Thesis notes

11th 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.

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

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 $\psi(G)$

Baseline datasets

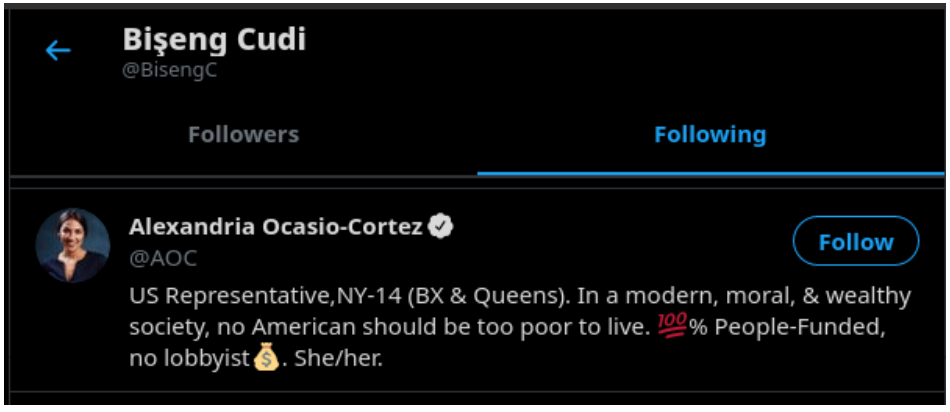
Reddit datasets:

- r/asktrumpsupporters, where $> 40\%$ of the nodes are labeled. Labels: Supporter, NonSupporter, Undecided. Missing labels are due to removed accounts and comments.
- r/debatereligion, where $\approx 1\%$ of the nodes are labeled. Labels: Christian, Muslim, ...
  Missing labels are mostly due to custom labels chosen by the user.

In both cases users declare their "position" through flairs.

Baseline datasets

- Look at the accounts a user is following



- Use the account name to look at the political party on wikipedia

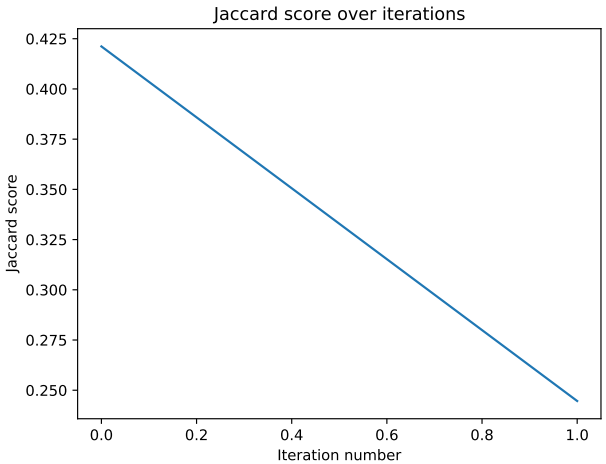
Baseline datasets

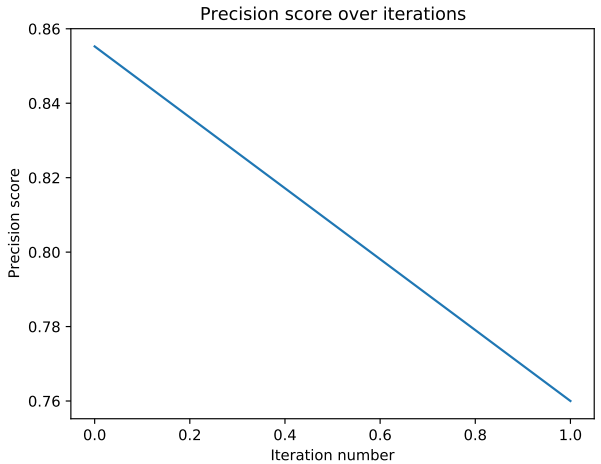
Use the party of the majority of the people a user is following for labeling him.

1. Construct the *Interaction graph* from @nytimes (or any other profile)
2. select k-core for reducing the number of users
3. label the nodes in the k-core

- For 200 contents from @nytimes the 4-core contains ≈ 1000 nodes, and around 50% of them is labeled
- of these labeled nodes, 80% is labeled as democrat and the remaining 20% as republican

Baseline results

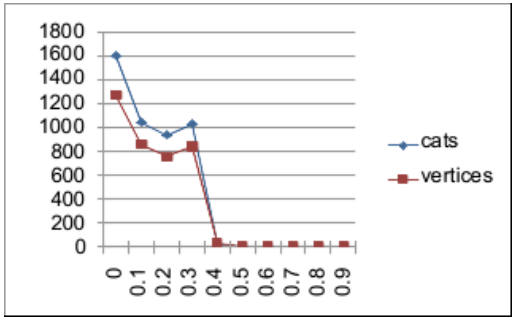




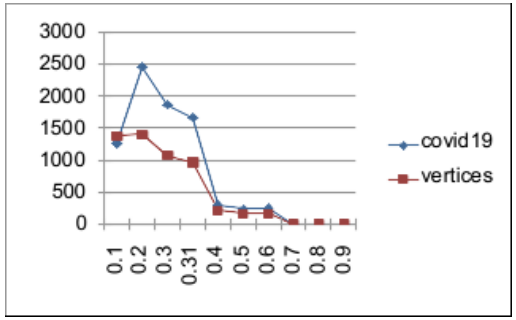
- Adjusted RAND score: 0.004424665696313976
- RAND score: 0.4451737676780167617
- Jaccard score: 0.1750972762645914 3

Dataset is skewed. Both in the first and second iteration the solution contains nodes whose majority is democrat

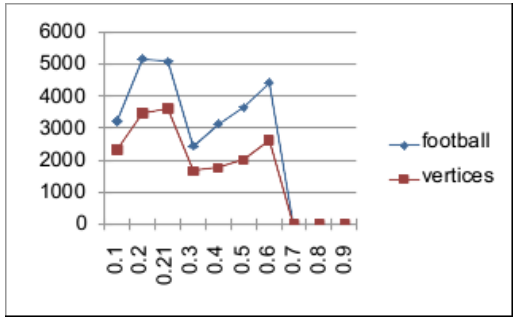
Relationship between alpha and Echo Chamber Score



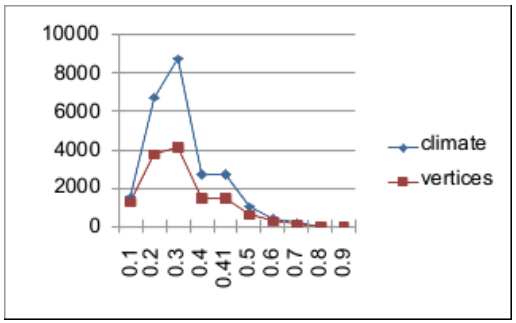
(a) r/cats



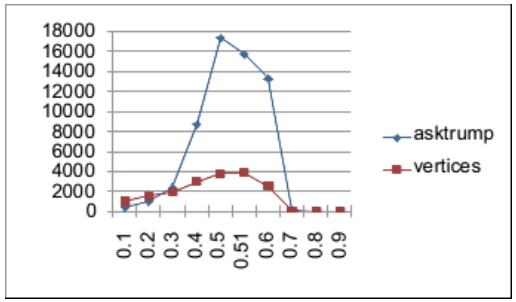
(b) r/covid19



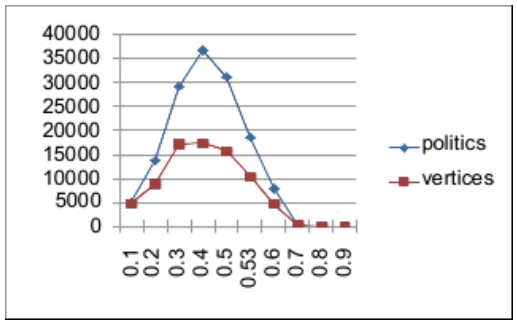
(c) r/football



(a) r/climate



(b) r/asktrumpsupporters



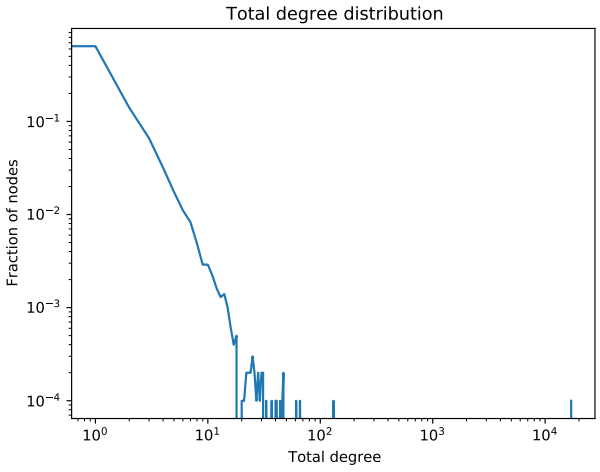
(c) r/politics

Analyzing @foxnews results

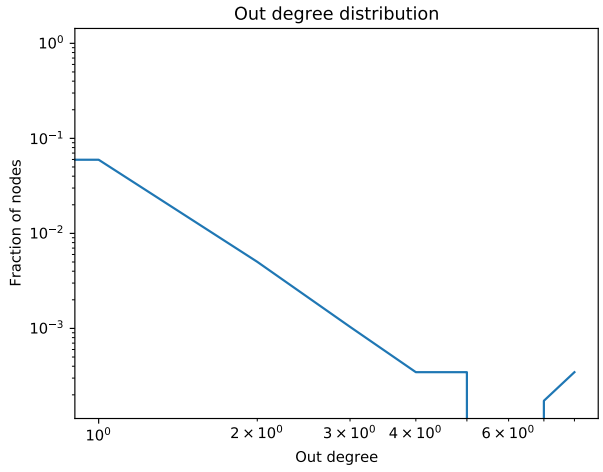
A graph from @foxnews, with 300 contents
α chosen as the median of the η of the contents, $\alpha = 0.58$.
The graph contains 21004 nodes and 44441 edges, $\xi(G) = 17473$ on 10017 vertices and define 598 components. In the original graph they were part of the same component.

- Average shortest path length: 1.031
- Median shortest path length: 1.0
- Average degree: 1.7
- Contributing threads: 155
- Number of threads: 320

Analyzing @foxnews results



(a) total degree distribution



(b) out degree distribution

Analyzing @foxnews results

Threads that contribute the most to the score:

Figure: 975

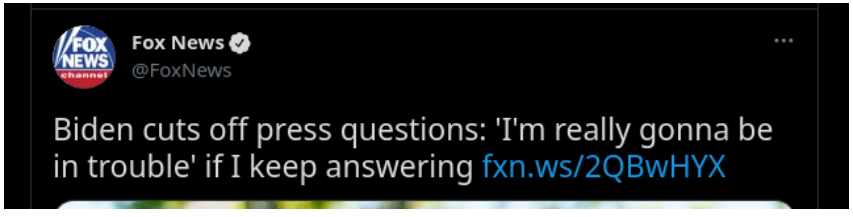


Figure: 870

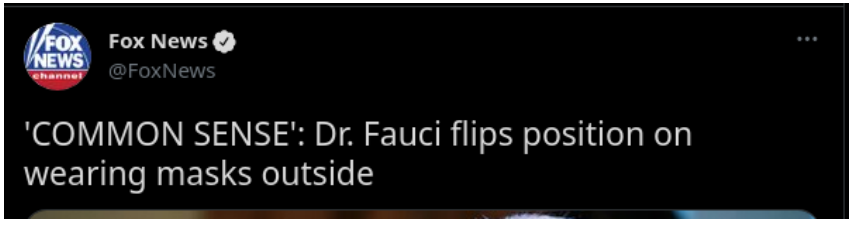


Figure: 803