

Link Communities

Yong-Yeol “YY” Ahn



SCHOOL OF INFORMATICS
AND COMPUTING

INDIANA UNIVERSITY

Bloomington

Most populated countries



1,300,000,000+



1,200,000,000+



300,000,000+

Most populated countries



1,300,000,000+



1,200,000,000+



800,000,000+



300,000,000+

Billions of people
recording their social life

in Bits.



300 million users

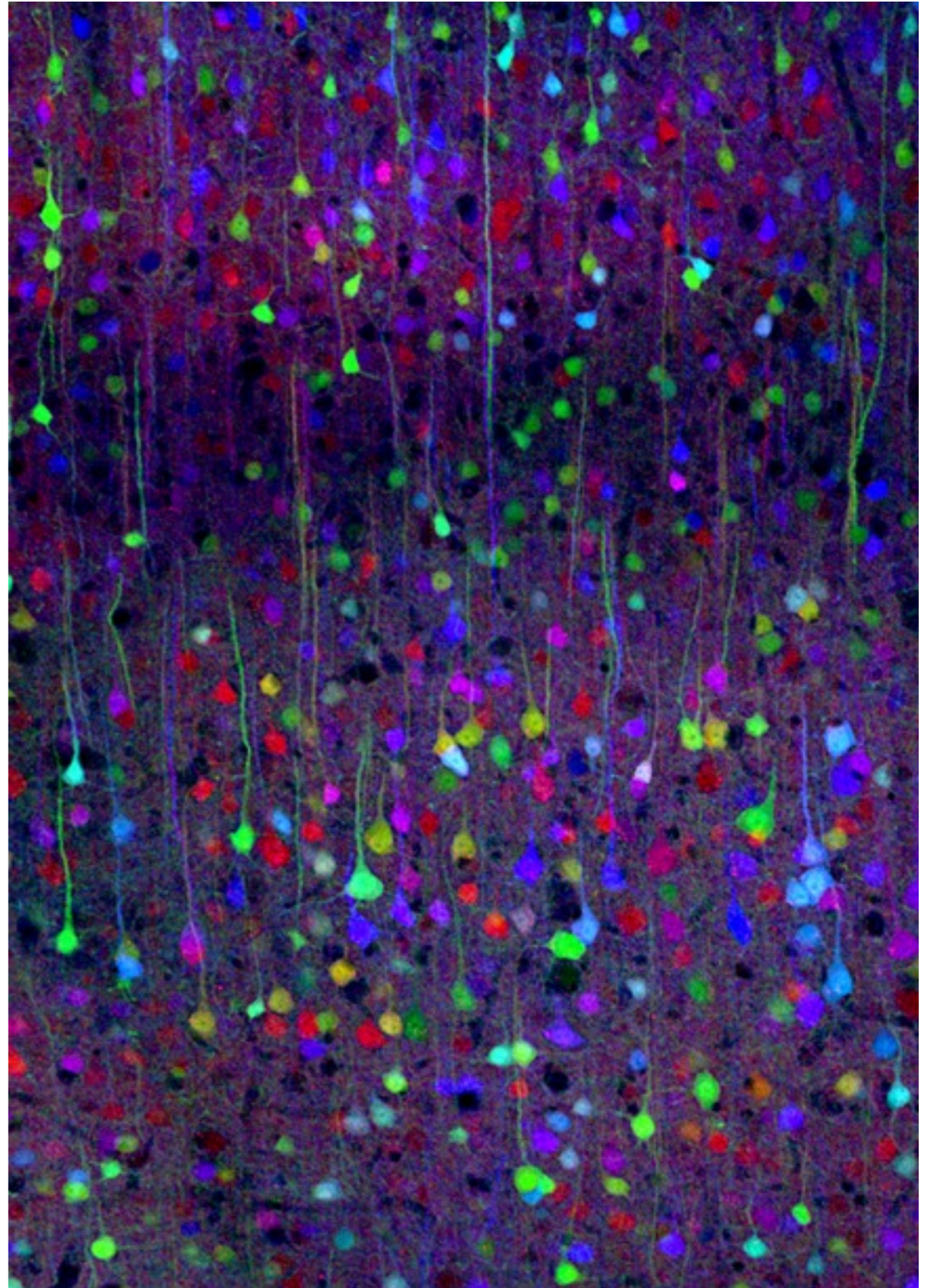
300 million tweets per day

**300 million people
publishing their life.**

48421551 48421561 48421571 48421581 48421591 48421601 48421611 48421621 48421631 48421641 48421651 48421661 48421671 48421681 48421691
tgagcagacctatataaagatgggttatgaagattcacacagcggctcatgcctgtgatcccagcactttgggaggctgaggcaagtggagcacctgagatcatgagttcaagaccagcctggccaacatggtgaaacccccatctctac
A.....T.....
tgacacagacctatataaagatgggttatgaagattcacacagcggctcatgcctgtgatcccagcactttgggaggctgaggcaagtggagcacctgagatcatgagttcaagaccagcctggccaacatggtgaaacccccatctctac
cagacctatataaagatgggttatgaagattcacacagcggctcatgcctgtgatcccagcactttgggaggctgaggcaagtggagcacctgagatcatgagttcaagaccagcctggccaacatggtgaaacccccatctctac
GACCTATATAAGATGGTTATGAAGATTCACACAGTGGCTC CCTGTGATCCCAGCACTTTGGGAGGCTGAGGCAAGTGGAG ACCTGAGATCATGAGTTCAAGACCAGCCTGGCAACATGG AACCCCATCTCTAC
TGAACAG ATATAAGATGGTTATGAAGATTCACACAGTGGCTCATGCC tgatcccagcactttgggagg TGGGCAAGTGGAGCACCTGAGATCATGAGTTCAAGACCA GCCAACATGGTGAACCCCATCTCTAC
TGAACAGAC TCAGATGGTTATGAAGATTCACACAGTGGCTCATGCCGT ATCCCAGCACTTTGGGAGGCTGAGGCAAGGGGAGCACCTG ATGAGTTCAAGACCAGCCTGGCCAACATGGTGAACCCCA CTCTAC
tgacacagccctata aagatgggttatgaagattcacacagcggctcatgcctgtg TCCCAGCACTTTGGGAGGCTGAGGCAAGTGGAGCACCTGA ATGAGTTCAAGACCAGCCTGGCCAACATGGTGAACCCCA TATA
tgacacagacctatata gatgggttatgaagattcacacagcggctcatgcctgtgat AGCACTTTGGGAGGCTGAGGCAAGGGGAGCACCTGA GAGTTCAAGACCAGCCTGGCCAACATGGTGAACCCCATC CTA
gacctatataaagatgggttatgaagattcacacagcggctc CTGTAATCCCATCACTTTGGGAGGCTGAGGCAAGTGGAGC CCTGAGATCATGAGTTCAAGA AGCCTGGCCAACATCGTGAACCCCATATCTAC
TGAACAGACCTATATAA TGGTTATGAAGATTCACACAGTGGCTCATGCCGTGATCC cacctttgggaggctgaggcaagtggagcacctgagatcat CAAGACCAGCCTGGCCAACATGGTGAACCCCATCTCTAC
TG ACCTATATAAGATGGTTATGAAGATTCACACAGTGGCTCA TGTGATCCCAGCACTTTGGGAGGCTGAGGCAAGTGGAGCA CTGAGATCACGAGTTCAAGACCAGCCTGGCCAACATGGTC AACCCCATCTCTAC
TGAACAGACCTATATAAGA GGTACGAAGATTCACACAGTGGCTCATGCCGTGATCCC cacctttgggaggctgaggcaagtggagcacctgagatcat AAGACCAGCCTGGCCAACATGGTGAACCCCATCTCTAC
tgacacagacctatataaagat gggttatgaagattcacacagcggctcatgcctgtgatcccag CTTTGGGAGGCTGAGGCAAGTGGAGCACCTGAGATCATGA agcctggccaacatggtgaaacccccatctctac
tgacacagacctatataaagatgggttatgaagattcacacagcggctcatgcctgtgatcccag GGGAGGCTGAGGCAAGTGGAGCACCTGAGATAATGAGTTC GCCTGGCCAACATGGTGAAC CCCATCTCTAC
TGAACAGACCTATATAAGATGGTTA agattcacacagcggctcatgcctgtgatcccagcacttt AGGCTGAGGCAAGTGGAGCACCTGAGATCATGAGTTCAAG CCTGGCCAACATGGTGAACCCCATCTCTAC
tgacacagacctatataaagatgggttatgaagattcacacagcggctcatgcctgtgatcccag GGTGAGGCAAGTGGAGCACCTGAGATCATGAGTTCAAGAC CCAACATGGTGAACCCCATCTCTAC
TGAACAGACCTATATAAGA TGGTTAT CAGTGGCTCATGCCGTGAT ACTTTGGGAGGCTGAGGCAAGTGGAGCACCTGAGATCATG CAACATGGTGAACCCCATCTCTAC
TGAACAGACCTATATAAGATGGTTATGAAG CAGTGGCTCATGCCGTGATC ACTTCGGGAGGCTGAGGCAAGTGGAGCACCTGAGATCATG AACATGGTGAACCCCATCTCTAC
TGAACAGACCTATATAAGATGGTTATGAAGAT CAGTGGCTCATGCCGTGATCC CCTTCGGGAGGCTGAGGCAAGTGGAGCACCTGAGATCATG ACATGGTGAACCCCATCTCTAC
TGAACAGACCTATCTAAGATGGTTATGAAGATT GCGGCTCATGCCGTGATC CTTTGGGAGGCTGAGGCAAGTGGAGCACCTGAGATCATGA ACATGGTGAACCCCATCTATA
TGAACAGACCTATATAAGATGGTTATGAAGATTC CTCTTGCCGTGATCCCAGCACTTTGGGAGGCTGACGCAA TGGAGCACCTGAGATCATGAGTTCAAGACCAGCCTGGCCA TGGTGAACCCCATCTCTAC
TGAACAGACCTATATAAGATGGTTATGAAGATTC CTCATGCCGTGATCCCAGCACTTTGGGAGGCTGAGGCAA TGGAGCACCTGAGATCATGAGTTCAAGACCAGCCTGGCCA GGTGAACCCCATCGCTAC
TGAACAGACCTATATAAGATGGTTATGAAGATTCA GTGATCCCAGCACTTTGGGAGGCTGAGGCAAGTGGAGCAC GATCATGAGTTCAAGACCAGCCTGGCCAACATGGTGAAC ccatctctac
AGATGGTTATGAAGATTCACACAGTGGCTCATGCCGTG GA CCAGCACTTTGGGAGGCTGAGGCAAGTGGAGTACCTGAGA GAGTTCAAGACCAGCCTGGCCAACATGGTGAACCCCATC TAC
ACA TGGTTATGAAGATTCACACAGTGGCTCATGCCGTG GA CTTTGGGAGGCTGAGGCAAGTGGAGCACCTGAGATCATGA CATGGTGAACCCCATCTCTAC
GGTTATGAAGATTCACACAGTGGCTCATGCCGTG GATCCC CTC TGGGAGGCTGAGGCAAGTG agcactgagatcatgagttcaagaccagcctg*caaca! tgaacccccatctctac
TATGAAGATTCACACAGTGGCTCA gatcccagcactttgggaggctgaggcaagtggagcacct agttcaagaccagcctggccaacatggtgaaacccccatctctac
ATGAAGATTCACACAGTGGCTCATGCCGTGATCCCAGCA TCTGGGAGGCTGAGGCAAGTGGAGCACCTGAGATCATGAG CATGGTGAACCCCATCTCTAC
gatcccagcactttgggaggctgaggcaagtggagcacct CATGGTGAACCCCATCTCTAC
atcccagcactttgggaggctgaggcaagtggagcacctg CATGGTGAACCCCATCTCTAC
TCTGAGAGGCTGAGGCAAGTGGAGCACCTGAGATCATGAG GTGAACCCCATCTCTAC
GGGATGCTTAGTCAATTGTAGCACCTGAGATCATGAGTTC GTGAACCCCATCTCTAC
aggctgaggcaagtggagcacctgagatcatgagttcaag gttgaaacccccatctctac
tggggcaagtggagcacctgagatcatgagttcaagacca gttgaaacccgtgtctctac
tgaggcaagtggagcacctgagatcatgagttcaagacca GAAATCCCATCTCTAC
GAGGCAAGTGGAGCACCTGAGATCATGAGTTCAAGACCAG GAAACCCCATCTCTAC
AGGCAAGTGGAGCACCTGAGATCATGAGTTCAAGACCAG GAAACCCCATCTCTAC
aggcaacttgagctccgagatcatgagttcaagaccagc gaaacccccatctctac
GCAAGTGGAGCACCTGAGATCA AACCCCATCTCTAC
CAAGTGGAGCACCTGAGATCATGAGTTCAAGACCAGCCTG AATCCCATCTCTAC
caagtggagcacctgagatcatgagttcaagaccagcctg aacccccatctctac
AAGTGGAGCACCTGAGATCATGAGTTCAAGACCAGCCTGG AACCCCATCTCTAC
AGTGGAGCACCTGAGATCATGAGTTCAAGACCAGCCTGGC ACCCCGTTCTAC
AGTGCAGCACCTGAGATCATGAGTTCAAGACCAGCCTGGC accccatctctac
GTGGAGCACCTGAGATCATGAGTTCAAGACCAGCATGGCC CCCCATCTCTAC
GGAGCACCTGAGATCATGAGTTCAAGACCAGCCTGGCCAA CATCTCTAC
ggagcacctgagatcatgagttcaagaccagggggccaa CATCTCTAC
ggagcacctgagatcatgagttcaagaccagcctggccaa CGTCTCTAC
GAGCACCTGAGATCATGAGTTCAAGACCAGCCTGGCCAA CATCTCTAC

40421551 40421561 40421571 40421581 40421591 40421601 40421611 40421621 40421631 40421641 40421651 40421661 40421671 40421681 40421691
tgagcagacctatataaagatgggttatgaagattcacacagcggctcatgcctgtgatcccagcactttgggaggctgaggcaagtggagcacctgagatcatgagttcaagaccagcctggccaacatggtgaaaccccatctctac
A.....T.....
tgacagacctatataaagatgggttatgaagattcacacagcggctcatgcctgtgatcccagcactttgggaggctgaggcaagtggagcacctgagatcatgagttcaagaccagcctggccaacatggtgaaaccccatctctac
cagacctatataaagatgggttatgaagattcacacagcggctcatgcctgtgatcccagcactttgggaggctgaggcaagtggagcacctgagatcatgagttcaagaccagcctggccaacatggtgaaaccccatctctac
GACCTATATAAGATGGTTATGAAGATTCACACAGTGGCTC CCTGTGATCCCAGCACTTTGGGAGGCTGAGGCAAGTGGAG ACCTGAGATCATGAGTTC AAGACCAGCC TGGCAACATGG AACCCCATCTCTAC
TGAACAG ATATAAGATGGTTATGAAGATTCACACAGTGGCTCATGCC tgatcccagcactttgggagg TGGGCAAGTGGAGCACCTGAGATCATGAGTTC AAGACCA GCCAACATGGTGAACCCCATCTCTAC
TGAACAGAC TCAGATGGTTATGAAGATTCACACAGTGGCTCATGCCGT ATCCCAGCACTTTGGGAGGCTGAGGCAAGGGGAGCACCTG ATGAGTTC AAGACCAGCC TGGCAACATGGTGAACCCCA CTCTAC
tgacagccctataaagatgggttatgaagattcacacagcggctcatgcctgtgatcccagcactttgggaggctgaggcaagtggagcacctgagatcatgagttcaagaccagcctggccaacatggtgaaaccccatctctac
tgacagacctatataaagatgggttatgaagattcacacagcggctcatgcctgtgatcccagcactttgggaggctgaggcaagtggagcacctgagatcatgagttcaagaccagcctggccaacatggtgaaaccccatctctac
gacctatataaagatgggttatgaagattcacacagcggctc CTGTAATCCCATCACTTTGGGAGGCTGAGGCAAGTGGAGC CCTGAGATCATGAGTTC AAGA AGCC TGGCAACATCGTGAACCCCATATCTAC
TGAACAGACCTATATAA TGGTTATGAAGATTCACACAGTGGCTCATGCCGTGATCC cacctttgggaggctgaggcaagtggagcacctgagatcat CAAGACCAGCC TGGCAACATGGTGAACCCCATCTCTAC
TG ACCTATATAAGATGGTTATGAAGATTCACACAGTGGCTCA TGTGATCCCAGCACTTTGGGAGGCTGAGGCAAGTGGAGCA CTGAGATCACGAGTTC AAGACCAGCC TGGCAACATGGTC AACCCCATCTCTAC
TGAACAGACCTATATAAGA GGTACGAAGATTCACACAGTGGCTCATGCCGTGATCCC cacctttgggaggctgaggcaagtggagcacctgagatcat AAGACCAGCC TGGCAACATGGTGAACCCCATCTCTAC
tgacagacctatataaagatgggttatgaagattcacacagcggctcatgcctgtgatcccagcactttgggaggctgaggcaagtggagcacctgagatcatgagttcaagaccagcctggccaacatggtgaaaccccatctctac
tgacagacctatataaagatgggttatgaagattcacacagcggctcatgcctgtgatcccagcactttgggaggctgaggcaagtggagcacctgagatcatgagttcaagaccagcctggccaacatggtgaaaccccatctctac
TGAACAGACCTATATAAGATGGTTATGAAG CAGTGGCTCATGCCGTGAT AC TTTGGGAGGCTGAGGCAAGTGGAGCACCTGAGATCATG CAACATGGTGAACCCCATCTCTAC
TGAACAGACCTATATAAGATGGTTATGAAG CAGTGGCTCATGCCGTGATC AC TTTGGGAGGCTGAGGCAAGTGGAGCACCTGAGATCATG AACATGGTGAACCCCATCTCTAC
TGAACAGACCTATATAAGATGGTTATGAAGAT CAGTGGCTCATGCCGTGATCC CCTTTGGGAGGCTGAGGCAAGTGGAGCACCTGAGATCATG ACATGGTGAACCCCATCTCTAC
TGAACAGACCTATATAAGATGGTTATGAAGATT GCGGCTCATGCCGTGATC CTTTGGGAGGCTGAGGCAAGTGGAGCACCTGAGATCATG ACATGGTGAACCCCATCTCTAC
TGAACAGACCTATATAAGATGGTTATGAAGATTC CTGTTGATCCCACACTTTGGGAGGCTGAGGCAAGTGGAGCACCTGAGATCATGAGTTC AAGACCAGCC TGGCAACATGGTGAACCCCATCTCTAC
TGAACAGACCTATATAAGATGGTTATGAAGATTC CTGTTGATCCCACACTTTGGGAGGCTGAGGCAAGTGGAGCACCTGAGATCATGAGTTC AAGACCAGCC TGGCAACATGGTGAACCCCATCTCTAC
TGAACAGACCTATATAAGATGGTTATGAAGATTCA GTGATCCCACACTTTGGGAGGCTGAGGCAAGTGGAGCACCTGAGATCATGAGTTC AAGACCAGCC TGGCAACATGGTGAACCCCATCTCTAC
AGATGGTTATGAAGATTCACACAGTGGCTCATGCCGTGATCCC CACTTTGGGAGGCTGAGGCAAGTGGAGCACCTGAGATCATGAGTTC AAGACCAGCC TGGCAACATGGTGAACCCCATCTCTAC
ACA TGGTTATGAAGATTCACACAGTGGCTCATGCCGTGATCCC CACTTTGGGAGGCTGAGGCAAGTGGAGCACCTGAGATCATGAGTTC AAGACCAGCC TGGCAACATGGTGAACCCCATCTCTAC
GGTTATGAAGATTCACACAGTGGCTCATGCCGTGATCCC CACTTTGGGAGGCTGAGGCAAGTGGAGCACCTGAGATCATGAGTTC AAGACCAGCC TGGCAACATGGTGAACCCCATCTCTAC
TATGAAGATTCACACAGTGGCTCATGCCGTGATCCC CACTTTGGGAGGCTGAGGCAAGTGGAGCACCTGAGATCATGAGTTC AAGACCAGCC TGGCAACATGGTGAACCCCATCTCTAC
ATGAAGATTCACACAGTGGCTCATGCCGTGATCCC CACTTTGGGAGGCTGAGGCAAGTGGAGCACCTGAGATCATGAGTTC AAGACCAGCC TGGCAACATGGTGAACCCCATCTCTAC
gatcccagcactttgggaggctgaggcaagtggagcacct
atcccagcactttgggaggctgaggcaagtggagcacctg
TCTGAGAGGCTGAGGCAAGTGGAGCACCTGAGATCATGAG
GGGATGCTTAGTCAATTGTAGCACCTGAGATCATGAGTTC
aggctgaggcaagtggagcacctgagatcatgagttcaag
tgggcaagtggagcacctgagatcatgagttcaagacca
tgggcaagtggagcacctgagatcatgagttcaagacca
GAGGCAAGTGGAGCACCTGAGATCATGAGTTC AAGACCAG
AGGCAAGTGGAGCACCTGAGATCATGAGTTC AAGACCAGC
aggcaagtggagcacctgagatcatgagttcaagaccagc
GCAAGTGGAGCACCTGAGATCA
CAAGTGGAGCACCTGAGATCATGAGTTC AAGACCAGCC TG
caagtggagcacctgagatcatgagttcaagaccagcctg
AAGTGGAGCACCTGAGATCATGAGTTC AAGACCAGCC TG
AGTGGAGCACCTGAGATCATGAGTTC AAGACCAGCC TGGC
AGTGCAGCACCTGAGATCATGAGTTC AAGACCAGCC TGGC
GTGGAGCACCTGAGATCATGAGTTC AAGACCAGCATGGCC
GGAGCACCTGAGATCATGAGTTC AAGACCAGCC TGGCAA
ggagcacctgagatcatgagttcaagaccagggggccaa
ggagcacctgagatcatgagttcaagaccagcctggccaa
GAGCACCTGAGATCATGAGTTC AAGACCAGCC TGGCAA

\$1000



BIG DATA

BIG DATA



INFORMATION





BIG DATA

SOCIETY

LIFE

BIG DATA

ECONOMY

LIFE
SOCIETY
ECONOMY

BIG DATA

LIFE
SOCIETY
ECONOMY

BIG DATA

COMPLEX
SYSTEMS

COMPLEX SYSTEMS

COMPLEX SYSTEMS

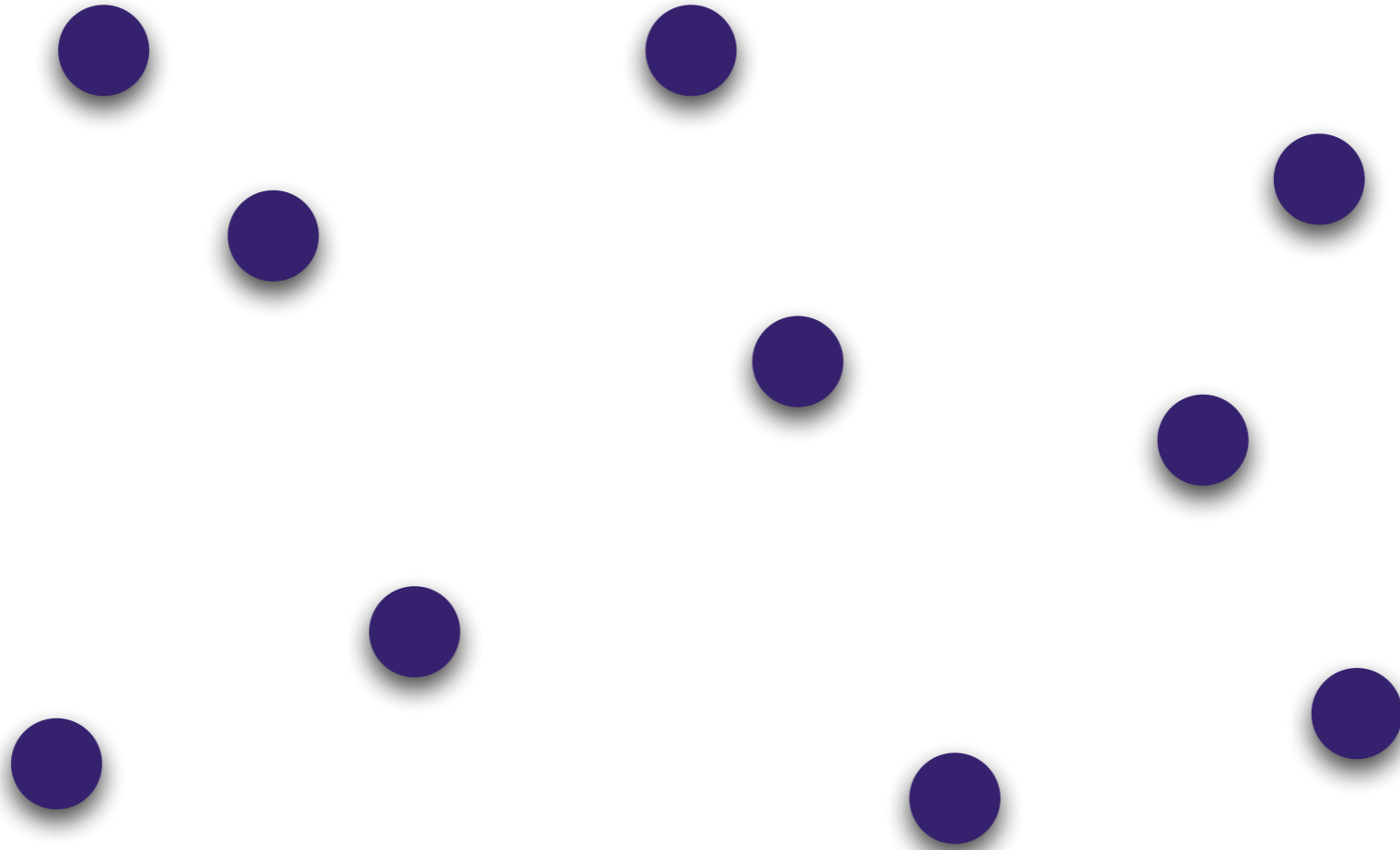
MANY parts,

INTERACTING with each other

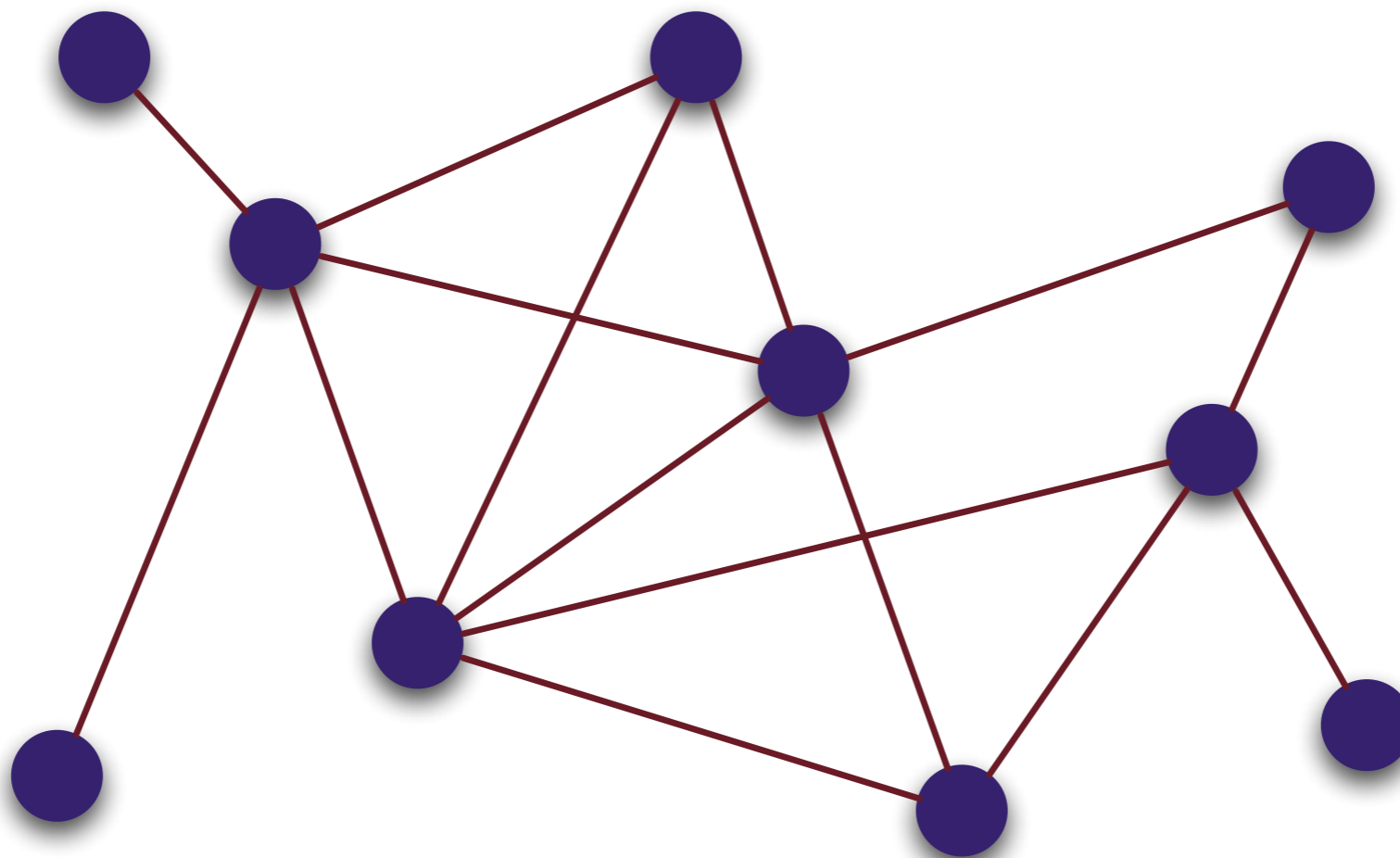
in **NON-TRIVIAL WAYS**

NETWORKS

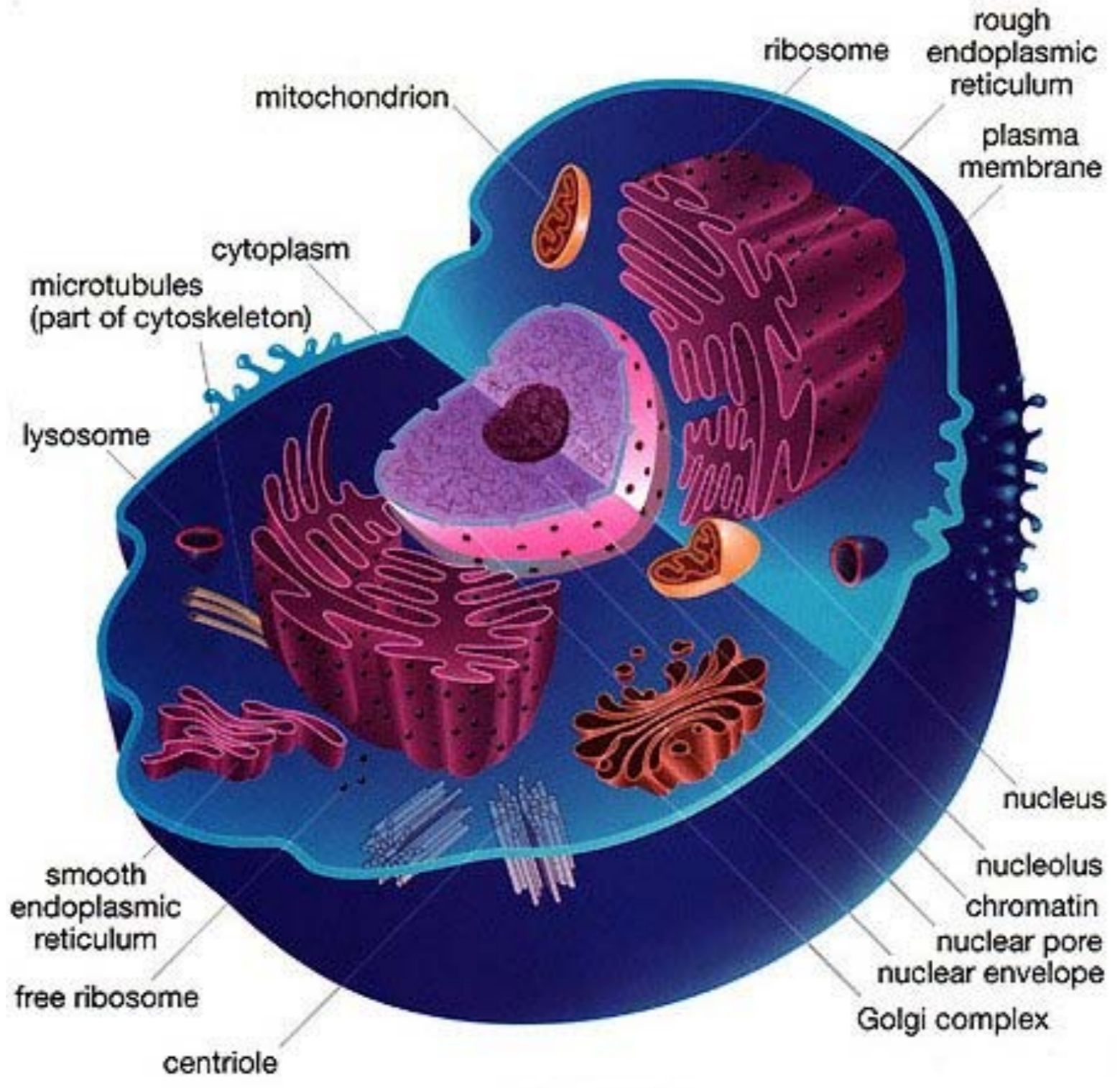


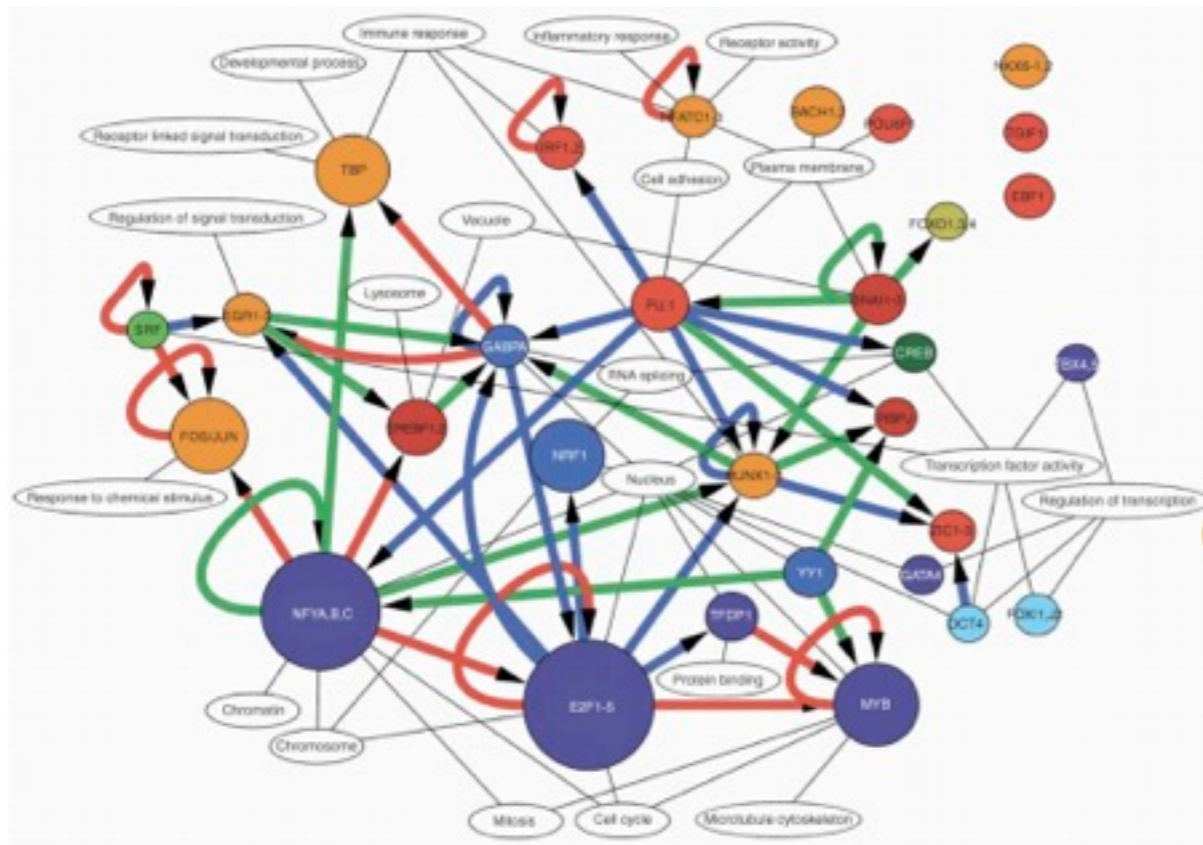


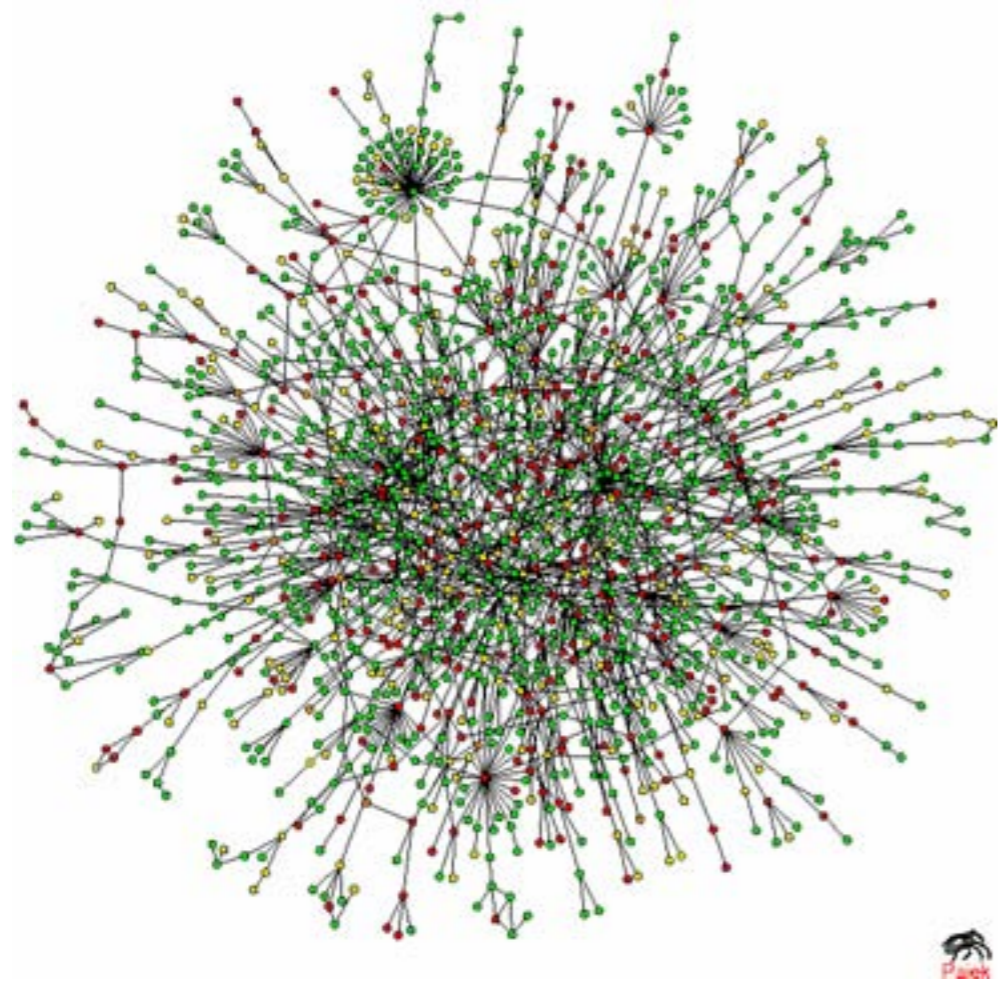
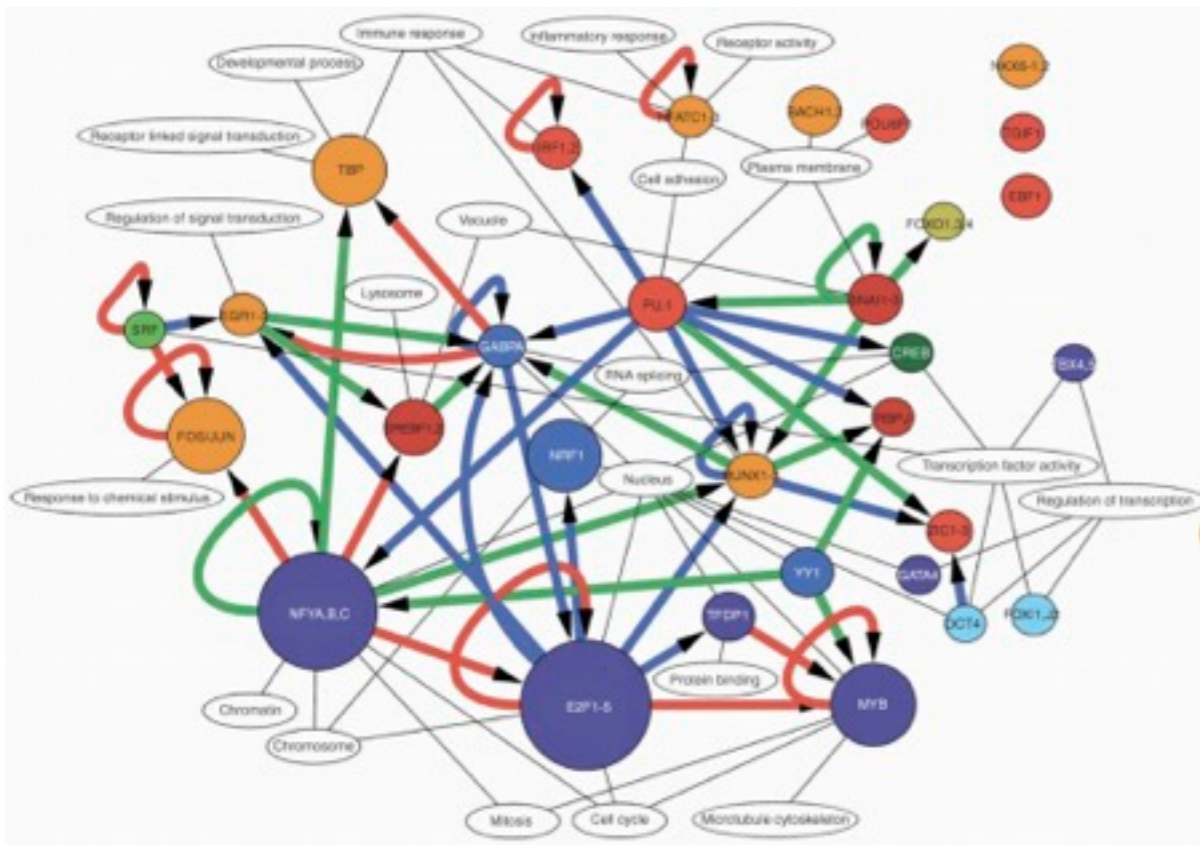
Nodes

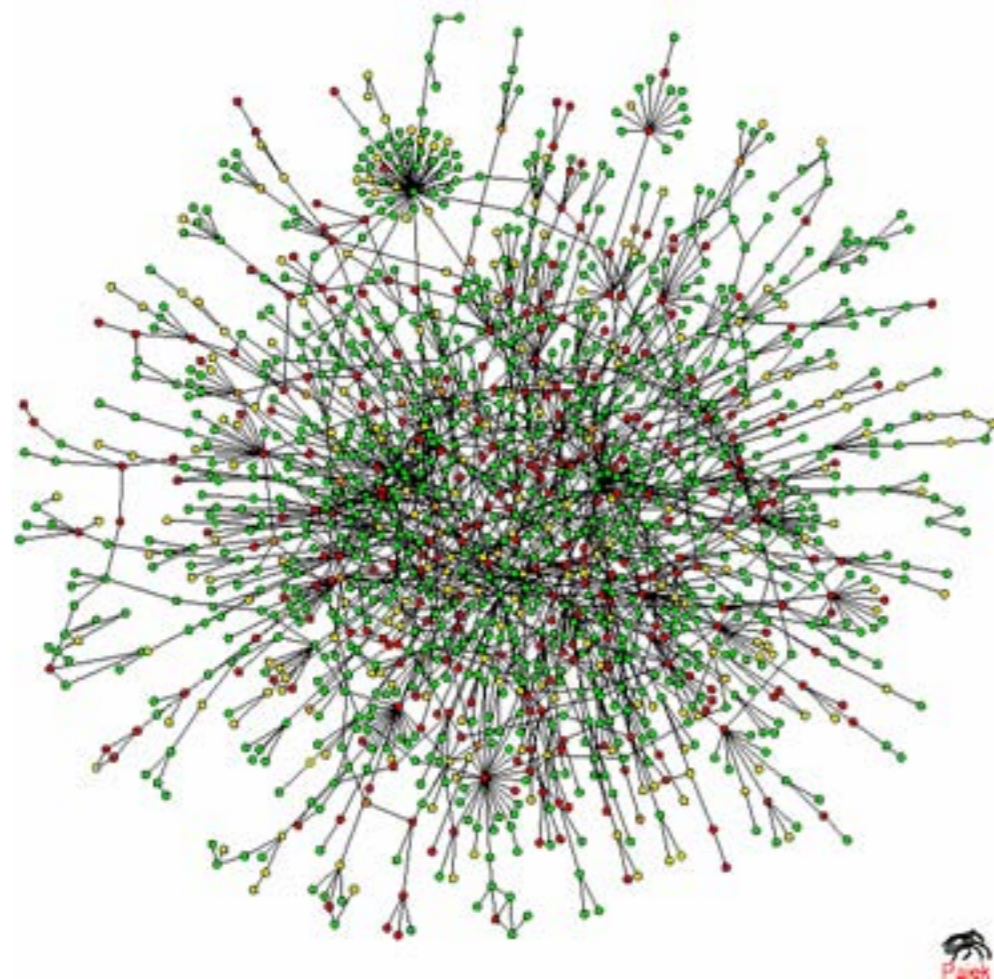
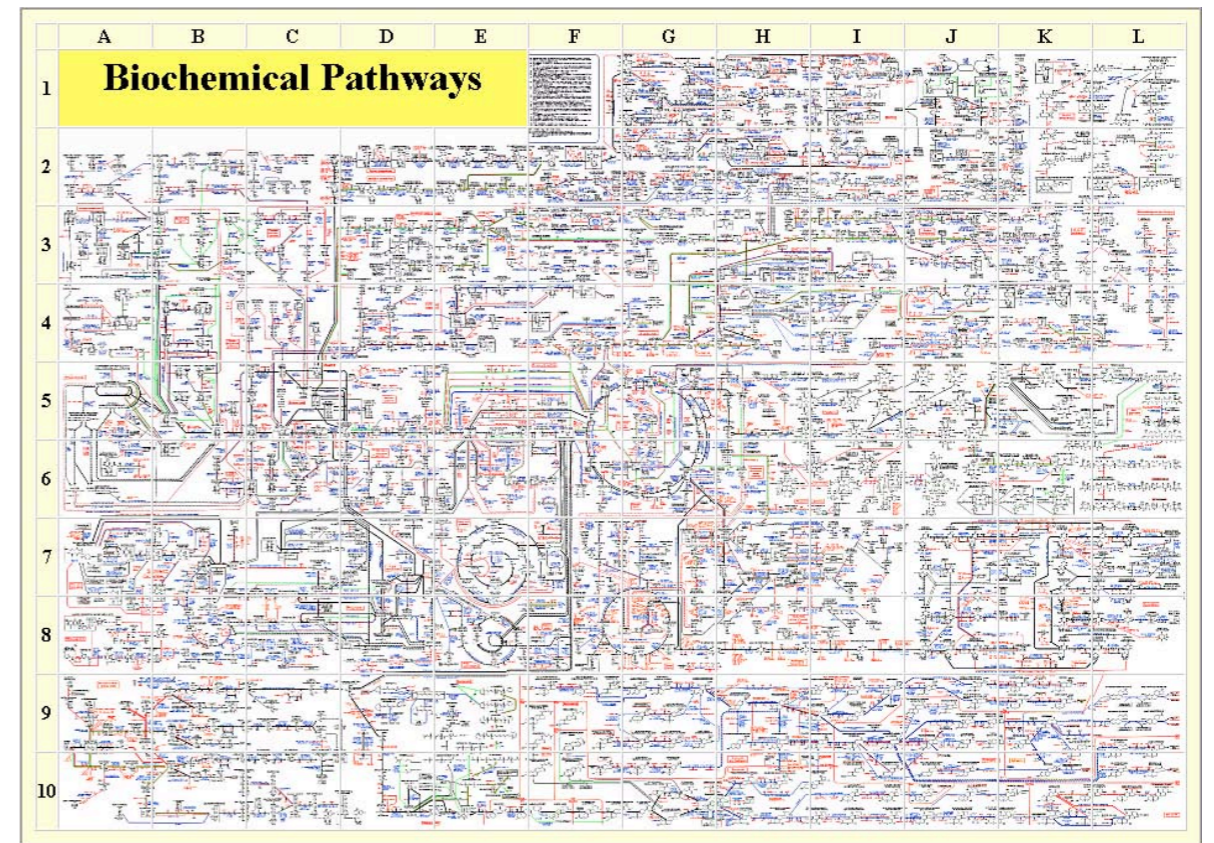
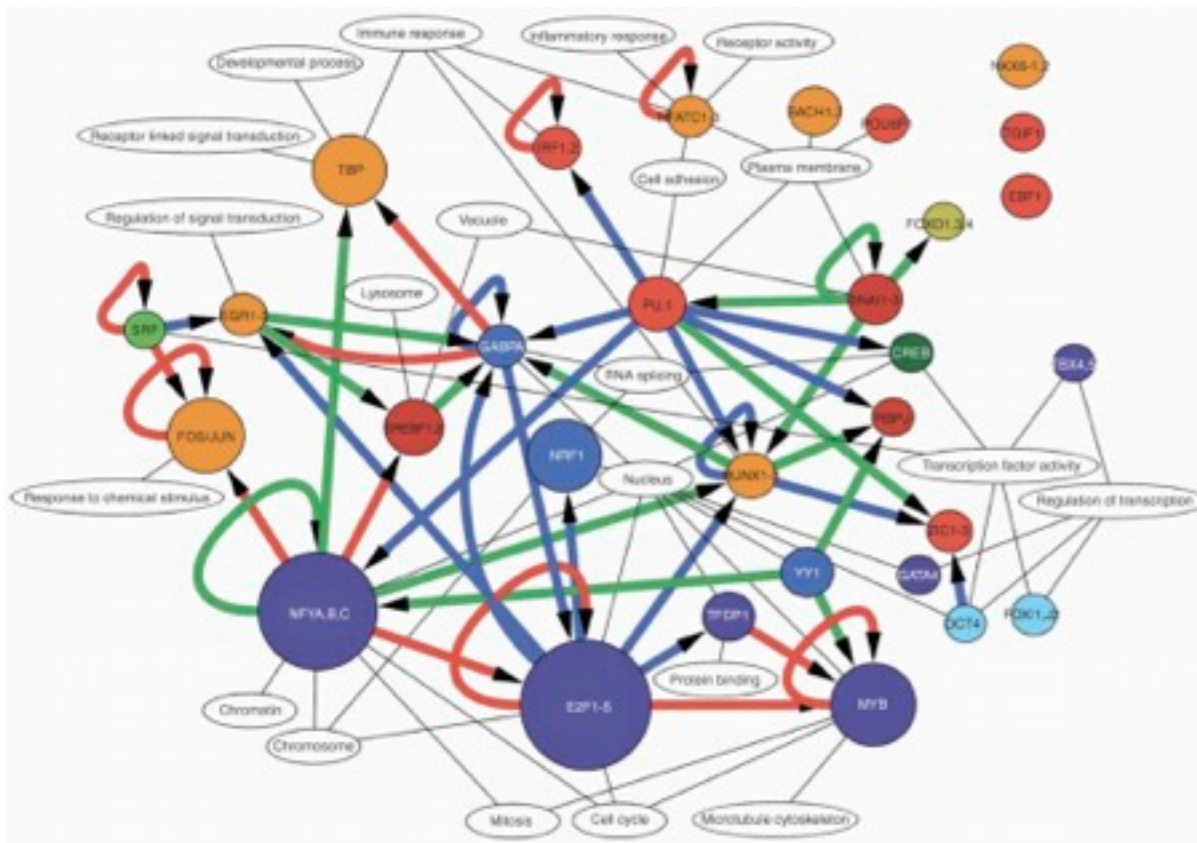


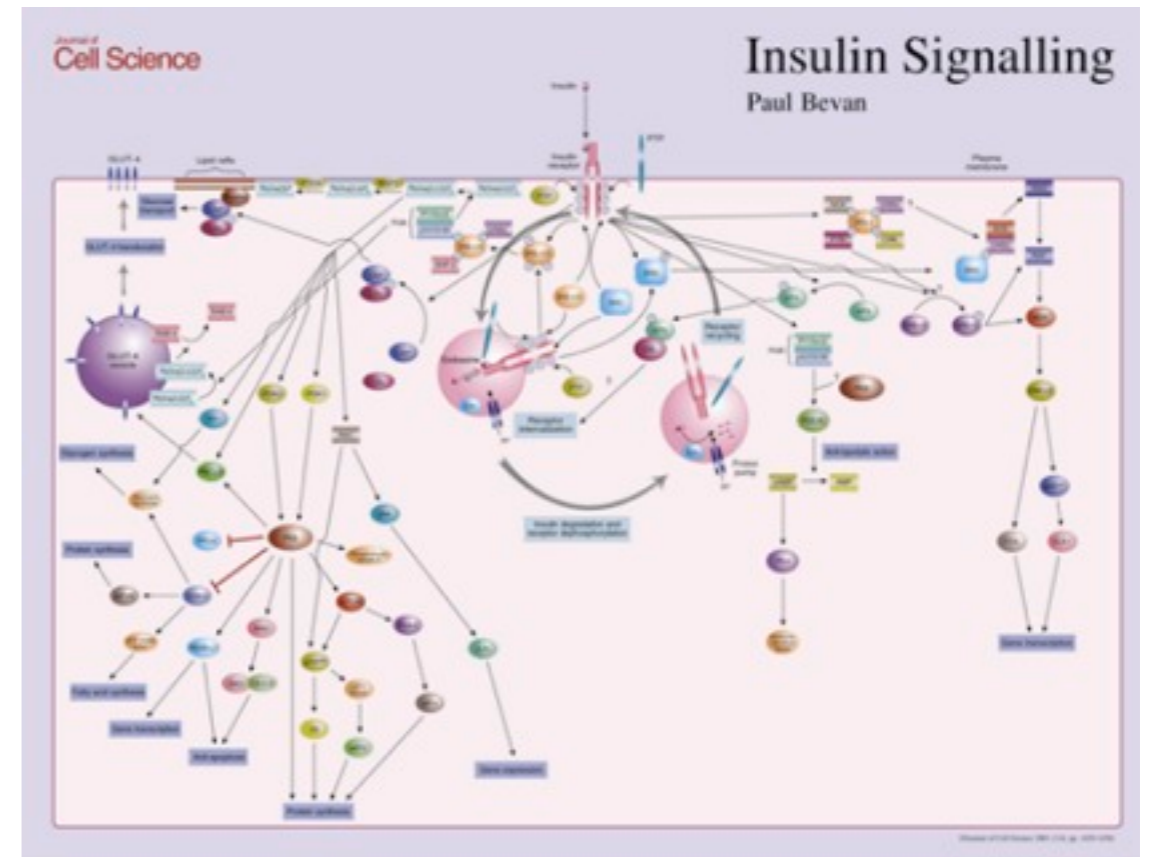
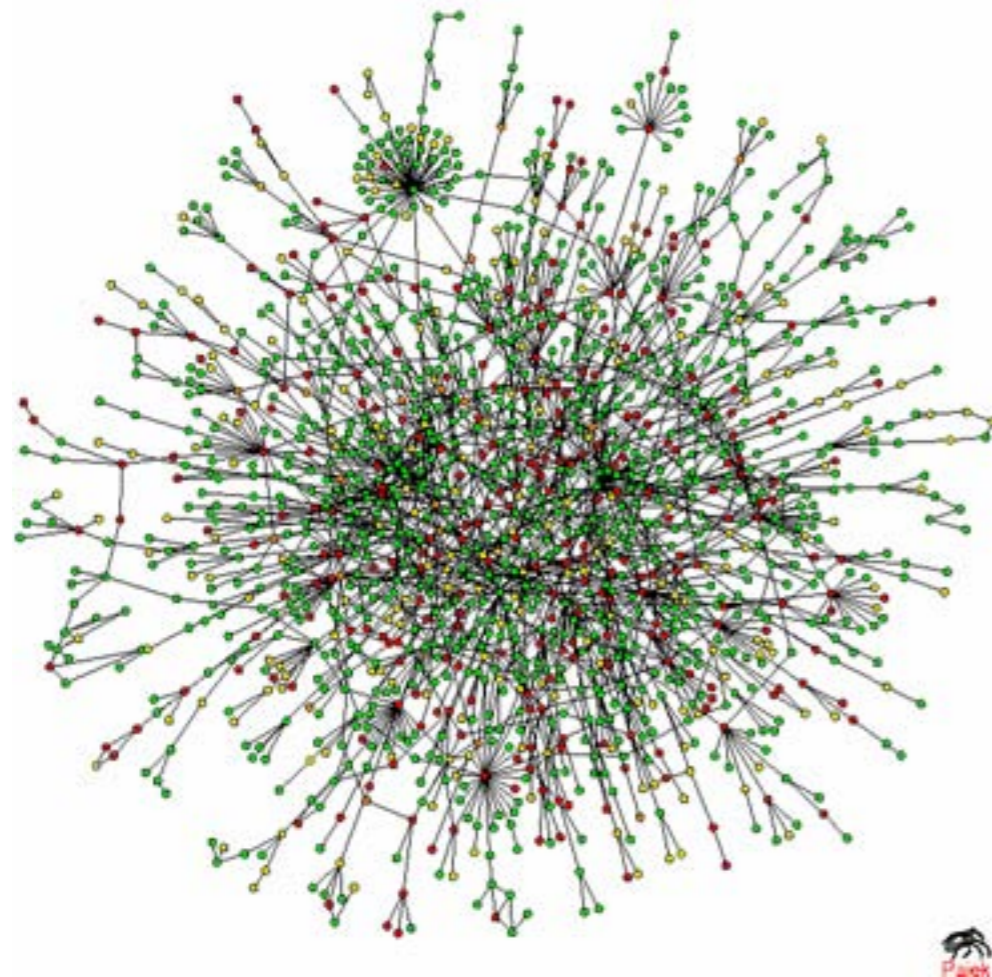
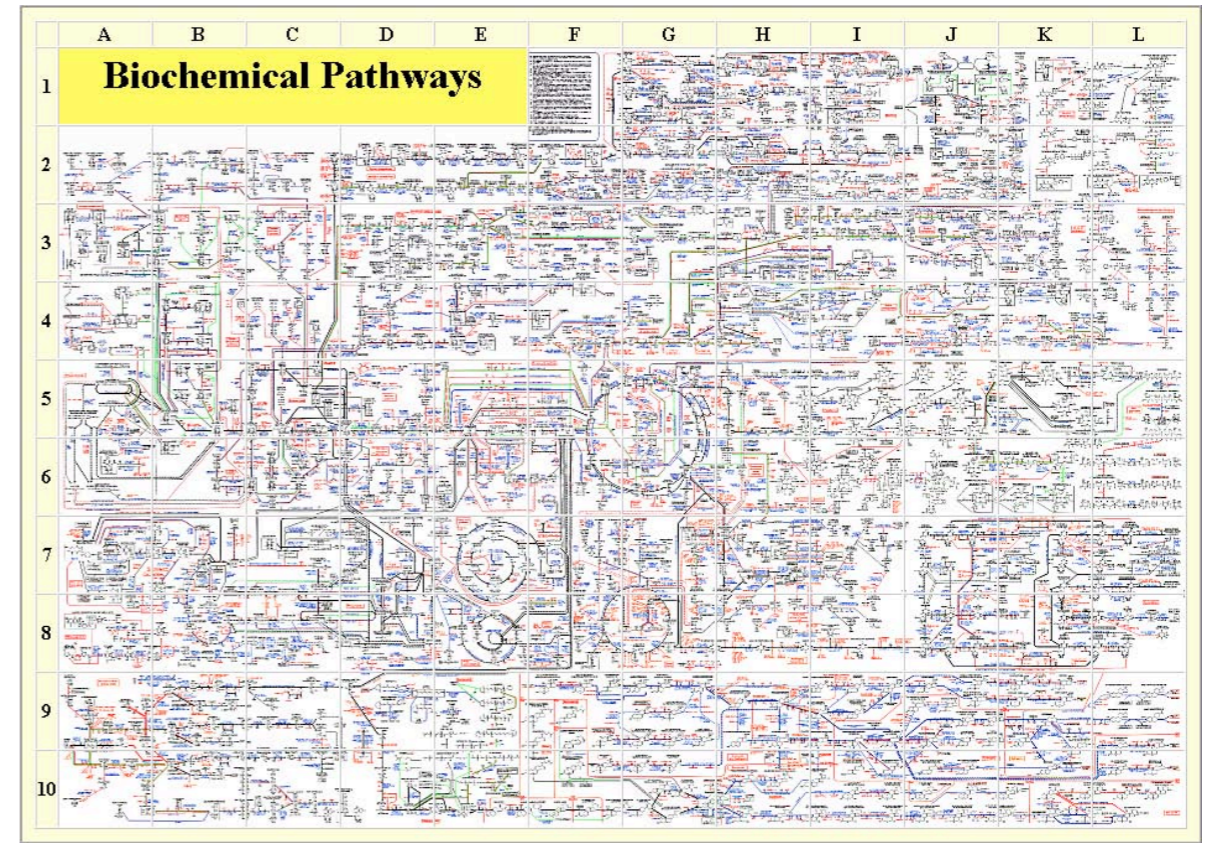
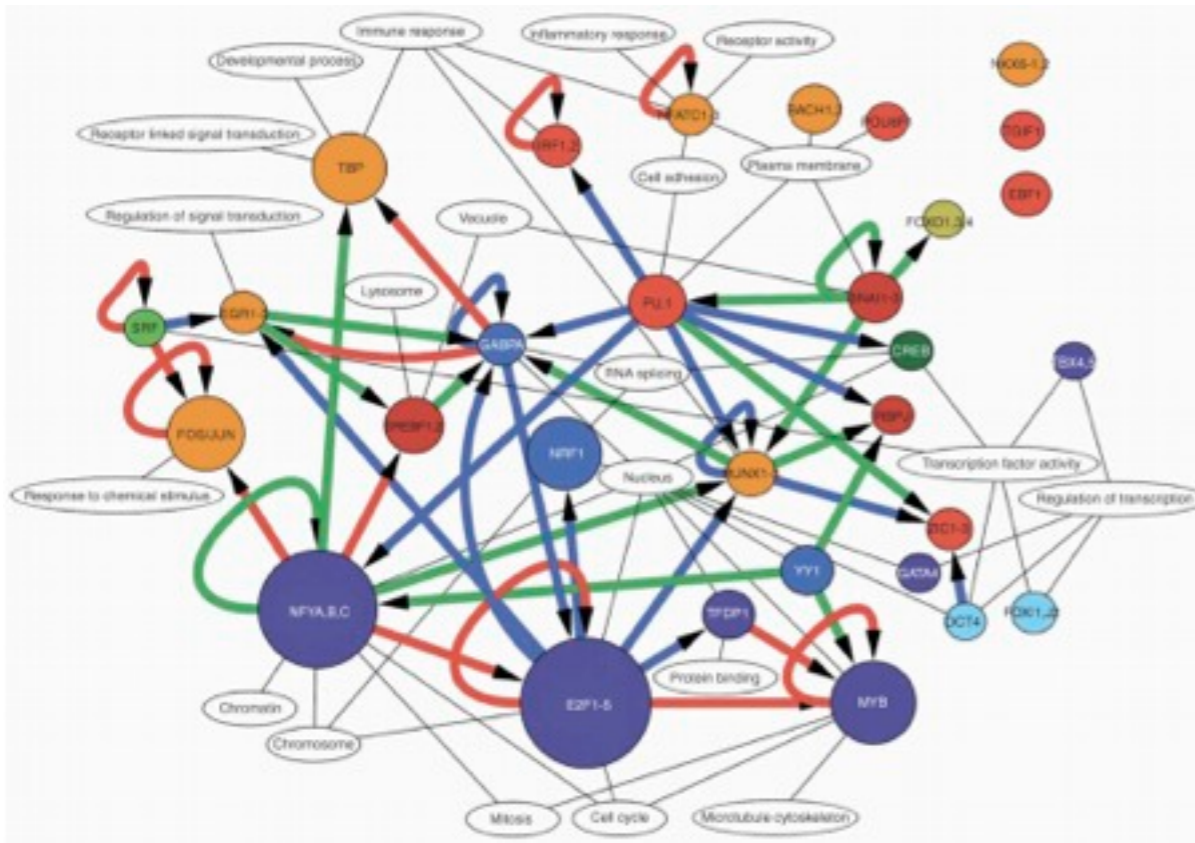
Links (edges) between nodes

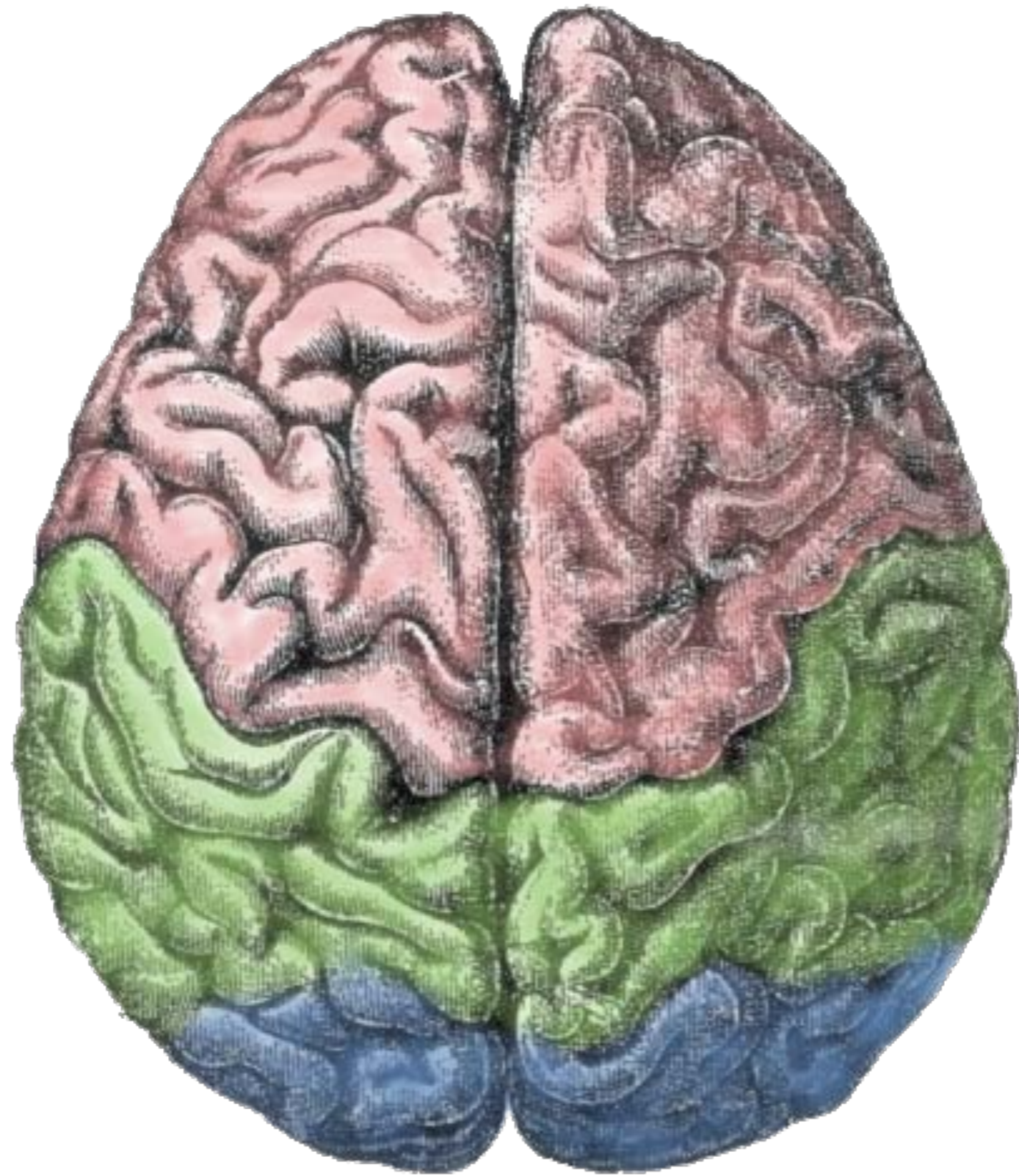
















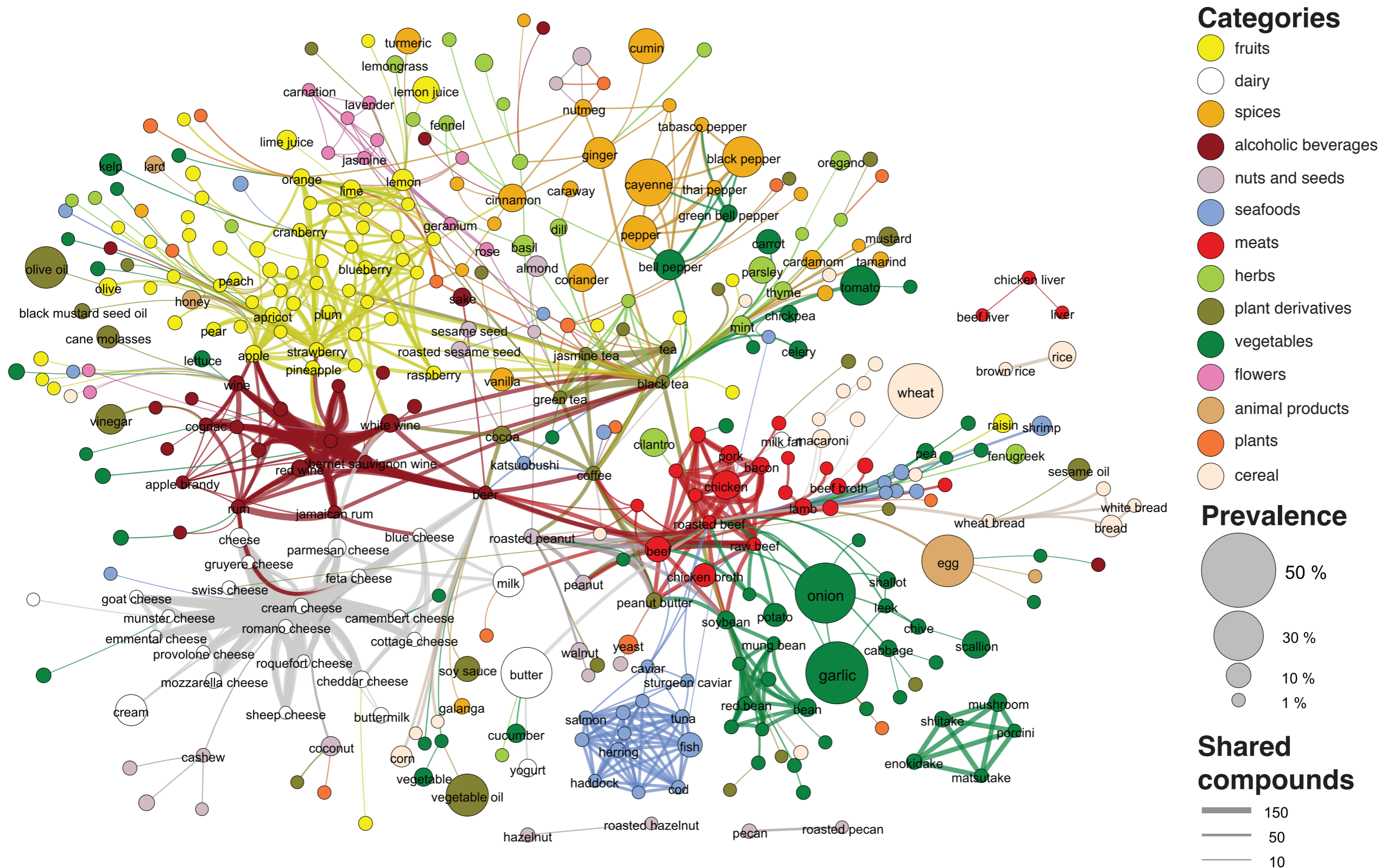
facebook

December 2010



Saturday, May 19, 12

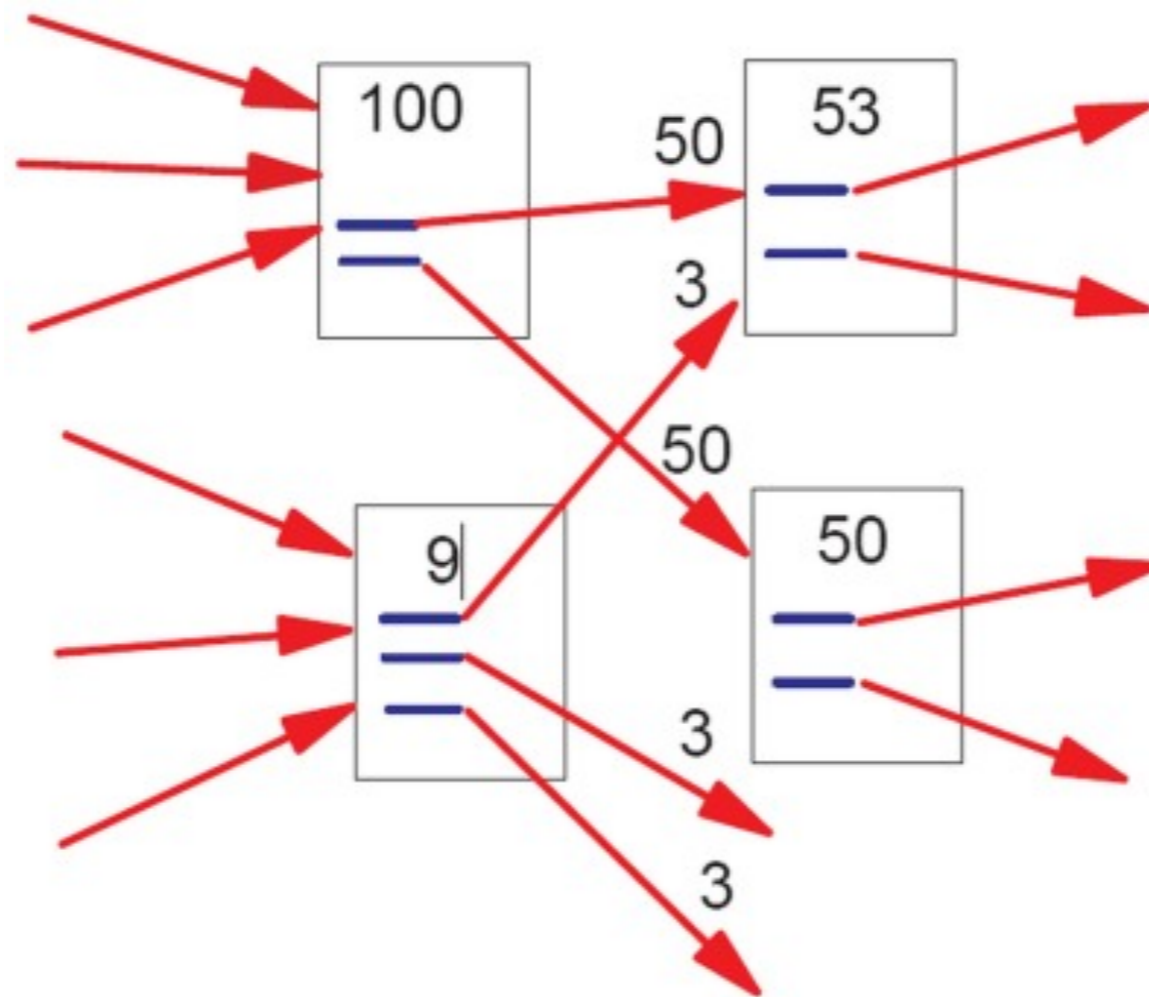




Y.-Y. Ahn, S. Ahnert, J. P. Bagrow, A.-L. Barabási, *Sci. Rep.* 2011

So what?

Google

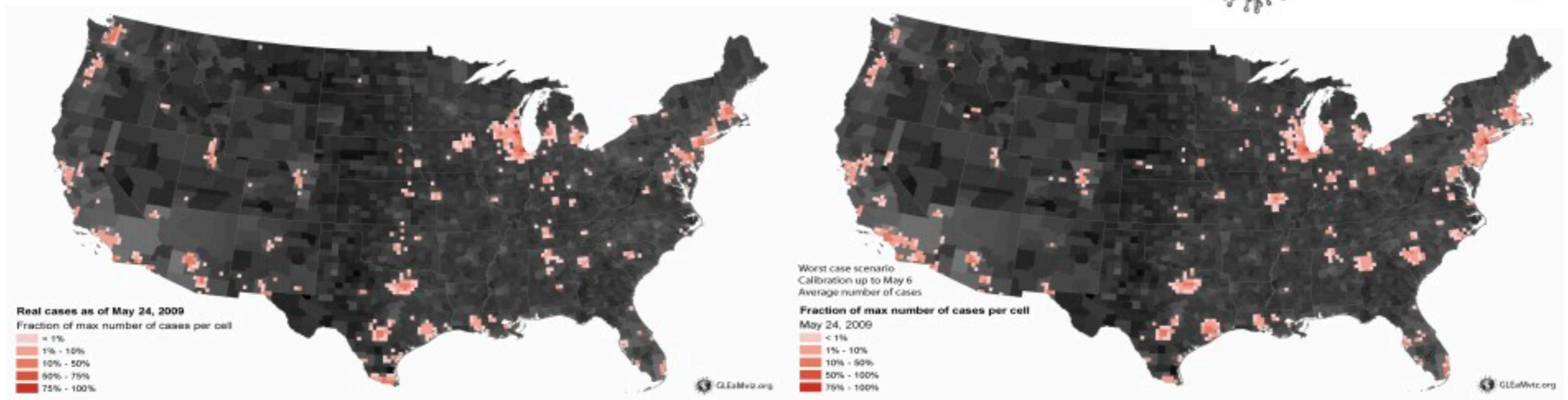


Pagerank =
Random walk problem on a network

H1N1 Pandemic prediction



GLEaMviz.org



Real

Prediction

Reaction-diffusion system with
transportation networks

Can we understand a
complex system

without knowing the **structure**
of it?

NETWORKS



Modular Structure

Local

Global

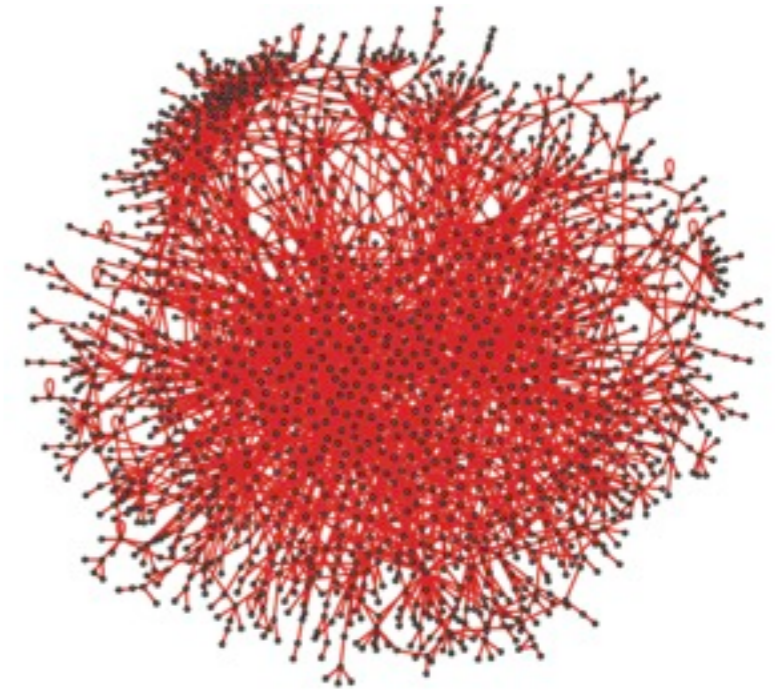




Local

Global

degree
clustering
motifs



Local

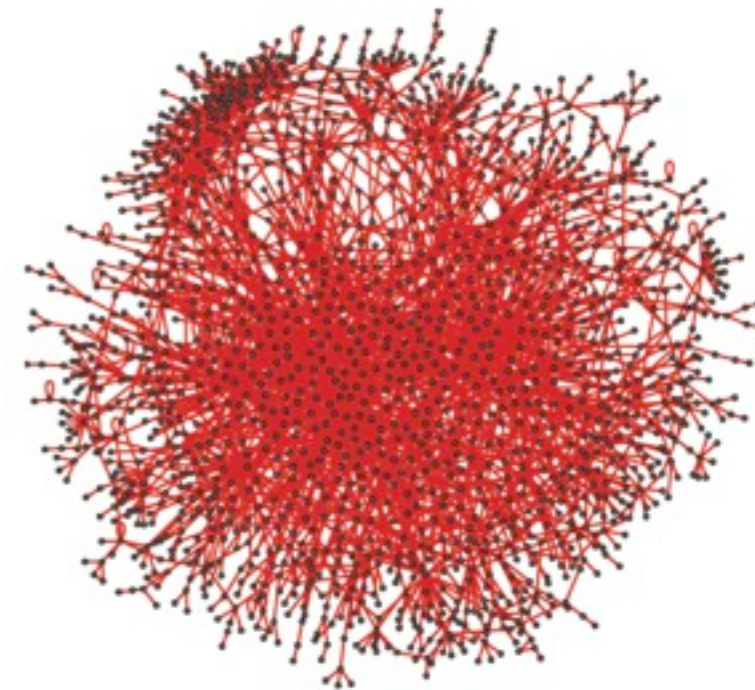
Global

motifs
degree
clustering

degree
distribution
Robustness



?



Local

Global

motifs
degree
clustering

degree
distribution
Robustness

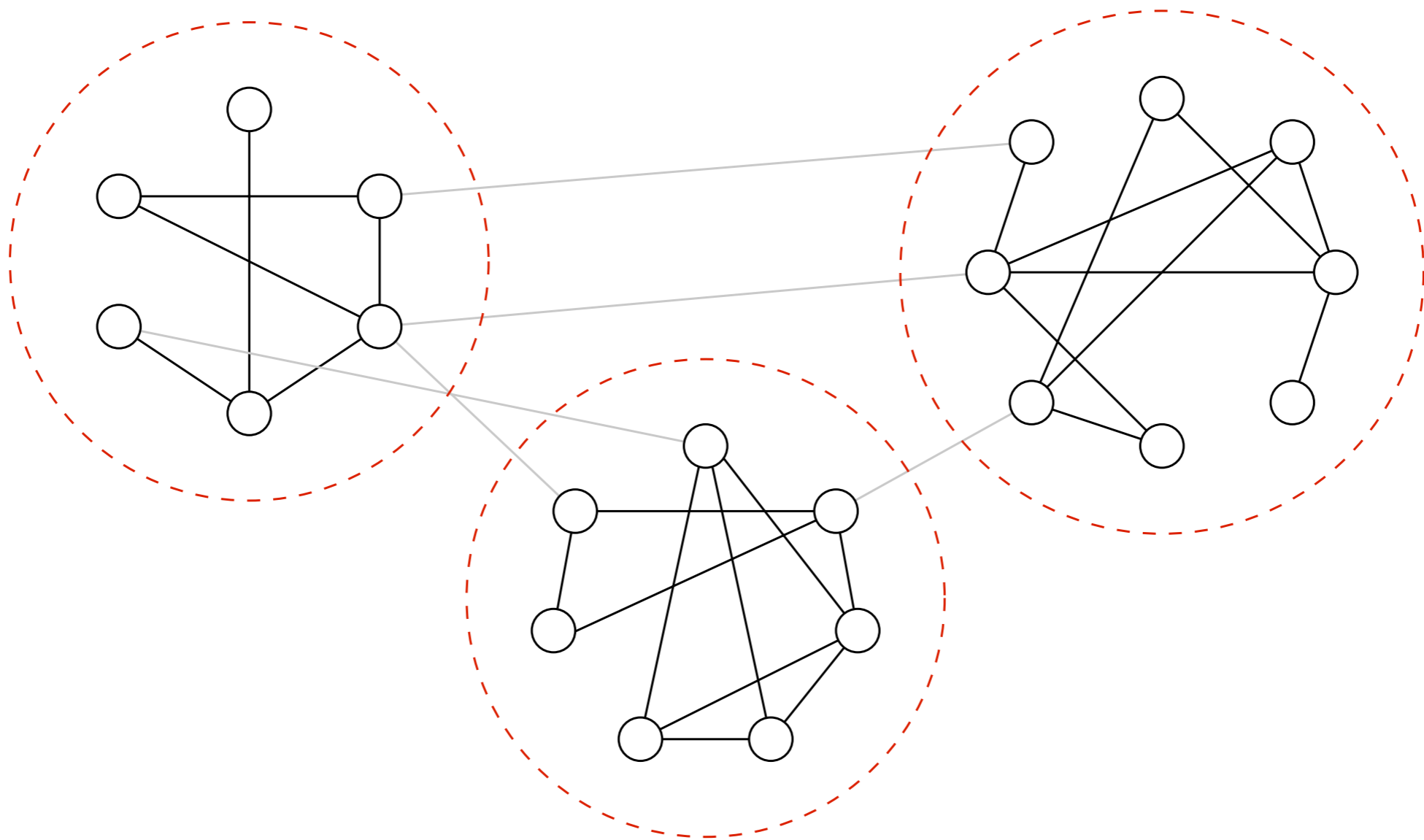
Network Communities





“a group of densely interconnected nodes”

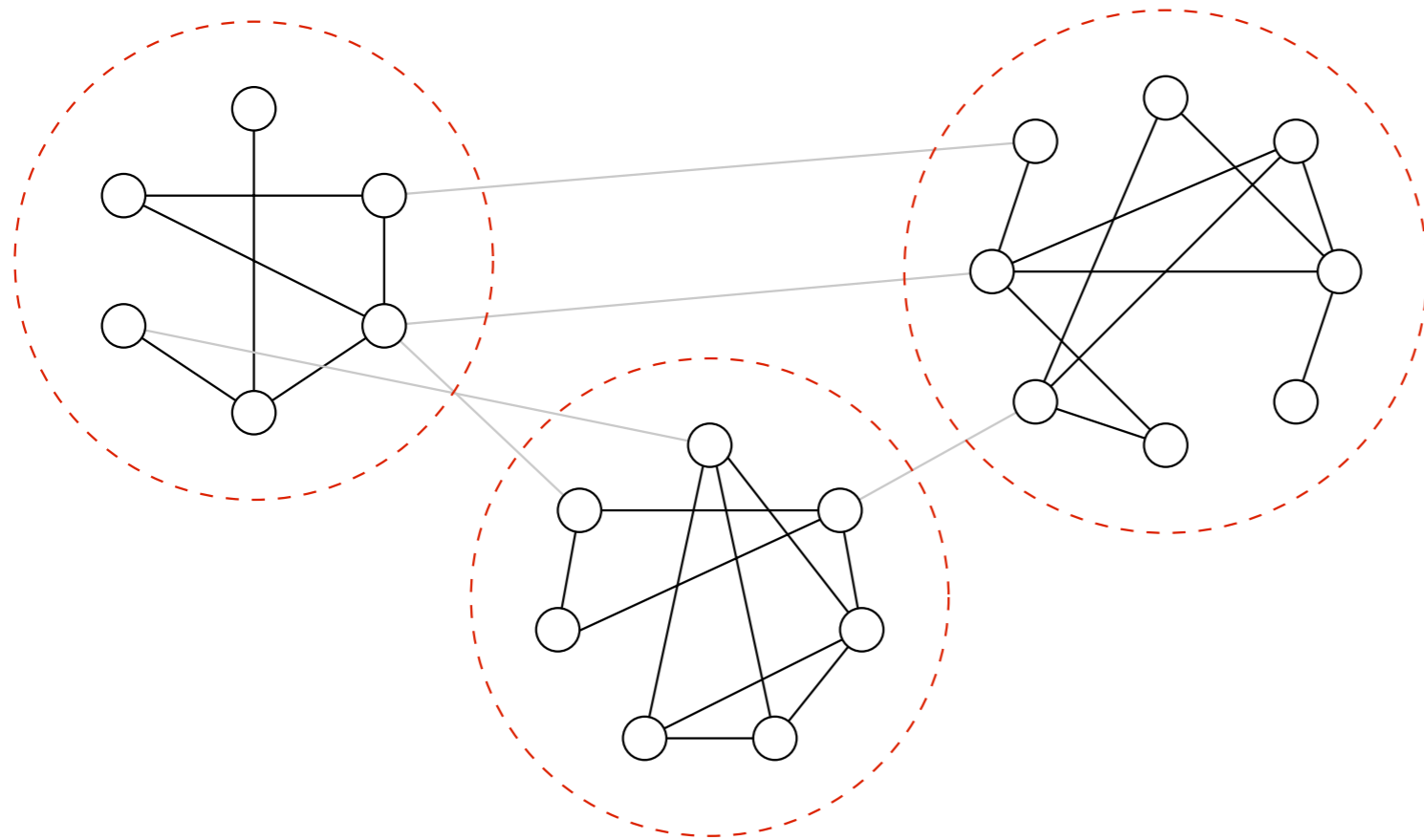
“a group of densely interconnected nodes”



Modularity

$$Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j)$$

M. Girvan and M. E. J. Newman, *PNAS* (2002)



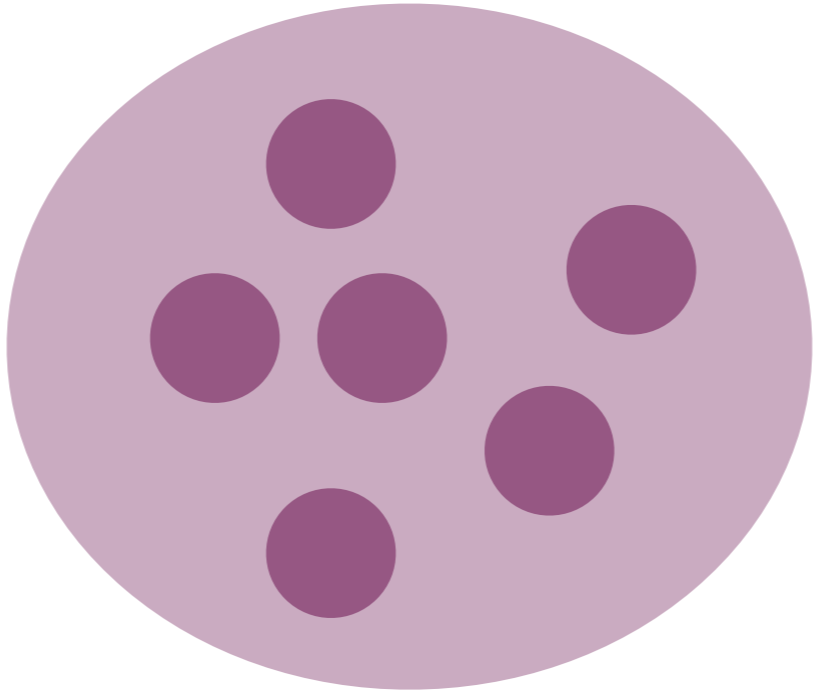
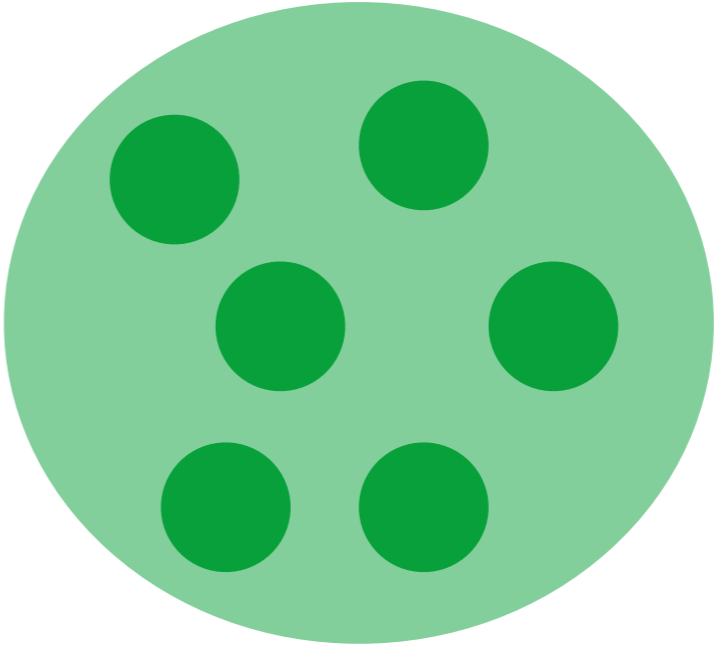
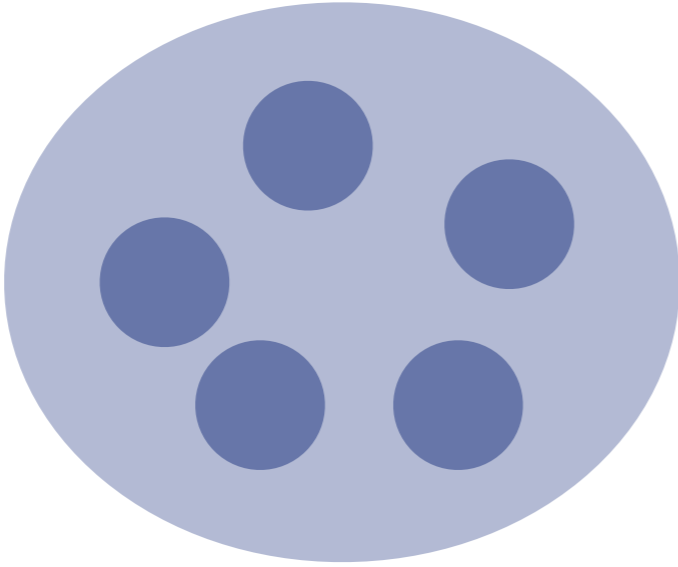
$$Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j)$$

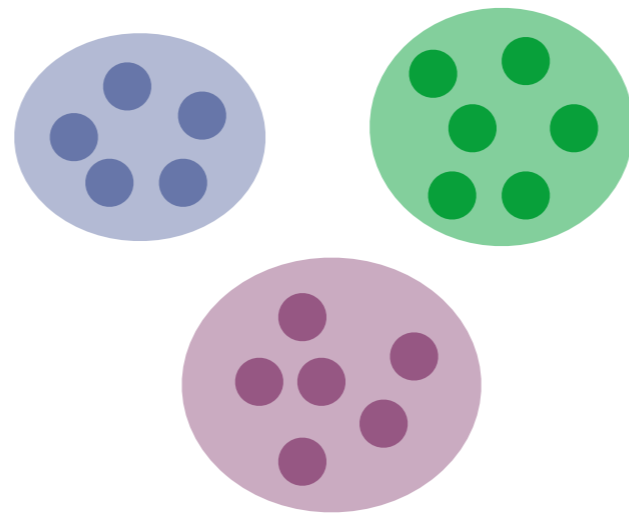
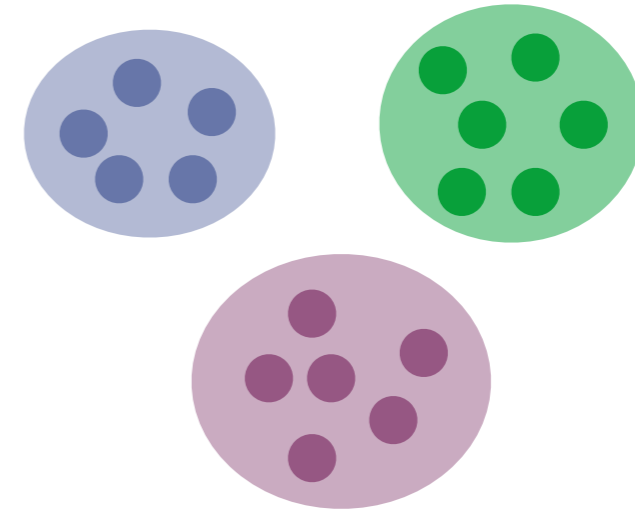
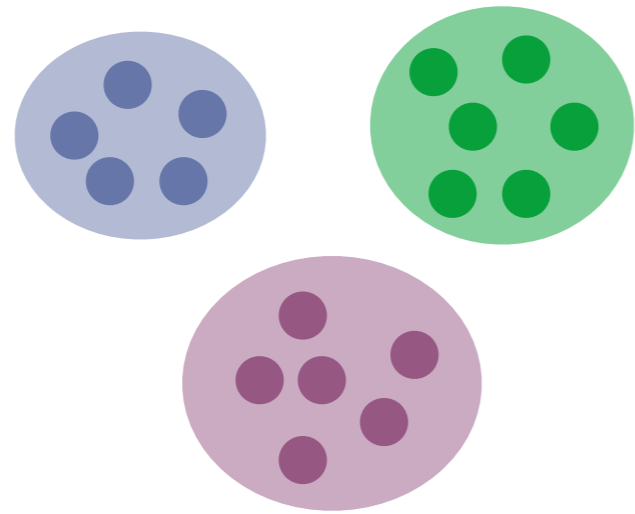
Hundreds of community detection methods

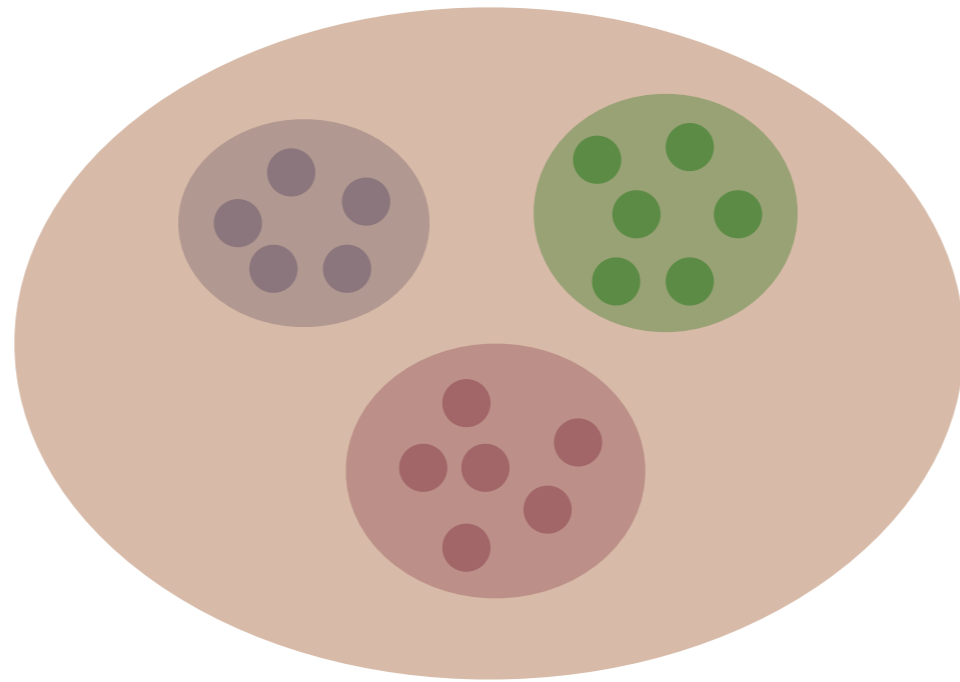
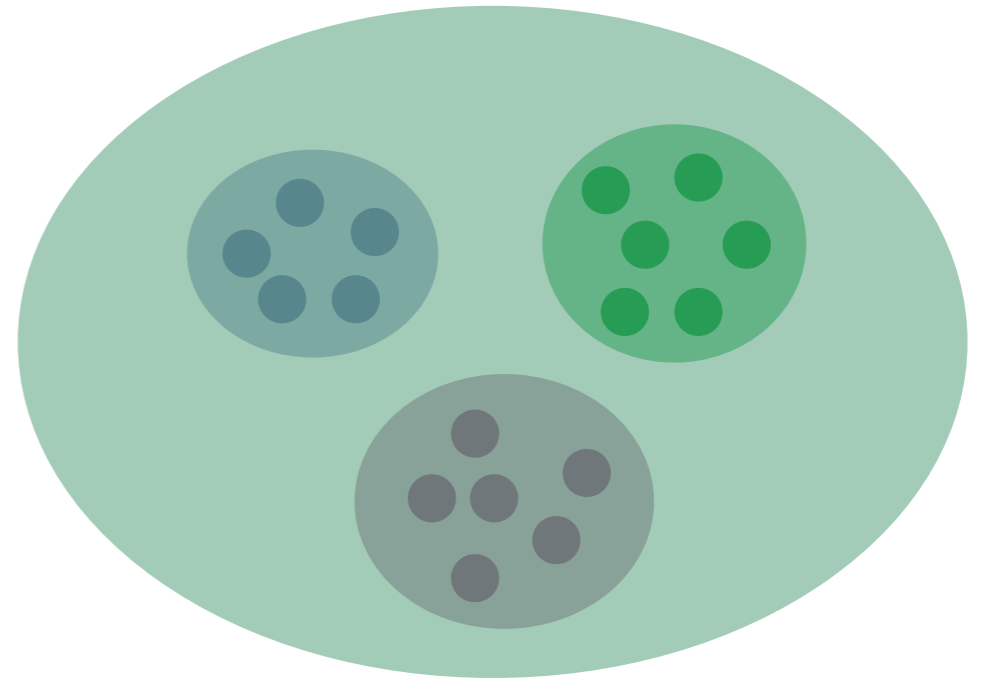
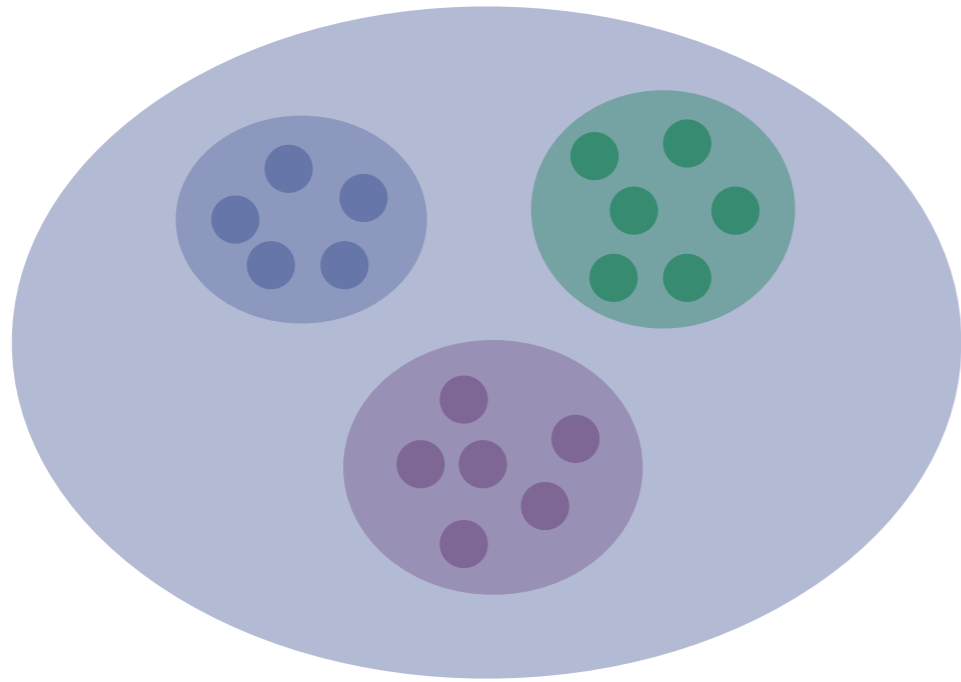
Then, why bother?

Hierarchy & Overlap

Hierarchy

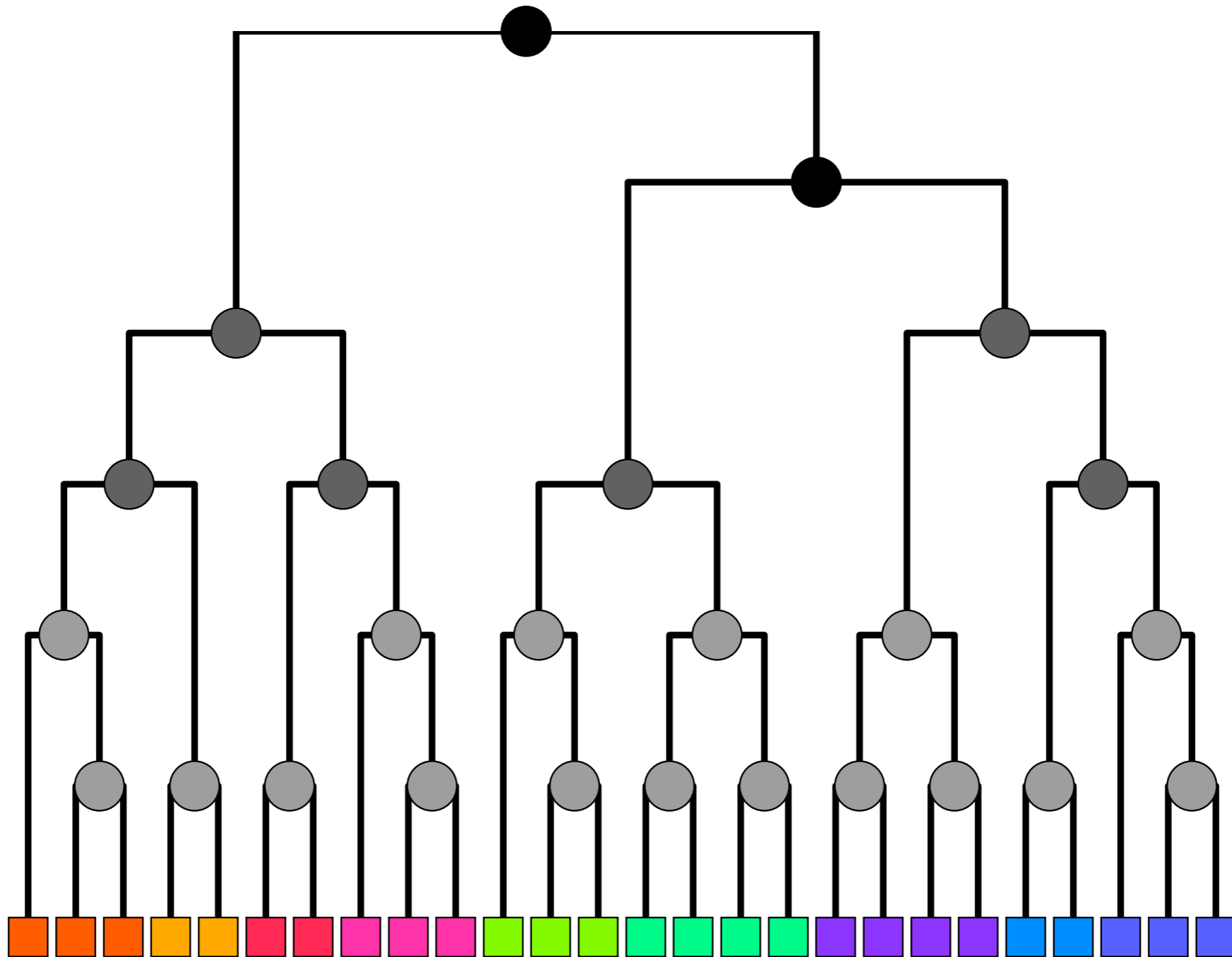




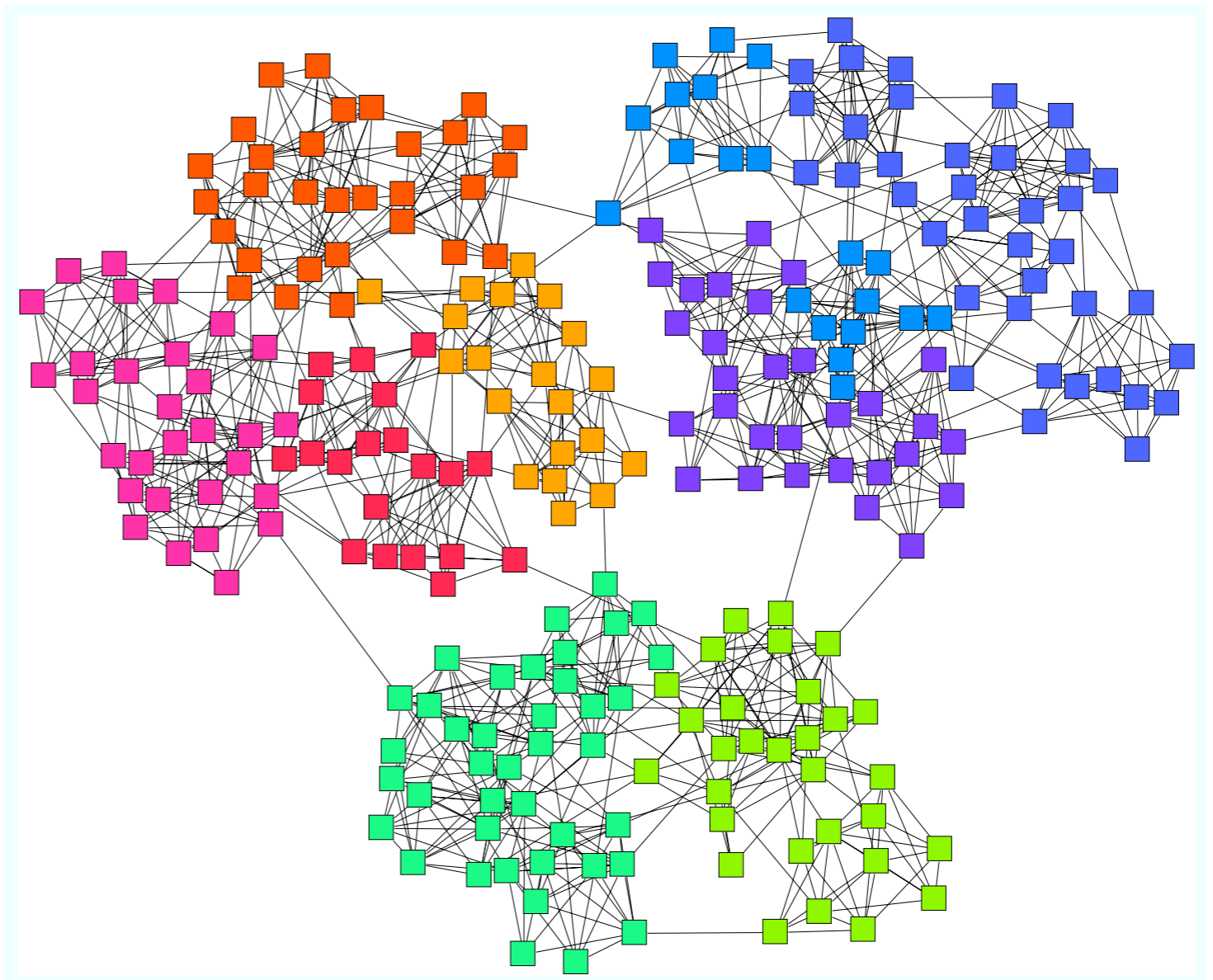
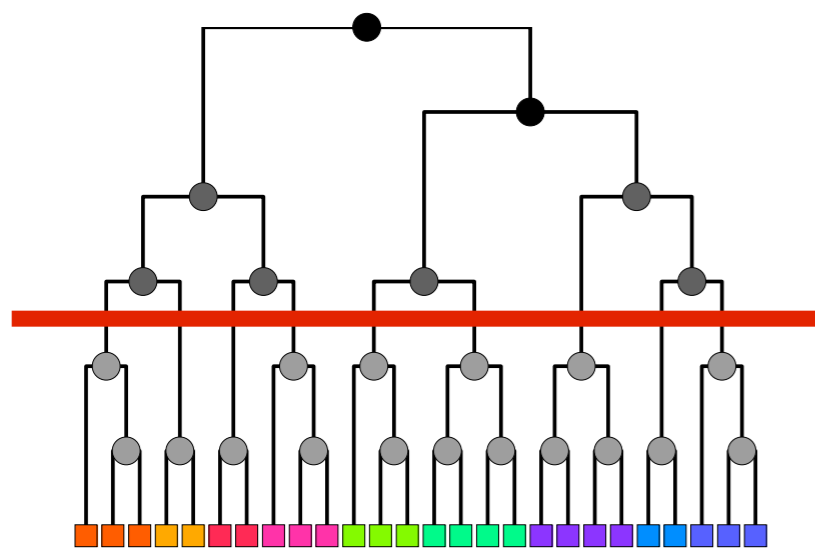


**Hierarchy implies
communities.**

Hierarchical Random Graph model



A. Clauset, C. Moore, and M. E. J. Newman, *Nature* (2008)



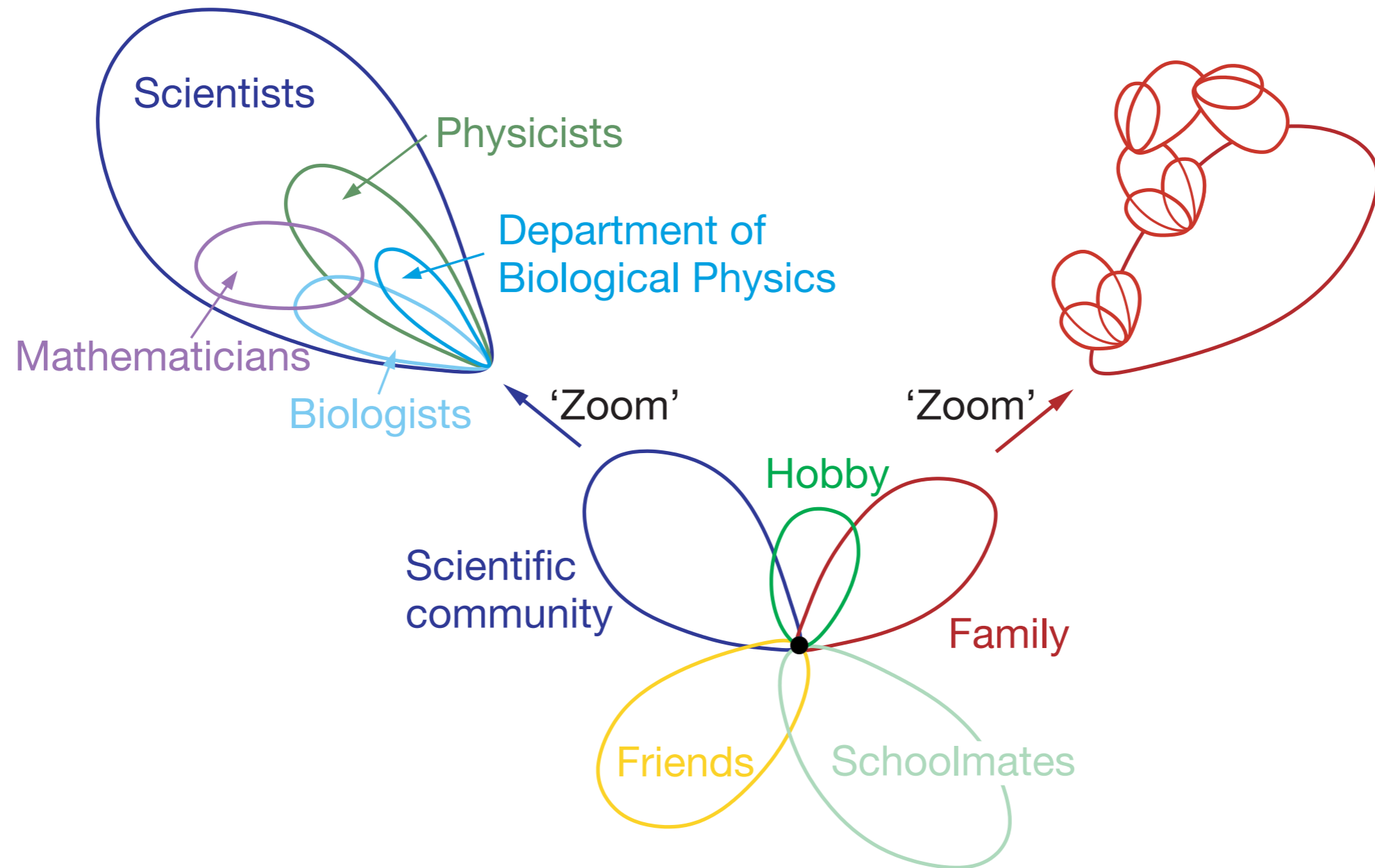
A. Clauset, C. Moore, and M. E. J. Newman, *Nature* (2008)

Hierarchical community structure

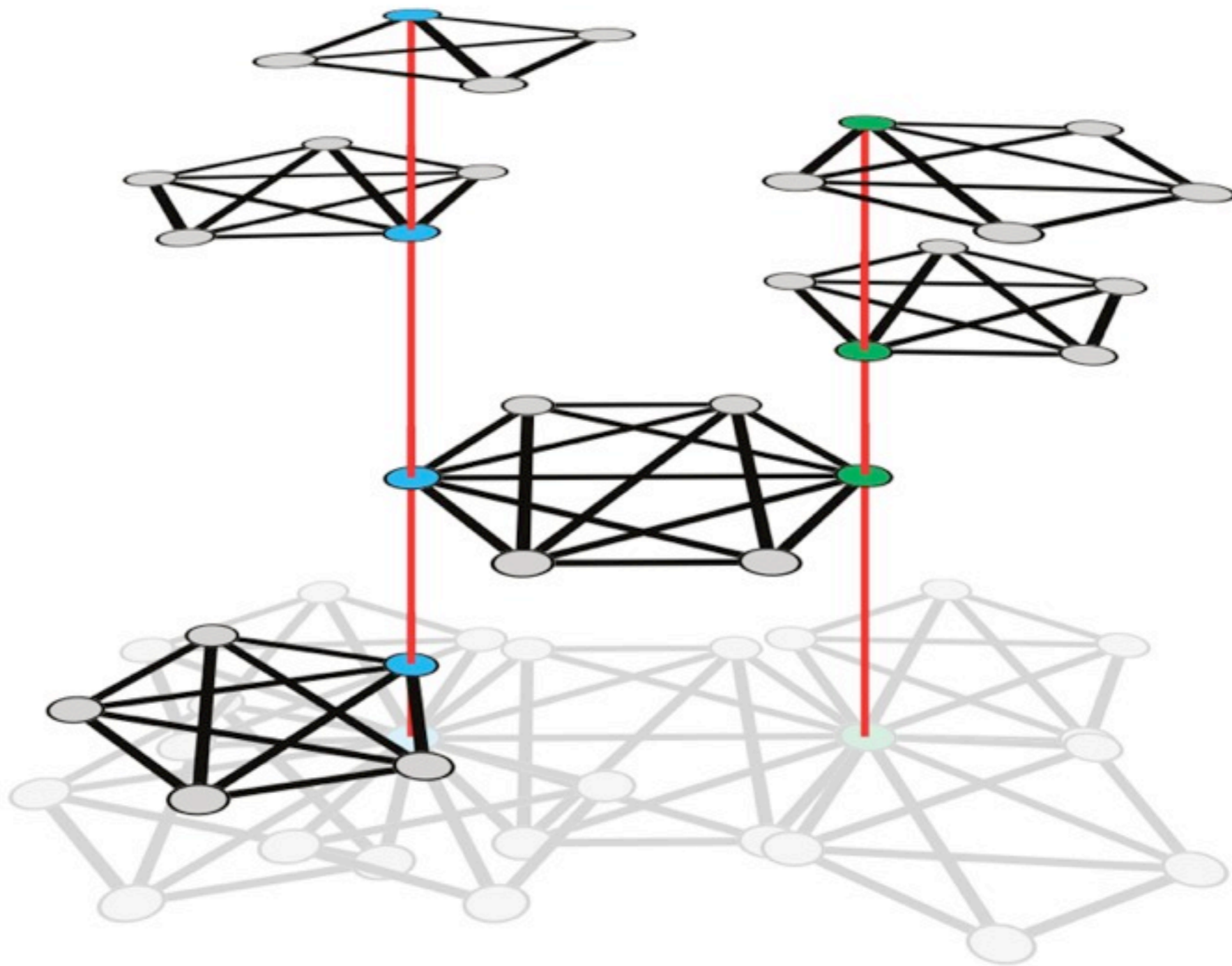
Hierarchy → Communities

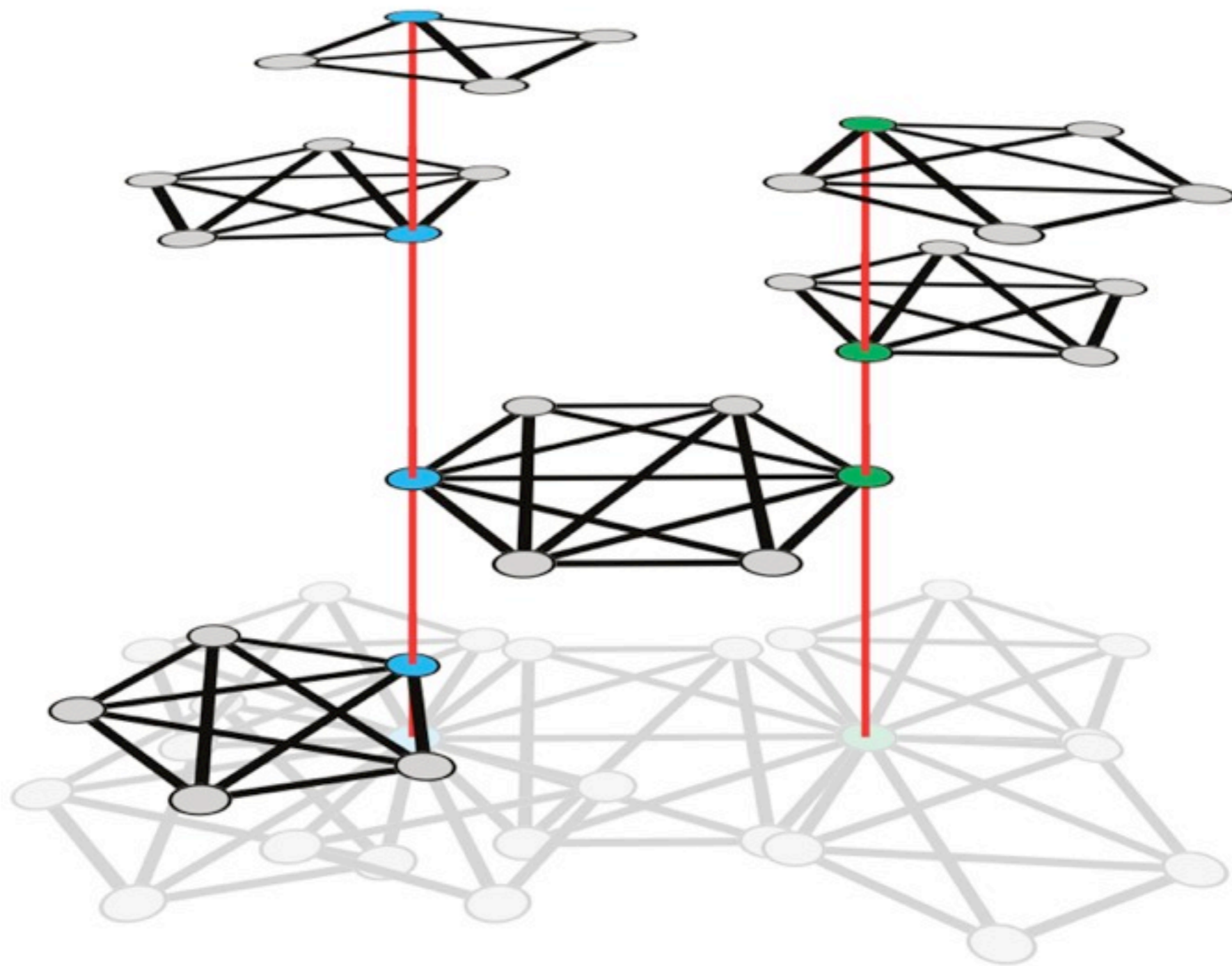


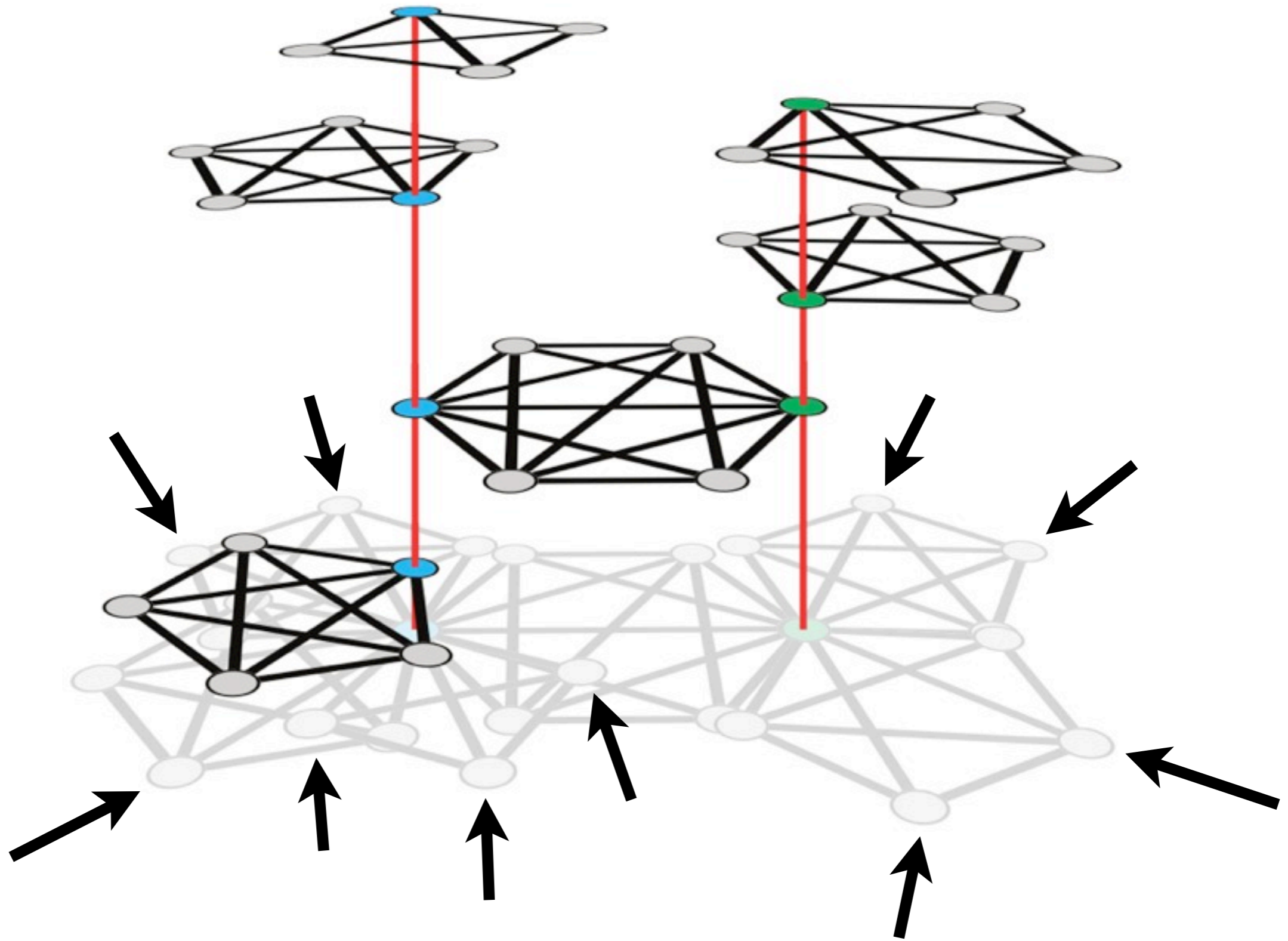
BUT,



G. Palla, I. Derényi, I. Farkas & T. Vicsek, *Nature* (2005)

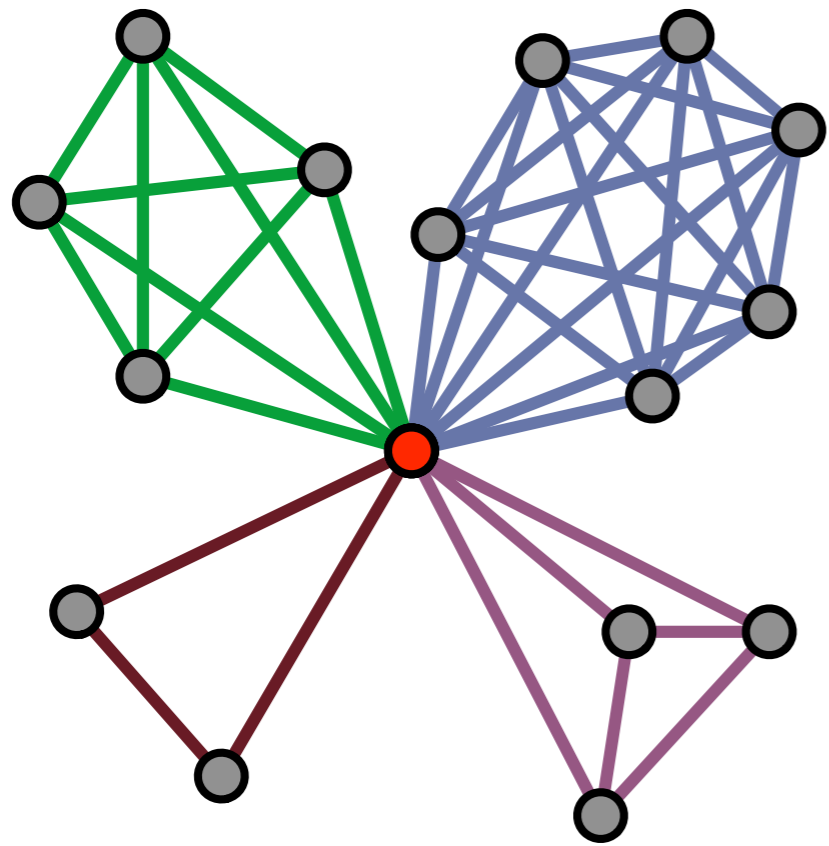






**Overlap is
pervasive.**

Overlap is
pervasive.



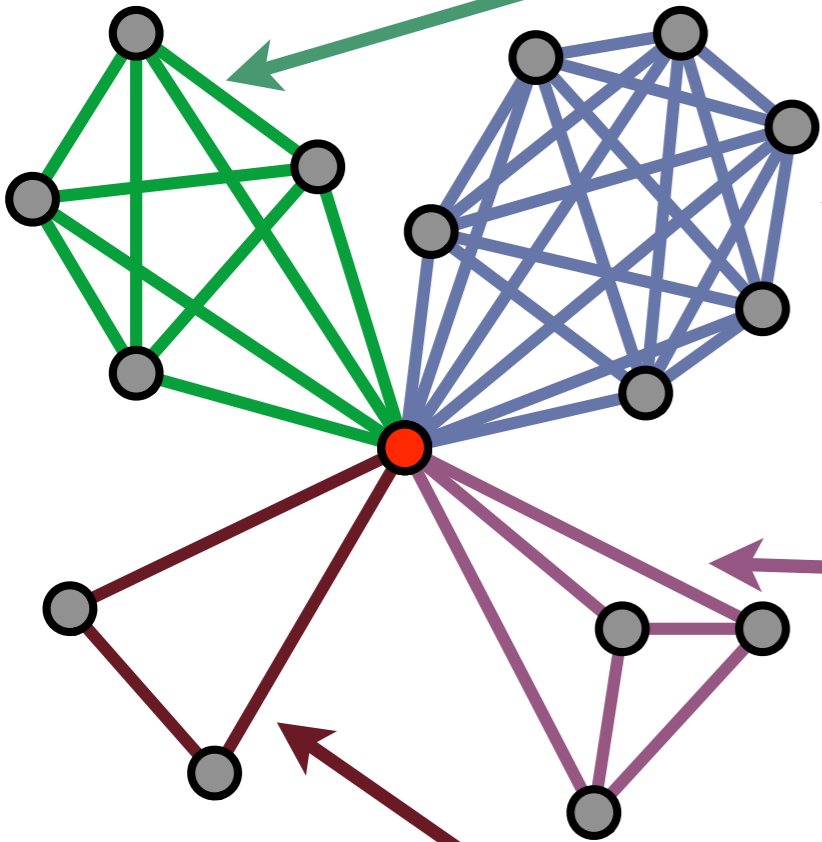
Multiple Contexts

Multiple Contexts

Multiple Contexts

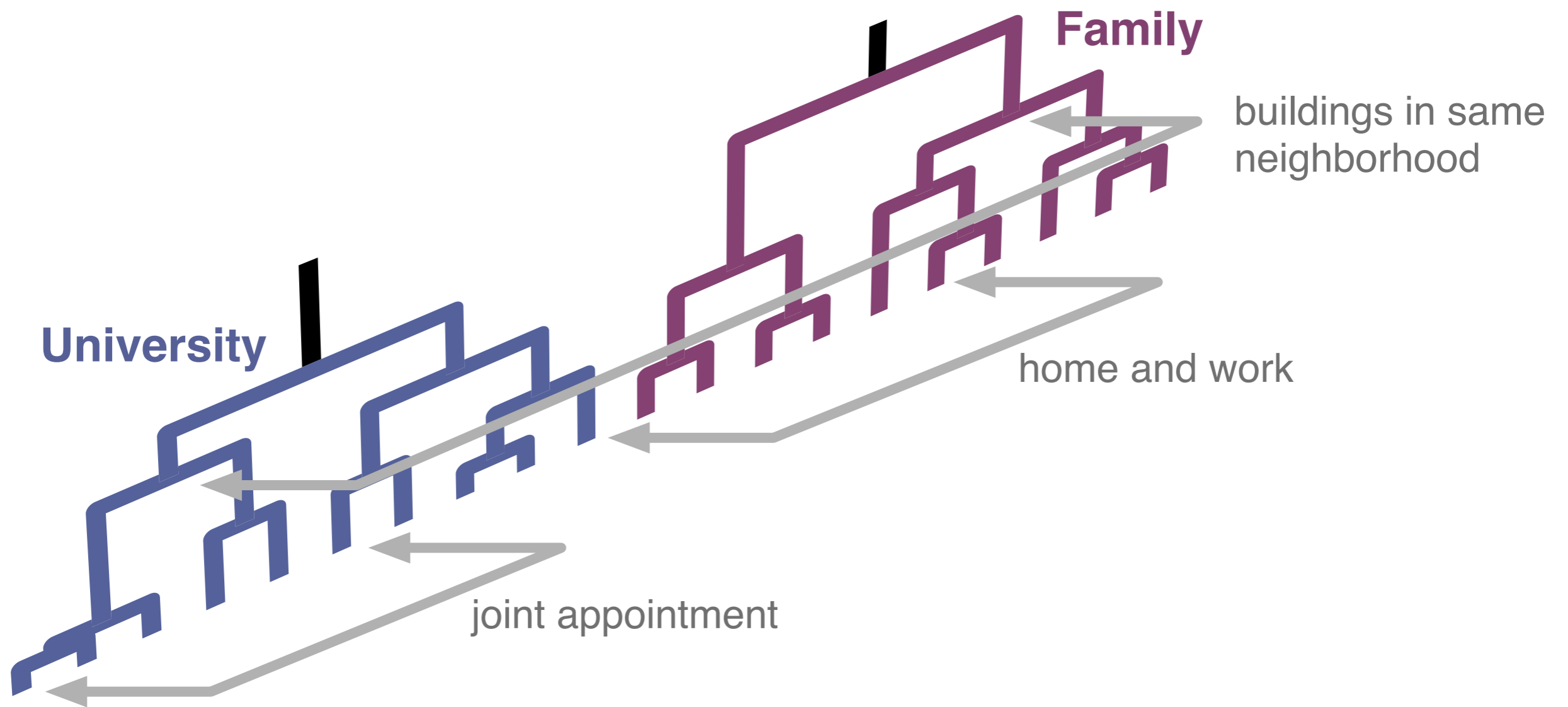
Multiple Contexts

Multiple Contexts



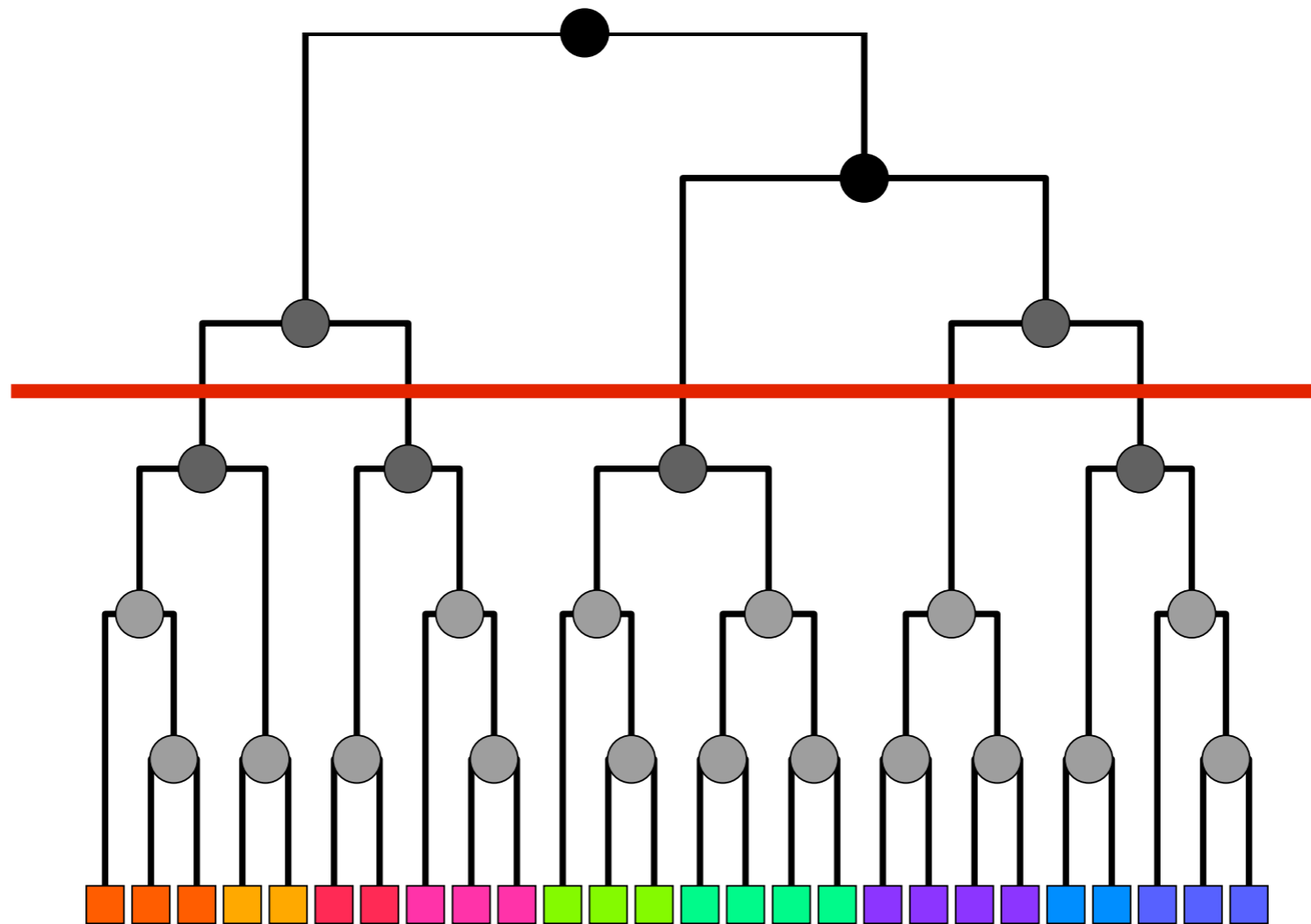
“Role theory”

<http://www.youtube.com/watch?v=SxuYdzs4SS8>

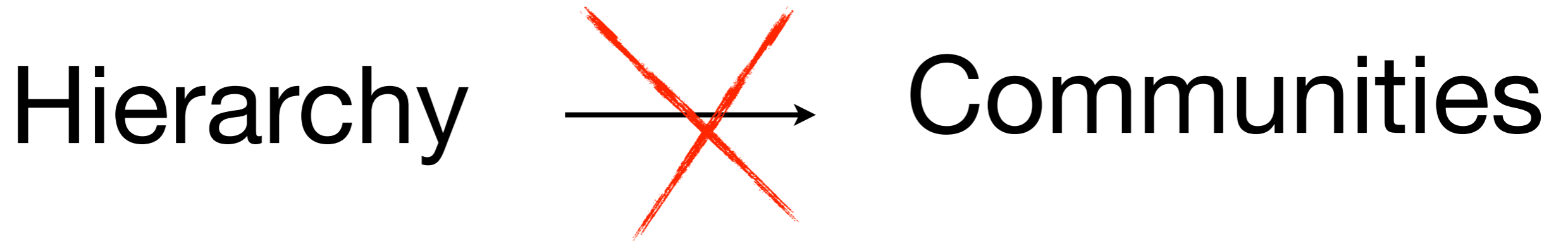


It is **impossible** to obtain a single dendrogram.

Hierarchy implies disjoint communities.

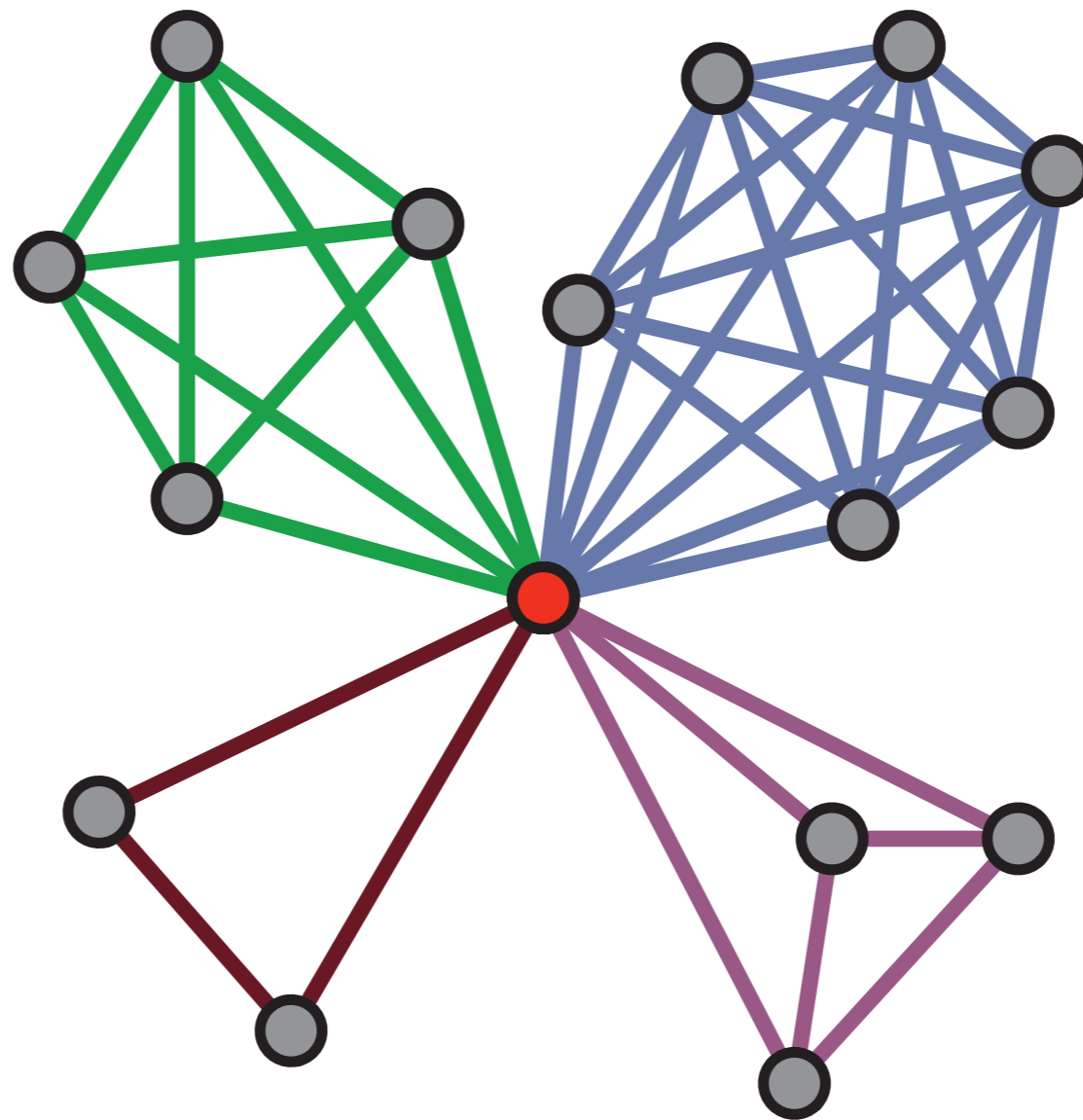


Hierarchical community structure

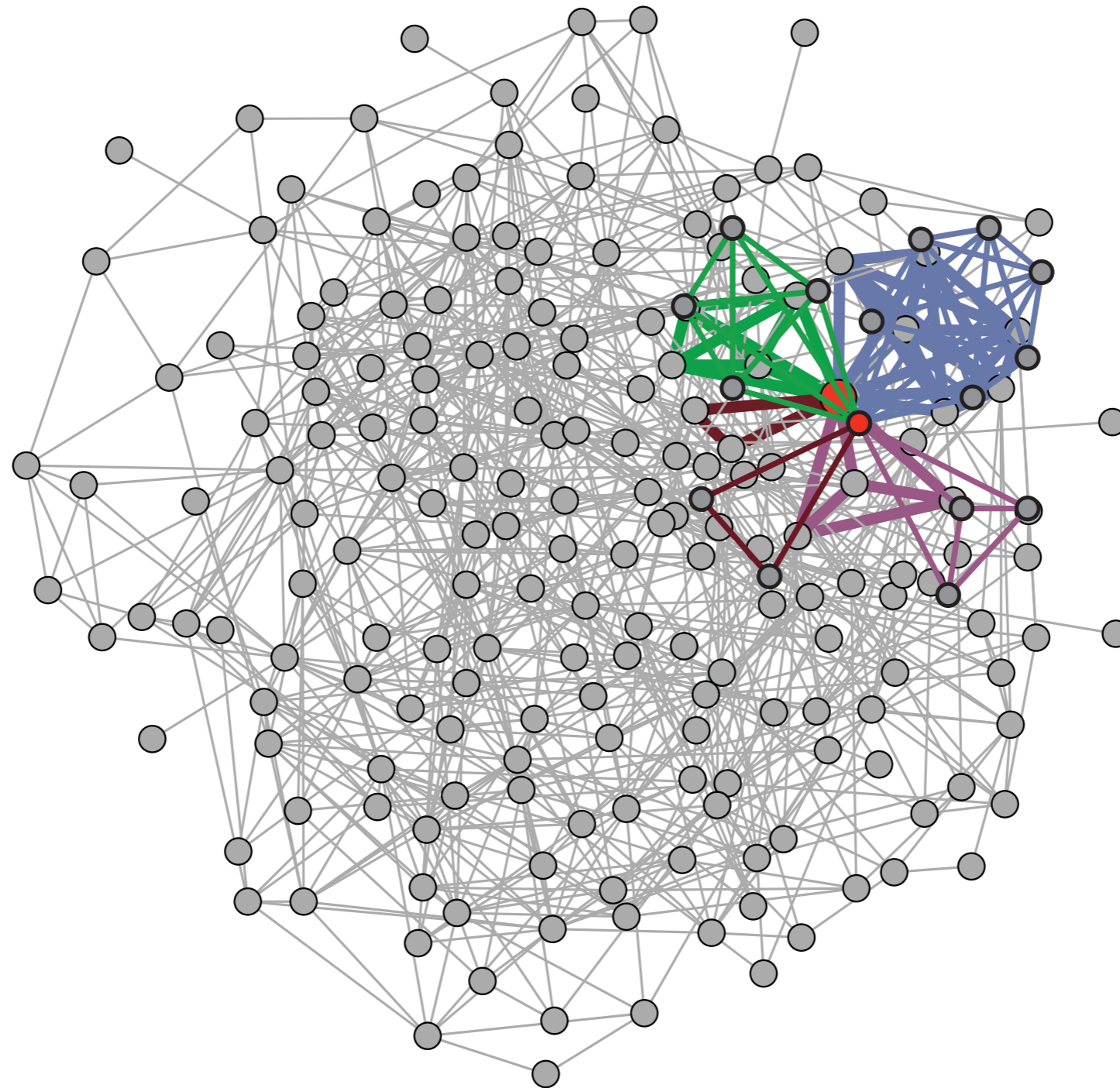


Another consequence

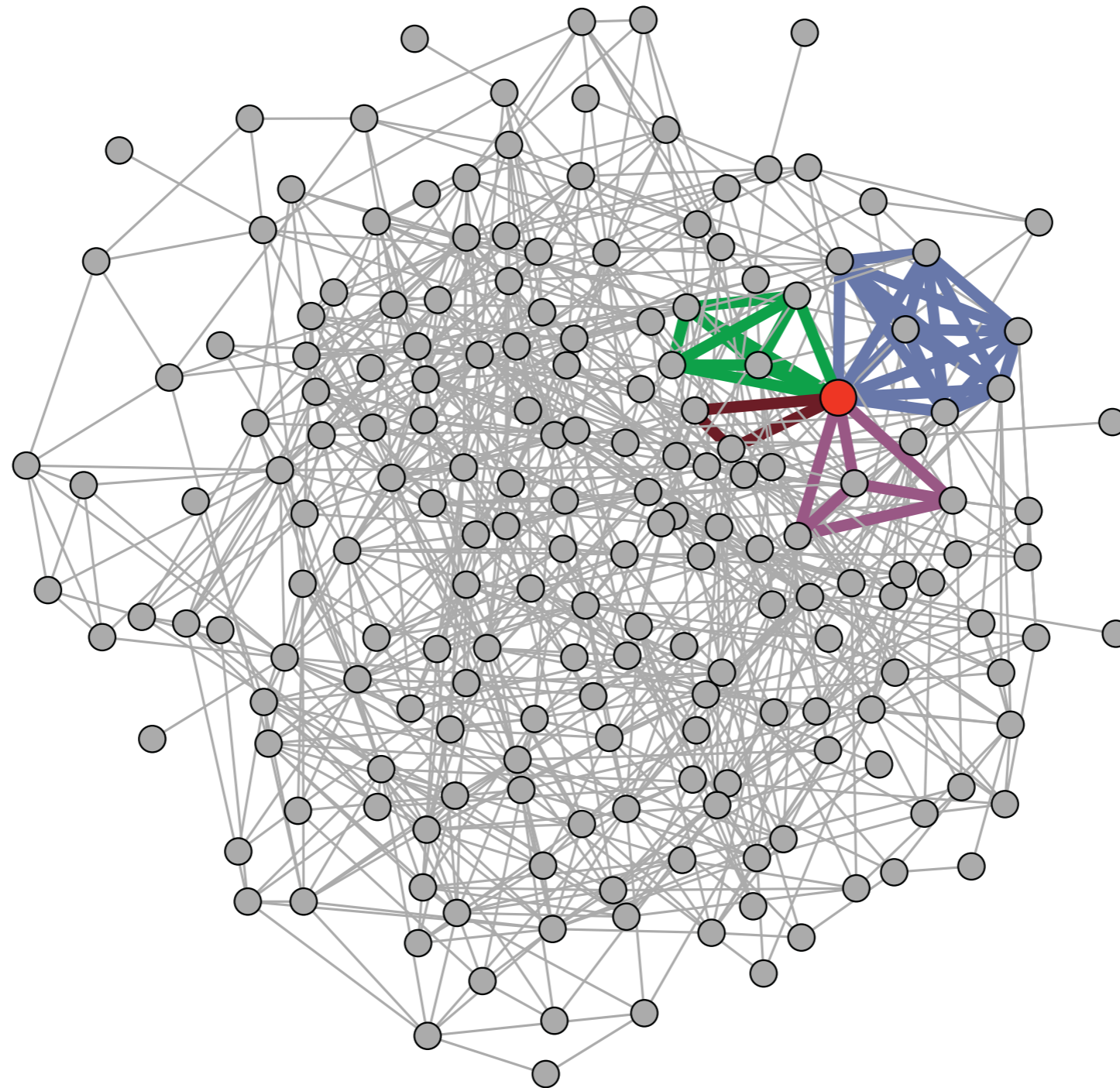
Simple local structure

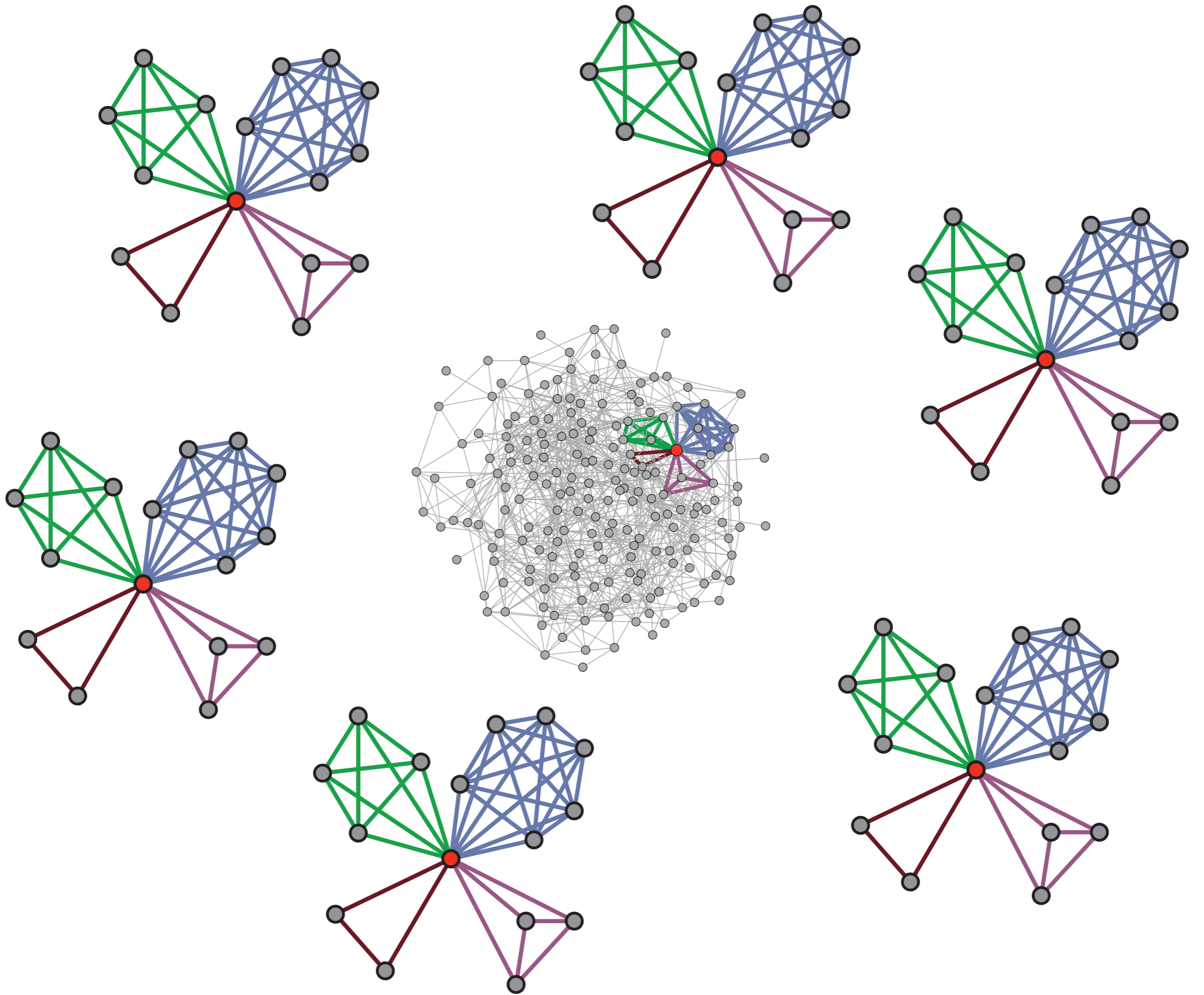


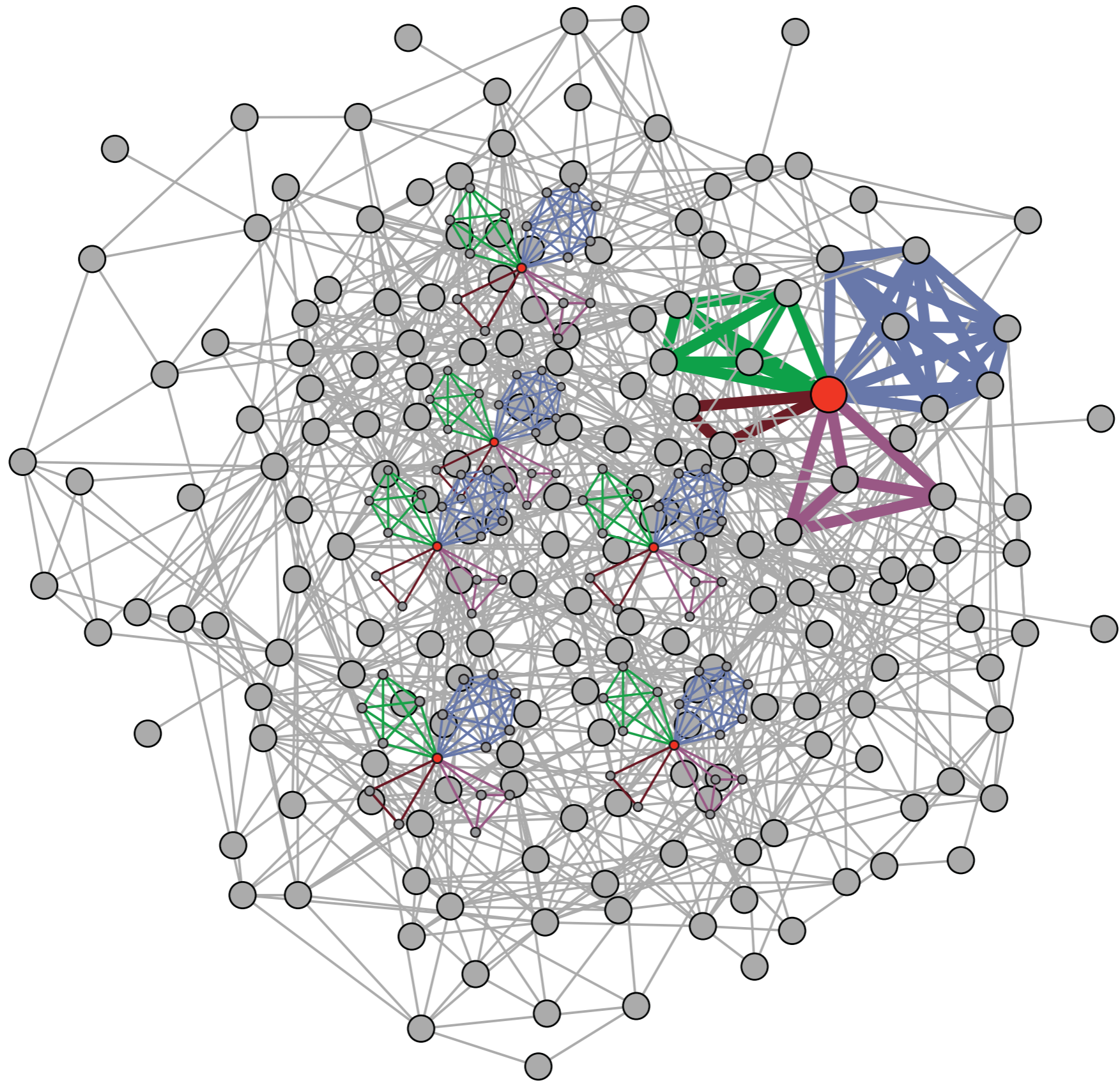
Complex global structure

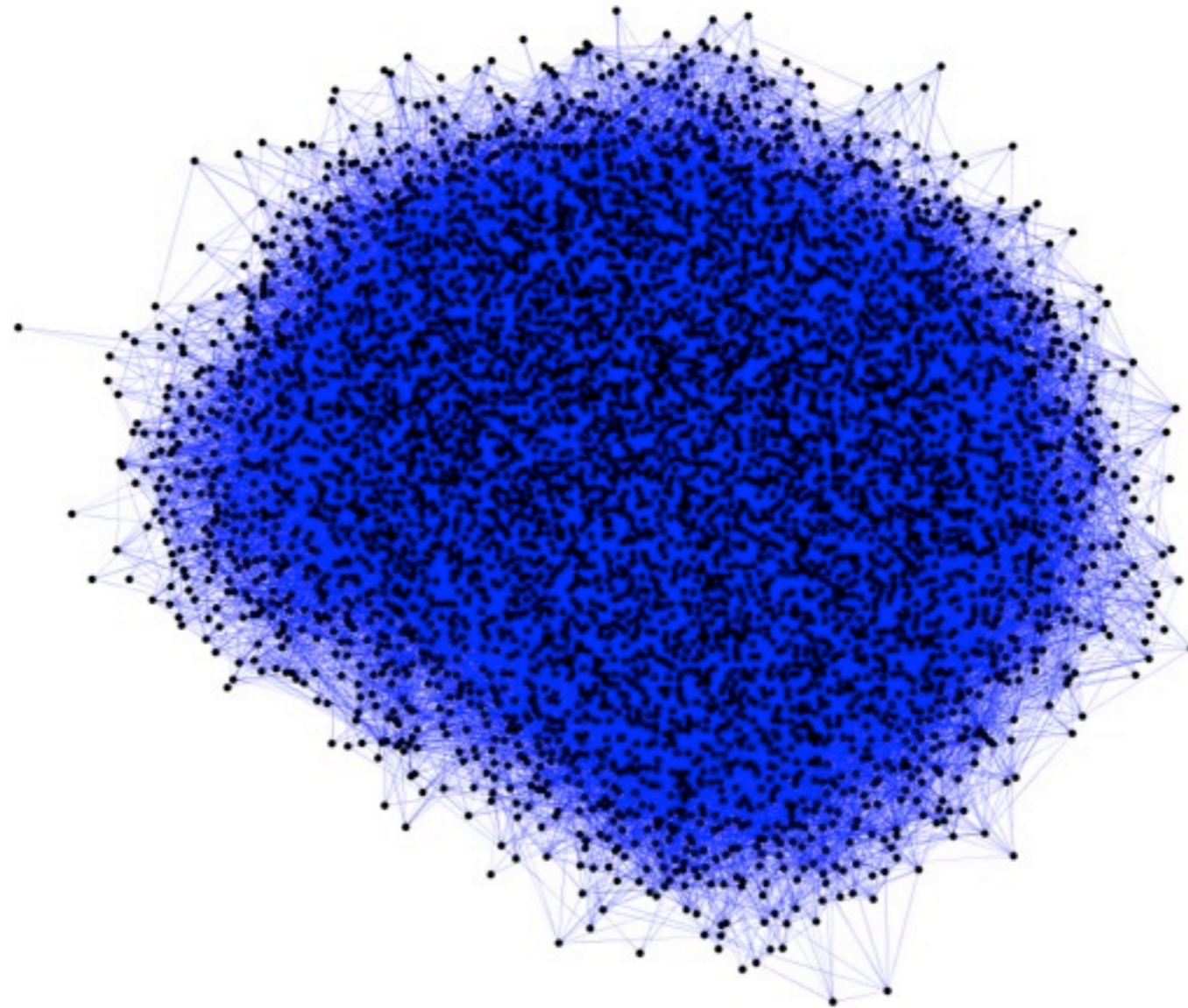


Complex global structure



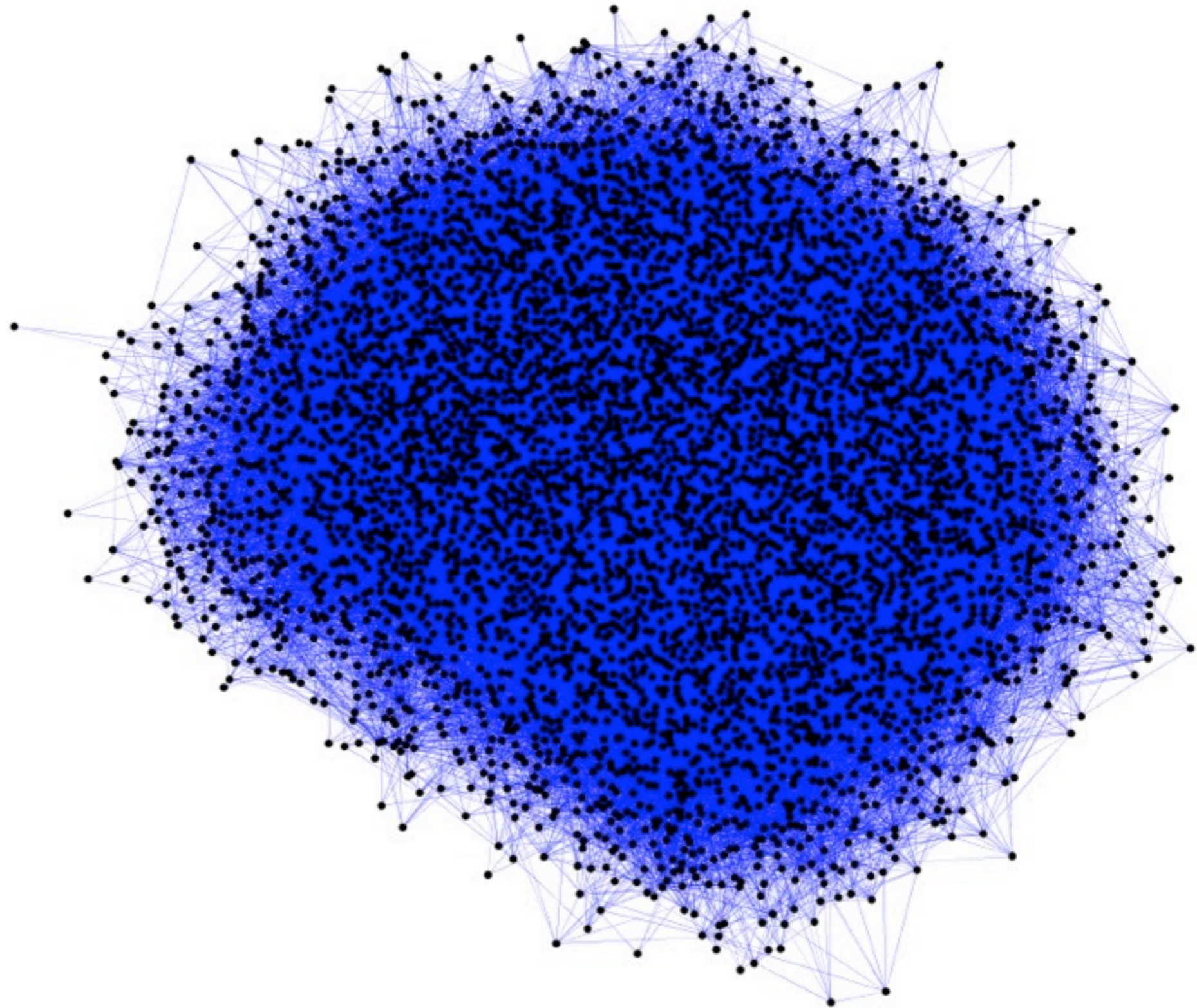




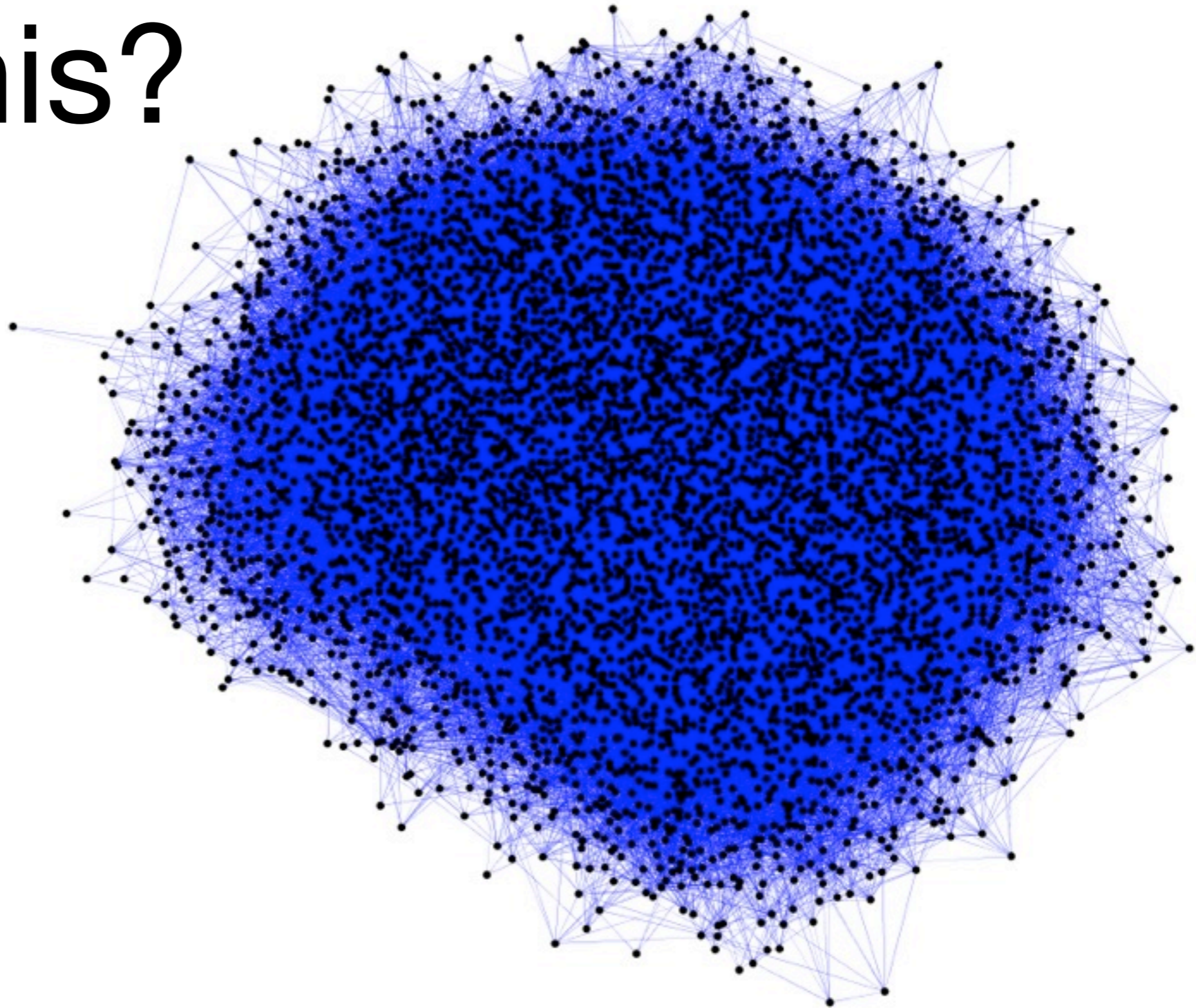


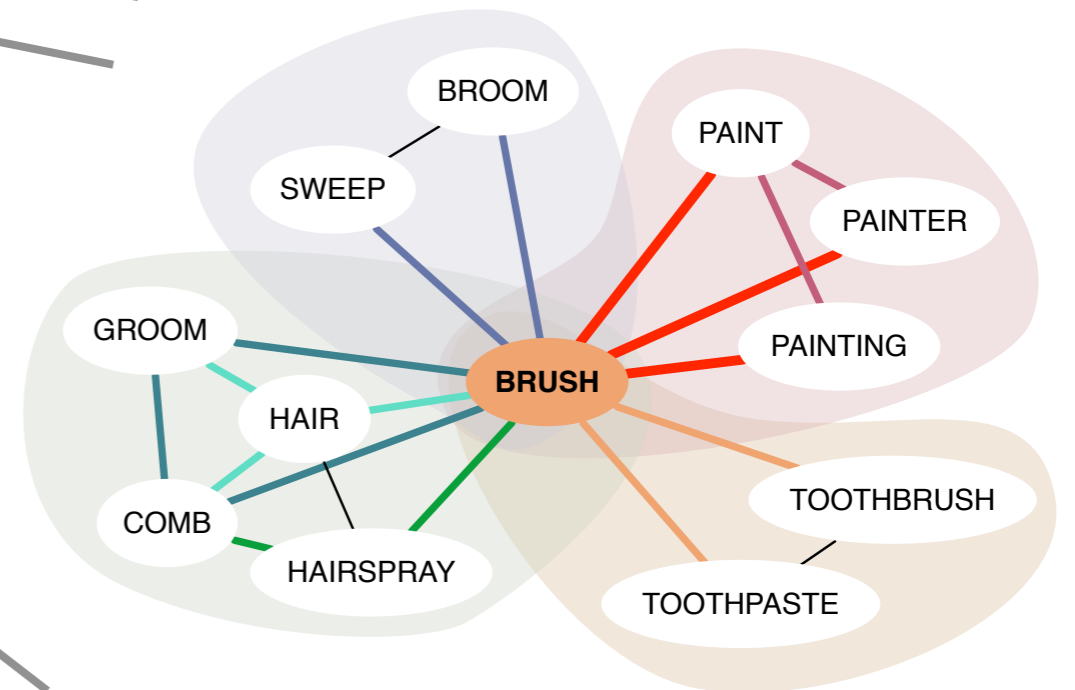
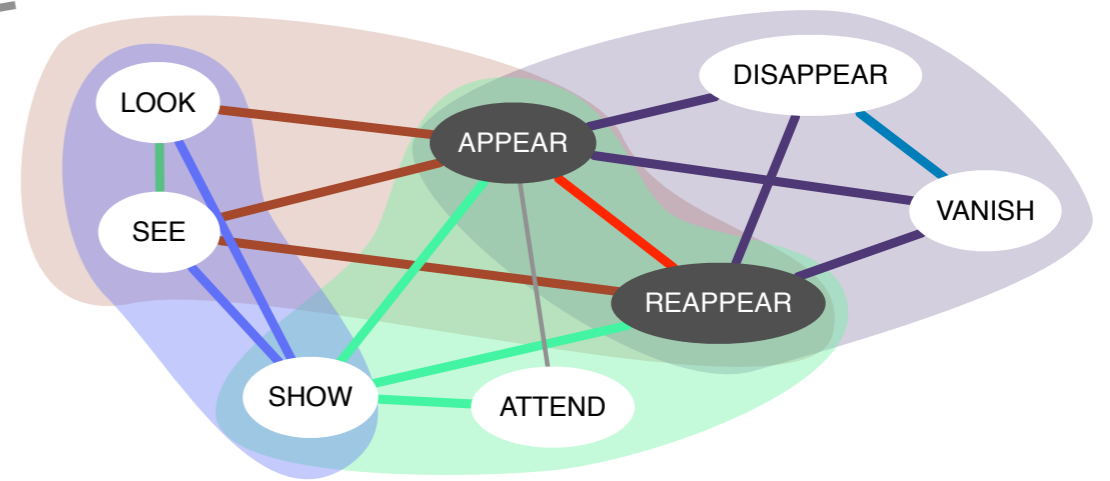
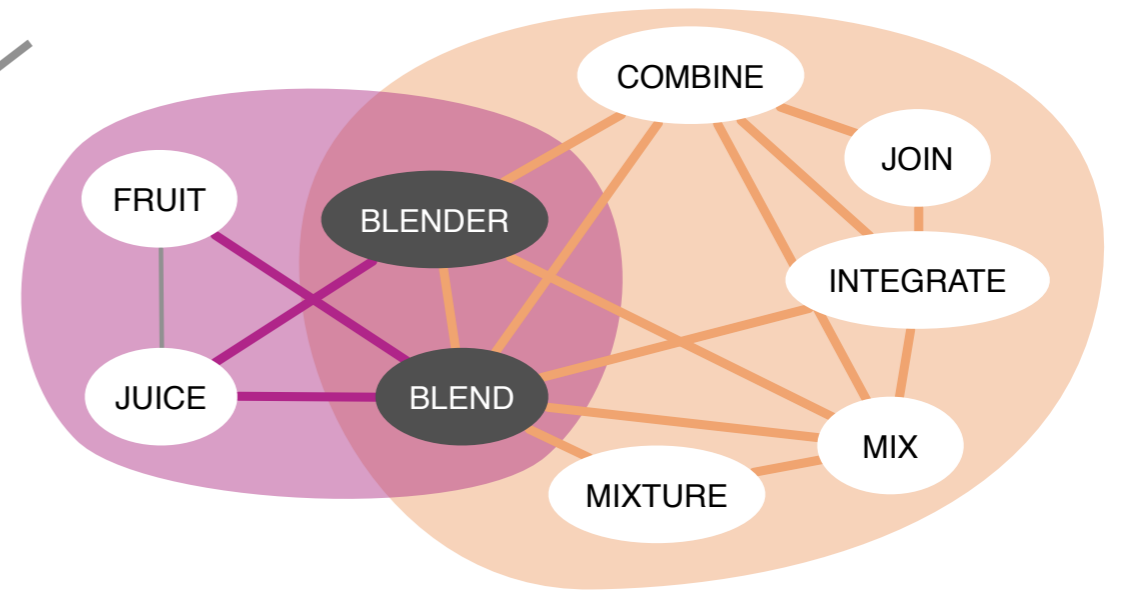
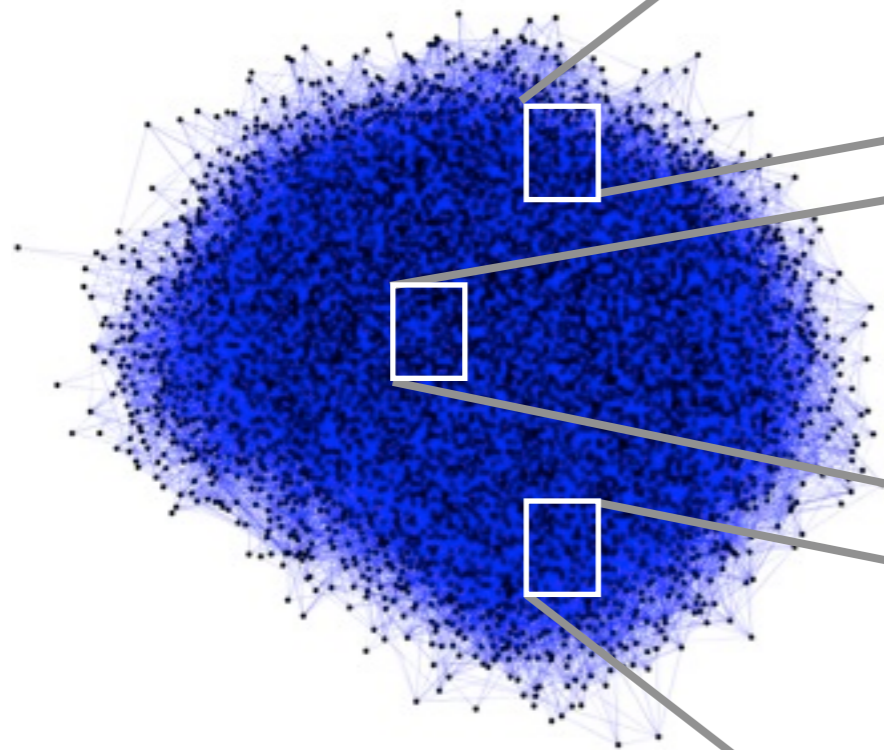
This is a modular network.

What is this?



What the xxxx is this?



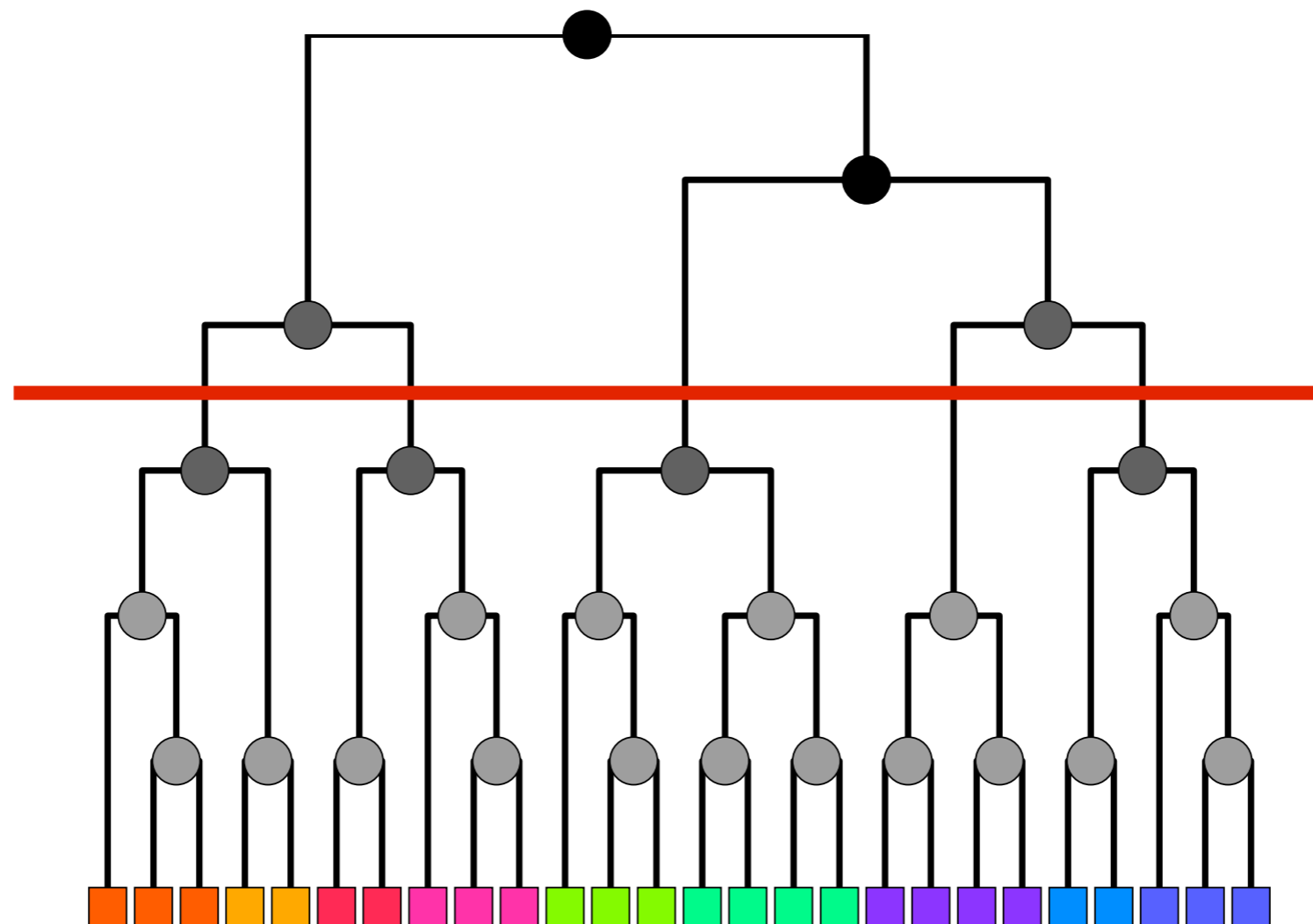


Here is the
PROBLEM.

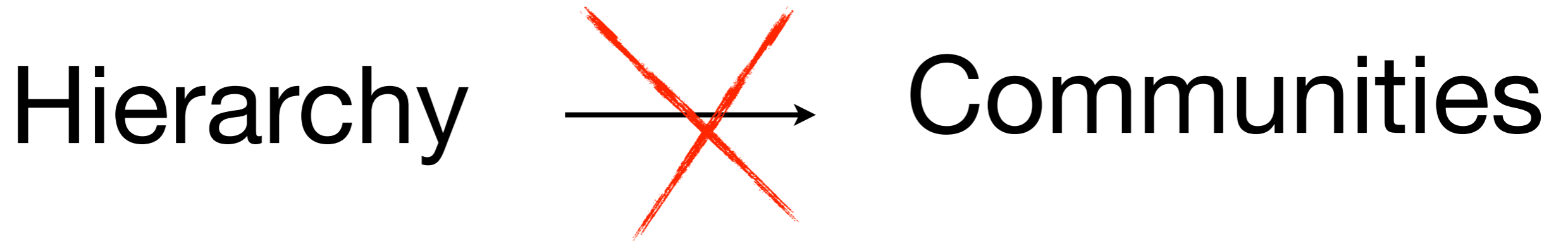
Communities exist.

**Hierarchical structure
exists.**

Hierarchy implies disjoint communities.



Hierarchical community structure



Hopeless?

Solution:
Use **LINKS**

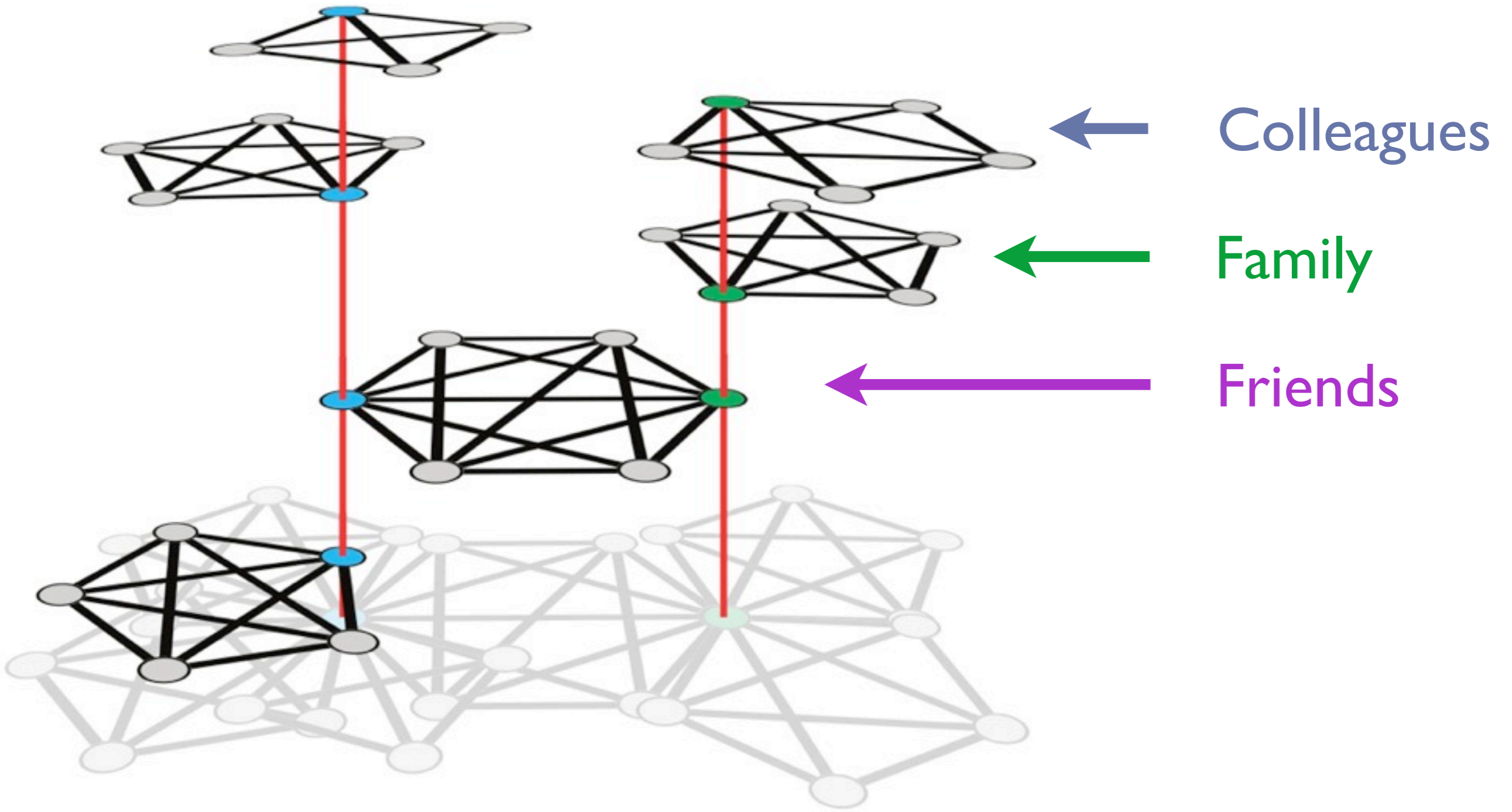
Solution:
Use **LINKS**

Solution:
Use **Links**

“a group of densely
interconnected nodes”

Our solution:
Use **Links**

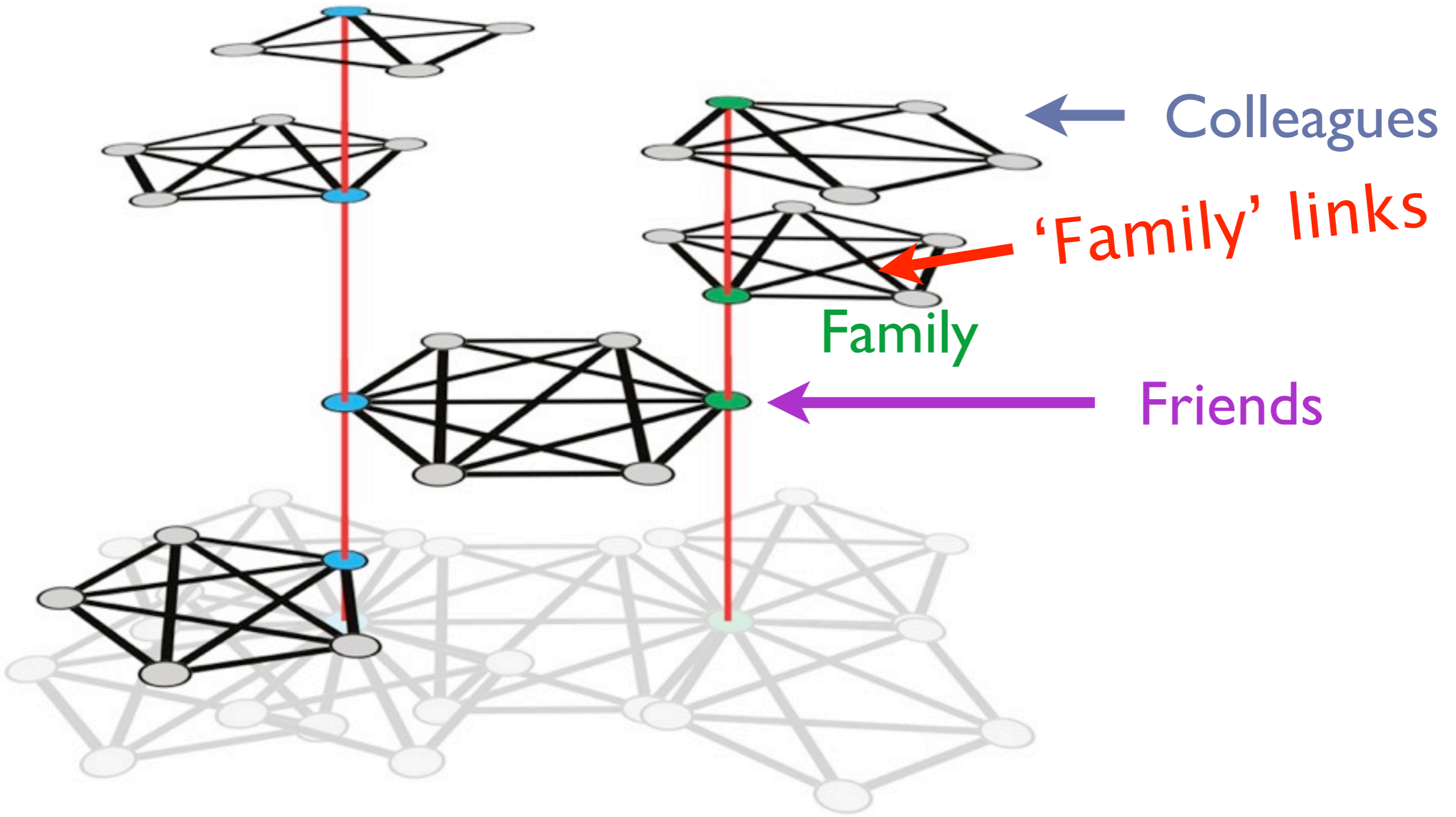
“a group of densely
interconnected **LINKS**”

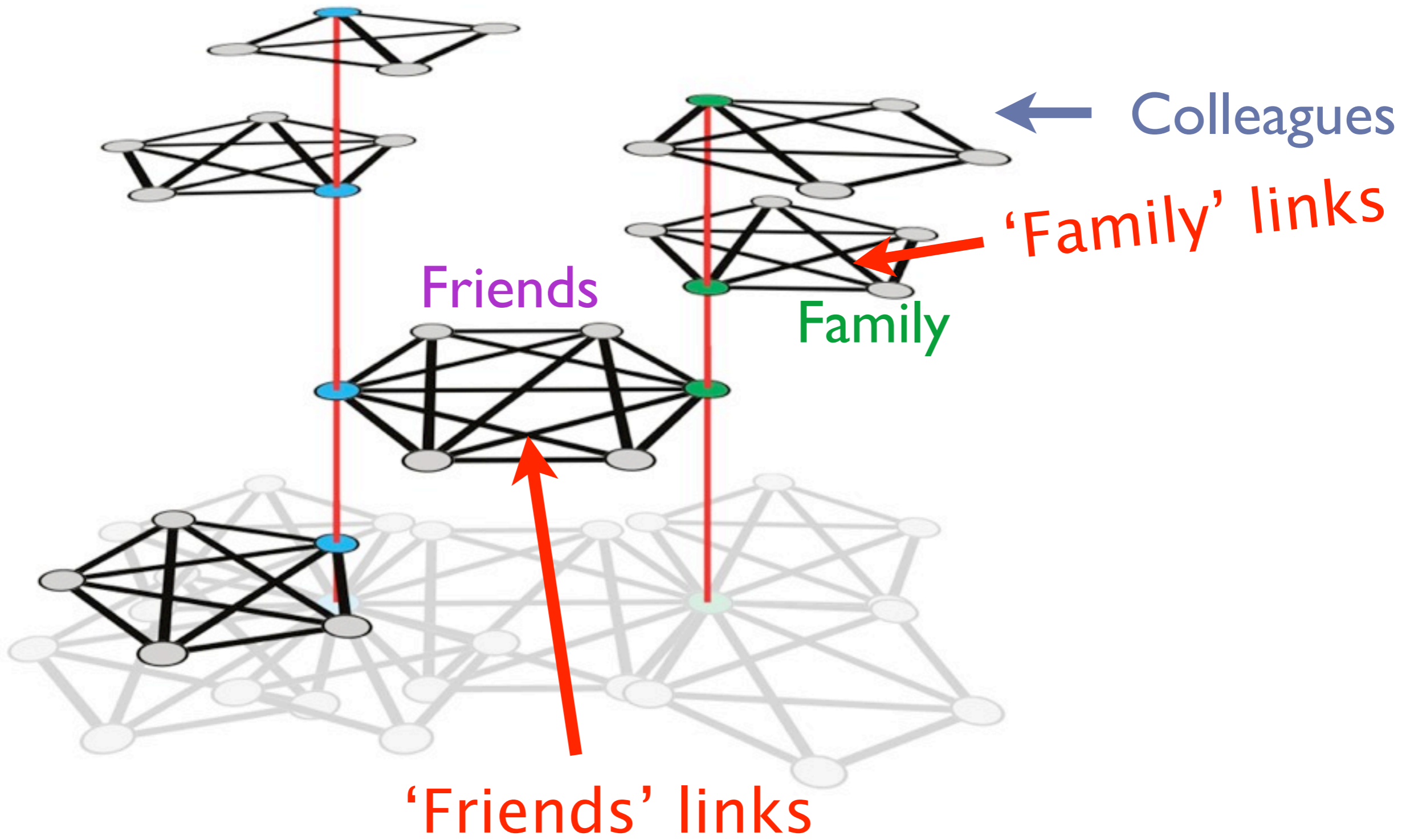


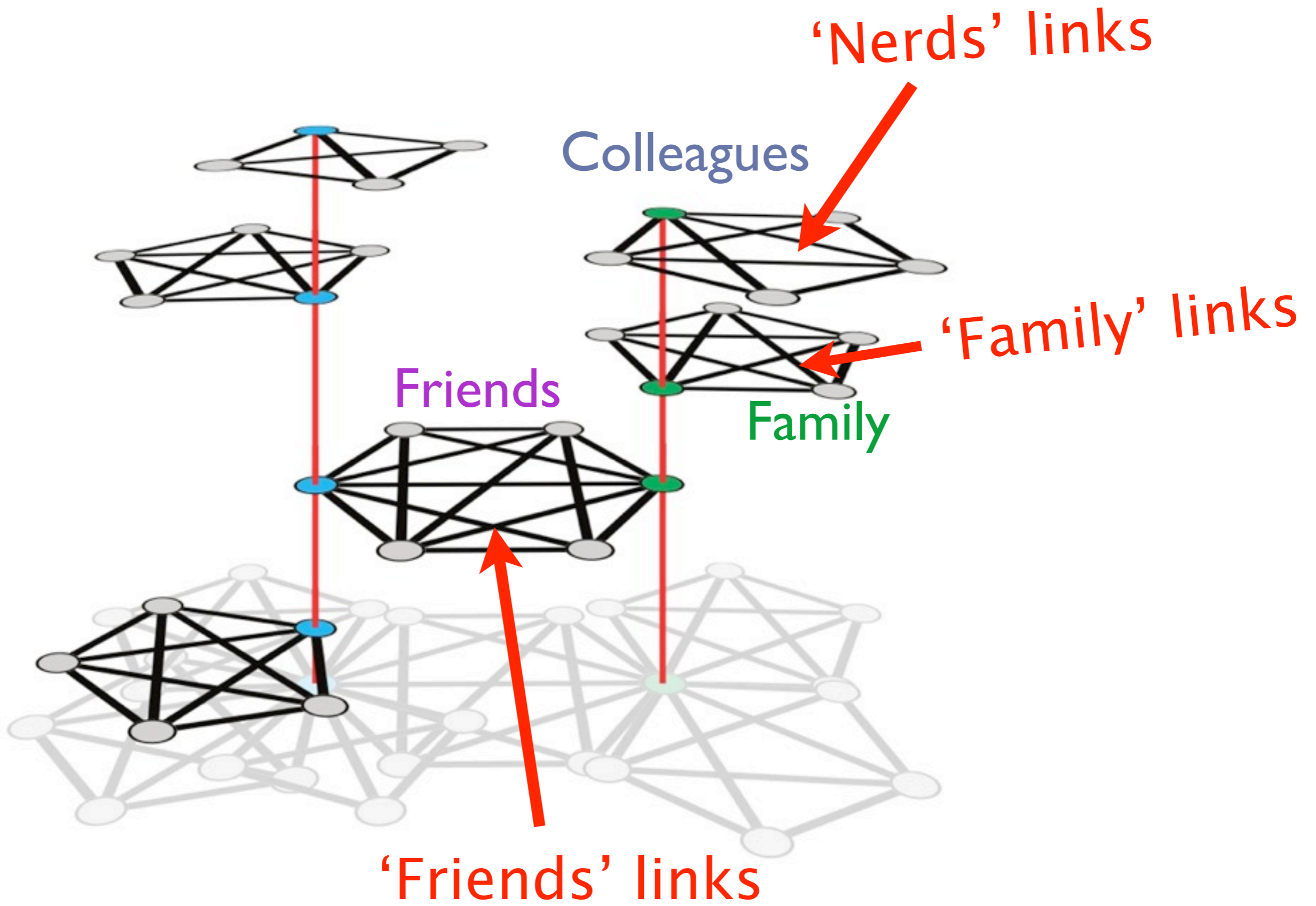
Colleagues

Family

Friends



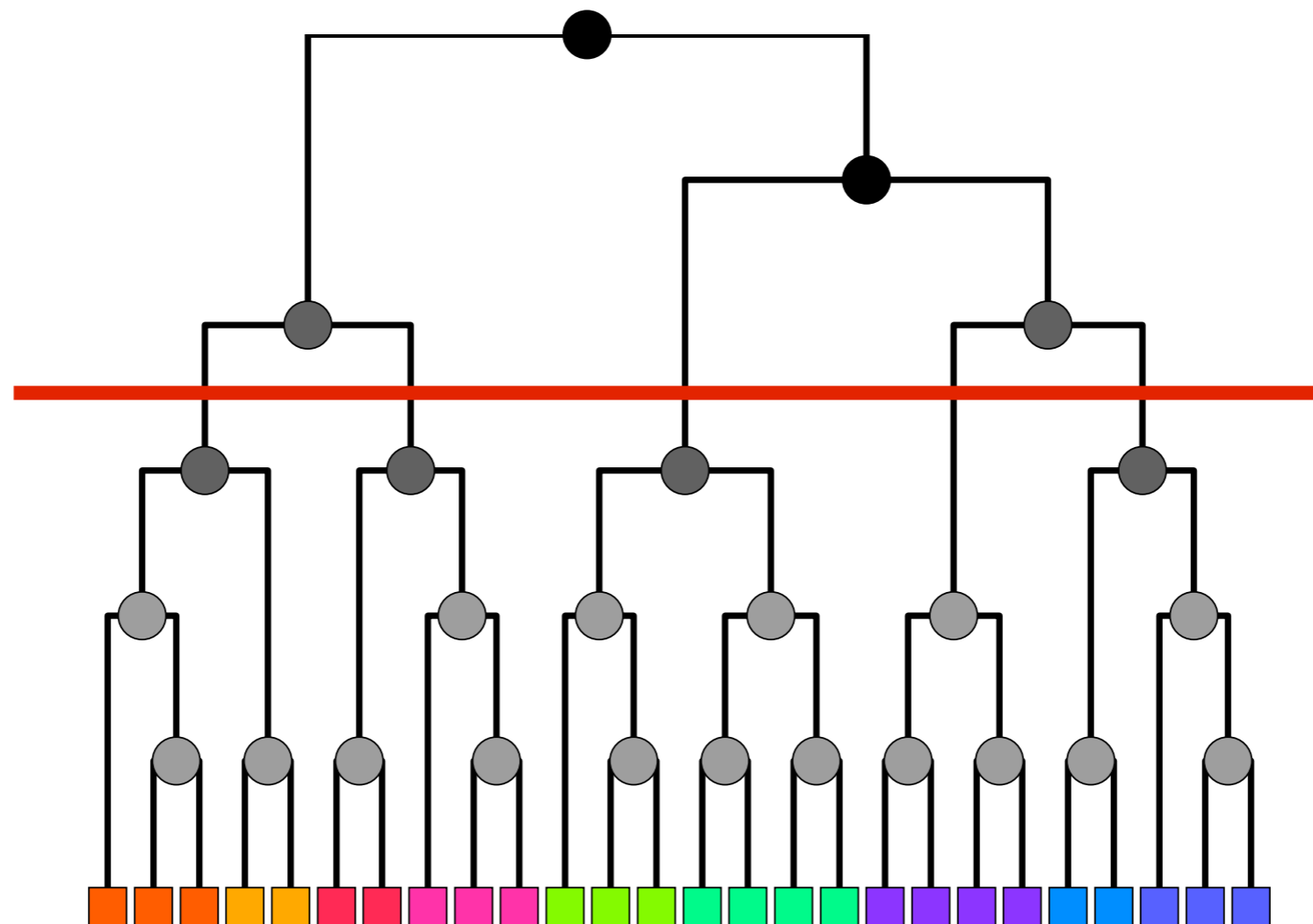




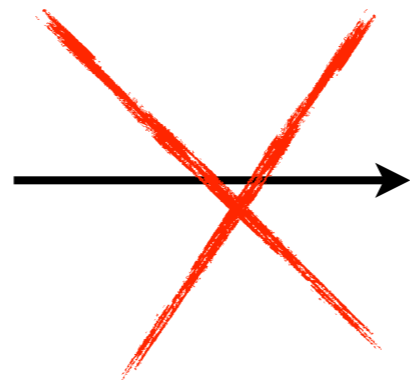
Nodes: multiple membership

Links: unique membership

Hierarchy implies disjoint communities.



Hierarchy



Communities

Hierarchy → **Communities**

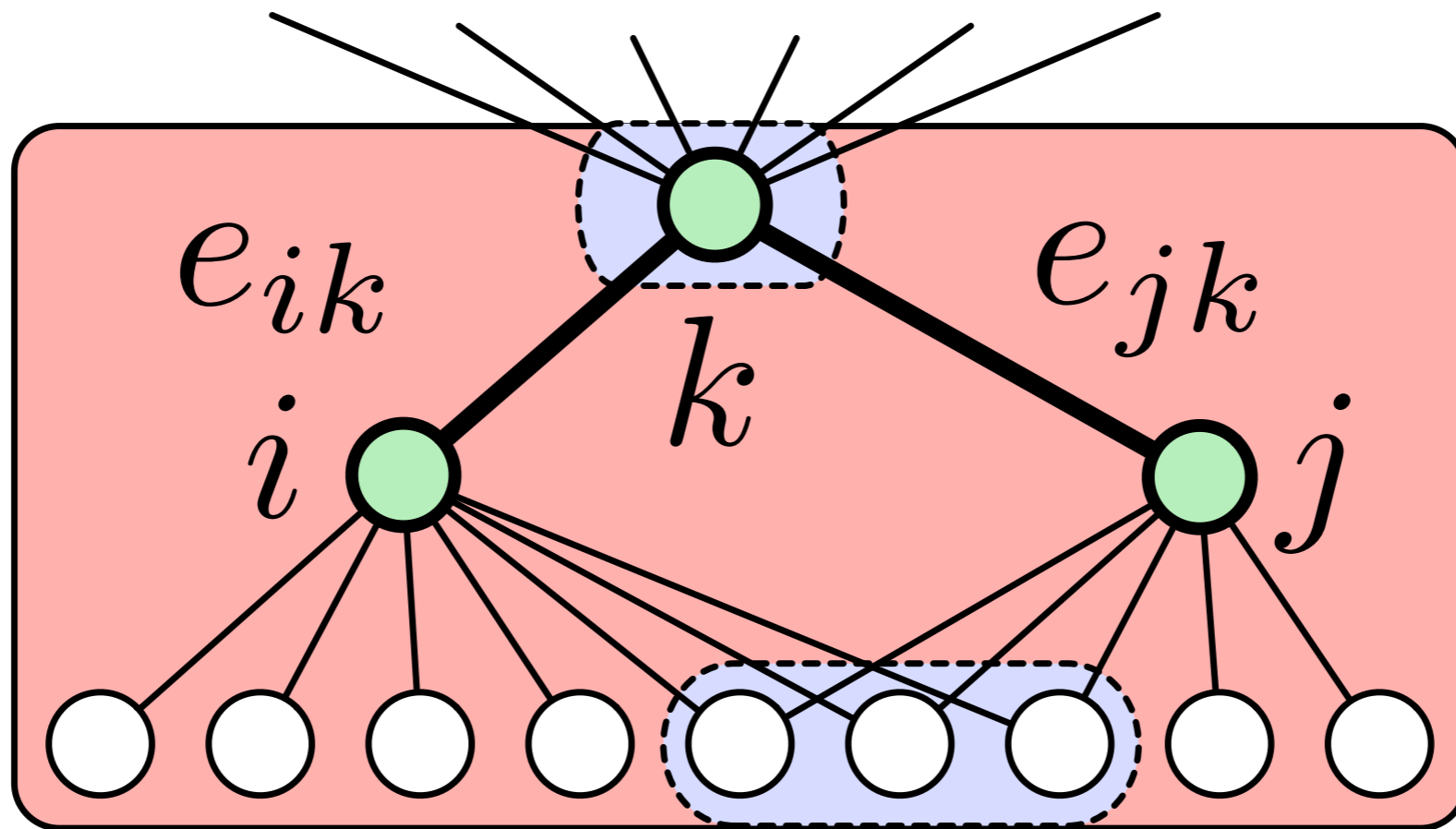
RECONCILIATION

So, How?

Similarity between links

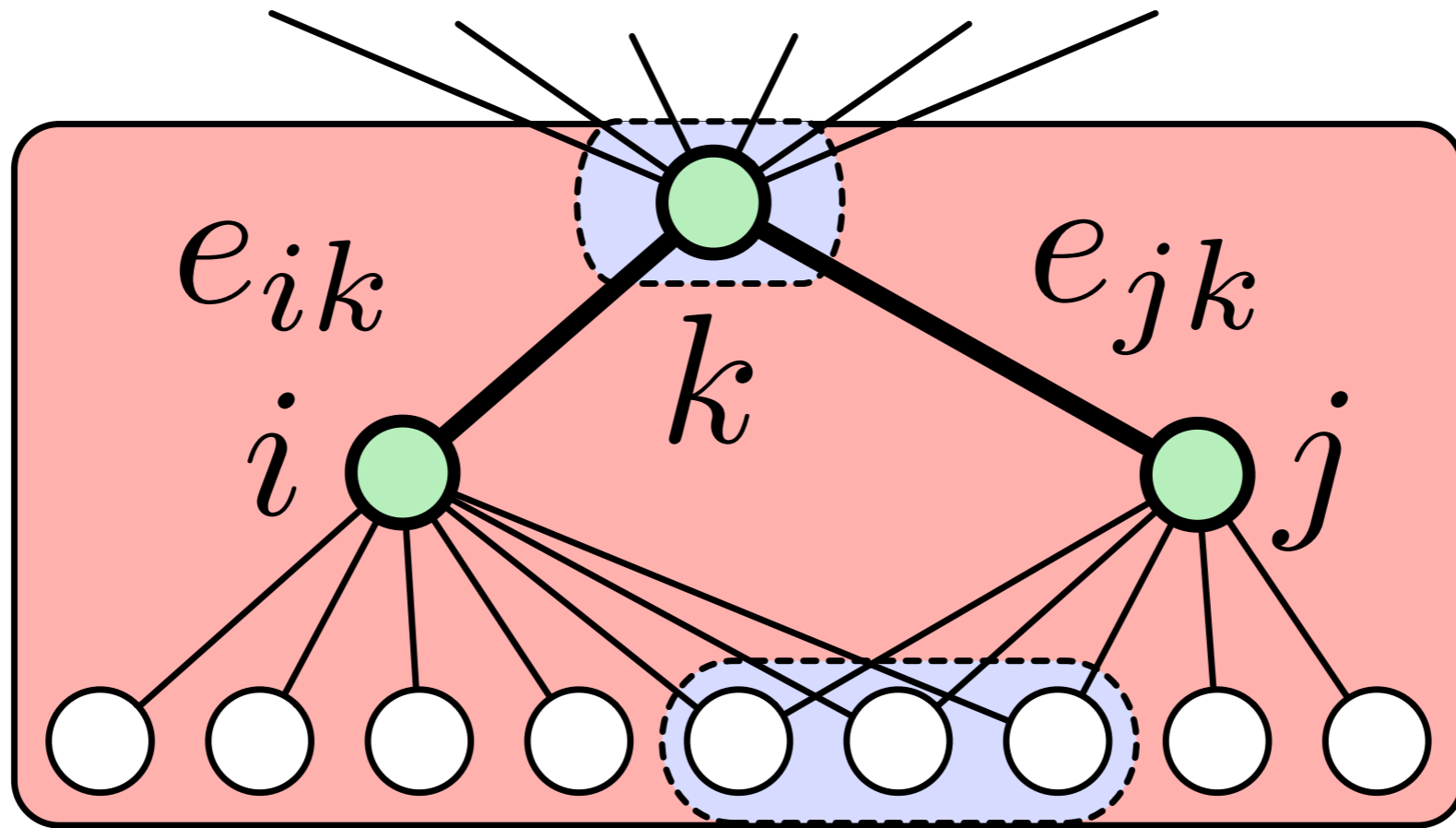


Hierarchical Clustering



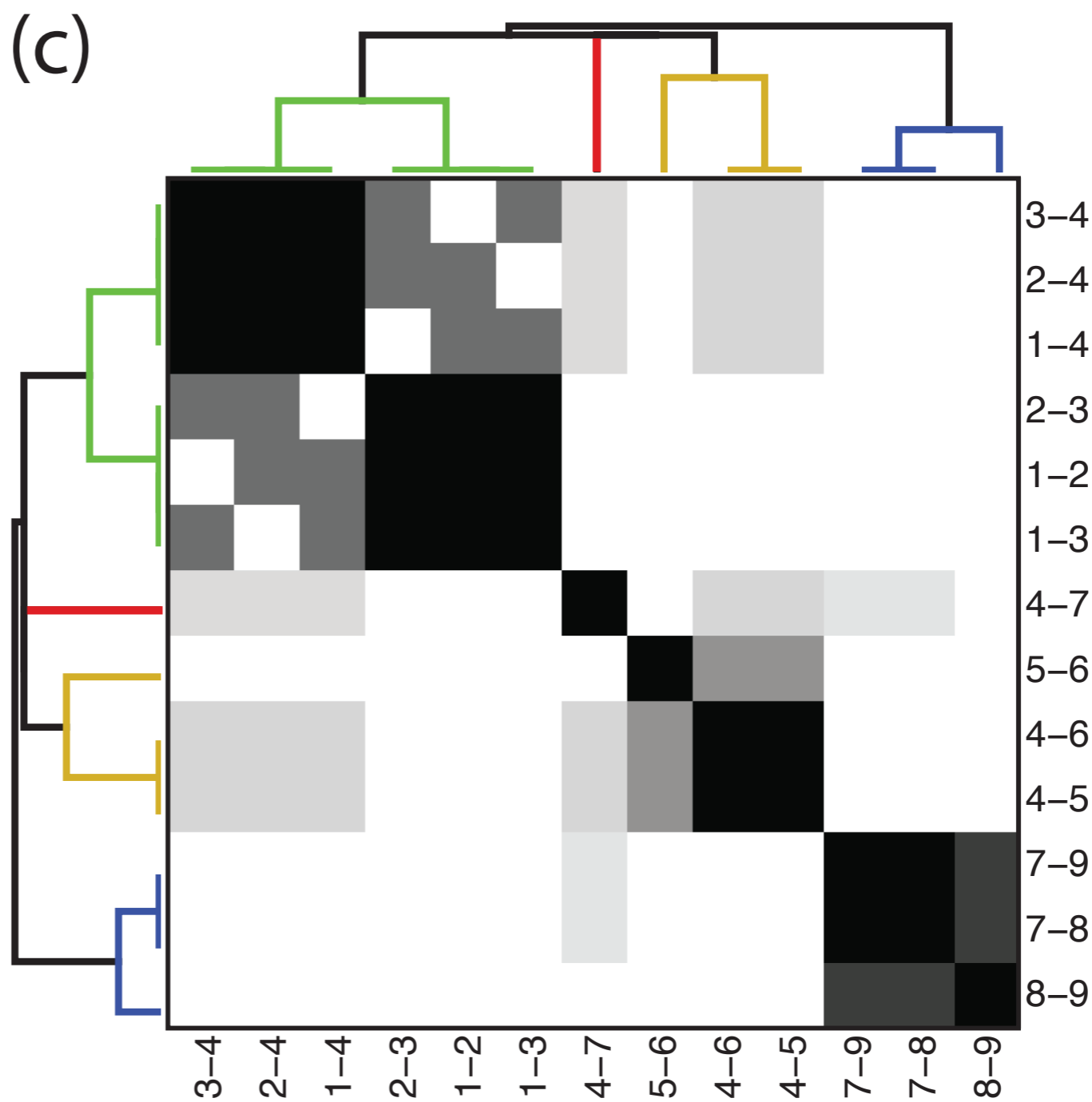
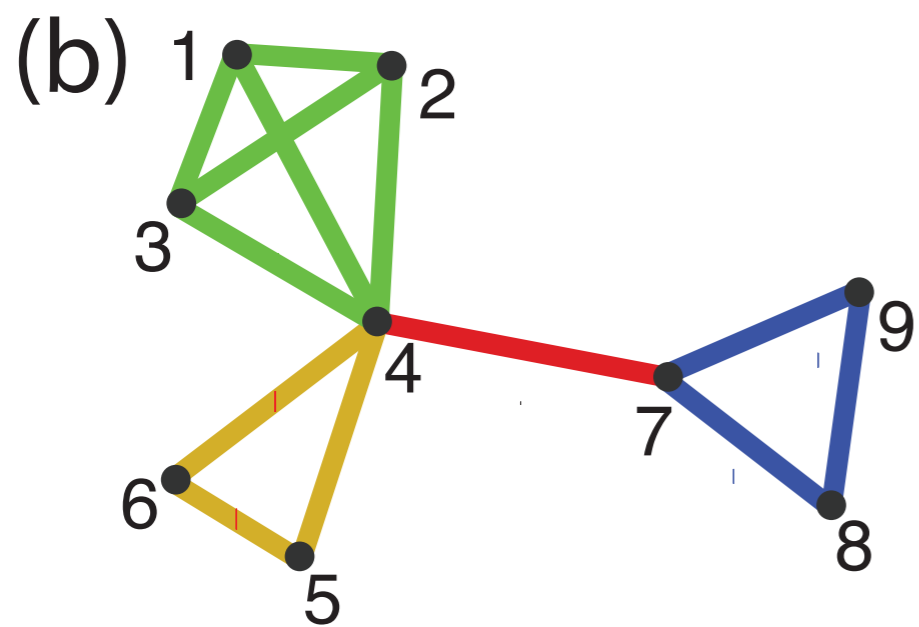
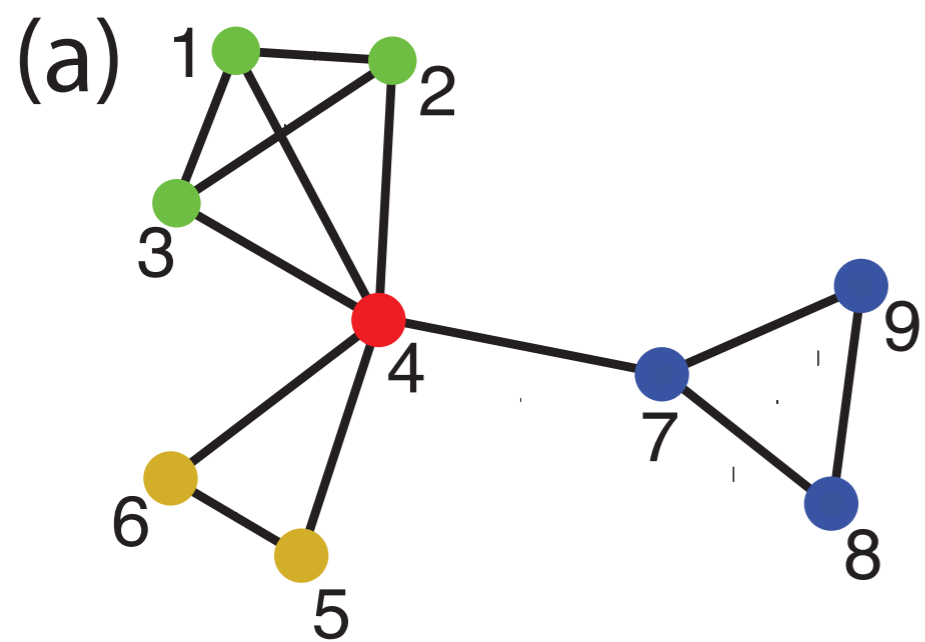
$$n_+(i) \equiv \{x \mid d(i, x) \leq 1\}$$

$$S(e_{ik}, e_{jk}) = \frac{|n_+(i) \cap n_+(j)|}{|n_+(i) \cup n_+(j)|}$$

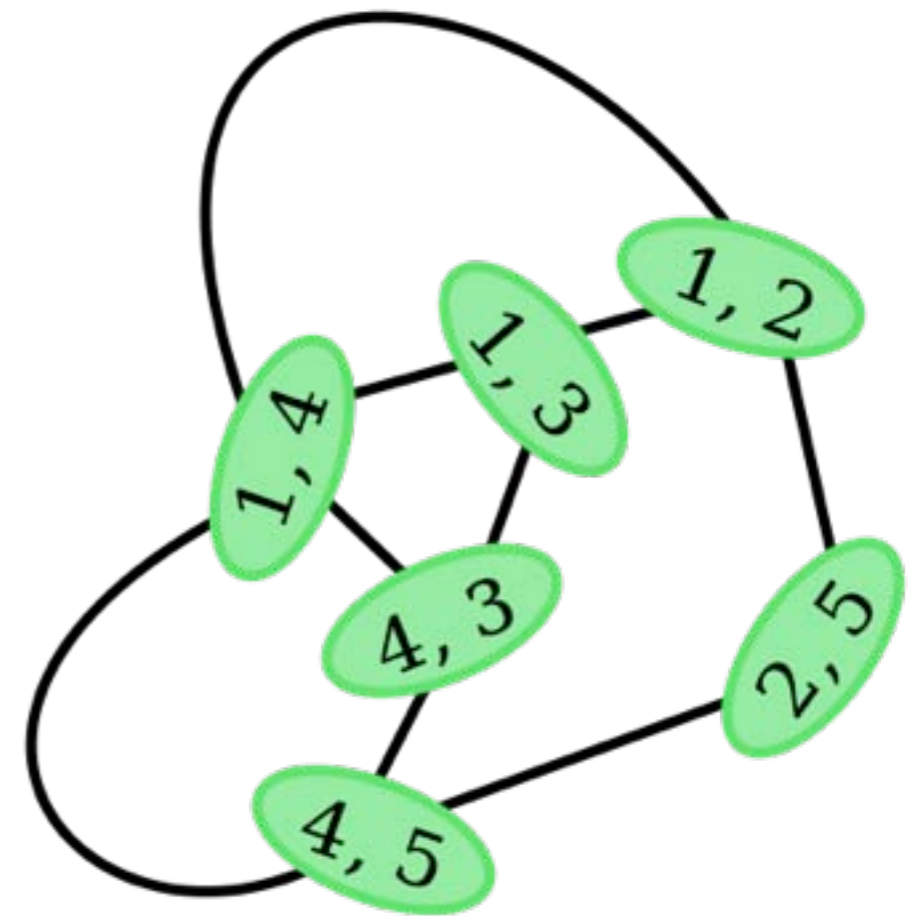
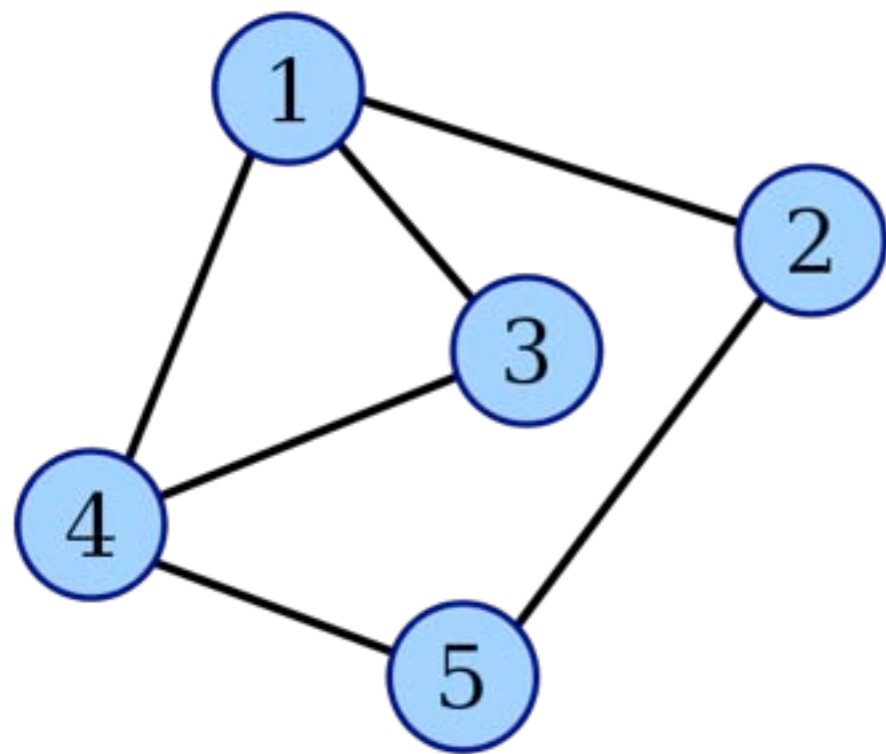


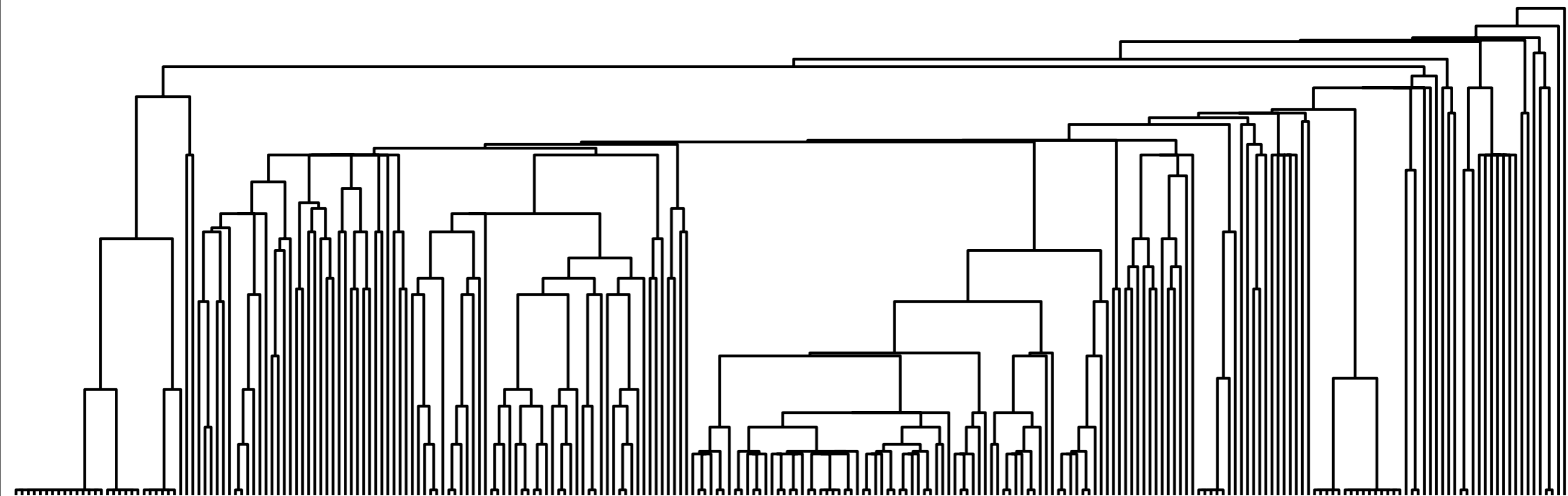
$$n_+(i) \equiv \{x \mid d(i, x) \leq 1\}$$

$$S(e_{ik}, e_{jk}) = \frac{|n_+(i) \cap n_+(j)|}{|n_+(i) \cup n_+(j)|} = \frac{4}{12}$$

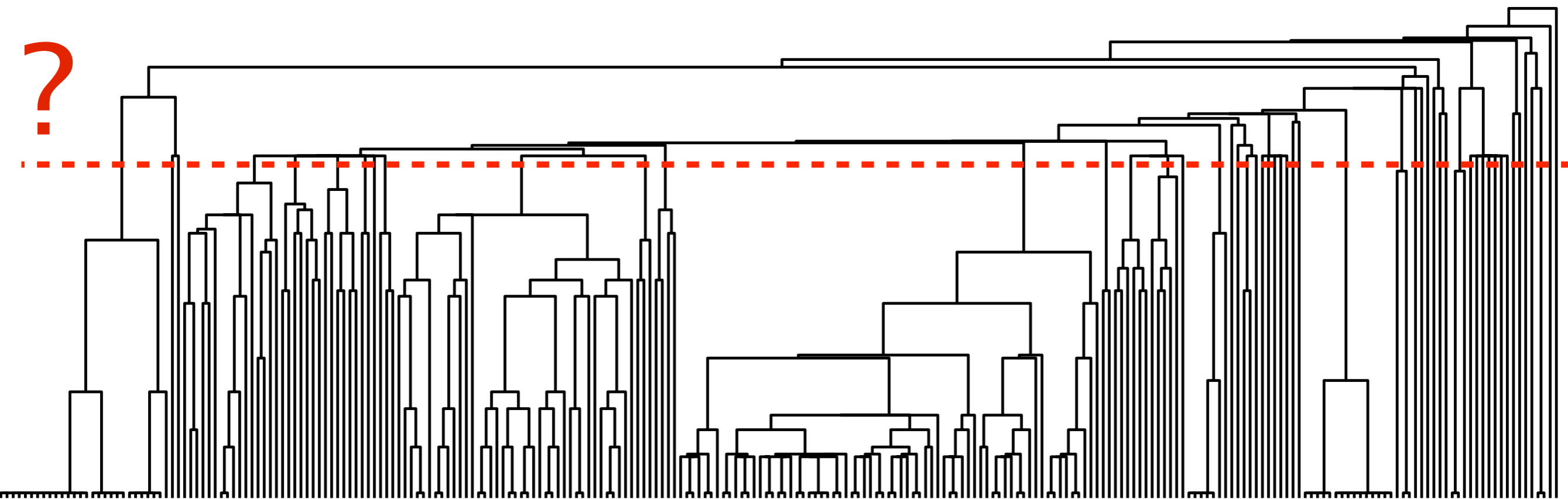


Line graph transformation





?

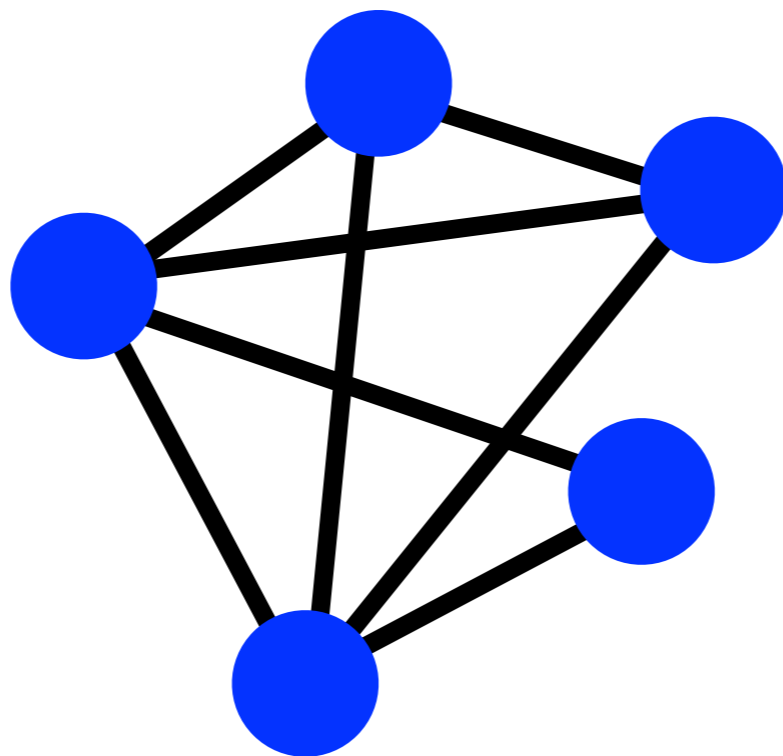


Partition Density

Community c has m_c edges and n_c induced nodes

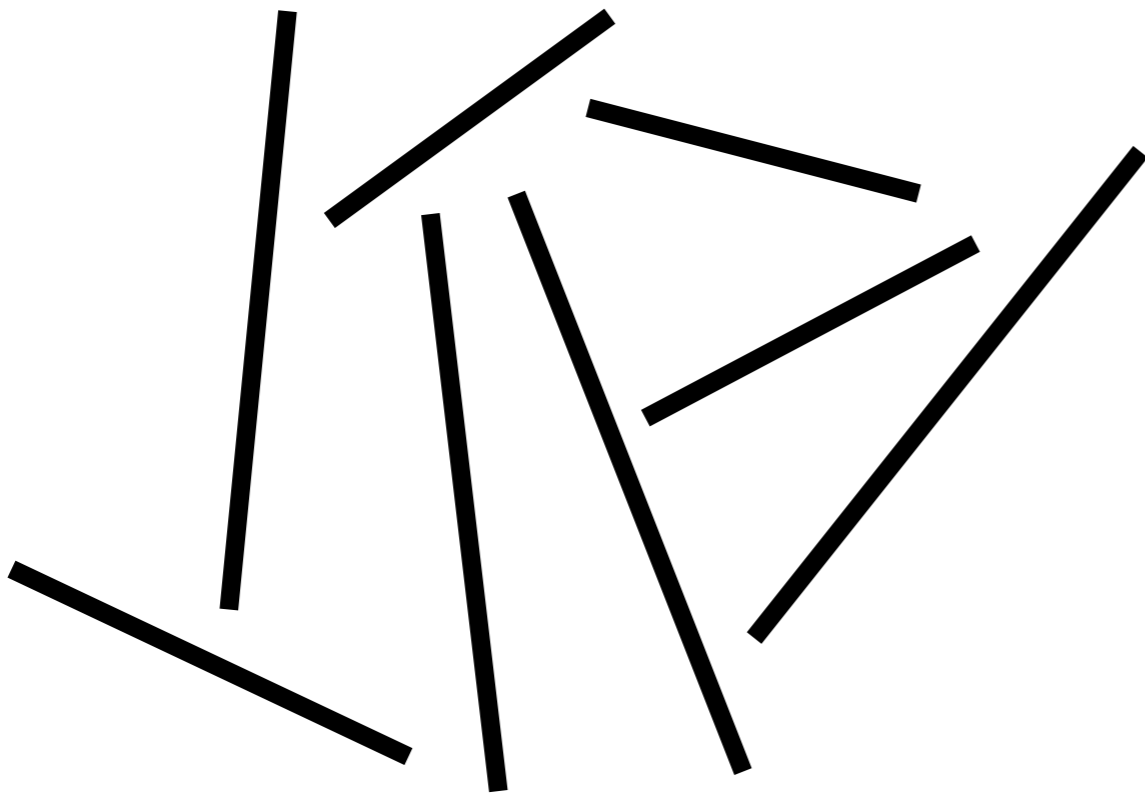
Partition Density

Community c has m_c edges and n_c induced nodes

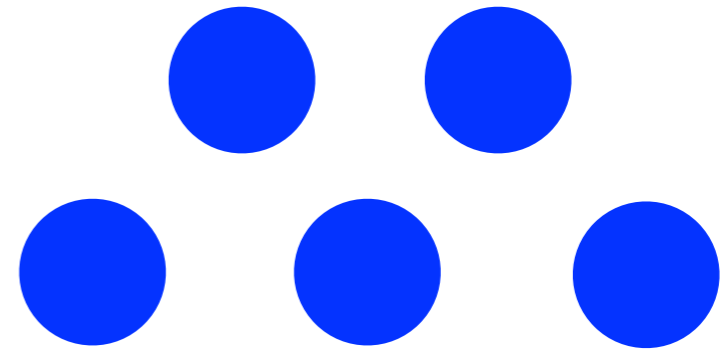


Partition Density

Community c has m_c edges and n_c induced nodes



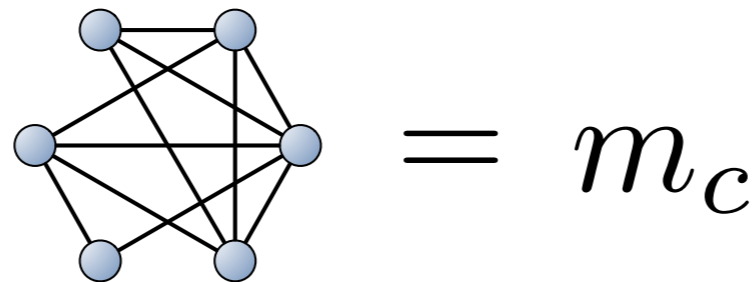
$$m_c = 8$$



$$n_c = 5$$

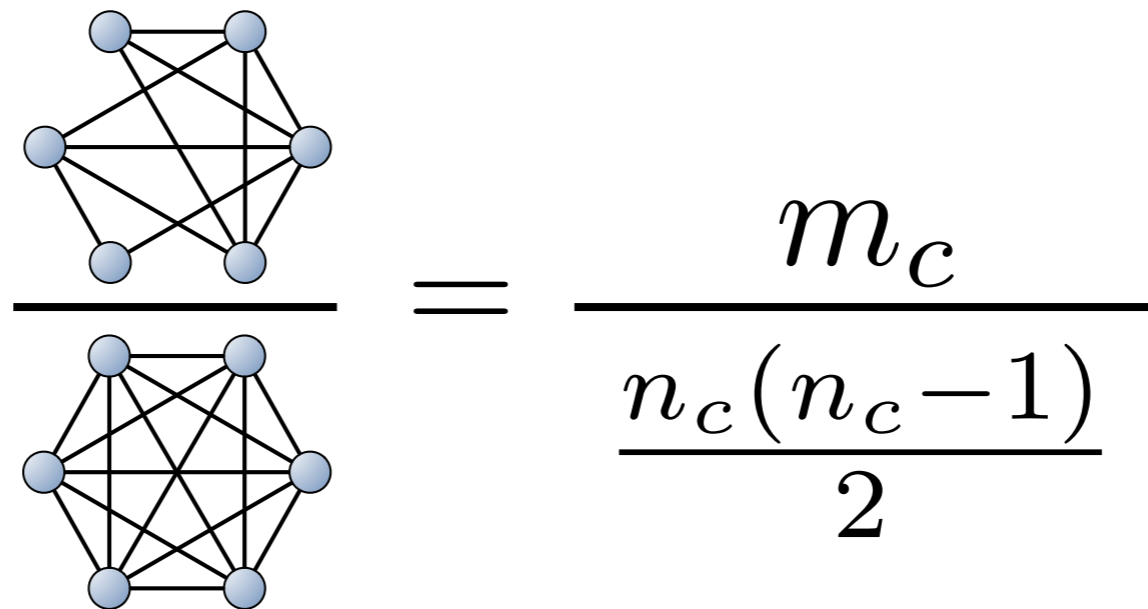
Partition Density

Community c has m_c edges and n_c induced nodes



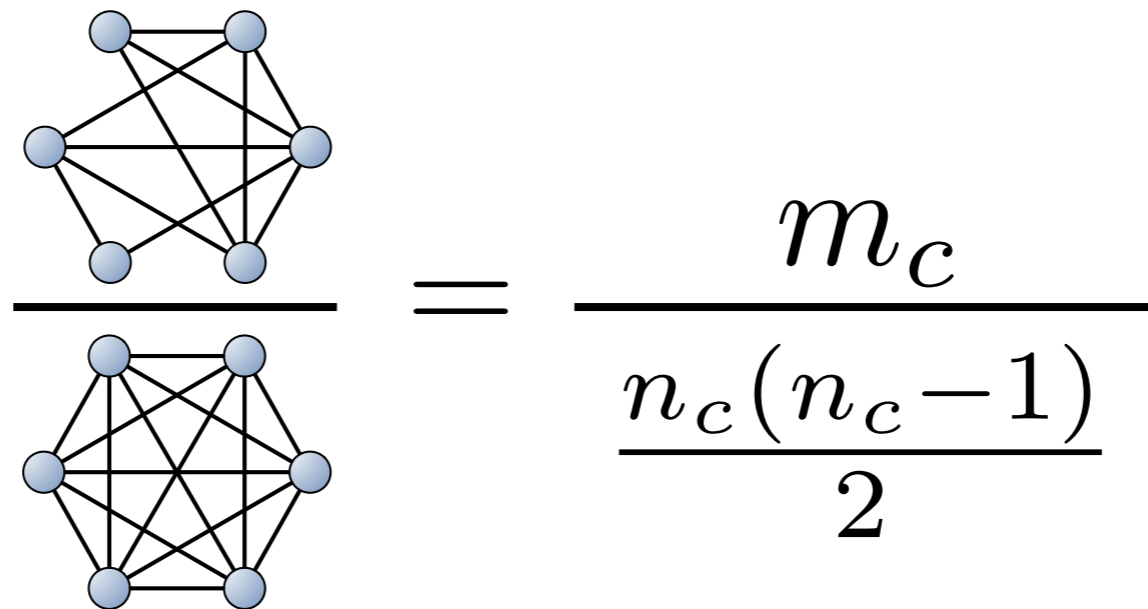
Partition Density

Community c has m_c edges and n_c induced nodes


$$\frac{\text{Complete Graph with 5 nodes}}{\text{Complete Graph with 6 nodes}} = \frac{m_c}{\frac{n_c(n_c-1)}{2}}$$

Partition Density

Community c has m_c edges and n_c induced nodes


$$\frac{\text{Graph with } m_c \text{ edges}}{\text{Complete Graph } K_{n_c}} = \frac{m_c}{\frac{n_c(n_c-1)}{2}}$$



A **single** link is **maximally dense**

Partition Density

Community c has m_c edges and n_c induced nodes

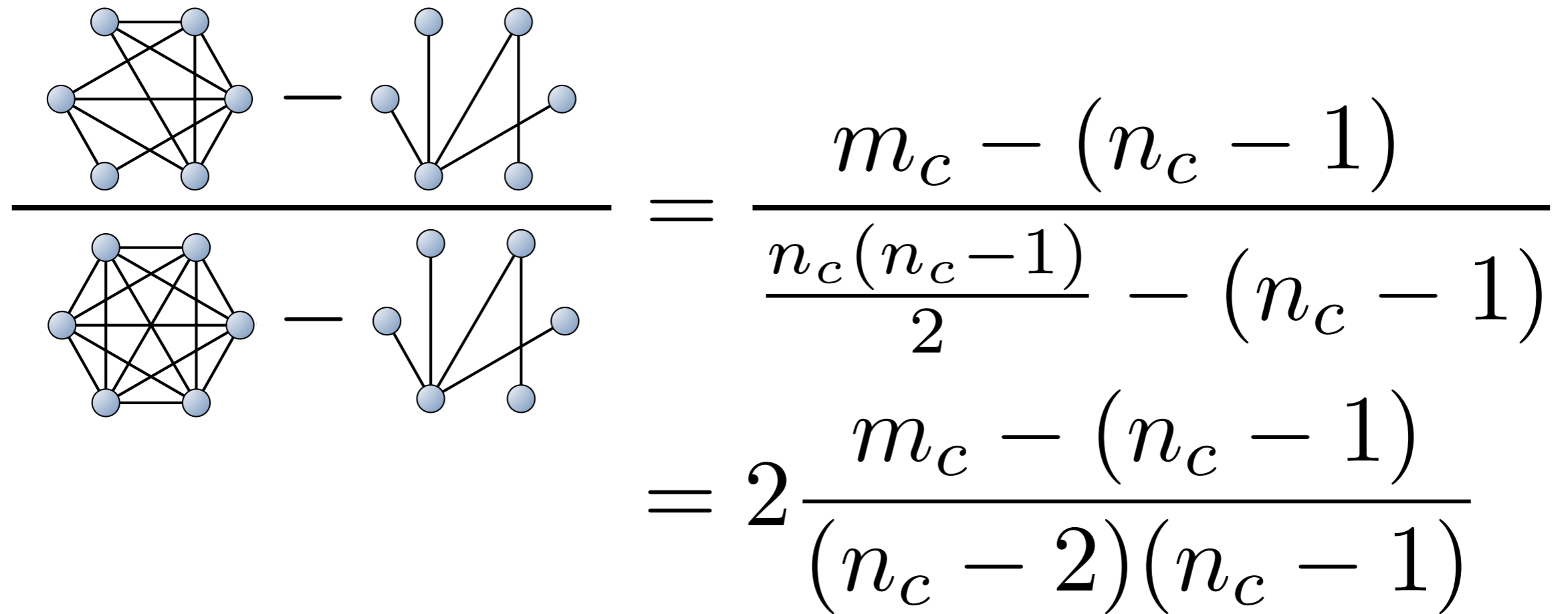
$$\frac{\text{Graph with } m_c \text{ edges} - \text{Star graph with } n_c - 1 \text{ edges}}{\text{Complete graph with } \frac{n_c(n_c-1)}{2} \text{ edges} - \text{Star graph with } n_c - 1 \text{ edges}} = \frac{m_c - (n_c - 1)}{\frac{n_c(n_c-1)}{2} - (n_c - 1)}$$

Partition Density

$$\frac{\text{[Diagram: 6 nodes in a circle, all connected to each other] - \text{[Diagram: 6 nodes in a circle, one node connected to all others]}}{\text{[Diagram: 6 nodes in a circle, all connected to each other] - \text{[Diagram: 6 nodes in a circle, one node connected to all others]}} = \frac{m_c - (n_c - 1)}{\frac{n_c(n_c - 1)}{2} - (n_c - 1)}$$

$$= 2 \frac{m_c - (n_c - 1)}{(n_c - 2)(n_c - 1)}$$

Partition Density



$$\frac{m_c - (n_c - 1)}{\frac{n_c(n_c - 1)}{2} - (n_c - 1)} = 2 \frac{m_c - (n_c - 1)}{(n_c - 2)(n_c - 1)}$$

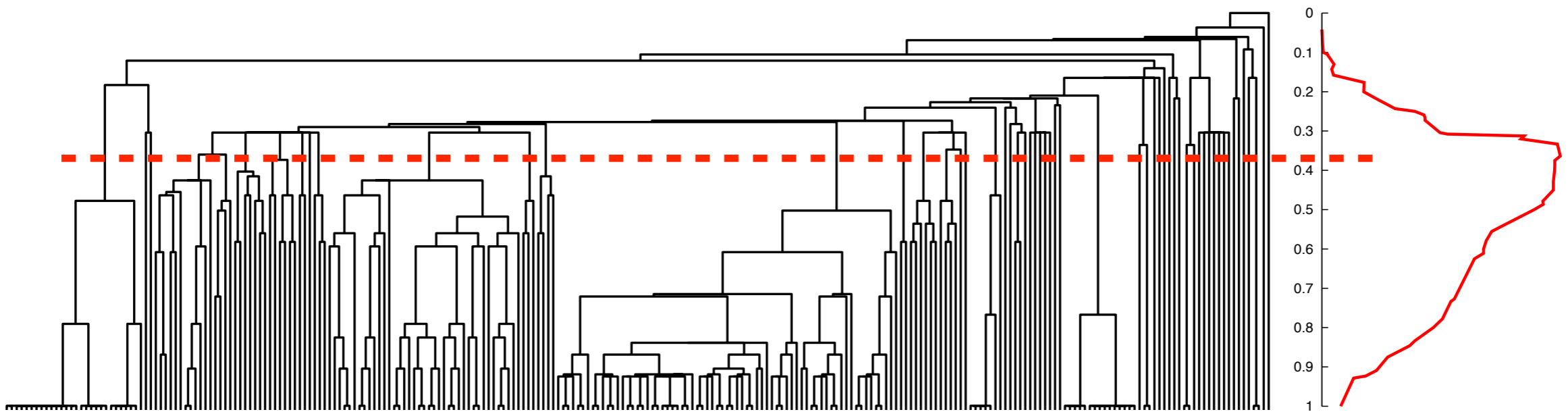
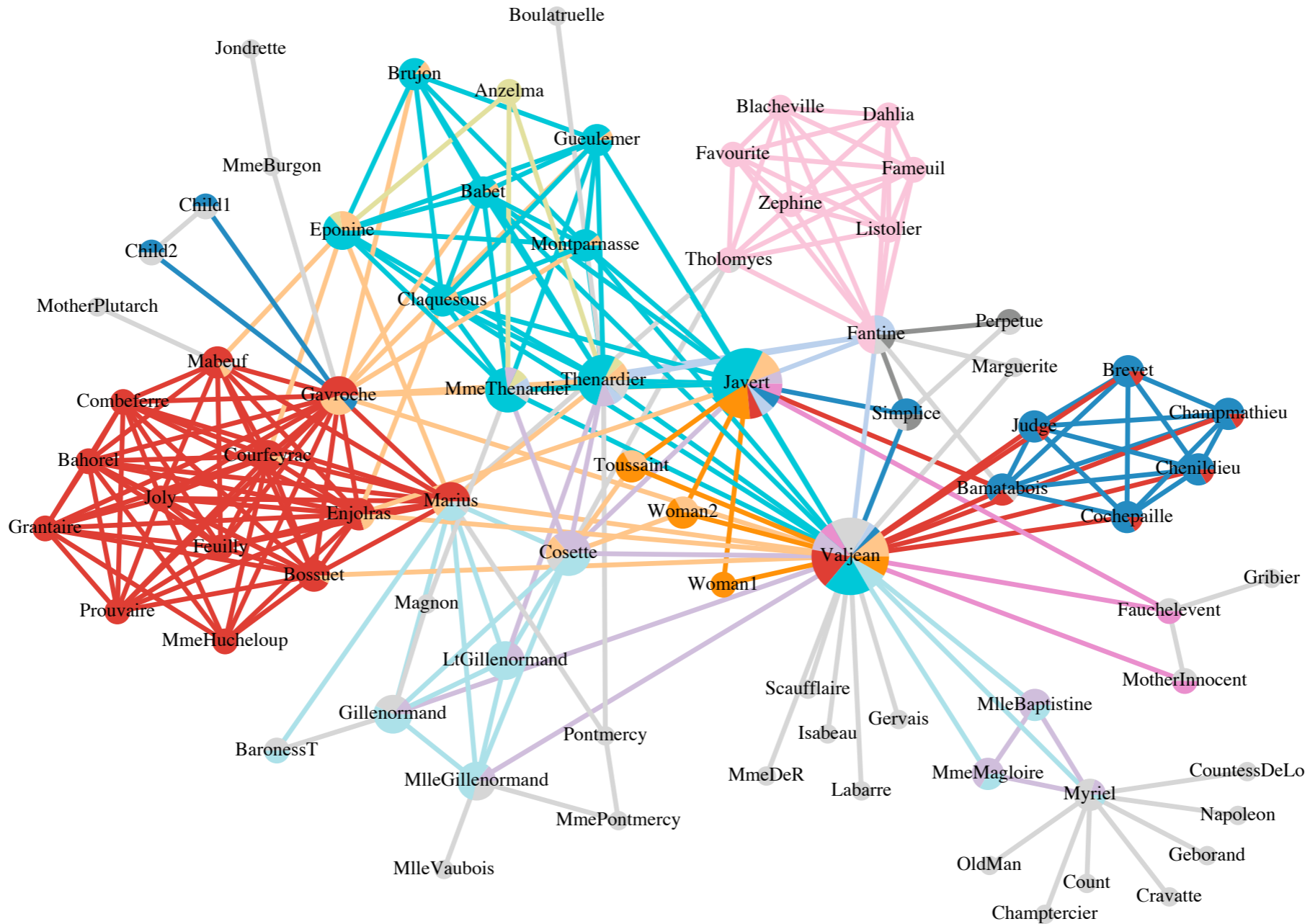
$$D \equiv \frac{2}{M} \sum_c m_c \frac{m_c - (n_c - 1)}{(n_c - 2)(n_c - 1)}$$

Partition Density

$$D \equiv \frac{2}{M} \sum_c m_c \frac{m_c - (n_c - 1)}{(n_c - 2)(n_c - 1)}$$

objective function!

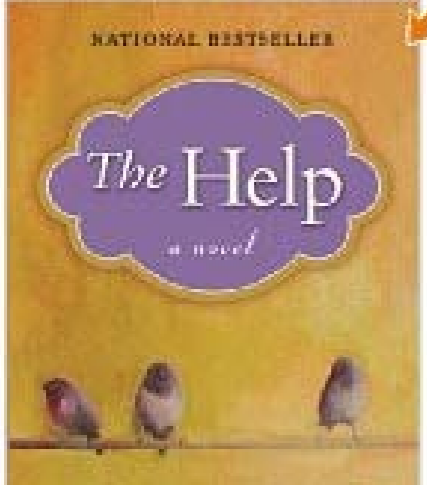
overlapping communities \longrightarrow **well posed** optimization



Does it really work?

Metadata

Click to **LOOK INSIDE!**



The Help (Ha
~ Kathryn Stockett |
Key Phrases: relaxing
★★★★★ (1,346)

List Price: ~~\$24.95~~
Price: **\$9.50**
You Save: **\$15.45**


In Stock.
Ships from and sold

Look for Similar Items by Subject

Search **Books** by subject:

- [Fiction](#)
- [Fiction - General](#)
- [Fiction / Literary](#)
- [Literary](#)
- [American Historical Fiction](#)
- [African American women](#)
- [Civil rights movements](#)

Customers Who Bought This Item Also Bought



The Guernsey Literary and Potato Peel Pie Soc... by Mary Ann Shaffer
★★★★★ (1,065) \$6.98

Sarah's Key by Tatiana de Rosnay
★★★★☆ (352) \$8.37

The Girl with the Dragon Tattoo (Vintage) by Stieg Larsson
★★★★☆ (700) \$6.00

Olive K by Eliza
★★★★

Tags Customers Associate with This Product [\(What's this?\)](#)

Click on a tag to find related items, discussions, and people.

- [race relations](#) (188)
- [historical fiction](#) (162)
- [deep south](#) (154)
- [civil rights movement](#) (119)
- [mississippi](#) (109)
- [civil rights](#) (108)
- [debut novel](#) (83)
- [integration](#) (66)

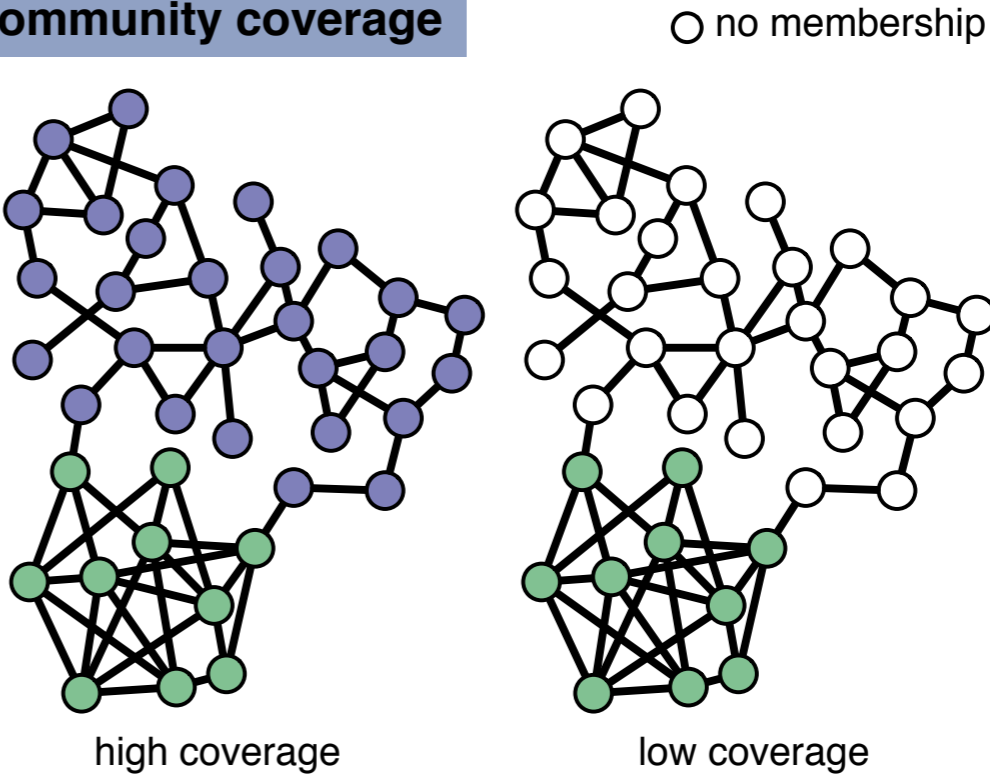
Your tags: [Add your first tag](#)

Quantitative Evaluation Framework

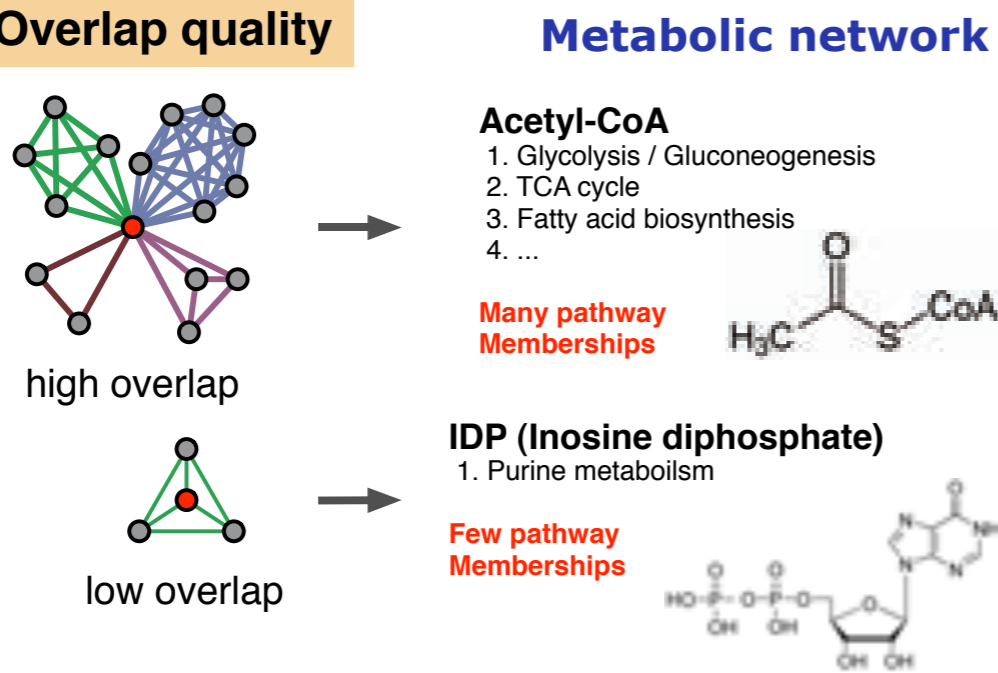
Community quality



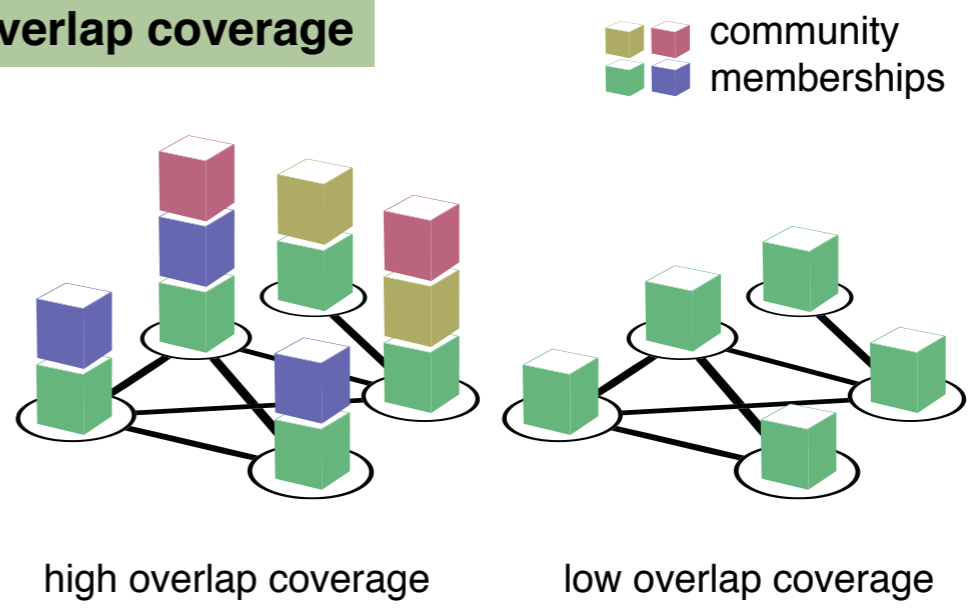
Community coverage



Overlap quality



Overlap coverage



network	description	N	$\langle k \rangle$	metadata	
				community	overlap
PPI (Y2H)	PPI network of <i>S. cerevisiae</i> obtained by yeast two-hybrid (Y2H) experiment [3]	1647	3.06	Set of each protein's known functions (GO terms) ^a	The number of GO terms
PPI (AP/MS)	Affinity purification mass spectrometry (AP/MS) experiment	1004	16.57	GO terms	GO terms
PPI (LC)	Literature curated (LC)	1213	4.21	GO terms	GO terms
PPI (all)	Union of Y2H, AP/MS, and LC PPI networks ^b	2729	8.92	GO terms	GO-terms
Metabolic	Metabolic network (metabolites connected by reactions) of <i>E. coli</i>	1042	16.81	Set of each metabolite's pathway annotations (KEGG) ^c	The number of KEGG pathway annotations
Phone	Social contacts between mobile phone users [15, 16, 17]	885989	6.34	Each user's most likely geographic location	Call activity (number of phone calls ^d)
Actor	Film actors that appear in the same movies during 2000–2009 [18]	67411	8.90	Set of plot keywords for all of the actor's films	Length of career (year of first role)
US Congress	Congressmen who co-sponsor bills during the 108th US Congress [19, 20]	390	38.95	Political ideology, from the common space score [21, 22]	Seniority (number of congresses served)
Philosopher	Philosophers and their philosophical influences, from the English Wikipedia ^e	1219	9.80	Set of (wikipedia) hyperlinks exiting in the philosopher's page	Number of wikipedia subject categories
Word Assoc.	English words that are often mentally associated [23]	5018	22.02	Set of each word's senses, as documented by WordNet ^f	Number of senses
Amazon.com	Products that users frequently buy together	18142	5.09 ^g	Set of each product's user tags (annotations)	Number of product categories

Measures

- overlap coverage
- ▒ community coverage
- overlap quality
- community quality

Methods

- L – Links
- C – Clique Percolation
- G – Greedy Modularity
- I – Infomap

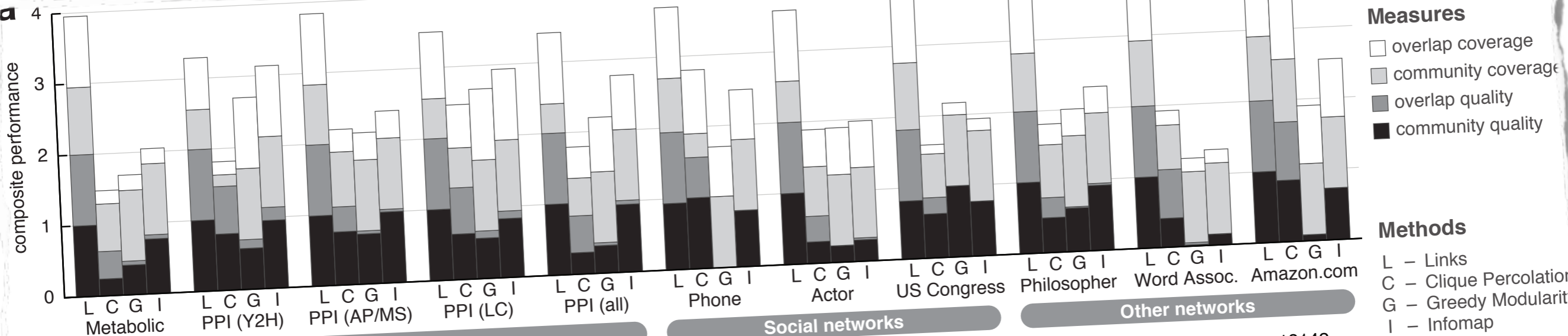
G I
n.com



42

9

a



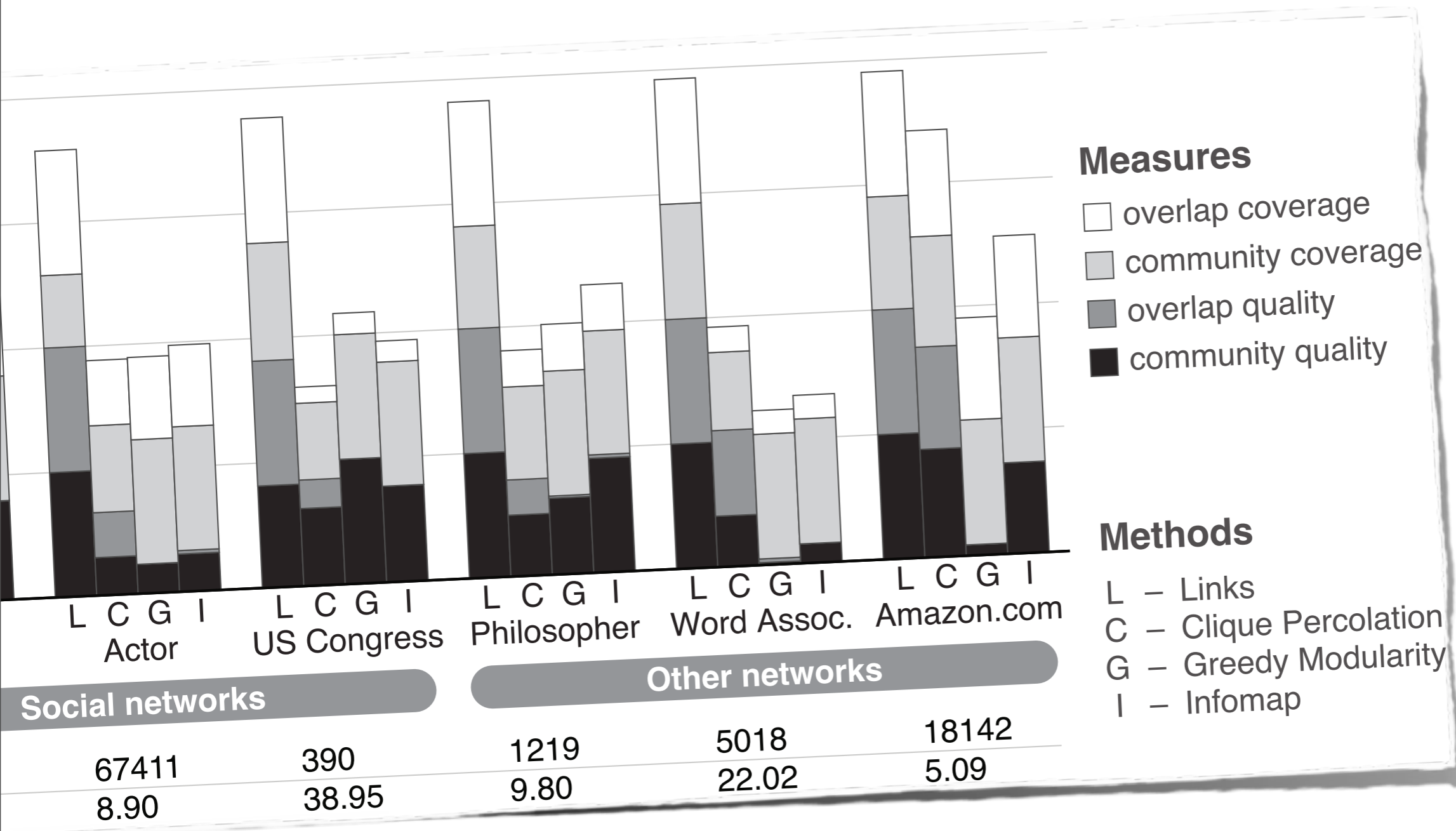
Measures

- overlap coverage
- community coverage
- overlap quality
- community quality

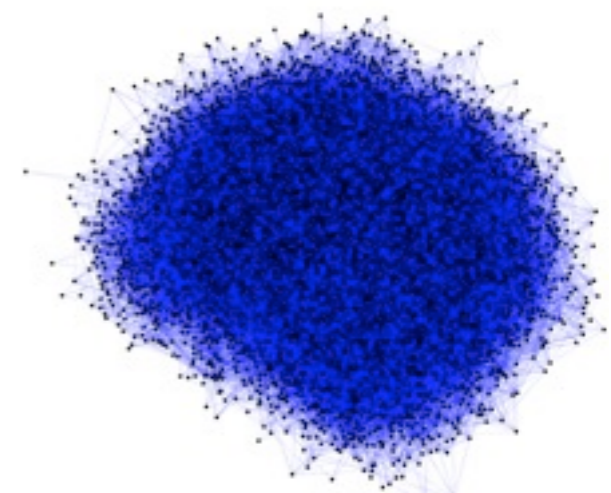
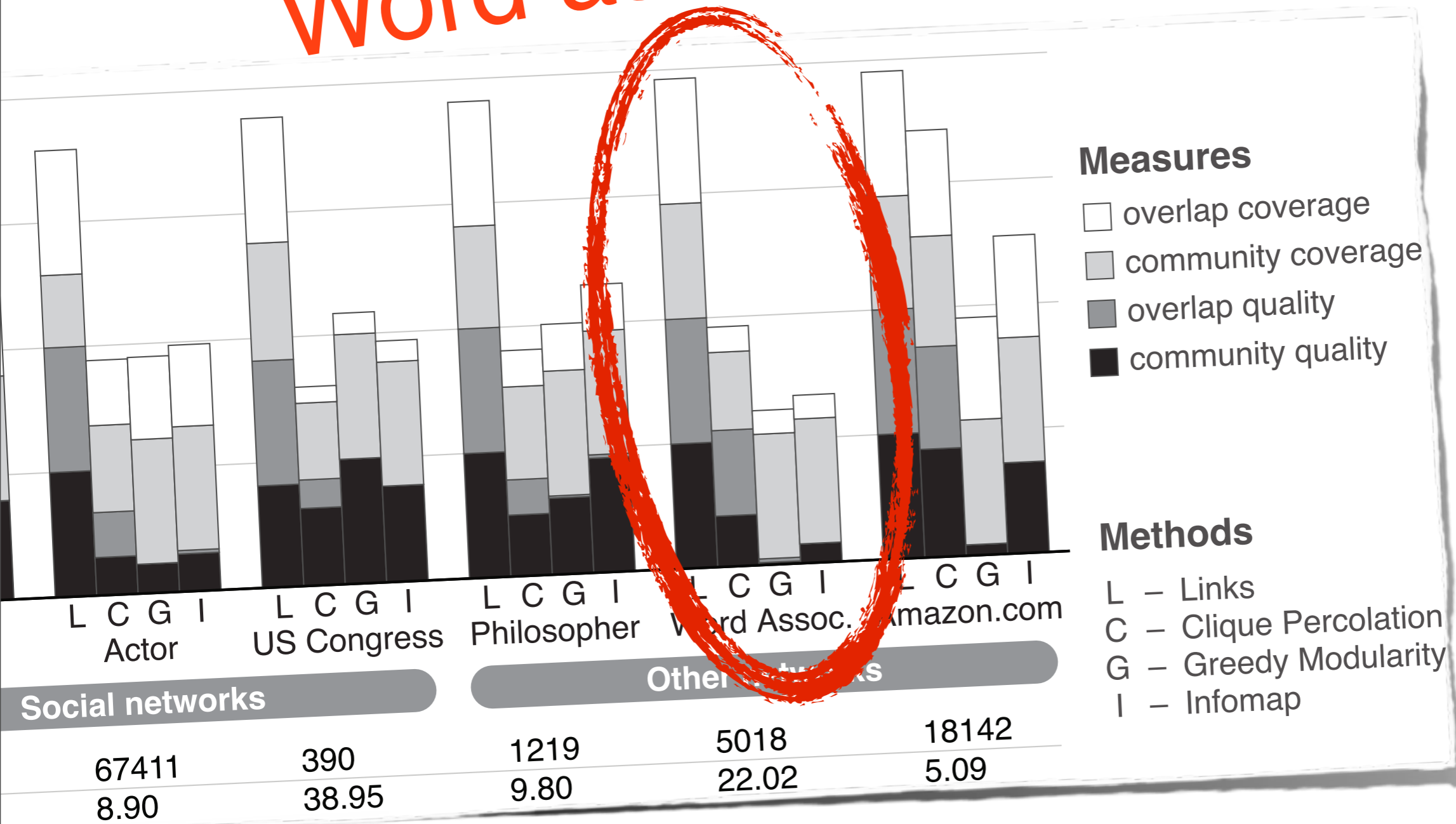
Methods

- L – Links
- C – Clique Percolation
- G – Greedy Modularity
- I – Infomap

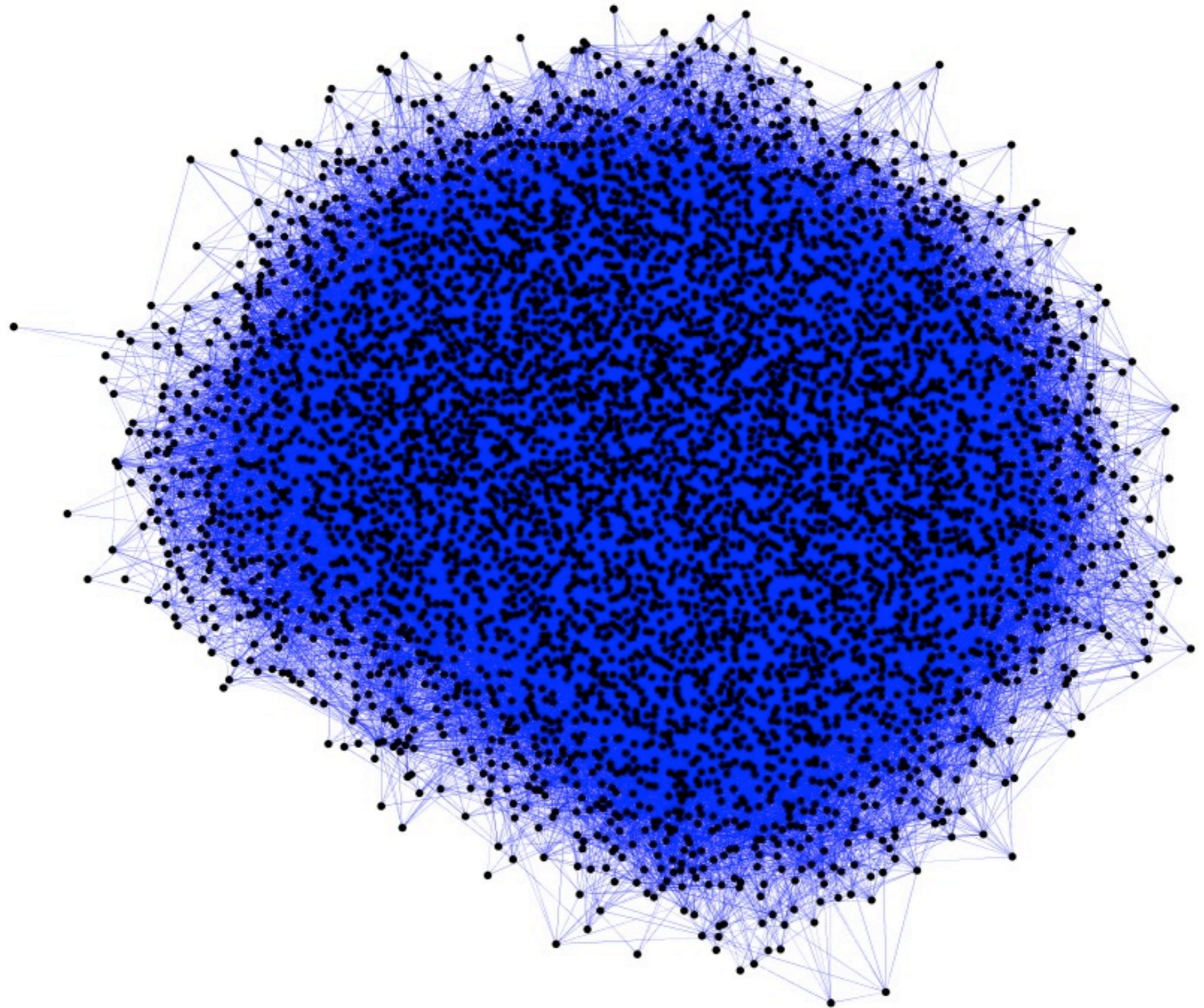
	Biological networks				Social networks			Other networks			
	Metabolic	PPI (Y2H)	PPI (AP/MS)	PPI (LC)	PPI (all)	Phone	Actor	US Congress	Philosopher	Word Assoc.	Amazon.com
<i>N</i>	1042	1647	1004	1213	2729	885989	67411	390	1219	5018	18142
$\langle k \rangle$	16.81	3.06	16.57	4.21	8.92	6.34	8.90	38.95	9.80	22.02	5.09

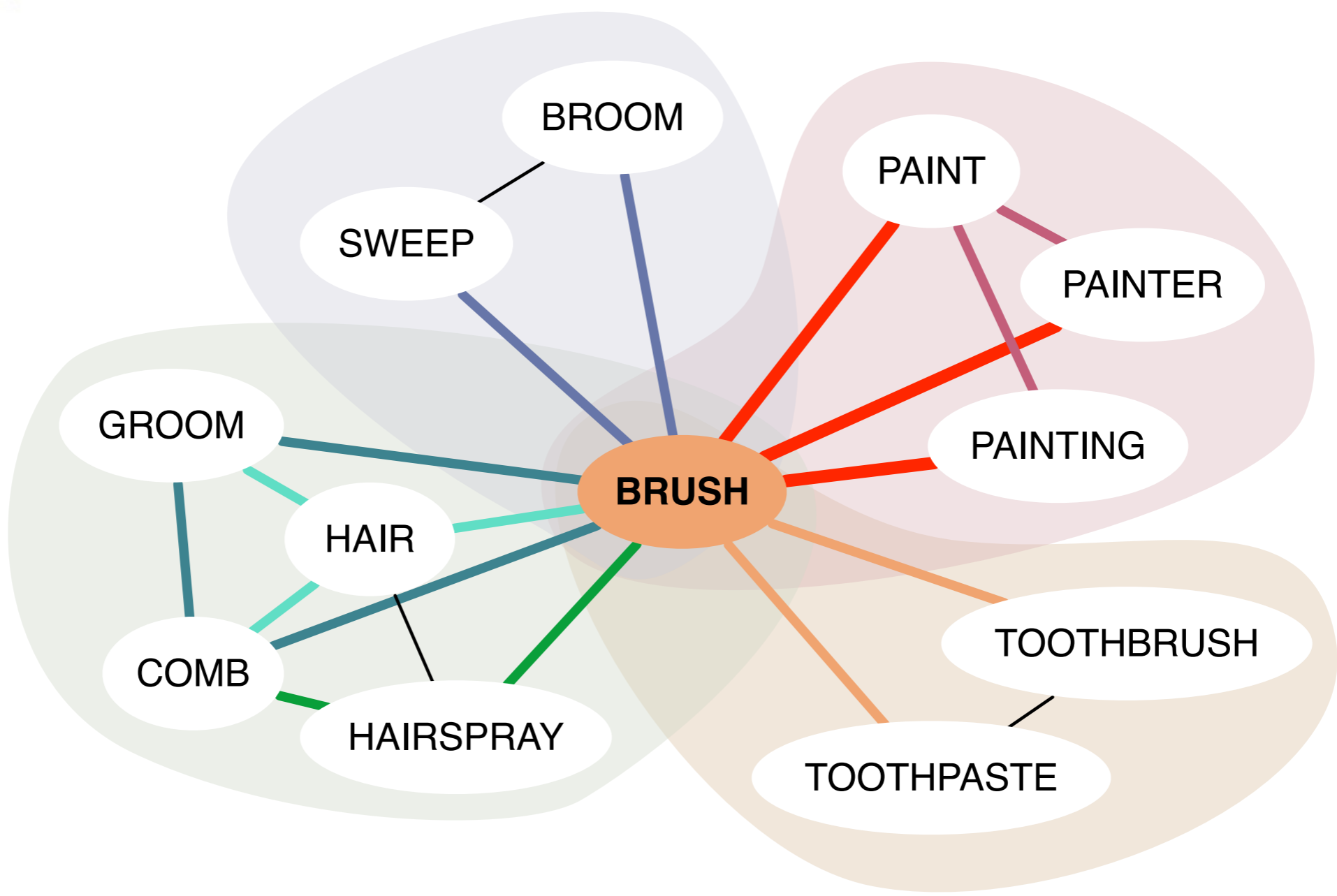
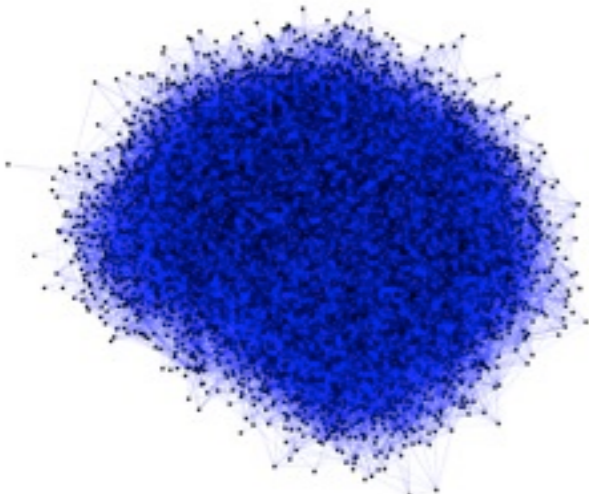


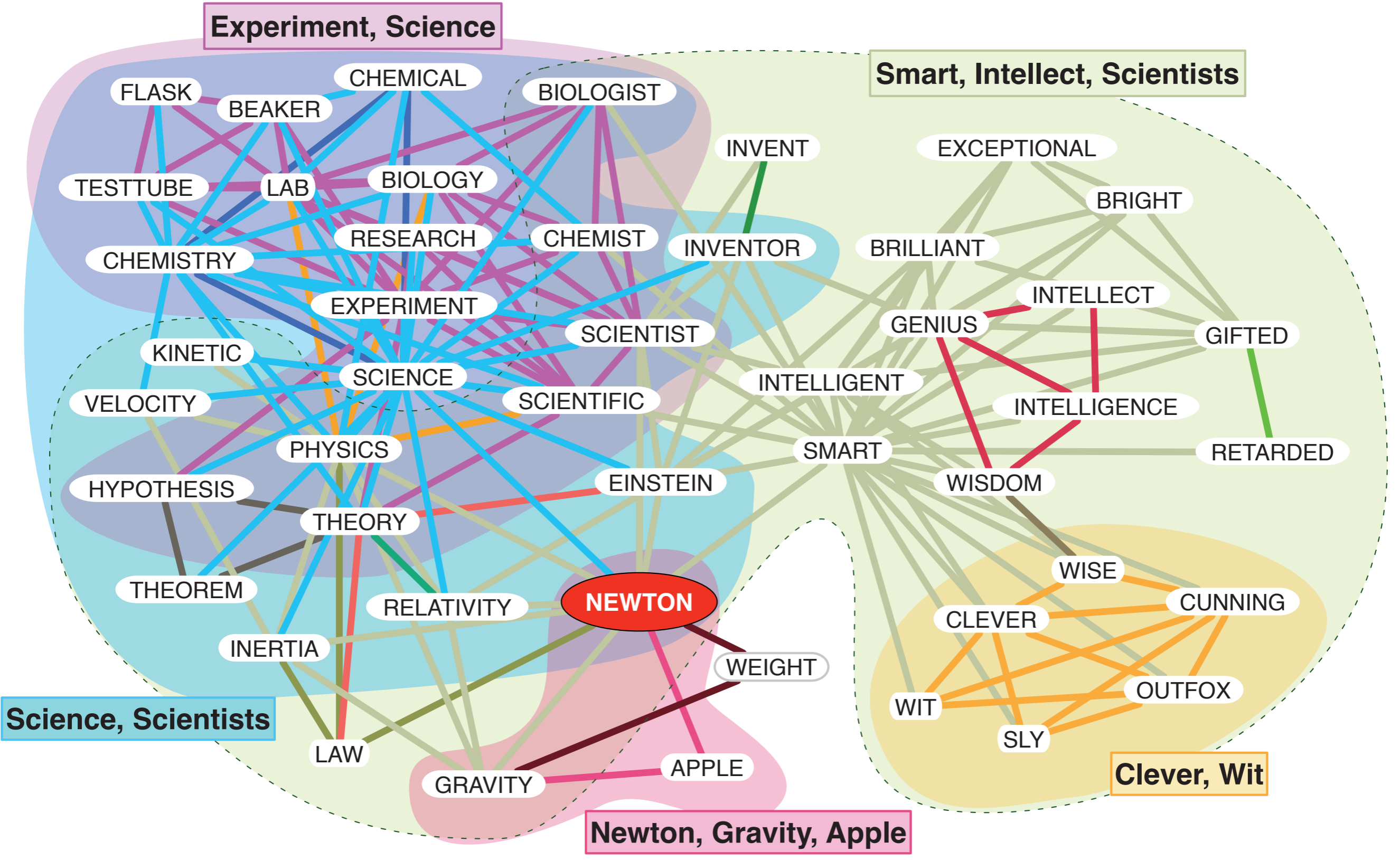
Word assoc.

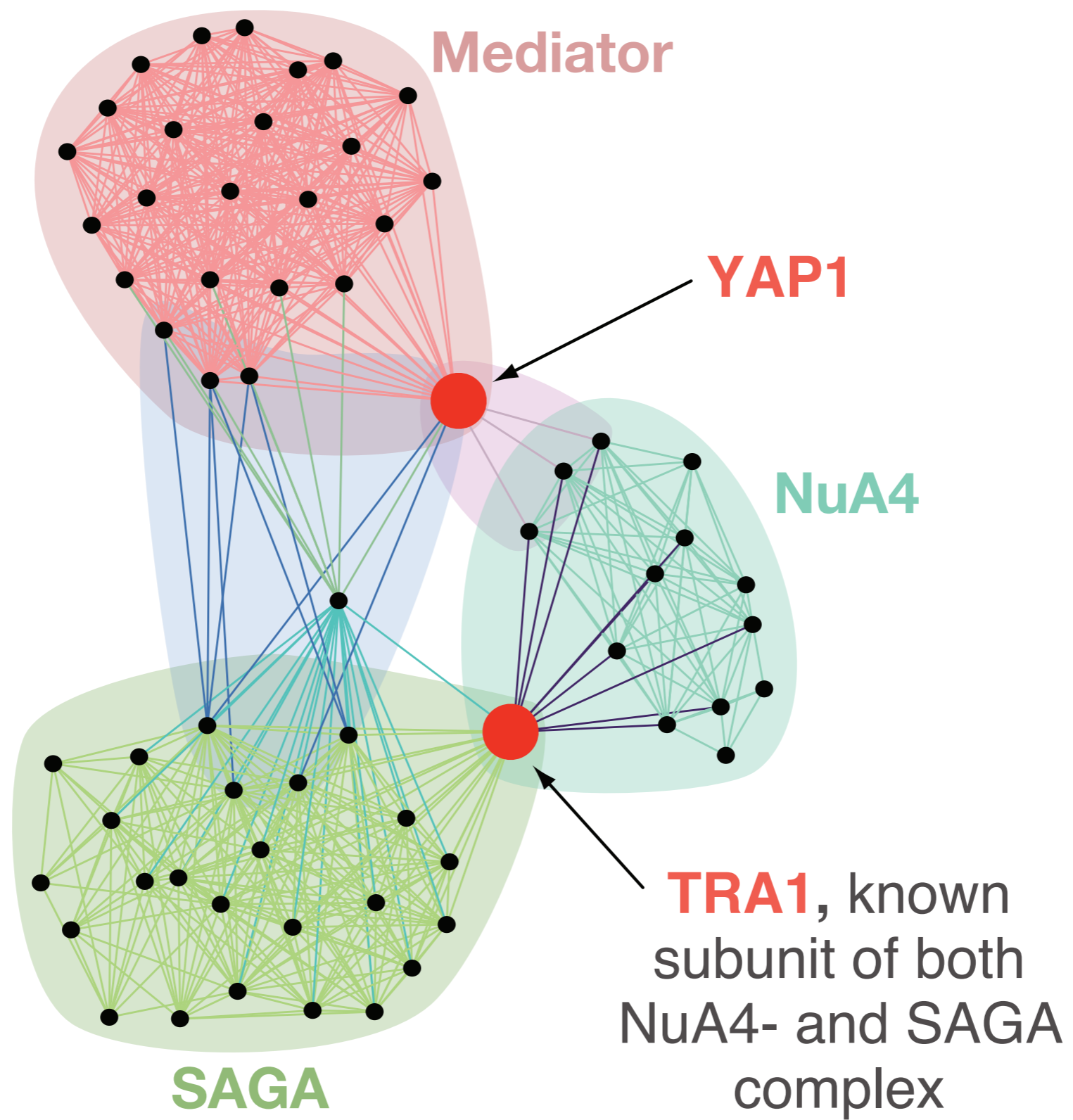


Examples

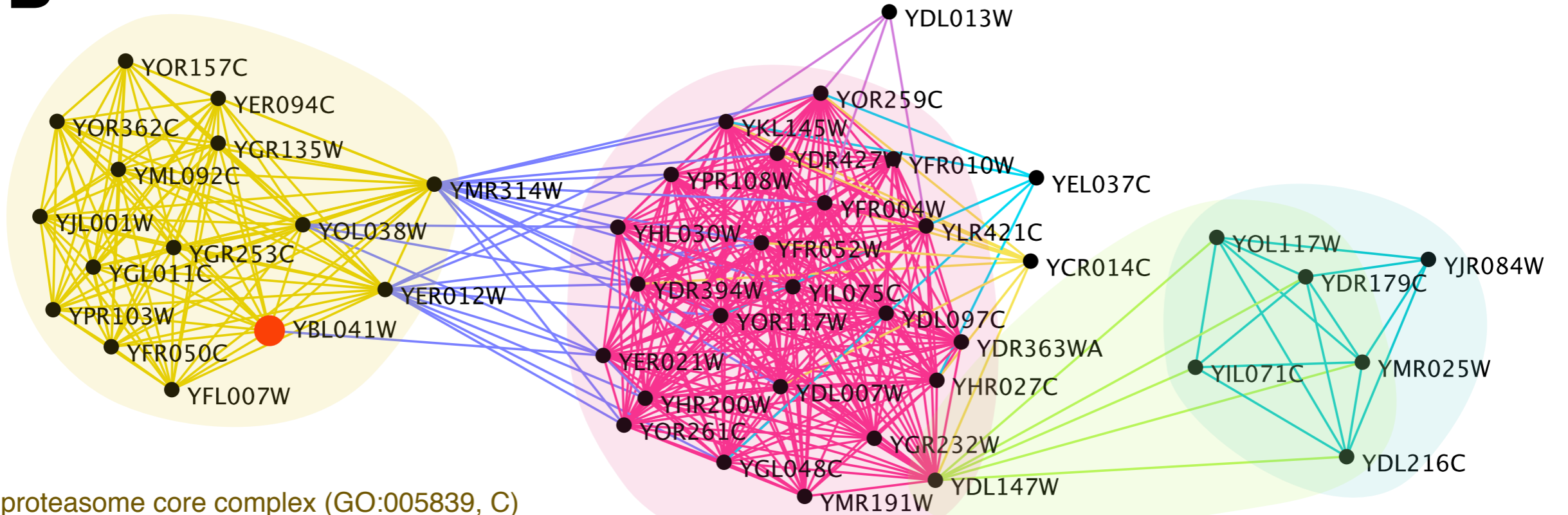








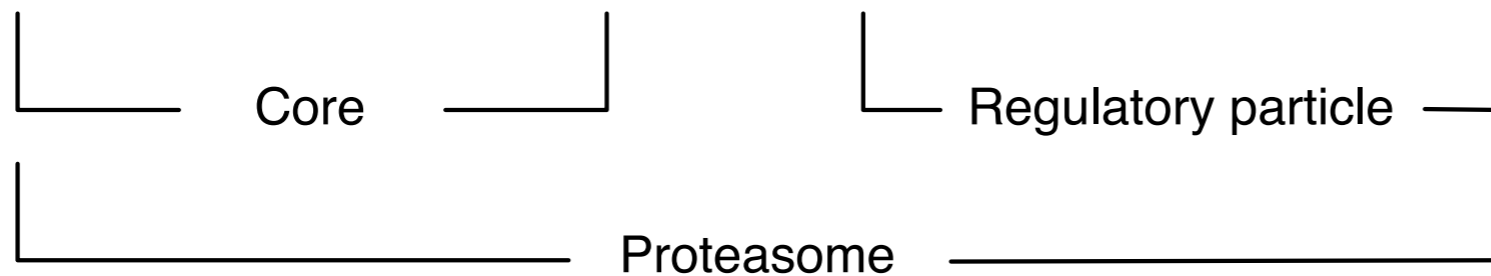
B



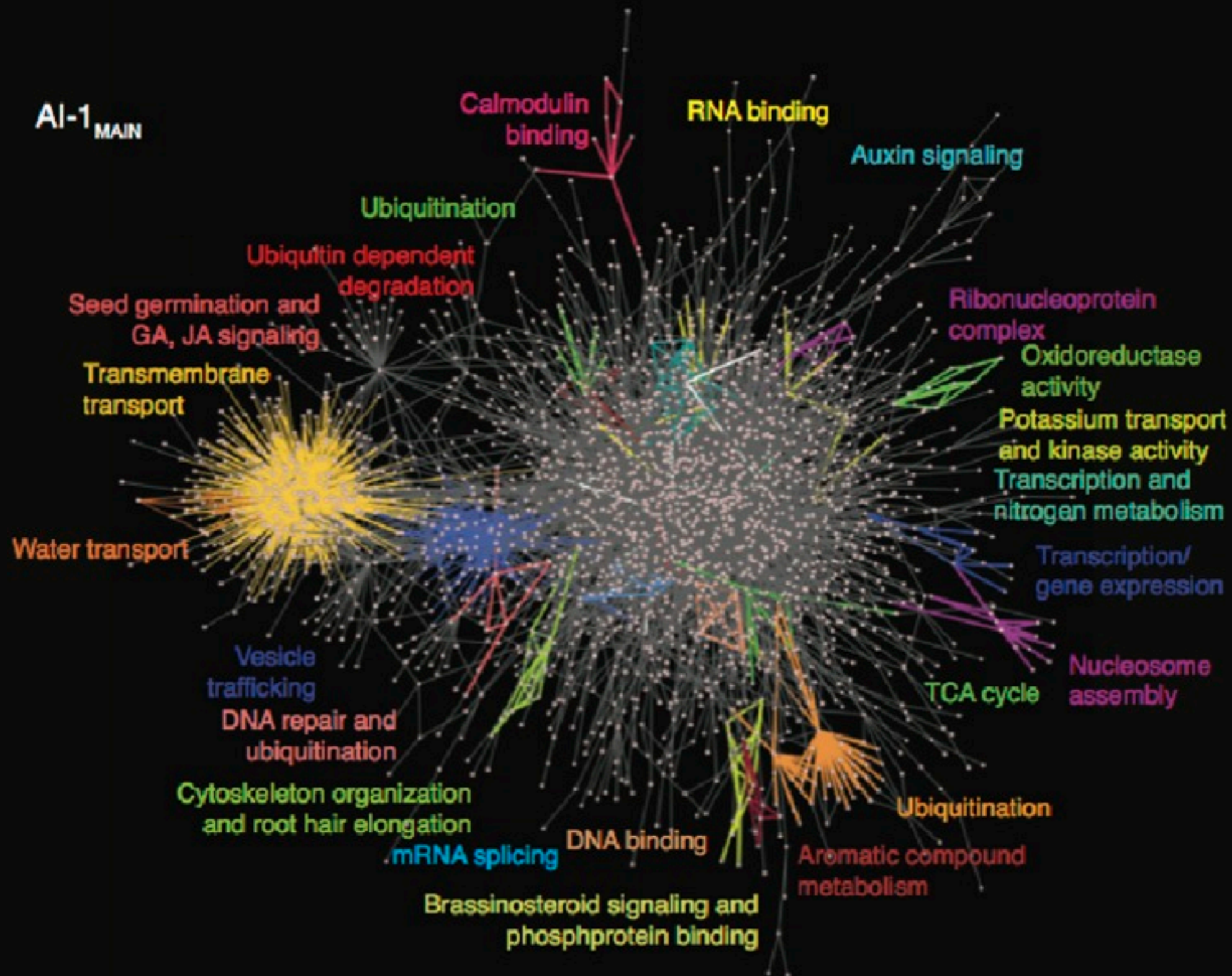
proteasome core complex (GO:005839, C)
 threonine-type endopeptidase activity (GO:0004298, F)
 ubiquitin-dependent protein catabolic process (GO:0006511, P)

proteasome regulatory particle (GO:0005838, C)
 ubiquitin-dependent protein catabolic process (GO:0006511, P)
 endopeptidase activity (GO:0004175, F)

Signalosome (GO:0008180, C)
 Protein deneddylation (GO:0000338, P)

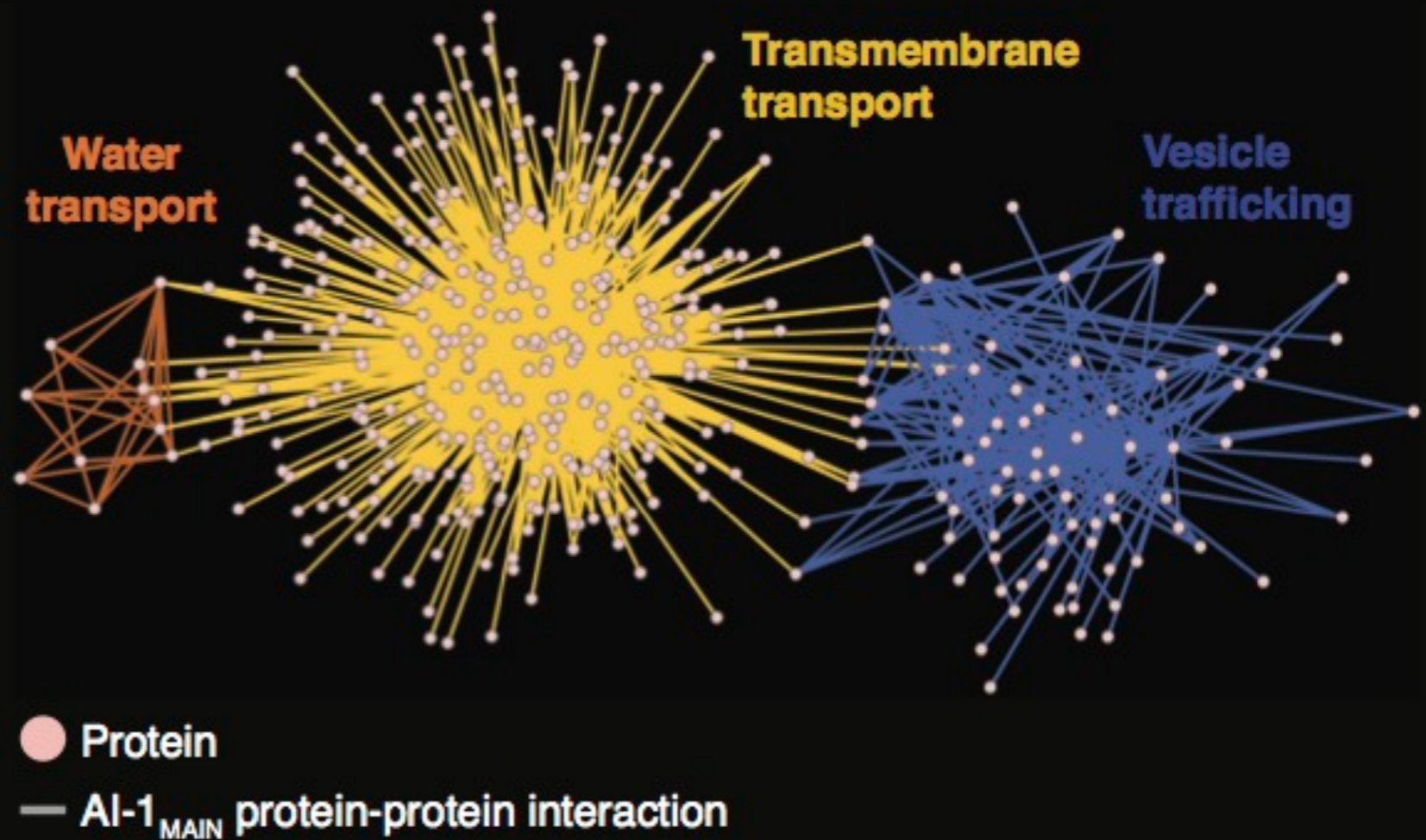
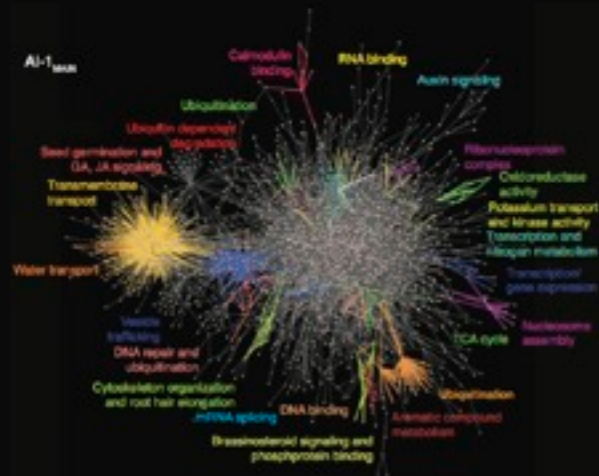


The first plant (genomic scale) interactome

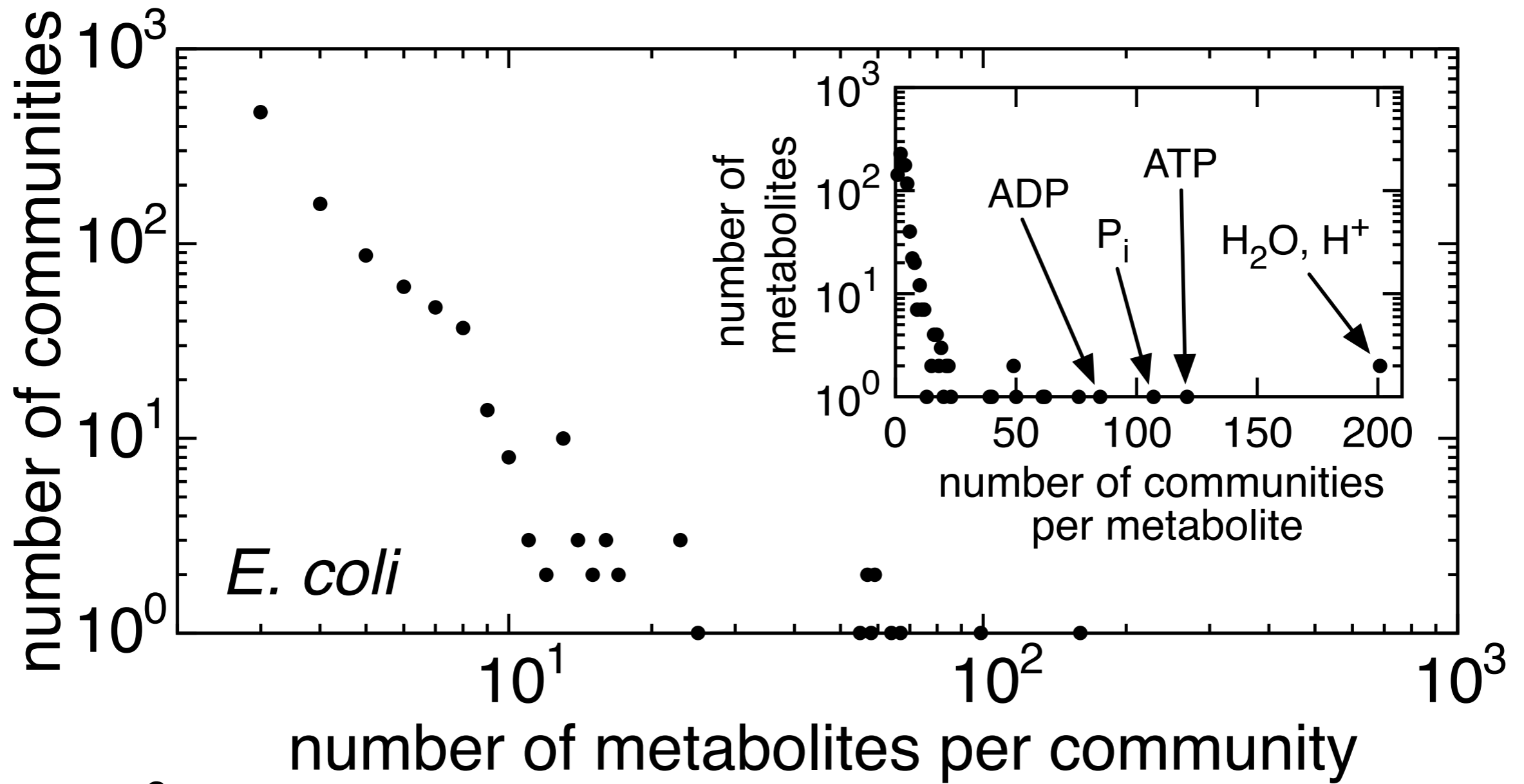


Arabidopsis Interactome Mapping Consortium, *Science*, 2011

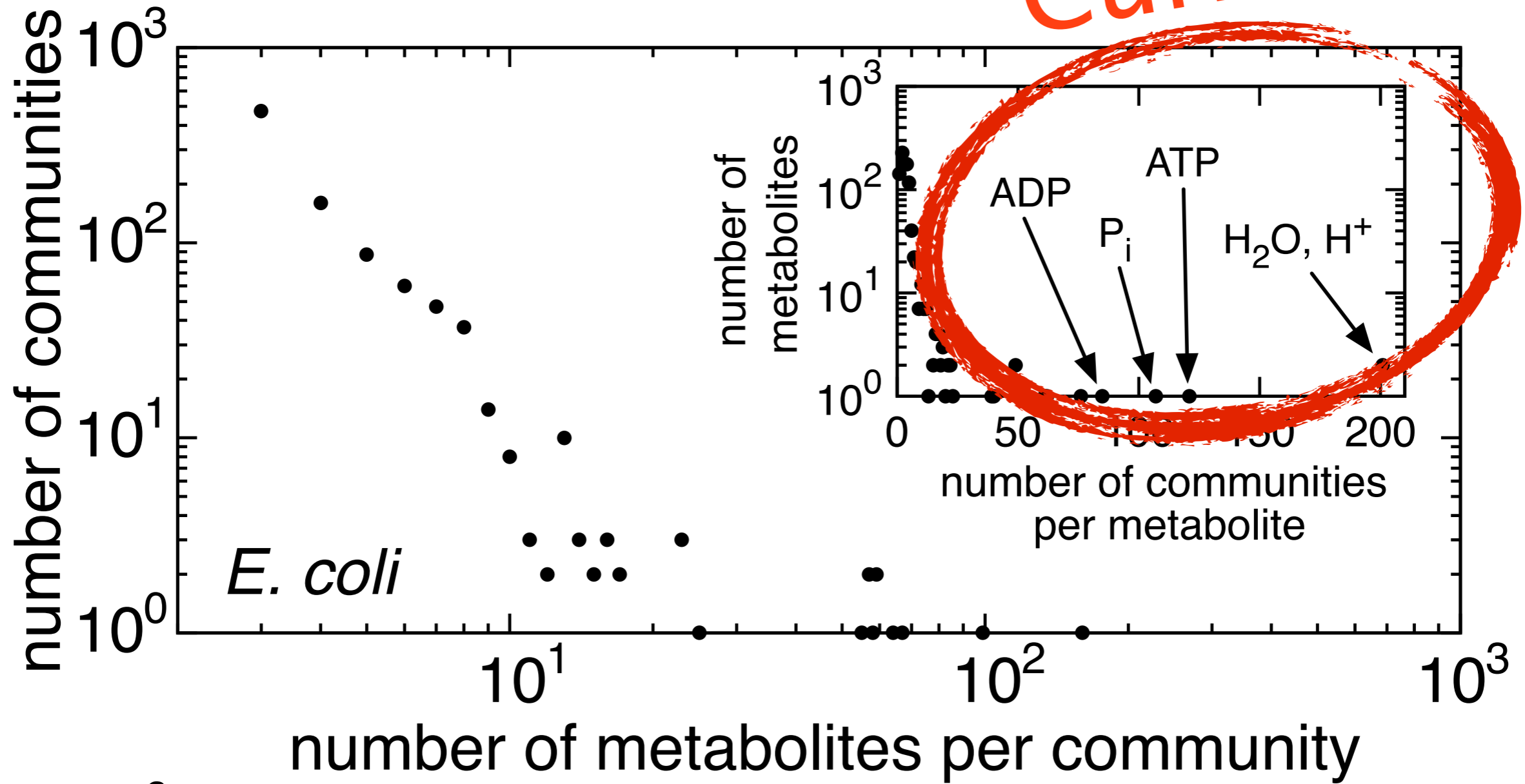
The first plant (genomic scale) interactome



Arabidopsis Interactome Mapping Consortium, *Science*, 2011



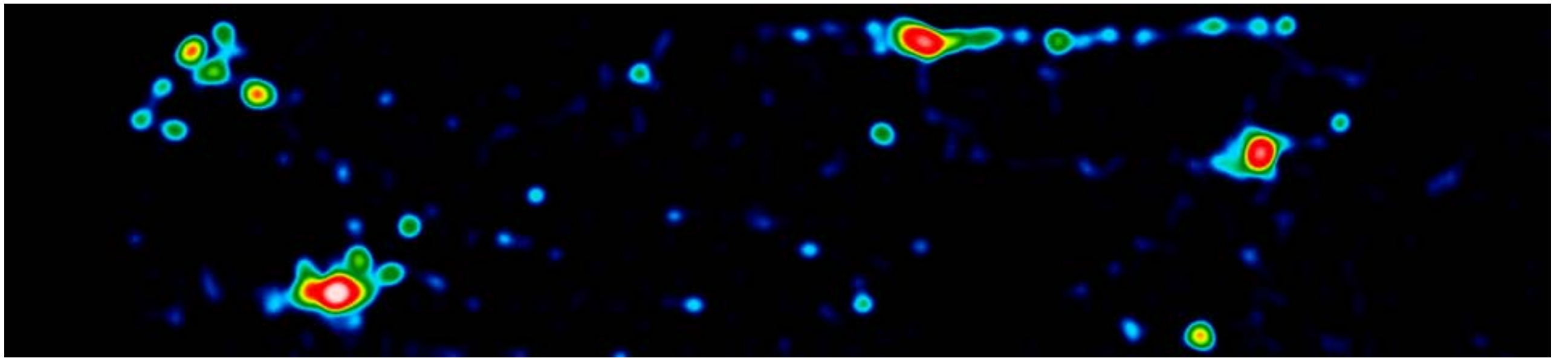
Currencies

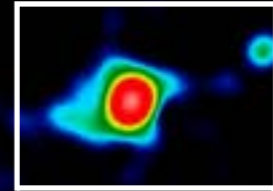
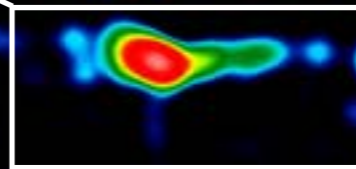
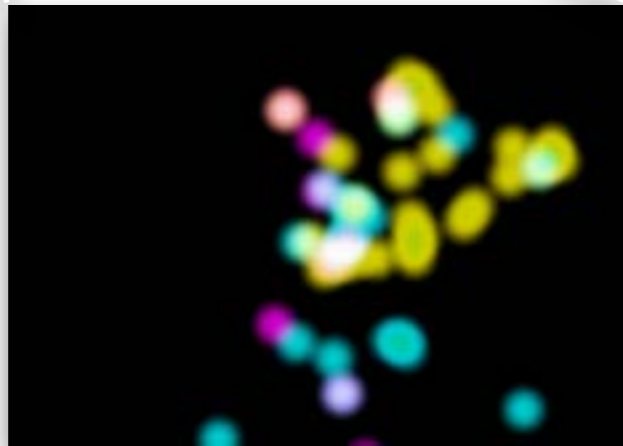
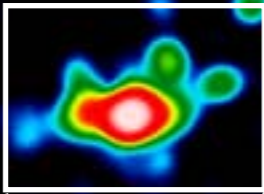
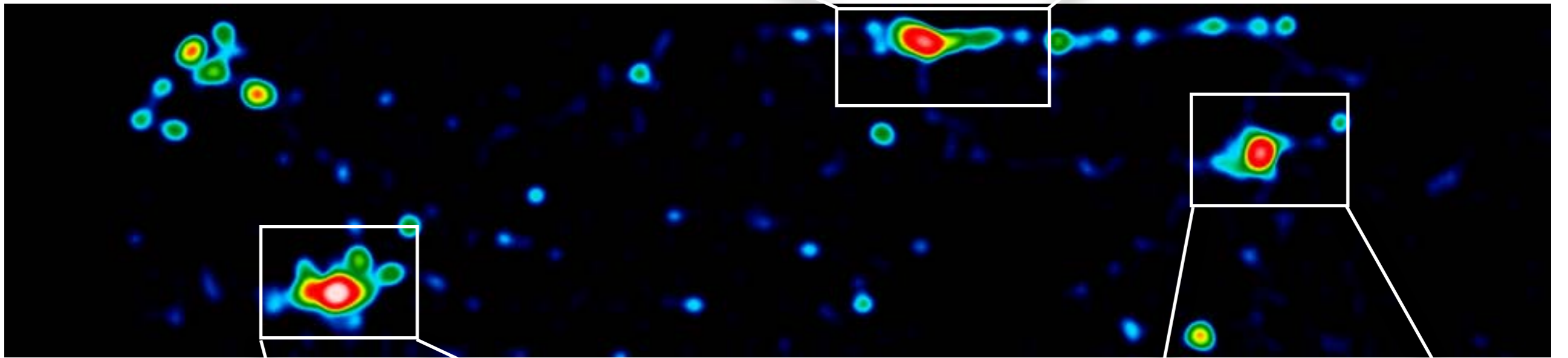
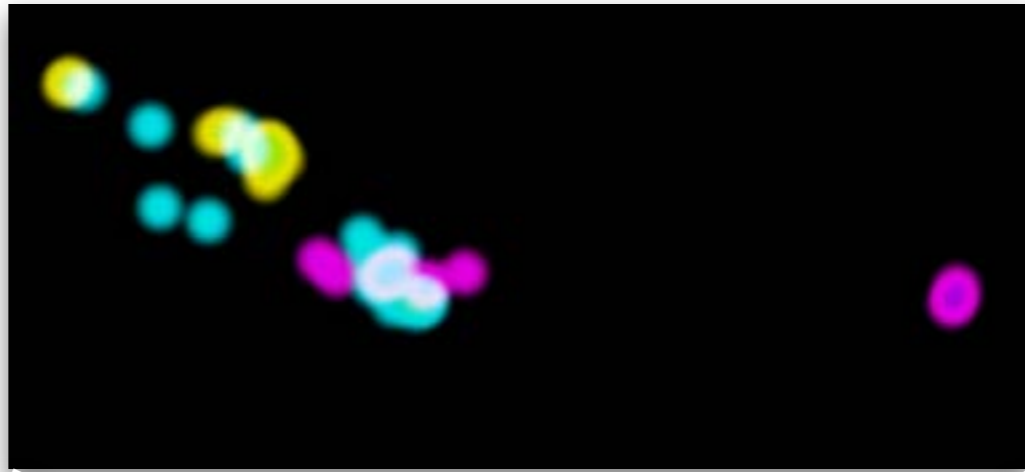


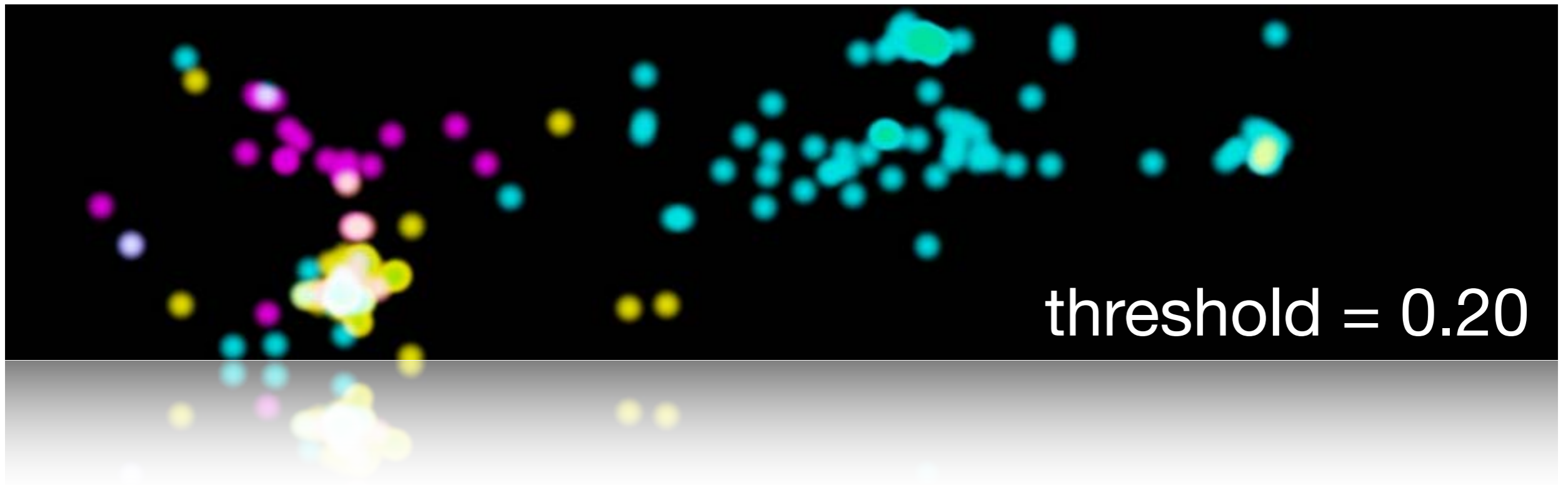
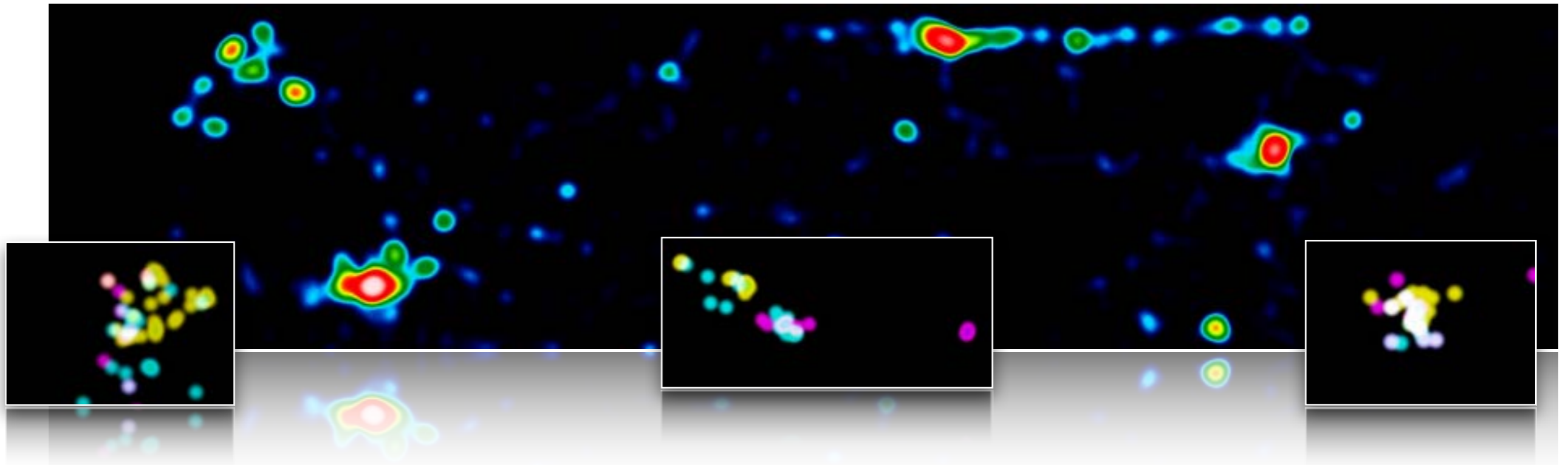
Hierarchical organization

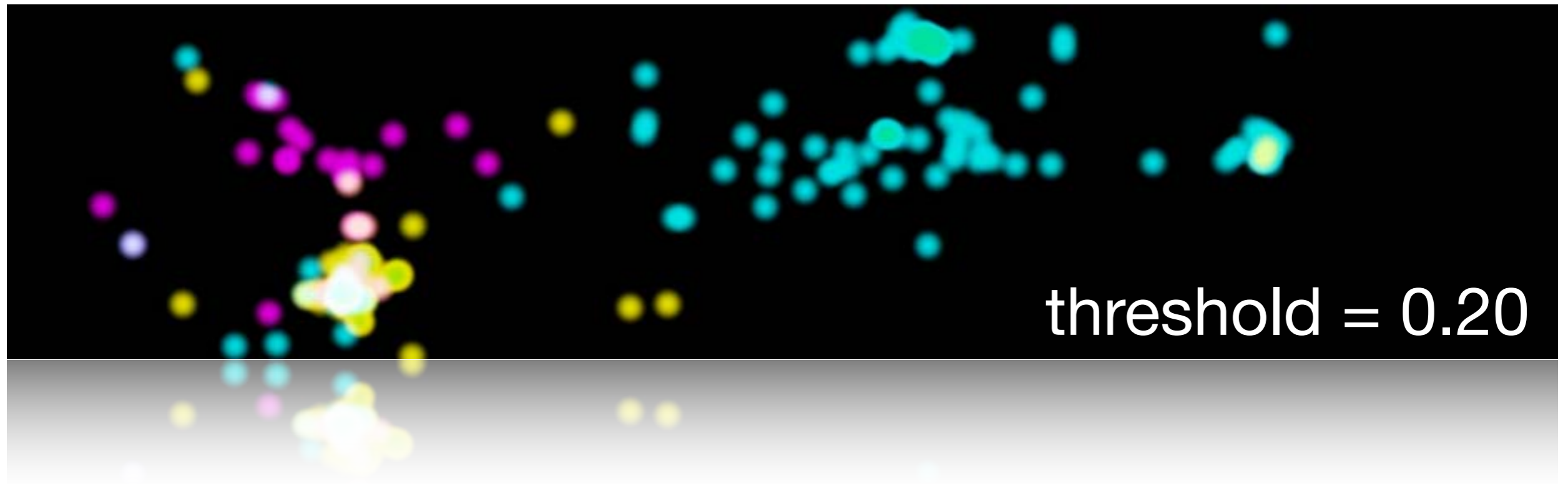
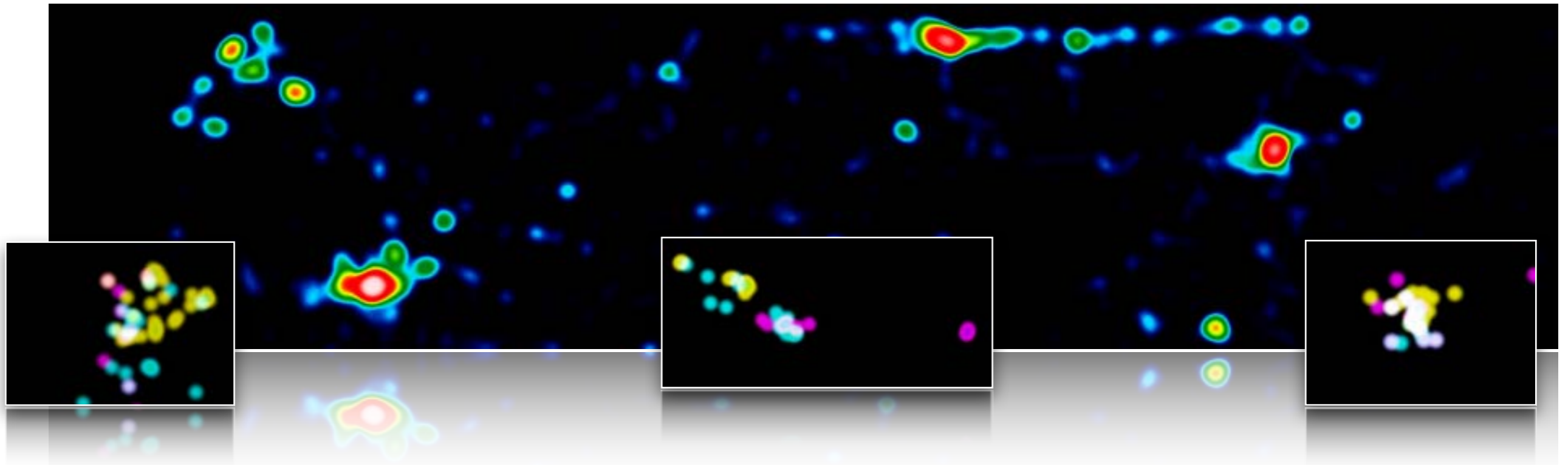
~600k nodes
~3M edges

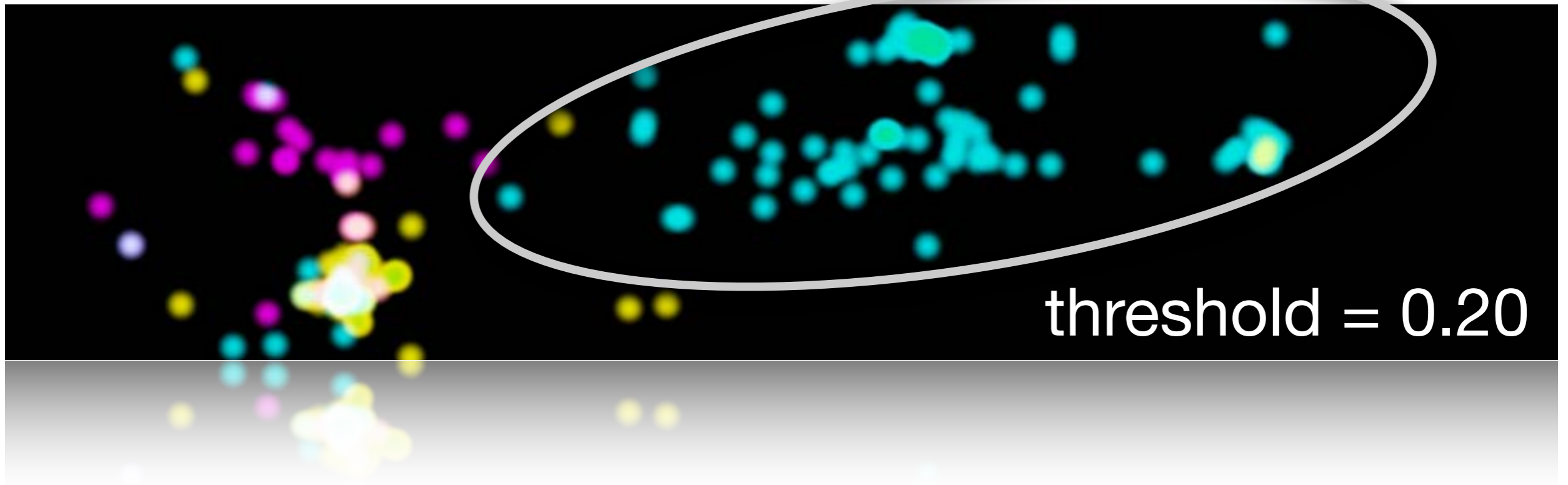
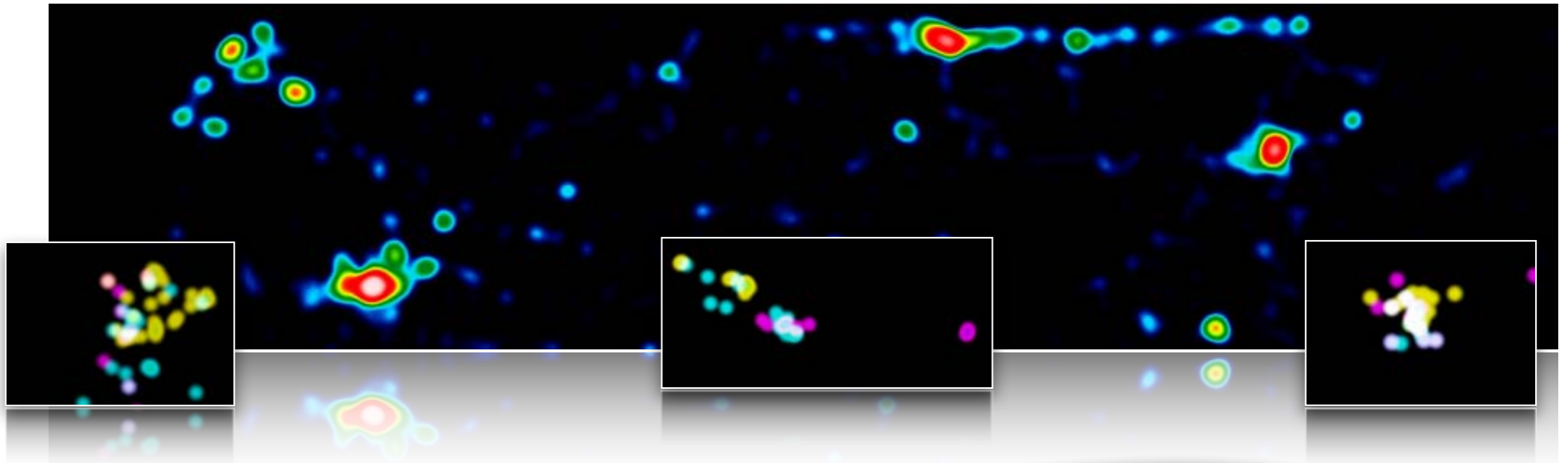




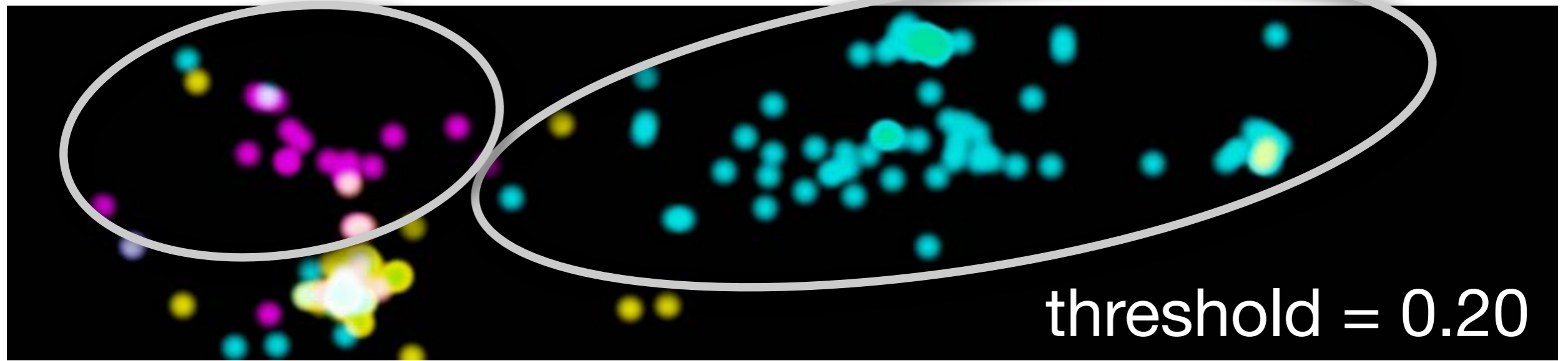
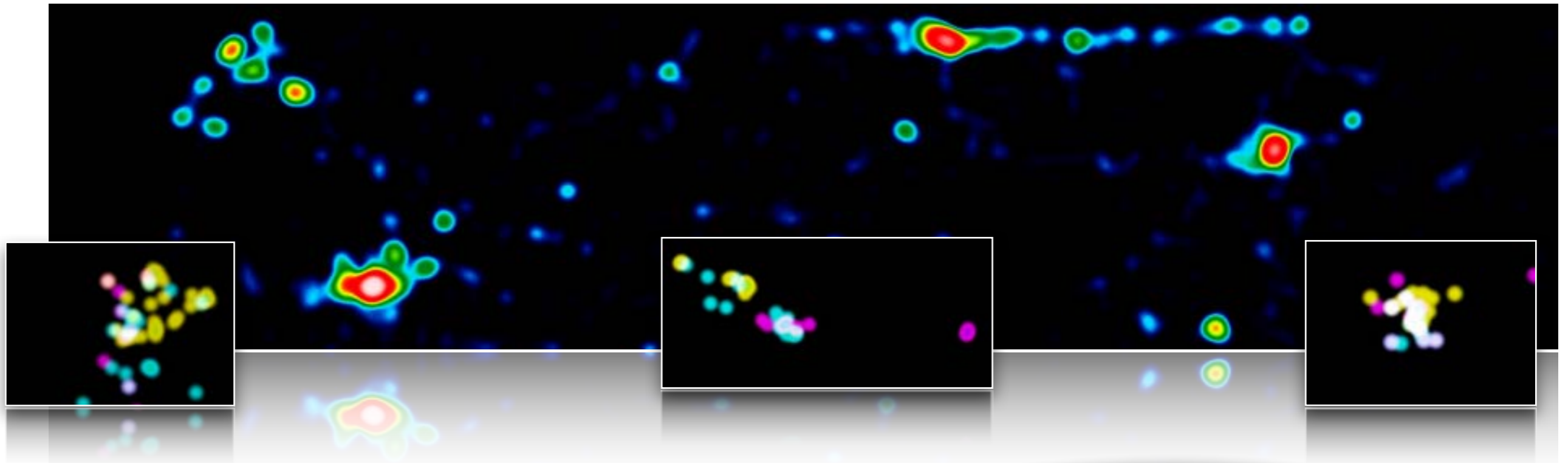




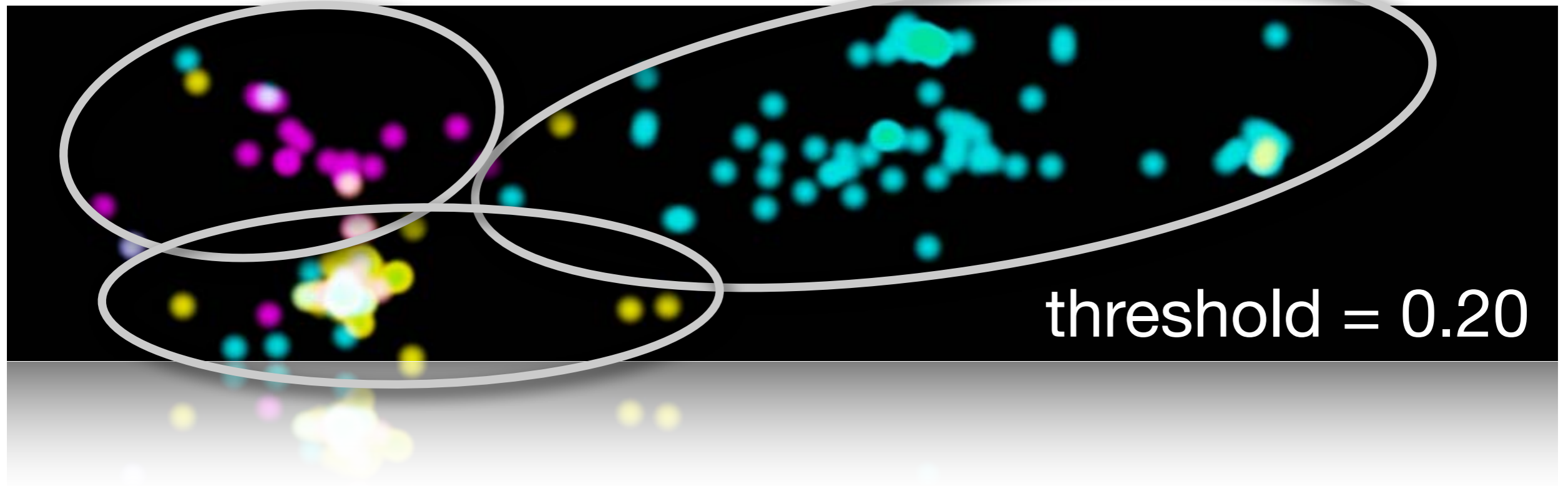
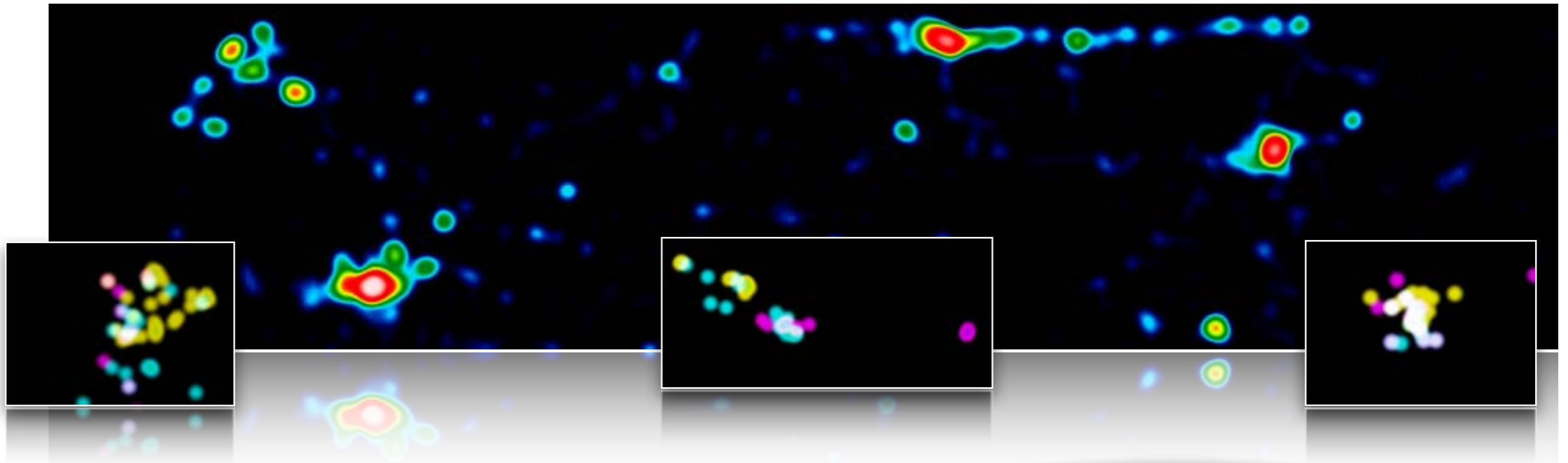


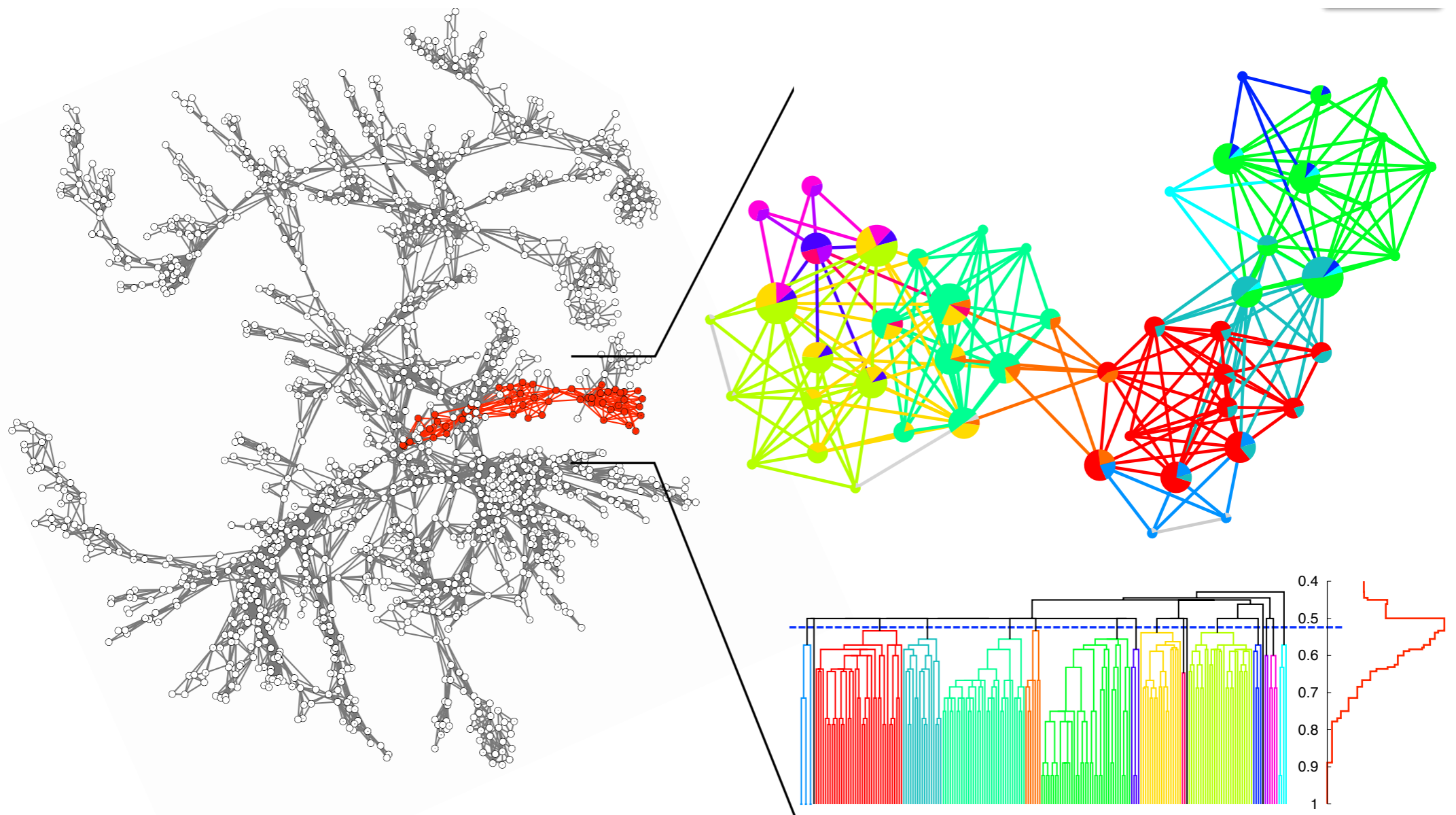
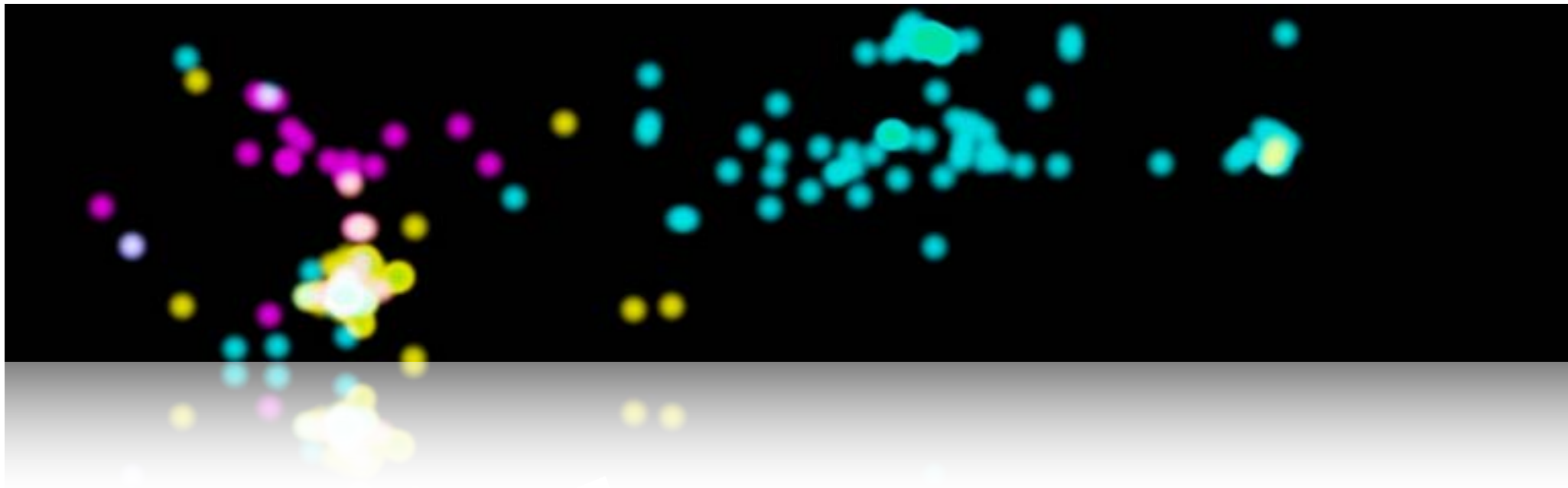


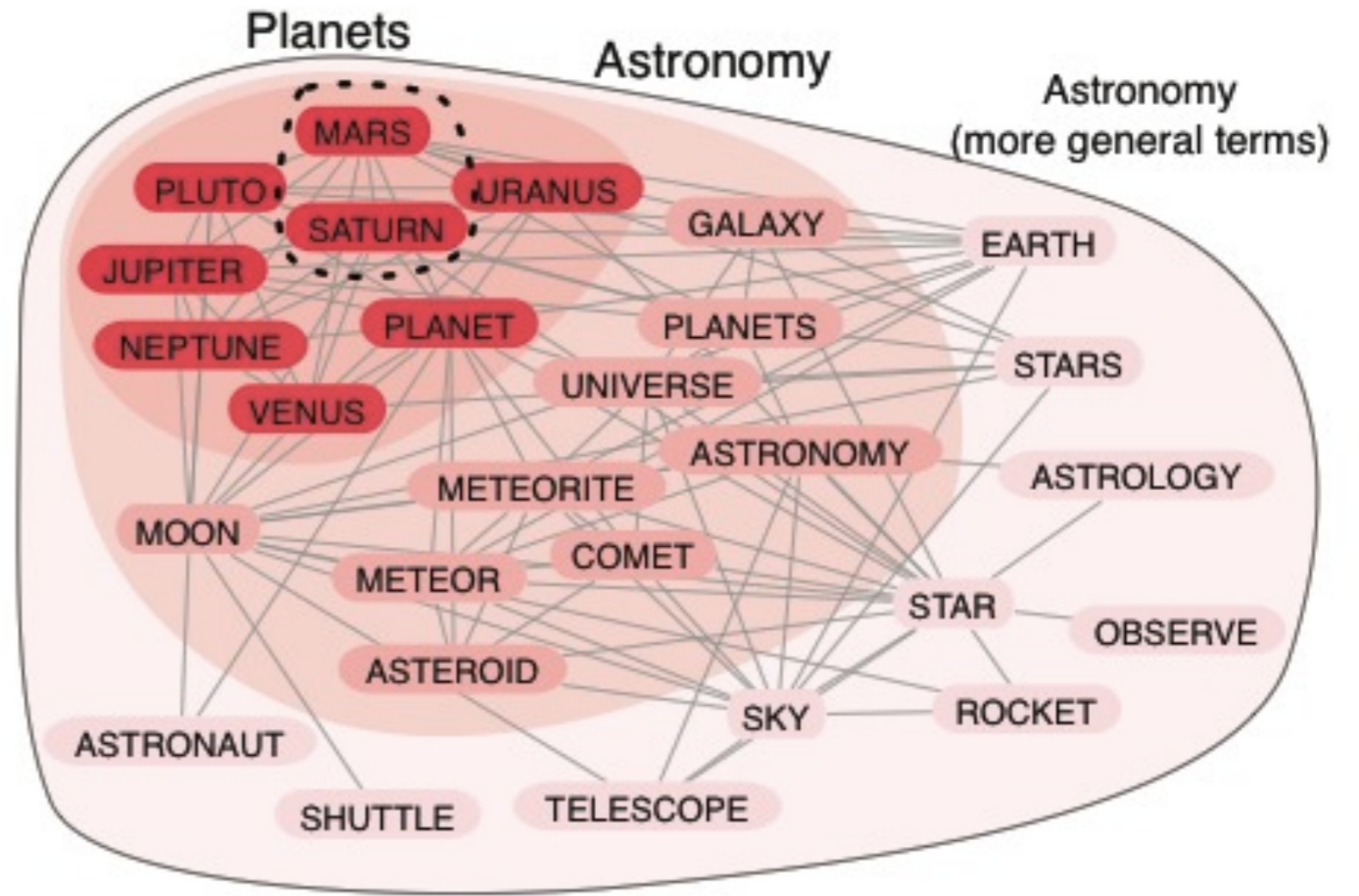
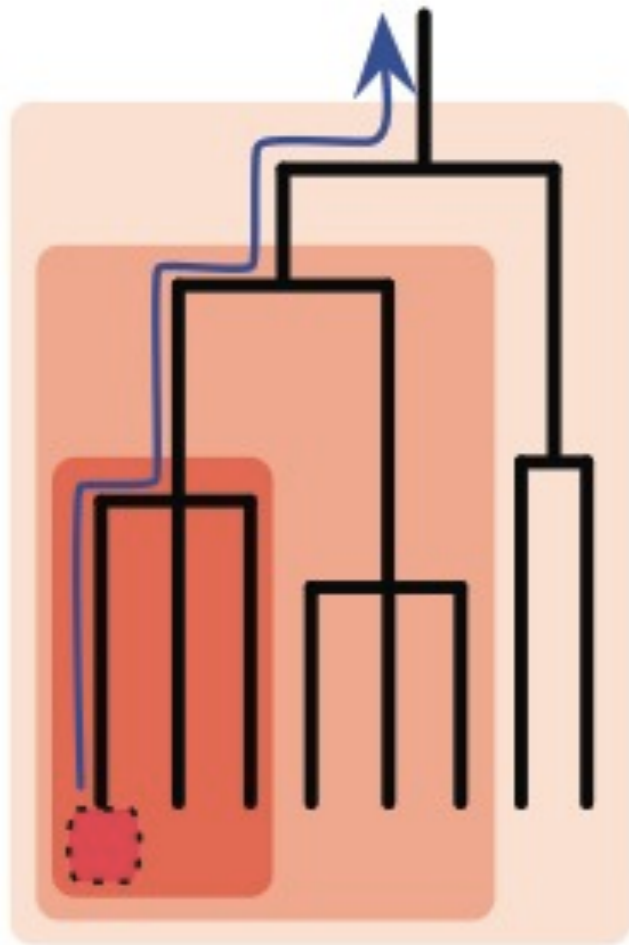
threshold = 0.20



threshold = 0.20

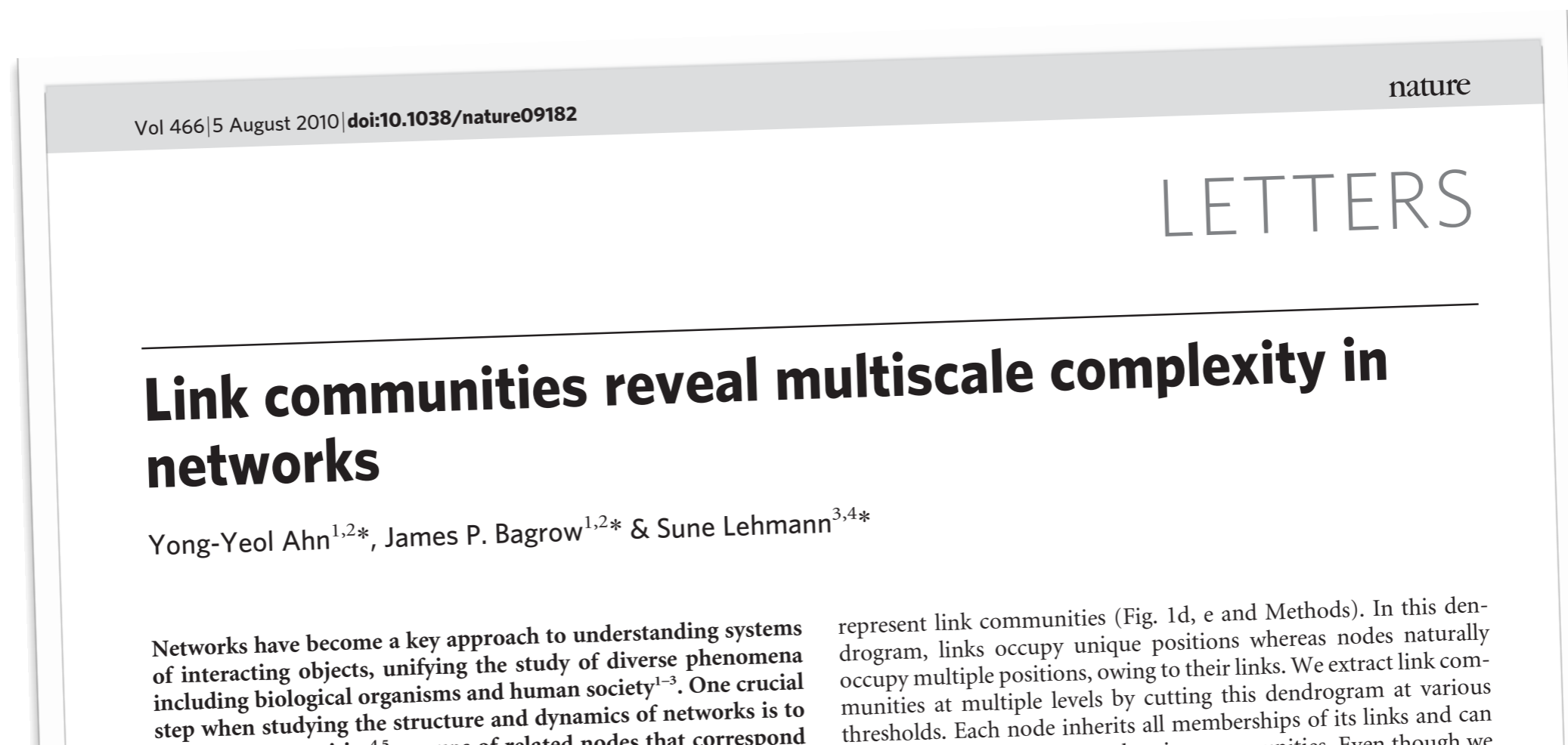






Summary

- Networks matter.
- Link (edge) perspective is useful.



Acknowledgements



James P. Bagrow



Sune Lehmann

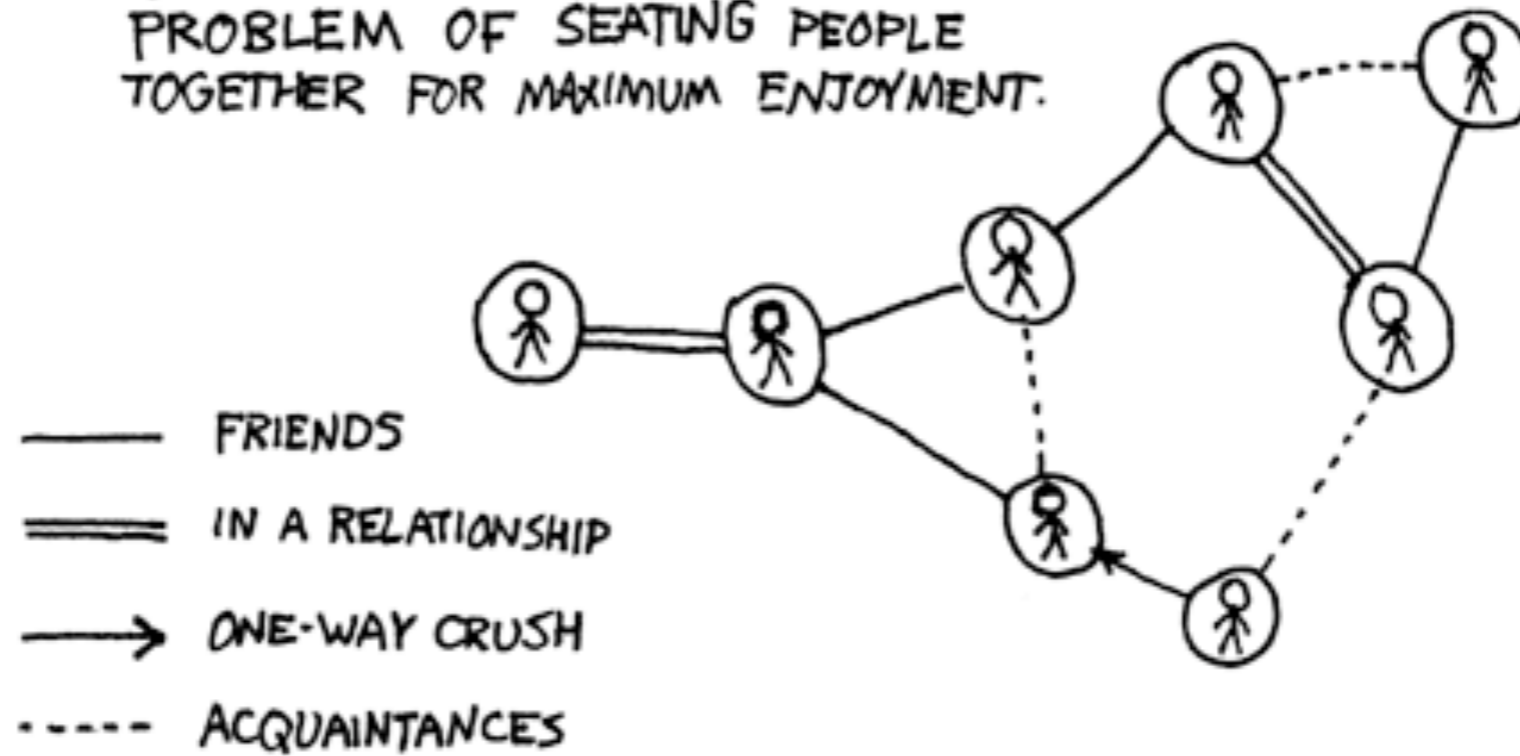


Albert-László Barabási

T. S. Evans, R. Lambiotte

Line Graphs, Link Partitions and
Overlapping Communities

AT THE MOVIES, I GET FRUSTRATED
WHEN WE FILE INTO OUR ROW
HAPHAZARDLY, IGNORING THE
COMPUTATIONALLY DIFFICULT
PROBLEM OF SEATING PEOPLE
TOGETHER FOR MAXIMUM ENJOYMENT.

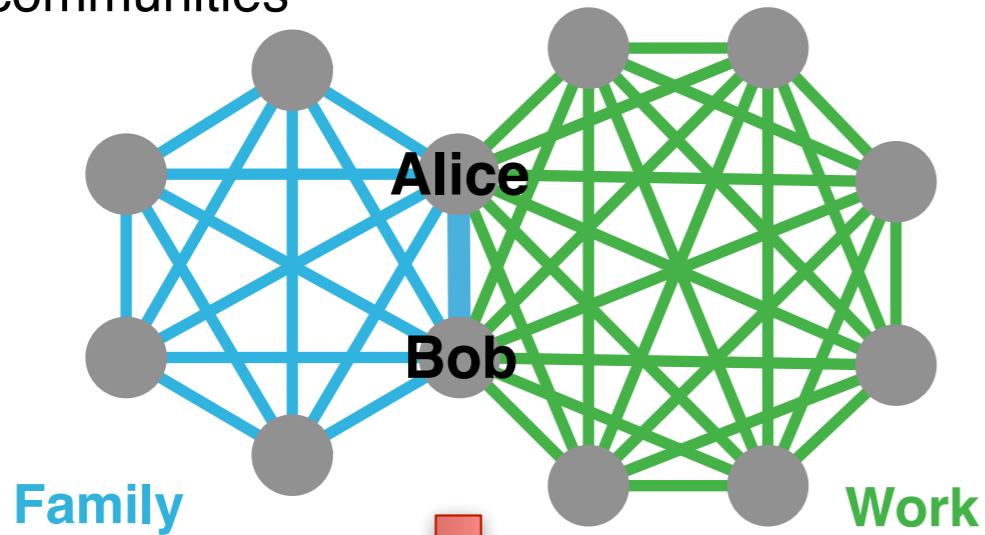


GUYS! THIS IS NOT
SOCIAALLY OPTIMAL!

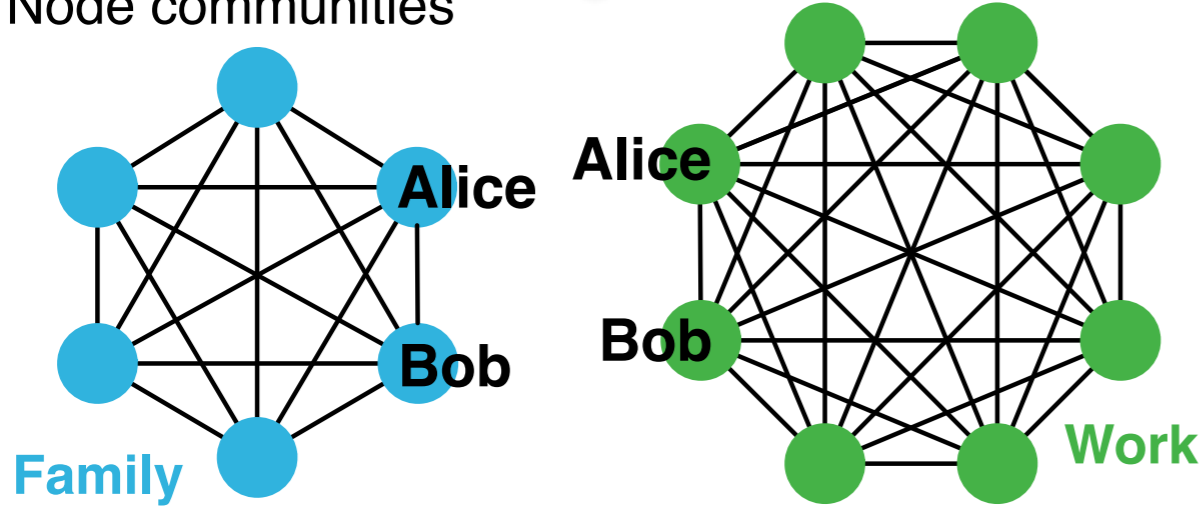


a Spouses **Alice** and **Bob** also work together

Link communities



Node communities



The **Alice-Bob link** was placed in **family** but both **home** and **work** relationships are identified

b Word Association examples

