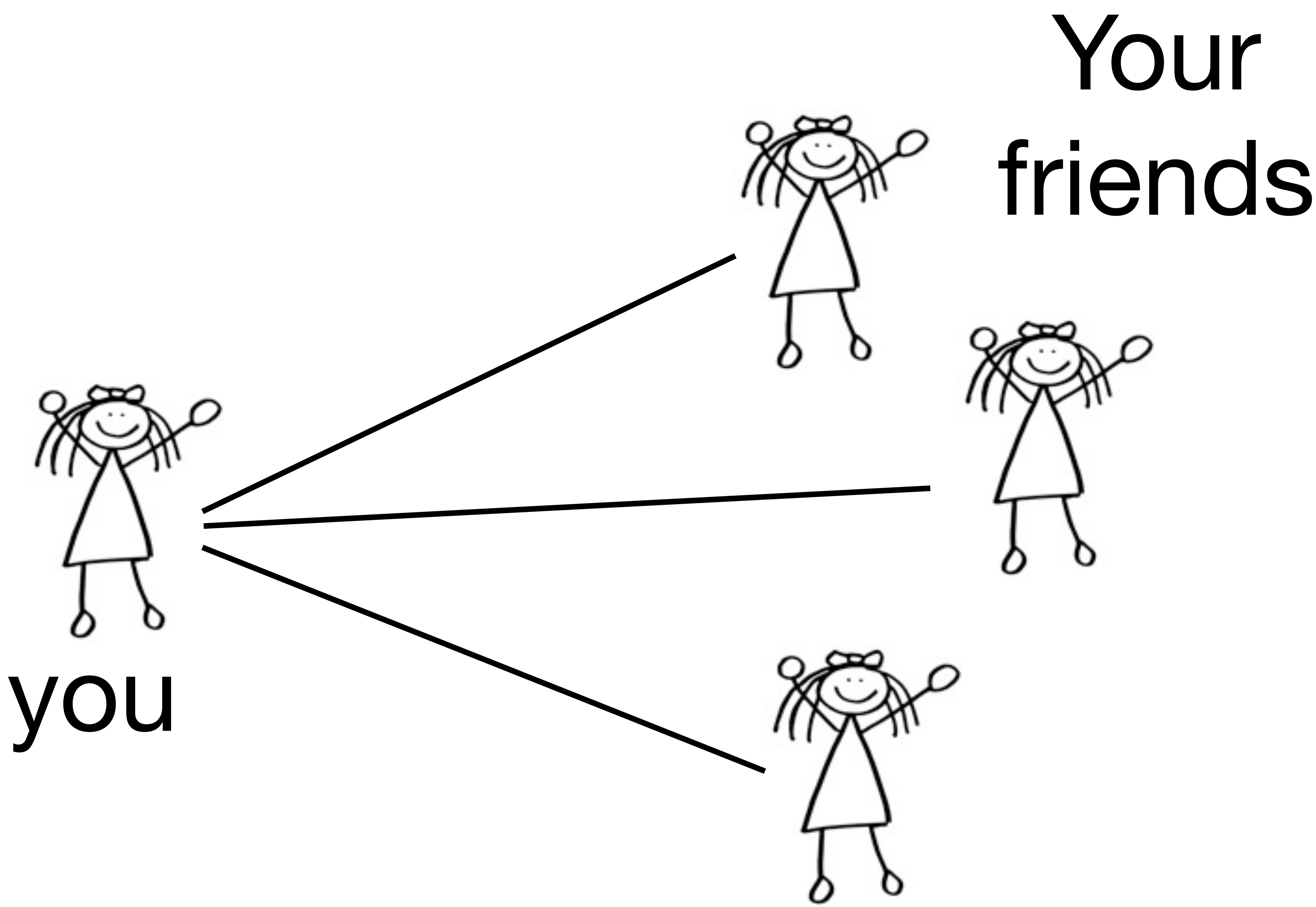
Introduction to Network Science

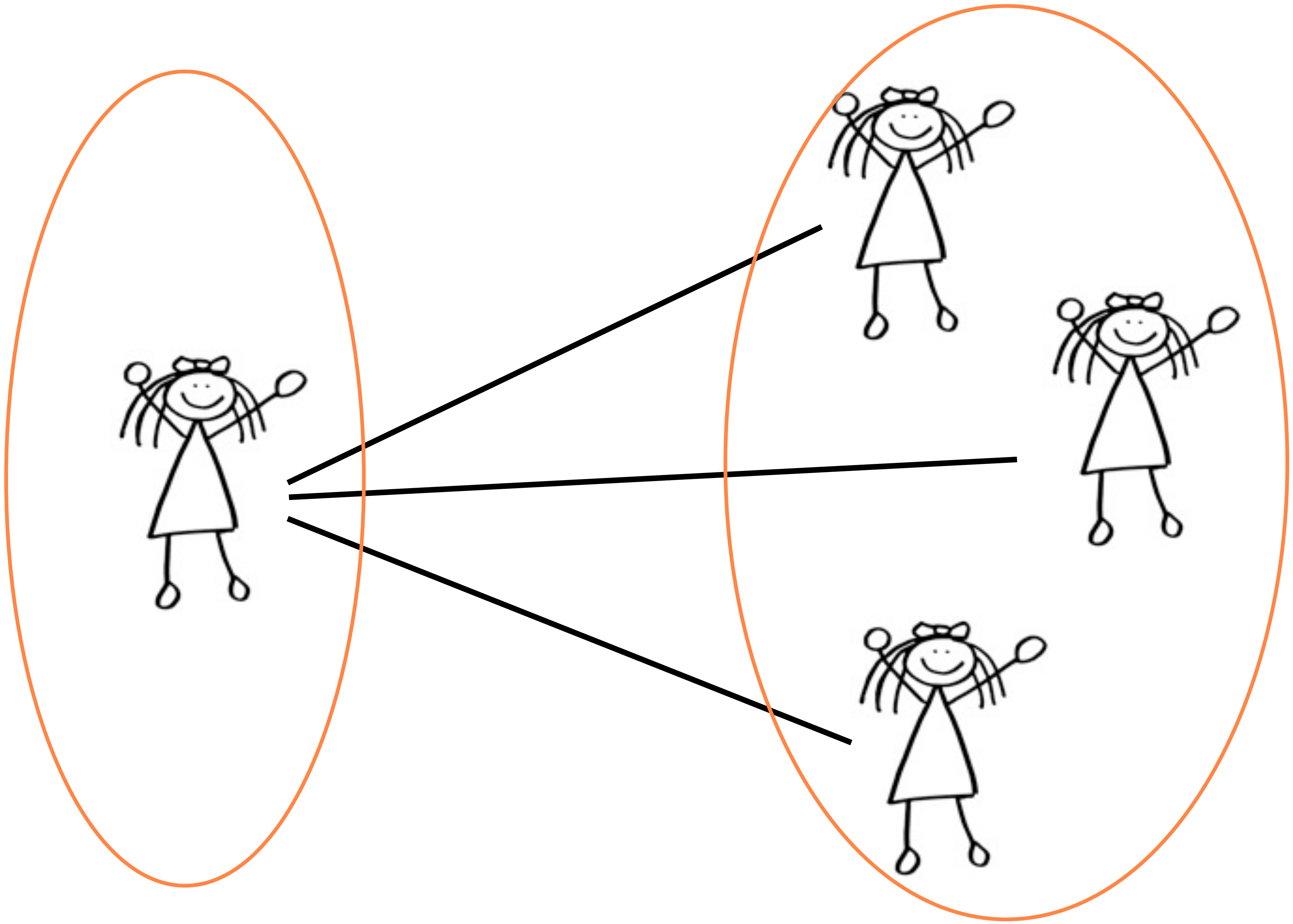
Yong-Yeol "YY" Ahn

SCHOOL OF INFORMATICS AND COMPUTING

INDIANA UNIVERSITY

Bloomington


Your
friends
you



Q: Who are more popular, you or your friends (on average)?

1. You

2. Same

3. Your friends

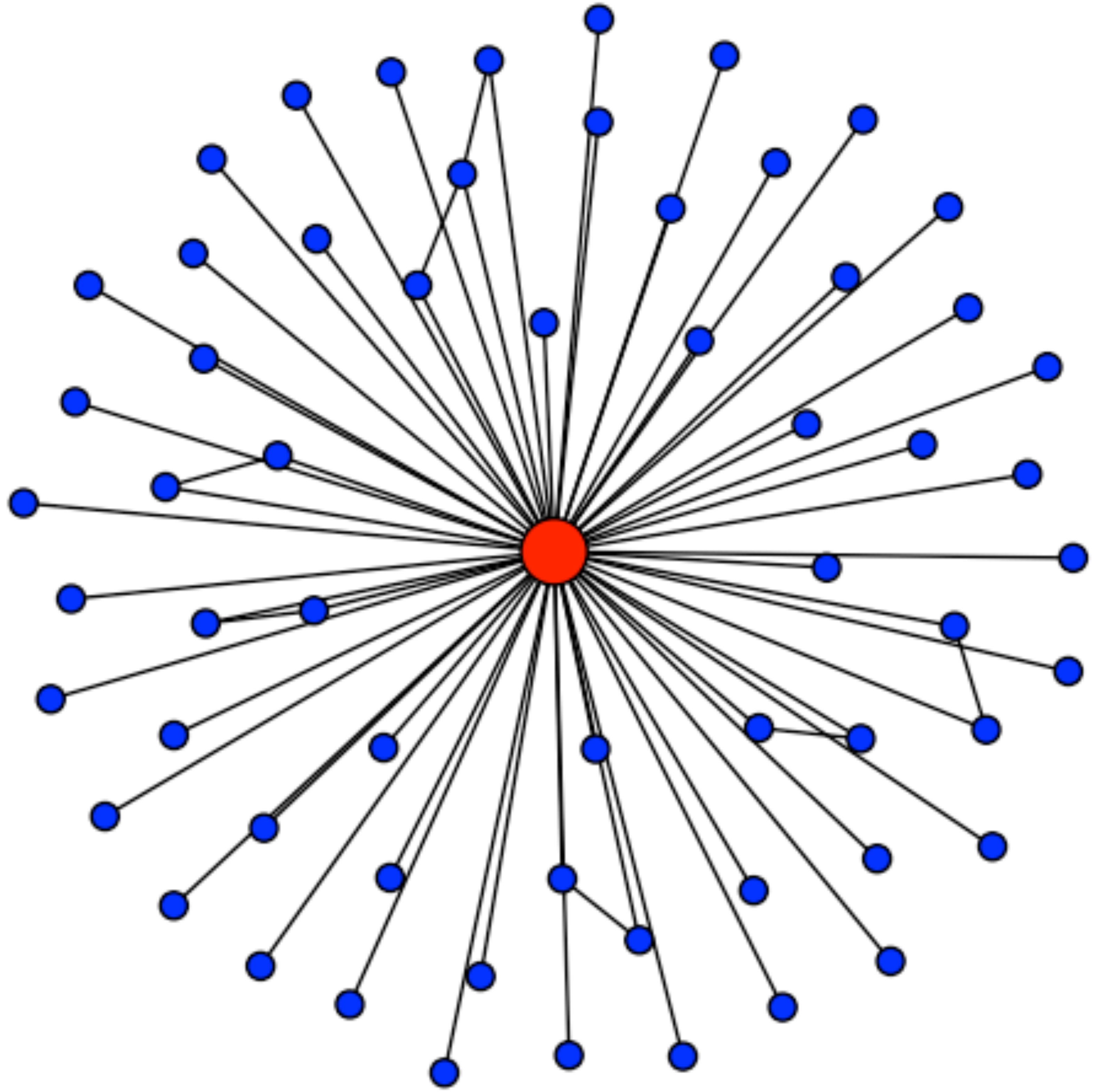
1. You

2. Same

3. Your friends

How can ***everyone*** feel that his/her friends are more popular?

“Friendship Paradox”



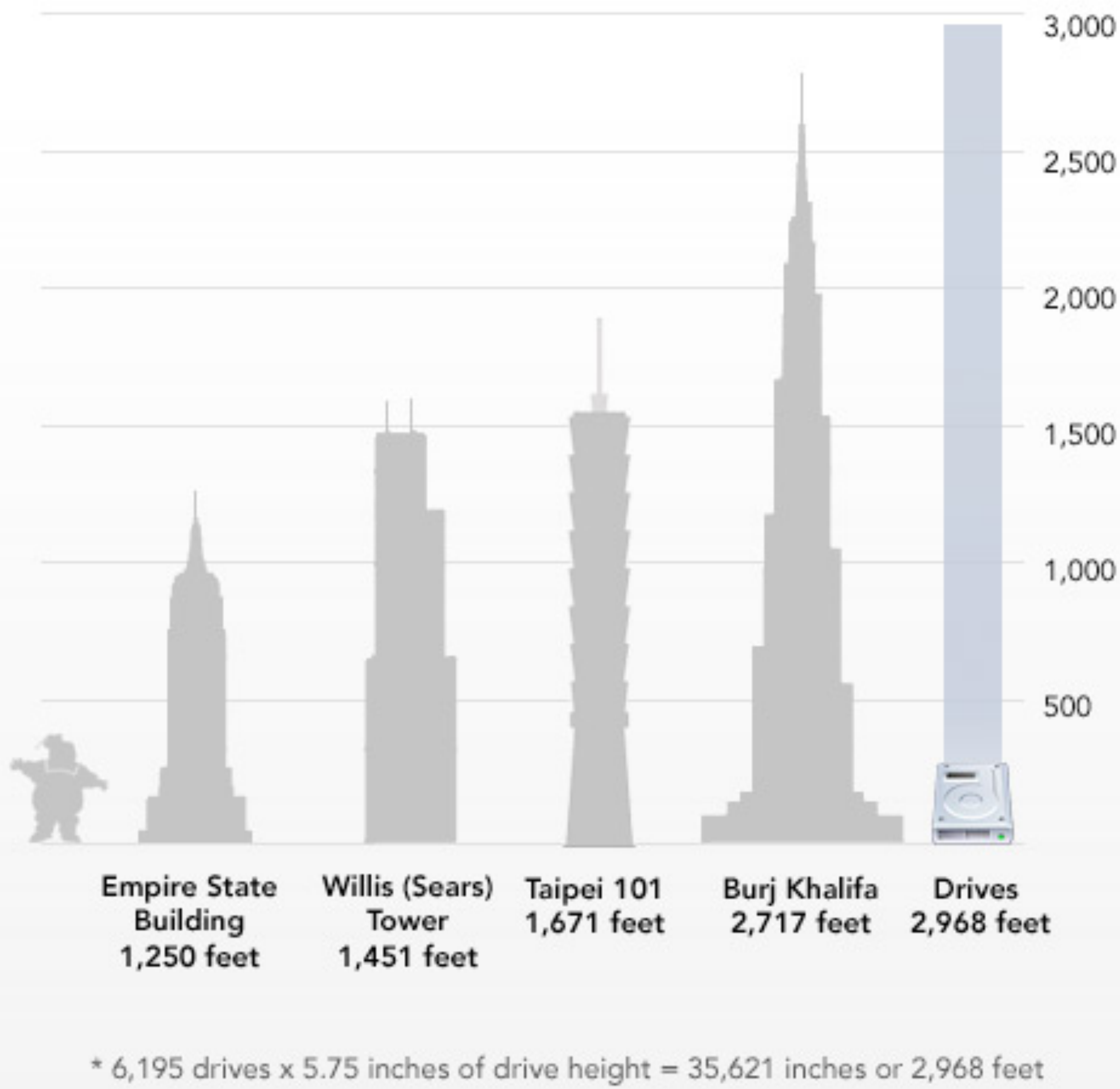
When are we living?

Google processes

20+ petabytes

per day

10 petabytes



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+
facebook	900,000,000+
	500,000,000+
	300,000,000+

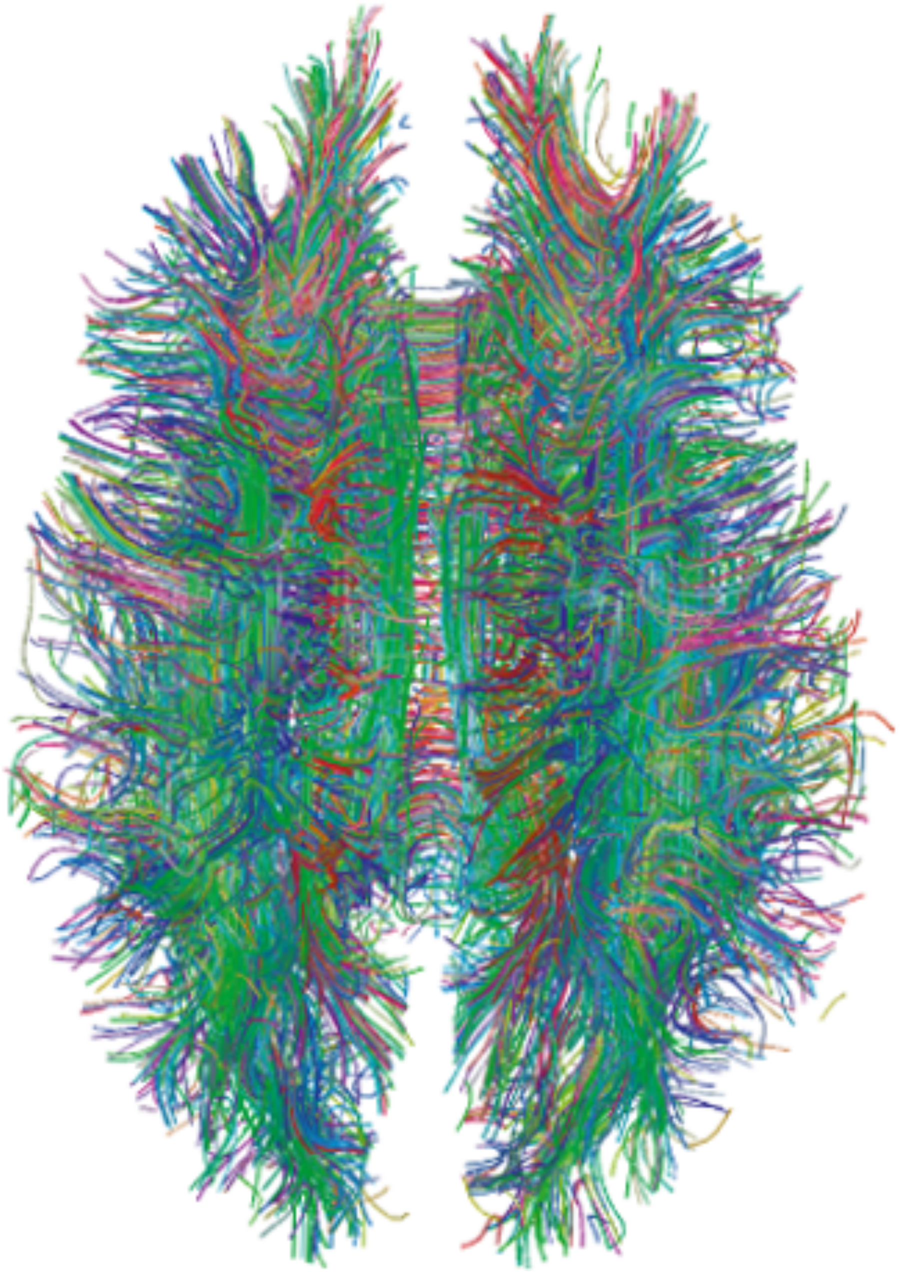
Billions of people

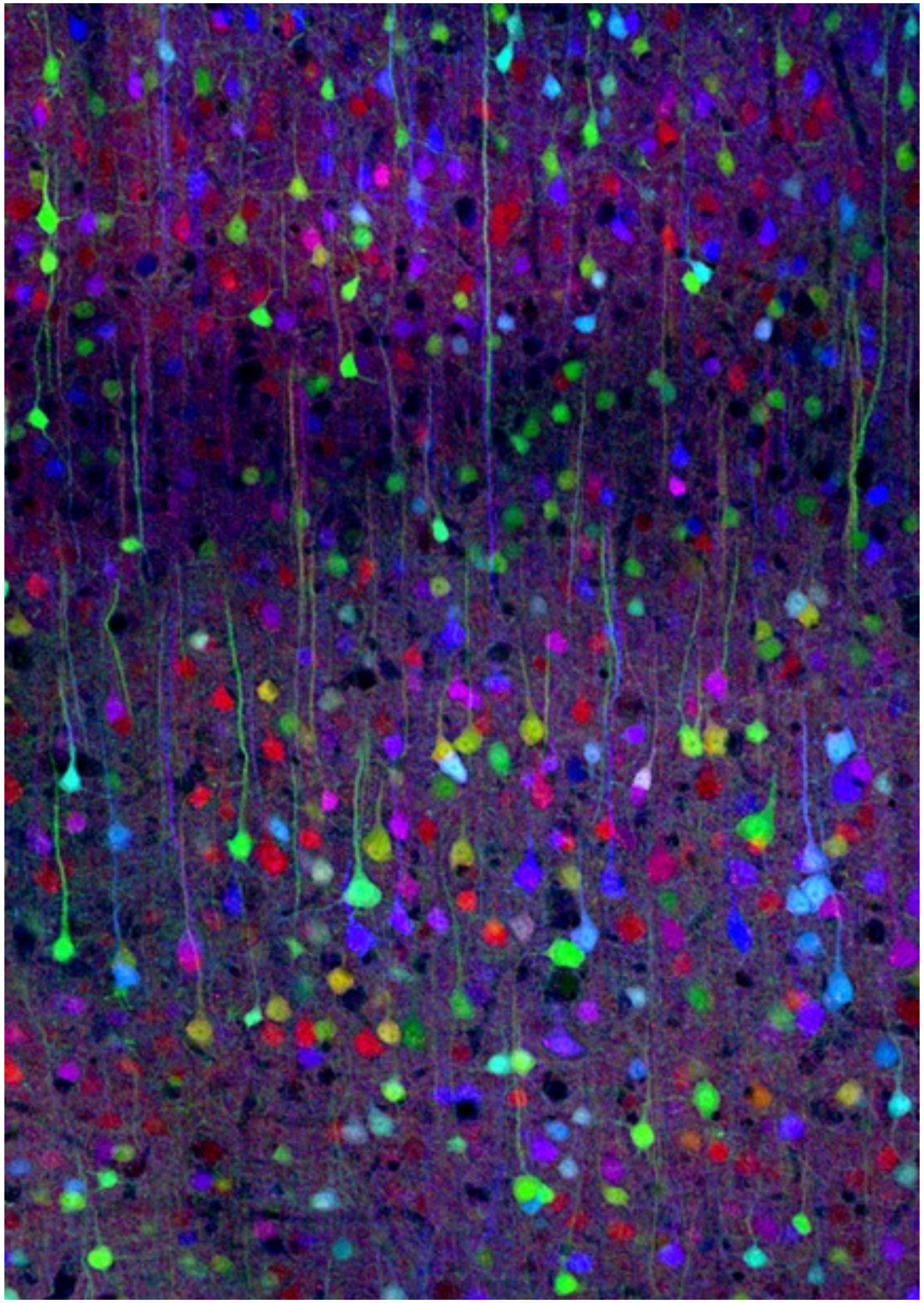
recording their social life

in **Bits**.


40421551 40421561 40421571 40421581 40421591 40421601 40421611 40421621 40421631 40421641 40421651 40421661 40421671 40421681


40421551 40421561 40421571 40421581 40421591 40421601 40421611 40421621 40421631 40421641 40421651 40421661 40421671 40421681
$1000





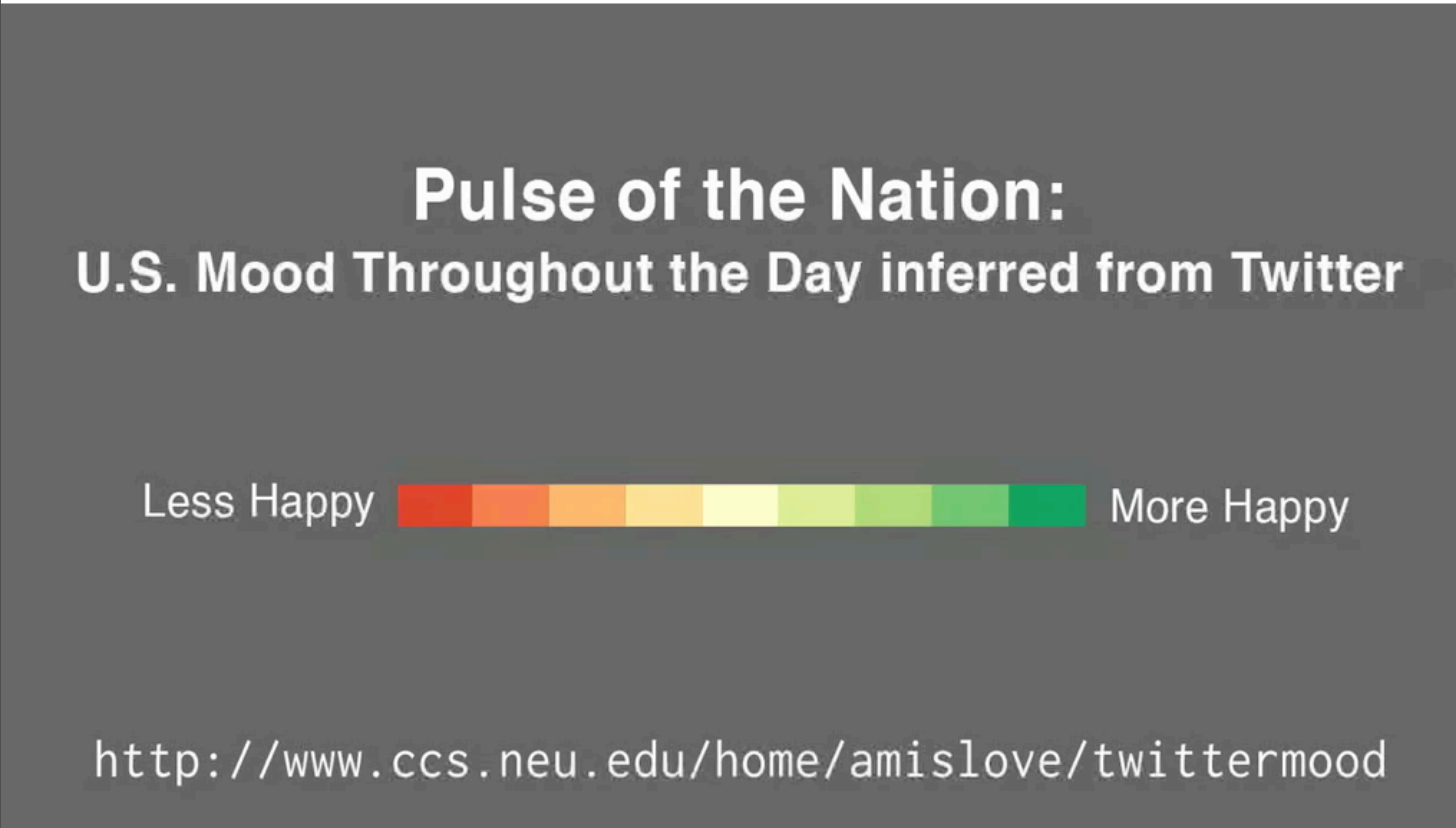
BIG DATA

BIG DATA

$\downarrow$

INFORMATION


GOLD
"SWISS PRALINES"
NET WT
300g - 10.5oz
GOLDKENN


Pulse of the Nation:
U.S. Mood Throughout the Day inferred from Twitter
Less Happy
More Happy
http://www.ccs.neu.edu/home/amislove/twittermood

BIG DATA

SOCIETY

LIFE

BIG DATA

ECONOMY

LIFE
SOCIETY
ECONOMY

BIG DATA

LIFE
SOCIETY
ECONOMY

BIG DATA

COMPLEX
SYSTEMS

COMPLEX SYSTEMS

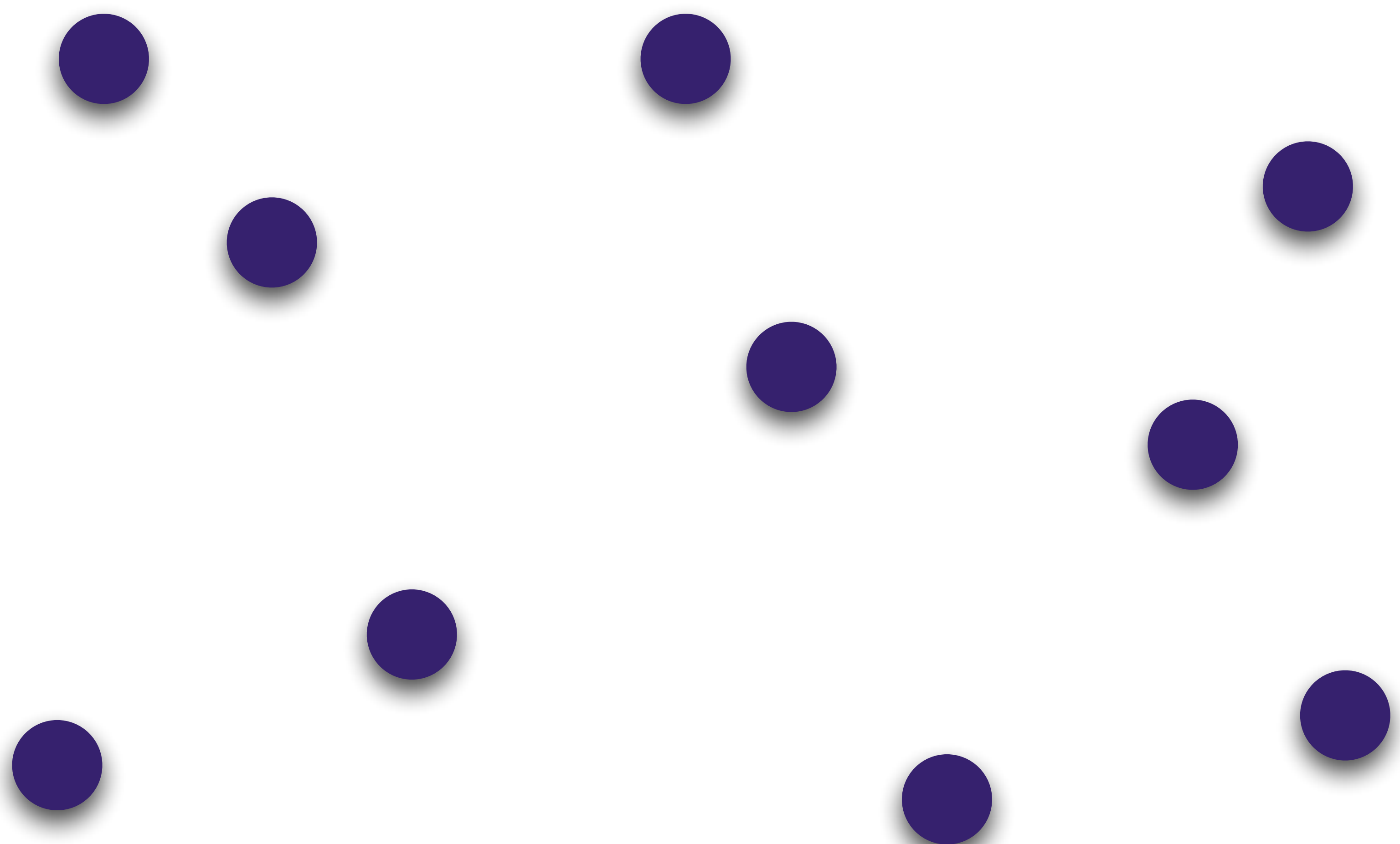
COMPLEX SYSTEMS

MANY parts,

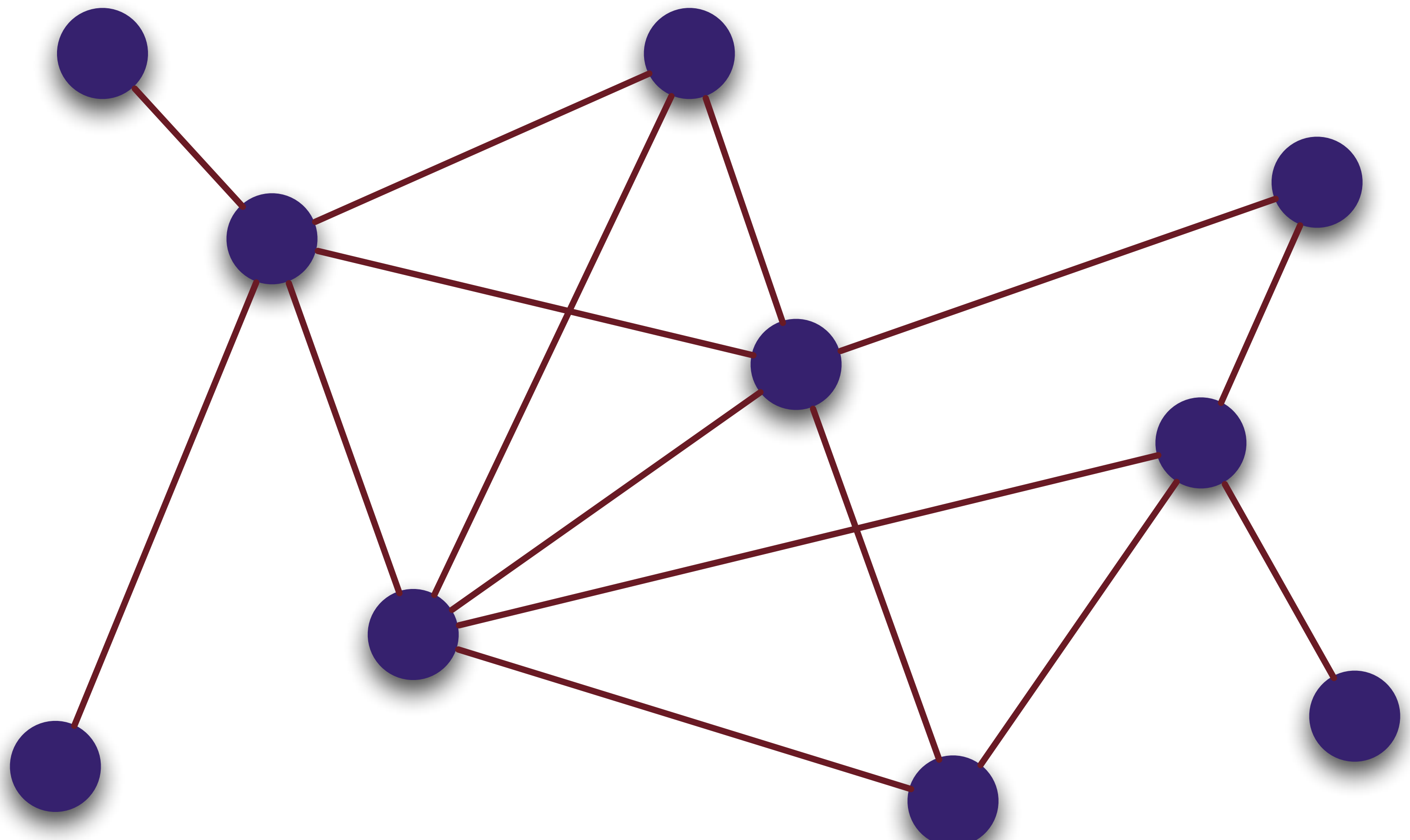
INTERACTING with each other

in **NON-TRIVIAL WAYS**

NETWORKS

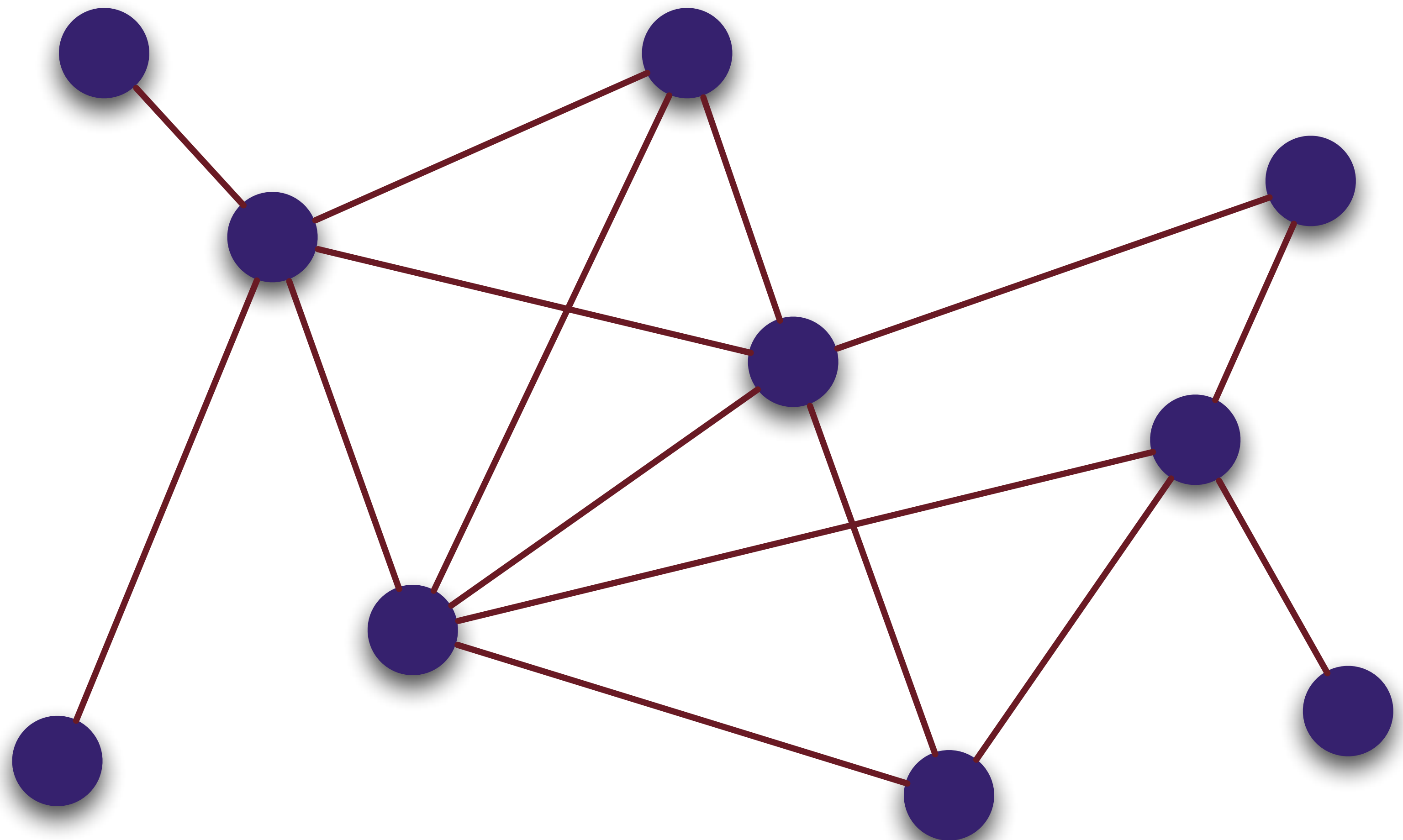


Nodes

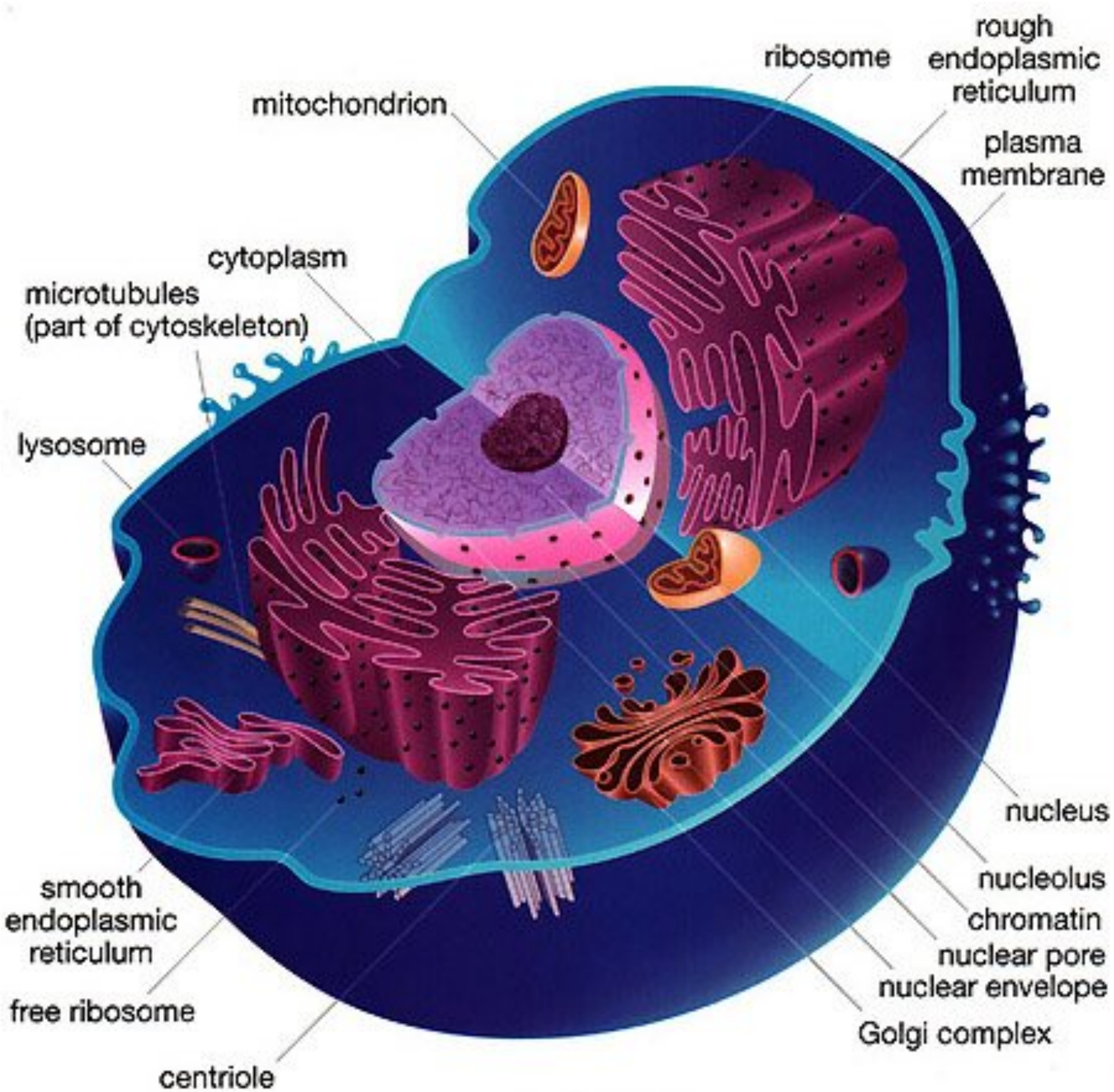


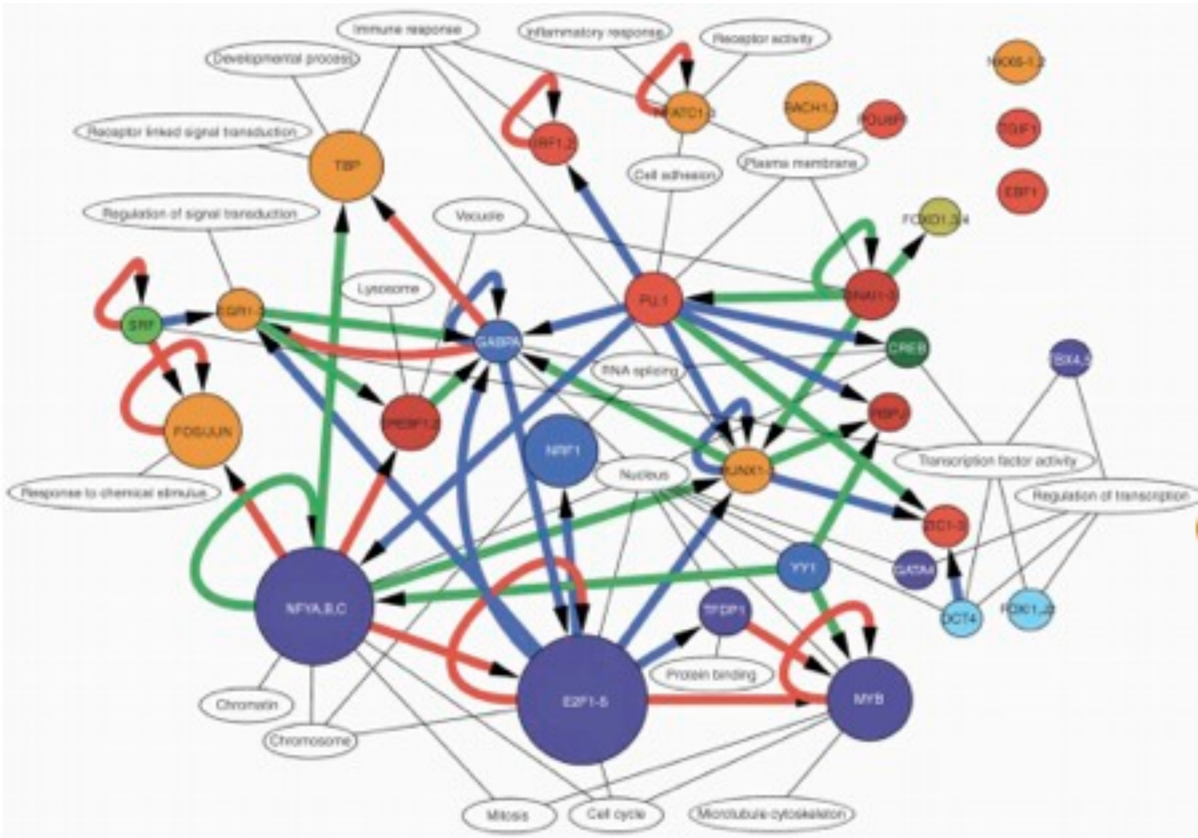
Links (edges) between nodes

Degree: # of neighbors

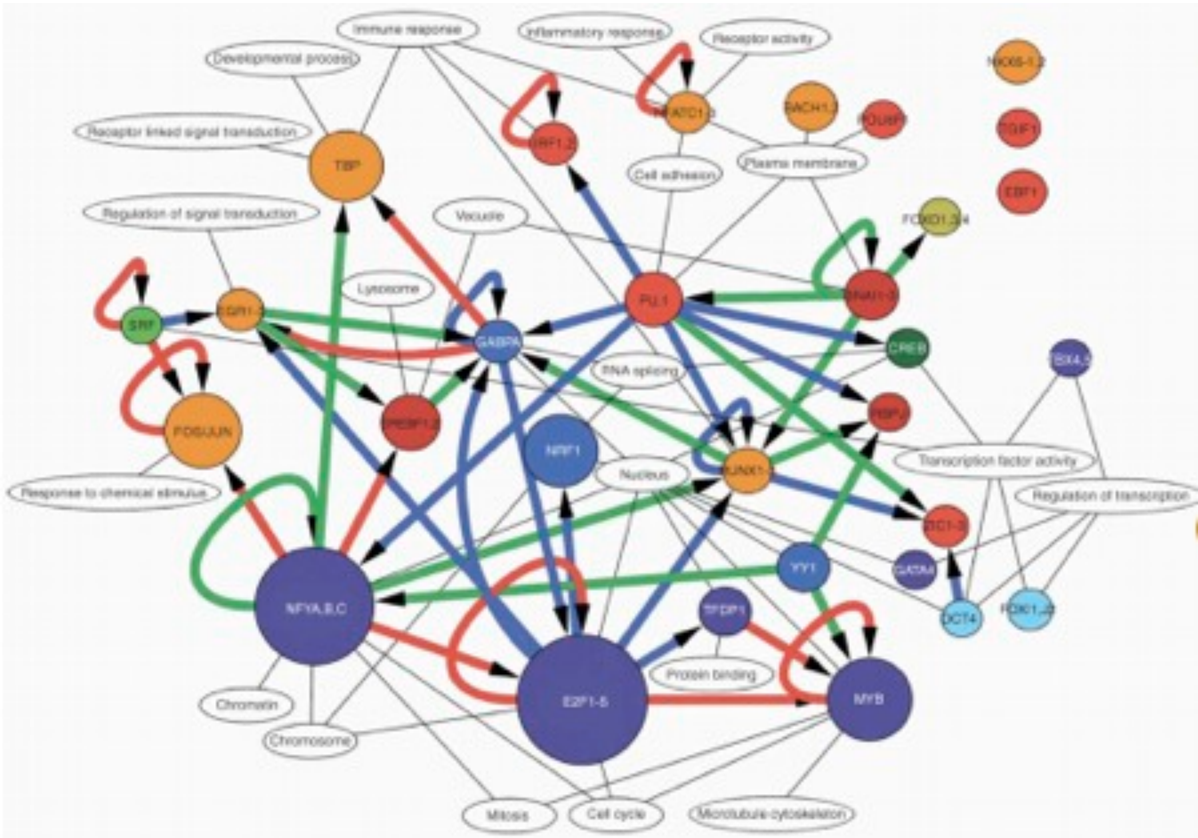


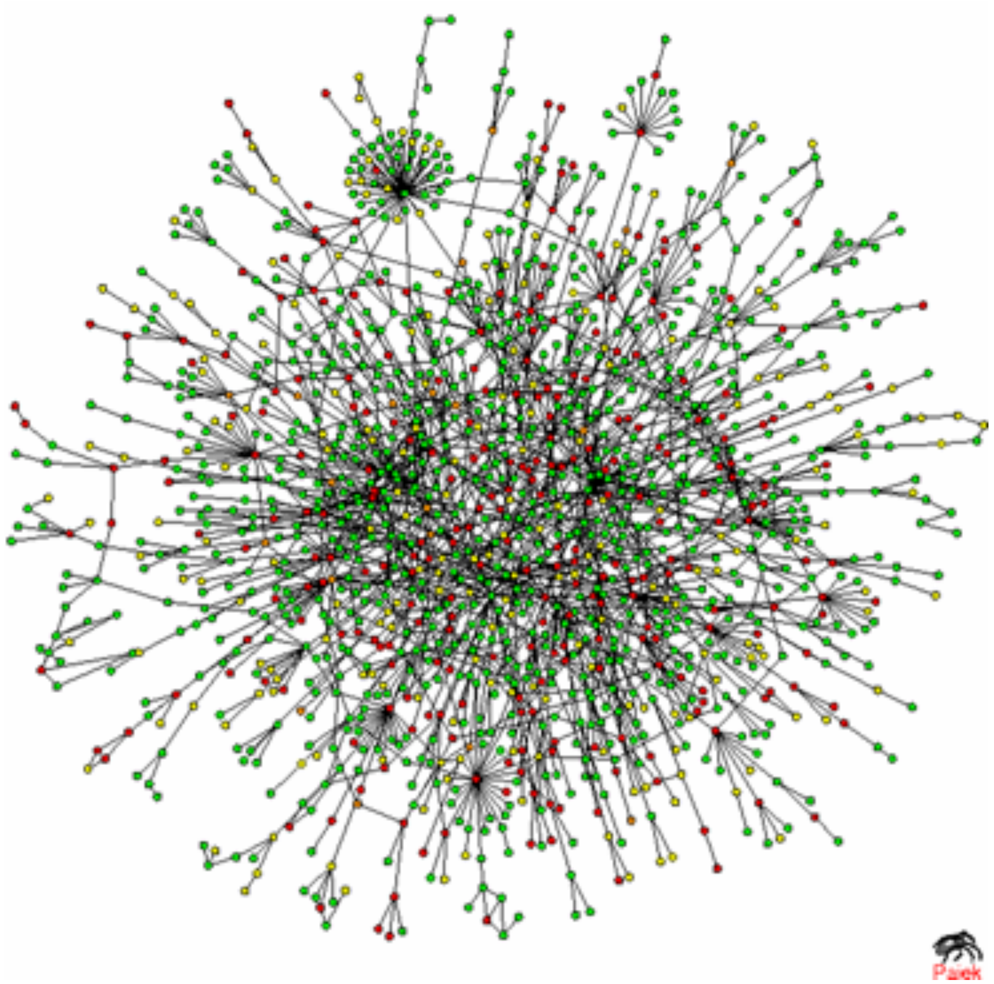
Links (edges) between nodes

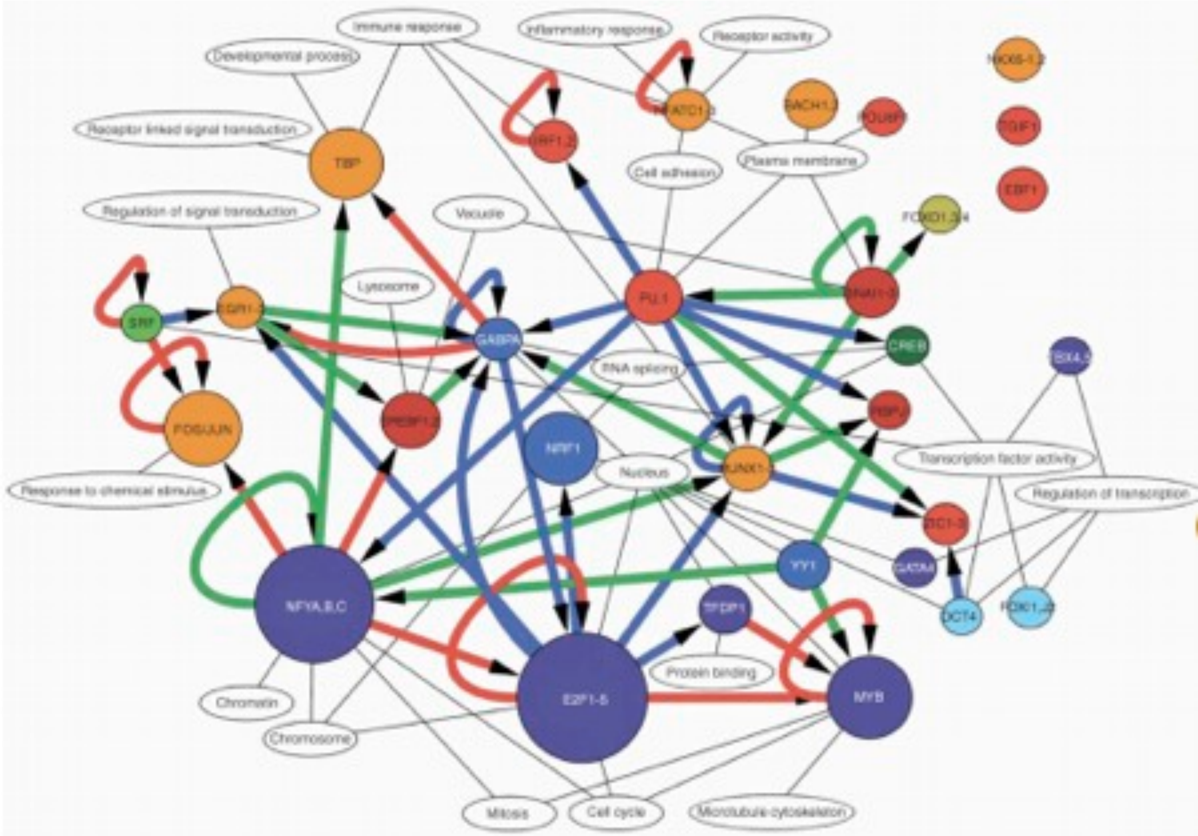

ribosome
rough endoplasmic reticulum
mitochondrion
plasma membrane
cytoplasm
microtubules (part of cytoskeleton)
lysosome
nucleus
nucleolus
chromatin
nuclear pore
nuclear envelope
Golgi complex
smooth endoplasmic reticulum
free ribosome
centriole

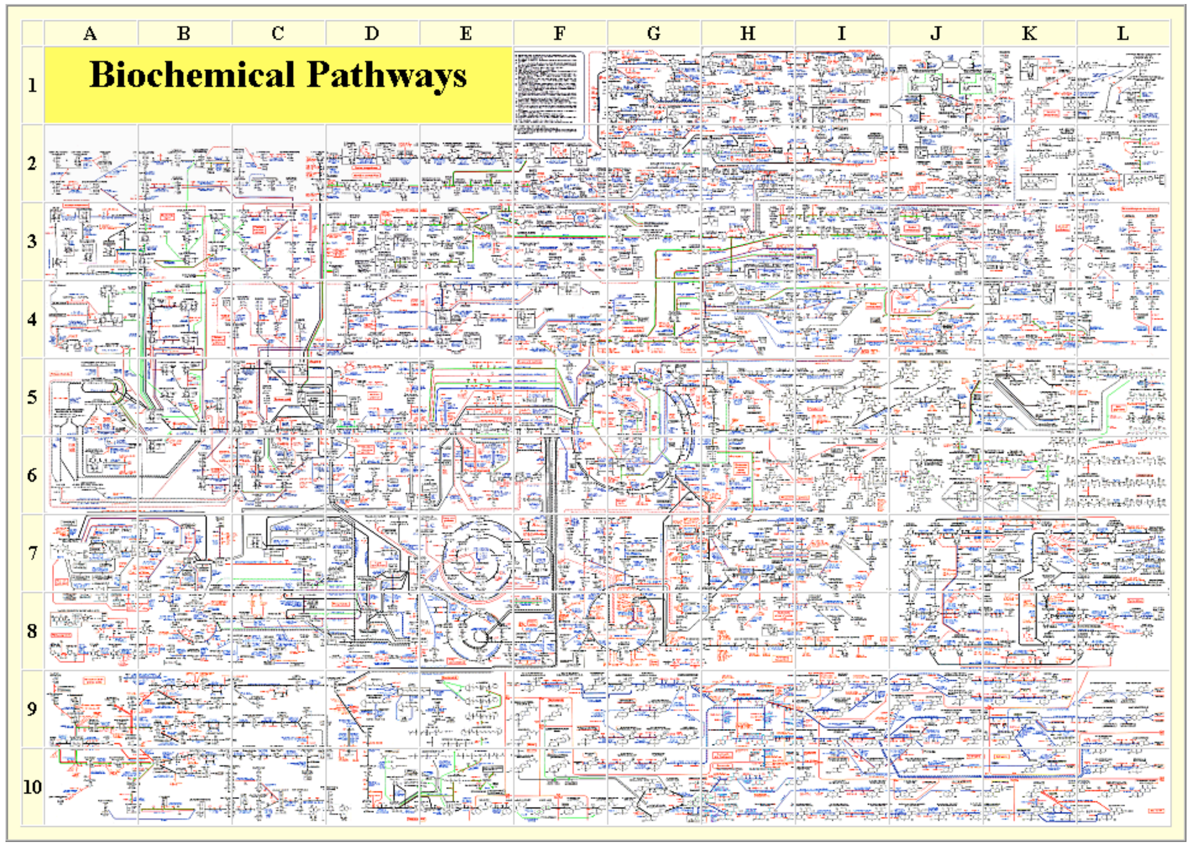


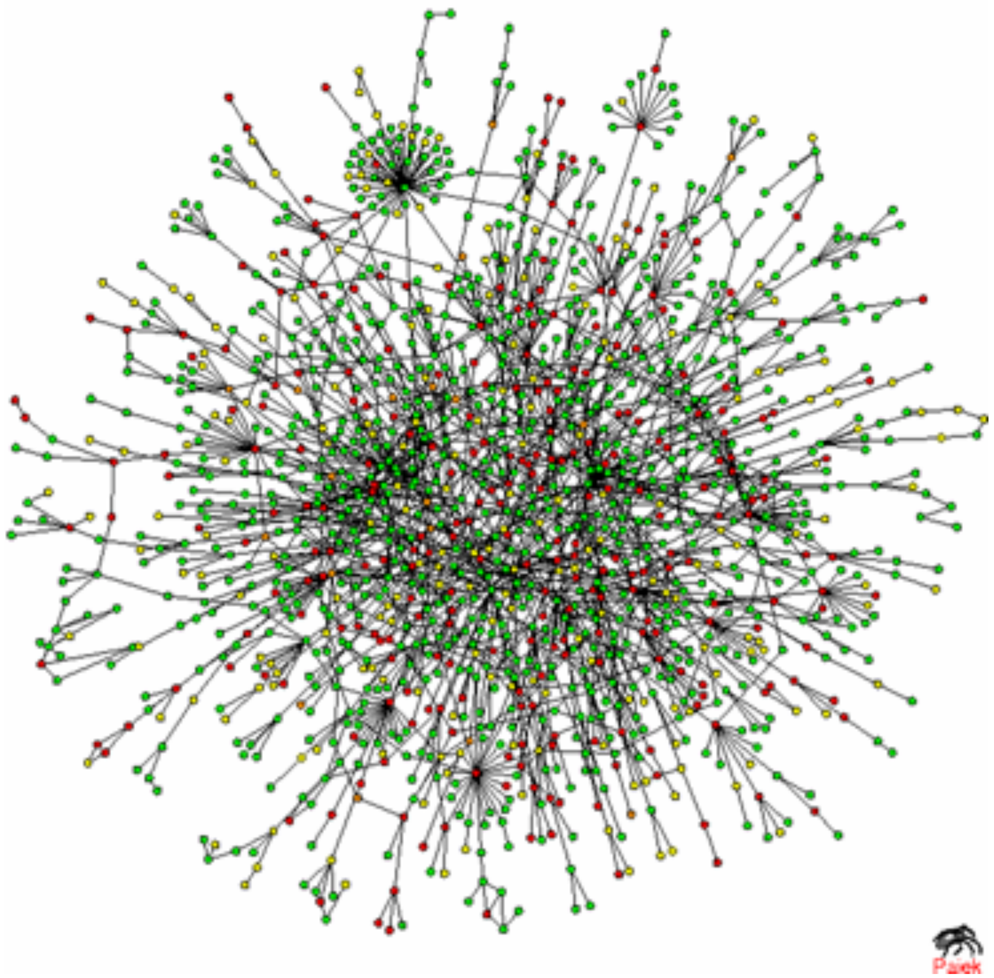
Immune response
Inflammatory response
Receptor activity
Developmental process
Receptor linked signal transduction
TBP
Cell adhesion
Plasma membrane
EBF1
Regulation of signal transduction
Vacuole
Lysosome
SRF
PU.1
GABPA
CREB
RNA splicing
NRF1
Nucleus
Transcription factor activity
Regulation of transcription
Response to chemical stimulus
NFYA,B,C
YY1
GATA4
OCT4
TFDP1
Protein binding
E2F1-6
MYB
Chromatin
Chromosome
Mitosis
Cell cycle
Microtubule cytoskeleton

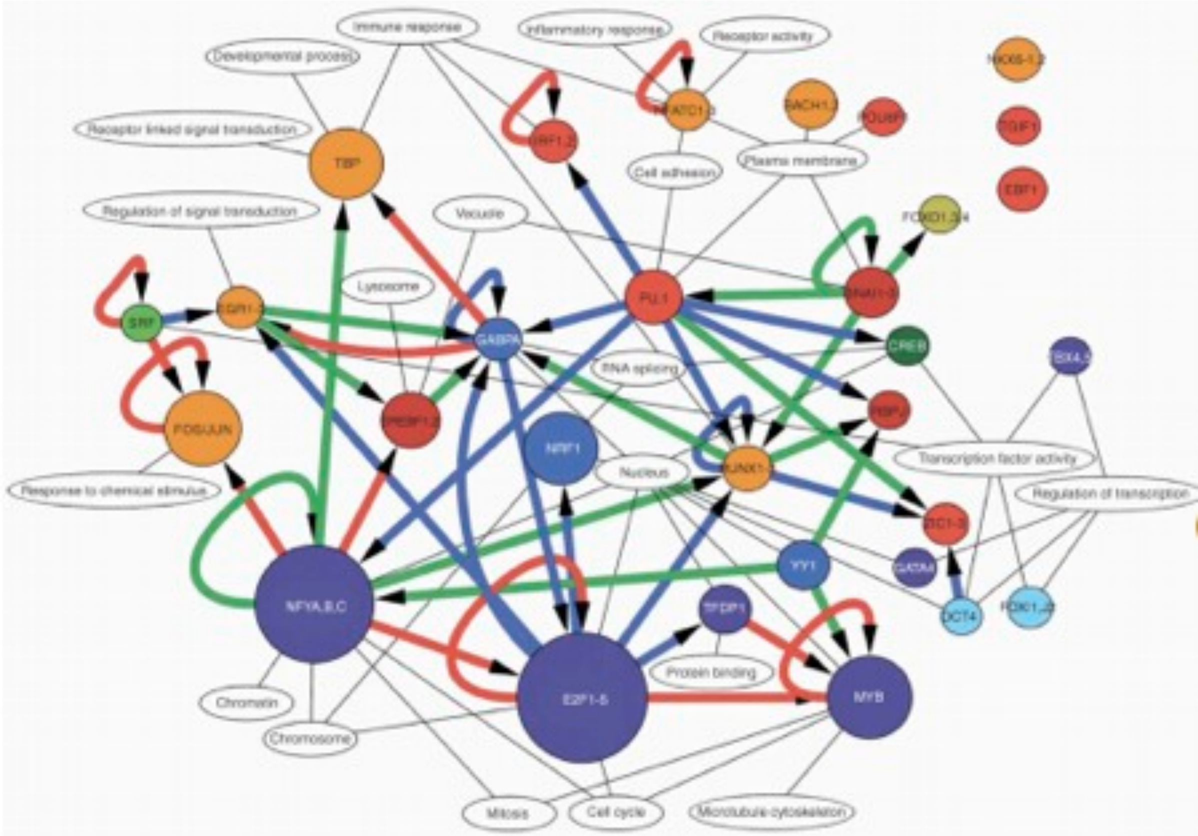

Immune response
Inflammatory response
Receptor activity
Developmental process
Receptor linked signal transduction
TBP
Cell adhesion
Plasma membrane
Regulation of signal transduction
Vacuole
Lysosome
SRF
GABPA
PU.1
CREB
RNA splicing
NRF1
Nucleus
Transcription factor activity
Regulation of transcription
Response to chemical stimulus
YY1
GATA4
OCT4
NFYA,B,C
TFDP1
Protein binding
E2F1-5
MYB
Chromatin
Chromosome
Mitosis
Cell cycle
Microtubule cytoskeleton

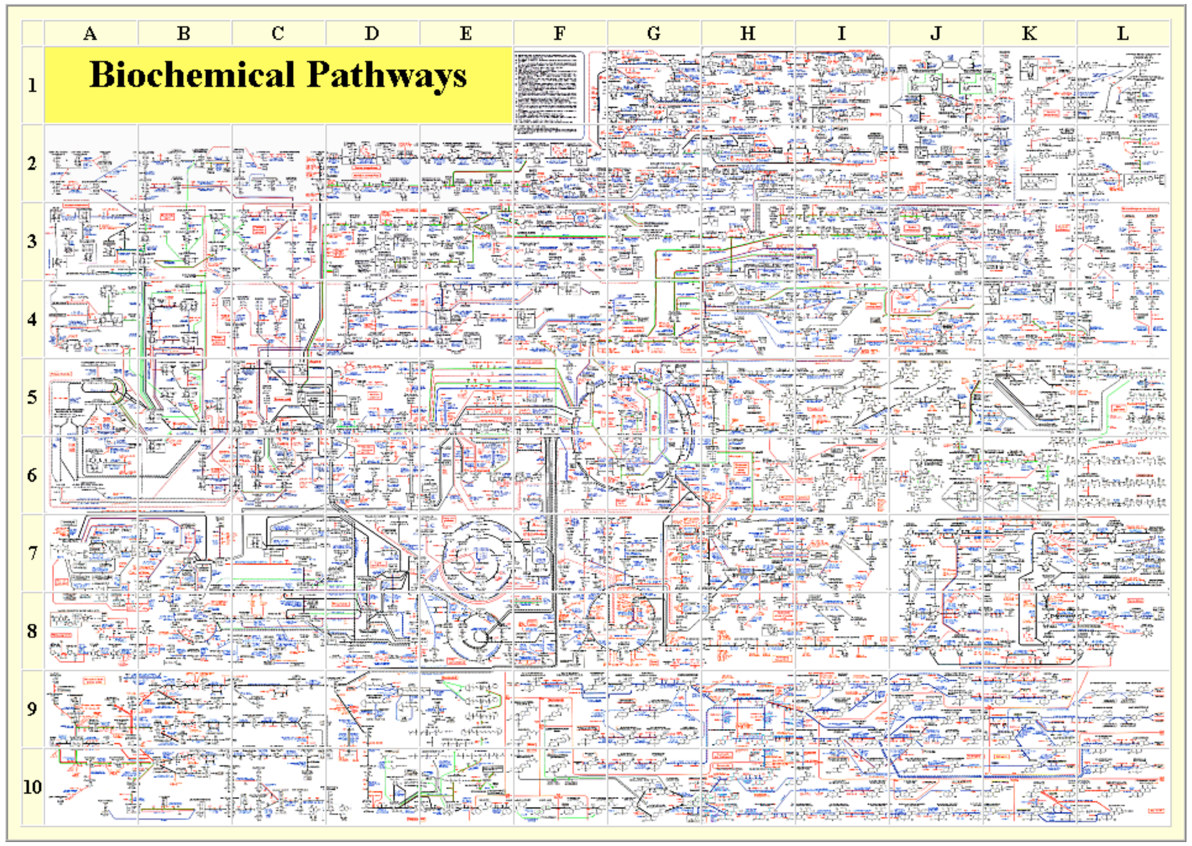

Pajek

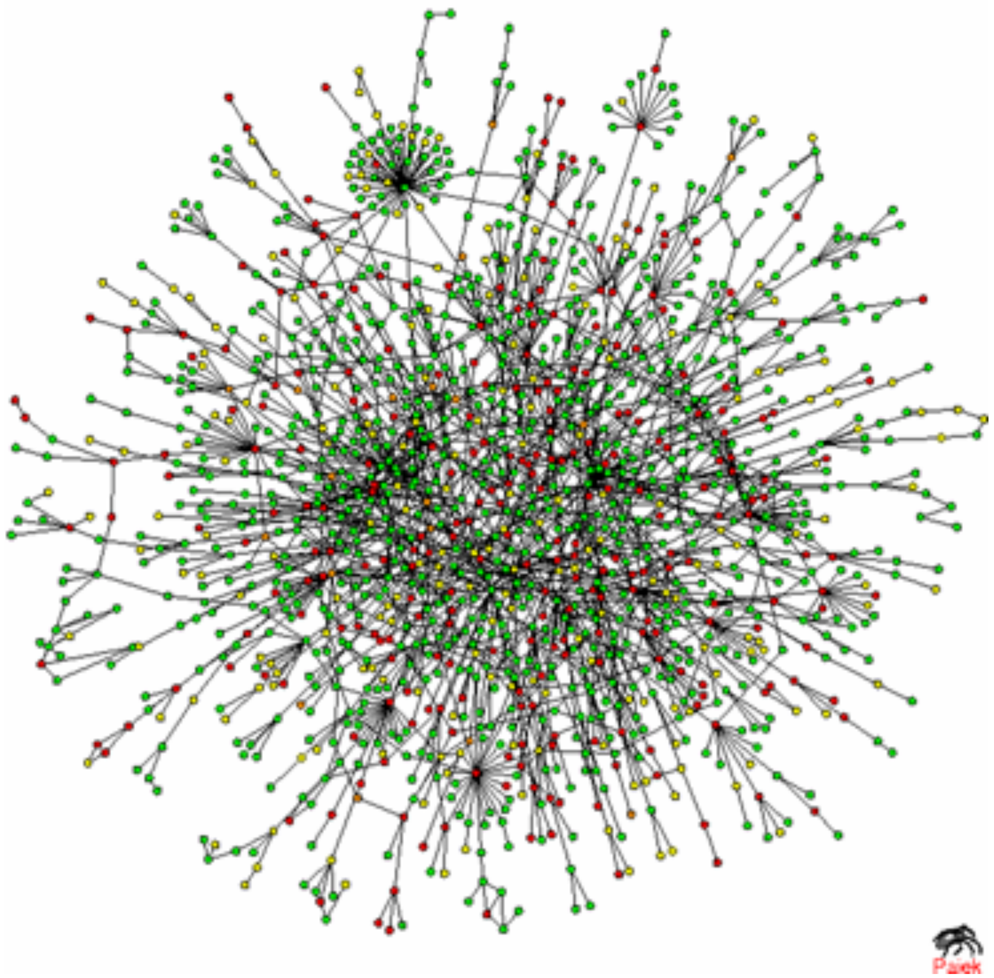

Immune response
Inflammatory response
Receptor activity
Developmental process
Receptor linked signal transduction
TBP
Cell adhesion
Plasma membrane
Regulation of signal transduction
Vacuole
Lysosome
SRF
GABPA
PU.1
CREB
RNA splicing
FOS/JUN
NRF1
Nucleus
Transcription factor activity
Response to chemical stimulus
Regulation of transcription
YY1
NFYA,B,C
TFDP1
Protein binding
E2F1-5
MYB
Chromatin
Chromosome
Mitosis
Cell cycle
Microtubule cytoskeleton

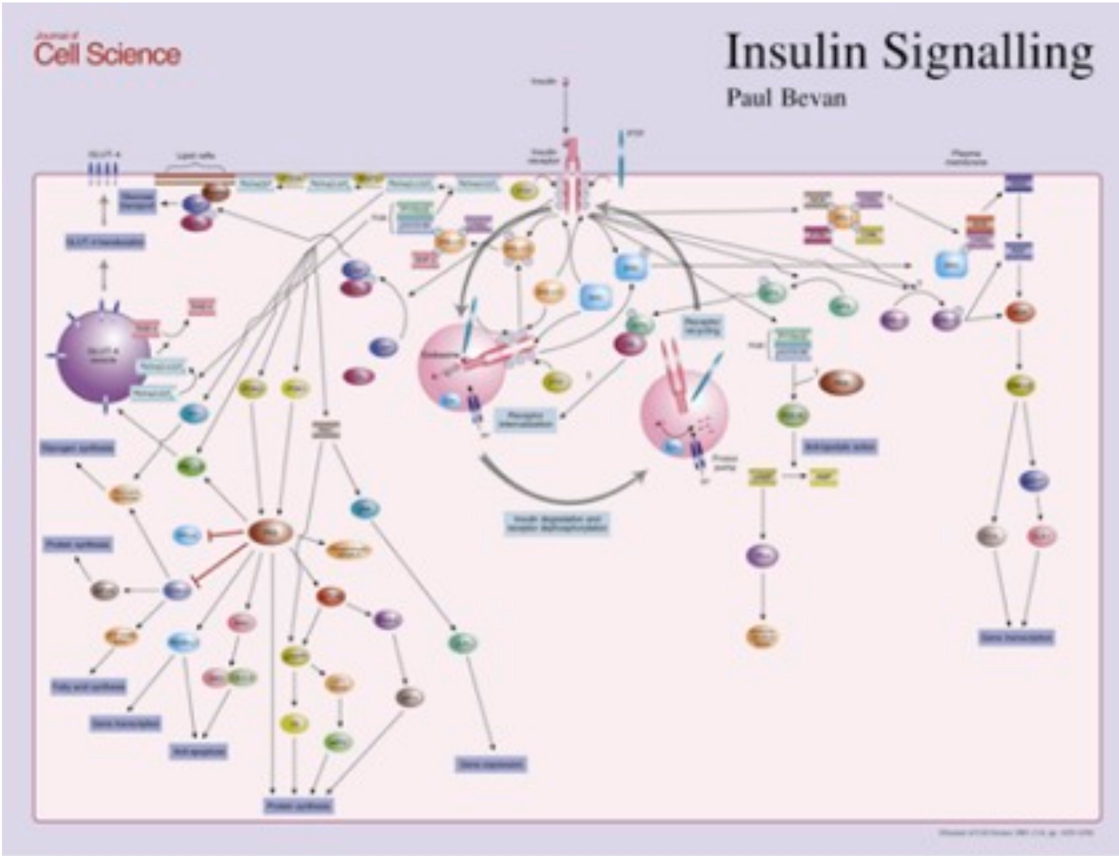

Biochemical Pathways
A
B
C
D
E
F
G
H
I
J
K
L
1
2
3
4
5
6
7
8
9
10

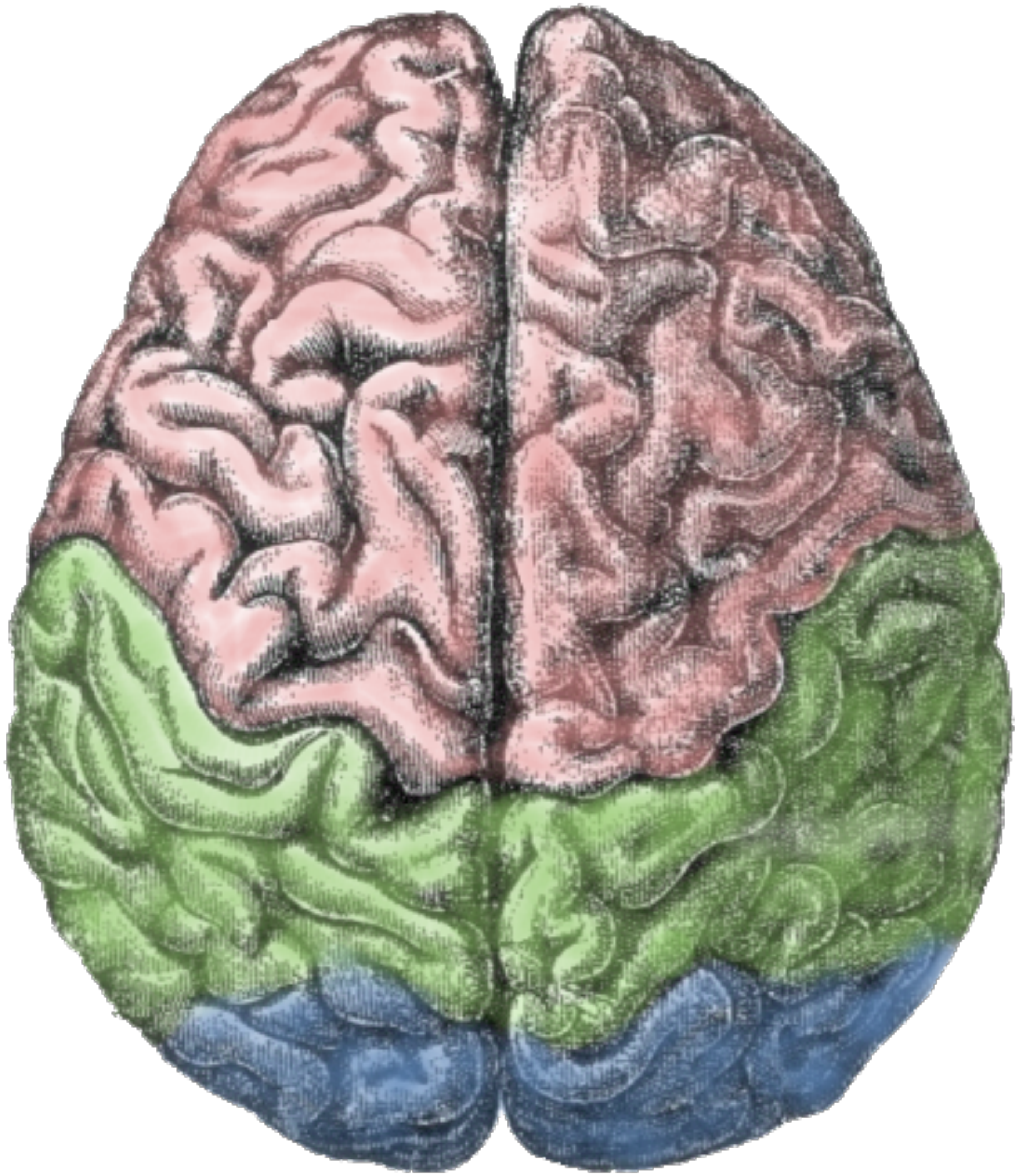



Immune response
Inflammatory response
Receptor activity
Developmental process
Receptor linked signal transduction
TBP
Cell adhesion
Plasma membrane
Regulation of signal transduction
Vacuole
Lysosome
SRF
GABPA
PU.1
CREB
RNA splicing
FOS/JUN
NRF1
Nucleus
Transcription factor activity
Response to chemical stimulus
Regulation of transcription
YY1
NFYA,B,C
TFDP1
Protein binding
E2F1-5
MYB
Chromatin
Chromosome
Mitosis
Cell cycle
Microtubule cytoskeleton

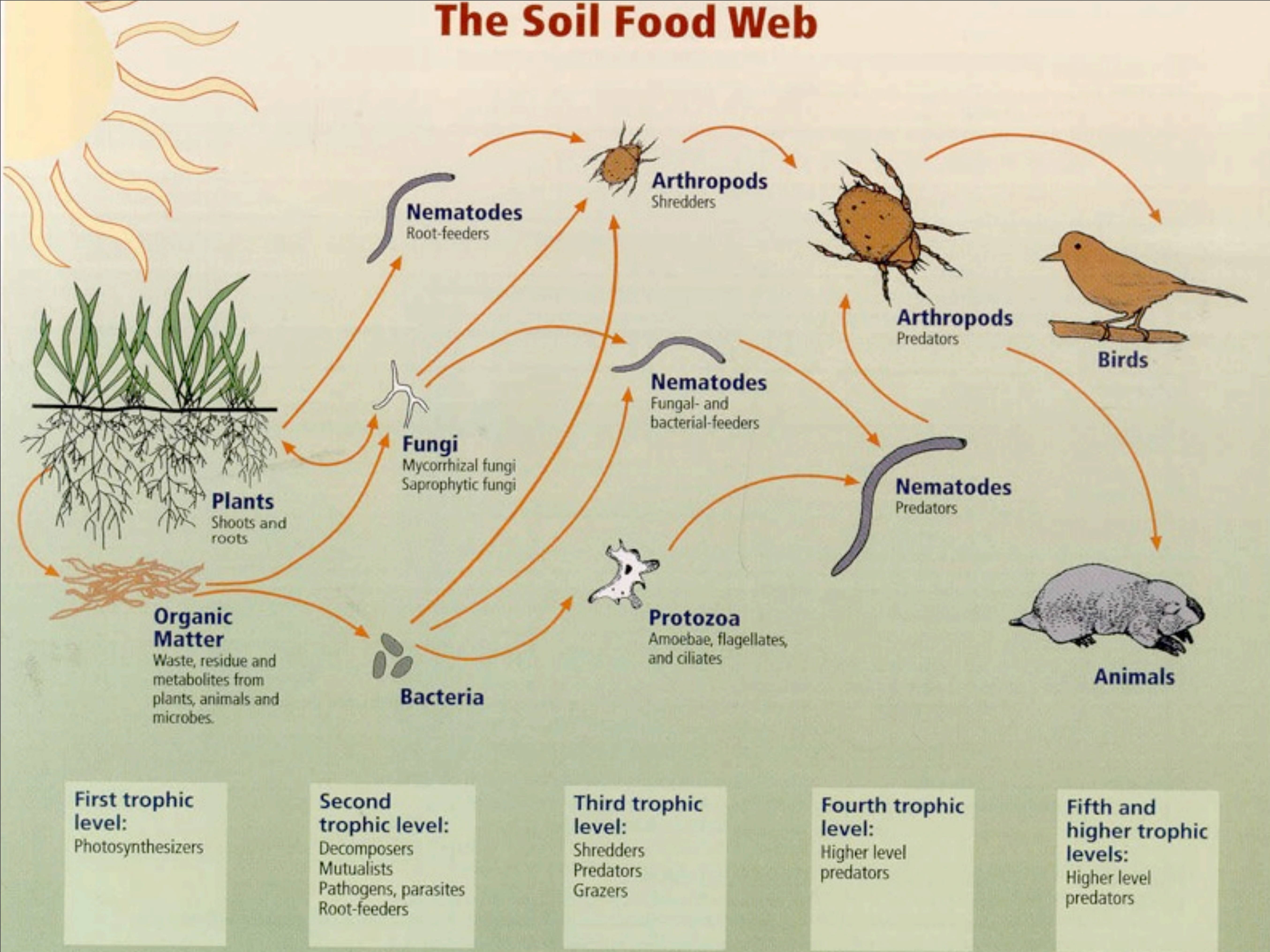

Biochemical Pathways
A
B
C
D
E
F
G
H
I
J
K
L
1
2
3
4
5
6
7
8
9
10

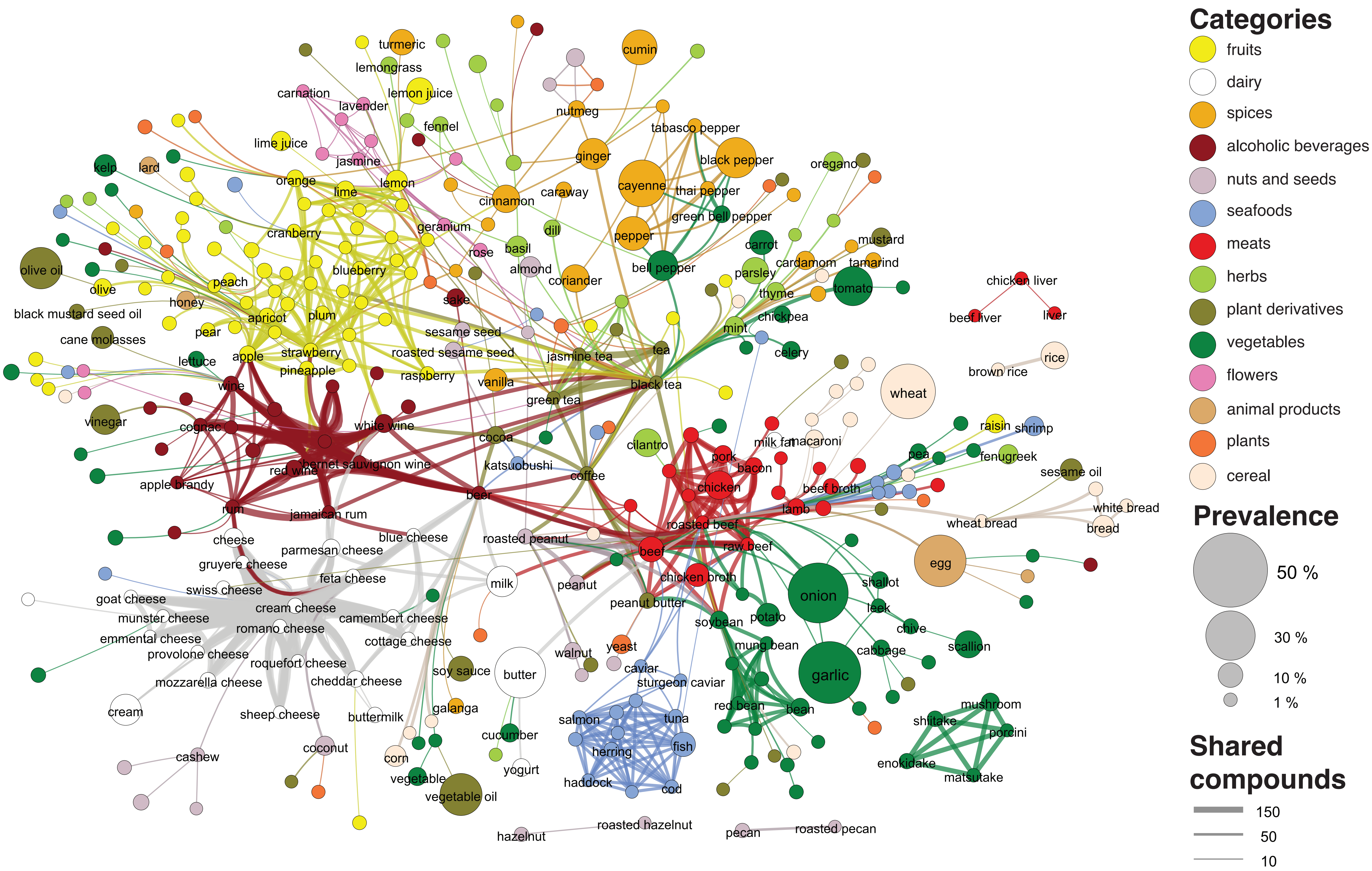



Cell Science
Insulin Signalling
Paul Bevan




facebook
December 2010

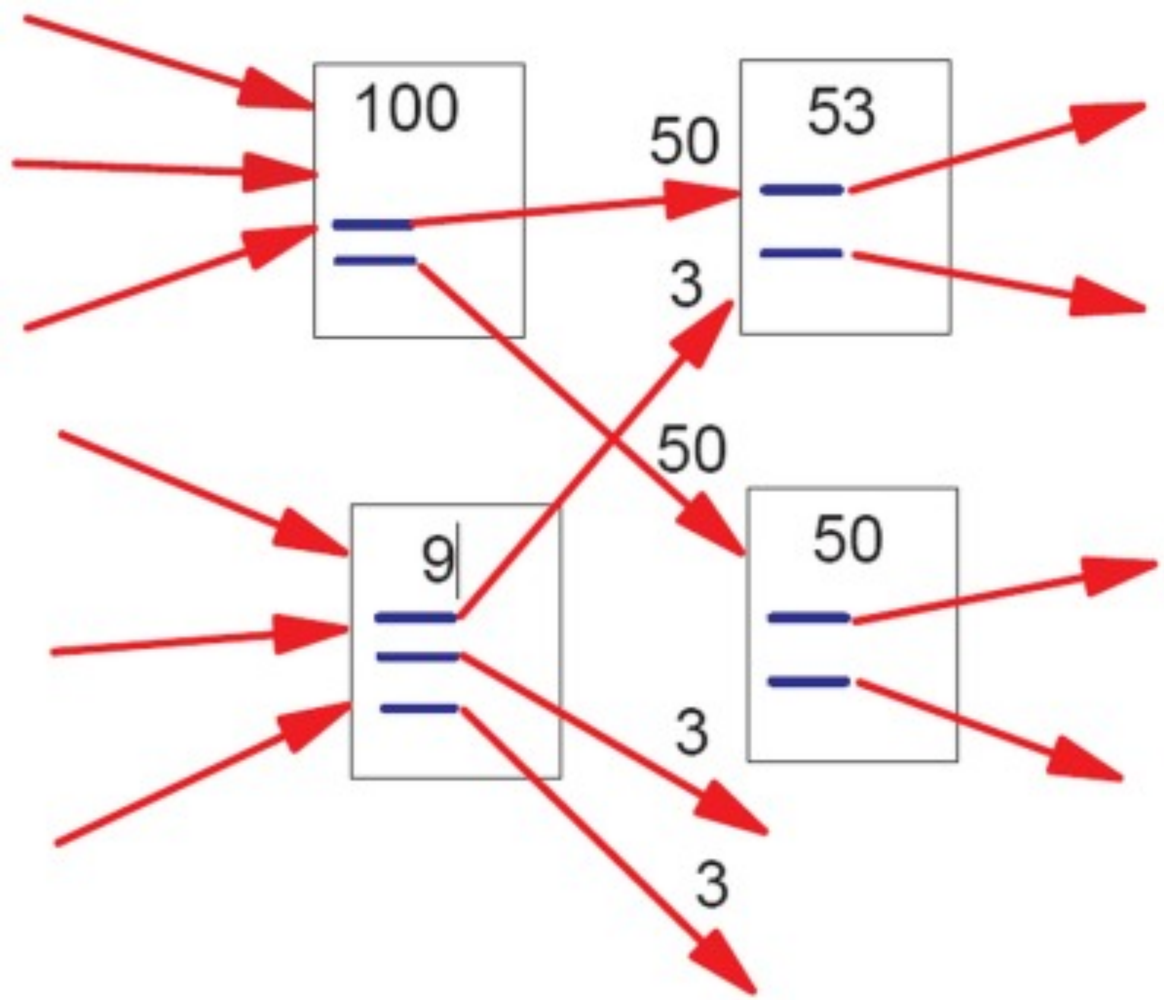

The Soil Food Web
Arthropods
Shredders
Nematodes
Root-feeders
Arthropods
Predators
Birds
Nematodes
Fungal- and
bacterial-feeders
Fungi
Mycorrhizal fungi
Saprophytic fungi
Nematodes
Predators
Plants
Shoots and
roots
Organic
Matter
Waste, residue and
metabolites from
plants, animals and
microbes.
Protozoa
Amoebae, flagellates,
and ciliates
Bacteria
Animals
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?

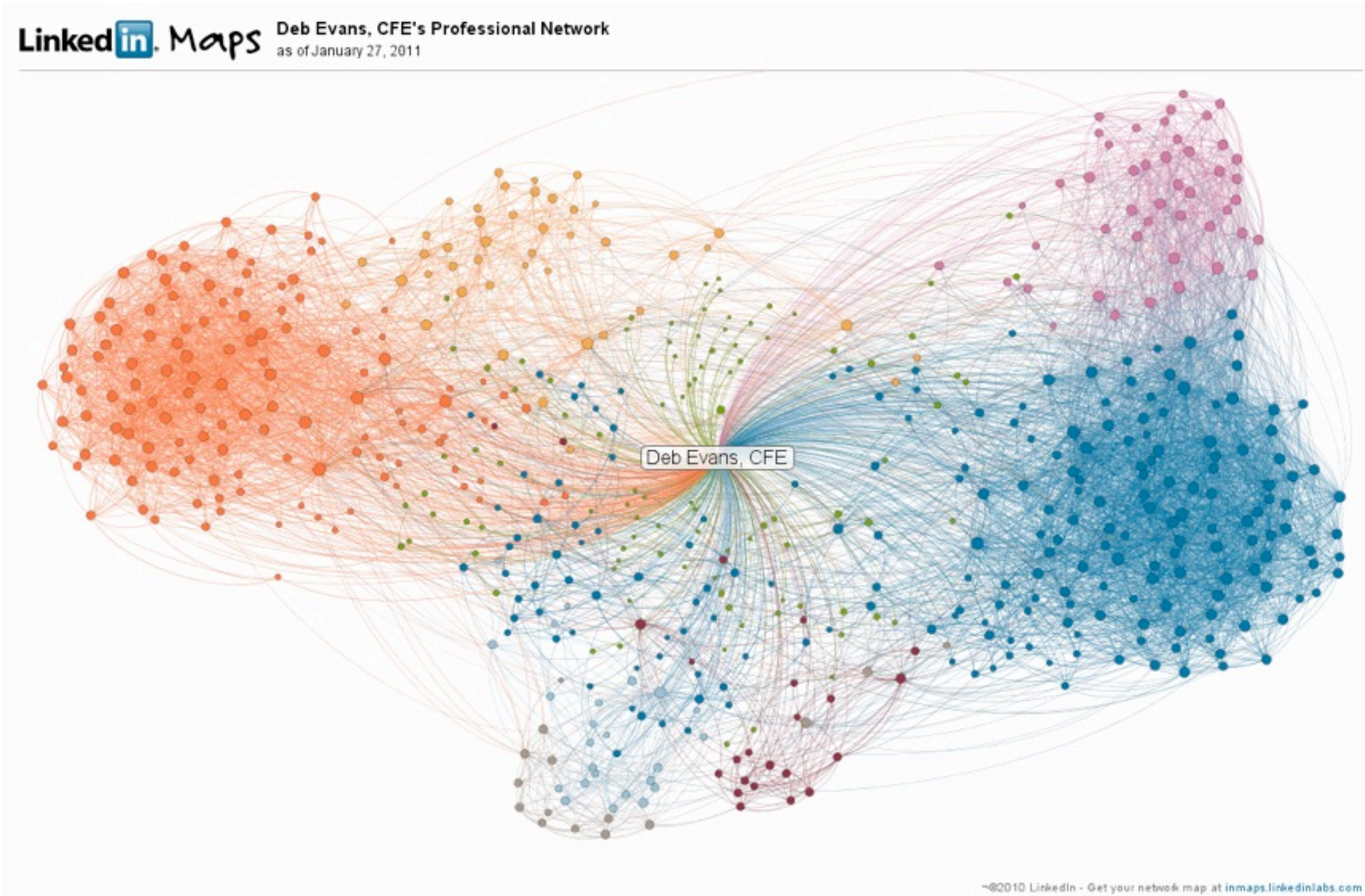
Google



Pagerank =

Random walk problem on a network

LinkedIn


LinkedIn Maps
Deb Evans, CFE's Professional Network
as of January 27, 2011
Deb Evans, CFE
©2010 LinkedIn - Get your network map at inmaps.linkedinlabs.com


facebook
December 2010


nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

INSIGHT

CAMPAIGN TRIAL

Close friendships beat weaker Facebook links in test of election-day influence PAGES 212 & 295

MICROWAVE WARFARE

UNEXPLODED MYTH?

The Pentagon's quest for the e-bomb

PAGE 198

RESEARCH CAREERS

PICKING WINNERS

A formula to predict your future h-index

PAGE 201

LITERATURE

FORGOTTEN FUTURES

When science fiction came of age

PAGE 204

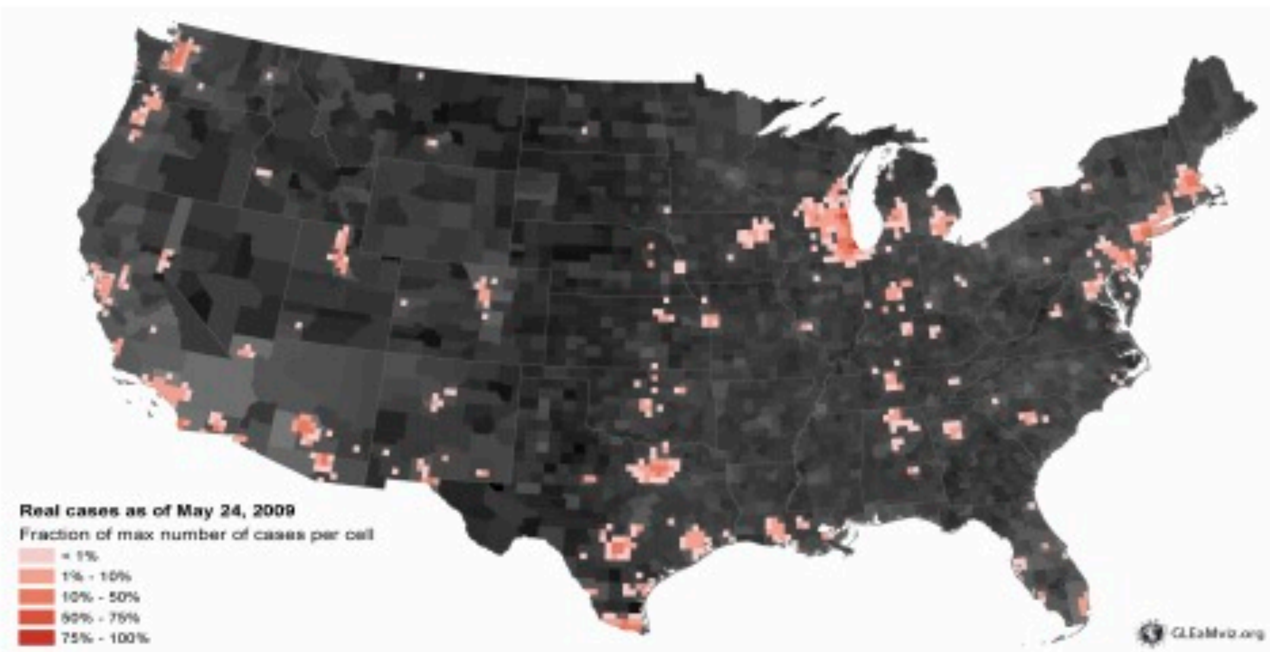
NATURE.COM/NATURE

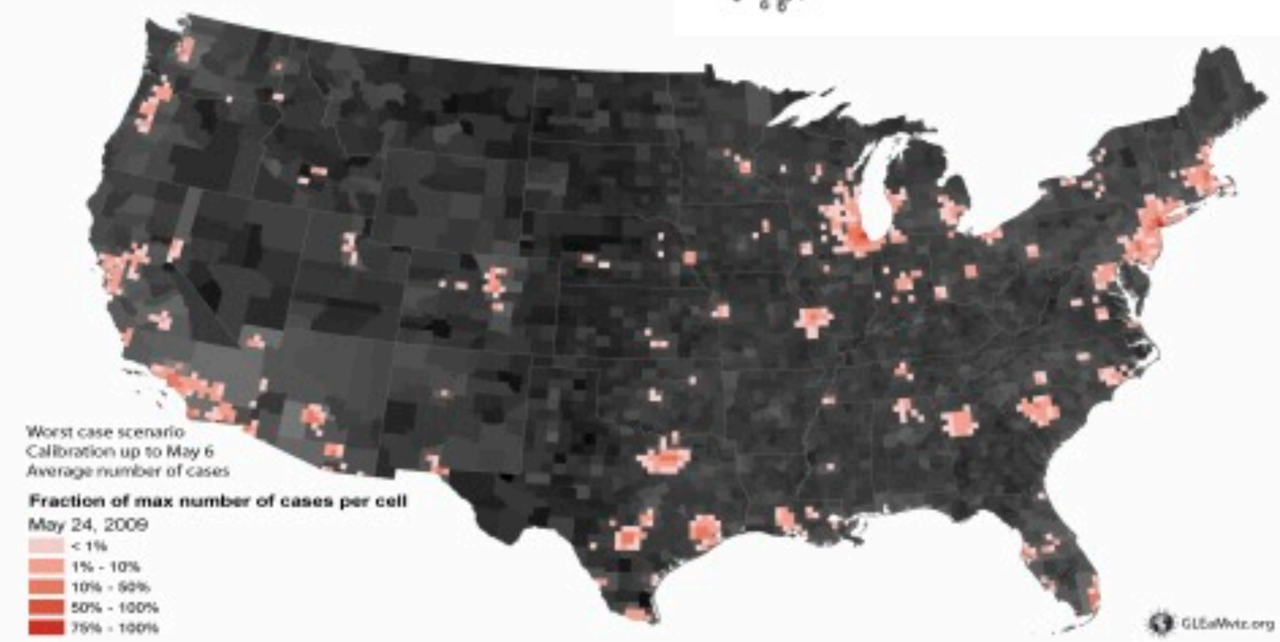
13 September 2012 £10

Vol. 489, No. 7415

H1N1 Pandemic prediction

GLEaMviz.org

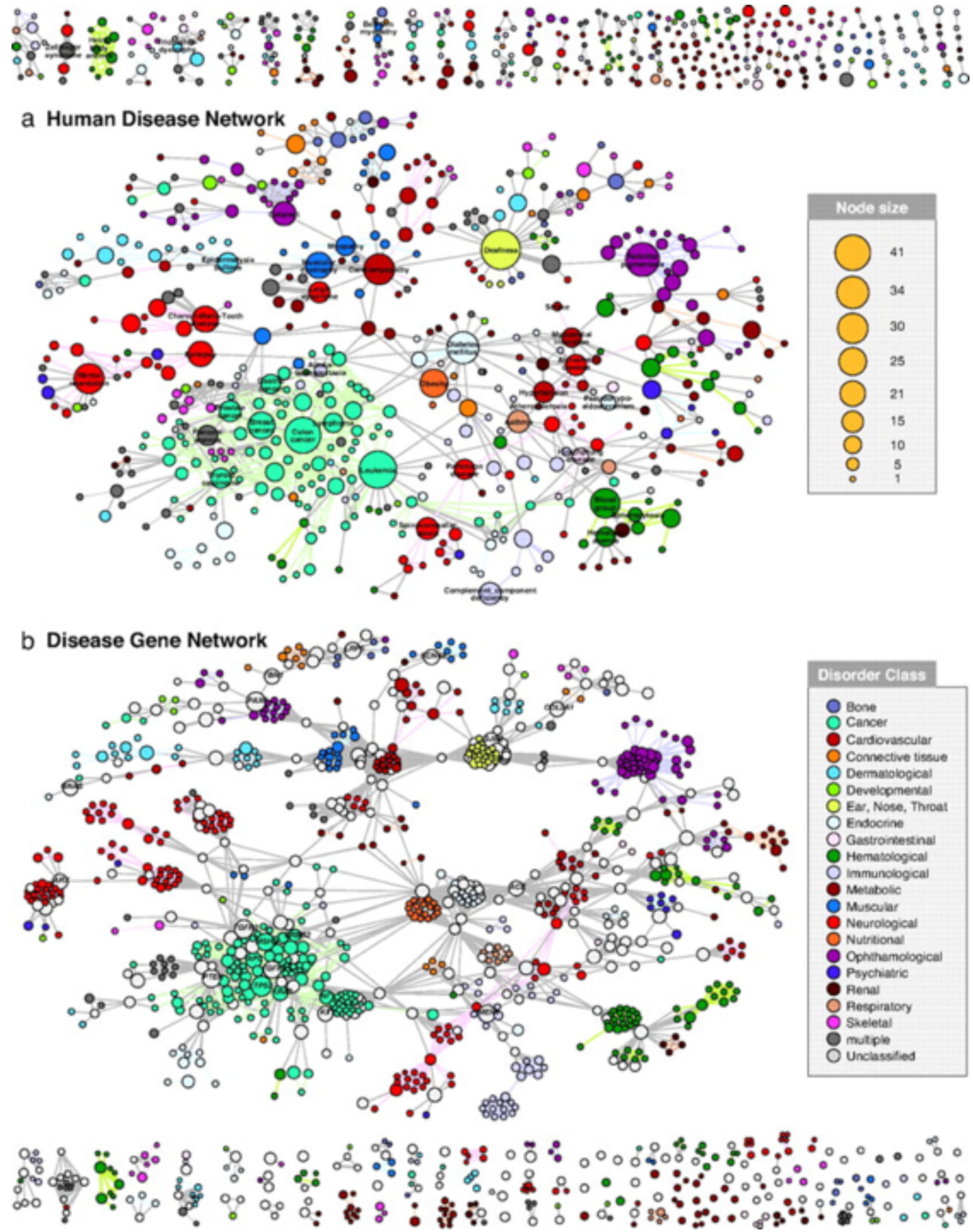




Real

Prediction

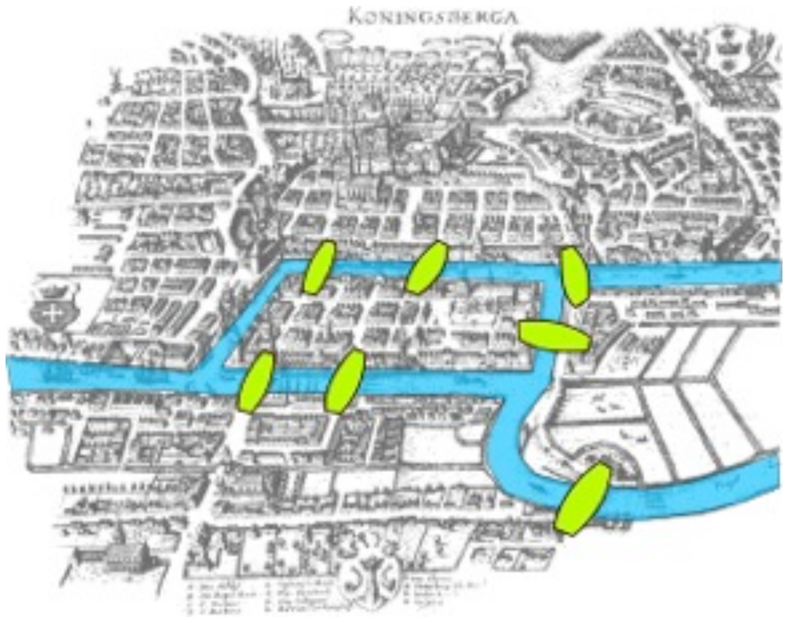
Reaction-diffusion system with
transportation networks

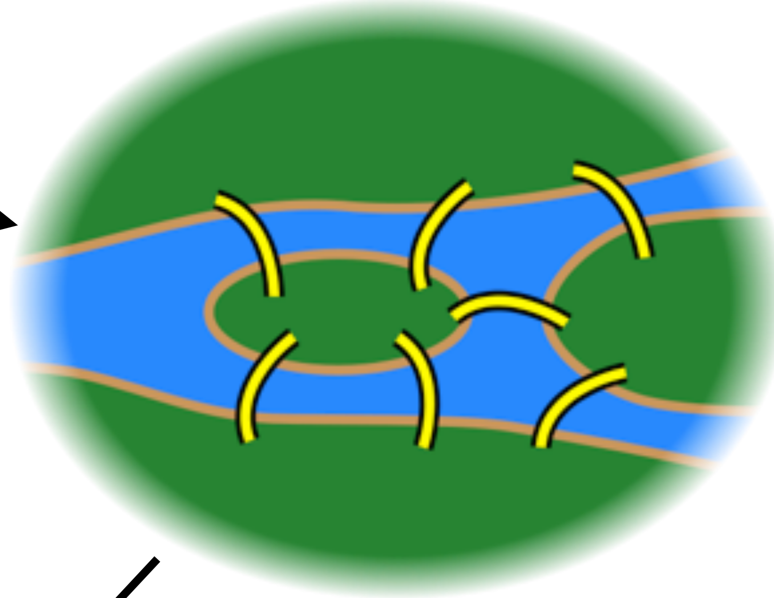

a Human Disease Network
Node size
41
34
30
25
21
15
10
5
1
b Disease Gene Network
Disorder Class
Bone
Cancer
Cardiovascular
Connective tissue
Dermatological
Developmental
Ear, Nose, Throat
Endocrine
Gastrointestinal
Hematological
Immunological
Metabolic
Muscular
Neurological
Nutritional
Ophthamological
Psychiatric
Renal
Respiratory
Skeletal
multiple
Unclassified

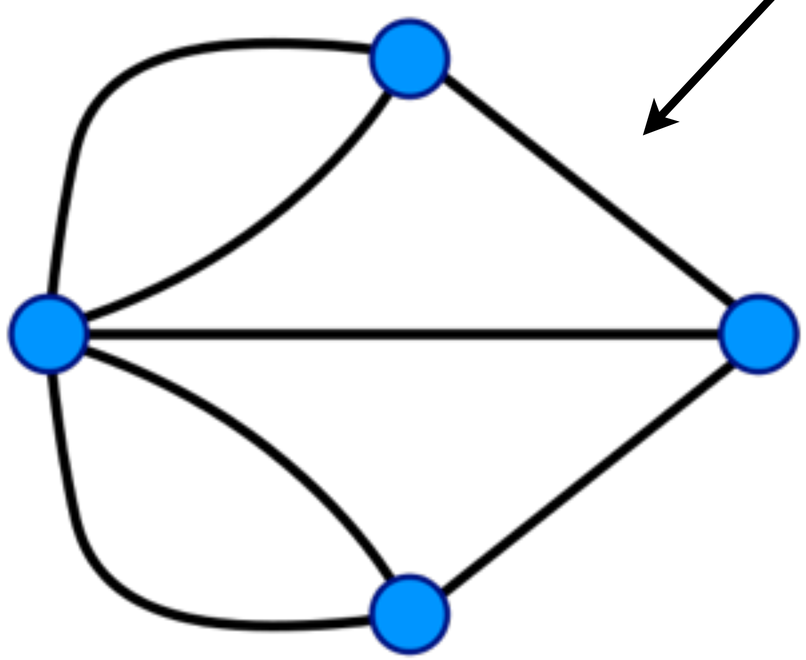
Can we understand a
complex system

without knowing the **structure**
of it?

NETWORKS







Graph Theory

Leonhard Euler

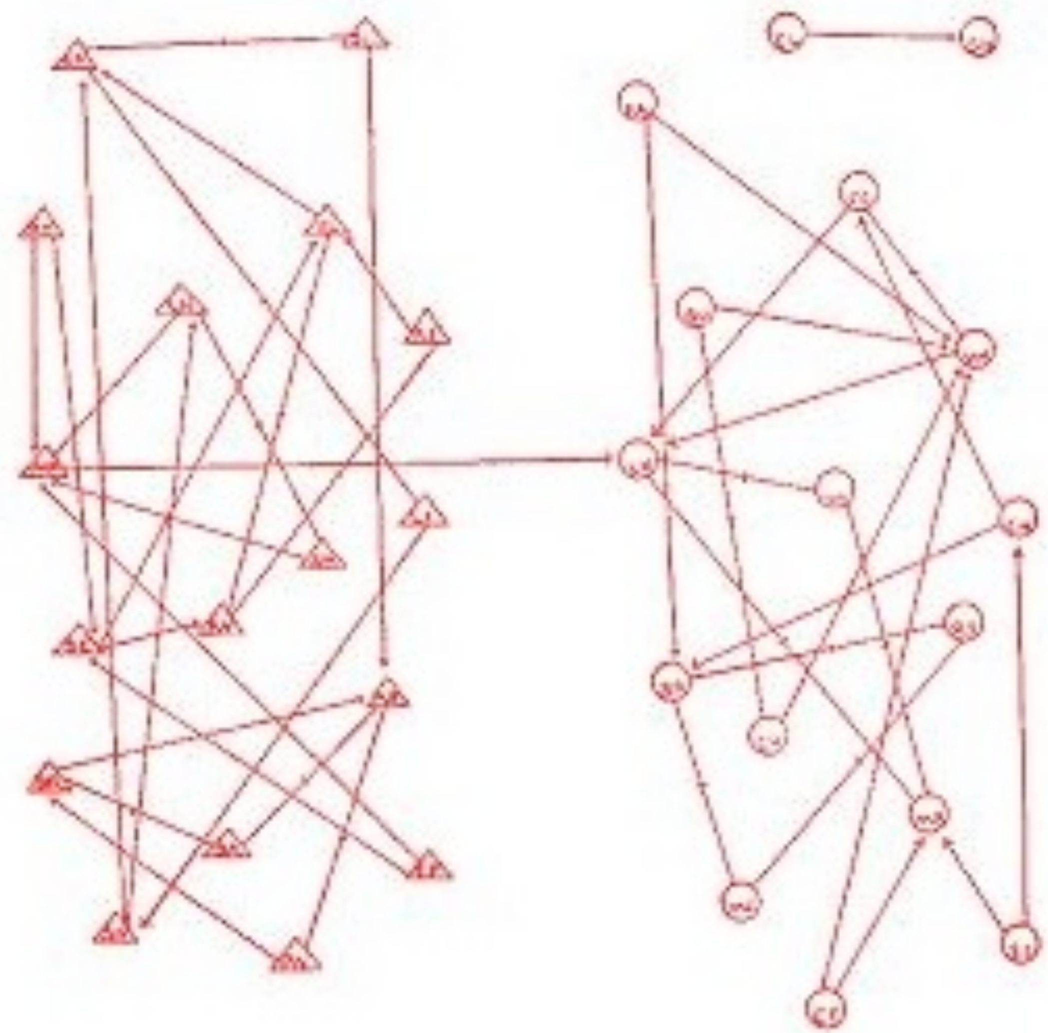
"Sociogram"

EMOTIONS MAPPED BY NEW GEOGRAPHY

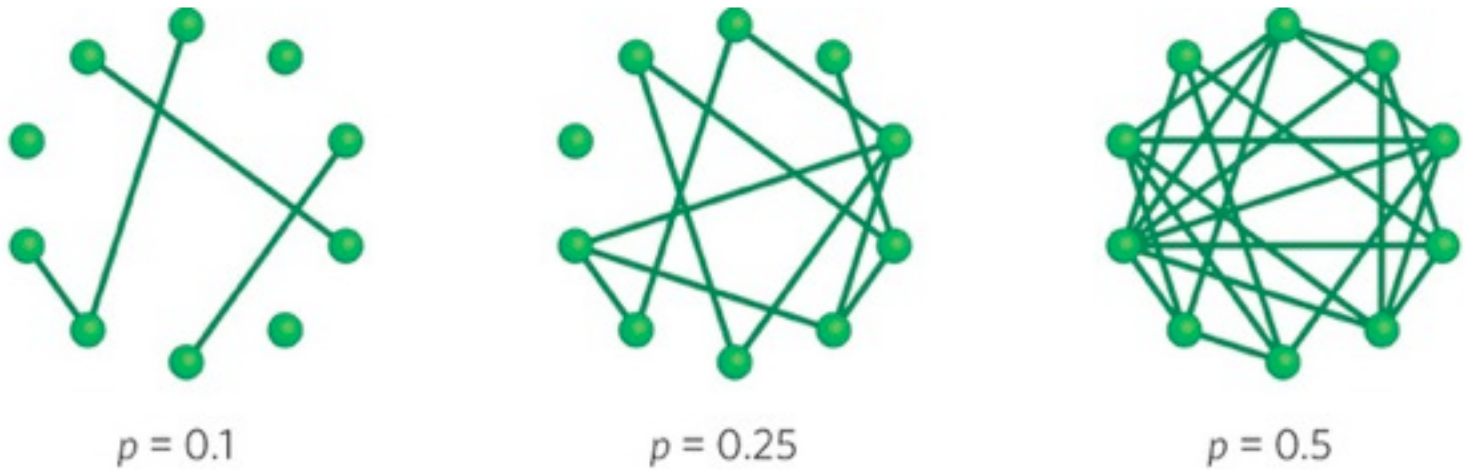
Charts Seek to Portray the Psychological Currents of Human Relationships.

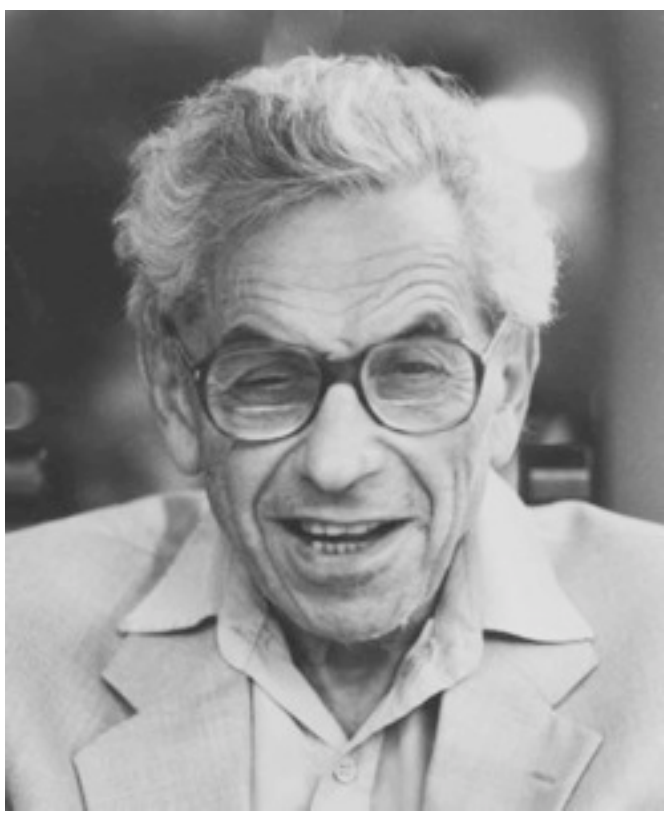
New York Times

April 3, 1933



What’s the structure of networks?





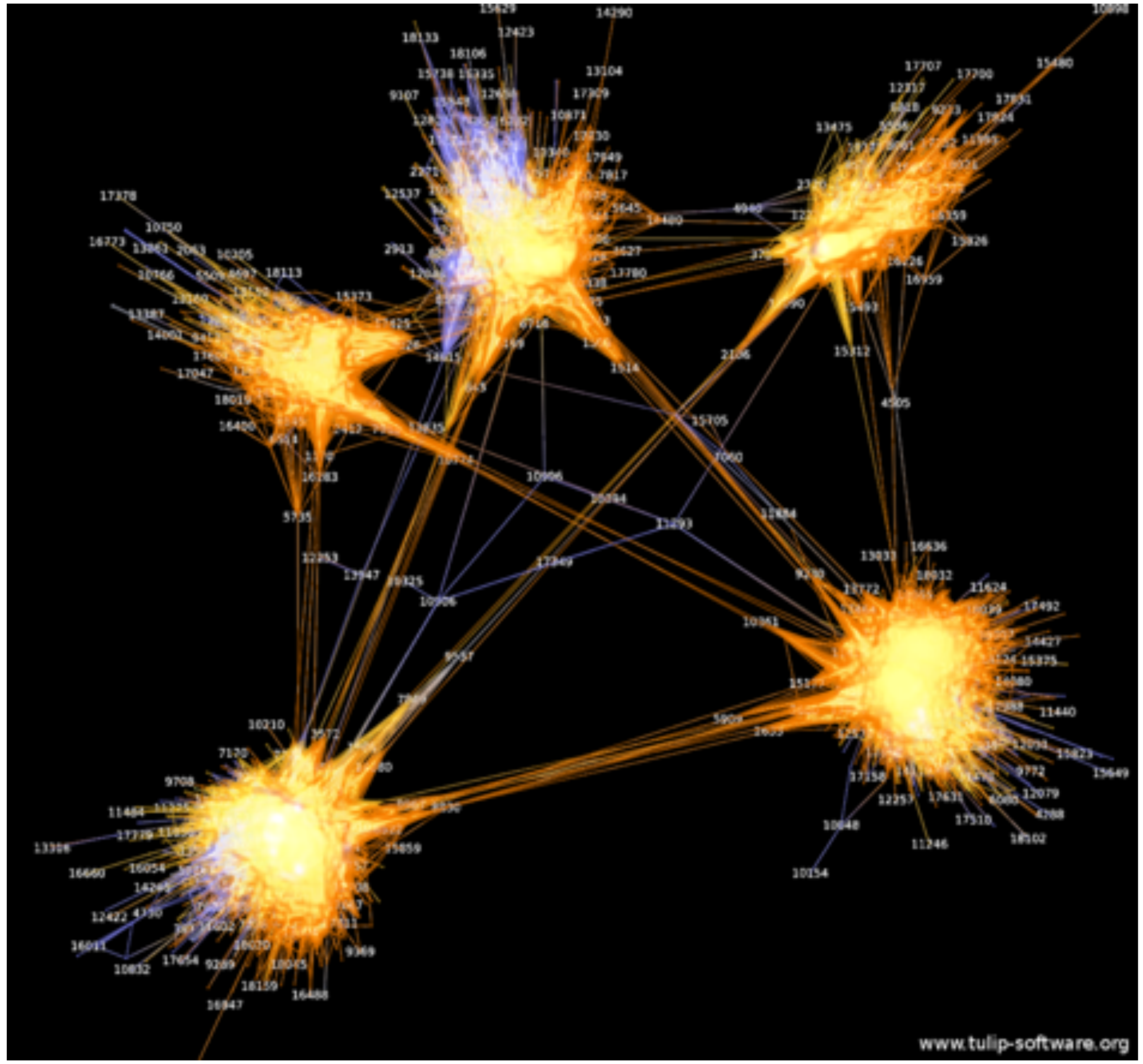
Paul Erdős

Alfréd Rényi

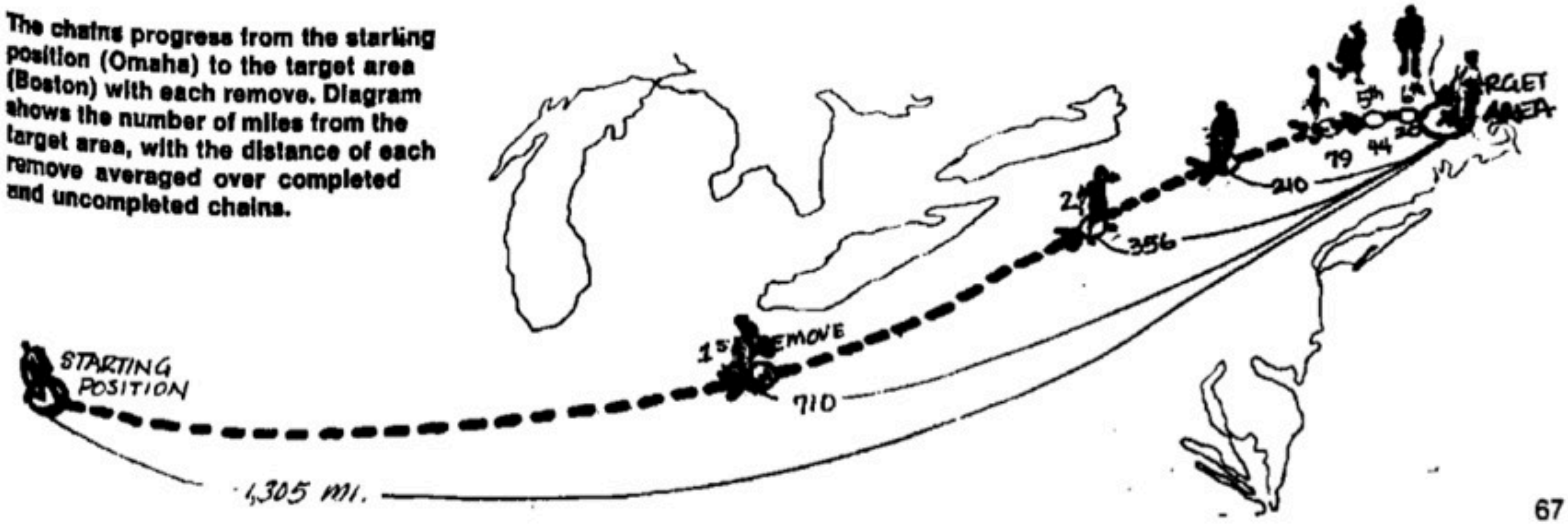
Clustering

Small-world

Heterogeneity

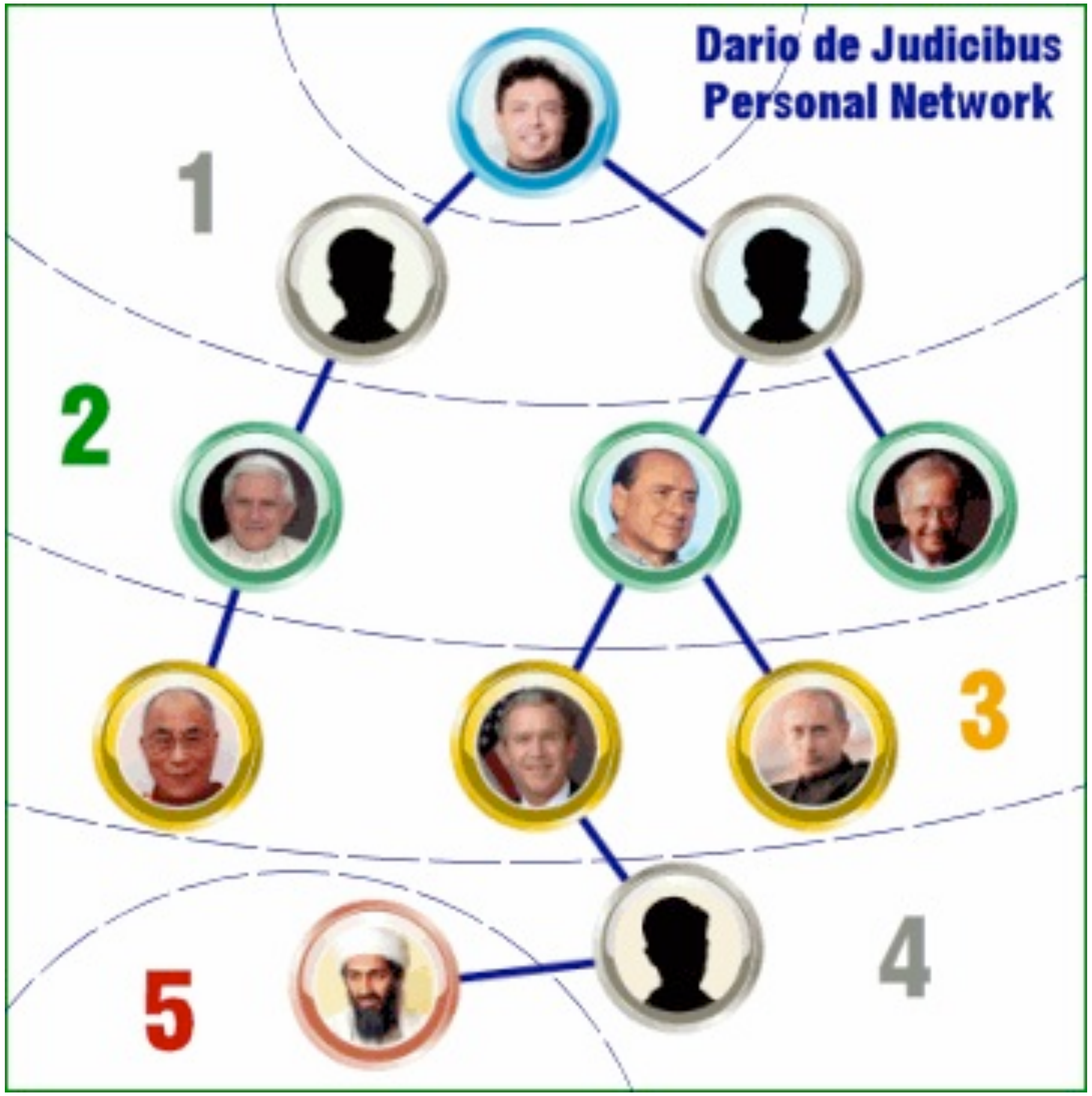


It’s not random! We form clusters.



“Small world experiment”

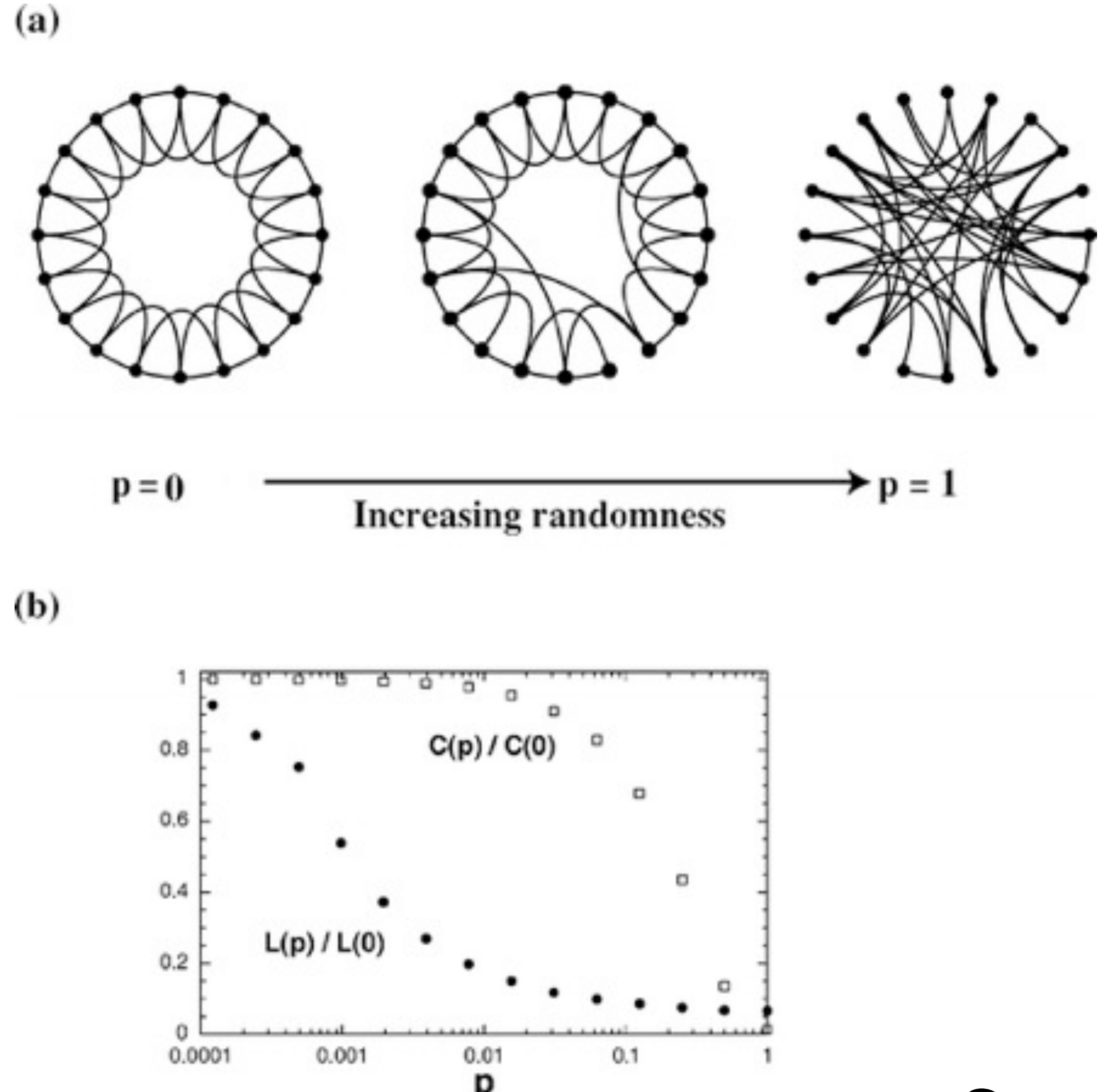
Stanley Milgram



http://oracleofbacon.org/index.php

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

Duncan J. Watts



Steven H. Strogatz

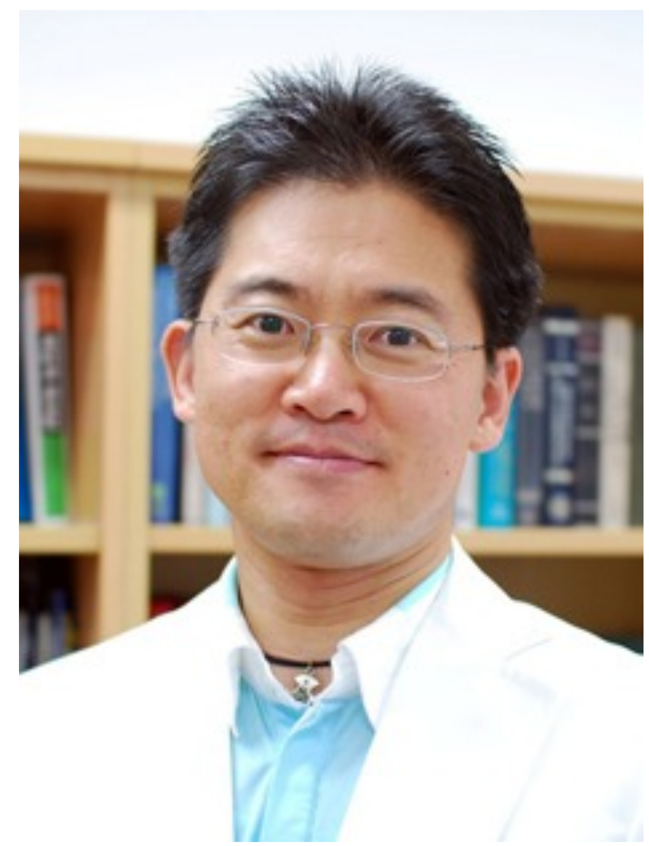
Watts and Strogatz model

Watts & Strogatz, Nature 1998

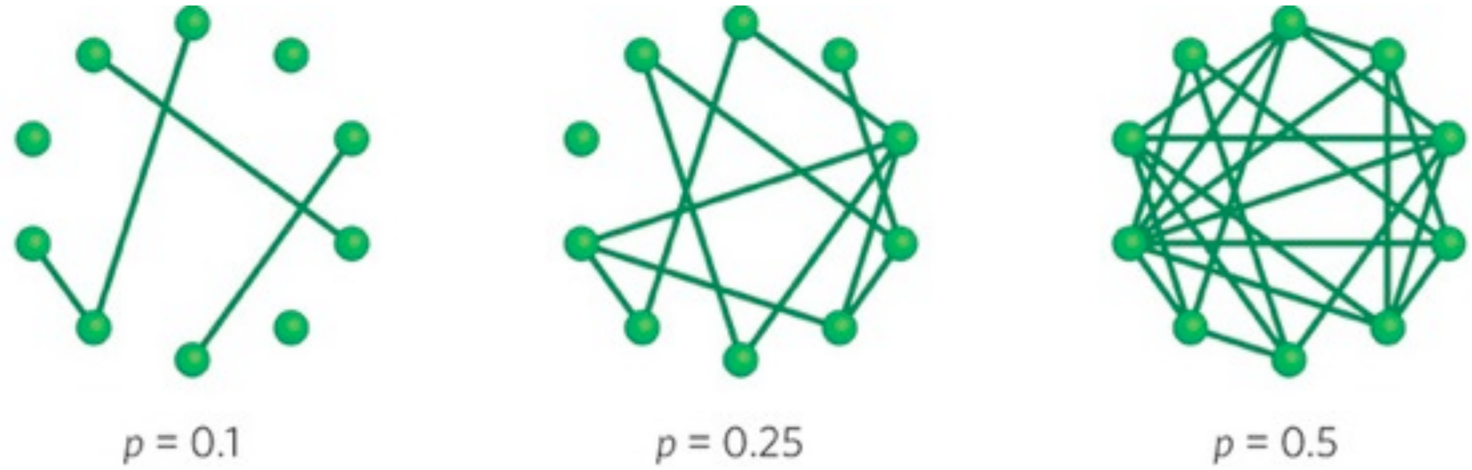
Networks are heterogeneous!

Albert-László Barabási

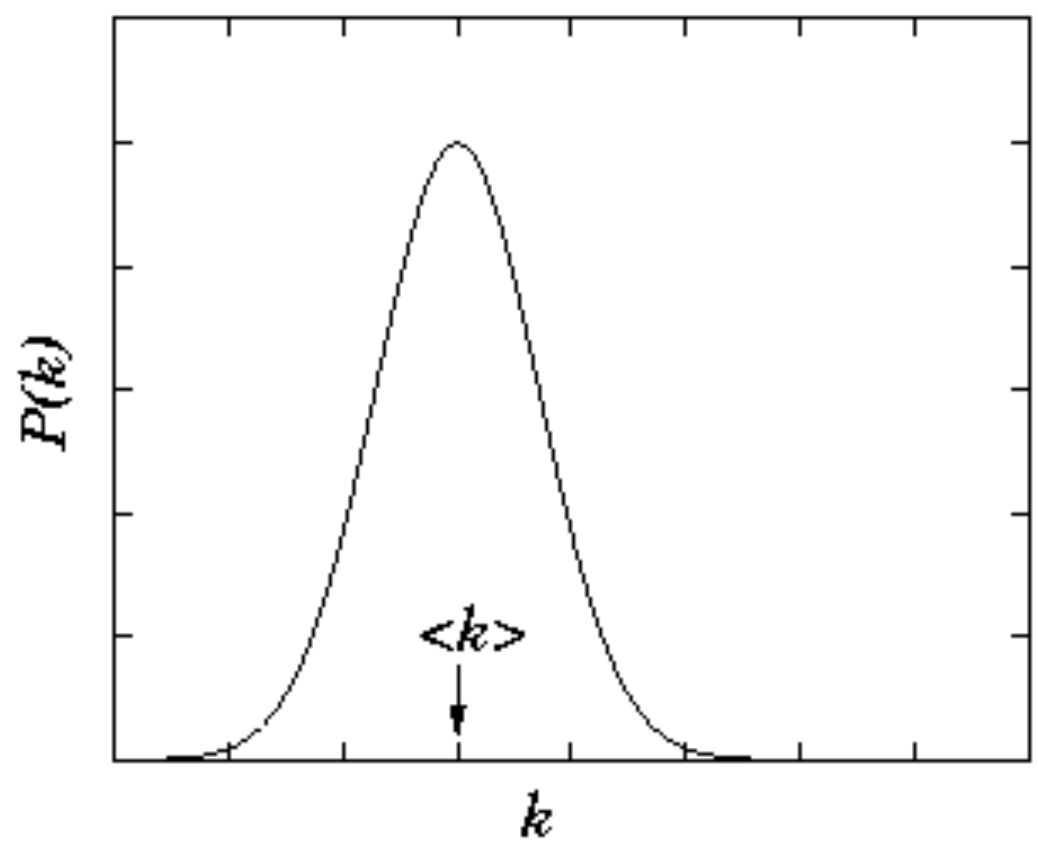
Réka Albert



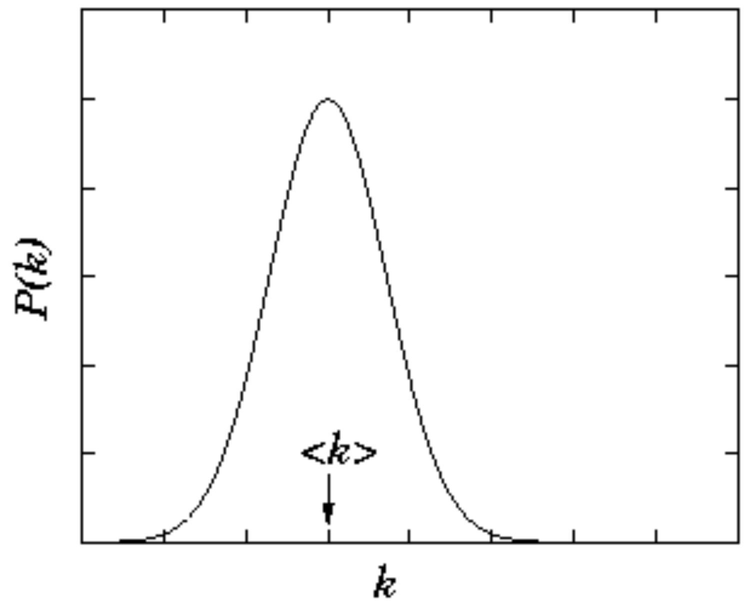
Hawoong Jeong



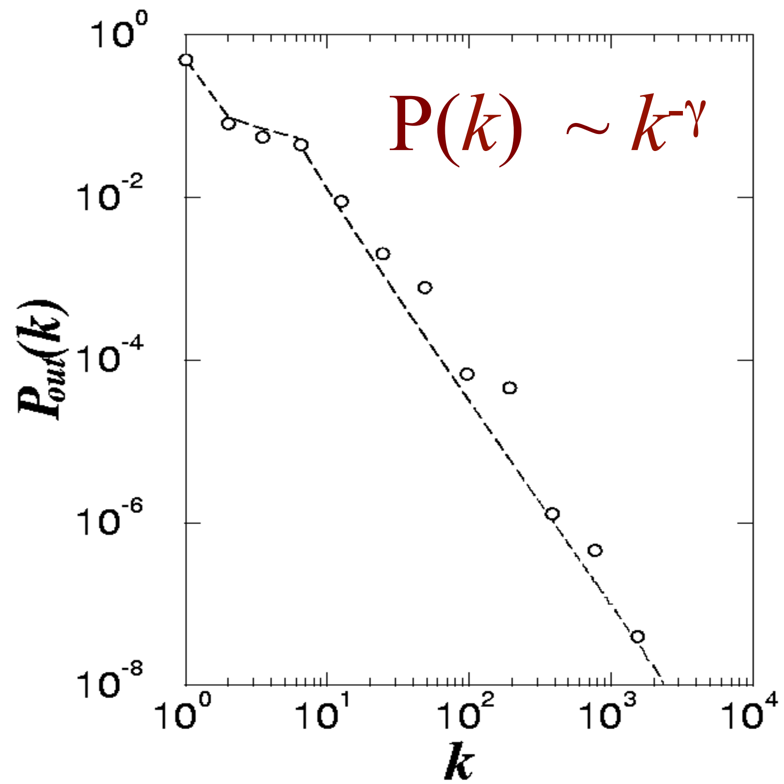
Poisson distribution



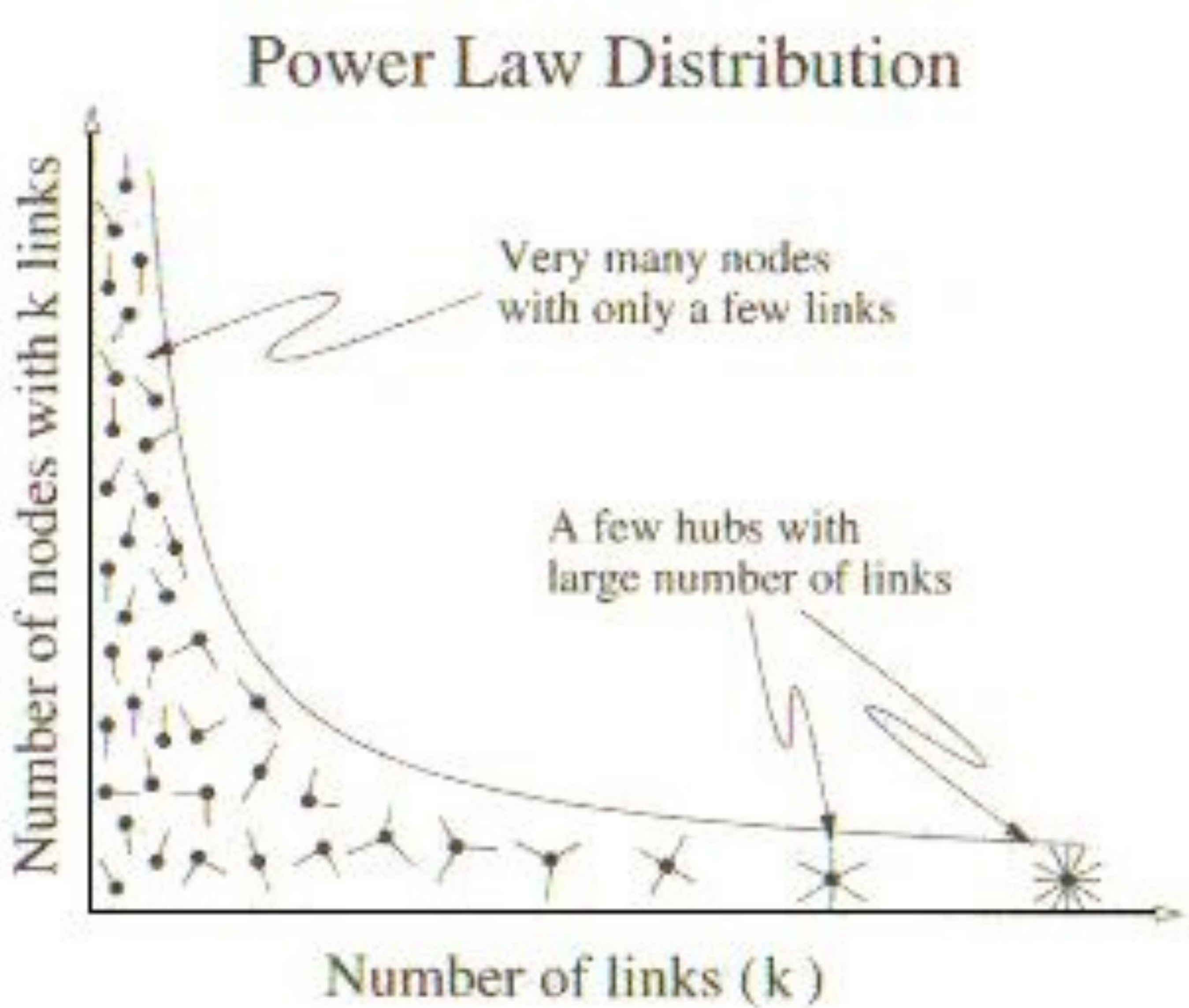
Degree: # of neighbors

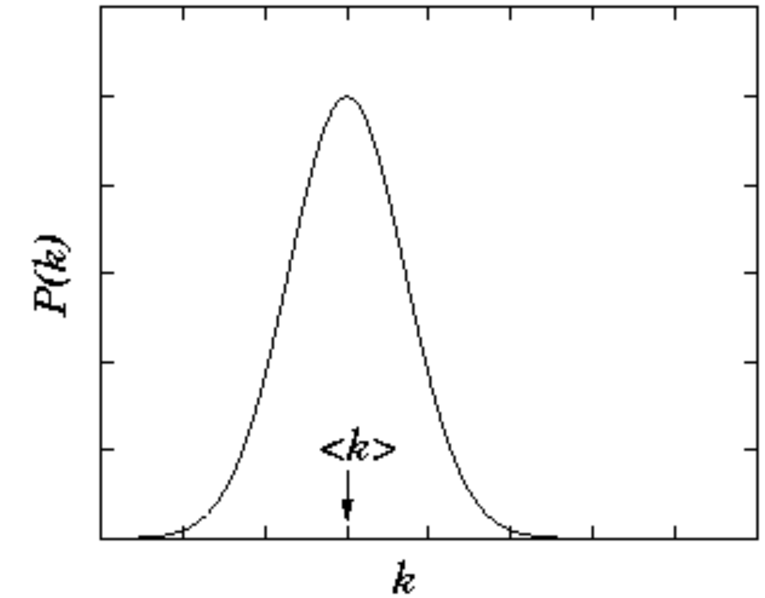


Expected

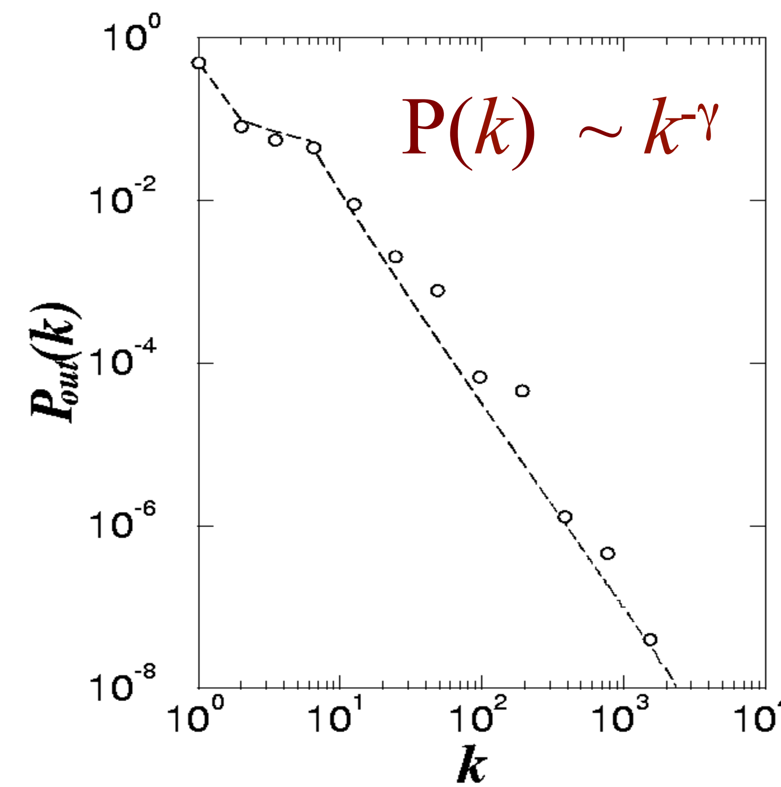


Found

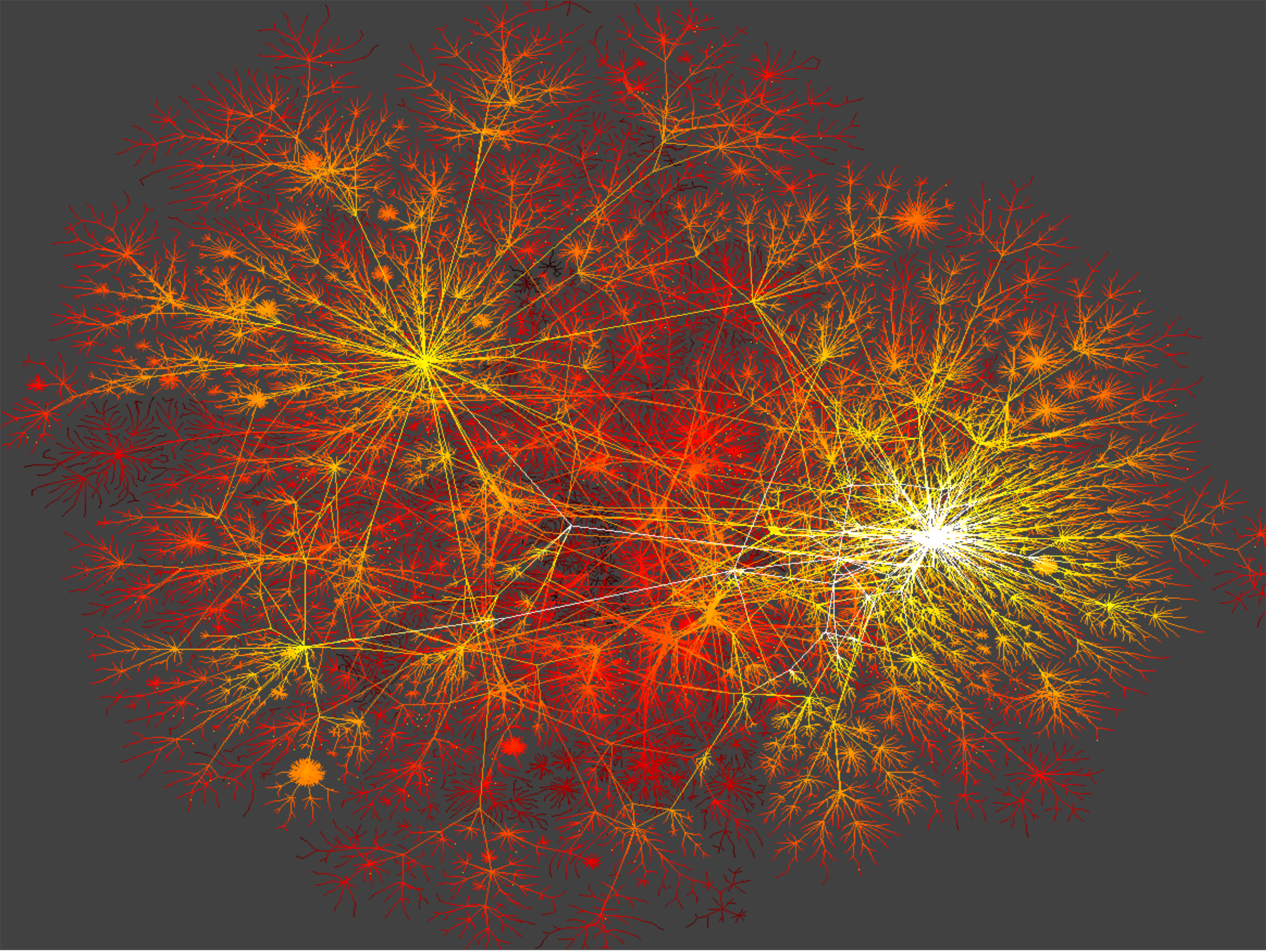


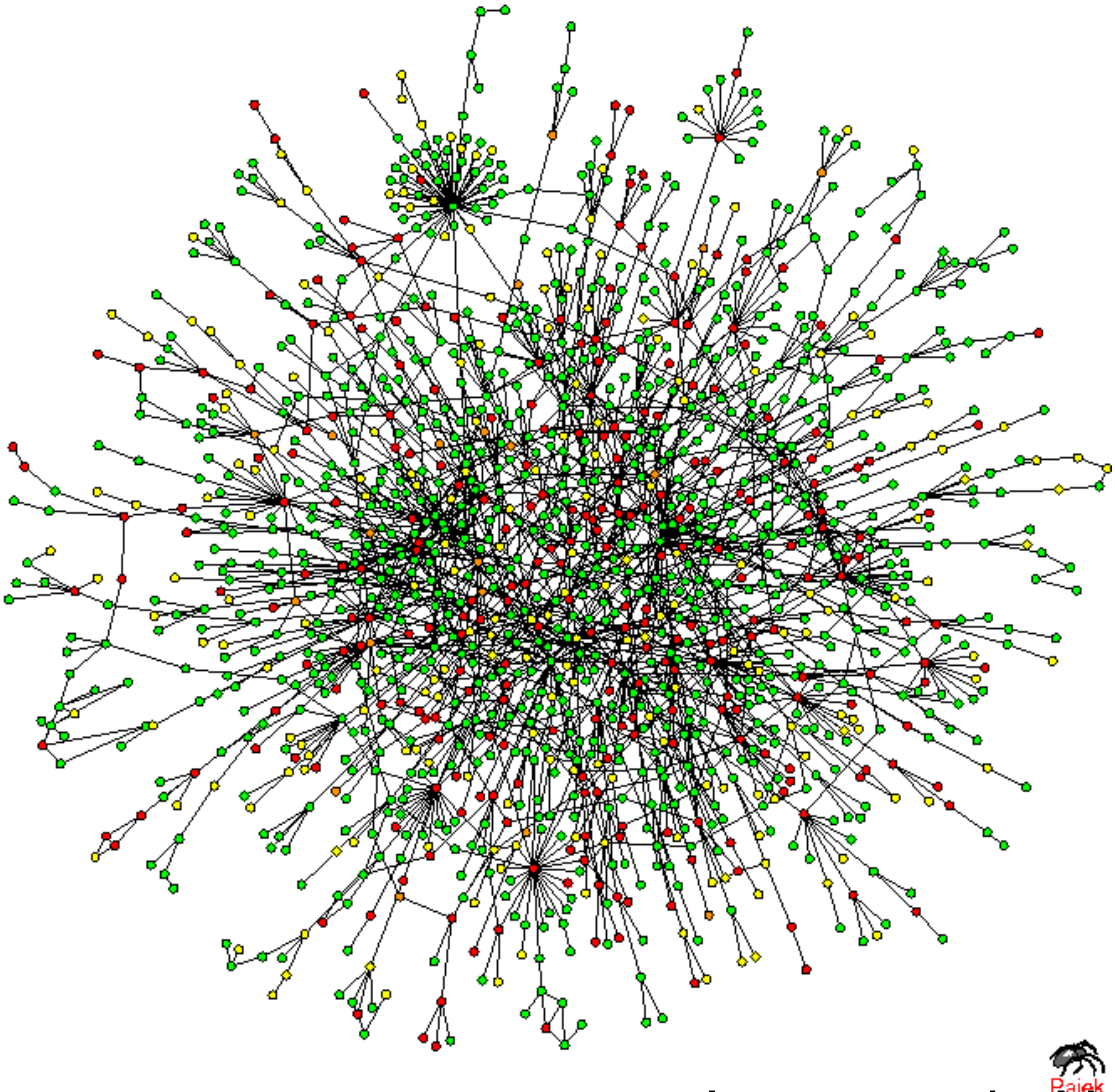


Expected

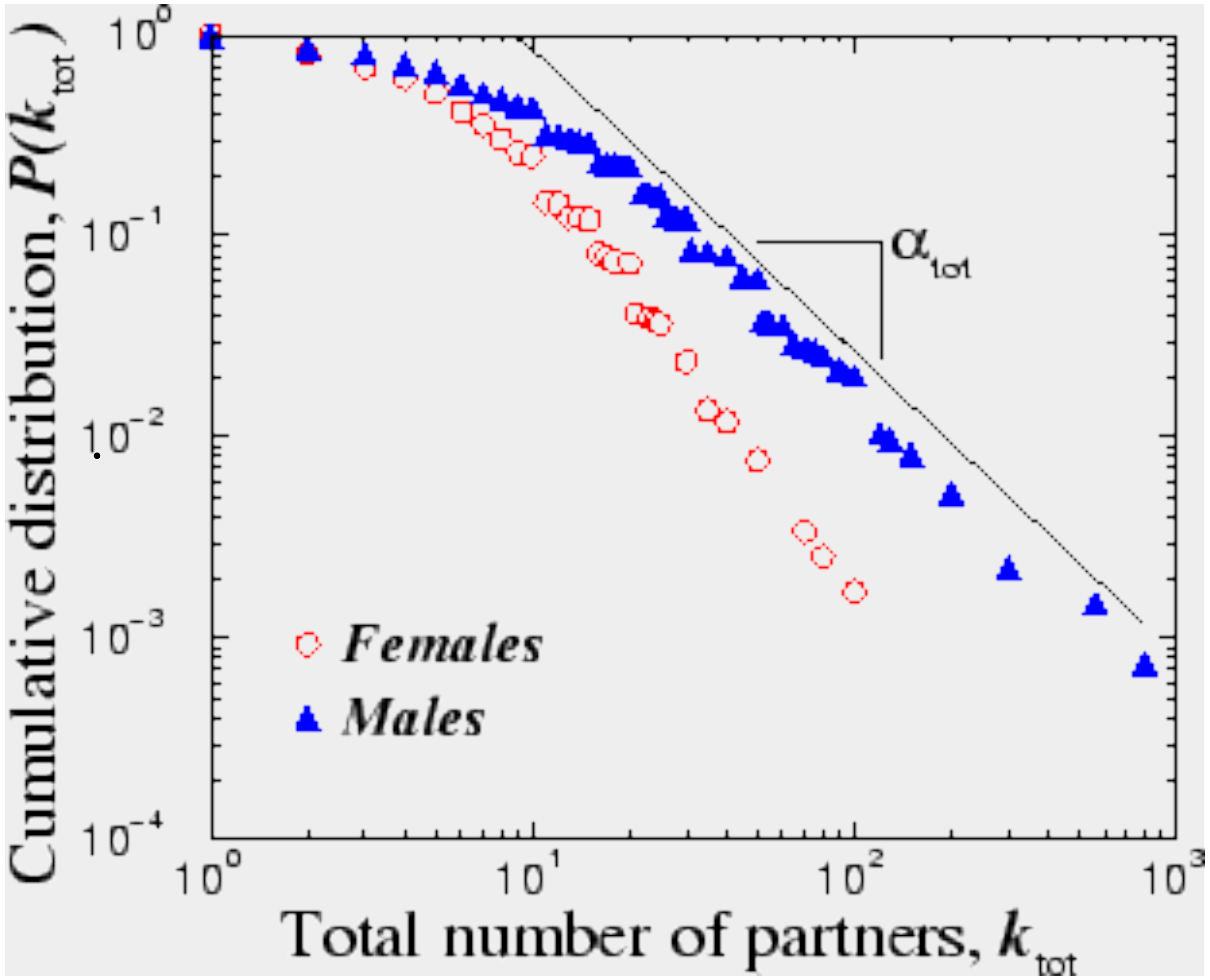


Found





Jeong et al., Nature 2001

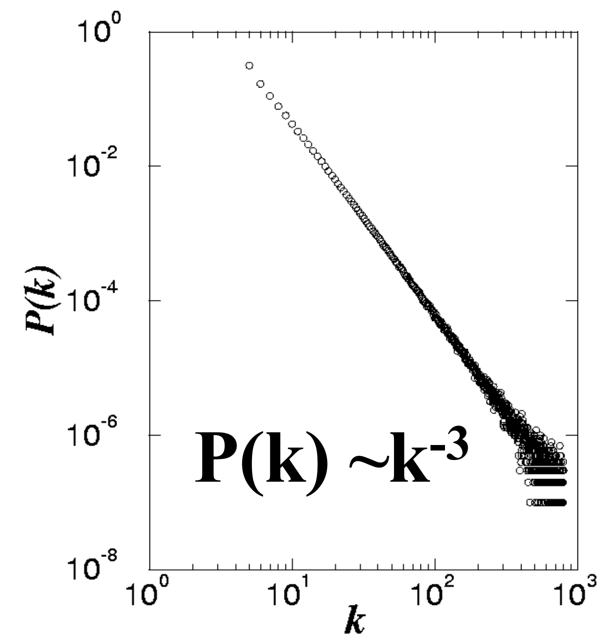


Liljeros et al., Nature 2001

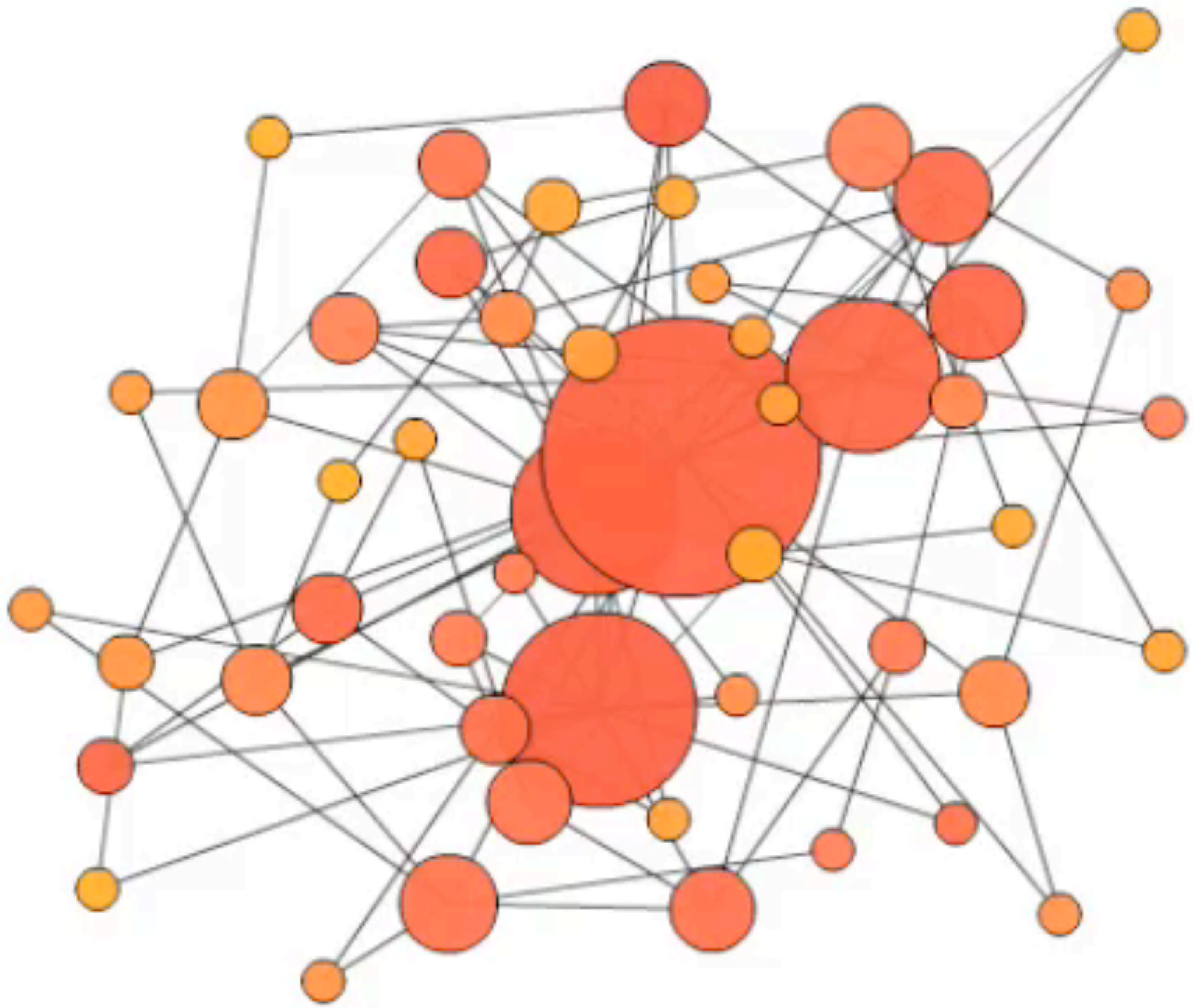
GROWTH:
add a new node with m links

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

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

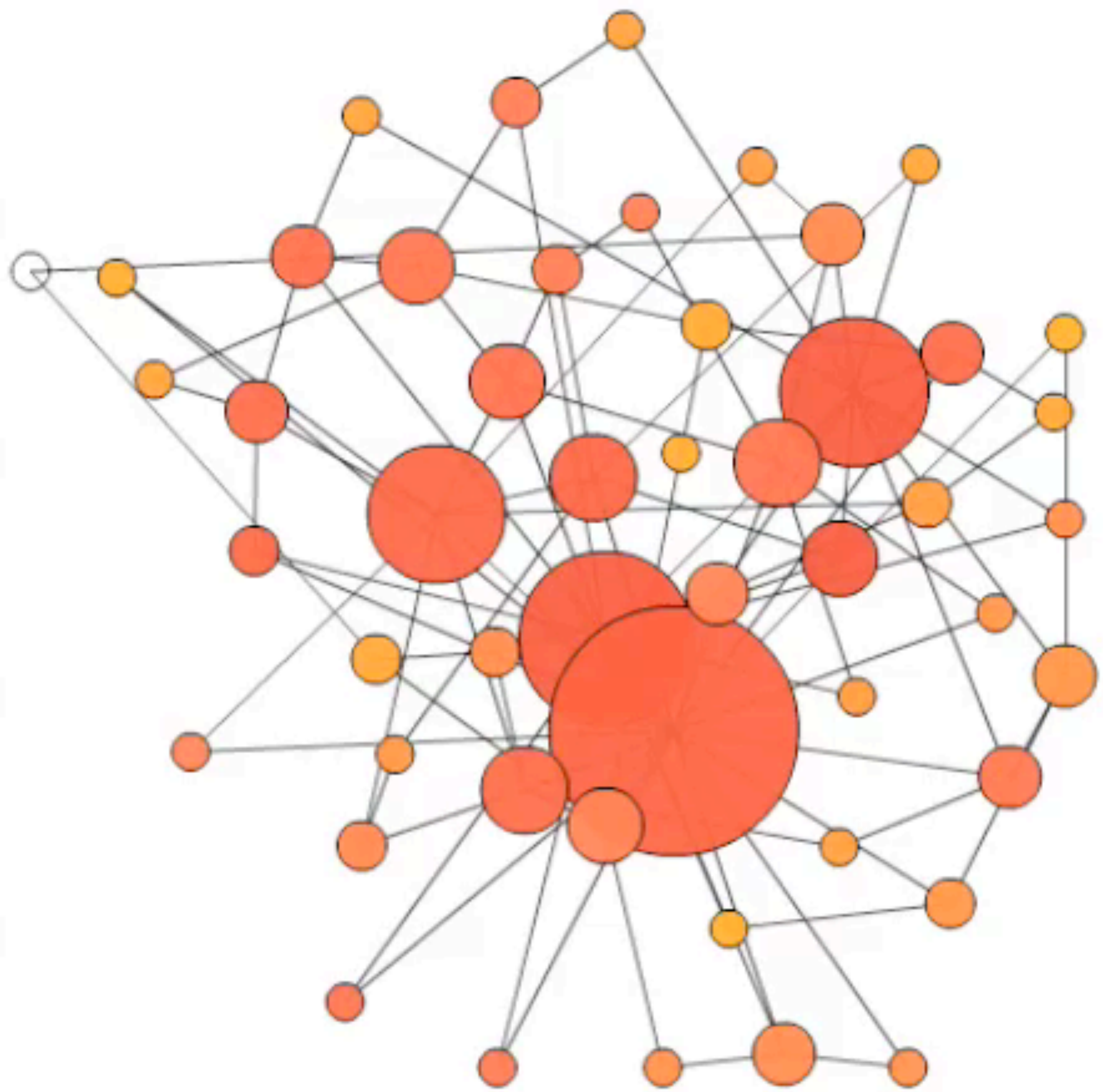


Error



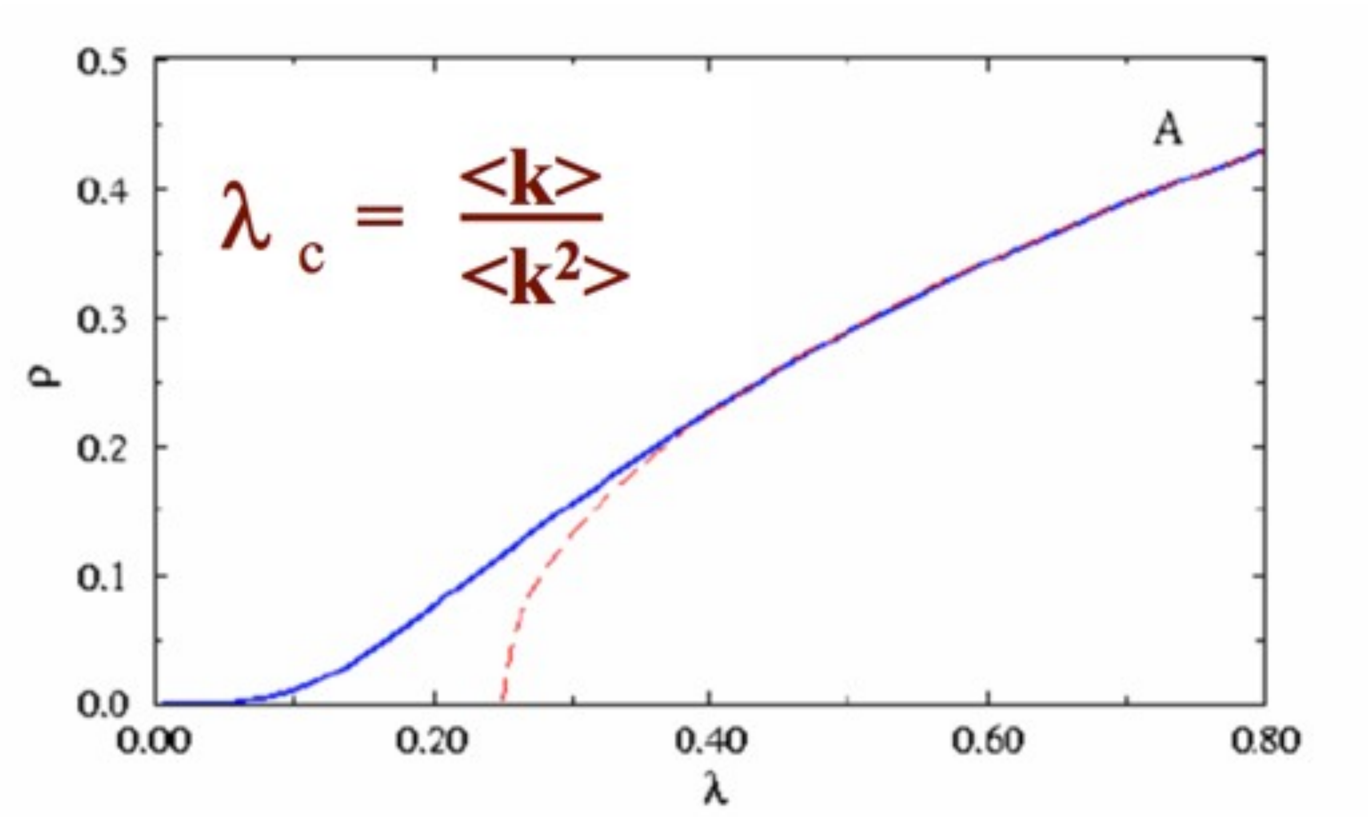
Error

Attack



Attack

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



Alessandro Vespignani

Epidemic spreading: “following links”

Epidemic spreading: “following links”

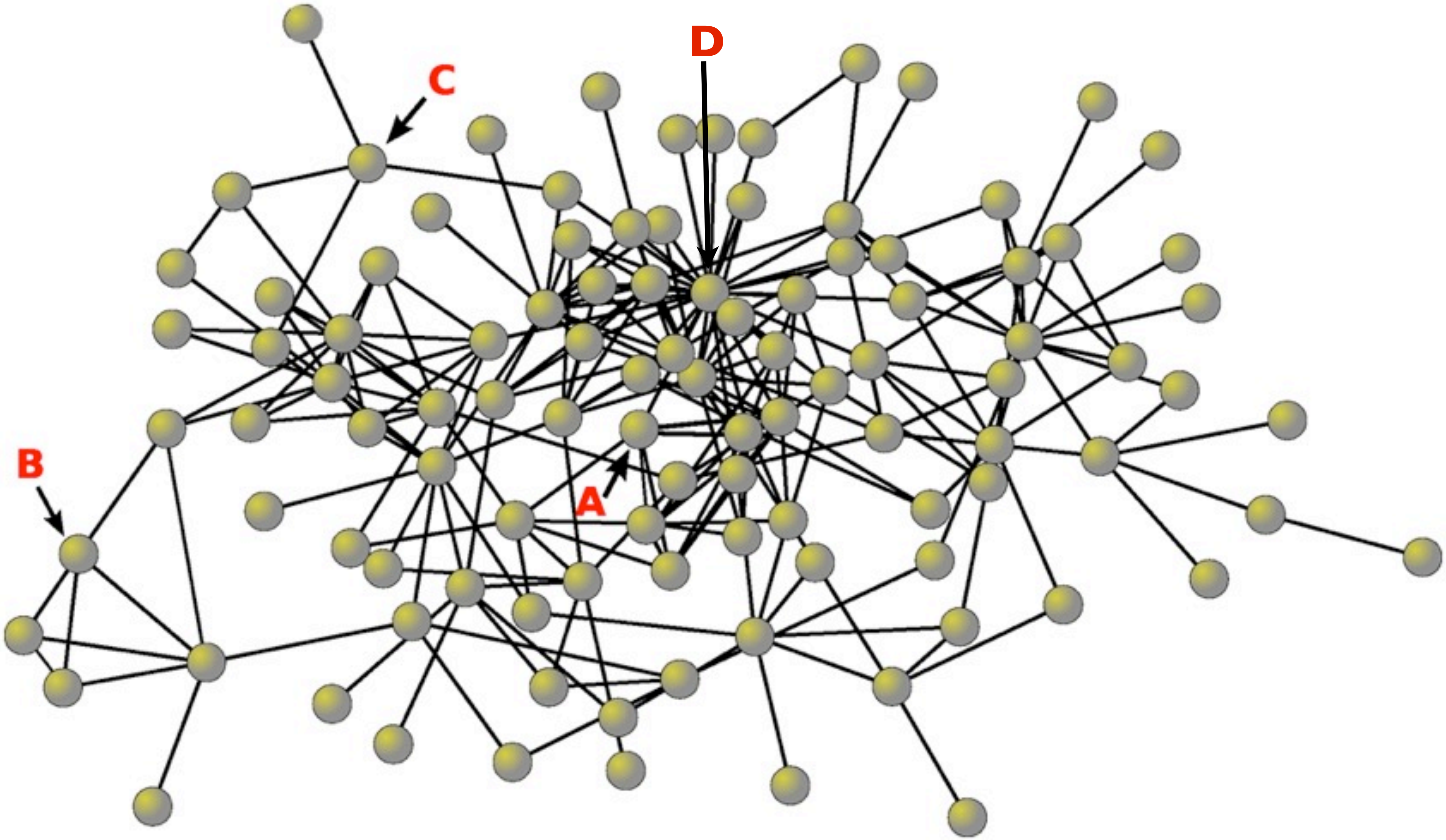
“Friendship Paradox”

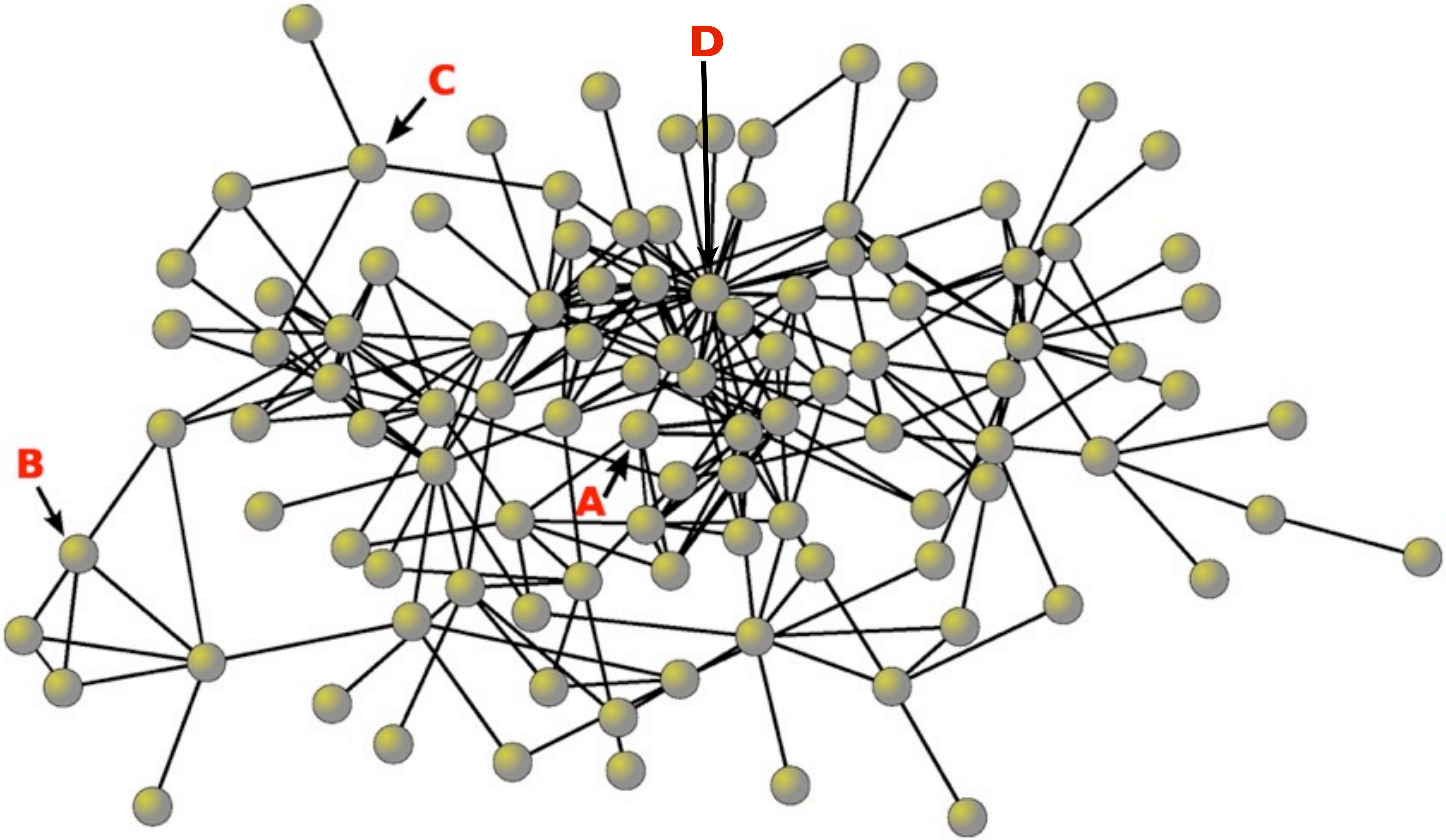
Epidemic spreading: "following links"

"Friendship Paradox"

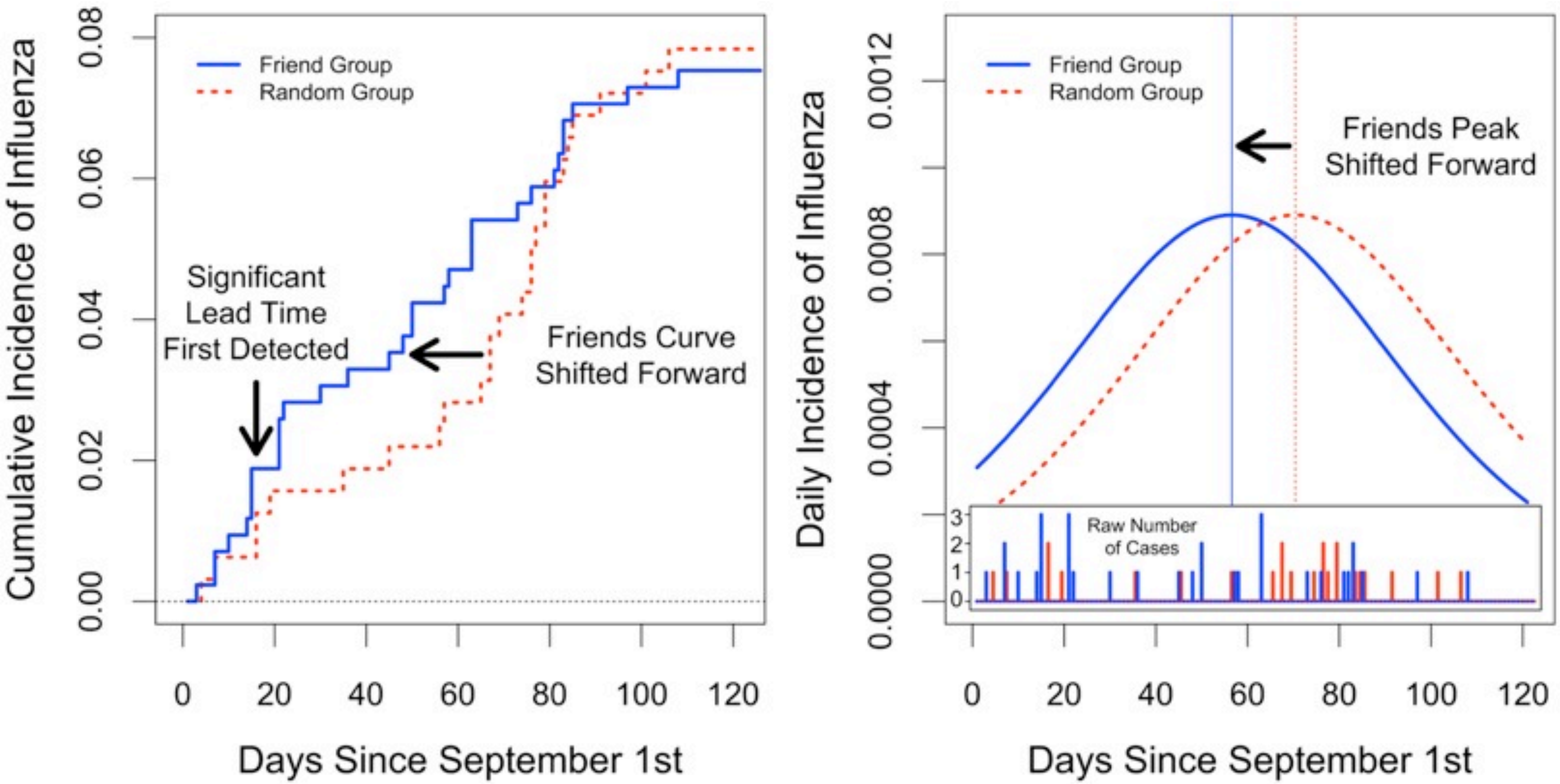
The disease quickly get to the **hubs**

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


D
C
B
A



“Hubs”

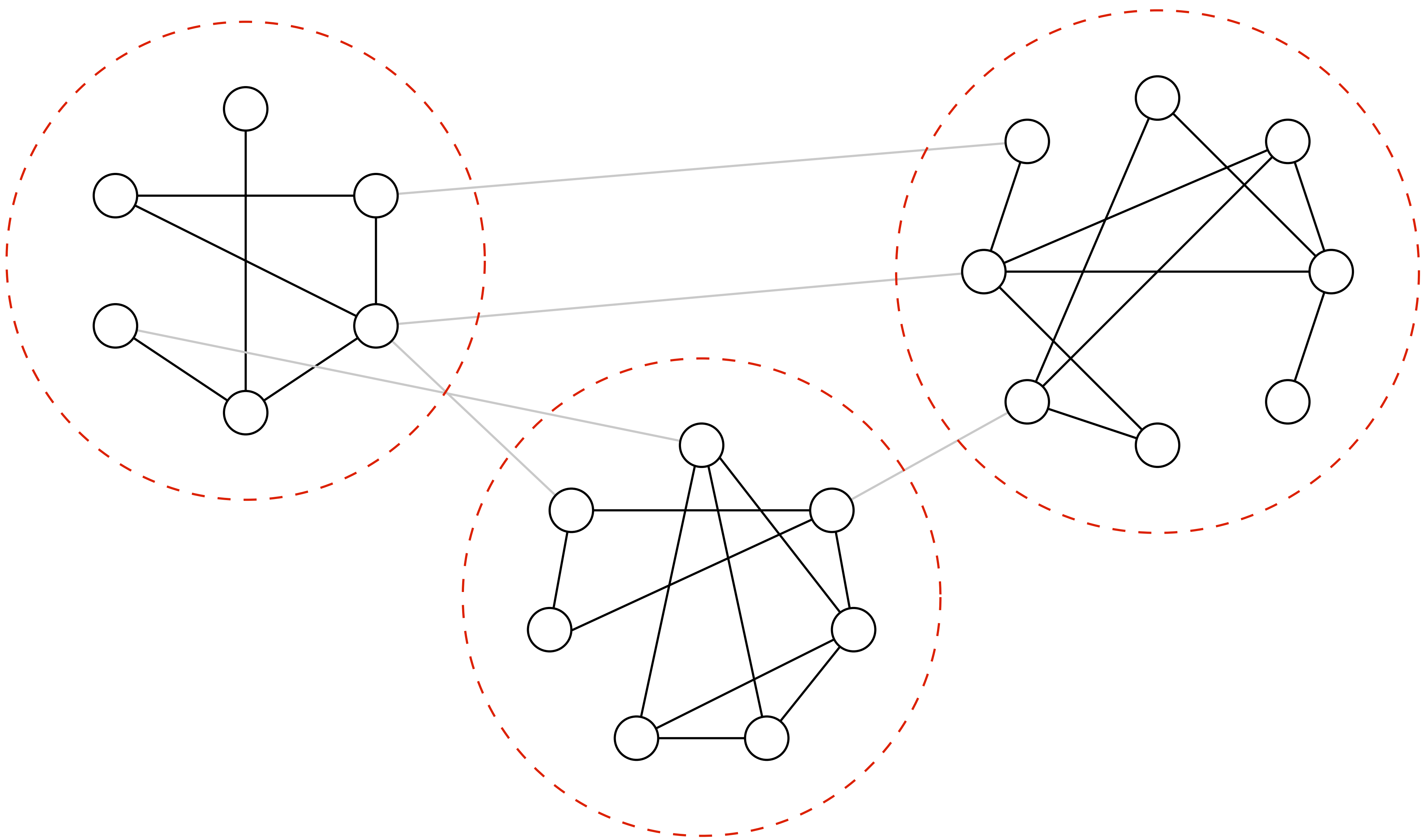


Random person -> immunize a random friend of the person (but **not the original one!**)

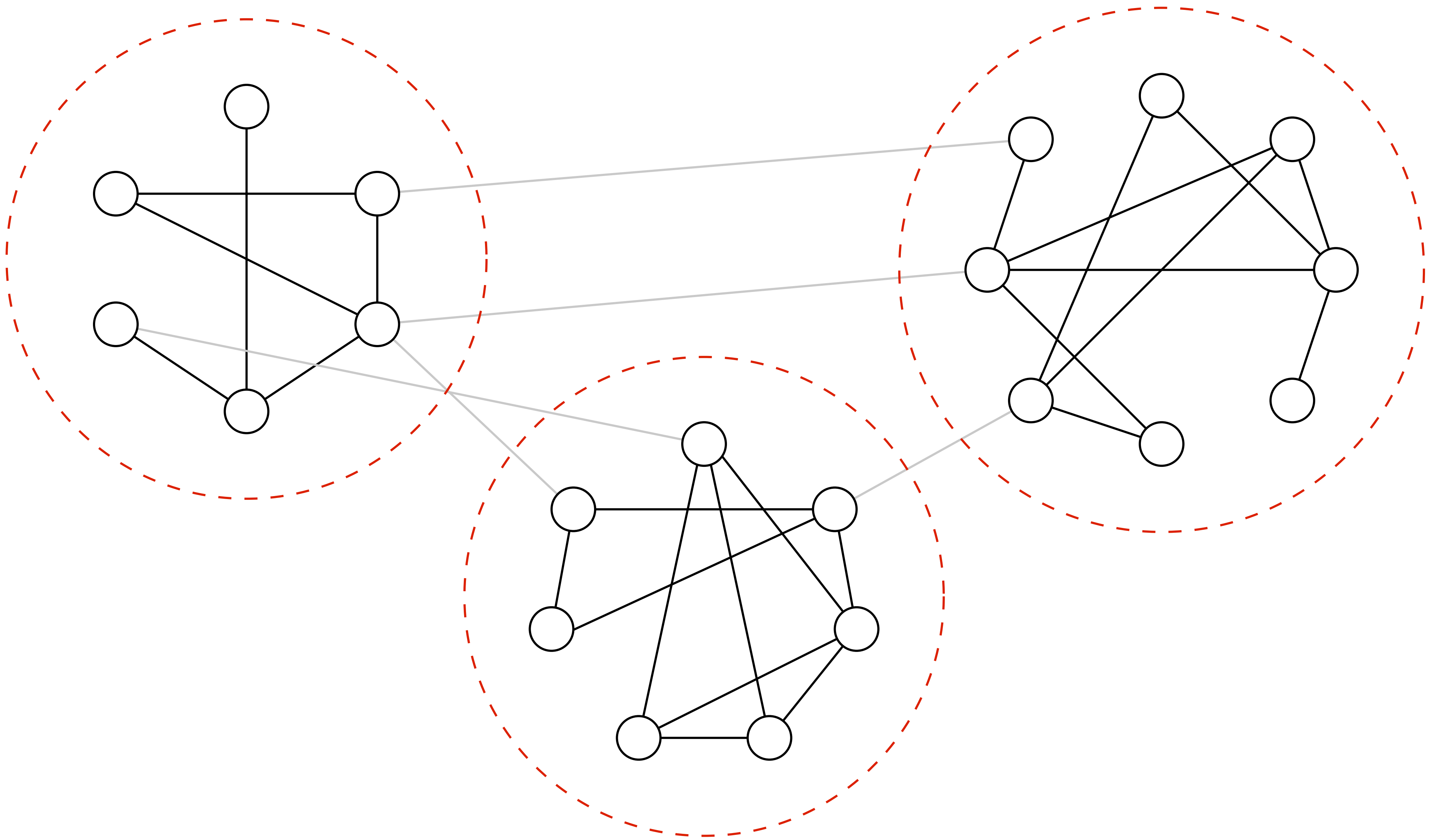
Communities

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

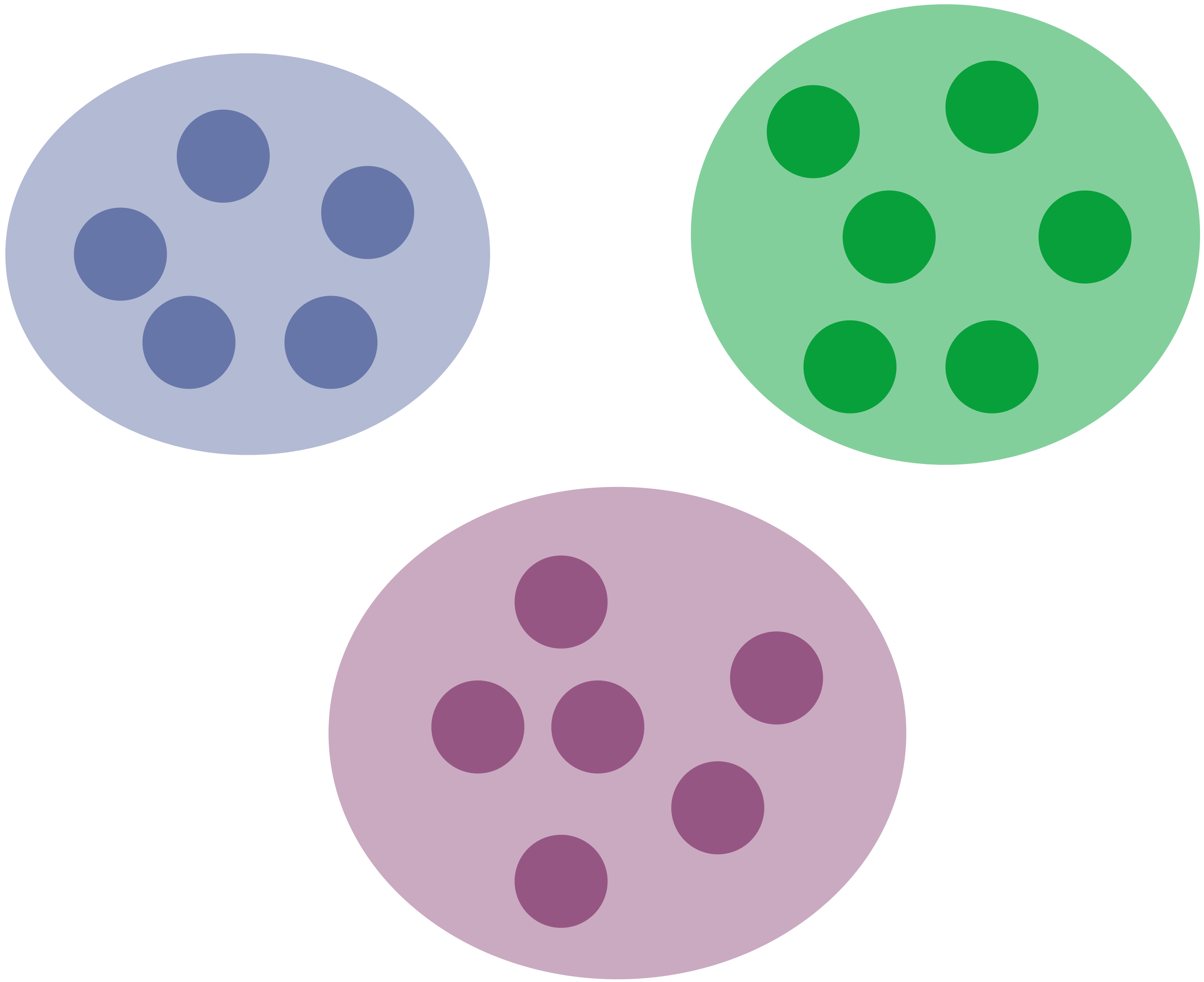
“a group of densely interconnected nodes”

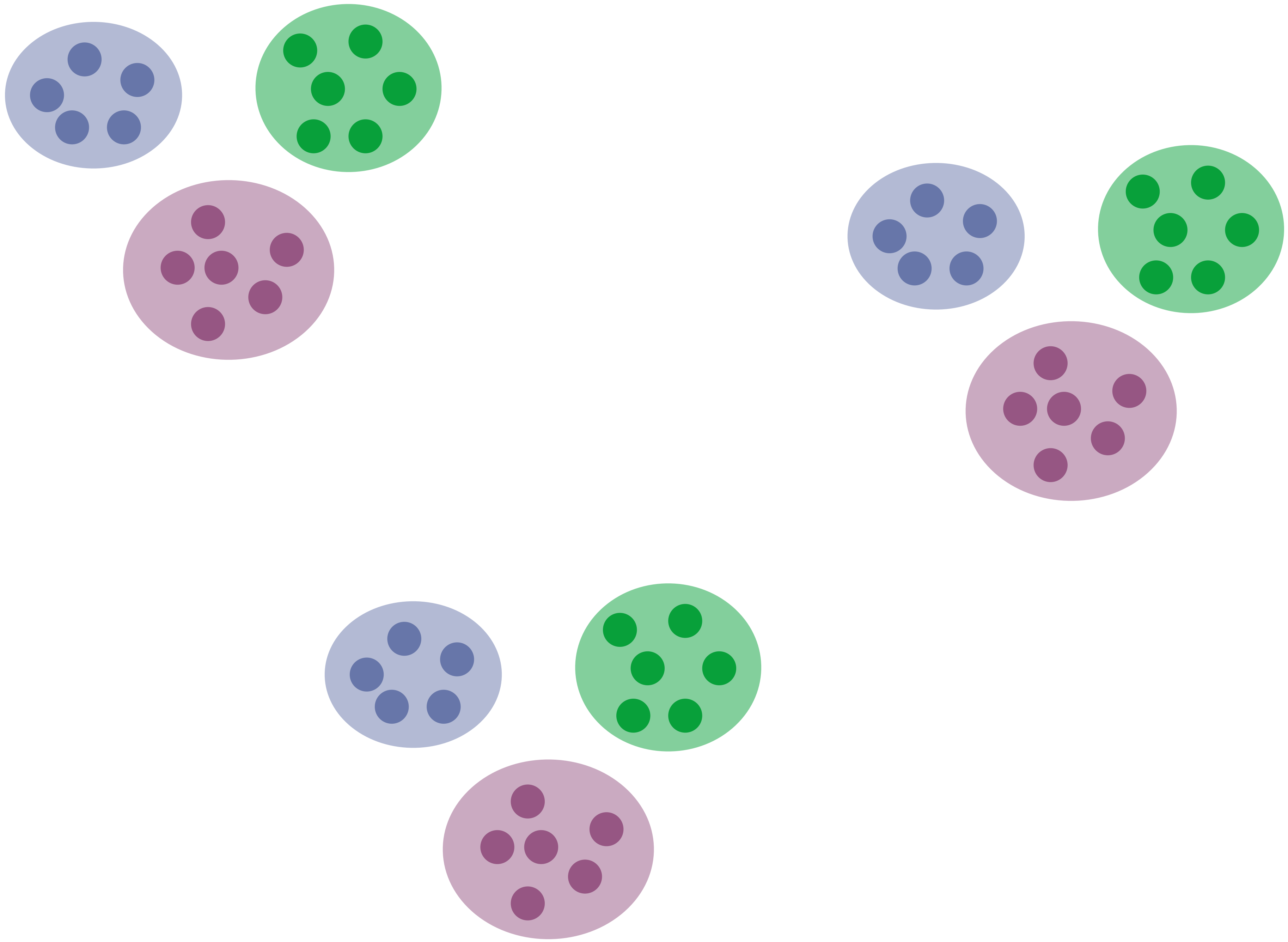


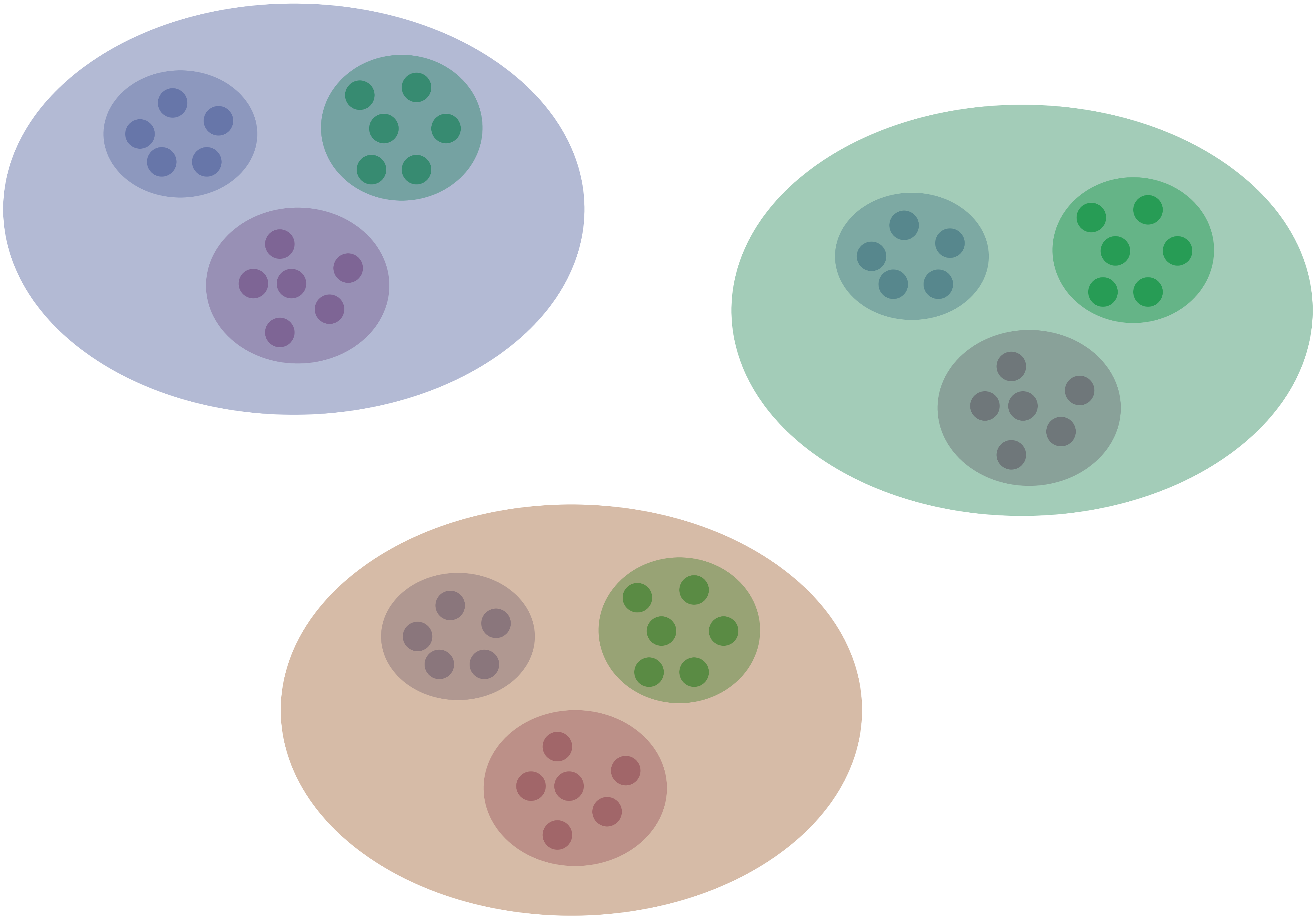
"a group of densely interconnected nodes"



Hierarchy

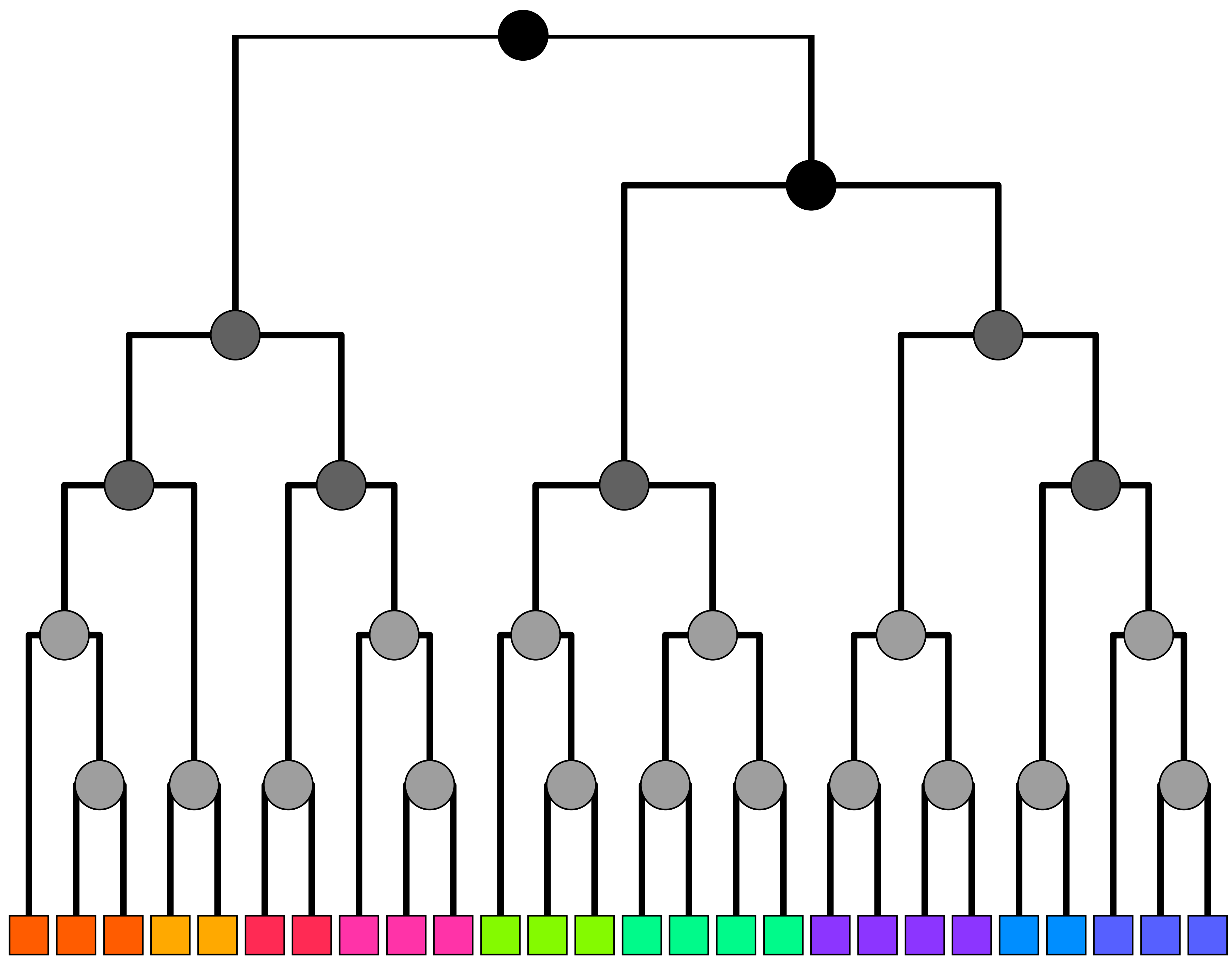




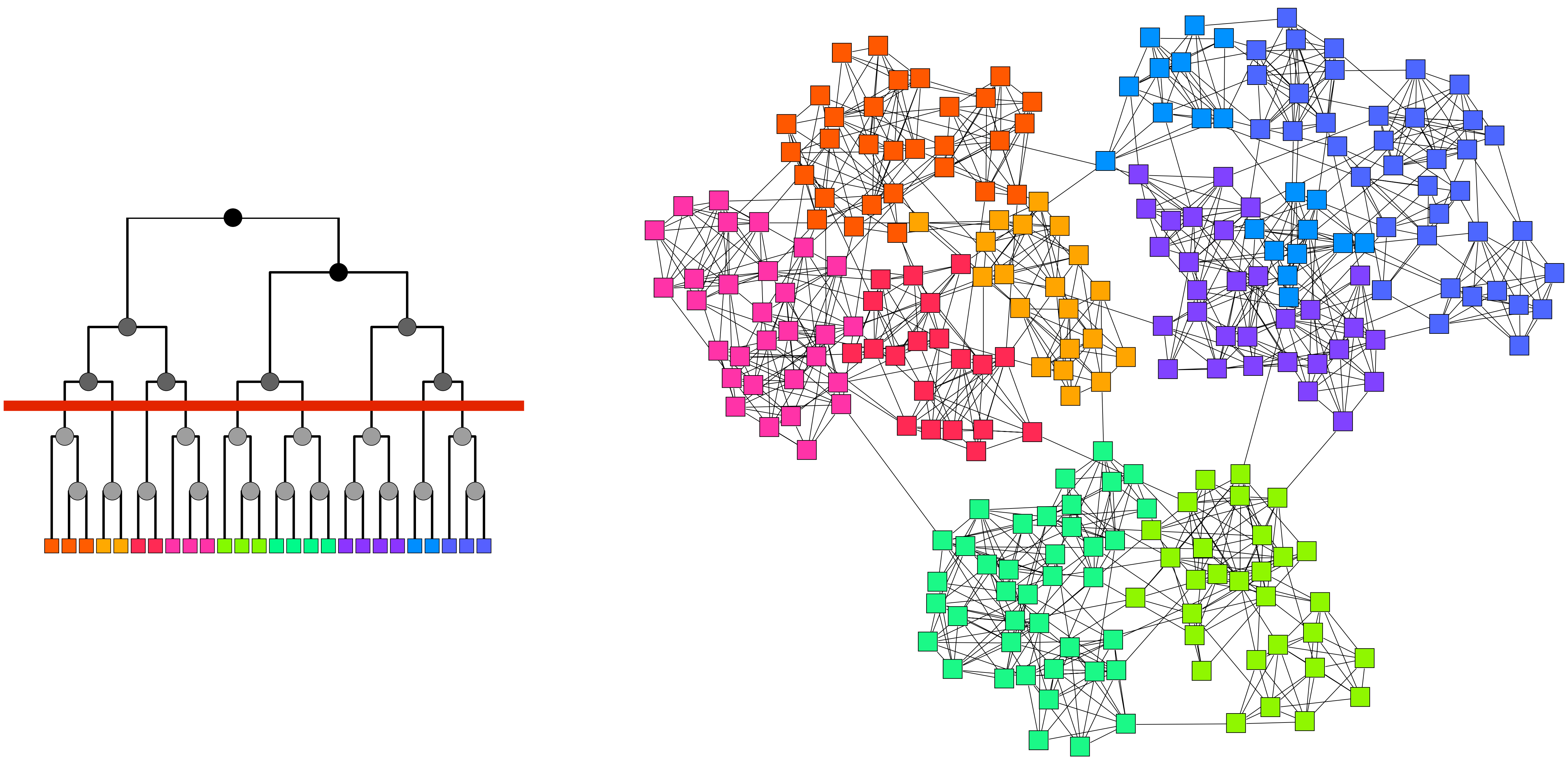


Hierarchy implies **communities.**

Hierarchical Random Graph model



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

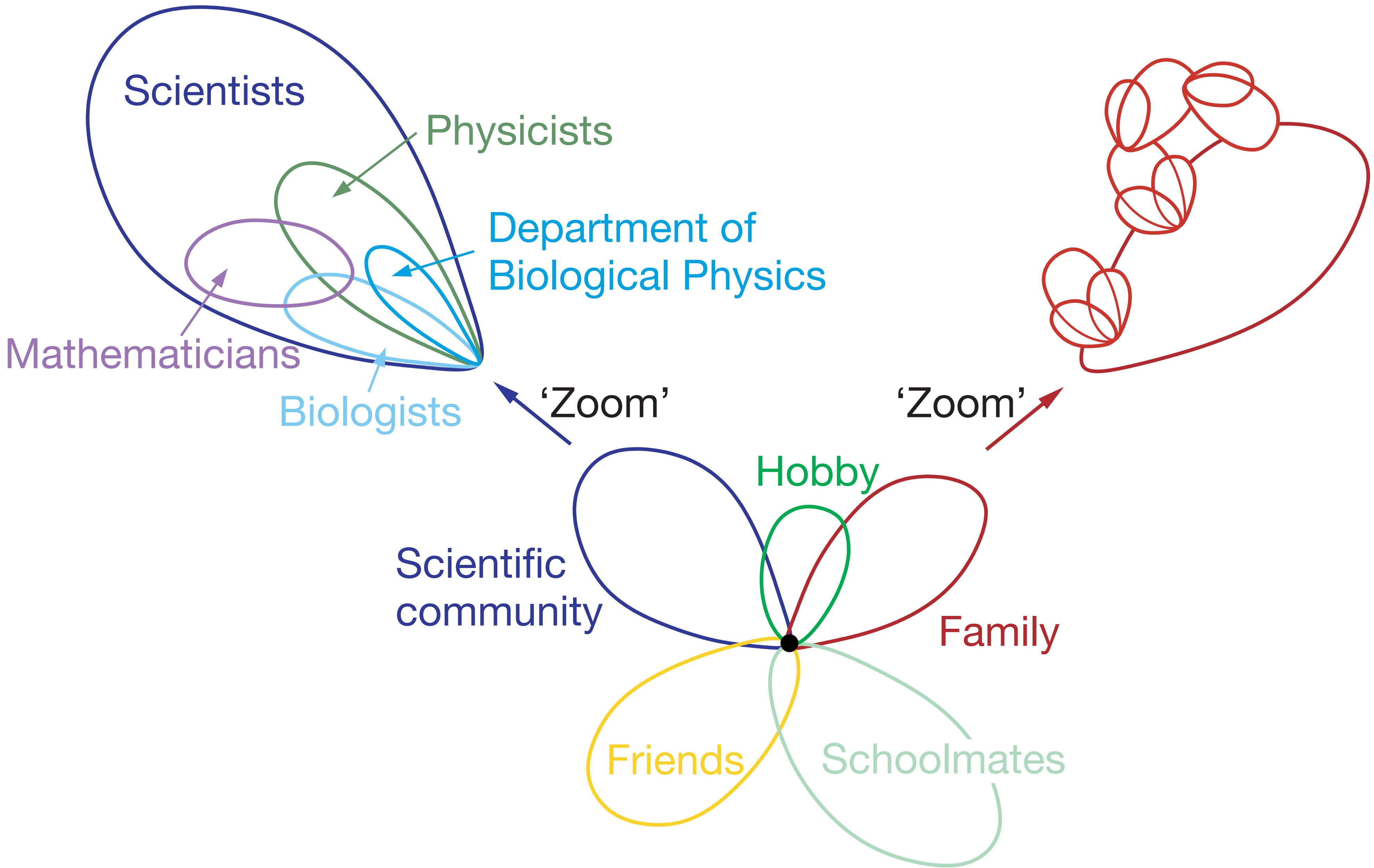


A. Clauset, C. Moore, and M. E. J. Newman, *Nature* (2008)

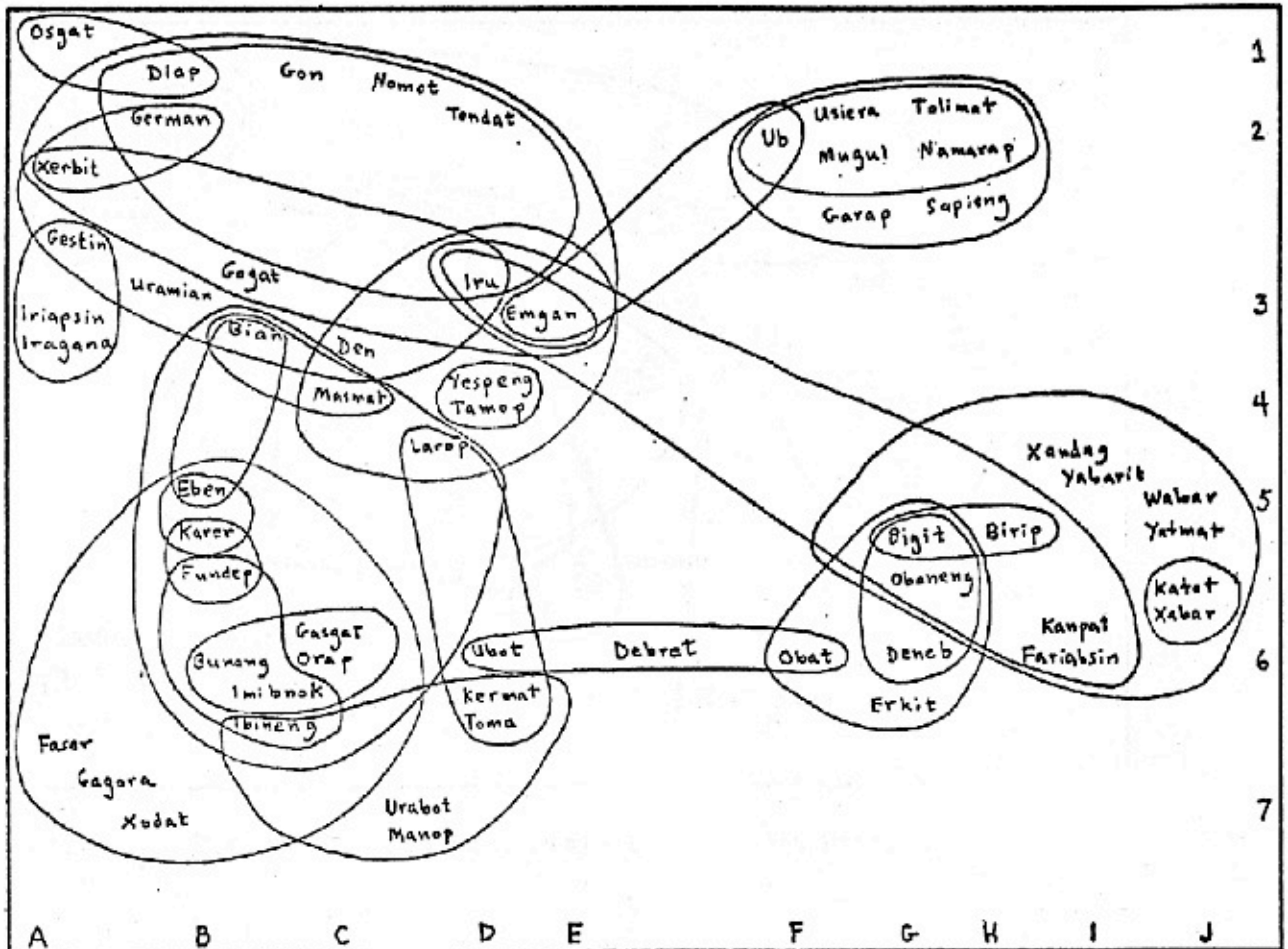
Hierarchical community structure

Hierarchy $\longrightarrow$ Communities

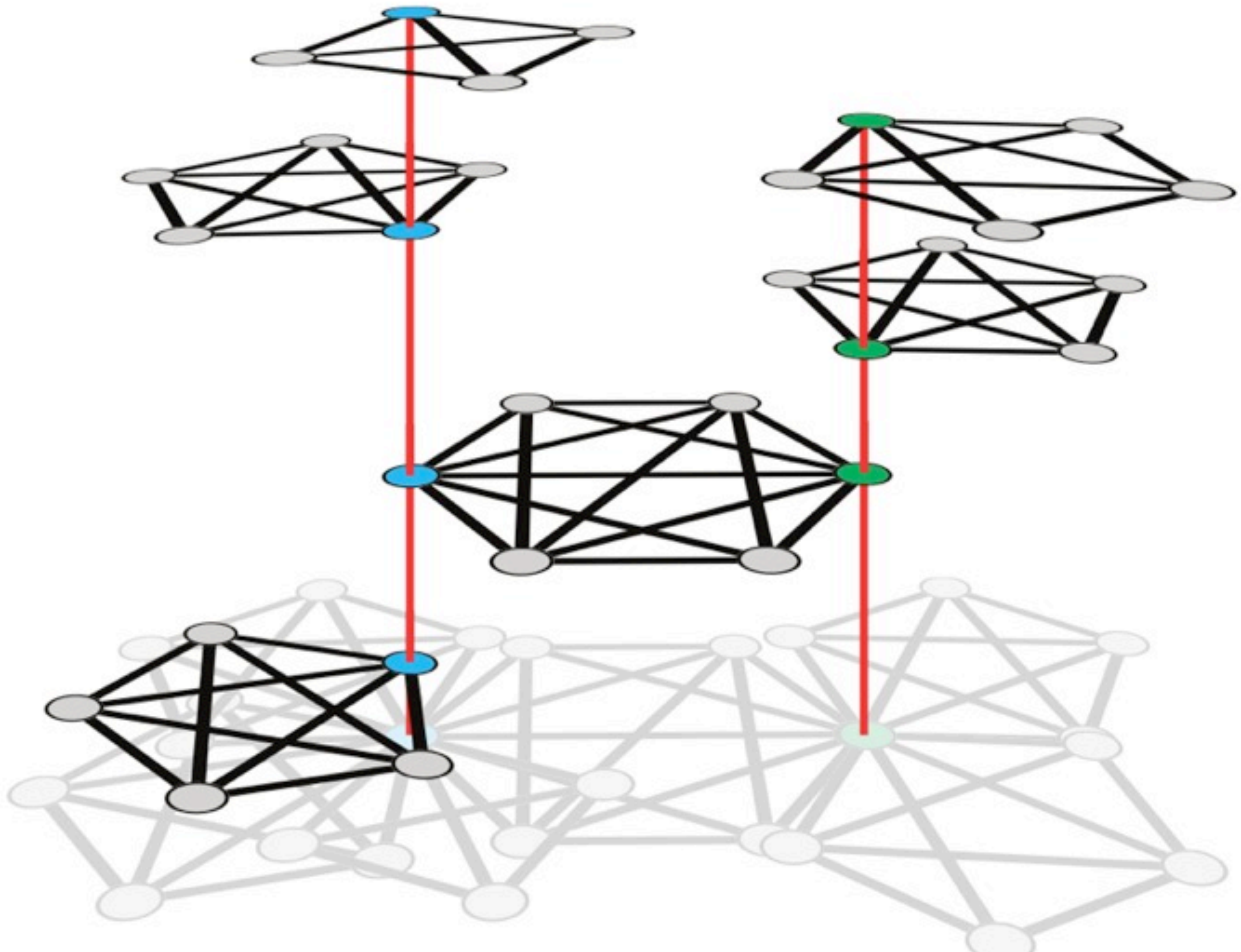
BUT,

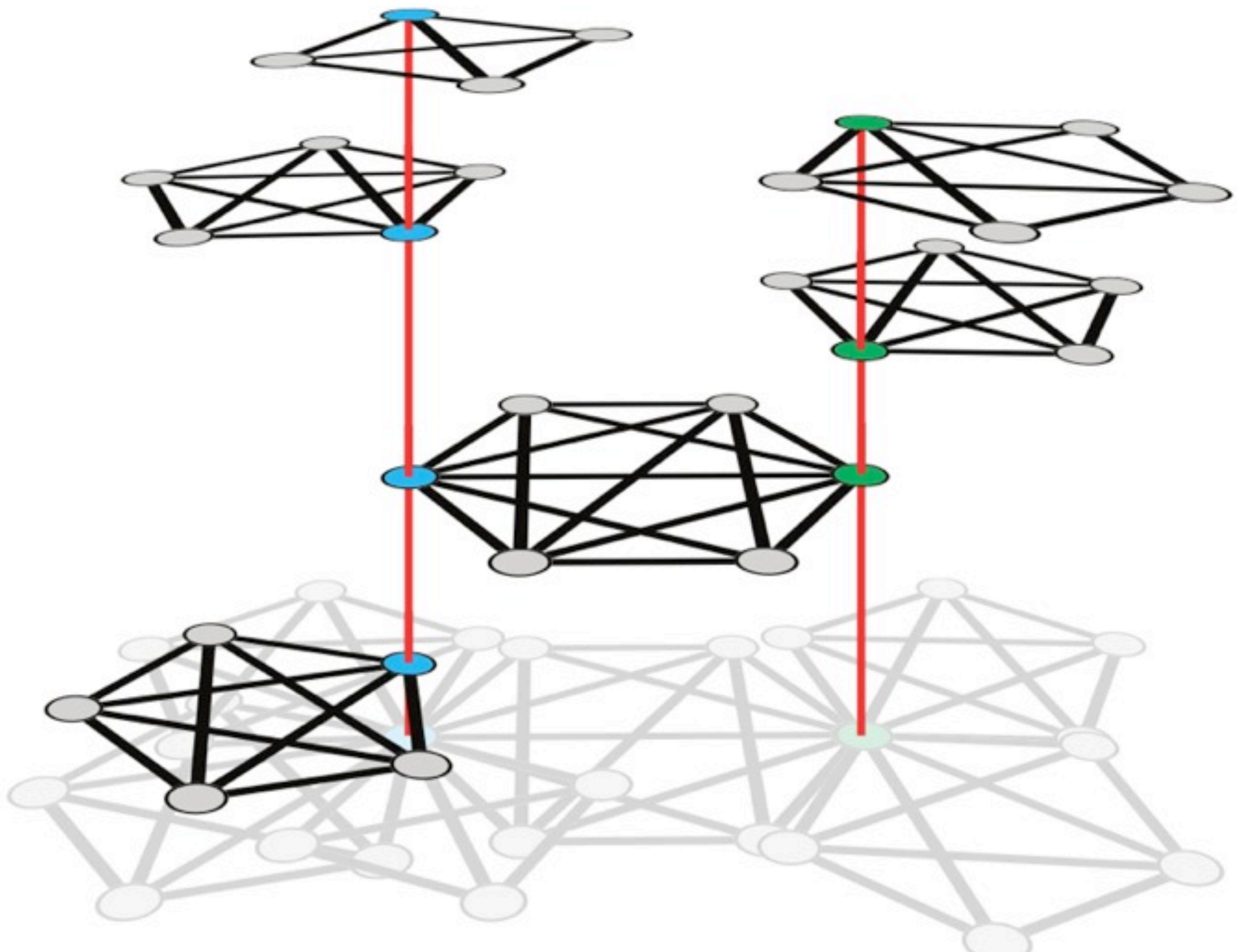


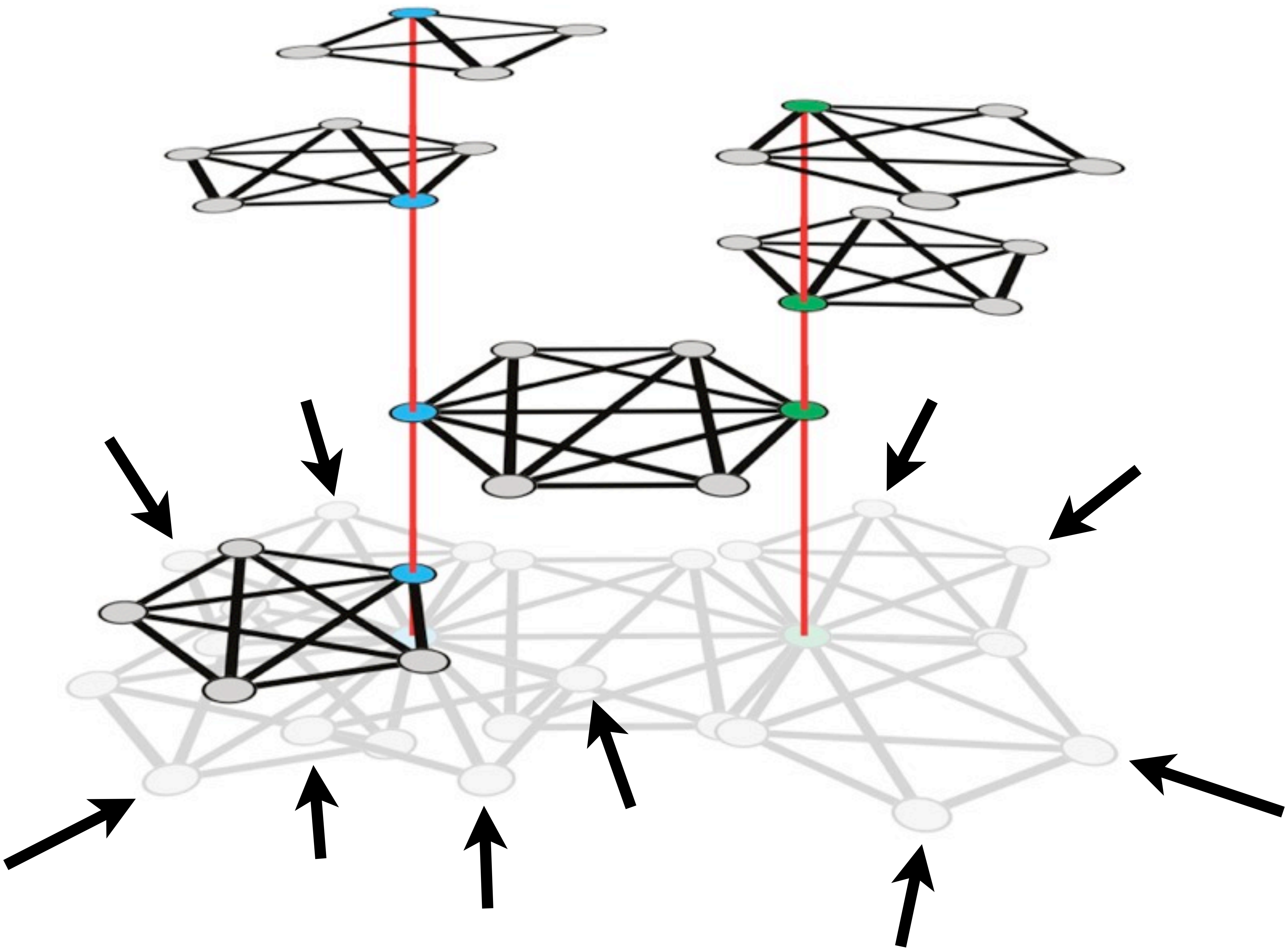
G. Palla, I. Derényi, I. Farkas & T. Vicsek, *Nature* (2005)



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

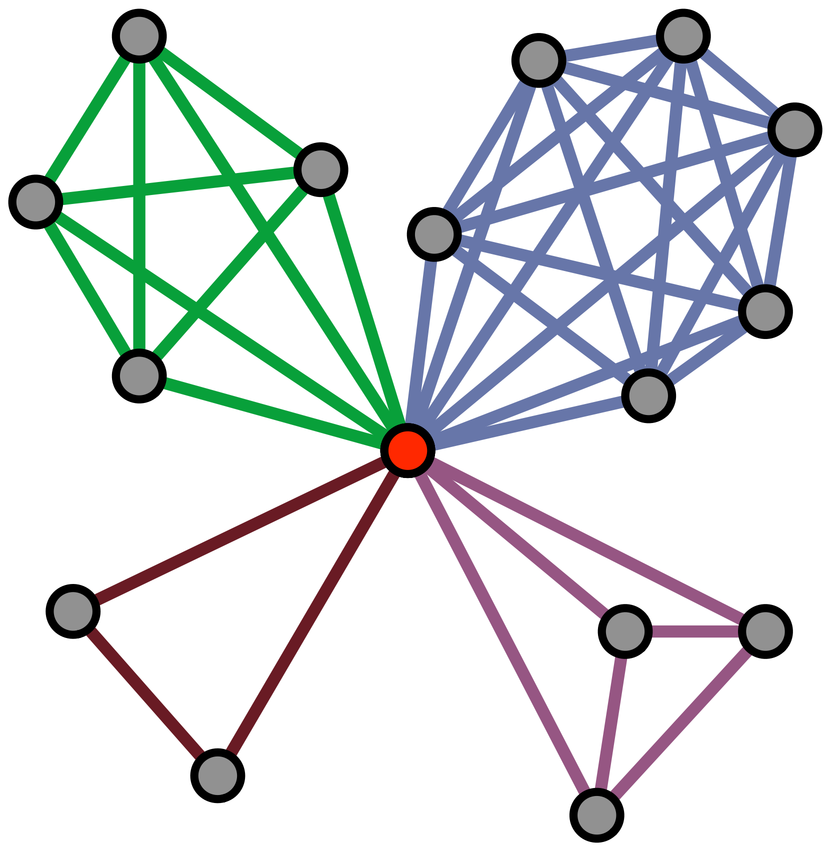




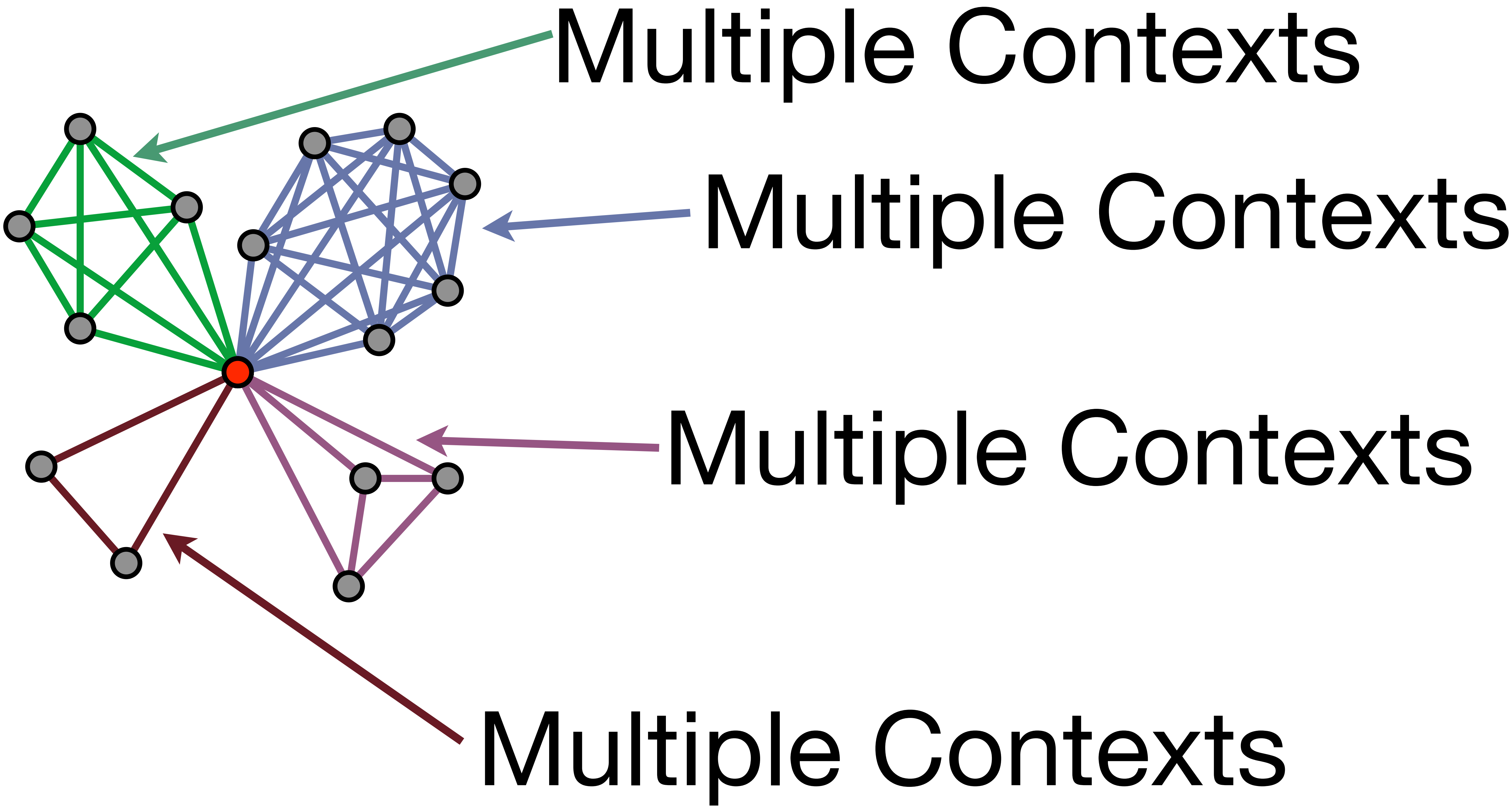


Overlap is **pervasive.**

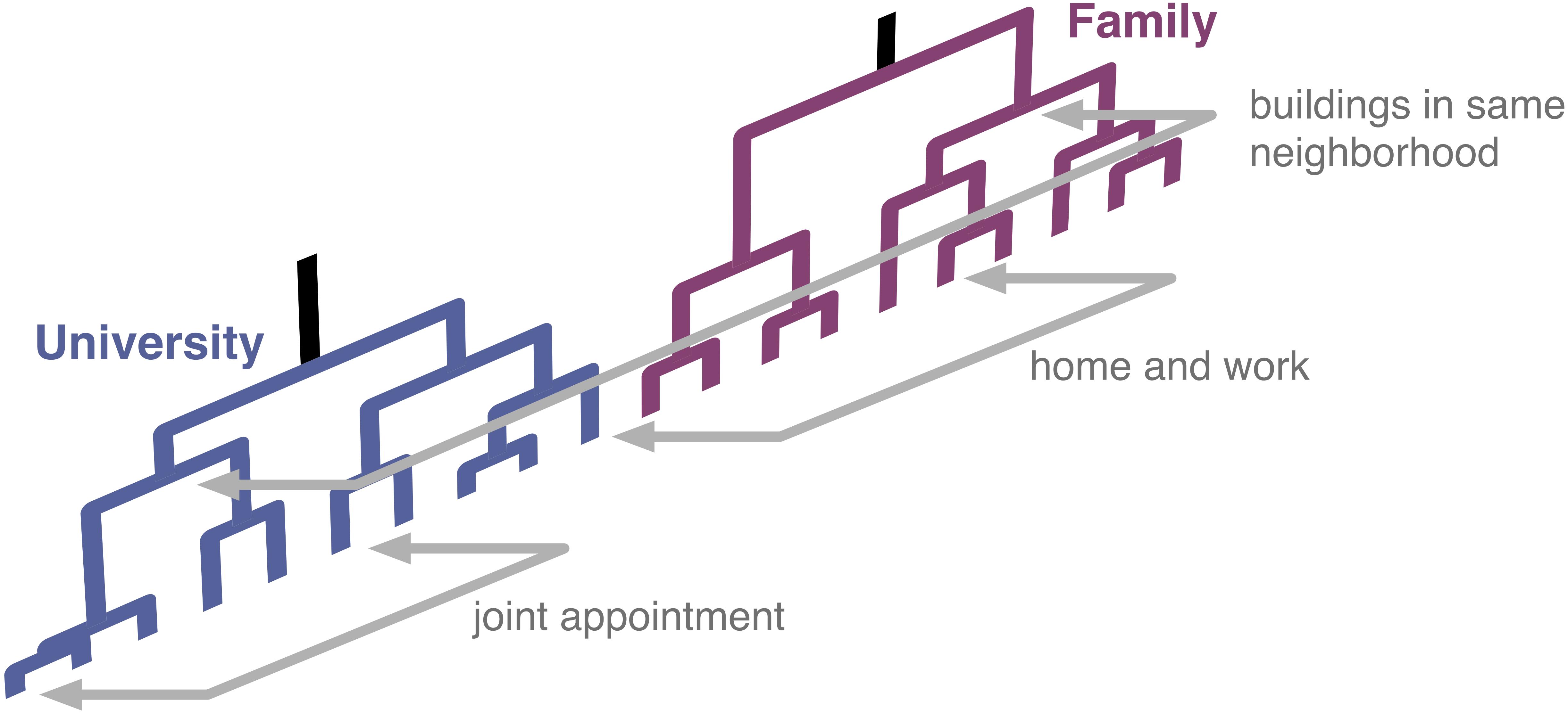
Overlap is **pervasive.**



Multiple Contexts


Multiple Contexts
Multiple Contexts
Multiple Contexts
Multiple Contexts

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

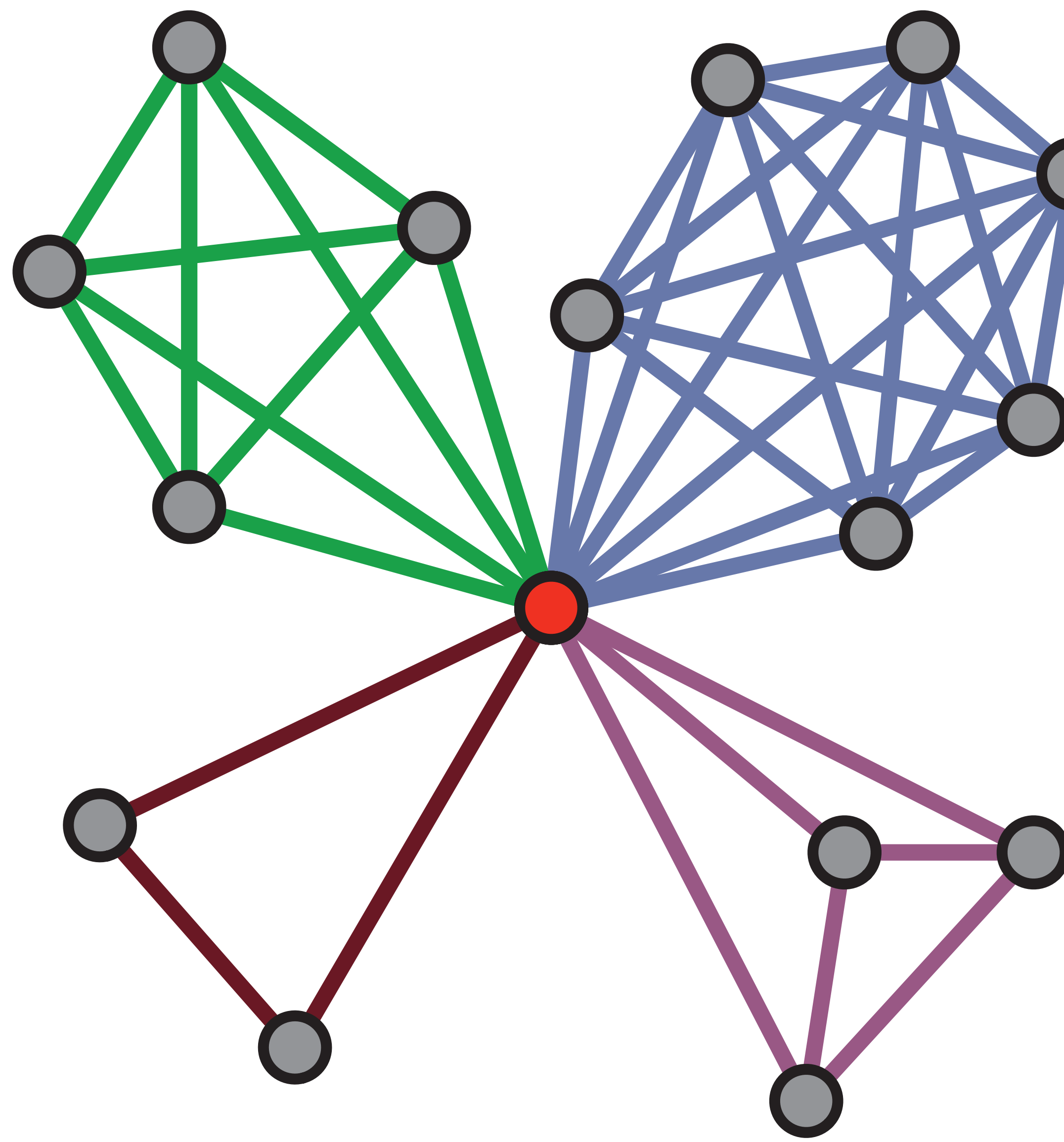


It is **impossible** to obtain a single dendrogram.

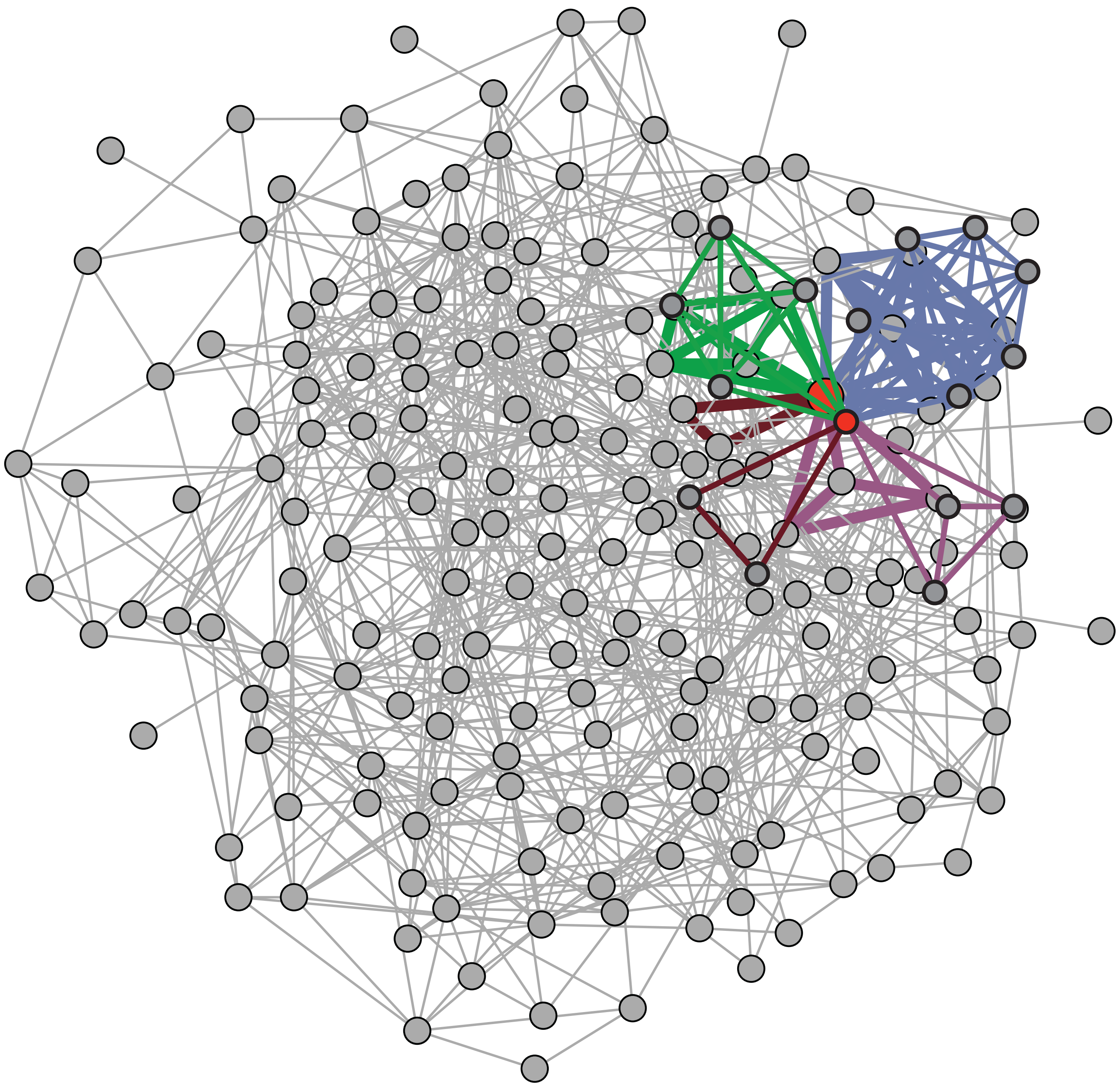
Hierarchical community structure

Hierarchy $\not\longrightarrow$ Communities

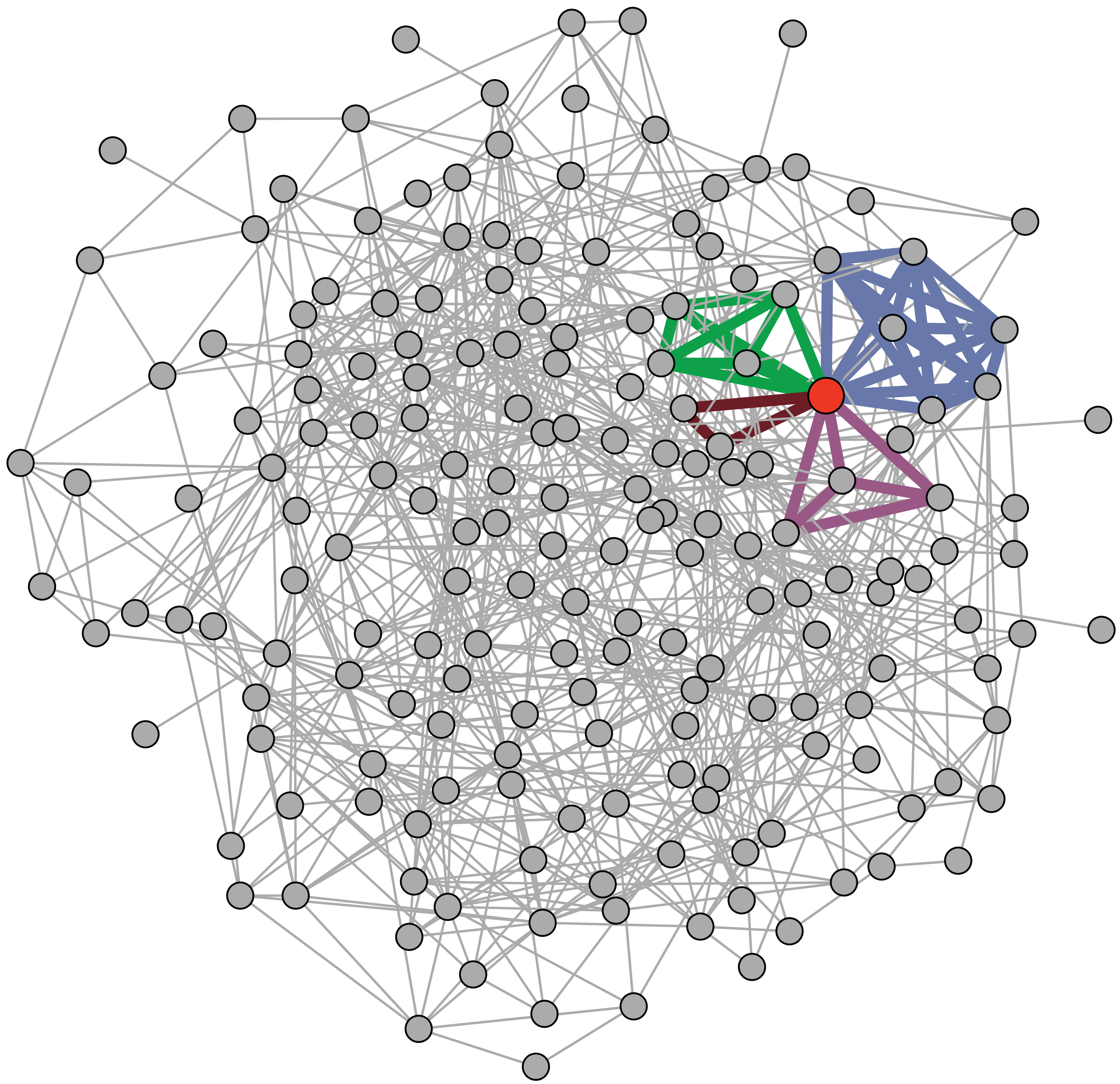
Simple local structure

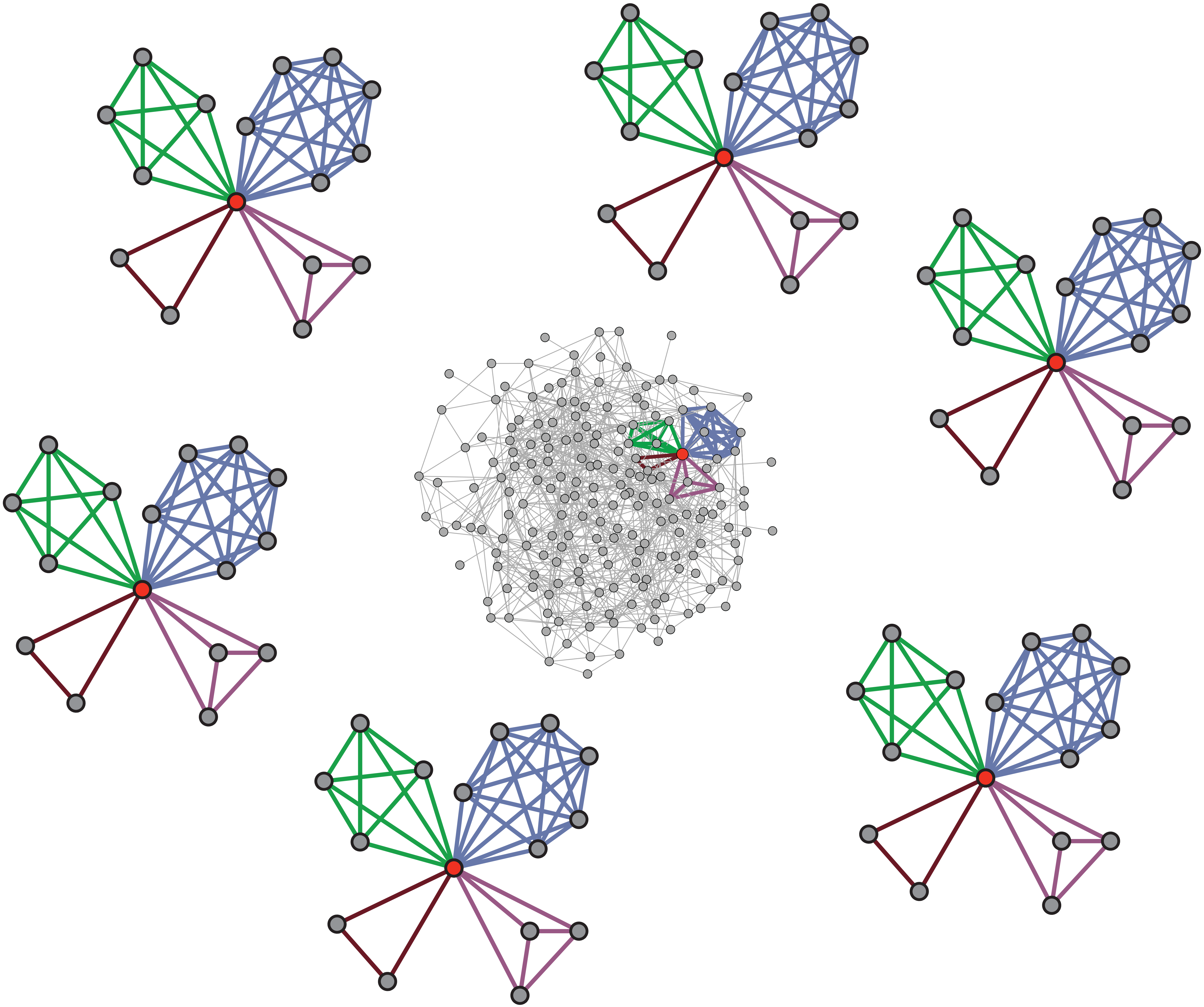


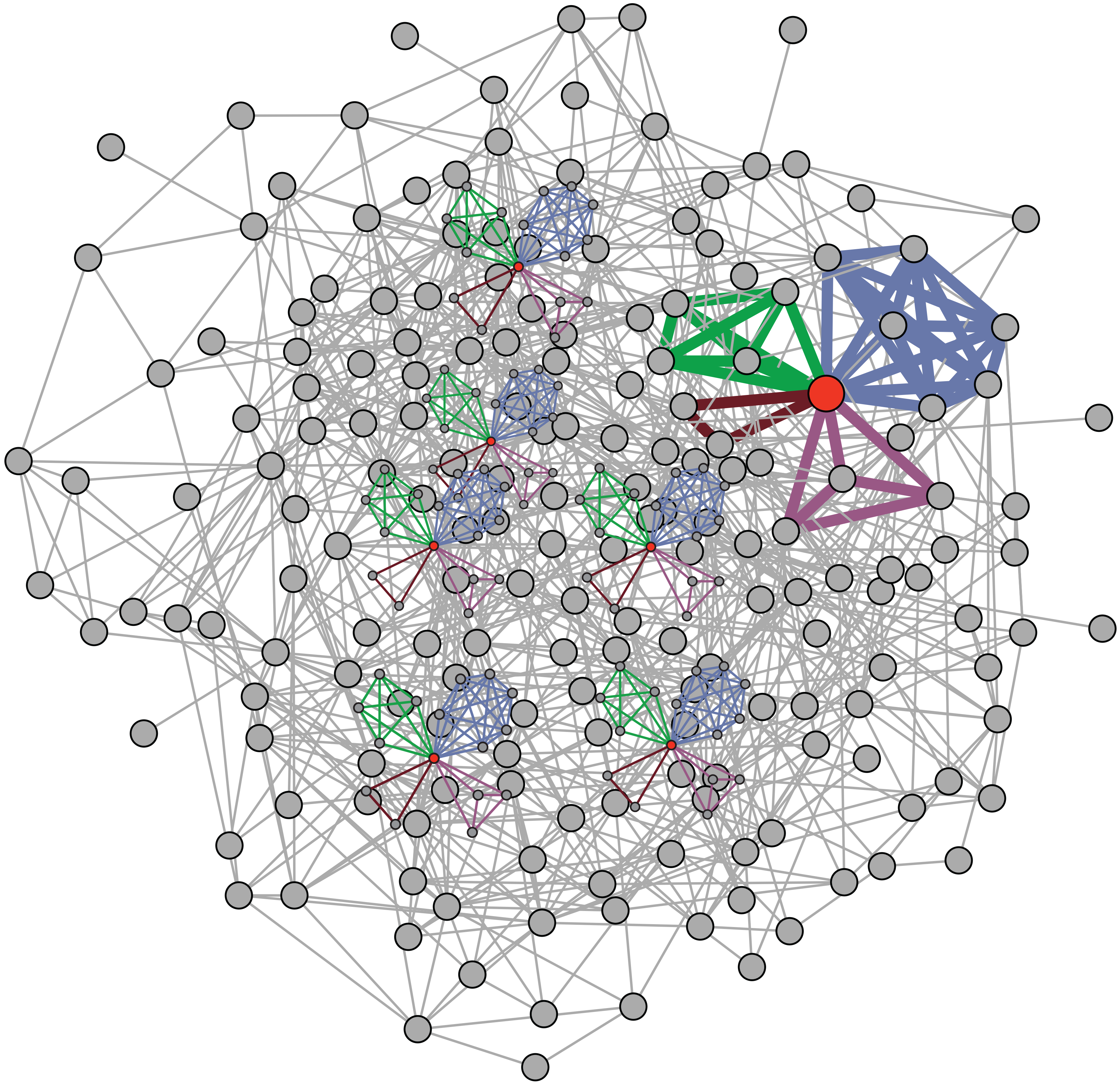
Complex global structure

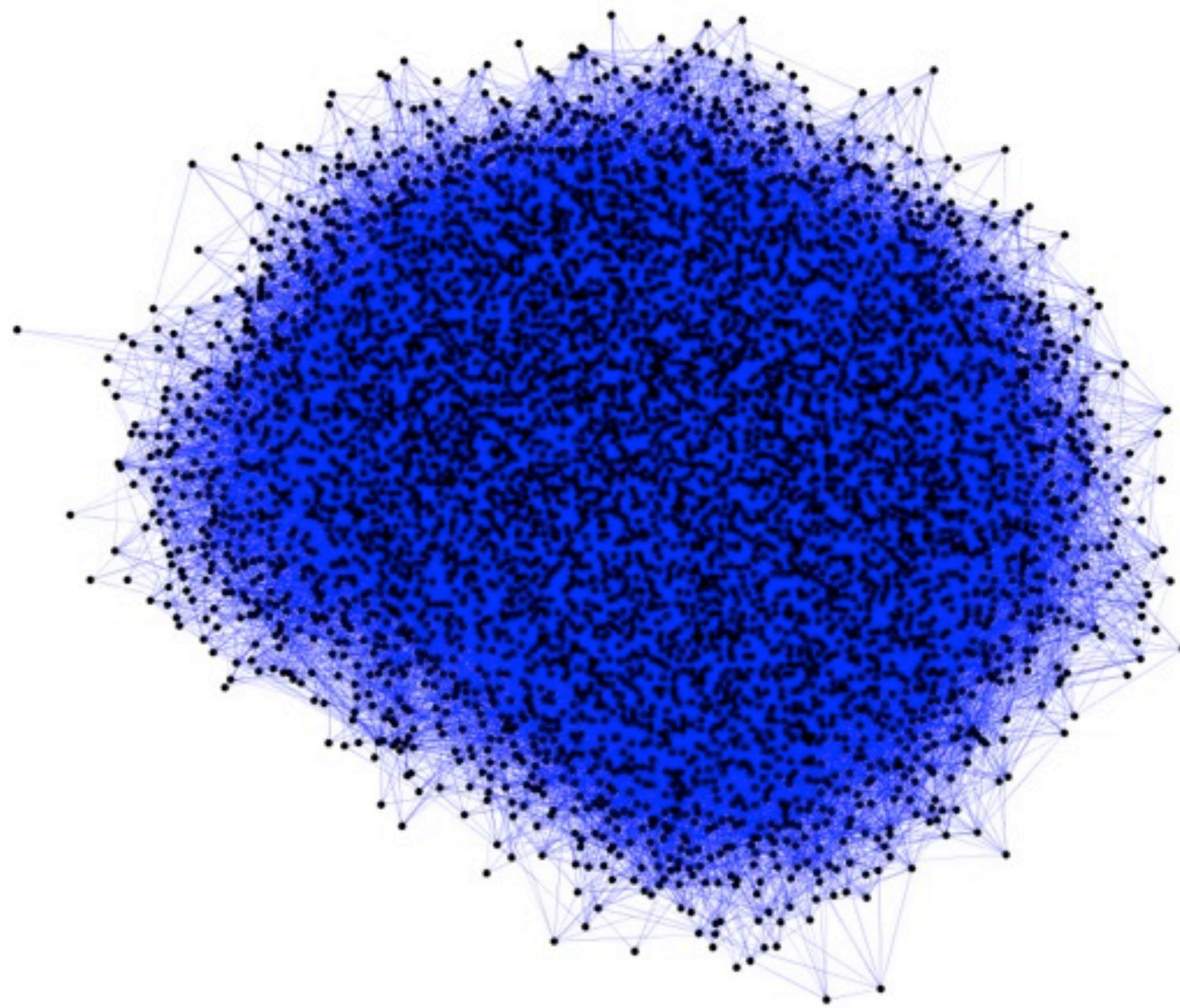


Complex global structure



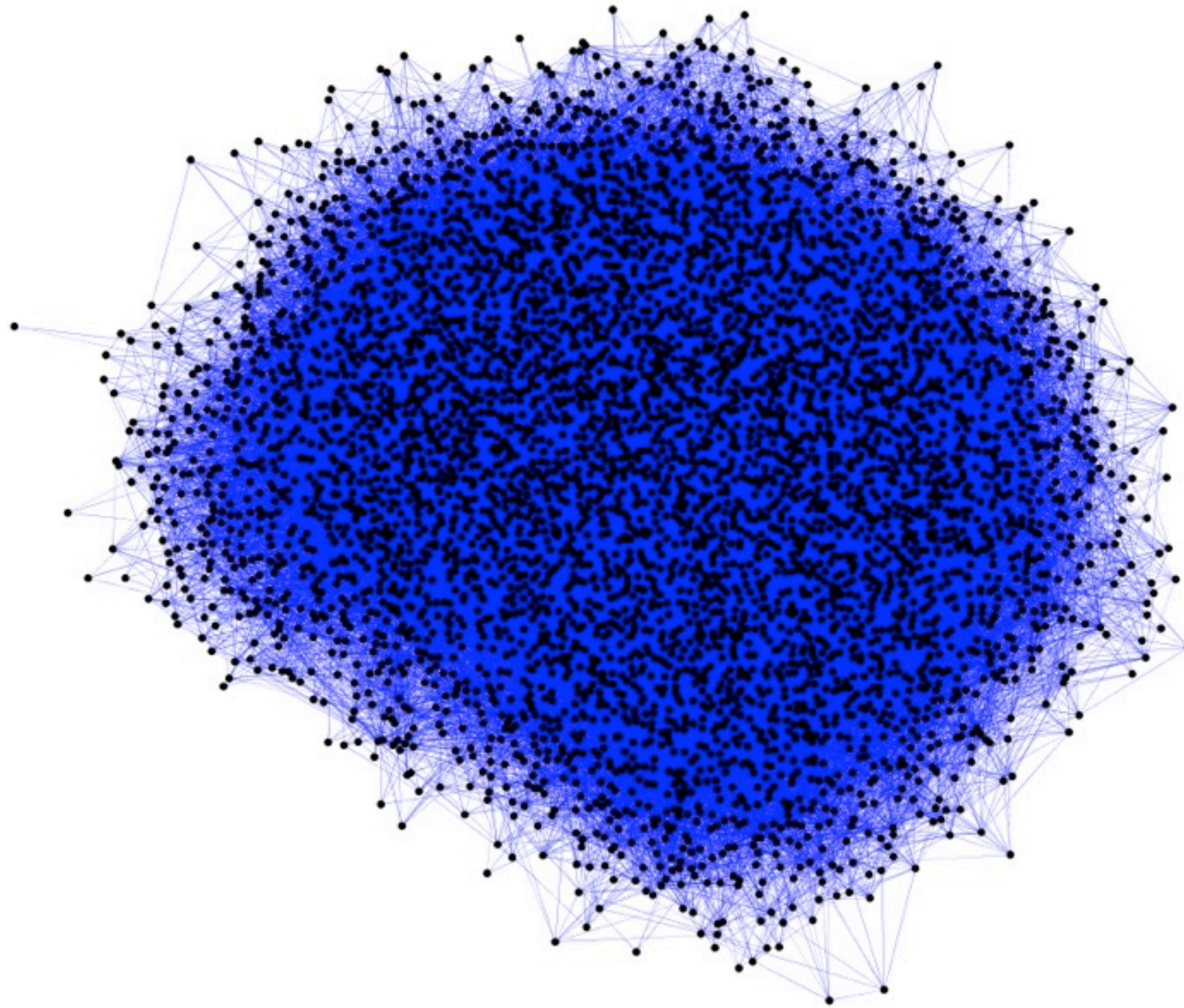




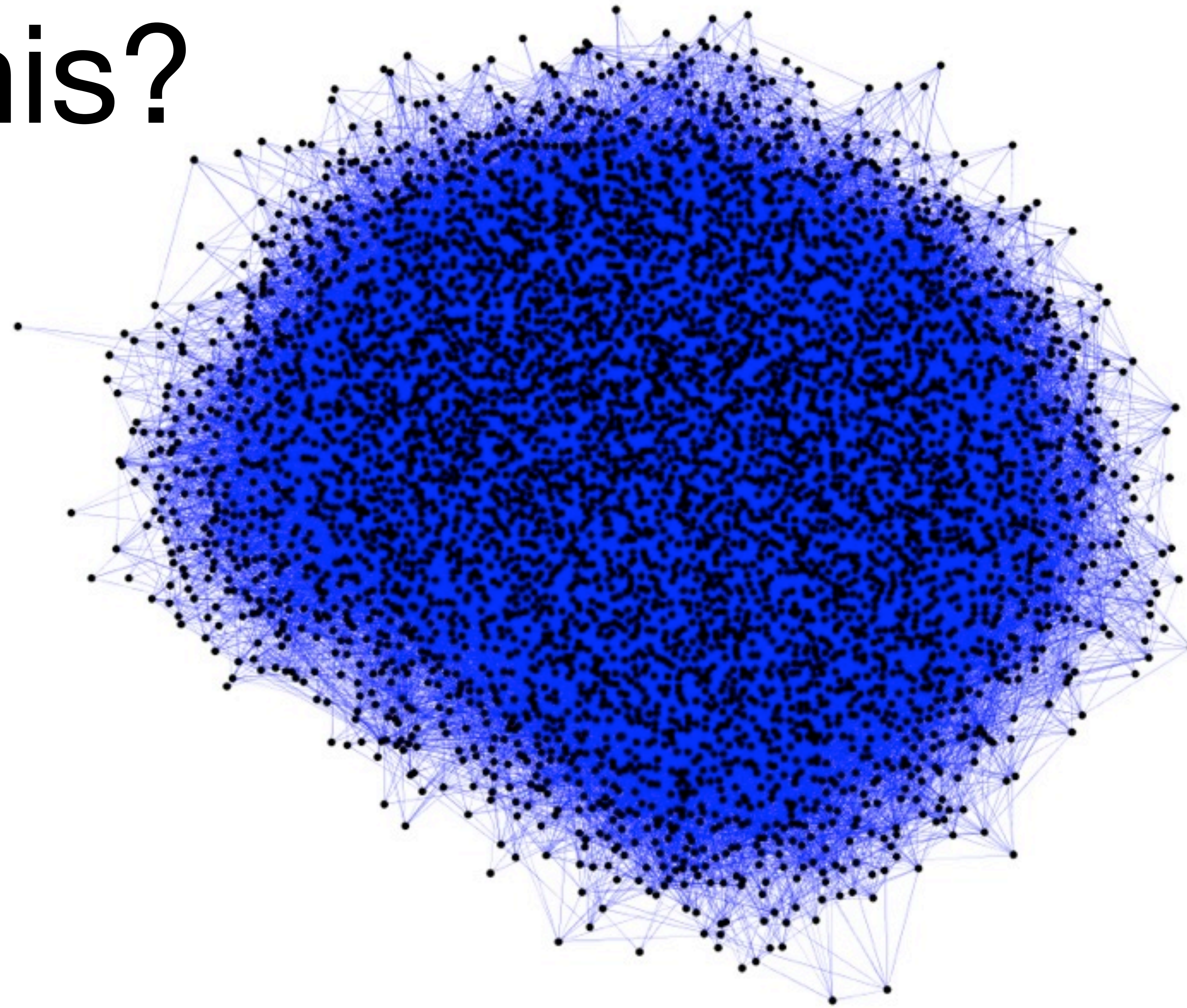


This is a modular network.

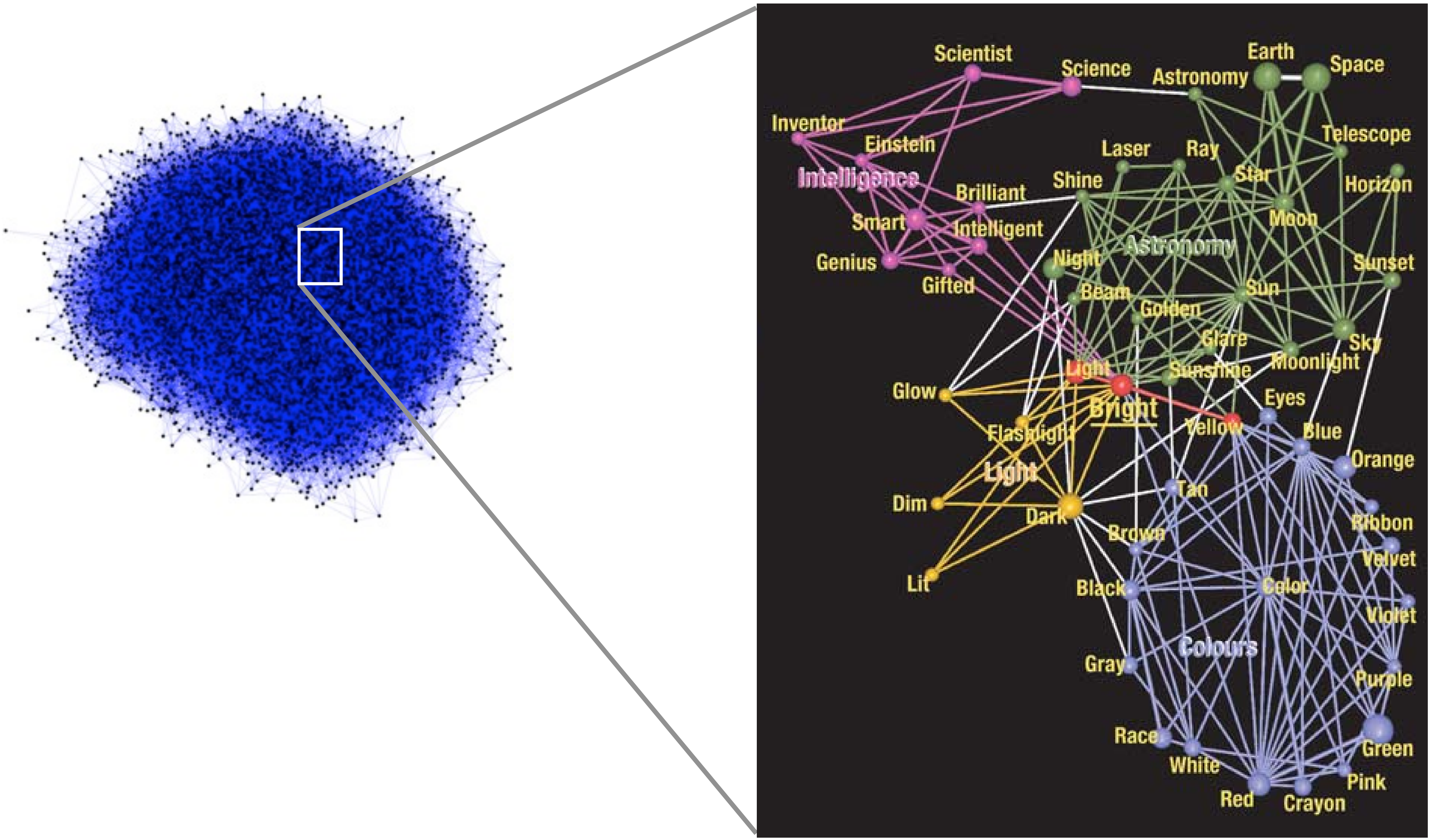
What is this?



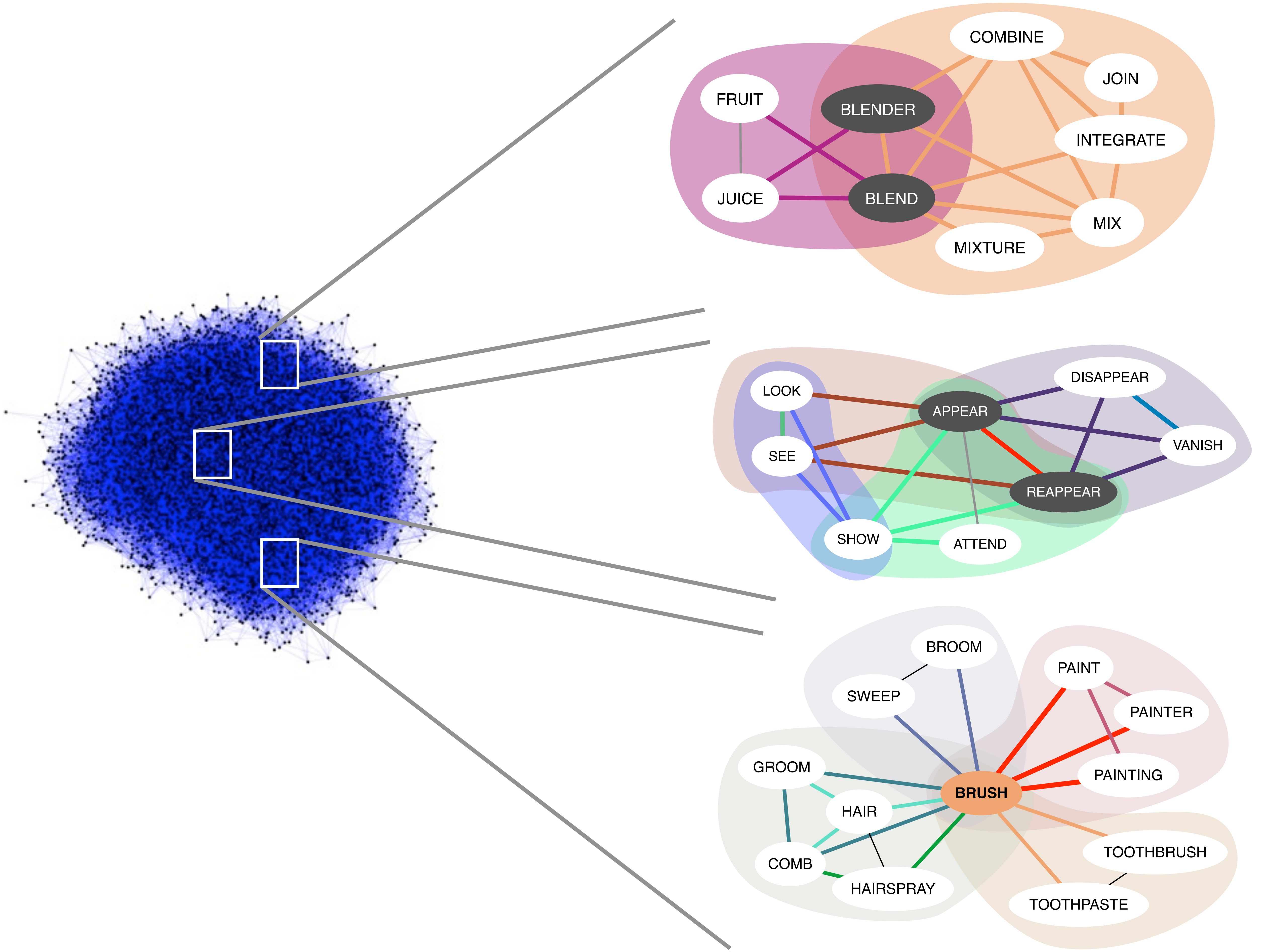
What the xxxx is this?



Word association network: Network of "commonly associated English words"



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


COMBINE
JOIN
FRUIT
BLENDER
INTEGRATE
JUICE
BLEND
MIX
MIXTURE
LOOK
DISAPPEAR
APPEAR
VANISH
SEE
REAPPEAR
SHOW
ATTEND
BROOM
PAINT
SWEEP
PAINTER
GROOM
PAINTING
BRUSH
HAIR
TOOTHBRUSH
COMB
HAIRSPRAY
TOOTHPASTE

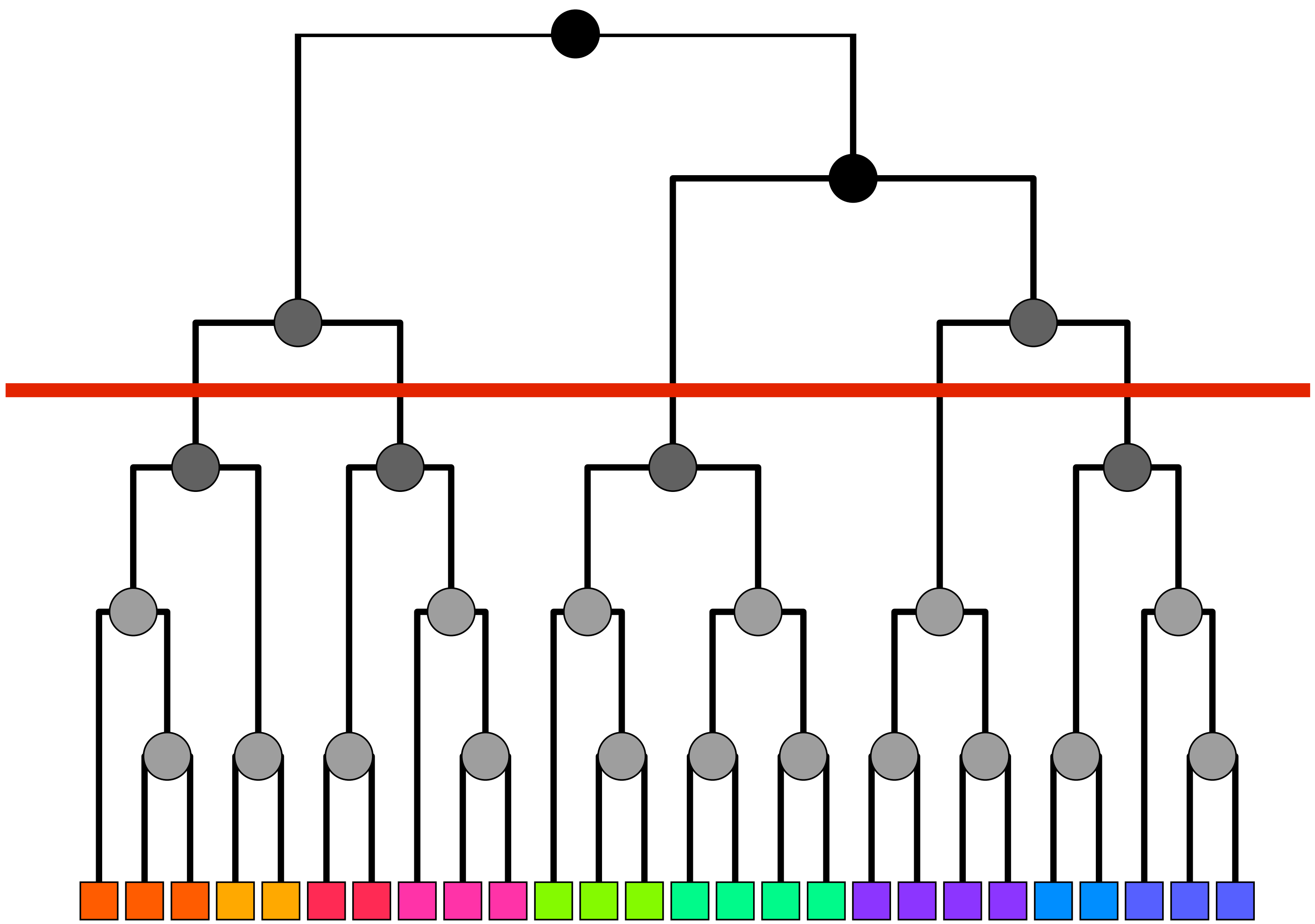
Here is the

PROBLEM.

Communities exist.

Hierarchical structure exists.

Hierarchy implies **disjoint communities.**



Hierarchical community structure

Hierarchy $\not\longrightarrow$ Communities

Hopeless?

Solution:

Use **LINKS**

Solution:

Use **LINKS**

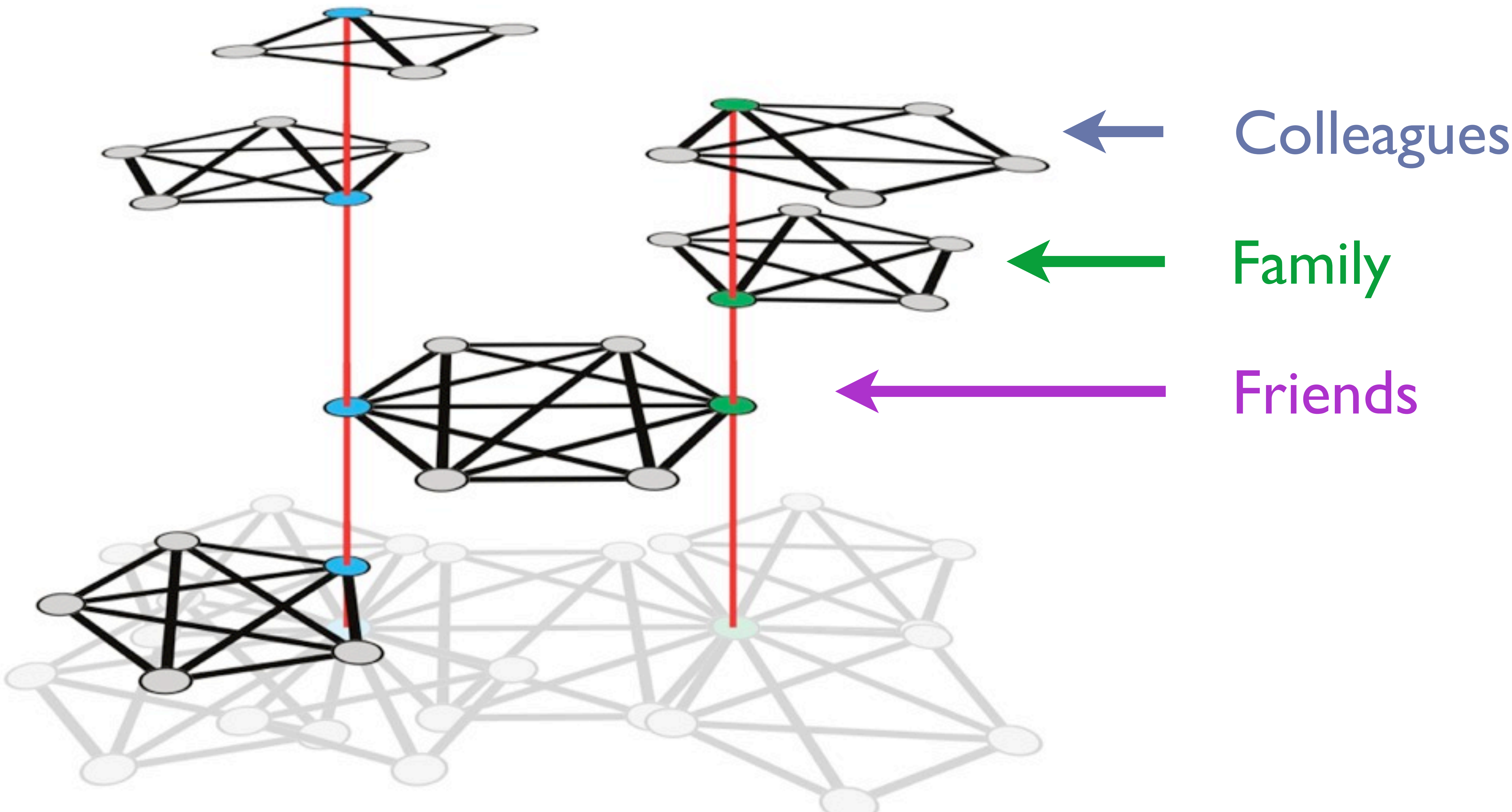
Solution:
Use **Links**

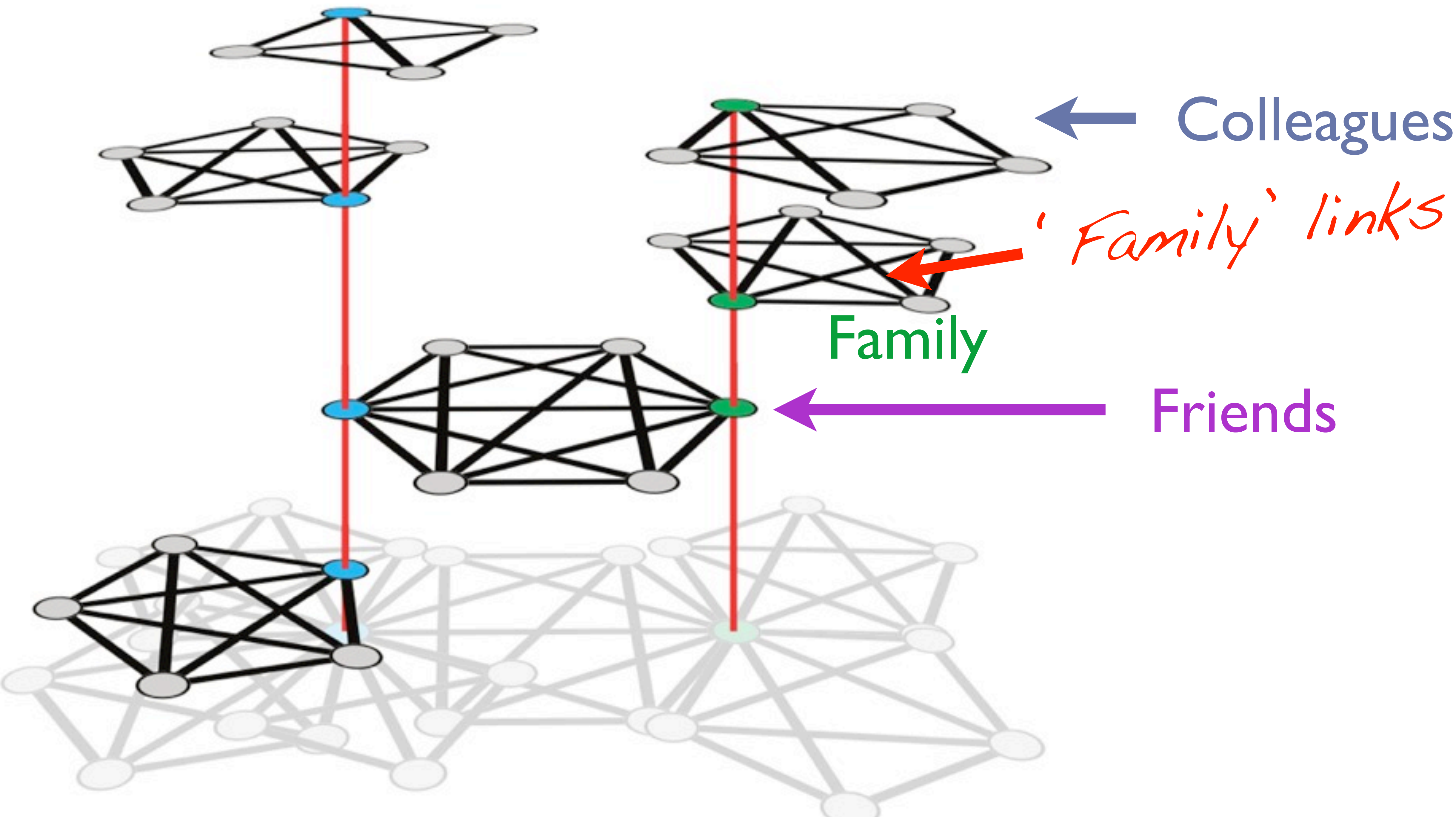
"a group of densely interconnected nodes"

Our solution:
Use **Links**

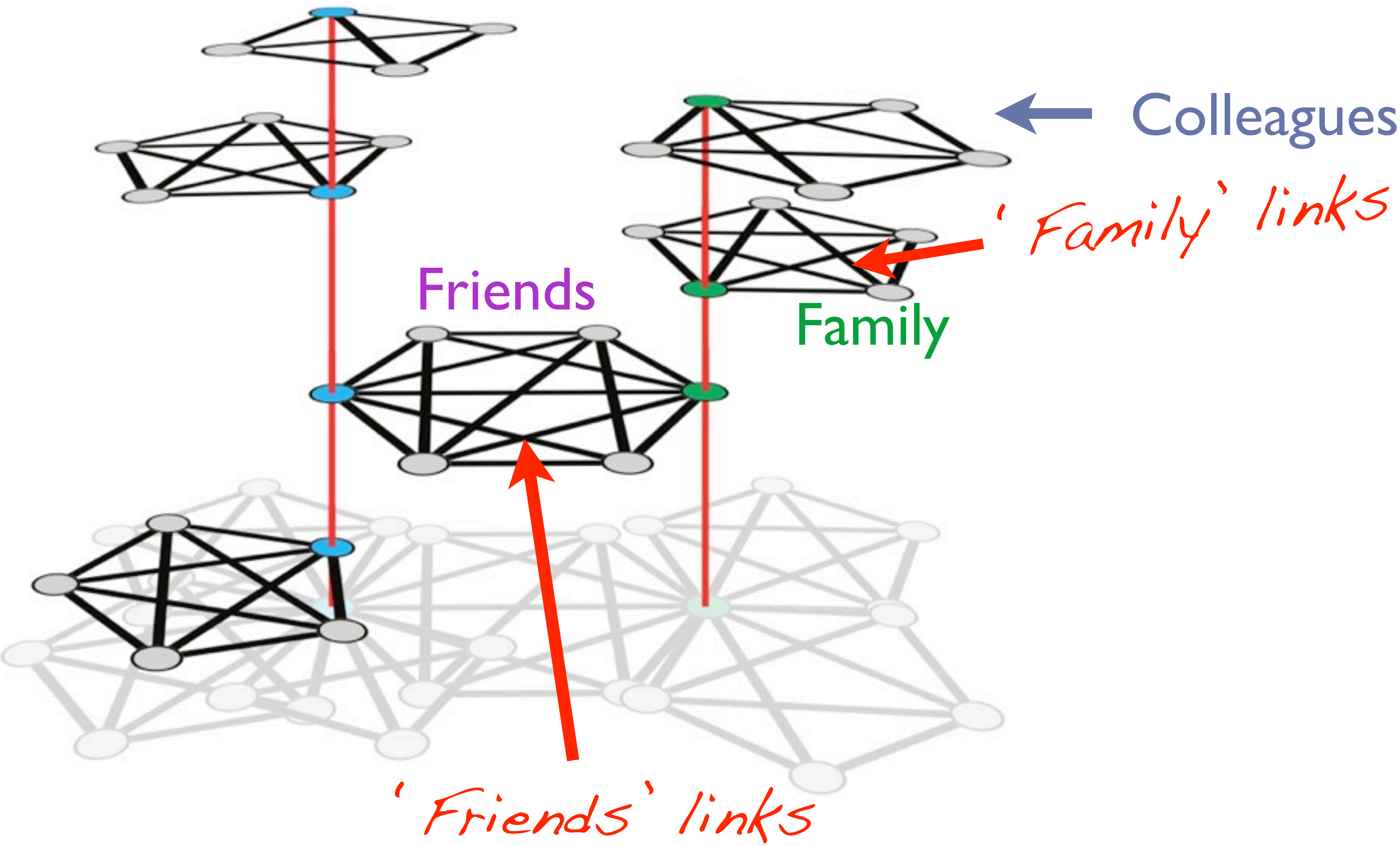
"a group of densely interconnected ~~nodes~~ **LINKS**"

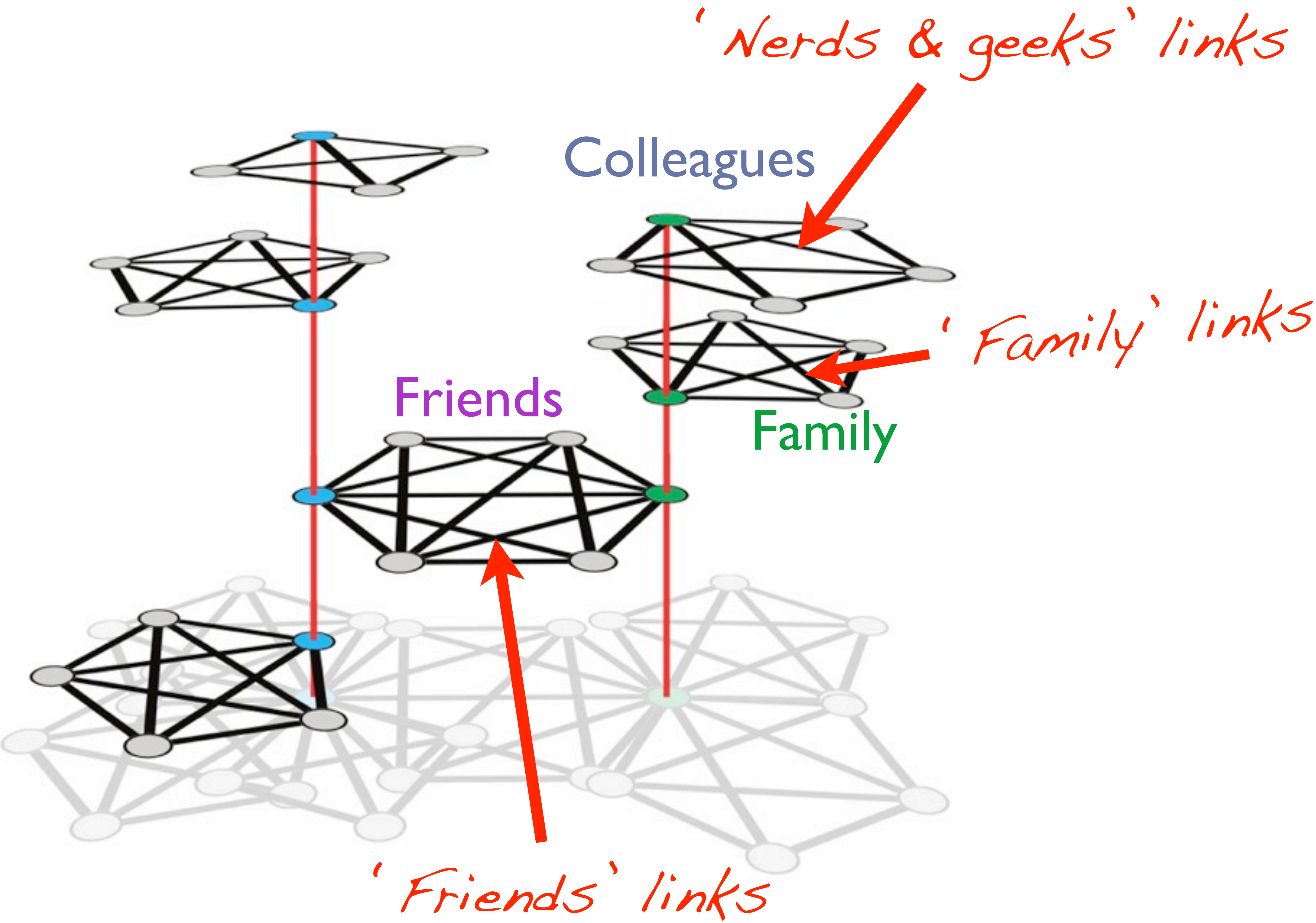
LINK communities


Colleagues
Family
Friends



Colleagues
'Family' links
Family
Friends

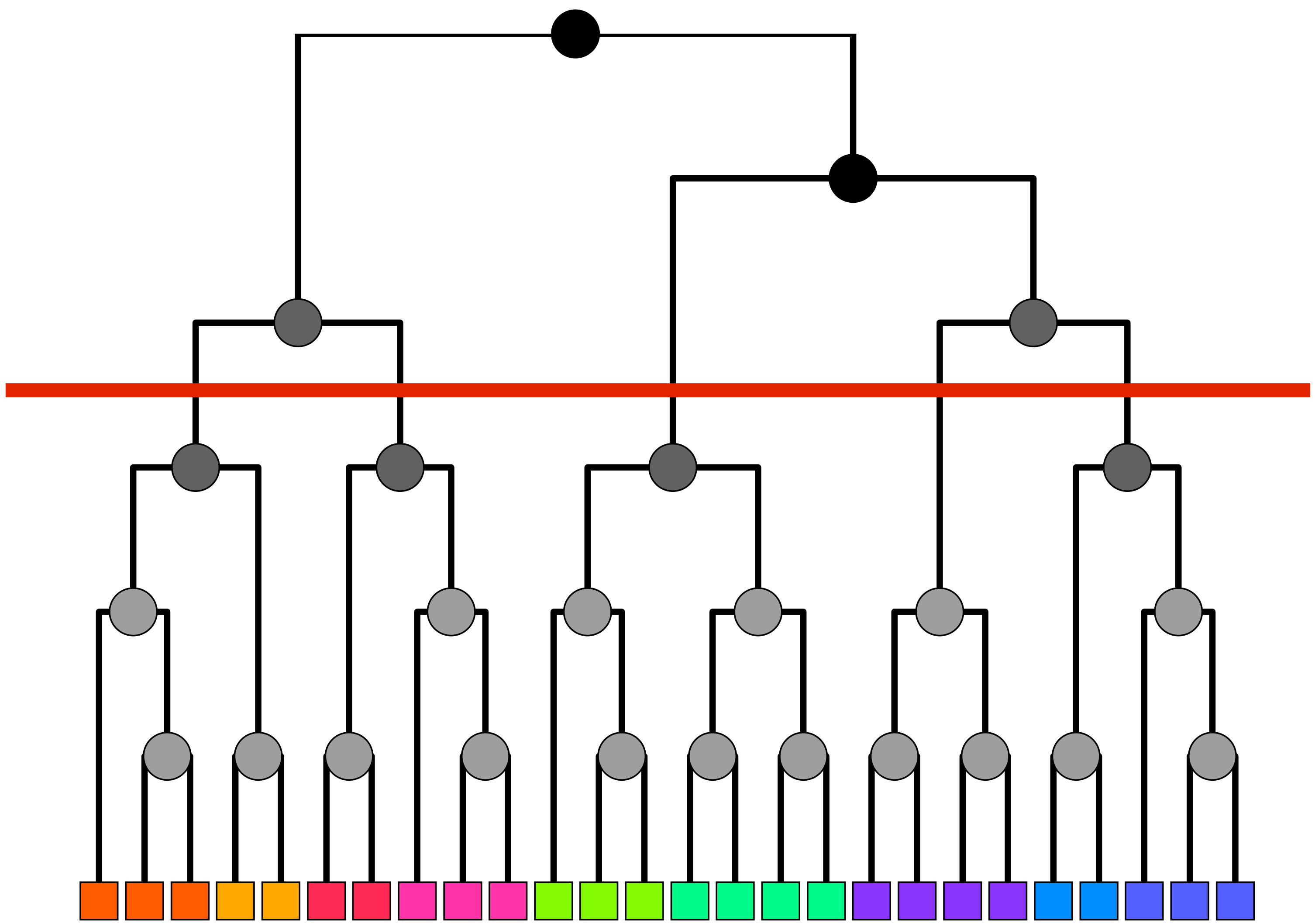

Colleagues
'Family' links
Friends
Family
'Friends' links


'Nerds & geeks' links
Colleagues
'Family' links
Friends
Family
'Friends' links

Nodes: multiple membership

Links: unique membership

Hierarchy implies **disjoint communities.**



Hierarchy $\cancel{\longrightarrow}$ Communities

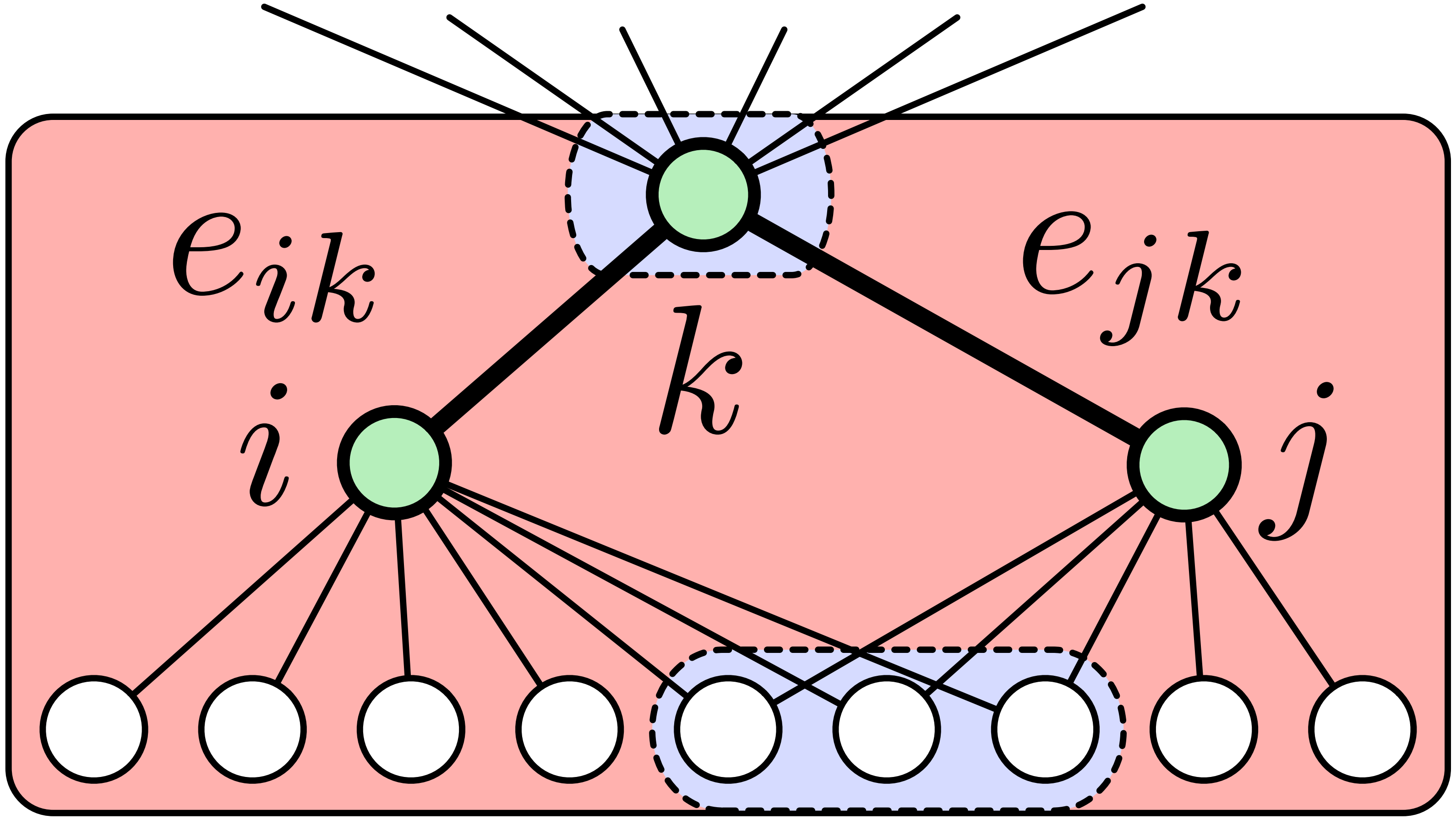
Hierarchy $\longrightarrow$ Communities

So, How?

Similarity between links

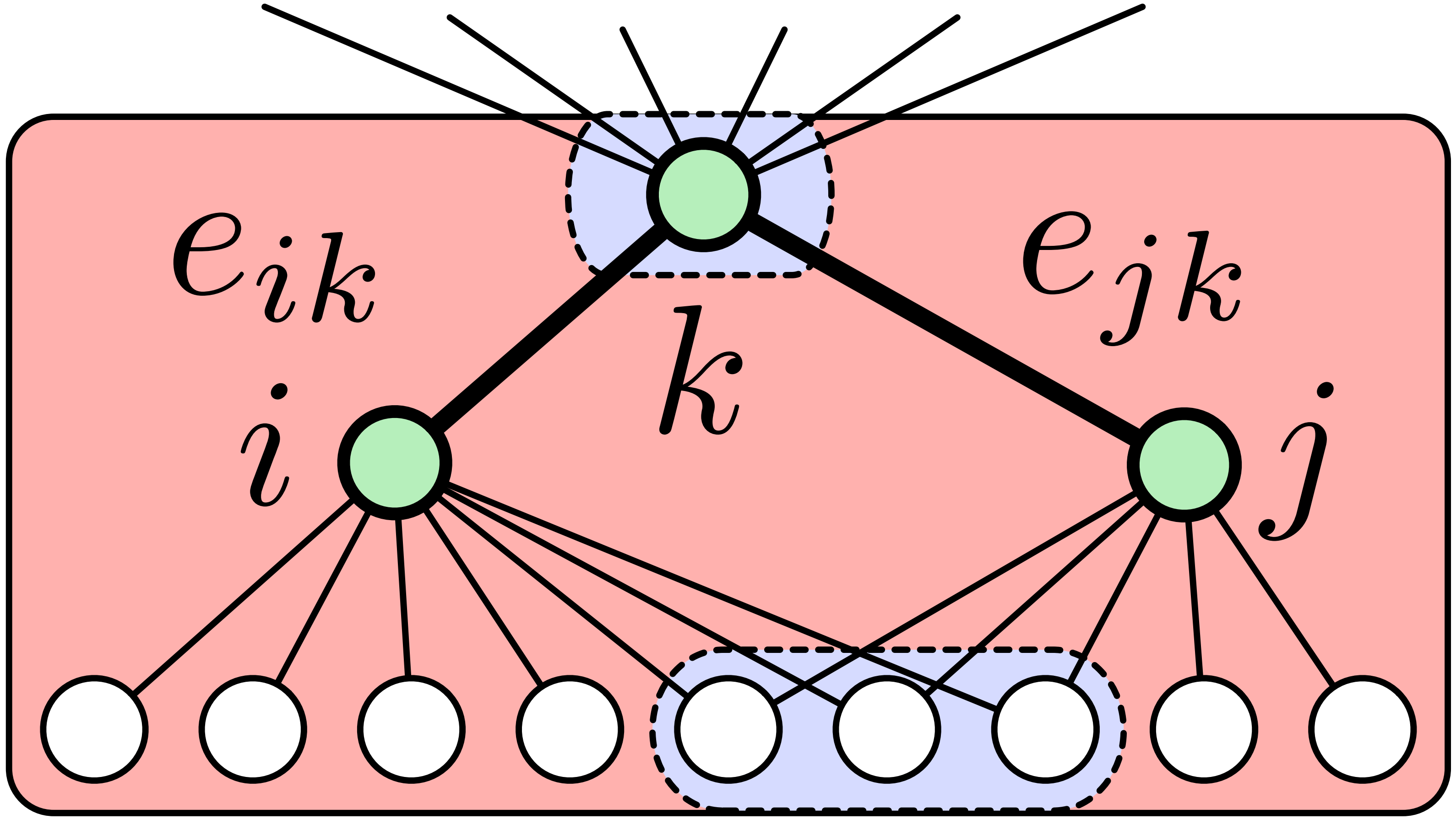
↓

Hierarchical Clustering



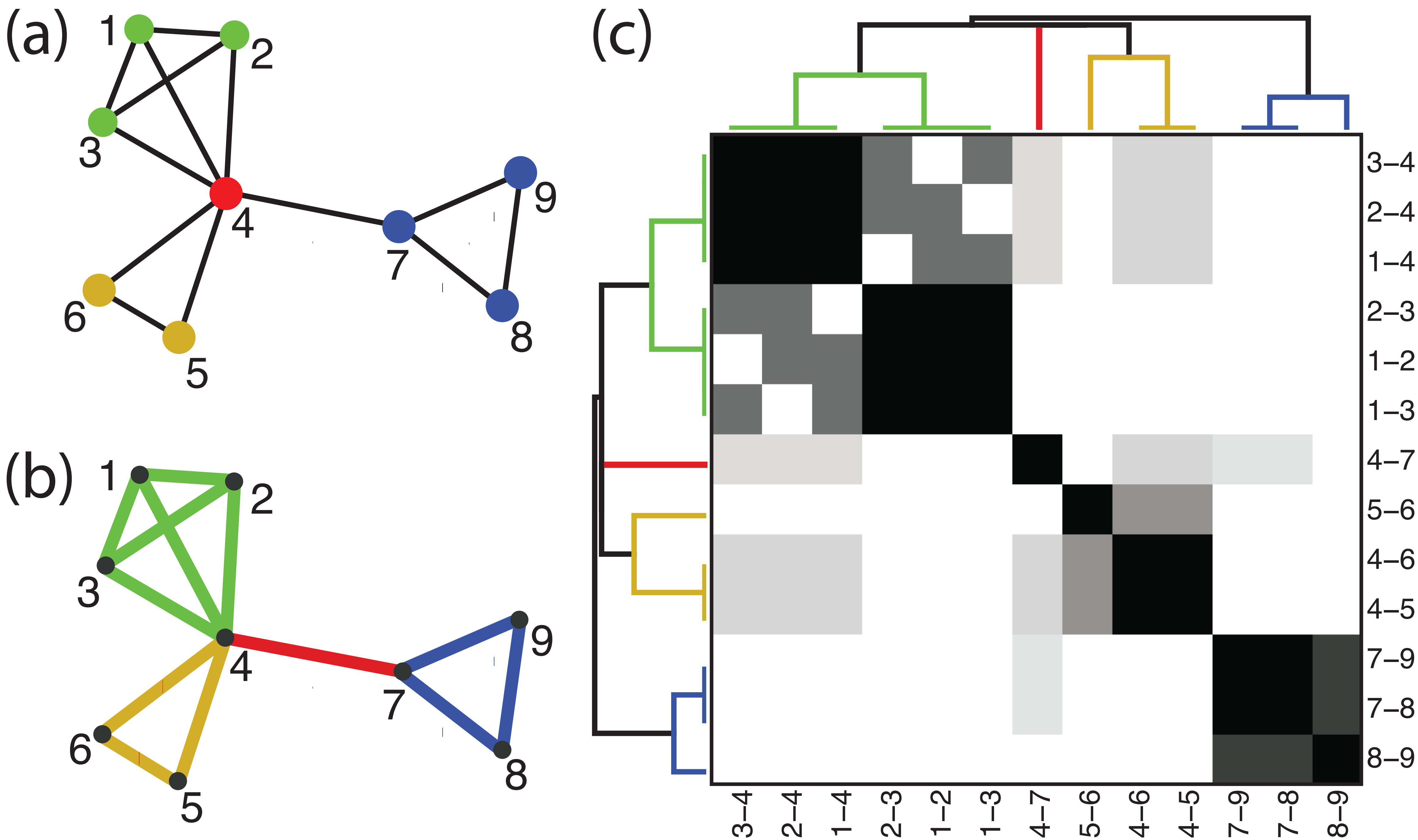
$$n_+(i) \equiv \{x \mid d(i,x) \leq 1\}$$

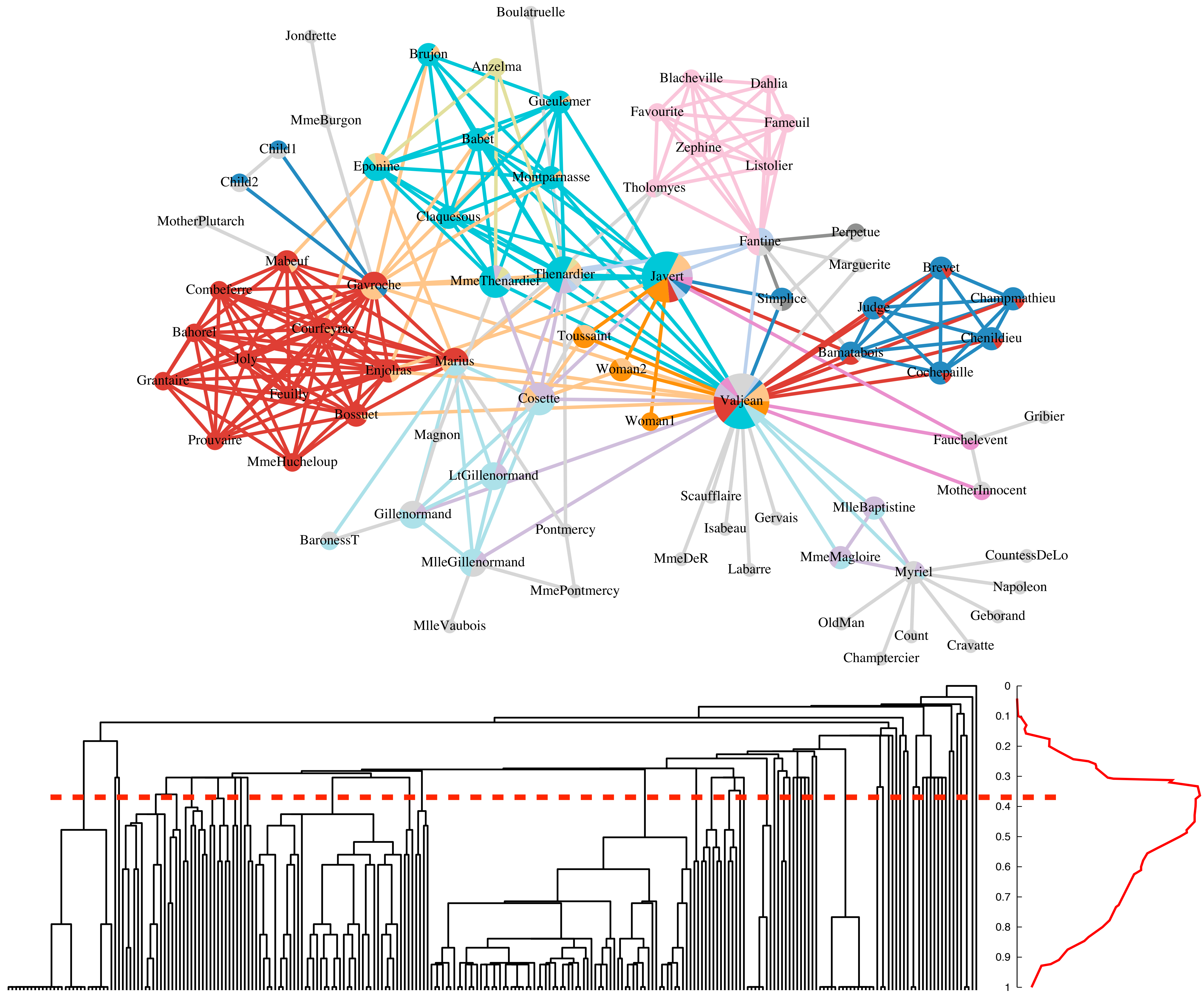
$$S(e_{ik}, e_{jk}) = \frac{|n_+(i) \cap n_+(j)|}{|n_+(i) \cup n_+(j)|}$$



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

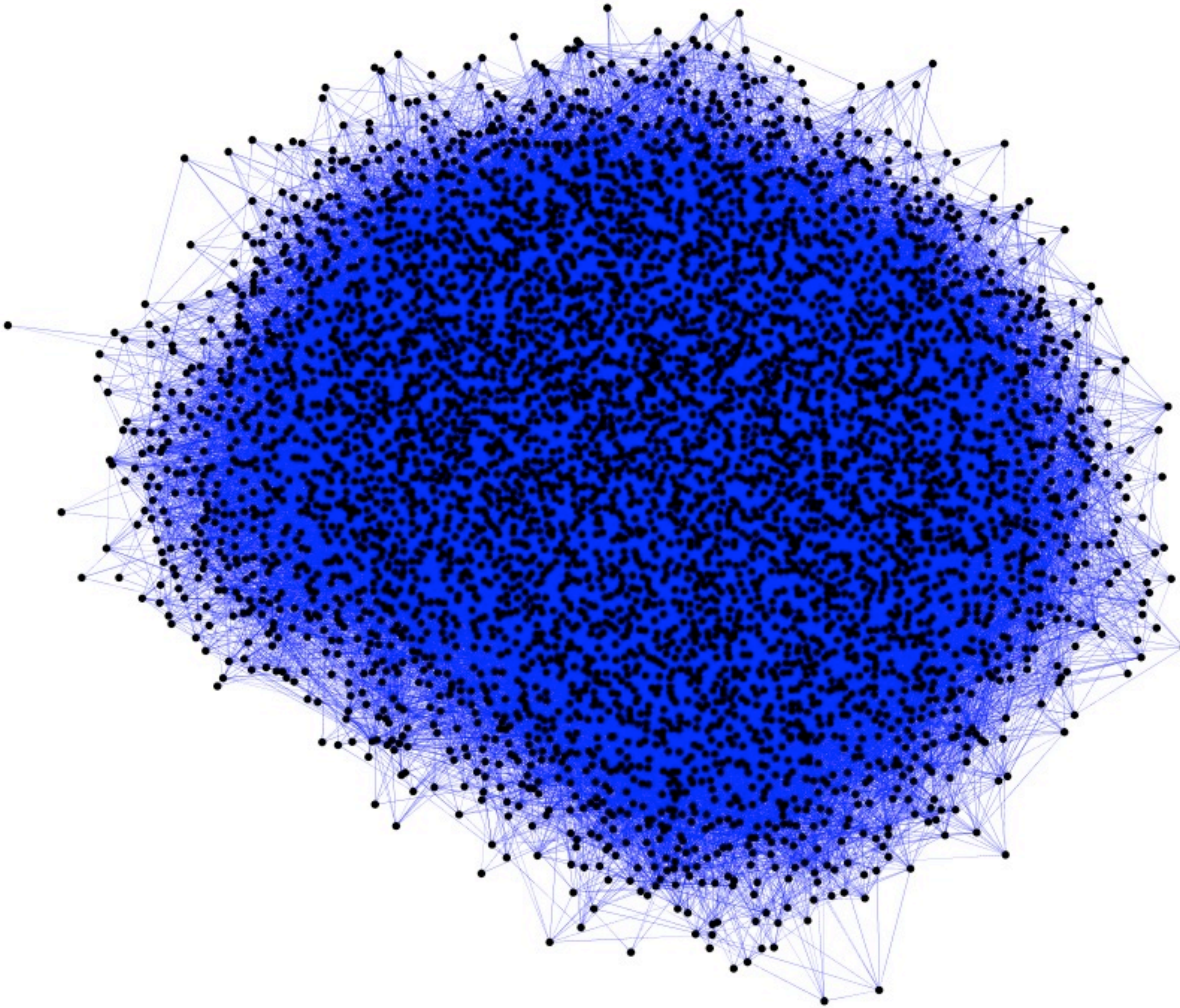
$$S(e_{ik}, e_{jk}) = \frac{|n_+(i) \cap n_+(j)|}{|n_+(i) \cup n_+(j)|} \quad \frac{4}{12}$$

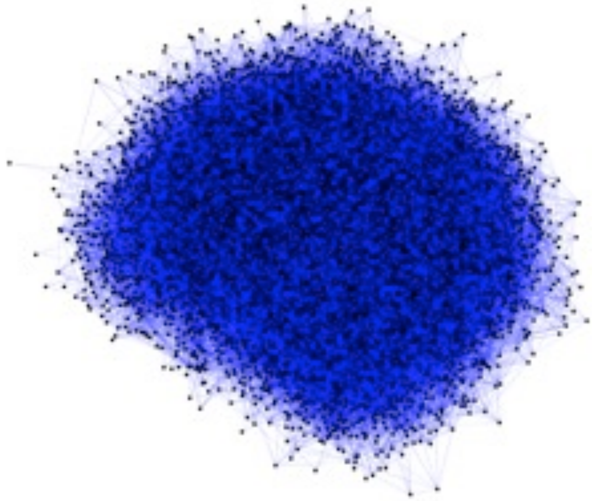

(a)
1
2
3
4
5
6
7
8
9
(b)
1
2
3
4
5
6
7
8
9
(c)
3–4
2–4
1–4
2–3
1–2
1–3
4–7
5–6
4–6
4–5
7–9
7–8
8–9

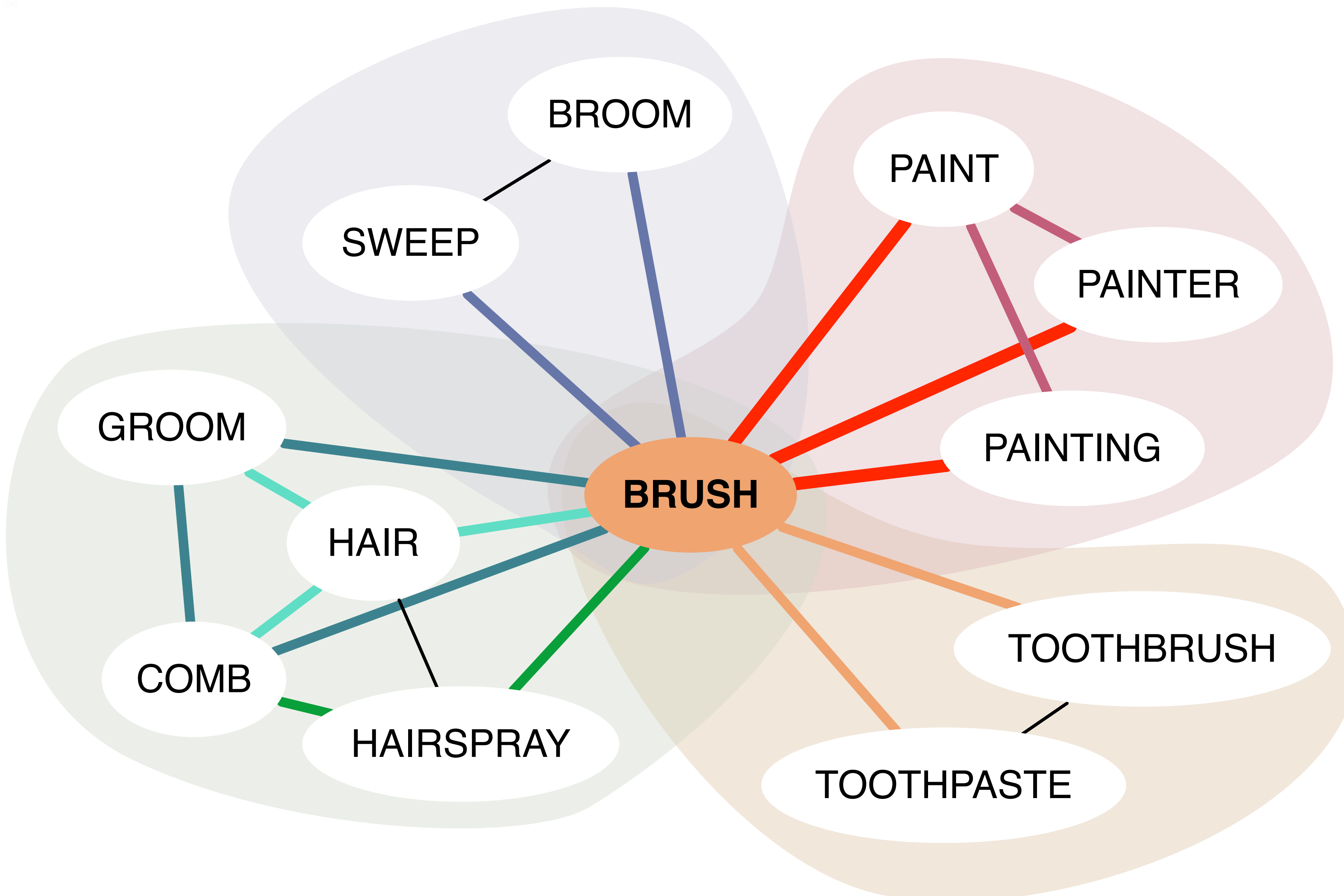


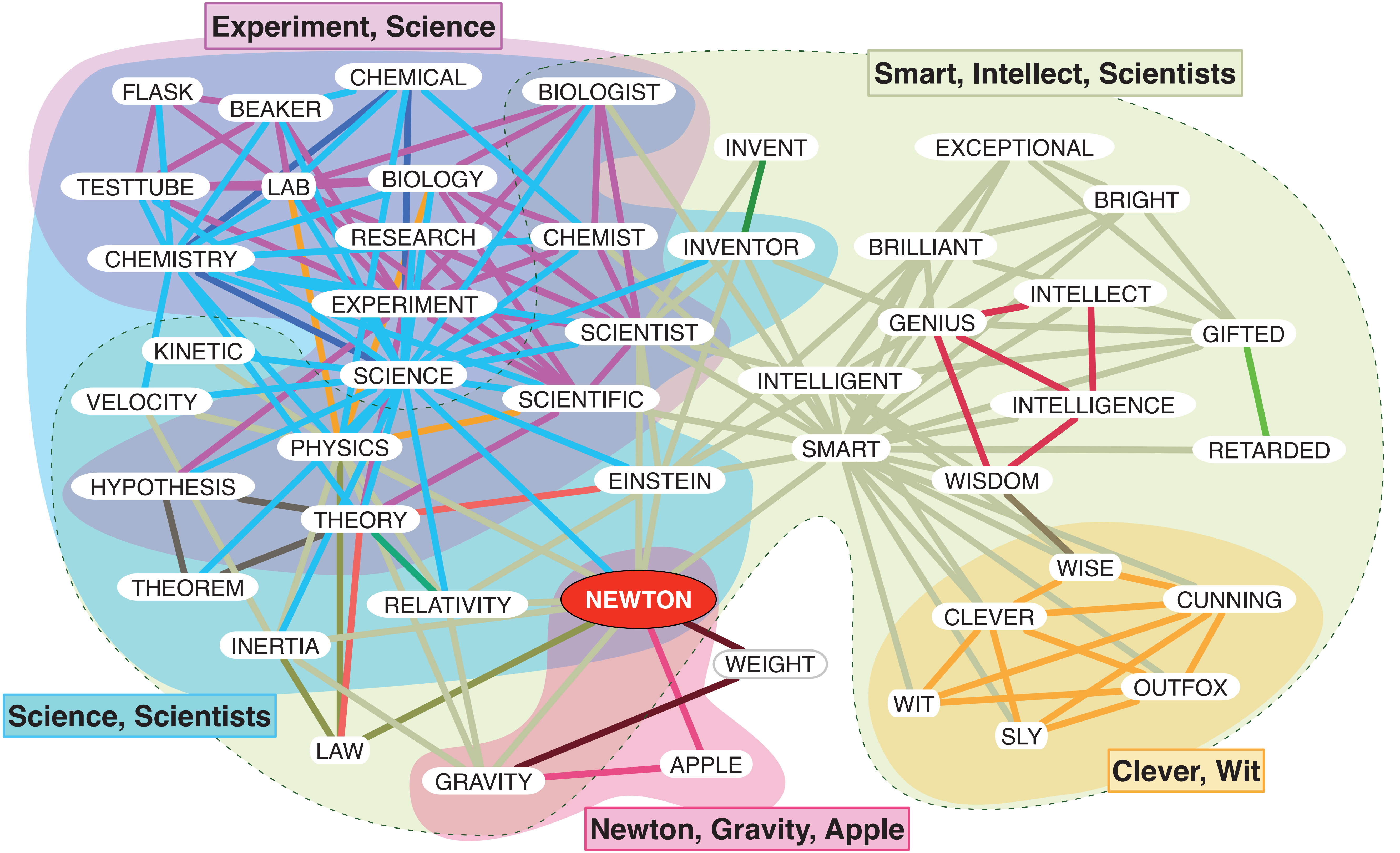
Boulatruelle
Jondrette
Brujon
Anzelma
Gueulemer
Blacheville
Dahlia
Favourite
Fameuil
MmeBurgon
Babet
Zephine
Child1
Eponine
Listolier
Montparnasse
Child2
Tholomyes
Claquesous
MotherPlutarch
Fantine
Perpetue
Mabeuf
Marguerite
Brevet
Gavroche
MmeThenardier
Thenardier
Javert
Combeferre
Simplice
Champmathieu
Judge
Courfeyrac
Toussaint
Bahorel
Chenildieu
Bamatabois
Joly
Woman2
Marius
Enjolras
Cochepaille
Grantaire
Feuilly
Cosette
Valjean
Bossuet
Woman1
Gribier
Magnon
Prouvaire
Fauchelevent
MmeHucheloup
LtGillenormand
MotherInnocent
Scaufflaire
MlleBaptistine
Gillenormand
Gervais
Isabeau
Pontmercy
BaronessT
CountessDeLo
MmeDeR
MmeMagloire
MlleGillenormand
Labarre
Myriel
Napoleon
MmePontmercy
Geborand
OldMan
MlleVaubois
Count
Cravatte
Champtercier
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Does it really work?

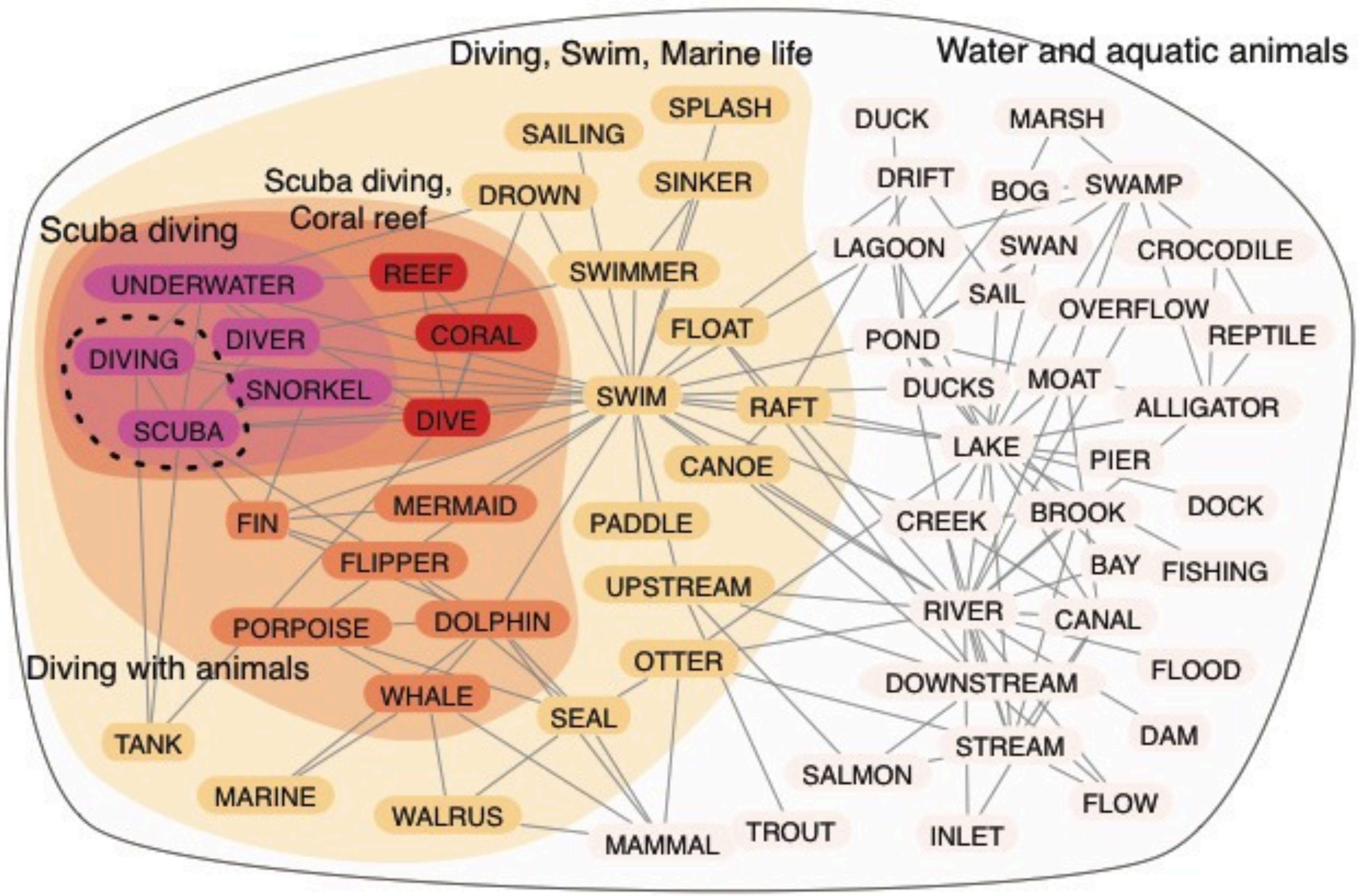





BROOM
SWEEP
PAINT
PAINTER
PAINTING
GROOM
HAIR
BRUSH
COMB
HAIRSPRAY
TOOTHBRUSH
TOOTHPASTE

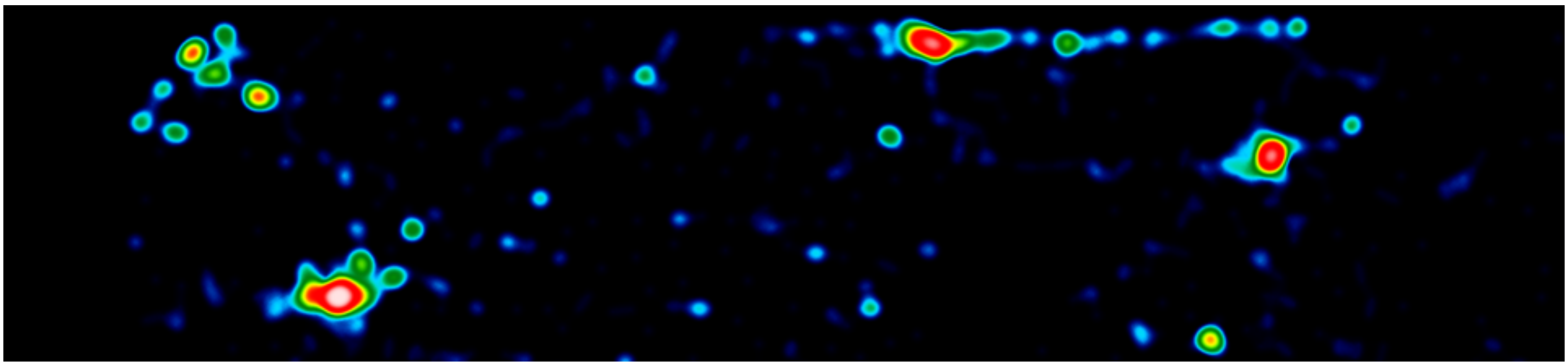


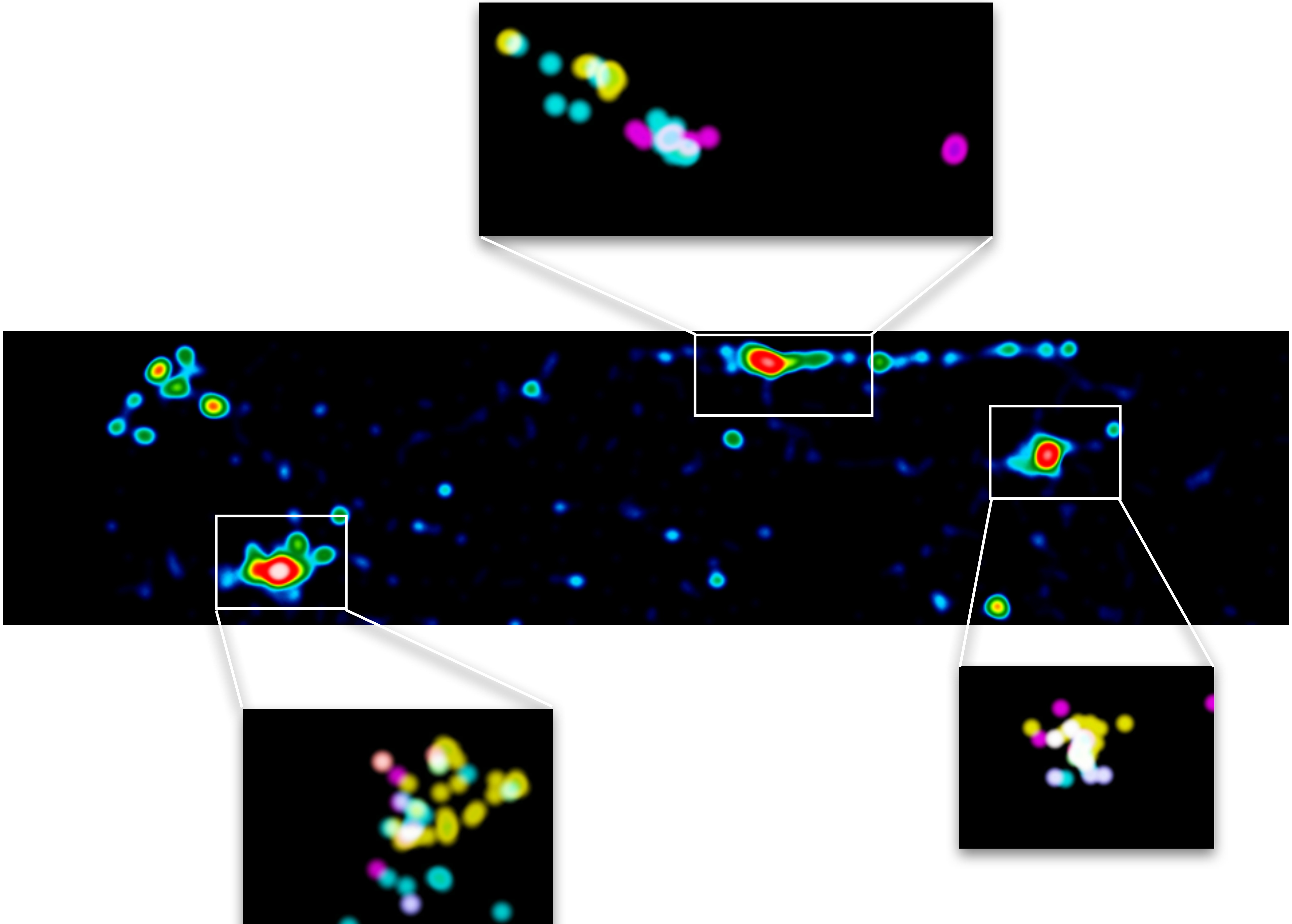
Experiment, Science
Smart, Intellect, Scientists
Science, Scientists
Newton, Gravity, Apple
Clever, Wit
FLASK
BEAKER
CHEMICAL
BIOLOGIST
TESTTUBE
LAB
BIOLOGY
CHEMISTRY
RESEARCH
CHEMIST
EXPERIMENT
SCIENTIST
KINETIC
SCIENCE
VELOCITY
SCIENTIFIC
PHYSICS
HYPOTHESIS
THEORY
EINSTEIN
THEOREM
RELATIVITY
NEWTON
INERTIA
WEIGHT
LAW
GRAVITY
APPLE
INVENT
INVENTOR
EXCEPTIONAL
BRIGHT
BRILLIANT
INTELLECT
GENIUS
GIFTED
INTELLIGENT
INTELLIGENCE
SMART
RETARDED
WISDOM
WISE
CUNNING
CLEVER
OUTFOX
WIT
SLY

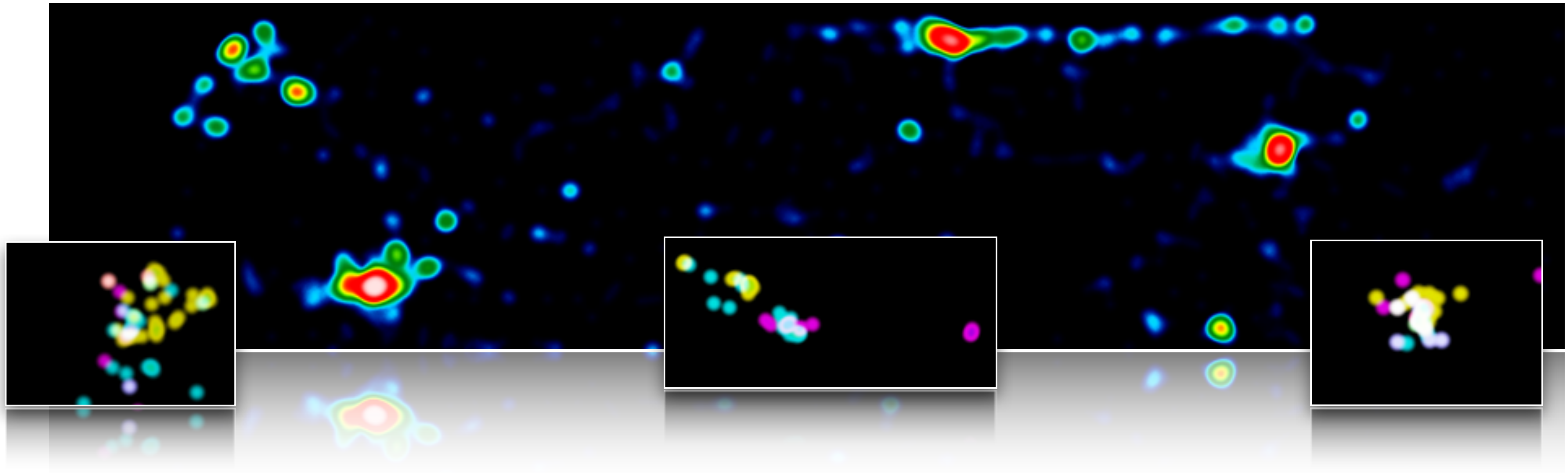

Diving, Swim, Marine life
Water and aquatic animals
Scuba diving
Scuba diving,
Coral reef
Diving with animals
SPLASH
SAILING
SINKER
DROWN
SWIMMER
UNDERWATER
REEF
CORAL
DIVER
DIVING
SNORKEL
SCUBA
DIVE
FLOAT
SWIM
RAFT
CANOE
MERMAID
FIN
PADDLE
FLIPPER
UPSTREAM
PORPOISE
DOLPHIN
OTTER
WHALE
SEAL
TANK
MARINE
WALRUS
MAMMAL
TROUT
DUCK
MARSH
DRIFT
BOG
SWAMP
LAGOON
SWAN
CROCODILE
SAIL
OVERFLOW
POND
REPTILE
DUCKS
MOAT
ALLIGATOR
LAKE
PIER
DOCK
CREEK
BROOK
BAY
FISHING
RIVER
CANAL
FLOOD
DOWNSTREAM
DAM
STREAM
SALMON
FLOW
INLET

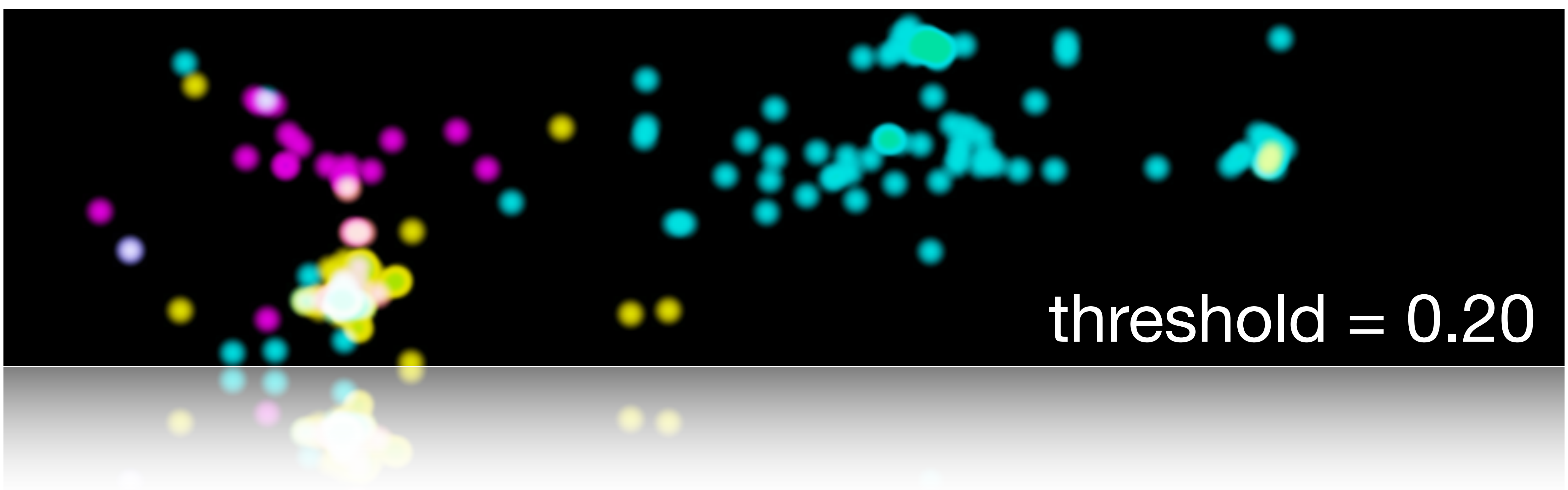
~600k nodes

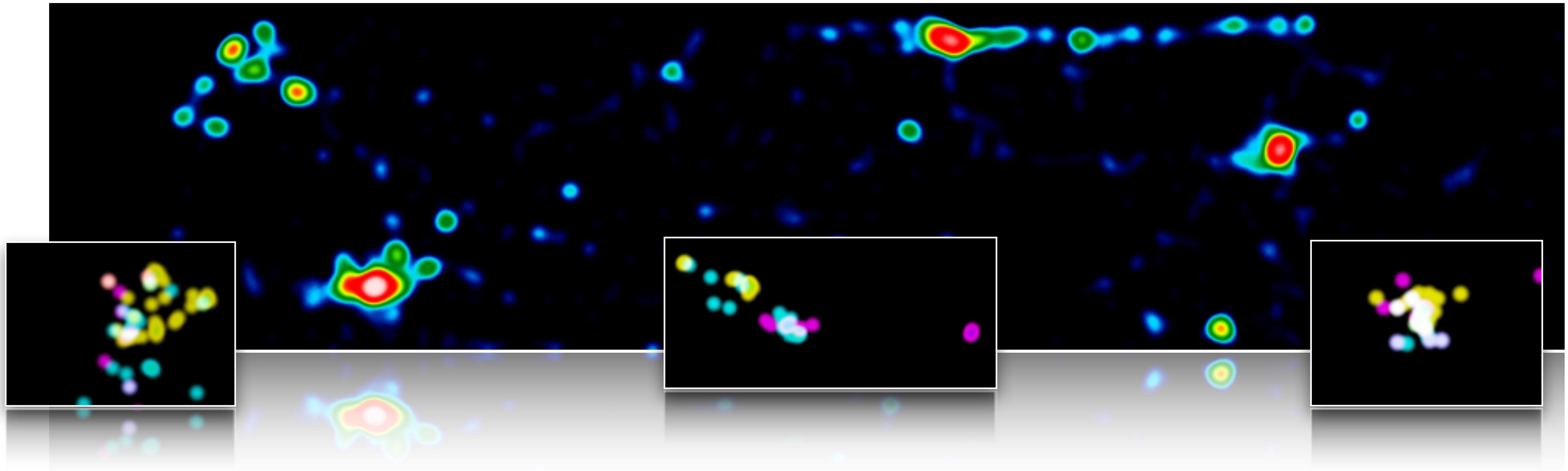
~3M edges

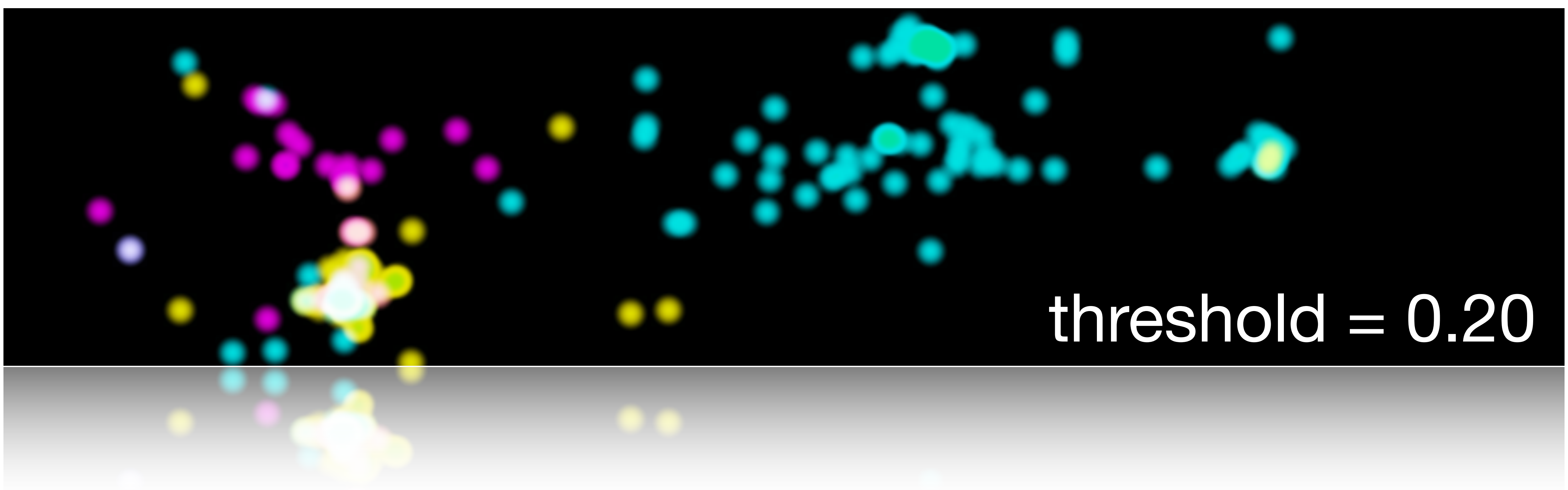


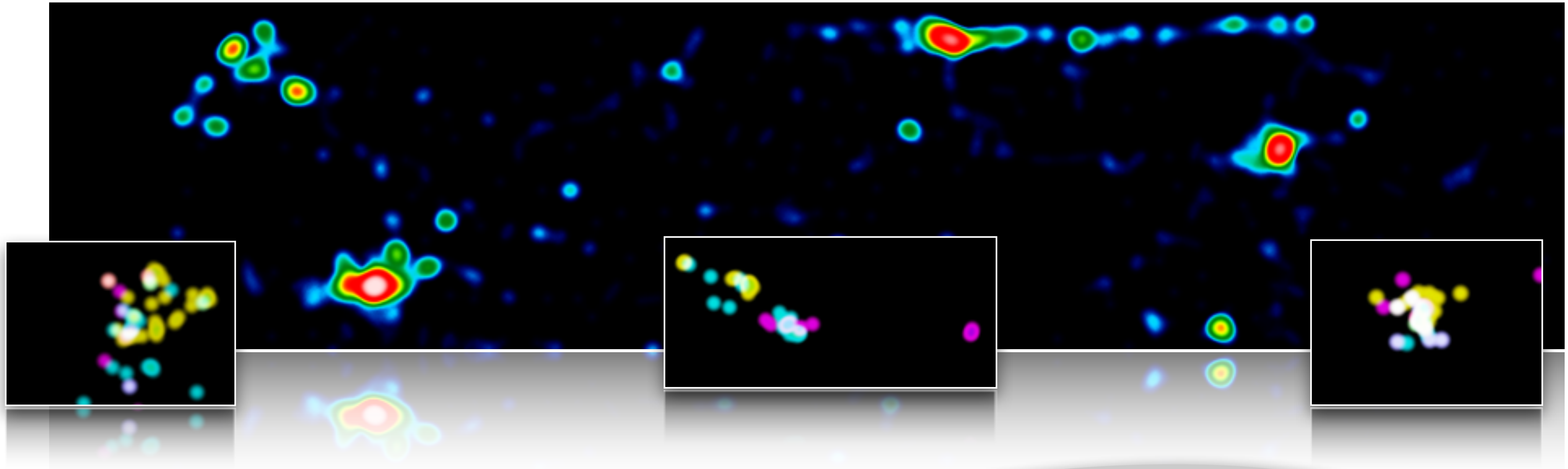


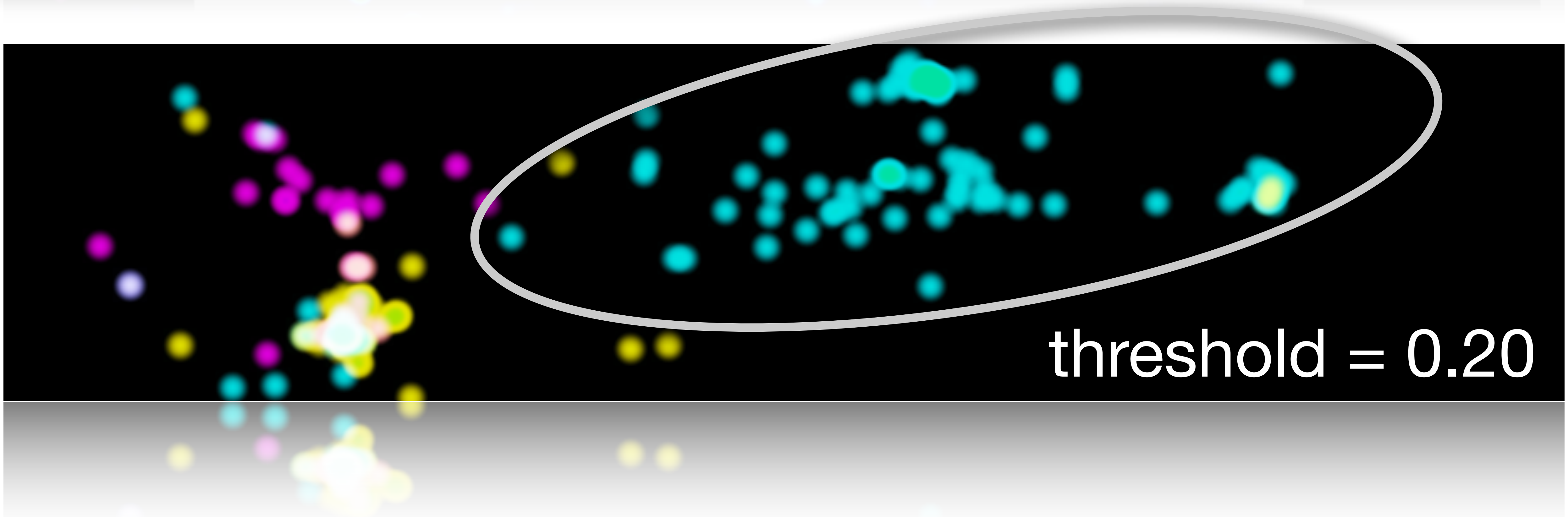


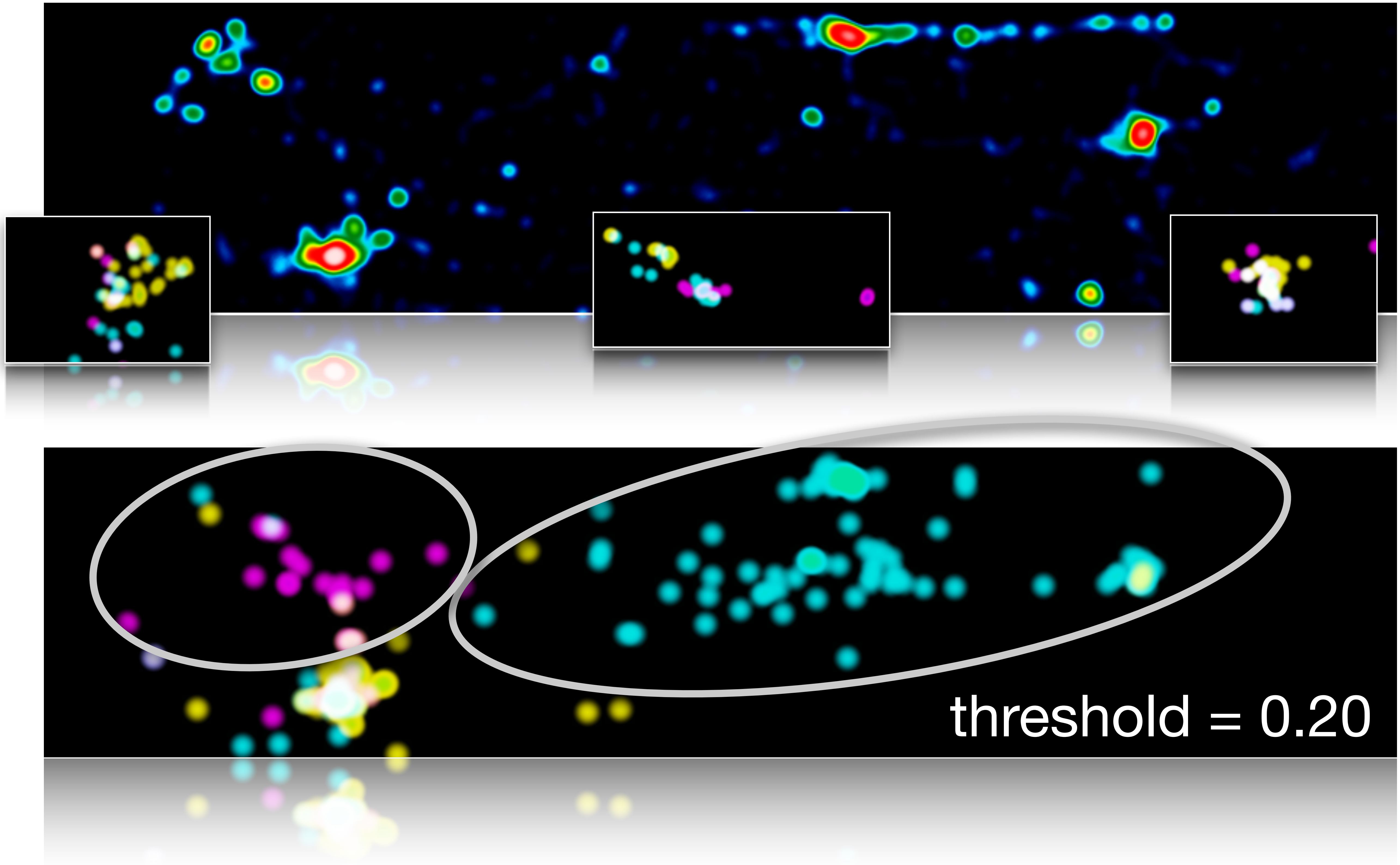

threshold = 0.20

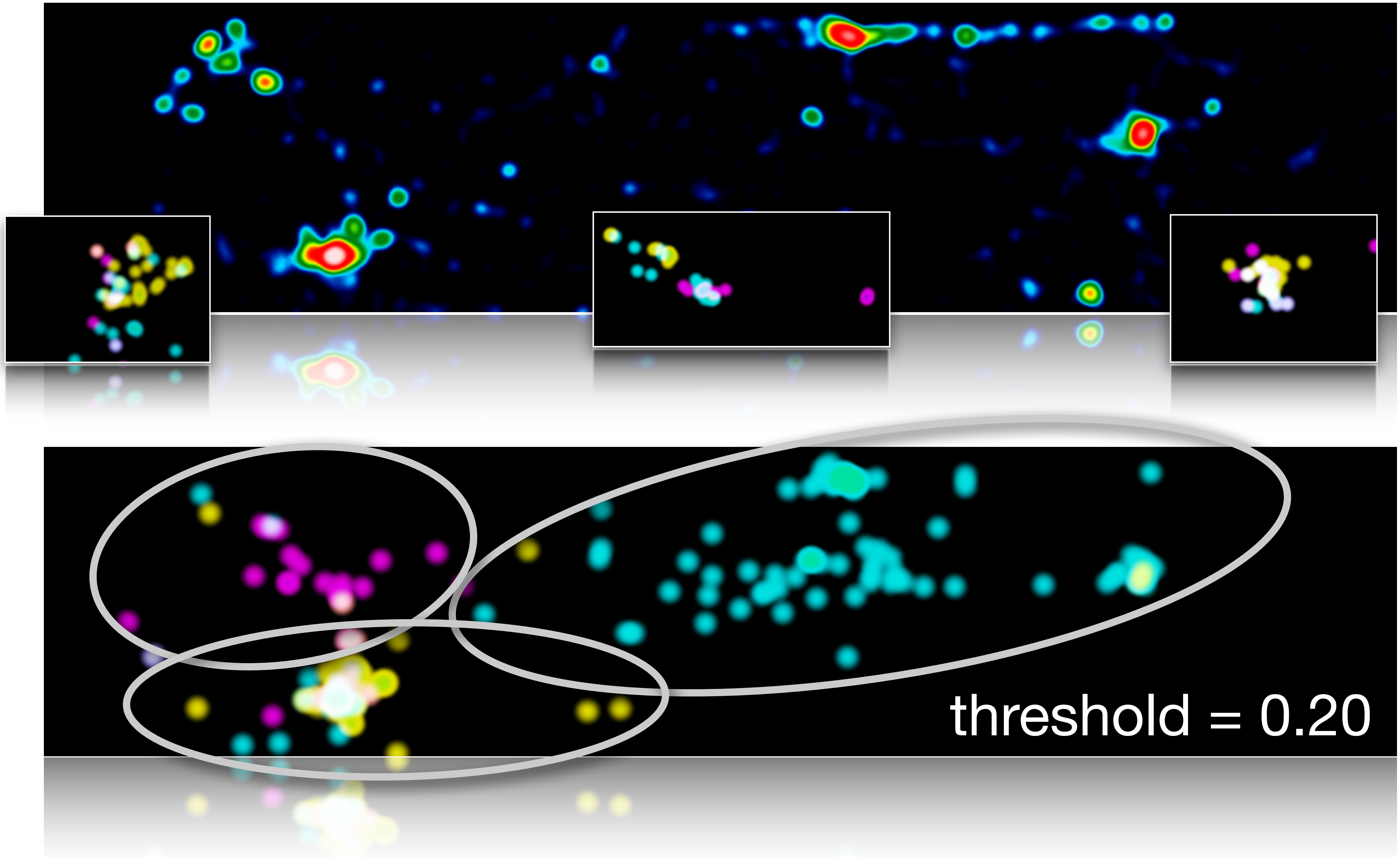


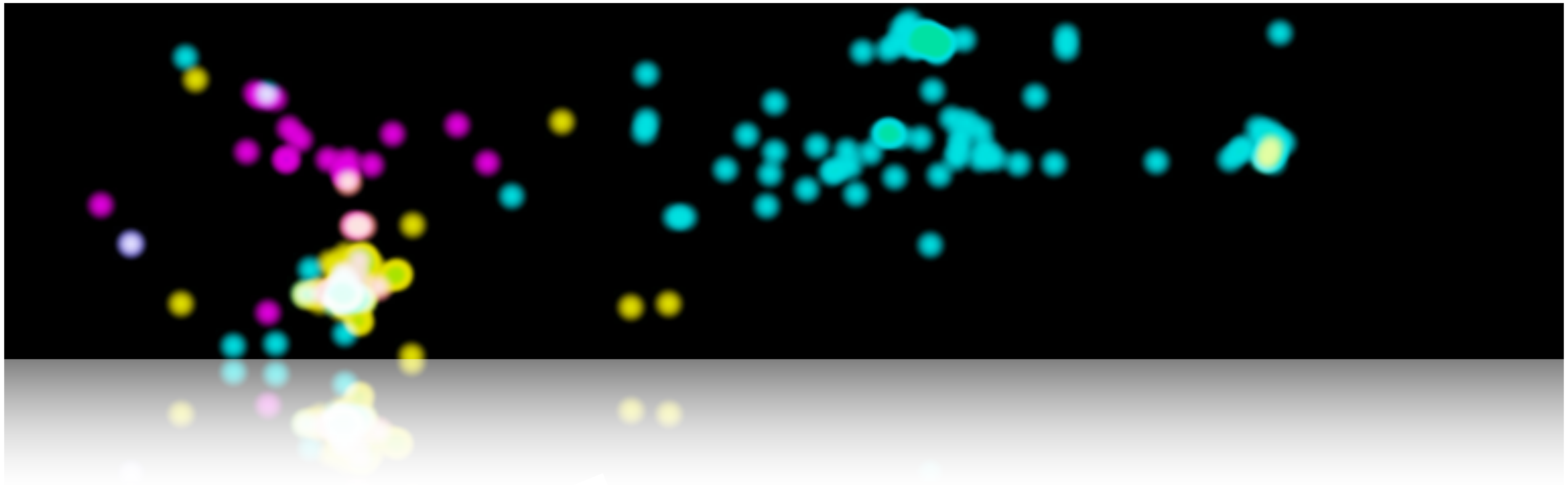

threshold = 0.20

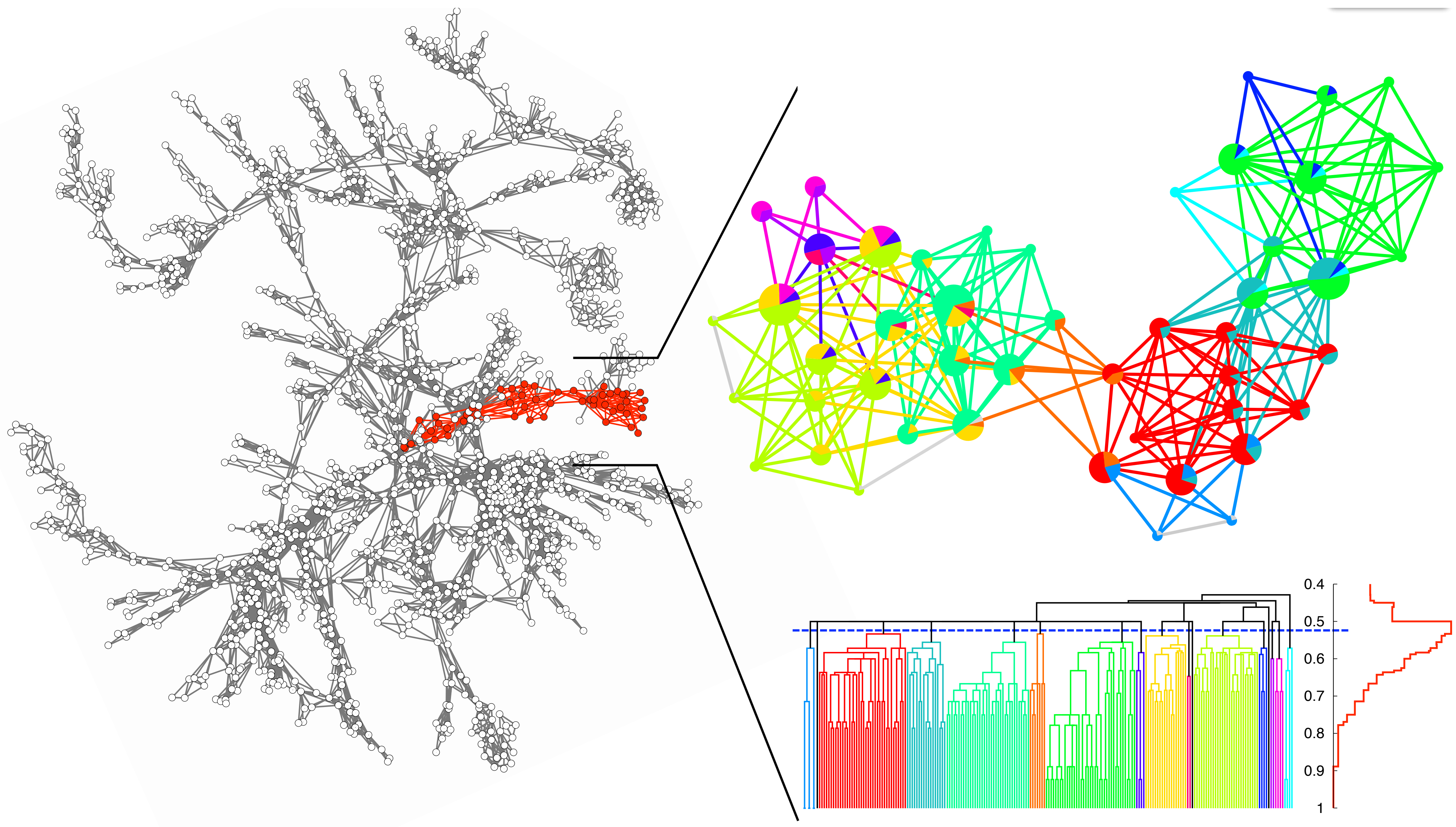



threshold = 0.20


threshold = 0.20


threshold = 0.20




0.4
0.5
0.6
0.7
0.8
0.9
1

Summary

- Networks matter.
- Particularly in the age of big data and social networks.
- Many interesting problems waiting for you!

Resources

- http://yongyeol.com/courses/2012S-I590/
- http://yongyeol.com/w/index.php?title=Network_science
- http://en.wikipedia.org/wiki/Network_science

- http://yongyeol.com
- yyahn@indiana.edu
- @yy