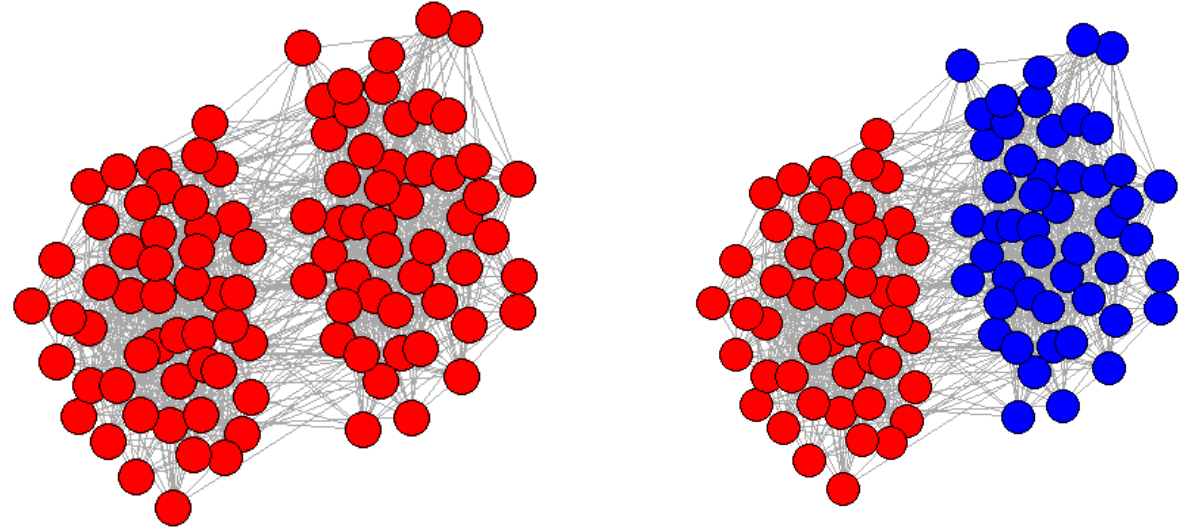# COMMUNITY DETECTION

# Community detection

- What is a community?

- Types of communities?
  - Political networks?
  - Facebook?
  - Twitter?

- How do we find them?
  - GN – not covered
  - Hierarchical clustering
  - Modularity
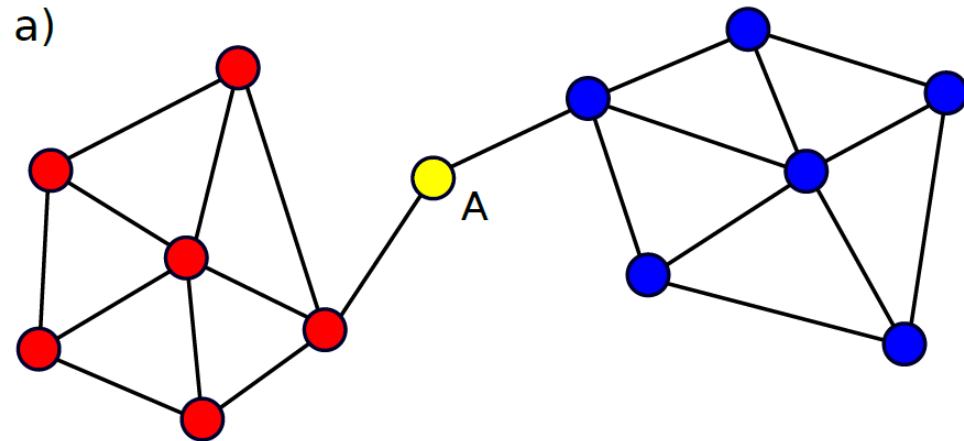  - Flow



Community detection in graphs

Santo Fortunato *

Community detection in networks: A user guide

Santo Fortunato [a,b,*], Darko Hric [b]

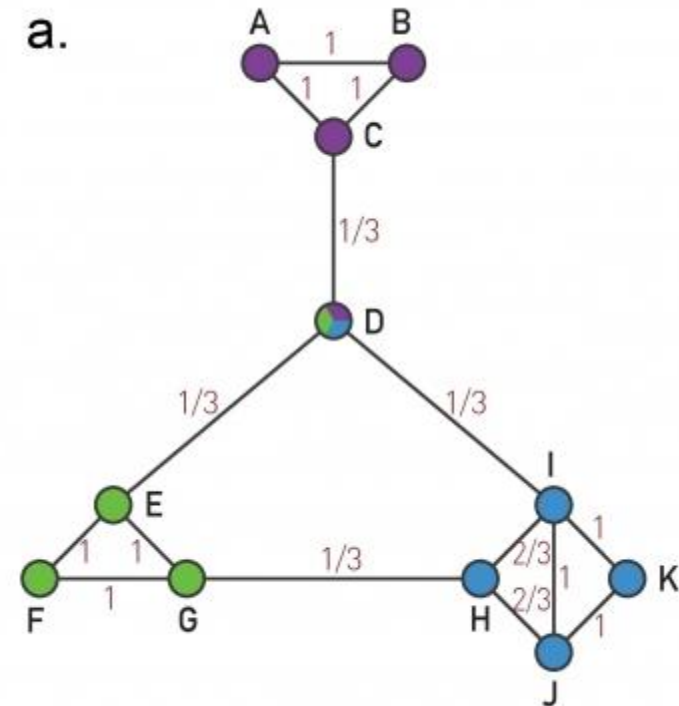Community detection in networks: Structural communities versus ground truth

Darko Hric, Richard K. Darst, and Santo Fortunato

# Girvan-Newman

- Based on shortest path length
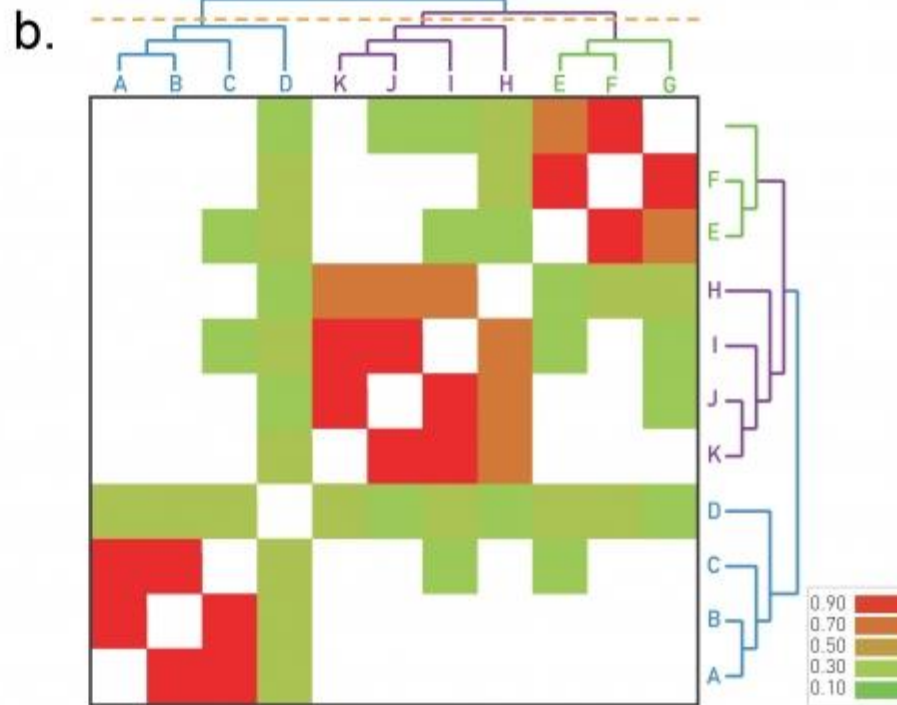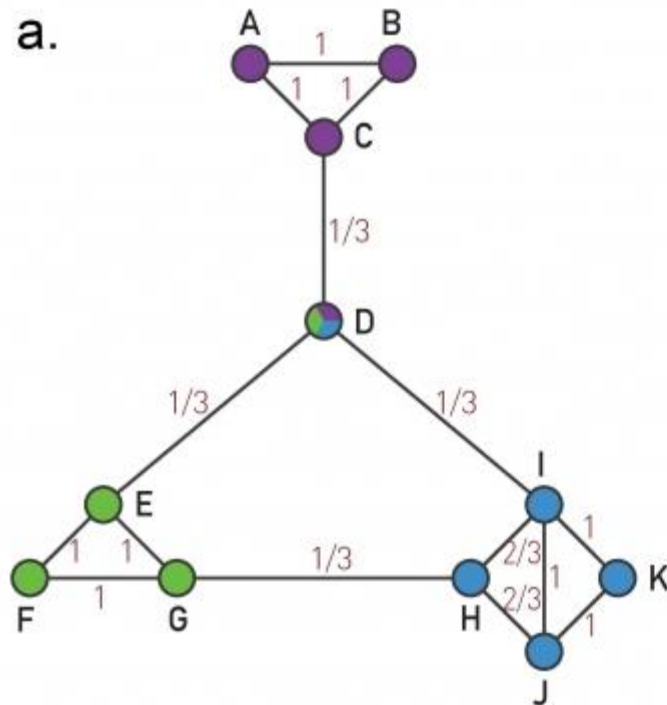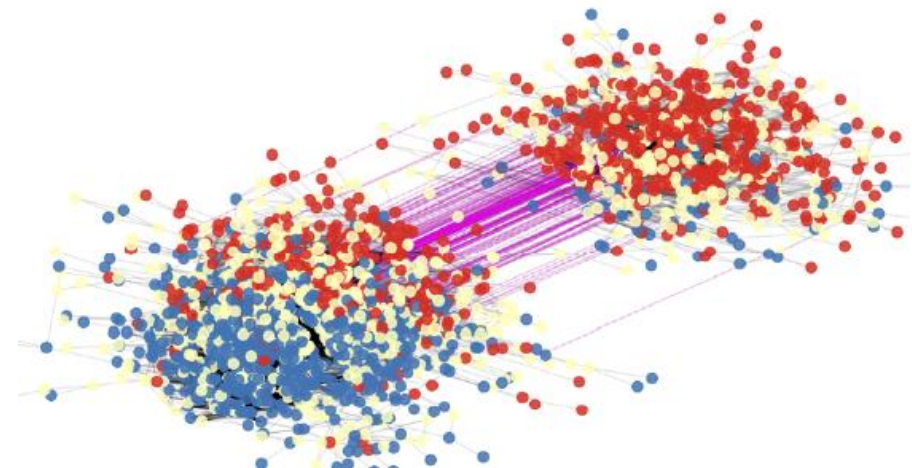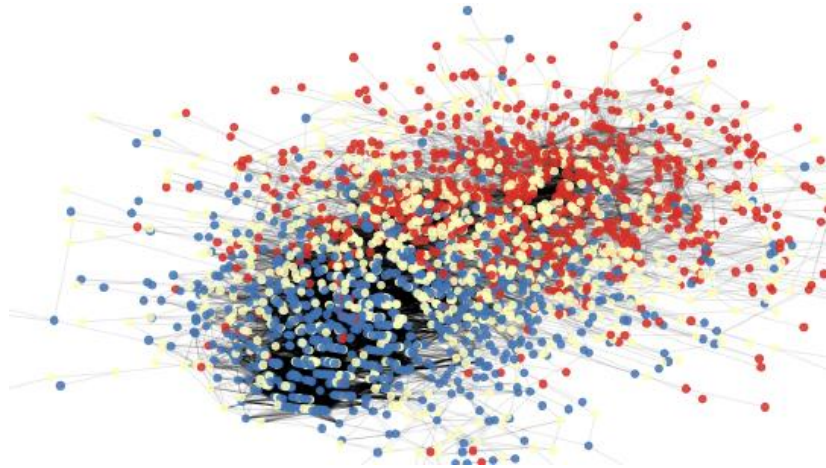


a)

# Hierarchical clustering

- Similarity based
  - Shared average cluster similarity
  - Aggregated using a linkage function

# Hierarchical clustering

- Similarity based
  - Shared average cluster similarity



DETECTING OPINION-BASED GROUPS AND POLARISATION
IN SURVEY-BASED ATTITUDE NETWORKS AND ESTIMATING
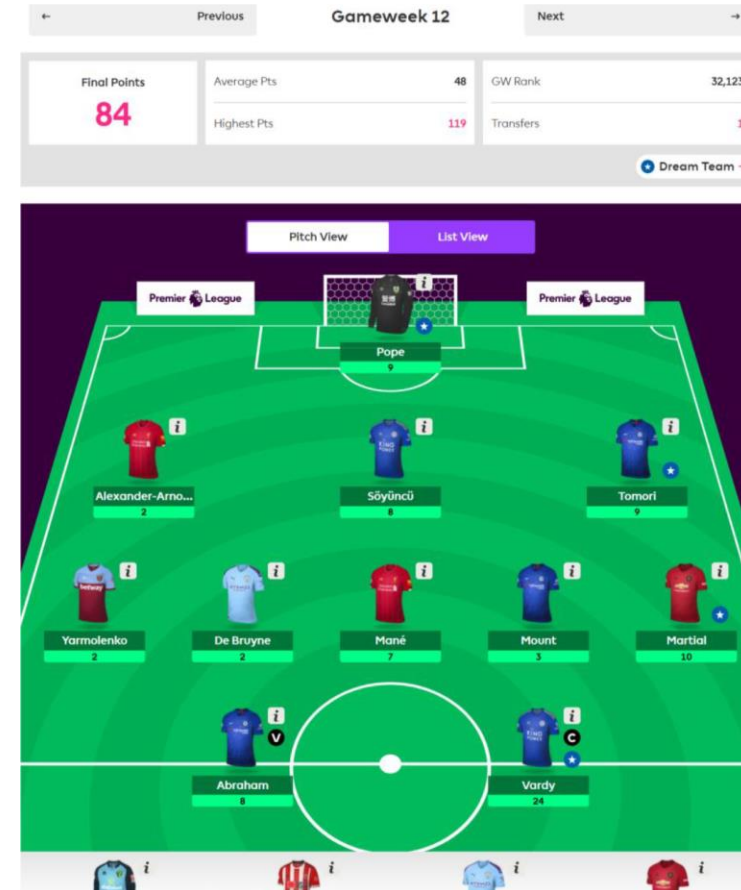QUESTION RELEVANCE

# Hierarchical clustering

- Current application
  - Fantasy sports

# Hierarchical clustering

- Slightly different network…
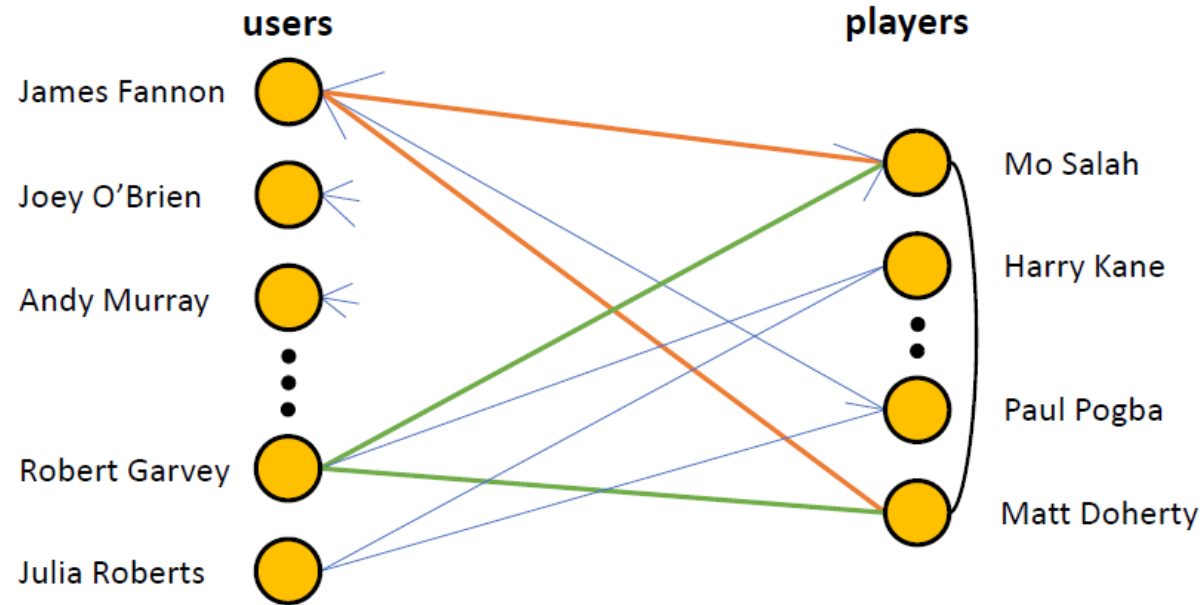
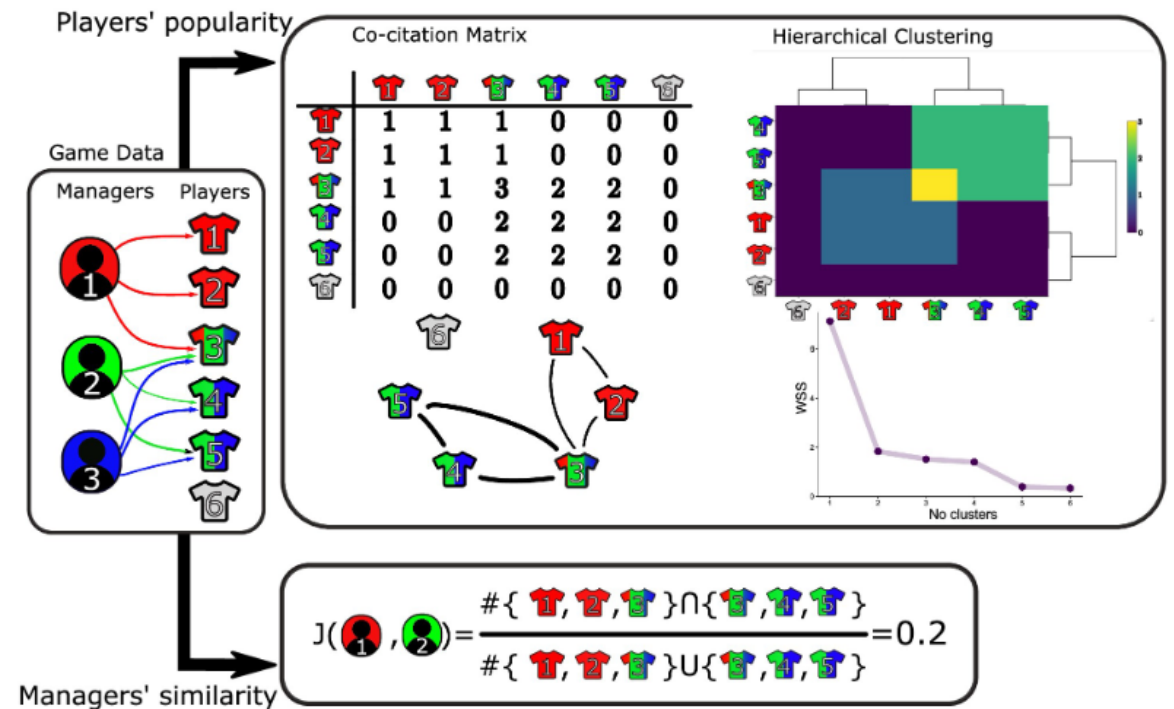# Hierarchical clustering

- One type of node
- Weighted
- Temporal

RESEARCH ARTICLE

Identification of skill in an online game: The case of Fantasy Premier League

Joseph D. O'Brien *, James P. Gleeson, David J. P. O'Sullivan

# Hierarchical clustering
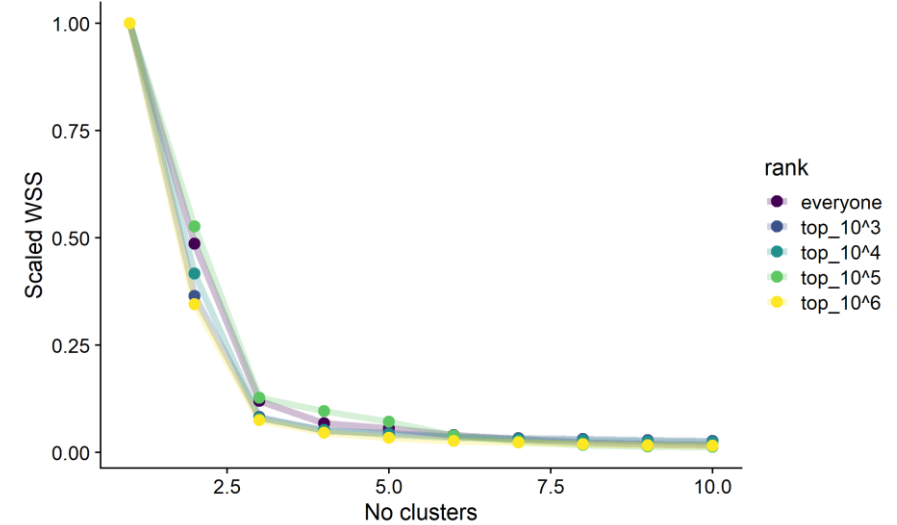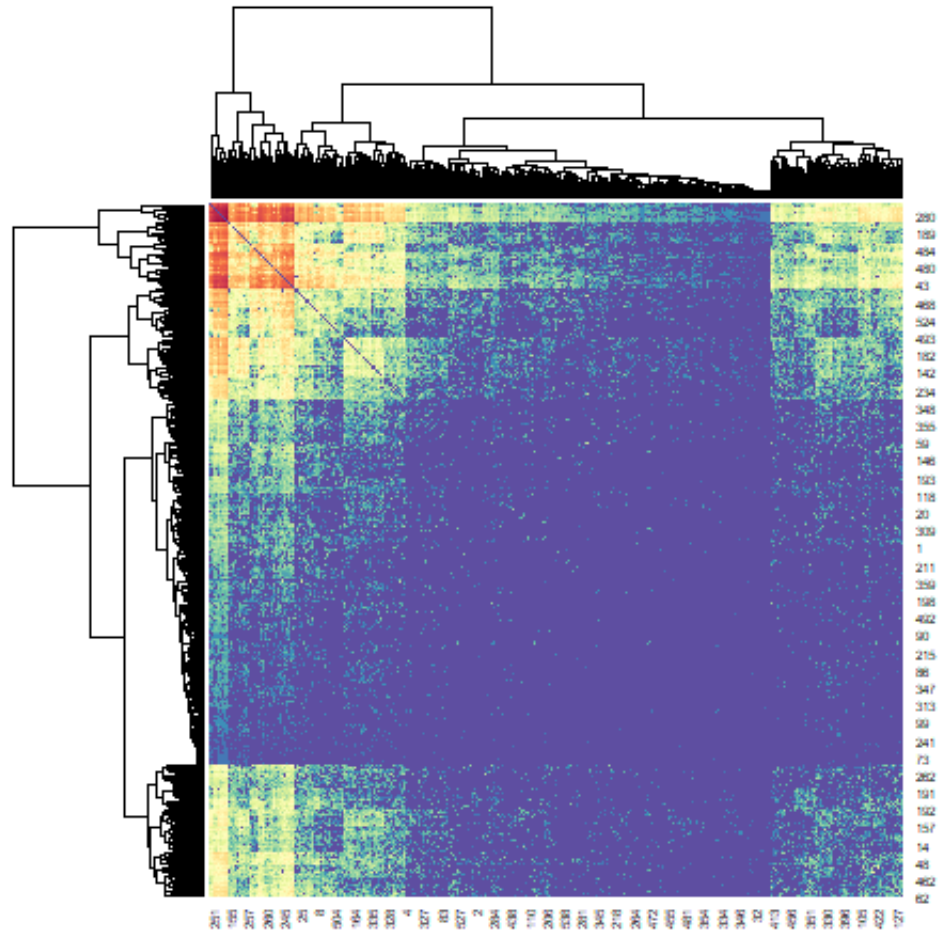
# Modularity, briefly

- 'more density connected internally than externally'

- Null model of a random graph to quantify quality of a community:

$$Q = \frac{1}{2m} \sum_{ij} (A_{ij} - \frac{k_i k_j}{2m}) \delta(c_i, c_j)$$

- Optimizing this is NP-hard!
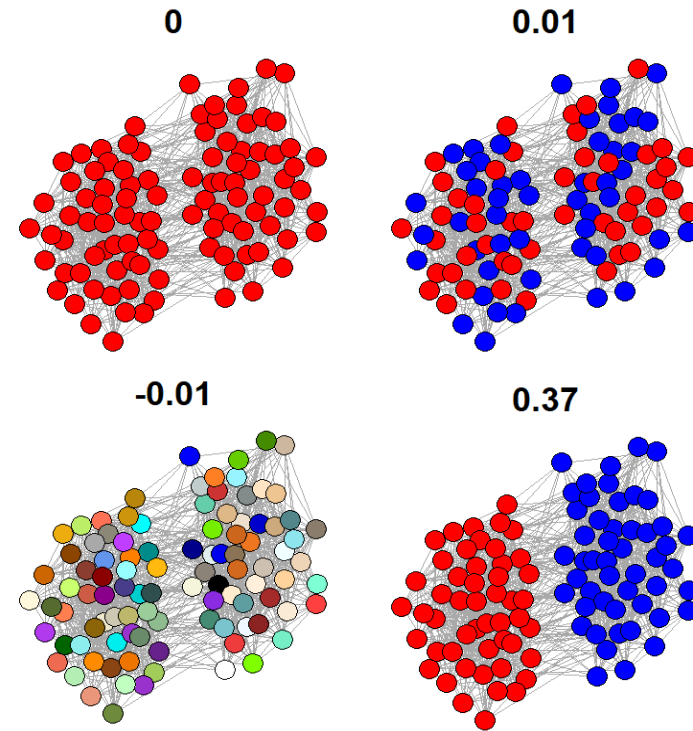
# Modularity, briefly

- Louvain Algorithm
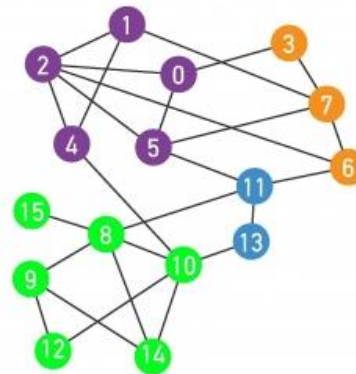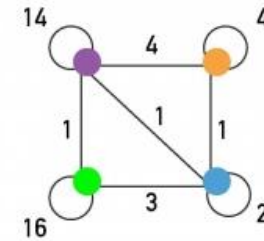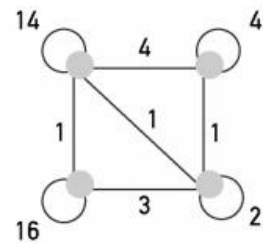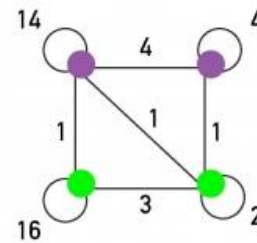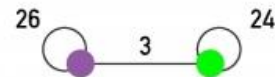
# Markov stability, again, briefly

- Flow based methods
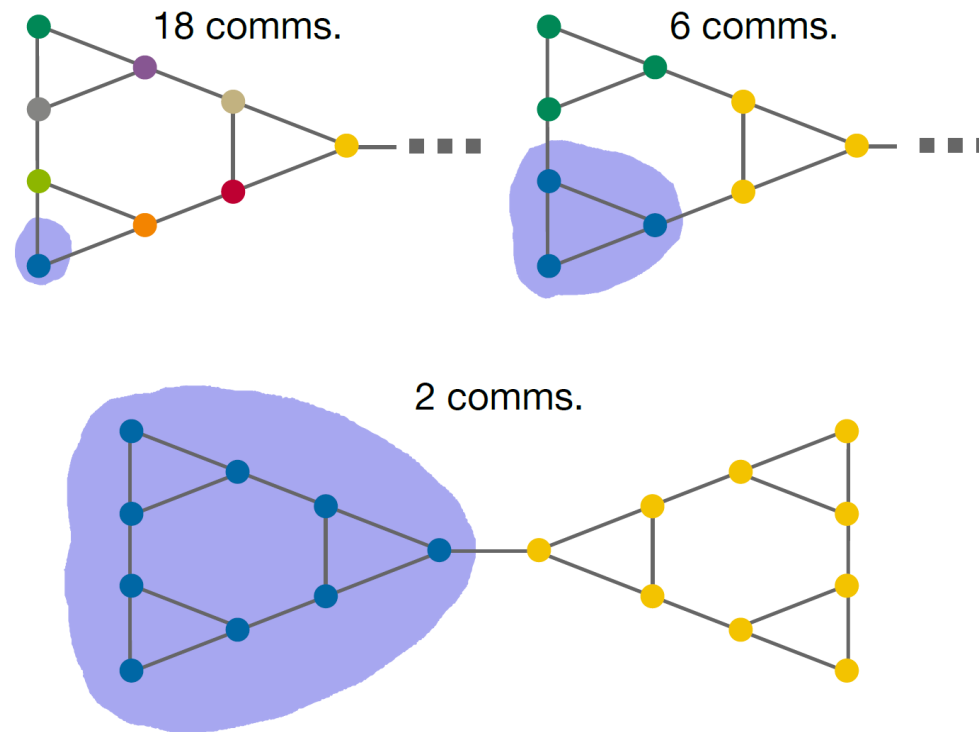
# Again, another little break with R

- So what did this analysis actually look like…

- 4_community_dectection.r

# Community Detection

- Markov stability on mentions and follower networks

- Mentions
  - Partition into 2 communities
- Follower
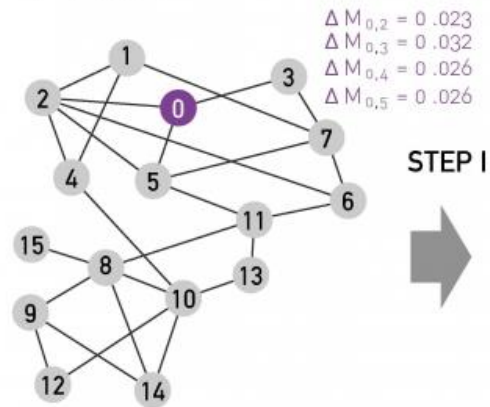  - Partition into 6 communities

- How could we combine information from both networks?

- Like a unique partition that reflects both previous partitions
  - Refine both then cluster back together

Mention Network Partition

Follower Network Partition

# Finding Yes & No Supporters



Mention Network Partition

Follower Network Partition

Intersection of partitions

Cluster partitions using homophily variables

Final Partition

# Community Detection

- Markov stability on mentions and follower networks

- Mentions
  - Partition into 2 communities
- Follower
  - Partition into 6 communities

- How could we combine information from both networks?

- Like a unique partition that reflects both previous partitions
  - Refine both then cluster back together

Mention Network Partition

Follower Network Partition

# Finding Yes & No Supporters

- Use sentiment and community detection



(a) Mention Network Partition
(b) Follower Network Partition
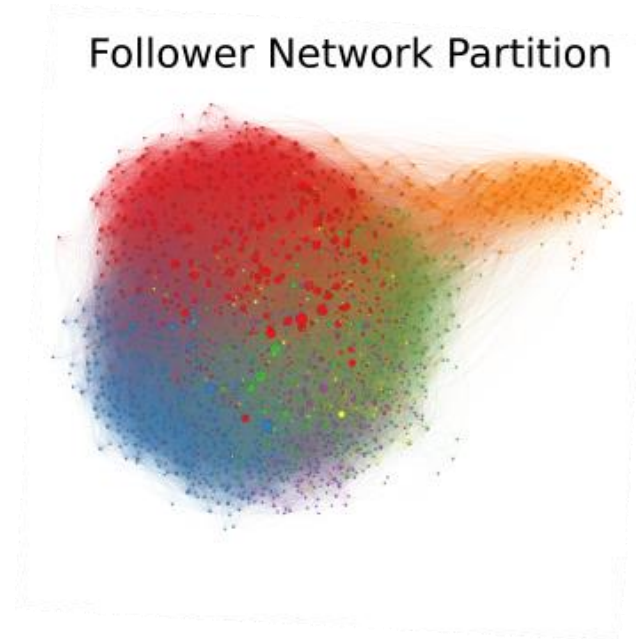(c) Intersection of partitions
(d) Average User Sentiment — Positive / Negative
(e) Chosen Number of Clusters; Within sum of squares vs Number of clusters
(f) Average user sentiment in vs Average user sentiment out
(g) Final Partition — $CC_1$, $CC_2$, $CC_3$
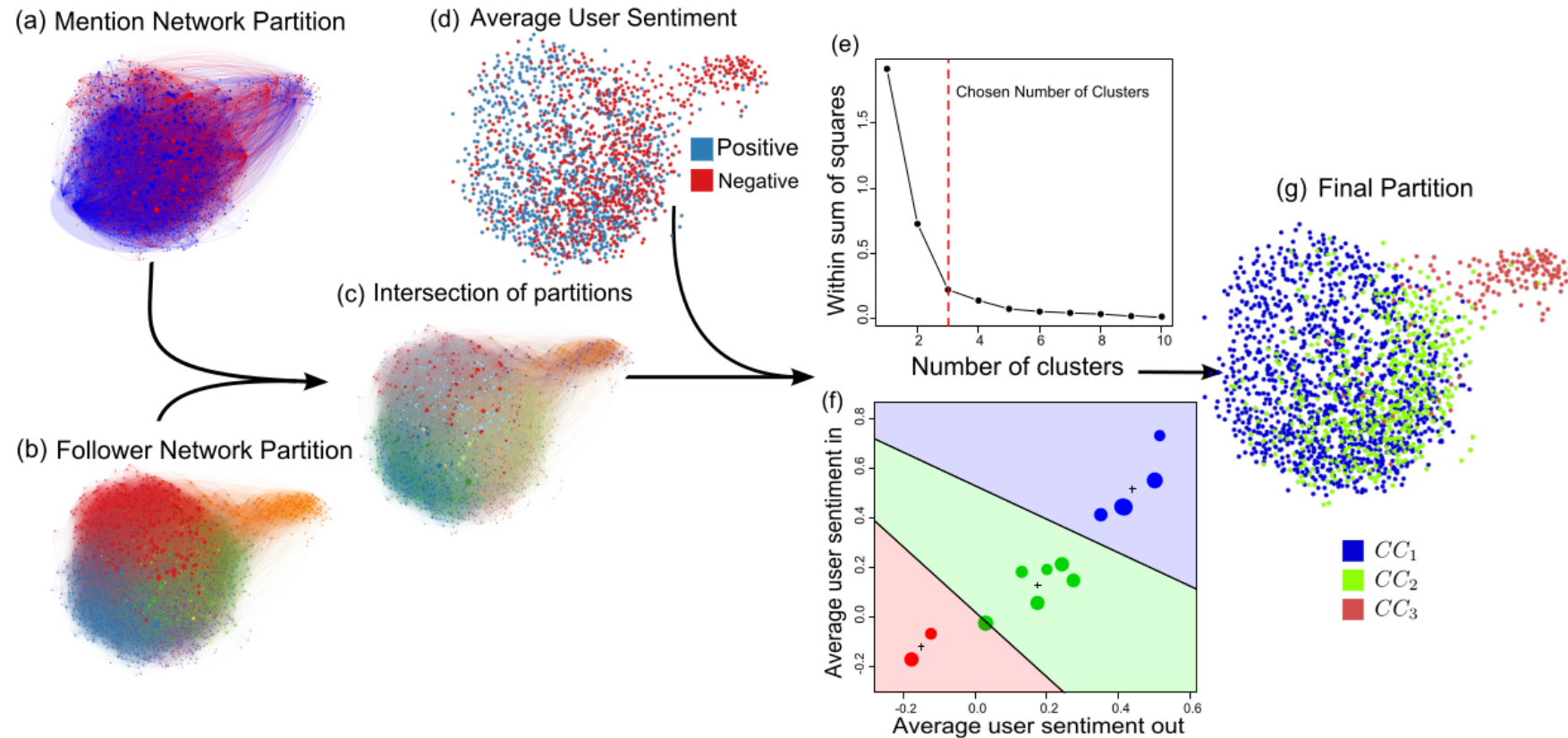
- What does makes up these three groups?

# Community Detection and Sentiment

- What about yes and no voters?
  - From each CC select random sample (20%) and classify
    - Hand classify as yes, no leaning or unaligned

- How do we classify
  - Look at the tweets the user has sent
  - Profile description

| | | Community cluster | | | |
|---|---|---|---|---|---|
| | | $CC_1$ | $CC_2$ | $CC_3$ | Total |
| | Yes | 183 | 114 | 6 | 303 |
| Alignment | No | 1 | 2 | 23 | 26 |
| | Unaligned | 21 | 5 | 3 | 29 |
| | Total | 205 | 121 | 32 | 358 |

# Again, another little break with R

- So what did this analysis actually look like…

- 5_community_aggregation.r

# HOW DID THE COMMUNITIES TALK?
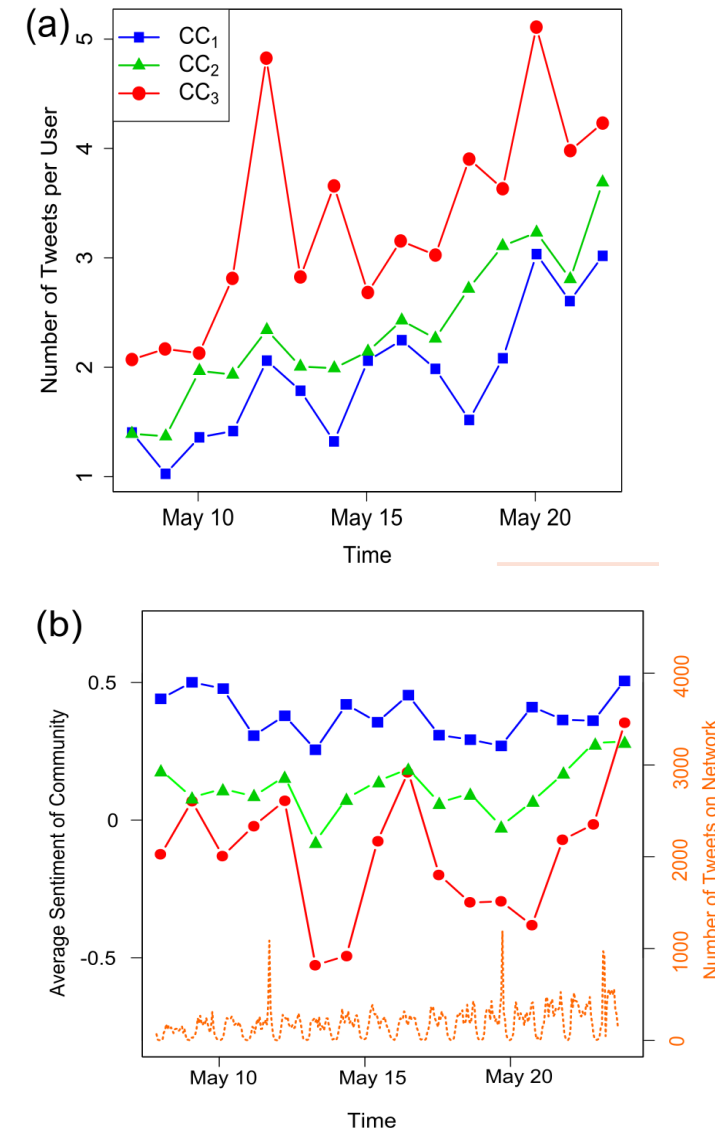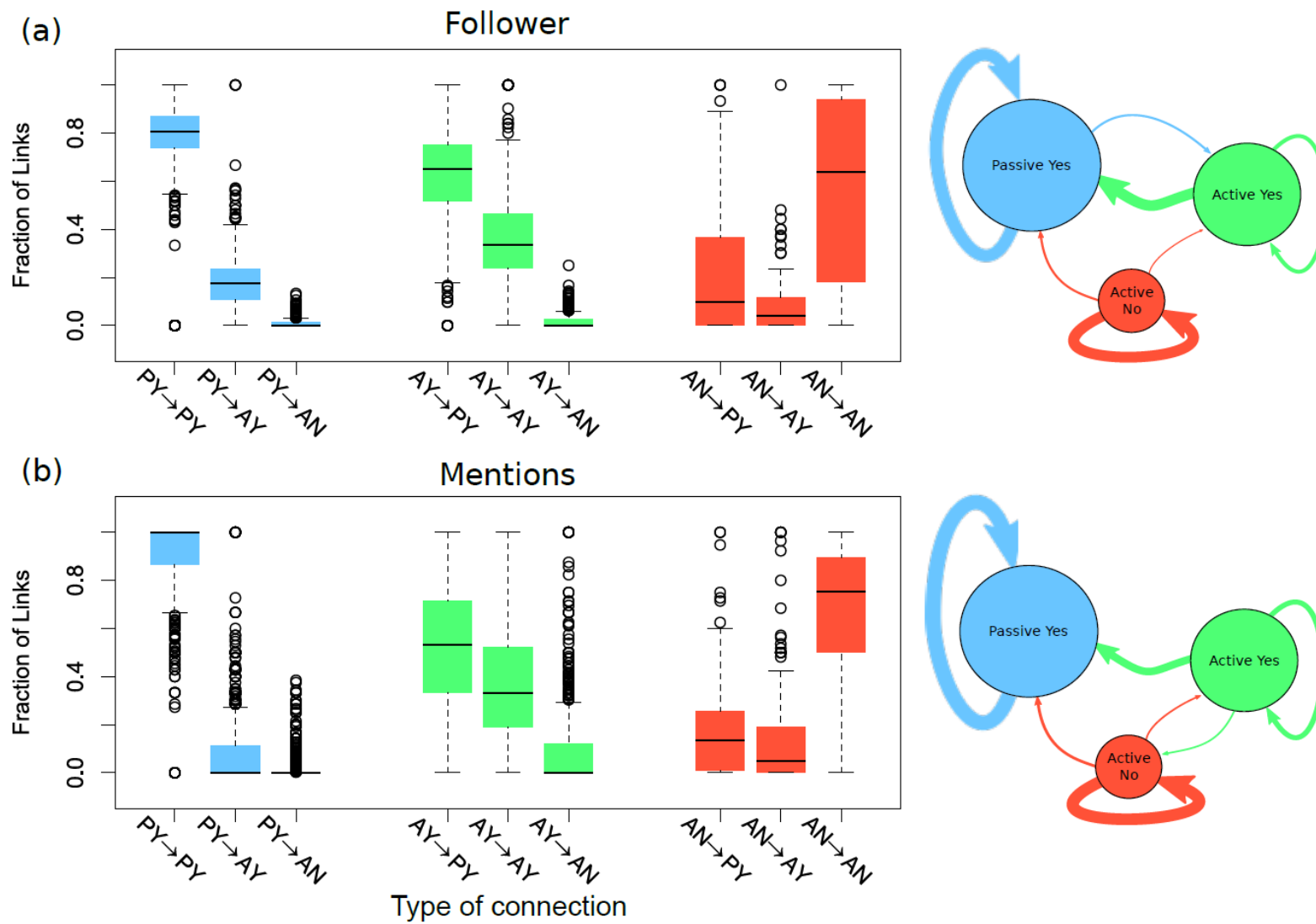
# Community Detection and Sentiment

- What about yes and no voters?
  - From each CC select random sample (20%) and classify
    - Hand classify as yes, no leaning or unaligned

- How do we classify
  - Look at the tweets the user has sent
  - Profile description

|  | | Community cluster | | | |
|---|---|---|---|---|---|
|  |  | $CC_1$ | $CC_2$ | $CC_3$ | Total |
|  | Yes | 183 | 114 | 6 | 303 |
| Alignment | No | 1 | 2 | 23 | 26 |
|  | Unaligned | 21 | 5 | 3 | 29 |
|  | Total | 205 | 121 | 32 | 358 |

- Redefine CC's
  - Accuracy 89%
  - Balanced Accuracy 81%

- How do these groups interact?

**Final Partition**

$CC_3$

$CC_2$

$CC_1$

**Final Partition**

Active No

Active Yes
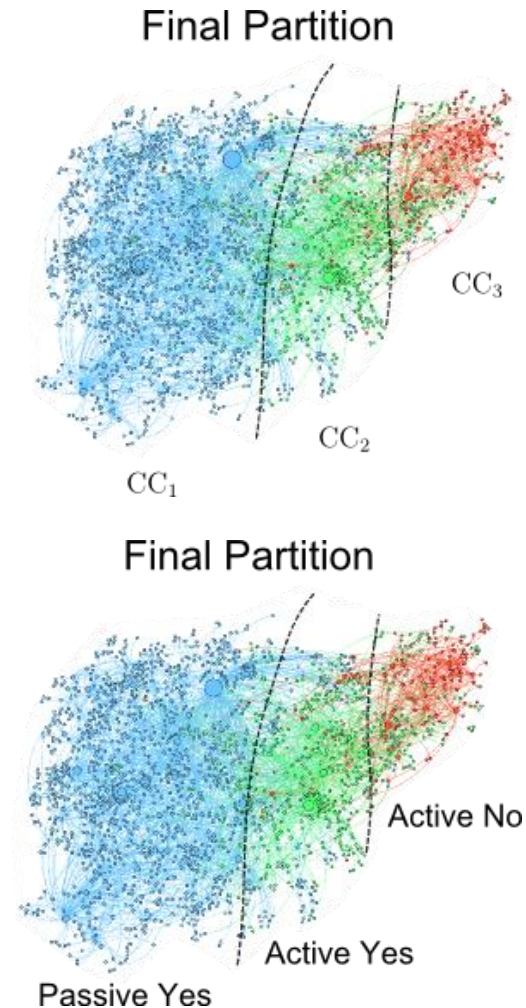
Passive Yes

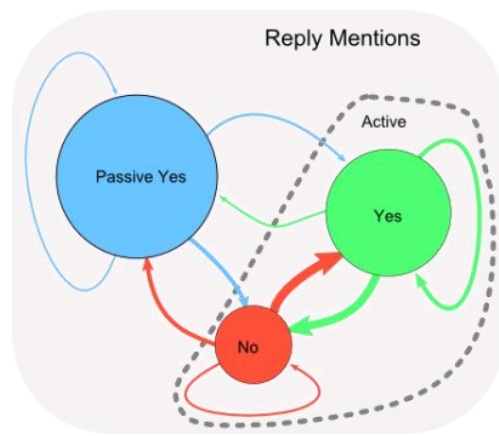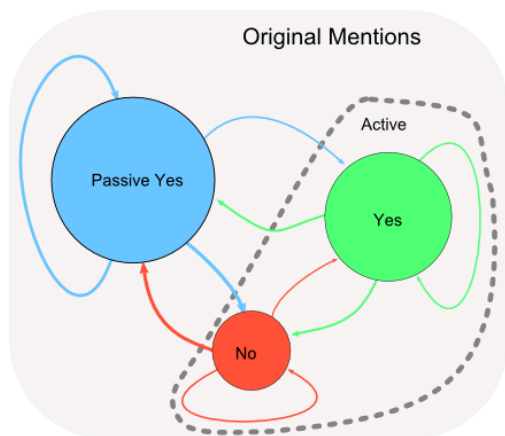# Activity between community clusters

# Activity between community clusters

| From | To | Mention tweets | | | Proportion of cluster's tweets | | |
|------|-----|---------|-------|---------|----------|-------|---------|
| | | Original | Reply | Retweet | Original | Reply | Retweet |
| Passive Yes | Passive Yes | 5302 | 1755 | 21740 | 0.16 | 0.06 | 0.68 |
| Passive Yes | Active Yes | 206 | 306 | 2152 | 0.01 | 0.01 | 0.07 |
| Passive Yes | Active No | 139 | 168 | 308 | 0.00 | 0.00 | 0.01 |
| Active Yes | Passive Yes | 1200 | 1205 | 10130 | 0.05 | 0.05 | 0.45 |
| Active Yes | Active Yes | 380 | 1935 | 4648 | 0.02 | 0.09 | 0.21 |
| Active Yes | Active No | 286 | 1948 | 601 | 0.01 | 0.09 | 0.03 |
| Active No | Passive Yes | 361 | 458 | 753 | 0.04 | 0.06 | 0.09 |
| Active No | Active Yes | 47 | 939 | 257 | 0.01 | 0.12 | 0.03 |
| Active No | Active No | 310 | 649 | 4345 | 0.04 | 0.08 | 0.54 |
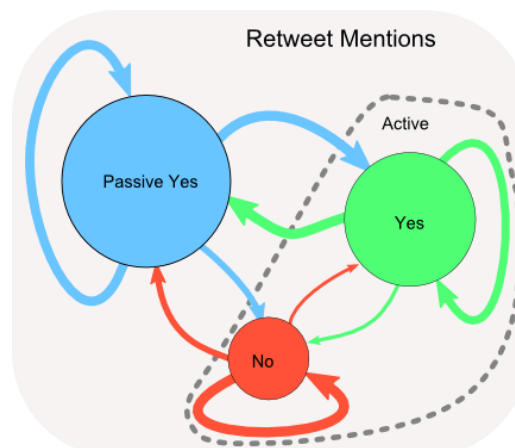
# Activity between community clusters
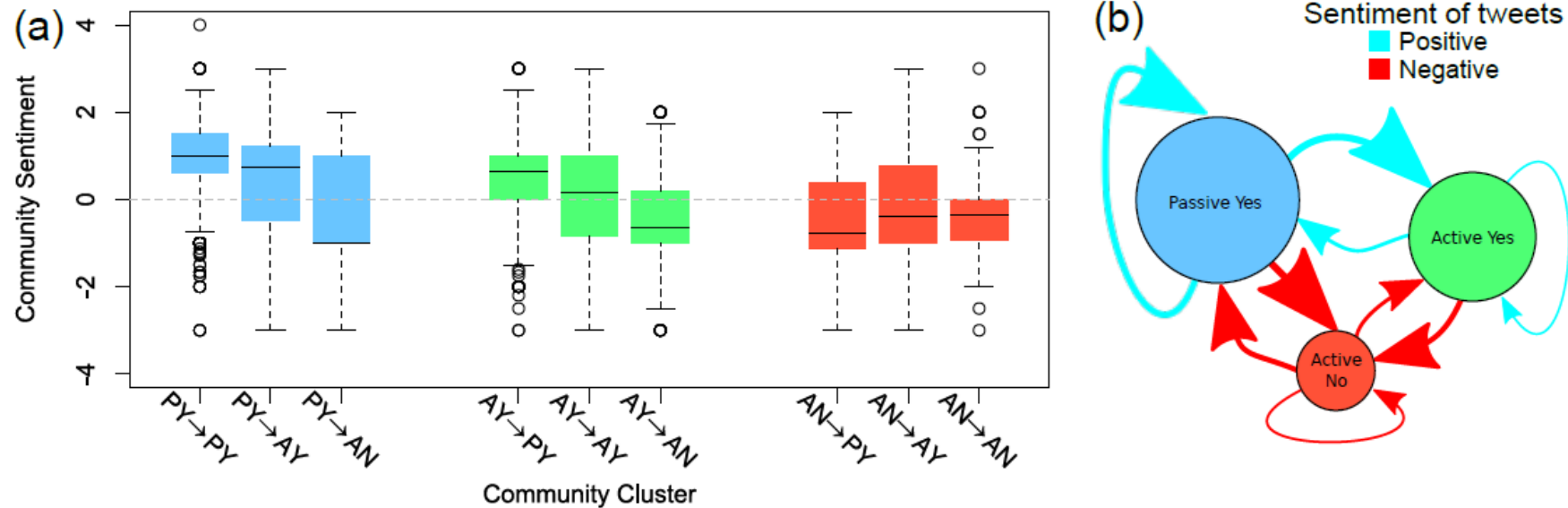


- Any thought?
  - How would you extend this work?

# SPREADING PROCESS ON NETWORKS