Random Models of 21st Century Networks and Their Connectivity Structure

Bruce Reed

COPPE, UFRJ

Rio de Janeiro

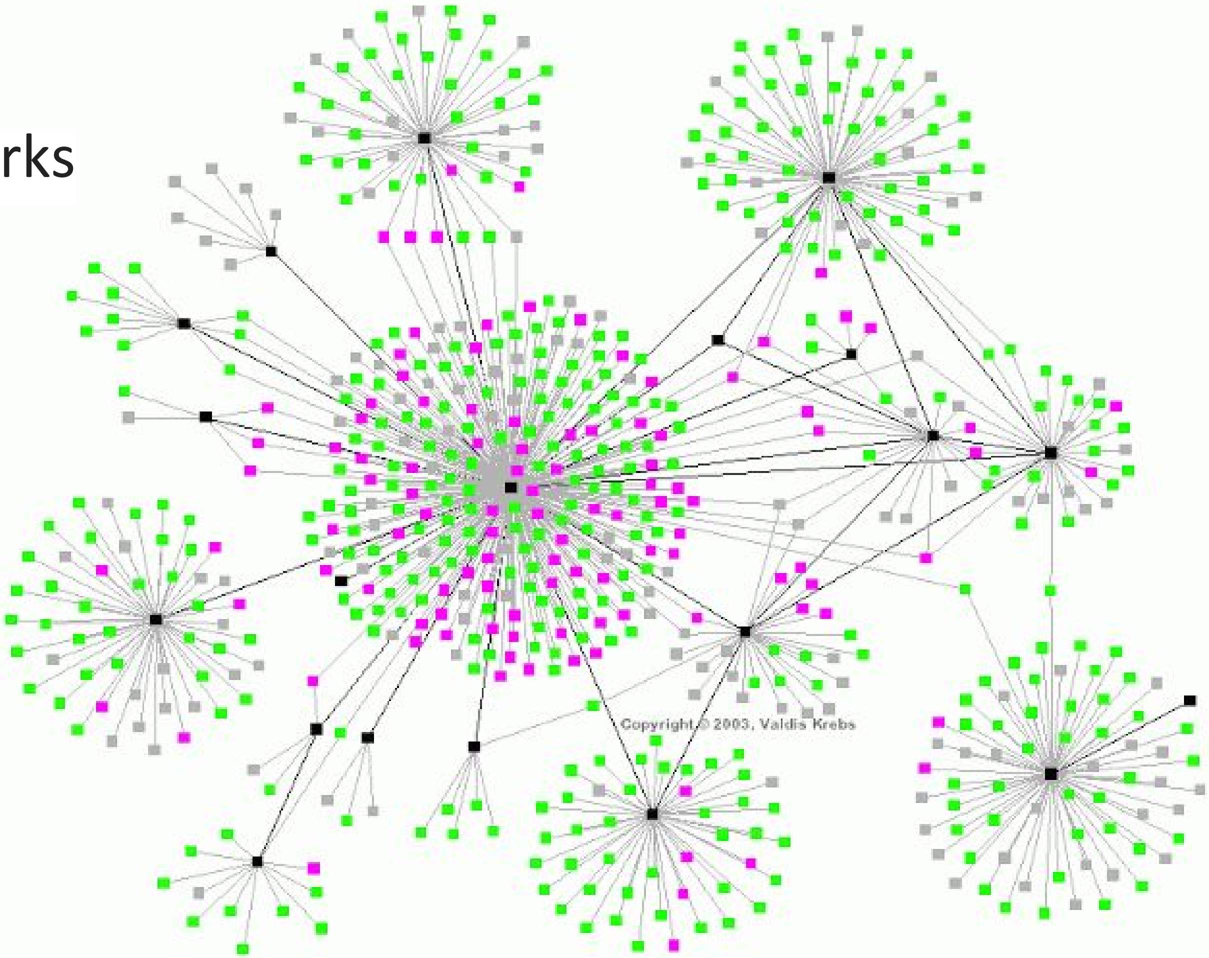
March 15, 2017

Random Networks as Controls

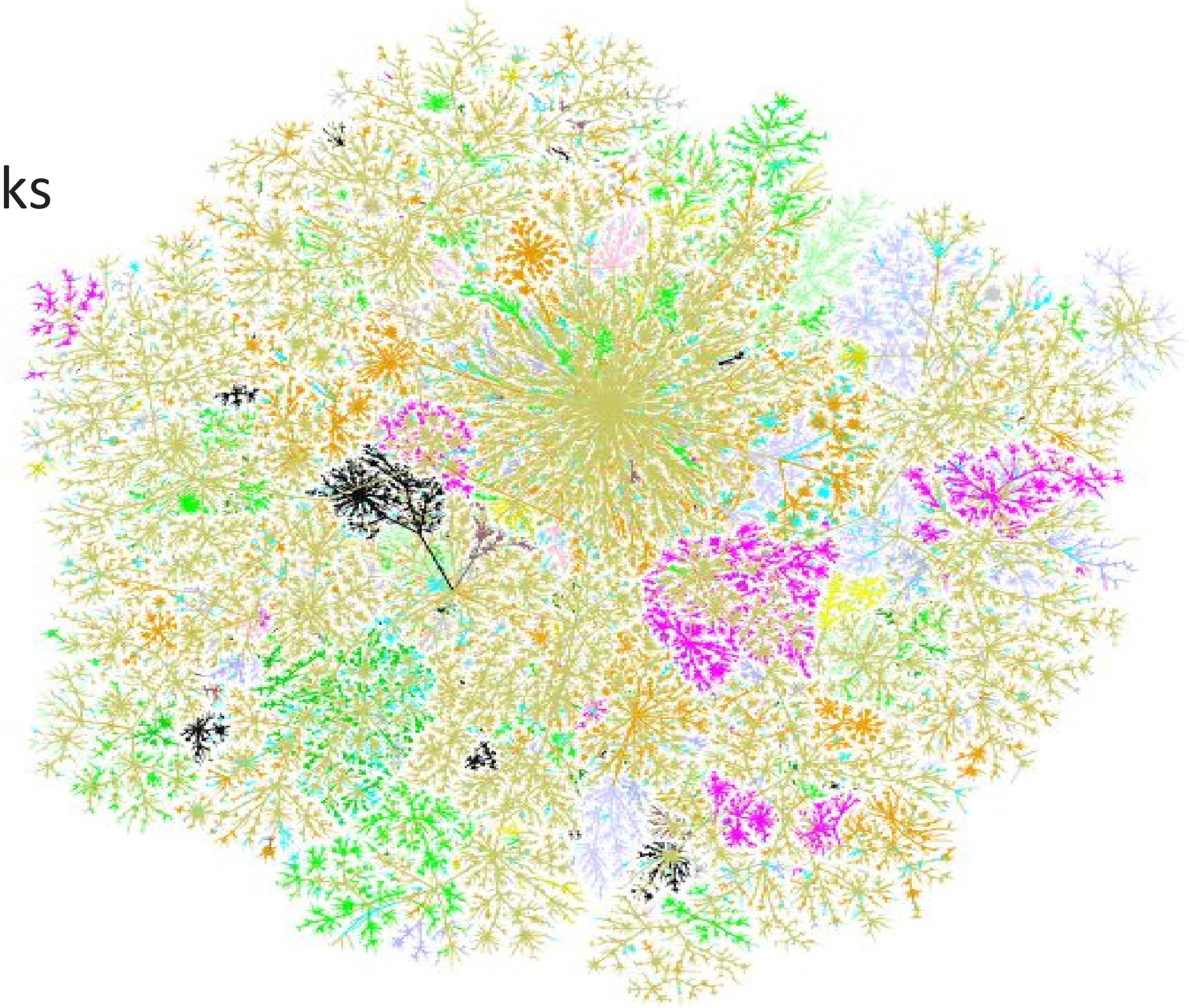
A common technique to analyze the properties of a single network is to use statistical randomization methods to create a reference network which is used for comparison purposes.

Mondragon and Zhou, 2012.

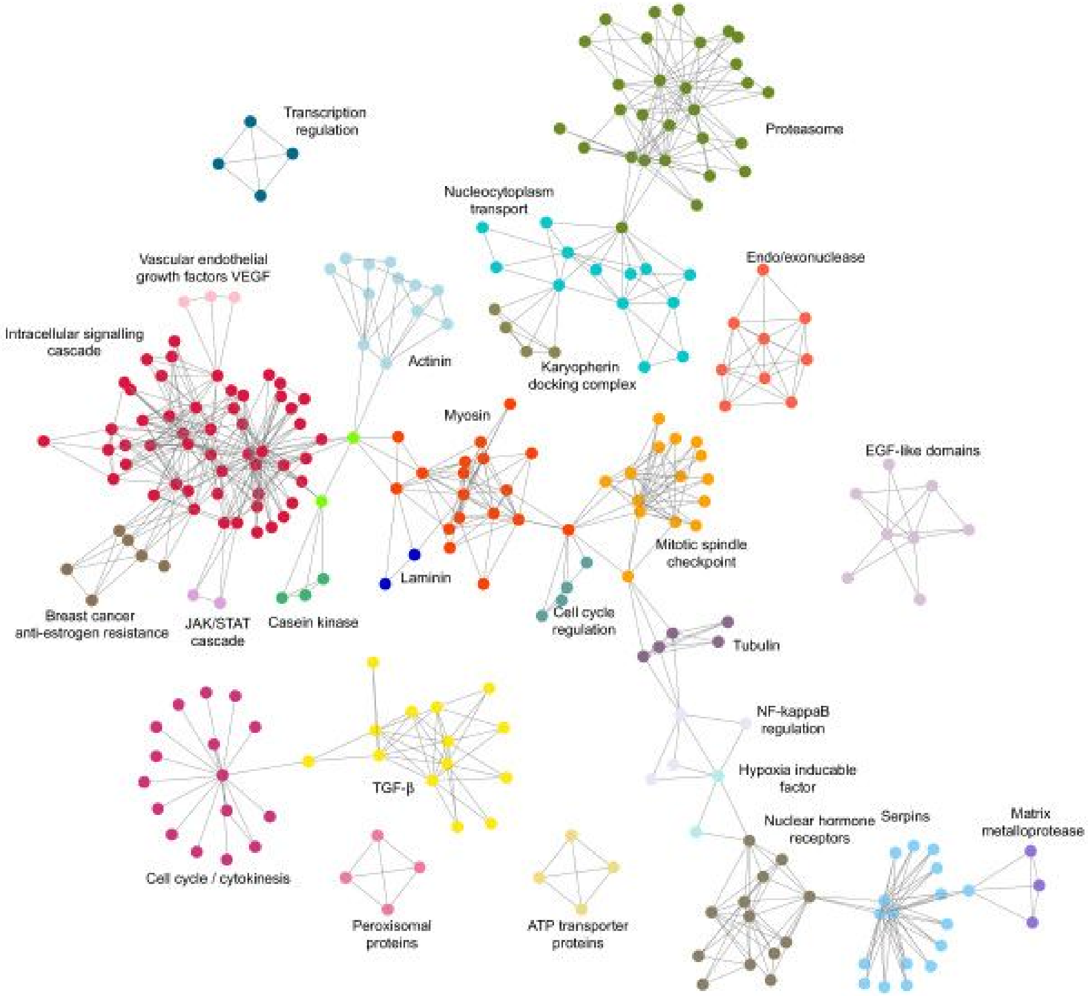
Looking for Clusters I: Epidemiological Networks



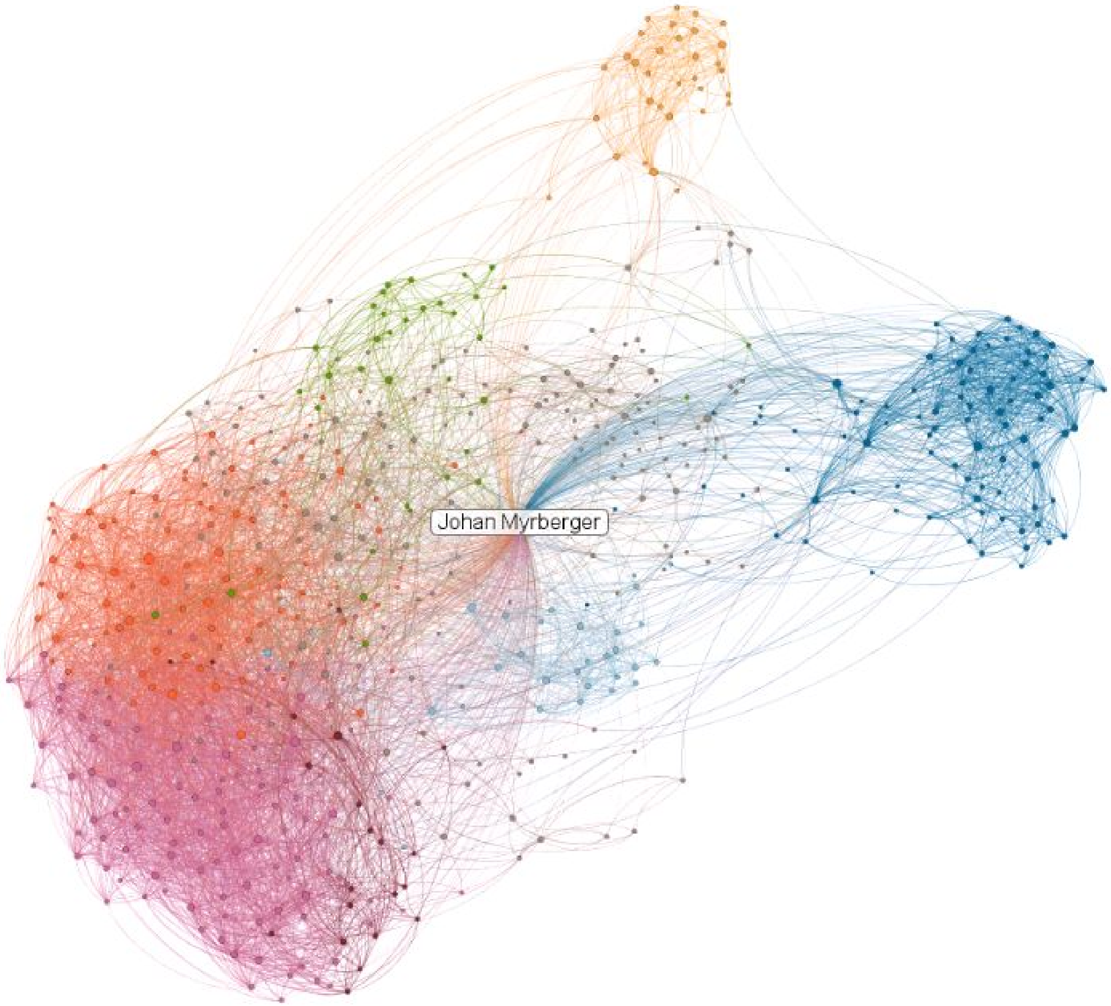
Looking for Clusters II: Communication Networks



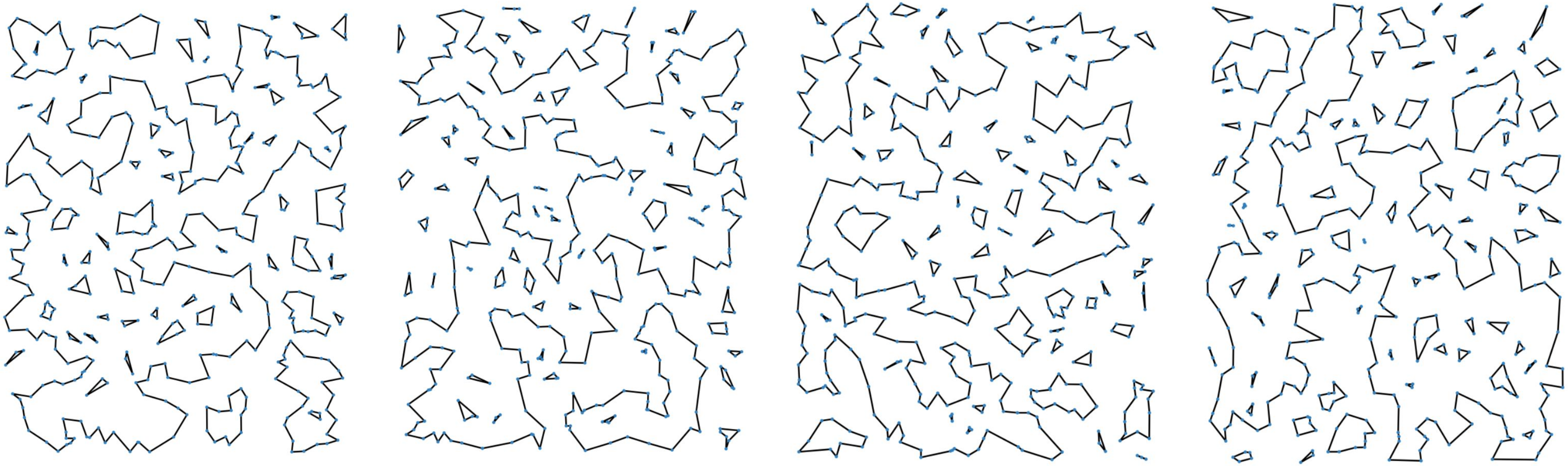
Looking for Clusters III: Biological Networks



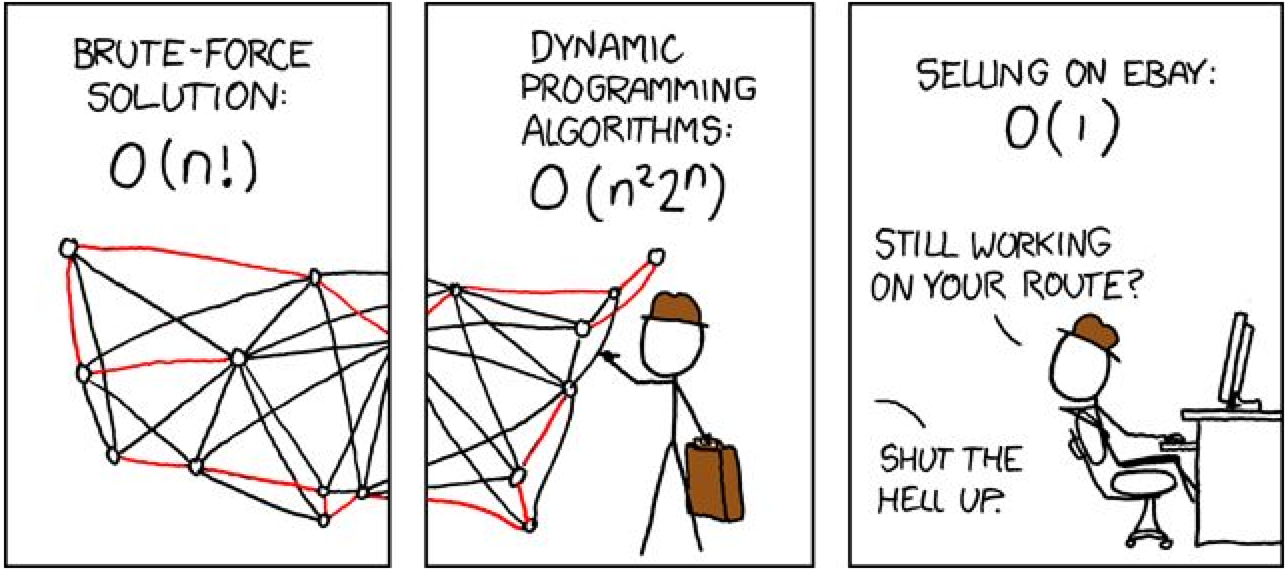
Looking for Clusters IV: Social Networks

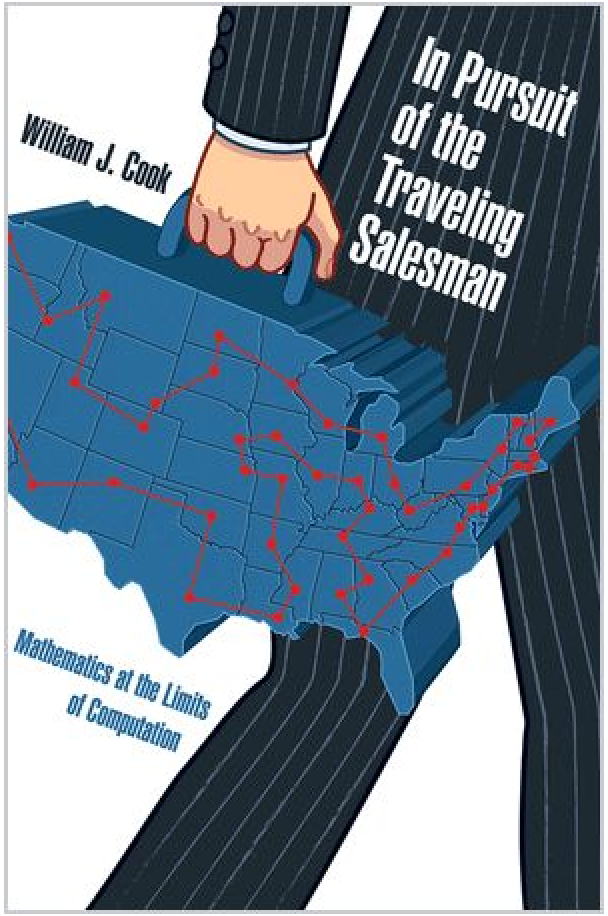


Looking for Clusters V: Euclidean 2-factors

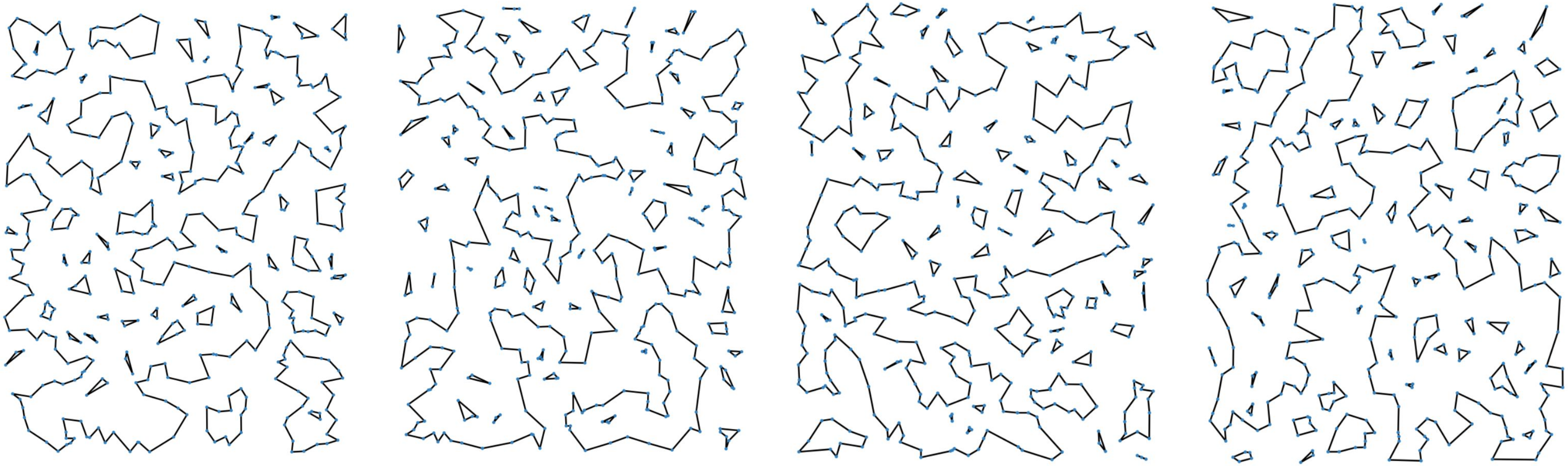


The Travelling Salesman Problem

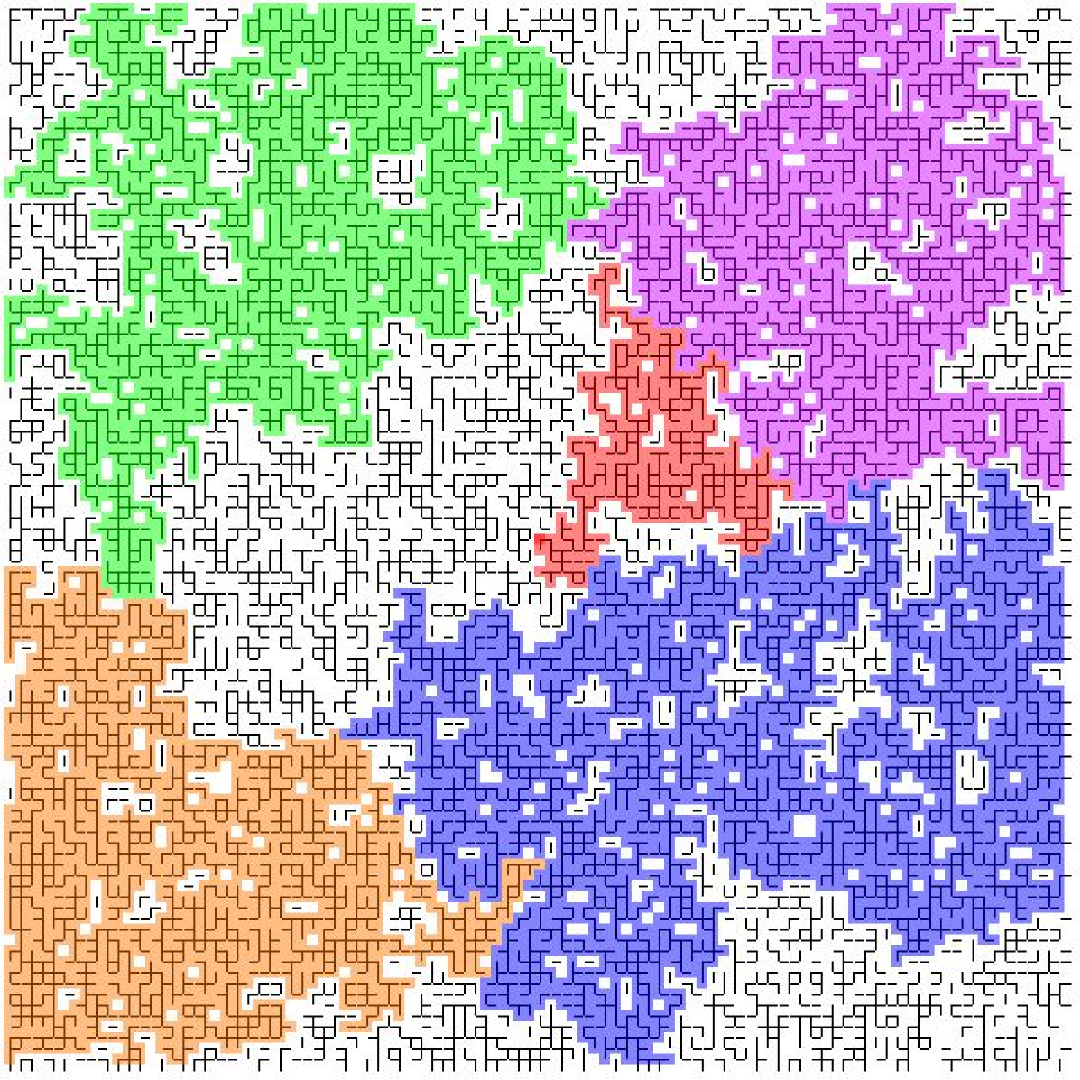



In Pursuit of the Traveling Salesman
William J. Cook
Mathematics at the Limits of Computation

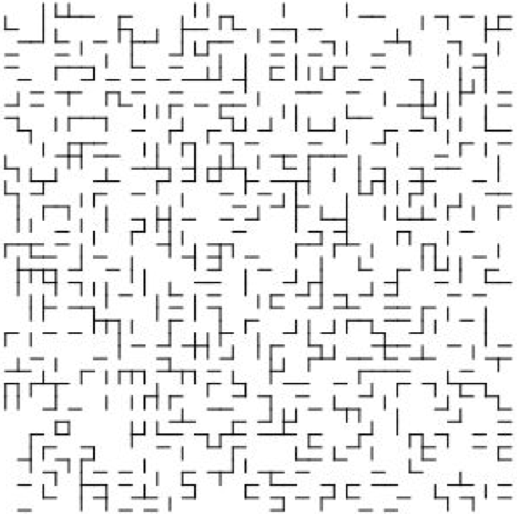
Looking for Clusters V: Euclidean 2-factors



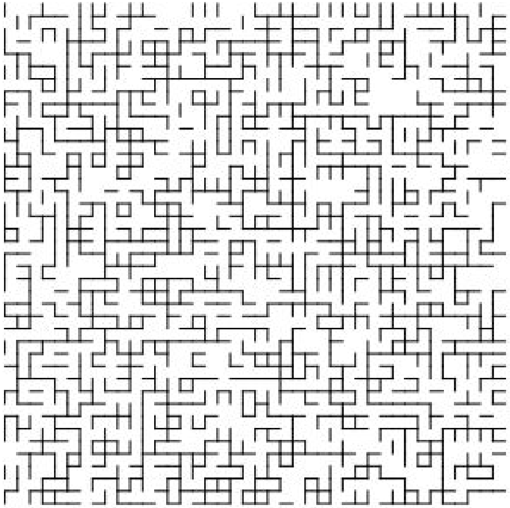
Looking for Clusters VI: Percolation



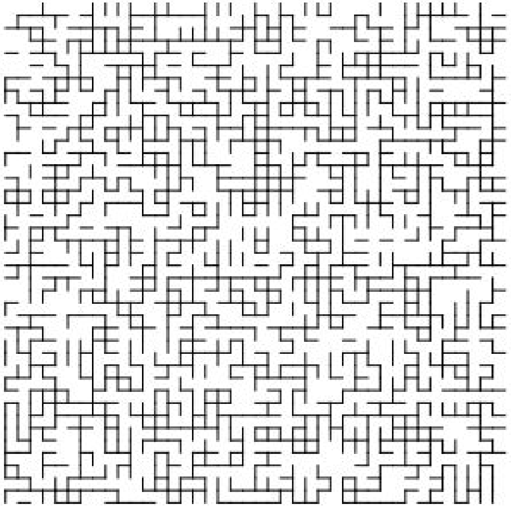
More Edges Means More Clustering



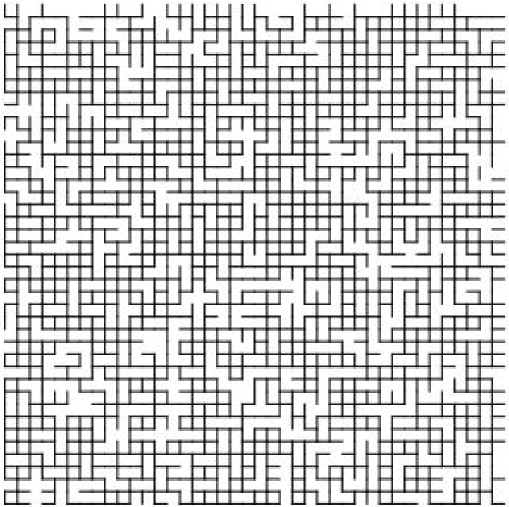
p=0.25



p=0.48

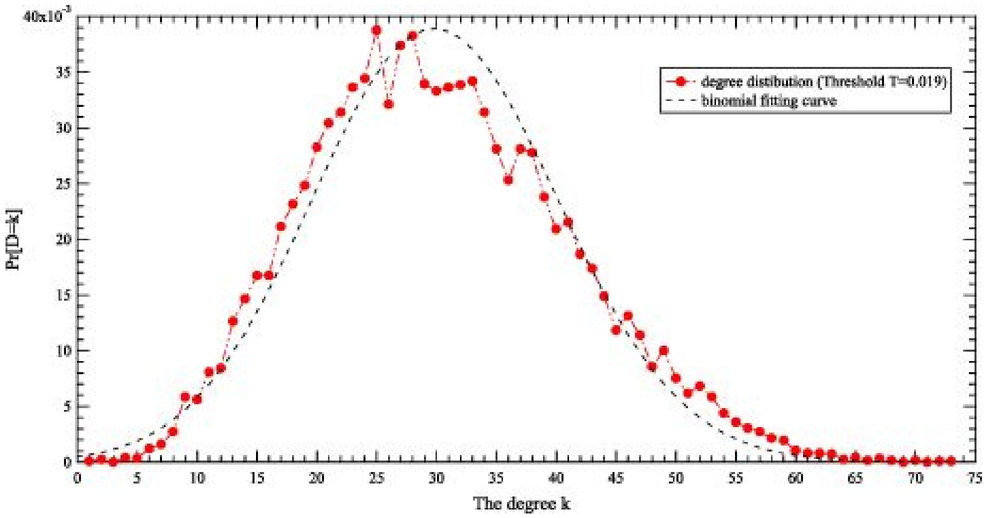


p=0.52



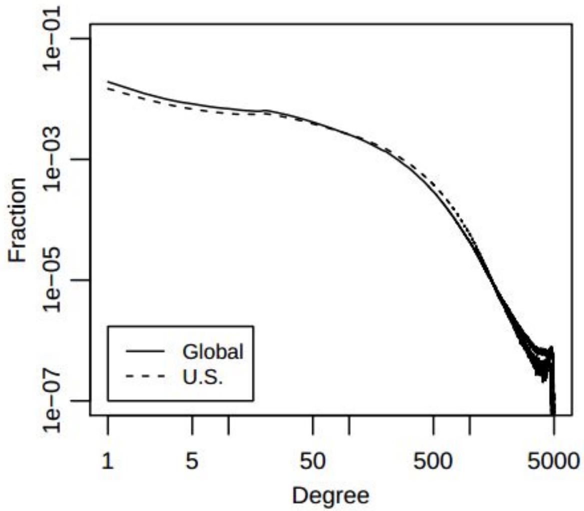
p=0.75

Degree Distributions Differ

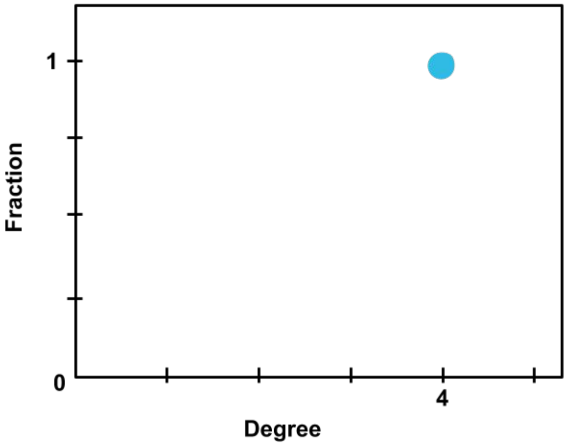


Classic Erdős-Renyi Model

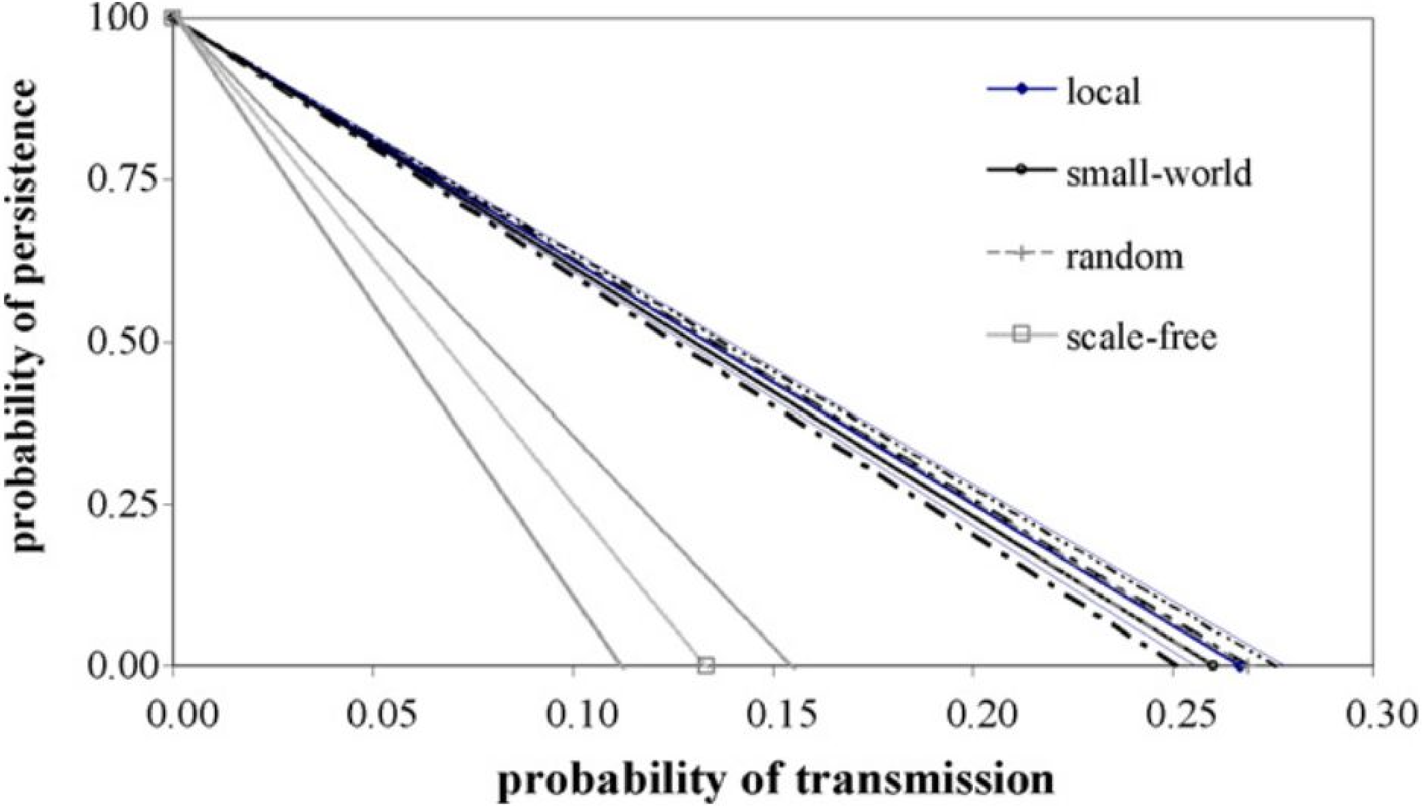
Facebook Friends



Lattice



Network Structure Affects Cluster Size



Random Networks as Controls

A common technique to analyze the properties of a single network is to use statistical randomization methods to create a reference network which is used for comparison purposes.

Mondragon and Zhou, 2012.

What is the probability a uniformly chosen graph on a given degree sequence has a linear sized component?

What is the probability a uniformly chosen graph on a given degree sequence has a linear sized component?

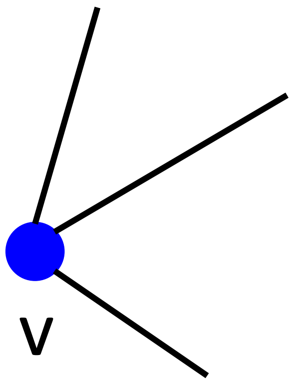
For a sequence D of nonzero degrees, G(D) is a uniformly chosen graph with degree sequence D.

What is the probability a uniformly chosen graph on a given degree sequence has a linear sized component?

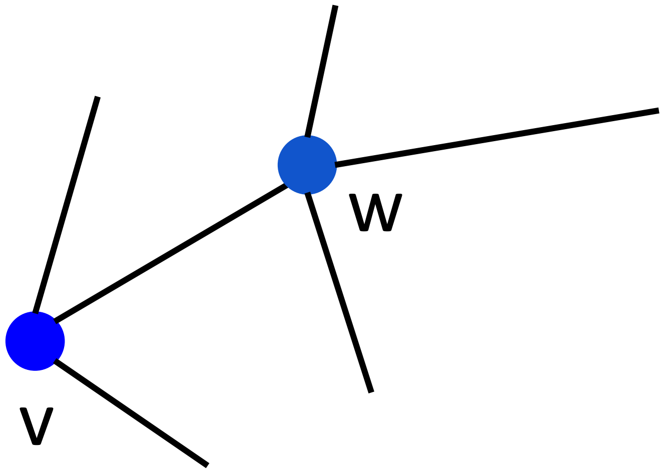
For a sequence D of nonzero degrees, G(D) is a uniformly chosen graph with degree sequence D.

Will assume D is non-decreasing and all degrees are positive.

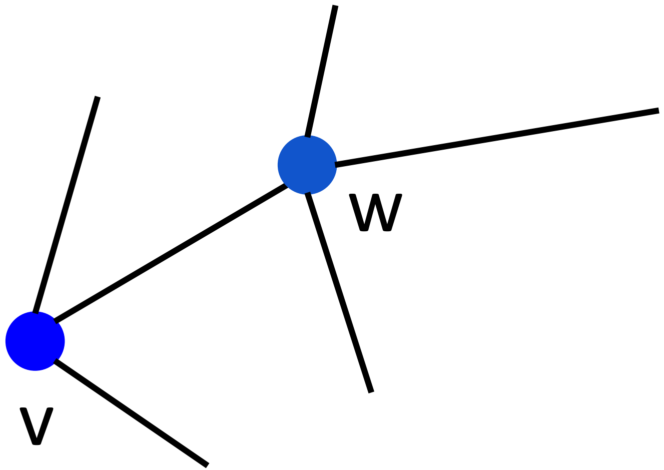
A Heuristic Argument



A Heuristic Argument



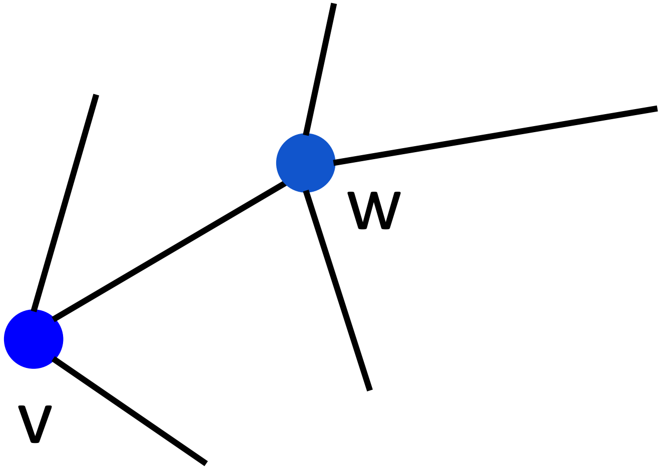
A Heuristic Argument



Change in number of open edges:

$$d(w) - 2$$

A Heuristic Argument



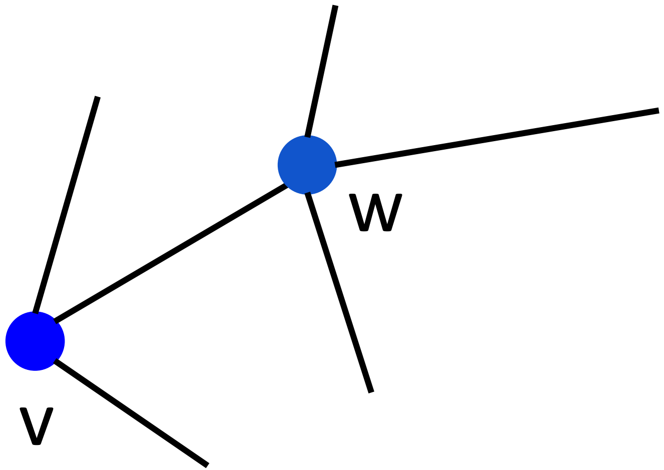
Change in number of open edges:

$$d(w) - 2$$

Probability pick w about

$$d(w) \,/\, \sum_u d(u)$$

A Heuristic Argument



Change in number of open edges:

$$d(w) - 2$$

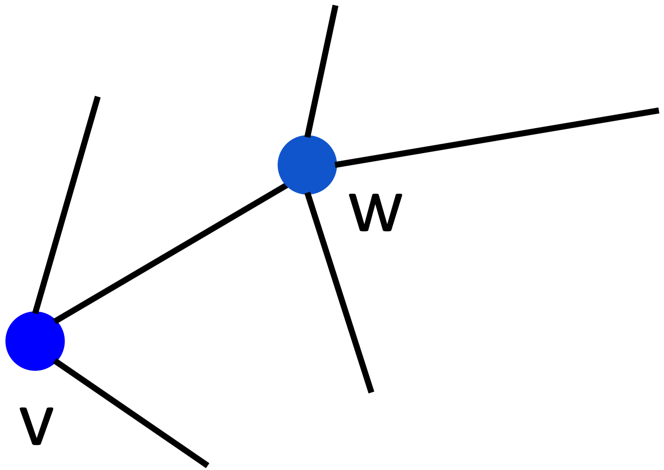
Probability pick w about

$$d(w) \,/\, \sum_u d(u)$$

Expected change about

$$\sum_u d(u)(d(u) - 2) \,/\, \sum_u d(u)$$

A Heuristic Argument



Change in number of open edges:

$$d(w) - 2$$

Probability pick w about

$$d(w) / \sum_u d(u)$$

Expected change about

$$\sum_u d(u)(d(u) - 2) / \sum_u d(u)$$

Giant Component if and only if $\sum_u d(u)(d(u)-2)$ is positive??

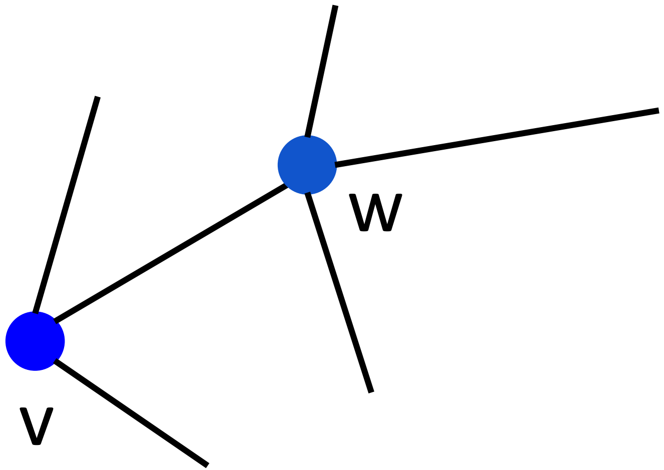
Molloy-Reed(1995) Result

Under considerable technical conditions including maximum degree at most $n^{1/8}$:

$$\sum_u d(u)(d(u) - 2) > \varepsilon n$$ implies a giant component exists.

$$\sum_u d(u)(d(u) - 2) < -\varepsilon n$$ implies no giant component exists.

Turning The Heuristic Argument Into A Proof



Change in number of open edges:

$$d(w) - 2$$

Probability pick w about

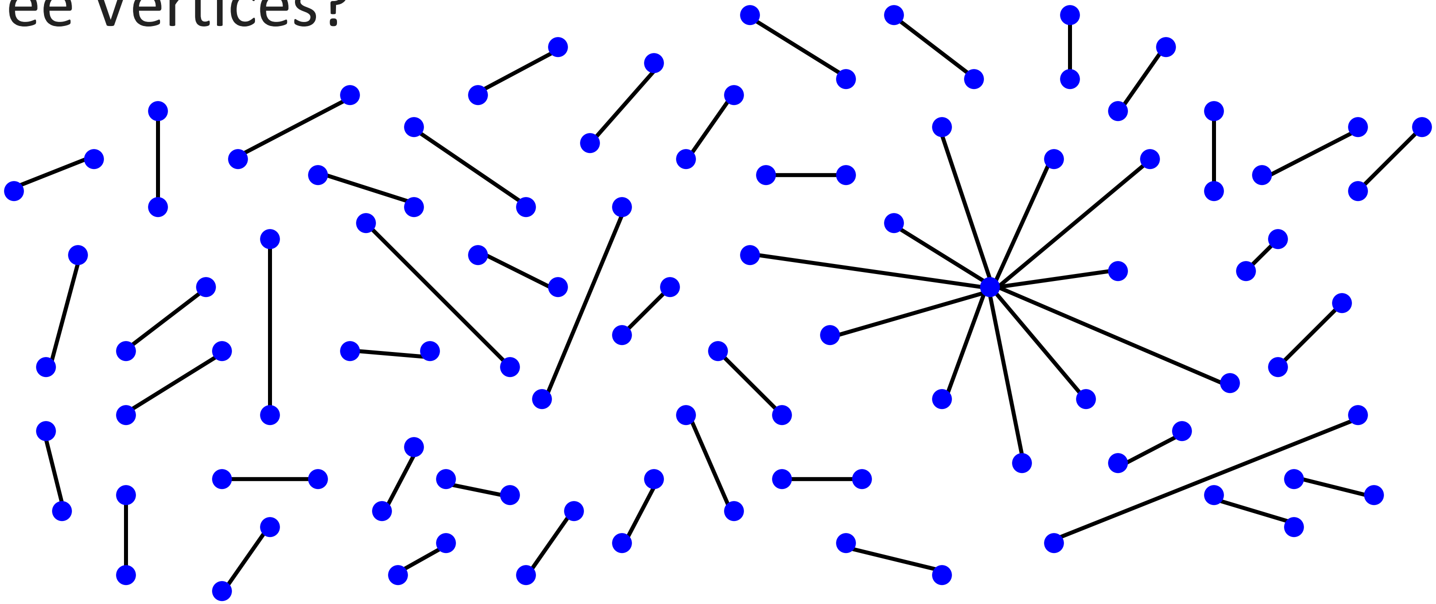
$$d(w) / \sum_u d(u)$$

Expected change about

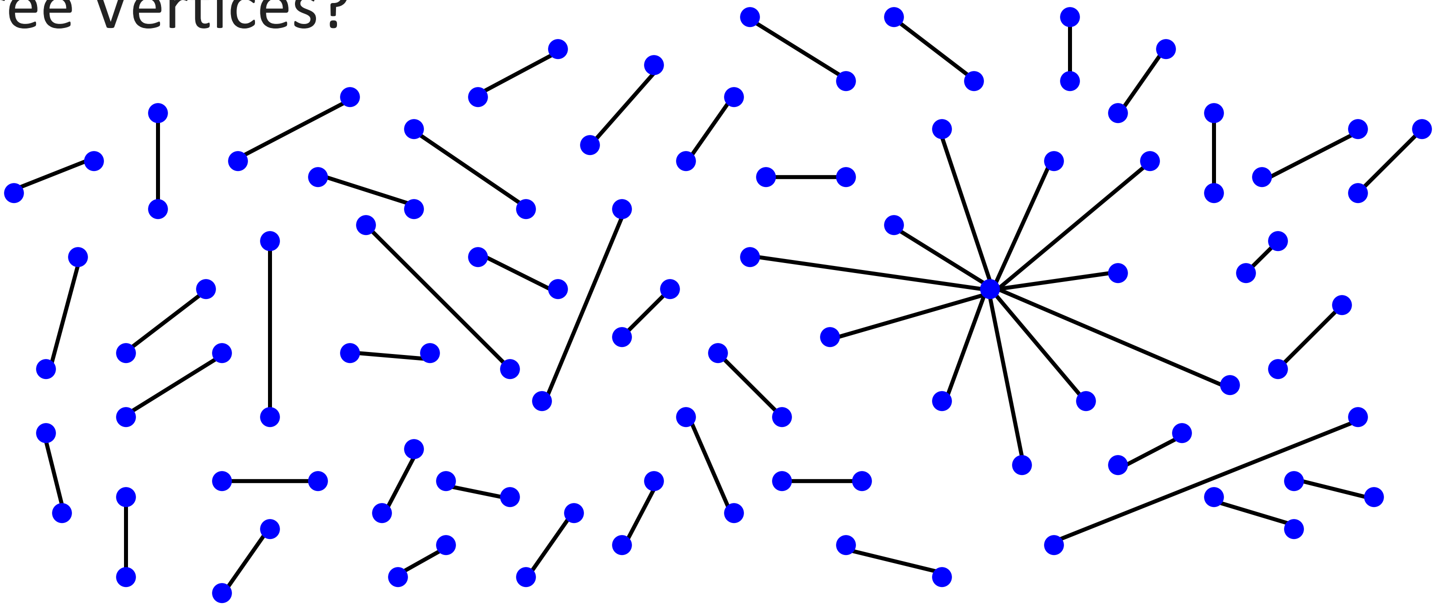
$$\sum_u d(u)(d(u) - 2) / \sum_u d(u)$$

Giant Component if and only if $\sum_u d(u)(d(u)-2)$ is positive??

Why Can't We Prove The Result For Graphs With High Degree Vertices?

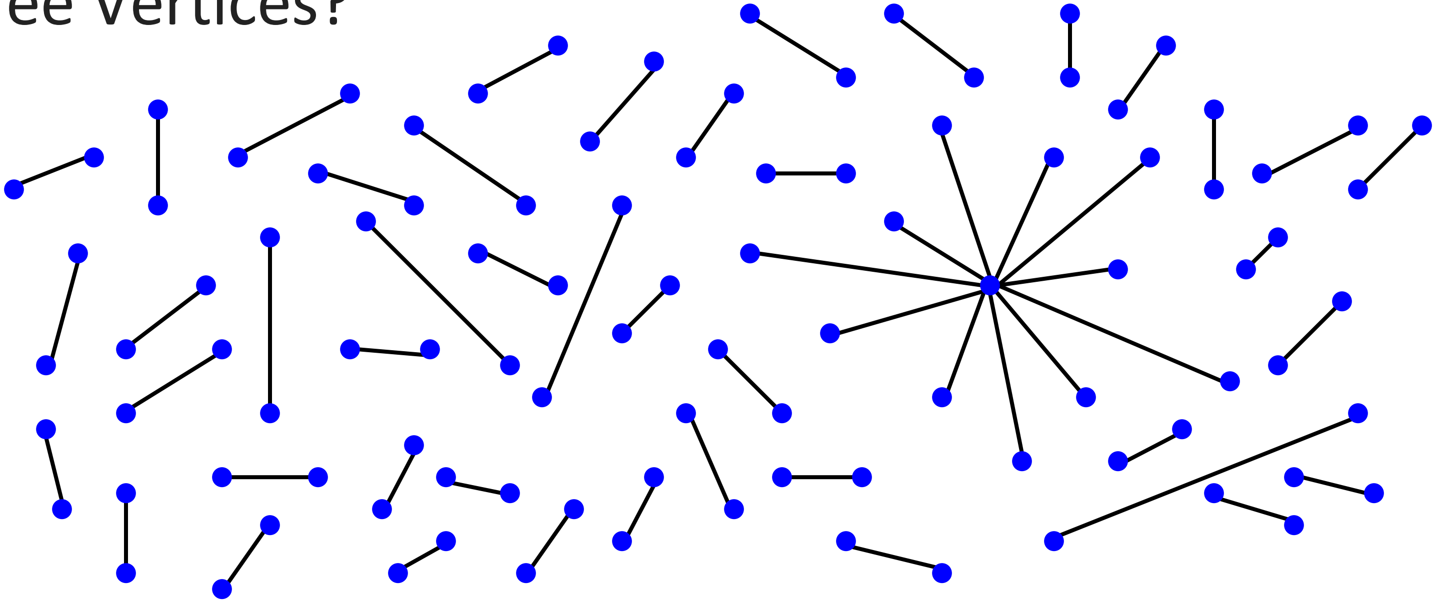


Why Can't We Prove The Result For Graphs With High Degree Vertices?



Because it is false.

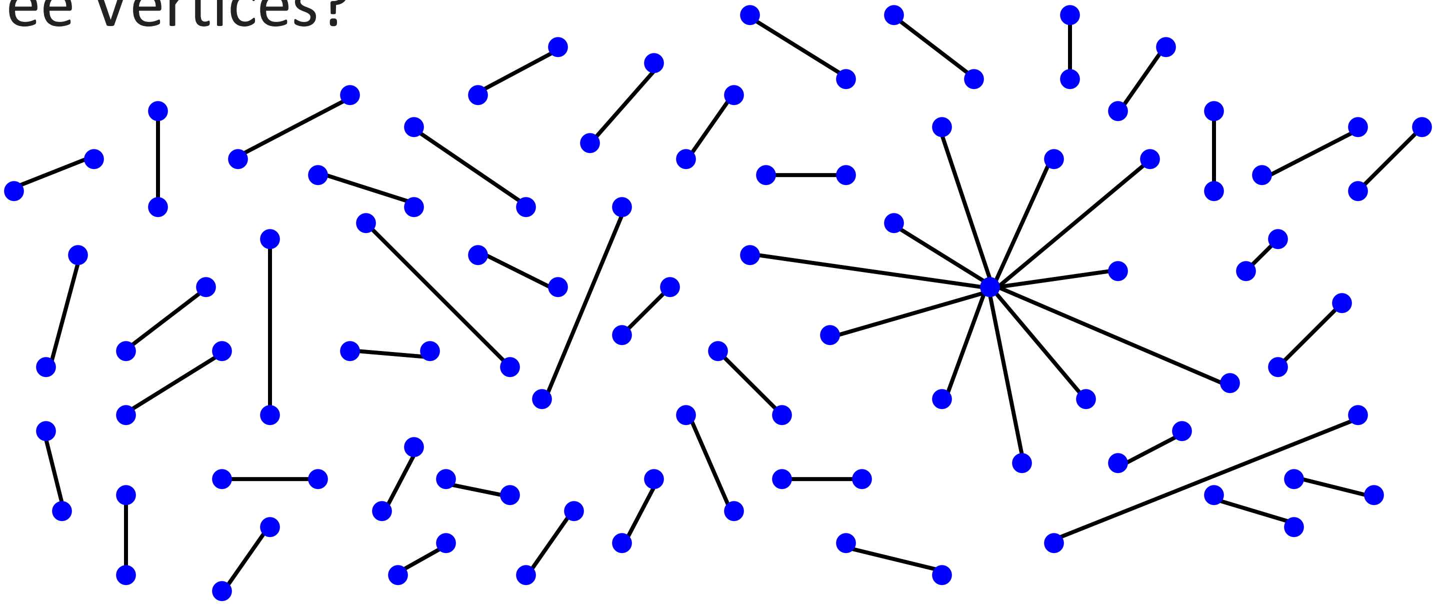
Why Can't We Prove The Result For Graphs With High Degree Vertices?



Cannot translate results from the multigraph model

.

Why Can't We Prove The Result For Graphs With High Degree Vertices?



Cannot translate results from the multigraph model.

Hard to prove concentration results.

OUR QUESTION REVISITED

Does a uniformly chosen graph on a given degree sequence have a giant component?

For a sequence D of nonzero degrees, G(D) is a uniformly chosen graph with degree sequence D.

Will assume D is non-decreasing and all degrees are positive.

Four Definitions

M is the sum of the degrees in D which are not 2.

D is f-well behaved if M is at least $f(n)$.

$$j_D = \min\left(i \text{ s.t. } \sum_{j=1}^{i} d_j(d_j - 2) > 0, n\right)$$

$$R_D = \sum_{j_D}^{n} d_j$$

One Crucial Observation

$\sum_{j=1}^{n} d(u)(d(u)-2)$ is at least R_D

One Crucial Observation

$$\sum_{j=1}^{n} d(u)(d(u)-2) \text{ is at least } R_D$$

and for some ɣ > 0 remains above $R_D/2$ until the sum of the degrees of the vertices explored is at least $ɣR_D$.

One Crucial Observation

$\sum_{j=1}^{n} d(u)(d(u)-2)$ is at least R_D

and for some ɣ > 0 remains above $R_D/2$ until the sum of the degrees of the vertices explored is at least $ɣR_D$.

But goes negative once all the vertices with index > j_D are explored.

Two Theorems

Theorem 1: For any $f \longrightarrow \infty$ and $b \longrightarrow 0$, if a well behaved degree distribution D satisfies $R_D \leq b(n)M$ then G(D) has no giant component

.

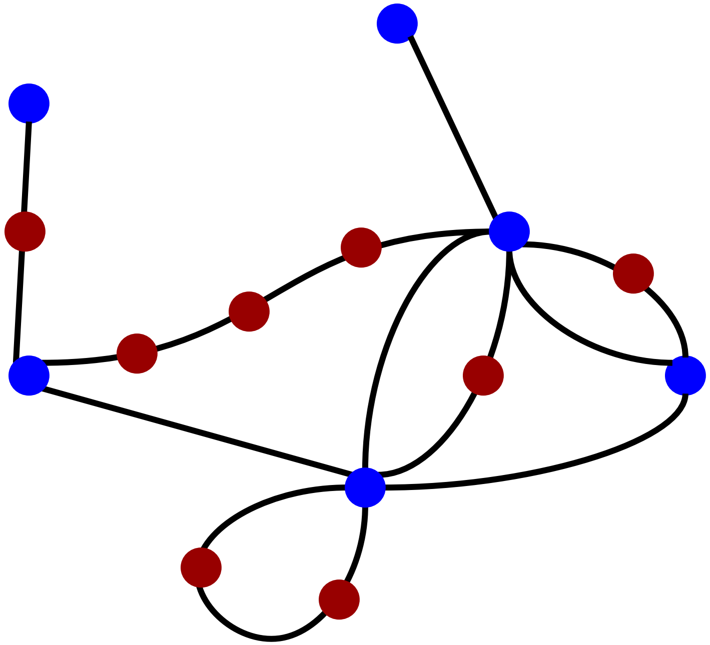
Two Theorems

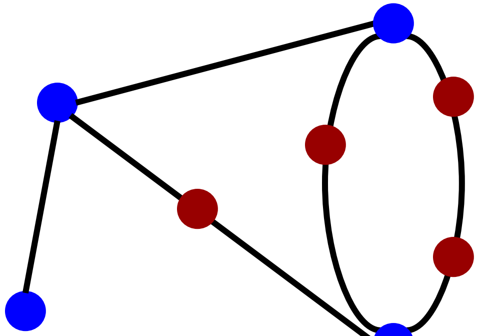
Theorem 1: For any $f \longrightarrow \infty$ and $b \longrightarrow 0$, if a well behaved degree distribution D satisfies $R_D \leq b(n)M$ then G(D) has no giant component.

Theorem 2: For any $f \longrightarrow \infty$ and $\varepsilon > 0$ if a well behaved degree distribution D satisfies $R_D \geq \varepsilon M$ then G(D) has a giant component

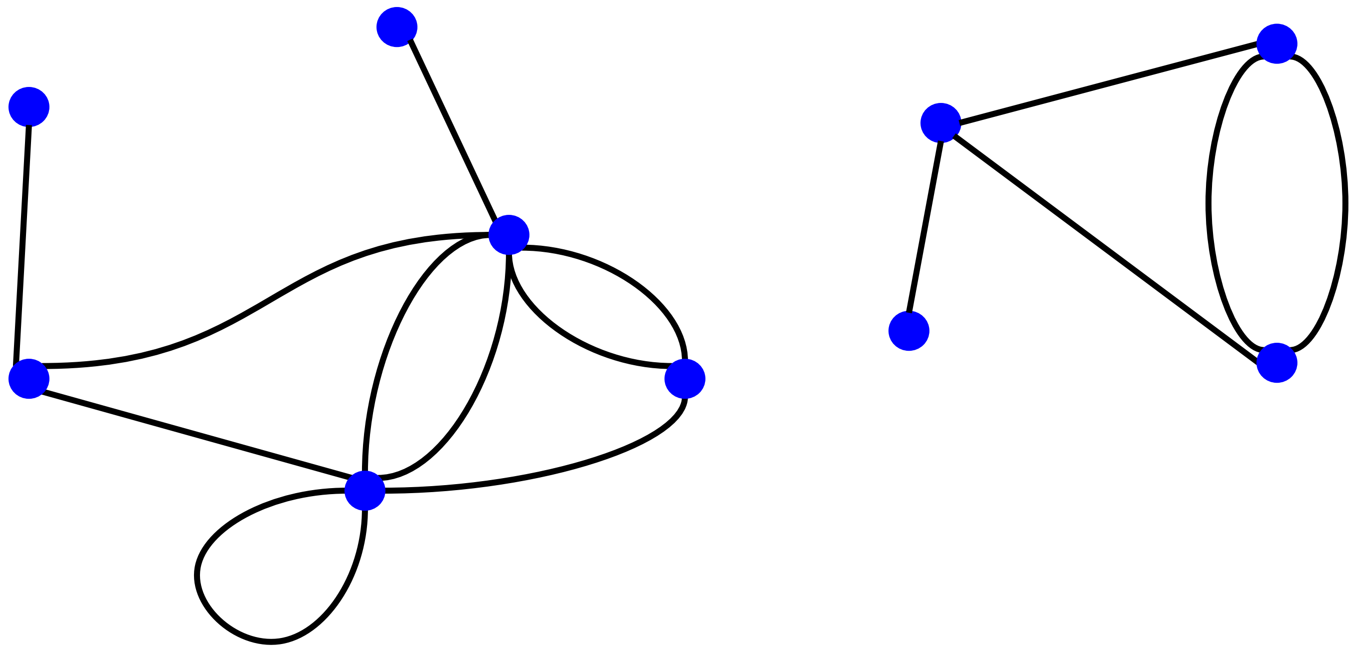
(Joos, Perarnau-Llobet, Rautenbach, Reed 2015)

Why we focus on M and not n

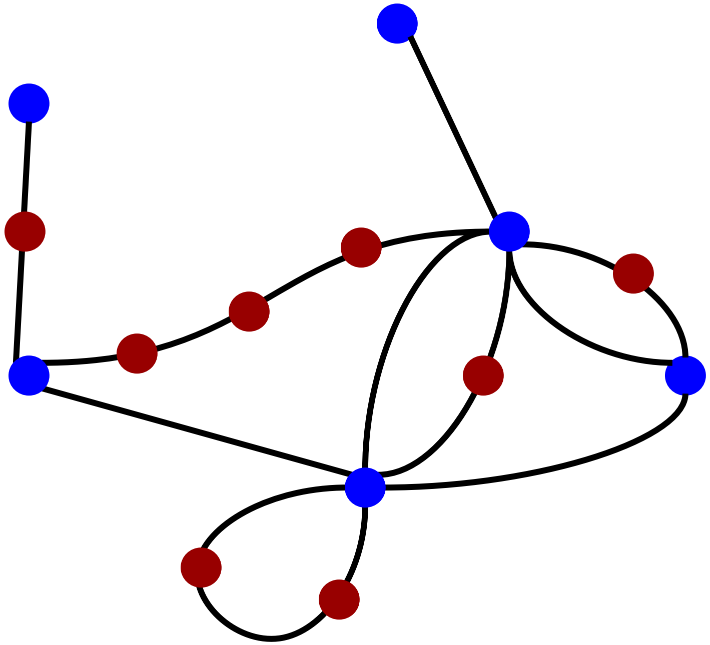


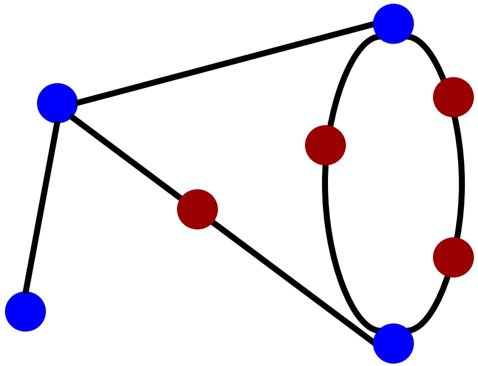


Why we focus on M and not n



Why we focus on M and not n





What About Badly Behaved Graphs?

Badly Behaved graphs do not have 0-1 Behaviour

Badly Behaved graphs do not have 0-1 Behaviour

For all $0<\varepsilon<1$, the probability of a component of size at least εn lies between c and $1-c$ for some constant c between 0 and 1.

Badly Behaved graphs do not have 0-1 Behaviour

For all $0<\varepsilon<1$, the probability of a component of size at least εn lies between c and 1-c for some constant c between 0 and 1.

If all vertices of degree 2 just taking a random 2-factor.

Badly Behaved graphs do not have 0-1 Behaviour

For all $0<\varepsilon<1$, the probability of a component of size at least εn lies between c and 1-c for some constant c between 0 and 1.

If all vertices of degree 2 just taking a random 2-factor.

If M is at most some constant b, with probability $p(b)>0$ all but $\varepsilon n/2$ of the vertices lie in cyclic components.

Two Theorems

Theorem 1: For any $f \rightarrow \infty$ and $b \rightarrow 0$, if a well behaved degree distribution D satisfies $R_D \leq b(n)M$ then G(D) has no giant component.

Theorem 2: For any $f \rightarrow \infty$ and $\varepsilon > 0$ if a well behaved degree distribution D satisfies $R_D \geq \varepsilon M$ then G(D) has a giant component

(Joos, Perarnau-Llobet, Rautenbach, Reed 2015)

Differences in the Proof

Determine if there is a component K of the multigraph obtained by suppressing degree 2 vertices satisfying:

(*) $|E(K)| > \varepsilon' M$.

Differences in the Proof

Determine if there is a component K of the multigraph obtained by suppressing degree 2 vertices satisfying:

(*) $|E(K)| > \varepsilon' M$.

Use a combinatorial switching argument to obtain bounds on edge probabilities in this multigraph.

Differences in the Proof - When No Giant Component Exists

Begin the random process with a large enough set of high degree vertices that our process has negative drift.

Differences in the Proof - When No Giant Component Exists

Begin the random process with a large enough set of high degree vertices that our process has negative drift.

Show drift becomes more and more negative over time, if the process does not die out.

Differences in the Proof - When A Giant Component Exists

Focus on the set $H = \{v \mid d(v) > (\sqrt{M})/\log(M)\}$

Differences in the Proof - When A Giant Component Exists

Focus on the set H = {v | d(v) > (√M)/log(M)}

We can show, using our combinatorial switching argument, that depending on the sum of the sizes of the components intersecting H, either

(a) there is a giant component containing all of H , or
(b) we can reduce to a problem with H empty.

Demonstrating The Switching Argument

Demonstrating The Switching Argument

Theorem: If $|E|>8n \log n$ then,

$$\text{Prob}(G \text{ has a component with } (1-o(1))n \text{ vertices}) = 1-o(1).$$

Future Work

Tight bounds on the size of the largest component in terms of R_D

Thank you for your attention!