Thesis notes

13th July

The Echo Chamber Problem

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

$$\xi(U) = \sum_{C \in \hat{\mathcal{C}}} \sum_{T[U] \in S_C(U)} (|T^+[U]| - |T^-[U]|) \tag{1}$$

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 $\xi(G)$

Purity scores

New score for evaluating our method

$$\text{Purity}(U) = \frac{\#\text{ nodes with majority label}}{|U|} \tag{2}$$

Evaluation algorithm

Evaluation algorithm for a graph $G = (V, E)$ with $\mathcal{I}$ communities.

Algorithm 1: Clustering process

foreach $i \in \mathcal{I}$ **do**

$U \leftarrow$ solve ECP on G ;

// Remove edges contributing to $\xi(U)$;

$E \leftarrow E \setminus \{e_{ij} \in E_k, T_k \in \mathcal{S}_C(U), C \in \hat{\mathcal{C}}\}$;

$l \leftarrow$ majority label of users U in $\mathcal{L}$;

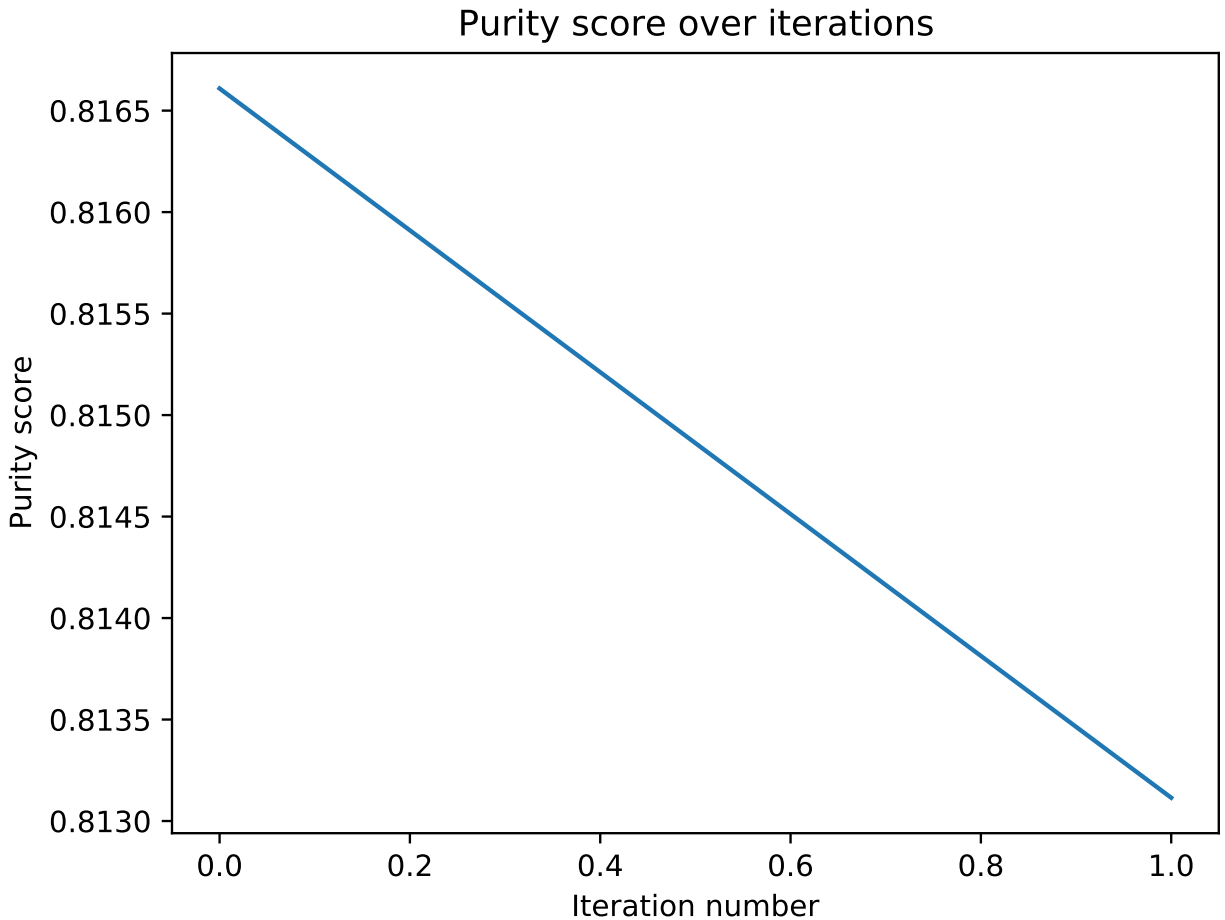
$U_m \leftarrow$ largest component in $G[U]$;

$P_i = \text{Purity}(U_m)$

end

Scores of @nytimes

Dataset has 80% with one label and the remaining 20% having another label



Possible new paths

- Try to reduce noise in the data
  - Do not consider "neutral" comments
  - Limit comments to certain depth
  - More data