August 1st, 2025

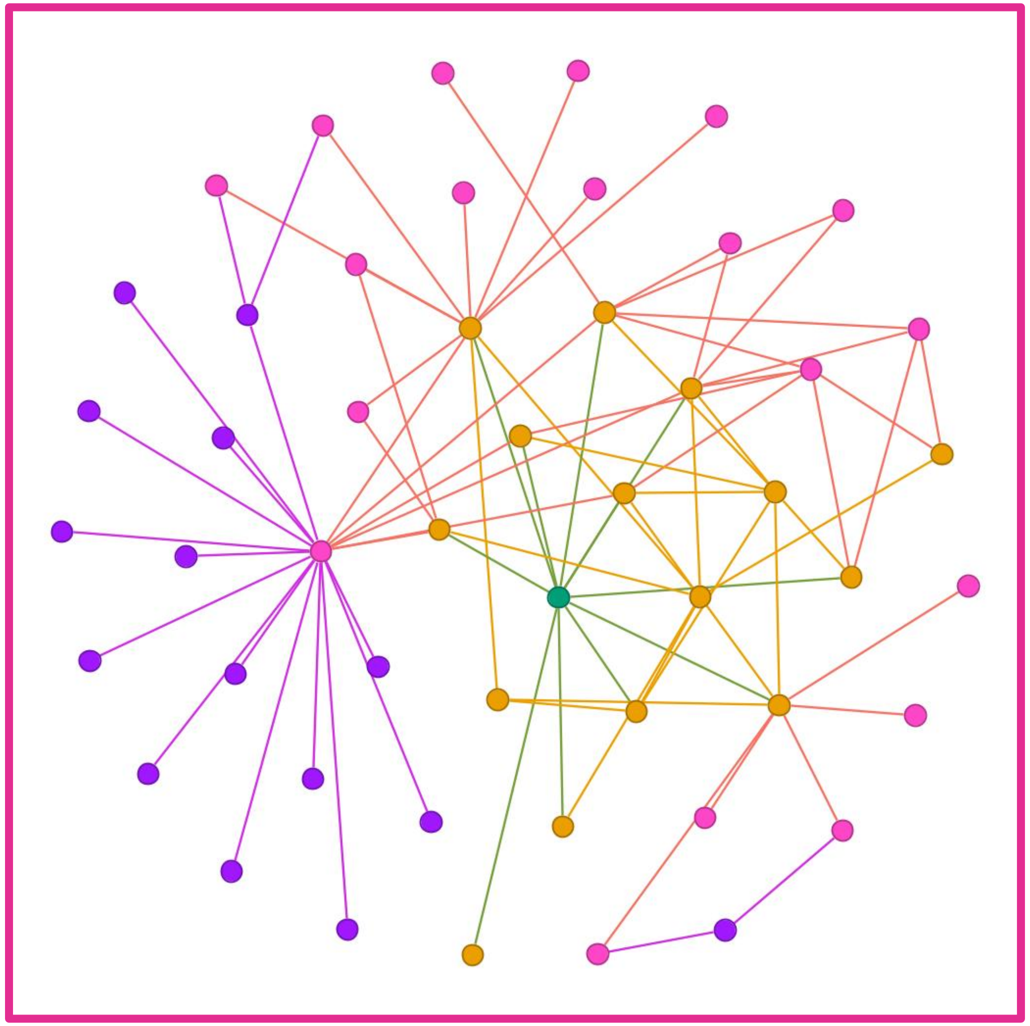
INFORMATION DIFFUSION IN COMPLEX NETWORKS

A Proportion-Based Approach

Megan Bryson under the supervision of Dr. Meger

OUTLINE

- Introduction
- Past work
- Definition of our model
- Results
- Conclusion

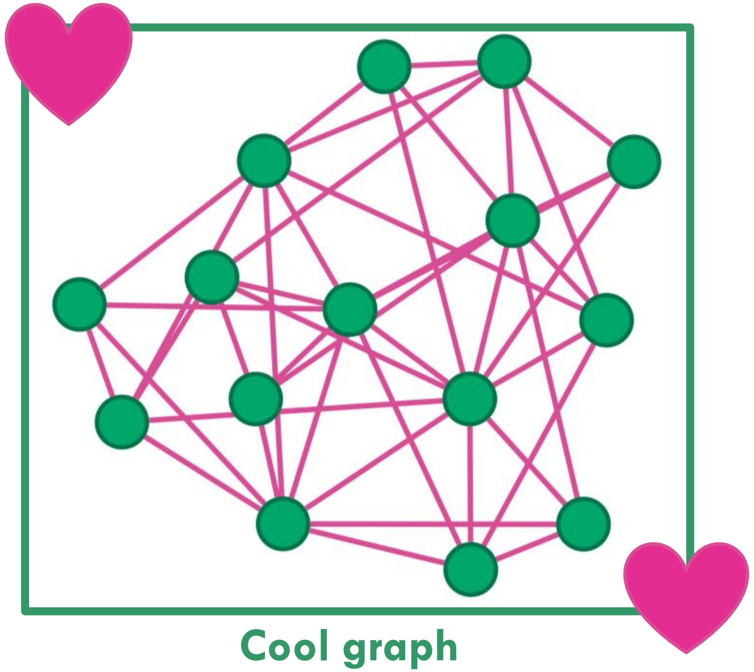


INTRODUCTION

WHAT IS A COMPLEX NETWORK?

- A type of graph (no, not *that* type of graph)
- All complex networks are graphs, *not all* graphs are complex networks
- Both are comprised of **edges** and **nodes** (see Cool graph)
- Complex networks have specific properties but can be hard to define precisely as we are trying to define a thing that exists already, instead starting with our definitions and creating something new

Boring graph

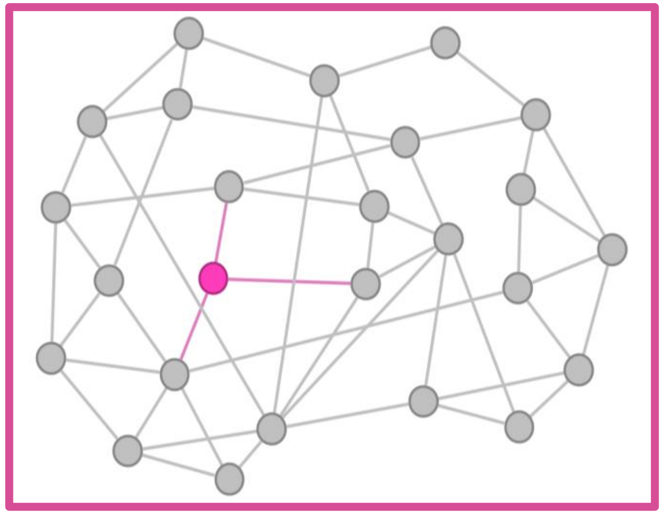


Cool graph

NODE PROPERTIES

Degree

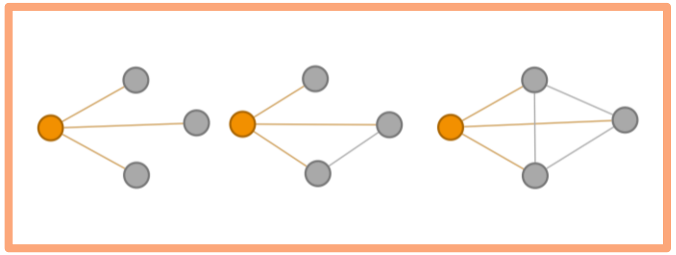
How many edges a node is incident to



The pink node has degree 3

Clustering Coefficient

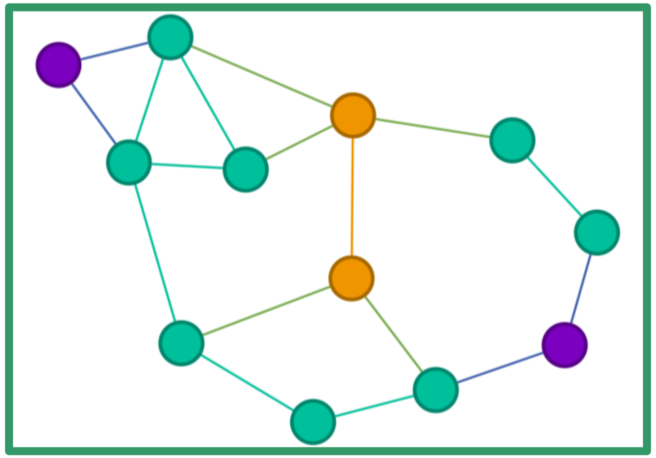
Proportion of neighbours that are also connected to each other, compared to the maximum possible



The orange node has C = 0, C = 1/3 and C = 1 respectively

Eccentricity

The maximum path length from your starting node to all other nodes in the graph



Orange, green and purple nodes have eccentricity 3, 4 and 5 respectively

4 PRINCIPLES OF COMPLEX NETWORKS

Large

- 100s, 1000s or 1,000,000,000s of nodes

Evolving

- Nodes and edges can be both created and deleted as the graph grows

The small world phenomenon

- Small distance (two nodes are typically joined by a short chain of mutual acquaintances)
- Clustering effect (two nodes who share a common neighbor are more likely to know each other)

Power law degree distribution

- There are few nodes of high degree, and many nodes with low degrees

Dr. Chung

4 PRINCIPLES OF COMPLEX

Actually, I said the second property was ‘sparse’, but contemporary works often swap it for ‘evolving’…

Dr. Chung

Large

- 100s, 1000s or 1,000,000,000s of nodes

Evolving

- Nodes and edges can be both created and deleted as the graph grows

The small world phenomenon

- Small distance (two nodes are typically joined by a short chain of mutual acquaintances)
- Clustering effect (two nodes who share a common neighbor are more likely to know each other)

Power law degree distribution

- There are few nodes of high degree, and many nodes with low degrees

~~4~~ 5 PRINCIPLES OF COMPLEX NETWORKS

Large

- 100s, 1000s or 1,000,000,000s of nodes

Evolving

- Nodes and edges can be both created and deleted as the graph grows

The small world phenomenon

- Small distance (two nodes are typically joined by a short chain of mutual acquaintances)
- Clustering effect (two nodes who share a common neighbor are more likely to know each other)

Power law degree distribution

- There are few nodes of high degree, and many nodes with low degrees

Densification

- As the graph continues to grow, the ratio between edges and nodes increases

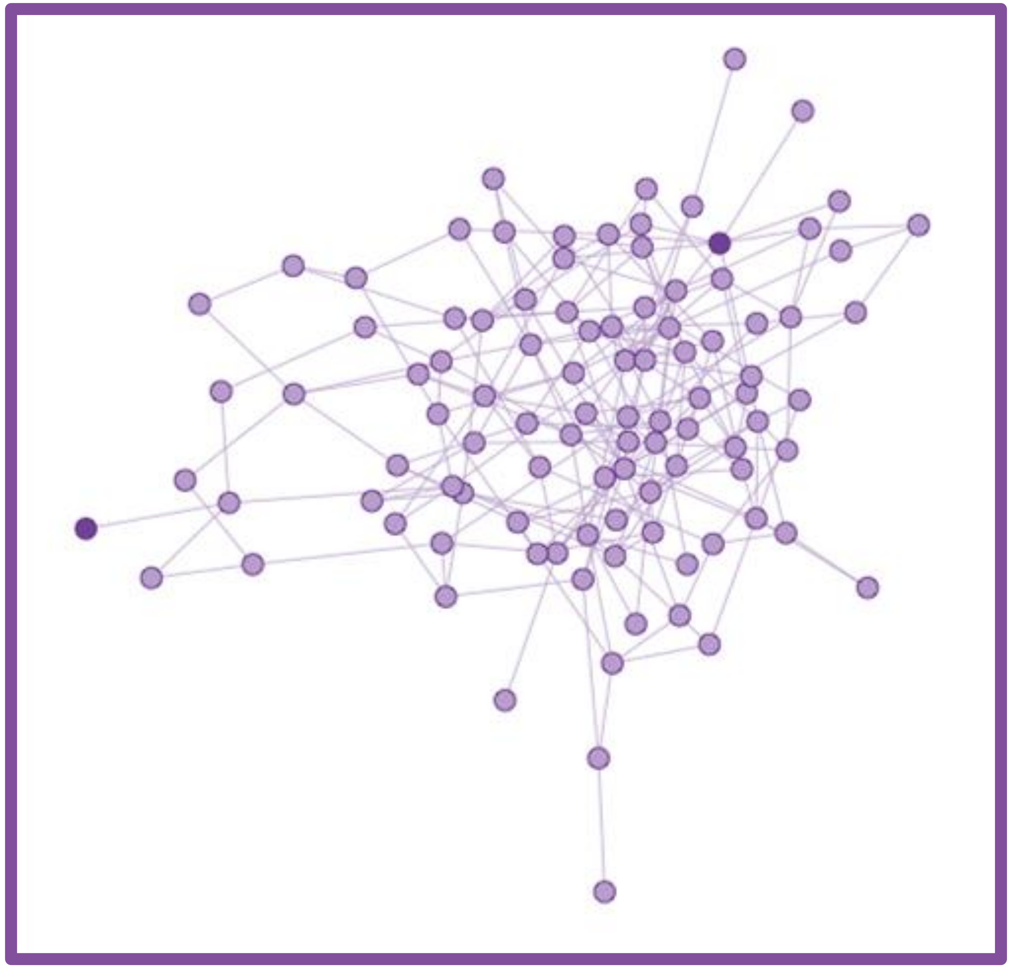
Dr. Chung

The community*

* Mostly:
J. Leskovec,
J. Kleinberg,
C. Faloutsos,

WHAT IS INFORMATION DIFFUSION TO US?

- How does information spread throughout a graph?
- Examples:
  - Disease spreading across a community
  - A rumour spreading through a school
  - Fire spreading in a forest



DIFFERENCE BETWEEN DETERMINISTIC AND PROBABILISTIC

Probabilistic

Things will be different every time

- Probability is added to accurately model real-world randomness leading to highly dynamic processes
- Examples:
  - proportions of edges are kept
  - nodes are created with some probability

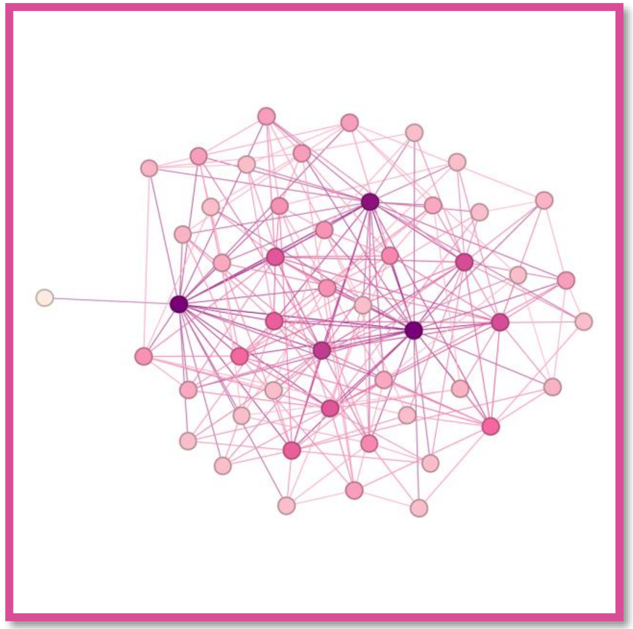
Deterministic

Same decisions will be made every time

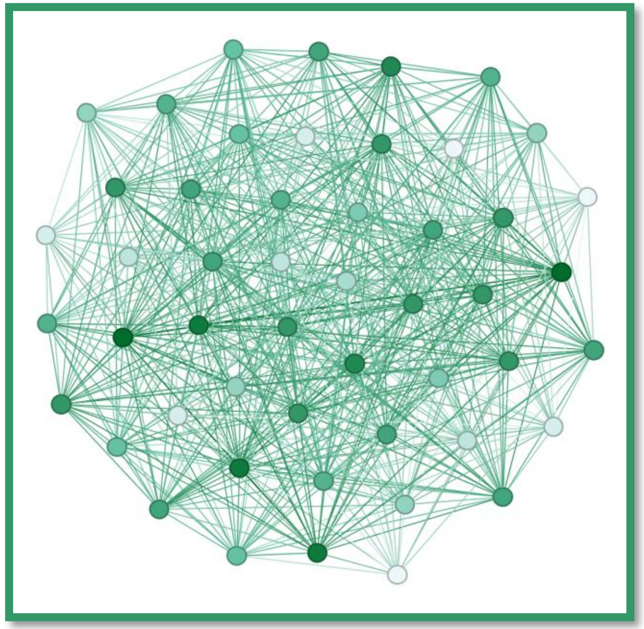
- Many decisions are made with an absolute answer
- If questions arise, there is an algorithm to determine the outcome, that does not rely on chance

PAST WORK

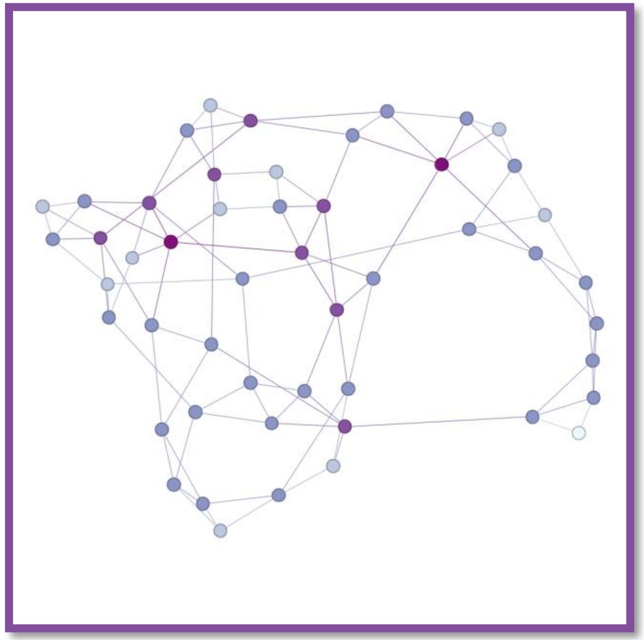
Barabási–Albert



Erdős–Rényi-Gilbert

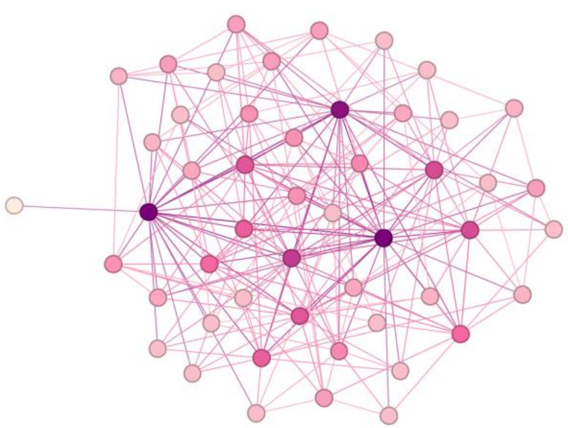


Watts–Strogatz



Barabási–Albert

Example



N = 50, M = 5

Fun facts

- Often abbreviated as "BA" or "PA"
- Created by Réka Albert and Albert-László Barabási in 1999
- Albert and Barabási were the first to use the term "preferential attachment"
- Average degree: 2**M**

- Final graph will have **N** nodes
- Grows by one node each step
- Each new node will have **M** edges, attaching to **M** existing nodes
- The **M** new edges are chosen with *Preferential Attachment*
- This means, nodes with higher degree have a higher probability of being chosen
- Scale-free

$$P(v) = \frac{\deg(v)}{2|E(G)|}$$

Watts–Strogatz

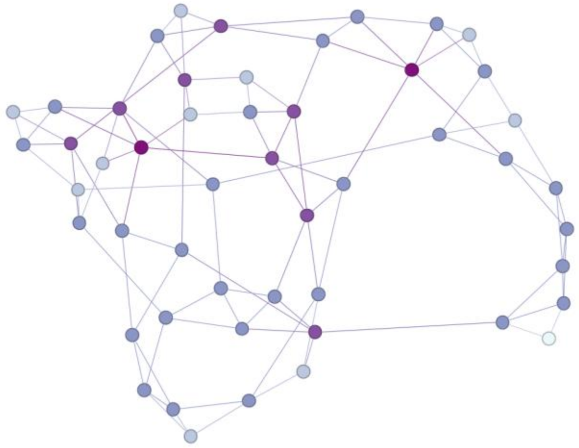

Dr. Watts


Dr. Strogatz

Fun facts

- Abbreviated as “WS”
- Created by Duncan Watts and Steven Strogatz in 1998
- Strogatz was Watts doctoral supervisor!
- Average degree: 2**K** when **ß** is zero
- This is the only model we use with 3 variables

Example


N = 50, K = 5, ß = 0.2

- Small world graph (the first one ever!)
- High clustering
- Has **N** nodes, originally connected to **K** neighbours in a ring lattice
- Edges have probability **ß** to be replaced by a random edge (non-self loop, non-duplicate)
- When **ß** = 0, the graph is just a ring lattice, when **ß** = 1 , the graph is completely random
- When **ß** = 1, the graph is ~ an G(N,P) graph

Example

N = 50, K = 3, ß = 0.2

Erdős–Rényi-Gilbert

- Most famous graph generation model
- Final graph will have **N** nodes
- Each edge has probability **P** of connecting two nodes, this is independent of other nodes connections
- Low clustering
- Binomial degree distribution

Dr. Rényi

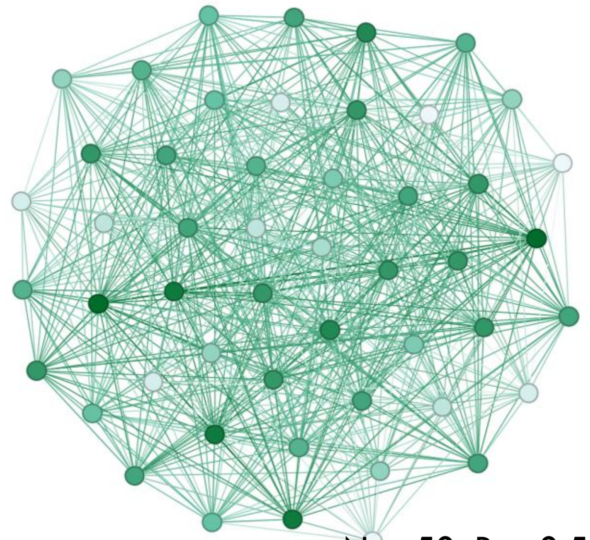
Dr. Erdős

Dr. Gilbert

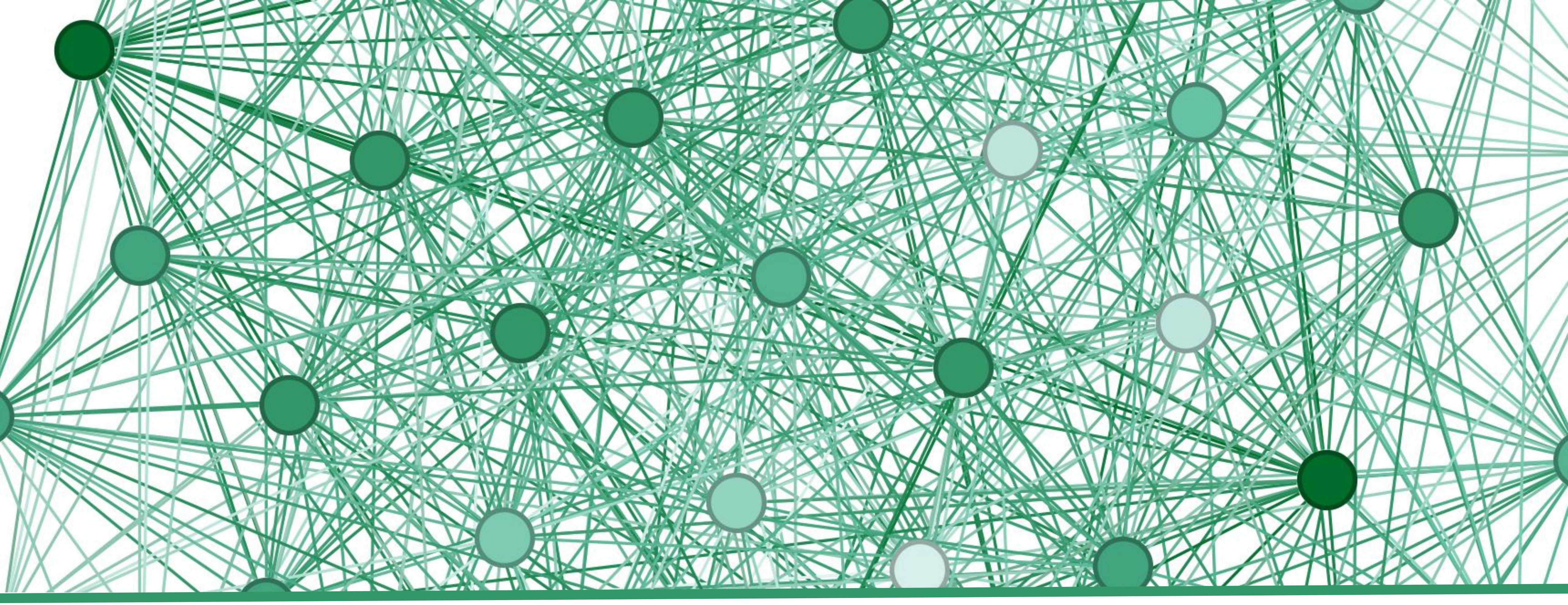
Fun facts

- Often called G(**N**,**P**)
- Popularized by Paul Erdős and Alfréd Rényi in 1959
- "Erdős–Rényi" can refer to two models, making it important to specify which you are using! We use the one constructed by Edgar Gilbert (if you ever use G(**N**,**P**) you are too!)
- Average degree: **NP**

Example



N = 50, P = 0.5



DEFINITION OF OUR MODEL

PROPORTIONAL INFLUENCE MODEL

- Created to model how influence moves throughout a graph
- Stemmed from the idea that "if I'm influenced by many sources, each of them is individually less influential"
- For example: if I have 100 friends and 3 tell me to read a book, I might. But if I have 4 friends and 3 tell me read a book, I'm a lot more likely to

Requirements to run the algorithm:

- A fully connected graph of size at least 3
- All of the nodes of the graph in an ordered list
- 2 (separate) starting nodes, in our cases we've been using 1 red node and 1 blue

Dr. Meger

Megan Bryson

Fun facts

- Created by Megan Bryson and Dr Erin Meger in 2024
- Dr Meger is Megan's supervisor
- They are both speaking at this conference! You just heard her speak!
- They both did their undergrads at WLU, which is also where they met
- All her grad students are named Me(a)gan, she maintains this is a coincidence

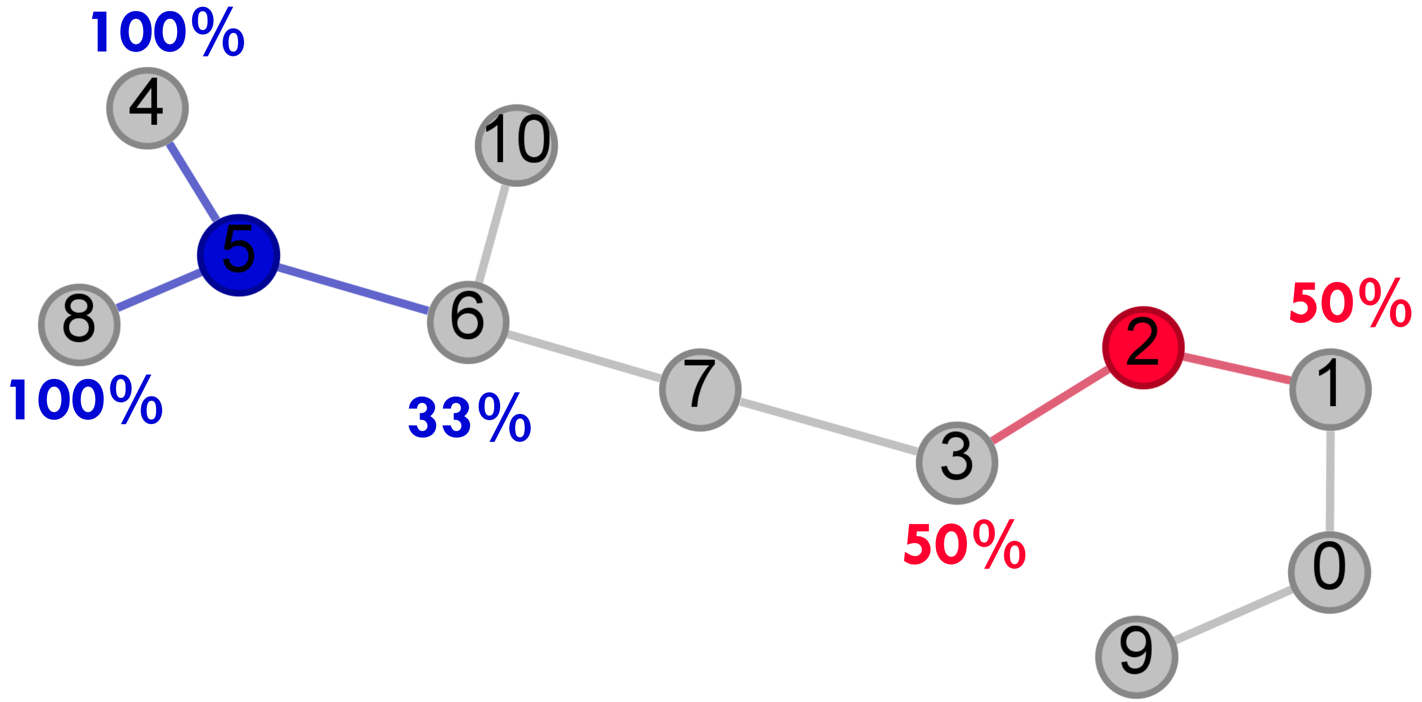
HOW IT WORKS *

- Assign all nodes a color:
  - All nodes begin grey
  - Seed: select a set of nodes to be the red seed and the blue seed (we will start by having seeds of size 1)
- At each step, we select a single node to be influenced
  - Determine which node receives the most "pressure to change' (highest influence calculated based on a proportion of neighbours exerting influence),
  - It receives the color of this pressure
  - Note: conflicts may arise, see right box
- Continue coloring nodes one at a time until the entire graph is colored.
  - If a tie is still present at the end of colouring, nodes in even slots in the ordered list will turn red and odd will turn blue (this is to keep the algorithm deterministic), then the algorithm is restarted on any nodes that were blocked the stall
- We check the final proportion of red and blue nodes

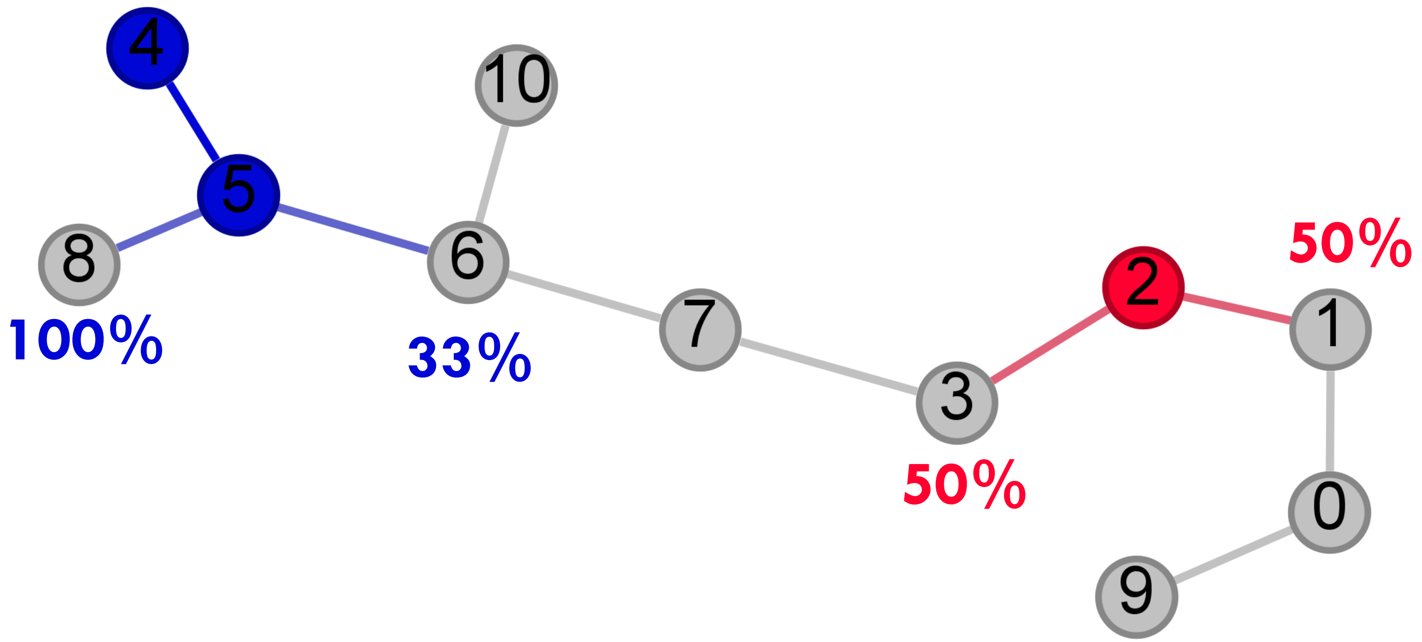
- Once coloured blue or red, the colour is fixed
- In the event of a tie between several nodes receiving the same amount of influence, the one appearing first in our ordered list will go first
- If a node has equal influence from red and blue (2 blue neighbours, 2 red neighbours), it will be skipped over during the current step in favour of the next influenced node colouring instead

* In words, a visual demo will be shown shortly

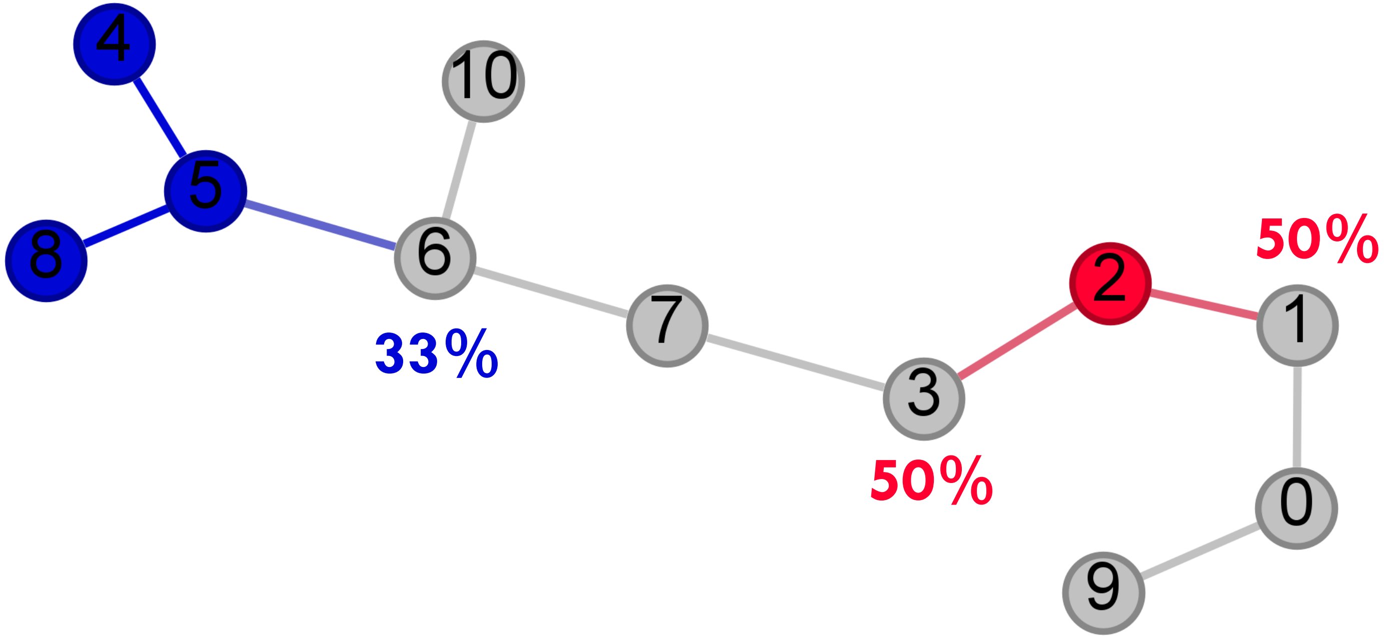
STEP 1



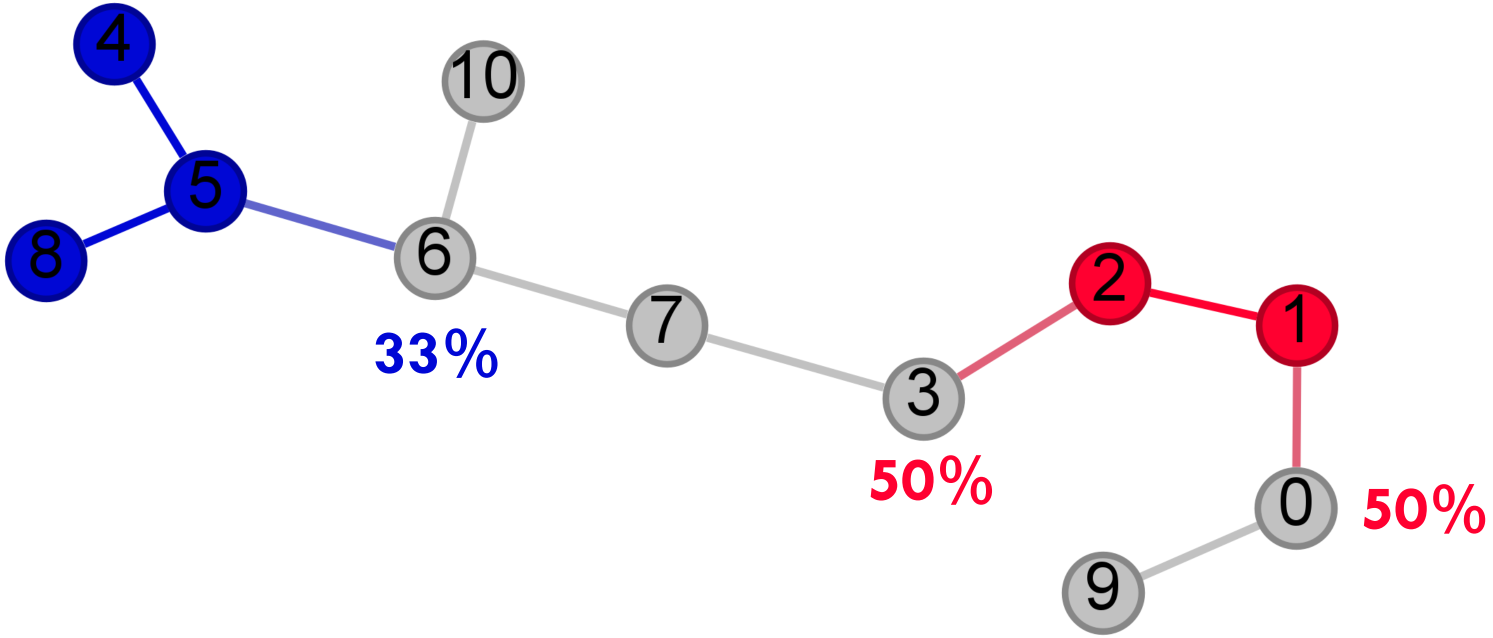
STEP 2



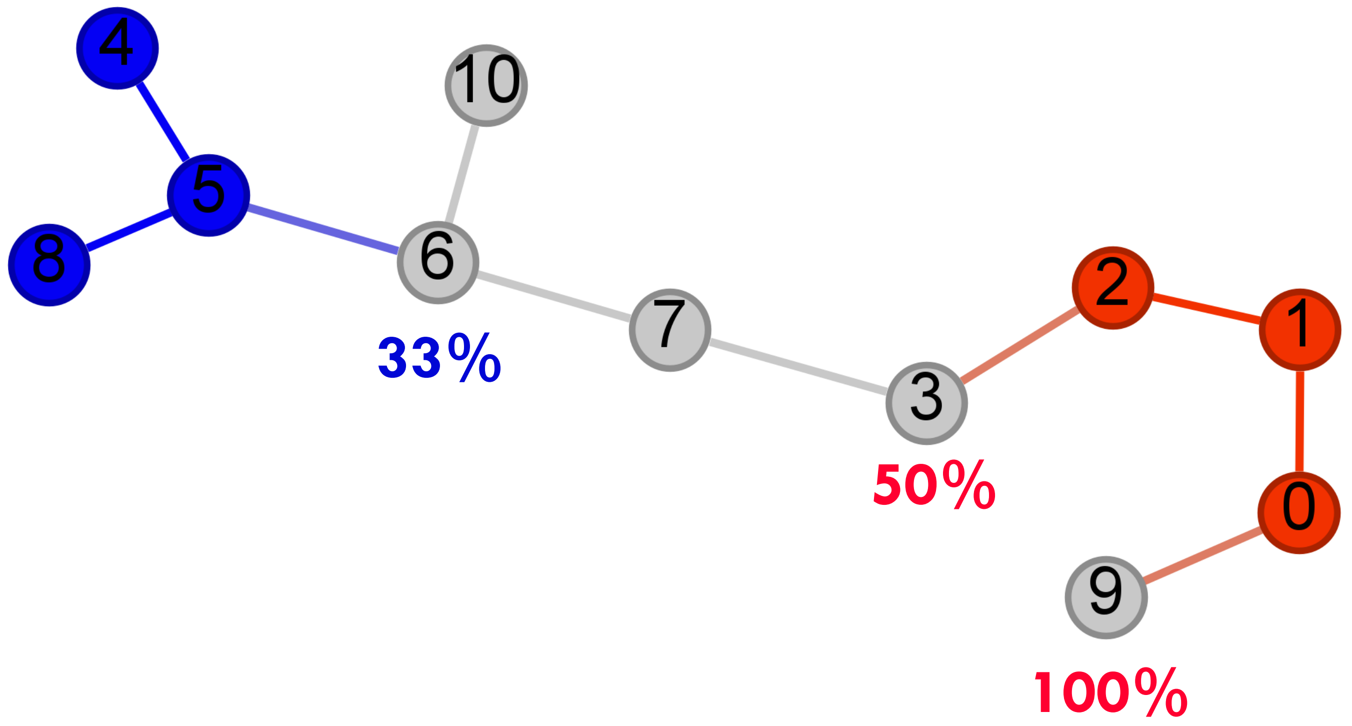
STEP 3



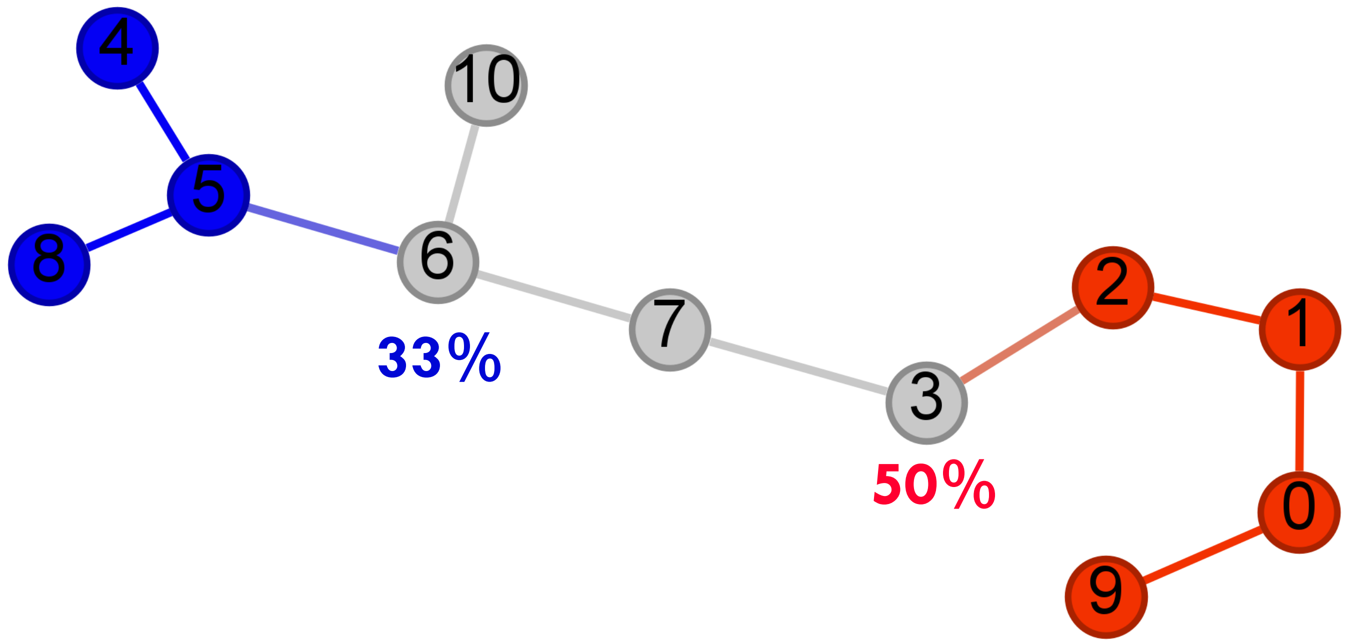
STEP 4



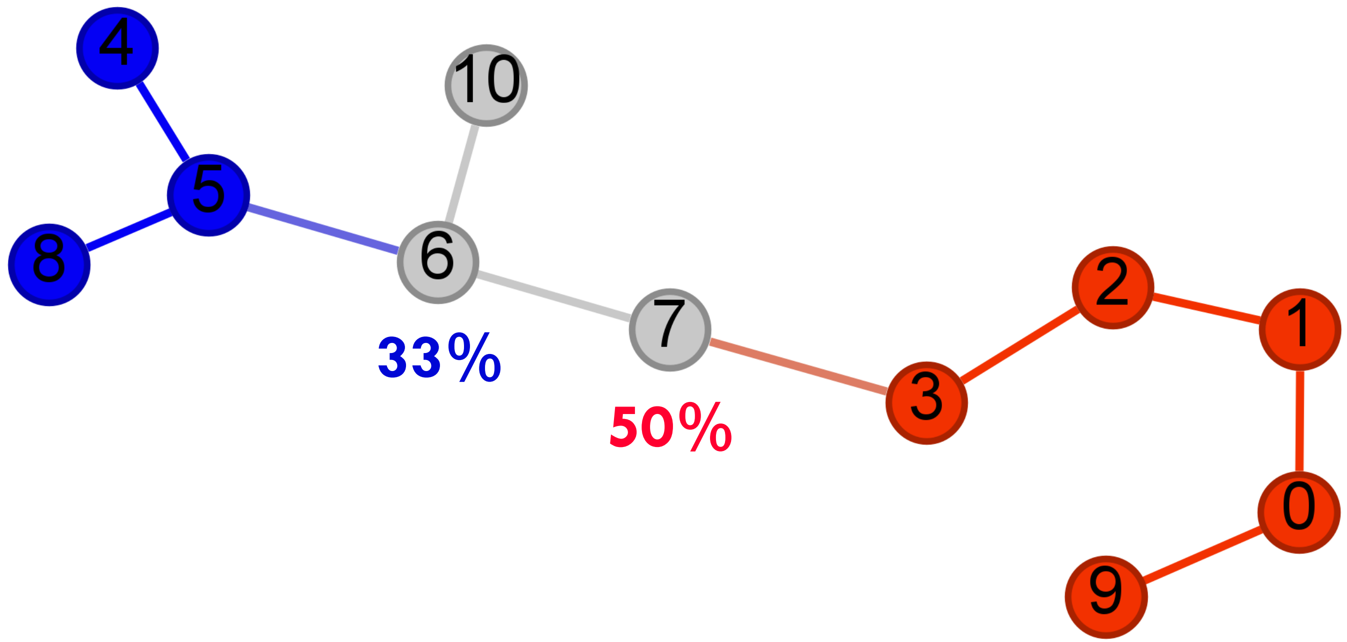
STEP 5



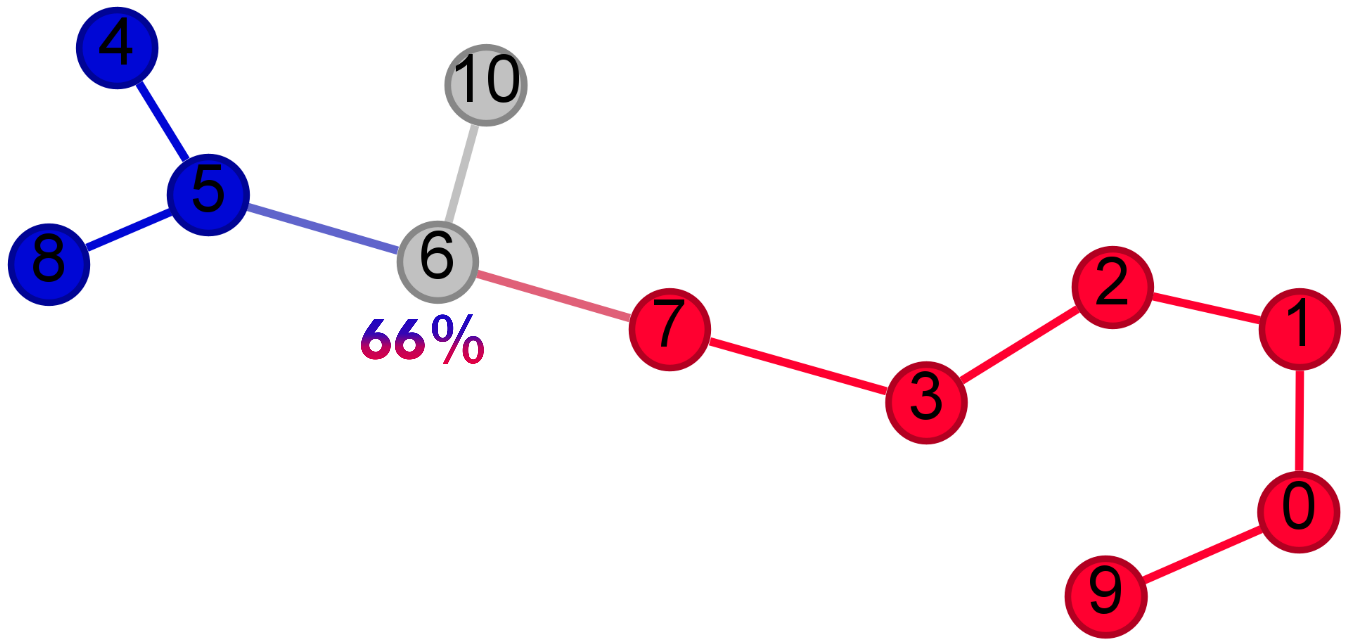
STEP 6



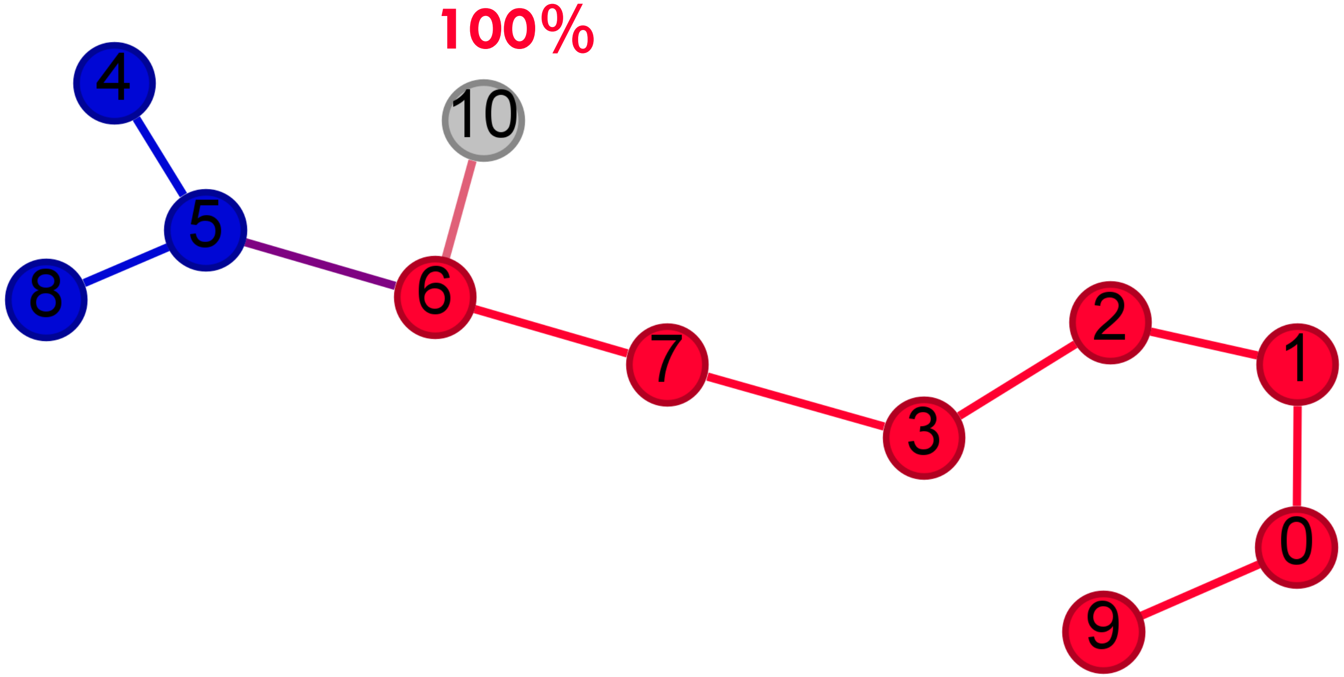
STEP 7



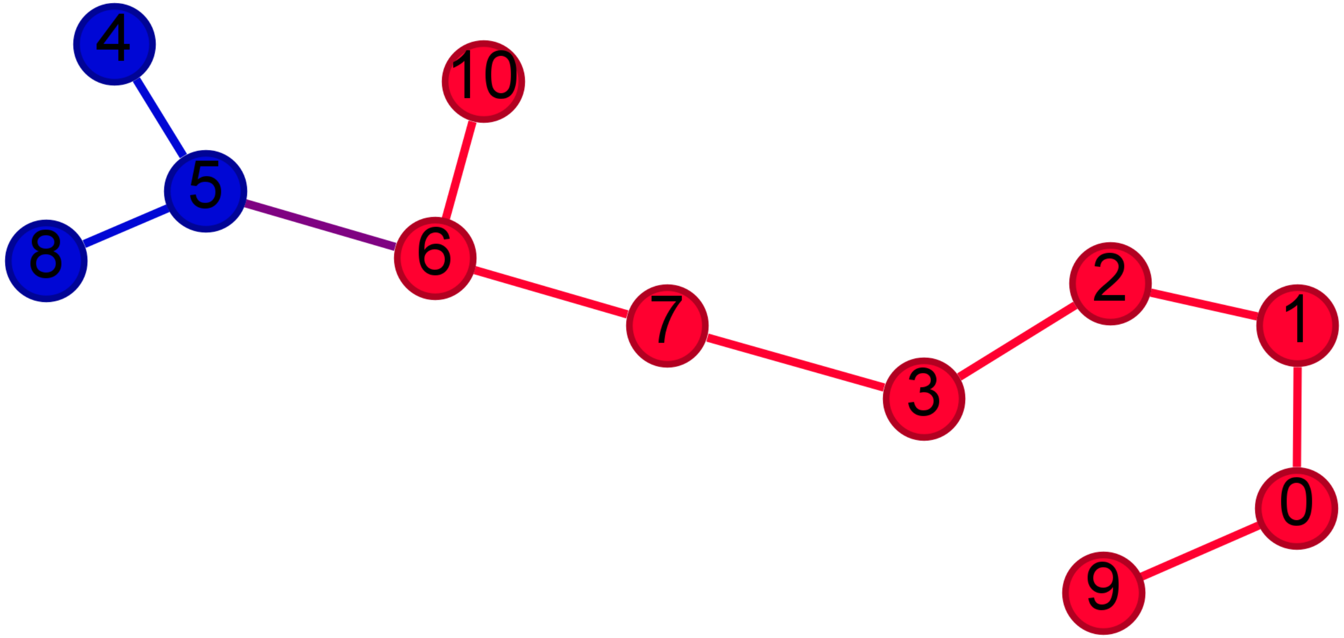
STEP 8



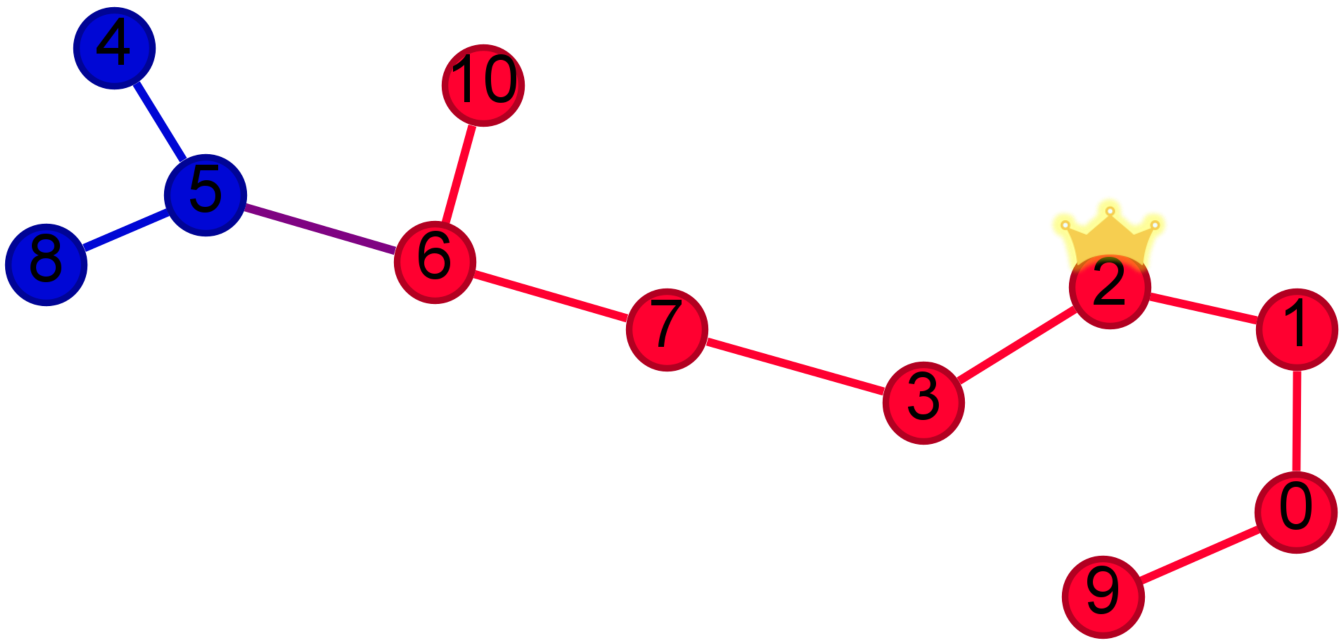
STEP 9



STEP 10

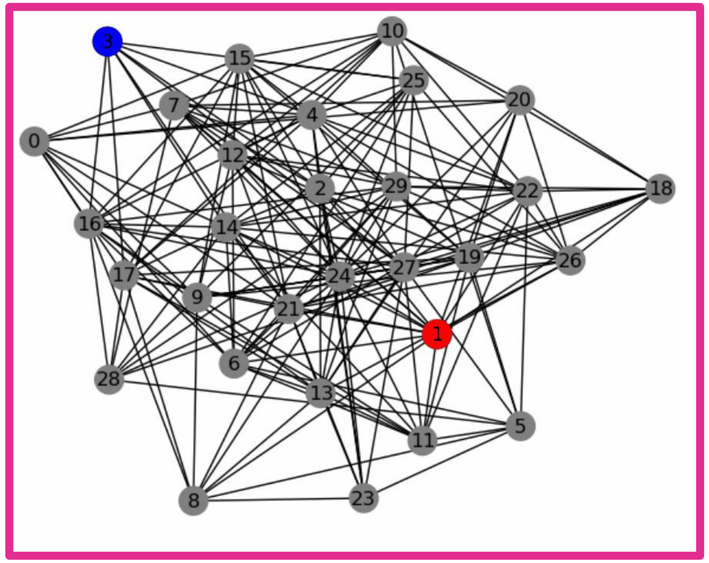


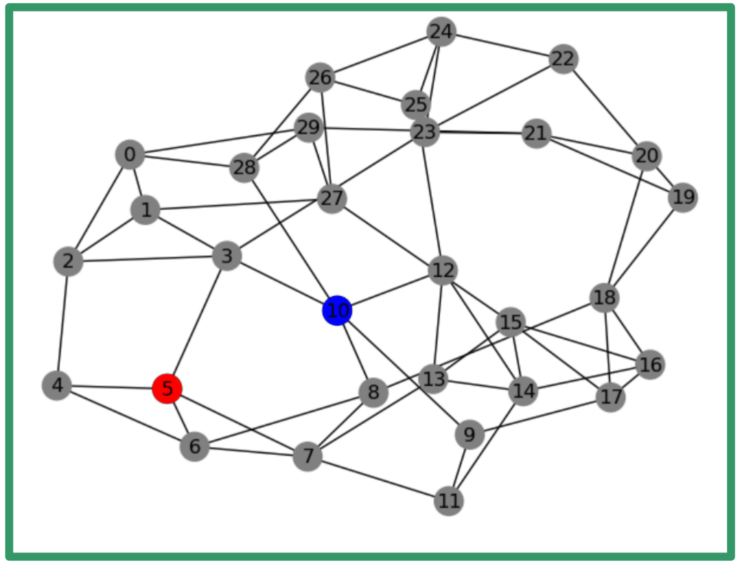
RED WINS!

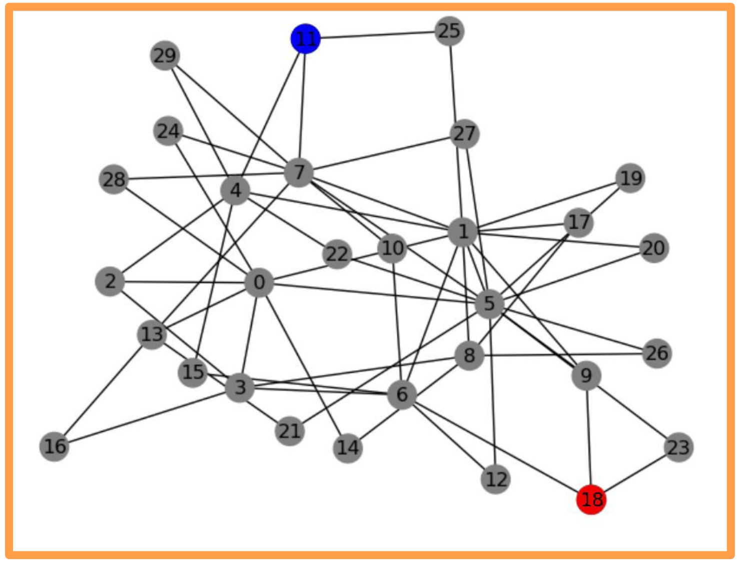


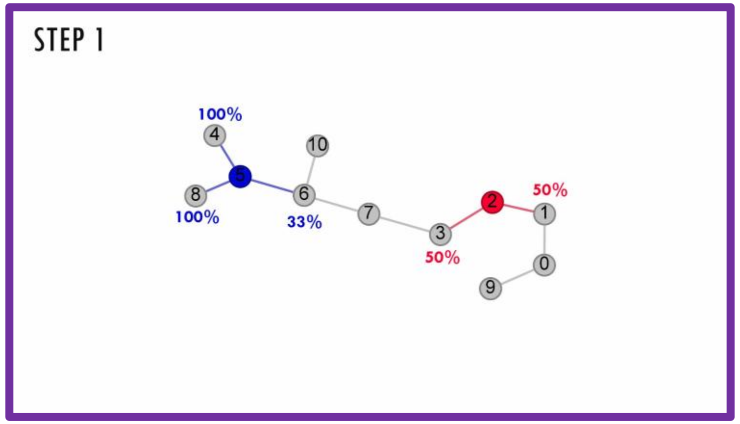
ANY QUESTION ON HOW THE ALG WORKS?

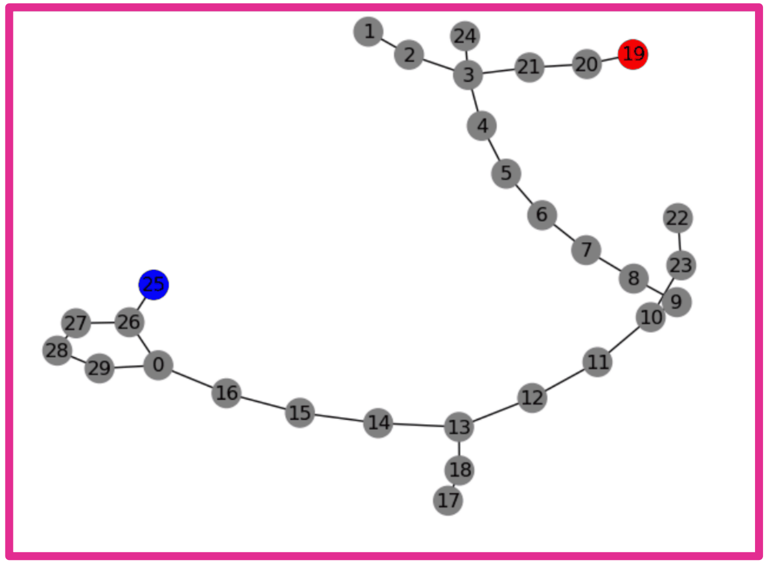
(WE GO OVER VARIANTS AT THE END)

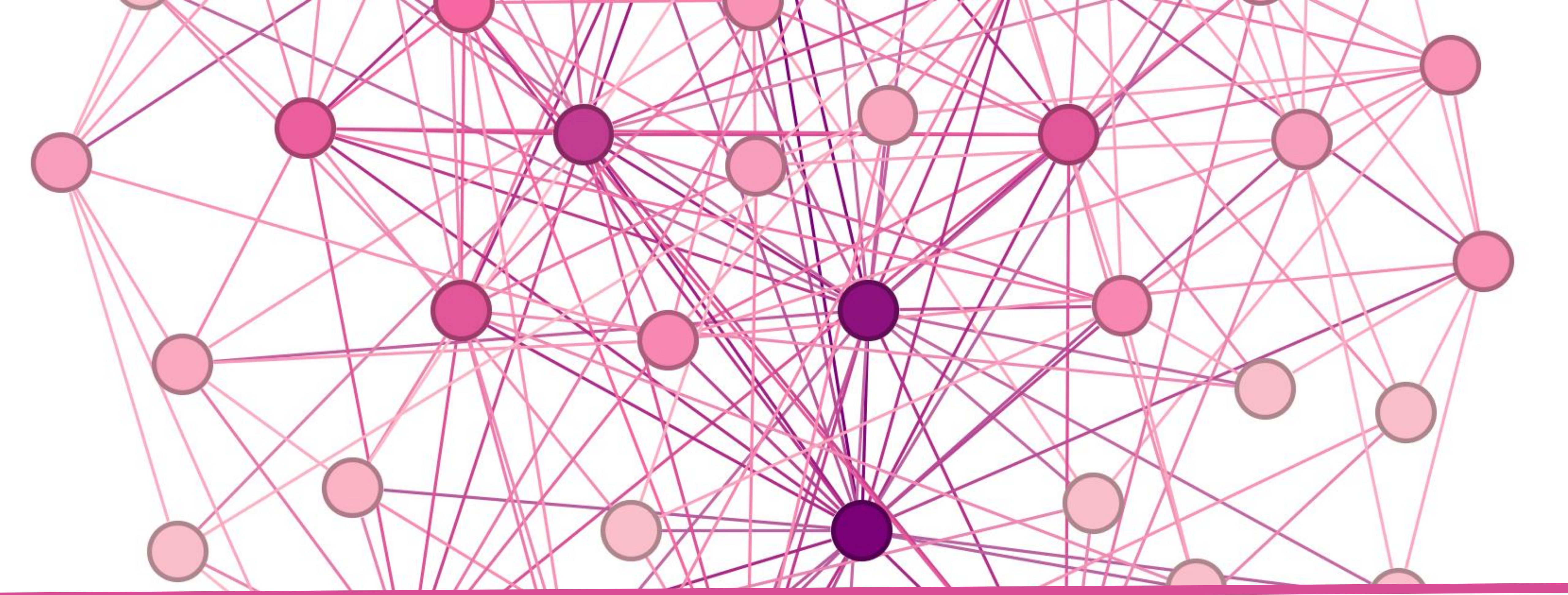












RESULTS

WHAT TEST ARE WE RUNNING?

- We run the algorithm 100 times on 4 varied graph generations, for each of the 3 models (Meaning the algorithm was run **147,000** total times across **1,200** unique graphs with a total of **60,000** nodes)
- Every combination of **Selected Nodes** is run on each graph
  - In the event of a tie, all combinations are also run (ie if 3 nodes tie for highest degree, and 2 nodes tie for lowest degree, all possible matchings are tested). These are then normalized
- Results are then calculated based on the percentage of tests won

Selected Nodes

- Eccentricity
  - Highest
  - Lowest
  - Approximately middle
- Degree
  - Highest
  - Lowest
  - Approximately middle
- Clustering coefficient
  - Highest
  - Lowest
  - Approximately middle
- Randomly chosen

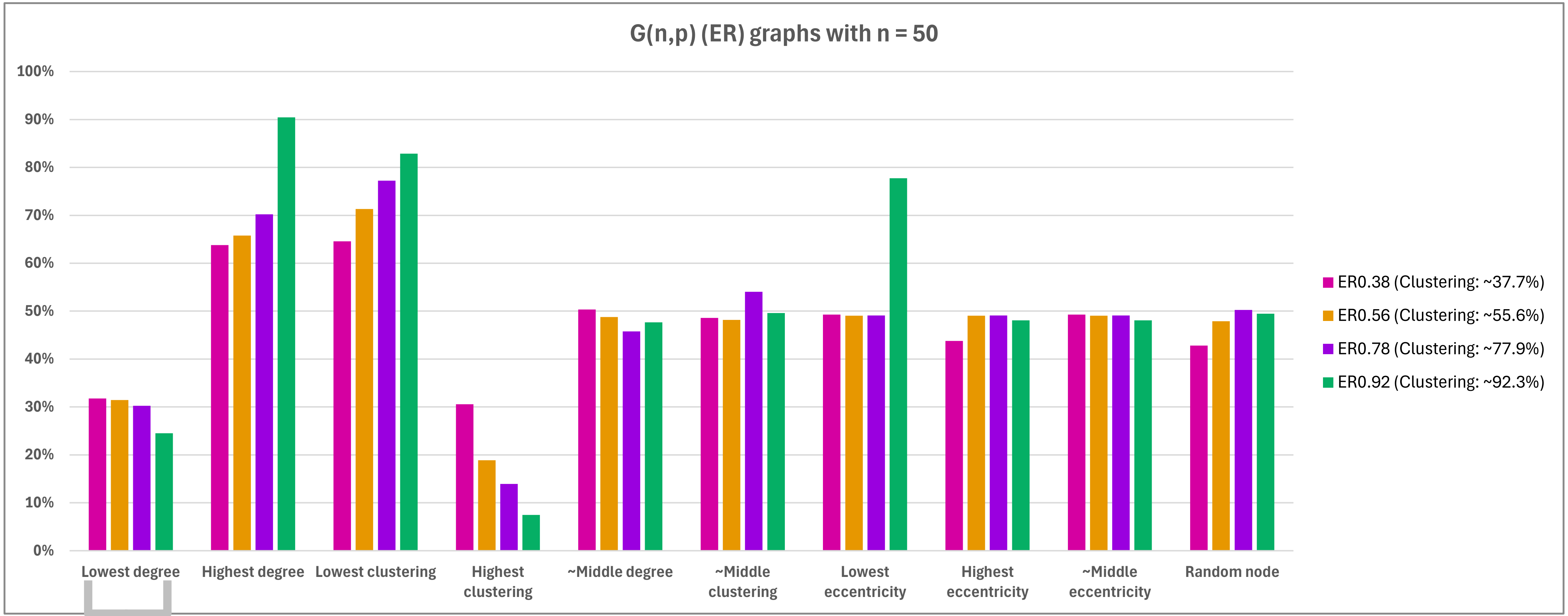
ON WHICH GRAPHS?

- All graphs are fully connected and have 50 nodes
- Hundreds of graphs of various parameters across the 3 models were generated and analyzed
- 4 benchmark graphs for each model were selected, based on having approximately the same average clustering
- Clustering and number of nodes were chosen for their impact on graph structure
- We do this to have some standardization between graphs allowing us to compare tests on different generations models in a meaningful way

WS, 50, X, 0.2	X = 10
Trial 1	0.35553
Trial 2	0.4091
Trial 3	0.39647
Trial 4	0.35668
Trial 5	0.39307
Trial 6	0.38821
Trial 7	0.40394
Trial 8	0.38189
Trial 9	0.38281
Trial 10	0.32851
Average	0.37962
Rounded average	0.382

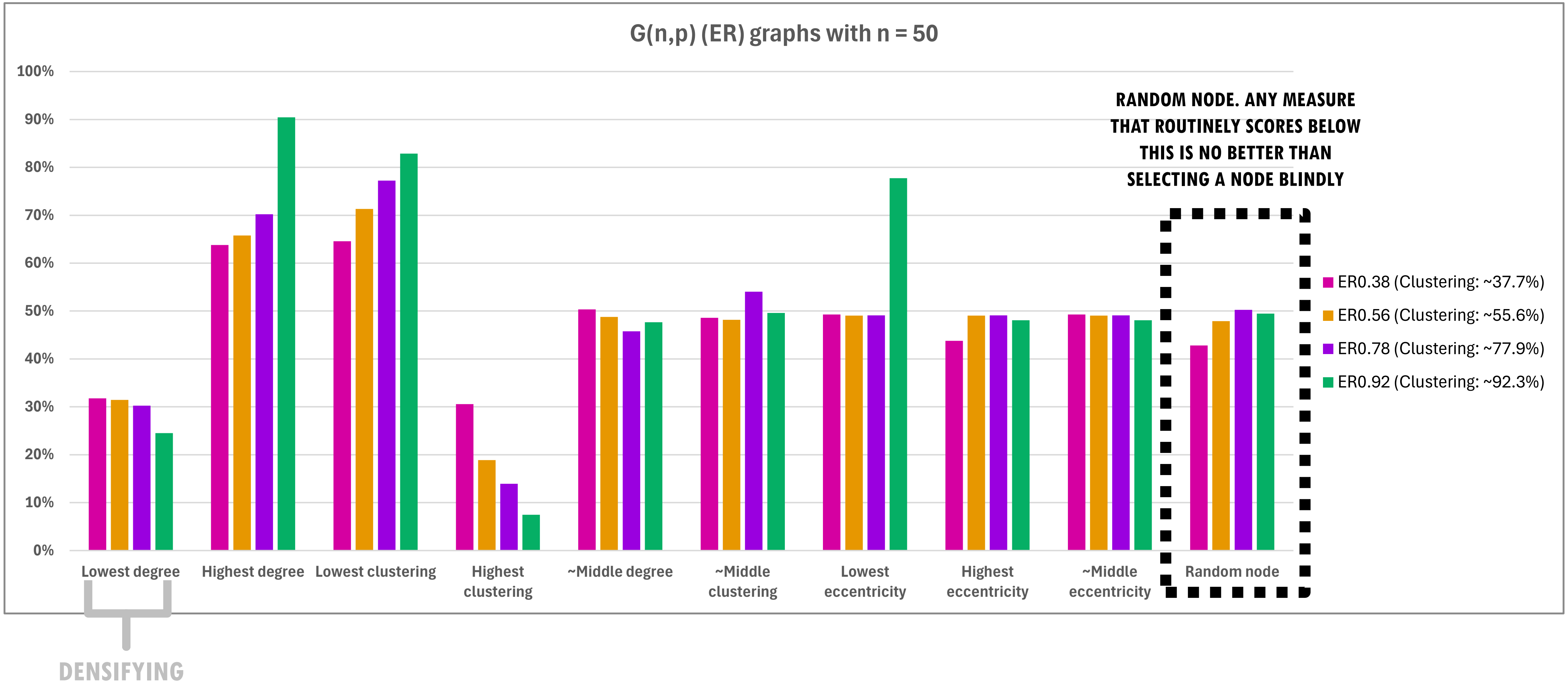
Chosen params	~ Clustering	In %
WS(50, 10 , 0.2)	0.382	38.2%
ER(50, 0.38)	0.377	37.7%
PA(50,8)	0.384	38.4%

ERDŐS-RÉNYI-GILBERT

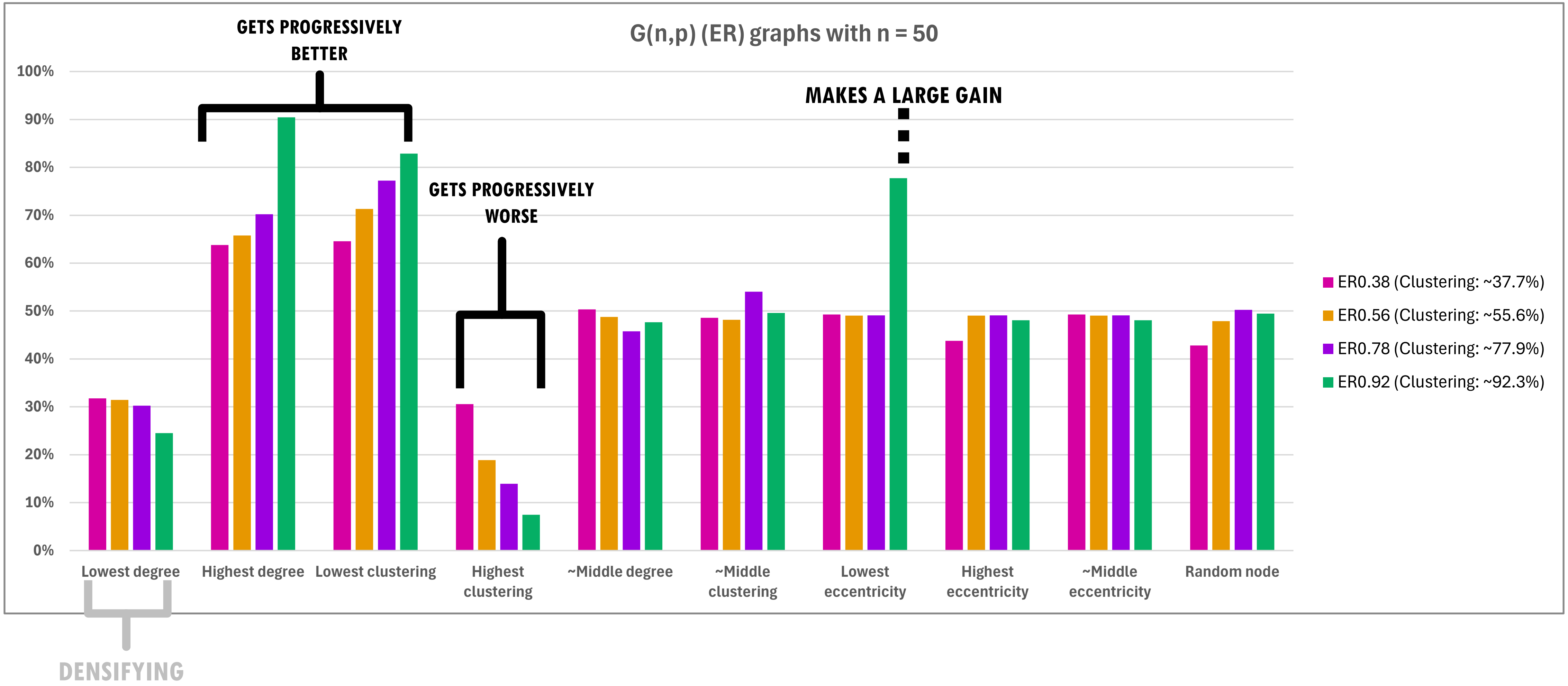


DENSIFYING

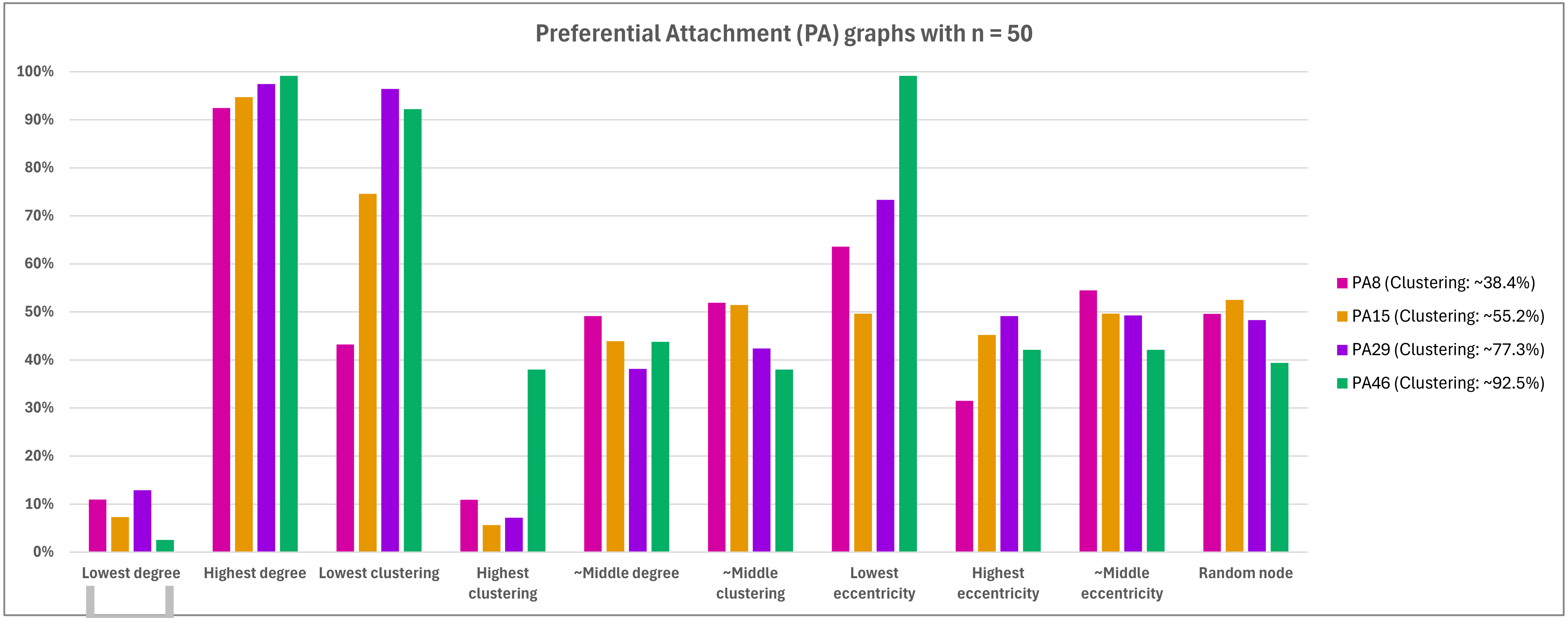
ERDŐS-RÉNYI-GILBERT



ERDŐS-RÉNYI-GILBERT



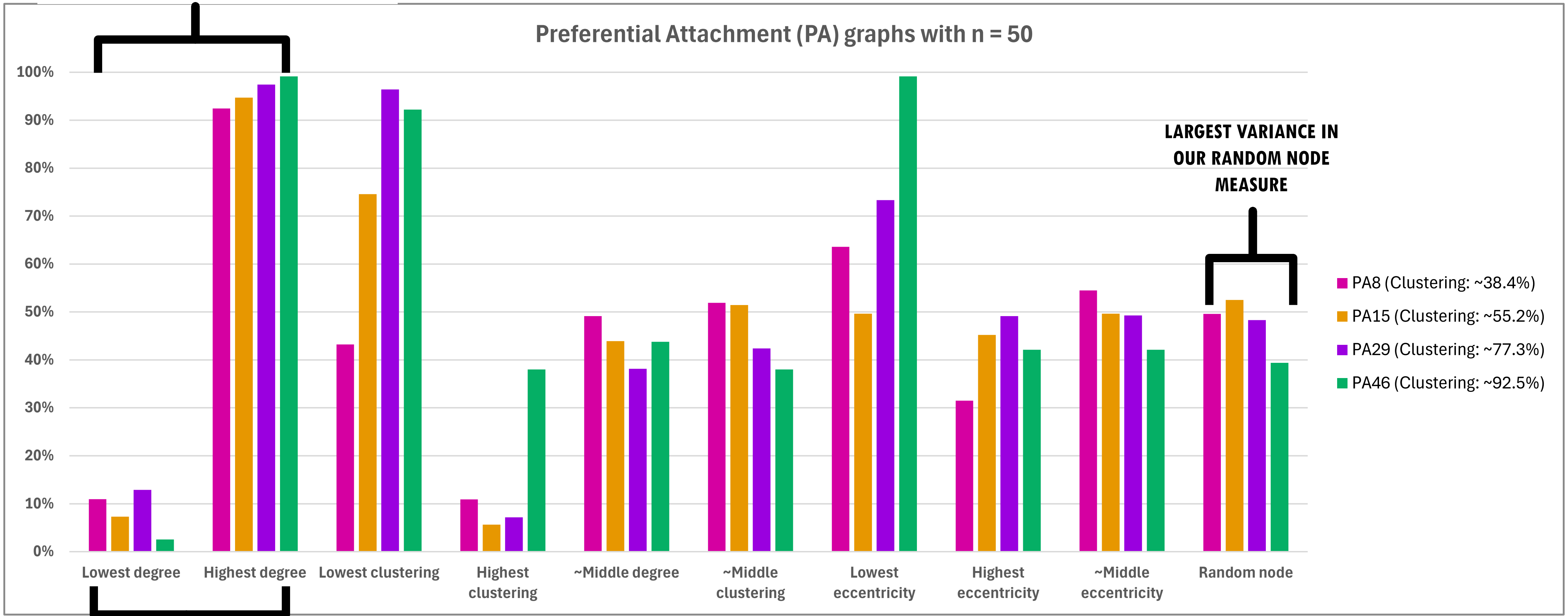
BARABÁSI—ALBERT



DENSIFYING

BARABÁSI–ALBERT

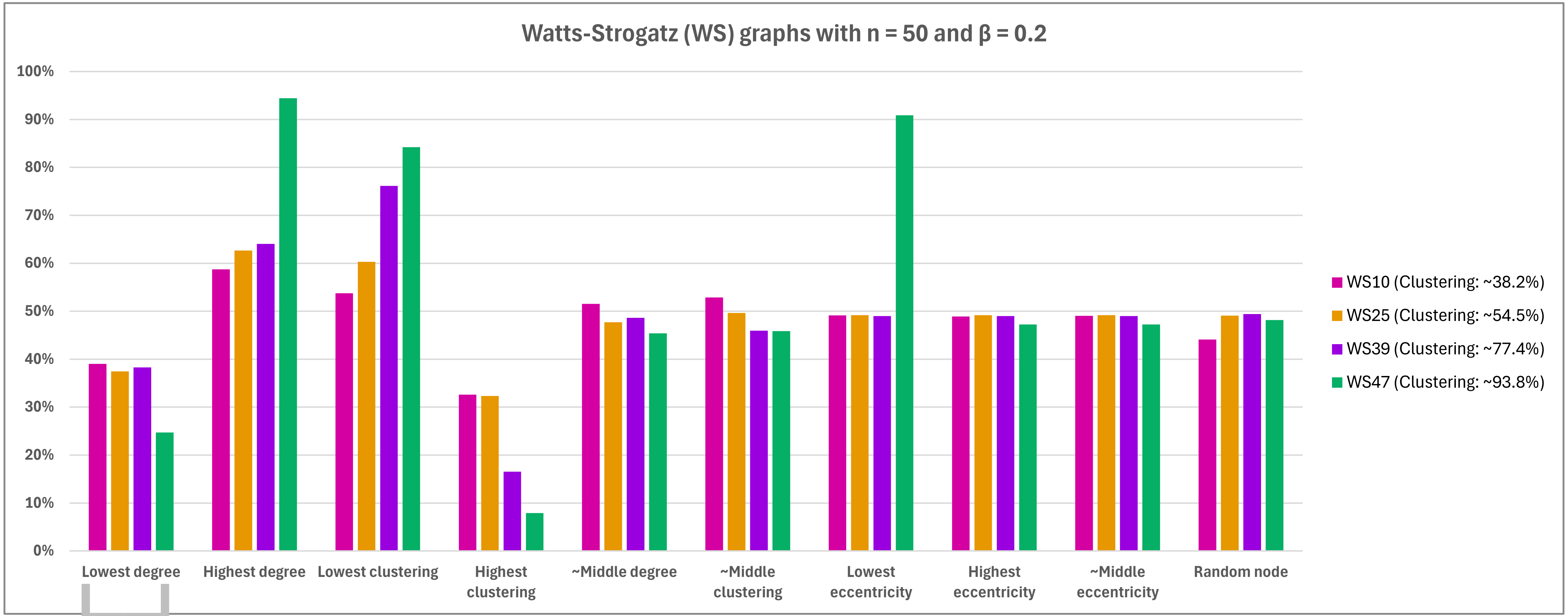
LARGEST GAP BETWEEN TWO MEASURES



LARGEST VARIANCE IN OUR RANDOM NODE MEASURE

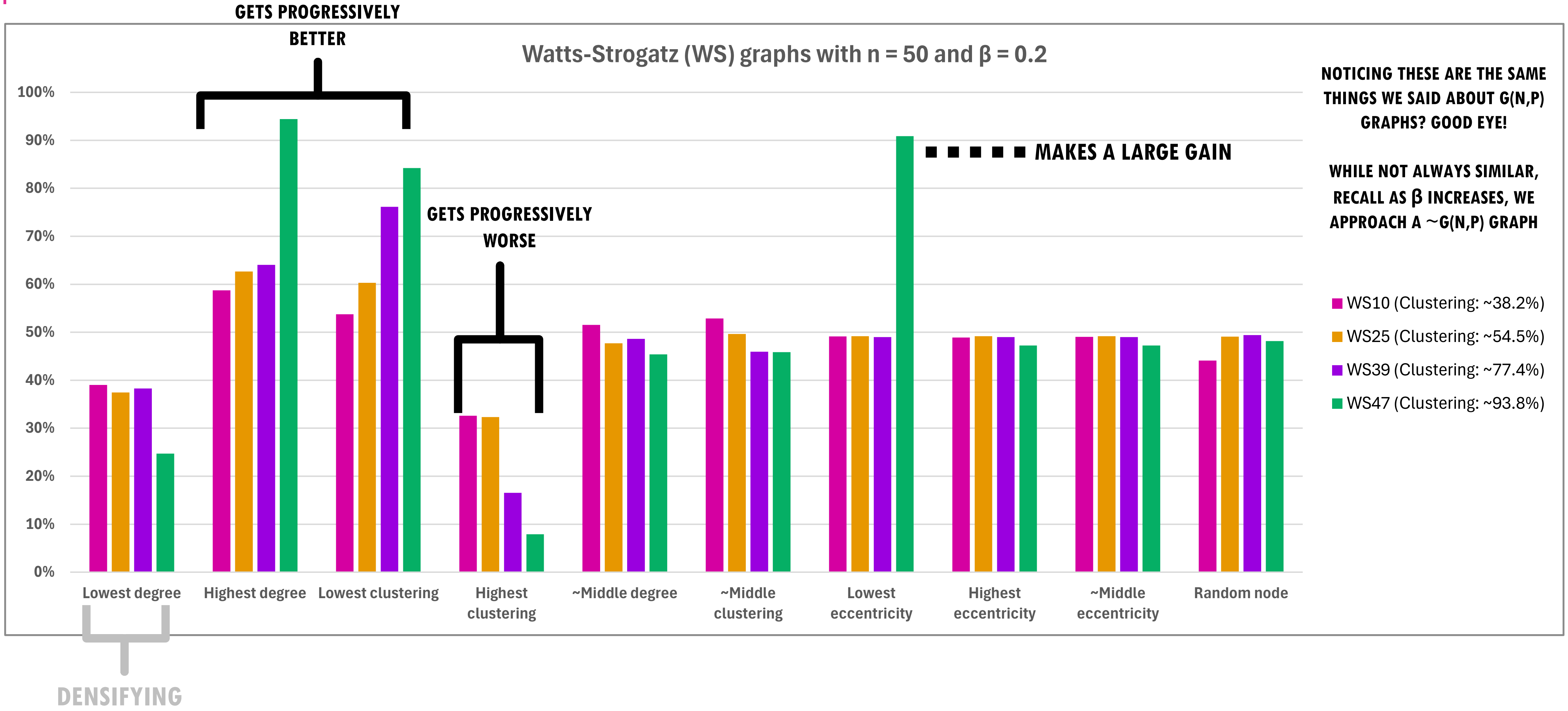
HIGHEST AND LOWEST SUCCESS RATE FOR A MEASURE OVER ALL GENERATION MODELS

WATTS–STROGATZ



DENSIFYING

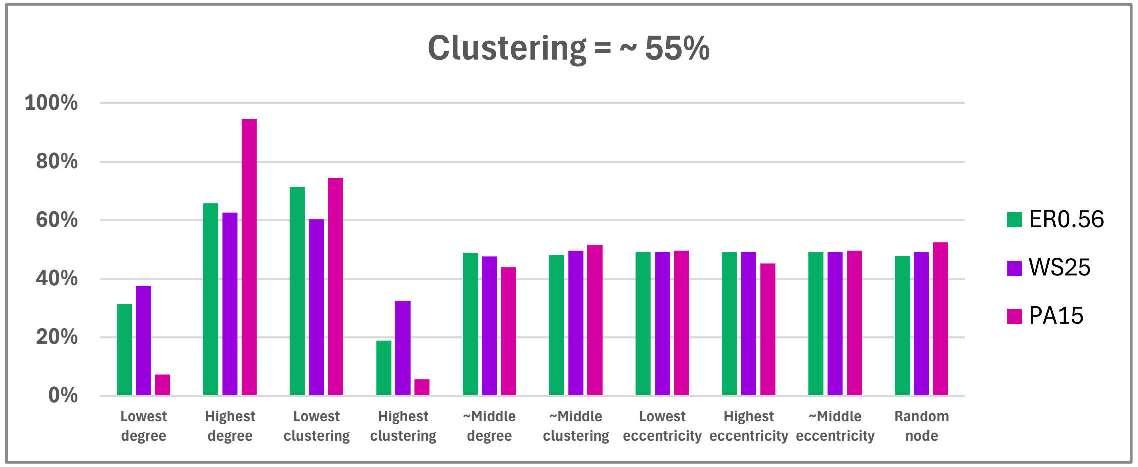
WATTS—STROGATZ

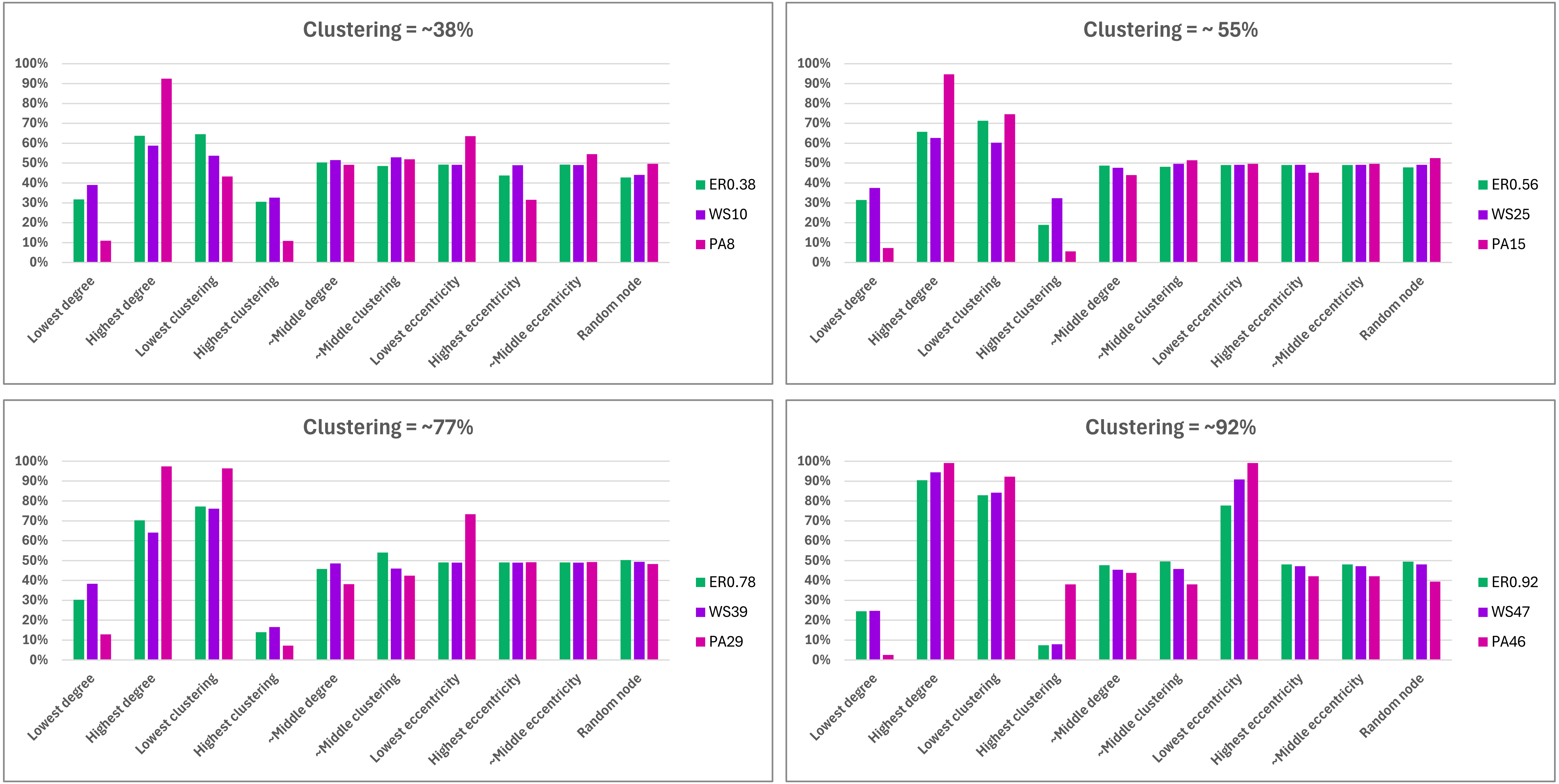


NOTICING THESE ARE THE SAME THINGS WE SAID ABOUT G(N,P) GRAPHS? GOOD EYE!

WHILE NOT ALWAYS SIMILAR, RECALL AS β INCREASES, WE APPROACH A ~G(N,P) GRAPH

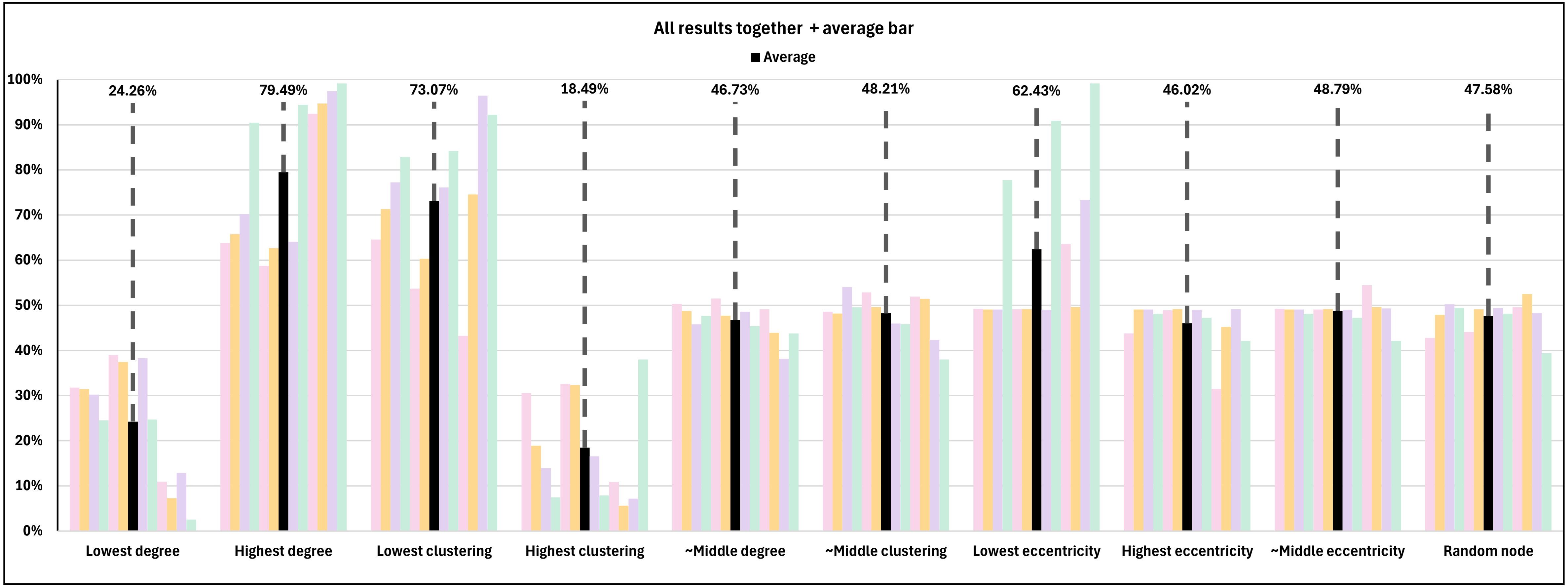
BY CLUSTERING

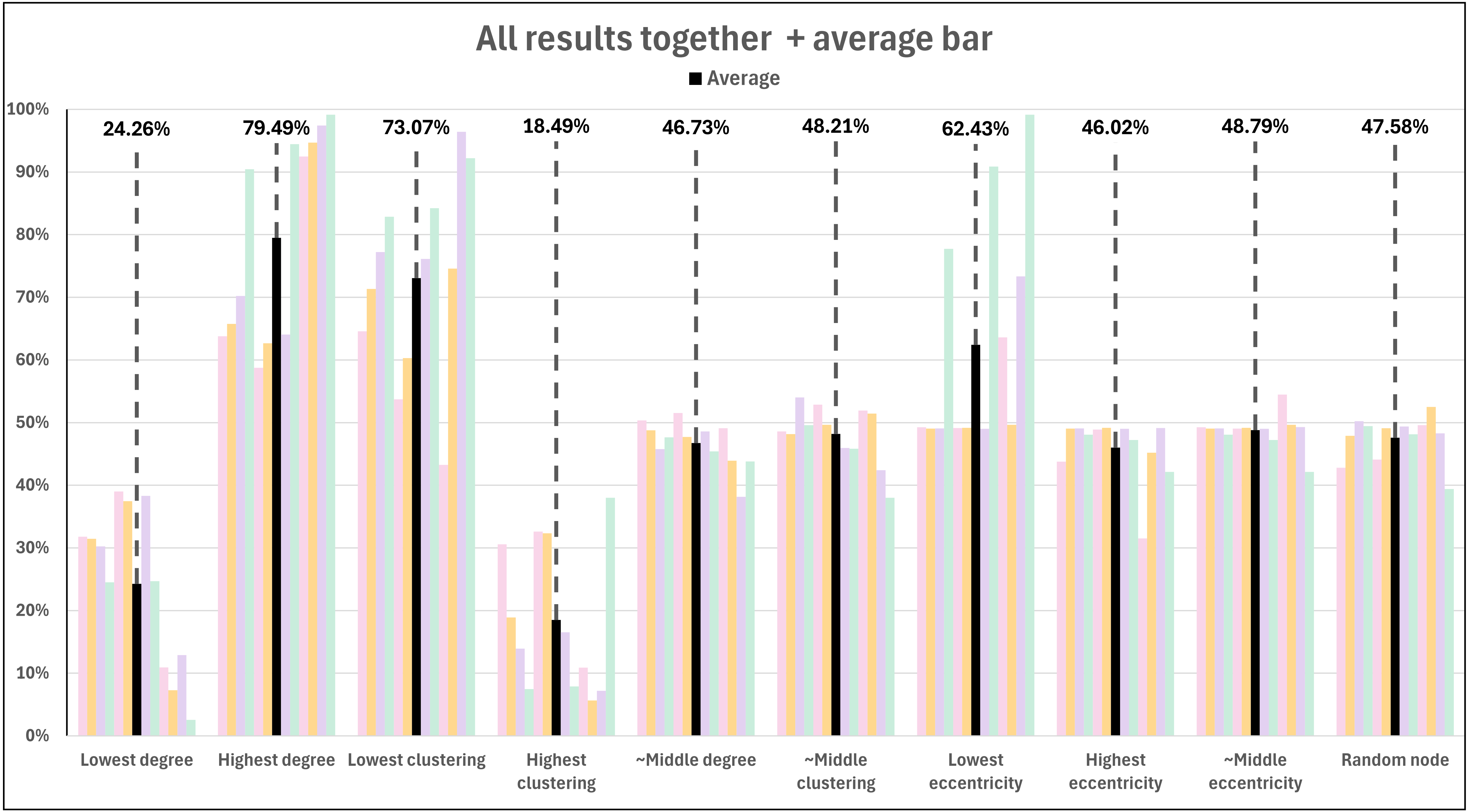


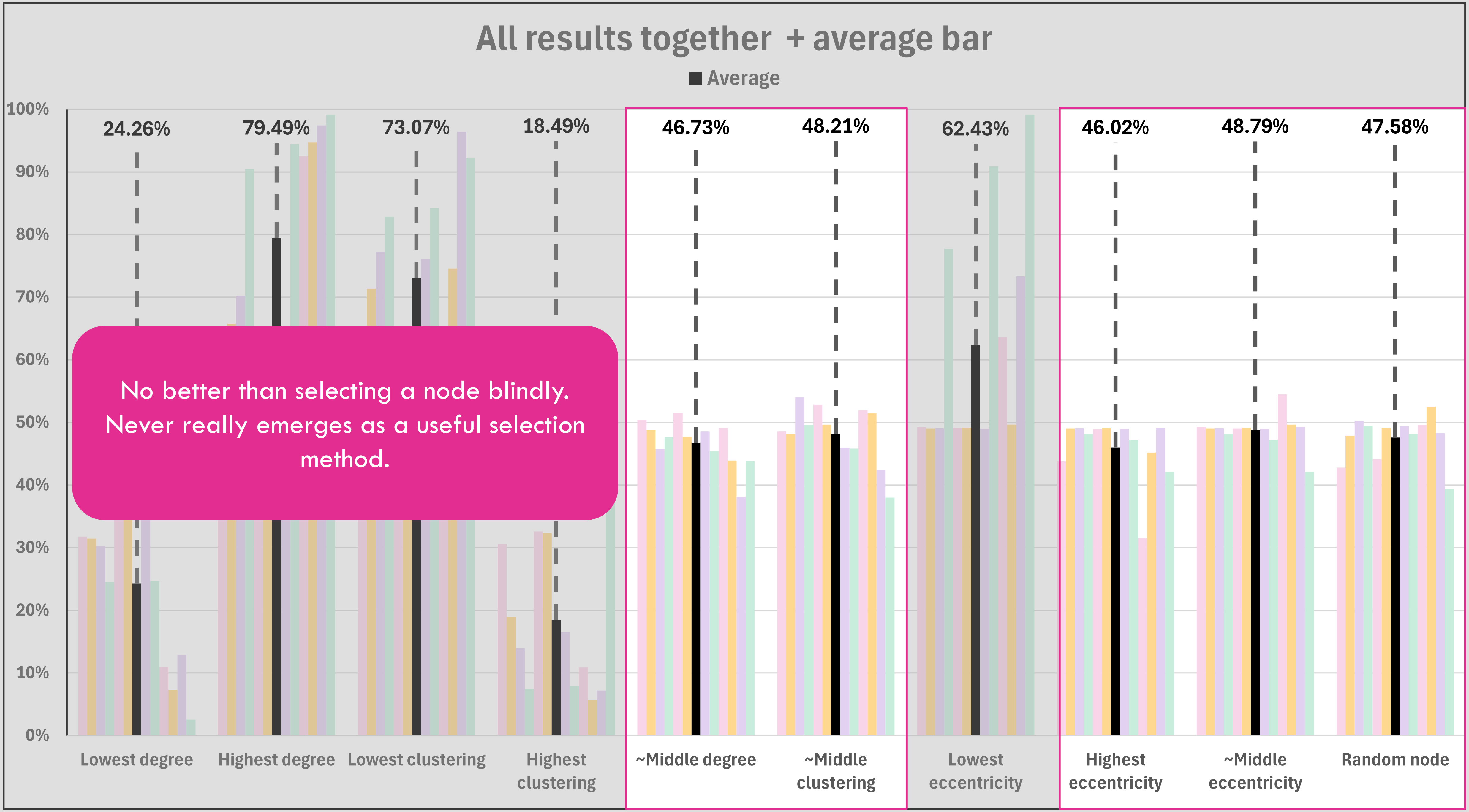


Clustering = ~38%
ER0.38
WS10
PA8
Clustering = ~ 55%
ER0.56
WS25
PA15
Clustering = ~77%
ER0.78
WS39
PA29
Clustering = ~92%
ER0.92
WS47
PA46
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
Lowest degree
Highest degree
Lowest clustering
Highest clustering
~Middle degree
~Middle clustering
Lowest eccentricity
Highest eccentricity
~Middle eccentricity
Random node

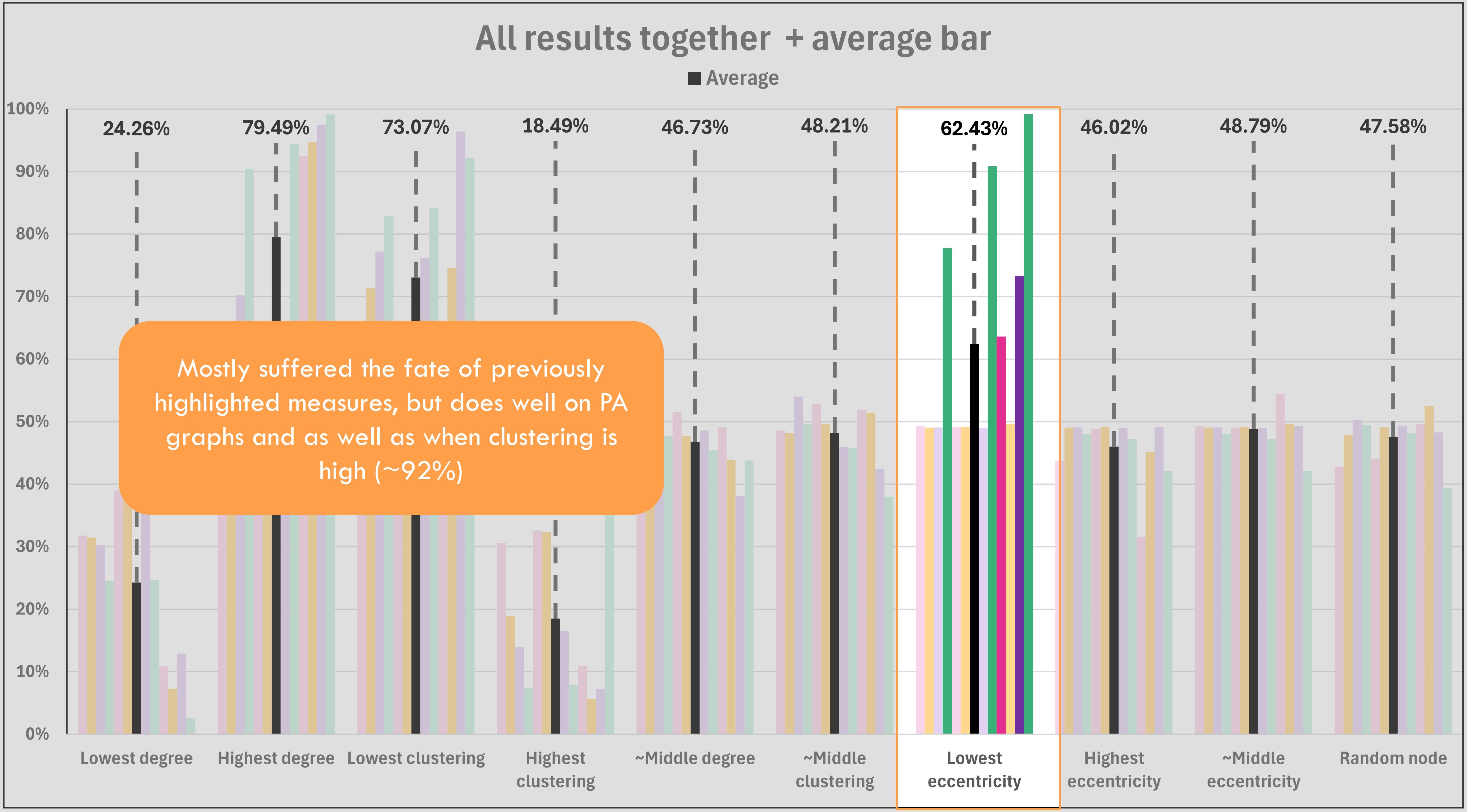
EVERYTHING + AVERAGE



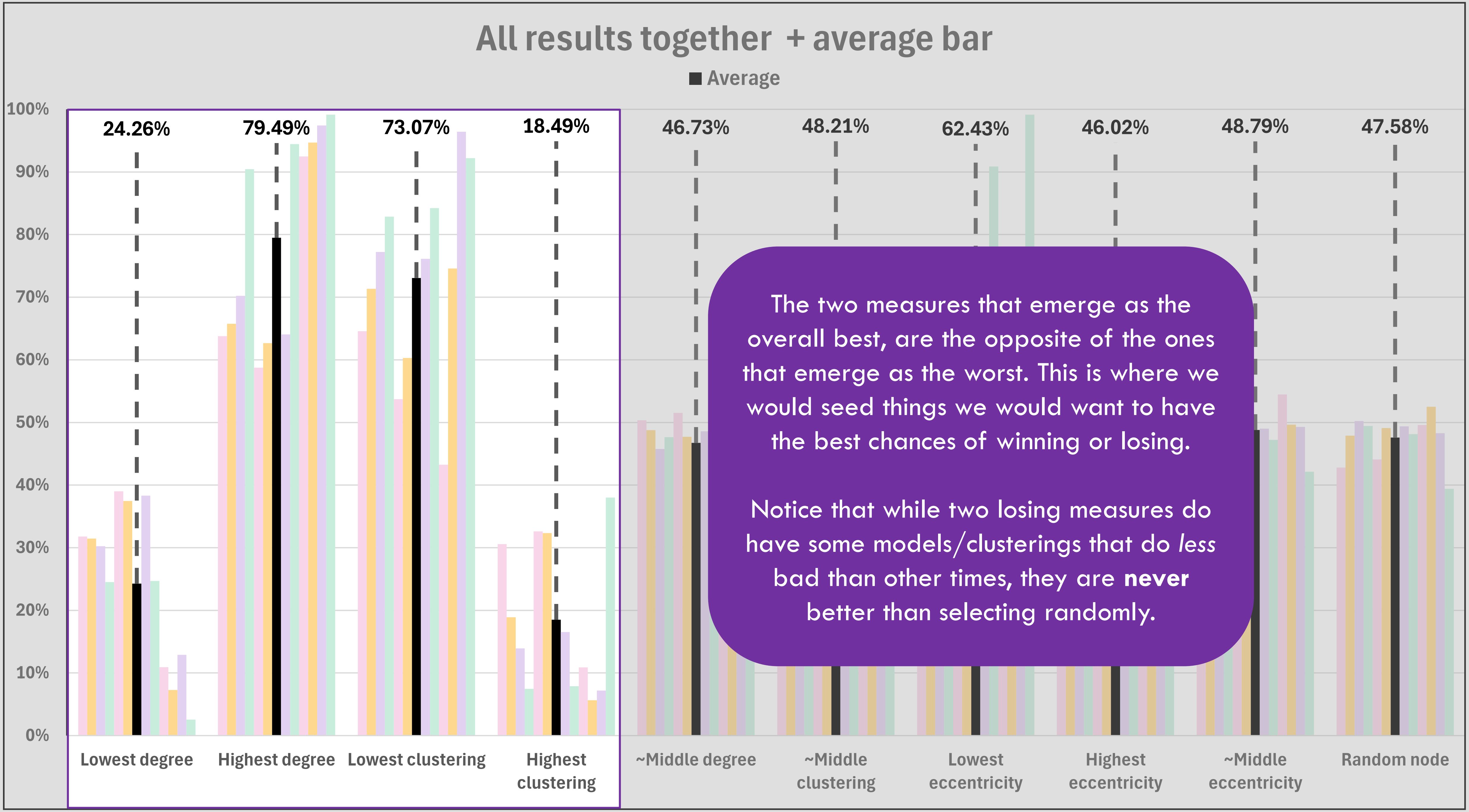

All results together + average bar
Average
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
24.26%
79.49%
73.07%
18.49%
46.73%
48.21%
62.43%
46.02%
48.79%
47.58%
Lowest degree
Highest degree
Lowest clustering
Highest clustering
~Middle degree
~Middle clustering
Lowest eccentricity
Highest eccentricity
~Middle eccentricity
Random node

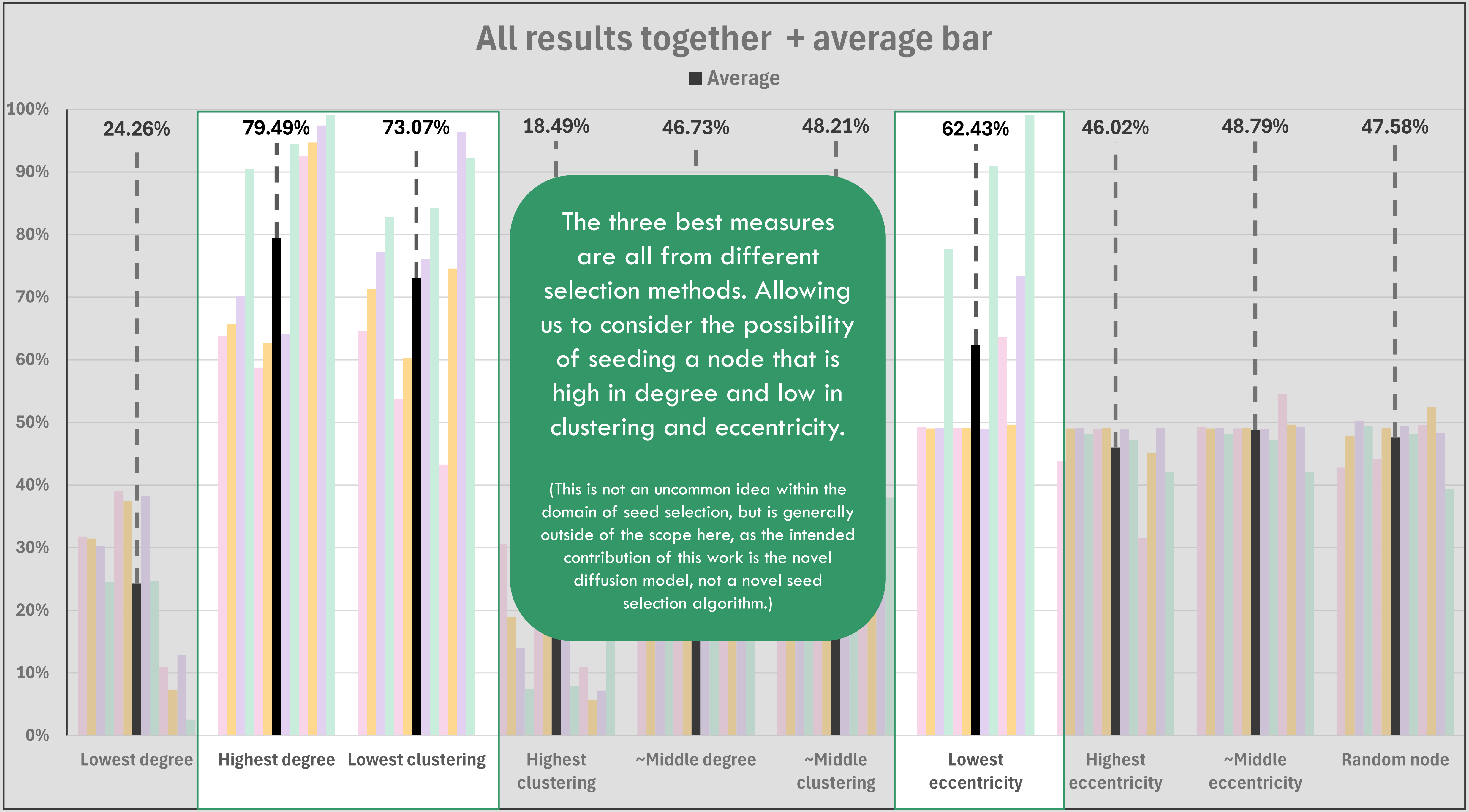


No better than selecting a node blindly. Never really emerges as a useful selection method.



Mostly suffered the fate of previously highlighted measures, but does well on PA graphs and as well as when clustering is high (~92%)

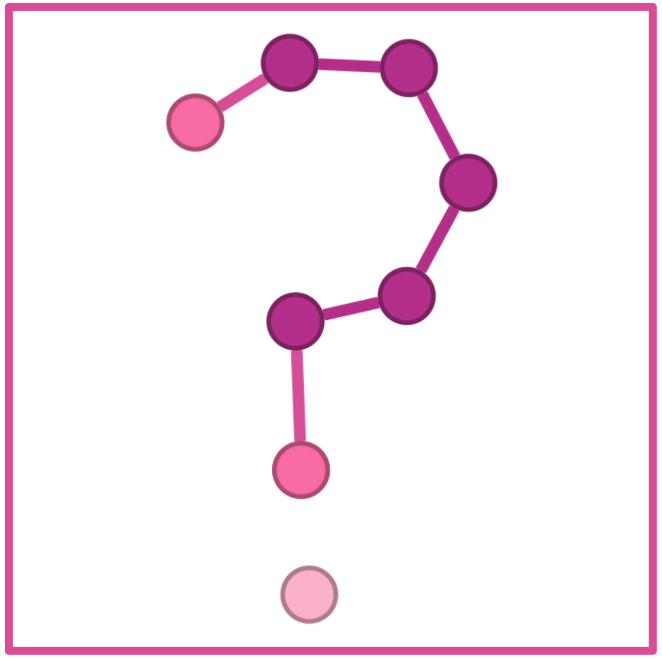

All results together + average bar
Average
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
24.26%
79.49%
73.07%
18.49%
46.73%
48.21%
62.43%
46.02%
48.79%
47.58%
Lowest degree
Highest degree
Lowest clustering
Highest clustering
~Middle degree
~Middle clustering
Lowest eccentricity
Highest eccentricity
~Middle eccentricity
Random node
The two measures that emerge as the overall best, are the opposite of the ones that emerge as the worst. This is where we would seed things we would want to have the best chances of winning or losing.
Notice that while two losing measures do have some models/clusterings that do *less* bad than other times, they are **never** better than selecting randomly.

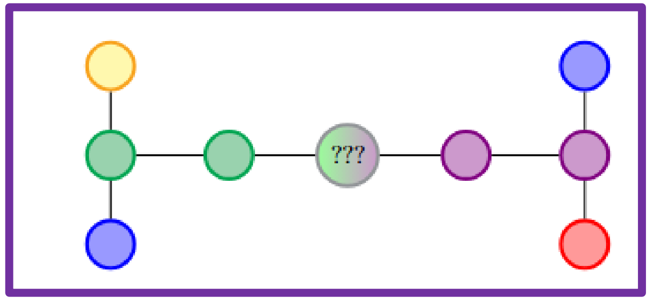

All results together + average bar
Average
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
24.26%
79.49%
73.07%
18.49%
46.73%
48.21%
62.43%
46.02%
48.79%
47.58%
The three best measures are all from different selection methods. Allowing us to consider the possibility of seeding a node that is high in degree and low in clustering and eccentricity.
(This is not an uncommon idea within the domain of seed selection, but is generally outside of the scope here, as the intended contribution of this work is the novel diffusion model, not a novel seed selection algorithm.)
Lowest degree
Highest degree
Lowest clustering
Highest clustering
~Middle degree
~Middle clustering
Lowest eccentricity
Highest eccentricity
~Middle eccentricity
Random node

CONCLUSION

FUTURE WORK: VARIANTS

- We have lots!!
- Here are a select few discussed in the work:
  - Seeding more than 2 colours
  - Expanding seed set to be larger than 1 per colour
  - Allowing stalls to be solved with a probabilistic coin flip
  - Limiting the nodes that can be coloured to only those over a threshold (modified Linear Threshold Model, see Kempe 2003)
  - Have stalled nodes remain uncoloured, either failing or simply not fully colouring graph
  - Allowing nodes stalled between red and blue to become purple (hybrid colours)
  - Allowing nodes to revaluate their colour choice (especially if we allow purple to spawn on temporarily stalled nodes)
  - Have nodes exist as percentages of the seeded colours (i.e. 25% red, 25% yellow, 50% blue)





THANK YOU TO...

Dr. Stoica

Dr. Watts

Dr. Meger

Dr. Gilbert

Hermie

SIAM/CIAMS

Dr. Kempe

Dr. Albert

Dr. Strogatz

Queen's University

Dr. Rényi

Dr. Erdős

NetworkX

Dr. Barabási

WLU

Dr. Chung

...AND MANY MORE!

THANK YOU

FOR LISTENING!

If you have any questions, feel free ask away and I will answer them to the best of my ability!

Relevant files available at:
github.com/meganbryson/ProportionBasedAlg
Visit meganbryson.ca if you want to recuse me from unemployment

Special thank you to my friends and family for their continued support!