Thesis notes

18th May

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.

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

The Echo Chamber Problem

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

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

where $|T^-[U]|$ and $|T^+[U]|$ denotes the number of negative and positive edges induced in the subgraph, respectively.

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

Computing exactly the score

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

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

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

$$\sum_{ij \in E^-(T_k)} x_{ij}^k - \alpha \sum_{ij \in E(T_k)} x_{ij}^k \le 0 \qquad \forall T_k \in \mathcal{T}_C, C \in \hat{\mathcal{C}} \tag{6}$$

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

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

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

For $\alpha \le 0.5$. Only the objective function needed to be corrected.

Computing exactly the score

For $\alpha > 0.5$ additional constraints and variables are needed.

$$\sum_{ij \in E^-(T_k)} a_{ij}^k - \alpha \sum_{ij \in E(T_k)} a_{ij}^k \geq -N_k z_k \qquad \forall T_k \in \mathcal{T}_C, C \in \hat{\mathcal{C}} \quad (10)$$

$$a_{ij}^k \geq -1 + y_i + y_j \qquad \forall ij \in E(T_k), T_k \in \mathcal{T}_C, C \in \hat{\mathcal{C}} \quad (11)$$

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

$$0 \leq a_{ij}^k \leq 1 \qquad \forall ij \in E(T_k), T_k \in \mathcal{T}_C, C \in \hat{\mathcal{C}} \quad (13)$$

$$0 \leq z_k \in \{0, 1\} \qquad \forall T_k \in \mathcal{T}_C, C \in \hat{\mathcal{C}} \quad (14)$$

Where a_{ij}^k encodes the information "e_{ij}^k is induced by the chosen set of vertices".

Computing exactly the score

This because for $\alpha > 0.5$ there may be non controversial threads whose total contribution to the score is negative. Without this constraint the model would prefer "considering" them as controversial, i.e. setting z_k to 0.

With the previous new constraint if the vertices induce a subgraph which is not controversial then z_k is forced to be 1, and consequently the corresponding x_{ij}^k.

N_k can be chosen to be $\alpha \cdot |E^-(T_k)|$, which is the minimum value achievable by the LHS.

A model for the Echo Chamber Problem

Each node has a group assignment and there are probabilities of positive and negative edges ω_{rs}^{+} and ω_{rs}^{+}, respectively.

1. Generate the *follow* graph G by using a SBM with parameters $\{\phi_{rs}\}$.
2. Each node can be active with probability β_a
3. Any active node activates his inactive neighbours in G with probability β_n
4. active nodes interact according to the categorical $(\omega_{rs}^{+}, \omega_{rs}^{-}, 1 - \omega_{rs}^{+} - \omega_{rs}^{-})$ otherwise (at least one of the 2 nodes is inactive) with categorical $(\theta\omega_{rs}^{+}, \theta\omega_{rs}^{-}, 1 - \theta(\omega_{rs}^{+} + \omega_{rs}^{-}))$, $\theta \leq 1$

A parametrized model (1)

Parameter choice:

$$\phi_{rs} = \begin{cases} 1 & \text{if } r = s \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

Users follow other all and only users in the same community.

$\beta_a = 1$, $\beta_n = 1$: all users interact on each post.

$$\omega^{+}_{rs} = \begin{cases} 1 - x & \text{if } r = s \\ x & \text{otherwise} \end{cases} \quad \omega^{-}_{rs} = \begin{cases} x & \text{if } r = s \\ 1 - x & \text{otherwise} \end{cases} \tag{16}$$

where the *noise* x is distributed according to a Truncated Normal distribution with $\mu = 0$, standard deviation σ and truncated within the $[0, 1]$ interval.

A parametrized model (2)

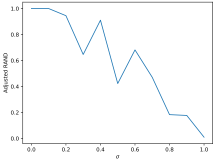
If there is no *noise* the model will produce a graph with communities corresponding to positive cliques while each node will be connected to all nodes in the other communities with negative edges.

As the noise increases interaction between users in the same community and in different communities will become less distinguishable.

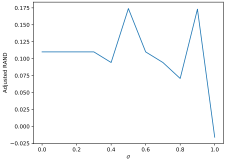
The tests have been carried on very small graphs since the exact model was also used, 10 nodes per community and 3 threads.

A parametrized model - results

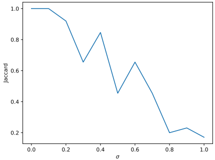
Even in absence of noise the approximation algorithm is not able to cluster all nodes correctly, differently from the exact MIP model.



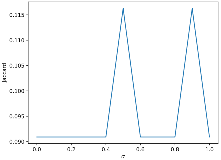
(a) Adj RAND, MIP



(b) Adj RAND, approx.



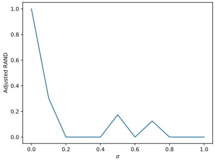
(c) Jaccard, MIP



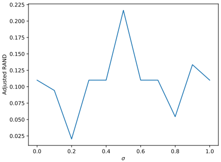
(d) Jaccard, approx.

A parametrized model - results

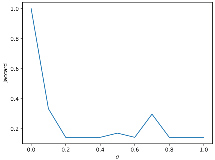
MIP is also more robust to noise than the other variants of the problem.



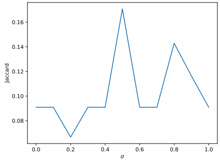
(a) Adj RAND, Densest subgraph on Threads



(b) Adj RAND, O^2-BFF



(c) Jaccard, Densest subgraph on Threads



(d) Jaccard, O^2-BFF

Solving exactly the D-ECP

$$\textit{maximize} \sum_{T_k \in \mathcal{T}_C,\ C \in \hat{\mathcal{C}}} \Big(\sum_{ij \in E^+(T_k)} x_{ij}^k - \sum_{ij \in E^-(T_k)} x_{ij}^k \Big) \tag{17}$$

$$x_{ij}^k \leq y_i, \quad x_{ij}^k \leq y_j \qquad \forall ij \in E(T_k), T_k \in \mathcal{T}_C, C \in \hat{\mathcal{C}} \tag{18}$$

$$a_{ij}^k \geq -2 + b_i + b_j + z_k, \quad z_k \geq a_{ij}^k \qquad \forall ij \in E(T_k), T_k \in \mathcal{T}_C, C \in \hat{\mathcal{C}} \tag{19}$$

$$\sum_{ij \in E^-(T_k)} x_{ij}^k - \alpha \sum_{ij \in E(T_k)} x_{ij}^k \leq 0 \qquad \forall T_k \in \mathcal{T}_C, C \in \hat{\mathcal{C}} \tag{20}$$

$$\sum_{i \in V} y_i \leq 1 \tag{21}$$

$$a_{ij}^k \leq b_i, \quad a_{ij}^k \leq b_j \quad \forall ij \in E(T_k), T_k \in \mathcal{T}_C, C \in \hat{\mathcal{C}} \tag{22}$$

$$y_i \geq 0, \quad b_i \in \{0,1\}, \quad b_i \geq y_i \qquad \forall i \in V \tag{23}$$

$$y_i \geq -1 + b_i + y_j \qquad \forall (i,j) \in V \tag{24}$$

$$x_{ij}^k \geq -1 + a_{ij}^k + y_i \quad \forall ij \in E(T_k), T_k \in \mathcal{T}_C, C \in \hat{\mathcal{C}} \tag{25}$$

$$x_{ij}^k \geq -1 + a_{ij}^k + y_j \quad \forall ij \in E(T_k), T_k \in \mathcal{T}_C, C \in \hat{\mathcal{C}} \tag{26}$$

$$x_{ij}^k \geq 0, \quad a_{ij}^k \in \{0,1\}, \quad a_{ij}^k \geq x_{ij}^k \qquad \forall ij \in E(T_k), T_k \in \mathcal{T}_C, C \in \hat{\mathcal{C}} \tag{27}$$

$$0 \leq z_k \leq 1 \qquad \forall T_k \in \mathcal{T}_C, C \in \hat{\mathcal{C}} \tag{28}$$

For $\alpha \leq 0.5$.

Solving exactly the D-ECP

$$\text{maximize} \sum_{T_k \in \mathcal{T}_C,\ C \in \hat{\mathcal{C}}} \Big(\sum_{ij \in E^+(T_k)} x_{ij}^k - \sum_{ij \in E^-(T_k)} x_{ij}^k \Big) \tag{29}$$

$$x_{ij}^k \leq y_i, \quad x_{ij}^k \leq y_j \qquad \forall ij \in E(T_k), T_k \in \mathcal{T}_C, C \in \hat{\mathcal{C}} \tag{30}$$

$$a_{ij}^k \geq -1 + b_i + b_j \qquad \forall ij \in E(T_k), T_k \in \mathcal{T}_C, C \in \hat{\mathcal{C}} \tag{31}$$

$$-N_k z_k \leq \sum_{ij \in E^-(T_k)} x_{ij}^k - \alpha \sum_{ij \in E(T_k)} x_{ij}^k \leq 0 \qquad \forall T_k \in \mathcal{T}_C, C \in \hat{\mathcal{C}} \tag{32}$$

$$\sum_{i \in V} y_i \leq 1 \tag{33}$$

$$a_{ij}^k \leq b_i, \quad a_{ij}^k \leq b_j \quad \forall ij \in E(T_k), T_k \in \mathcal{T}_C, C \in \hat{\mathcal{C}} \tag{34}$$

$$y_i \geq 0, \quad b_i \in \{0,1\}, \quad b_i \geq y_i \qquad \forall i \in V \tag{35}$$

$$y_i \geq -1 + b_i + y_j \qquad \forall (i,j) \in V \tag{36}$$

$$x_{ij}^k \geq -1 + a_{ij}^k + z_k + y_i \quad \forall ij \in E(T_k), T_k \in \mathcal{T}_C, C \in \hat{\mathcal{C}} \tag{37}$$

$$x_{ij}^k \geq -1 + a_{ij}^k + z_k + y_j \quad \forall ij \in E(T_k), T_k \in \mathcal{T}_C, C \in \hat{\mathcal{C}} \tag{38}$$

$$x_{ij}^k \geq 0, \quad a_{ij}^k \in \{0,1\}, \quad a_{ij}^k \geq x_{ij}^k \qquad \forall ij \in E(T_k), T_k \in \mathcal{T}_C, C \in \hat{\mathcal{C}} \tag{39}$$

$$0 \leq z_k \leq 1 \qquad \forall T_k \in \mathcal{T}_C, C \in \hat{\mathcal{C}} \tag{40}$$

For $\alpha > 0.5$.