

Mapping Echo Chambers in large Networks

Giorgio De Magistris, Sergio Pedri, Pietro Spadaccino

[GitHub](#)

Echo Chambers in Social Networks

“An echo chamber is situation a in which beliefs are amplified or reinforced by communication and repetition inside a closed system”

- In social networks, echo chambers are communities of users, characterized by many intra-community and few inter-community contacts
- This positive-feedback between users in an echo chamber can bring social polarization and extremism

Mapping the echo-chamber: detecting and characterizing partisan networks on Twitter

Reference Paper

- Use fake news dataset and link analysis to assign a “score” to each domain cited by tweets
- Analyze the network of Twitter users to map echo chambers spreading disinformation
- Use word2Vec to find the topics discussed in these misleading echo chambers

Our work

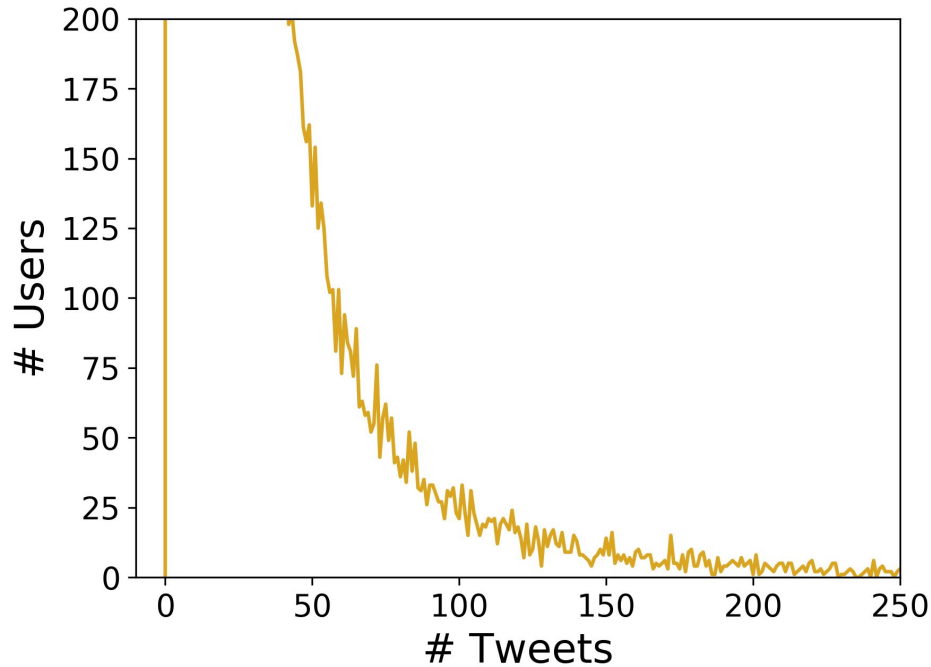
- Collect tweets about climate change
- Find polarization (ie. groups of users sharing the same opinion) inside the network about that topic
- Find which are the terms assuming different meaning in each echo chamber, or being associated to a different set of other terms

Dataset description

- The **dataset** contains tweets unique *ids* related to “climate change” topic [1], gathered from 2017 to 2019, containing ~39M tweets
- Our “hydrated” dataset (from tweet id to whole tweet) contains ~40K *.json* files, each containing ~100 tweets, totalling ~4M tweets of ~1.1M distinct users
- Each tweet is represented as JSON element, containing a set of attributes such as its unique id, the id of the user that created it, mentions, retweet information (if present), URLs embedded in the tweet text and more

[1] <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/5QCCUU>

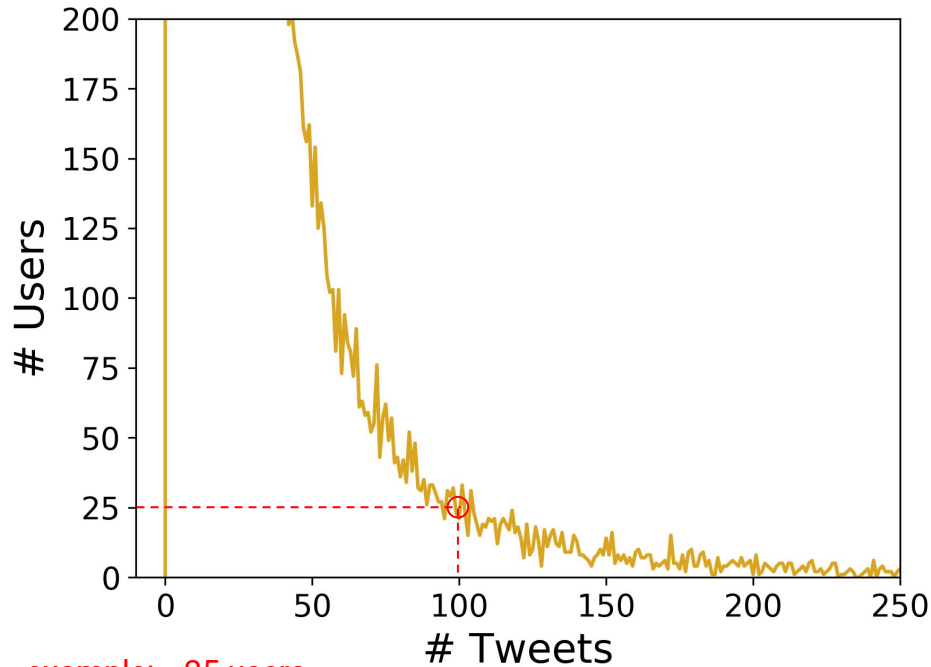
Dataset characterization



Some insights:

- **0.77M** of users tweeted only 1 tweet
- **90%** of users tweeted 4 or less tweets
- **99%** of users tweeted 26 or less tweets

Dataset characterization



example: ~25 users
tweeted ~100 tweets

Some insights:

- **0.77M** of users tweeted only 1 tweet
- **90%** of users tweeted 4 or less tweets
- **99%** of users tweeted 26 or less tweets

Outline

In order to identify the echo chambers and their related deviating words we followed these steps:

1. “Hydrate” the tweet dataset
2. Build the user graph
3. Echo chambers identification
4. Train word embeddings with word2vec on resulting echo chambers

Graph building

Input: *Tweet dataset*

Output: *Undirected weighted graph of twitter users*

- Users = Nodes
- Add (weighted undirected) edges between users if:
 - Direct **retweet** between users
 - Tweet **mentions**
 - Tweet **reply**
 - **Common websites domains** cited in tweet text (not so easy)

Louvain method (our implementation)

Input: *Twitter users graph*

Output: *Communities (echo chambers), using Louvain method [2]*

Initialize by assigning a different community to each node

Phase 1: For each node u , find (if it exists) the neighbour v such that the modularity increase obtained by moving u in the community of v is maximized. Repeat until is no longer possible to increase modularity

Phase 2: Create a new graph having one node per community found in Phase 1 and one edge between two communities $C1$ and $C2$ which weight is the sum of the weights of the edges in the previous graph from nodes in $C1$ to nodes in $C2$, Repeat Phase 1 using this new graph

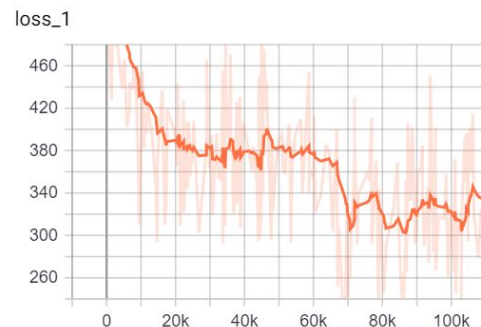
- Where modularity $Q \in [-1, 1]$ measures the density of links inside communities as compared to links between communities

$$Q = \frac{1}{2m} \sum_{i,j} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j)$$

Word-to-Vec

Input: *All tweets, and list of detected echo chambers*

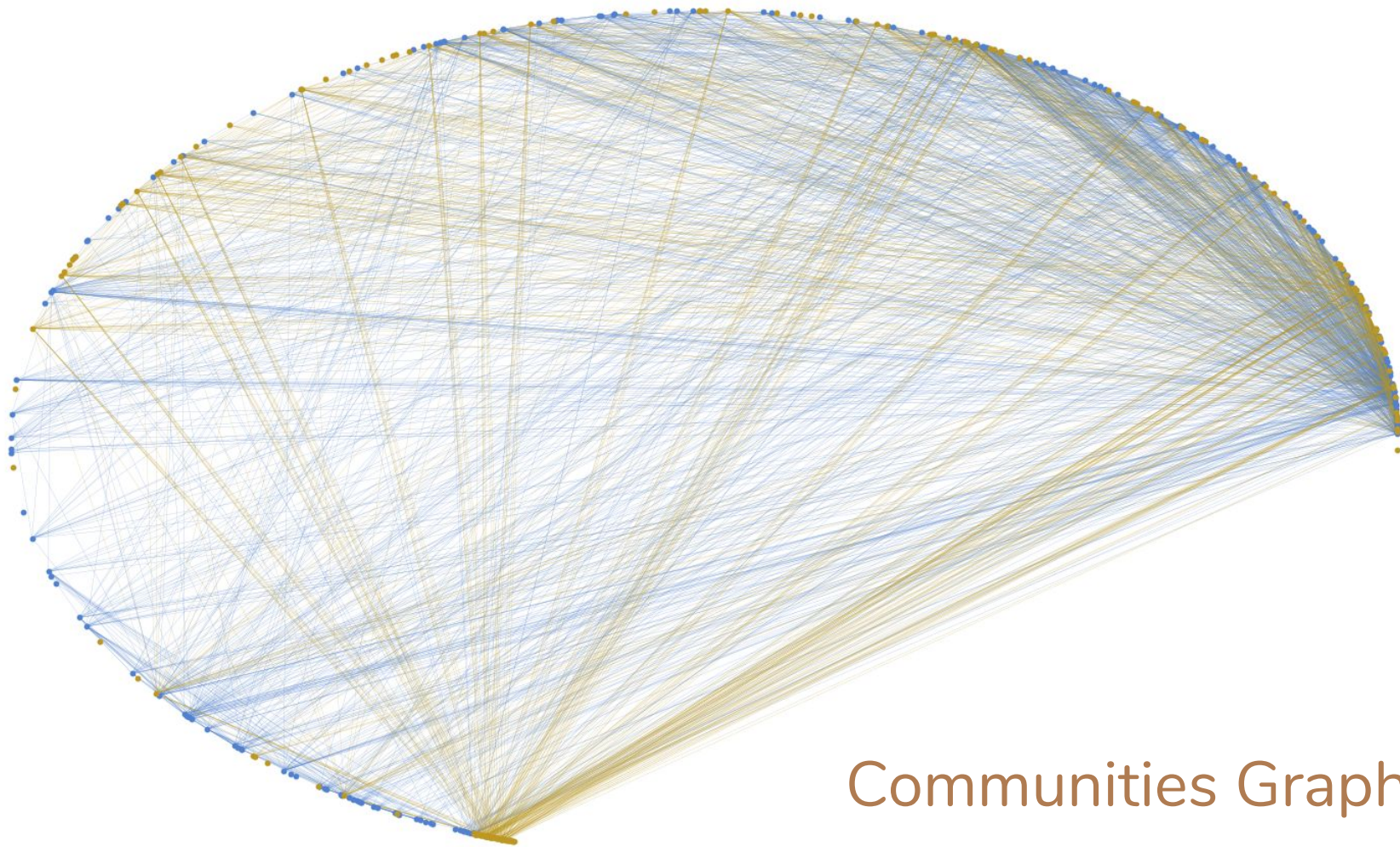
Output: *word embeddings*



Preprocessing by extracting all the individual tokens from the tweets corpus to build a dataset and a dictionary to use in the word-to-vec network. Some manual fine-tuning is done as well

Initial training (in skip-gram mode) on the whole corpus of downloaded tweets

Community training: For each extracted community, the network is further trained on the subset of tweets from those users, starting from the embeddings generated in the first step



Communities Graph

- @realDonaldTrump: This very expensive GLOBAL WARMING bullshit has got to stop. Our planet is freezing, record low temps, and our GW scien...
- Enough with the global warming nonsense: It's freezing out here! -- Earth Changes <https://t.co/gGhBFsAP1s>
- @KeiraSavage00 Proves it's a load of crap the Greens, Labor and Liberal parties are talking about with climate warming
- @SkyNewsAust @DeeDeeDunleavy because George Soro and pals have been brainwashing the public about this global warming conspiracy for decades
- Right wingers hold more store on bullshit and less on science. This may be the cause of the dilemma in question
- @realDonaldTrump: It's pretty funny that the people that don't think global warming is real are the same people that think Noah's Arc is a true story
- From yesterday: We've been getting warnings about human-induced climate change for over a hundred years - <https://t.co/h8GVLvMTcJ>
- Big Oil says government climate change lawsuits are a "conspiracy" | Via QZ <https://t.co/qZb4Utt3hZ>

