Thesis notes

13th 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.

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

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 $\xi(G)$

Computing exactly the score

$$\text{maximize} \sum_{ij \in E(T_k), T_k \in \mathcal{T}_C, C \in \hat{\mathcal{C}}} x_{ij}^k \tag{3}$$

$$x_{ij}^k \leq y_i,\ x_{ij}^k \leq y_j,\ x_{ij}^k \leq z_k \qquad \forall ij \in E(T_k), T_k \in \mathcal{T}_C, C \in \hat{\mathcal{C}} \tag{4}$$

$$x_{ij}^k \geq -2 + y_i + y_j + z_k \qquad \forall ij \in E(T_k), T_k \in \mathcal{T}_C, C \in \hat{\mathcal{C}} \tag{5}$$

$$\sum_{ij \in E^-(T_k)} x_{ij}^k - \alpha \sum_{ij \in E(T_k)} x_{ij}^k \leq -\alpha z_k \qquad \forall T_k \in \mathcal{T}_C, C \in \hat{\mathcal{C}} \tag{6}$$

$$y_i \in \{0,1\} \qquad \forall i \in V \tag{7}$$

$$0 \leq x_{ij}^k \leq 1 \qquad \forall ij \in E(T_k), T_k \in \mathcal{T}_C, C \in \hat{\mathcal{C}} \tag{8}$$

$$0 \leq z_k \leq 1 \qquad \forall T_k \in \mathcal{T}_C, C \in \hat{\mathcal{C}} \tag{9}$$

The Densest Echo Chamber Problem

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

$$\psi(U) = \sum_{C \in \hat{\mathcal{C}}} \sum_{T[U] \in S_C(U)} \frac{|T[U]|}{|U|} \tag{10}$$

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

Echo Chamber Problem inapproximability

Theorem

Echo Chamber Problem (ECP) has no $n^{1-\epsilon}$-approximation algorithm for any ϵ unless $\mathcal{P} = \mathcal{NP}$

Theorem

Densest Echo Chamber Problem (D-ECP) has no $n^{1-\epsilon}$-approximation algorithm for any ϵ unless $\mathcal{P} = \mathcal{NP}$

Solving exactly the D-ECP

$$\text{maximize} \sum_{ij \in E(T_k), T_k \in \mathcal{T}_C, C \in \hat{\mathcal{C}}} x_{ij}^k \tag{11}$$

$$x_{ij}^k \leq y_i, \quad x_{ij}^k \leq y_j \qquad \forall ij \in E(T_k), T_k \in \mathcal{T}_C, C \in \hat{\mathcal{C}} \tag{12}$$

$$a_{ij}^k \geq -2 + b_i + b_j + z_k, \quad z_k \geq a_{ij}^k \qquad \forall ij \in E(T_k), T_k \in \mathcal{T}_C, C \in \hat{\mathcal{C}} \tag{13}$$

$$\sum_{ij \in E^-(T_k)} a_{ij}^k - \alpha \sum_{ij \in E(T_k)} a_{ij}^k \leq -\alpha z_k \qquad \forall T_k \in \mathcal{T}_C, C \in \hat{\mathcal{C}} \tag{14}$$

$$\sum_{i \in V} y_i \leq 1 \tag{15}$$

$$y_i \geq 0, \quad b_i \in \{0,1\}, \quad b_i \geq y_i \qquad \forall i \in V \tag{16}$$

$$x_{ij}^k \geq 0, \quad a_{ij}^k \in \{0,1\}, \quad a_{ij}^k \geq x_{ij}^k \qquad \forall ij \in E(T_k), T_k \in \mathcal{T}_C, C \in \hat{\mathcal{C}} \tag{17}$$

$$0 \leq z_k \leq 1 \qquad \forall T_k \in \mathcal{T}_C, C \in \hat{\mathcal{C}} \tag{18}$$

Computing exactly the score

$y_i > 0 \implies b_i = 1$ means that $y_i \in U$.
$z_k = 1$ means that thread k is non controversial.
$x_{ij}^k > 0 \implies a_{ij}^k = 1$ means that the edge contributes to the score.

The following constraints enforce that only edges ij whose both vertices are in U and are associated to a non controversial thread can contribute. Also if one edge contributes to the score $\implies$ thread is not controversial.

$$x_{ij}^k \leq y_i \quad \forall ij \in E(\hat{\mathcal{C}})$$

$$x_{ij}^k \leq y_j \quad \forall ij \in E(\hat{\mathcal{C}})$$

$$a_{ij}^k \leq z_k \quad \forall ij \in E(T_k), T_k \in \mathcal{T}_C, C \in \hat{\mathcal{C}}$$

Computing exactly the score

Also, considering edges contributing to the score, they must not produce a controversial thread

$$\sum_{ij \in E^-(T_k)} a_{ij}^k - \alpha \sum_{ij \in E(T_k)} a_{ij}^k \leq -\alpha z_k \quad \forall T_k \in \mathcal{T}_C, C \in \hat{\mathcal{C}}$$

If $T_k[U]$ is controversial, then x_{ij} must be 0 $\forall ij \in E(T_k[U])$.

In general either all $ij \in E(T_k[U])$ are 0 or they are 1.

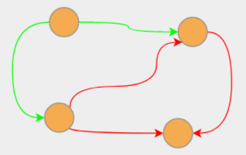
$$a_{ij}^k \leq z_k \quad \forall ij \in E(T_k), T_k \in \mathcal{T}_C, C \in \hat{\mathcal{C}}$$

$$a_{ij}^k \geq -2 + b_i + b_j + z_k \quad \forall ij \in E(T_k), T_k \in \mathcal{T}_C, C \in \hat{\mathcal{C}}$$

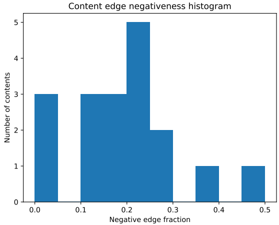
Computing exactly the score

Density introduced by

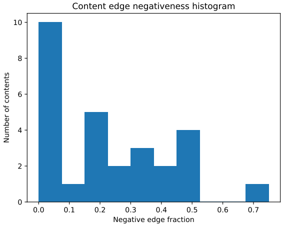
$$\sum_{i \in V} y_i \leq 1$$



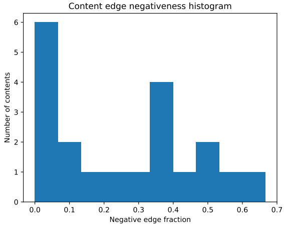
The datasets - negative edge fractions for contents



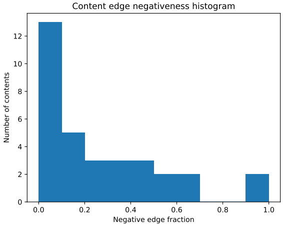
(a) @emanews



(b) @bbcscience



(c) @bbcentertainment



(d) @bbctech

An initial implementation - results

- Beta algorithm was repeated $\sqrt{n}$ times for a graph with n nodes

Table: Echo chamber scores, greedy approach

Source	|V|	|E|	$(\xi_\beta(G), \beta)$	$\xi_{peel}(G)$
@emanews	1226	1842	(0, *)	0
@bbcscience	477	388	(3, 0.9) $\boxed{2,1}$	7 $\boxed{444,4}$
@bbcentertainment	220	183	(21, 1.0)	16
@bbctech	793	719	(101, *) $\boxed{96,1}$	107 $\boxed{132,25}$

Table: Echo chamber scores, MIP approaches

Source	$\xi_{MIP}(G)$	$\xi_{MIPr}(G)$	$\xi_{MIPr\_alg}(G)$
@emanews	0	1.43	0
@bbcscience	7 $\boxed{12,5}$	10.76	7 12,5
@bbcentertainment	34 $\boxed{35,6}$	41.69	34 $\boxed{35,6}$
@bbctech	309 $\boxed{309,296}$	326.63	309 $\boxed{309,300}$

An initial implementation - timings

- Timings are reported in seconds.

Table: Echo chamber timings, greedy approach

Source	|V|	|E|	Beta	Peeling
@emanews	1226	1842	0.5	18062
@bbcscience	477	388	0.05	678
@bbcentertainment	220	183	0.1	81
@bbctech	793	719	80	4031

Table: Echo chamber timings, MIP approaches

Source	MIP	MIPr	MIPr alg
@emanews	3.77	0.08	2.8
@bbcscience	0.82	0.07	0.26
@bbcentertainment	0.79	0.09	1.04
@bbctech	2.39	1.76	0.38