Depolarising Social Networks: Optimisation of Exposure to Adverse Opinions in the Presence of a Backfire Effect

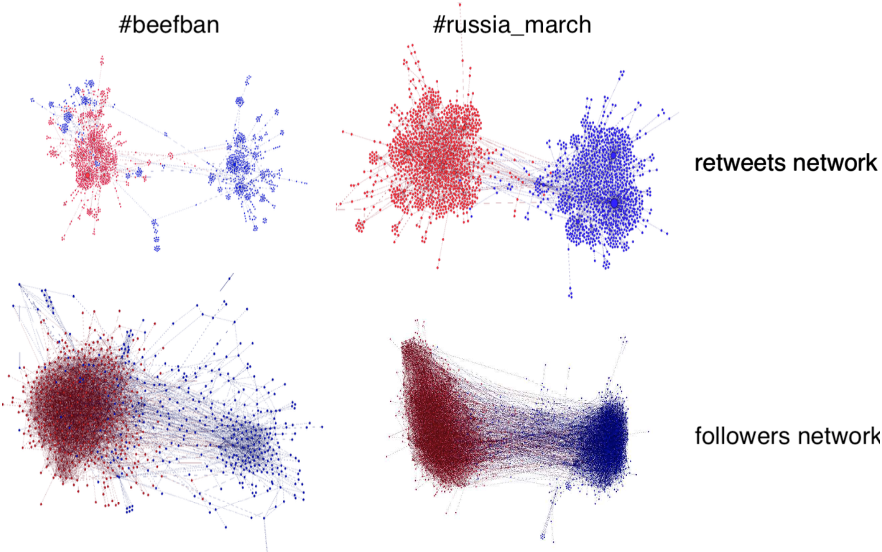
Antoine Vendeville   Benjamin Guedj   Shi Zhou

Department of Computer Science
Centre for Doctoral Training in Cybersecurity
University College London

April 12th, 2022

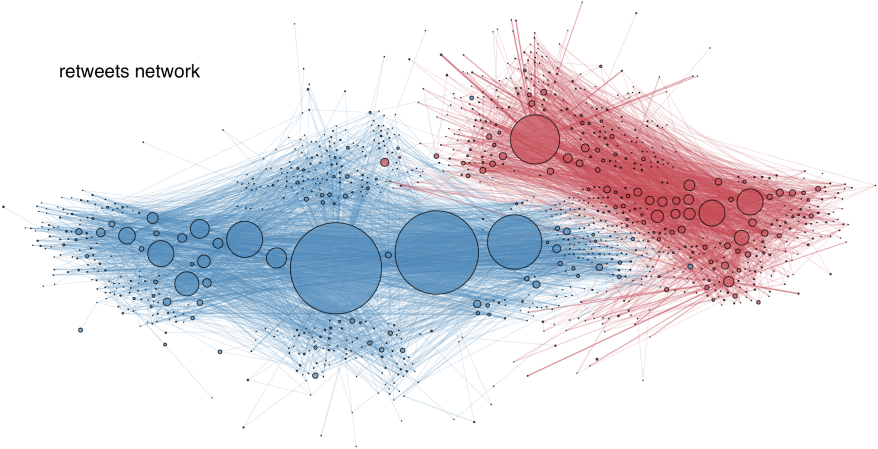
Outline

Polarisation and echo chambers (1)



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

Polarisation and echo chambers (2)



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`

Objective

Can we open up echo chambers by promoting diversity of opinions and cross-cutting links?

Constraint: backfire effect.

Outline

Setting

- social network
- N users
- individual opinions in $\{0, 1\}$

Original Model

Setting

- social network
- N users
- individual opinions in $\{0, 1\}$

Dynamics

At each step, an user chosen uniformly at random adopts the opinion of a random neighbour of them.

If the graph is connected, everyone eventually agrees.

Framework

- Connected, directed, weighted user graph.
- (z_0, z_1) zealots: in-degree 0.
- $F = N - z_0 - z_1$ free users.

Voter Model with Zealots

Framework

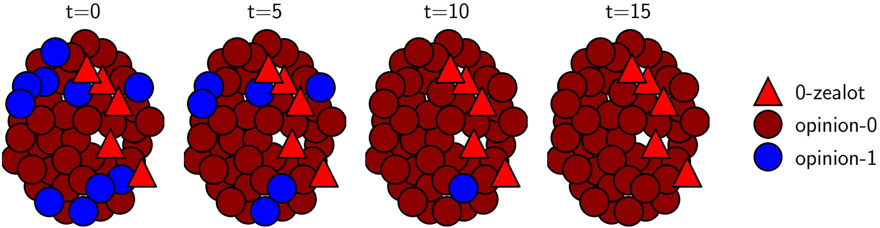
- Connected, directed, weighted user graph.
- (z_0, z_1) zealots: in-degree 0.
- $F = N - z_0 - z_1$ free users.

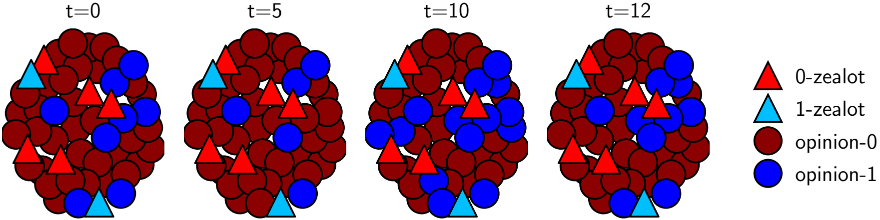
Long-time behaviour

1. $z_0 > 0, z_1 = 0 \Rightarrow$ almost surely everyone eventually has opinion 0.
2. $z_0 = 0, z_1 > 0 \Rightarrow$ almost surely everyone eventually has opinion 1.
3. $z_0 > 0, z_1 > 0 \Rightarrow$ equilibrium state.

We assume $z_0 + z_1 > 0$.

Example


t=0
t=5
t=10
t=15
0-zealot
opinion-0
opinion-1


t=0
t=5
t=10
t=12
0-zealot
1-zealot
opinion-0
opinion-1

Average opinion at equilibrium

Directed, weighted graph (Masuda, 2015)

For a free user i

$$x_i^* = \frac{\sum_{j \in \mathcal{F}} w_{ij} x_j^* + z_{1,i}}{d_i}. \quad (1)$$

Average opinion at equilibrium

Directed, weighted graph (Masuda, 2015)

For a free user i

$$x_i^* = \frac{\sum_{j \in \mathcal{F}} w_{ij} x_j^* + z_{1,i}}{d_i}. \quad (1)$$

Complete, unweighted graph

$$x_i^* = \frac{z_1}{z_0 + z_1}. \quad (2)$$

Goal

z_0 is given. Find the optimal $z_1 \geqslant 0$ to maximise either

1. Average diversity of opinions, or
2. Average proportion of active links.

Backfire effect: αz_1 free nodes will become 0-zealots $(\alpha \leqslant 1)$.

Outline

Equilibrium diversity

Average diversity of opinion at equilibrium

$$\sigma = 4\bar{x}^*(1 - \bar{x}^*) \in [0, 1] \tag{3}$$

where

$$\bar{x}^* = \frac{1}{N}\mathbf{1}^\top[L + \mathrm{diag}(Z_0 + Z_1)]^{-1} Z_1 \mathbf{1} + \frac{z_1}{N} \tag{4}$$

is the average opinion at equilibrium over the whole network.

Maximisation

Given z_0, there exists z_1 for which $\sigma = 1$.

Equilibrium diversity (complete network)

Average diversity of opinion at equilibrium

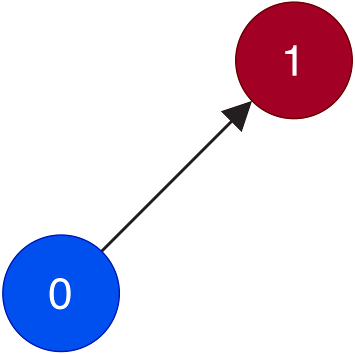
$$\sigma = \frac{4 z_0 z_1}{(z_0 + z_1)^2}. \tag{5}$$

Maximisation

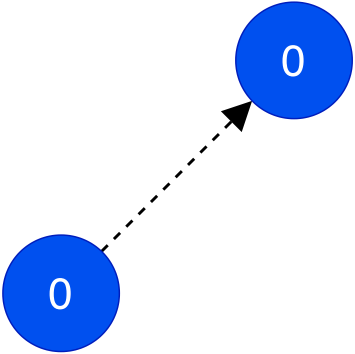
Given z_0 and backfire effect α, we find $z_1^\star$ that maximises σ via $\frac{\partial \sigma}{\partial z_1} = 0$.

Outline

Active links



active link



inactive link

Density of active links at equilibrium

q_{ij}: probability of finding $j \rightarrow i$ active.

Density of active links at equilibrium

q_{ij}: probability of finding $j \to i$ active.

Single edge

A mean-field approximation of q is given by the solution of a linear system that depends on edge weights and equilibrium opinions x^*.

$$q_{ij}(d_i + d_j) - \sum_{k \in \mathcal{F} \setminus \{i,j\}} (w_{ik} q_{jk} + w_{jk} q_{ik}) = \tilde{z}_j x_i^* + \tilde{z}_i x_j^* + z_{1,i} + z_{1,j} \qquad (6)$$

where $\tilde{z}_k = z_{0,k} - z_{1,k}$.

Density of active links at equilibrium

q_{ij}: probability of finding $j \to i$ active.

Single edge

A mean-field approximation of q is given by the solution of a linear system that depends on edge weights and equilibrium opinions x^*.

$$q_{ij}(d_i + d_j) - \sum_{k \in \mathcal{F} \setminus \{i,j\}} (w_{ik} q_{jk} + w_{jk} q_{ik}) = \tilde{z}_j x_i^* + \tilde{z}_i x_j^* + z_{1,i} + z_{1,j} \quad (6)$$

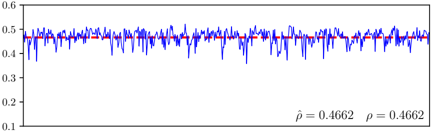
where $\tilde{z}_k = z_{0,k} - z_{1,k}$.

Density of active links

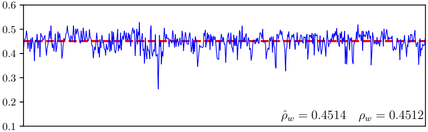
ρ: average of all the q_{ij}.
ρ_w: weighted average.

Numerical example

Erdös-Rényi graph with random uniform weights

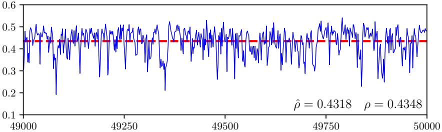


ρ

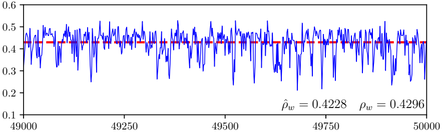


ρ_w

Barabasi-Albert graph with random exponential weights



ρ



ρ_w

Maximisation on complete networks

On a complete, unweighted user graph:

$$\rho = \frac{2z_0 z_1 (N - z_0 - z_1)}{(N-1)(z_0 + z_1)(z_0 + z_1 + 1)}. \tag{7}$$

Maximisation

Given z_0 and backfire effect α, we find $z_1^\star$ that maximises ρ via $\frac{\partial \rho}{\partial z_1} = 0$.

Outline

Dataset

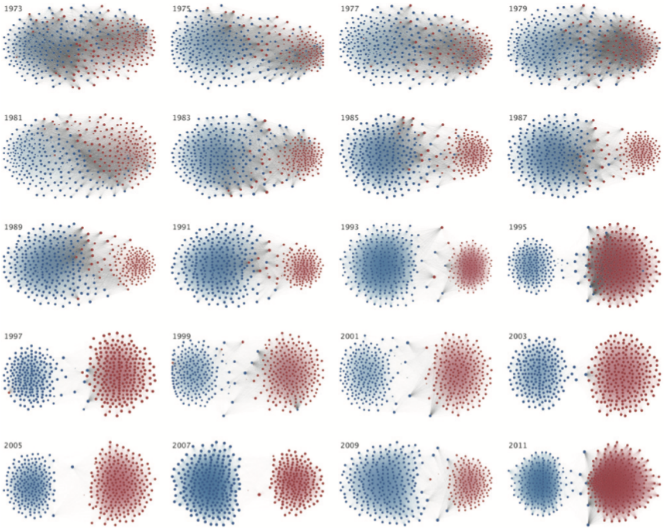
- Composition of the US House of Representatives since 1947.
- $K = 38$ Houses total.
- D_k: number of Democrat seats in the k^{th} House.
- R_k: number of Republican seats in the k^{th} House.

Fundamental hypothesis

Assume D_k, R_k correspond to punctual observations of the Voter Model.

$$D_k + R_k > 95\% \text{ for all } k.$$

Illustration



Adris *et al.*(2015). The Rise of Partisanship and Super-Cooperators in the U.S. House of Representatives. PLOS ONE 10(4): e0123507. `https://doi.org/10.1371/journal.pone.0123507`

Number of zealots

$$(\hat{z}_D, \hat{z}_R) = \underset{(z_D, z_R) \in Z}{\text{argmin}} \quad \frac{|\hat{\sigma} - \sigma| + |\hat{\rho} - \rho|}{2} \tag{Q}$$

Empirical opinion diversity

$$\hat{\sigma} = \frac{4}{K} \sum_{k=1}^{K} \frac{D_k R_k}{(D_k + R_k)^2}. \tag{8}$$

Empirical active links density

$$\hat{\rho} = \frac{1}{K} \sum_{k=1}^{K} \frac{2 D_k R_k - D_k z_R - R_k z_D}{N_k (N_k - 1)}. \tag{9}$$

Solve by exhaustive search.

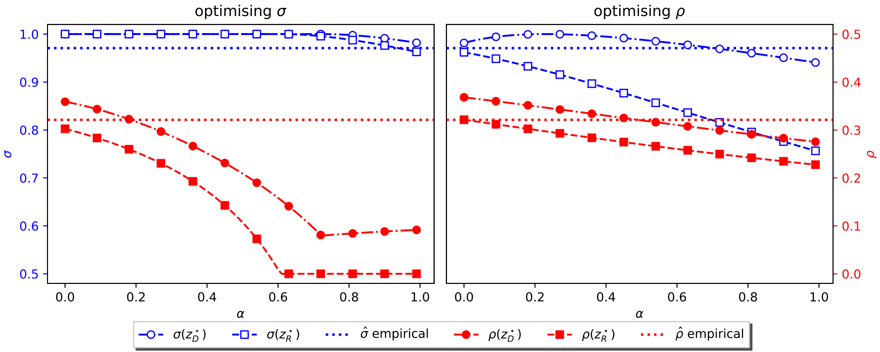
Number of zealots

$$(\hat{z}_D, \hat{z}_R) = \underset{(z_D, z_R) \in Z}{\mathrm{argmin}} \quad \frac{|\hat{\sigma} - \sigma| + |\hat{\rho} - \rho|}{2} \tag{Q}$$

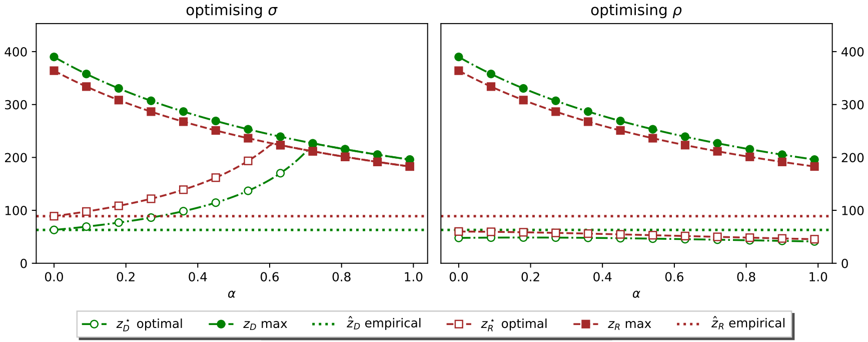
Estimates

- $(\hat{z}_D, \hat{z}_R) = (89, 63)$
- $(\hat{\sigma}, \hat{\rho}) = (0.97, 0.32)$
- $\frac{|\hat{\sigma}-\sigma|+|\hat{\rho}-\rho|}{2} = 3.8 \times 10^{-5}$

Maximisation of σ and ρ



Maximisation of σ and ρ



Conclusion

What we did...

- Closed formula for the average opinion diversity at equilibrium.
- Closed formula for the average density of active links at equilibrium.
- How to maximise them on a complete graph.
- Application to an American politics dataset.

What's next...

- Maximisation on directed, weighted, non-complete user graphs.
- Optimise on the edge weights $\equiv$ platform recommendations.
- Data from online social networks.