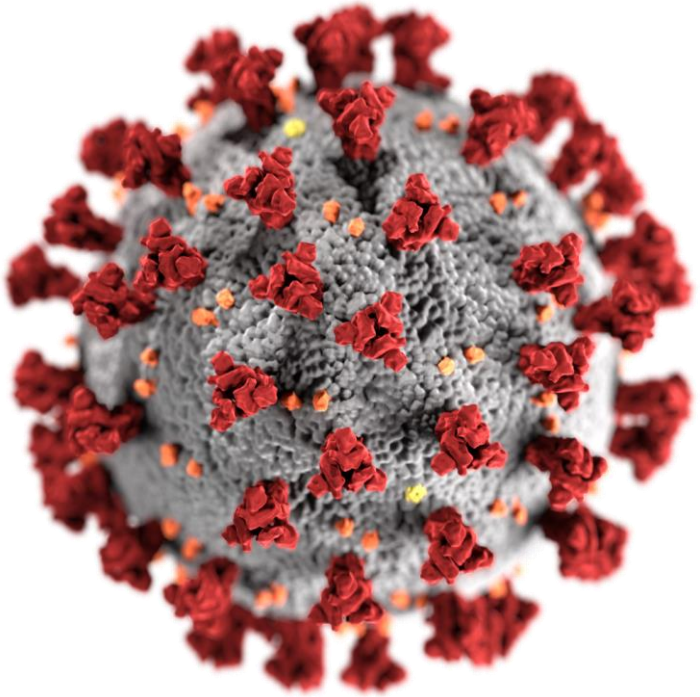
Social Networks: The Tracking of COVID-19 Misinformation Using NLP and Graph-based Approaches

S. Toyonaga
A. Kartha

Motivation

- The Covid-19 pandemic has brought about millions of death globally
- Misinformation plays a big role through **vaccine hesitancy** and undermining of the virus's impact
- Social media has been used to spread **misinformation** at unprecedented rates, primarily through bots and influencers



YORK U

Sub Questions to guide research

1. What kinds of users participate in the dissemination of misinformation?

2. What Natural Language Processing (NLP) based patterns can be drawn from malicious users and their tweets/posts?

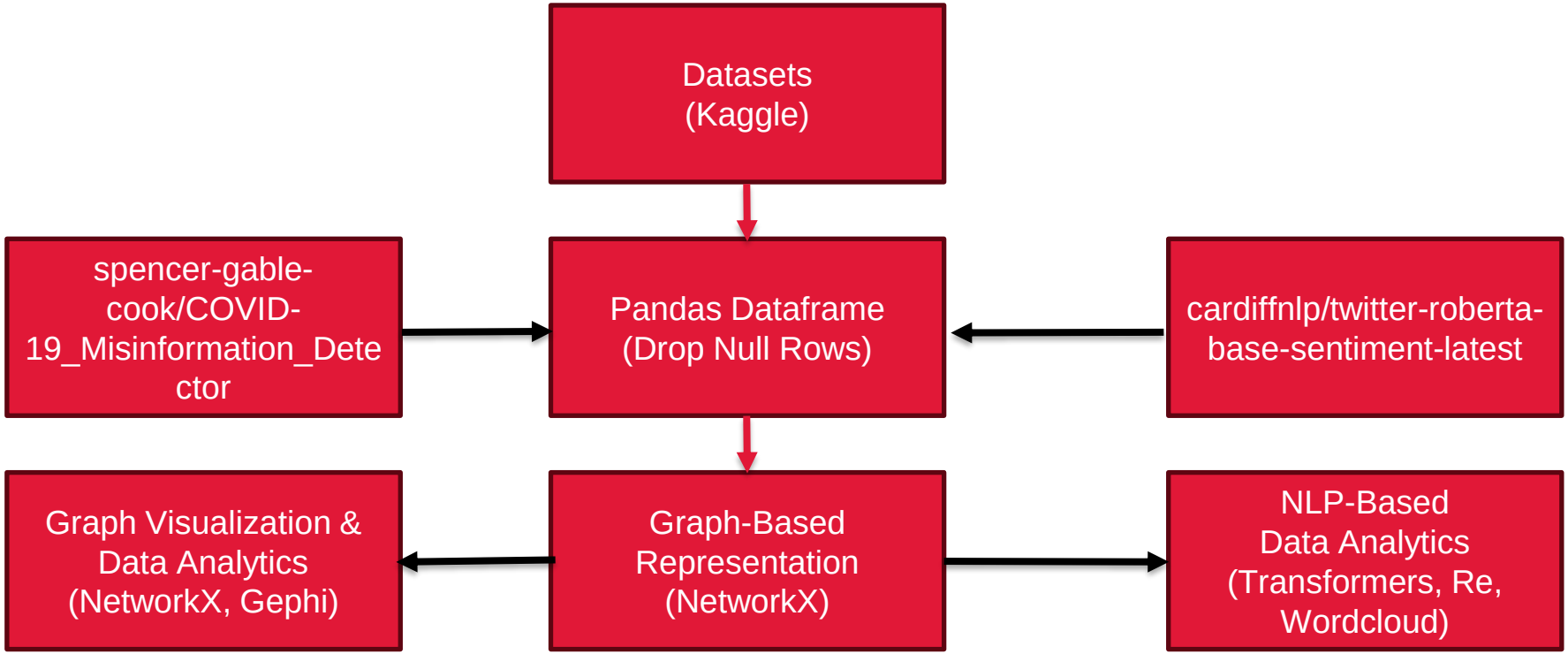
3. How is misinformation propagated in terms of a longitudinal time frame?

4. What are noticeable differences and similarities in how misinformation spreads between Reddit and Twitter?

Problem Definition

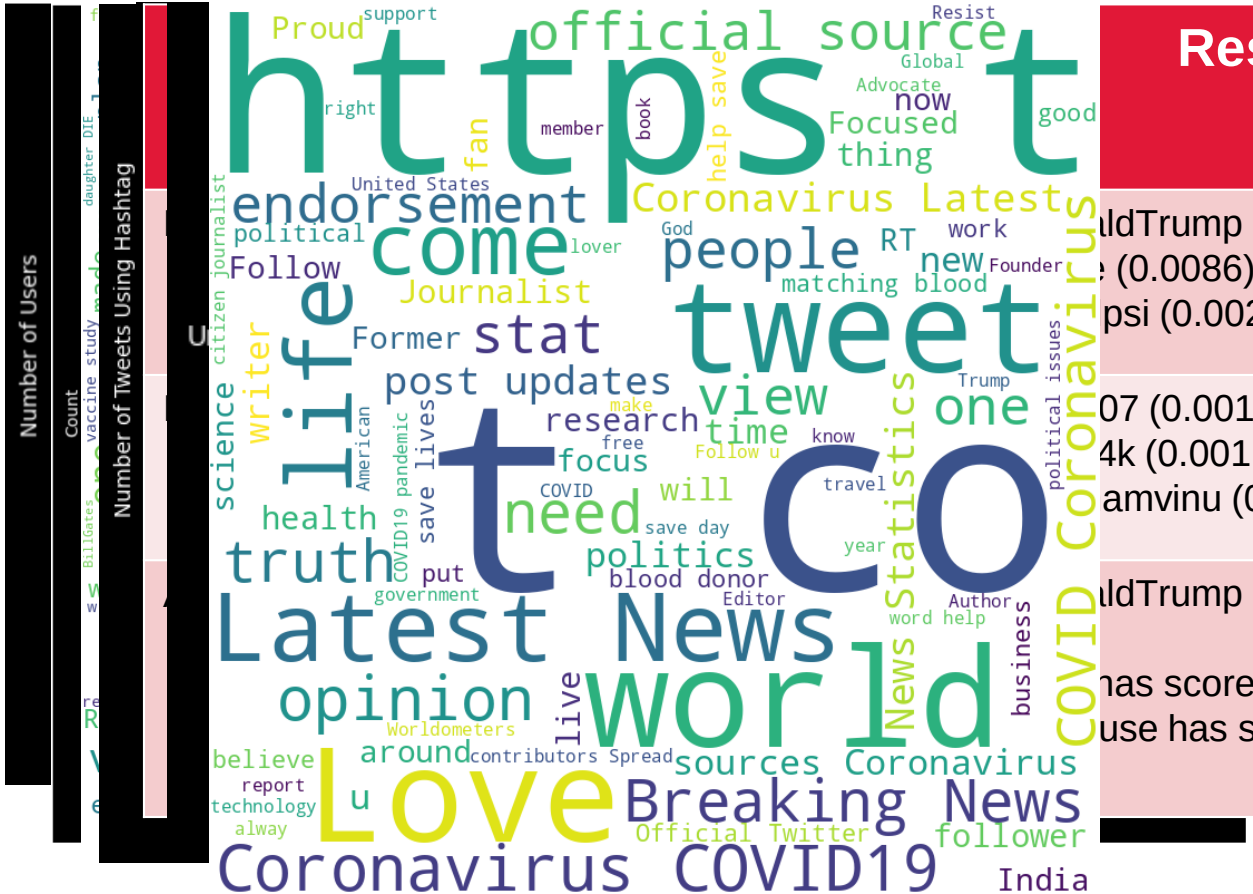
- Our aim is to create a **directed** social network graph **G = (V, E)** where:
    - **V** represents users who are participating in the spread of misinformation
    - **E** represents the relationships between users who share, view, or distribute misinformation

Methodology (Twitter)



YORK U

Results and Analysis (Twitter)



Results

ldTrump (0.0115)
(0.0086)
psi (0.0025)

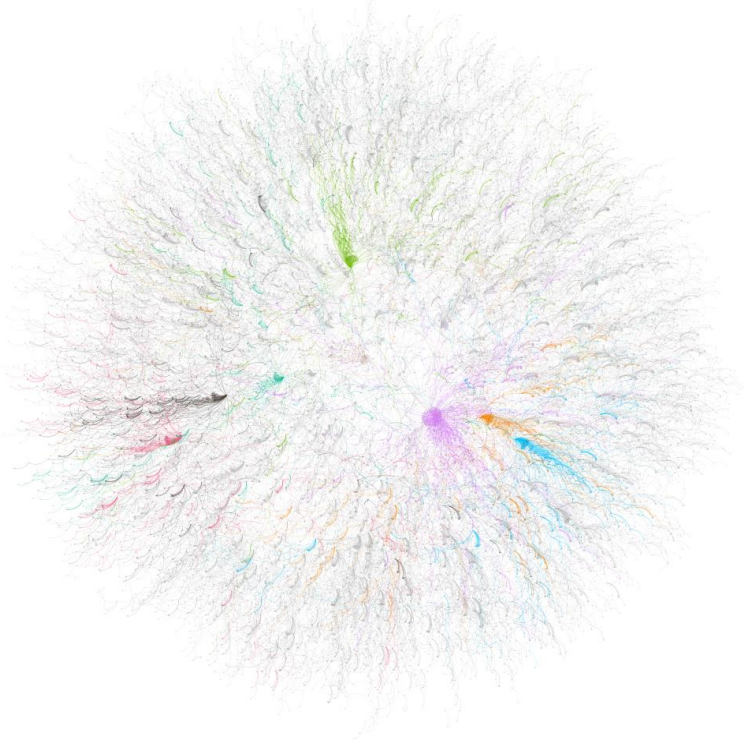
07 (0.0013)
4k (0.0011)
amvinu (0.001)

ldTrump has score contribution

has score contribution 0.0116
use has score contribution

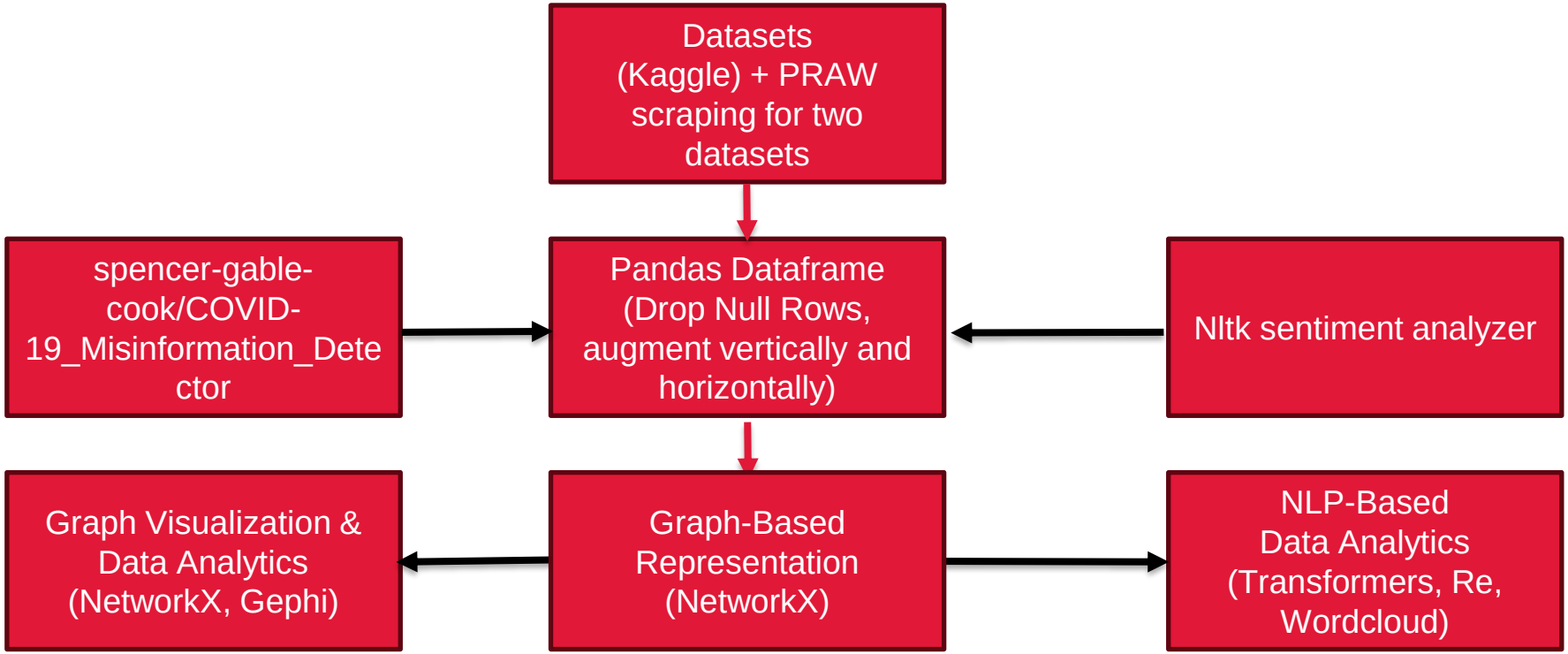
YORK U

Results and Analysis (Twitter) - 2



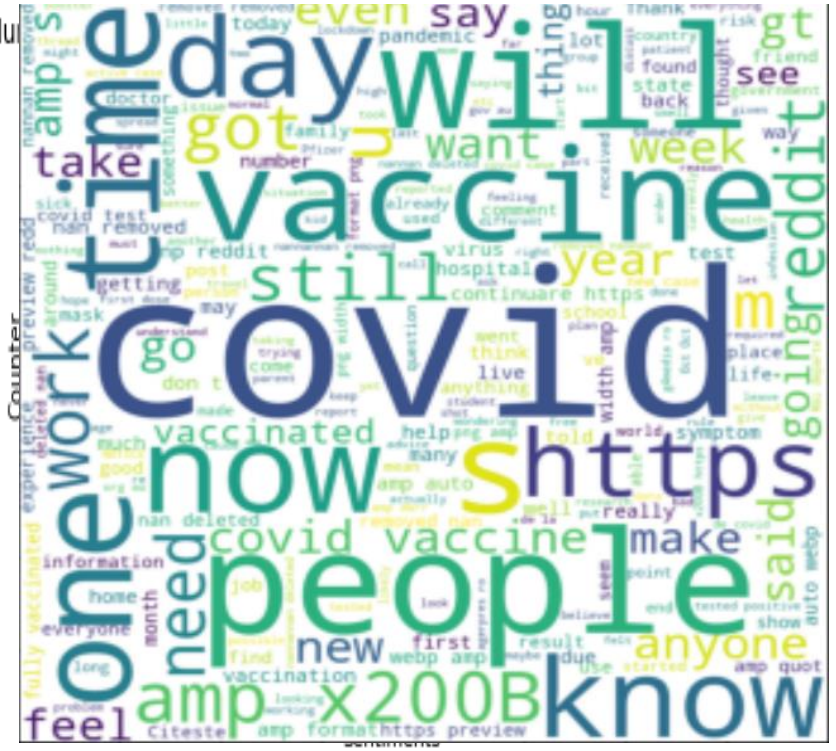
Property	Value
Nodes	35,147
Edges	35,146
Degree Distribution	2.99
Triangles	0
Clustering Coefficient	0
Modularity	0.980
Communities	112
Diameter	31

Methodology (Reddit)



YORK U

Results and Analysis (Reddit)



YORK U

QA LLM (In-Progress)

- The idea is to have a question answering system that can supplement our data visualization and analytics system
- Example of input – output combination based on results received so far
- Split into various difficulties

Simple QA:

- Q: Which user has the highest Twitter score for PageRank?
- A: @realDonaldTrump with 0.0115

Complex QA (Logic + reasoning) :

- Q: By what percent does the first and second place differ for Authorities score contribution for Twitter?
- A: There is an increase of ((0.2262-0.0116)/0.0116) x 100 = 1850 %

- ^ The following involves 'Chain-of-thought-prompting' which tackles tasks that involve arithmetic and logical reasoning. Showing HOW results are derived is important.

Conclusion and limitations

- Spread of misinformation is prevalent
- Politicians / media personalities dominate twitter
  - Expected with figures like Trump
- Differences between reddit and twitter:
  - Reddit seemed to have much more neutrality in posts and comments as compared to twitter
  - Twitter is more prone the spreading of misinformation
  - Reddit doesn't have as many clear relations between entities (No tags a factor)


That's all Folks!