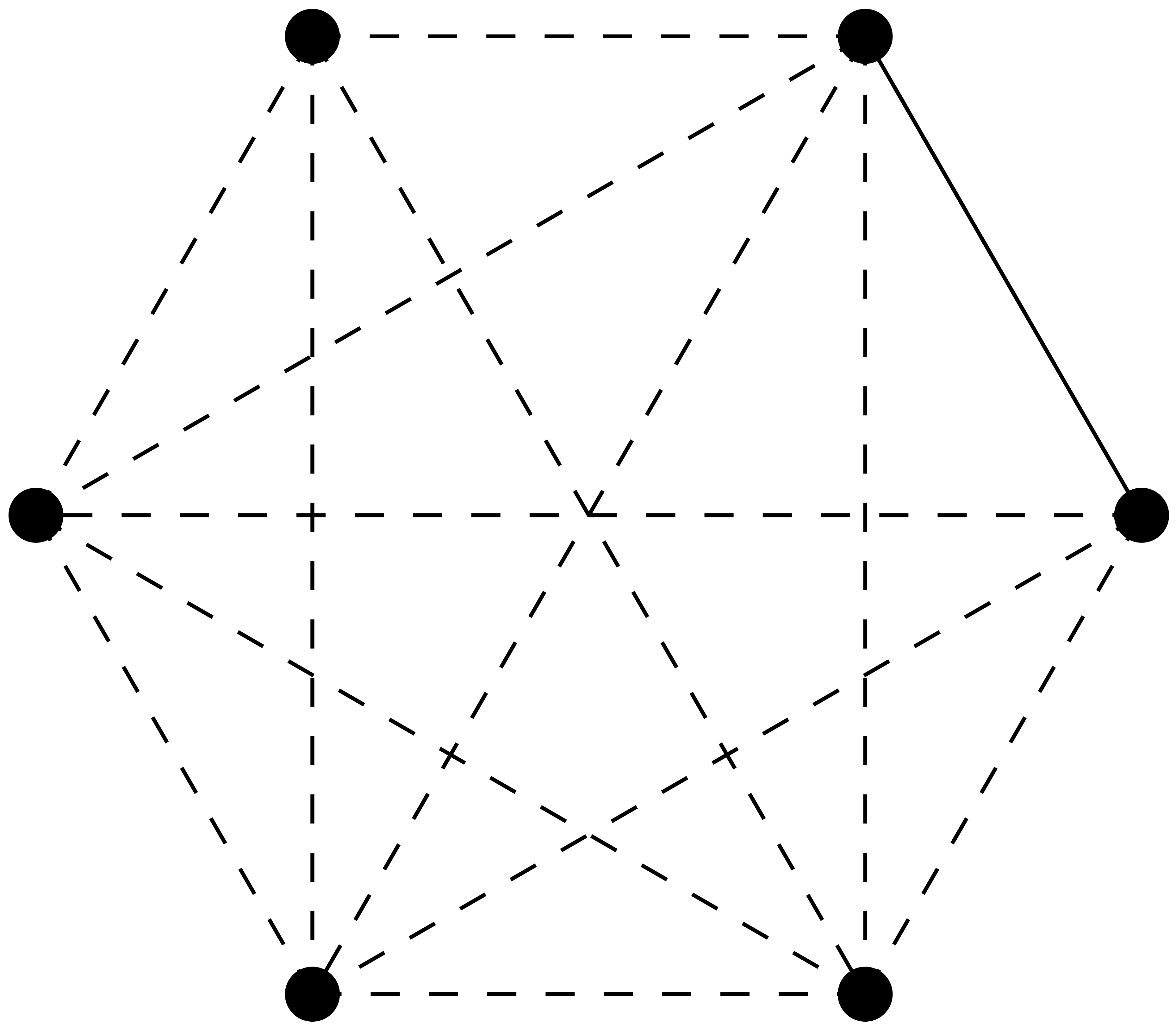
Erdos-Renyi graphs



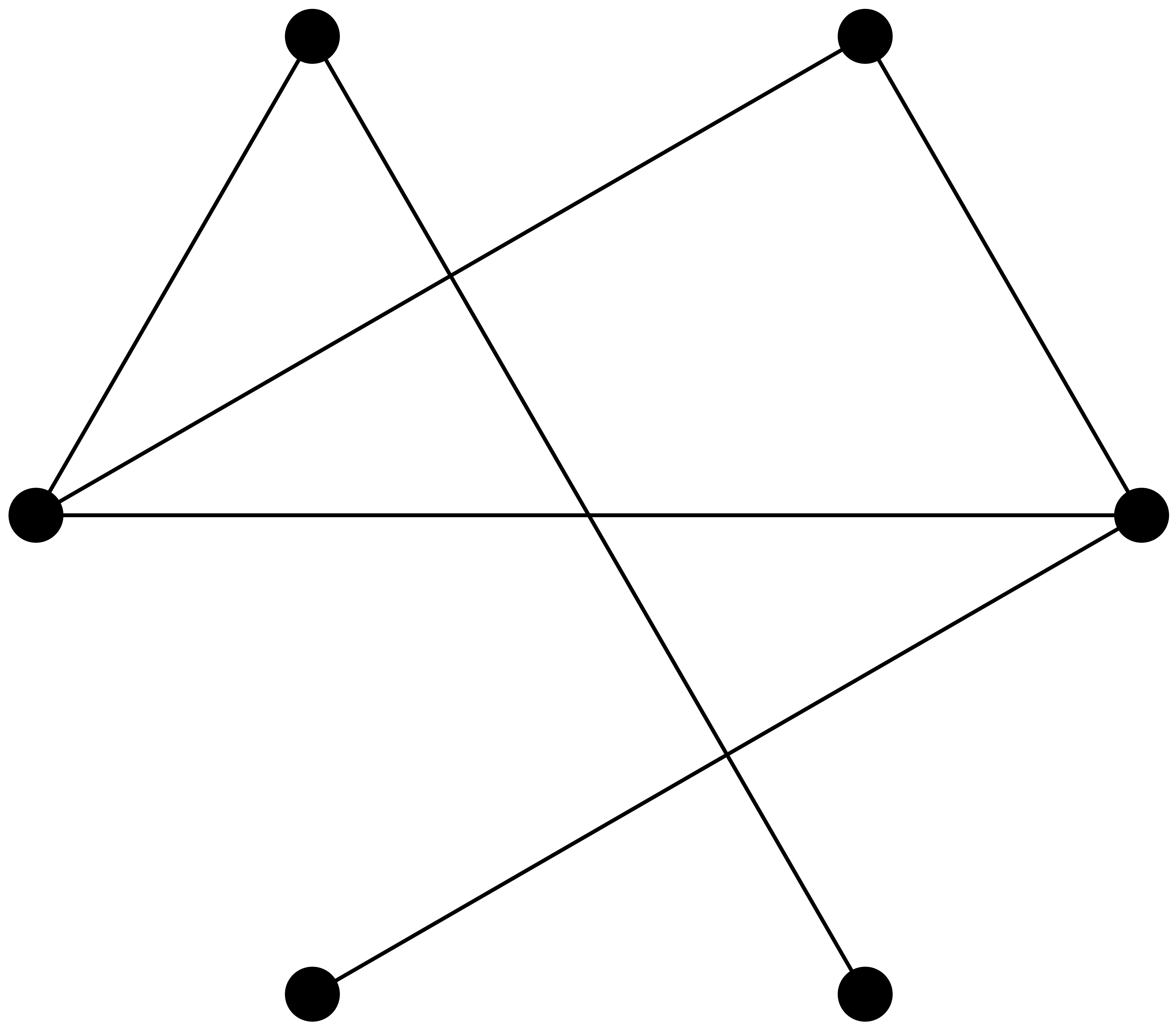
Erdos-Renyi graphs



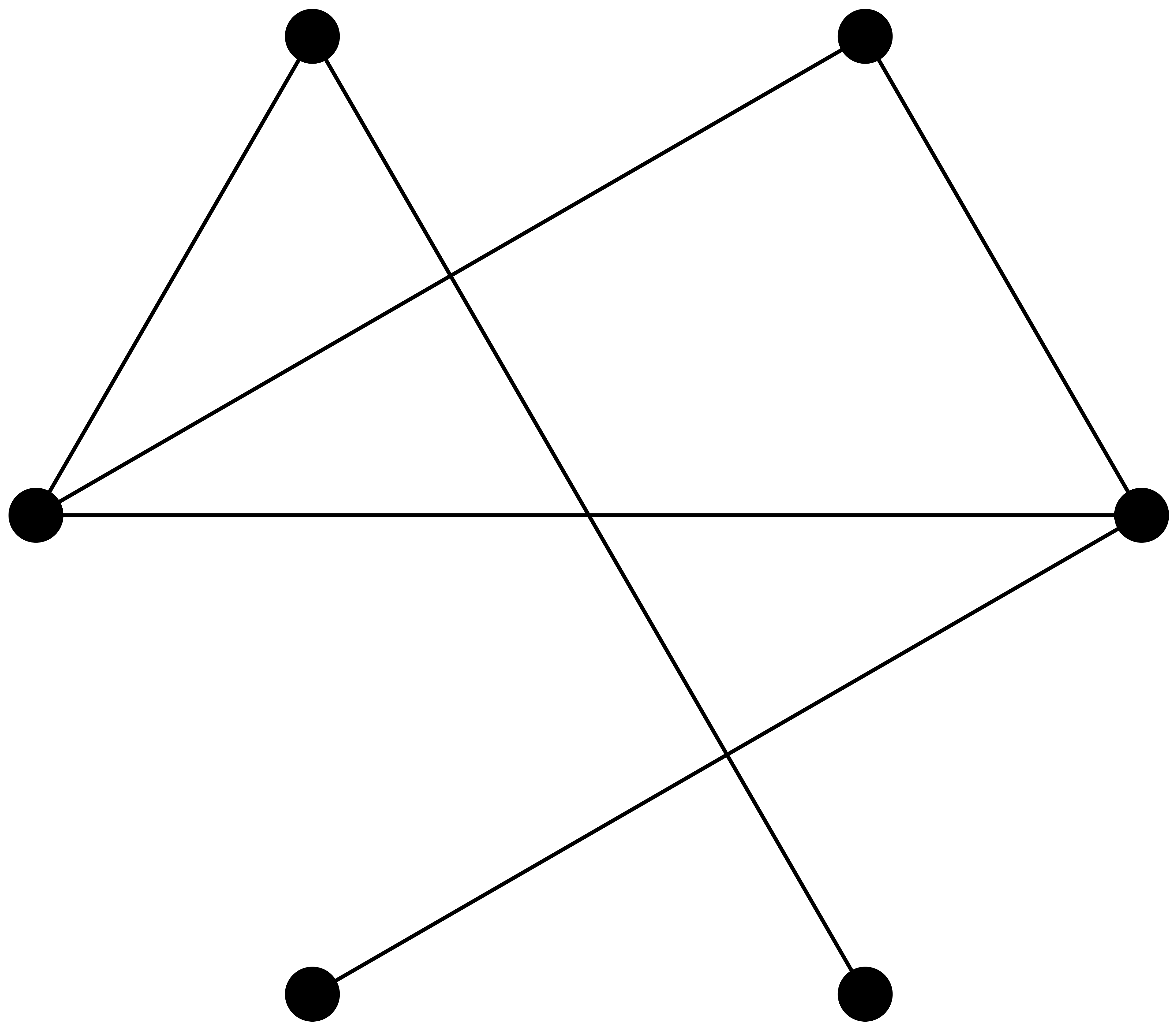
Erdos-Renyi graphs



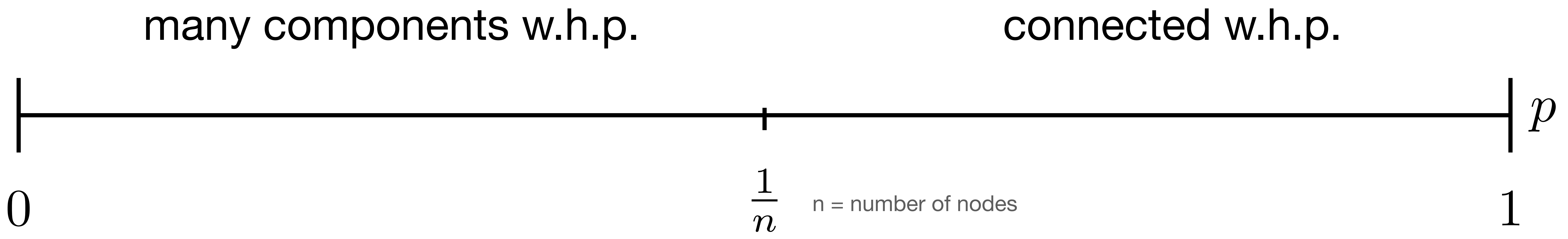
Erdos-Renyi graphs



Erdos-Renyi graphs

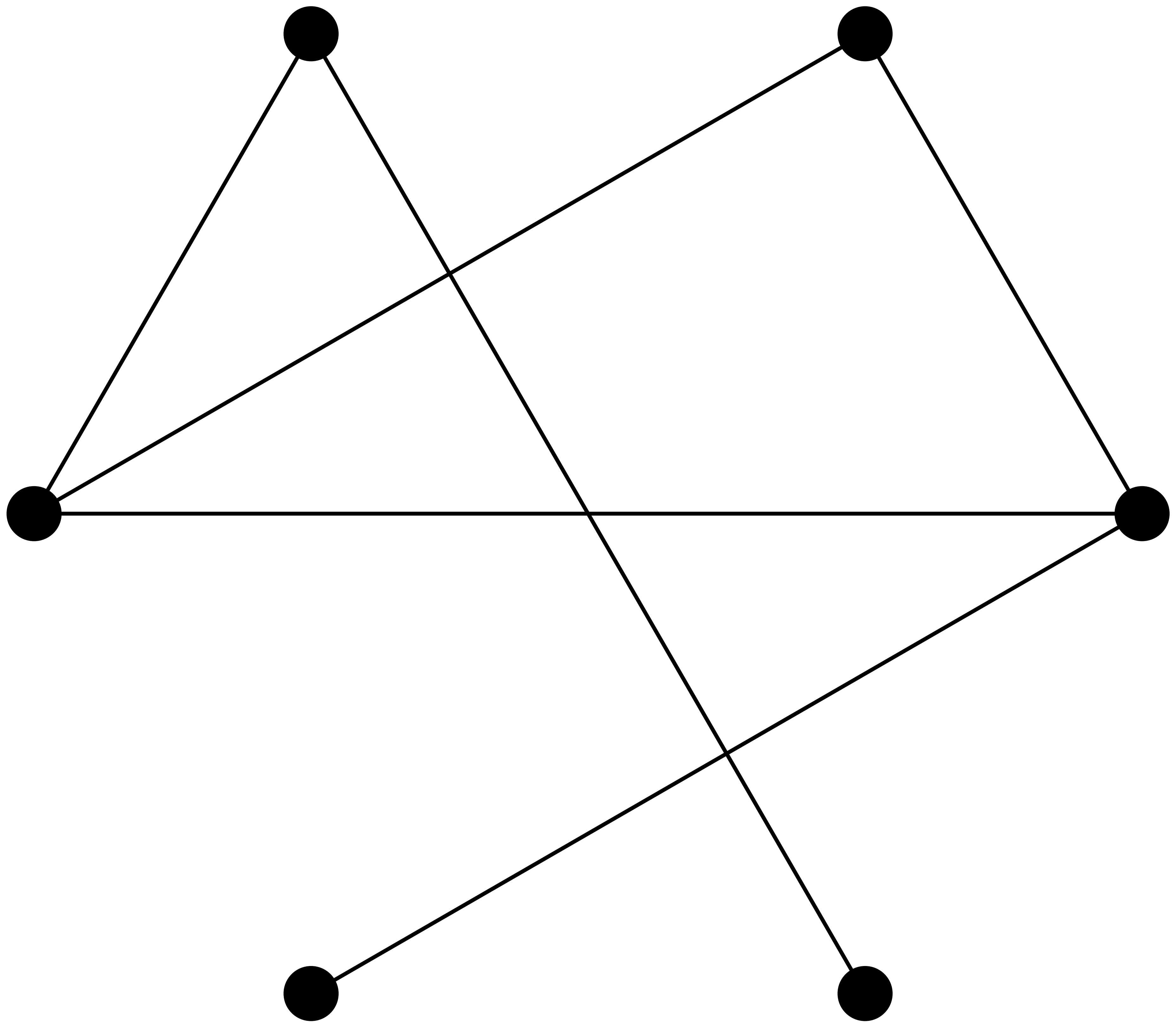


Phase Transition
[Erdos-Renyi 1960]

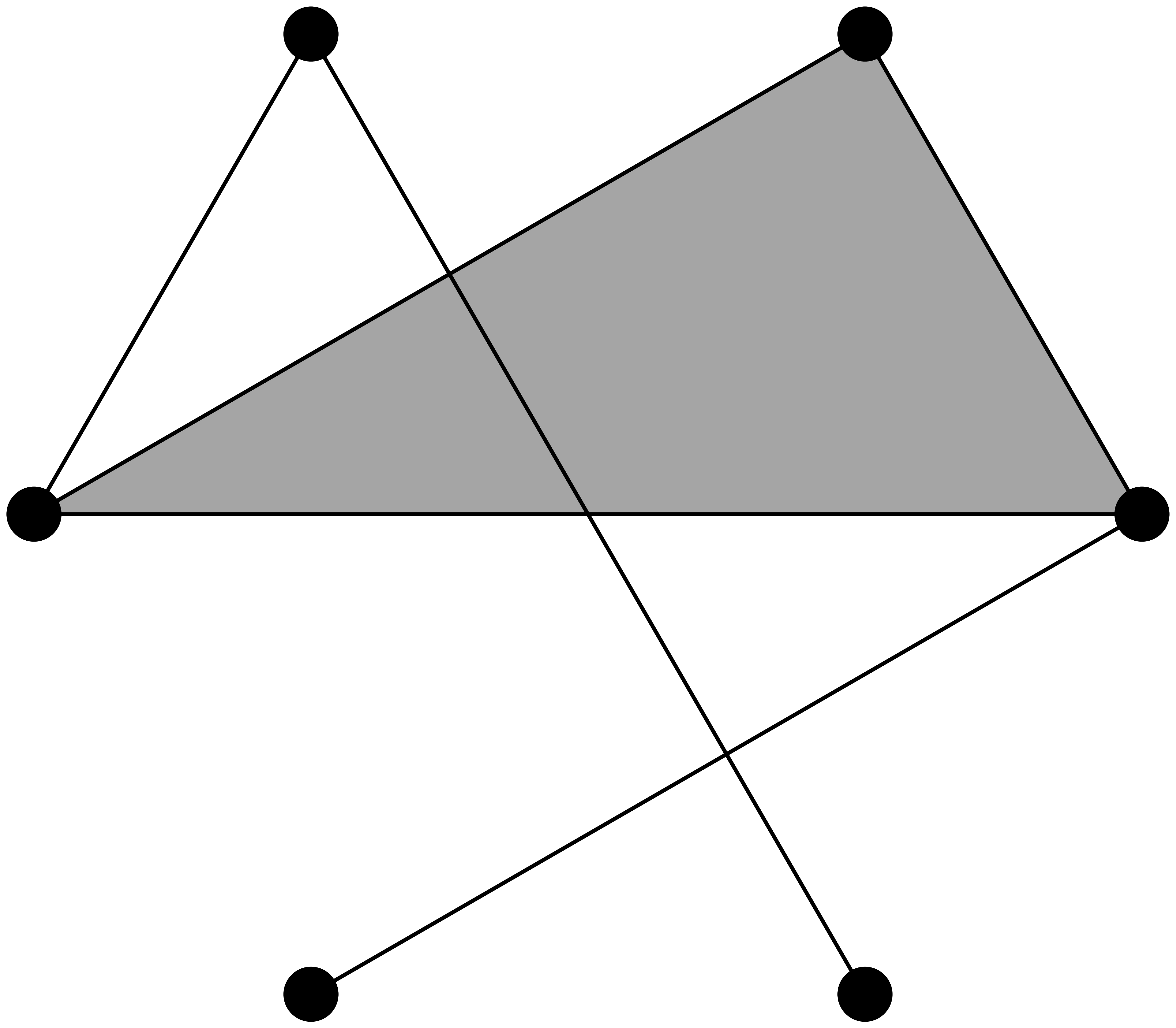


all log terms and constants forgone

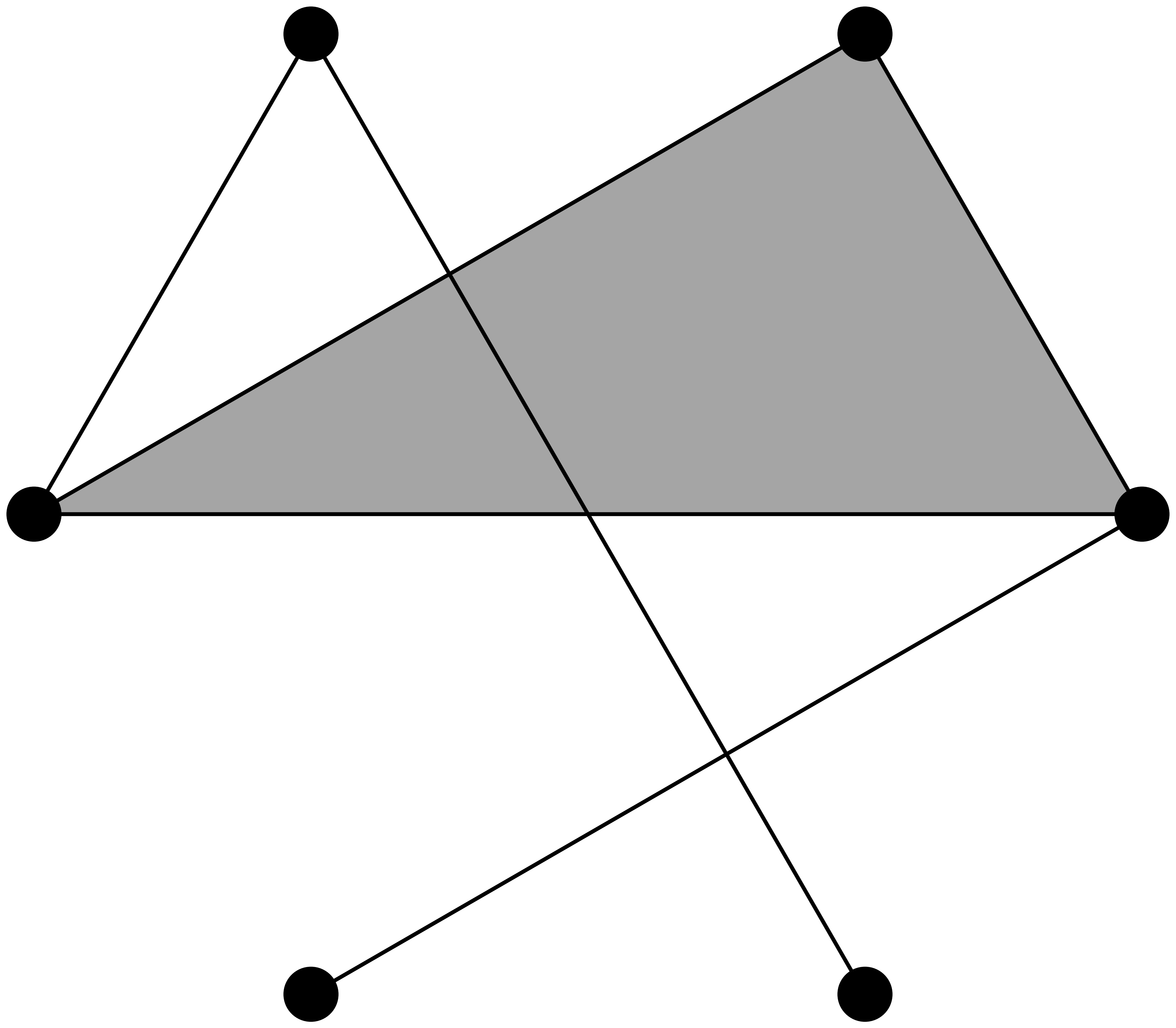
Erdos-Renyi Clique Complex

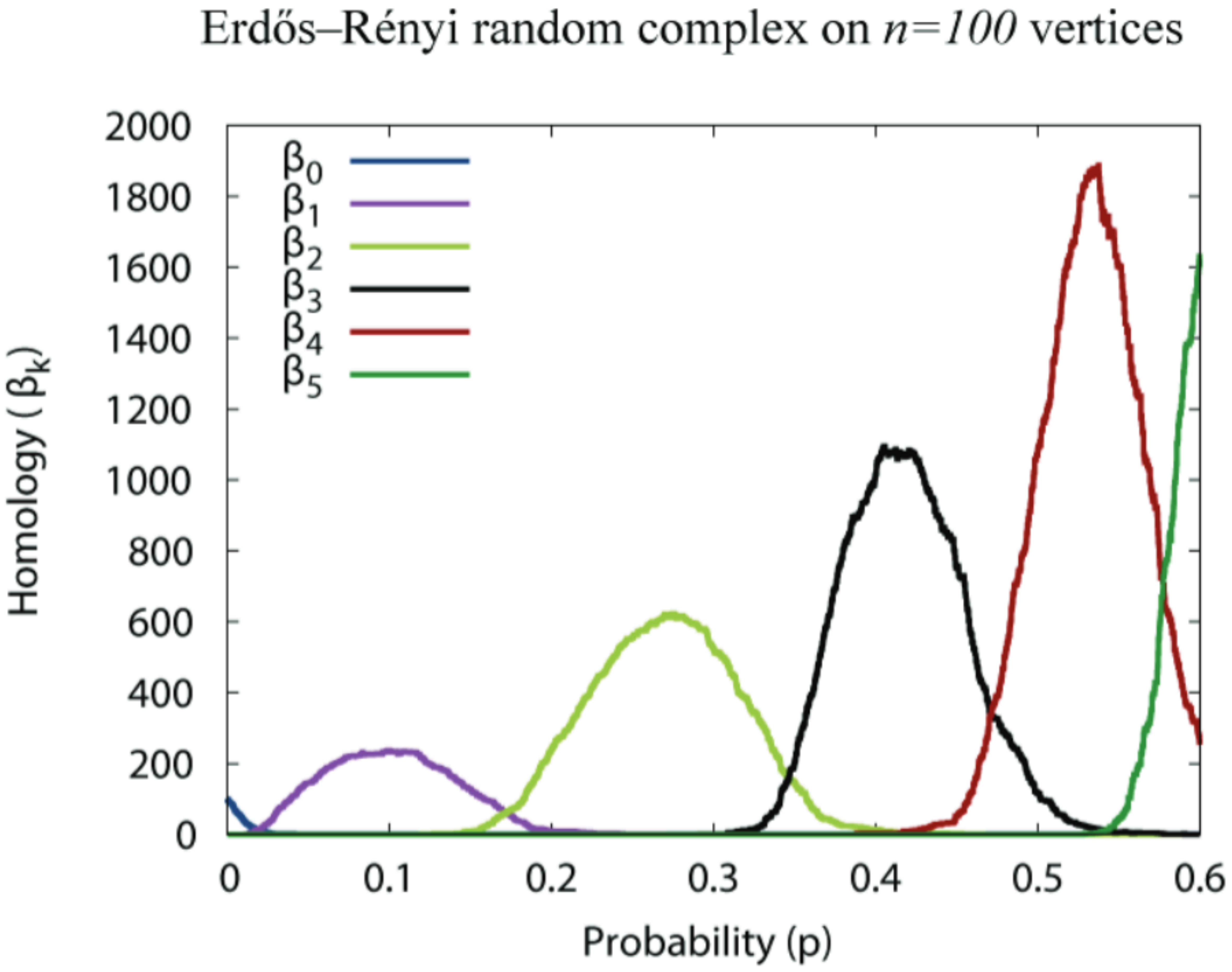


Erdos-Renyi Clique Complex



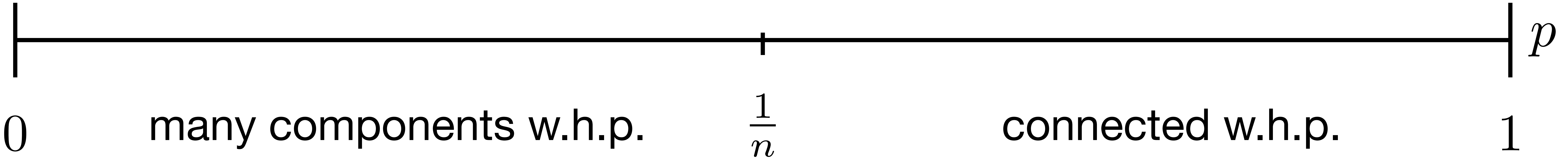
Betti Numbers





computation and plotting done by Zomorodian

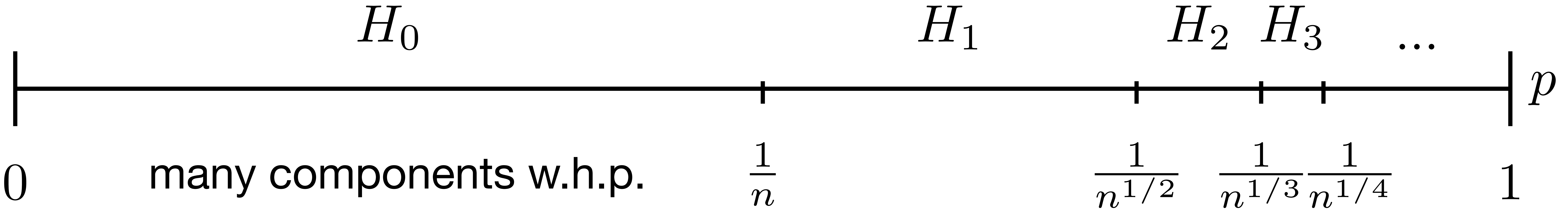
Phase Transition
[Erdos-Renyi 1960]



n = number of nodes
all log terms and constants forgone

Phase Transition

[Kahle 2009, 2014]

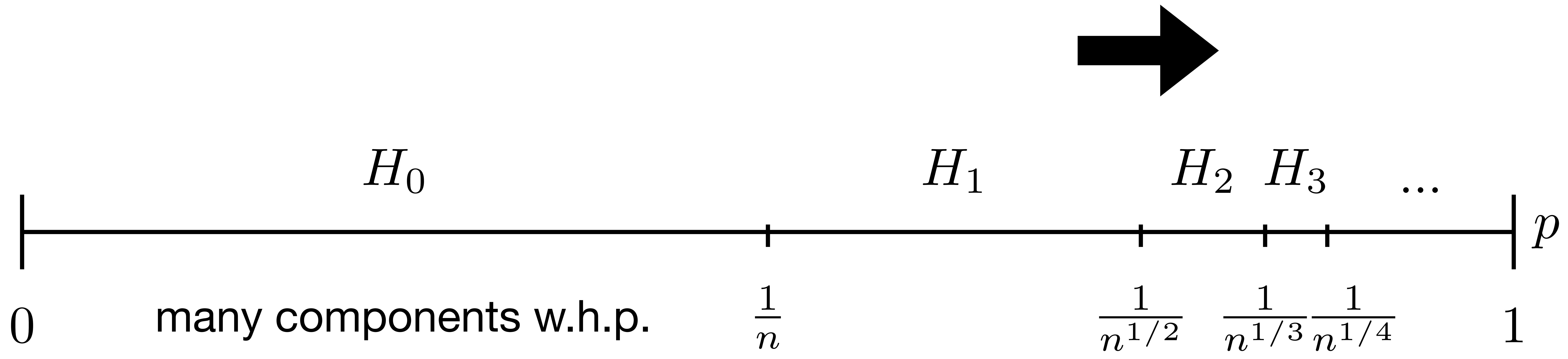


n = number of nodes
all log terms and constants forgone

Phase Transition

[Kahle 2009, 2014]

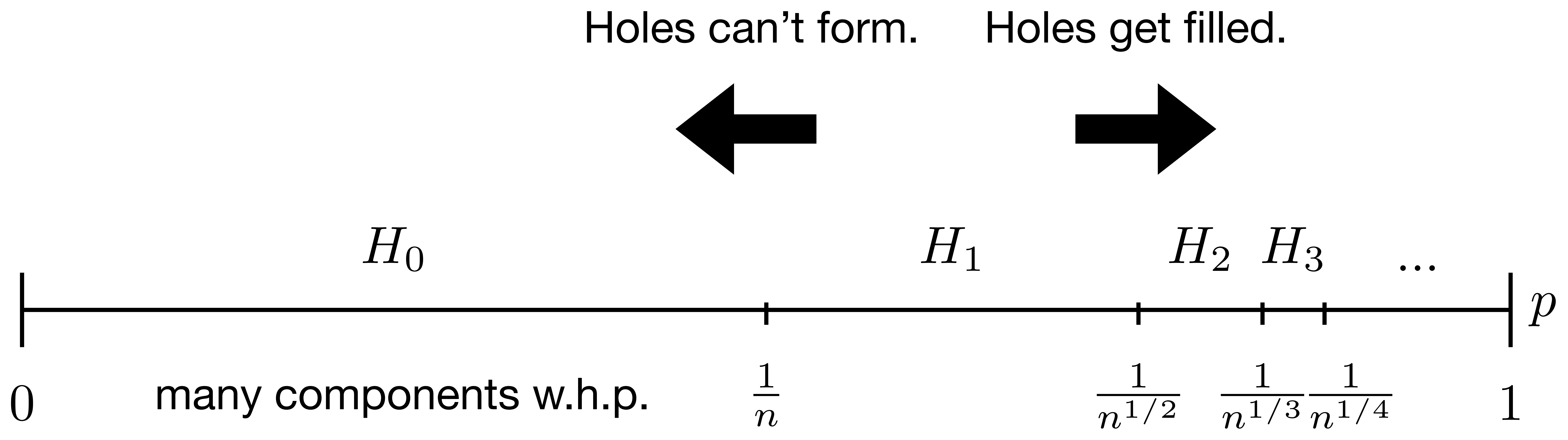
Holes get filled.



n = number of nodes
all log terms and constants forgone

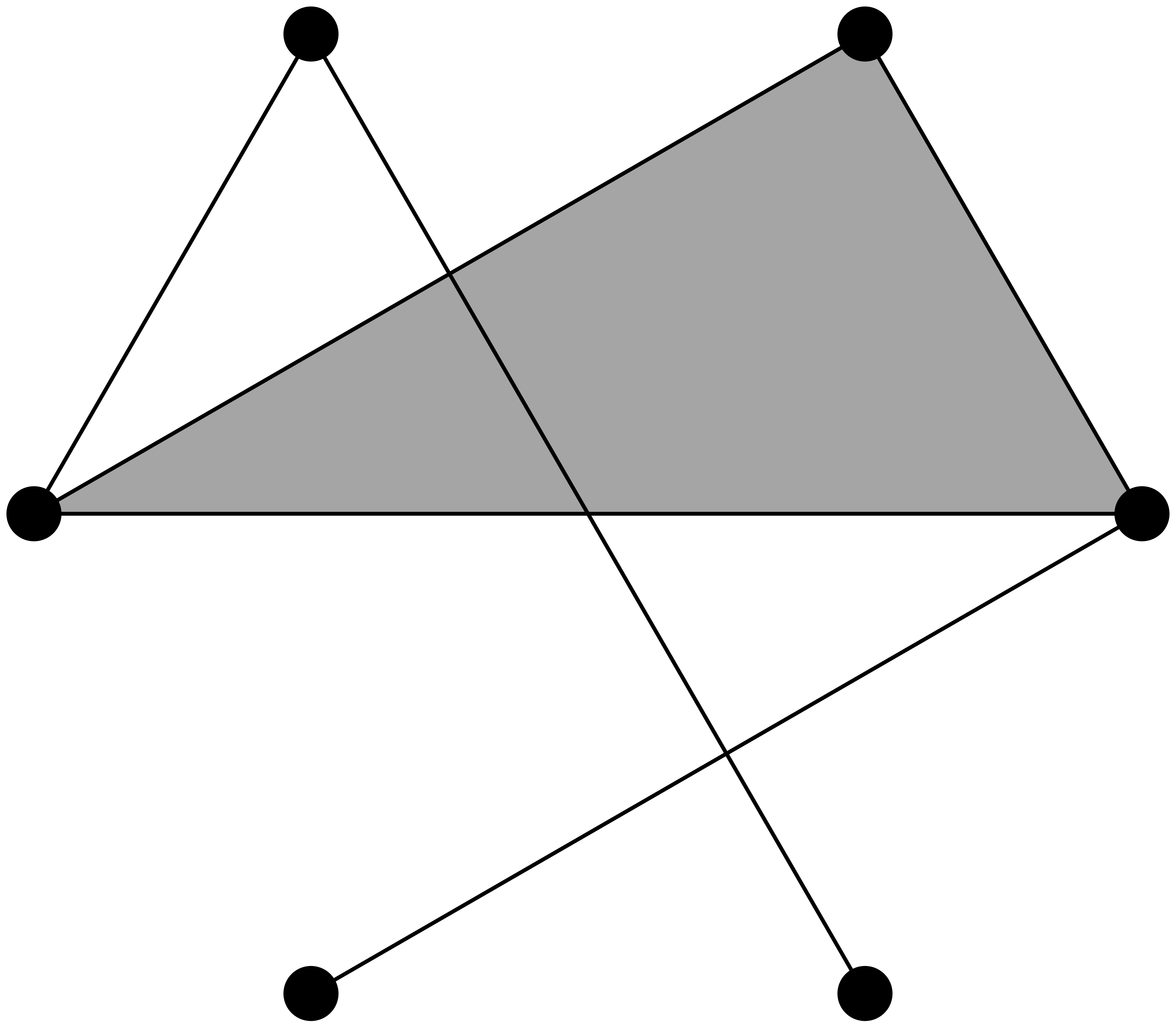
Phase Transition

[Kahle 2009, 2014]



n = number of nodes
all log terms and constants forgone

Erdos-Renyi Clique Complex



Geometric Complexes

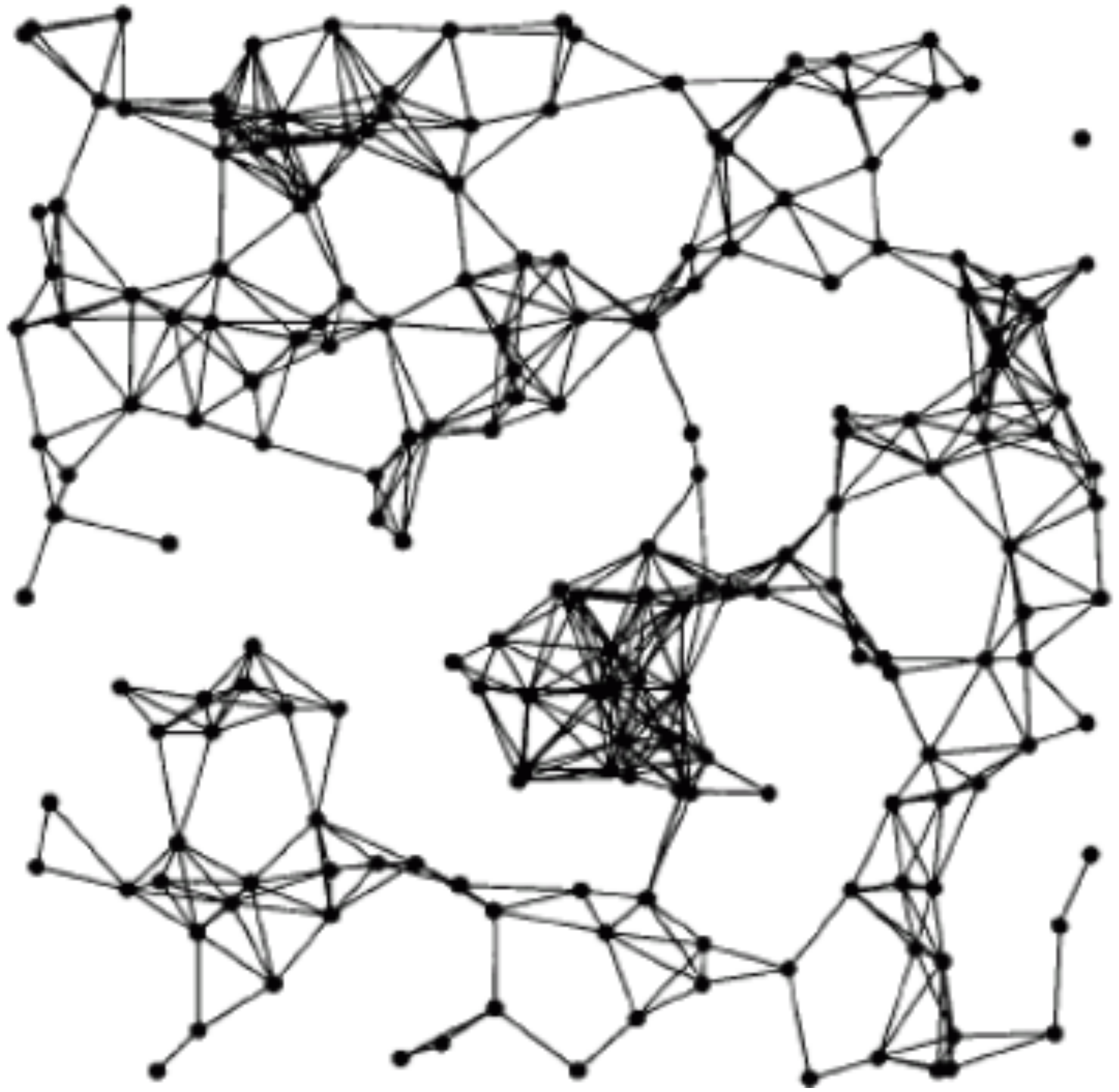


image credit: Penrose

Geometric Complexes

- Rips
- Cech

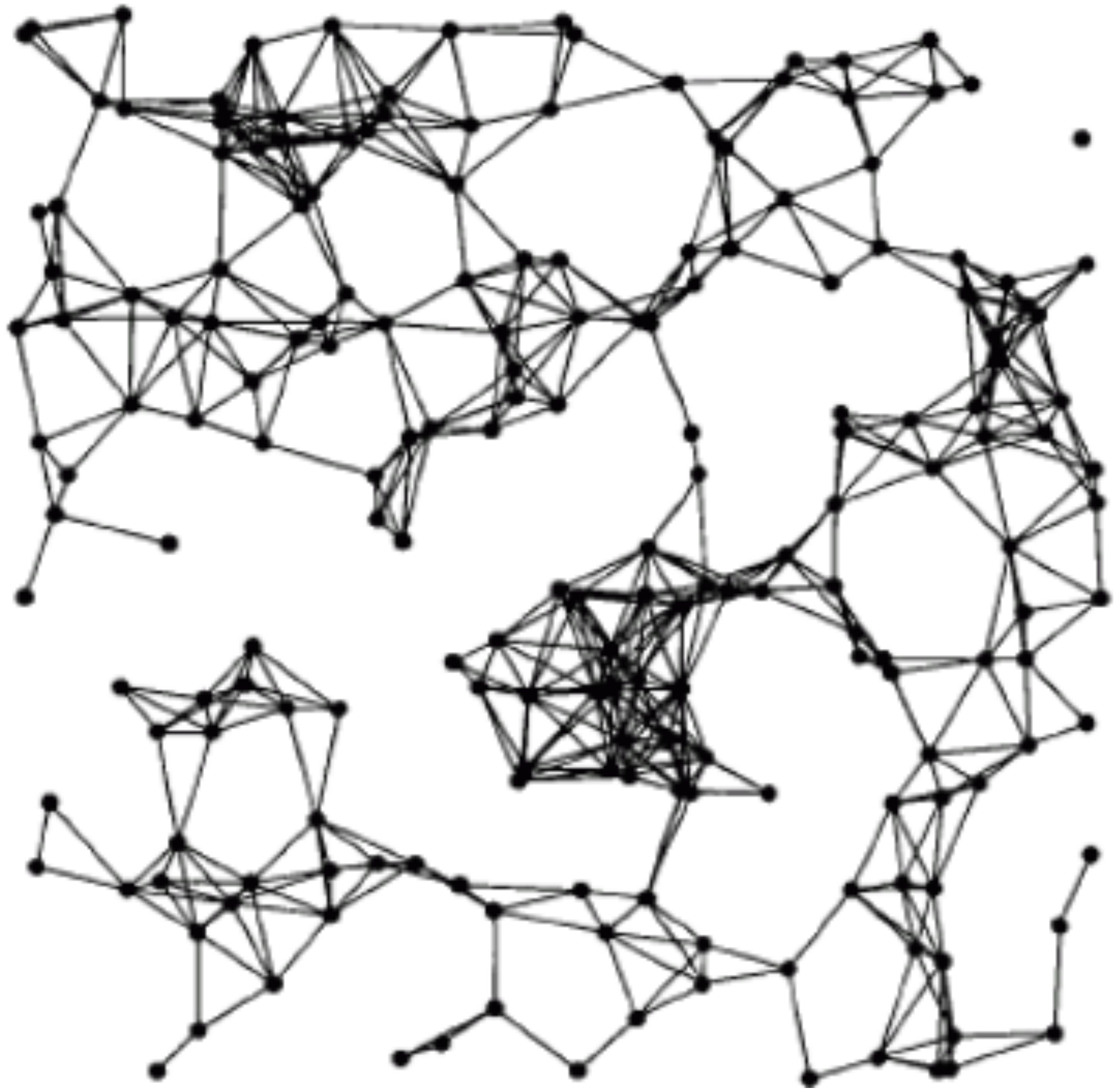


image credit: Penrose

Geometric Complexes

- Rips (clique)
- Cech

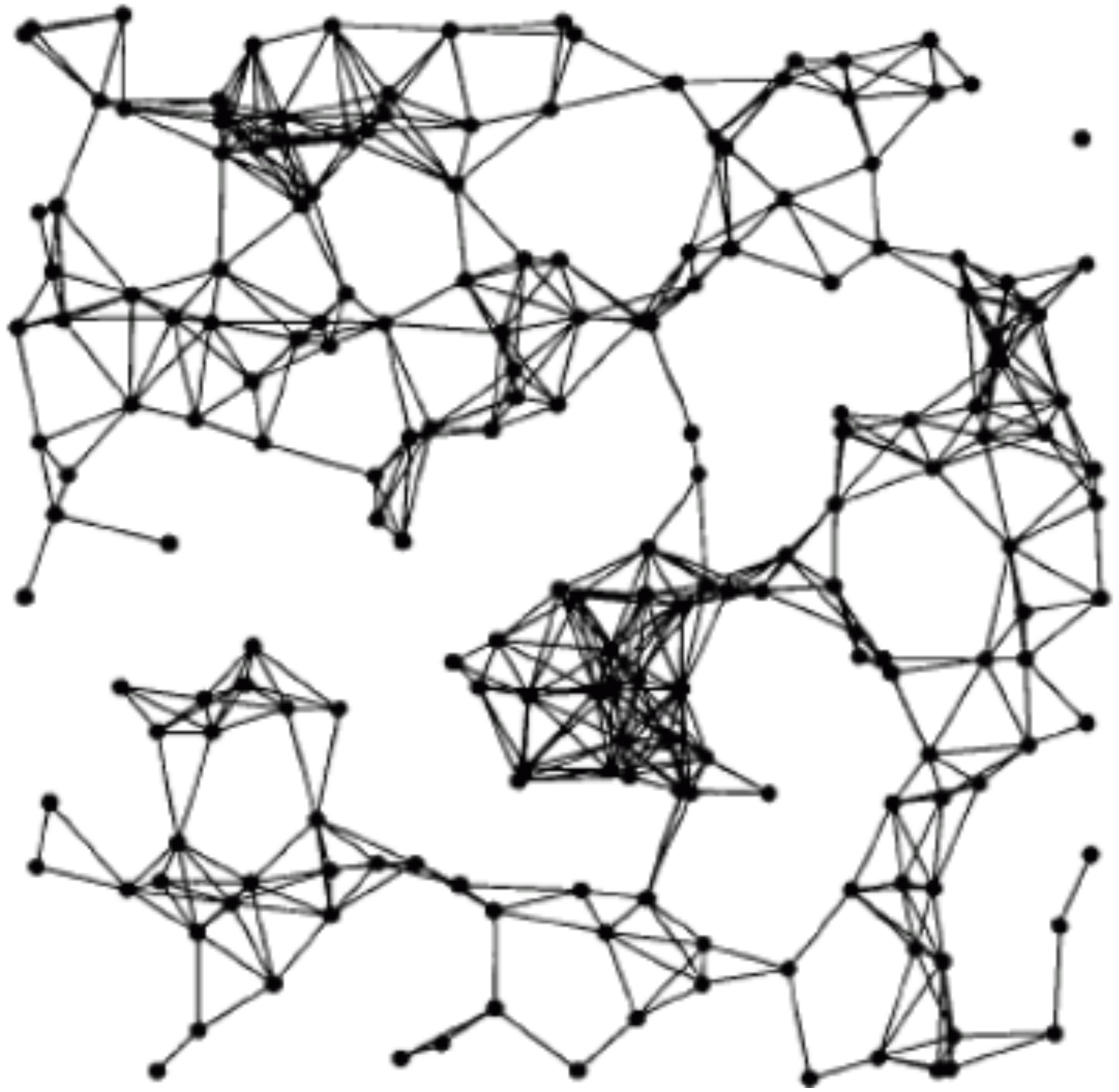
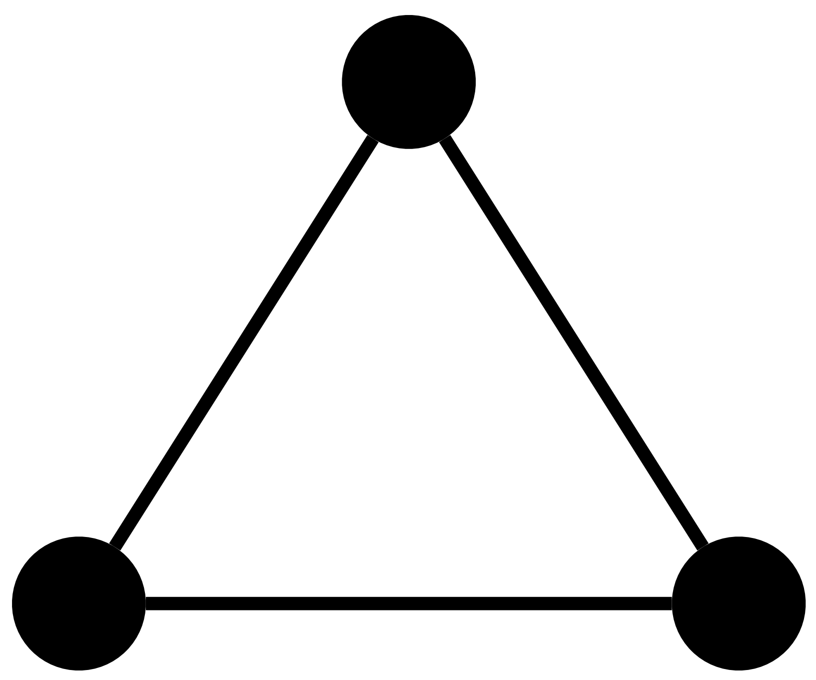


image credit: Penrose

Geometric Complexes

- Rips (clique)
- Cech



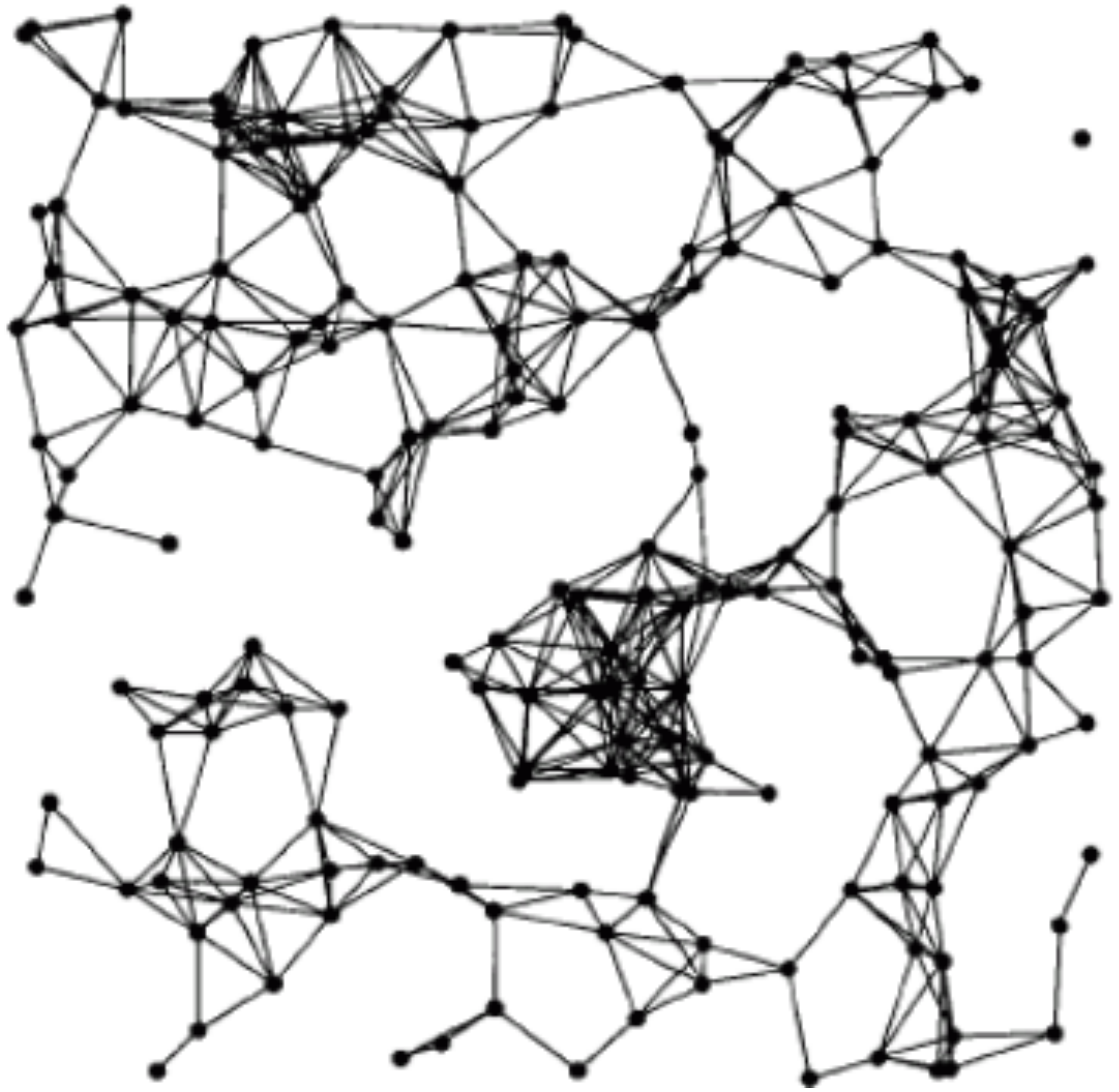
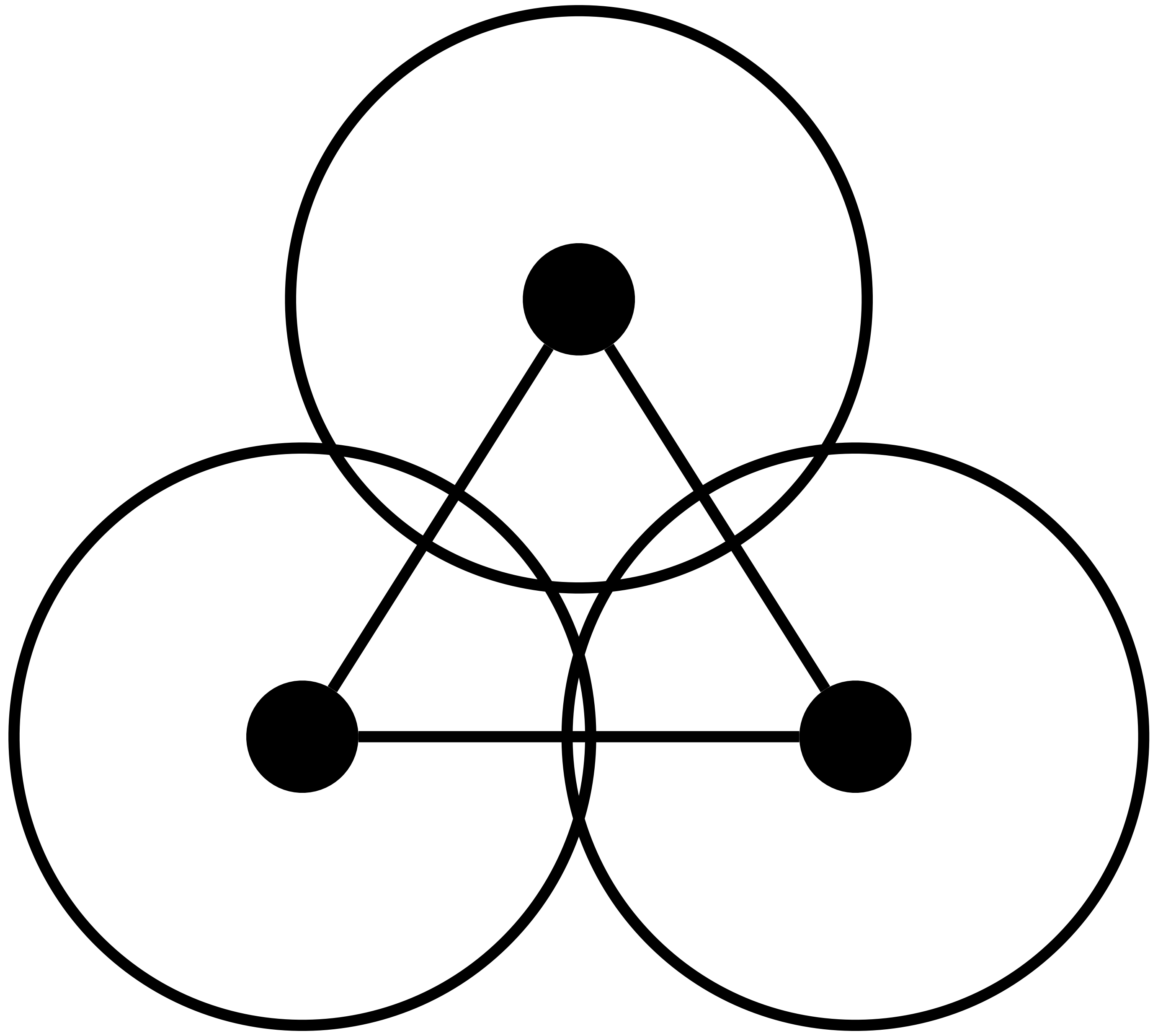


image credit: Penrose

Geometric Complexes

- Rips (clique)
- Cech



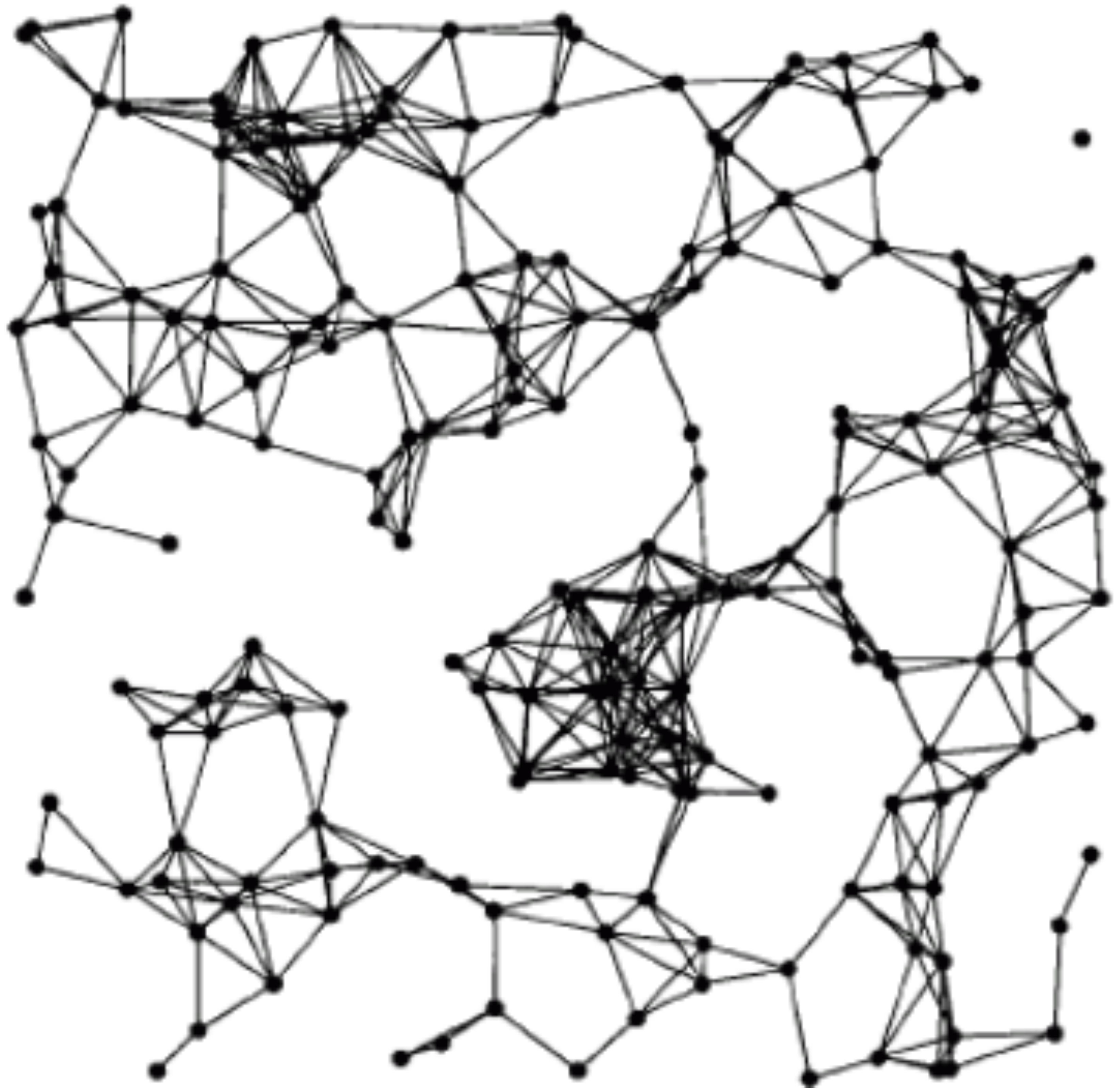


image credit: Penrose

Expected Betti numbers at dimension k

[Kahle 2011]

Expected Betti numbers at dimension k
[Kahle 2011]

- n, the number of points

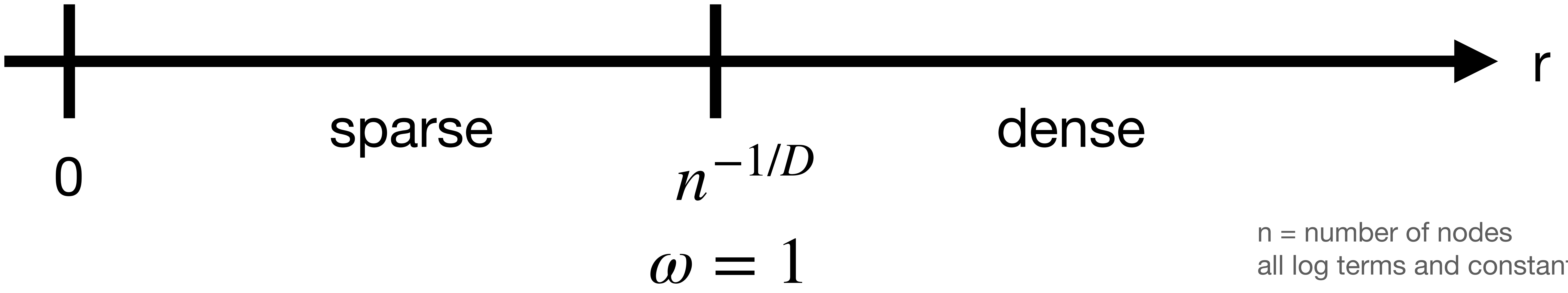
Expected Betti numbers at dimension k

[Kahle 2011]

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

Expected Betti numbers at dimension k
[Kahle 2011]

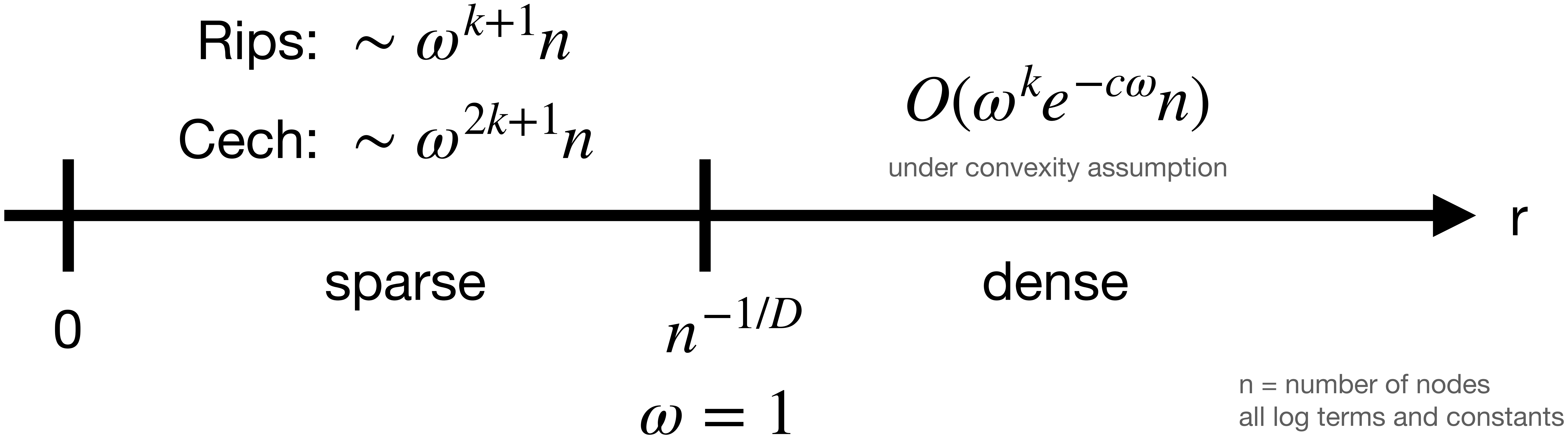
- n, the number of points
- $\omega = nr^D$, where D is the ambient dimension



n = number of nodes
all log terms and constants forgone

Expected Betti numbers at dimension k
[Kahle 2011]

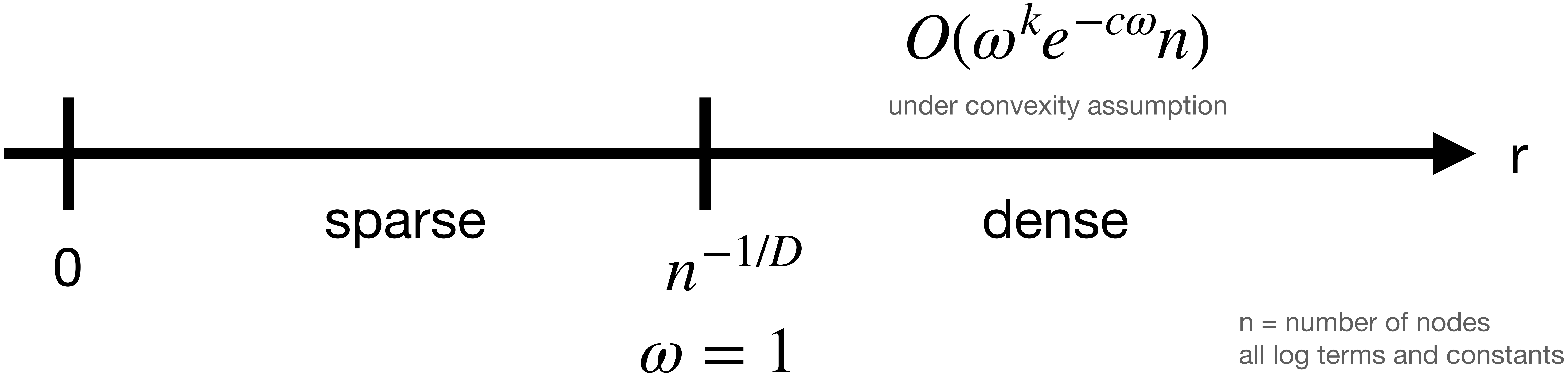
- n, the number of points
- $\omega = nr^D$, where D is the ambient dimension



n = number of nodes
all log terms and constants forgone

Expected Betti numbers at dimension k
[Kahle 2011]

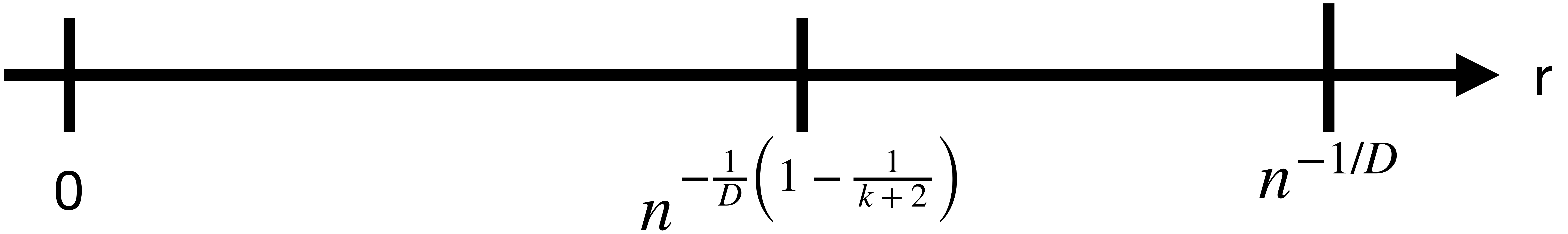
- n, the number of points
- $\omega = nr^D$, where D is the ambient dimension
- $E\beta_k(\text{Cech}) \sim \omega^{2k+1}n$



n = number of nodes
all log terms and constants forgone

Expected Betti numbers at dimension k
[Kahle 2011]

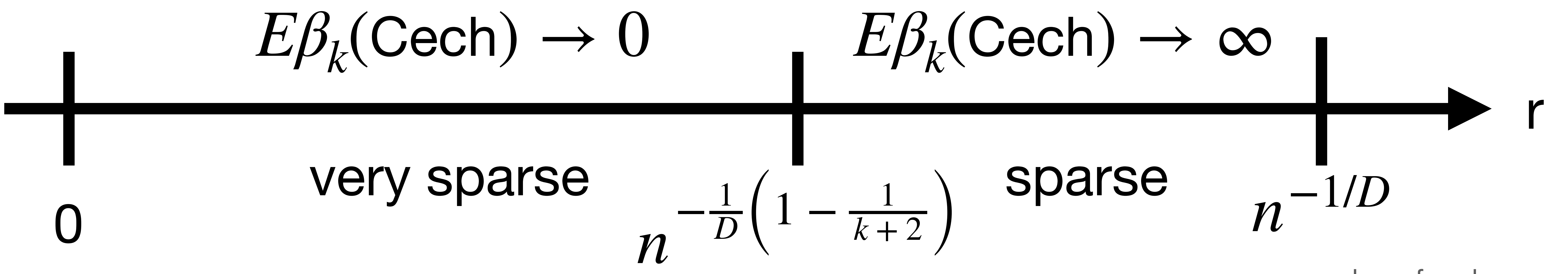
- n, the number of points
- $\omega = nr^D$, where D is the ambient dimension
- $E\beta_k(\text{Cech}) \sim \omega^{2k+1}n$



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

Functional Convergence at dimension k?

[Thomas and Owada 2020]

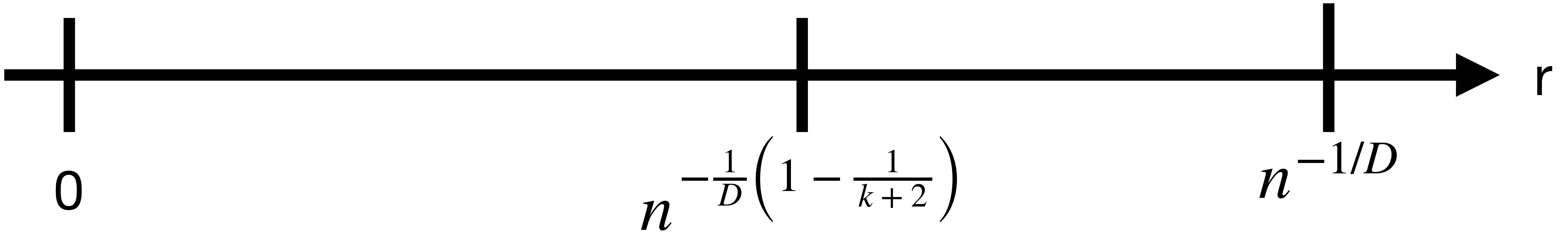
Functional Convergence at dimension k?
[Thomas and Owada 2020]

- Cech: weak convergence in finite-dimensional sense

Functional Convergence at dimension k?

[Thomas and Owada 2020]

- Cech: weak convergence in finite-dimensional sense

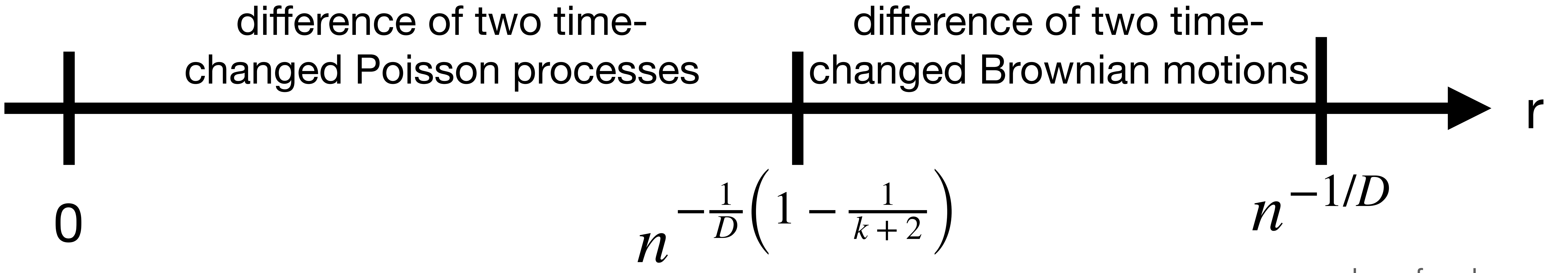


n = number of nodes
all log terms and constants forgone

Functional Convergence at dimension k?

[Thomas and Owada 2020]

- Cech: weak convergence in finite-dimensional sense

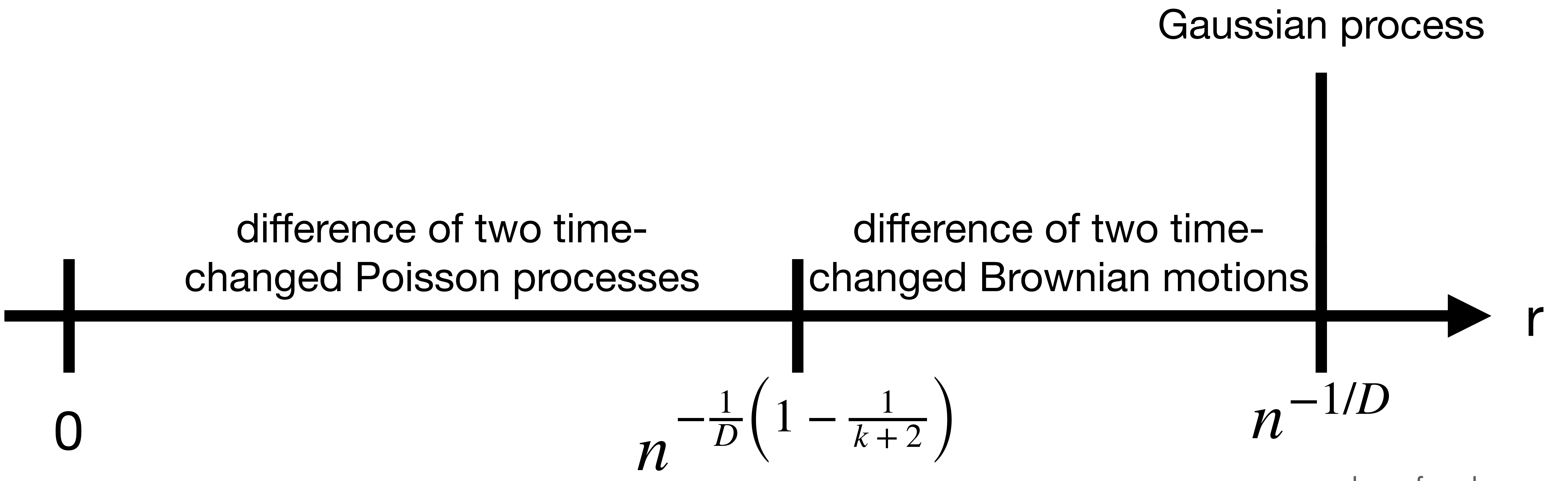


n = number of nodes
all log terms and constants forgone

Functional Convergence at dimension k?

[Thomas and Owada 2020]

- Cech: weak convergence in finite-dimensional sense



n = number of nodes
all log terms and constants forgone

Geometric Complexes

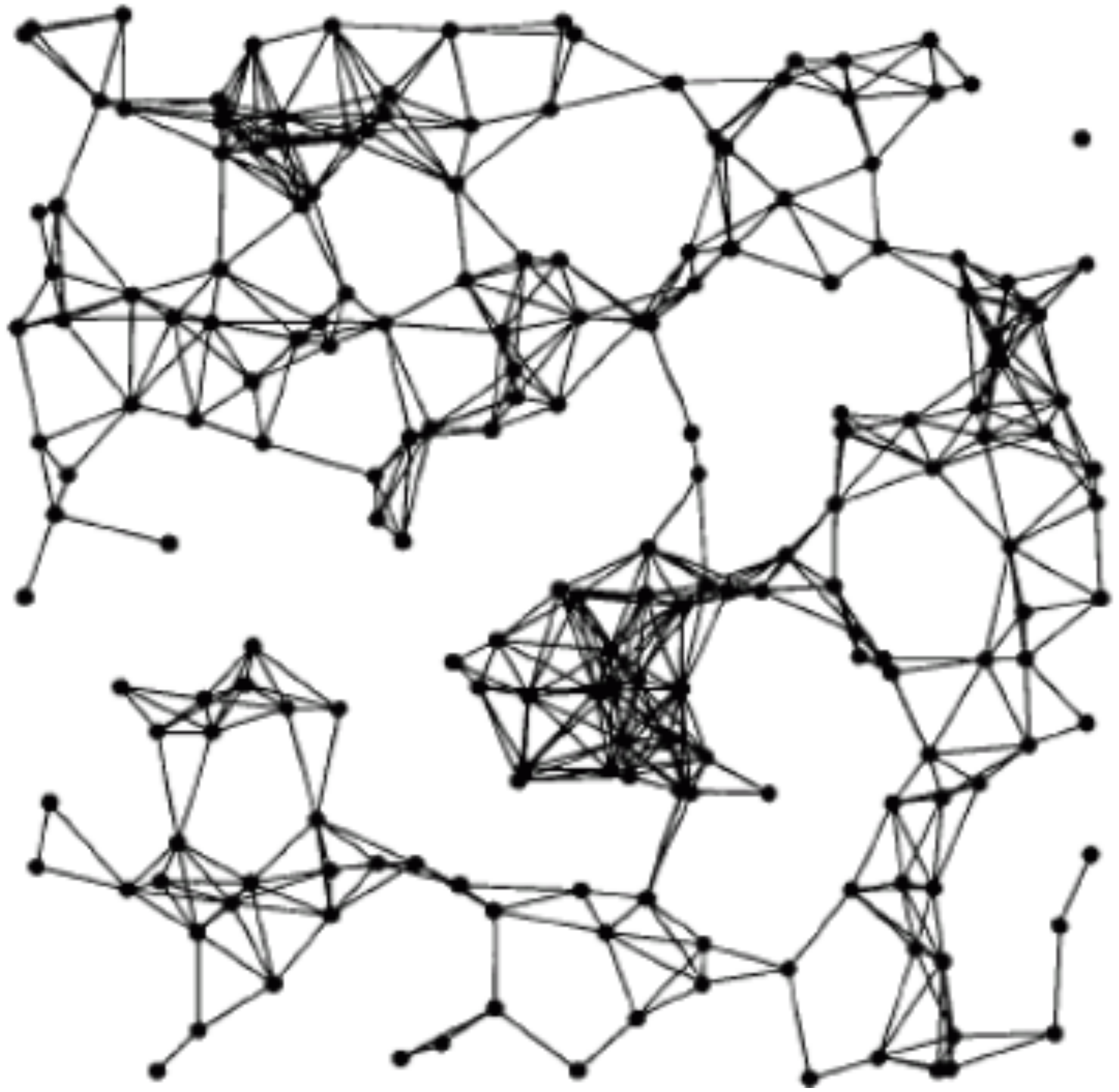
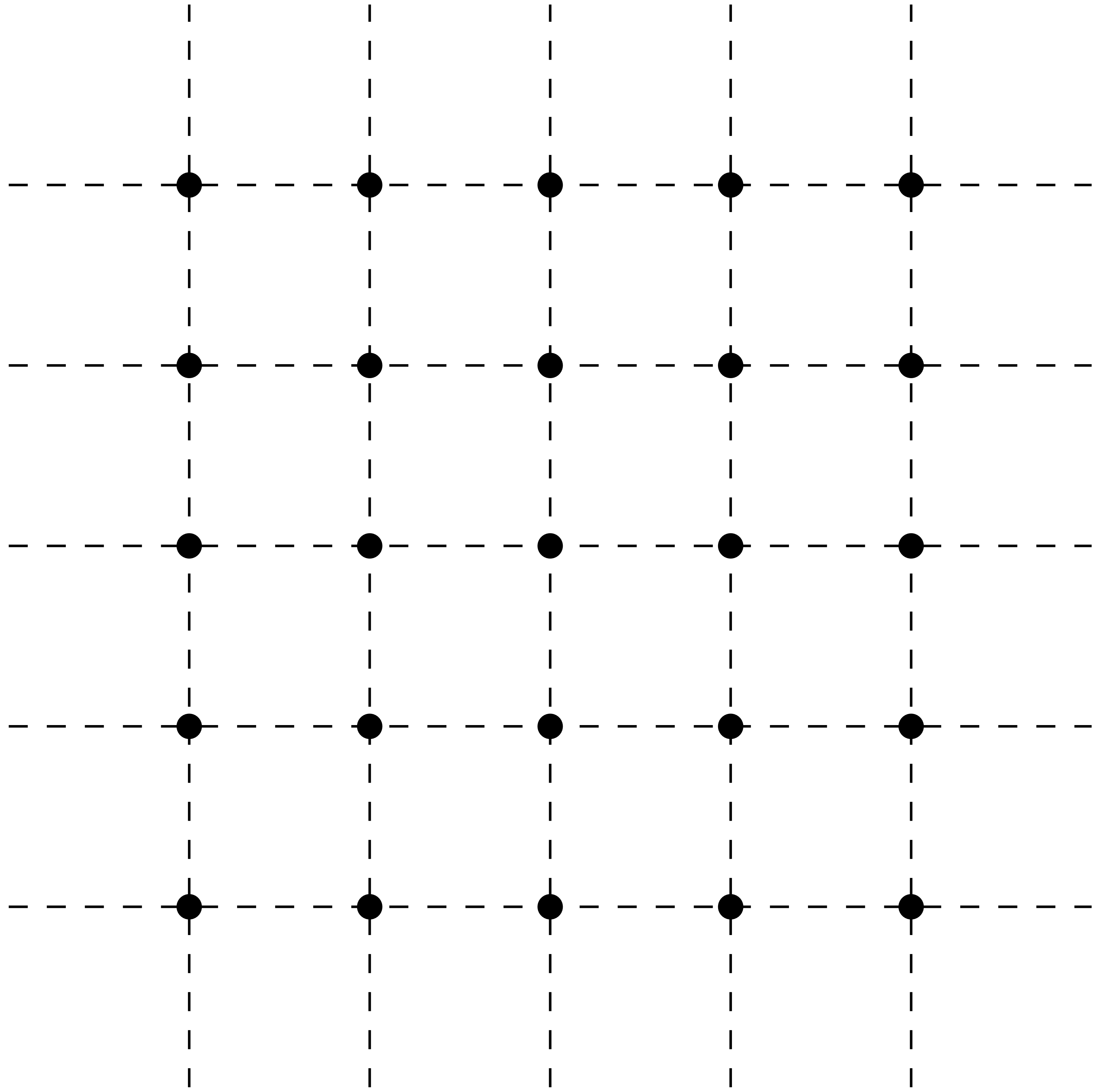
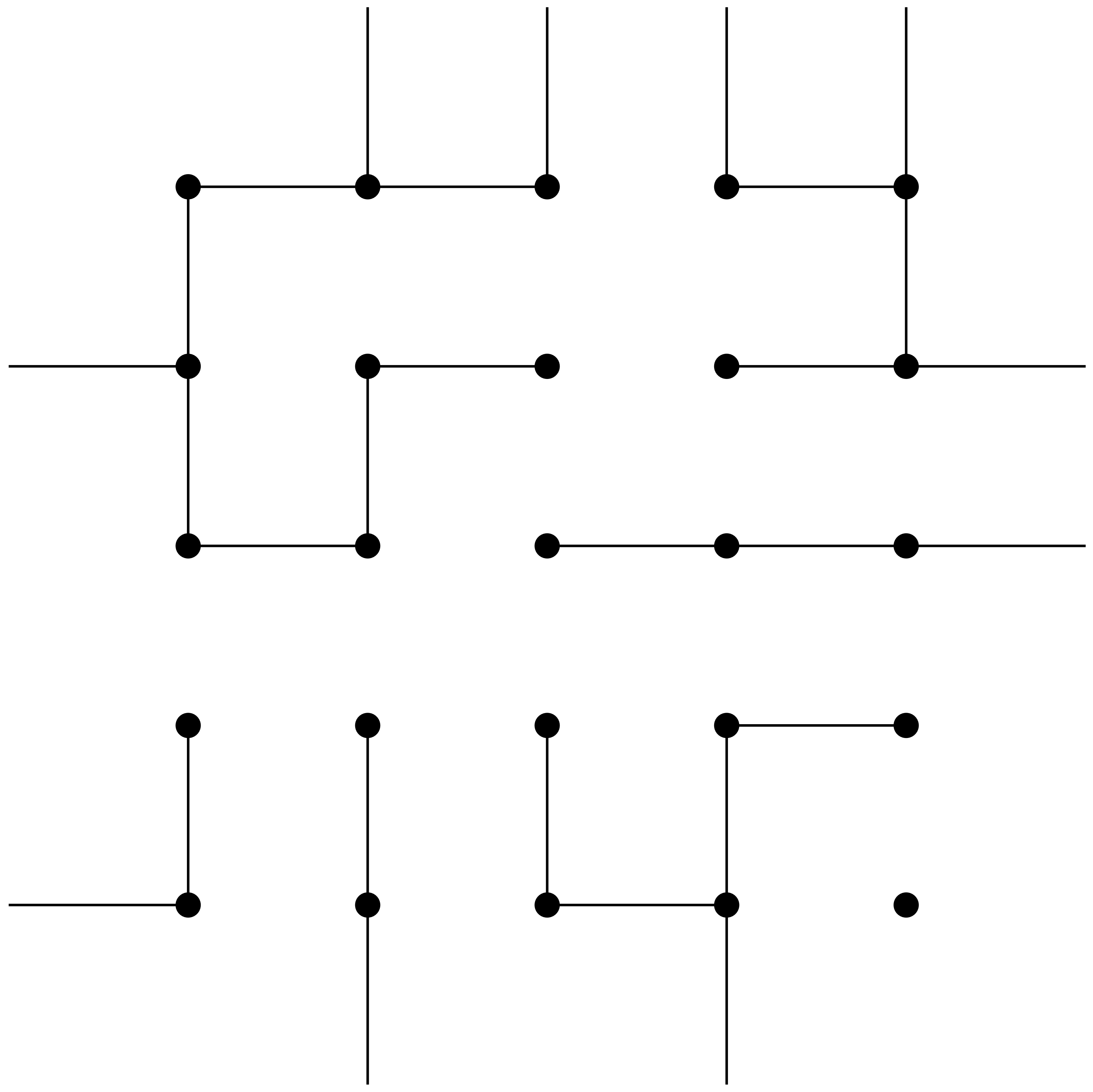


image credit: Penrose

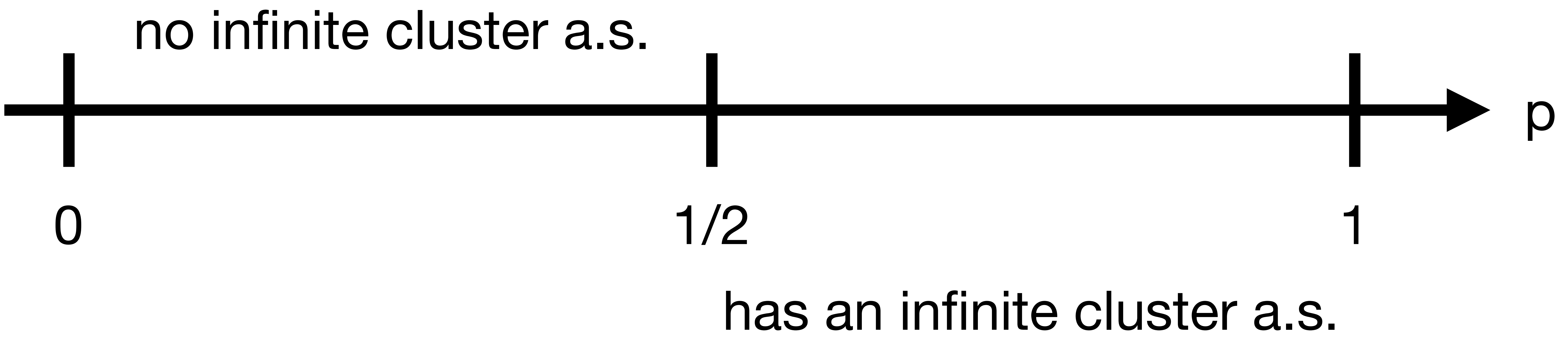
Bernoulli Bond Percolation



Bernoulli Bond Percolation

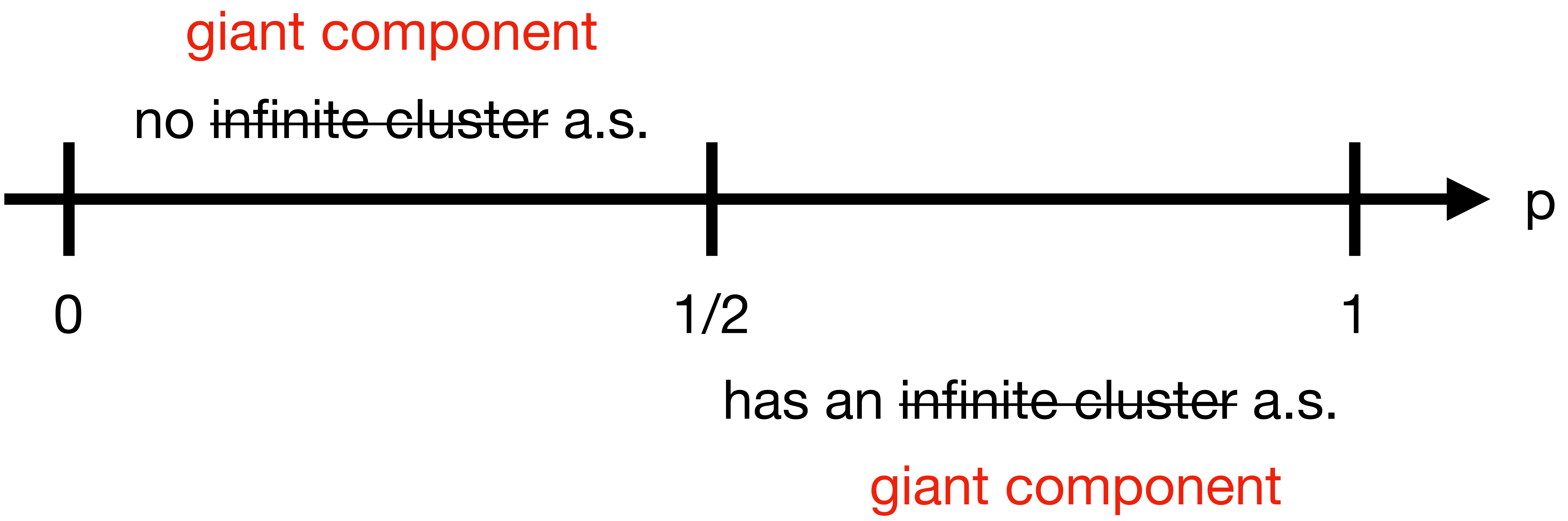


Phase Transition
[Harris 1960, Kesten 1980]



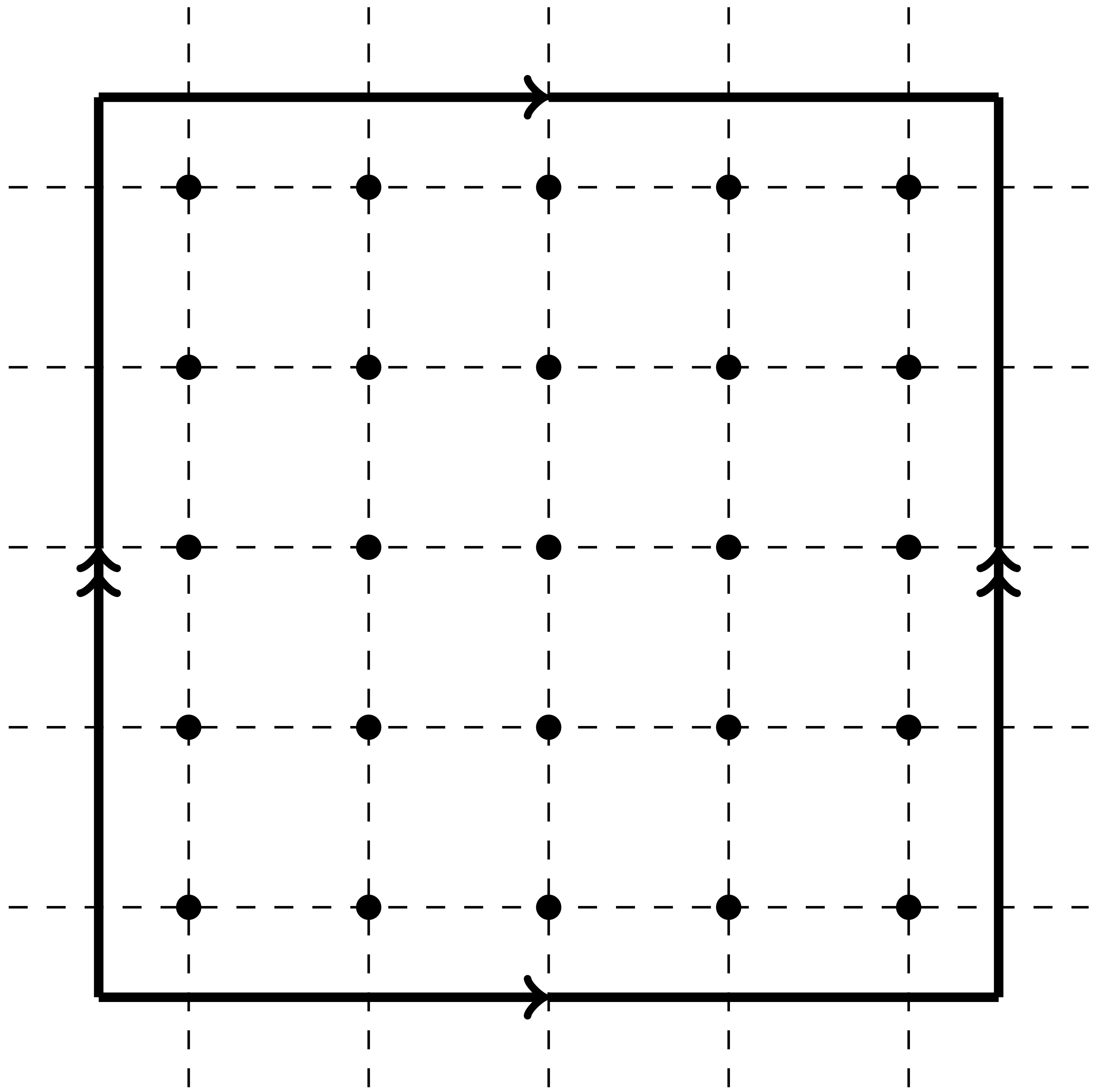
Phase Transition

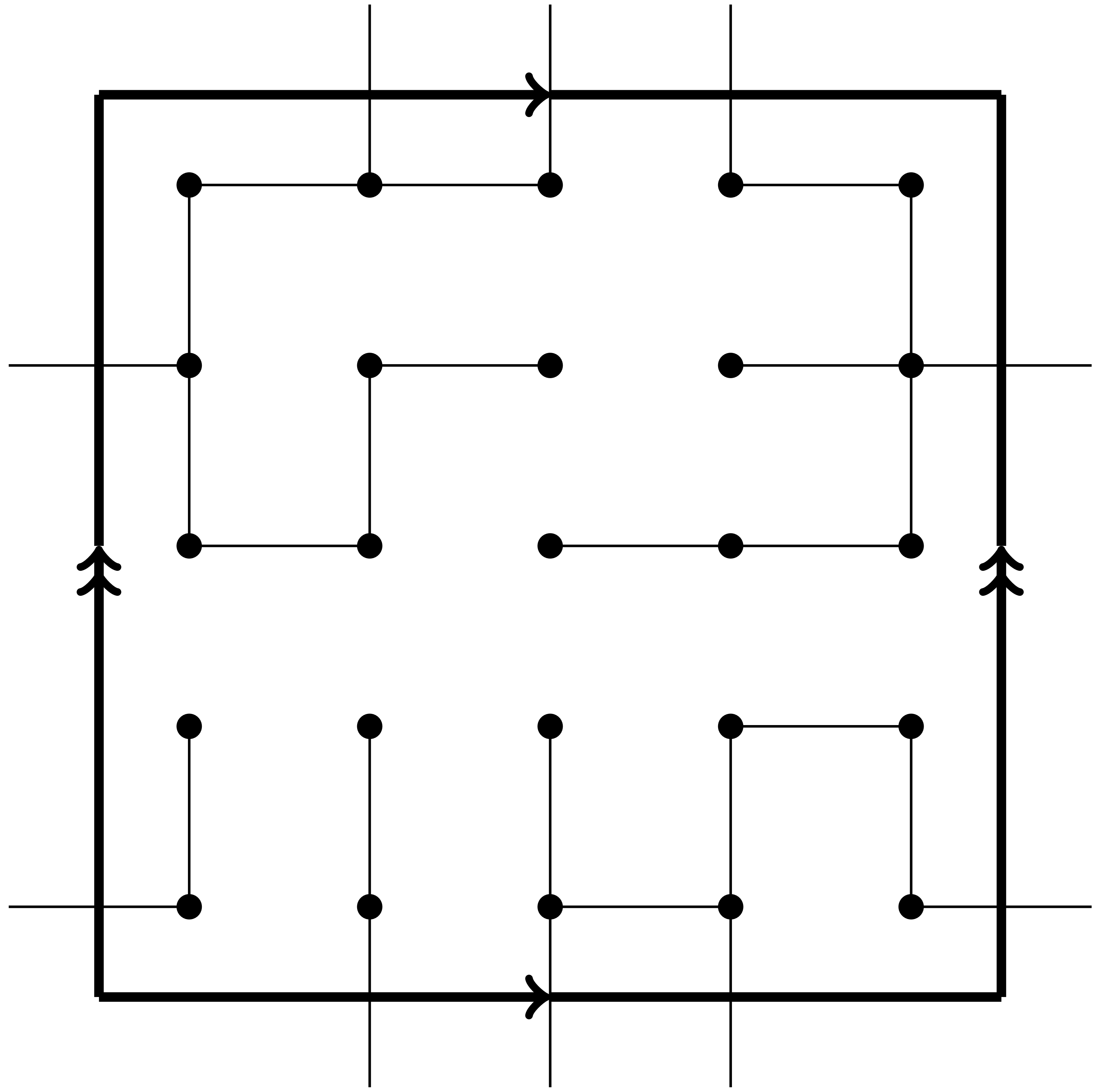
[Harris 1960, Kesten 1980]



Giant Cycles?

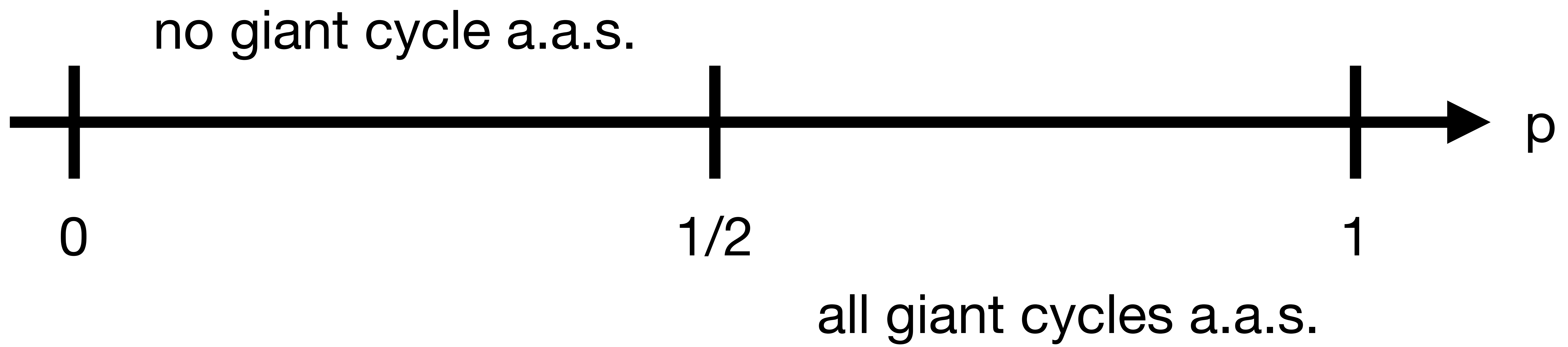
Bernoulli Bond Percolation



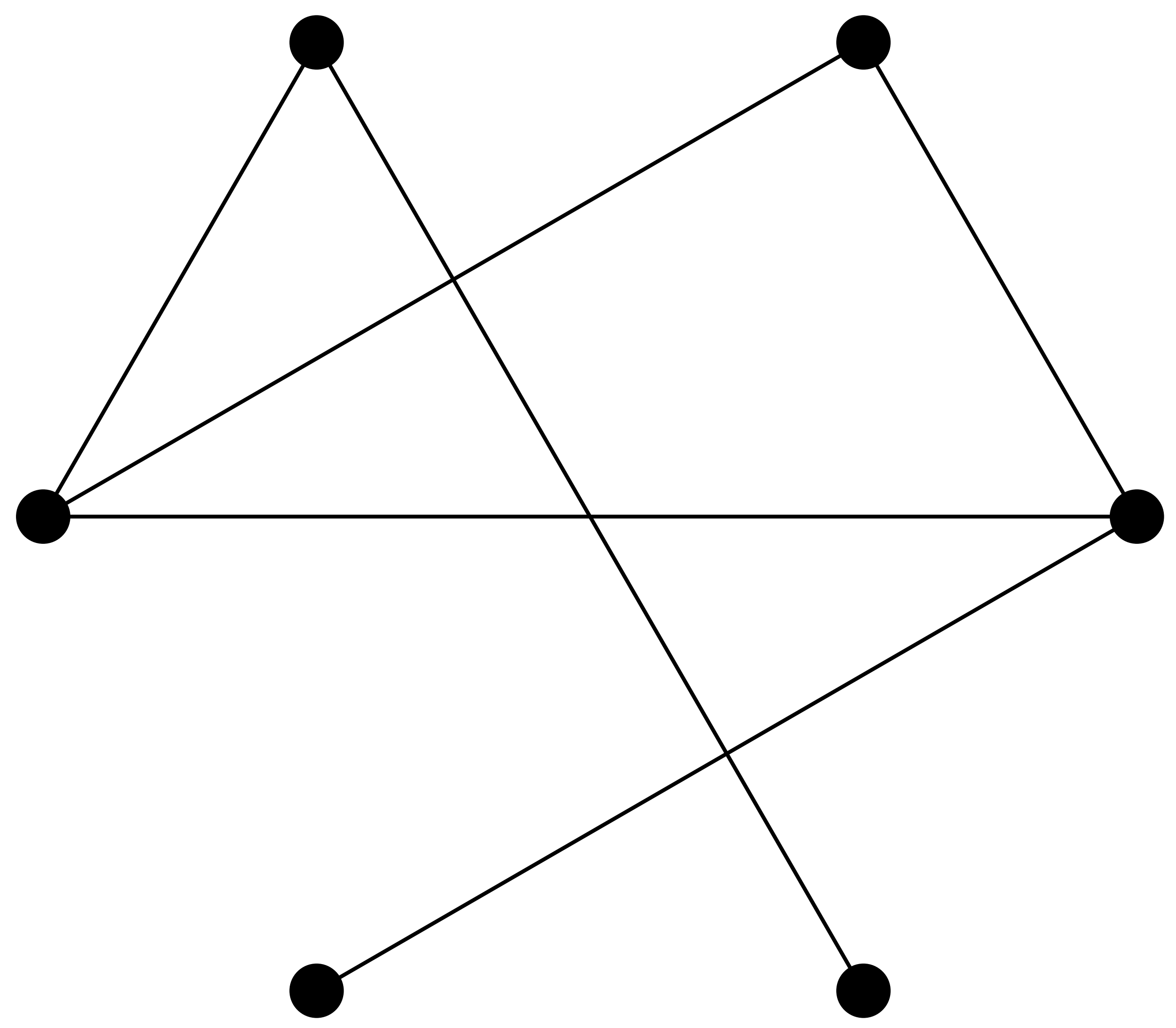


Phase Transition

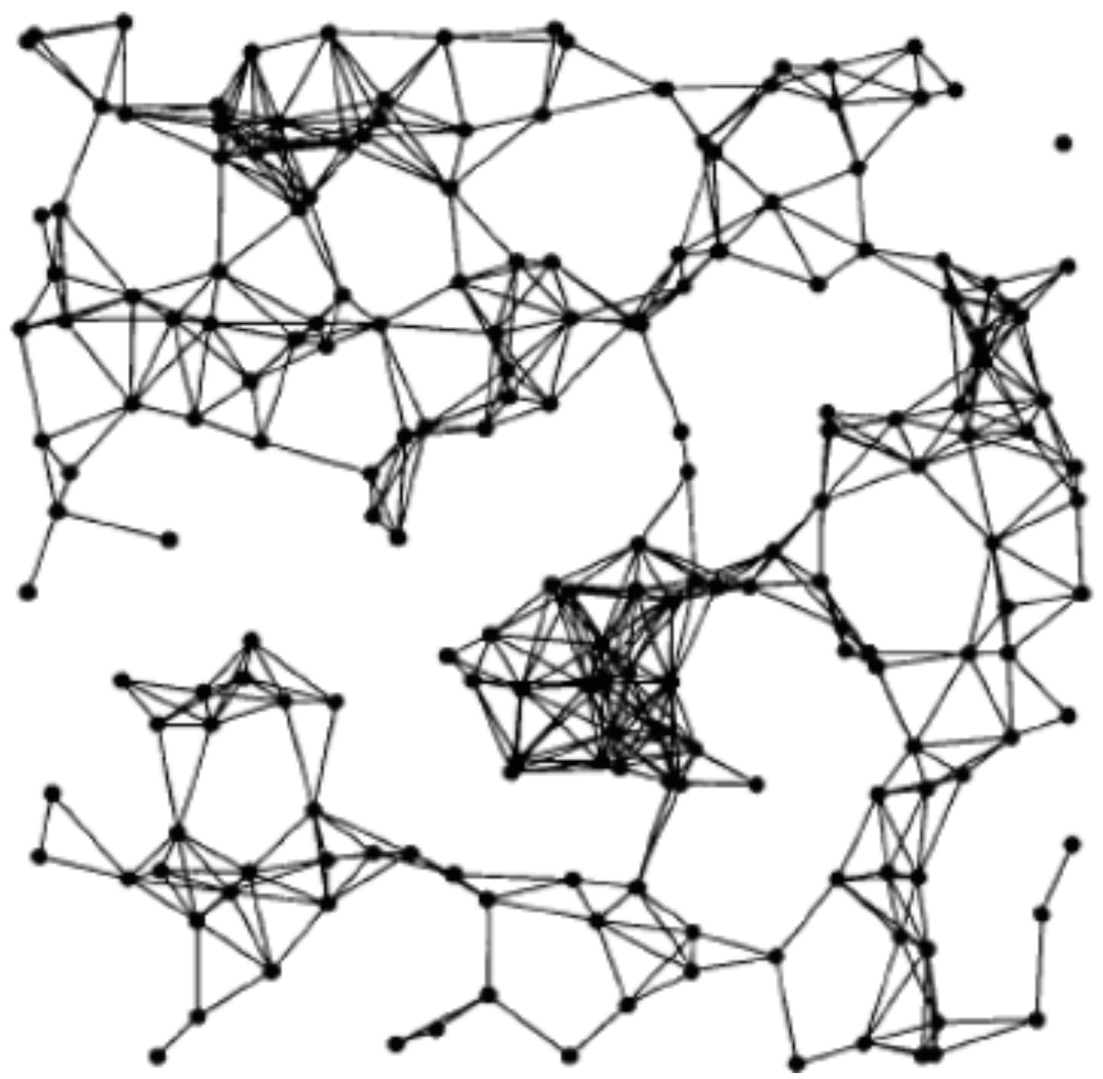
[Duncan-Kahle-Schweinhart, 2021]



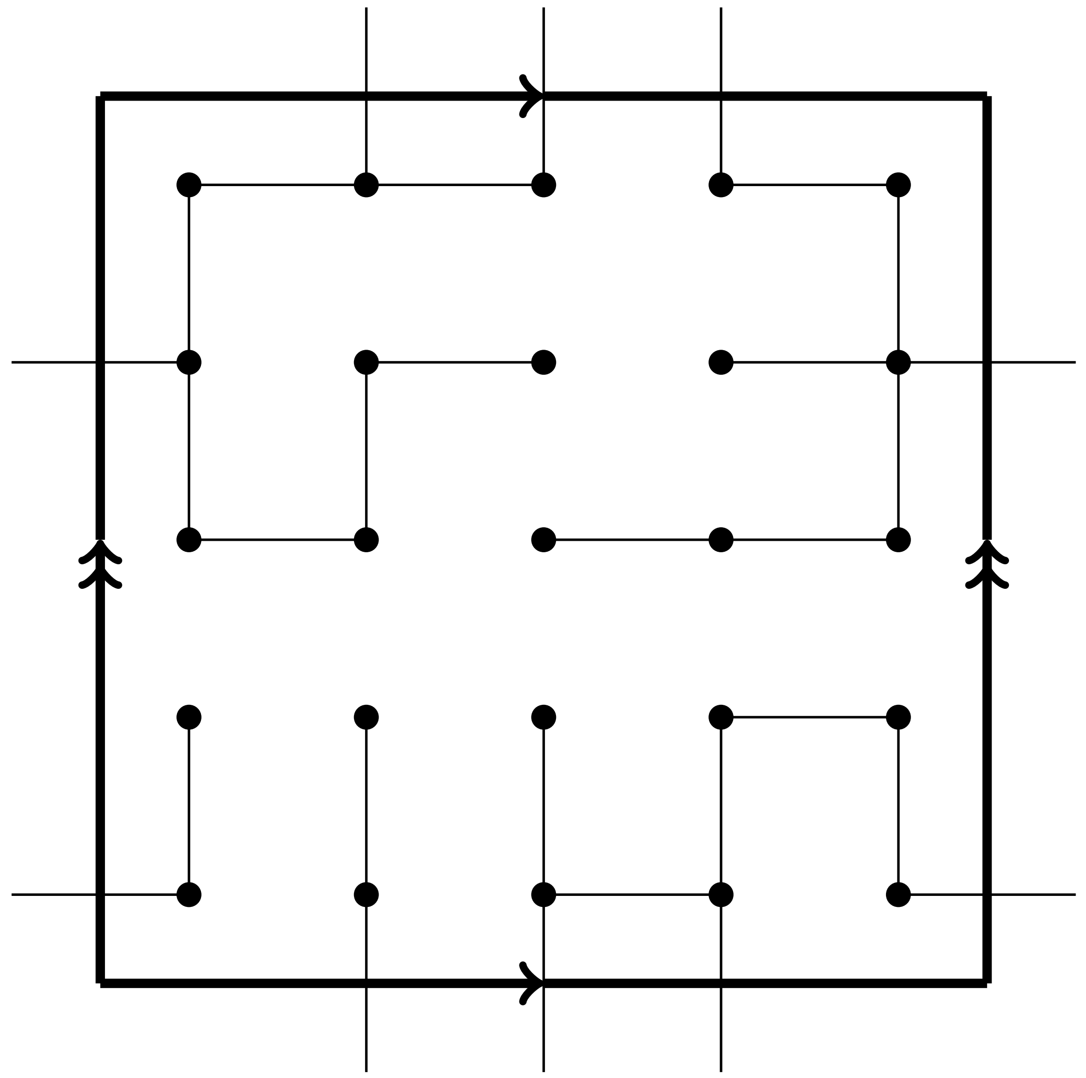
Tapas at Random Topology



Erdo-Renyi Complexes

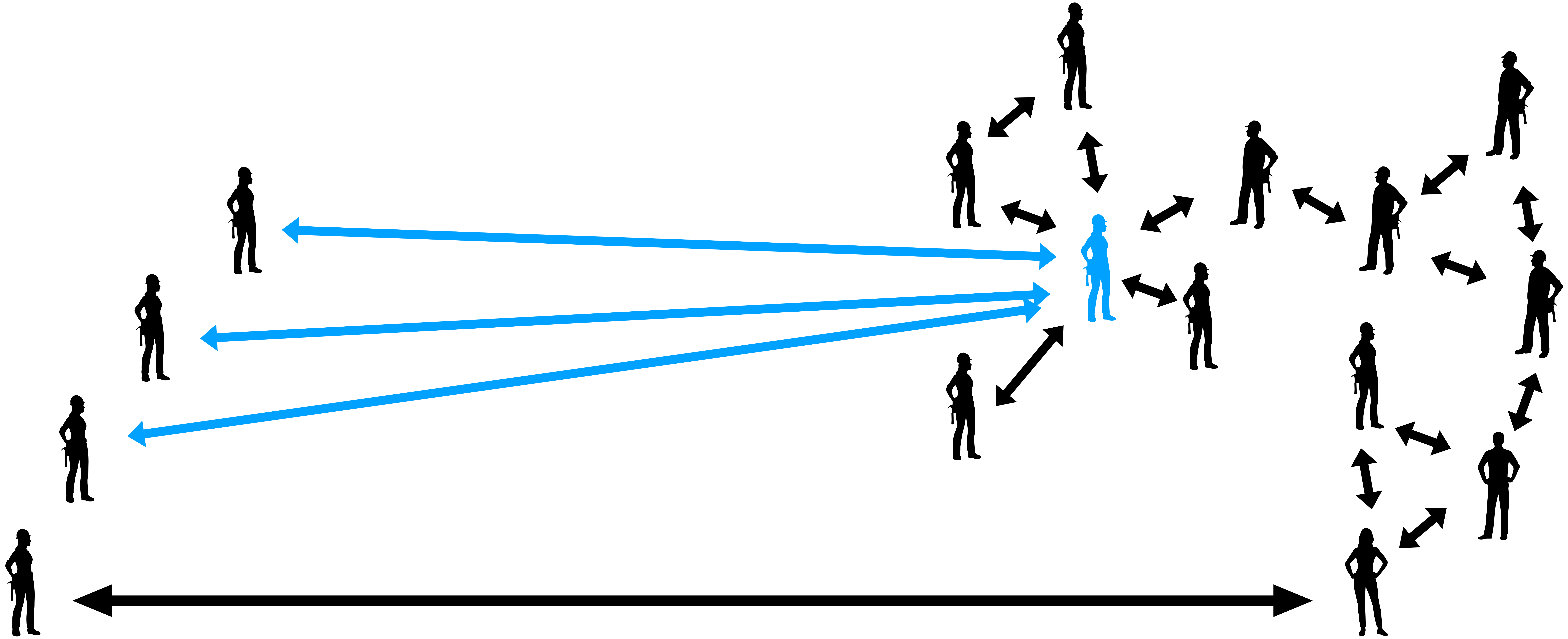


Geometric Complexes



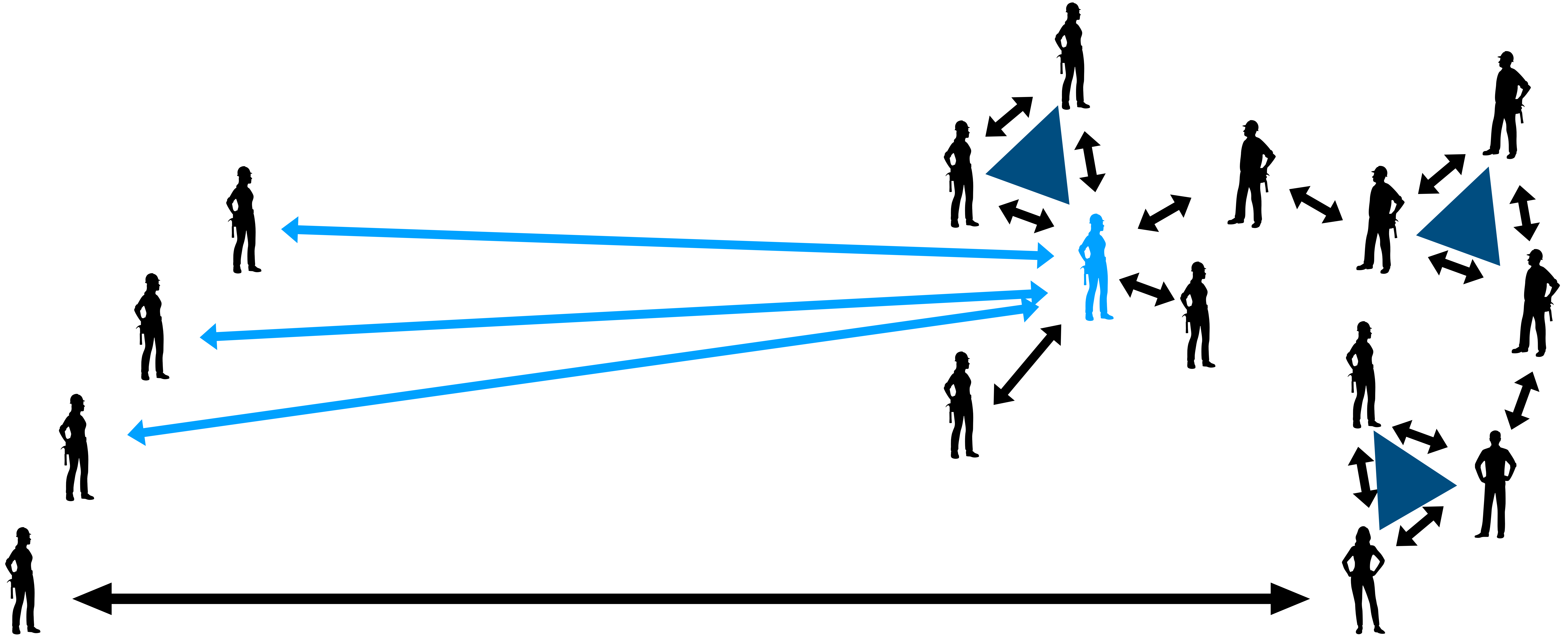
Topological Percolation

Betti numbers count repeated connections “in all dimensions”.



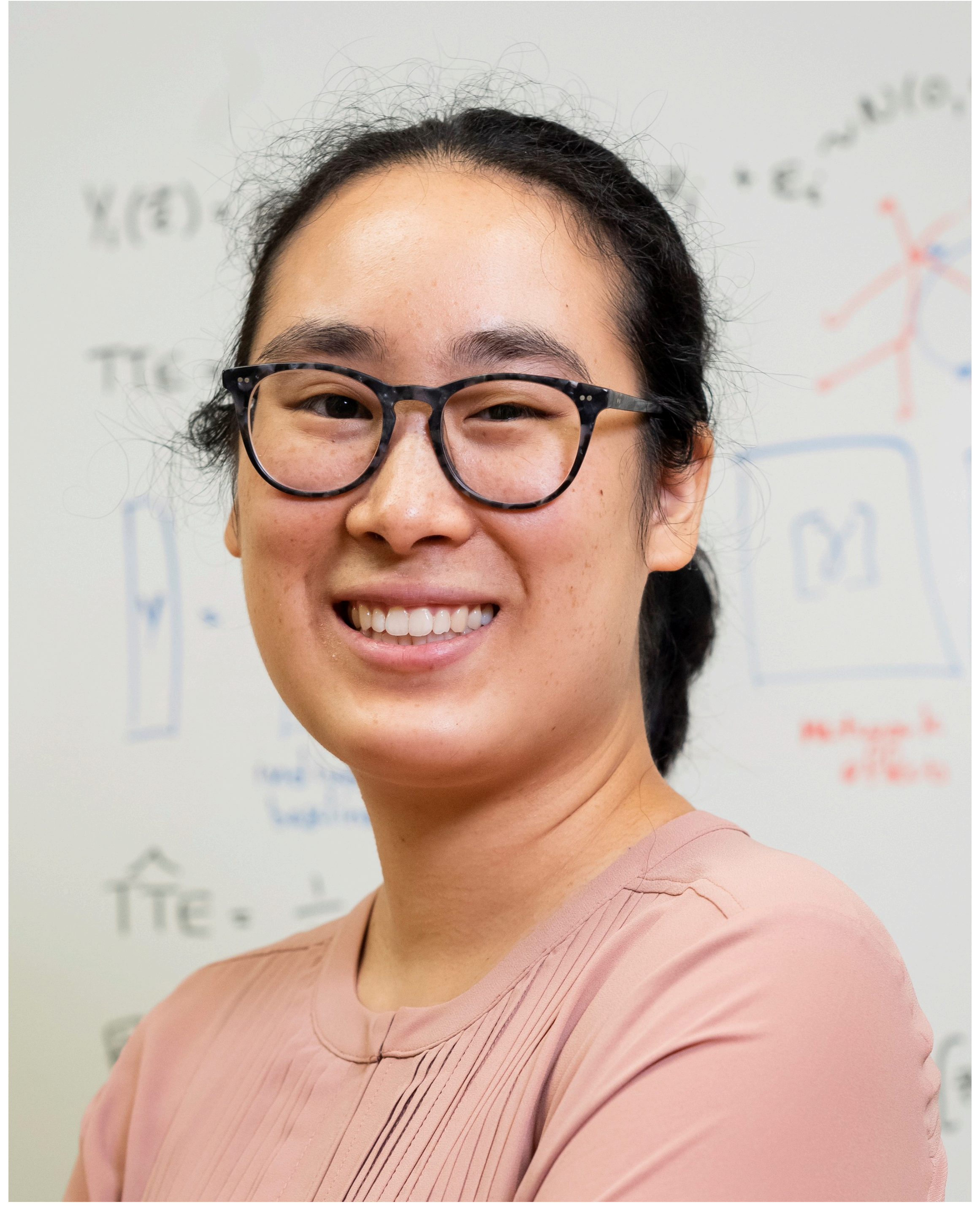
Clique Complex

aka Flag Complex



III Topology of Preferential Attachment

My Lovely Collaborators



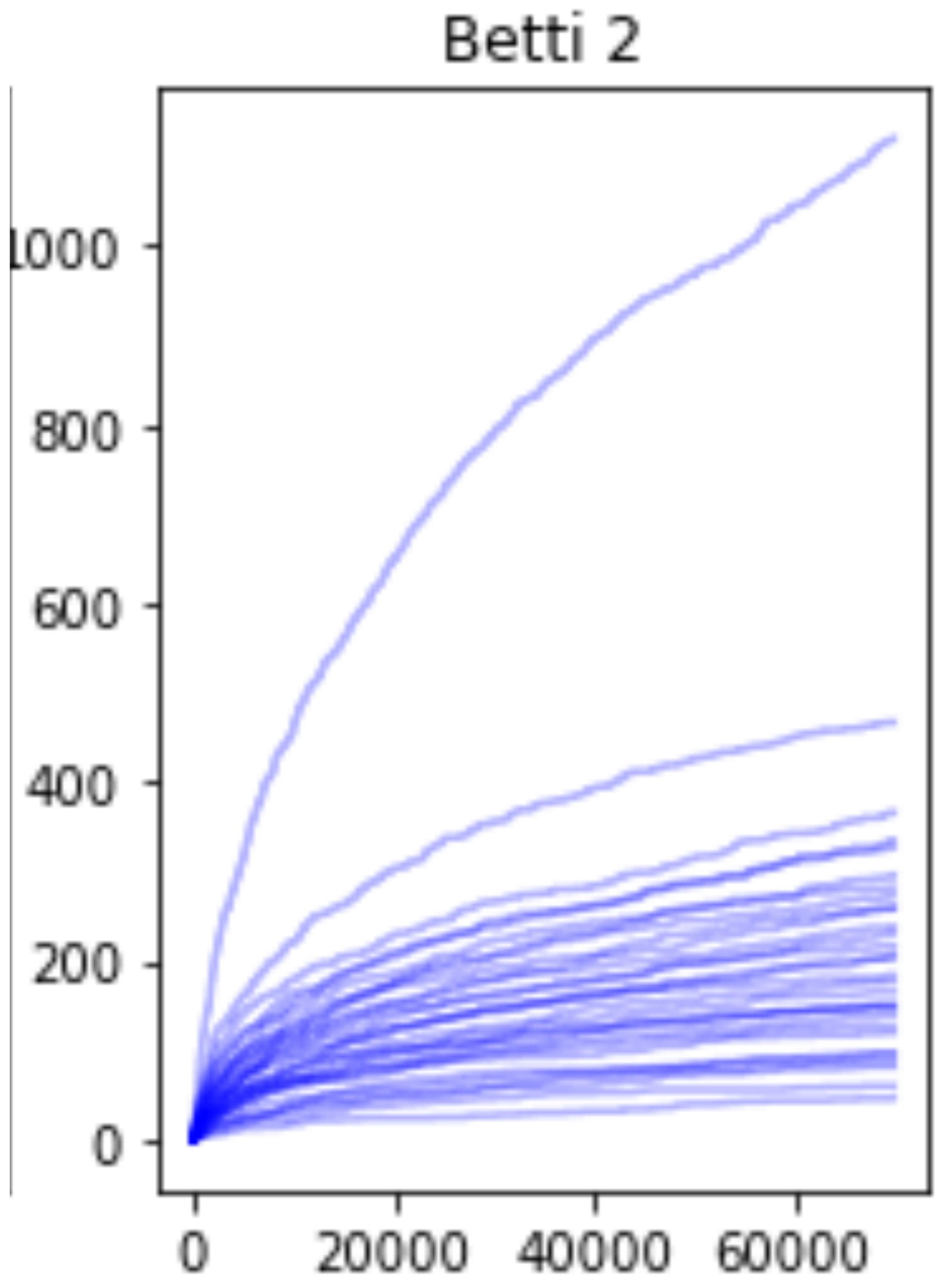
Christina Lee Yu

Gennady Samorodnitsky

Rongyi He (Caroline)

Expected Betti Number $E[\beta_q]$

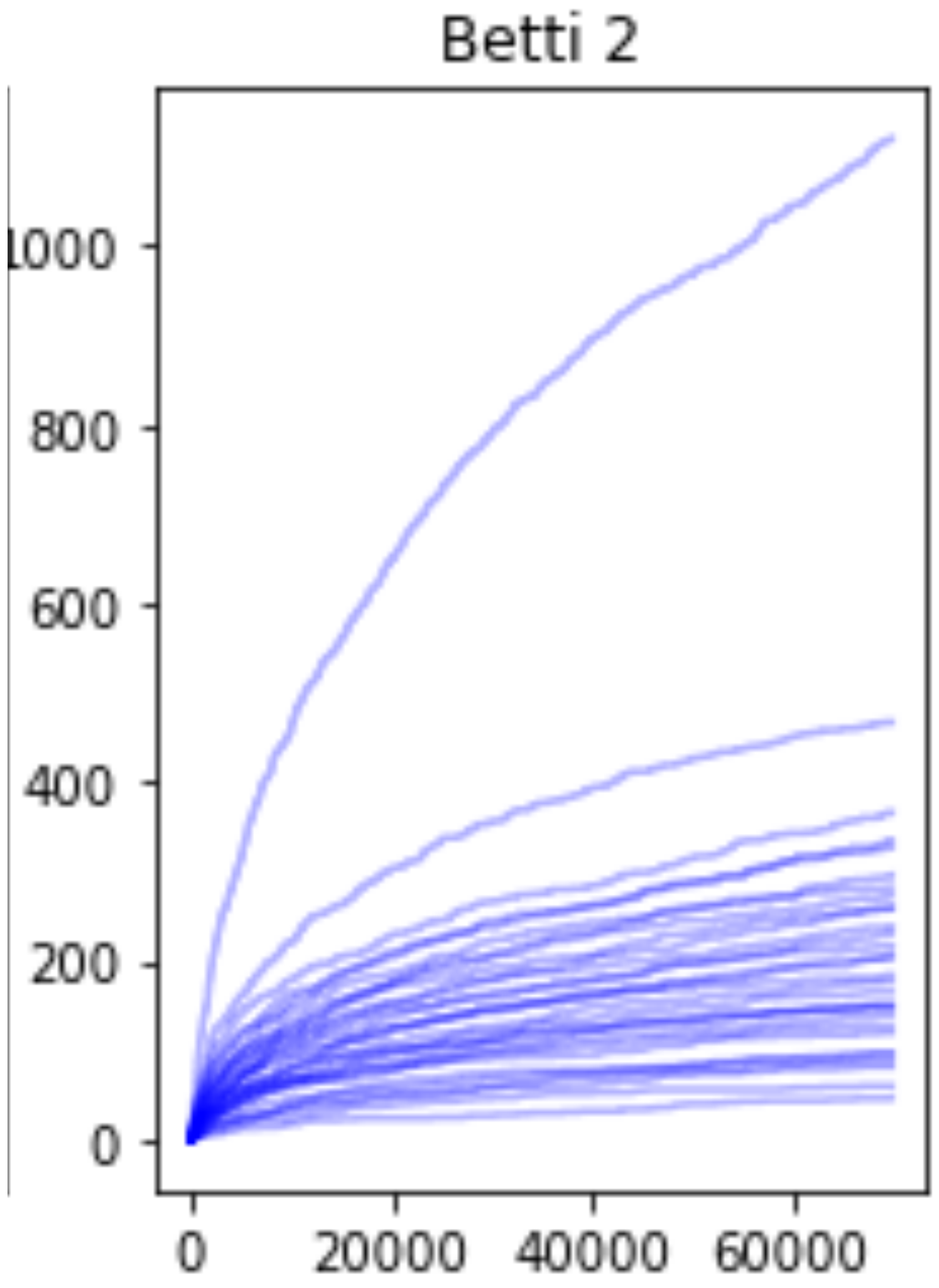
Expected Betti Number $E[\beta_q]$



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

Expected Betti Number $E[\beta_q]$

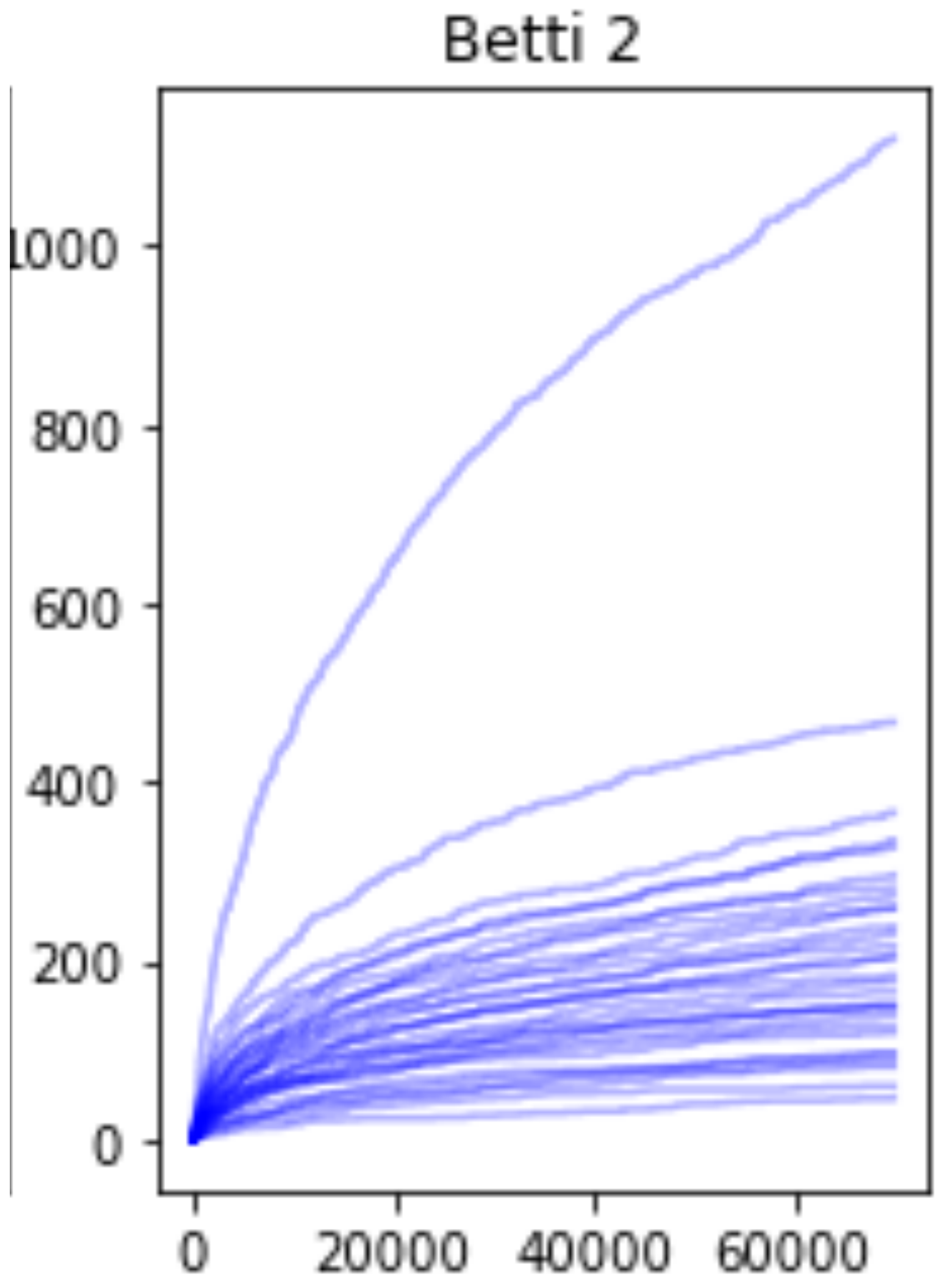
- increasing trend



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

Expected Betti Number $E[\beta_q]$

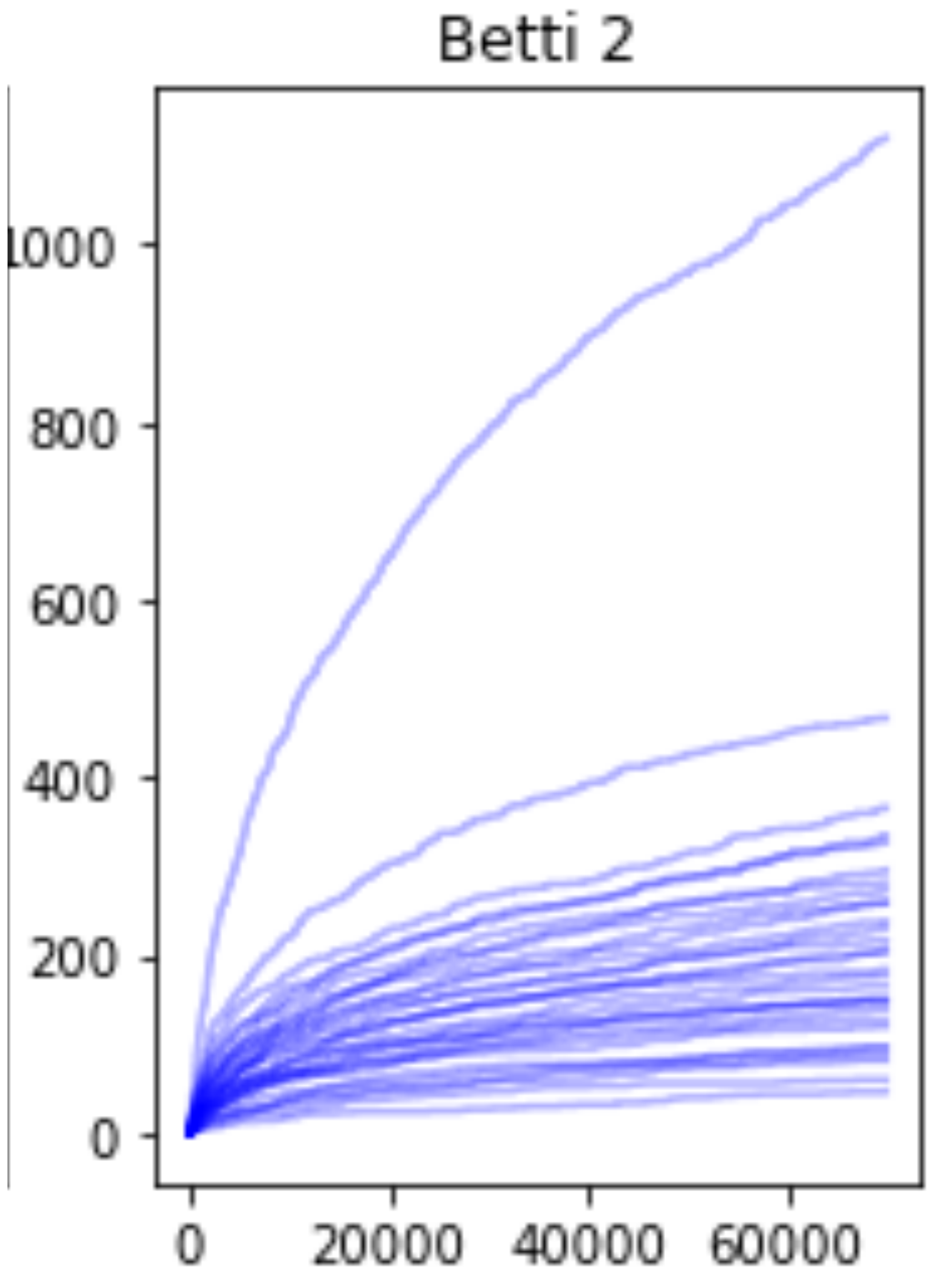
- increasing trend
- concave growth



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

Expected Betti Number $E[\beta_q]$

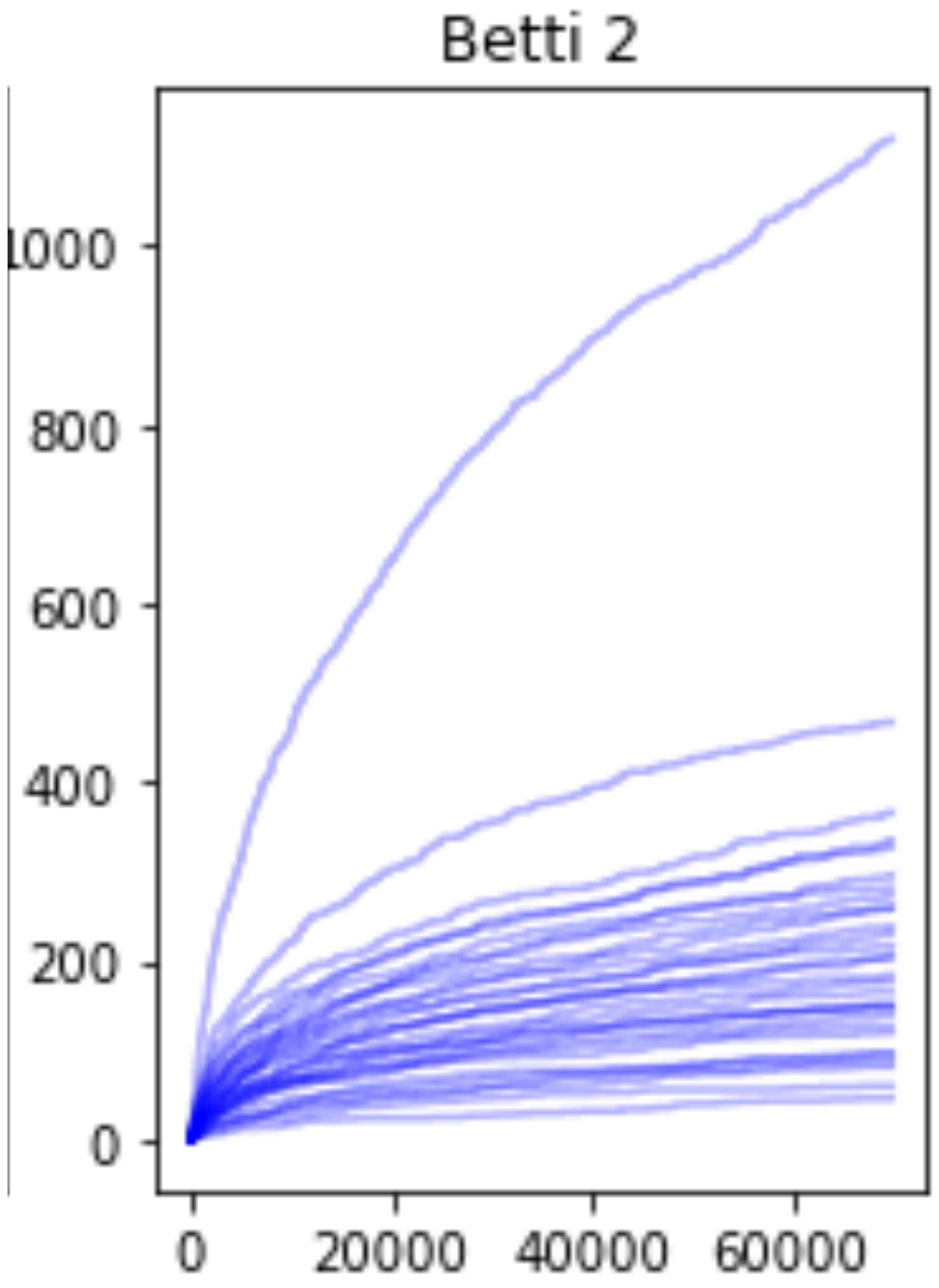
- increasing trend
- concave growth
- outlier



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

Expected Betti Number $E[\beta_q]$

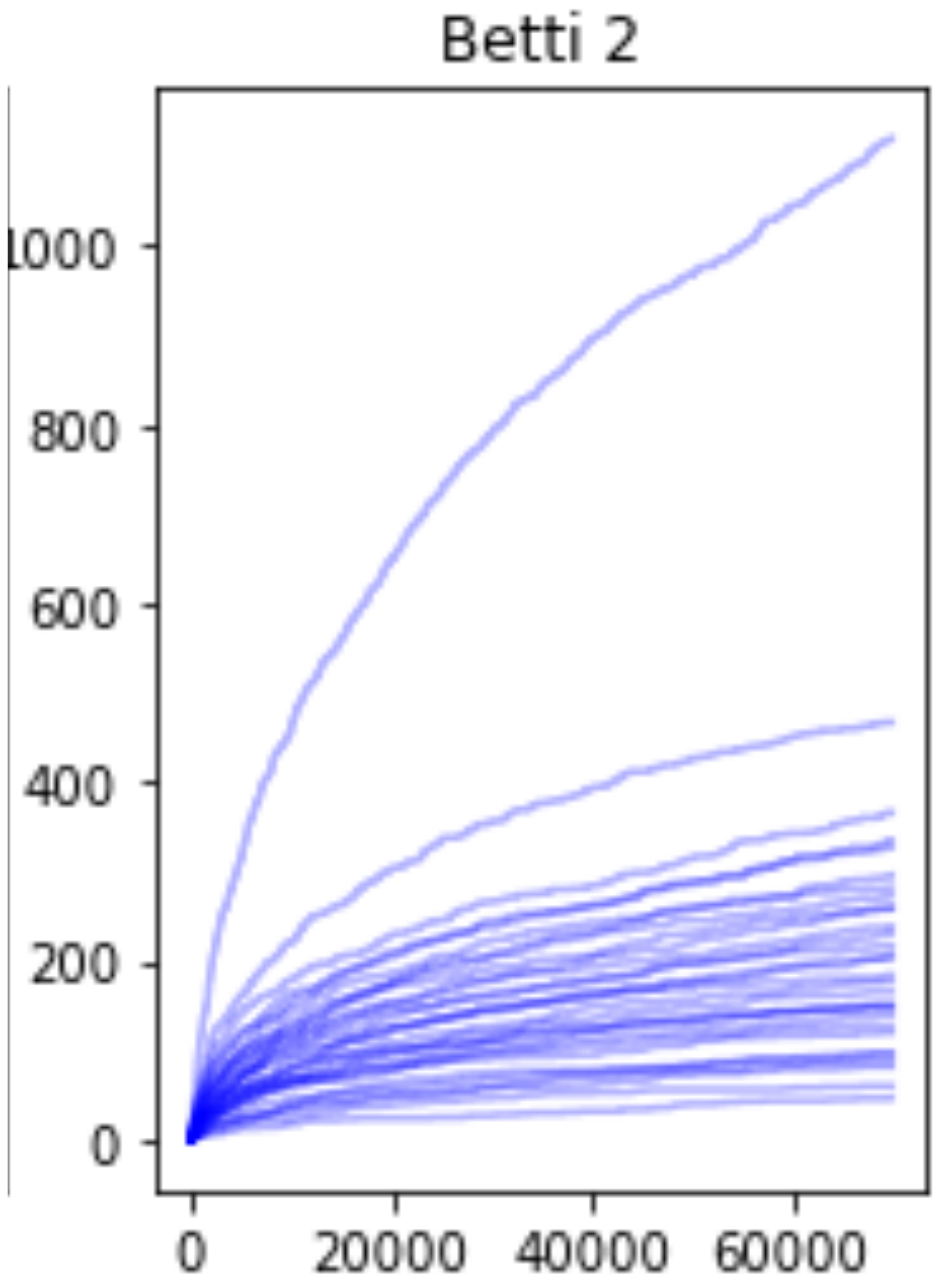
- $c(\text{num of nodes}^{1-4x}) \leq E[\beta_2] \leq C(\text{num of nodes}^{1-4x})$
  under mild assumptions
  - $x \in (0,1/2)$ depends on pref. attachment strength



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

Expected Betti Number $E[\beta_q]$

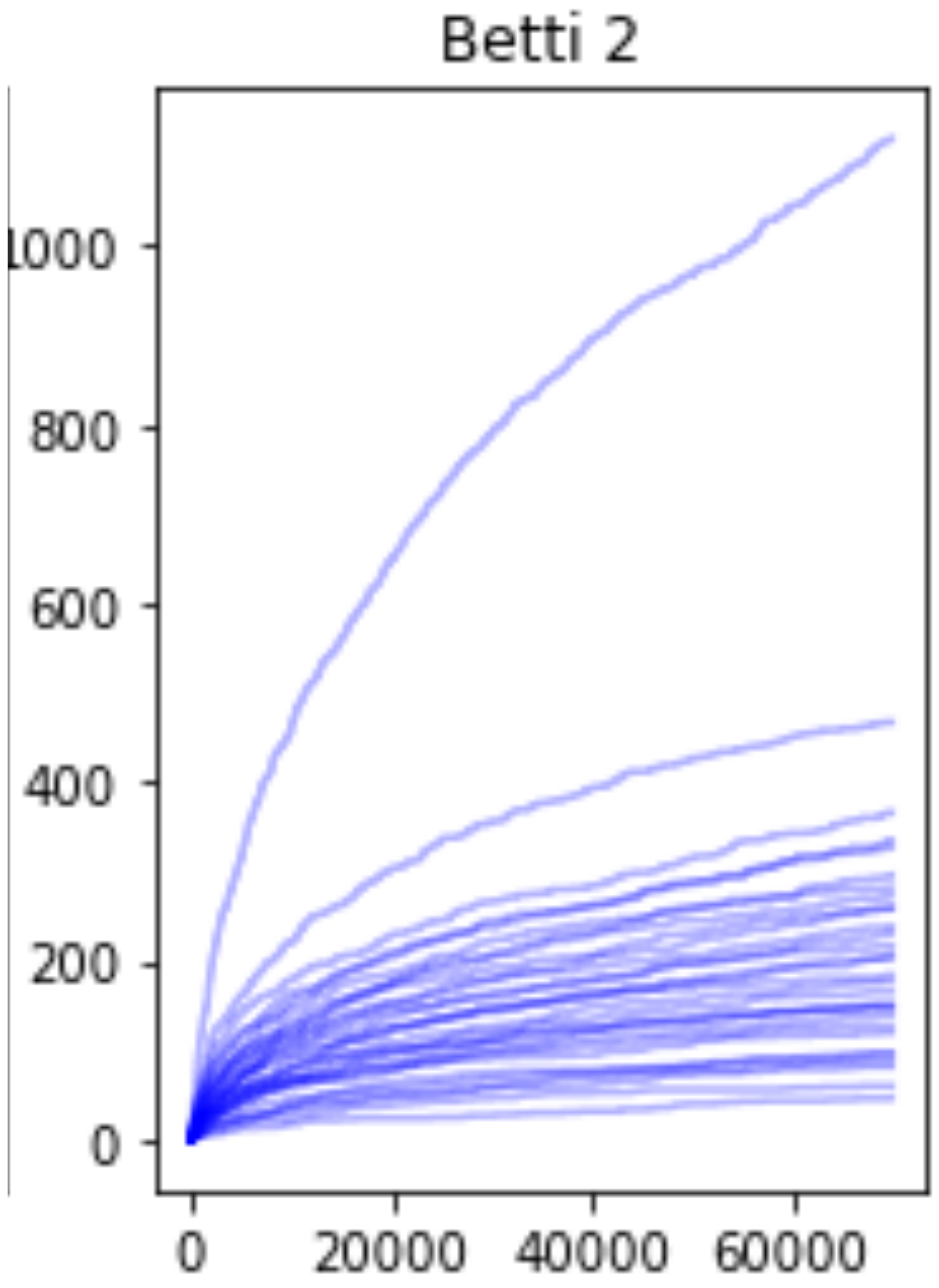
- c(num of nodes^{1-4x}) $\leq E[\beta_2] \leq C$(num of nodes^{1-4x})
  under mild assumptions
  - $x \in (0,1/2)$ depends on pref. attachment strength
  - If $1 - 4x < 0$, then $E[\beta_2] \leq C$.



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 pref. attachment strength
    - If $1 - 4x < 0$, then $E[\beta_2] \leq C$.
- $c(\text{num of nodes}^{1-2qx}) \leq E[\beta_q] \leq C(\text{num of nodes}^{1-2qx})$ for $q \geq 2$ if $1 - 2qx > 0$



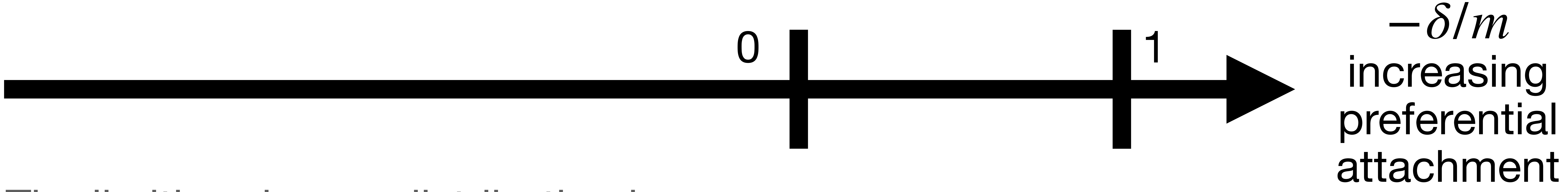
Different curves, different random seeds.
All curves have the same model parameters.

Phase transition

Recall

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

m = number of edges per new node



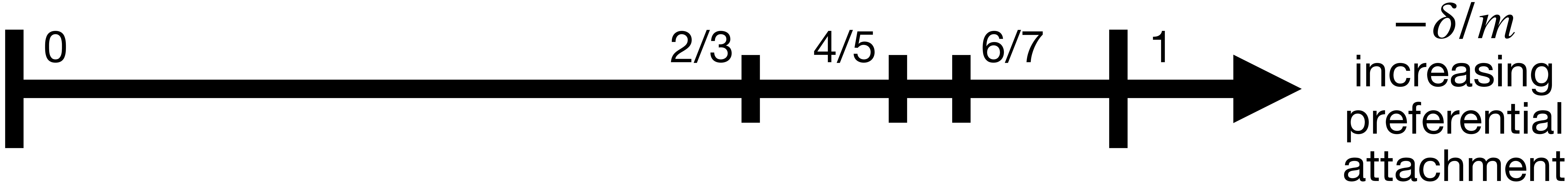
The limiting degree distribution has …

Phase transition

Recall

P(attaching to v) ∝ degree + δ

m = number of edges per new node

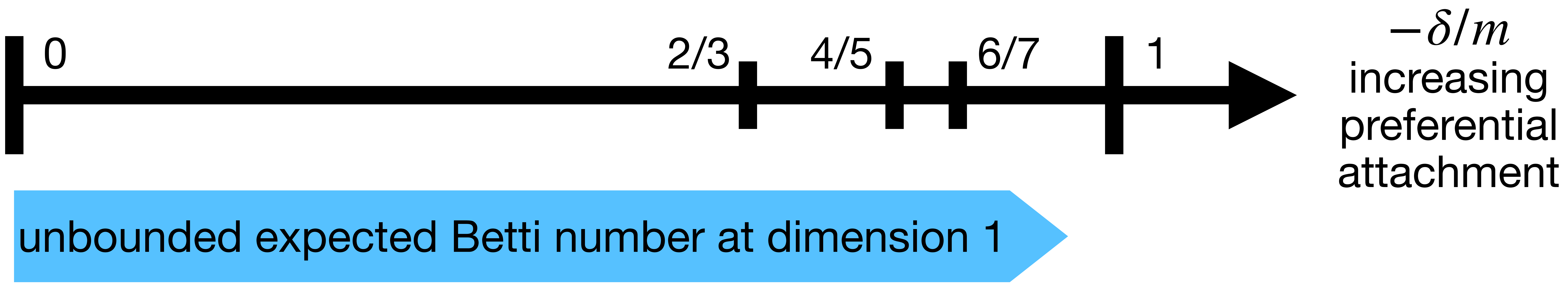


Phase transition

Recall

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

m = number of edges per new node

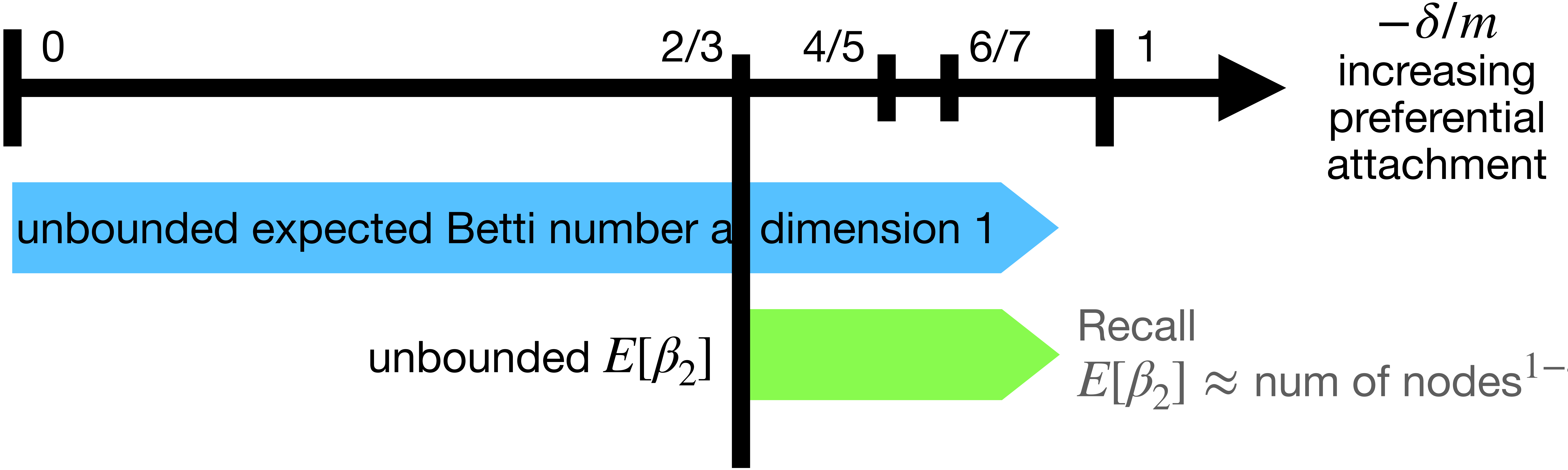


Phase transition

Recall

P(attaching to v) ∝ degree + δ

m = number of edges per new node



Recall

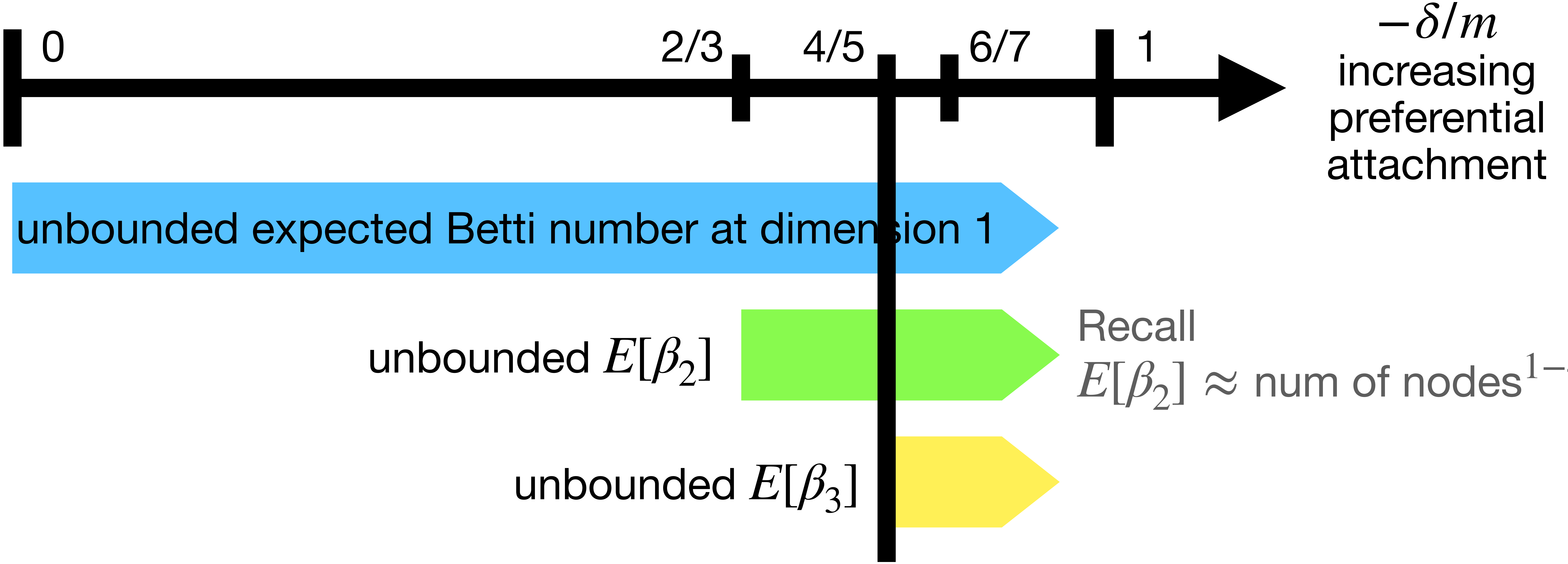
$E[\beta_2] \approx$ num of nodes^{1-4x}

Phase transition

Recall

P(attaching to v) ∝ degree + δ

m = number of edges per new node

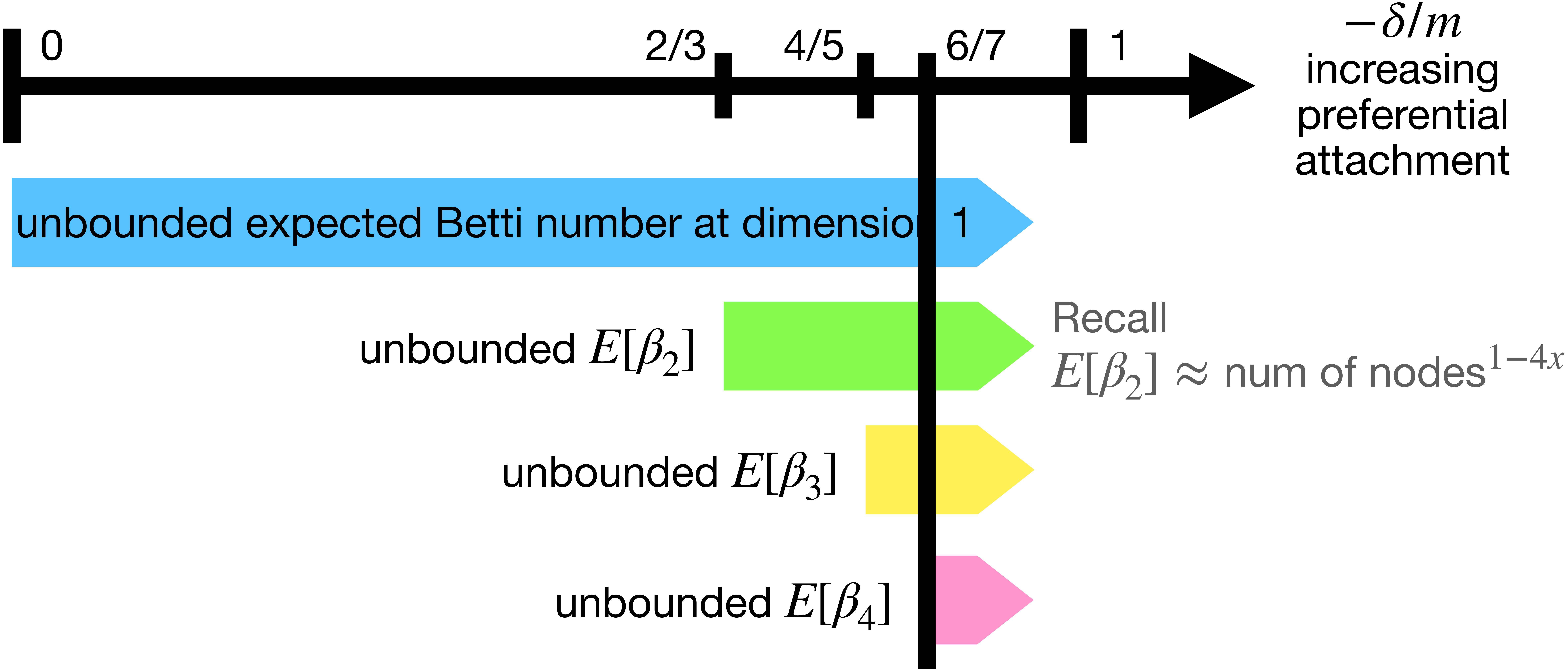


Recall

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

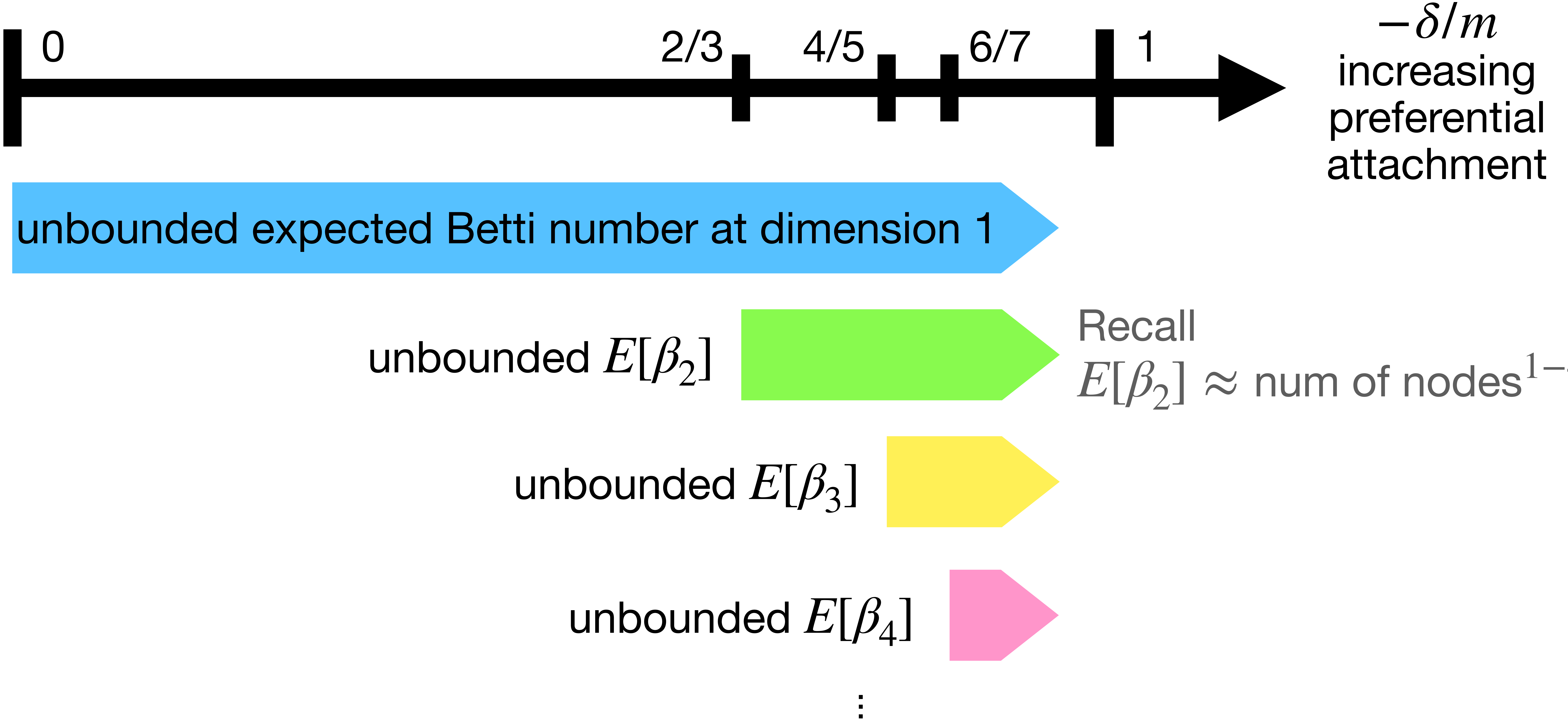
Phase transition

Recall
P(attaching to v) ∝ 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

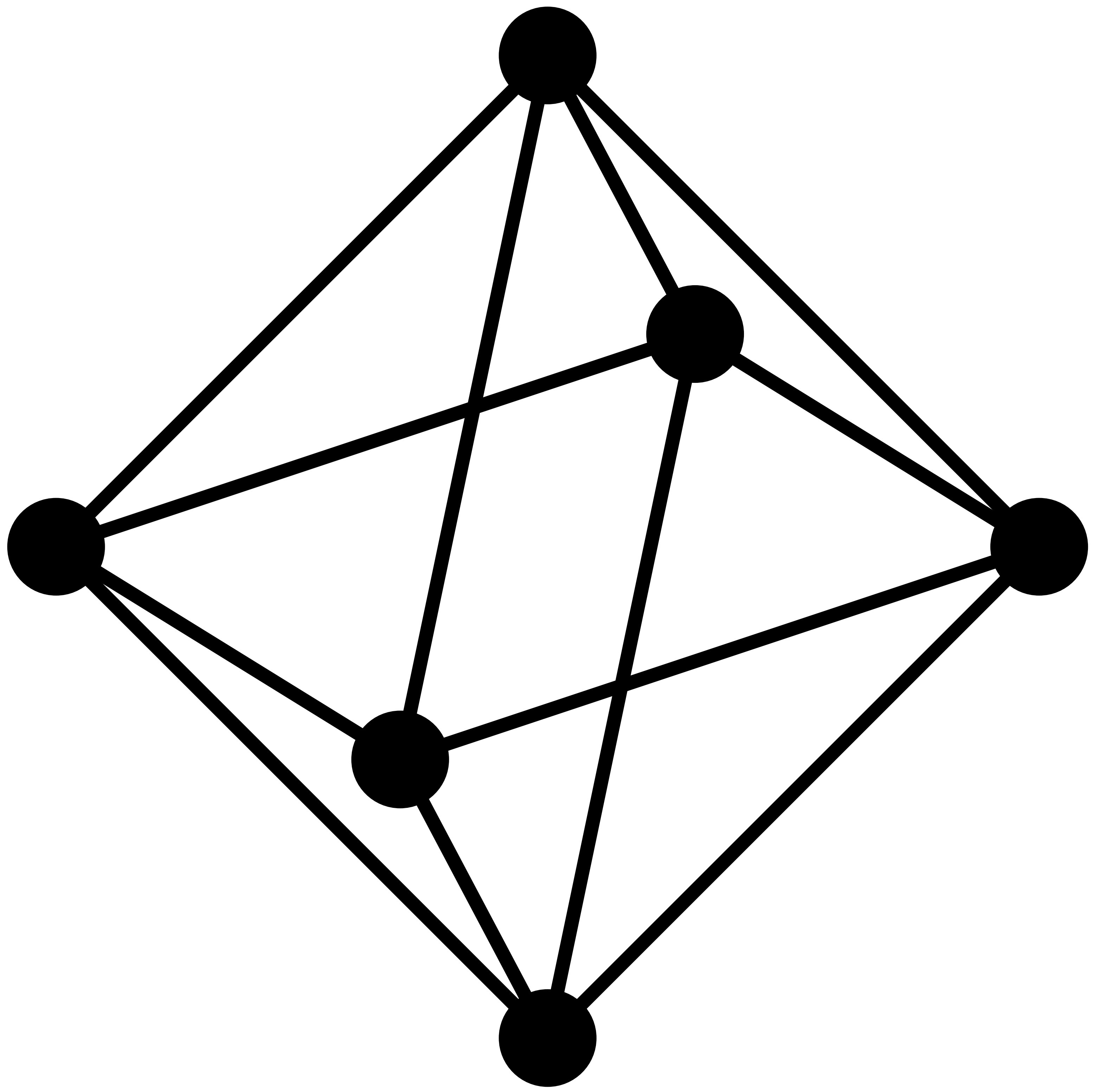


Recall
$E[\beta_2] \approx$ num of nodes^{1-4x}

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

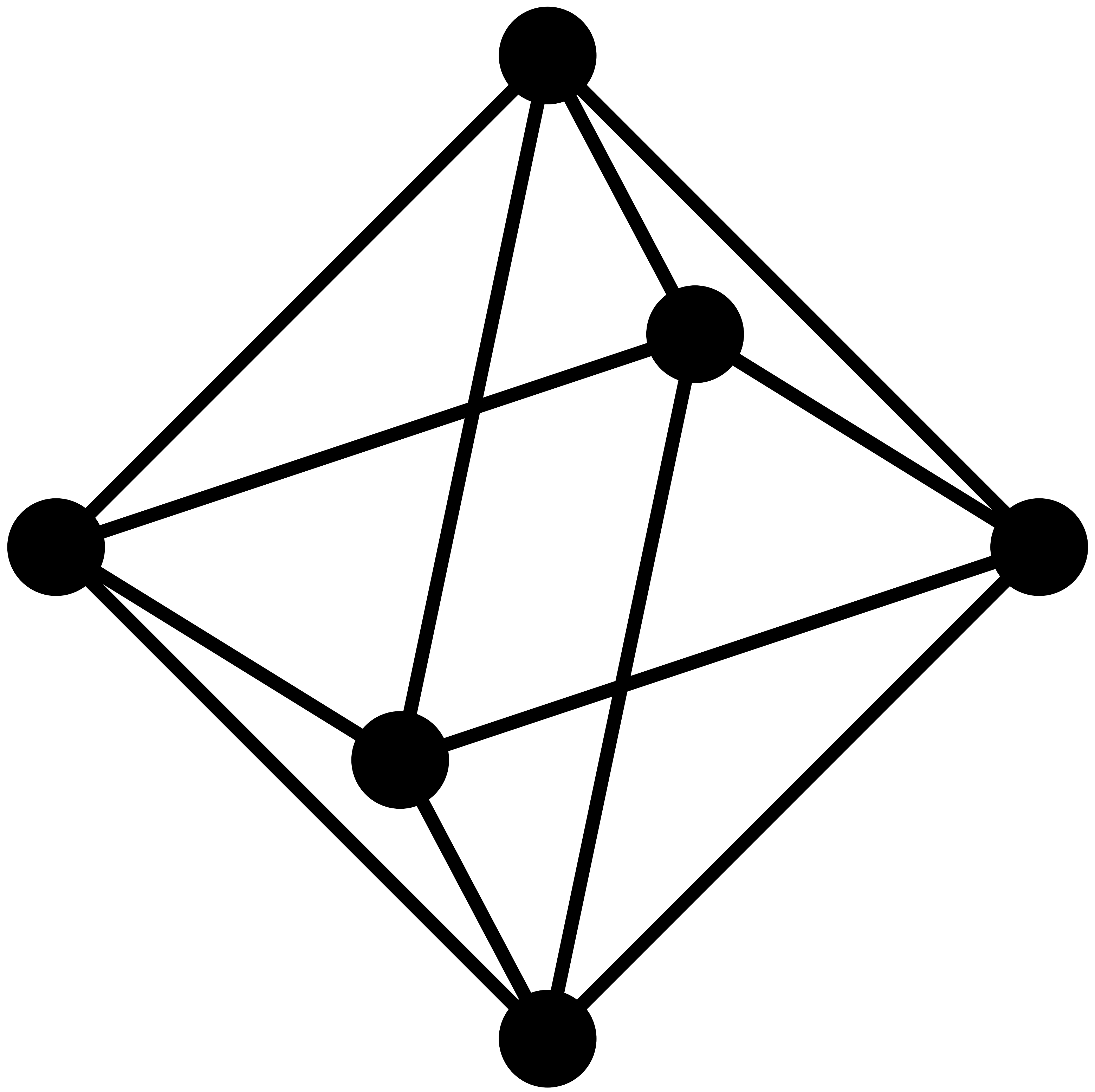
Proof?

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

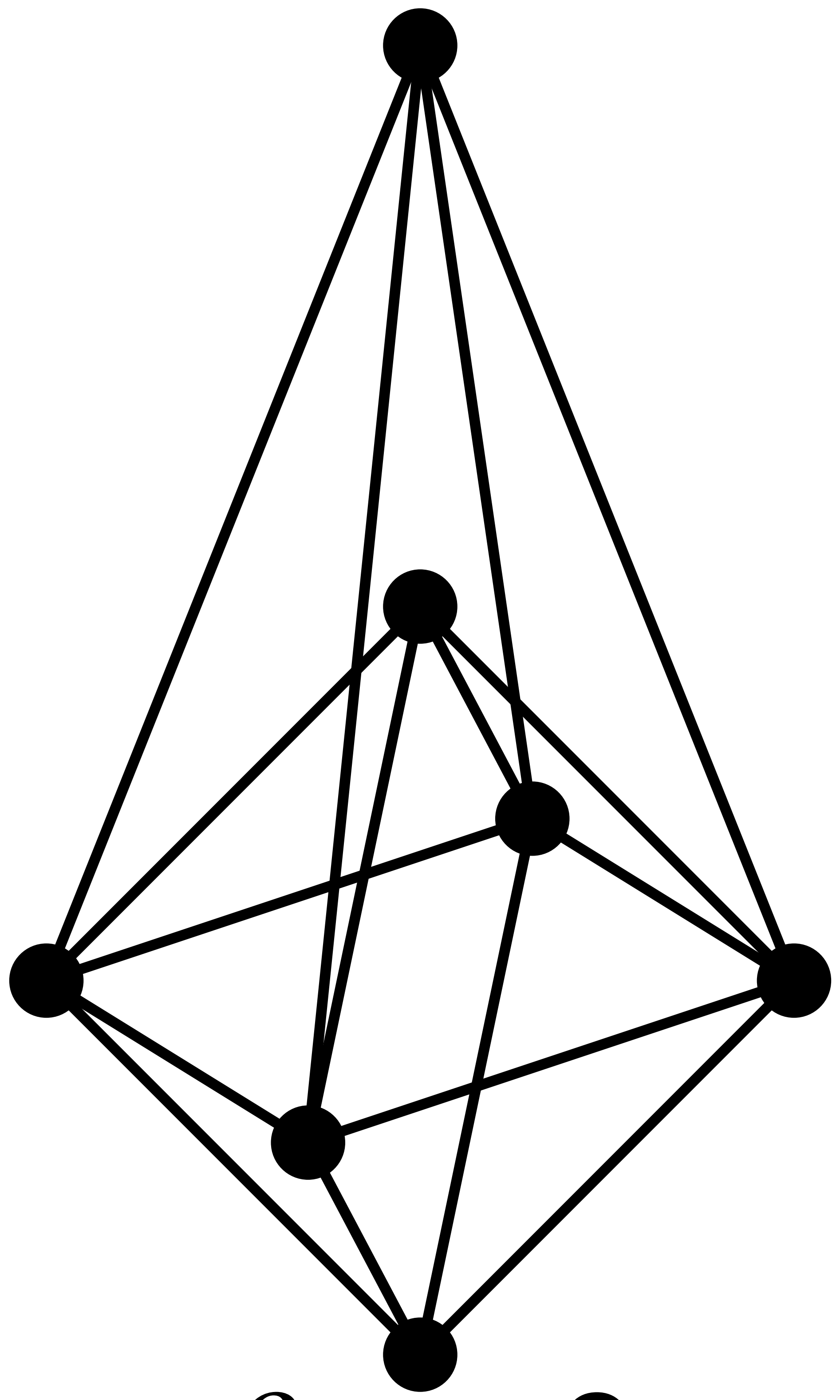


$\beta_2 = 1$

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

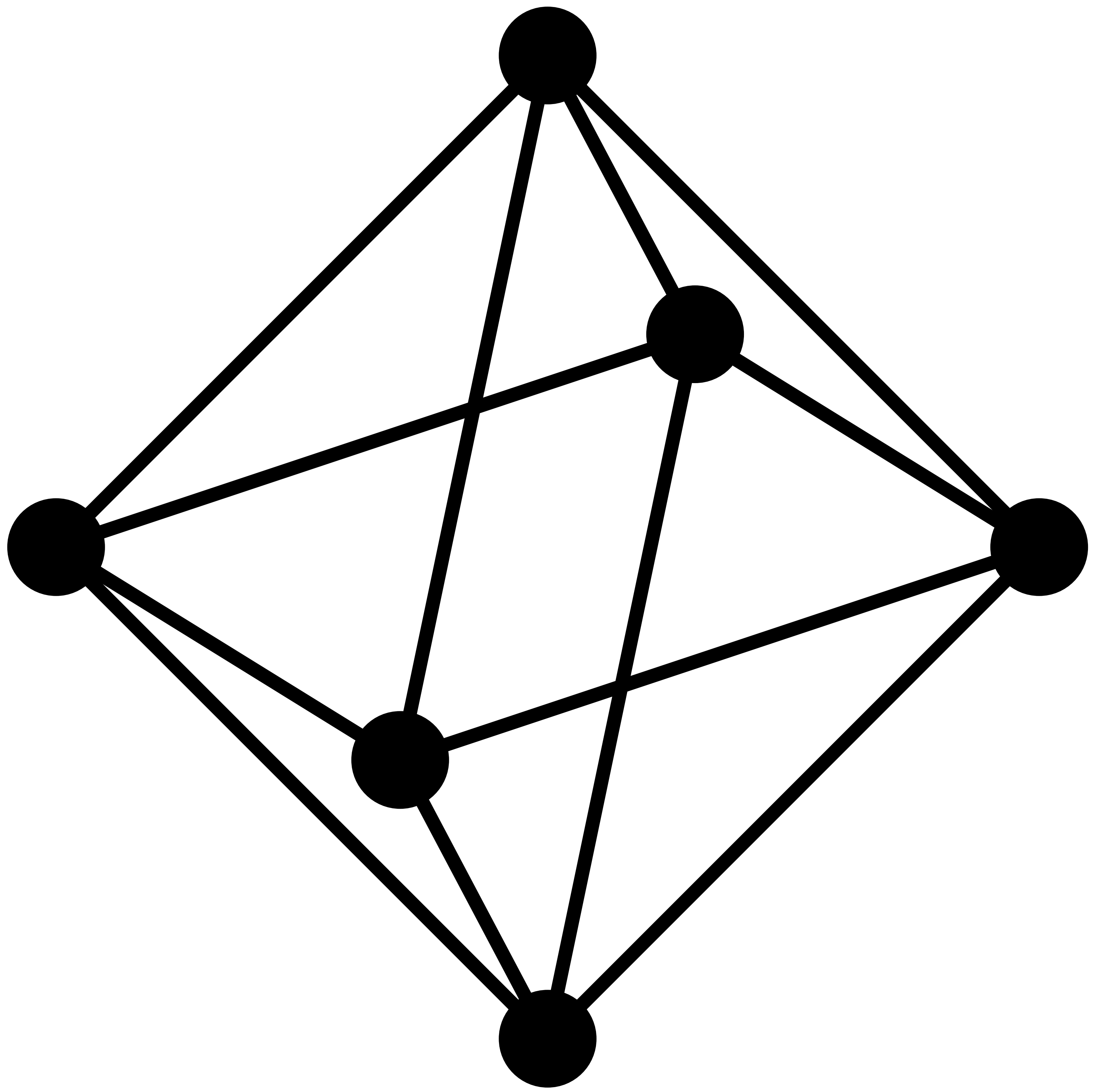


$\beta_2 = 1$

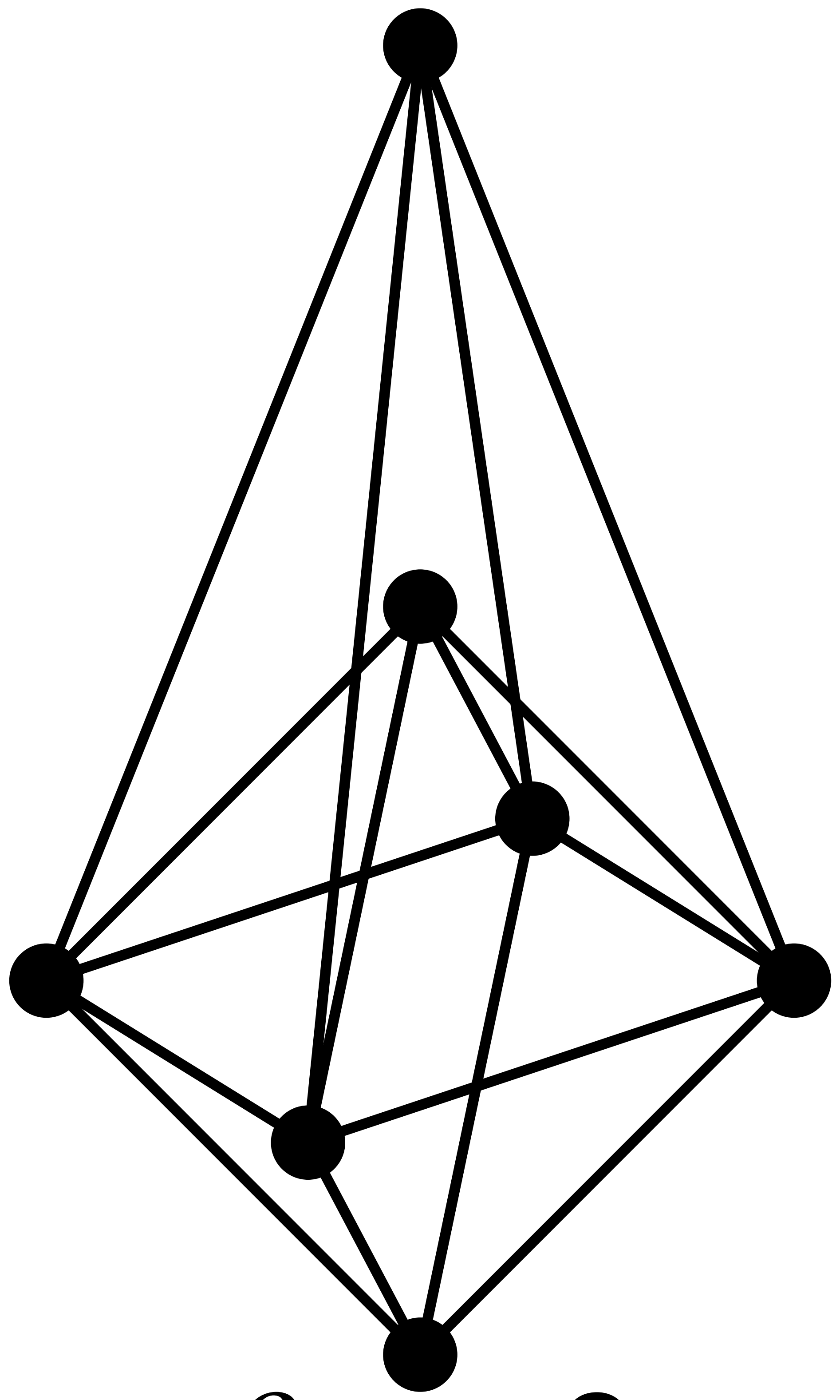


$\beta_2 = 2$

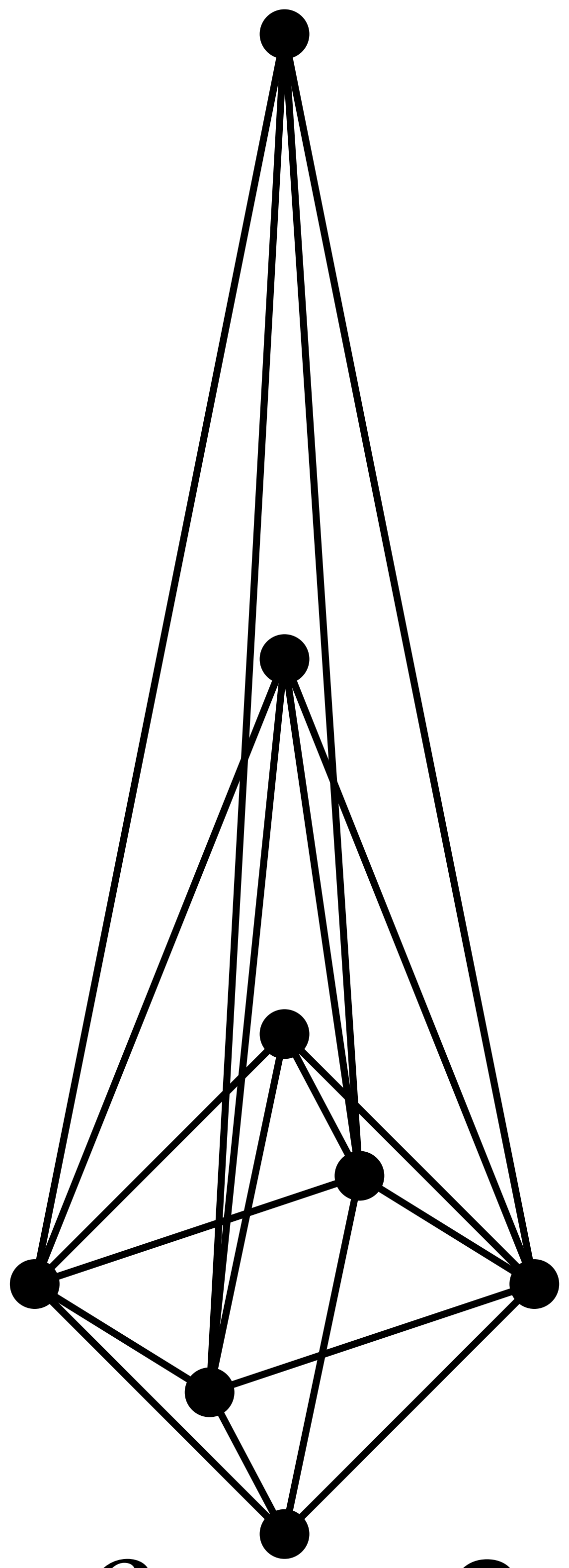
Proof of $E[\beta_2] \approx$ num of nodes^{1-4x}



$\beta_2 = 1$



$\beta_2 = 2$



$\beta_2 = 3$

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

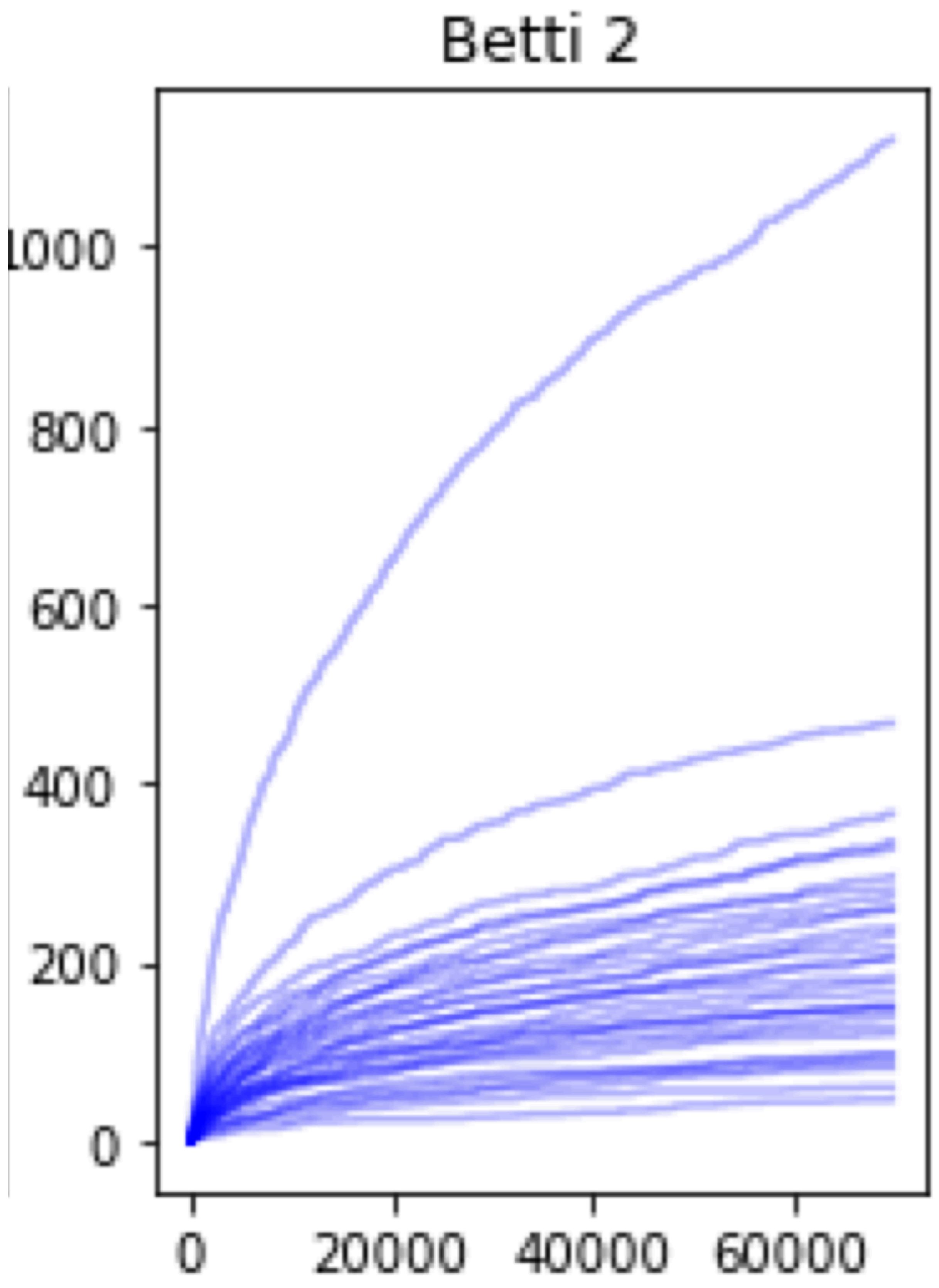
Subtleties

- 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}

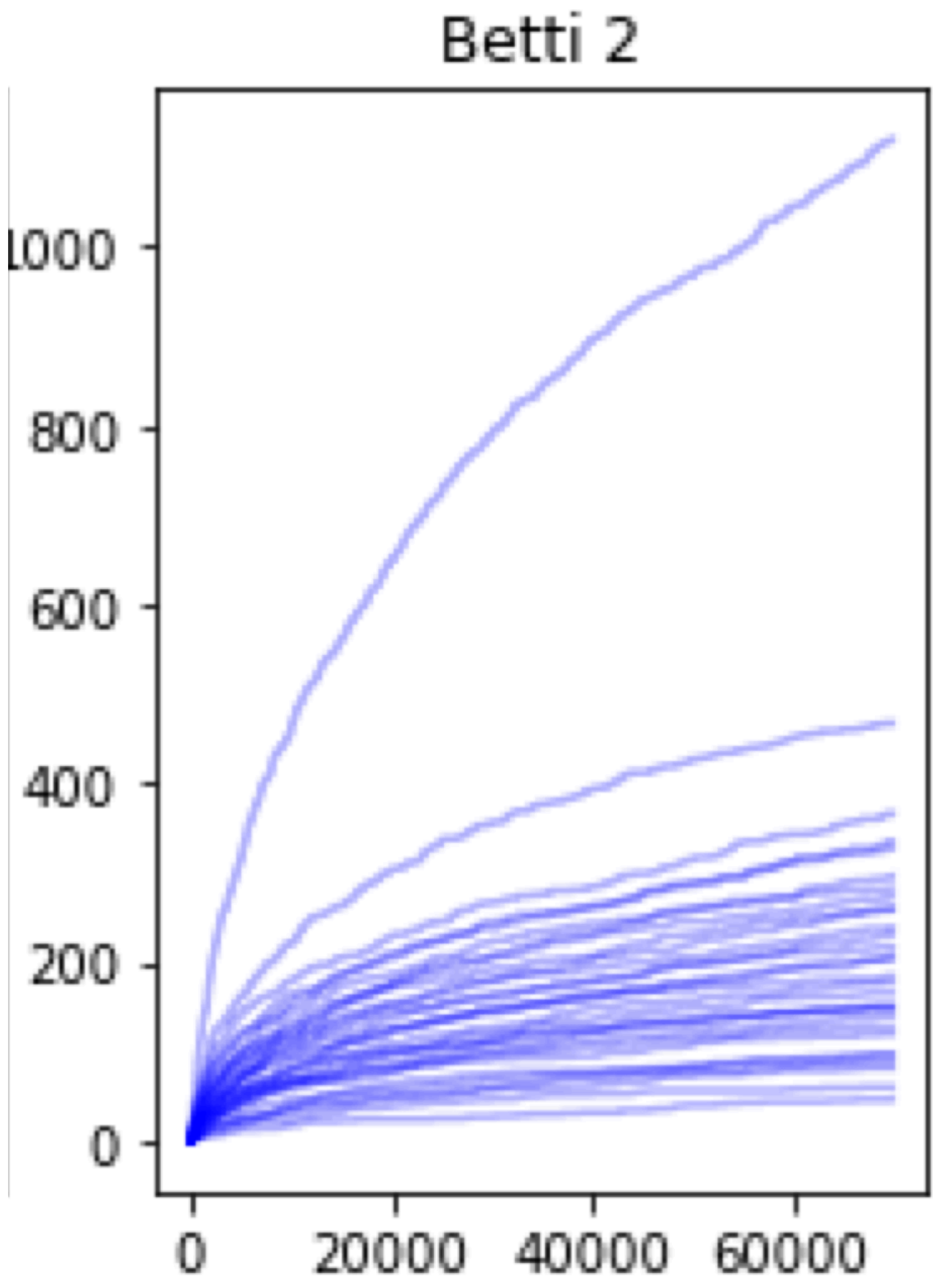
In practice???

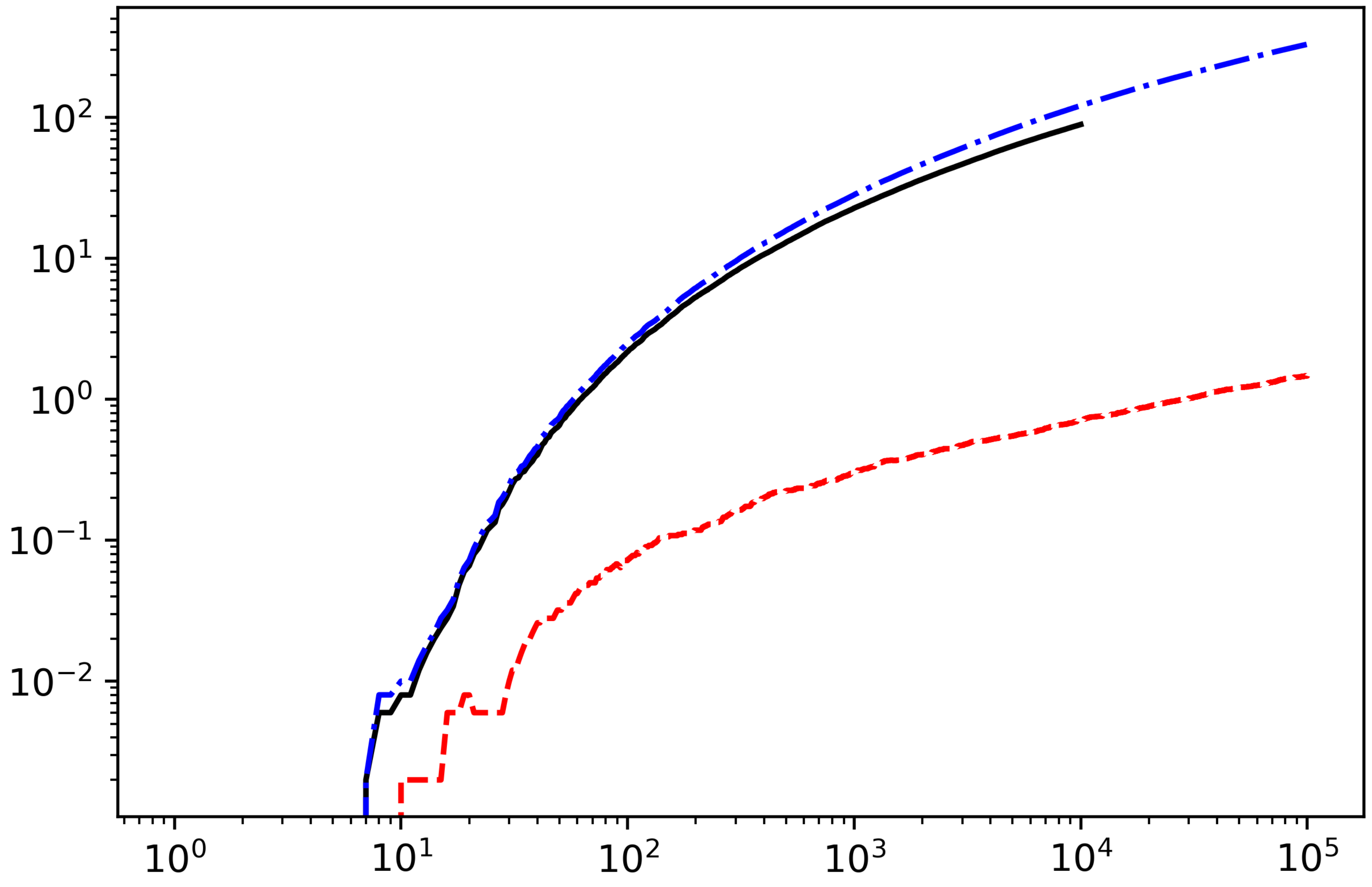
$E[\beta_2] \approx$ **num of nodes**$^{1-4x}$



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

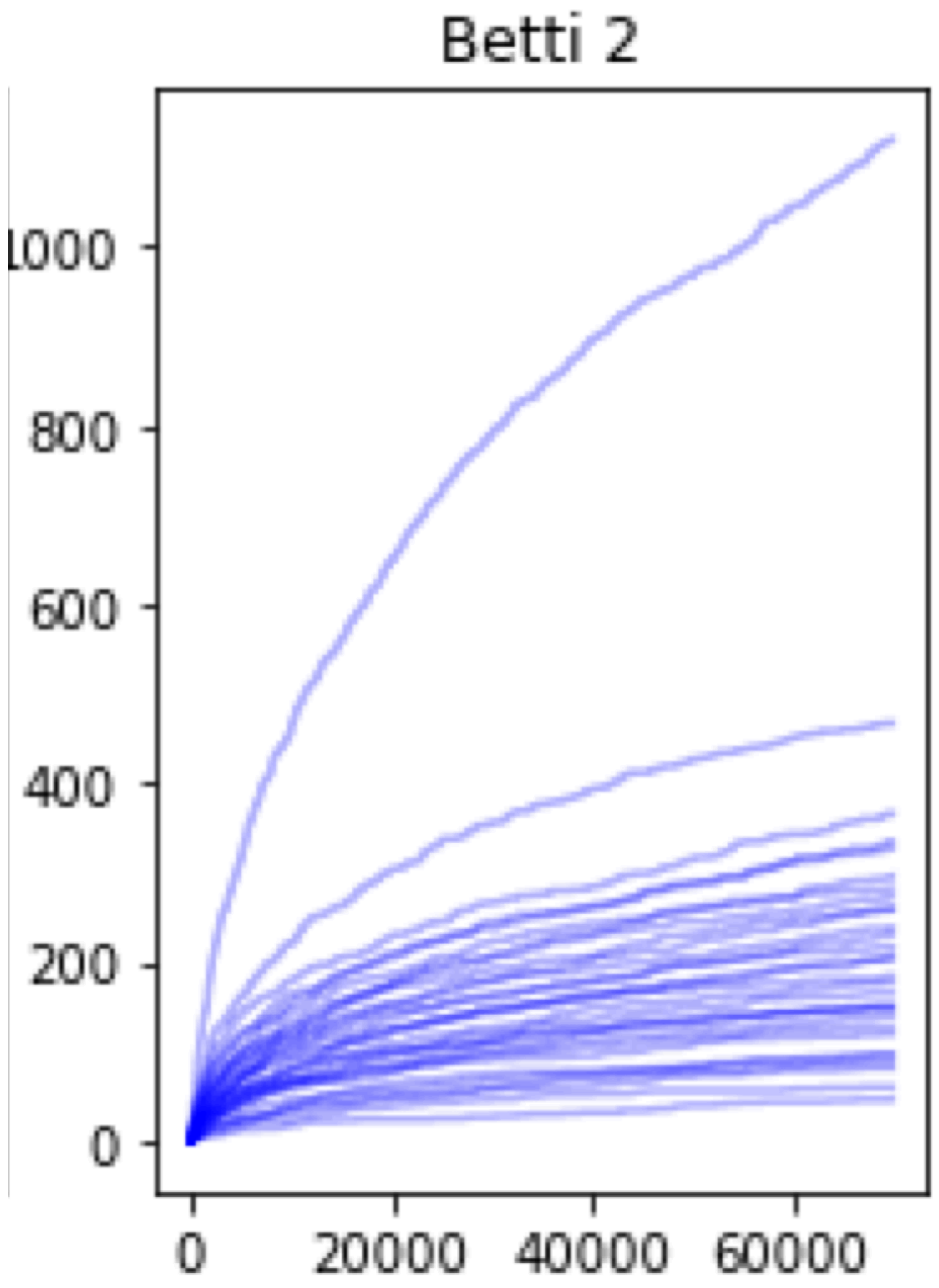
$$\log E[\beta_2] \approx (1-4x)\log(\textbf{num of nodes})$$

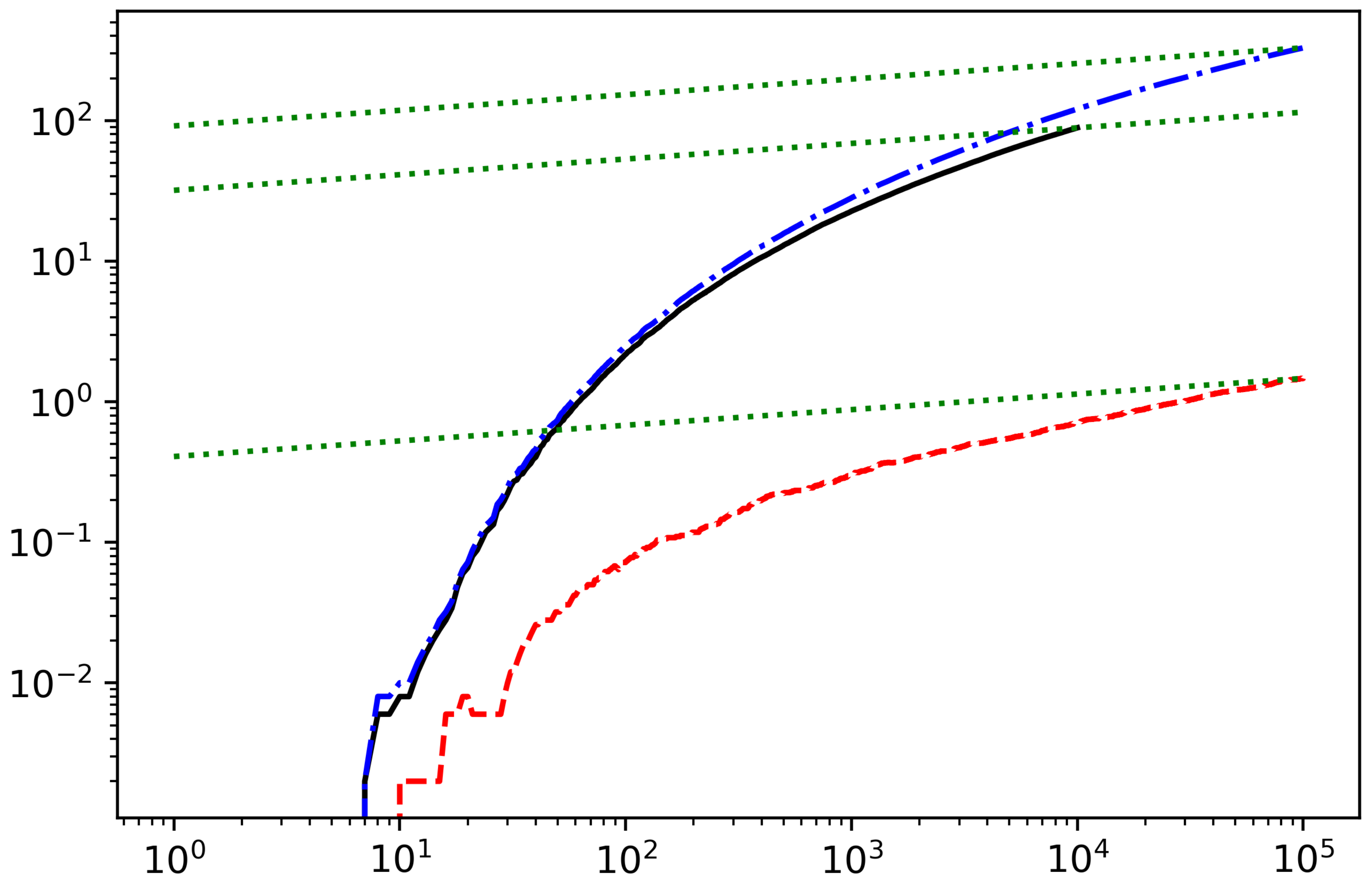




$$E[\beta_2] \approx \textbf{num of nodes}^{1-4x}$$

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





IV. 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.

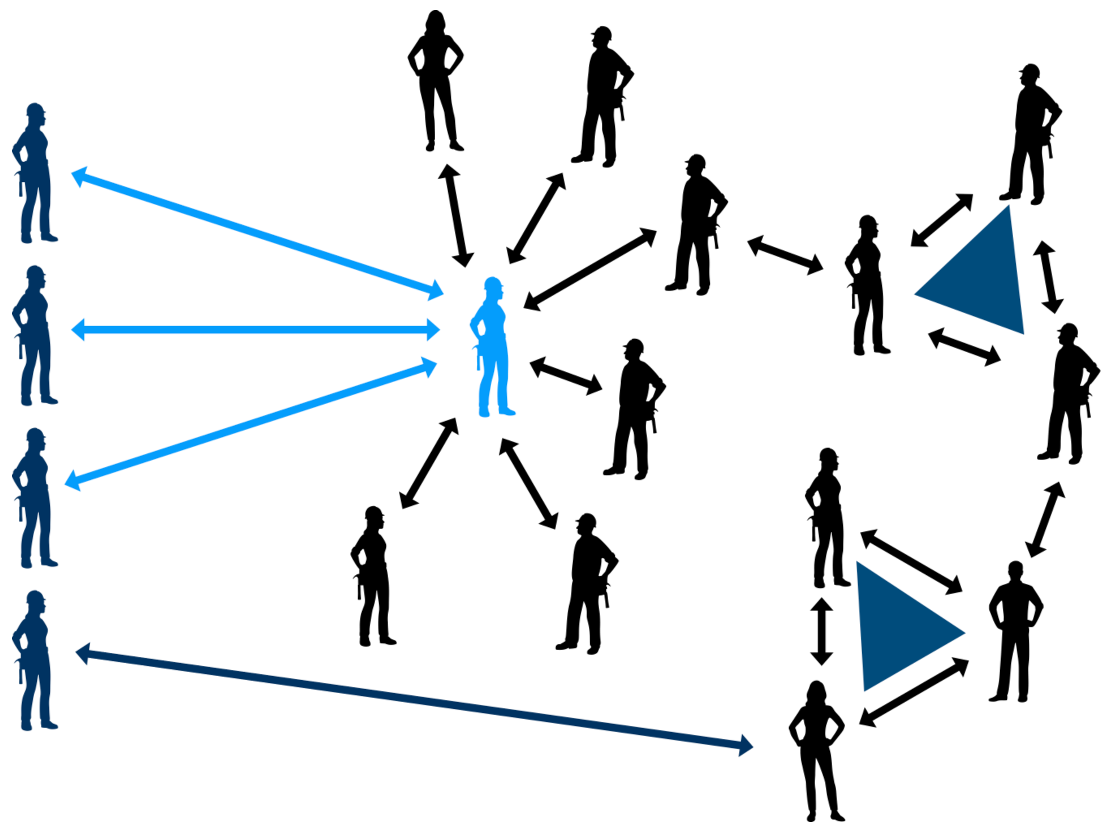
What did we learn today?

- Random topology is cool.
- Preferential attachment graph has interesting topology.

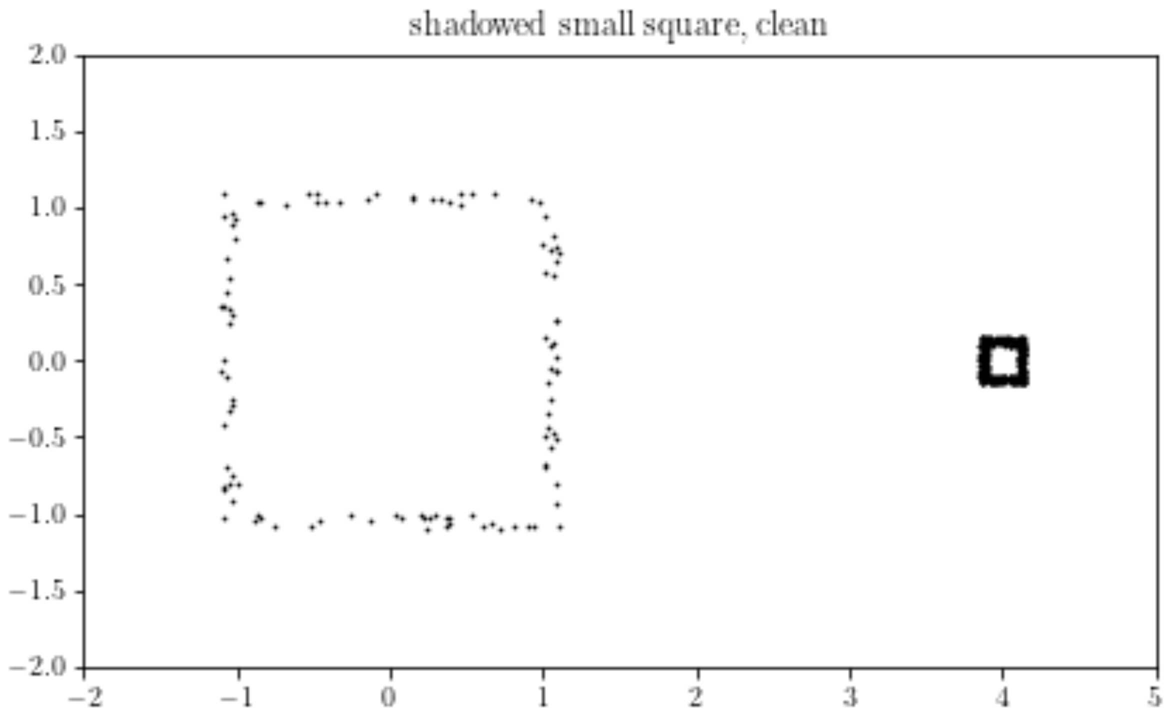
What did we learn today?

- 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



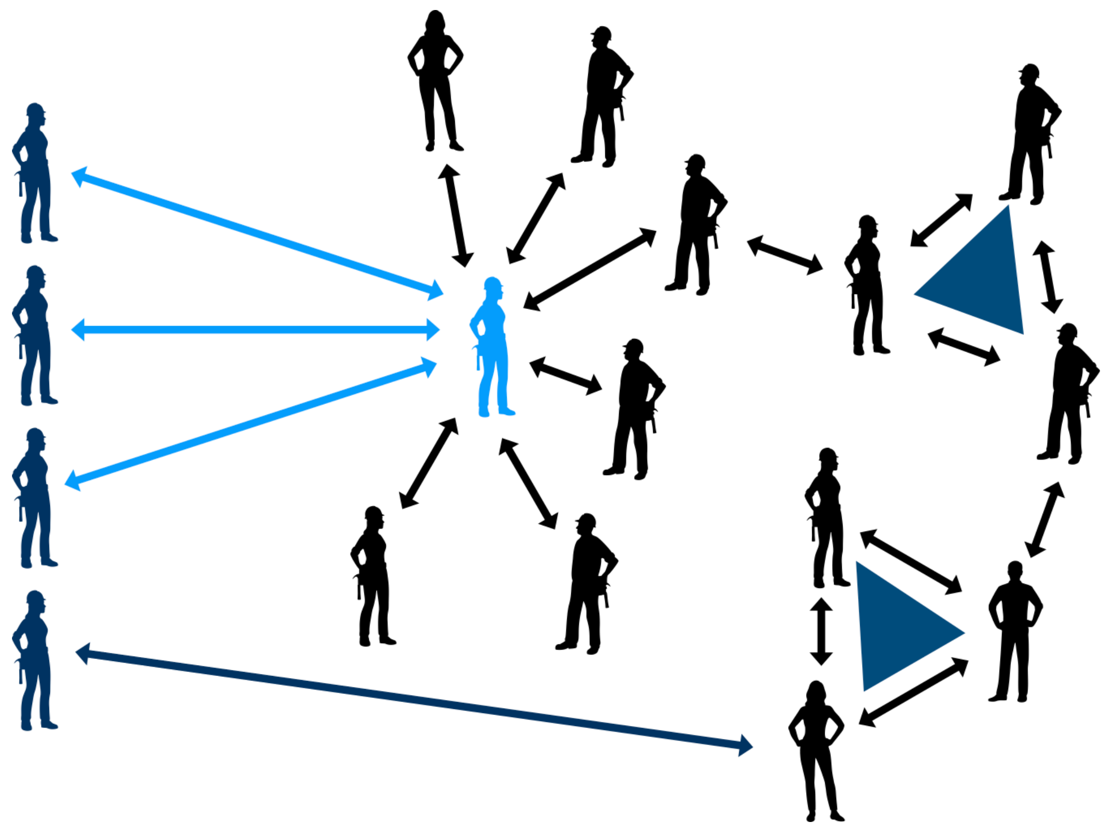
arxiv paper



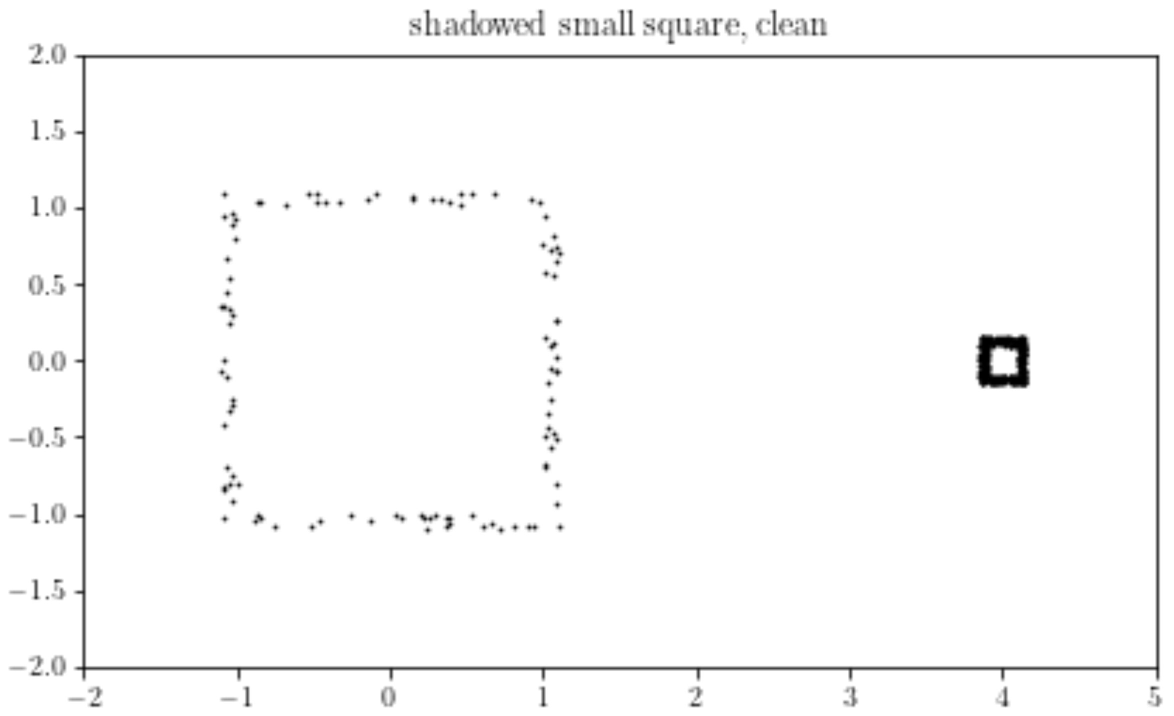
my video about small holes

Thank you!

Chunyin Siu cs2323@cornell.edu
Cornell University



arxiv paper



my video about small holes