

HUMAN GENE

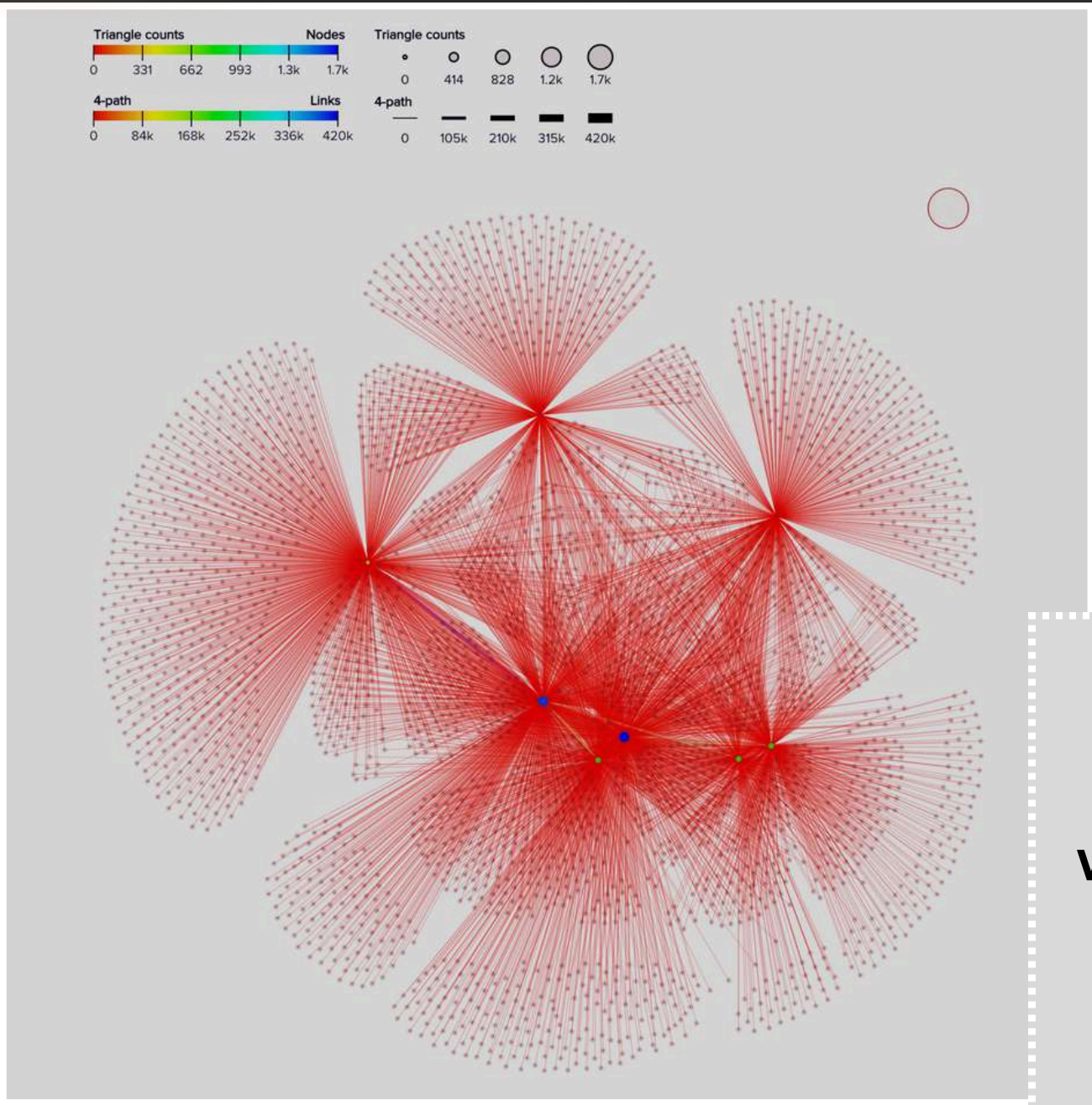
CO-EXPRESSION NETWORK
ANALYSIS

Presented by

Anastasiia Mazur

NETWORK DESCRIPTION

PROBLEM STATEMENT



What does an edge mean?

- Statistical association (correlation of expression profiles)
- NOT causal influence
- NOT physical protein interaction

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

Undirected

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

Unweighted*

Weights used only for sparsification

Unipartite

All nodes are genes

REPRESENTATION

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	
<k>	26.55	26.52	
k_max	2,097	50	×42
C_global	0.047	0.002	×26
⟨l⟩	3.41	3.23	×1.06
l_max	8	5	×1.6

- 1

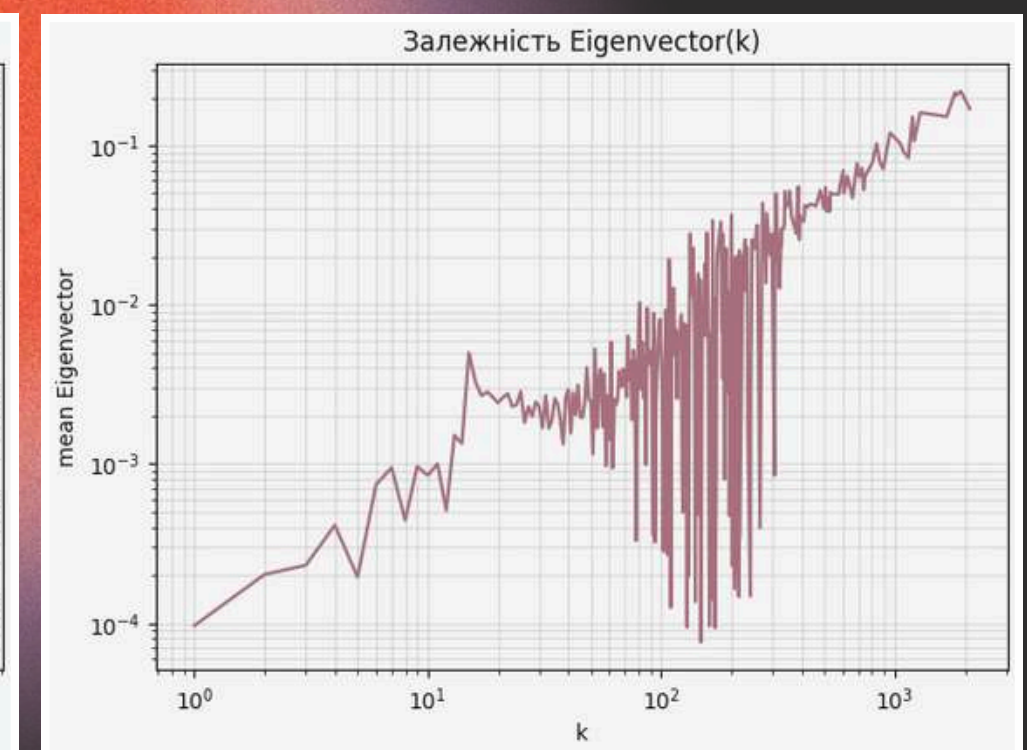
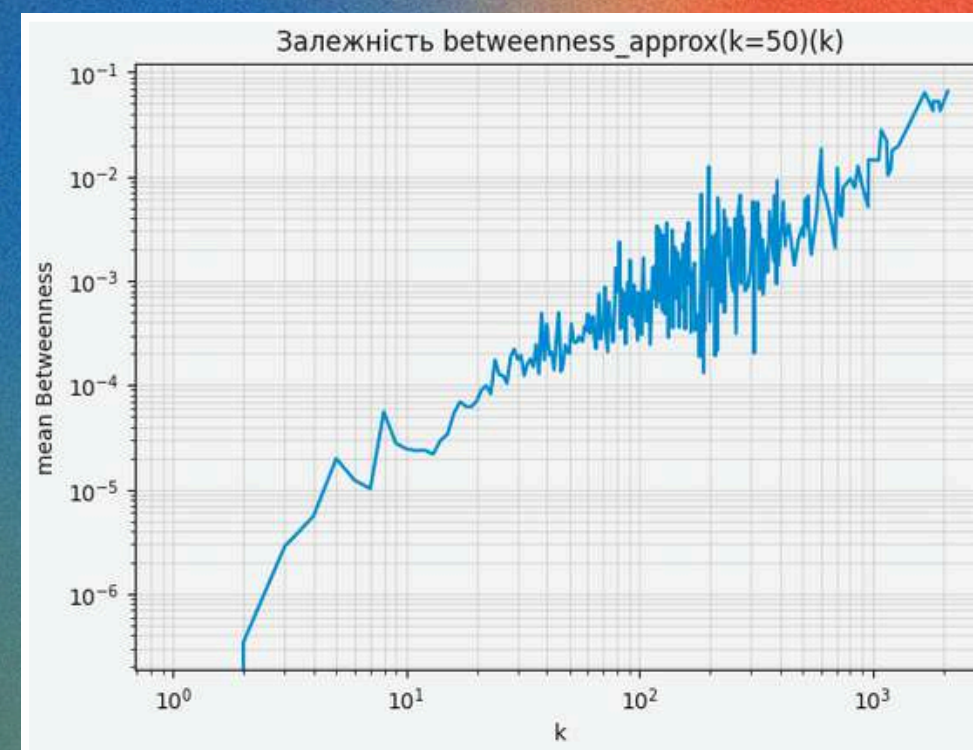
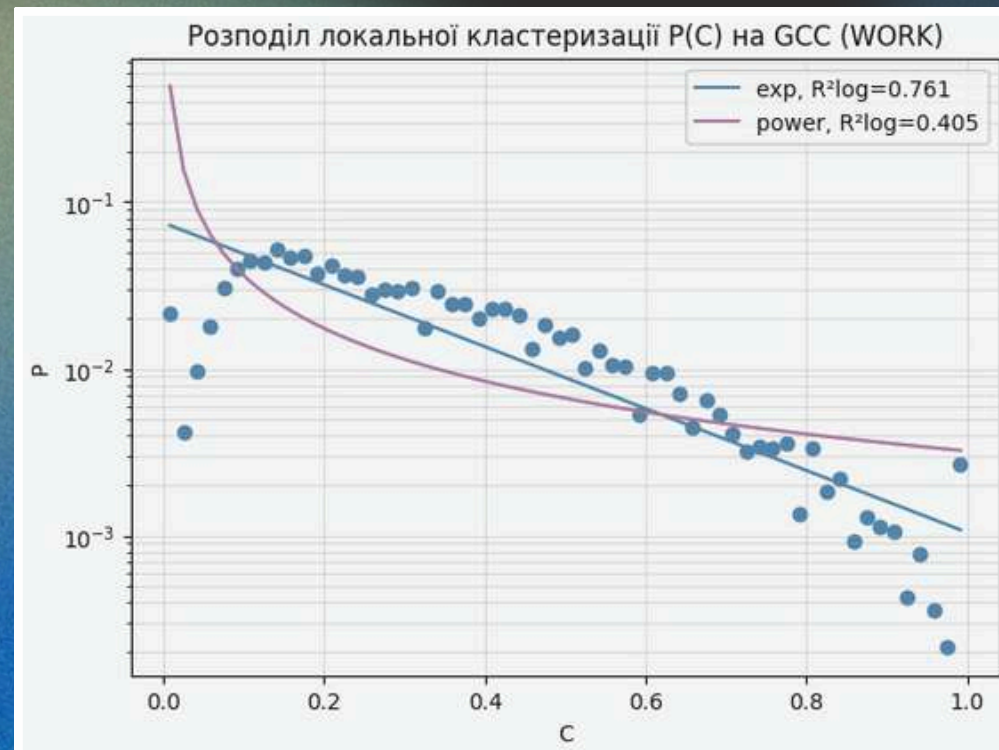
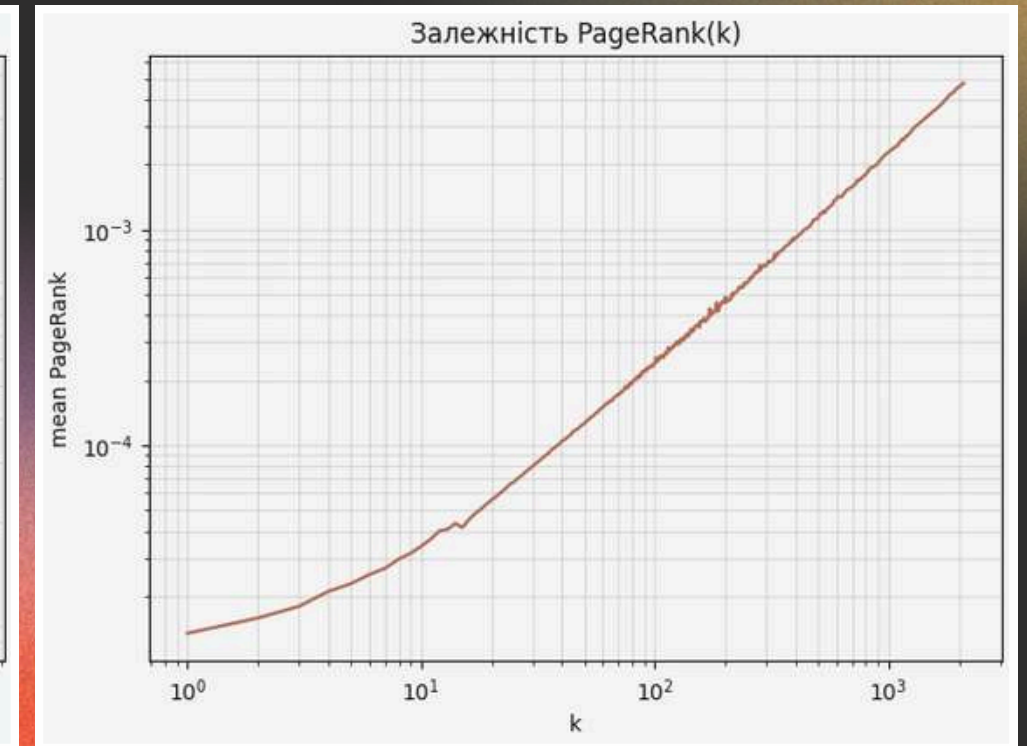
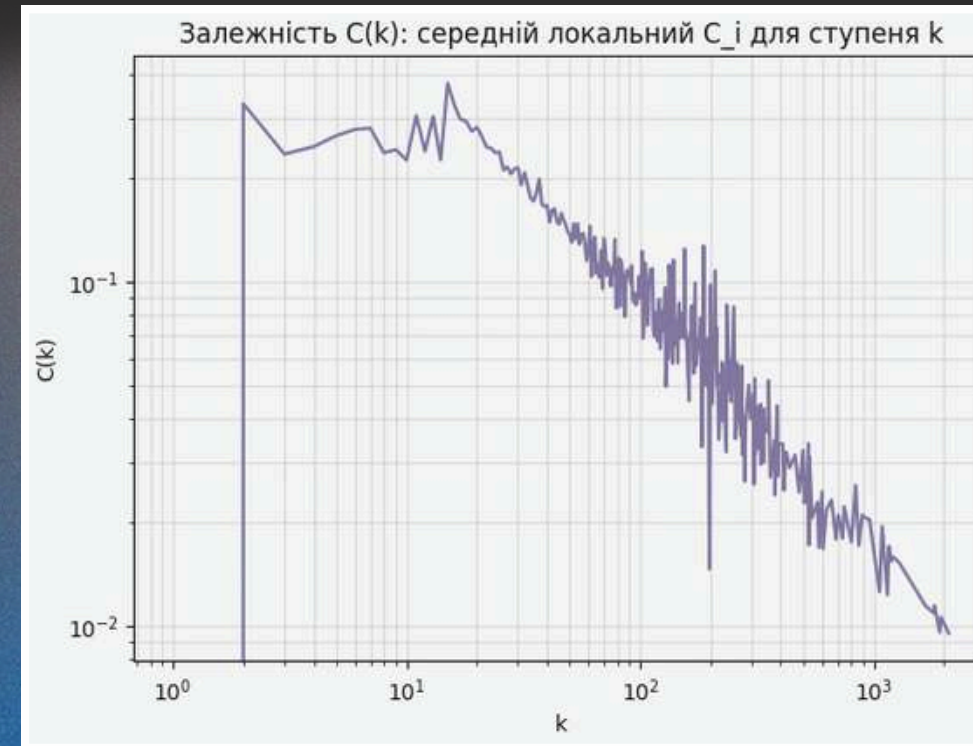
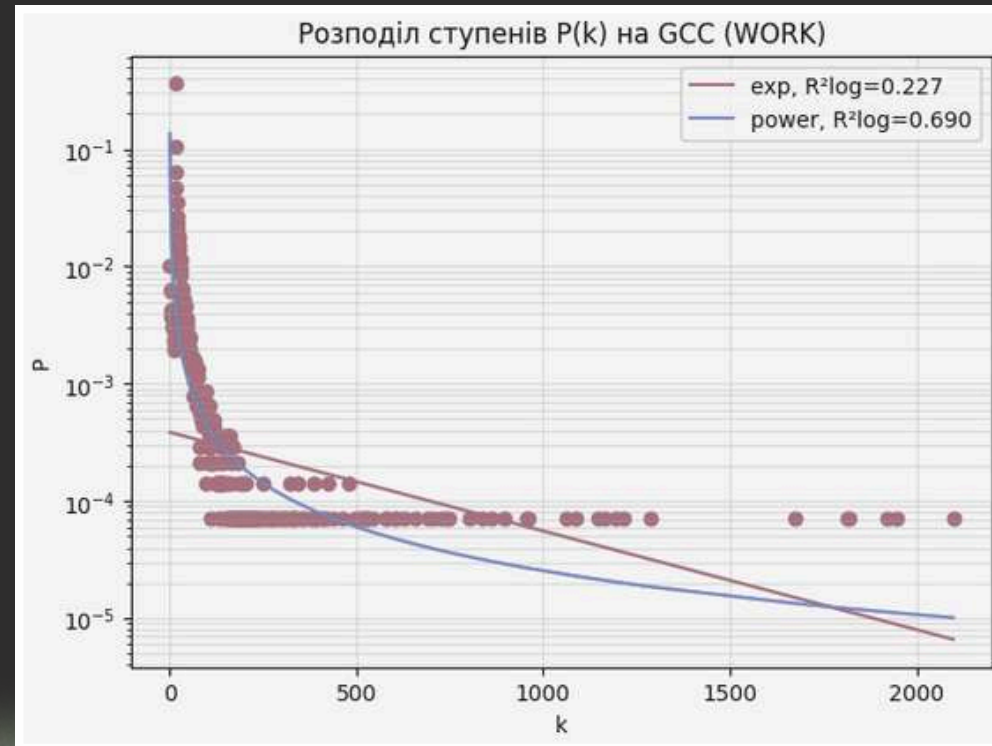
Super-hubs exist (k_max 42× larger)
- 2

26× more triangles (modularity)
- 3

Small-world: ⟨l⟩ ≈ 3.4 with 14K nodes

DEGREE DISTRIBUTION

CLUSTERING & CENTRALITY



1

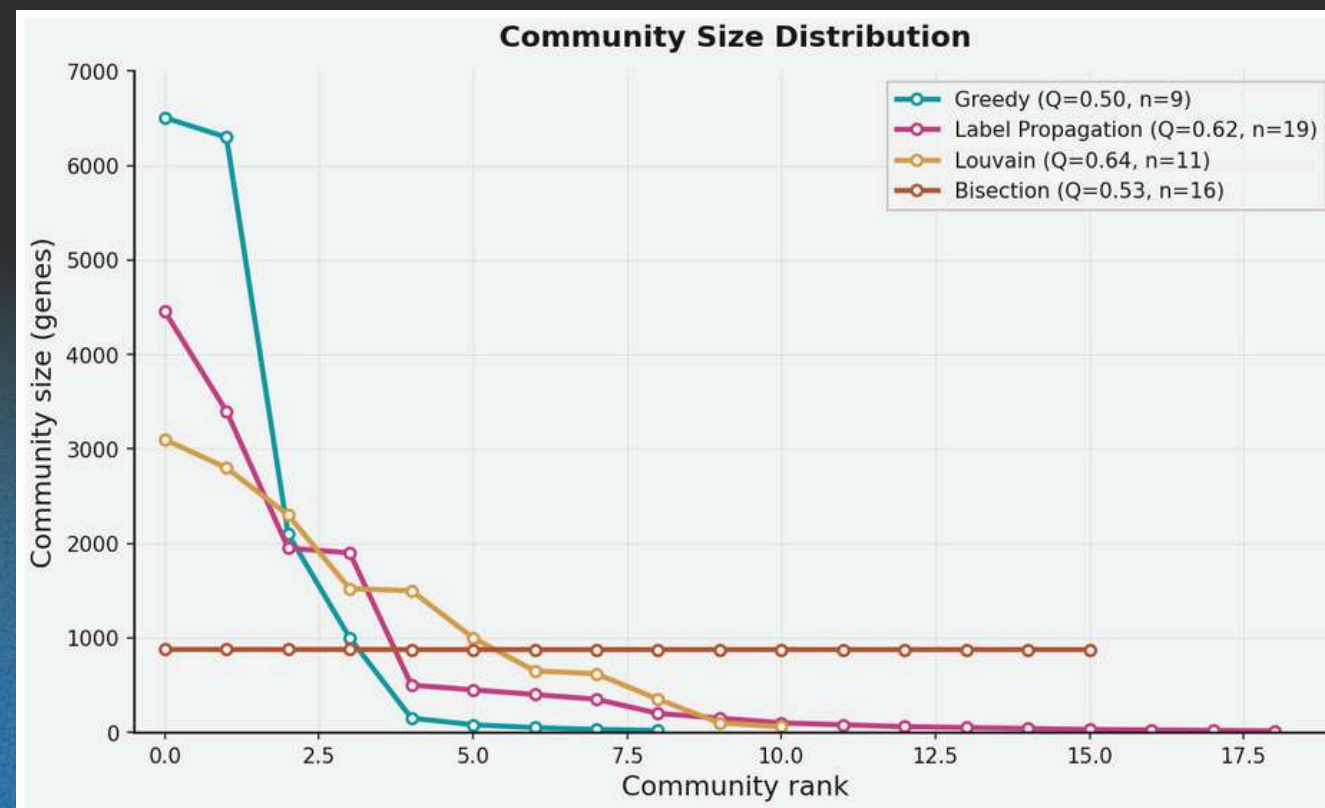
Greedy Modularity is dominated by 2 giant communities

2

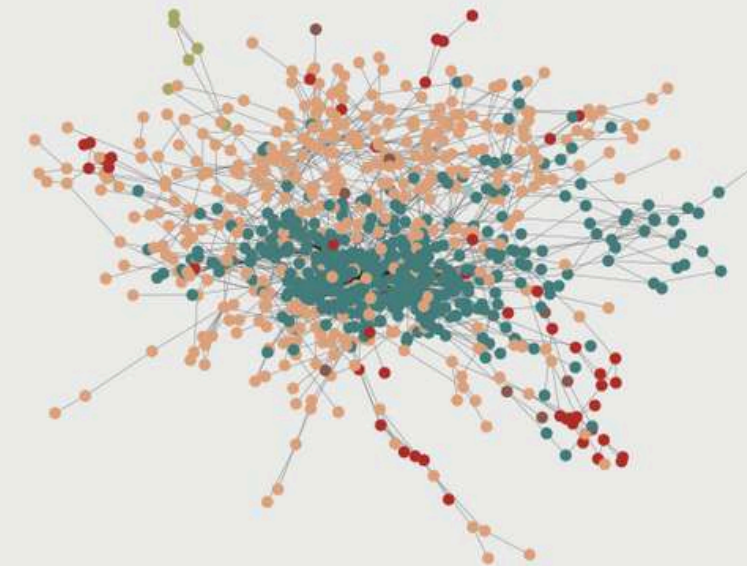
Louvain achieves highest Q with balanced community sizes

3

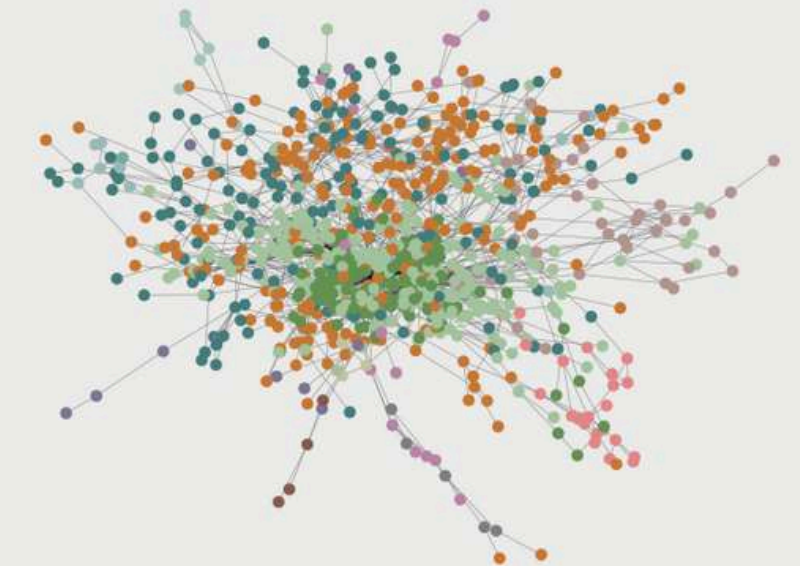
Bisection forces equal splits and loses natural structure



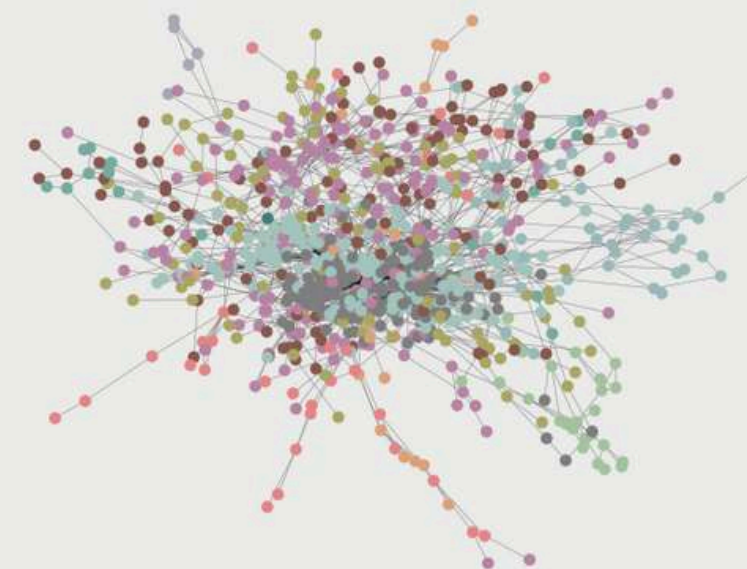
Communities on WORK GCC (greedy) (sample=967, edges=1593)



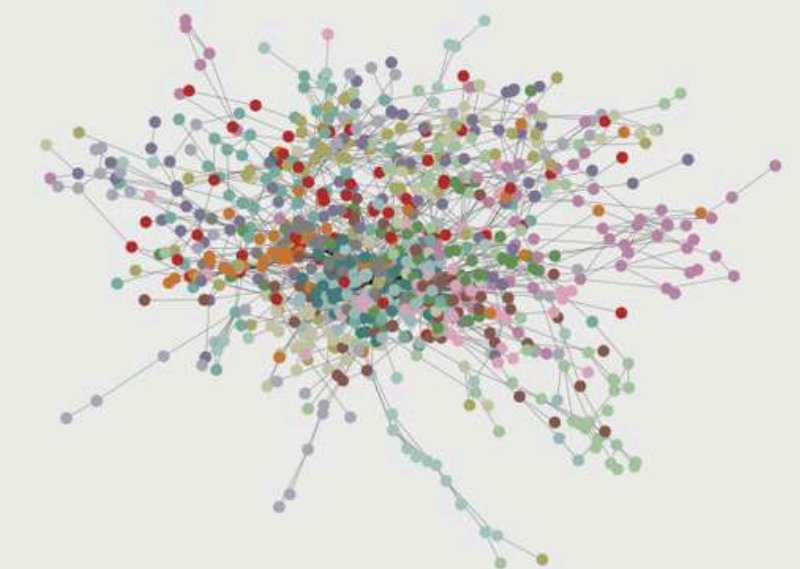
Communities on WORK GCC (label_prop) (sample=967, edges=1593)



Communities on WORK GCC (louvain) (sample=967, edges=1593)

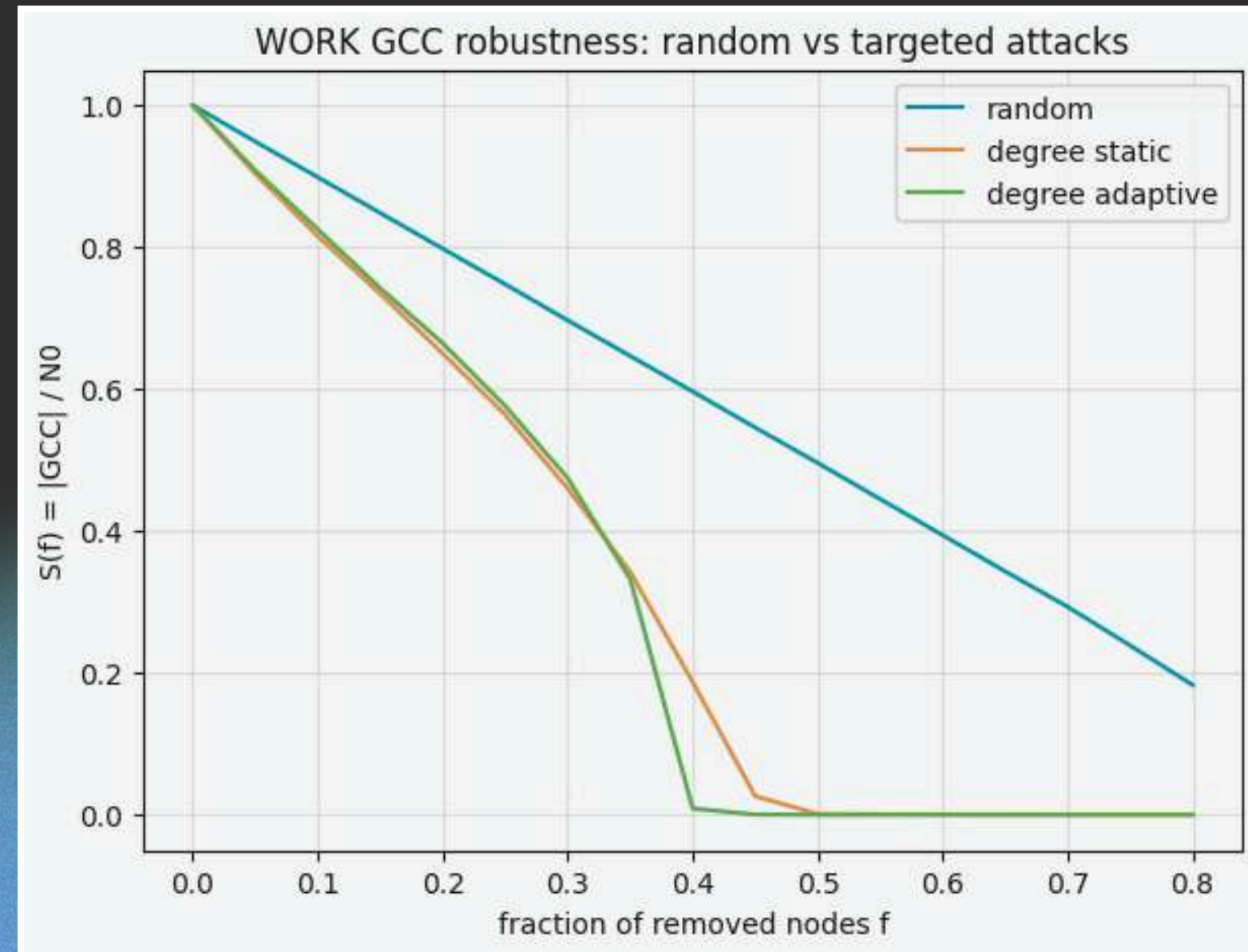


Communities on WORK GCC (bisection_recursive) (sample=967, edges=1593)



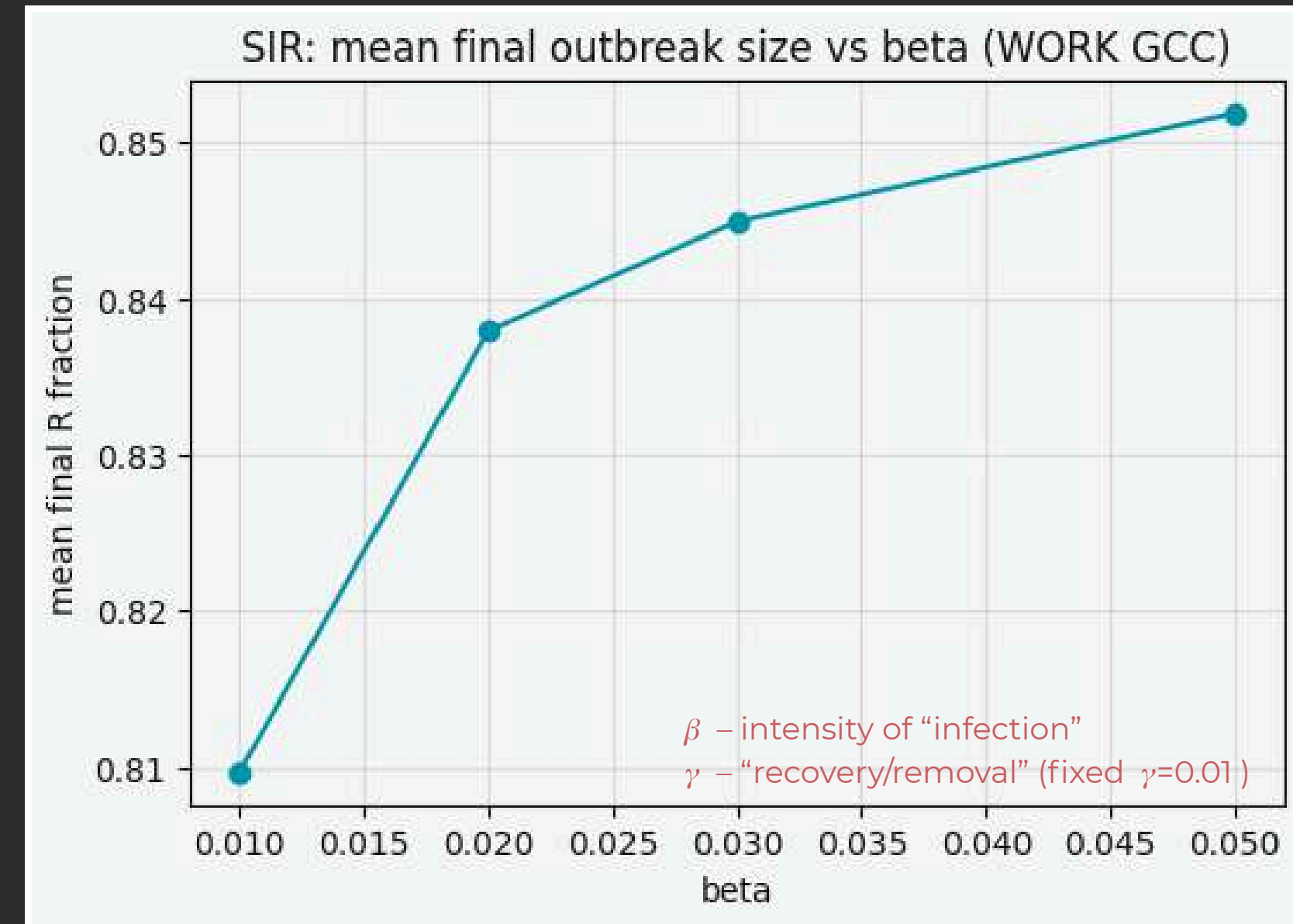
COMMUNITY DETECTION

ROBUSTNESS TO ATTACKS



Axis x : proportion of removed vertices f

SPREADING SIMULATION



For $\gamma=0.01$ and growth β :

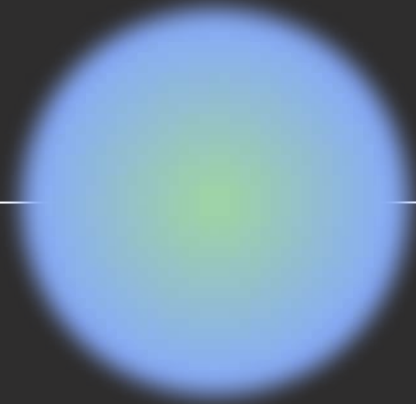
final_R_frac_mean increases approximately from 0.81→0.85

peak_I_frac_mean increases $\approx 0.78 \rightarrow \approx 0.92$

t_peak_mean decreases sharply 44→13

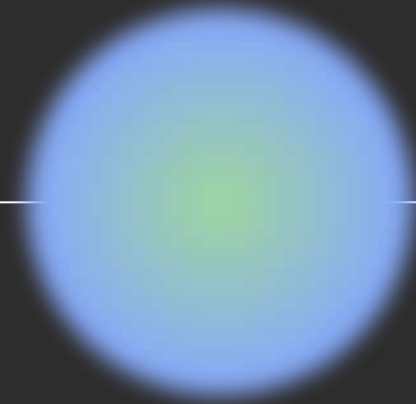
CONCLUSIONS

NON-RANDOM STRUCTURE



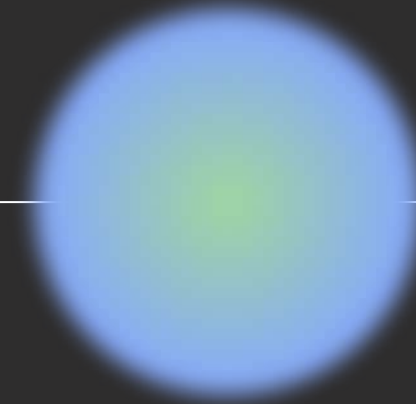
Compared to an ER baseline, the network shows extremely stronger heterogeneity and clustering (hub-dominated + highly clustered), indicating real biological organization rather than noise

HIERARCHICAL MODULARITY



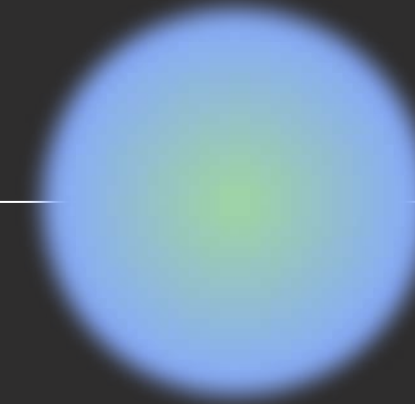
Genes form dense functional modules connected by a small set of hub “bridge” genes (high modularity $Q \approx 0.5-0.6Q$)

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