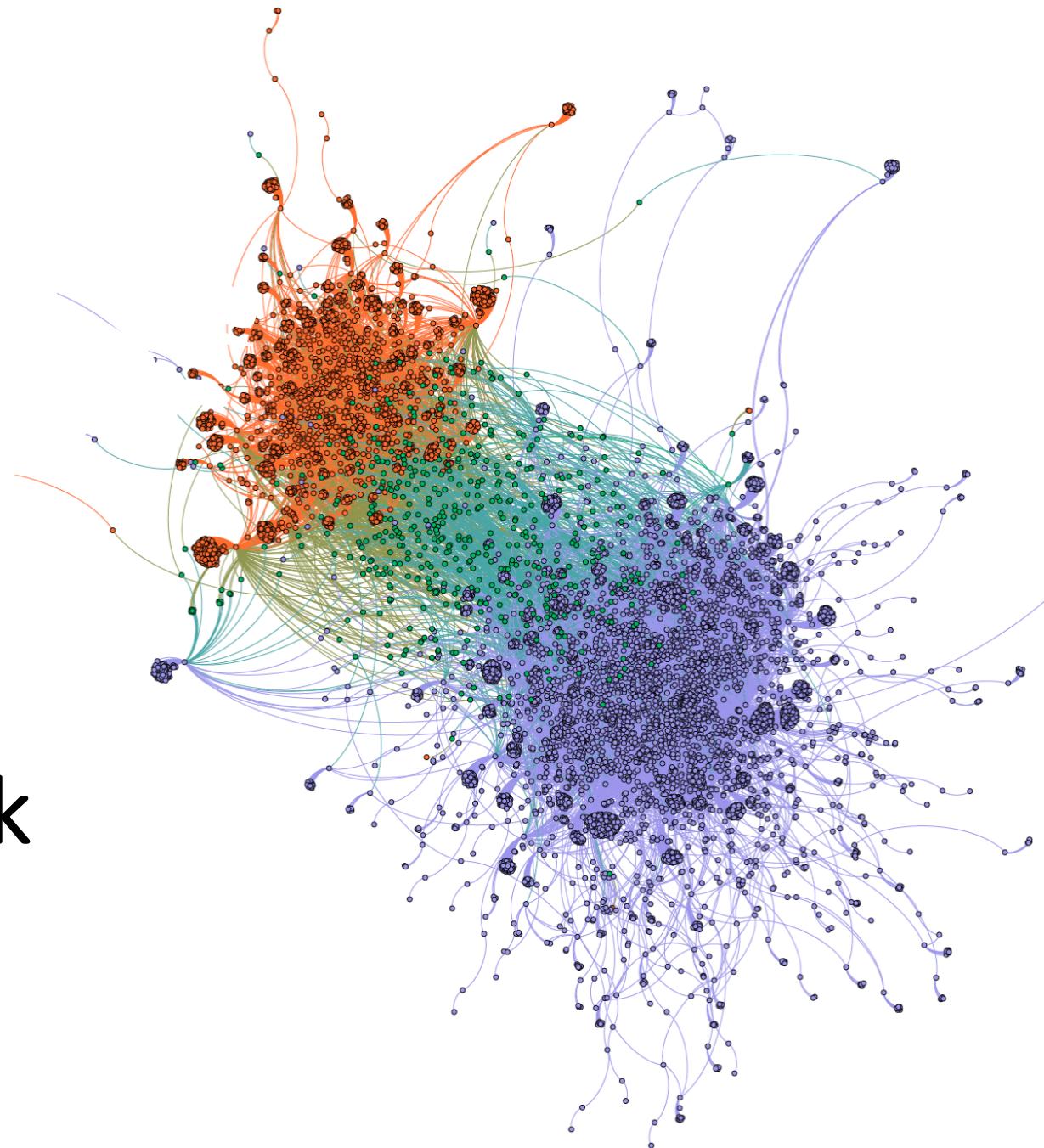


COVID-19 Information Spread around Political Engagement on Facebook

08.Mar.2021

POOREUMOE KIM

Tobias Eder



Introduction

- Facebook has more than 2 billion user accounts
- Little detailed study of how COVID-19 information is spread in the massive social network
- Primary diffusion metric: share count
- Data collection by CrowdTangle APIs

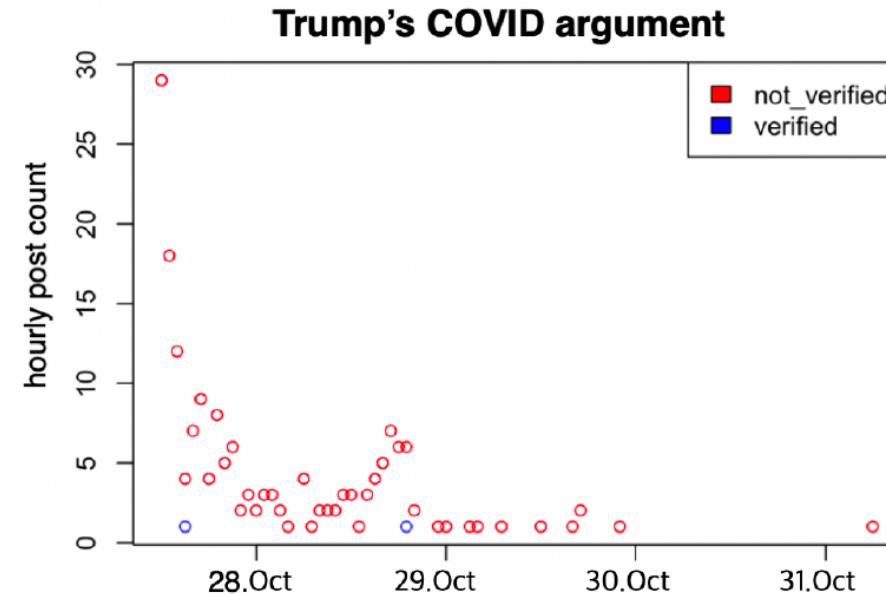
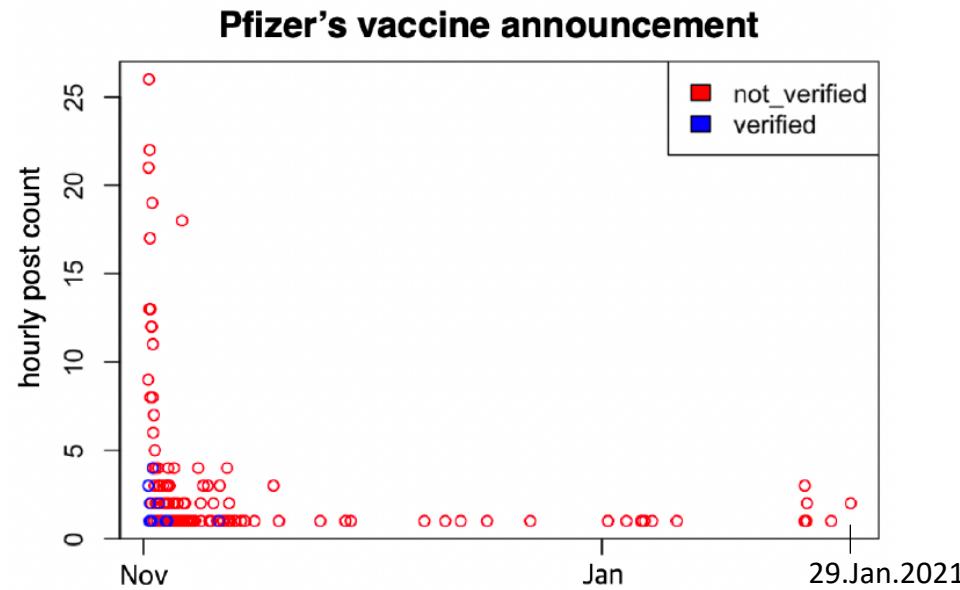
Research Questions

1. How does a general sharing pattern of COVID-19 posts appears?
2. How does COVID-19 information spread through linked posts? Especially from official sources to unofficial on Facebook?
3. What are the main COVID-19 topics and trends in 2020? How did users react to them?
4. Can we predict a post's long-term future share counts with its initial behavior? If then, which feature plays a crucial role?

Research Questions

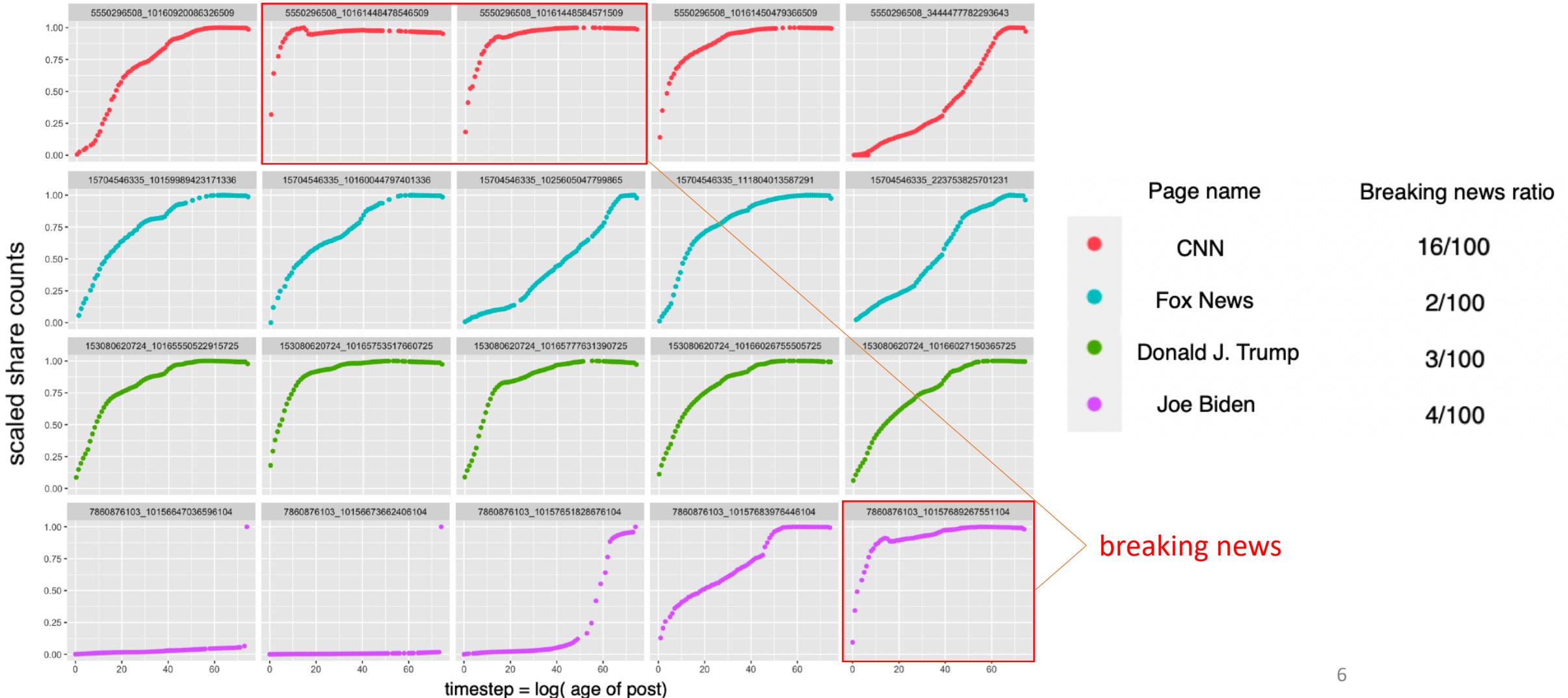
1. How does a general sharing pattern of COVID-19 posts appears?
✓ Sharing pattern exploration
2. How does COVID-19 information spread through linked posts? Especially from official sources to unofficial on Facebook?
✓ Network mapping
3. What are the main COVID-19 topics and trends in 2020? How did users react to them?
✓ Topic modeling
4. Can we predict a post's long-term future share counts with its initial behavior? If then, which feature plays a crucial role?
✓ Prediction modeling

Sharing Pattern Exploration



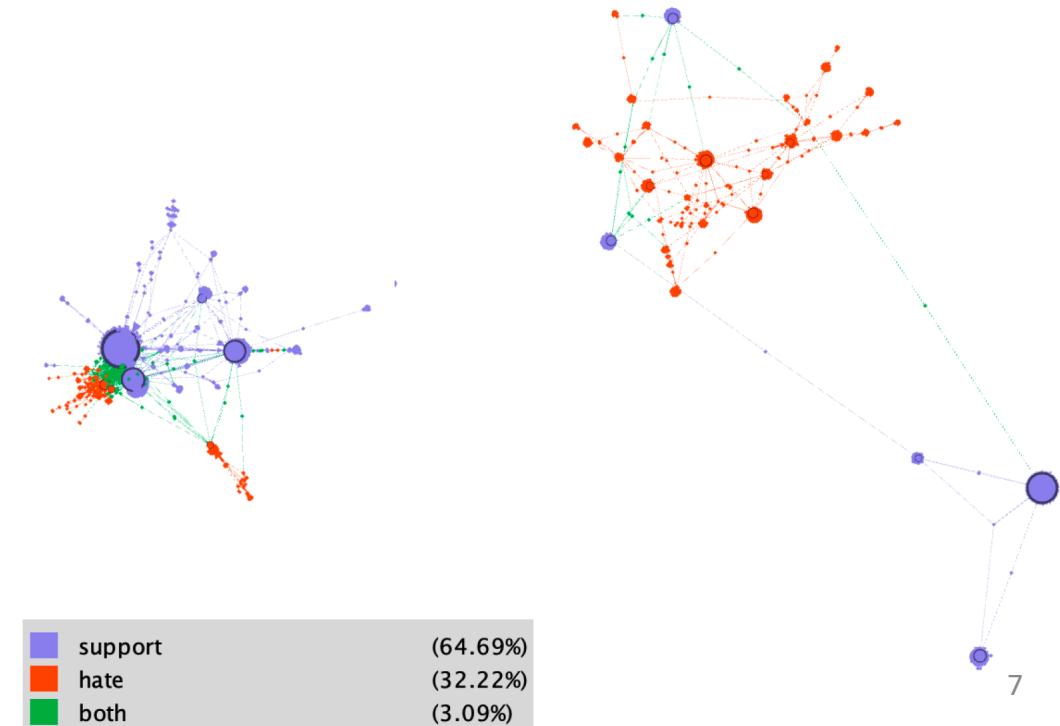
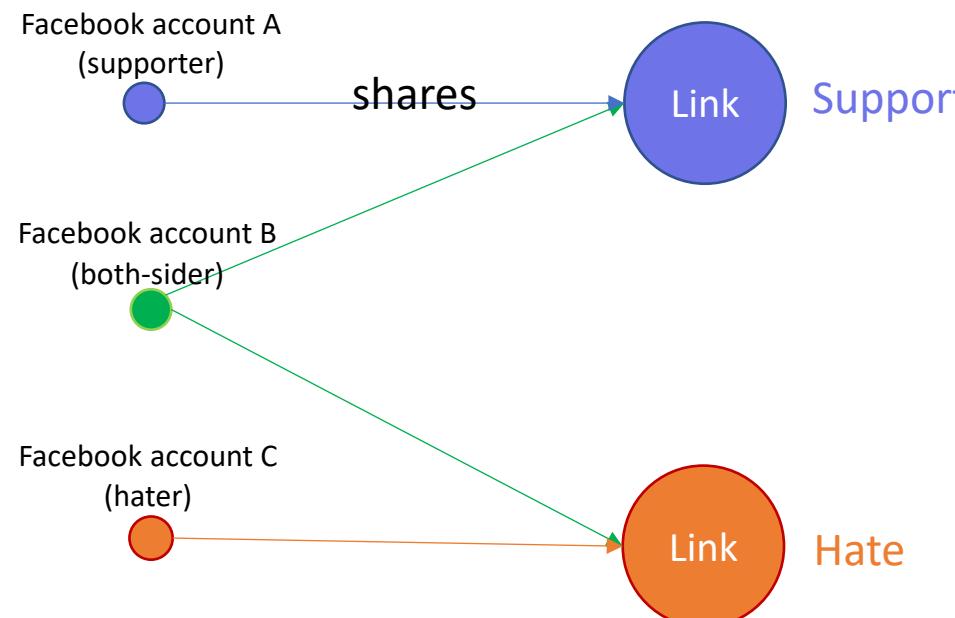
	Pfizer's announcement	Trump's argument
Verified accounts ratio	24/430	2/182
Sharing period	81 days	5 days

Sharing Pattern Exploration

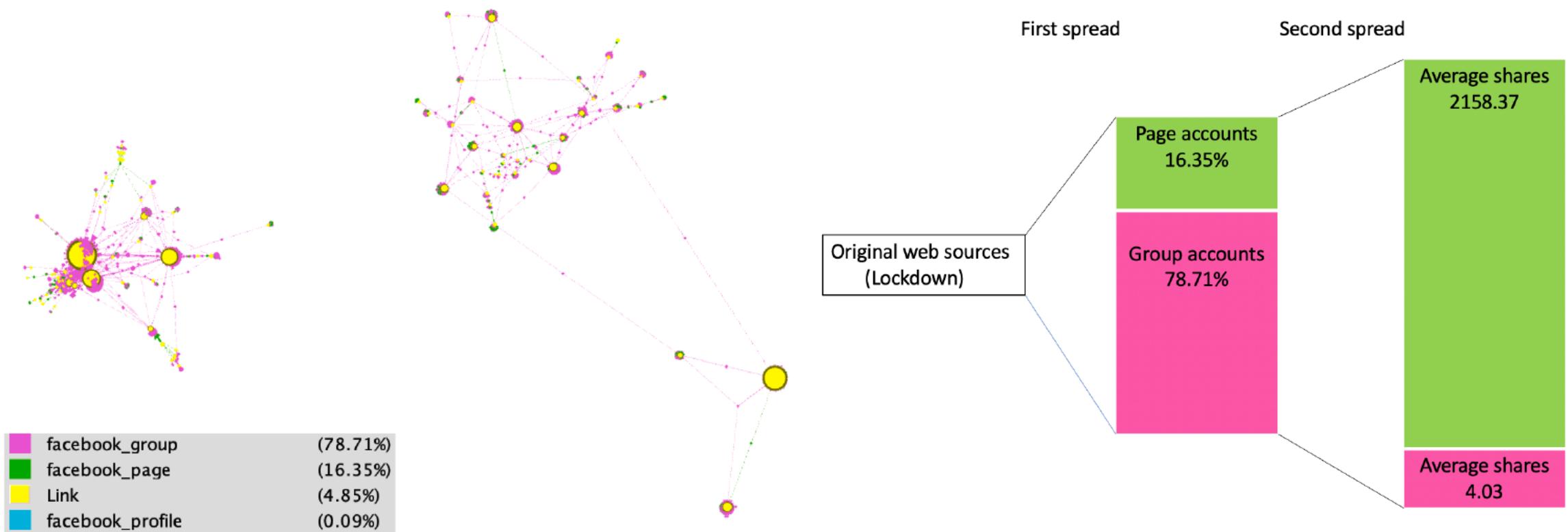


Network Mapping

- The 100 + 100 most interacted posts in 2020 by the queries:
 - lockdown AND (positive OR lucky OR play OR safe) -> lockdown supporters
 - lockdown AND (liberty OR unconstitutional) -> lockdown haters

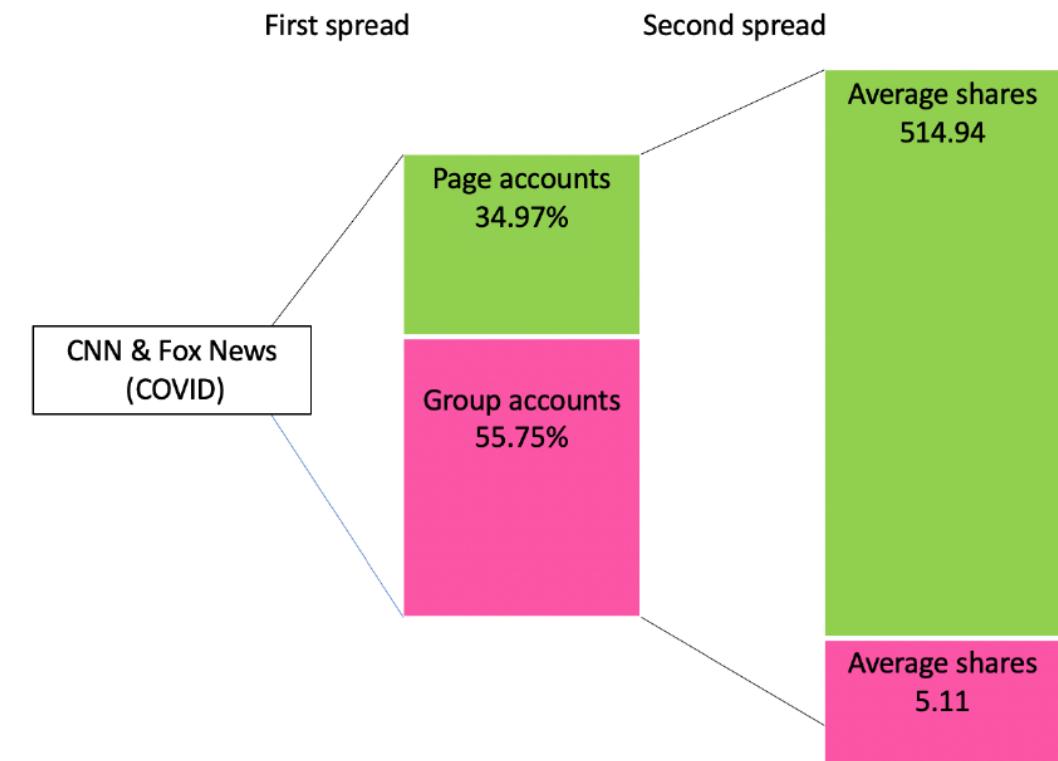
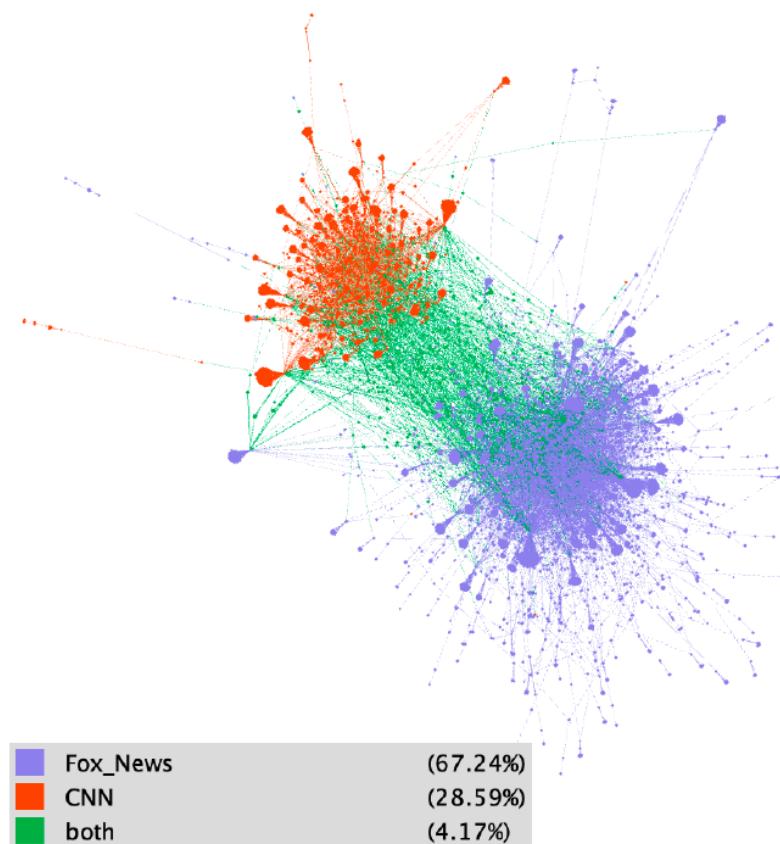


Network Mapping



Network Mapping

- The 100+100 most shared posts in 2020 from CNN and Fox News pages, with the query 'COVID.'



Network Mapping

- CNN's COVID-19 information spreads **1.58 times more** than Fox News' despite the number of sharing accounts in the network.
- Total shares of 1,000 COVID-19 posts from CNN and Fox News that were the most shared in 2020
 - CNN: 3,162,516
 - Fox News: 2,001,144

Topic Modeling

- Dataset: 3,000 most shared COVID-19 posts on CNN and Fox News' Facebook pages in 2020
- Model test:
Latent Dirichlet Allocation (LDA),
Cluwords,
Contextualized Topic Modeling(CTM)
- NPMI achieved the most considerable correlation to human topic coherence ratings

(G. Bouma, “Normalized (pointwise) mutual information in collocation extraction,” Proceedings of GSCL, pp. 31–40, 2009)

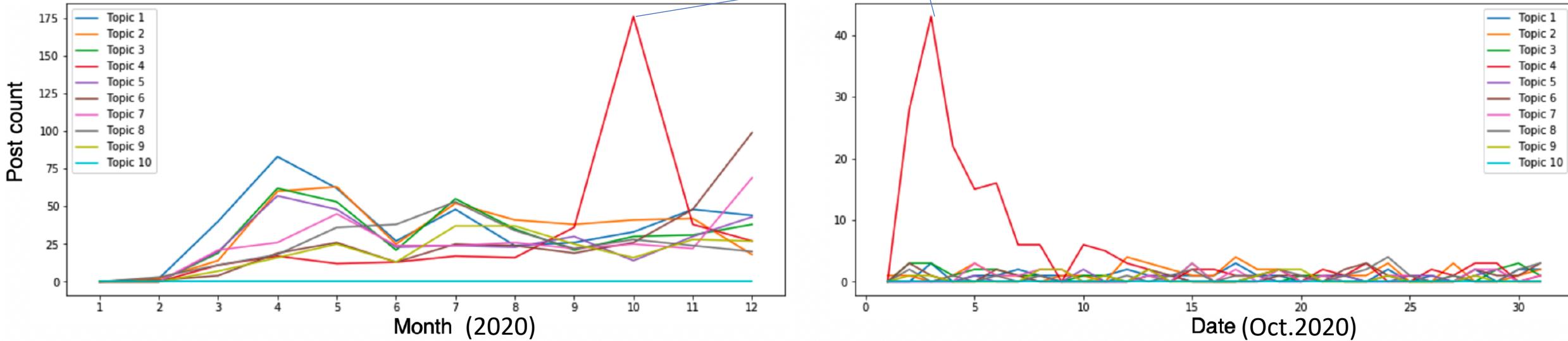
	LDA	Cluwords	CTM
<i>C_{UCI}</i>	-0.34	-10.83	-0.42
<i>C_{UMass}</i>	-2.85	-15.39	-3.58
<i>C_{NPMI}</i>	0.12	-0.38	0.21

Topic	Component words
1	provides, holds, andrew, looking, governor
2	study, likely, disease, blood, infection
3	died, test, police, year, three
4	trump, president, donald, biden, white
5	workers, kids, parents, back, care
6	vaccine, pfizer, vaccines, administration, biotech
7	relief, bill, senate, economic, package
8	social, health, content, source, medium
9	social, content, source, fbcnn, medium
10	reported, states, number, infections, cases

Ten topics made by CTM

Topic Modeling

Topic 4: trump, president, donald, biden, white



'JUST IN: Former White House adviser Kellyanne Conway says she **has tested positive for Covid-19**'

"Friday morning on "America's Newsroom," Chris Wallace weighed in **on President Trump's COVID-19 diagnosis** after moderating the first presidential debate on Tuesday. Former Vice President Joe Biden revealed Friday afternoon that he tested negative for COVID-19."

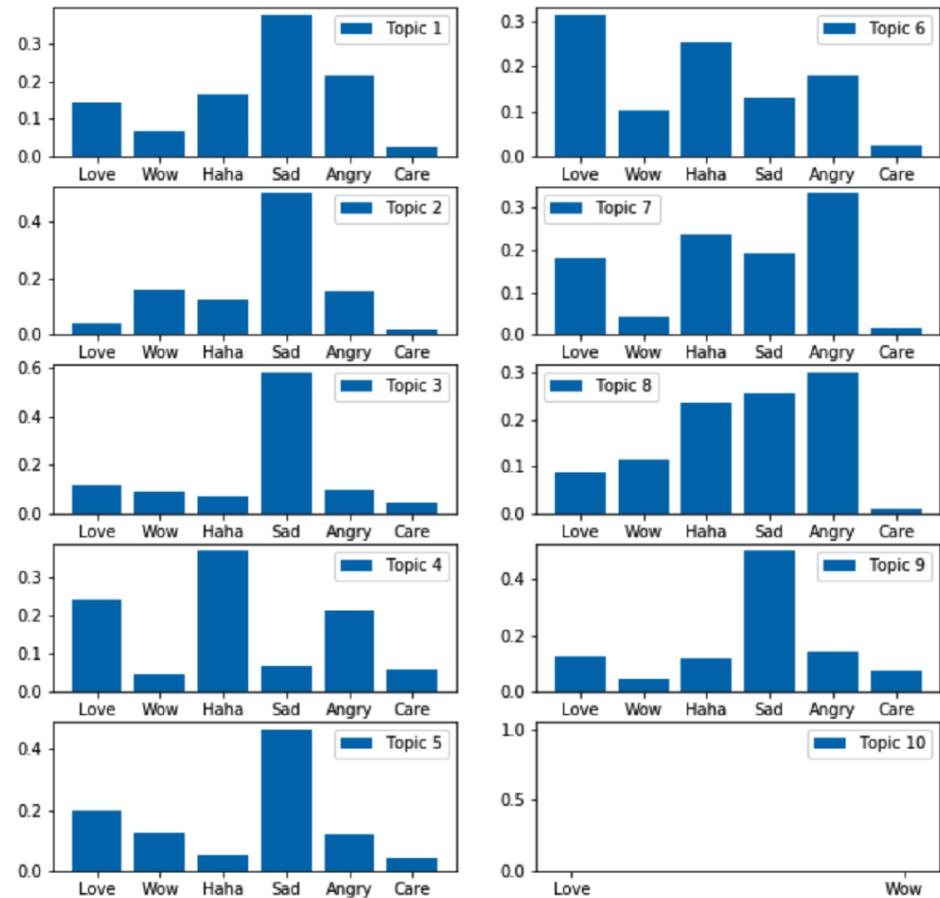
"Joe Biden's campaign is pulling all negative ads Friday following **President Trump's diagnosis with Covid-19**, a campaign aide says"

The three most shared posts on 3.Oct

(a) Ten topics from CNN and Fox News

Topic	Component words
1	provides, holds, andrew, looking, governor
2	study, likely, disease, blood, infection
3	died, test, police, year, three
4	trump, president, donald, biden, white
5	workers, kids, parents, back, care
6	vaccine, pfizer, vaccines, administration, biotech
7	relief, bill, senate, economic, package
8	social, health, content, source, medium
9	social, content, source, fbcnn, medium
10	reported, states, number, infections, cases

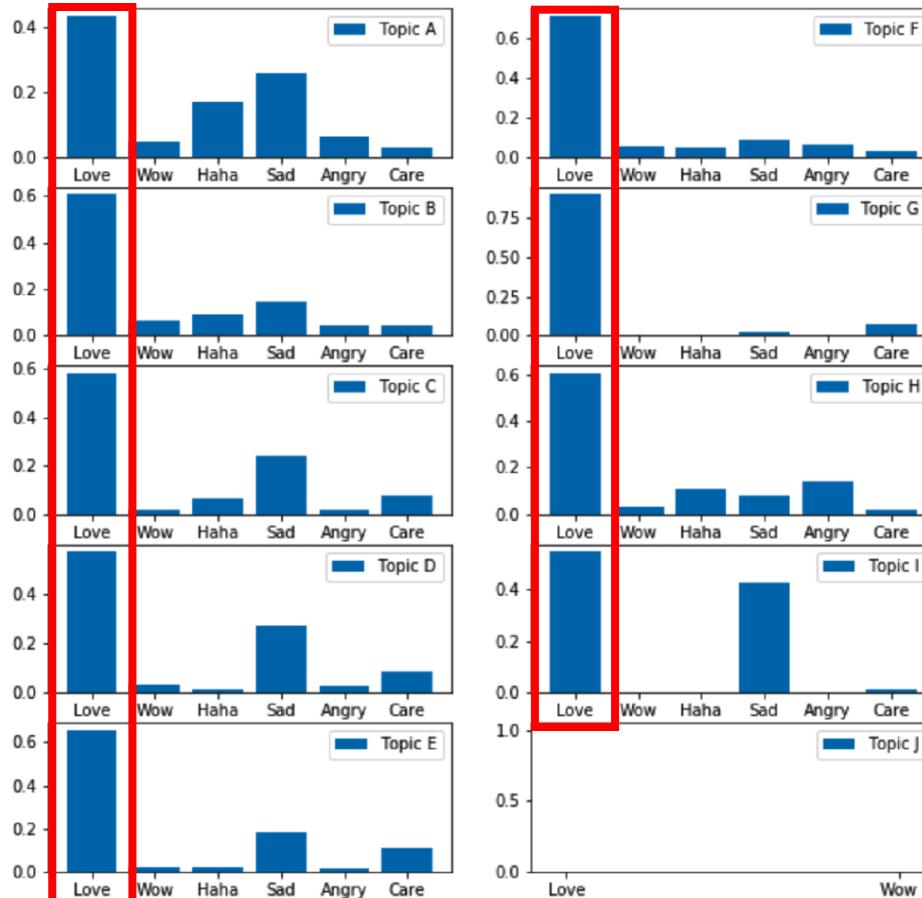
Sentiment distribution by topics on CNN and Fox News posts about COVID-19



(b) Ten topics from group posts

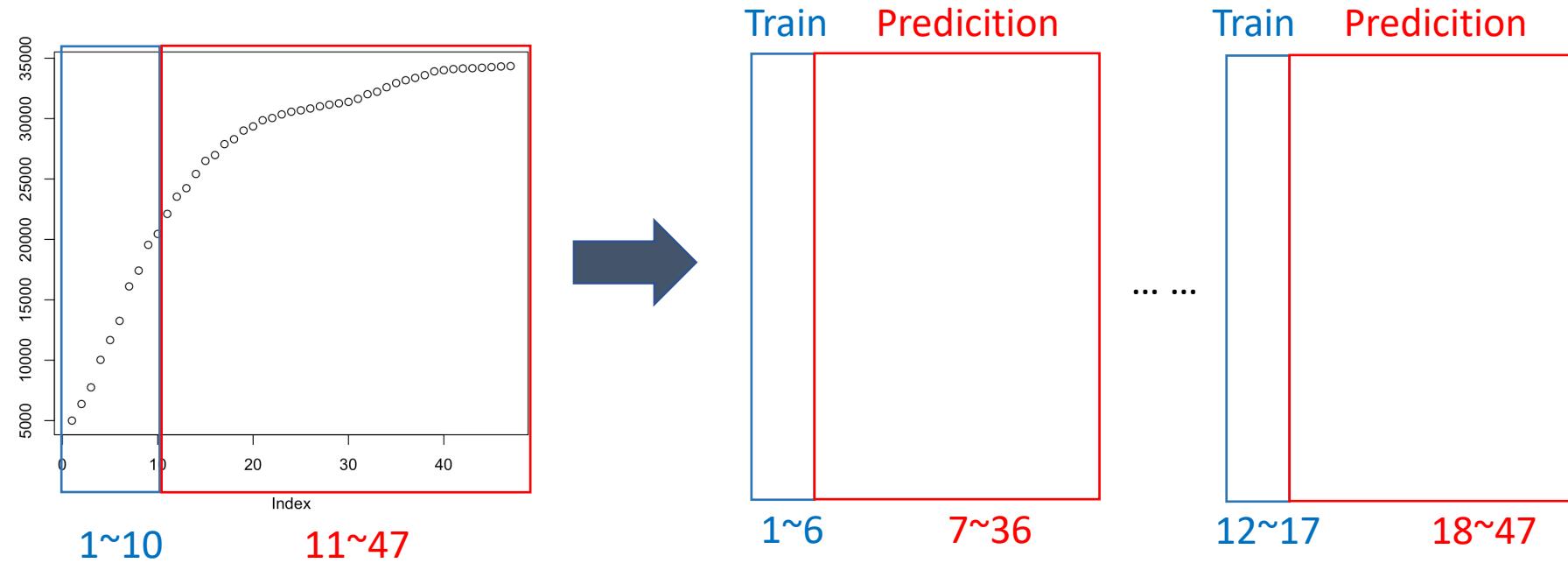
Topic	Component words
A	changing, sound, wore, thanksgiving, copied
B	mexican, advance, cheese, double, posting
C	pastor, testimony, soon, shame, wore
D	test, patients, hospital, positive, tested
E	people, virus, like, home, even
F	public, must, health, community, government
G	jesus, pray, lord, amen, heart
H	world, trump, country, president, lockdown
I	million, name, share, fraud, organize
J	please, family, would, home, keep

Sentiment distribution by topics on Facebook group posts about COVID-19



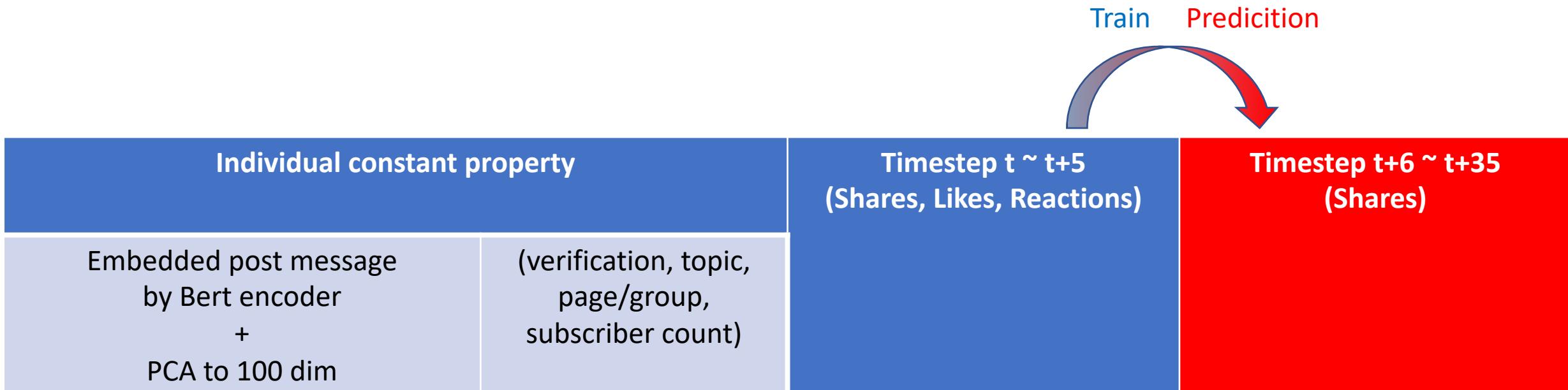
Prediction Modeling

- crawling 6 x 100 most interacted posts by the following queries:
 - 1) COVID and mask, 2) COVID and lockdown, 3) COVID and vaccine
 - 4) COVID and Trump, 5) COVID and Biden, 6) COVID and Jesus
- Slice the post histories by time window (size = 6)
-> 50,280 records



Prediction Modeling

- Feature Engineering



* Timestep = $\log(\text{age of post})$

Prediction Modeling

- Model comparison

	Linear regression	XGBoost	NN (1 FC hidden, 300ep)	NN (2 FC hidden, 300ep)	NN (2 FC hidden, 2000ep)	
Validation RMSE	6111.9370	3143.1568	6027.5431	1514.8007	1053.5369	
Hidden layers' output sizes	-	-	100	100, 50	100, 50	

*NN configuration

```
nnmodel = Sequential()
nnmodel.add(Dense(100, activation='relu'))
nnmodel.add(Dense(50, activation='relu'))
nnmodel.add(Dense(target_split_point, activation='relu'))
nnmodel.compile(loss='mse', metrics=[tf.keras.metrics.RootMeanSquaredError()], optimizer='adam')
# fit network
history = nnmodel.fit(X_train, y_train, epochs=2000, batch_size=100,
                      validation_data=(X_vali, y_vali),
                      verbose=1,
                      shuffle=True)
```

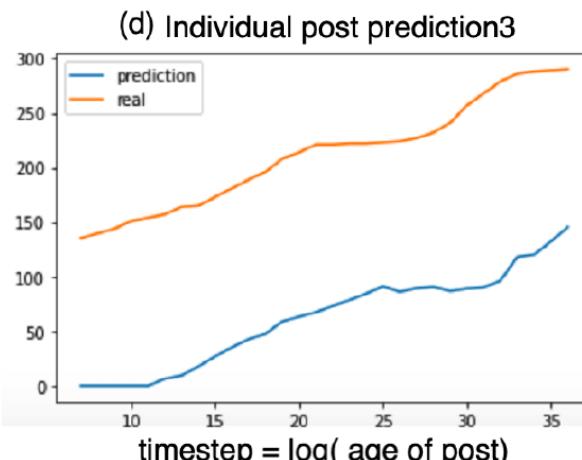
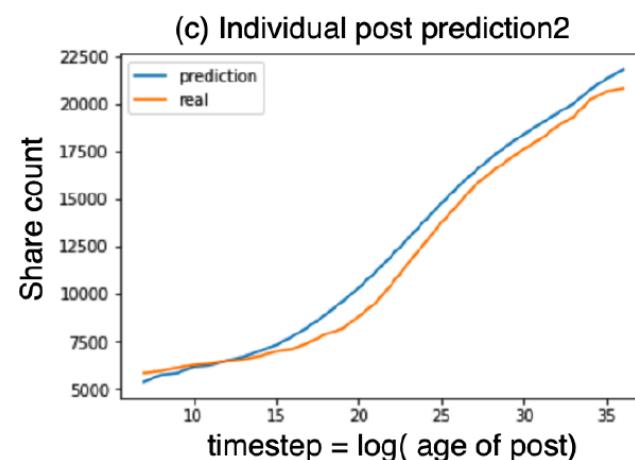
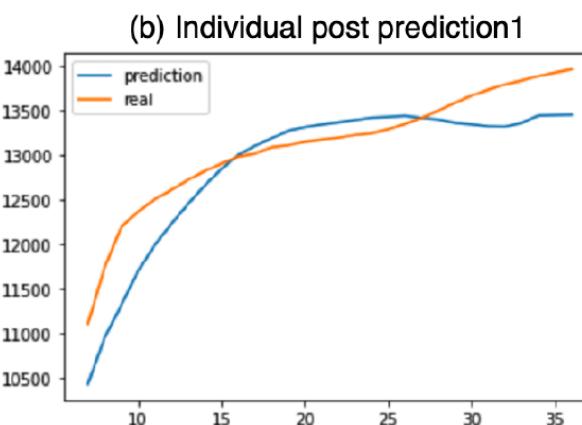
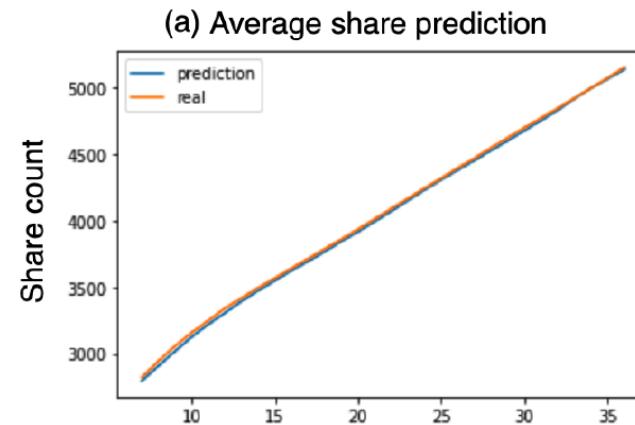
Prediction Modeling

- Feature importance by perturbation (Experiment on train dataset)

Feature	Importance
actual.shareCount6	1134.8206
actual.shareCount5	667.8565
group	382.0848
page	375.1222
... ...	
emb23	28.7548431
actual.likeCount4	28.4147295
actual.sadCount4	27.7950195
score4	27.636936
actual.angryCount5	26.3365968
actual.loveCount2	25.096232
actual.hahaCount4	23.5013038
actual.angryCount4	21.0586289
actual.hahaCount3	18.5350365

Prediction Modeling

- Final prediction on Test dataset



Questions?

[Article on Github \(Link\)](#)



python™

crowdtangle

TensorFlow

scikit
learn

Gephi

