OK Boomer: Probing the socio-demographic Divide in Echo Chambers

Henri-Jacques Geiß[+], Flora Sakketou[*], Lucie Flek[*]

[+] Department of Computer Science, Technical University of Darmstadt

[*] Conversational AI and Social Analytics (CAISA) Lab, University of Marburg,

Philipps Universität Marburg

TECHNISCHE UNIVERSITÄT DARMSTADT

CAISA Lab
https://caisa-lab.github.io/

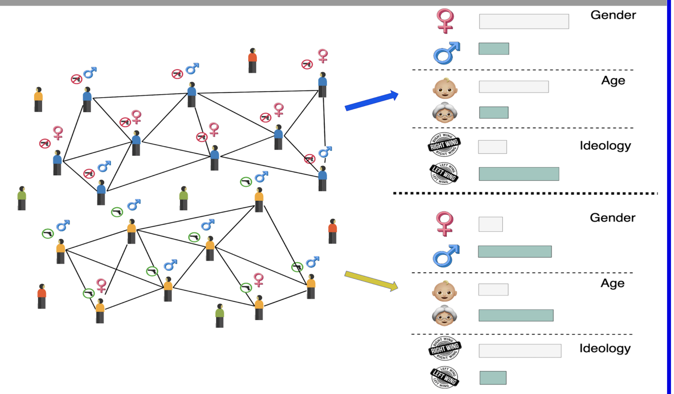
Motivation

- "Birds of a feather flock together"
- Social networks can develop echo chambers with posts reinforcing users' stance
- What is the relation between how often communities interact, what the users' opinions are, and how homogeneous their socio-demographic traits are?

Research Framework

We combine:

- Network-based metrics for community detection
- Annotated user stances
- Classifier-based estimation of gender, age and ideology



Data

- 8 contemporary discussion topics on Reddit
- The 640 manually annotated users of the SPINOS dataset were used as stance samples

Topic	Reddit #Users	Reddit #Posts
Abortion	3,747	631,177
Brexit	2,857	423,294
Capitalism	2,757	418,476
Climate change	1,117	269,032
Feminism	3,613	510,768
Gun control	5,192	667,477
Veganism	1,467	277,786
Nuclear-Energy	535	157,082
Total	**20,571**	**2,716,998**

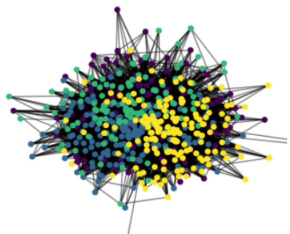
Echo Chamber Detection

- Social Interaction Graph
- Louvain CD Algorithm for community partitions (modularity to determine # of communities)
- Echo Chamber quantified by separability, expansion and density

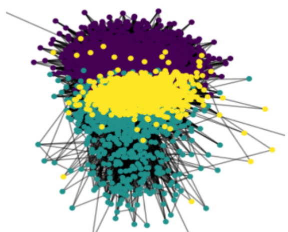
$$\text{separability} = s(c) = \frac{i_c}{o_c}$$

$$\text{expansion} = e(c) = \frac{o_c}{n_c}$$

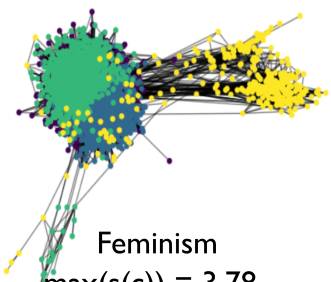
$$\text{density} = d(c) = \frac{i_c}{n_c(n_c - 1) \times 0.5}$$



Brexit
max(s(c)) = 0.446



Abortion
max(s(c)) = 2.02



Feminism
max(s(c)) = 3.78

Socio-demographic Prediction

- Interpretable **gender**, **age** and **ideology** classifiers trained on data from Twitter
- Quality checks by regex pattern

	LinSVM	LogReg	RForest	Base	
Gender (Class-Balanced Down)					
tf-idf	0.735	0.693	0.769	0.5	male/ female
word2vec	0.696	0.659	0.728	0.5	
unigrams	**0.786**	0.7652	0.756	0.5	
Age (Class-Balanced Down)					
tf-idf	0.549	0.51	0.542	0.33	< 30
word2vec	0.516	0.492	0.546	0.33	< 45
unigrams	**0.577**	0.56	0.564	0.33	> 45
Ideology (Class-Balanced Up)					
tf-idf	**0.587**	0.574	0.506	0.33	liberal/
word2vec	0.563	0.557	0.524	0.33	neutral/
unigrams	0.585	0.6	0.516	0.33	conservative

Results

	Stance	Stance σ	Gender	Age	Ideology
Separability	0.483	0.317	0.630	0.110	0.498
Expansion	-0.549	-0.090	-0.403	-0.170	-0.585

Pearson's correlation of the s(c) and e(c) values of the detected communities to (a) their mean stance, (b) their stance standard deviation and (c) their deviations from the full in-topic socio-demographics

- More 'echo-chambered' communities show an increased separation in their stance and gender and ideology profiles
- These results reinforce the call for incorporating socio-demographic and network information into data sets and models for tasks like sentiment analysis and stance prediction.