Clusters of Nodes in Networks

Clusters

Grouping nodes in networks

Why?

- Groups are a basic theoretical component of social structure.
- Cohesion, unity, identity, …
- Divisions, conflict, hierarchy, …

How?

- Generally: clusters are groups of nodes that tend to connect more to each other than to others

Different formalizations of the basic idea:

- *Embedded cliques*
- Overlapping/hierarchical groups
- Partition of entire network

Clusters

Grouping nodes in networks

Why?

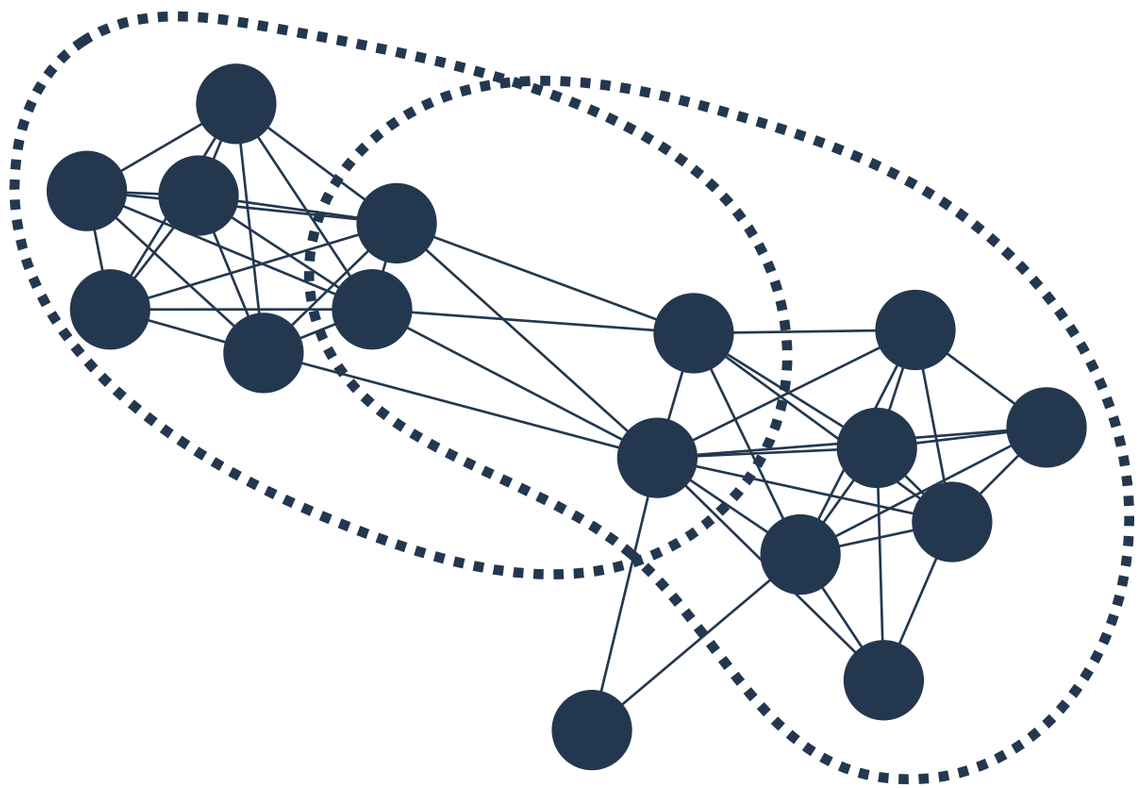
- Groups are a basic theoretical component of social structure.
- Cohesion, unity, identity, …
- Divisions, conflict, hierarchy, …

How?

- Generally: clusters are groups of nodes that tend to connect more to each other than to others

Different formalizations of the basic idea:

- Embedded cliques
- *Overlapping/hierarchical groups*
- Partition of entire network



Clusters

Grouping nodes in networks

Why?

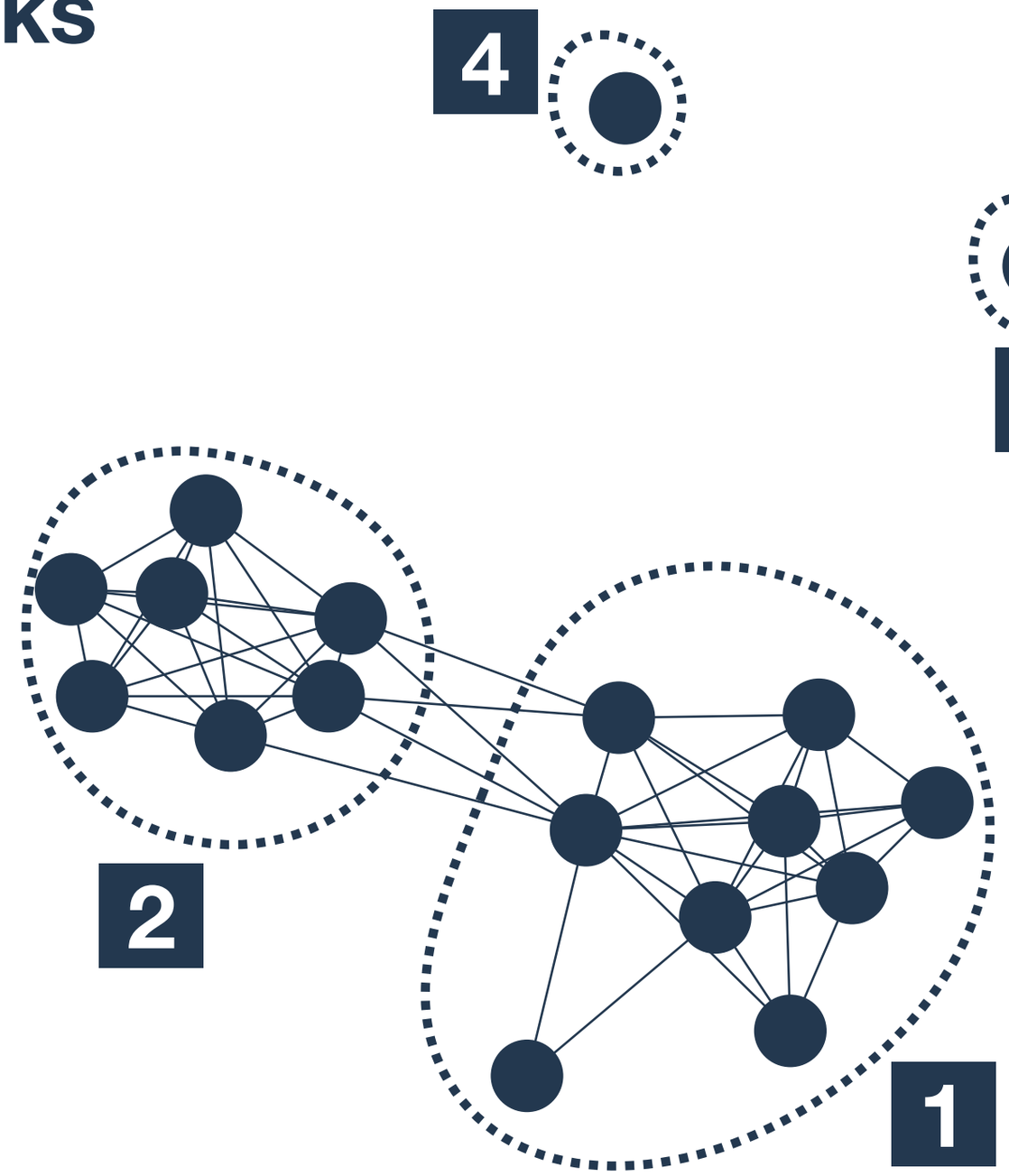
- Groups are a basic theoretical component of social structure.
- Cohesion, unity, identity, …
- Divisions, conflict, hierarchy, …

How?

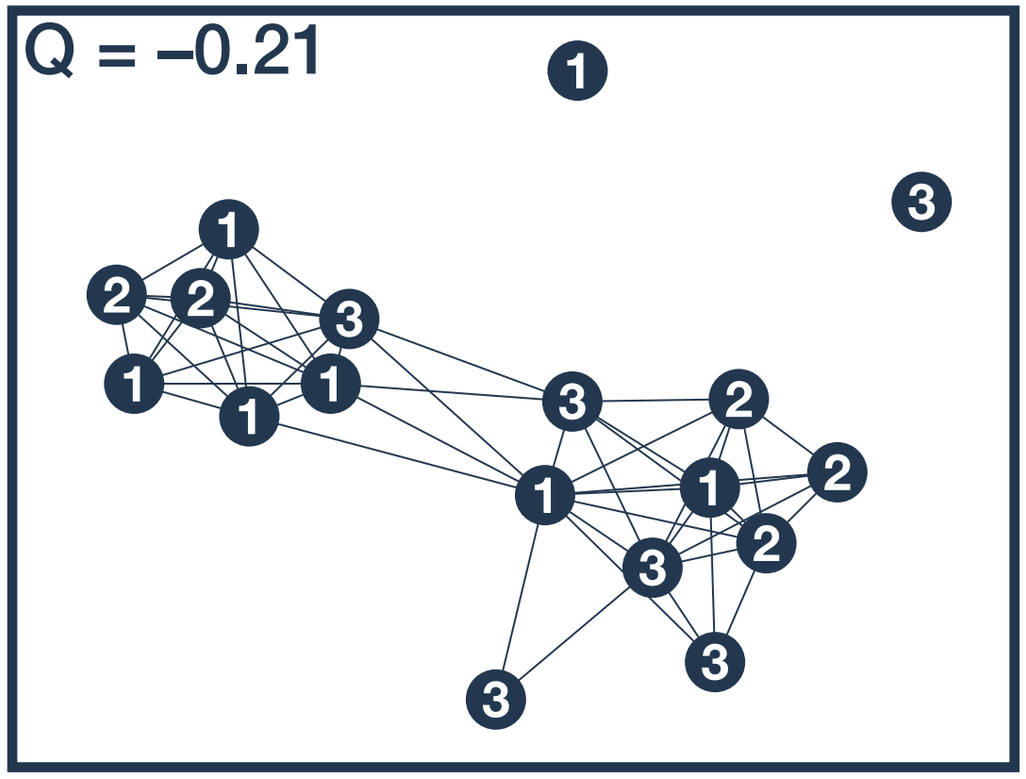
- Generally: clusters are groups of nodes that tend to connect more to each other than to others

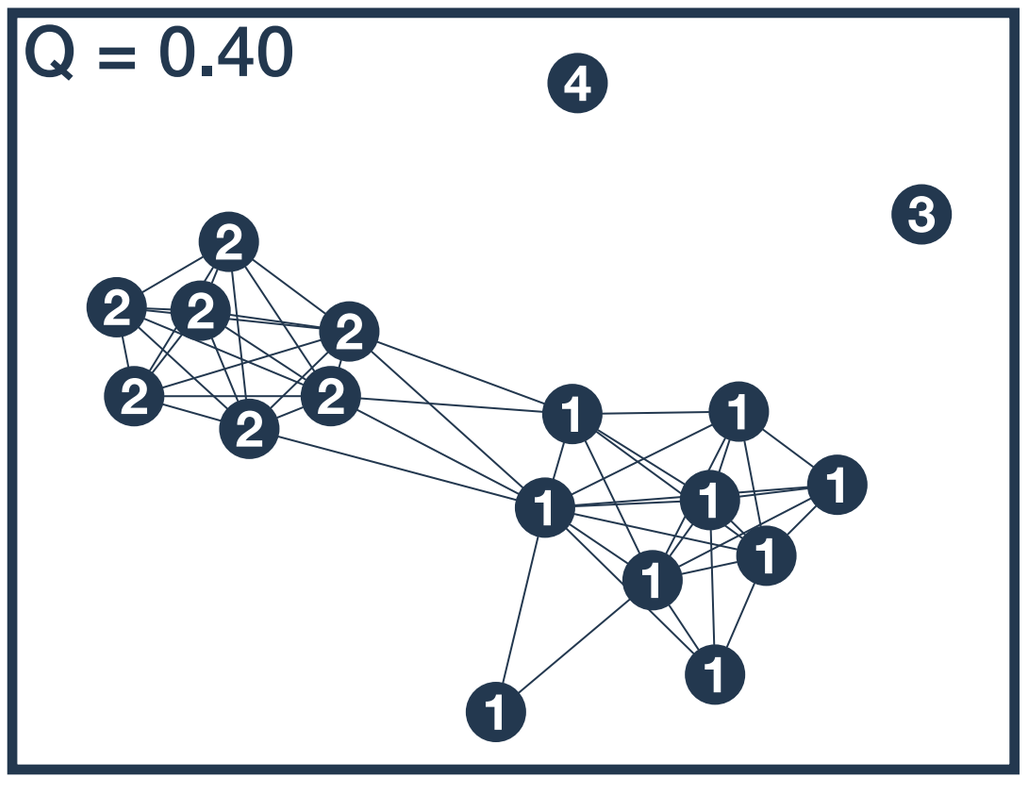
Different formalizations of the basic idea:

- Embedded cliques
- Overlapping/hierarchical groups
- ***Partition of entire network***



Modularity





Partitioning a network

- A *partitioning* of a network is any labeling scheme that assigns every node exactly one label.

Modularity (Q) is one measure of 'goodness' of a partitioning

- For any specific partitioning of a network, the *modularity* of that partitioning measures the degree to which edges tend to stay within a partition.
- For a partitioning with high modularity, edges will tend to connect nodes with the same label.
- Ranges from –0.5 (very bad fit) to 1.0 (very good fit)

Modularity maximization

- Clustering strategy that finds the partitioning that has the highest possible modularity

Clustering algorithms

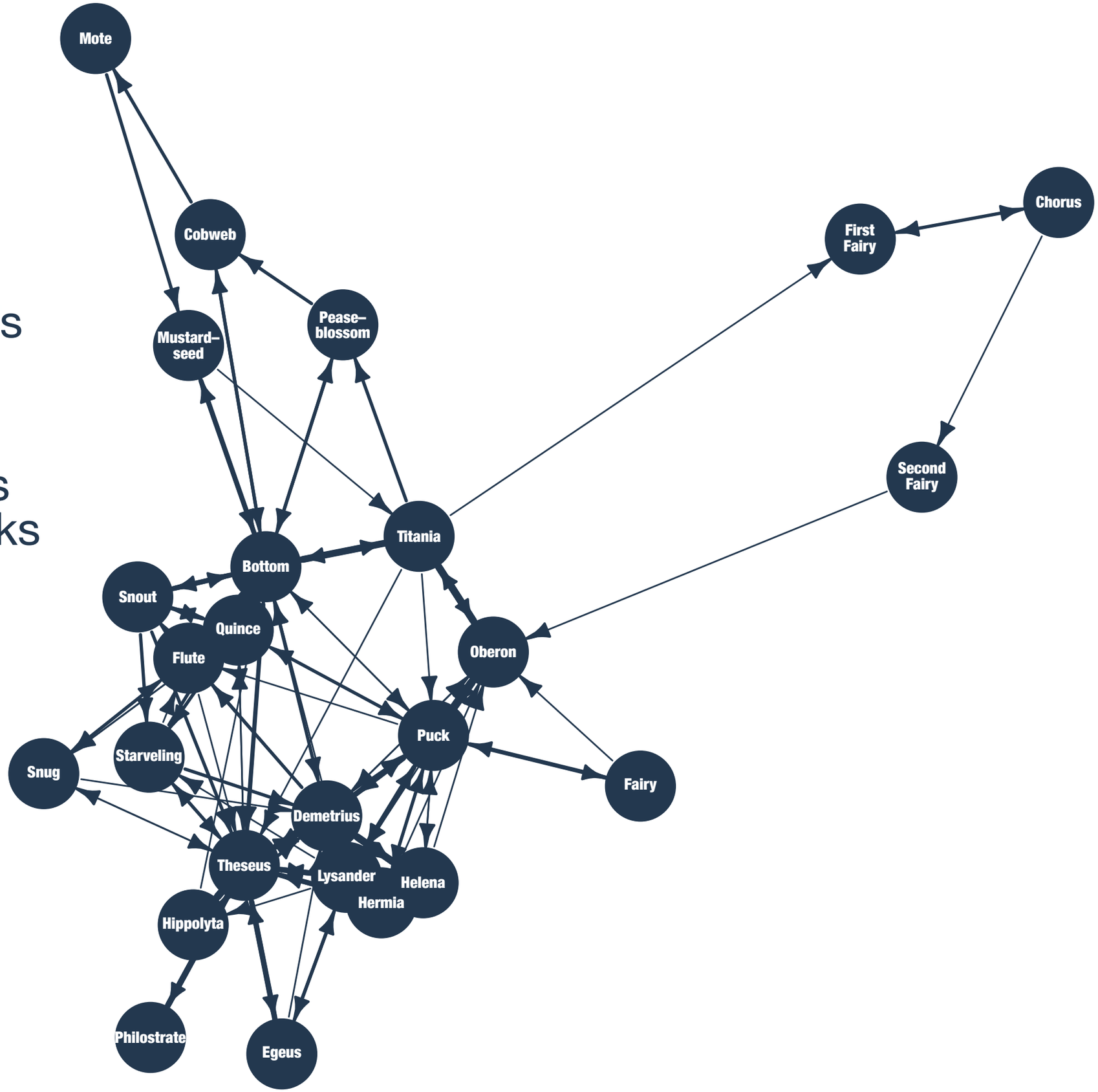
A Midsummer Night's Dream

Clustering algorithms

A Midsummer Night's Dream

Character network

- Directed edges indicate number of times one character's line immediately preceded another's
- E.g. Cobweb speaks and then Mote speaks
- Rough proxy for interaction



Clustering algorithms

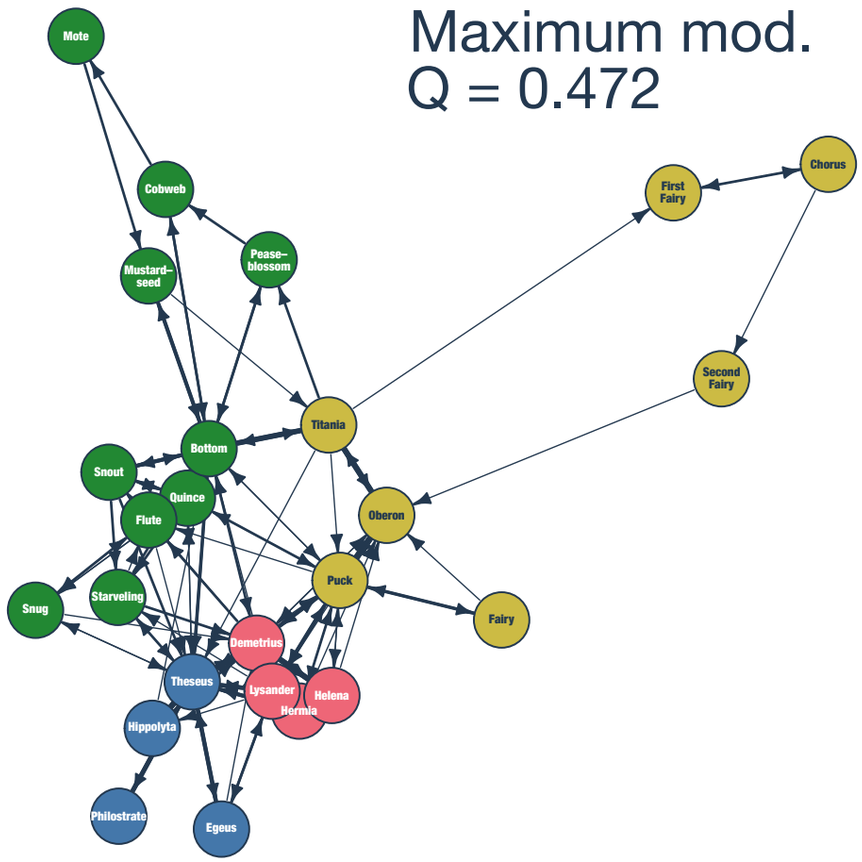
A Midsummer Night's Dream

Maximum modularity clusters

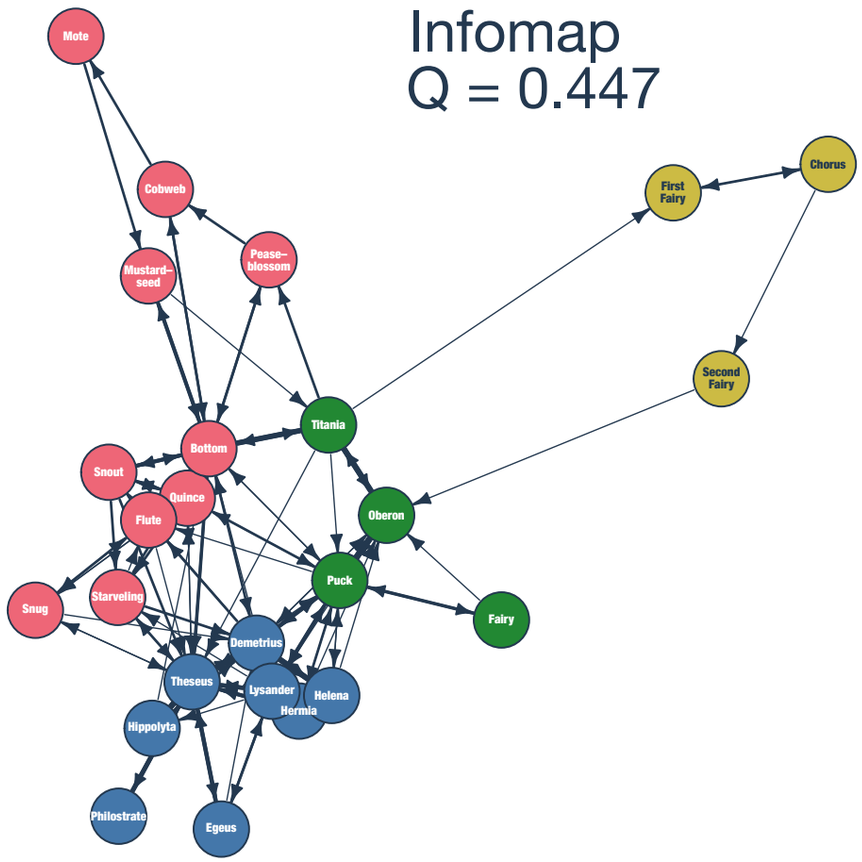
- Q = 0.472

Clustering algorithms

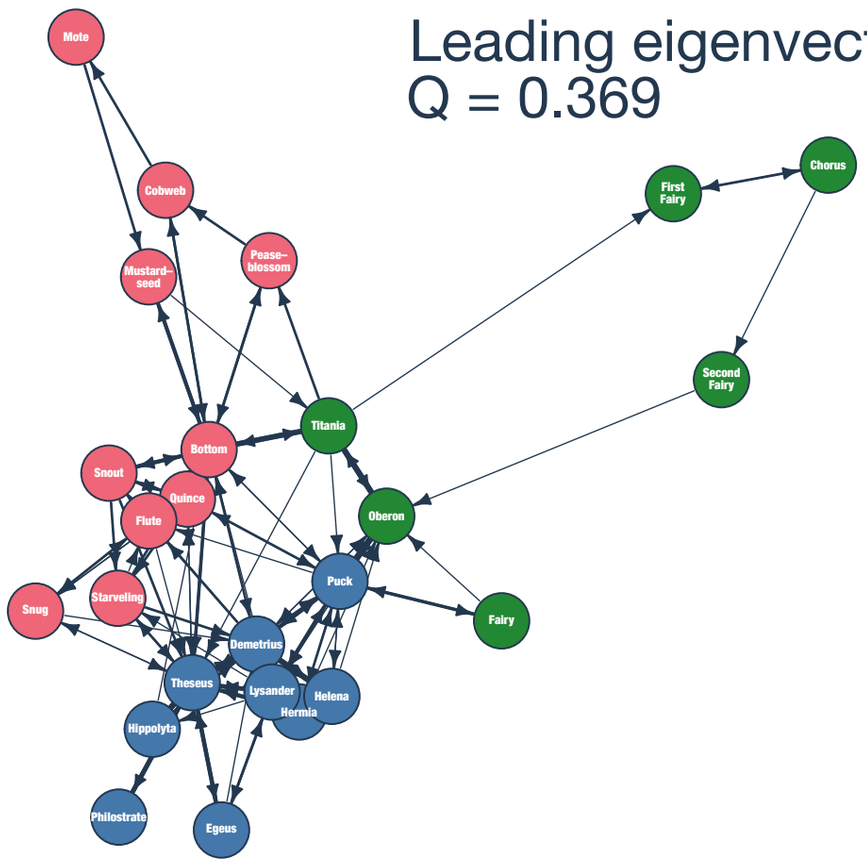
Maximum mod.
Q = 0.472



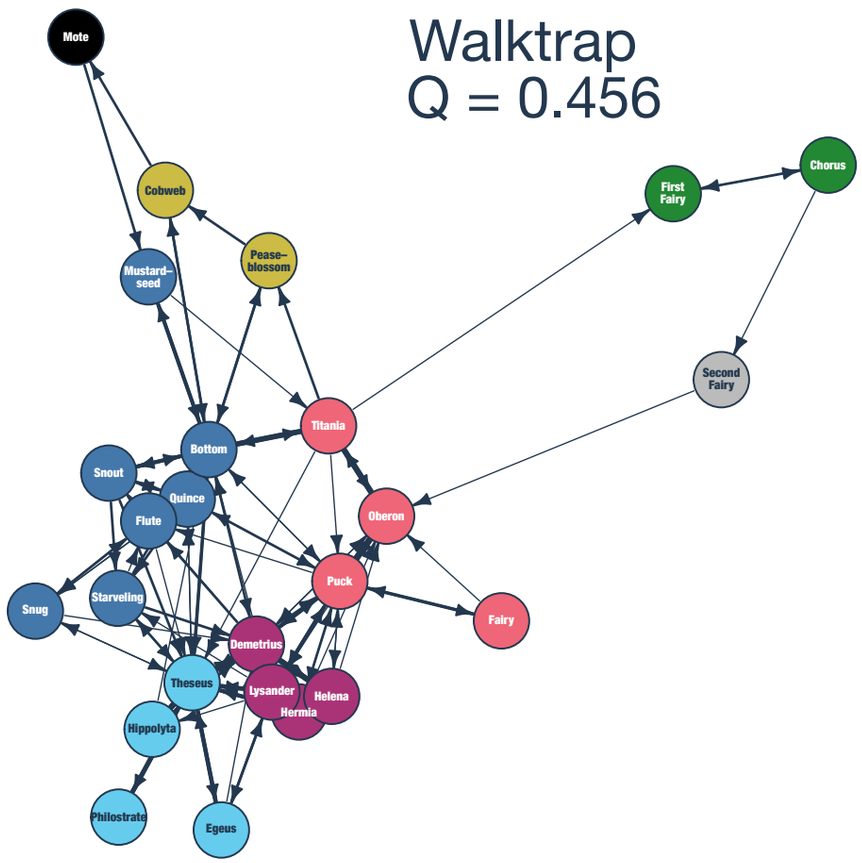
Infomap
Q = 0.447



Leading eigenvector
Q = 0.369

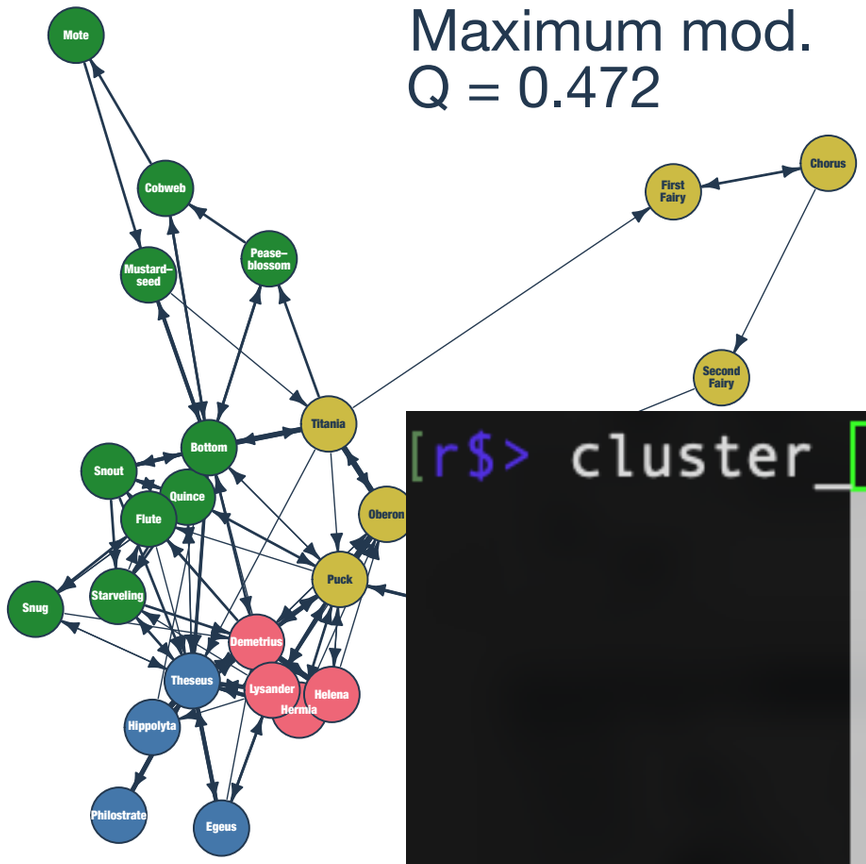


Walktrap
Q = 0.456



Clustering algorithms

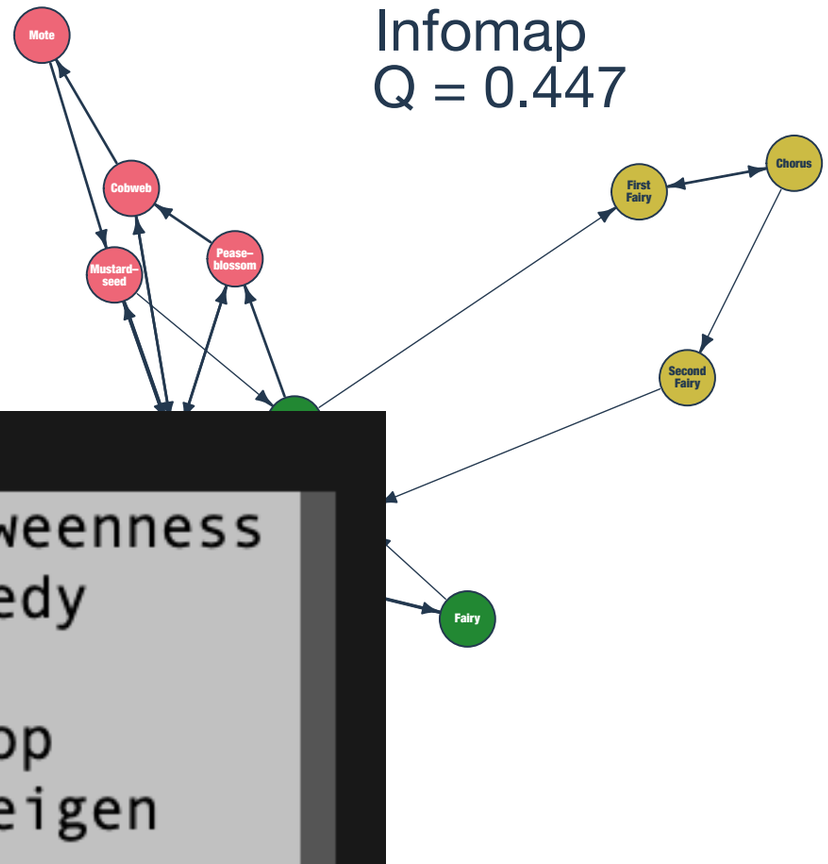
Maximum mod.
Q = 0.472



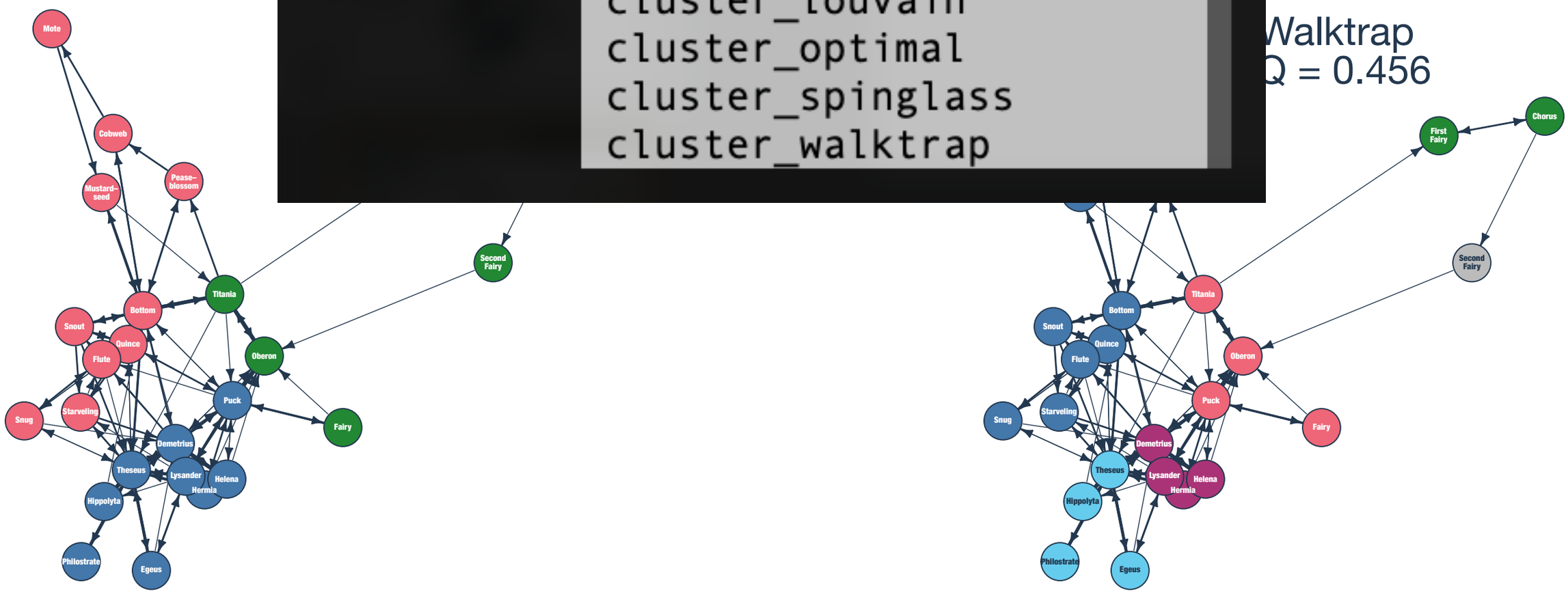

```
[r$> cluster_
```

```
cluster_edge_betweenness
cluster_fast_greedy
cluster_infomap
cluster_label_prop
cluster_leading_eigen
cluster_louvain
cluster_optimal
cluster_spinglass
cluster_walktrap
```

Infomap
Q = 0.447



Walktrap
Q = 0.456



Scientific Consensus

The Temporal Structure of Scientific Consensus Formation

Shwed and Bearman (2010)

Crash course on the sociology of science

S&B:

- Scientific consensus is contingent on broader societal discourse
- Therefore there is not a uniform progression toward consensus
- S&B investigate this by using *citation networks* to measure consensus over time

Citation networks

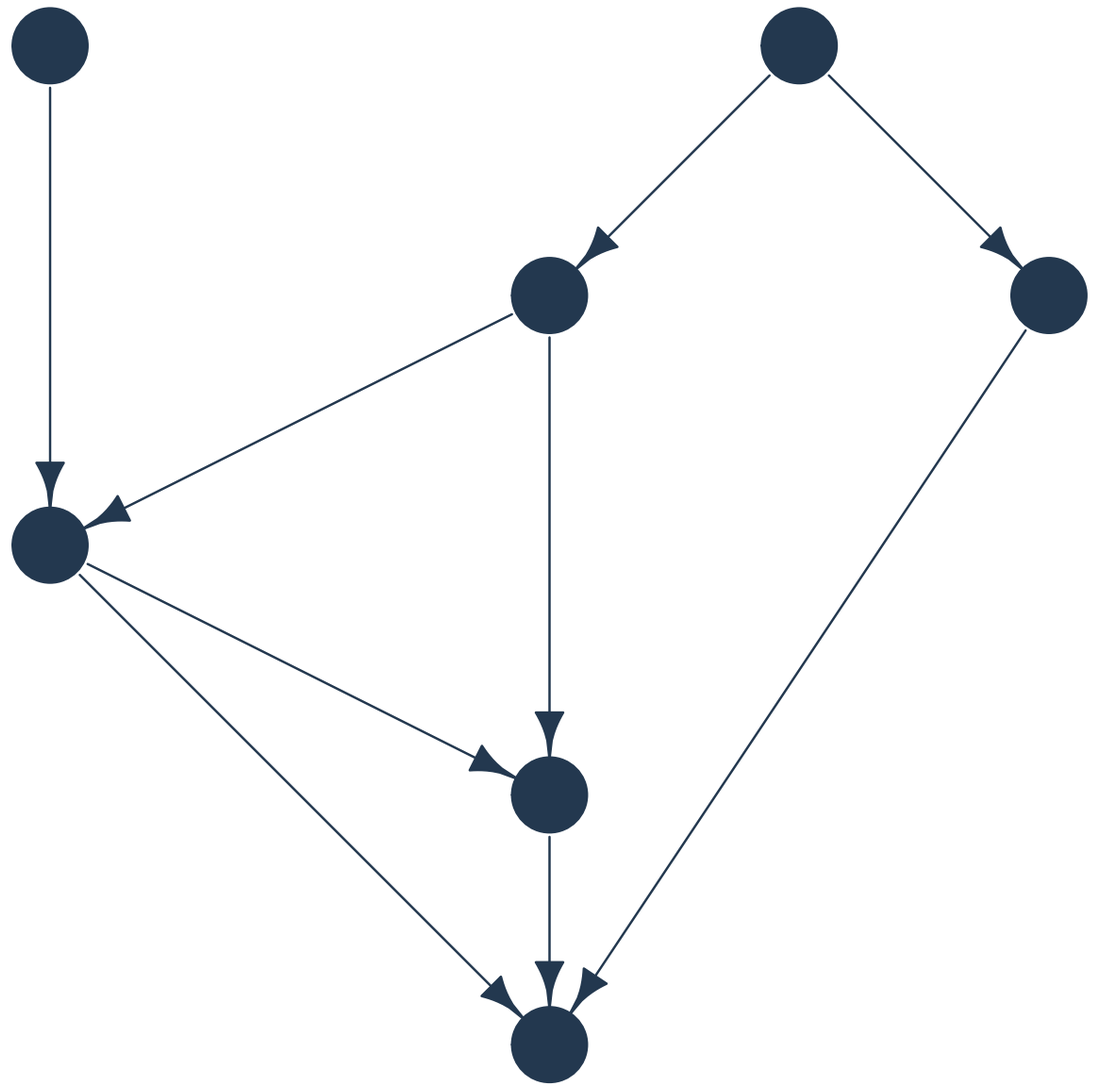
Measuring relations between scholarly publications

Citation network

- Vertices are publications (articles, books, conference papers, etc.)
- Directed edges represent citation
- Temporality imposes structure

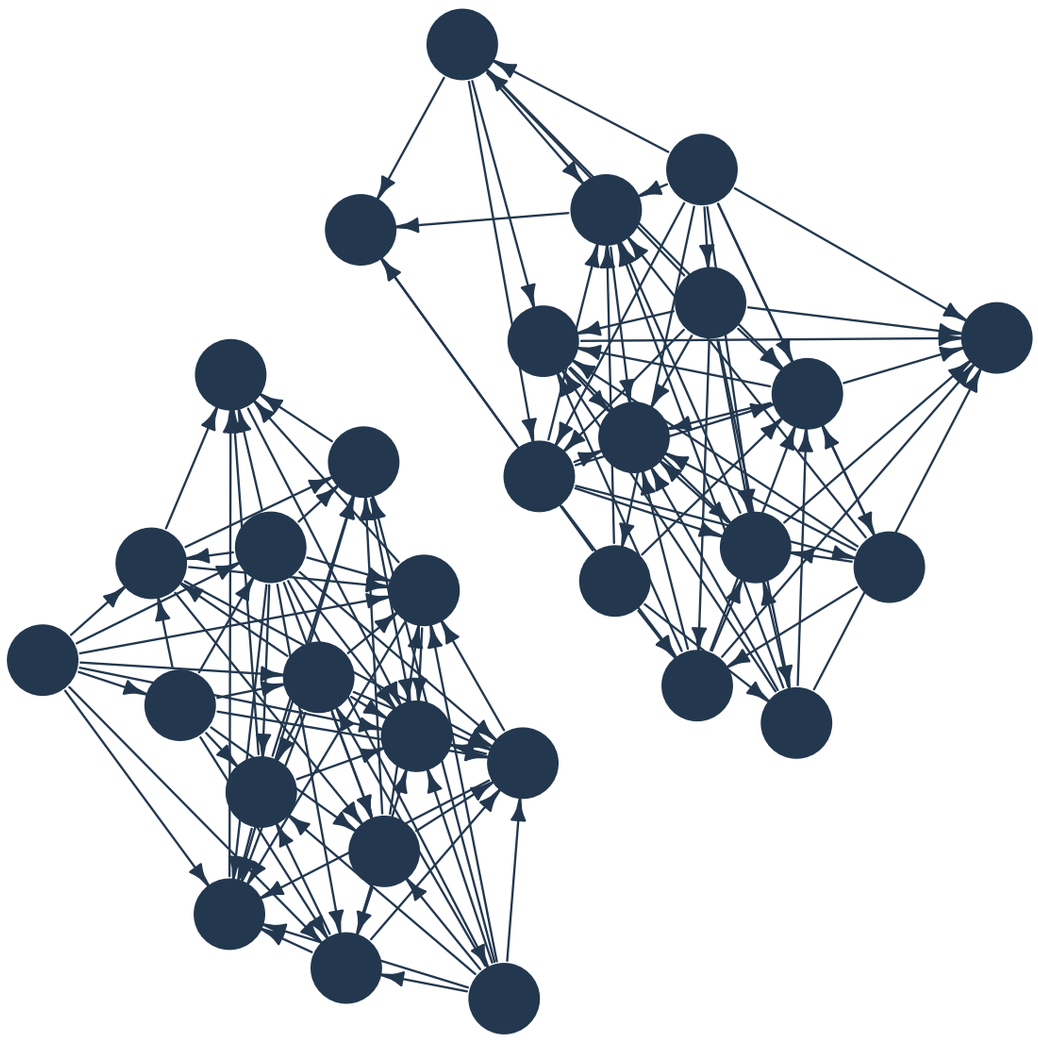
Citations as relations

- Scientific knowledge is not purely cumulative
- Citation indicates similarity of theories, methods, assumptions, etc.

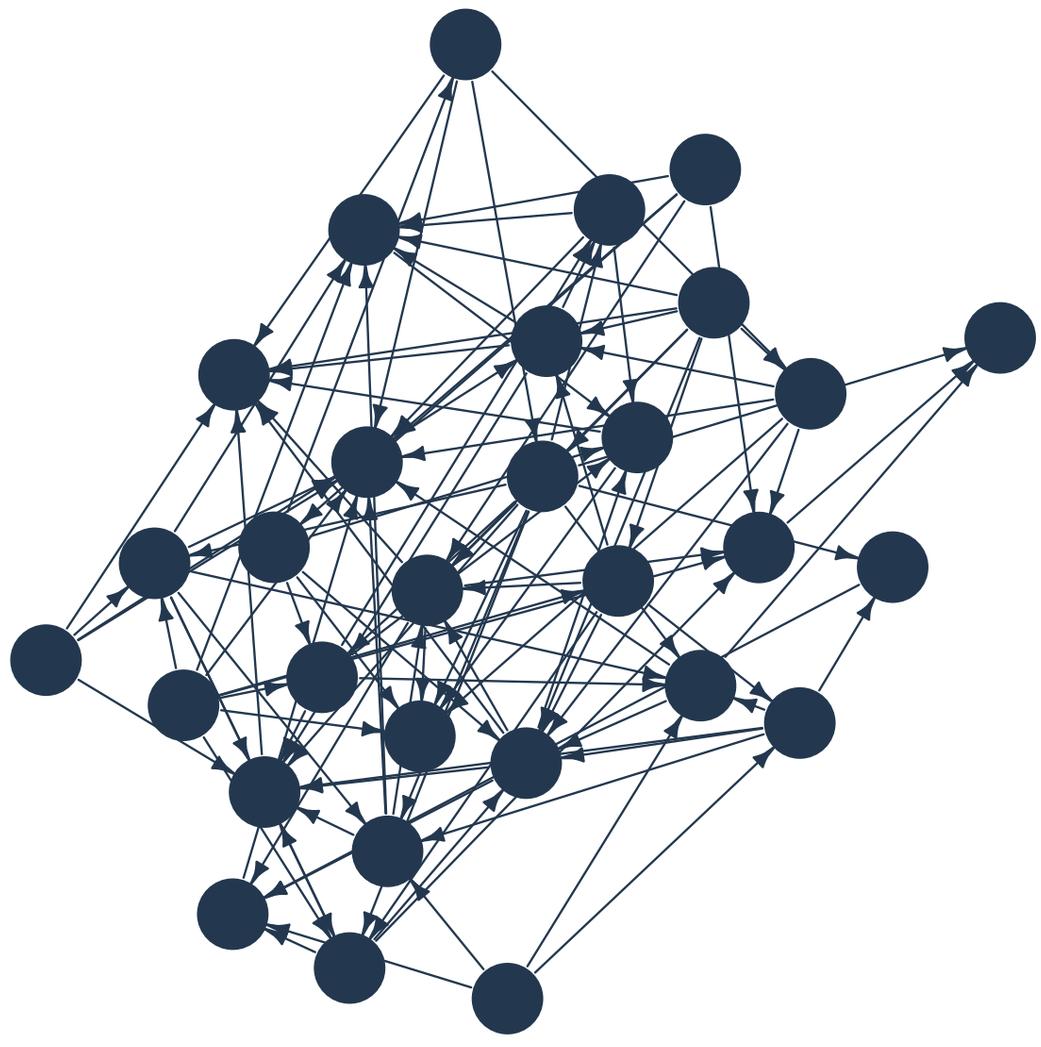


Measuring consensus

Two hypothetical citation networks



Q = 0.5
(epistemic rivalry)



Q = 0.05
(epistemic consensus)