Who are the opinion leaders on COVID-19?

Shahriar Yazdipour
Komal Tanaji Abhang
Jessica Alejandra Grau Chopite

Technische Universität Ilmenau
July 2020

Relevance of the study

Traditionally, Opinion Leaders (OLs) are a minority of members within a society, who possess qualities that make them **persuasive in spreading ideas to others**. Their opinions greatly affect others on social networks.

However, sometimes ordinary internet users can produce and effectively broadcast to mass audiences. These users can even **create alternative sources of content that can influence mass media agendas.**

Relevance of the study

OLs can help to spread important messages regarding COVID-19, as they can support the **promotion of social harmony, stability and the latest events**.

Studying OLs' influence patterns could help us understanding why certain information is adopted faster than other, and consequently, could help to **create more effective communications** regarding the virus in the upcoming future.

(Chen et al., 2014; Meeyoung Cha et al., 2010; Xu et al., 2014)

3

Research Questions

RQ: Who are the Opinion Leaders on COVID-19?

Who are the Opinion Leaders on COVID-19?

Step 1: Filtering the data

Step 2: Preprocessing the data

Step 3: Social graph formation

Step 4: Social network analysis
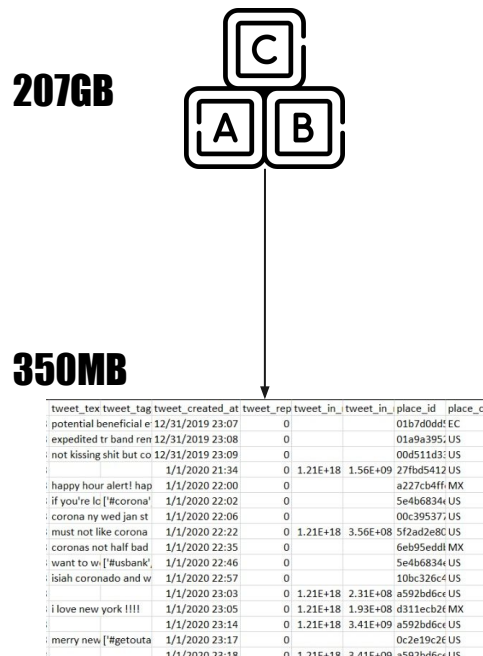
Step 1: Filtering the data

Threading vs Multiprocessing

- 12 Processes for 12 files
- One Thread for each Chunk of Data
- ~9100 Threads used
- Less than a day to filter the data
- Reduce Size from 207GB to ~2.6GB

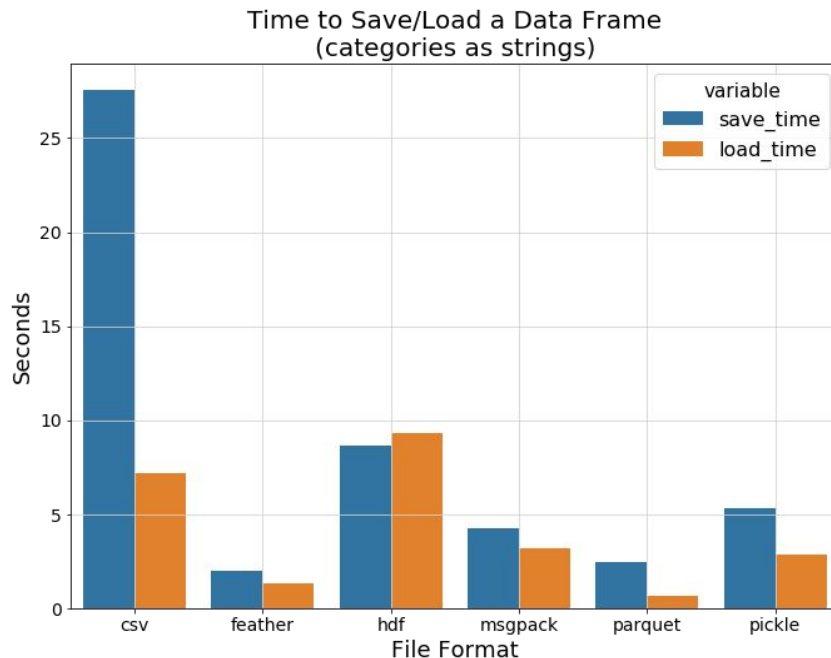| USER	PR	NI	VIRT	RES	SHR	S	%CPU	%MEM	TIME+	COMMAND
shya6478	20	0	1683404	75852	11004	R	79.0	0.3	0:19.42	python3.6
vaka5268	20	0	2895852	1.0g	10520	R	77.7	4.4	645:41.20	python3.6
tiwa8375	20	0	1940476	68628	9116	R	71.0	0.3	2873:52	python3.6
shya6478	20	0	1683916	76092	11004	R	71.0	0.3	0:19.27	python3.6
shya6478	20	0	1683112	74920	11004	R	68.7	0.3	0:19.27	python3.6
shya6478	20	0	1682328	74520	11004	R	66.0	0.3	0:19.09	python3.6
shya6478	20	0	1682940	74788	11004	R	63.0	0.3	0:19.48	python3.6
kash1240	20	0	6164140	4.1g	10304	R	61.0	17.6	276:29.63	python3.6
shya6478	20	0	1681204	73504	11004	R	53.0	0.3	0:17.54	python3.6
shya6478	20	0	1680492	73484	11004	R	53.0	0.3	0:19.75	python3.6
shya6478	20	0	1680268	72372	11036	R	51.3	0.3	0:14.72	python3.6
shya6478	20	0	1678792	71688	11004	R	35.7	0.3	0:10.90	python3.6
shya6478	20	0	1678812	71808	11004	R	31.3	0.3	0:10.93	python3.6
shya6478	20	0	610376	75000	5256	S	12.7	0.3	7:08.61	jupyterhub-s
shya6478	20	0	1742240	83548	24344	S	5.3	0.3	0:04.05	python3.6
root	20	0	1065048	36996	5968	S	1.0	0.2	42:41.65	node

Step 2: Preprocessing

207GB

- Collect COVID-19 Tweets based on Tweet Content using Regex filters
- Drop Extra Properties
- Reduce Size from ~2.6GB to ~350MB

350MB

tweet_tex	tweet_tag	tweet_created_at	tweet_rep	tweet_in_	tweet_in_	place_id	place_c
potential beneficial e		12/31/2019 23:07	0			01b7d0dd!	EC
expedited tr	band rem	12/31/2019 23:08	0			01a9a395;	US
not kissing shit but co		12/31/2019 23:09	0			00d511d3!	US
		1/1/2020 21:34	0	1.21E+18	1.56E+09	27fbd5412	US
happy hour alert! hap		1/1/2020 22:00	0			a227cb4ffi	MX
if you're lc	['#corona'	1/1/2020 22:02	0			5e4b6834(	US
corona ny wed jan st		1/1/2020 22:06	0			00c395373	US
must not like corona		1/1/2020 22:22	0	1.21E+18	3.56E+08	5f2ad2e80	US
coronas not half bad		1/1/2020 22:35	0			6eb95eddl	MX
want to w	['#usbank'	1/1/2020 22:46	0			5e4b6834(	US
isiah coronado and w		1/1/2020 22:57	0			10bc326c4	US
		1/1/2020 23:03	0	1.21E+18	2.31E+08	a592bd6c(	US
i love new york !!!!		1/1/2020 23:05	0	1.21E+18	1.93E+08	d311ecb2(	MX
		1/1/2020 23:14	0	1.21E+18	3.41E+09	a592bd6c(	US
merry new	['#getouta	1/1/2020 23:17	0			0c2e19c2(	US
		1/1/2020 23:18	0	1.21E+18	3.41E+09	a592bd6c(	US

Step 2: Preprocessing

- Filter Big JSON files to multiple Feather
  - Faster to read and write
- Preprocess on feather files save to CSV
  - Easier for everyone to read
  - Built in support in all graph analysis
    programs, like Gephi

https://towardsdatascience.com/the-best-format-to-save-pandas-data-414dca023e0d

Time to Save/Load a Data Frame
(categories as strings)

8

Why a Social Network Analysis?

The positions of users in online social networks, such as Twitter,indicate their connectivity

Users' social connectivity can be measured by different types of centrality measures like degree , betweenness and closeness centrality.

In a social network analysis, a central network position indicates **high centrality**, and a central and strategic network position implies that there is a **control of information flow and an influence of others attitudes and behaviors.**

(Xu et al., 2014)

Step 3: Graph Formation

Relation 1: Tweet Reply Relation

- Source Node : Tweet ID
- Destination Node : Reply_to_status_ID

Relation 2: User Reply Relation

- Source Node : User ID
- Destination Node : Reply_to_user_ID

Step 4: Social Network Analysis

Centrality Measures:

1. Degree Centrality
   - Number of connections
2. Betweenness Centrality
   - number of shortest paths from all nodes to all others that pass through a particular node
3. Closeness Centrality
   - average length of the shortest paths between a specific node and all other nodes in the graph

Results of Graph Formation:

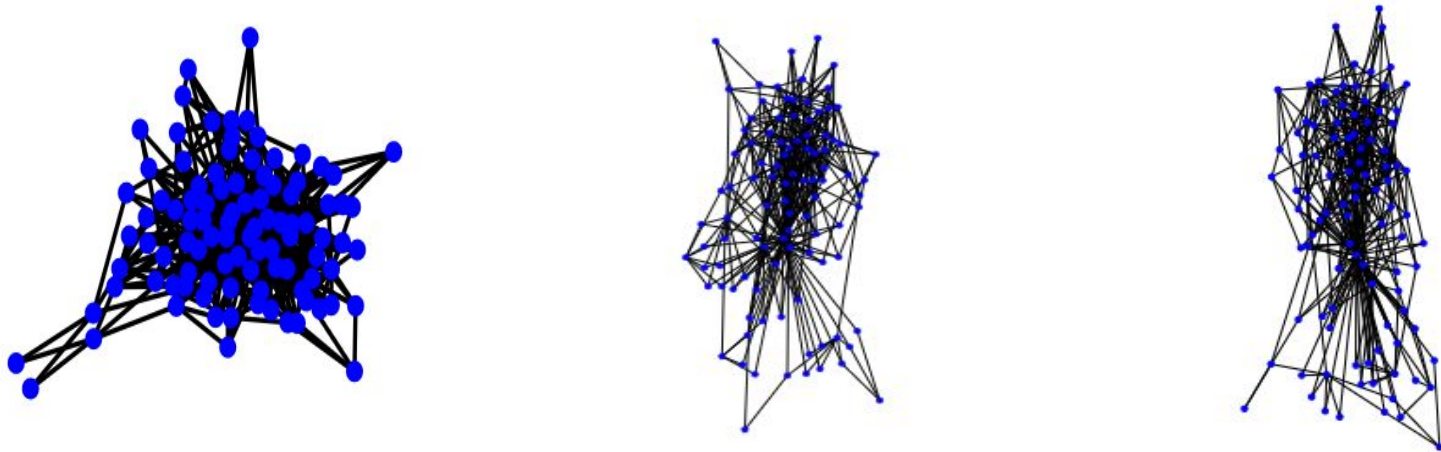Graph for User Reply Relation: Higher connectivity is equivalent to more influence in network

Fig1: Sub-graph for highest 100 users of degree, betweenness and closeness centrality

Results of Social Network Analysis:

Relation	Centrality Measures	Highest	Most Influential User
User Relation	Degree	0.39	Original User_id = 25073877
	Betweenness	0.66	
	Closeness	0.44	

What do we expect?

The main expectations of this research are

to understand predictors of opinion leadership on Twitter regarding COVID-19;

to give an insight on why certain content regarding COVID-19 spreads easier than other

Issues

1. Reply_count is always 0, so there is no access to this property.

2. Limitations with the data access in general.

Future Approach:

1. Planning to consider more influence metrics

2. Categorizing user profiles, to explore whether the most influential tweets are from organizations or if they are ordinary internet users' content

3. Exploring the sentiment of Opinion Leaders' tweets

Libraries we used

- Numpy / Pandas
- Tweet-preprocessor
  - URLs
  - Hashtags
  - Mentions
  - Emojis
- Dask
  - It provides advanced parallelism for analytics, enabling performance at scale
  - Operations that were slow on Pandas, like iterating through row-by-row, remain slow on Dask
- NetworkX
  - Python library for studying graphs and networks.

Thank you for your attention

Contact:
jessica-alejandra.grau-chopite@tu-ilmenau.de
komal-tanaji.abhang@tu-ilmenau.de
shahriar.yazdipour@tu-ilmenau.de

References

Chen, Y., Wang, X., Tang, B., Xu, R., Yuan, B., Xiang, X., & Bu, J. (2014). Identifying Opinion Leaders from Online Comments. In H. Huang, T. Liu, H.-P. Zhang, & J. Tang (Eds.), *Communications in Computer and Information Science: Vol. 489. Social media processing: National Conference, SMP 2014, Beijing, China, November 1-2. Proceedings* (1st ed., Vol. 489, pp. 231–239). Springer. https://doi.org/10.1007/978-3-662-45558-6_21

Meeyoung Cha, Hamed Haddadi, Fabrício Benevenuto, & Gummadi, K. P. (2010). Measuring user influence in twitter: The million follower fallacy. In *ICWSM 2010 - Proceedings of the 4th International AAAI Conference on Weblogs and Social Media*.

Xu, W. W., Sang, Y., Blasiola, S., & Park, H. W. (2014). Predicting Opinion Leaders in Twitter Activism Networks. *American Behavioral Scientist*, *58*(10), 1278–1293. https://doi.org/10.1177/0002764214527091