

CS185C Final Project

Power of BTS ARMY for Social Change Envisaged by Twitter Network Analysis

Vrinda, Inhee & Anirudh

<https://github.com/TheChirpyWitch/BLMAndBTSArmy>

Power of **BTS ARMY** for Social Change Envisaged by Twitter Network Analysis

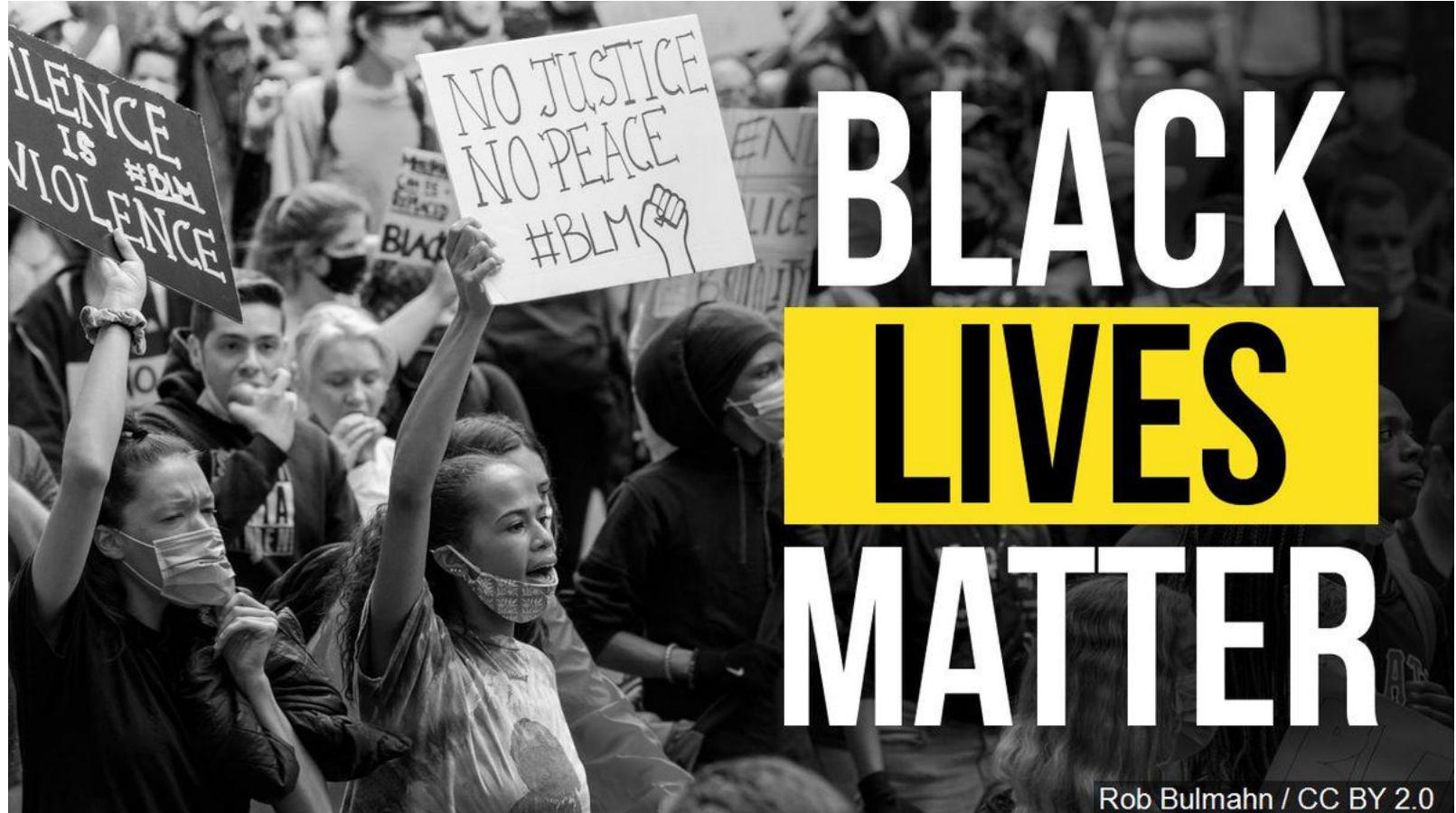
1. Motivation

Vrinda

Inhee

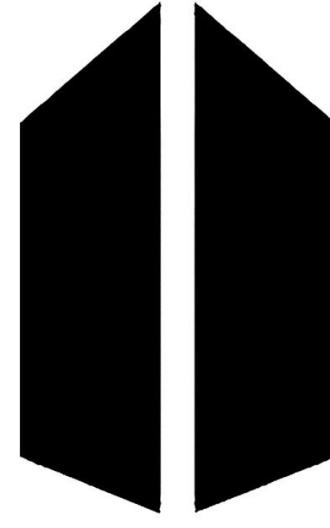
Anirudh

Motivation



Rob Bulmahn / CC BY 2.0

Motivation



Motivation

- In June 2020, **BTS** donated one million dollars for **Black Lives Matter**.
- **ARMY** had mobilised via **Twitter** and decided to match the donation.
- They used Twitter hashtags **#MatchAMillion**, **#Match1Million** and **#MatchTheMillion** to spread word about the campaign.
- Within **24 hours**, the fundraising account @OneInAnARMY announced that they had met their goal of 1 million dollars for BLM.

Motivation

- BTS Army is known for supporting various **causes**, so the initial campaign was not surprising.
- What was surprising was the **speed** at which this feat had been completed.
- This 24 hour window became the focal point of our analysis.

Topics Discussed

1. Motivation
2. Twitter Data Crawling of #MatchMillion for 24 Hours
3. Data Pre-processing:
 - a. Retweets
 - b. Followers
 - c. User description
4. Networks Structural Analysis
 - a. Retweet Network
 - b. Followers Network
5. Dynamic Retweet Network (Evolution over 24 Hours)

Topics Discussed

5. Reasons for quick spread:

- a. Power Law Analysis: Influential nodes
- b. User Engagement Metric: Common Neighbors Algorithm
- c. Local Bridges Among Communities: High Connectedness
- d. Bridge Nodes : Bilingual Translators

6. Linguistic Analysis

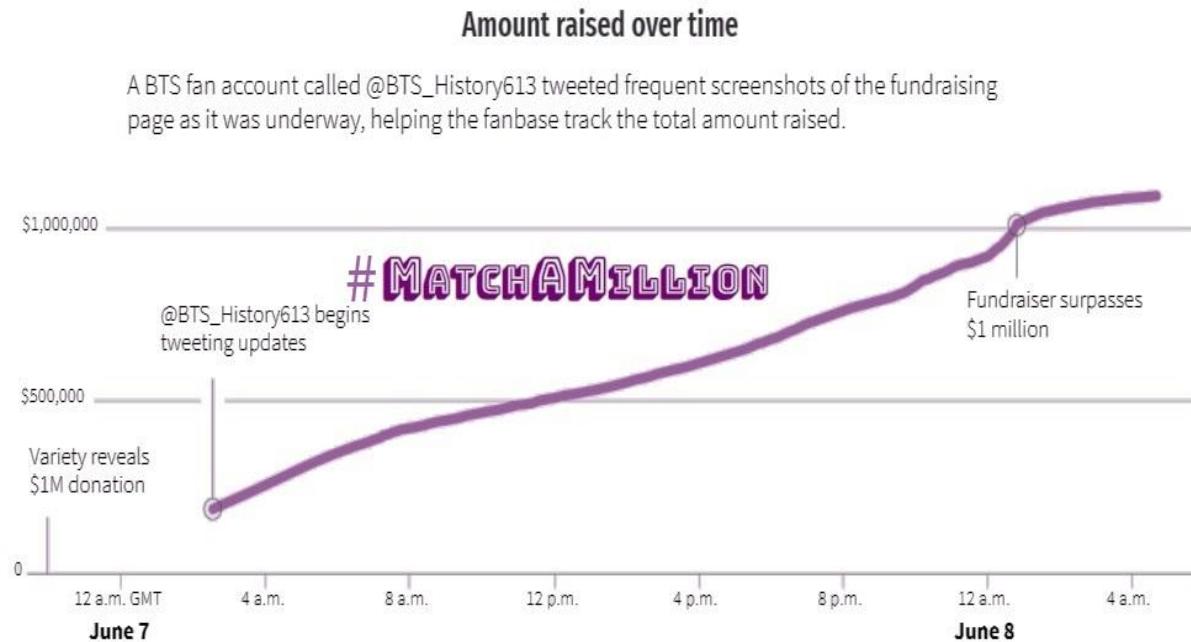
- a. Understanding the dataset using word clouds
- b. Sentiment Analysis of tweets
- c. NLP-based Behavioral/Functional Analysis
- d. Semantic Network Analysis
- e. Fan Classification: BTS Army or not?

7. Future Works --- Vrinda

2. Twitter Data Crawling for **#MatchMillion** for 24 Hours

Twitter Data Crawling for #MatchAMillion for 24 Hours

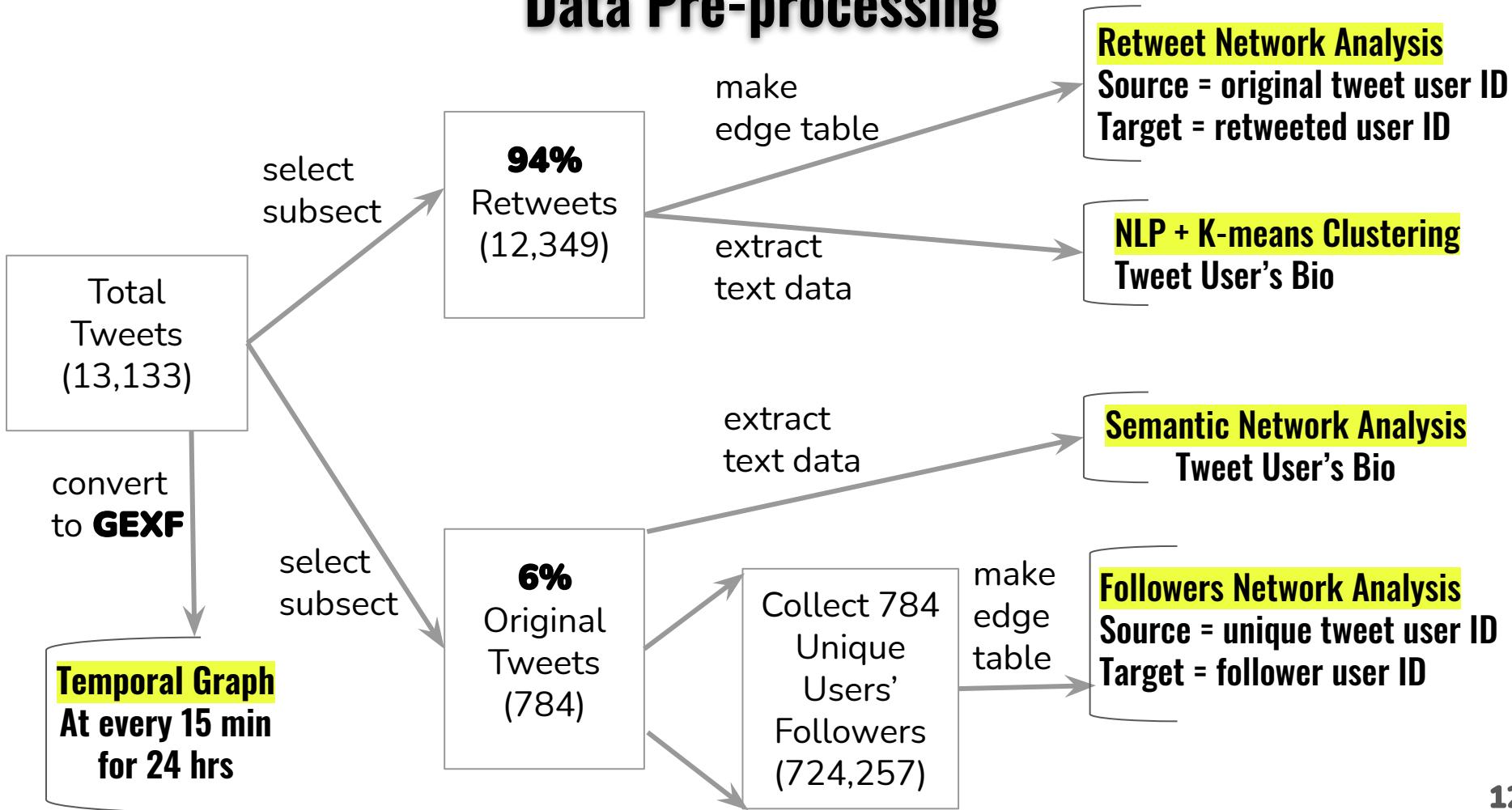
- Scrapped the Twitter data for one hashtag #MatchAMillion for 24 hours time window at every 15 minutes interval between 06/07/2020 - 06/08/2020.
- Total of 13,133 data points collected.



3. Data Pre-processing:

- ✓ Retweets
- ✓ Followers
- ✓ Temporal Data
- ✓ User description/bio

Data Pre-processing

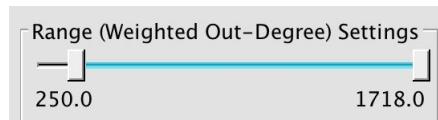


4. Networks Structural Analysis

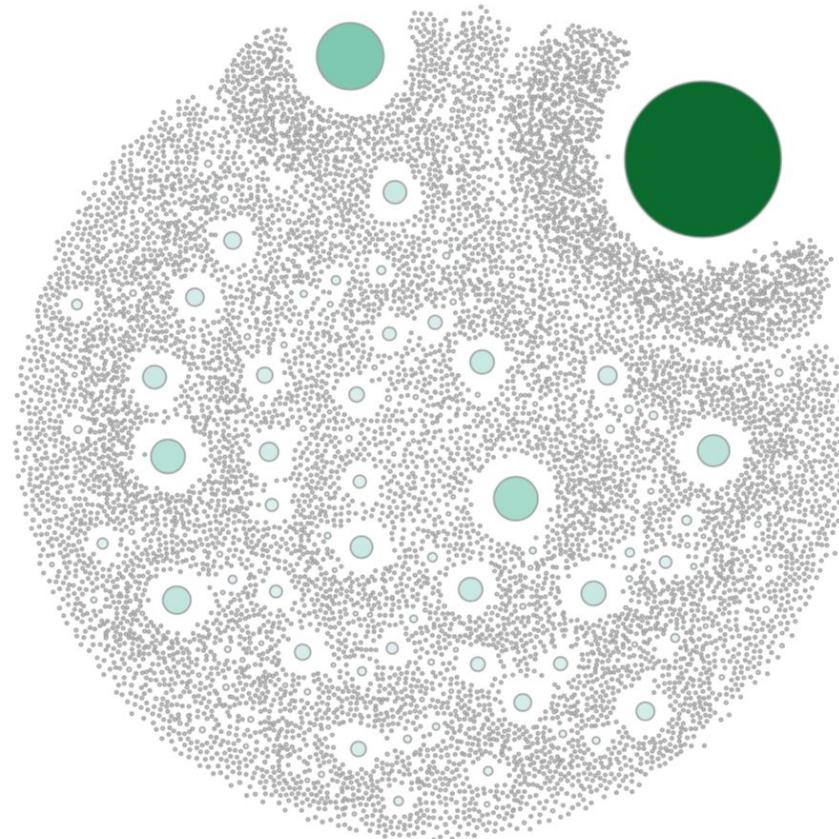
- ✓ Retweet Network
- ✓ Followers Network

Retweet Network - Characterization of Structure

- Directed, weighted graph
- Nodes = 10,309
- Edges = 12,050
- Connected components = 258
- Network diameter = 6
- Max out-degree = 1,718
- The 5 nodes with the highest Out-degree > 250

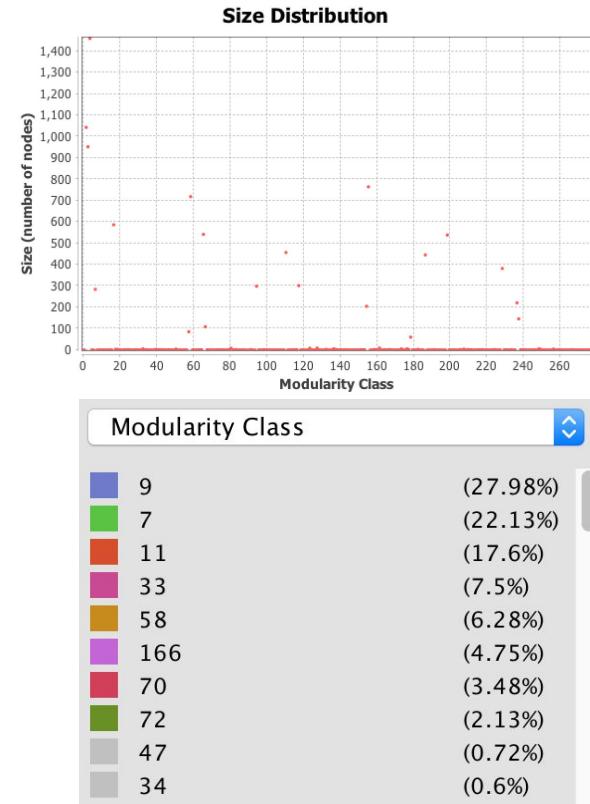
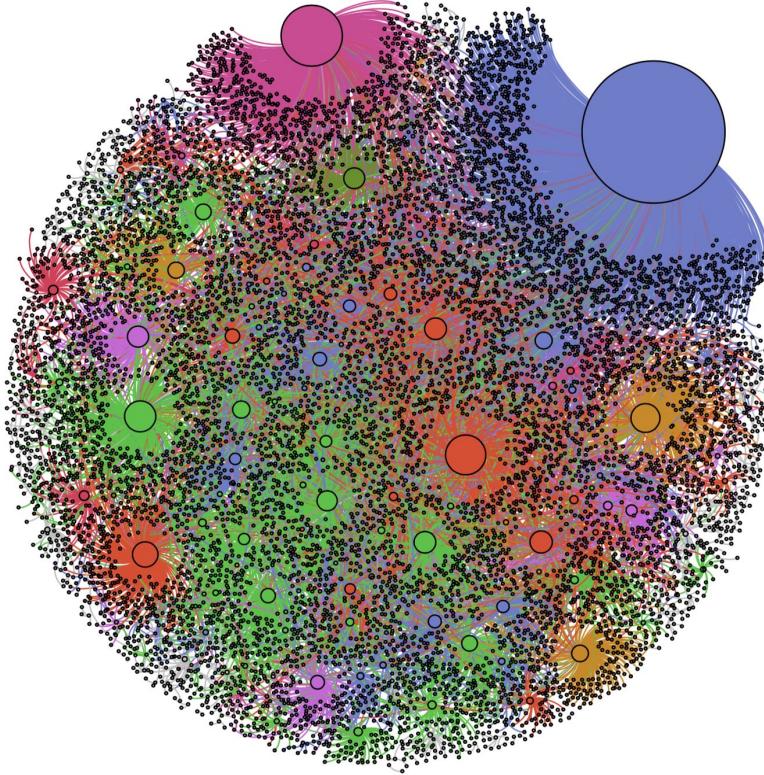


972611390757310464
1146555342286
107072215
1120396122172
1228903054872702978
2226802341058344182610300930



Retweet Network - Community Detection by Louvain Algorithm

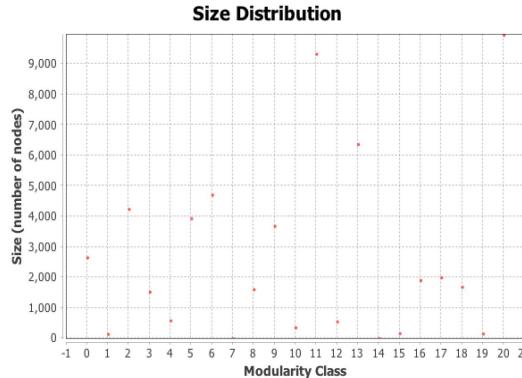
- Bottom-up algorithm; agglomerative clustering
- Node color = Community
- Node size = Degrees



- Approx. 10 communities with ≥ 100 nodes of total 10,309 nodes

Followers Network - Characterization of Structure

- Randomly select 100 users (original tweet writers) and their followers
- Nodes = 55,638 | 70,537
- Edges = 65,697 | 80,073
- Connected components = 4 | 5
- Network diameter = 3 | 3
- Node color = Community
- Node size = Degrees



Followers network shows
indeed **stronger ties** than
Retweet network w/ 21
communities! (<100)

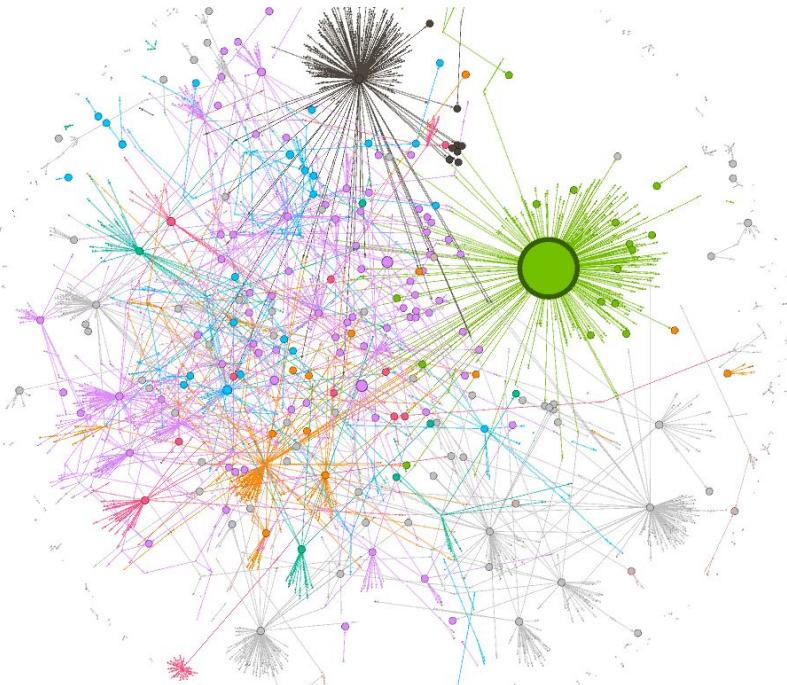
5. Dynamic Retweet Network (Evolution over 24 Hours)

Dynamic Evolution of Retweet Network over 24 Hours

Start

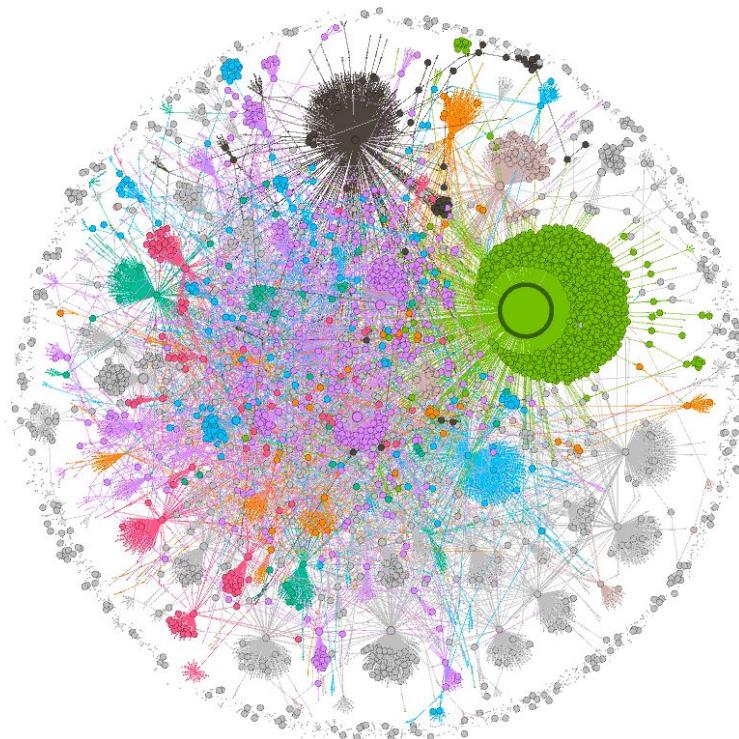


12 hour



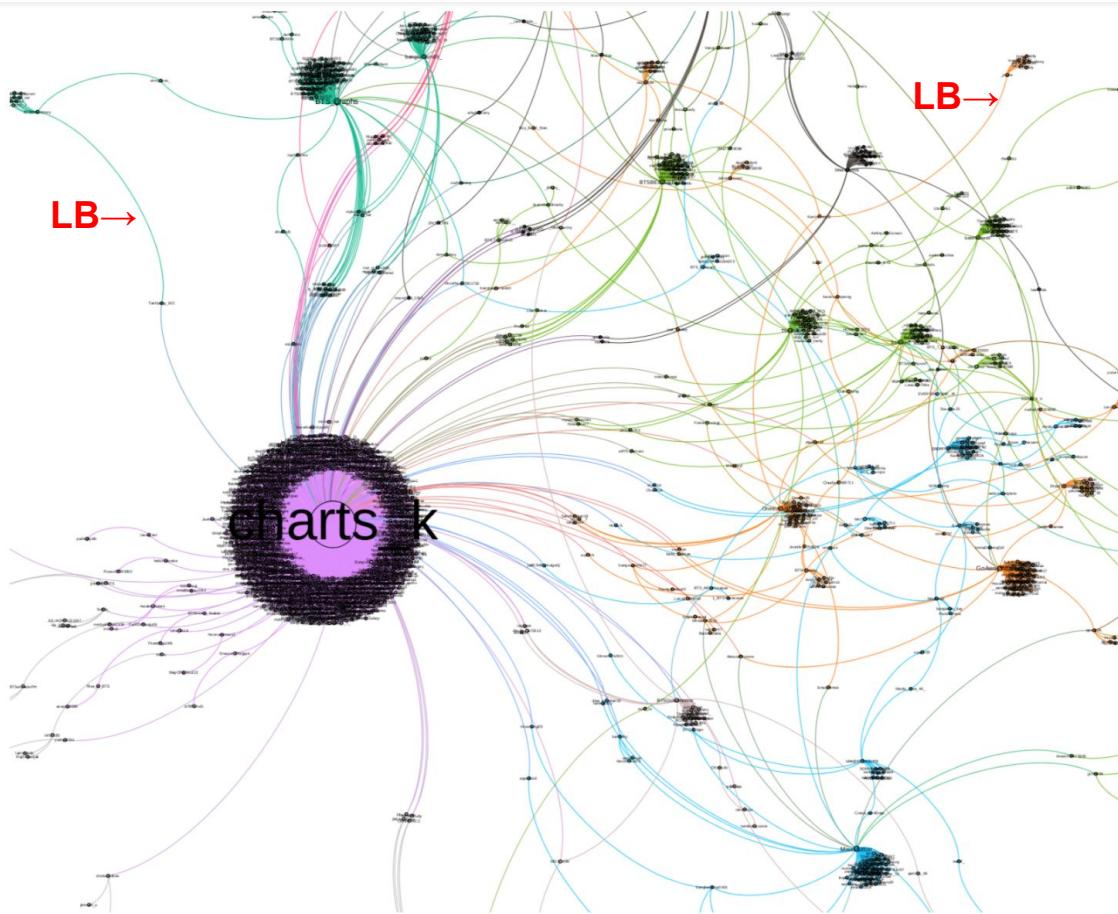
Dynamic Evolution of Retweet Network over 24 Hours

24 hour



Time Snapshot Reveals Local Bridge between Communities

- Snapshot on time interval 4:43 PM-5:00 PM reveals local bridges (LB)
- Majority cluster is centered at user "charts_k"



6. Reasons for quick spread:

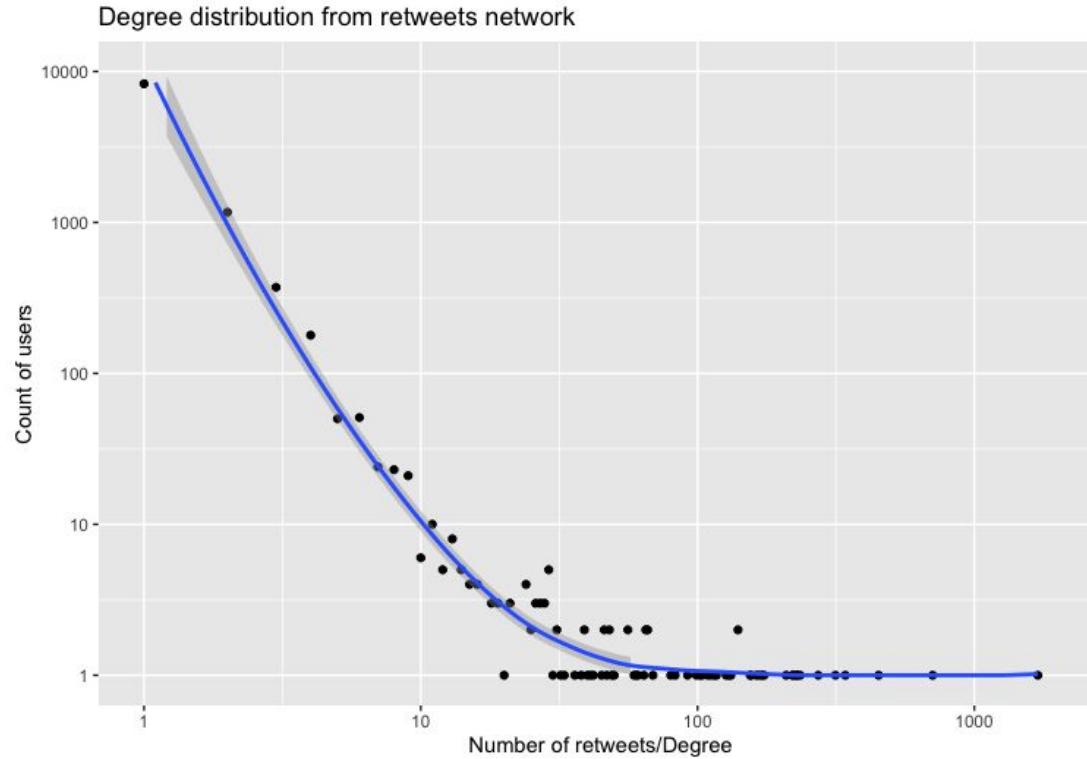
- ✓ Power Law Analysis: Influential Nodes User
- ✓ Engagement Metric: Common Neighbors Algorithm
- ✓ Local Bridges Among Communities: High
- ✓ Connectedness Bridge Nodes : Bilingual Translators

Reasons For Quick Spread: Power Law Analysis

1) Retweet Network:

- Total Nodes: 10,309
- Max degree of non-fan tweet: 372
- Max degree of fan tweet: 8,285
- Most influential tweets:
 - @charts_k: 8,285
 - @OneInAnARMY: 1,168
 - @BTS_History613: 1,168
 - @GoAwayWithJae: 372

Reasons For Quick Spread: Power Law Analysis



General Retweet
Threshold: 10

Our Retweets
Network
Threshold: 30

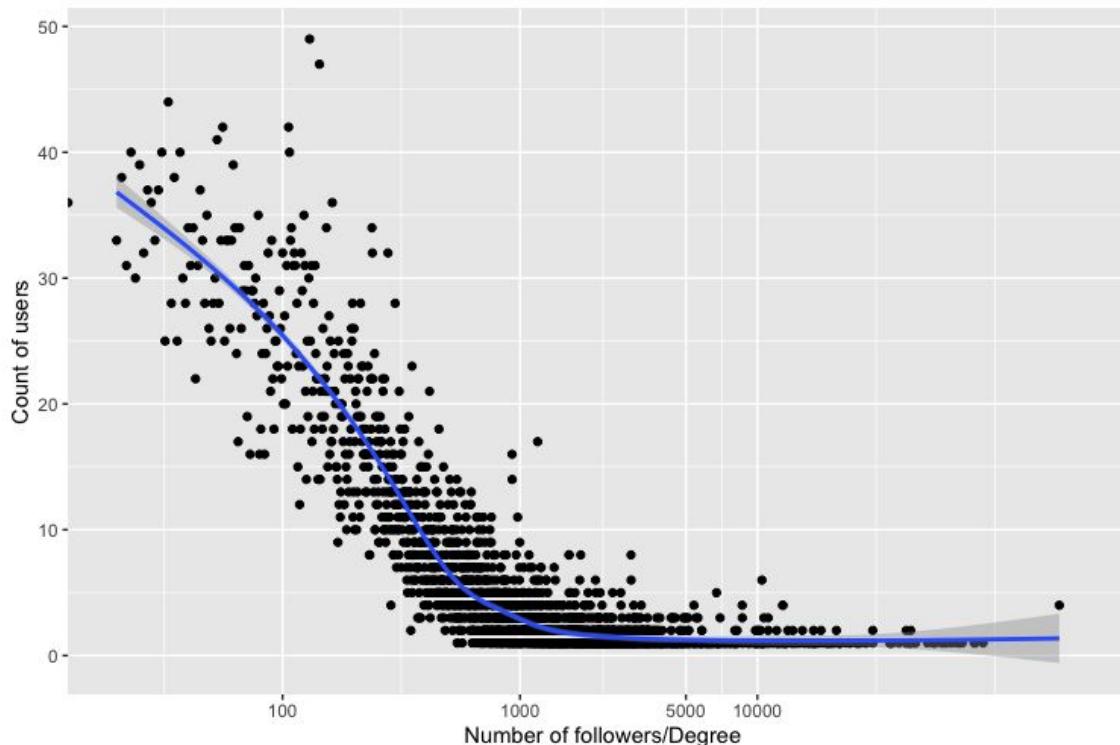
Reasons For Quick Spread: Power Law Analysis

2) Follower Network:

- Total Nodes: 10,956
- Max degree of non-fan account: 1,424,816
- Max degree of fan account: 279,062
- Most influential users:
 - @TheCut: 1,424,816
 - @charts_k: 279,062
 - @BTS_ARMYLeague: 262,027

Reasons For Quick Spread: Power Law Analysis

Degree distribution from followers network



General

Threshold:
1000

Our Followers
Network

Threshold:
3000

Reasons For Quick Spread: High Engagements

- Engagement: Relation between static followers and dynamic real-time retweets.
- General Common Neighbors Algorithm
 - Based on similarity between neighbourhoods of two nodes u and v
 - $\text{Similarity}(u, v) = |\Gamma(u) \cap \Gamma(v)|$
 - $\Gamma(u)$ and $\Gamma(v)$ are neighbours of u and v

Reasons For Quick Spread: High Engagements

- Adapted Common Neighbors Algorithm
 - Based on similarity between neighbourhoods of the same node in two networks
 - $\text{Similarity}(u_i, v_i) = |\Gamma(u_i) \cap \Gamma(v_i)|$
 - $\Gamma(u_i)$: neighbourhood of node i in retweets network
 - $\Gamma(v_i)$: neighbourhood of node i in followers network
 - $\Gamma(u_i) \cap \Gamma(v_i)$: number of followers who retweeted

Reasons For Quick Spread: High Engagements

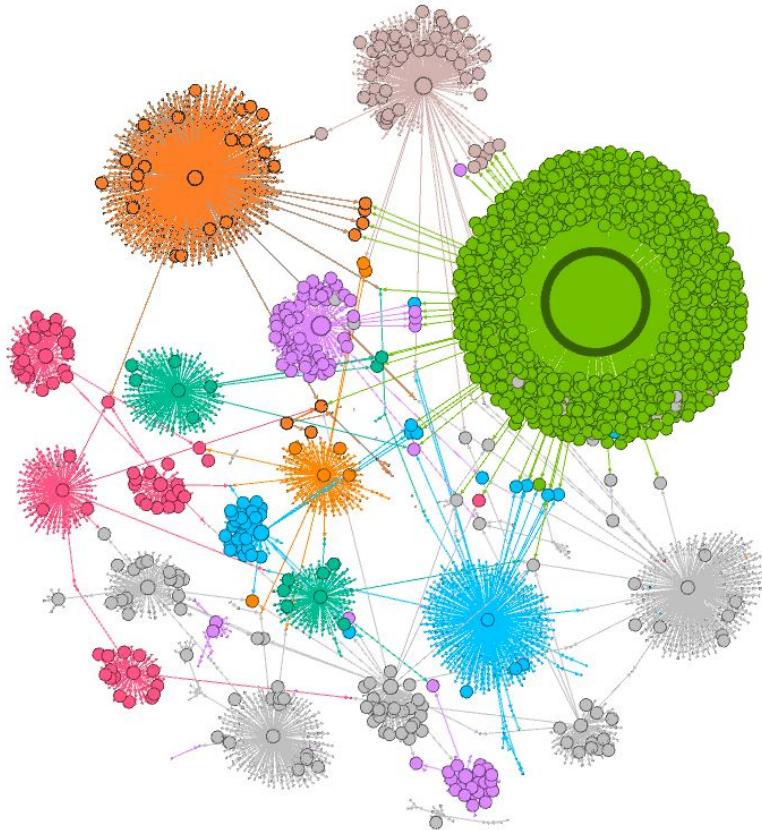
User engagement

$$= \frac{\text{No. of retweets by followers}}{\text{No. of followers}}$$

$$= \frac{|\Gamma(u_i) \cap \Gamma(v_i)|}{\sum(v_i)}$$

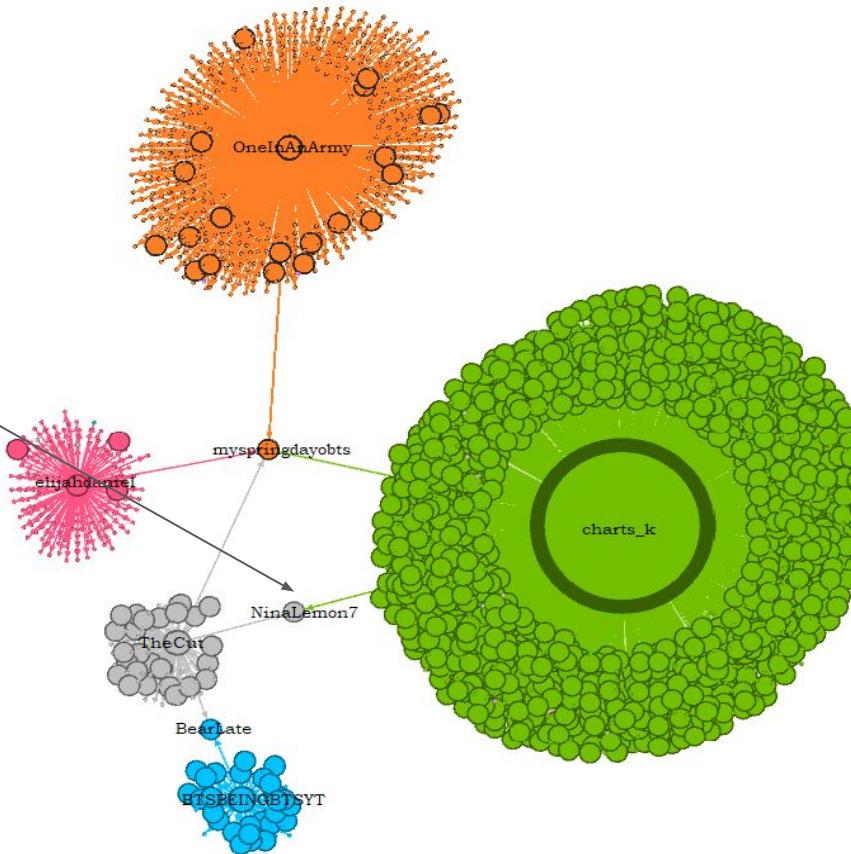
user_name	engagement
NakitaGroenewa1	0.25
BTS_History613	0.0964
smilingbangtans	0.071428571
bones2blossoms	0.066667
leaoutsold	0.056338
kimia_82j	0.052632
joonhopekook	0.050658
kimpark2017	0.046358

Reasons For Quick Spread: Local Bridges Among Communities



Reasons For Quick Spread: Connectedness Bridge Nodes

Bilingual translator:
@NinaLemon7



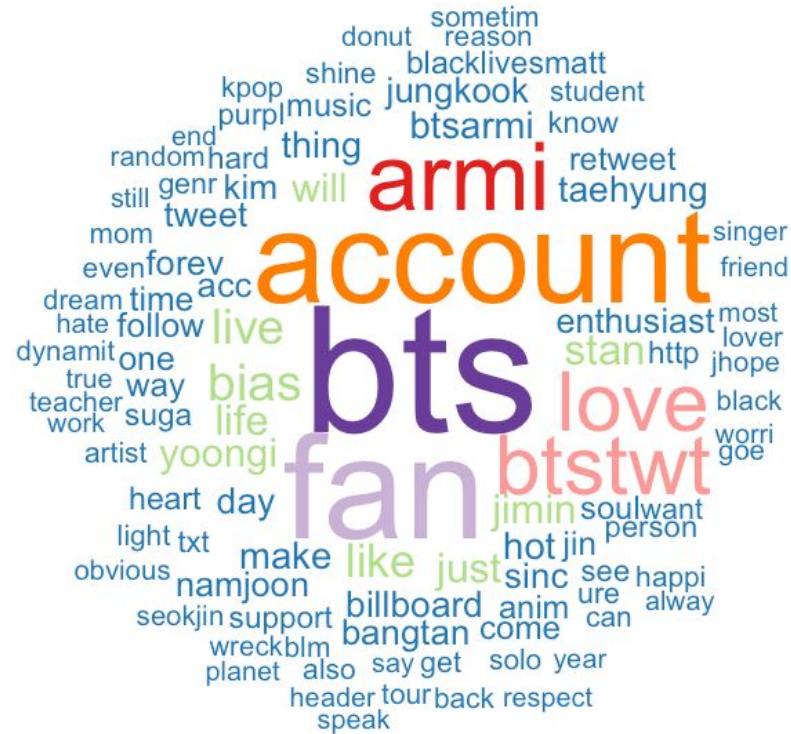
7. Linguistic Analysis

- ✓ Word Cloud
- ✓ Sentiment Analysis of tweets
- ✓ NLP-based Behavioral/Functional Analysis
- ✓ Semantic Network Analysis for Themes
- ✓ Fan Classification: BTS Army or not?

Linguistics: Understanding Dataset by Word Clouds

- Most frequent words in user_descriptions:

- **bts: 205**
- **fan: 164**
- **account: 144**
- armi: 97
- btstwt: 81
- love: 81
- bias 39
- live 36
- like 33
- stan 28



Linguistics: Understanding Dataset by Word Clouds

- Most frequent words in tweets:

- **matchamillion 513**
- **donat 224**
- **btstwt 205**
- **armi 185**
- blacklivesmatt 96
- can 89
- bts 88
- proud 58
- oneinanarmi 55
- million 49



Linguistics: Understanding Dataset by Word Clouds

Difference between user_description and tweets

user_description	tweets
Constant for a long time	Real-time
More likely to describe interests, likes, dislikes	More likely to describe spontaneous feelings and thoughts
Reflects the user	Reflects the state of the network
Can be used to check whether a user belongs to a particular community	Can be used to understand the news propagation and overall health of a network

Linguistics: Sentiment Analysis using Naive Bayes

Positive Tweets	Negative Tweets
537/842 ~ 63.7%	305/842 ~ 26.3%
<ul style="list-style-type: none">oneinanarmy i have a lot of money atms so this be all i could do but happy to help matchamillion	<ul style="list-style-type: none">listen ... i 'm broke blacklivesmatter matchamillion

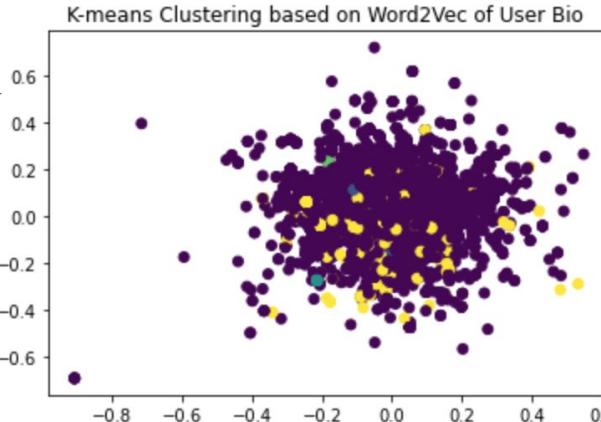
The figure consists of two word clouds side-by-side. Both word clouds have 'matchamillion' as the central, largest word. The left word cloud, representing positive tweets, contains words like 'blacklivesmatter', 'btstwt', 'donate', 'army', 'you', 'for', 'please', 'link', 'close', 'see', 'army', 'you', 'reach', 'goal', 'thankyou', 'nts', 'btstarmy', 'all', 'armys', 'oneinanarmy', 'sleep', 'have', 'help', 'omg', 'say', 'insane', 'after', 'time', 'are', 'feel', 'para', 'wake', 'hour', 'less', 'matchamillion', 'today', 'contribute', 'surpass', 'miss', 'from', 'gon', 'matchamillion', 'before', 'wake', 'over', 'miss', 'time', 'when', 'fandom', 'want', 'theresay', 'dollar', 'even', 'blacklivesmatter', 'some', 'count', 'money', 'let', 'day', 'yes', 'right', 'million', 'for', 'but', 'btstwt', 'almost', 'take', 'now', 'they', 'love', 'are', 'just', 'btstwt', 'could', 'what', 'surpass', 'who', 'too', 'from', 'gon', 'matchamillion', 'before', 'wake', 'over', 'miss', 'time', 'when', 'fandom', 'want', 'theresay', 'dollar', 'even', 'blacklivesmatter', 'some', 'count', 'money', 'let', 'day', 'yes', 'right', 'million', 'for', 'but', 'btstwt', 'almost', 'take', 'now', 'they', 'love', 'are', 'just', 'btstwt', 'could', 'what', 'every', 'give', 'really', 'contribute', 'link', 'para', 'than', 'get', 'have', 'all', 'the', 'goal', 'guys', 'insane', 'close', 'feel', 'army', 'not', 'please', 'black', 'know', 'able', 'our', 'help', 'much', 'already', 'less', 'oneinanarmy', 'btstarmy', 'see', 'matchthemillion', 'more', 'today', 'amount', 'hour', 'until', 'thankyou', 'nts', 'organization'.

Linguistics: NLP-based Behavioral/Functional Analysis

- Word Embeddings of All Tweet User's Bio Text by **Word2Vec**
- K-means Clustering with Features of Word2Vec

"sugar & spice"
"i miss @bts_twt #BlackLivesMatter #BLM"
"2014. THE GENRE IS BTS. YOONMIN VISUAL."
"whip and neigh neigh lesbian for goro majima..."
"I look forward to the day you will stand and ..."
...
"\ud83d\udc9cPhilippine ARMY\ud83d\udc9c #SanD..."
"\ud211\ud29c\ud28f \ud051\ud0f \ud026a \ud29f..."
"mono by rm disciple. \ud0dc\ud0dc\udaf9 amour..."
"\ud83d\udc9c In this Bangtan Sonymeondan ish f..."
"\ud351 \ud571\ud02b0\ud1d52\ud1d58\ud1d4d\ud02b0\ud02b0..."

[-0.1875 -0.22558594 0.
-0.01281738 0.0324707 0.
0. -0.16210938 0.28125
-0.22558594 0. 0.
-0.16992188 0. 0.03320312
-0.14648438 -0.22558594 0.28125
0.00393677 0. 0.00393677
0.10107422 0.05834961 -0.02941895
-0.17285156 -0.17578125 -0.06591797
0.1875 -0.07666016 0.0324707]



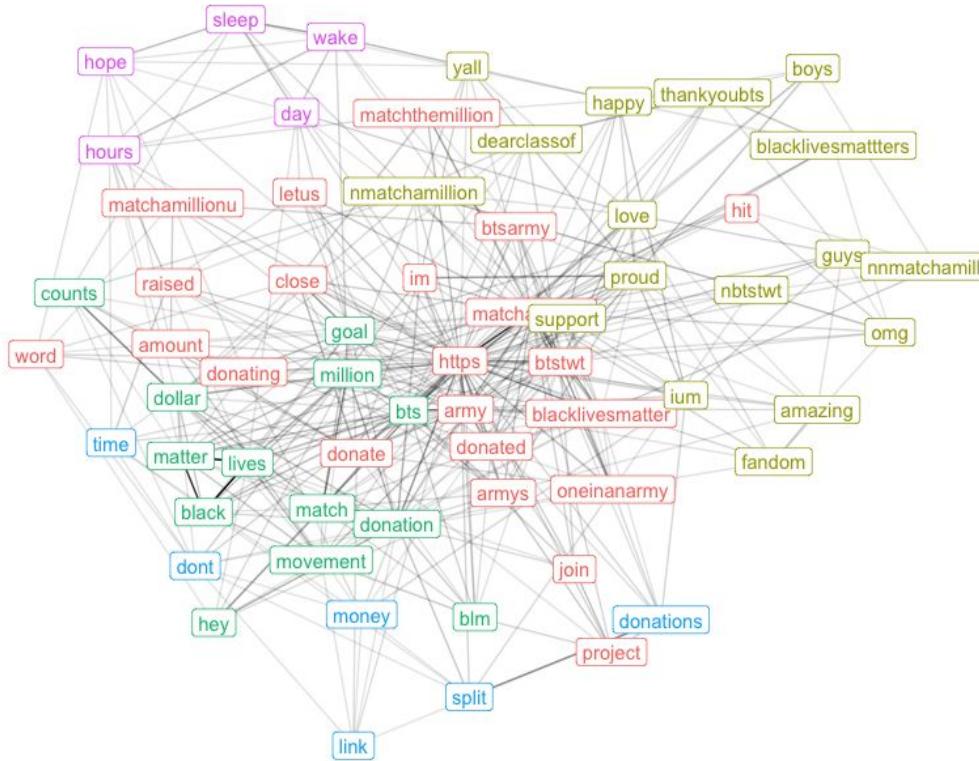
- Too fuzzy high-dimensional vectors do not discriminate
- Better to use explicit “Semantic Network Analysis for Themes”

Linguistics: Semantic Network Analysis for themes

- Filter words to get top 80% most frequently occurring words
- Find similarity between these words - cosine similarity is used here
- Plot a Word Similarity graph
- Run a clustering algorithm on the graph to find groups - igraph's cluster_walktrap is used here
- Each group would represent a theme

Linguistics: Semantic Network Analysis for themes

Clustering on
word similarity
graph for
tweets



Linguistics: Semantic Network Analysis for themes (1-gram)

Group 1: A call for action	Group 2: Overwhelming positive support for the cause	Group 3: Includes the tweets relating to goals in monetary collection	Group 4: Includes time management in the collection	Group 5: Is about experience of people involved
amount, army, armys, blacklivesmatter, btsarmy, btstwt, close, donate, donated, donating, hit, https, im, join, letus, word matchamillion, matchthemillion, oneinanarmy, project, raised	amazing, blacklivesmatters, boys, dearclassof, fandom, guys, happy, ium, love, bts, nmatchamillion, omg, proud, support, thankyoubts, yall	black, blm, bts, counts, dollar, donation, goal, hey, lives, match, matter, million, movement	donations, dont, link, money, split, time	day, hope, hours, sleep, wake

Linguistics: Semantic Network Analysis for themes (2-grams)

Clustering on
2-gram
similarity
graph



Linguistics: Semantic Network Analysis for themes (2-grams)

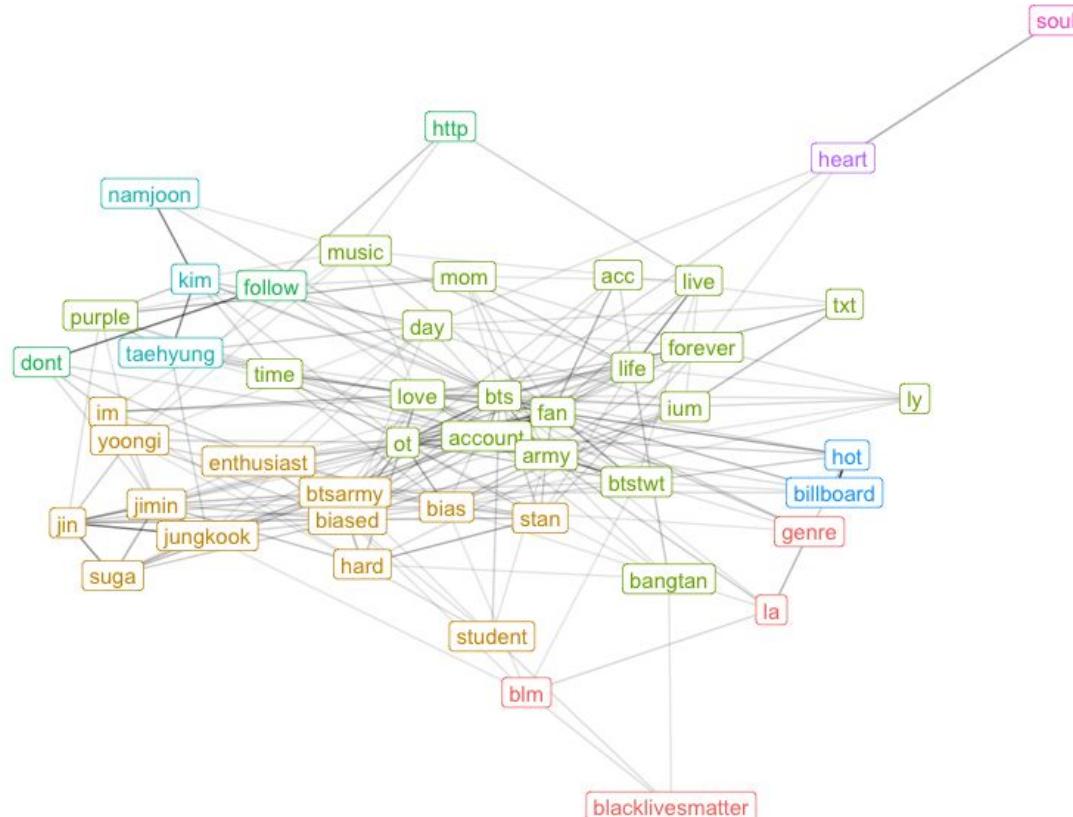
Group 1:	Group 2:	Group 3:
Encouraging tweets for armys and links shared	Personal accounts of donations	Day is almost at an end, fans asking each other to pull through
almost there, at k, blacklivesmatter btstwt, blacklivesmatter https, btstwt oneinanarmy, get it, i donated, in the, less than, matchamillion blacklivesmatter, matchamillion blacklivesmatters, matchamillion https, oneinanarmy btstwt, over k, than k, thank you, the matchamillion, this is, matchamillion, want to, will be	an army, btstwt matchamillion, but i, for matchamillion, going to, i could, i just, just donated, the word, to help, to matchamillion, up to, wake up, what i	a million, are so, army matchamillion, close to, matchamillion matchthemillion, matchthemillion https, so close, to the, we are, were so

Linguistics: Semantic Network Analysis for themes (user_desc)

Group 4:	Group 5:	Group 6:	Group 7:	Group 8:
Spreading news about the hashtag	Pride for the community's work	Encouraging others to donate politely	Tagging #MatchAMillion and BTS for engagements	Sending love to each other
black lives, bts army, donation to, lives matter, to match, trying to, we can	i am, im so, of this, part of, proud of, proud to, so proud, to be	able to, if you, to donate, you can	btstwt https, matchamillion btstwt	i love, love you, so much

Linguistics: Semantic Network Analysis for themes (user_desc)

Clustering on
word similarity
graph for
user_description



Linguistics: Semantic Network Analysis for themes

Group 1: Support for BLM	Group 2: Common army lingo	Group 3: Common army-adjacent lingo	Group 4: Directs user to a different account	Group 5: BTS members	Group 6: Relating to song charts	Group 7: Talking about interests
blacklivesmatter , blm , genre, la	bias, biased, btsarmy , enthusiast, hard, im, jimin, jin, jungkook, stan, student, suga, yoongi	acc, account , army, bangtan, bts, btstwt, day, fan, forever, ium, life, live, love, ly, mom, music, ot, purple, time, txt	dont, follow, http	kim, namjoon, taehyung	billboard, hot, song	heart, soul

Linguistics: Fan Classification: BTS Army or not?

- Dataset used: user_description
- Classification algorithm:
 - Keywords mined from various BTS Army resources and previous semantic analysis
 - Keywords assigned weights according to NER rules and ground truth
 - User_description tokens compared against keywords and assigned weights accordingly
 - Total weight for a user > 0.2 implies BTS Army account
- High number of true positives
- High number of false positives

8. Future Works

Vrinda

Inhee

Anirudh

Future Works

- Homophily of nodes using similarity metrics
- Fan classification using supervised learning
- Exploring Bridge Nodes using text analysis
- Affiliation network analysis for follower and retweet networks

Power of **BTS ARMY** for Social Change Envisaged by Twitter Network Analysis

Thank You!

Vrinda

Inhee

Anirudh