

# HUMAN GENE

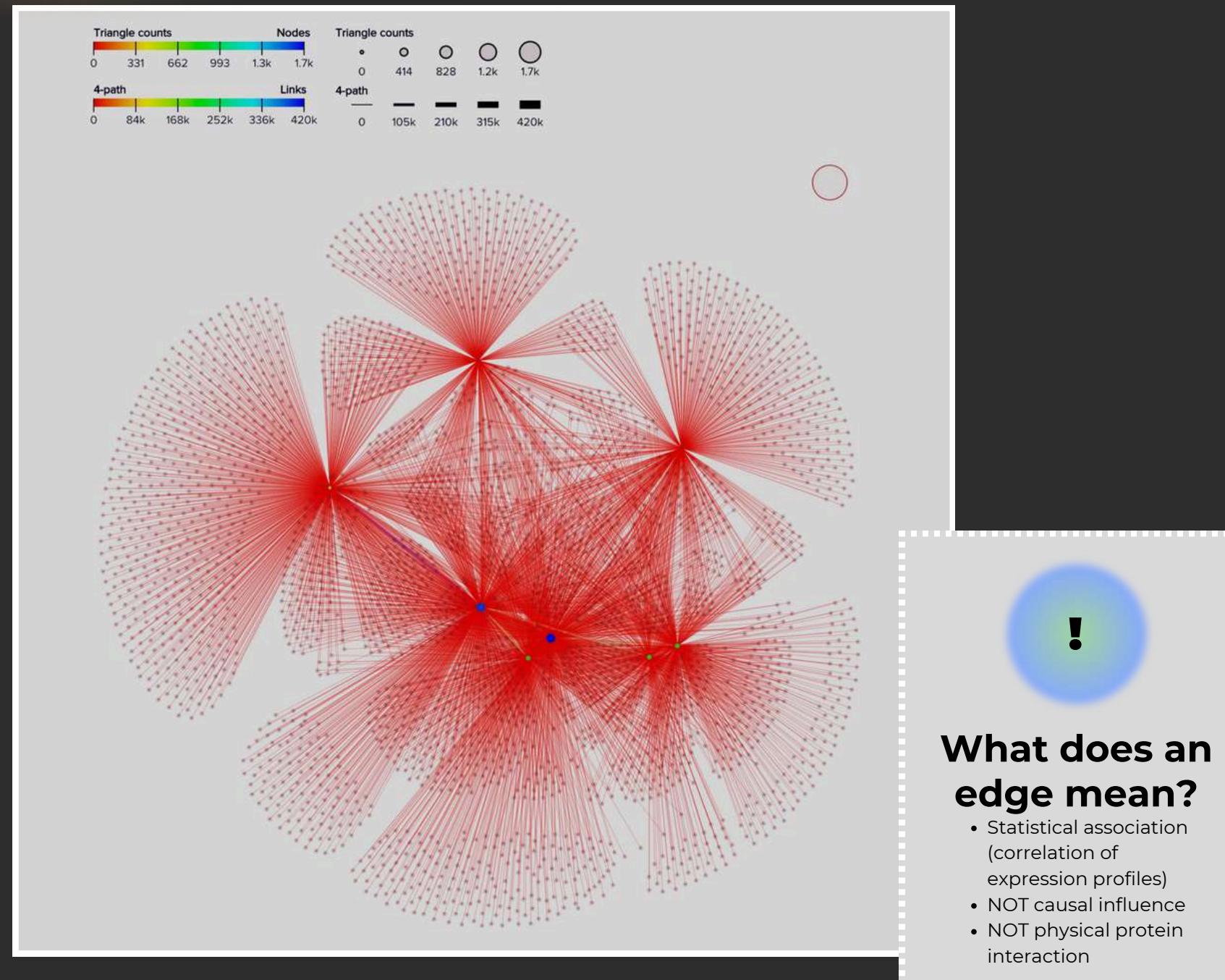
CO-EXPRESSION NETWORK  
ANALYSIS

Presented by

Anastasiia Mazur

# NETWORK DESCRIPTION

## PROBLEM STATEMENT



Characteristic	Value
Source	Network Repository (Human gene2)
Type	Co-expression network
Nodes	14,340 genes
Edges	9,027,024 connections
Density	0.092 ( $\approx$ 9% of all pairs)

# REPRESENTATION

**Undirected**

Correlation is symmetric:  $\text{cor}(A,B) = \text{cor}(B,A)$

**Unweighted\***

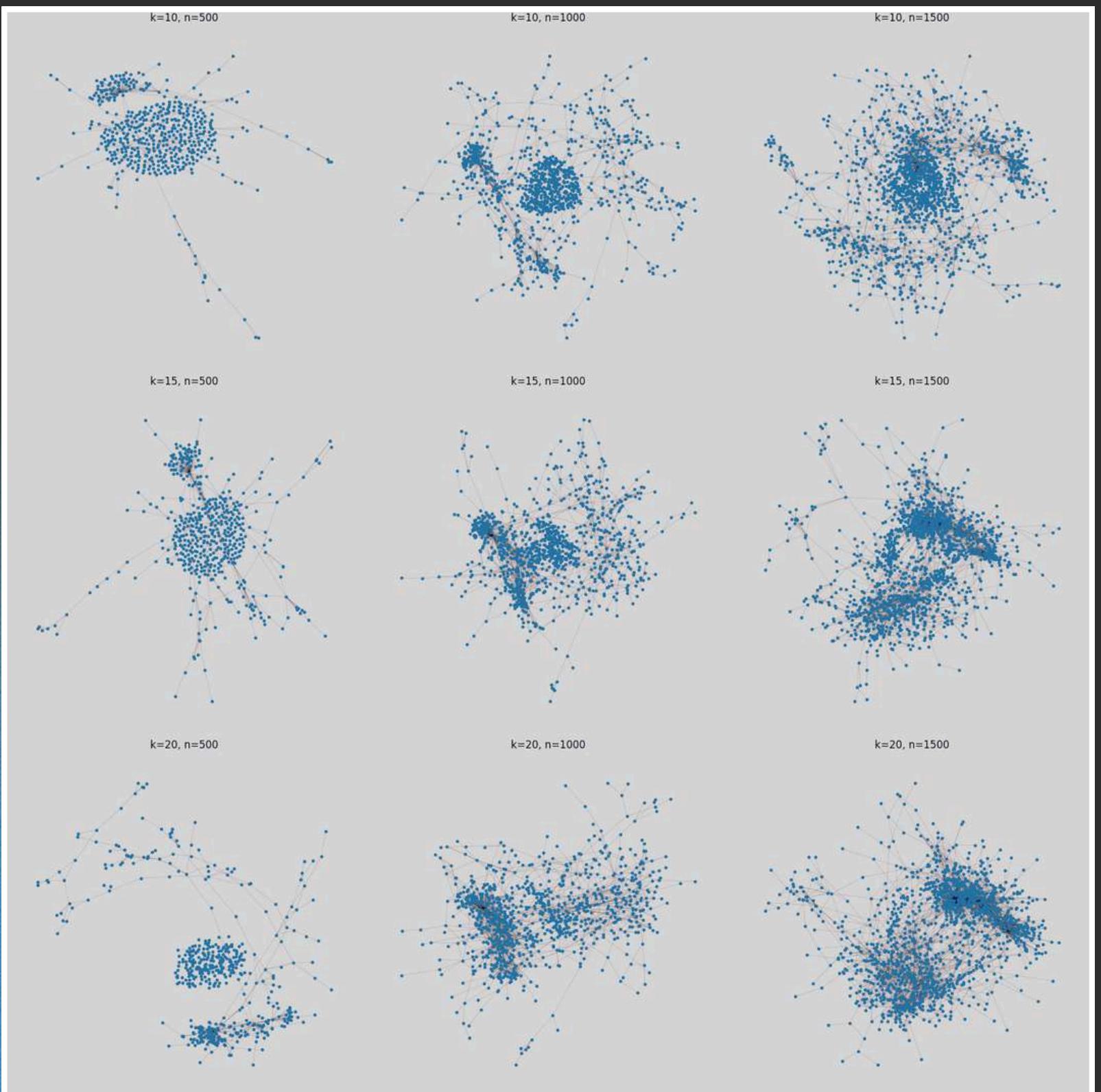
Weights used only for sparsification

**Unipartite**

All nodes are genes



# DENSITY PROBLEM



Original network:  
 $\langle k \rangle \approx 1,300$  neighbors per gene

- 1 Metrics become "trivial" (everyone is close to everyone)
- 2 Hard to reveal real structure

Method	Result
Top-q% by weight	GCC = 23-67% — network fragments!
kNN ( $k=15$ )	GCC = 97.8% — preserves connectivity

**Chosen:**  $k = 15$  (trade-off between connectivity and structure)

# SPARSIFICATION

# COMPARISON

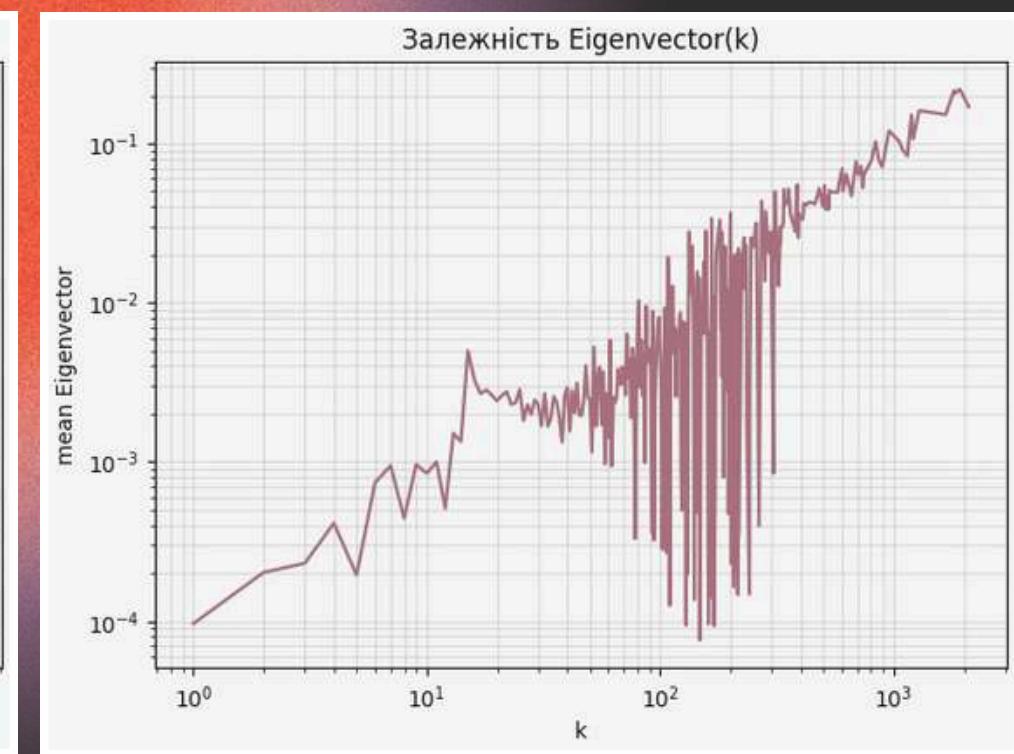
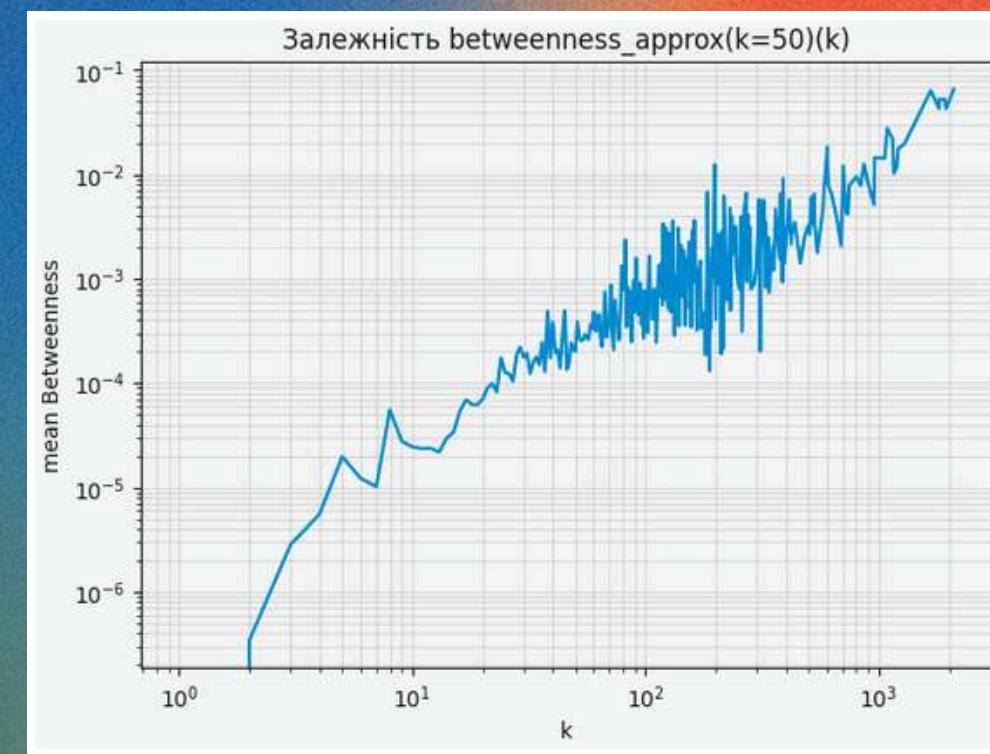
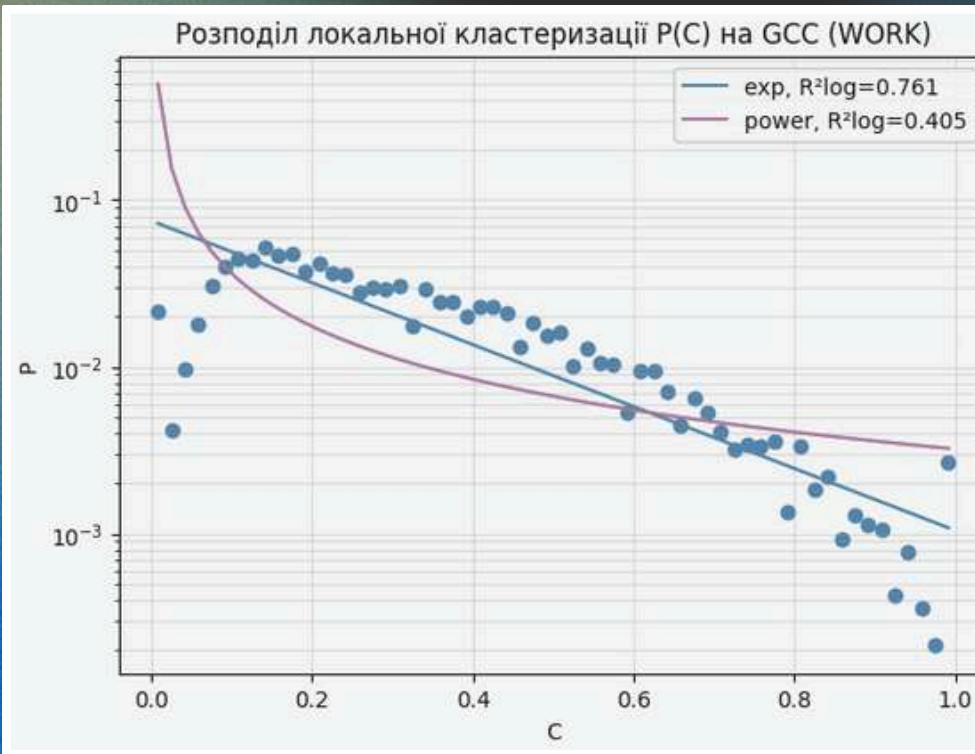
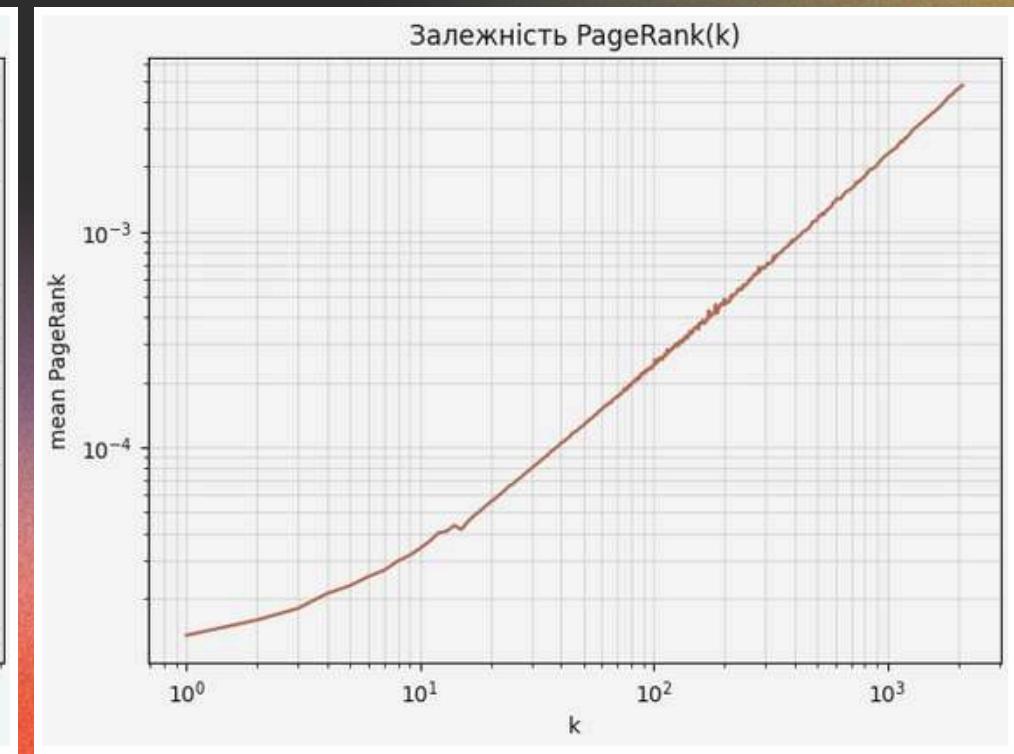
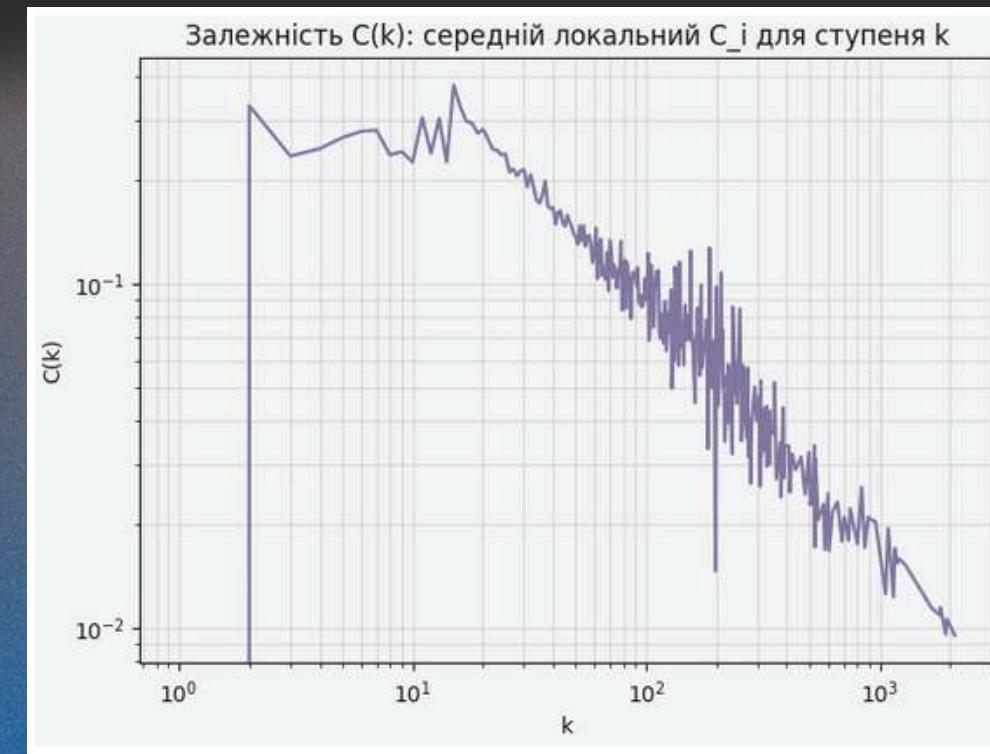
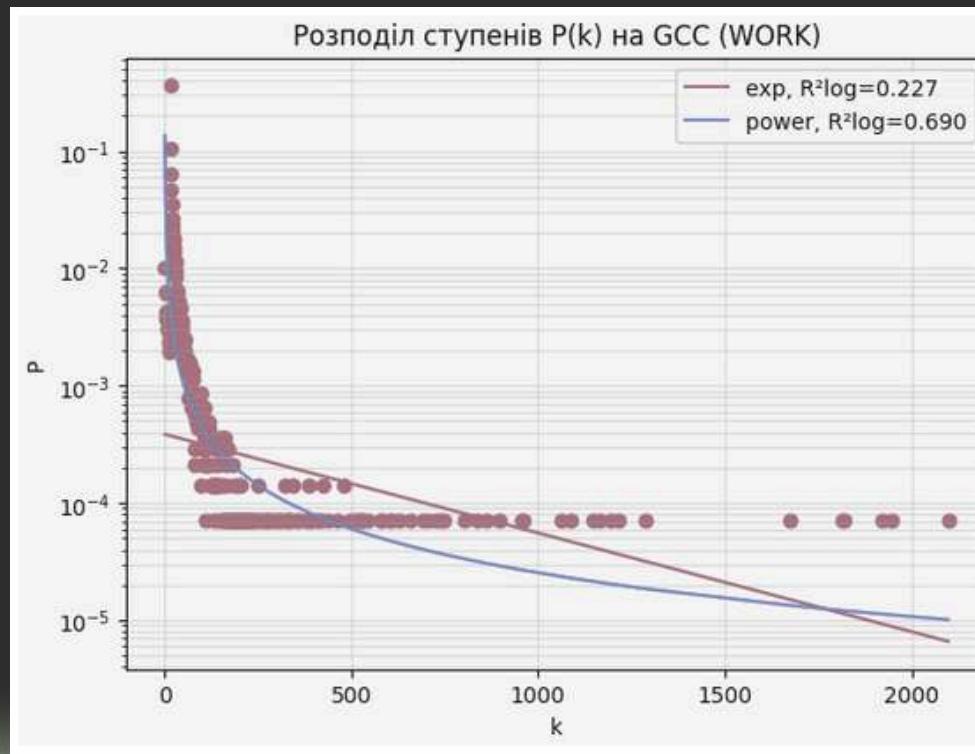
## RANDOM

## GRAPH

Metric	Gene Network	ER (random)	Ratio
M	190,332	190,173	
$\langle k \rangle$	26.55	26.52	
k_max	<b>2,097</b>	50	<b>×42</b>
C_global	<b>0.047</b>	0.002	<b>×26</b>
$\langle l \rangle$	3.41	3.23	×1.06
l_max	8	5	×1.6

- 1 Super-hubs exist ( $k_{\max}$  42× larger)
- 2 26× more triangles (modularity)
- 3 Small-world:  $\langle l \rangle \approx 3.4$  with 14K nodes

# DEGREE DISTRIBUTION CLUSTERING & CENTRALITY



1

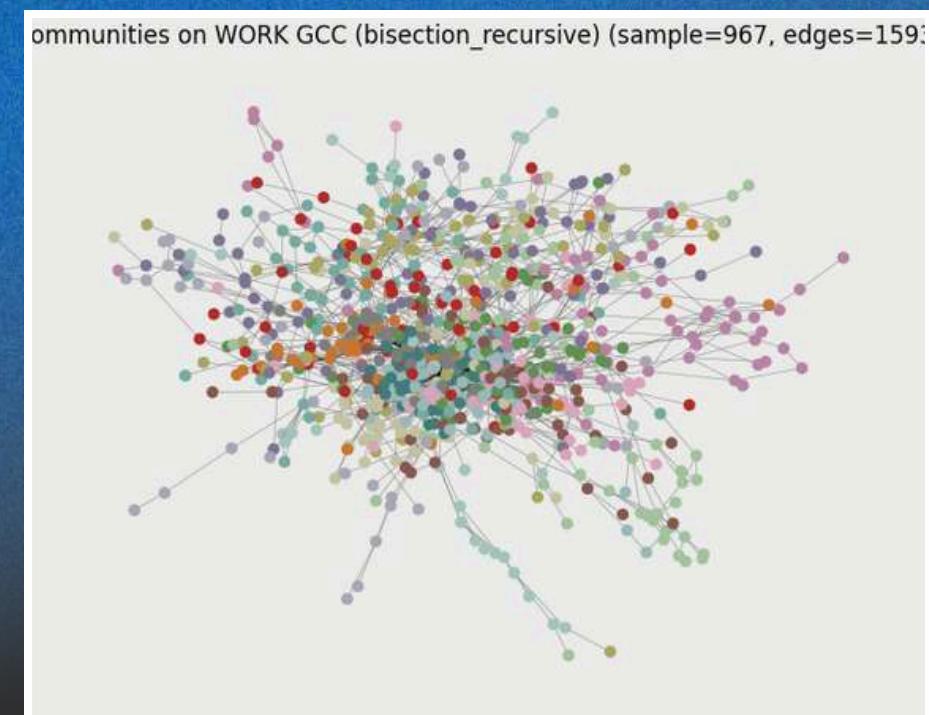
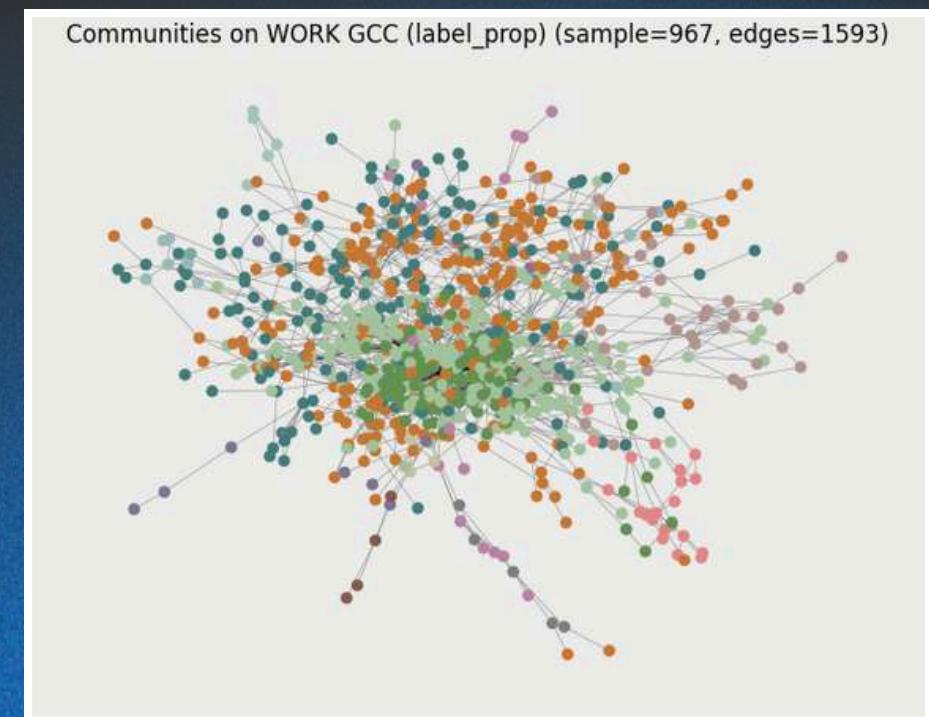
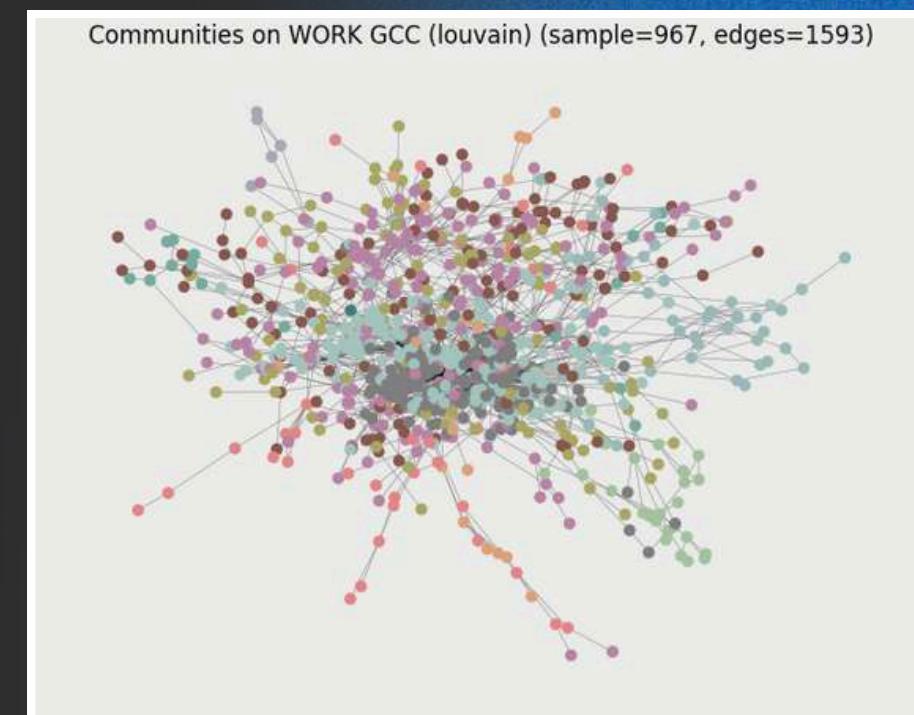
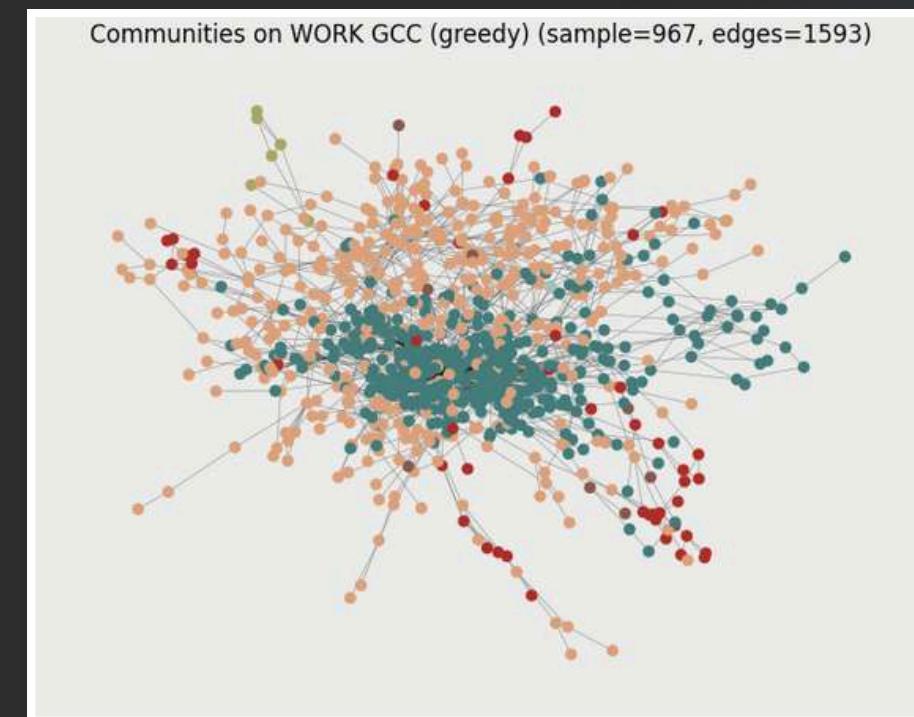
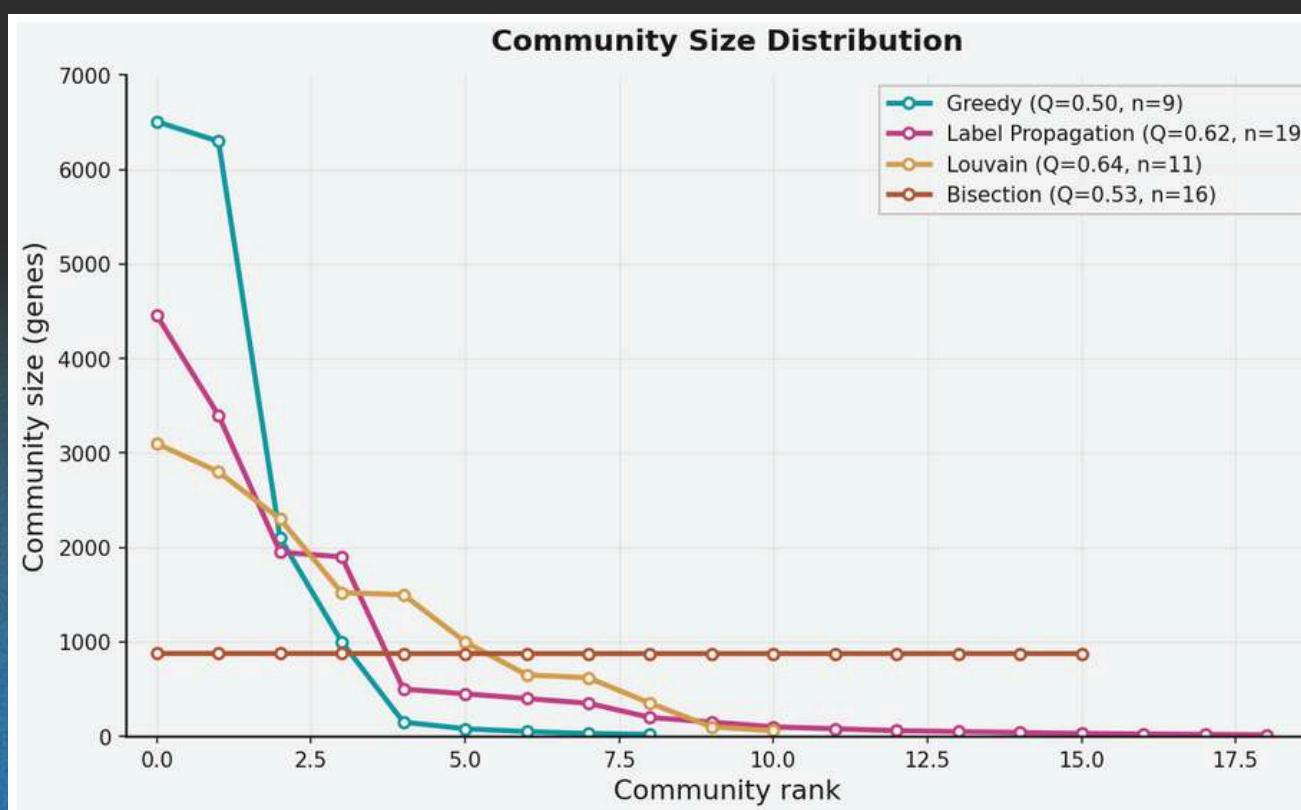
Greedy Modularity is dominated by 2 giant communities

2

Louvain achieves highest Q with balanced community sizes

3

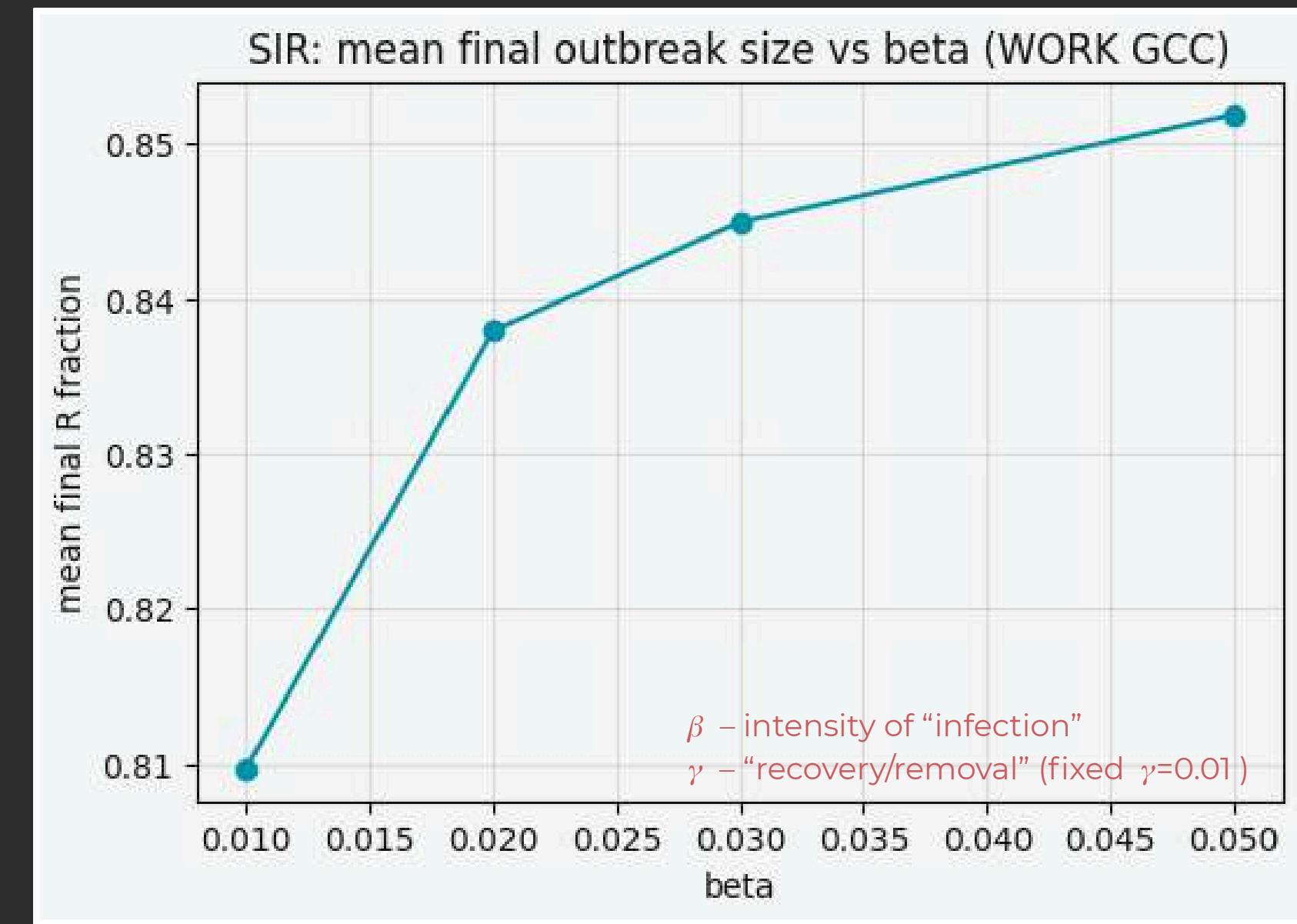
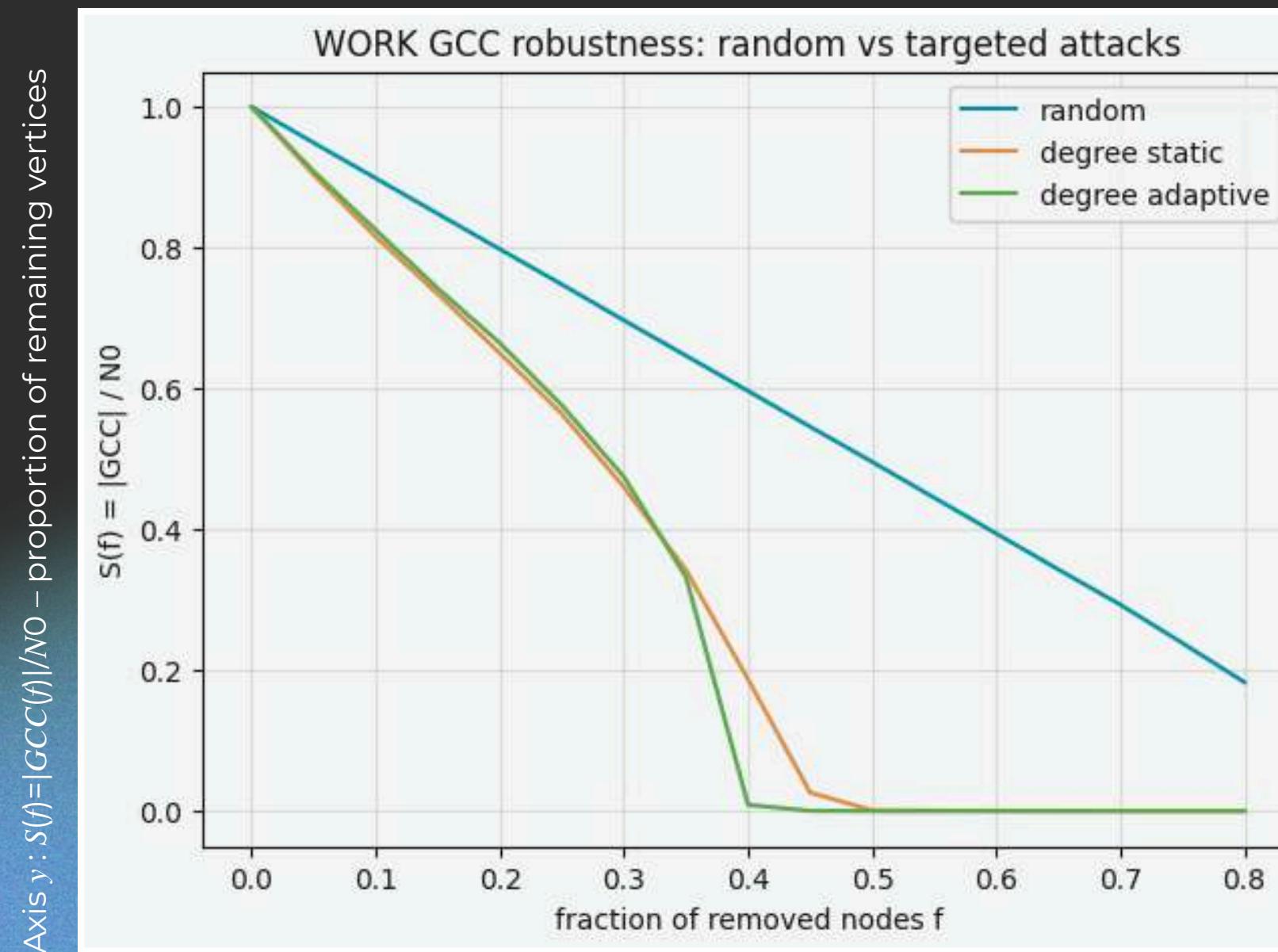
Bisection forces equal splits and loses natural structure



# COMMUNITY DETECTION

# ROBUSTNESS TO ATTACKS

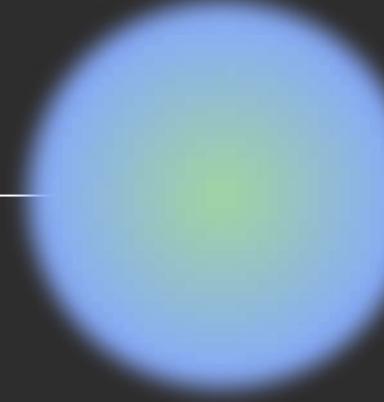
# SPREADING SIMULATION



final\_R\_frac\_mean increases approximately from 0.81 $\rightarrow$ 0.85  
peak\_I\_frac\_mean increases  $\approx$ 0.78 $\rightarrow$  $\approx$ 0.92  
t\_peak\_mean decreases sharply 44 $\rightarrow$ 13

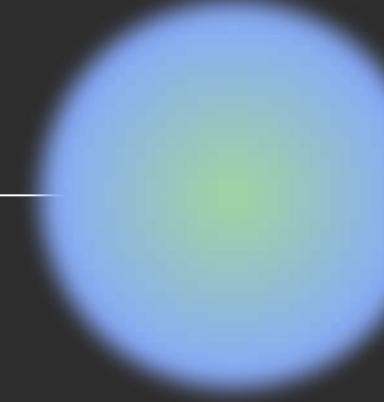
# CONCLUSIONS

## NON-RANDOM STRUCTURE



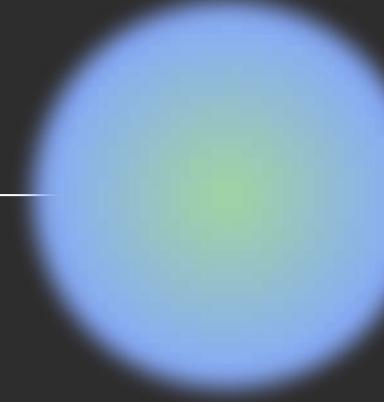
Compared to an ER baseline, the network shows extremely stronger heterogeneity and clustering (hub-dominated + highly clustered), indicating real biological organization rather than noise

## HIERARCHICAL MODULARITY



Genes form dense functional modules connected by a small set of hub “bridge” genes (high modularity  $Q \approx 0.5\text{--}0.6Q$ )

## SMALL-WORLD INTEGRATION



Short path lengths ( $\langle l \rangle \approx 3.4$ ) enable fast global connectivity while preserving local specialization inside modules

## HUB-DRIVEN ROBUSTNESS & SPREADING



The network is robust to random failures but vulnerable to targeted removal of hubs; perturbations can propagate widely (>80% coverage), highlighting the systemic impact of key genes

# THANKS

Presented by

**Anastasiia Mazur**