An Iterative Link Analysis Method for Finding and Ranking Influencers on Social Media

Hong Qu, Min Gong, Qingtao Cao
Northeastern University

Contagion on Complex Social Systems 2022
August 10, 2022

Outline

Real World
Anti-Vaccine
Discourse

Simulation
Agent-based Model

Evaluate Results

Who are anti-vaccine influencers?

Full Fact @FullFact

A clip of Dr Aseem Malhotra on GB News talking about research linking heart disease to the Covid-19 vaccines has been viewed over two million times.

There are serious concerns about the quality of the research mentioned.

fullfact.org
Concerns raised about legitimacy of research linking vaccines and heart attack...
It's been claimed new research shows a link between Covid-19 vaccines and heart attacks. The research has serious flaws.

HEALTH / CORONAVIRUS

Concerns raised about legitimacy of research linking vaccines and heart attacks

30 NOVEMBER 2021

WHAT WAS CLAIMED

A study found the risk of a group of patients having a heart attack within five years increased from 11% to 25% after they were given their mRNA Covid-19 vaccine.

OUR VERDICT

Serious concerns have been raised as to the quality of the research. Its publisher notes it may contain "potential errors".

See the action we've taken as a result of this fact check.

> *"What this research has shown is that markers associated with increasing the risk of heart attack and probably even progression of underlying heart disease in people who have already got some heart disease. There's been a significantly increased risk from 11% at five years, the risk of heart attack, to 25%."*
>
> DR ASEEM MALHOTRA, 25 NOVEMBER 2021.

A clip of health campaigner and cardiologist Dr Aseem Malhotra on GB News talking about claims linking the Covid-19 vaccines to heart attacks has gone viral on Twitter and been viewed at least one million times.

Not Your Mom Retweeted In-degree

Dr Aseem Malhotra
@DrAseemMalhotra

BREAKING:

Repeat Booster Shots Spur European Warning on Immune-System Risks

'European Union regulators warned that frequent Covid-19 booster shots could adversely affect the immune system and may not be feasible'

It's all starting to unravel

bloomberg.com
Frequent Boosters Spur Warning on Immune Response
European Union regulators warned that frequent Covid-19 booster shots could adversely affect the immune response and may not be feasible.

9:40 PM · Jan 11, 2022 · Twitter for iPhone

3,943 Retweets **383** Quote Tweets **7,853** Likes

Dr Clare Craig (not one of her impersonators) Retweeted In-degree

Dr Aseem Malhotra
@DrAseemMalhotra

Out-degree

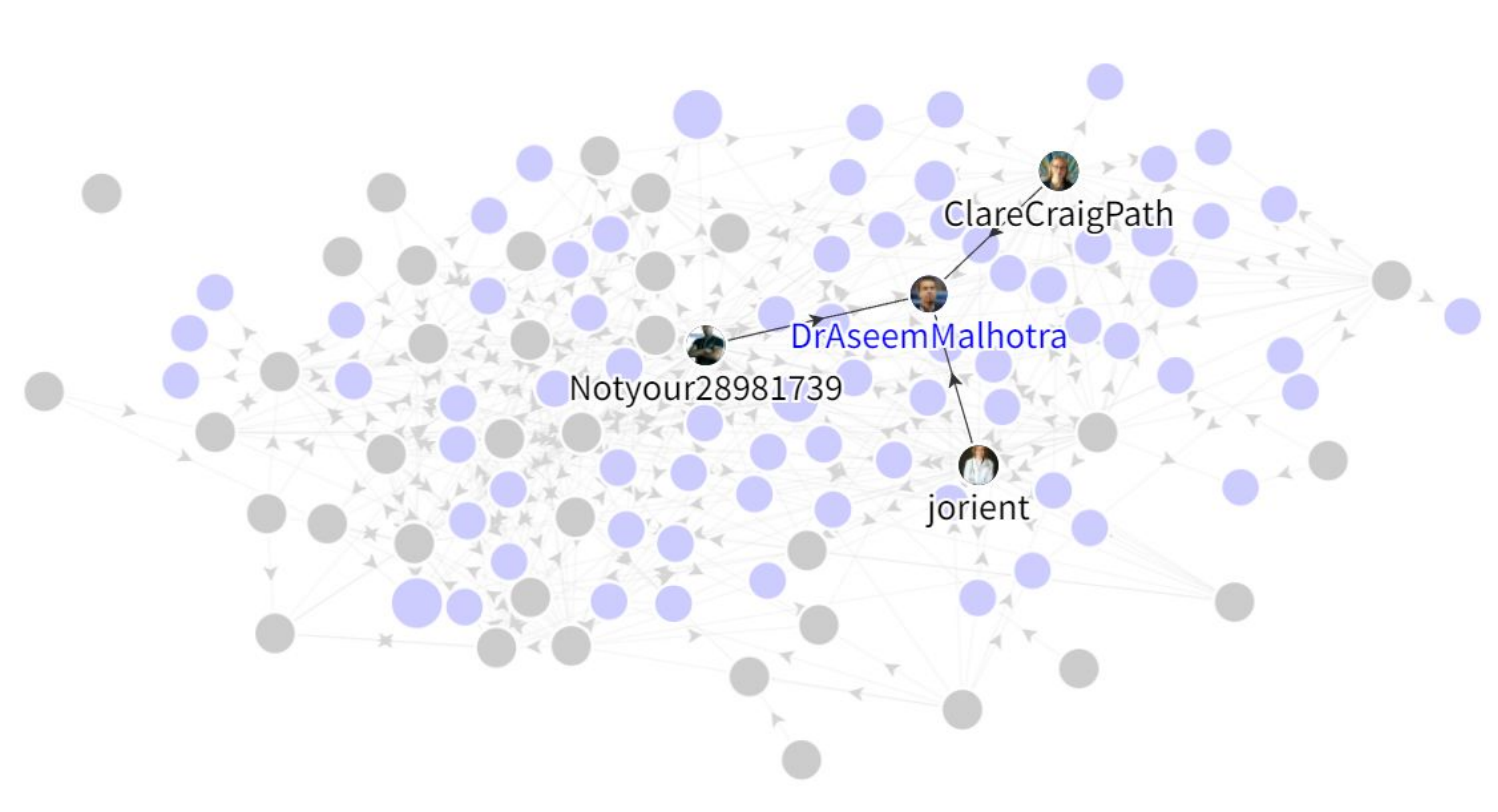
Brilliant. @sajidjavid PLEASE U TURN on vaccine mandate for #NHS staff. It is not ethical, will result in catastrophic staff shortages, and is not backed by the BMA nor the Academy of Medical Royal Colleges

#informedconsent

> **Adam Brooks** @EssexPR · Nov 30, 2021
> 🚨US JUDGE BLOCKS VACCINE MANDATE FOR HEALTH CARE WORKERS
>
> "If human nature & history teach anything, it is that civil liberties face grave risks when governments proclaim indefinite states of emergency"
>
> Judge Doughty wrote in his order blocking Vaccine mandate.

6:44 AM · Dec 1, 2021 · Twitter for iPhone

1,079 Retweets **32** Quote Tweets **3,118** Likes

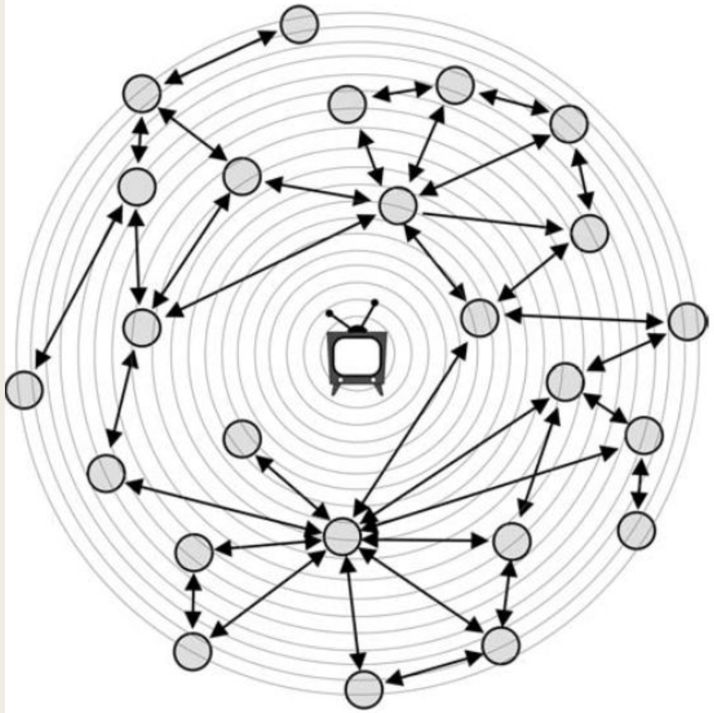

ClareCraigPath
DrAseemMalhotra
Notyour28981739
jorient

Influentials, Networks, and Public Opinion Formation

DUNCAN J. WATTS
PETER SHERIDAN DODDS*

FIGURE 2

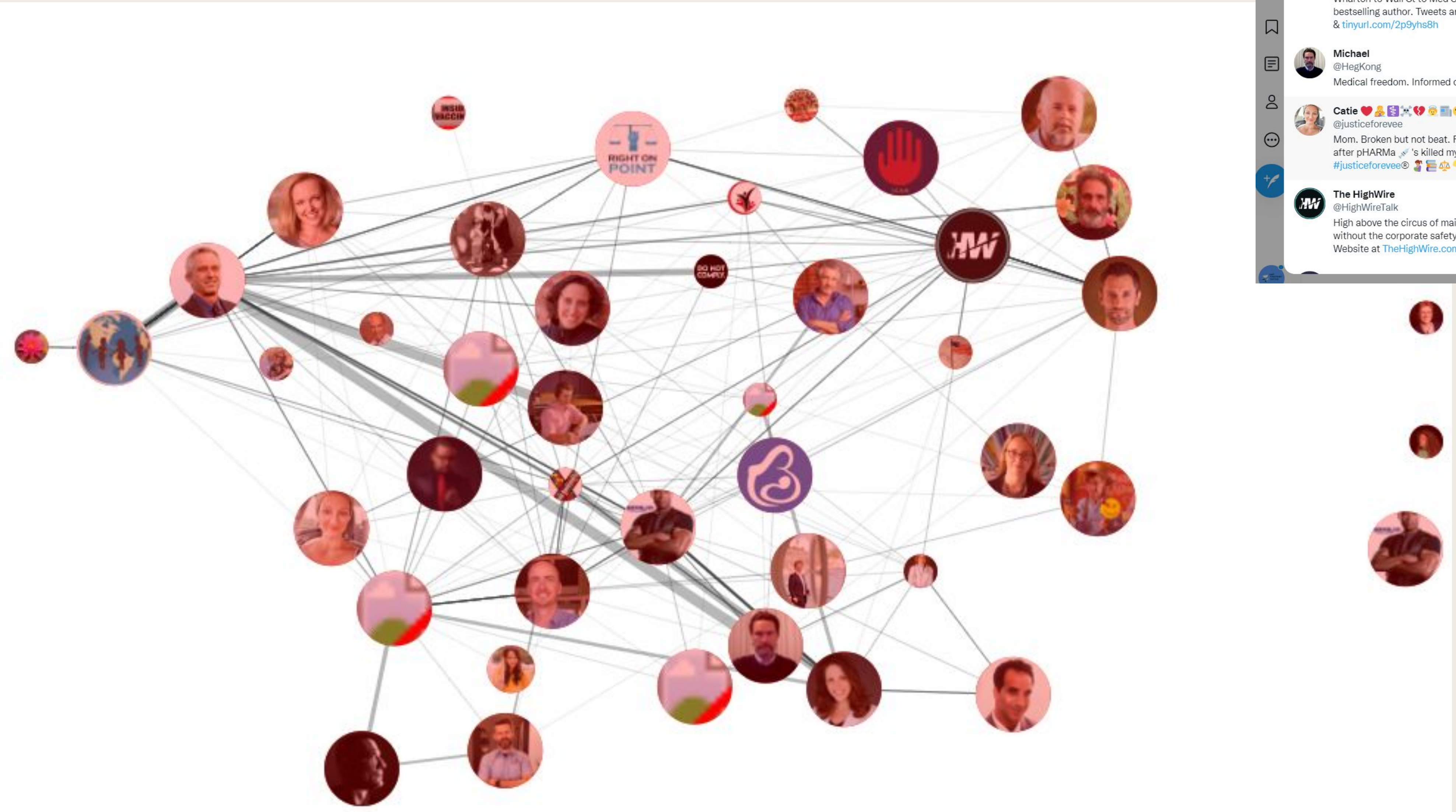
SCHEMATIC OF NETWORK MODEL OF INFLUENCE

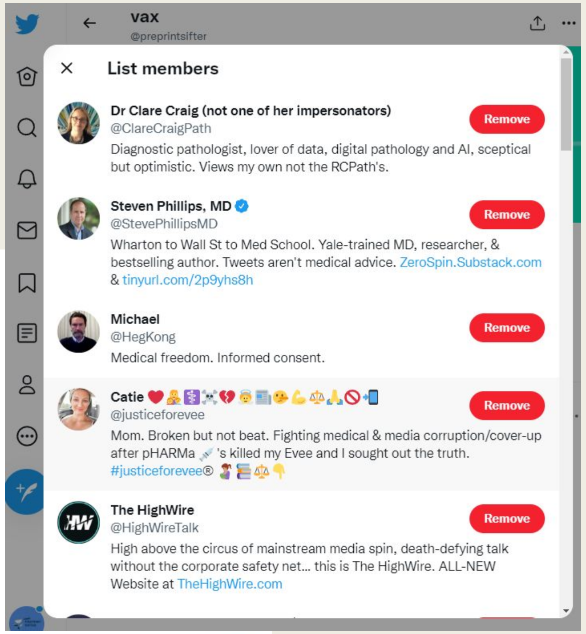


A central idea in marketing and diffusion research is that influentials—a minority of individuals who influence an exceptional number of their peers—are important to the formation of public opinion. Here we examine this idea, which we call the "influentials hypothesis," using a series of computer simulations of interpersonal influence processes. Under most conditions that we consider, we find that large cascades of influence are driven not by influentials but by a critical mass of easily influenced individuals. Although our results do not exclude the possibility that influentials can be important, they suggest that the influentials hypothesis requires more careful specification and testing than it has received.

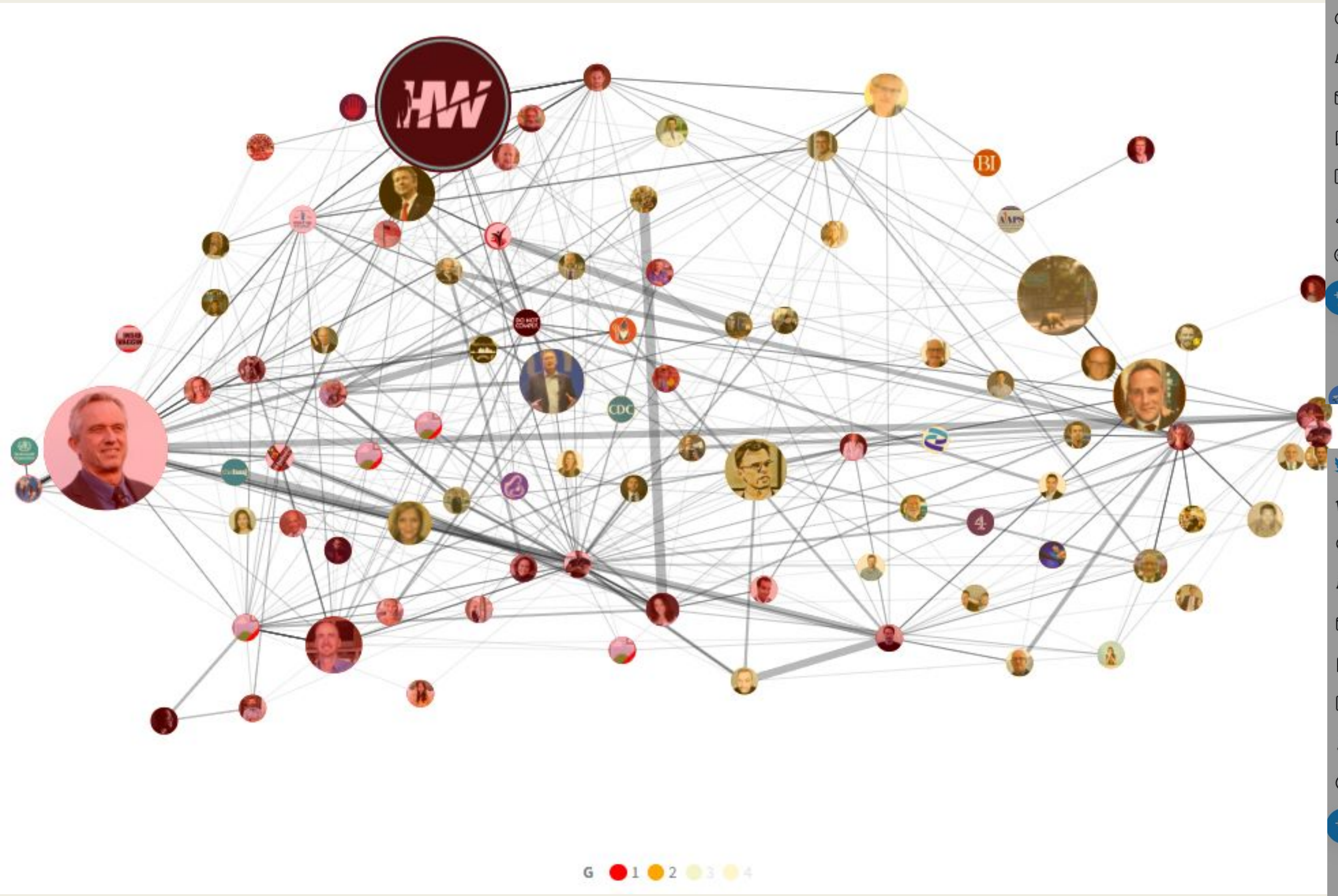
"Influentials, by definition, are relatively rare—here they constitute just 10% of the population—thus they are necessarily more difficult to locate than average individuals and possibly more difficult to mobilize also"

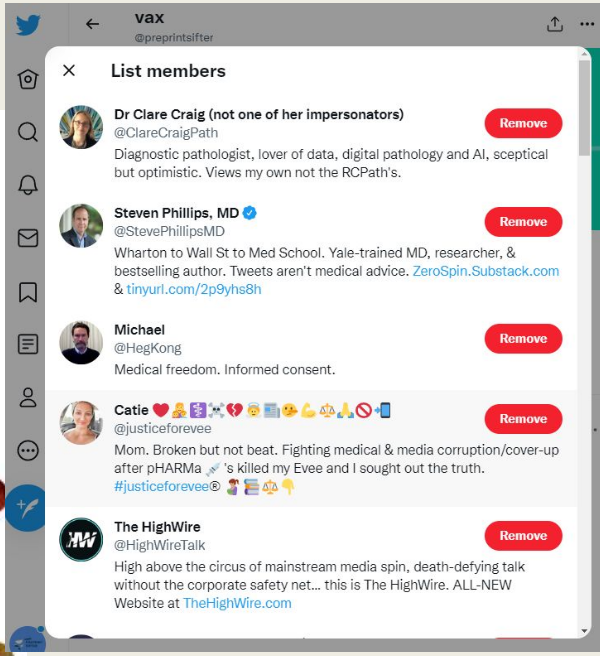
Seed list of 30 anti-vax users selected by health experts

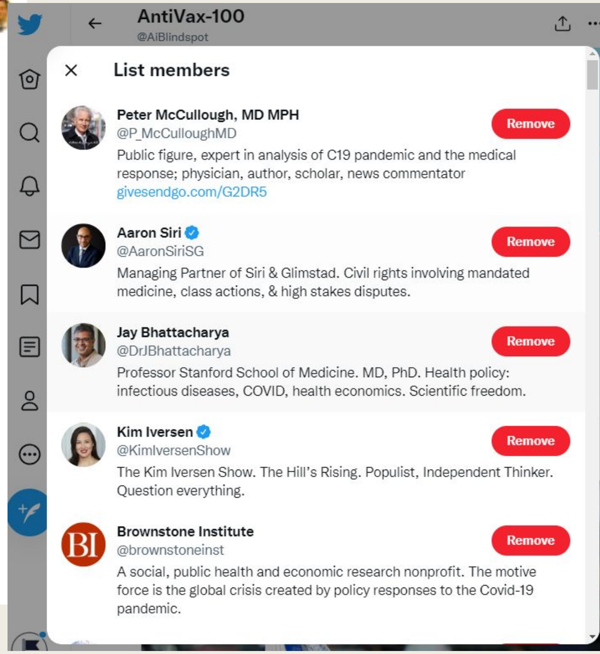


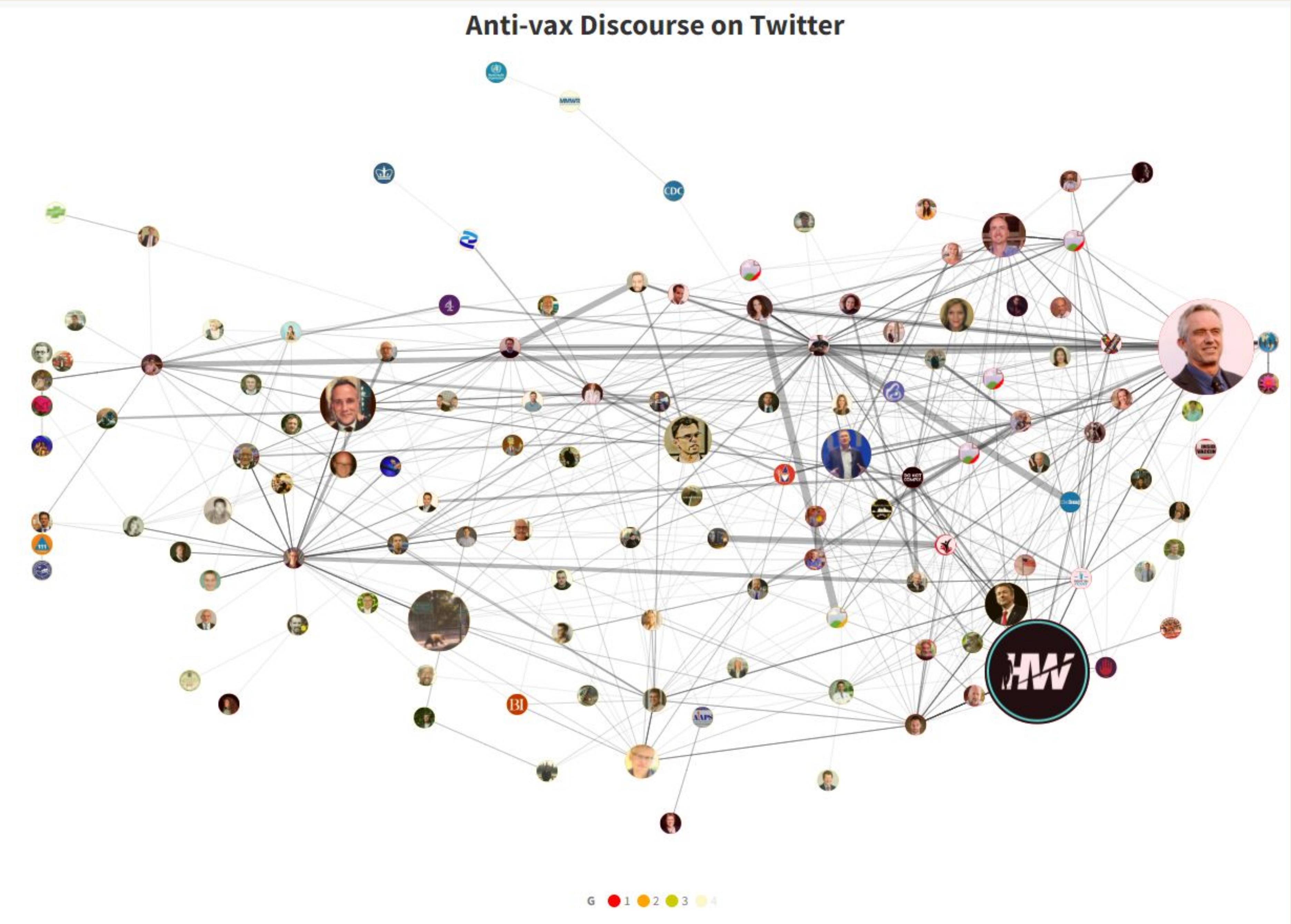


https://twitter.com/i/lists/1333771036274331654/members


HW
BI
CDC
G
1
2

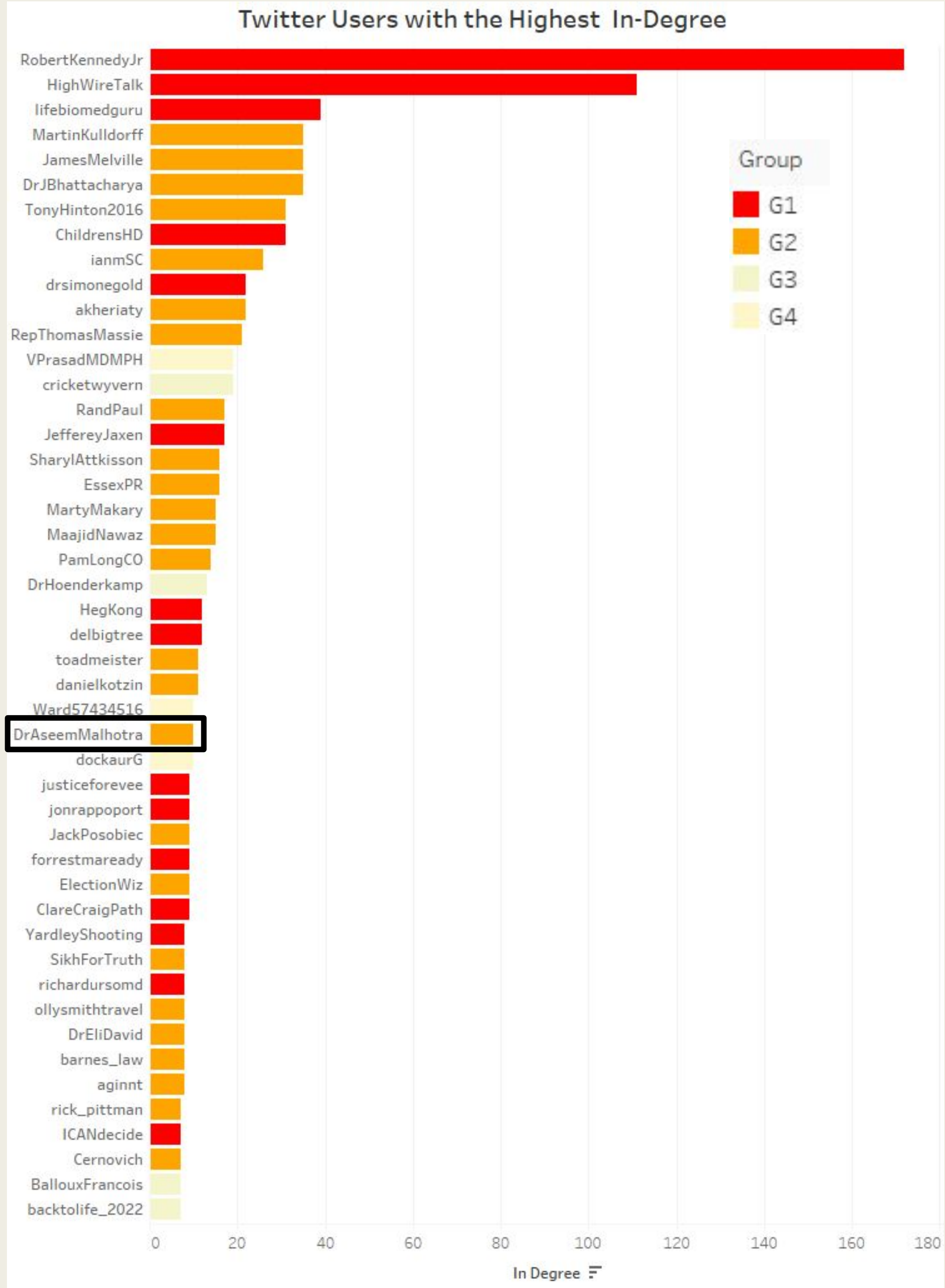

vax
@preprintsifter
List members
Dr Clare Craig (not one of her impersonators)
@ClareCraigPath
Diagnostic pathologist, lover of data, digital pathology and AI, sceptical but optimistic. Views my own not the RCPath's.
Remove
Steven Phillips, MD
@StevePhillipsMD
Wharton to Wall St to Med School. Yale-trained MD, researcher, & bestselling author. Tweets aren't medical advice. ZeroSpin.Substack.com & tinyurl.com/2p9yhs8h
Remove
Michael
@HegKong
Medical freedom. Informed consent.
Remove
Catie
@justiceforevee
Mom. Broken but not beat. Fighting medical & media corruption/cover-up after pHARMa 's killed my Evee and I sought out the truth. #justiceforevee®
Remove
The HighWire
@HighWireTalk
High above the circus of mainstream media spin, death-defying talk without the corporate safety net… this is The HighWire. ALL-NEW Website at TheHighWire.com
Remove

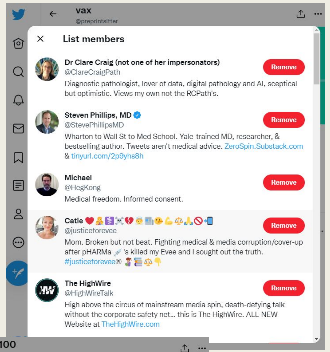

AntiVax-100
@AIBlindspot
List members
Peter McCullough, MD MPH
@P_McCulloughMD
Public figure, expert in analysis of C19 pandemic and the medical response; physician, author, scholar, news commentator givesendgo.com/G2DR5
Remove
Aaron Siri
@AaronSiriSG
Managing Partner of Siri & Glimstad. Civil rights involving mandated medicine, class actions, & high stakes disputes.
Remove
Jay Bhattacharya
@DrJBhattacharya
Professor Stanford School of Medicine. MD, PhD. Health policy: infectious diseases, COVID, health economics. Scientific freedom.
Remove
Kim Iversen
@KimIversenShow
The Kim Iversen Show. The Hill's Rising. Populist, Independent Thinker. Question everything.
Remove
Brownstone Institute
@brownstoneinst
A social, public health and economic research nonprofit. The motive force is the global crisis created by policy responses to the Covid-19 pandemic.
Remove

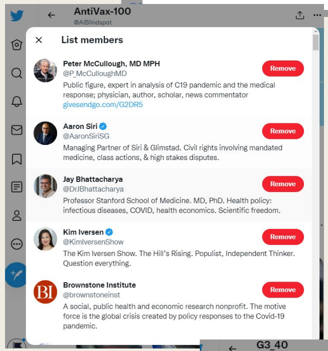


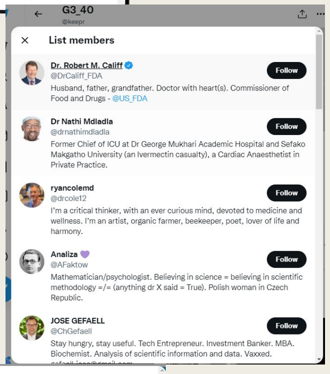
https://public.flourish.studio/visualisation/9620477/

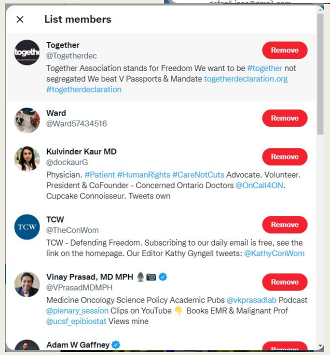
Generation 1	Generation 2	Generation 3	Generation 4
mercola	Channel4News	CityJournal	VPrasadMDMPH
forrestmaready	SteveBakerHW	abiroberts	HowardGriffiths
TannersDad	disclosetv	cvpayne	sajidjavid
doctorsensation	spectator	buckyouhorses	Ward57434516
jonrappoport	ianmSC	cricketwyvern	awgaffney
RobertKennedyJr	bmj_latest	alanvibe	JuliaHB1
med1cinewoman	nypost	CDCMMWR	ECDC_EU
ICANdecide	toadmeister	brianvastag	TheConWom
drsimonegold	zerohedge	BenGoldsmith	Togetherdec
picphysicians	ShannonBream	BuffyWicks	dockaurG
waynerohde	JamesMelville	BucksCouncil	
lifebiomedguru	ezralevant	DrPanMD	
jorient	TuckerCarlson	ajlamesa	
ritamollerpalma	drdavidsamadi	derekjamesfrom	
delbigtree	ollysmithtravel	doctorcall	
DrWakefield	NICKIMINAJ	DavidAnber	
JenniferMarguli	EssexPR	eduortu	
YardleyShooting	AAPSonline	buckscgs	
Peeps_TV	MaajidNawaz	ColumbiaMed	
AlietaEck	ClayTravis	ake2306	
JeffereyJaxen	akheriaty	AnnieWBelle	
ClareCraigPath	ZubyMusic	Azeem_Majeed	
StevePhillipsMD	MartyMakary	DrHoenderkamp	
HegKong	jimmy_dore	EdMHill	
justiceforevee	TheMarieOakes	ATSoos	
HighWireTalk	nadhimzahawi	AgJury	
richardursomd	rick_pittman	DrCaliff_FDA	
ChildrensHD	aginnt	drnathimdladla	
	SharylAttkisson	drcole12	
	indepdubnrth	AFaktow	
	RandPaul	ChGefaell	
	SenRonJohnson	d5_rss	
	ToniaBuxton	DrTessaT	
	kksheld	BallouxFrancois	
	DrAseemMalhotra	backtolife_2022	
	JackPosobiec	AnimalsRule9	
	NeilClark66	CC_CRF	
	RepThomasMassie		
	davidkurten		
	MythinformedMKE		
	danielkotzin		
	DrCharlesL		
	barnes_law		
	MartinKulldorff		
	NBSaphierMD		

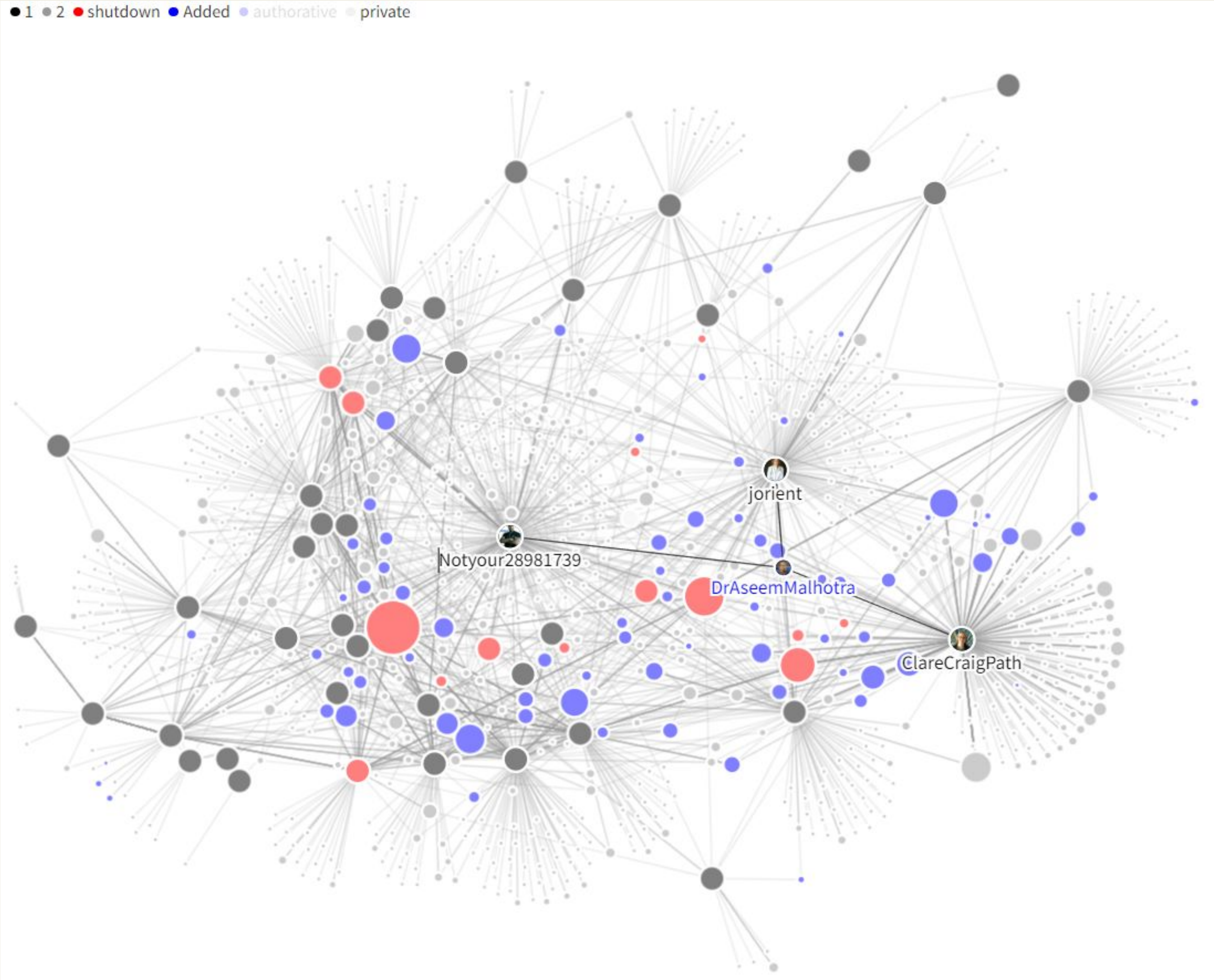











1
2
shutdown
Added
authorative
private
jorient
Notyour28981739
DrAseemMalhotra
ClareCraigPath

Design of Agent-based Model Simulation

- Why
  - *High volume and velocity of real-world data*
  - *Lack of access to “God view” data*
- Goal
  - *To mimic misinformation spread on Twitter*
  - *To test our proposed strategy of exploring and finding influencers*
- How
  - *Nodes: 500*
  - *Messages: 50*

Parallel to real world dynamics

1. Among all users, only a small portion is popular.
2. In a specific topic, there is usually a group of users who are very active.
3. A message always roots from one “news maker” originally.
4. In a fake news community, most of users believe in it, fake news usually spreads more broadly than true news; there are very few authorities may get involved, who won’t spread misinformation.
5. Users who have higher level of willingness may have higher probability to retweet the posts.

1 and 2 Node’s attribute: popularity

3 Each message starts from one user.

4 Message’s attribute: fakeness

4 Node’s attribute: credibility

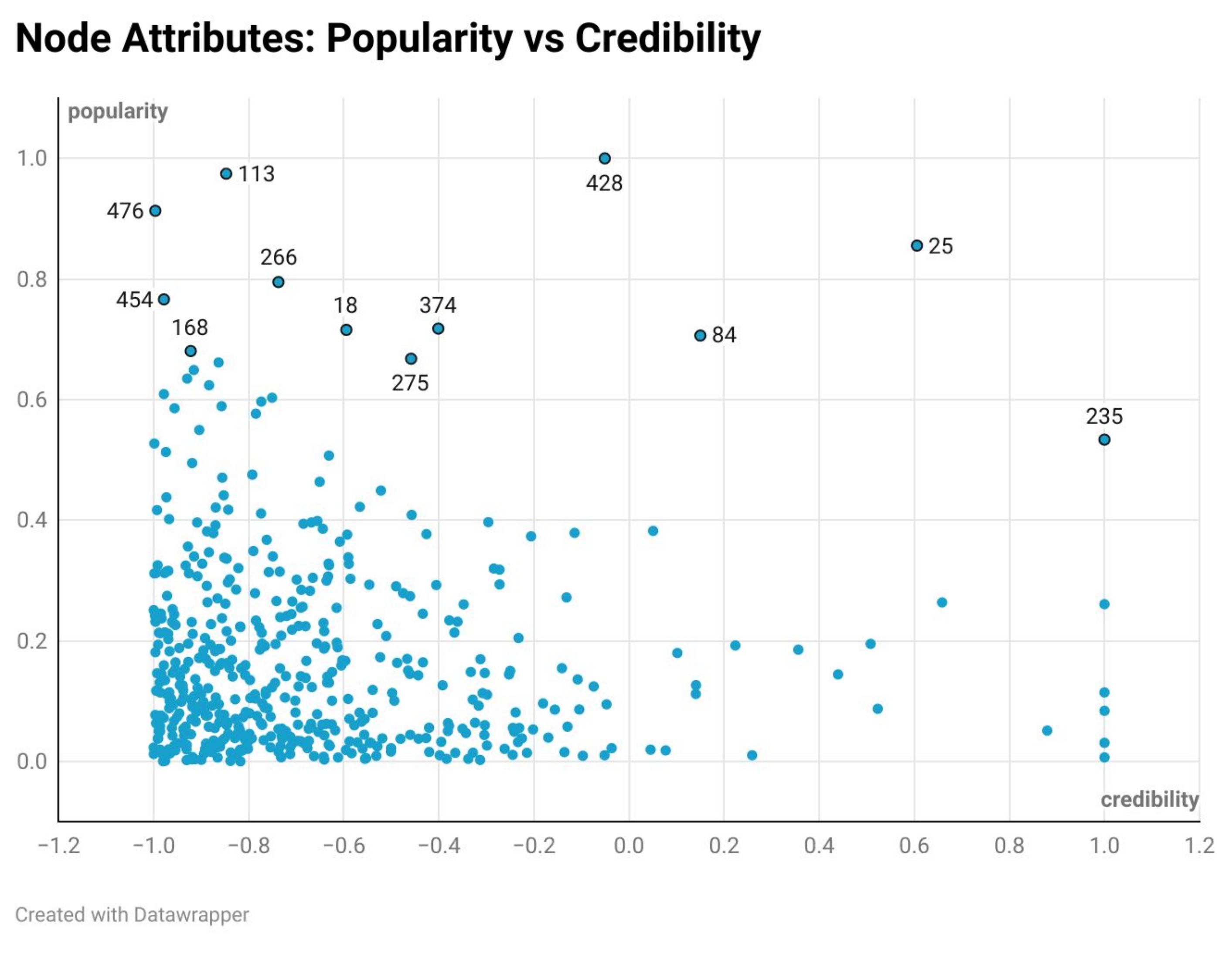
5 Node’s attribute: willingness

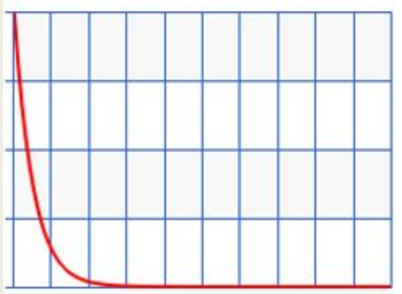
Simulation settings

- Nodes: users, 500
  - ***Popularity***: [0,1], denotes the popularity of this node, eg. number of followers
  - ***Willingness***: [0,1], the willingness of this node to retweet a message
  - ***Credibility***: [-1,1], a **higher (positive) value denotes higher tendency to share true news**, a lower (negative) value denotes higher tendency to share fake news. Purposely define 1% nodes with very high credibility (stands for CDC, FDA…), the rest follow a long-tail distribution (tail on -1).

- Messages: 50, each message starts from one of the top 10 popular users
  - ***Fakeness***: [-1,1], -1 denotes a completely true news, 1 denotes a completely fake news

Distribution of Node Attributes





Network Construction

Q: How does a message spread among users? (how do we create edges between nodes?)

If willingness(j) + popularity(i) - credibility(j) * fakeness(m) >= threshold

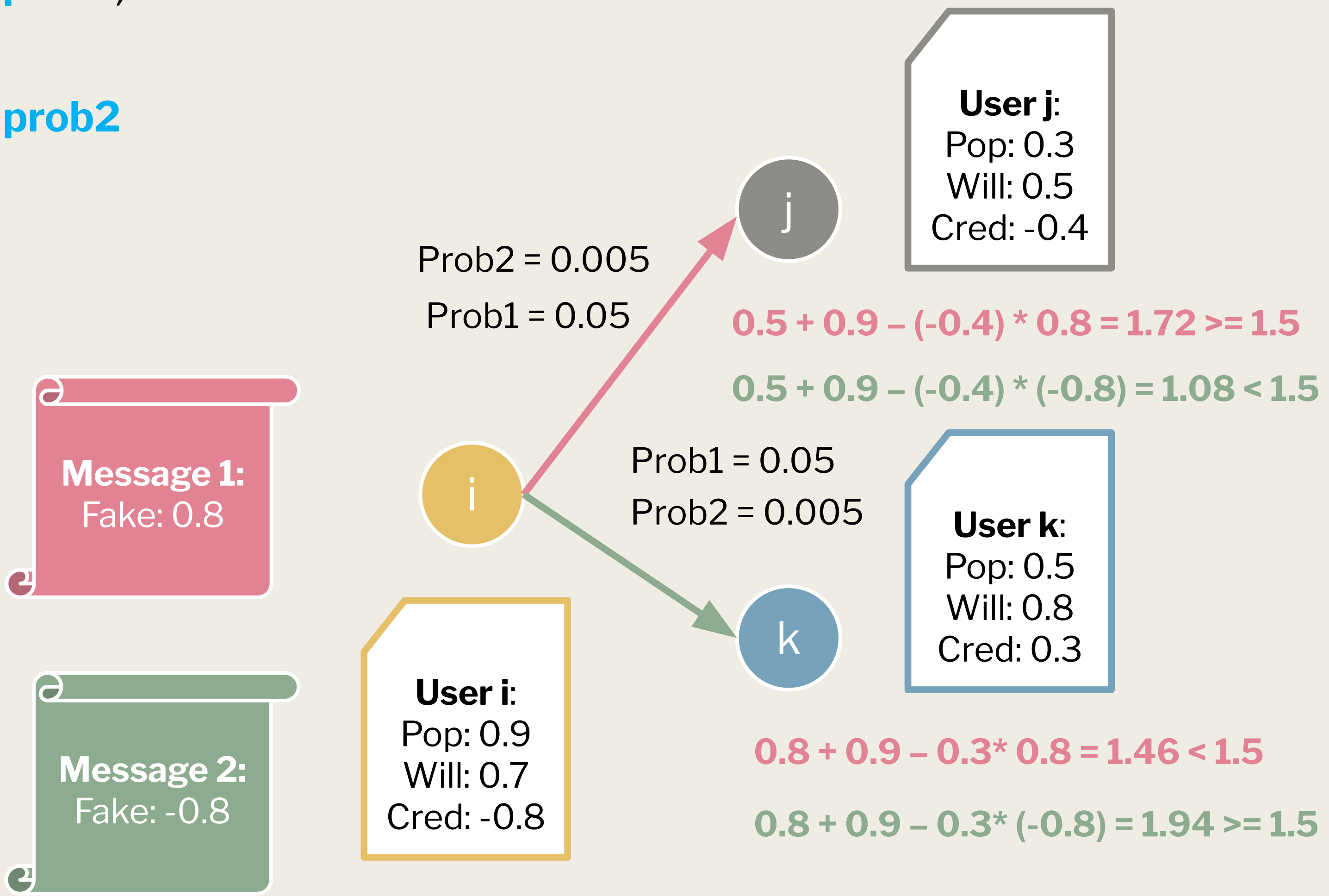
node i -> node j with a probability prob1;

Otherwise,

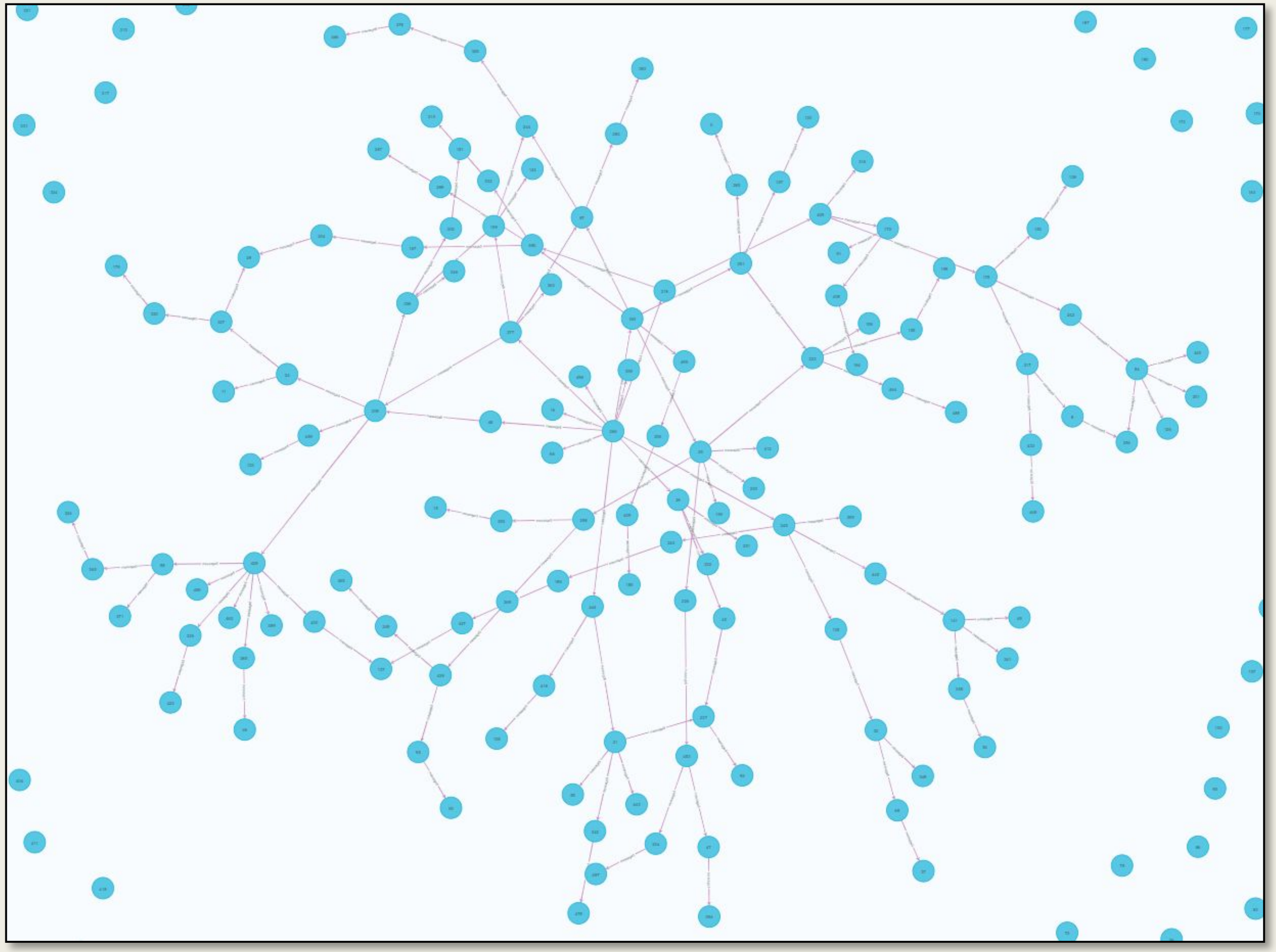
node i -> node j with a probability prob2

Eg. A popular user(i) made up and posted a rumor…

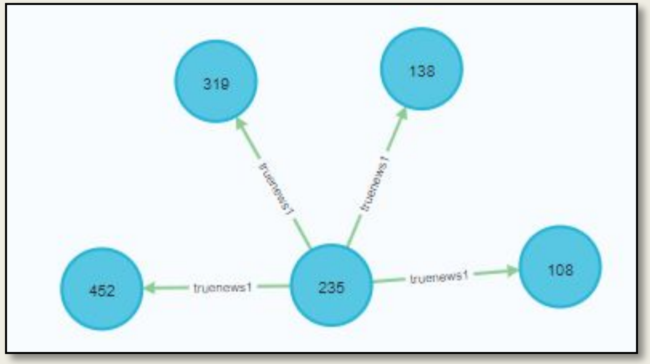
Threshold = 1.5
Prob1 = 0.05
Prob2 = 0.005



Edge generation process



- 50 fake news messages (fakeness in [0,1]) starting from one of the top 10 popular users
- For each news, stop spreading when no new edges can be generated in one iteration
- We also tried adding 3 true news, which didn't spread far in this network

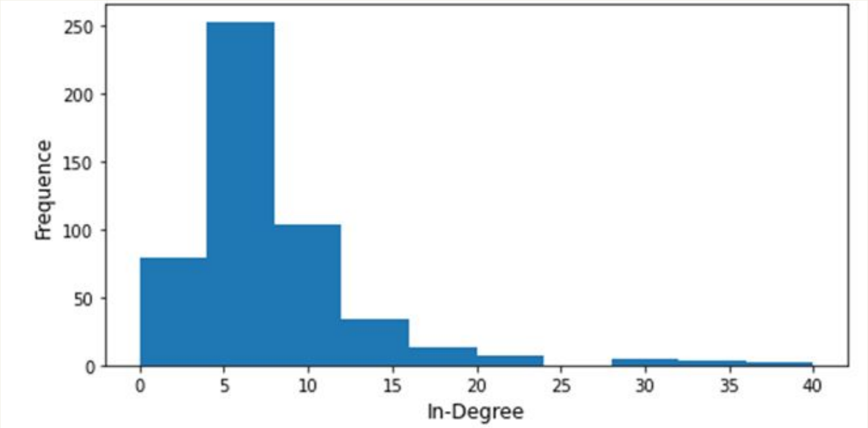


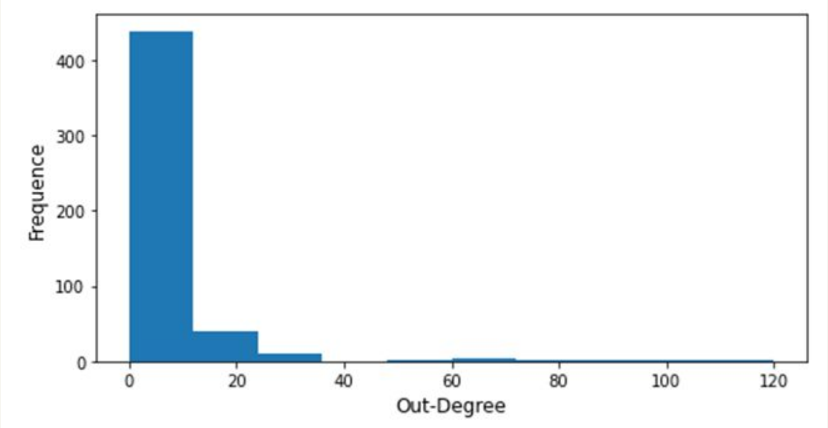
Simulation results

The correlation between different features of the whole retweet network

	willingness	credibility	popularity	inde_agg	outde_agg	in_eigenvector	out_eigenvector	betweenness	influence
willingness	1.000000	0.055650	0.004634	0.690344	0.201964	0.647649	0.223838	0.512264	0.026859
credibility	0.055650	1.000000	-0.015215	-0.267911	-0.068225	-0.233063	-0.071508	-0.112013	0.039558
popularity	0.004634	-0.015215	1.000000	-0.003843	0.679483	0.020513	0.670067	0.385535	0.611816
inde_agg	0.690344	-0.267911	-0.003843	1.000000	0.282686	0.932895	0.302729	0.699396	0.026847
outde_agg	0.201964	-0.068225	0.679483	0.282686	1.000000	0.263692	0.948549	0.718013	0.904479
in_eigenvector	0.647649	-0.233063	0.020513	0.932895	0.263692	1.000000	0.271190	0.664959	0.030347
out_eigenvector	0.223838	-0.071508	0.670067	0.302729	0.948549	0.271190	1.000000	0.718564	0.830836
betweenness	0.512264	-0.112013	0.385535	0.699396	0.718013	0.664959	0.718564	1.000000	0.475861
influence	0.026859	0.039558	0.611816	0.026847	0.904479	0.030347	0.830836	0.475861	1.000000

In and Out Degrees distributions





Exploring the network using a random walk

1. Select ***N*** users as the first generation (G1).

2. Search the retweet messages of G1, and related retweeted users. Then, we can generate the retweet network where G1 users connect with new users.

3. Apart from ***N*** G1 users, select ***M*** new users in the retweet network based on users' in-degree. These ***M*** new users becomes G2 users.

4. Repeat step 2: Search the retweet messages of G1 and G2, and related retweeted users. Then, select ***M*** new users among the new users as G3 users.

5. With the same logic, we can create ***K*** generations.

6. Finally, we search the re-tweet messages of all ***K*** generations' users, and related retweeted new users.

Snowball sampling to explore the graph

Simulation data:
- 500 users
- 3660 links
- 50 different messages

Exploration:
- Randomly select 10 users as G1
- Adding new 20 users ***<u>who has most in-degree</u>*** as the next generation users
- Total 5 generations (total 90 users)

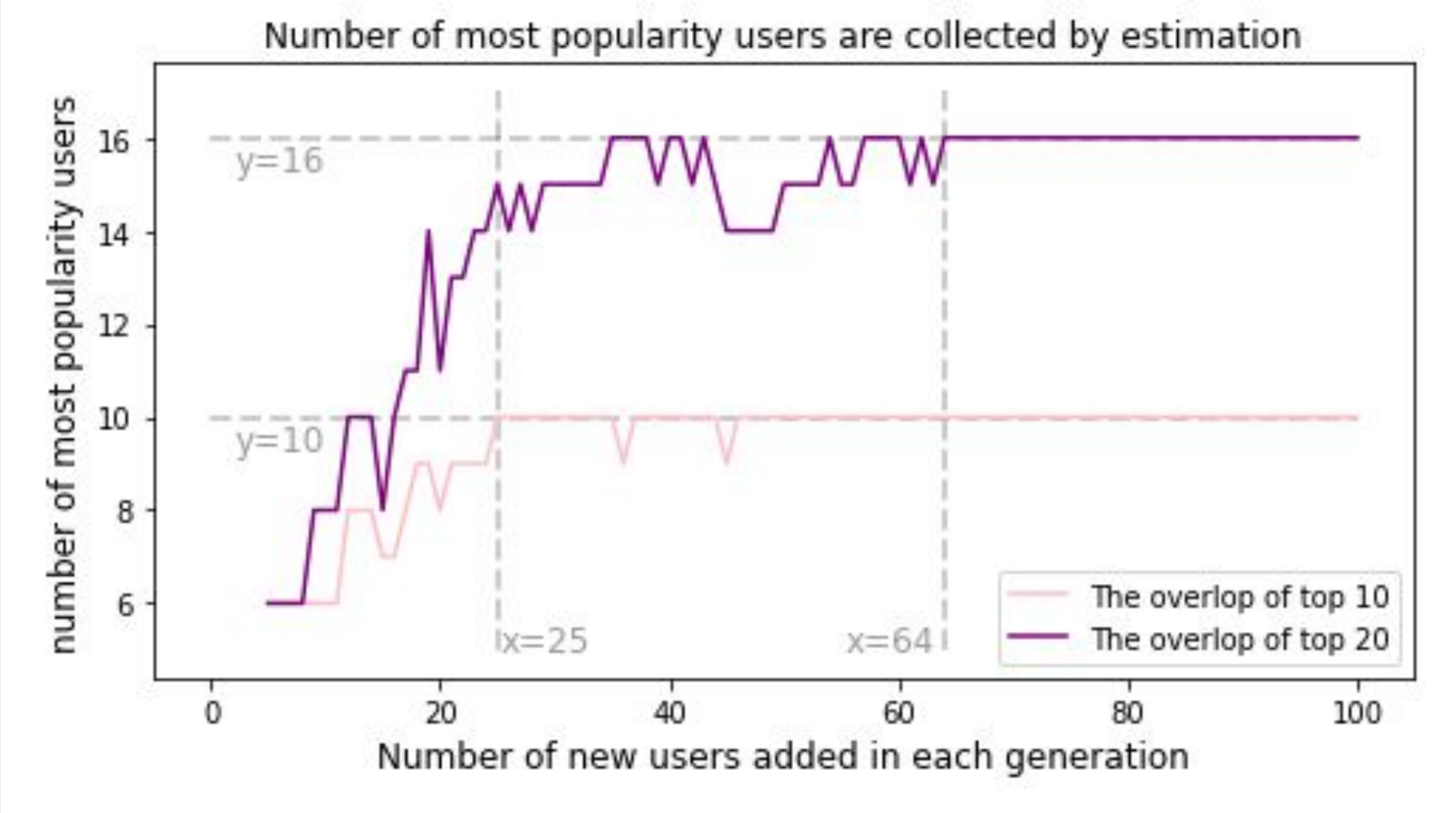
Sensitivity analysis

The impact of the <u>number of new added users in each generation</u> on the <u>number of high poluarity users found.</u>

Exploration:

① Randomly select 10 users as G1. ②Adding new ***X*** users in each generation. ③ Total 5 generations (total 10 + 4***X*** users).

Results:



The top 10 popularity users are always can be found.

The most number of top 20 popularity users can be found is 16.

Summary

- It is difficult to investigate misinformation spreading on social media using the real-world data due to lack of access to high volume, high velocity data.
- We proposed a **link analysis method** for finding and ranking influencers
    a. *Start with a seed list of influencers and snowball sample to iteratively grow the sample*
    b. *Create a model representing three properties of agents (popularity, willingness, credibility) and generate graphs by spreading 50 messages in a simulation*
    c. *Rank influencers by the nodes' in-degree*
        - *Simulation found the top 10 and 16 out of top 20 popular nodes*
- A estimator for influence:
    - ***Influence = (in-degree/out-degree) * total-degree * generation*100***

Future Direction

- Estimate the Influence and Credibility values for all nodes in the simulation
- Re-apply agent based model method on analyze empirical networks to automate/augment social network analysis with heuristics:
  a. What cutoff (in-degree threshold) should be used for adding users into the sample?
  b. How to identify high credibility actors and spawn a subgraph/community?
- Build a data pipeline for seeding, sampling, capturing, and visualizing temporal graphs of any controversial discourse on social media in real time