Opening up Echo Chambers via Optimal Content Recommendation

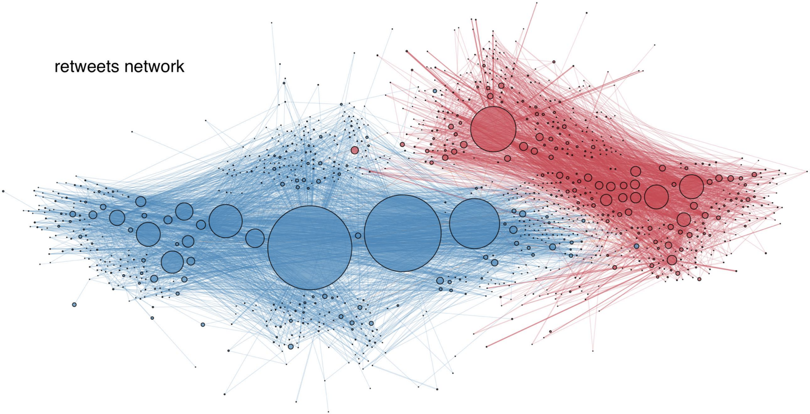
A. Vendeville[1,2], A. Giovanidis[3], E. Papanastasiou[3] and B. Guedj[1,2]

[1]University College London, [2]Inria, [3]Sorbonne University

September 29, 2022

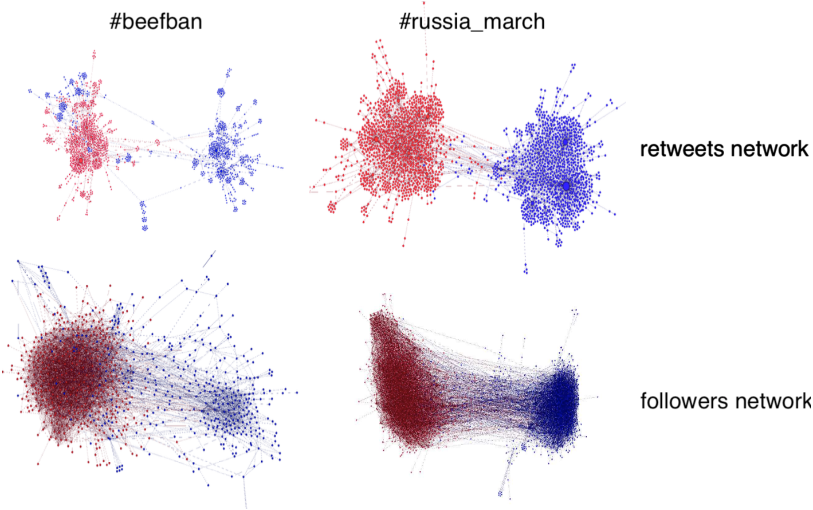
What is an echo chamber?

Echo chambers



Weber *et al.*(2020). #ArsonEmergency and Australia's "Black Summer": Polarisation and Misinformation on Social Media. MISDOOM 2020.
https://doi.org/10.1007/978-3-030-61841-4_11

Echo chambers



Garimella *et al.*(2016). Quantifying Controversy in Social Media. WSDM ’16.
`https://doi.org/10.1145/2835776.2835792`.

Consequences...

- opinion polarisation
- extremism
- fake news
- conspiracy theories

Consequences...

- opinion polarisation
- extremism
- fake news
- conspiracy theories

Need to open up the echo chambers!

The #Elysée2017fr dataset

Fraisier, O., Cabanac, G., Pitarch, Y., Besançon, R., Boughanem, M.: #Elysée2017fr: The 2017 French Presidential Campaign on Twitter. In: Proceedings of the 12th International AAAI Conference on Web and Social Media (2018).

The #Elysée2017fr dataset

Fraisier, O., Cabanac, G., Pitarch, Y., Besançon, R., Boughanem, M.: #Elysée2017fr: The 2017 French Presidential Campaign on Twitter. In: Proceedings of the 12th International AAAI Conference on Web and Social Media (2018).

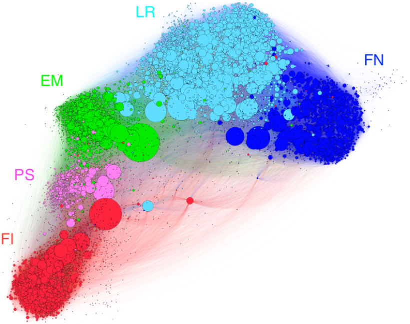
- 2.4M tweets
- 7.7M retweets
- 22,853 profiles
- November 2016 - May 2017
- **known political affiliations FI,PS,EM,LR,FN**

The #Elysée2017fr dataset

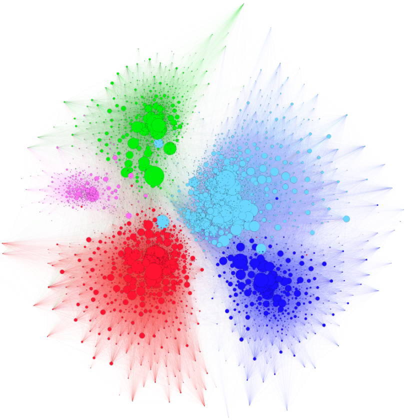
Fraisier, O., Cabanac, G., Pitarch, Y., Besançon, R., Boughanem, M.: #Elysée2017fr: The 2017 French Presidential Campaign on Twitter. In: Proceedings of the 12th International AAAI Conference on Web and Social Media (2018).

- 2.4M tweets
- 7.7M retweets
- 22,853 profiles
- November 2016 - May 2017
- **known political affiliations FI,PS,EM,LR,FN**

Followers graph: 8,277 users and 975,168 edges

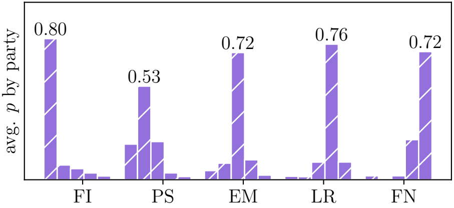


Followers graph



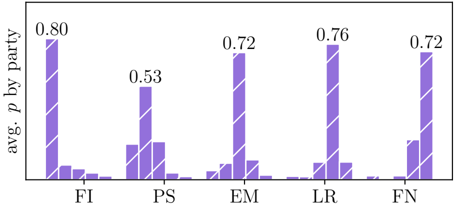
Retweet graph

Echo chambers in #Elysée2017fr



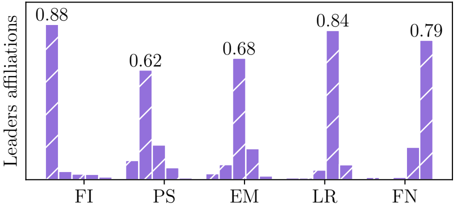
Distribution of content users are exposed to.

Echo chambers in `#Elysée2017fr`



Distribution of content users are exposed to.

Not surprising...



Quantifying content diversity

For user n:

$$\Phi_n = \frac{S}{S-1} \sum_{s=1}^{S} p_s^{(n)} (1 - p_s^{(n)}). \tag{1}$$

$p_s^{(n)}$: average proportion of content from party s on the newsfeed of n.
$S = 5$: number of parties.

Quantifying content diversity

For user n:

$$\Phi_n = \frac{S}{S-1} \sum_{s=1}^{S} p_s^{(n)}(1 - p_s^{(n)}). \tag{1}$$

$p_s^{(n)}$: average proportion of content from party s on the newsfeed of n.
$S = 5$: number of parties.

How to increase Φ_n with recommendations?

Diffusion model

- Strongly connected network of N users.
- Self-posting rates $\lambda_s^{(n)}$.
- Re-posting rates $\mu^{(n)}$.
- Newsfeeds of finite size.
- Posts appear on the newsfeeds of followers and replace a random item.
- Repost uniformly at random amongst newsfeed items.

Giovanidis, A., Baynat, B., Magnien, C., Vendeville, A.: Ranking online social users by their influence. IEEE/ACM Transactions on Networking 29(5), 2198–2214 (2021)

Balance of opinions on newsfeeds

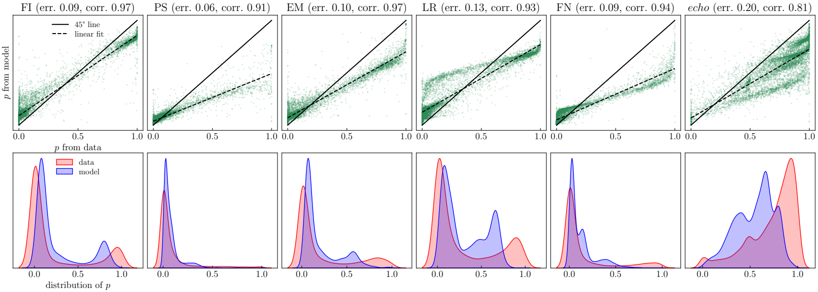
At equilibrium $p_s^{(1)}, \ldots, p_s^{(N)}$ are solution of the following linear system:

for $n = 1, \ldots, N$,

$$p_s^{(n)} \sum_{k \in \mathcal{L}^{(n)}} (\lambda^{(k)} + \mu^{(k)}) = \sum_{k \in \mathcal{L}^{(n)}} (\lambda_s^{(k)} + \mu^{(k)} p_s^{(k)}). \qquad (2)$$

- Assuming the user graph is strongly connected and at least one user has $\lambda > 0$, the system has a unique solution.
- Computed via power iteration.

Empirical evaluation



Method to increase diversity

Goal: maximise average diversity of content on the newsfeeds.

Method to increase diversity

Goal: maximise average diversity of content on the newsfeeds.

Method: insert posts into the newsfeeds (recommendation).

Method to increase diversity

Goal: maximise average diversity of content on the newsfeeds.

Method: insert posts into the newsfeeds (recommendation).

- $x_s^{(n)}$: rate at which we insert posts from party s into n's newsfeed
- B budget: no more than a proportion B of recommended content on newsfeeds

Method to increase diversity

Goal: maximise average diversity of content on the newsfeeds.

Method: insert posts into the newsfeeds (recommendation).

- $x_s^{(n)}$: rate at which we insert posts from party s into n's newsfeed
- B budget: no more than a proportion B of recommended content on newsfeeds

Objective: find $x_s^{(n)}$ for all n, s to maximise average diversity under budget B.

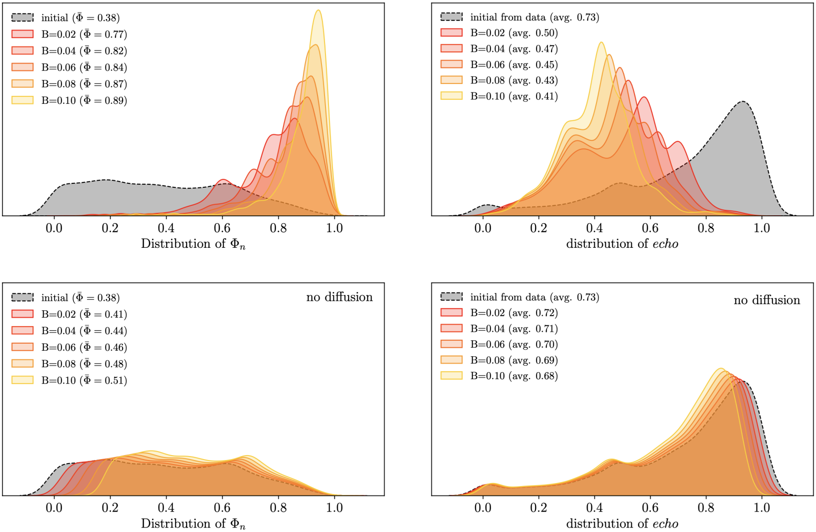
Optimisation problem

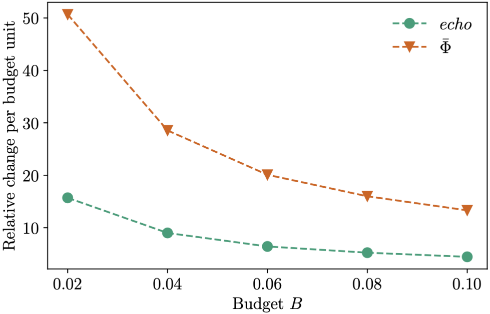
$$
\begin{aligned}
\underset{x,p}{\operatorname{argmax}} \quad & \frac{1}{N} \sum_{n} \Phi_n \\
\text{s.t.} \quad & \text{for all } n, s: \\
& \underbrace{\frac{p_s^{(n)}}{1-B} \sum_{k \in \mathcal{L}^{(n)}} (\lambda^{(k)} + \mu^{(k)}) = x_s^{(n)} + \sum_{k \in \mathcal{L}^{(n)}} (\lambda_s^{(k)} + \mu^{(k)} p_s^{(k)}),}_{\textit{model equation}} \\
& \underbrace{\sum_{s} x_s^{(n)} = \frac{B}{1-B} \sum_{k \in \mathcal{L}^{(n)}} (\lambda^{(k)} + \mu^{(k)}),}_{\textit{budget constraint}} \\
& x_s^{(n)}, p_s^{(n)} \geq 0.
\end{aligned}
$$

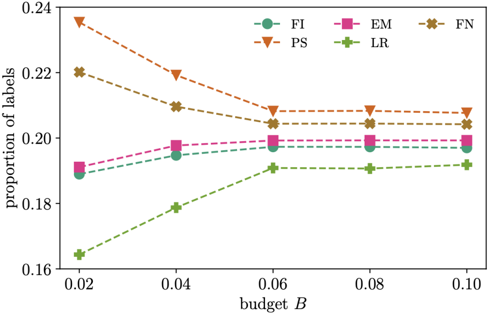
Optimisation problem

- quadratic objective with linear constraints
- 83K variables
- 50K constraints
- Gurobi solver (barrier algorithm)
- runtime $\sim$10min

Now let's see the results...


initial ($\bar{\Phi} = 0.38$)
B=0.02 ($\bar{\Phi} = 0.77$)
B=0.04 ($\bar{\Phi} = 0.82$)
B=0.06 ($\bar{\Phi} = 0.84$)
B=0.08 ($\bar{\Phi} = 0.87$)
B=0.10 ($\bar{\Phi} = 0.89$)
0.0 0.2 0.4 0.6 0.8 1.0
Distribution of Φ_n
initial from data (avg. 0.73)
B=0.02 (avg. 0.50)
B=0.04 (avg. 0.47)
B=0.06 (avg. 0.45)
B=0.08 (avg. 0.43)
B=0.10 (avg. 0.41)
0.0 0.2 0.4 0.6 0.8 1.0
distribution of *echo*
initial ($\bar{\Phi} = 0.38$)
B=0.02 ($\bar{\Phi} = 0.41$)
B=0.04 ($\bar{\Phi} = 0.44$)
B=0.06 ($\bar{\Phi} = 0.46$)
B=0.08 ($\bar{\Phi} = 0.48$)
B=0.10 ($\bar{\Phi} = 0.51$)
no diffusion
0.0 0.2 0.4 0.6 0.8 1.0
Distribution of Φ_n
initial from data (avg. 0.73)
B=0.02 (avg. 0.72)
B=0.04 (avg. 0.71)
B=0.06 (avg. 0.70)
B=0.08 (avg. 0.69)
B=0.10 (avg. 0.68)
no diffusion
0.0 0.2 0.4 0.6 0.8 1.0
distribution of *echo*

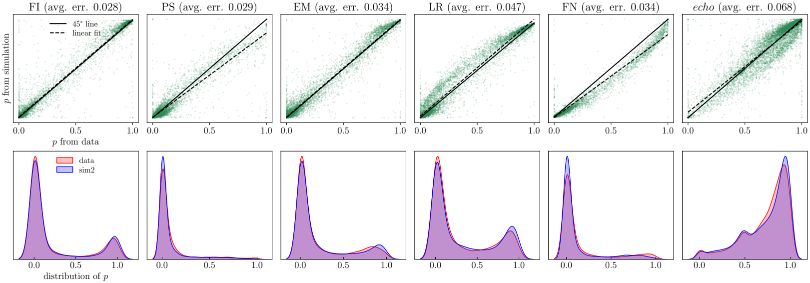

echo
$\bar{\Phi}$
Relative change per budget unit
Budget B
0.02
0.04
0.06
0.08
0.10
10
20
30
40
50


FI
EM
FN
PS
LR
proportion of labels
budget B
0.16
0.18
0.20
0.22
0.24
0.02
0.04
0.06
0.08
0.10

Further research

- Model accuracy vs empirical values...
- Backfire effect: limit the amount of cross-cutting content?
- enforce equality in the share of recommendations dedicated to each party
- other methods: content filtering, users recommendations...

Model simulation with preferential reposting



Thank you!

Budget constraint

$$\sum_s x_s^{(n)} = B\left(\sum_s x_s^{(n)} + \sum_{k\in\mathcal{L}^{(n)}} (\lambda^{(k)} + \mu^{(k)})\right) \tag{3}$$

$$\implies \sum_s x_s^{(n)} = \frac{B}{1-B} \sum_{k\in\mathcal{L}^{(n)}} (\lambda^{(k)} + \mu^{(k)}) \tag{4}$$

Model equations

$$p_s^{(n)} \left(\sum_s x_s^{(n)} + \sum_{k \in \mathcal{L}^{(n)}} (\lambda^{(k)} + \mu^{(k)}) \right) = x_s^{(n)} + \sum_{k \in \mathcal{L}^{(n)}} (\lambda_s^{(k)} + \mu^{(k)} p_s^{(k)}) \tag{5}$$

$$\implies \frac{p_s^{(n)}}{1 - B} \sum_{k \in \mathcal{L}^{(n)}} (\lambda^{(k)} + \mu^{(k)}) = x_s^{(n)} + \sum_{k \in \mathcal{L}^{(n)}} (\lambda_s^{(k)} + \mu^{(k)} p_s^{(k)}) \tag{6}$$