Relevance of Negative Links in Graph Partitioning

Rosa Figueiredo[1]

Joint work with:
Lúcia Drummond[2] Yuri Frota[1] Vincent Labatut[1]
Mario Levorato[2] Israel Mendonça[1] Philippe Michelon[1]

1 LIA, University of Avignon, France.

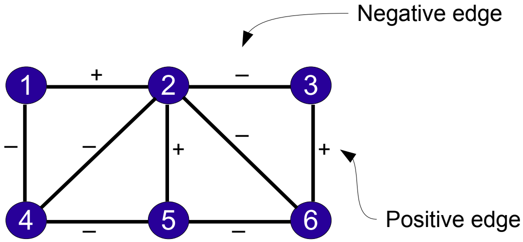
2 Dept. of Computer Science, Fluminense Federal University, Brazil.

Avignon, Mars 2016

Signed Graph

- $G = (V, E, s)$:
  - (V, E) is an undirected graph,
  - $s : E \to \{+, -\}$ is a function that assigns a sign to each edge in E.

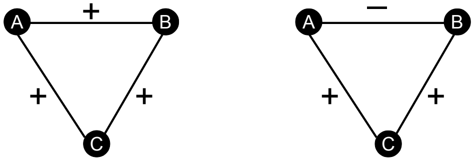
- E^-: set of negative edges.
  E^+: set of positive edges.



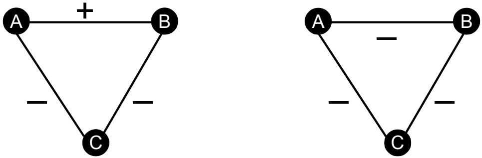
Structural Balance

[Heider, 1946]:

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



(a) Balanced (b) Not balanced

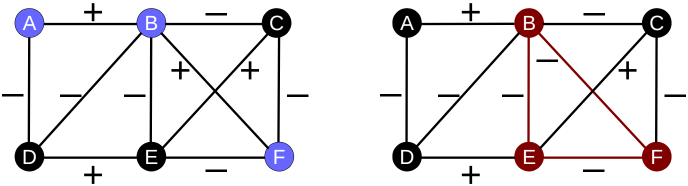


(c) Balanced (d) Not balanced

Structural Balance

[Cartwright and Harary, 1956]:

- The group can be partitioned into two mutually antagonistic subgroups each having internal solidarity.



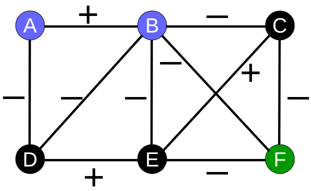
(a) Balanced signed graph: $S = \{A,B,F\}$ and $S' = \{C,D,E\}$

(b) Not balanced signed graph

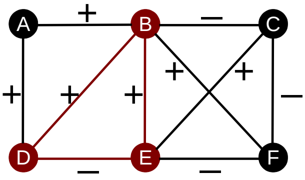
Structural Balance

[Davis, 1967]:

- Balanced social group = Clusterable signed graph.
- Two or more mutually antagonistic subgroups each having internal solidarity.



(a) Clusterable signed graph:
S_1 = {A,B}, S_2 = {C,D,E} e S_3 = {F}



(b) Not clusterable

Structural Balance

- Applications:
  - Social networks: [Doreian and Mrvar, 1996], [Leskovec et al., 2010], [Facchetti et al., 2011], [Srinivasan, 2011]...
  - Efficient document classification: [Bansal et al., 2002], [Zhang et al., 2008].
  - Financial networks: [Harary et al., 2003], [Huffner et al., 2010].
  - Biological networks: [DasGupta et al., 2007], [Huffner et al., 2010].

Structural Balance

- Applications:
  - Social networks: [Doreian and Mrvar, 1996], [Leskovec et al., 2010], [Facchetti et al., 2011], [Srinivasan, 2011]...
  - Efficient document classification: [Bansal et al., 2002], [Zhang et al., 2008].
  - Financial networks: [Harary et al., 2003], [Huffner et al., 2010].
  - Biological networks: [DasGupta et al., 2007], [Huffner et al., 2010].

- Most of the signed networks are not balanced!

- **How to evaluate balance/imbalance in a signed network?**

How to evaluate balance/imbalance in a signed network?

Solving a NP-hard combinatorial optimization problem:

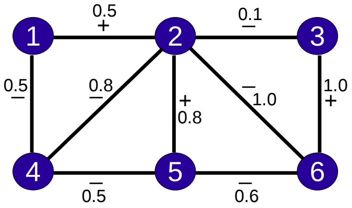
- Computing the line index of balance
  [Facchetti et al., 2011].
- Maximum balanced subgraph problem
  [Figueiredo and Frota, 2012].
- Correlation Clustering (CC) problem
  [Doreian and Mrvar, 1996].
- Relaxed Correlation Clustering problem
  [Doreian and Mrvar, 2009, Figueiredo and Moura, 2013].

Signed Graph

- $G = (V, E, s)$:
  - (V, E) is an undirected graph,
  - $s : E \to \{+, -\}$ is a function that assigns a sign to each edge in E.

- E^-: set of negative edges.
  E^+: set of positive edges.

- w_e: nonnegative edge weight associated with each $e \in E$.



Correlation Clustering Problem

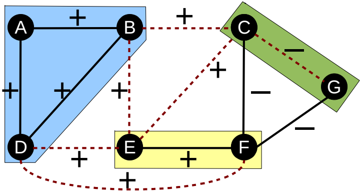
- Partition $P = \{S_1, S_2, \ldots, S_l\}$ of V

$$\Omega^+(S_i, S_j) = \sum_{e \in E^+ \cap E[S_i:S_j]} w_e$$

$$\Omega^-(S_i, S_j) = \sum_{e \in E^- \cap E[S_i:S_j]} w_e$$

- Imbalance

$$I(P) = \sum_{1 \le i \le l} \Omega^-(S_i, S_i) + \sum_{1 \le i < j \le l} \Omega^+(S_i, S_j).$$



$w_e = 1, \forall e \in E$. $I(P) = 1 + 5 = 6$.

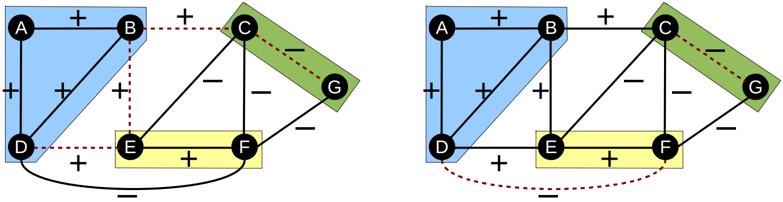
Correlation Clustering Problem

Definition

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

- [Doreian and Mrvar, 2009]:
  - revisited the definition of imbalance,
  - includes mediation between mutually hostile partitions,
  - includes internal subgroup hostility.

$$w_e = 1, \forall e \in E.$$



$I(P) = 1 + 3 = 4$ $\qquad$ $RI(P) = 1 + 1 = 2$

Relaxed Correlation Clustering Problem

Definition

Consider a signed graph $G = (V, E, s)$ with a nonnegative weight associated with each $e \in E$. The relaxed correlation-clustering problem is the problem of finding a partition P of V such that the imbalance $RI(P)$ is minimized.

Correlation-Clustering Problem - ILP formulation

$$x_{ij} = \begin{cases} 0 & \text{if vertex } i \text{ and } j \text{ are in a common set,} \\ 1 & \text{otherwise.} \end{cases}$$

$$\begin{aligned}
\text{minimize} \quad & \sum_{(i,j)\in E^-} w_{ij}(1 - x_{ij}) + \sum_{(i,j)\in E^+} w_{ij}x_{ij} & & \\
\text{subject to} \quad & x_{ip} + x_{pj} \geq x_{ij}, & \forall\, i,p,j \in V, & \quad (1)\\
& x_{ij} = x_{ji}, & \forall\, i,j \in V, & \quad (2)\\
& x_{ij} \in \{0,1\}, & \forall\, i,j \in V. & \quad (3)
\end{aligned}$$

- Disadvantage: for larger instances ($n > 200$), the number of restrictions in the formulation grows and the solver is unable to find an optimal solution within the time limit.

CC Problem - Heuristics

Heuristics applied to the CC problem:

- [Elsner and Schudy, 2009]: VOTE-BOEM (Constructive)
- [Doreian and Mrvar, 1996, Batagelj and Mrvar, 2014]: Constructive + Local Search
- [Zhang et al., 2008]: Genetic algorithm
- [Drummond et al., 2013]: **GRASP**
- [Levorato et al., 2015]: **Iterated Local Search**.

CC Problem - Heuristics

Heuristics applied to the CC problem:

- [Elsner and Schudy, 2009]: VOTE-BOEM (Constructive)
- [Doreian and Mrvar, 1996, Batagelj and Mrvar, 2014]: Constructive + Local Search
- [Zhang et al., 2008]: Genetic algorithm
- [Drummond et al., 2013]: **GRASP**
- [Levorato et al., 2015]: **Iterated Local Search**.

Metaheuristics:

- Higher-level procedure designed to find, generate or select a lower-level heuristic (partial search algorithm) to solve an optimization problem.
- Iteratively try to improve a candidate solution with regard to a given measure of quality.

Iterated Local Search

Developed by [Lourenço et al., 2003], is comprised of 4 modules:

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

Experimental results: test instances

- 22 small-sized instances, frequently used in the **literature** of structural balance [Brusco, 2003, Doreian and Mrvar, 2009].
- random instances with $n \in \{100, 200, 300, 400, 600\}$, varying network density d and negative graph density $d^- = |E^-|/|E|$.
- 63 medium-sized social networks based on **United Nations General Assembly (UNGA)** annual voting records [Macon et al., 2012]:
  - Between 1946 and 2008;
  - $\approx$ 190 vertices.
- 10 larger signed networks (with n from 200 to 10000 vertices) extracted from the **Slashdot**[1] website.

[1] `http://www.slashdot.com` - from February 21 2009 with 82,144 vertices and 549,202 edges.

Experimental results - ILP×Heuristics

Random Instances with $|V| = 100$:

Instance			ILP		GRASP			ILS		
$\|E\|$	d	d^-	BestSol	Time	AvgI(P)	Gap%I(P)	AvgTime	AvgI(P)	Gap%I(P)	AvgTime
990	0.1	0.2	**198**	71.49	198	0.0%	17.98	198	0.0%	1.0
		0.5	292	1339.70	238	-18.49%	130.89	236.4	-19.04%	1.6
		0.8	**50**	308.74	73.2	46.40%	384.91	62.8	25.60%	2.8
1980	0.2	0.2	**396**	82.50	396	0.00%	16.12	396	0.00%	0.9
		0.5	780	933.03	586.8	-24.77%	249.92	589.6	-24.41%	1.9
		0.8	272	709.02	225.2	-17.21%	1044.48	216.4	-20.44%	5.5
4950	0.5	0.2	**990**	60.42	990	0.00%	18.06	990	0.00%	0.8
		0.5	2234	1267.70	1845.6	-17.39%	424.53	1851.2	-17.14%	3.0
		0.8	858	641.85	750	-12.59%	2973.79	741.6	-13.57%	9.8
7920	0.8	0.2	**1584**	33.16	1584	0.00%	19.78	1584	0.00%	1.0
		0.5	3624	1542.02	3134.4	-13.51%	591.94	3148.8	-13.11%	3.7
		0.8	1476	689.52	1324	-10.30%	3601.58	1311.6	-11.14%	12.7

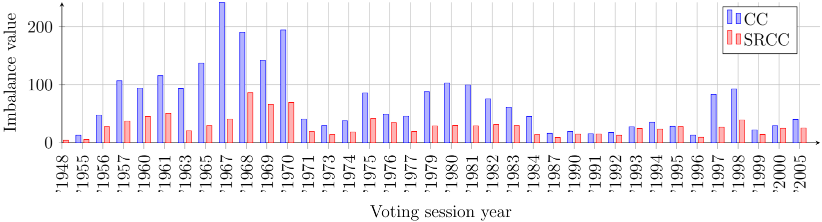
Experimental results: Slashdot test instances

Slashdot	GRASP		ILS		Gap	
n	BestSol	AvgTime	BestSol	AvgTime	% BestSol	AvgTime
200	45,0	1.39	45,0	2.05	0.00%	0.67
300	54,0	1.91	54,0	2.58	0.00%	0.67
400	57,0	2.63	57,2	3.77	0.35%	1.14
600	109,0	4.86	109,2	3.99	0.18%	-0.87
800	240,0	13.21	240,0	7.51	0.00%	-5.71
1000	600,0	23.69	600,0	12.91	0.00%	-10.79
2000	2186,0	232.48	2187,2	47.80	0.05%	-184.68
4000	6202,6	1415.45	6213,0	371.06	0.17%	-1044.39
8000	16082,6	7030.32	16073,2	1699.38	-0.06%	-5330.93
10000	20594,6	7200.49	20594,8	2782.59	0.00%	-4417.90
Avg	-	1592.64	-	493.36	0.07%	-1099.28

Number of vertices: *n*;
BestSol: value of the best solution found within time limit;
AvgTime: average time spent on 5 independent executions of each heuristic;
Gap: difference between solution value or time, between ILS and GRASP.

Experimental results - CC×RCC

UNGA instances :



Experimental results - CC on UNGA instances

- **1946-1953:** USA and Cuba in the same group.
- **1954 and beyond:** Cuban revolution, countries in opposite clusters.

Experimental results - CC on UNGA instances

- **1946-1953:** USA and Cuba in the same group.
- **1954 and beyond:** Cuban revolution, countries in opposite clusters.

- **1962:** Cuban missil crisis - bipolarity evident during the Cold War

Cluster A	Cluster B
USA, most Latin American countries, Western Europe, Japan, Taiwan, India, Australia and other Pacific Countries	Russia, Cuba, Poland, Hungary, Czechoslovakia, Albania, Yugoslavia, Bulgaria, Ukraine and many African countries

Experimental results - CC on UNGA instances

- **1946-1953:** USA and Cuba in the same group.
- **1954 and beyond:** Cuban revolution, countries in opposite clusters.

- **1962:** Cuban missil crisis - bipolarity evident during the Cold War

Cluster A	**Cluster B**
USA, most Latin American countries, Western Europe, Japan, Taiwan, India, Australia and other Pacific Countries	Russia, Cuba, Poland, Hungary, Czechoslovakia, Albania, Yugoslavia, Bulgaria, Ukraine and many African countries

- **1974:** Apartheid - South Africa appears isolated.

Experimental results - CC on UNGA instances

- **1946-1953:** USA and Cuba in the same group.
- **1954 and beyond:** Cuban revolution, countries in opposite clusters.

- **1962:** Cuban missil crisis - bipolarity evident during the Cold War

Cluster A	Cluster B
USA, most Latin American countries, Western Europe, Japan, Taiwan, India, Australia and other Pacific Countries	Russia, Cuba, Poland, Hungary, Czechoslovakia, Albania, Yugoslavia, Bulgaria, Ukraine and many African countries

- **1974:** Apartheid - South Africa appears isolated.

- **2006-2008:** Gaza conflict - Israel and USA appear together, isolated inside a group, with the rest of the world in another group.

Experimental results - RCC on UNGA instances

- **1987:** First Intifada started

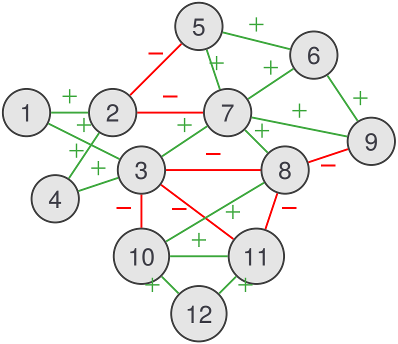
Cluster A	**Cluster B**	**Cluster C**
Canada, Ireland, Netherlands, Belgium, Luxembourg, France, Spain, Portugal, German Federal Republic, Italy, Norway, Denmark, Iceland, Japan, Australia and New Zealand	USA, Dominica, the UK and Israel	another one with 138 countries
+ internal: 100% **- external: 94%**		

Relevance of Negative Links

- Negative links are costly $\rightarrow$ relevance for graph partitioning?

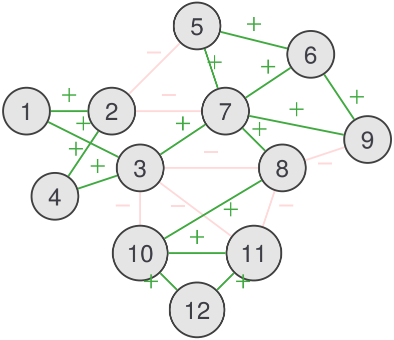
Relevance of Negative Links

- Negative links are costly → relevance for graph partitioning?
- Work by Esmailian *et al*. [Esmailian et al., 2014]



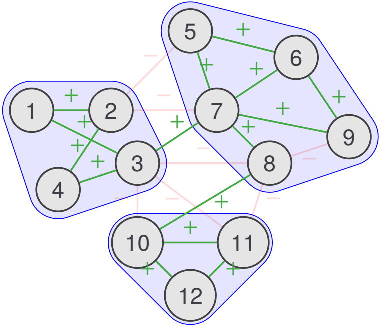
Relevance of Negative Links

- Negative links are costly → relevance for graph partitioning?
- Work by Esmailian *et al*. [Esmailian et al., 2014]
  - Community detection on positive links only



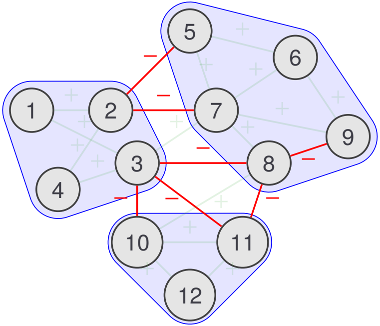
Relevance of Negative Links

- Negative links are costly → relevance for graph partitioning?
- Work by Esmailian *et al*. [Esmailian et al., 2014]
  - Community detection on positive links only



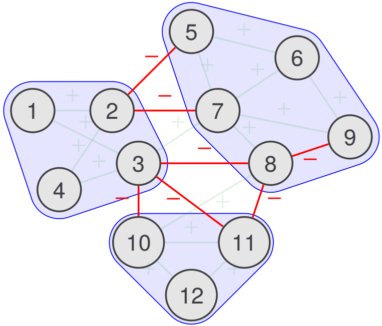
Relevance of Negative Links

- Negative links are costly → relevance for graph partitioning?
- Work by Esmailian *et al*. [Esmailian et al., 2014]
  - Community detection on positive links only
  - Study the community-wise location of negative links



Relevance of Negative Links

- Negative links are costly → relevance for graph partitioning?
- Work by Esmailian *et al*. [Esmailian et al., 2014]
  - Community detection on positive links only
  - Study the community-wise location of negative links
  - → Most are between communities or non-significant



Relevance of Negative Links

- Negative links are costly → relevance for graph partitioning?
- Work by Esmailian *et al*. [Esmailian et al., 2014]
  - Community detection on positive links only
  - Study the community-wise location of negative links
  - → Most are between communities or non-significant

- Limitations:
  - Only 2 datasets, both social networking services (Epinions and Slashdot)
  - Imbalance assessed only locally

Relevance of Negative Links

- Negative links are costly → relevance for graph partitioning?
- Work by Esmailian *et al*. [Esmailian et al., 2014]
  - Community detection on positive links only
  - Study the community-wise location of negative links
  - → Most are between communities or non-significant
- Limitations:
  - Only 2 datasets, both social networking services (Epinions and Slashdot)
  - Imbalance assessed only locally
- Proposed method:
  - Community detection by solving the correlation clustering problem
  - Consider a different dataset, modeling a different type of relationships

Data Extraction

- Raw data:
  - Nature: Voting activity at the European Parliament
  - Source: VoteWatch Europe
  - Period: 7th term (June 2009–June 2014)
  - Size: 840 MEPs, 1426 documents, 21 topics

Data Extraction

- Raw data:
  - Nature: Voting activity at the European Parliament
  - Source: VoteWatch Europe
  - Period: 7th term (June 2009–June 2014)
  - Size: 840 MEPs, 1426 documents, 21 topics
- Voting Agreement Index:
  - Compares two MEPs
  - Ranges from -1 to $+1$
  - Document-wise agreement averaged over all documents
    - Agreement: $+1$ (FOR VS. FOR, AGAINST VS. AGAINST)
    - Disagreement: -1 (FOR VS. AGAINST)
    - Undetermined: 0 (ABSTAIN/ABSENT VS. $*$)

Data Extraction

- Raw data:
  - Nature: Voting activity at the European Parliament
  - Source: VoteWatch Europe
  - Period: 7th term (June 2009–June 2014)
  - Size: 840 MEPs, 1426 documents, 21 topics
- Voting Agreement Index:
  - Compares two MEPs
  - Ranges from -1 to $+1$
  - Document-wise agreement averaged over all documents
    - Agreement: $+1$ (FOR VS. FOR, AGAINST VS. AGAINST)
    - Disagreement: -1 (FOR VS. AGAINST)
    - Undetermined: 0 (ABSTAIN/ABSENT VS. $*$)
- Networks:
  - Nodes: Members of the European Parliament (MEPs)
  - Weighted: voting agreement index values
  - Total: 264 (time $\times$ topics)

Partitioning Algorithms

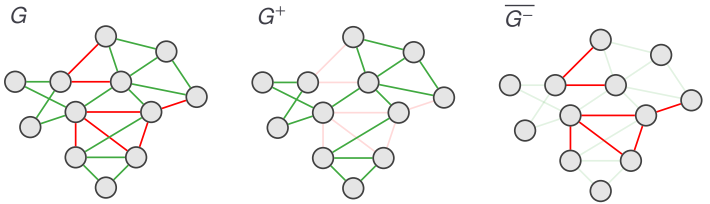
- Correlation clustering (G)
  - Parallel Iterated Local Search

Partitioning Algorithms

- Correlation clustering (G)
  - Parallel Iterated Local Search
- Community detection (G^+ and $\overline{G^-}$)
  - InfoMap [Rosvall and Bergstrom, 2008]
  - EdgeBetweenness [Newman and Girvan, 2004]
  - WalkTrap [Pons and Latapy, 2005]
  - FastGreedy [Clauset et al., 2004]

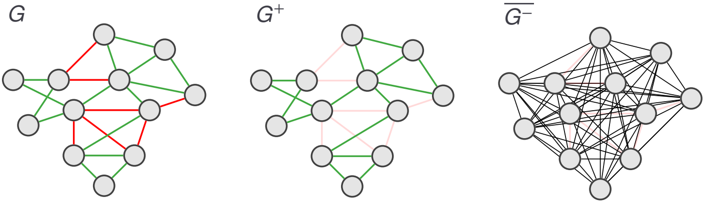
Partitioning Algorithms

- Correlation clustering (G)
  - Parallel Iterated Local Search
- Community detection (G^+ and $\overline{G^-}$)
  - InfoMap [Rosvall and Bergstrom, 2008]
  - EdgeBetweenness [Newman and Girvan, 2004]
  - WalkTrap [Pons and Latapy, 2005]
  - FastGreedy [Clauset et al., 2004]



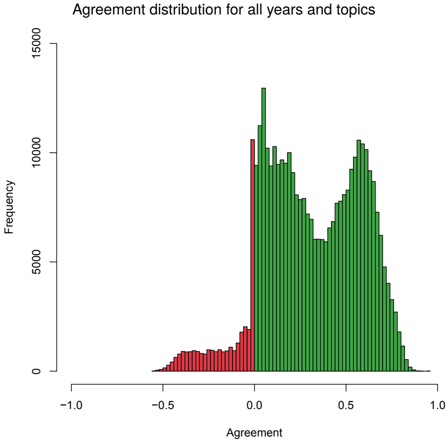
Partitioning Algorithms

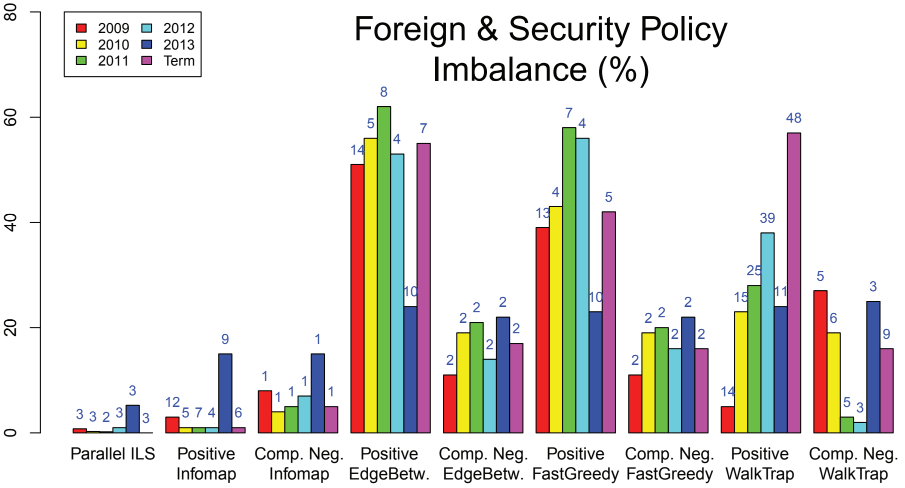
- Correlation clustering (G)
  - Parallel Iterated Local Search
- Community detection (G^+ and $\overline{G^-}$)
  - InfoMap [Rosvall and Bergstrom, 2008]
  - EdgeBetweenness [Newman and Girvan, 2004]
  - WalkTrap [Pons and Latapy, 2005]
  - FastGreedy [Clauset et al., 2004]



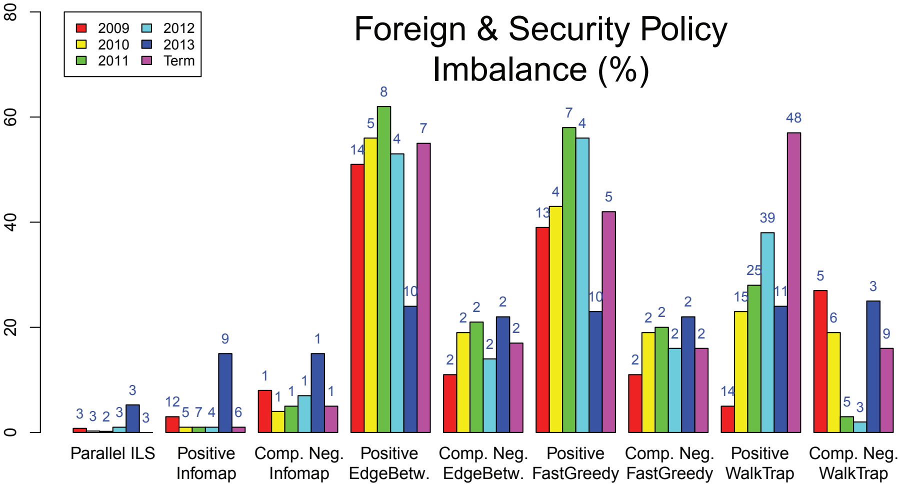
Extracted Networks

- Same observations for all topics/durations
- Positive side: bimodal distribution
  - Left peak: certain MEPs are frequently absent
  - Right peak: most MEPs often vote similarly
- Negative side: less extreme values
  - Clear majority, in average




Foreign & Security Policy
Imbalance (%)
2009
2010
2011
2012
2013
Term
0
20
40
60
80
Parallel ILS
Positive Infomap
Comp. Neg. Infomap
Positive EdgeBetw.
Comp. Neg. EdgeBetw.
Positive FastGreedy
Comp. Neg. FastGreedy
Positive WalkTrap
Comp. Neg. WalkTrap

Partition Comparison



Partition comparison: near-zero NMI

Conclusion

- Considering negative links on our dataset leads to:
  - Lower imbalance (at least 3 times better)
  - Only InfoMap outputs CC-like results (imbalance)
  - Different partitions (fewer clusters)

- Contradiction with Esmailian *et al*.'s conclusions

- Perspectives:
  - Consider more data, different types of networks (collection)
  - Exhaustive exploration of vote-based extraction methods
  - Political interpretation of the VoteWatch results

References:

Bansal, N., Blum, A., and Chawla, S. (2002).
Correlation clustering.
In *Proceedings of the 43rd annual IEEE symposium of foundations of computer science*, pages 238–250, Vancouver, Canada.

Batagelj, V. and Mrvar, A. (2014).
Pajek wiki.

Brusco, M. (2003).
An enhanced branch-and-bound algorithm for a partitioning problem.
British Journal of Mathematical and Statistical Psychology, 56:83–92.

Cartwright, D. and Harary, F. (1956).
Structural balance: A generalization of heiders theory.
Psychological Review, 63:277–293.

Clauset, A., Newman, M. E. J., and Moore, C. (2004).
Finding community structure in very large networks.
Physical Review E, 70(6):066111.

DasGupta, B., Encisob, G. A., Sontag, E., and Zhanga, Y. (2007).

Algorithmic and complexity results for decompositions of biological networks into monotone subsystems.
BioSystems, 90:161–178.

Davis, J. (1967).
Clustering and structural balance in graphs.
Human Relations, 20:181–187.

Doreian, P. and Mrvar, A. (1996).
A partitioning approach to structural balance.
Social Networks, 18:149–168.

Doreian, P. and Mrvar, A. (2009).
Partitioning signed social networks.
Social Networks, 31:1–11.

Drummond, L., Figueiredo, R., Frota, Y., and Levorato, M. (2013).
Efficient solution of the correlation clustering problem: An application to structural balance.
In *On the Move to Meaningful Internet Systems: OTM 2013 Workshops*, pages 674–683. Springer.

Elsner, M. and Schudy, W. (2009).
Bounding and comparing methods for correlation clustering beyond ilp.
In *ILP'09 Proceedings of the Workshop on Integer Linear Programming for Natural Language Processing*, pages 19–27.

Esmailian, P., Abtahi, S. E., and Jalili, M. (2014).
Mesoscopic analysis of online social networks: The role of negative ties.
Physical Review E, 90(4):042817.

Facchetti, G., Iacono, G., and Altafini, C. (2011).
Computing global structural balance in large-scale signed social networks.
In *Proceedings of the National Academy of Sciences of the United States of America*, volume 108, pages 20953–20958.

Figueiredo, R. and Frota, Y. (2012).
The maximum balanced subgraph of a signed graph: applications and solution approaches.
Paper submitted.

Figueiredo, R. and Moura, G. (2013).
Mixed integer programming formulations for clustering problems related to structural balance.
Social Netorks.

Harary, F., Lim, M., and Wunsch, D. C. (2003).
Signed graphs for portfolio analysis in risk management.
IMA Journal of Management Mathematics, 13:1–10.

Heider, F. (1946).
Attitudes and cognitive organization.
Journal of Psychology, 21:107–112.

Huffner, F., Betzler, N., and Niedermeier, R. (2010).
Separator-based data reduction for signed graph balancing.
Journal of Combinatorial Optimization, 20:335–360.

Leskovec, J., Huttenlocher, D., and Kleinberg, J. (2010).
Signed networks in social media.
In *CHI'10 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1361–1370.

Levorato, M., Drummond, L., Figueiredo, R., and Frota, Y. (2015).
Balance evaluation of signed social networks through correlation clustering problems.
Submitted to Computers and OR.

Lourenço, H., Martin, O., and Stutzle, T. (2003).
Handbook of Metaheuristics, chapter Iterated Local Search, pages 1355–1377.
Kluwer Academic Publishers, Norwell, MA.

Macon, K., Mucha, P., and Porter, M. (2012).
Community structure in the united nations general assembly.
Physica A: Statistical Mechanics and its Applications, 391:343–361.

Newman, M. E. J. and Girvan, M. (2004).
Finding and evaluating community structure in networks.
Physical Review E, 69(2):026113.

Pons, P. and Latapy, M. (2005).
Computing communities in large networks using random walks.
Lecture Notes in Computer Science, 3733:284–293.

Rosvall, M. and Bergstrom, C. T. (2008).
Maps of random walks on complex networks reveal community structure.
Proceedings of the National Academy of Sciences, 105(4):1118.

Srinivasan, A. (2011).
Local balancing influences global structure in social networks.
In *Proceedings of the National Academy of Sciences of the United States of America*, volume 108, pages 1751–1752.

Zhang, Z., Cheng, H., Chen, W., Zhang, S., and Fang, Q. (2008).
Correlation clustering based on genetic algorithm for documents clustering.
In *IEEE Congress on Evolutionary Computation*, pages 3193–3198.

Thank you for your attention!

Experimental results: implementation details

- GRASP and ILS heuristics:
  - Implemented in ANSI C++.
  - Heuristic outcomes represent the average of 5 independent runs.

- Mathematical formulation:
  - Xpress Mosel 3.2.0.

- All experiments were performed on:
  - Cluster with 42 nodes, each one with two Intel Xeon QuadCore 2.66GHz processors and 16Gb of RAM under Linux (Red Hat 5.3).