Integrated Twitter Analysis to Distinguish Systems Thinkers at Various Levels: A Case Study of COVID-19

harun pirim
industrial and systems engineering
mississippi state university

Introduction

- The current literature review identifies research gaps relevant to evaluating if system-thinking skills of experts can be correlated with Twitter analysis.
- Microscale features emphasize certain node or edge characteristics such as the most critical person or relationship,

- Macro-scale features reveal information about the network itself, such as the density of connections in the network.

- Mesoscale features existing in between provide distinct information such as communities inside the network.

social network analysis

- Some examples of social networks include social media networks, business networks, and information circulation networks.

- Malik et al. evaluated the human advice-seeking behavior of primary health care (PHC) physician and finding ways to strengthen and enhance this system to better meet the provider needs using social network analysis.

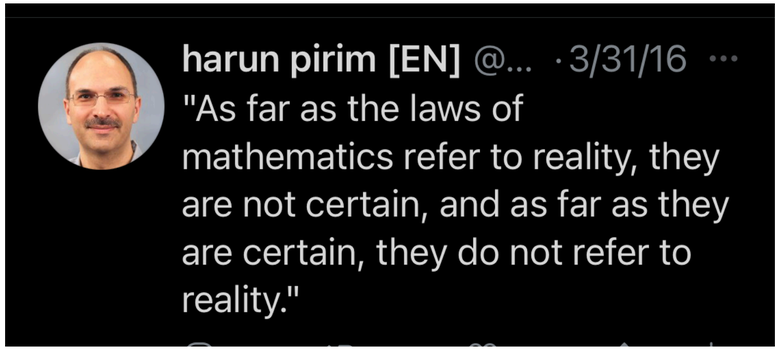
- Priven and Sacks adopted the social network analysis is their study to measure and analyze the communication level between subcontractors.

twitter analysis

- Liu et al. investigated diabetes-related individuals on Twitter by depicting the frequency and timing of diabetes-related tweets, the geography of tweets, and the type of members over a 2-year test of 10% of all tweets.

- William et al. explore the content and structure of online word-of-mouth (eWOM) and its effect at a tourism destination when a festival is staged using the leveraging of a combination of social network analysis and text analysis. The results showed that people generally and Twitter users precisely trust seemingly disinterested opinions from other individuals outside their actual social network.

twitter analysis



- Twitter, standing as one of the most widely used social media, provides various data such as tweet and retweet statistics, demographics, relations to followers, hashtags for particular topics.

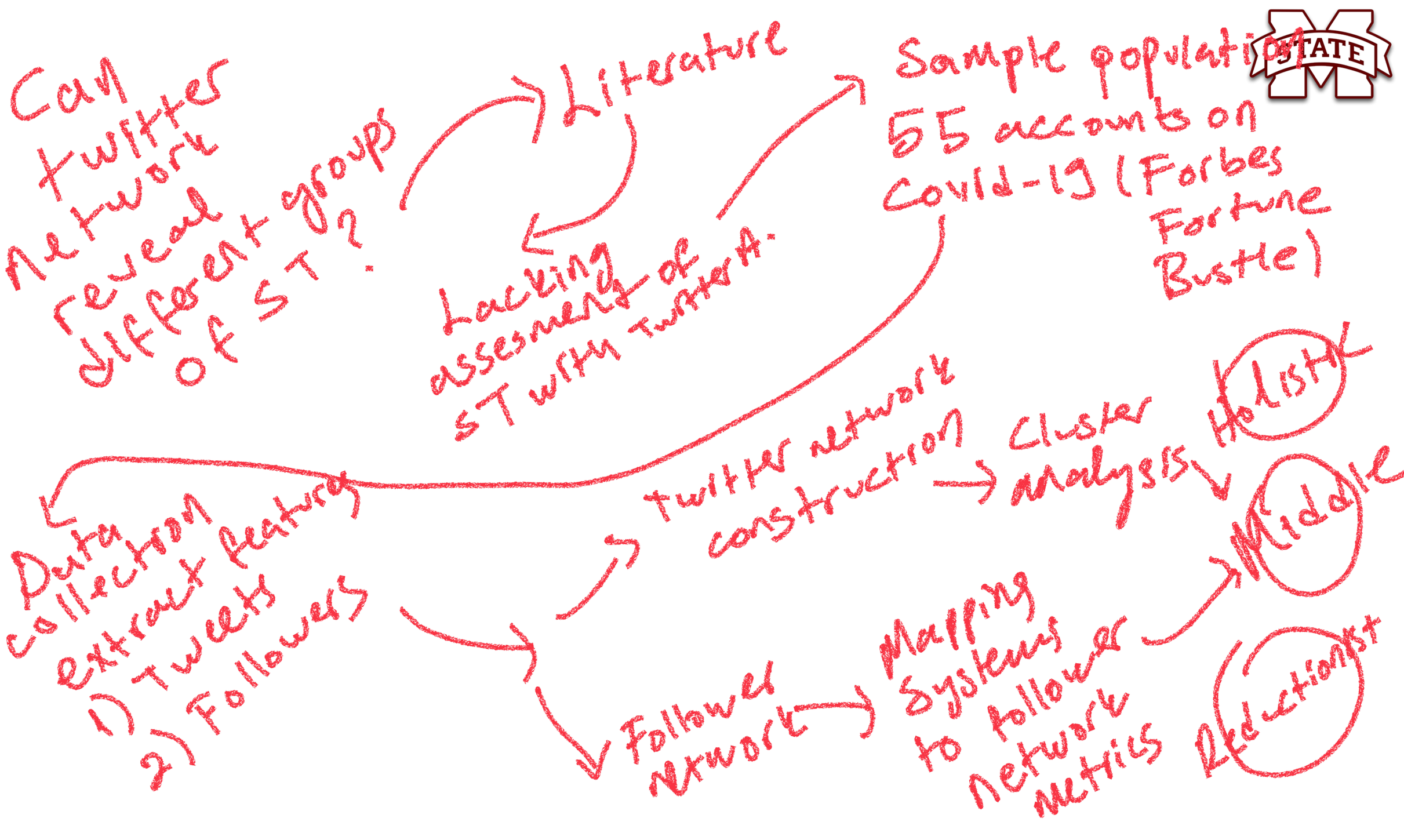
methodology and data collection

a team developer access to twitter through Application Programming Interface (API). After a few rounds of review, Twitter grants us team access to collect data. Twitter, by default, has some restrictions for tweets. The first restriction is it gives a maximum of 5000 tweets for each individual account. The second restriction is that we can access the information of 15 followers of each account every 15 minutes, which makes the follower analysis time consuming and difficult.

Table A1. The search thread for systematic literature search

Search criteria	**Search thread**
Search words	(twitter analysis) AND ((systems thinking)) AND (tweet)
Exact keywords (limited to)	"Social Media" OR "Social Networking (online)" OR "Twitter" OR "Human" OR "Decision Making" OR "Behavioral Research" OR "Social Network" OR "Sentiment Analysis" OR "Social Networks" OR "Social Network Analysis."
Document type (limited to)	Journal article
Language (limited to)	English
Subject area (limited to)	Engineering

The worldwide outbreak of the COVID-19 pandemic, a vital and complex problem, emerged as a potential use case to investigate the relationship between systems thinkers' twitter responses to complex problems and their systems thinking capabilities.


Can twitter network reveal different groups of ST?
Literature
Lacking assessment of ST with twitter.
Sample population
55 accounts on Covid-19 (Forbes Fortune Bustle)
Data collection
extract features
1) Tweets
2) Followers
Twitter network construction
Cluster analysis
Holistic
Middle
Reductionist
Follower network
Mapping systems to follower network metrics

“rtweet” library have been used to collect the data from Twitter. “create_token” function using four API credentials from Twitter, including

consumer key, consumer secret key, access token, and access secret, have been utilized, followed by the usage of “get_timeline” function to extract tweets of 55 identified Twitter accounts.

First, the potentially important information from each twitter account was extracted, such as the name of the account, id, screen name, location, description, follower counts, friends count, listed count, favorites count, status counts, and some other information.

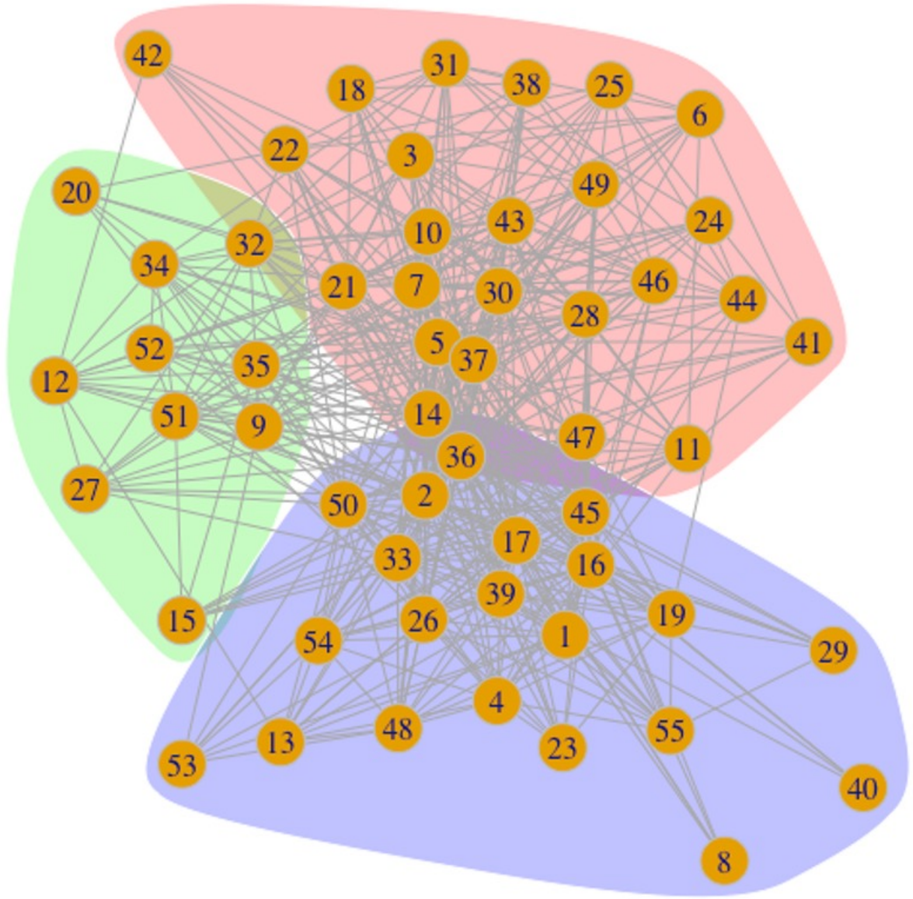
Then, all the tweets for each account categorized in the following three classes, starting by the1) organic tweets (excluding replies and retweets), 2) replies, and 3) retweets.

Main categories of Twitter features	Selected Twitter features
Twitter account metrics	followers_count
	friends_count
	listed_count
	favoritess_count
	statuses_count
	# of all tweets
	# of Organic tweets
	# of Retweets
	# of Replies
	% Organic tweets
	% Retweets
	% Replies


All tweets
(including retweets and replies)
favorite count
Ave top 10
SD top 10
Median
Mean
Max
retweet count
Ave top 10
SD top 10
Median
Mean
Max
quoted_favorite_count
Median
Mean
Max
quoted_retweet_count
Median
Mean
quoted_followers_count
Median
Mean
Max
quoted_friends_count
Median
Mean
Max
quoted_statuses_count
Median
Mean
Max
retweet_favorite_count
Median
Mean
Max
retweet_retweet_count
Median
Mean
Max
retweet_followers_count
Median
Mean
Max
retweet_friends_count
Median
Mean
Max
retweet_statuses_count
Median
Mean
Max

network construction

- The twitter network is constructed, making use of 104 features extracted from 55 twitter accounts.
- Normalized Gower distance is used to quantify pairwise relationships among 55 accounts. Distance values are converted to similarity matrix to construct the weighted network.
- The weighted network is binarized retaining relationships stronger than the average. The community structure finding algorithm by Blondel et al. [24] is applied to find similar groups of accounts in the network. The algorithm detected three groups

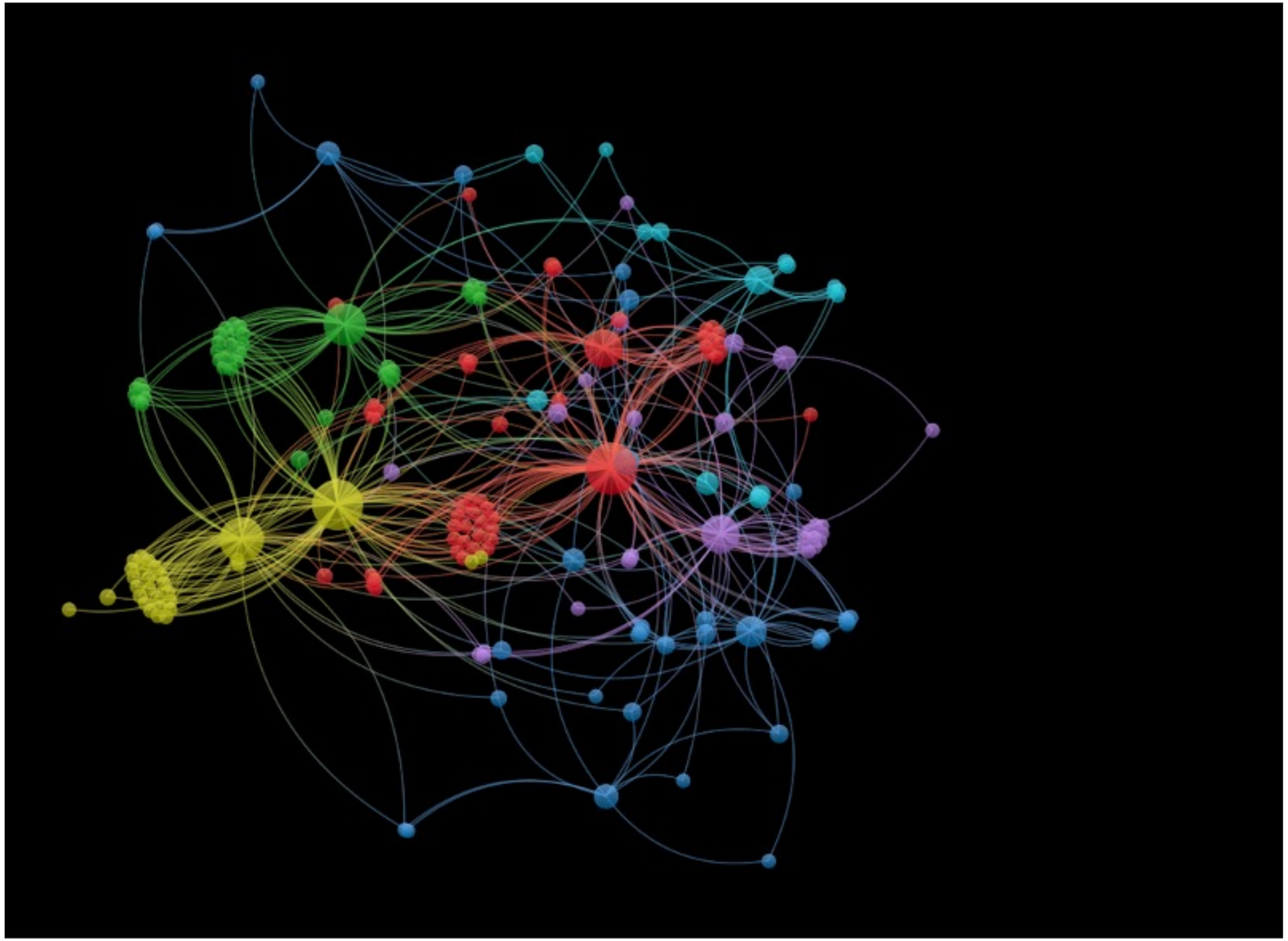

42
31
18
38
25
6
22
3
49
20
43
24
10
32
34
46
21
7
30
44
28
41
5
52
37
35
12
14
51
9
47
11
36
27
50
2
45
17
33
16
39
15
26
19
54
1
29
4
48
55
13
23
53
40
8

follower network

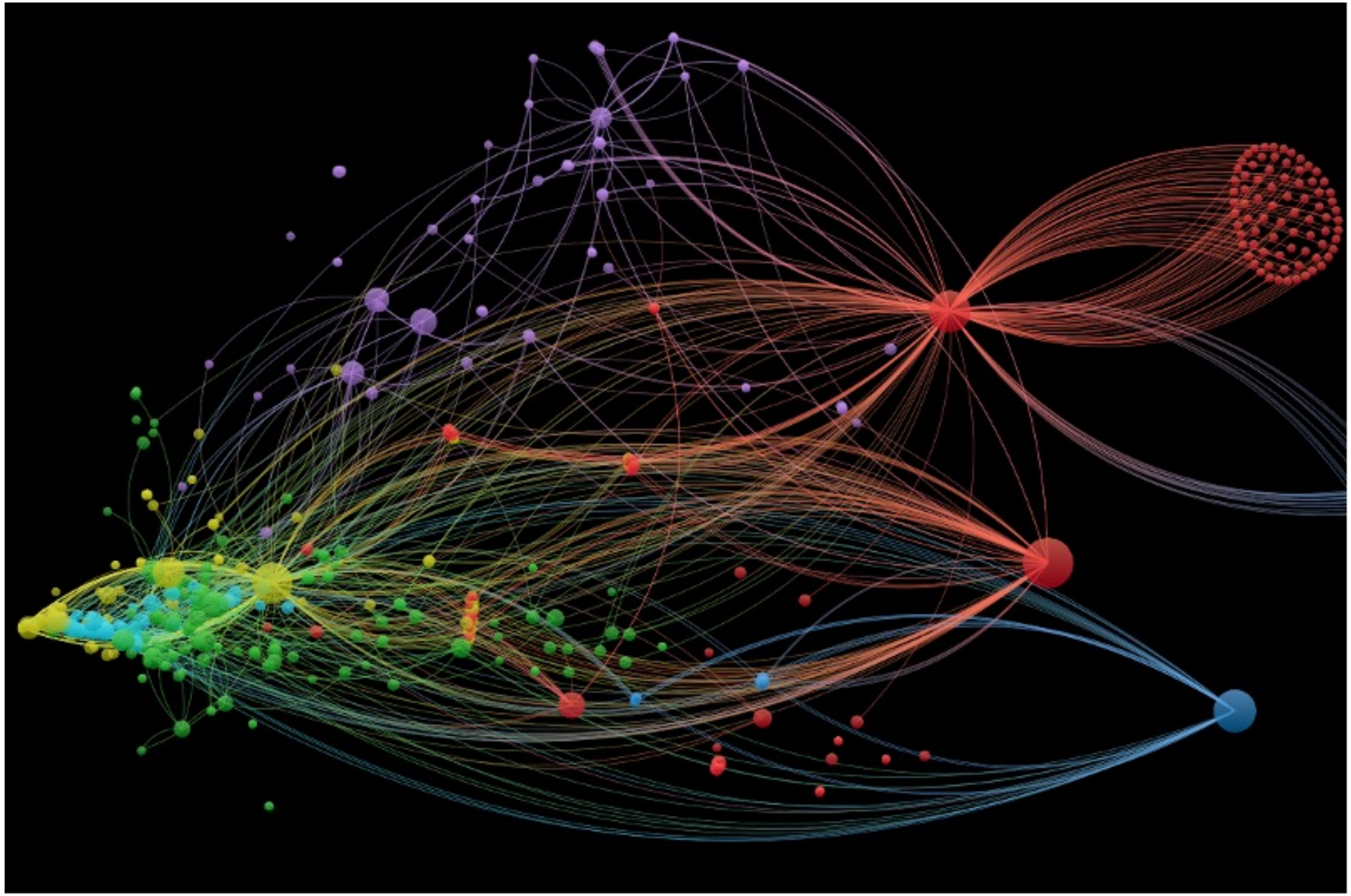
Due to the main restriction of Twitter regarding followers' data (accessing 15 followers of a Twitter account every 15 minutes), only a sample of 45 followers is selected to perform the follower network analysis.

After collecting the followers' data for each Twitter account, we extracted some information regarding the number of followers, following, and friends of each account.

we go only as deep as two layers of followers. For example, the University of Washington Virology's followers, as well as followers of those followers,



UW virology follower network



WHO

Four centrality measures, namely, Betweenness centrality, Closeness centrality, Degree centrality, and Eigen Centrality, were chosen for followers' network analysis, and the summary statistics for each measure (e.g., mean centrality, SD centrality, median centrality, min centrality, and max centrality) were generated as network features.

The node-level metrics, such as "this account follows how many accounts," "number of this account's friends," "number of nodes in the network of this account," and "number of edges in the network of this account," were added to the features of followers' network.

systems thinking mapping

Table 1: PCA results of the centrality measures.

Importance of components	Betweenness centrality	Closeness centrality	Degree centrality		Eigen Centrality
Number of composite variables emerged	1	1	2		1
Standard deviation (greater than 1 is preferred)	1.96	2.00	1.63	1.12	1.90
Proportion of variance explained (more than 60% is preferred)	77%	80%	53%	25%	90%

five composite centrality variables, as well as four node-level metrics, applied to interpret the followers' characteristic of each of 55 twitter accounts.

The Eigen Centrality calculates the impact of a node (here, a Twitter account) on the connection with other nodes (e.g., followers Twitter accounts) of the network system

The Eigen Centrality has a hypothetical positive correlation with all seven ST skills dimensions' definitions.

For example, the Twitter account that scores high Eigen Centrality is well-connected with its followers and has well-linked followers; the individual (who owns the Twitter account) might have a high level of Interaction and Autonomy skills based on the operational definitions

Dimension	Less Systemic (reductionist)	More Systemic (holistic)
Level of Complexity: Comfort with multidimensional problems and limited system understanding	**Simplicity (S)**: Avoid uncertainty, work on linear problems, prefer the best solution, and prefer small-scale problems.	**Complexity (C)**: Expect uncertainty, work on multidimensional problems, prefer a working solution, and explore the surrounding environment.
Level of Autonomy: Balance between local-level autonomy versus system integration	**Autonomy (A)**: Preserve local autonomy, tend more to an independent decision and local performance level.	**Integration (G)**: Preserve global integration, tend more to a dependent decision and global performance.
Level of Interaction: Interconnectedness in coordination and communication among multiple systems	**Isolation (N)**: Inclined to local interaction, follow a detailed plan, prefer to work individually, enjoy working in small systems, and interested more in cause-effect solution.	**Interconnectivity (I)**: Inclined to global interactions, follow a general plan, work within a team, and interested less in identifiable cause-effect relationships
Level of Change: Comfort with rapidly shifting systems and situations	**Resistance to Change (V)**: Prefer taking few perspectives into consideration, over-specify requirements, focus more on internal forces, like short-range plans, tend to settle things, and work best in a stable environment.	**Tolerant of Change (Y)**: Prefer taking multiple perspectives into consideration, underspecify requirements, focus more on external forces, like long-range plans, keep options open, and work best in a changing environment.
Level of Uncertainty: Acceptance of unpredictable situations with limited control	**Stability (T)**: Prepare detailed plans beforehand, focus on the details, uncomfortable with uncertainty, believe the work environment is under control, and enjoy objectivity and technical problems.	**Emergence (E)**: React to situations as they occur, focus on the whole, comfortable with uncertainty, believe the work environment is difficult to control, enjoy subjectivity and non-technical problems.
Level of Hierarchical View: Understanding system behavior at the whole versus part level	**Reductionism (R)**: Focus on particulars, prefer analyzing the parts for better performance.	**Holism (H)**: Focus on the whole, interested more in the big picture, interested in concepts and abstract meaning of ideas.
Level of Flexibility: Accommodation of change or modifications in systems or approach	**Rigidity (D)**: Prefer not to change, like a determined plan, open to new ideas, motivated by routine.	**Flexibility (F)**: Accommodating to change, like a flexible plan, open to new ideas, and unmotivated by routine.

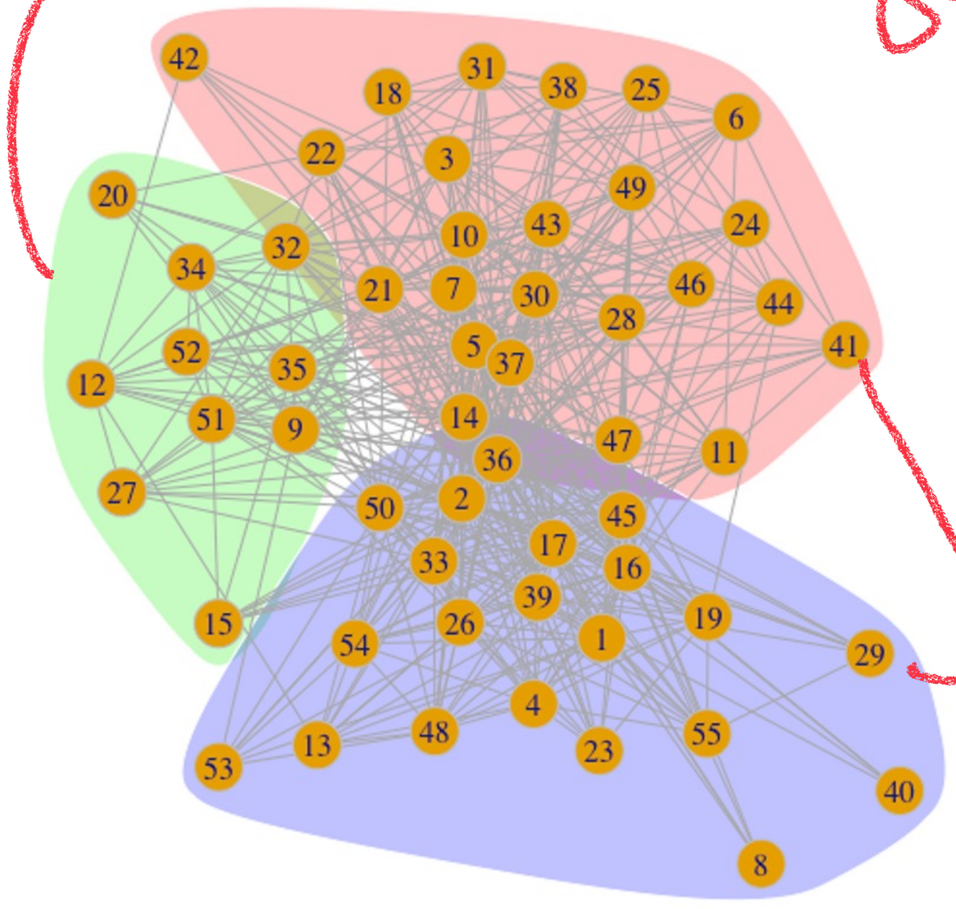
	Statistics	Betweenness centrality	Closeness centrality	Degree centrality 1	Degree centrality 2	Eigen Centrality	# of following accounts	# of friends' accounts	# of network nodes	# of network edges
Cluster 1 (n=24) Orange Circles	*M*	59.08%	7.36%	64.32 %	8.24%	14.79 %	1472.7	162.8	855.4	2110.0
	SD	30.09%	8.18%	9.98%	4.12%	7.94%	1211.5	453.2	529.8	1348.1
Cluster 2 (n=10) Green	*M*	80.30%	14.73 %	72.72 %	8.00%	21.94 %	696.3	3.2	459.1	1110.8
	SD	15.01%	11.82 %	7.55%	7.48%	14.92 %	577.6	3.4	311.0	764.5
Cluster 3 (n=21) Purple	*M*	81.76%	25.30 %	72.10 %	16.09 %	32.99 %	1761.7	9.1	284.5	691.0
	SD	16.64%	29.91 %	20.96 %	22.27 %	27.82 %	1673.2	17.9	211.6	527.0

High scores in centrality measures of follower analysis associated with high ST capability of individuals.

Twitter accounts in cluster 3 have relatively higher centrality measures, holistic thinker cluster.

Twitter accounts in cluster 1 called reductionist thinker Twitter accounts due to low scores pertaining to centrality measures of their follower network analysis.

This study's results shows the follower network analysis might be able to cluster systemic thinking capabilities of individuals based on their tweets in the case of the COVID-19 pandemic.



middle system thinkers

enriched by holistic systems thinkers

reductionist system thinkers

conclusion

- Using follower analysis and Twitter analysis to understand systems thinking of people, especially the influencers and celebrities, and how their role is important to spread the true news and knowledge to the community.

- How systems thinking is related to have a better network of followers, which promotes the more efficient and effective transformation of information and knowledge to the community.

- As the level of systems thinking skills of individuals can be enhanced, the social media activity of the individuals can be improved. This is very important since the research shows there is a necessity to create a safe and healthy virtual environment for everybody, so everyone can express their opinions and beliefs in the direction of the community's values.

references

- Asmat U Malik, Cameron D Wills, Saima Hamid, Anar Ulikpan, and Peter S Hill. 2014. Advancing the application of systems thinking in health: advice seeking behavior among primary health care physicians in Pakistan. Health Research Policy and Systems, 12(1), 43.

- Vitaliy Priven and Rafael Sacks. 2015. Effects of the Last Planner System on Social Networking among construction trade crews. Journal of Construction Engineering and Management, 141(6), 04015006.

- Yang Liu, Qiaozhu Mei, David A Hanauer, Kai Zheng, and Joyce M Lee. 2016. Use of social media in the diabetes community: An exploratory analysis of diabetes-related tweets. JMIR diabetes, 1(2), e4.

- Nigel L. Williams, Alessandro Inversini, Dimitrios Buhalis, and Nivole Ferdinand. 2015. Community crosstalk: an exploratory analysis of destination and festival eWOM on Twitter. Journal of Marketing Management, 31(9-10), 1113-1140

work in progress

prediction of hypothetical proteins using deep learning and network science (gene sequence data)

BRAIN FUNCTIONAL CONNECTIVITY PATTERN RECOGNITION FOR ATTENTION-DEFICIT/HYPERACTIVITY DISORDER DIAGNOSIS (brain correlation data)