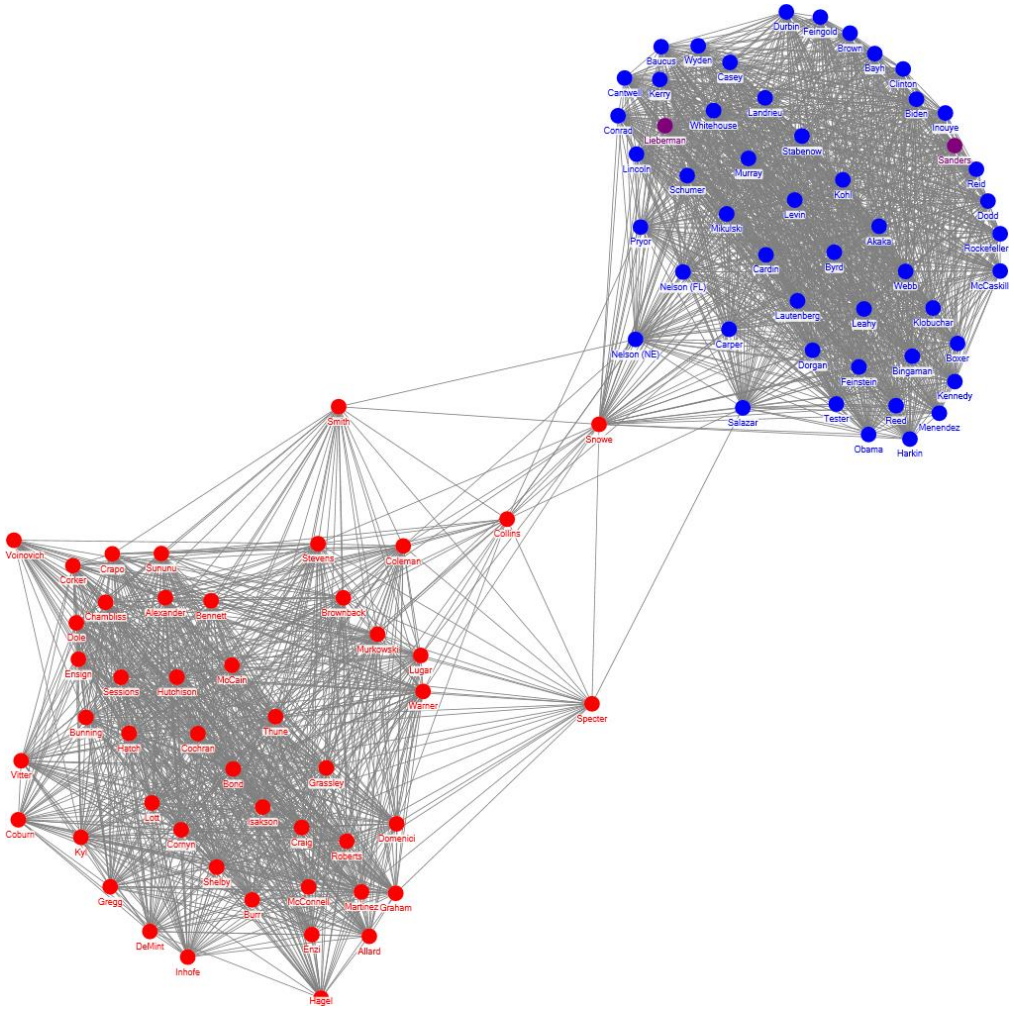
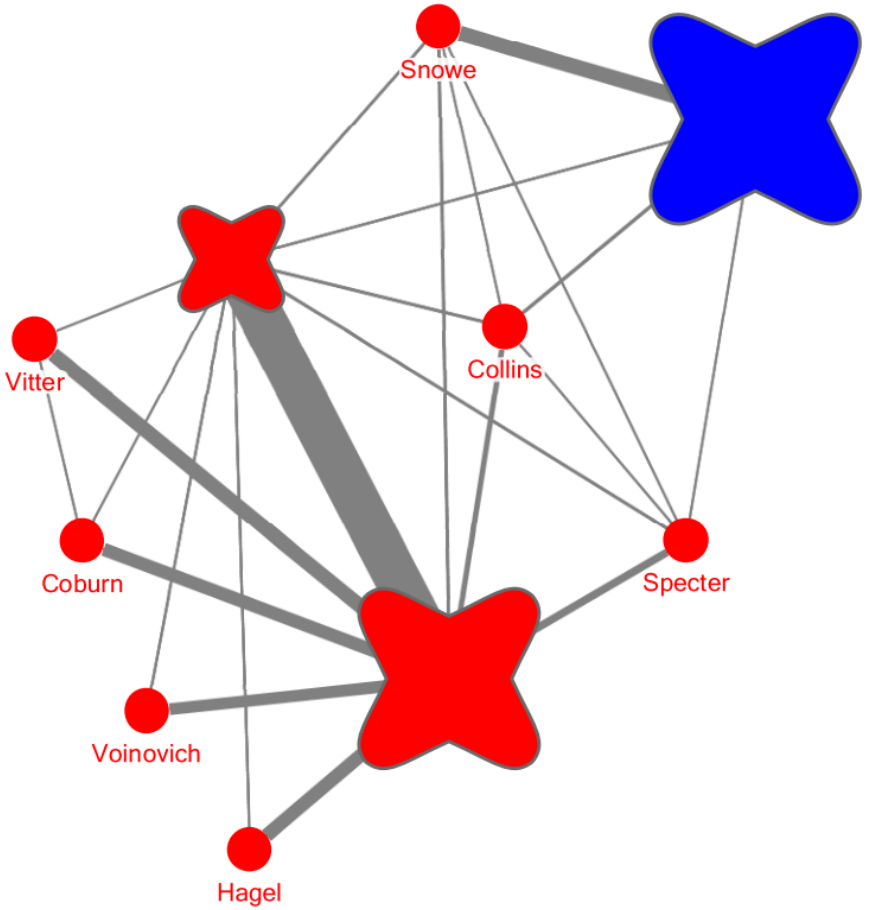
CS 7295-01 Special Topics on Visualization in Network Science





Professor Cody Dunne

https://codydunne.github.io/cs7295-f17/

c.dunne@northeastern.edu

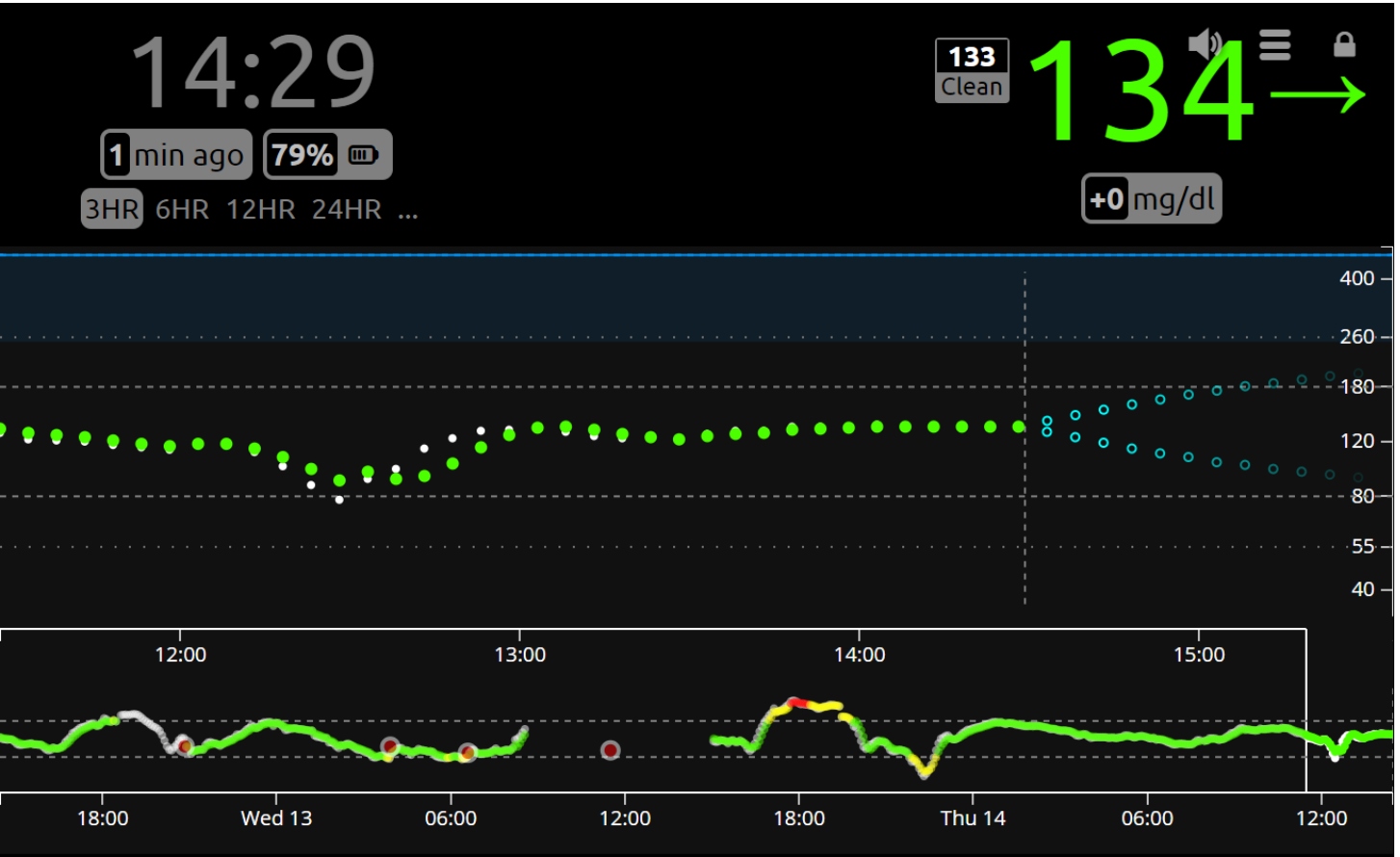
Logistics

Logistics

- Announcements, questions on Piazza
  - Secondary: Office hours, email
- Everyone has a homework 2 team?
- Numbering of HW, project assignments vs. numbering of reading discussions, quizzes
- Grade components updated on syllabus, homework, and project pages
- Grades entered in Blackboard throughout semester

https://codydunne.github.io/cs7295-f17/
http://blackboard.neu.edu

Speak your name!



Grad. Cert. in Digital Humanities

Can complete in addition to your degree. It's an opportunity for you to study with the NULab's interdisciplinary faculty and to join Northeastern's rapidly-growing digital humanities community.

- Students take 12 semester hours total:
  - three semesters of a one-hour project seminar workshop (NULab project seminar (INSH 7910) has fall openings)
  - introduction to digital humanities class
  - two electives

- Final project, part of a thesis/dissertation or part of on a larger DH project, e.g.:
  - Women Writers Project http://wwp.northeastern.edu/
  - Viral Texts http://viraltexts.northeastern.edu/

- Reach out to Sarah Connell sa.connell@northeastern.edu, assistant director of NULab http://www.northeastern.edu/nulab

http://northeastern.edu/nulab/dh-certificate

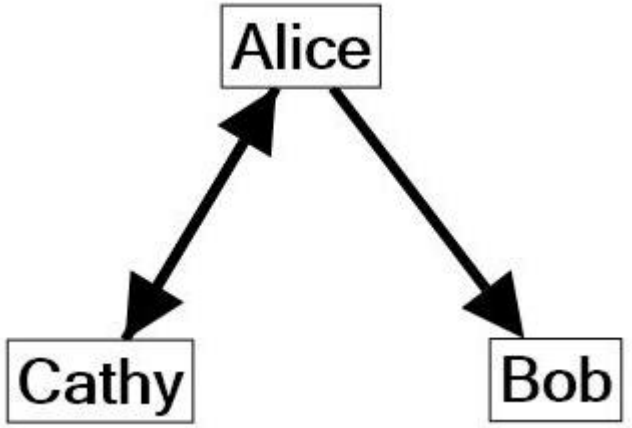
Upcoming Deadlines

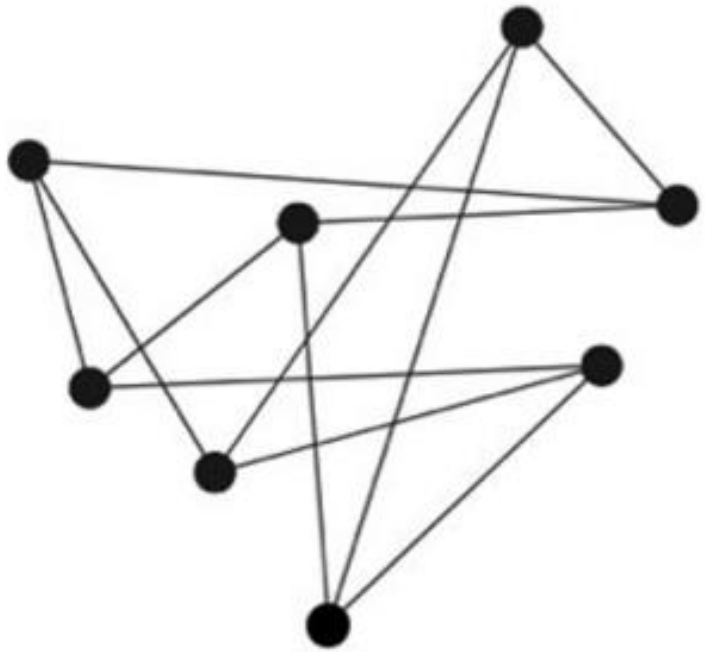
https://codydunne.github.io/cs7295-f17/schedule/

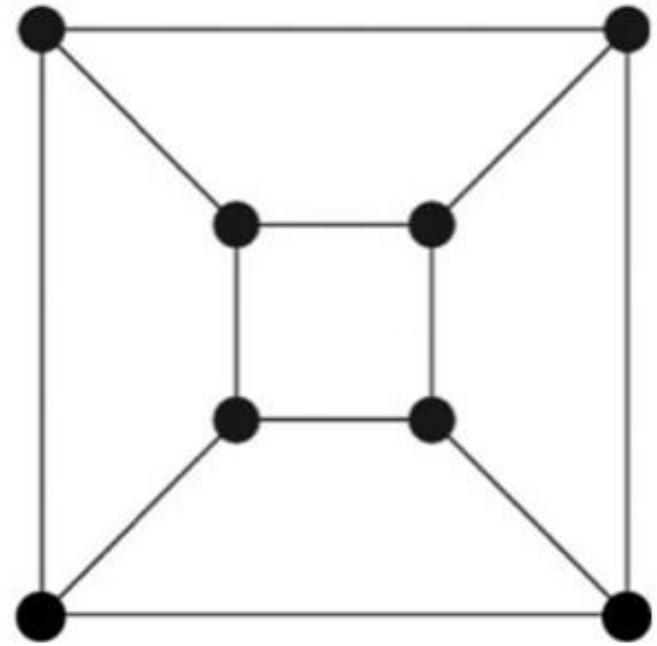
Quiz!

Graph ≈ Network
Node ≈ Vertex ≈ Entity
Edge ≈ Link ≈ Relationship

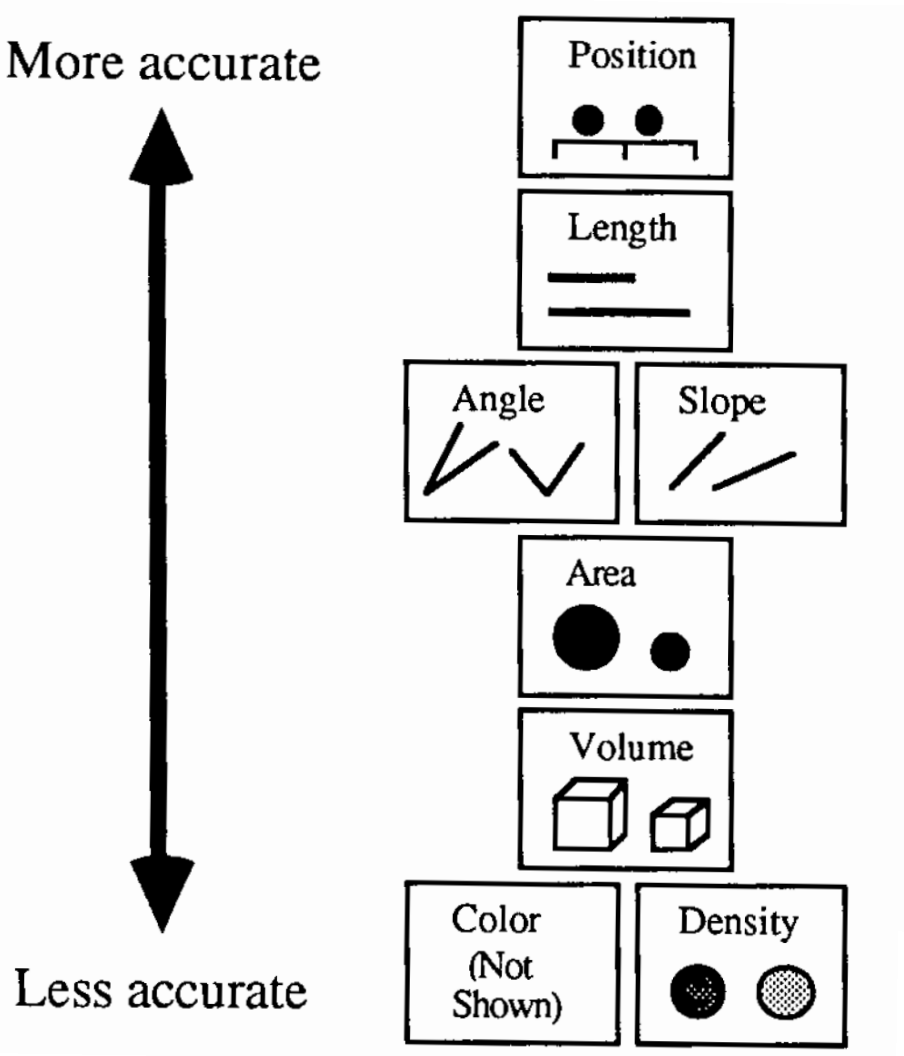
Node 1	Node 2
Alice	Bob
Alice	Cathy
Cathy	Alice

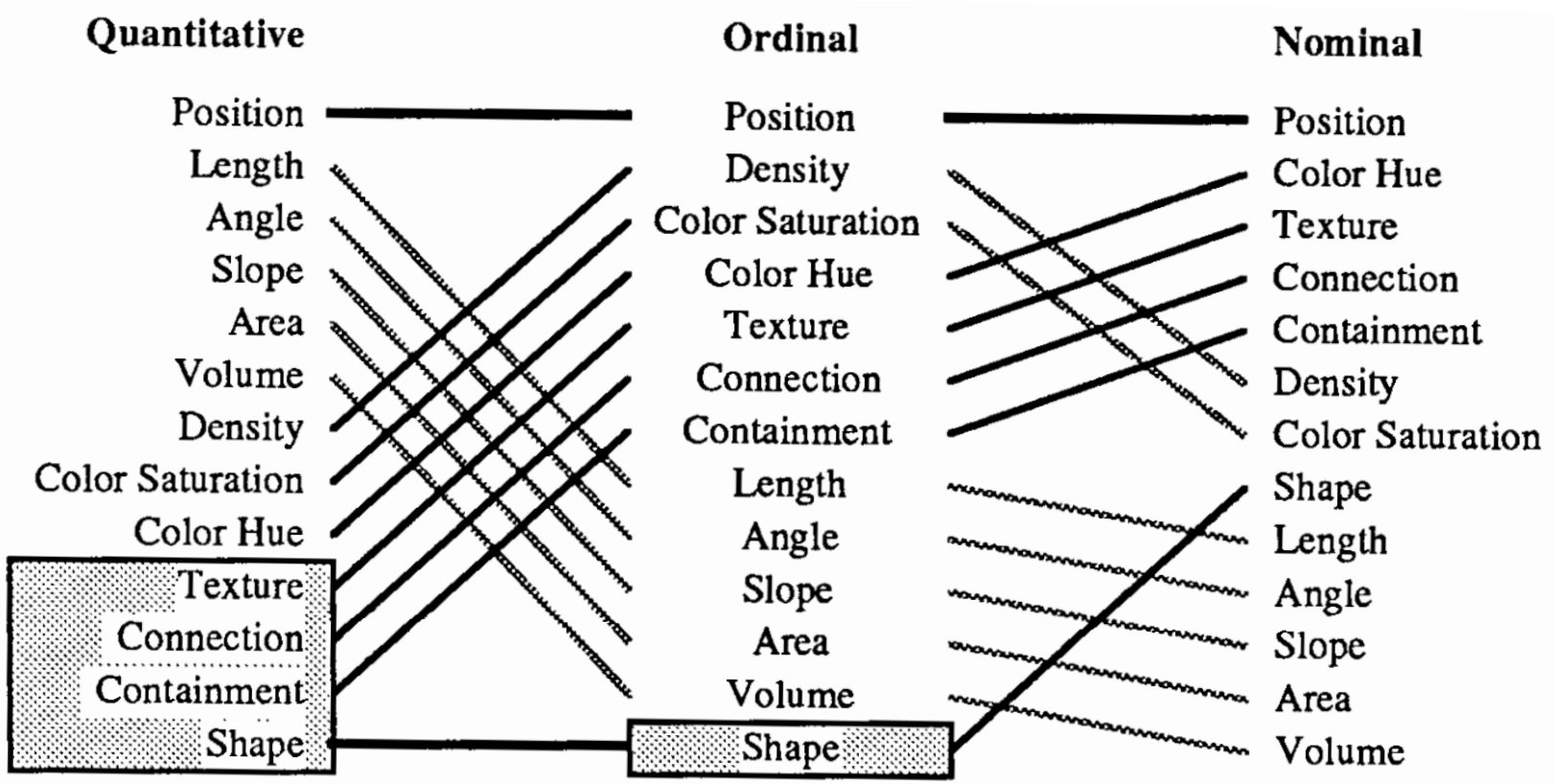






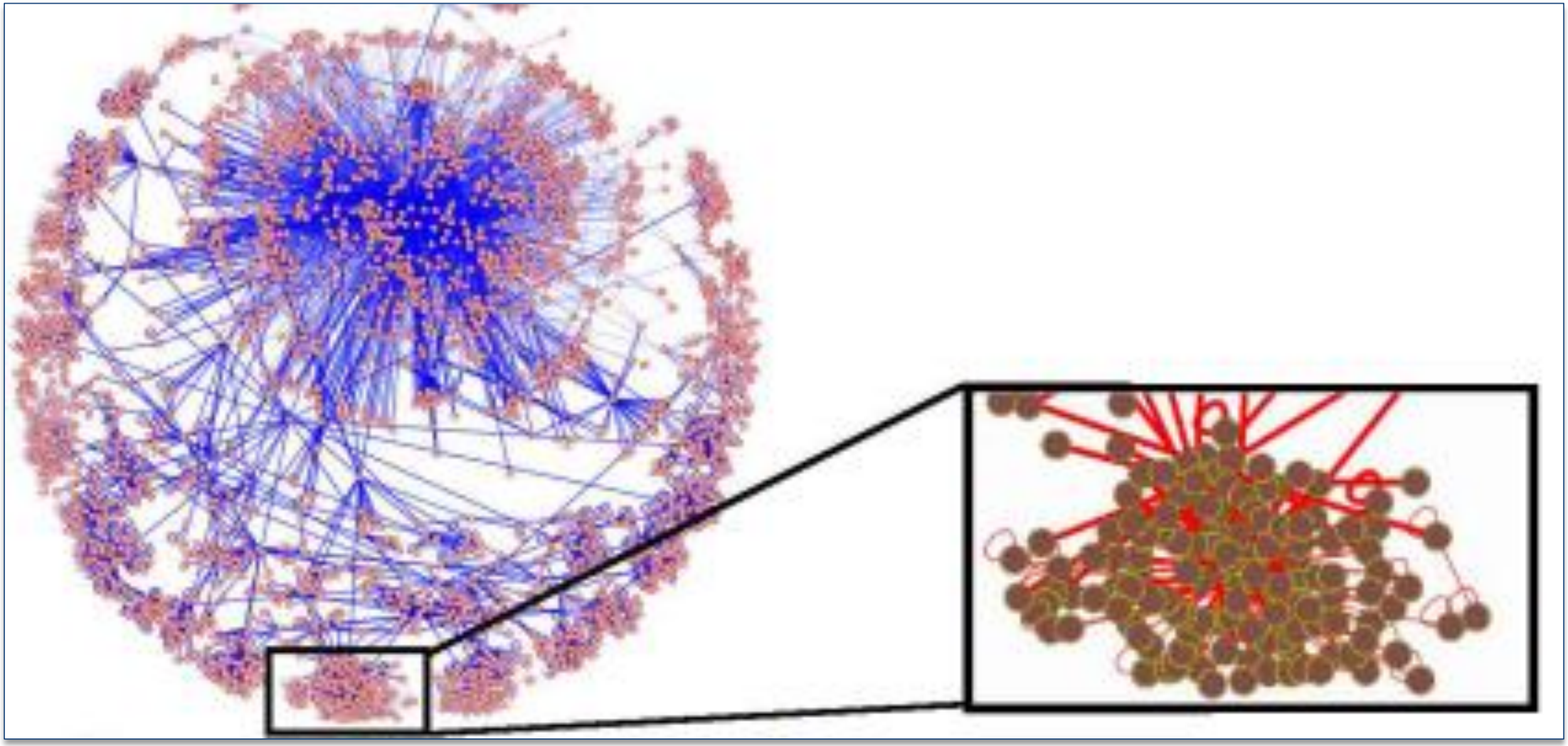
Quantitative Tasks



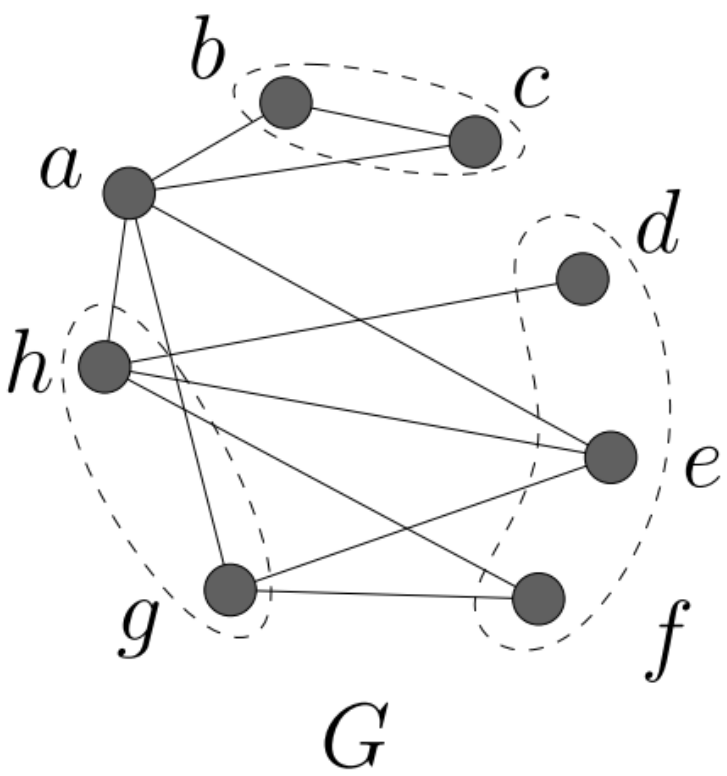


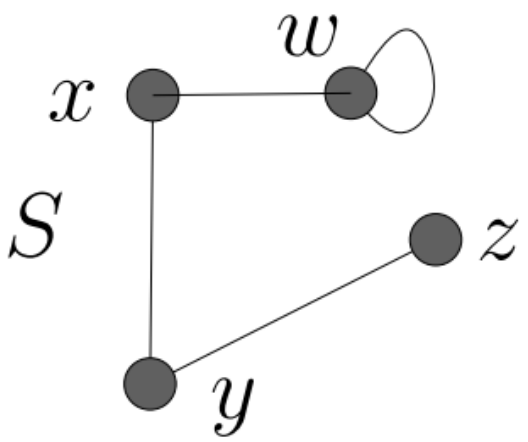
Mackinlay, 1986

Topology Aggregation



Navlakha et al., 2008





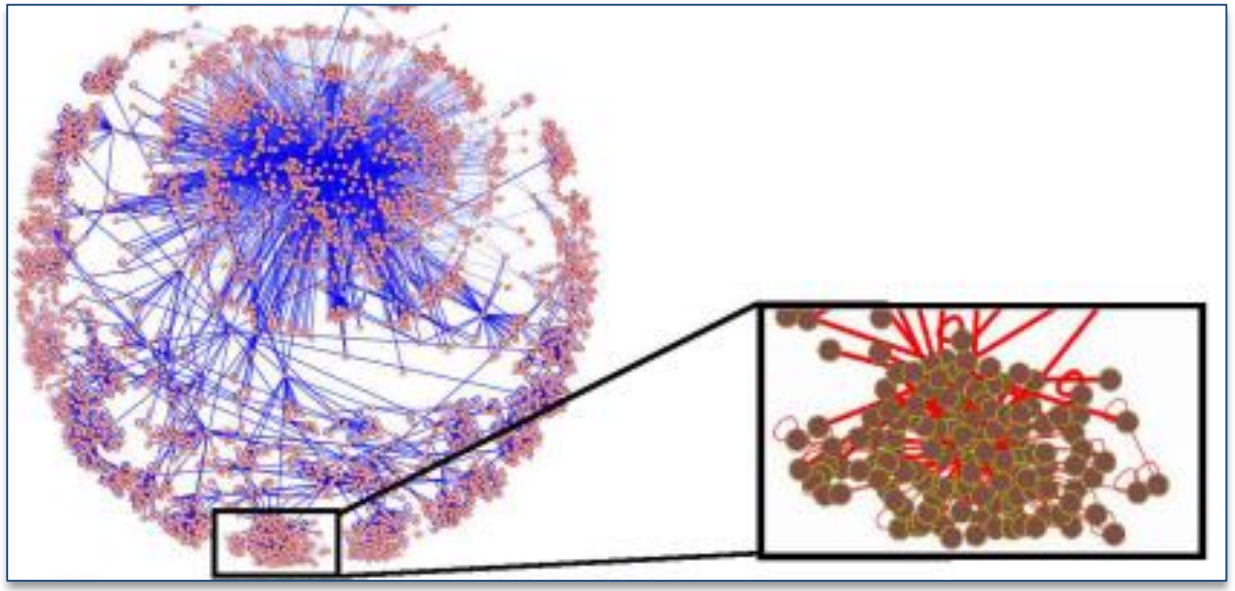
C

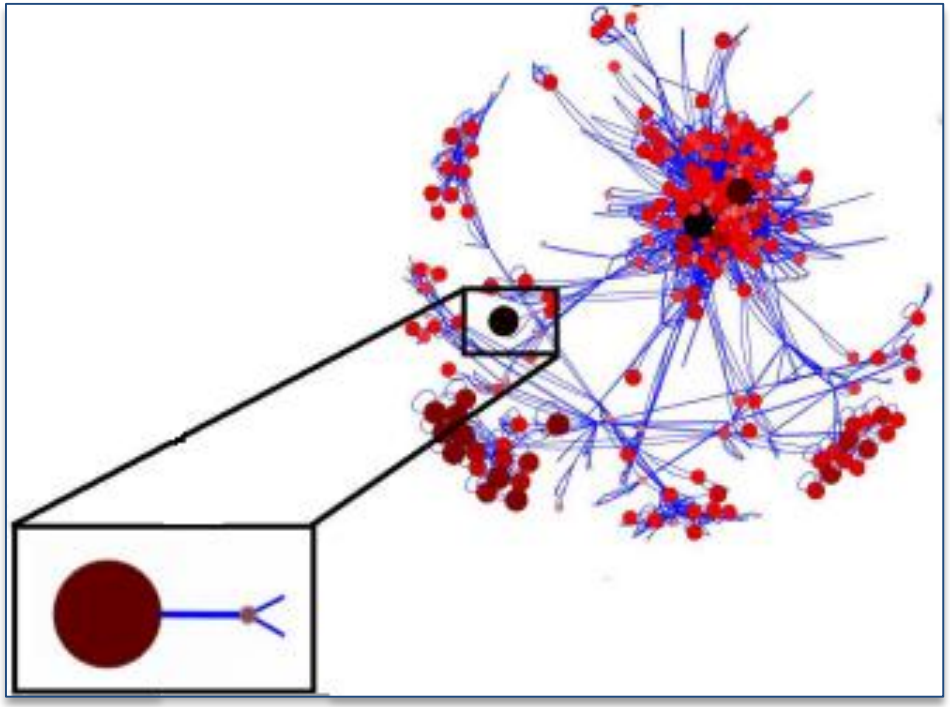
$+(a, e)$
$-(g, d)$

$A_w = \{b, c\}$	$A_x = \{a\}$
$A_y = \{g, h\}$	$A_z = \{d, e, f\}$

Figure 1: The two part graph representation. The LHS shows the original graph, while the RHS contains the graph summary (S), corrections (C), and the supernode mapping.

Navlakha et al., 2008





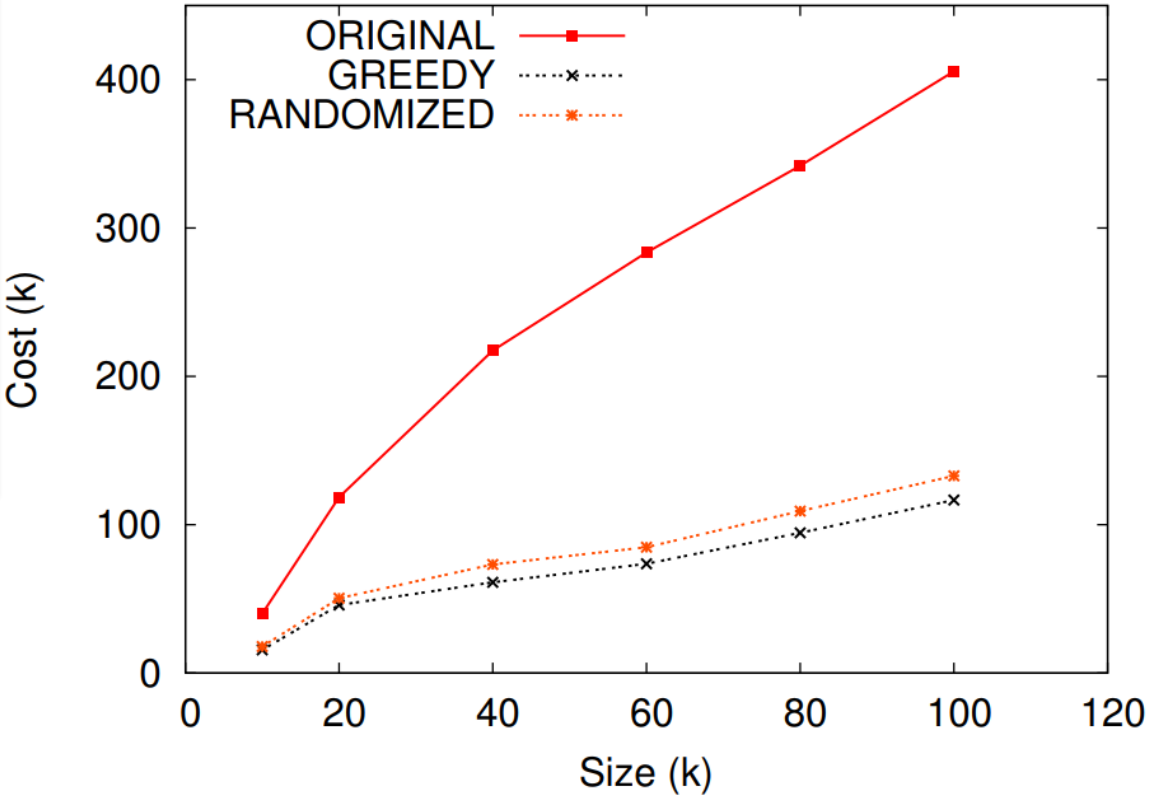


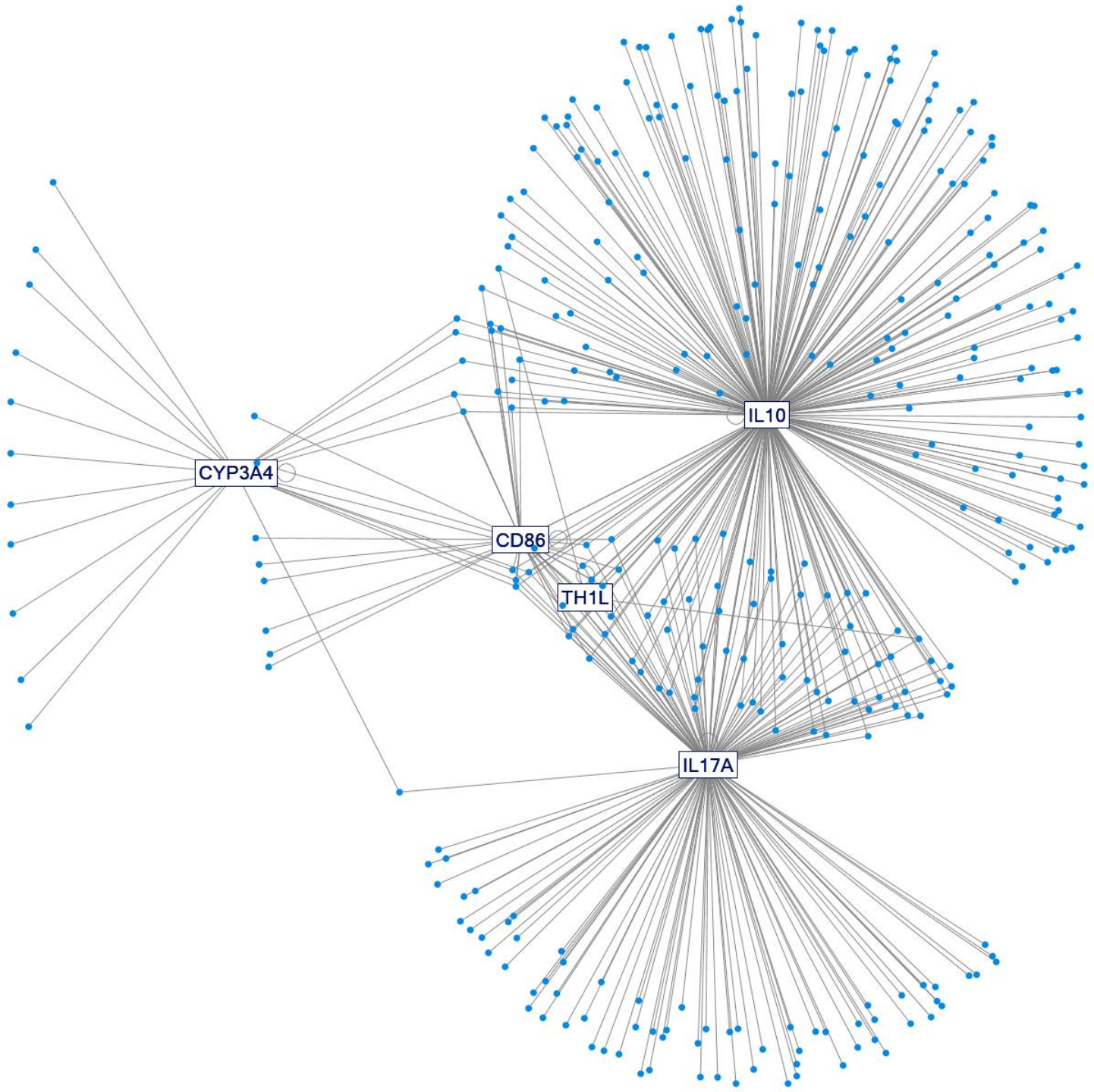
Figure 3: Comparison of costs of representations computed by GREEDY and RANDOMIZED on the CNR dataset.

Navlakha et al., 2008

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

Query genes

Conditions



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

Query genes

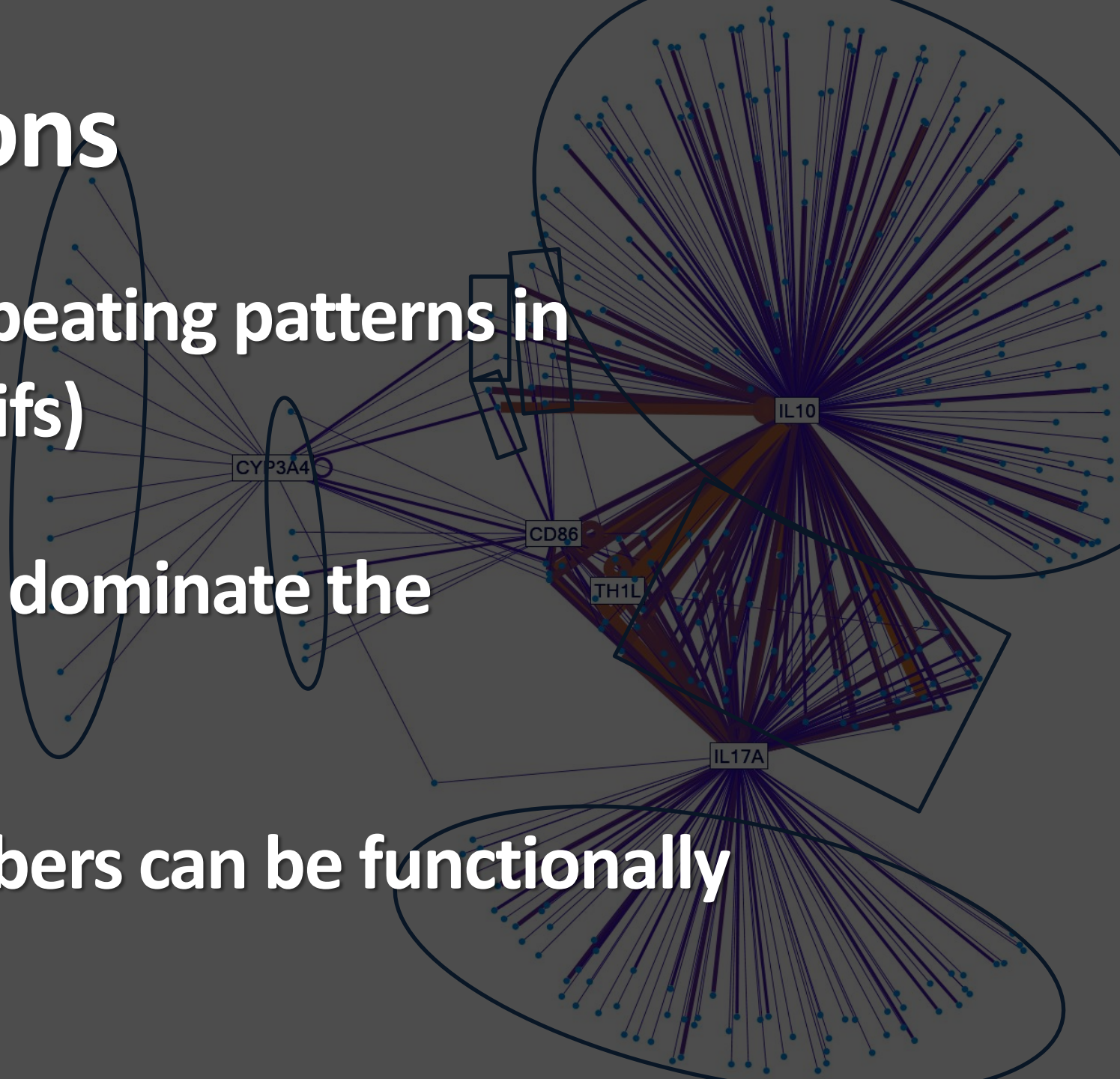
Conditions

Observations

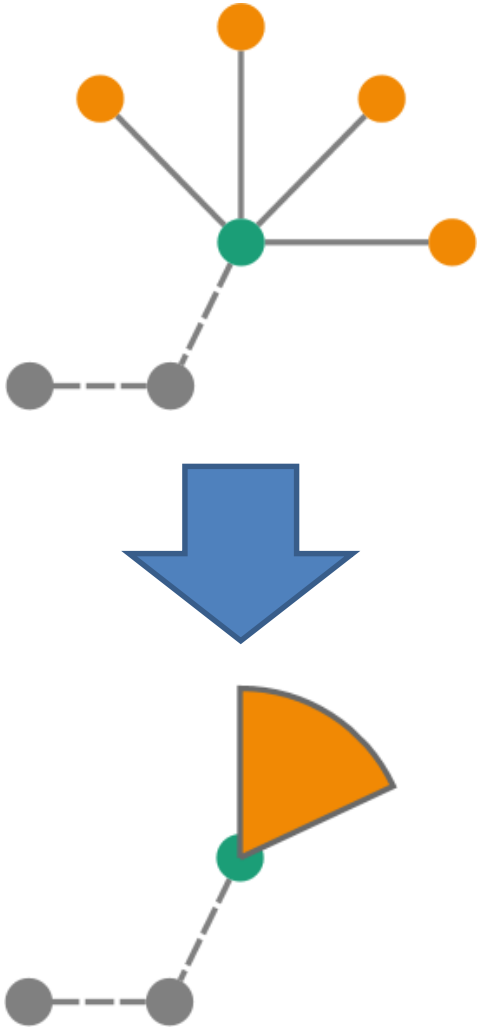
1: There are repeating patterns in networks (motifs)

2: Motifs often dominate the visualization

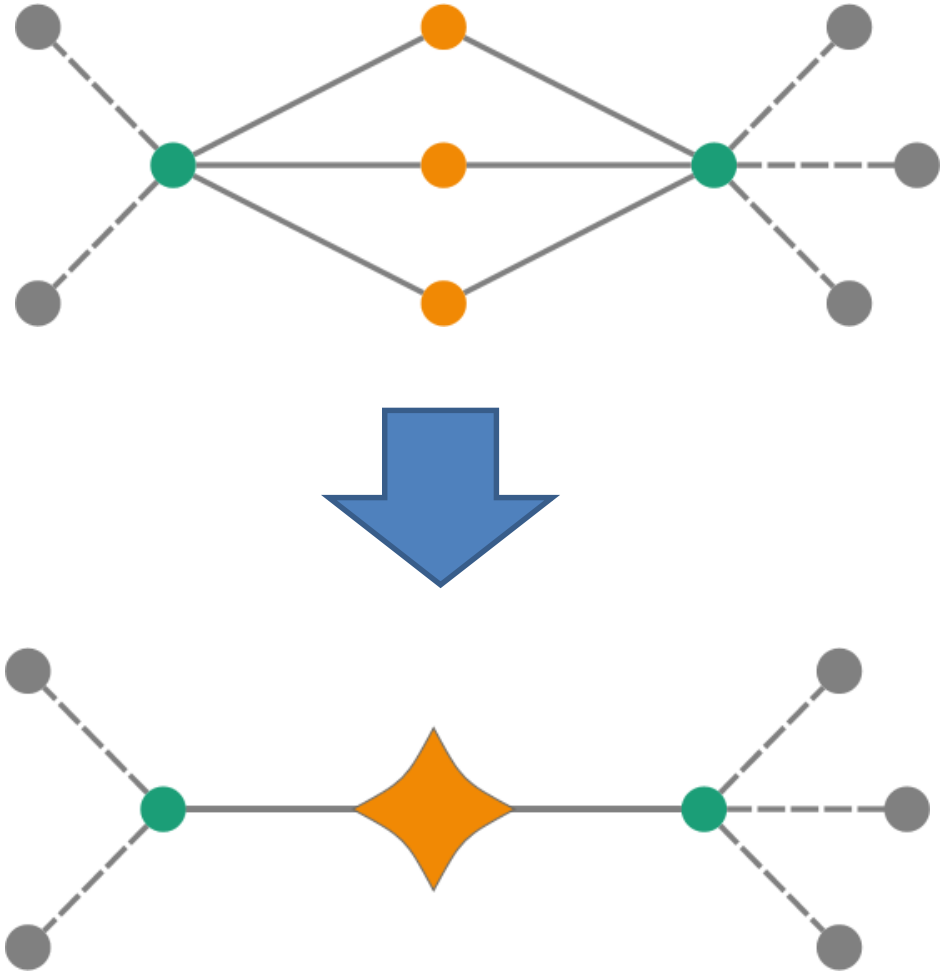
3: Motifs members can be functionally equivalent



Fan Motif

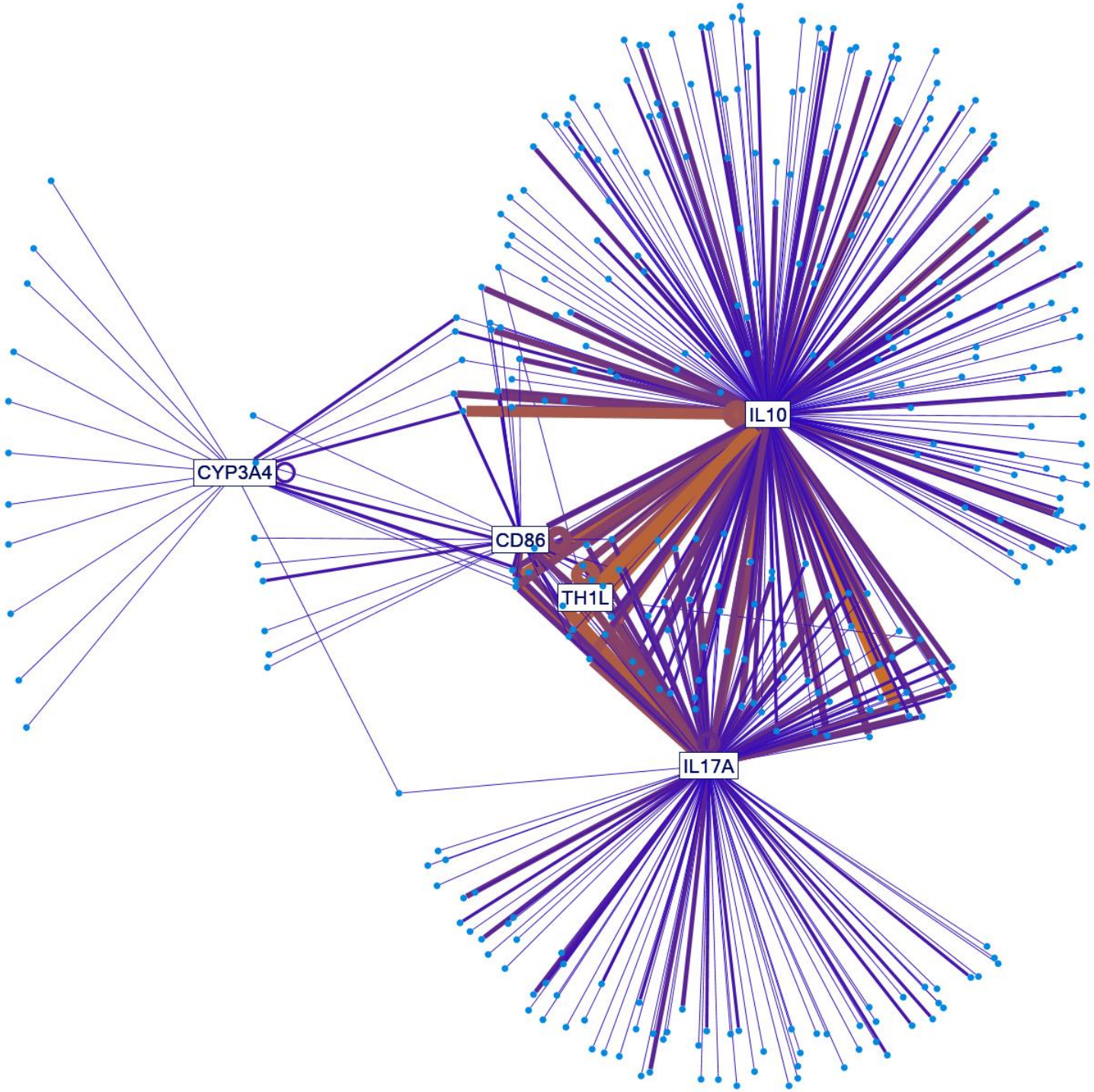


2-Connector Motif



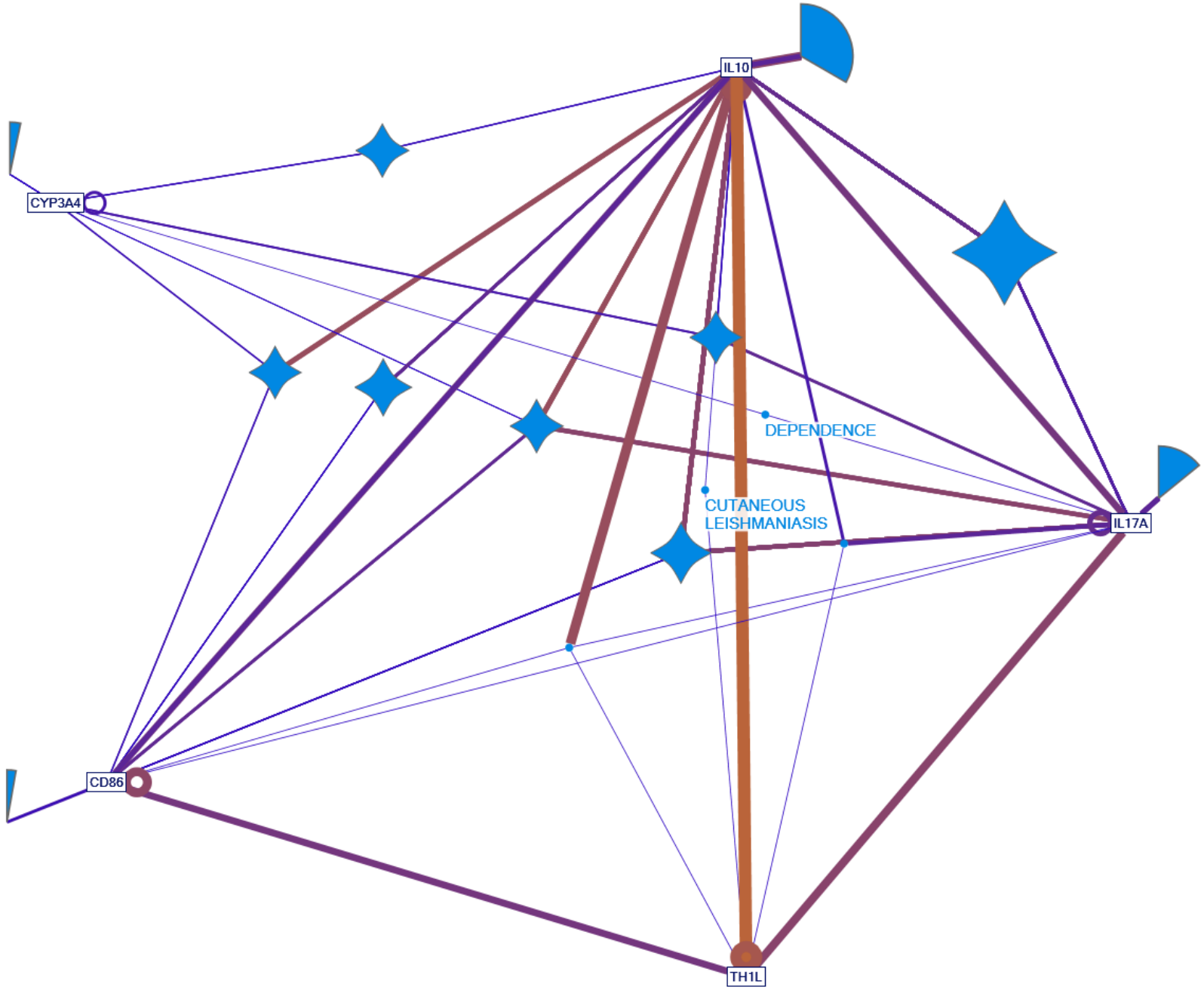
Query genes

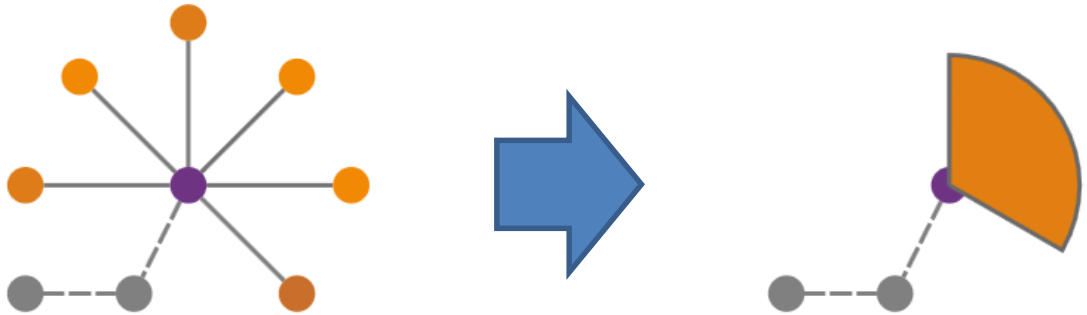
Conditions

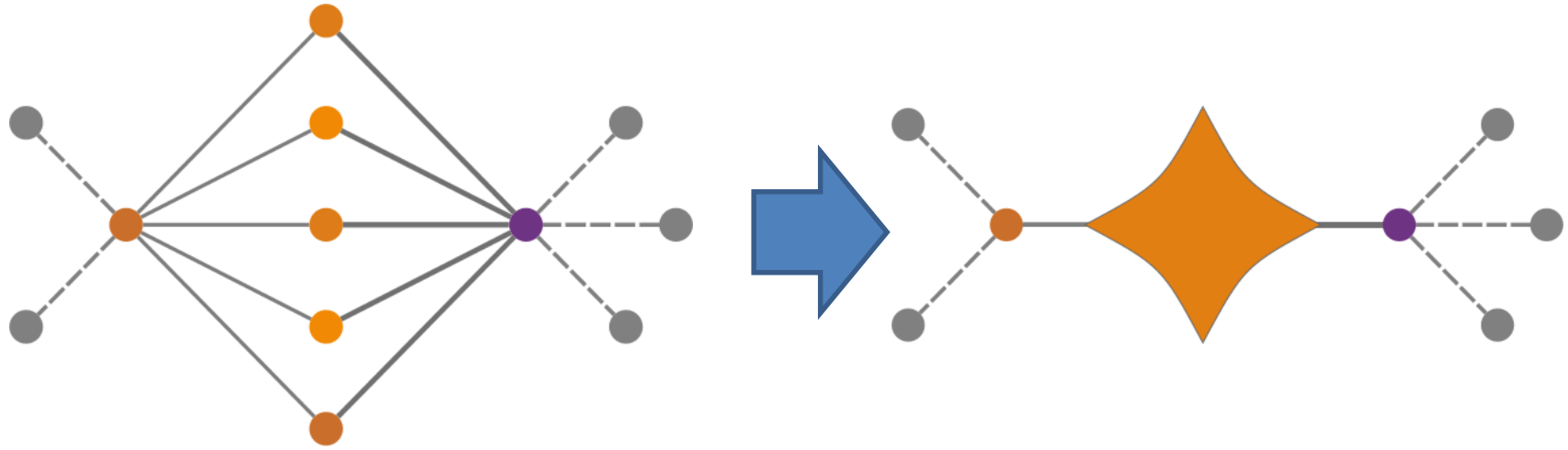


Query genes

Conditions, some in motifs

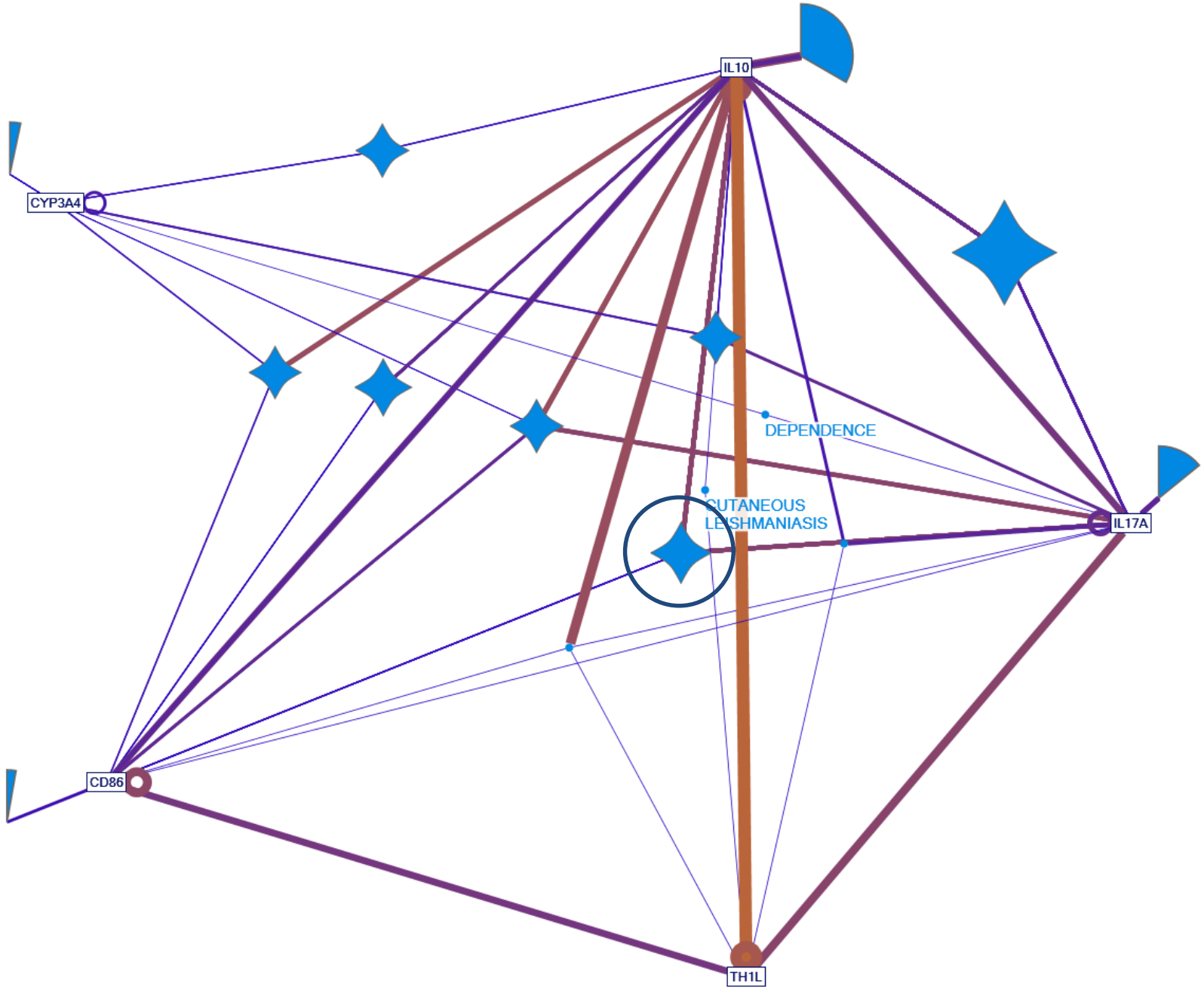






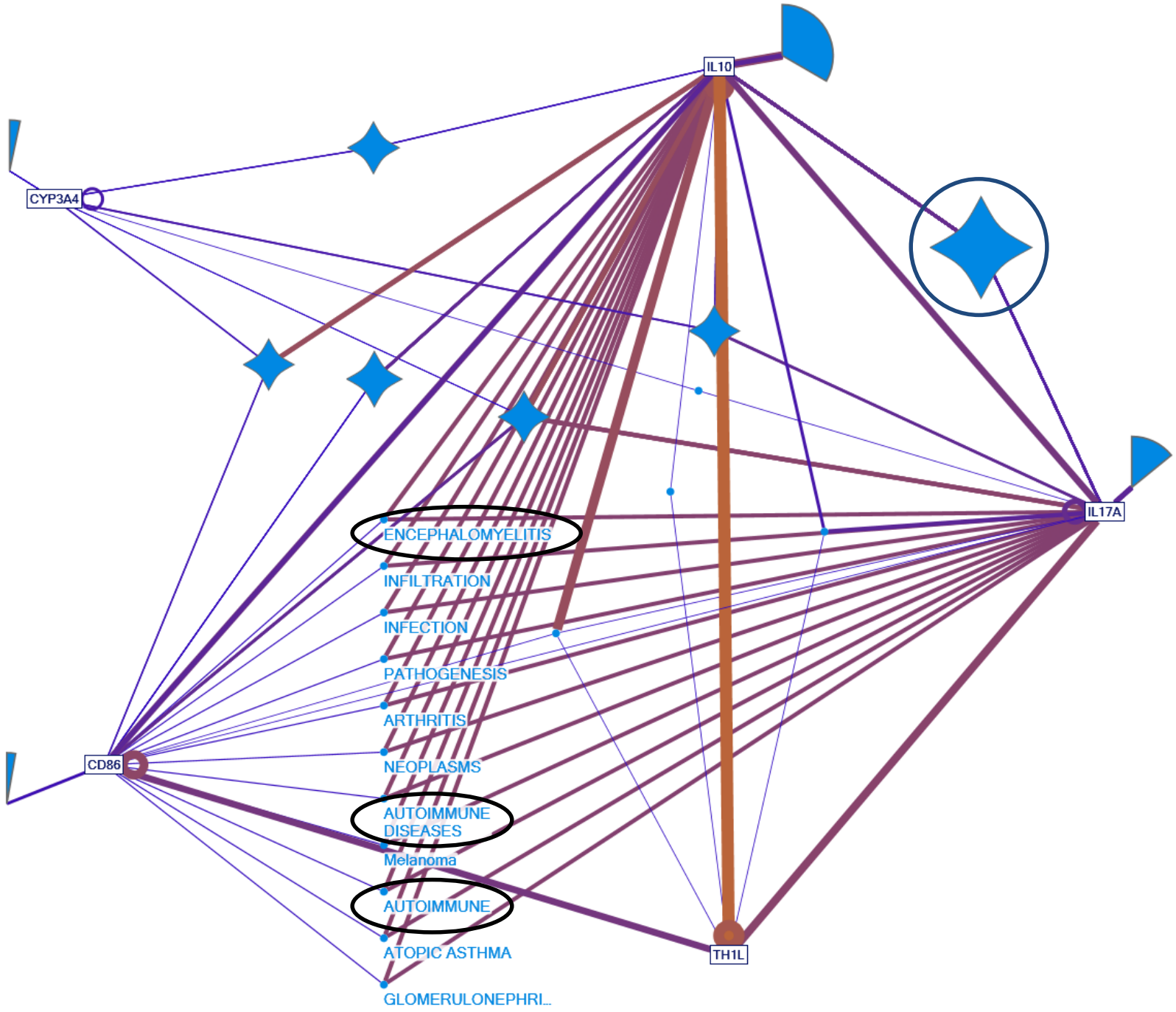
Query genes

Conditions, some in motifs



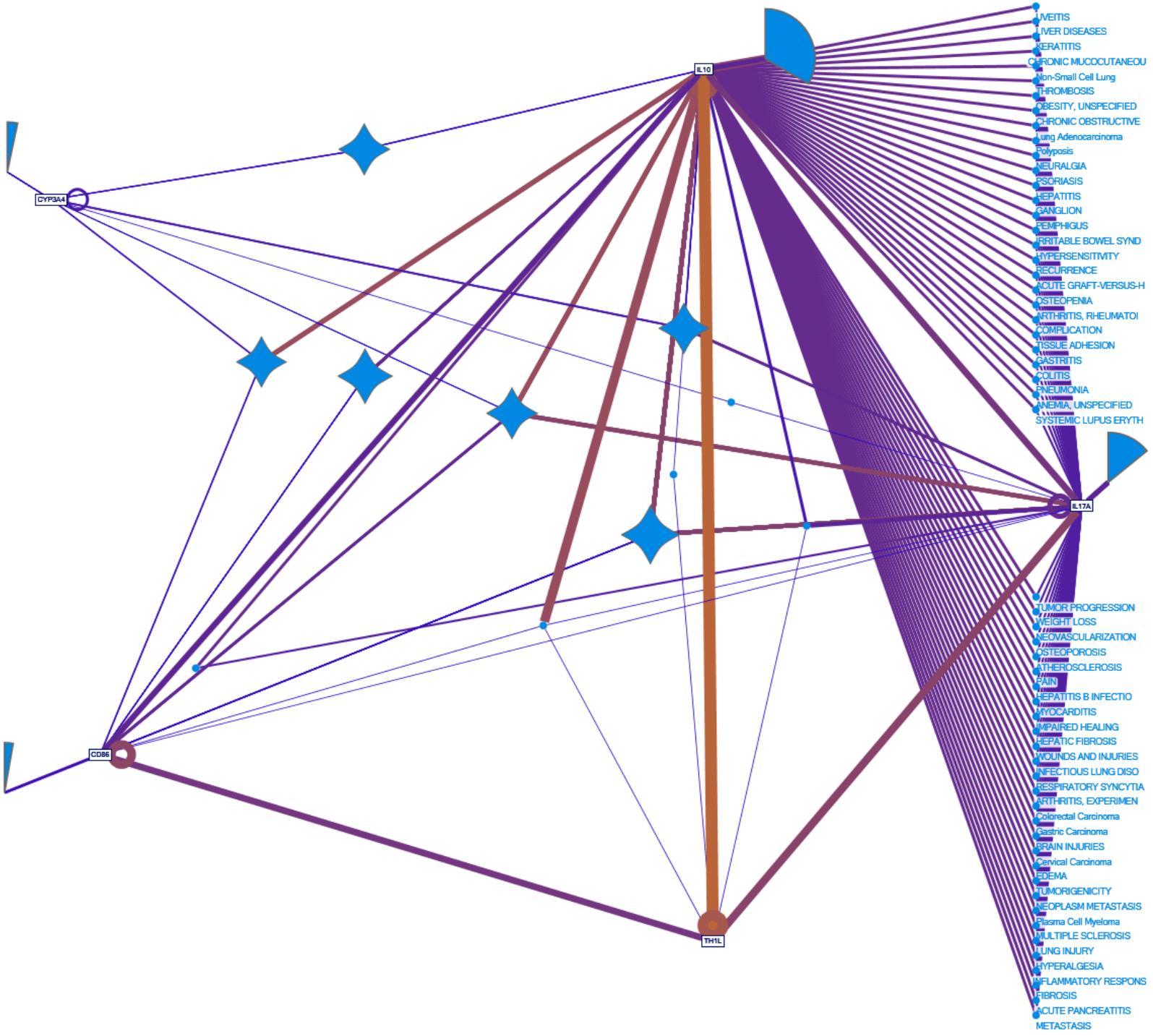
Query genes

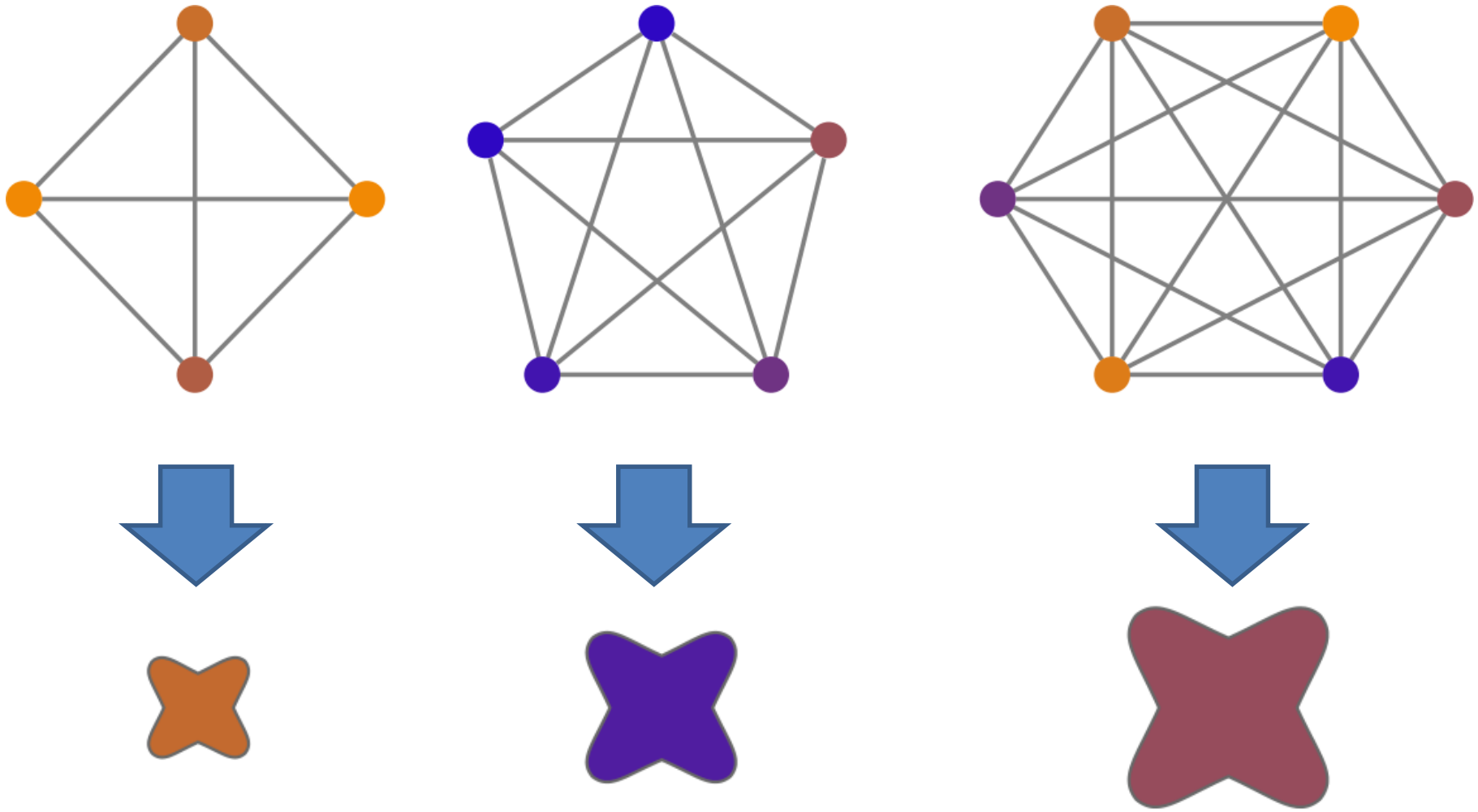
Conditions, some in motifs



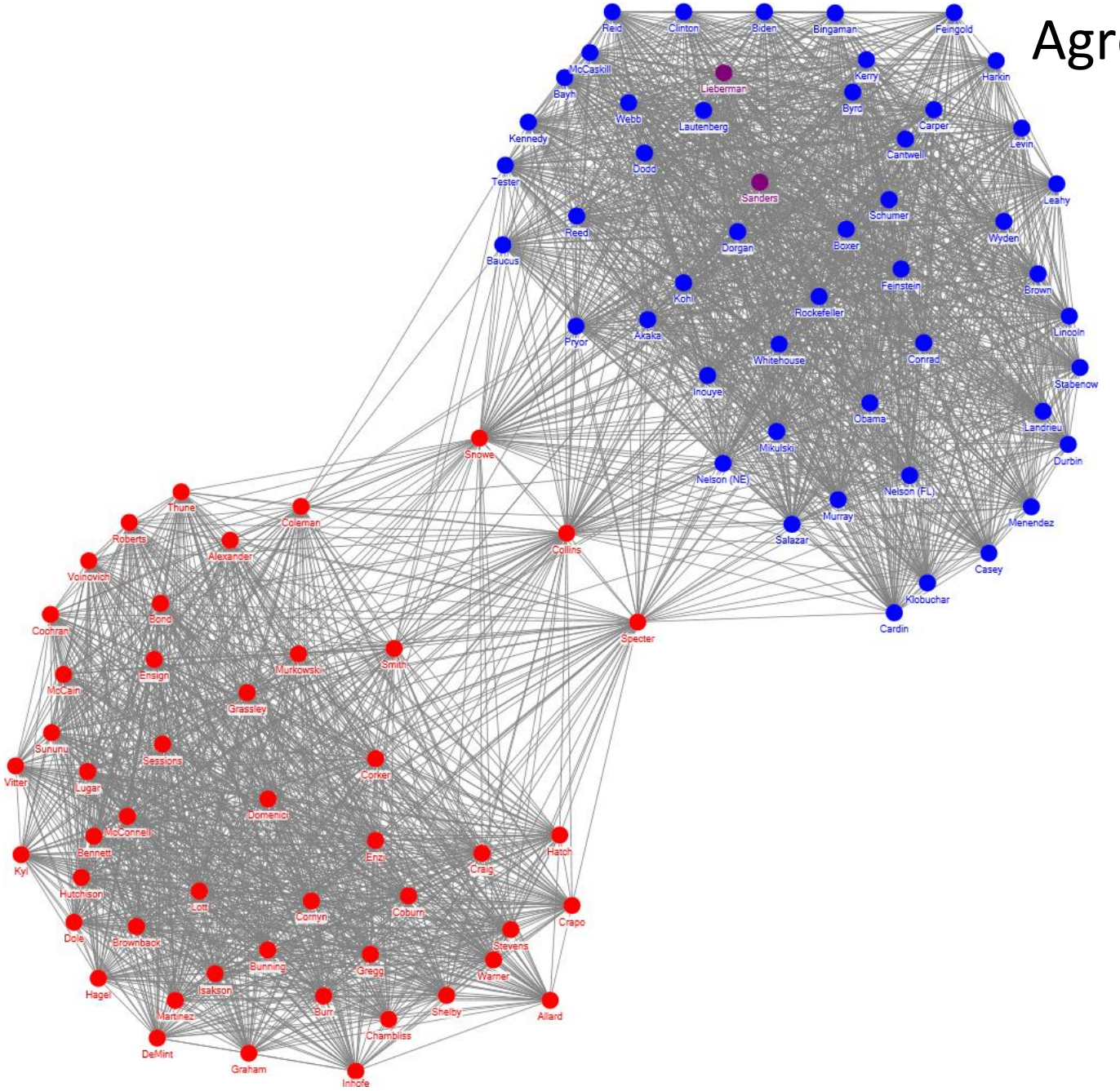
Query genes

Conditions, some in motifs

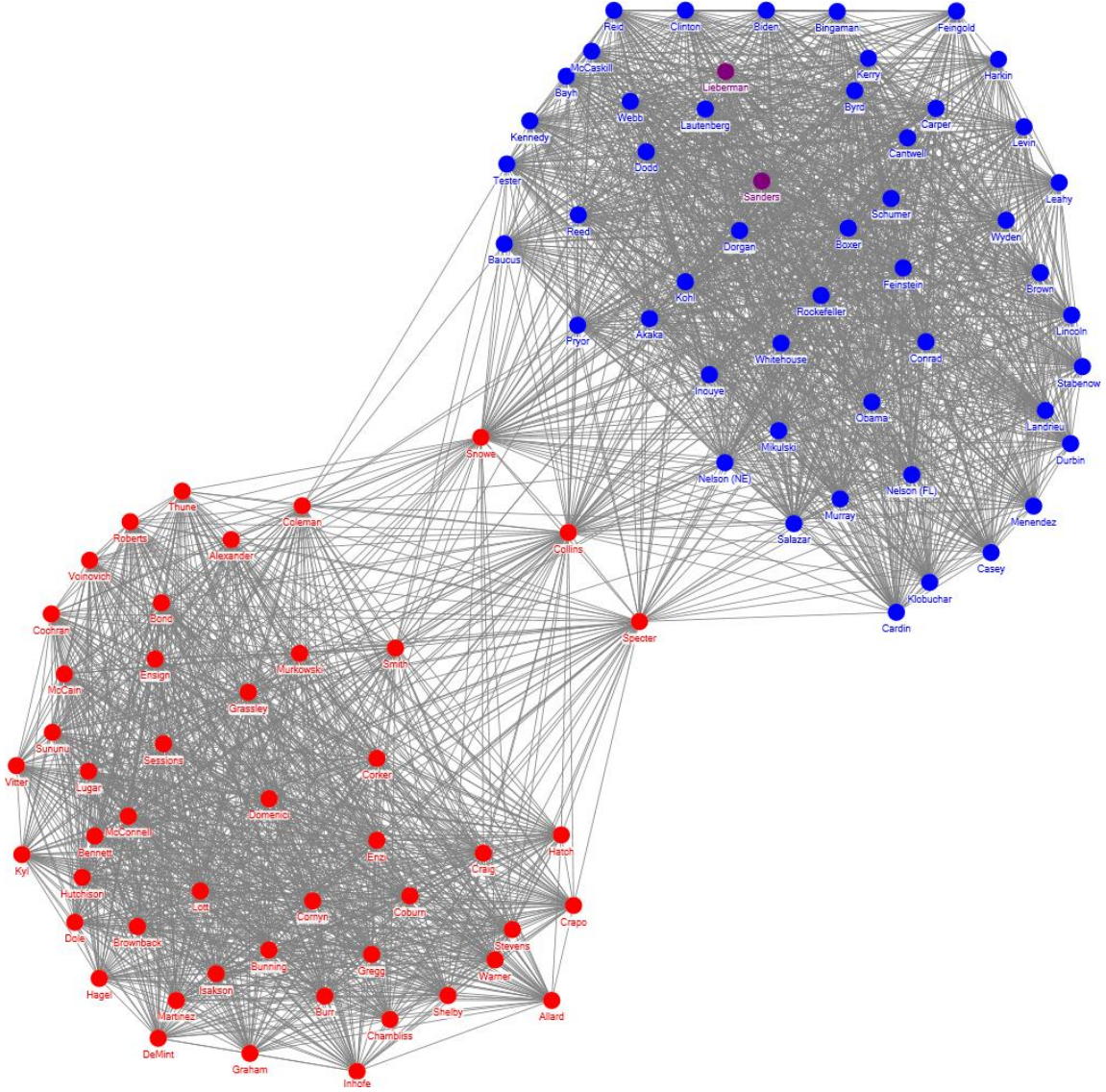


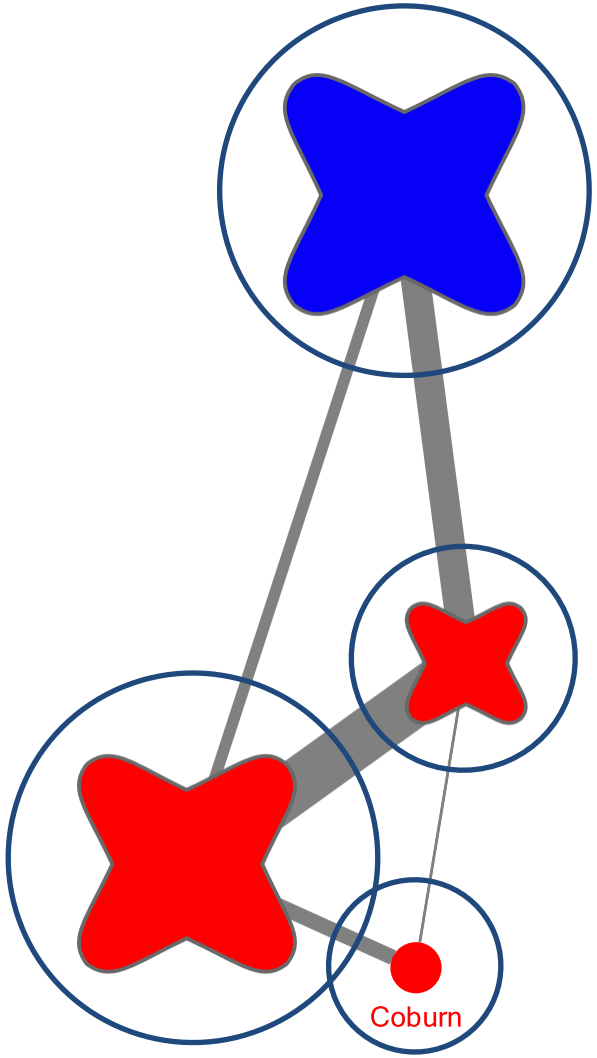


Senate Co-Voting: 65% Agreement

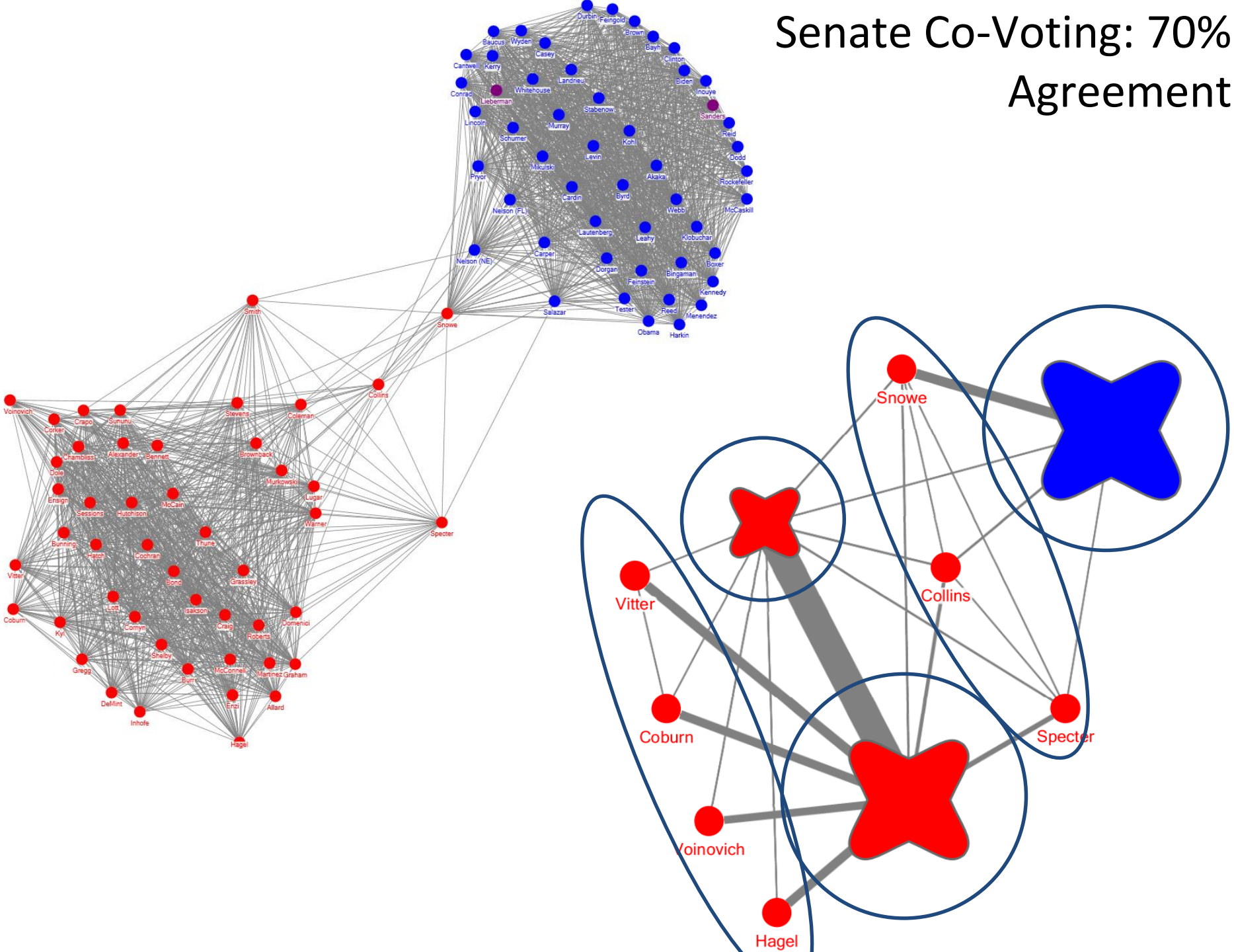


Senate Co-Voting: 65% Agreement

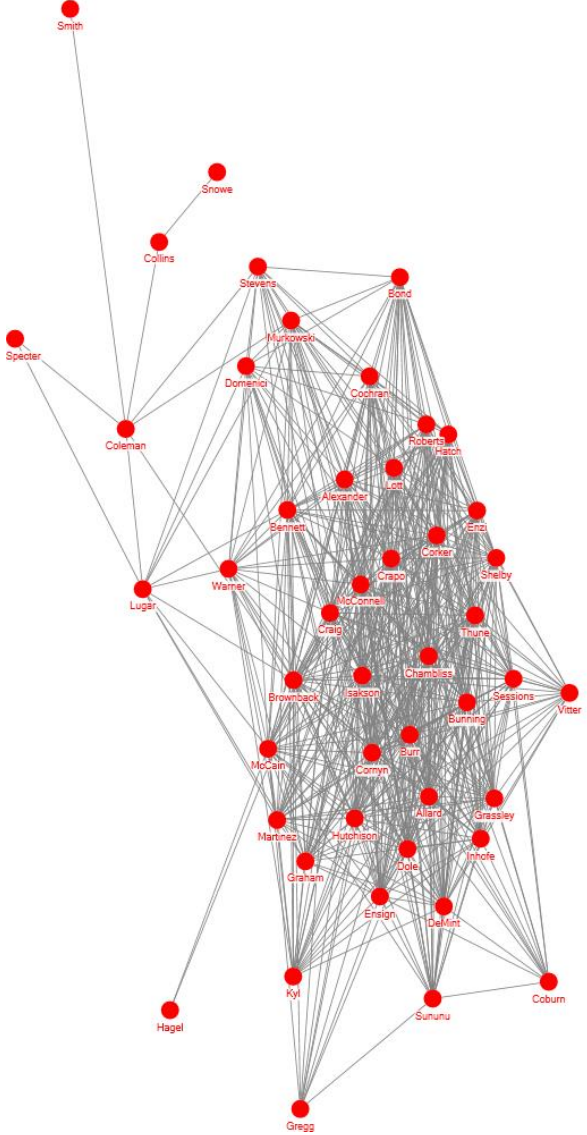


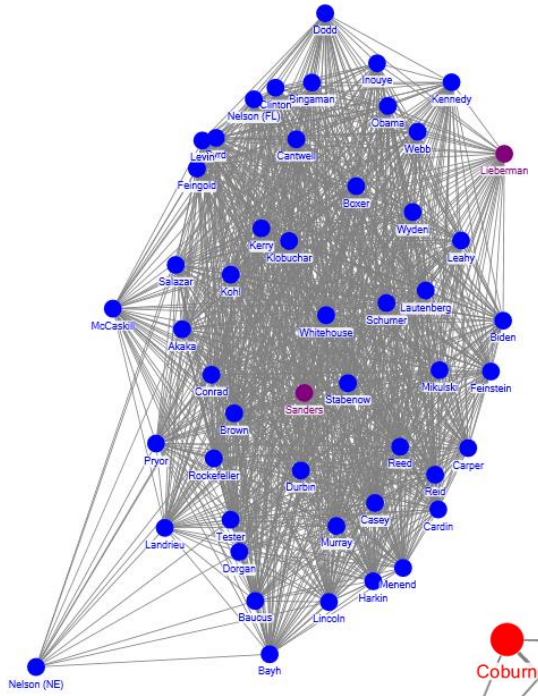


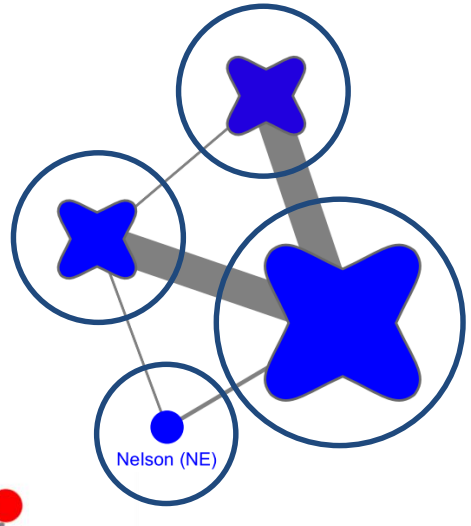
Senate Co-Voting: 70% Agreement

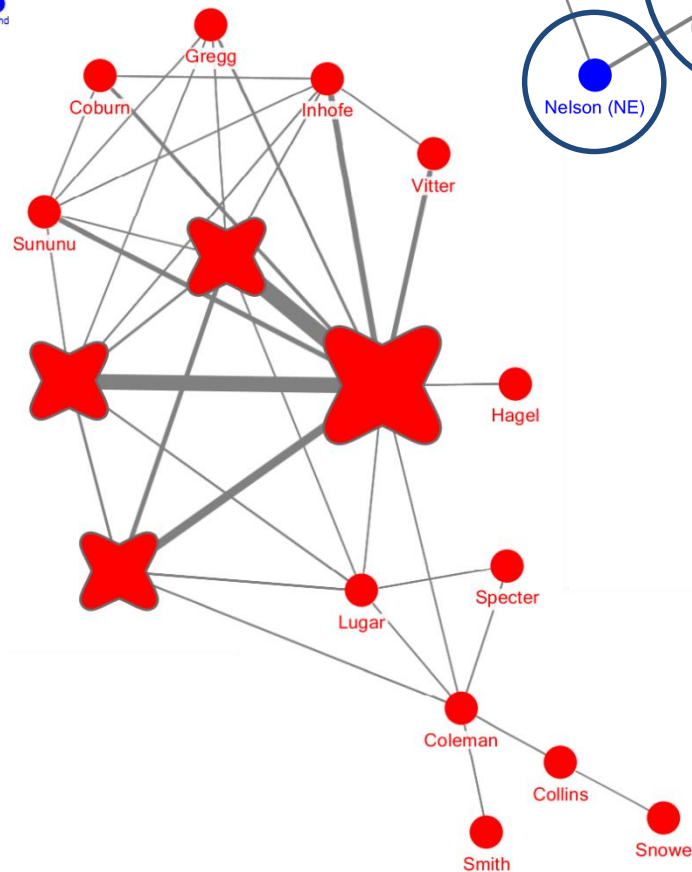


Senate Co-Voting: 80% Agreement

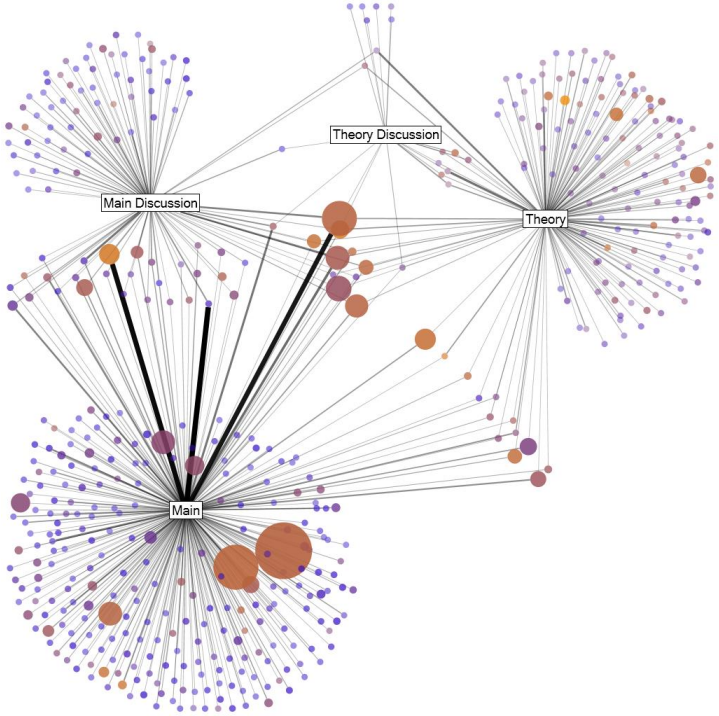




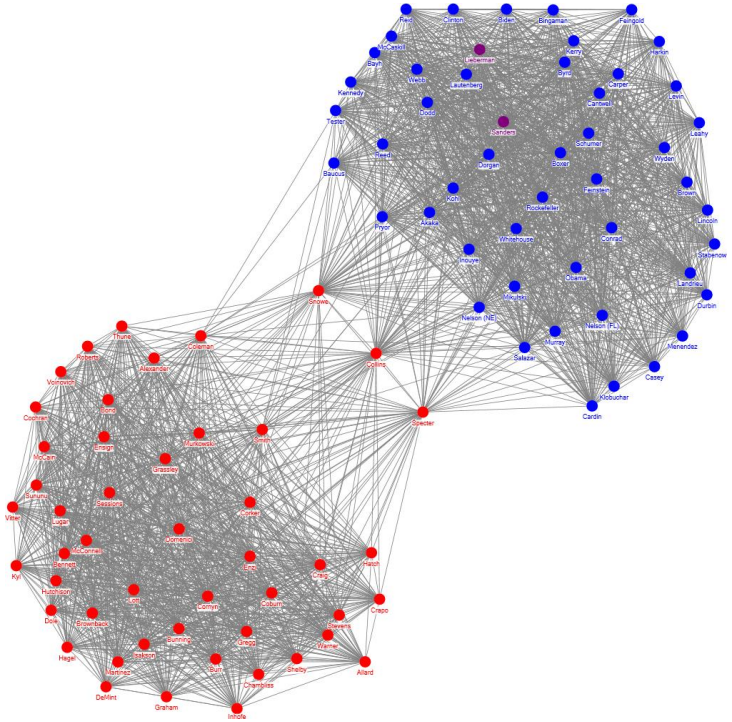




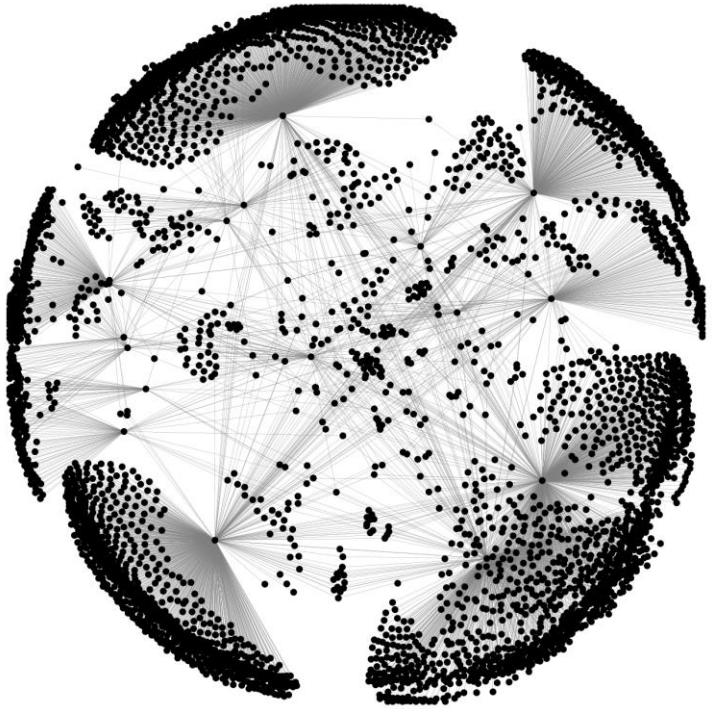
Wiki

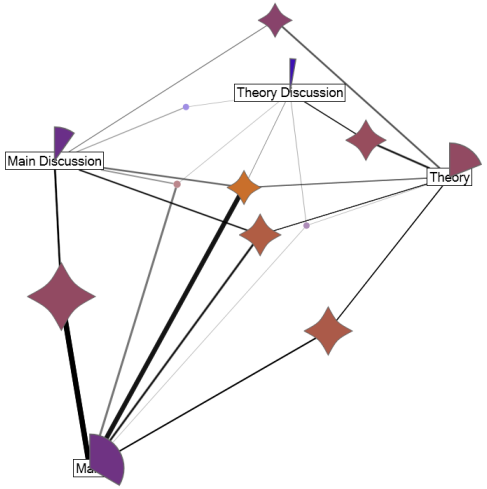


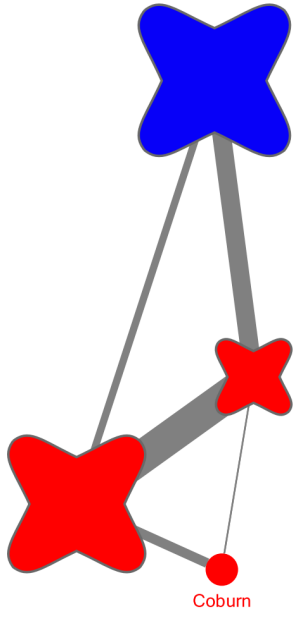
Senate

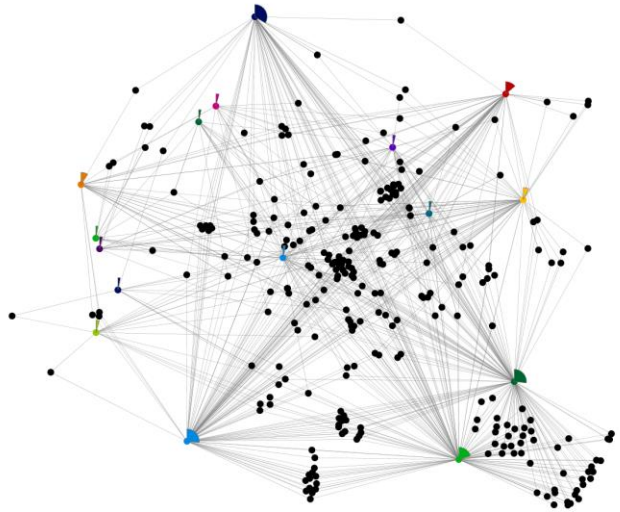


Web Crawl









Controlled Experiment

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

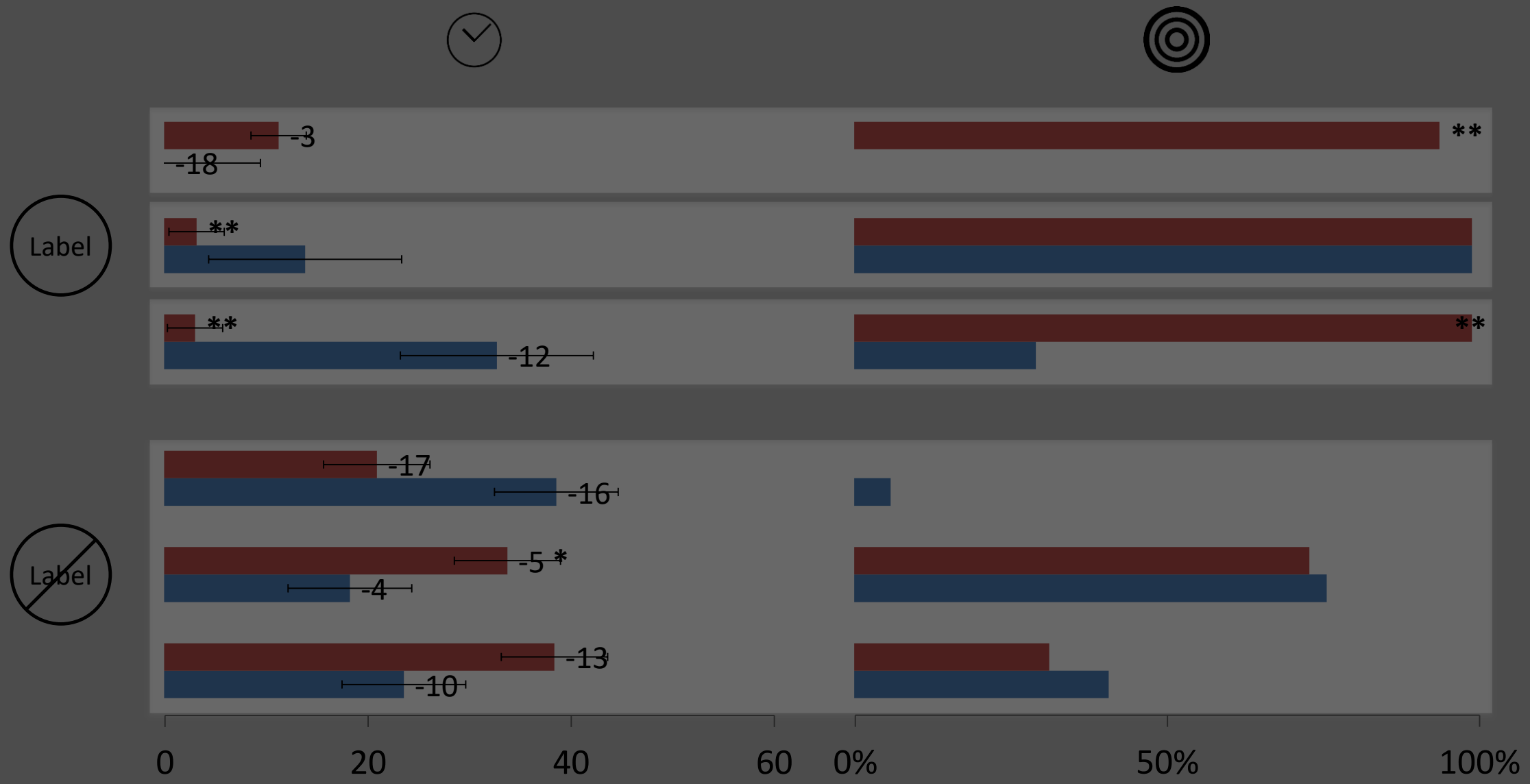
Controlled Experiment - Tasks

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

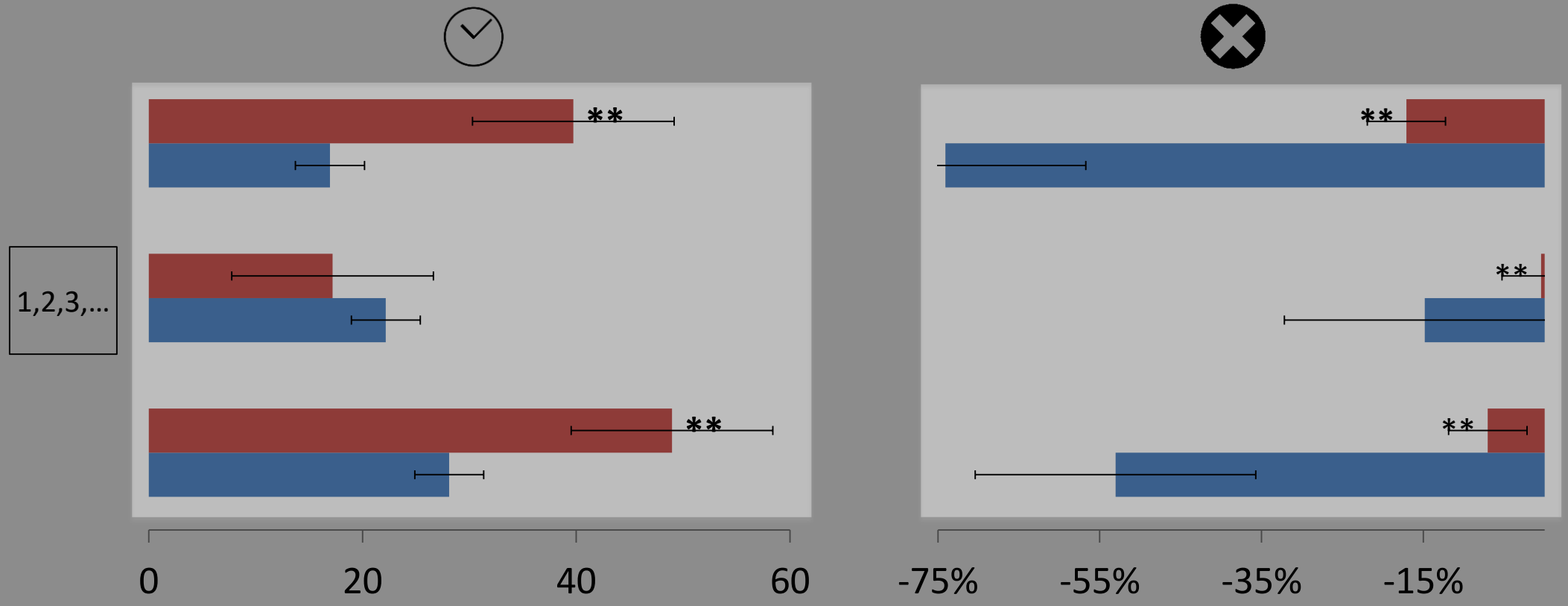


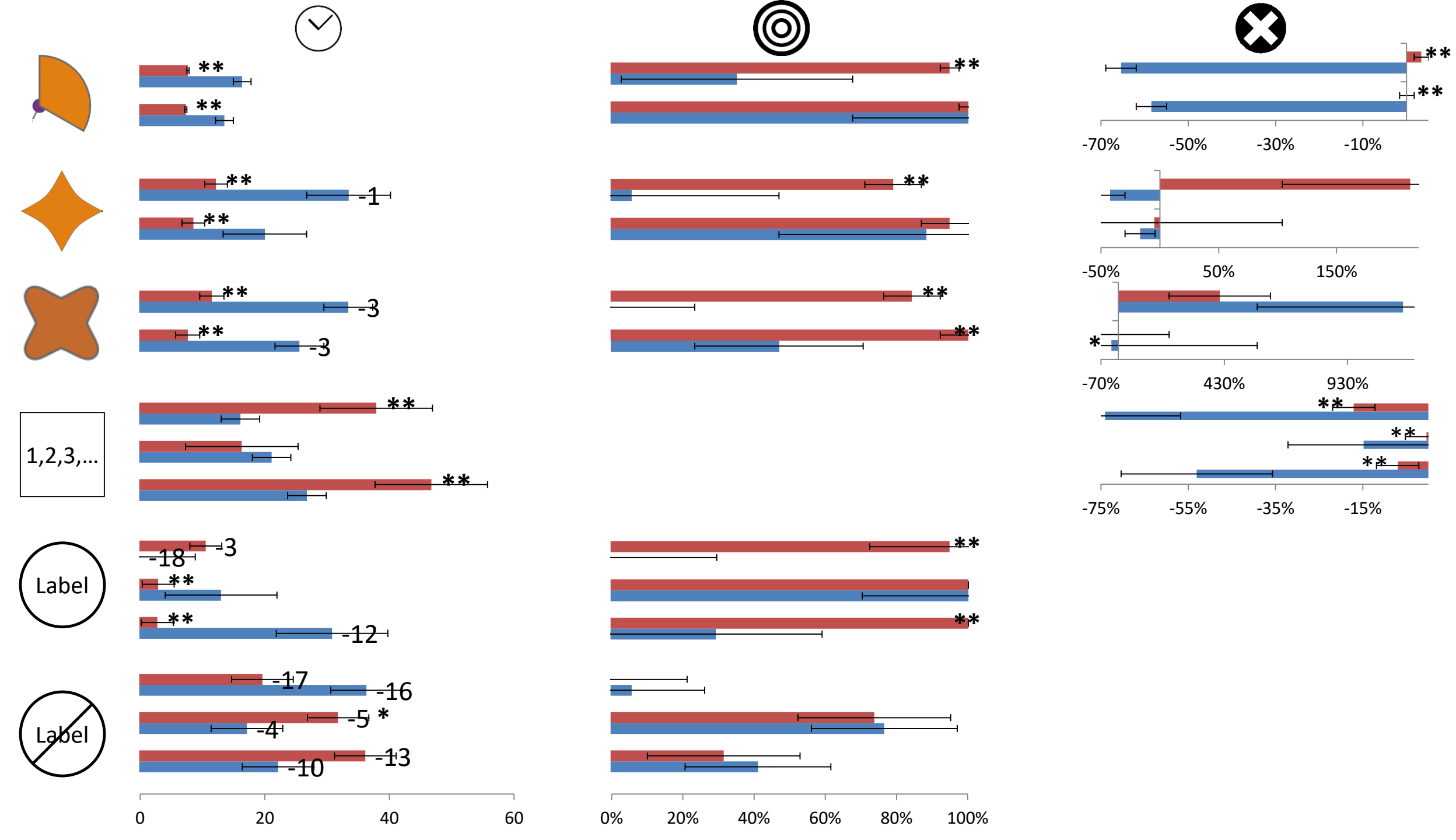
Estimating Node Count

Time in seconds & Error

Plain

Simplified





1,2,3,...
Label
Label
0
20
40
60
0%
20%
40%
60%
80%
100%
-70%
-50%
-30%
-10%
-50%
50%
150%
-70%
430%
930%
-75%
-55%
-35%
-15%

Motif Simplification

- **Algorithms** for detecting fans, connectors, and cliques
- Publicly available **implementation in NodeXL**: nodexl.codeplex.com
- **Case studies** in political science, sociology, urban planning, medical informatics, intelligence analysis…
- **Controlled experiment** with 36 users showed that motif simplification improves user task performance

Dunne C and Shneiderman B (2013), *"Motif simplification: improving network visualization readability with fan, connector, and clique glyphs"*, In CHI `13.

Shneiderman B and Dunne C (2012), *"Interactive network exploration to derive insights: Filtering, clustering, grouping, and simplification"*, In Graph Drawing `12.

Dunne C, Shneiderman B and Johnson T (2014), *"Understanding patterns in patient discharge summaries using network analysis"*. University of Maryland. Human-Computer Interaction Lab Tech Report No. (HCIL-2014-06).