Misinformation Flow in Social Networks

Group -14

MEMBERS

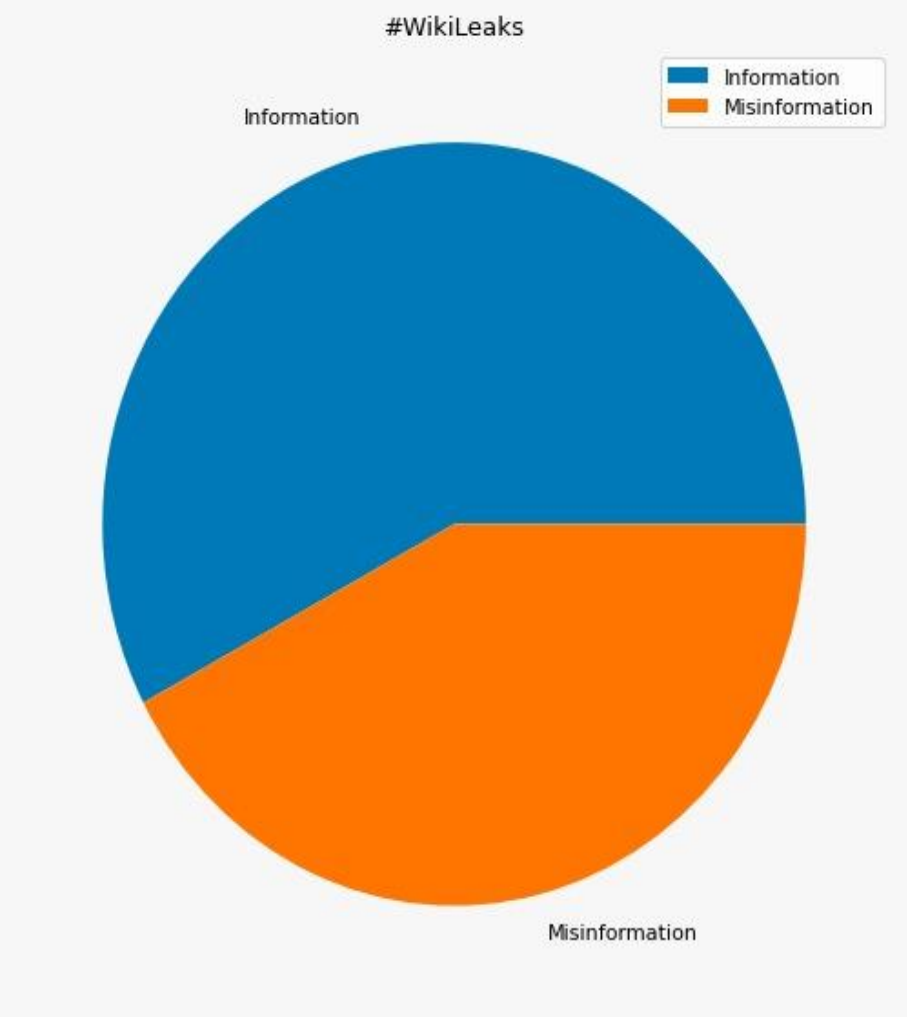
- ADITYA SHARMA
- ANNAPOORANI. A
- PRATIK HAWARE
- SHLOK BANSAL
- KSHITIJ GUPTA
- KAMAL PHOOLWANI
- SARTHAK VERMA
- ALOK JAIN
- ASHISH RAI
- RAHUL KATYAL

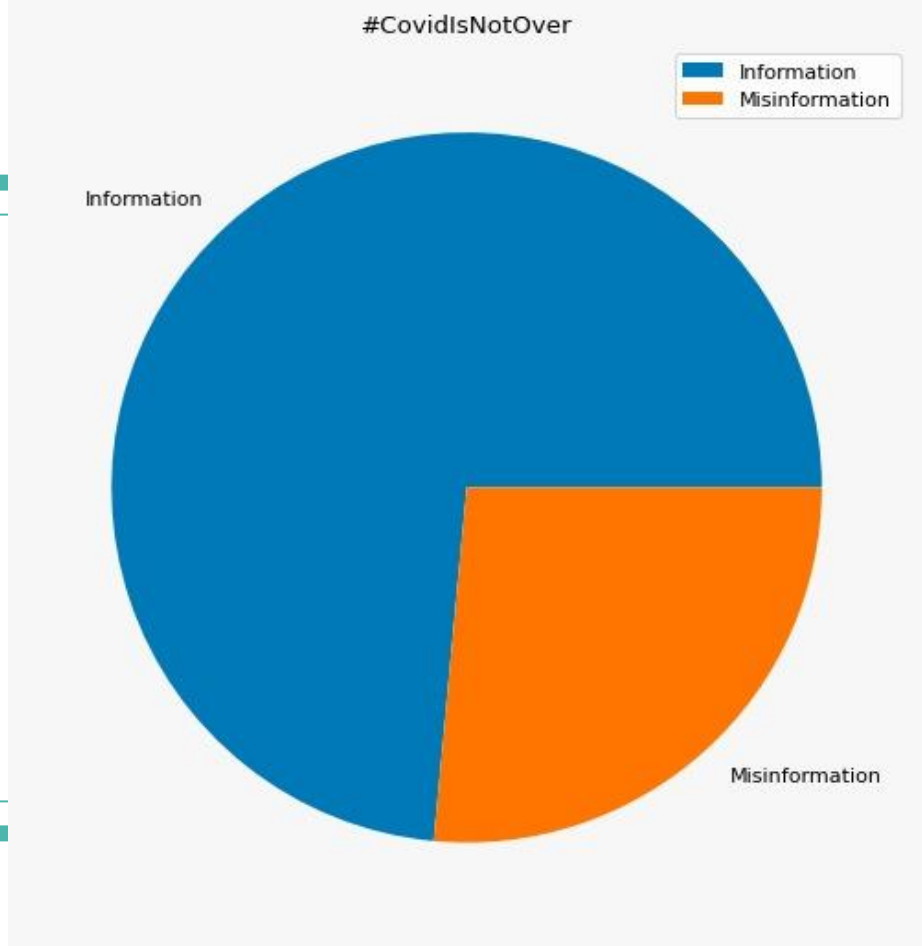
	A	B	C	D
1	Topics	Hashtags	Frequency of Misinformation	
		#COVID19	250	
		#COVID	215	
		#Chinavirus	300	
		#Wuhanvirus	350	
2	COVID-19	#covidisnotover	410	
		#ukrainerussianwar	450	
3	War	#ukrainewar	400	
4	Wikileaks	#wikileaks	100	
		#football	50	
		#cricket	75	
5	Sports	#tennis	50	
6				

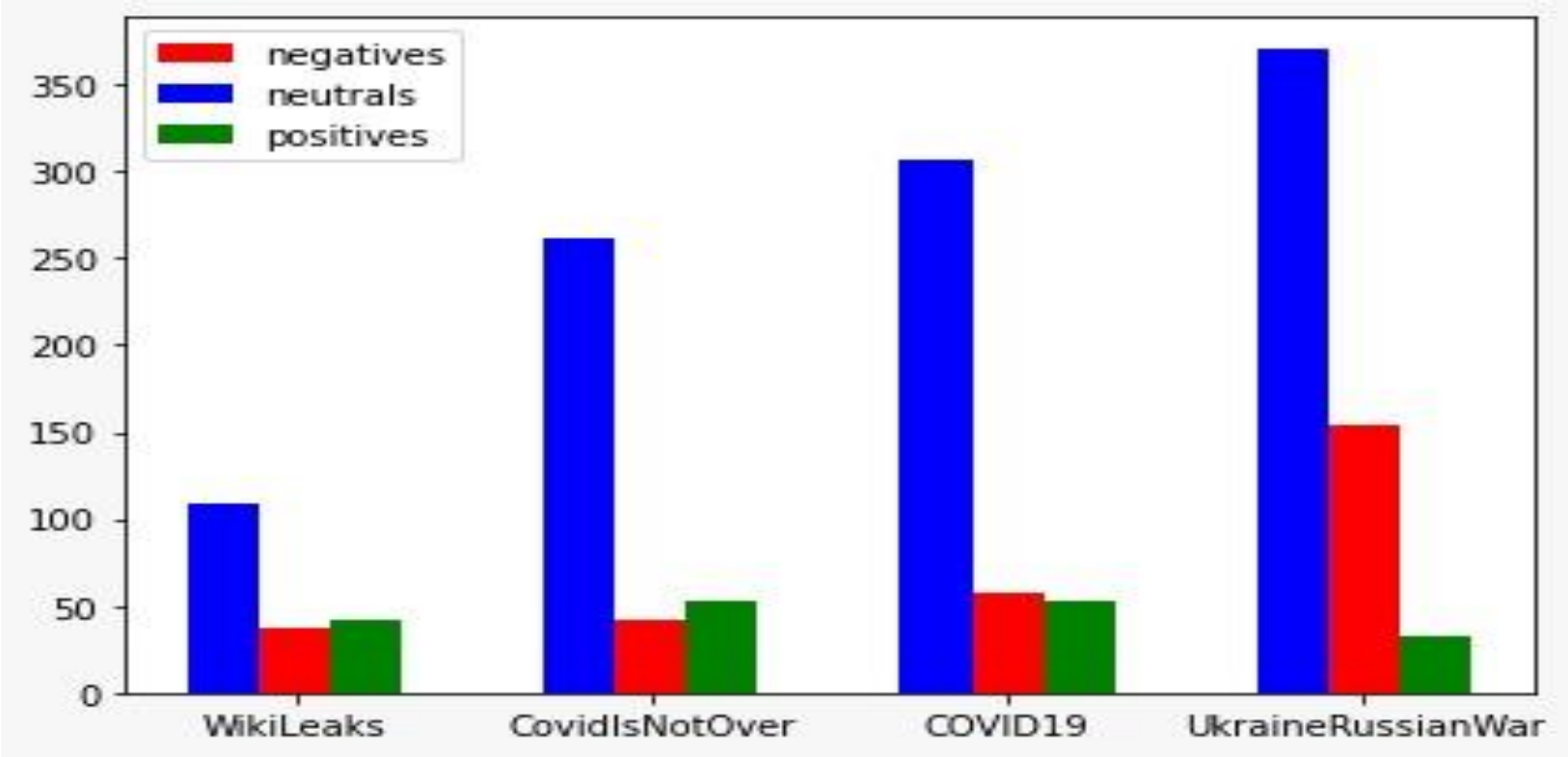
	B	C	D	E	F	G	
	Tweet content	User	Urls list	Hashtags list	@ list	ID	created_a
0	Felt a bit off the last day or so, did test and Well, bollocks. Still, few days o	backudog123		CovidIsNotOver COVID COVID19		1.50596E+18	
1	Yeah you are.	BetterWthRachel	https://twitter.com/BCisvulnerable/sta	CovidIsNotOver COVID19		1.50596E+18	
2	1 out of 3 of my customers have been masked coming in today. Thank	OldManGrumpee		Ontario COVID CovidIsNotOver		1.50596E+18	
6	80% of my boys classes taught by supply today!! And 1/3 of his classmates	ElizStanley_	https://twitter.com/helen_chap/status	CovidIsNotOver		1.50596E+18	
8	You are going to have to pee higher on them & their families	BritWeatherSvs		Science CovidIsNotOver	BorisJohnson metpoliceuk	1.50596E+18	
9	@TheLastPersont2 This government is acting completely irresponsible this	markeverett1977		CovidIsNotOver	TheLastPersont2	1.50596E+18	
14	.@RishiSunak .@10DowningStreet	maraudersmap	https://twitter.com/BeckettUnite/stat	ToriesPartiedWhilePeopleDied C	RishiSunak 10DowningStree	1.50596E+18	
16	@h2ofizz @Tyhj34777949 @curiocat13 @fascinatorfun You can't argue w	patanna		COVIDIOTS CovidIsNotOver	h2ofizz Tyhj34777949 curio	1.50596E+18	
17	Over 1% of the Irish population caught COVID in the last 5 days. Hardly kn	akellyucc	https://twitter.com/fergalbowers/stat	CovidIsNotOver		1.50596E+18	
18	Scotland's record for Covid hospitalisation broken, and by some margin. W	ZeroCovidScot	https://twitter.com/TravellingTabby/st	CovidIsNotOver	HumzaYousaf	1.50596E+18	
24	@NadineDorries He does shame us! #CovidIsNotOver #ToryCovidDisaster	BobbyWotnotArt		CovidIsNotOver ToryCovidDisas	NadineDorries	1.50596E+18	
28	@AmitAryaMD There is no need to drop masks. Covid is not over! #WearA	fcktrwing		WearAMask COVIDisAirborne Cc	AmitAryaMD	1.50596E+18	
29	@blogTO Douglas Ford is playing politics instead of listening to the science	fcktrwing		WearAMask COVIDisAirborne Cc	blogTO	1.50596E+18	
38	@BrianHTweed ðŸ¤¦â€â™‚ï¸ Get well soon! ðŸ¤žðŸ¤ž	Justice30		CovidIsNotOver	BrianHTweed	1.50596E+18	
39	@NaheedD âœ‹ Covid is not over just because Douglas Ford says so. #We	fcktrwing		WearAMask COVIDisAirborne Cc	NaheedD	1.50596E+18	
40	Just home after a doctor's appt and trip to Shoppers to pick up meds and	oconnor_gmail		onpoli maskmandate WearAMask	CovidIsNotOver covidonta	1.50596E+18	
45	@ReliableCA @PeterHire69 @AratxuS I'm not saying there is. But I'm	KatyTreverton		CovidIsNotOver	ReliableCA PeterHire69 Ara	1.50596E+18	
47	@StevenDelDuca @OntLiberal Well said. Ty for pointing it out. #VoteFord	fcktrwing		VoteFordOut2022 NeverVoteCo	StevenDelDuca OntLiberal	1.50596E+18	
48	@Karenlovecheese #CoronavirusUK âš ï¸ #Covid19UK	Justice30		CoronavirusUK Covid19UK Covic	Karenlovecheese	1.50596E+18	
49	A Grounding Practice that Soothes my Soul when Depression Kicks In. http	Debbie_Aruta	https://elejrnl.com/?p=3311746	CovidIsNotOver MentalHealthA	elephantjournal	1.50596E+18	
55	I'm not and will never be vaccinated but I will continue to #WearAMask	jjab__84		WearAMask CovidIsNotOver BA2Variant		1.50596E+18	
64	Can @UKLabour @Keir_Starmer push for answer pls	TheJusticeGame2	https://twitter.com/JujuliaGrace/statu	CovidIsNotOver Covid19UK	UKLabour Keir_Starmer	1.50596E+18	
69	#Kharkiv. #Russian missile hit civilians in the supermarket	AbeedahAkhtar		Kharkiv Russian StandWithUkraine	StopWar StopRussia Azad	1.50596E+18	
71	Another #Covid_19 variant is coming #HereWeGo #CovidIsNotOver	rasberuk		Covid_19 HereWeGo CovidIsNotOver		1.50596E+18	
74	A couple of days ago @PremierScottMoe & @MerrimanPaul said	actagainstcovid	https://bit.ly/3ufsJFb	covid19sk covidisnotover	PremierScottMoe Merrimar	1.50595E+18	
75	This! Overheard multiple people say they canâ€™t wait until a store	BrookePercy1	https://twitter.com/chantalsback/stat	CovidIsNotOver WearAMask		1.50595E+18	
76	Yup, my mask is staying on.	Nicwitchy	https://twitter.com/click4amanda/stat	CovidIsNotOver WearAMask		1.50595E+18	
79	This is exactly why I still mask, wash often and socially distancing!! #Covid	LizMarinTRU	https://twitter.com/cha_myoung/statu	CovidIsNotOver BA2Variant		1.50595E+18	
80	Just finished shopping at Walmart - didn't see a single unmasked face	kylejhutton		CovidIsNotOver onpoli cdnpoli	fordnation	1.50595E+18	
81	Dream after listening to "Oedipus King" while having covid:	scotjess3		CovidIsNotOver GreekMyths ClassicsTwitter		1.50595E+18	

CovidIsNotOver_new

Accessibility: Unavailable


#WikiLeaks
Information
Misinformation
Information
Misinformation


#CovidIsNotOver
Information
Misinformation
Information
Misinformation


negatives
neutrals
positives
350
300
250
200
150
100
50
0
WikiLeaks
CovidIsNotOver
COVID19
UkraineRussianWar

PIPELINE

Collection

Collecting tweets related to a certain topic via scraping or API

Filtering

Tweets that can be seen as valid misinformation tweets for our purposes are filtered out

Network Analysis

Analyse diffusion of information in network structure

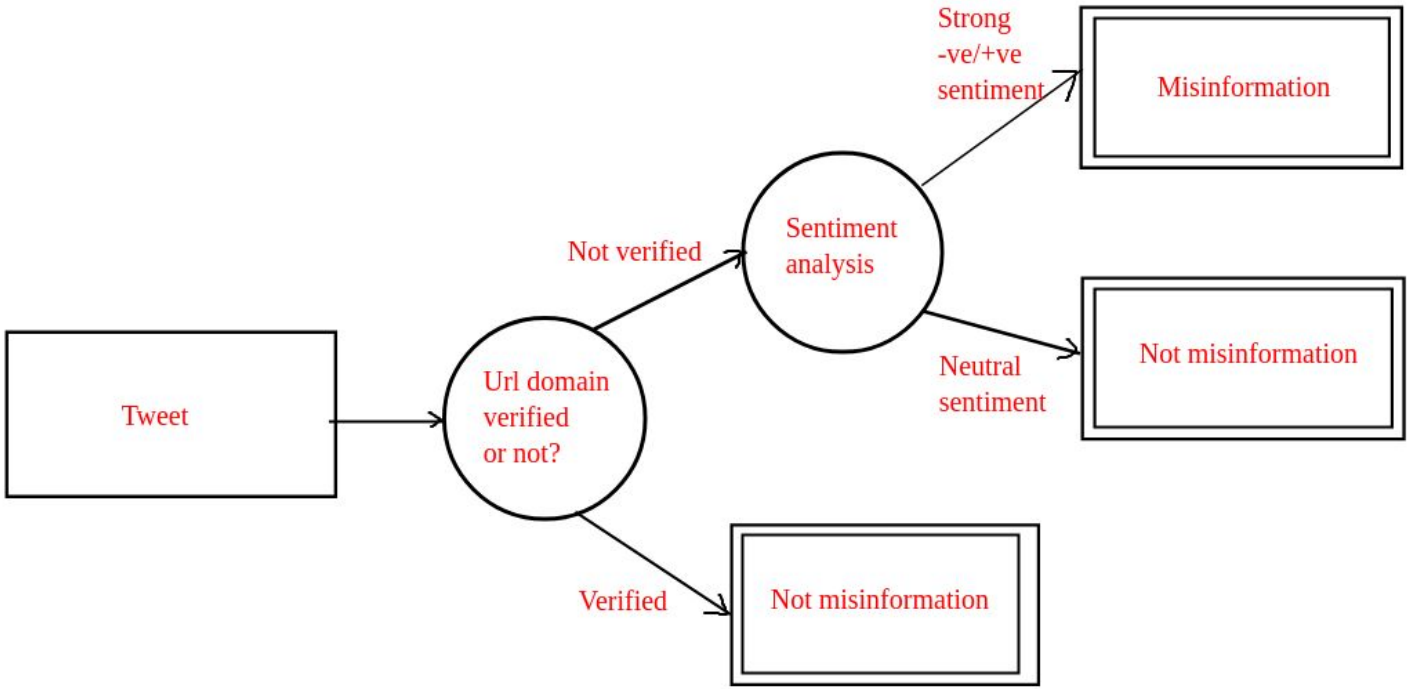
Content Analysis

Use sentiment analysis, topic analysis and clustering . Then use GLM for to examine the differences in diffusion characteristics

Labeling Tweets As Misinformation

- A list of trusted handles is maintained for initial misinformation validation.
- If the tweet has cited to any of the sites or handle available in the trusted list then the information is deemed to be a valid information.
- If not, sentiment analysis is done. Neutral sentiment indicates valid information, strong sentiments indicate misinformation possibility

Labeling Tweets As Misinformation (Pt.2)



T - Test

- The T-Test or the Student's test is used to reject a null-hypothesis or not reject it.
- We will be using the independent T-Test to determine the diffusion characteristics and draw a comparison, showing that there is a difference in the tweets, likes or comments between the misinformation and the true information.
- *T-value = (difference between group means)/(variability between the groups)*
- The factors like number of tweets, p-value, degree of freedom, mean, variance, skewness and kurtosis can help us understand the spread of diffusion of misinformation and analyse the characteristics of such networks, and the properties of true and misinformation.

	Attribute	Retweet true info	Retweet misinfo	Favourite true info	Favourite misinfo	Comments true info	Comments misinfo
0	Mean	61832.200000	10178.400000	11364.000000	1918.800000	210267.800000	37539.600000
1	Max	206891.000000	27585.000000	11364.000000	1918.800000	210267.800000	37539.600000
2	Min	4886.000000	1647.000000	1047.000000	89.000000	16810.000000	4171.000000
3	Std	75435.728679	9047.437617	10920.036465	1472.532838	228218.530014	36378.919465
4	Skew	1.843974	1.812885	1.370968	0.374266	1.607976	1.932215
5	Kurtosis	-0.130264	-0.079894	-0.511332	-1.449218	-0.309785	0.028726

Formula: t-val =difference between the means of group/ variability of the groups

Null hypothesis: There is no significant statistical difference between the retweets, favourites or comments between true and misinformation.

Observations: The t-test gives t-values 1.52, 1.92 and 1.67 for retweets, favourites and comments. The number of degrees of freedom we obtained is 8 and on taking a significance level of 5% as a one-tailed test (p=0.05). We find our critical value is 1.86. Thus, our t-value is greater than our critical value in the number of favourites, so we can reject our null hypothesis for favourites and accept it for the other two cases.

Conclusion: The diffusion of misinformation in favourites is statistically much significant from the diffusion of true information in favourites.

Analysis: The Independent Sampled T-test was performed. The increased **kurtosis** indicate that there is increase in the tail data greater than normal distribution.

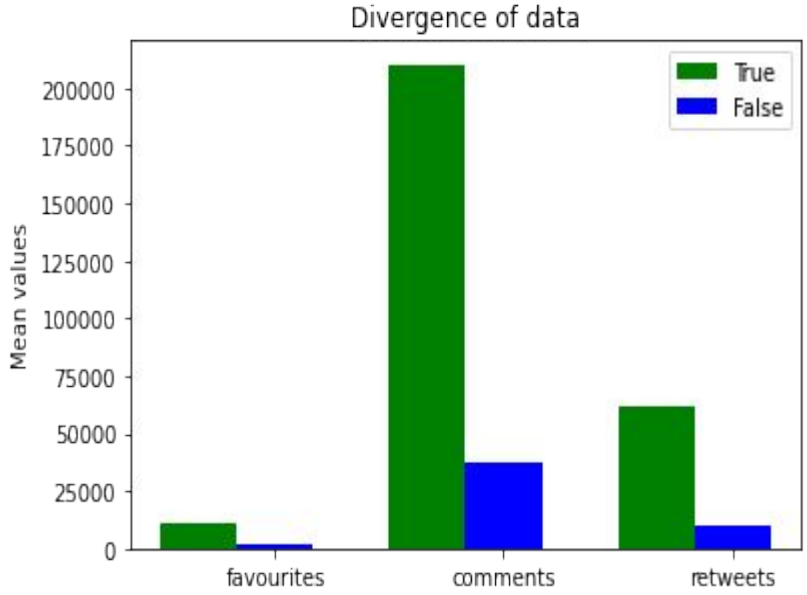
The lower value shows lower values in tail data. Thus increased kurtosis indicates **extreme data,** which we see only in the comments**.**

Skewness has all values positive, this shows that our bell curve is distorted to the right. So there is a distortion of data.

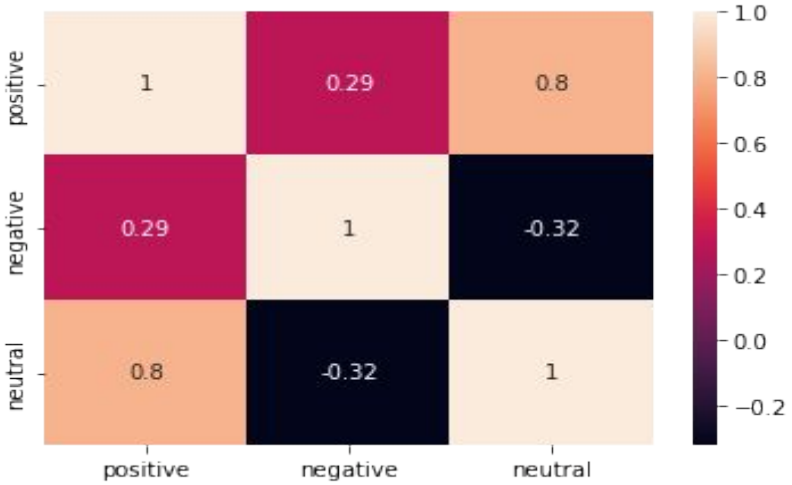
Divergence is data is observed by this graph. We see that the maximum divergence occurs because of the difference in the means of true and false information, especially in the number of comments.

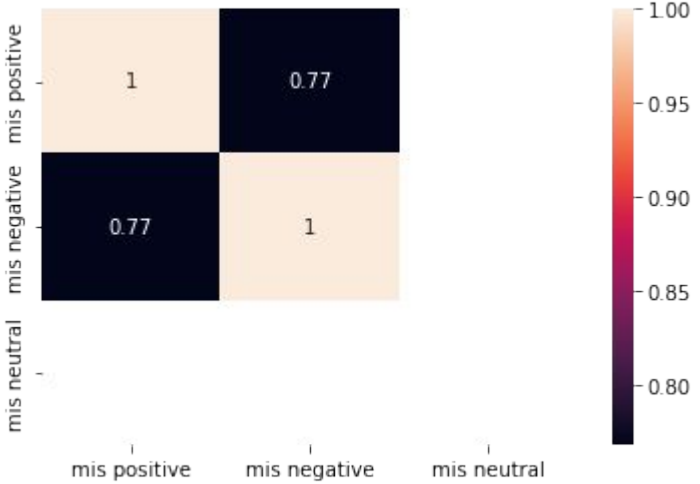
So diffusion index is higher in retweets and comments when compared to favourites.

We see that overall, true information diffuses more that mis-information.



Correlation





The correlation graph is made on the various sentiments (positive,negative and neutral) of the users. We see that a positive correlation indicates that information is **more likely to be retweeted.**

A negative correlation is lesser retweeted. This analysis gives the relation between **veracity** and network diffusion characteristics. We see that there is no neutral misinformation in our dataset.

We use the **Pearson's coefficient** as a metric of correlation. It uses covariance and gives the relationship between the variables.

The closer the relationship, the stronger the veracity and network diffusion of the information.

Pearson's Point Binomial Correlation Coefficient is used when a variable is dichotomous and this is a special case of the normal Pearson's coefficient.

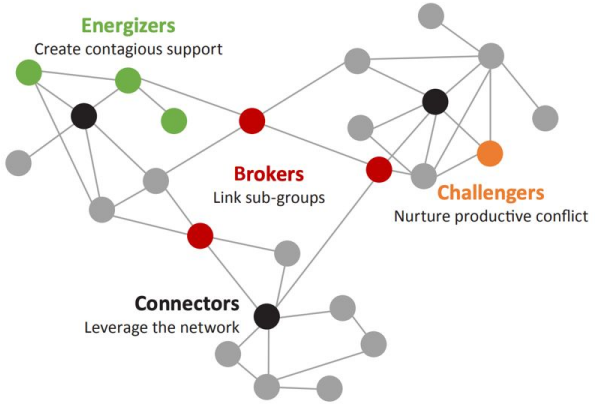
Formula:

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

Network Analysis

- Network Analysis helps us in deep understanding the structure of a relationship in social networks, a structure or process of change in natural phenomenons, or even the analysis of biological systems of organisms.

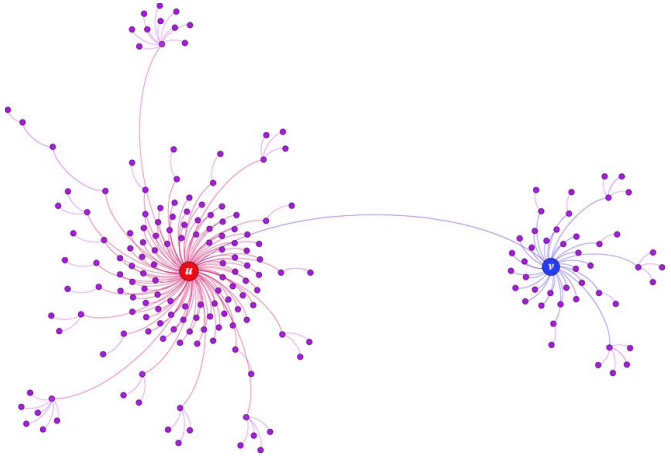
- Use cases:
  - Identifying the most influential person/people in a group
  - Defining characteristics of groups of users
  - Prediction of suitable items for users



Diffusion Networks

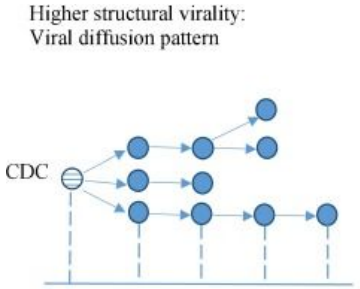
- Network diffusion captures the underlying mechanism of how events propagate throughout a complex network.

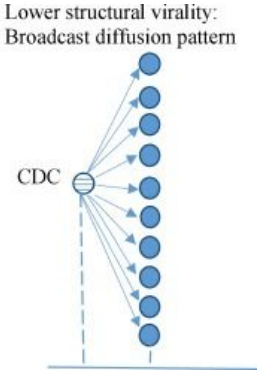
- Questions in case of misinformation:
  - How fast will it spread?
  - How will the system as a whole react?
  - What are the hubs in the network that are more connected and therefore more important than others



Diffusion Characteristics For Misinformation Tweets

- No. of Likes
- No. of Retweets
- No. of Comments
- Range of retweet (Timestamp)
- The Structural Virality (Goel. et al., 2016)



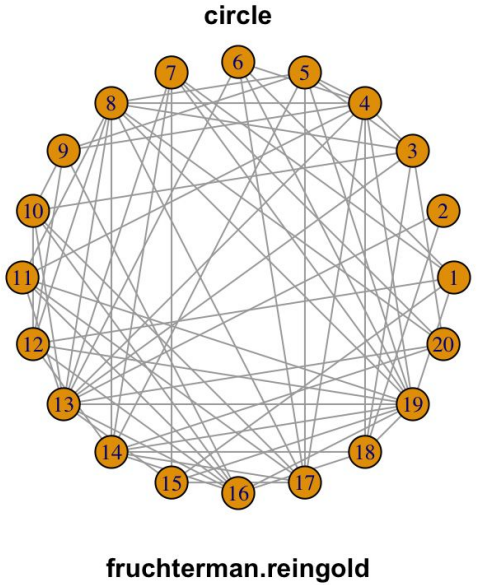


Diffusion Network Challenges

- We'll be using Fruchterman Reingold Layout for graph of forwarding network. Reasons:
    - Shows spread/flow of information between nodes
    - Equidistant nodes mean clear edges between nodes.
    - Different classes of nodes and their edges can be compared more clearly

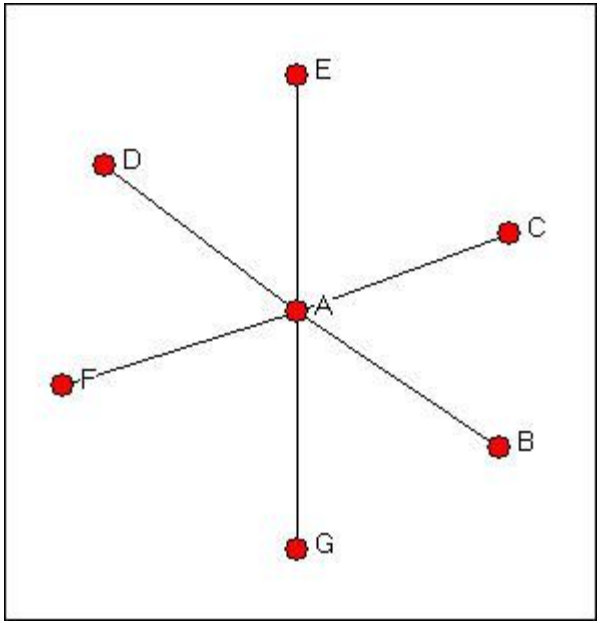
- Individual tweets would be used as nodes where user engagement is used as a metric to highlight certain tweets

- Retweeting relationships would be treated as edges in the network.

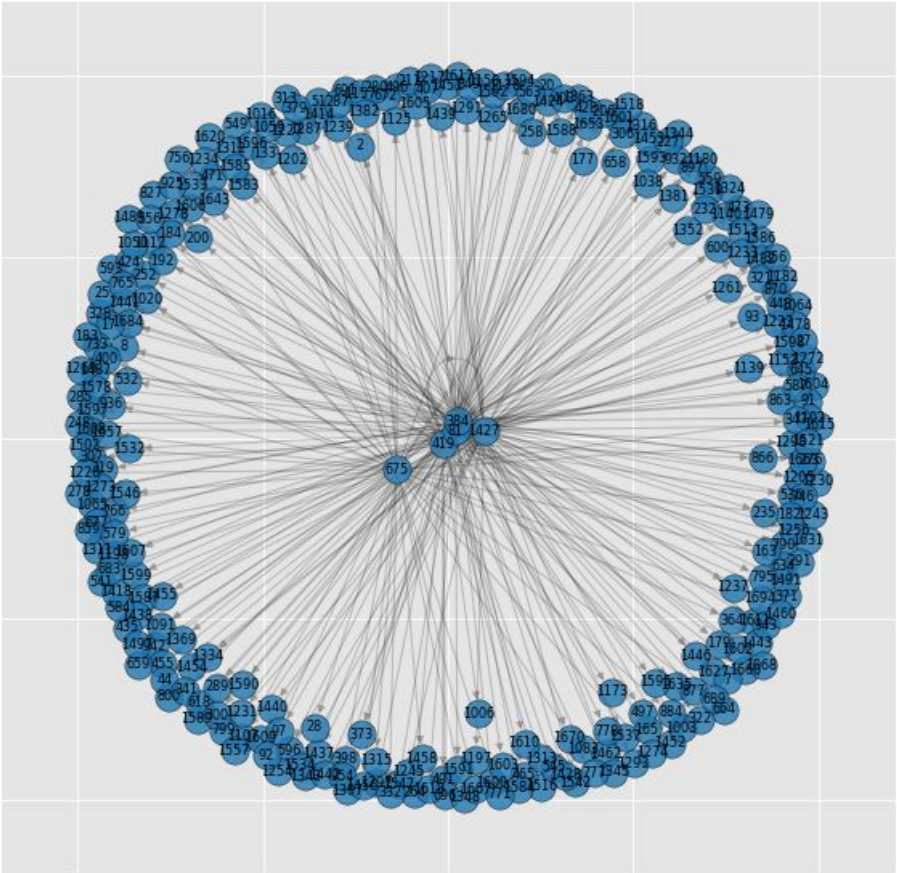


Diffusion Network Challenges

- Since API only shows original tweet for retweeted tweets. The network would form star topologies and may not be able to accurately show diffusion structure.
- To show diffusion better distance from centre would be proportional to time of tweet.
- All intermediary retweets would be considered as nodes in path.
- We'll use Weighted Softmax Function to Resolve retweet paths



Building Retweet Cascades



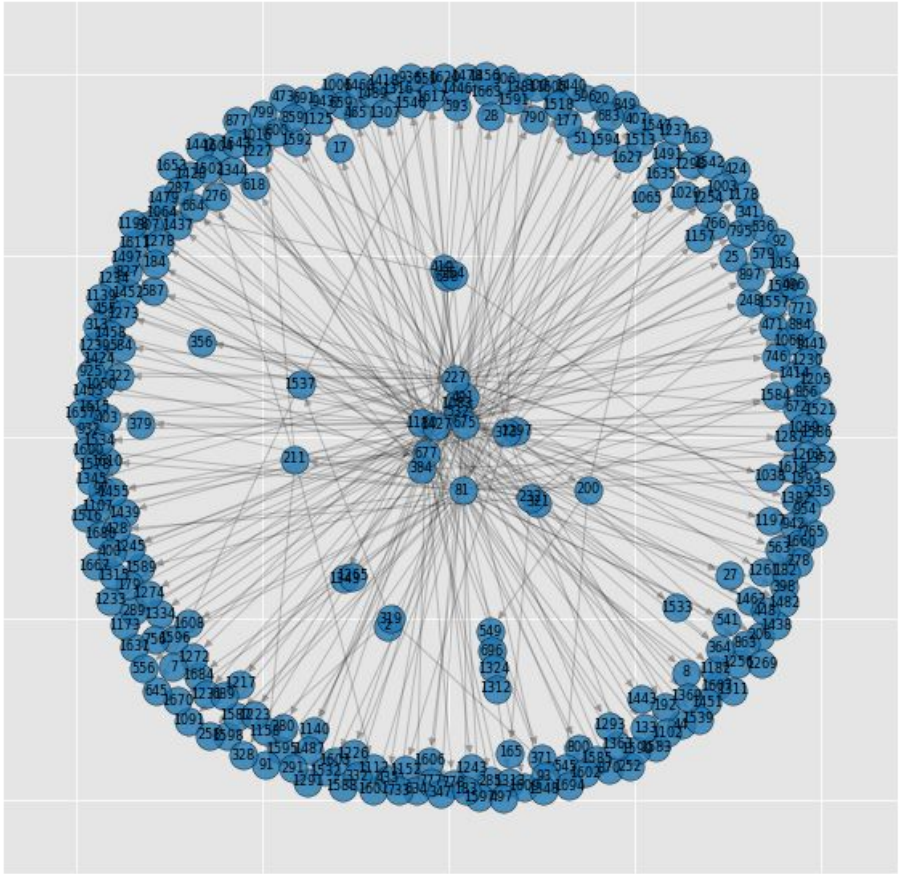
Building Retweet Cascades

$$p_{ij} = \frac{m_i e^{-r(t_j - t_i)}}{\sum_{k=1}^{j-1} m_k e^{-r(t_j - t_k)}}$$

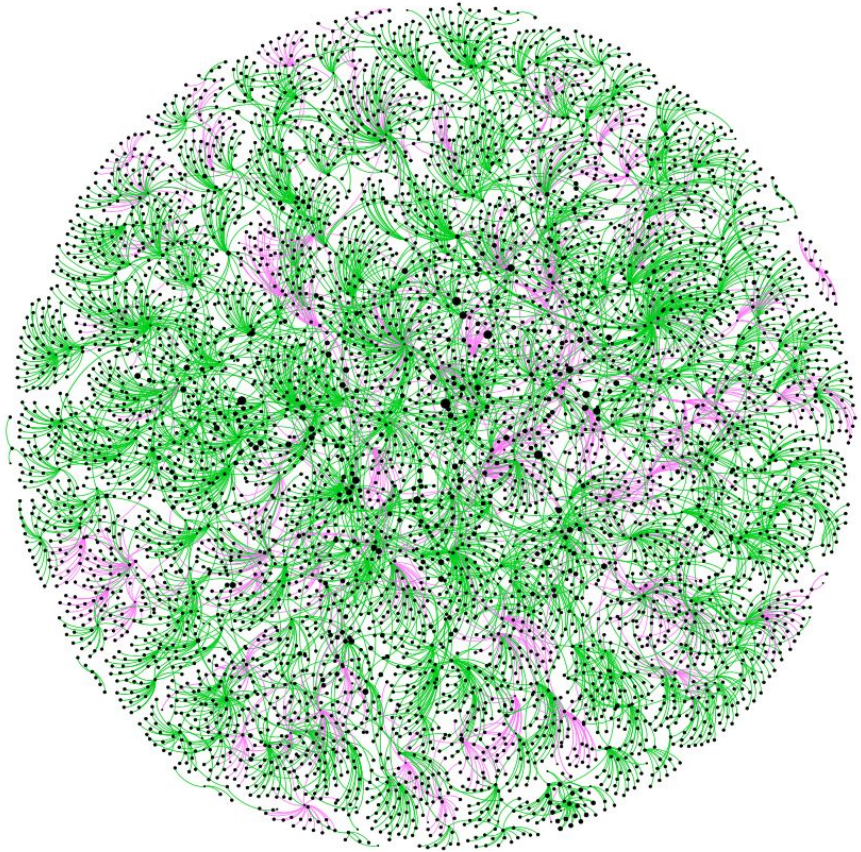
1. Probability of Retweeting Drop Exponentially with Time Difference

2. Users Prefer Locally Influential users

Building Retweet Cascades



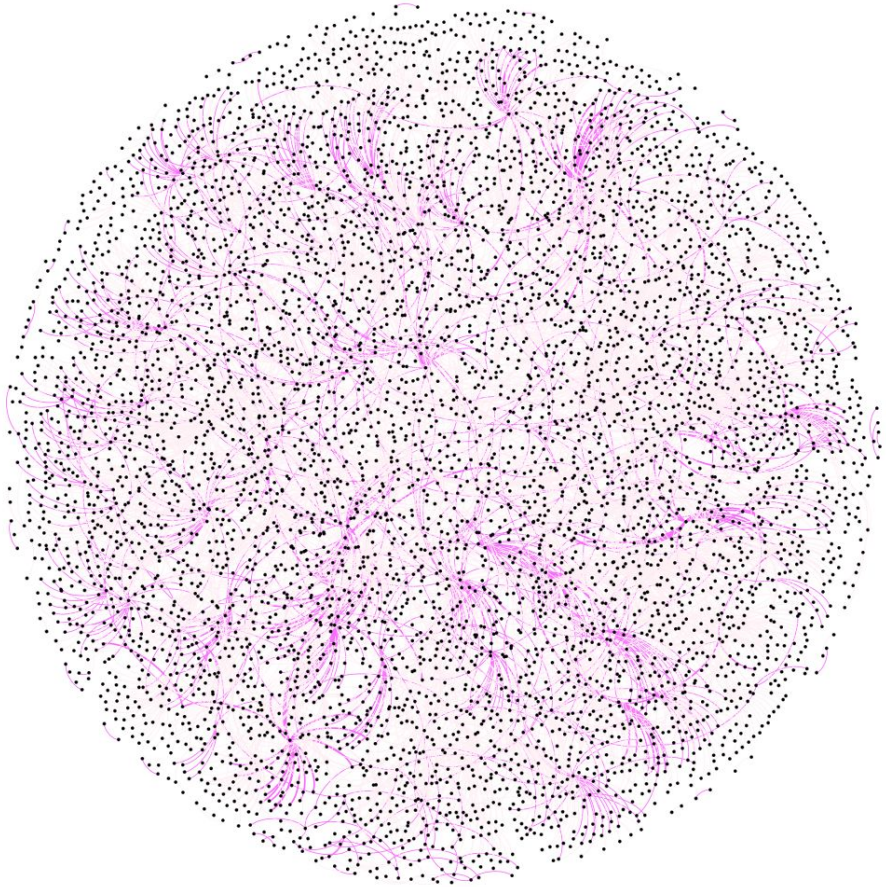
Misinformation vs True Information



Misinformation

True Information

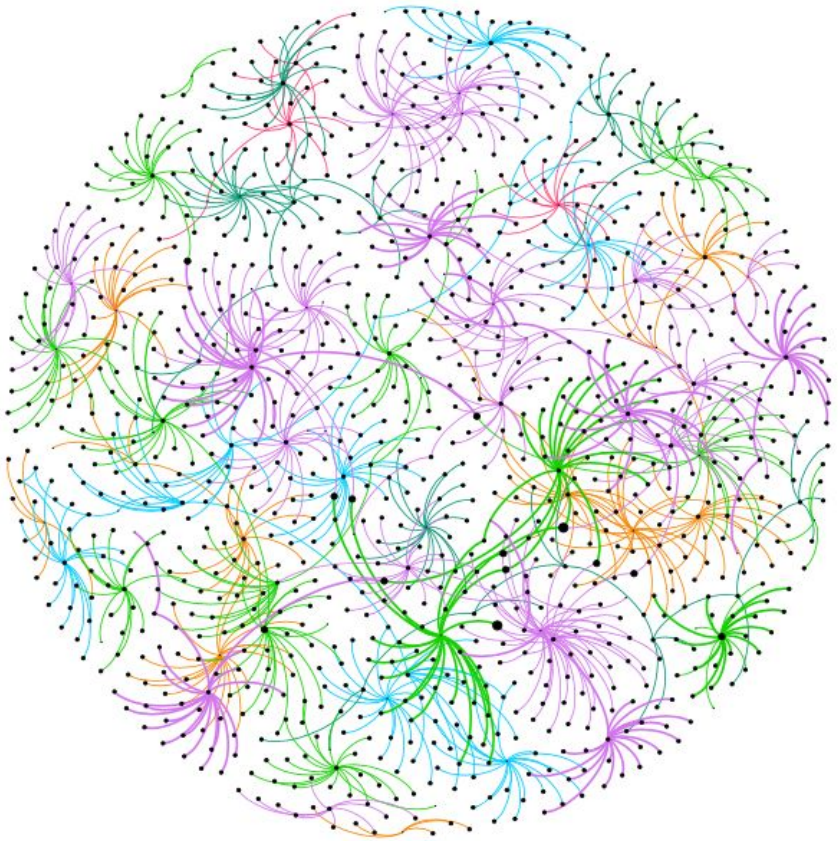
Misinformation vs True Information



Misinformation

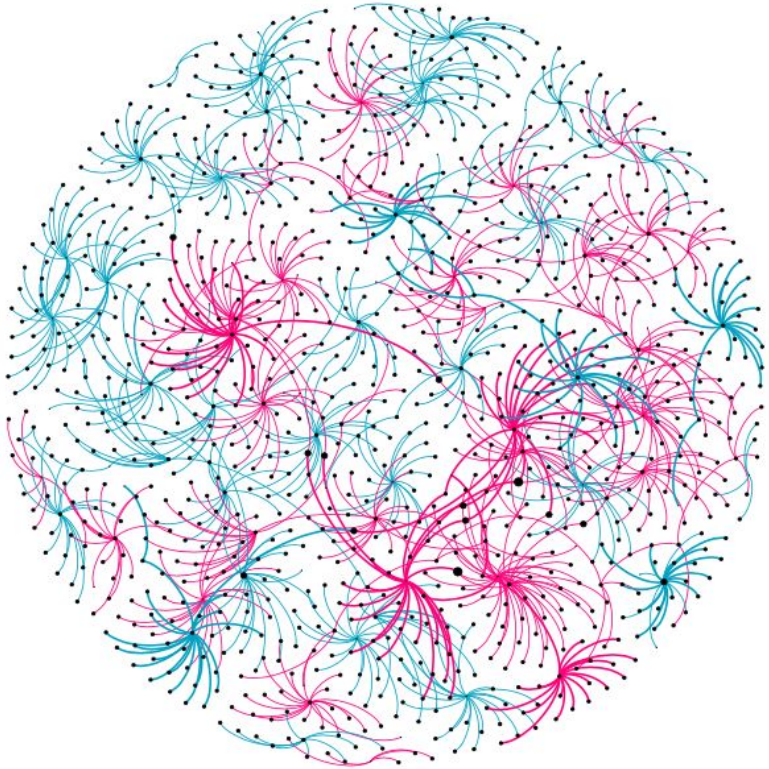
True Information

Misinformation By Topics



	Avg. Cascades
Religion	128
Health	97
Celebrities	44
War	73
Politics	108
Crisis	83

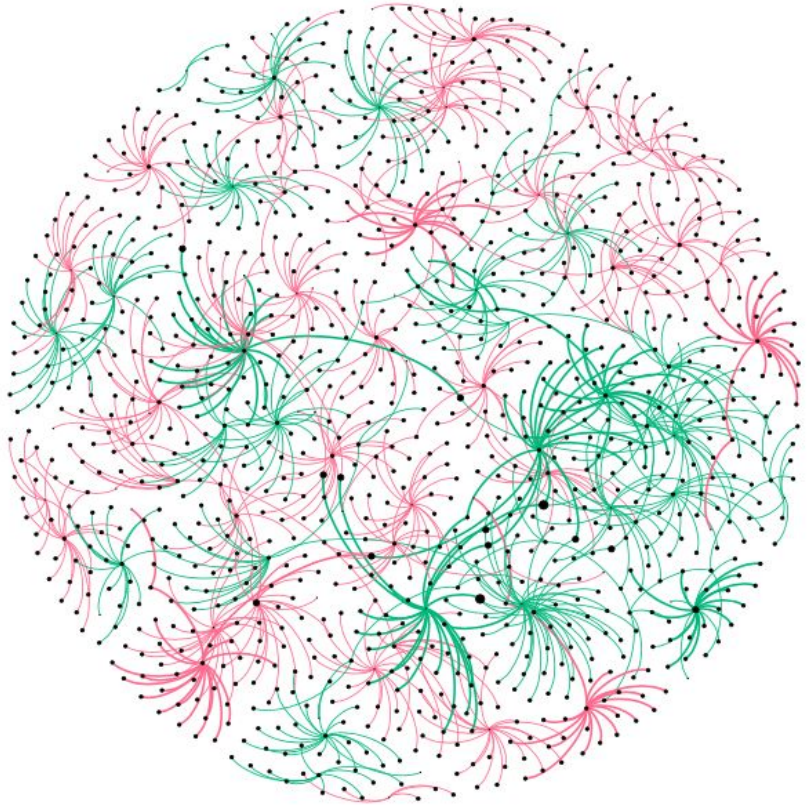
Misinformation By Sentiment



Negative

Positive

Misinformation By Media Presence

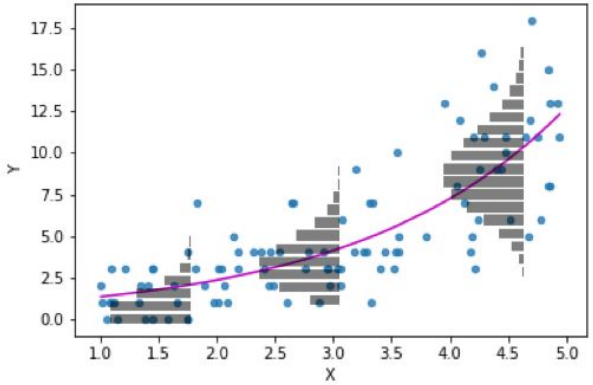


Media Present

Not Present

General linear Model

- The General Linear Model (GLM) is a useful framework for comparing how several variables affect different continuous variables.
- It is mainly used for model specification. It helps us to arrive at the exact equation that will most accurately summarize the data, in other words it allows us to summarise the research outcomes.
- It is important as incorporation of the wrong model will introduces biases and will not be able to describe the dataset accurately.
- We will use the General Linear Model (GLM) to examine the differences in diffusion characteristics between different categories of topics in true information and misinformation.



General linear Model

- Based on expected number of favorites is more likely to be spreading misinformation than information.
- Tweet classified as belonging to health, based on number of favorites is more likely to be spreading information rather than misinformation.
- The number of retweets and comments, the diffusion spread is varied and means are at different levels for each category.

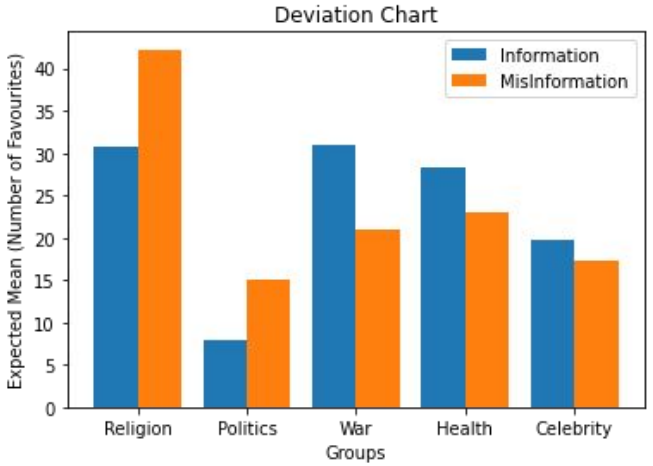


Figure : Mean Number of favorites

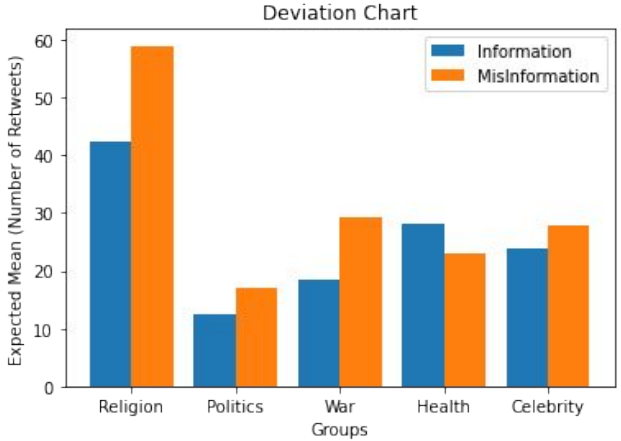


Figure : Mean Number of Retweets

THANK YOU

References:

- https://www.researchgate.net/figure/Diffusion-network-examples-The-network-visualization-in-a-shows-that-rumors-involve-a_fig2_312367197
- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6231818/
- https://www.sciencedirect.com/science/article/pii/S0747563218303613#bib19
- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7123536/