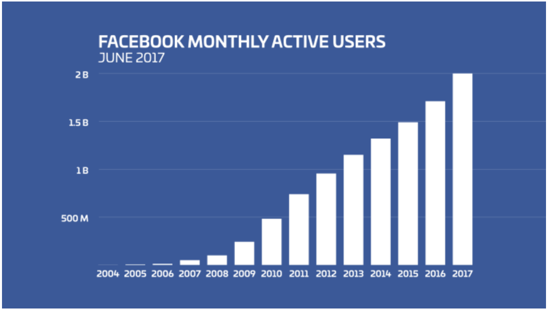
Minimizing Polarization and Disagreement in Social Networks

Cameron Musco Chris Musco **Charalampos E. Tsourakakis**
MIT MIT Boston University

WWW 2018

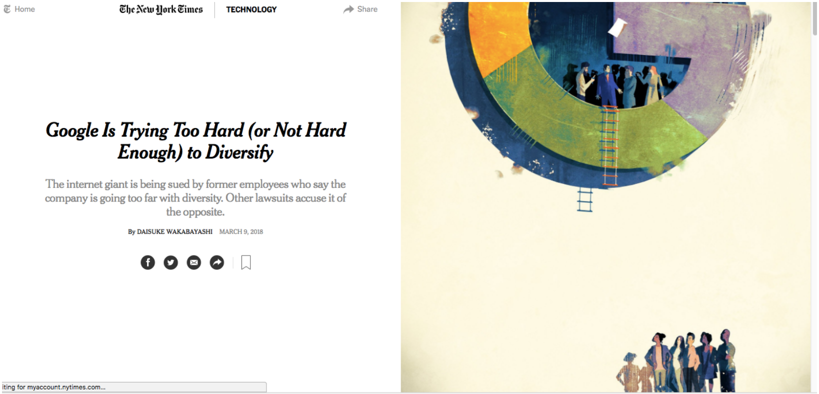
April 25th, 2018

Online social media



Online social media

- **Fierce** debates take place online



Home

The New York Times | TECHNOLOGY

Share

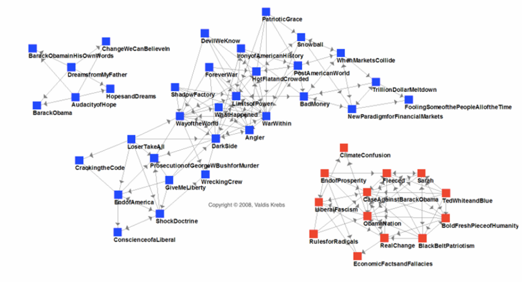
Google Is Trying Too Hard (or Not Hard Enough) to Diversify

The internet giant is being sued by former employees who say the company is going too far with diversity. Other lawsuits accuse it of the opposite.

By DAISUKE WAKABAYASHI MARCH 9, 2018

iting for myaccount.nytimes.com...

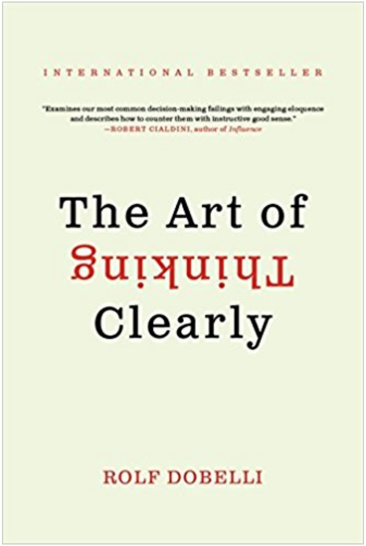
Human biases

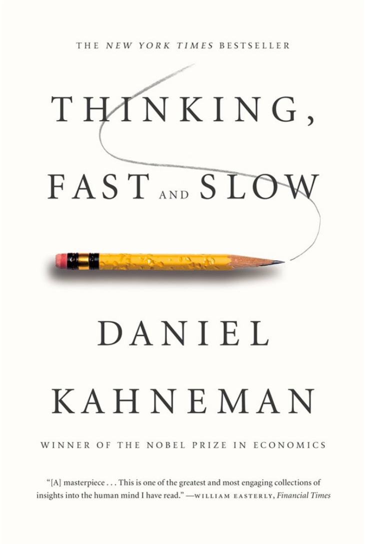


Source: Valdis Krebs, `http://www.orgnet.com/divided.html`

- Political books co-purchase graph
- Three connected components, corresponding to the two big parties
  - Those who buy books for Obama don't other political books

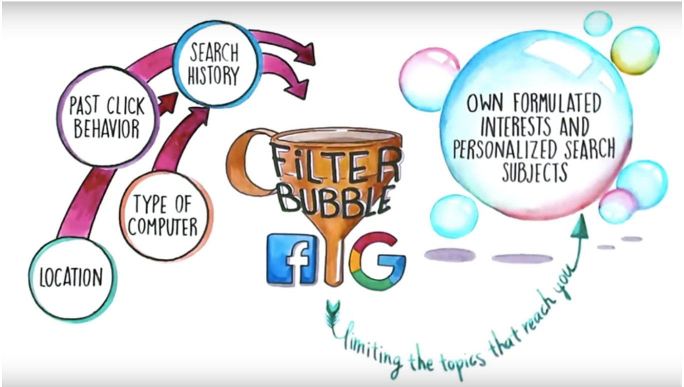
(Human biases)



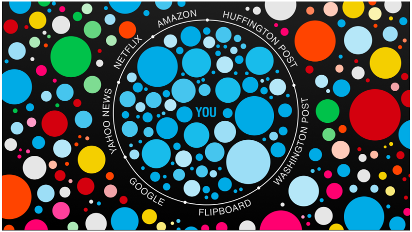


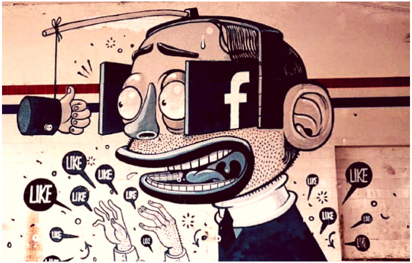
Filter bubbles

Mark Zuckerberg principle: "A squirrel dying in front of your house may be more relevant to your interests right now than people dying in Africa."



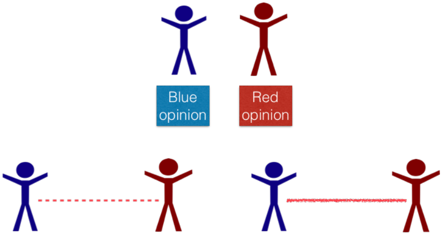
Filter bubble and echo chambers





Echo chamber: A situation in which information, ideas, or beliefs are amplified or reinforced by communication and repetition inside a defined system.

Disagreement and Polarization



- Suppose we have two humans with two opposite opinions on a certain topic.
- **Question:** Should we recommend a link between the two?
- Approach 1: No! No disagreement is caused between the two!
- Approach 2: Yes! Through an exchange of arguments they may end up approaching each other, i.e., become less polarized!

Opinion Dynamics

- Opinion dynamics model social learning processes.
- Survey by Mossel and Tamuz [Mossel et al., 2017]
- DeGroot model: Describes how a set of individuals can reach consensus [DeGroot, 1974]

Setup

- Social network $G(V, E, w)$.
- Node opinions at time 0: $s : V \rightarrow [0, 1]$.
- Basic idea: People re-peatedly average their neighbors actions
- Convergence guaranteed. [1]
- For more, see tutorial by Garimella, Morales, Gionis, Mathioudakis `http://gvrkiran.github.io/polarization/`

[1] The underlying Markov chain is irreducible and a-periodic.

Opinion Dynamics

- **Friedkin Johnsen model**: Each node i maintains a persistent internal opinion s_i.
- **Social network**: $G(V, E, w)$, where $w_{ij} \geq 0$ is the weight on edge $(i, j) \in E$ and $N(i)$ denotes the neighborhood of node i
- **Repeated averaging**: Each node i updates its expressed opinion z_i

$$z_i = \frac{s_i + \sum_{j \in N(i)} w_{ij} z_j}{1 + \sum_{j \in N(i)} w_{ij}}.$$

- **Equilibrium**: $z^* = (I + L)^{-1} s$

Key Question

- Given n agents, each with its own initial opinion that reflects its core value on a topic,
- and an opinion dynamics model (**Friedkin Johnsen model**)
- **what** is the structure of a connected social network with a given total edge weight that minimizes *polarization* and *disagreement* simultaneously?

Formalizing the key question

- Disagreement of an edge (u, v): squared difference between the opinions of u, v at equilibrium: $d(u, v) = w_{uv}(z_u^* - z_v^*)^2$
  We define **total disagreement** $D_{G,s}$ as:

$$D_{G,s} = \sum_{(u,v)\in E} d(u, v). \quad \text{[Disagreement]} \qquad (1)$$

- **Polarization**: Let $\bar{z} = z^* - \frac{{z^*}^T \vec{1}}{n}\vec{1}$. Then the polarization $P_{G,s}$ is defined to be:

$$P_{G,s} = \sum_{u\in V} \bar{z}_u^2 = \bar{z}^T \bar{z} \quad \text{[Polarization]} \qquad (2)$$

Polarization-Disagreement index

Polarization-Disagreement index is the objective we care about.

$$\mathcal{I}_{G,s} = P_{G,s} + D_{G,s} \quad \text{[Polarization-Disagreement index]}$$

Example.

- Three agents, with innate opinions $s = [0, 0, 1]$.
- We wish to recommend one link with weight 1 between these three agents.

Recommended link	$P_{G,s}$	$D_{G,s}$	$\mathcal{I}_{G,s}$
$(1, 2)$	0.667	0	0.667
$(1, 3)$	0.111	0.222	0.333
$(2, 3)$	0.111	0.222	0.333

Some observations

Recall, $\bar{z}$ is the centered equilibrium vector. We make some important observations:

- Observation 1: $D_{G,s} = \sum_{(u,v)\in E} w_{uv}(\bar{z}_u - \bar{z}_v)^2$.
- Observation 2: $D_{G,s} = {z^*}^T L z^* = \bar{z}^T L \bar{z}$
- Observation 3: Let $\bar{s} = s - \frac{s^T\vec{1}}{n}\vec{1}$ be the mean-centered innate opinion vector. Then, $\bar{z} = (I + L)^{-1}\bar{s}$.

Formal Statement

- Given our observations 1,2,3,
- our key question becomes equivalent to the following optimization problem

$$\begin{array}{ll} \min_{L \in \mathbb{R}^{n\times n}} & \bar{z}^T \bar{z} + \bar{z}^T L \bar{z} \\ \text{subject to} & L \in \mathcal{L} \\ & Tr(L) = 2m \end{array}$$

Convexity

Lemma

The objective $\bar{z}^T\bar{z} + \bar{z}^T L\bar{z}$ is a convex *function of the edge weights in the graph G corresponding to the Laplacian L.*

- To see why, recall that $\bar{z} = (I+L)^{-1}s$, and notice that we can rewrite the objective as follows:

$$\begin{aligned}\bar{z}^T\bar{z} + \bar{z}^T L\bar{z} &= \bar{s}^T(I+L)^{-1}(I+L)^{-1}\bar{s} + \bar{s}^T(I+L)^{-1}L(I+L)^{-1}\bar{s} \\ &= \bar{s}^T(I+L)^{-1}(I+L)(I+L)^{-1}\bar{s} = \bar{s}^T(I+L)^{-1}\bar{s},\end{aligned}$$

- and $f(L) = (I+L)^{-1}$ is matrix-convex

Convexity – Algorithm

- Therefore, our problem can be solved in polynomial time using off-the-shelf first or second order gradient methods.

- We use gradient descent in our experiments. The gradient with respect to weight w_i

$$\frac{\partial s^T (I+L)^{-1} s}{\partial w_i} = -s^T (I+L)^{-1} b_i b_i^T (I+L)^{-1} s$$

- where $L = Bdiag(w)B^T$, and b_i be the i-th column of B.

Non-convexity

- Perhaps surprisingly, a slightly more general form of our objective, where one of the two terms is multiplied by any factor $\rho \geq 0$ (i.e., polarization and disagreement are weighted differently), is not convex!

Theorem

Let $\rho \bar{z}^T \bar{z} + \bar{z}^T L \bar{z}, \rho \geq 0$ be our objective. In general, the objective is a non-convex function of the edge weights.

Proof by counterexample!

Sparsity lemma

Theorem

There always exists a graph G with $O(n/\epsilon^2)$ edges that achieves polarization-disagreement index $\mathcal{I}_{G,s}$ within a multiplicative $(1+\epsilon+O(\epsilon^2))$ factor of optimal for our problem

- Proof based on spectral sparsifiers [Spielman and Srivastava, 2011, Spielman and Teng, 2011, Batson et al., 2012, Spielman and Teng, 2014].
- In our experiments, we use the Spielman-Srivastava sparsification that produces graphs with $O(n \log n/\epsilon^2)$ edges.

Key Question II

- Given n agents, each with its own initial opinion that reflects its core value on a topic,
- a weighted social network G
- an opinion dynamics model (**Friedkin Johnsen model**),
- and a budget $\alpha > 0$,

- **how should we change** the initial opinions using total opinion mass at most α in order to minimize *polarization* and *disagreement* simultaneously?

Formal Statement II

An initial mathematical formalization of the problem follows

$$
\begin{aligned}
\min_{ds \in \mathbb{R}^n} \quad & \bar{z}^T \bar{z} + \bar{z}^T L \bar{z} \\
\text{subject to} \quad & z^* = (I + L)^{-1}(s + ds) \\
& \bar{z} = z^* - \frac{\vec{1}^T z^*}{n} \vec{1} \\
& \vec{1}^T ds \geq -\alpha \\
& ds \leq \vec{0} \\
& s + ds \geq \vec{0}
\end{aligned}
$$

Formal Statement II

Proposition: The formulation of Key Question II is solvable in polynomial time.

Claim (details omitted): We can simplify our formulation to the following convex (quadratic form) formulation:

$$
\begin{aligned}
\text{minimize} \quad & (s + ds)^T (I + L)^{-1} (s + ds) \\
\text{subject to} \quad & ds \leq \vec{0} \\
& \vec{1}^T ds \geq -\alpha \\
& s + ds \geq \vec{0}
\end{aligned}
$$

Experimental Setup

- Datasets. We use two datasets collected by [De et al., 2014].
  1. *Twitter:* $n = 548$ nodes, and $m = 3{,}638$ edges, opinions on the Delhi legislative assembly elections of 2013
  2. *Reddit:* $n = 556$ nodes and $m = 8{,}969$ edges, topic of interest *politics*
- Preprocessing. Opinions extracted from text using NLP and sentiment analysis. Details in [De et al., 2014].
- Machine specs. All experiments were run on a laptop with 1.7 GHz Intel Core i7 processor and 8GB of main memory.
- Code. Our code was written in Matlab. Our code is publicly available at `https://github.com/tsourolampis/polarization-disagreement`.

Some findings I

Twitter		Reddit	
$\mathcal{I}_{Twitter,s}$	199.84	$\mathcal{I}_{Reddit,s}$	138.02
# Edges	3,638	# Edges	8,969
$\mathcal{I}_{G^*,s}$	30.113	$\mathcal{I}_{G^*,s}$	0.0022
# Edges	120,000	# Edges	103,050
$\mathcal{I}_{\tilde{G}^*,s}$	30.114	$\mathcal{I}_{\tilde{G}^*,s}$	0.0022
# Edges	3,455	# Edges	7,521

- Remark: Third row shows the objective and the number of edges for the sparsified optimal solution G^*

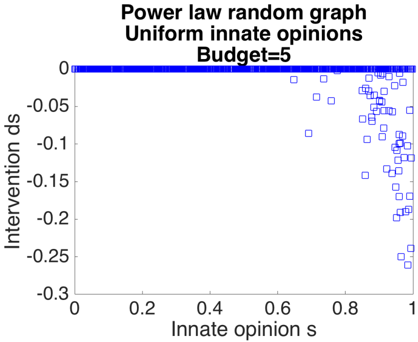
Some findings I

- Lots of controlled experiments in the paper.
- Average polarization-disagreement indices over 5×5 experiments, over 5 random innate opinion vectors and 5 random graphs.

	ER(0.5)	PL(2)	PL(2.5)	PL(3)	Proposed method L^*	Proposed method $\tilde{L}^*$-sparsified
$s \sim PL(1.5)$	14.38	16.10	22.06	53.05	11.60	11.60
$s \sim PL(2)$	25.98	45.16	72.11	107.23	19.24	19.27
$s \sim PL(2.5)$	94.87	103.62	121.21	166.38	85.55	85.56

Some findings II

- Lots of controlled experiments in the paper.



Power law random graph (slope 2), uniform innate opinions, budget $\alpha = 5$

Conclusions

Summary

- We provide the first formulation for finding an optimal topology which minimizes the sum of polarization and disagreement under a popular opinion formation model.
- We prove various facts about our objective (e.g., sparsity lemma).
- We provide efficient optimization procedures.
- We conduct both controlled experiments, and on real-world data.

Open Problems

- The same questions we asked here can be also asked using other opinion formation models.
- How good approximation is an expander graph to our objective?
- Approximate non-convex objective?

Thank you! Questions?

email: babis@seas.harvard.edu

github: https://github.com/tsourolampis

web page: http://tsourakakis.com

project web page:
https://tsourakakis.com/opinion-dynamics/

references I

Batson, J., Spielman, D. A., and Srivastava, N. (2012).
Twice-Ramanujan sparsifiers.
SIAM Journal on Computing, 41(6):1704–1721.

De, A., Bhattacharya, S., Bhattacharya, P., Ganguly, N., and Chakrabarti, S. (2014).
Learning a linear influence model from transient opinion dynamics.
In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, pages 401–410. ACM.

DeGroot, M. H. (1974).
Reaching a consensus.
Journal of the American Statistical Association, 69(345):118–121.

references II

Mossel, E., Tamuz, O., et al. (2017).
Opinion exchange dynamics.
Probability Surveys, 14:155–204.

Spielman, D. A. and Srivastava, N. (2011).
Graph sparsification by effective resistances.
SIAM Journal on Computing, 40(6):1913–1926.

Spielman, D. A. and Teng, S.-H. (2011).
Spectral sparsification of graphs.
SIAM Journal on Computing, 40(4):981–1025.

Spielman, D. A. and Teng, S.-H. (2014).
Nearly linear time algorithms for preconditioning and solving symmetric, diagonally dominant linear systems.
SIAM Journal on Matrix Analysis and Applications, 35(3):835–885.