Thesis notes

25th 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.

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

The Echo Chamber Problem

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

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

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

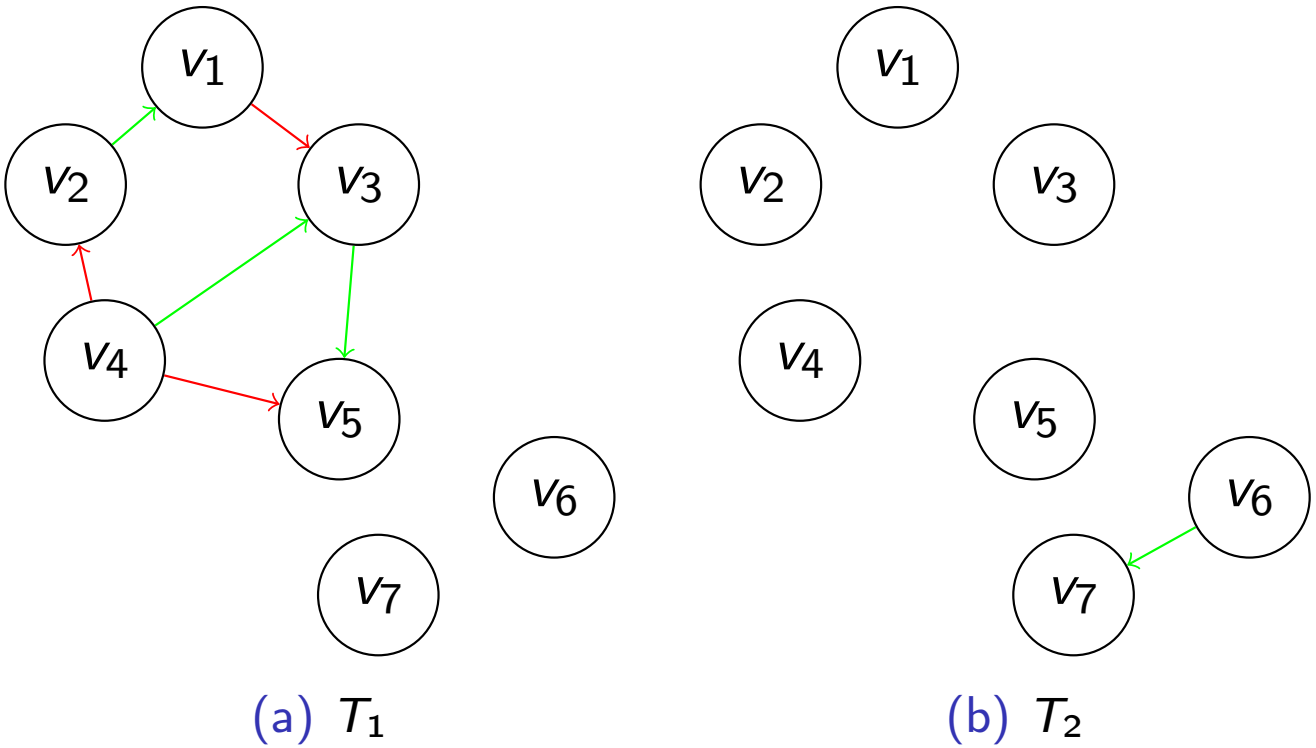
The set of users maximing the expression is denoted as $\hat{U}$ and the corresponding score is $\xi(G)$

The rounding algorithm

Algorithm 1: Rounding algorithm

$\hat{G} \leftarrow$ empty graph ;
$\hat{V} \leftarrow$ vertices of $\hat{G}$;
$S = 0$
foreach $e_{ij}^k \in \tilde{E}$ **do**
  $\hat{V} \leftarrow \hat{V} \bigcup \{v_i\}$ **if** $v_i \notin \hat{V}$;
  $\hat{V} \leftarrow \hat{V} \bigcup \{v_j\}$ **if** $v_j \notin \hat{V}$;
  $S \leftarrow \max(S, \xi(\hat{V}))$
  foreach *component* C *in* $\hat{G}$ **do**
    $S \leftarrow \max(S, \xi(C))$
  end
end
return S ;

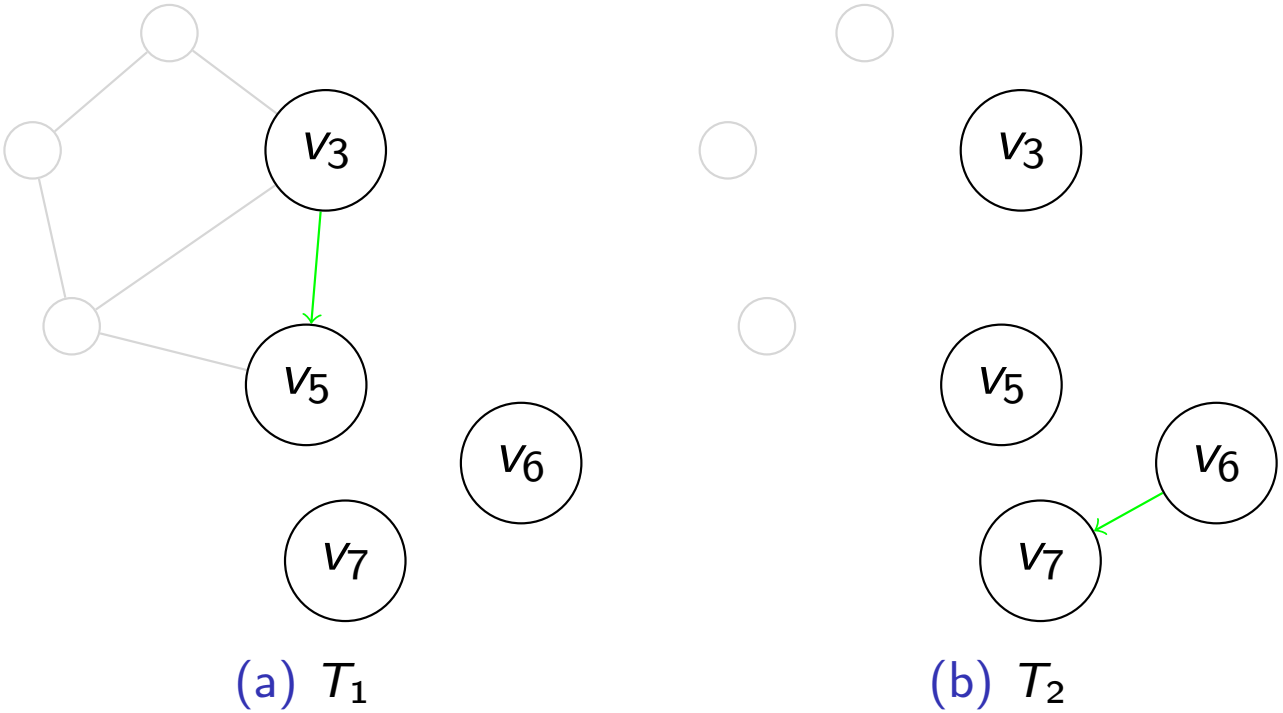
Example - Original Graph



(a) T_1 (b) T_2

Figure: Example original *Interaction Graph G*

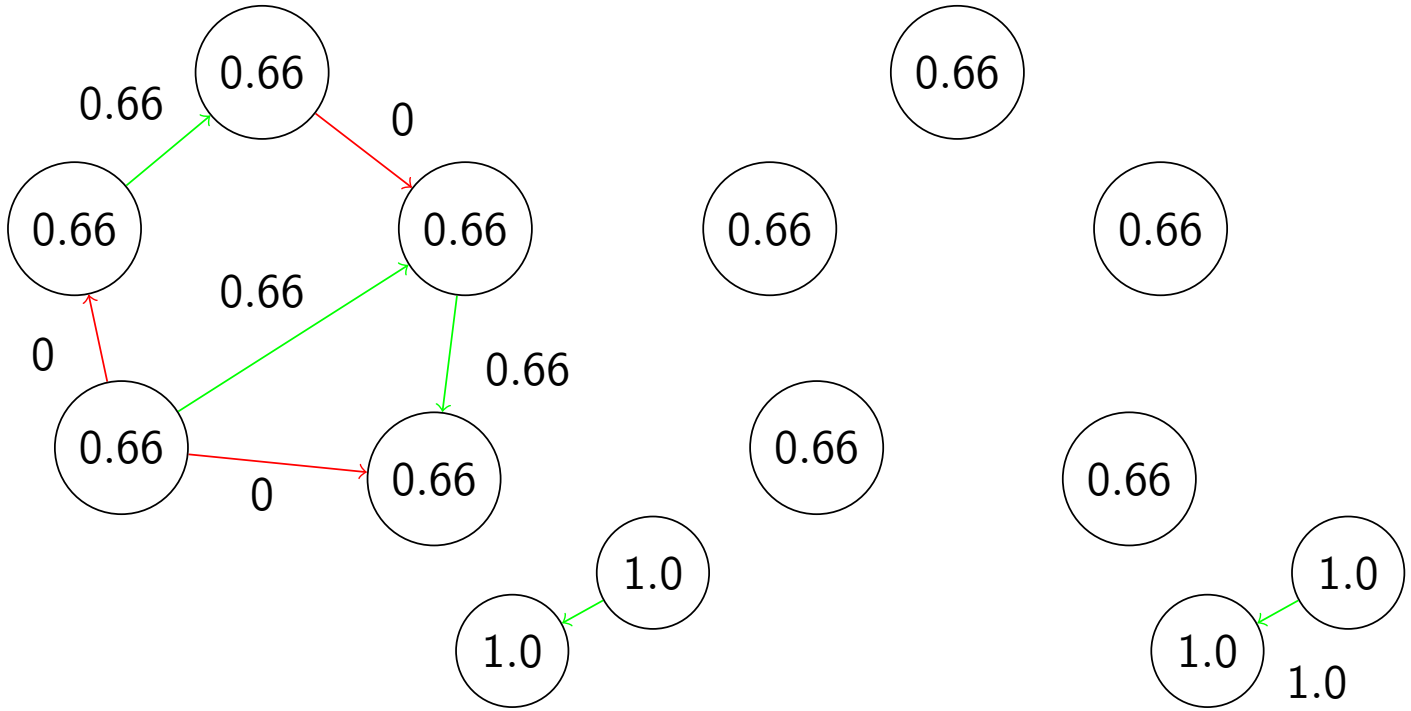
Example - Exact solution



(a) T_1 (b) T_2

Figure: Exact solution of the example in 1, $\alpha = 0.4$

Example - Relaxation solution



(a) T_1, where $z_1 = 0.66$ (b) T_2, where $z_2 = 1.0$

Figure: Solution of the relaxation of G of 1, $\alpha = 0.4$

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:

$$\phi_{rs} = \begin{cases} 1 & \text{if } r = s \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

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

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

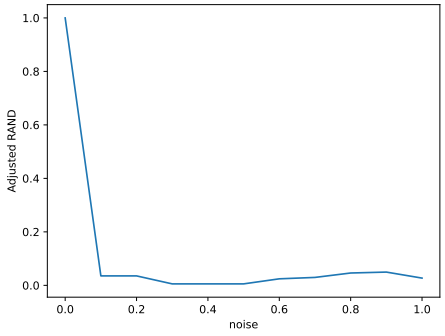
$$\omega_{rs}^{+} = \begin{cases} 1 - x & \text{if } r = s \\ x & \text{otherwise} \end{cases} \quad \omega_{rs}^{-} = \begin{cases} x & \text{if } r = s \\ 1 - x & \text{otherwise} \end{cases} \tag{4}$$

A parametrized model (2)

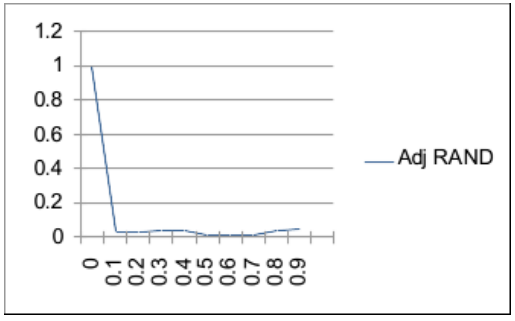
The tests have been carried on very small graphs since the exact model was also used, 10 nodes per community and 3 threads.

A parametrized model - results

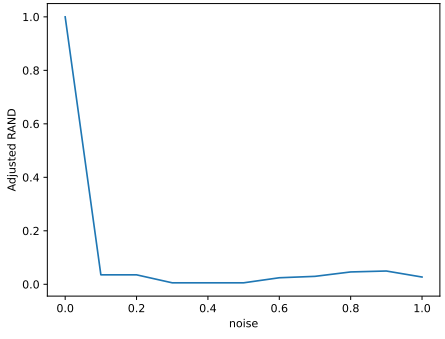
Even in absence of noise the approximation algorithm is not able to cluster all nodes correctly, differently from the exact MIP model.



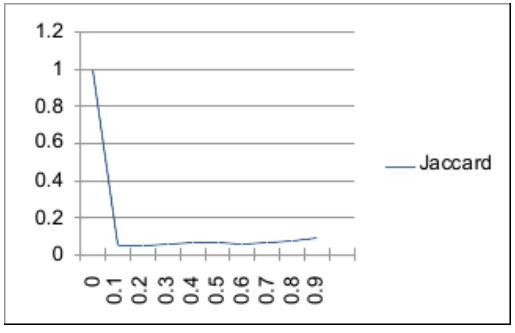
(a) Adj RAND, MIP



(b) Adj RAND, approx.



(c) Jaccard, MIP



(d) Jaccard, approx.

Analyzing @nytimes results

A graph from @nytimes, with 500 contents
α chosen as the median of the η of the contents, $\alpha = 0.44$.
The graph contains $\approx$ 83000 nodes and $\approx$ 120000 edges,
$\xi(G) \approx 30000$ on 27000 vertices and define $\approx$ 1500 components.
In the original graph there were 829 components.

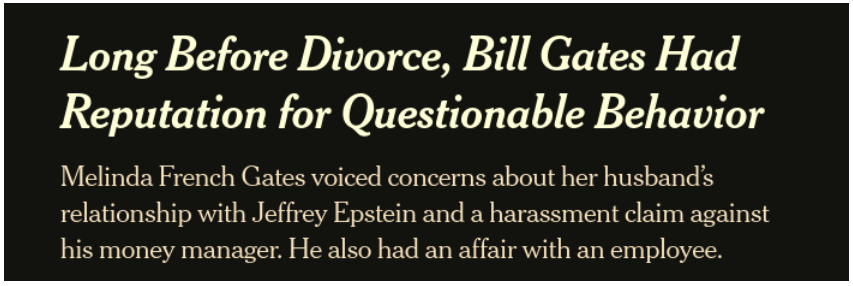
- Average shortest path length: 1.94
- Median shortest path length: 1.0
- Average degree: 1.35
- Contributing threads: 2200
- Number of threads in the graph: 6246

Analyzing @nytimes results

Contents that contribute the most to the score:

Figure: 3047

Figure: 2431



Analyzing @nytimes results - examples

It is in general not easy to easily distinguish Echo Chamber since

- components are very small and so there are very few comments
- not all edges are related to the same thread
- comments are reply to other comments which are missing in the Echo chamber, making it difficult to catch up with the discussion

The resulting communities and thread sometimes contains also noise introduced by images and people speaking languages different from English.

Analyzing @nytimes results - examples

Usually echo chambers contains 2 or 3 comments which show the same opinion while the rest of the comments and users in the chamber do not really show their alignment. Example, about Bill Gates's divorce:

1. B to A: "Well this is not the problem, the main one is his friendship with Epstein and many of his travels in the Lolita's xxxpress"
2. C to A: "Fantasme ! ?"
3. D to A: "don't worry his wife will have a big slice of his cake!"
4. E to A: "Philanthropie..."
5. F to A: "Wait till his lawyer kicks up some juicy bits"