

TECHNICAL UNIVERSITY MUNICH

TUM Data Innovation Lab

A Network Analytical take on the European Parliament

- Authors: Abinav Ravi Venkatakrishnan
Niklas Schmidt
- Mentor: Dr. Mirco Schönfeld (Professorship for Computational Social Science & Big Data)
- Co-Mentor: Laure Vuaille (Department of Mathematics)
- Project Lead: Dr. Ricardo Acevedo Cabra (Department of Mathematics)
- Supervisor: Prof. Dr. Massimo Fornasier (Department of Mathematics)

Motivation

Find Hidden Agendas

Motivation

Find Hidden Agendas

What?

How?

Motivation

Find Hidden Coalition

What?

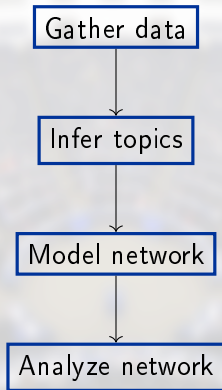
How?

Hidden Agenda

Following goal in non-obvious
manner

Hidden Coalition

Collaboration not apparent by
direct work



Gather data

```
graph TD; A[Gather data] --> B[Infer topics]; B --> C[Model network]; C --> D[Analyze network];
```

Infer topics

Model network

Analyze network

Gathering Data

	date	speechnr	agenda	name	nationality	party	euparty	text
0	1999-07-21	en.19990721.1.3-001	Address by the President	Nicole Fontaine	France	Union pour la démocratie française	Group of the European People's Party (Christia...	Ladies and gentlemen; once again; I should ilk...
1	1999-07-21	en.19990721.1.3-003	Address by the President	Nicole Fontaine	France	Union pour la démocratie française	Group of the European People's Party (Christia...	I thank the President-in-Office of the Council.
2	1999-07-21	en.19990721.1.3-005	Address by the President	Nicole Fontaine	France	Union pour la démocratie française	Group of the European People's Party (Christia...	I am truly grateful; Mr Commissioner Marin.
3	1999-07-21	en.19990721.2.3-006	Approval of the Minutes	Nicole Fontaine	France	Union pour la démocratie française	Group of the European People's Party (Christia...	The Minutes of the last sitting have been dist...
4	1999-07-21	en.19990721.2.3-007	Approval of the Minutes	Marie-Hélène Gillig	France	Parti socialiste	Group of the Party of European Socialists	(FR) Madam President; with regard to the Minut...

<http://linkedpolitics.ops.few.vu.nl>

Preprocessing

	date	speechnr	agenda	name	nationality	party	euparty	text
0	1999-07-21	en.19990721.1.3-001	Address by the President	Nicole Fontaine	France	Union pour la démocratie française	Group of the European People's Party (Christia...	[address, convey, heartfelt, trust, show, elec...
1	1999-07-21	en.19990721.1.3-003	Address by the President	Nicole Fontaine	France	Union pour la démocratie française	Group of the European People's Party (Christia...	[address, presidentinoffic]
2	1999-07-21	en.19990721.1.3-005	Address by the President	Nicole Fontaine	France	Union pour la démocratie française	Group of the European People's Party (Christia...	[address, truil, grate, marr]
3	1999-07-21	en.19990721.2.3-006	Approval of the Minutes	Nicole Fontaine	France	Union pour la démocratie française	Group of the European People's Party (Christia...	[approv, last, distribut, comment]
4	1999-07-21	en.19990721.2.3-007	Approval of the Minutes	Marie-Hélène Gillig	France	Parti socialiste	Group of the Party of European Socialists	[approv, regard, yesterday, provid, inform, re...

Gather data



Infer topics



Model network



Analyze network

Topic Modelling

- Latent Dirichlet allocation (LDA)
- Idea: Find topics in texts by assigning word probabilities to topics

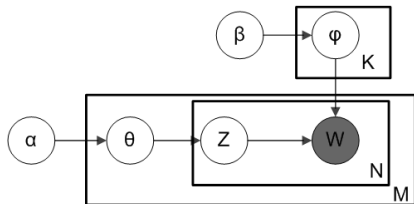
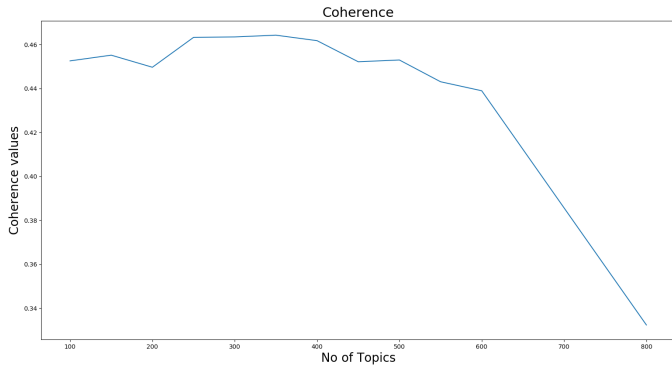


Figure:

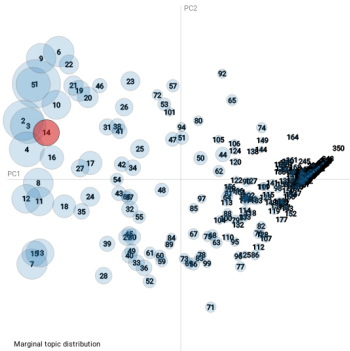
https://upload.wikimedia.org/wikipedia/commons/4/4d/Smoothed_LDA.png

Optimal Number of Topics



Topic Visualisation

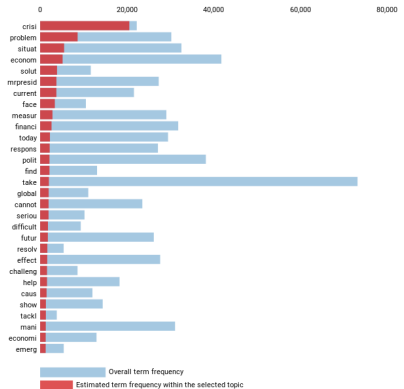
Intertopic Distance Map (via multidimensional scaling)



Marginal topic distribution



Top-30 Most Relevant Terms for Topic 14 (1.8% of tokens)



1. $s_{allency}(term, w) = frequency(w) * [\sum_i p(i|w) * \log(p(i|w)/p(i))]$ for topics i ; see Chuang et al. (2012)
 2. $relevance(term, w | topic t) = \lambda * p(w | t) + (1 - \lambda) * p(w | t)/p(w)$; see Sievert & Shirley (2014)

Example



Figure: Danielle Auroi

Madam President, President Prodi, ladies and gentlemen, after hearing Mr Prodi's proposals, I am utterly astounded by the position of the PPE and the PSE on food **safety**. Perhaps they do not feel capable of putting forward concrete proposals today, but we do. That is why we wished to propose a **resolution** for, throughout **Europe**, the series of **scandals** which have occurred means that, today, the citizens and consumers no longer have any confidence in their farmers. The quibbling involved in stating that the Committee ...

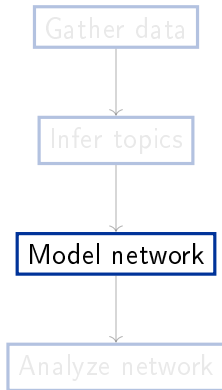
Inferred Topics

LDA result

- (0, '0.475*"strategi" + 0.152*"lisbon" + 0.055*"object" + 0.030*"implement" + 0.029*"competit"')
- (1, '0.153*"indian" + 0.130*"threeyear" + 0.094*"empir" + 0.066*"disintegr" + 0.053*"overshadow"')
- (2, '0.028*"develop" + 0.020*"econom" + 0.018*"area" + 0.017*"support" + 0.011*"increas"')
- (3, '0.211*"diseas" + 0.089*"prevent" + 0.073*"vaccin" + 0.062*"infect" + 0.053*"spread"')
- (4, '0.307*"polish" + 0.239*"domest" + 0.093*"beekeep" + 0.091*"gross" + 0.069*"default"')
- (5, '0.503*"medium" + 0.085*"televis" + 0.069*"broadcast" + 0.040*"audiovisu" + 0.036*"guinea"')
- (6, '0.202*"marginalis" + 0.154*"worsen" + 0.105*"antidiscrimin" + 0.066*"perpetu" + 0.062*"michel"')
- (7, '0.134*"volatil" + 0.104*"tight" + 0.097*"minimis" + 0.079*"roughli" + 0.079*"inher"')
- (8, '0.161*"hamper" + 0.148*"smallscal" + 0.091*"anticorrupt" + 0.083*"bolster" + 0.077*"adr"')
- (9, '0.075*"cooper" + 0.032*"develop" + 0.032*"instrument" + 0.022*"coordin" + 0.018*"effect"')

Inference

	name	date	topic
0	Marie-Noëlle Lienemann	1999-07-01	[(23, 0.05672398), (38, 0.016829032), (73, 0.0...
1	Gerhard Schmid	1999-07-01	[(141, 0.07714286), (242, 0.5914885), (257, 0....
2	Hanja Maij-Weggen	1999-07-01	[(36, 0.019324558), (109, 0.020725463), (111, ...
3	Ingo Friedrich	1999-07-01	[(60, 0.08798485), (110, 0.022743504), (144, 0...
4	Hans-Peter Martin	1999-07-01	[(146, 0.28848597), (238, 0.023322258), (242, ...

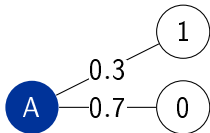


Network modelling

name	date	topic
A	1999-07-01	[(0,0.7), (1,0.3)]
B	1999-07-01	[(0,0.7), (1,0.3)]
C	1999-07-01	[(1,0.7), (2,0.3)]
D	1999-07-01	[(2,1.0)]
A	1999-08-01	[(0,0.5), (1,0.5)]
B	1999-08-01	[(0,0.3), (1,0.2), (2,0.5)]
C	1999-08-01	[(1,0.5), (2,0.5)]
A	1999-09-01	[(0,0.8), (1,0.2)]

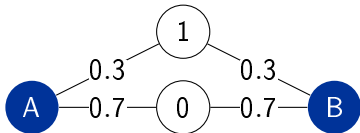
Network modelling

name	date	topic
A	1999-07-01	[(0,0.7), (1,0.3)]
B	1999-07-01	[(0,0.7), (1,0.3)]
C	1999-07-01	[(1,0.7), (2,0.3)]
D	1999-07-01	[(2,1.0)]
A	1999-08-01	[(0,0.5), (1,0.5)]
B	1999-08-01	[(0,0.3), (1,0.2), (2,0.5)]
C	1999-08-01	[(1,0.5), (2,0.5)]
A	1999-09-01	[(0,0.8), (1,0.2)]



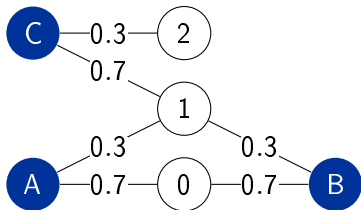
Network modelling

name	date	topic
A	1999-07-01	[(0,0.7), (1,0.3)]
B	1999-07-01	[(0,0.7), (1,0.3)]
C	1999-07-01	[(1,0.7), (2,0.3)]
D	1999-07-01	[(2,1.0)]
A	1999-08-01	[(0,0.5), (1,0.5)]
B	1999-08-01	[(0,0.3), (1,0.2), (2,0.5)]
C	1999-08-01	[(1,0.5), (2,0.5)]
A	1999-09-01	[(0,0.8), (1,0.2)]



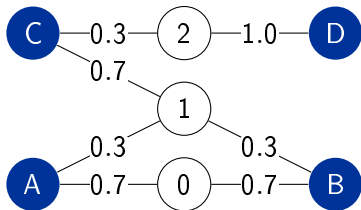
Network modelling

name	date	topic
A	1999-07-01	[(0,0.7), (1,0.3)]
B	1999-07-01	[(0,0.7), (1,0.3)]
C	1999-07-01	[(1,0.7), (2,0.3)]
D	1999-07-01	[(2,1.0)]
A	1999-08-01	[(0,0.5), (1,0.5)]
B	1999-08-01	[(0,0.3), (1,0.2), (2,0.5)]
C	1999-08-01	[(1,0.5), (2,0.5)]
A	1999-09-01	[(0,0.8), (1,0.2)]



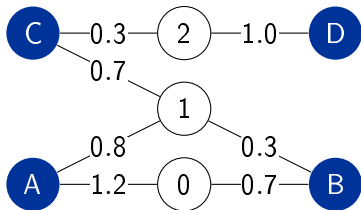
Network modelling

name	date	topic
A	1999-07-01	[(0,0.7), (1,0.3)]
B	1999-07-01	[(0,0.7), (1,0.3)]
C	1999-07-01	[(1,0.7), (2,0.3)]
D	1999-07-01	[(2,1.0)]
A	1999-08-01	[(0,0.5), (1,0.5)]
B	1999-08-01	[(0,0.3), (1,0.2), (2,0.5)]
C	1999-08-01	[(1,0.5), (2,0.5)]
A	1999-09-01	[(0,0.8), (1,0.2)]



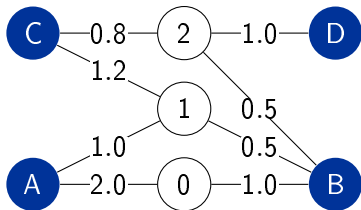
Network modelling

name	date	topic
A	1999-07-01	[(0,0.7), (1,0.3)]
B	1999-07-01	[(0,0.7), (1,0.3)]
C	1999-07-01	[(1,0.7), (2,0.3)]
D	1999-07-01	[(2,1.0)]
A	1999-08-01	[(0,0.5), (1,0.5)]
B	1999-08-01	[(0,0.3), (1,0.2), (2,0.5)]
C	1999-08-01	[(1,0.5), (2,0.5)]
A	1999-09-01	[(0,0.8), (1,0.2)]



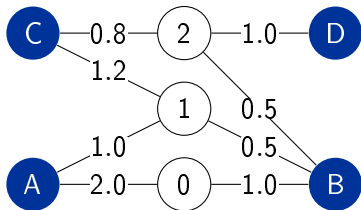
Network modelling

name	date	topic
A	1999-07-01	[(0,0.7), (1,0.3)]
B	1999-07-01	[(0,0.7), (1,0.3)]
C	1999-07-01	[(1,0.7), (2,0.3)]
D	1999-07-01	[(2,1.0)]
A	1999-08-01	[(0,0.5), (1,0.5)]
B	1999-08-01	[(0,0.3), (1,0.2), (2,0.5)]
C	1999-08-01	[(1,0.5), (2,0.5)]
A	1999-09-01	[(0,0.8), (1,0.2)]



Network modelling

name	date	topic
A	1999-07-01	[(0,0.7), (1,0.3)]
B	1999-07-01	[(0,0.7), (1,0.3)]
C	1999-07-01	[(1,0.7), (2,0.3)]
D	1999-07-01	[(2,1.0)]
A	1999-08-01	[(0,0.5), (1,0.5)]
B	1999-08-01	[(0,0.3), (1,0.2), (2,0.5)]
C	1999-08-01	[(1,0.5), (2,0.5)]
A	1999-09-01	[(0,0.8), (1,0.2)]



$$\begin{bmatrix} 0 \\ 1.2 \\ 0.8 \end{bmatrix} \text{C}$$

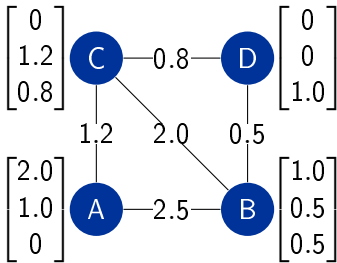
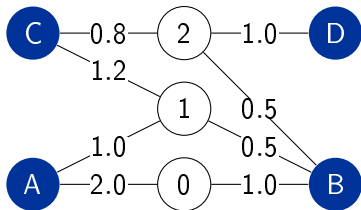
$$\text{D} \begin{bmatrix} 0 \\ 0 \\ 1.0 \end{bmatrix}$$

$$\begin{bmatrix} 2.0 \\ 1.0 \\ 0 \end{bmatrix} \text{A}$$

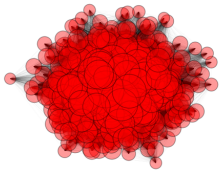
$$\text{B} \begin{bmatrix} 1.0 \\ 0.5 \\ 0.5 \end{bmatrix}$$

Network modelling

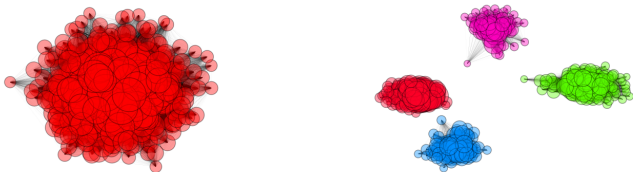
name	date	topic
A	1999-07-01	[(0,0.7), (1,0.3)]
B	1999-07-01	[(0,0.7), (1,0.3)]
C	1999-07-01	[(1,0.7), (2,0.3)]
D	1999-07-01	[(2,1.0)]
A	1999-08-01	[(0,0.5), (1,0.5)]
B	1999-08-01	[(0,0.3), (1,0.2), (2,0.5)]
C	1999-08-01	[(1,0.5), (2,0.5)]
A	1999-09-01	[(0,0.8), (1,0.2)]



Community Detection



Community Detection



$$Q = \frac{1}{2m} \sum_{v,u \in V} \left(a_{vu} - \frac{k_v k_u}{2m} \right) \delta(c_v, c_u)$$

Gather data

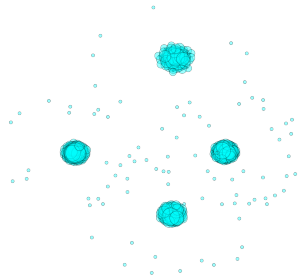
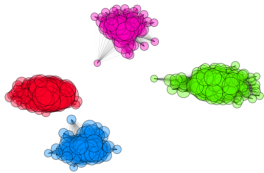
```
graph TD; A[Gather data] --> B[Infer topics]; B --> C[Model network]; C --> D[Analyze network];
```

Infer topics

Model network

Analyze network

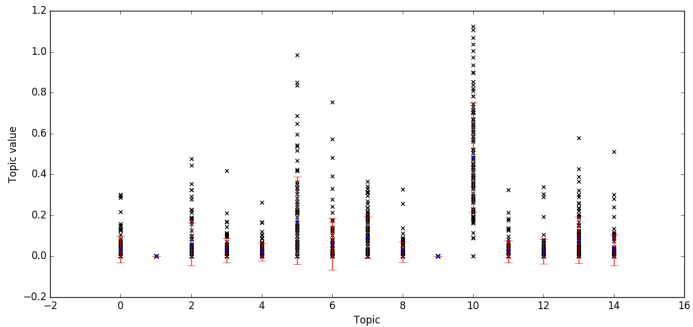
Outlier Detection



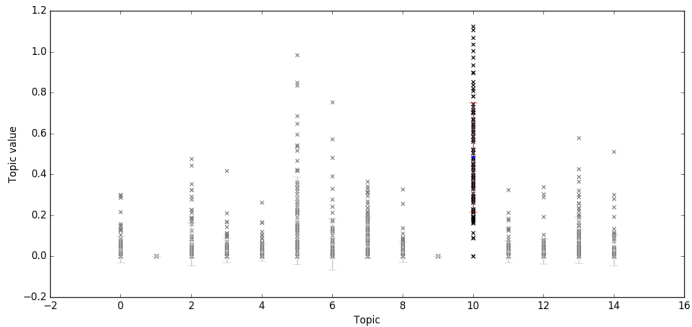
Outlier

- Girvan-Newman algorithm
- High topic overlap \Leftrightarrow Large edge weight
- Mismatching data

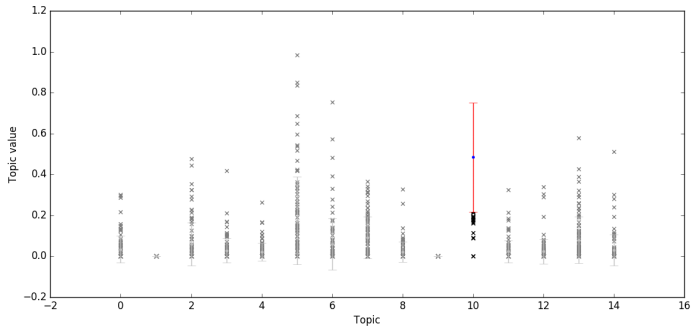
Topic Distribution



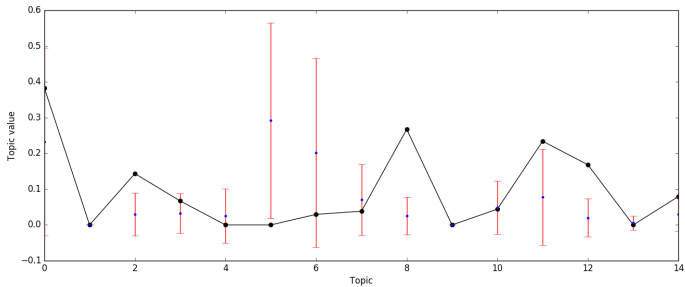
Topic Distribution



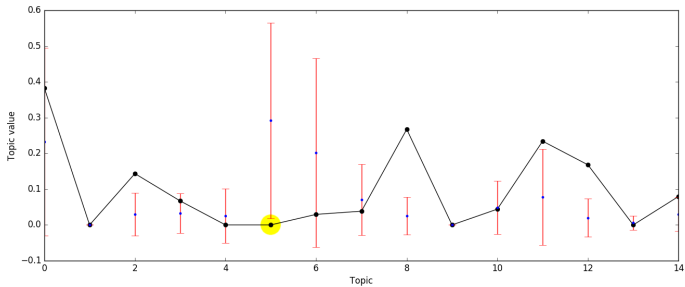
Topic Distribution



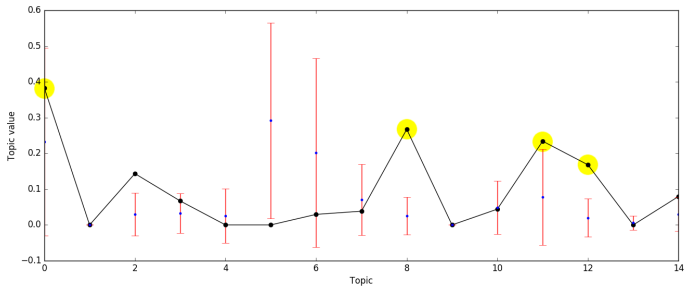
Outlier vs. Community



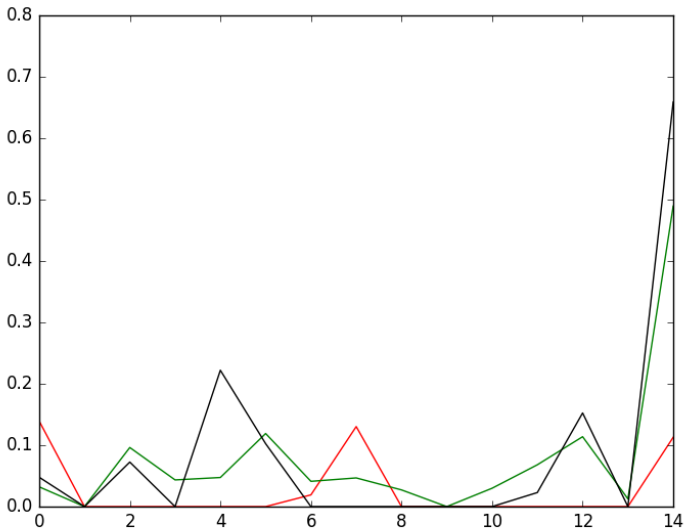
Outlier vs. Community



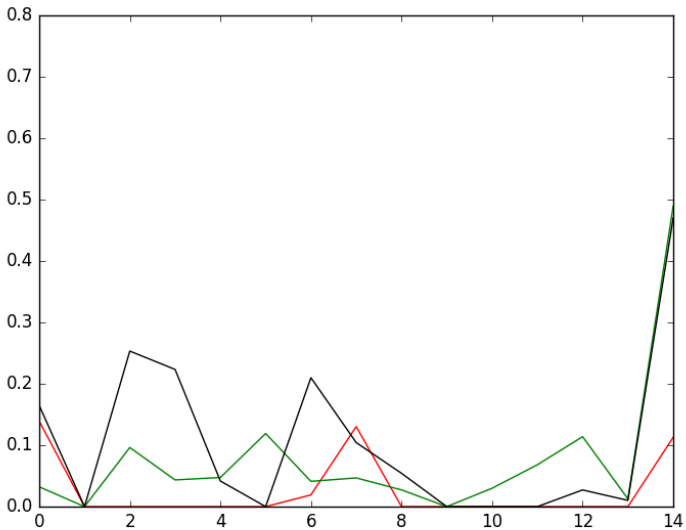
Outlier vs. Community



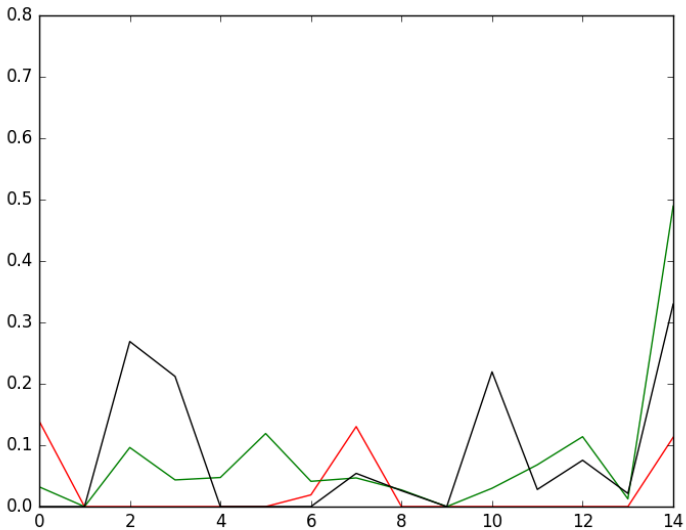
Outlier vs. Neighbours



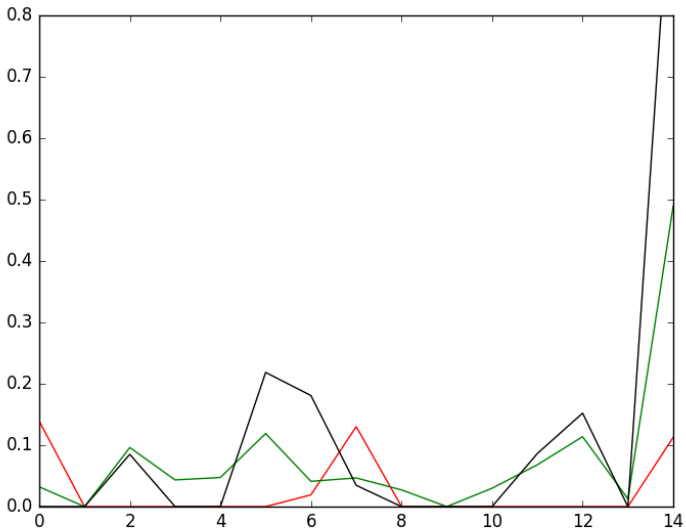
Outlier vs. Neighbours



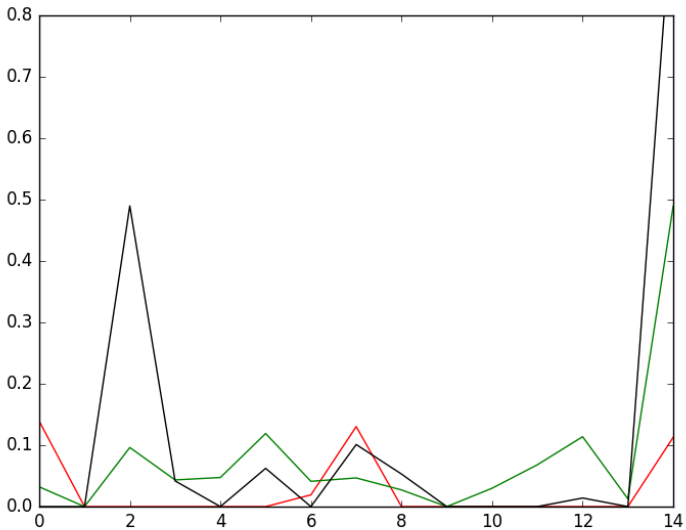
Outlier vs. Neighbours



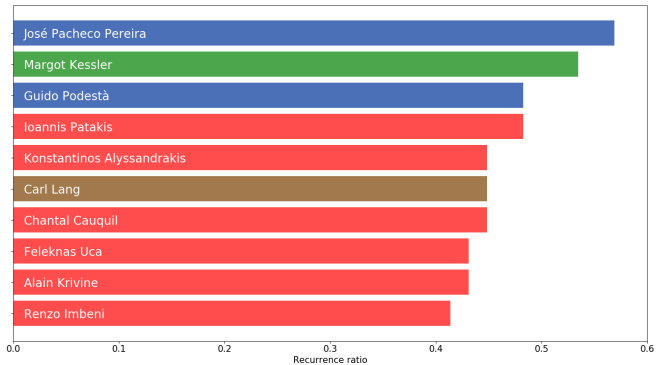
Outlier vs. Neighbours



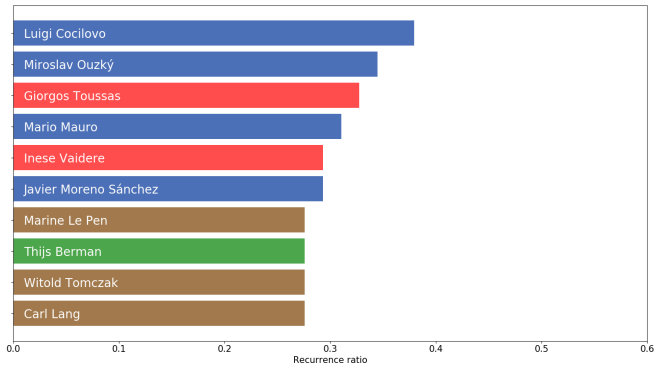
Outlier vs. Neighbours



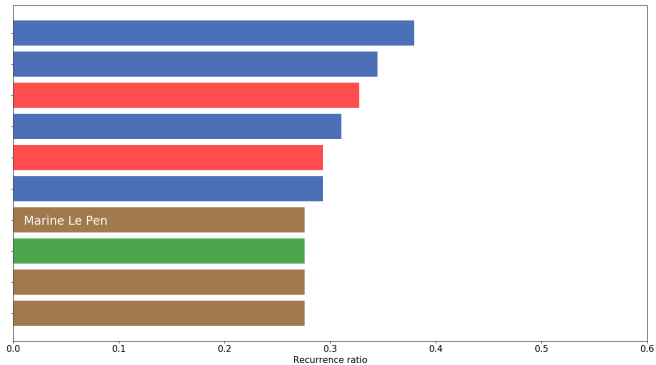
Results Hidden Agenda



Results Hidden Agenda



Results Hidden Agenda



Hidden Community Detection

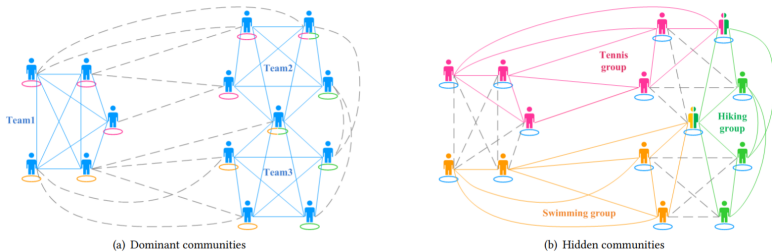


Figure: From K.He et.al, Hidden Community Detection in Social Networks,2017

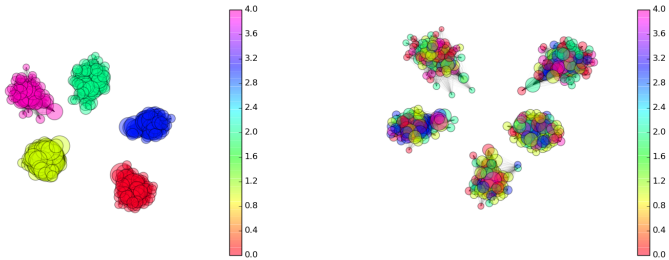
HiCoDe Algorithm

- Apply the base algorithm - Louvain Algorithm
- Calculate the modularity
- Weaken the structure by using refinement algorithms such as remove edge or reduce edge
- Repeat until appropriate layers

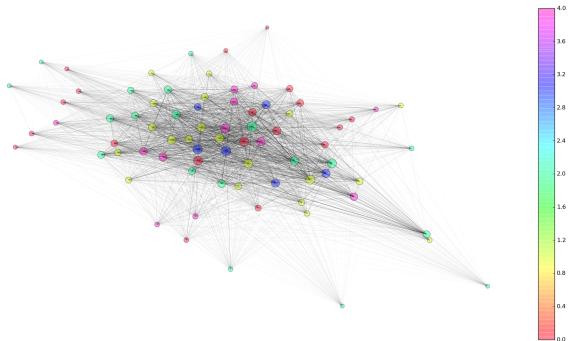
Number of Layers

- Calculate the modularity for dominant community Q_0
- Perform T iterations of refinement and calculate modularity for each iteration Q_T
- Calculate average improvement ratio of modularity per iteration. as $R_T = \frac{\sum_{t=1}^T Q_T}{Q_0 T}$
- Choose layer which has highest R_T

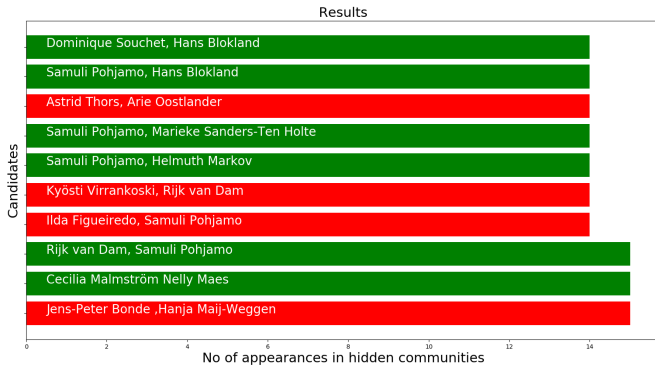
Hidden Community Detection



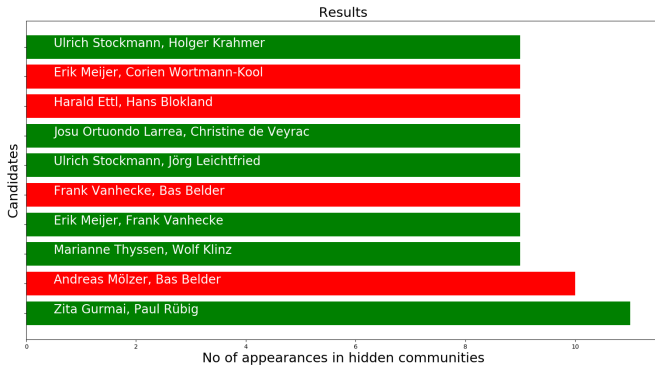
One Hidden Community



Results Hidden Coalition

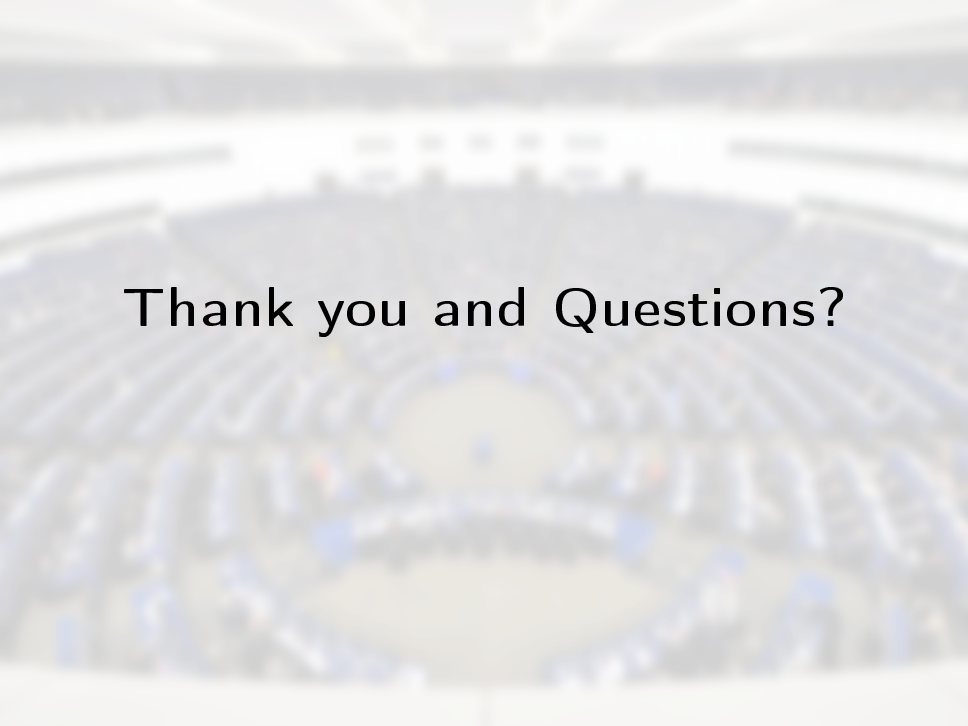


Results Hidden Coalition



Outlook

- Translation
- Metadata
- Hollistic Community Outlier



Thank you and Questions?