Signed Graph Analysis for the Interpretation of Voting Behavior

Nejat Arinik
Rosa Figueiredo
Vincent Labatut

Laboratoire Informatique d'Avignon, University of Avignon, France.

Graz, OCT 11 - 12, 2017

Outline

Context

- Partitioning European Parliament (EP) vote networks
  - Extracting signed networks
  - Partitioning signed networks

Our previous works	Limitations	How to deal with?

Context

- Partitioning European Parliament (EP) vote networks
  - Extracting signed networks
  - Partitioning signed networks

Our previous works	Limitations	How to deal with?
1. Relevance of Negative Links in Graph Partitioning [Mendonça et al., 2015] 		

Context

- Partitioning European Parliament (EP) vote networks
  - Extracting signed networks
  - Partitioning signed networks

Our previous works	Limitations	How to deal with?
1. Relevance of Negative Links in Graph Partitioning [Mendonça et al., 2015] 	• uncomplete data  • fully-connected networks  • interpretation problems 	

Context

- Partitioning European Parliament (EP) vote networks
  - Extracting signed networks
  - Partitioning signed networks

Our previous works	Limitations	How to deal with?
1. Relevance of Negative Links in Graph Partitioning [Mendonça et al., 2015] 	• uncomplete data 	
2. Brazilian Congress structural balance analysis [Levorato and Frota, 2017] 	• fully-connected networks 	
	• interpretation problems 	

Context

- Partitioning European Parliament (EP) vote networks
  - Extracting signed networks
  - Partitioning signed networks

Our previous works	Limitations	How to deal with?
1. Relevance of Negative Links in Graph Partitioning [Mendonça et al., 2015] 	• uncomplete data 	
2. Brazilian Congress structural balance analysis [Levorato and Frota, 2017] 	• fully-connected networks  	
	• interpretation problems  	

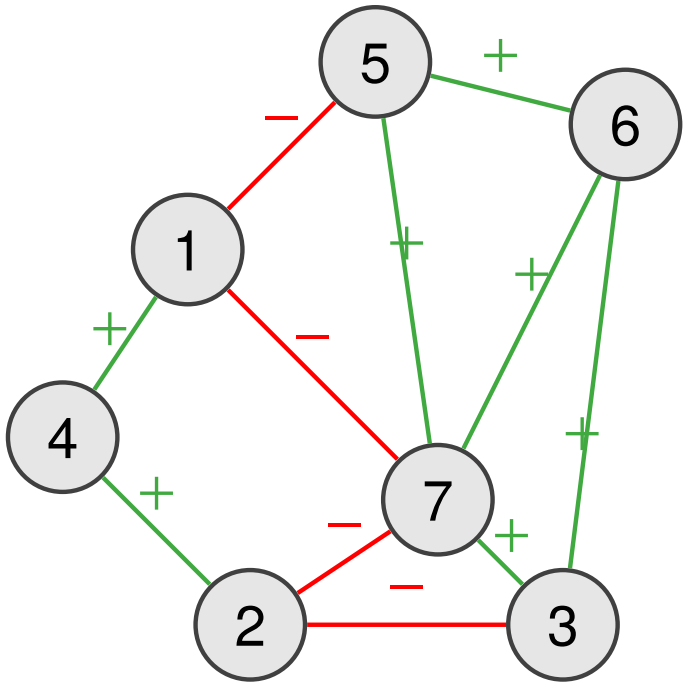
Context

- Partitioning European Parliament (EP) vote networks
  - Extracting signed networks
  - Partitioning signed networks

Our previous works	Limitations	How to deal with?
1. Relevance of Negative Links in Graph Partitioning [Mendonça et al., 2015] 	• uncomplete data 	• changing data source
2. Brazilian Congress structural balance analysis [Levorato and Frota, 2017] 	• fully-connected networks  	• filtering step
	• interpretation problems  	• deeper analysis on selected cases

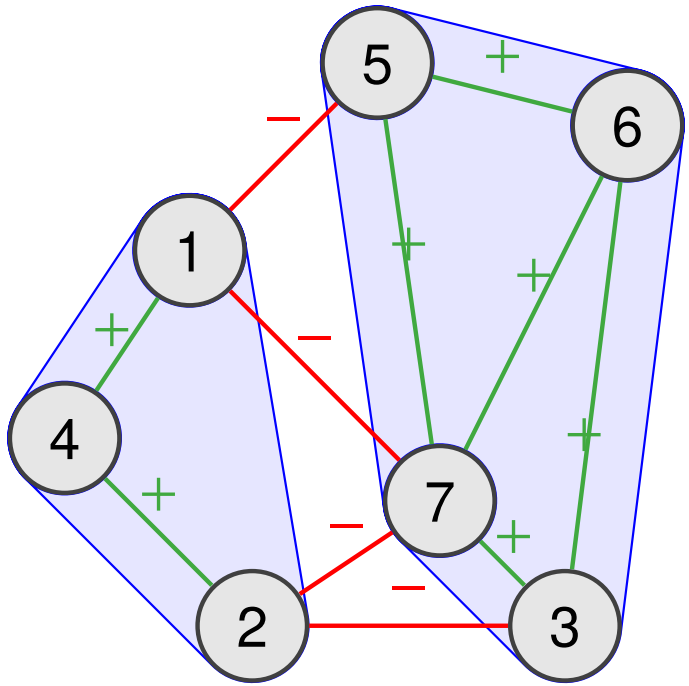
Structural Balance and Signed Graph Partitioning

- **Signed** graphs



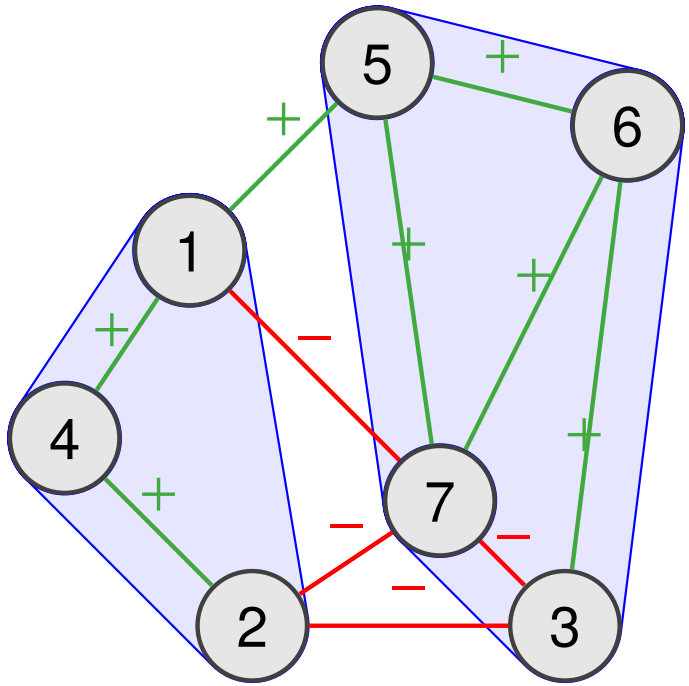
Structural Balance and Signed Graph Partitioning

- **Signed** graphs
- **Structural Balance**: Partitioning into two [Heider, 1946] or more [Davis, 1967] mutually hostile subgroups each having internal solidarity



Structural Balance and Signed Graph Partitioning

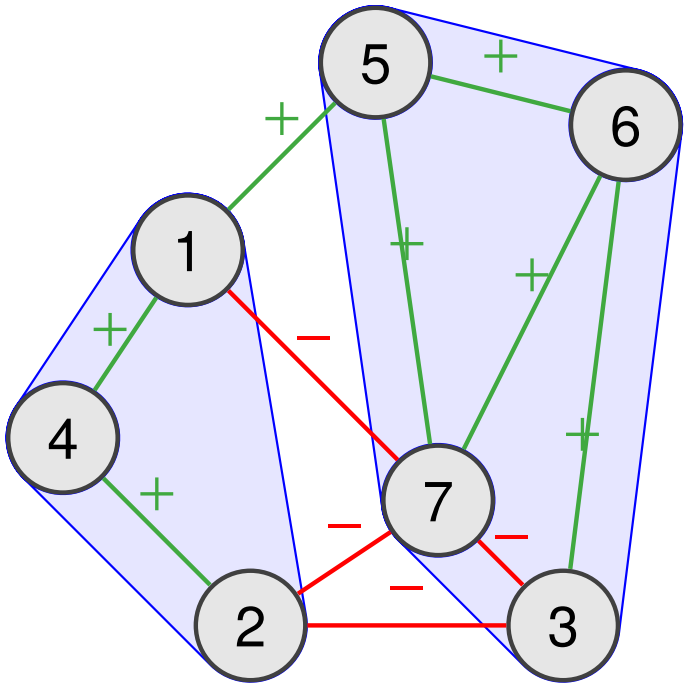
- **Signed** graphs
- **Structural Balance**: Partitioning into two [Heider, 1946] or more [Davis, 1967] mutually hostile subgroups each having internal solidarity
- Most real networks are not structurally balanced → need to measure graph **imbalance**



Measuring imbalance - CC problem

Definition

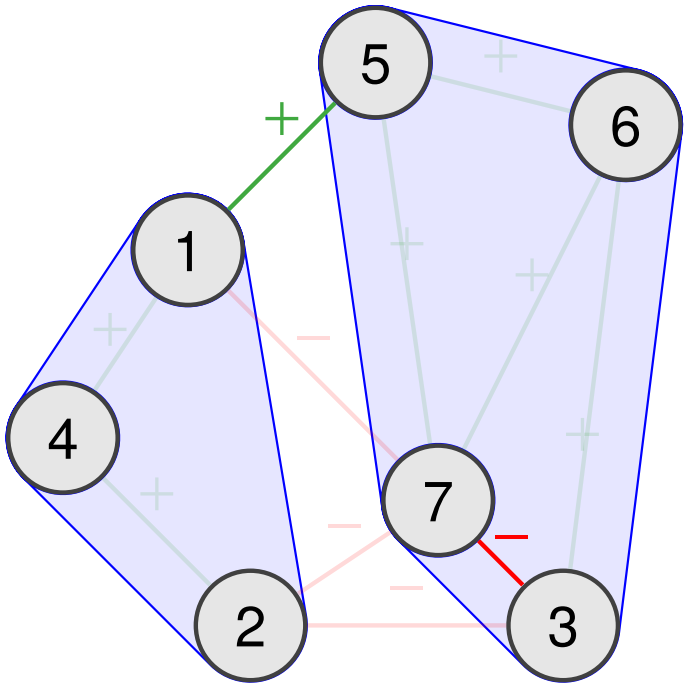
Consider a signed graph $G = (V, E, s)$ with a nonnegative weight associated with each $e \in E$. The Correlation Clustering (CC) problem is the problem of finding a partition P of V such that the imbalance $I(P)$ is minimized.



Measuring imbalance - CC problem

Definition

Consider a signed graph $G = (V, E, s)$ with a nonnegative weight associated with each $e \in E$. The Correlation Clustering (CC) problem is the problem of finding a partition P of V such that the imbalance $I(P)$ is minimized.

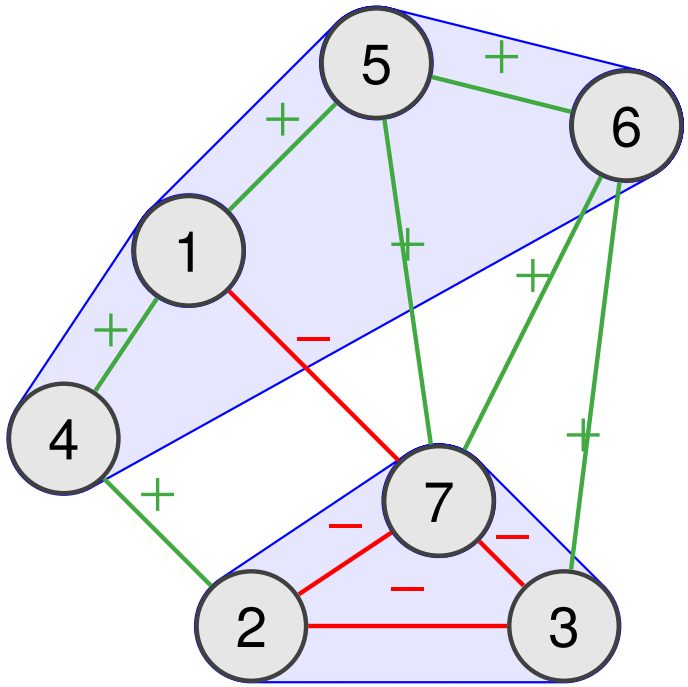


$$I(P) = 0 + 1 + 1 = 2$$

Measuring imbalance - RCC problem

Definition

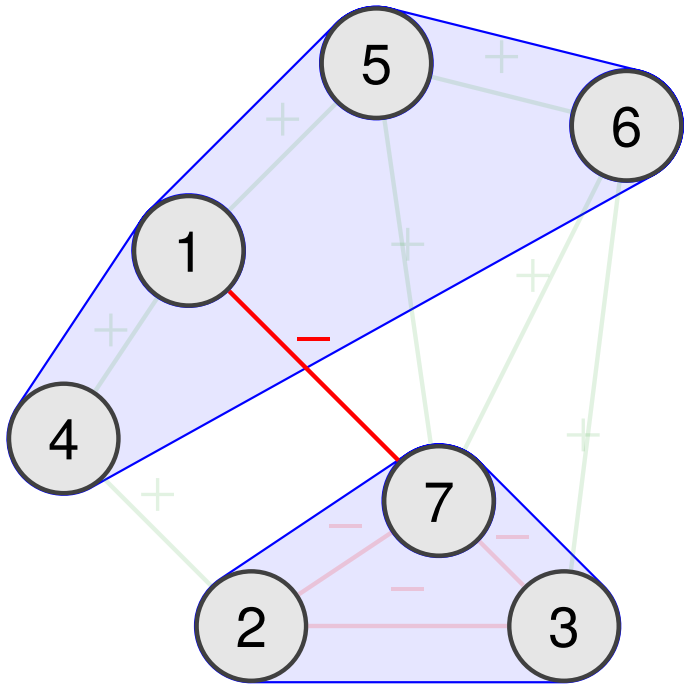
Consider a signed graph $G = (V, E, s)$ with a nonnegative weight associated with each $e \in E$ and an integer value satisfying $1 \leq k \leq n$. The Relaxed CC problem is the problem of finding a partition P of V, with at most k sets, such that the imbalance RI(P) is minimized.



Measuring imbalance - RCC problem

Definition

Consider a signed graph $G = (V, E, s)$ with a nonnegative weight associated with each $e \in E$ and an integer value satisfying $1 \leq k \leq n$. The Relaxed CC problem is the problem of finding a partition P of V, with at most k sets, such that the imbalance $RI(P)$ is minimized.



$$RI(P) = 0 + 0 + 1 = 1$$

Computational Experiments

- Dataset description, Construction of vote graphs and filtering step
- Evaluation of filtering step on graph structure and partitioning algorithms

Data Extraction

- Raw data (from itsyourparliament.eu):
  - Nature: Voting activity at the European Parliament
  - Period: 7^{th} term (June 2009–June 2014)
  - Size: 840 MEPs, 1426 documents, 21 topics

Data Extraction

- Raw data (from itsyourparliament.eu):
  - Nature: Voting activity at the European Parliament
  - Period: 7^{th} term (June 2009–June 2014)
  - Size: 840 MEPs, 1426 documents, 21 topics
- Voting Agreement Index (VAI) [Mendonça et al., 2015]:
  - Compares two MEPs

	For	Abstain	Against
For	+1	0	−1
Abstain	0	+1	0
Against	−1	0	+1

Data Extraction

- Raw data (from itsyourparliament.eu):
  - Nature: Voting activity at the European Parliament
  - Period: 7^{th} term (June 2009–June 2014)
  - Size: 840 MEPs, 1426 documents, 21 topics
- Voting Agreement Index (VAI) [Mendonça et al., 2015]:
  - Compares two MEPs

	For	Abstain	Against
For	+1	0	−1
Abstain	0	+1	0
Against	−1	0	+1

- Networks:
  - Nodes: Members of the European Parliament (MEPs)
  - Weighted: VAI values (document-wise agreement averaged over all documents) → Ranges from −1 to +1
  - Dimensions: member country × EP political group × time × topics
    - For instance, documents voted by French MEPs on Agriculture in 2012-2013

Data Extraction

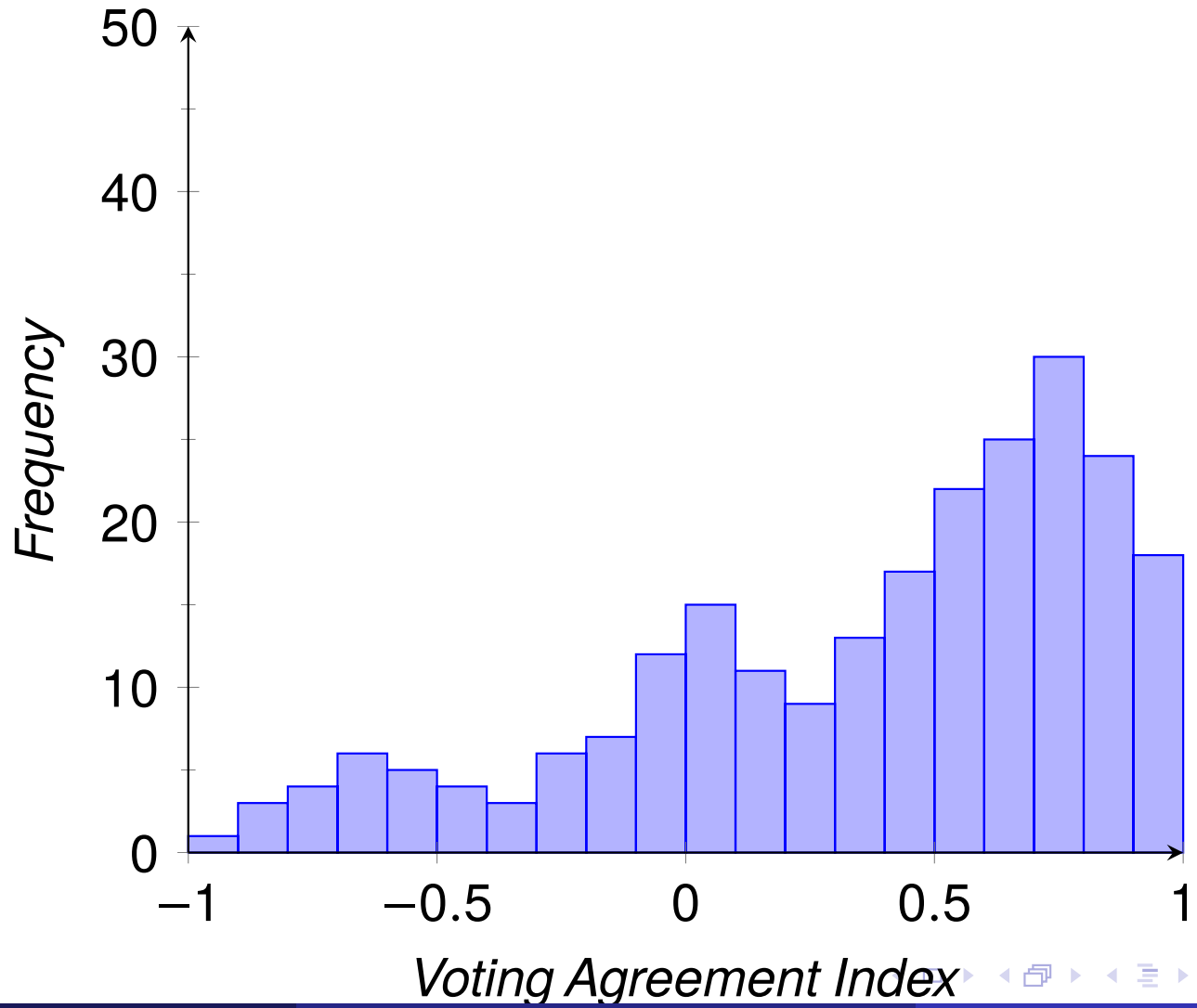
- Raw data (from itsyourparliament.eu):
  - Nature: Voting activity at the European Parliament
  - Period: 7^{th} term (June 2009–June 2014)
  - Size: 840 MEPs, 1426 documents, 21 topics
- Voting Agreement Index (VAI) [Mendonça et al., 2015]:
  - Compares two MEPs

	For	Abstain	Against
For	+1	0	−1
Abstain	0	+1	0
Against	−1	0	+1

- Networks:
  - Nodes: Members of the European Parliament (MEPs)
  - Weighted: VAI values (document-wise agreement averaged over all documents) → Ranges from −1 to +1
  - Dimensions: member country × EP political group × time × topics
    - For instance, documents voted by French MEPs on Agriculture in 2012-2013
- Filtering step

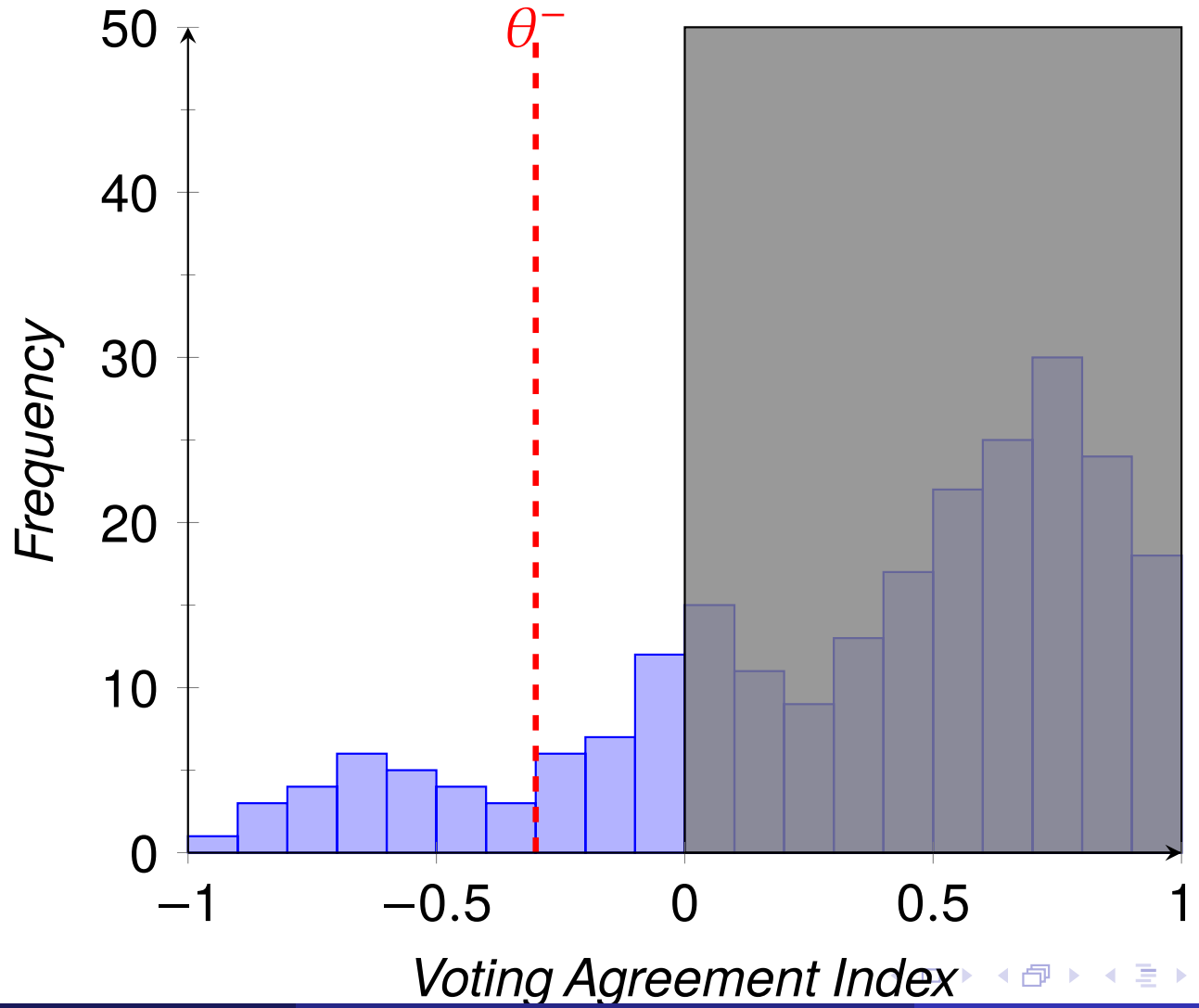
France, AGRI, 2012–2013: Filtering Step

- Filtering step aims at ignoring the values too close to zero, deemed non-significant



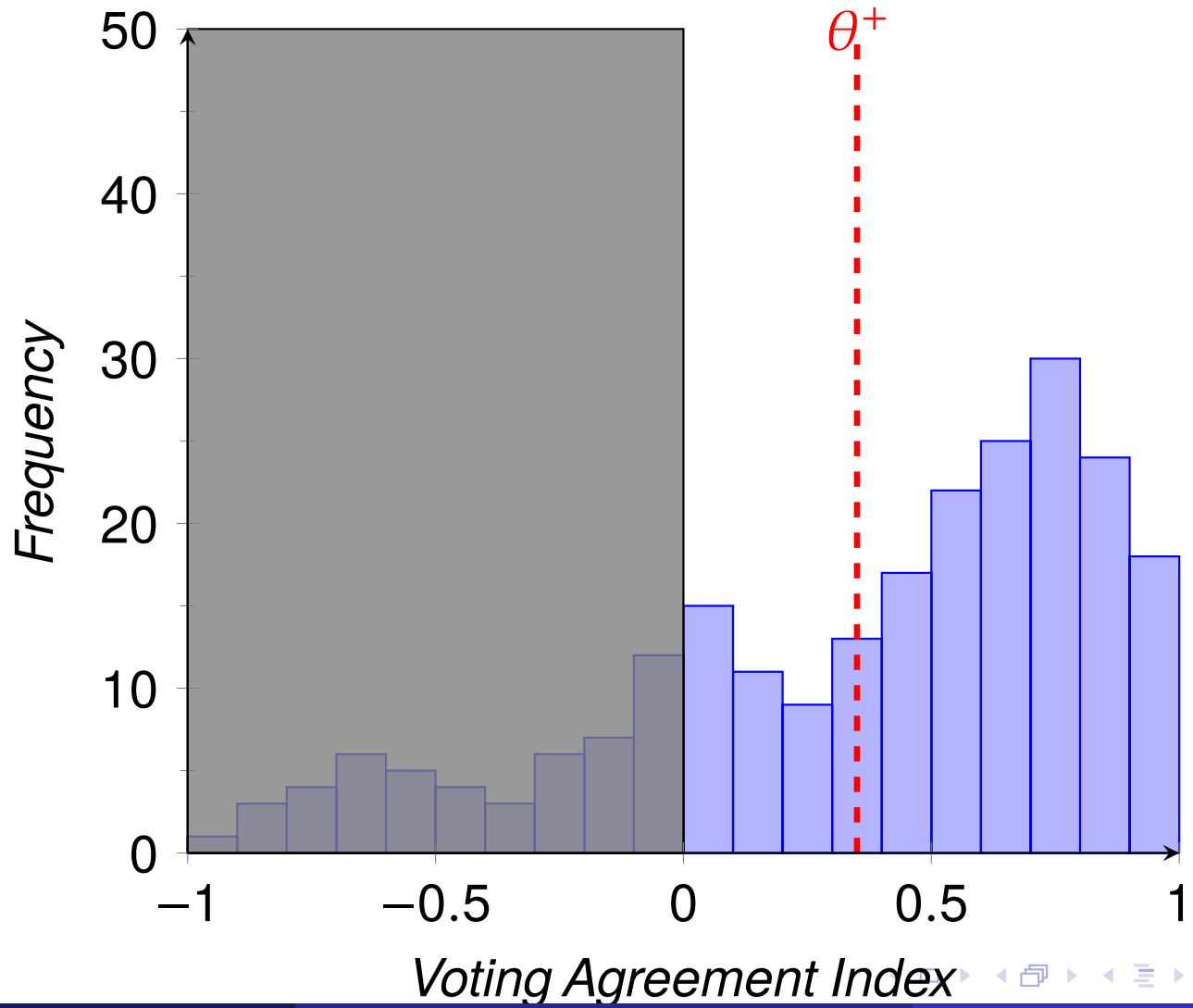
France, AGRI, 2012–2013: Filtering Step

- Filtering step aims at ignoring the values too close to zero, deemed non-significant



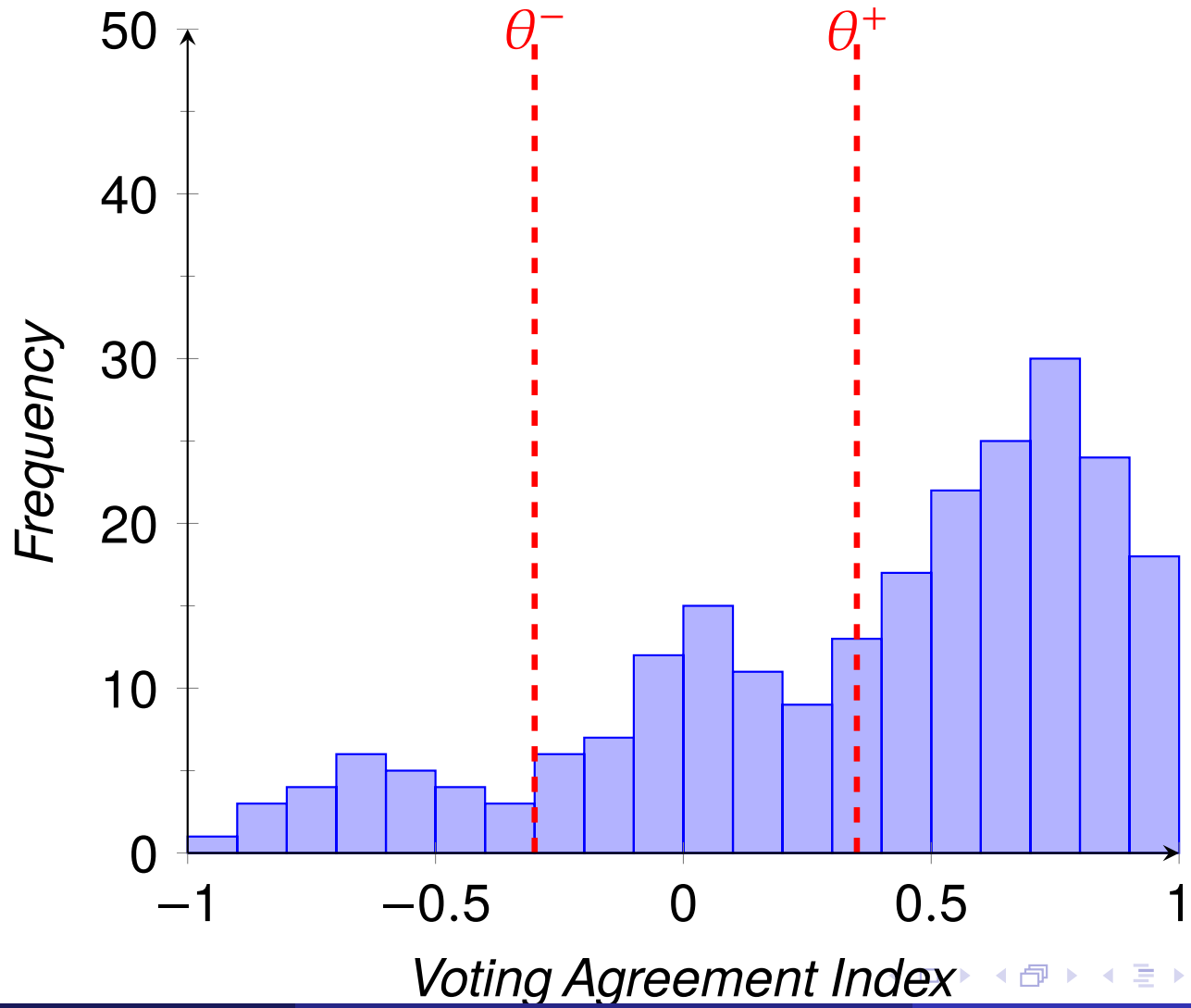
France, AGRI, 2012–2013: Filtering Step

- Filtering step aims at ignoring the values too close to zero, deemed non-significant



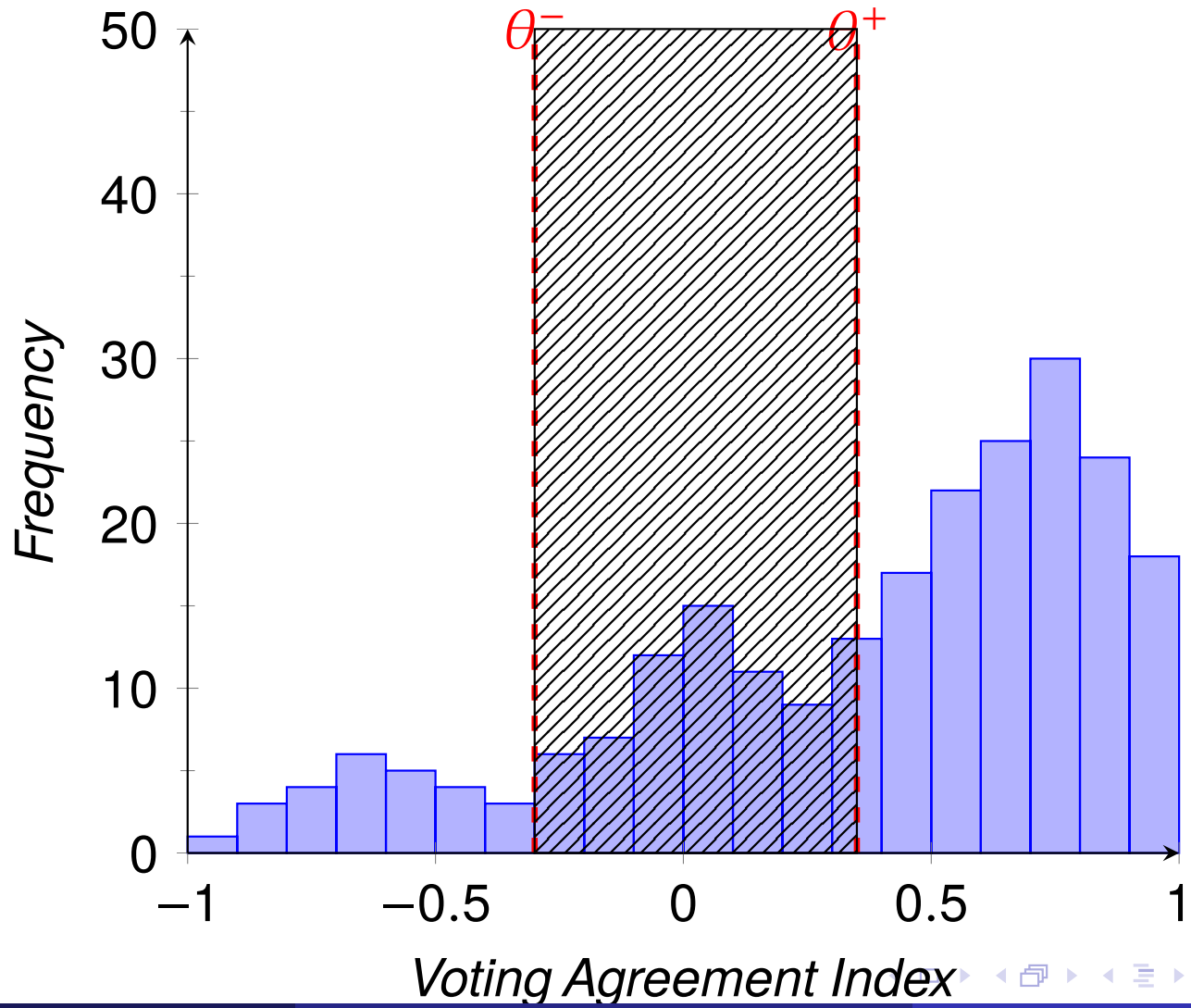
France, AGRI, 2012–2013: Filtering Step

- Filtering step aims at ignoring the values too close to zero, deemed non-significant



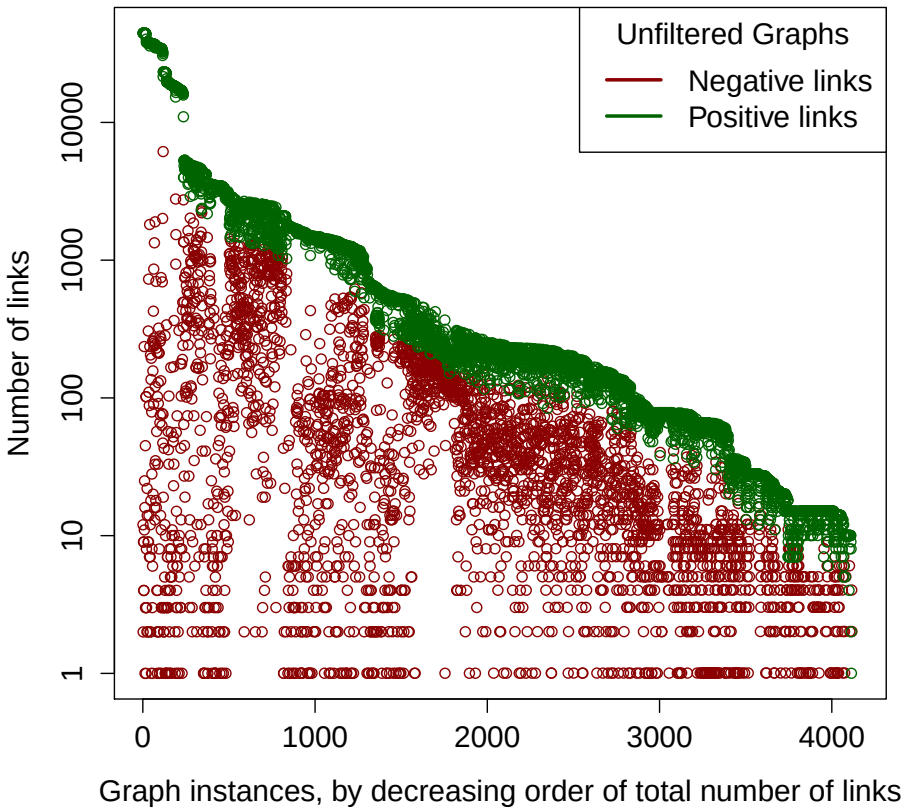
France, AGRI, 2012–2013: Filtering Step

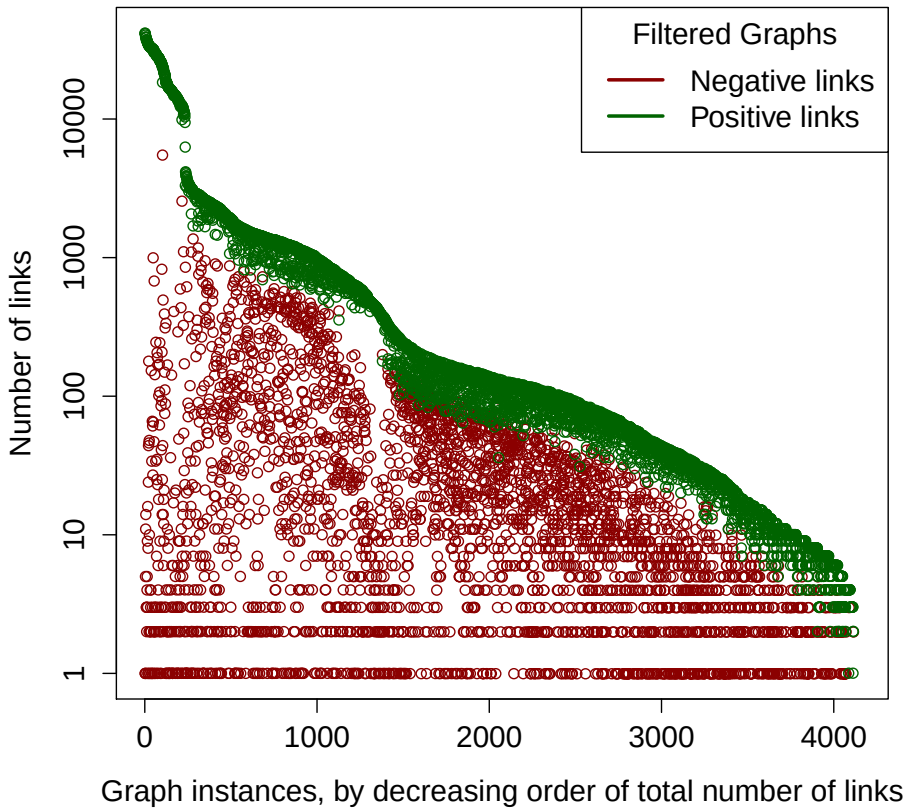
- Filtering step aims at ignoring the values too close to zero, deemed non-significant



Effect of the filtering on Graph Structure

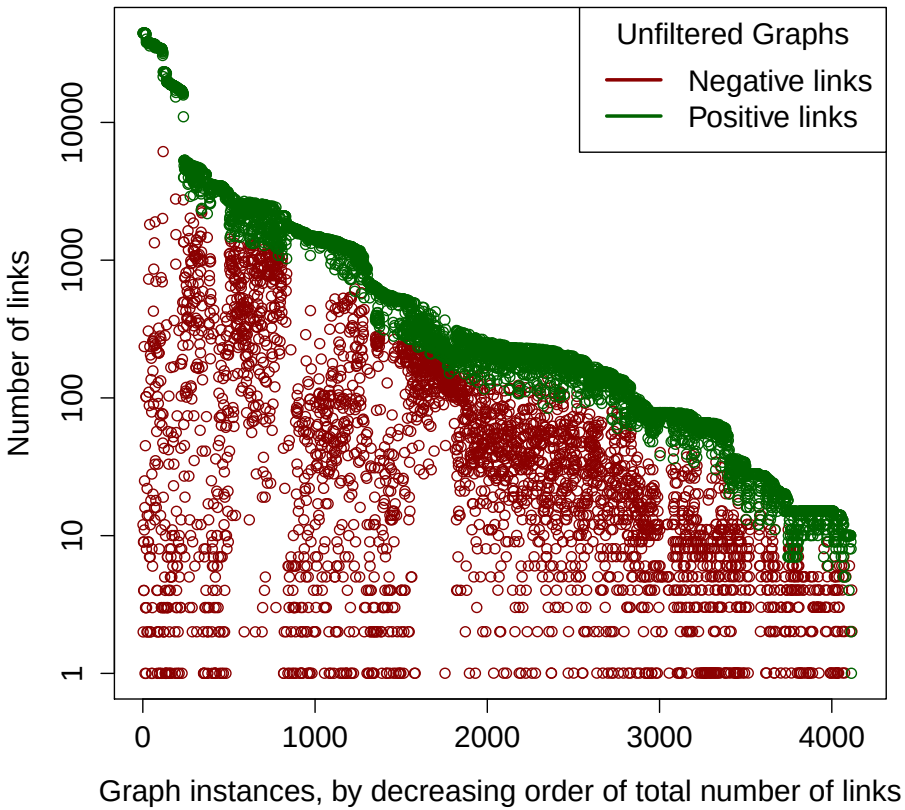
- Filtering removes in average: 43% of the links, 26% of the network weight

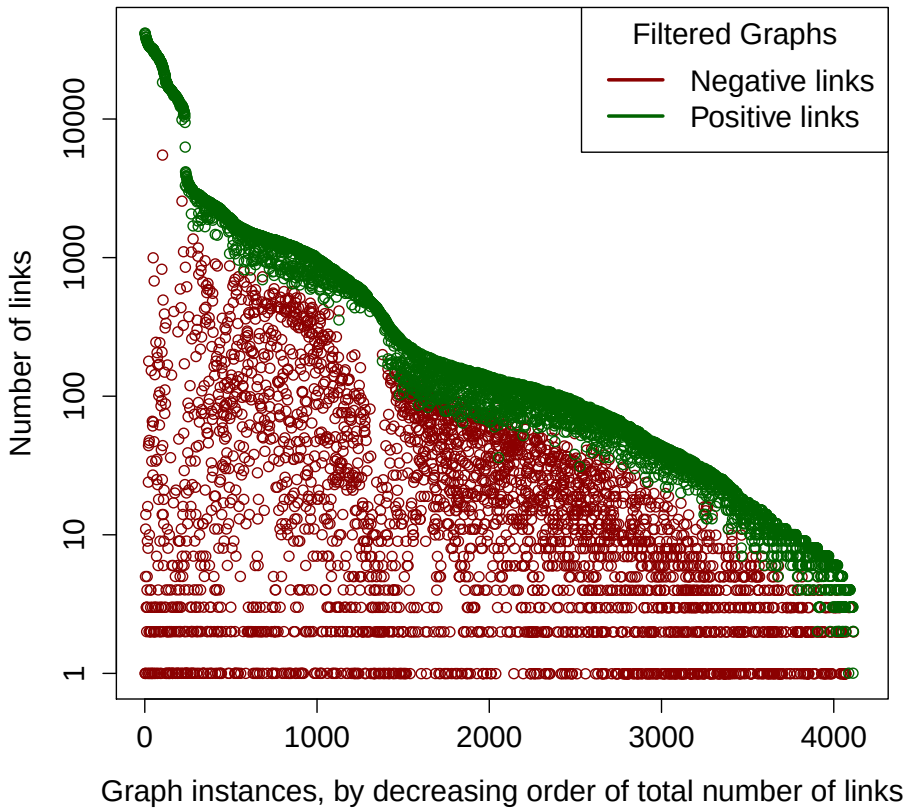




Effect of the filtering on Graph Structure

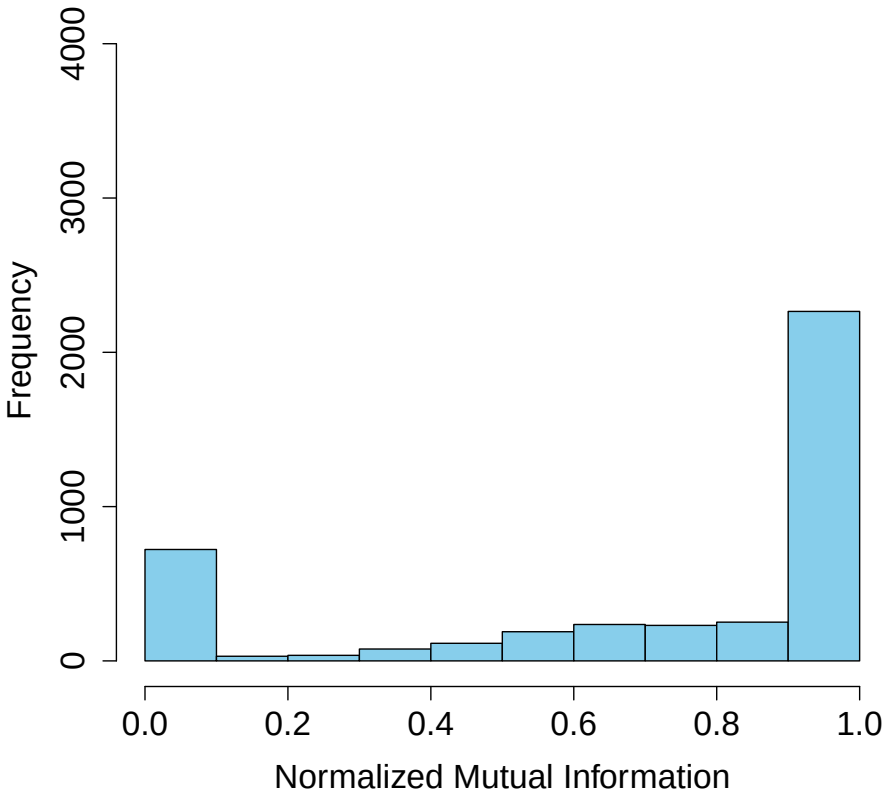
- Filtering removes in average: 43% of the links, 26% of the network weight
- Connectivity of filtered networks: 66% still connected, 23% divided into 2 components (but the giant component)

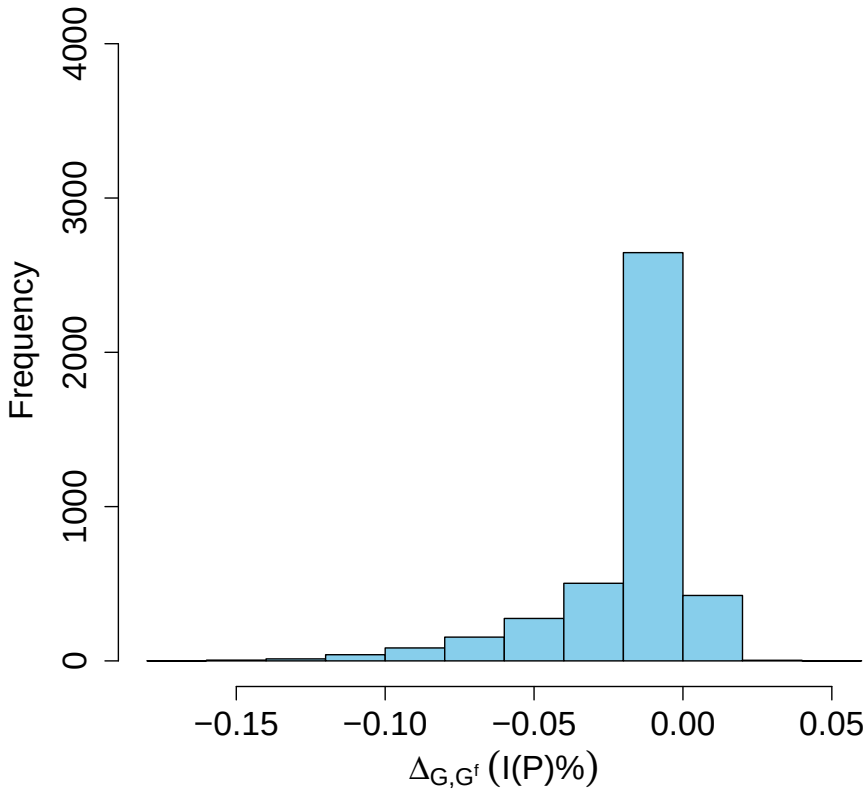




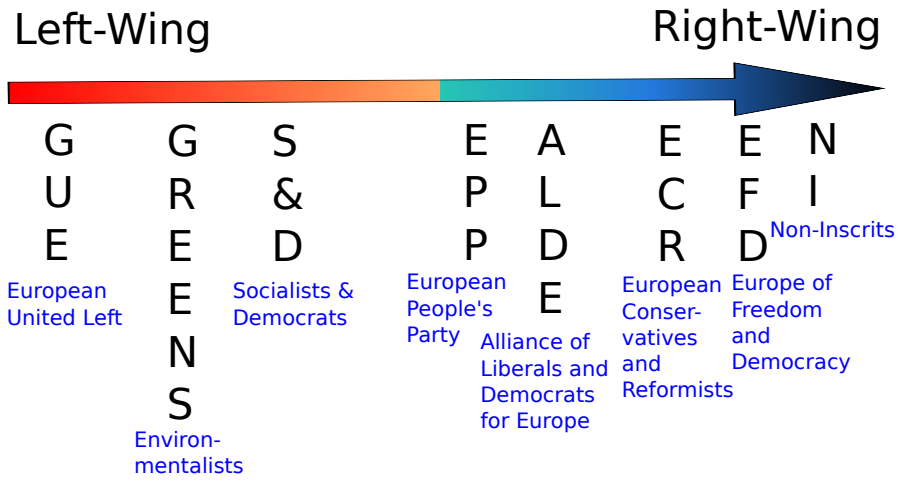
Effect of the filtering on Partitioning Algorithms

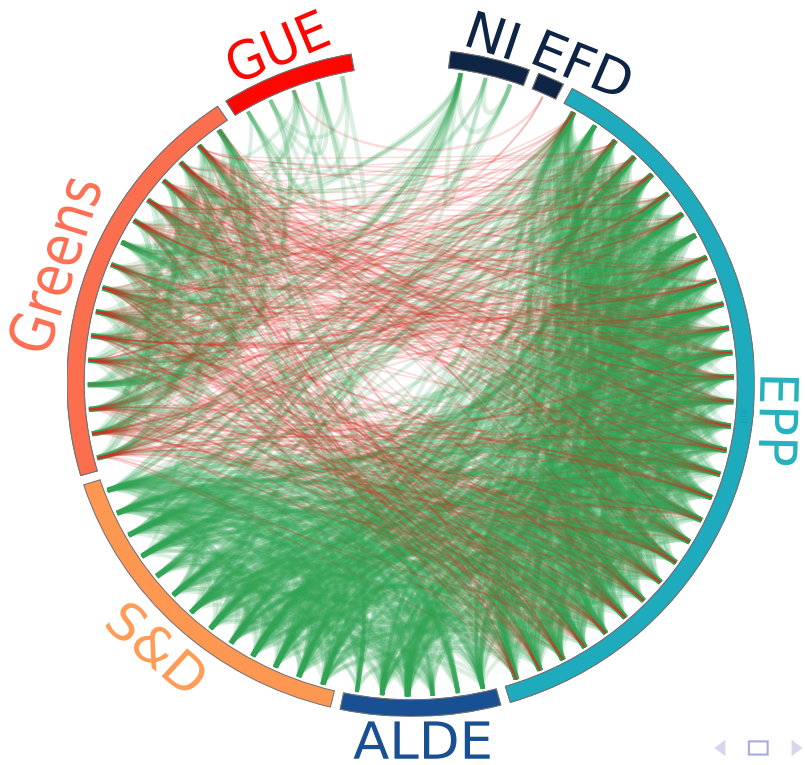
- Ex-CC before vs. after filtering
- The NMI is close to 1 for most instances ($\geq$0.8 for 61% of them)
- The quality in terms of I(P)% does not change much with filtering





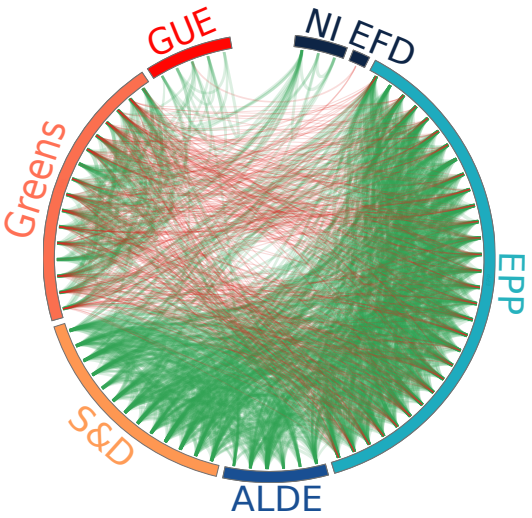
Interpretation of a vote network: France, AGRI, 2012–13





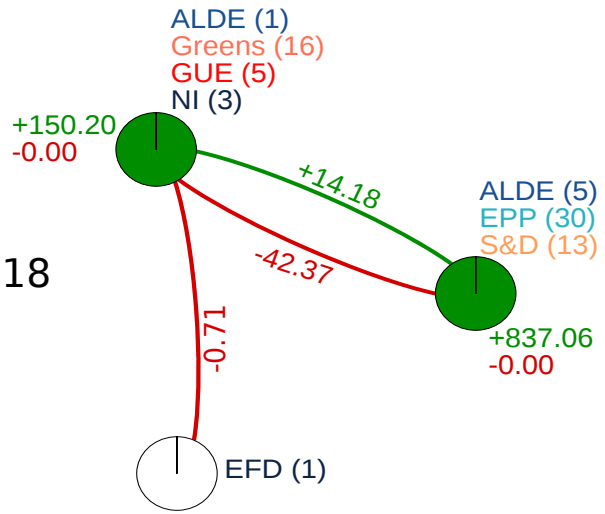
Interpretation of a vote network: France, AGRI, 2012–13

a)



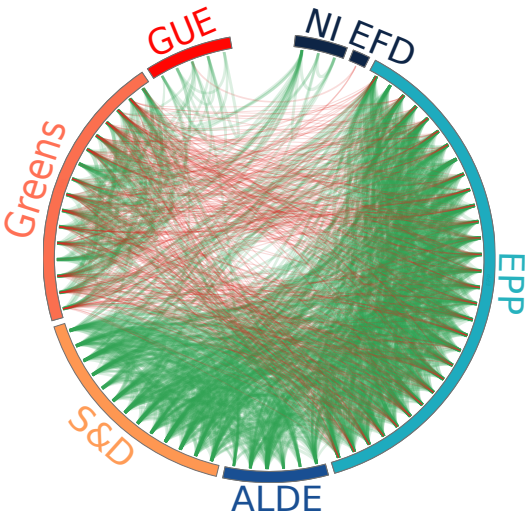
b) CC

I(P)=14.18



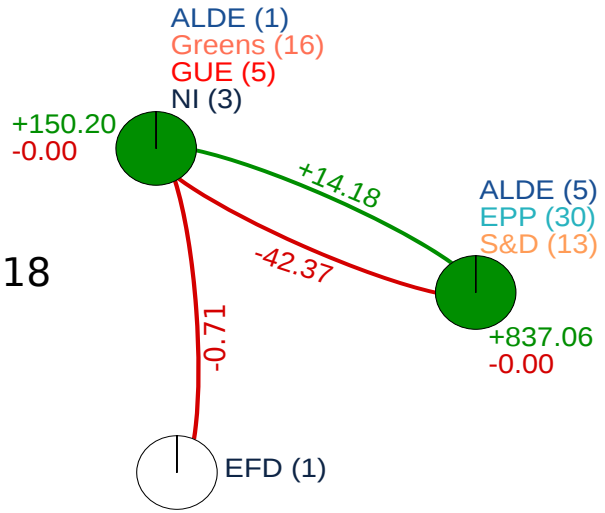
Interpretation of a vote network: France, AGRI, 2012–13

a)



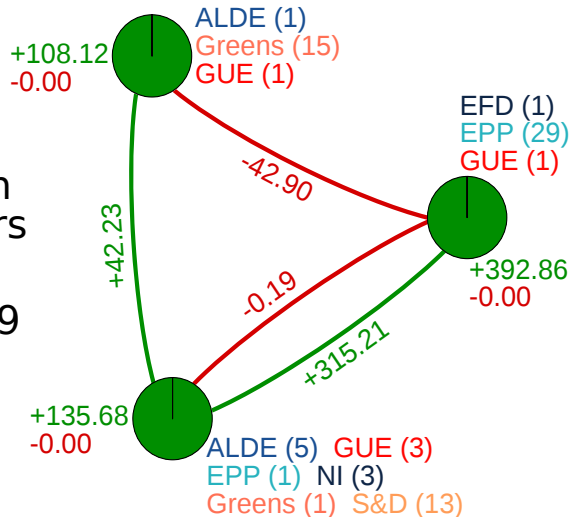
b) CC

I(P)=14.18



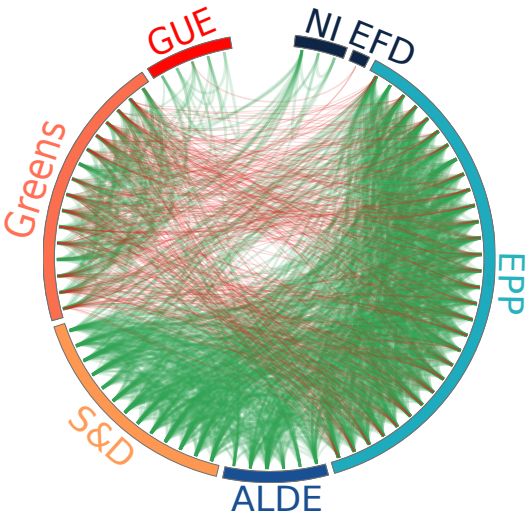
c) RCC with 3 clusters

RI(P)=0.19



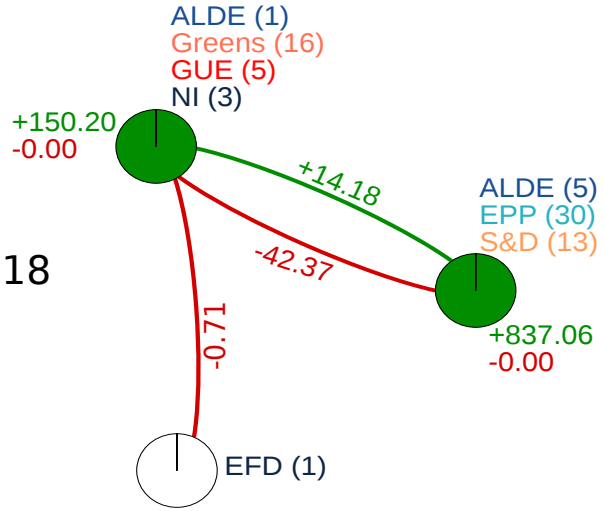
Interpretation of a vote network: France, AGRI, 2012–13

a)



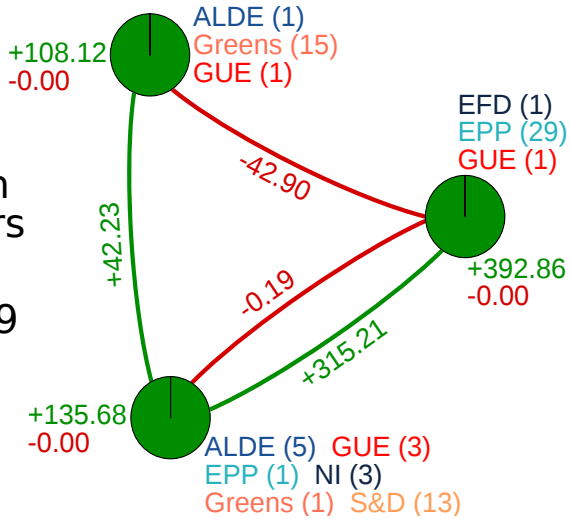
b) CC

I(P)=14.18



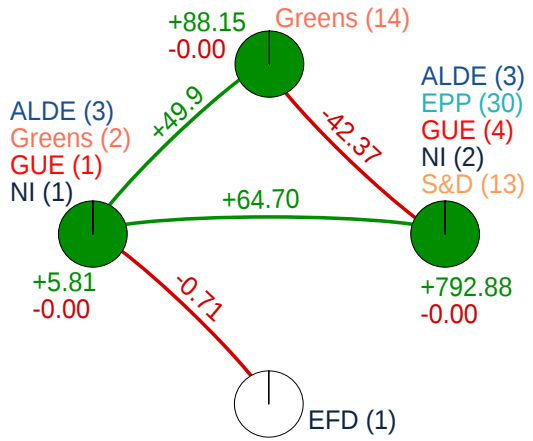
c) RCC with 3 clusters

RI(P)=0.19



d) RCC with 4 clusters

RI(P)=0



Conclusion & Further research

- Methodological:
  - Other Structural Balance problems
  - Detailed experiments based on randomly generated signed networks
- Application:
  - More complete interpretations via Manifestos of EU parties and the parties' policy preferences $\longrightarrow$ Comparative Manifestos Project (CMP).

Thank you for your attention!

References:

Davis, J. (1967).

Clustering and structural balance in graphs.

Human Relations, 20:181–187.

Heider, F. (1946).

Attitudes and cognitive organization.

Journal of Psychology, 21:107–112.

Levorato, M. and Frota, Y. (2017).

Brazilian congress structural balance analysis.

Journal of Interdisciplinary Methodologies and Issues in Science, 2.

Mendonça, I., Figueiredo, R., Labatut, V., and Michelon, P. (2015).

Relevance of negative links in graph partitioning: A case study using votes from the european parliament.

In *2015 Second European Network Intelligence Conference, ENIC 2015, Karlskrona, Sweden, September 21-22, 2015*, pages 122–129.

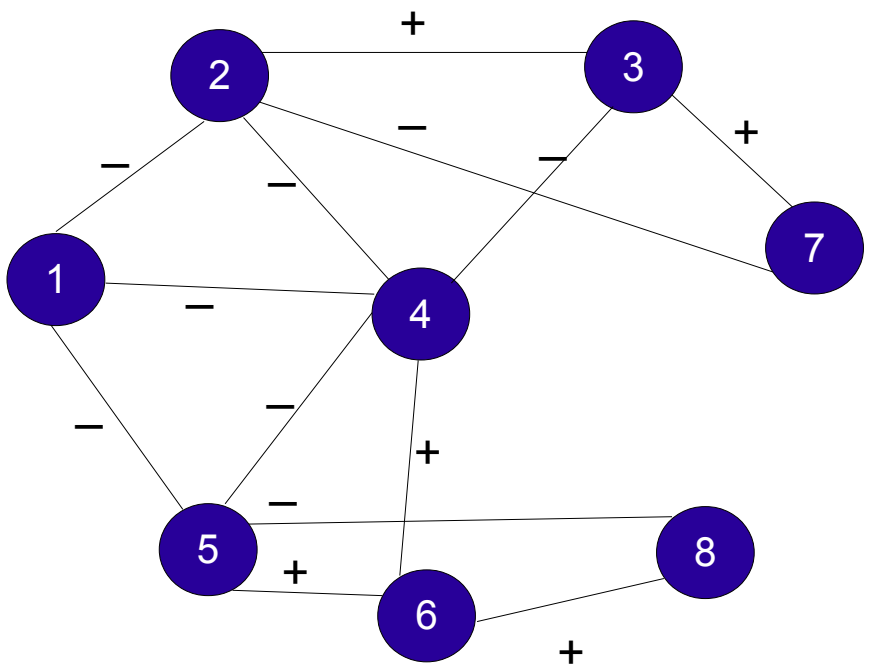
Iterated Local Search

It is comprised of 4 modules:

1. Constructive phase;
2. Local search;
3. Perturbation;
4. Acceptance criterion.

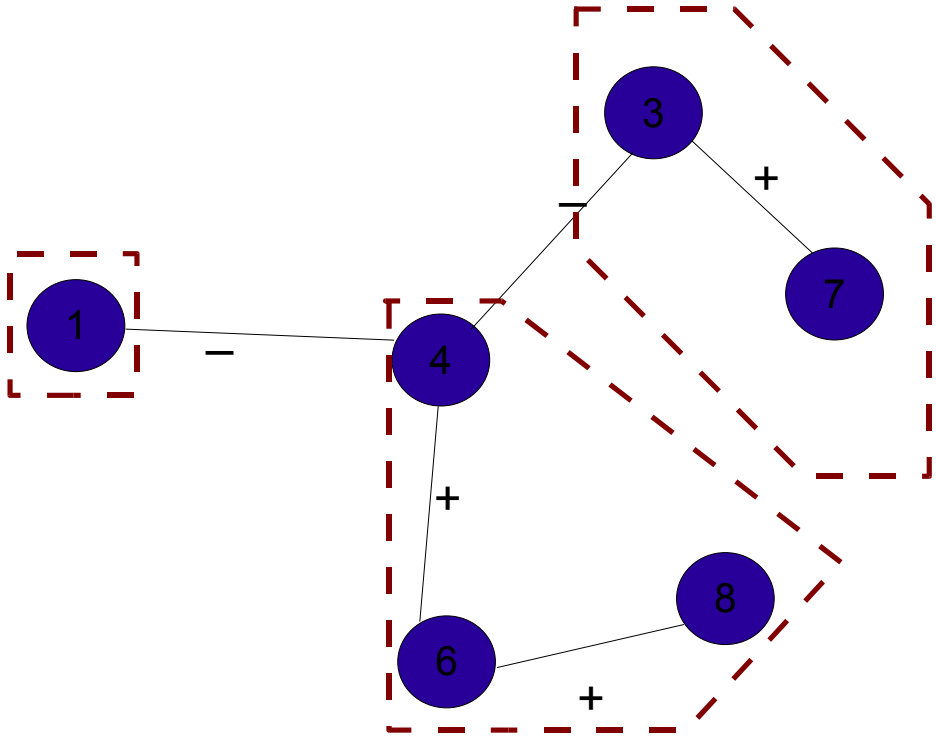
Measure ↔ Graph optimization problem

$|V|$ in a k balanced subgraph ↔ Maximum k-balanced subgraph Problem



Measure ↔ Graph optimization problem

$|V|$ in a k balanced subgraph ↔ Maximum k-balanced subgraph Problem



Measure ↔ Graph optimization problem

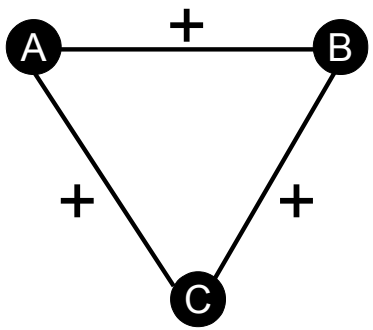
Definition

Consider a signed graph $G = (V, E, s)$ and an integer value satisfying $1 \leq k \leq n$. The Maximum k-Balanced Subgraph problem is the problem of finding a subgraph $H = (V', E', s)$ of G such that H is k-balanced and maximizes the cardinality of the vertex set V'.

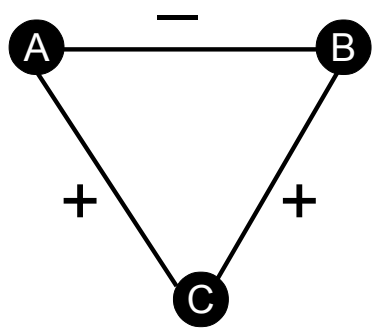
Structural Balance

[Heider, 1946]:

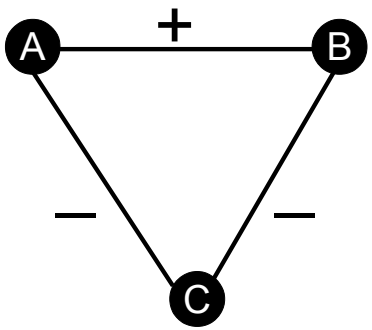
- People strive for cognitive balance in their network of likes and dislikes.



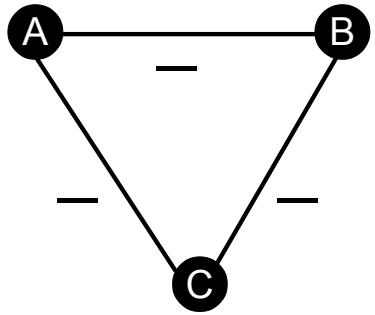
(a) Balanced



(b) Not balanced



(c) Balanced

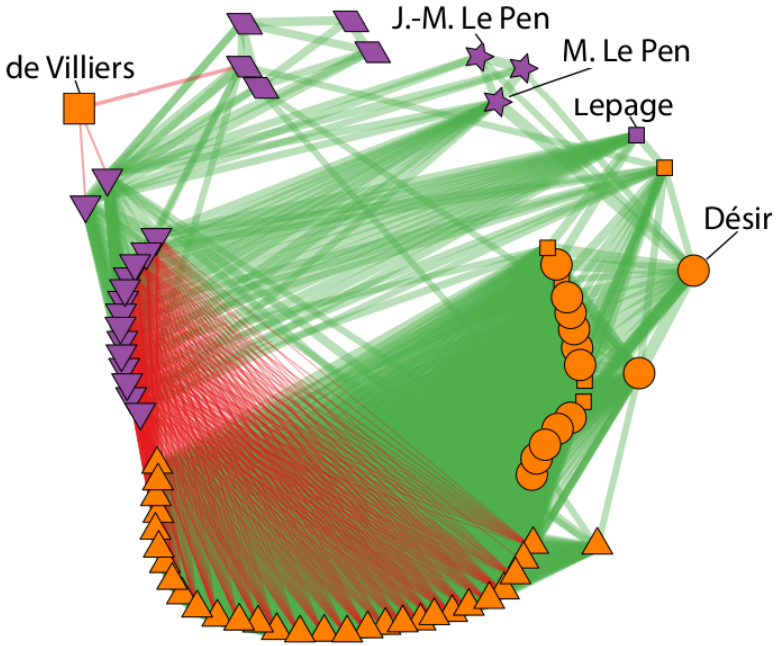


(d) Not balanced

Comparison between a community detection and signed graph partitioning methods

ILS-CC:

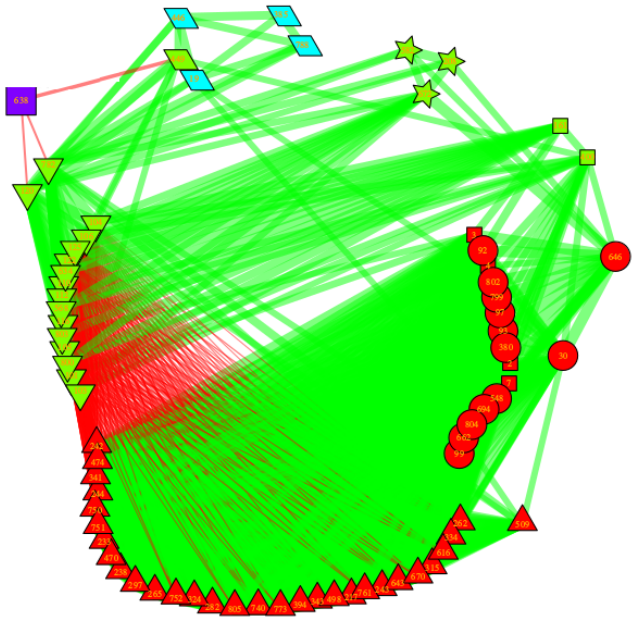
- Method designed for signed graphs
- Negative links taking into account



imbalance (count) = 14.18

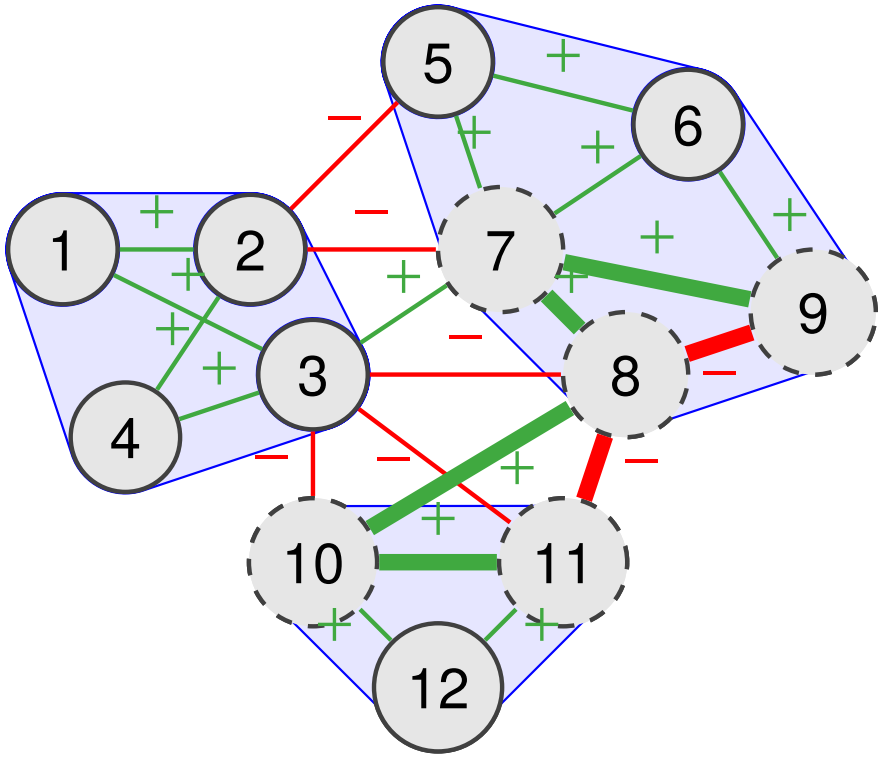
Infomap:

- Method designed for unsigned graphs
- Negative links not taking into account

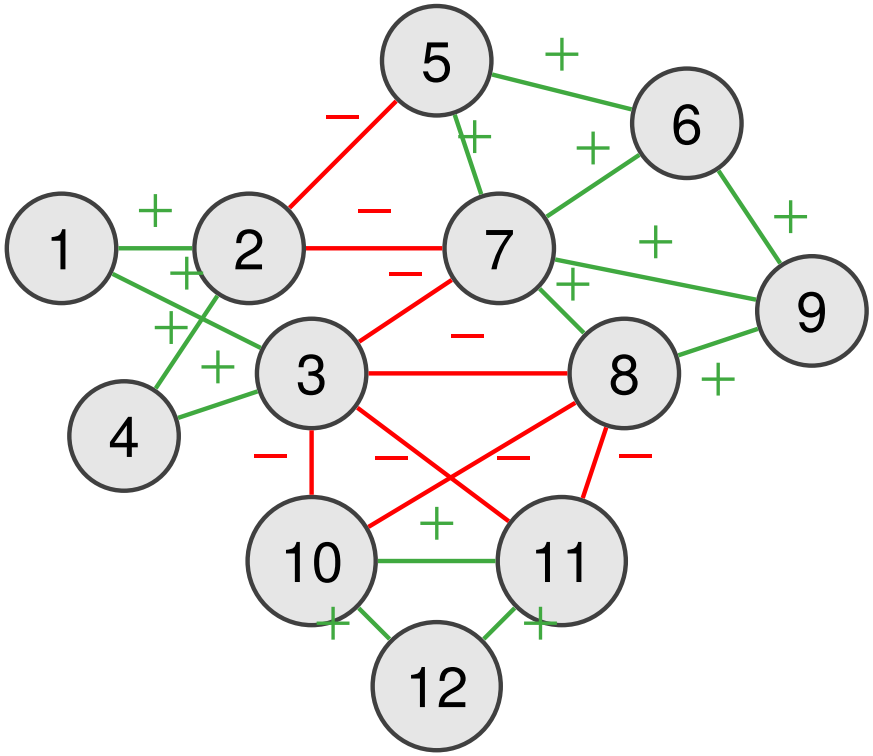


imbalance (count) = 21.70

Strucutural Balance (checking of local property)



An example of link prediction



An example of link prediction