1 Echo Chamber Problem

("Echo Chamber Problem" might not be the best name, but I did not come up with another one.)

We propose to use our data to find echo chambers inside social networks. The general idea is that some content is controversially discussed inside a social network, but that inside echo chambers we do not see this controversy. Thus, the general idea is as follows: Look at contents in social networks that are controversially discussed *globally* and then find subgraphs in which the same contents are discussed non-controversially. These subgraphs are the echo chambers.

We will now make this notion more formal, but it requires a few definitions.

- Suppose there is a signed, directed, multi-graph $G = (V, E^+, E^-)$. In the graph, the vertices correspond to users of a social network and the edges correspond to positive/negative interactions of the different users. We will call G the *interaction graph*.
- Consider a set of contents $\mathcal{C}$, e.g., newspaper articles that are shared on a social network.
- For each content $C \in \mathcal{C}$, there exists a set of threads $\mathcal{T}_C$. Each thread is $T \in \mathcal{T}_C$ is a subgraph of G, i.e., $T \subseteq G$, such that in the initial post of the thread a user shared the content C and then other users replied to this post. The edges are directed outwards from the replying user; the edges have signs based on whether the interaction was positive/negative. Furthermore, we assume that the whole graph is the union of the threads, i.e., $G = \bigcup_{C \in \mathcal{C}} \bigcup_{T \in \mathcal{T}_C} T$.
- Fix some $\alpha \in [0, 1]$. We denote the fraction of negative edges of a content C and thread T as $\eta(C)$ and $\eta(T)$, respectively. We now define *(non-)controversial* threads and contents:
  - Intuitively, a thread is controversial if there are many negative interactions. Formally, we say that a thread T is *controversial* if $\eta(T) \in [\alpha, 1]$. Otherwise, the thread is *non-controversial* (in this case, the thread contains mostly positive edges).
  - Intuitively, a content is controversial if it triggers a lot of negative interactions in the whole network. Formally, we say that a content C is *controversial* if the fraction of negative edges in $\bigcup_{T \in \mathcal{T}_C} T$ is in the interval $[\alpha, 1]$, i.e. $\eta(C) \in [\alpha, 1]$. Otherwise, the content is *non-controversial*. We write $\hat{\mathcal{C}} \subseteq \mathcal{C}$ to denote the set of controversial contents.
- Now our intuition is that inside *echo chambers* there should be little controversy because users have similar opinions and there is little rebuttal. In particular, echo chambers should satisfy that content is which is *globally* controversial is discussed in a non-controversial way *inside* the echo chamber.
- We will now formalize this notion. We will use the following notation, where $U \subseteq V$ is a set of vertices:
  - For a thread T, we write $T[U]$ to denote the induced subgraph of T only containing vertices in U. As above, we say that $T[U]$ is *controversial* if the fraction of negative edges in $T[U]$ is in the interval $[\alpha, 1]$.
  - We write $|T(U)|$ to denote the number of edges in $T(U)$ (irrespective of their sign).
  - For a controversial content $C \in \hat{}$

$Candasetofusers U \subseteq V$, we write $\mathcal{S}_C(U)$ to denote all threads $T[U]$ such that $T \in \mathcal{T}_C$ and $T[U]$ is non-controversial. Note that, as desired, $\mathcal{S}_C(U)$ contains only threads are *locally* non-controversial and that $\mathcal{S}_C(U)$ is only defined for contents which are *globally* controversial.

– For a set of users $U \subseteq V$, we say that its *echo chamber score* $\xi(U)$ is defined as

$$\xi(U) = \sum_{C \in}$$

C$\sum_{T[U] \in \mathcal{S}_C(U)} |T[U]|$.Note that in the first sum we only take into account contents that are globally controversial; this choice was made to ensure that "wholesome" content has no effect on the echo chamber score.

- We can now define the **Echo Chamber Problem**:
  - *Input:* An interaction graph $G = (V, E^+, E^-)$, a set of contents $\mathcal{C}$, a set of threads $\mathcal{T}_C$ for each $C \in \mathcal{C}$ and a parameter $\alpha \in [0, 1]$.
  - *Goal:* Find a set of users $U \subseteq V$ that maximizes the echo chamber score $\xi(U)$.

2 Comments

- I do not insist that this is the perfect problem formulation that we should study. I just think that it might be a starting point for a discussion.
- In the echo chamber score, we could as well say that we only count the number of positive edges in $T[U]$.
- The problem formulation has the benefit that we do not have to try to infer the leaning of users towards contents.
- We could assume that we also have access to the *follow graph*, i.e., the graph that encodes which user follows which other user. In that case, we could define communities based on the follow graph and then check which of the communities has the highest echo chamber score. We could find the communities inside the follow graph by simply using existing community detection algorithms. This would also allow us to check whether the echo chamber score is useful in practice without having to solve the optimization problem.
- A problem with the above definition of the echo chamber problem that it does not take into account the follow graph at all. For example, suppose that in a graph all users interact with each other and half them is left-wing and the other one is right-wing. Then the above problem could simply pick one of the two communities (inside which should be mostly positive interactions) and claim that it is an echo chamber. But that is not really true because (by assumption) all users interact and debate with each other regularly.
- We could also extend the definition of *controversial* and *non-controversial* to the interval $[\alpha, 1 - \alpha]$ instead of $[0, 1 - \alpha]$. This would resemble the fact that there might be threads such that all users from an echo chamber reply negatively. In other words, there is still a clear "majority" reaction. This might, however, come at the risk of finding opposing groups which contain a lot of negative edges; this would not really go too well with the usual notion of an echo chamber.

3 Approaches

For a graph G we denote with $\xi(G)$ the maximum echo chamber score, and $\hat{U}$ the corresponding set of users.

3.1 Greedy algorithm

The behaviour of the algorithm depends on the choice of β, regulating the density of the resulting set of users (for smaller values an higher density is to be expected, generally)

Algorithm 1: Greedy algorithm

$U = \{$ random node $\}$;
while $\xi(U)$ *can be increased by adding a node* **do**
  With probability β add to U the node increasing the score $\xi(U)$ the most;
  With probability $(1-\beta)$ remove from U the node contributing less to the score $\xi(U)$.
  This node will be ignored in the next iteration;
end

The process is repeated with different starting nodes.

A variant takes into a account the fraction of positive edges of a vertex in sampling the initial node (the higher the fraction, the more likely it is to be sampled).

3.2 Integer Linear Programming model

We define the following ILP model to compute exactly the maximing score for a graph G.

We denote as $E(\hat{\mathcal{C}})$ the set of edges associated to controversial contents.

$$maximize \sum_{ij \in E(\hat{\mathcal{C}})} x_{ij} \tag{1}$$

$$x_{ij} \leq y_i \quad \forall ij \in E(\hat{\mathcal{C}}) \tag{2}$$

$$x_{ij} \leq y_j \quad \forall ij \in E(\hat{\mathcal{C}}) \tag{3}$$

$$\sum_{ij \in E^-(T_k)} x_{ij} - \alpha \sum_{ij \in E(T_k)} x_{ij} \leq M_k(1 - z_k) \quad \forall T_k \in \mathcal{T}_C, C \in \mathcal{C} \tag{4}$$

$$\sum_{ij \in E(T_k)} x_{ij} \leq N_k z_k \tag{5}$$

$$x_{ij} \in \{0,1\} \quad \forall ij \in E(\hat{\mathcal{C}}) \tag{6}$$

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

$$z_k \in \{0,1\} \quad \forall T_k \in \mathcal{T}_C, C \in \mathcal{C} \tag{8}$$

The model associates to each vertex a variable y_i, to each edge a variable x_{ij} and to each thread associated to a controversial content a variable z_k. If the variable associated to a vertex is 1 then it belongs to the set of users U considered for the current solution; similarly, if the variable associated to an edge is 1 it is counted in the score (= objective function): This happens only for links induced by the chosen vertices (due to 2 and 3).

Constraints 4 and 5 exclude edges associated to controversial threads from contributing to the objective: for a non controversial thread T_k by definition it will be $\eta(T) \leq \alpha$, i.e.

$$\frac{\sum_{ij \in E^-(T_k)} x_{ij}}{\sum_{ij \in E(T_k)} x_{ij}} \leq \alpha \tag{9}$$

which can be written as

$$\sum_{ij \in E^-(T_k)} x_{ij} - \alpha \sum_{ij \in E(T_k)} x_{ij} \leq 0 \tag{10}$$

Instead, for a controversial thread

$$\sum_{ij \in E^-(T_k)} x_{ij} - \alpha \sum_{ij \in E^(T_k)} x_{ij} > 0 \tag{11}$$

So, considering again Equation 4, for a controversial thread it will necessarily be $z_k = 0$. But, because of constraint Equation 5, all the edges associated to that thread T_k needs to be 0, making that an invalid solution. This means that a solution in which edges associated to a controversial thread are set to 1 is an invalid solution.

The model finds a solution whose value of the objective function corresponds to $\xi(G)$ and the corresponding set of users is the set of vertices whose $y_i = 1$.

The choice of M_k and N_k can simply be m (the number of edges of the graph G) to produce a valid formulation.