

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

## Motivation

In June 2020, the international sensation **BTS** (a group of 7 South Korean musicians) donated one million dollars for **Black Lives Matter**. Within 24 hours, their **fandom** called **ARMY** had mobilised via **Twitter** and matched BTS's one million dollar donation by using social media networks under the hashtags **#MatchAMillion**, **#Match1Million** and **#MatchTheMillion** as reported [here](#). [1]

This collective effort sparked the idea for our project. If you are on Twitter, you would have come across the BTS ARMY at some point. Their [mobilisation power for social causes](#) [2] is unlike any other group on the **internet with multiple collectives hosting fundraisers for various causes - but we want back that claim with numbers.**

We also want to analyse how BTS ARMY's reach led to this unprecedented donation and how it inspired people who might have been living on the other side of the world to donate for a worthy cause. As pointed out in this [article](#), [3] **ARMY** had help from other **fanbases** and people who were passionate about **Black Lives Matter Movement**. We want to try to see how these two clusters came together along with analysing some salient features of this **network using data scraped from Twitter.**

## Proposed Method

To achieve the goal outlined in the above section, we will follow these seven steps:

- 1) Twitter data crawling;
- 2) Construction of a network;
- 3) Measurement of structural metrics of the network ;
- 4) Community detection to infer the behavior/function from the network structure;
- 5) Monitor community evolution by constructing a dynamic temporal network (our tweet data has 24-hour span); and
- 6) Evaluation of each detected community by constructing wordcloud (expected to see a distinct dominant word from each community).
- 7) Based on the findings from the steps above, we will conduct further measurement of influence and homophily.[7] Basically we would like to observe how the influence and homophily generate similarity in our twitter-based social network.

## I. Creation of a social network

1. Data Scraping from Twitter API
2. Setting nodes and edges to create a network using:
  - a. Twitter's follower rules to obtain a directed network
  - b. NLP-based phrase matching using word embeddings

All three team members have already applied to the Twitter API Developer Accounts [4], having been granted the required credentials. We thus have already started to scrap the Twitter data for one of hashtags. To build a social network we should prepare the two types of data tables:

(1) "Node" table data: twitter user IDs can serve as Node information; we also compute some NLP-based numerics of each user (such as TF-IDF, word embeddings) from their recent 20 tweets or their profile description to classify whether the user might belong to BTS fans or belong to the BLM group.

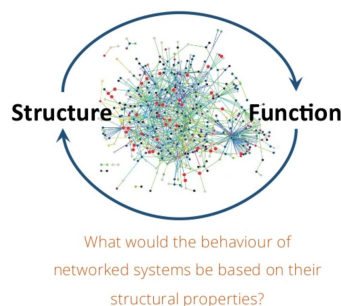
(2) "Edge" table data: connectivity between users (source and target) via retweeting will serve as directed Edge information; also, connectivity between users sharing same hash-tags (either BTS or BLM) will also serve as undirected but weighted (counted by co-occurrence of different hash-tags) Edge information. Basically, we will convert Tweets text to numeric vectors using NLP Embedding Methods to prepare for the node attribution data.

## II. Characterization of the network

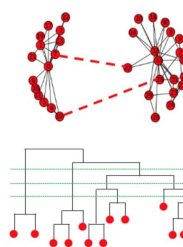
1. Number of nodes, edges and connected components.
2. The diameter (longest shortest path).
3. The five nodes with the highest clustering coefficients.
4. The five nodes with highest betweenness centrality (node betweenness).
5. Assortativity of the nodes.
6. Community detection for BTS ARMY using:
  - a. Node Similarity using NLP based word embeddings (ML-based clustering)
  - b. Girvan-Newman algorithm (edge-betweenness-based)
  - c. Louvain algorithm (modularity-maximization-based)
  - d. Hierarchical clustering (distance-matrix-based)

### Network Structure

- Path length
- Diameter
- Clustering coefficient
- Degree distribution
- Centrality measures



### Quantifying Community Structures



#### Divisive

- Girvan-Newman Algorithm (edge betweenness)

#### Agglomerative

- Hierarchical clustering (distance matrix)
- Louvain Method (modularity-maximization)

On one hand, from the network structure-based metrics, we can detect communities by mainly three methods: 1) divisive, edge betweenness-based “Girvan-Newman” algorithm; 2)

modularity-maximization-based “Louvain” algorithm; and 3) distance-matrix-based “Hierarchical” clustering.

On the other hand, we can also detect communities using the ML-based, unsupervised “Clustering” method using the pre-computed node attributes such as word embedding vectors.

Major difference between **community detection** and **clustering** is that in community detection, individuals are connected to others via a network of links, whereas in clustering, data points are not embedded in a network. [7]

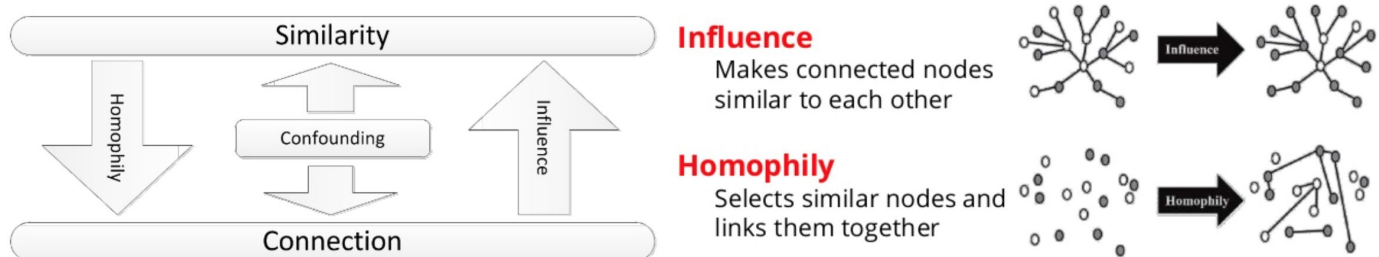
### III. Visualization of the network

1. A static visualisation of the networks using iGraph
2. Network evolution using temporal networks using Gephi
3. Community evaluation using word clouds using iGraph

We will collect #MatchAMillion and #BlackLivesMatter hashtags for 24 hours time windows between 2020-06-07-2020-06-08. From the collected tweet data, we will extract retweeting connectivity in the form of three columns: source userID; target userID; time point (at every 15 minutes interval). We then create the dynamic (temporal) network using Gephi software. User's geographic location, if available, will be mapped onto the nodes and evolution of modularity (mapped onto the edges) will be observed over 24 hours. [8-10]

### IV. Report of results

1. Collating a report with the observations from above.
2. Novelties in the project included:
  - a. Identify key players by measuring influence:
    - a) Within the fandom
    - b) Outside the fandom
  - b. Identifying ARMY accounts by measuring homophily
  - c. Sentiment analysis of key ARMY accounts: During this 24 hour period, the fandom went through a lot of collective emotions, and we want to analyse those.



As shown in figures above, “Influence” and “Homophily” are main driving social forces connecting individuals. By measuring influence and homophily from our constructed tweeter based

network, we would expect to observe similarities between ARMY fans, also observe mutual influences between ARMY fans to BlackLivesMatter supporters for social goodness.

## References

[1] BTS ARMY Matched The Group's \$1 Million Black Lives Matter Donation, Proving The Positive Power Of Fandoms

<https://www.forbes.com/sites/bryanrolli/2020/06/08/bts-army-black-lives-matter-1-million-donation/#4cdc05f56465>

[2] BTS' ARMY are flipping conversations about stan culture, one good deed at a time

<https://thefortyfive.com/opinion/bts-army-charity-work/>

[3] How the South Korean band's fanbase – known as ARMY – raised over \$1 million for the Black Lives Matter movement, mostly in just one day.

<https://graphics.reuters.com/GLOBAL-RACE/BTS-FANS/nmopajgmxxva/>

[4] Get started with Twitter APIs and tools Apply for access

<https://developer.twitter.com/en/apply-for-access>

<https://medium.com/swlh/extracting-tweets-using-twitter-premium-search-api-and-python-2d025144e8a4>

[5] Embedding-as-Service : One-Stop Solution to encode sentence to fixed length vectors from various embedding techniques

<https://pypi.org/project/embedding-as-service/>

[6] Universal-Sentence-Encoder: Collection of universal sentence encoders trained on a variety of data.

<https://tfhub.dev/google/collections/universal-sentence-encoder/1>

[7] Zafarani, R., Abbasi, M., & Liu, H. (2014). Social Media Mining: An Introduction. Cambridge: Cambridge University Press. doi:10.1017/CBO9781139088510

[8] Mapping users (nodes) onto geographic location

<https://gephi.org/plugins/#/plugin/geolayout-plugin>

[9] Creating a simple dynamic network

<https://seinecle.github.io/gephi-tutorials/generated-html/creating-a-simple-dynamic-network.html>

[10] Converting a network with dates into a dynamic network

<https://seinecle.github.io/gephi-tutorials/generated-html/converting-a-network-with-dates-into-dynamic.html>