Communities and Diffusion of Culture

Yong-Yeol "YY" Ahn
@yy

SCHOOL OF INFORMATICS AND COMPUTING
INDIANA UNIVERSITY
Bloomington

CNetS Center for Complex Networks and Systems Research

INDIANA UNIVERSITY
NETWORK INSTITUTE

NAZIS SUCK

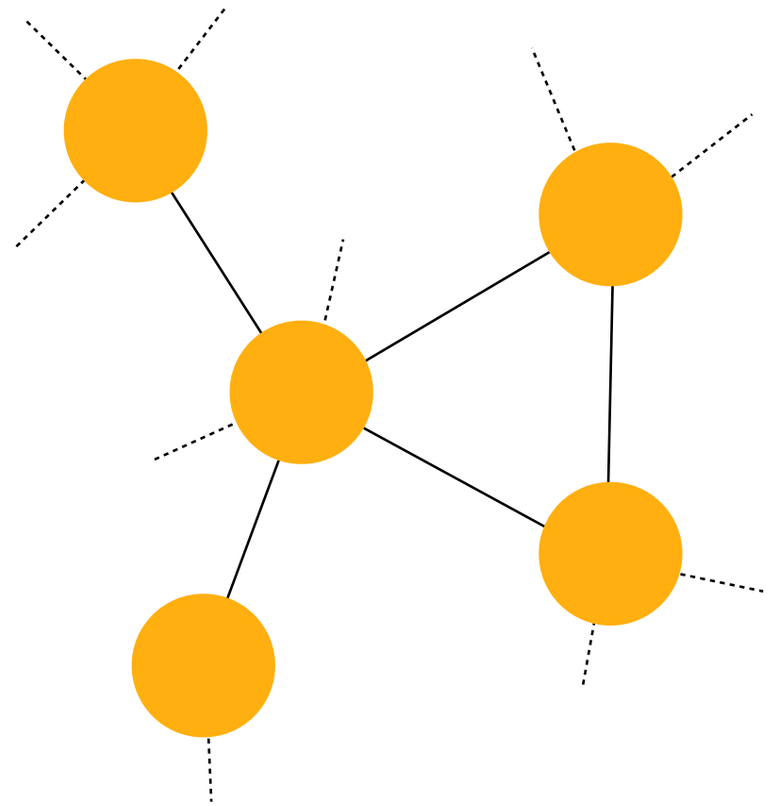
at giving directions

Cultural problems

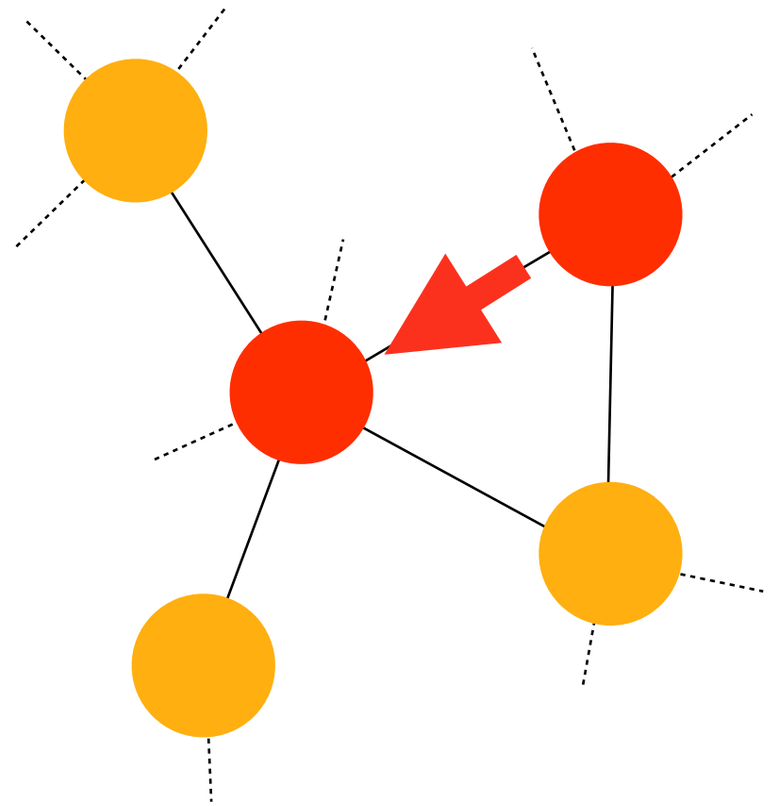
Cultural problems

How **ideas and information** spread?

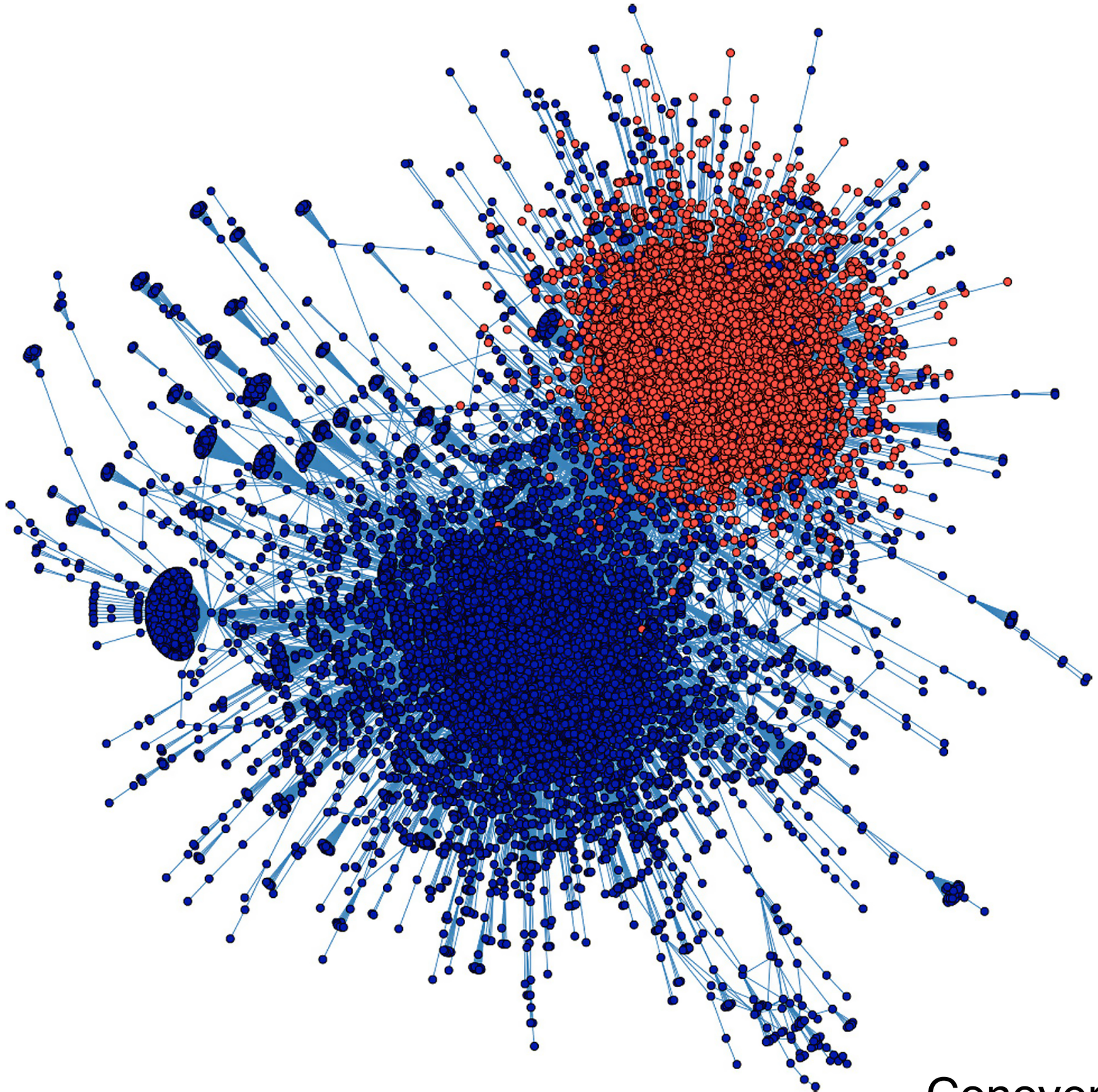
‘How one **influences**
the others’



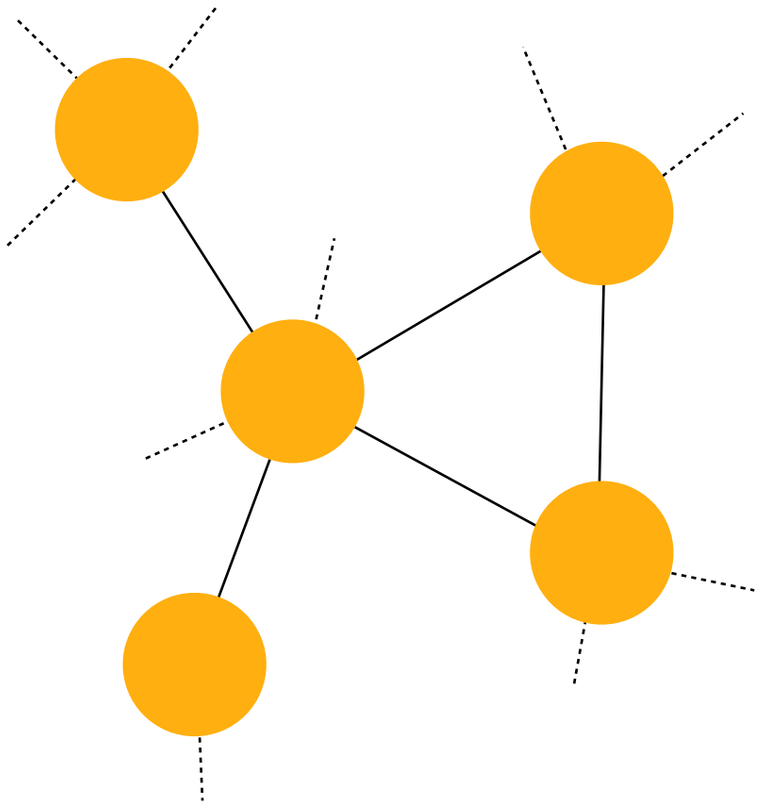
Network
(Social Structure)



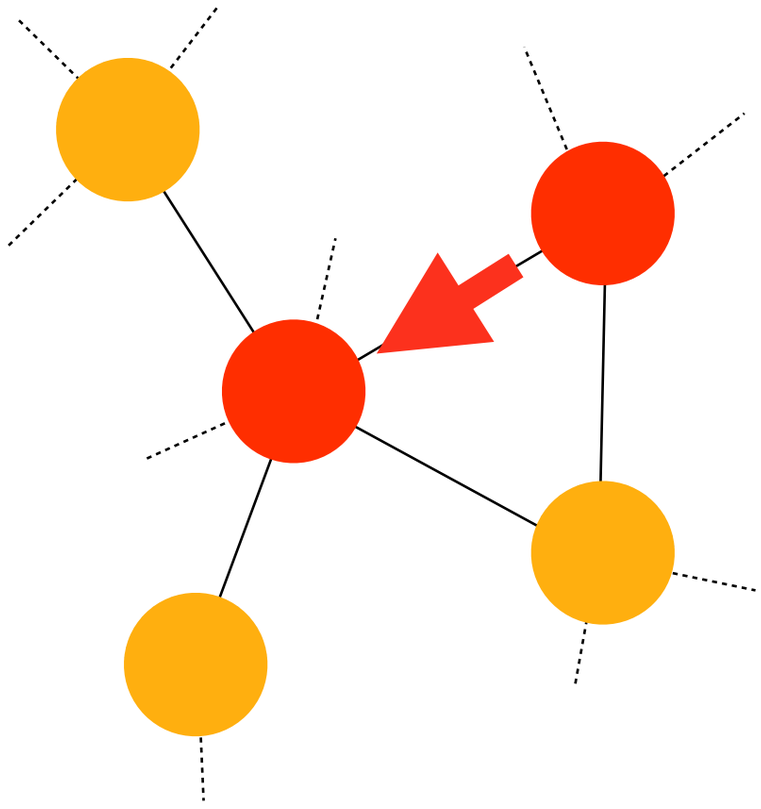
Spreading



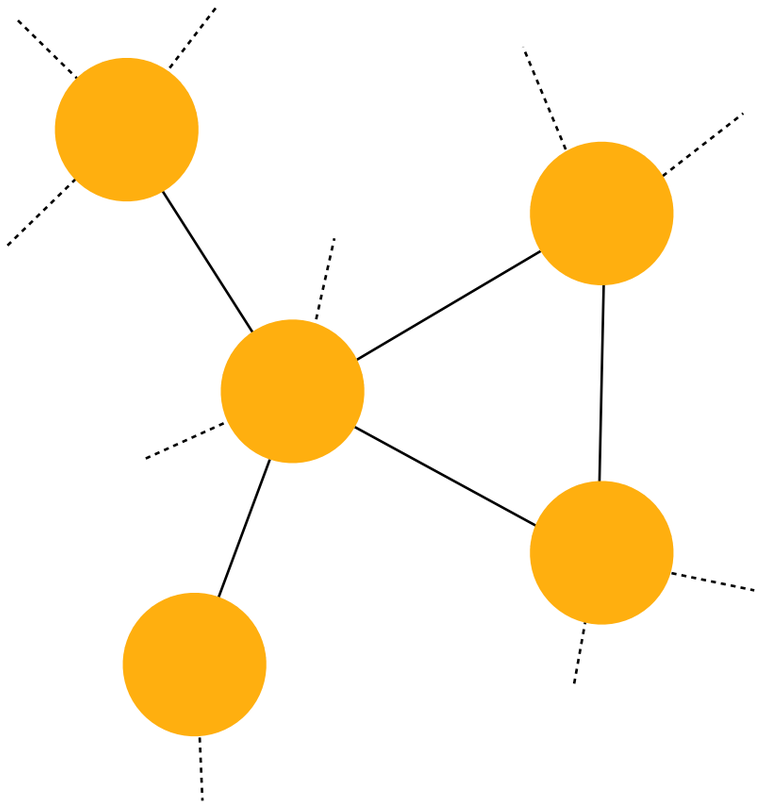
Conover et al., 2012



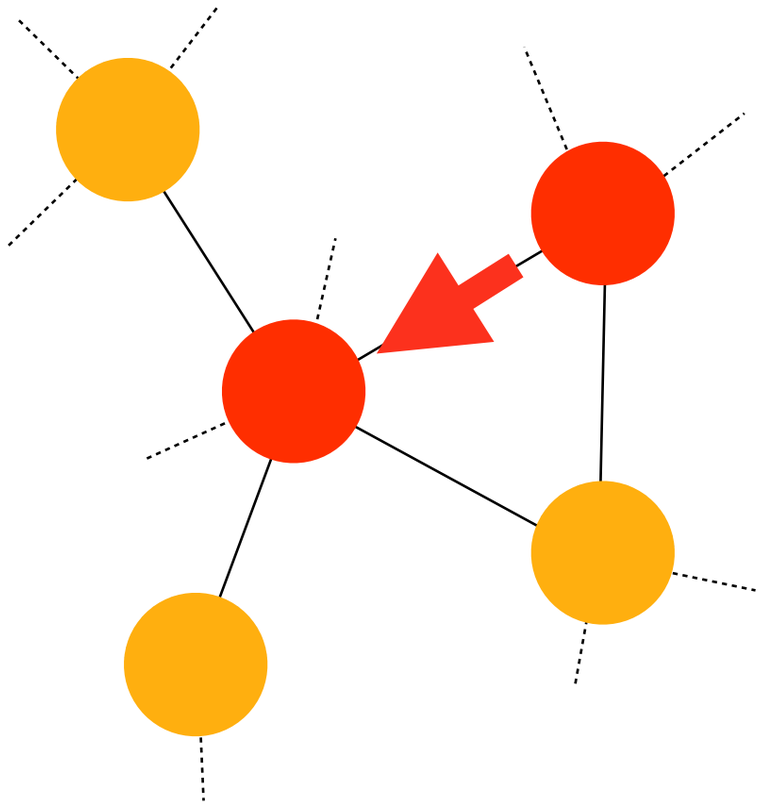
Network
(structure)



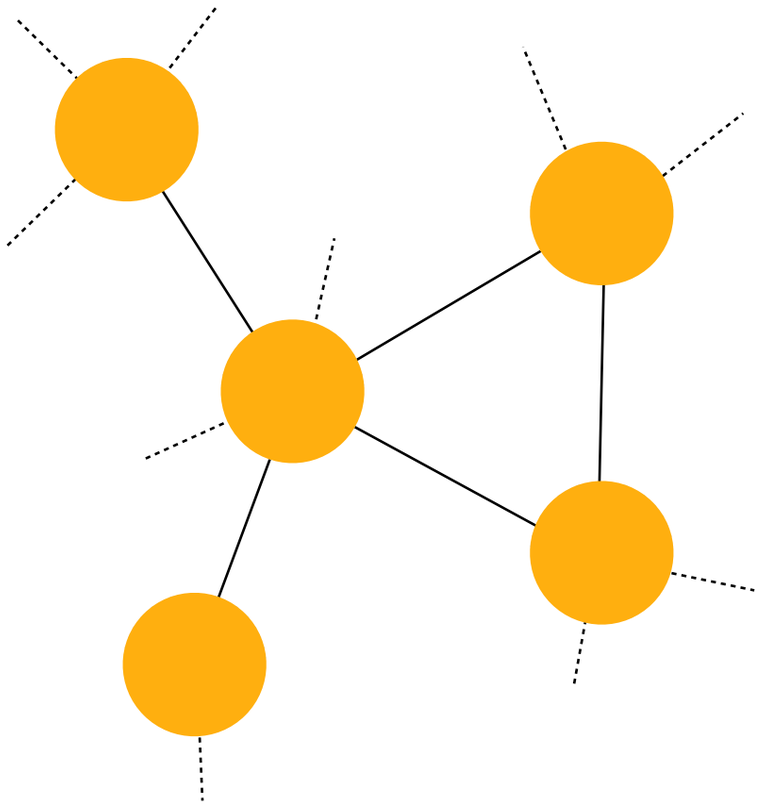
Diffusion
(Dynamics)

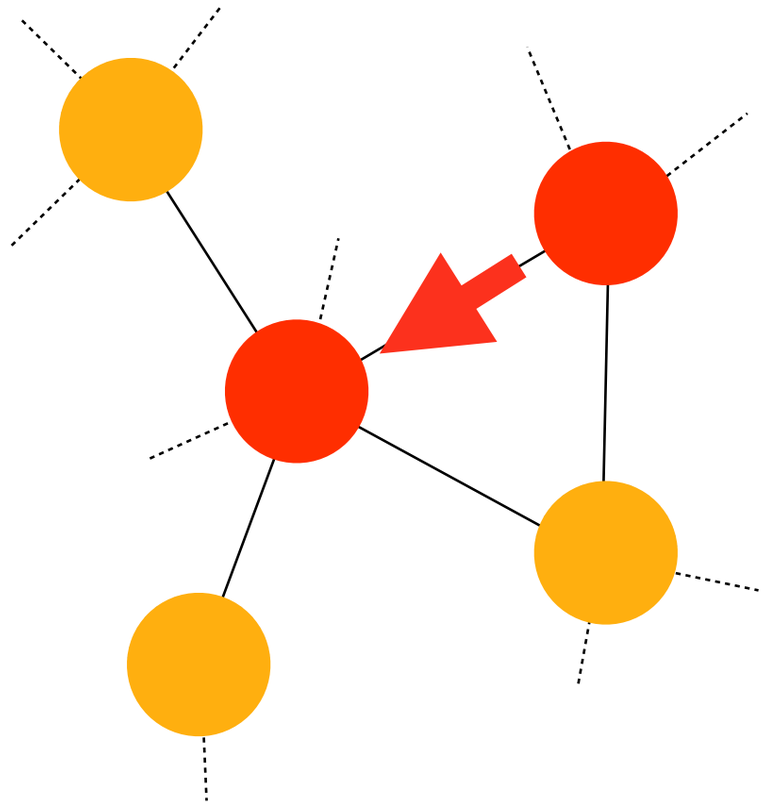


Network
(structure)

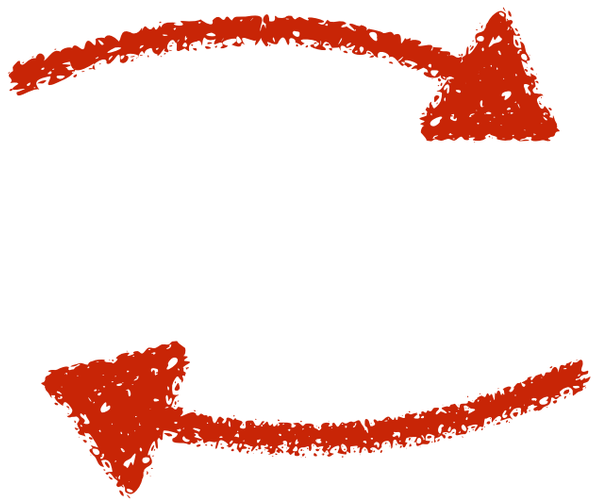


Diffusion
(Dynamics)



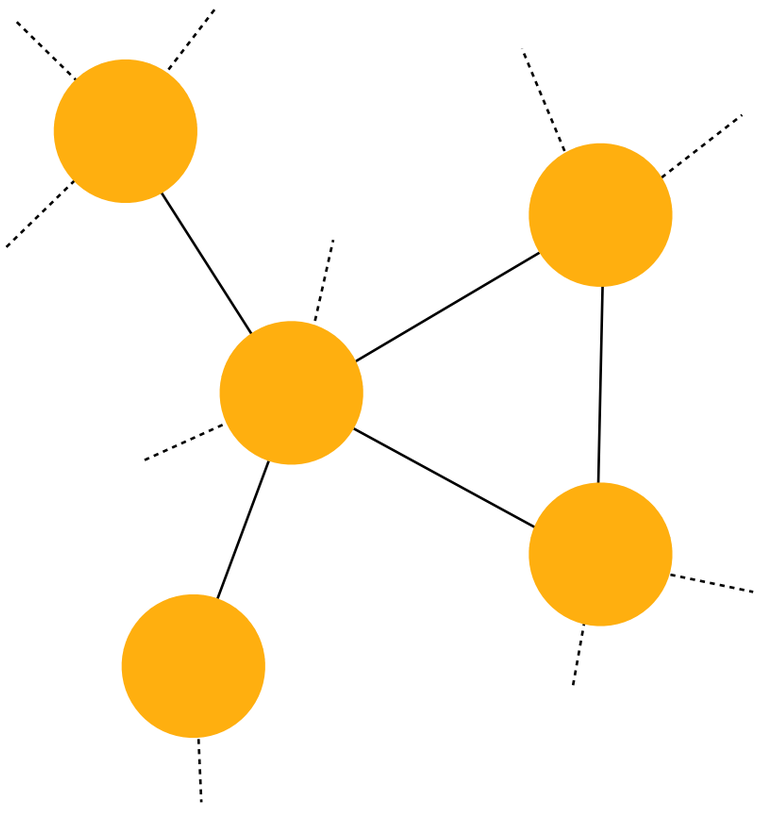


Network
(structure)



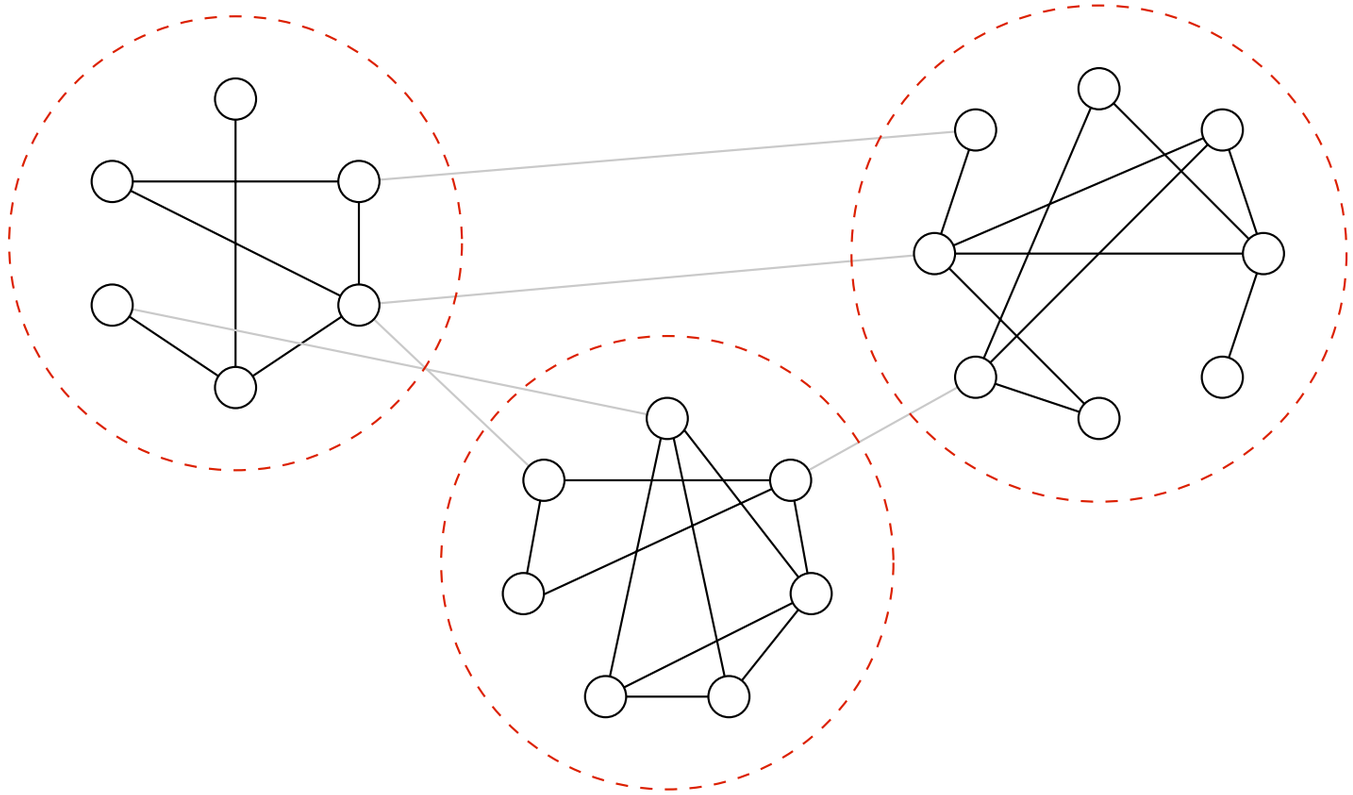
Diffusion
(Dynamics)

How are real networks organized?
How can we reveal them?



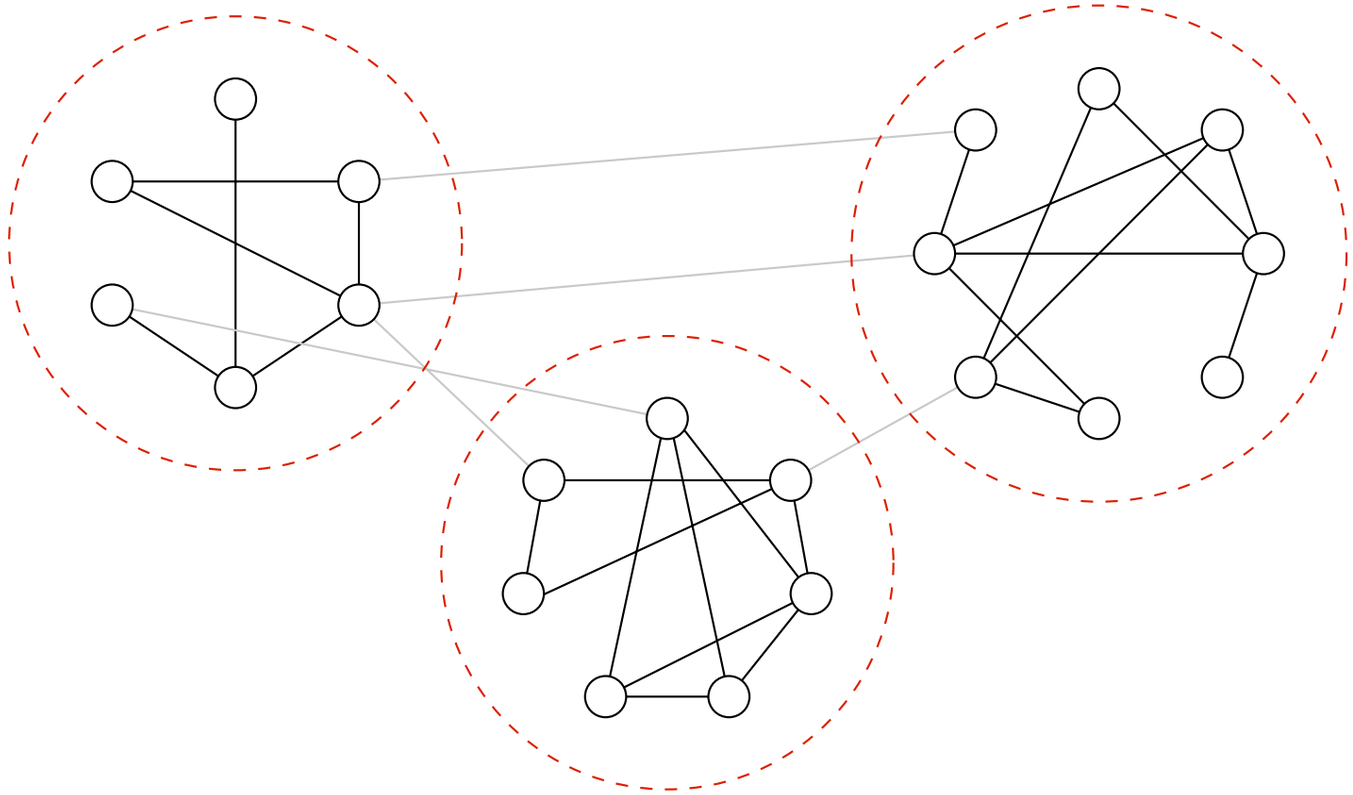
“Network Community”

“a group of densely interconnected nodes”

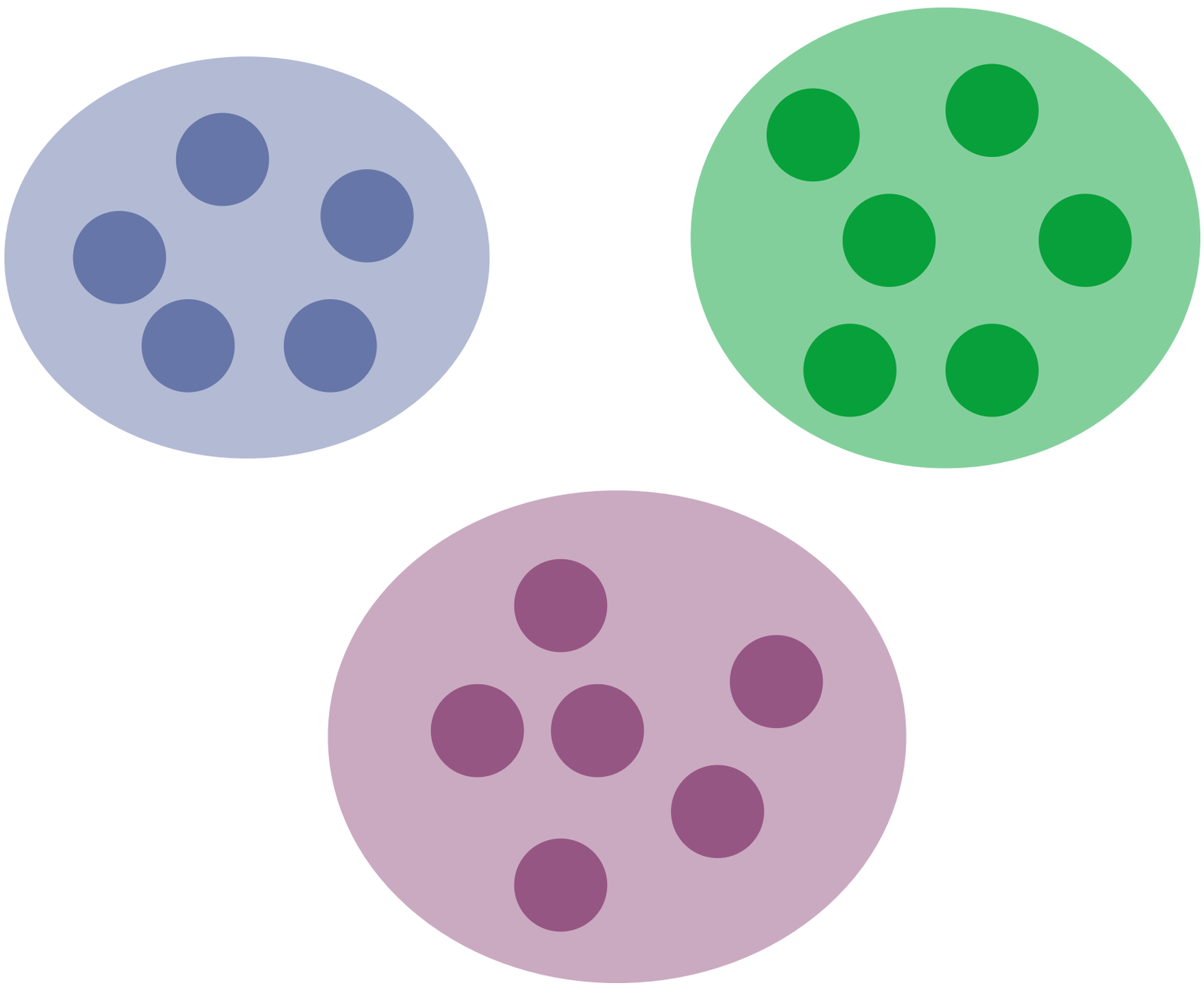


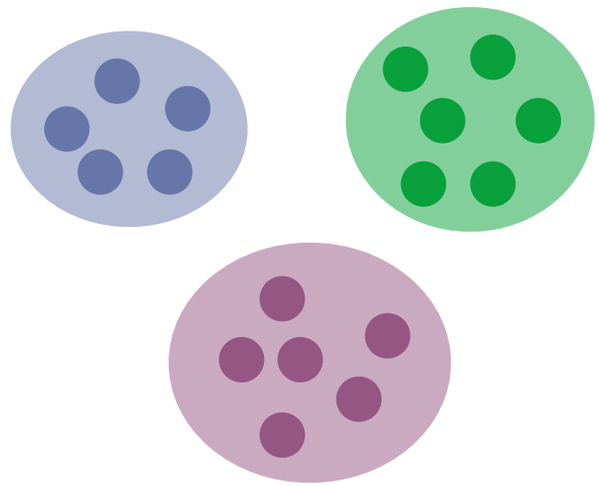
“Network Community”

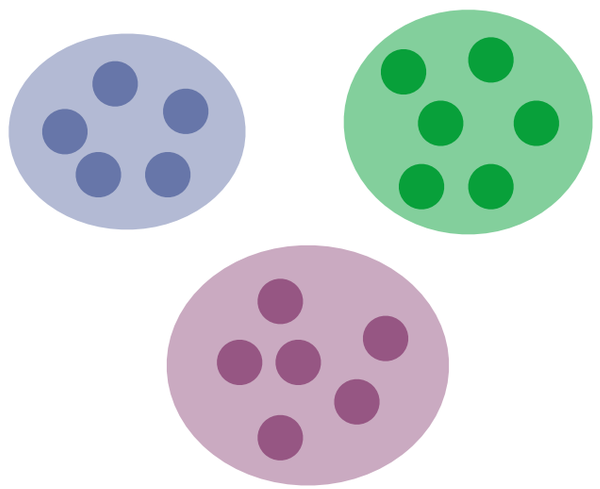
“a group of densely interconnected nodes”

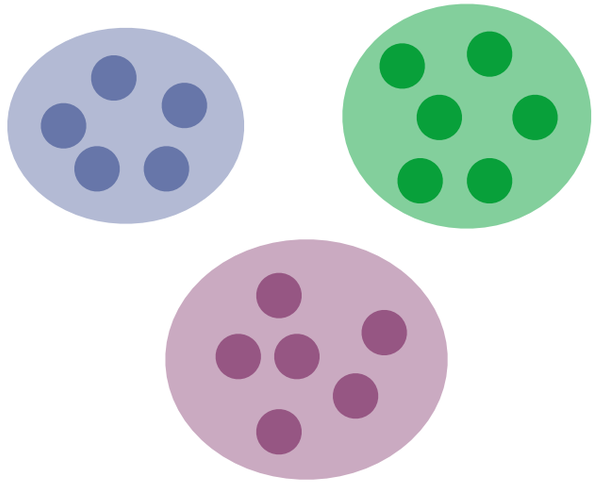


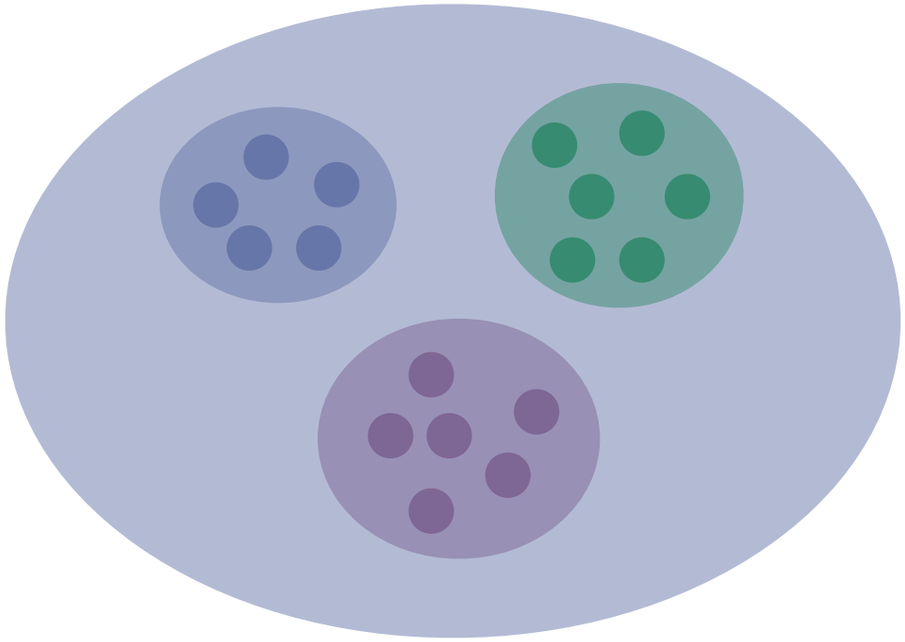
Hierarchy

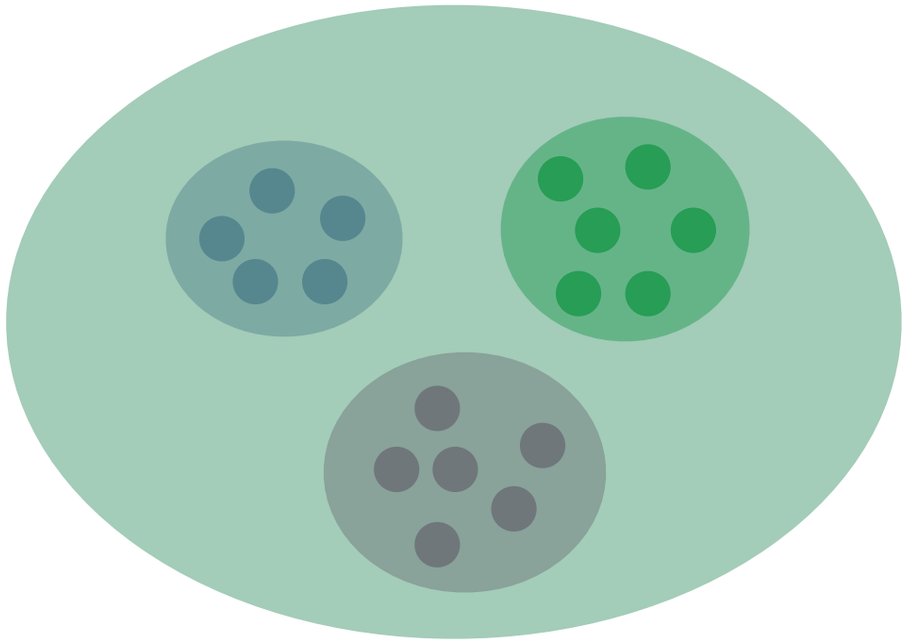


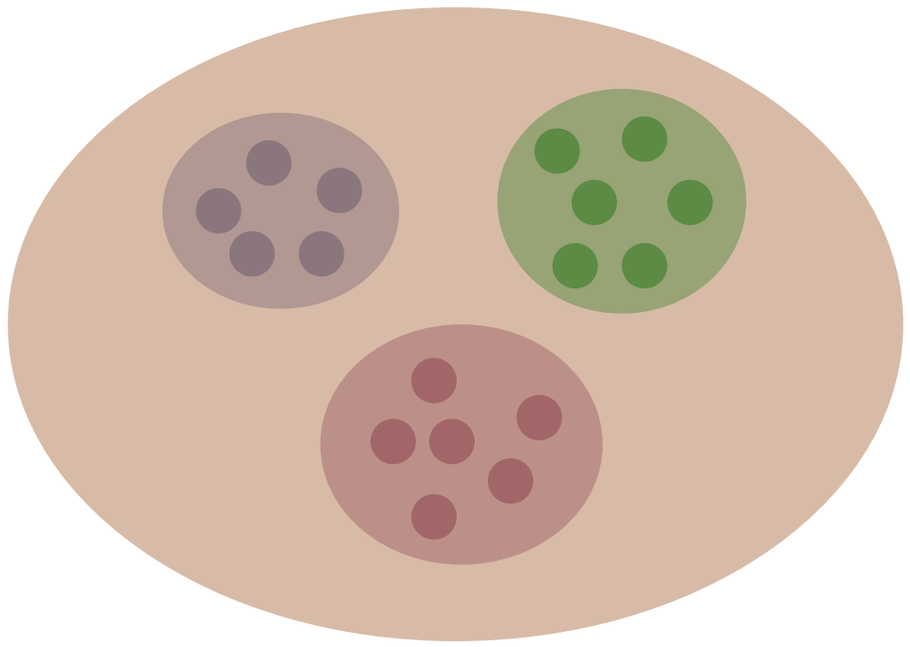






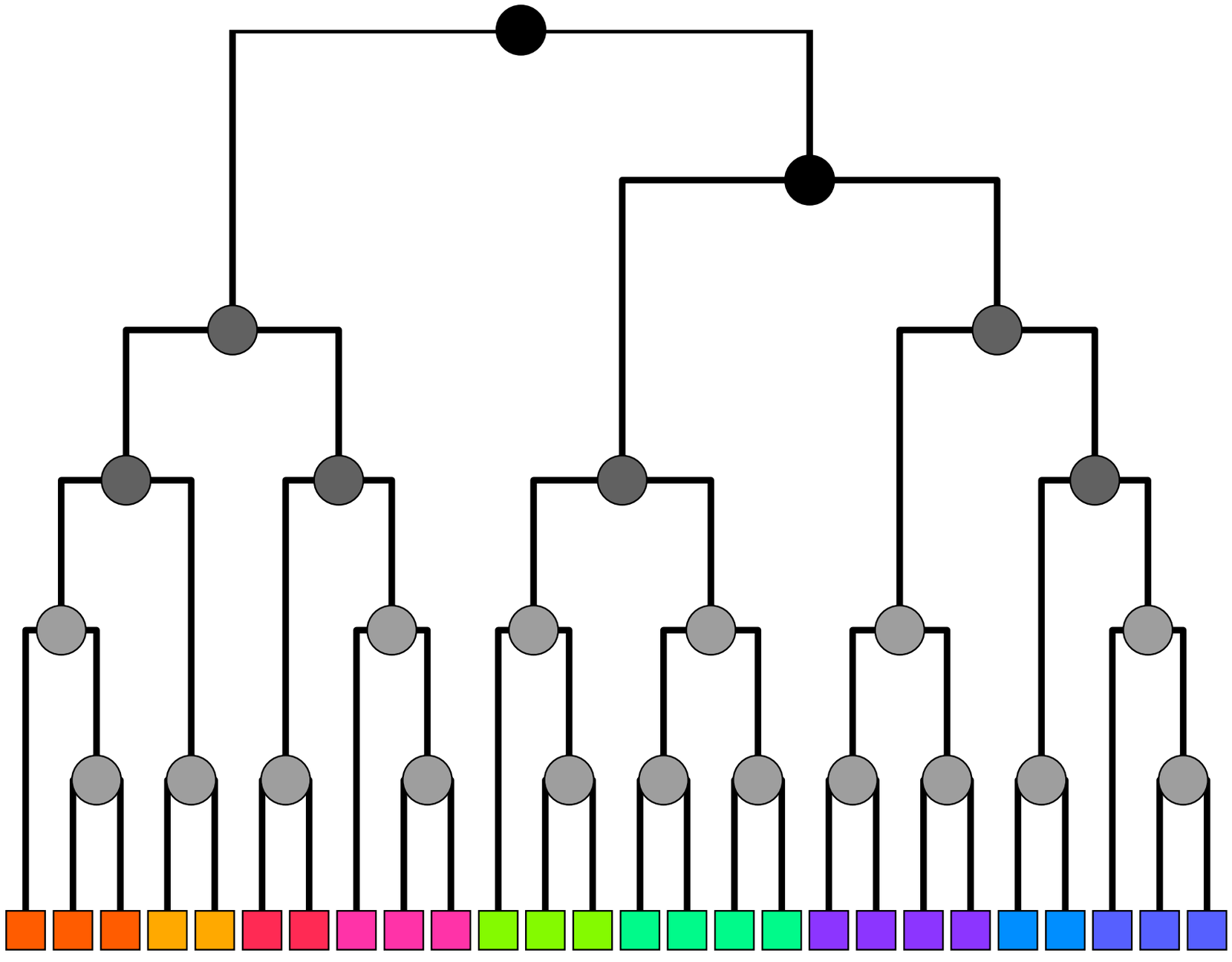




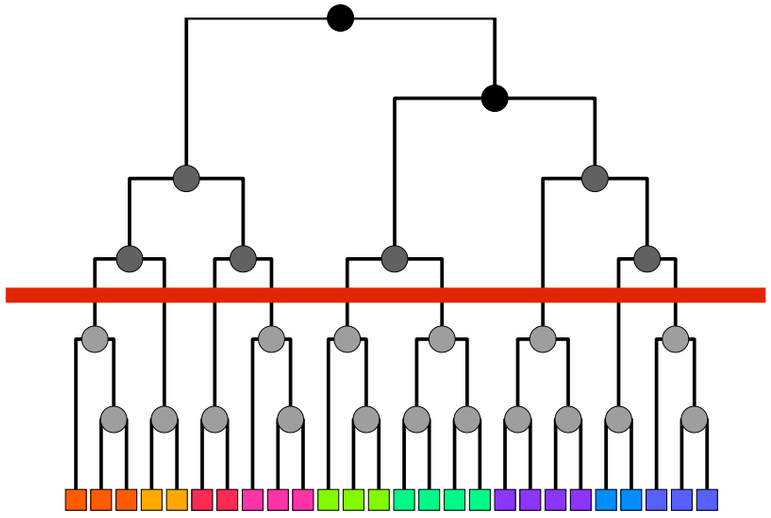


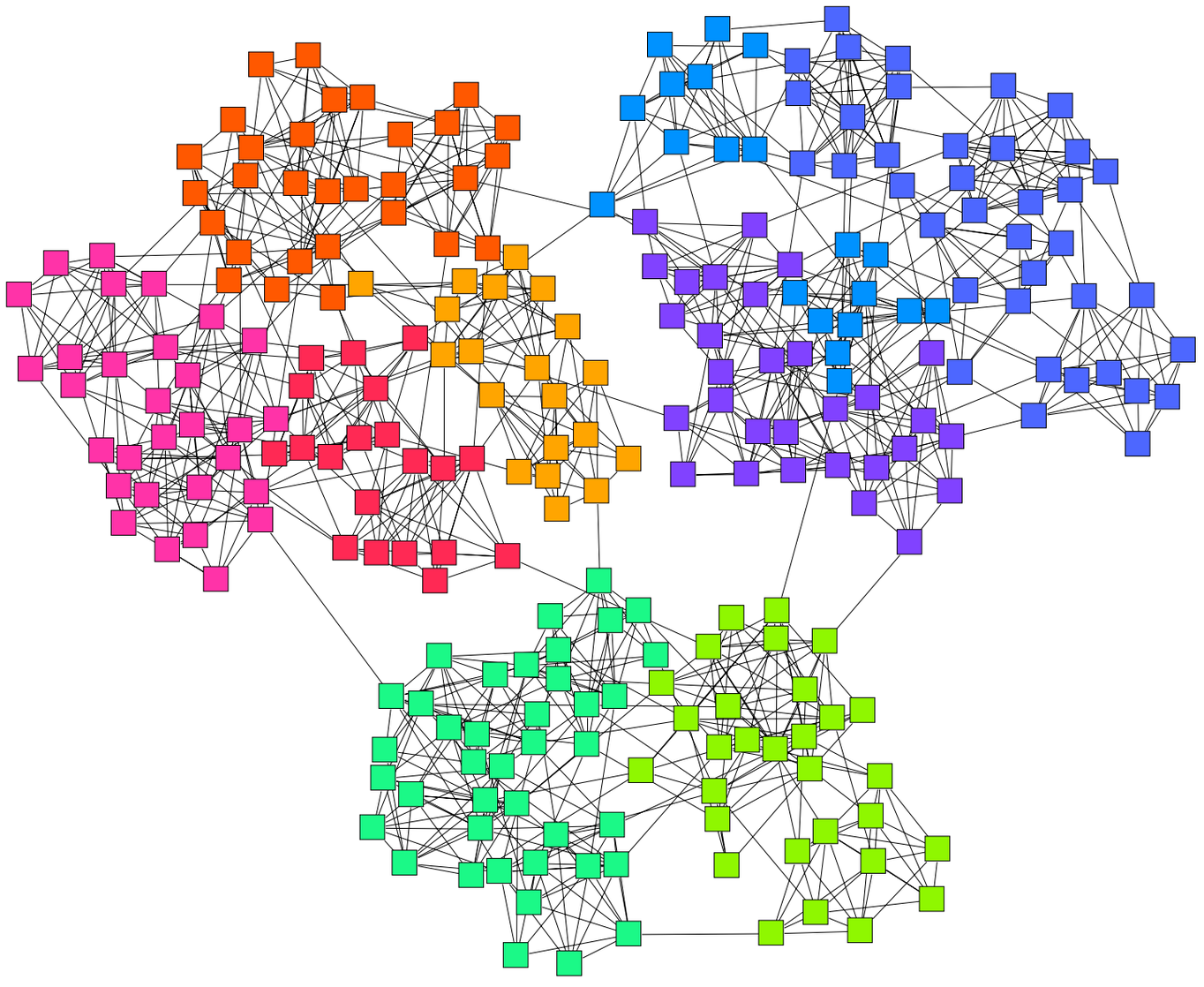
Hierarchy implies **communities**.

Hierarchical Random Graph model



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



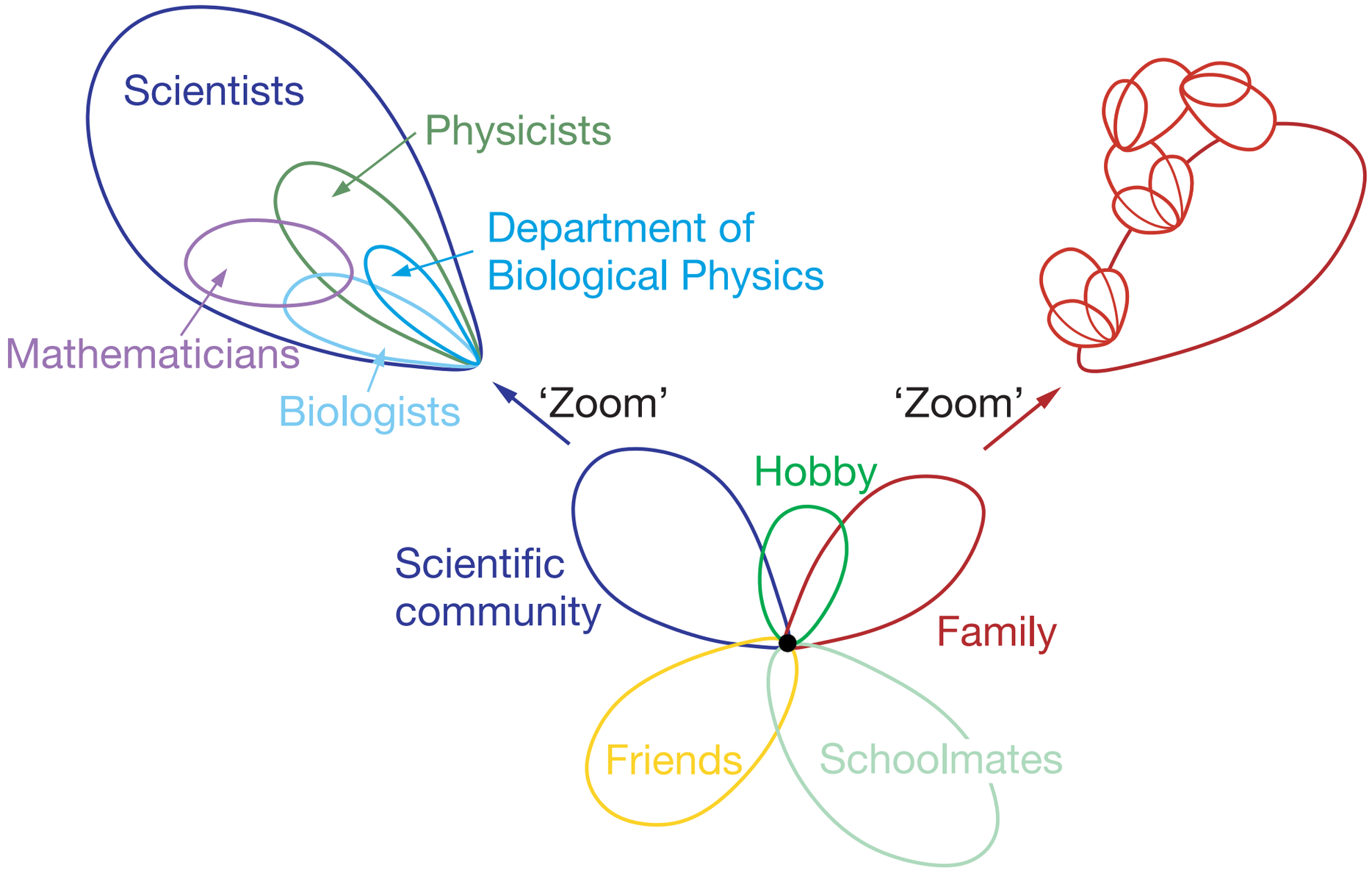


A. Clauset, C. Moore, and M. E. J. Newman, *Nature* (2008)

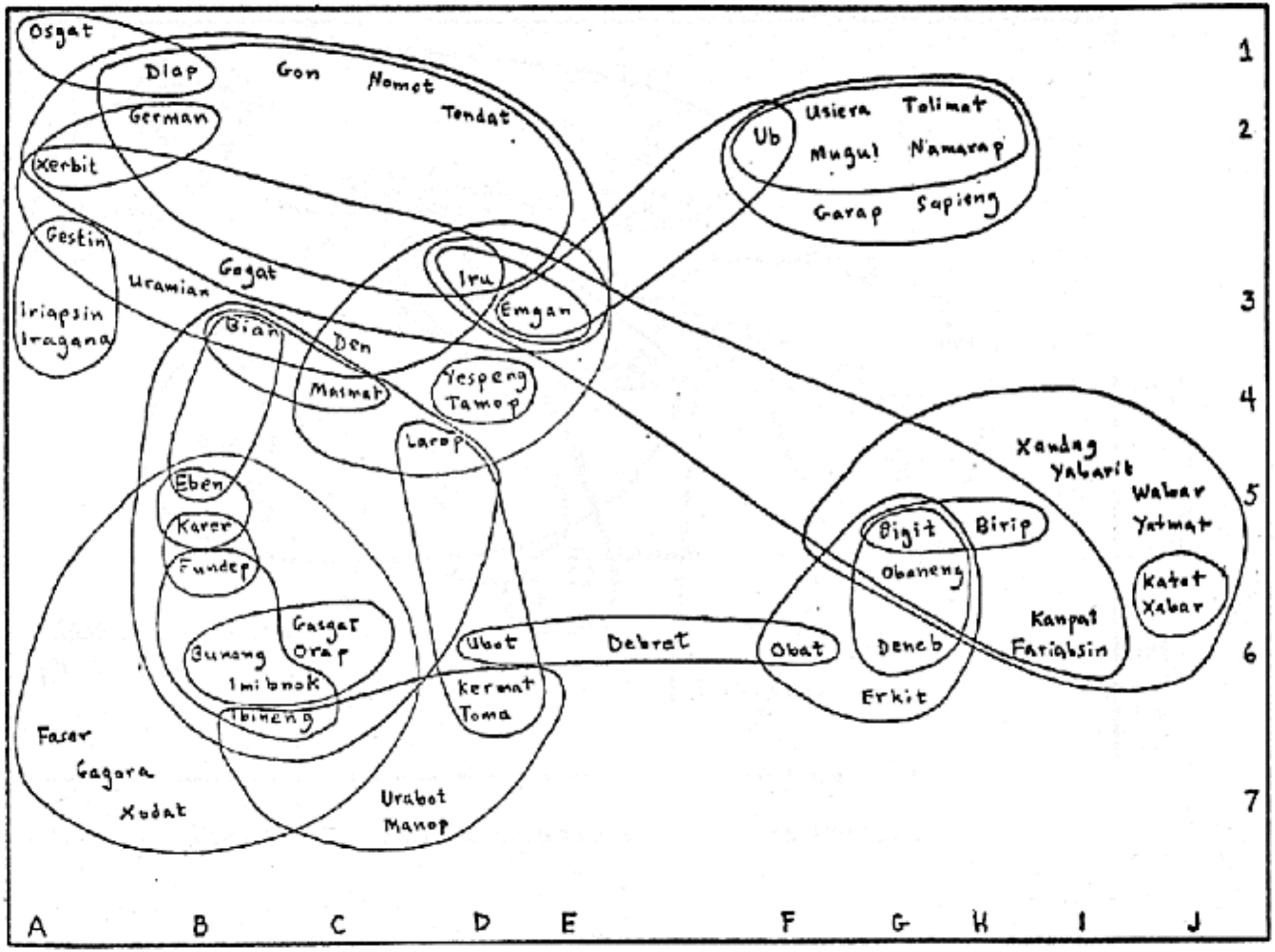
Hierarchical community structure

Hierarchy ⟶ Communities

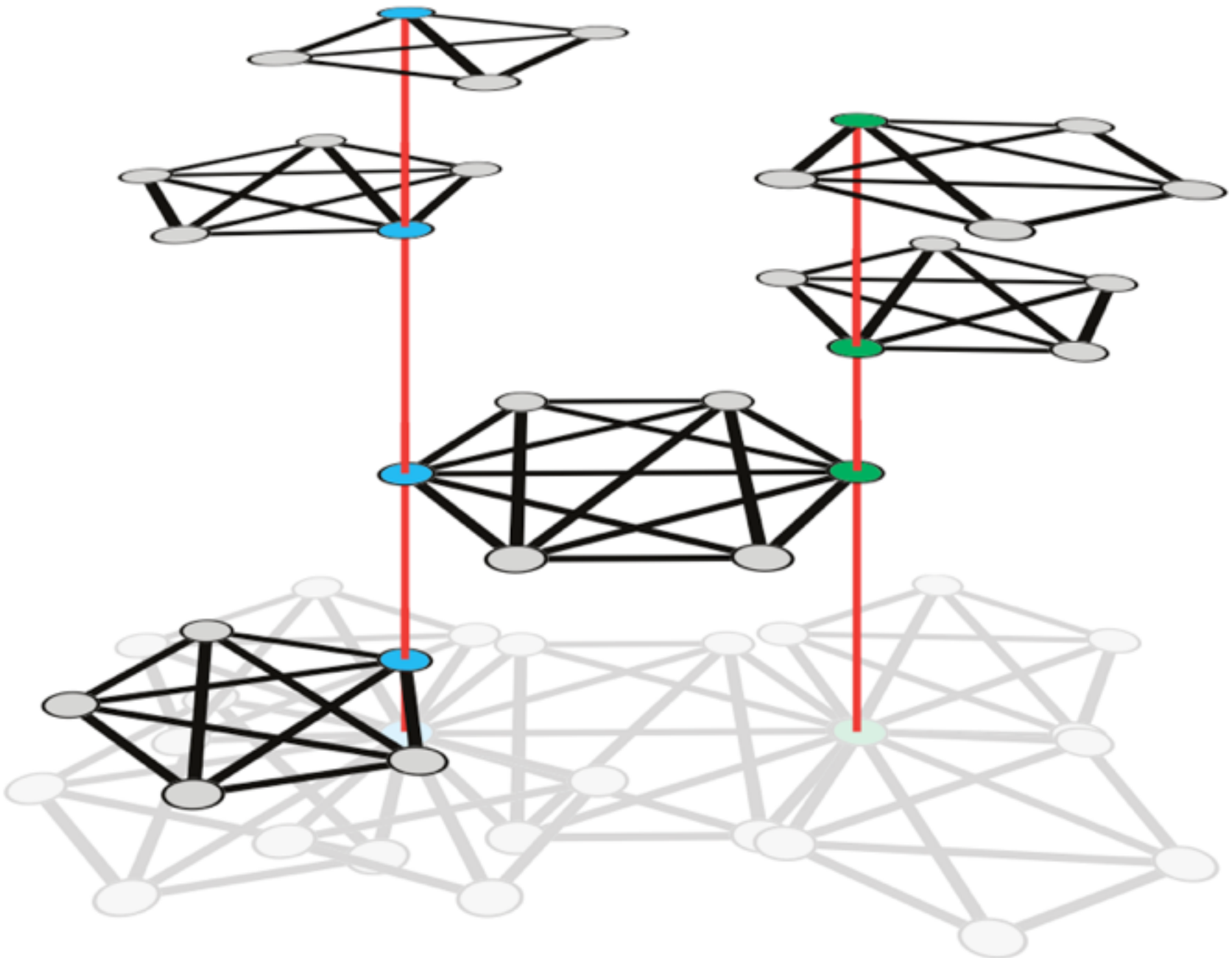
Overlap

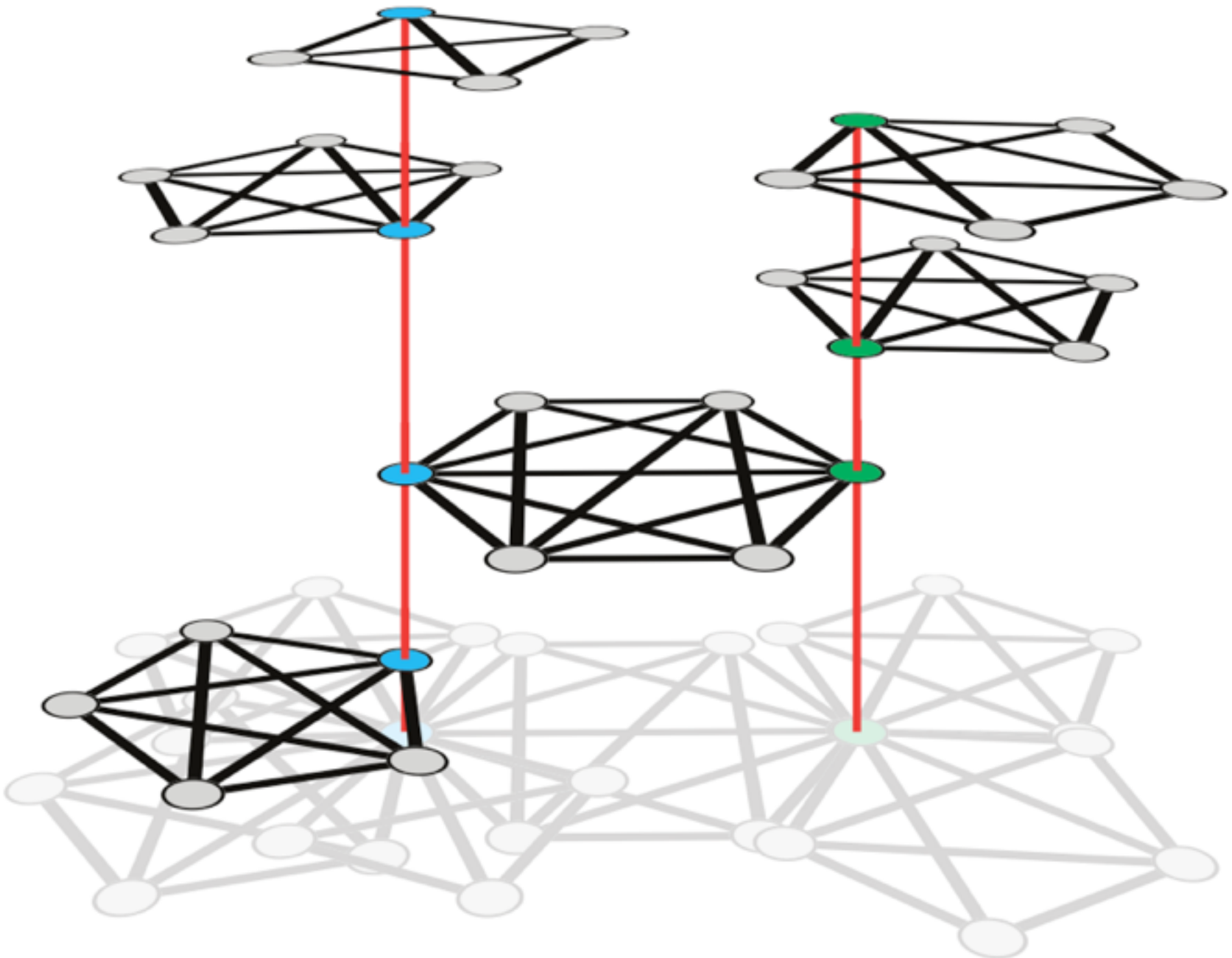


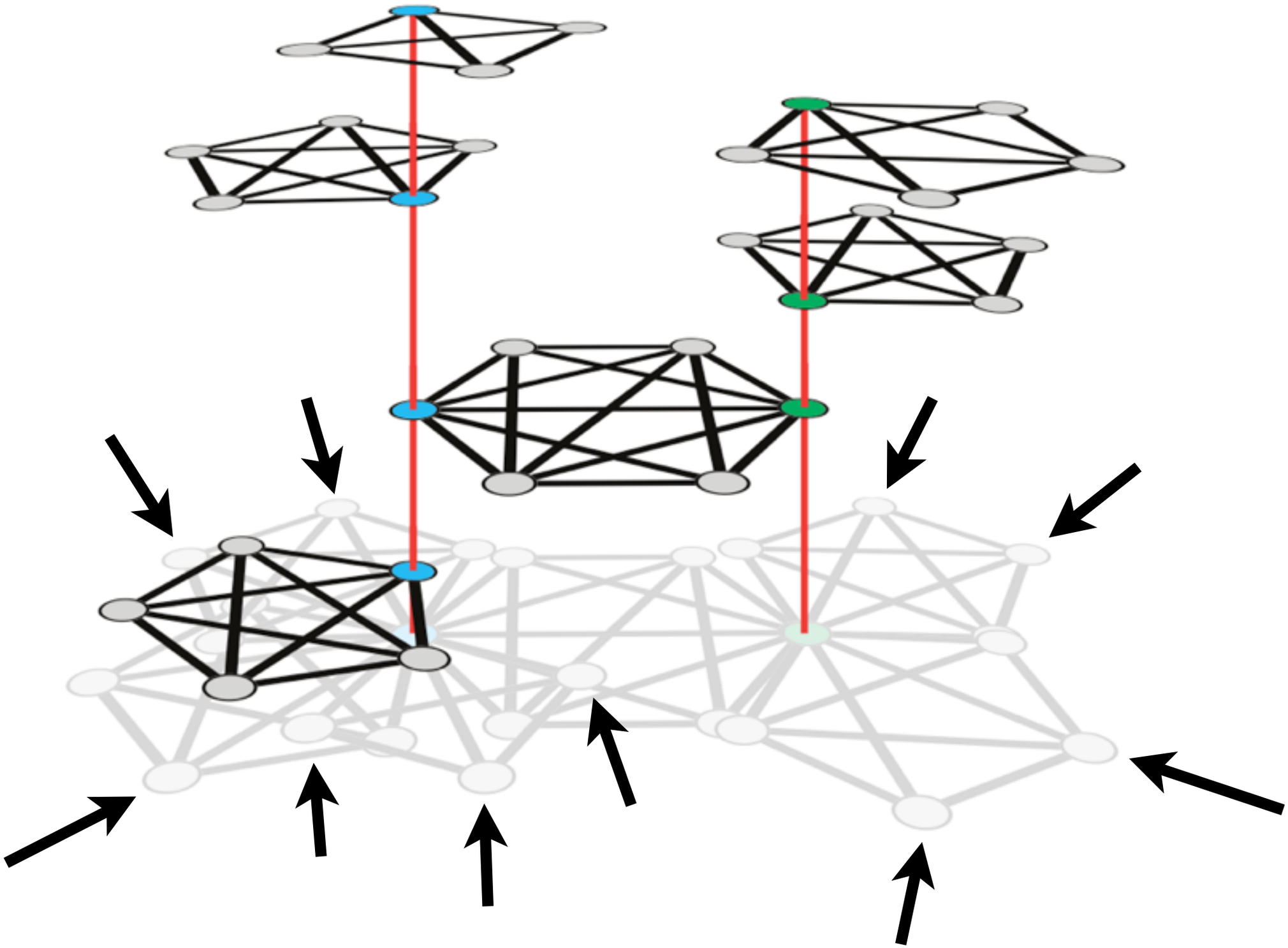
G. Palla, I. Derényi, I. Farkas & T. Vicsek, *Nature* (2005)



Arnold Perey, Social organization of *Oksapmin*, Papua New Guinea (1973)





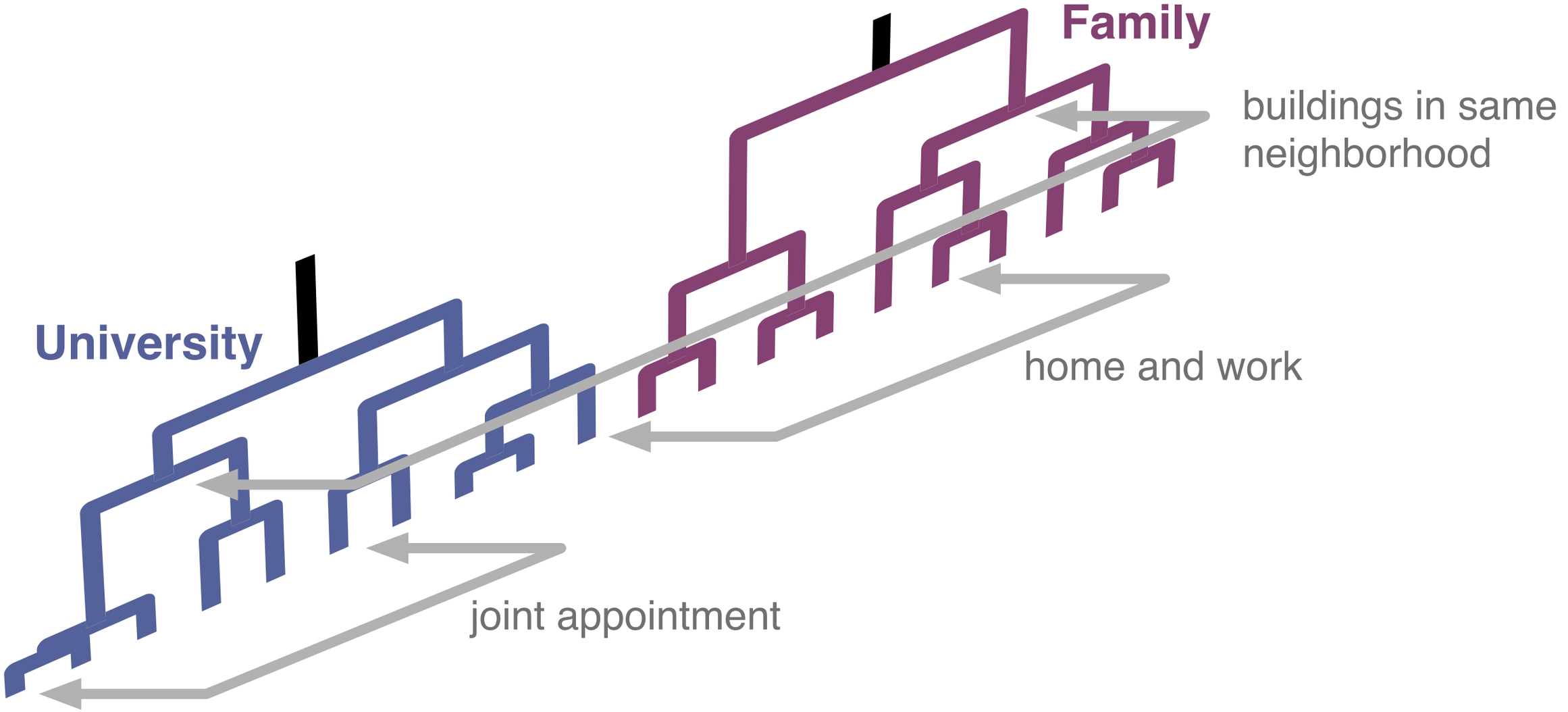


Seinfeld - Independent George

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

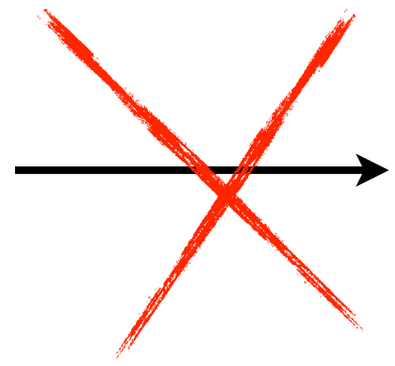
Overlap is

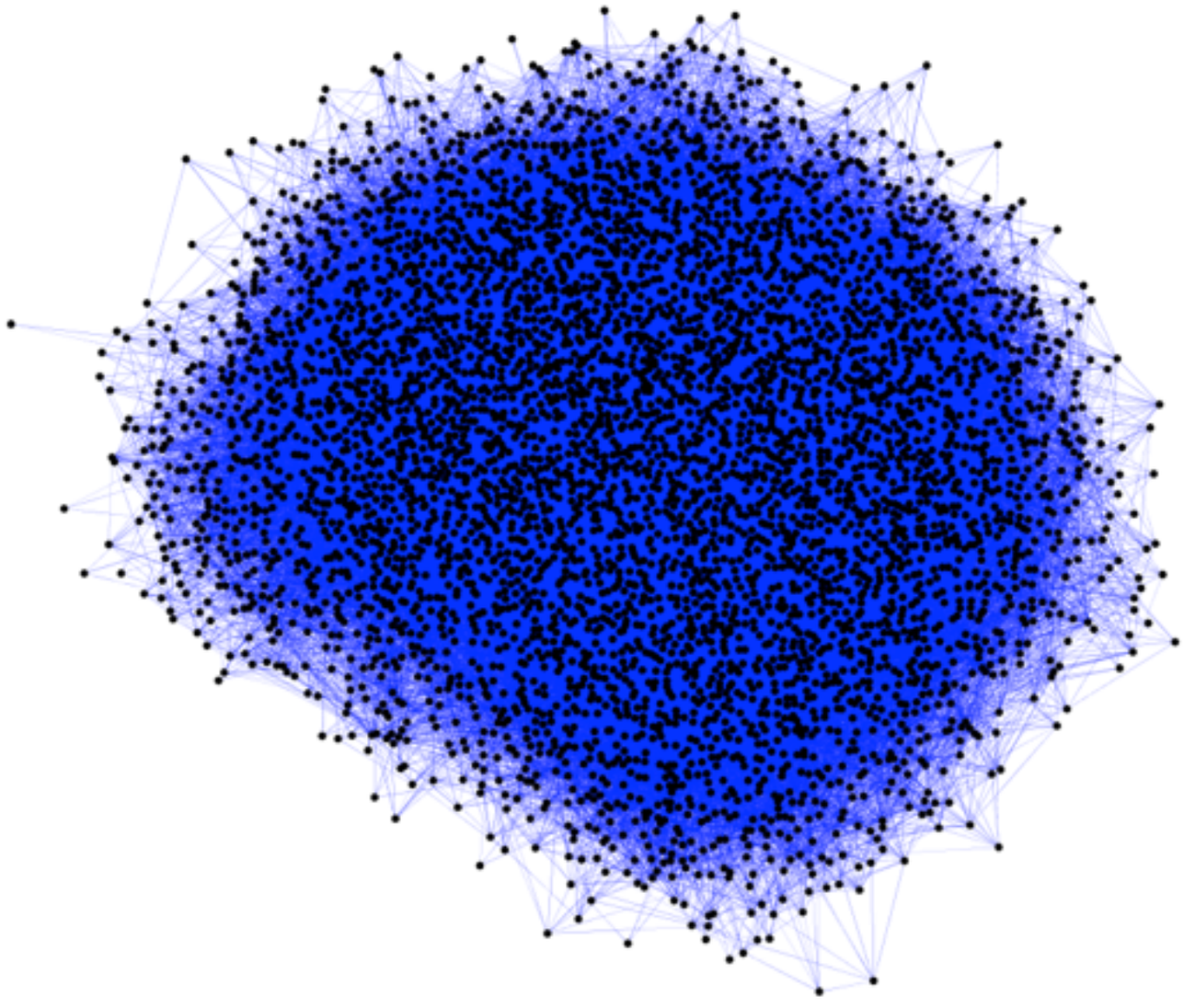
pervasive.



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

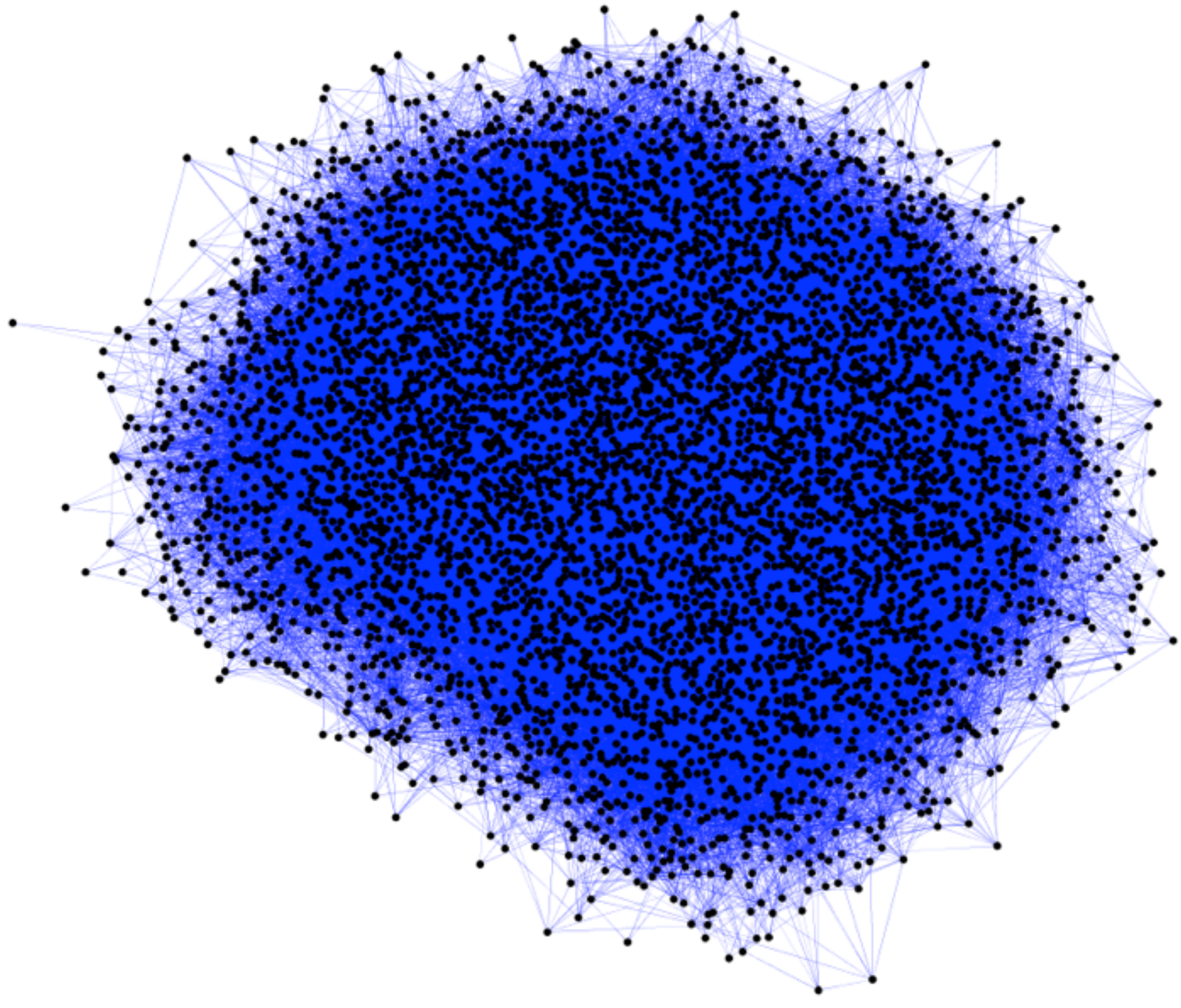
Hierarchical community structure

Hierarchy  Communities

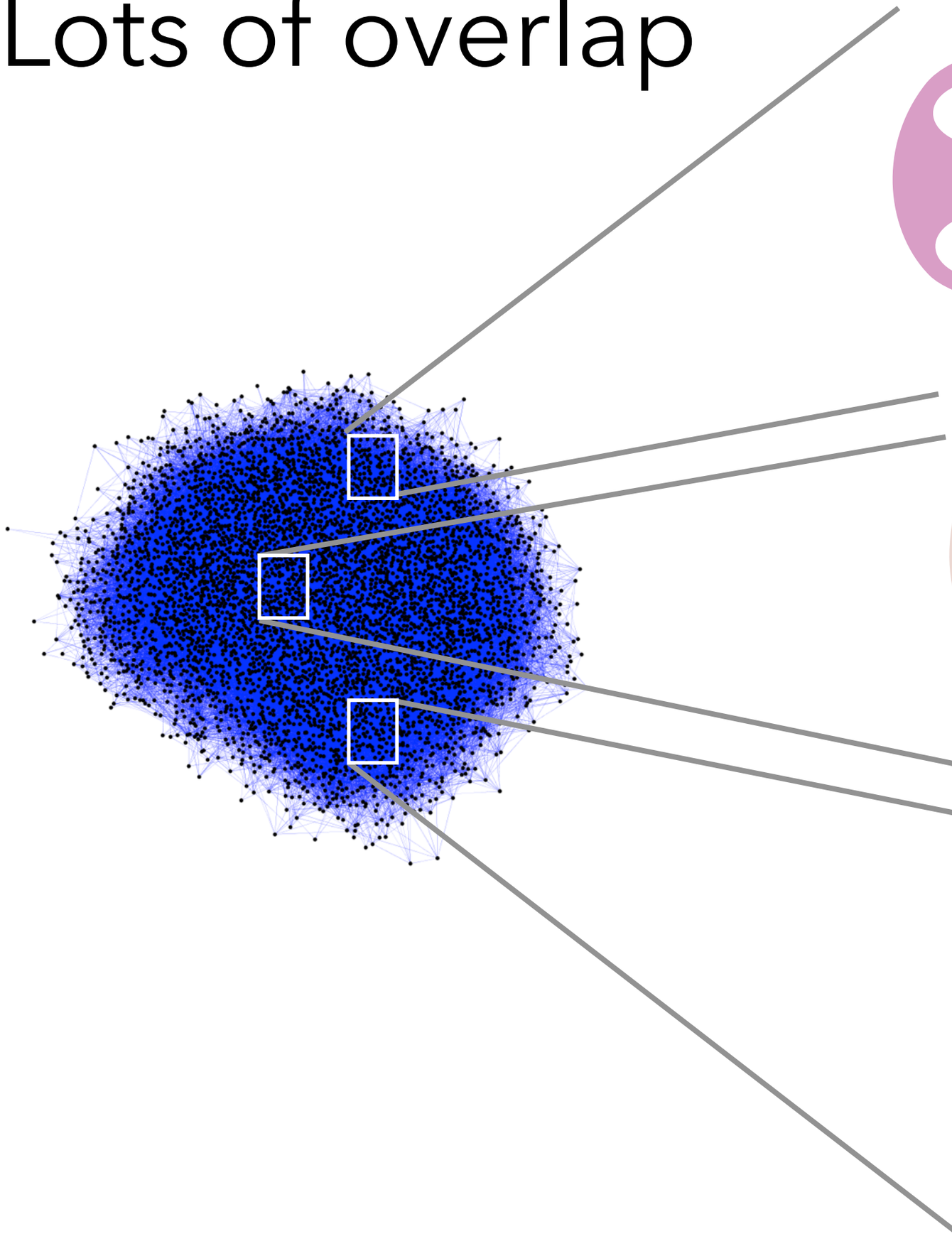


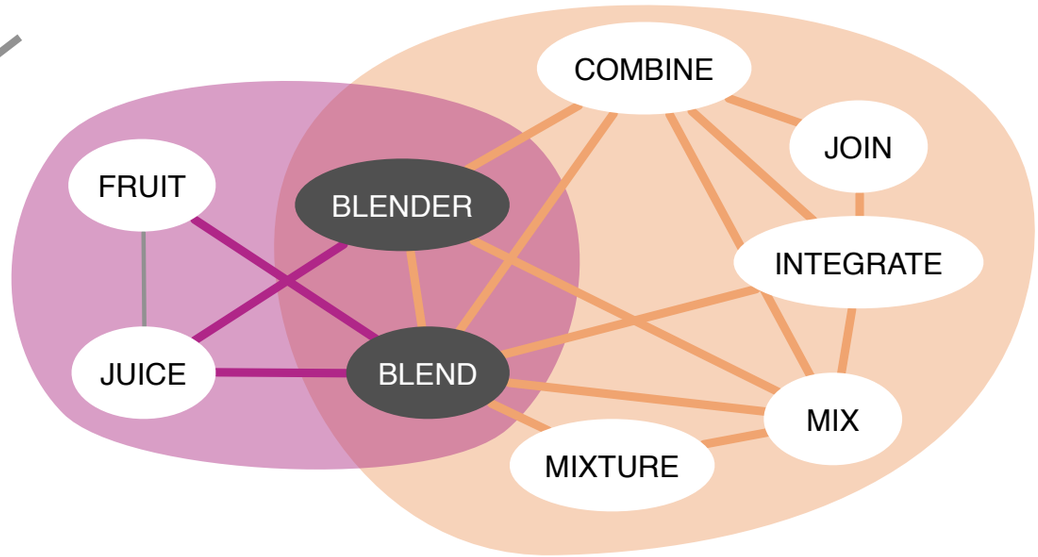
This is a modular network.

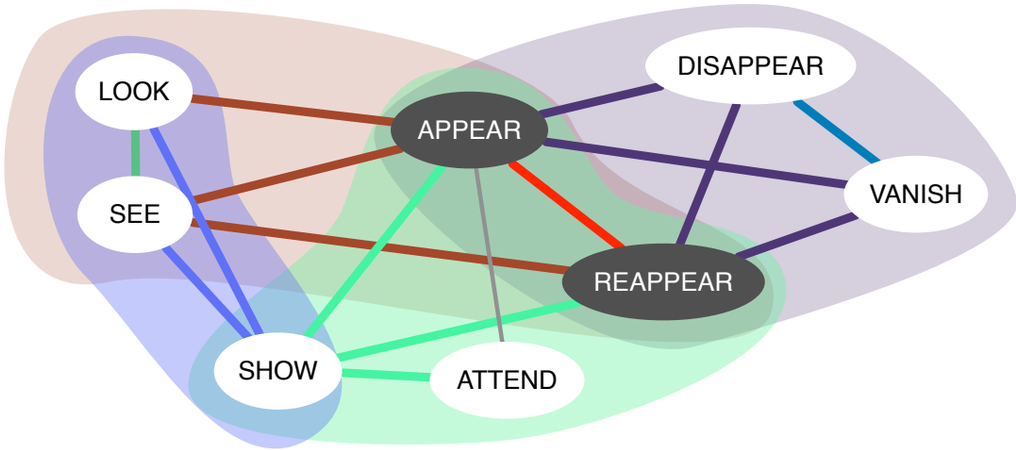
What is this?

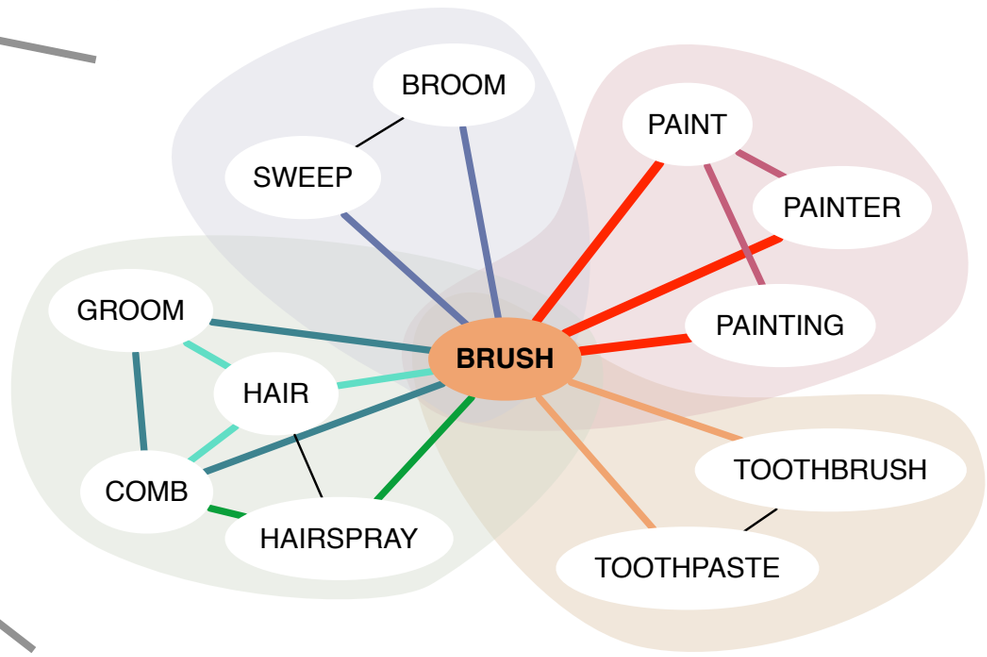


Lots of overlap

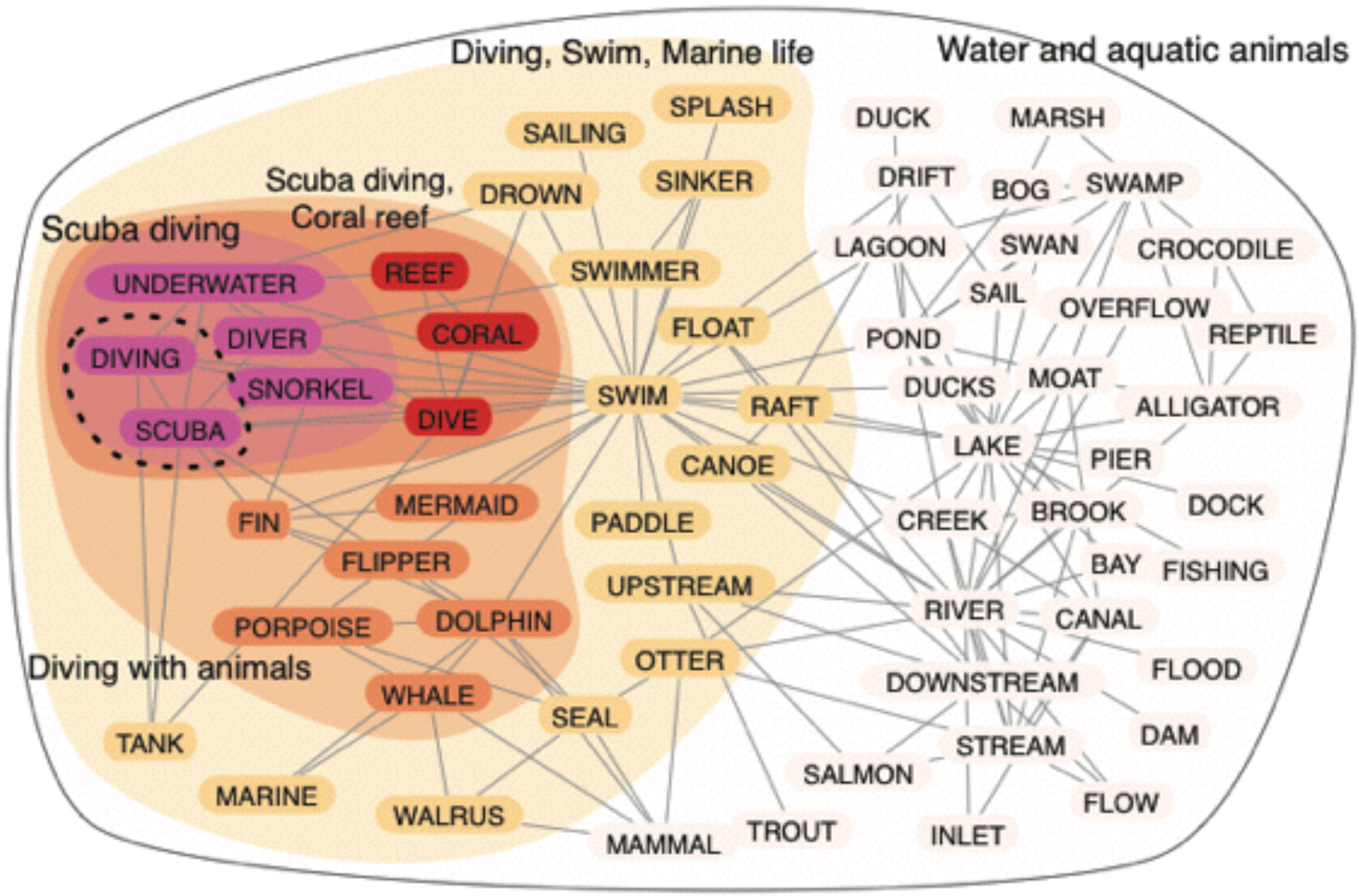








and hierarchy

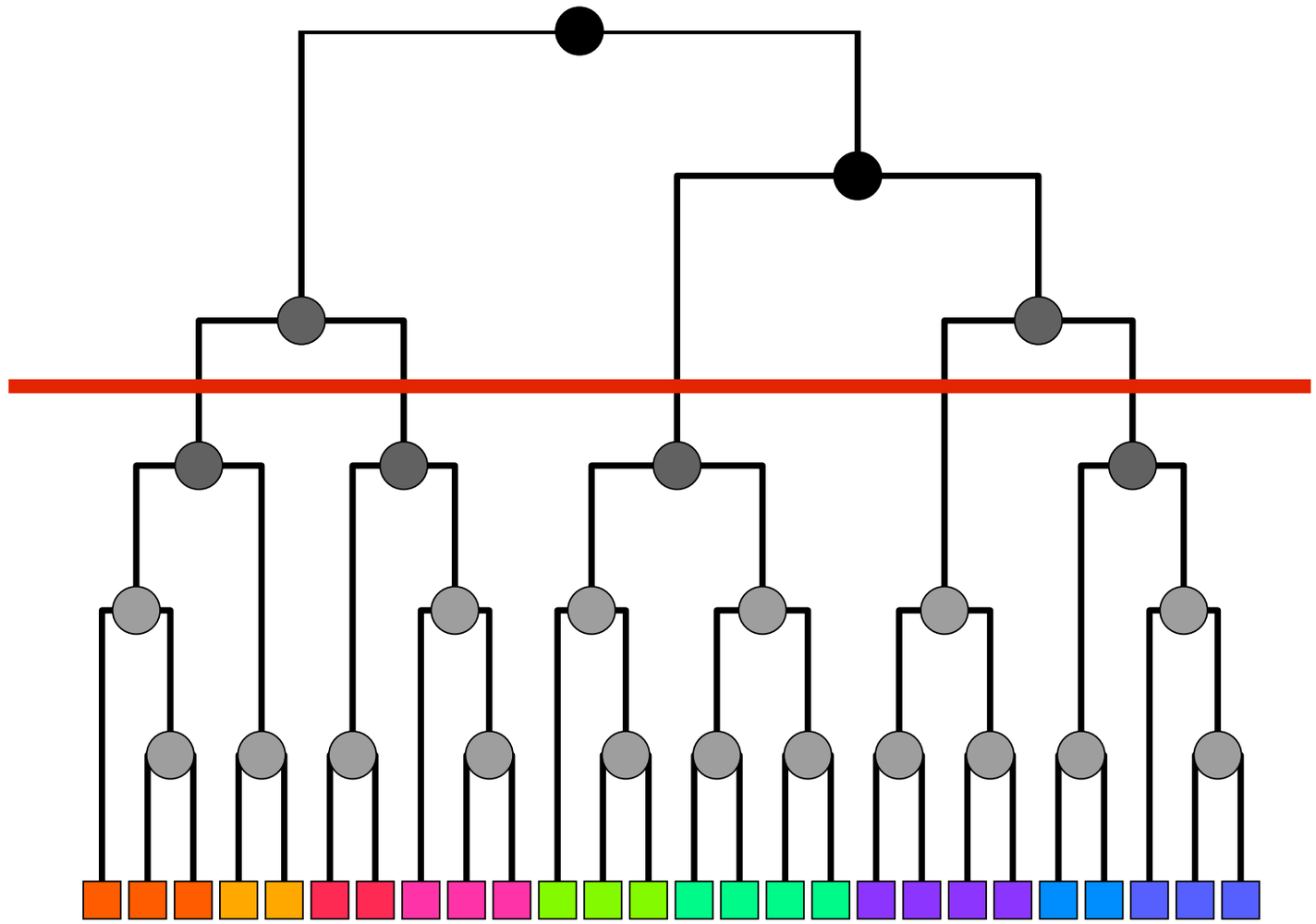


Here is the **PROBLEM.**

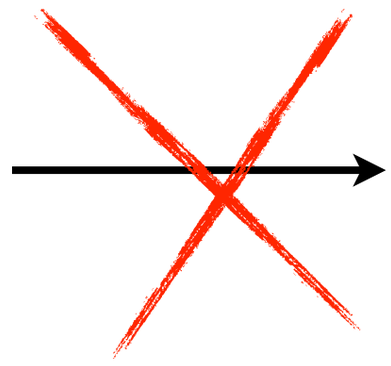
Communities overlap.

Hierarchical structure exists.

Hierarchy implies **disjoint communities.**



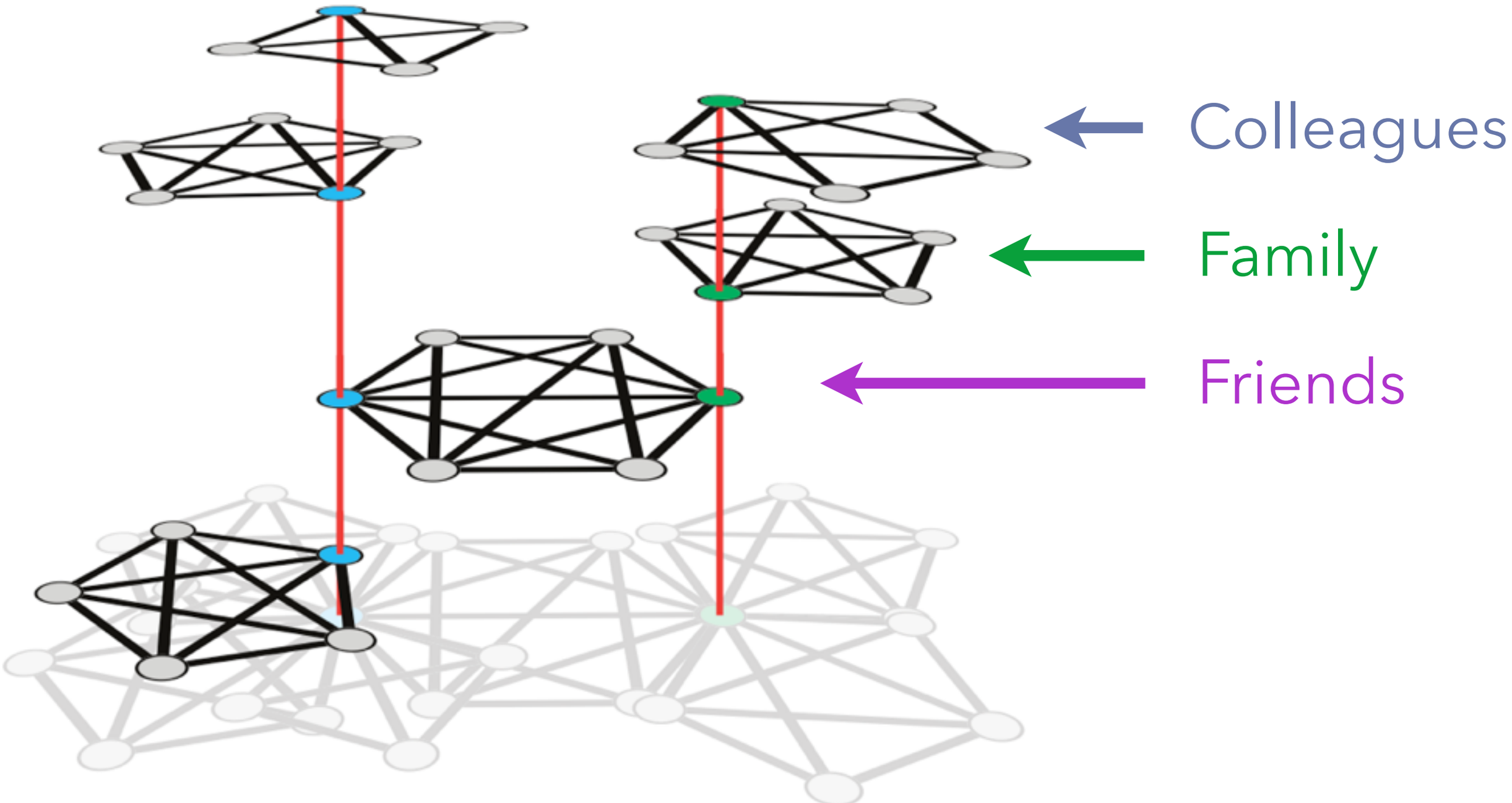
Hierarchical community structure

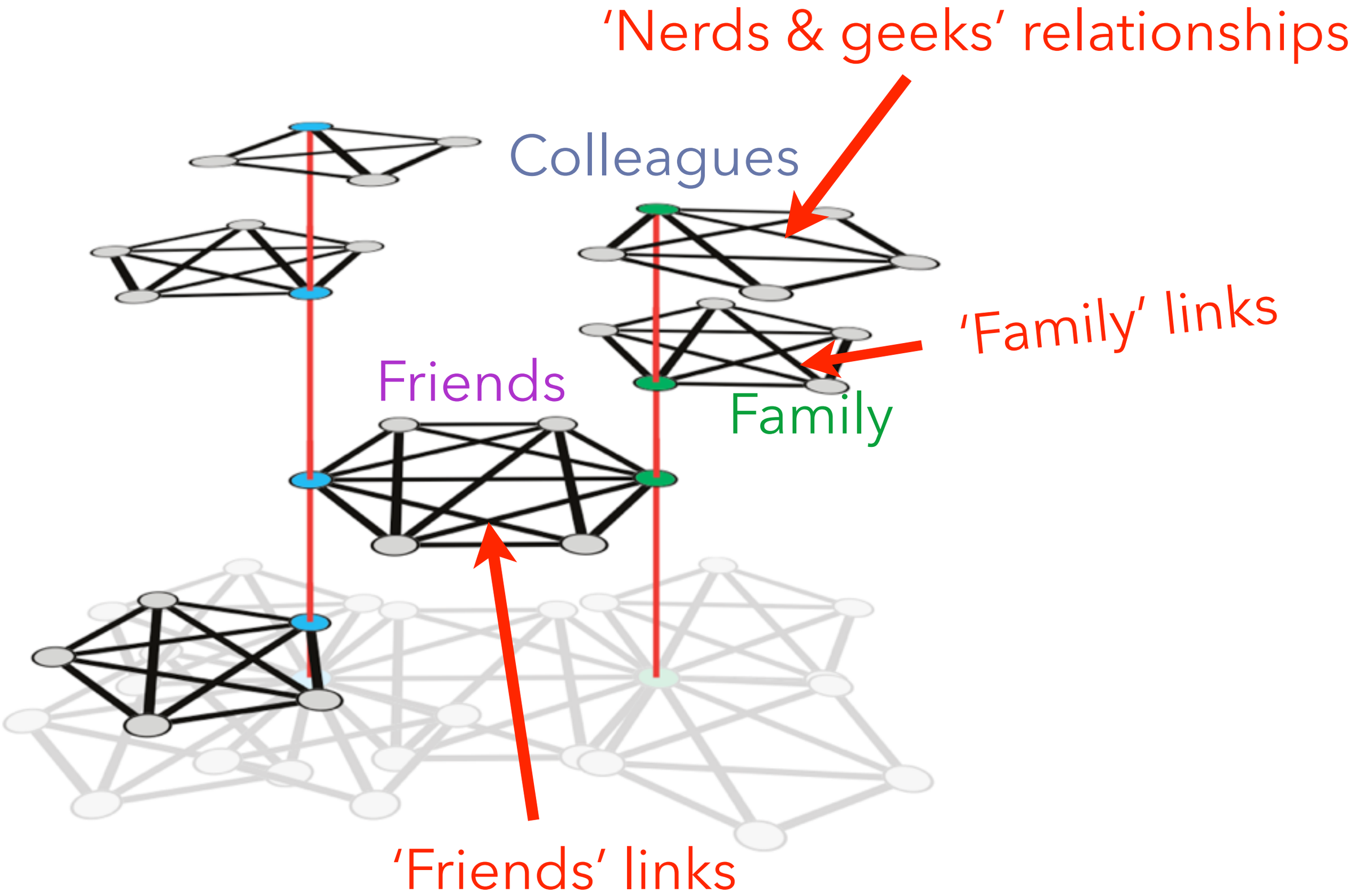
Hierarchy  Communities

How can we discover both **overlap** and **hierarchy**?

How can we discover both **overlap** and **hierarchy**?

link communities


Colleagues
Family
Friends


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

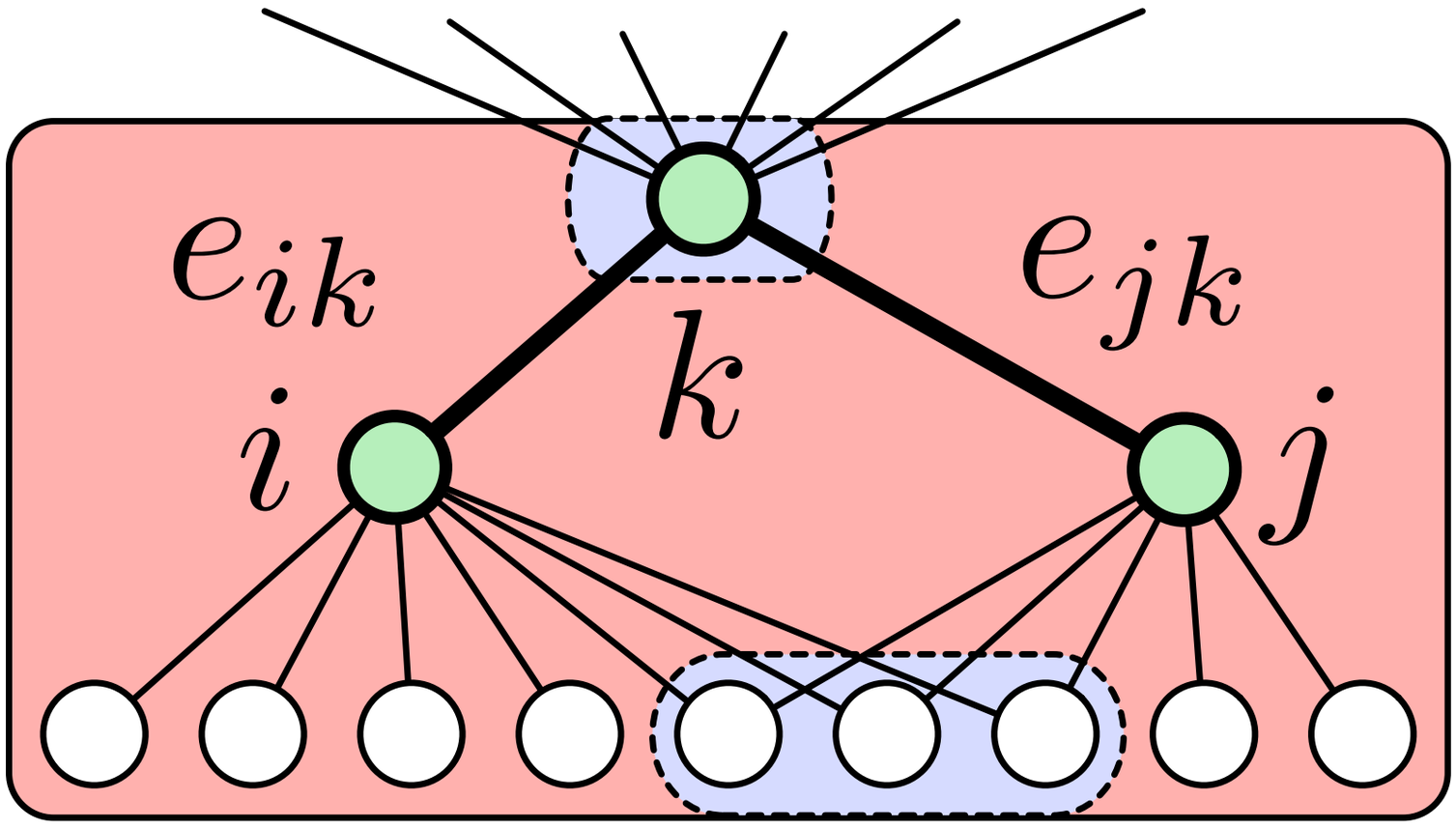
Nodes: multiple membership

Links: **unique** membership

Nodes: multiple membership

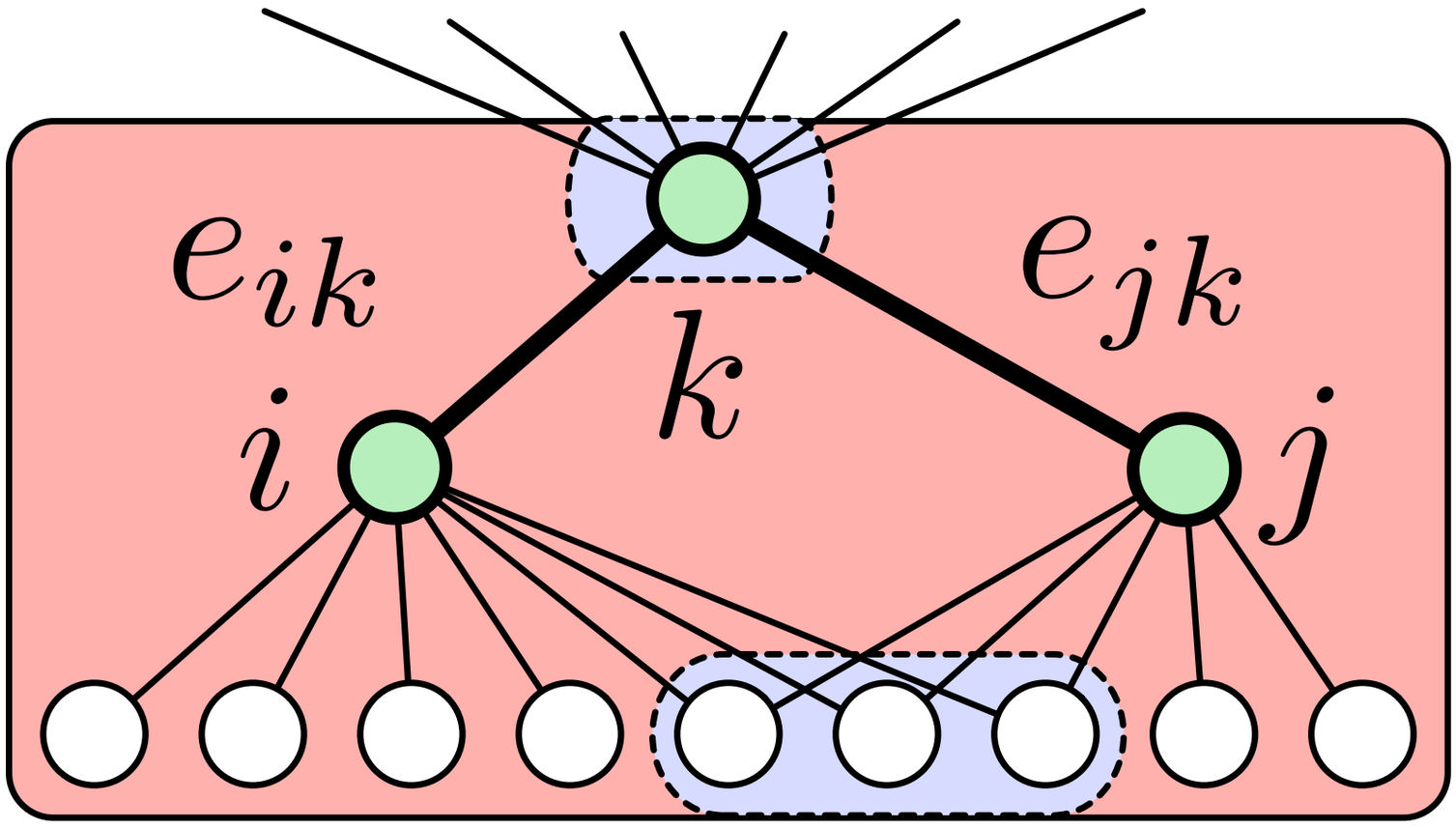
Links: **unique** membership

Then, why don't we define communities in terms of links (edges)?



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

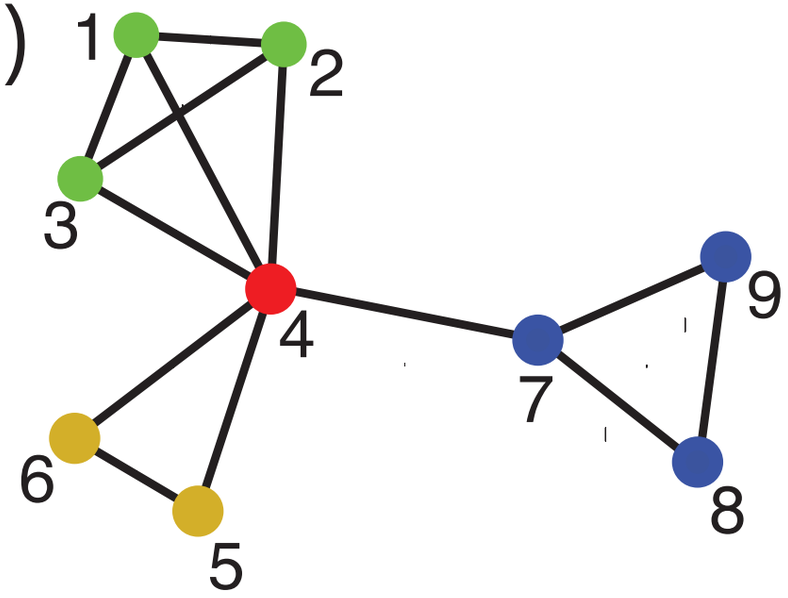
$$S(e_{ik}, e_{jk}) = \frac{|n_+(i) \cap n_+(j)|}{|n_+(i) \cup n_+(j)|}$$



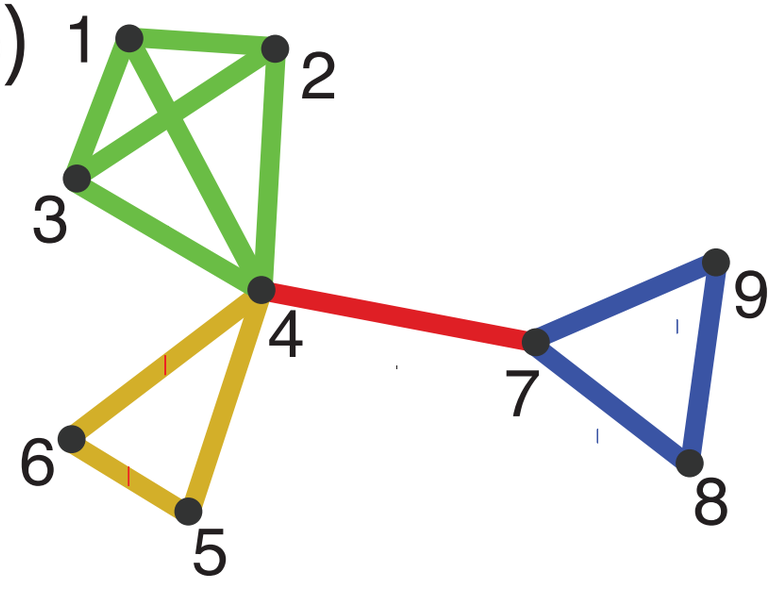
$$n_+(i) \equiv \{x \mid d(i,x) \leq 1\}$$

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

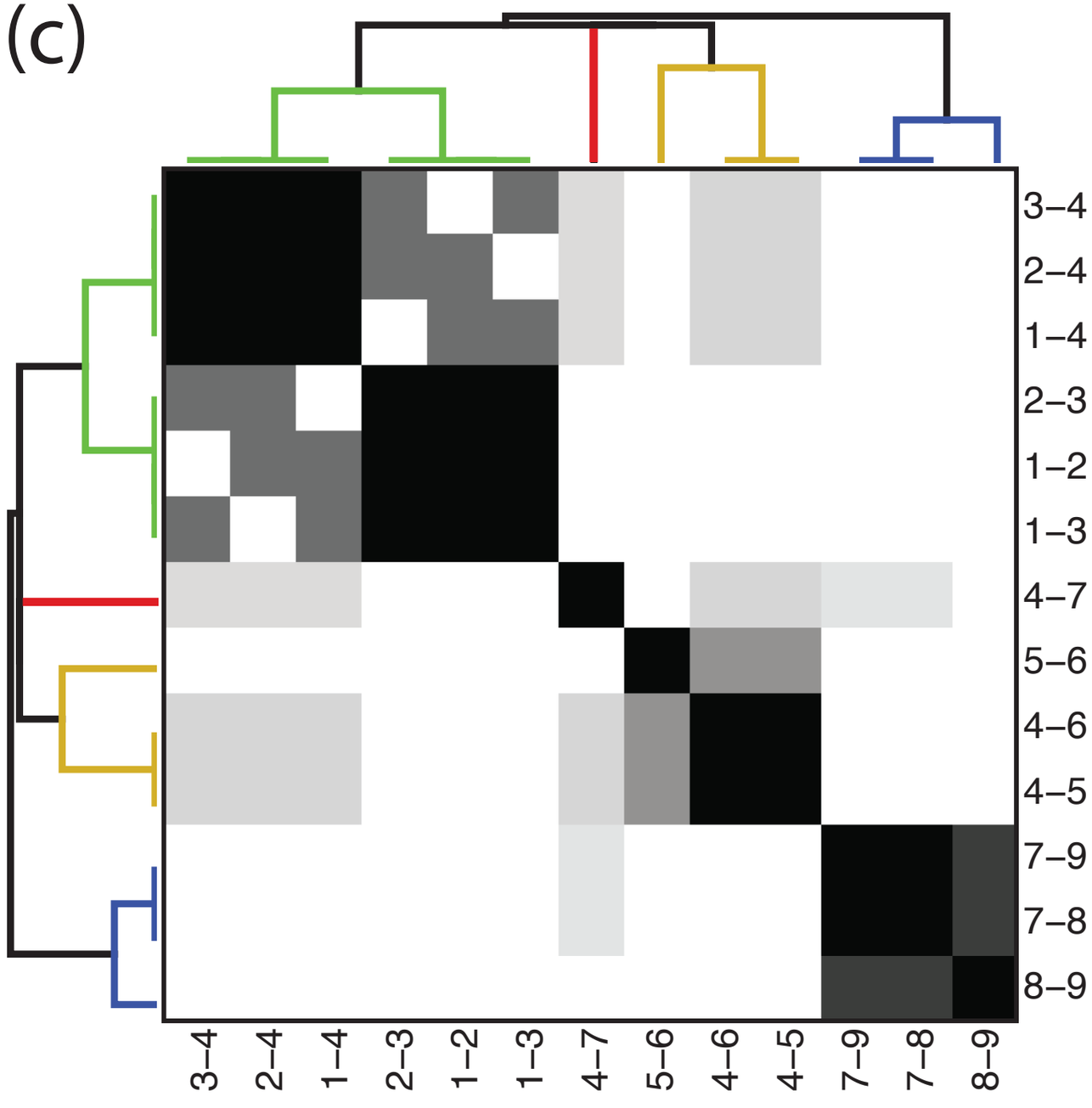
(a)



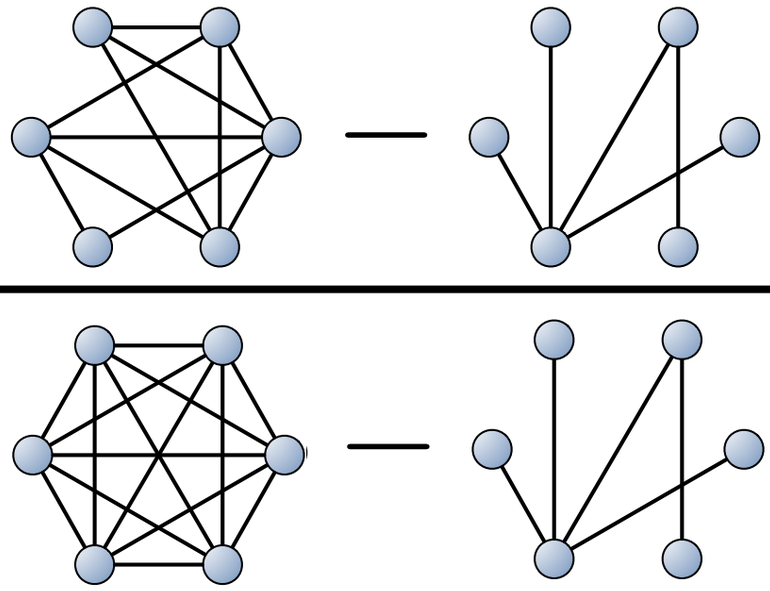
(b)



(c)



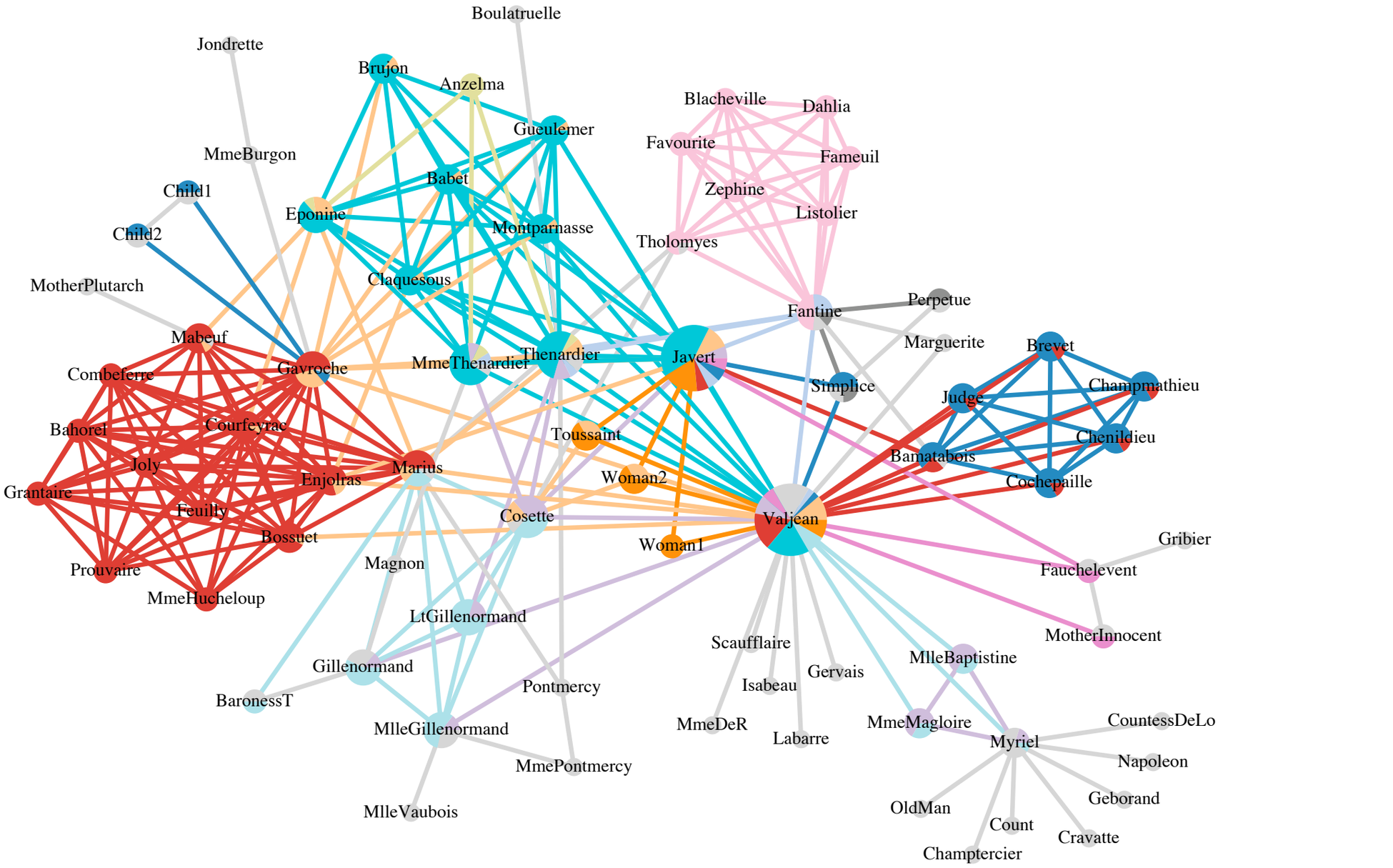
Partition Density

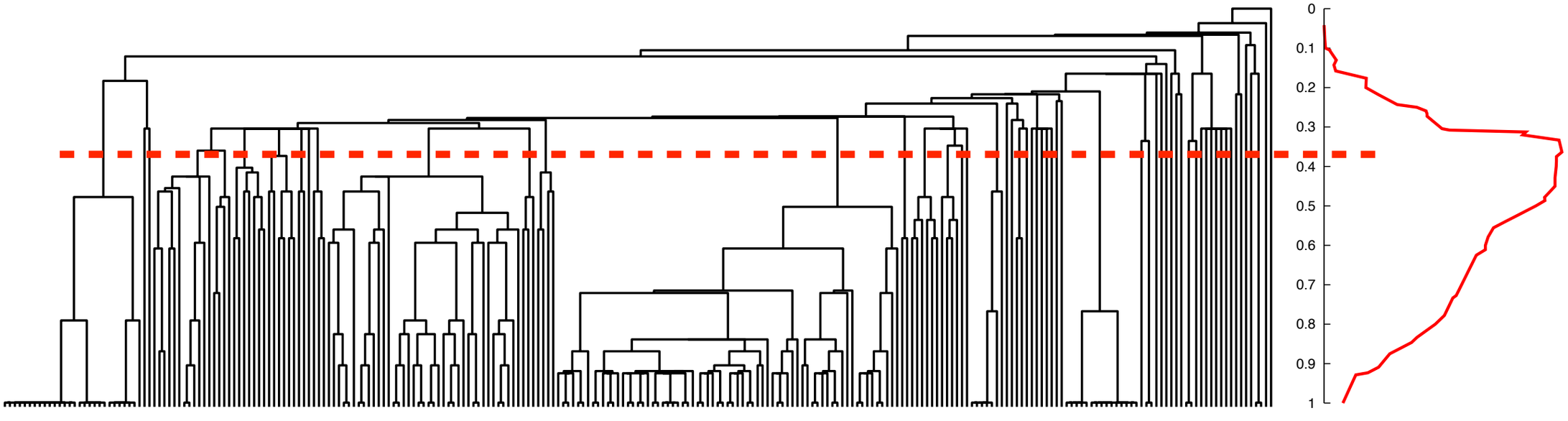


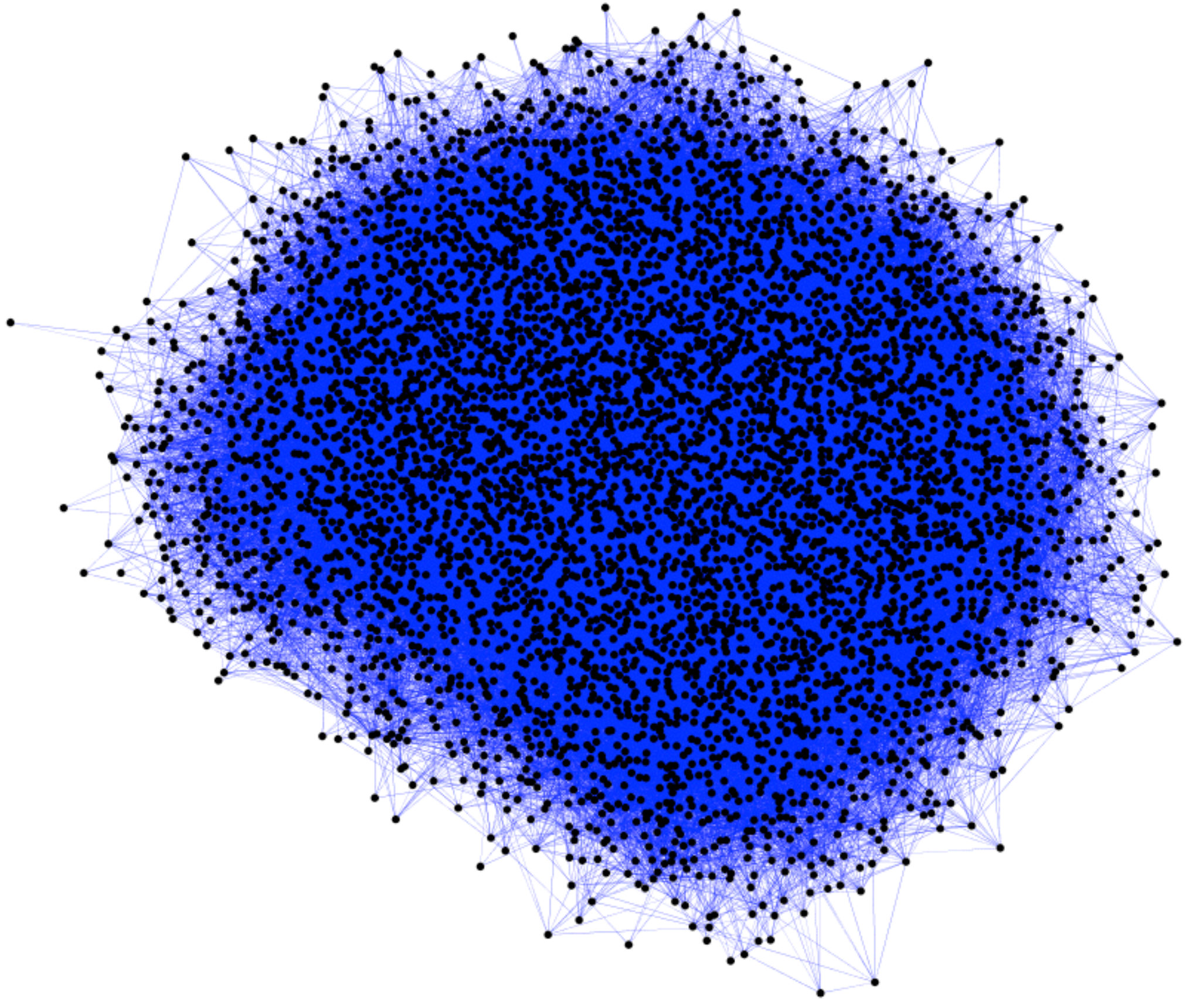
$$= \frac{m_c - (n_c - 1)}{\frac{n_c(n_c-1)}{2} - (n_c - 1)}$$

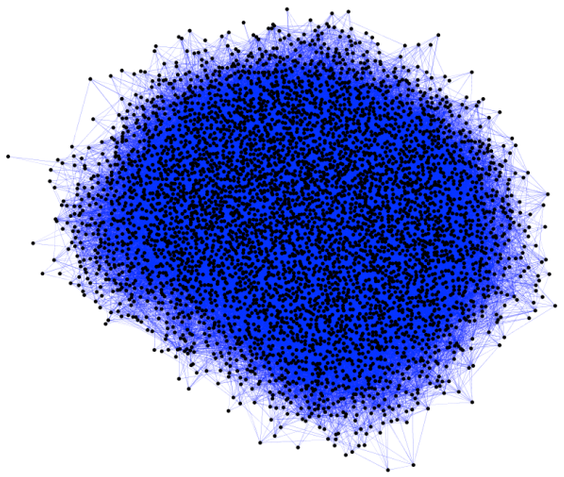
$$= 2\frac{m_c - (n_c - 1)}{(n_c - 2)(n_c - 1)}$$

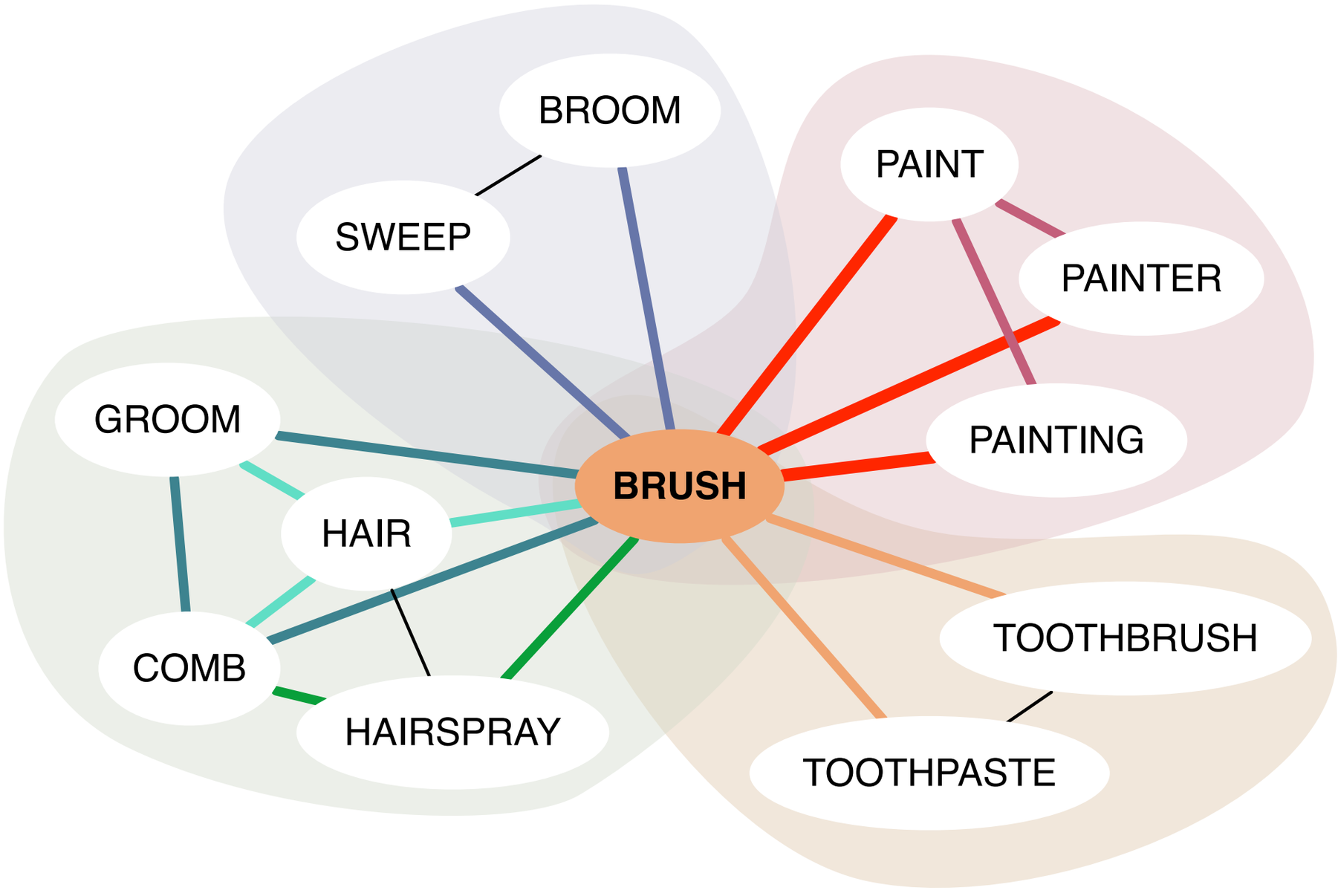
$$D \equiv \frac{2}{M} \sum_c m_c \frac{m_c - (n_c - 1)}{(n_c - 2)(n_c - 1)}$$

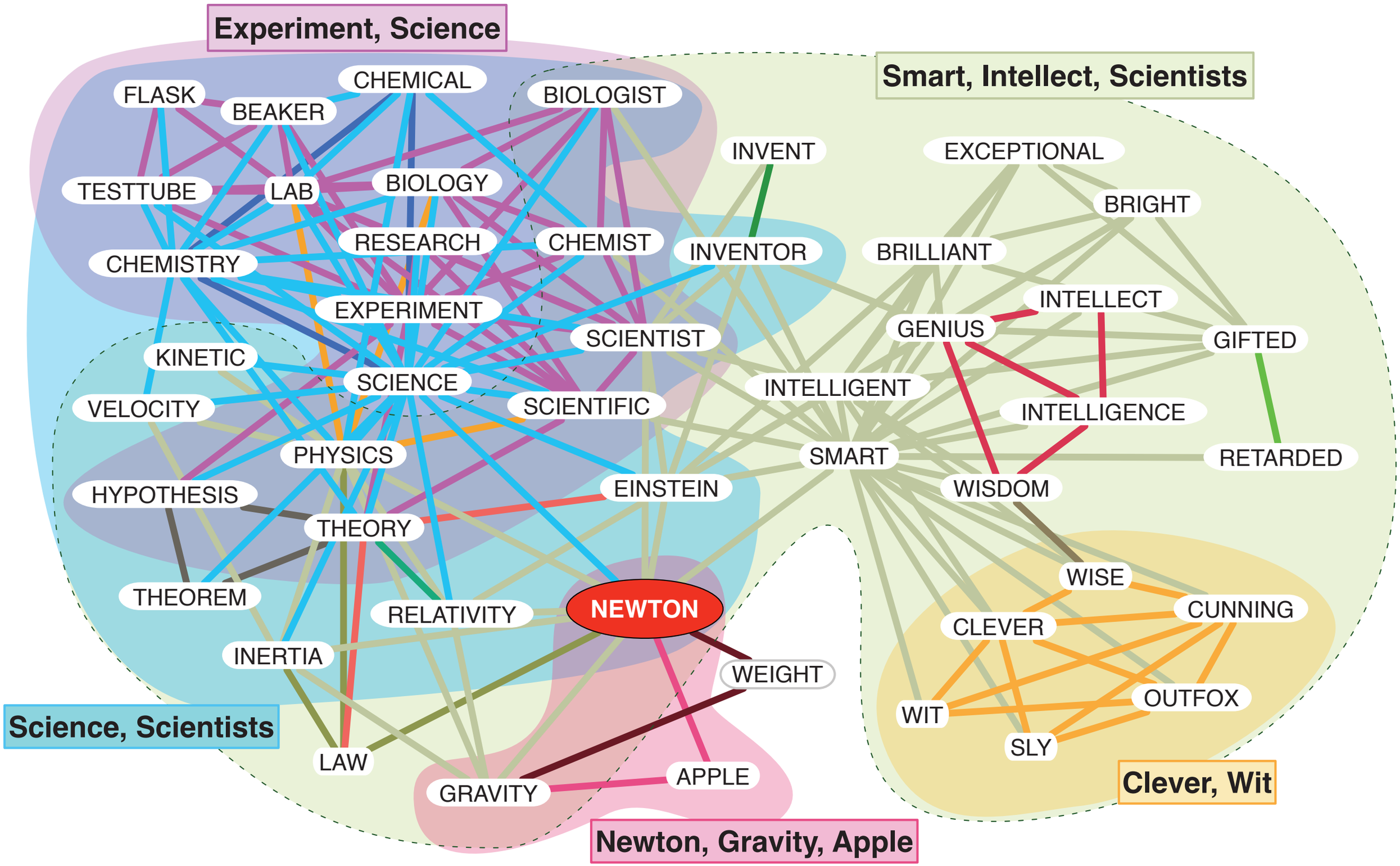

Boulatruelle
Jondrette
Brujon
Anzelma
Gueulemer
MmeBurgon
Babet
Child1
Eponine
Montparnasse
Child2
Claquesous
MotherPlutarch
Mabeuf
Gavroche
MmeThenardier
Thenardier
Javert
Combeferre
Courfeyrac
Bahorel
Joly
Enjolras
Marius
Grantaire
Feuilly
Bossuet
Prouvaire
MmeHucheloup
Toussaint
Woman2
Cosette
Woman1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


0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

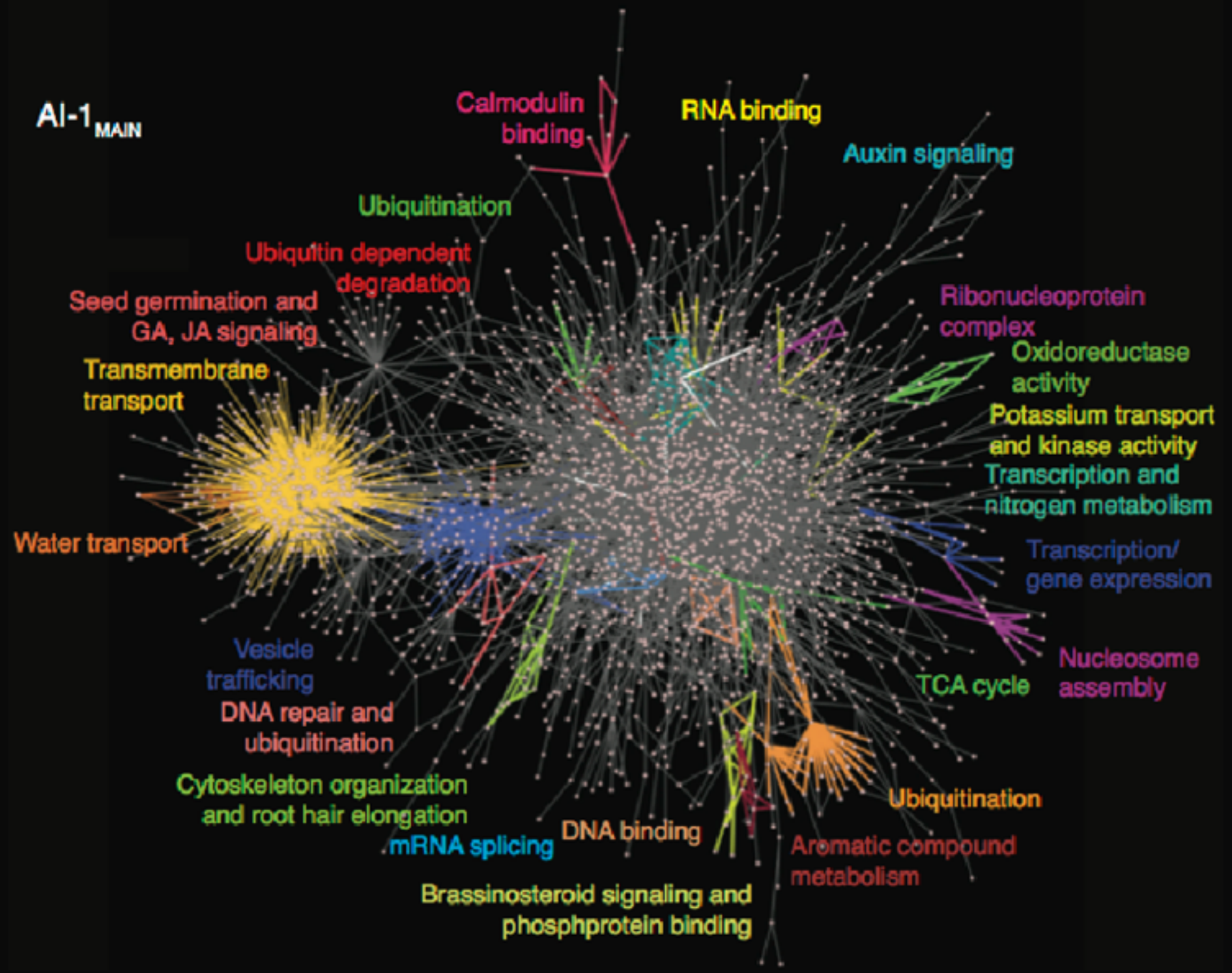





BROOM
SWEEP
PAINT
PAINTER
GROOM
PAINTING
BRUSH
HAIR
TOOTHBRUSH
COMB
HAIRSPRAY
TOOTHPASTE

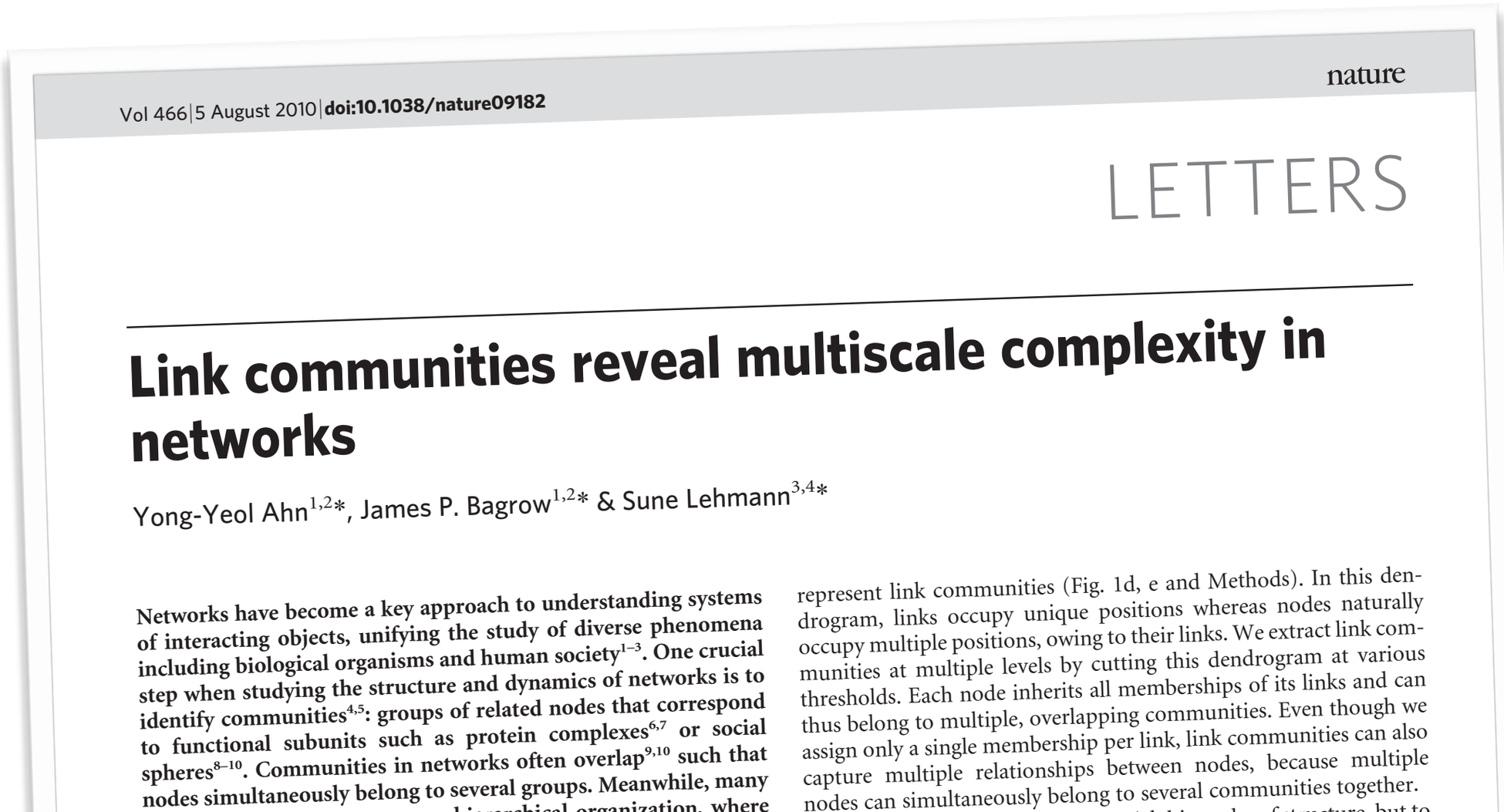

Experiment, Science
Smart, Intellect, Scientists
FLASK
BEAKER
CHEMICAL
BIOLOGIST
INVENT
EXCEPTIONAL
TESTTUBE
LAB
BIOLOGY
BRIGHT
RESEARCH
CHEMIST
INVENTOR
BRILLIANT
CHEMISTRY
INTELLECT
EXPERIMENT
GENIUS
GIFTED
SCIENTIST
KINETIC
SCIENCE
INTELLIGENT
VELOCITY
SCIENTIFIC
INTELLIGENCE
PHYSICS
SMART
RETARDED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

The first plant (genomic scale) interactome



Arabidopsis Interactome Mapping Consortium, *Science, 2011*

Relationship vs. entities



Vol 466|5 August 2010|doi:10.1038/nature09182

nature

LETTERS

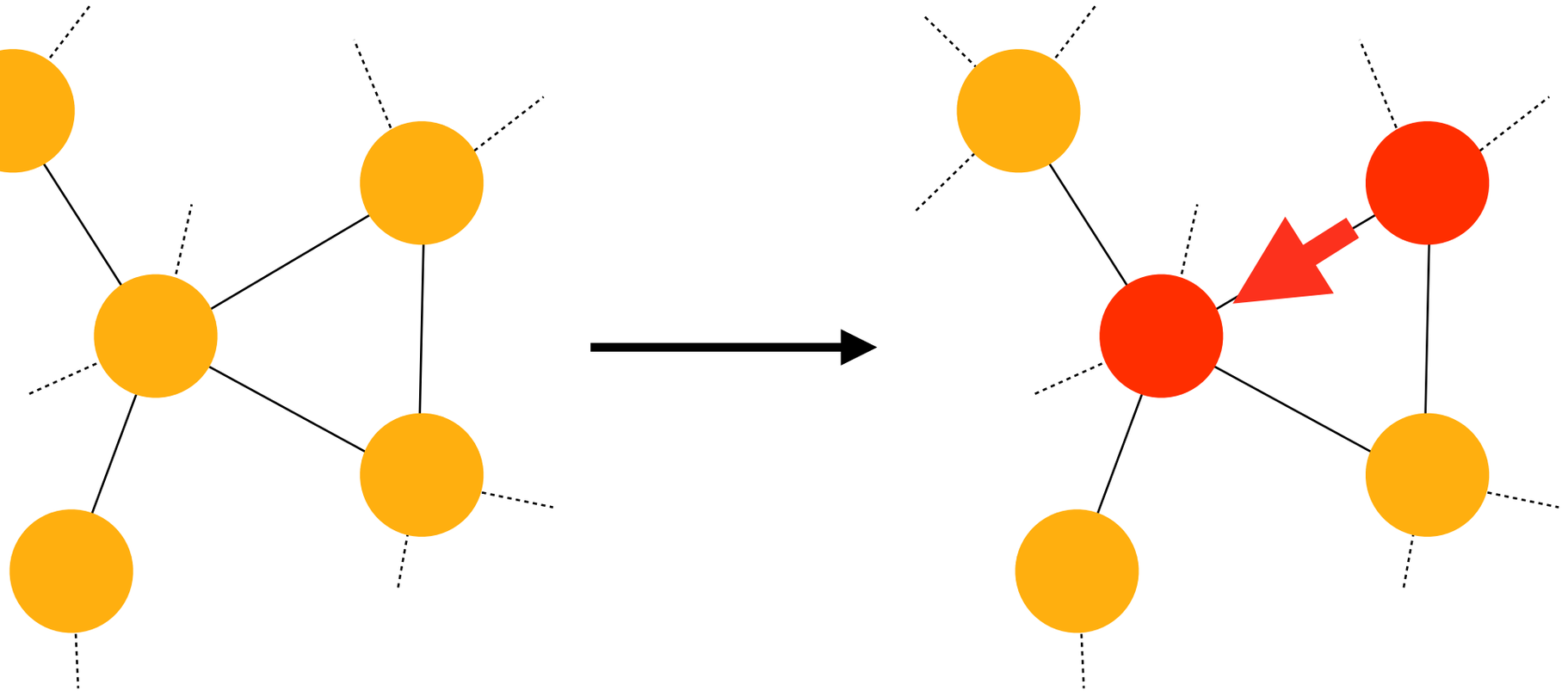
Link communities reveal multiscale complexity in networks

Yong-Yeol Ahn[1,2]*, James P. Bagrow[1,2]* & Sune Lehmann[3,4]*

Networks have become a key approach to understanding systems of interacting objects, unifying the study of diverse phenomena including biological organisms and human society[1–3]. One crucial step when studying the structure and dynamics of networks is to identify communities[4,5]: groups of related nodes that correspond to functional subunits such as protein complexes[6,7] or social spheres[8–10]. Communities in networks often overlap[9,10] such that nodes simultaneously belong to several groups. Meanwhile, many

represent link communities (Fig. 1d, e and Methods). In this dendrogram, links occupy unique positions whereas nodes naturally occupy multiple positions, owing to their links. We extract link communities at multiple levels by cutting this dendrogram at various thresholds. Each node inherits all memberships of its links and can thus belong to multiple, overlapping communities. Even though we assign only a single membership per link, link communities can also capture multiple relationships between nodes, because multiple nodes can simultaneously belong to several communities together.

#2 Communities and Contagion



Cultural diffusion = Infectious diseases?

GENERALIZATION OF EPIDEMIC THEORY

AN APPLICATION TO THE TRANSMISSION OF IDEAS

By Dr. WILLIAM GOFFMAN
Center of Documentation and Communication Research, School of Library Science, Western Reserve University

AND

Dr. VAUN A. NEWILL
School of Medicine, Western Reserve University, Cleveland, Ohio

ONE of the most fundamental problems in the field of information retrieval is that of determining the circumstances under which it might be necessary to introduce an information retrieval system as an aid to a given population of scientists. It is proposed that this problem be examined in terms of the transmission and development of ideas within a population. Specifically, the transmission of ideas within a population will be treated as if it were the transmission of an infectious disease, that is, in terms of an epidemic process. An attempt will be made to indicate the role of information retrieval in the development of such a process.

The Epidemic Model

Since the spread of disease in a population is to be our model for the transmission of ideas, it is appropriate to discuss the essential principles pertinent to this issue. These principles are a part of epidemiology. The necessary elements involved in the process of the spread of an infectious disease are those of : (1) a specified population ; (2) an exposure to infectious material. The

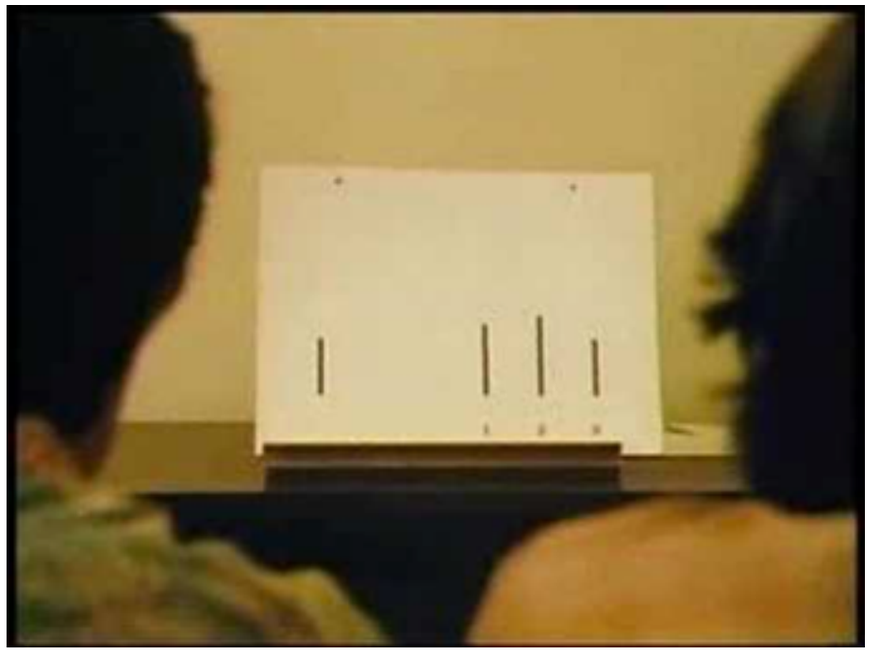
Transmission of Ideas as an 'Epidemic' Process

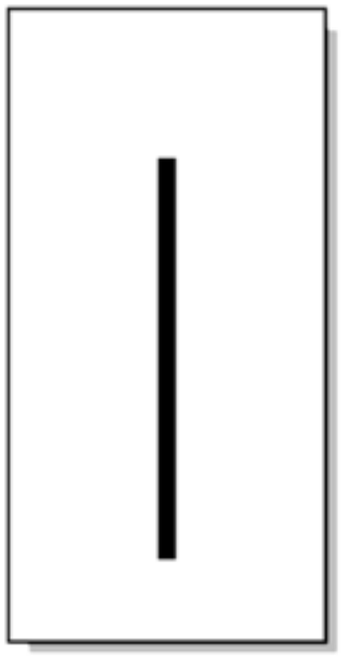
In general, the 'epidemic' process can be characterized as one of transition from one state (susceptible) to another (infective) where the transition is caused by exposure to some phenomenon (infectious material). The process need not be restricted to infectious disease but is a more general abstract process that might be applied to many situations. All that is needed is the appropriate interpretation of the process elements, that is, susceptibles, infectives, removals, infectious material, intermediary host, latency period, disease, etc.

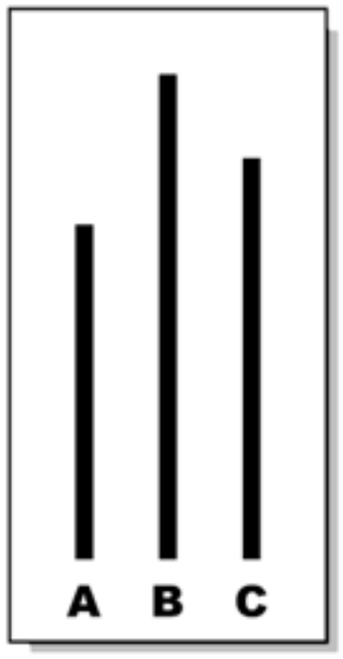
People are susceptible to certain ideas and resistant to others. Once an individual is infected with an idea he may in turn, after some period of time, transmit it to others. Such a process can result in an intellectual 'epidemic' (Table 1). For example, consider the development of psychoanalysis in the early part of this century. Freud was no less host to the infectious material of the 'disease' of psychoanalysis than the person carrying the organism capable of transmitting a cold, nor is his writing less of a 'vector' carrying the 'infectious material' than the mos-

Maybe not

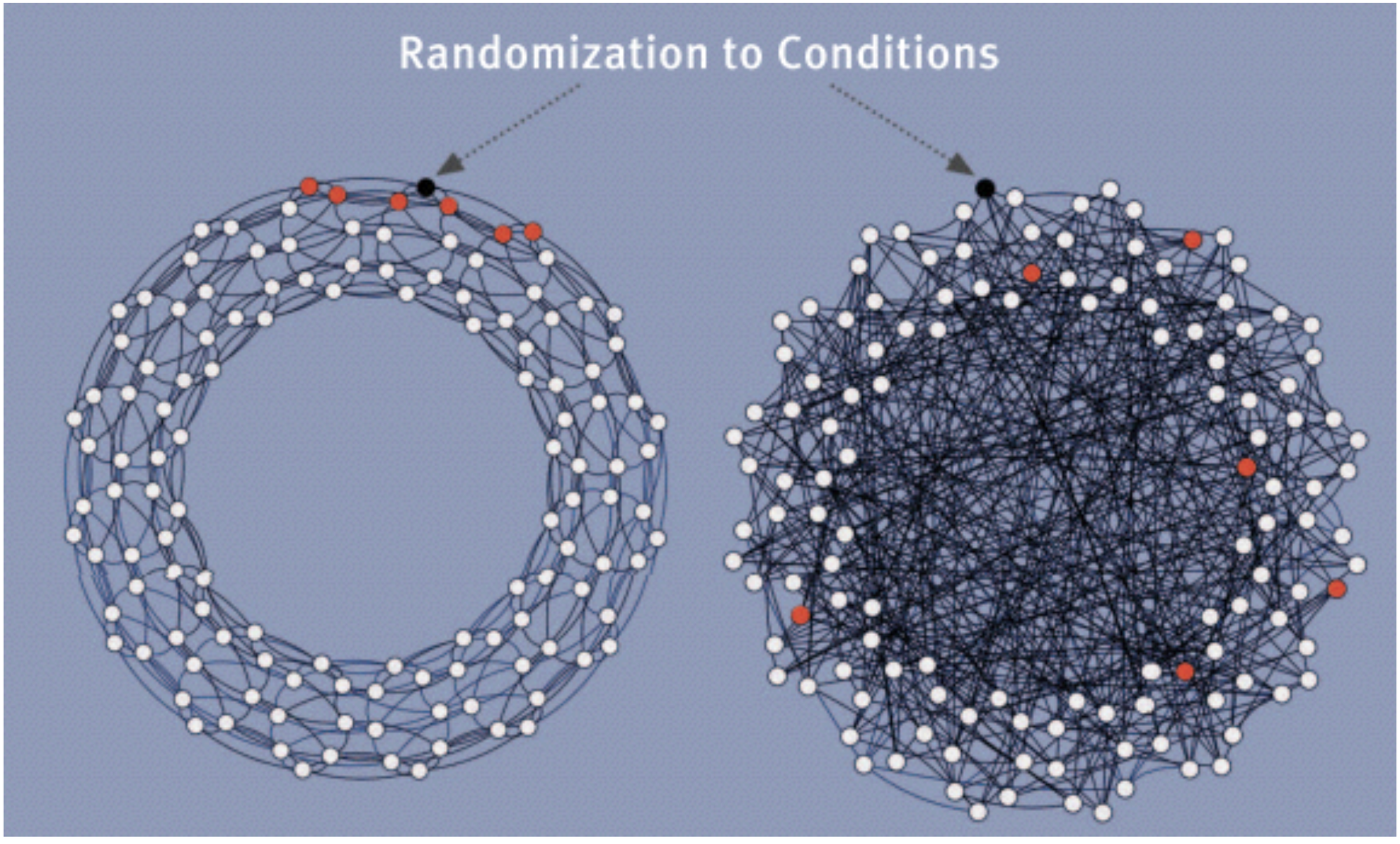
Asch conformity experiment







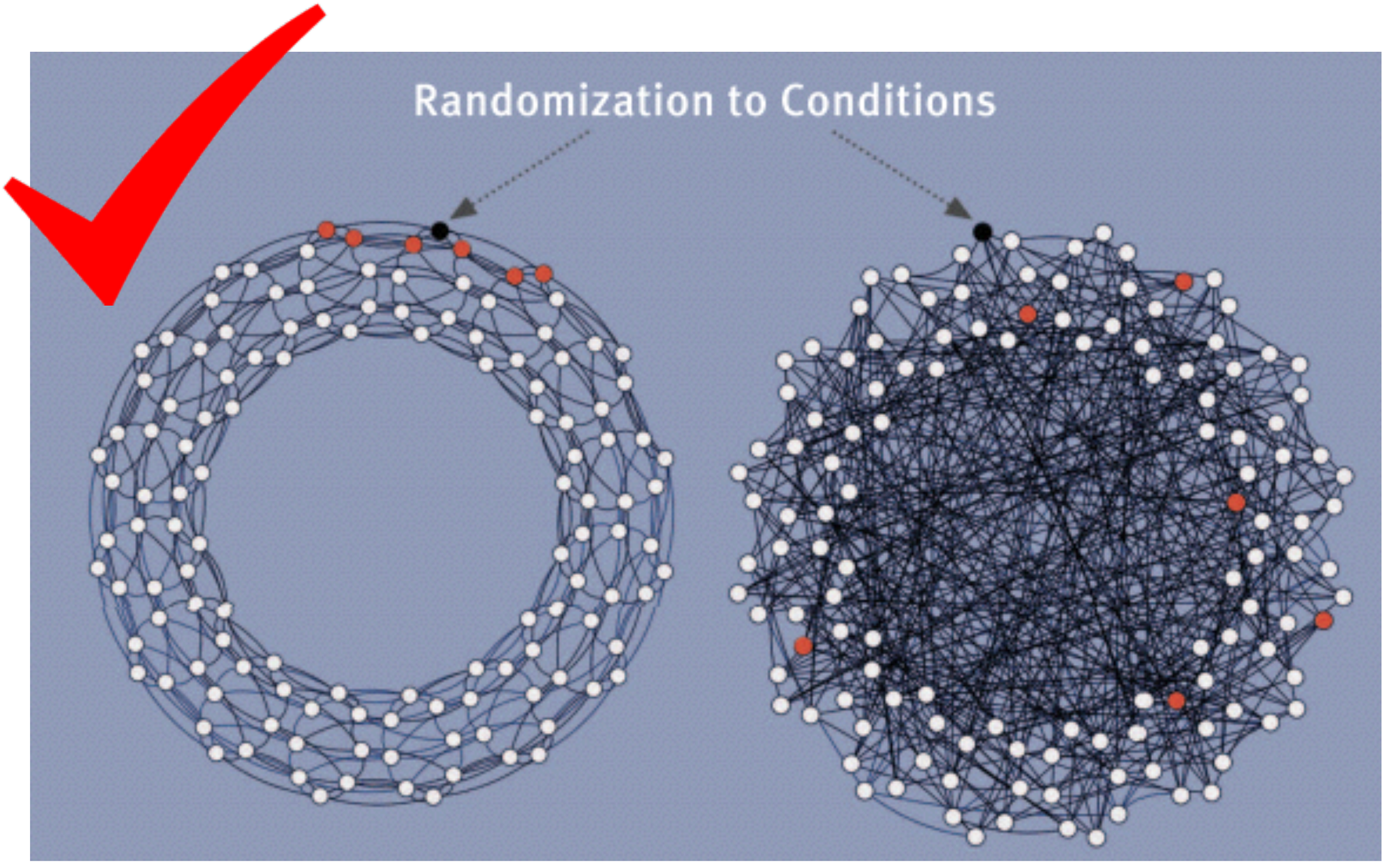
More people exert greater pressure



"Large"

"Small"

Centola, Science, 2010

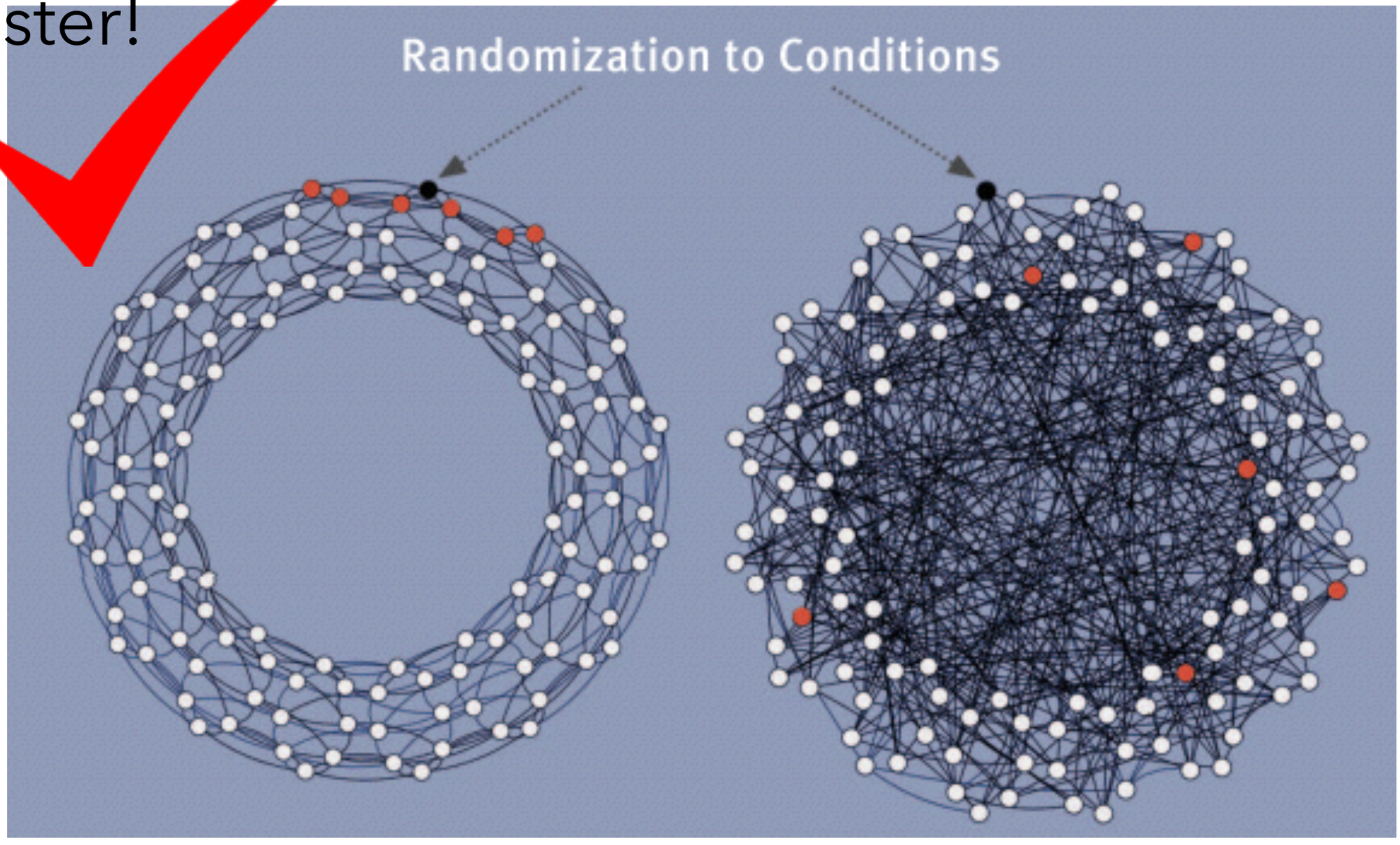


"Large"

"Small"

Centola, Science, 2010

Faster!



"Large"

"Small"

Centola, Science, 2010

Complex Contagion

三人成虎

三人成虎

three

三人成虎

three people

construct

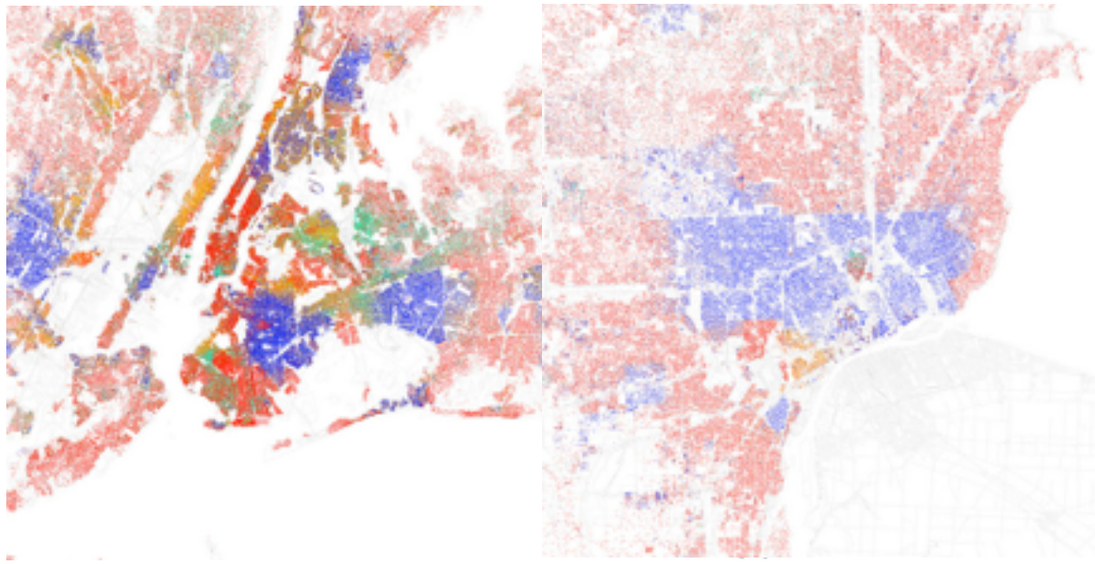
三人成虎

three people

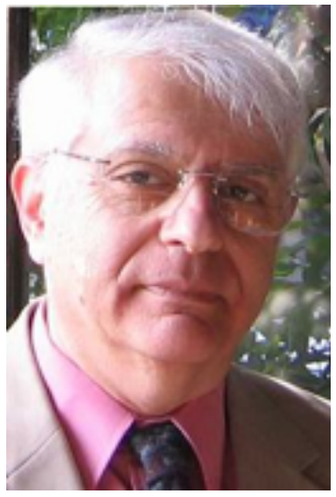
construct

三人成虎

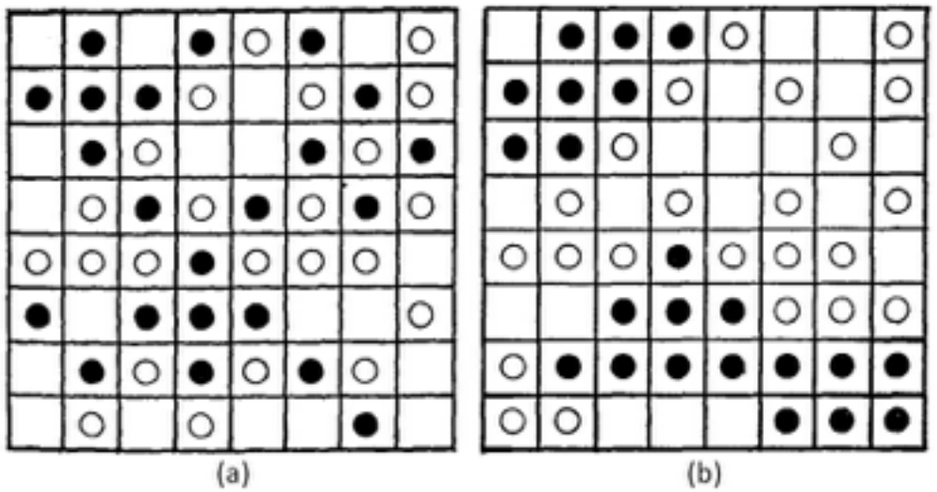
three people tiger



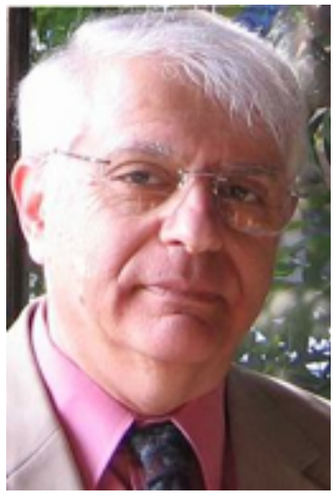
Thomas Schelling



Mark Granovetter



Thomas Schelling



Mark Granovetter

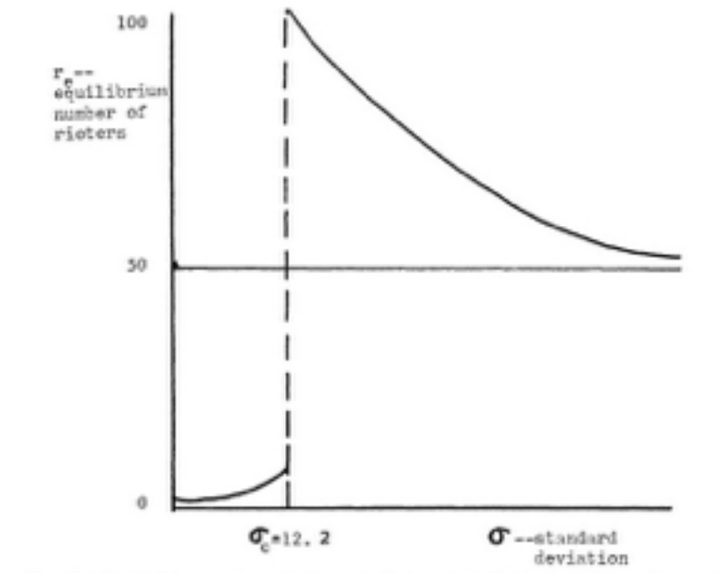
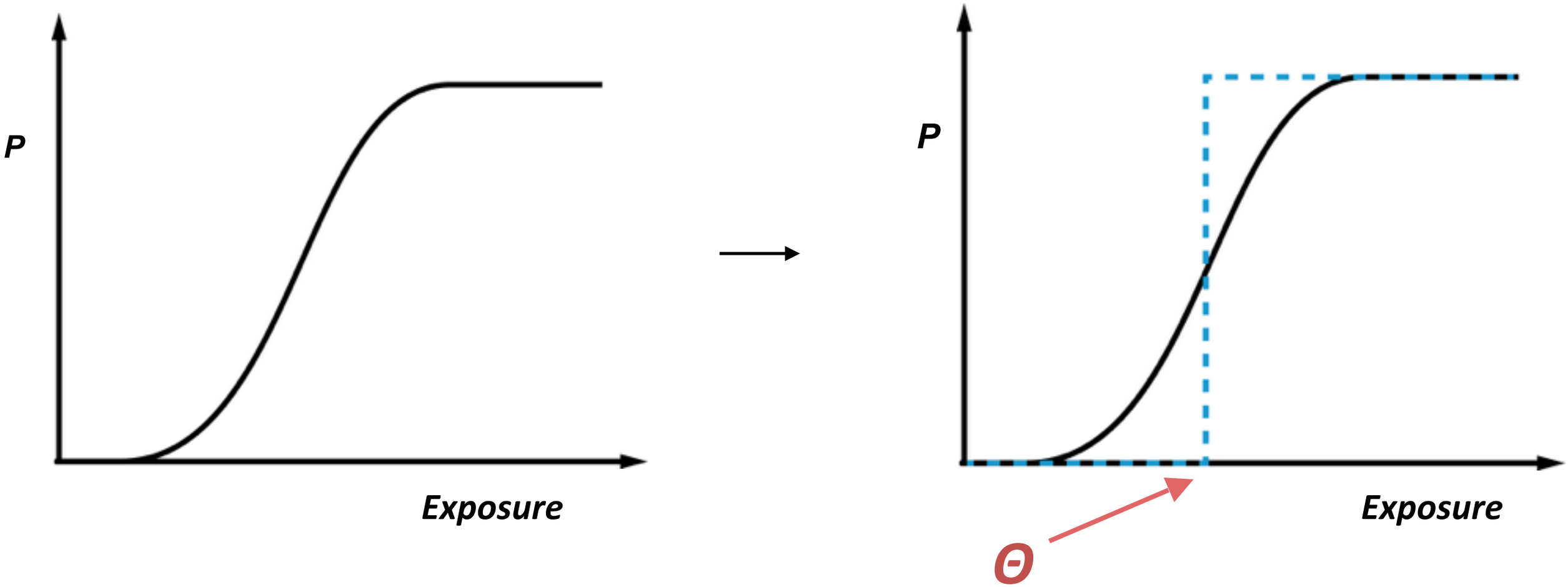
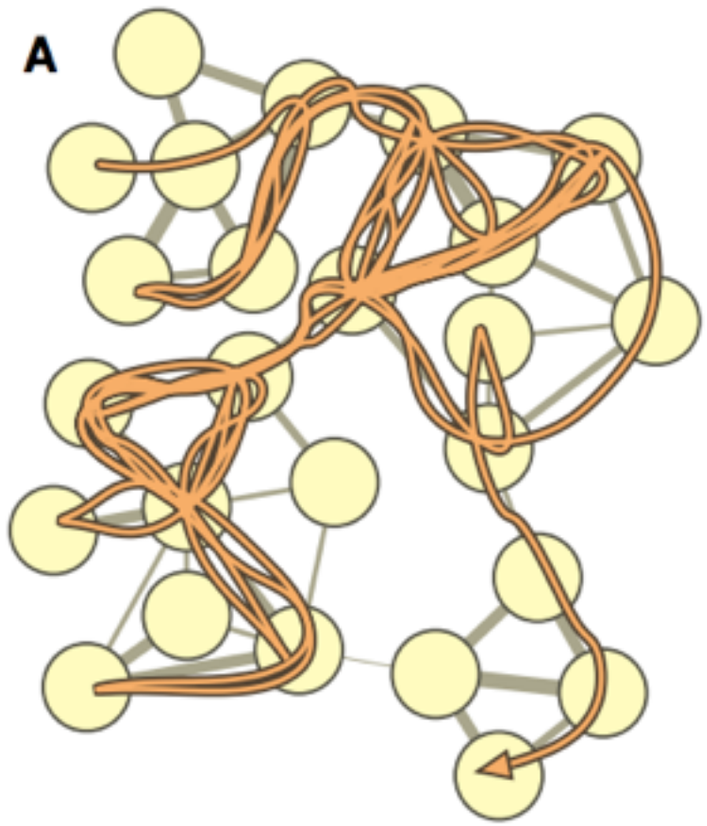


FIG. 2.—Equilibrium number of rioters plotted against standard deviation of normal distributions of thresholds with mean = 25, N = 100.

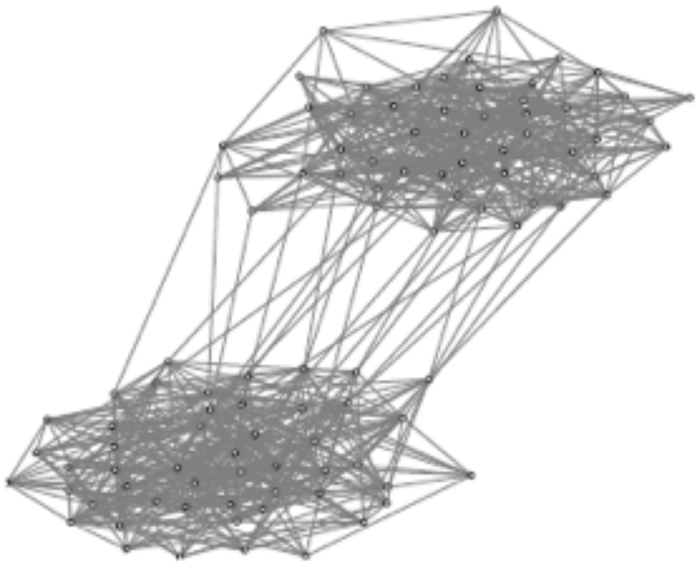
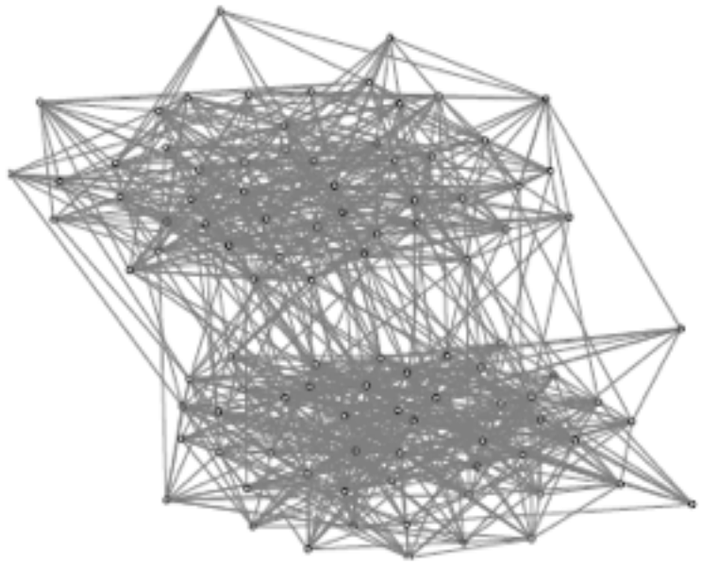
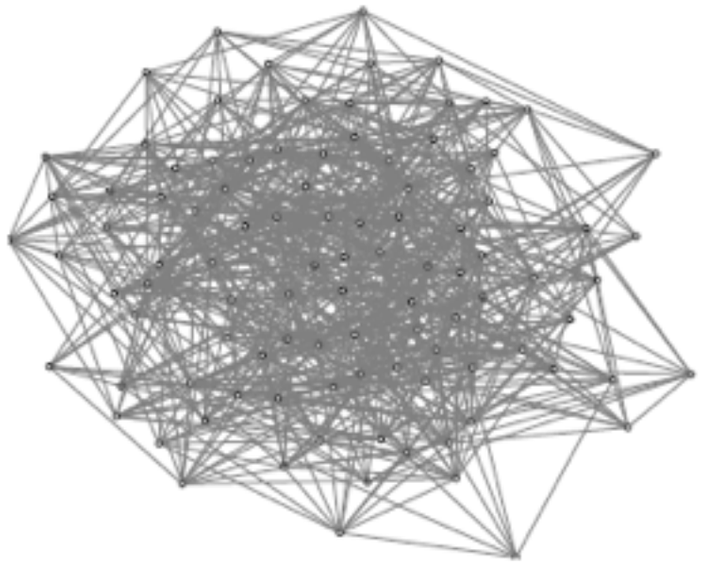
‘Threshold’ model

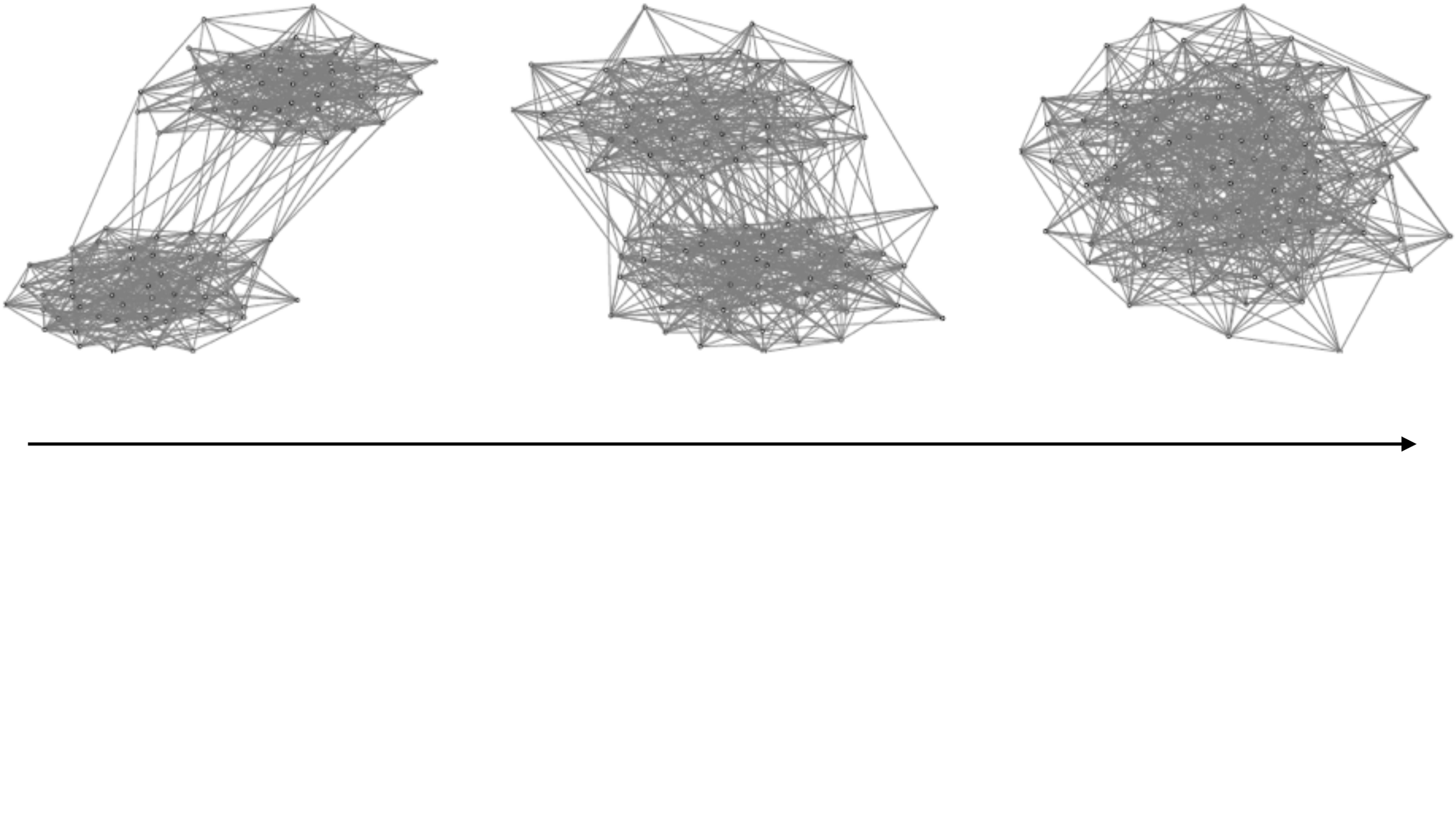


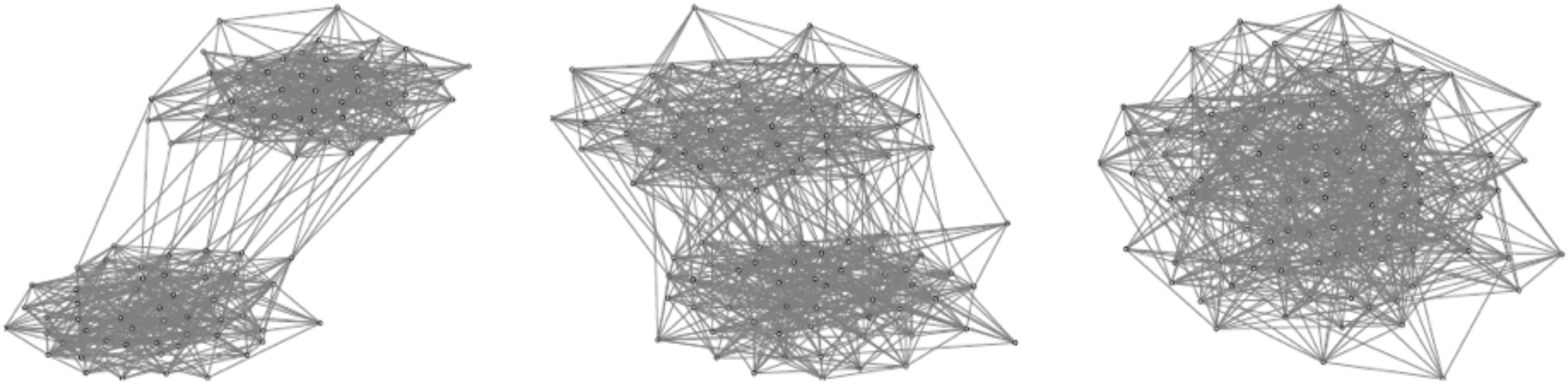
So how communities affect complex contagion?



"Of course communities should **trap** contagion"

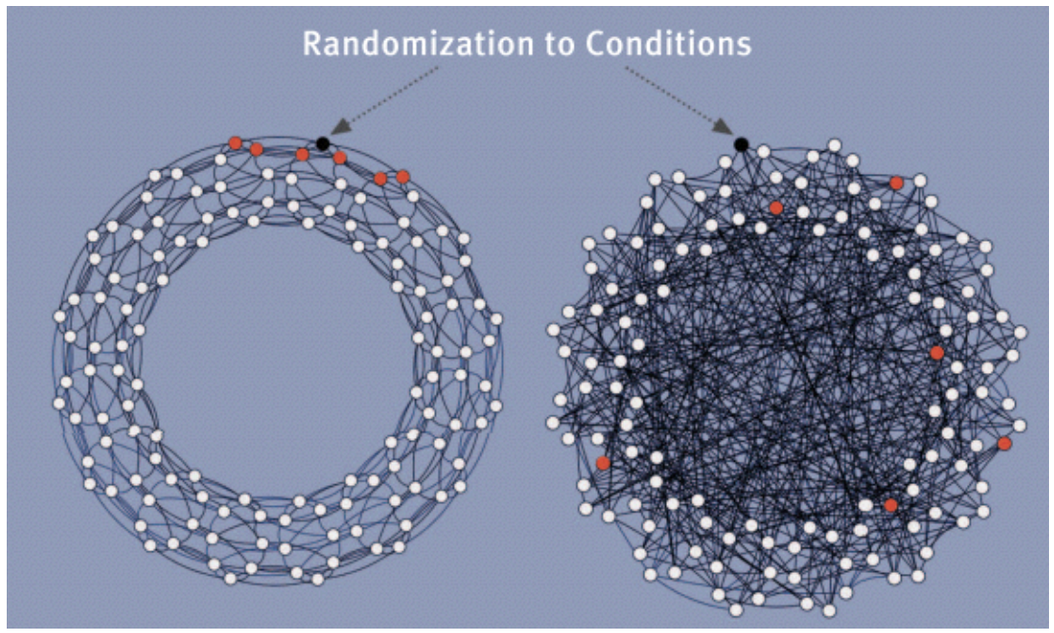




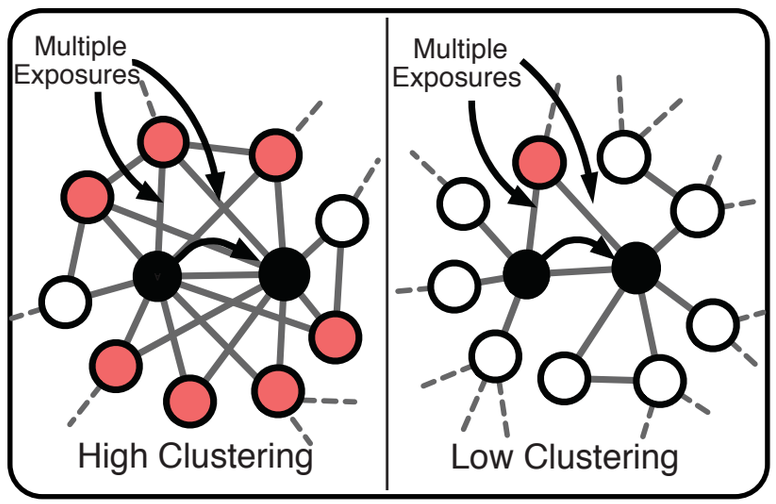




Better diffusion?

Not necessarily

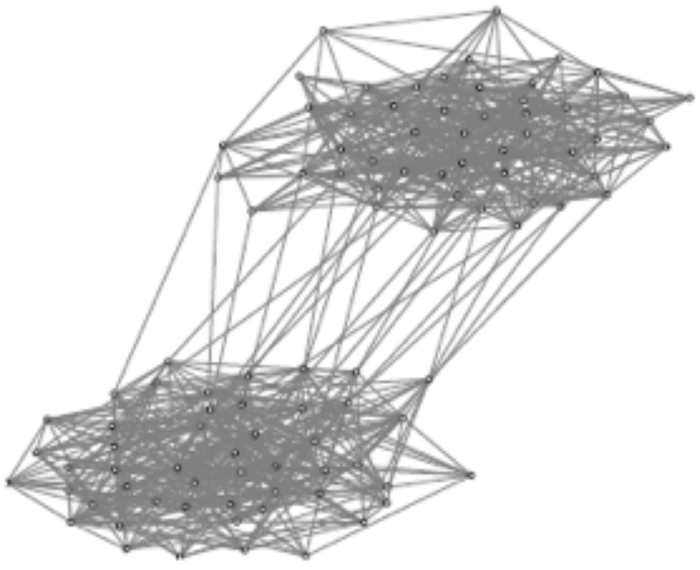
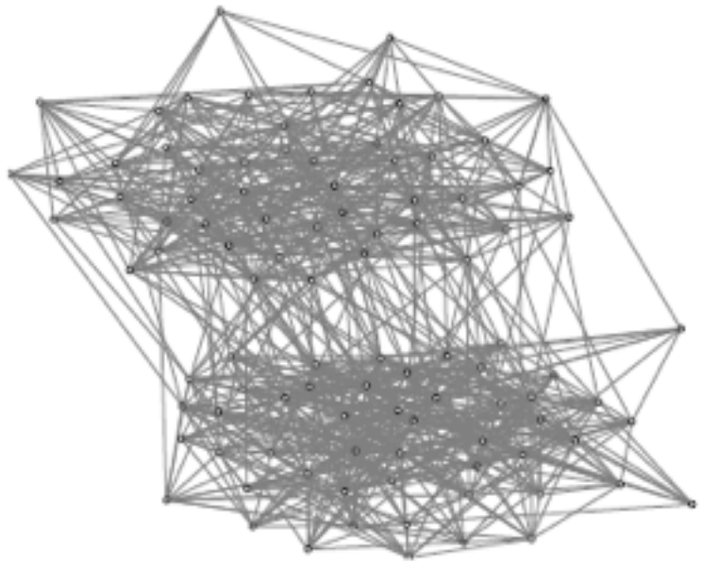
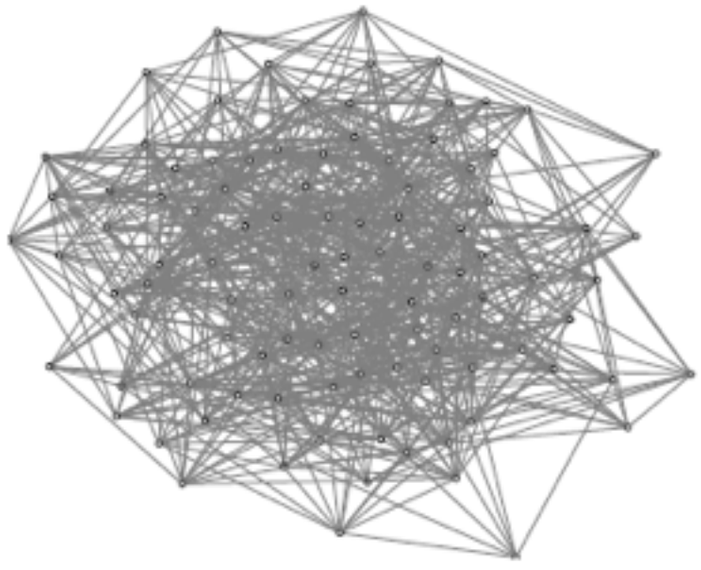




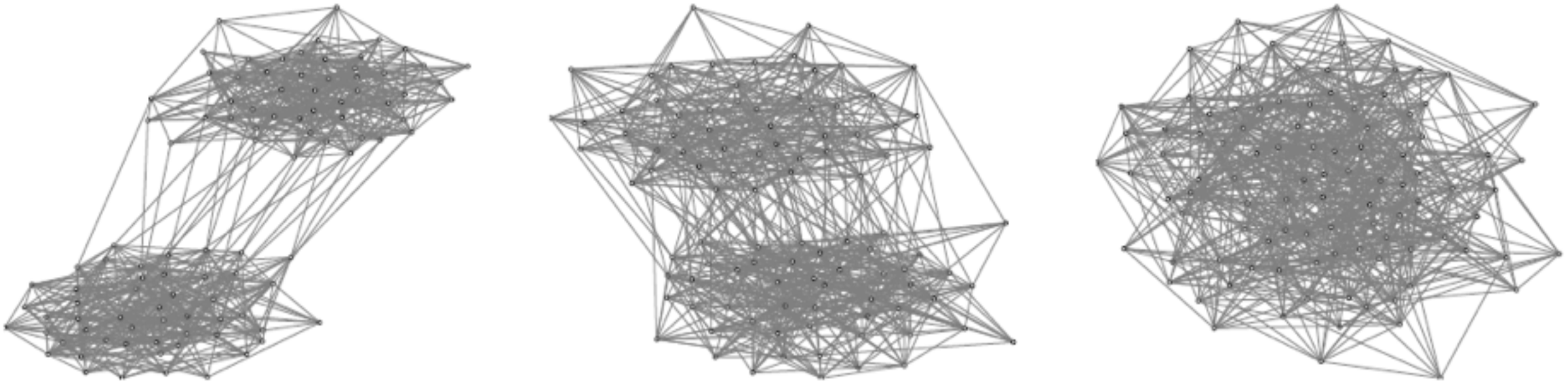
Communities,

while **hindering** between-community spreading,

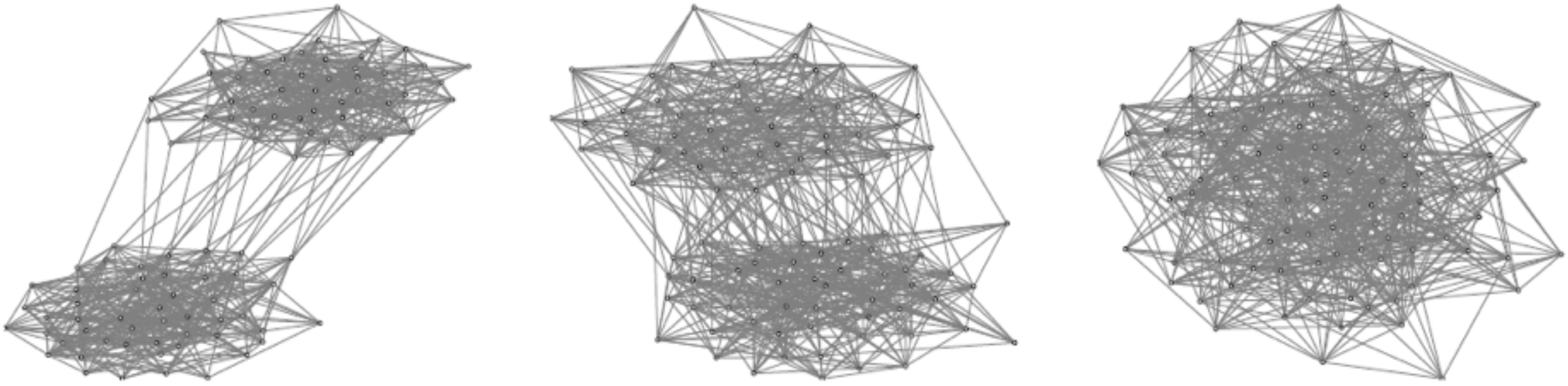
can **enhance** within-community spreading

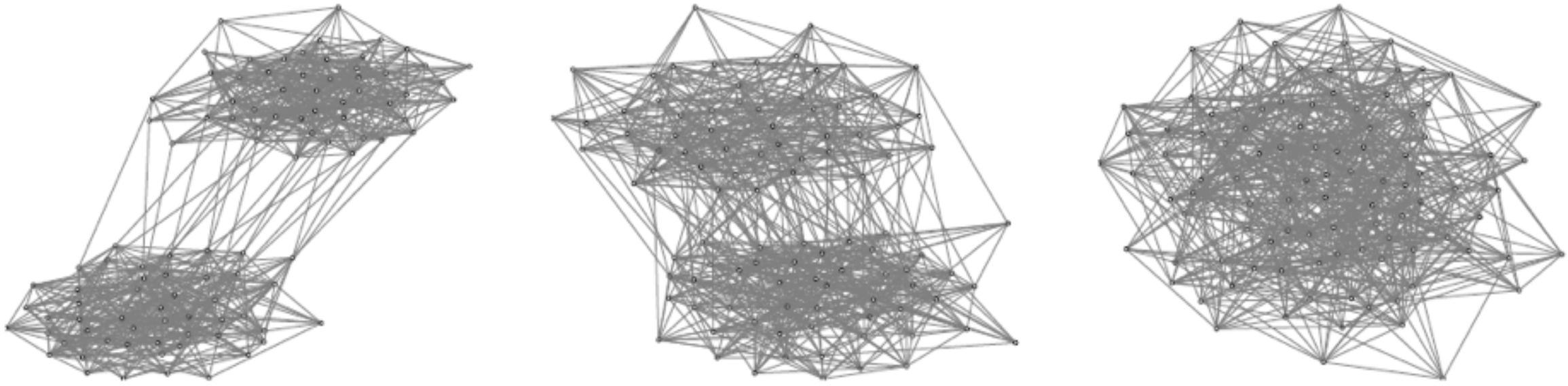
Better **inter-community** spreading



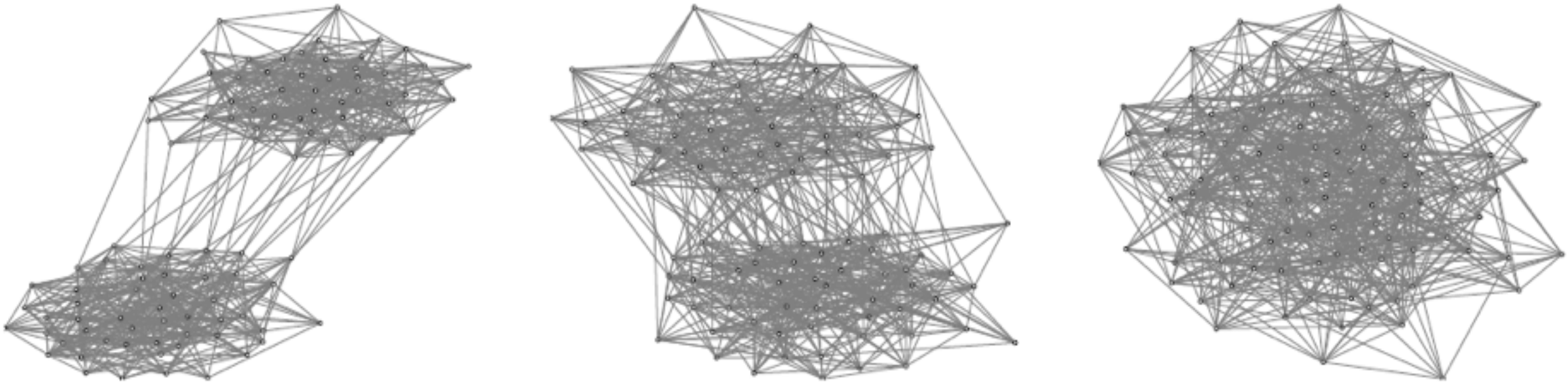
Better **inter-community** spreading



Better **inter-community** spreading

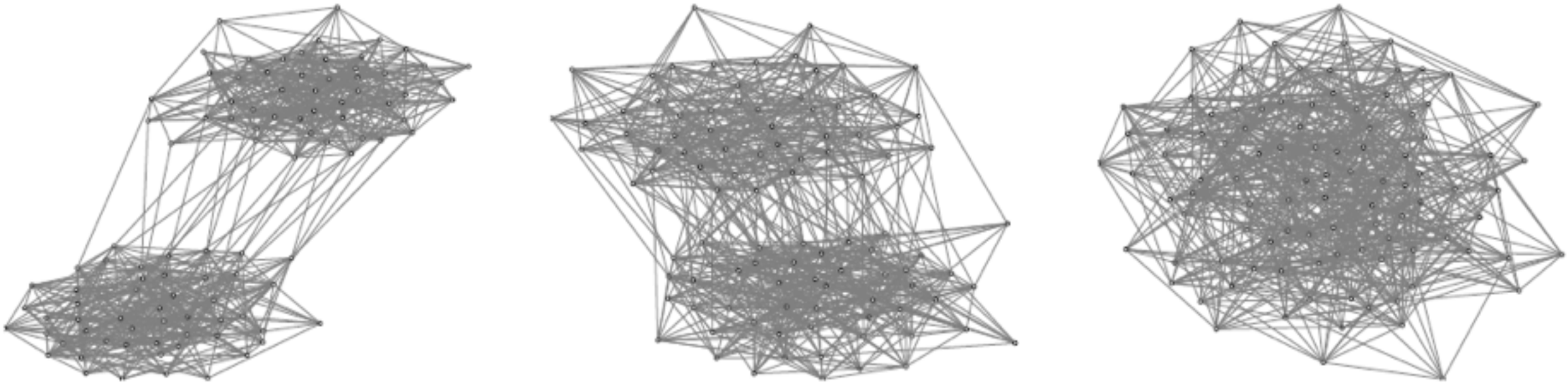


Better **inter**-**community**
spreading



Better **intra**-**community**
spreading

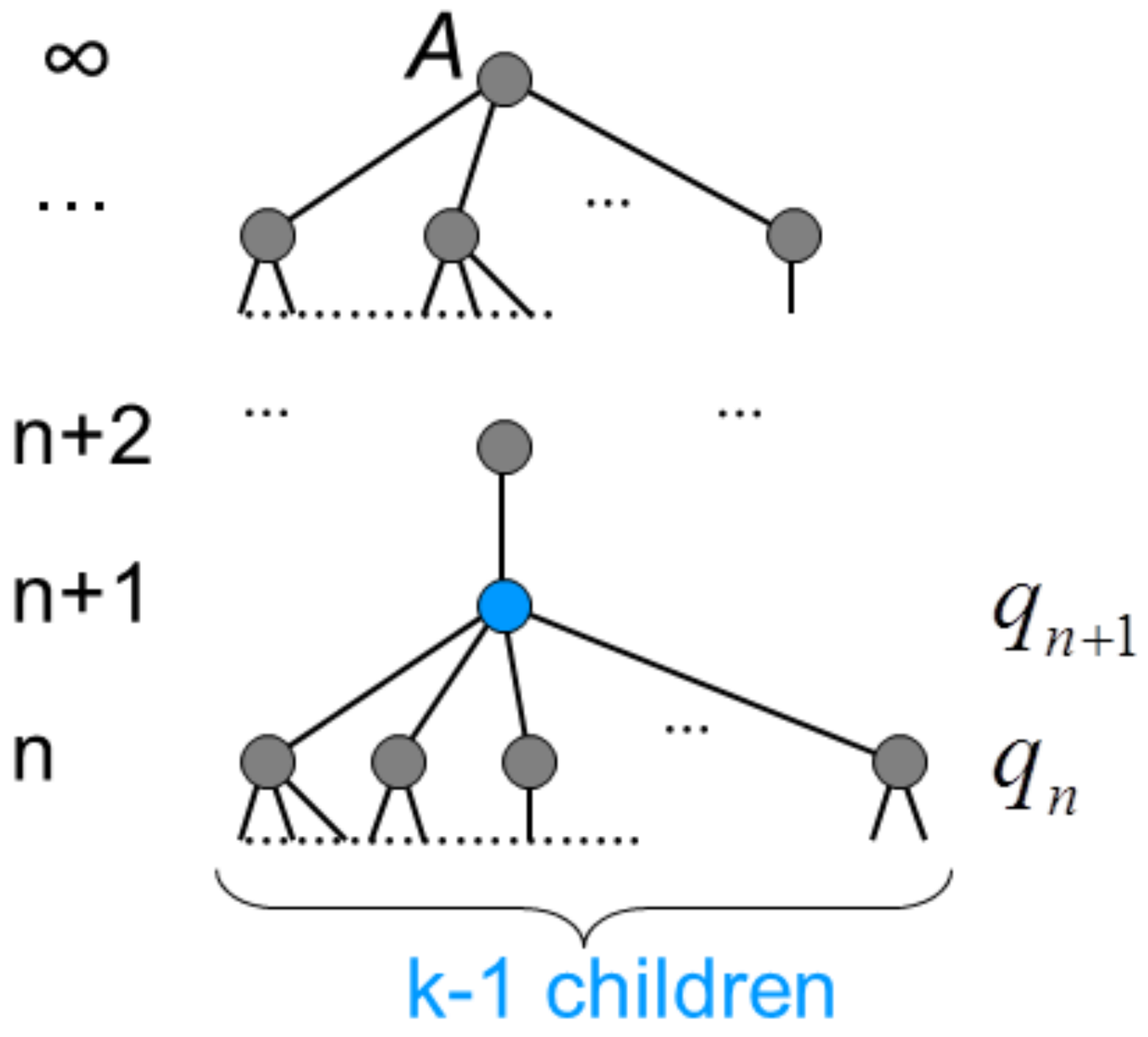
Better **inter**-**community** spreading



Better **intra**-**community** spreading

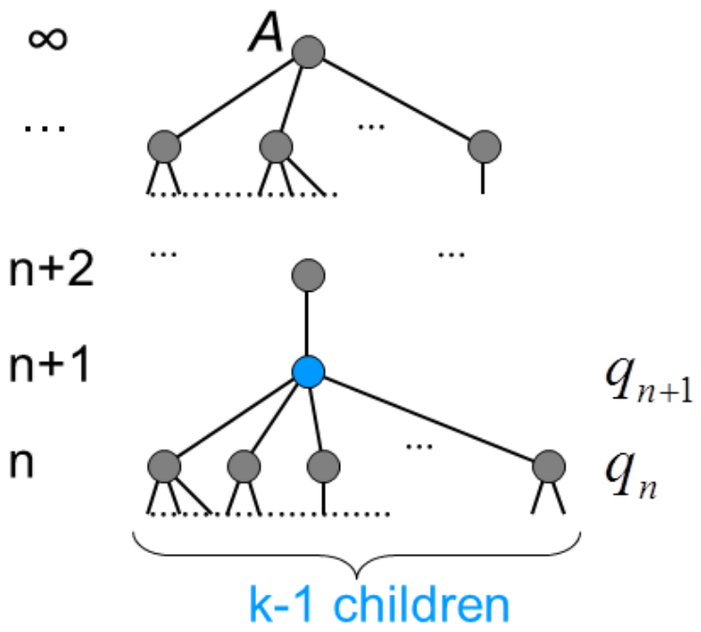
Optimal?

Tree Like Approximation (TL)



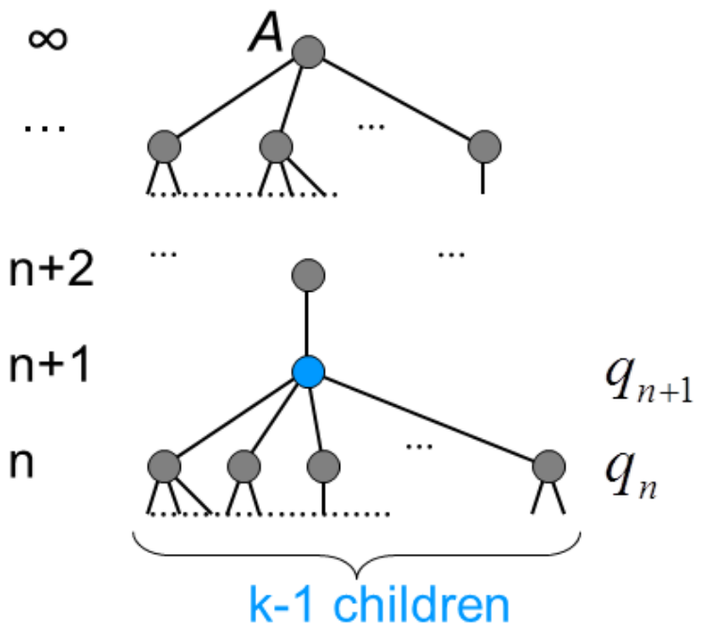
Gleeson, 2011.

Tree Like Approximation (TL)



$$q_{n+1} = \rho_0 + (1-\rho_0)\sum_{k=1}^{\infty} \frac{k}{z} p_k \sum_{m=0}^{k-1} \binom{k-1}{m} q_n^m (1-q_n)^{k-1-m} F(m,k)$$

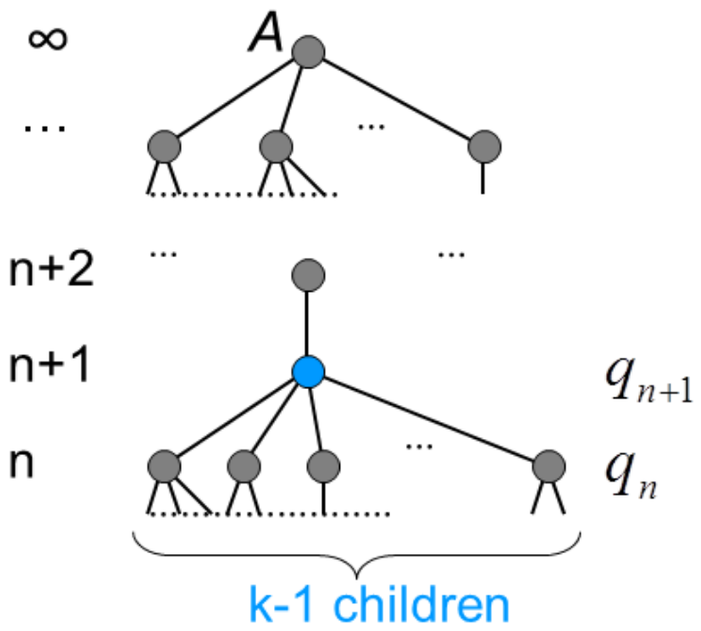
Tree Like Approximation (TL)



$$q_{n+1} = \rho_0 + (1-\rho_0)\sum_{k=1}^{\infty} \frac{k}{z} p_k \sum_{m=0}^{k-1} \binom{k-1}{m} q_n^m (1-q_n)^{k-1-m} F(m,k)$$

seed

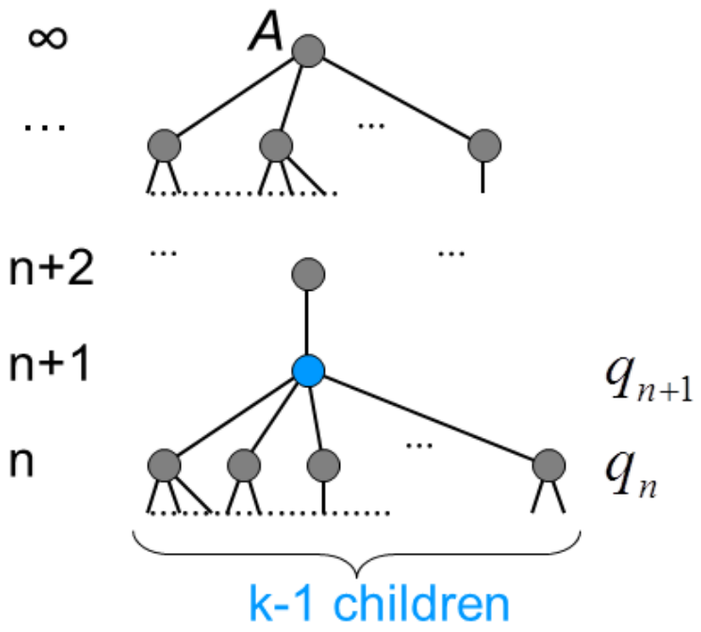
Tree Like Approximation (TL)



$$q_{n+1} = \rho_0 + (1-\rho_0)\sum_{k=1}^{\infty} \frac{k}{z} p_k \sum_{m=0}^{k-1} \binom{k-1}{m} q_n^m (1-q_n)^{k-1-m} F(m,k)$$

seed non-seeds

Tree Like Approximation (TL)

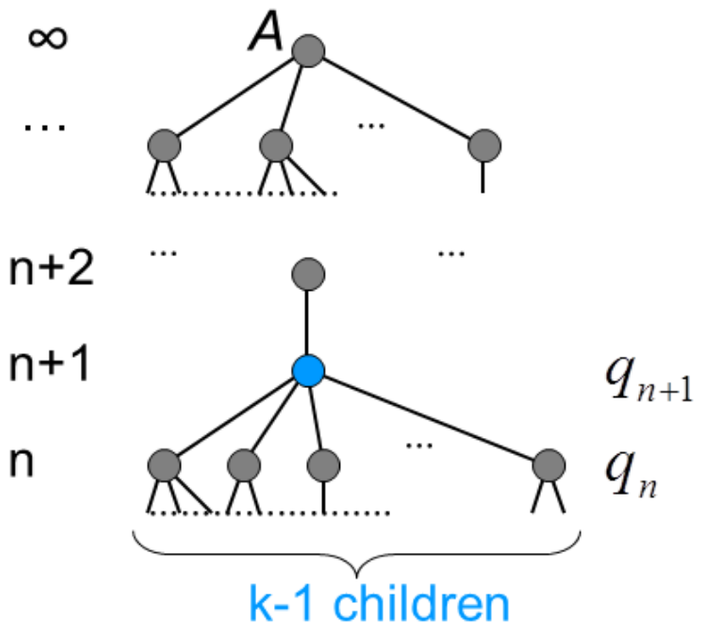


Excess degree distribution

$$q_{n+1} = \rho_0 + (1-\rho_0)\sum_{k=1}^{\infty} \frac{k}{z} p_k \sum_{m=0}^{k-1} \binom{k-1}{m} q_n^m (1-q_n)^{k-1-m} F(m,k)$$

seed non-seeds

Tree Like Approximation (TL)



Excess degree distribution

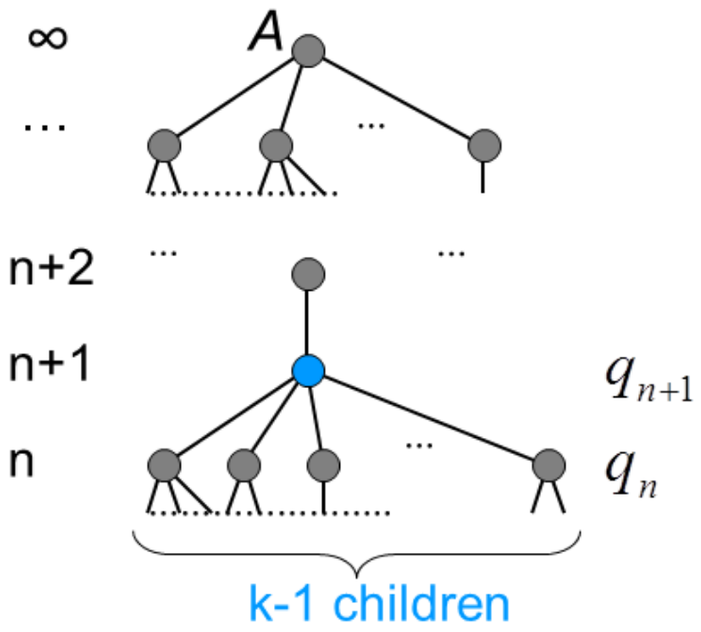
$$q_{n+1} = \rho_0 + (1-\rho_0)\sum_{k=1}^{\infty} \frac{k}{z} p_k \sum_{m=0}^{k-1} \binom{k-1}{m} q_n^m (1-q_n)^{k-1-m} F(m,k)$$

seed

non-seeds

For m larger than the threshold,

Tree Like Approximation (TL)



Excess degree distribution

Probability of such cases

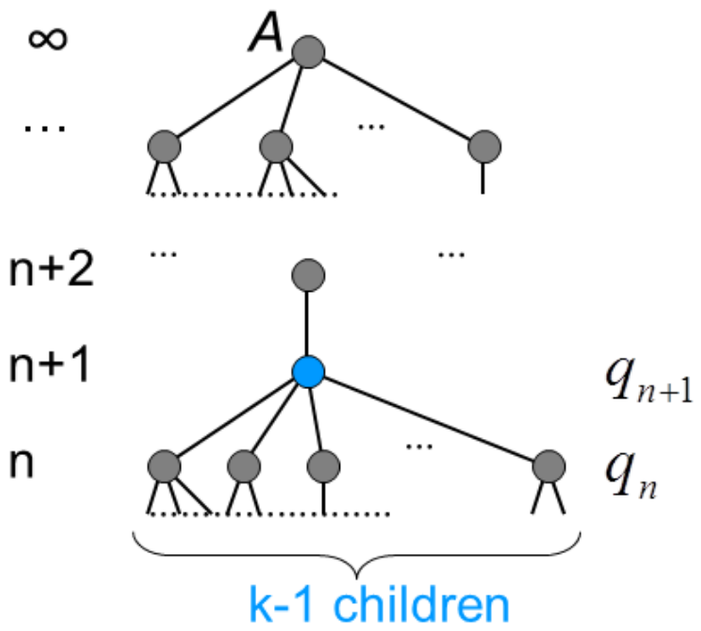
$$q_{n+1} = \rho_0 + (1-\rho_0)\sum_{k=1}^{\infty}\frac{k}{z}p_k\sum_{m=0}^{k-1}\binom{k-1}{m}q_n^m(1-q_n)^{k-1-m}F(m,k)$$

seed

non-seeds

For m larger than the threshold,

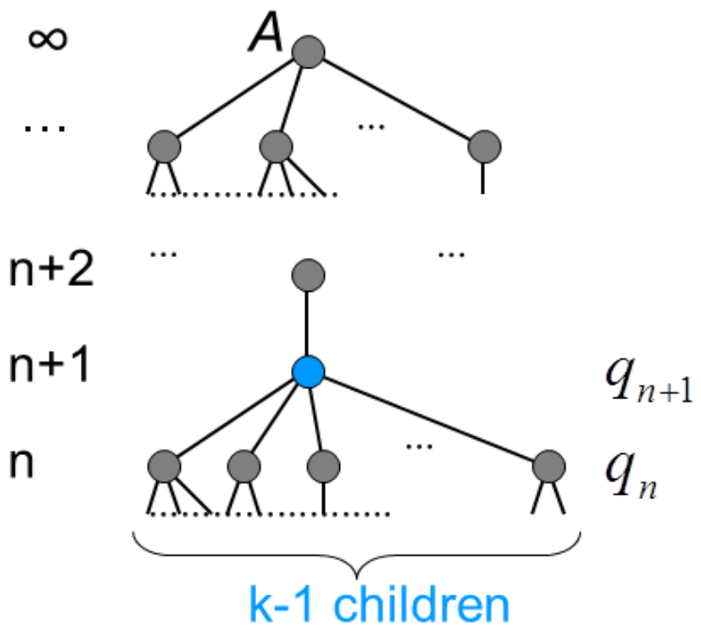
2. Tree Like Approximation (TL)



$$q_{n+1} = \rho_0 + (1-\rho_0)\sum_{k=1}^{\infty} \frac{k}{z} p_k \sum_{m=0}^{k-1} \binom{k-1}{m} q_n^m (1-q_n)^{k-1-m} F(m,k)$$

$$\rho = \rho_0 + (1-\rho_0)\sum_{k=0}^{\infty} p_k \sum_{m=0}^{k} \binom{k}{m} q_\infty^m (1-q_\infty)^{k-m} F(m,k)$$

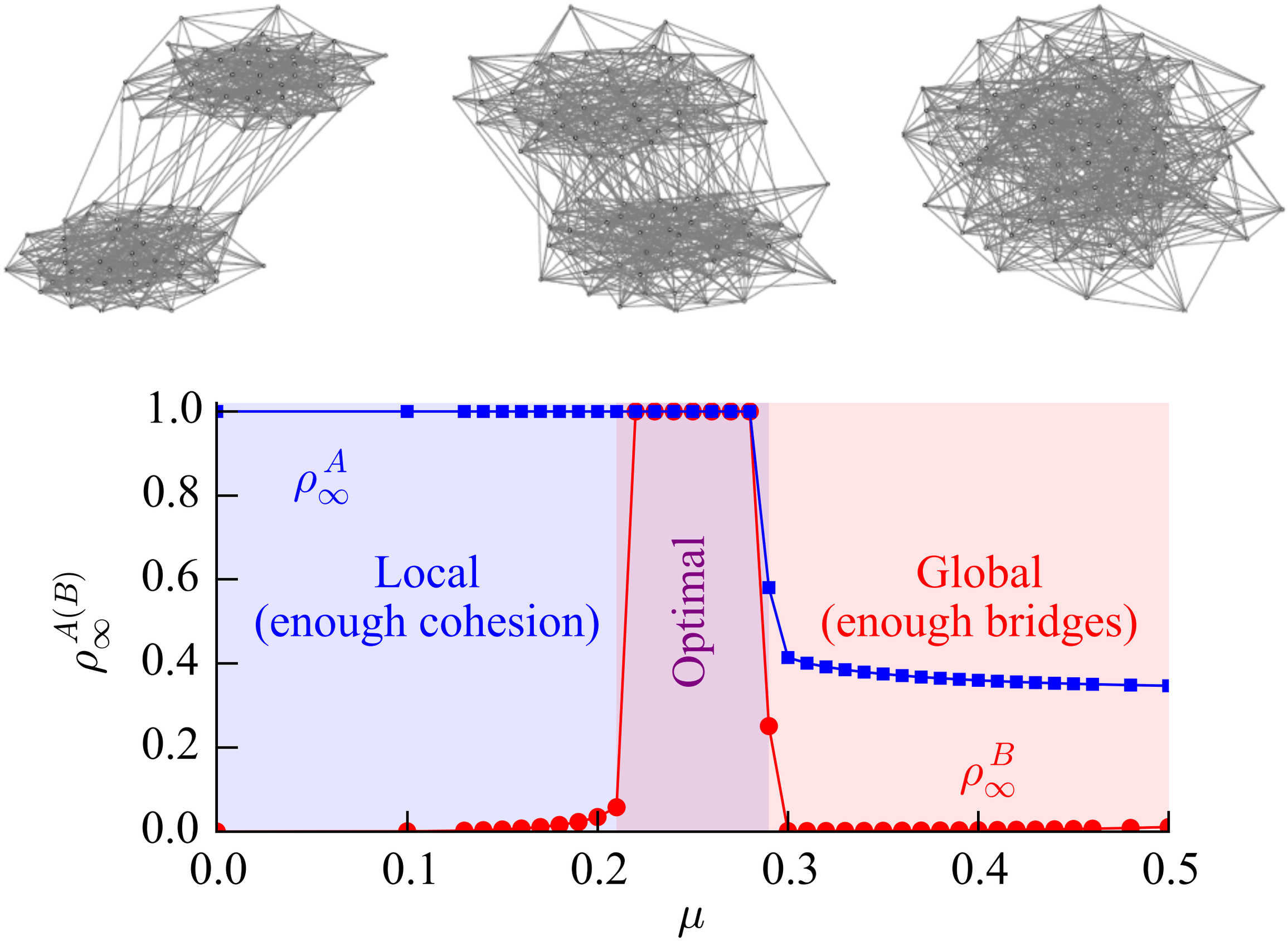
2. Tree Like Approximation (TL)



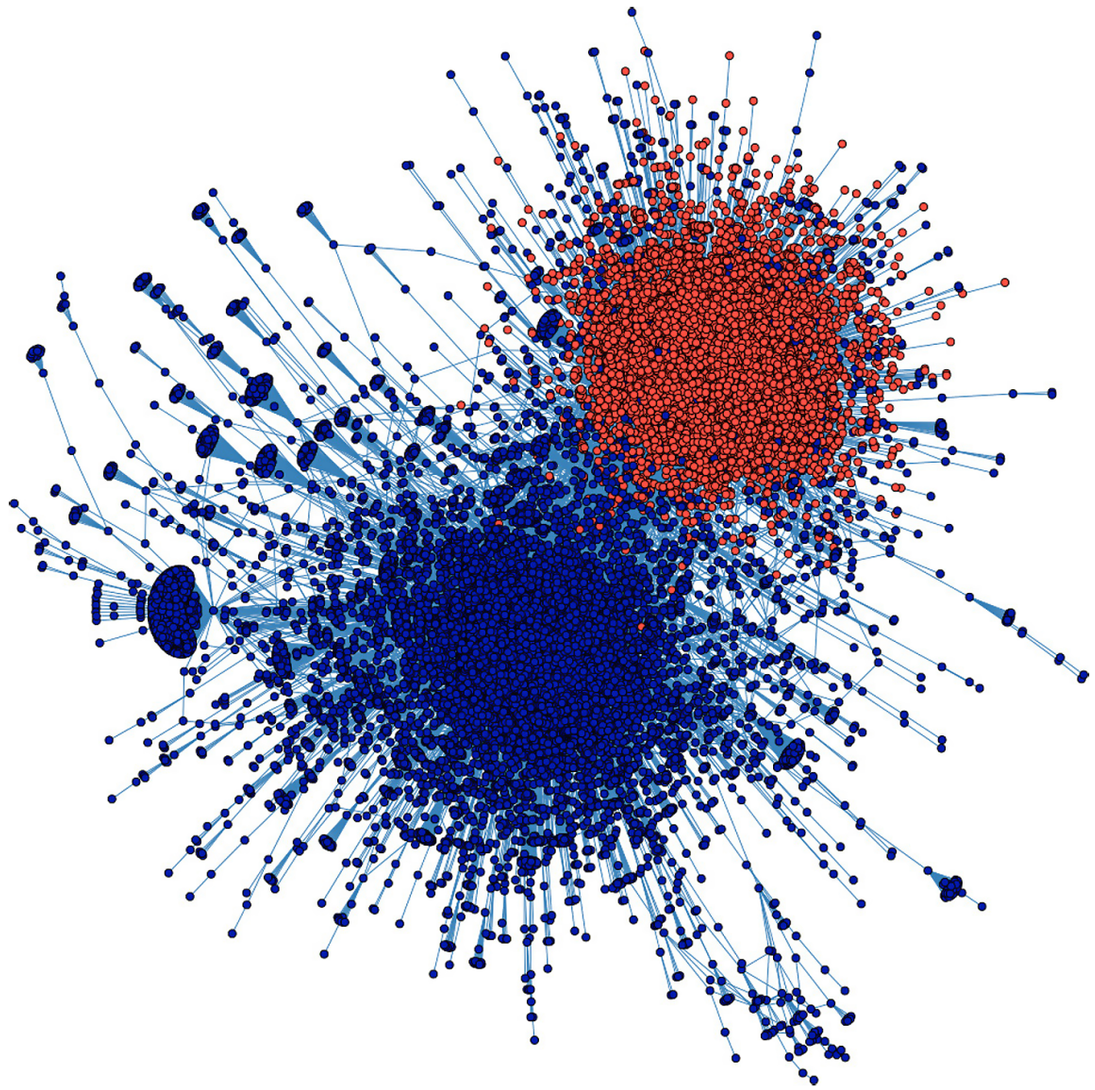
$$q_{n+1} = \rho_0 + (1-\rho_0)\sum_{k=1}^{\infty} \frac{k}{z} p_k \sum_{m=0}^{k-1} \binom{k-1}{m} q_n^m (1-q_n)^{k-1-m} F(m,k)$$

Degree distribution

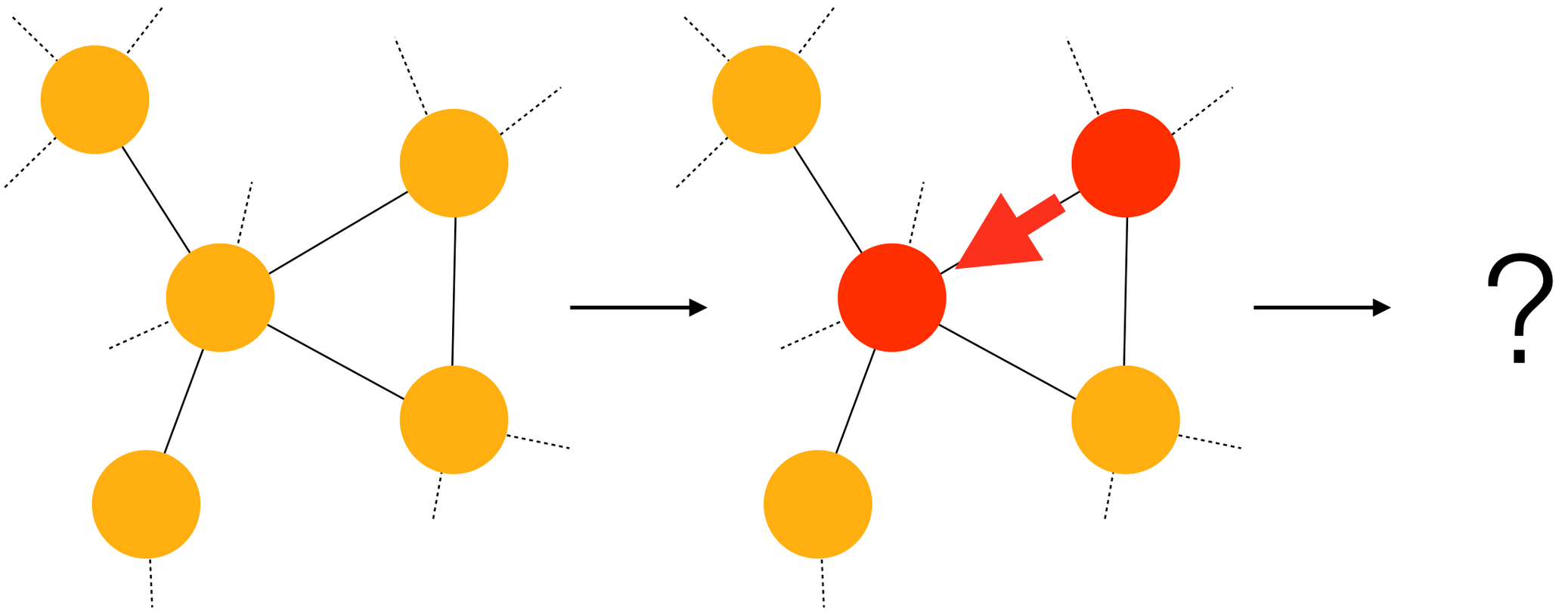
$$\rho = \rho_0 + (1-\rho_0)\sum_{k=0}^{\infty} p_k \sum_{m=0}^{k} \binom{k}{m} q_\infty^m (1-q_\infty)^{k-m} F(m,k)$$

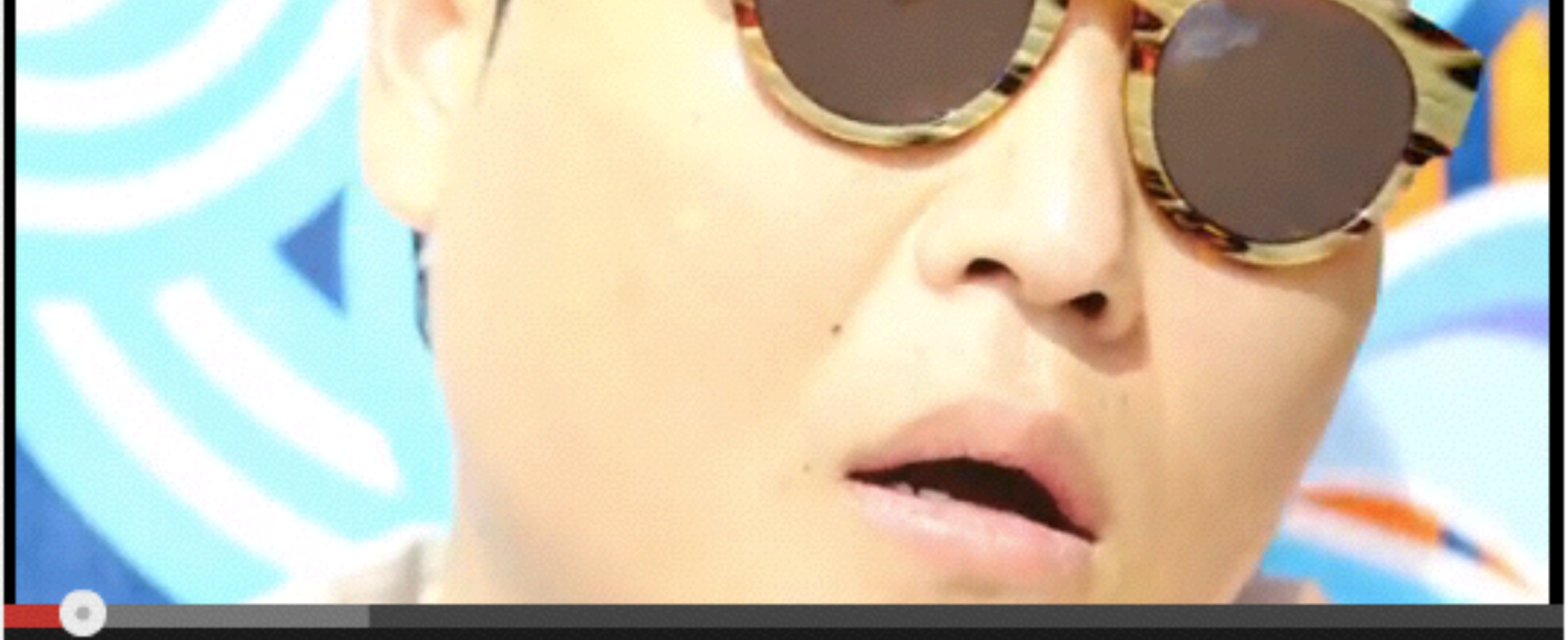

1.0
0.8
0.6
0.4
0.2
0.0
$\rho_\infty^{A(B)}$
ρ_∞^A
Local
(enough cohesion)
Optimal
Global
(enough bridges)
ρ_∞^B
0.0
0.1
0.2
0.3
0.4
0.5
μ

Even strong communities can **enhance** spreading



#3 Viral memes





0:08 / 4:12

PSY - GANGNAM STYLE (강남스타일) M/V

officialpsy · 82 videos

Subscribe 7,013,662

2,004,782,088

8,359,033 1,042,809

Why? How?

The song is addictive, MV is funny, the dance move is great, ...

NUMBERS

Homo Narrativus and the Trouble with Fame

Networks: We think that fame is deserved. We are wrong.

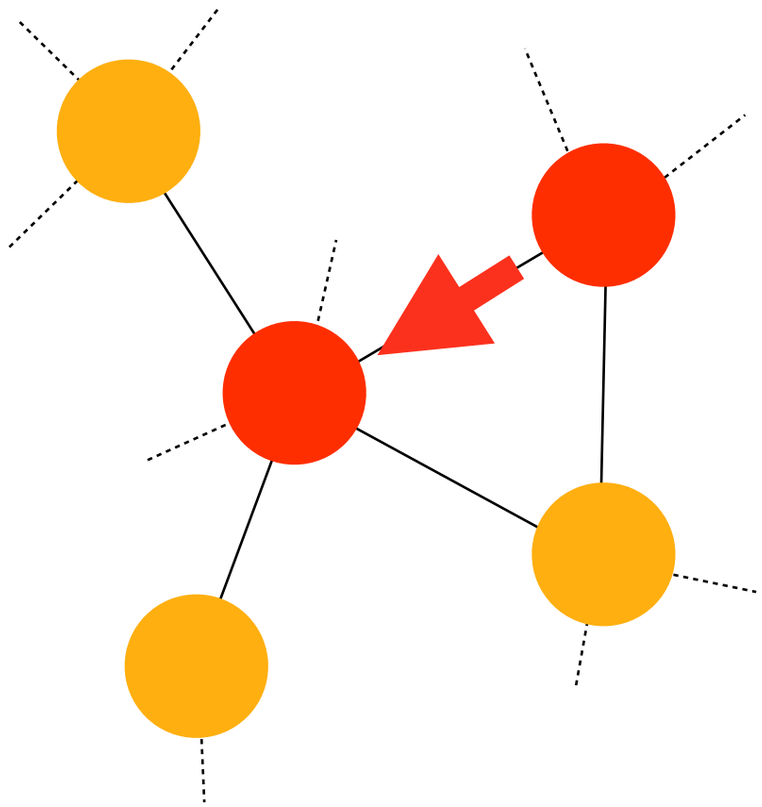
◆

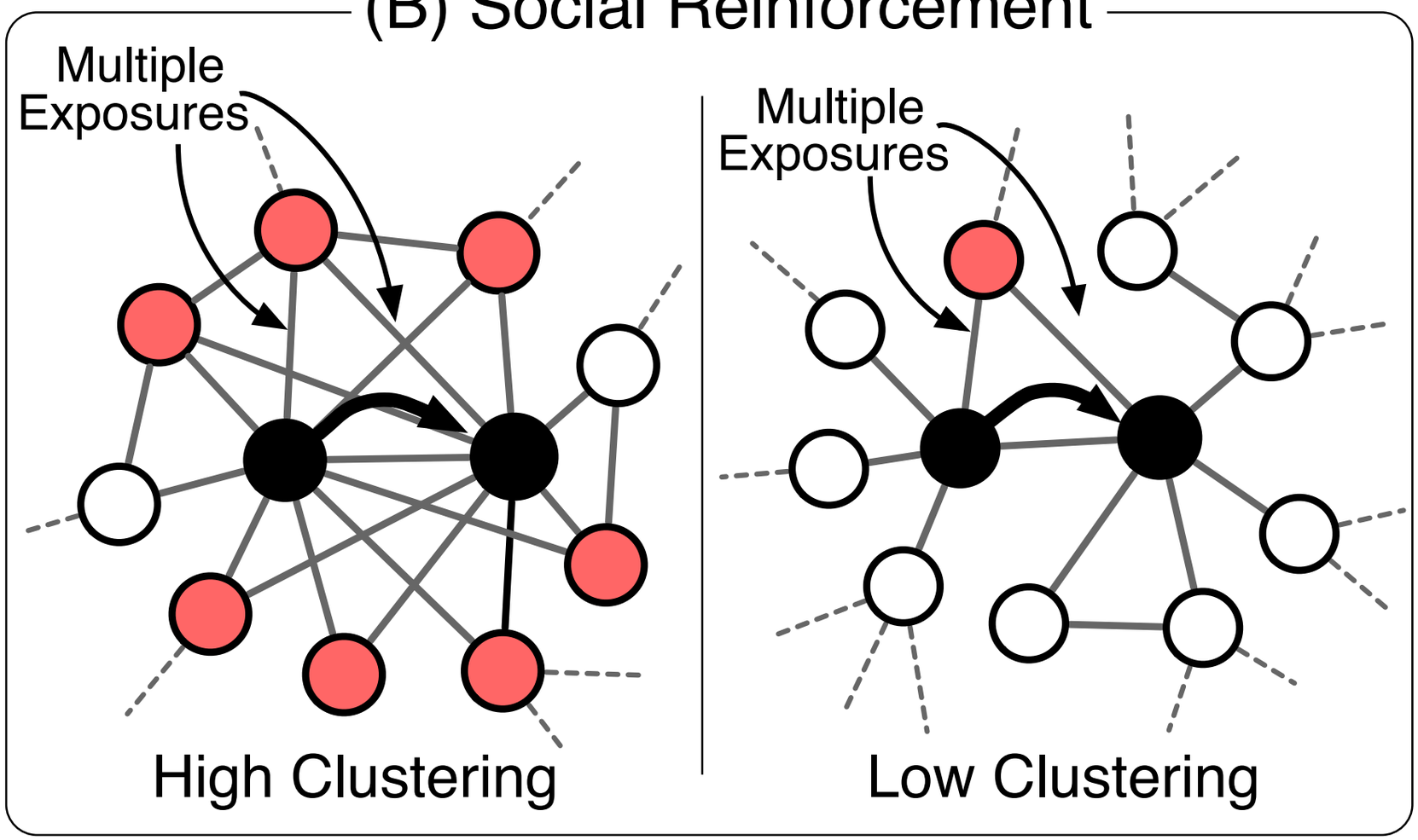
BY PETER SHERIDAN DODDS
ILLUSTRATION BY DANIEL ZALKUS

Homo

"Fame has much less to do with intrinsic quality than we believe it does, and much more to do with the characteristics of the people among whom fame spreads."

BY PETER SHERIDAN DODDS
ILLUSTRATION BY DANIEL ZALKUS



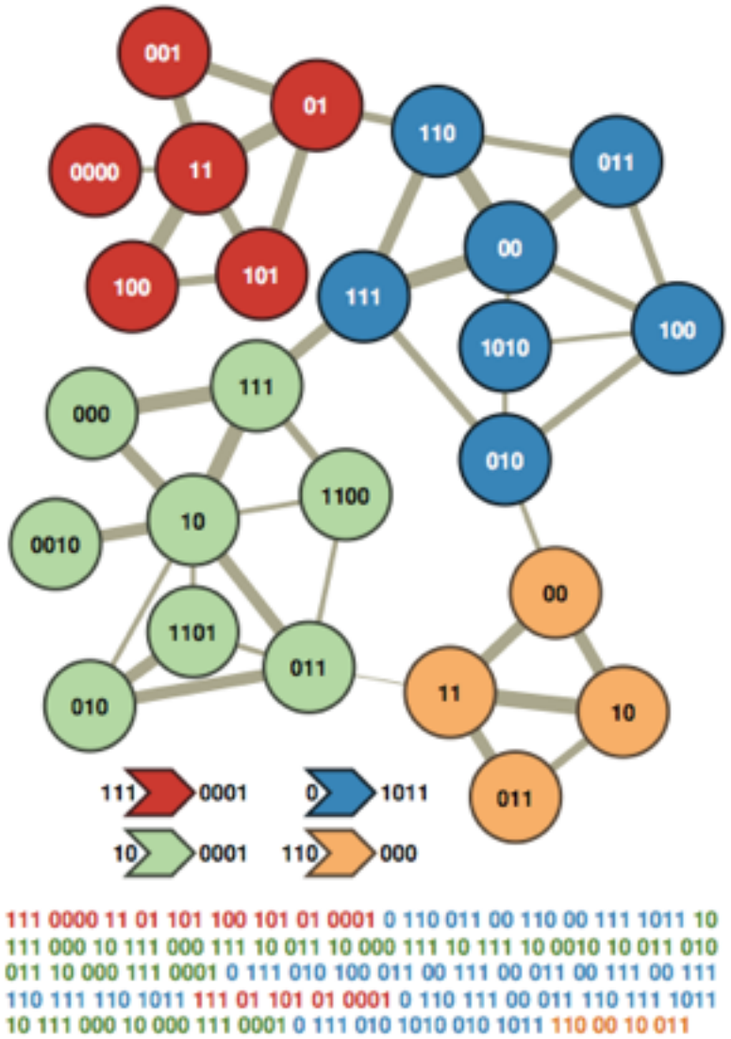


communities can **enhance** complex contagion

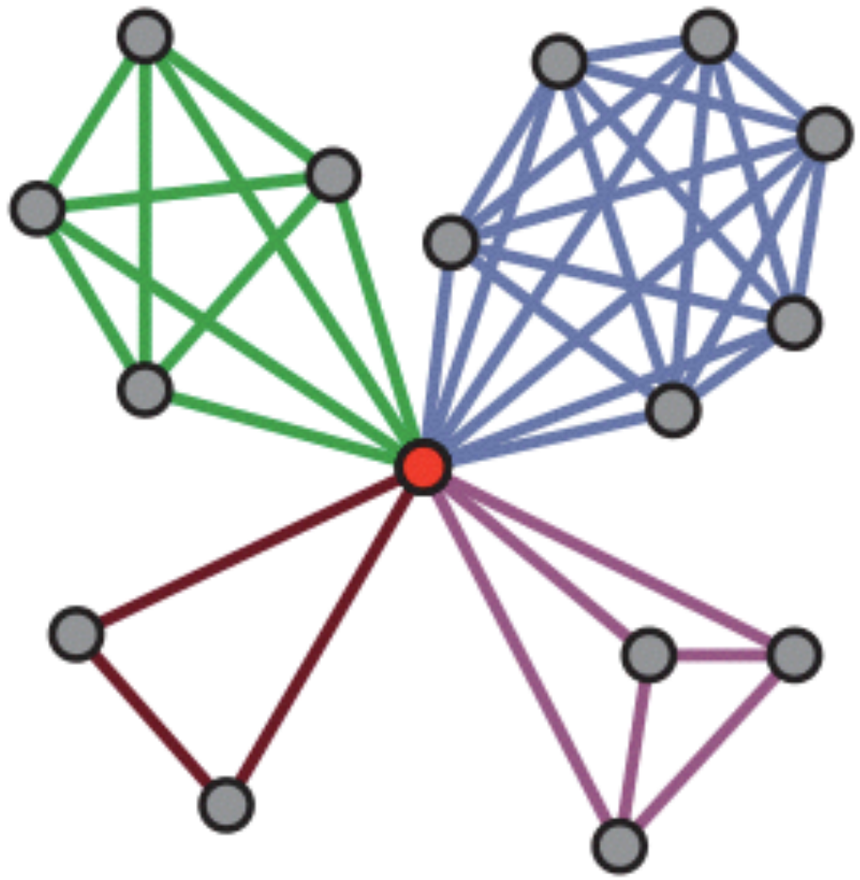
Prediction

If memes are complex contagions,
there will be strong **concentration**
of memes in **communities**.

Two community detection methods



Infomap (Rosvall & Bergstrom, 2008)

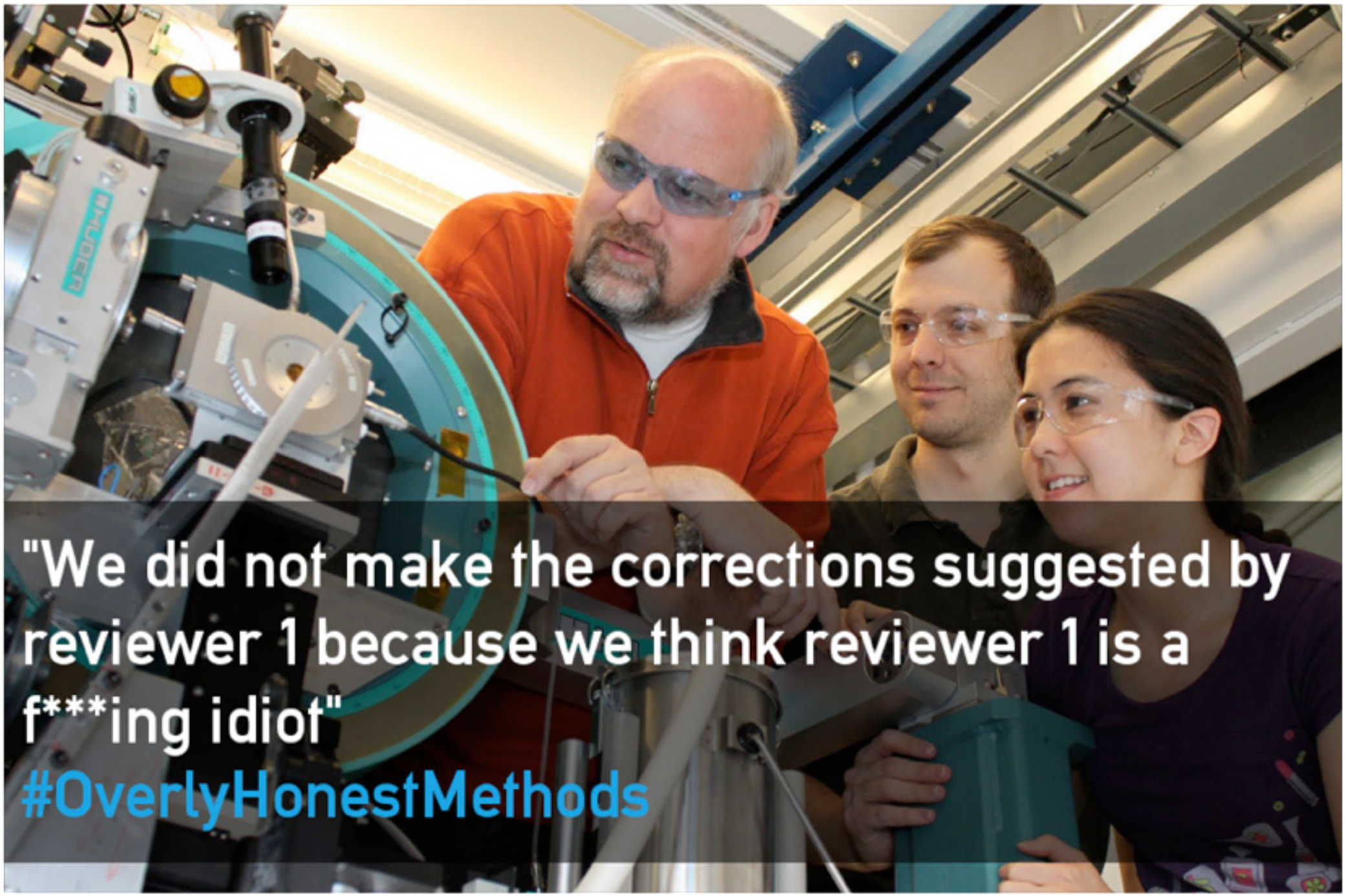


Link clustering (Ahn, Bagrow, Lehmann, 2010)

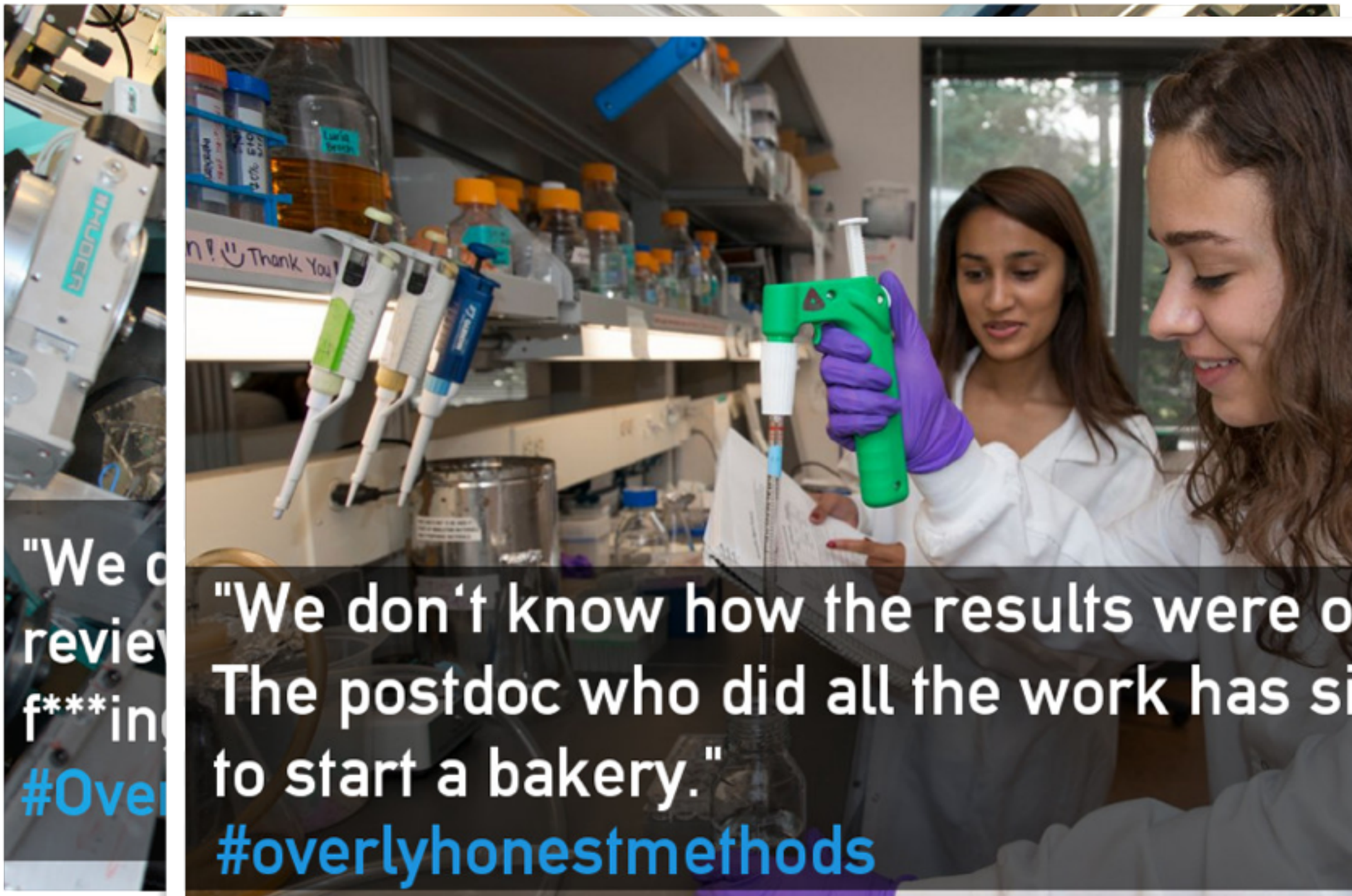
120 million tweets
(Mar 24 – Apr 25, 2012)

600k users, only
reciprocal edges.

Hashtags ~ Memes



Hashtags ~ Memes



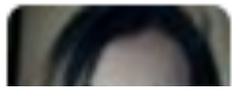
Sarah Kendzior
@sarahkendzior

Follow

Here's a paper to cite: mine. #SixWordPeerReview

6:16 PM - 21 Jan 2014

35 RETWEETS **44** FAVORITES



Sarah Kendzior

Follow

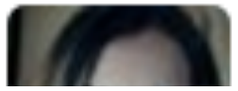
Eric Pollitt
@StaphEric

Follow

I came, I saw, I rejected #SixWordPeerReview

7:36 PM - 21 Jan 2014

3 RETWEETS **6** FAVORITES



Sarah Kendzior

Follow

Eric Pollitt

@StaphEric

Follow

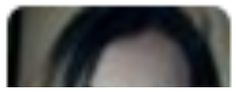
Leonsaviour

@leonsaviour

Follow

My insomnia is gone. Thank you. #SixWordPeerReview

7:40 PM - 21 Jan 2014



Sarah Kendzior

Follow

Eric Pollitt
@StaphEric

Follow

Leonsaviour
@leonsaviour

Follow

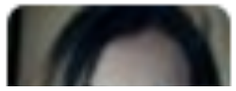
Nancy Owens
@nancyjowens

Follow

Too long; didn't read. Looks good. #SixWordPeerReview

11:51 PM - 21 Jan 2014

1 FAVORITE



Sarah Kendzior

Follow

Eric Pollitt
@StaphEric

Follow

Leonsaviour
@leonsaviour

Follow

Nancy Owens
@nancyjowens

Follow

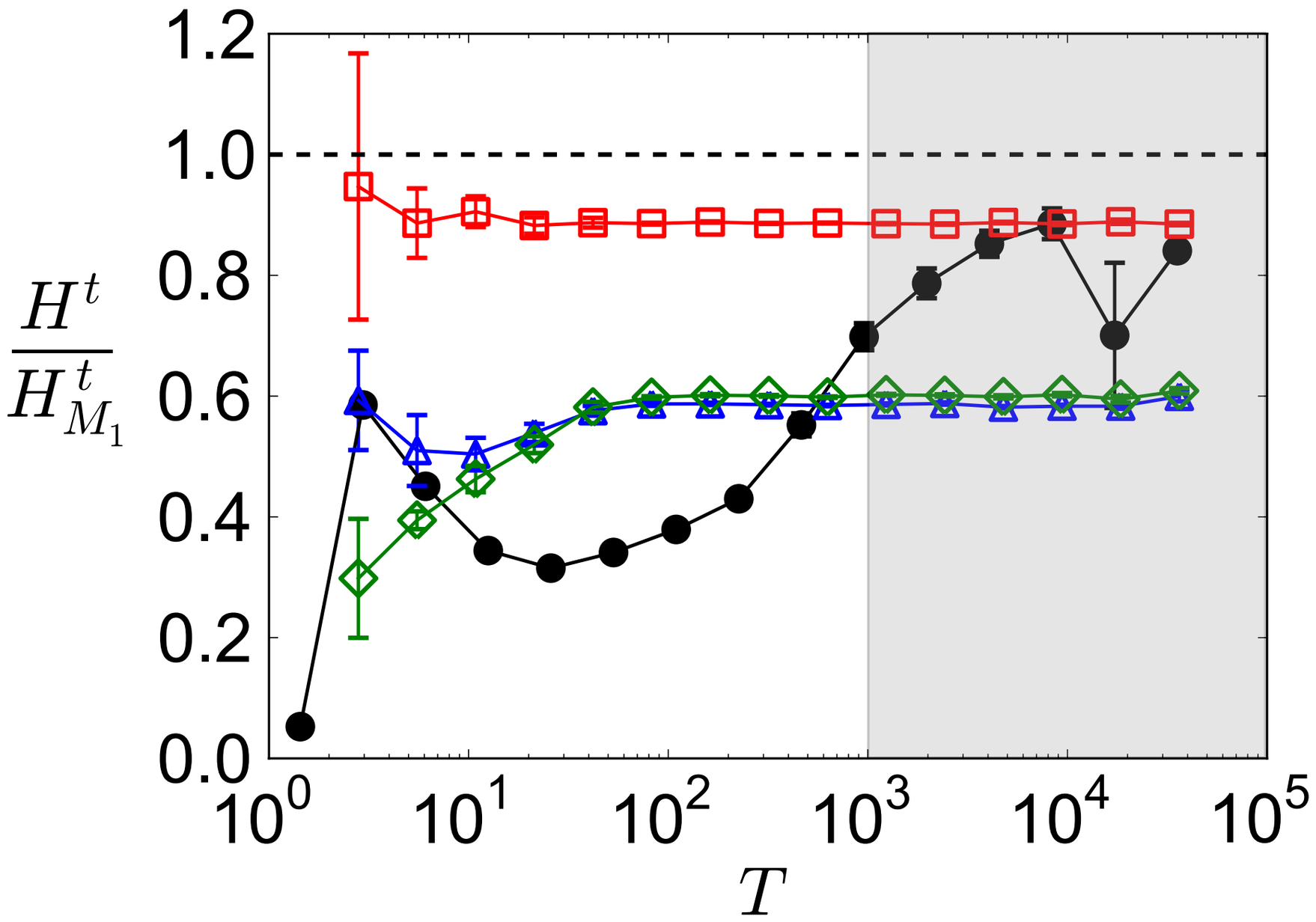
Tyler Schnoebelen
@TSchnoebelen

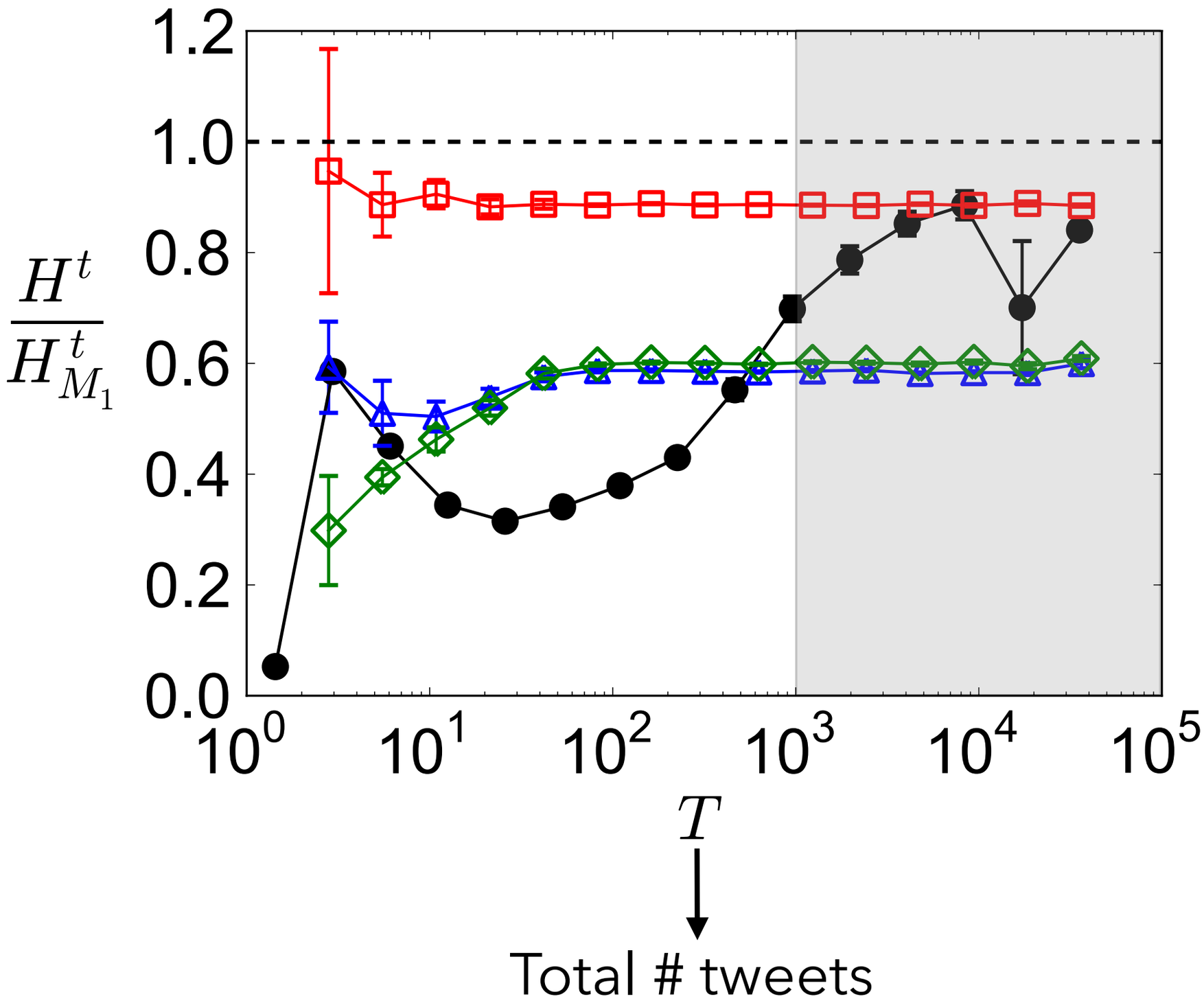
Follow

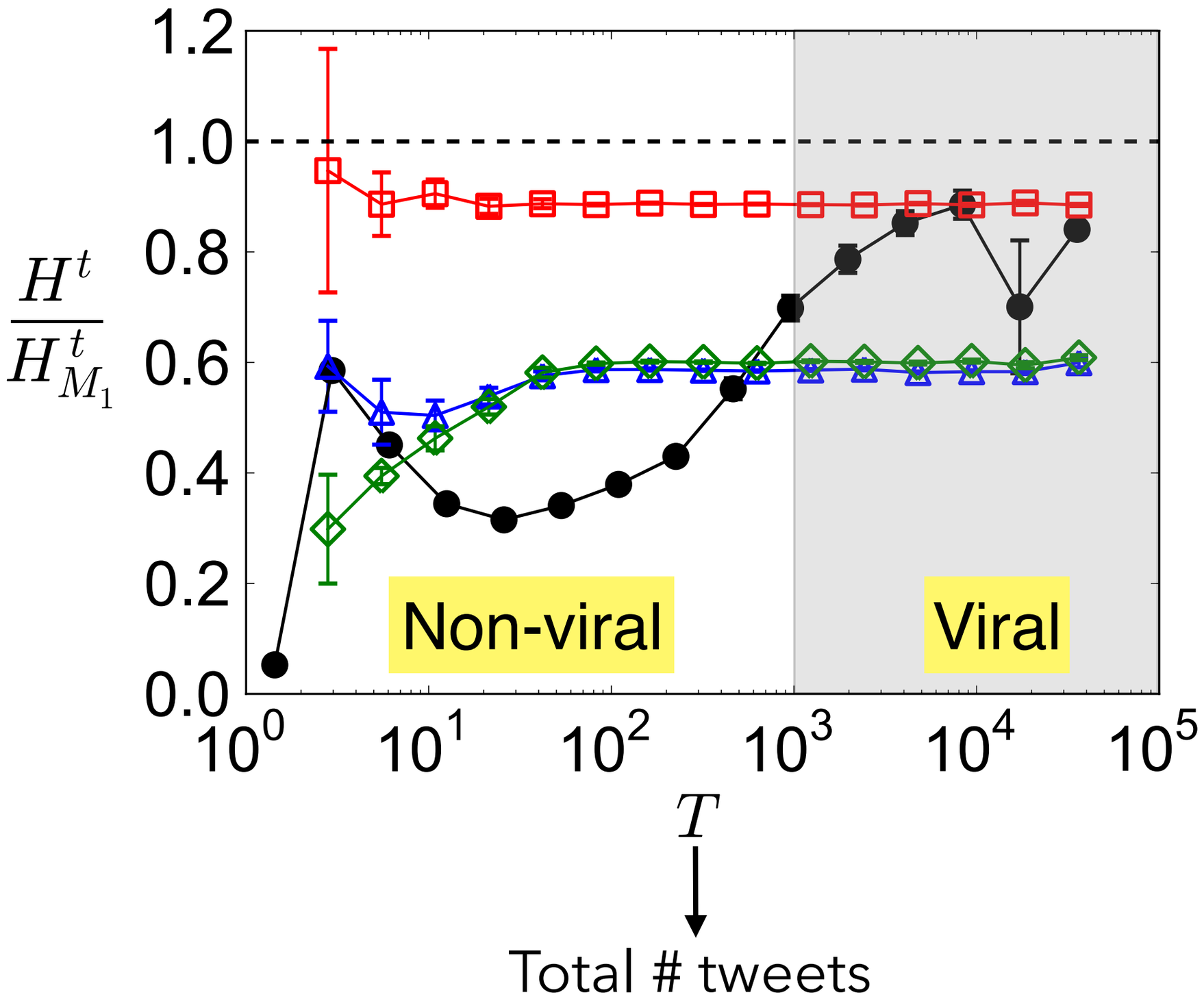
Your bibliography is a giant selfie. #SixWordPeerReview

12:59 AM - 22 Jan 2014

74 RETWEETS **53** FAVORITES

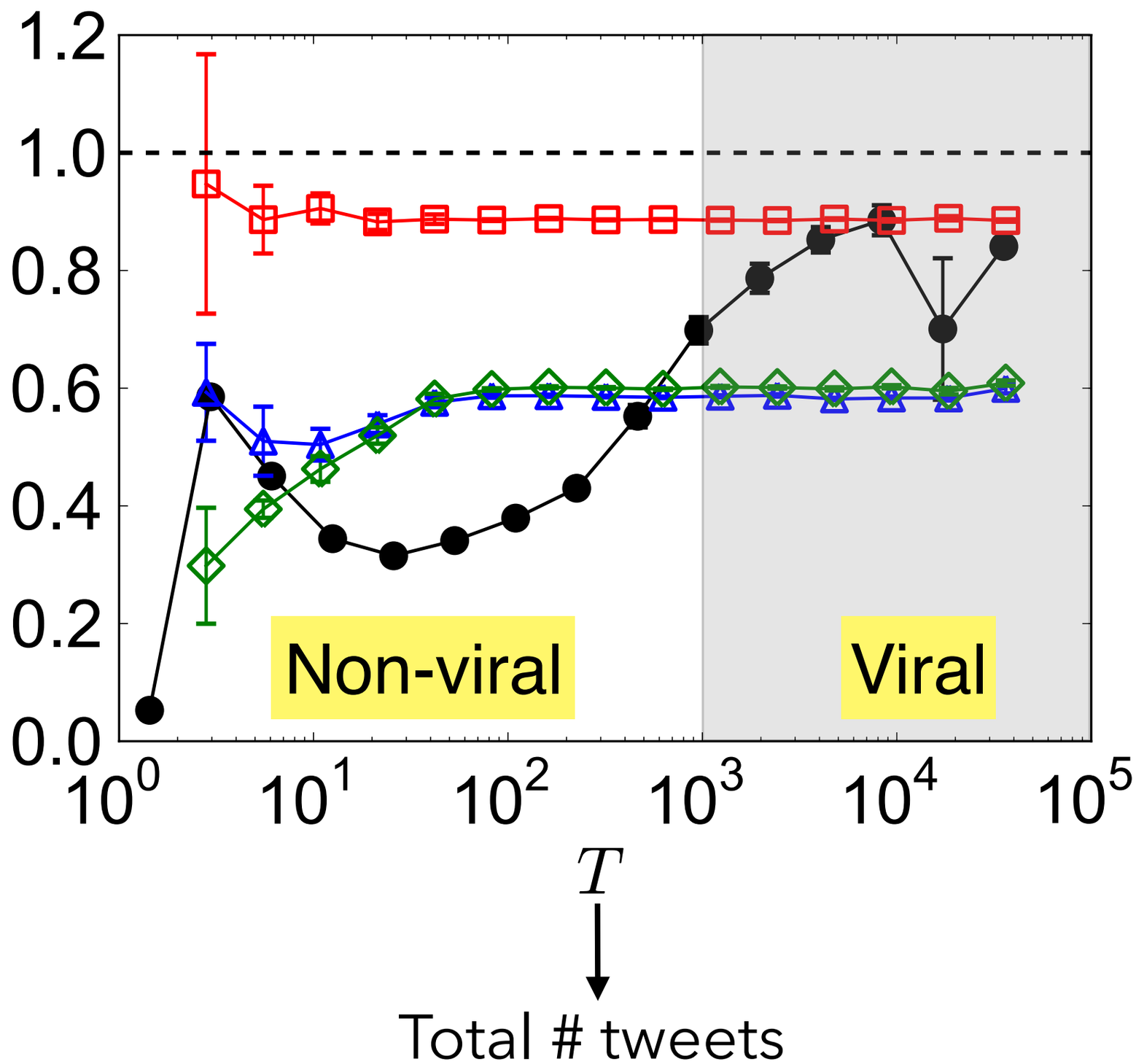

1.2
1.0
0.8
0.6
0.4
0.2
0.0
$\frac{H^t}{H^t_{M_1}}$
10^0
10^1
10^2
10^3
10^4
10^5
T


1.2
1.0
0.8
0.6
0.4
0.2
0.0
$\frac{H^t}{H^t_{M_1}}$
10^0
10^1
10^2
10^3
10^4
10^5
T
Total # tweets


1.2
1.0
0.8
0.6
0.4
0.2
0.0
$\frac{H^t}{H^t_{M_1}}$
Non-viral
Viral
10^0
10^1
10^2
10^3
10^4
10^5
T
Total # tweets

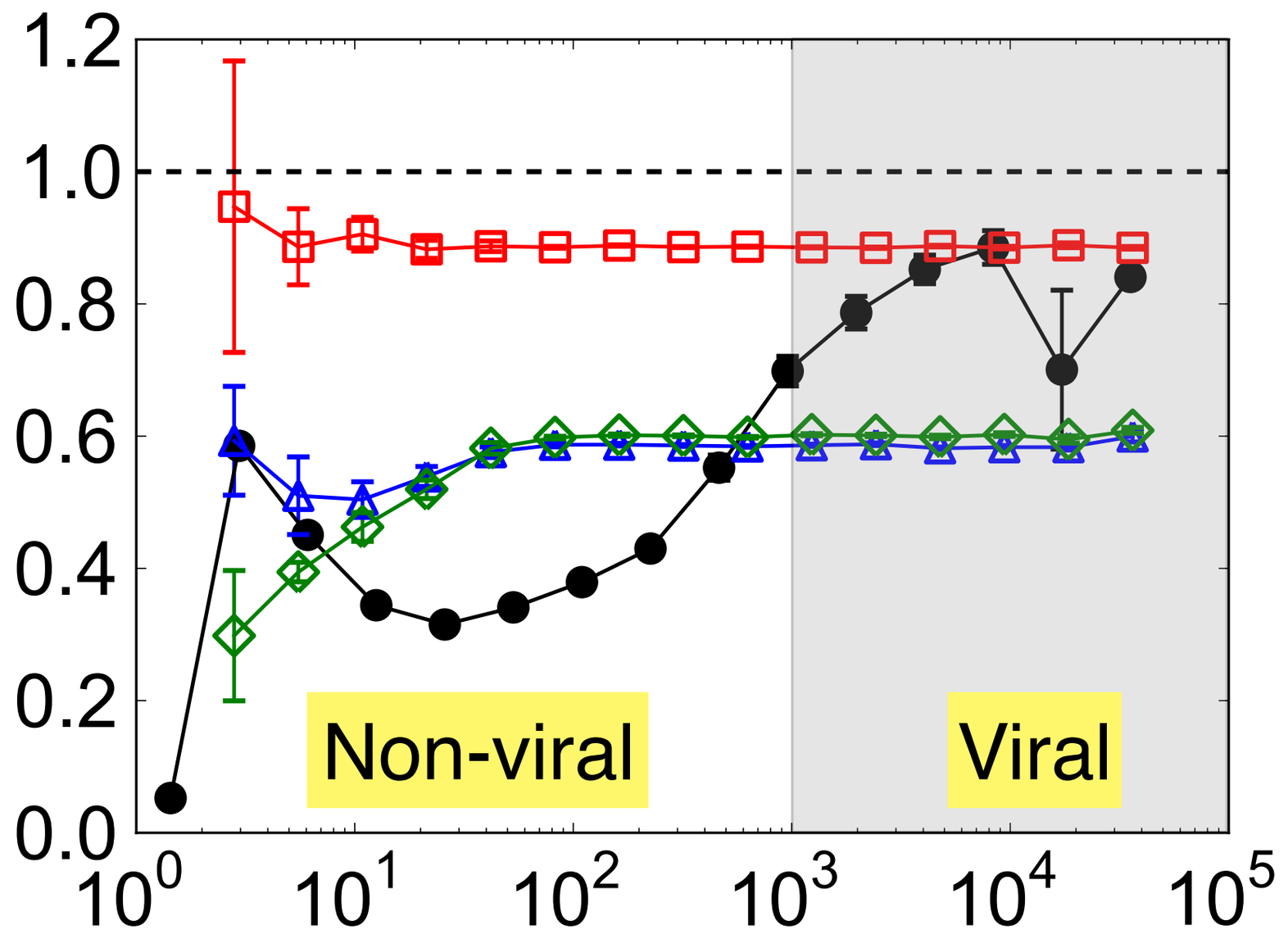
Entropy of # tweets distributed in different communities

$$\frac{H^t}{H^t_{M_1}}$$



Entropy of # tweets distributed in different communities

$$\frac{H^t}{H^t_{M_1}}$$



M1: Random distribution

M2: Random diffusion

M3: Social reinforcement

M4: Homophily

Viral memes spread like **diseases**.

Viral memes spread like **diseases**.

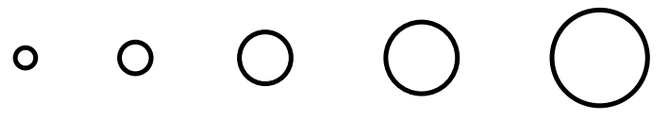
Uninteresting memes are easily ‘**trapped**’ by communities while viral memes are not.

Old

New

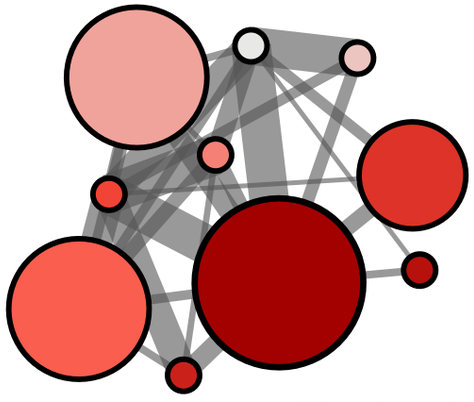
Less dominant

More dominant



(A) #ThoughtsDuringSchool

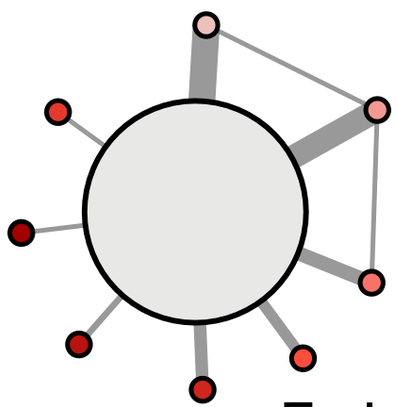
30 tweets



Early Stage

(B) #ProperBand

30 tweets



Early Stage

Old

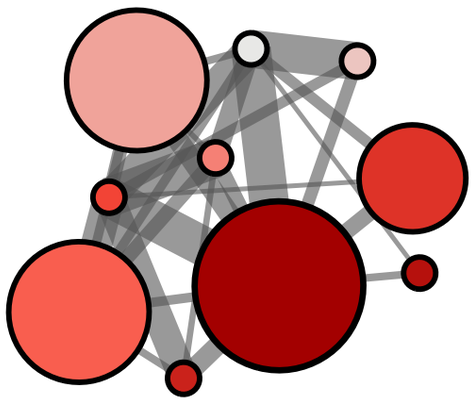
New

Less dominant

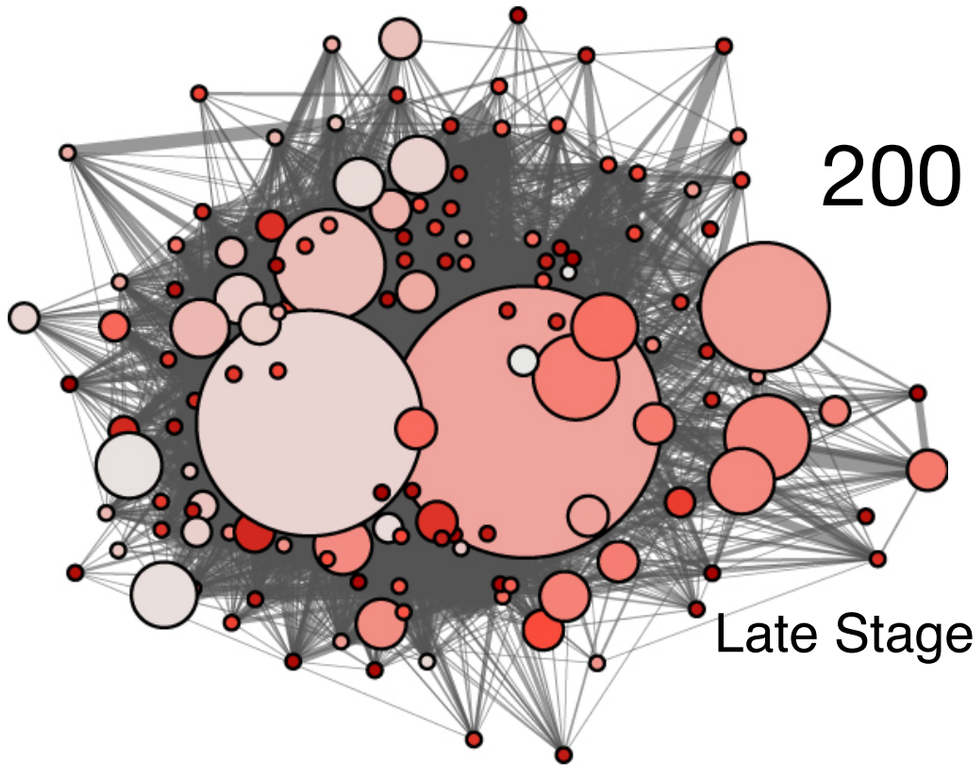
More dominant

(A) #ThoughtsDuringSchool

30 tweets



Early Stage

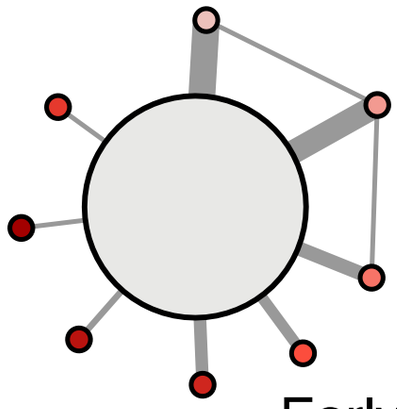


200 tweets

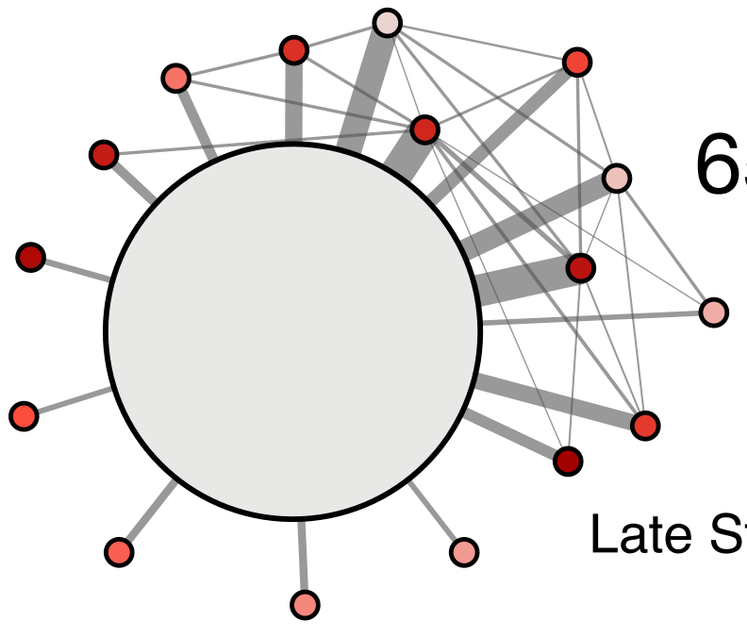
Late Stage

(B) #ProperBand

30 tweets



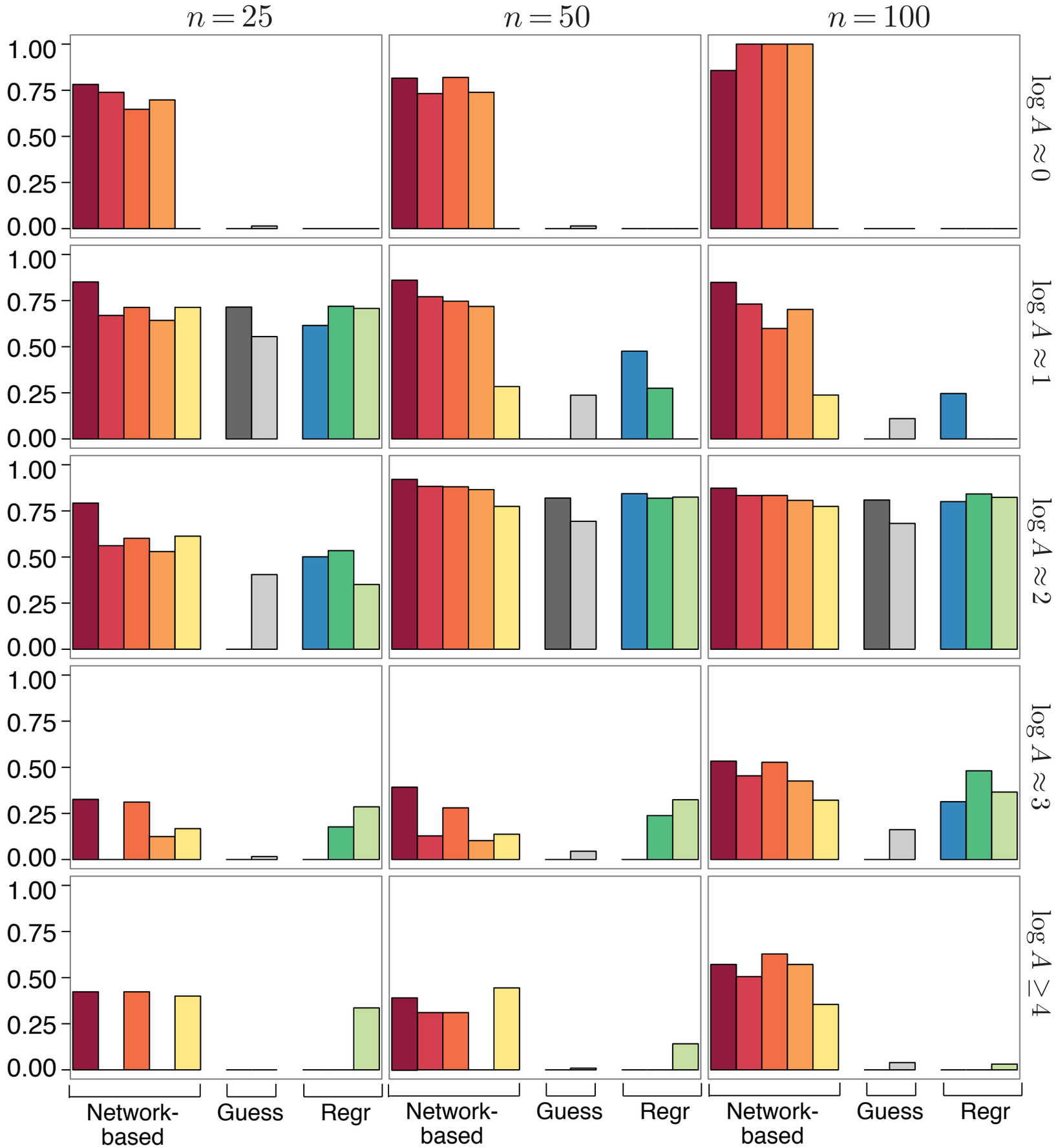
Early Stage



65 tweets

Late Stage

F1



Our model

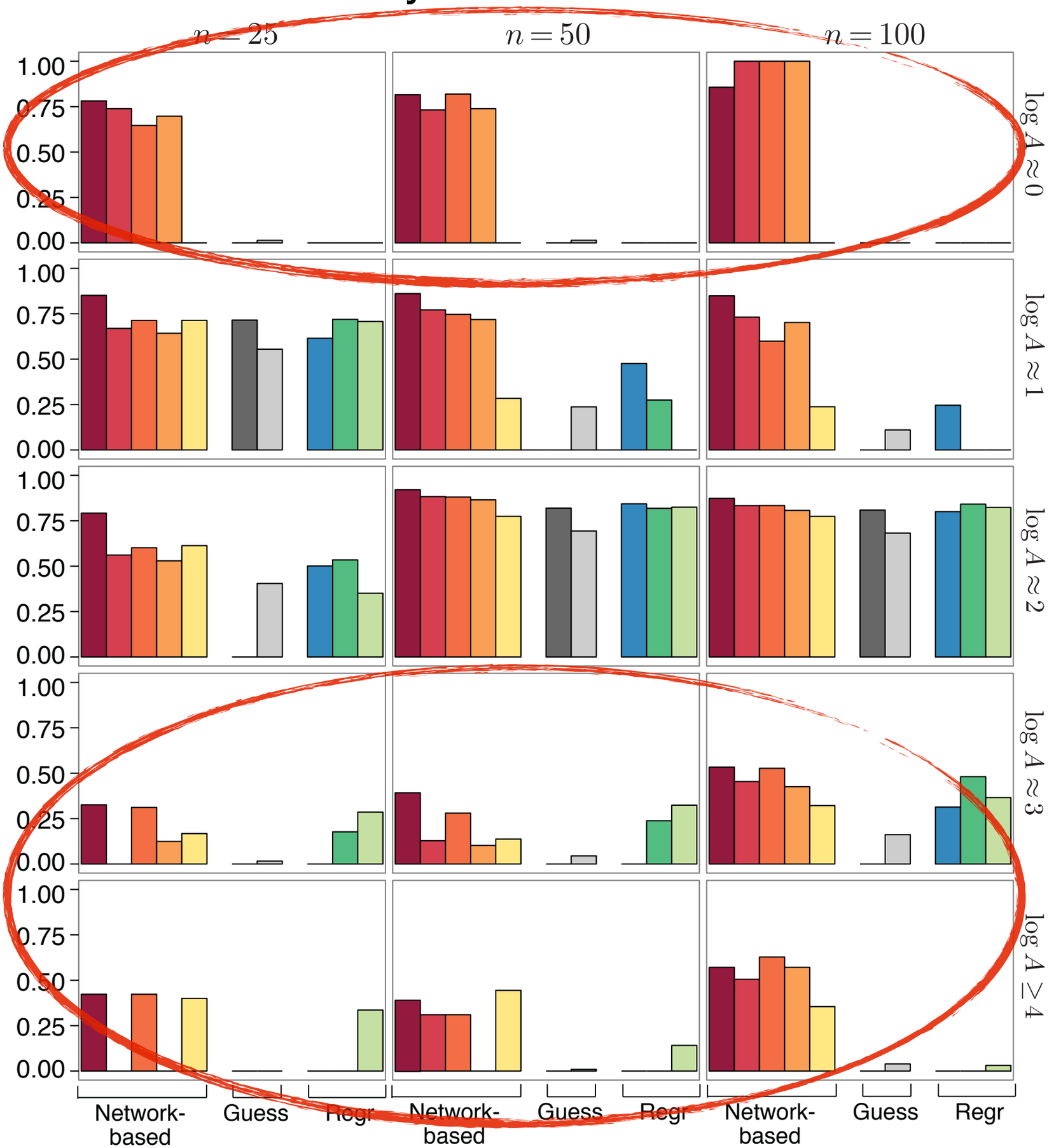
- All, P_n
- Basic
- Community
- Distance
- Timing

- Random, B_1
- Majority, B_2

- Influence, B_3
- LN($\tau = 7$), B_4
- ML($\tau = 7$), B_5

(Weng et al. 2014)

F1



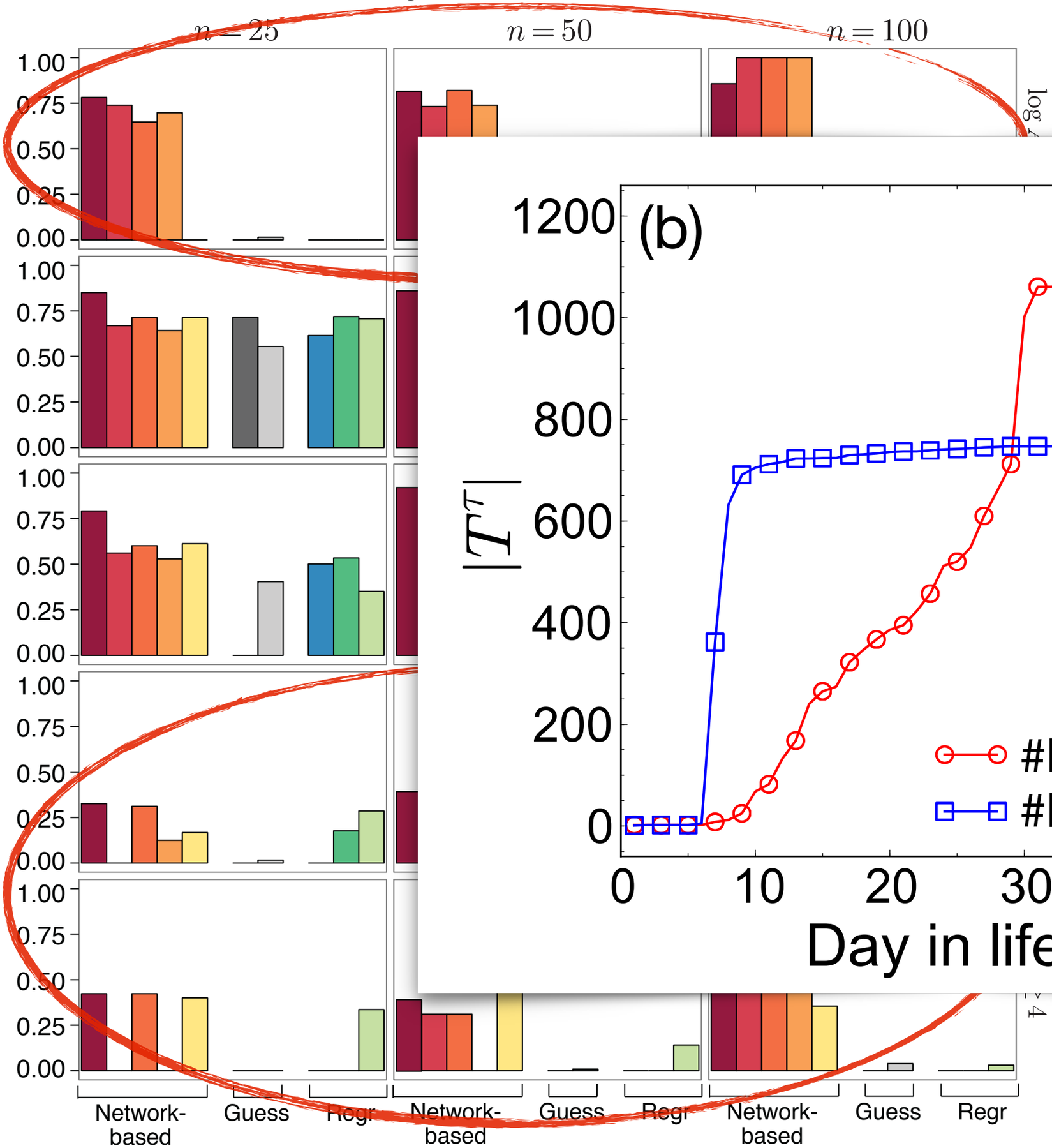
Our model

- All, P_n
- Basic
- Community
- Distance
- Timing
- Random, B_1
- Majority, B_2
- Influence, B_3
- LN($\tau = 7$), B_4
- ML($\tau = 7$), B_5

(Weng et al. 2014)

Early tweets tracked

F1



(Weng et al. 2014)

Examining spreading patterns in terms of **communities**

allows us to distinguish different dynamics (simple vs. complex)

#4 Cultural diffusion of food

“Tell me what you eat, and I will tell you what you are.”

Jean Anthelme Brillat-Savarin
(1755-1826)

What do we eat?

Tender hamburger bun, made from scratch and toasted in beef suet

Hamburger glaze of suet, pureed tomato confit, beef stock, and smoked salt

Maitake mushroom, sauteéd in beef suet

Romaine lettuce infused sous vide with liquid hickory smoke

Vacuum-compressed heirloom tomato

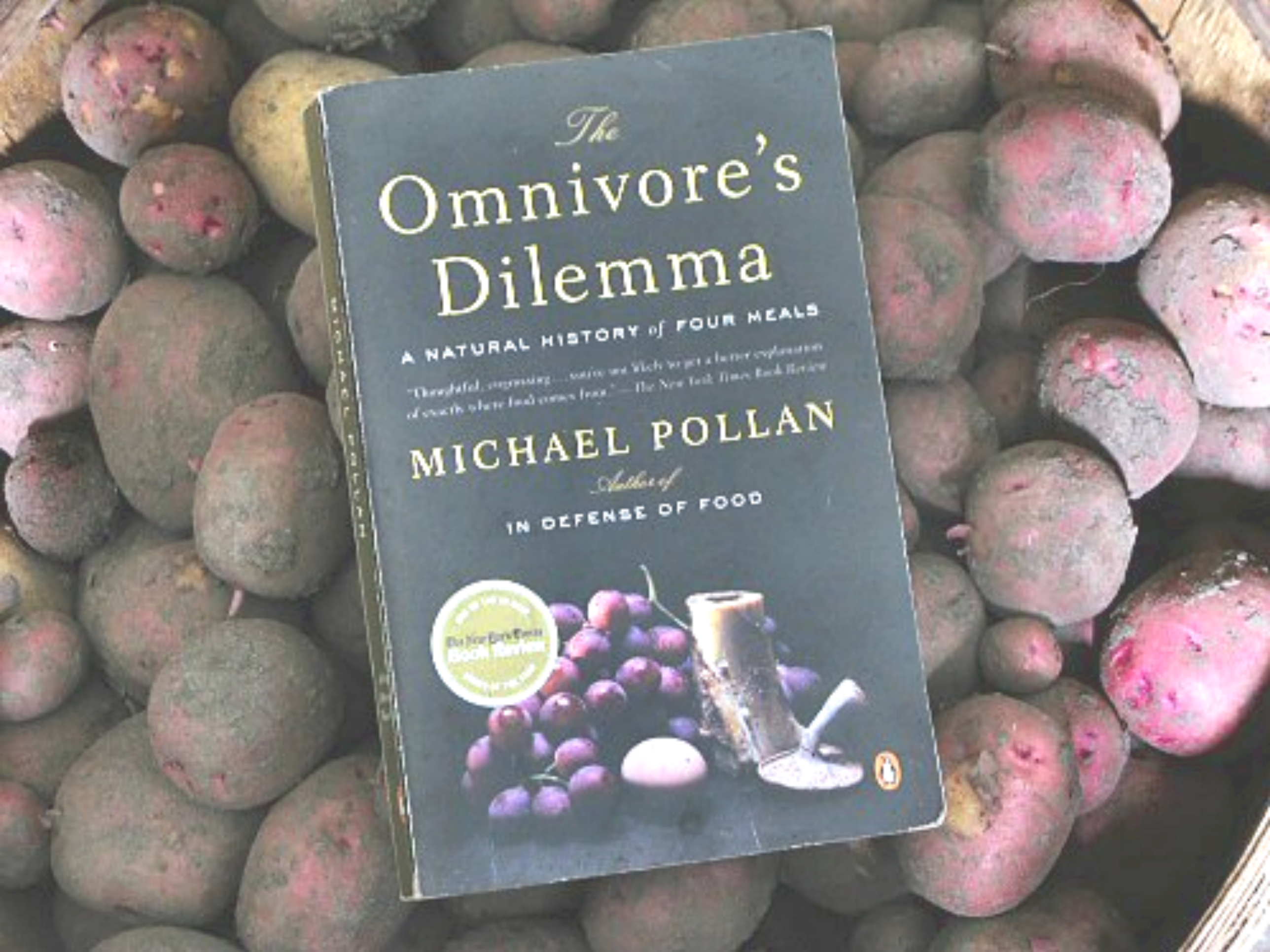
Cheese single made from aged Emmental, Comté, and wheat ale

Short-rib patty ground to vertically align the grain

Crimini mushroom ketchup with honey, horseradish, fish sauce, ginger, and allspice

We are

Omnivores


The
Omnivore's
Dilemma
A NATURAL HISTORY of FOUR MEALS
MICHAEL POLLAN
Author of
IN DEFENSE OF FOOD


Friendship
All Natural
Sour
Cream
TABASCO

POISONOUS & PSYCHOTROPIC MUSHROOMS

How do we choose?

Why do we eat X?

Because it’s delicious!

Why is it delicious?

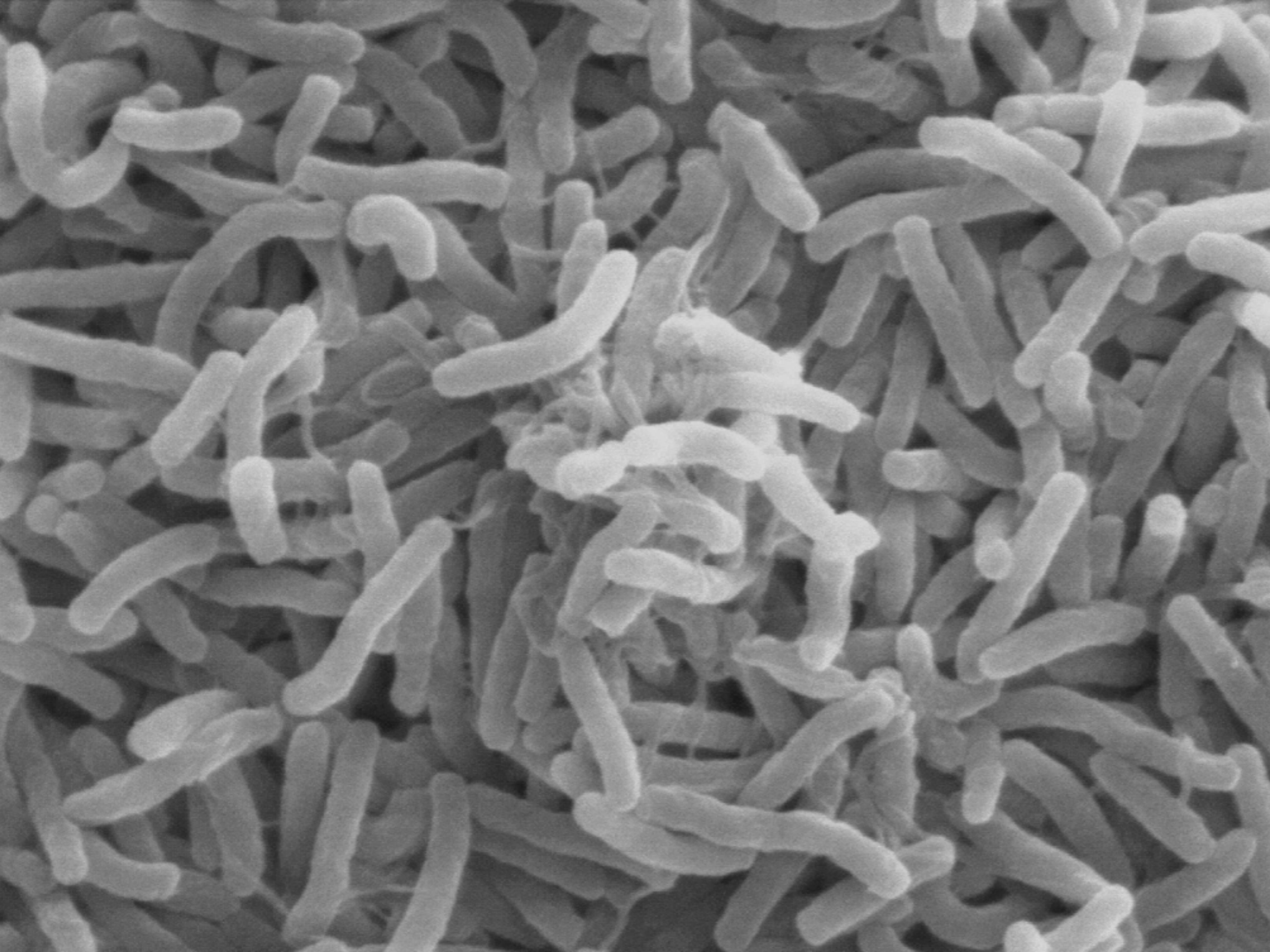
Energy!

Sweet + Fat

= **AWESOME**

Energy

Why do we eat spices?



Darwinian Gastronomy: Why We Use Spices

Spices taste good because they are good for us

Paul W. Sherman and Jennifer Billing

Humans have borrowed plants' chemical "recipes" for evolutionary survival for use in cuisine to combat foodborne microorganisms and to reduce food poisoning

Spices are plant products used in flavoring foods and beverages. For thousands of years, aromatic plant materials have been used in food preparation and preservation, as well as for embalming, in areas where the plants are native, such as Hindustan and the Spice Islands (Govindarajan 1985, Dillon and Board 1994). During and after the Middle Ages, seafarers such as Marco Polo, Ferdinand Magellan, and Christopher Columbus undertook hazardous voyages to establish

fruits of herbaceous plants (Figure 1). Cookbooks generally distinguish between seasonings (spices used in food preparation) and condiments (spices added after food is served), but not between herbs and spices. However, herbs, which are defined botanically (as plants that do not develop woody, persistent tissue), usually are called for in their fresh state, whereas spices generally are dried (Figure 2). Salt is sometimes thought of as a spice, but it is a mineral.

Each spice has a unique aroma and flavor, which derive from com-

vinian Gastronomy:
hy We Use Spices

ste good because they are good for us

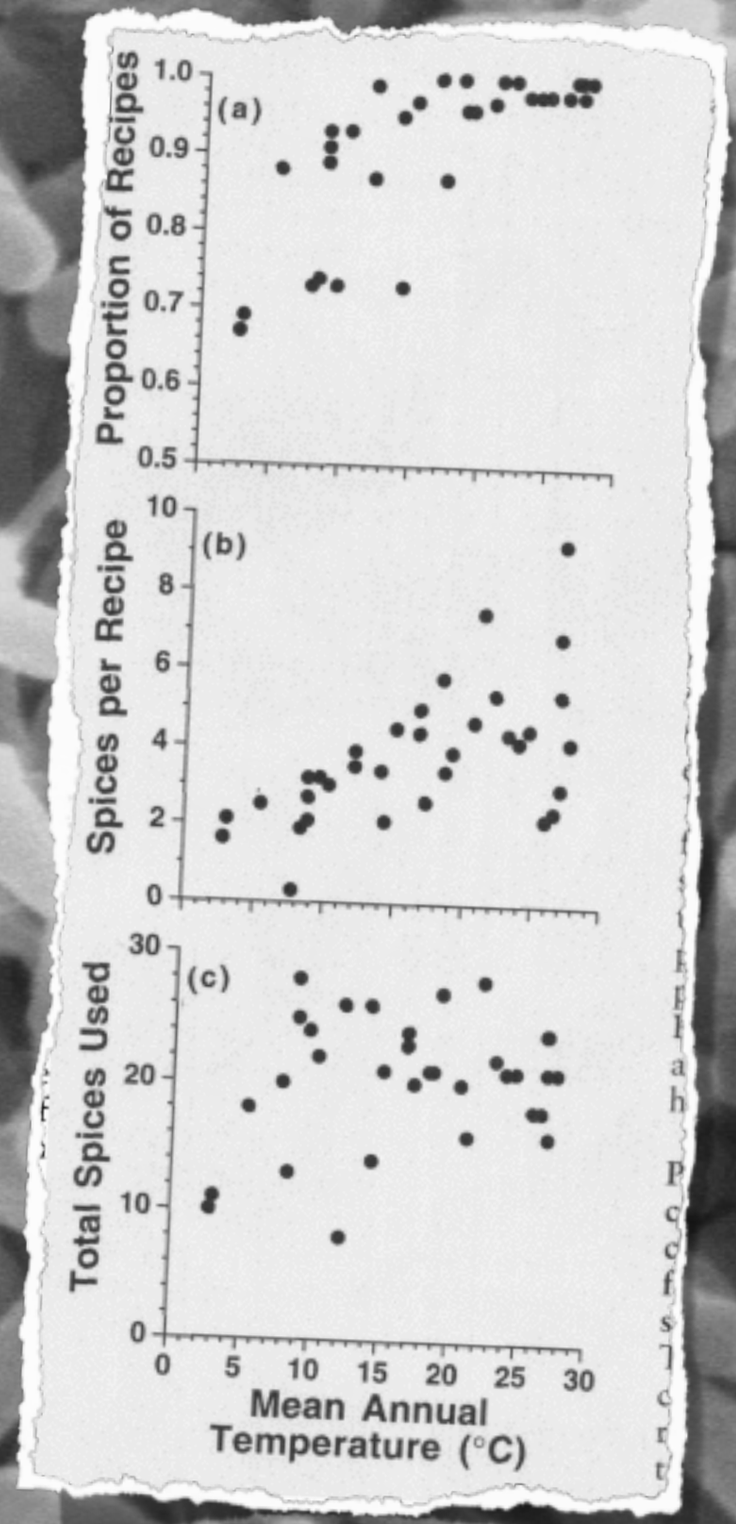
Paul W. Sherman and Jennifer Billing

Humans have borrowed plants' chemical "recipes" for evolutionary survival for use in cuisine to combat foodborne microorganisms and to reduce food poisoning

sed in
erages.
, aro-
n used
eserva-
ing, in
native,
pice Is-
Dillon
nd after
such as
lagellan,
s under-
establish

fruits of herbaceous plants (Figure 1). Cookbooks generally distinguish between seasonings (spices used in food preparation) and condiments (spices added after food is served), but not between herbs and spices. However, herbs, which are defined botanically (as plants that do not develop woody, persistent tissue), usually are called for in their fresh state, whereas spices generally are dried (Figure 2). Salt is sometimes thought of as a spice, but it is a mineral.

Each spice has a unique aroma and flavor, which derive from com-



In victu veritas

Harold McGee

A survey of spice use around the world concludes that spices serve the adaptive purpose of reducing food-borne disease. It highlights, however, the need for further research *in victu* — in food itself — rather than *in vitro*.

Like all animals, human beings eat to live; and food preparation is a cultural behaviour that presumably contributes to our fitness by making plant and animal tissues more nourishing. But food preparation in many cultures has become far more elaborate than simple survival would seem to require. Why do we invest so much effort in adorning and transforming our cereals, tubers, meats and milks? In particular, why do humans bother to flavour their foods with nutritionally insignificant quantities of herbs and spices? And why does tropical heat seem to foster especially pungent cuisines?

In an extensive literature survey and correlational study, published in *The Quarterly Review of Biology*[1], Billing and Sherman affirm the relationship between climate and spiciness. They attribute it to what they consider the primary adaptive value of seasoning — that antimicrobial compounds concentrated in spices reduce the incidence of food-borne disease.

Drawing on 93 cookbooks covering 36 countries, Billing and Sherman analysed 4,578 meat, poultry and seafood recipes for the number and kinds of spices included. (They use 'spice' to signify all plant flavourings, whether Far Eastern natives such as pepper and cloves, central Asian garlic and onions, Mediterranean herbs such as thyme and oregano, or central American chillies. Quantities were not considered.) They found a strong correlation between the mean annual temperature of a given country — an index of the rate at which foods will spoil there — and the mean number of spices added to its flesh dishes, which ranged from two in Norway and three in Ireland to ten in

but found no significant correlation between the production of a spice in a given country and its consumption in that country. They also reject the adaptiveness of using spices to disguise spoilage, which they point out would increase the likelihood of food poisoning. They conclude that although the proximate reason for spice use is to make food more palatable, the "ultimate reason is most likely that spices help cleanse food of pathogens and thereby contribute to the health, longevity, and reproductive success of people who find their flavors enjoyable".

Food historians and other writers have often casually remarked that spices help preserve foods, so a systematic approach to the subject is most welcome. But correlational studies can only be as reliable as their data sets, and the data chosen in this study are not well suited to shed new light on spicy heat.

The large recipe database turns out to be a narrow source of evidence for the 'ultimate' purpose of an ancient habit across all cultures and historical periods. No cookbook consulted antedates 1945, by which time any biologically determined patterns of spice use have long been obscured by migrations of peoples and plants, technical advances in agriculture, transport and food handling, and

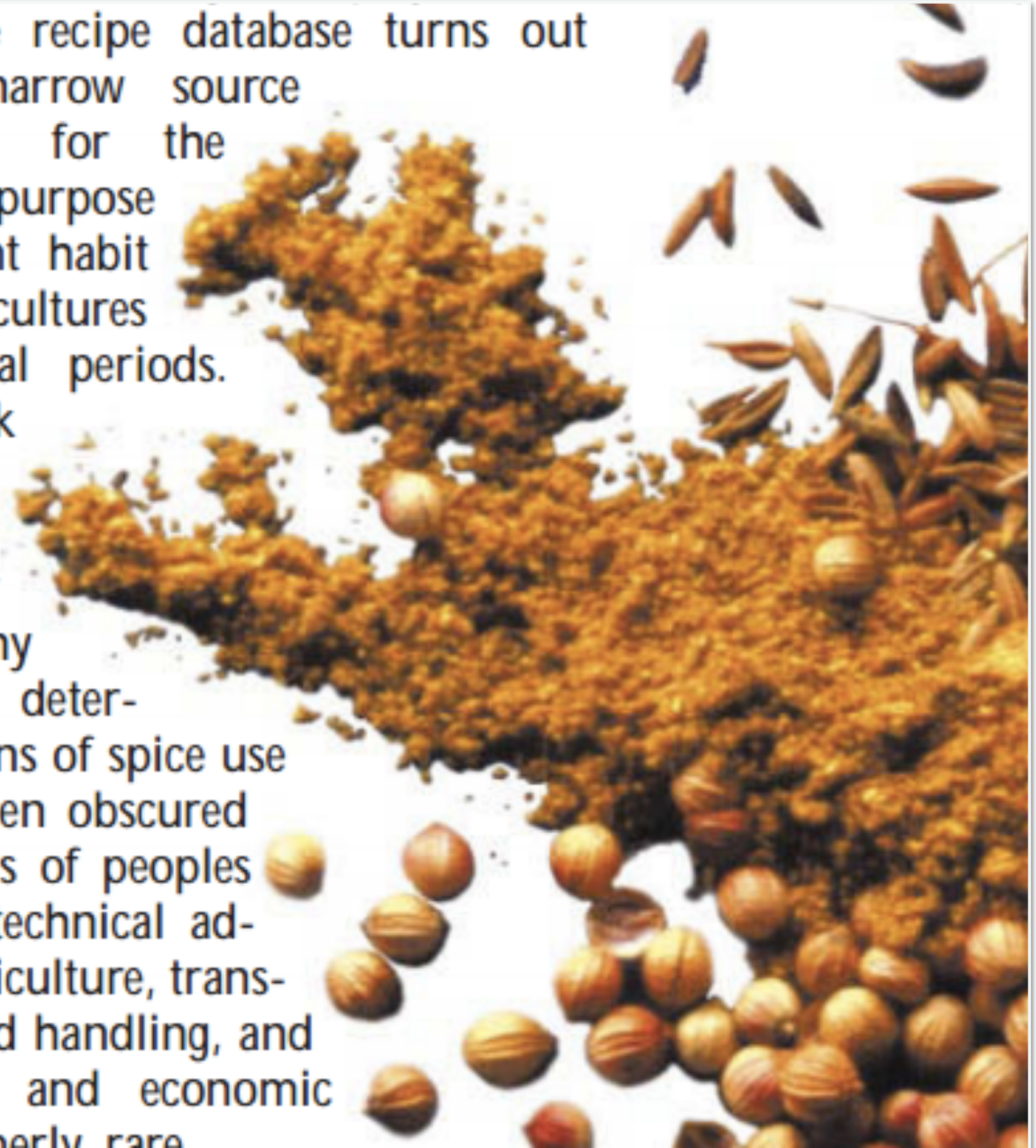
the null hypothesis survives pretty well unscathed.

A second limitation of the recipe database is that the authors consulted only English-language cookbooks. The ingredient lists in African and Asian books appear to have been modified for Western readers and their larders (no indigenous African spices make the list, nor do Japanese wasabi or shiso, or Chinese star anise or Szechwan pepper, which is neither *Piper* nor *Capsicum*). Together with the exclusive emphasis on meat and fish dishes, this Western orientation may also give undue weight to recipes of the affluent few who are best able and most likely to use ancillary ingredients. For example, an Indian study[2] found that rural labourers near Hyderabad tend to spice their predominantly cereal regimen more heavily than the urban middle-class did its broader diet, but with a more limited palette of chillies, tamarind and turmeric. In Billing and Sherman's tabulation of spice use, this sample of the average Indian diet would rank closer to Ireland than to the India of the cookbooks.

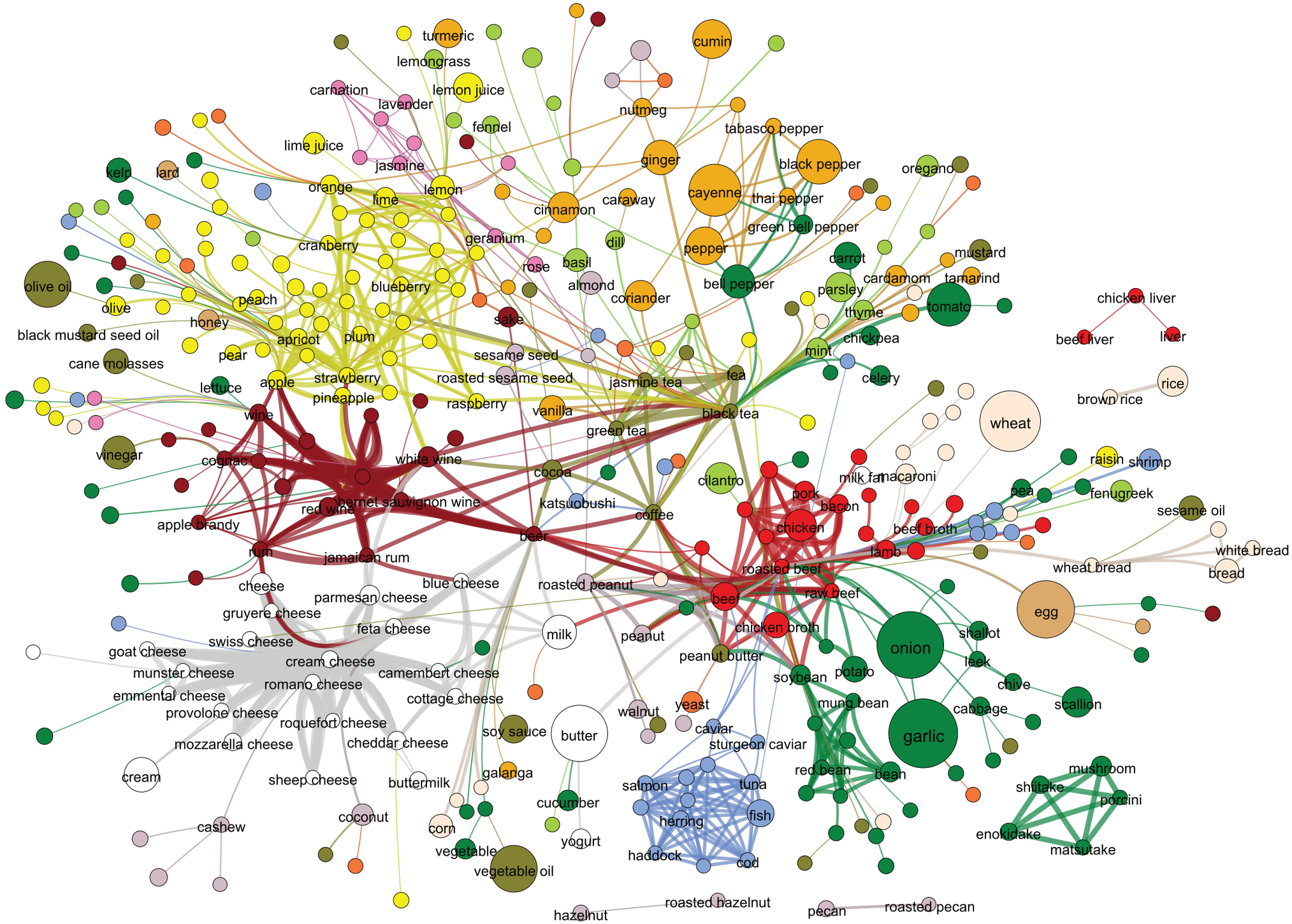
The most serious weakness in Billing and Sherman's case for their hygienic hypothesis is the way in which they represent the antibacterial activities of spices. Studies in this area have used a great variety of experimental conditions. The authors there-

ANTHONY BLAKE PHOTO LIBRARY/MVEREHURST

The large recipe database turns out to be a narrow source of evidence for the 'ultimate' purpose of an ancient habit across all cultures and historical periods. No cookbook consulted ante-dates 1945, by which time any biologically determined patterns of spice use have long been obscured by migrations of peoples and plants, technical advances in agriculture, transport and food handling, and other social and economic factors. Formerly rare



Are there any other **principles** that transcend individuals and cultures?

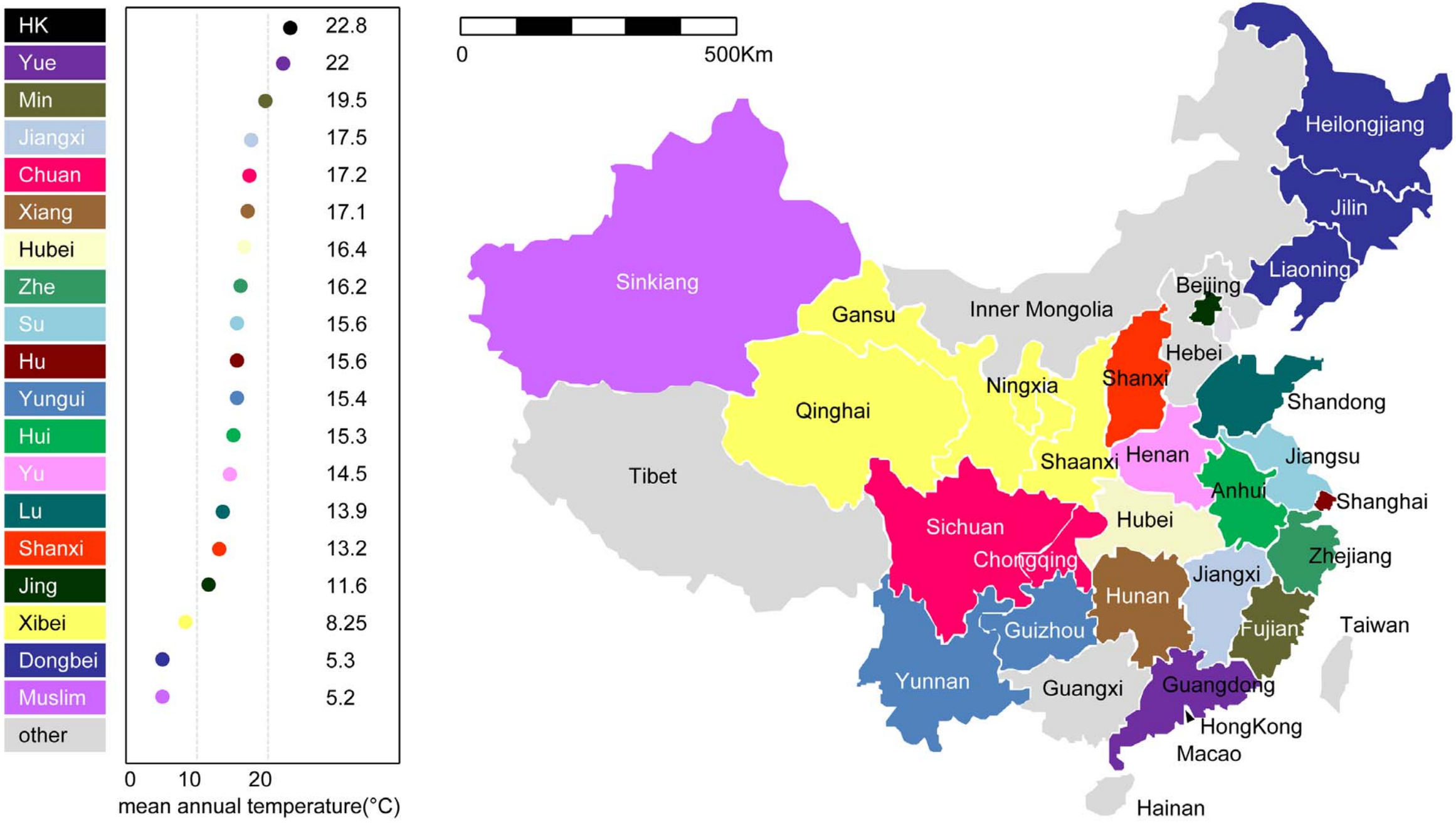

turmeric
lemongrass
carnation
lavender
lemon juice
cumin
nutmeg
fennel
tabasco pepper
lime juice
jasmine
ginger
black pepper
oregano
kelp
lard
orange
lime
lemon
cayenne
thai pepper
caraway
cinnamon
green bell pepper
geranium
dill
pepper
cranberry
rose
basil
carrot
mustard
olive oil
blueberry
almond
bell pepper
parsley
cardamom
tamarind
coriander
peach
olive
tomato
thyme
sake
honey
black mustard seed oil
apricot
plum
chickpea
pear
sesame seed
mint
cane molasses
tea
celery
roasted sesame seed
jasmine tea
lettuce
apple
strawberry
pineapple
raspberry
vanilla
black tea
wine
green tea
wheat
vinegar
cognac
white wine
cocoa
cilantro
milk fat
macaroni
raisin
shrimp
pea
fenugreek
red wine
bernet sauvignon wine
katsuobushi
coffee
pork
bacon
sesame oil
apple brandy
chicken
beef broth
bear
lamb
white bread
rum
jamaican rum
roasted beef
wheat bread
bread
cheese
blue cheese
roasted peanut
beef
raw beef
parmesan cheese
egg
gruyere cheese
feta cheese
milk
peanut
chicken broth
onion
shallot
swiss cheese
goat cheese
peanut butter
leek
cream cheese
camembert cheese
soybean
potato
munster cheese
romano cheese
chive
emmental cheese
cottage cheese
mung bean
cabbage
scallion
provolone cheese
yeast
walnut
roquefort cheese
soy sauce
butter
caviar
garlic
mozzarella cheese
cheddar cheese
sturgeon caviar
red bean
bean
cream
sheep cheese
buttermilk
galanga
salmon
tuna
mushroom
shiitake
porcini
cucumber
herring
fish
coconut
cashew
corn
yogurt
enokidake
matsutake
vegetable
haddock
cod
vegetable oil
roasted hazelnut
pecan
roasted pecan
hazelnut
chicken liver
beef liver
liver
brown rice
rice
Categories
fruits
dairy
spices
alcoholic beverages
nuts and seeds
seafoods
meats
herbs
plant derivatives
vegetables
flowers
animal products
plants
cereal
Prevalence
50 %
30 %
10 %
1 %
Shared compounds
150
50
10

What shapes recipes?

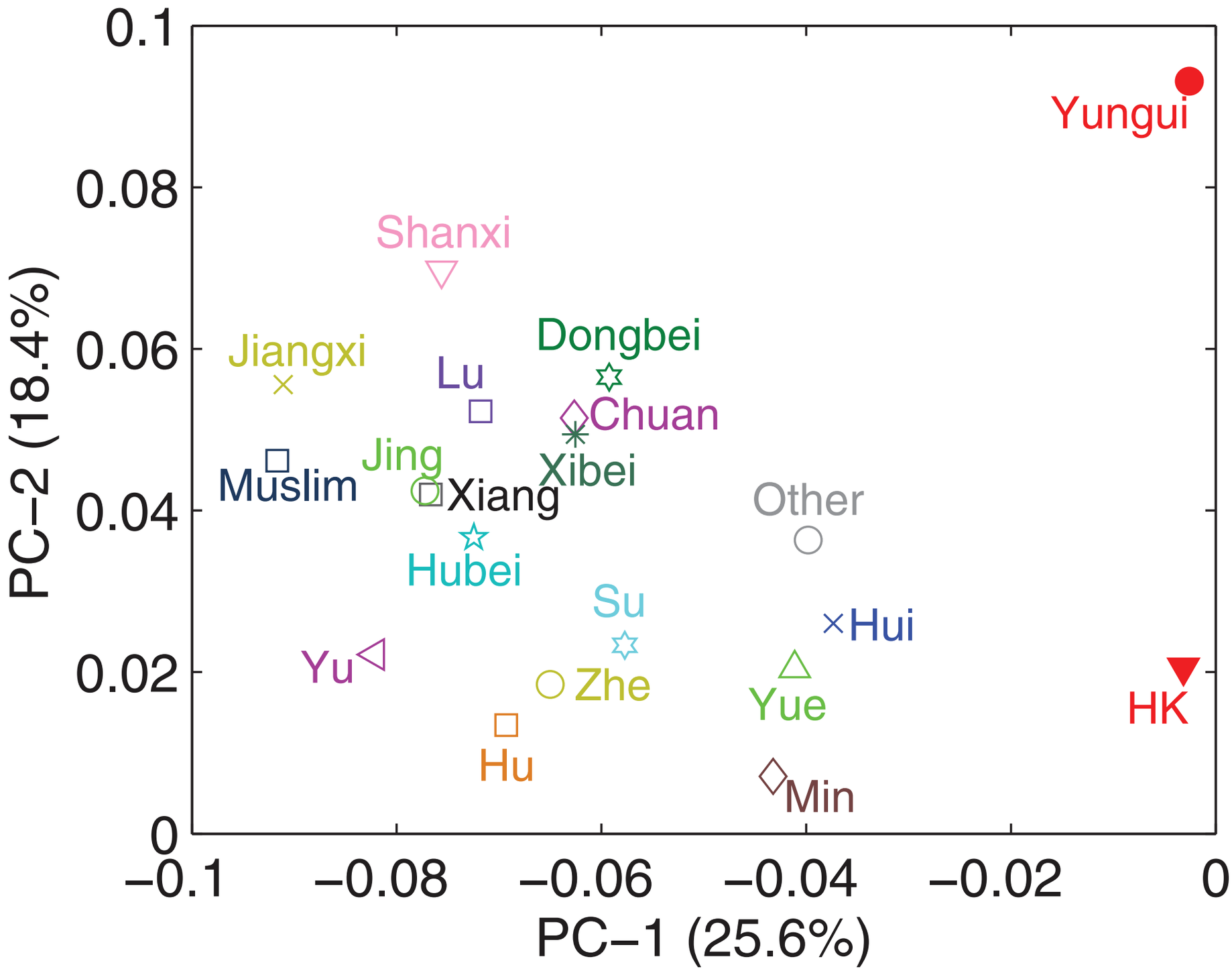
Climate?

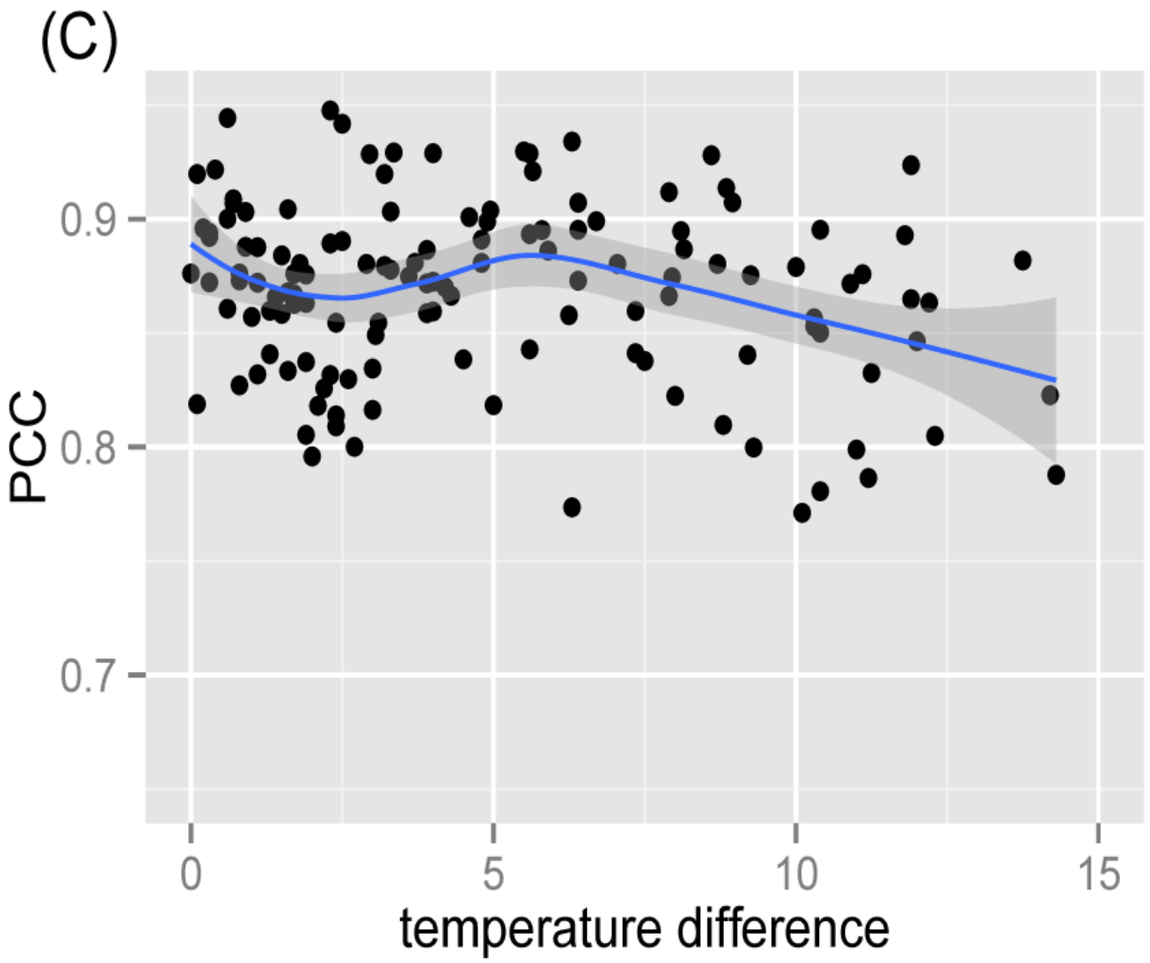
Cultural interactions?

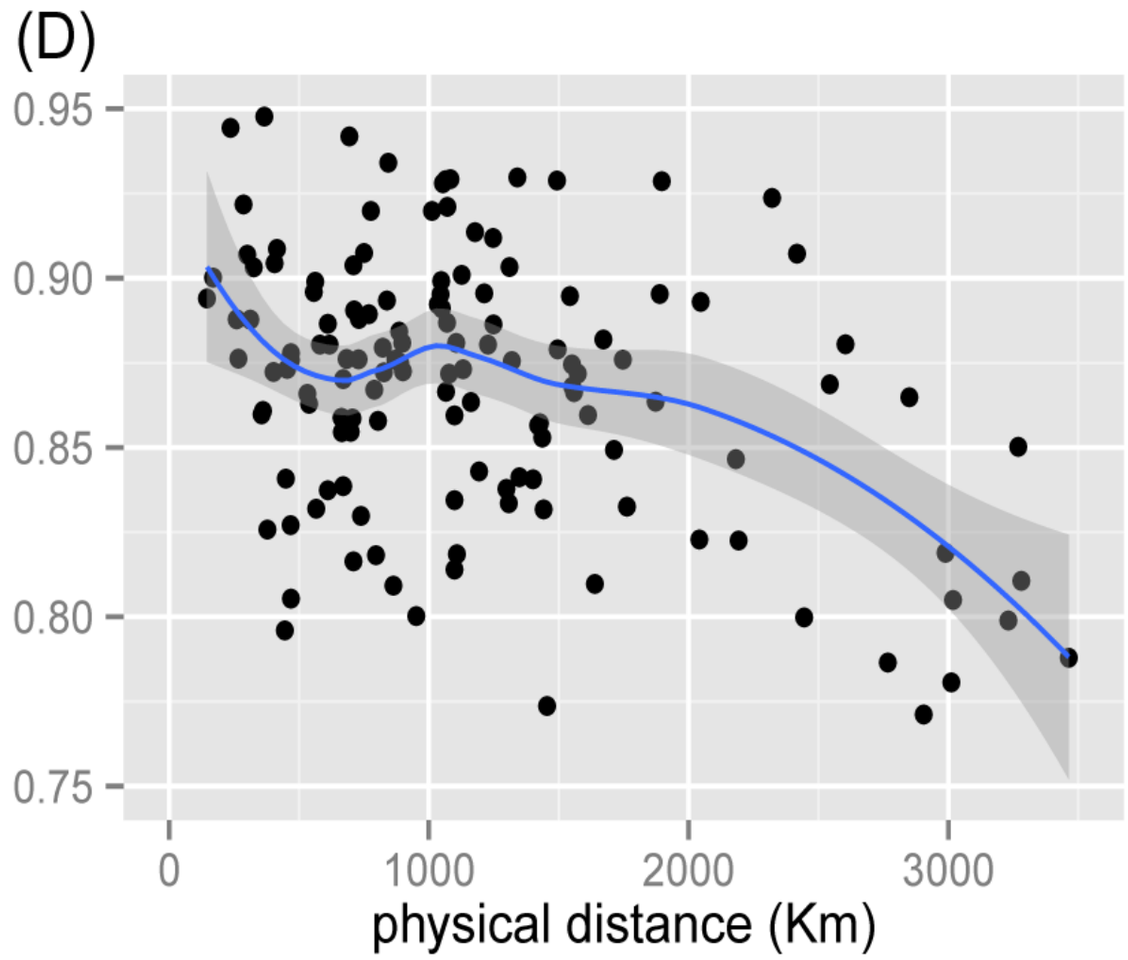
Let’s look at regional recipes



Represent each cuisine as a TF-IDF vector of ingredient usage frequency


0.1
0.08
0.06
0.04
0.02
0
PC–2 (18.4%)
–0.1
–0.08
–0.06
–0.04
–0.02
0
PC–1 (25.6%)
Yungui
Shanxi
Jiangxi
Lu
Dongbei
Chuan
Jing
Xibei
Muslim
Xiang
Other
Hubei
Su
Hui
Yu
Zhe
Yue
HK
Hu
Min


(C)
PCC
0.9
0.8
0.7
0
5
10
15
temperature difference


(D)
0.95
0.90
0.85
0.80
0.75
0
1000
2000
3000
physical distance (Km)

Temperature and proximity are correlated.

Partial correlation:

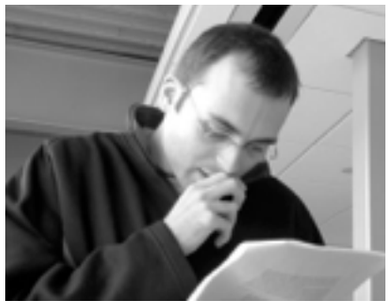
The effect of

temperature vanish.

The cultural diffusion seems to be an important driving factor.

What is “traditional” Korean (Italian, …) cuisine anyway?

http://yongyeol.com @yy



James P. Bagrow

Azadeh Nematzadeh

Emilio Ferrara

Alessandro Flammini

Sune Lehmann

Lilian Weng

Fil Menczer

Yu-Xiao Zhu

Sebastian Ahnert