Community detection

LEARNING OUTCOMES

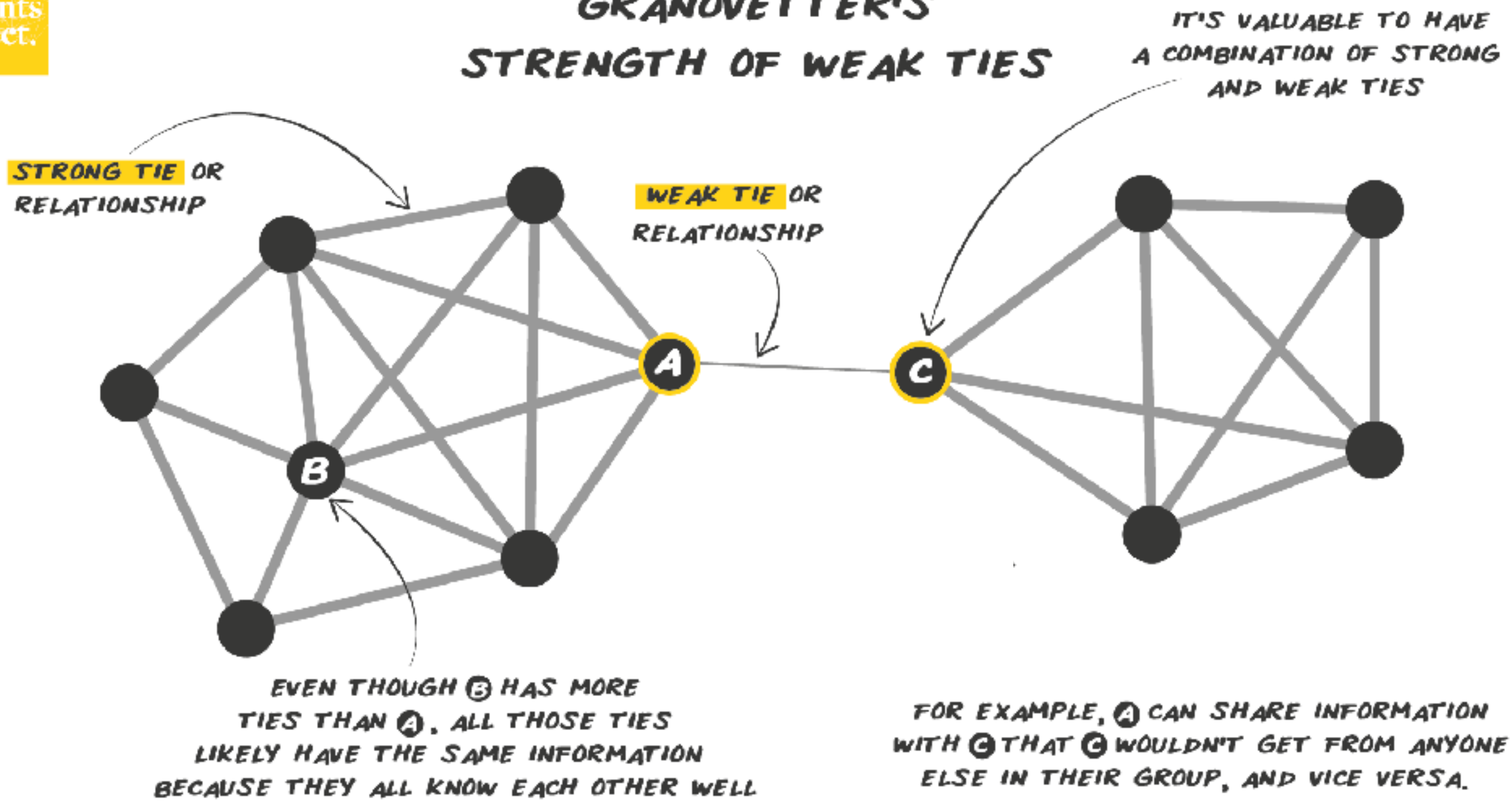
UNDERSTAND WHAT **COMMUNITIES** ARE

BE ABLE TO **DESCRIBE A NETWORK** IN TERMS OF COMMUNITIES

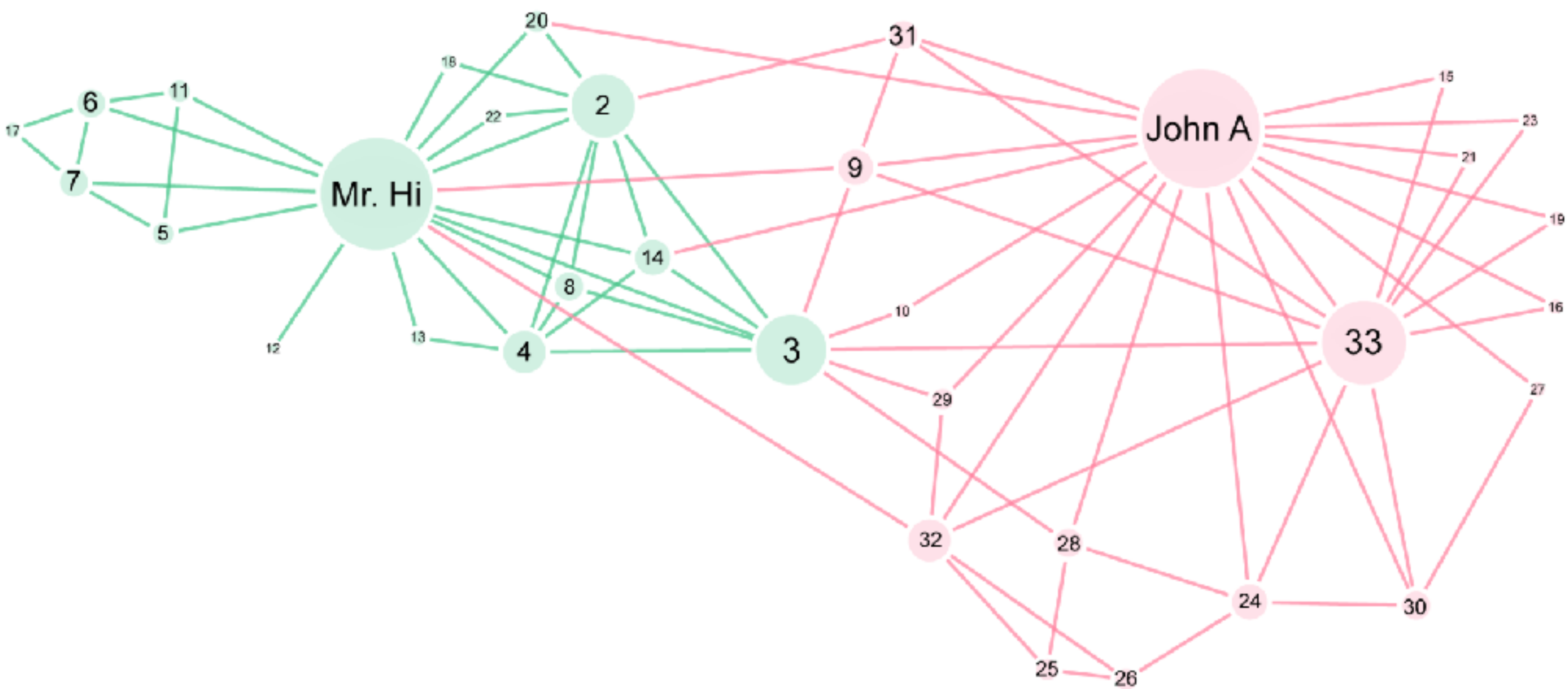
LEARN DIFFERENT TYPES OF COMMUNITY **CLASSIFICATIONS**

The Reliants Project.

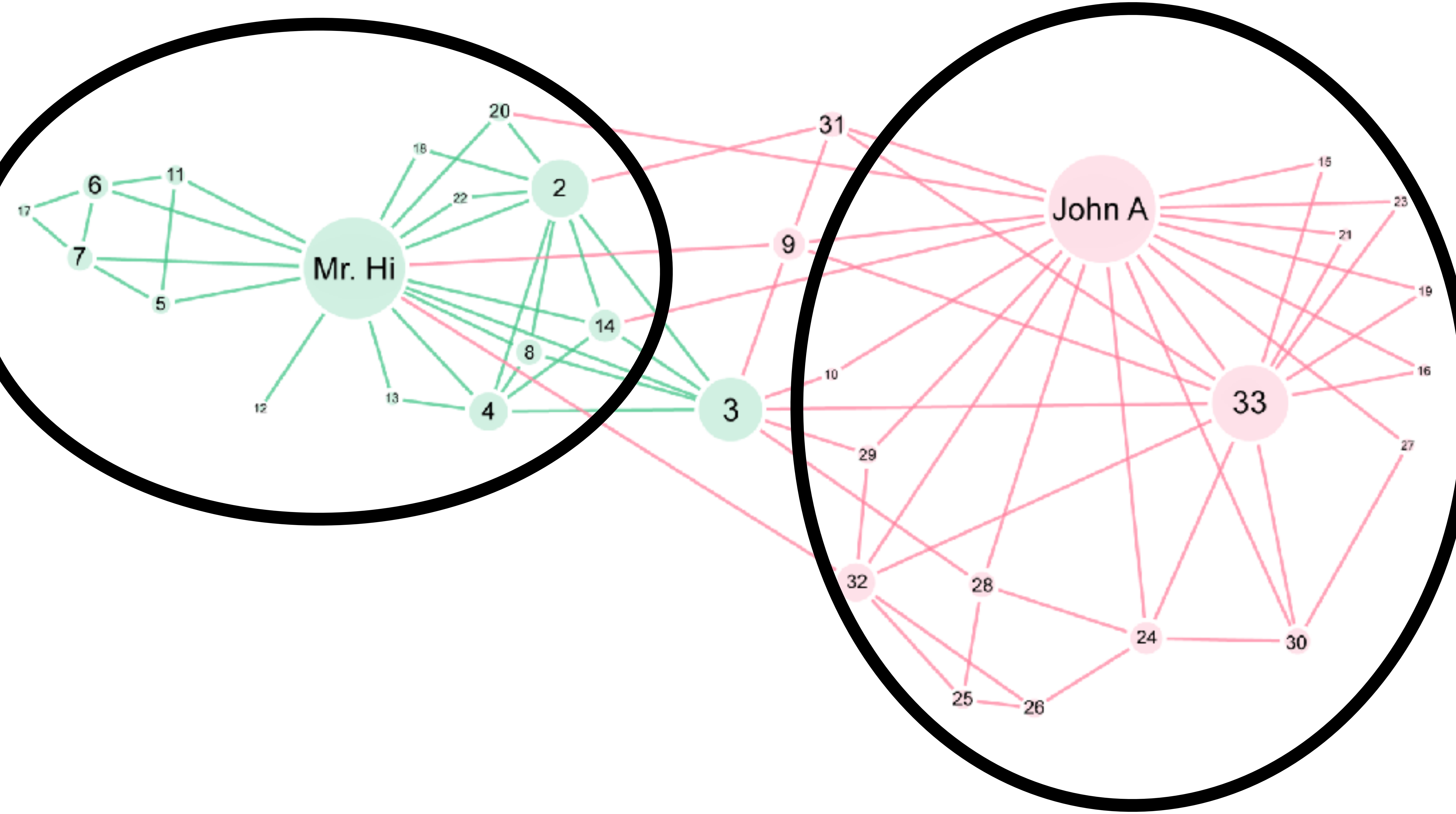
GRANOVETTER'S STRENGTH OF WEAK TIES




ALL VALLEY KARATE CHAMPIONSHIP
SEMI-FINALS
FINALS



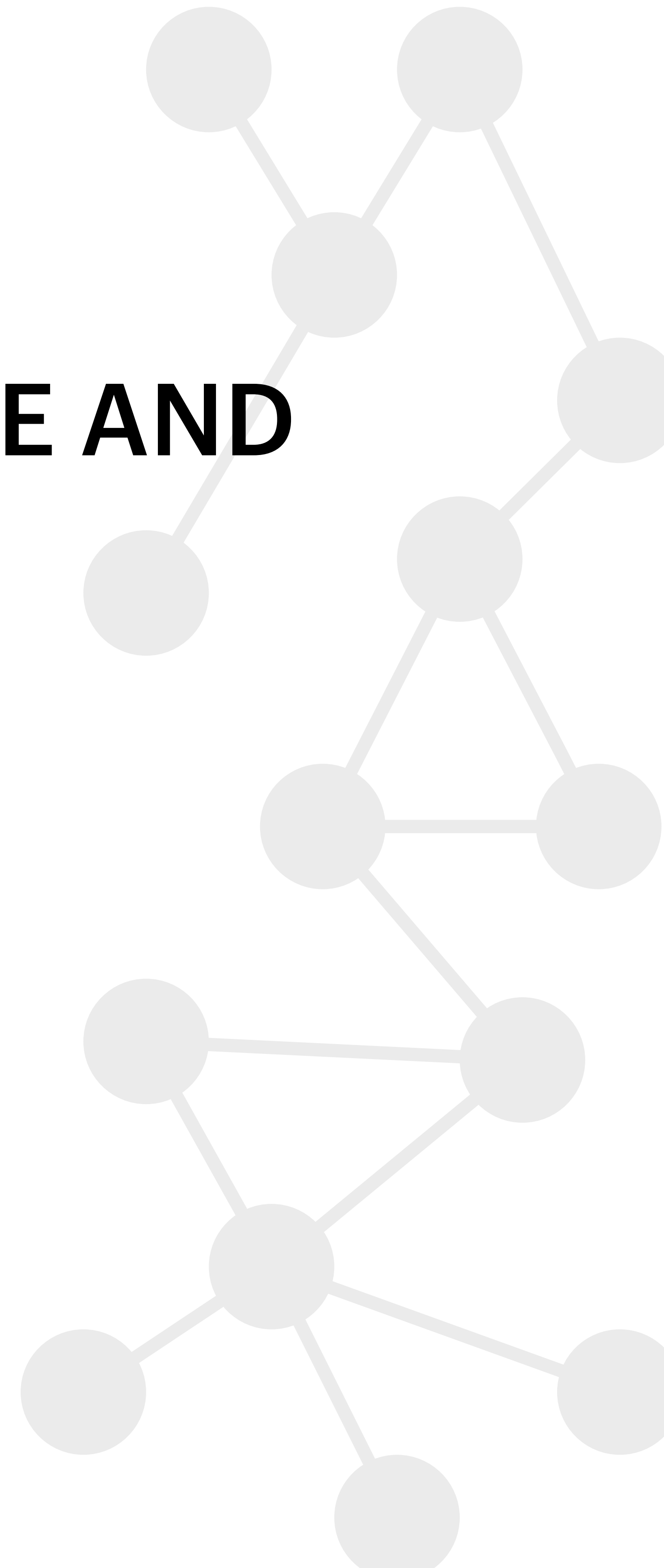
20
31
18
6
11
2
22
John A
17
15
23
21
7
9
Mr. Hi
19
5
14
8
16
10
33
12
13
4
3
27
29
32
28
24
30
25
26



20
31
18
6
11
2
22
John A
15
23
17
21
7
9
Mr. Hi
5
19
14
8
16
10
12
13
4
3
33
29
27
32
28
24
30
25
26

DEFINITIONS

INTERNAL AND EXTERNAL DEGREE:
THE NUMBER OF NEIGHBOURS INSIDE AND OUTSIDE THE COMMUNITY



DEFINITIONS

i is called internal node of community **c** if

$$k_i^{ext} = 0 \text{ **And** } k_i^{int} > 0$$

i is called boundary node of community **c** if

$$k_i^{ext} > 0 \text{ **And** } k_i^{int} > 0$$

DEFINITIONS

INTERNAL AND EXTERNAL DEGREE:
THE NUMBER OF NEIGHBOURS INSIDE AND OUTSIDE THE COMMUNITY

$$k_i = k_i^{int} + k_i^{ext}$$

DEFINITIONS

INTERNAL AND EXTERNAL DEGREE:
THE NUMBER OF NEIGHBOURS INSIDE AND OUTSIDE THE COMMUNITY

NUMBER OF INTERNAL LINKS:
THE NUMBER OF LINKS BETWEEN NODES WITHIN THE COMMUNITY

DEFINITIONS

INTERNAL AND EXTERNAL DEGREE:

THE NUMBER OF NEIGHBOURS INSIDE AND OUTSIDE THE COMMUNITY

NUMBER OF INTERNAL LINKS:

THE NUMBER OF LINKS BETWEEN NODES WITHIN THE COMMUNITY

COMMUNITY DEGREE:

THE SUM OF DEGREE OF ALL THE NODES IN THE COMMUNITY

DEFINITIONS

COMMUNITY DEGREE:
THE SUM OF DEGREE OF ALL THE NODES IN THE COMMUNITY

$$k_C = \sum_{i \in C} k_i$$

DEFINITIONS

INTERNAL AND EXTERNAL DEGREE:
THE NUMBER OF NEIGHBOURS INSIDE AND OUTSIDE THE COMMUNITY

NUMBER OF INTERNAL LINKS:
THE NUMBER OF LINKS BETWEEN NODES WITHIN THE COMMUNITY

COMMUNITY DEGREE:
THE SUM OF DEGREE OF ALL THE NODES IN THE COMMUNITY

INTERNAL LINK DENSITY:
DENSITY THAT CONSIDERS ONLY LINKS BETWEEN MEMBERS OF THE COMMUNITY

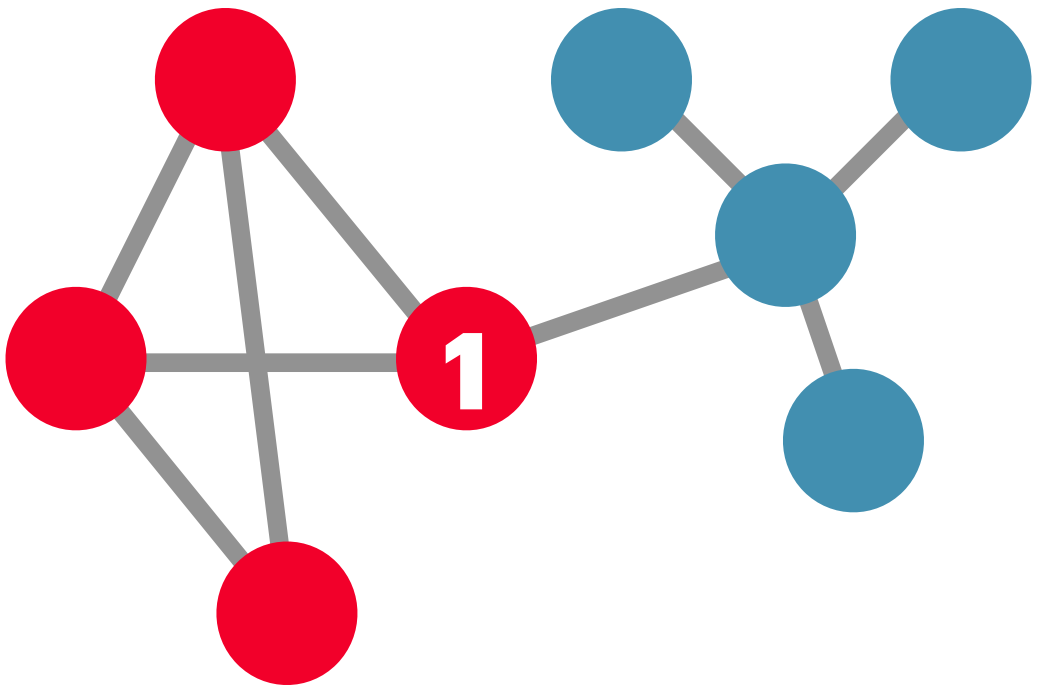
DEFINITIONS

INTERNAL LINK DENSITY:

DENSITY THAT CONSIDERS ONLY LINKS BETWEEN MEMBERS OF THE COMMUNITY

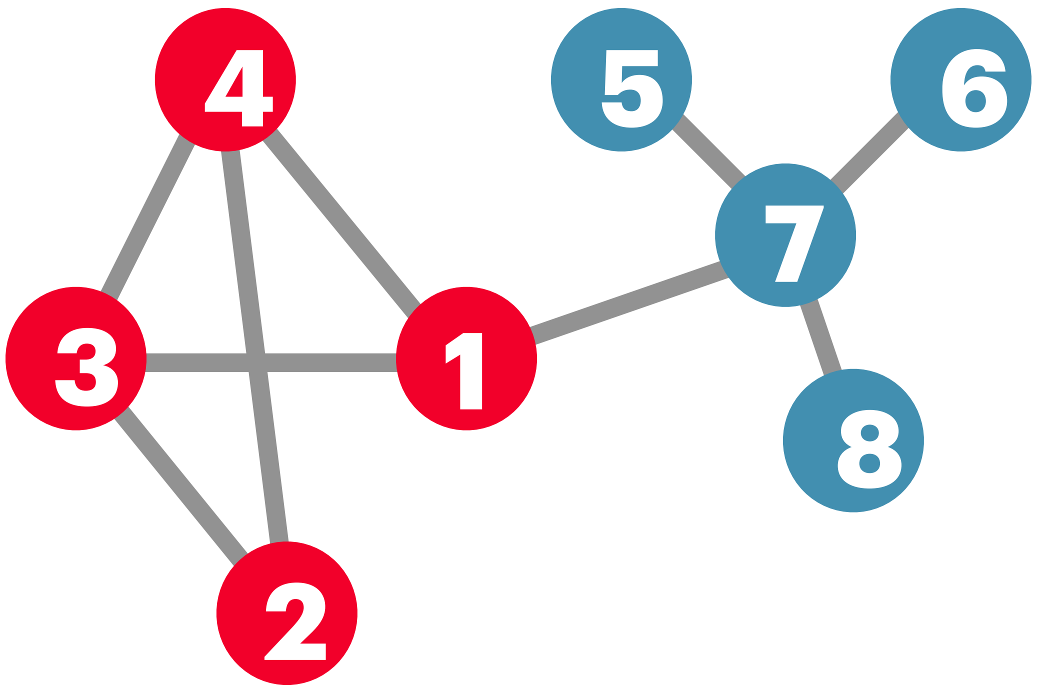
$$\delta_C^{int} = \frac{L_C}{L_C^{max}} = \frac{2L_C}{N_C(N_C - 1)}$$

DEFINITIONS



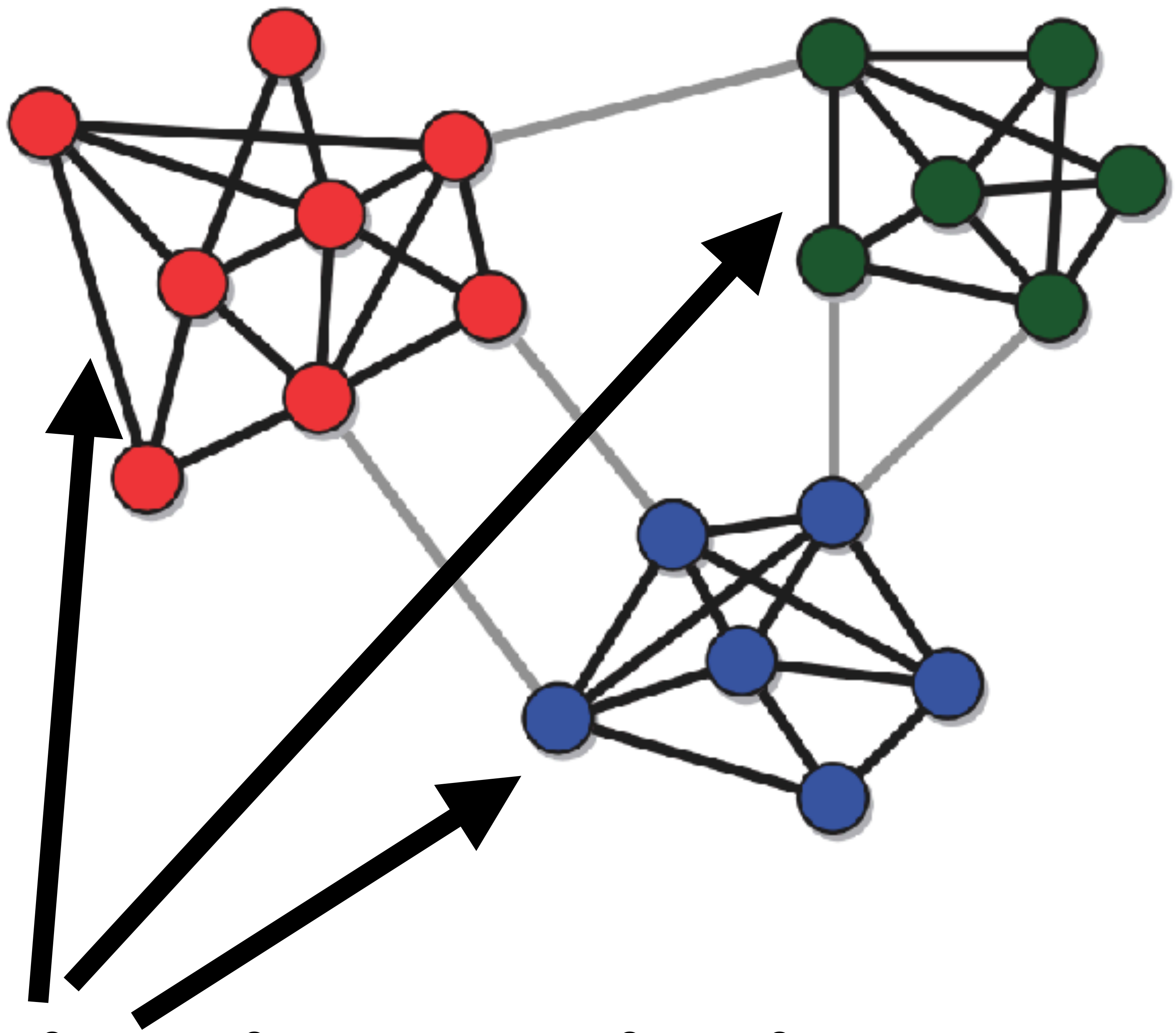
? $k_1^{ext}, k_1^{int}, \delta_{red}^{int}, k_{blue}$

DEFINITIONS



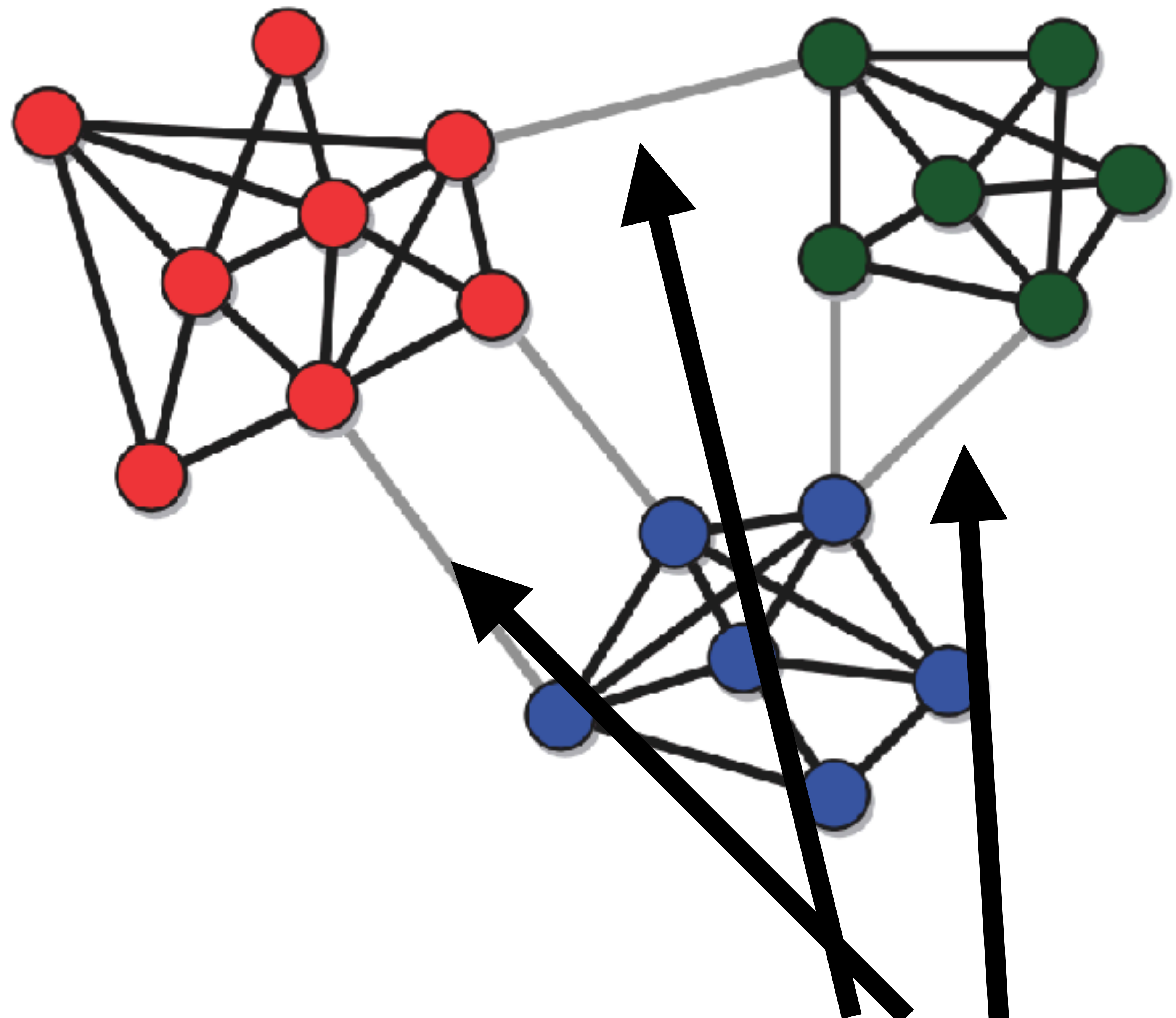
Which nodes are the boundary nodes?

DEFINITIONS



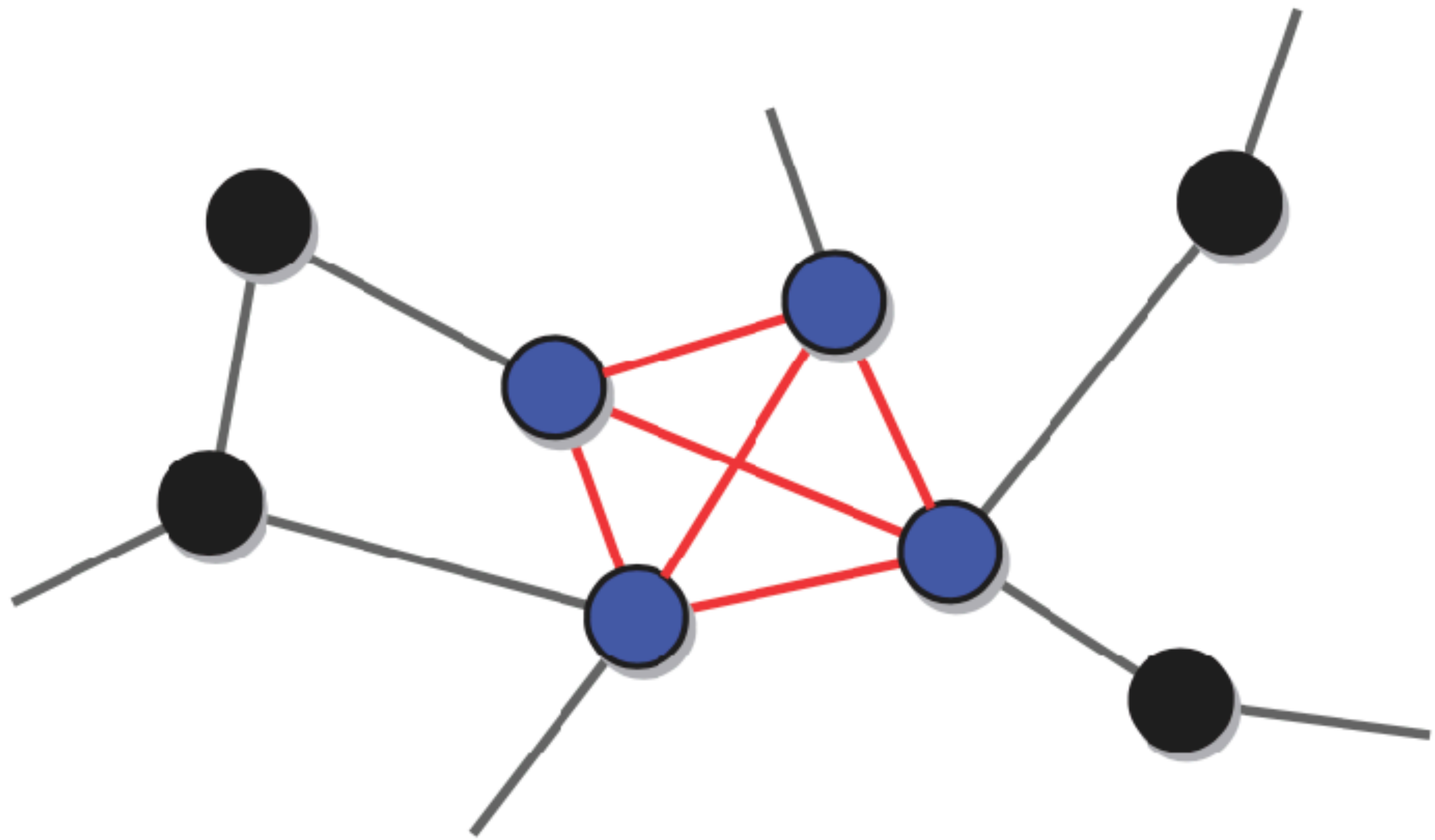
high cohesion, high separation

DEFINITIONS



high cohesion, high separation

DEFINITIONS



clique (a fully connected subgraph)

DEFINITIONS

Stricter definition of community

Strong community: $\forall i \in C : k_i^{int} > k_i^{ext}$

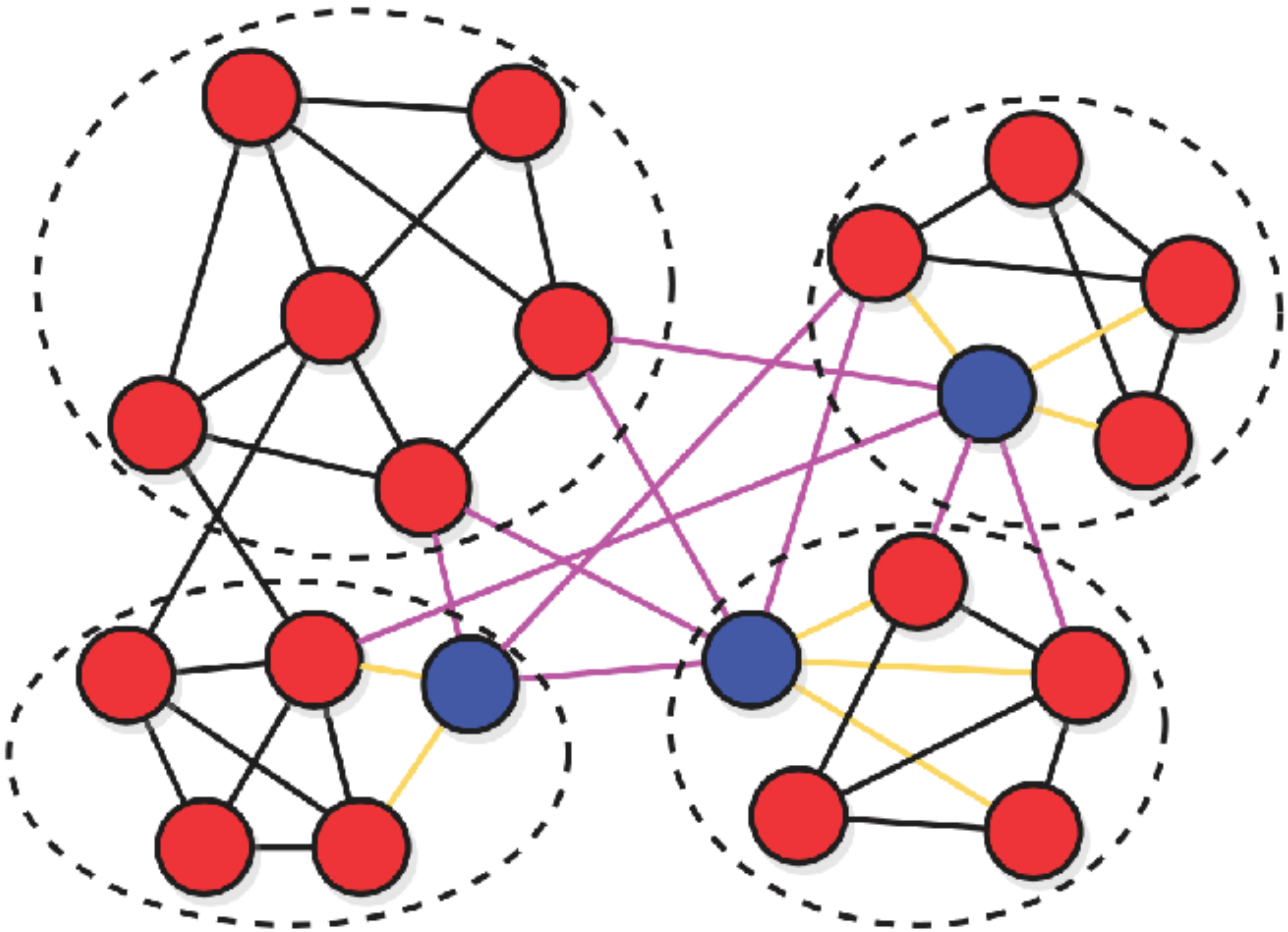
Weak community: $\sum_{i \in C} k_i^{int} > \sum_{i \in C} k_i^{ext}$

DEFINITIONS

Less strict definition of community

Strong community: $\forall i \in C, \forall v \in V : k_i^{int} > k_i^{ext,v}$

Weak community: $\forall v \in V : \sum_{i \in C} k_i^{int} > \sum_{i \in C} k_i^{ext,v}$



Strong and weak communities. The four subnetworks enclosed in the dashed contours are weak communities according to both definitions we have given. They are also strong communities according to the less stringent definition, as the internal degree of each node exceeds the number of links joining the node with those of every other community. However, three of the subnetworks are not strong communities in the more stringent sense, because some nodes (in blue) have external degree larger than their internal degree (the internal and external links of these nodes are colored in yellow and magenta, respectively). Adapted from Fortunato and Hric (2016).

PARTITIONS

A PARTITION IS A DIVISION OF THE NETWORK IN COMMUNITIES

PARTITIONS

SUPPOSE YOU HAVE A NETWORK **G** WITH **10 NODES** 1,2,…,10

{1,2,…,10}

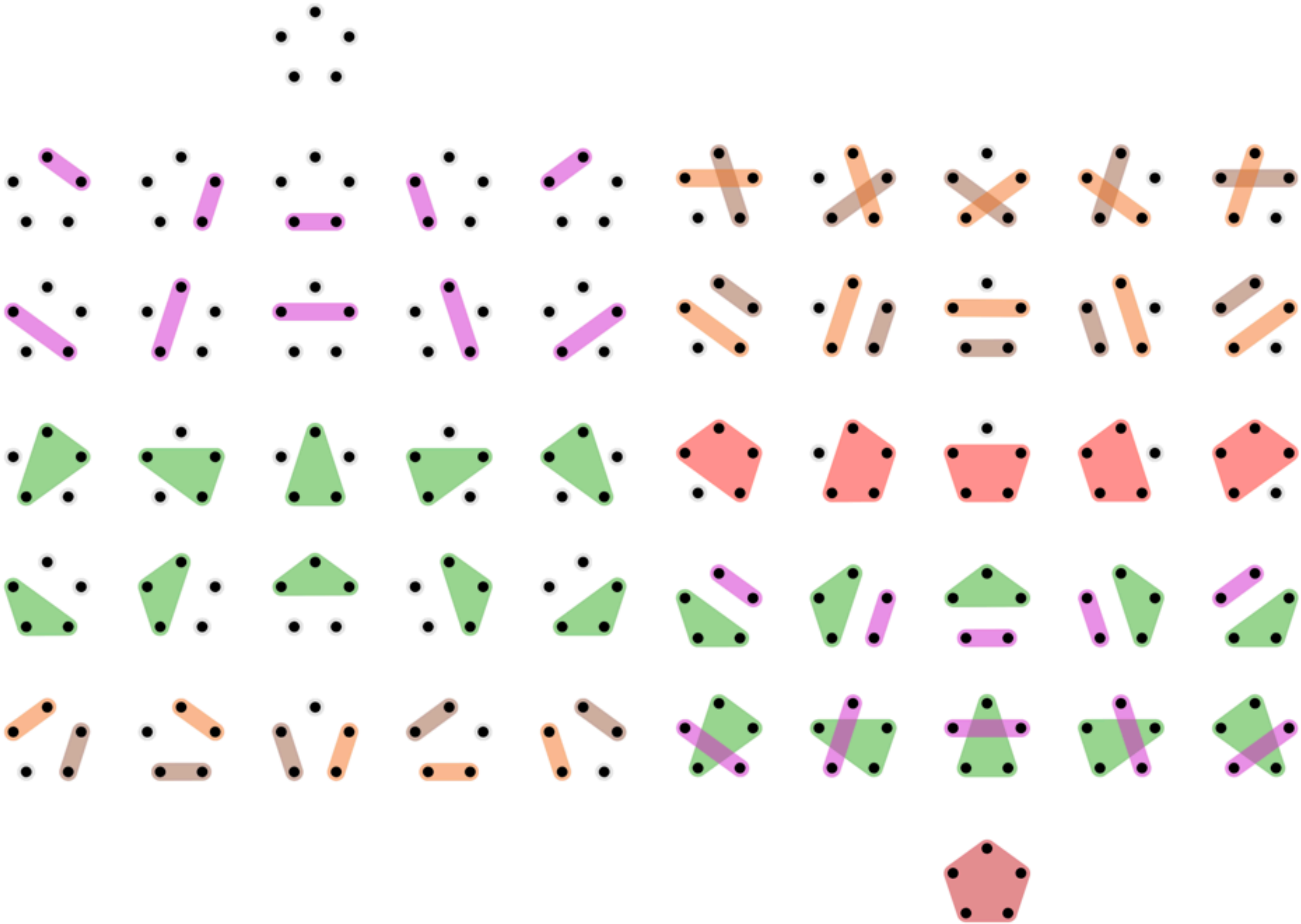
{1} {2} {3} … {10}

{1,2} {3,6,9} {5,8,10} {7,4}

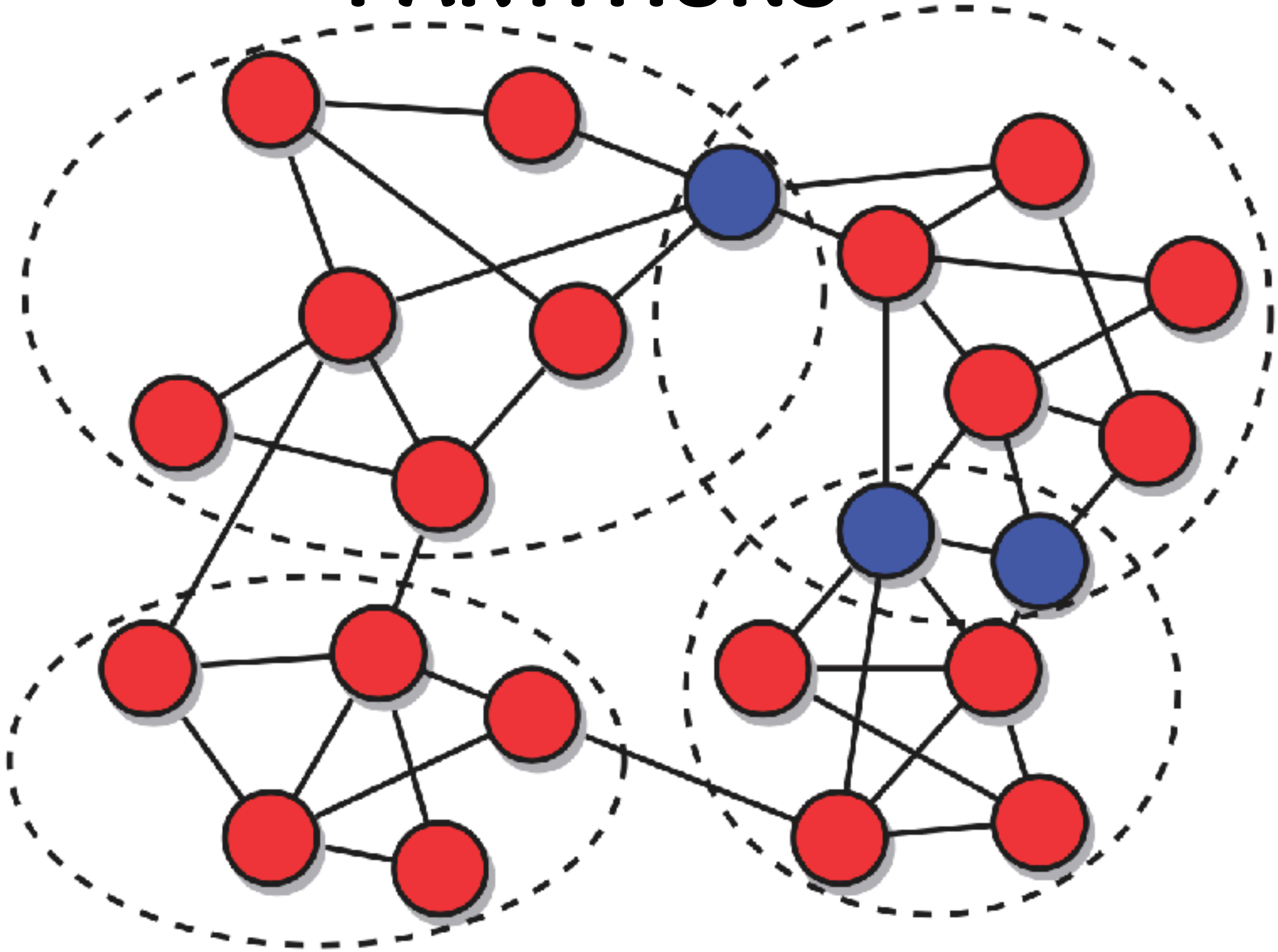
THESE ARE ALL VALID PARTITIONS

The number of possible partitions grows super-exponentially

PARTITIONS



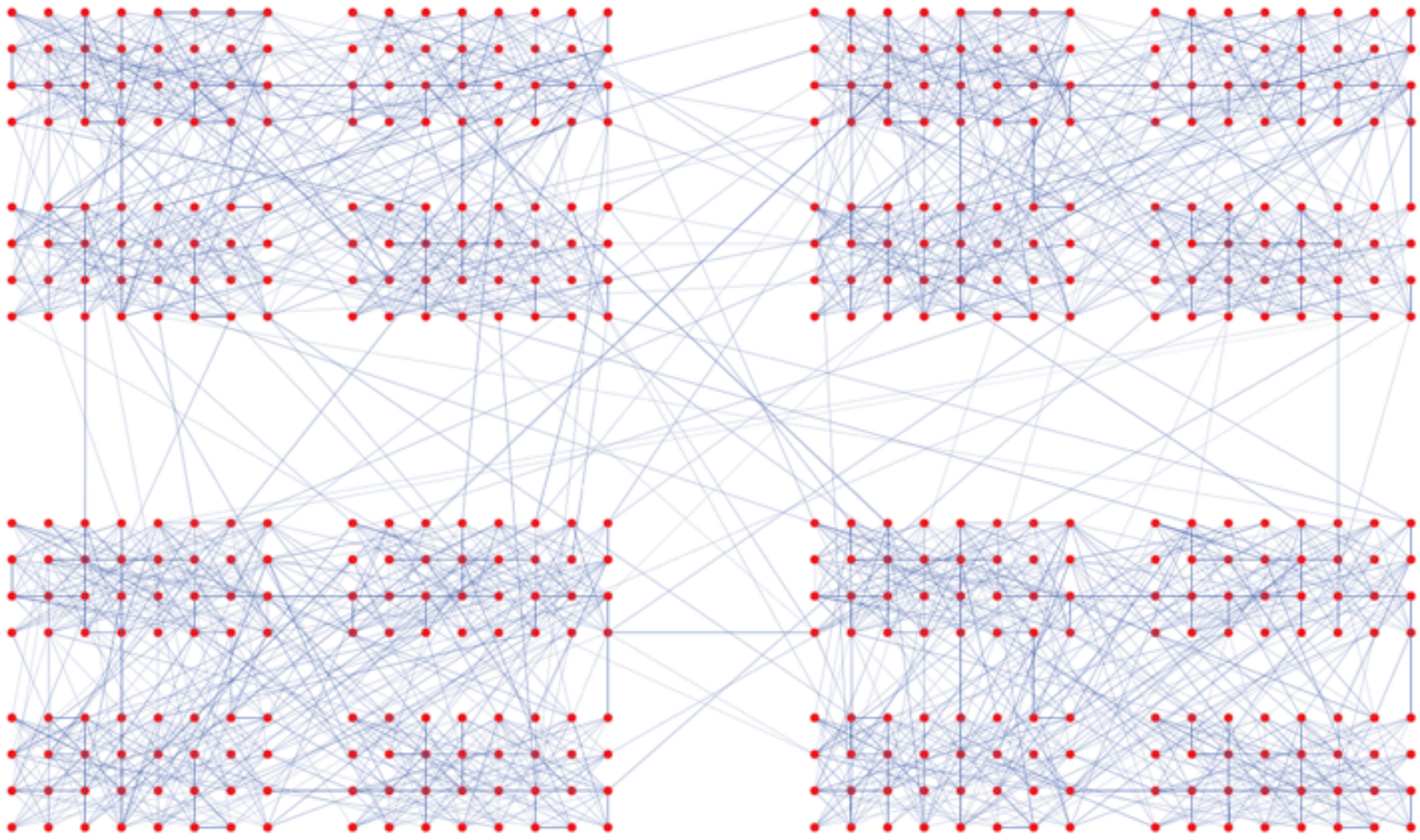
PARTITIONS



COMMUNITIES CAN OVERLAP

(You are part of different communities, think about it)

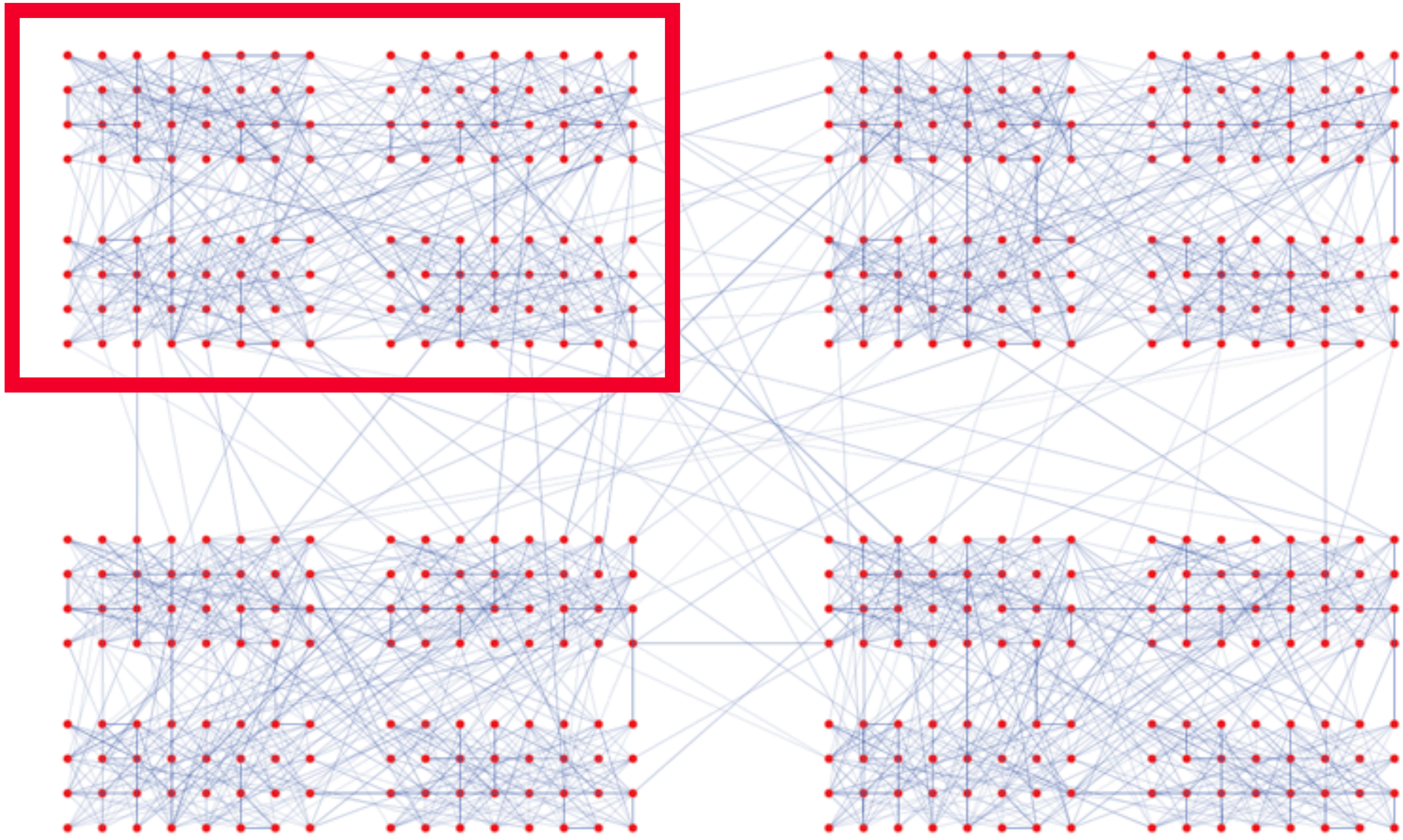
PARTITIONS



COMMUNITIES CAN BE HIERARCHICAL

(There might be communities within communities)

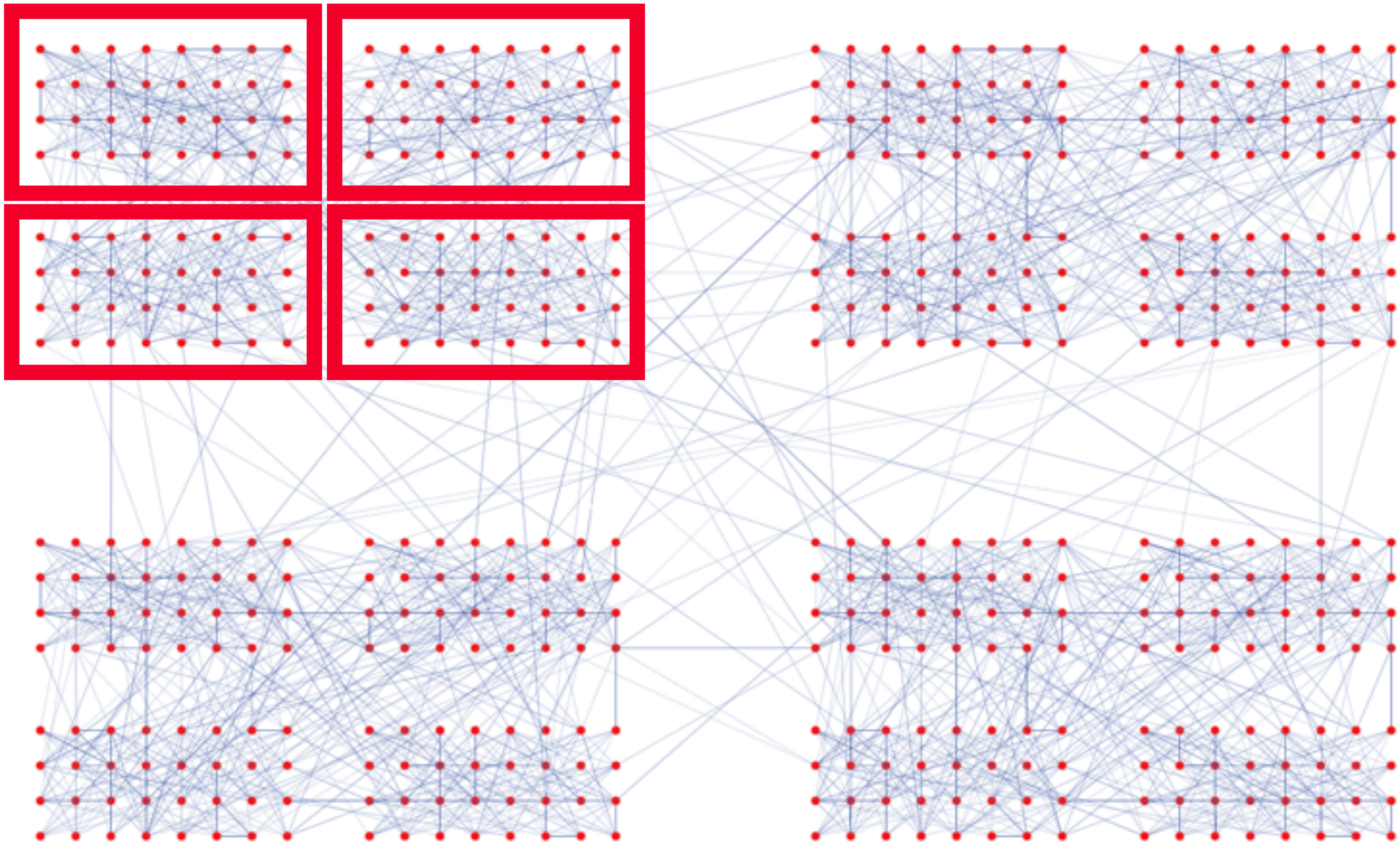
PARTITIONS



COMMUNITIES CAN BE HIERARCHICAL

(There might be communities within communities)

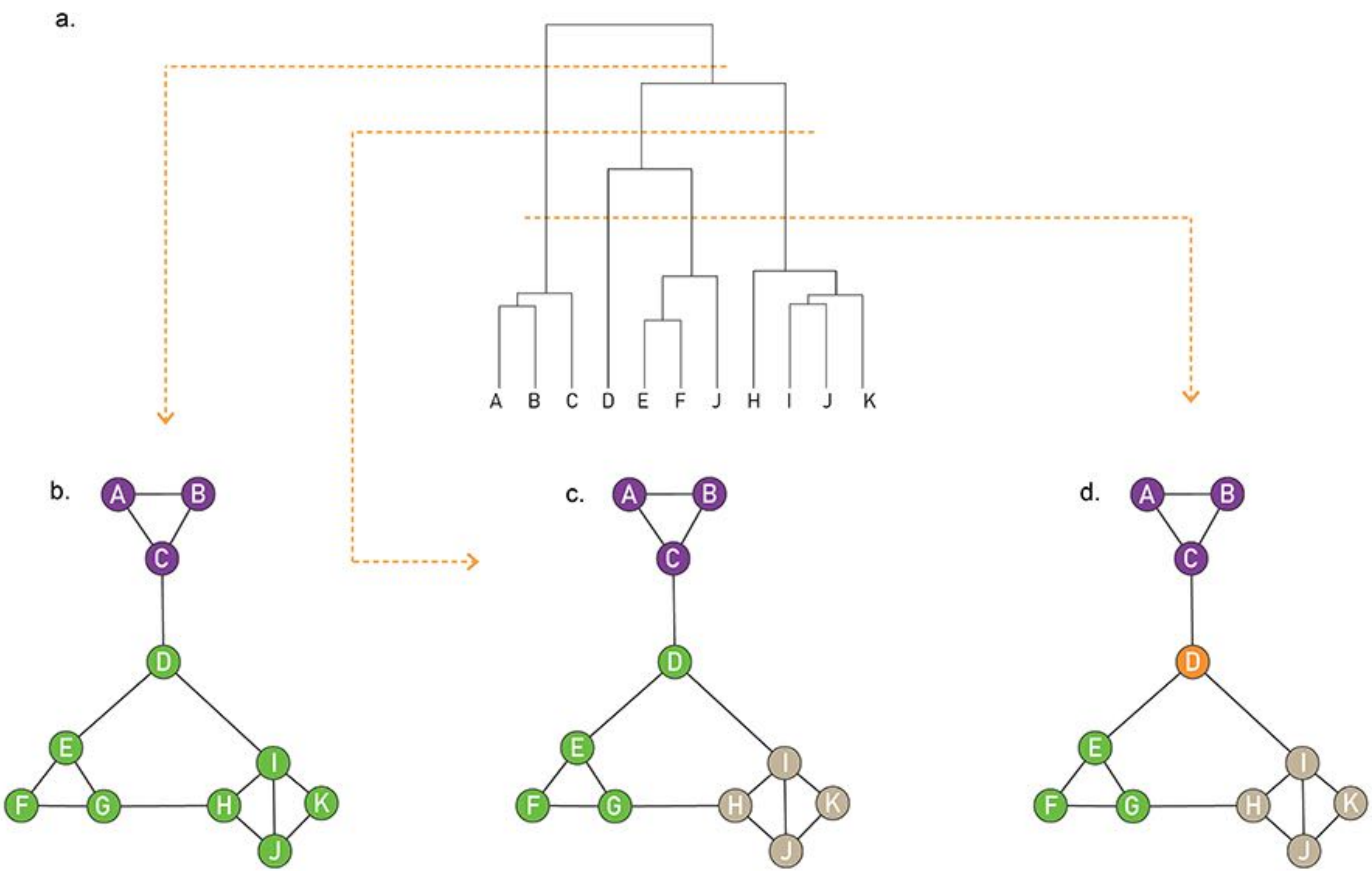
PARTITIONS



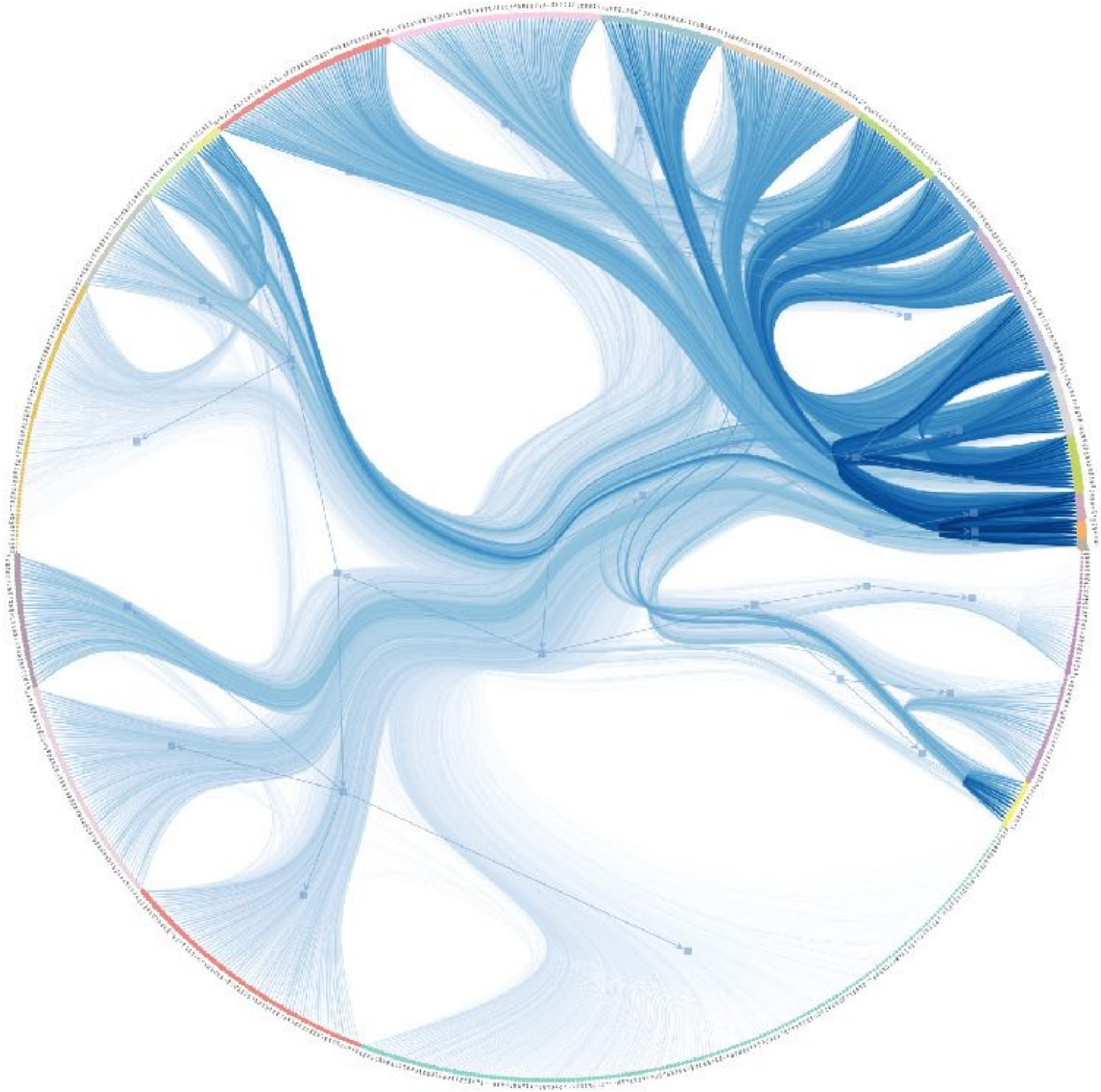
COMMUNITIES CAN BE HIERARCHICAL

(There might be communities within communities)

PARTITIONS

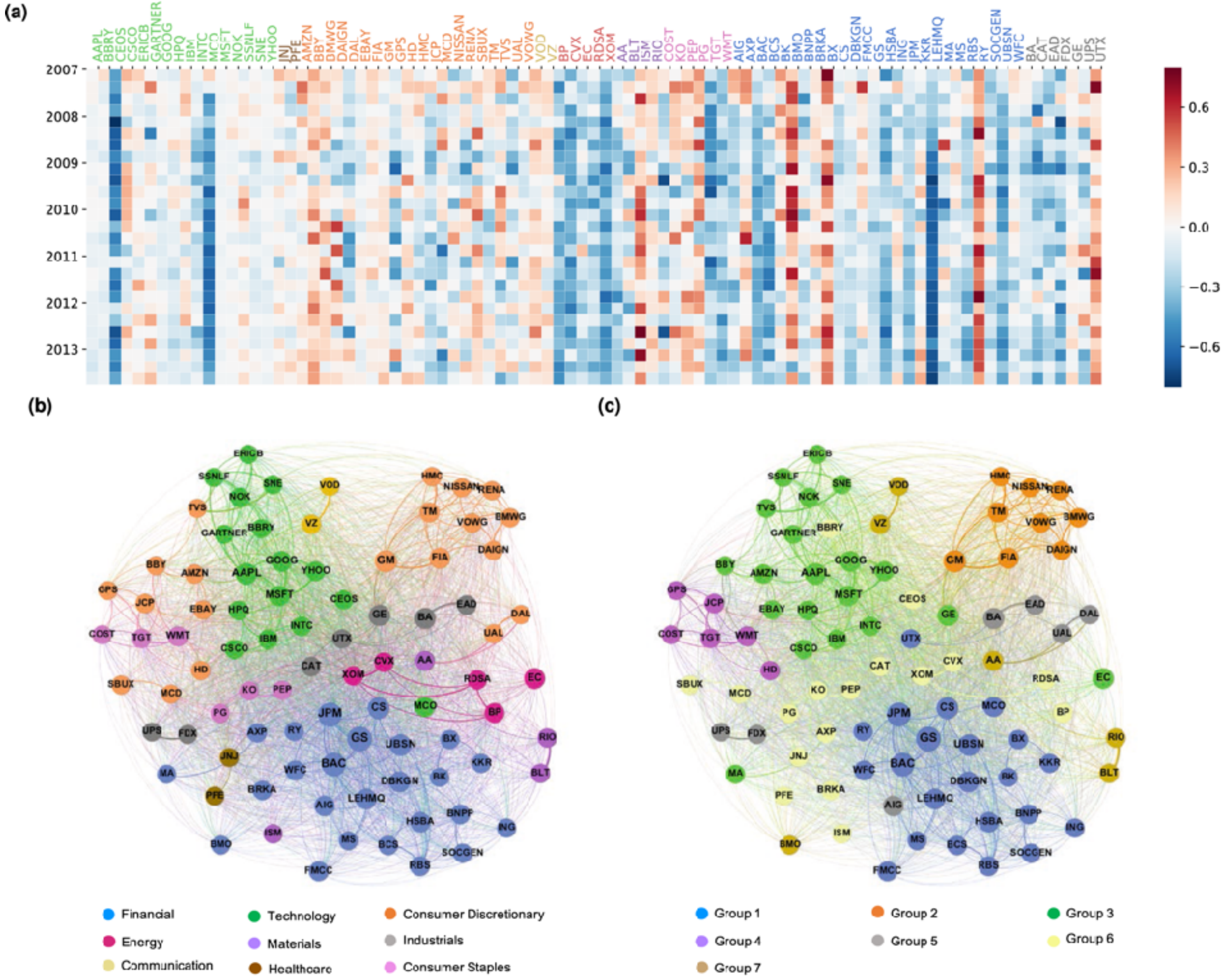


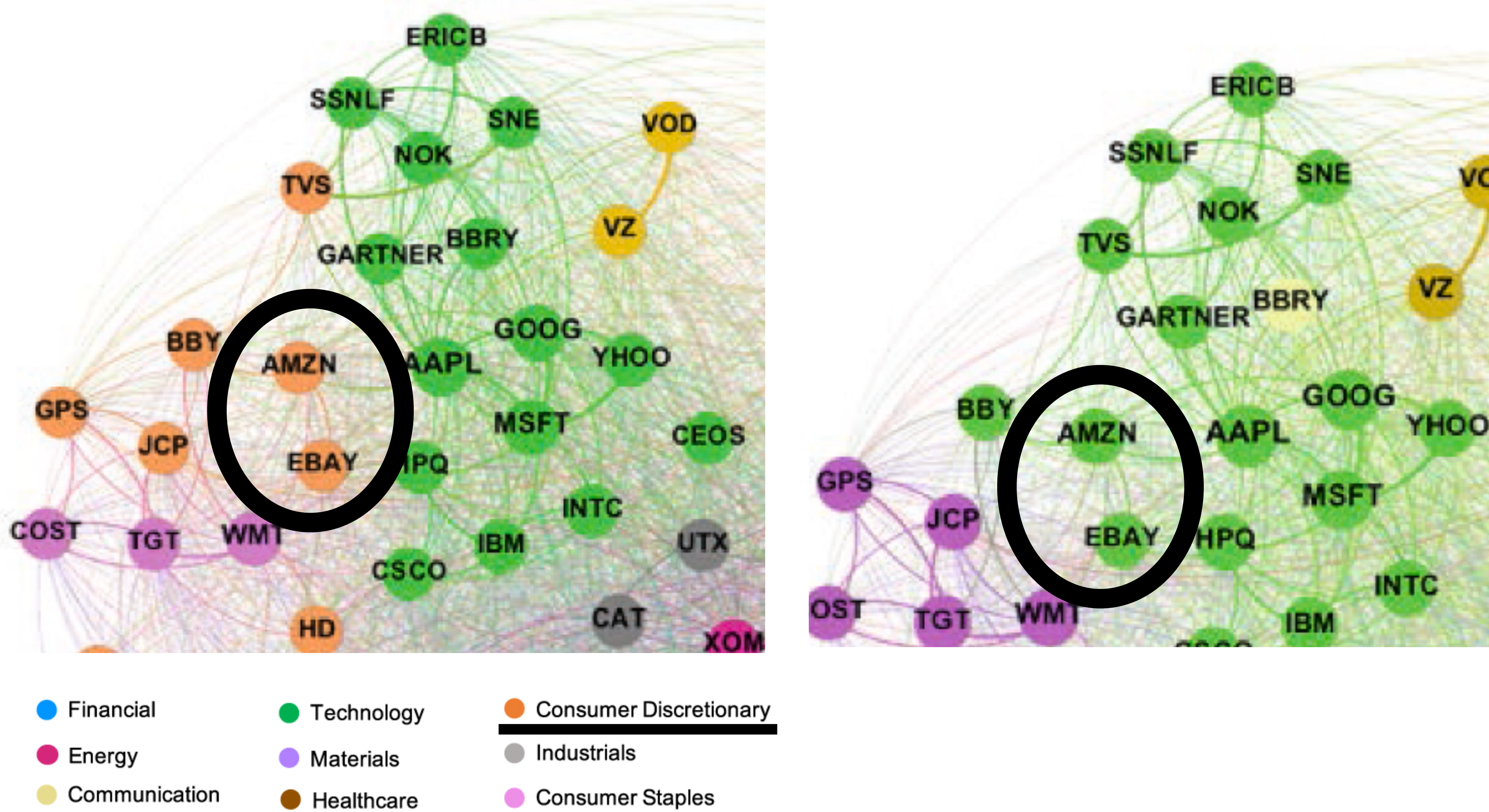
Dendrogram



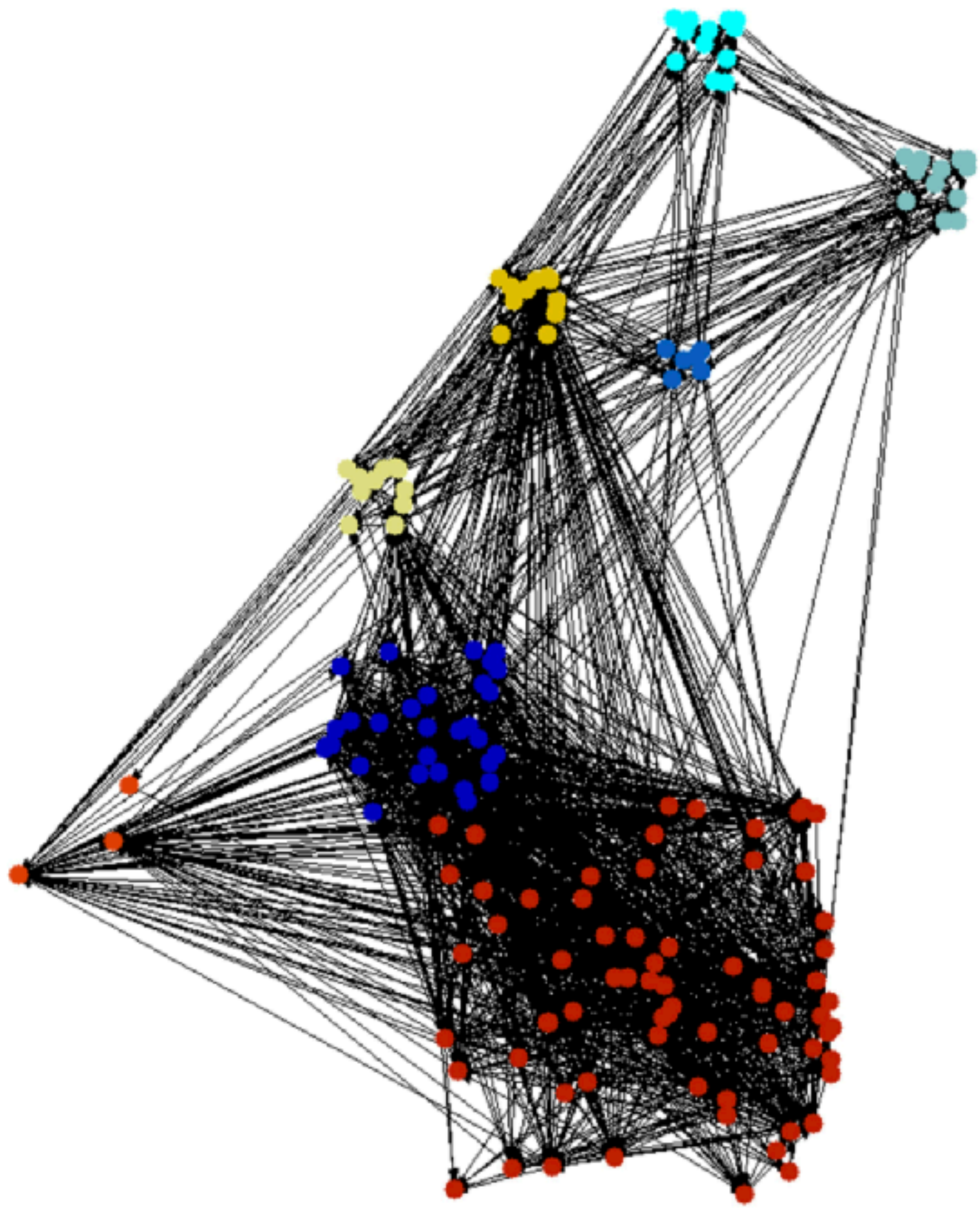
EXERCISE

MAKE SOME EXAMPLES
OF SOCIAL AND FINANCIAL
NETWORKS WITH COMMUNITIES

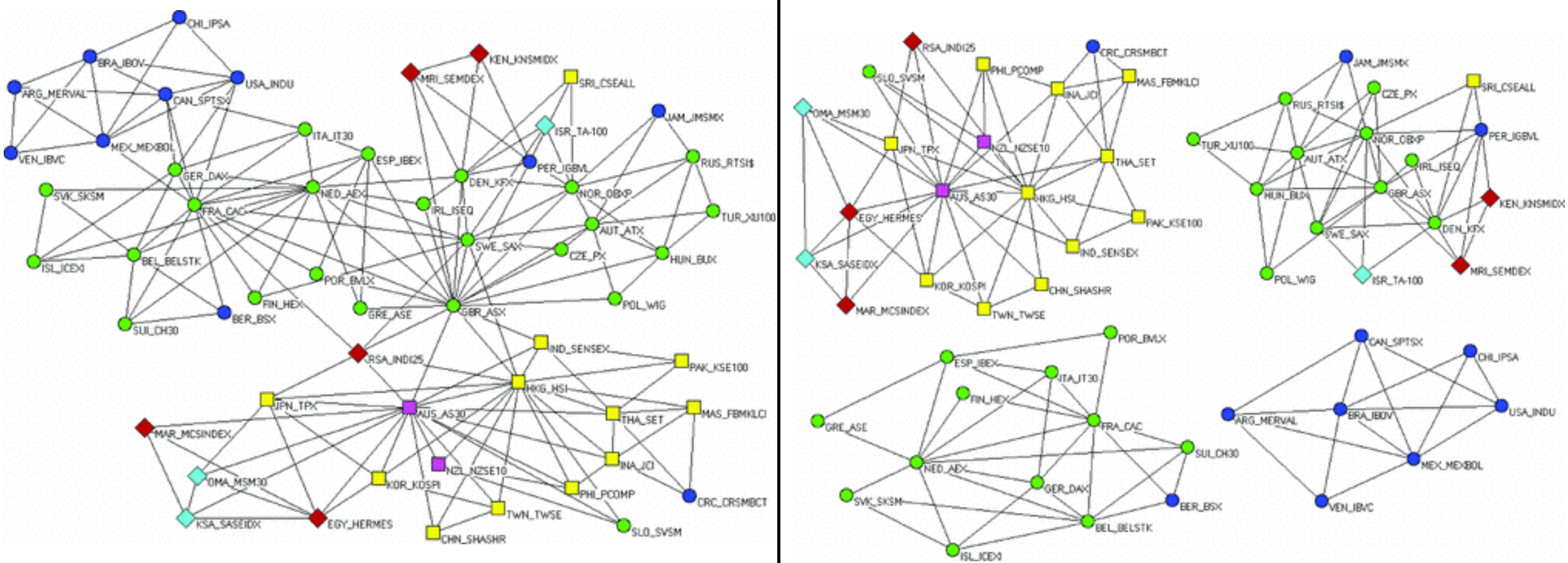

(a)
2007
2008
2009
2010
2011
2012
2013
0.6
0.3
0.0
−0.3
−0.6
(b)
(c)
Financial
Technology
Consumer Discretionary
Energy
Materials
Industrials
Communication
Healthcare
Consumer Staples
Group 1
Group 2
Group 3
Group 4
Group 5
Group 6
Group 7



ERICB
SSNLF
SNE
VOD
NOK
TVS
VZ
GARTNER
BBRY
BBY
AMZN
AAPL
GOOG
YHOO
GPS
JCP
MSFT
CEOS
EBAY
HPQ
INTC
COST
TGT
WMT
IBM
UTX
CSCO
HD
CAT
XOM
ERICB
SSNLF
SNE
NOK
TVS
VZ
GARTNER
BBRY
BBY
AMZN
AAPL
GOOG
YHOO
GPS
JCP
MSFT
EBAY
HPQ
INTC
OST
TGT
WMT
IBM
Financial
Technology
Consumer Discretionary
Energy
Materials
Industrials
Communication
Healthcare
Consumer Staples



Pecora N, Rovira Kaltwasser P, Spelta A (2017) Correction: Discovering SIFIs in Interbank Communities. PLOS ONE 12(4): e0176542.



Song et al. (2011) Evolution of worldwide stock markets, correlation structure, and correlation-based graphs Phys. Rev. E 84, 026108