node (not allowing self-links). We treated our network as indirect one, so the average degree of nodes should be 8.404. Then we made two models with the same number of nodes of our network (24819), but setting the average degree k=8 for having the lower bound and k=9 as upper bound. Doing this we obtain different number of links for each parameter k equal to 8 gave us 99276 links instead k=9 had 124095. As predictable no one of the two limits gave the same number of our real network. For a Regular Lattice, Small World Regime and ER we calculated the major characteristics setting for each the upper and lower bounds.

Table 3: Table following Watts-Strogatz Model

	Lattice	ER	Small World
lower, upper	k = 8 | k = 9	k = 8 | k = 9	k = 8 | k = 9
p	0	1	0.37
<C>	0.6428 | 0.6667	0.0003 | 0.0004	0.1644 | 0.1695
dens	0.0003 | 0.0004	0.0003 | 0.0004	0.0003 | 0.0004
k	8 | 9	8 | 9	8 | 9
<d>	1551.625 | 1241.400	5.173 | 4.686	5.688 | 5.133
distr	Dirac delta func	poisson	poisson
connected	yes	yes	yes

Lower and upper refer to the boundaries, p is the probability to have a link with a distant node, <C> represent the average clustering coefficient, dens is the density of the network, k is the average degree distribution, distr represent which degree distribution follow the model, connected if all the nodes are connected

In one hand, Regular Lattice have an average clustering coefficient significantly higher then our network, in other hand ER have an average clustering coefficient lower then our Small World Regime is near to our network in the average clustering and the average shortest path, characteristic that is also similar to our graph and ER model. Small World Graph is the model more similar to our Subreddit network, but the density is different so we have to use another kind of network to explore this feature.

Comparison with Barbarasi-Albert Model

Scale-free networks are a type of network characterized by the presence of large hubs, that are a nodes highly connected to other nodes in the network. The presence of hubs will give the degree distribution a long tail, indicating the presence of nodes with a much higher degree than most other nodes. To understand if our network is Scale Free we have initially created an artificial Barabasi-Albert graphs, indirect. We create one of BA model putting the same number of nodes of our network and a number of links for each node equal to 8 (the average degree of our network if will be indirect). In figure 4 it's visible the BA model degree distribution, on the left, in the center the cumulative distribution function (CDF) and, on the right, the complementary cumulative distribution function (CCDF) or simply the tail distribution. Figure 5 is the computation of our network, considering it as indirect, using the same methods and graph. Figure 6 is the degree distribution of our real direct network. Then, Figure 7 represent the distribution of in and out degree in our network.

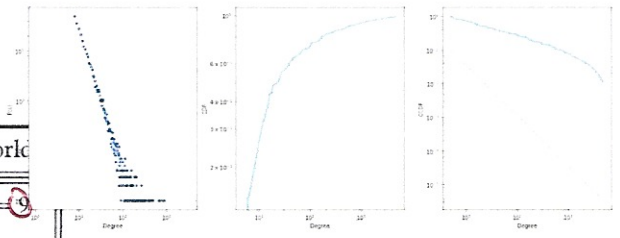

Figure 4: BA model degree distribution
The syntetich Network has alpha=2.505 and sigma=0.049.

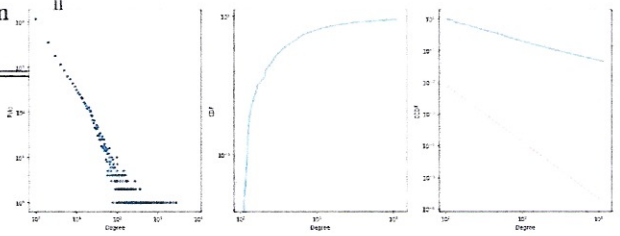

Figure 5: Subreddit network as indirect, degree distribution
Our network (transformed in indirect) has alpha=2.78 and sigma=0.168

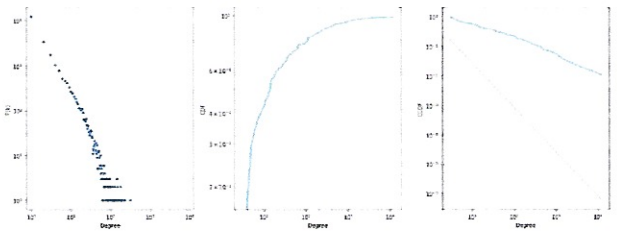

Figure 6: Subreddit network direct
The Directed Network has alpha=2.51 and, sigma=0.054

In conclusion, our network is a scale free network, since the alpha value falls within the range between 2 and 3. That's means in our network there is a presence of hubs as was already pointed out in the first part of the network characterisation.

5 TASK 2: SPREADING DIFFUSION

Here we analyze how information considered as infection, are spreading around our network and compare it with all other synthetic models built before (ER model, BA model, WS model).

We starting our analysis, using the Threshold model introduced by Granovetter[3]. The model works in this way: each node has a threshold; during a generic iteration every node is observed and if the percentage of its infected neighbors is greater than its threshold it becomes infected as well. Using this method we assumed that a meme or a trend as already infected the first percent of our network, for a node is very easy to adopt a 0.2 threshold, although, we iterate the process for 100 times.

"""""(insert img threshold figure ln47)"""""

With this setting the cascade is completed in ER and WS models, instead of our network and in the BA model, the infection stops almost immediately. So, for BA and our network we have tried to start a 10 % of infection supposing that a meme was already very popular. In this case, the simulation shows that the cascade happens for the BA models but fails to fully happen in our network, it becoming stationary when reach something more then the 80 % of nodes, and thus showing the presence of a cluster with density 1 to 0.2 in our network.

Then we try to explore if there were some differences using the SI model introduced by Kermack[4]. During the course of an epidemics, a node changes its status from Susceptible (S) to Infected (I). SI assumes that if, during a generic iteration (we set 1500), a susceptible node comes into contact with an infected one, it becomes infected with probability β (in our case settled as 0.001). Once a node becomes infected, it stays infected and At the beginning of the epidemics we assumed the 1% in the network was already infected.

"""(insert img from ln65)"""

The speed of the spreading varies across the models; from faster to slower, we have BA, ER, WS and our network. So due to the fact that our network is the slower, we want to observe what could append in the SIS model[5]. The model is same as the one above but a node, instead of remain infected, can switch again to susceptible with probability λ. Here we consider the 1 % of nodes already infected, the infection rate equal to the previous one, a recovery rate of 0.005 and $\lambda = \frac{0.01}{0.005} = 2$ (a situation where an outbreak would happen). Iterating all for 600 times.

"""(insert img from ln 84)"""

Here as the one above our network still reach at least 60% of nodes. That's curious, it seems that something in our network doesn't allow to reach the total number of nodes, it could be due to the conformation of our network or lacking of information because we doesn't the complete network of Reddit.

Also we considered a SIS where $\lambda < 1$ so an endemic state is reached ($\mu < \beta\langle k\rangle$;$\langle k\rangle$is the average degree for our network). We maintain all parameters same as the previous simulation, but changing $\lambda = \frac{0.005}{0.01}$ and adding 0.01 < 0.005*8.63.

"""(insert img ln105)"""

What We wanted also observe in SIS, a situation where a "disease-free" condition was reached. The condition to achieve this is having $\mu > \beta\langle k\rangle$. Setting all parameters with 70% of infected nodes at the beginning, the $\beta = 0.001$, $\gamma = 0.05$(the recovery rate) We iterate it 100 times.

(insert img ln124)

By this graphs we could understand that the infection just could fall down because there aren't anymore much subreddits to infect.

It is possible, inside of a subreddit, a post arrived and it is considered interesting by the users of that subreddit, maybe they will re-post it in other, spreading the content, and they will be interested on that posts for a certain number of days. After a while, users will be fed up of that specific topic and introduce other new posts. For observe this, we have used the SIR model[6]. This models follow more or less this concept that a subreddit could be infected, be infected for a certain amount of time and then forget the infection and be healthy again. Here we simulate using this parameters: $\lambda > 1$ over the epidemic threshold, infeted at the beginning 1%, $\gamma = 0.005$, $\beta = 0.01$, iterated 1000 times.

"""(insert img ln150)"""

Finally, a simulation where $\lambda < 1$, so below the epidemic threshold, with a infected rate at the beginning of 1 %, $\gamma = 0.01$, $\beta = 0.005$ and iterating it 1000 times.

"""(insert img ln174)"""

From the comparison between models, it appears that the spread of the "contagion" in our real network is limited in terms of infected nodes compared to all the other models; furthermore it seems, in the most of cases, that, in our network, the epidemic is slower in spread in comparison to other models considered.

[3] M. Granovetter, "Threshold models of collective behavior," The American Journal of Sociology, vol. 83, no. 6, pp. 1420–1443, 1978

[4] W. O. Kermack and A. McKendrick, "A Contribution to the Mathematical Theory of Epidemics," Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character, vol. 115, no. 772, pp. 700–721, Aug. 1927.

[5] W. O. Kermack and A. McKendrick, "A Contribution to the Mathematical Theory of Epidemics," Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character, vol. 115, no. 772, pp. 700–721, Aug. 1927

[6] W. O. Kermack and A. McKendrick, "A Contribution to the Mathematical Theory of Epidemics," Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character, vol. 115, no. 772, pp. 700–721, Aug. 1927

that on average share more crossposts then original ones? As the analysis above we started with the twenty-fifth more active subreddits that share not original content.

"""(inserire istogrammi)"""

Very low and high-medium number of members used to share not original content, most of them are media share typo but community typo are more then news subreddits. Themes of this subreddit are politics, melting pot and meme the majority but we could notice that there are no science subreddits but emerges personal one. Those are private subreddits created and operated by a single redditor, in which all posts are crossposts from other subreddits. The reason why those exist is because there is a limit of 1000 posts a single redditor can save and once this limit is reached a personal subreddit is created in order to continue to save posts.

""" (inserire istogrammi)"""

If we took a large picture, taking into account the top one-hundred reposter subreddits, what seems to be different is the presence in only one subreddit with number of member over 10M and the majority of them are in the range 100k-1M.This may be caused by the fact that those smaller subreddits have less active users and are at the same time more niche, so less original content are created.

As showing until now we could say the major part of original content are created in subreddits with and high number of members and then they finish in small subreddits. The major part of the subreddits' contents are media share, following news and discussion on topic common between members. In our network there isn't a dominant theme in fact melting pot represent a category of subreddits that share variable contents, as was predictable, but is quite common on on our network find posts related on politics and memes. So if there isn't a preferred theme that could have more success then other, how a post could became viral and be cross-posted?

For try to figure how this could happen, we were thinking that probably more up-votes a posts have and higher is the probability of this post to be saw and crossposted. So we choose the first one thousand viral posts (having the highest number of up-votes) and we check in how many other subreddits they were posted.

""" insert img """

Surprisingly the number of subreddits its very low, infact the majority of the posts were crossposted in only 1 other subreddit. We make another test using the top ten thousand viral posts for compare the result but once again the number of crosspost sill very low. We try to watch his distribution but no significant change was found.

"""(manca l'immagine della distribuzione)"""

According to our observations it's not so easy for a post became viral and be cross-posted in hundred and hundred subreddits, probably it's not enough have a lot of up-votes, in part that could due because Reddit have a particular structure and his users participate and follow only what they are interested on. Other motivation could be because a post need a sum of circumstances for having a chance to became viral. Furthermore it also possible that in Reddit there aren't BOTs that spam contents in different subreddits.

Old posts revival

During our data collection we noticed some hot posts dated before 2021. That capture our curiosity because we take for granted that the hotted posts will be only the recent one due to the constant bombing of contents that is typical in social media network communities. For this reason we decided to analyze better this kind of posts. What we have done, was picked posts and filtrated in order to select only the ones scraped from the "hot" list of a subreddit. Opposed to all the other cross-posted posts collected after a "hot" crosspost was found. We ordered them by the data of posting and took all posts published before the 2021 (using R) After that, with a python script, the members of the subreddits where those posts had been submitted to were collected with a script [7]. Therefore, subspre2021 contains the number of members of a subreddit with "hot" posts dated older than 2021, 4 months prior to the network scraping. Using that information we plot a distribution. It was immediately clear that the most part of subreditts had a quite modest number of members. Eliminated the 132 outliers, this is the distribution:

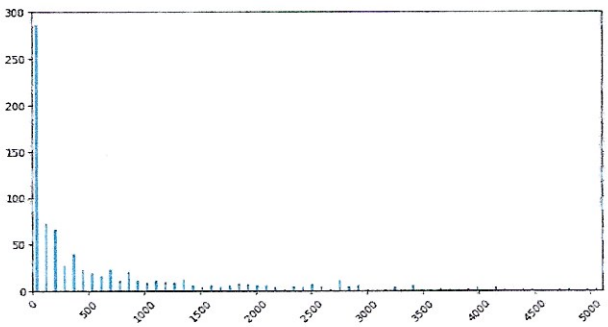

Figure 10: Distribution subreddits/members

Having very old cross-posts among the "hot" ones appears to be a sign of a not very active subreddit. In fact, our code, once find a subreddit, scroll all posts inside it; and, if the subreddit is not very active, so it goes back in time until reach the "hot" posts. Randomly checking the content of

[7] the number of members of the selected subreddits was collected approximately a month and a half after the initial data scraping. Four subreddits from the partial dataset had been banned in the meanwhile. They have been ignored for the data collection.

~~our~~ posts, we realised that some of them were pinned[8] and included in the data set. With a python script, we retrieved all the pinned posts and for each subreddit, counted the number of posts not pinned and plotted a distribution.

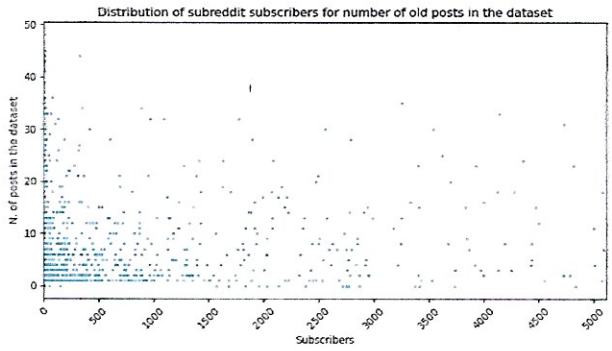

Figure 11: Distribution subreddits posts

As already explain above, the imagine well confirmed that small subreddits have less active users, in fact it is possible to observe an high distribution of old posts in them.

Game Stop Case, spreading in our network

8 DISCUSSION

REFERENCES

[8] A pinned post is a social media post saved to the top of a page or profile on Facebook, Twitter, Reddit and so on. Pinning a post is a great way to feature an important announcement or highlight some of your best content.