Measuring the Diversity of Shared News and the Effect of Recommendation Algorithms

Pedro Ramaciotti Morales, Robin Lamarche-Perrin

{firstname.lastname-lastname}@lip6.fr

LIP6-UMPC, Sorbonne Université, Paris.

Leipzig, June 20, 2018

Introduction

- Project ALGODIV: measuring the diversity of the digital public space and the effects of recommender systems (RS).

Introduction

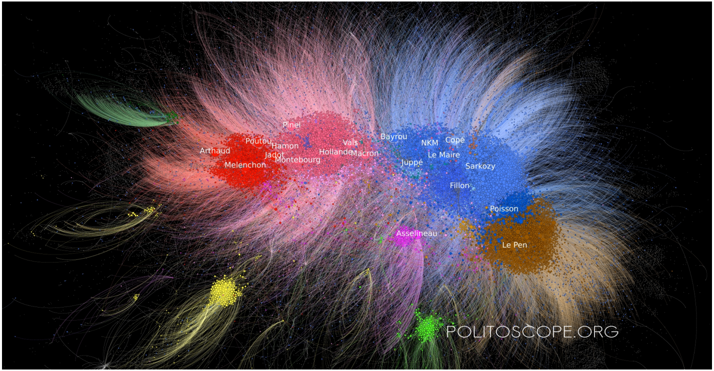
- Project ALGODIV: measuring the diversity of the digital public space and the effects of recommender systems (RS).
- Aim of this presentation: present a particular approach to the problem of measuring diversity, use it to explore the diversity of news sharing on Twitter, and to explore the effects of RS.

Introduction

- Project ALGODIV: measuring the diversity of the digital public space and the effects of recommender systems (RS).
- Aim of this presentation: present a particular approach to the problem of measuring diversity, use it to explore the diversity of news sharing on Twitter, and to explore the effects of RS.
- Structure of this presentation:
  1. Using Politoscope data to build classification of media sources in the political community space.
  2. A graph- and information-theoretical approach to the problem of measuring diversity.
  3. Analysing the diversity of recommendations of media sources using RS.

News Sharing on Twitter

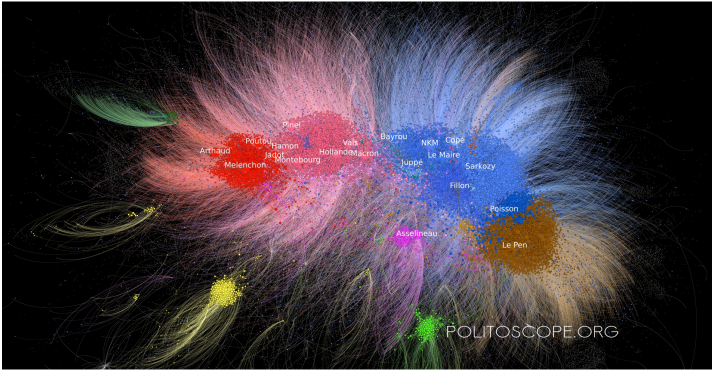
Politoscope dataset: data on media sources shared on Twitter during the 2017 French presidential election campaign and communitarisation of users.



[Gaumont, Panahi, Chavalarias 2017]

News Sharing on Twitter

Politoscope dataset: data on media sources shared on Twitter during the 2017 French presidential election campaign and communitarisation of users.



[Gaumont, Panahi, Chavalarias 2017]

How to use the community structure of the user-network to produce a classification of media sources in a political community space?

News Sharing on Twitter

In the dataset, entries register users belonging to communities sharing media sources. This allows us to create a **source-community mention matrix**.

News Sharing on Twitter

In the dataset, entries register users belonging to communities sharing media sources. This allows us to create a **source-community mention matrix**.

For a period under analysis (1 month before the first round)...

comm.	user	source	weight
Fil	u1	Figaro	0.5
Mac	u3	Le Monde	1.0
Mel	u2	BFMTV	1.0
Ham	u1	FranceInfo	0.5
Fil	u4	Figaro	1.0
⋮	⋮	⋮	⋮

News Sharing on Twitter

In the dataset, entries register users belonging to communities sharing media sources. This allows us to create a **source-community mention matrix**.

For a period under analysis (1 month before the first round)...

comm.	user	source	weight
Fil	u1	Figaro	0.5
Mac	u3	Le Monde	1.0
Mel	u2	BFMTV	1.0
Ham	u1	FranceInfo	0.5
Fil	u4	Figaro	1.0
⋮	⋮	⋮	⋮

We assemble the mention matrix and normalize for each source.

News Sharing on Twitter

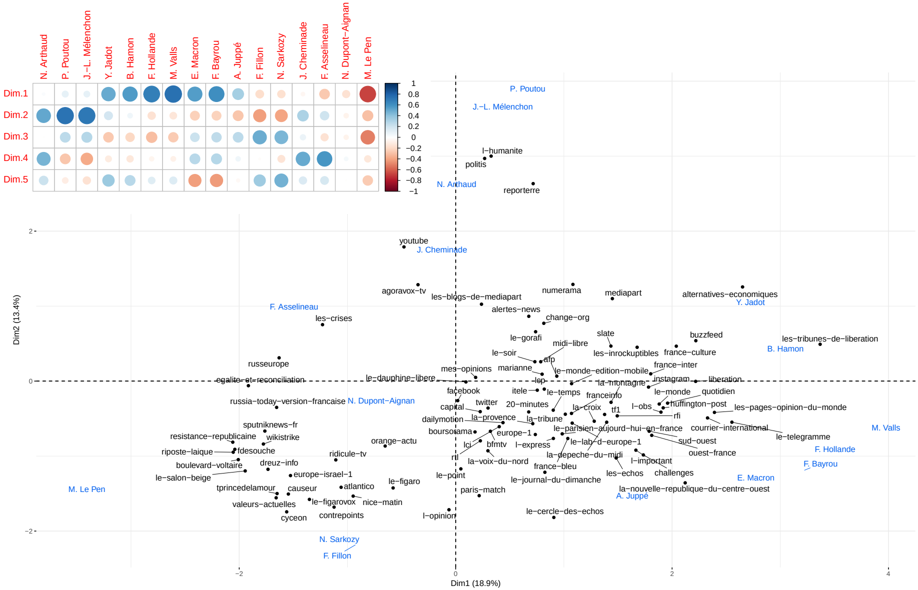
In the dataset, entries register users belonging to communities sharing media sources. This allows us to create a **source-community mention matrix**.

For a period under analysis (1 month before the first round)...

comm.	user	source	weight
Fil	u1	Figaro	0.5
Mac	u3	Le Monde	1.0
Mel	u2	BFMTV	1.0
Ham	u1	FranceInfo	0.5
Fil	u4	Figaro	1.0
⋮	⋮	⋮	⋮

We assemble the mention matrix and normalize for each source.

	Art	Ass	Bay	Che	DAi	Fil	Ham	Hol	Jad	Jup	LPe	Mac	Mel	Pou	Sar	Val
twitter	0.01	0.03	0.07	0.01	0.03	0.14	0.03	0.04	0.03	0.02	0.19	0.07	0.18	0.03	0.07	0.04
youtube	0.04	0.10	0.01	0.03	0.02	0.04	0.02	0.01	0.02	0.00	0.12	0.01	0.50	0.06	0.02	0.01
facebook	0.01	0.03	0.05	0.01	0.03	0.10	0.03	0.04	0.03	0.02	0.34	0.05	0.13	0.03	0.06	0.03
le-figaro	0.01	0.03	0.05	0.01	0.02	0.30	0.01	0.04	0.01	0.03	0.18	0.05	0.07	0.01	0.14	0.03
bfmtv	0.03	0.05	0.08	0.01	0.03	0.10	0.04	0.06	0.04	0.02	0.17	0.08	0.18	0.02	0.05	0.05
le-monde	0.01	0.03	0.14	0.01	0.02	0.04	0.06	0.10	0.06	0.04	0.05	0.14	0.15	0.02	0.03	0.09
franceinfo	0.01	0.03	0.10	0.01	0.02	0.07	0.04	0.07	0.04	0.04	0.15	0.10	0.17	0.02	0.04	0.07
lafranceinsoumise.fr	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.93	0.04	0.00	0.01
liberation	0.01	0.01	0.11	0.01	0.01	0.08	0.09	0.08	0.09	0.02	0.06	0.11	0.21	0.03	0.02	0.08
en-marche.fr	0.00	0.00	0.31	0.00	0.00	0.00	0.00	0.17	0.00	0.02	0.01	0.31	0.00	0.00	0.00	0.16
le-parisien	0.01	0.03	0.09	0.01	0.02	0.12	0.05	0.06	0.05	0.04	0.16	0.09	0.12	0.02	0.07	0.06
l-obs	0.01	0.01	0.13	0.00	0.02	0.06	0.04	0.09	0.04	0.03	0.08	0.13	0.21	0.02	0.04	0.09
valeurs-actuelles	0.00	0.01	0.01	0.00	0.02	0.34	0.00	0.00	0.00	0.02	0.41	0.01	0.02	0.00	0.15	0.00
huffington-post	0.01	0.01	0.10	0.01	0.01	0.08	0.07	0.08	0.07	0.03	0.04	0.10	0.26	0.02	0.04	0.07
le-point	0.01	0.02	0.11	0.00	0.02	0.18	0.03	0.07	0.03	0.02	0.12	0.11	0.11	0.02	0.11	0.06

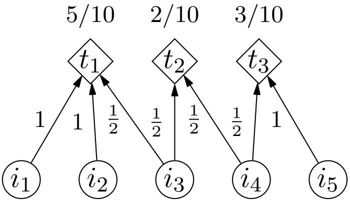

N. Arthaud
P. Poutou
J.–L. Mélenchon
Y. Jadot
B. Hamon
F. Hollande
M. Valls
E. Macron
F. Bayrou
A. Juppé
F. Fillon
N. Sarkozy
J. Cheminade
F. Asselineau
N. Dupont–Aignan
M. Le Pen
Dim.1
Dim.2
Dim.3
Dim.4
Dim.5
1
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
−0.8
−1
Dim1 (18.9%)
Dim2 (13.4%)
l–humanite
politis
reporterre
youtube
agoravox–tv
les–blogs–de–mediapart
numerama
mediapart
alternatives–economiques
alertes–news
change–org
les–crises
le–gorafi
midi–libre
slate
buzzfeed
les–tribunes–de–liberation
les–inrockuptibles
france–culture
le–soir
afp
russeurope
mes–opinions
marianne
le–monde–edition–mobile
france–inter
le–dauphine–libere
egalite–et–reconciliation
lep
la–montagne
instagram
liberation
facebook
itele
le–temps
franceinfo
le–monde
quotidien
russia–today–version–francaise
capital
twitter
20–minutes
la–croix
l–obs
huffington–post
les–pages–opinion–du–monde
tf1
rfi
dailymotion
la–provence
la–tribune
sputniknews–fr
le–parisien–aujourd–hui–en–france
courrier–international
resistance–republicaine
boursorama
europe–1
le–telegramme
wikistrike
orange–actu
lci
bfmtv
l–express
le–lab–d–europe–1
sud–ouest
riposte–laique
fdesouche
ouest–france
ridicule–tv
rtl
la–depeche–du–midi
boulevard–voltaire
dreuz–info
la–voix–du–nord
l–important
france–bleu
les–echos
challenges
le–salon–beige
europe–israel–1
le–point
le–figaro
le–journal–du–dimanche
tprincedelamour
causeur
atlantico
paris–match
la–nouvelle–republique–du–centre–ouest
valeurs–actuelles
le–figarovox
nice–matin
contrepoints
l–opinion
le–cercle–des–echos
cyceon
2
0
−2
4

Measuring Diversity

Information theory offers quantitave measures of diversity in situations that can be modeled as elements belonging to categories. These are widely used in ecology, sociology, economics (e.g. richness, entropy, Simpson index, Herfindahl index).

Measuring Diversity

Information theory offers quantitave measures of diversity in situations that can be modeled as elements belonging to categories. These are widely used in ecology, sociology, economics (e.g. richness, entropy, Simpson index, Herfindahl index).

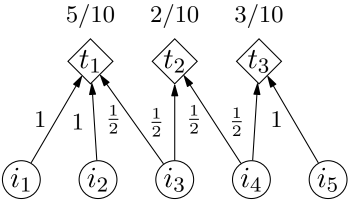


Type-proportional abundance of items:

$$p = (5/10, 2/10, 3/10)$$

Measuring Diversity

Information theory offers quantitave measures of diversity in situations that can be modeled as elements belonging to categories. These are widely used in ecology, sociology, economics (e.g. richness, entropy, Simpson index, Herfindahl index).



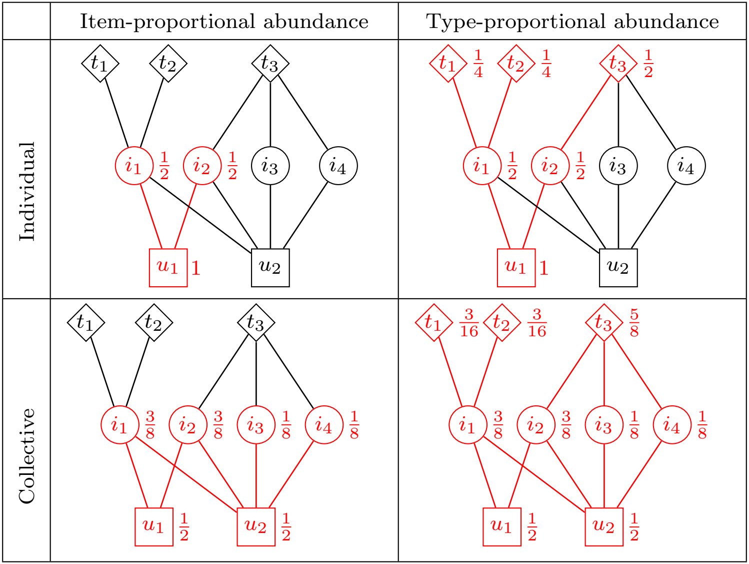
Type-proportional abundance of items:

$$p = (5/10, 2/10, 3/10)$$

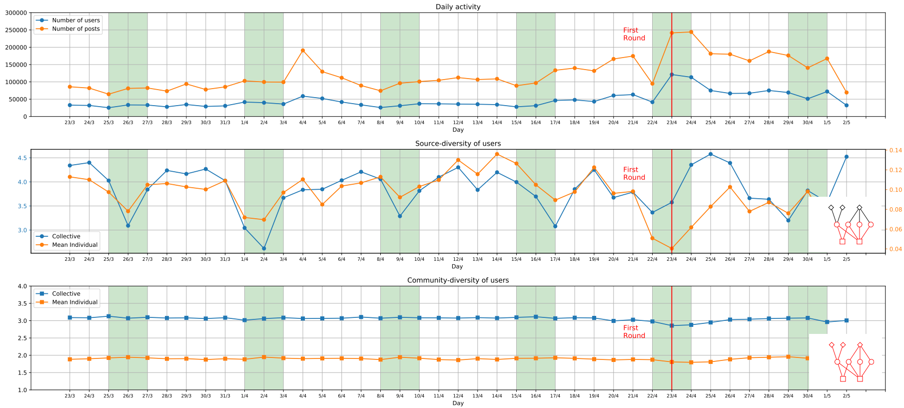
A diversity measure is a function $\mathcal{D} : p \longmapsto d \in \mathbb{R}_+$. A commonly used diversity measure is the Shannon Entropy:

$$\mathcal{D}(p) = -\sum_{i=1}^{n} p_i \log_2 (p_i)$$

Measuring Diversity



Measuring Diversity



1. News Sharing on Twitter

2. Measuring Diversity

3. Recommender Systems

- Interest on RS, particularly in domains of societal relevance.

Recommender Systems

- Interest on RS, particularly in domains of societal relevance.
- What are RS? Algorithmic systems that analyse previous behaviour to predict prefereces/interests on items, in order to produce recommendations.

Recommender Systems

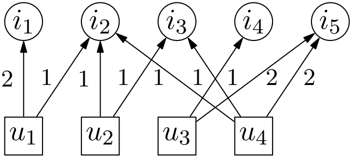
- Interest on RS, particularly in domains of societal relevance.
- What are RS? Algorithmic systems that analyse previous behaviour to predict prefereces/interests on items, in order to produce recommendations.
- Why are they needed? Large set of items, limited attention/time.

Recommender Systems

- Interest on RS, particularly in domains of societal relevance.
- What are RS? Algorithmic systems that analyse previous behaviour to predict prefereces/interests on items, in order to produce recommendations.
- Why are they needed? Large set of items, limited attention/time.
- RS comprise mainly Content-Based & Collaborative Filtering (user-item matrix).
  - Neighborhood Methods (Memory-based CF).
  - Matrix Factorization Methods (Model-based CF).

Recommender Systems

Analyzing users' previous choices of media sources, encoded in a user-item matrix.

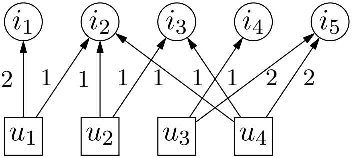


$$R \in \mathbb{R}^{|\mathcal{U}|\times|\mathcal{I}|}$$

$$\begin{array}{c|ccccc|} & i_1 & i_2 & i_3 & i_4 & i_5 \\ \hline u_1 & 2 & 1 & 0 & 0 & 0 \\ u_2 & 0 & 1 & 1 & 0 & 0 \\ u_3 & 0 & 0 & 0 & 1 & 2 \\ u_4 & 0 & 1 & 1 & 0 & 2 \\ \hline \end{array}$$

Recommender Systems

Analyzing users' previous choices of media sources, encoded in a user-item matrix.



$R \in \mathbb{R}^{|\mathcal{U}| \times |\mathcal{I}|}$

	i_1	i_2	i_3	i_4	i_5
u_1	2	1	0	0	0
u_2	0	1	1	0	0
u_3	0	0	0	1	2
u_4	0	1	1	0	2

Item-Based Collaborative Filtering

	i_1	i_2	i_3	i_4	i_5
i_1	1.0	0.6	0.0	0.0	0.0
i_2	0.6	1.0	0.8	0.0	0.4
i_3	0.0	0.8	1.0	0.0	0.5
i_4	0.0	0.0	0.0	1.0	0.7
i_5	0.0	0.4	0.5	0.7	1.0

Item similarity matrix:

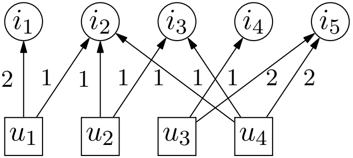
$$S \in \mathbb{R}^{|\mathcal{I}| \times |\mathcal{I}|}$$

$$\hat{R}_{u,i} = \frac{\sum\limits_{j \in \mathrm{Clique}(i)} S_{i,j} \cdot R_{u,j}}{\sum\limits_{j \in \mathrm{Clique}(i)} |S_{i,j}|}$$

Main parameter: clique size.

Recommender Systems

Analyzing users' previous choices of media sources, encoded in a user-item matrix.



$R \in \mathbb{R}^{|\mathcal{U}| \times |\mathcal{I}|}$

	i_1	i_2	i_3	i_4	i_5
u_1	2	1	0	0	0
u_2	0	1	1	0	0
u_3	0	0	0	1	2
u_4	0	1	1	0	2

Item-Based Collaborative Filtering

	i_1	i_2	i_3	i_4	i_5
i_1	1.0	0.6	0.0	0.0	0.0
i_2	0.6	1.0	0.8	0.0	0.4
i_3	0.0	0.8	1.0	0.0	0.5
i_4	0.0	0.0	0.0	1.0	0.7
i_5	0.0	0.4	0.5	0.7	1.0

Item similarity matrix:

$$S \in \mathbb{R}^{|\mathcal{I}| \times |\mathcal{I}|}$$

$$\hat{R}_{u,i} = \frac{\sum_{j \in \text{Clique}(i)} S_{i,j} \cdot R_{u,j}}{\sum_{j \in \text{Clique}(i)} |S_{i,j}|}$$

Main parameter: clique size.

Matrix Factorisation

Latent factors $\mathcal{F}$,

$$U \in \mathbb{R}^{|\mathcal{U}| \times |\mathcal{F}|},\ I \in \mathbb{R}^{|\mathcal{I}| \times |\mathcal{F}|},$$

$$\hat{R} = U \cdot I^T.$$

$$\min_{U,I} \sum_{(u,i) \text{ s.t. } R_{u,i} \neq 0} \left(R_{u,i} - U_u I_i^T\right)^2 + reg.$$

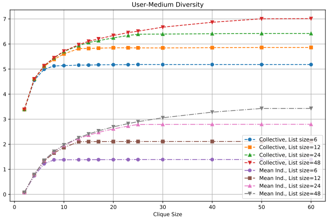
Main parameter: num. of latent factors.

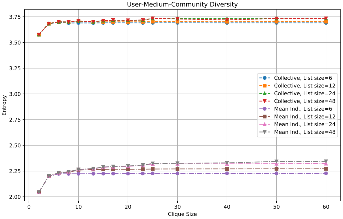
Diversity of rec. based on Politoscope data (April 23rd, 2017).

Source-diversity of users

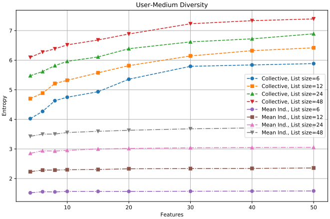
Community-diversity of users

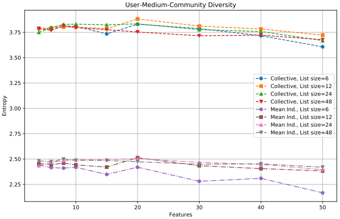
IBCF

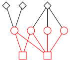


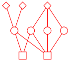


Matrix Factorisation









Conclusions & Future Work

- Locating/classifying media sources in the space of political communities using Politoscope data.

Conclusions & Future Work

- Locating/classifying media sources in the space of political communities using Politoscope data.
- Measuring the diversity in a network of users, items, and classifications: graph- & information-theoretical approach.

Conclusions & Future Work

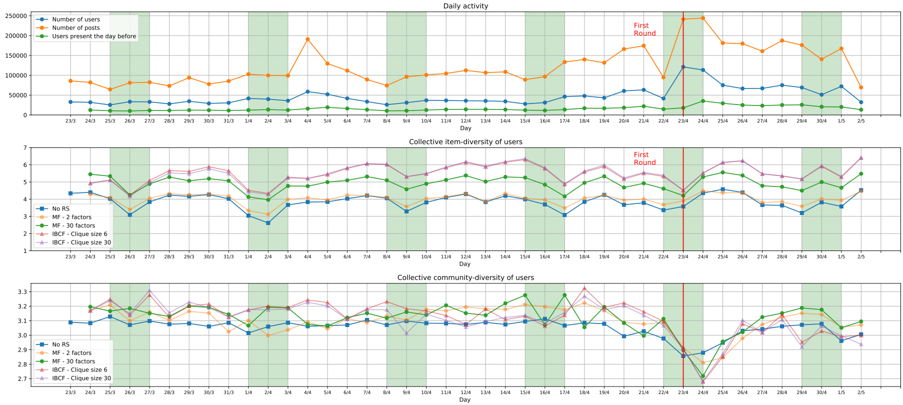
- Locating/classifying media sources in the space of political communities using Politoscope data.
- Measuring the diversity in a network of users, items, and classifications: graph- & information-theoretical approach.
- Measuring the diversity of users in the space of media sources, and in the space and of political communities.

Conclusions & Future Work

- Locating/classifying media sources in the space of political communities using Politoscope data.
- Measuring the diversity in a network of users, items, and classifications: graph- & information-theoretical approach.
- Measuring the diversity of users in the space of media sources, and in the space and of political communities.
- Measuring the diversity of recommendations using ubiquitous RS on Politoscope .

Conclusions & Future Work

- Locating/classifying media sources in the space of political communities using Politoscope data.
- Measuring the diversity in a network of users, items, and classifications: graph- & information-theoretical approach.
- Measuring the diversity of users in the space of media sources, and in the space and of political communities.
- Measuring the diversity of recommendations using ubiquitous RS on Politoscope .
- Future work: simulating the evolution of the diversity under the influence of RS.


Daily activity
Number of users
Number of posts
Users present the day before
First
Round
Day
Collective item-diversity of users
No RS
MF - 2 factors
MF - 30 factors
IBCF - Clique size 6
IBCF - Clique size 30
Collective community-diversity of users

Thank you for your attention.