# Do humans and bots tweet differently? Analyzing Twitter data about COVID-19 Vaccination

Yuming Fang, Wenhao Li, Hao Xu, Shuxi Wu

# The Prevalence of Social Bots

- In health field, e-cigarettes promotion, polarized vaccination discussions, uncredible health claims about cannabis

- Amplifying and perpetuating misinformation (Schuchard et al., 2019), disturbing online discussions (Allem, Escobedo, & Dharmapuri, 2020), exacerbating polarization and conflicts (Stella, Ferrara, & De Domenico, 2018), and manipulating public opinions (Cheng, Luo, & Yu, 2020)

# Background

- Date: January 21-28, 2021

    - The earliest batch of COVID-19 vaccines (incl. Pfizer, Moderna, AstraZeneca) rolled out in different places around the world.

    - The Biden Administration’s COVID-19 response plan:
      Jan. 20: Inauguration of the Biden administration
      Jan. 26: “The administration is working to buy 200 million more COVID-19 vaccine doses.” - NPR

# Research Questions

**RQ 1:** Are there any differences in network structures in terms of their connectivity in human-like Twitter handles and bot-like Twitter handles?

**RQ 2:** How did human-like Twitter handles and bot-like Twitter handles interact with each other?

**RQ 3:** What were the main topics/themes in the tweets posted by human-like Twitter handles and bot-like Twitter handles?

# Method

**Data Collection:**

- Pulled tweets with keyword and hashtag query
- Dataset of 120,000 tweets, published from Jan. 21, 2021 to Jan. 28, 2021


**Data processing:**

- Dropped all “verified” twitter accounts to exclude institutional accounts
- Filtered out tweets with no engagement (retweet or quote)
- Converted probability of bot to a binary variable (threshold = 0.5)
- For topic modeling: filtered out language not classified as English

# Method cont.

**Data analysis:**

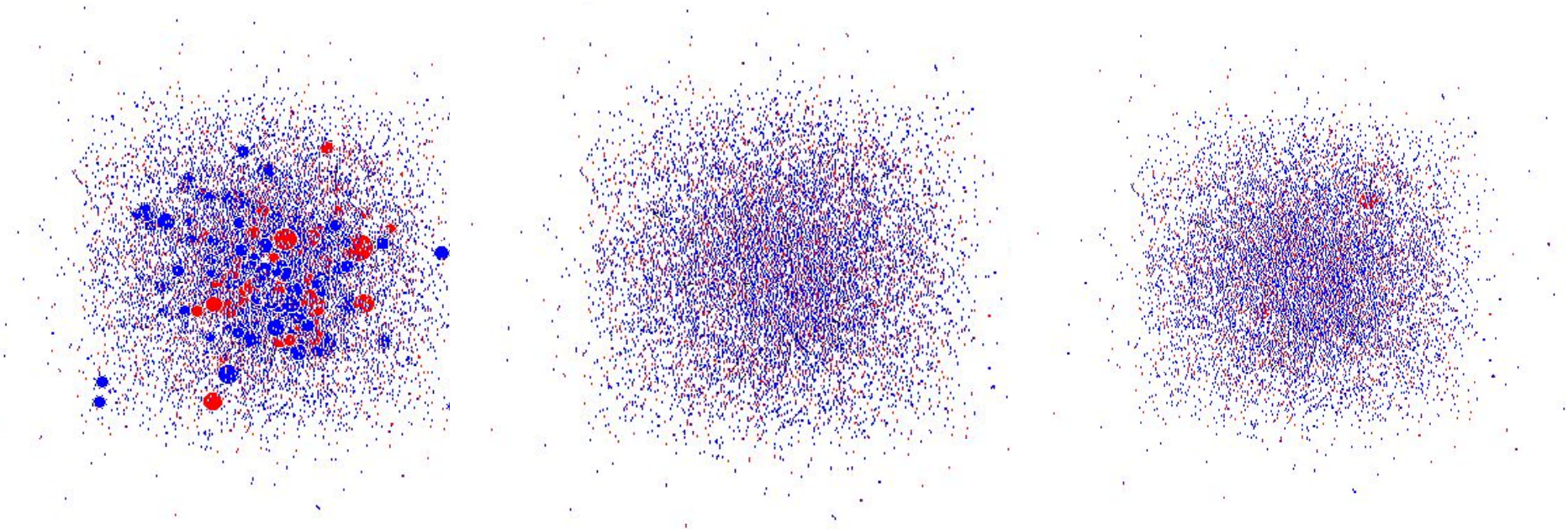
- Social network analysis
  - Combined dataset in r
  - Separate dataset (human and bot) in gephi
- Topic modeling
  - Combined dataset, covariate with human/bot binary variable


**Next steps:**

- Separate human and bot topic modeling analysis
- Reduce topic (K) number for topic modeling

# SNA in r: combined dataset visuals

**In-degree** **Out-degree** **Betweenness**



- Human-like Twitter handles
- Bot-like Twitter handles

# SNA in R: statistics of combined dataset

Mean in-degree: 0.943 (max 2000)

Mean out-degree: 0.943 (max 124)

Mean closeness: 0.920 (max 1)

Mean betweenness: 0.077 (max 210)

Mean distance: 1.077

Graph density: 0.0000104

Diameter: 5

Reciprocity: 0.000304

Transitivity: 0.0000199

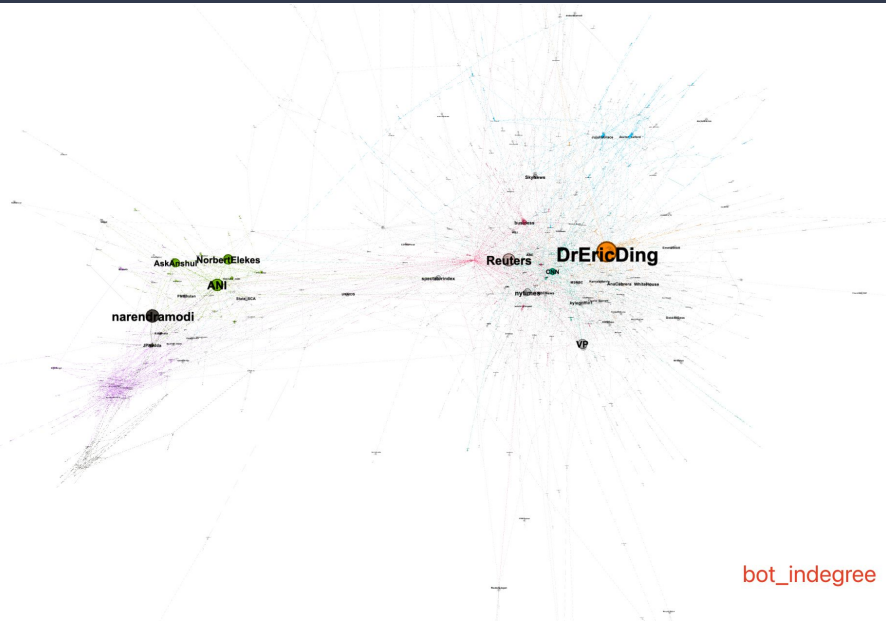
| | name | in_degree |
|---|---|---|
| 1 | DrEricDing | 2000 |
| 2 | Reuters | 1017 |
| 3 | EmmaScott | 865 |
| 4 | VP | 793 |
| 5 | NorbertElekes | 777 |
| 6 | JujuliaGrace | 688 |
| 7 | ANI | 666 |
| 8 | nytimes | 659 |
| 9 | CNN | 621 |
| 10 | spectatorindex | 589 |

| | name | out_degree |
|---|---|---|
| 1 | CoronaUpdateBot | 124 |
| 2 | viralvideovlogs | 77 |
| 3 | vmrwanda | 42 |
| 4 | YouCanFoolMost | 27 |
| 5 | All435Reps | 26 |
| 6 | monitor_PH | 19 |
| 7 | 62_elizay | 16 |
| 8 | john1966olsen | 16 |
| 9 | lizditz | 15 |
| 10 | Evelyn74415780 | 1 |

| | name | bt_degree |
|---|---|---|
| 1 | BpsmithUk | 210.0 |
| 2 | PaulPaul49 | 180.0 |
| 3 | DrAmirKhanGP | 174.0 |
| 4 | SanayaAdhikary | 142.5 |
| 5 | ProfKevinFenton | 132.0 |
| 6 | DataDrivenMD | 92.0 |
| 7 | DipKumarDey11 | 89.0 |
| 8 | TapasJana_bjp | 88.0 |
| 9 | mojos55 | 83.0 |
| 10 | conarck | 80.0 |

# SNA in Gephi: Human-like vs. bot-like handles

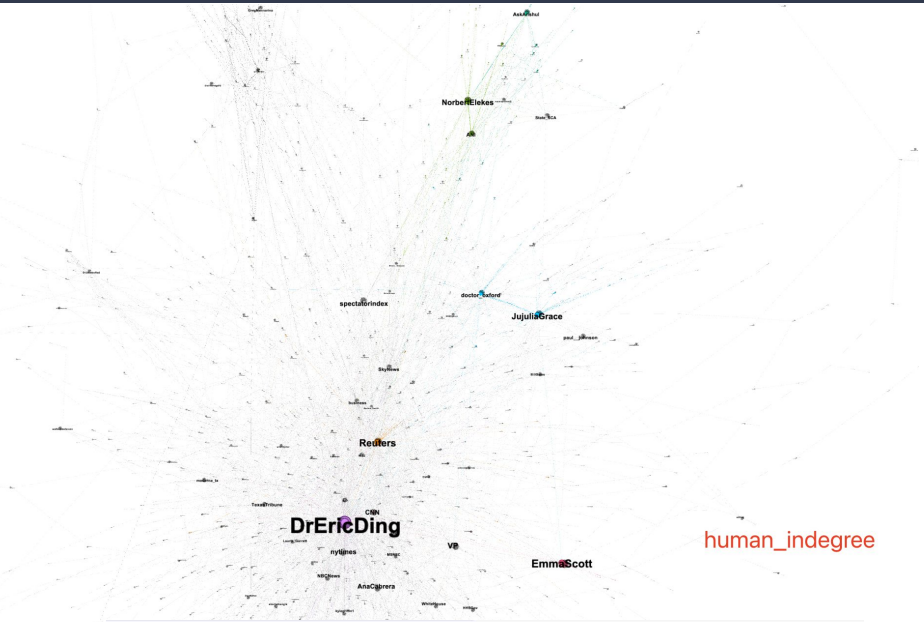
## In-degree: Bot-like handles




bot_indegree

| Network Overview | | | Node Overview | | |
|---|---|---|---|---|---|
| Average Degree | 0.919 | Run | Avg. Clustering Coefficient | 0 | Run |
| Avg. Weighted Degree | 0.919 | Run | Eigenvector Centrality | | Run |
| Network Diameter | 5 | Run | **Edge Overview** | | |
| Graph Density | 0 | Run | Avg. Path Length | 1.08 | Run |

## In-degree: Human-like handles

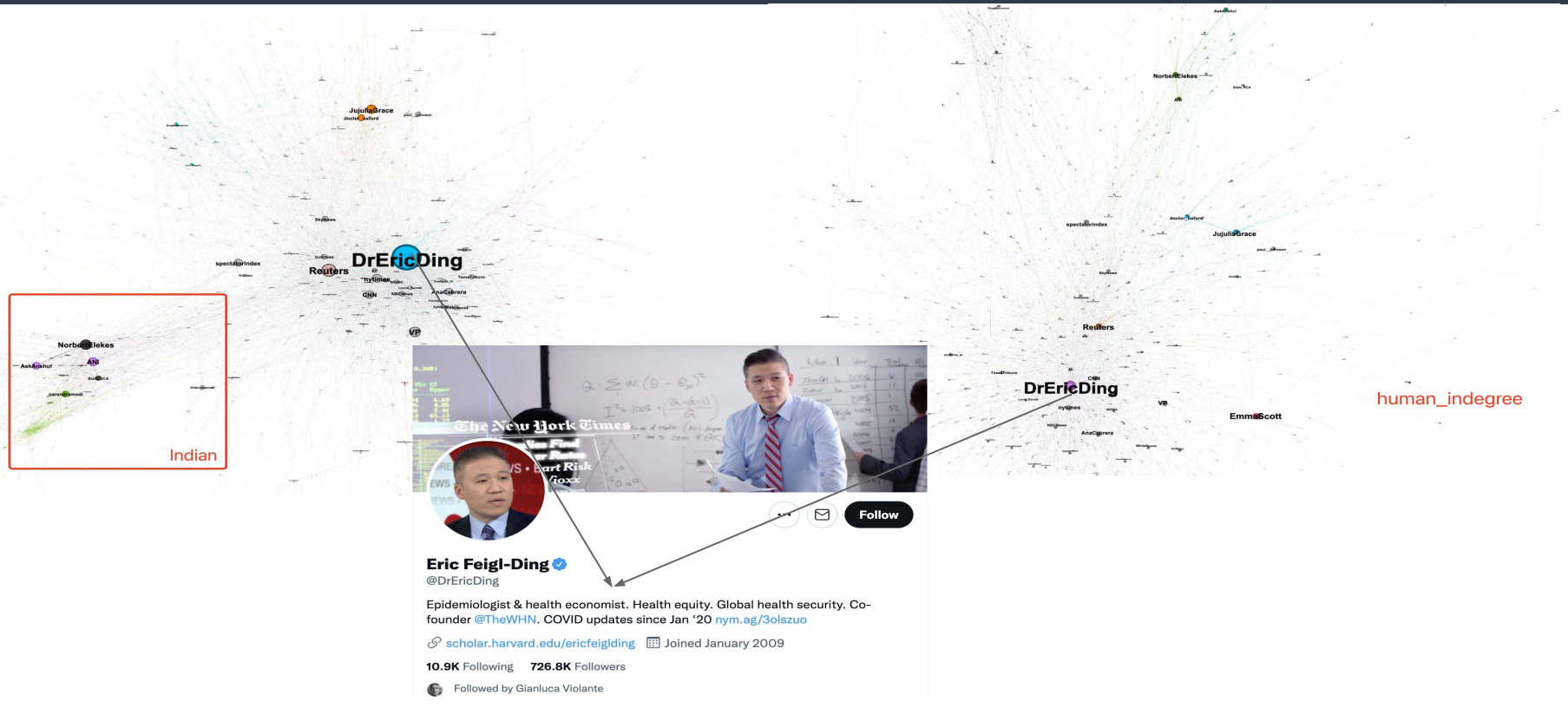



human_indegree

| Network Overview | | | Node Overview | | |
|---|---|---|---|---|---|
| Average Degree | 1.075 | Run | Avg. Clustering Coefficient | 0.041 | Run |
| Avg. Weighted Degree | 1.075 | Run | Eigenvector Centrality | | Run |
| Network Diameter | 4 | Run | **Edge Overview** | | |
| Graph Density | 0 | Run | Avg. Path Length | 1.183 | Run |

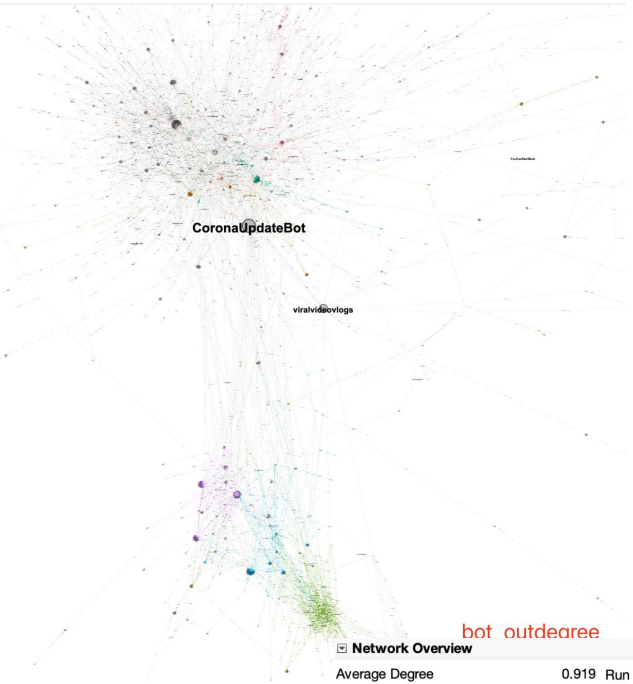
# SNA in Gephi: Human-like vs. bot-like handles

## In-degree comparison



# SNA in Gephi: Human-like vs. bot-like handles

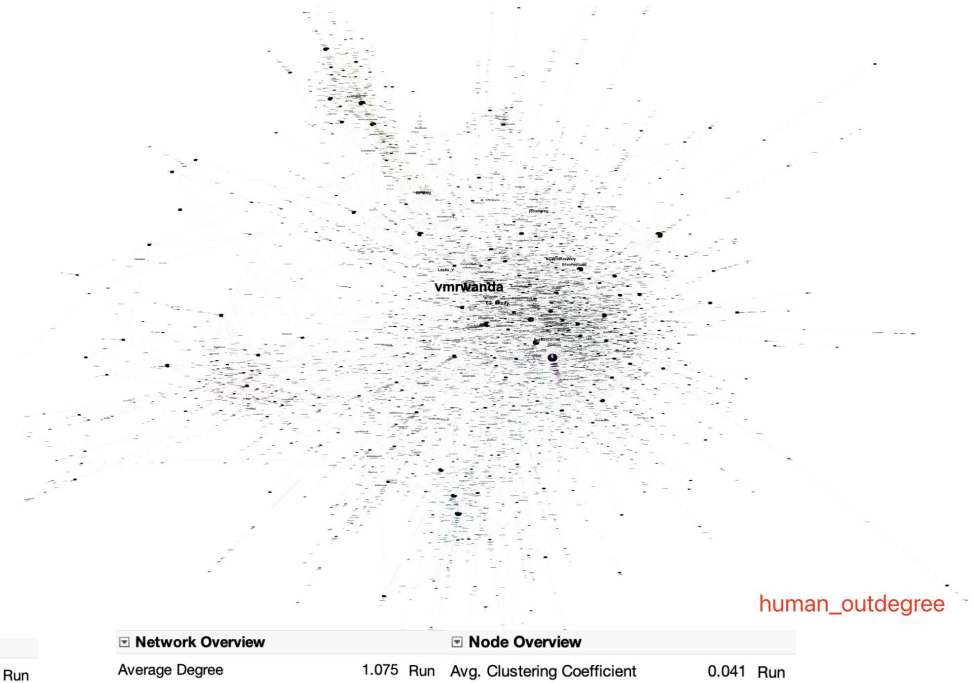
## Out-degree: Bot-like handles




bot_outdegree

| Network Overview | | | Node Overview | | |
|---|---|---|---|---|---|
| Average Degree | 0.919 | Run | Avg. Clustering Coefficient | 0 | Run |
| Avg. Weighted Degree | 0.919 | Run | Eigenvector Centrality | | Run |
| Network Diameter | 5 | Run | **Edge Overview** | | |
| Graph Density | 0 | Run | Avg. Path Length | 1.08 | Run |

## Out-degree: Human-like handles

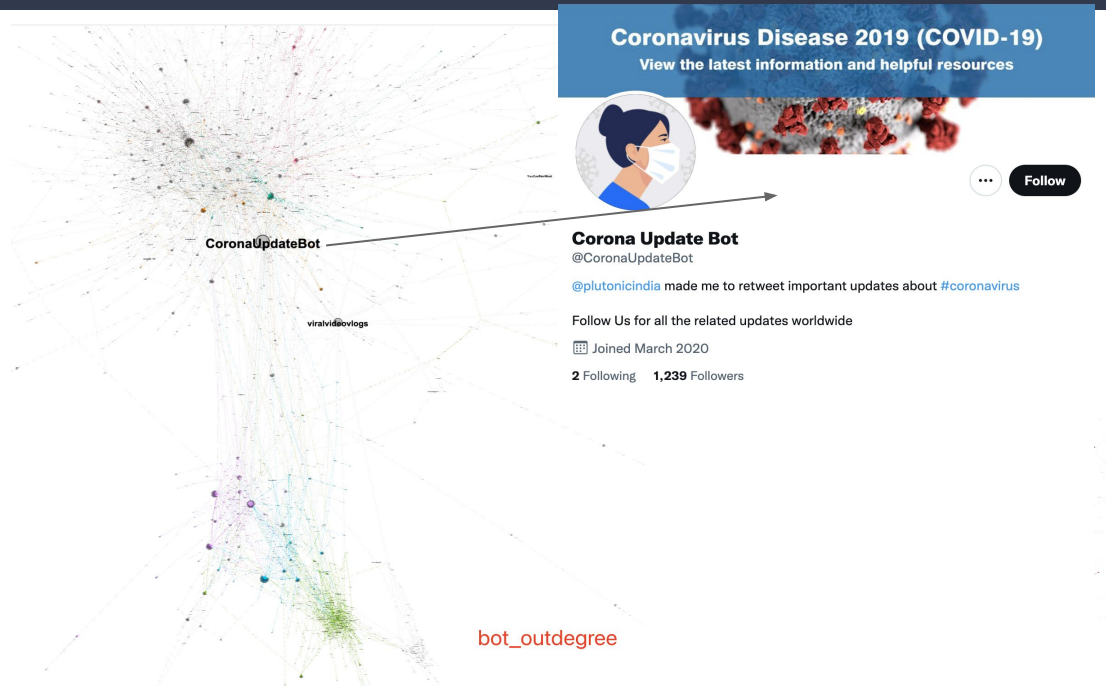



human_outdegree

| Network Overview | | | Node Overview | | |
|---|---|---|---|---|---|
| Average Degree | 1.075 | Run | Avg. Clustering Coefficient | 0.041 | Run |
| Avg. Weighted Degree | 1.075 | Run | Eigenvector Centrality | | Run |
| Network Diameter | 4 | Run | **Edge Overview** | | |
| Graph Density | 0 | Run | Avg. Path Length | 1.183 | Run |

# SNA in Gephi: Human-like vs. bot-like handles

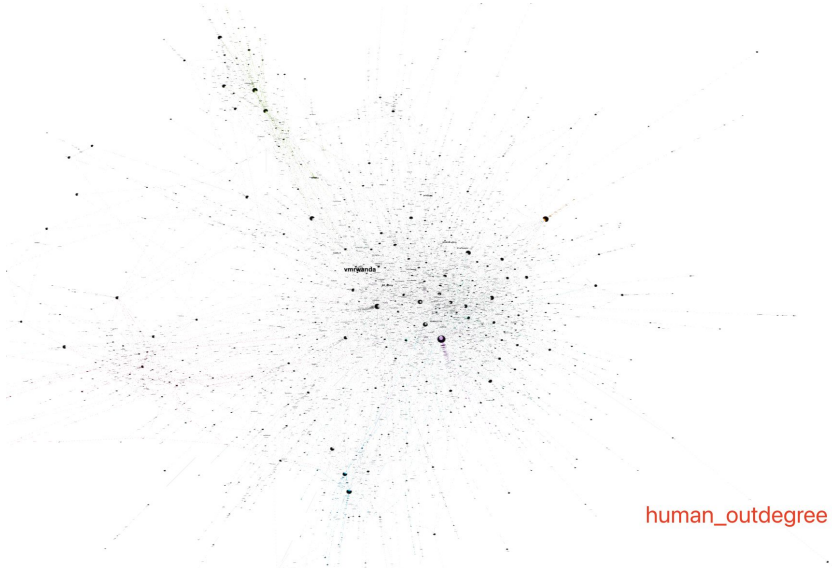
## Out-degree comparison







bot_outdegree




human_outdegree

# Topic modeling

**Basic parameters:**

- Lower.threshold = 63 (80% percentile), 20 (90% percentile)
- K = 12 (from threshold = 1000)

**Spoiler: A lot of limitations and confusions**

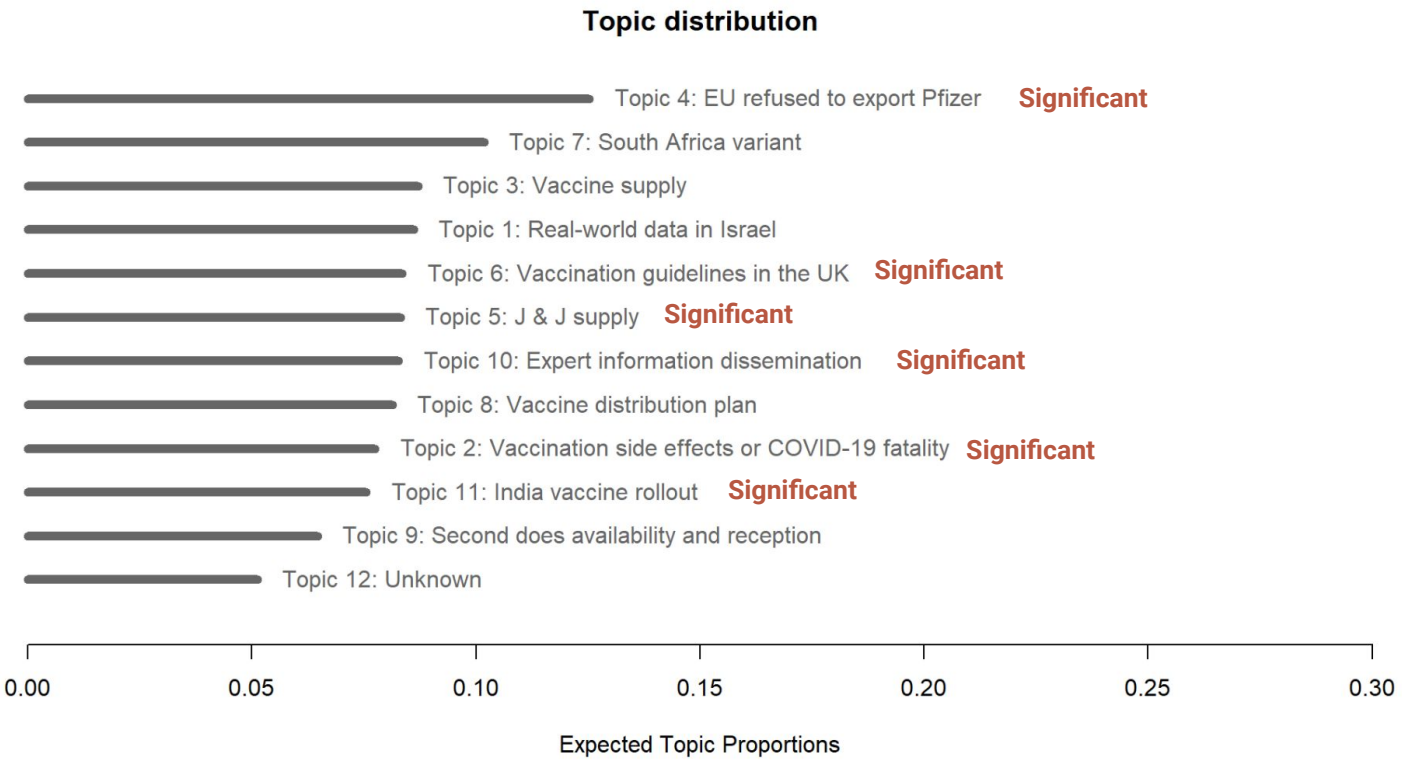
# Topic Modeling raw results

| | Topics 1 | Topics 2 | Topics 3 | Topics 4 | Topics 5 | Topics 6 | Topics 7 |
|---|---|---|---|---|---|---|---|
| 1 | said | county | workers | eu | just | nhs | south |
| 2 | us | officials | site | news | johnson | receive | variant |
| 3 | dr | many | countries | million | says | can | indias |
| 4 | day | people | also | doses | two | vaccinated | effective |
| 5 | israel | died | world | first | care | please | moderna |
| 6 | make | need | health | good | pfizer | get | uk |
| 7 | one | like | still | update | receiving | staff | variants |
| 8 | help | doses | + | country | th | know | new |
| 9 | take | receiving | data | state | back | b | free |
| 10 | government | government | supply | \| | first | doses | global |
| 11 | vaccinated | plan | back | help | government | two | breaking |
| 12 | world | state | th | government | announced | government | b |
| 13 | country | us | update | breaking | get | first | doses |
| 14 | first | say | state | announced | help | make | pfizer |
| 15 | pfizer | vaccinated | public | just | doses | still | health |
| 16 | just | update | government | two | make | one | announced |
| 17 | get | b | care | pfizer | dr | data | says |
| 18 | doses | two | vaccinated | us | health | need | first |
| 19 | two | first | doses | b | supply | us | just |
| 20 | news | get | good | new | still | people | government |

# Topic Modeling raw results

| | Topics 8 | Topics 9 | Topics 10 | Topics 11 | Topics 12 |
|---|---|---|---|---|---|
| 1 | biden | received | now | india | response |
| 2 | trump | second | year | pm | week |
| 3 | administration | dose | soon | drive | got |
| 4 | distribution | available | old | worlds | work |
| 5 | president | dont | im | rollout | minister |
| 6 | plan | government | fauci | pandemic | t |
| 7 | states | today | getting | w | public |
| 8 | announced | first | today | today | get |
| 9 | say | get | asked | th | first |
| 10 | new | know | first | take | one |
| 11 | \| | pfizer | b | government | health |
| 12 | million | say | pfizer | just | government |
| 13 | doses | people | help | first | back |
| 14 | news | one | update | \| | th |
| 15 | government | israel | government | b | dr |
| 16 | help | just | th | said | pfizer |
| 17 | b | doses | one | doses | new |
| 18 | state | announced | dr | country | announced |
| 19 | first | receiving | say | people | just |
| 20 | t | state | vaccinated | one | doses |

# Topic Modeling results (covariate with bot/human)



# (Preliminary) Conclusions

**RQ 1: Are there any differences in network structures in terms of their connectivity in human-like Twitter handles and bot-like Twitter handles?**

Bot-like handles in our dataset seems to have more than one network cluster, whereas human-like handles' network structure is fairly centralized.

**RQ 2: How did human-like Twitter handles and bot-like Twitter handles interact with each other?**

Human- and bot- like handles interact in a relatively equal way; it doesn't seem that discrimination exists between human- and bot-like handles.

**RQ 3: What were the main topics/themes in the tweets posted by human-like Twitter handles and bot-like Twitter handles?**

Trying to distinguish distinct themes from our topic modeling results were difficult; a next step may be to reduce topic numbers (k), or to enlist more NLP tools. There are some significant differences between human and bot in the themes they discuss, which can be further explored by running topic modeling separately.