Study of various forms polarization on the Reddit network

In this experiment we study various forms of polarization on Reddit.

-First we study ..

-

# Preprocessing

The process of the collection of the data has been carried out strategically. First, we find a controversial topic to begin with. For the sake of the experiment, I selected the topic of the COVID vaccines. However, the code can be easily changed, so it can be applied for any topic. I browsed among the most famous subreddits to collect the data. I concluded on these

* r/Coronavirus
* r/conspiracy
* r/science
* r/DebateVaccines

Those subreddits are not exclusively related to the vaccine, but there are many discussions about the vaccine among those.

So, after choosing the subreddits, I crawled on each one and processed on a specific way. For each subreddit I divide the data in three categories. *Top, Controversial, Both.* For the *Top* case, I put the filter *Top* that the reddit API provides and I collected the top submissions with the keyword ‘*vaccin’.* The same I did with the *Controversial* category.