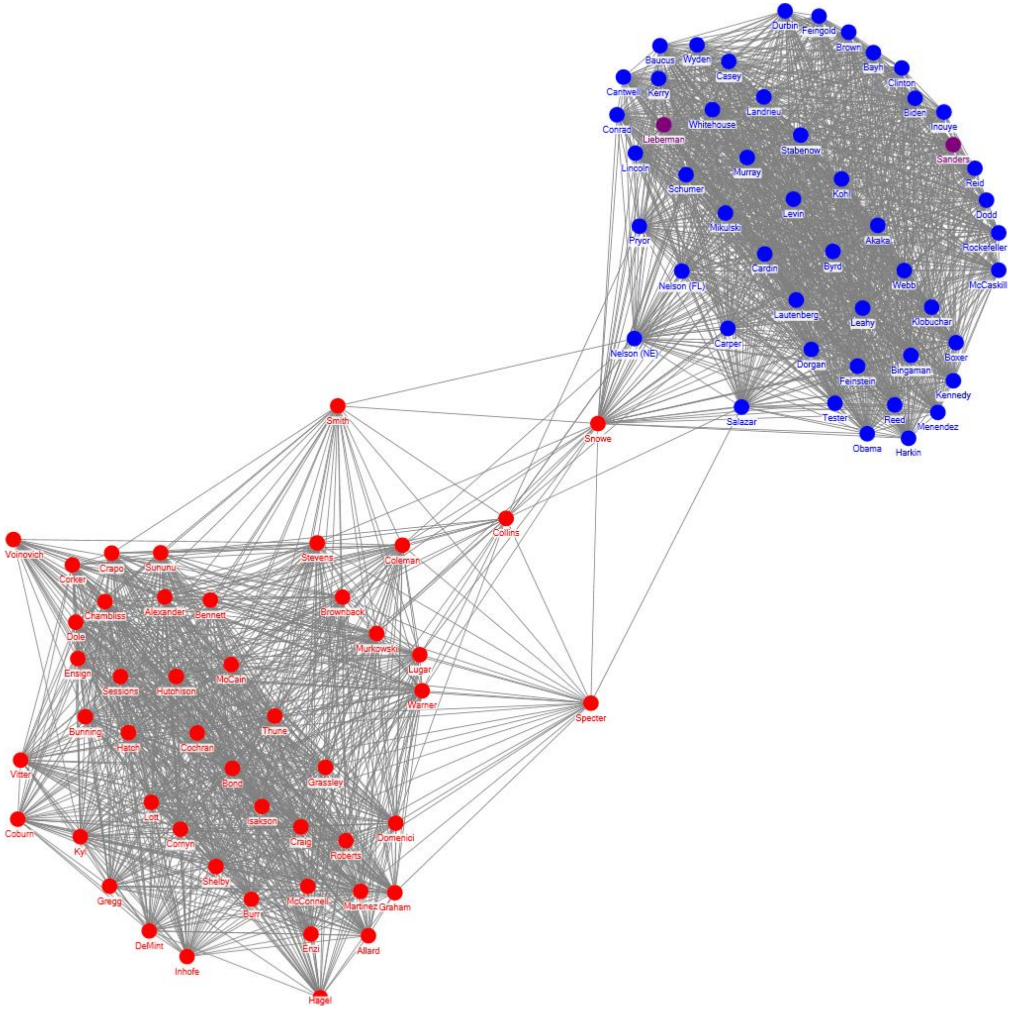
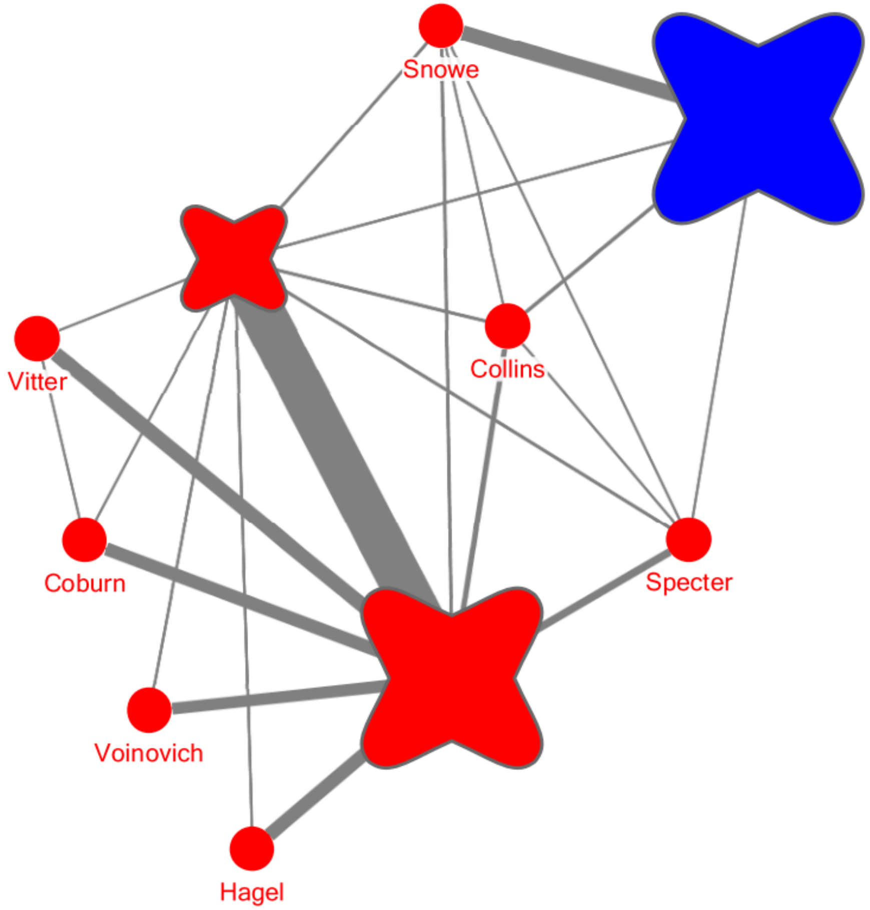
CS 7280-03 Special Topics on Visualization in Network Science

Lecture 3



Professor Cody Dunne

https://codydunne.github.io/cs7280-f16/
c.dunne@northeastern.edu



Reading Discussions

https://codydunne.github.io/cs7280-f16/schedule/
https://piazza.com/northeastern/fall2016/cs728003/

HW2 Tools & Teams

Projects

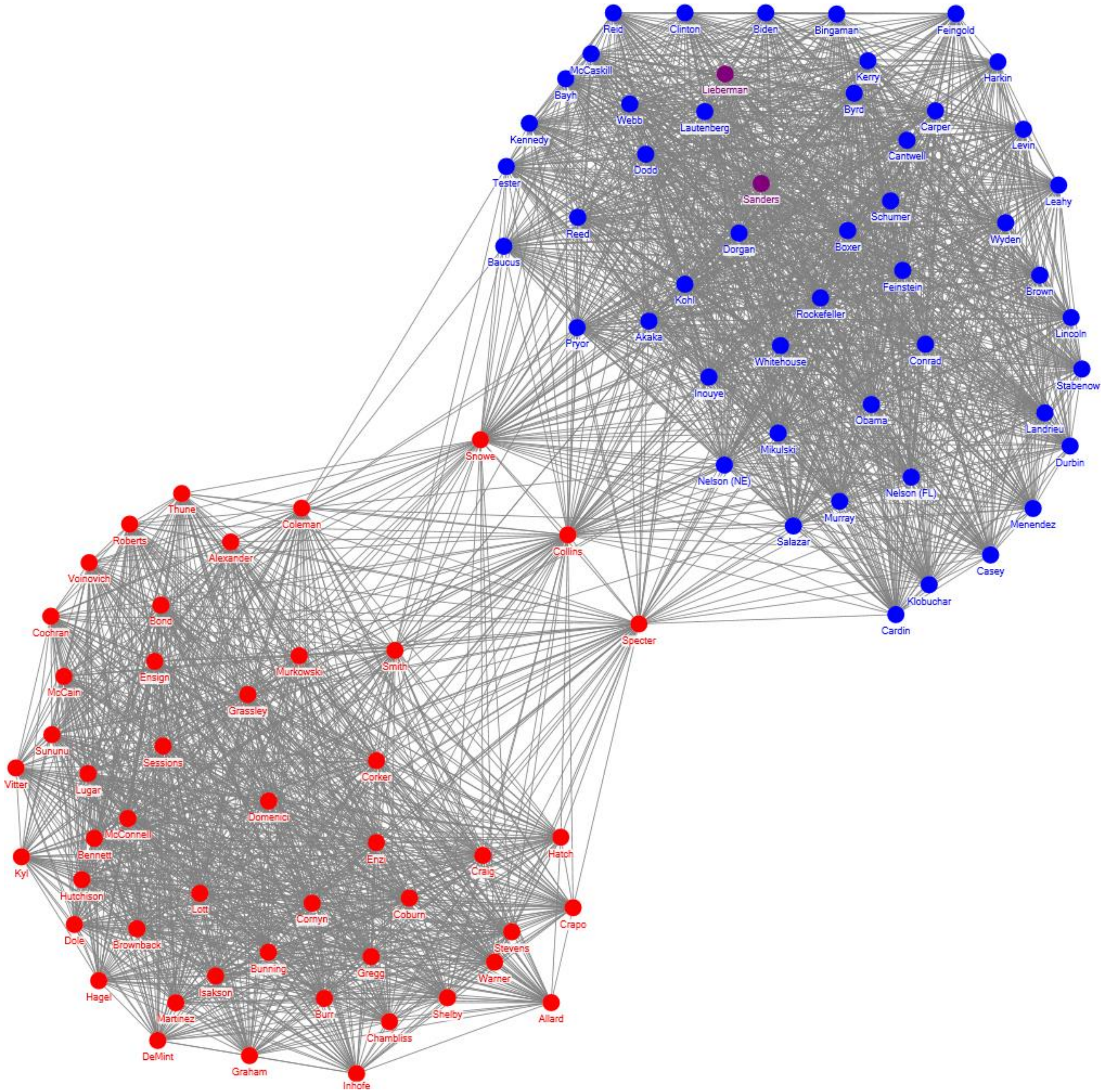
https://codydunne.github.io/cs7280-f16/project

Discussion:
Gestalt Principles for Node-Link Groups

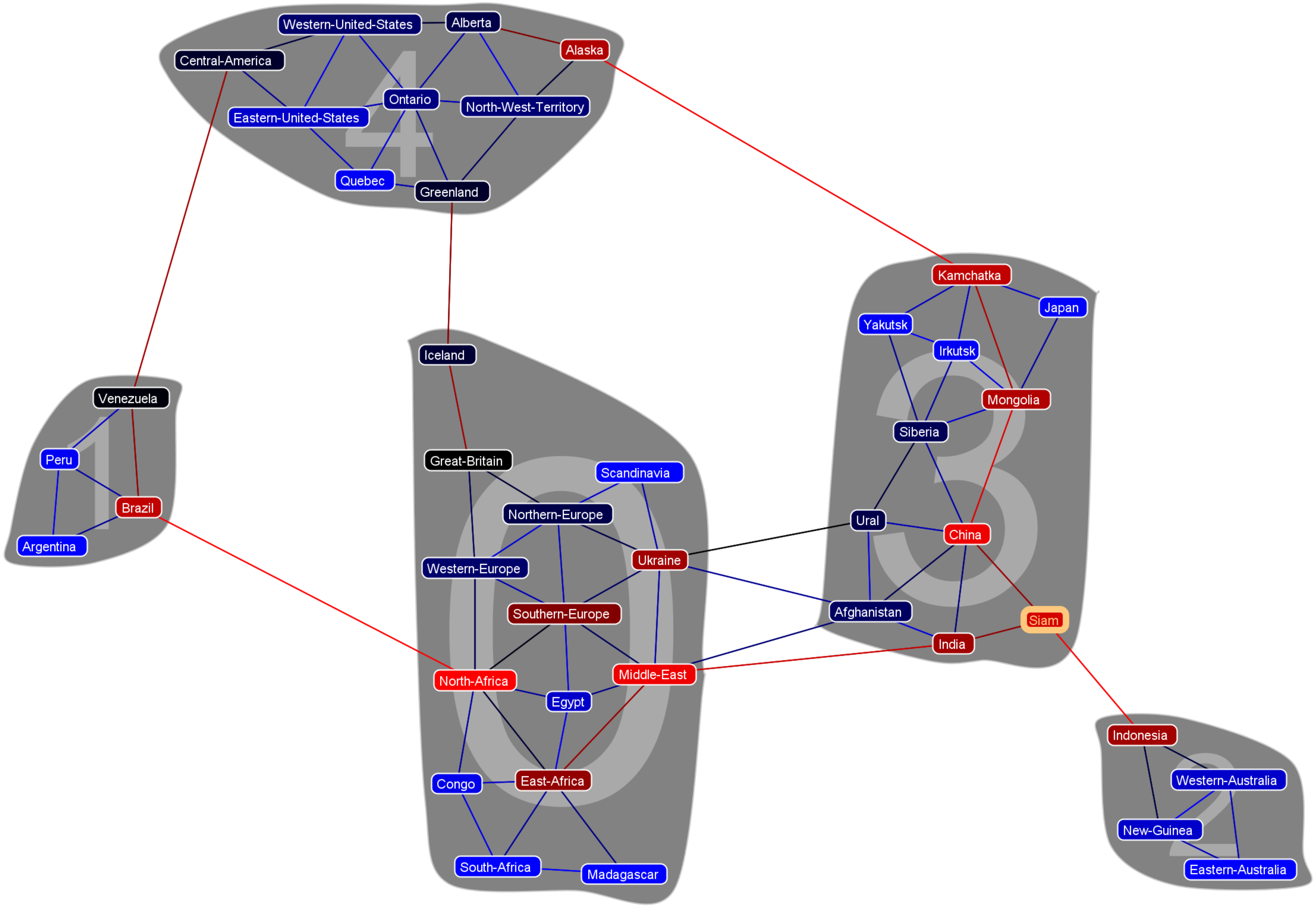
Gestalt Principles

similarity, proximity, connection, and enclosure

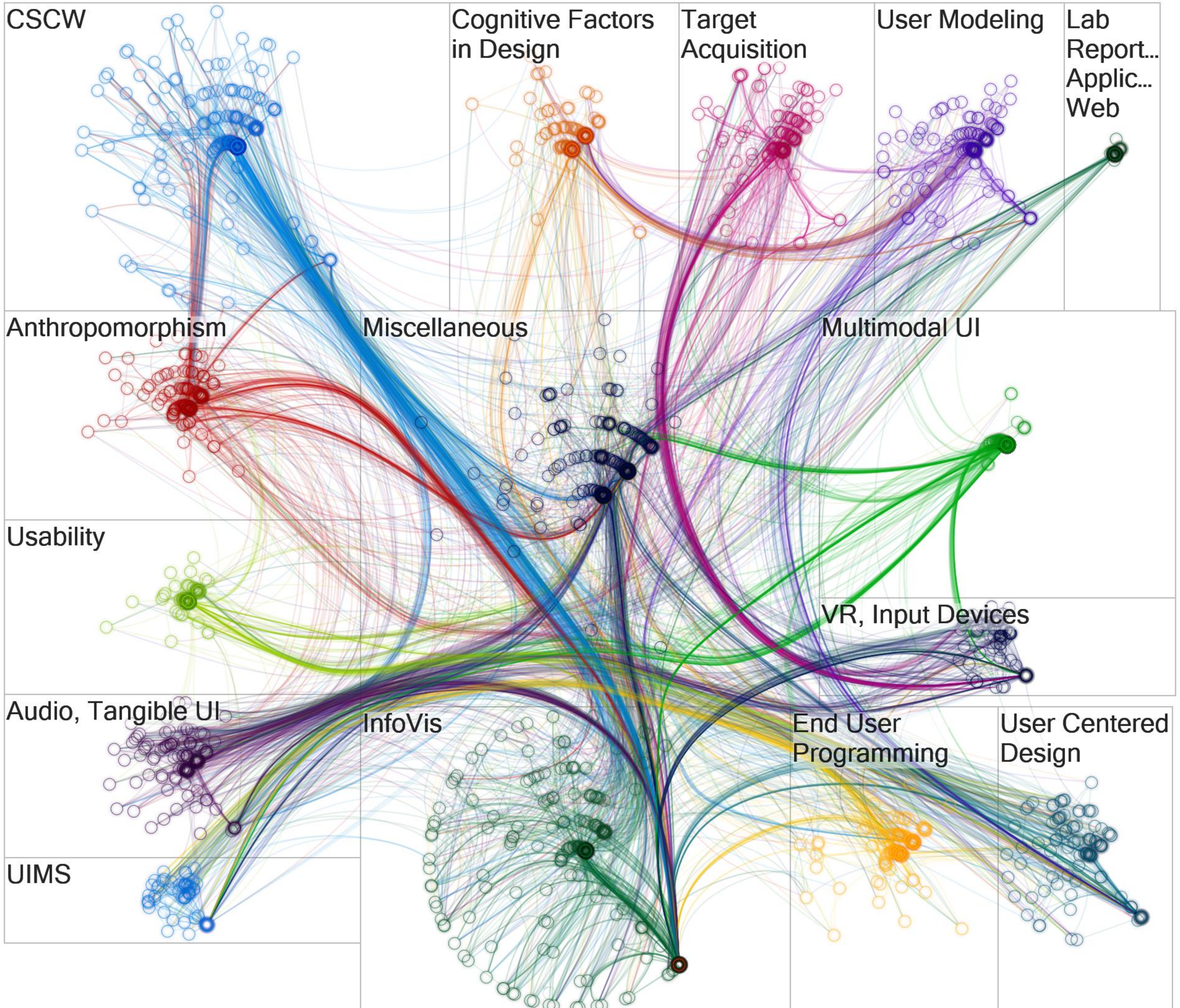
similarity, **proximity**, connection, and enclosure



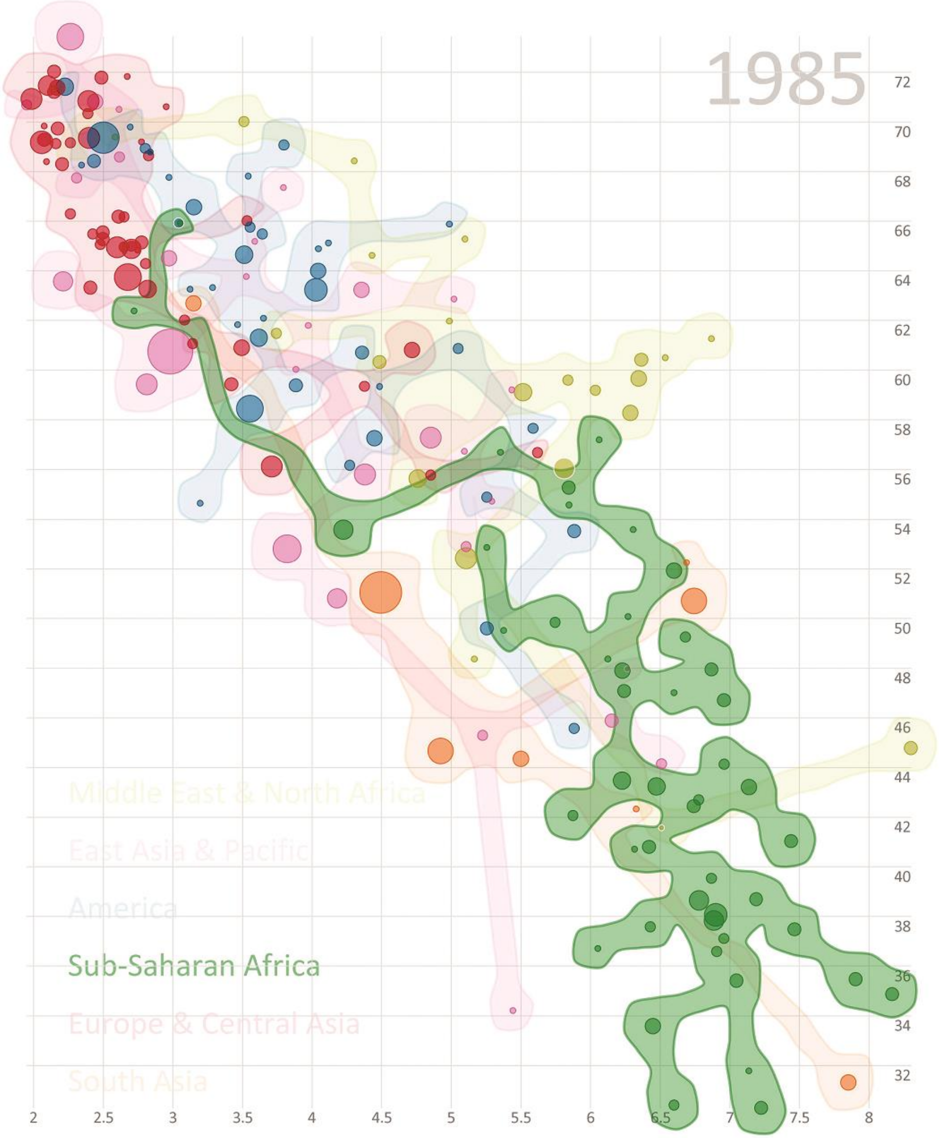
similarity, **proximity**, connection, and **enclosure**



similarity, **proximity**, connection, and **enclosure**

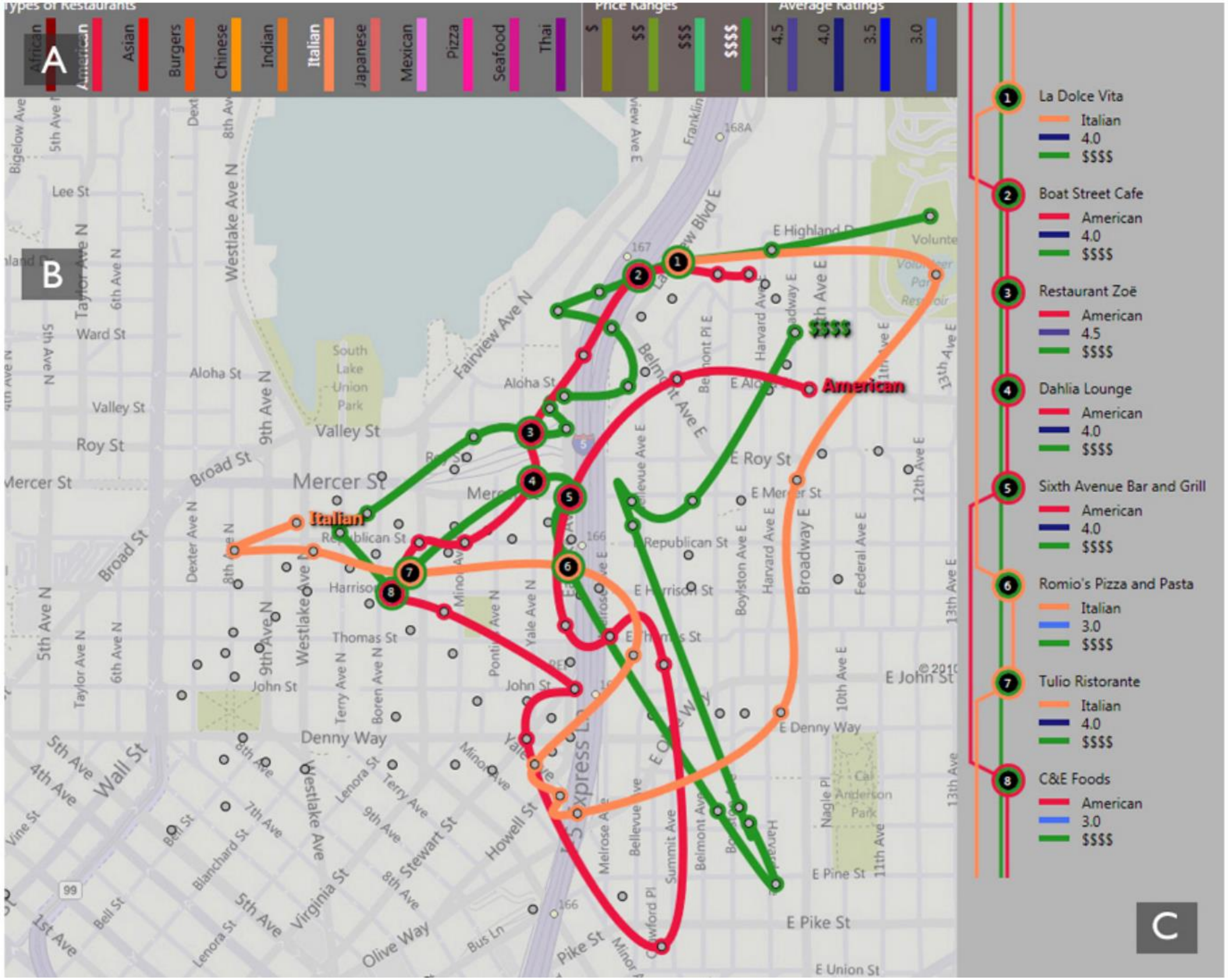


similarity, proximity, **connection**, and **enclosure**



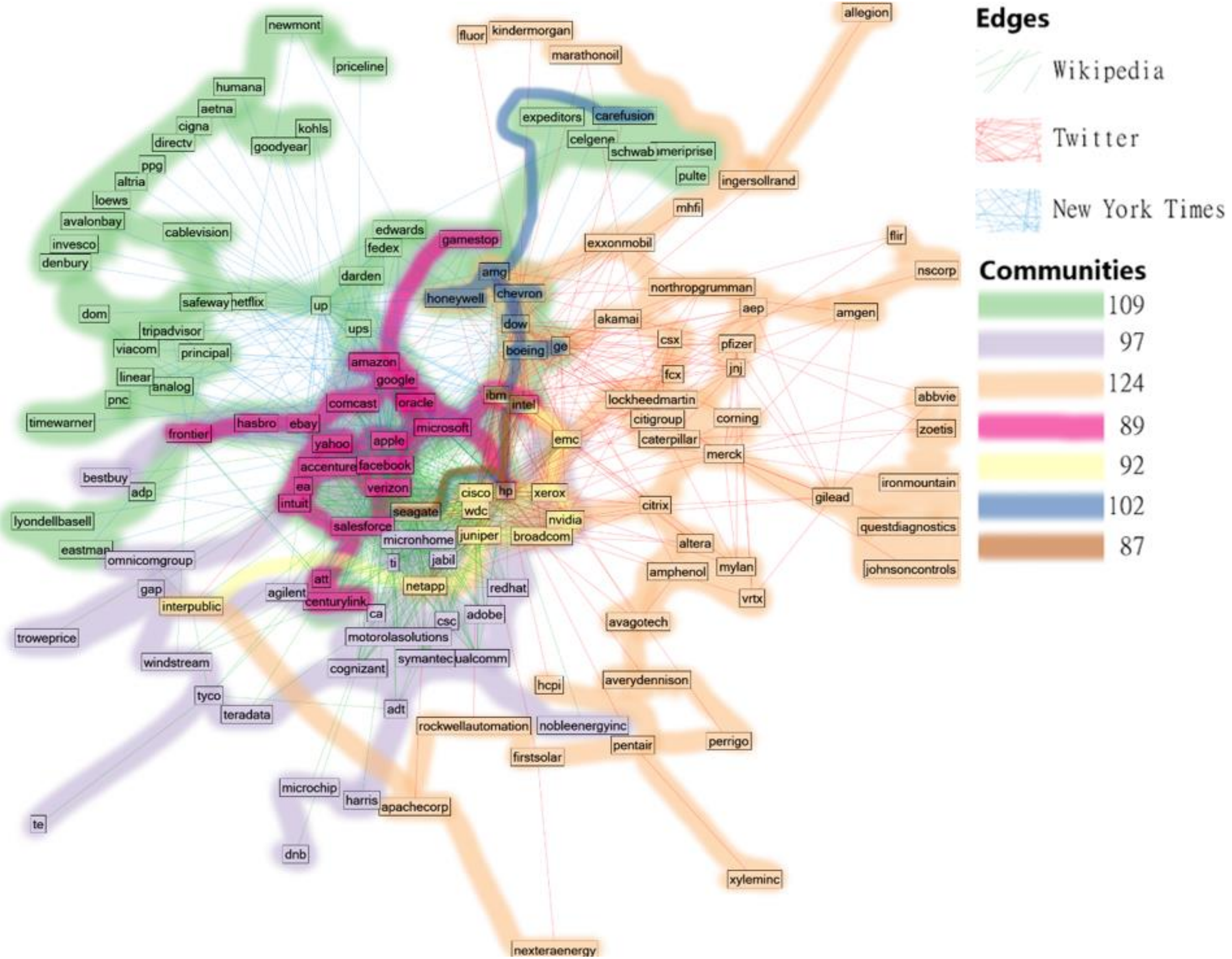
[Collins et al., 2009]

similarity, proximity, **connection**, and **enclosure**

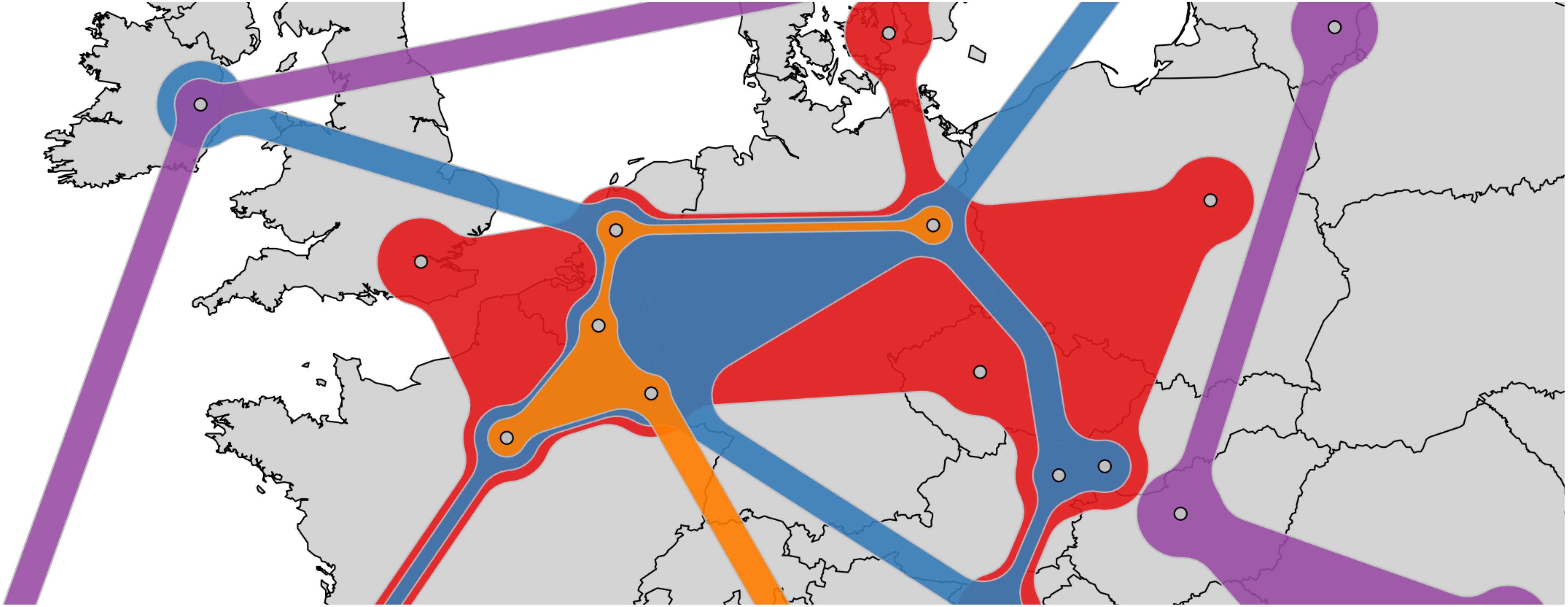


Alper et al., 2011

similarity, **proximity**, **connection**, and **enclosure**



similarity, **proximity**, **connection**, and **enclosure**



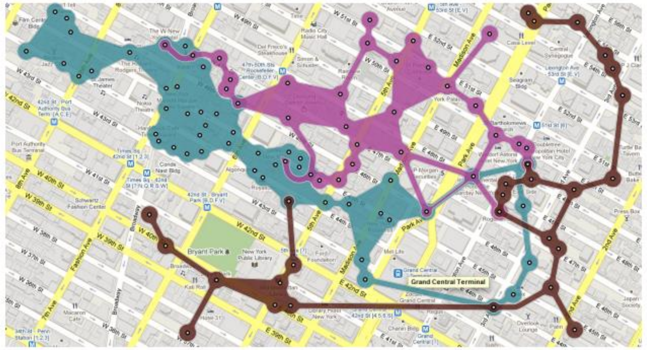
Muelemans et al., 2013

similarity, proximity, connection, and enclosure

(a) Bubble Sets

(b) Kelp Diagrams

(c) LineSets



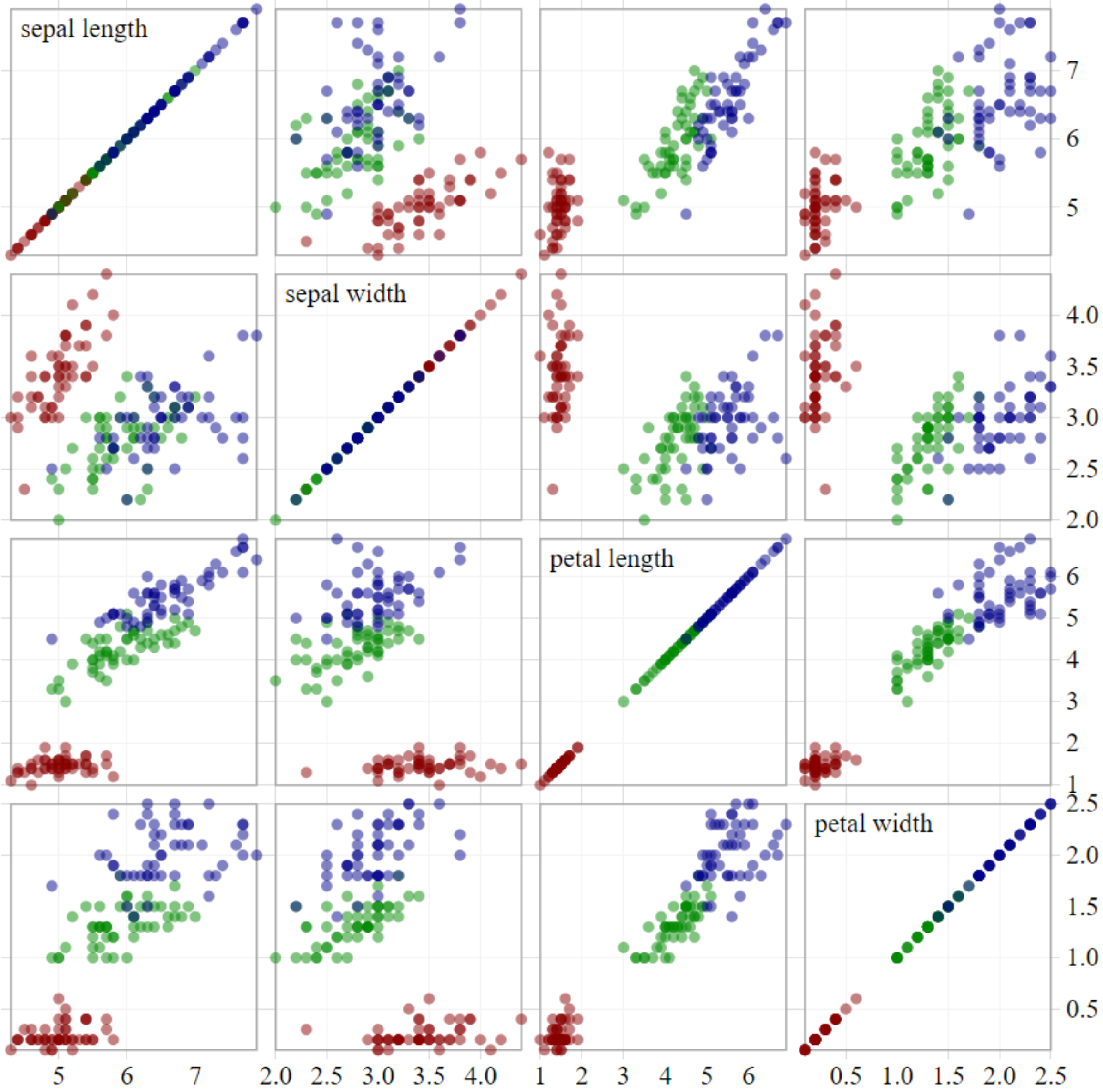
(d) KelpFusion (dense)

(e) KelpFusion (medium)

(f) KelpFusion (sparse)

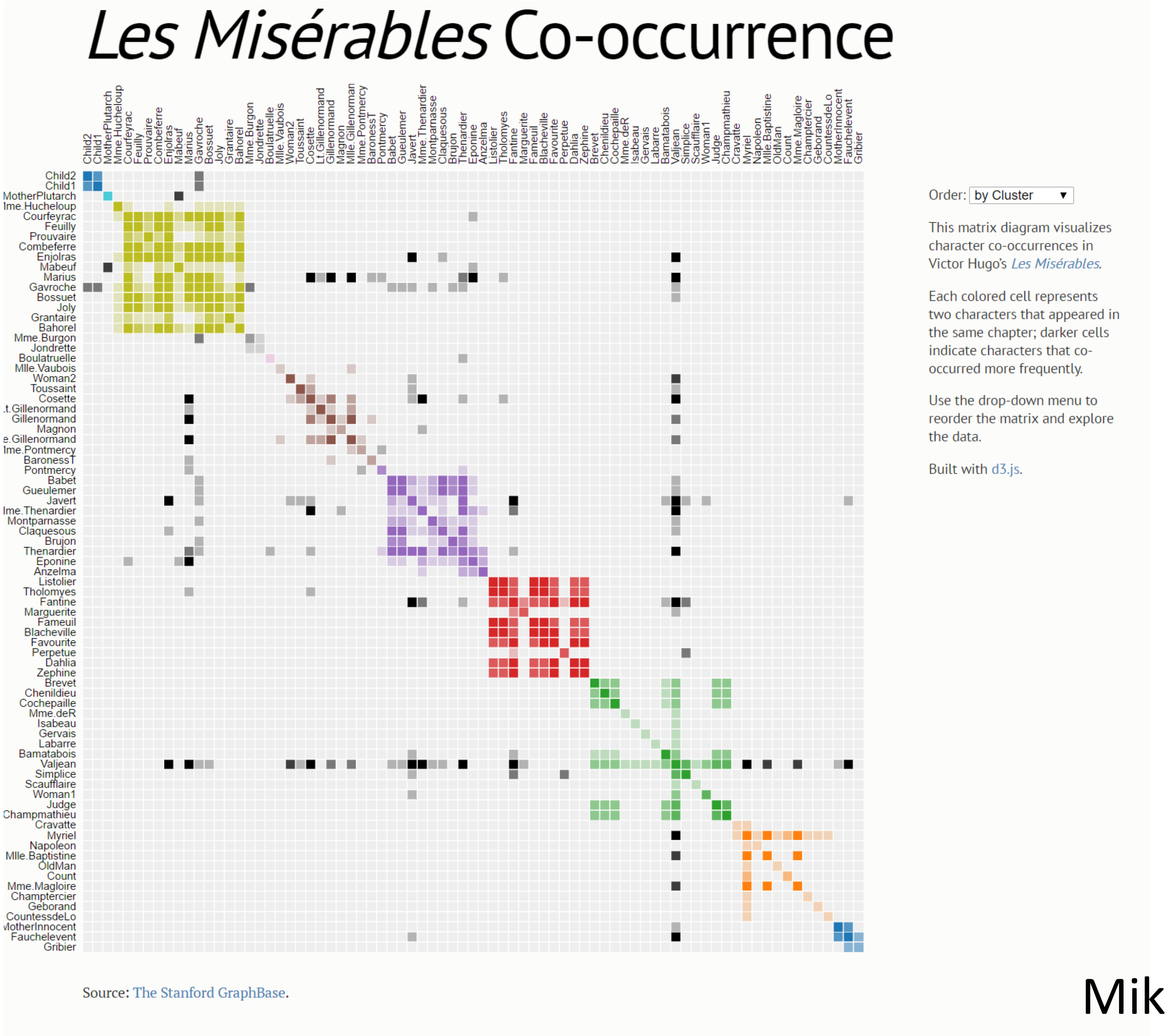
Muelemans et al., 2013

Discussion:
Matrix Seriation



● *Iris setosa*
● *Iris versicolor*
● *Iris virginica*

Edgar Anderson’s *Iris* data set
scatterplot matrix



Les Misérables Co-occurrence

Order: by Cluster ▾

This matrix diagram visualizes character co-occurrences in Victor Hugo's *Les Misérables*.

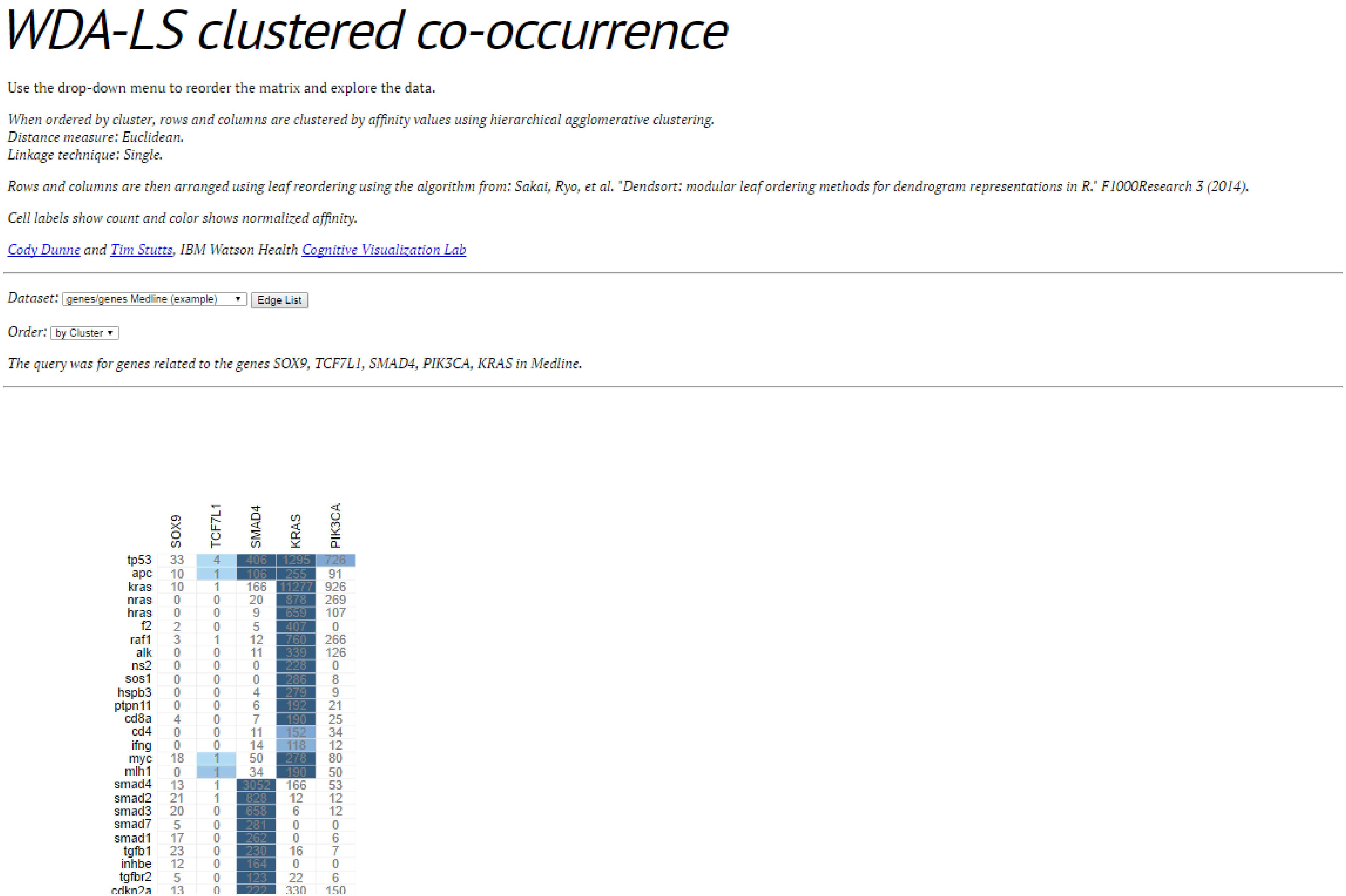
Each colored cell represents two characters that appeared in the same chapter; darker cells indicate characters that co-occurred more frequently.

Use the drop-down menu to reorder the matrix and explore the data.

Built with d3.js.

Source: The Stanford GraphBase.

Mike Bostock, 2012



WDA-LS clustered co-occurrence

Use the drop-down menu to reorder the matrix and explore the data.

When ordered by cluster, rows and columns are clustered by affinity values using hierarchical agglomerative clustering.
Distance measure: Euclidean.
Linkage technique: Single.

Rows and columns are then arranged using leaf reordering using the algorithm from: Sakai, Ryo, et al. "Dendsort: modular leaf ordering methods for dendrogram representations in R." F1000Research 3 (2014).

Cell labels show count and color shows normalized affinity.

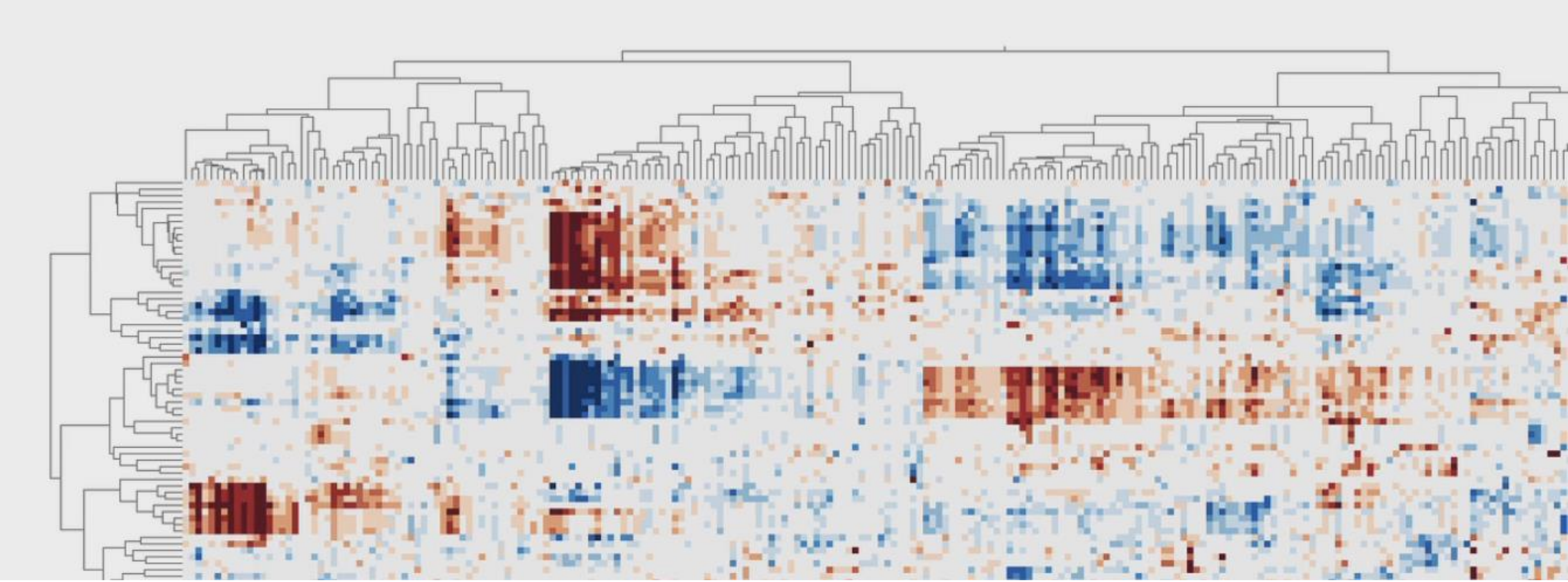
Cody Dunne and Tim Stutts, IBM Watson Health Cognitive Visualization Lab

Dataset: genes/genes Medline (example) ▾

Order: by Cluster ▾

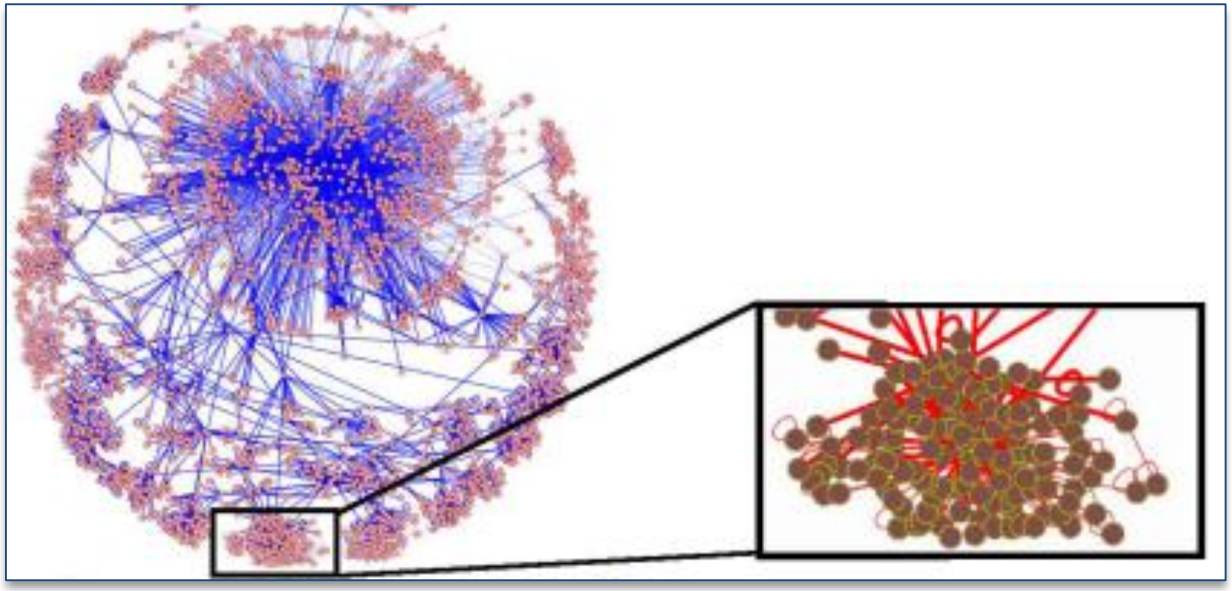
The query was for genes related to the genes SOX9, TCF7L1, SMAD4, PIK3CA, KRAS in Medline.

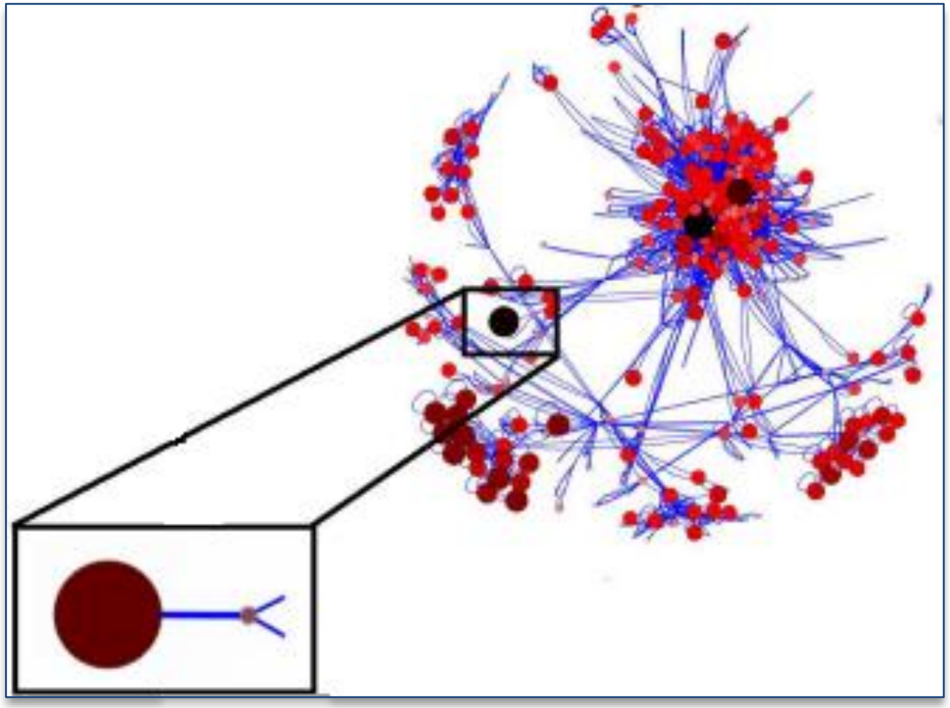
	SOX9	TCF7L1	SMAD4	KRAS	PIK3CA
tp53	33	4	406	1295	726
apc	10	1	106	255	91
kras	10	1	166	11277	926
nras	0	0	20	878	269
hras	0	0	9	659	107
f2	2	0	5	407	0
raf1	3	1	12	760	266
alk	0	0	11	339	126
ns2	0	0	0	228	0
sos1	0	0	0	286	8
hspb3	0	0	4	279	9
ptpn11	0	0	6	192	21
cd8a	4	0	7	190	25
cd4	0	0	11	152	34
ifng	0	0	14	118	12
myc	18	1	50	278	80
mlh1	0	1	34	190	50
smad4	13	1	3052	166	53
smad2	21	1	828	12	12
smad3	20	0	658	6	12
smad7	5	0	281	0	0
smad1	17	0	262	0	6
tgfb1	23	0	230	16	7
inhbe	12	0	164	0	0
tgfbr2	5	0	123	22	6
cdkn2a	13	0	222	330	150



Sakai et al., 2014

Topology Aggregation



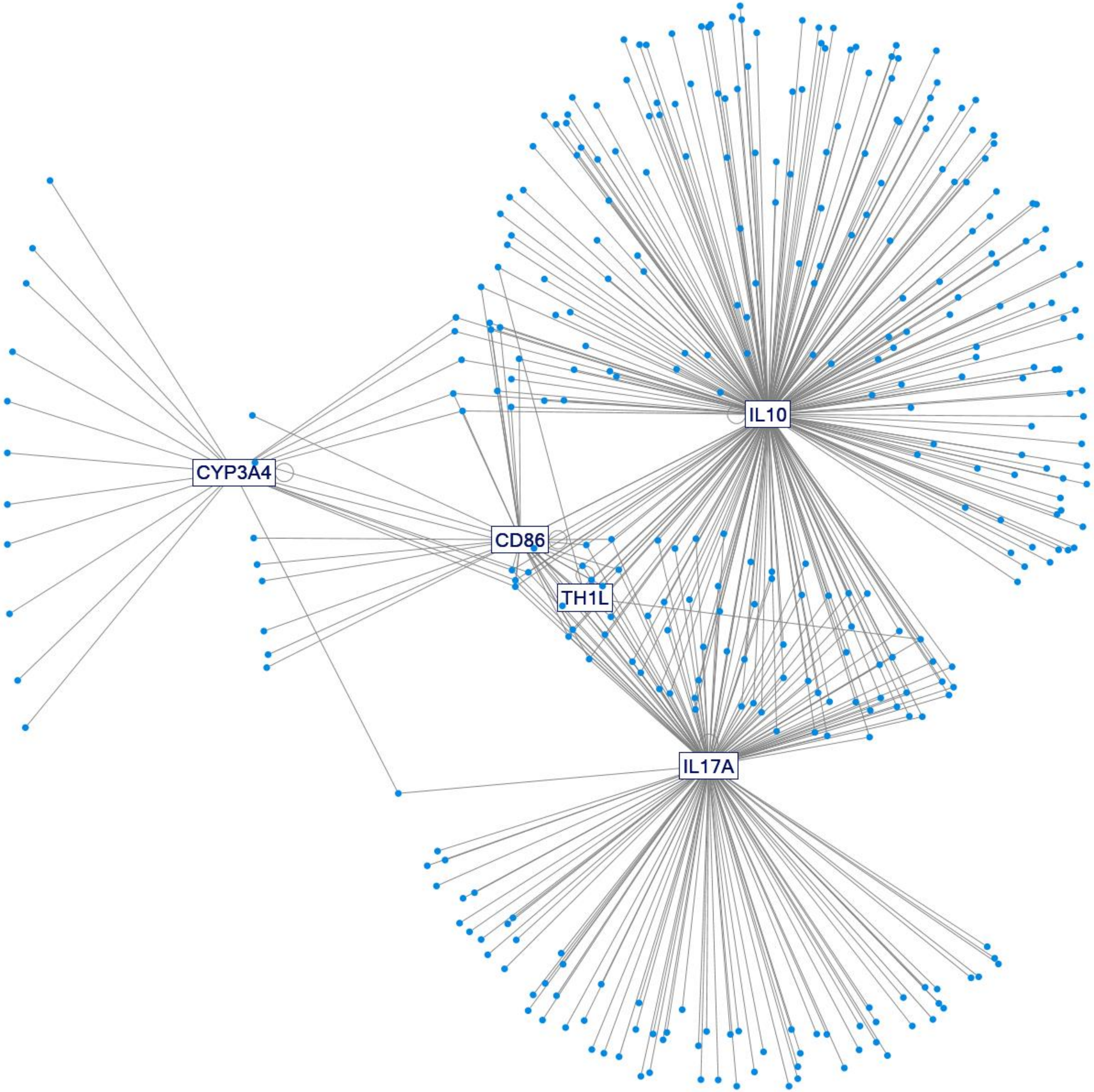


Navlakha et al., 2008

5 genes related to a target drug:
TH1L, IL10, CD86, IL17A, CYP3A4

Query genes

Conditions



Observations

1: There are repeating patterns in networks (motifs)

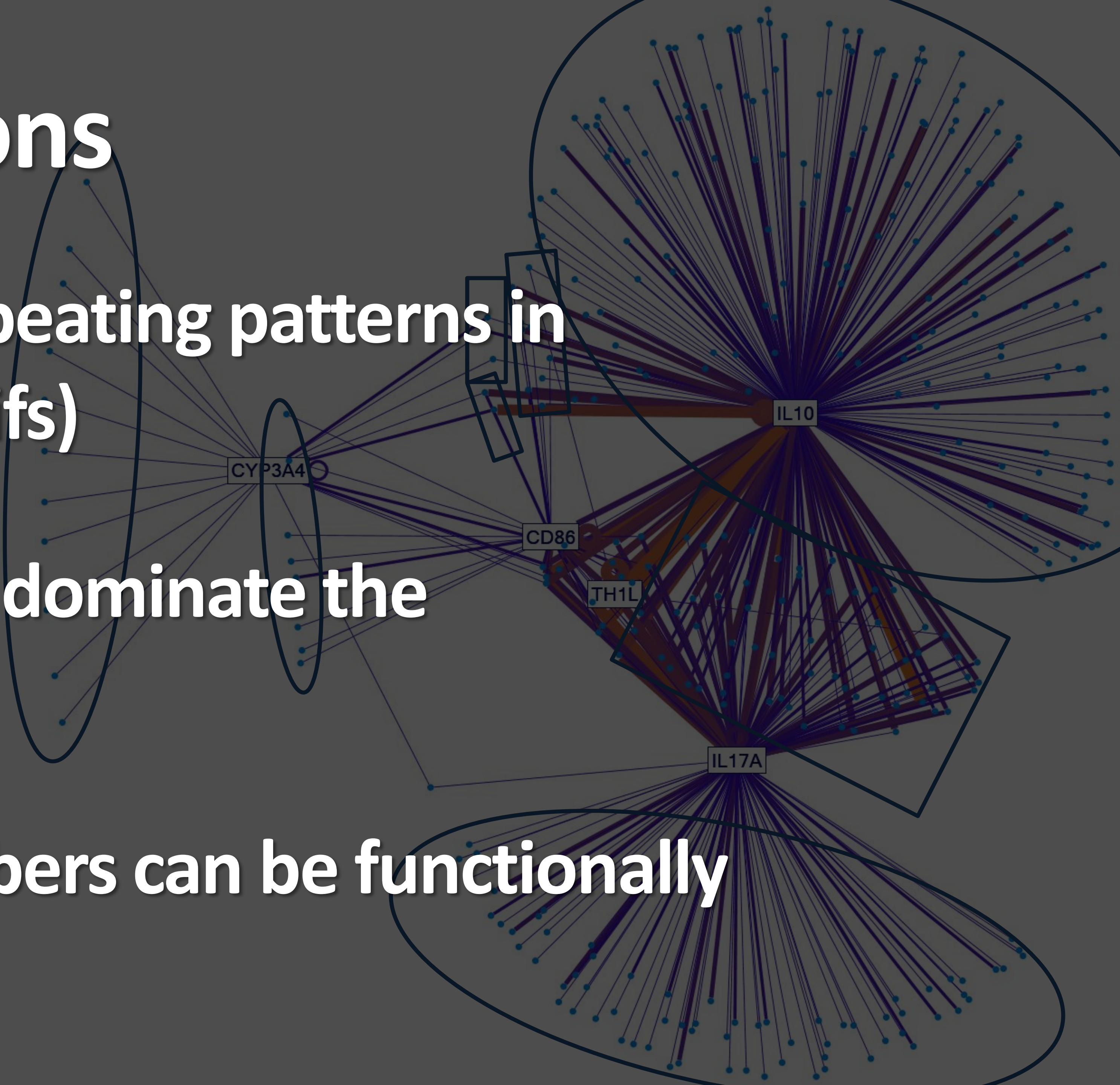
2: Motifs often dominate the visualization

3: Motifs members can be functionally equivalent

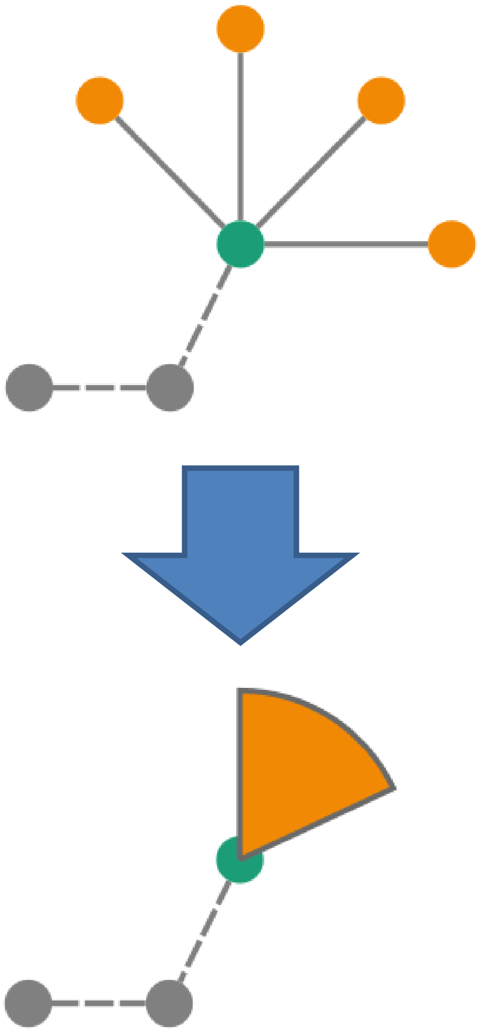
5 genes related to a target drug:
TH1L, IL10, CD86, IL17A, CYP3A4

Query genes

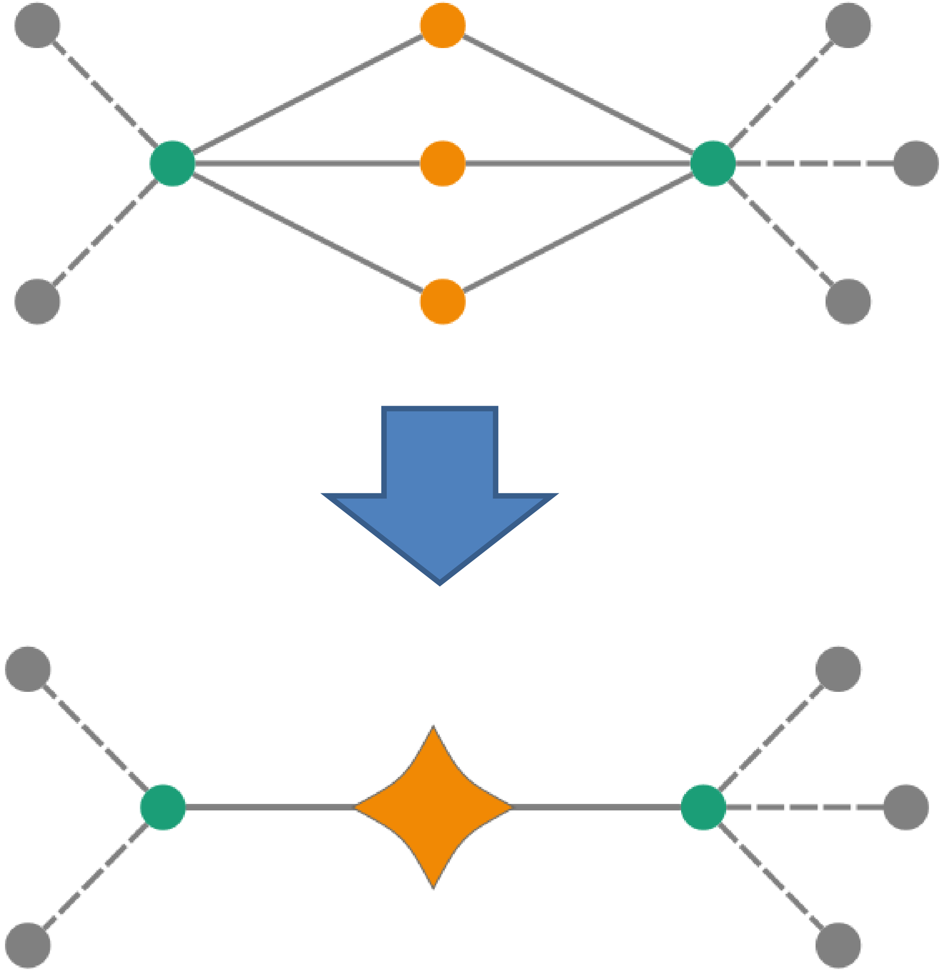
Conditions



Fan Motif

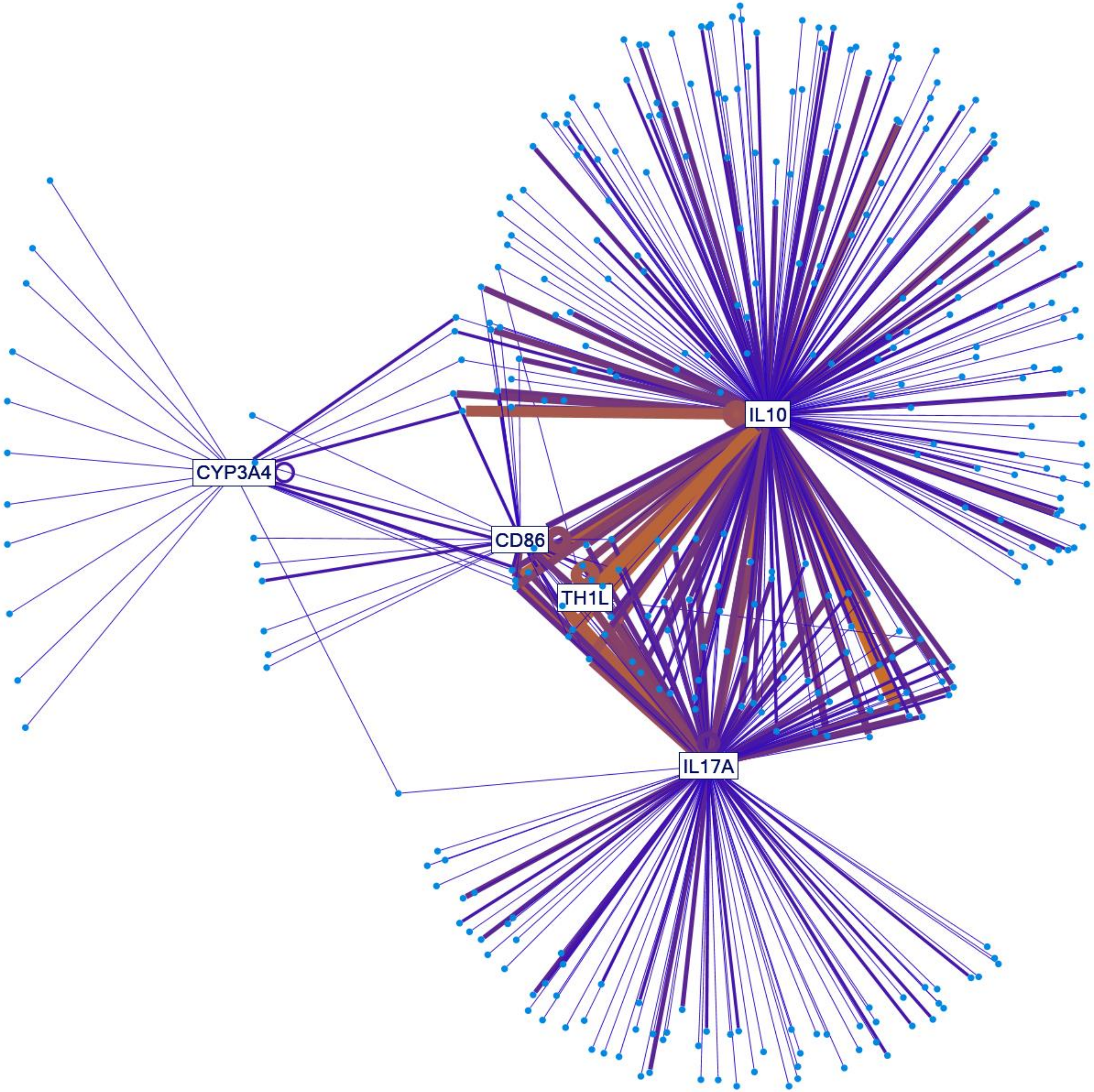


2-Connector Motif



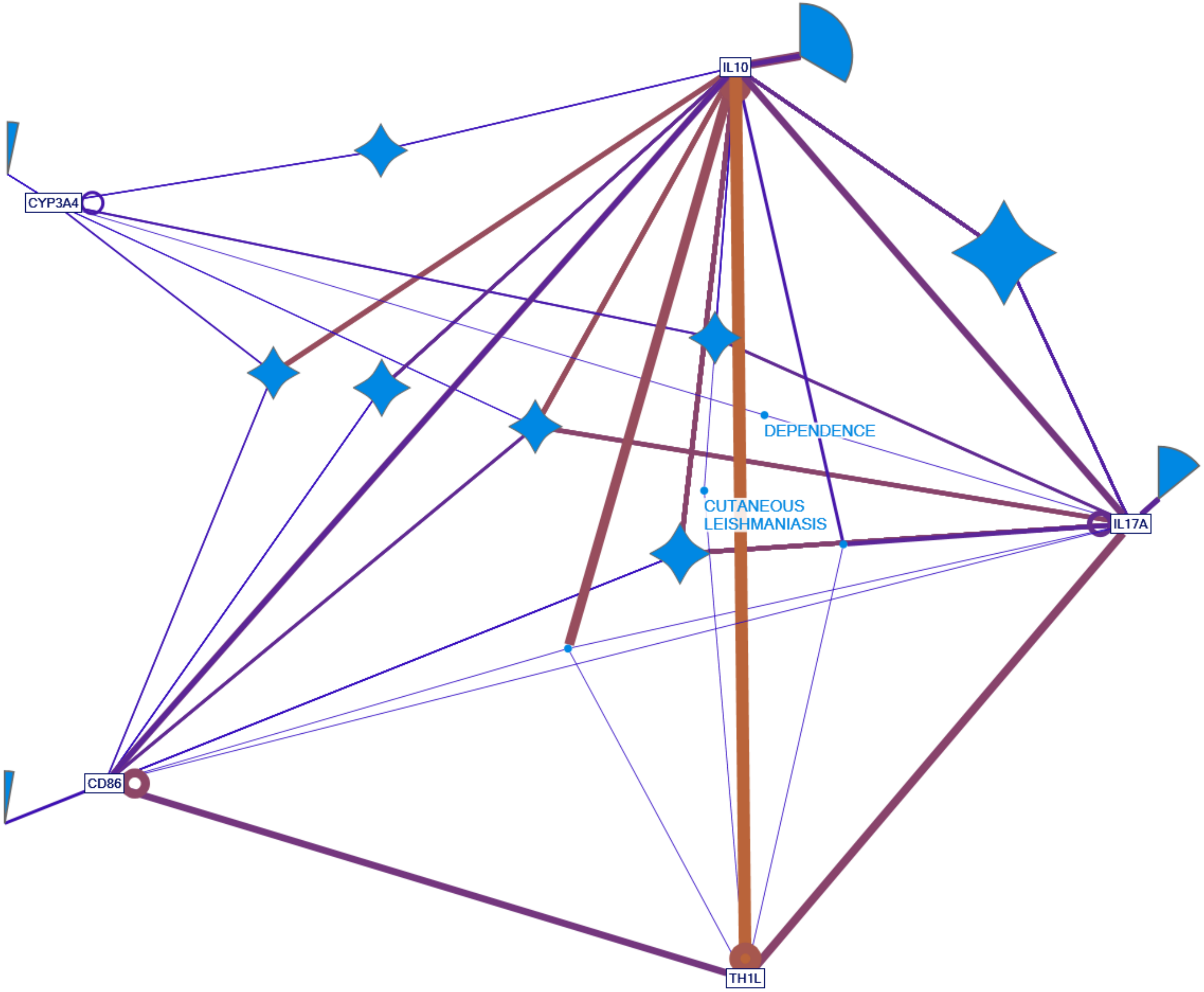
Query genes

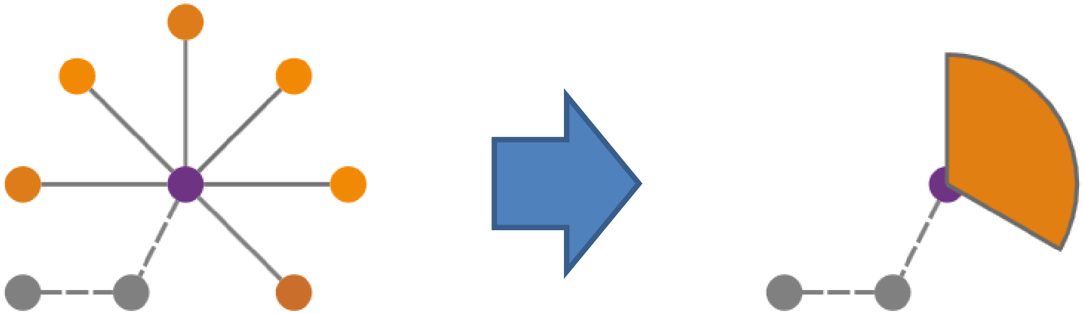
Conditions

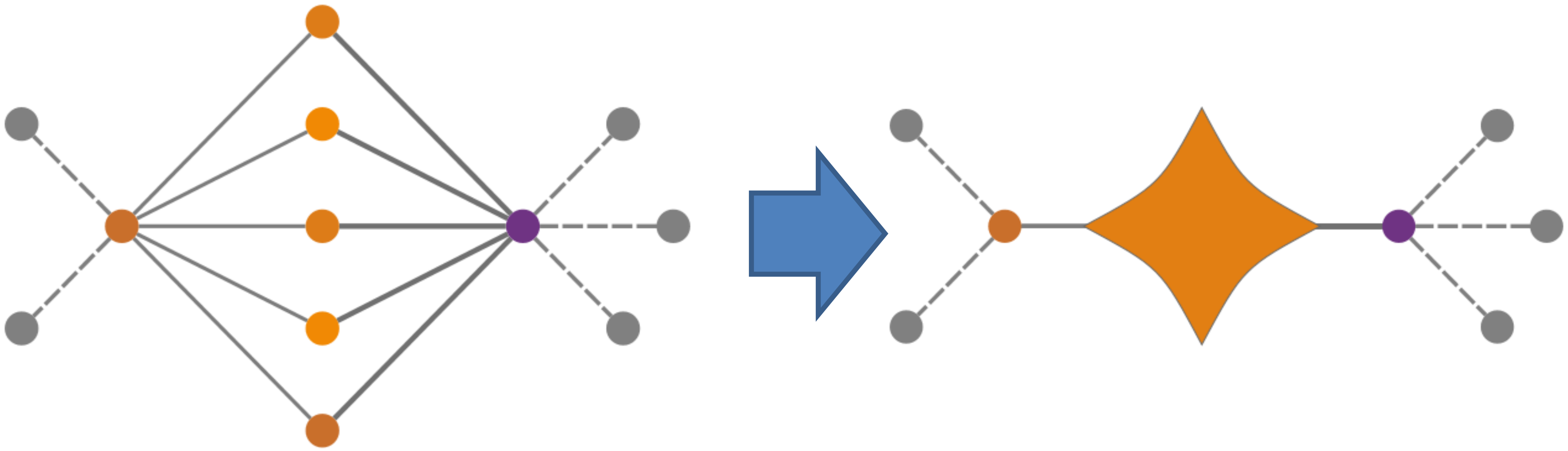


Query genes

Conditions,
some in motifs

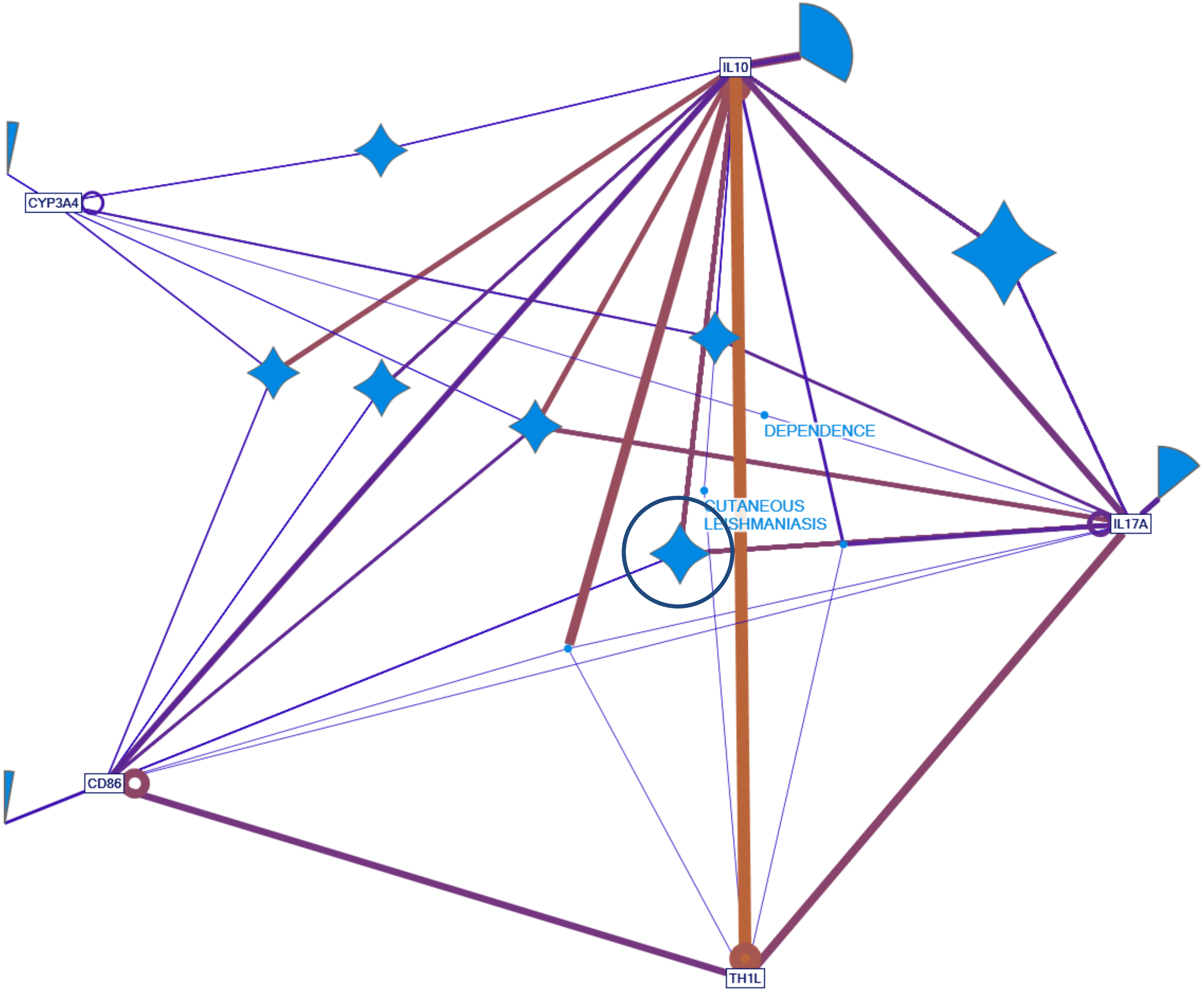






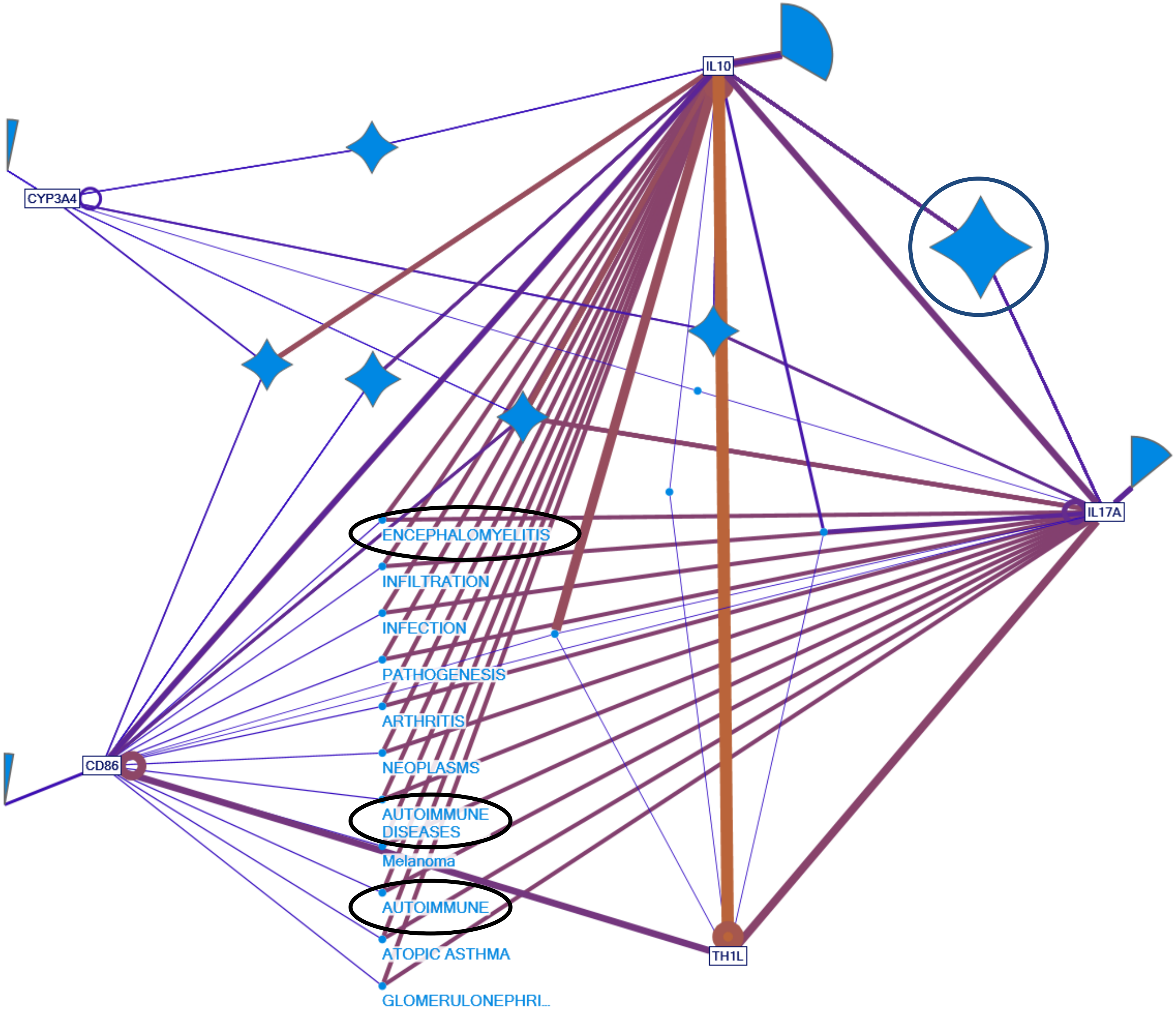
Query genes

Conditions,
some in motifs



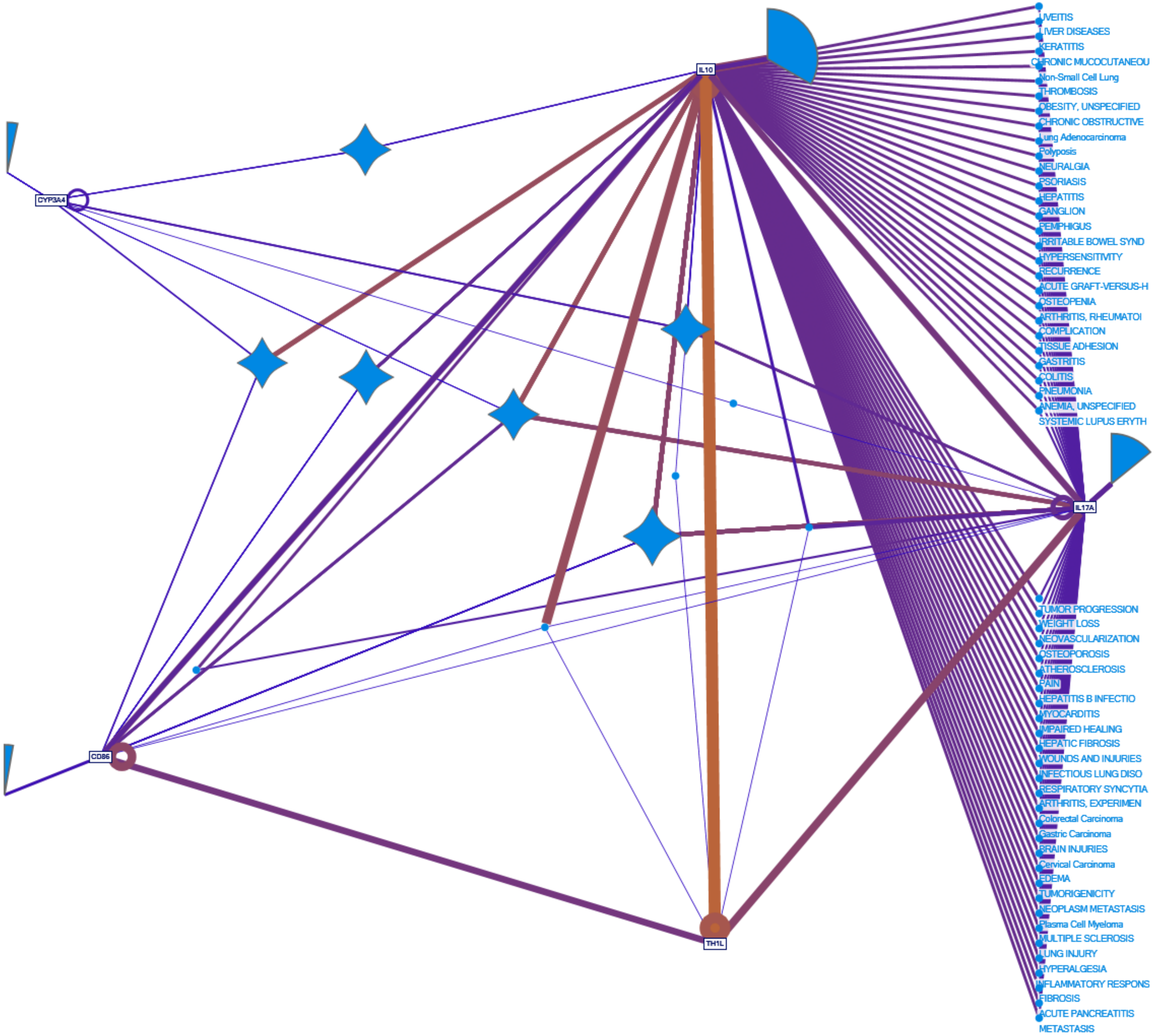
Query genes

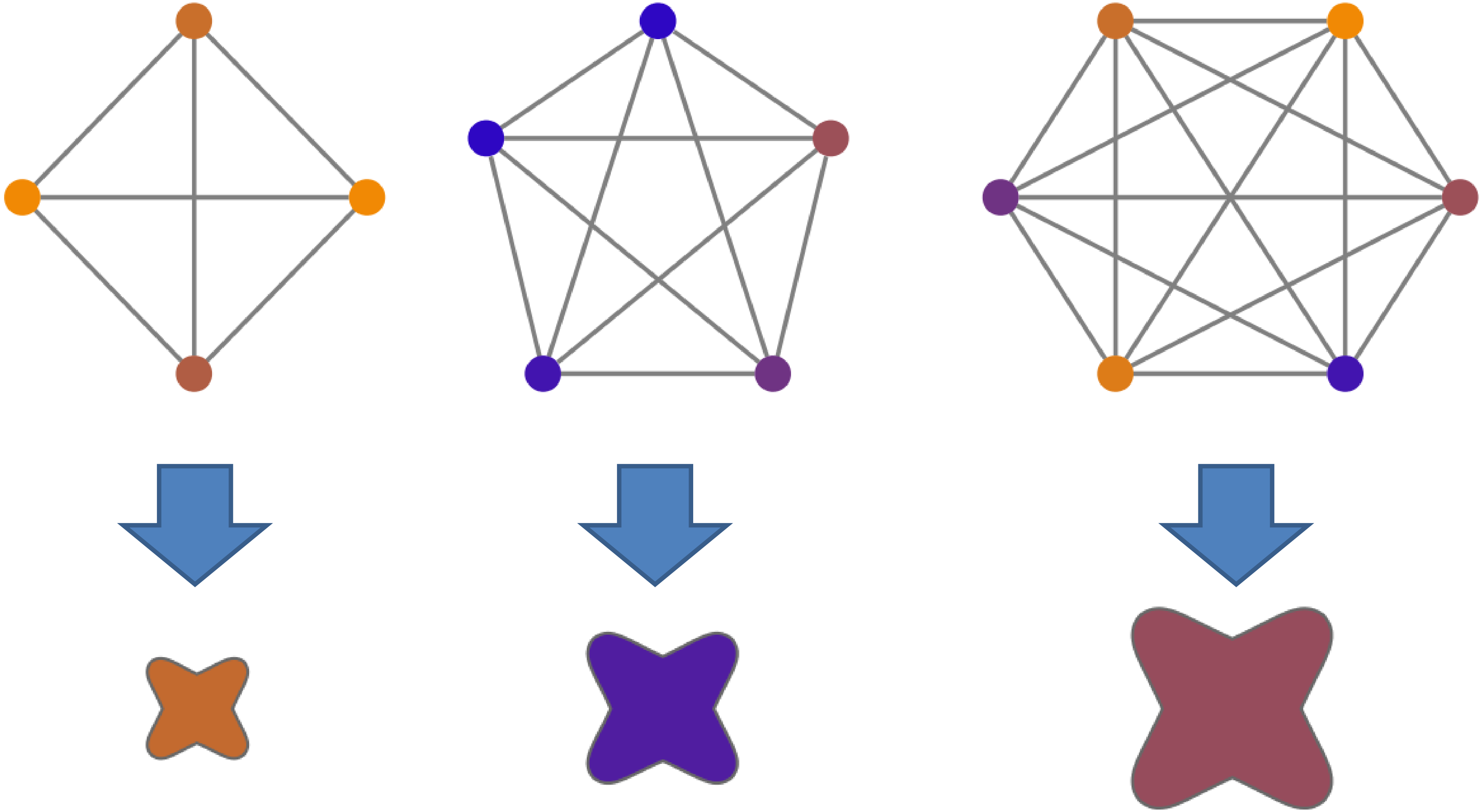
Conditions,
some in motifs



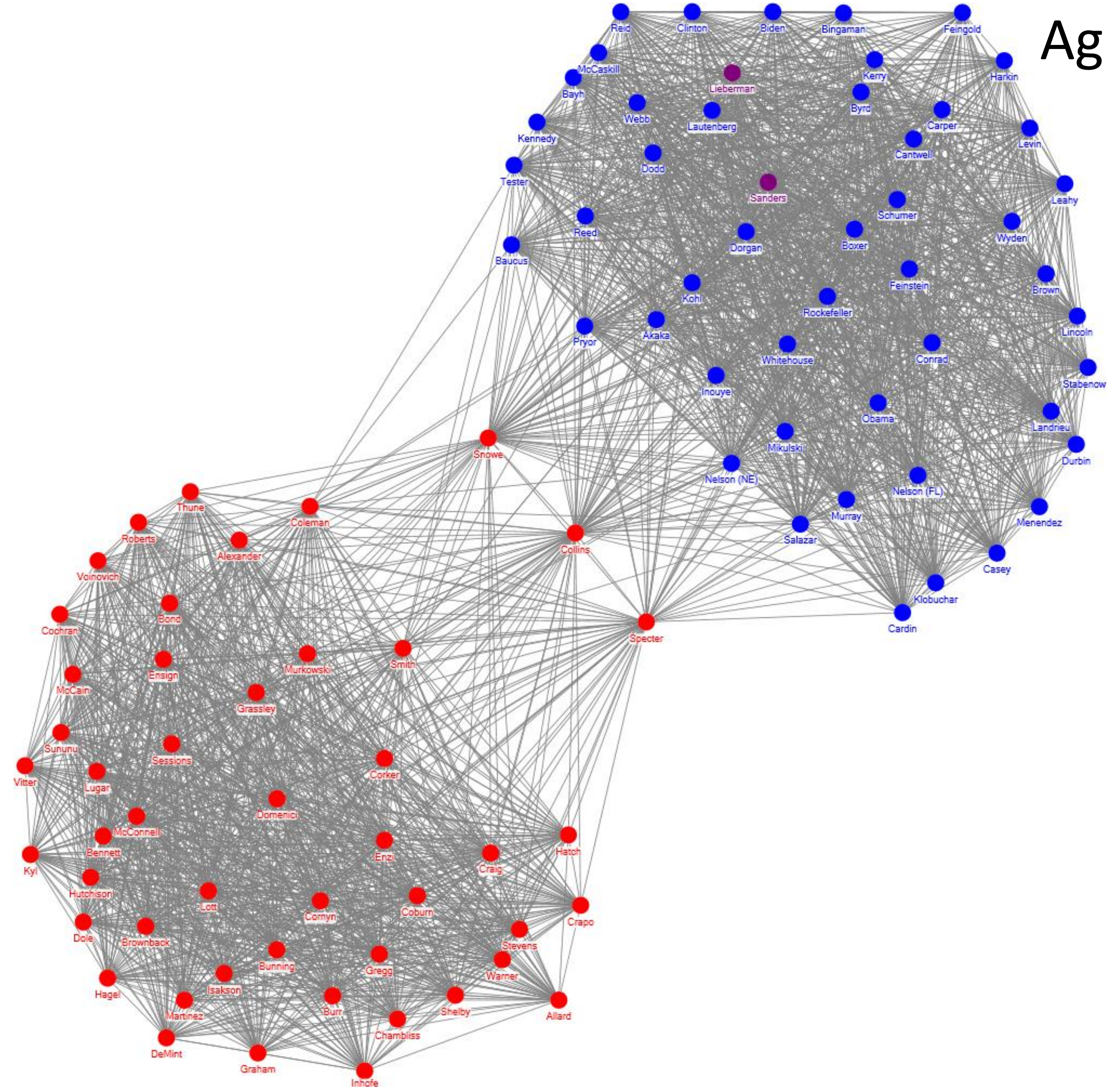
Query genes

Conditions,
some in motifs

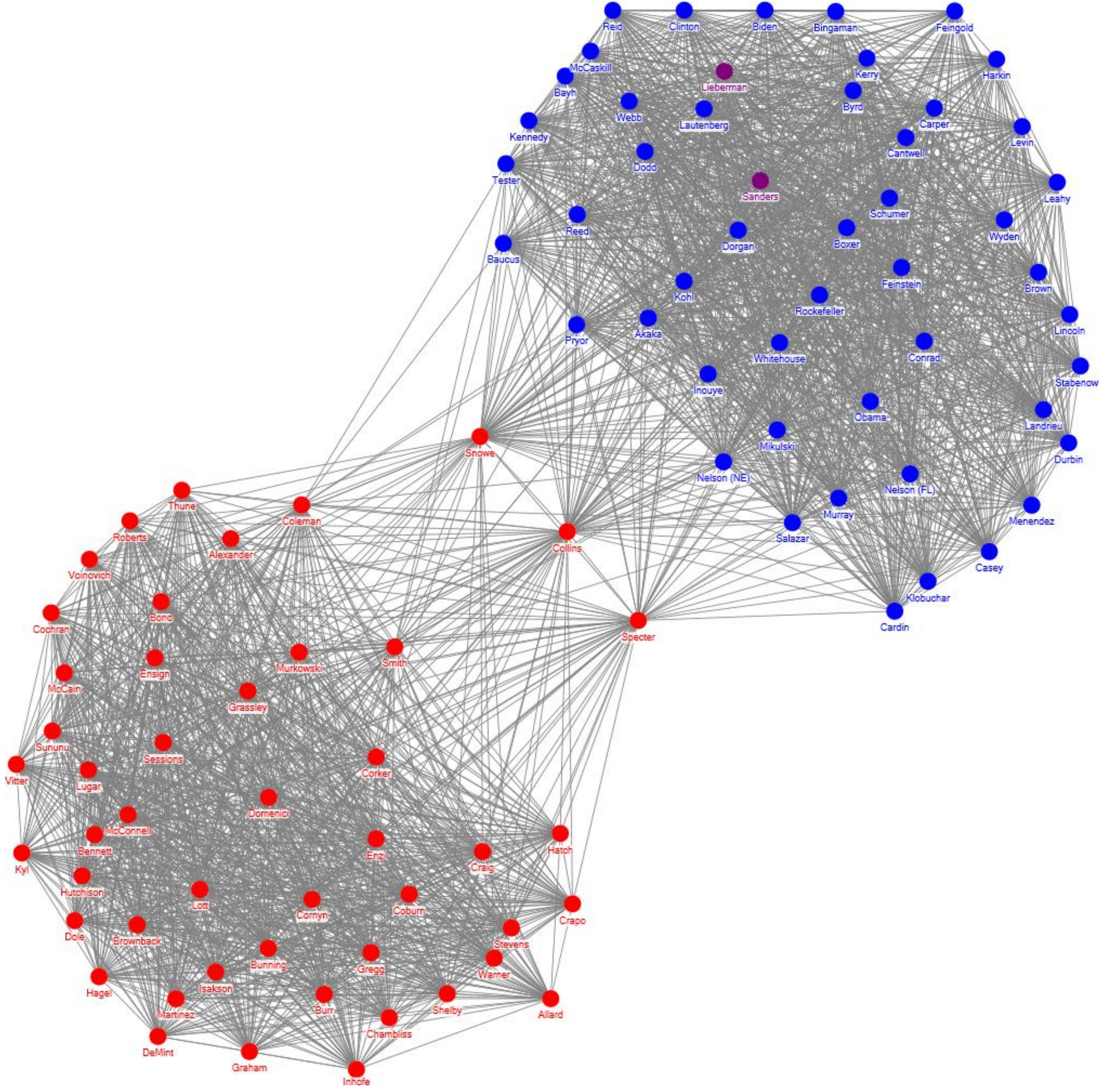


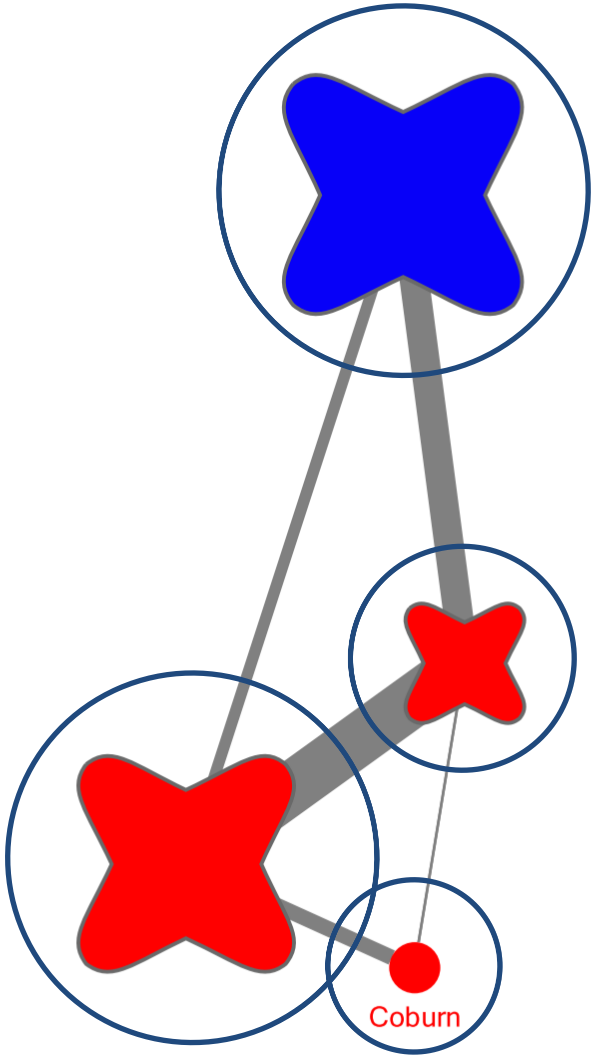


Senate Co-Voting: 65% Agreement

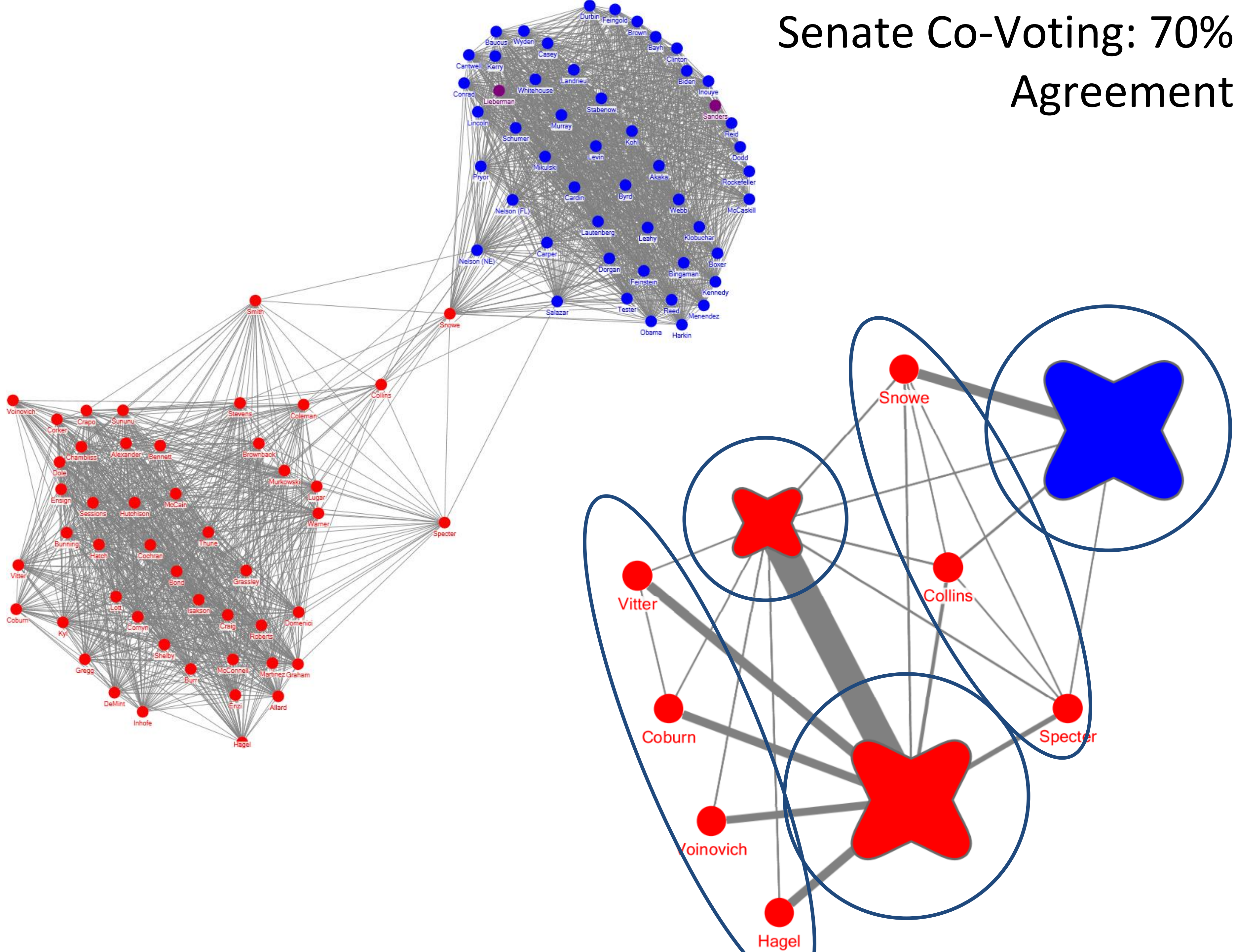


Senate Co-Voting: 65% Agreement

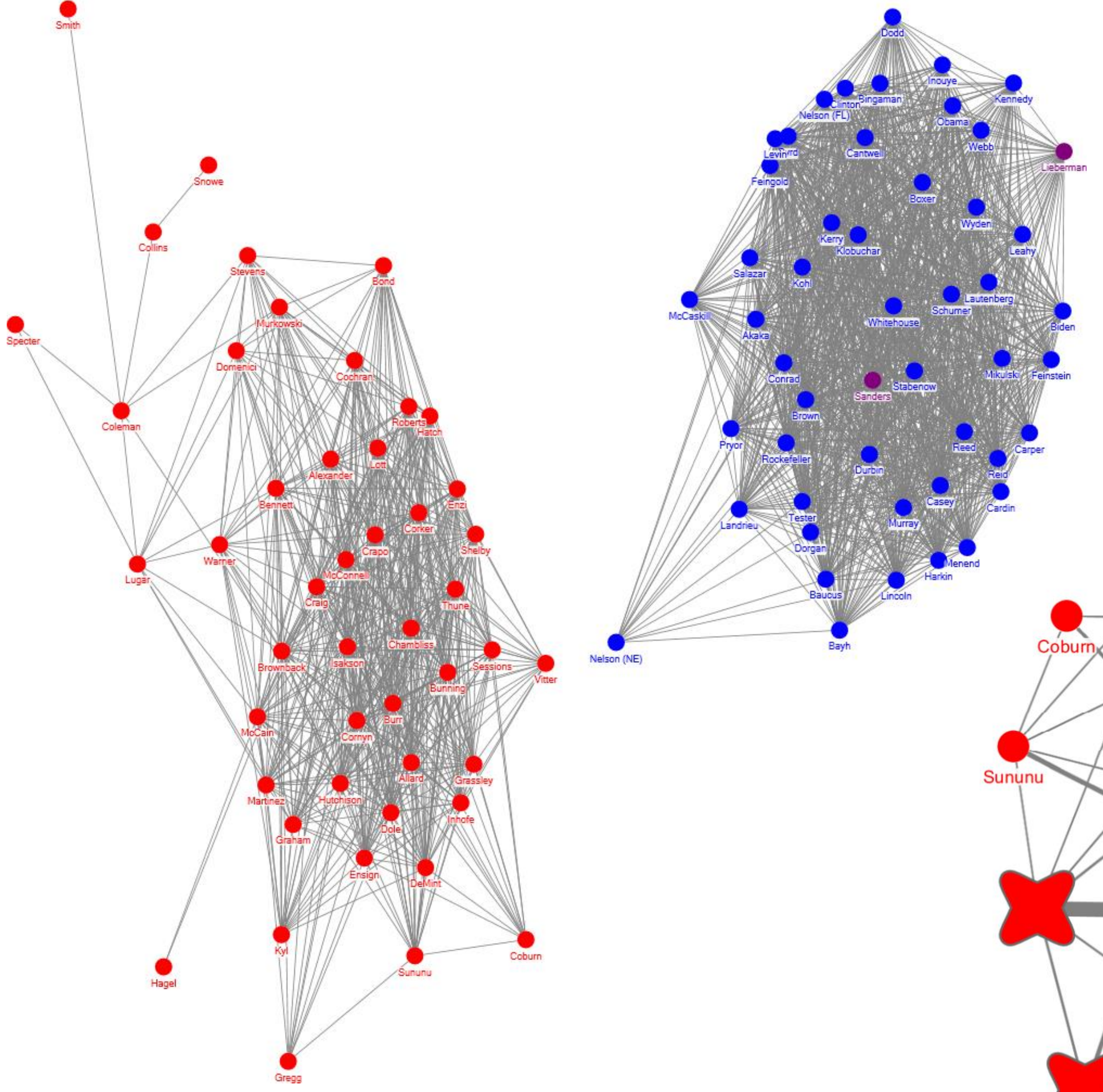


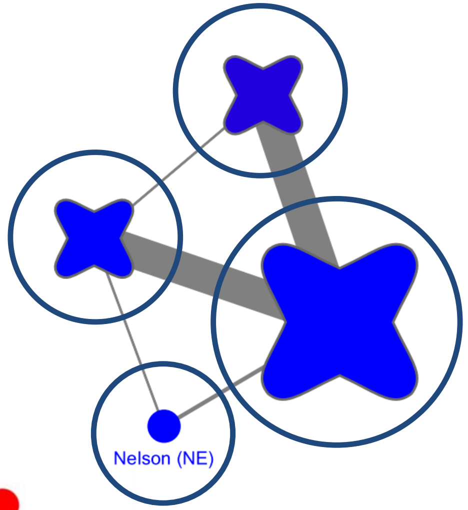


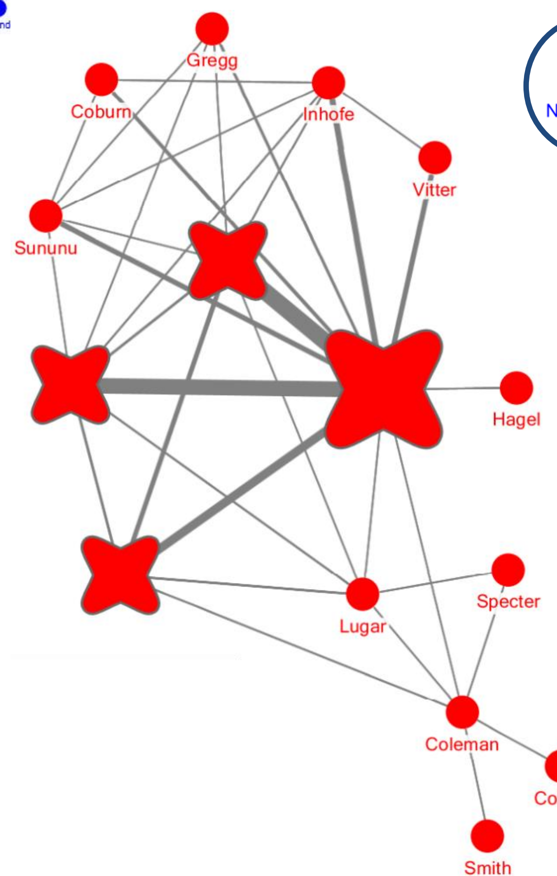
Senate Co-Voting: 70% Agreement

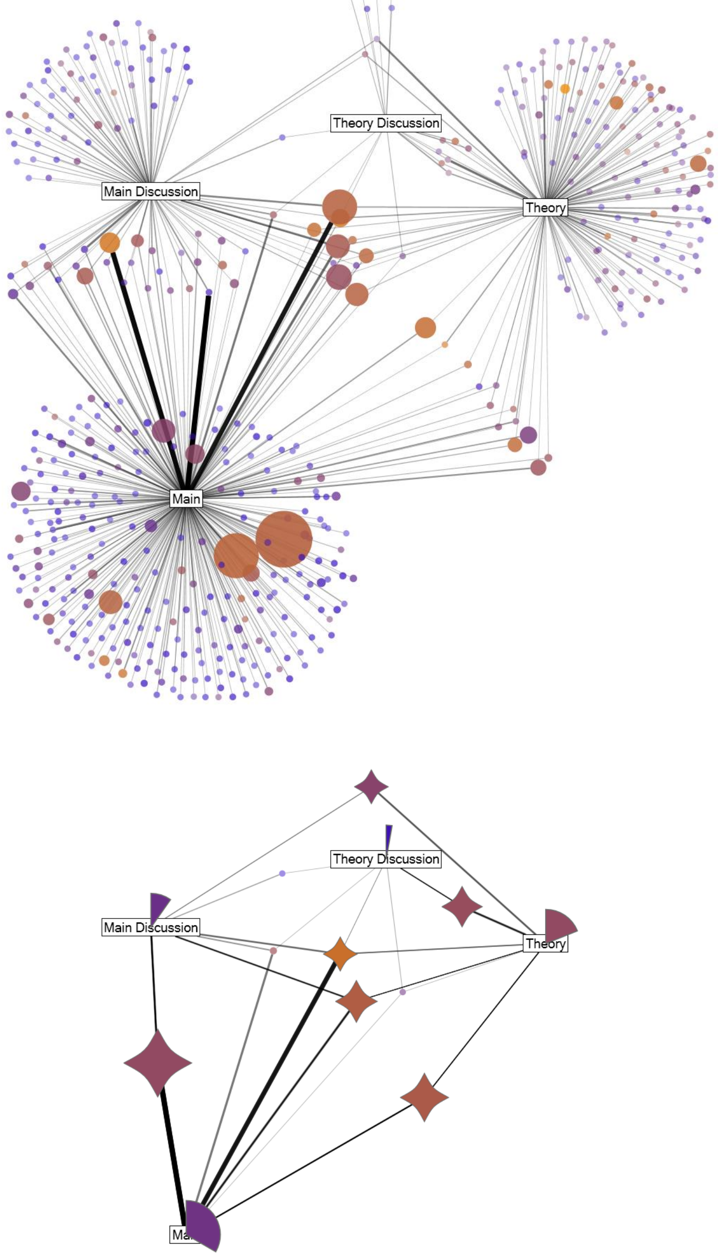


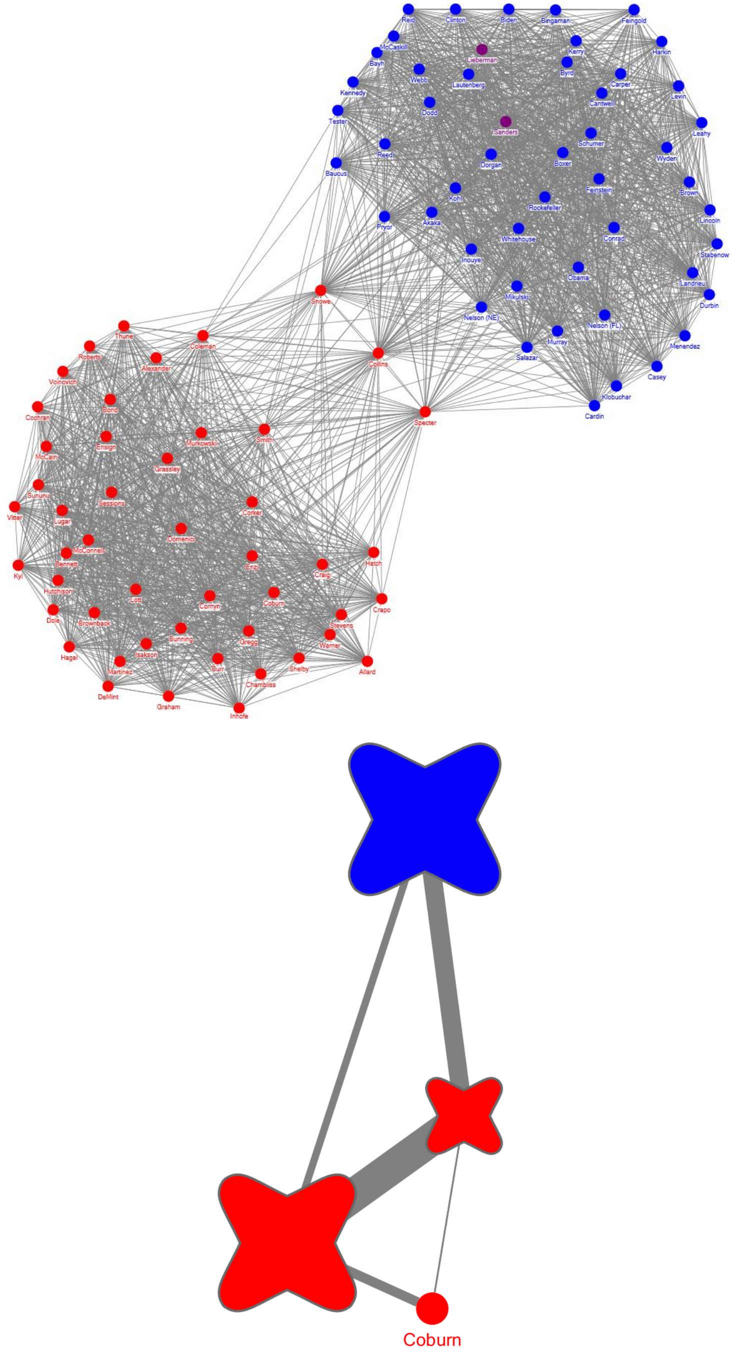
Senate Co-Voting: 80% Agreement

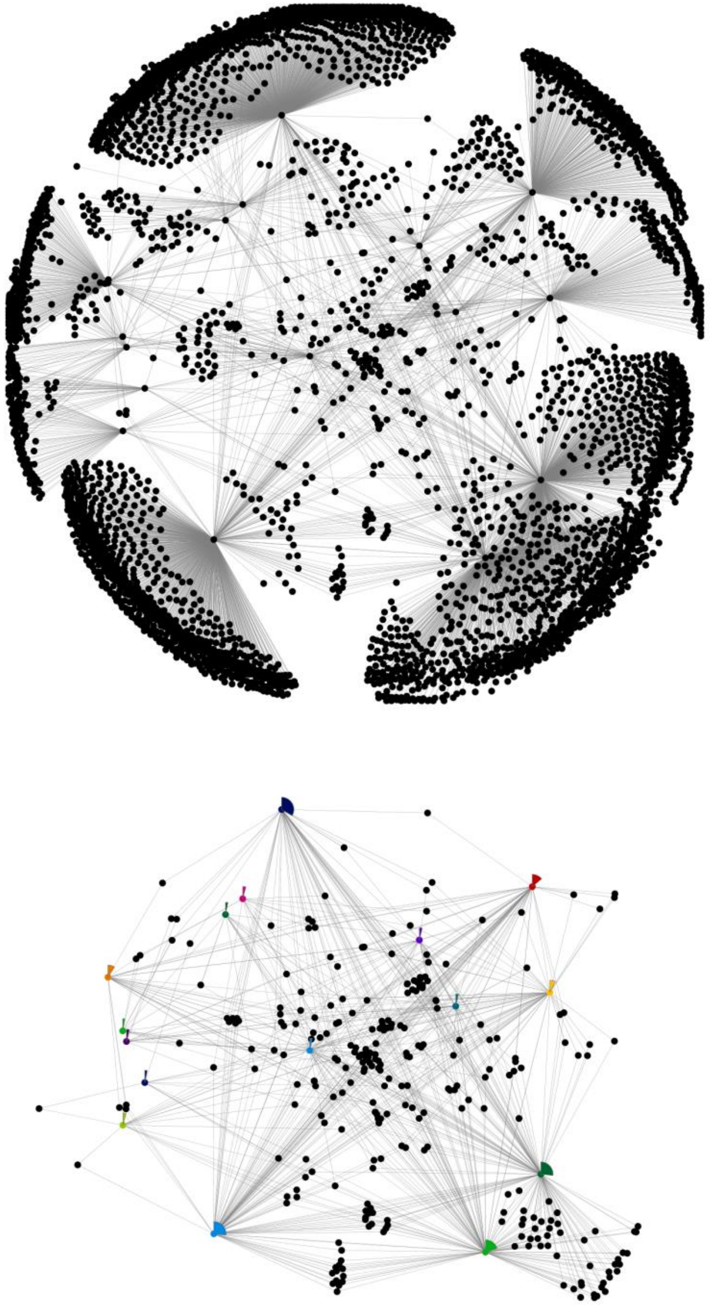







Wiki


Senate


Web Crawl

Controlled Experiment

- Participants: 2 pilot, 36 main
- Data: The Wiki, Senate, and Web networks
- Two groups: control and motif simplification
- 31 questions
- 45 minutes

Controlled Experiment - Tasks

Based on Lee et al. 2006 taxonomy:

1. Node count: About how many nodes are in the network?
2. Cut point: Which individual node would we remove to disconnect the most nodes from the main network?
3. Largest motif & size: Which is the largest (fan | connector | clique) motif and how many nodes does it contain?
4. Labels: Which node has the label “XXX”?
5. Shortest path: What is the length of the shortest path between the two highlighted nodes?
6. Neighbors: Which of the two highlighted nodes has more neighbors?
7. Common Neighbors: How many common neighbors are shared by the two highlighted nodes?
8. Common Neighbors: Which of these two pairs of nodes has more common neighbors?

Visible vs. Simplified Labels

Time in seconds & Accuracy

🟦 Plain

🟥 Simplified