

# Introduction to Network Science

Yong-Yeol “YY” Ahn



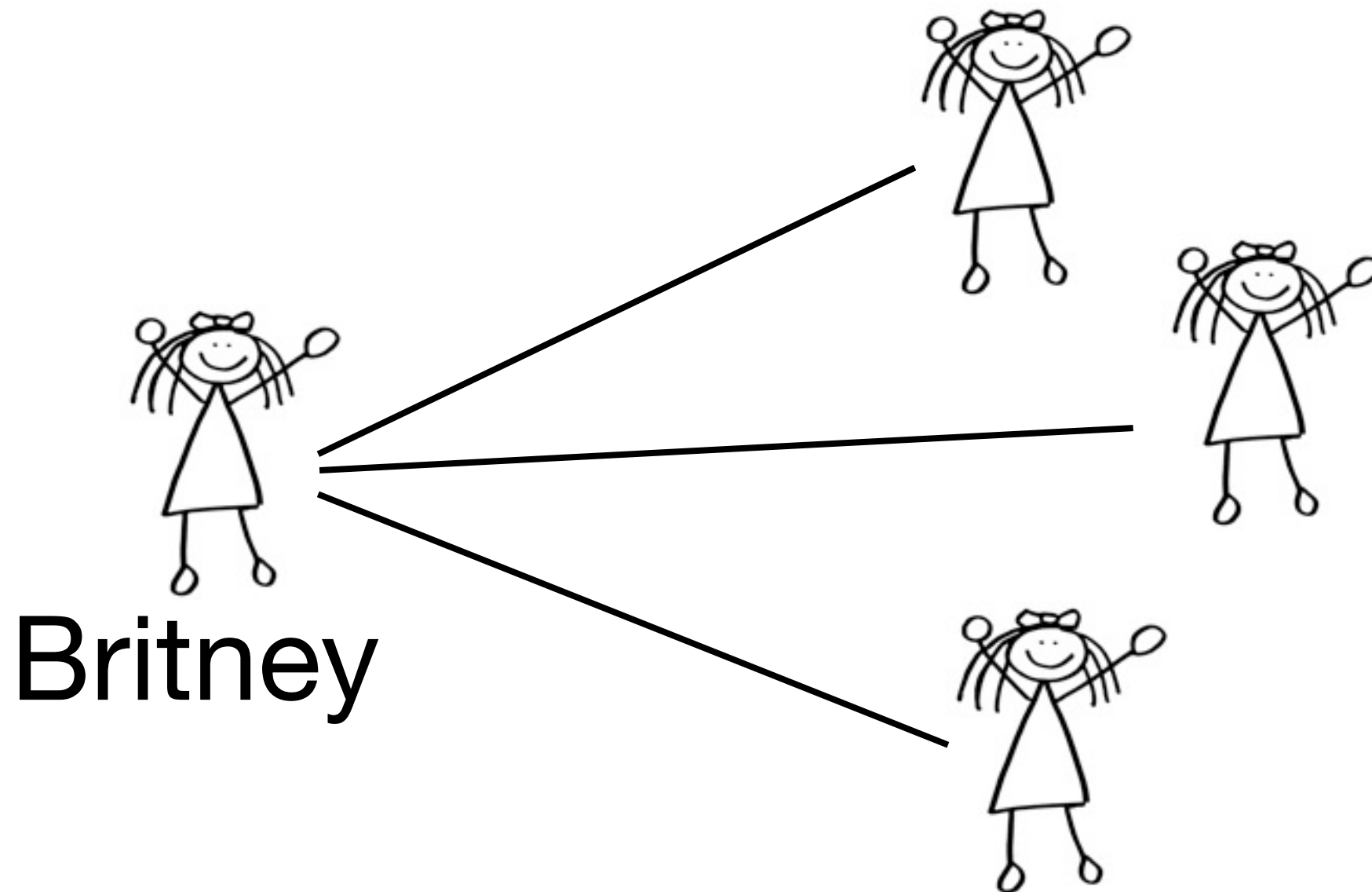
SCHOOL OF INFORMATICS  
AND COMPUTING

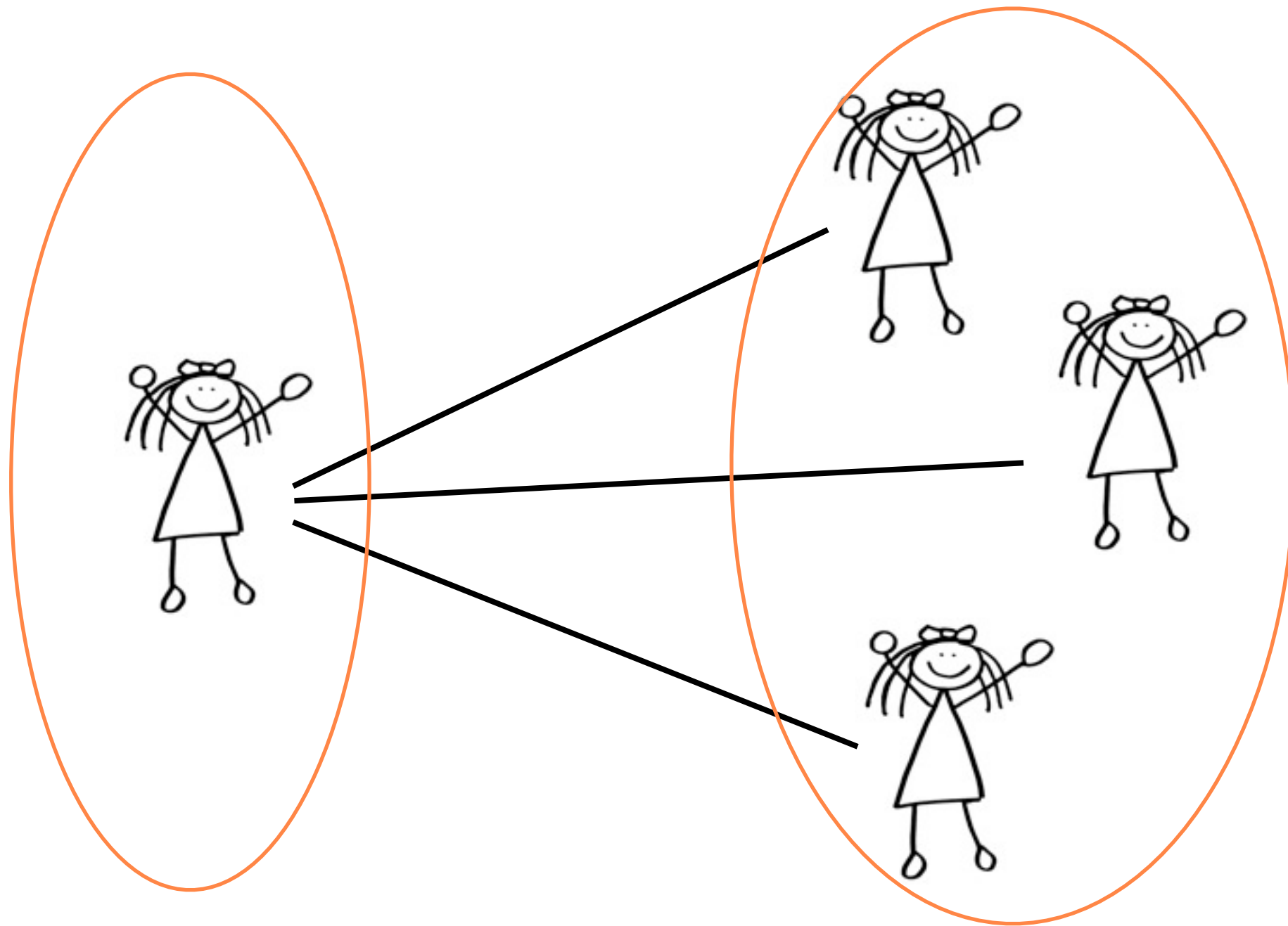
---

INDIANA UNIVERSITY

Bloomington

**Q:** randomly select someone,  
then look at her friends.





**Q:** Who will be more popular, Britney or her friends (on average, statistically speaking)?

1. Britney

2. Same

3. Her friends



# “Friendship Paradox”



# Getting to one of your friends

# Getting to one of your friends

## Following a link

# Getting to one of your friends

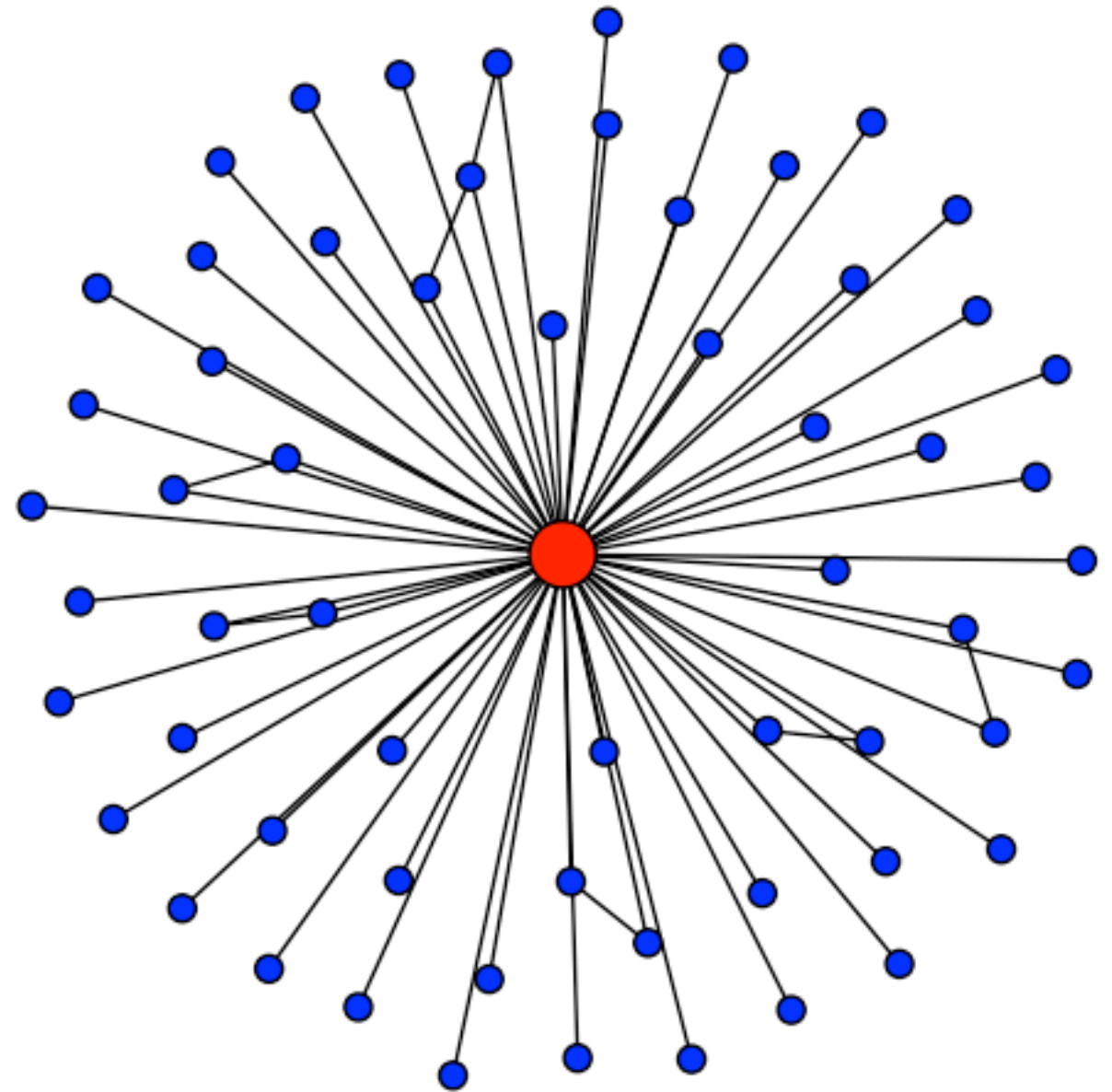
## Following a link

More likely to arrive at  
a hub because hubs  
have more links

# Getting to one of your friends

## Following a link

More likely to arrive at  
a hub because hubs  
have more links



***When*** are we living?

*“Every 2 days we create as much information as we did up to 2003.”*

Eric Schmidt, Google

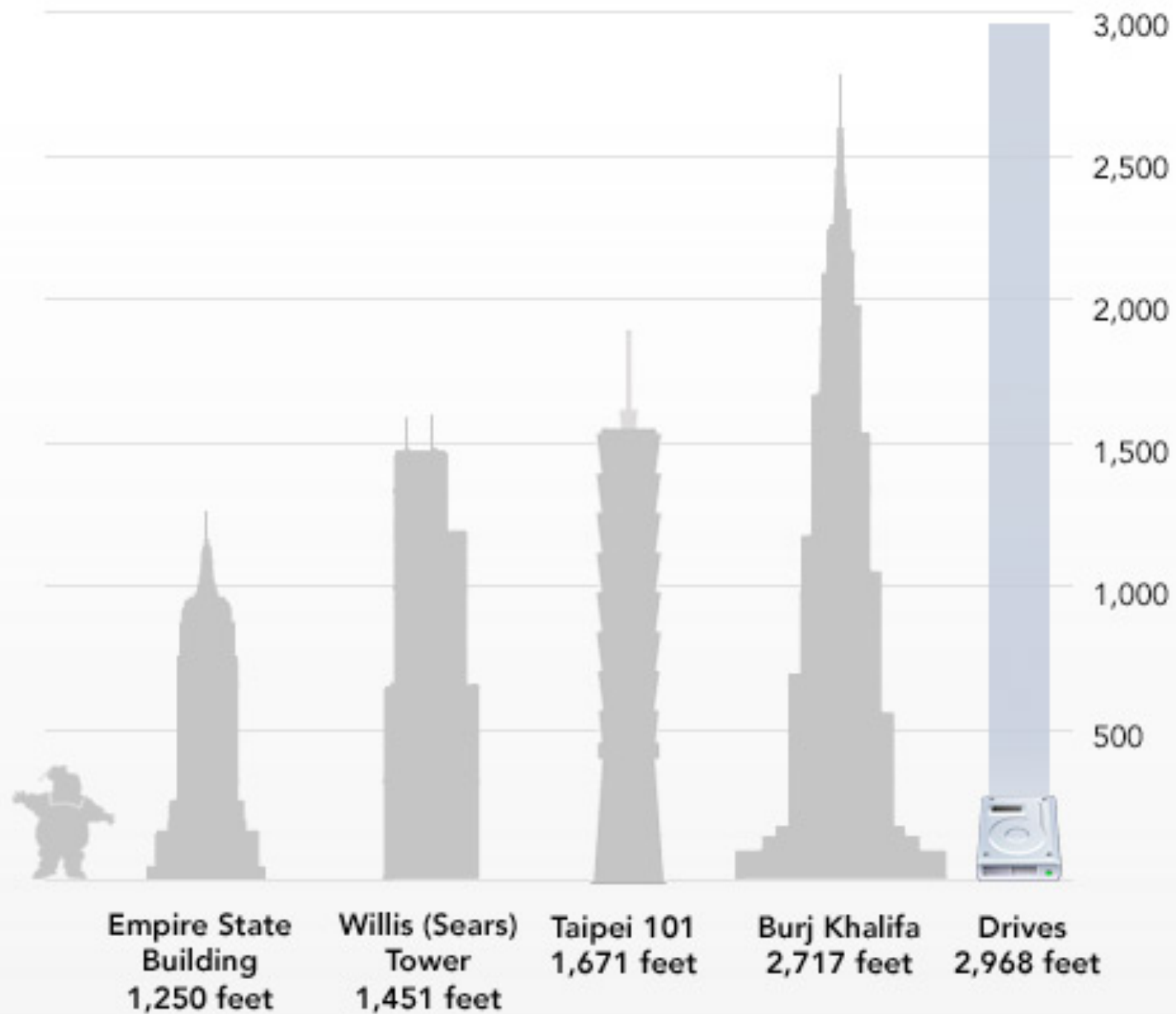
Google processes

**20+ petabytes**

per day



# 10 petabytes



\* 6,195 drives x 5.75 inches of drive height = 35,621 inches or 2,968 feet

# 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.**



What's happening?



Giving a talk @OhioState. Exciting! |



Columbus, OH

104

**Tweet**



300 million users

300 million tweets per day

**300 million people  
publishing their life.**



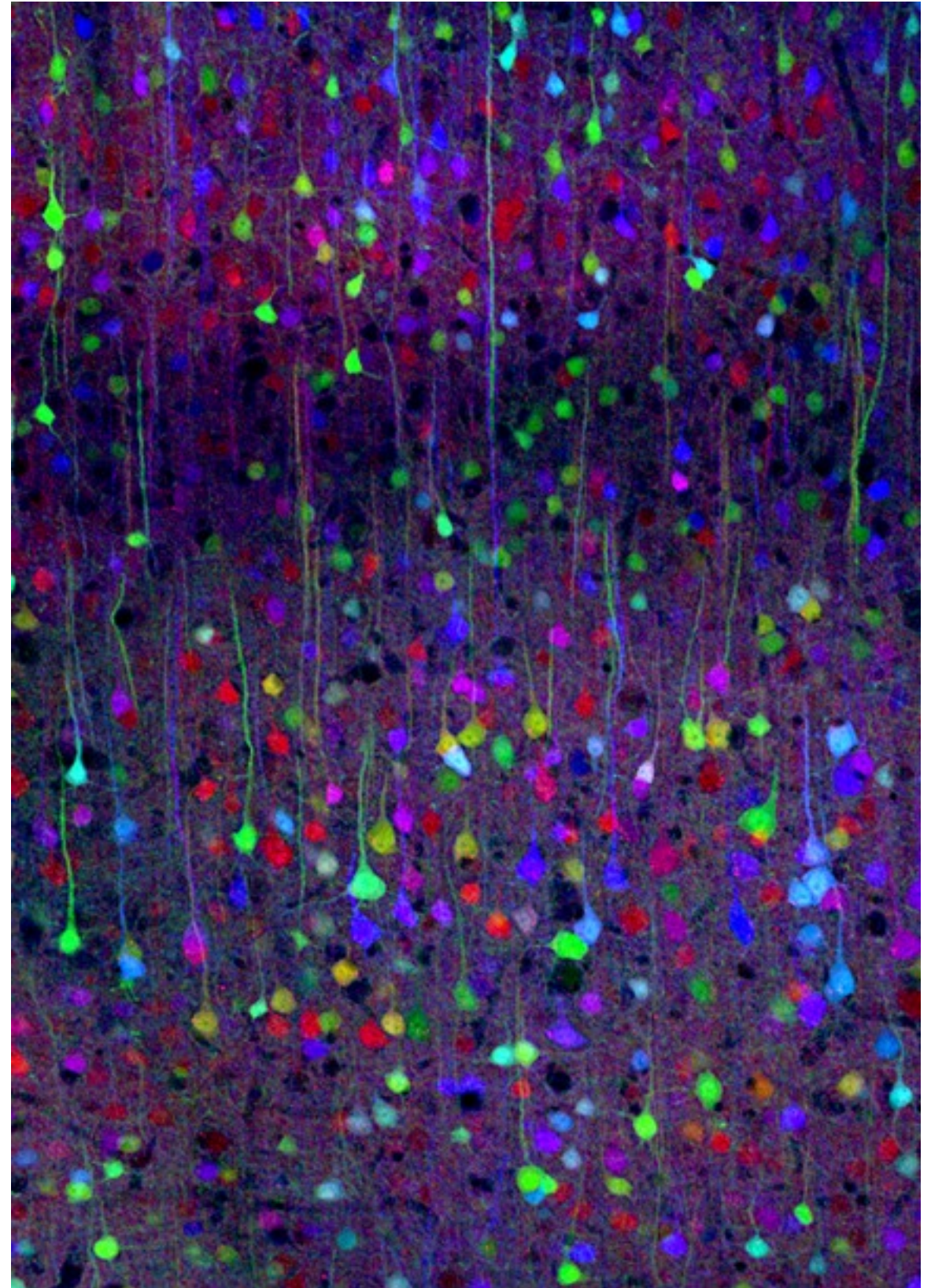
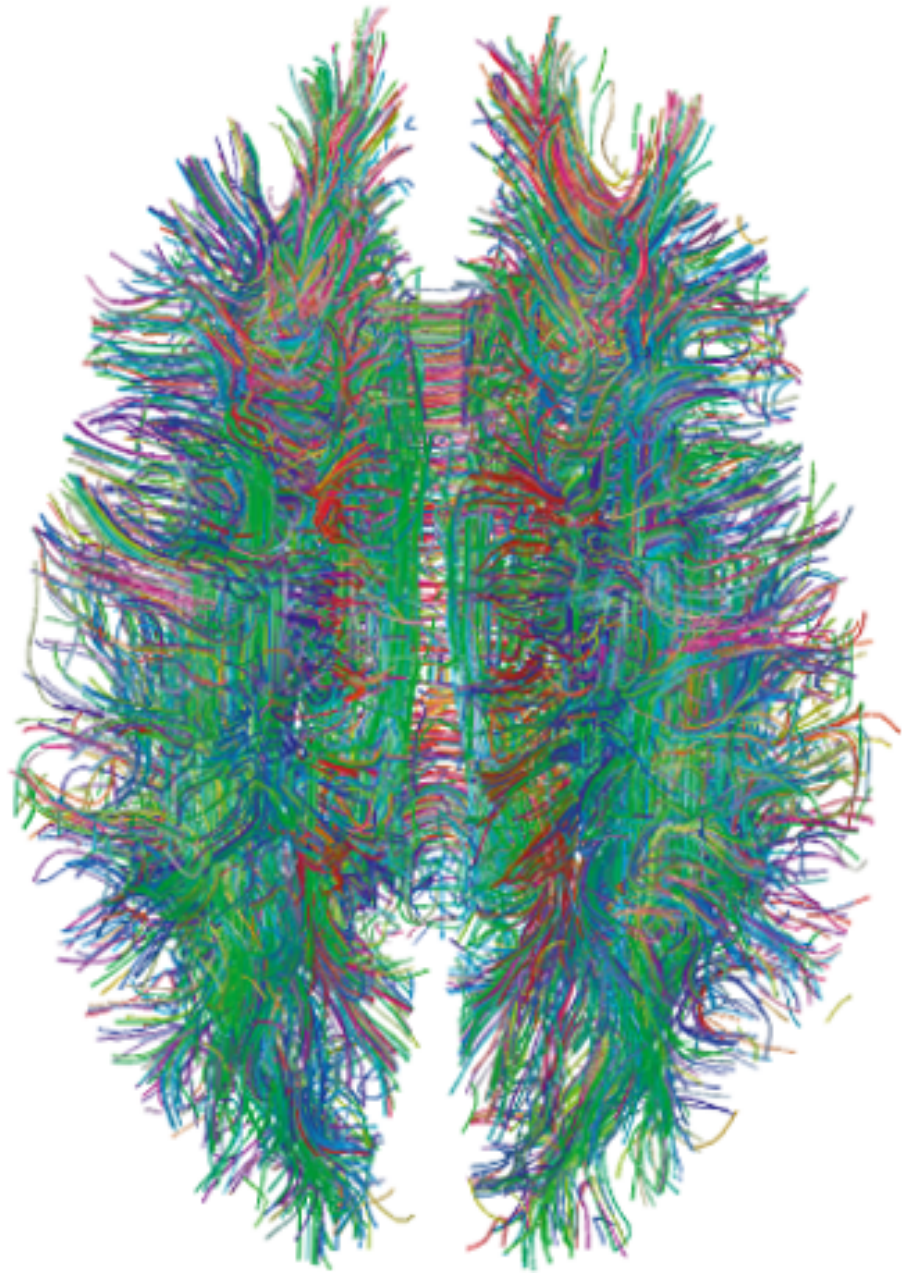
40421551 40421561 40421571 40421581 40421591 40421601 40421611 40421621 40421631 40421641 40421651 40421661 40421671 40421681 40421691  
tgagcagacctatataagatgggttatgaagattcacacagcggtcatgcctgtgatcccagcactttgggaggctgaggcaagtggagcacctgagatcatgagttcaagaccagcctggccaacatggtgaaaccccatctctac  
...A.....T.....  
tgacacagacctatataagatgggtt tgaagattcacacagtggtcatgcctgtgatcccagcac tgggaggctgagtcagtgaggcacctgagatcatgagtt ACCAGCCTGGCCAACATGGTGAAACCCCATCTCTAC  
cagacctatataagatgggtt aagatacacacagtggtcatgcctgtgatcccagcactt GGGAGGCTGAGGCAAGTGGAGCACCTGAGATCATGAGTTC cagcctggccaacatggtgaaaccccatctctac  
GACCTATATAAGATGGTTATGAAGATTCACACAGTGGCTC CCTGTGATCCCAGCACTTTGGGAGGCTGAGGCAAGTGGAG ACCTGAGATCATGAGTTCAAGACCAGCCTGGACAACATGG AACCCCATCTCTAC  
TGAACAG ATATAAGATGGTTATGAAGATTCACACAGTGGCTCATGCC tgatcccagcactttgggagg TGAGGCAAGTGGAGCACCTGAGATCATGAGTTCAAGACCA GCCAACATGGTGAAACCCCATCTCTAC  
TGAACAGAC TCAGATGGTTATGAAGATTCACACAGTGGCTCATGCCGT ATCCCAGCACTTTGGGAGGCTGAGGCAAGGGGAGCACCTG ATGAGTTCAAGACCAGCCTGGCCAACATGGTGAAACCCCA CTCTAC  
tgacacagccctata aagatgggttatgaagattcacacagtggtcatgcctgtg TCCCAGCACTTTGGGAGCCTGAGGCAAGTGGAGCACCTGA ATGAGTTCAAGACCAGCCTGGCCAACATGGTGAAACCCCA TATAC  
tgacacagacctatata gatgggttatgaagattcacacagtagctcatgcctgtgat AGCACTTTGGGAGGCTGAGGCAAGGGGAGCACGTGA GAGTTCAAGACCAGCCTGGCCAACATGGTGAAACCCCATC CTA  
gacctatataagatgggttatgaagattcacacagtggtc CTGTAATCCCATCACTTTGGGAGGCTGAGGCAAGTGGAGC CCTGAGATCATGAGTTCAAGA AGCCTGGCCAACATCGTGAAACCCCATATCTAC  
TGAACAGACCTATATAA TGGTTATGAAGATTCACACAGTGGCTCATGCCGTGATCC cactttgggatgctgaggcaagtggagcacctgagatcat CAAGACCAGCCTGGCCAACATGGTGAAACCCCATCTCTAC  
TG ACCTATATAAGATGGTTATGAAGATTCACACAGTGGCTCA TGTGATCCCAGCACTTTGGGAGGCTGAGGCAAGTGGAGCA CTGAGATCACGAGTTCAAGACCAGCCTGCCAACATGGTC AACCCCATCTCTAC  
TGAACAGACCTATATAAGA GGTACGAAGATTCACACAGTGGCTCATGCCGTGATCCC cacattgggaggctgaggcaagtggagcacctgagatcat AAGACCAGCCTGGCCAACATGGTGAAACCCCATCTCTAC  
tgacacagacctatataagat ttatgaagattcacacagtggtcatgcctgtgatcccag CTTTGGGAGGCTGAGGCAAGTGGAGCACCTGAGATCATGA agcctggccaacatggtgaaaccccatctctac  
tgacacagacctatataagatgggtt aagattcacacagtggtcatgccagtgatcccagcactt GGGAGGCTGAGGCAAGTGGAGCACCTGAGATAATGAGTTC GCCTGGCCAACATGGTGAAA CCCATCTCTAC  
TGAACAGACCTATATAAGATGGTTA agattcacacagaggctcatgcctgtgatcccagcacttt AGGCTGAGGCAAGTGGAGCACCTGAGATCATGAGTTCAAG CCTGGCCAACATGGTGAAACCCCATCTCTAC  
tgacacagacctatataagatgggtt TTCACACAGTGGCTCATGCCGTGATCCCAGCACCTTGGG GCTGAGGCAAGTGGAGCACCTGAGATCATGAGTTCAAGAC CCAACATGGTGAAACCCCATCTCTAC  
TGAACAGACCTATATAAGATGGTTAT CAGTGGCTCATGCCGTGAT ACTTTGGGAGGCTGAGGCAAGTGGAGCACCTGAGATCATG CAACATGGTGAAACCCCATCTCTAC  
TGAACAGACCTATATAAGATGGTTATGAAG CAGTGGCTCATGCCGTGATC ACTTCGGGAGGCTGAGGCAAGTGGAGCACCTGAGATCATG AACATGGTGAAACCCCATCTCTAC  
TGAACAGACCTATATAAGATGGTTATGAAGAT CAGTGGCTCATGCCGTGATCC CCTCTGGGAGGCTGAGGCAAGTGGAGCACCTGAGATCATG ACATGGTGAAACCCCATCTCTAC  
TGAACAGACCTATCTAAGATGGTTATGAAGATT GCGGCTCATGCCGTGATC CTTTGGGAGGCTGAGGCAAGTGGAGCACCTGAGATCATGA ACATGGTGAAACCCCATCTATA  
TGAACAGACCTATATAAGATGGTTATGAAGATTC CTCTTGCTGTGATCCCAGCACTTTGGGAGGCTGACGCAA TGGAGCACCTGAGATCATGAGTTCAAGACCAGCCTGGCCA TGGTGAAACCCCATCTCTAC  
TGAACAGACCTATATAAGATGGTTATGAAGATTC CTCATGCCGTGTGATCCCAGCACTTTGGGAGGCTGAGGCAA TGGAGCACCTGAGATCATGAGTTCAAGACCAGCCTGGCCA GGTGAAACCCCATCGCTAC  
TGAACAGACCTATATAAGATGGTTATGAAGATTCA GTGATCCCAGCACTTTGGGAGGCTGAGGCAAGTGGAGCAC GATCATGAGTTCAAGACCCGCTGGCCAACATGGTGAAAC ccatctctac  
AGATGGTTATGAAGATTCACACAGTGGCTCATGCCGTGTA CCAGCACTTTGGGAGGCTGAGGCAAGTGGAGTACCTGAGA GAGTTCAAGACCAGCCTGGCCAACATGGTGAAACCCCATC TAC  
ACATGGTTATGAAGATTCACACAGTGGCTCATGCCGTGTA CTTTGGGAGGCTGAGGCAAGTGGAGCACCTGAGATCATGA CATGGTGTAACCCCATCTCTAC  
GGTTATGAAGATTCACACAGTGGCTCATGCCGTGATCCC CTCTGGGAGGCTGAGGCAAGTG agcactgagatcatgagttcaagaccagcctg\*caacat tgaaccccatctctac  
TATGAAGATTCACACAGTGGCTCA gatcccagcactttgggaggctgaggcaagtggagcacct agttcaagaccagcctggccaacatggtgaaaccccatct TAC  
ATGAAGATTCACACAGTGGCTCATGCCGTGATCCCAGCA TCTGGGAGGCTGAGGCAAGTGGAGCACCTGAGATCATGAG CATGGTGAAACCCCATCTCTAC  
gatcccagctattttgggaggctgaggcaagtggagcacct CATGGTGAAACCCCATCTCT  
atcccagcactttgggaggctgaggcaagtggagcacctg CATGGTGAAACCCCATCTCTAC  
TCTGAGAGGCTGAGGCAAGTGGAGCACCTGAGATCATGAG GTGAAACCCCATCTCTAC  
GGGATGCTTAGTCAATTGTAGCACCTGAGATCATGAGTTC GTGAAACCCCATCTCTAC  
aggctgaggcaagtggagcacctgagatcatgagttcaag gtgaaaccccatctctac  
tggggcaagtggagcacctgagatcatgagttcaagacca gtgaaaccgtgtctctac  
tgaggcaagtggagcacctgagatcatgagttcaagacca GAAATCCCATCTCTAC  
GAGGCAAGTGGAGCACCTGAGATCATGAGTTCAAGACCAG GAAACCCCATCTCTAC  
AGGCAAGTGGAGCACCTGAGATCATGAGTTCAAGACCAG GAAACCCCATCTCTAC  
aggcaatttgagtcctgagatcatgagttcaagaccagc gaaaccccatctctg  
GCAAGTGGAGCACCTGAGATCA AACCCCATCTCTAC  
CAAGTGGAGCACCTGAGATCATGAGTTCAAGACCAGCCTG AATCCCATCTCTAC  
caagtggagcacctgagatcatgagttcaagaccagcctg aaccccatctctac  
AAGTGGAGCACCTGAGATCATGAGTTCAAGACCAGCCTGG AACCCCATCTCTAC  
AGTGGAGCACCTGAGATCATGAGTTCAAGACCAGCCTGGC ACCCCGTTTCTAC  
AGTGCAGCACCTGAGATCATGAGTTCAAGACCAGCCTGGC accccatctctac  
GTGGAGCACCTGAGATCATGAGTTCAAGACCAGCATGGCC CCCCATCTCTAC  
GGAGCACCTGAGATCATGAGTTCAAGACCAGCCTGGCCAA CATCTCTAC  
ggagcacctgagatgatgagttcaagaccaggggtggccaa CATCTCTAC  
ggagcacctgagatcatgagttcaagaccagcctggccaa CGTCTCTAC  
GAGCACCTGAGATCATGAGTTCAAGACCAGCCTGGCCAA CATCTCTAC



40421551 40421561 40421571 40421581 40421591 40421601 40421611 40421621 40421631 40421641 40421651 40421661 40421671 40421681 40421691  
tgagcagacctatataagatgggttatgaagattcacacagcggtcatgcctgtgatcccagcactttgggaggctgaggcaagtggagcacctgagatcatgagttcaagaccagcctggccaacatggtgaaaccccatctctac  
...A.....T.....  
tgaacagacctatataagatgggtt tgaagattcacacagtggtcatgcctgtgatcccagcac tgggaggctgagtcagtgaggcacctgagatcatgagtt ACCAGCCTGGCCAACATGGTGAAACCCCATCTCTAC  
cagacctatataagatgggtt aagatacacacagtggtcatgcctgtgatcccagcactt GGGAGGCTGAGGCAAGTGGAGCACCTGAGATCATGAGTTC cagcctggccaacatggtgaaaccccatctctac  
GACCTATATAAGATGGTTATGAAGATTCACACAGTGGCTC CCTGTGATCCCAGCACTTTGGGAGGCTGAGGCAAGTGGAG ACCTGAGATCATGAGTTCAAGACCAGCCTGGACAACATGG AACCCCATCTCTAC  
TGAACAG ATATAAGATGGTTATGAAGATTCACACAGTGGCTCATGCC tgatcccagcactttgggagg TGAGGCAAGTGGAGCACCTGAGATCATGAGTTCAAGACCA GCCAACATGGTGAAACCCCATCTCTAC  
TGAACAGAC TCAGATGGTTATGAAGATTCACACAGTGGCTCATGCCTGT ATCCCAGCACTTTGGGAGGCTGAGGCAAGGGGAGCACCTG ATGAGTTCAAGACCAGCCTGGCCAACATGGTGAAACCCCA CTCTAC  
tgaacagccctata aagatgggttatgaagattcacacagtggtcatgcctgtg TCCCAGCACTTTGGGAGCCTGAGGCAAGTGGAGCACCTGA ATGAGTTCAAGACCAGCCTGGCCAACATGGTGAAACCCCA TATAC  
tgaacagacctatata gatgggttatgaagattcacacagtagctcatgcctgtgat AGCACTTTGGGAGGCTGAGGCAAGGGGAGCACGTGA GAGTTCAAGACCAGCCTGGCCAACATGGTGAAACCCCATC CTA  
gacctatataagatgggttatgaagattcacacagtggtc CTGTAATCCCATCACTTTGGGAGGCTGAGGCAAGTGGAGC CCTGAGATCATGAGTTCAAGA AGCCTGGCCAACATCGTGAAACCCCATATCTAC  
TGAACAGACCTATATAA TGGTTATGAAGATTCACACAGTGGCTCATGCCTGTGATCC cactttgggatgctgaggcaagtggagcacctgagatcat CAAGACCAGCCTGGCCAACATGGTGAAACCCCATCTCTAC  
TG ACCTATATAAGATGGTTATGAAGATTCACACAGTGGCTCA TGTGATCCCAGCACTTTGGGAGGCTGAGGCAAGTGGAGCA CTGAGATCACGAGTTCAAGACCAGCCTGCCAACATGGTC AACCCCATCTCTAC  
TGAACAGACCTATATAAGA GGTTACGAAGATTCACACAGTGGCTCATGCCTGTGATCCC cacattgggaggctgaggcaagtggagcacctgagatcat AAGACCAGCCTGGCCAACATGGTGAAACCCCATCTCTAC  
tgaacagacctatataagat ttatgaagattcacacagtggtcatgcctgtgatcccag CTTTGGGAGGCTGAGGCAAGTGGAGCACCTGAGATCATGA agcctggccaacatggtgaaaccccatctctac  
tgaacagacctatataagatgggtt aagattcacacagtggtcatgccagtgatcccagcactt GGGAGGCTGAGGCAAGTGGAGCACCTGAGATAATGAGTTC GCCTGGCCAACATGGTGAAA CCCATCTCTAC  
TGAACAGACCTATATAAGATGGTTA agattcacacagggctcatgcctgtgatcccagcacttt AGGCTGAGGCAAGTGGAGCACCTGAGATCATGAGTTCAAG CCTGGCCAACATGGTGAAACCCCATCTCTAC  
tgaacagacctatataagatgggtta TTCACACAGTGGCTCATGCCTGTGATCCCAGCACCTTGGG GCTGAGGCAAGTGGAGCACCTGAGATCATGAGTTCAAGAC CCAACATGGTGAAACCCCATCTCTAC  
TGAACAGACCTATATAAGATGGTTAT CAGTGGCTCATGCCTGTGAT ACTTTGGGAGGCTGAGGCAAGTGGAGCACCTGAGATCATG CAACATGGTGAAACCCCATCTCTAC  
TGAACAGACCTATATAAGATGGTTATGAAG CAGTGGCTCATGCCTGTGATC ACTTCGGGAGGCTGAGGCAAGTGGAGCACCTGAGATCATG AACATGGTGAAACCCCATCTCTAC  
TGAACAGACCTATATAAGATGGTTATGAAGAT CAGTGGCTCATGCCTGTGATCC CCTCTGGGAGGCTGAGGCAAGTGGAGCACCTGAGATCATG ACATGGTGAAACCCCATCTCTAC  
TGAACAGACCTATCTAAGATGGTTATGAAGATT GCGGCTCATGCTGTTATC CTTTGGGAGGCTGAGGCAAGTGGAGCACCTGAGATCATGA ACATGGTGAAACCCCATCTATA  
TGAACAGACCTATATAAGATGGTTATGAAGATTC CTCCTGATCCCACACTTTGGGAGGCAAGTGGAGCACCTGAGATCATGAGTTCAAGACCAGCCTGGCCA TGGTGAAACCCCATCTCTAC  
TGAACAGACCTATATAAGATGGTTATGAAGATTC CTCCTGATCCCACACTTTGGGAGGCAAGTGGAGCACCTGAGATCATGAGTTCAAGACCAGCCTGGCCA GGTGAAACCCCATCGCTAC  
TGAACAGACCTATATAAGATGGTTATGAAGATTCA GTGATCCCACACTTTGGGAGGCTGAGGCAAGTGGAGCACCTGAGATCATGAGTTCAAGACCCGCTGGCCAACATGGTGAAAC ccatctctac  
AGATGGTTATGAAGATTCACACAGTGGCTCA GTGA CCAACTTTGGGAGGCTGAGGCAAGTGGAGCACCTGAGATCATGAGTTCAAGACCAGCCTGGCCAACATGGTGAAACCCCATC TAC  
ACATGGTTATGAAGATTCACACAGTGGCTCATG CTCGA CTTTGGGAGGCTGAGGCAAGTGGAGCACCTGAGATCATGAGTTCAAGACCAGCCTGGCCAACATGGTGAAACCCCATCTCTAC  
GGTTATGAAGATTCACACAGTGGCTCATG CTCGAATCCC CTCCTGAGGCTGAGGCAAGTGGAGCACCTGAGATCATGAGTTCAAGACCAGCCTGGCCAACATGGTGAAACCCCATCTCTAC  
TATGAAGATTCACACAGTGGCTCATG CTCGAATCCC CTCCTGAGGCTGAGGCAAGTGGAGCACCTGAGATCATGAGTTCAAGACCAGCCTGGCCAACATGGTGAAACCCCATCTCTAC  
ATGAAGATTCACACAGTGGCTCATG CTCGTGATCCCAGCA TCTGGGAGGCTGAGGCAAGTGGAGCACCTGAGATCATGAG CATGGTGAAACCCCATCTCTAC  
gatcccagctattttgggaggctgaggcaagtggagcacct CATGGTGAAACCCCATCTCTAC  
atcccagcactttgggaggctgaggcaagtggagcacctg CATGGTGAAACCCCATCTCTAC  
TCTGAGAGGCTGAGGCAAGTGGAGCACCTGAGATCATGAG GTGAAACCCCATCTCTAC  
GGGATGCTTAGTCAATTGTAGCACCTGAGATCATGAGTTC GTGAAACCCCATCTCTAC  
aggctgaggcaagtggagcacctgagatcatgagttcaag gtgaaaccccatctctac  
tggggcaagtggagcacctgagatcatgagttcaagacca gtgaaaccgtgtctctac  
tgaggcaagtggagcacctgagatcatgagttcaagacca GAAATCCCATCTCTAC  
GAGGCAAGTGGAGCACCTGAGATCATGAGTTCAAGACCAG GAAACCCCATCTCTAC  
AGGCAAGTGGAGCACCTGAGATCATGAGTTCAAGACCAG GAAACCCCATCTCTAC  
aggcaatttgagtcctgagatcatgagttcaagaccagc gaaaccccatctctac  
GCAAGTGGAGCACCTGAGATCA AACCCCATCTCTAC  
CAAGTGGAGCACCTGAGATCATGAGTTCAAGACCAGCCTG AATCCCATCTCTAC  
caagtggagcacctgagatcatgagttcaagaccagcctg aaccccatctctac  
AAGTGGAGCACCTGAGATCATGAGTTCAAGACCAGCCTGG AACCCCATCTCTAC  
AGTGGAGCACCTGAGATCATGAGTTCAAGACCAGCCTGGC ACCCCGTTTCTAC  
AGTGCAGCACCTGAGATCATGAGTTCAAGACCAGCCTGGC accccatctctac  
GTGGAGCACCTGAGATCATGAGTTCAAGACCAGCATGGCC CCCCATCTCTAC  
GGAGCACCTGAGATCATGAGTTCAAGACCAGCCTGGCCAA CATCTCTAC  
ggagcacctgagatgatgagttcaagaccagggaggccaa CATCTCTAC  
ggagcacctgagatcatgagttcaagaccagcctggccaa CGTCTCTAC  
GAGCACCTGAGATCATGAGTTCAAGACCAGCCTGGCCAA CATCTCTAC

\$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

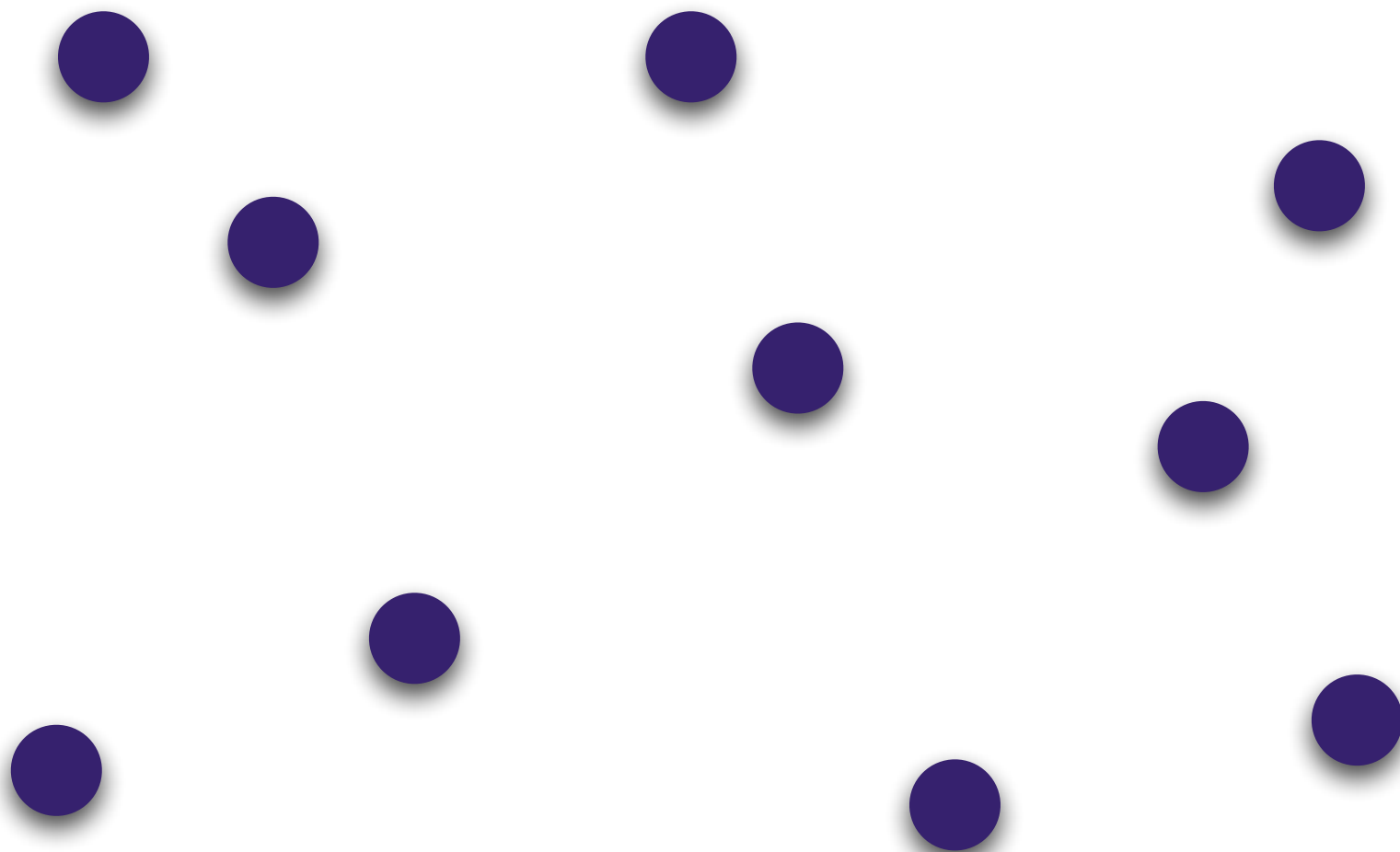
**MANY** parts,

**INTERACTING** with each other

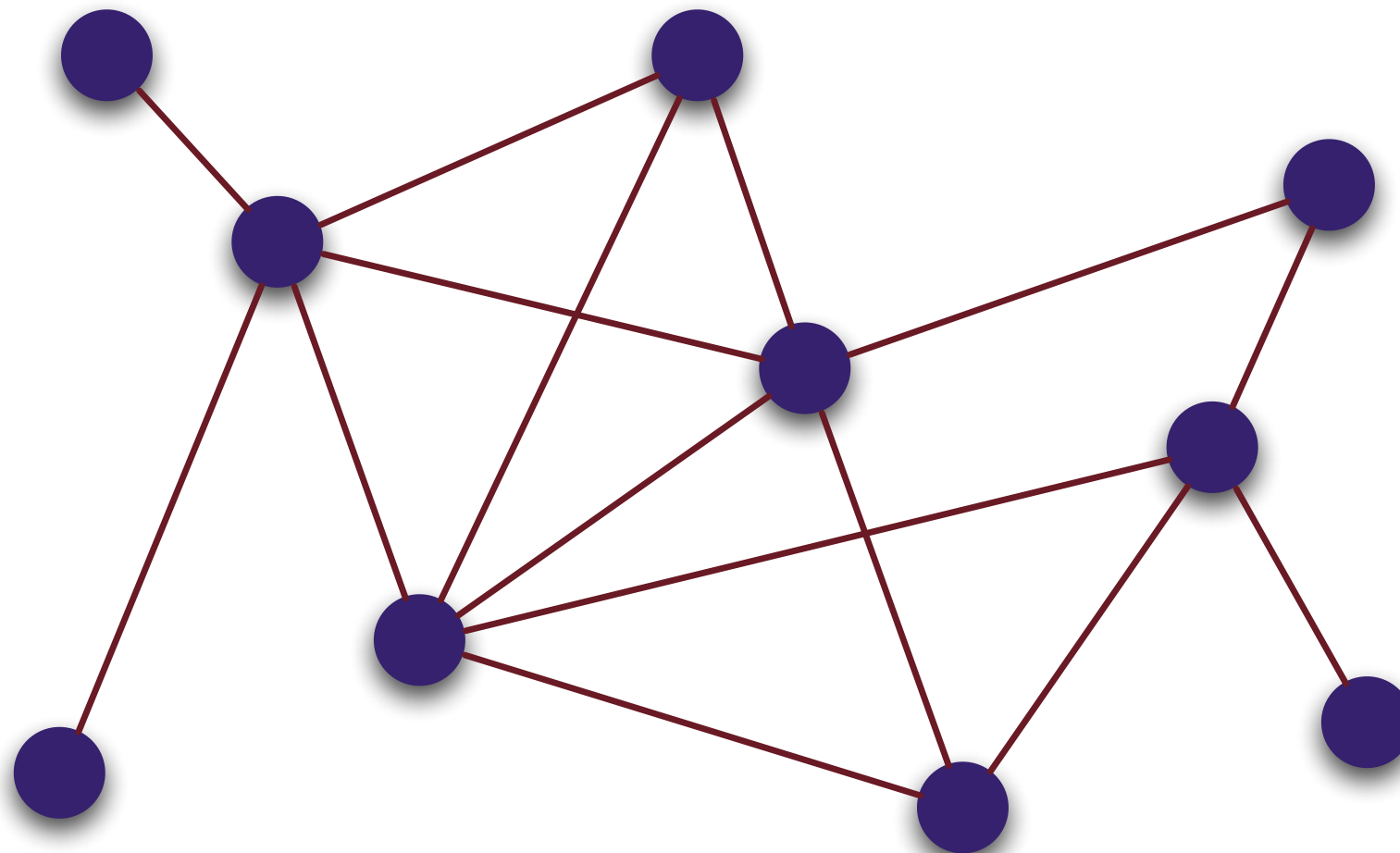
in **NON-TRIVIAL WAYS**

# NETWORKS



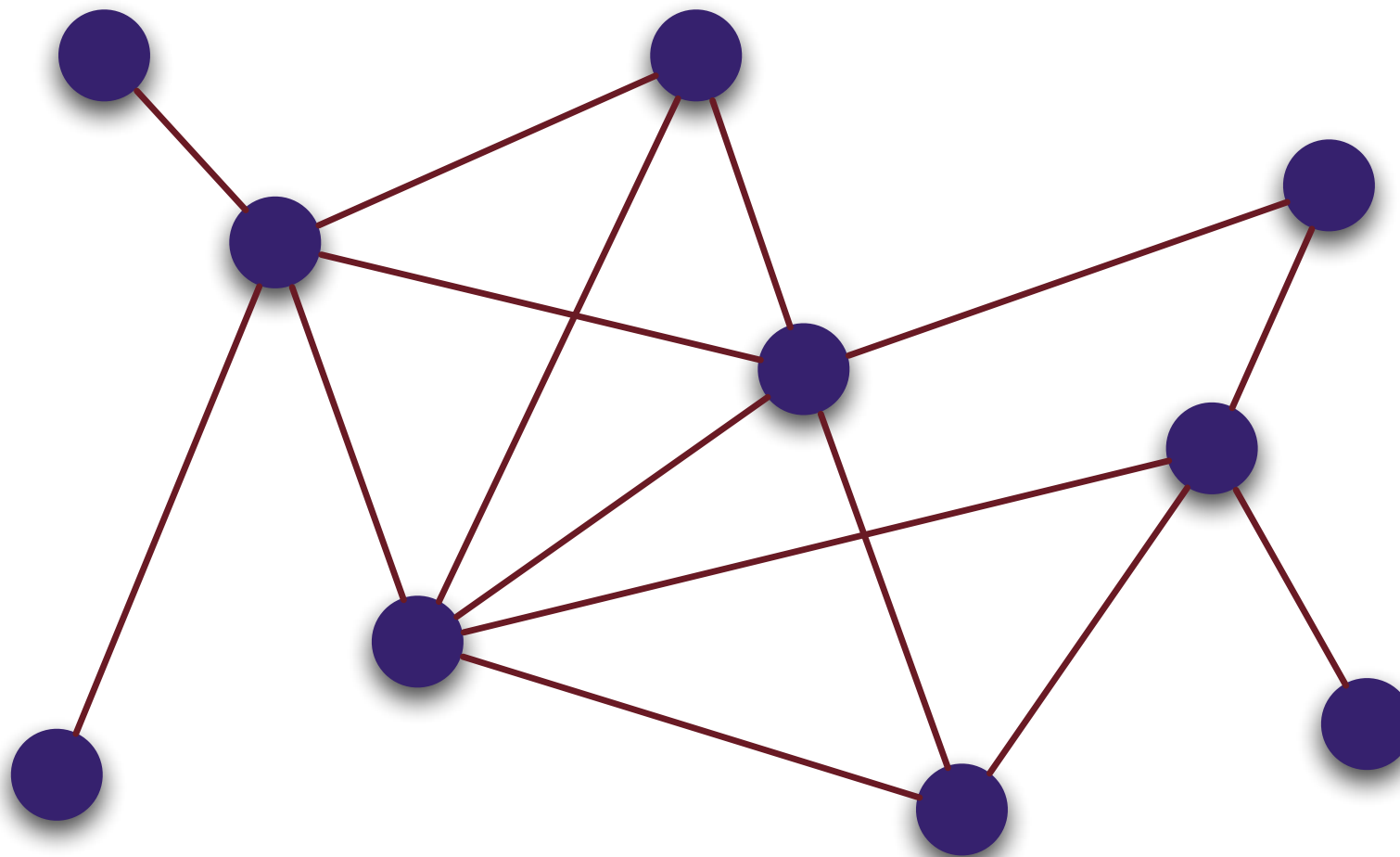


Nodes



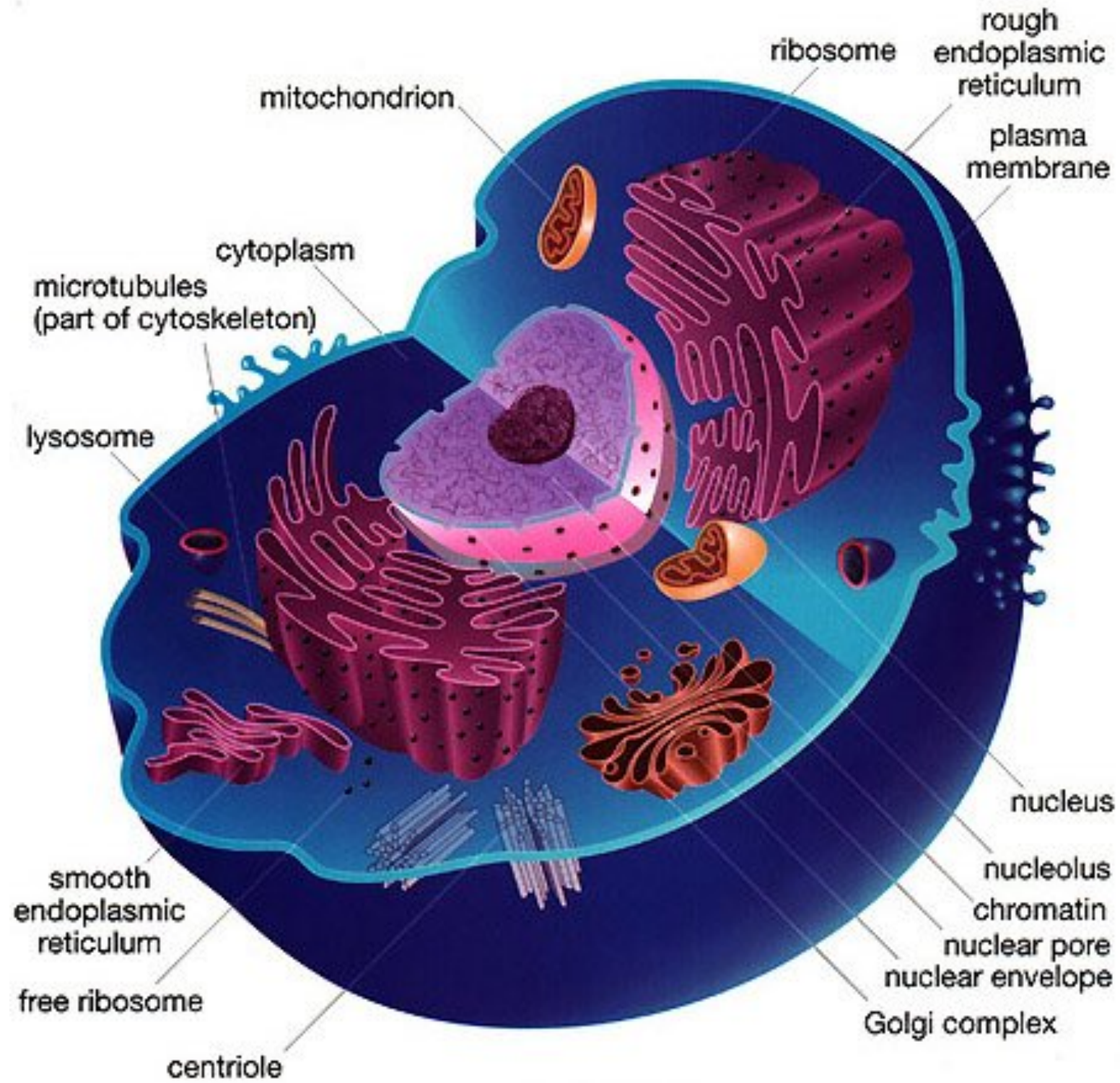
Links (edges) between nodes

Degree: # of neighbors

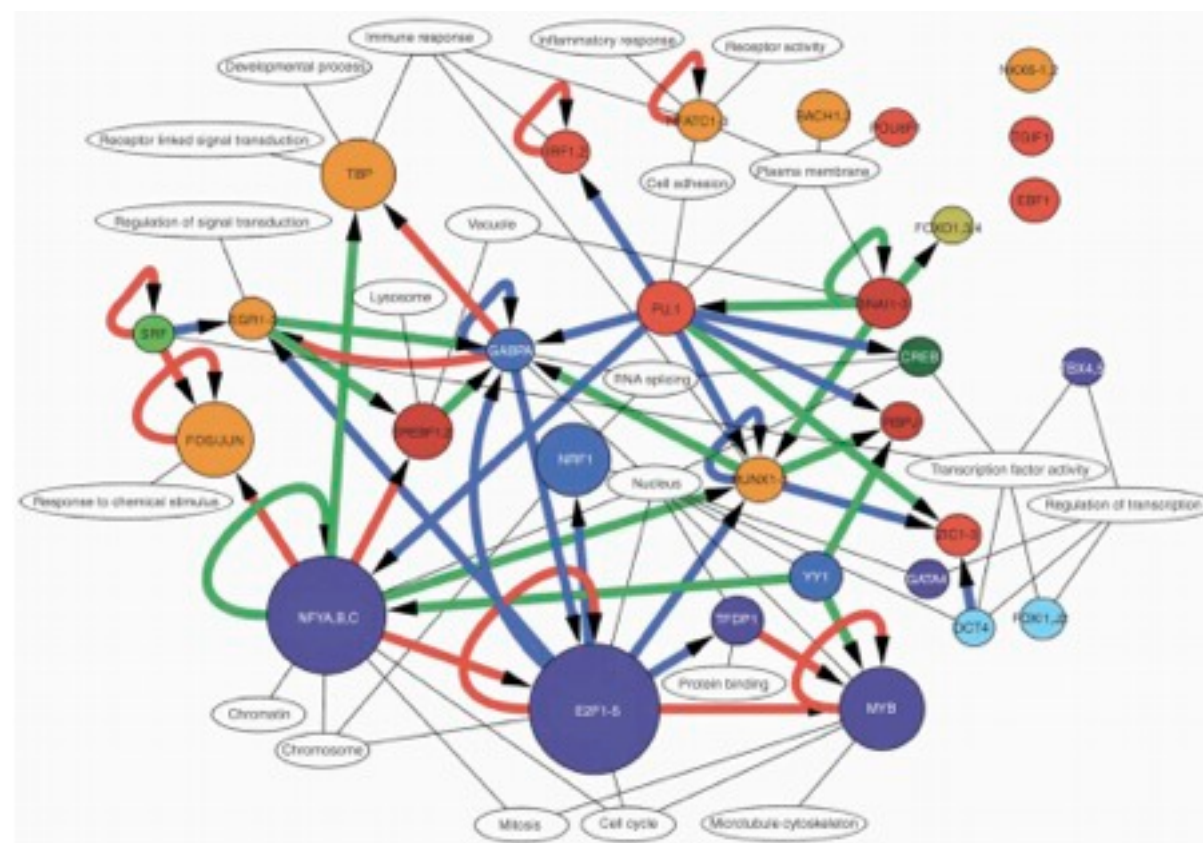


Links (edges) between nodes

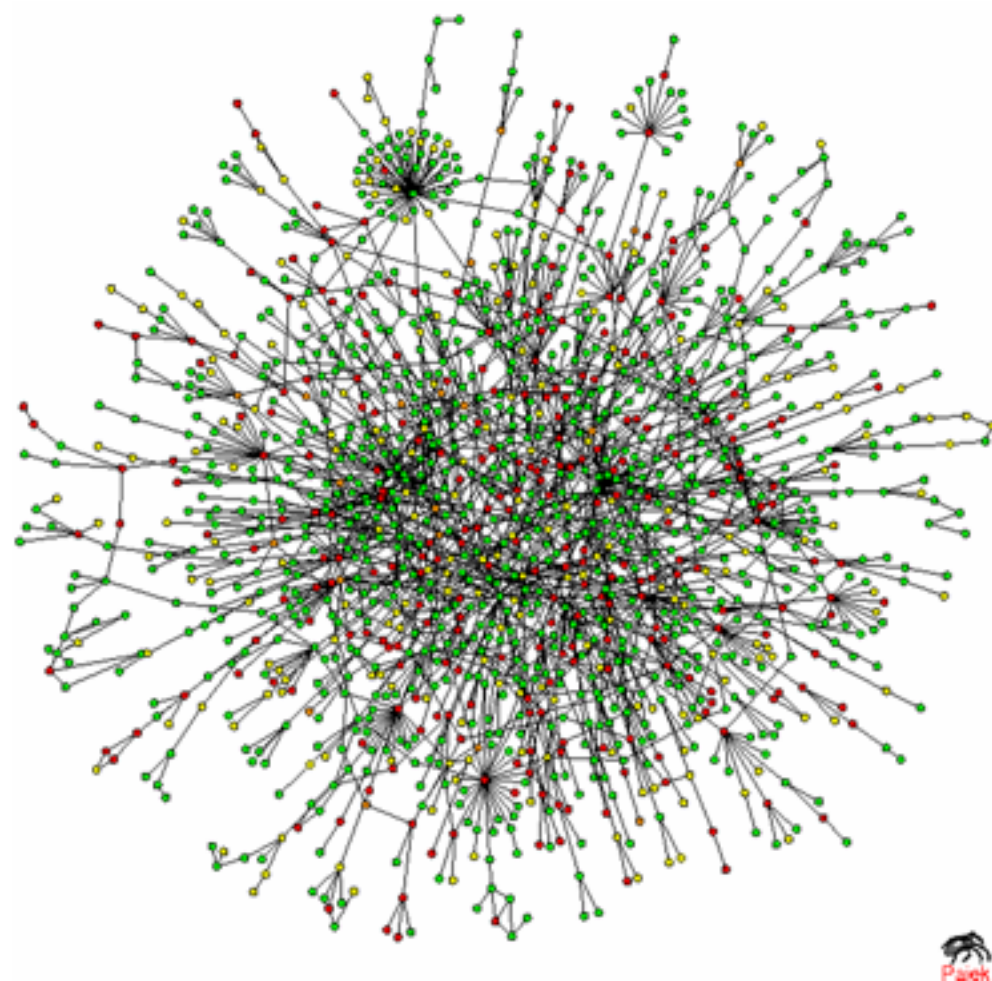
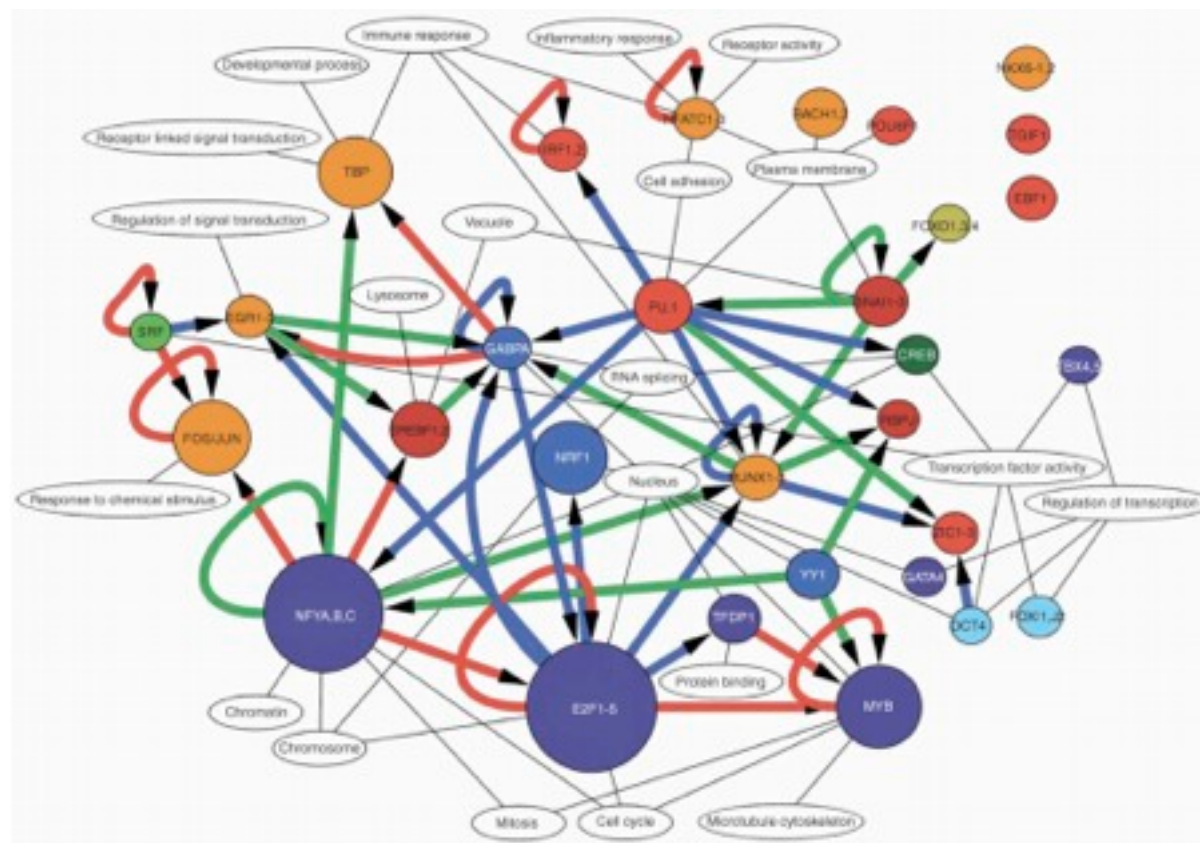




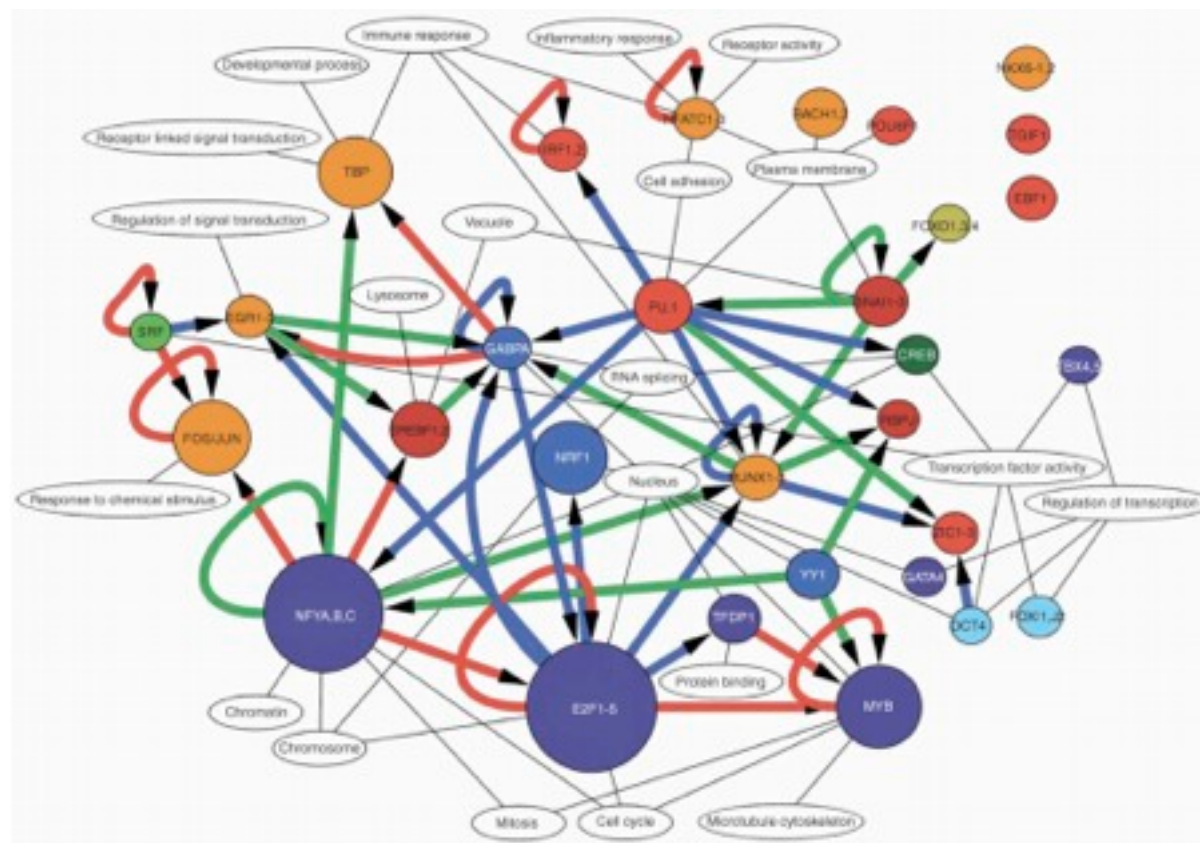




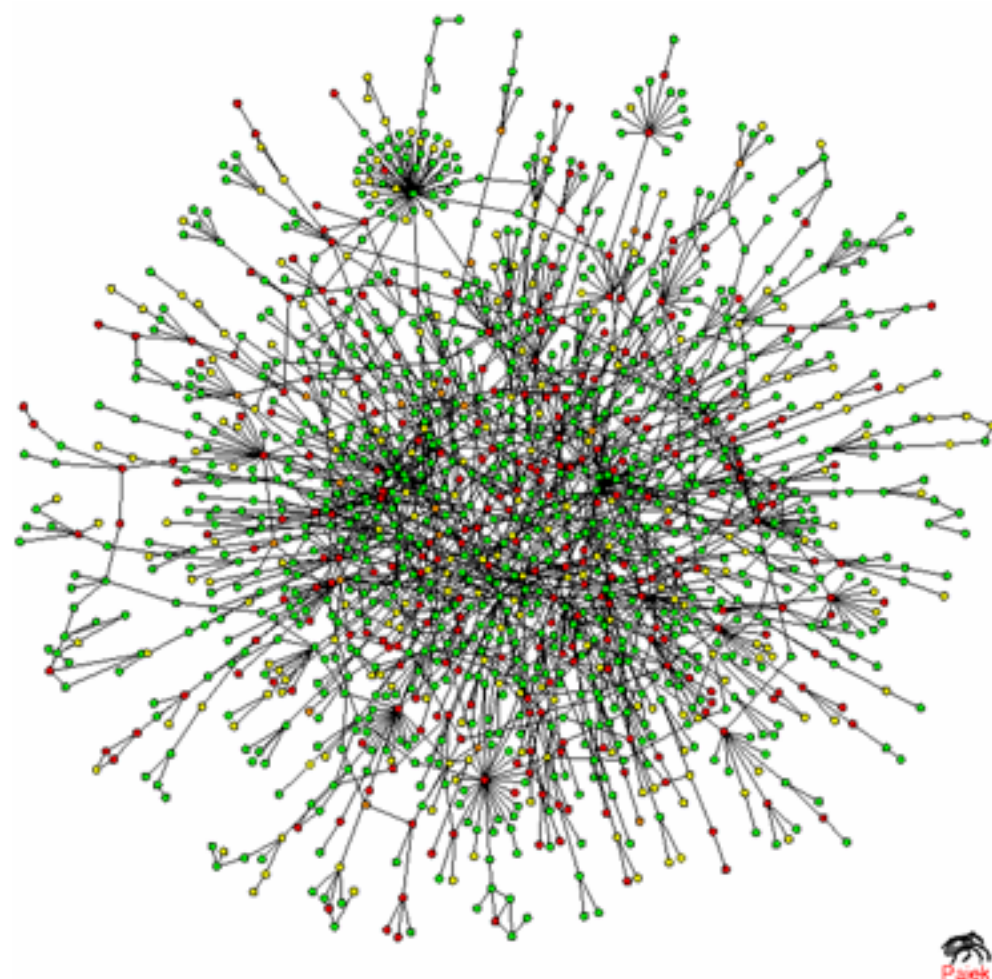




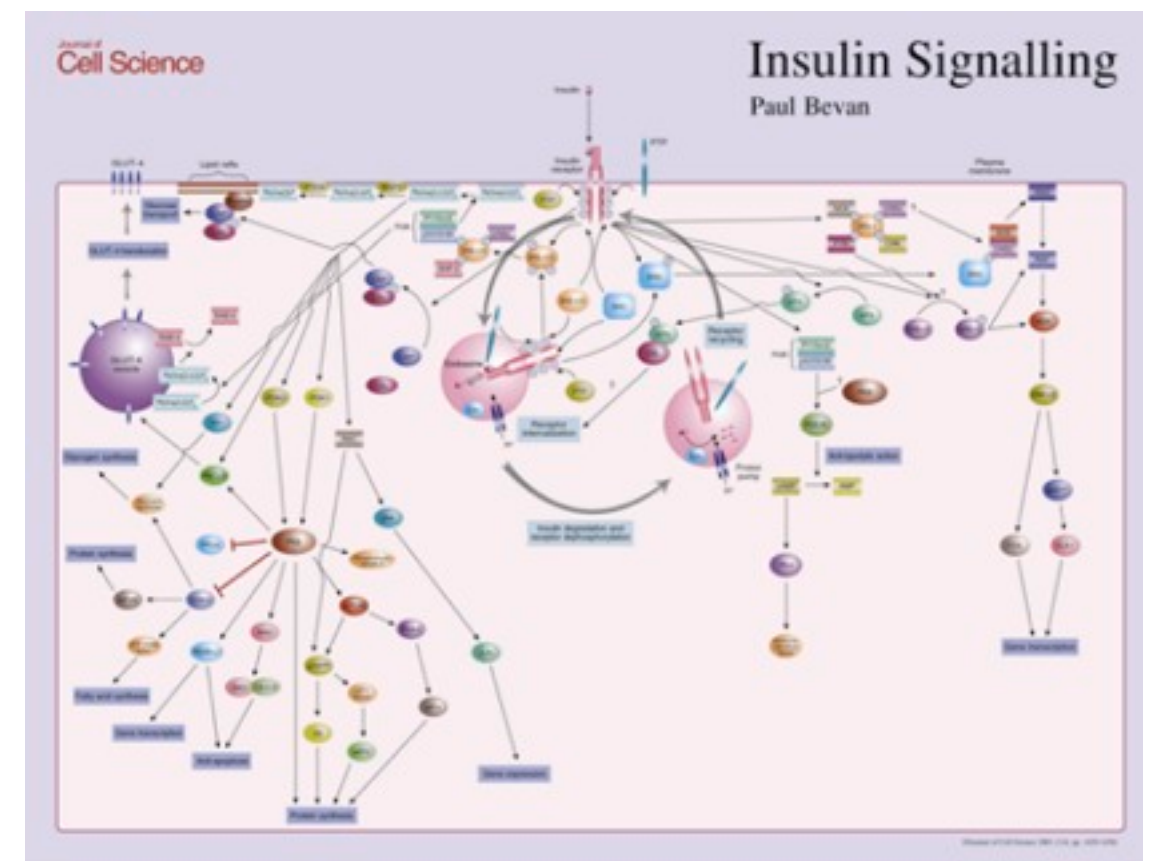
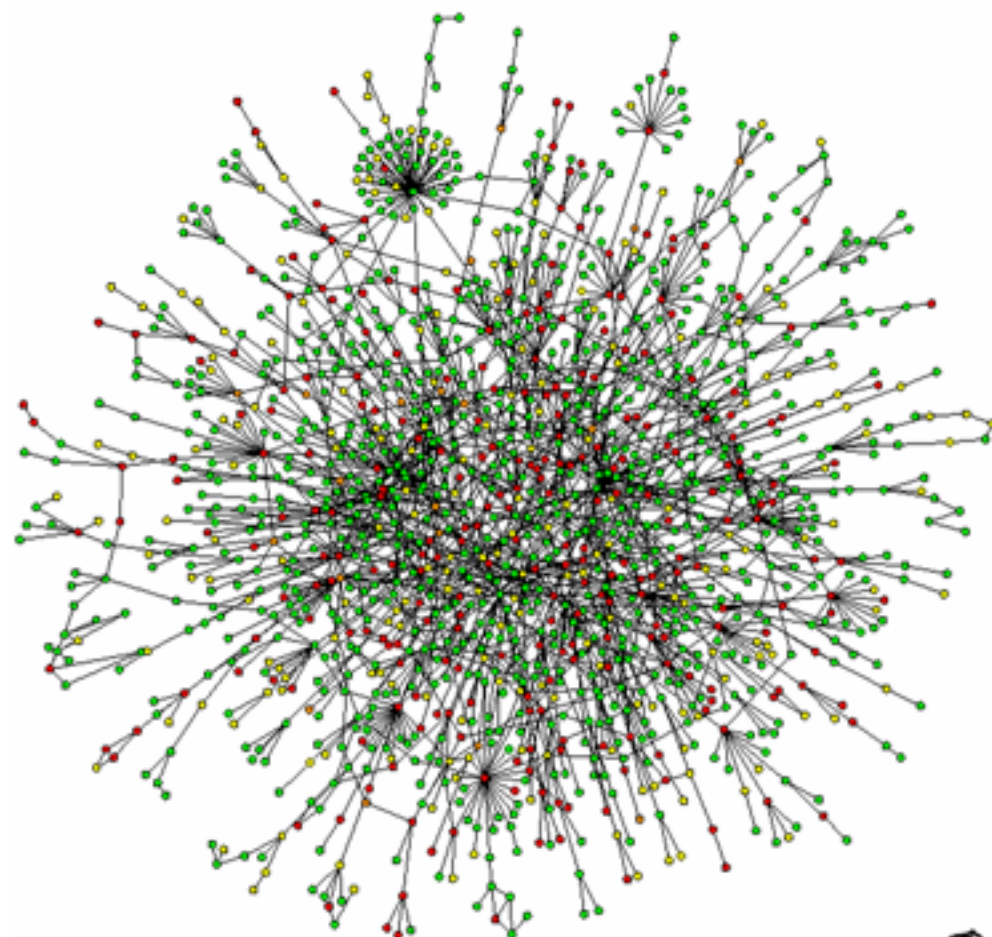
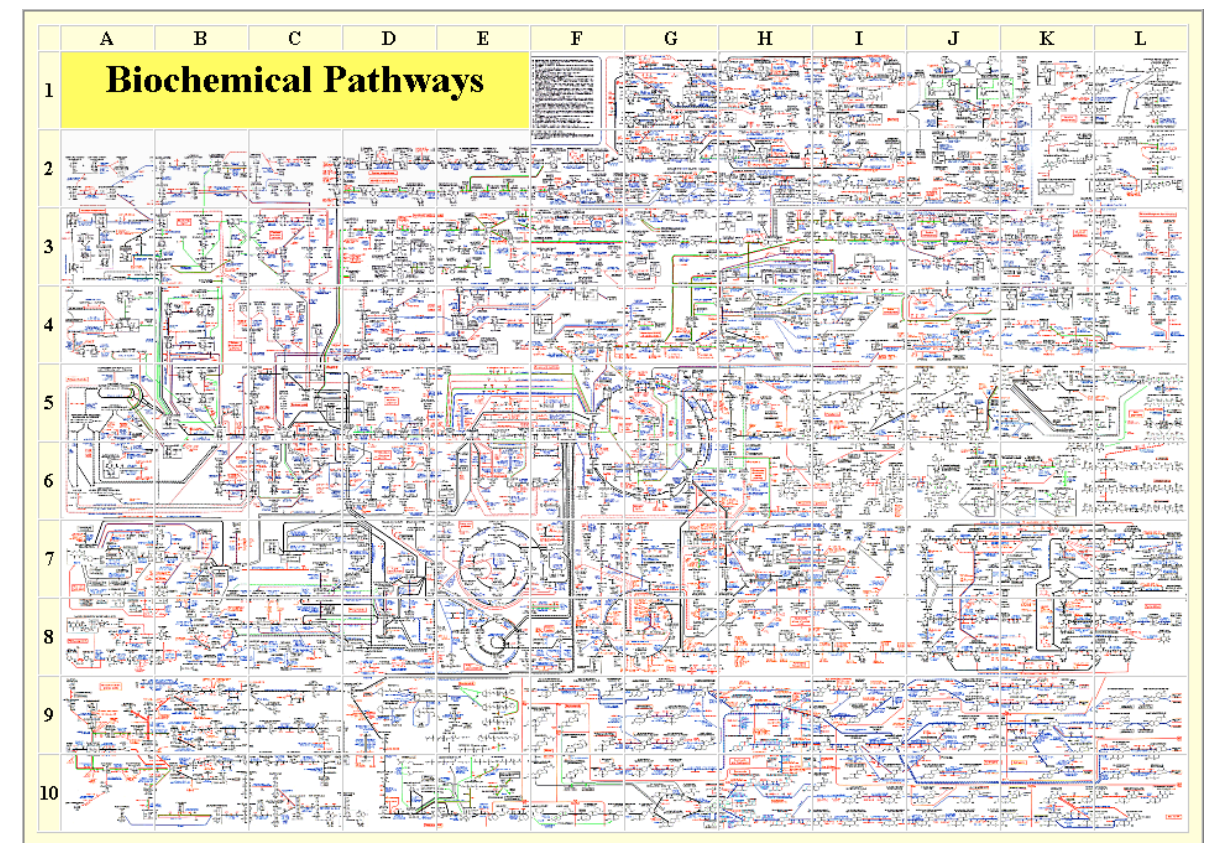
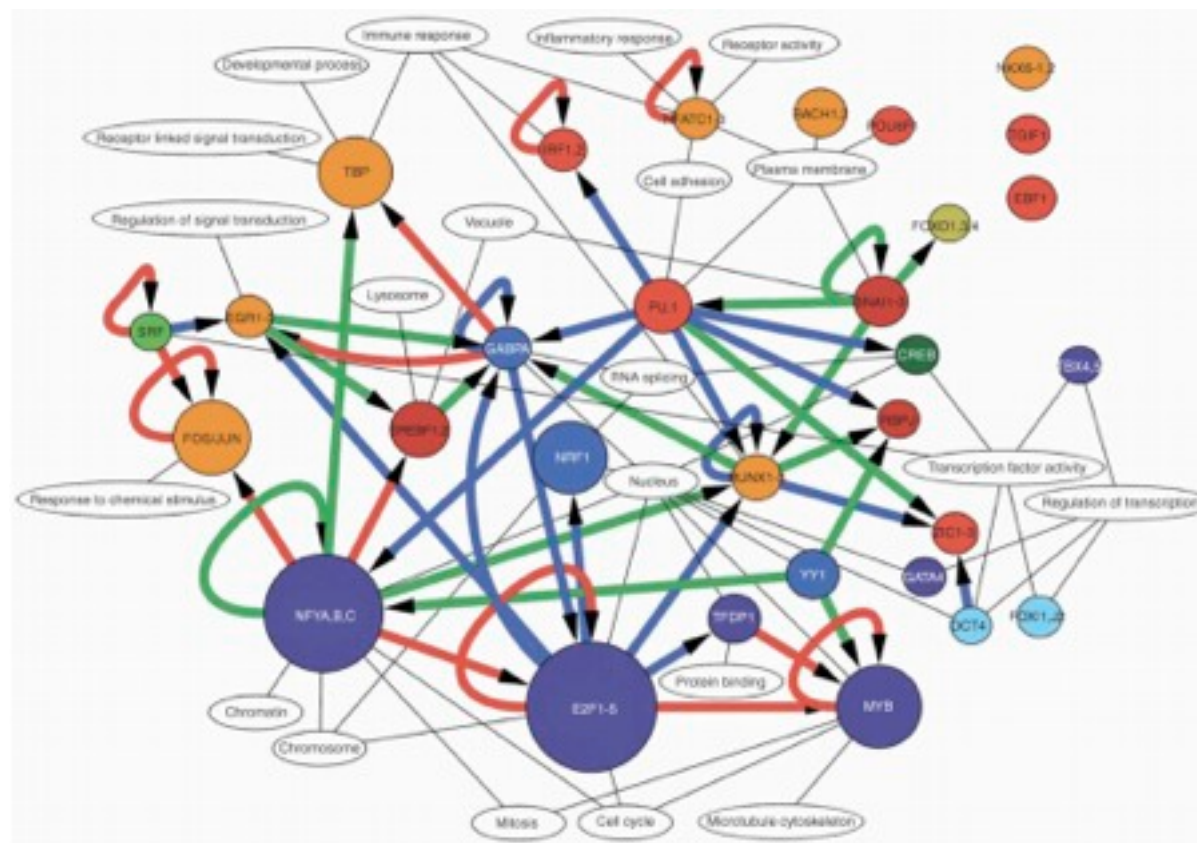




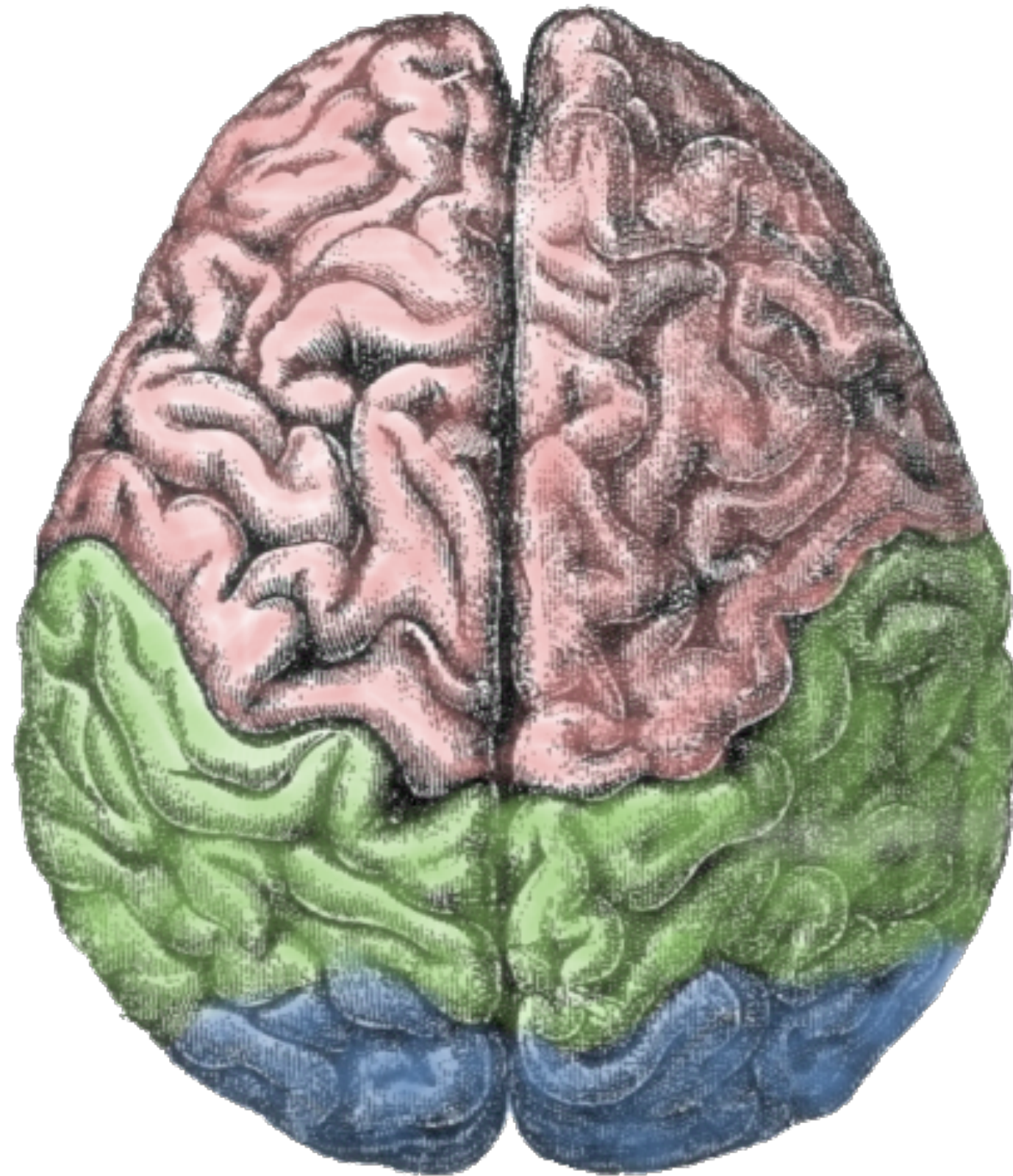
	A	B	C	D	E	F	G	H	I	J	K	L
1	<b>Biochemical Pathways</b>											
2												
3												
4												
5												
6												
7												
8												
9												
10												



















facebook

December 2010

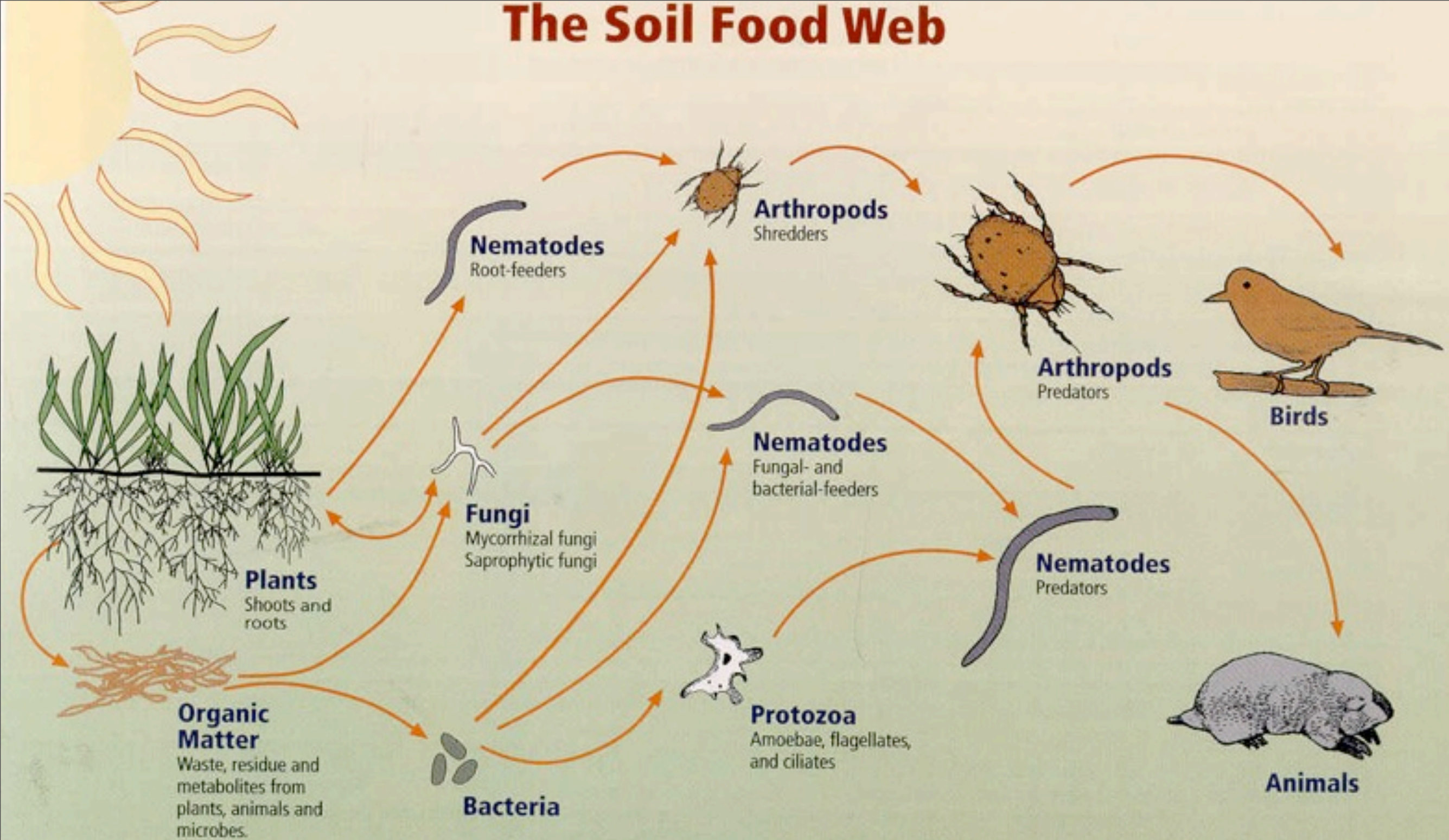




Friday, March 30, 12



# The Soil Food Web



**First trophic level:**  
Photosynthesizers

**Second trophic level:**  
Decomposers  
Mutualists  
Pathogens, parasites  
Root-feeders

**Third trophic level:**  
Shredders  
Predators  
Grazers

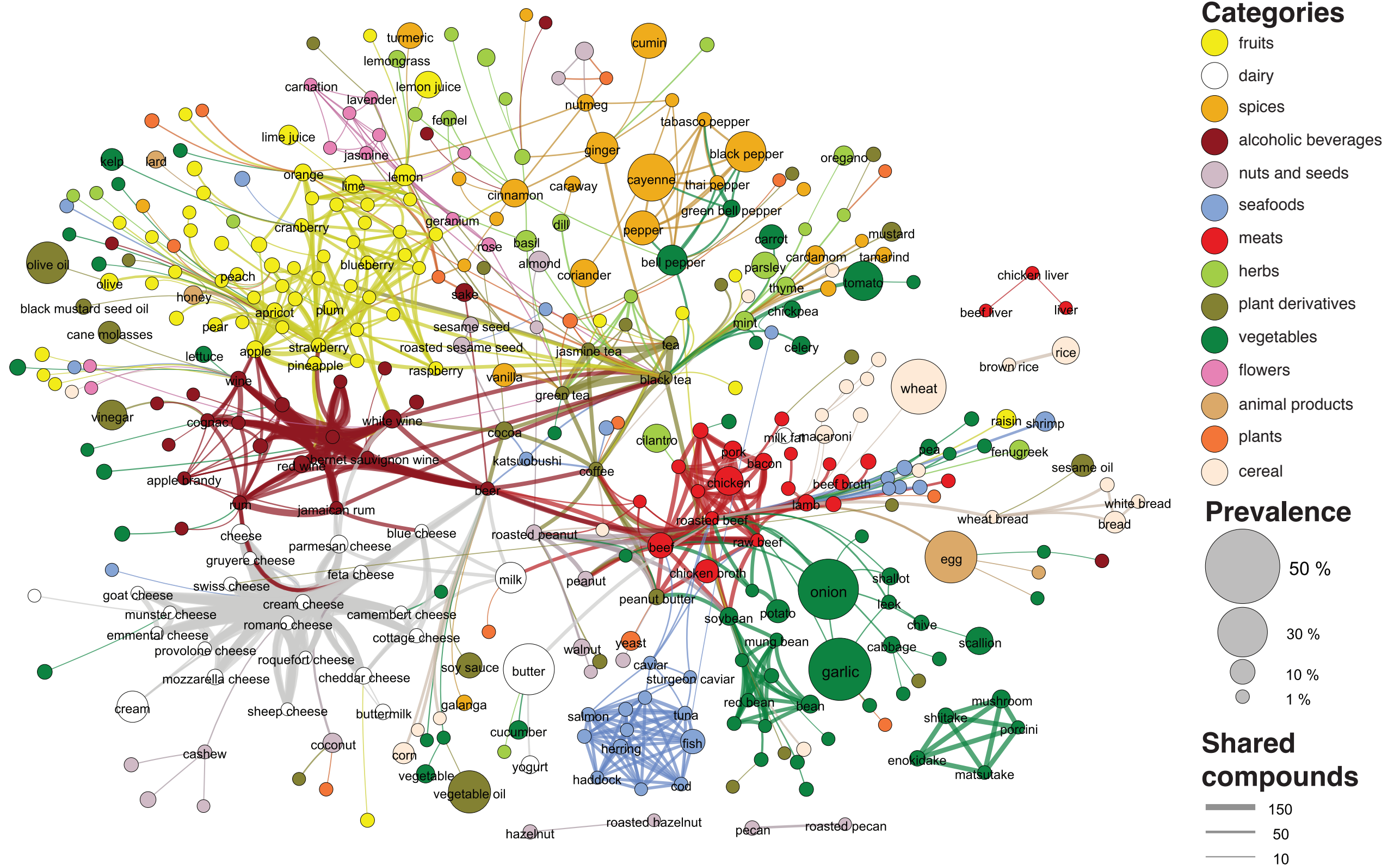
**Fourth trophic level:**  
Higher level predators

**Fifth and higher trophic levels:**  
Higher level predators





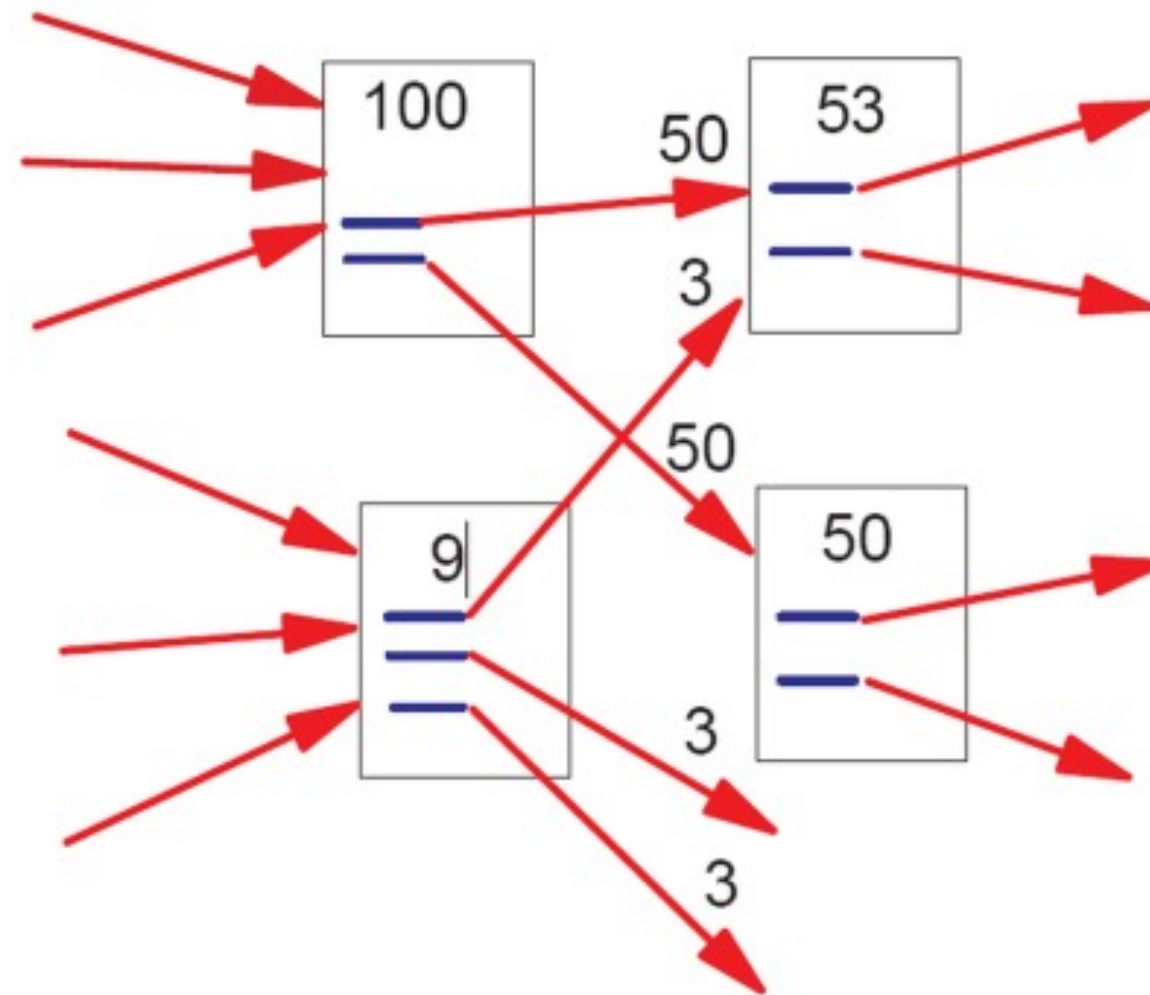




Y.-Y. Ahn, S. Ahnert, J. P. Bagrow, A.-L. Barabási, *Sci. Rep.* 2011

So what?

Google



**Pagerank =**  
Random walk problem on a network





facebook

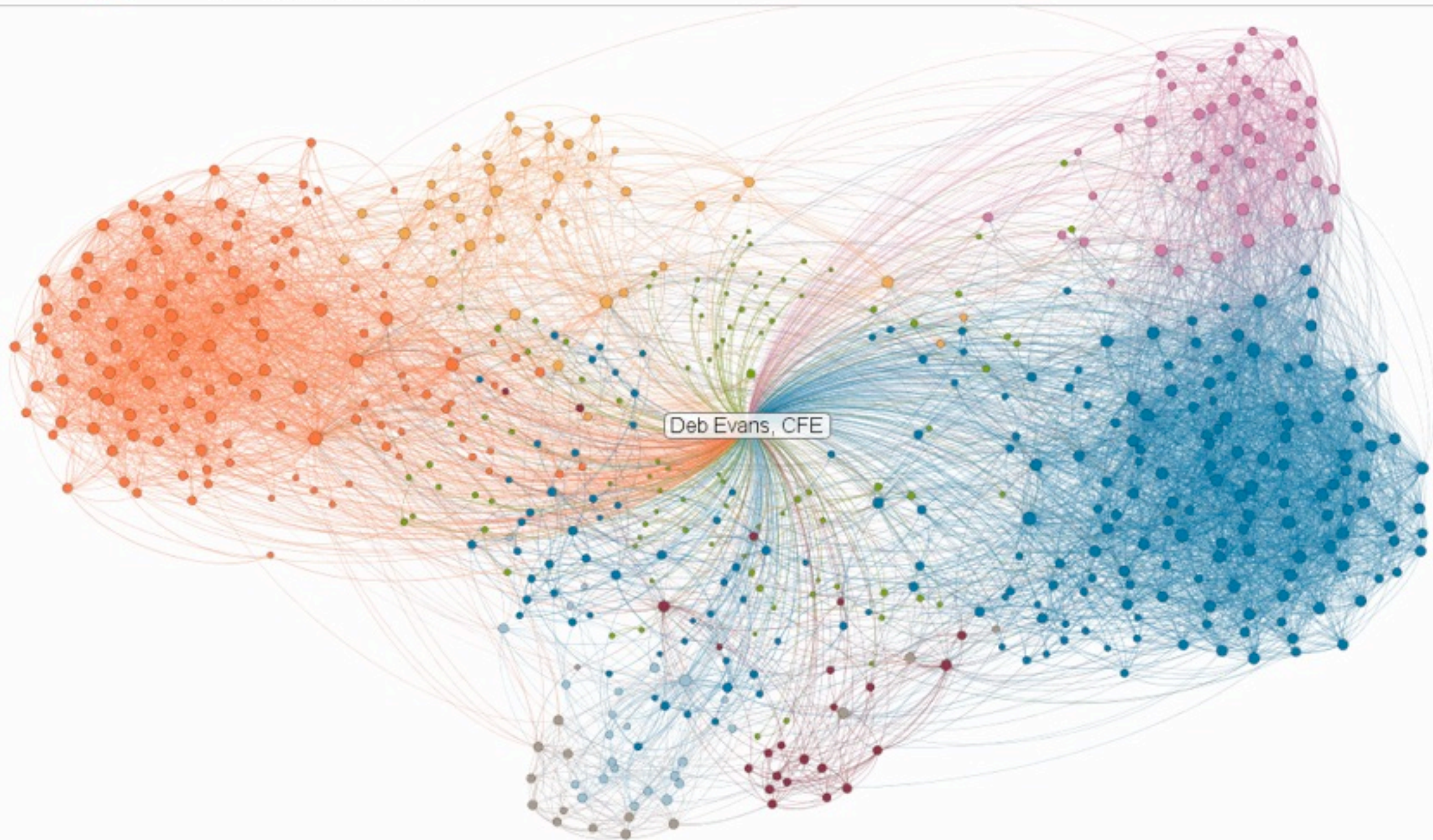
December 2010





LinkedIn Maps

Deb Evans, CFE's Professional Network  
as of January 27, 2011

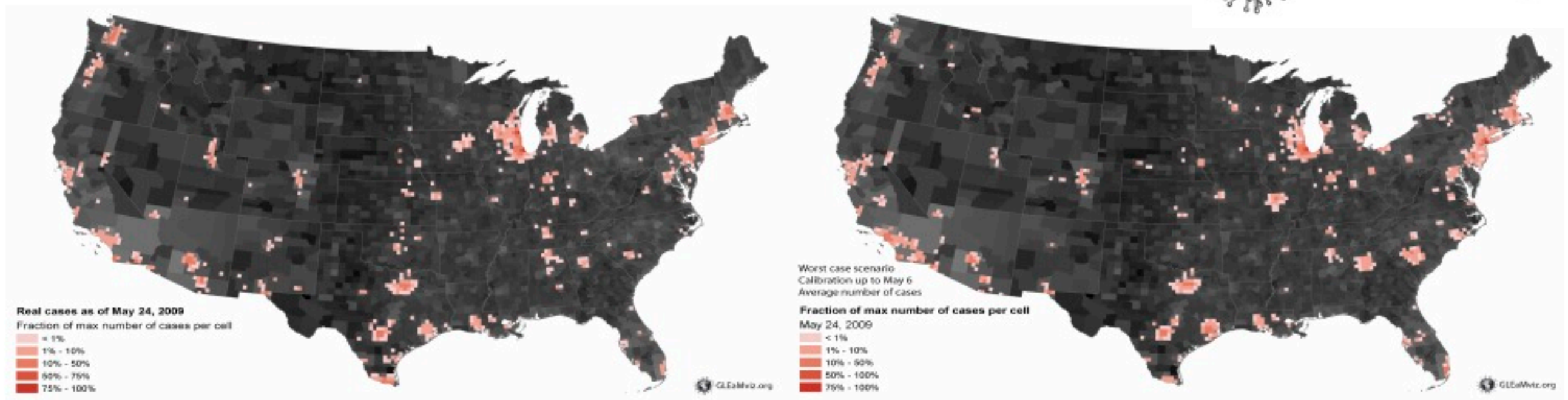


©2010 LinkedIn - Get your network map at [inmaps.linkedinlabs.com](http://inmaps.linkedinlabs.com)

# H1N1 Pandemic prediction



GLEaMviz.org



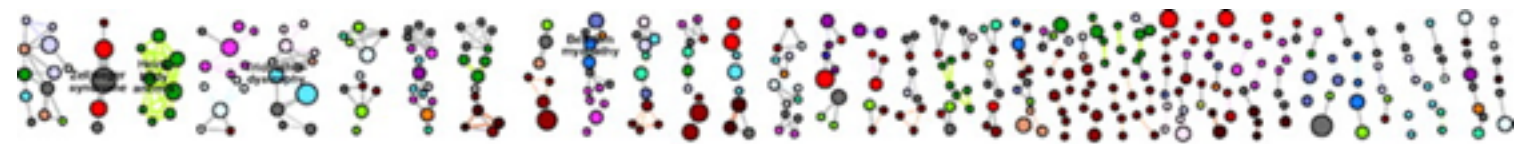
Real

Prediction

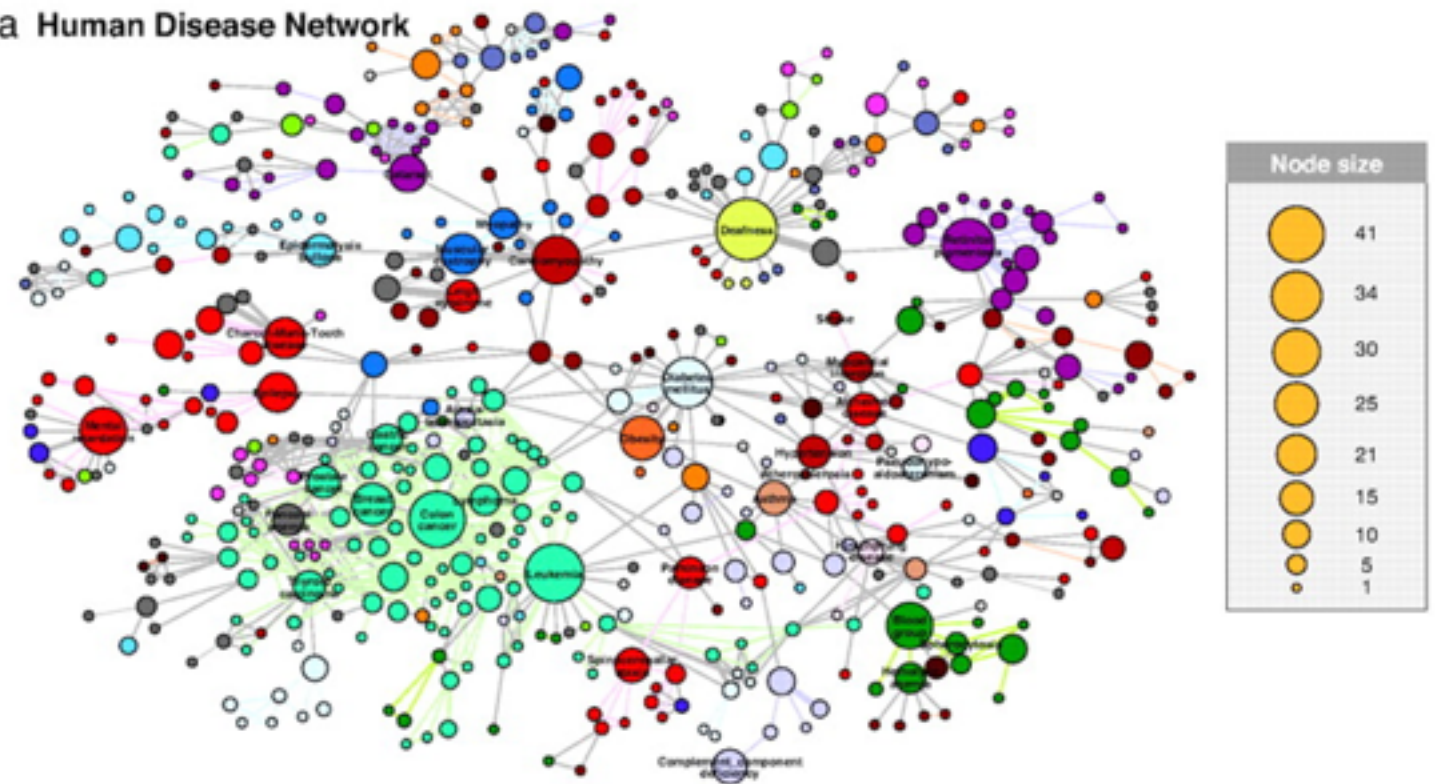
Reaction-diffusion system with  
transportation networks



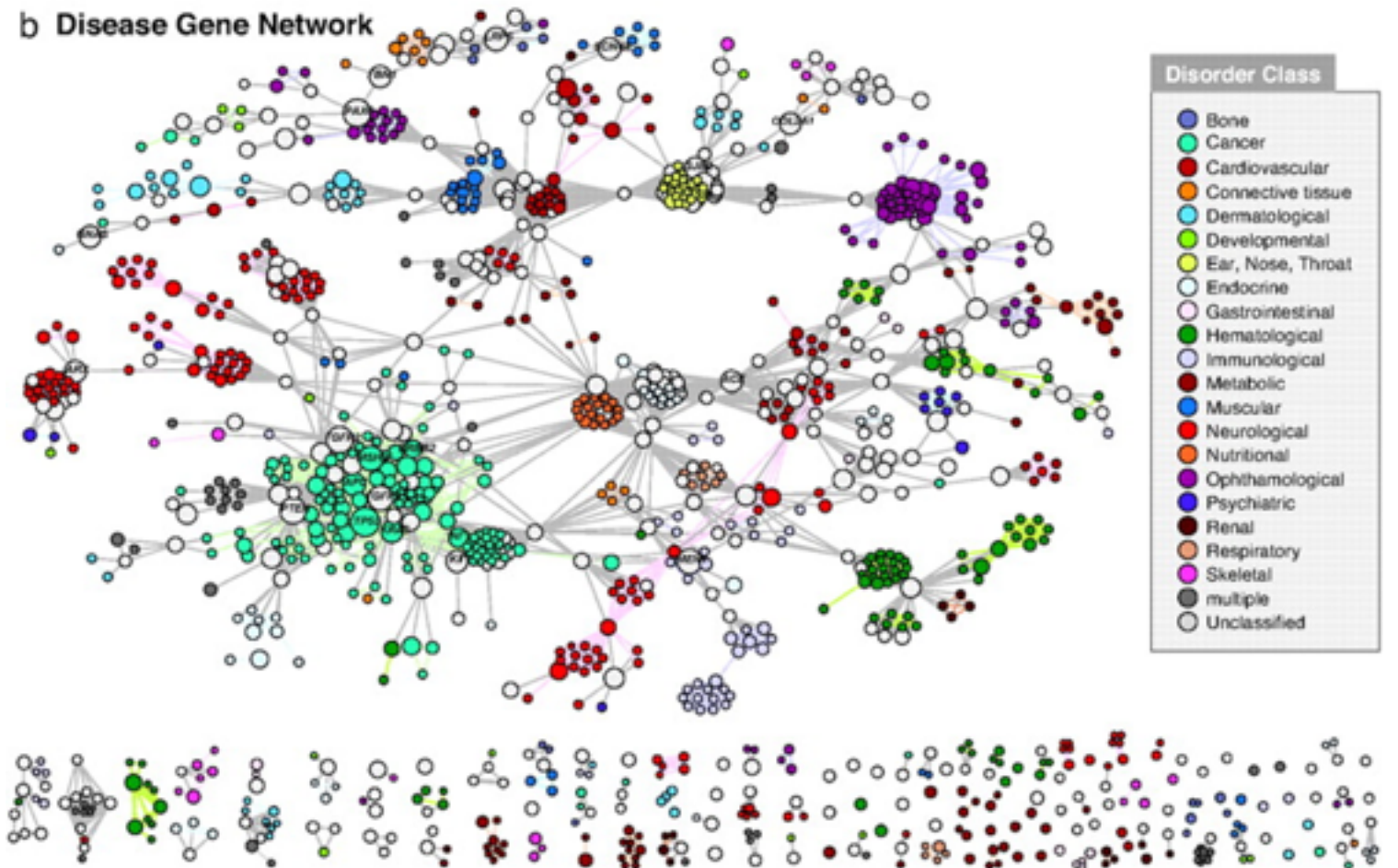




a Human Disease Network



b Disease Gene Network







# Blueprint for antimicrobial hit discovery targeting metabolic networks

Y. Shen<sup>a,b,1</sup>, J. Liu<sup>b,1</sup>, G. Estiu<sup>a</sup>, B. Isin<sup>b</sup>, Y-Y. Ahn<sup>c,d</sup>, D-S. Lee<sup>c,d,4</sup>, A-L. Barabási<sup>c,d</sup>, V. Kapatral<sup>e</sup>, O. Wiest<sup>a,2,3</sup>, and Z. N. Oltvai<sup>b,2,3</sup>

<sup>a</sup>Department of Chemistry and Biochemistry, University of Notre Dame, Notre Dame, IN, 46556; <sup>b</sup>Departments of Pathology and University of Pittsburgh, Pittsburgh, PA, 15261; <sup>c</sup>Center for Complex Network Research and Departments of Physics, Biology, and Northeastern University, Boston, MA 02115; <sup>d</sup>Center for Cancer Systems Biology, Dana-Farber Cancer Institute, Boston, MA 02115; and <sup>e</sup>Integrated Genomics, Inc., Chicago, IL 60612

Edited by H. Eugene Stanley, Boston University, Boston, MA, and approved November 9, 2009 (received for review August 18, 2009)

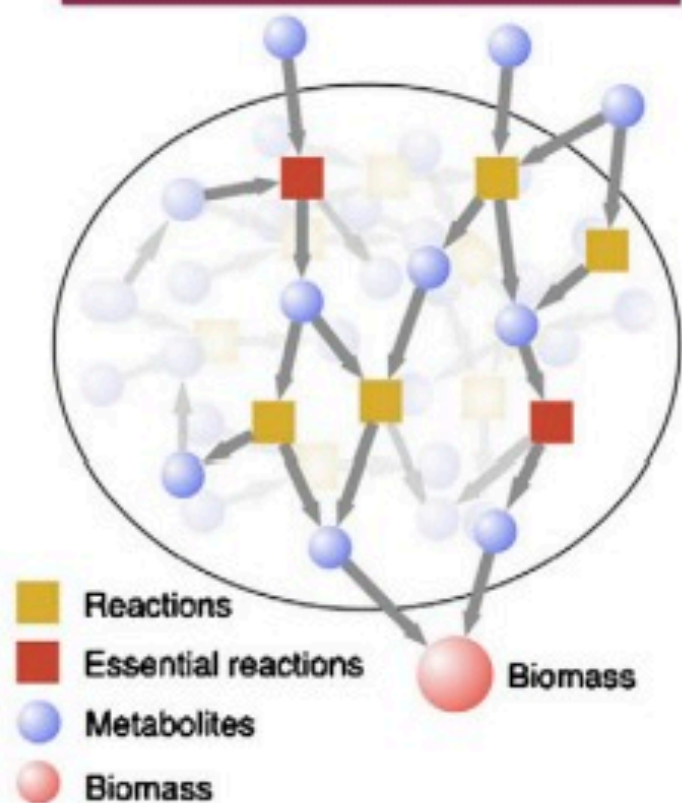
Advances in genome analysis, network biology, and computational chemistry have the potential to revolutionize drug discovery by combining system-level identification of drug targets with the atomistic modeling of small molecules capable of modulating their activity. To demonstrate the effectiveness of such a discovery pipeline, we deduced common antibiotic targets in *Escherichia coli* and *Staphylococcus aureus* by identifying shared tissue-specific

displays the main steps of our protocol. rapid discovery of new antibiotic hits against molecular targets addresses a critical need that is applicable to all diseases involving altered chemical reaction networks.

## Results

Identification of Antimicrobial Targets by

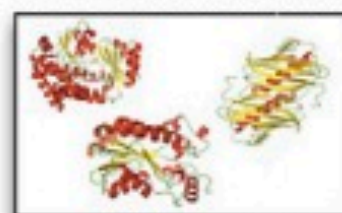
## Network-based target identification



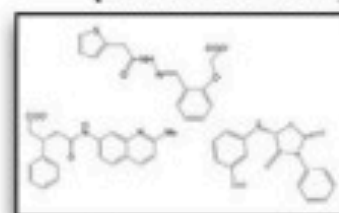
- Metabolic network reconstruction
- Essential reaction identification (FBA)

## Virtual screening

Target enzymes catalyzing the essential reactions

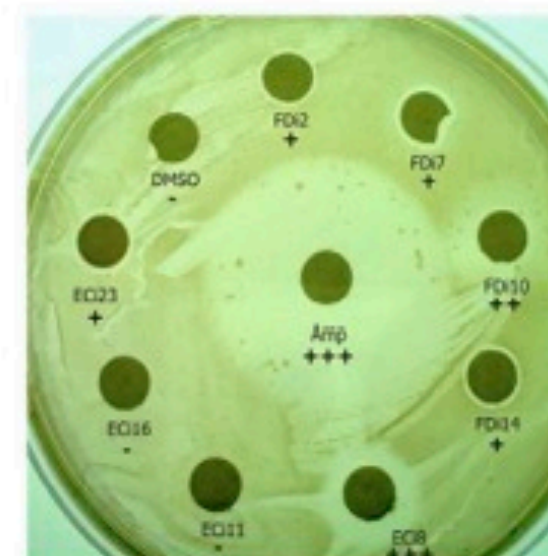


Drug candidates (small molecules)



- Retrieval or homology modeling of identified targets
- Virtual screening of  $10^6$  compounds for each target
- Cross docking with multiple scoring functions
- MM-PBSA ensemble re-scoring of selected candidates

## Experimental validation



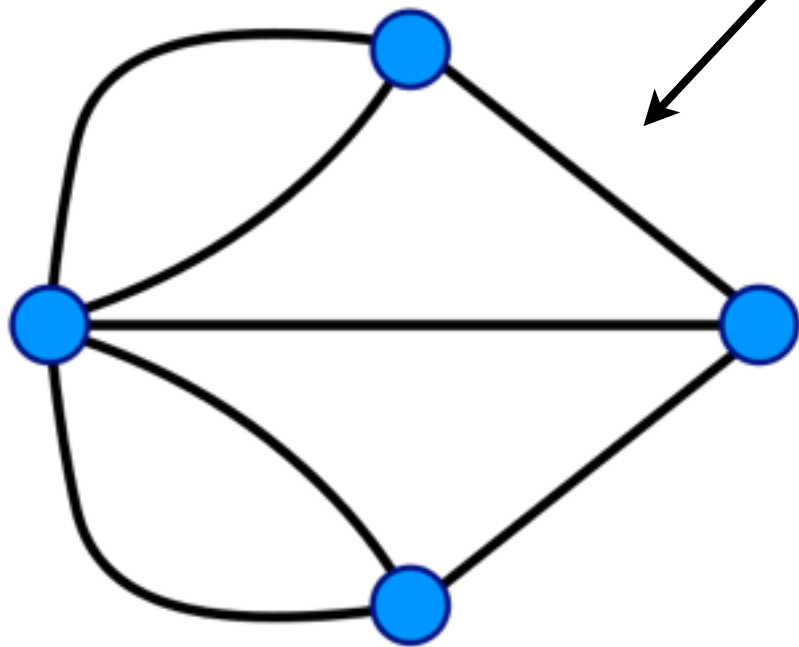
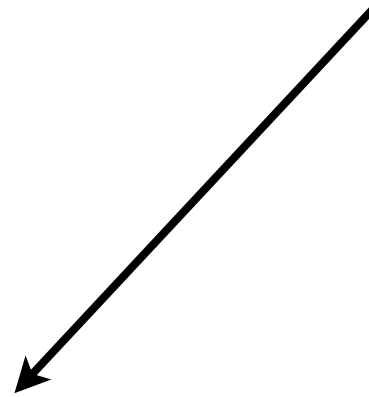
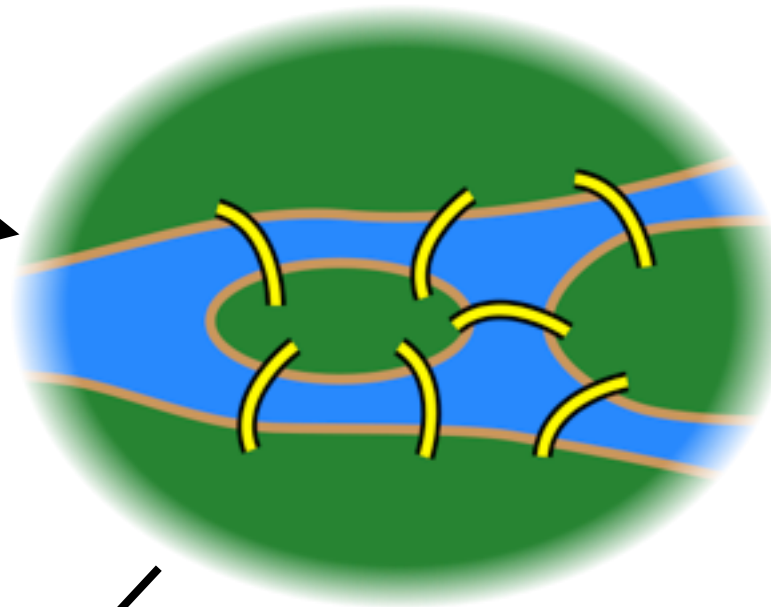
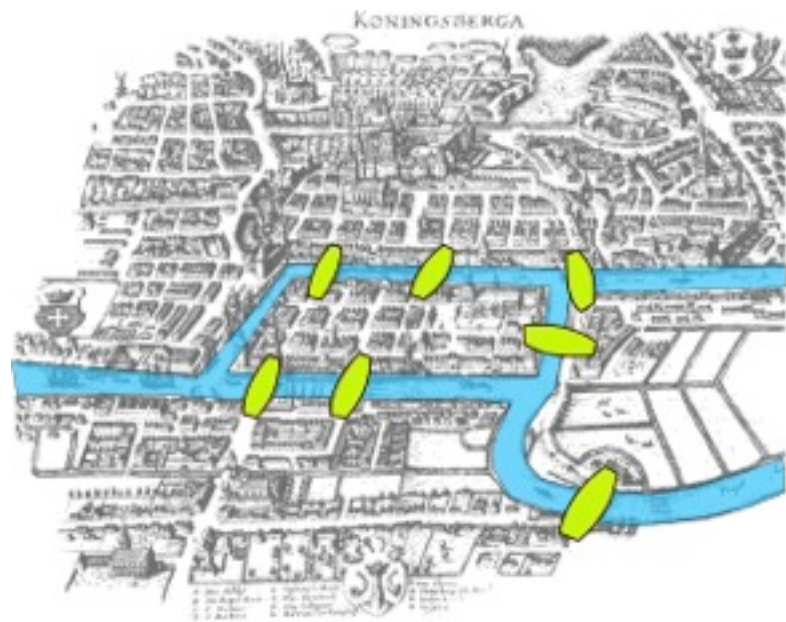
- *In vitro* enzyme activity assay
- *In vivo* viability assay
- *In vivo* cyto-toxicity assay



Can we understand a  
**complex system**  
without knowing the **structure**  
of it?

# NETWORKS





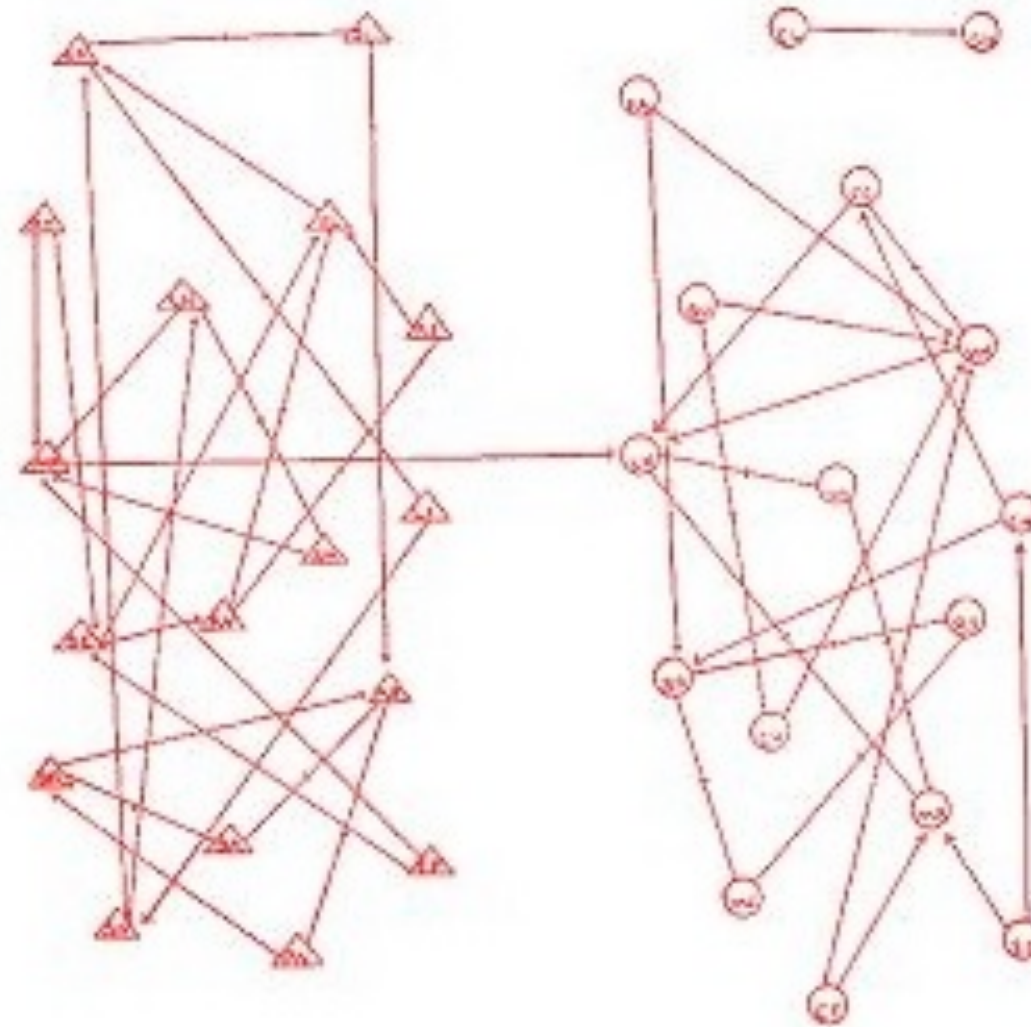
Leonhard Euler

# “Sociogram”

## EMOTIONS MAPPED BY NEW GEOGRAPHY

Charts Seek to Portray the  
Psychological Currents of  
Human Relationships.

**New York Times**  
April 3, 1933



# What's the structure of networks?





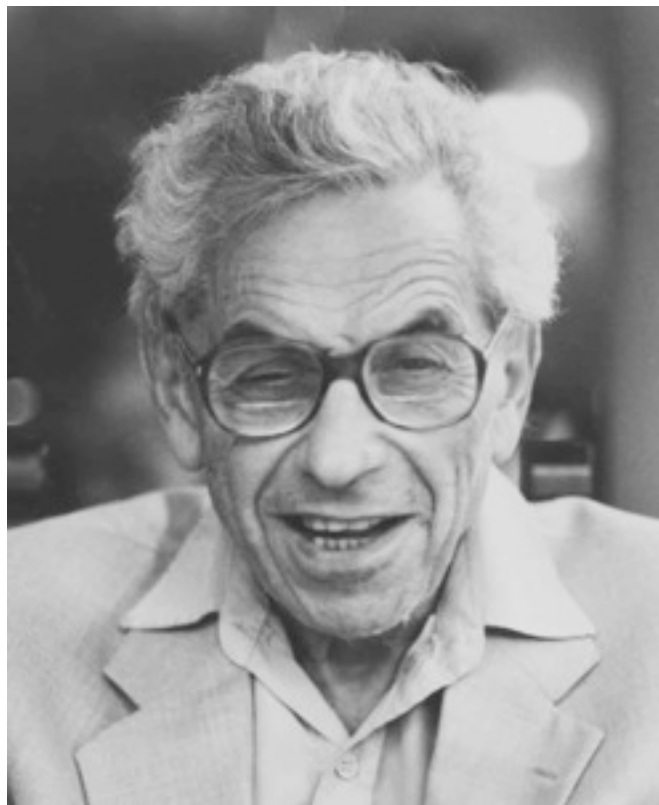
$p = 0.1$



$p = 0.25$



$p = 0.5$



Paul Erdős



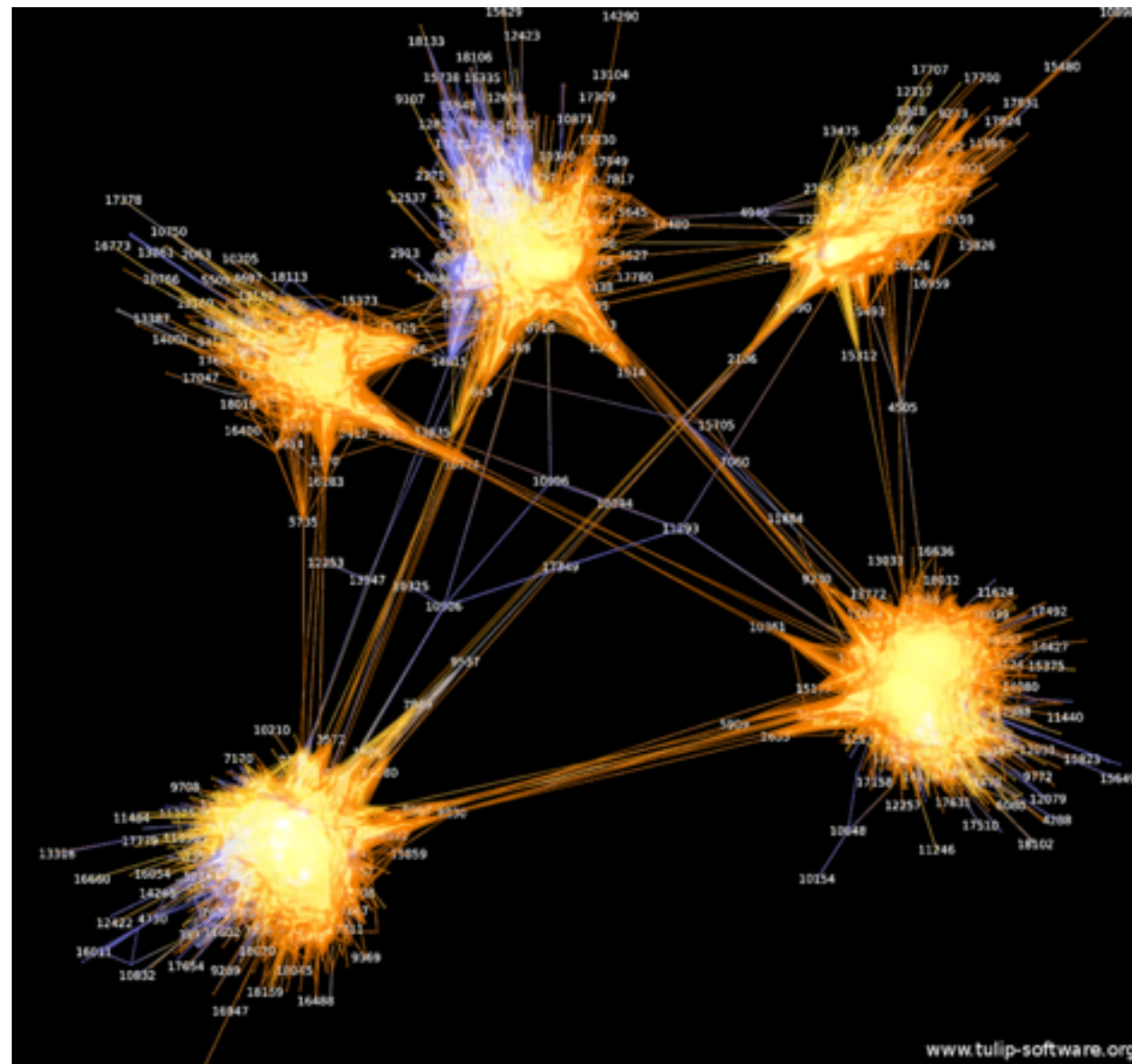
Alfréd Rényi

Clustering

Small-world

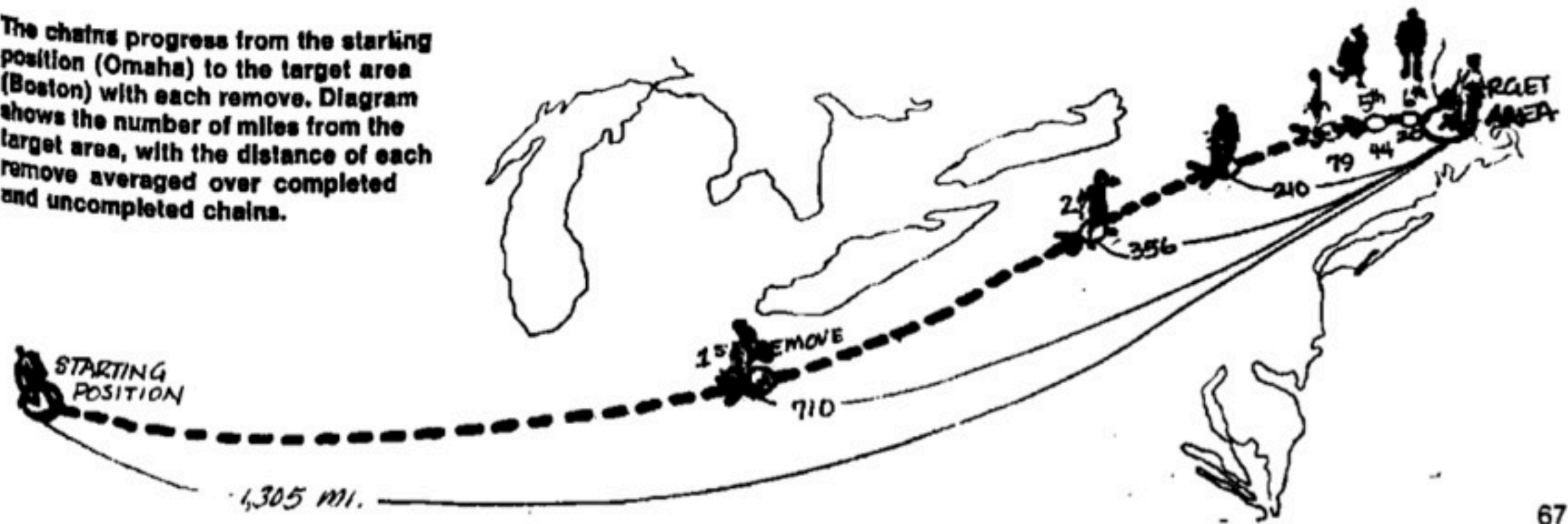
Heterogeneity





It's not random! We  
form clusters.

The chains progress from the starting position (Omaha) to the target area (Boston) with each remove. Diagram shows the number of miles from the target area, with the distance of each remove averaged over completed and uncompleted chains.



67

“Small world experiment”



Stanley Milgram

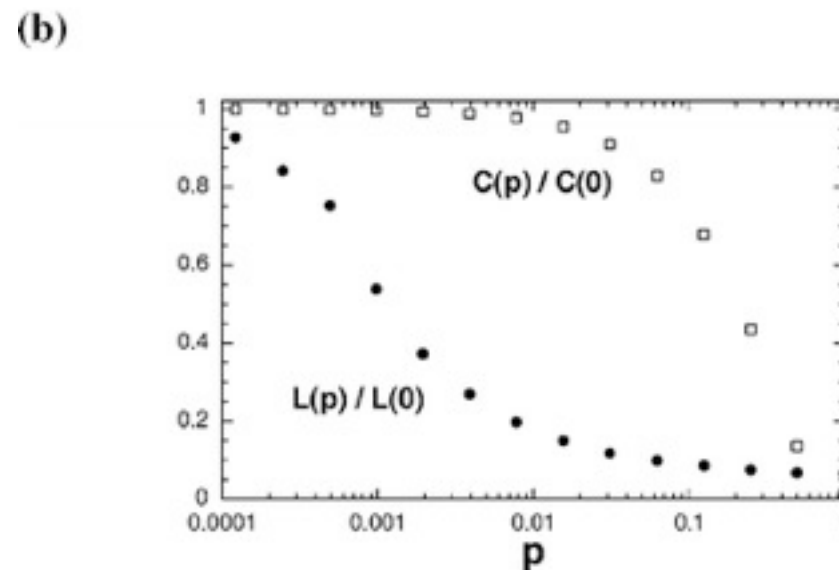
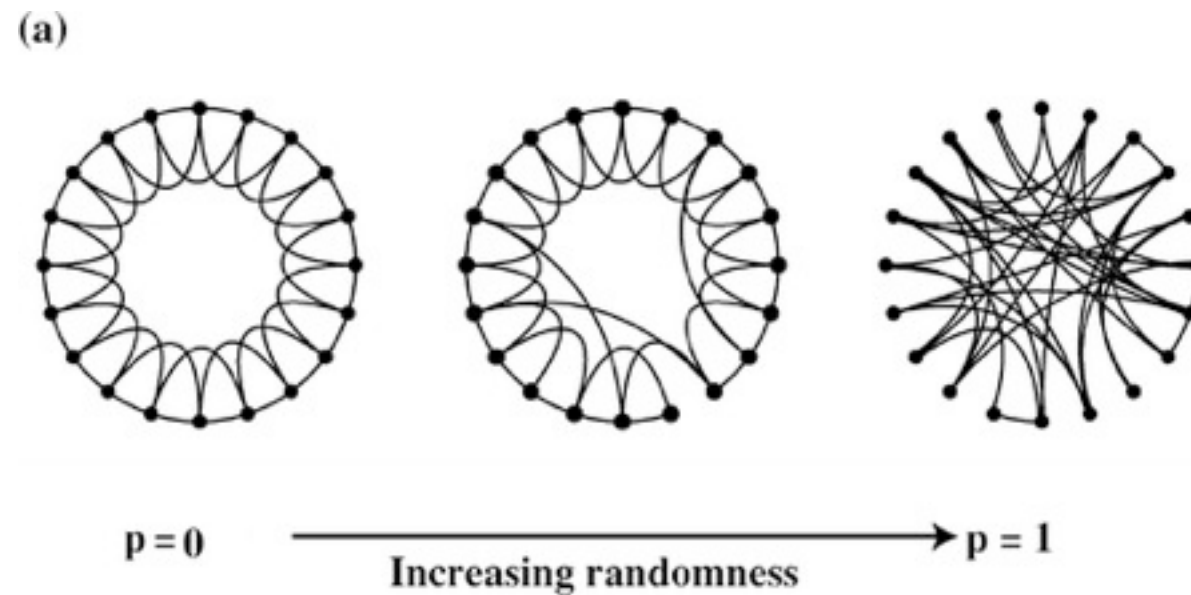


<http://oracleofbacon.org/index.php>

**We're clustered, but  
at the same time we  
are well-connected.**



# Duncan J. Watts



Steven H. Strogatz

Watts and Strogatz model

Watts & Strogatz, Nature 1998

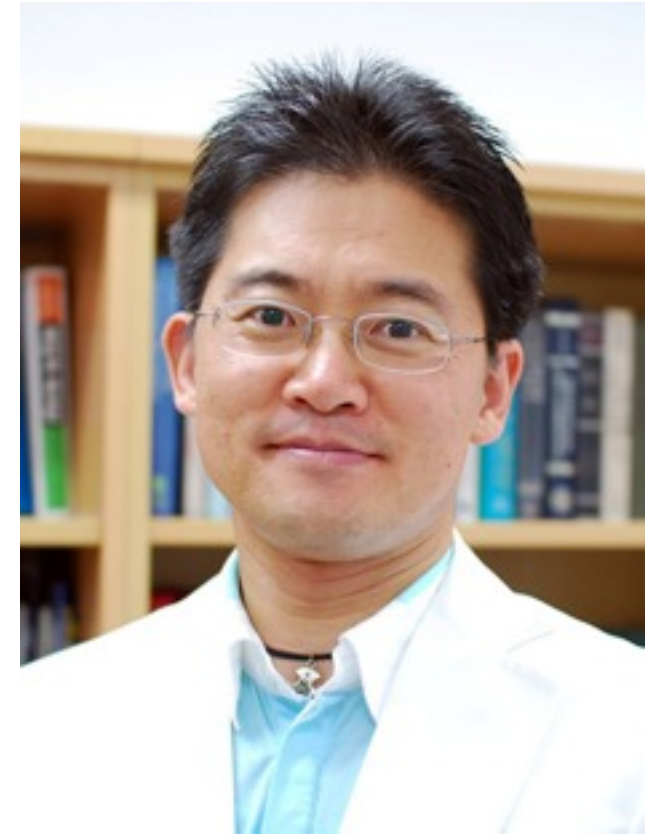
# Networks are heterogeneous!



Albert-László Barabási



Réka Albert



Hawoong Jeong



$p = 0.1$



$p = 0.25$

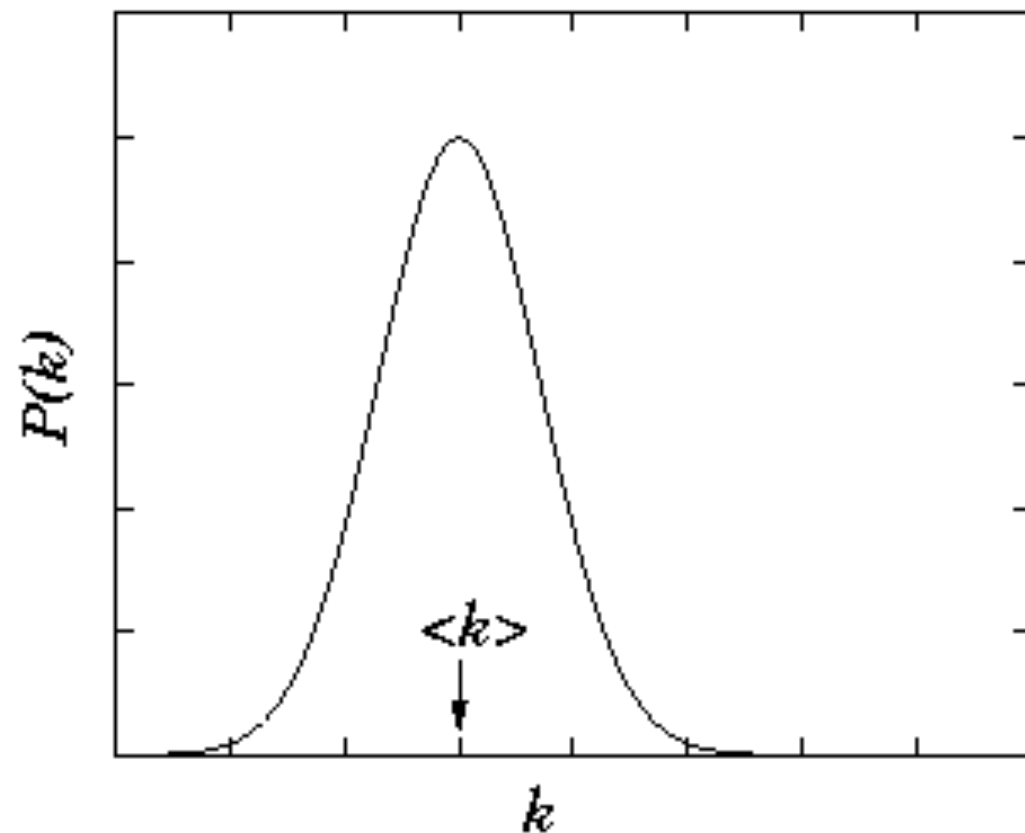


$p = 0.5$

## Poisson distribution

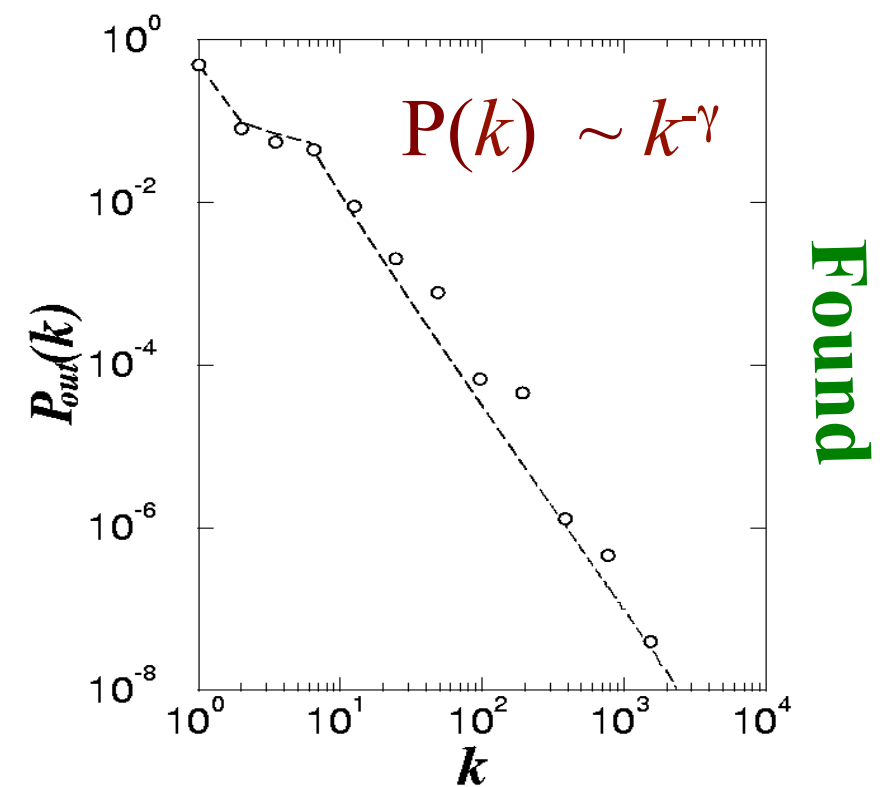
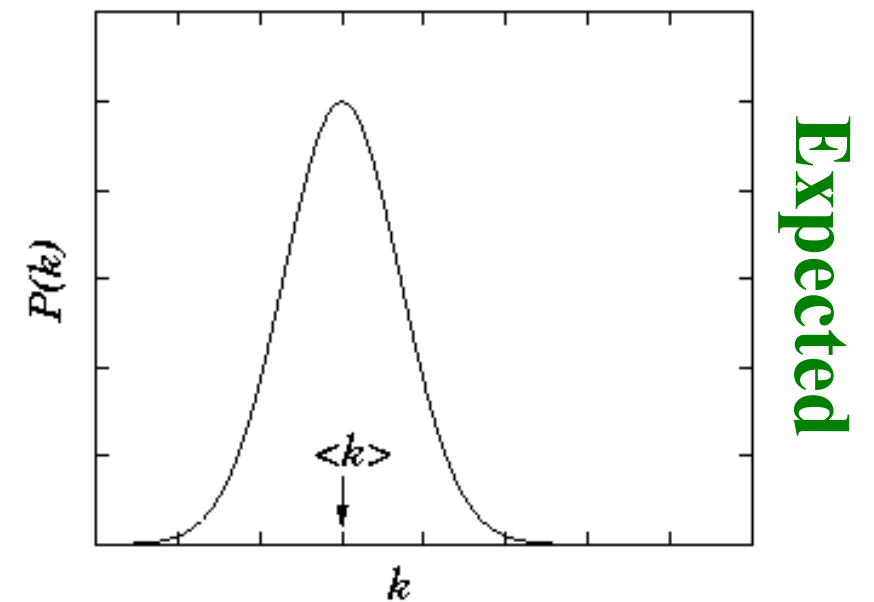


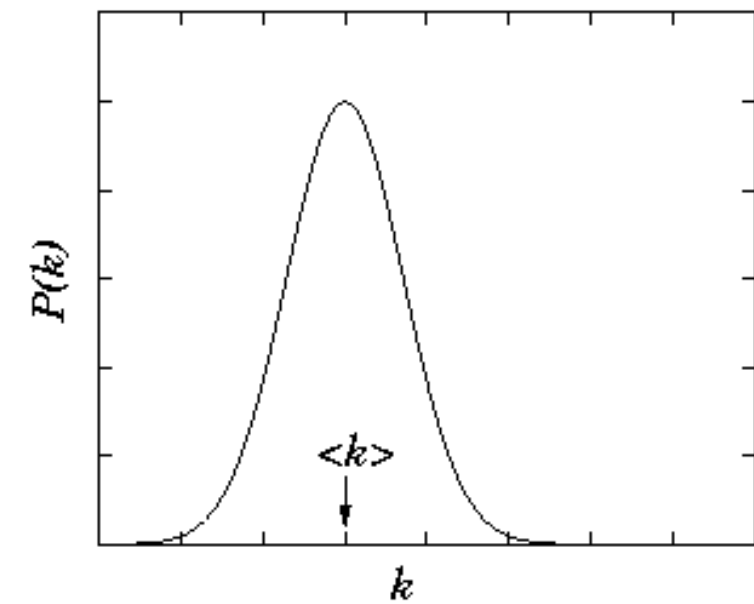
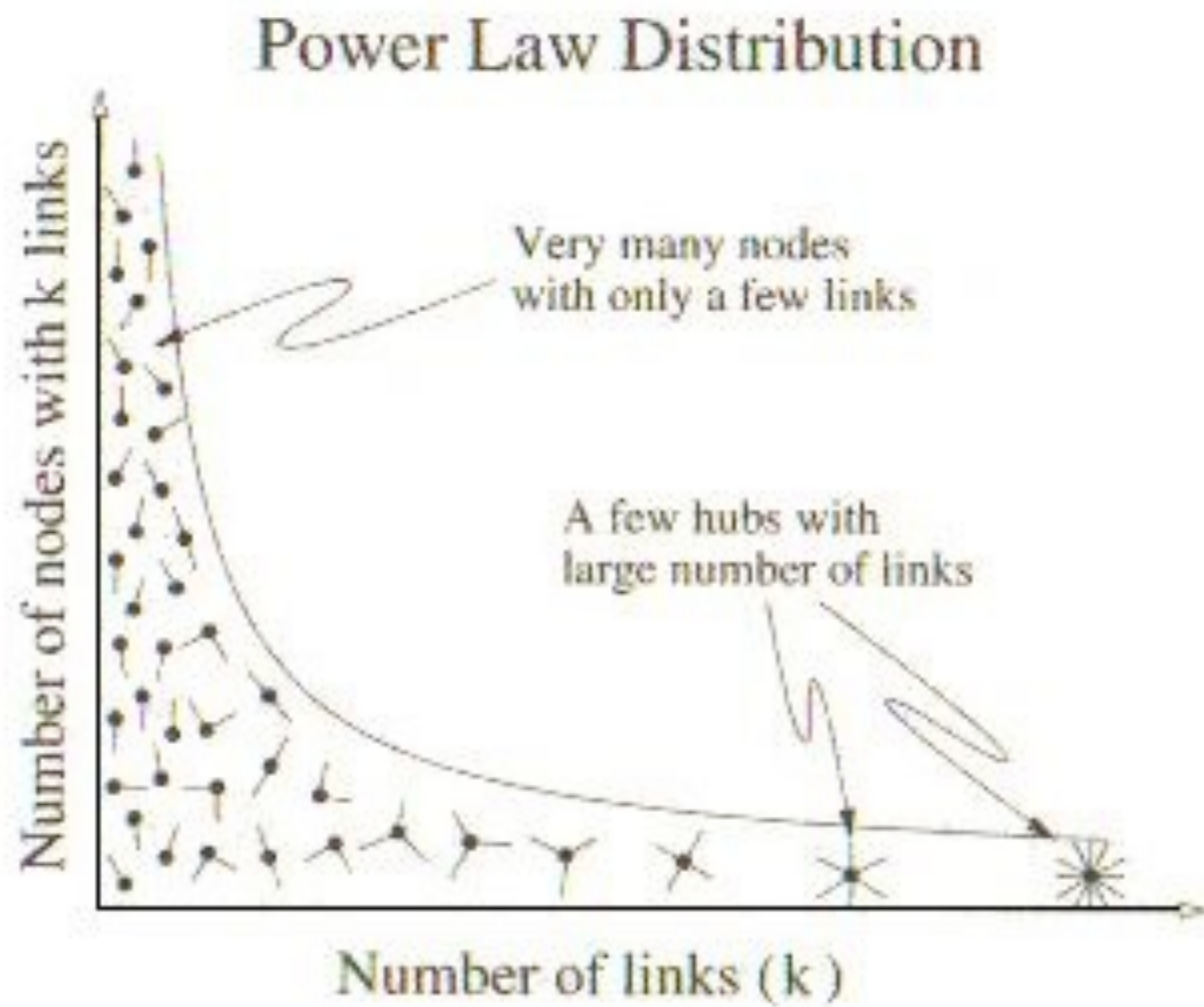
Degree: # of neighbors



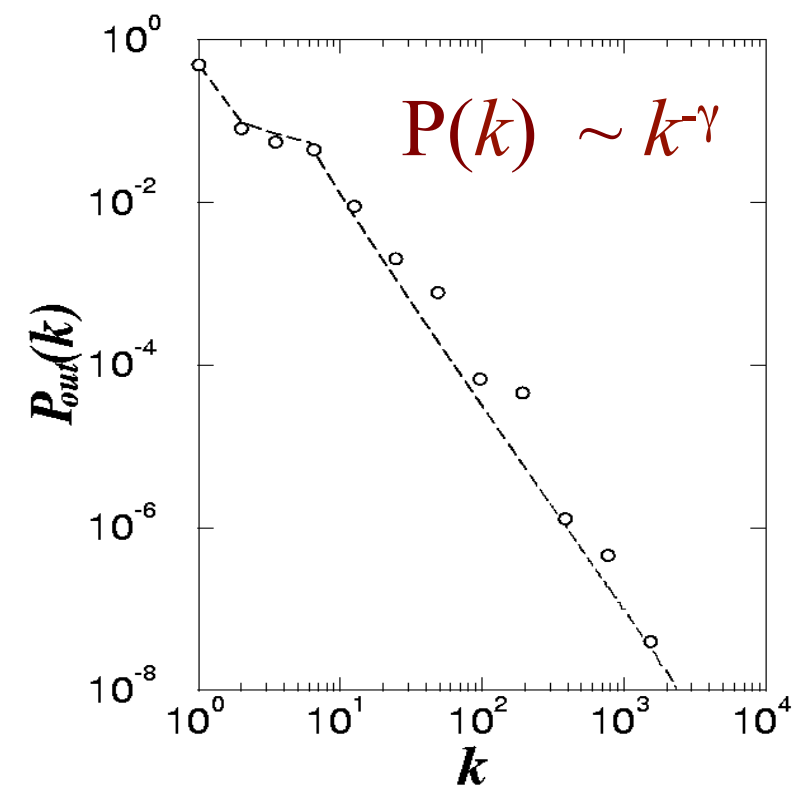


# What's the structure of networks?



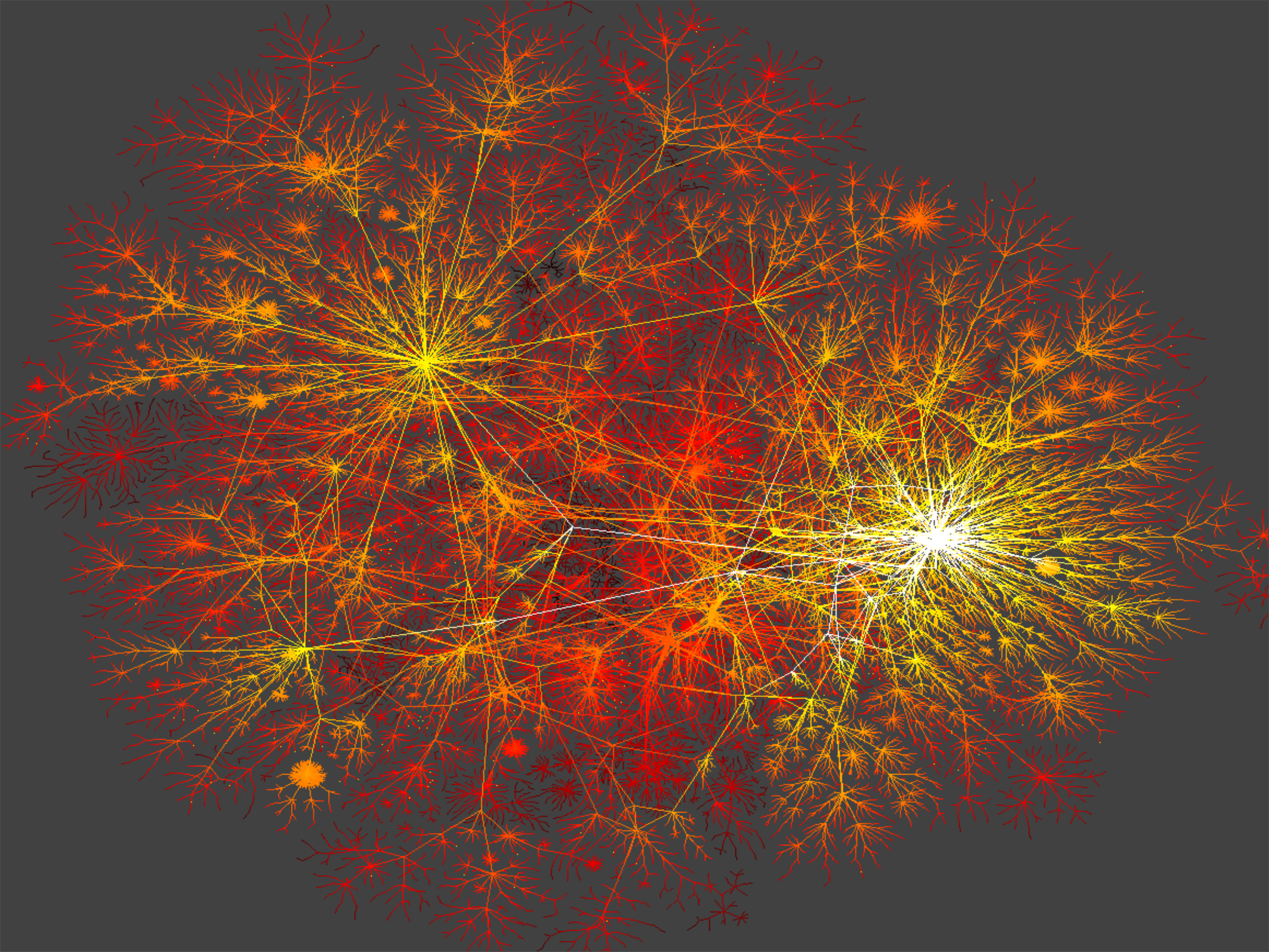


Expected

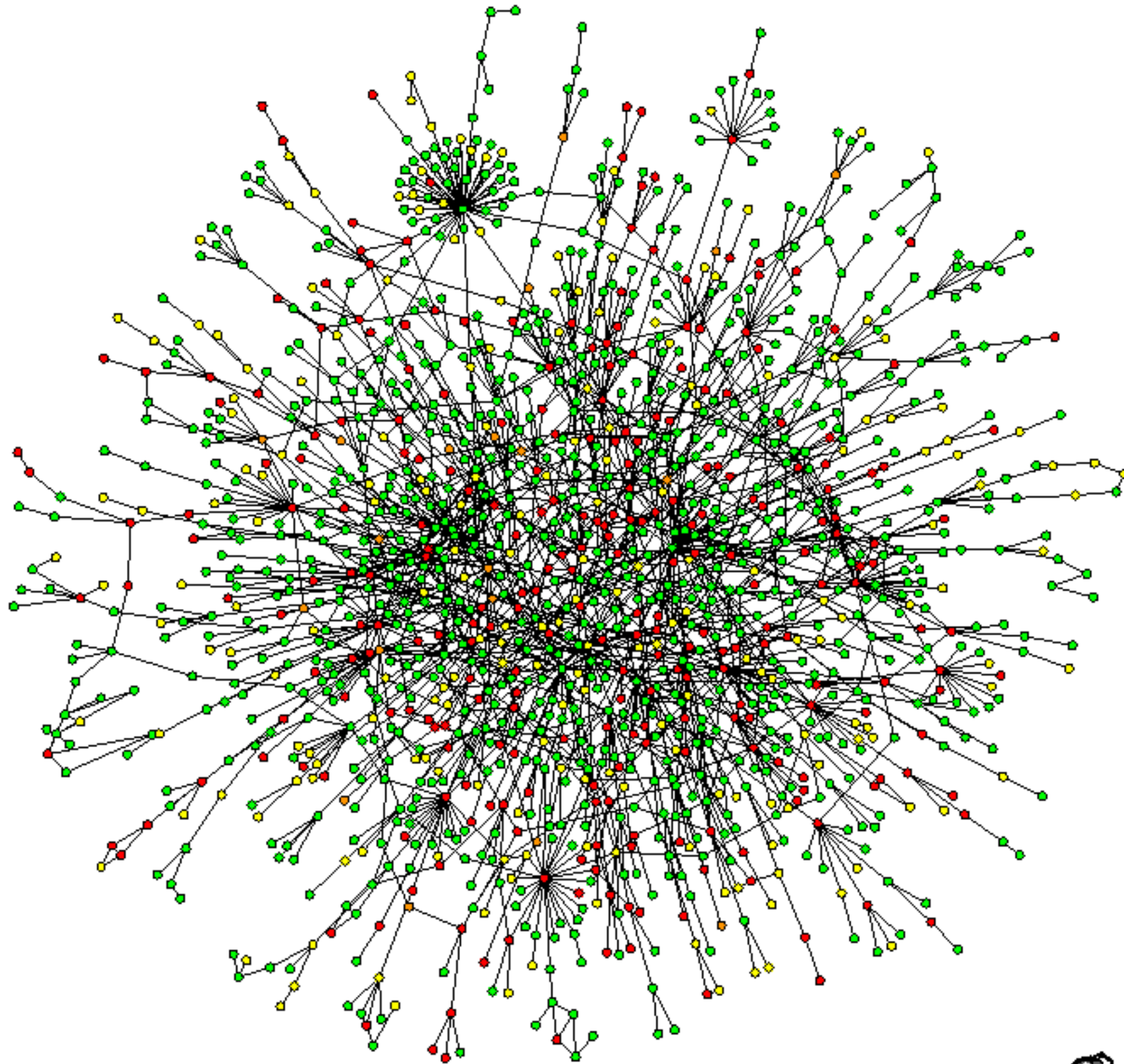


Found



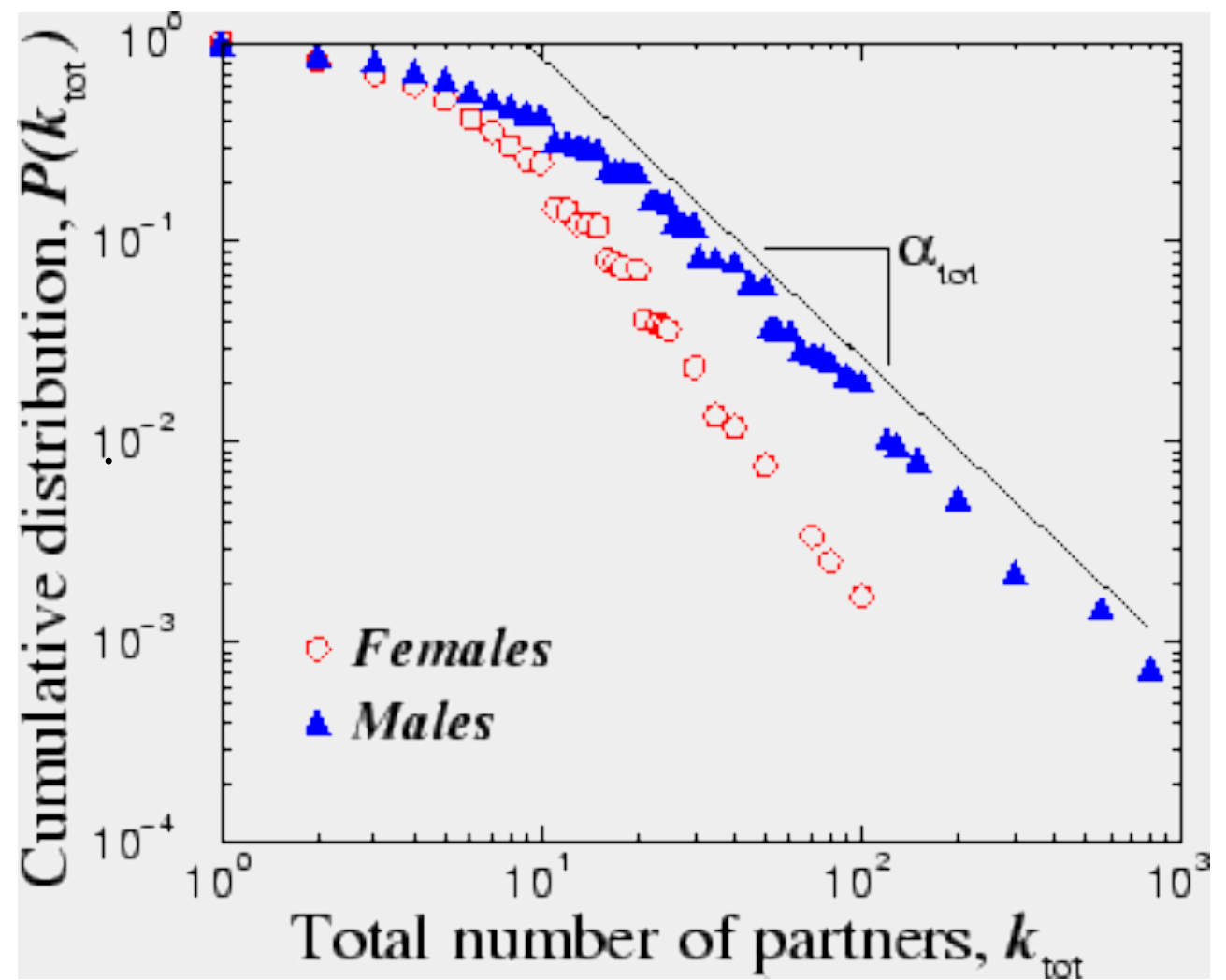






 Jeong et al., Nature 2001





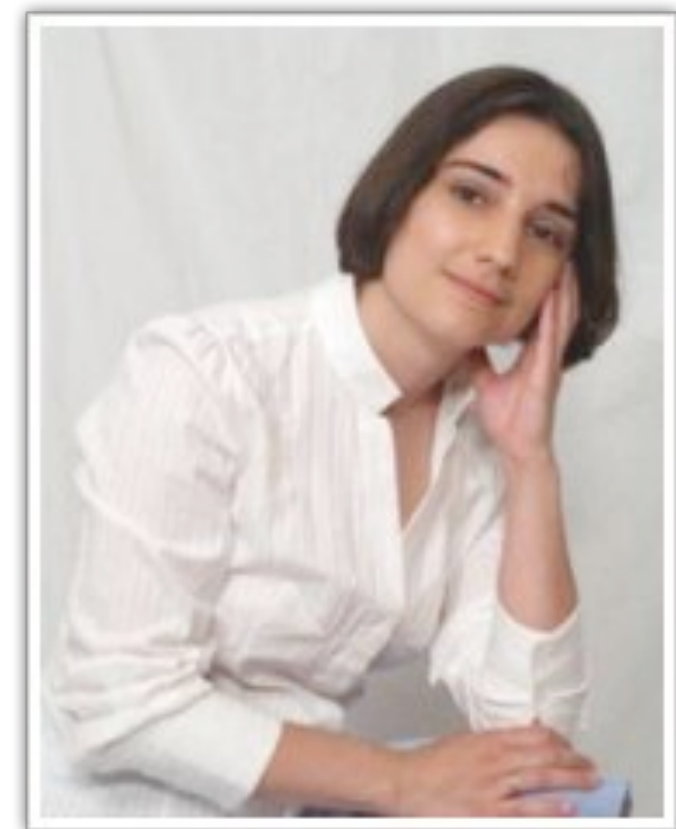
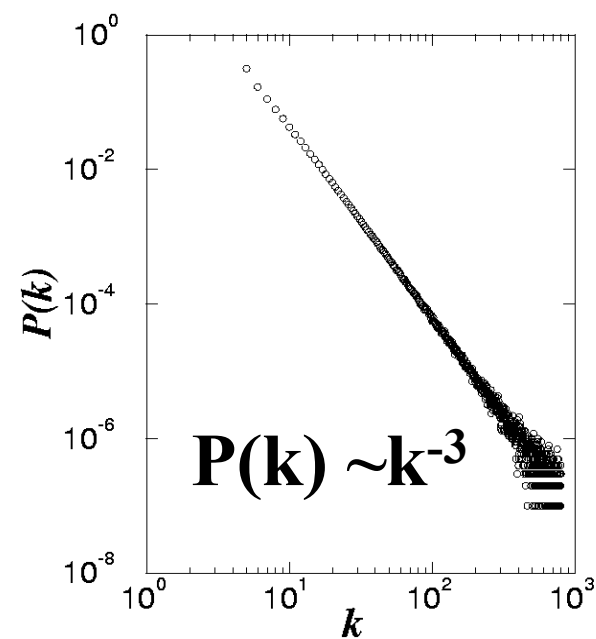
Liljeros et al., Nature 2001

### GROWTH:

add a new node with  $m$  links

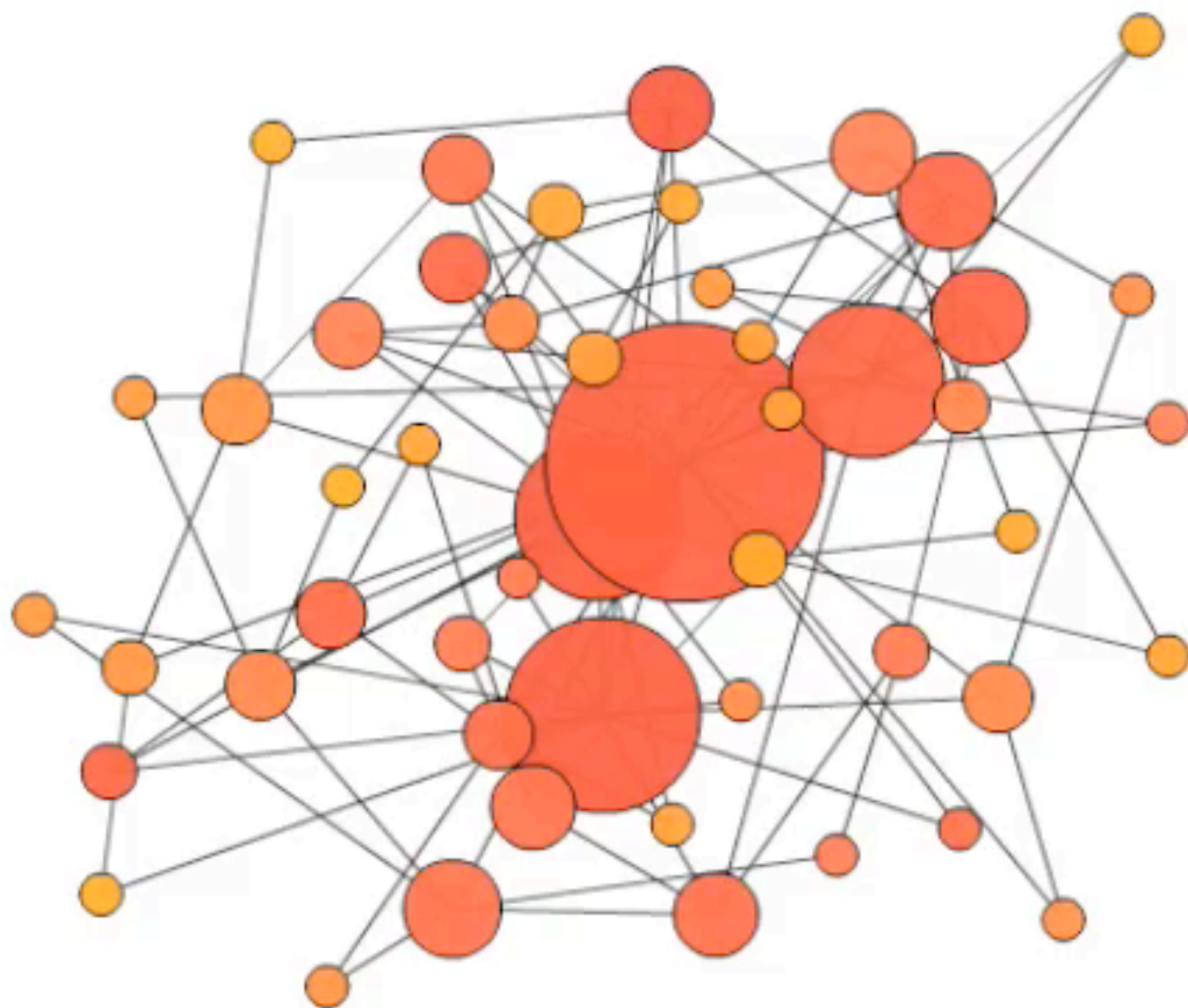
**PREFERENTIAL ATTACHMENT:** the probability that a node connects to a node with  $k$  links is proportional to  $k$ .

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}$$



# Error

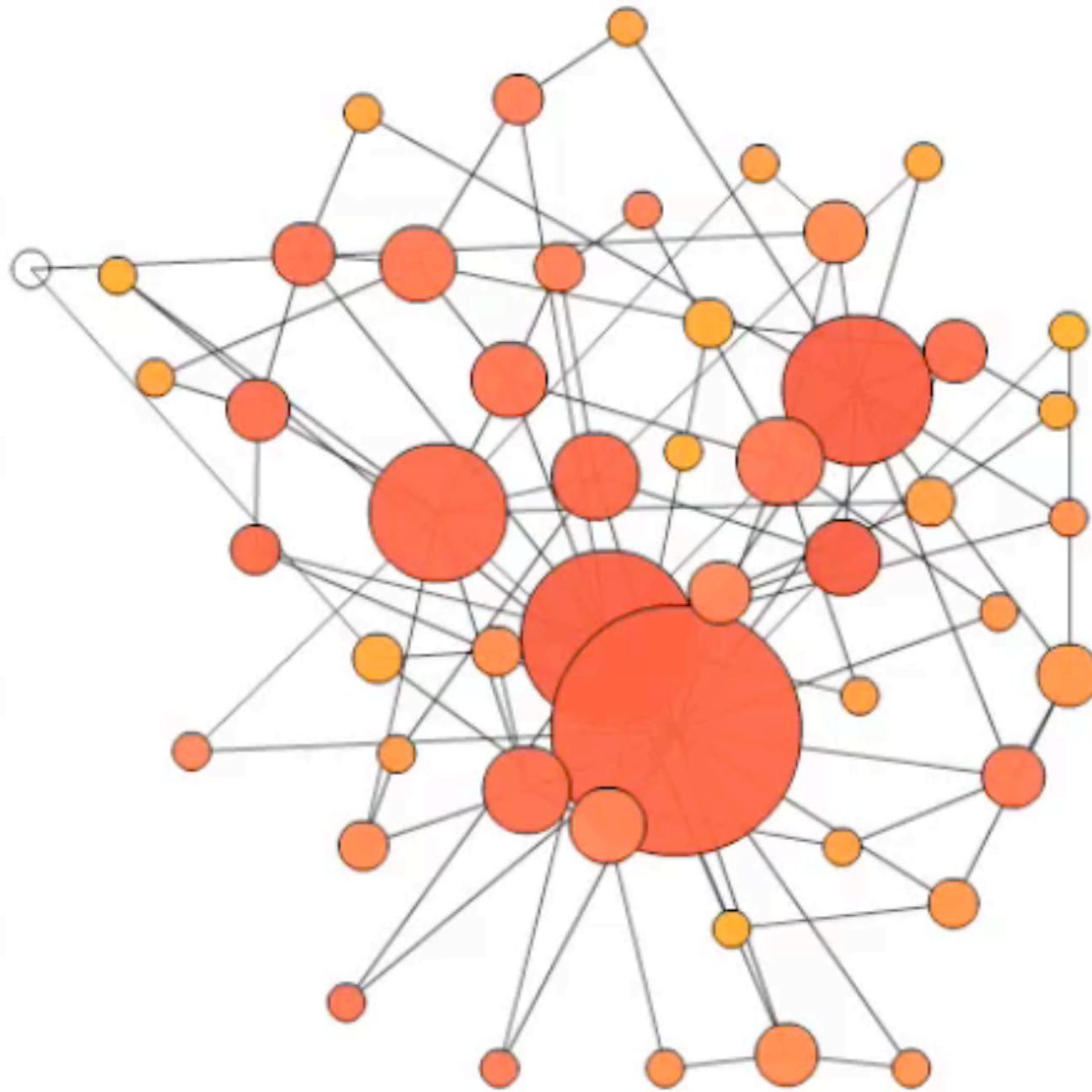




---

# Error

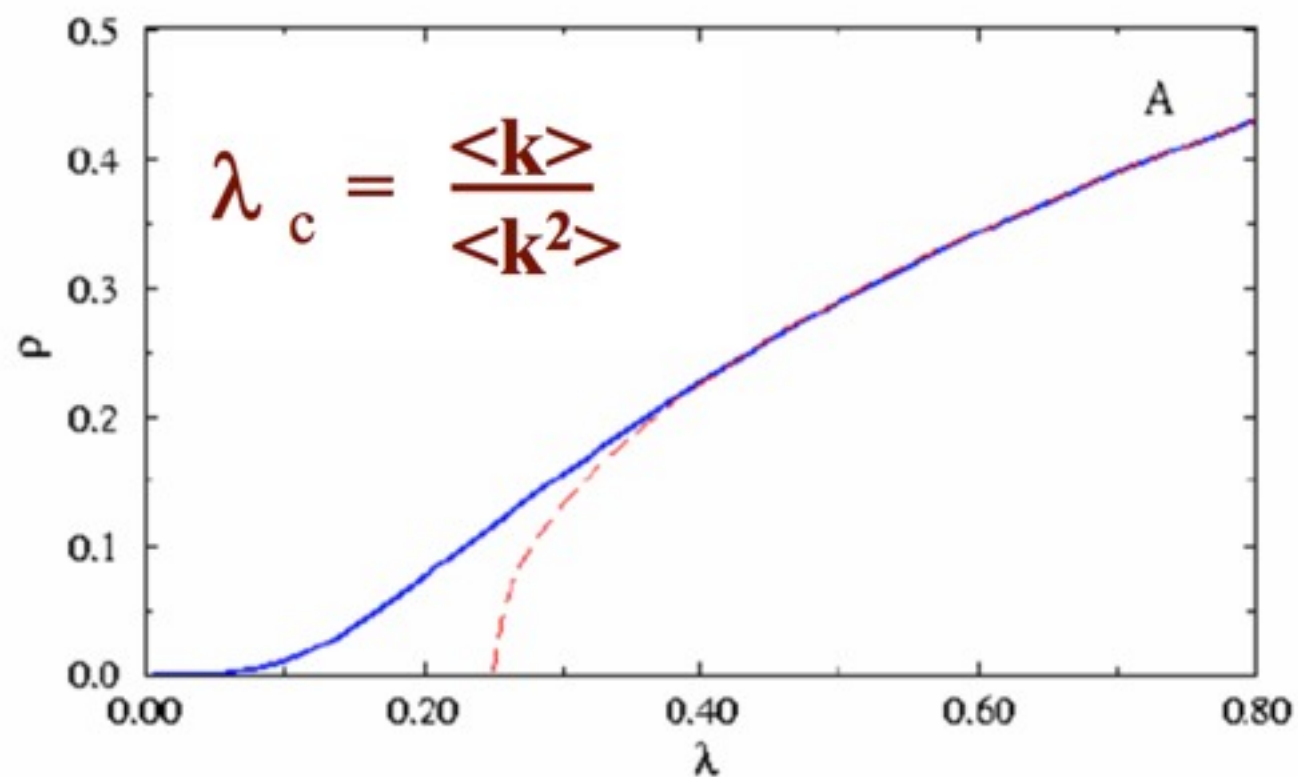
# Attack



# Attack



“We can’t block epidemic spreading on scale-free networks”



Alessandro Vespignani

# Epidemic spreading: “following links”

# Epidemic spreading: “following links”



“Friendship Paradox”



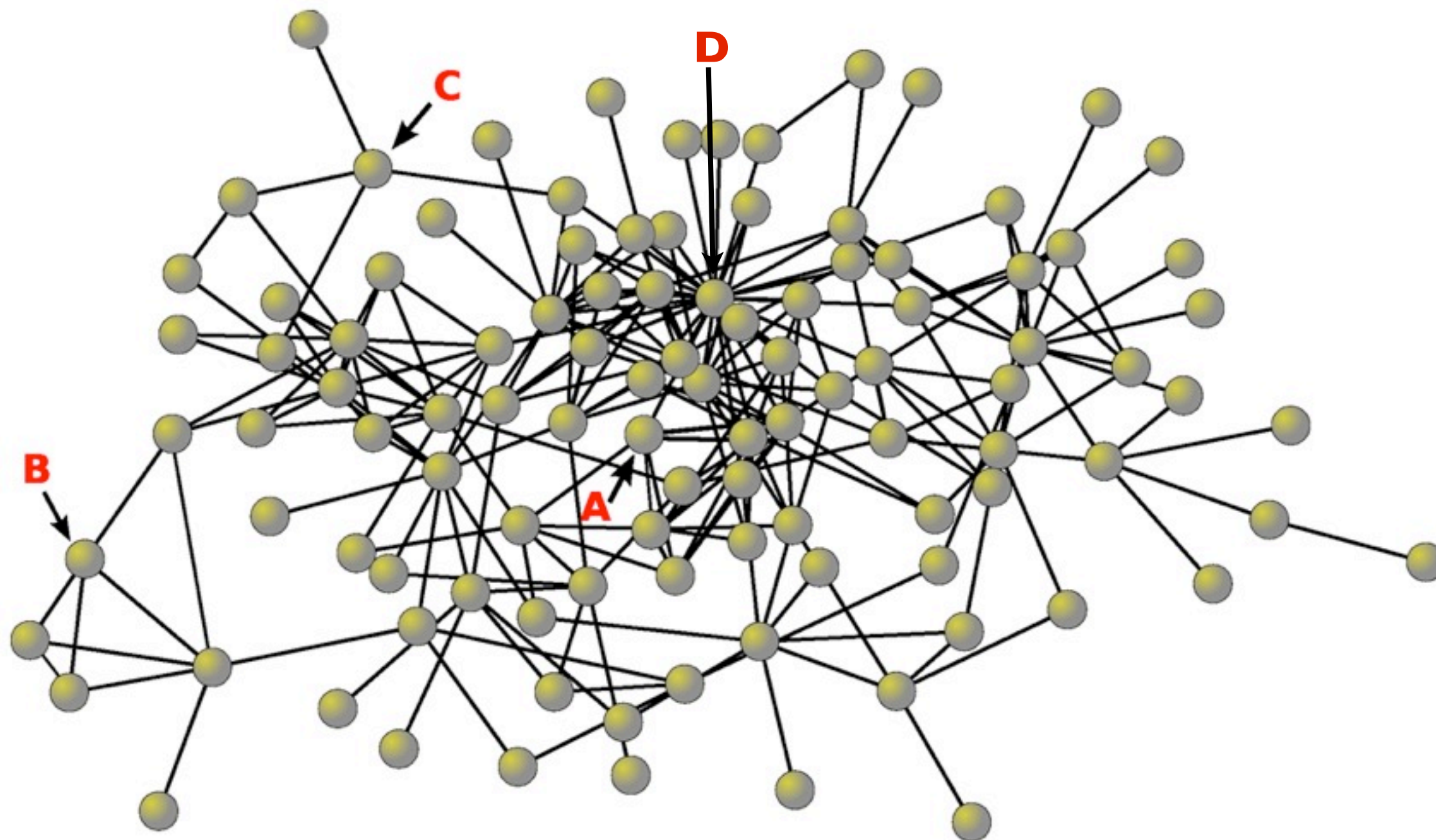
# Epidemic spreading: “following links”



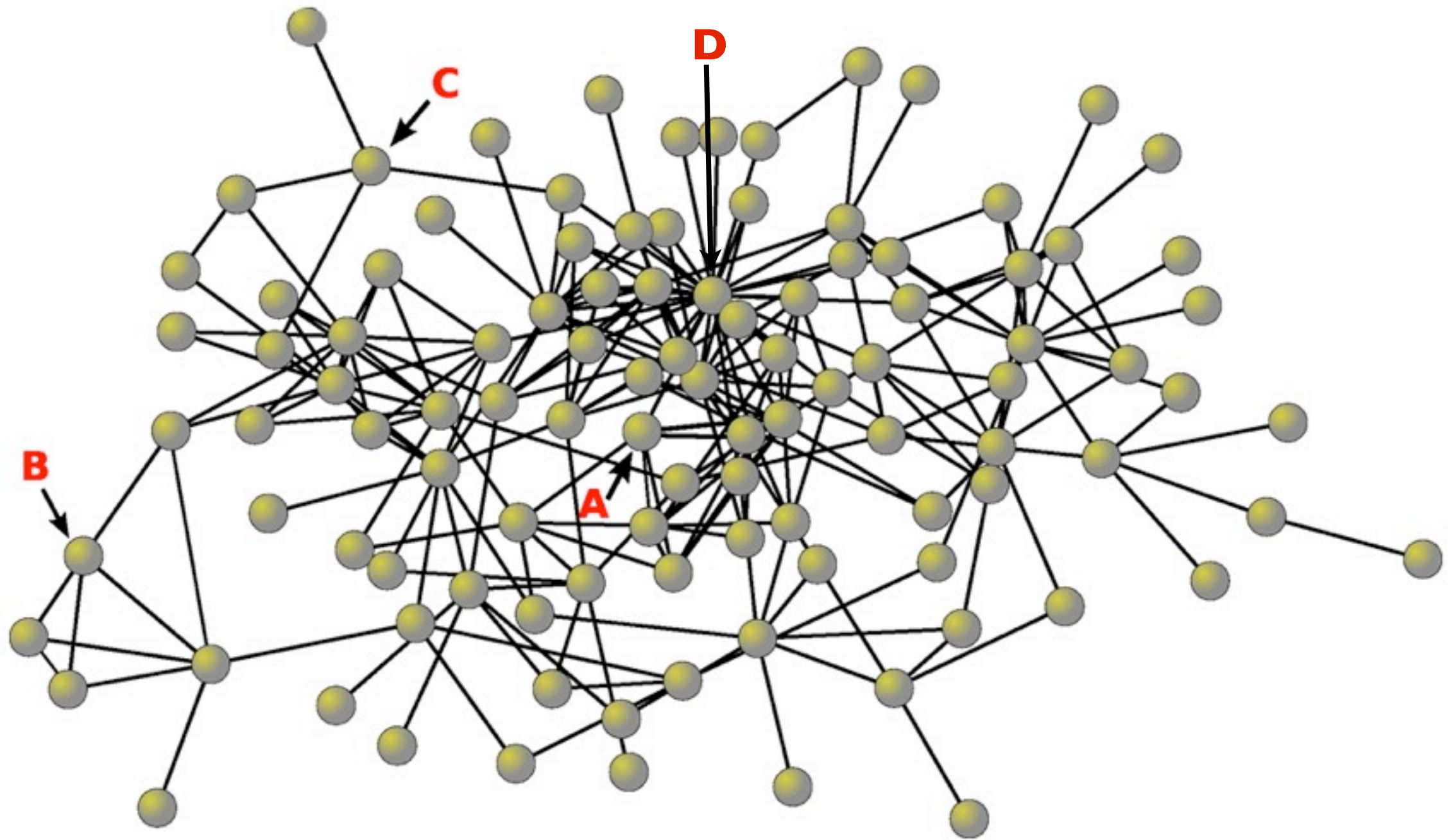
“Friendship Paradox”

The disease quickly get to  
the **hubs**

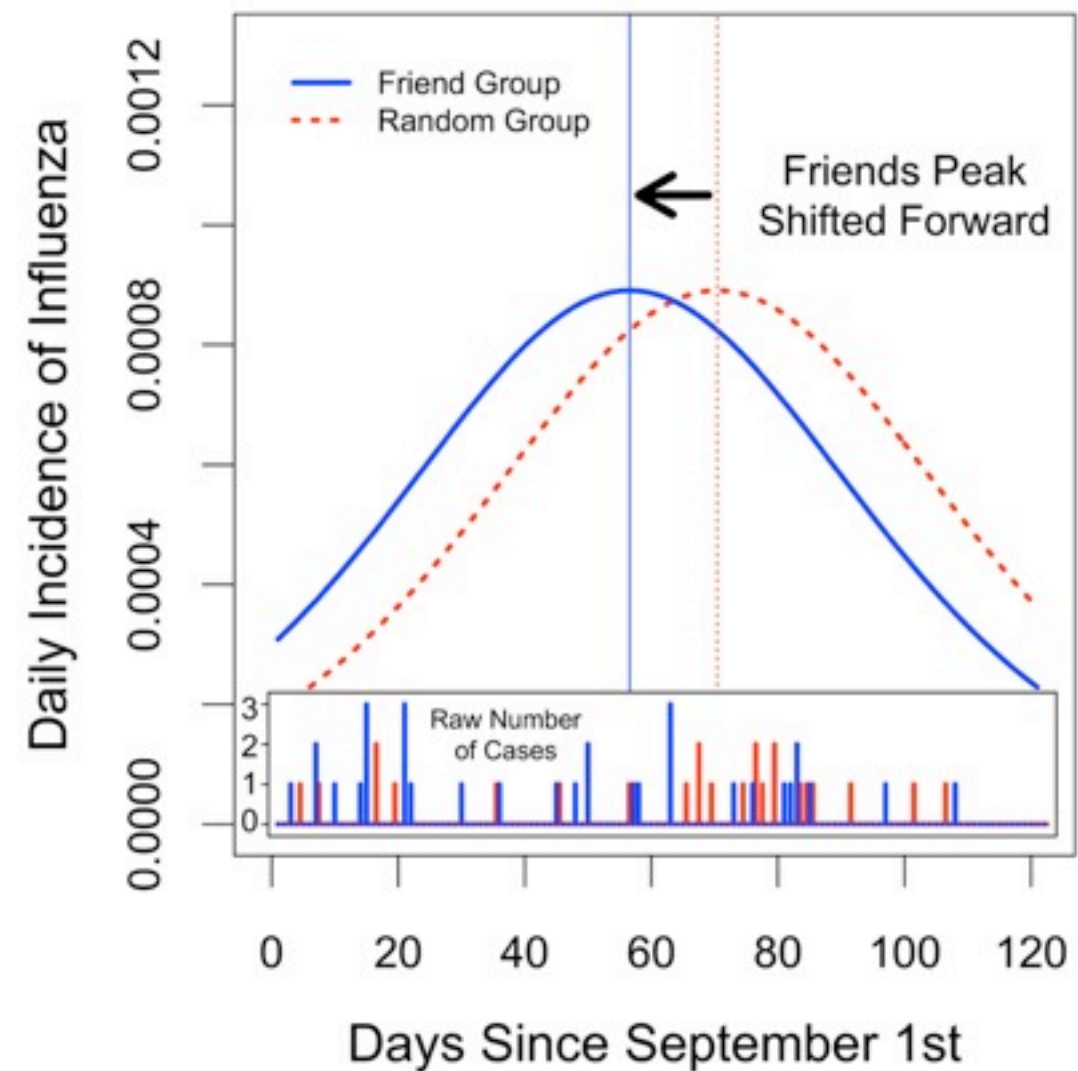
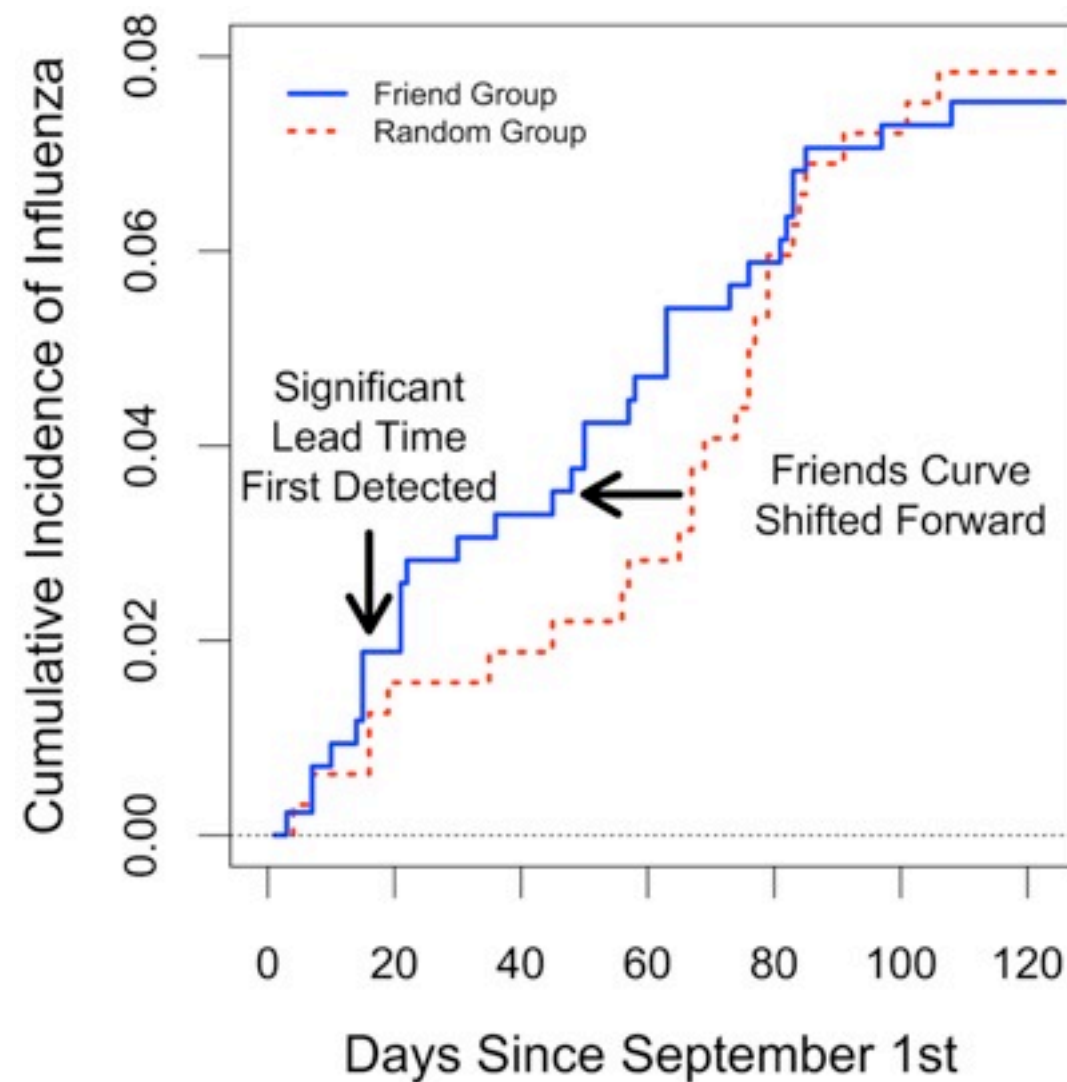
How to effectively  
**detect & prevent** the  
disease spreading?







**“Hubs”**



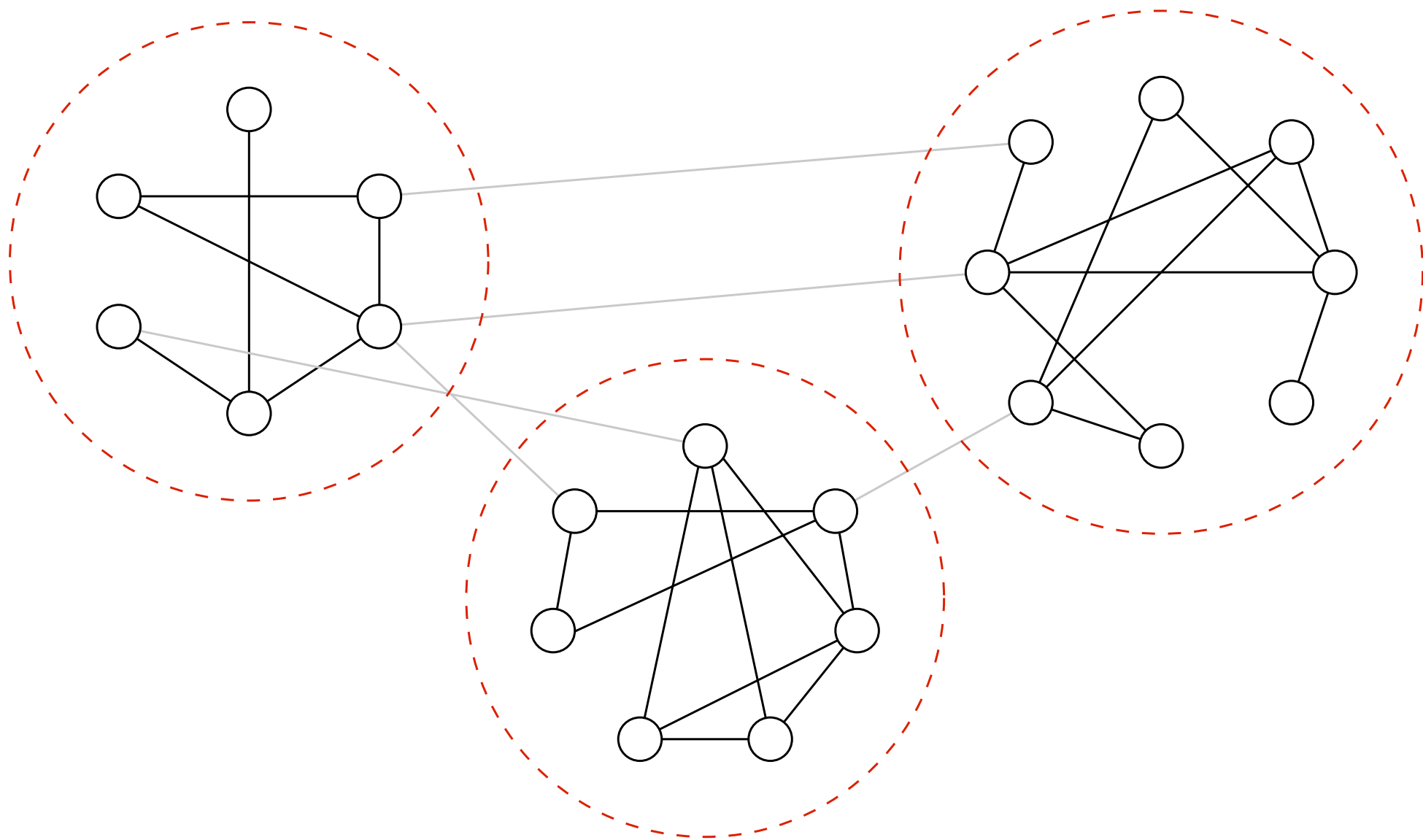
Random person -> immunize a random friend of the person (but **not the original person!**)

# Communities

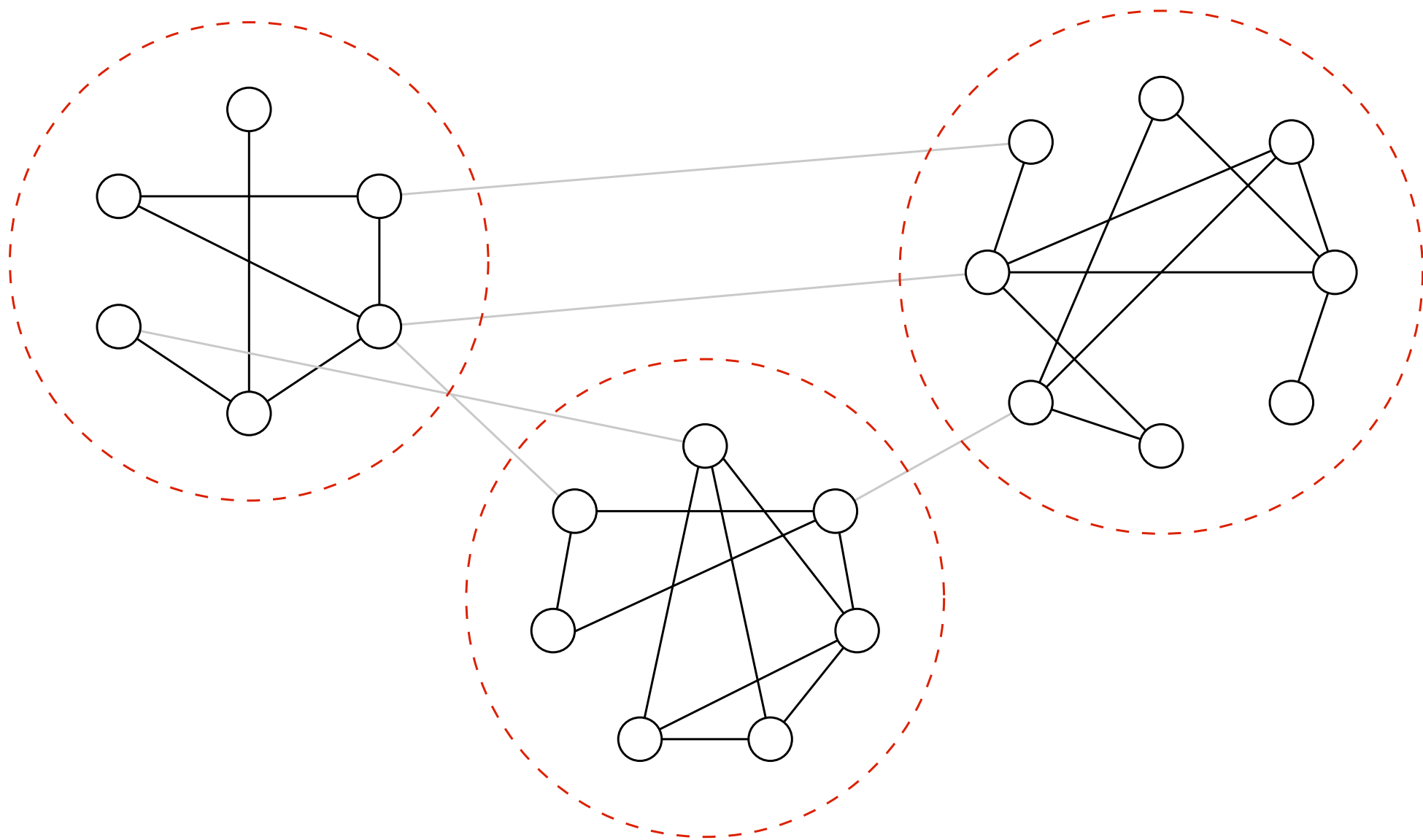


Networks are not just  
**clustered**, but form  
**communities**

“a group of densely interconnected nodes”



“a group of densely interconnected nodes”



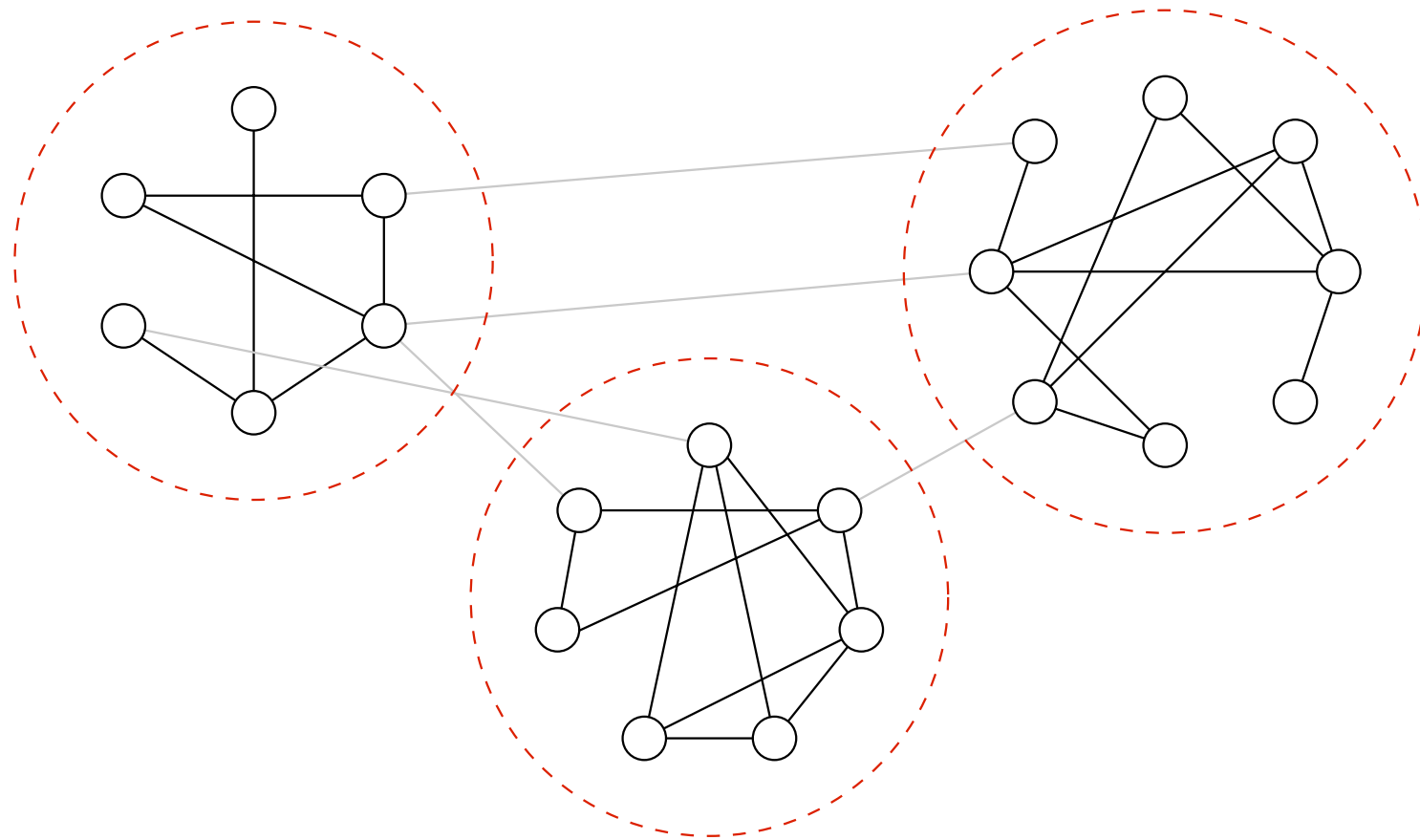


# How to find **communities?**

# 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

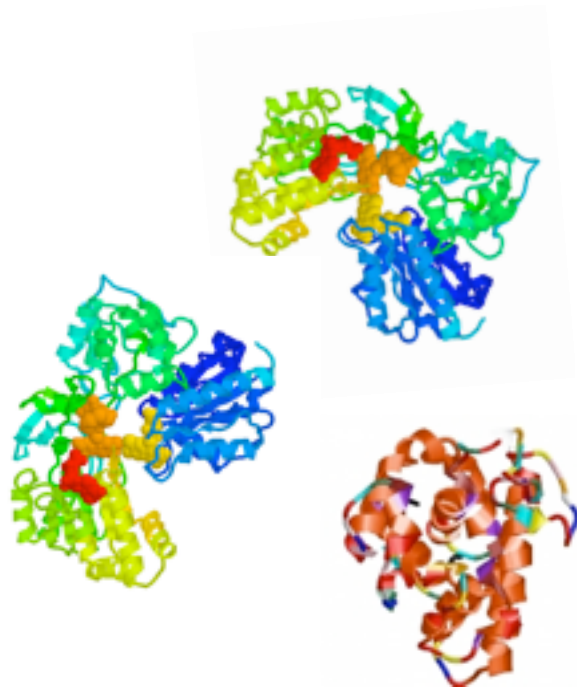


Then, why bother?

# Modular Structure







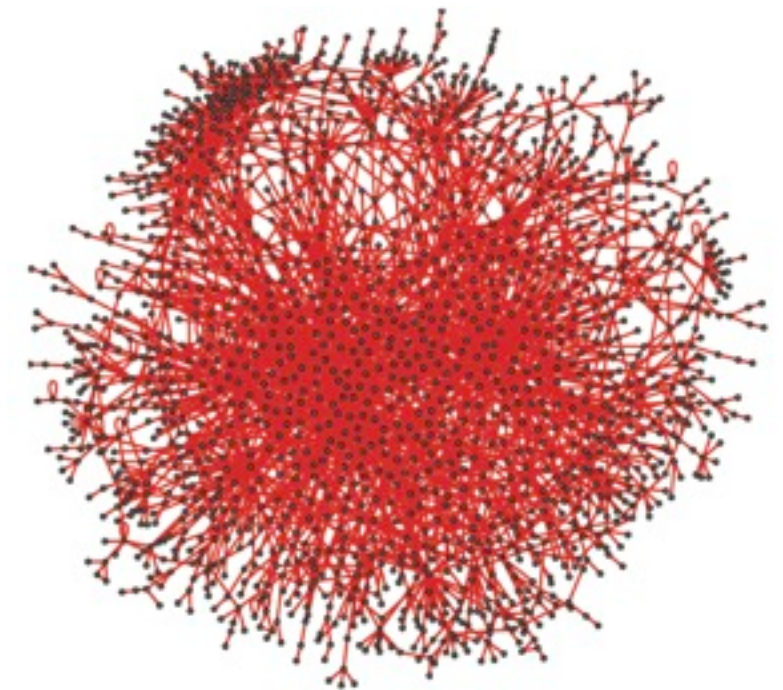
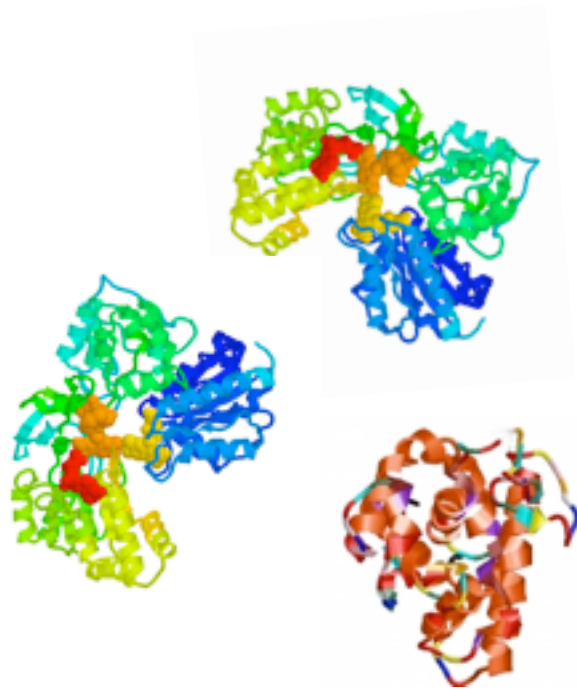
**Local**

**Global**

motifs

degree

clustering

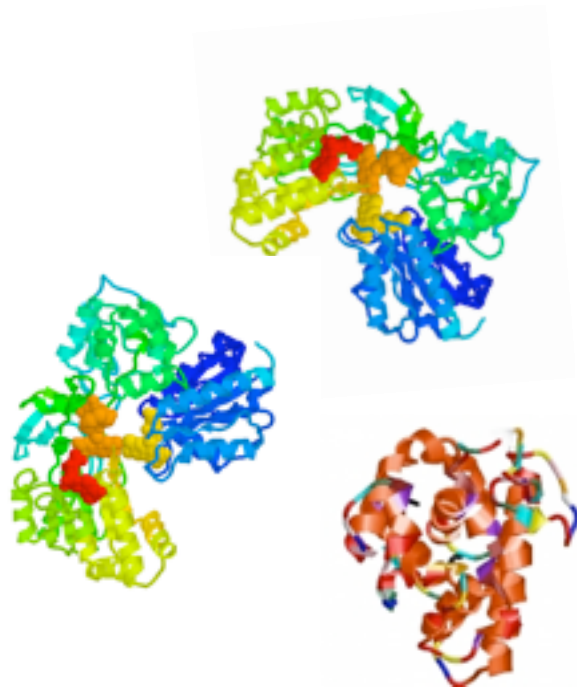


**Local**

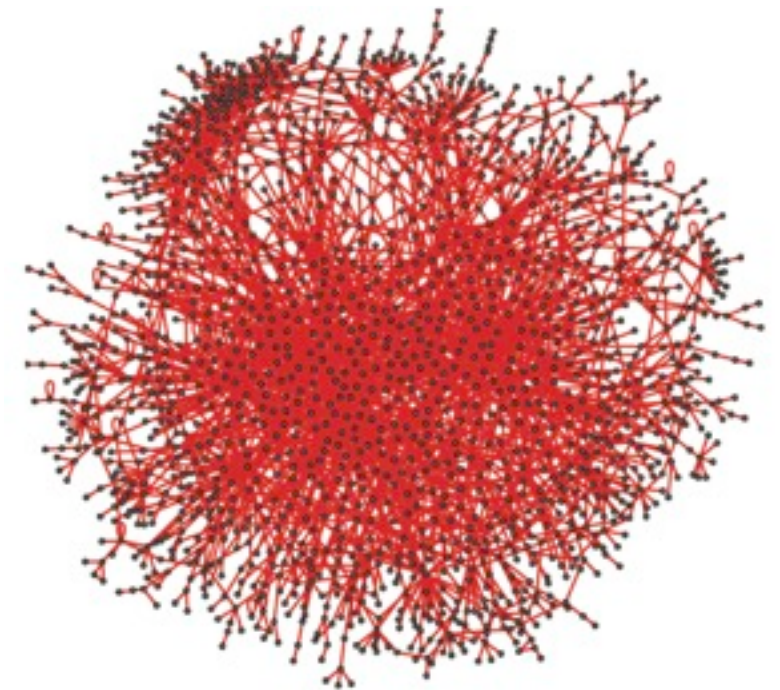
**Global**

motifs  
degree  
clustering

degree  
distribution  
Robustness



?



**Local**

**Global**

motifs  
degree  
clustering

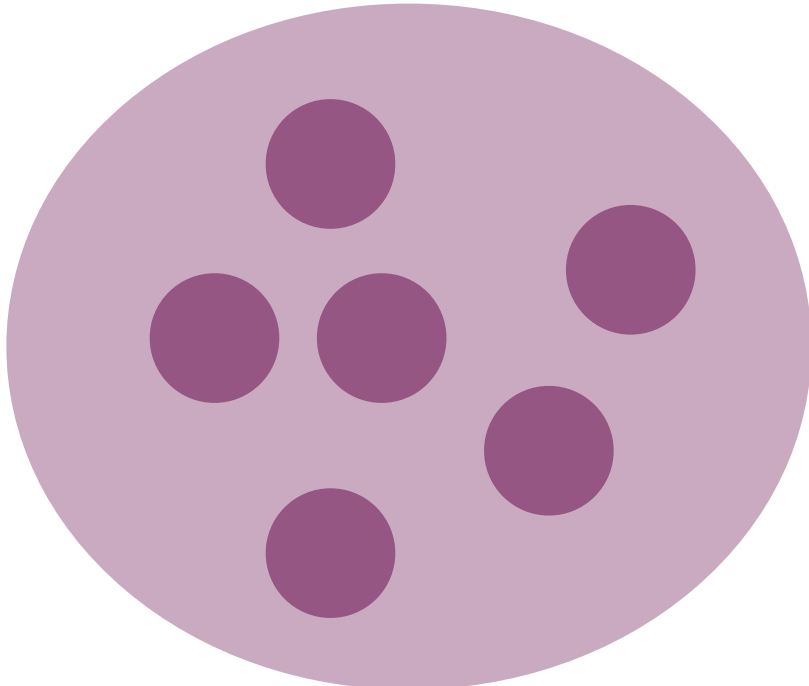
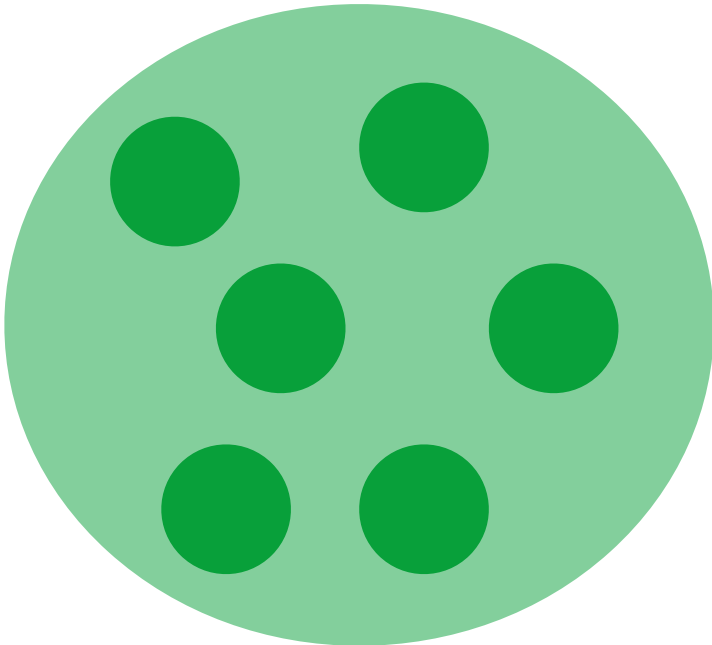
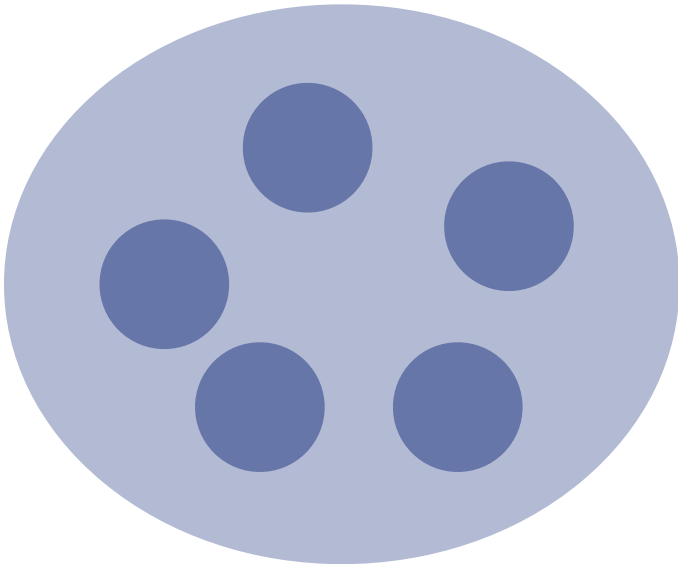
degree  
distribution  
Robustness

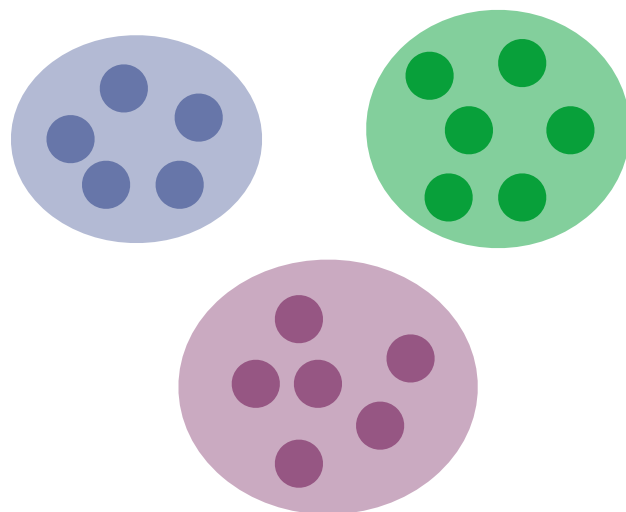
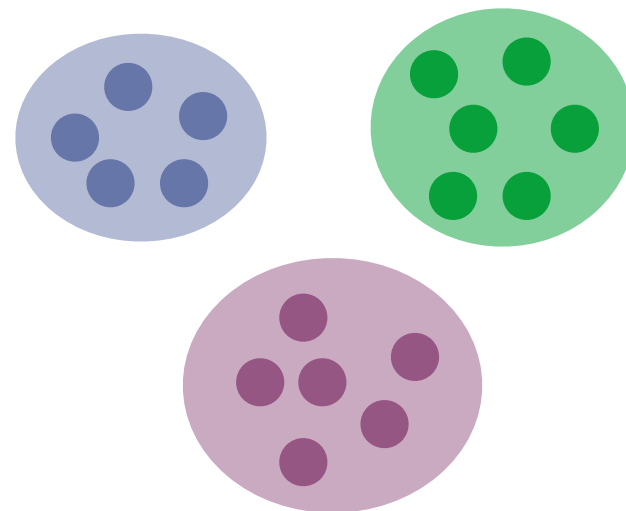
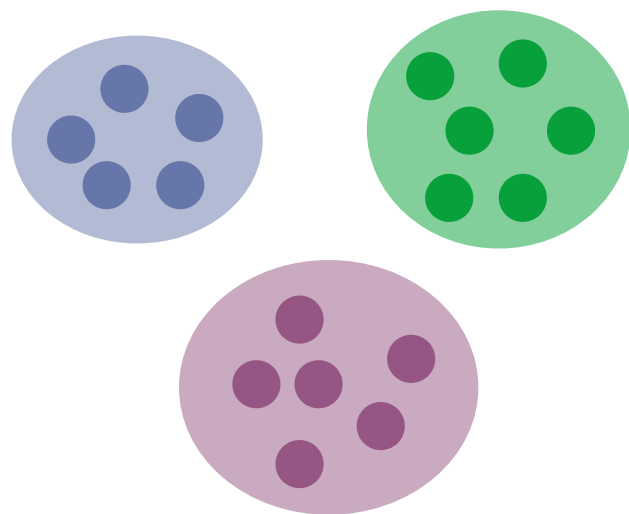
# Network Communities

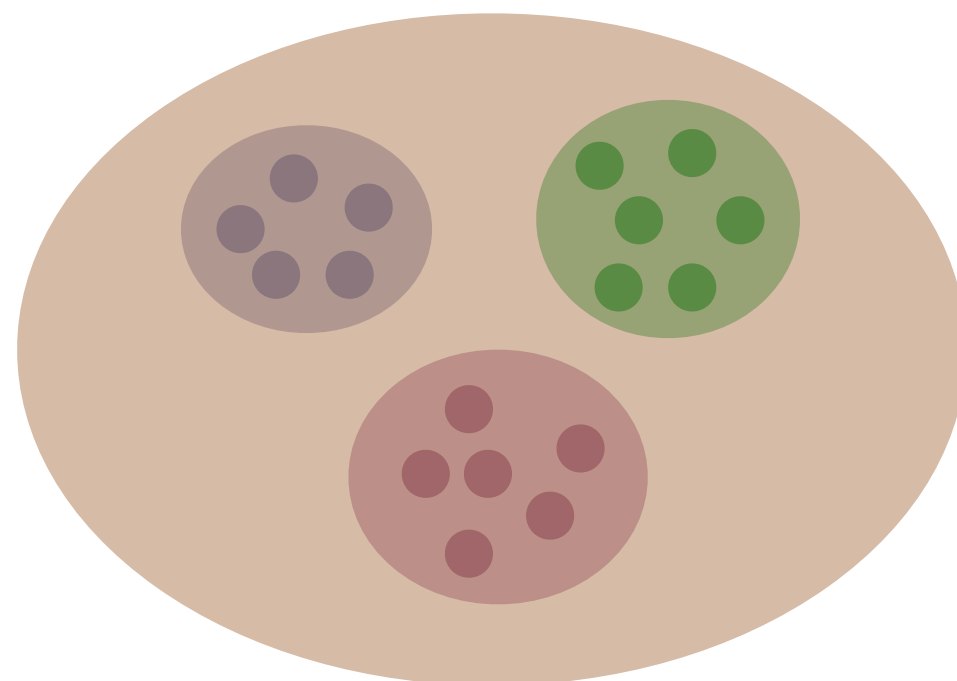
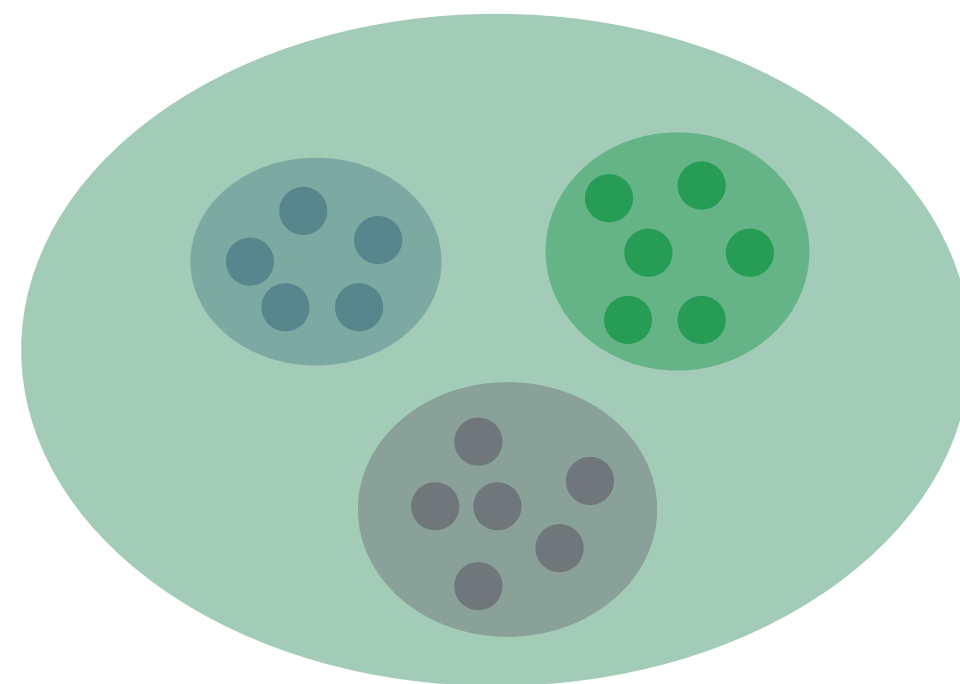
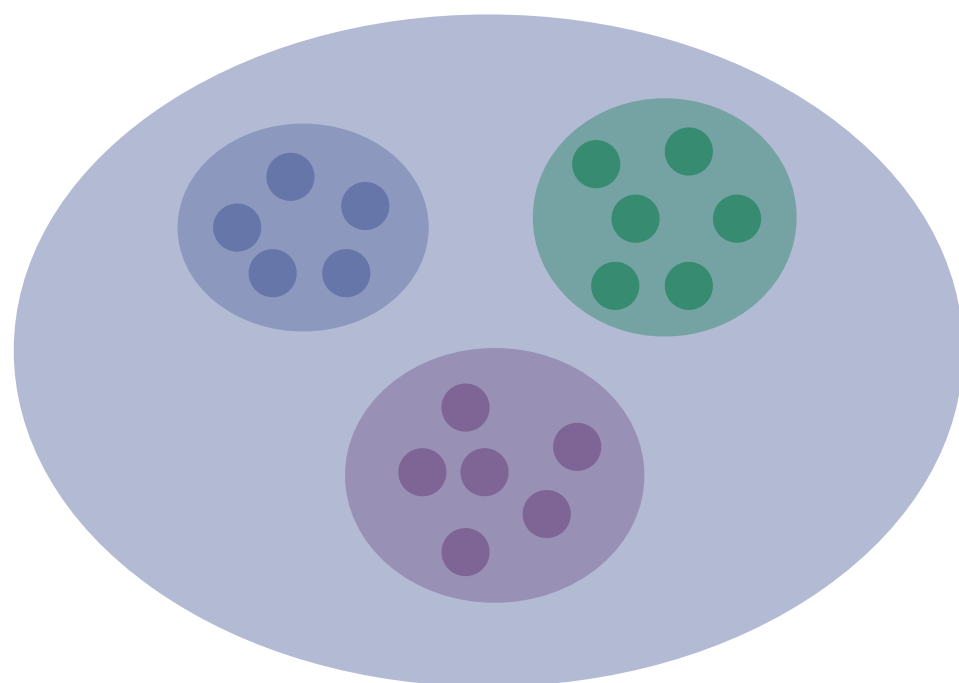




# Hierarchy



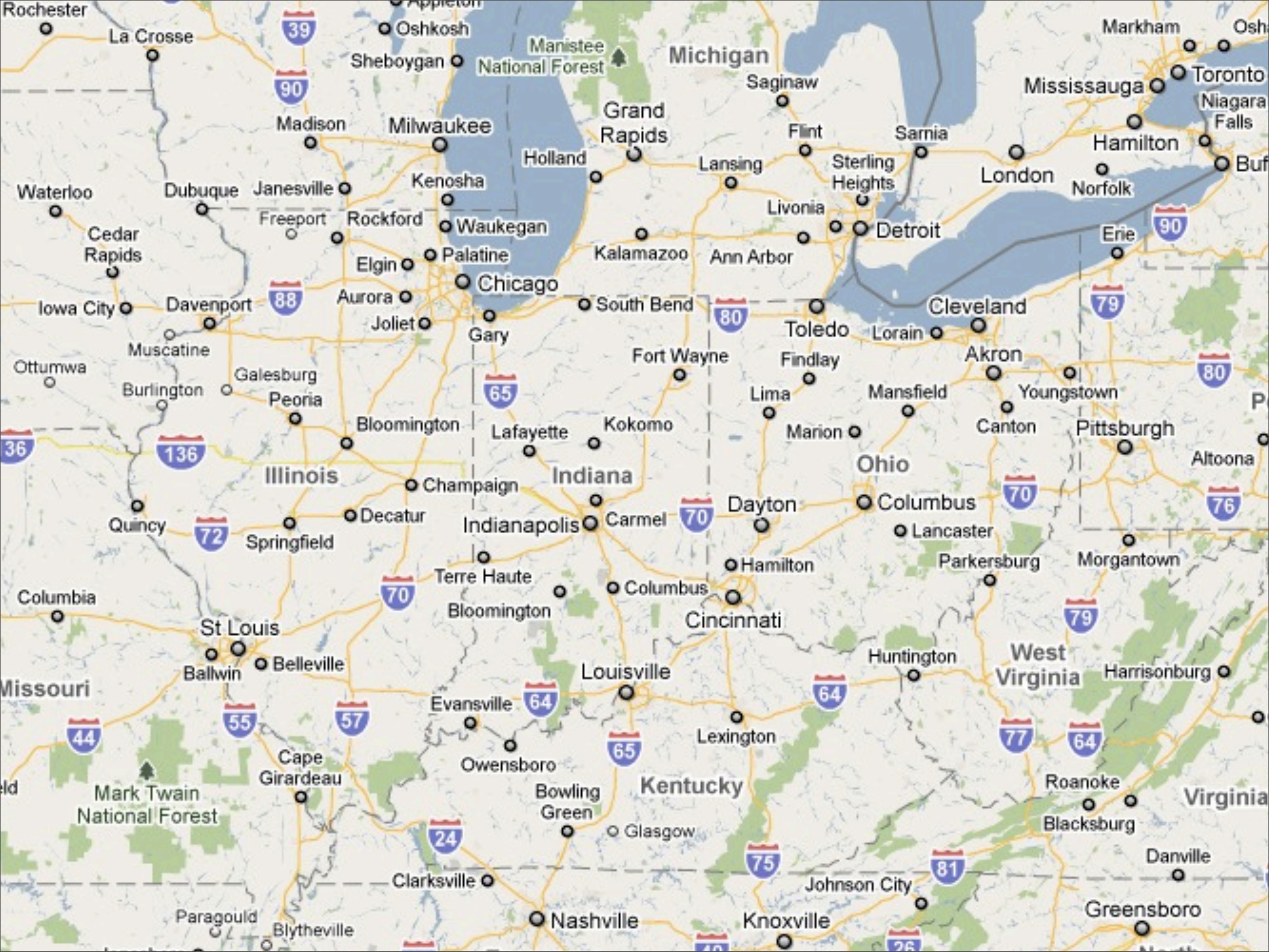




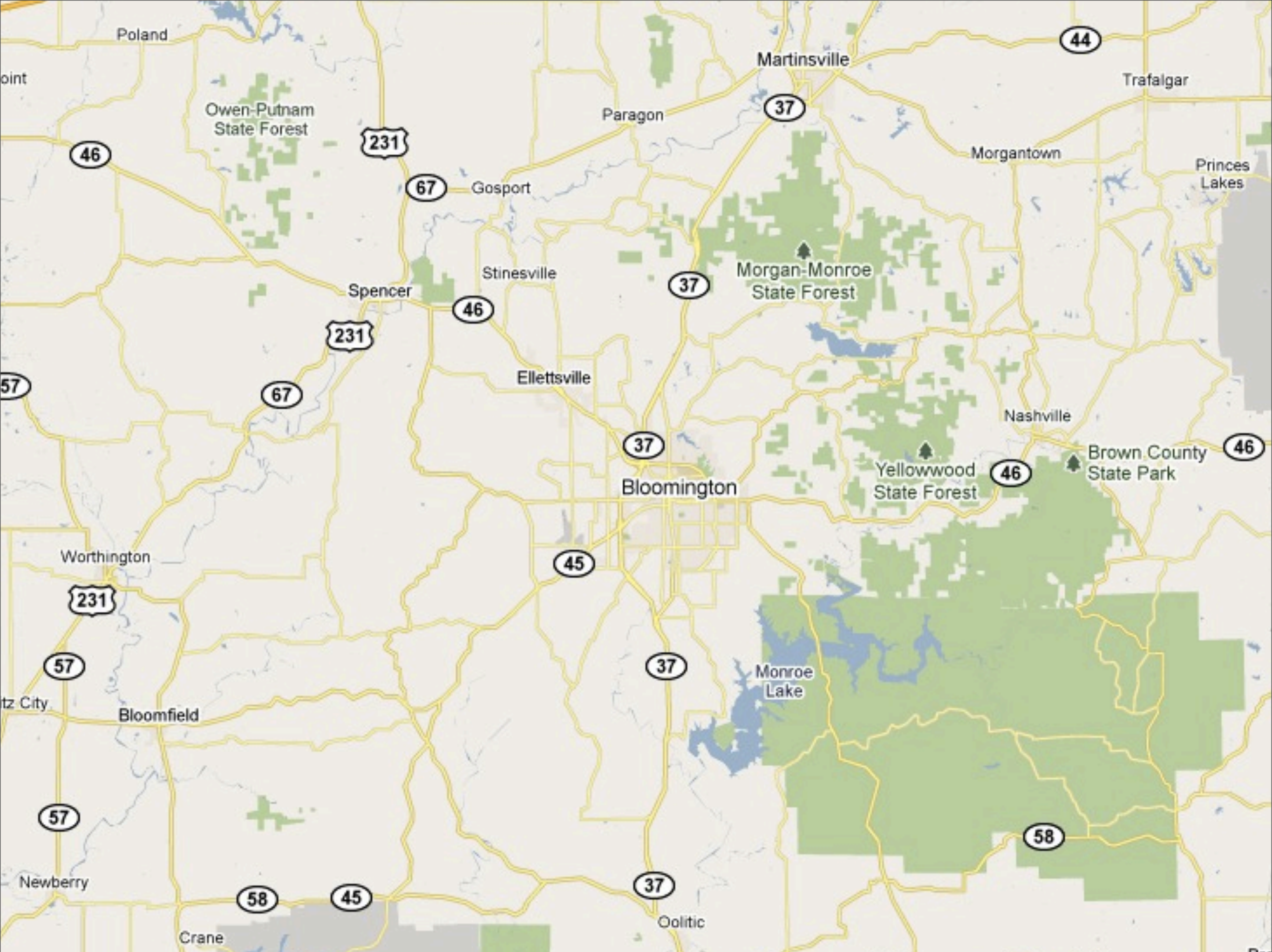








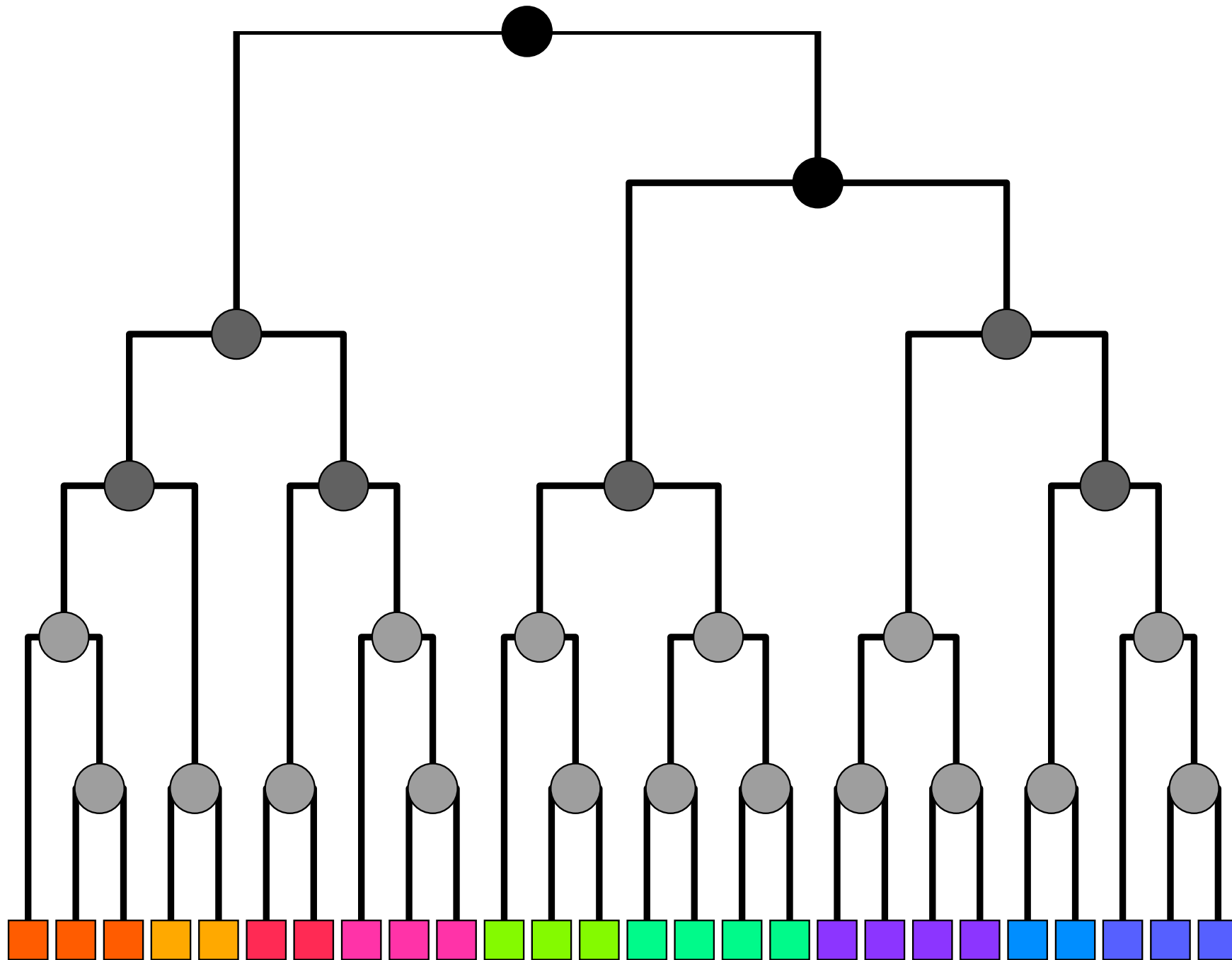




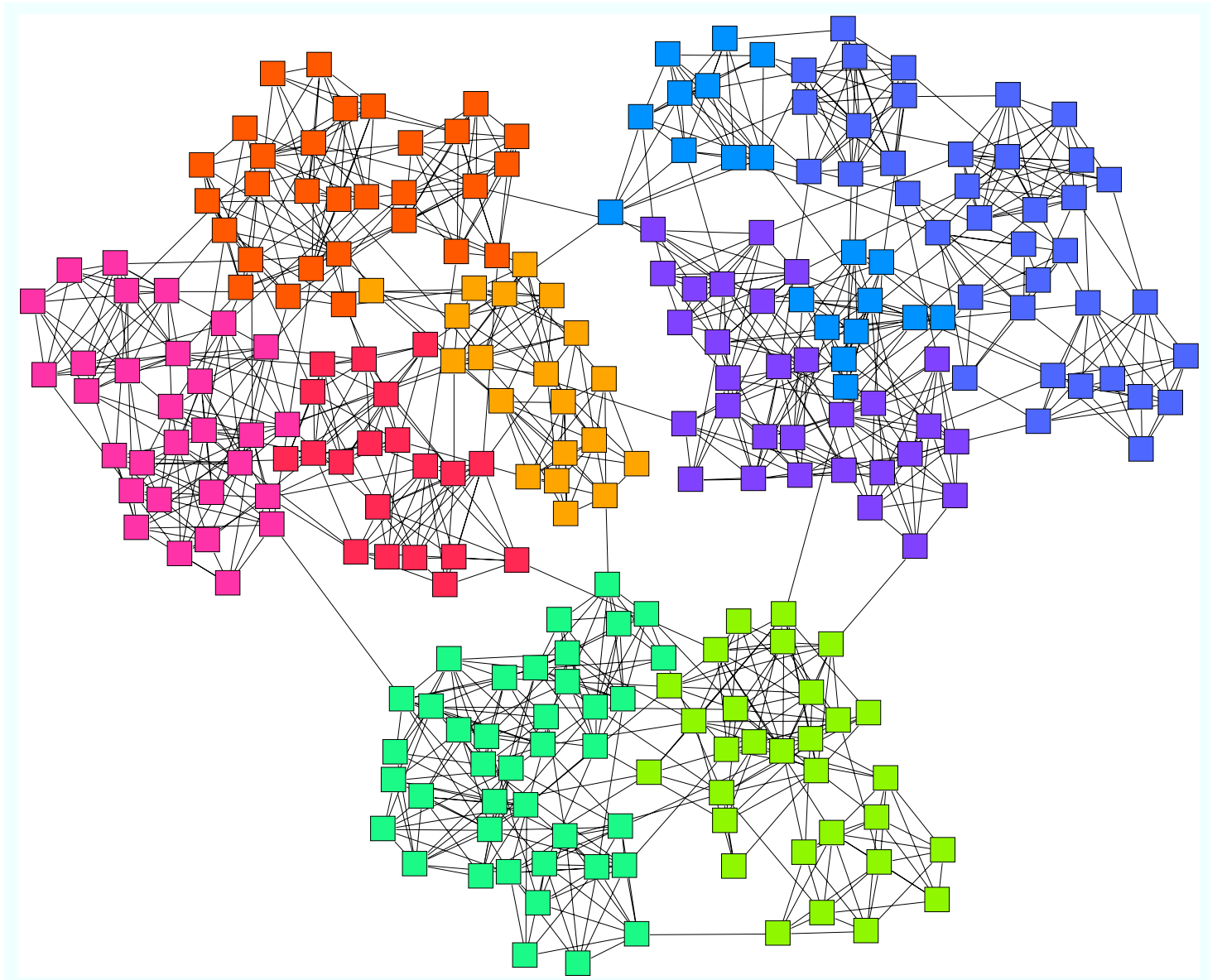
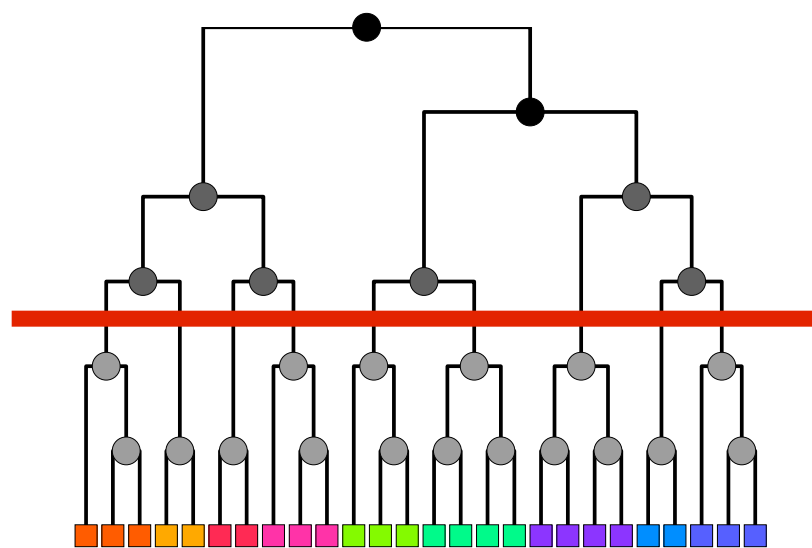
**Hierarchy implies  
communities.**



# 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)

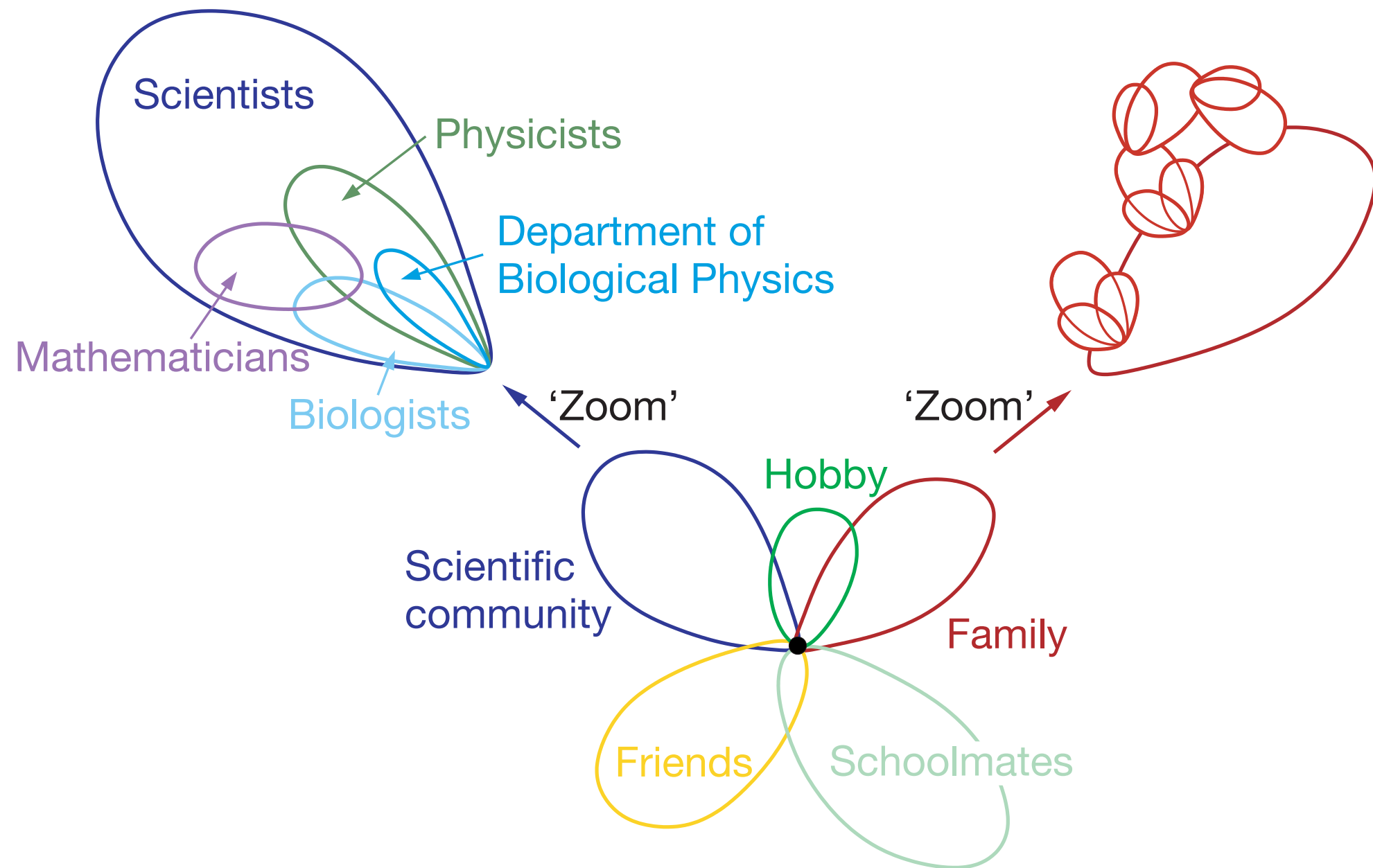
# Hierarchical community structure

Hierarchy → Communities



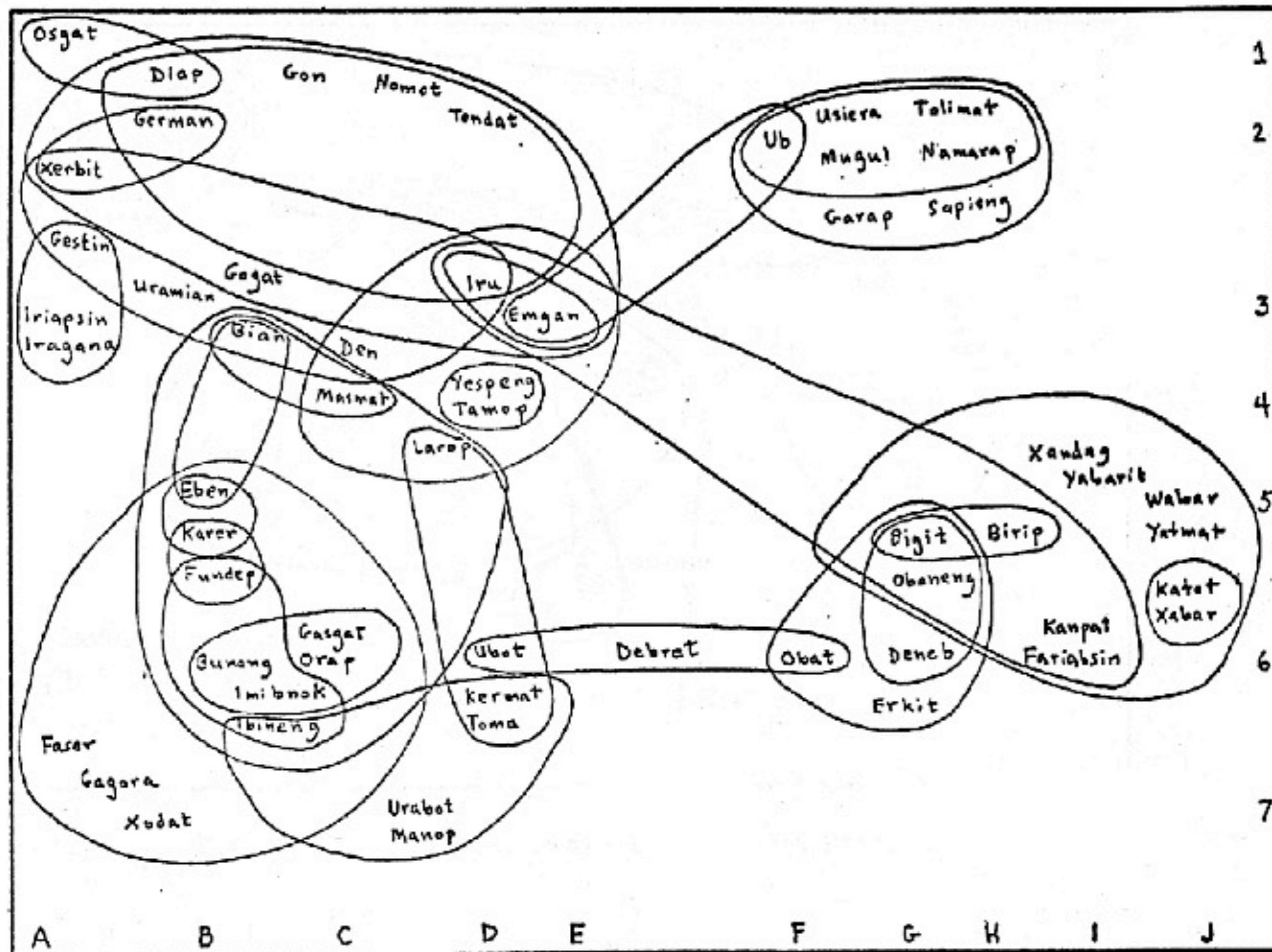


**BUT,**

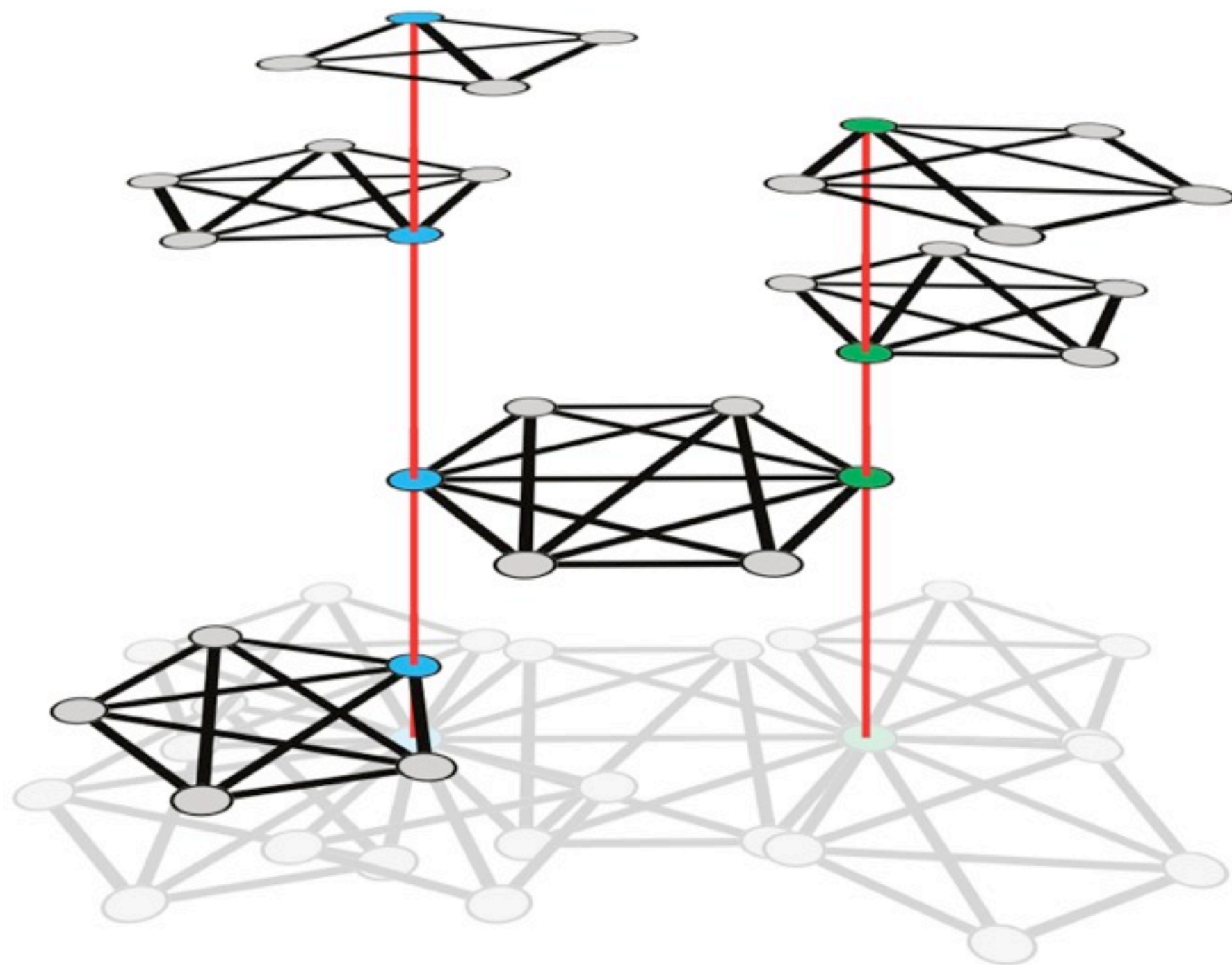


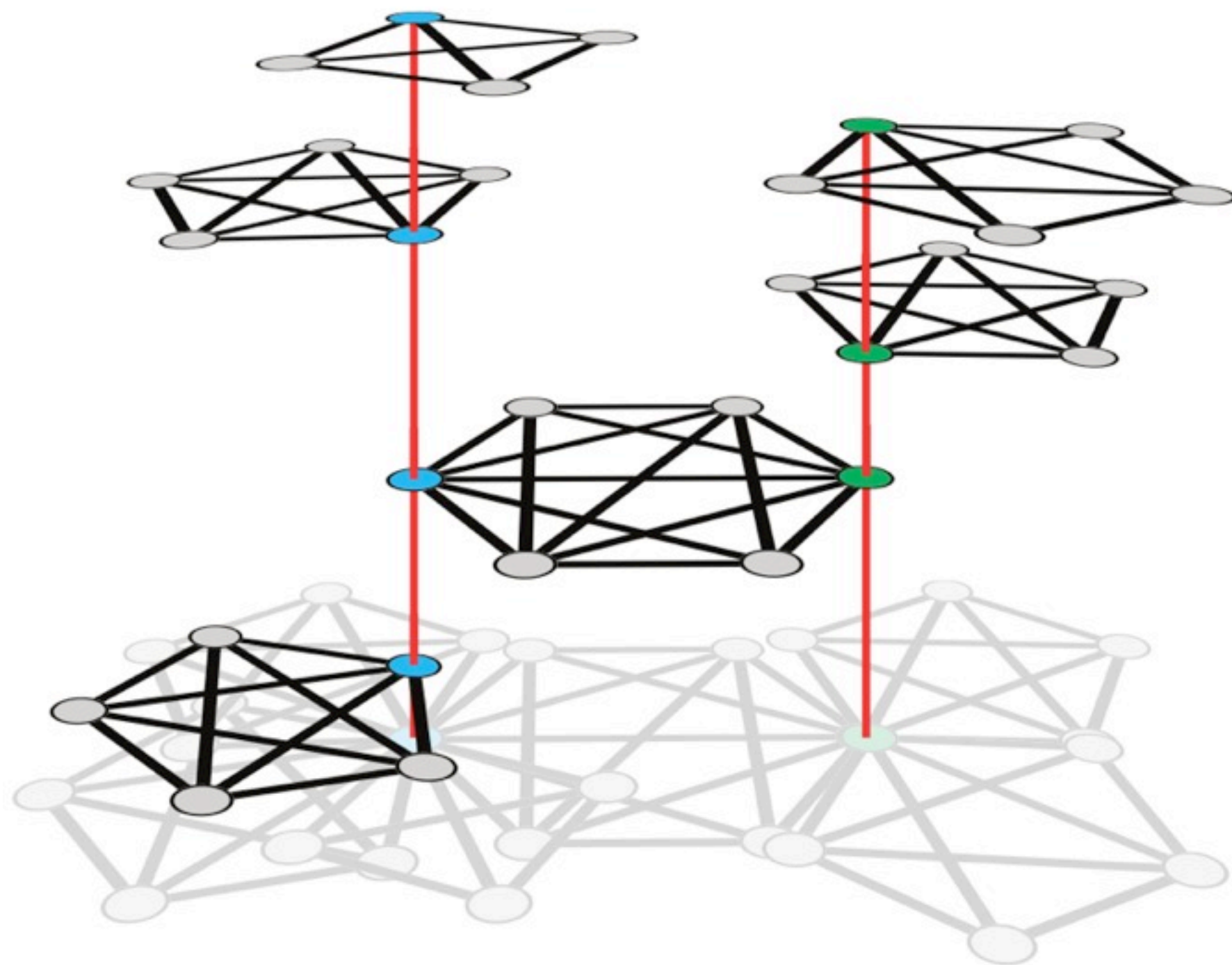
G. Palla, I. Derényi, I. Farkas & T. Vicsek, *Nature* (2005)



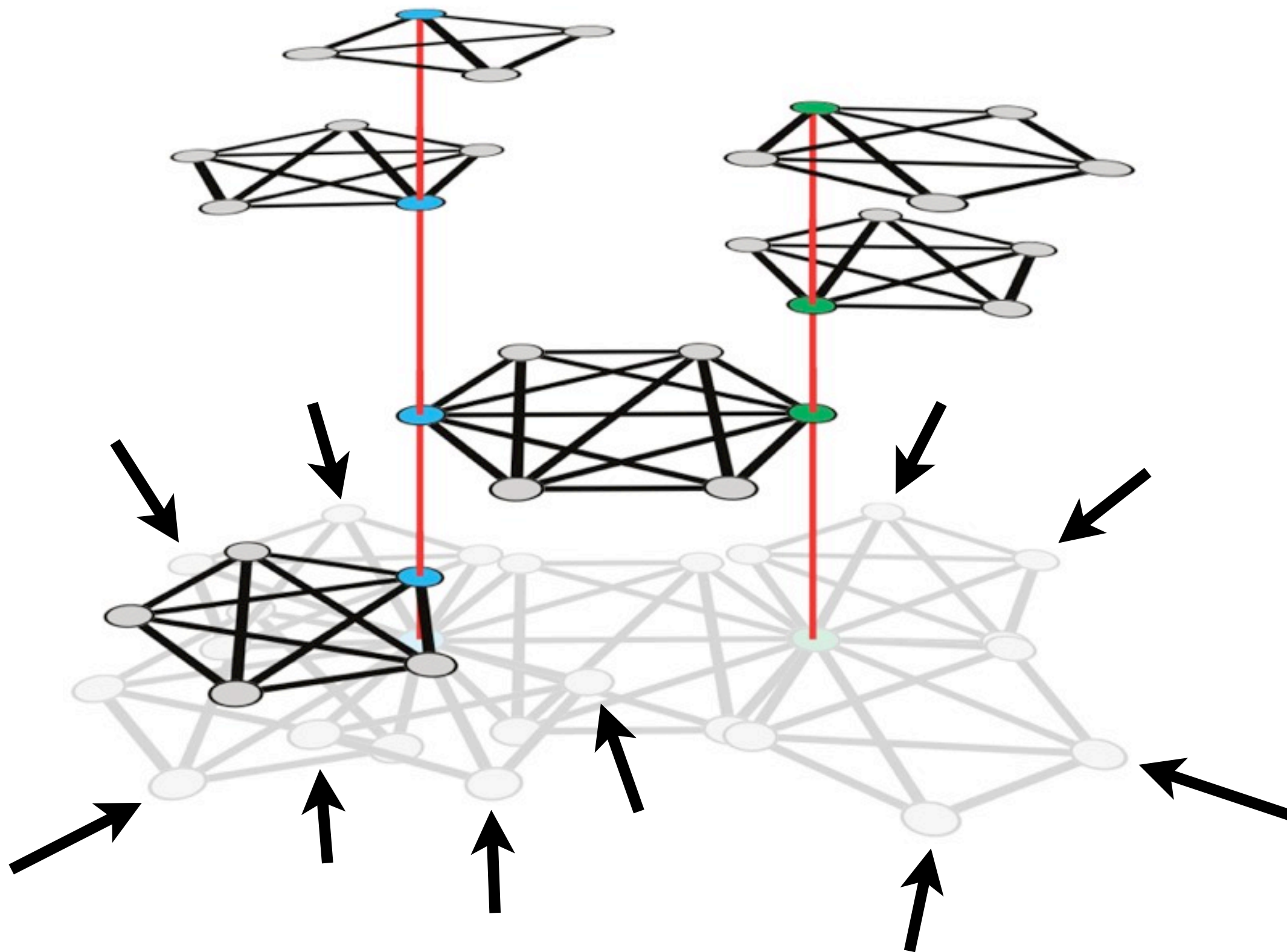


Arnold Perey, Social organization of *Oksapmin*, Papua New Guinea





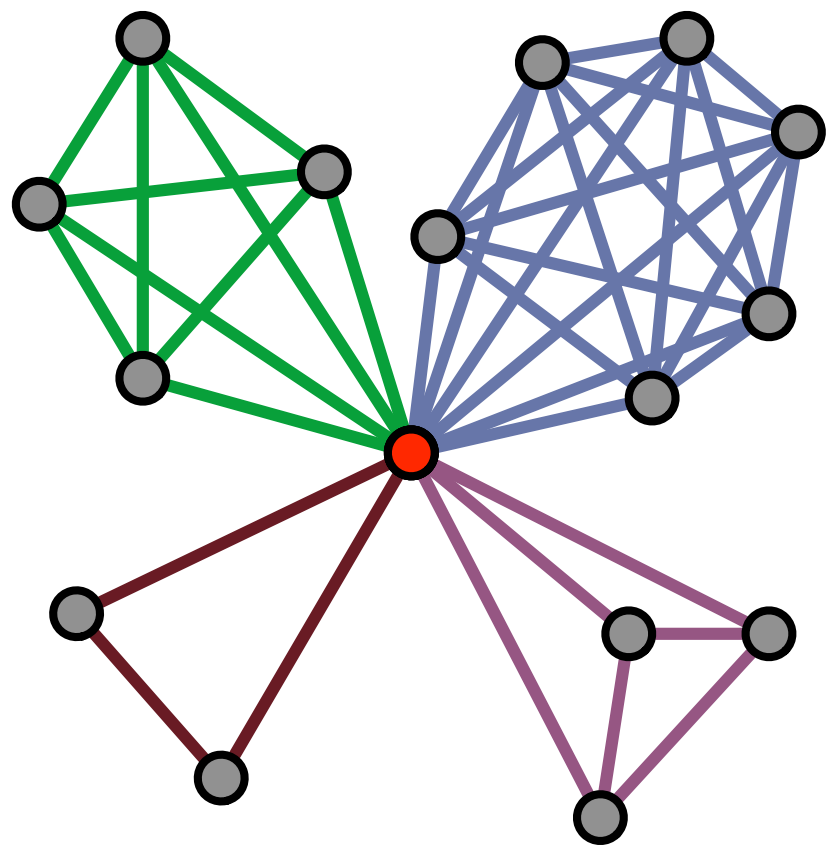




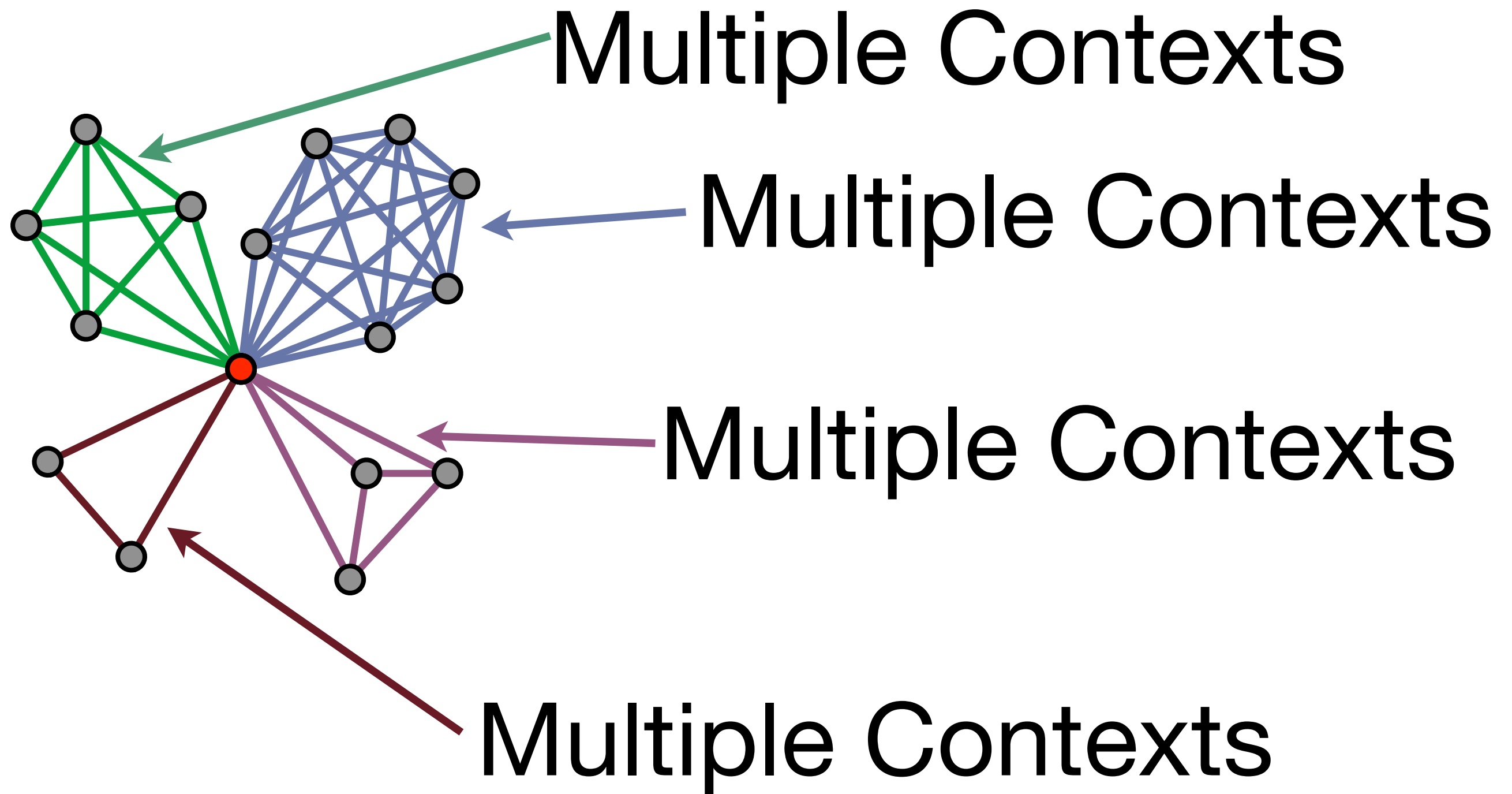
Overlap is  
pervasive.

Overlap is  
**pervasive.**

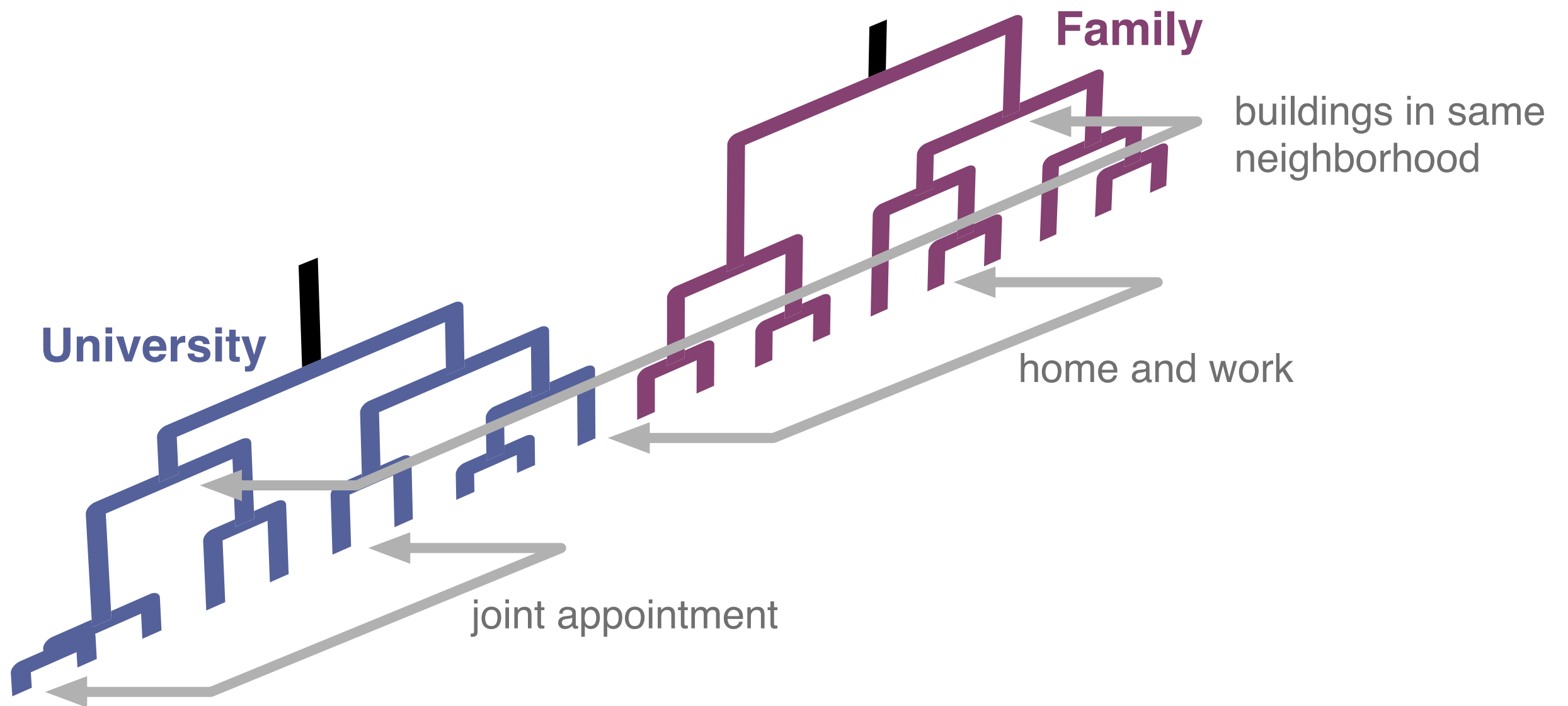




# Multiple Contexts



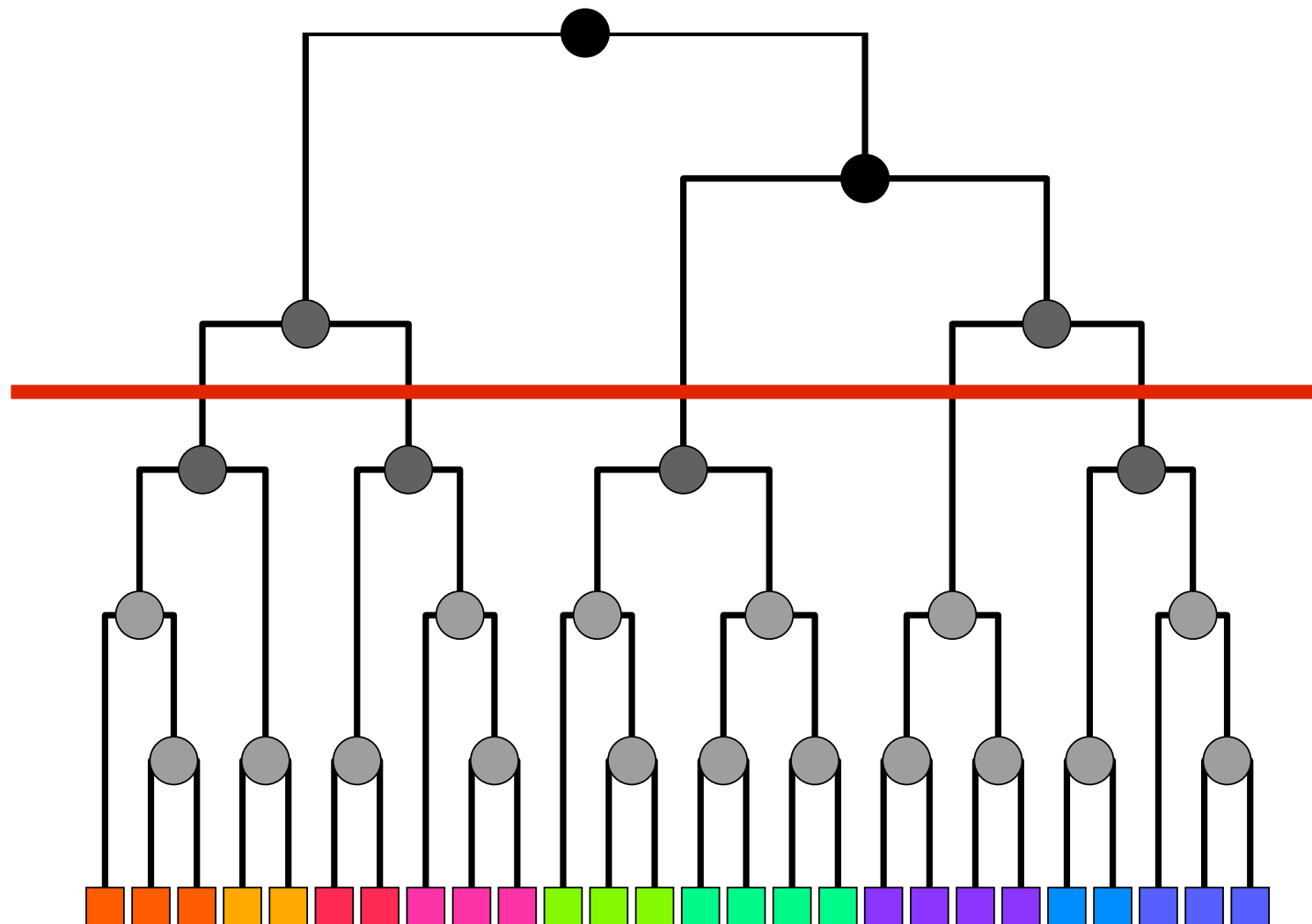
<http://www.youtube.com/watch?v=SxuYdzs4SS8>



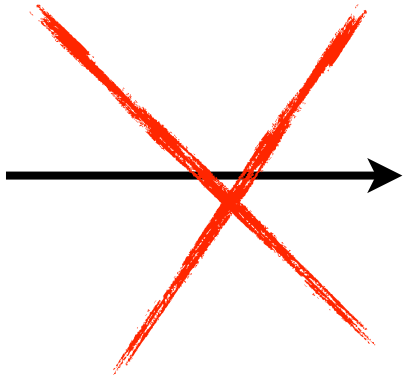
It is **impossible** to obtain a single dendrogram.



# Hierarchy implies disjoint communities.

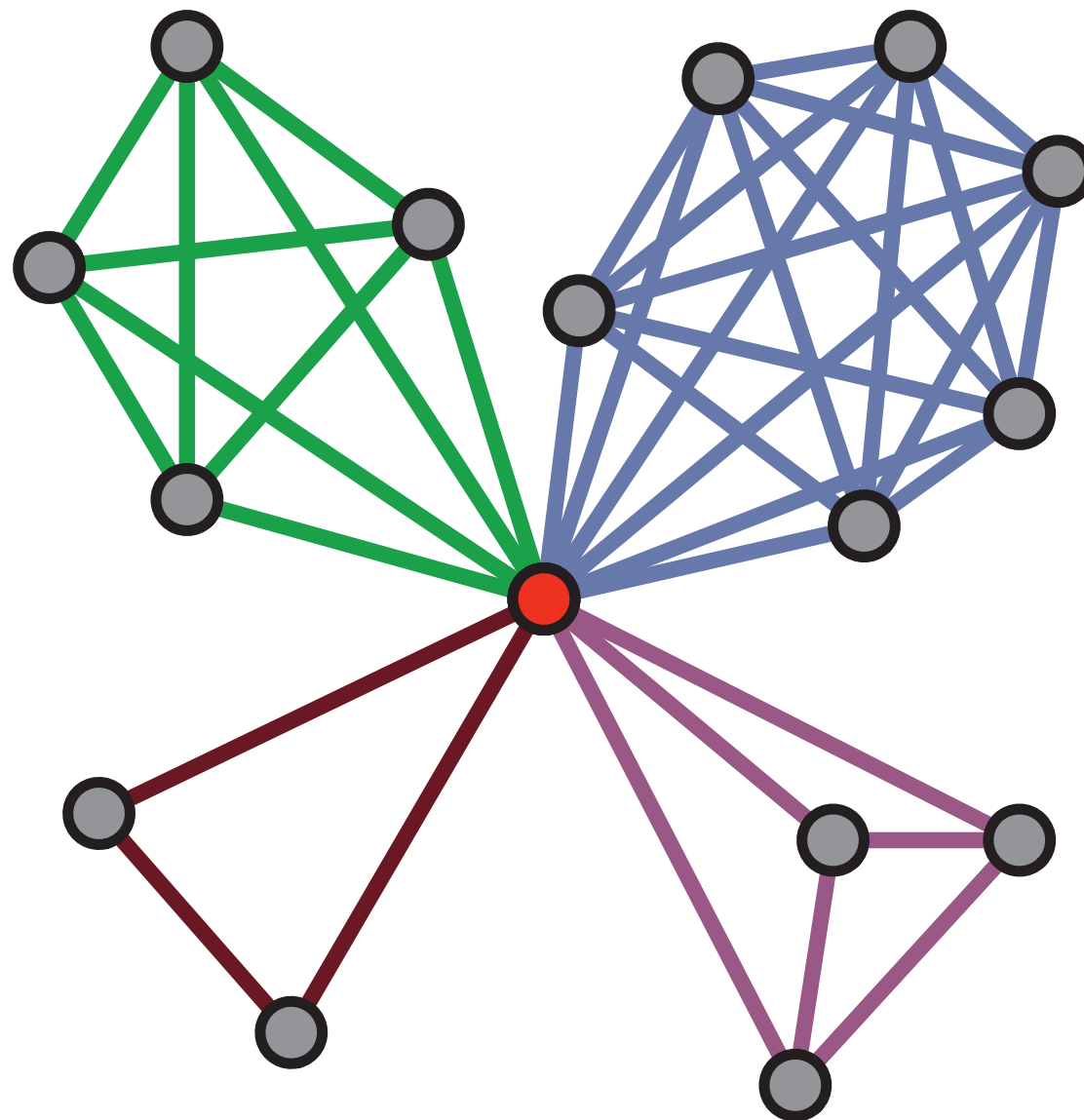


# Hierarchical community structure

Hierarchy  Communities

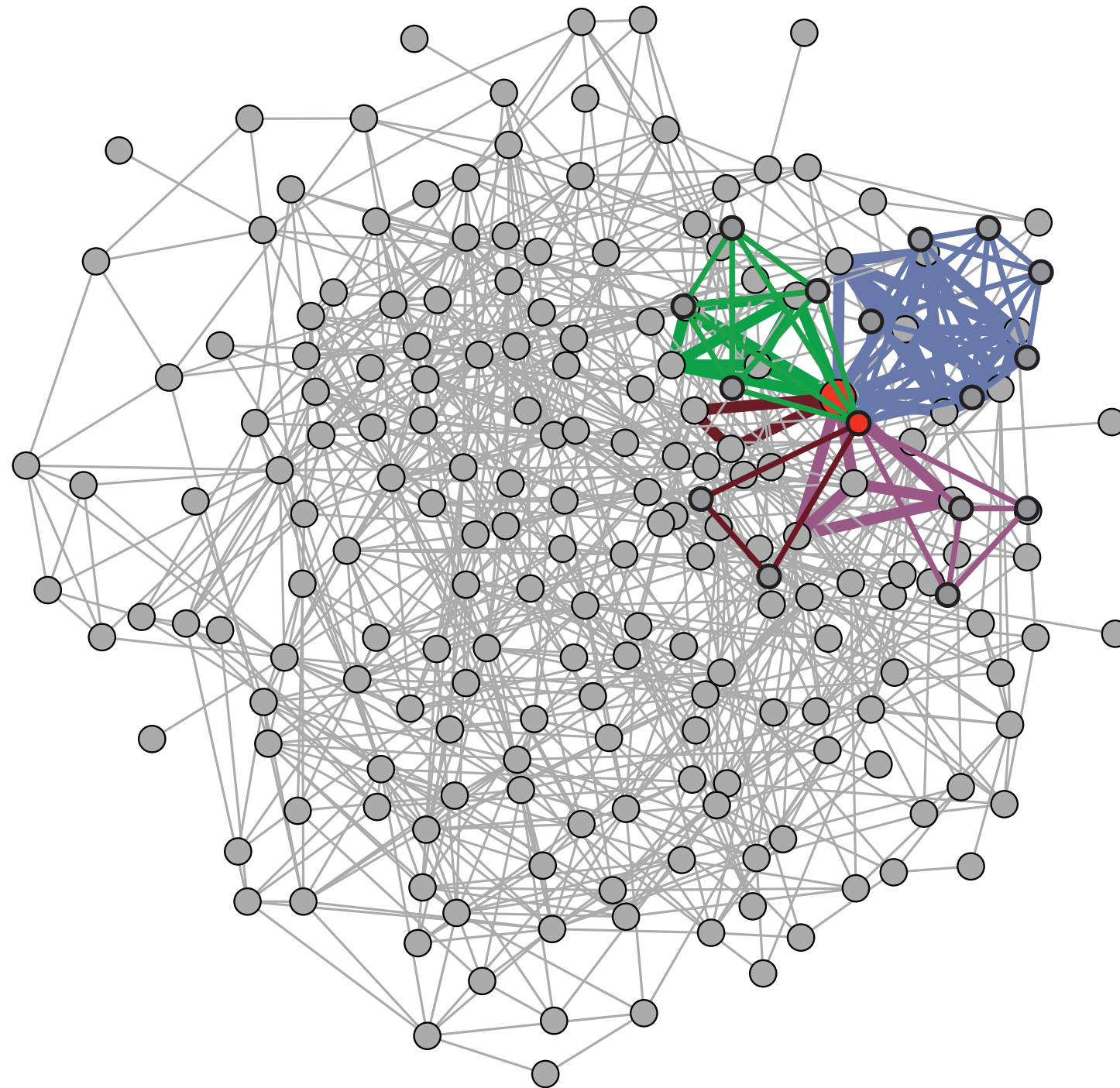
# Another consequence

# Simple local structure

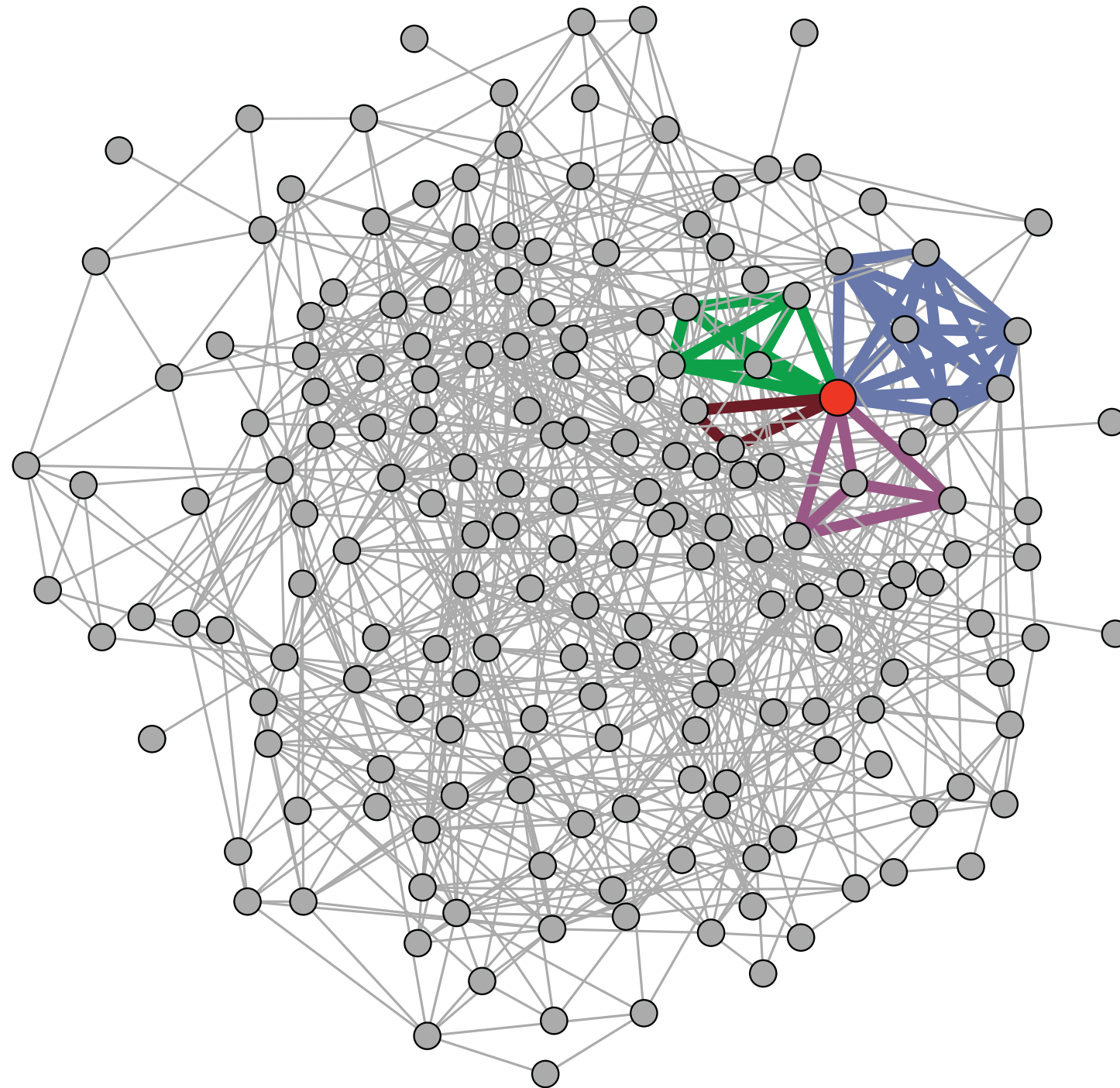


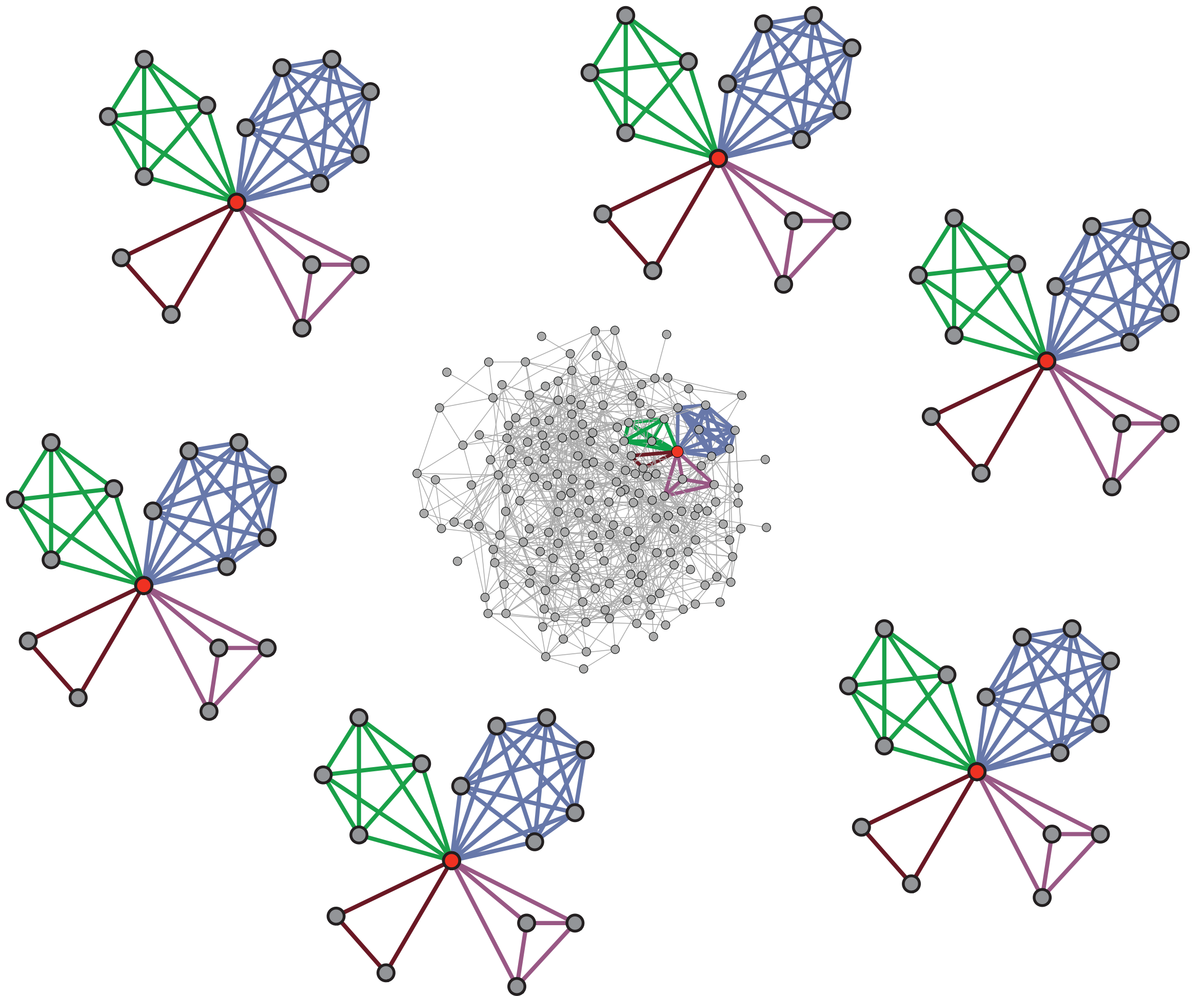


# Complex global structure

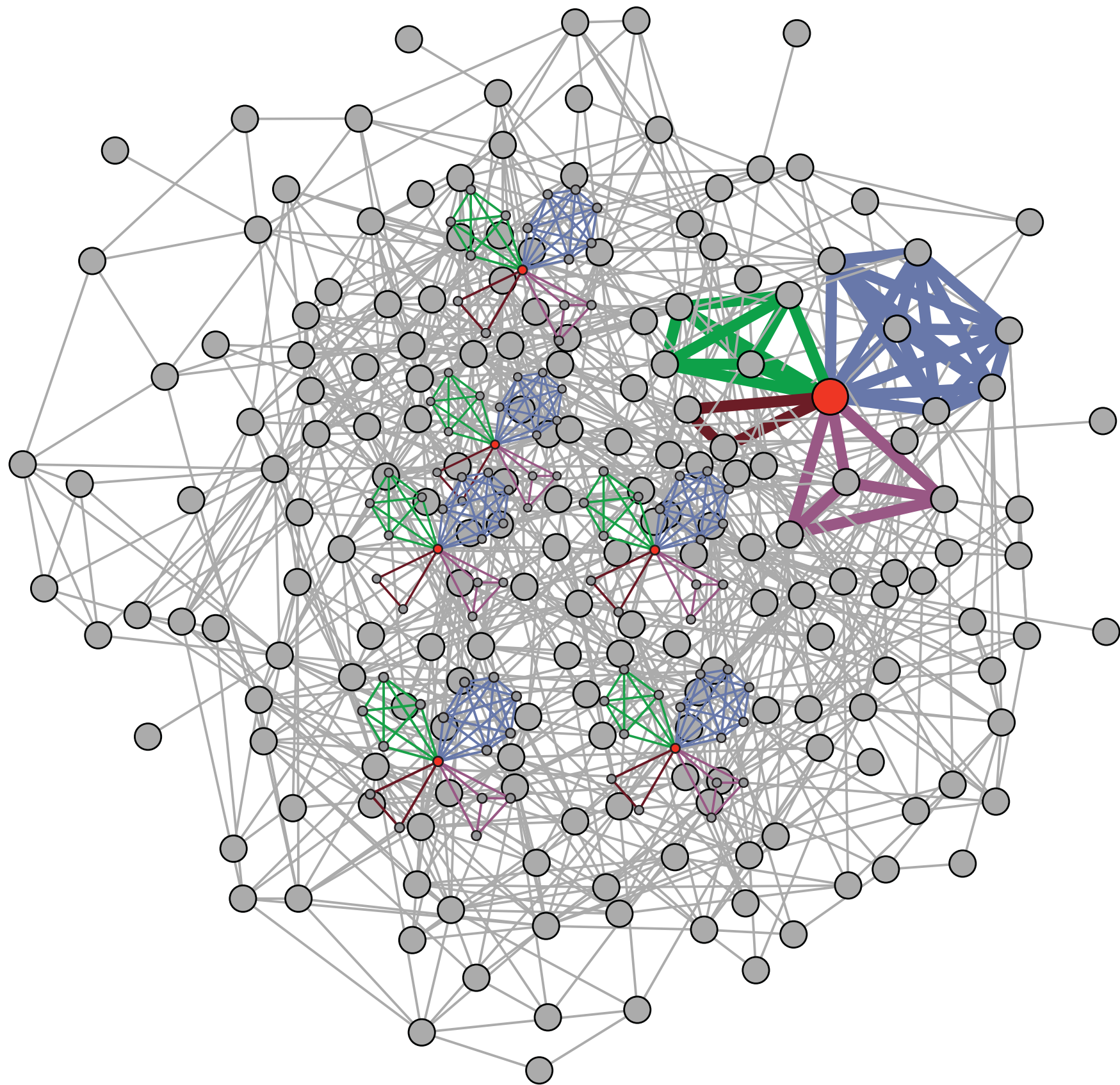


# Complex global structure

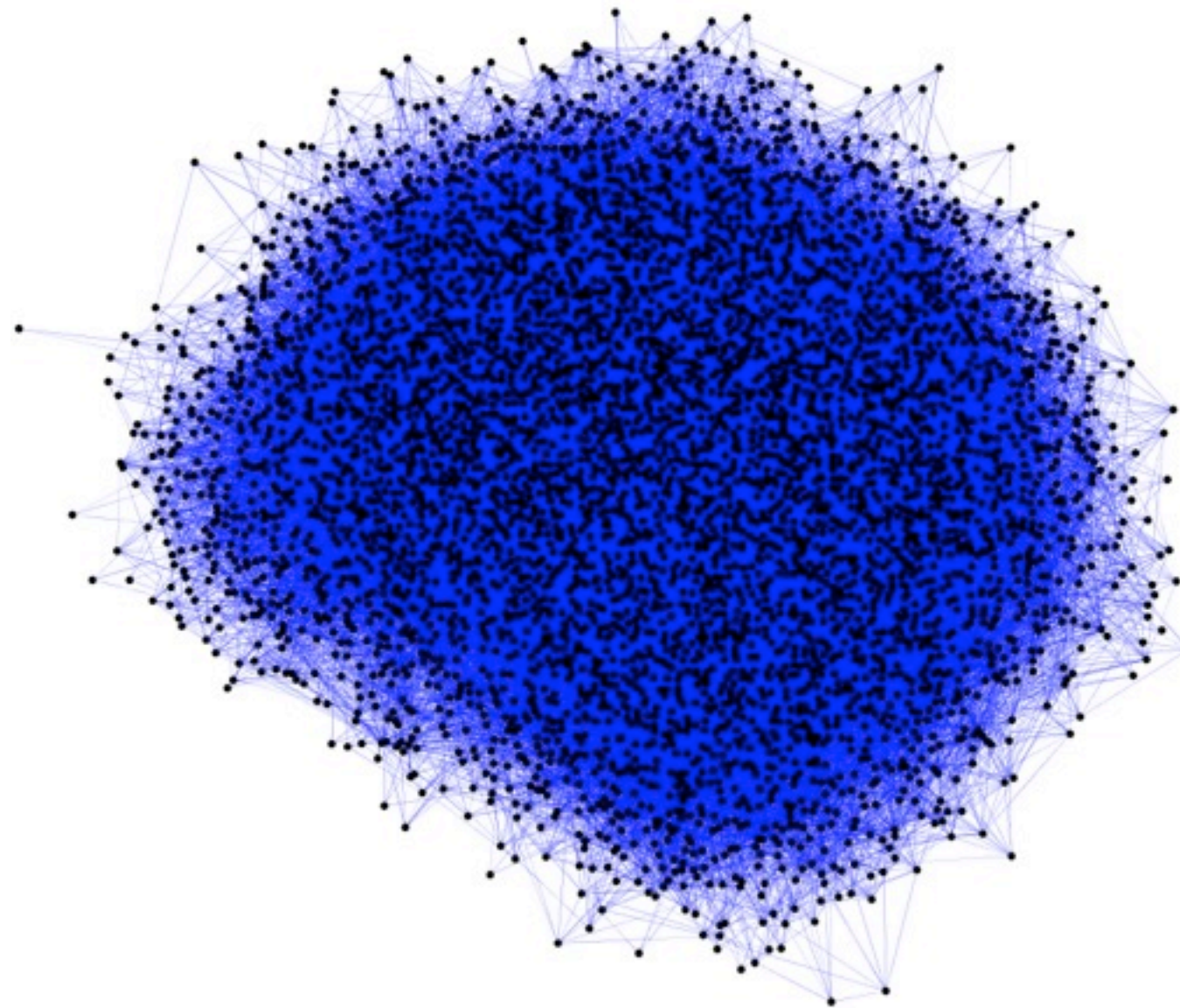








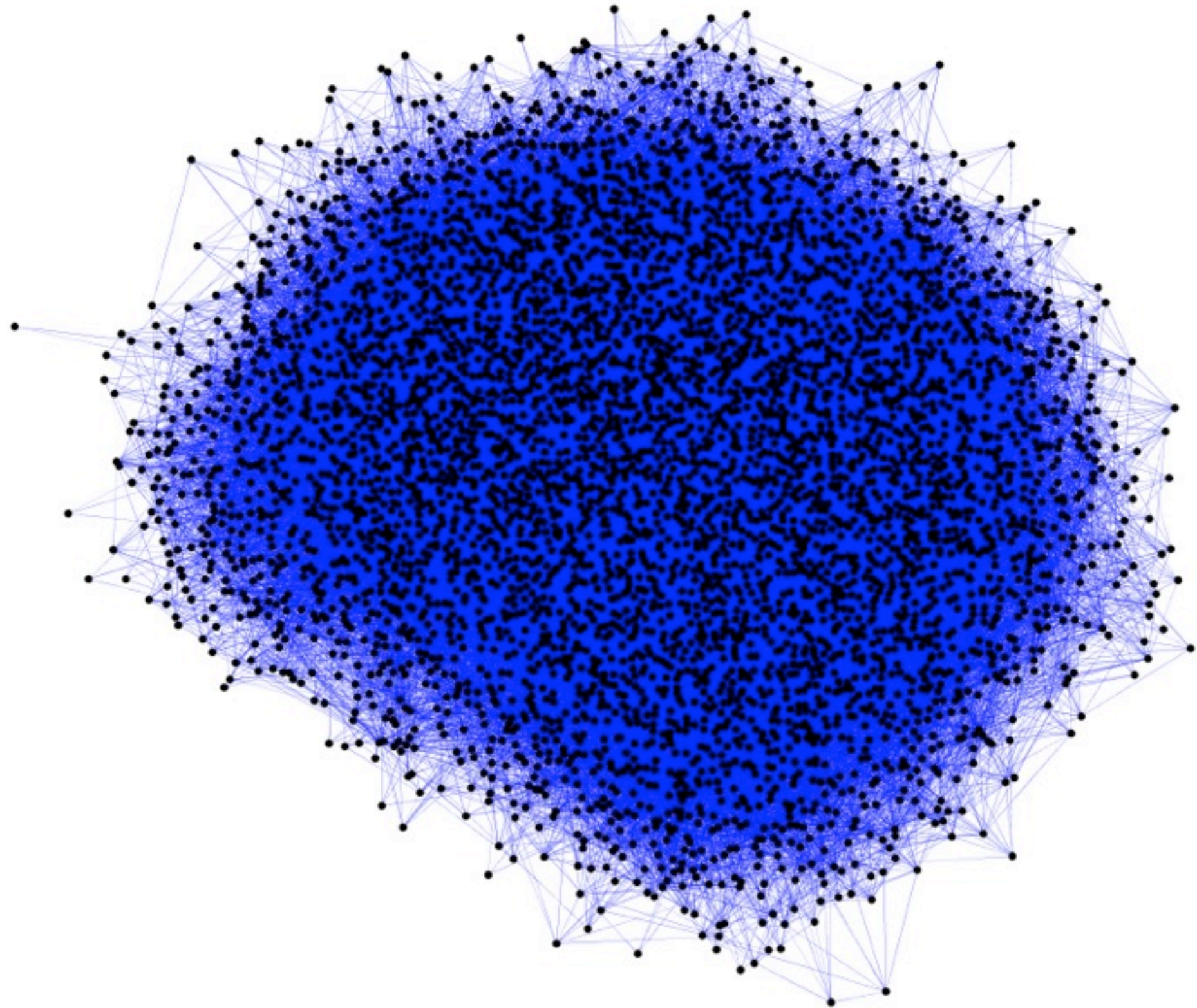




*This is a modular network.*

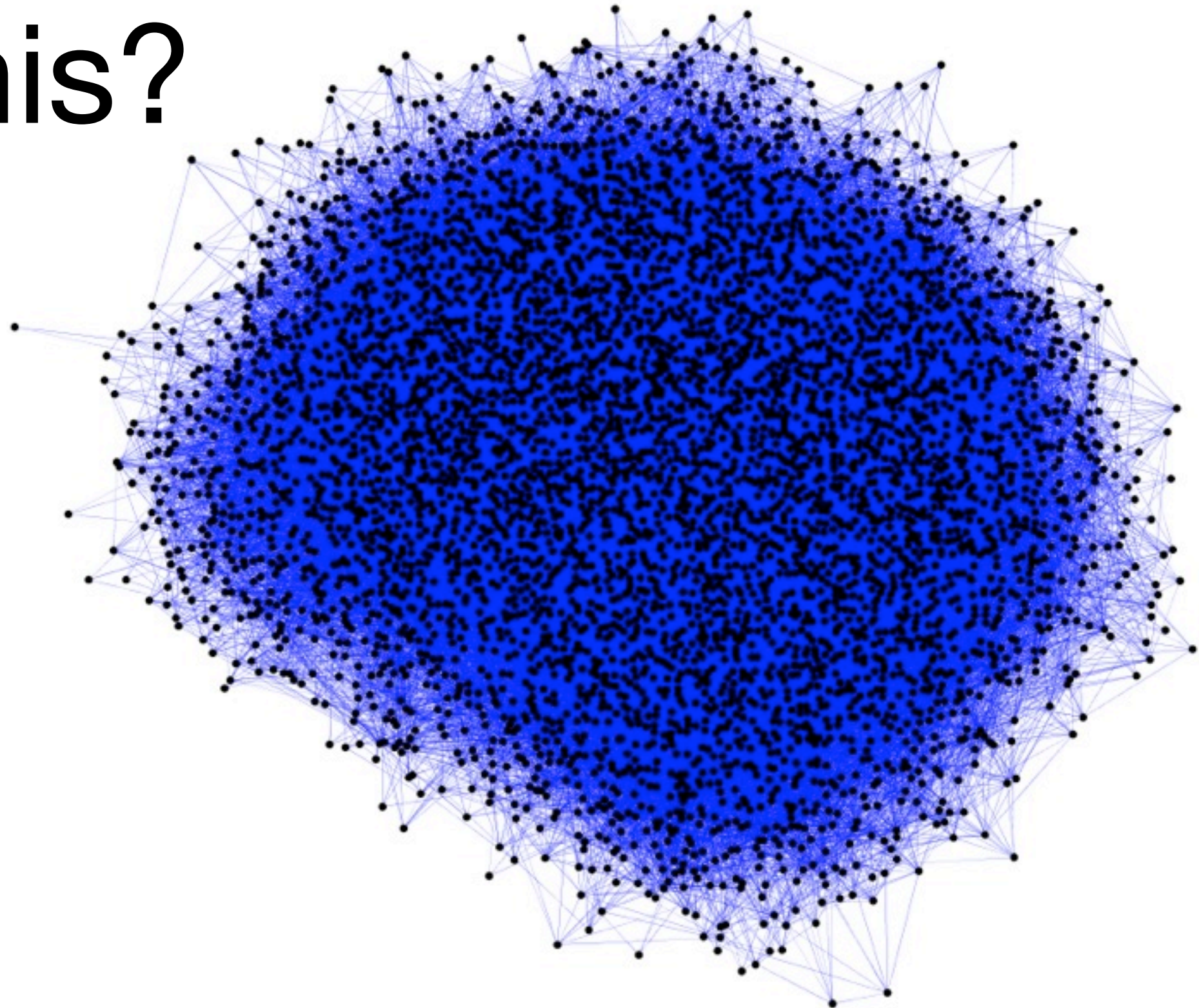


# What is this?



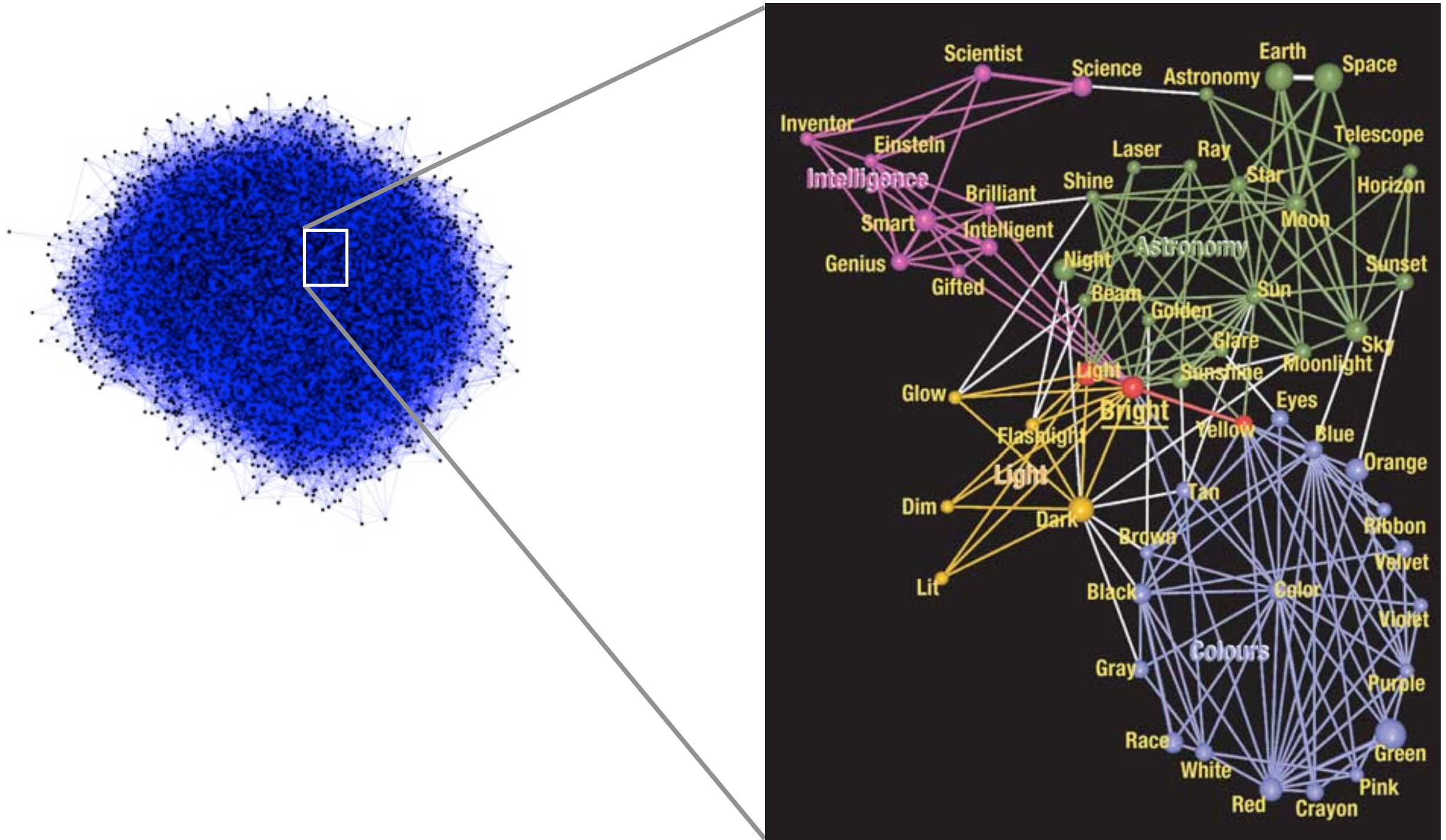


# What the xxxx is this?



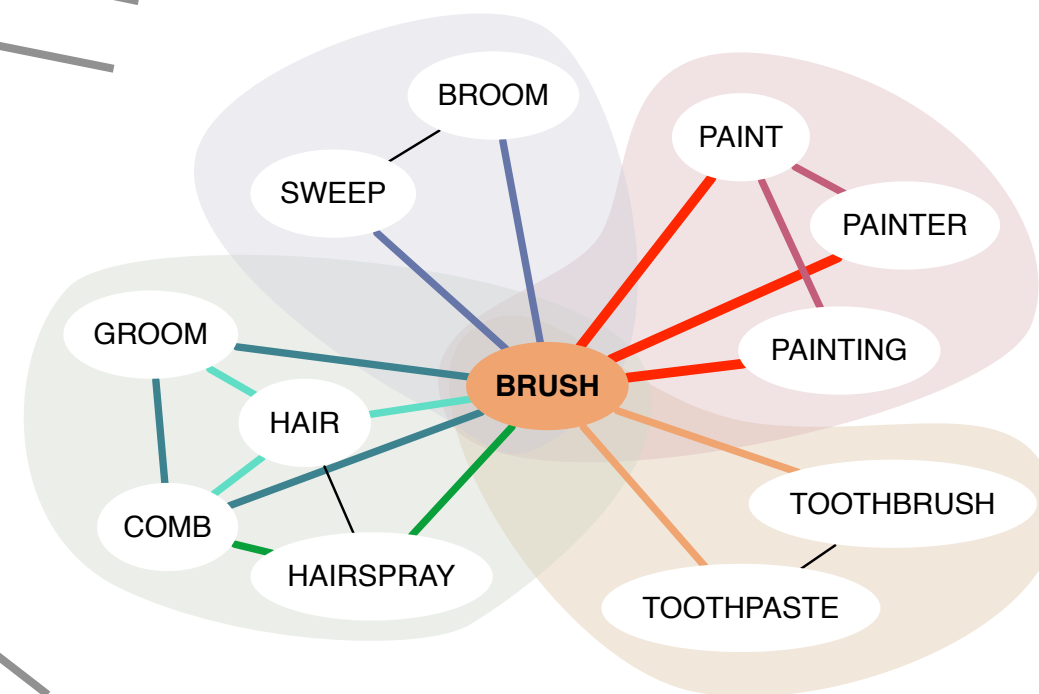
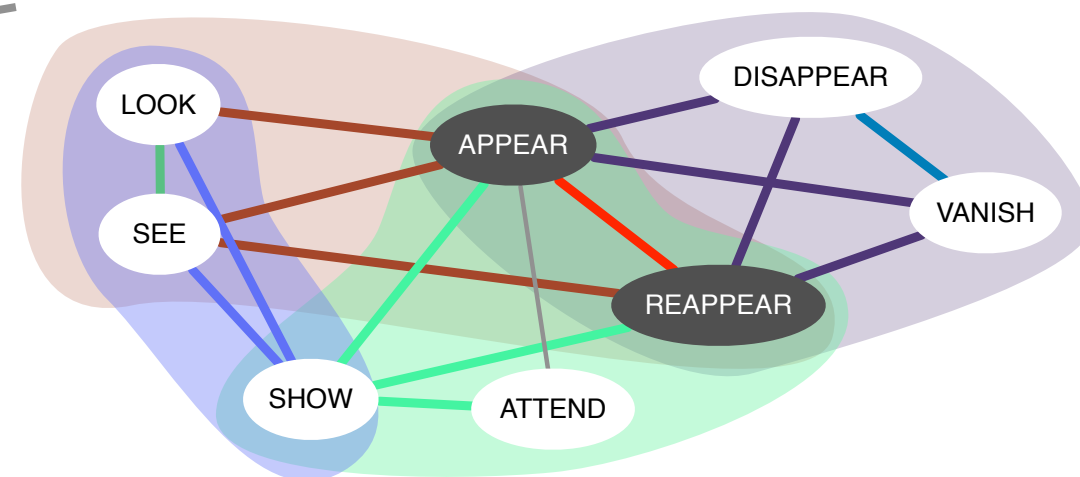
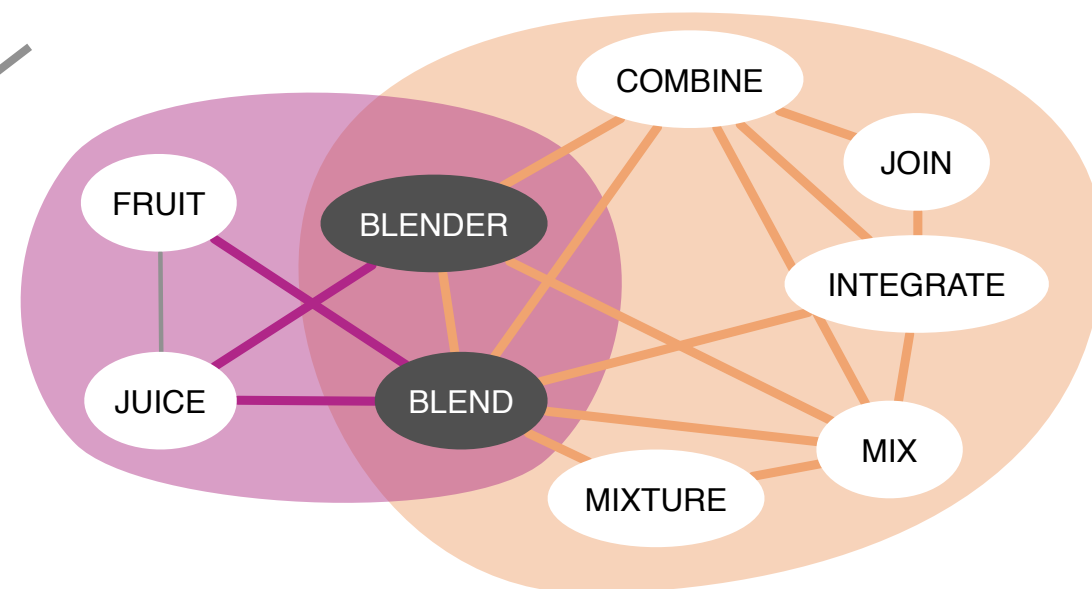
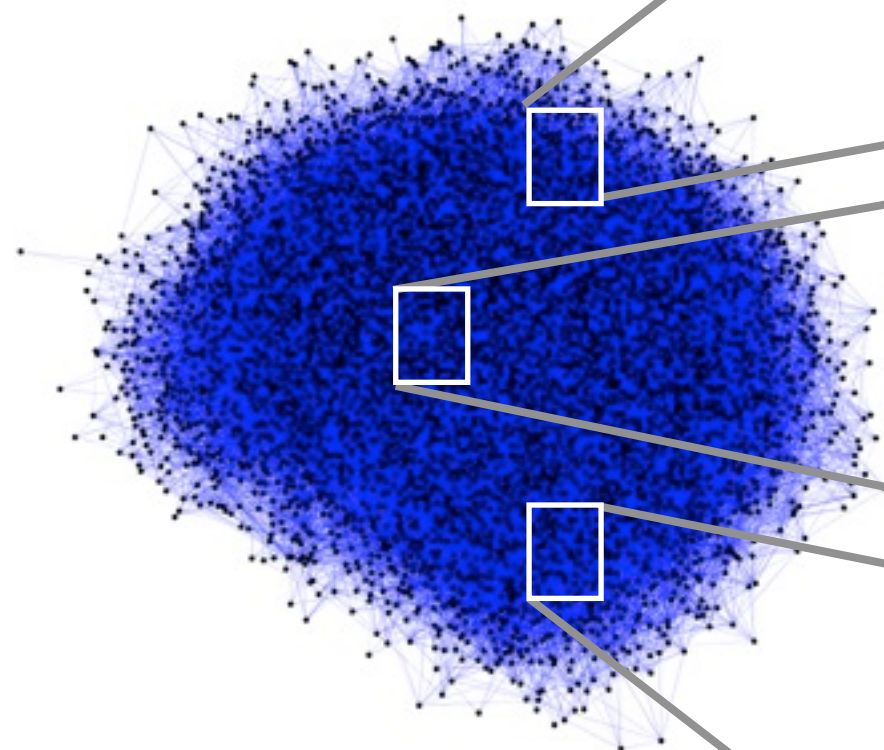


**Word association network:** Network of “commonly associated English words”



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





Here is the  
**PROBLEM.**

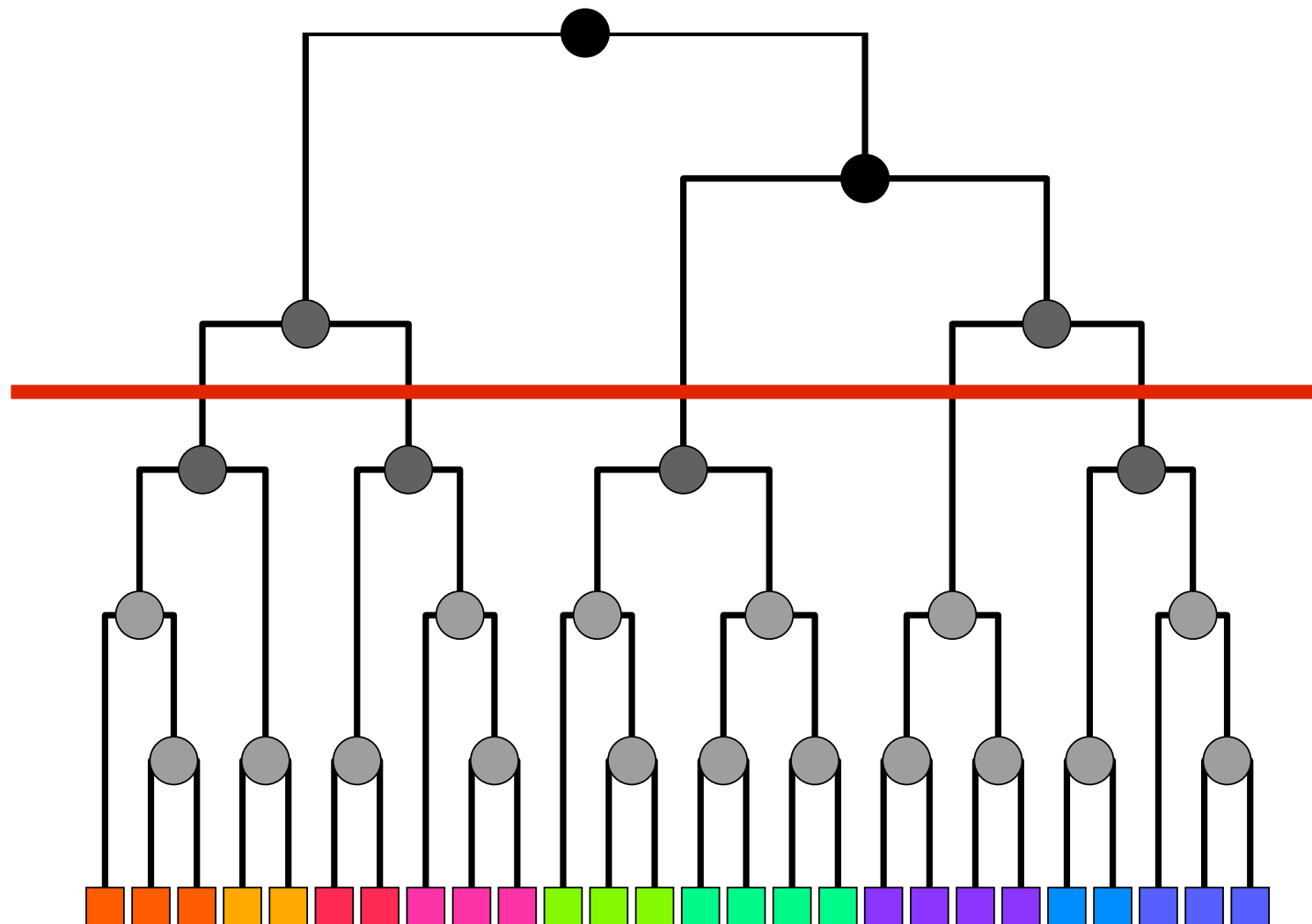


Communities exist.

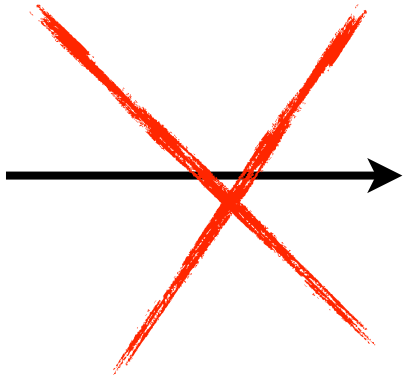
Hierarchical structure  
exists.



# Hierarchy implies disjoint communities.



# Hierarchical community structure

Hierarchy  Communities

# Hopeless?

Solution:  
Use **LINKS**



Solution:  
Use **LINKS**

# Solution: Use Links

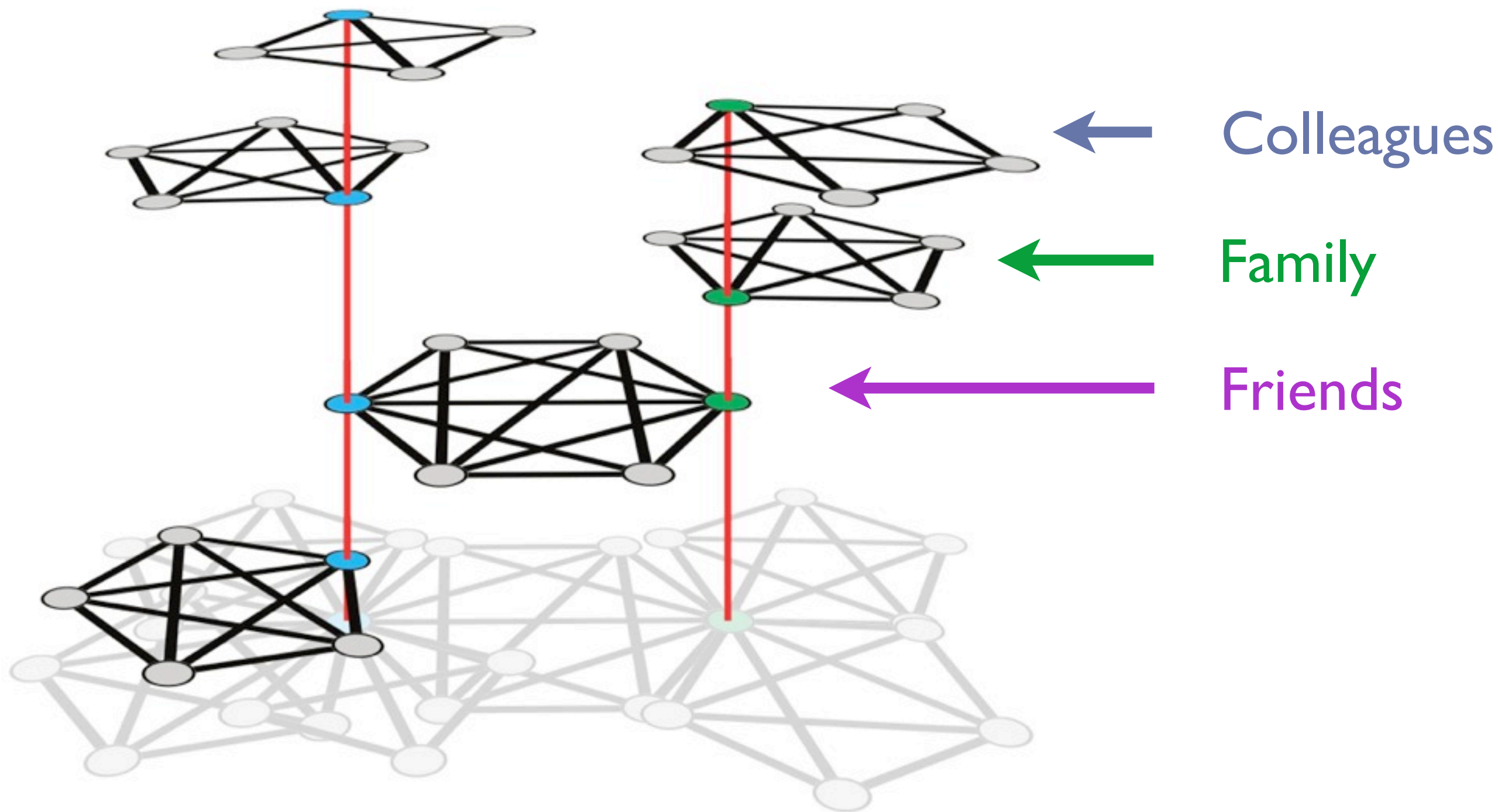
“a group of densely interconnected nodes”

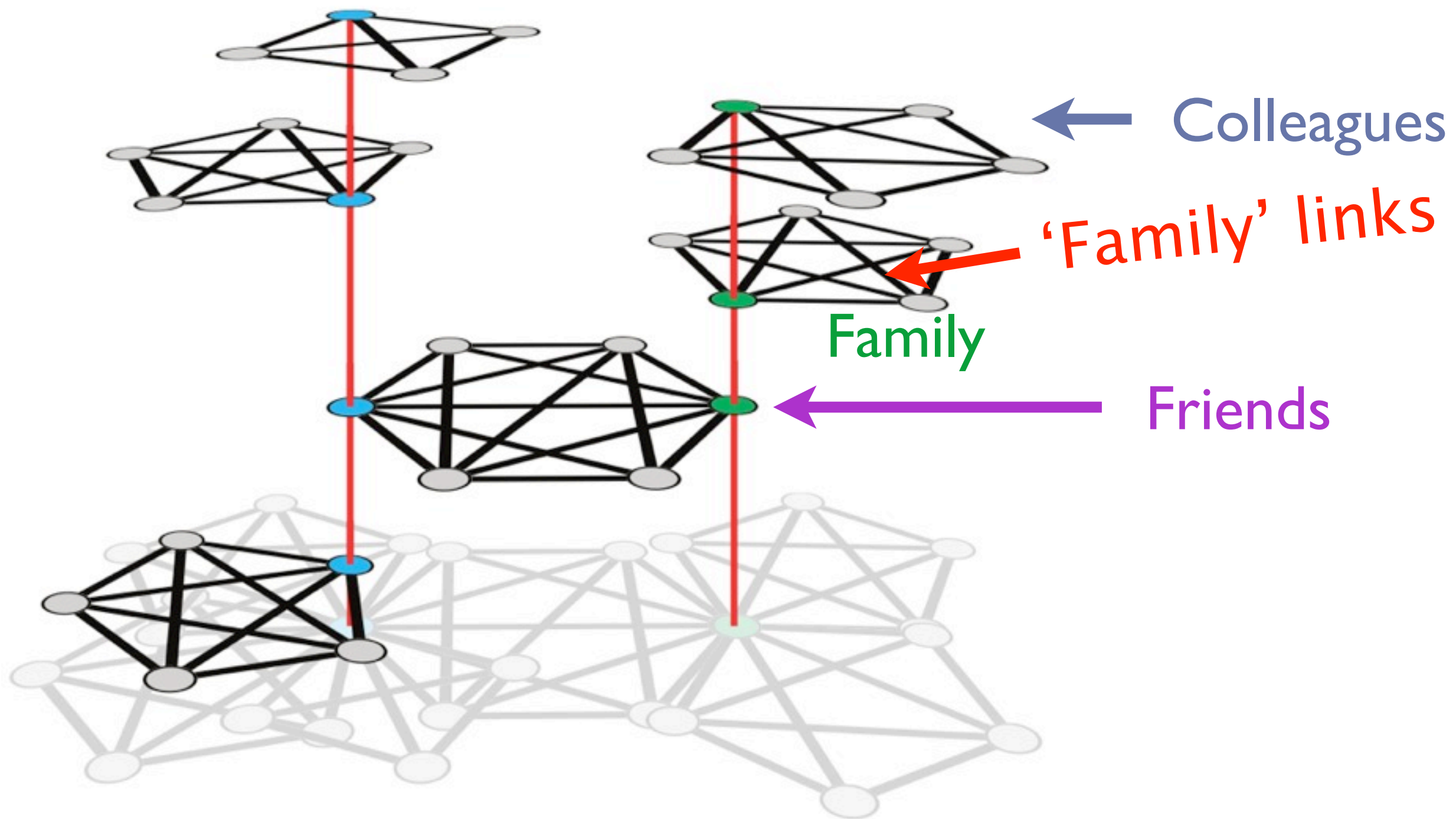
Our solution:  
Use Links

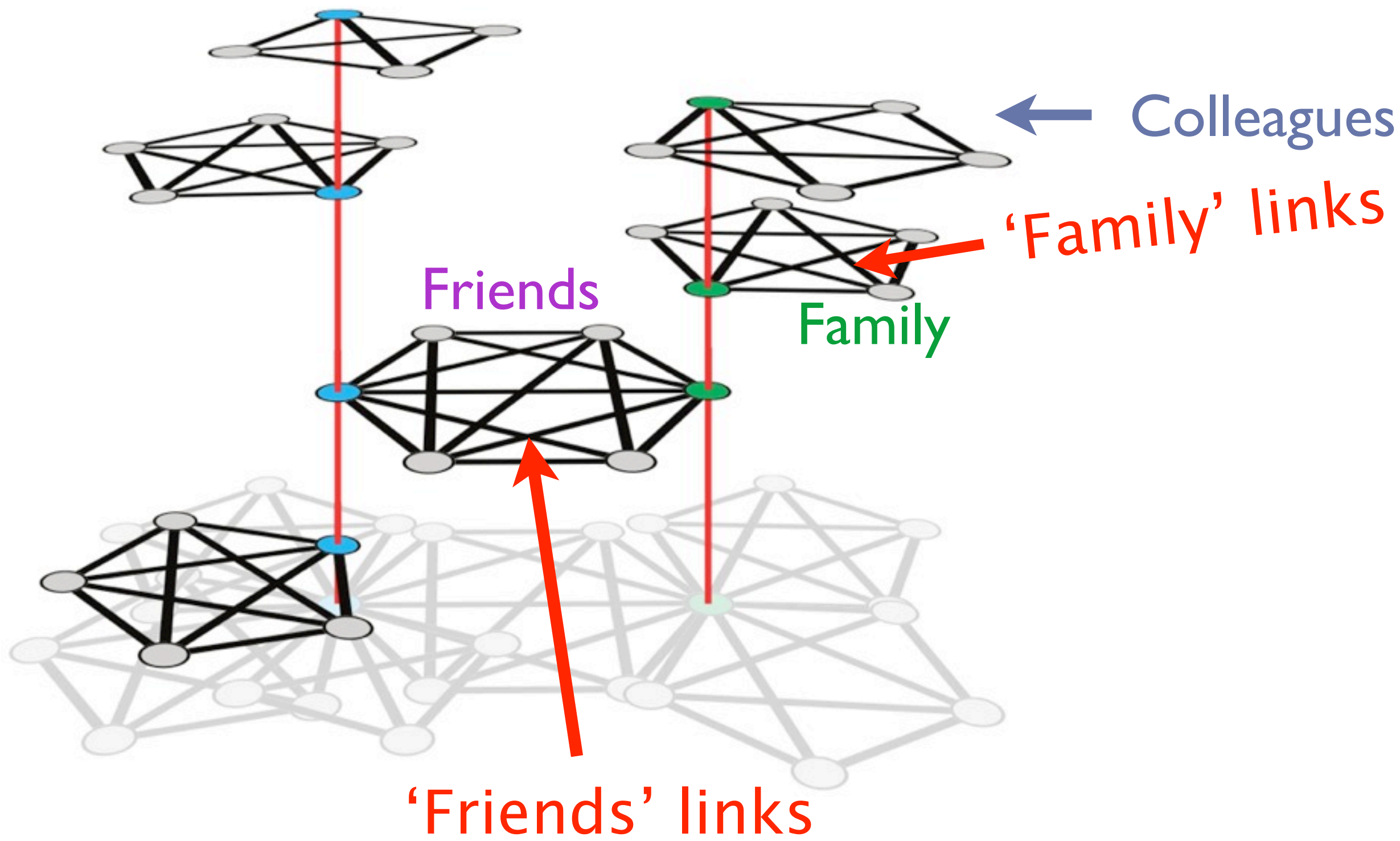
“a group of densely  
interconnected **LINKS**”

# LINK communities









‘Nerds & geeks’ links

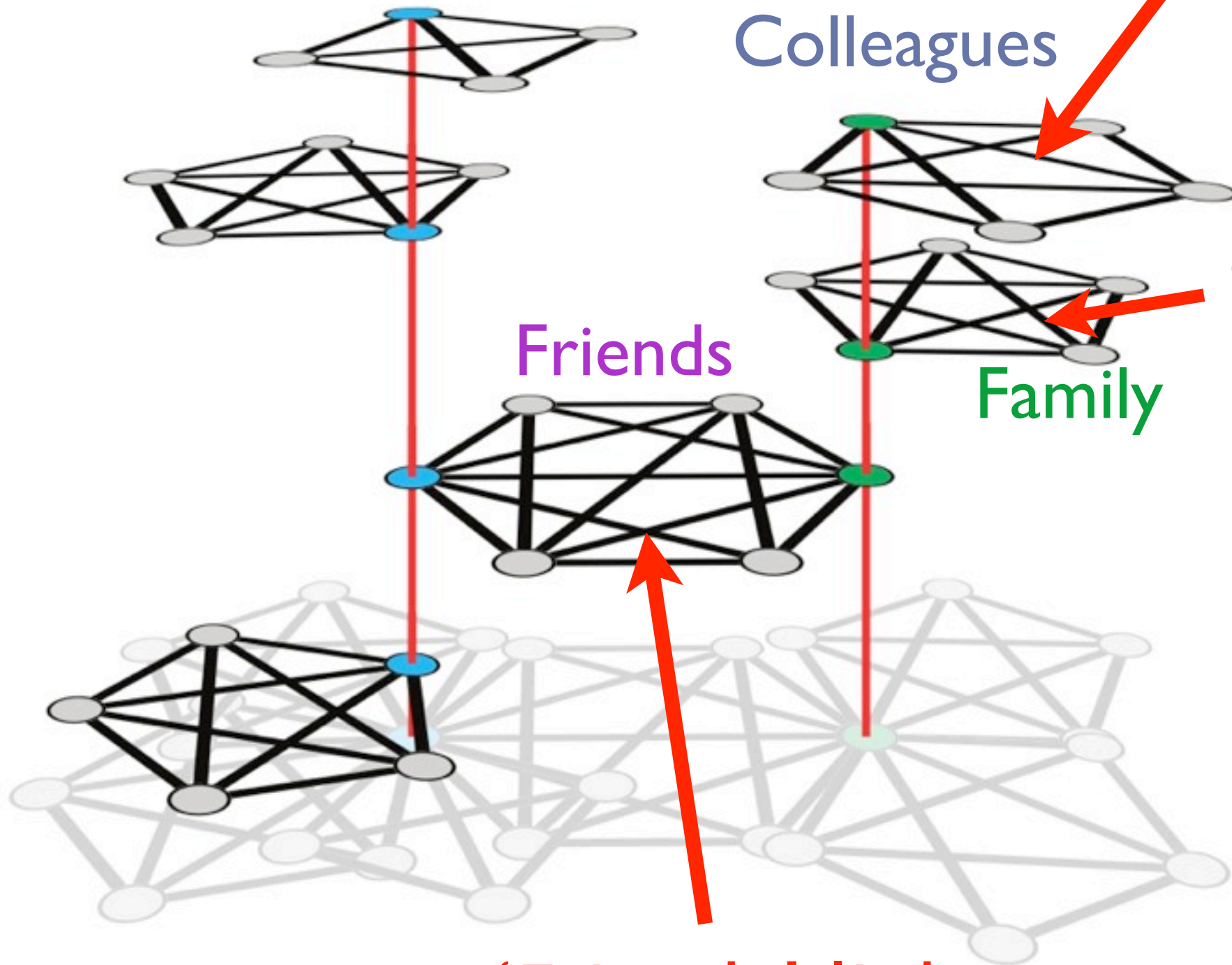
Colleagues

‘Family’ links

Friends

Family

‘Friends’ links

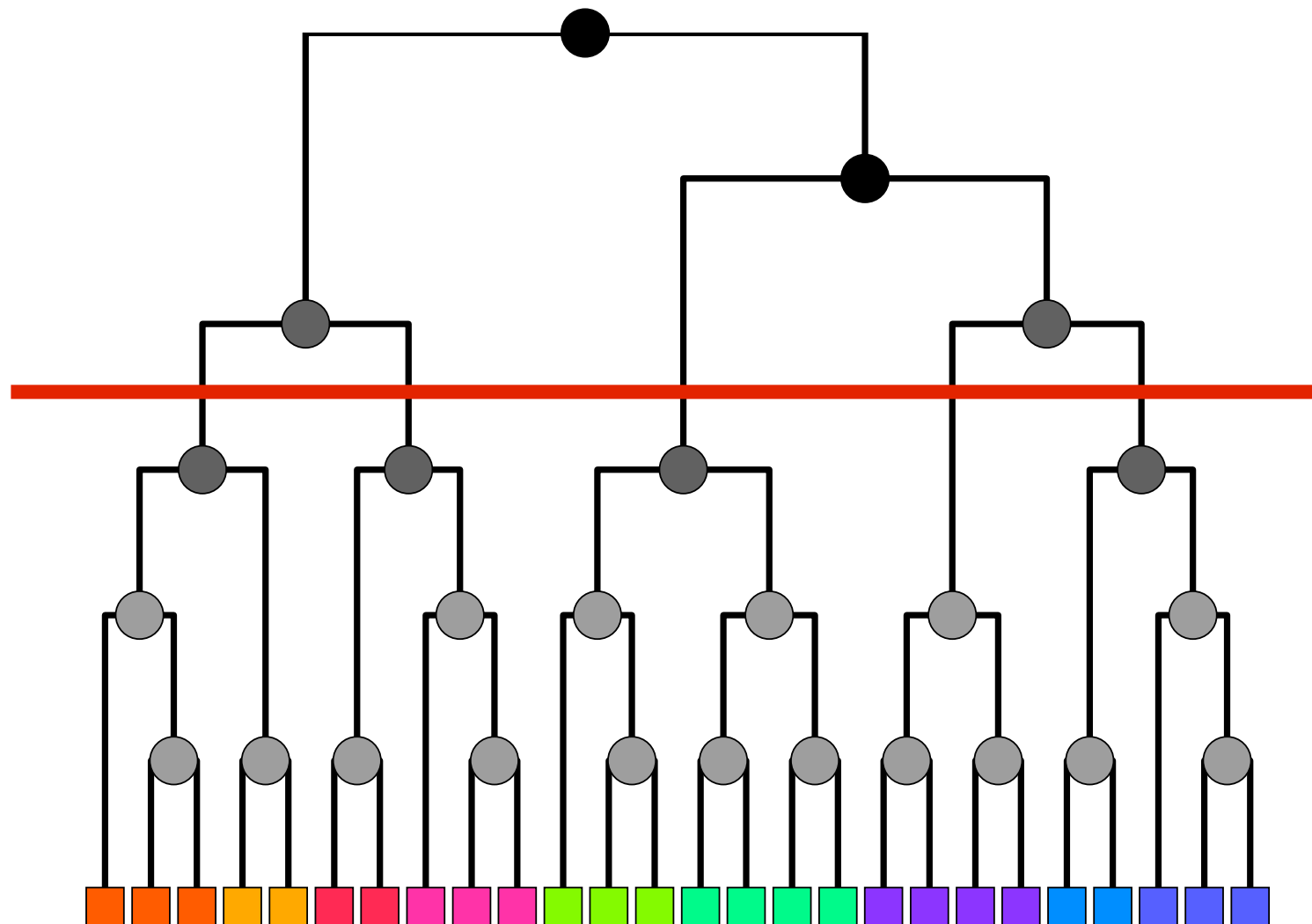


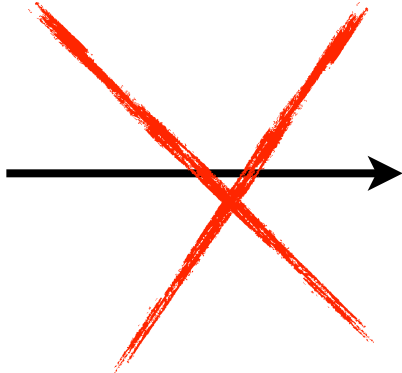


Nodes: multiple membership

Links: unique membership

# Hierarchy implies disjoint communities.



Hierarchy  Communities

Hierarchy —→ Communities





# RECONCILIATION

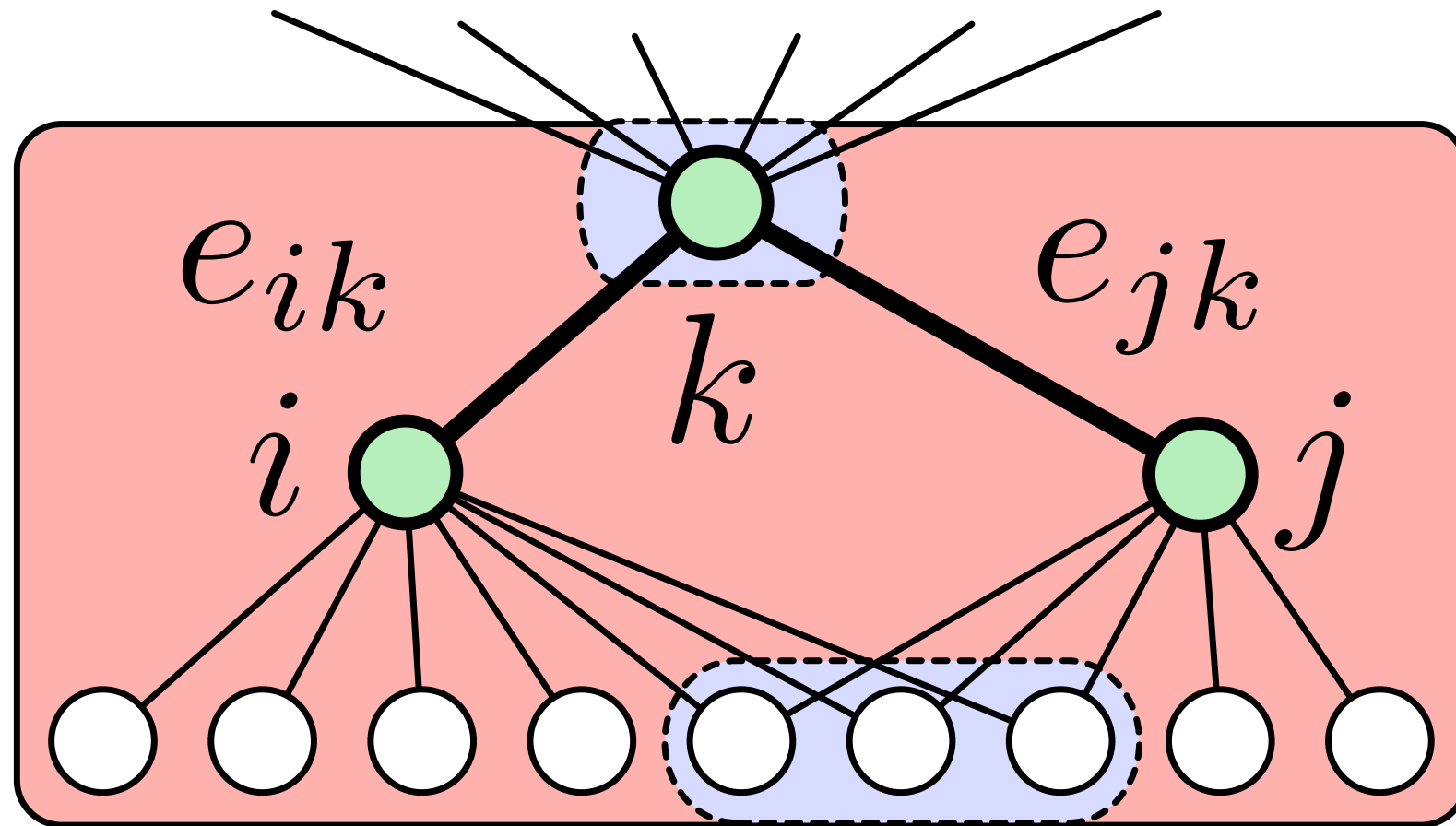
# So, How?

# Similarity between links



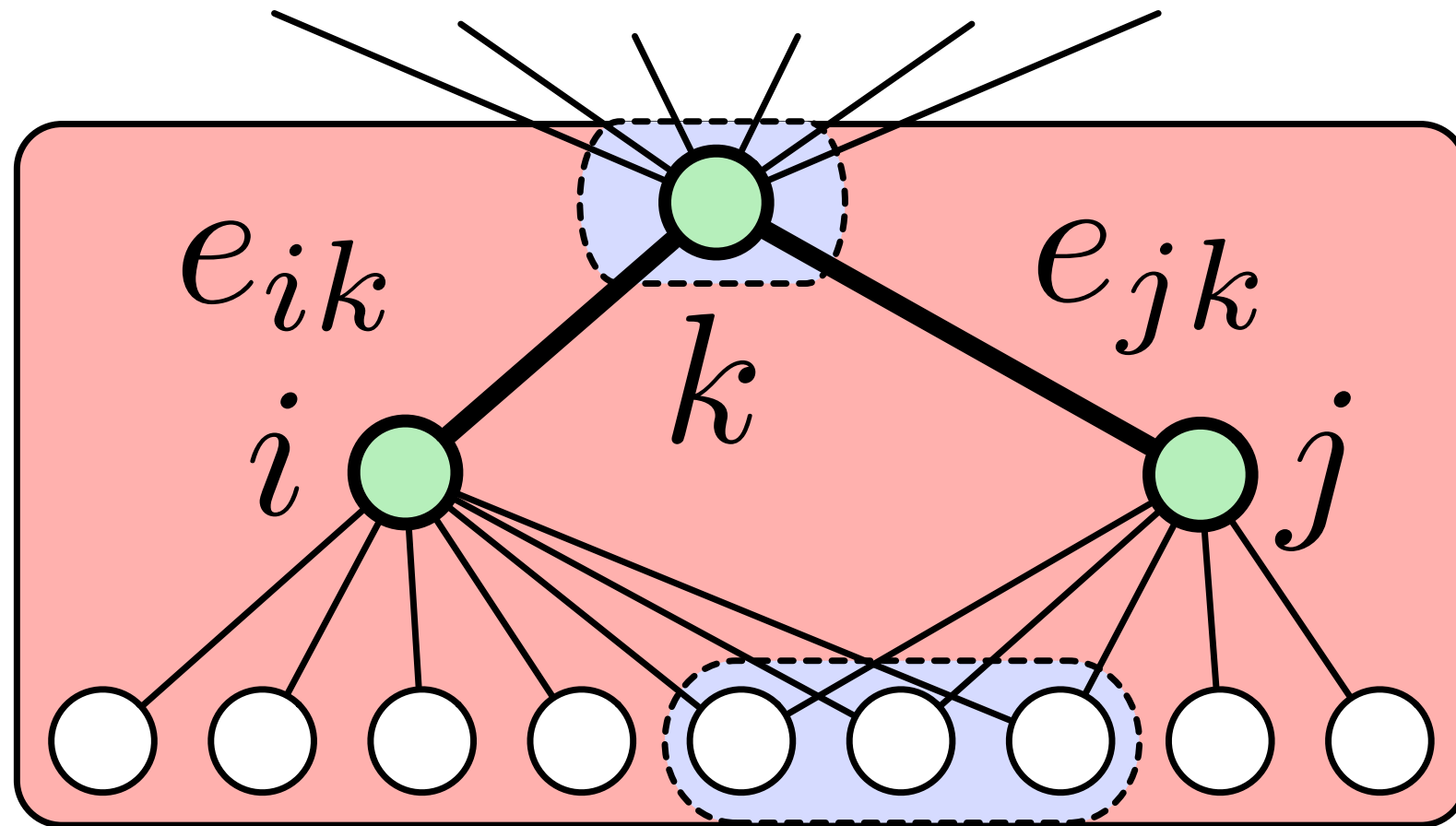
# Hierarchical Clustering





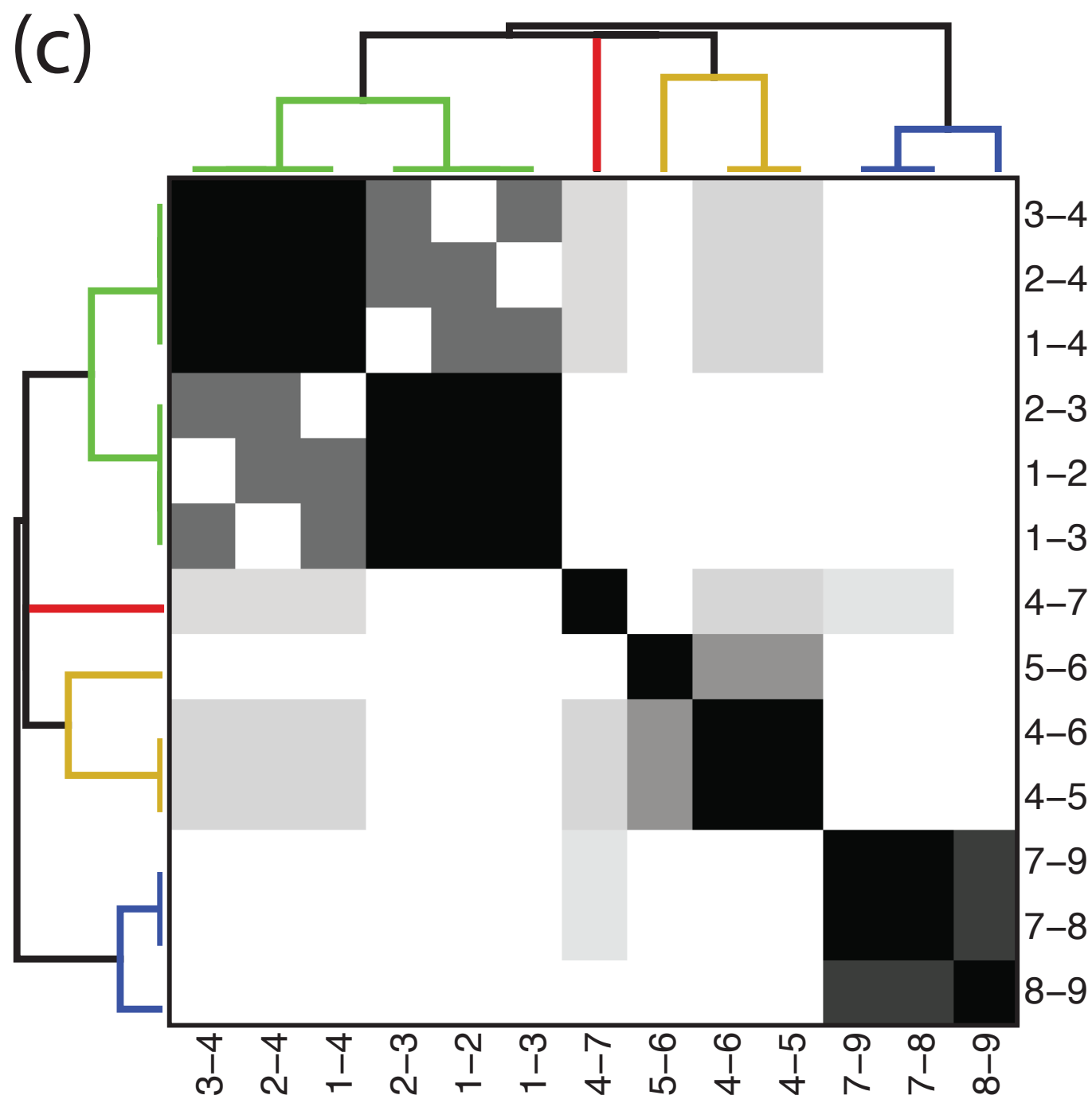
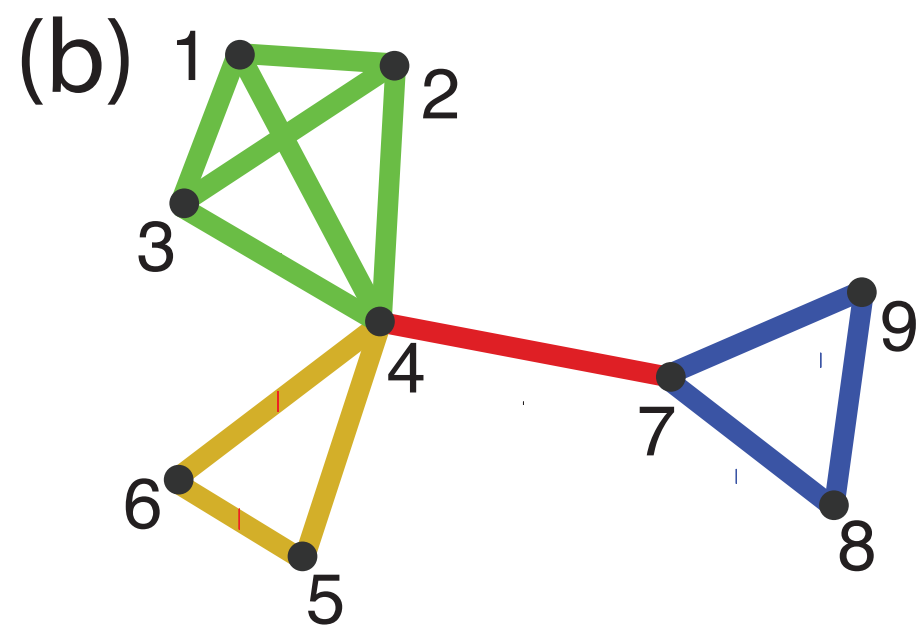
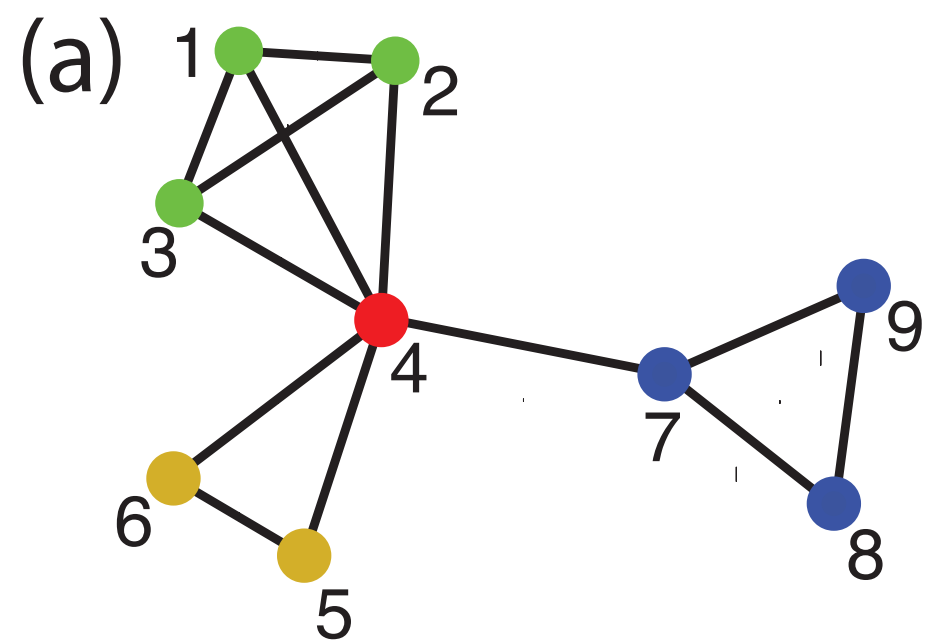
$$n_+(i) \equiv \{x \mid d(i, x) \leq 1\}$$

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



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

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



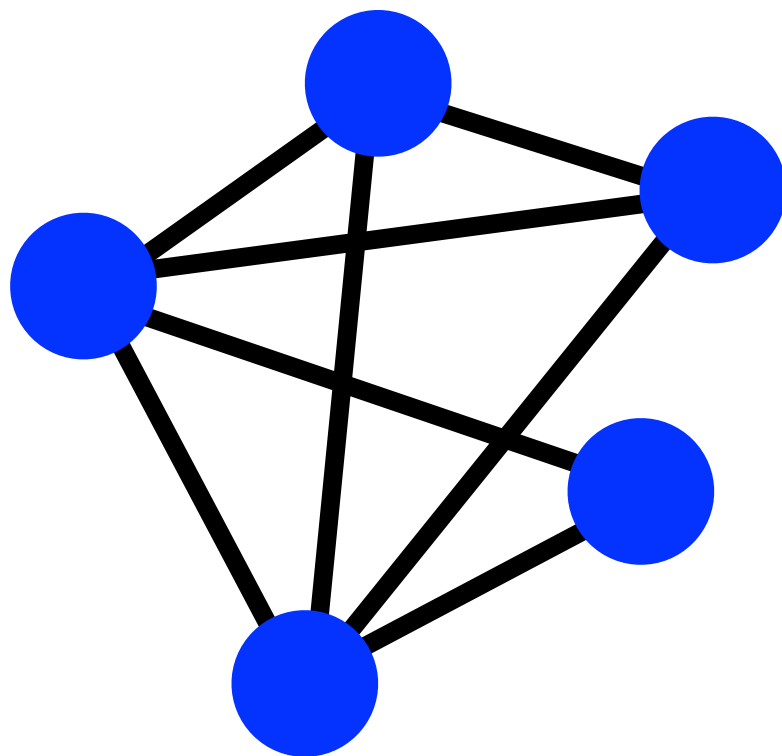
# Partition Density

Community  $c$  has  $m_c$  edges and  $n_c$  induced nodes



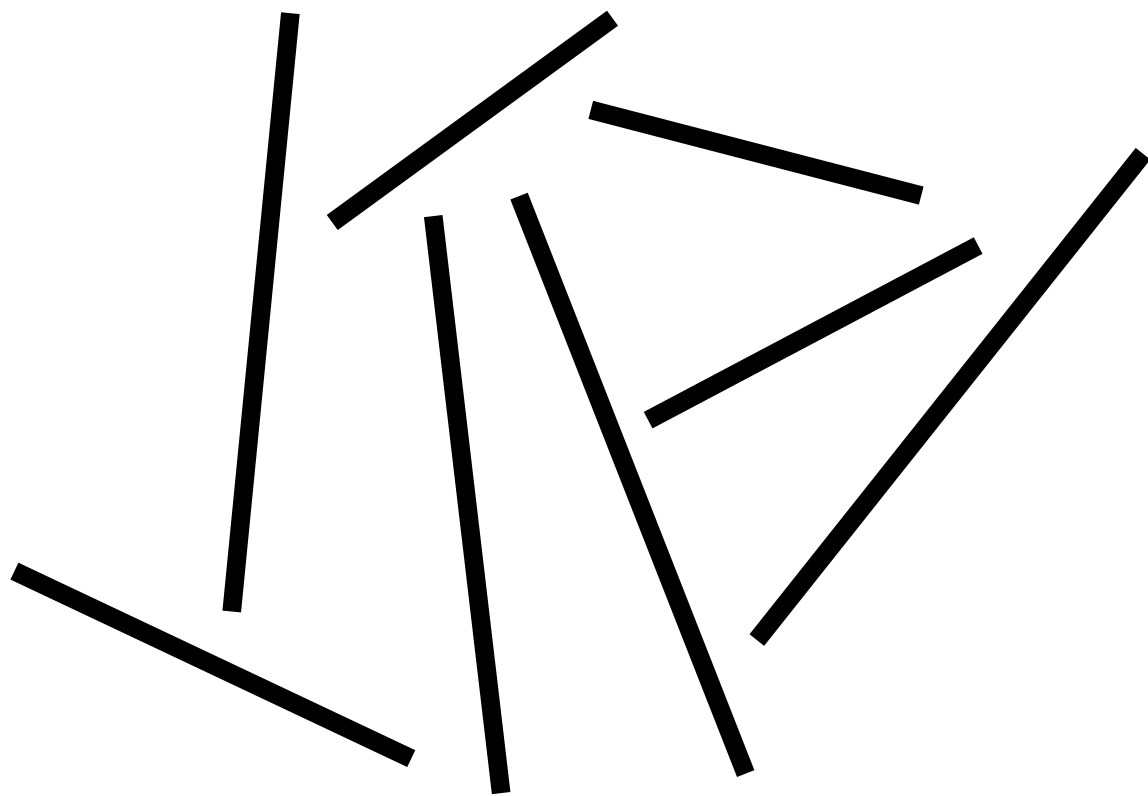
# Partition Density

Community  $c$  has  $m_c$  edges and  $n_c$  induced nodes

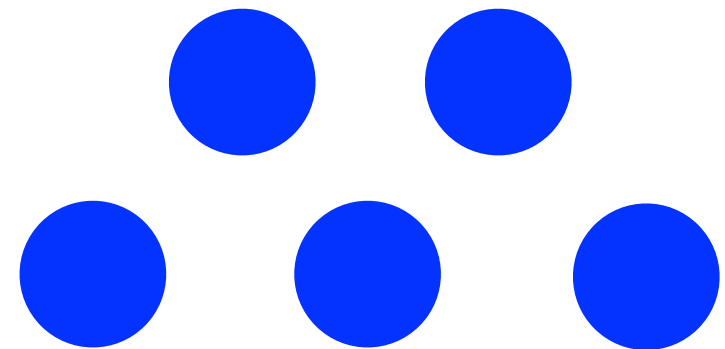


# Partition Density

Community  $c$  has  $m_c$  edges and  $n_c$  induced nodes



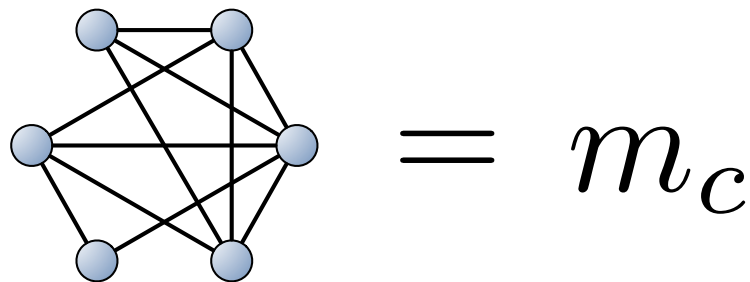
$$m_c = 8$$



$$n_c = 5$$

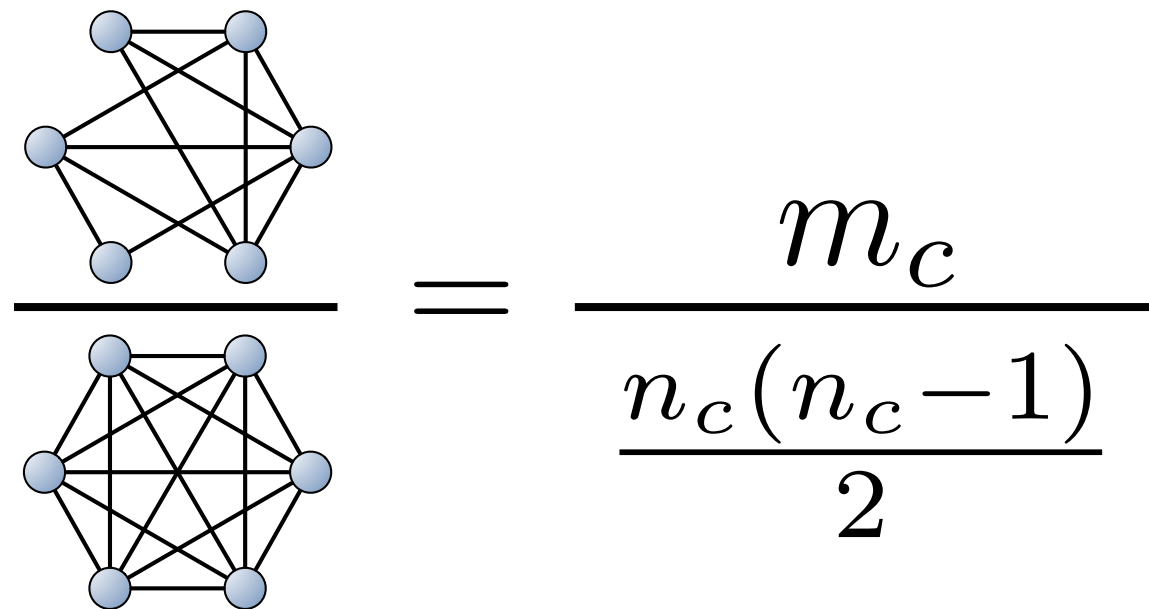
# Partition Density

Community  $c$  has  $m_c$  edges and  $n_c$  induced nodes



# Partition Density

Community  $c$  has  $m_c$  edges and  $n_c$  induced nodes



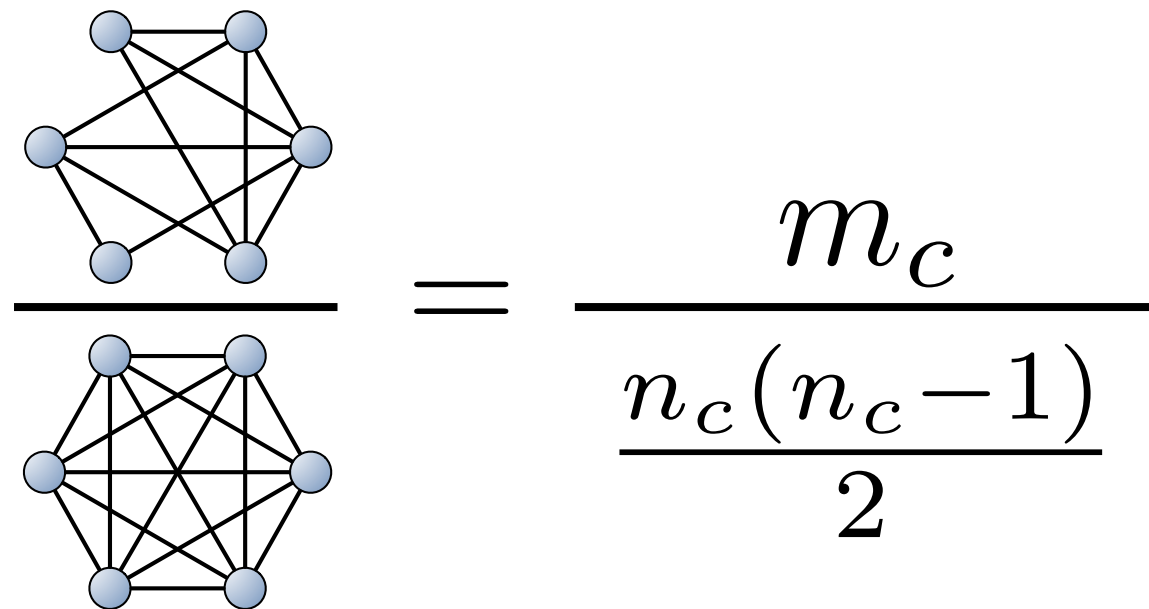
The diagram illustrates the formula for Partition Density. It consists of two complete graphs of 6 nodes each, represented by blue circles connected by black lines. The top graph is divided by a horizontal line, and the bottom graph is also divided by a horizontal line. To the right of these graphs is an equals sign followed by a fraction. The numerator of this fraction is  $m_c$ , and the denominator is  $\frac{n_c(n_c - 1)}{2}$ .

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



# Partition Density

Community  $c$  has  $m_c$  edges and  $n_c$  induced nodes

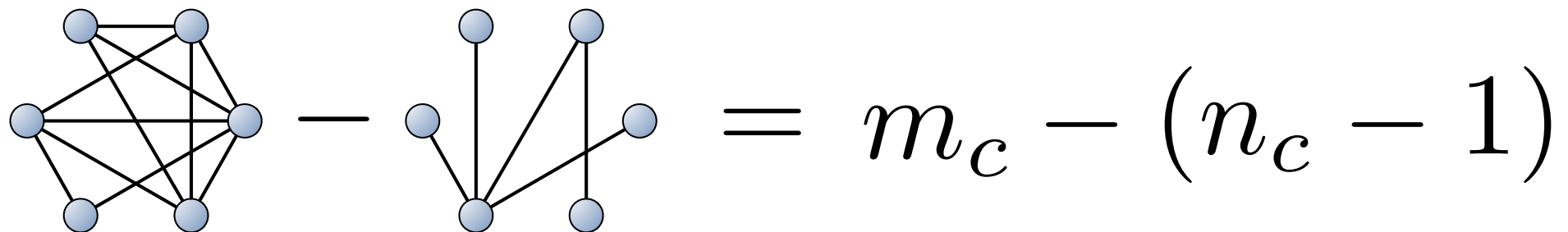

$$\frac{\text{Complete Graph } K_5}{\text{Complete Graph } K_6} = \frac{m_c}{\frac{n_c(n_c-1)}{2}}$$



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

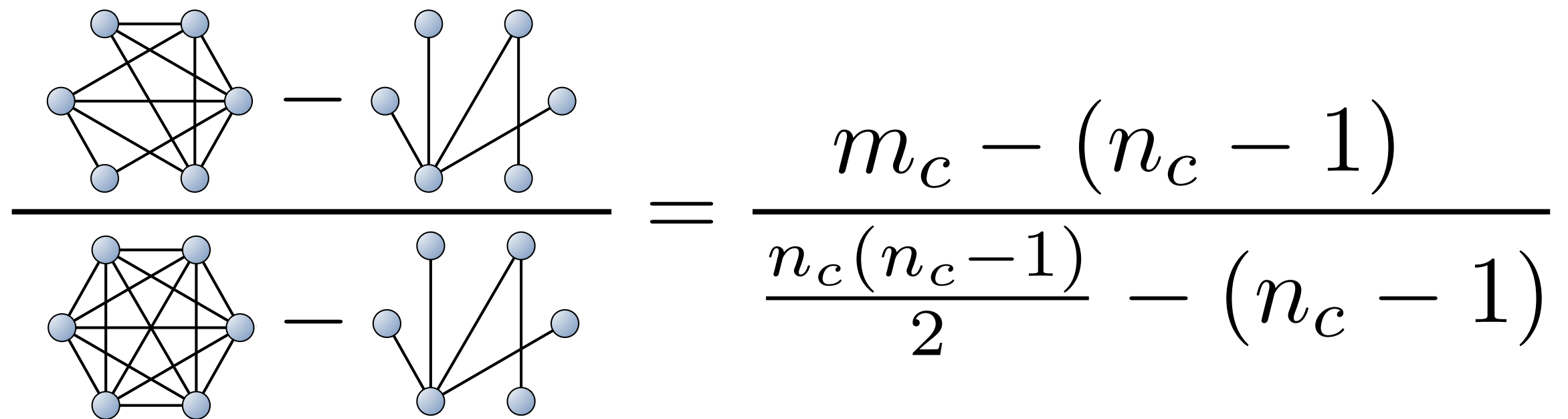
# Partition Density

Community  $c$  has  $m_c$  edges and  $n_c$  induced nodes


$$\text{Complete Graph } K_6 - \text{Star Graph } S_5 = m_c - (n_c - 1)$$

# Partition Density

Community  $c$  has  $m_c$  edges and  $n_c$  induced nodes

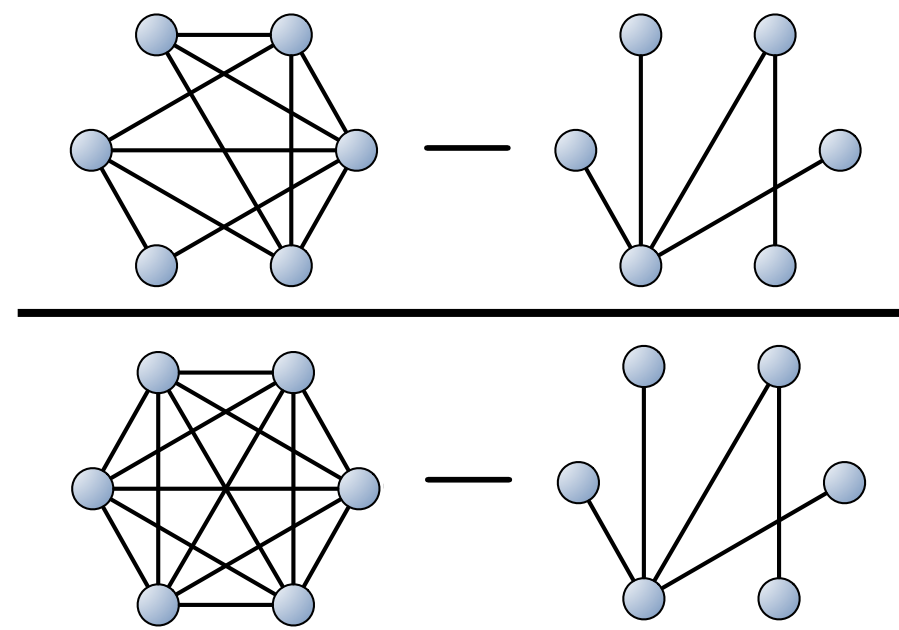
$$\frac{\text{Community } c \text{ edges} - \text{Community } c \text{ induced edges}}{\text{Complete graph } K_{n_c} \text{ edges} - \text{Community } c \text{ induced edges}} = \frac{m_c - (n_c - 1)}{\frac{n_c(n_c - 1)}{2} - (n_c - 1)}$$


# Partition Density

$$\begin{array}{c}
 \begin{array}{c} \text{Complete graph } K_6 \\ - \\ \text{Star graph } S_5 \end{array} \\
 \hline
 \begin{array}{c} \text{Complete graph } K_6 \\ - \\ \text{Star graph } S_5 \end{array}
 \end{array}
 = \frac{m_c - (n_c - 1)}{\frac{n_c(n_c - 1)}{2} - (n_c - 1)}
 = 2 \frac{m_c - (n_c - 1)}{(n_c - 2)(n_c - 1)}$$



# Partition Density



$$\frac{\text{Top Graph} - \text{Bottom Graph}}{\text{Top Graph} - \text{Bottom Graph}} = \frac{m_c - (n_c - 1)}{\frac{n_c(n_c - 1)}{2} - (n_c - 1)}$$

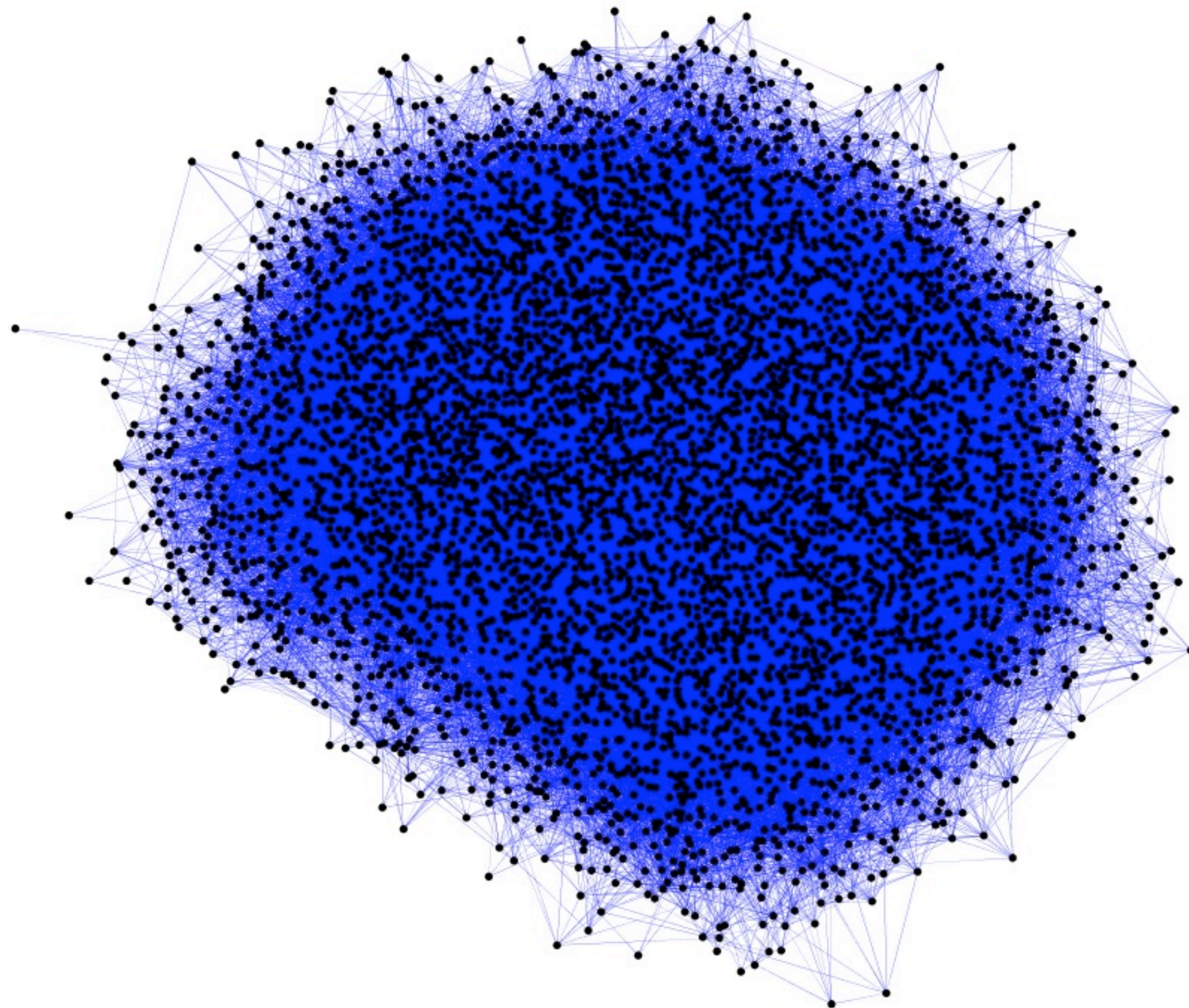
$$= 2 \frac{m_c - (n_c - 1)}{(n_c - 2)(n_c - 1)}$$

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

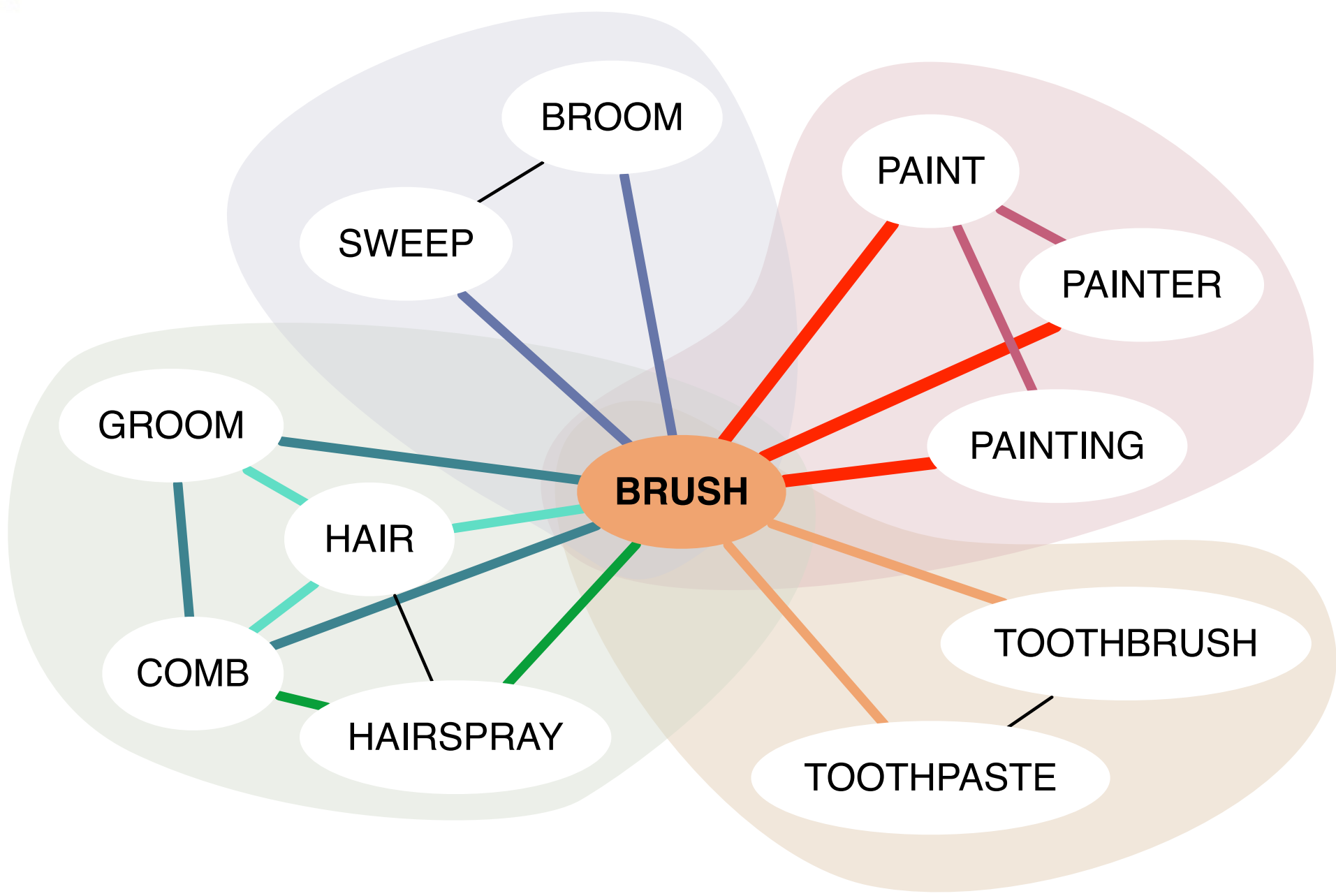
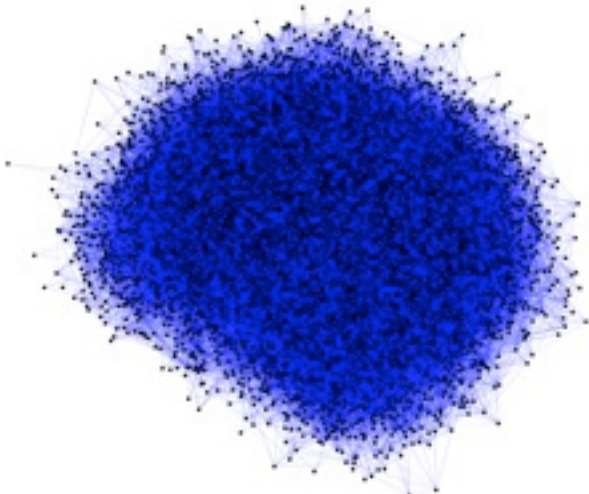


# Does it really work?



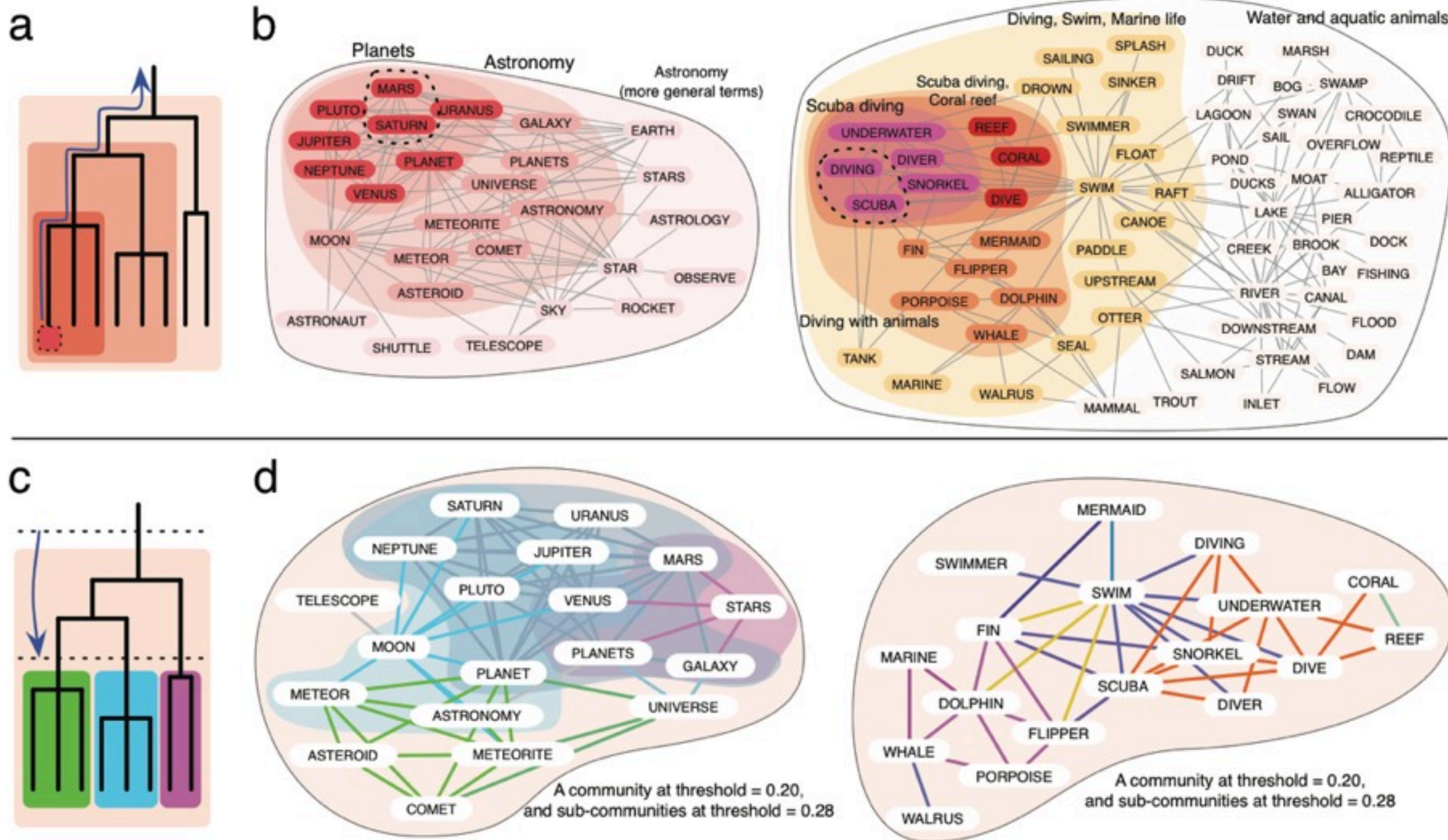




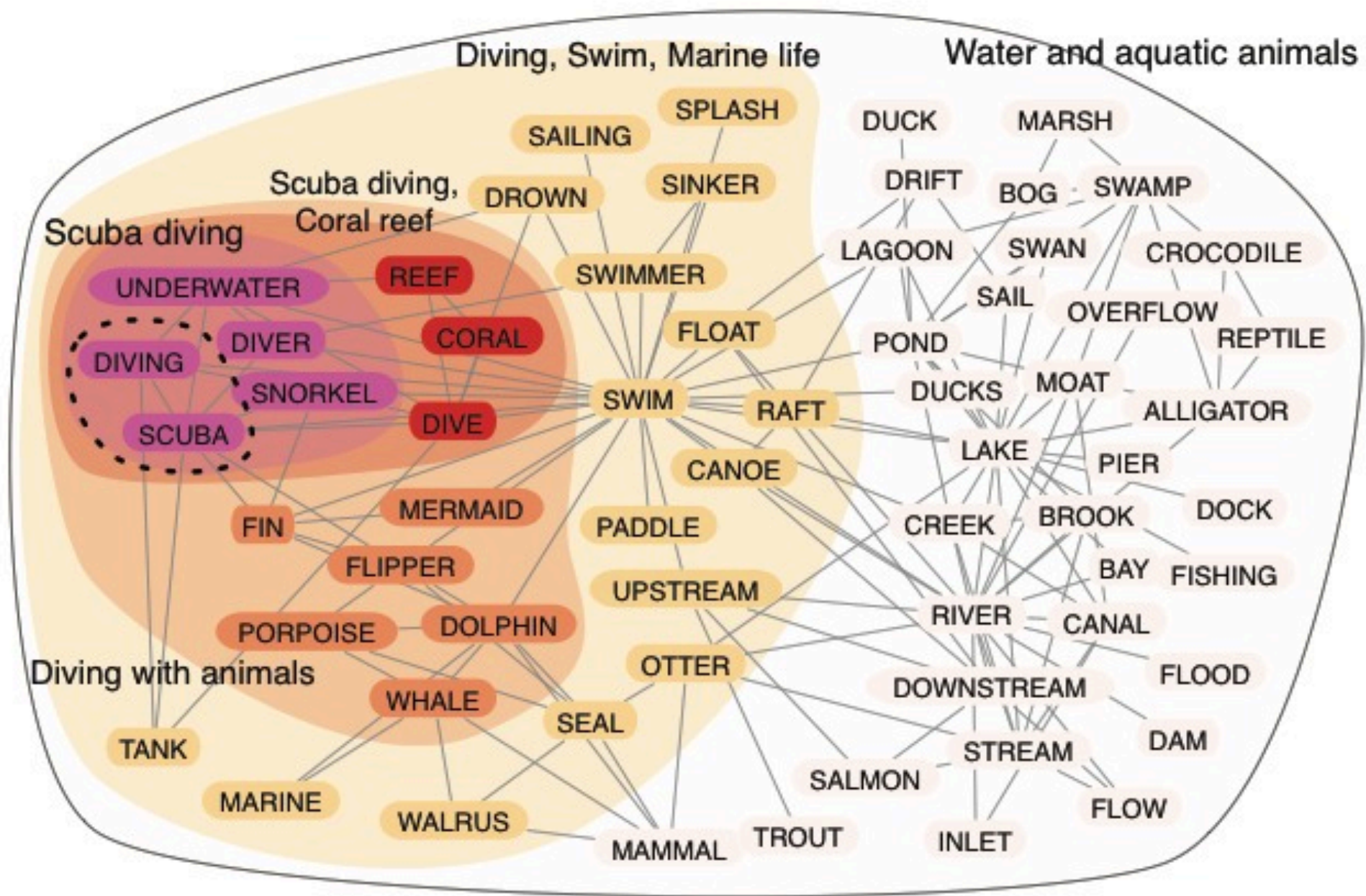




# Hierarchical organization

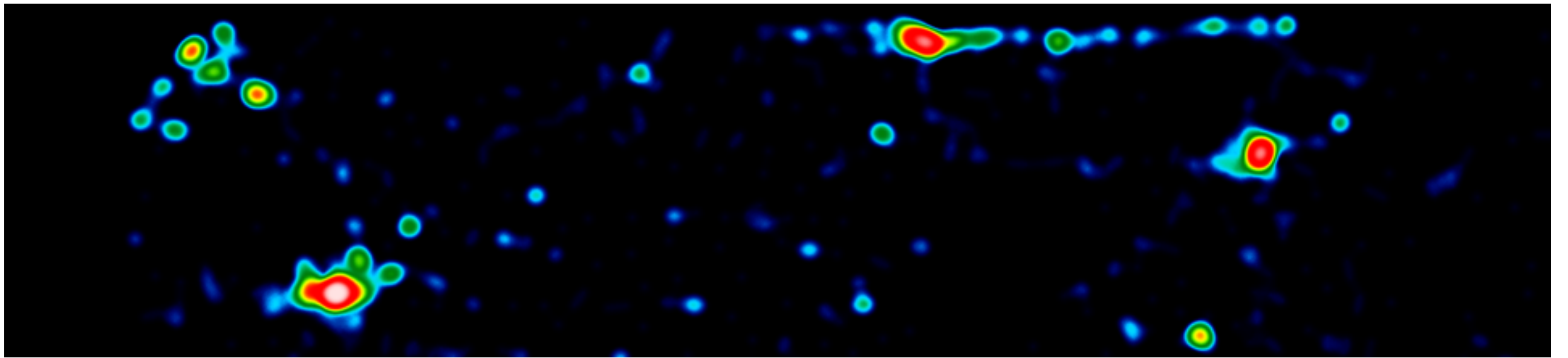


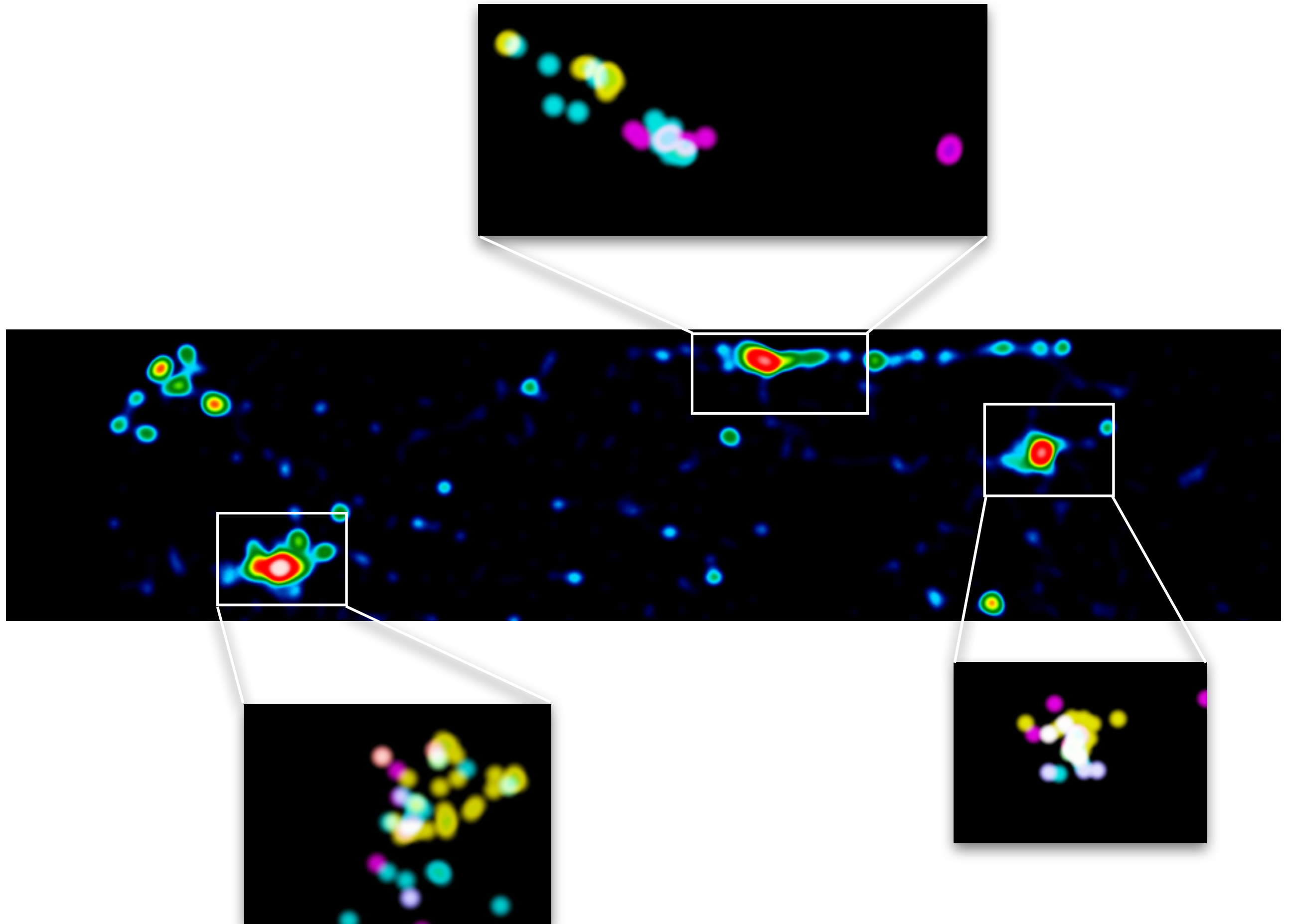




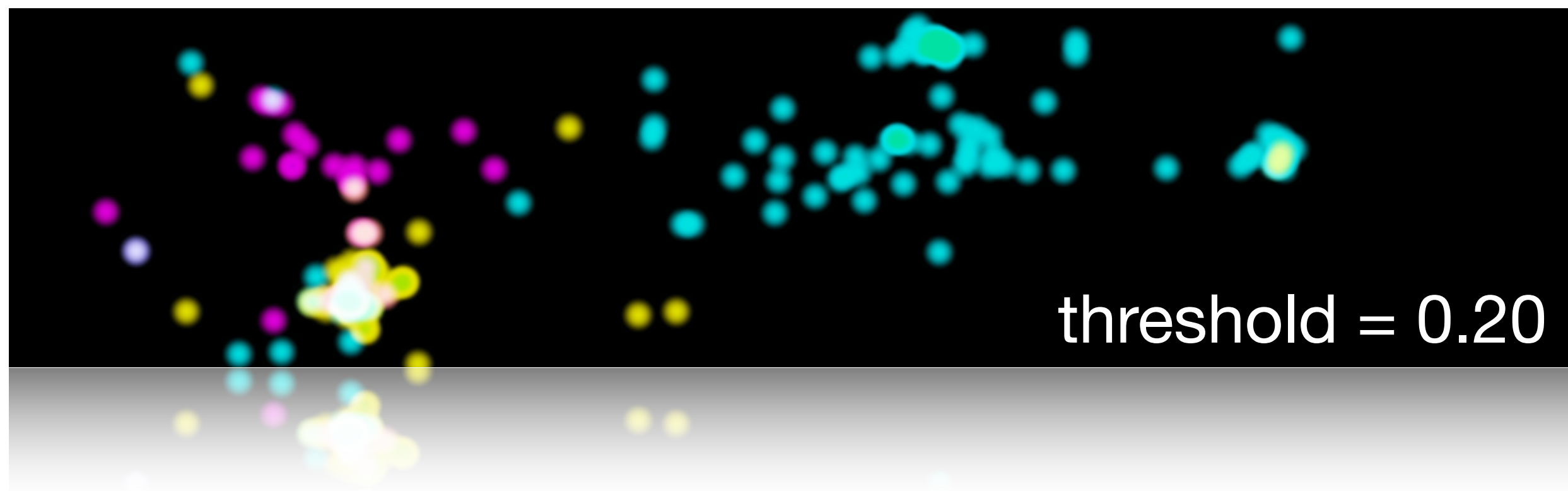
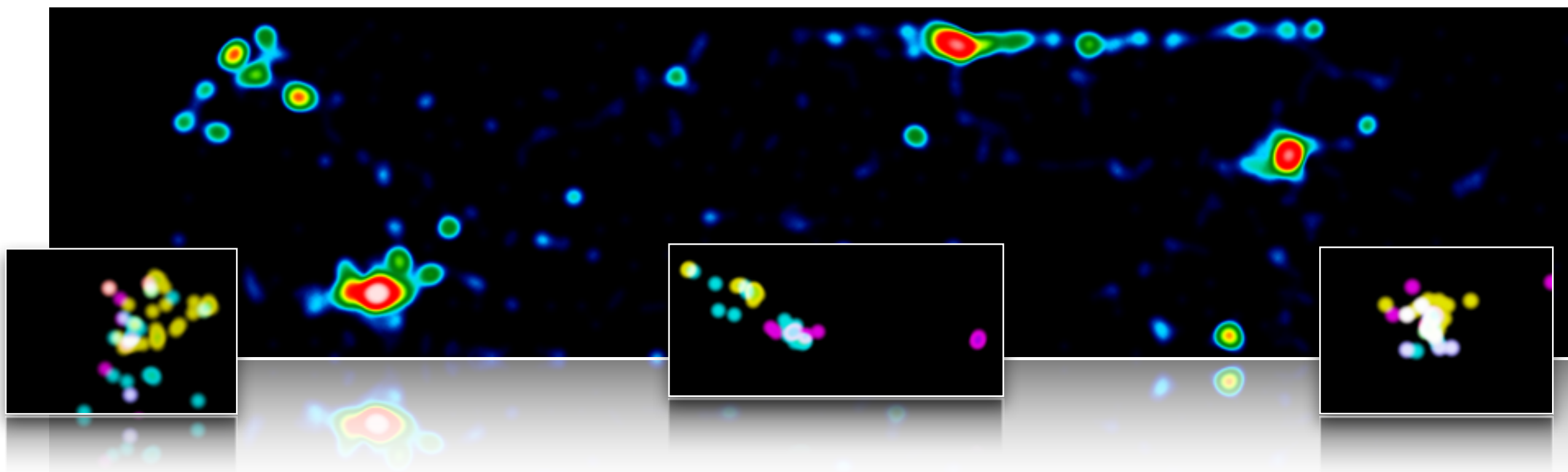
~600k nodes  
~3M edges

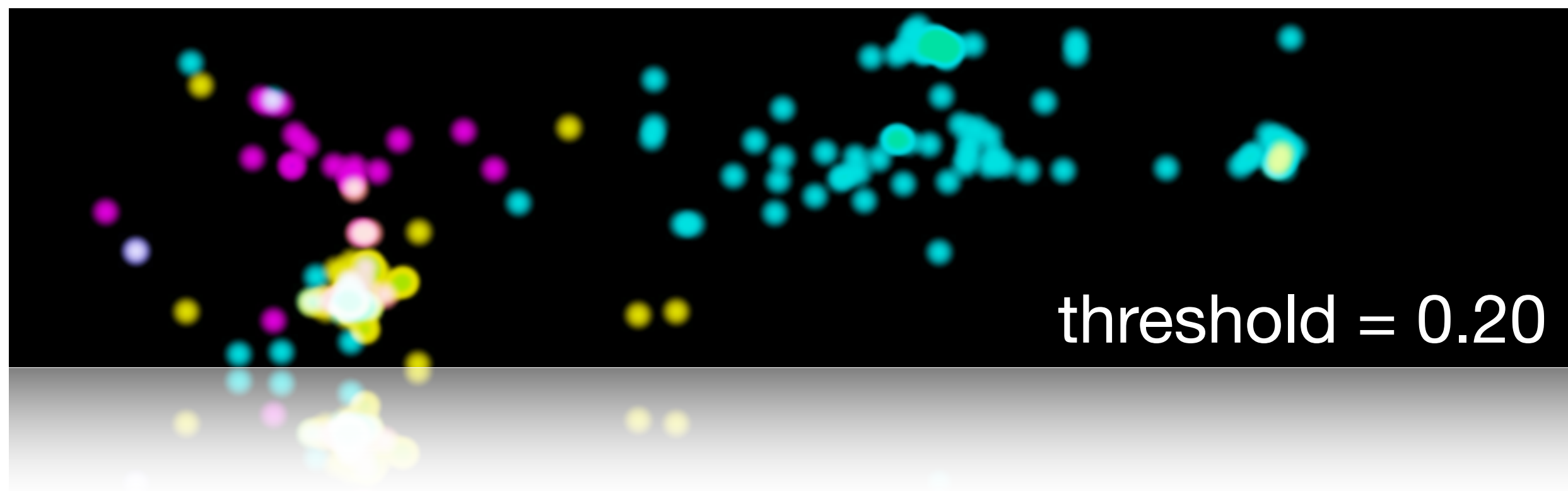
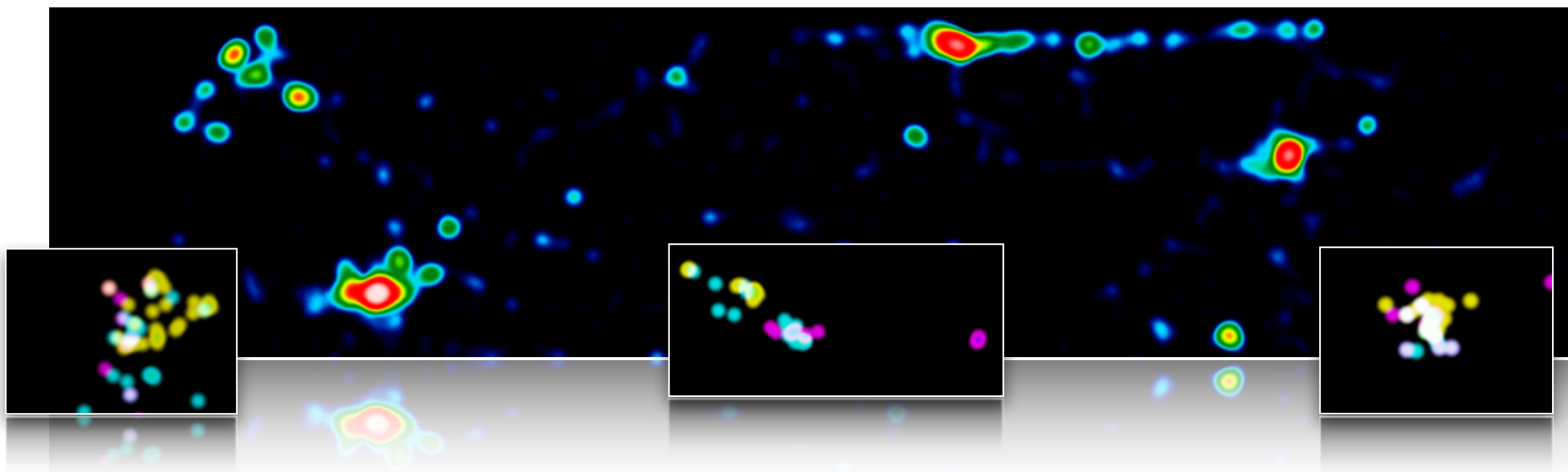


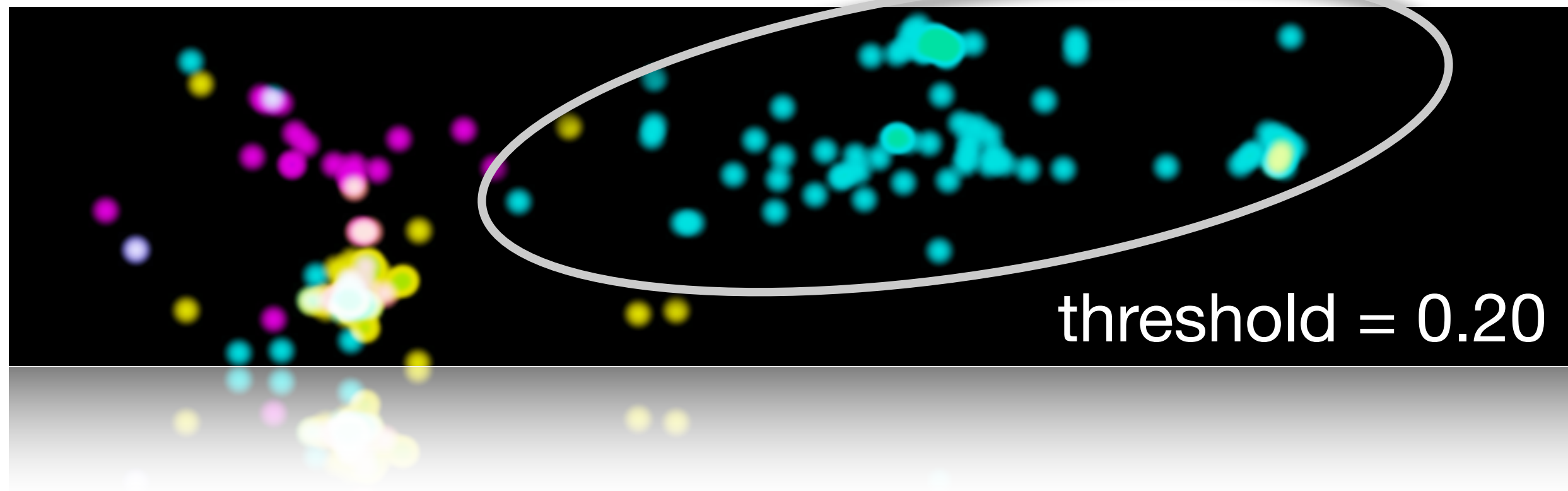
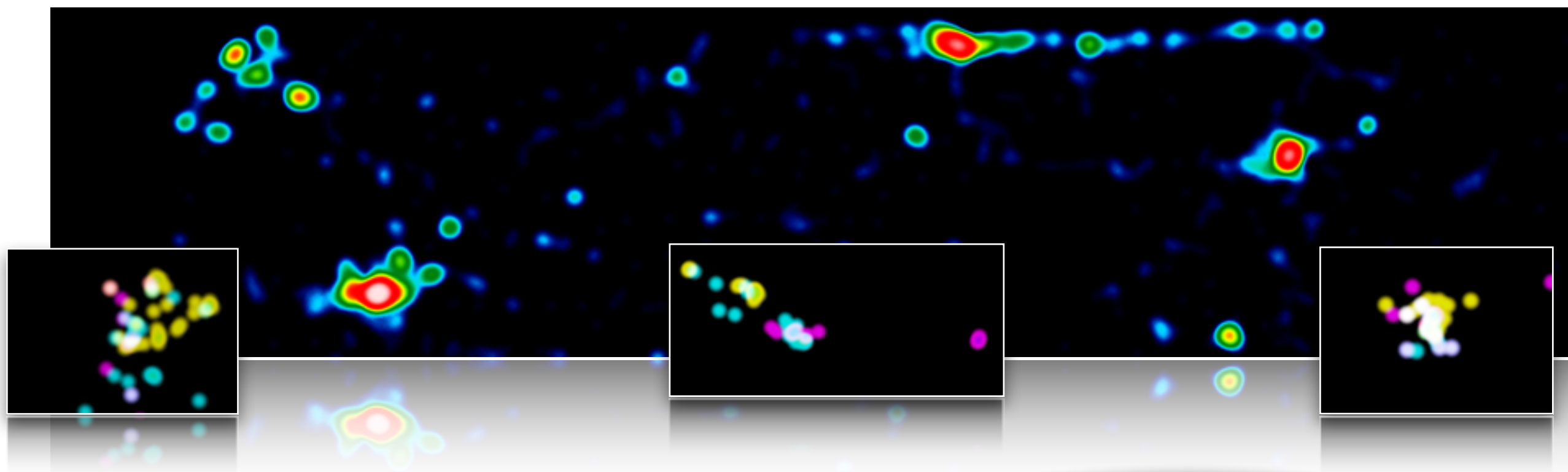


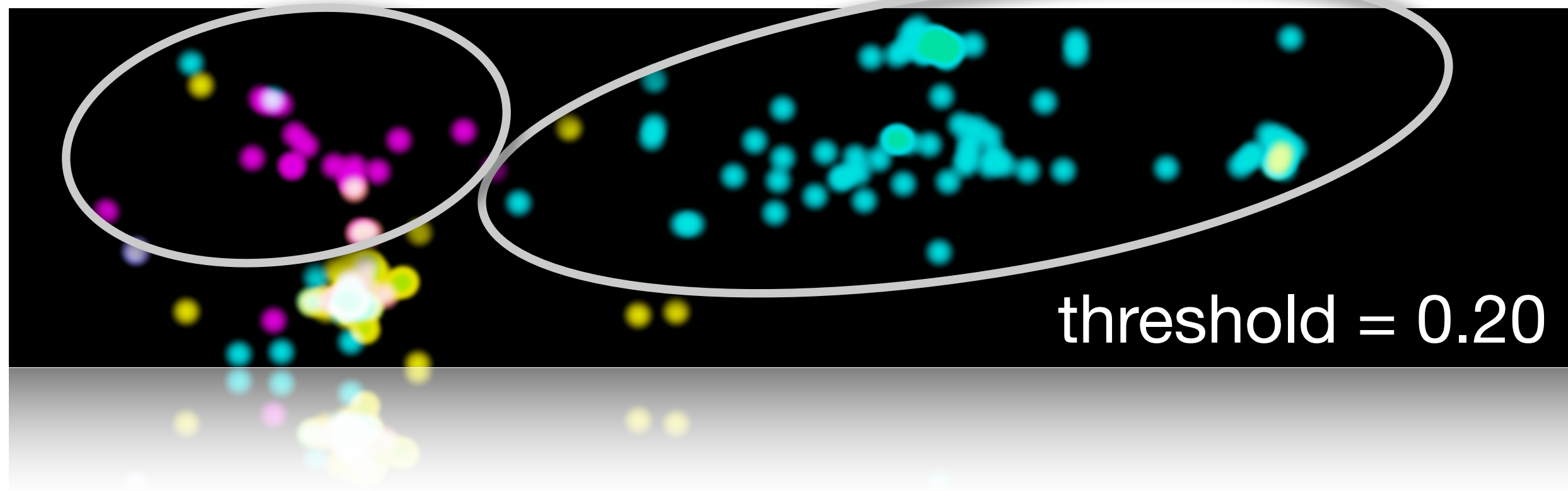
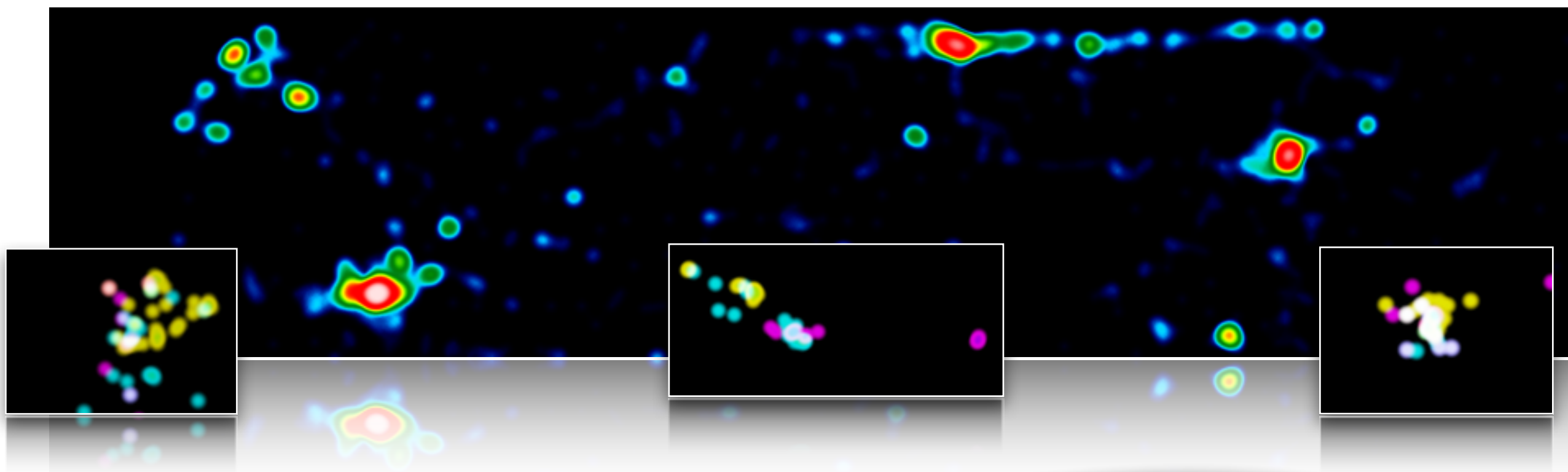




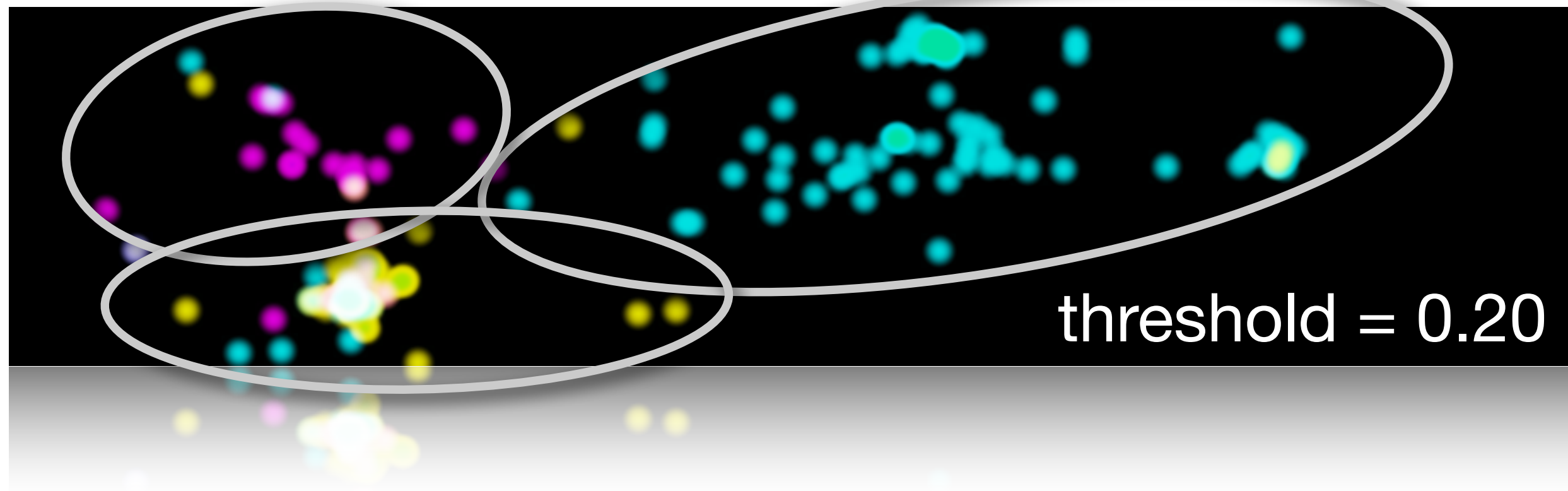
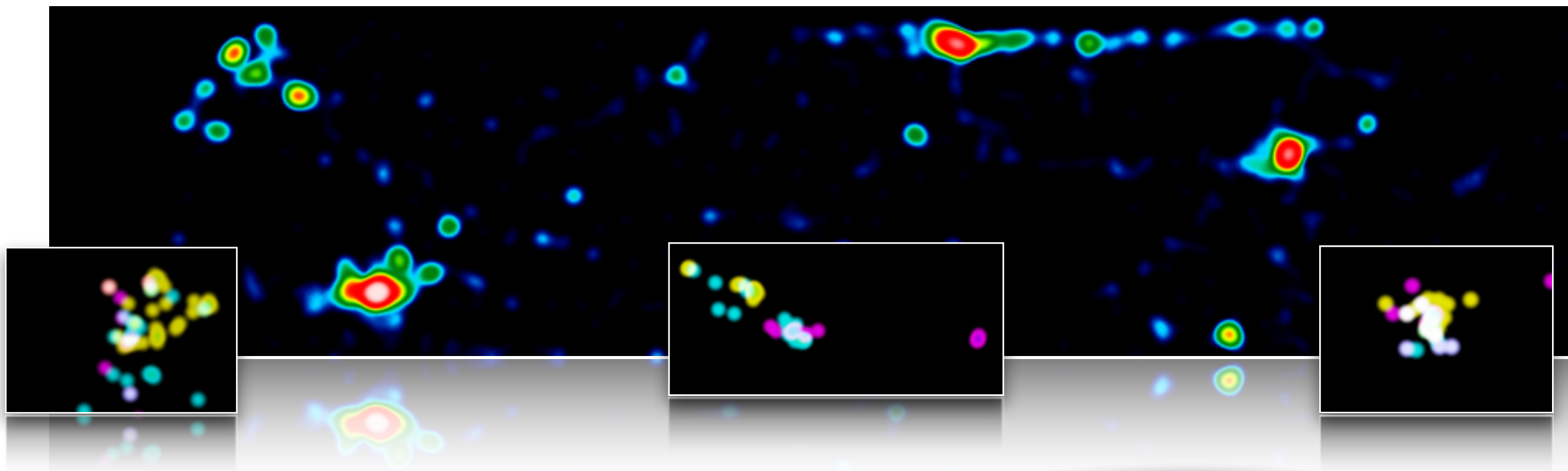




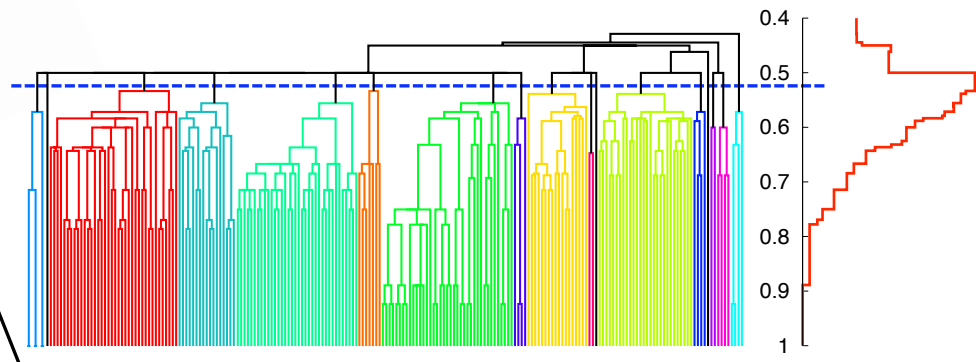
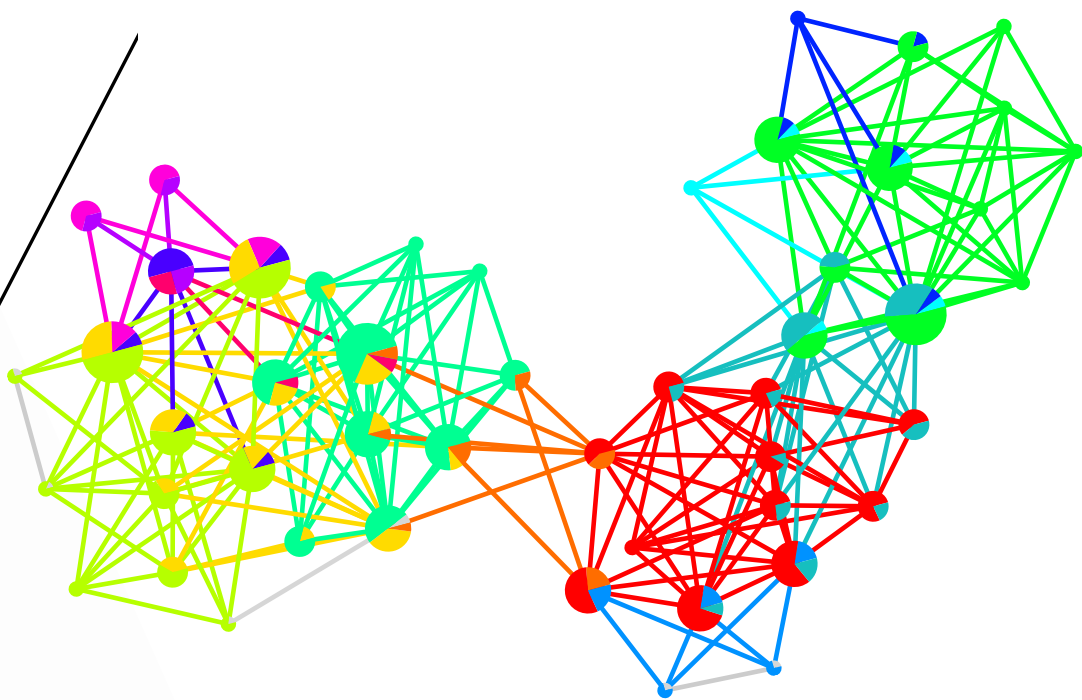
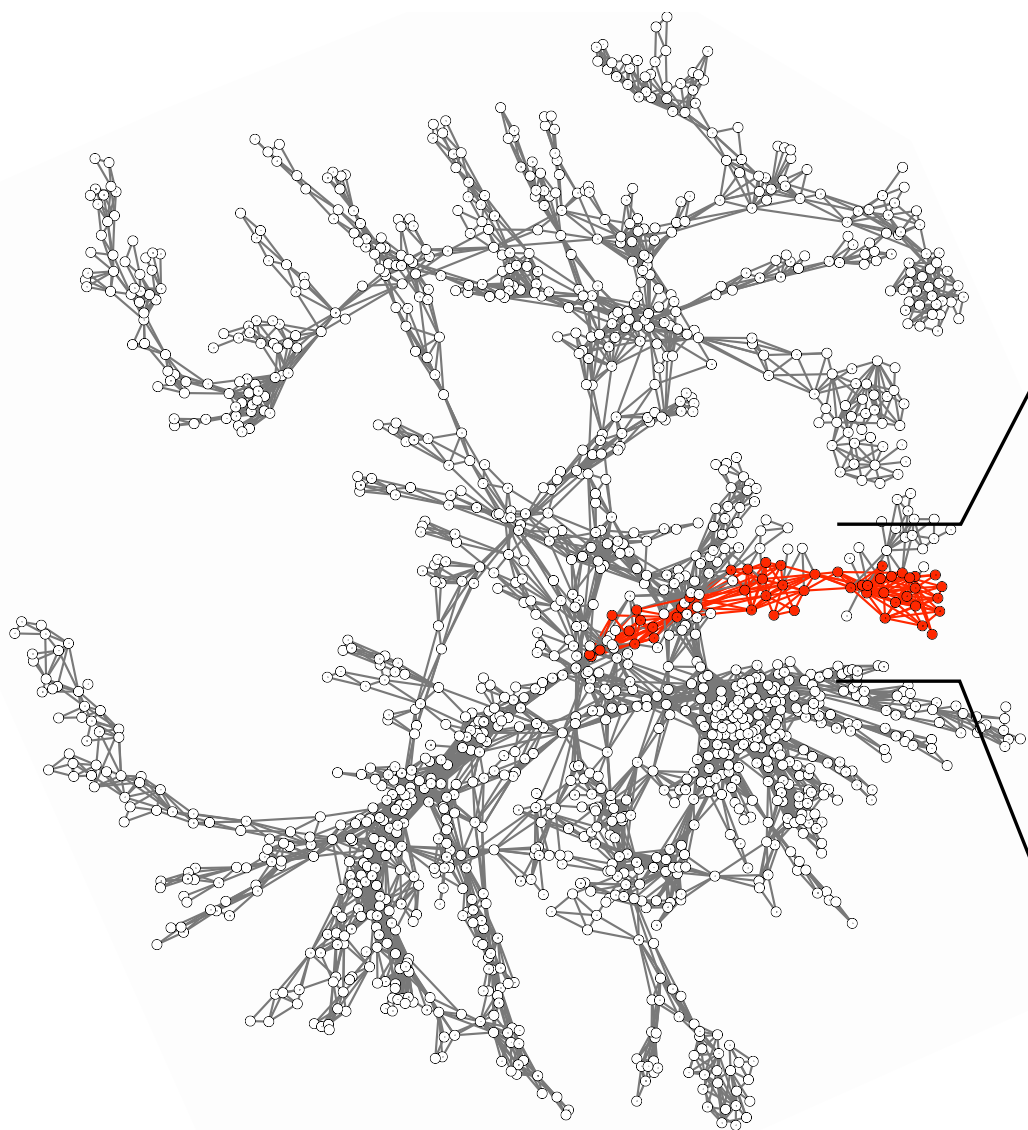
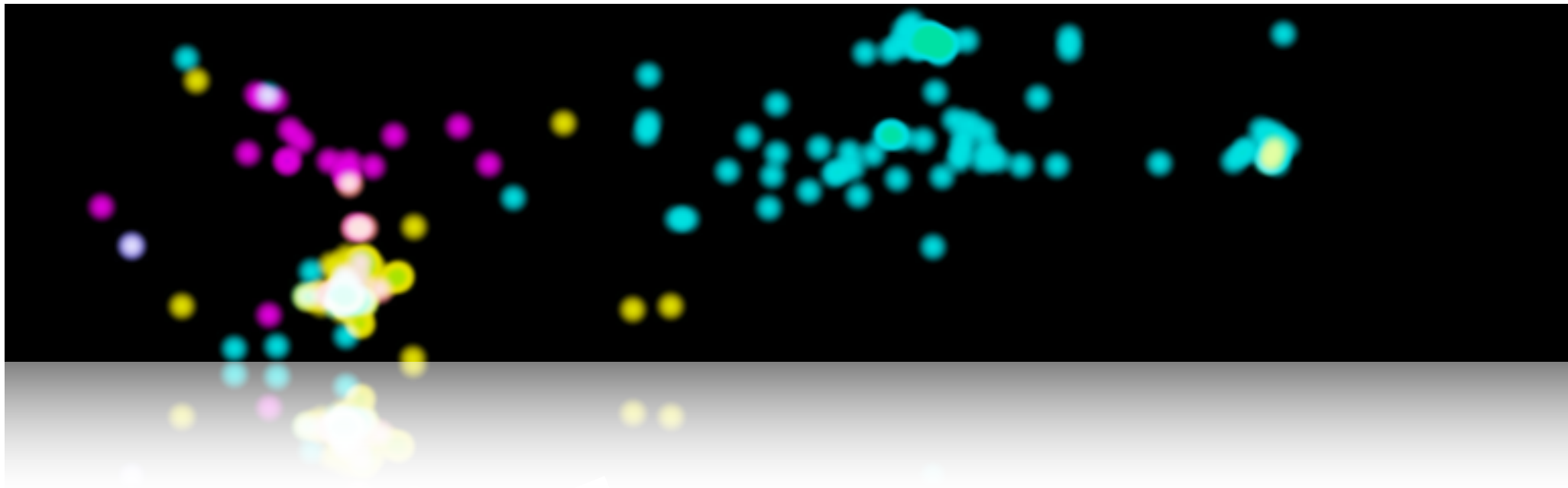








threshold = 0.20



# Resources

- <http://yongyeol.com/courses/2012S-I590/>
- [http://yongyeol.com/w/index.php?title=Network\\_science](http://yongyeol.com/w/index.php?title=Network_science)
- [http://en.wikipedia.org/wiki/Network\\_science](http://en.wikipedia.org/wiki/Network_science)
  
- <http://yongyeol.com>
- yyahn@indiana.edu