The Topology of Preferential Attachment

How Random Interaction Begets Holes

Chunyin Siu
Cornell University
cs2323@cornell.edu

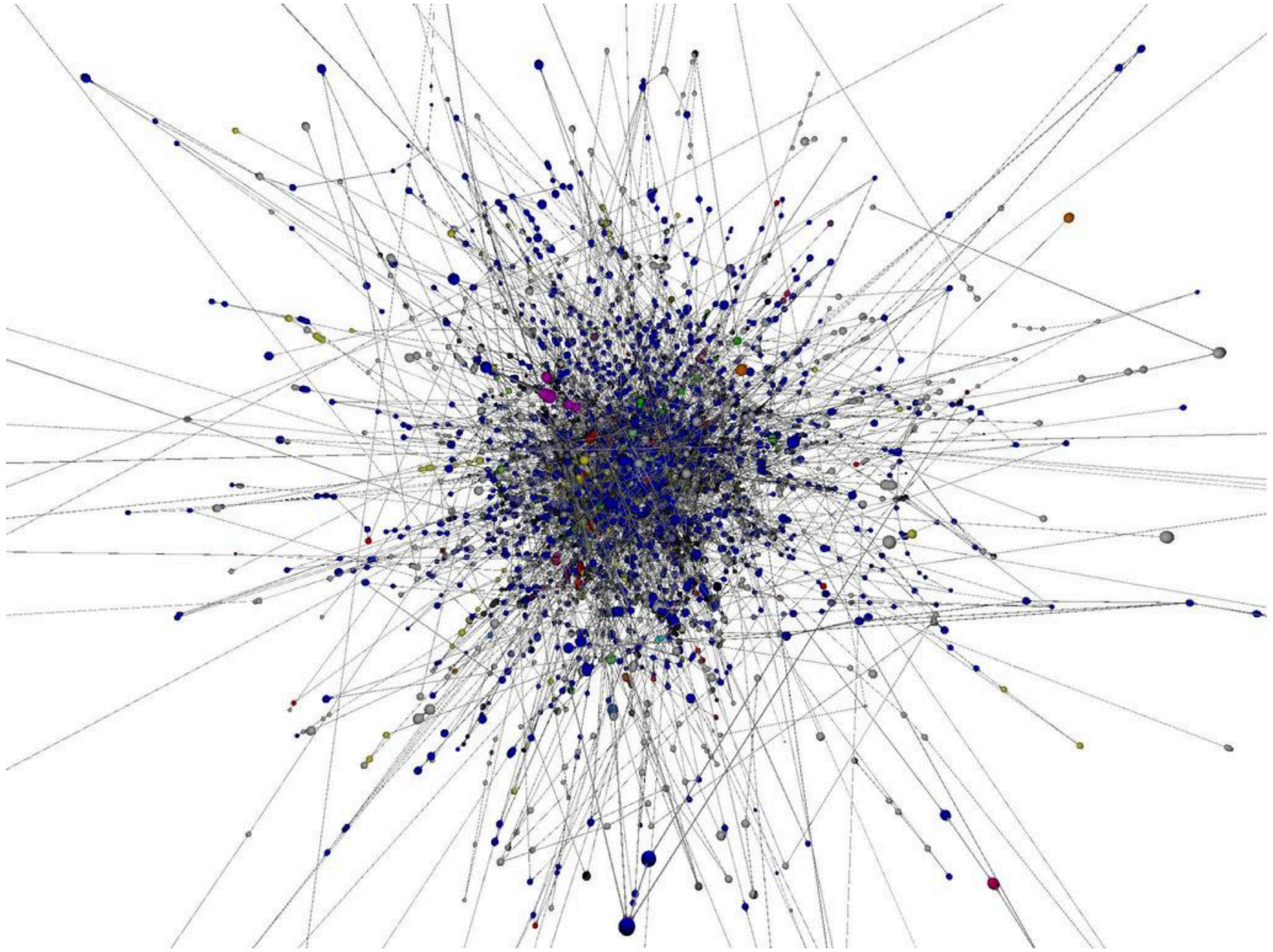
The Topology of Preferential Attachment — Theory and Computation

How Random Interaction Begets Holes

Chunyin Siu
Cornell University
cs2323@cornell.edu

postdoc for 24/25

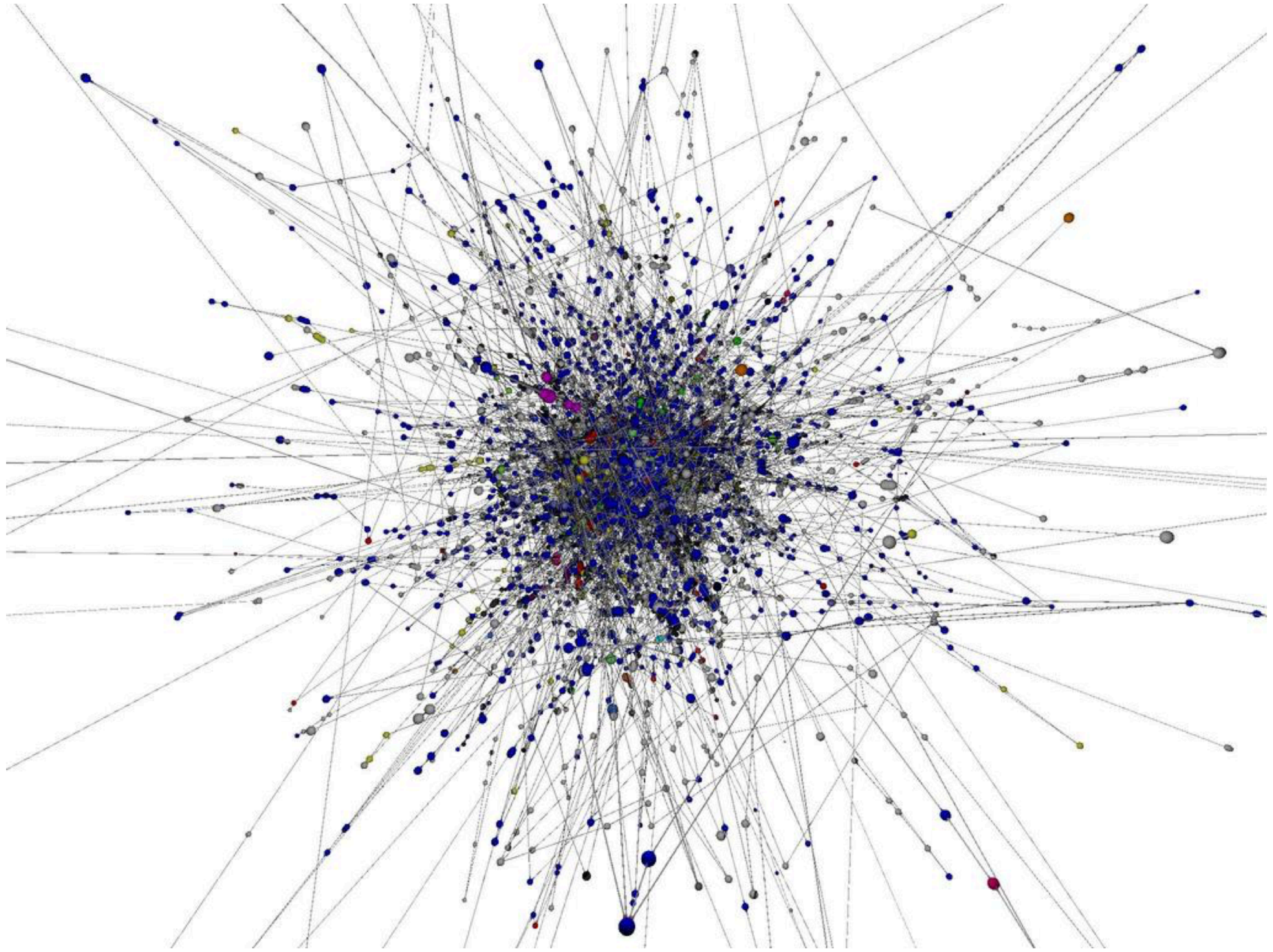
So, preferential attachment…



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

So, preferential attachment…

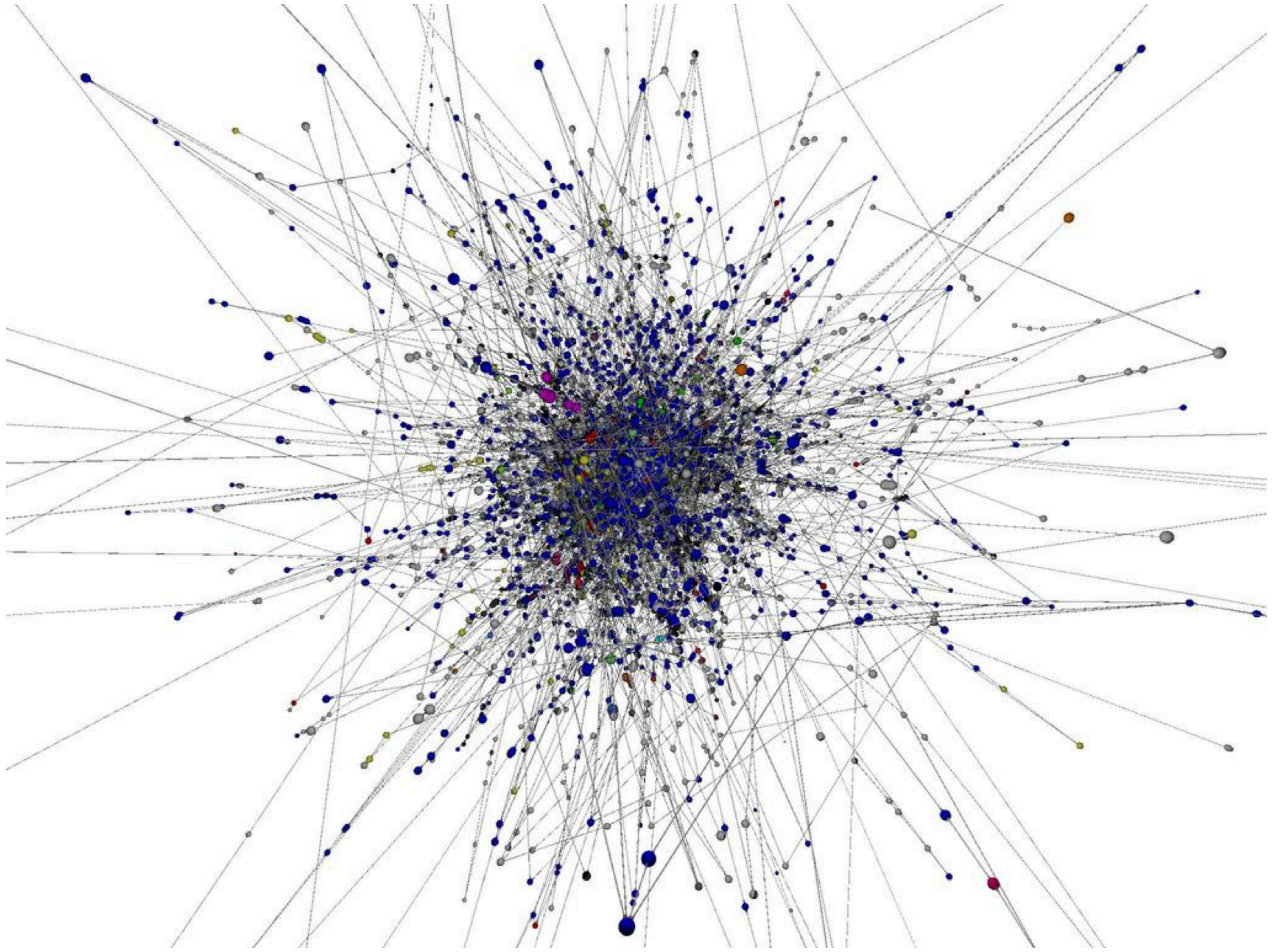
- Just a bouquet of circles?



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

So, preferential attachment…

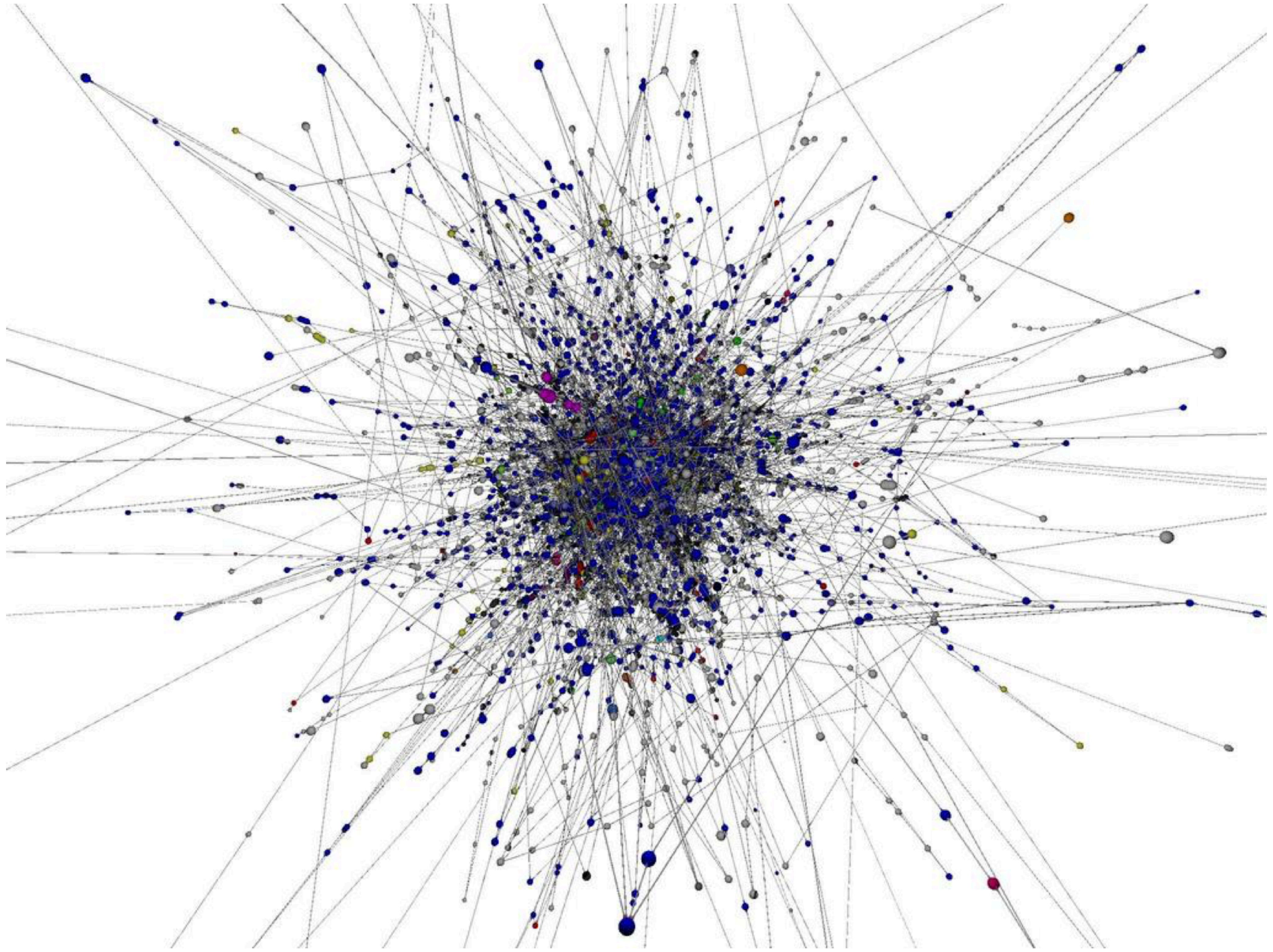
- Just a bouquet of circles?
- What is intrinsic and what is just random fluctuation?



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

So, preferential attachment…

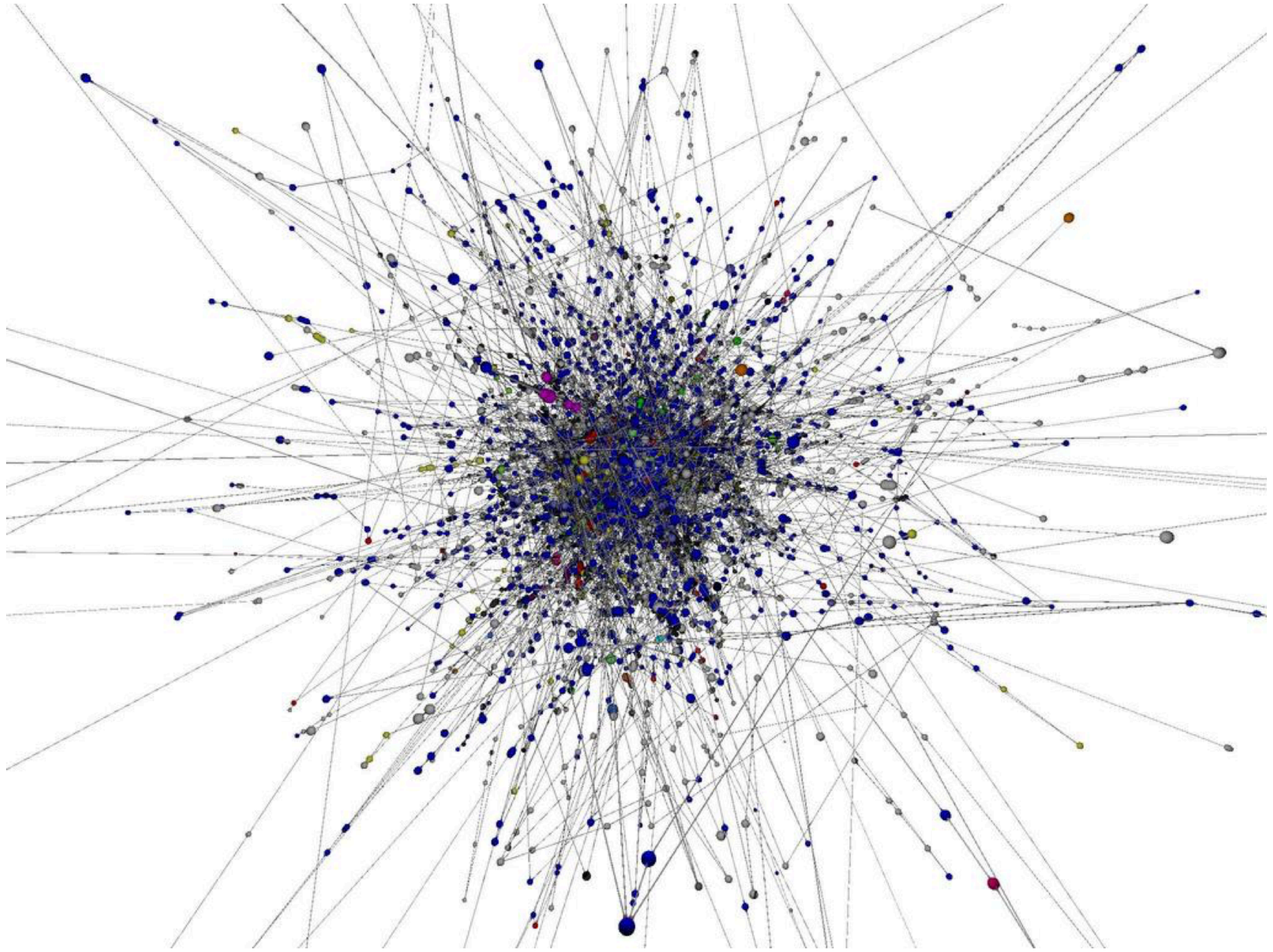
- Just a bouquet of circles?
- What is intrinsic and what is just random fluctuation?
- —> random topology



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

So, preferential attachment…

- Just a bouquet of circles?
- What is intrinsic and what is just random fluctuation?
- —> random topology
  - the random process of preferential attachment



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

Agenda

random topology

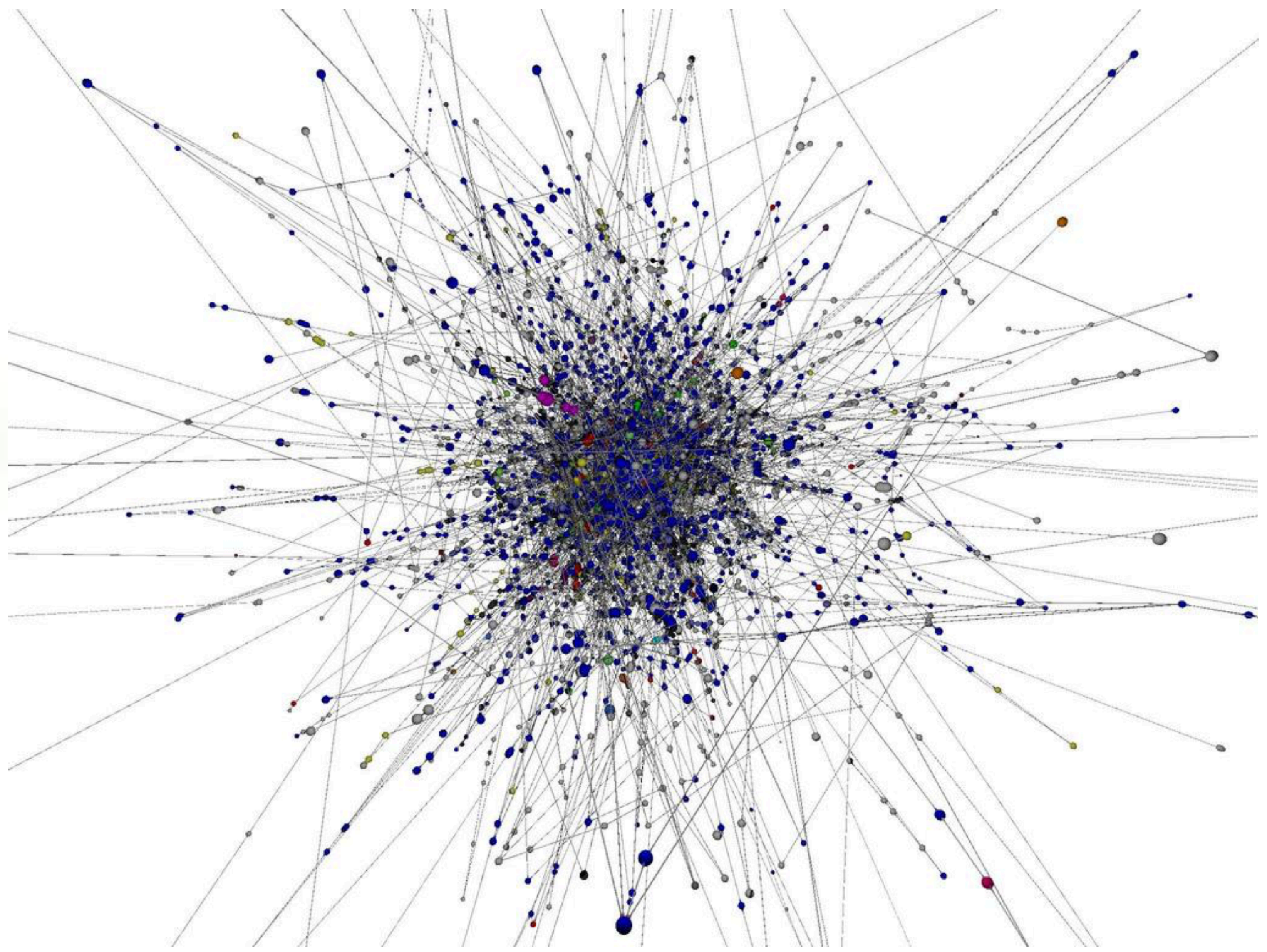
Agenda

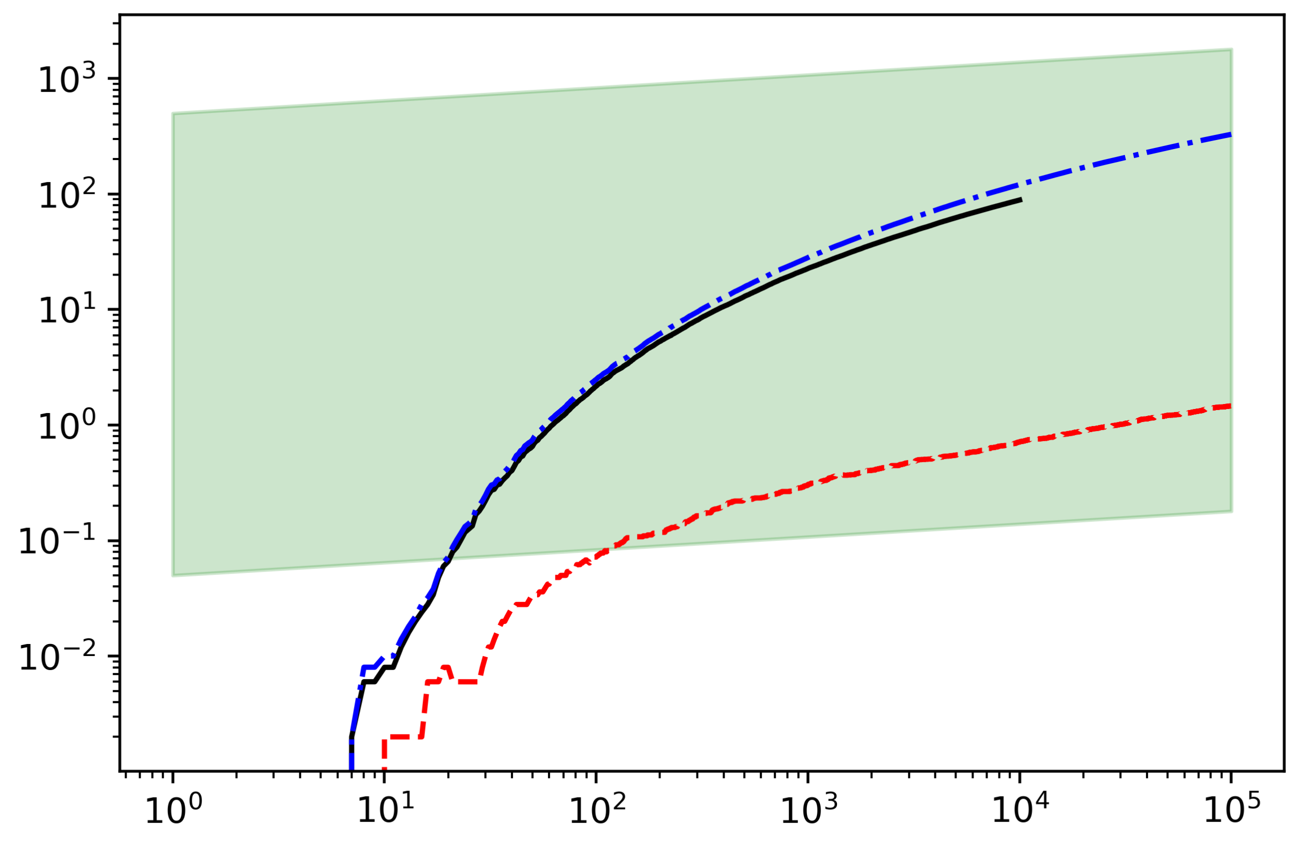
random topology

preferential attachment

Agenda


random topology

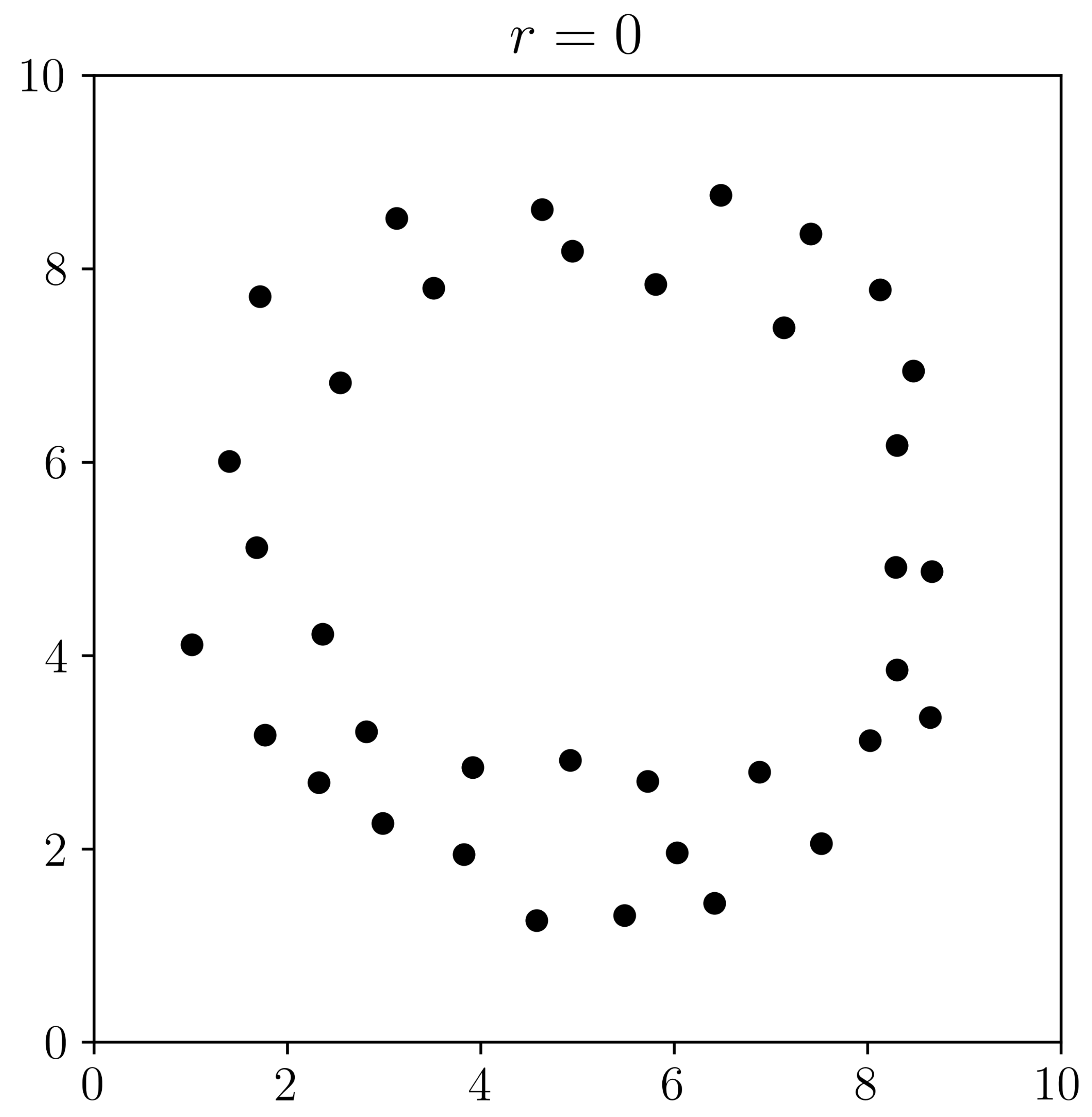

preferential attachment



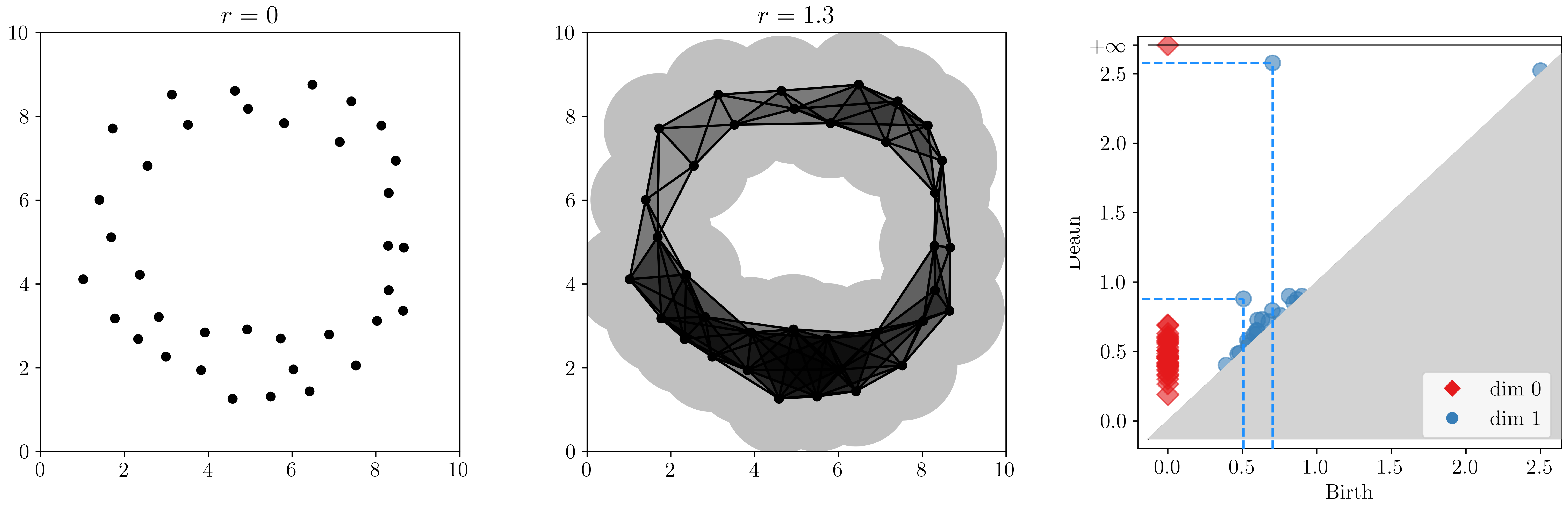
our result

I. A Probabilist’s Apology

Why Random Topology and What we Know

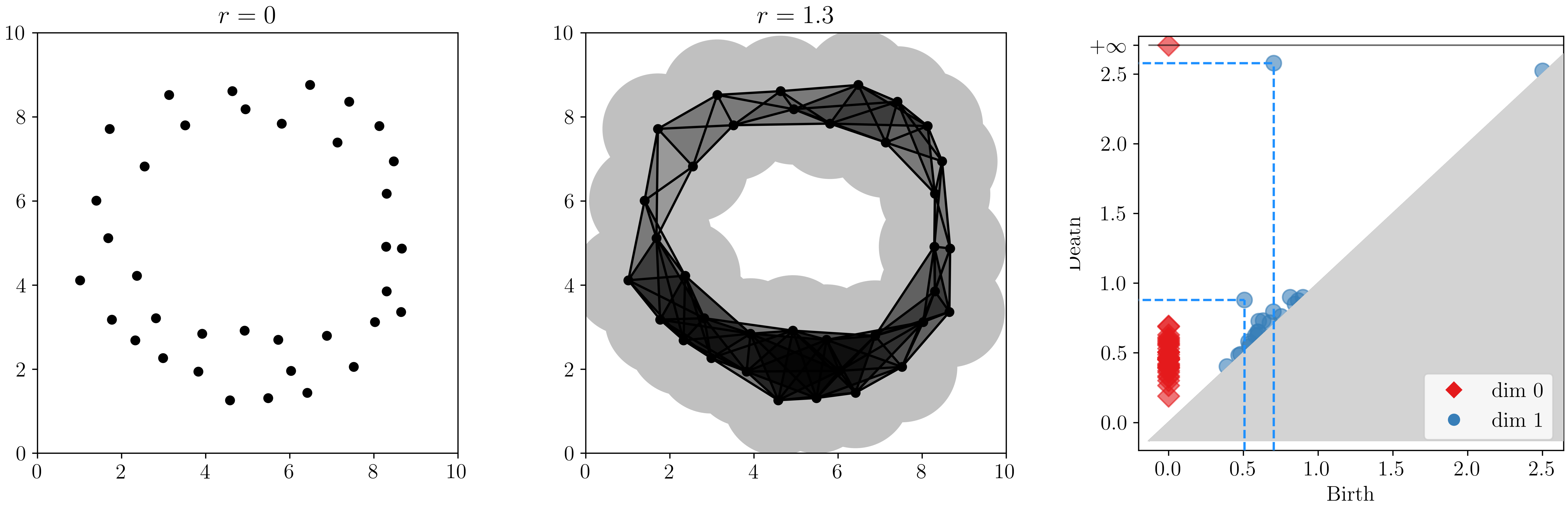


$r = 0$
0
2
4
6
8
10
0
2
4
6
8
10

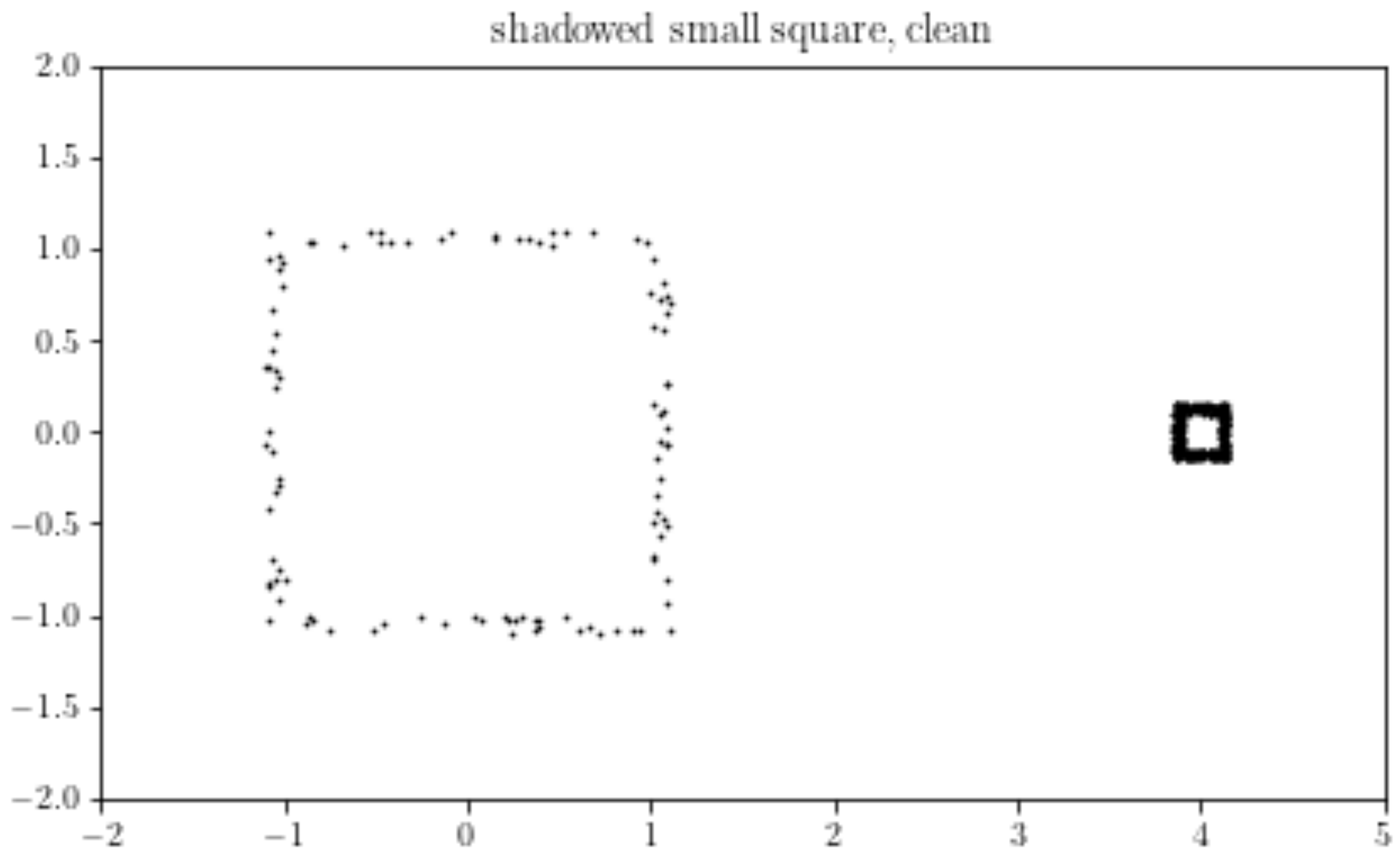


plots generated by Andrey Yao

Size is Signal

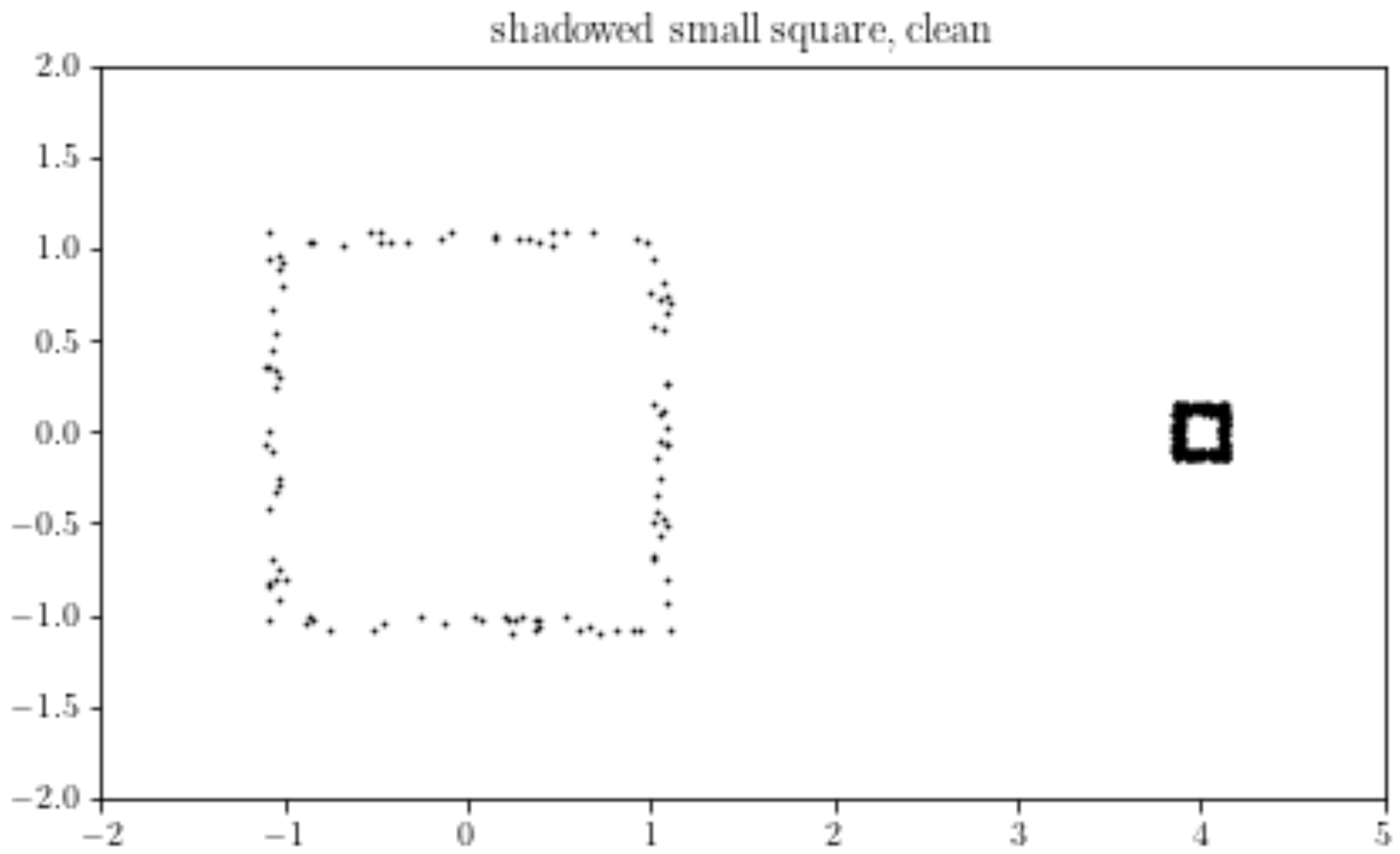


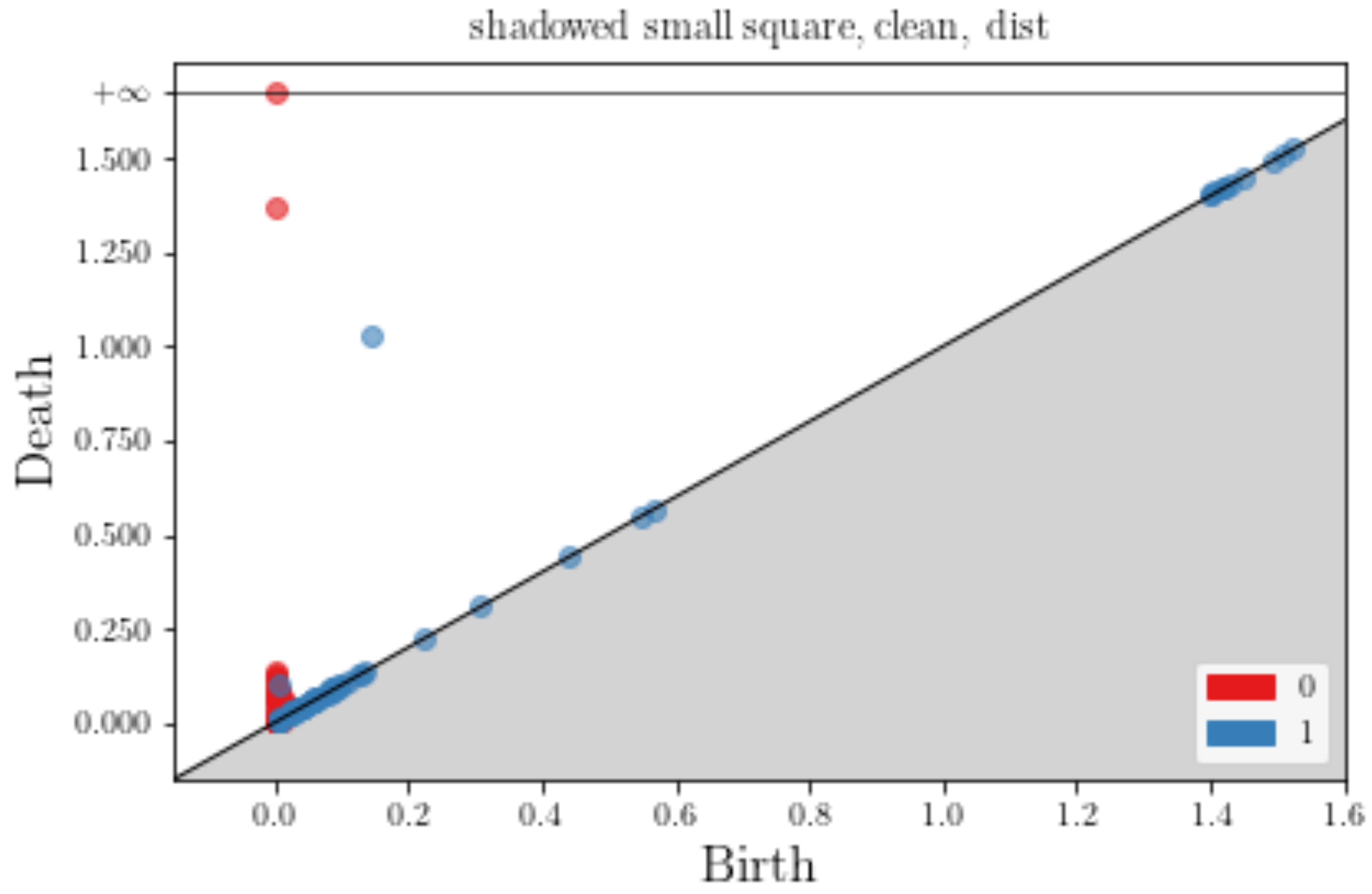
Or is it?



shadowed small square, clean
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
−2
−1
0
1
2
3
4
5

Or is it?


shadowed small square, clean
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
−2
−1
0
1
2
3
4
5

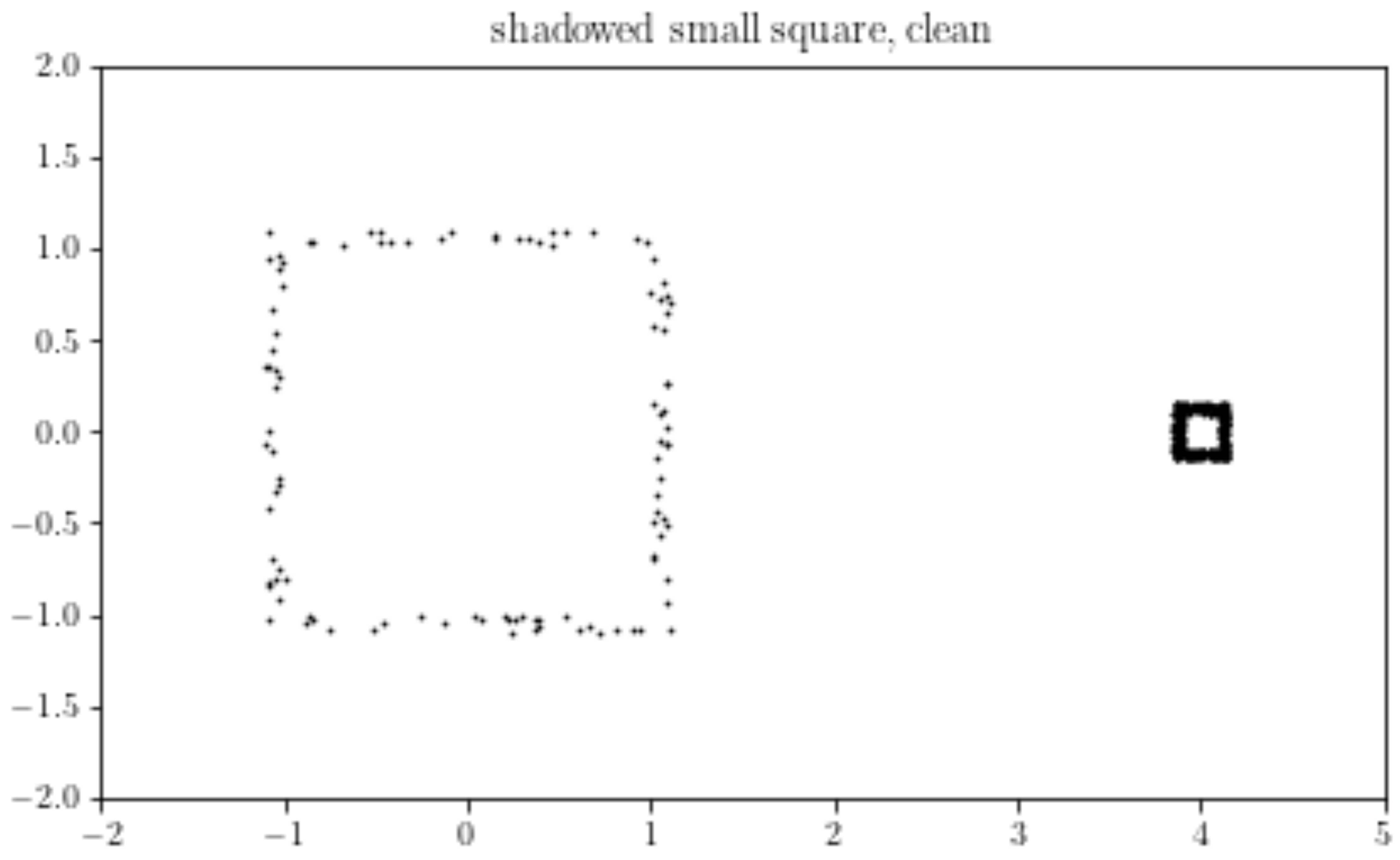

shadowed small square, clean, dist
+∞
1.500
1.250
1.000
0.750
0.500
0.250
0.000
Death
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
Birth
0
1

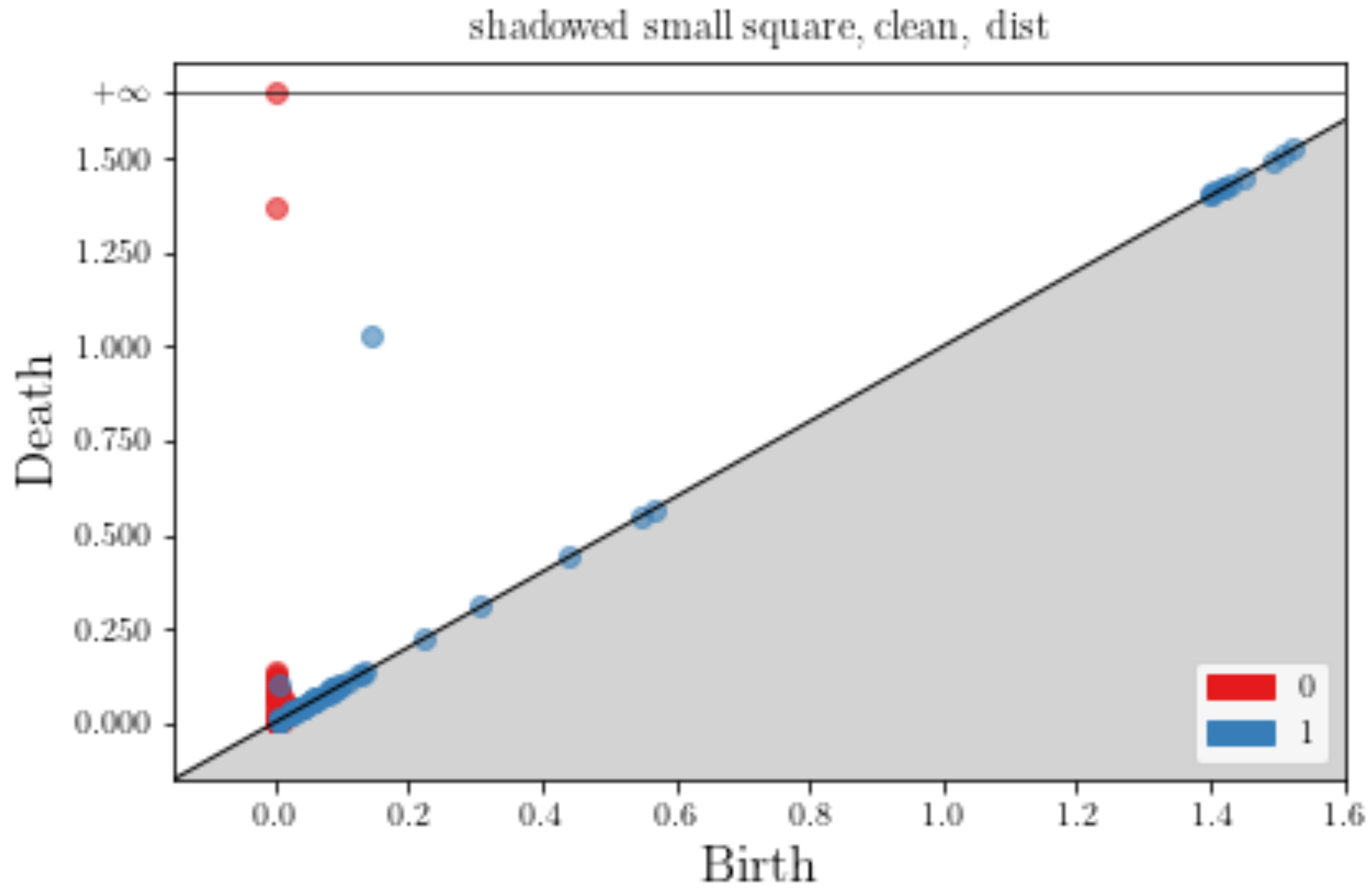
Size is Signal?

Surprise

~~Size~~ is Signal.

Random points don't do that.





Signal is what is not random.

Signal is what is not **random.**
So what is **random?**

What we know

[not meant to be complete]

What we know
[not meant to be complete]

- Erdos-Renyi clique complexes

What we know
[not meant to be complete]

- Erdos-Renyi clique complexes
  - Kahle 2009, 2014
  - Kahle and Meckes 2013
  - Costa et al 2015
  - Malen 2023
  - etc

What we know
[not meant to be complete]

- Erdos-Renyi clique complexes
  - Kahle 2009, 2014
  - Kahle and Meckes 2013
  - Costa et al 2015
  - Malen 2013
  - etc
- random geometric complexes

What we know
[not meant to be complete]

- Erdos-Renyi clique complexes
  - Kahle 2009, 2014
  - Kahle and Meckes 2013
  - Costa et al 2015
  - Malen 2013
  - etc

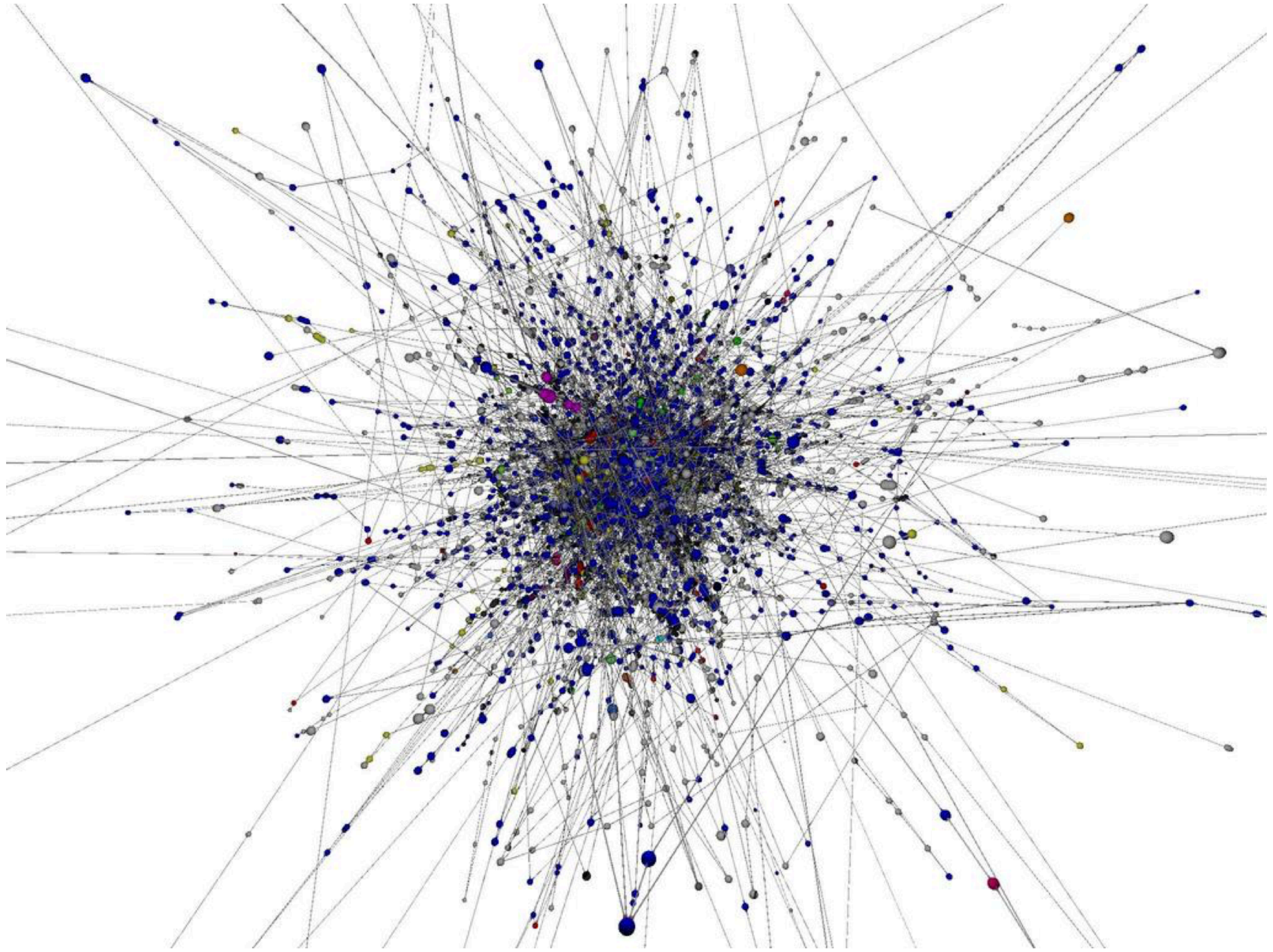
- random geometric complexes
  - Kahle 2011
  - Kahle and Meckes 2013
  - Yogeshwaran and Adler 2015
  - Bobrowski et al 2017
  - Hiraoka et al 2018
  - Thomas and Owada 2021a, b
  - Owada and Wei 2022
  - etc

II. Preferential Attachment

Beyond independence and homogeneity

Independent and identically distributed?

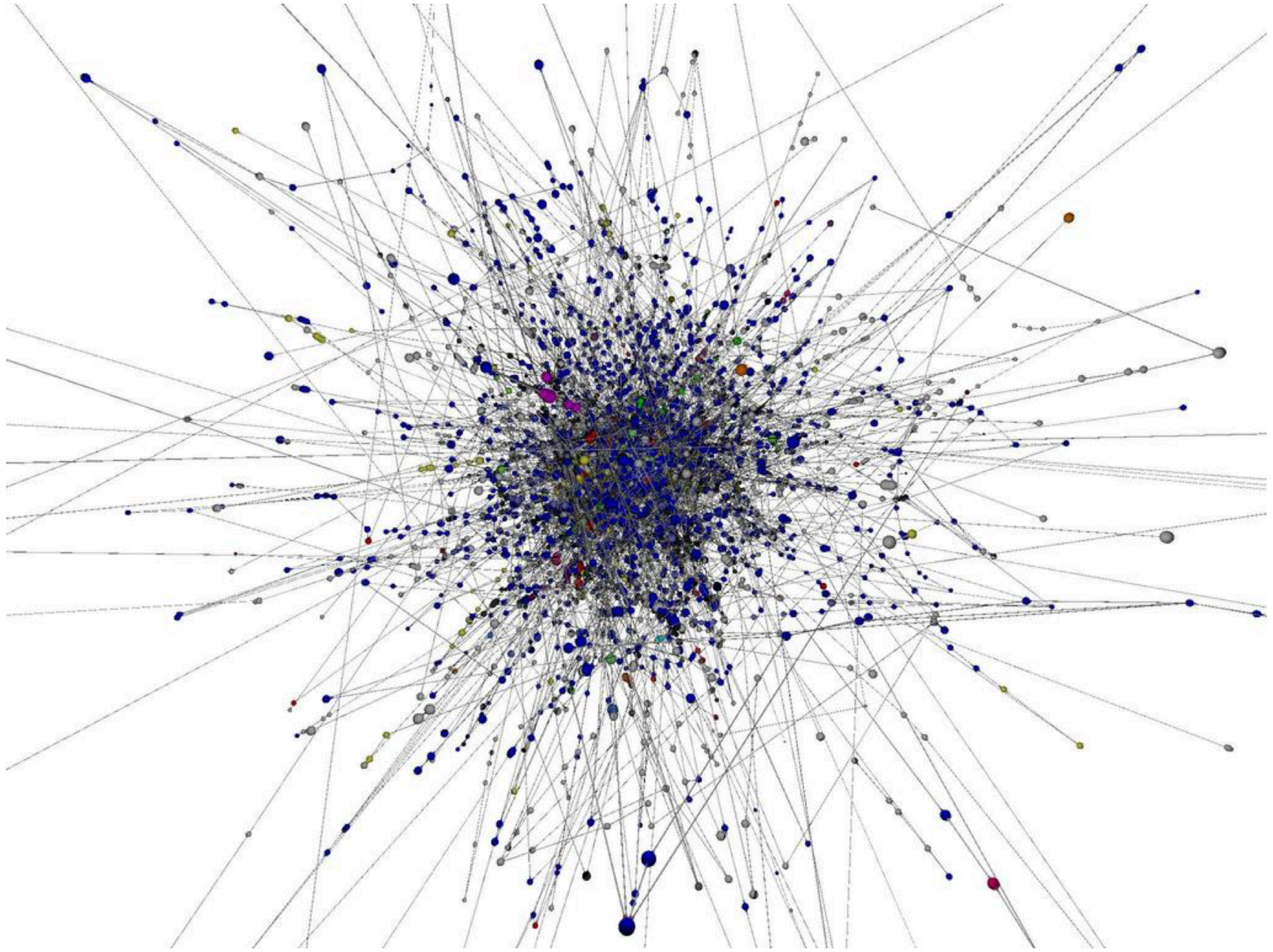
Independent and identically distributed?



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

Preferential Attachment

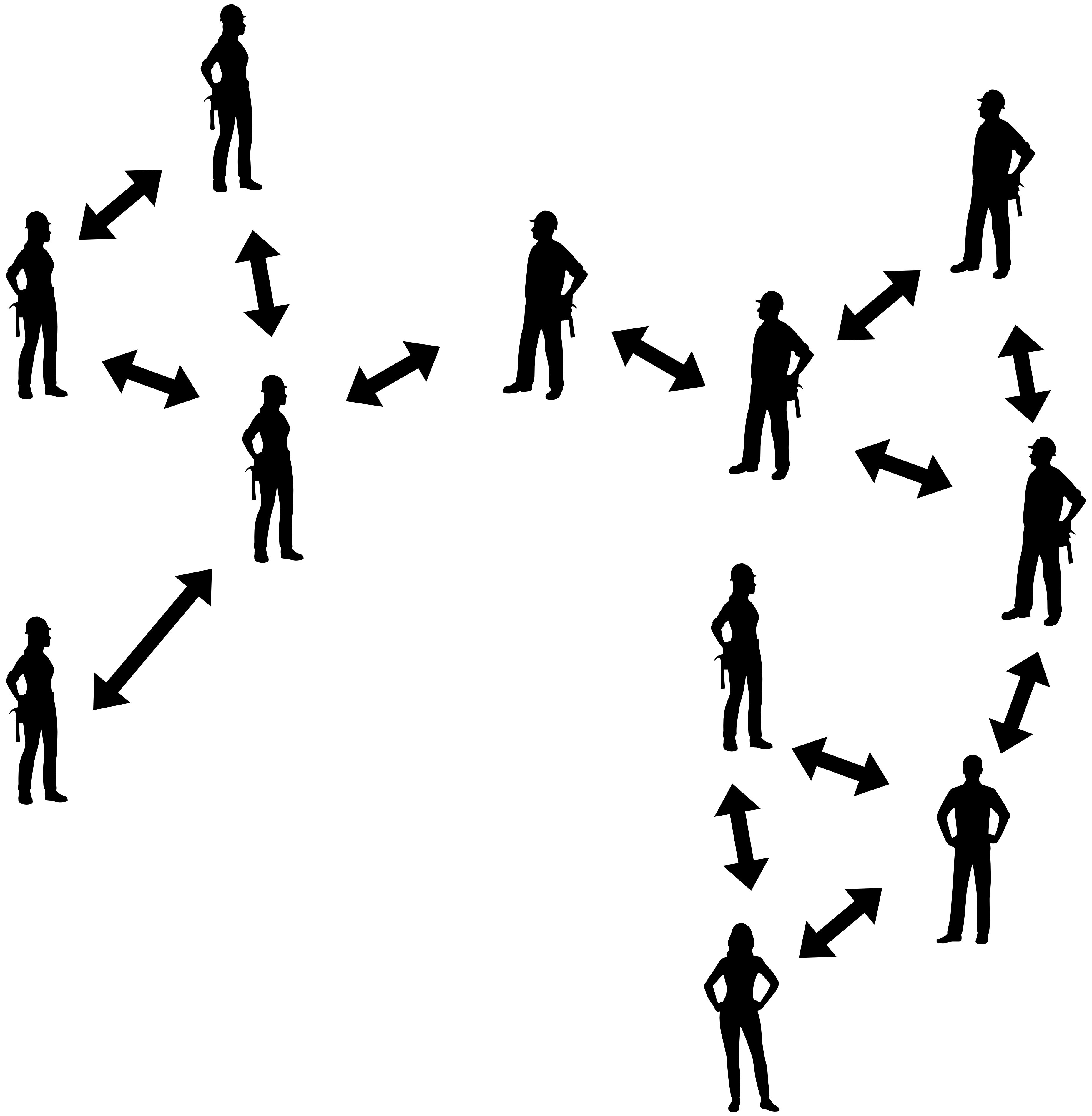
[Albert and Barabasi 1999]



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

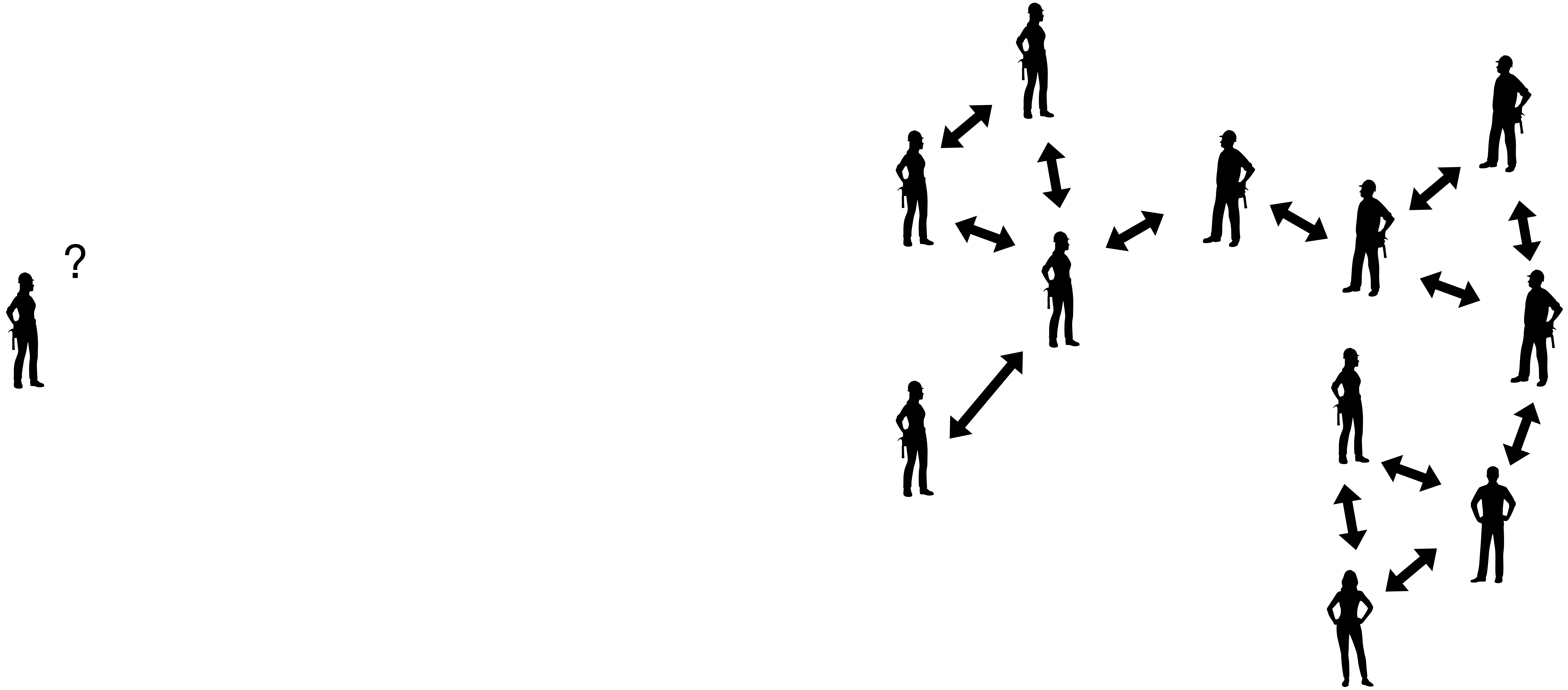
Preferential Attachment

[Albert and Barabasi 1999]



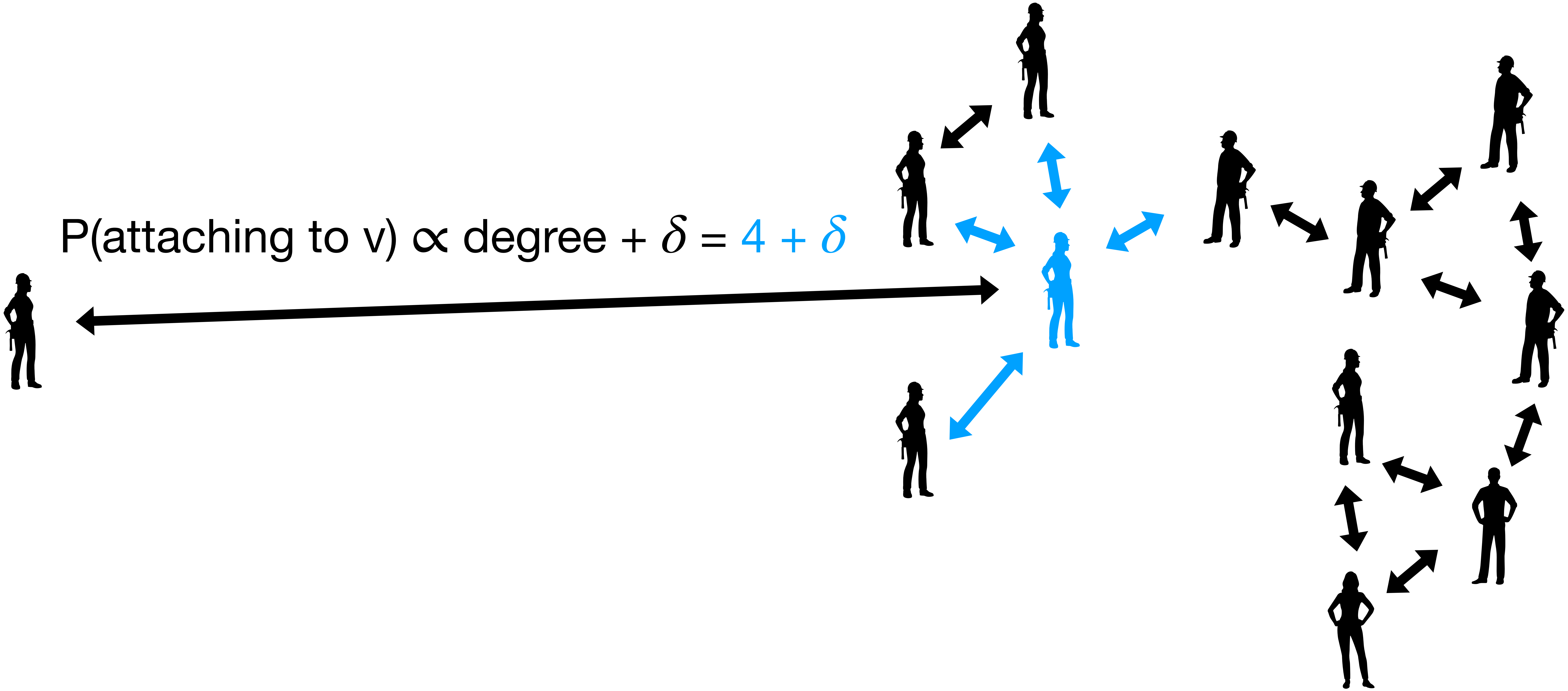
Preferential Attachment

[Albert and Barabasi 1999]



Preferential Attachment

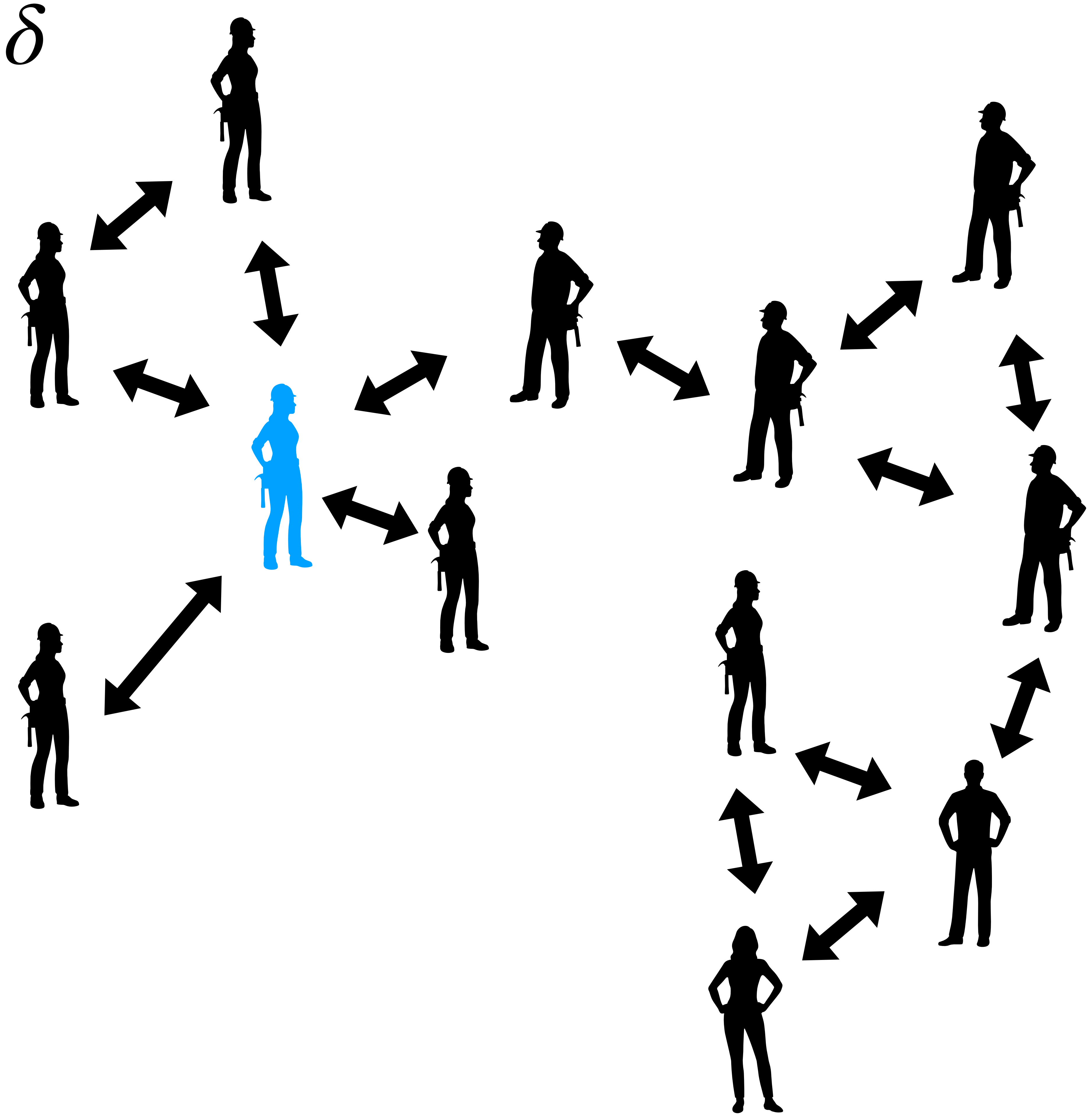
[Albert and Barabasi 1999]



Preferential Attachment

[Albert and Barabasi 1999]

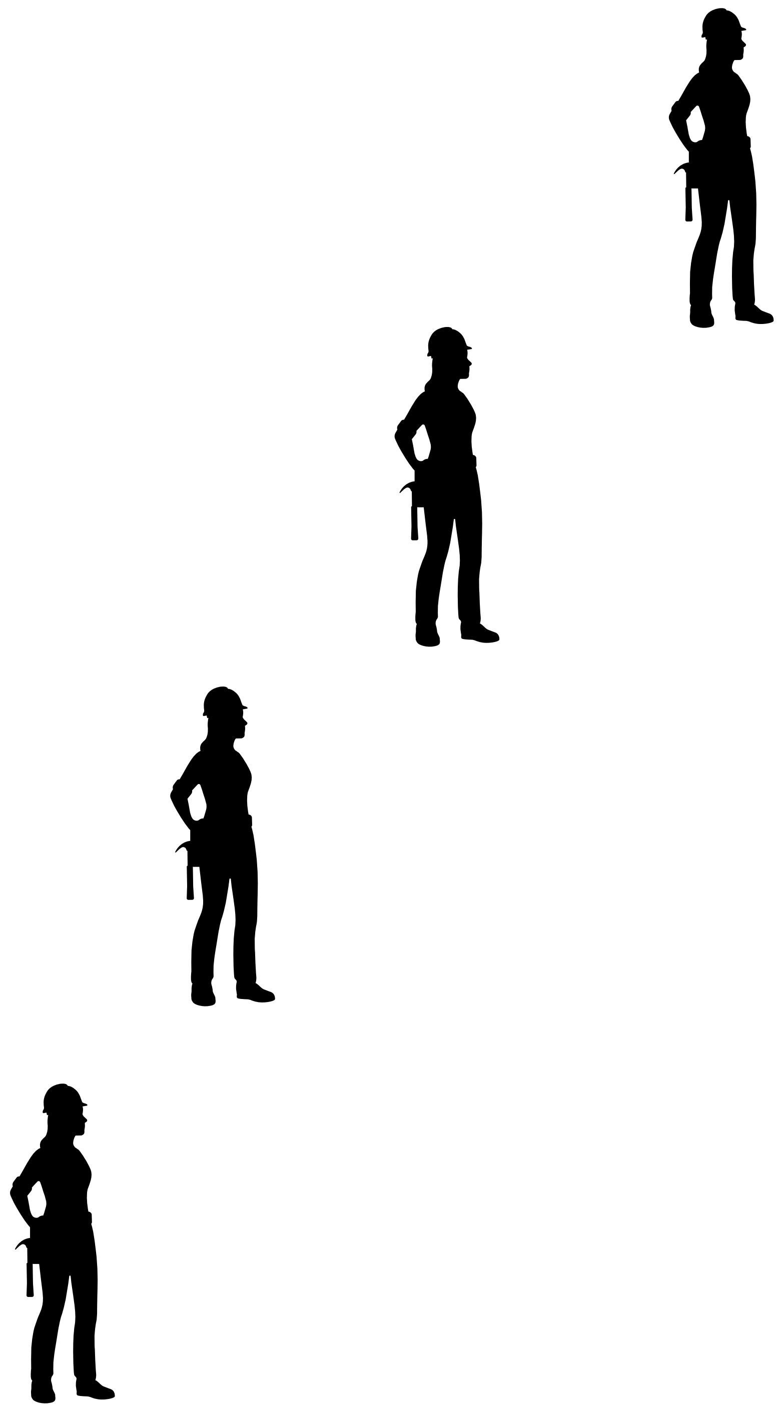
P(attaching to v) $\propto$ degree + a tuning parameter δ

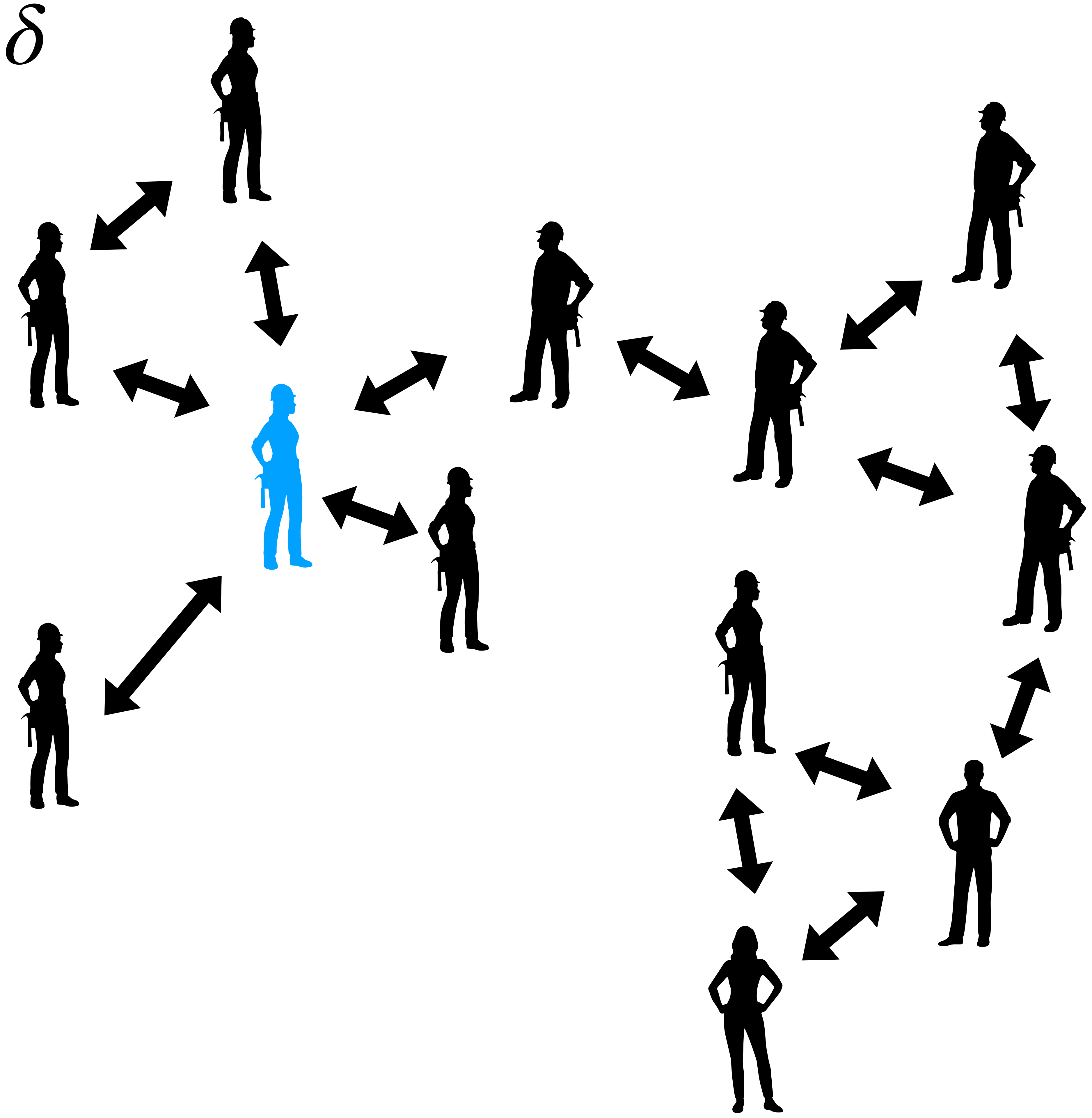


Preferential Attachment

[Albert and Barabasi 1999]

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

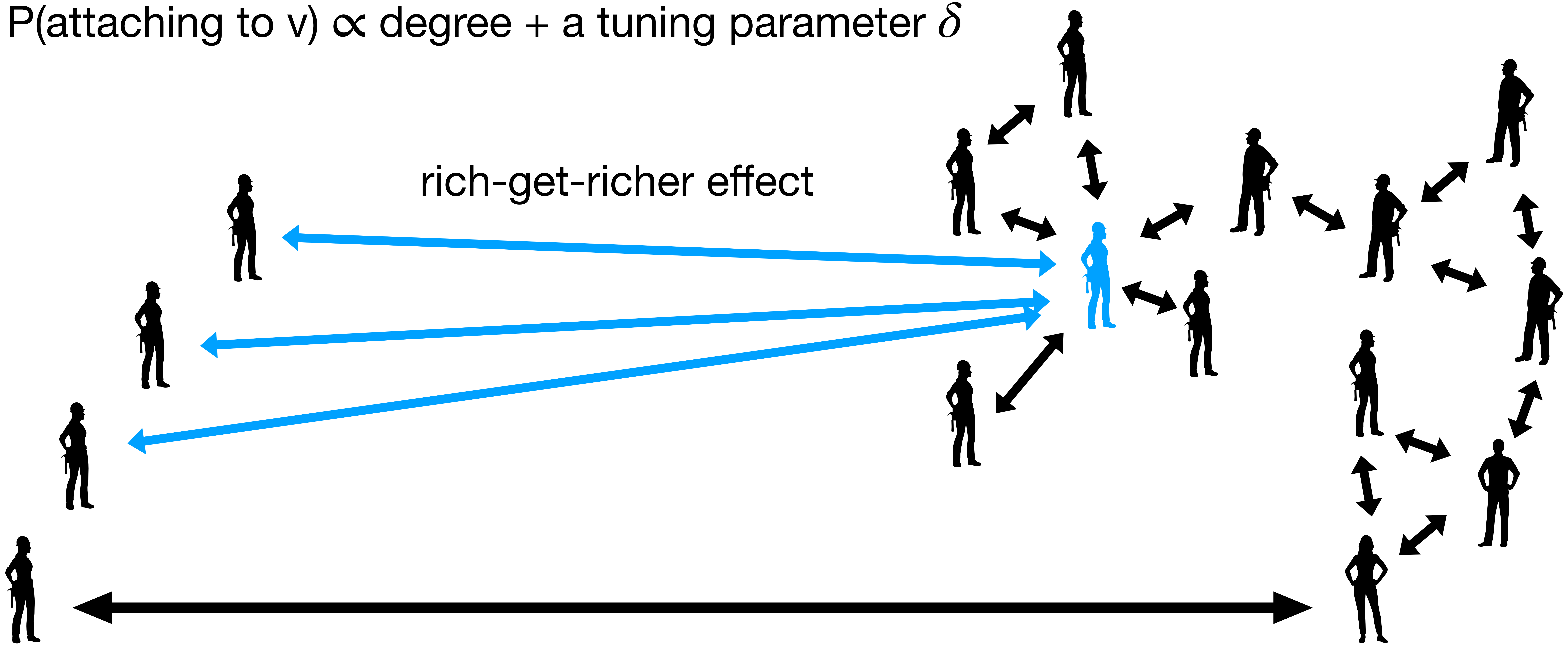




Preferential Attachment

[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?

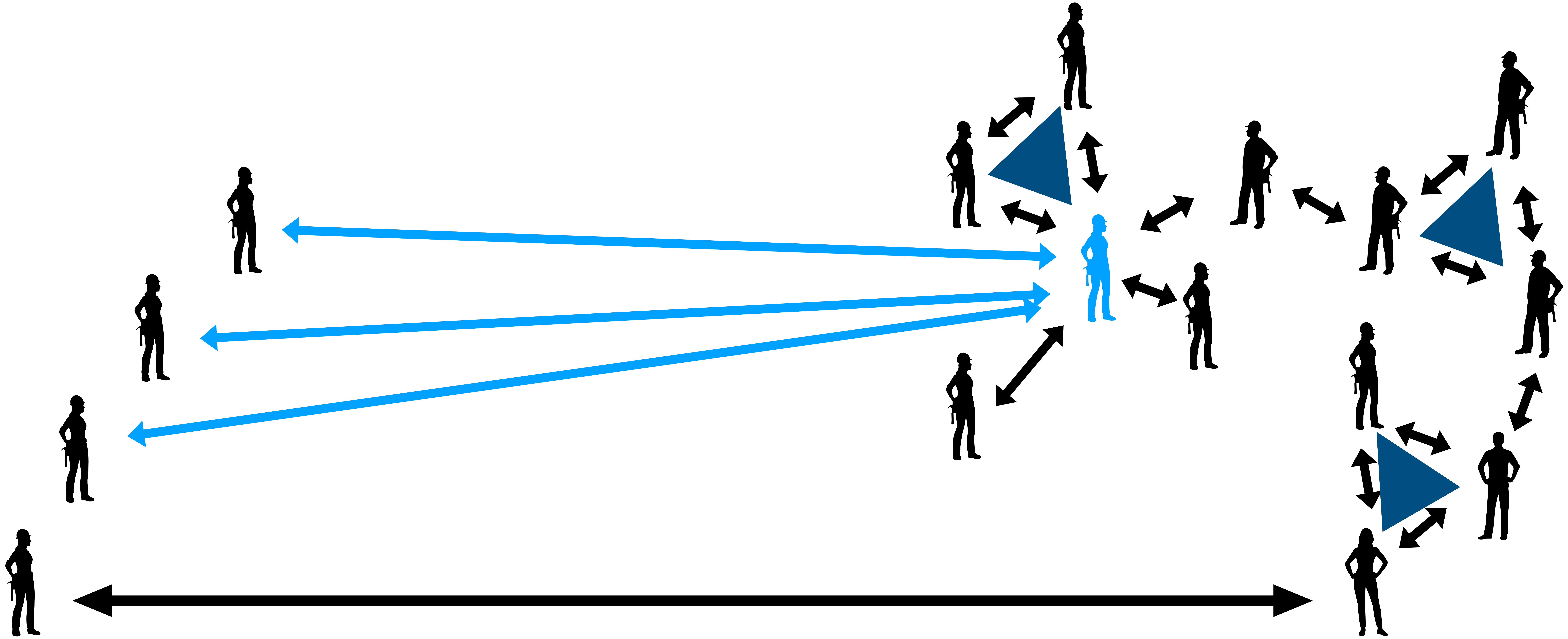
- triangle counts and clustering coefficient [Bollobas and Ridden 2002, Prokhorenkova et al 2013]
- subgraph counts [Garavaglia and Steghuis 2019]

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…

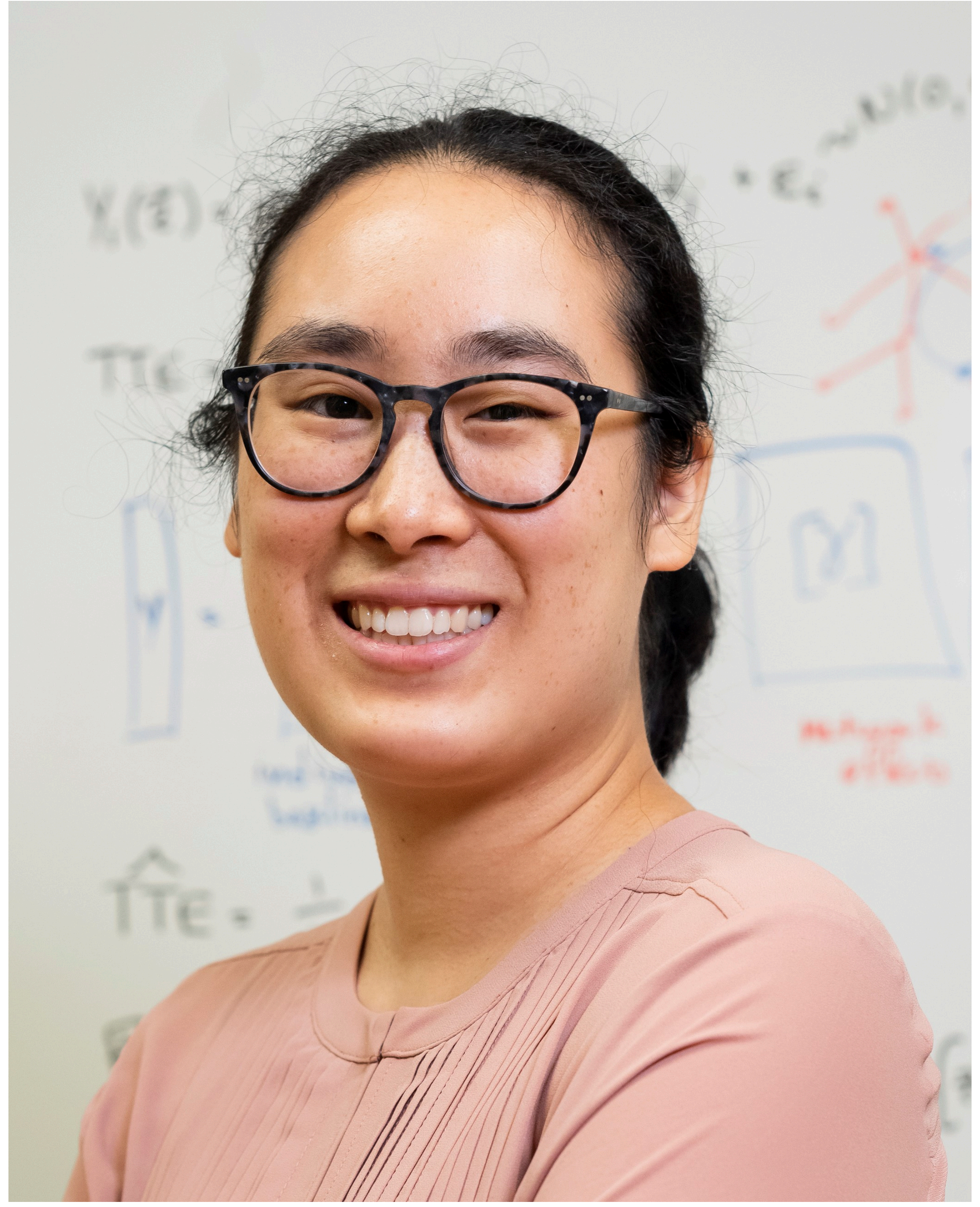
Clique Complex

aka Flag Complex



III Topology of Preferential Attachment

My Lovely Collaborators



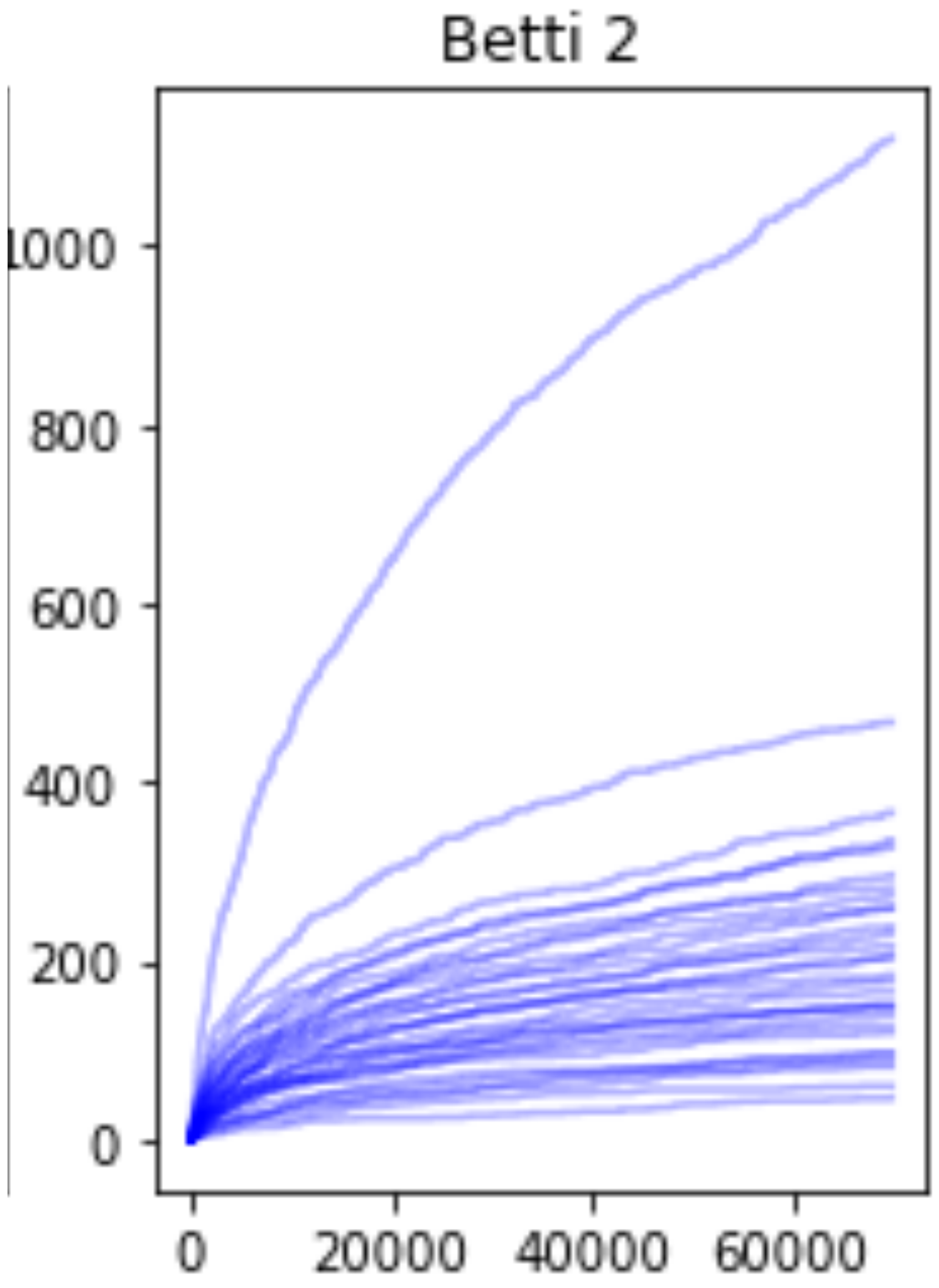
Christina Lee Yu

Gennady Samorodnitsky

Rongyi He (Caroline)

Expected Betti Number $E[\beta_q]$

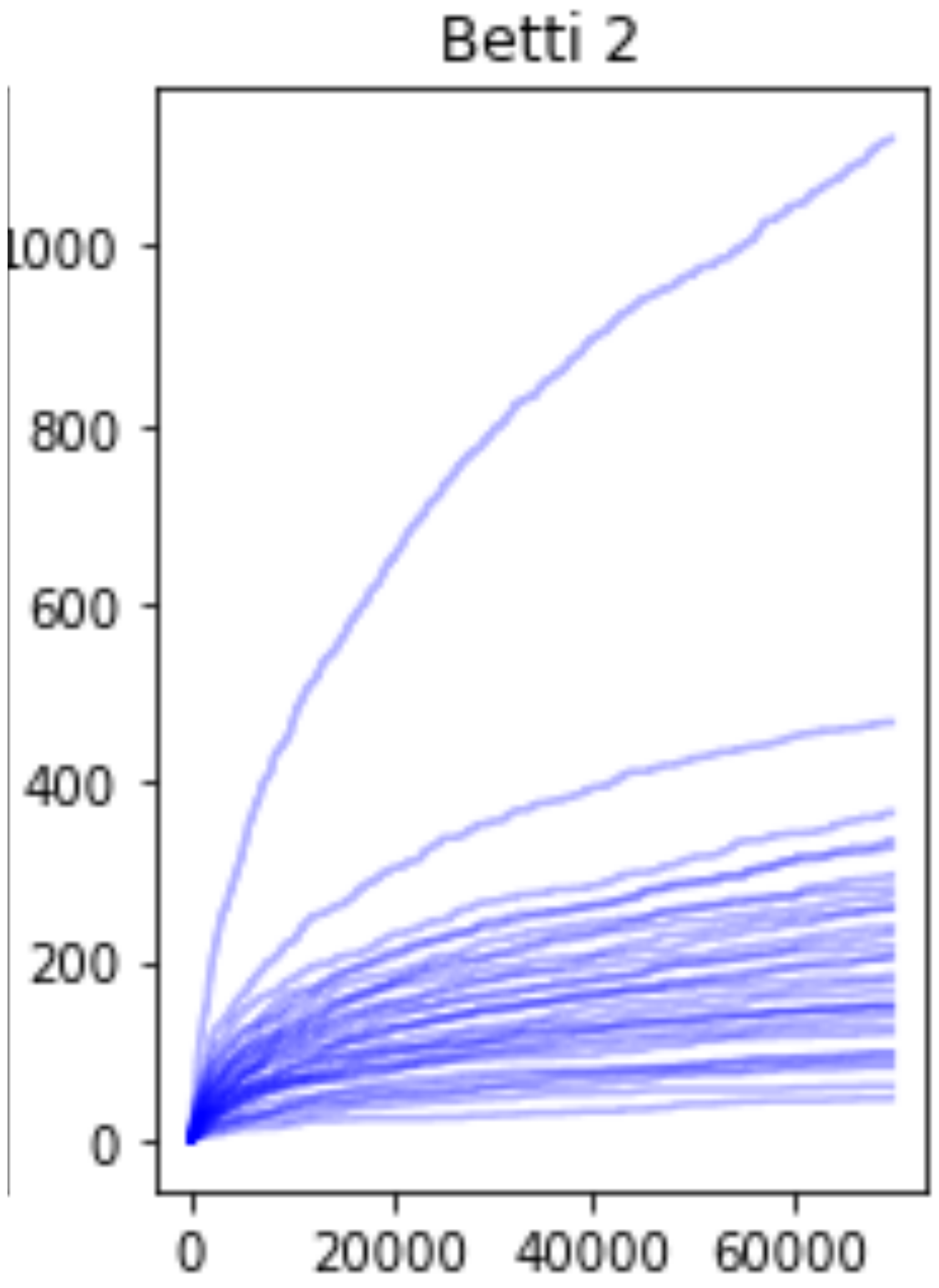
Expected Betti Number $E[\beta_q]$



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

Expected Betti Number $E[\beta_q]$

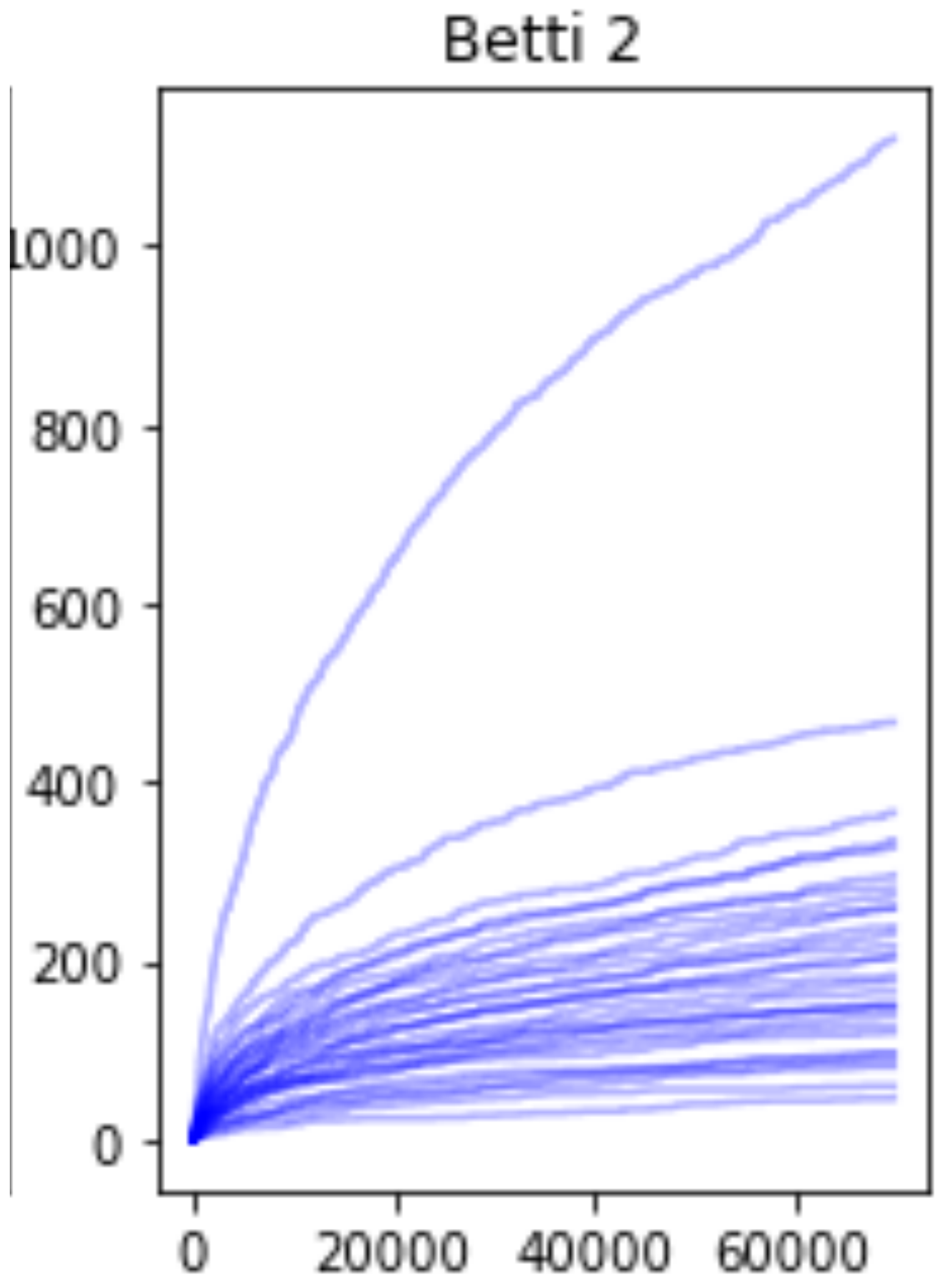
- increasing trend



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

Expected Betti Number $E[\beta_q]$

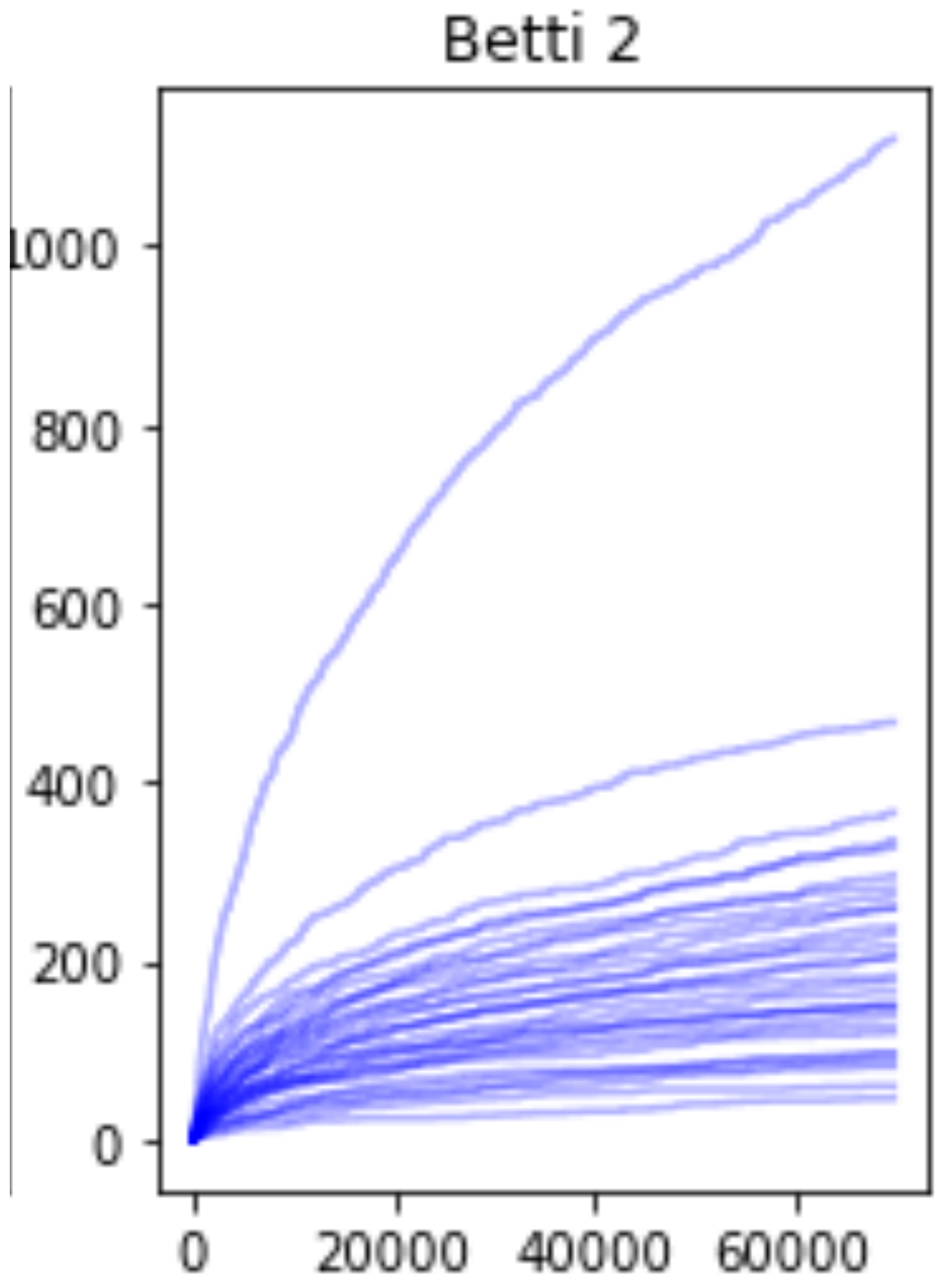
- increasing trend
- concave growth



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

Expected Betti Number $E[\beta_q]$

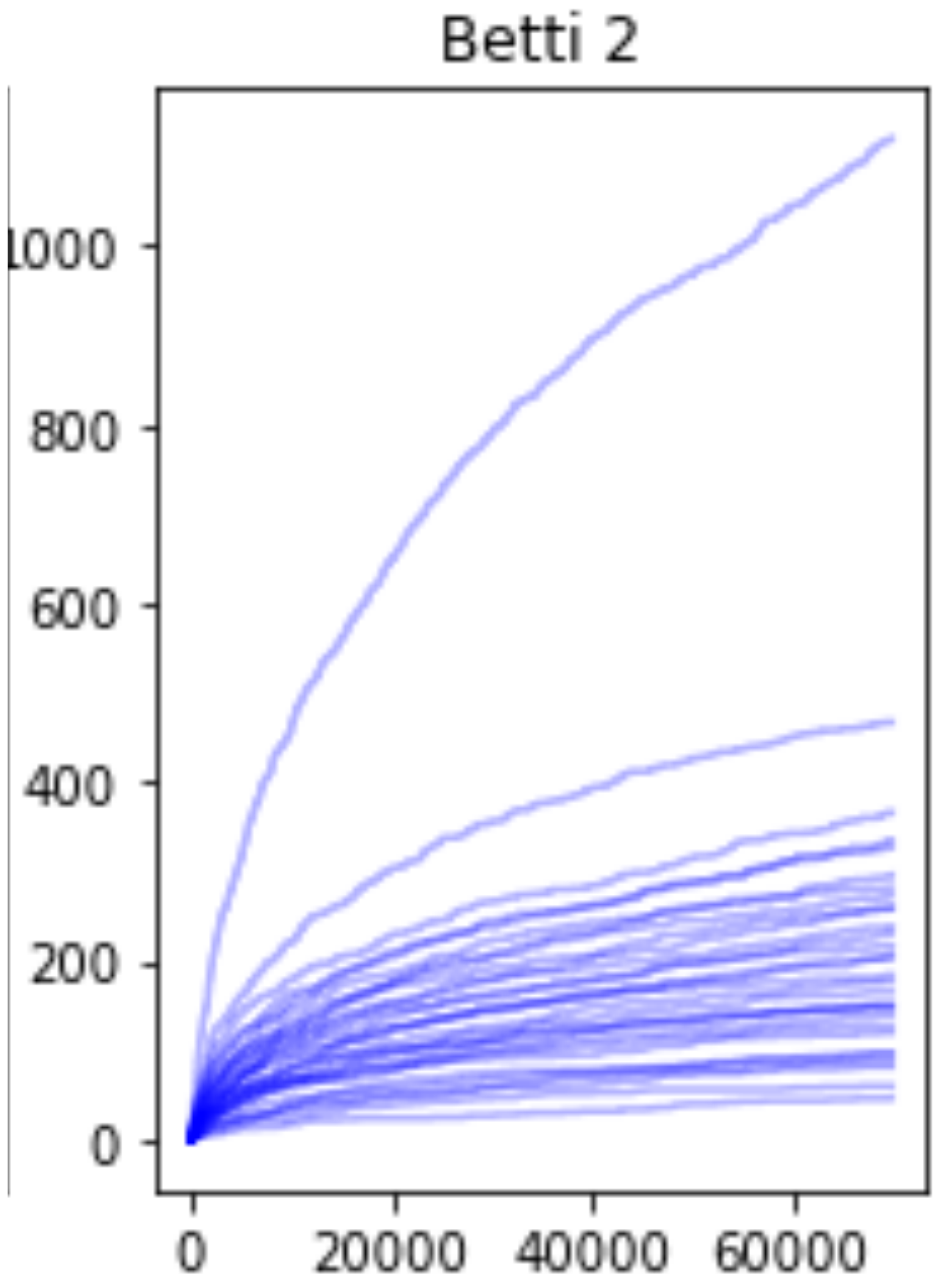
- increasing trend
- concave growth
- outlier



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

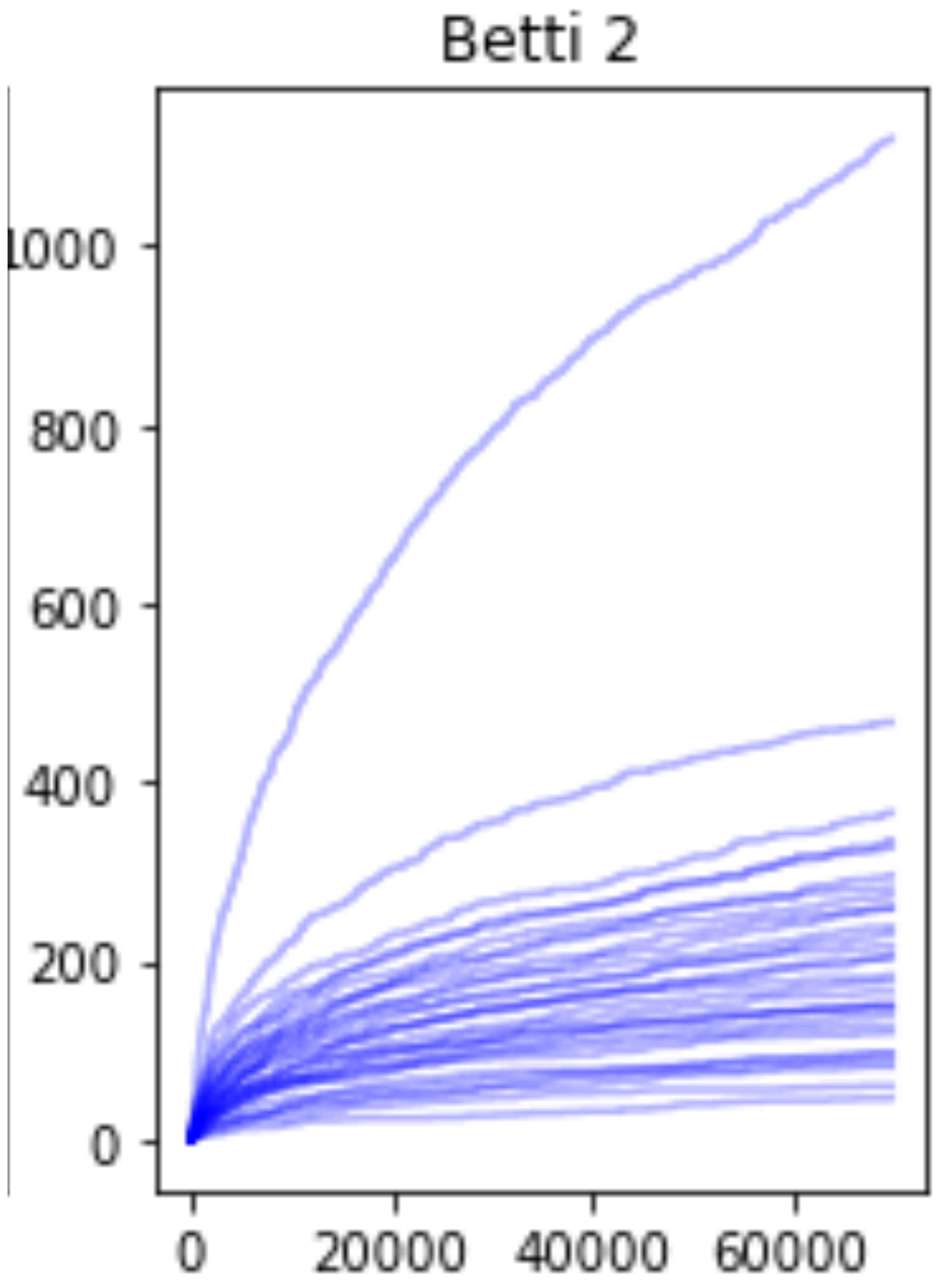
Expected Betti Number $E[\beta_q]$

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



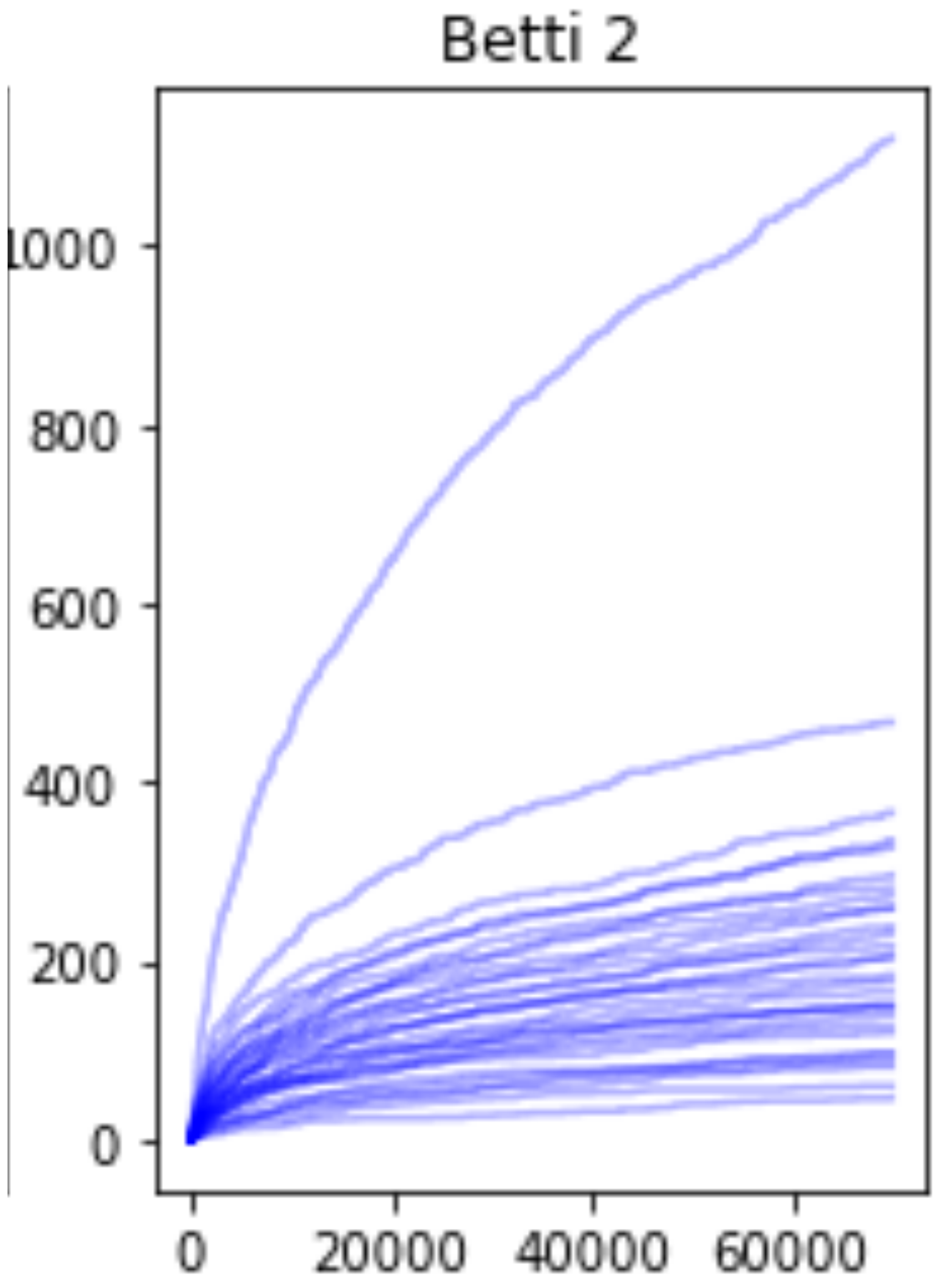
Expected Betti Number $E[\beta_q]$

- c(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 $E[\beta_2] \leq C$.



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 $q \geq 2$.

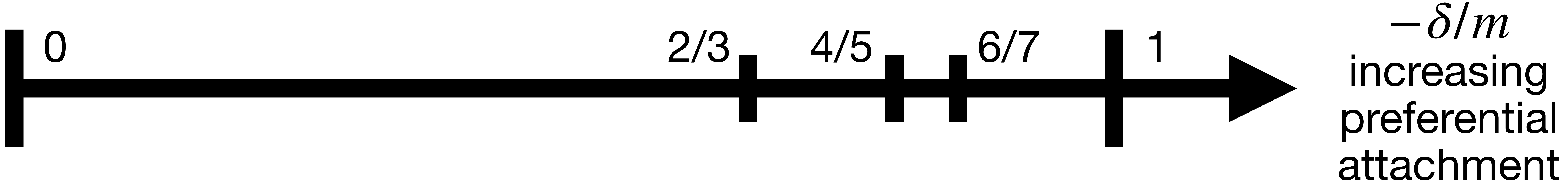


Phase transition

Recall

P(attaching to v) ∝ degree + δ

m = number of edges per new node

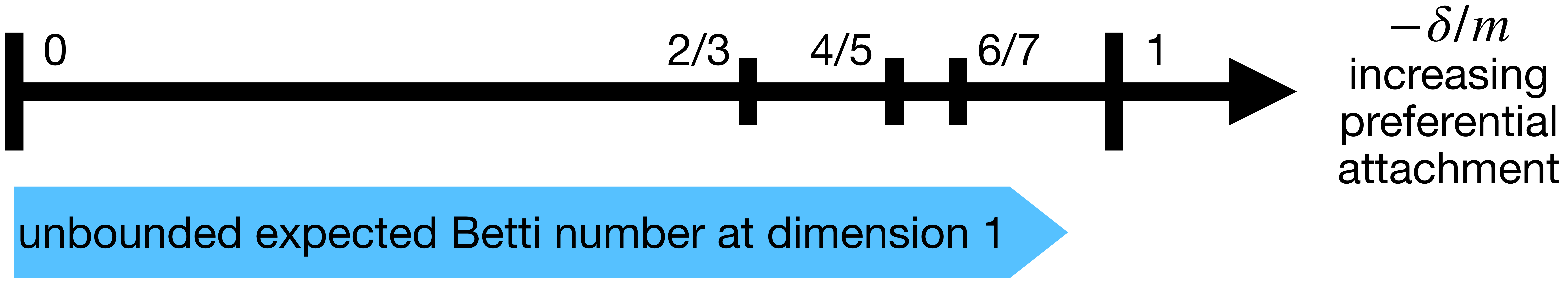


Phase transition

Recall

P(attaching to v) ∝ degree + δ

m = number of edges per new node

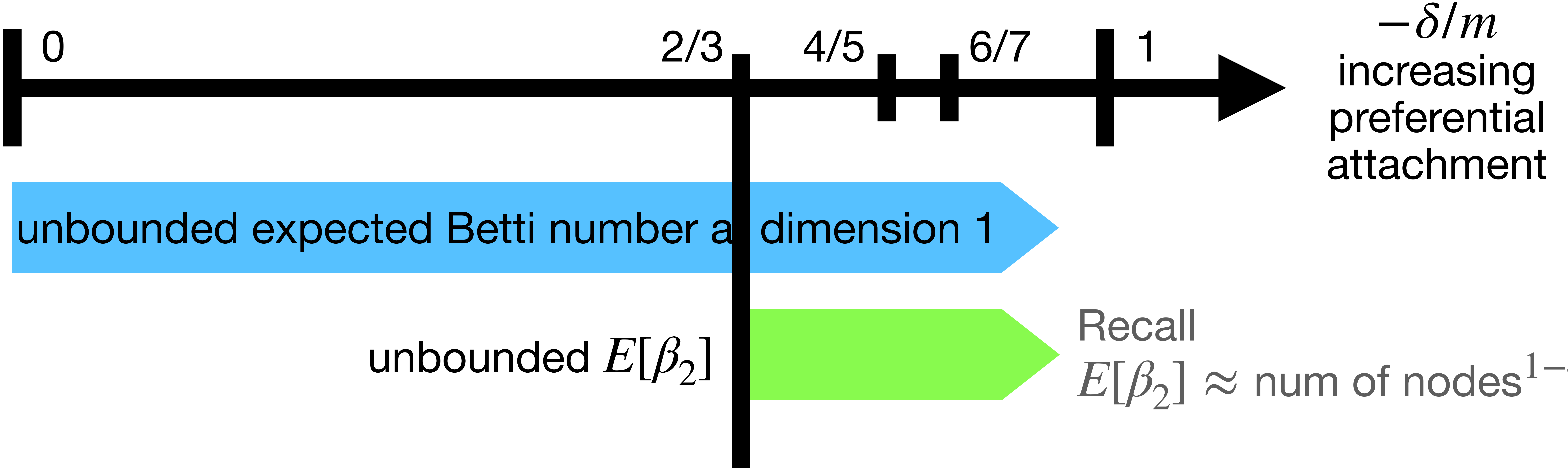


Phase transition

Recall

P(attaching to v) ∝ degree + δ

m = number of edges per new node



Recall

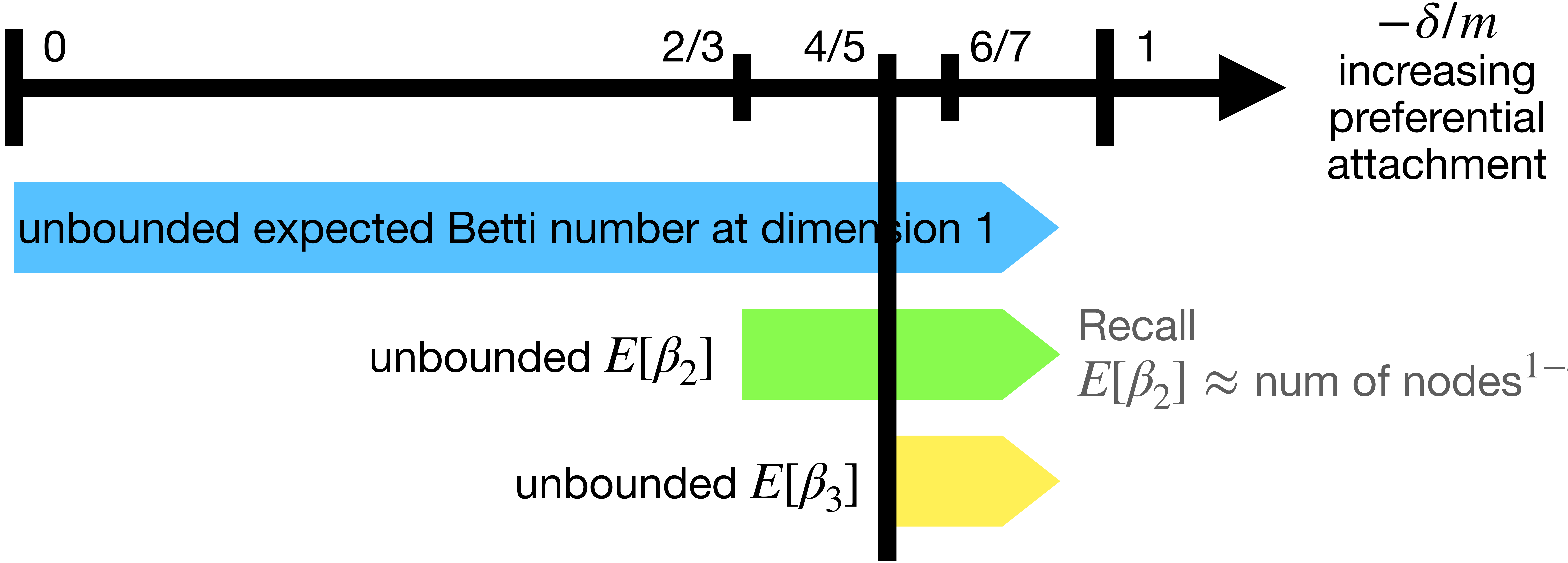
$E[\beta_2] \approx$ num of nodes^{1-4x}

Phase transition

Recall

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

m = number of edges per new node



Recall

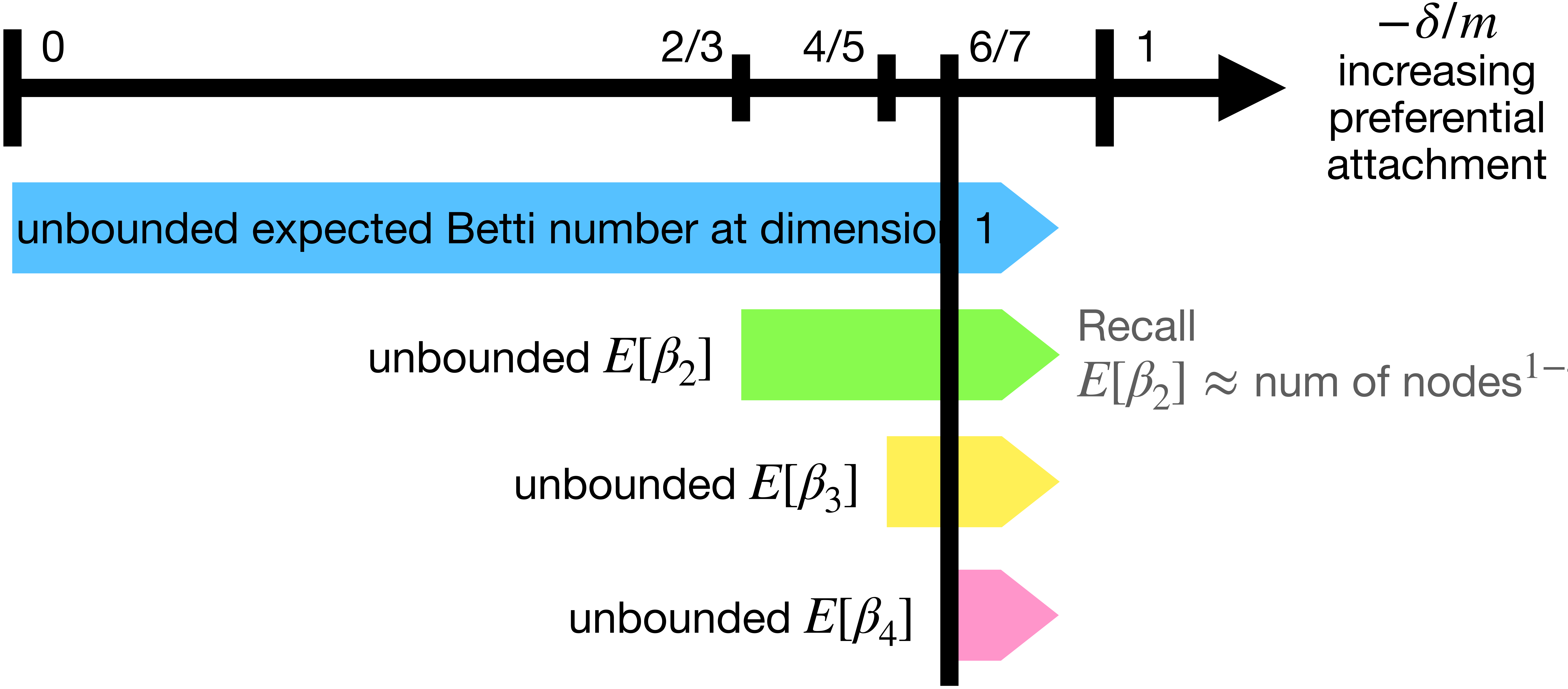
$E[\beta_2] \approx$ num of nodes^{1-4x}

Phase transition

Recall

P(attaching to v) ∝ degree + δ

m = number of edges per new node

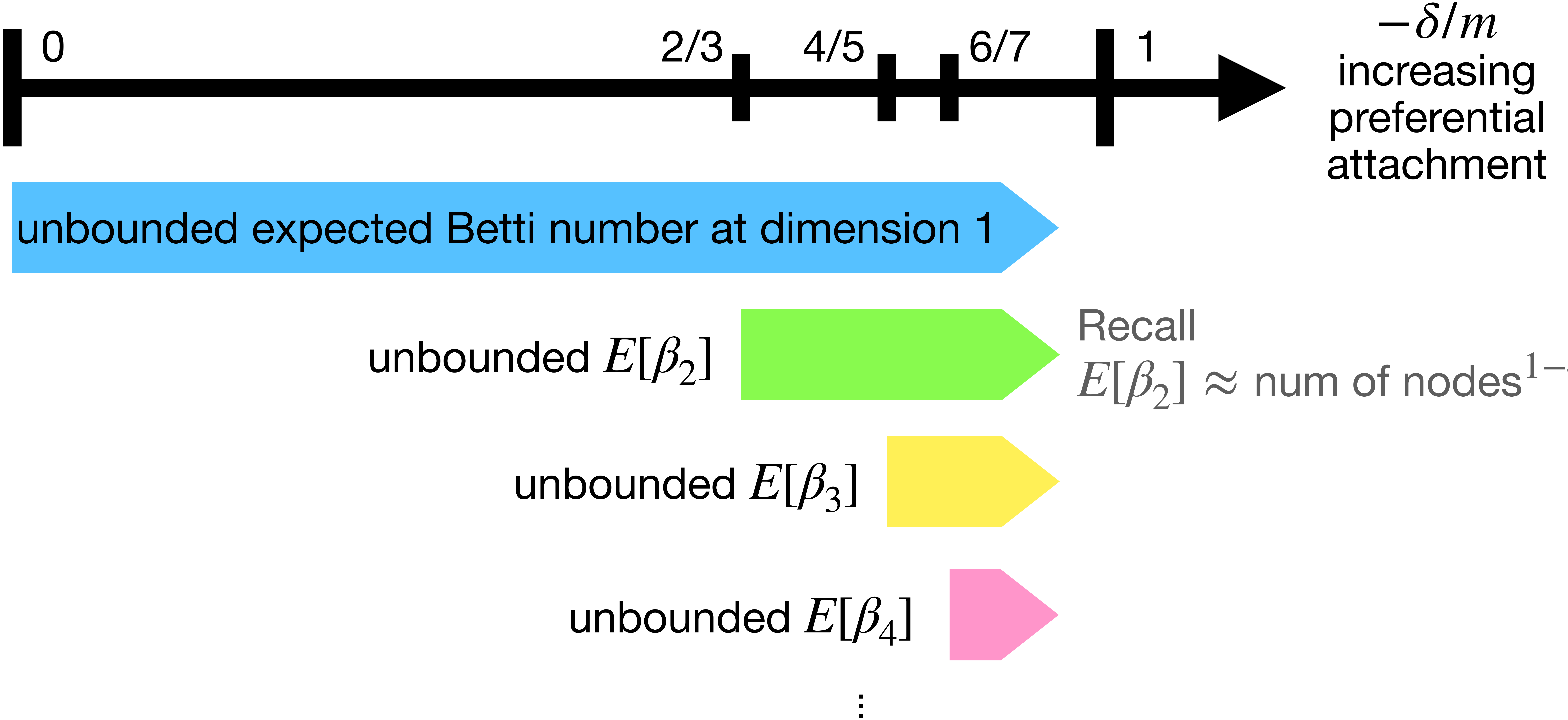


Recall

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

Phase transition

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

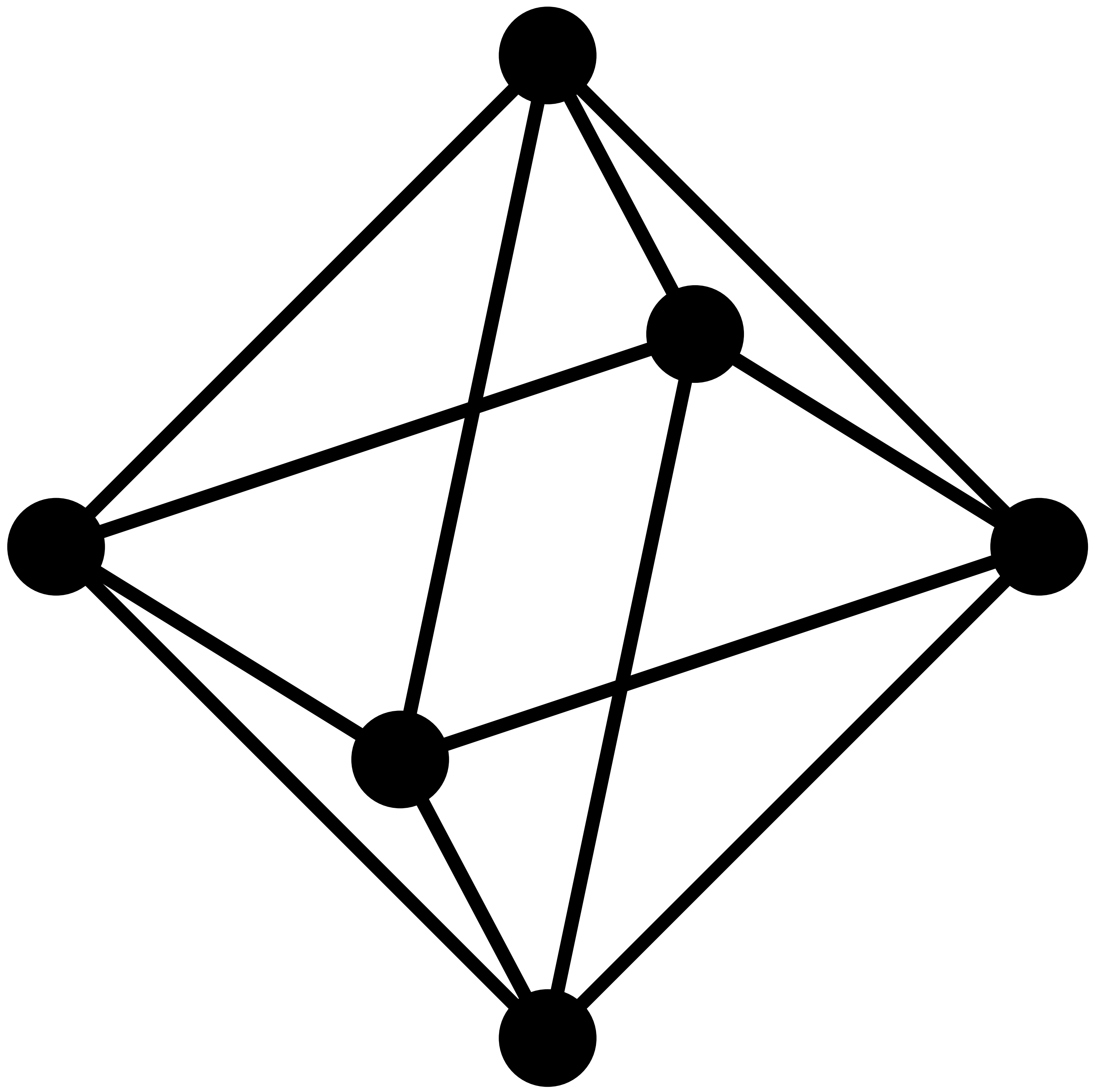


Recall
$E[\beta_2] \approx$ num of nodes$^{1-4\chi}$

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

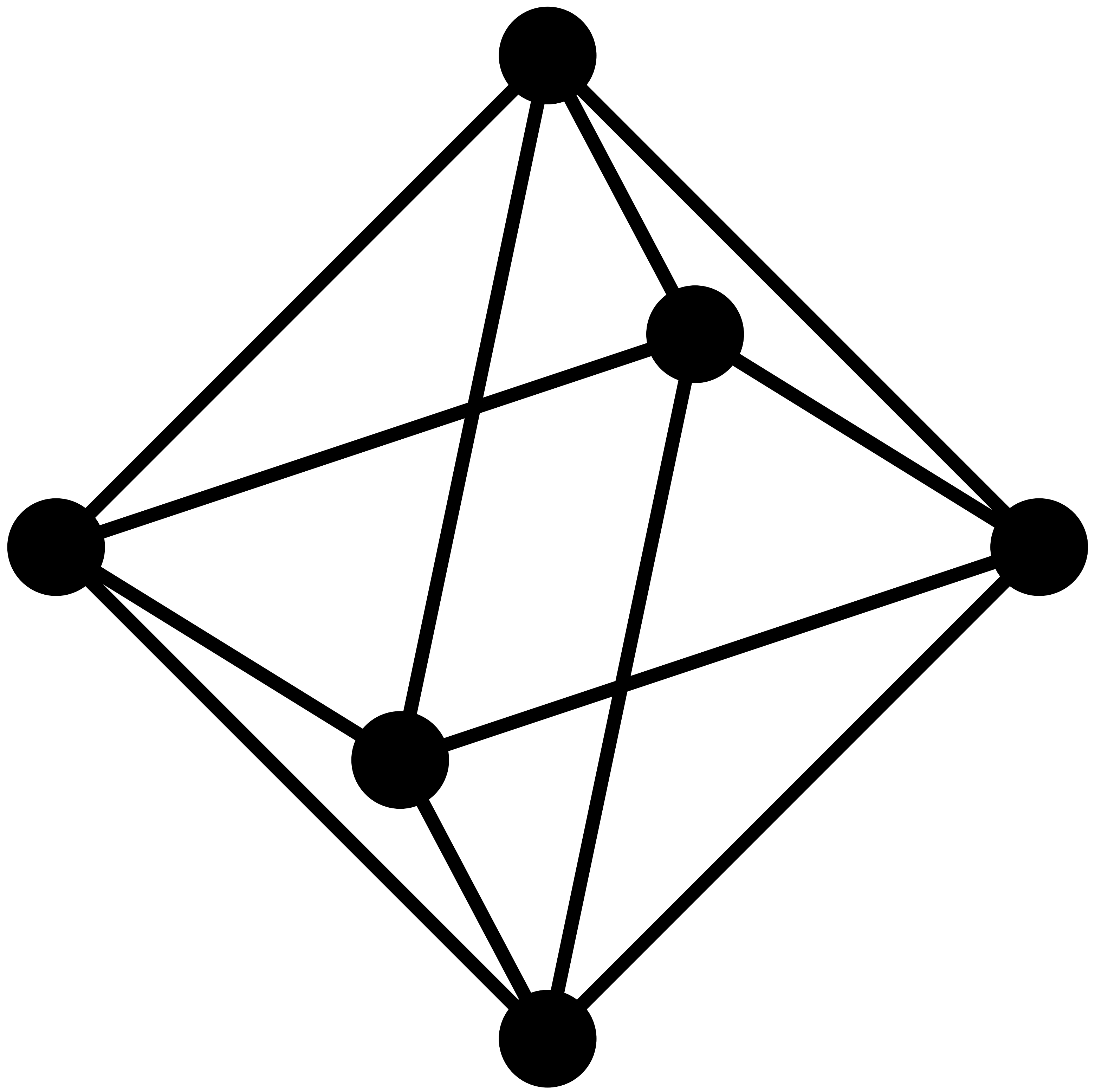
Proof?

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

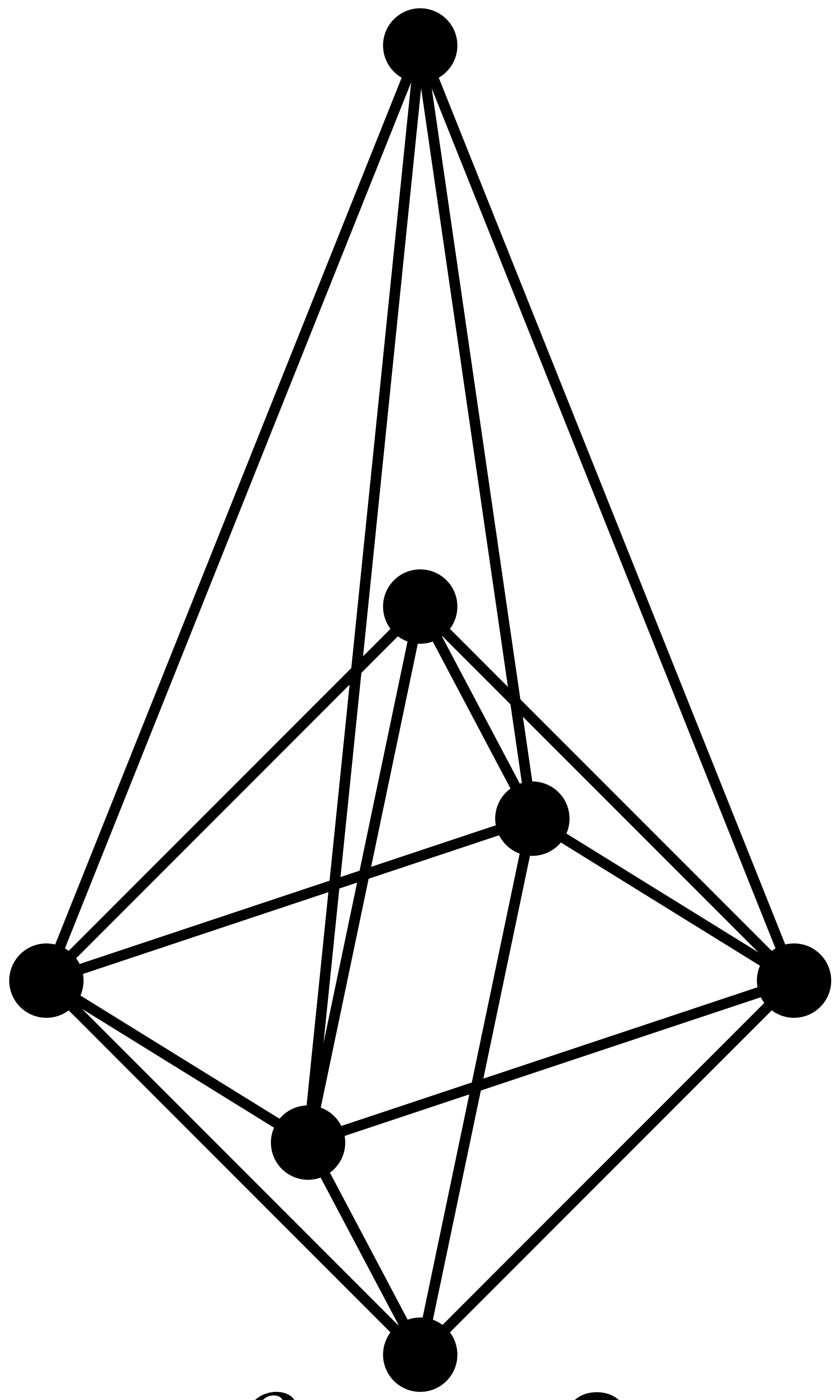


$\beta_2 = 1$

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

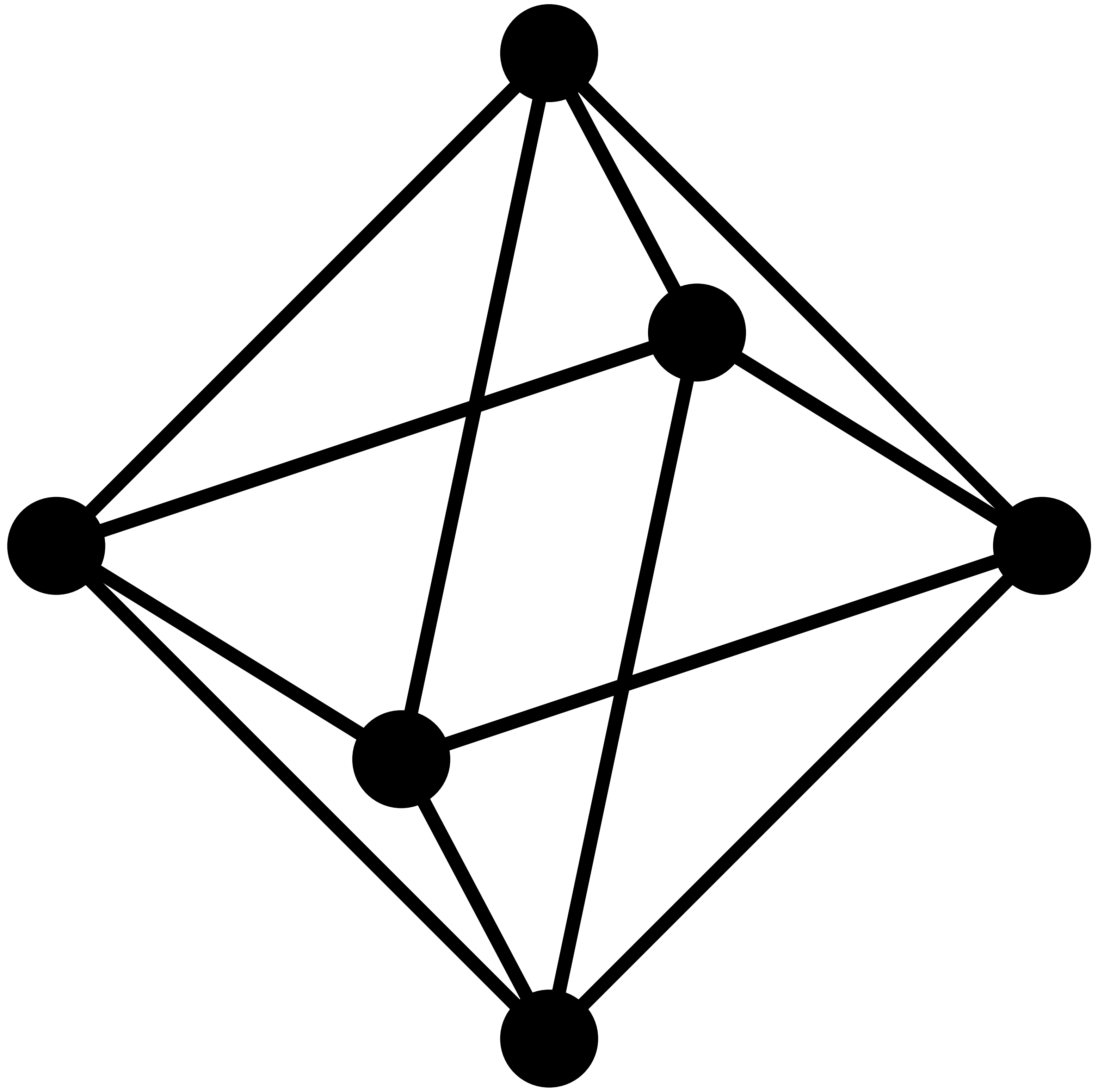


$\beta_2 = 1$

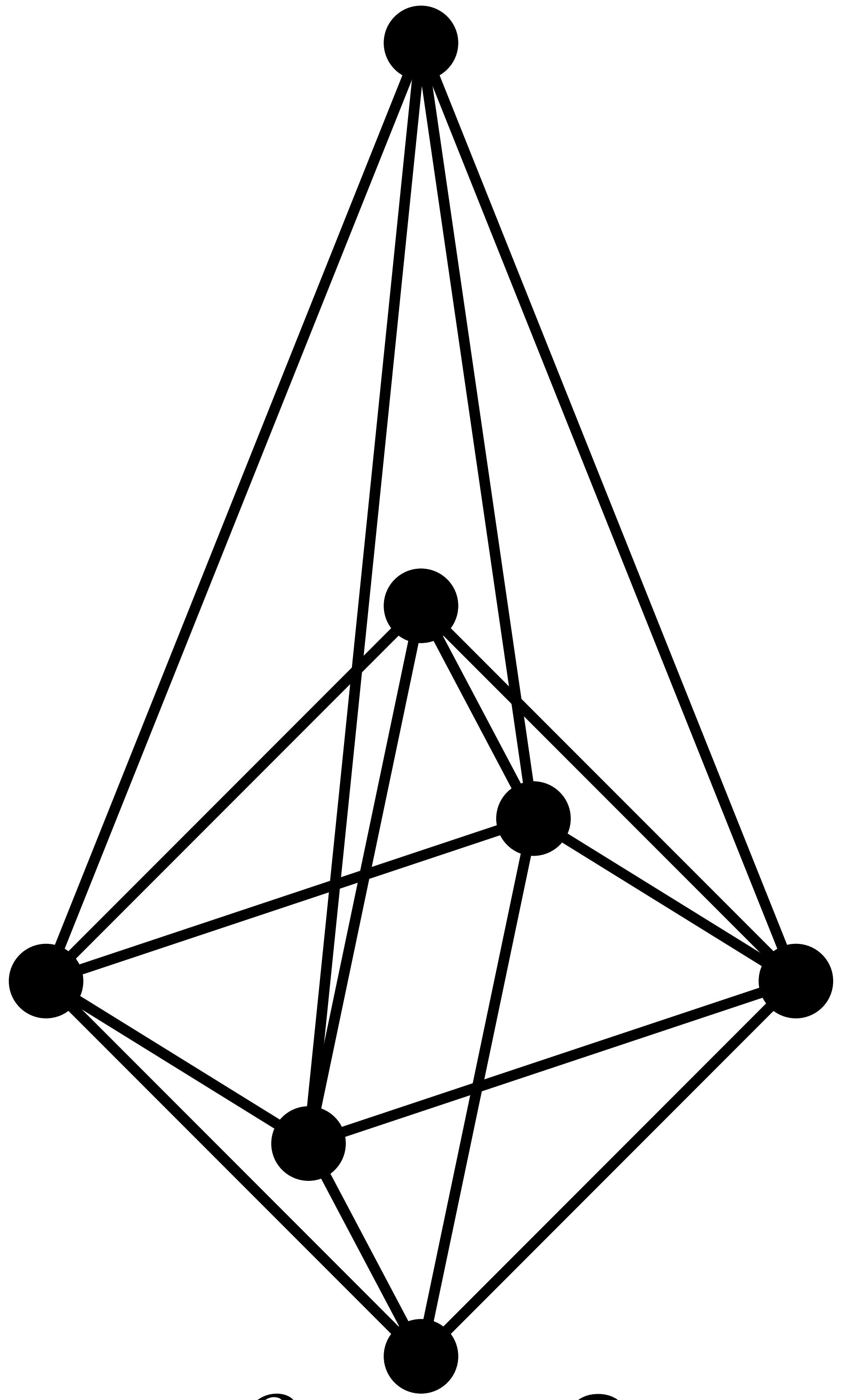


$\beta_2 = 2$

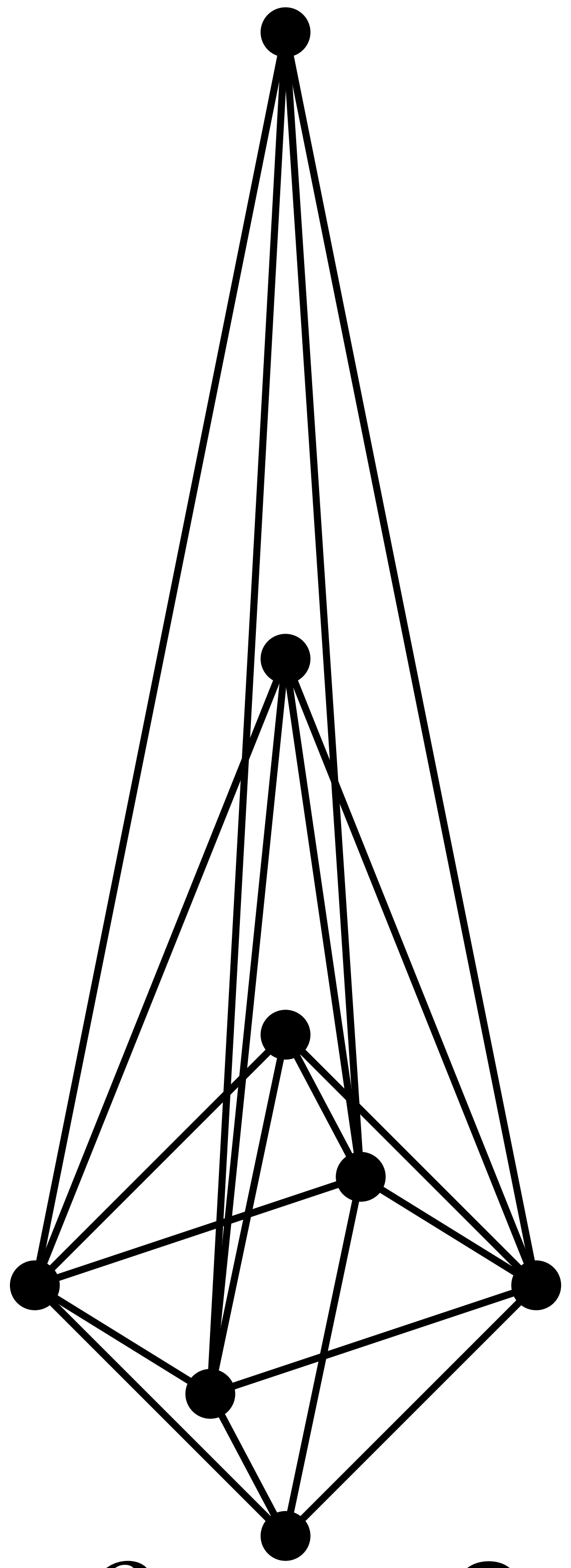
Proof of $E[\beta_2] \approx$ num of nodes^{1-4x}



$\beta_2 = 1$



$\beta_2 = 2$



$\beta_2 = 3$

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

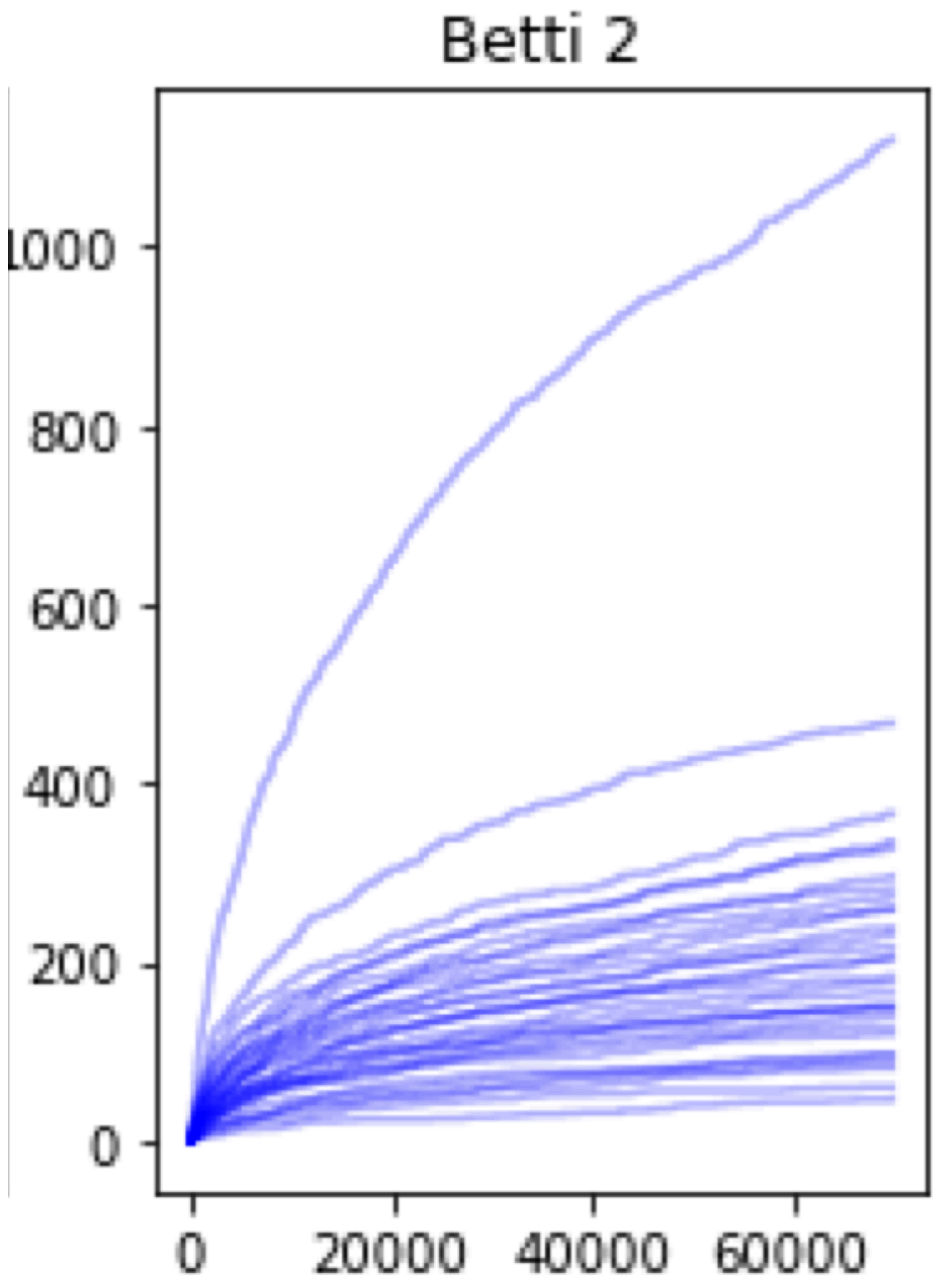
Subtleties

- 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}

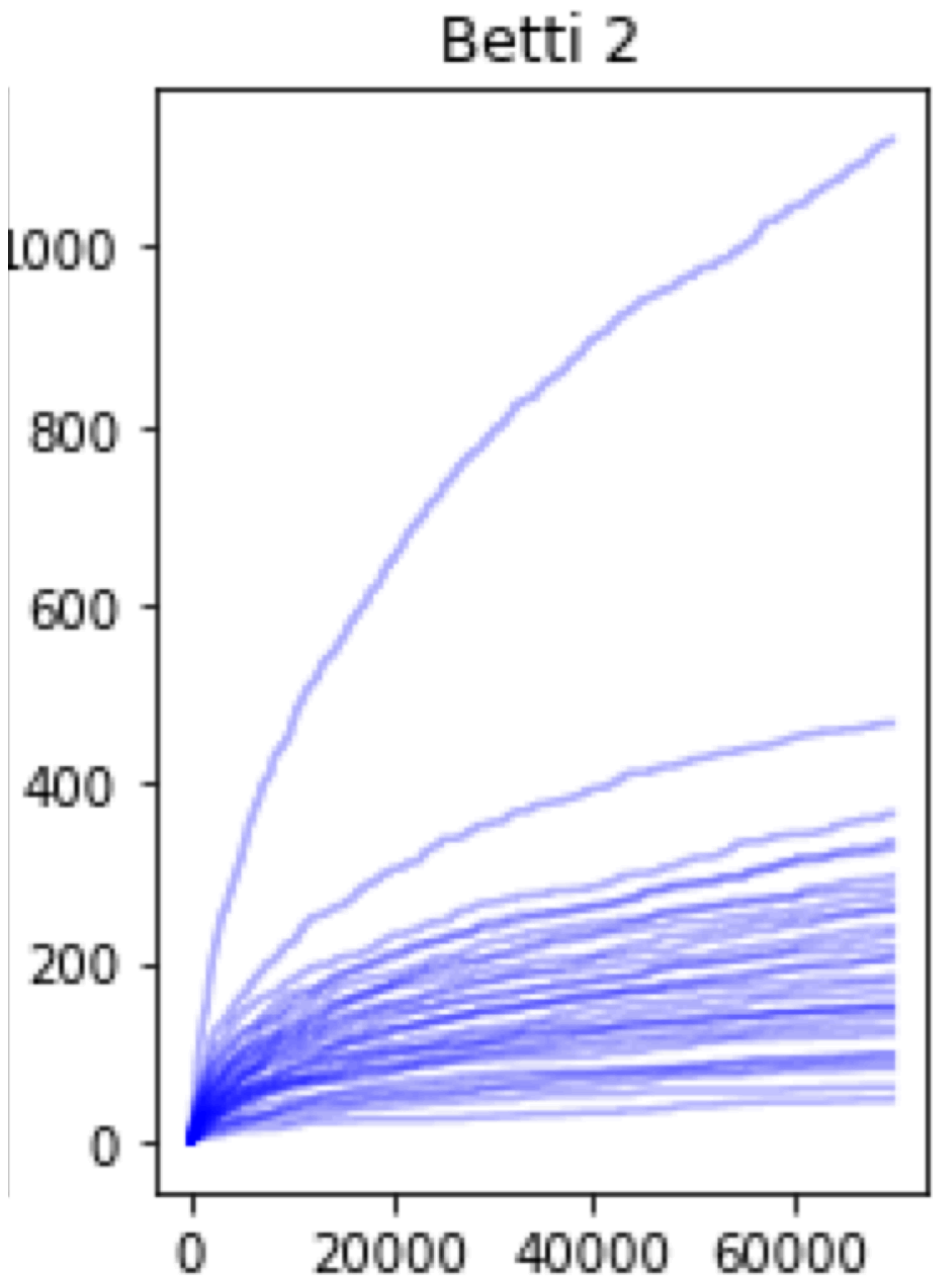
In practice???

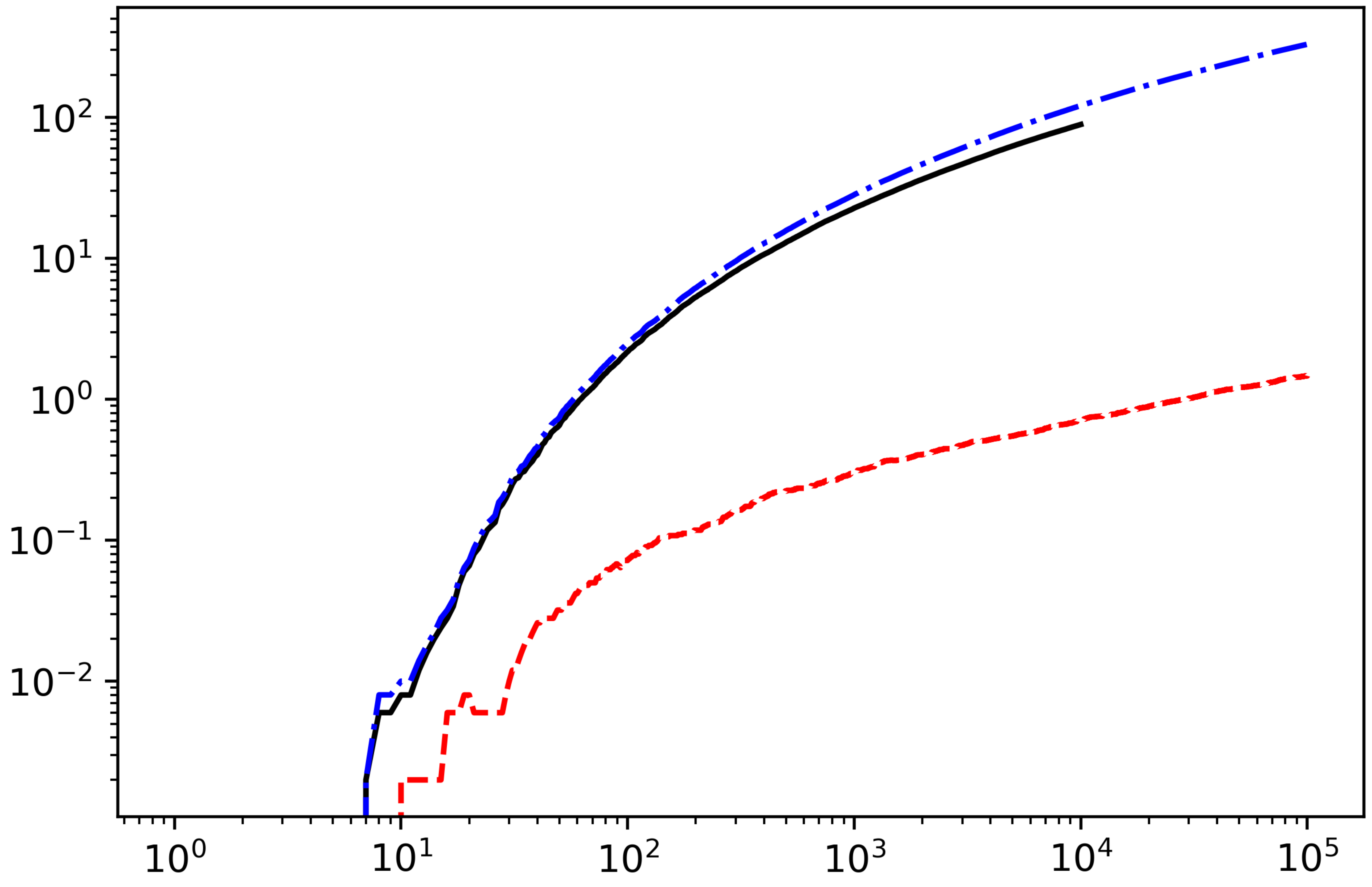
$E[\beta_2] \approx$ **num of nodes**$^{1-4x}$



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

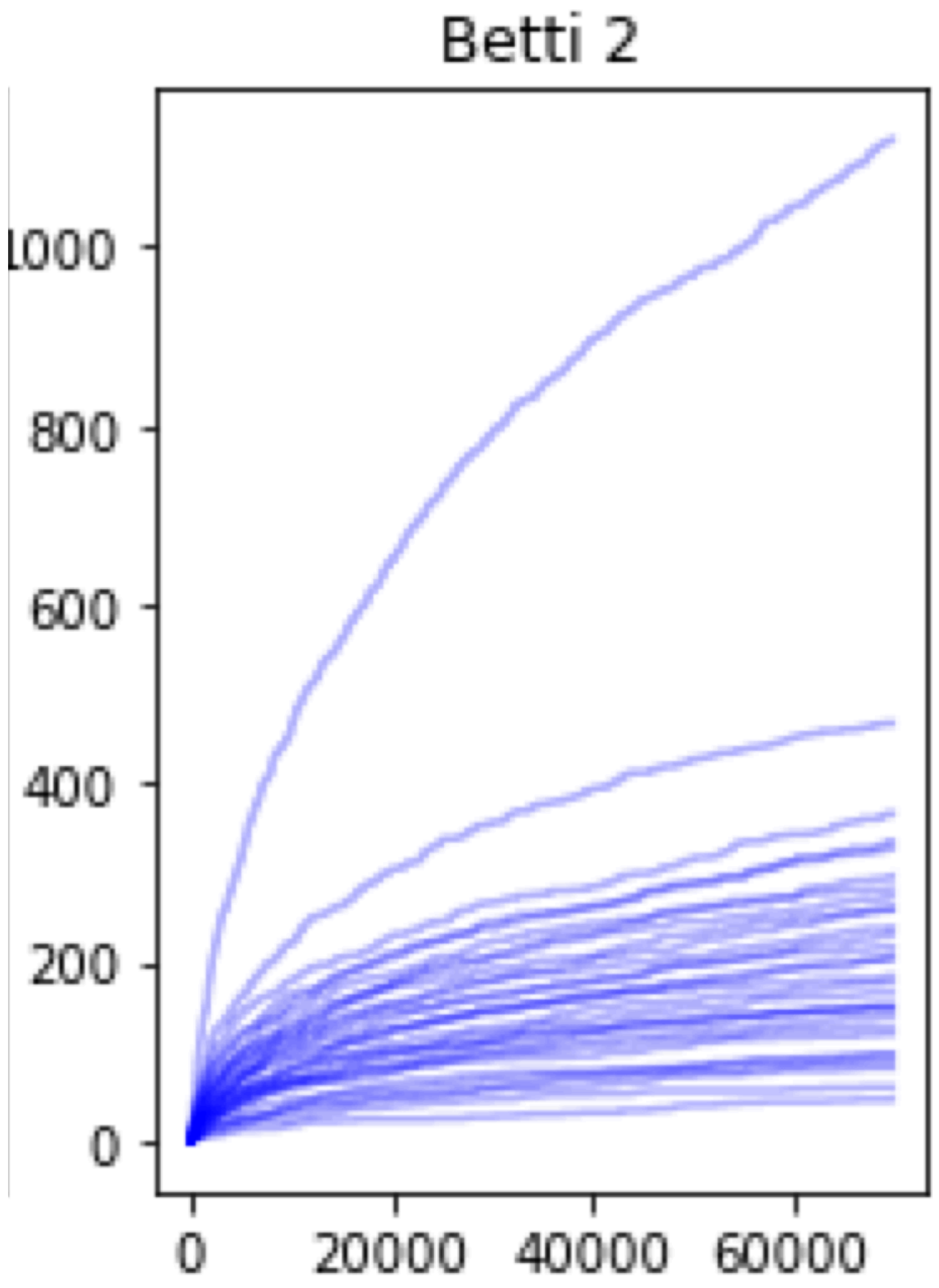
$$\log E[\beta_2] \approx (1-4x)\log(\textbf{num of nodes})$$

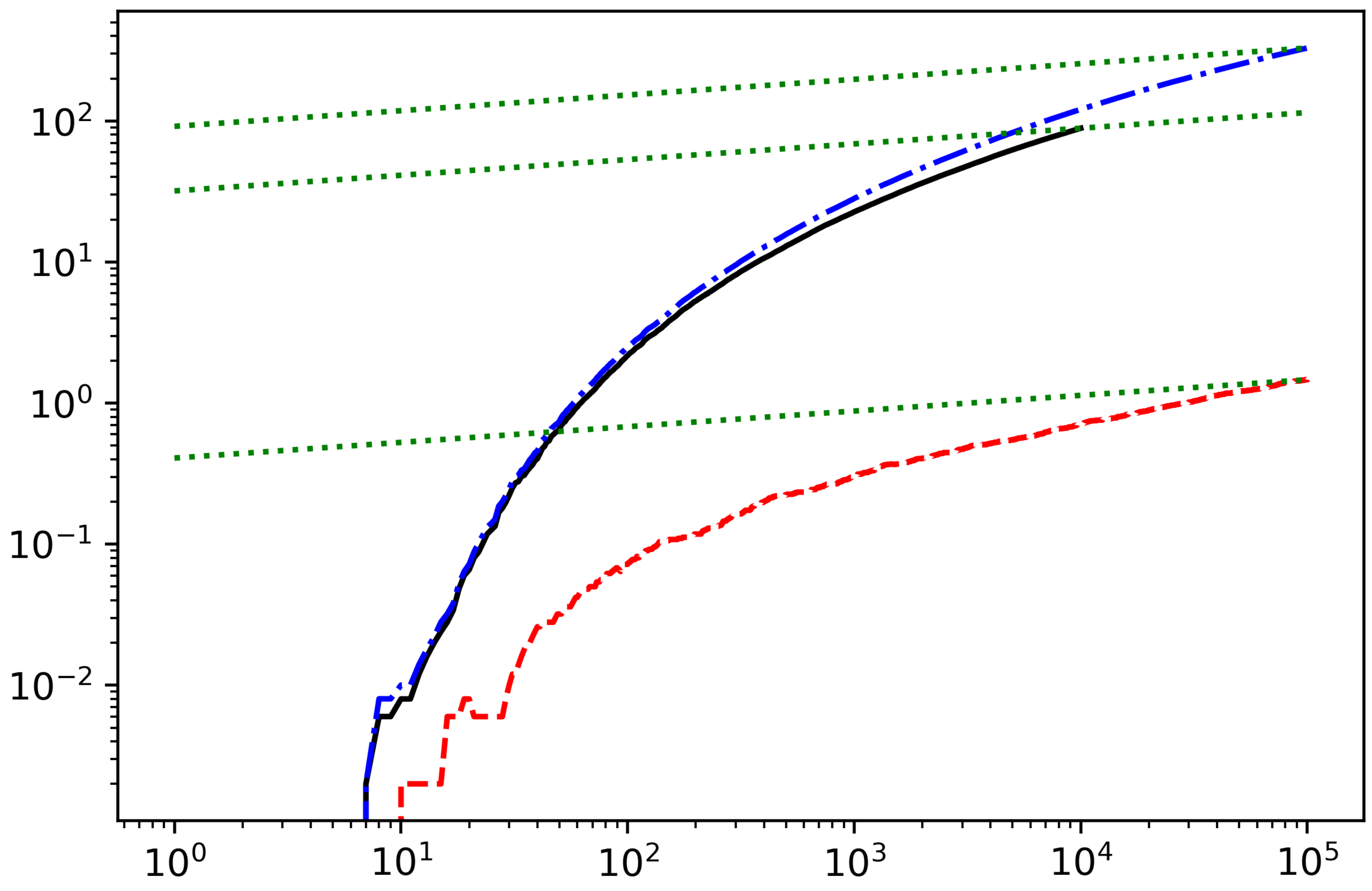




$$E[\beta_2] \approx \textbf{num of nodes}^{1-4x}$$

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

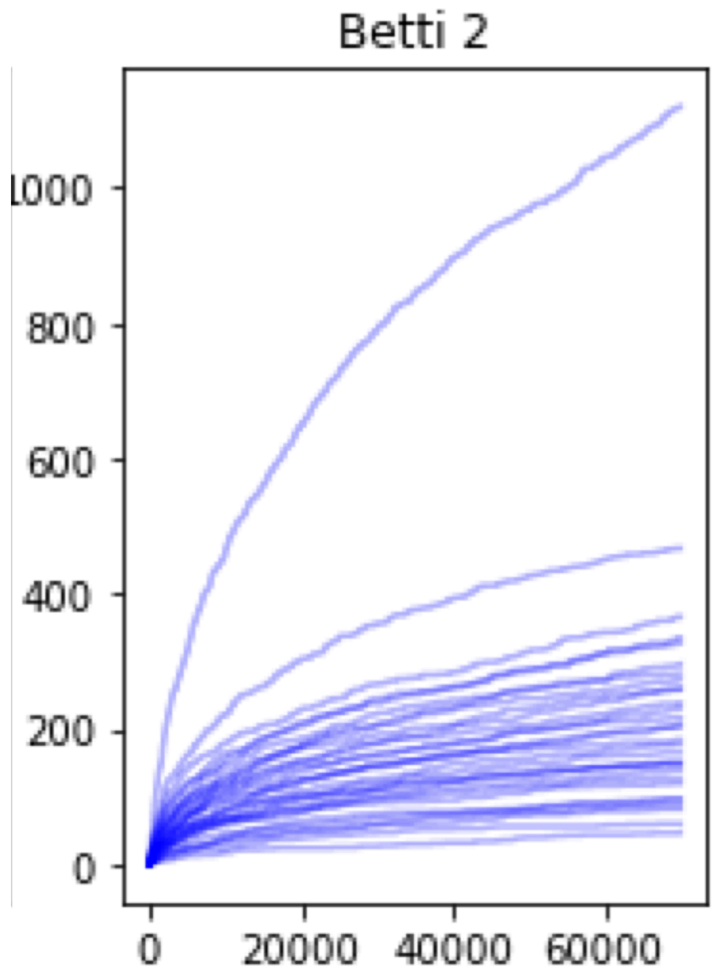


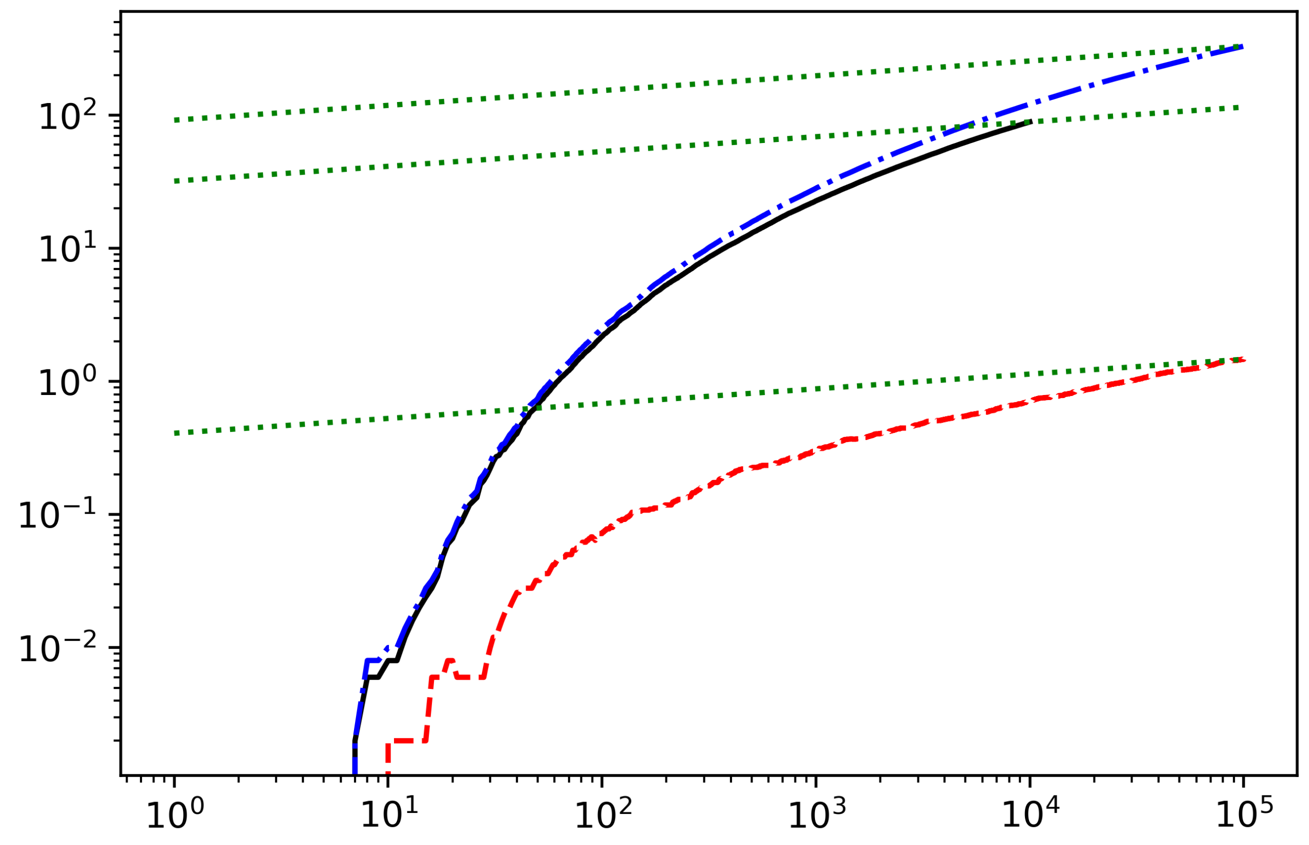


IV. Computation

Computational Challenges

- Ripser
- large graphs (1e4 ~ 1e5 nodes)
- large number of graphs (500 graphs)





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

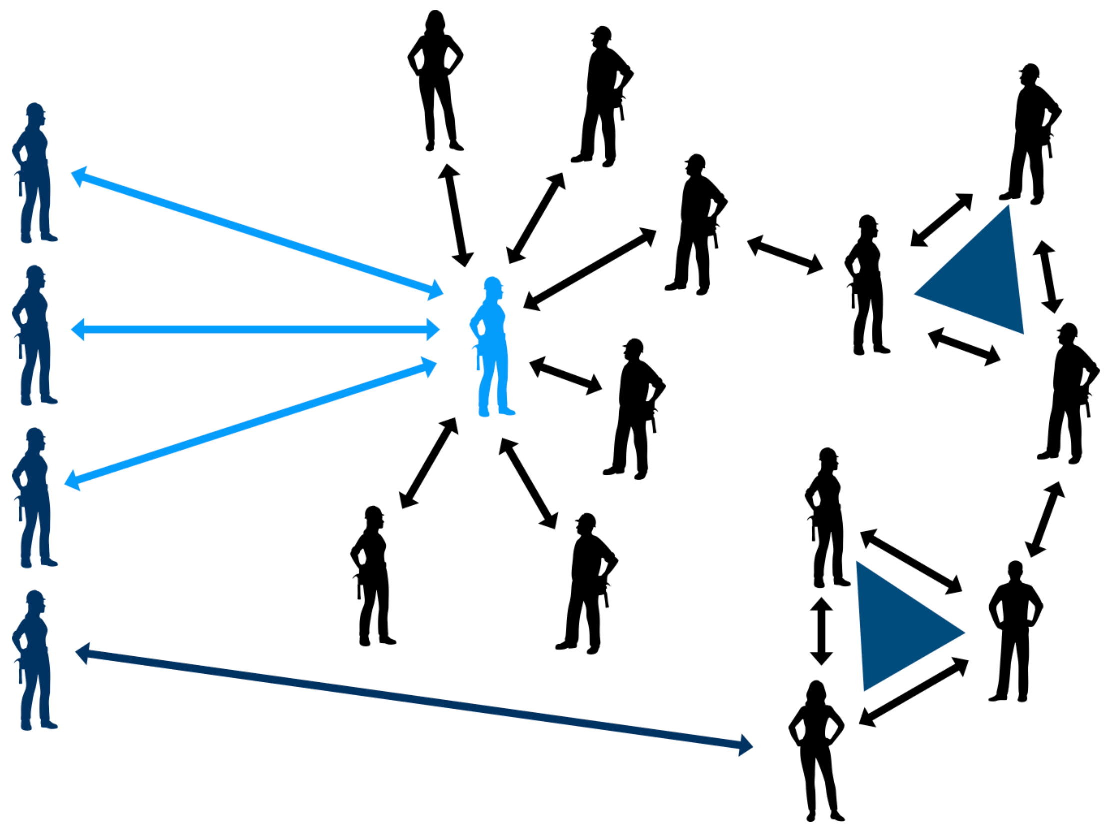
simplicial preferential attachment?

other non-homogeneous complexes?

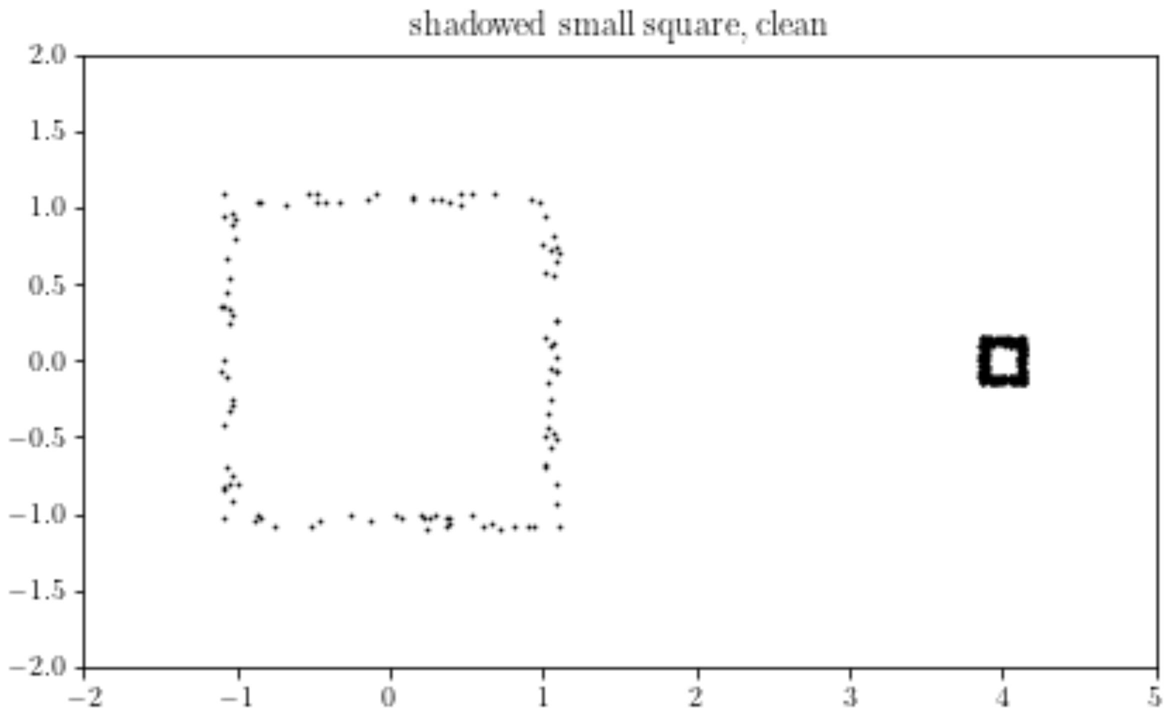
What did we learn today?

- 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



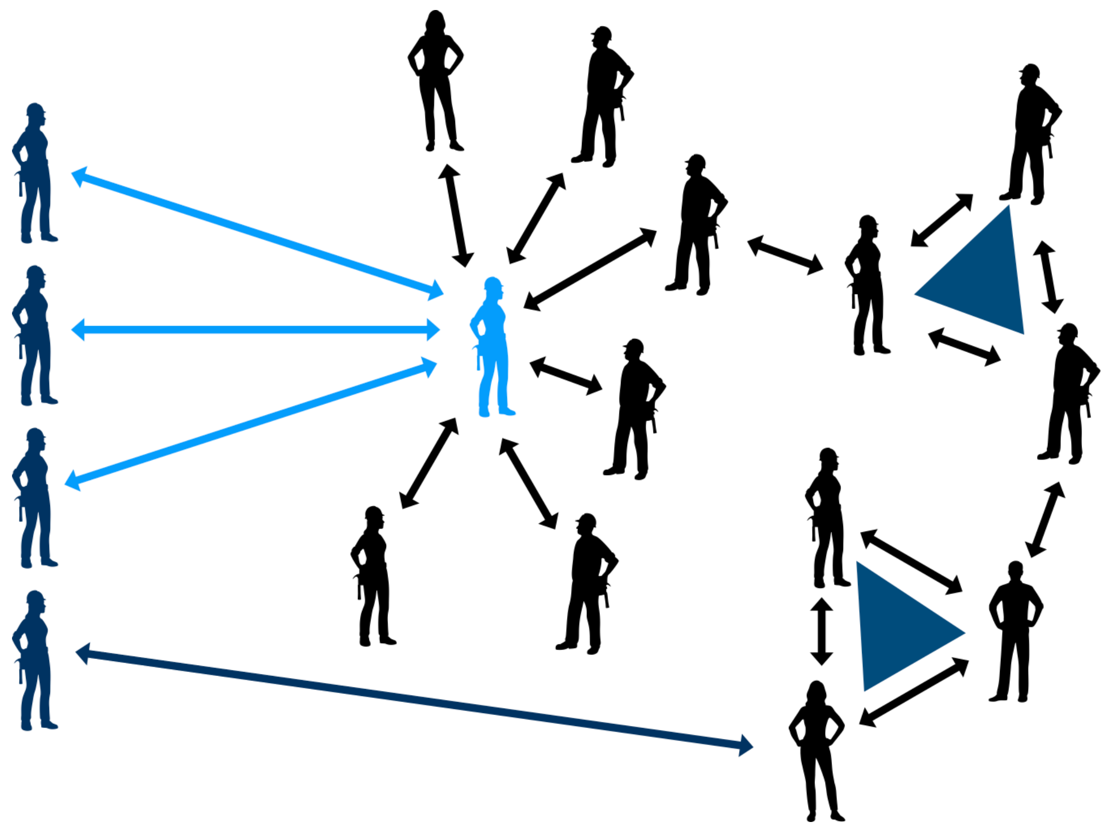
arxiv paper



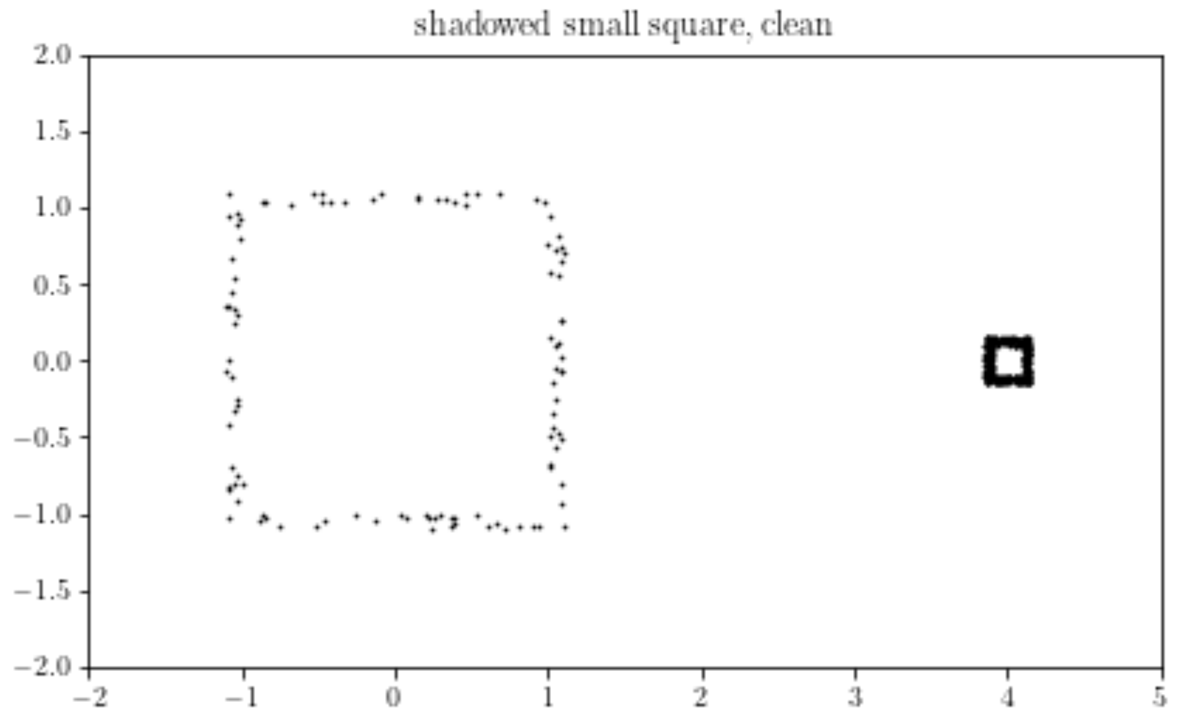
my video about small holes

Thank you!

Chunyin Siu cs2323@cornell.edu
Cornell University



arxiv paper



my video about small holes