Thesis notes

20th April

The Echo Chamber Problem - notation

- $G = (V, E^+, E^-)$ interaction graph
- $\mathcal{C}$ set of contents
- $C \in \mathcal{C}$ content, $\mathcal{T}_C$ set of threads associated with C. A thread $T \in \mathcal{T}_C$ is a subgraph of G
- $U \subseteq V$ subset of users, $T[U]$ subgraph of T induced by U. $|T(U)|$ is the number of edges of this subgraph

The Echo Chamber Problem - notation

- $\eta(C)$ fraction of negative edges associated with C (analogous definition for a thread T). Content (or thread) controversial if $\eta \in [\alpha, 1]$
- $\hat{\mathcal{C}} \subseteq \mathcal{C}$ set of *controversial* contents
- $\mathcal{S}_C(U)$ set of *non controversial* threads induced by U, for *controversial* contents, i.e.

$$\mathcal{S}_C(U) = \{\, T[U] \text{ s.t. } T[U] \text{ non controversial}, T \in \mathcal{T}_C, C \in \hat{\mathcal{C}}, U \subseteq V \} \tag{1}$$

The Echo Chamber Problem

Goal: given an interaction graph G, find $U \subseteq V$ maximing

$$\xi(U) = \sum_{C \in \hat{\mathcal{C}}} \sum_{T[U] \in S_C(U)} |T[U]| \tag{2}$$

The set of users maximing the expression is denoted as $\hat{U}$ and the corresponding score is $\xi(G)$

The Densest Echo Chamber Problem

Goal: given an interaction graph G, find $U \subseteq V$ maximing

$$\psi(U) = \sum_{C \in \hat{\mathcal{C}}} \sum_{T[U] \in S_C(U)} \frac{|T[U]|}{|U|} \tag{3}$$

The set of users maximing the expression is denoted as $\hat{U}$ and the corresponding score is $\psi(G)$

Unapproximable variations on the Echo Chamber problem

For any of these functions it is possible to find a number of negative edges to insert *s.t.* the problem resorts to the original one

$$\xi(U) = \sum_{C \in \mathcal{C}} \sum_{T \in \mathcal{T}_C} (\alpha - \eta(T[U]))|T[U]|$$

$$\xi(U) = \sum_{C \in \mathcal{C}} \sum_{T \in \mathcal{T}_C} (1 - \eta(T[U]))|T[U]|$$

$$\xi(U) = \sum_{C \in \mathcal{C}} \sum_{T \in \mathcal{T}_C} (1 - \eta(T[U]))^2|T[U]|$$

Same conclusion for the corresponding Densest version

A solvable Densest Echo Chamber problem (1)

Let $G = (V, E)$ be the interaction graph, $\delta(i, j)$ and $\delta^{-}(i, j)$ the sum of the edges and negative edges, respectively, between vertices v_i and v_j associated to controversial contents.

The graph $G_d = (V_d, E_d)$ is constructed as follows from G:

- for any vertex $v_i \in V$ add a corresponding vertex in V_d
- for any pair of vertices in G
  - let $\eta(i, j) := \frac{\delta^{-}(i,j)}{\delta(i,j)}$. If $\eta(i, j) \leq \alpha$ add a positive edge between v_i and v_j in G_d

Let $E_d[U]$ the set of edges induced on G_d by $U \subseteq V$. **Goal**: find U maximizing

$$\xi(U) = \frac{|E_d[U]|}{|U|} \tag{4}$$

A solvable Densest Echo Chamber problem (2)

Alternatives:

- Compute DCS-AM on G, where each snapshot corresponds to a content. Problems: graph may be too spars along contents, not solvable in polynomial time.
- Aggregate edges separately for each $T \in T_C, C \in \hat{\mathcal{C}}$, i.e. let $\delta_T(i,j)$ be $\delta(i,j)$ for the subgraph T:
  - let $\eta(T,i,j) := \frac{\delta_T^-(i,j)}{\delta_T(i,j)}$. If $\eta(T,i,j) \leq \alpha$ add a positive edge

A model for the Echo Chamber Problem

Model parameters:

- b_i, the group of each user i
- ω_{rs}^{+} and ω_{rs}^{-}, the probabilities of positive and negative edges, respectively, between users in group r and s ($\omega_{rs}^{+} + \omega_{rs}^{-} \leq 1$).

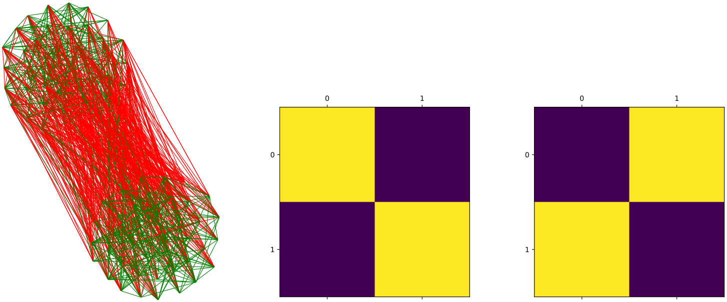
For each node pairing i, j consider their corresponding groups r and s and draw from the categorical distribution with parameters $(\omega_{rs}^{+}, \omega_{rs}^{-}, 1 - \omega_{rs}^{+} - \omega_{rs}^{-})$ to add an edge (or not).

Computing the score on the synthetic data (1)

The following graphs contains 2 communities of 40 vertices each and 10 threads, for $\alpha = 0.2$. The results have been computed with the non-exact algorithm.

Computing the score on the synthetic data (2)

First graph: mainly positive edges inside the groups and negative edges between groups.

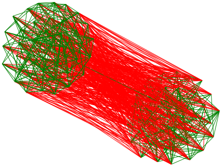


(a) Graph (b) ω_{rs}^{+} (c) ω_{rs}^{-}

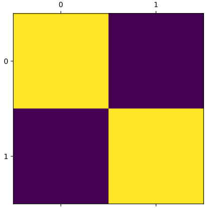
$|E| \approx 900, \ \eta(G) \approx 0.5, \ \bar{\xi}(G) = 32.$
Time (single iteration): 8.5 seconds.

Computing the score on the synthetic data (2)

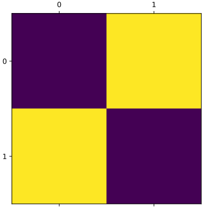
Second graph: much more positive edges inside groups and negative edges between groups.



(a) Graph



(b) ω_{rs}^{+}

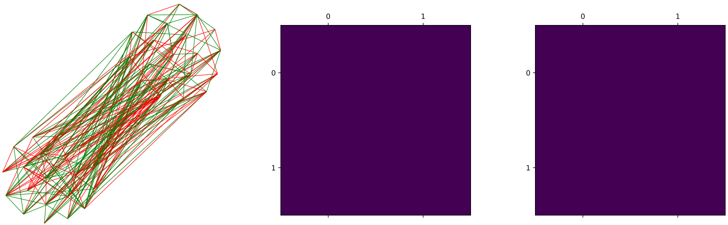


(c) ω_{rs}^{-}

$|E| \approx 900$, $\eta(G) \approx 0.5$, $\bar{\xi}(G) = 96.6$.
Time (single iteration): 8.9 seconds.

Computing the score on the synthetic data (2)

Third graph: equal distribution of positive and negative edges.



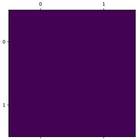
(a) Graph (b) ω_{rs}^{+} (c) ω_{rs}^{-}

$|E| \approx 830$, $\eta(G) \approx 0.5$, $\bar{\xi}(G) = 19.8$.
Time (single iteration): 8 seconds.

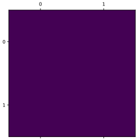
Computing the score on the synthetic data (4)

Fourth graph: positive and negative edges equally distributed in the graph but many more negative edges than positive ones.

(a) Graph



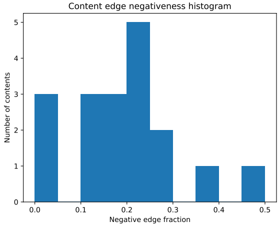
(b) ω_{rs}^{+}



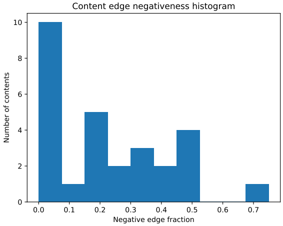
(c) ω_{rs}^{-}

$|E| \approx 800,\ \eta(G) \approx 0.75,\ \bar{\xi}(G) = 9.$
Time (single iteration): 7 seconds.

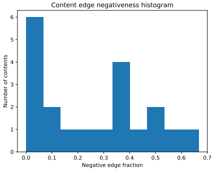
The datasets - negative edge fractions for contents



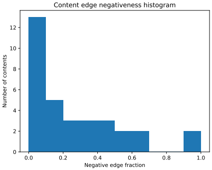
(a) @emanews



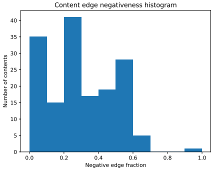
(b) @bbcscience



(c) @bbcentertainment



(d) @bbctech



(e) @bbcsport

An initial implementation - results

Table: Echo chamber scores, MIP approaches. Each results is a tuple with (score, $|U|$, number of contributing threads, time in seconds)

Source, $\lvert V\rvert$, $\lvert E\rvert$	$\xi_{MIP}(G)$	$\xi_{MIPr\_alg}(G)$	$\psi_{MIP}(G)$
@emanews, 1226, 1842	(0, 0, 0, 0.16)	(0, 0, 0, 0.14)	(0, 0, 0, 0.16)
@bbcscience, 447, 388	(7, 12, 5, 0.09)	(7, 12, 5, 0.05)	(0.75, 4, 1, 0.11)
@bbcentertainment, 220, 183	(34, 35, 6, 0.93)	(34, 35, 6, 1.14)	(1.5, 2, 1, 0.15)
@bbctech, 713, 719	(309, 295, 30, 1.5)	(309, 295, 30, 1.8)	(2.25, 4, 2, 0.76)
@bbcsport, 2140, 3100	(1030, 733, 48, 217)	(999, 723, 46, 18.6)	-