194.050 Social Network Analysis

Topic 04: Equivalence-based role mining

Final presentation

Babiy, Ivan

Chan, Yat Hin

Fan, Wing Yan

Muriasova, Karina

Olsiakova, Terezia

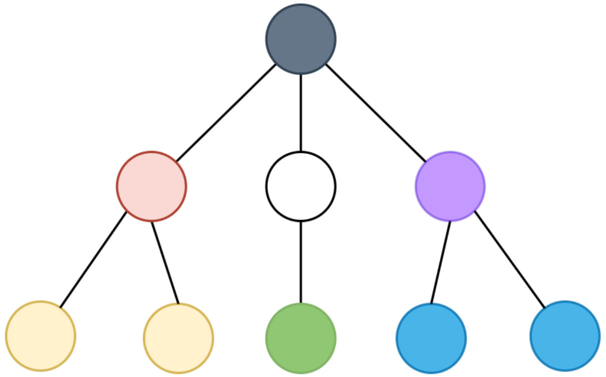
Outline

- Brief review of theoretical background
- Workflow overview
- Methodology
  - Subgraph extraction
  - Algorithms for different equivalence types
  - Clustering analysis
  - Role interpretation
- Limitations

Brief review of theoretical background

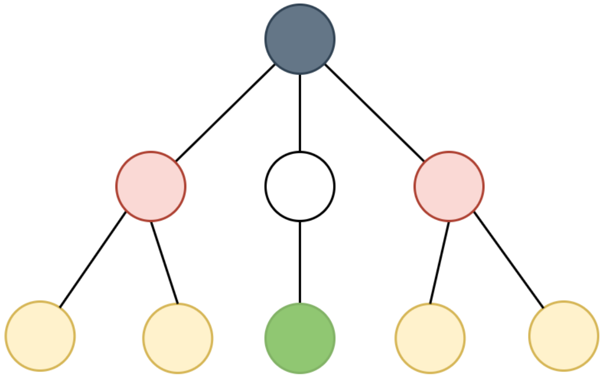
Comparing Positions: how nodes are embedded in its ego-network

Structural equivalence



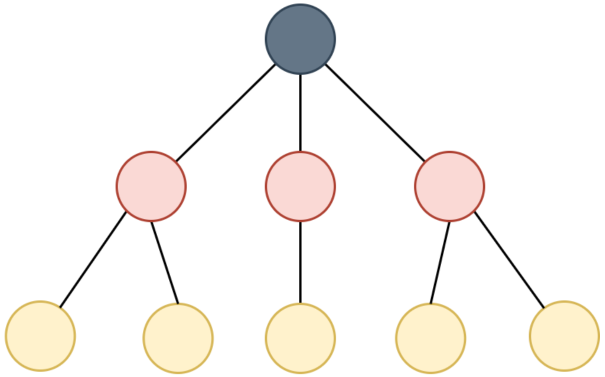
Identical composition
of neighbourhood

Automorphic equivalence



Identically-shaped neighbourhood
with same degree

Regular equivalence

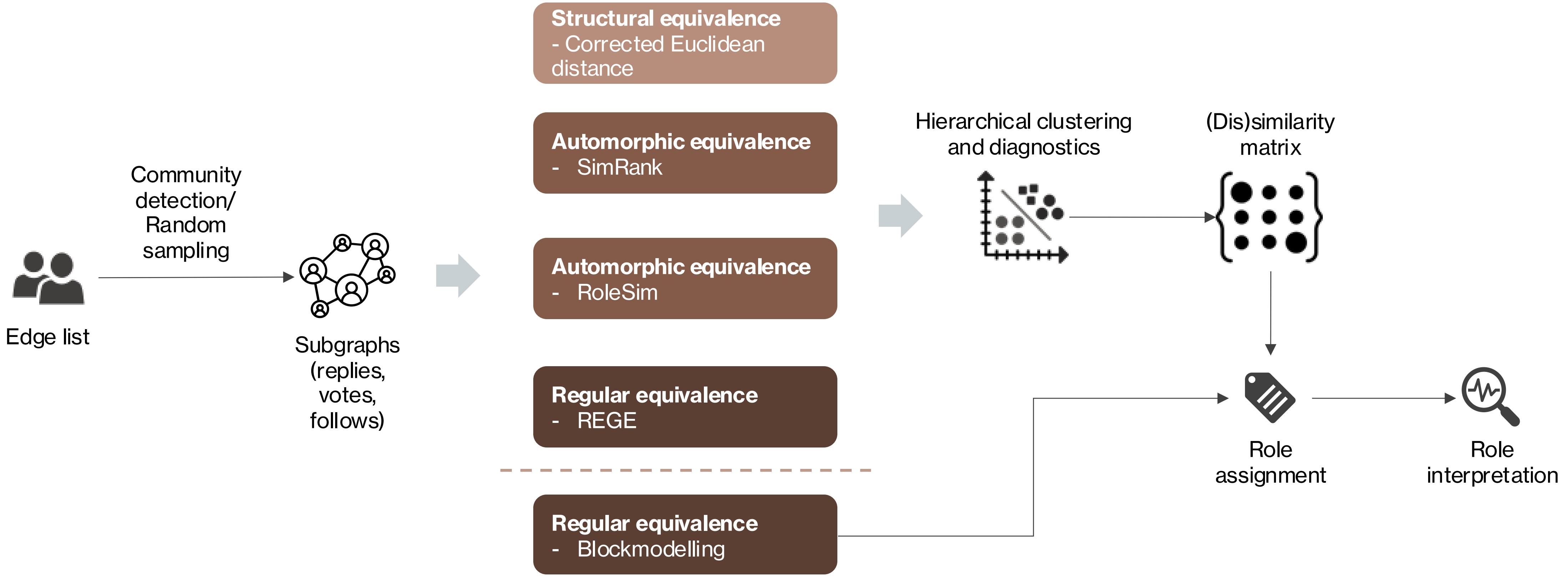


Existence of connections with
neighbour(s) of the same class

Most
restrictive

Least
restrictive

Workflow overview



Methology overview

Equivalence type	Algorithm	Subgraph	Output	Roles found
Structural	Corrected Euclidean distance	Small community	Similarity	Yes
Automorphic	SimRank	Random sampling	Similarity	No
Automorphic	RoleSim	Small community & random sampling	Similarity	No
Regular	REGE	Small community	Similarity	Yes
Regular	Blockmodelling	Small community	Role assignment	Yes

With reference to commonly predefined social network roles, we aim to find 3-8 groups/roles

Subgraph extraction with community detection

1. After creating directed graphs from the edge lists, we select only users who appear in all 3 graphs (follows, votes, replies)

2. We used modularity maximisation to detect communities in the follows graph, the sparest graph. The graph size on which all our algorithms can run within reasonable time is chosen

3. The same users in the community are extracted from all 3 graphs

type	nodes	edges
replies	253	1905
votes	253	6905
follows	253	436

Finding structural equivalence classes with Corrected Euclidean Distance

Key Concept

- Structural equivalence occurs when two nodes have identical relationships to all others, a rare condition in real-world networks

Corrected Euclidean Distance:

- Self-loops and reciprocal links
- Returns zero for equivalent nodes.

$$d_E(v_i, v_j; p) = \sqrt{\sum_{\substack{k=1 \\ k \neq i,j}}^{n} (Q_{ik} - Q_{jk})^2 + (Q_{ki} - Q_{kj})^2 + p((Q_{ii} - Q_{jj})^2 + (Q_{ij} - Q_{ji})^2}$$

Algorithm's possible results

- Dendrograms to visualize hierarchical clustering
- Blockmodels that categorize nodes into equivalence classes
- Role Assignments, offering insight into patterns and node functions

Hierarchical clustering Results

- Was applied but showed no strong equivalence clusters
- This highlights the network's heterogeneity or lack of distinct structural roles

Finding automorphic equivalence classes with SimRank and RoleSim

Automorphism

- Automorphic equivalence aims to capture nodes that are the same based on their neighborhoods

Similarity Measures

- SimRank – simple procedure, but does not satisfy necessary axioms for detecting automorphism
- RoleSim – more robust, guarantees the desired properties

Similarity matrices clustering results

- No meaningful clusters were be detected
- The network does not exhibit any such relationship where nodes participate in the network from similar contexts (neighborhoods)

Finding regular equivalence classes with REGE

- Consider "profile" of node's connection
- Suitable for directed data

Similarity Measures

- Utilise interim similarity scores as equivalence measures between iterations.
- Do not guarantee convergence to a similarity score, hence iterative approach
- Customary choice of iterations is 3

Similarity matrices clustering results

- Clusters of significant quality were detected for each graph type

Algorithm:

1. Define $\mathbf{M}^t$ as a matrix where $0 \leq \mathbf{M}^t_{ij} \leq 1$ quantifies the degree to which nodes v_i and v_j are regularly equivalent ($\mathbf{M}^t_{ij} = 1$ if they are perfectly equivalent, $\mathbf{M}_{ij} = 0$ if they are perfectly inequivalent) after the t^{th} iteration of REGE. Initialize $\mathbf{M}^0_{ij} = 1$ for all i and j and $t = 0$.
2. Select a maximum number of iterations x.
3. for $t = 0, 1 \ldots, x$ compute:

$$\mathbf{M}^{(t+1)}_{ij} = \frac{\sum_{k=1}^{g} \max_{m=1}^{g} \mathbf{M}^t_{km}({}_{ij}\mathcal{M}_{km} +_{ji} \mathcal{M}_{km})}{\sum_{k=1}^{g} \max^*_m ({}_{ij}\mathrm{Max}_{km} +_{ji} \mathrm{Max}_{km})}$$

4. return $\mathbf{M}^x$.

Clustering analysis

Motivation
- Difficult to find exact equivalence in real-life scenario

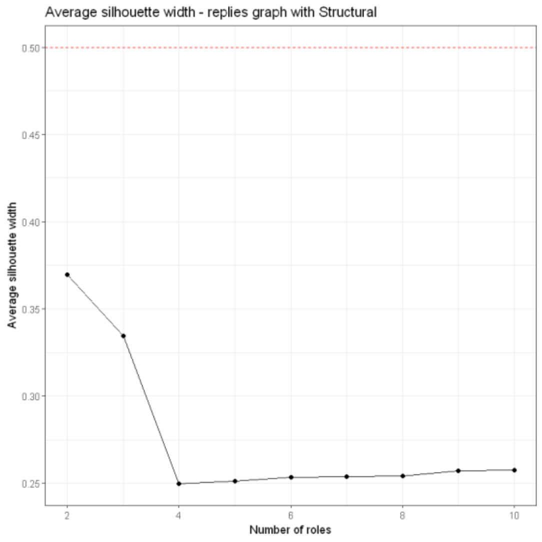
Setting:
- Approximation with Similarity matrices from algorithms

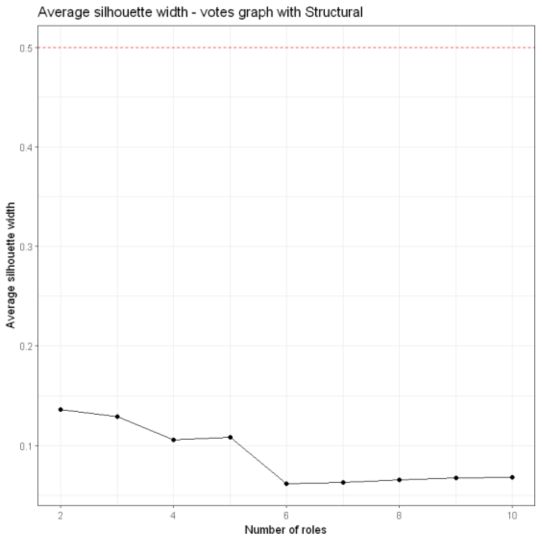
Approach: Agglomerative hierarchical clustering
- Similarity definition:
  Complete linkage / Ward's method
- Distance measure:
  Structural: Compatible variant of Euclidean distance
  Automorphic, Regular: Euclidean distance
- Cluster number decision and cluster quality evaluation:
  Based on average silhouette width supported by dendrogram

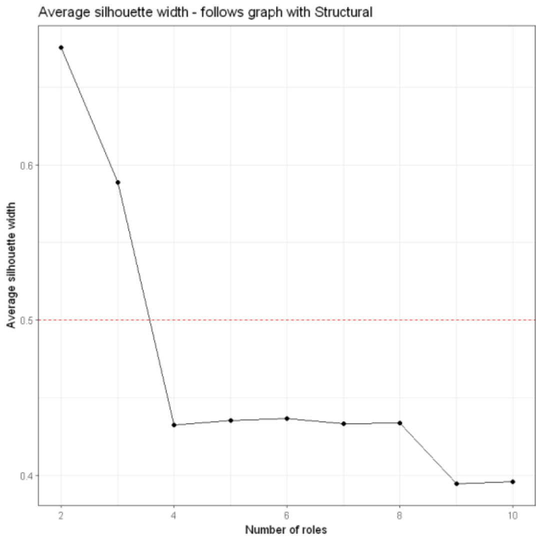
Clustering analysis

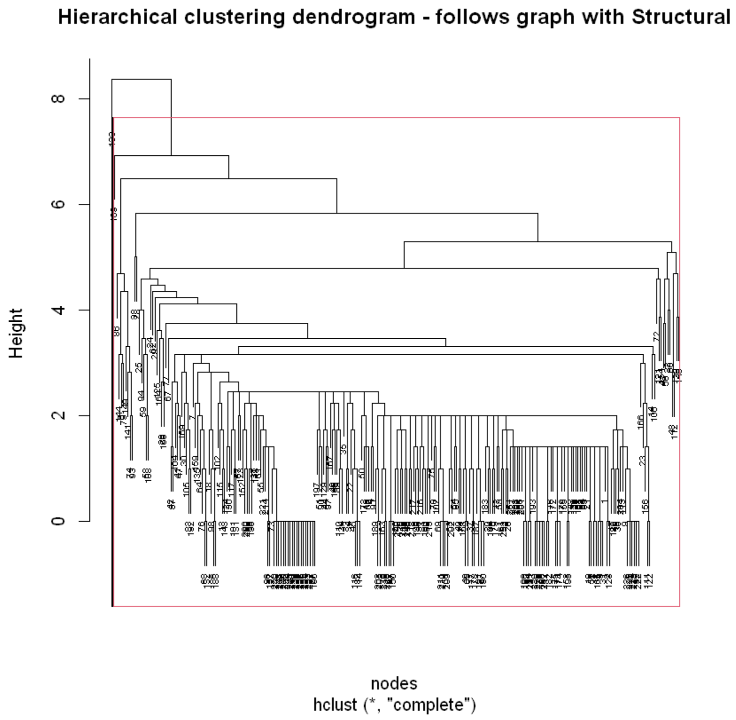
Structural equivalence

- Clusters identified but with insignificant number of members





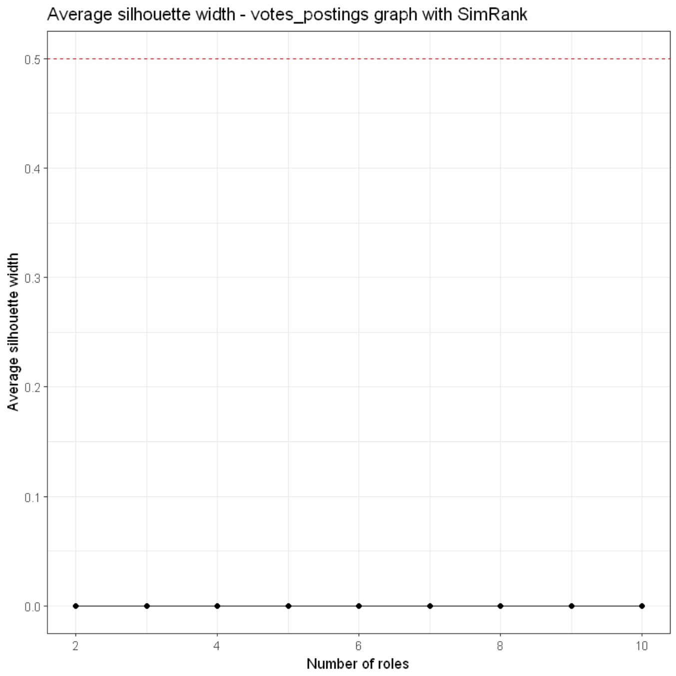


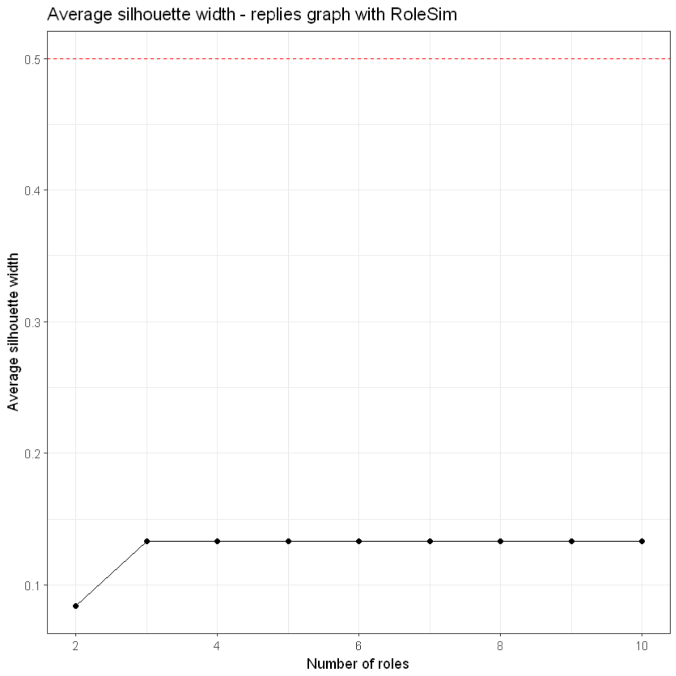


Clustering analysis

Automorphic equivalence – SimRank and RoleSim

- Either no clusters identified, or clusters identified but with insignificant number of members
- Examples:

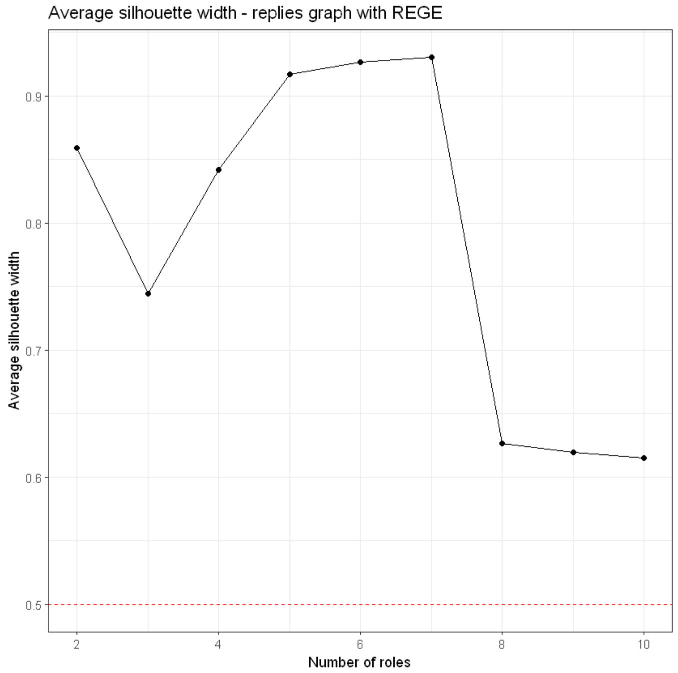


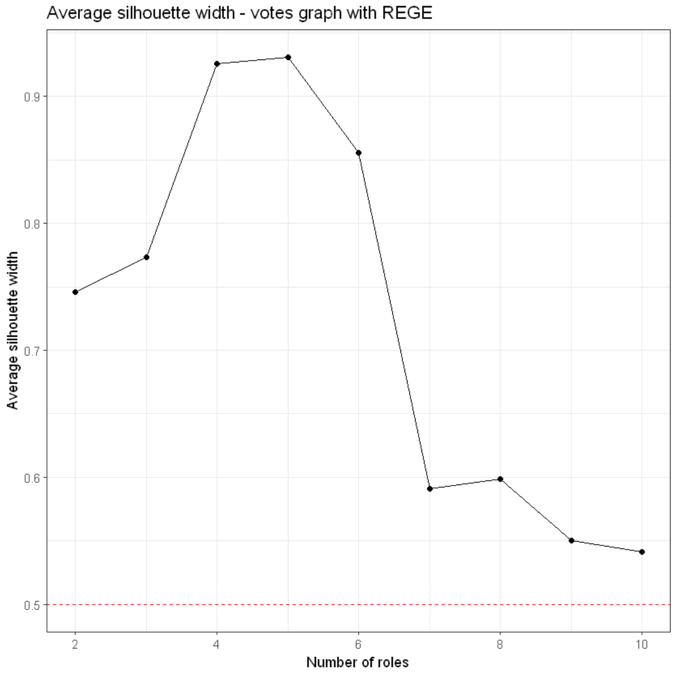


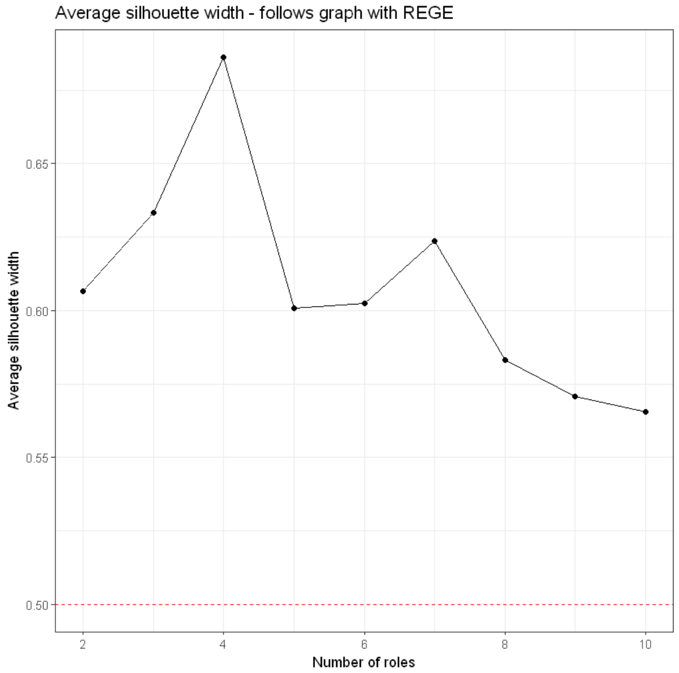
Clustering analysis

Regular equivalence - REGE

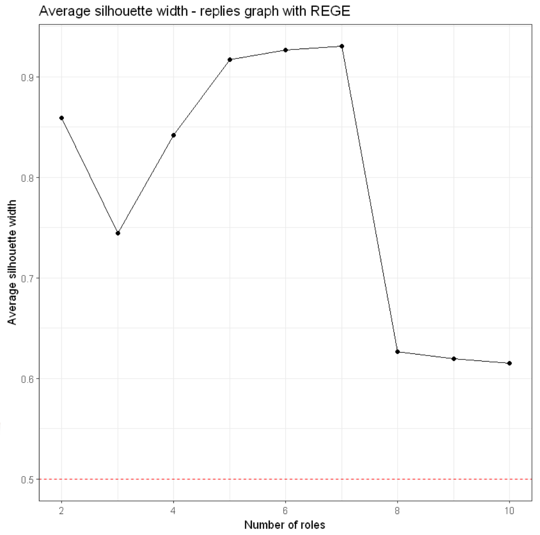
- Candidates with similar silhouette width in replies and votes graph for further decision





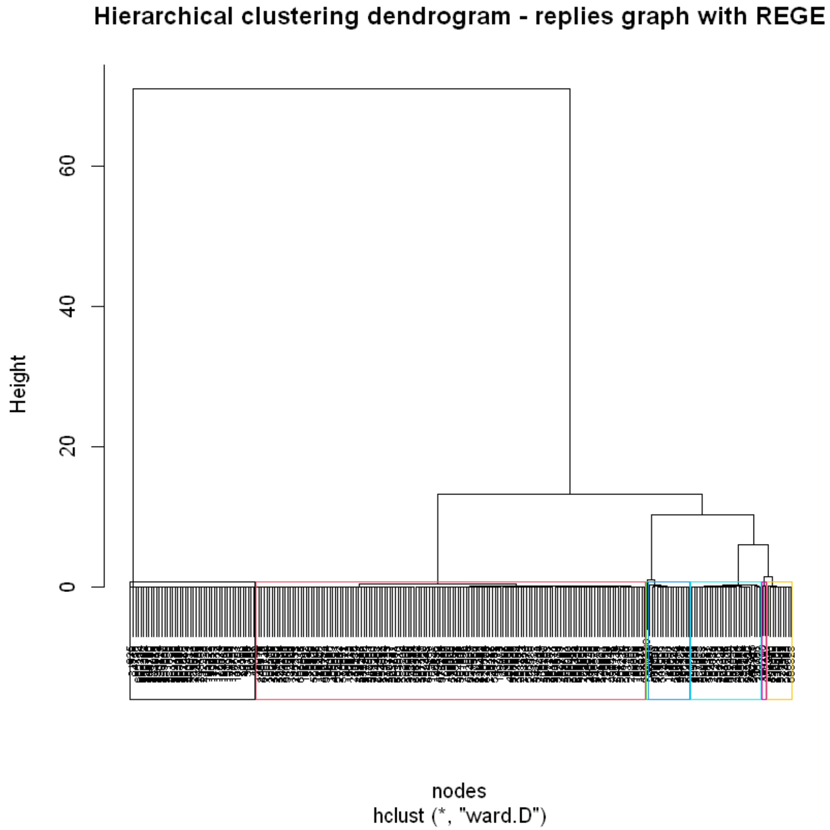


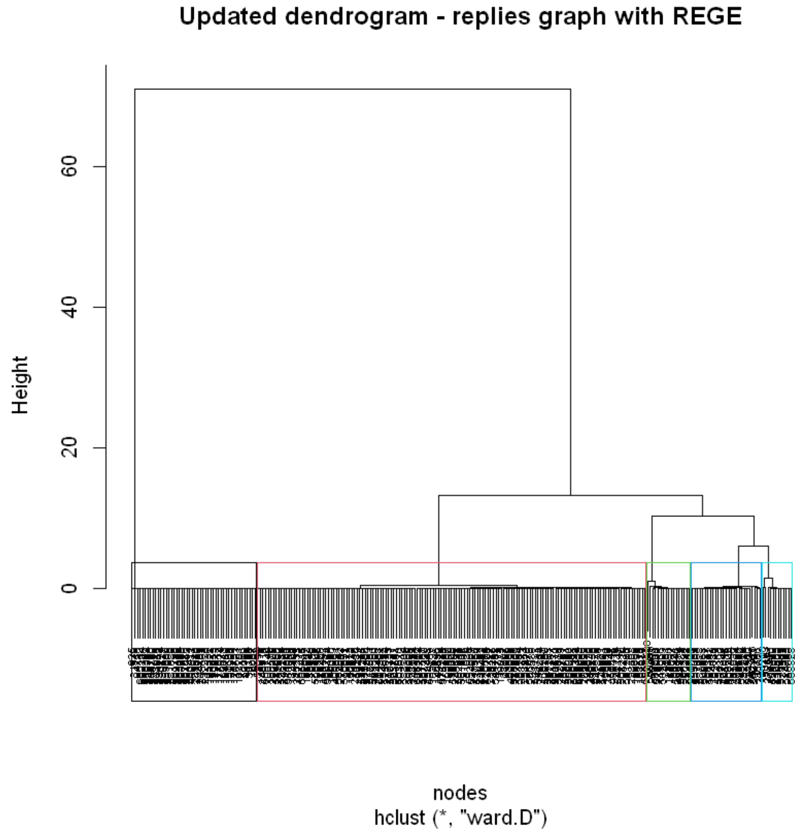
Clustering analysis



Regular equivalence - REGE

- Balance between role number and role size, e.g. in replies graph





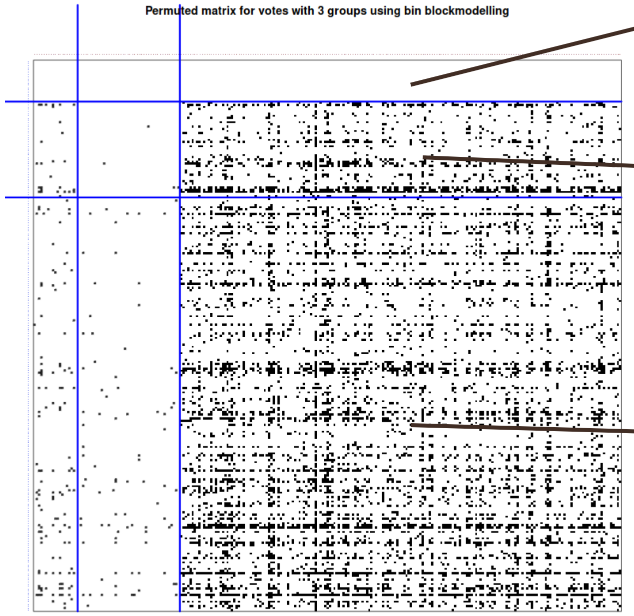
Finding regular equivalence classes with blockmodelling

Algorithm:

1. Define k classes $\mathcal{C} = \{C_1, \ldots C_k\}$ and randomly assign nodes into the k classes.
2. Permute $\mathbf{A}$ so that nodes in the same class are in consecutive rows.
3. For a block describing the relations between the members of classes R_i and R_j define the measure $d(R_i, R_j)$ as:

$$d(R_i, R_j) = \min \begin{cases} \text{\# of rows and columns of the block with at least one 1} \\ \text{\# of rows and columns of the block that only contain 0} \end{cases}$$

4. Initialize $S(\mathcal{C}) = \sum_{i,j} d(R_i, R_j)$.
5. For every transformation $t(\mathcal{C})$ that moves a node from one class to another or interchanges a node between two classes, compute $S(t(\mathcal{C}))$.
6. Let $S^* = \min_t S(t(\mathcal{C}))$ be the transformation with the smallest value of S.
7. If $0 < S^* < S$:
   a. Set $S = S^*$ and $\mathcal{C} = t(\mathcal{C})$.
   b. Repeat from step 5.
8. Otherwise, if $S = 0$ return $\mathcal{C}$.
9. After trying all transformations, return $\mathcal{C}$ with and label it `approximate solution`.



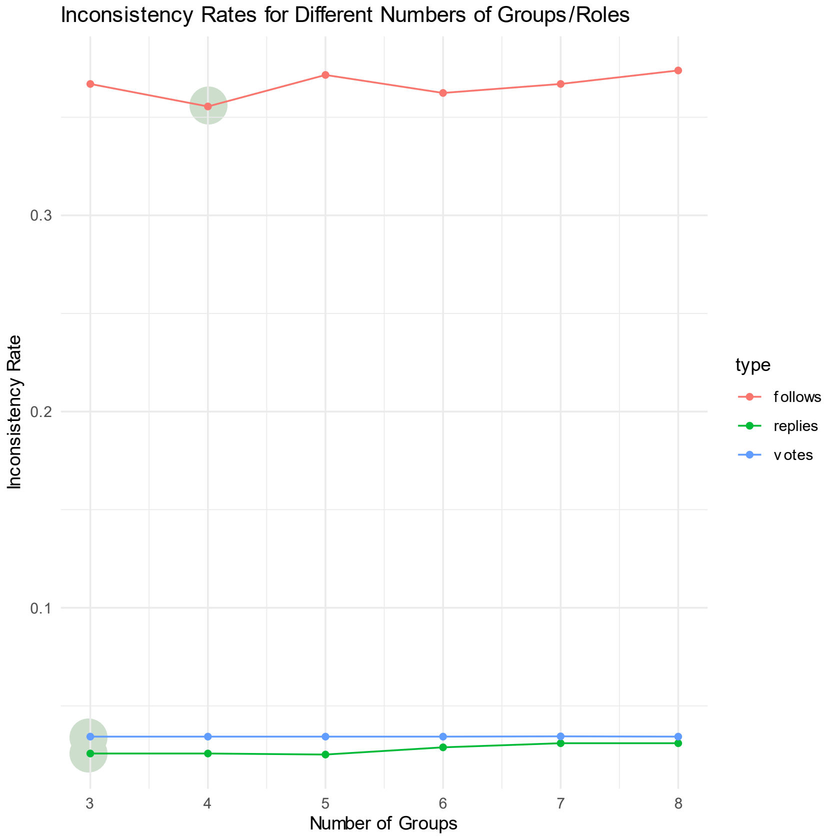
Users in group 1 have not voted for any users

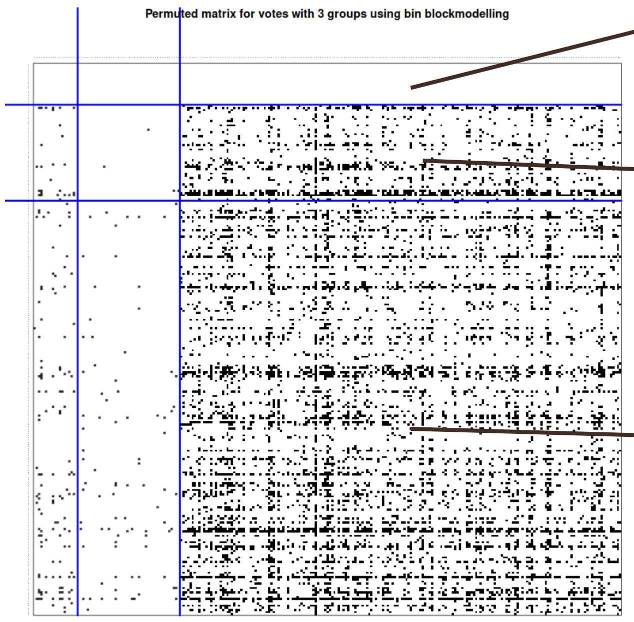
Users in group 2 have voted at least 1 user in group 3

Users in group 3 have voted and have been voted by at least 1 other user within the group

Connections in empty blocks and empty rows/columns in connected blocks are inconsistent with the role definition, to be minimised

Finding regular equivalence classes with blockmodelling





Users in group 1 have not voted for any users

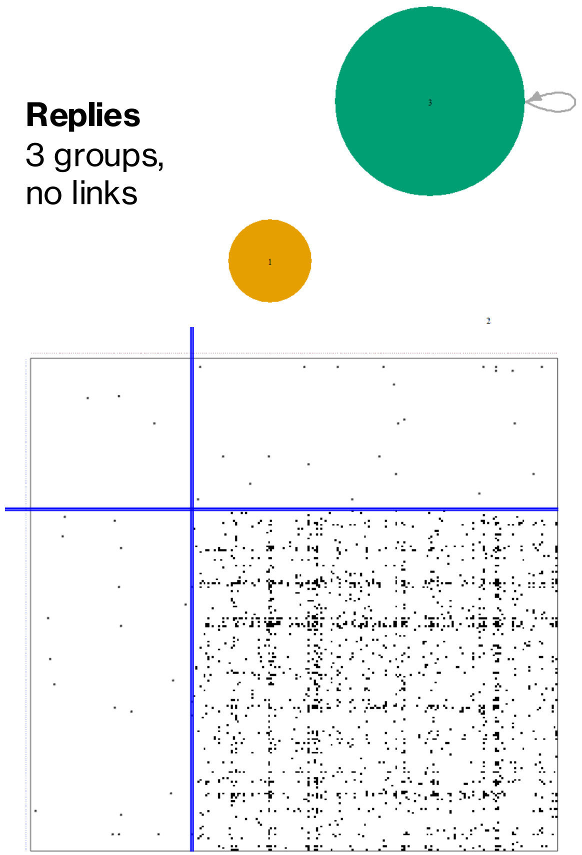
Users in group 2 have voted at least 1 user in group 3

Users in group 3 have voted and have been voted by at least 1 other user within the group

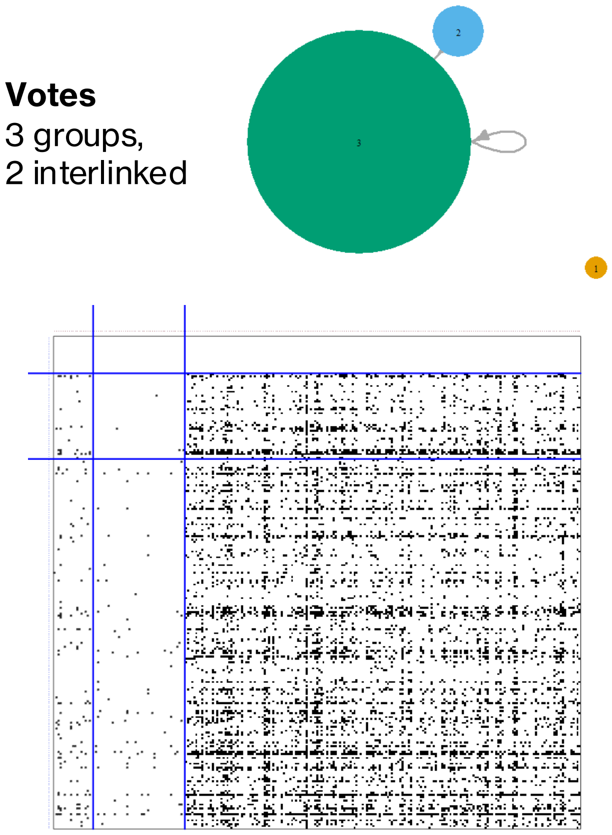
Connections in empty blocks and empty rows/columns in connected blocks are inconsistent with the role definition, to be minimised

Role interpretation for blockmodelling

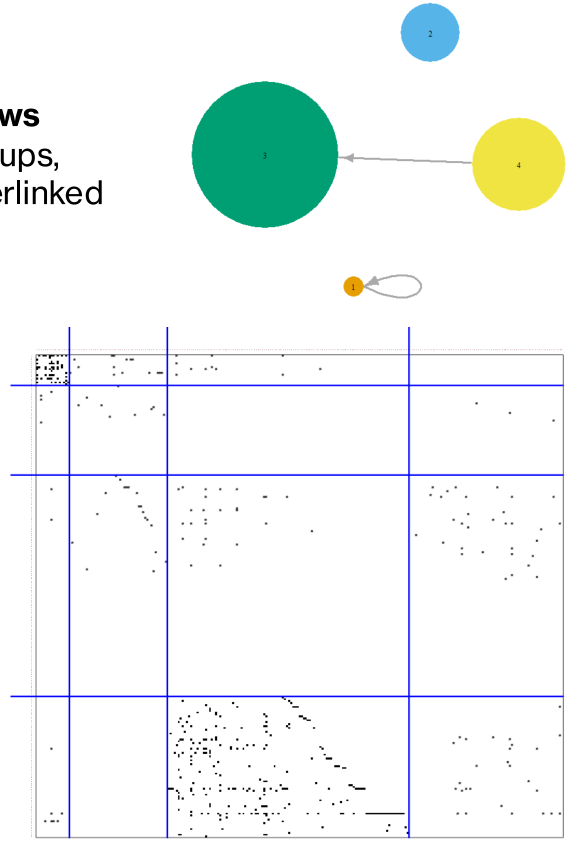
Replies
3 groups,
no links



Votes
3 groups,
2 interlinked



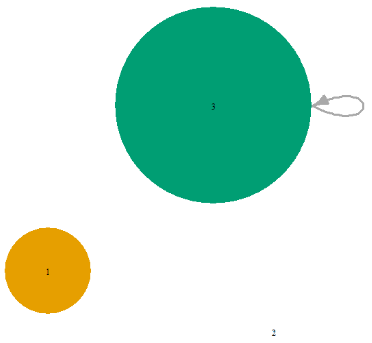
Follows
4 groups,
2 interlinked

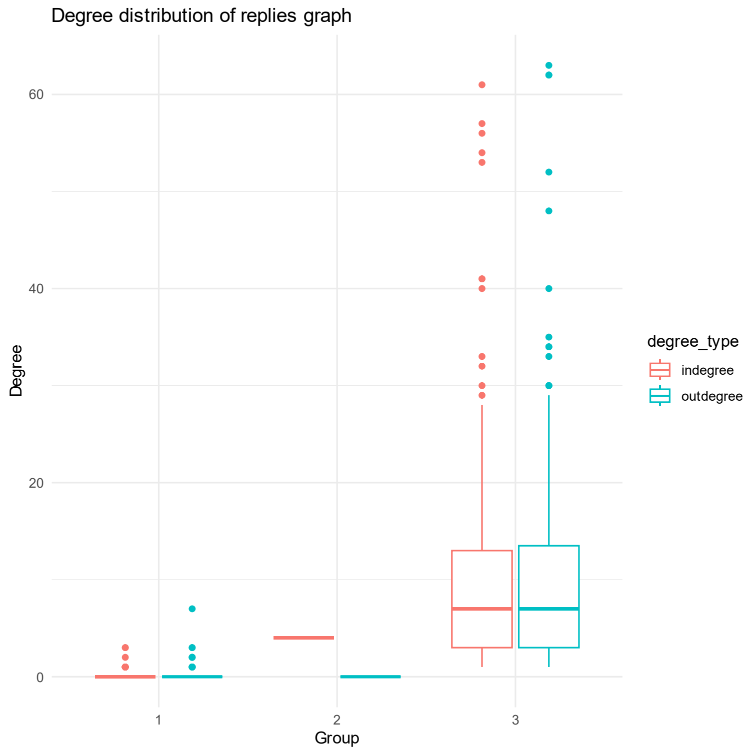


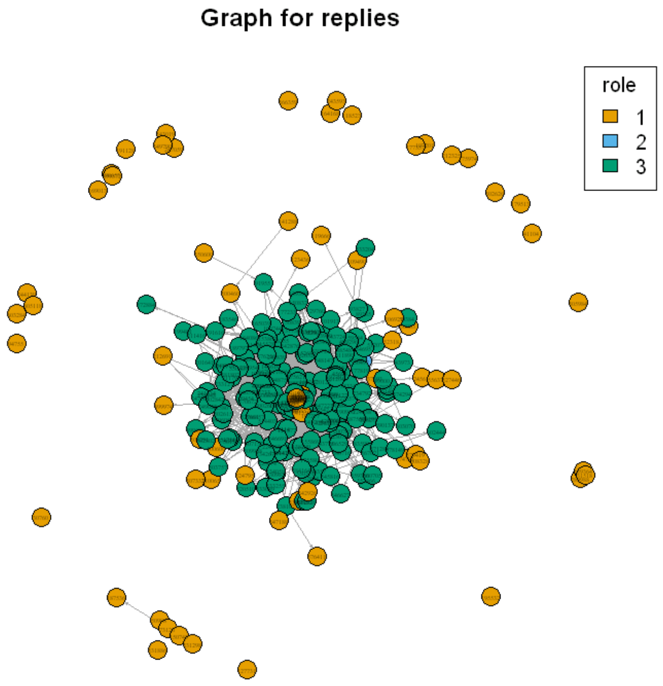
Role interpretation for blockmodelling

Replies

- Group 1
    - Some individuals not linked
    - Interacts with G3 – replies
- Group 2
    - Has only one user
- Group 3
    - Largest
    - Many interactions, also reflexive
    - role: conversation starters



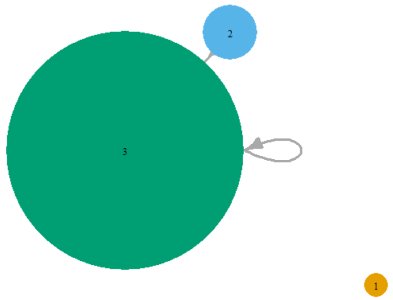


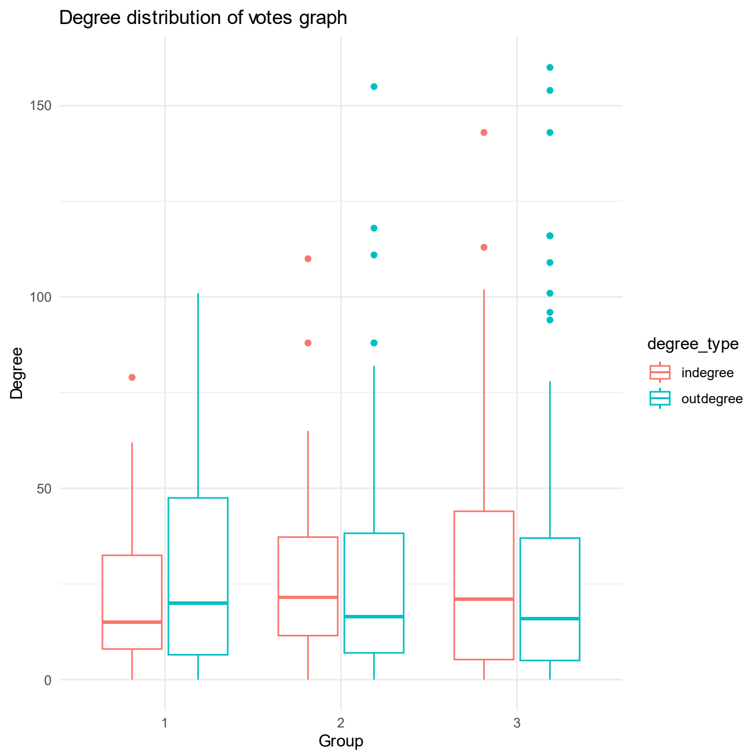


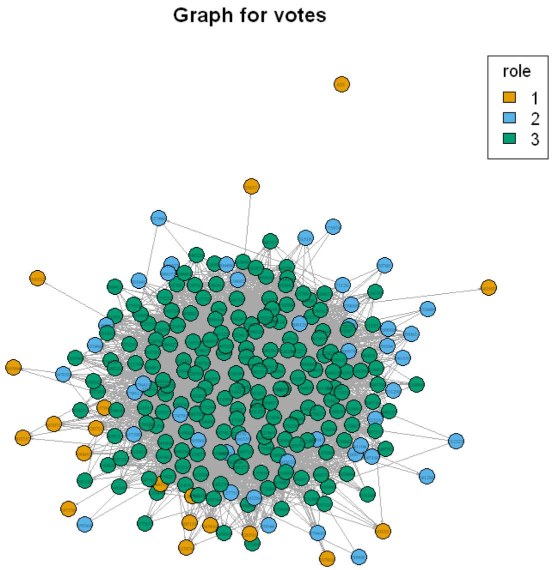
Role interpretation for blockmodelling

Votes

- Group 1
  - Smallest, few interactions
- Group 2
  - Interacts with G3, multiple interactions
- Group 3
  - More received votes than voted
- Weak resemblance to predefined roles



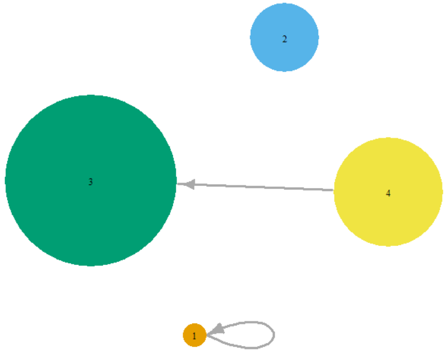


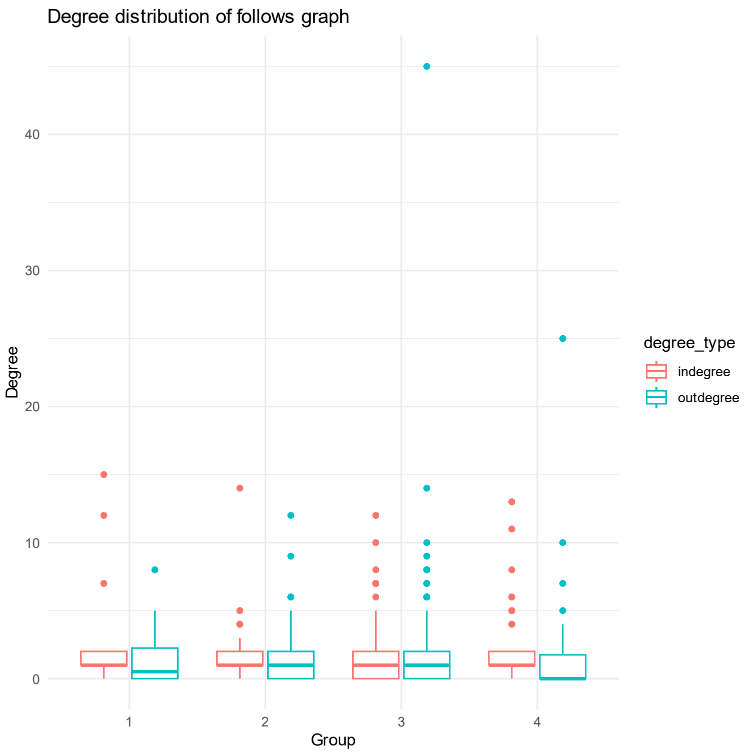


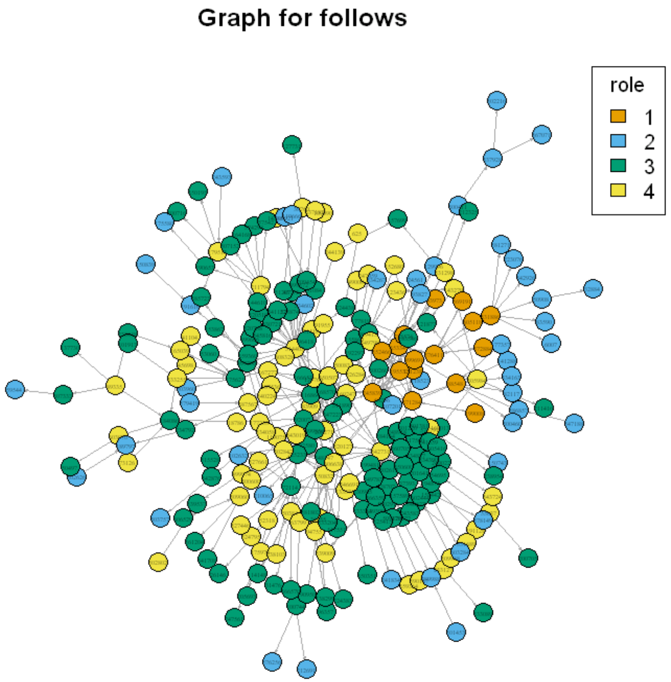
Role interpretation for blockmodelling

Follows

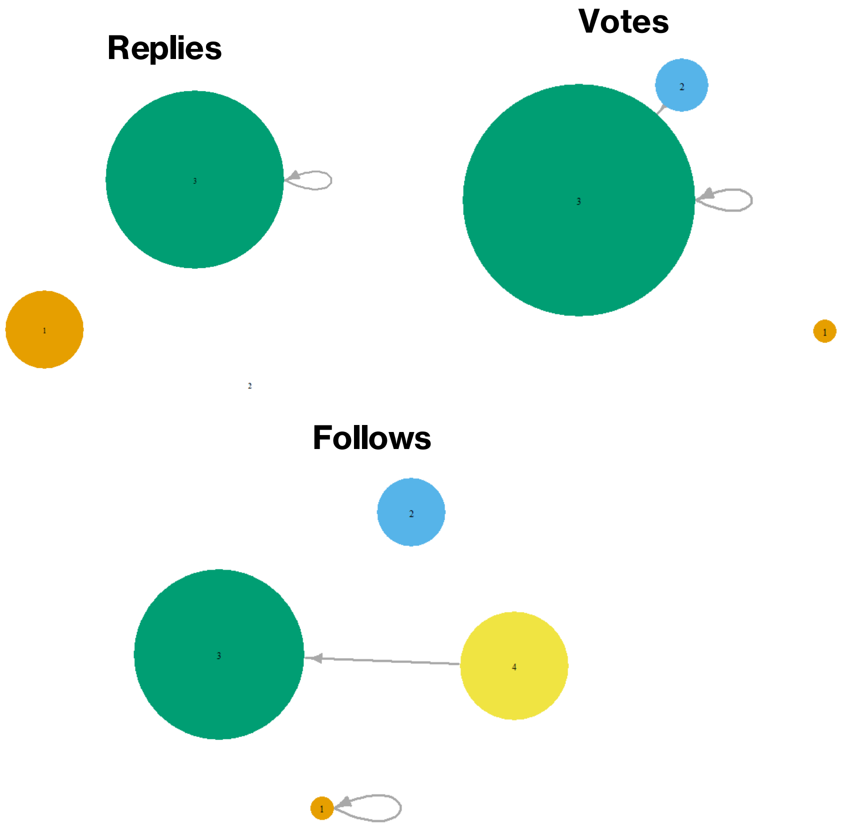
- Group 1
  - more isolated
- Group 2
  - on the periphery of the network
- Group 3 & Group 4
  - Interlinked
- No resemblance to predefined roles

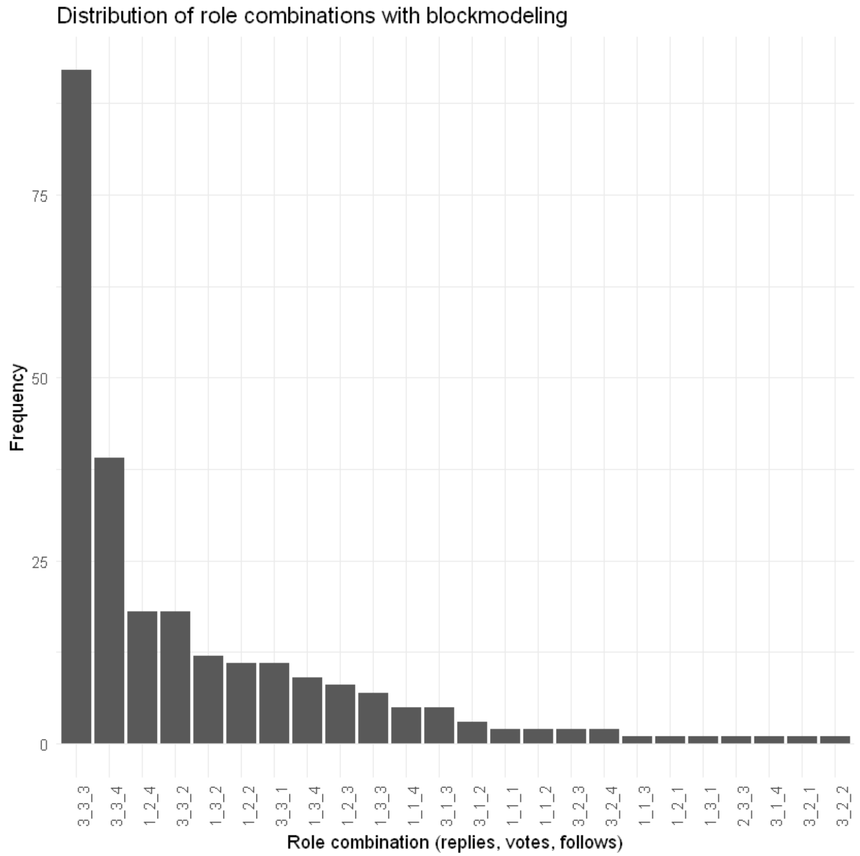






Role interpretation for blockmodelling



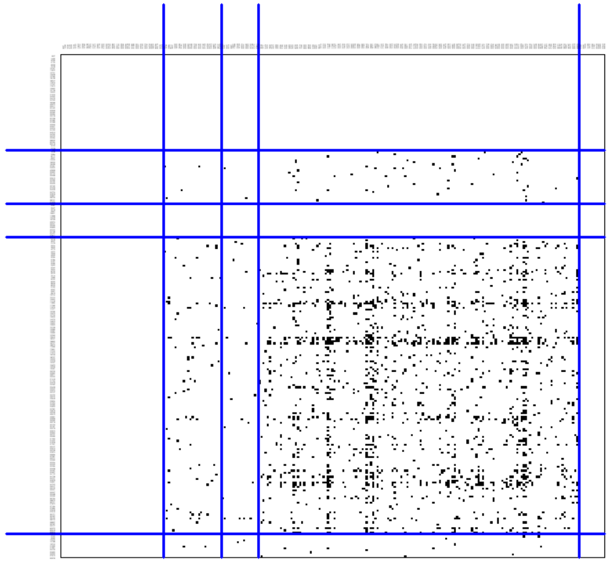


Role interpretation for REGE

Unlike the blockmodeling approach, the resulting roles are not as direct and interpretable

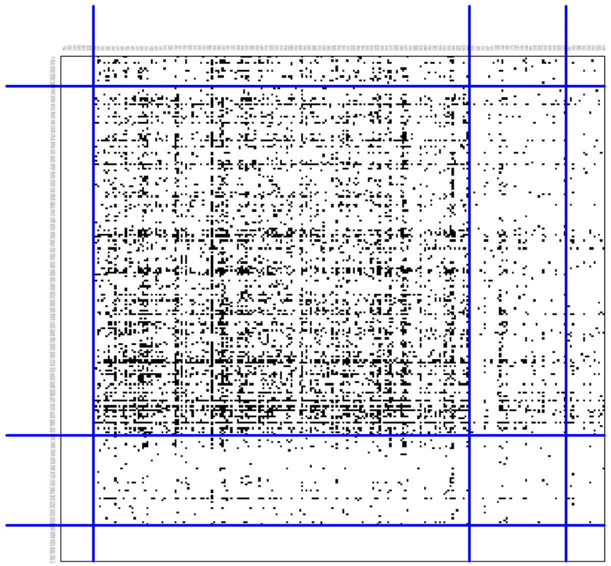
Replies
5 groups

Partitioned matrix for replies graph with REGE



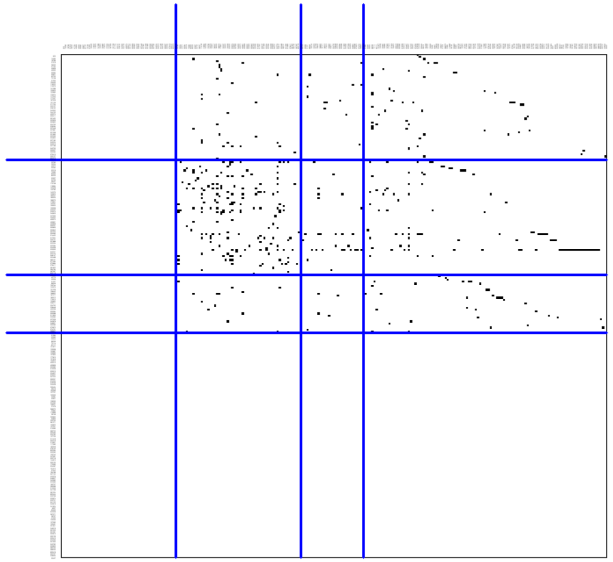
Votes
4 groups

Partitioned matrix for votes graph with REGE



Follows
4 groups

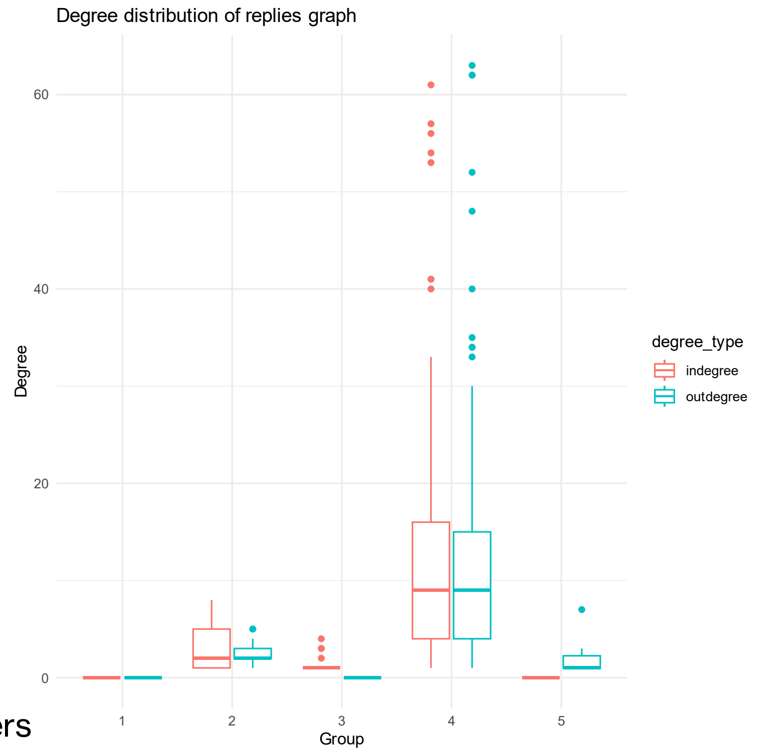
Partitioned matrix for follows graph with REGE

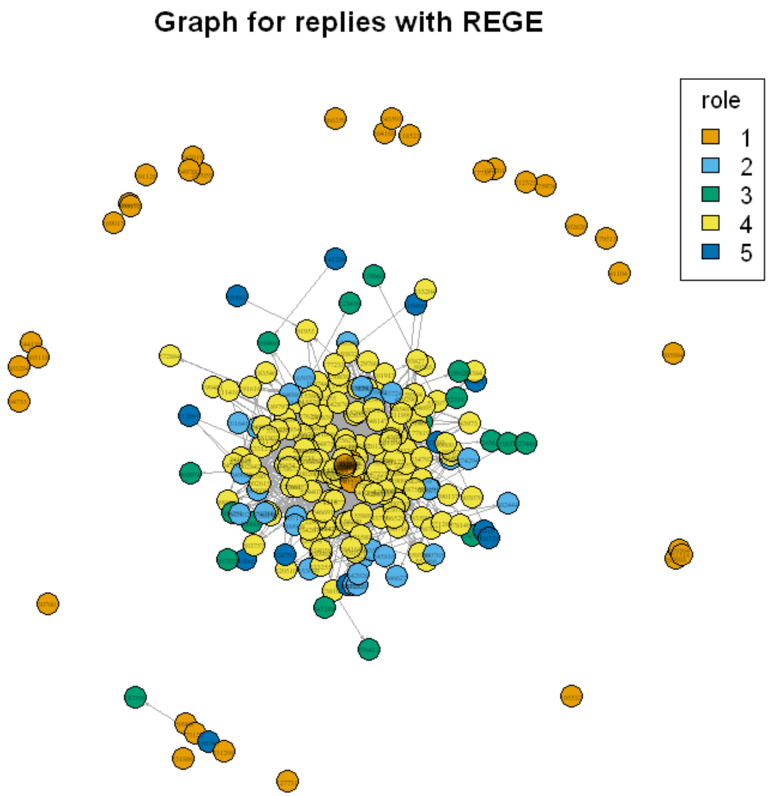


Role interpretation for REGE

Replies

- Group 1
  - Rather isolated, no or little interactions with other nodes
  - Potential role: inactive browser
- Group 2
  - Form a community with itself
  - Potential role: a small isolated online community
- Group 3
  - Do not reply to other nodes
  - Receive replies from other nodes
- Group 4
  - Largest group
  - Form a community with itself
  - Potential role: the majority active users
- Group 5
  - Reply to other nodes
  - Do not receive replies from other nodes

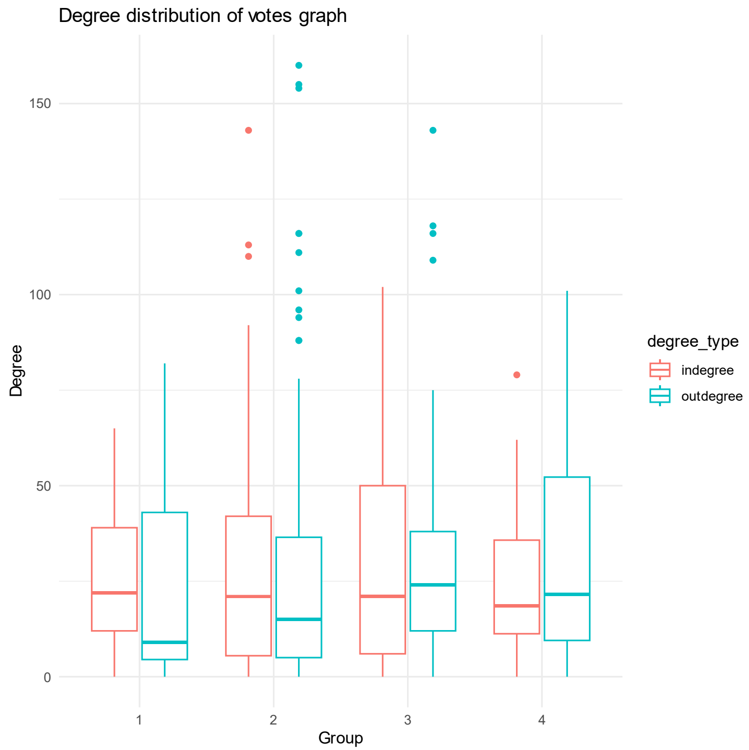


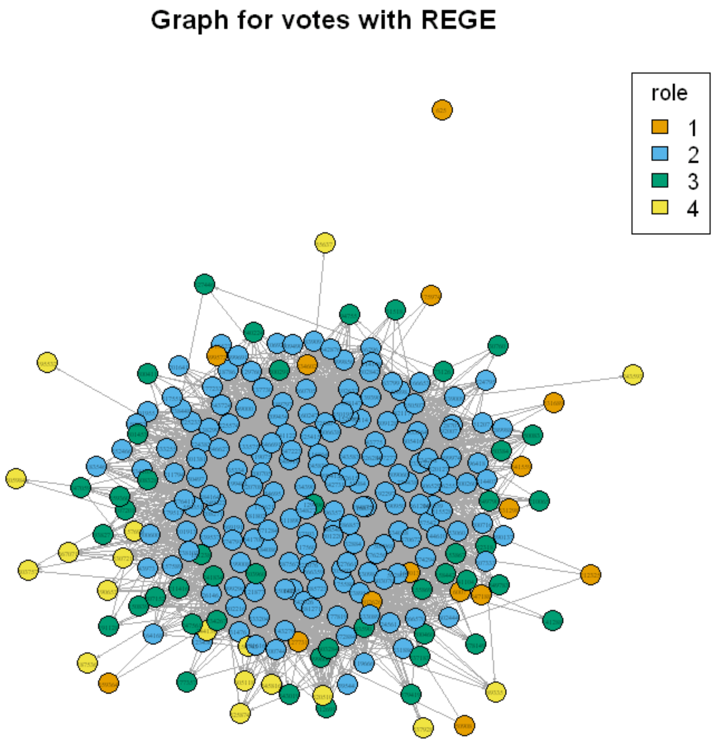


Role interpretation for REGE

Votes

- Group 1
  - Vote others
- Group 2
  - Largest
  - interlinked
- Group 3
  - Form a community
- Group 4
  - Receive votes

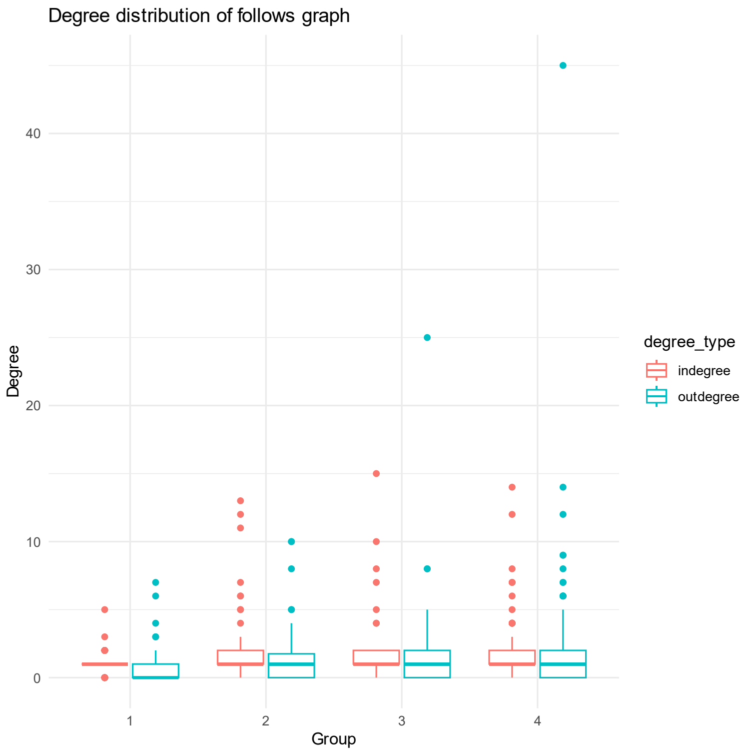


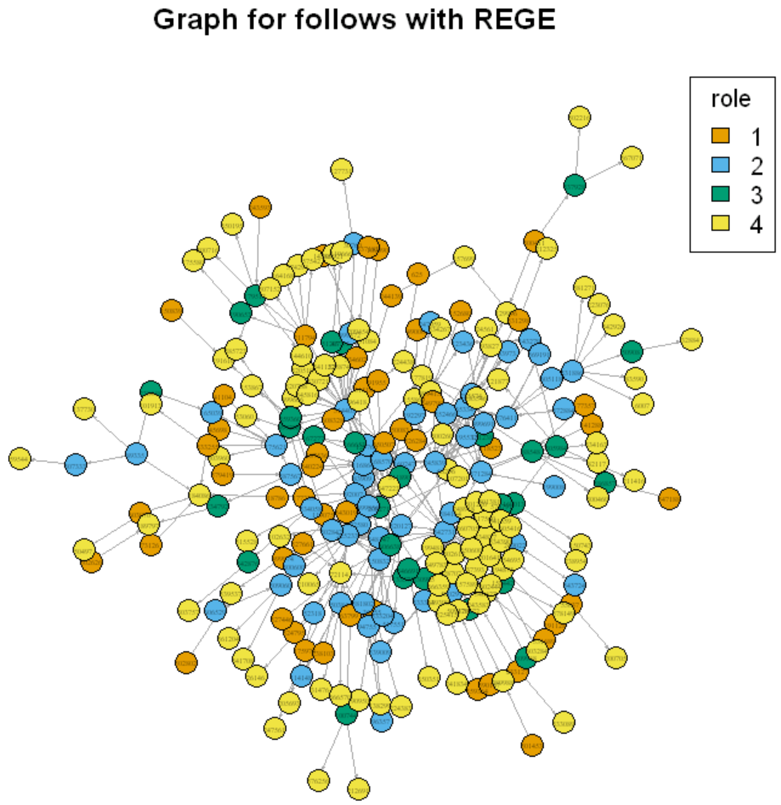


Role interpretation for REGE

Follows
More comparable group size

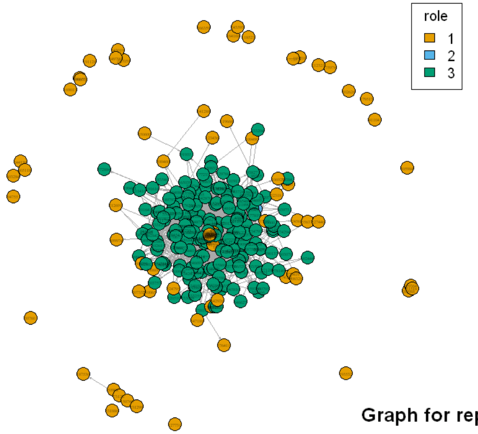
- Group 1
  - Follow other group
- Group 2
  - Follow other group and within its own group
- Group 3
  - Follow other group and
  - Being followed by other group
- Group 4
  - Followed by other group
  - Potential role: celebrities or influencers



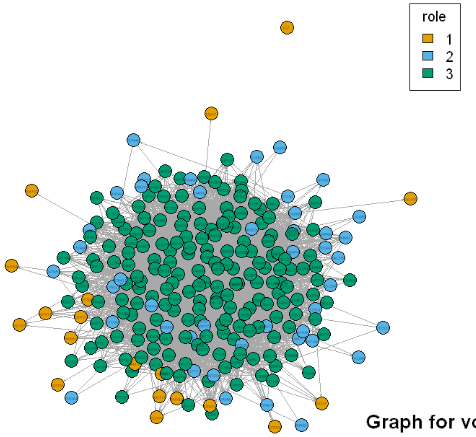


Overview of grouping by blockmodeling and REGE

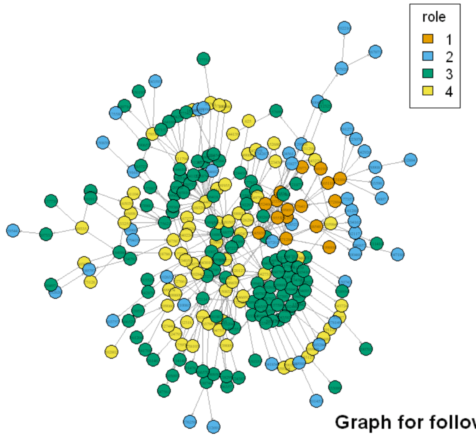
Graph for replies with blockmodeling



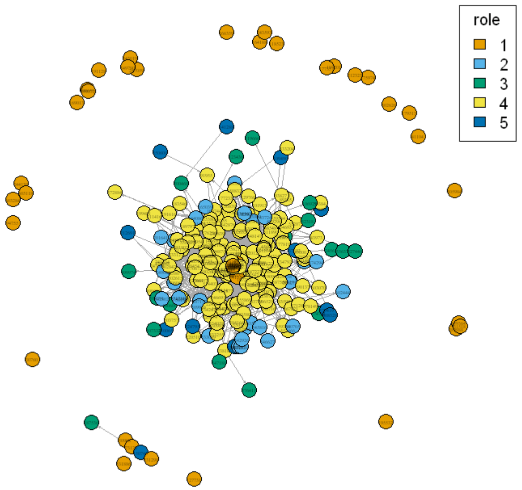
Graph for votes with blockmodeling



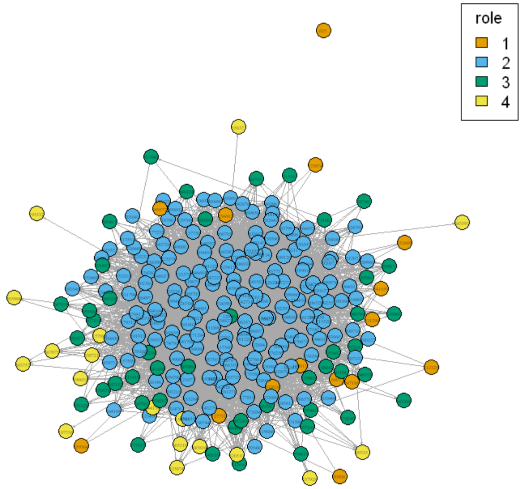
Graph for follows with blockmodeling



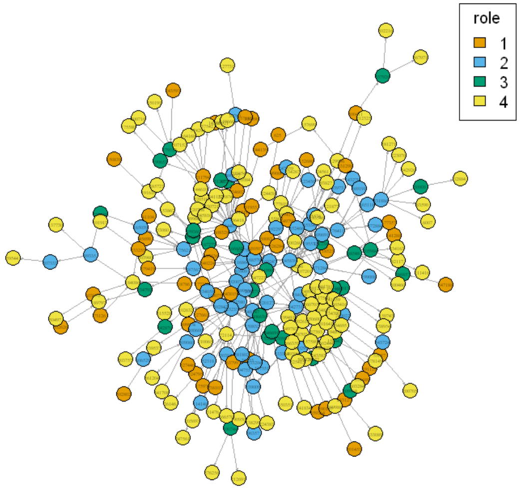
Graph for replies with REGE



Graph for votes with REGE



Graph for follows with REGE



Limitations

- Small graph
- Subgraphs are based on existing community detected in the follows graph, which might influence results
- Only unvalued graph used
- Interpretation based on metrics might not be intended for roles found by equivalence classes, since roles are based on node relationships, not on metrics

References

- Doran, D. (2017). Network Role Mining and Analysis. Springer International Publishing, Chapter 3.
- Murphy, P. (2019). An Introduction to Linear Regression in R. RPubs. https://rpubs.com/pjmurphy/436072
- Matjašič, Miha & Cugmas, Marjan & Žiberna, Aleš (2021). "blockmodeling: an R package for Generalized Blockmodeling," SocArXiv b8cxp, Center for Open Science.