Network Analysis to Discover Emerging Influencers & Superspreaders

TRUSTCON 2024

July 22, 2024

Hong Qu

Slides: https://tinyurl.com/superspreader2024

Hong Qu

hong_qu@hks.harvard.edu

- Adjunct lecturer at Harvard Kennedy School
- BA from Wesleyan University
- MS from UC Berkeley's School of Information
- YouTube startup engineer/designer
- Fellow at Stanford Human-Centered AI 2020-21
- PhD student in Network Science at Northeastern University (2020-202?): Political Communications

Network Analysis vs Network Science

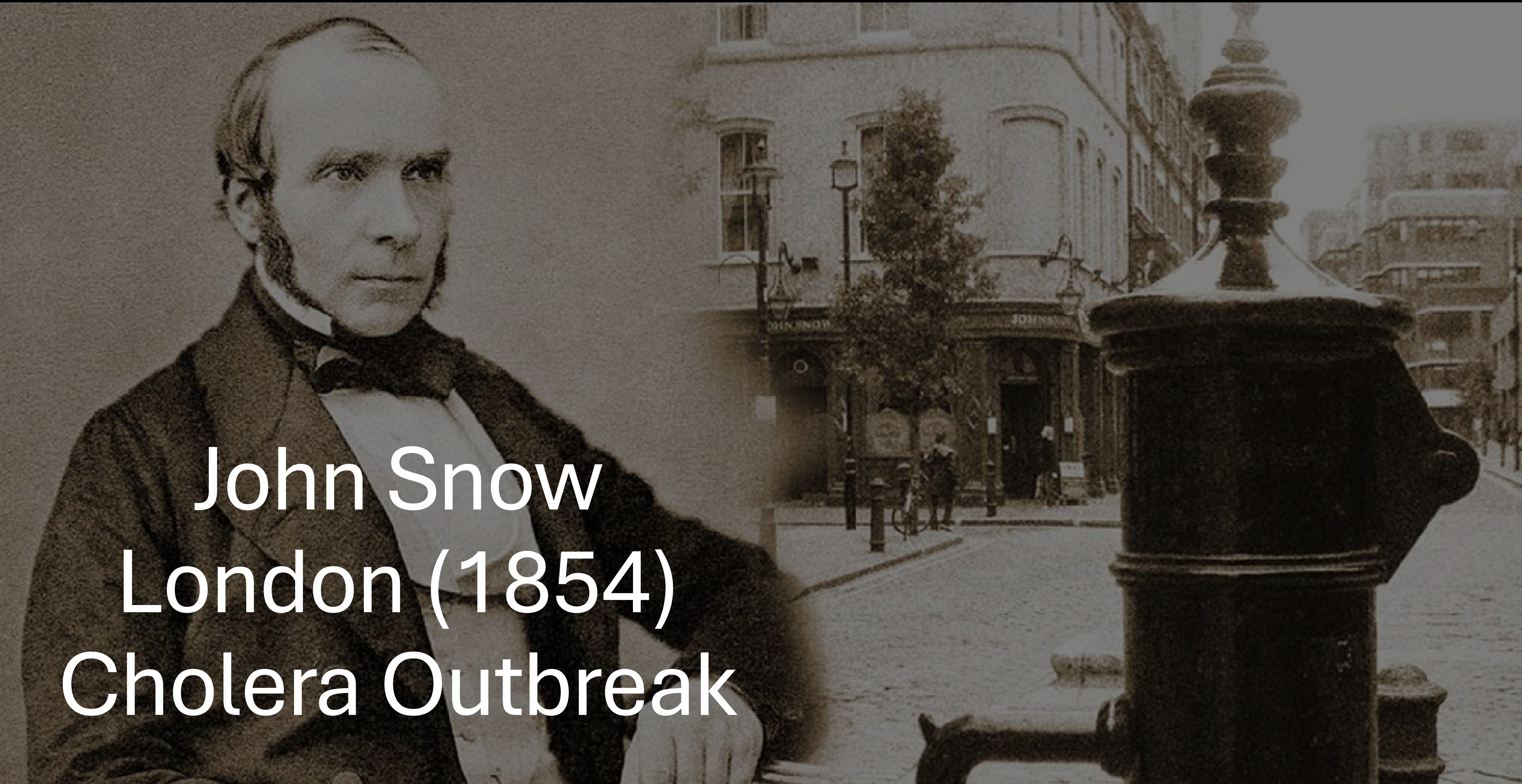
Prompt:
What is the difference between network analysis and network science?

ChatGPT response:

- Definition: Network science is an interdisciplinary field that studies complex networks in a holistic manner, encompassing theory, methods, and applications from various disciplines.
- Focus: It aims to understand the universal principles governing the structure, dynamics, and function of networks across different domains.
- Methods: It incorporates techniques from physics, mathematics, computer science, biology, and social sciences. It includes advanced modeling, simulations, and empirical studies.
- Applications: Network science applies to a broader range of networks, including social, biological, technological, informational, and more. It seeks to find commonalities and differences between these types of networks.

In summary, while network analysis is a crucial part of network science, the latter is a broader and more integrative field that seeks to uncover general principles applicable to various types of networks.

https://chatgpt.com/share/34c62fb4-bab0-497b-81c4-a8c099df4909

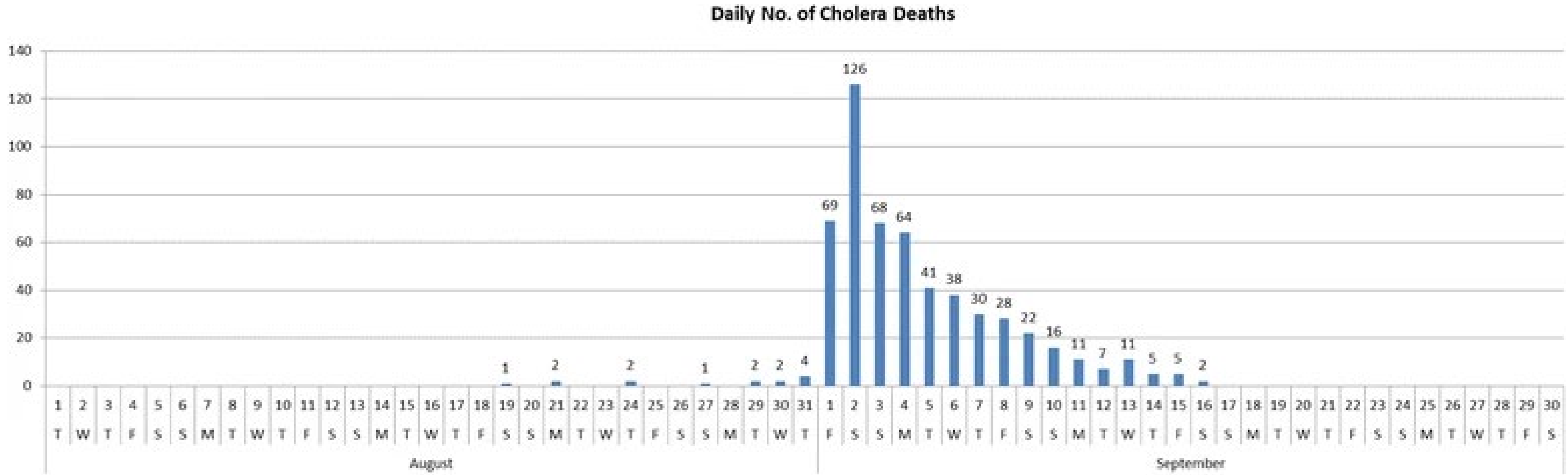

John Snow
London (1854)
Cholera Outbreak

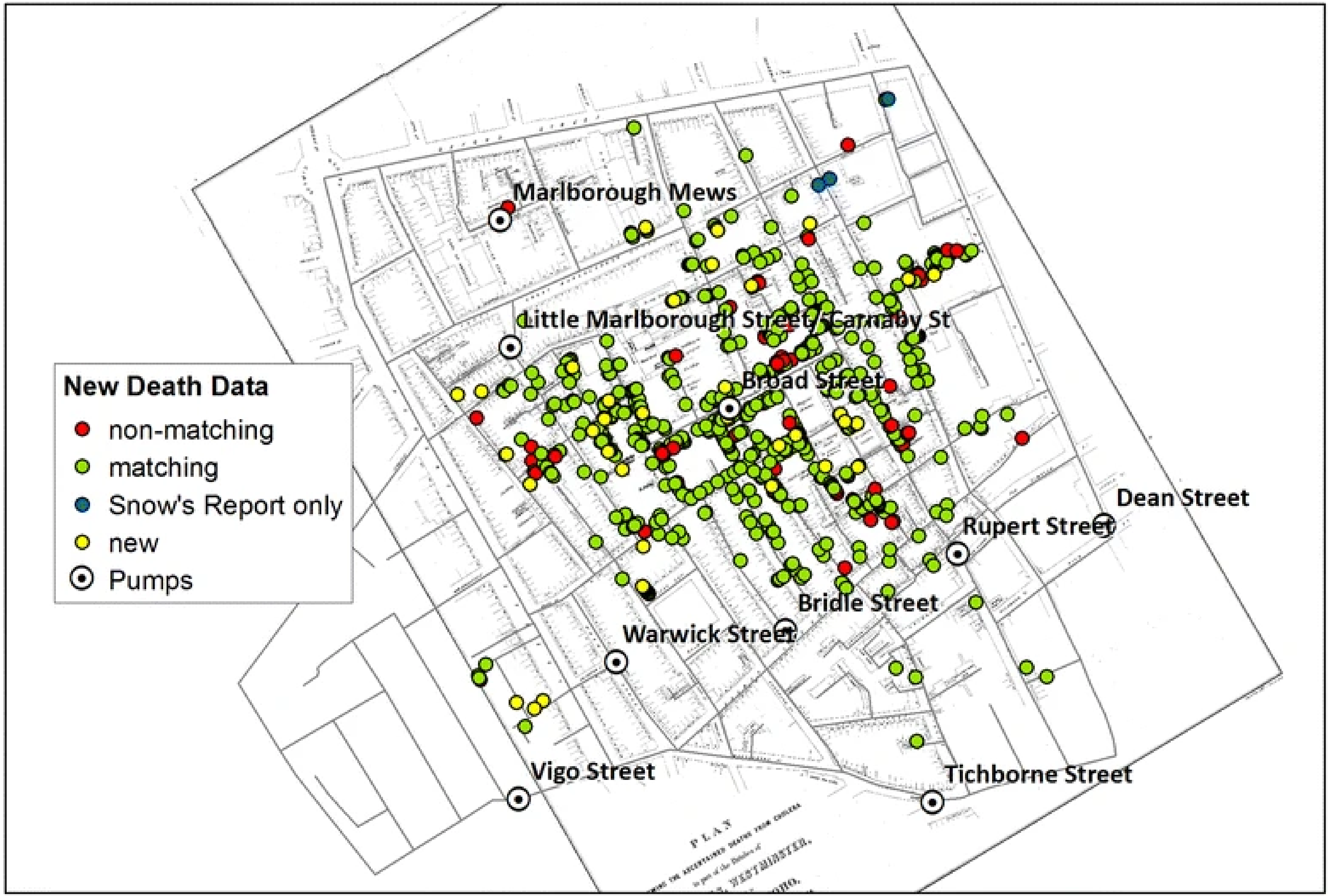
The mortality rates and the space-time patterns of John Snow's cholera epidemic map

Narushige Shiode[1*†], Shino Shiode[2†], Elodie Rod-Thatcher[2†], Sanjay Rana[2] and Peter Vinten-Johansen[3]

Abstract

Background: Snow's work on the Broad Street map is widely known as a pioneering example of spatial epidemiology. It lacks, however, two significant attributes required in contemporary analyses of disease incidence: population at risk and the progression of the epidemic over time. Despite this has been repeatedly suggested in the literature, no systematic investigation of these two aspects was previously carried out. Using a series of historical documents, this study constructs own data to revisit Snow's study to examine the mortality rate at each street location and the space-time pattern of the cholera outbreak.



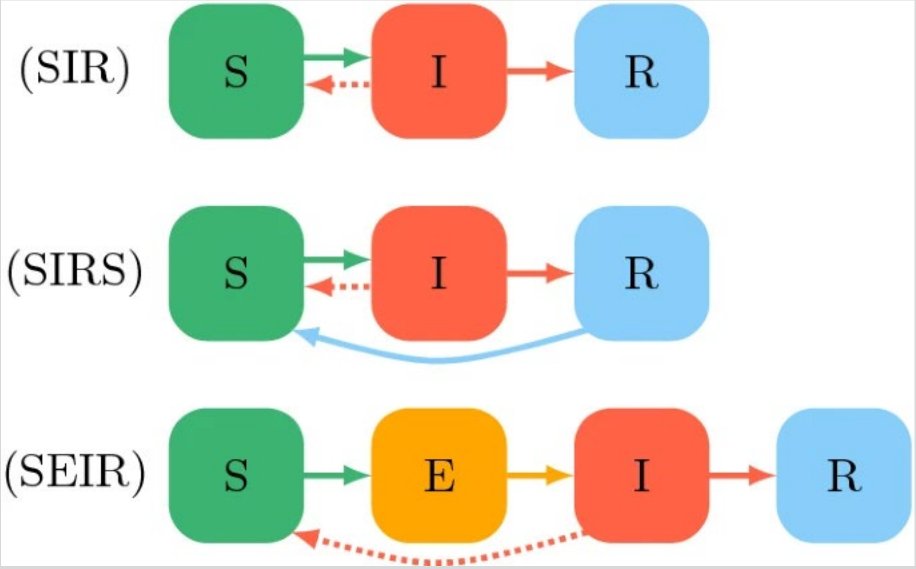

Marlborough Mews
Little Marlborough Street/Carnaby St
Broad Street
Dean Street
Rupert Street
Bridle Street
Warwick Street
Vigo Street
Tichborne Street
New Death Data
non-matching
matching
Snow's Report only
new
Pumps

Epidemiology

Discussion:

How might we adapt methods from epidemiology to study the **spread of ideas**?

- Are ideas more contagious?
- What is the difference in time scale?
- What is the difference in reach?
- Are there more superspreader?
- How do people build up immunity?
- Difficult to collect data on opinion/belief/stance
    - Diagnose someone's condition/state as:
    - S - susceptible
    - E- exposed
    - I - infected
    - R - recovered



Network **Topology**

Relationships are represented by nodes connected by edges.

For example:

- On Twitter, **follower - following** relationships
- During COVID, all your family, friends, coworkers, people you see regularly

Network **Dynamics**

The flow of “stuff” on a network over time.

For example:

- On Twitter, **retweeting** breaking news on Twitter
- During COVID, contact tracing the people an infected person came in contact with

Pagerank is network topology algorithm

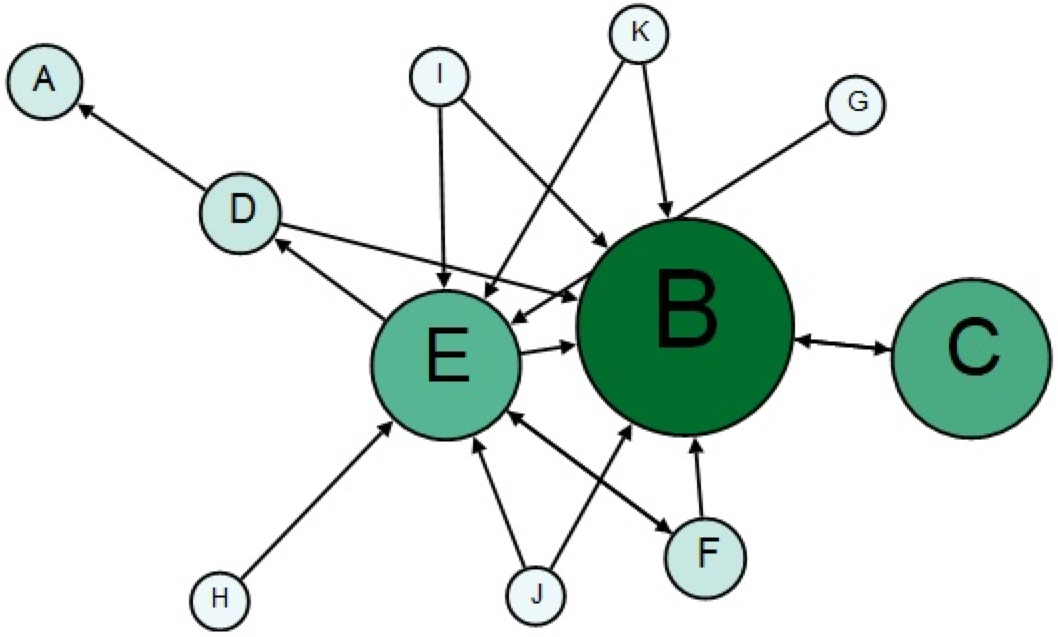
The PageRank Citation Ranking: Bringing Order to the Web

January 29, 1998

Abstract

The importance of a Web page is an inherently subjective matter, which depends on the readers interests, knowledge and attitudes. But there is still much that can be said objectively about the relative importance of Web pages. This paper describes PageRank, a method for rating Web pages objectively and mechanically, effectively measuring the human interest and attention devoted to them.

We compare PageRank to an idealized random Web surfer. We show how to efficiently compute PageRank for large numbers of pages. And, we show how to apply PageRank to search and to user navigation.



[How Does the PageRank Algorithm Work, Exactly?]

The Anatomy of a Large-Scale Hypertextual Web Search Engine

Sergey Brin and Lawrence Page

Difference between web links and social media

Discussion:
How could we use Pagerank to uncover superspreaders in social media?

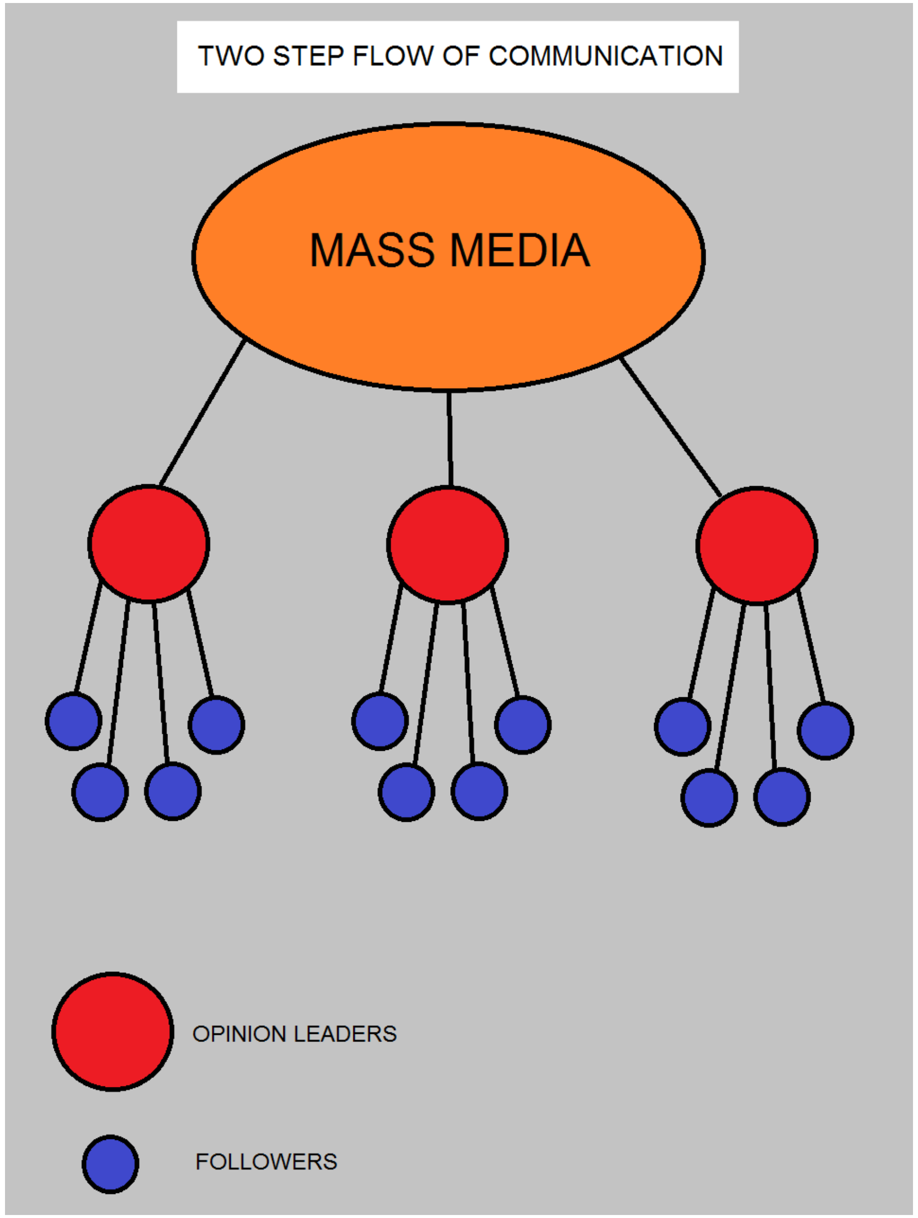
- What is the unit of analysis (aka node)?
- What is the relationship or interaction (aka link)?
- What is the time scale?

How Social Media Facilitates the Spread of Misinformation

Python code for agent-based model simulation:
https://colab.research.google.com/drive/1l31OXHH6sMWXDDd98uOdeu-xf6tL2b2q#scrollTo=9r5JRWgrztNe

FDR's Joint Address to Congress Leading to a Declaration of War Against Japan (1941)

On December 8, 1941 Roosevelt addressed a joint session of Congress and, via radio, the nation.



Two-step Flow of Communication Model, by Lazarsfeld and Katz (1955)

Generating synthetic networks using NetworkX

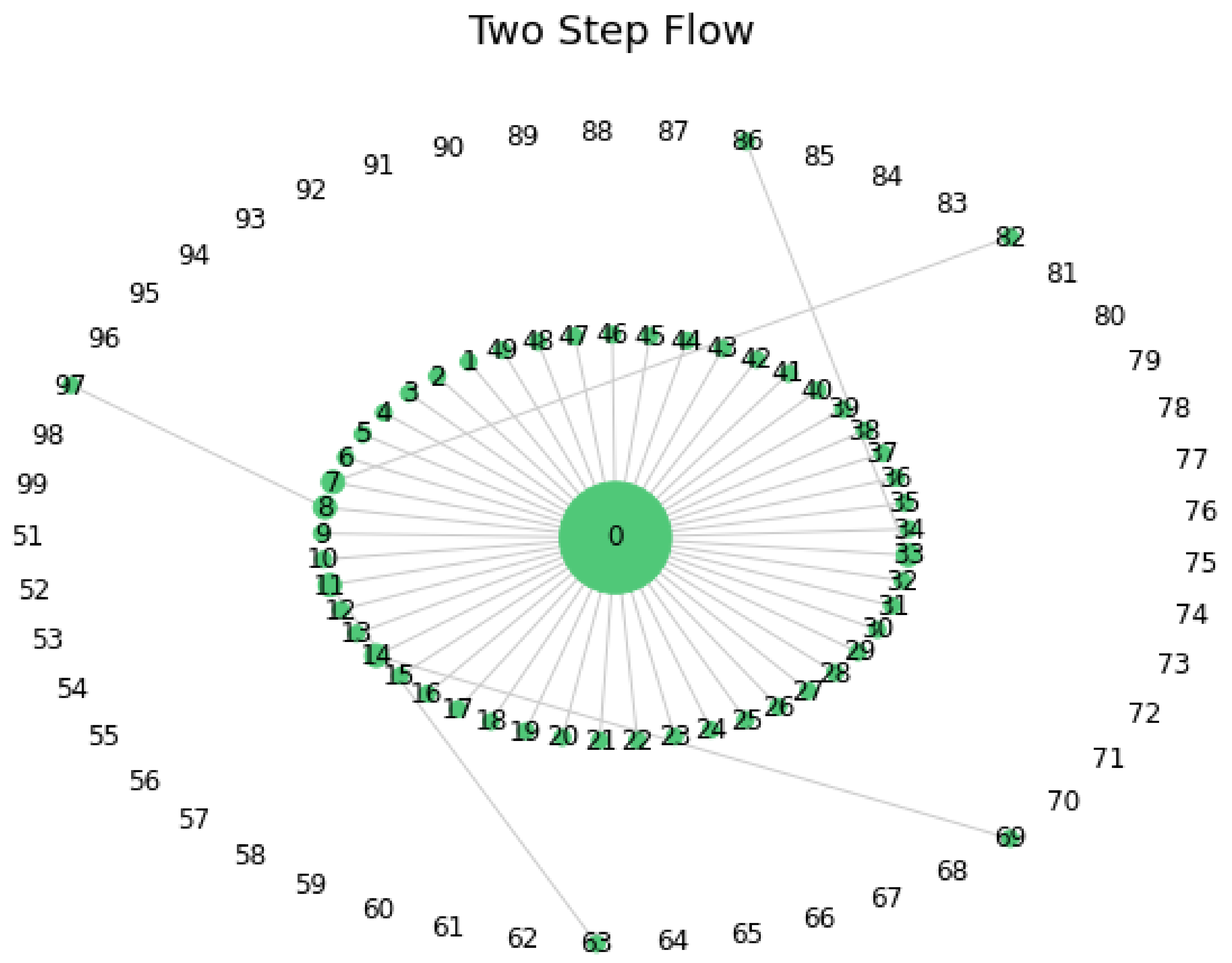
Python code:
[Generating Network Graphs]

```python
# prompt: Generate a scale free network with 1000 nodes and plot it, representating the nodes' degree as size

# Generate a scale-free network with 1000 nodes
G_scale_free = nx.barabasi_albert_graph(1000, 3)

# Calculate node degrees
degree_dist = dict(G_scale_free.degree)

# Plot the graph with node sizes representing degrees
plt.figure(figsize=(12, 8))
nx.draw(G_scale_free, edge_color='#CCCCCC', node_size=[v * 5 for v in degree_dist.values()],
        node_color='#50C878', with_labels=False)
plt.title('Scale Free Graph', size=18)
plt.show()
```

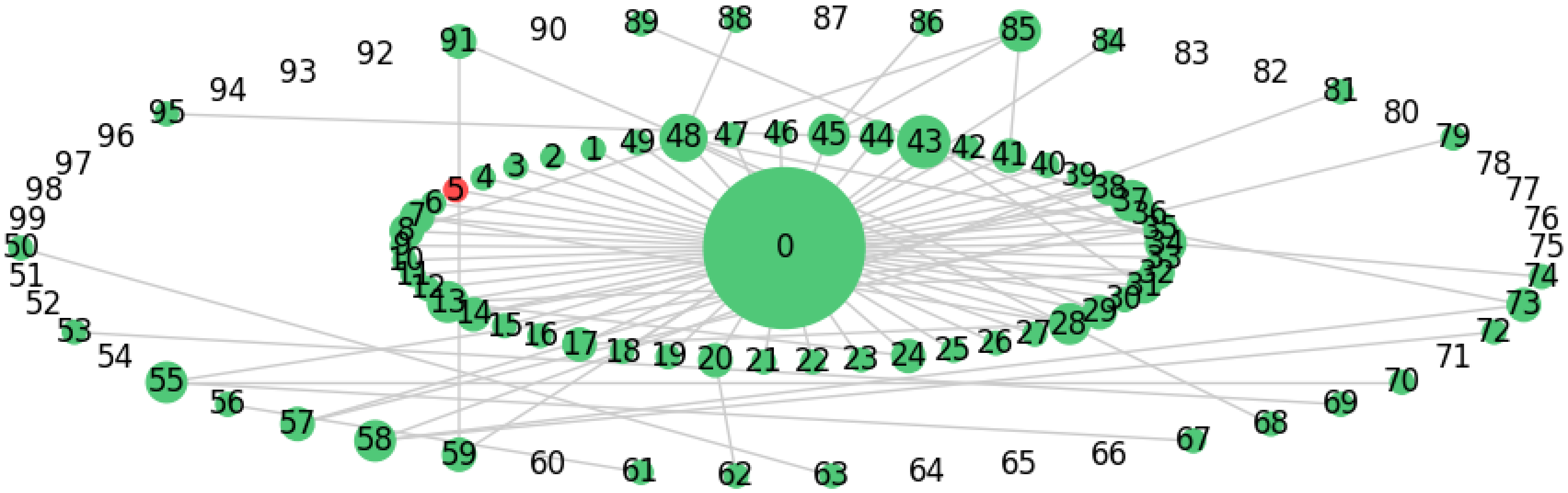
Agent based modeling using NetworkX

Python code:
agent-based model simulation of epidemiological spread of information on different network models

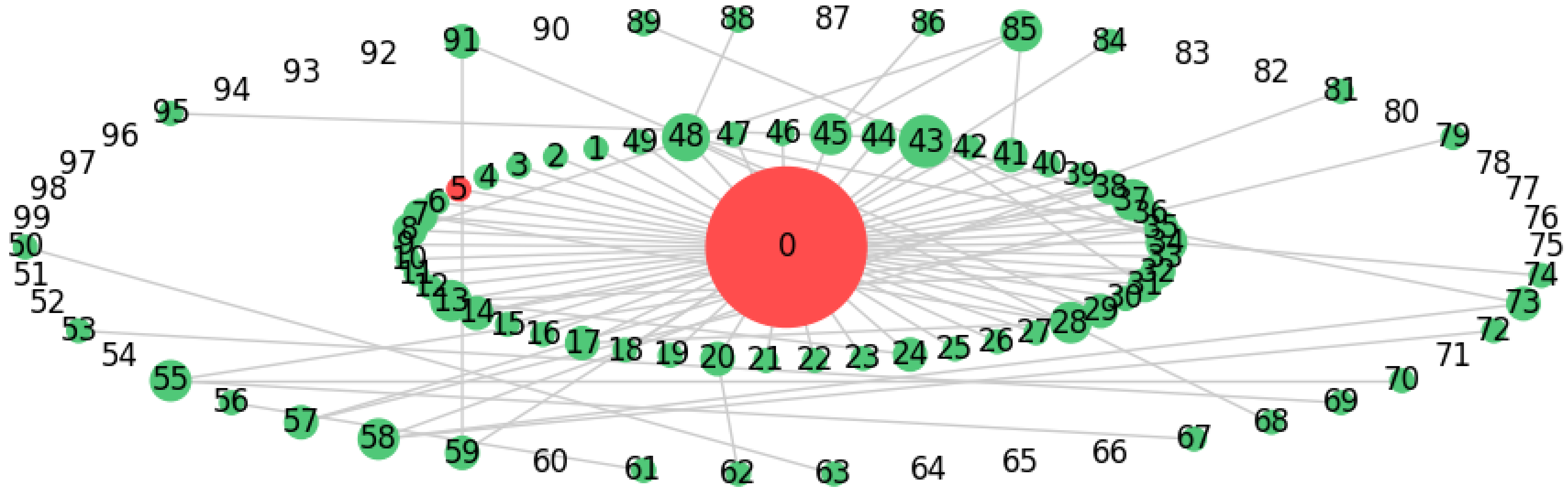
Broadcast with 2-step flow of communication



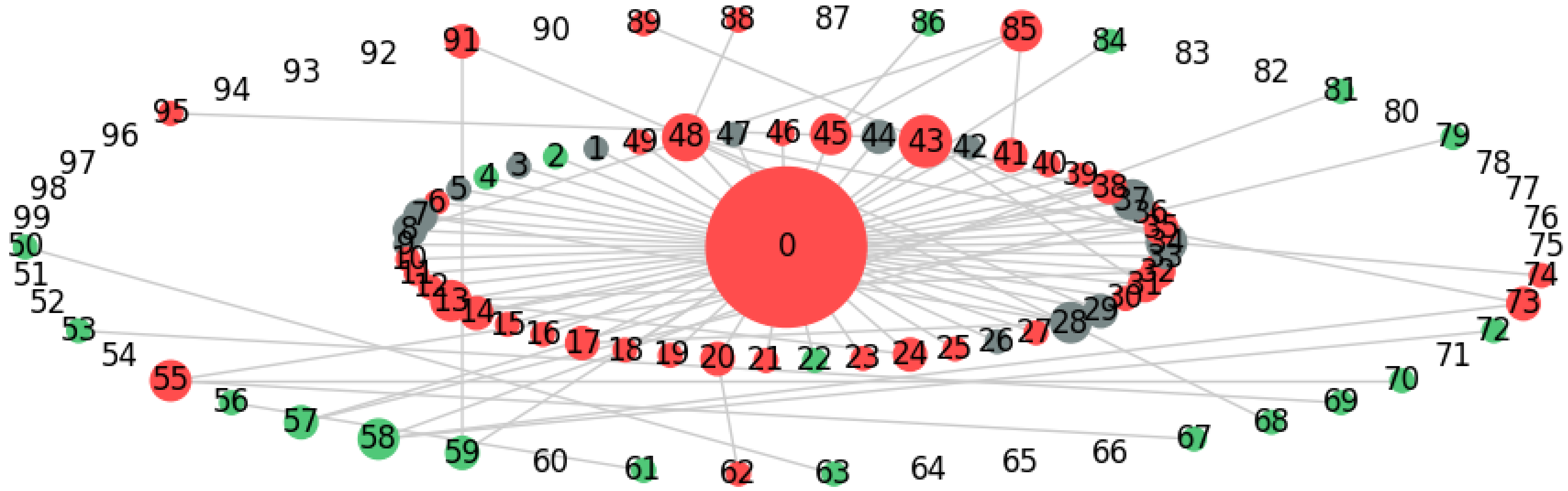
Nodes Status at Time = 0



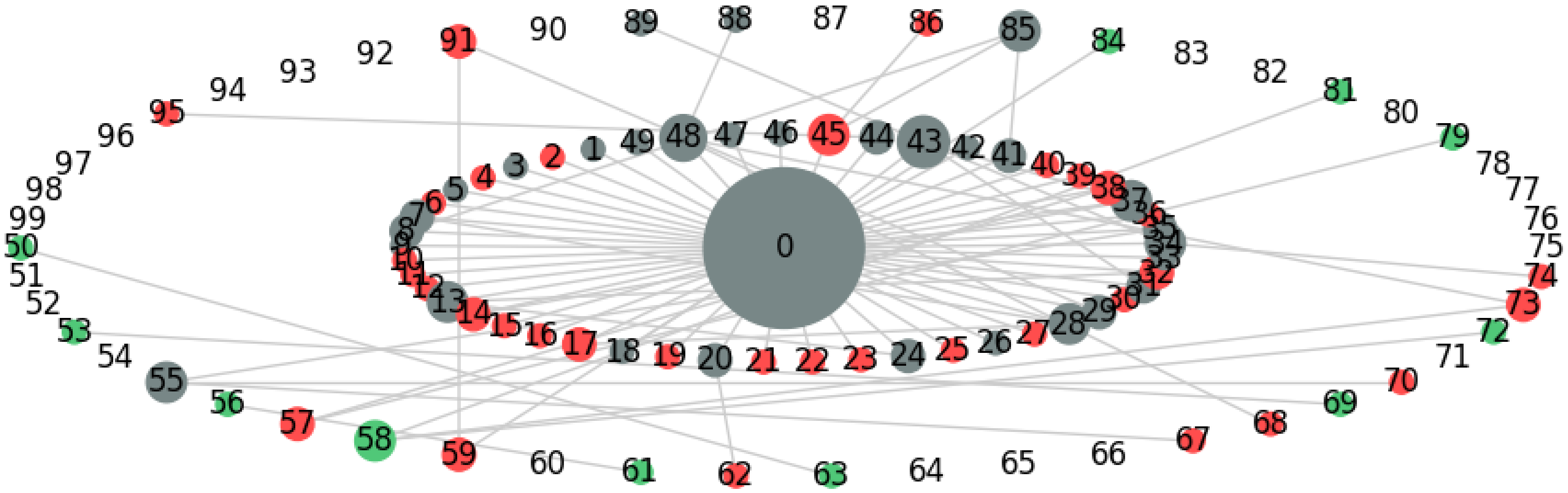
Nodes Status at Time = 1



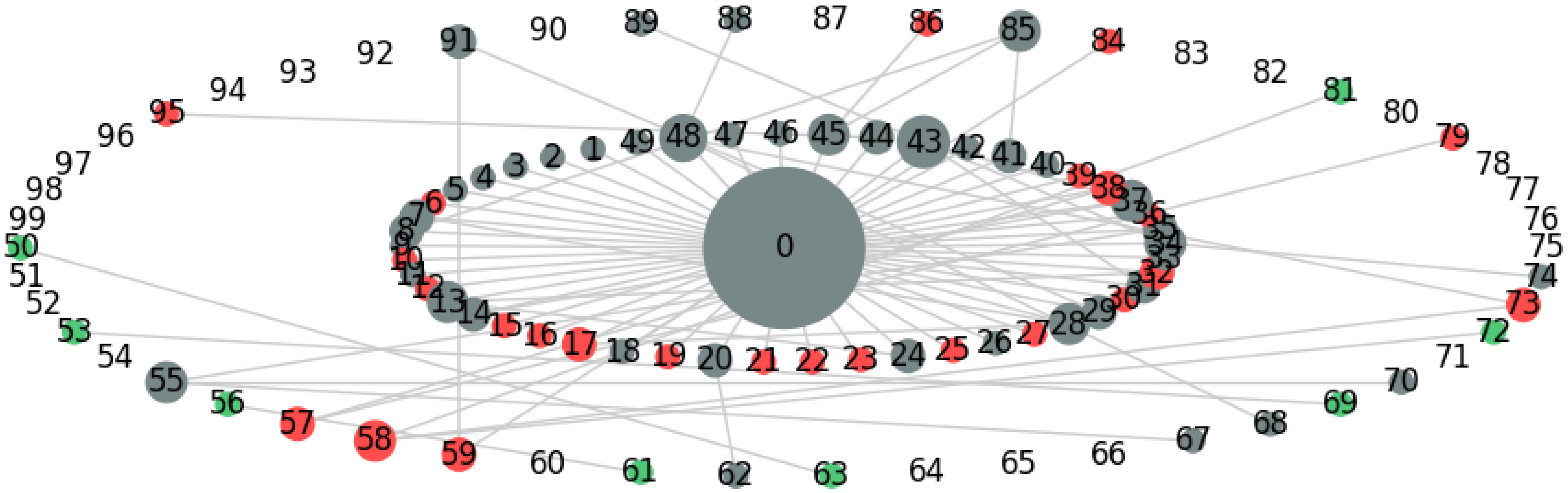
Nodes Status at Time = 3



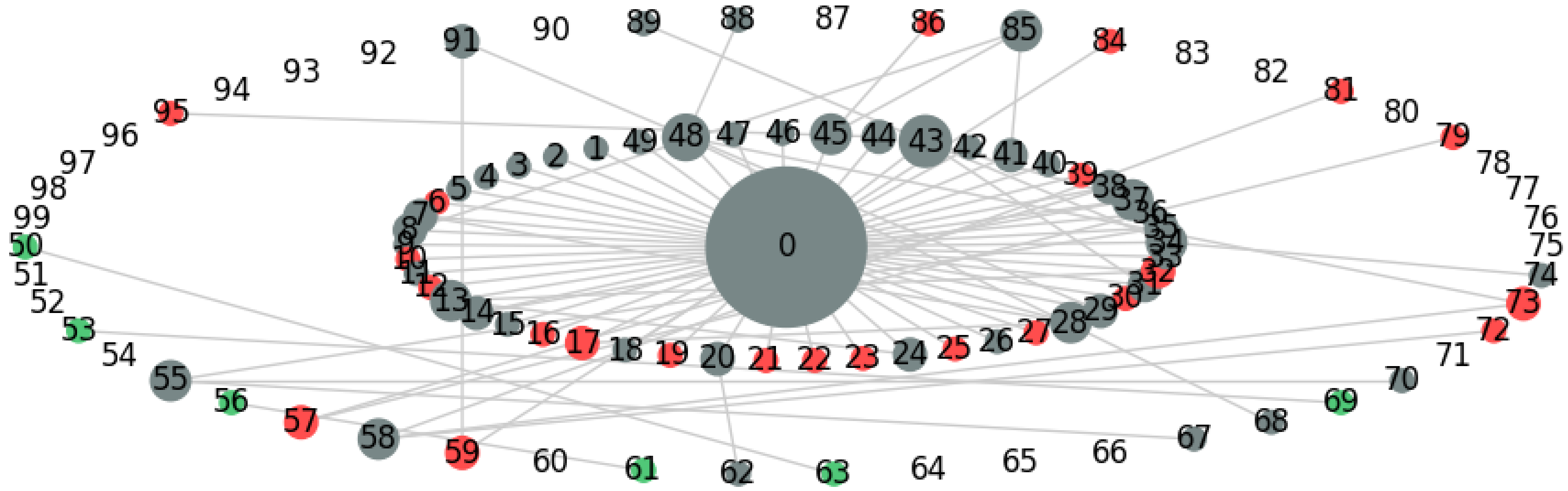
Nodes Status at Time = 4



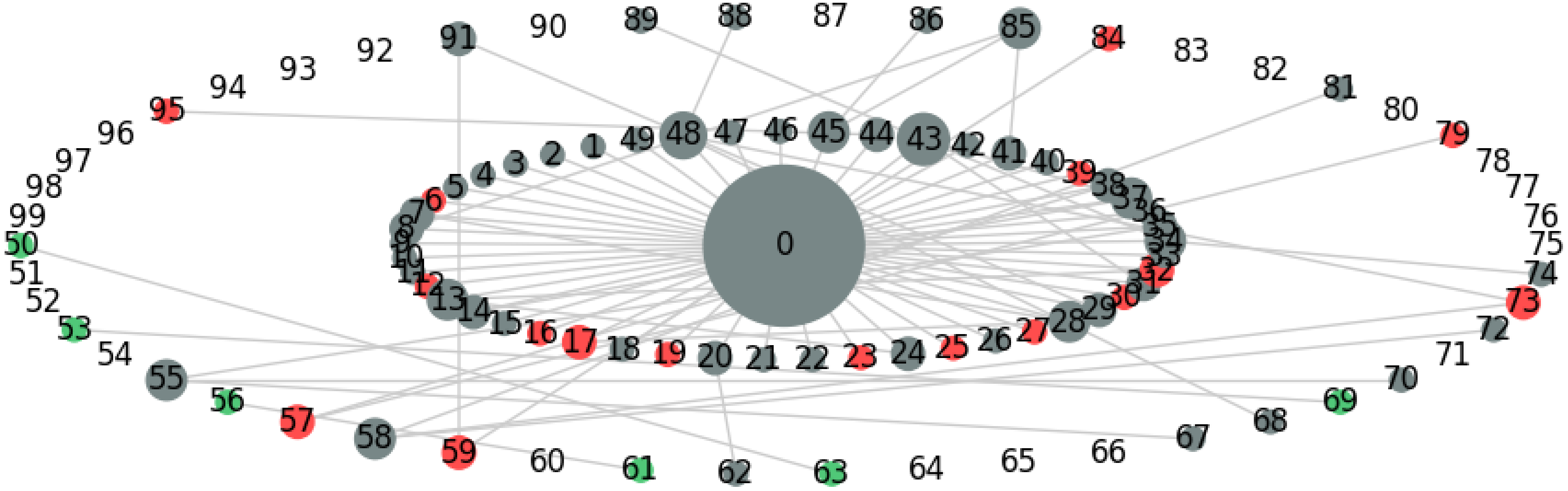
Nodes Status at Time = 5



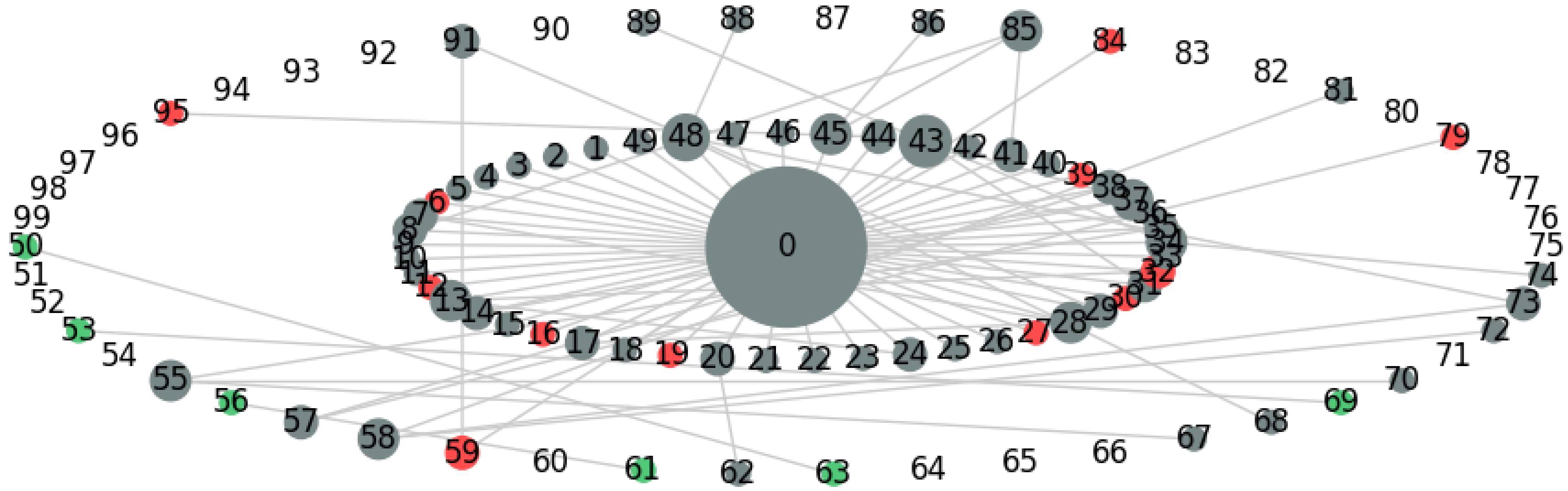
Nodes Status at Time = 6



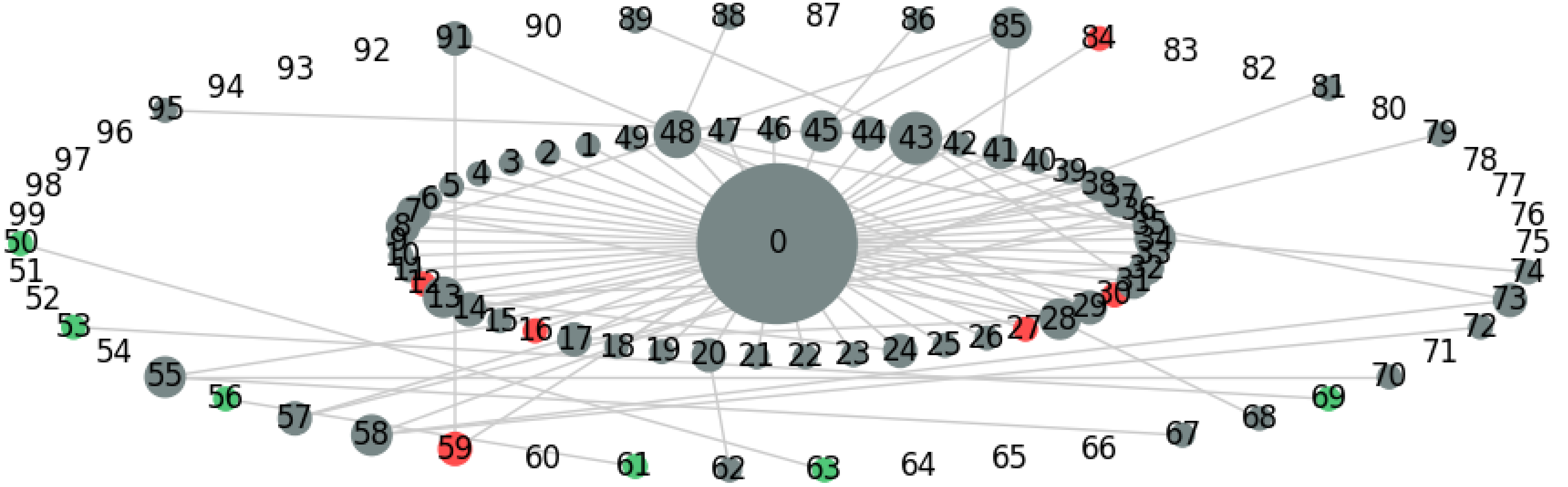
Nodes Status at Time = 7



Nodes Status at Time = 8



Nodes Status at Time = 9



Small World Networks

What happens after the new has been broadcasted?

How does the news information diffusion through communities?

Small World Network

"Networks are said to have a Small World property if any two nodes are connected by a short number of hops, typically less than six. This means it is possible to cross the network very quickly." – Future Learn

Letter | Published: 04 June 1998

Collective dynamics of 'small-world' networks

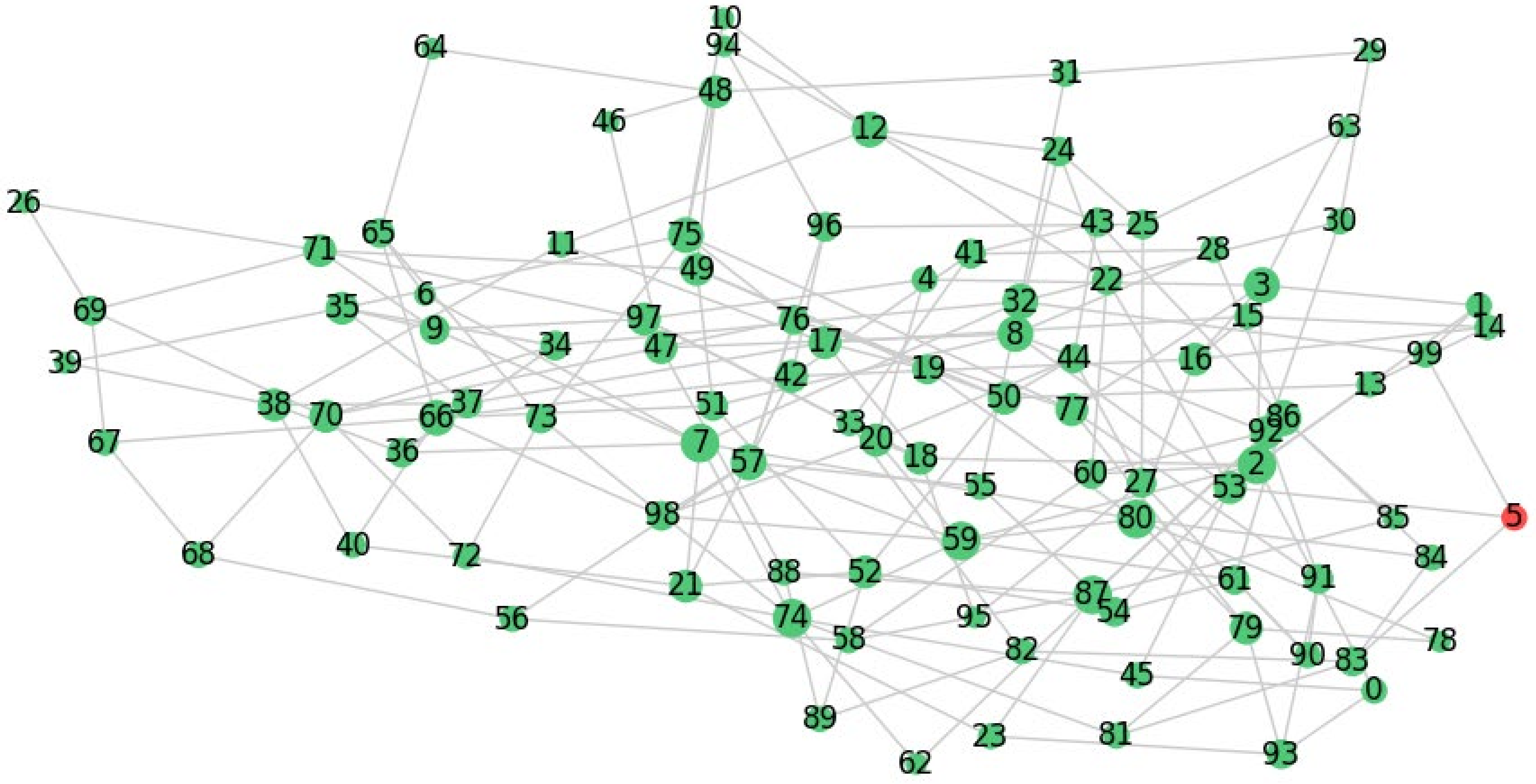
Duncan J. Watts & Steven H. Strogatz

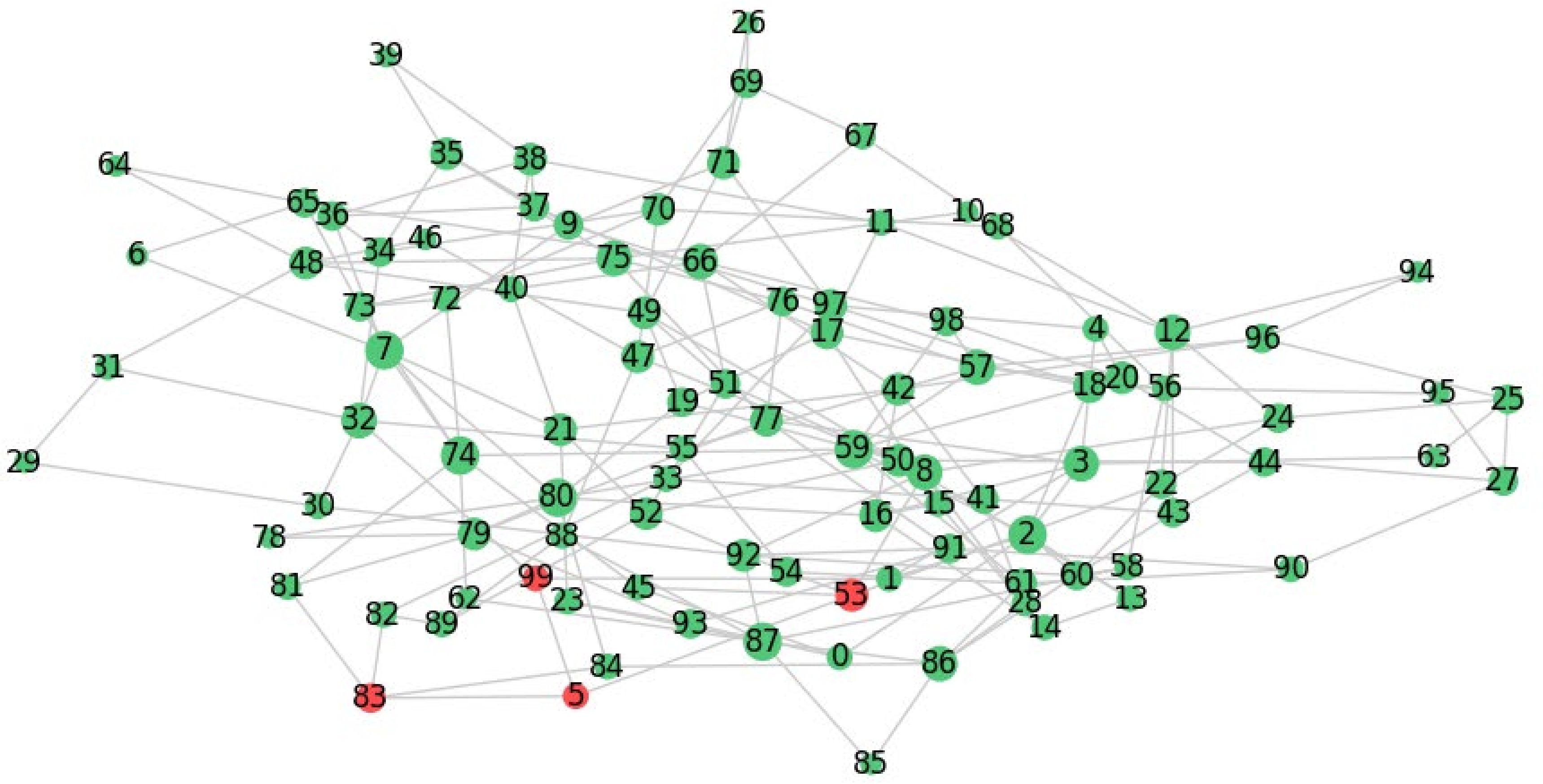
Nature **393**, 440–442 (1998) | Cite this article

Abstract

Networks of coupled dynamical systems have been used to model biological oscillators[1,2,3,4], Josephson junction arrays[5,6], excitable media[7], neural networks[8,9,10], spatial games[11], genetic control networks[12] and many other self-organizing systems. Ordinarily, the connection topology is assumed to be either completely regular or completely random. But many biological, technological and social networks lie somewhere between these two extremes. Here we explore simple models of networks that can be tuned through this middle ground: regular networks 'rewired' to introduce increasing amounts of disorder. We find that these systems can be highly clustered, like regular lattices, yet have small characteristic path lengths, like random graphs. We call them 'small-world' networks, by analogy with the small-world phenomenon[13,14] (popularly known as six degrees of separation[15]). The neural network of the worm *Caenorhabditis elegans*, the power grid of the

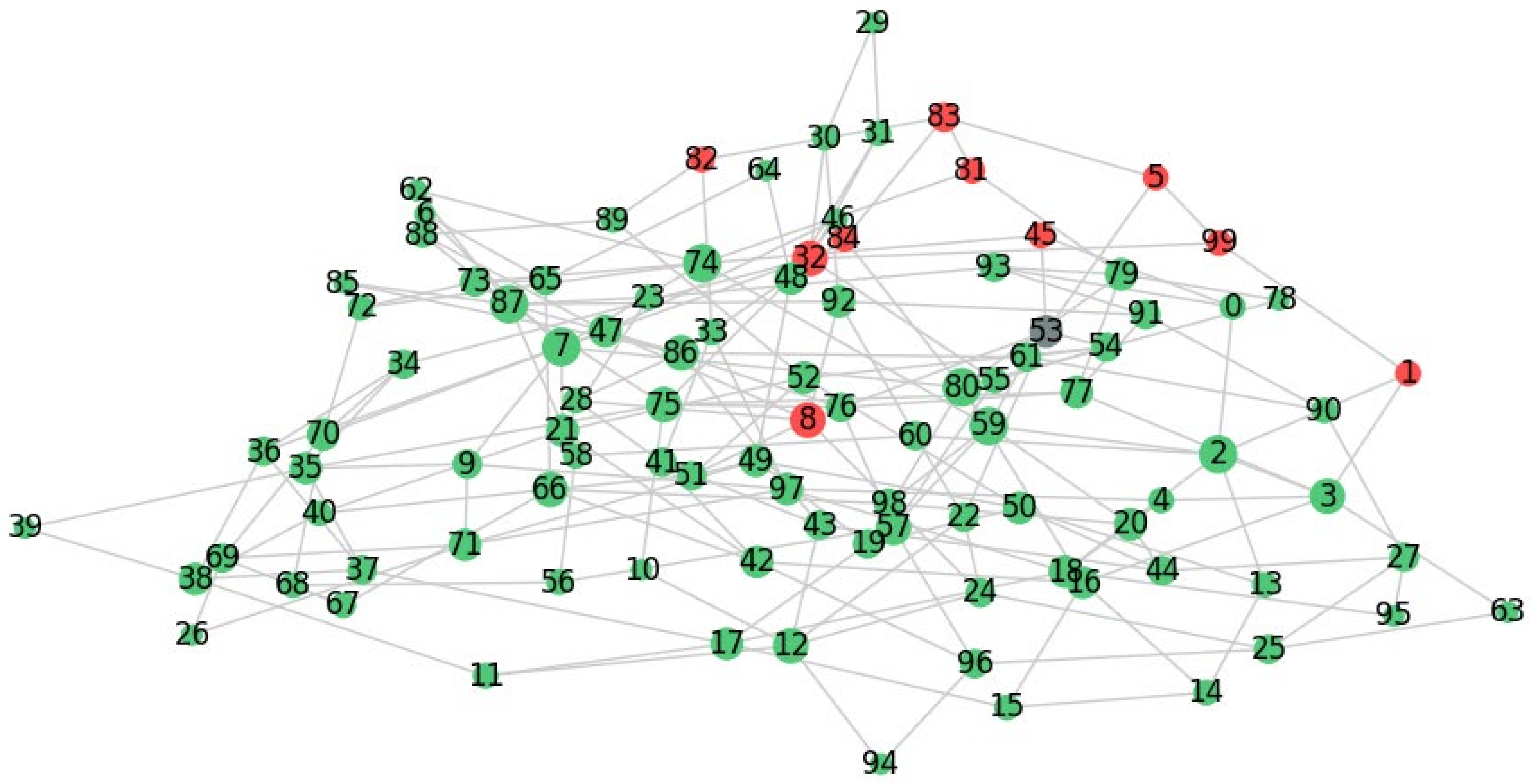
Nodes Status at Time = 0

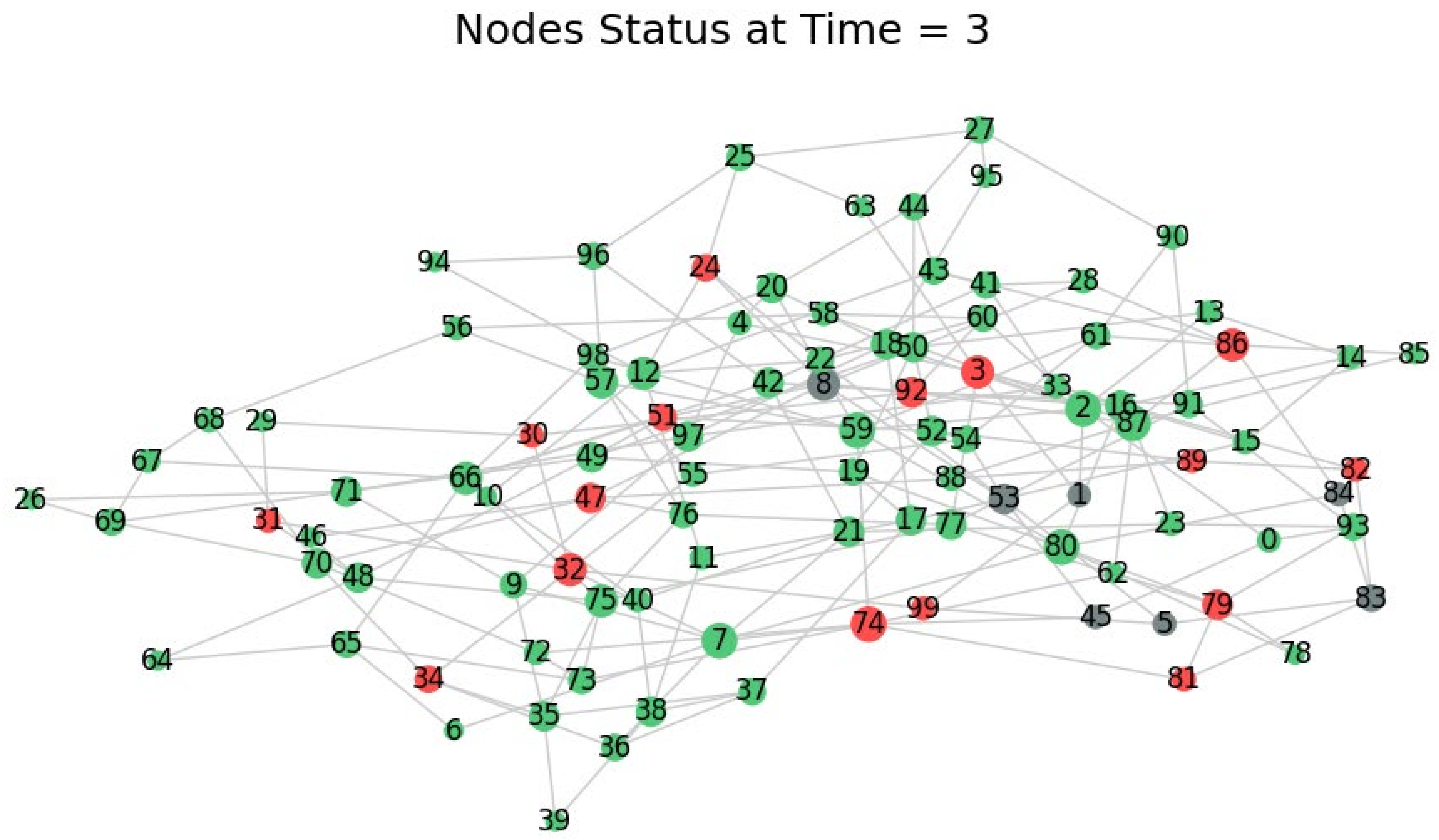




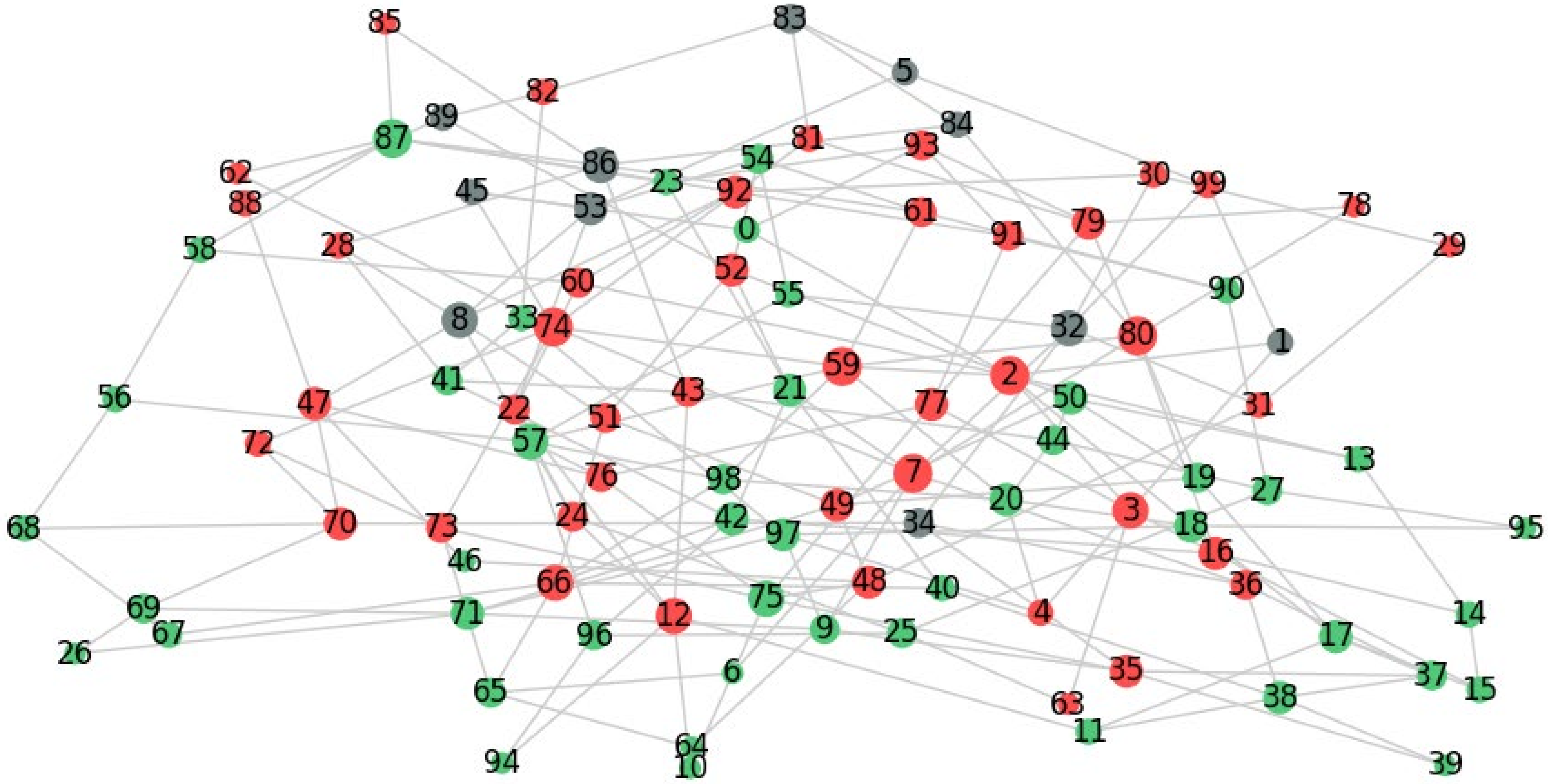
Nodes Status at Time = 1

Nodes Status at Time = 2

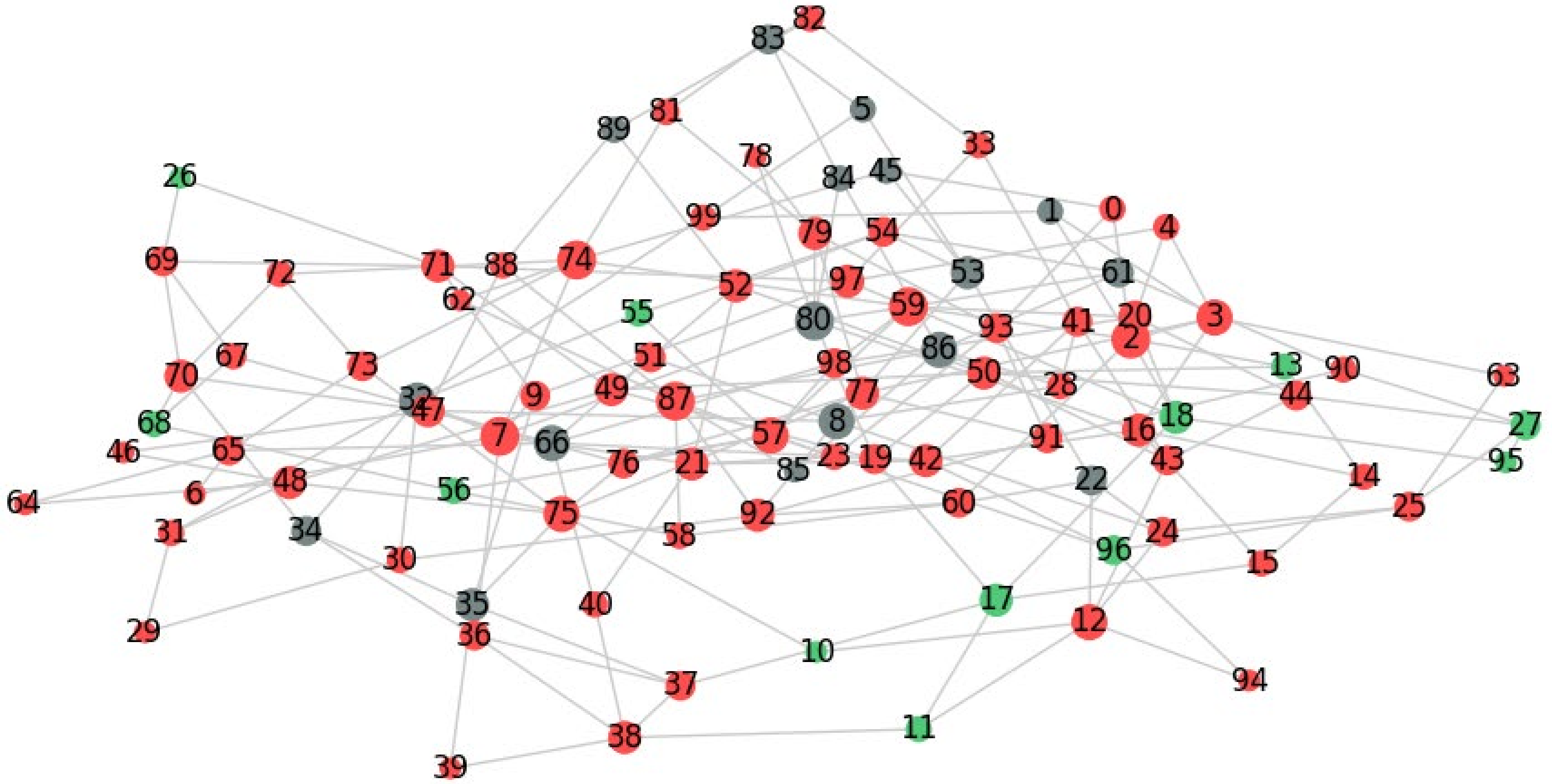



Nodes Status at Time = 3
27
25
95
63
44
90
94
96
24
20
43
41
28
13
58
60
4
56
61
86
14
85
18
50
98
12
22
3
57
42
8
92
33
2
16
91
87
68
29
51
30
97
59
52
54
15
67
49
89
82
55
19
66
88
26
71
10
47
53
1
84
69
31
76
17
77
23
46
21
80
0
93
70
48
11
32
62
9
75
40
99
79
83
45
5
74
7
65
72
78
64
34
73
81
37
38
35
6
36
39

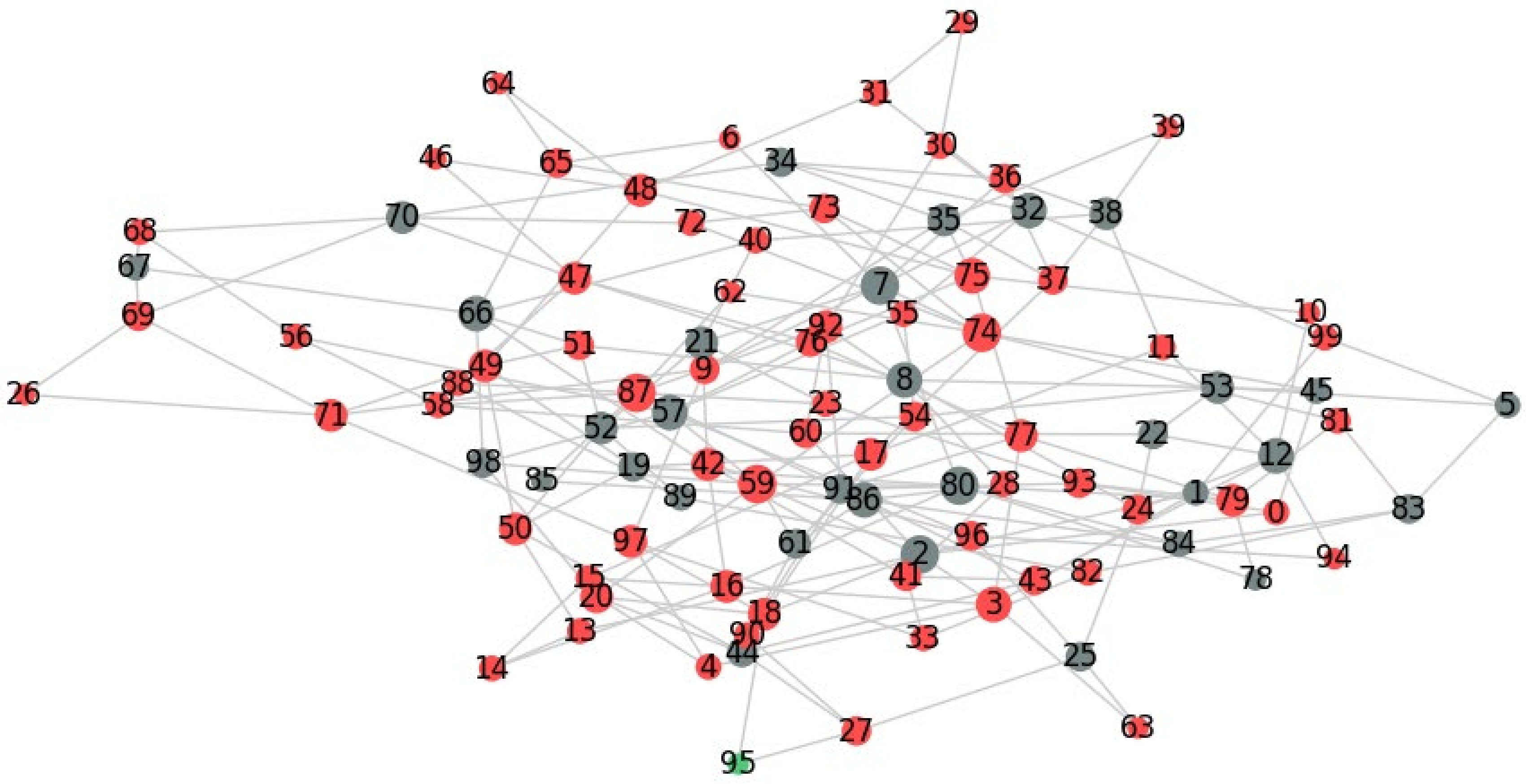
Nodes Status at Time = 4

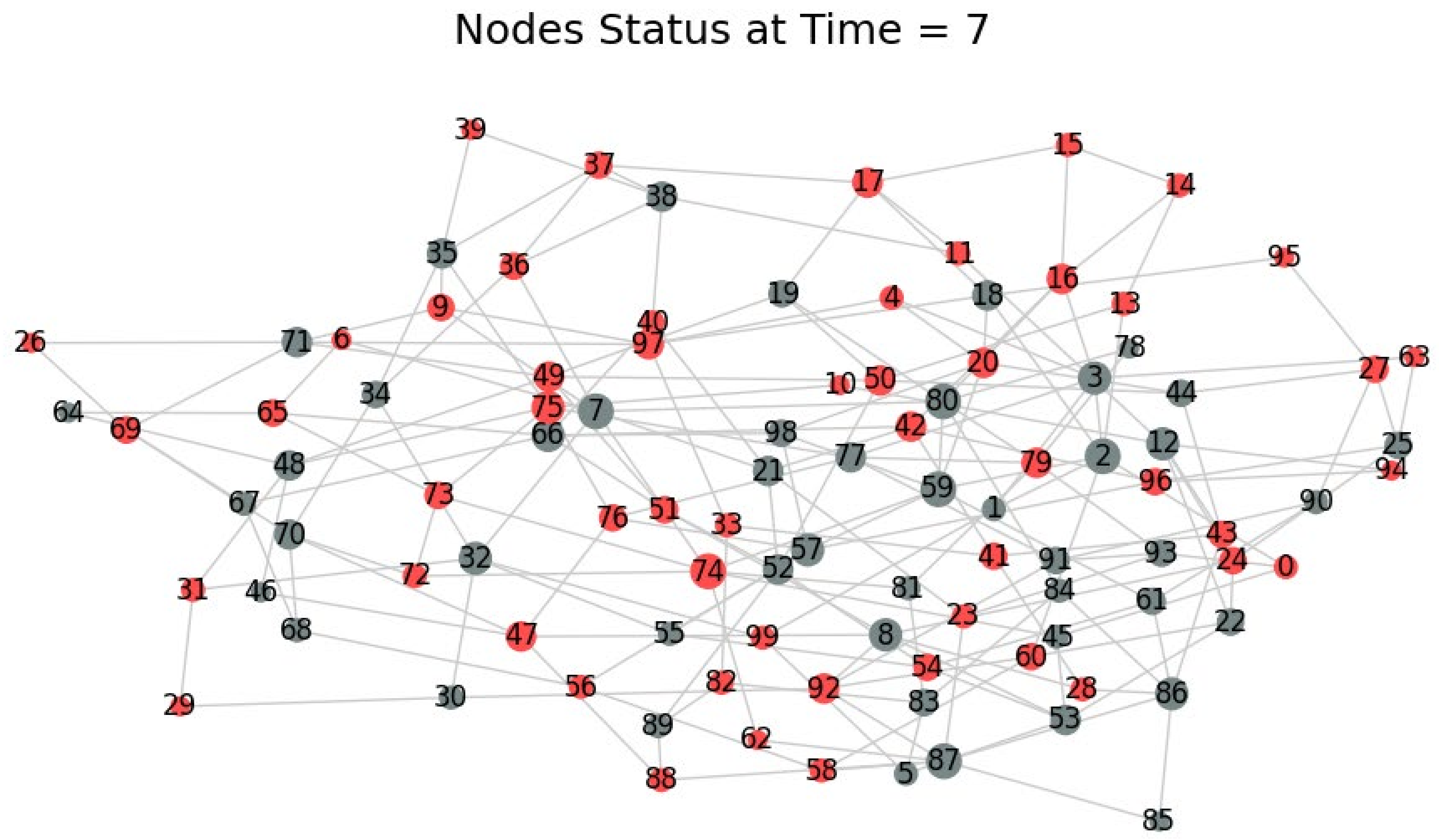


Nodes Status at Time = 5



Nodes Status at Time = 6



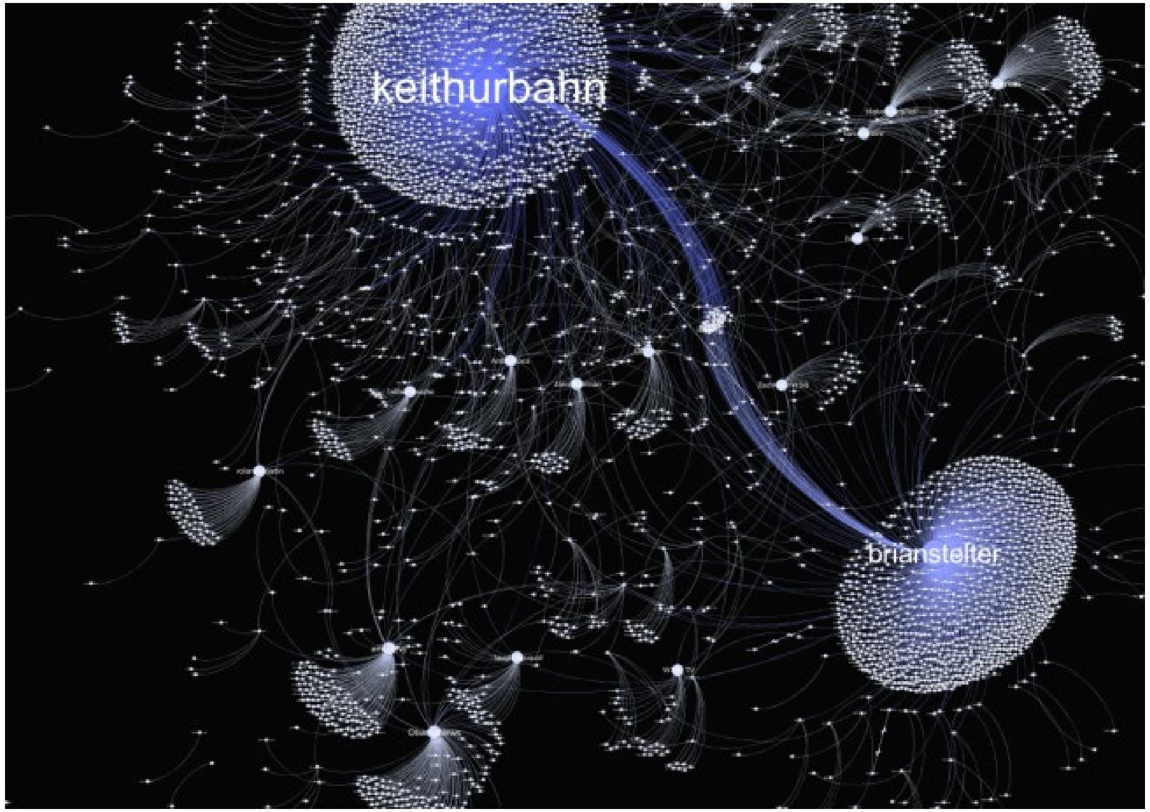

Nodes Status at Time = 7
39
37
15
17
14
38
35
36
11
95
16
19
4
18
13
9
40
26
71
6
97
78
63
49
10
50
20
3
27
44
34
75
7
80
64
69
65
66
98
42
48
77
12
2
25
21
79
94
96
67
73
59
51
76
33
1
90
70
57
43
32
91
93
24
0
72
74
52
31
46
81
84
61
23
22
68
47
55
99
8
45
60
54
28
86
82
56
92
83
29
30
53
89
62
87
58
5
88
85

Scale Free Networks

Turns out, one guy had the Twitter scoop on bin Laden's death, and it wasn't the guy who liveblogged the raid.

@keithurbahn aka Keith Urbahn, Chief of Staff at the office of Donald Rumsfeld, who a week ago had a little over 1,000 followers (which is respectable, but by no means a guaranteed viral indicator) set off a cascading domino effect with one tweet: "So I'm told by a reputable person they have killed Osama Bin Laden. Hot damn."

Stelter (aka @brianstelter) was the next big pinpoint of the Twitter effect. With 50,000 followers (now almost 56,000), Stelter "engaged his audience and yielded hundreds of retweets, furthering the spread of this message," which you can see in the visualization below.



[The Osama bin Laden Twitter explosion -- illustrated]

Scale Free Network

"A **Scale Free Network** is one in which the distribution of links to nodes follows a power law. The power law means that the vast majority of nodes have very few connections, while a few important nodes (we call them Hubs) have a huge number of connections."
– Future Learn

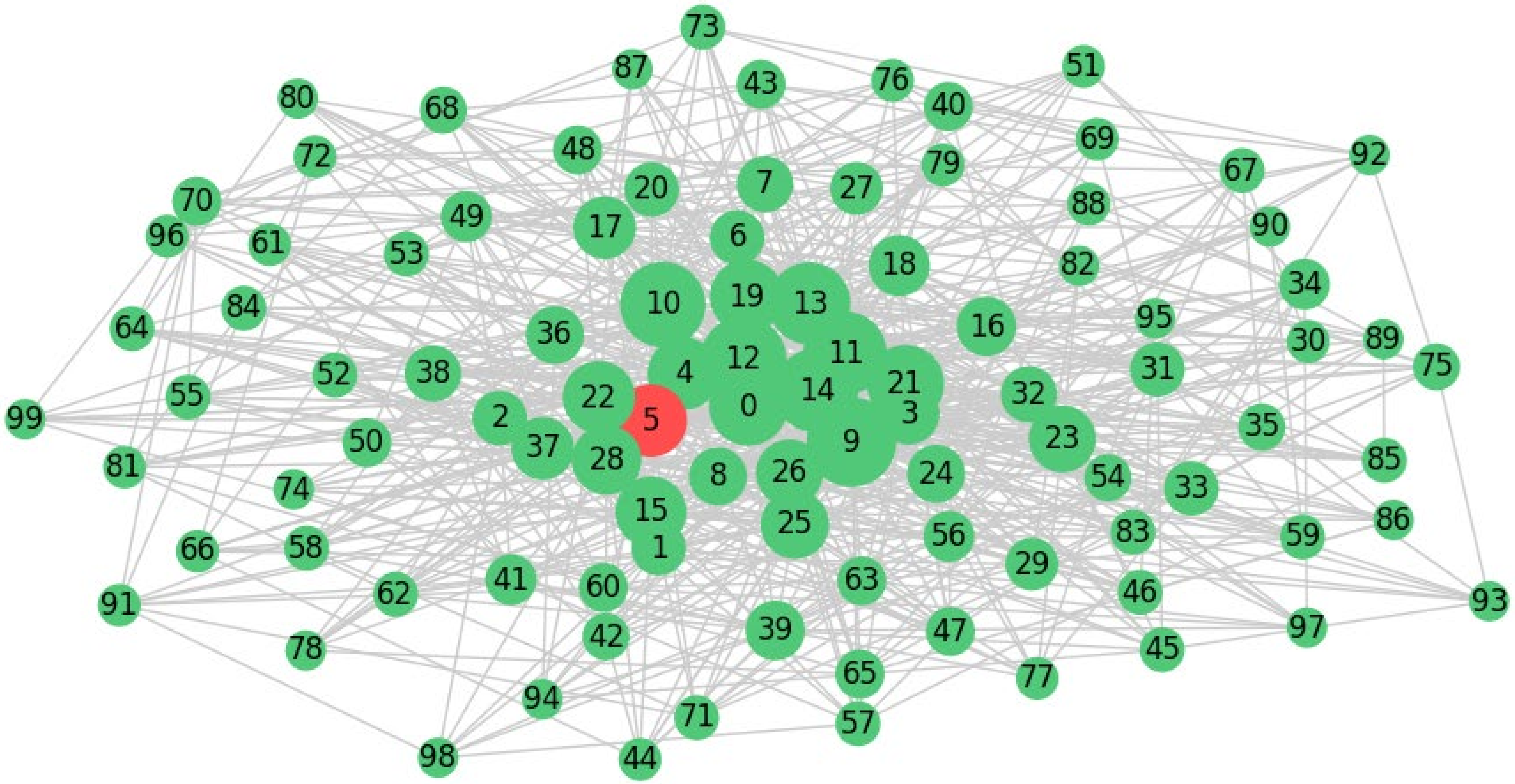
[Submitted on 21 Oct 1999]

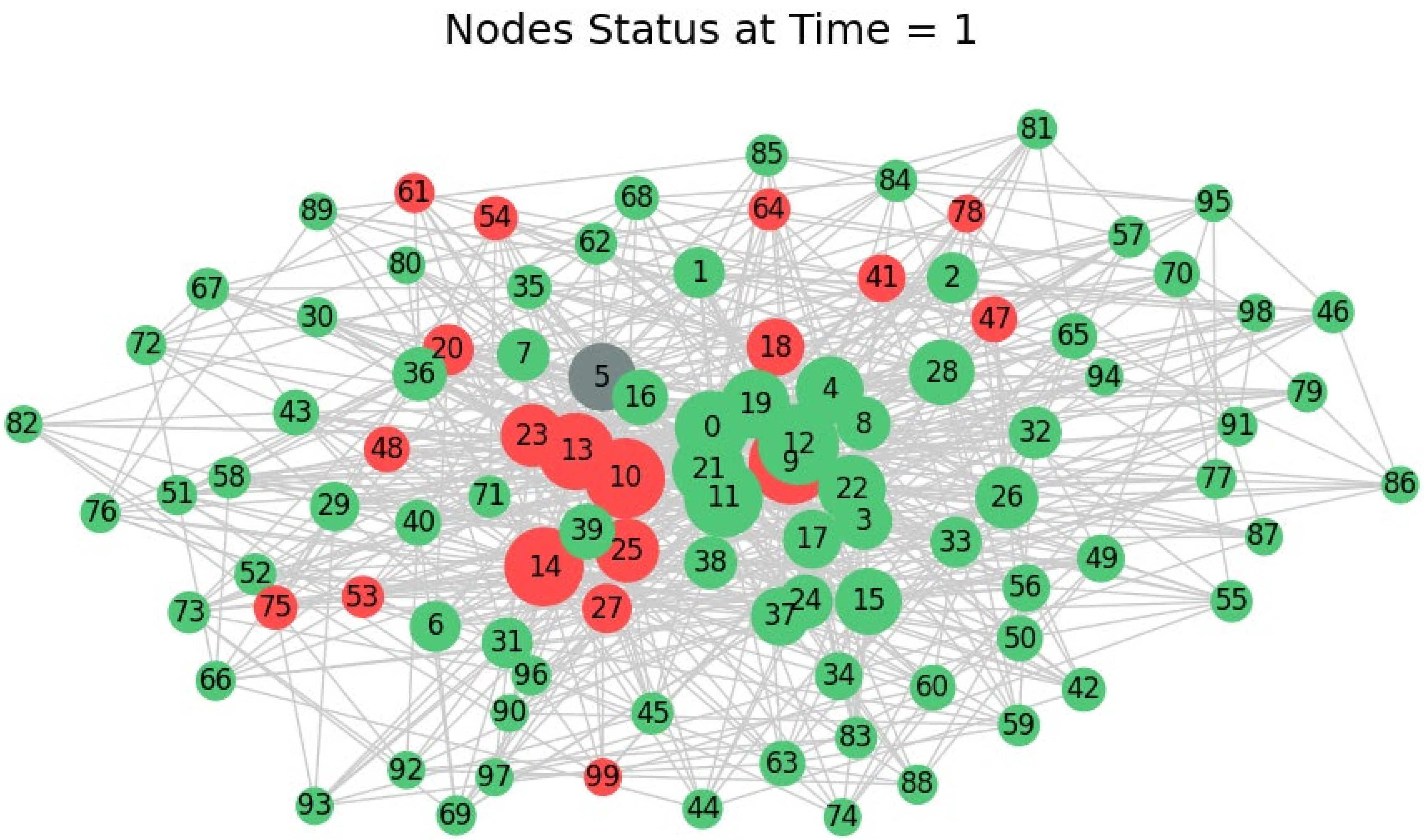
Emergence of scaling in random networks

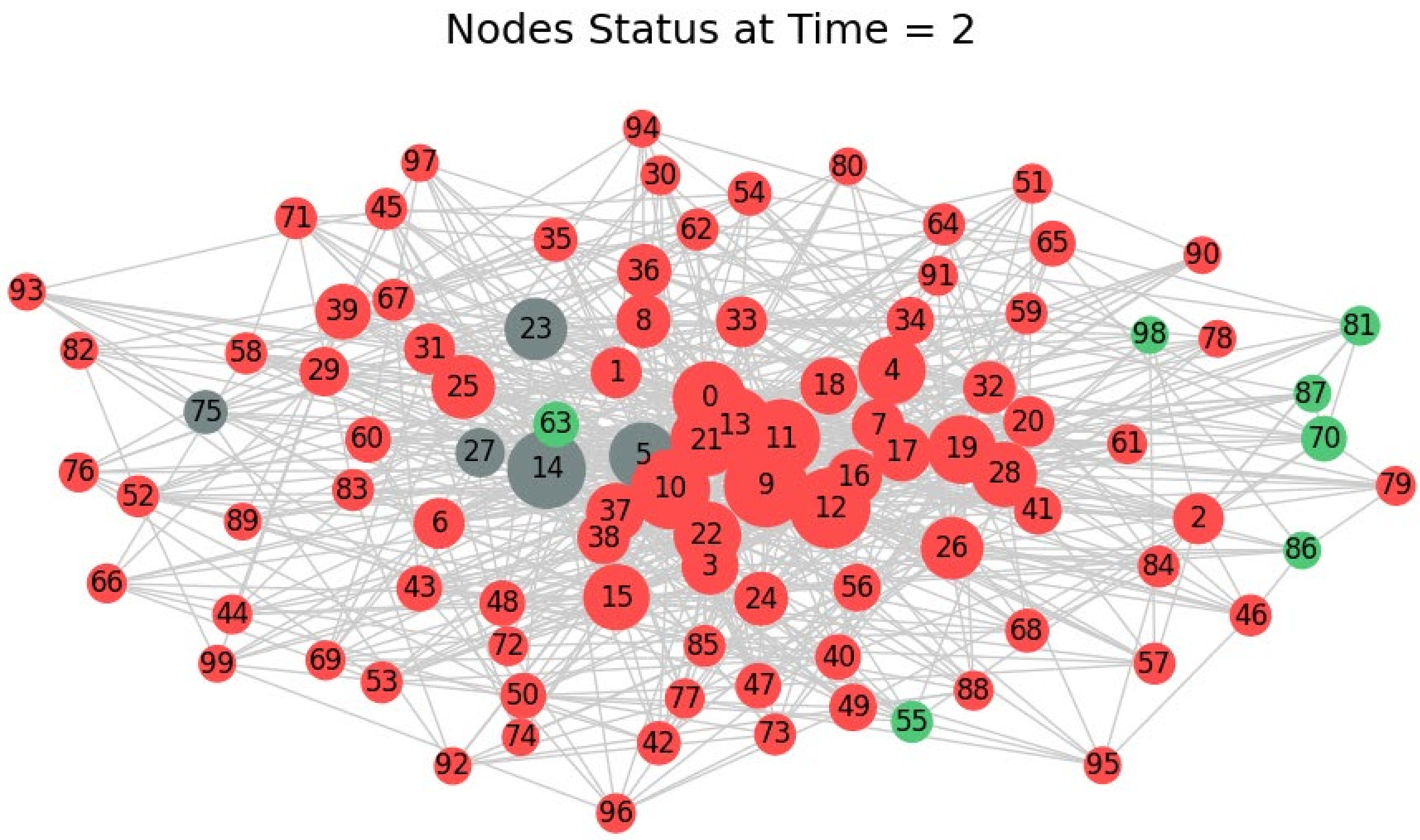
Albert-Laszlo Barabasi, Reka Albert (Univ. of Notre Dame)

Systems as diverse as genetic networks or the world wide web are best described as networks with complex topology. A common property of many large networks is that the vertex connectivities follow a scale-free power-law distribution. This feature is found to be a consequence of the two generic mechanisms that networks expand continuously by the addition of new vertices, and new vertices attach preferentially to already well connected sites. A model based on these two ingredients reproduces the observed stationary scale-free distributions, indicating that the development of large networks is governed by robust self-organizing phenomena that go beyond the particulars of the individual systems.

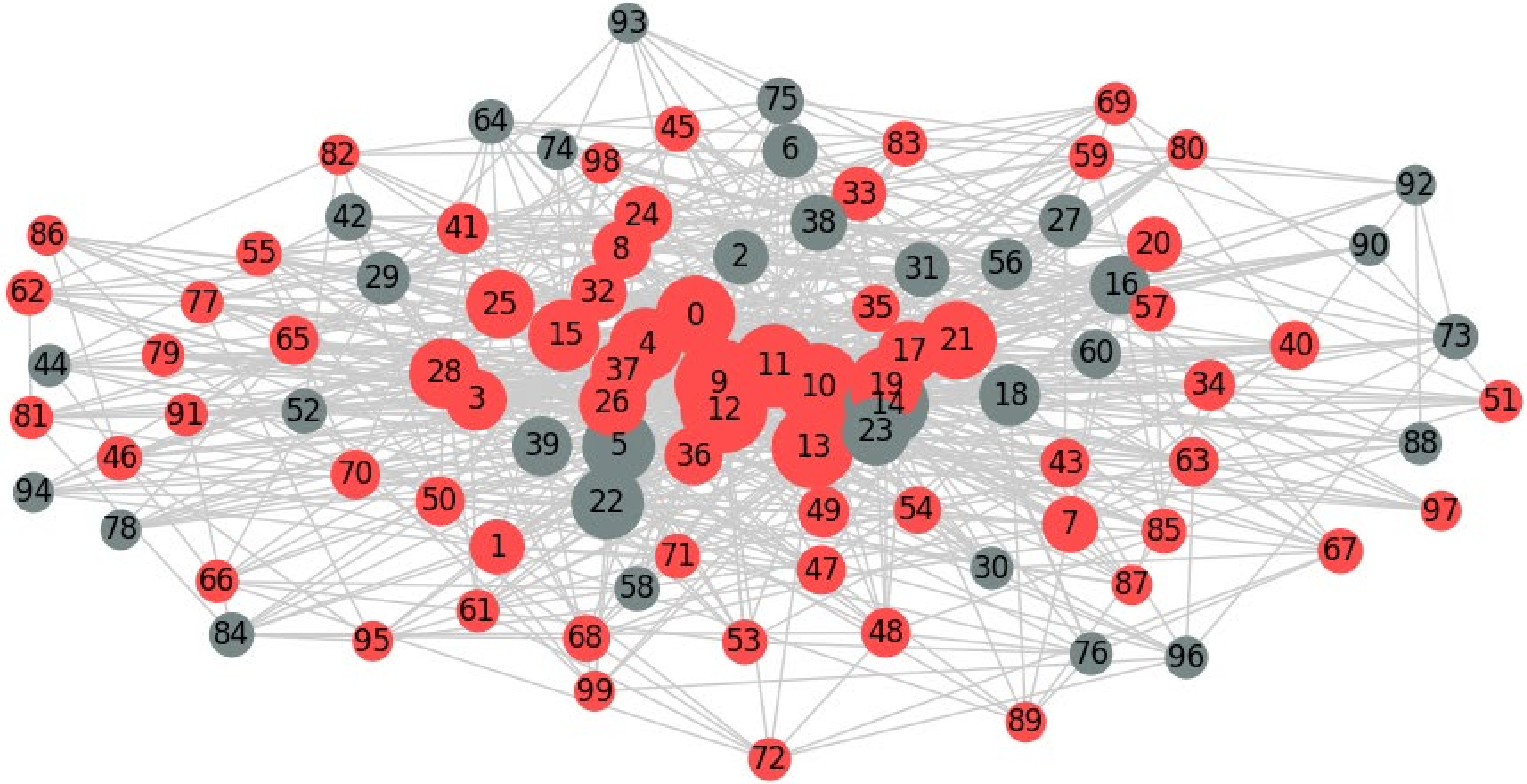
Nodes Status at Time = 0



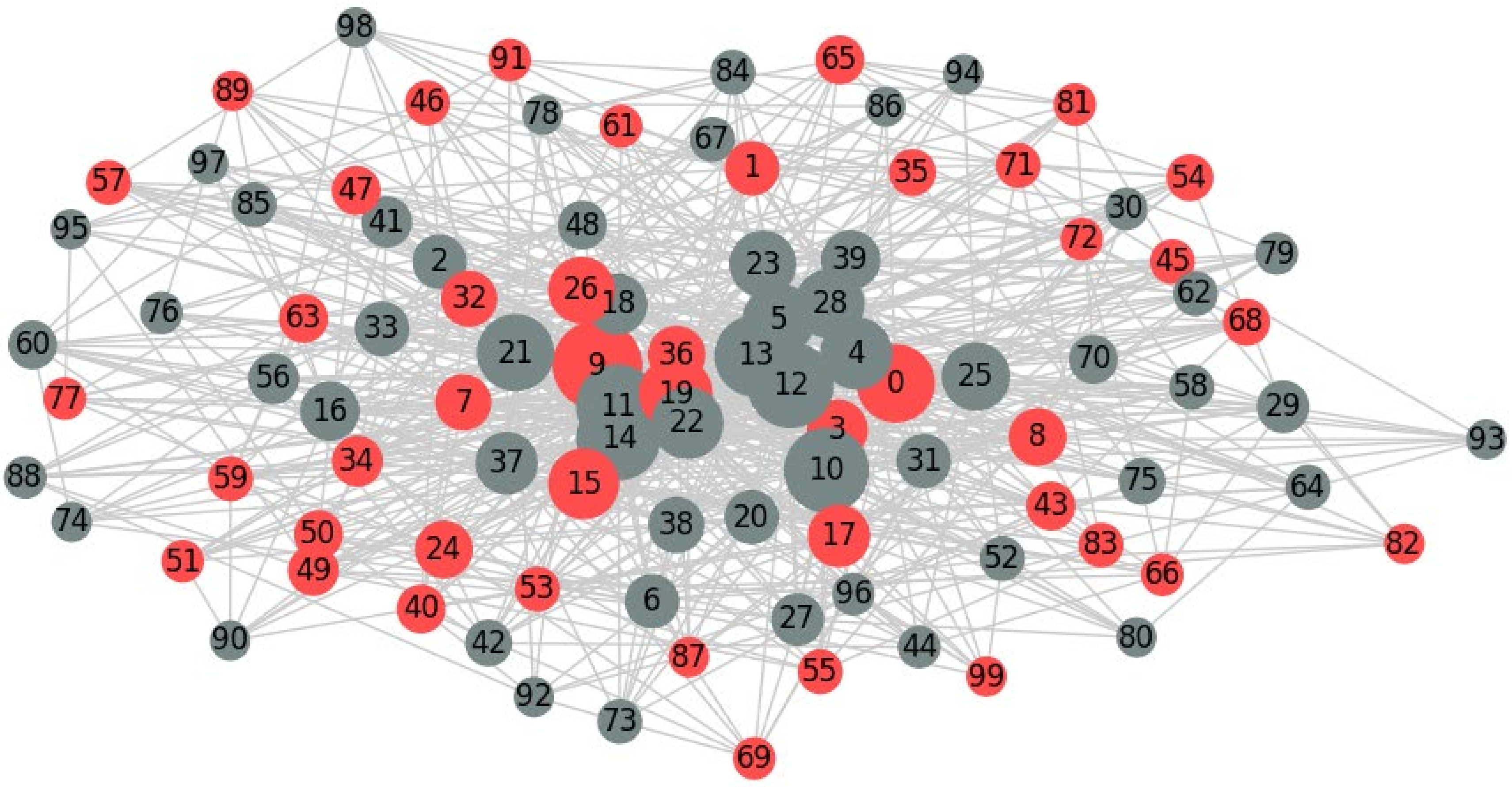

Nodes Status at Time = 1

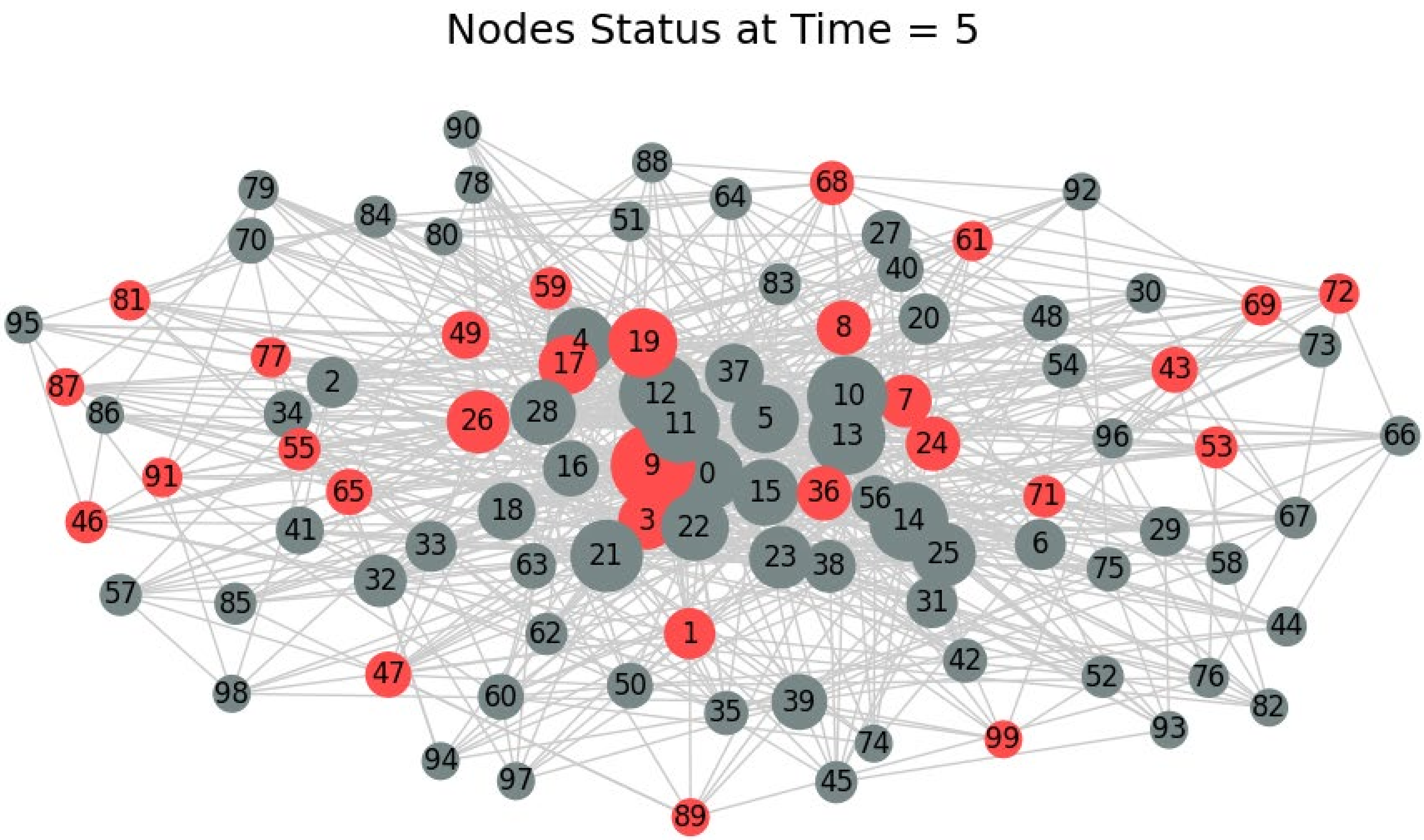

Nodes Status at Time = 2

Nodes Status at Time = 3

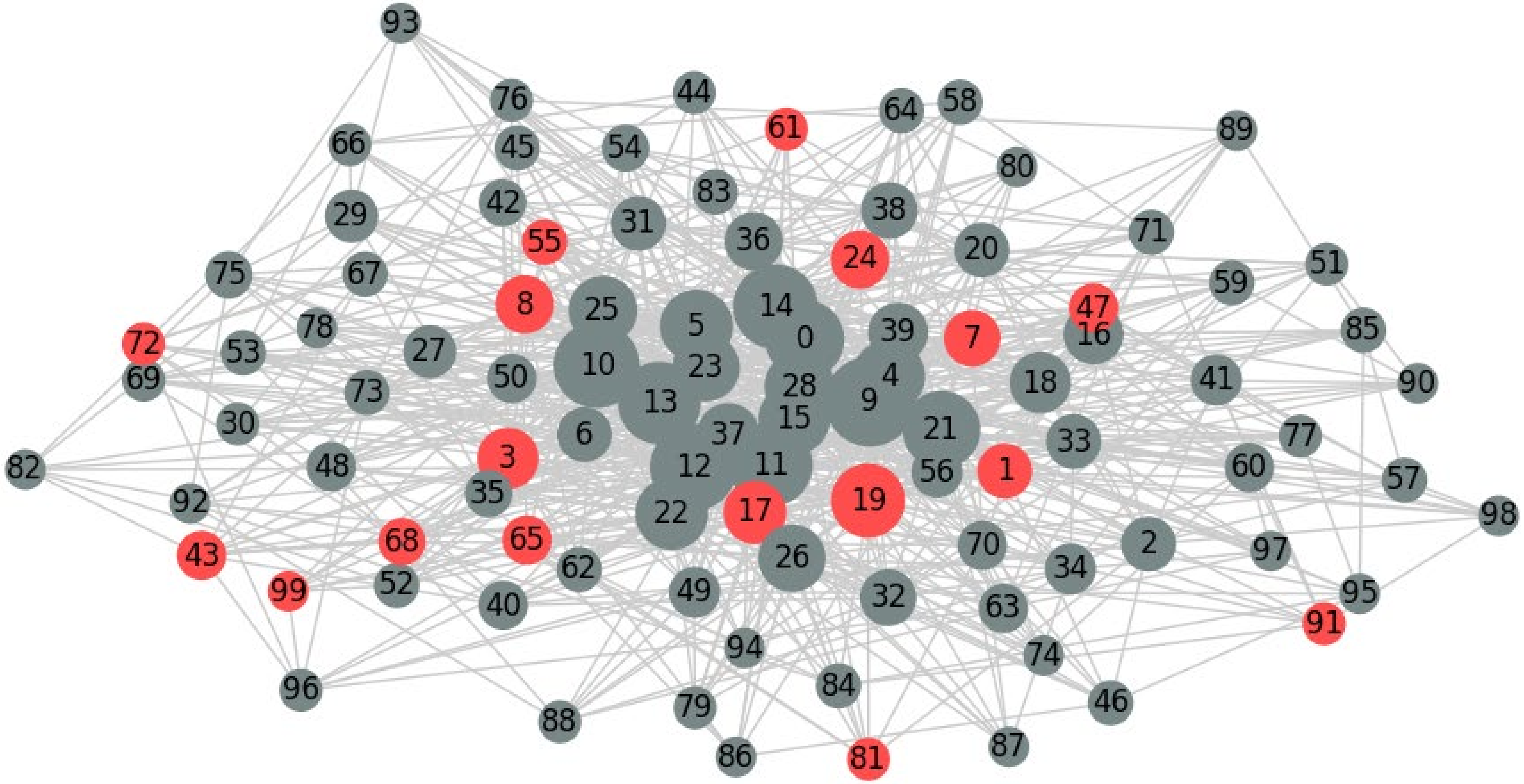


Nodes Status at Time = 4



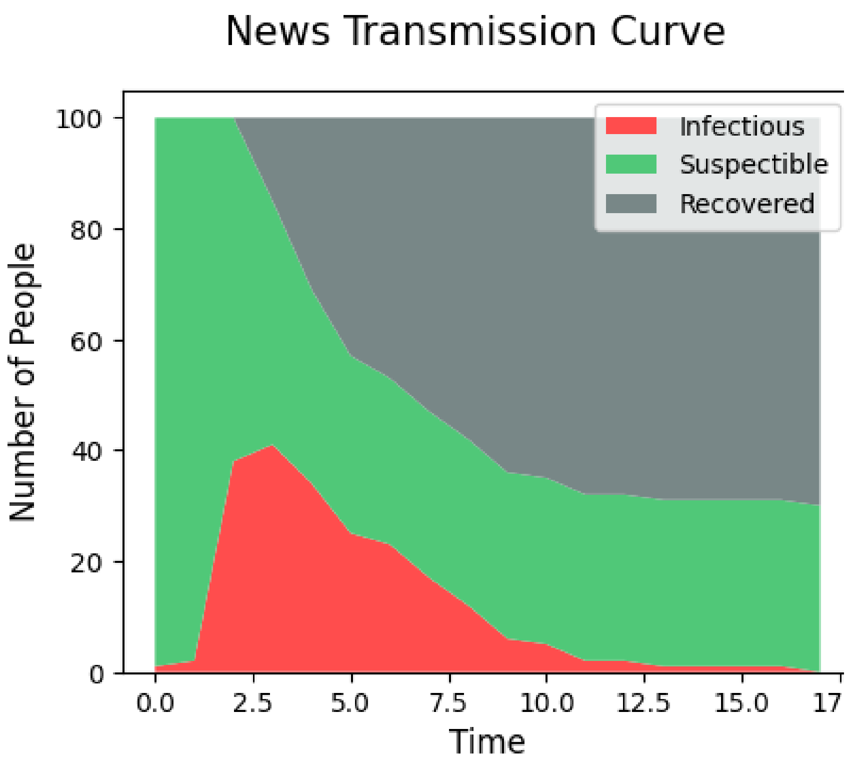

Nodes Status at Time = 5

Nodes Status at Time = 6

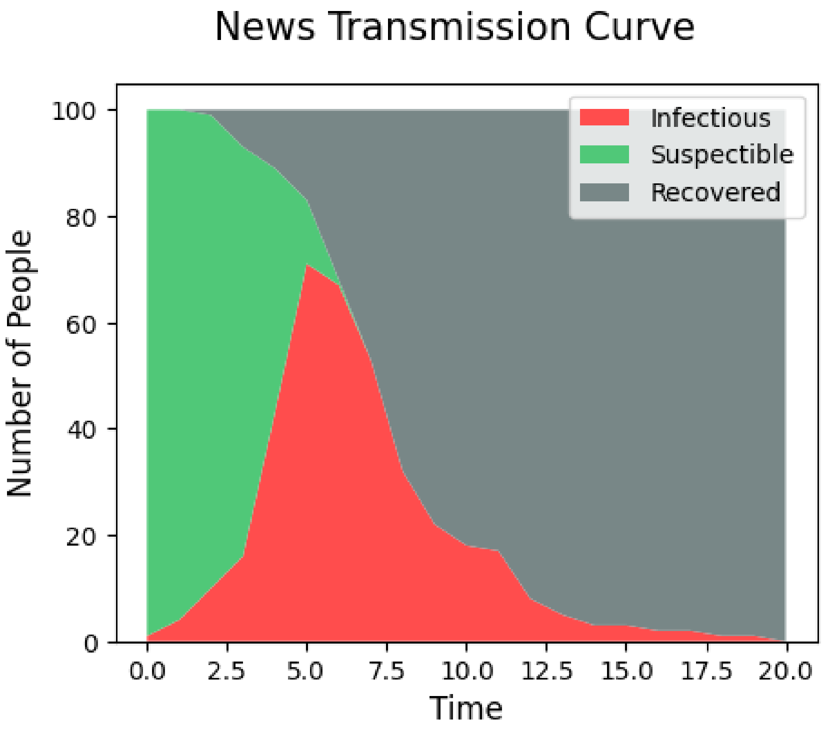


Information Diffusion Dynamics

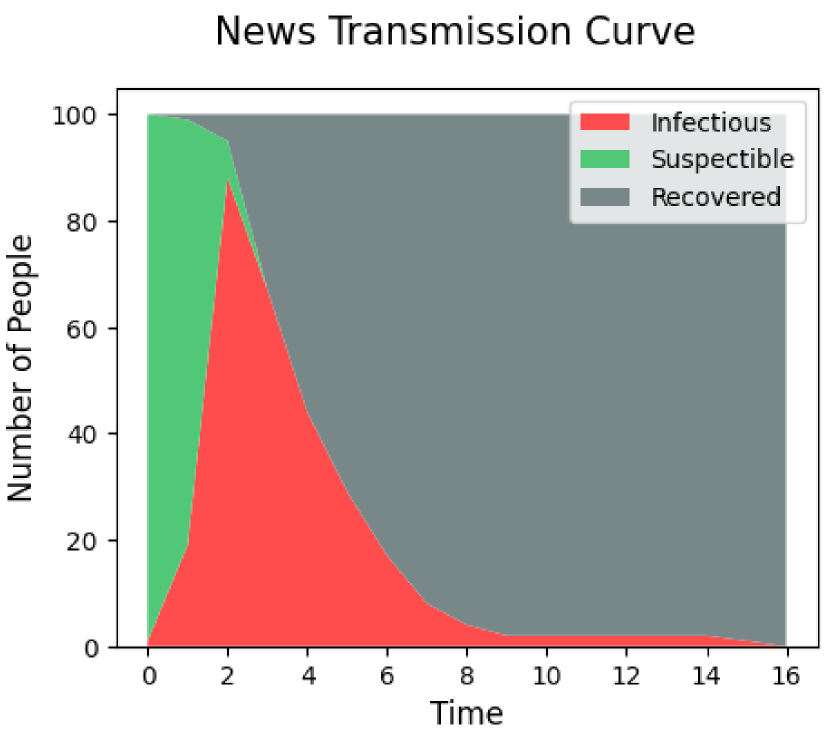
Star graph



Small world graph



Scale free graph



Infections Spread Fastest through Scale Free Networks

Results from agent-based model of viral diffusion on 3 different network models.

	Star	Small World	Scale Free
R0	2.2	5.3	19.6
Peak # of cases	41	71	88
Peak time period	3	5	2
Total time periods	17	20	16
# of individuals never infected	30	0	0

Source: Jupyter Notebook • Created with Datawrapper

Social Media Peer to Peer Branching Diffusion

Broadcast diffusion

- Star graph in a 2-step flow diffusion process
- News spreads through clusters of social groups in a small-world network

Social media diffusion

- Scale-free graph with multiples hubs
- Super-spreaders push rapid, widespread transmission
- Scale-free network structure and dynamics make it **difficult** to contain rapid spread of ideas, beliefs, and collective action

Superspreaders' effect on epidemic threshold

VOLUME 86, NUMBER 14 PHYSICAL REVIEW LETTERS 2 APRIL 2001

Epidemic Spreading in Scale-Free Networks

Romualdo Pastor-Satorras[1] and Alessandro Vespignani[2]

[1]*Departament de Física i Enginyeria Nuclear, Universitat Politècnica de Catalunya, Campus Nord, Mòdul B4, 08034 Barcelona, Spain*

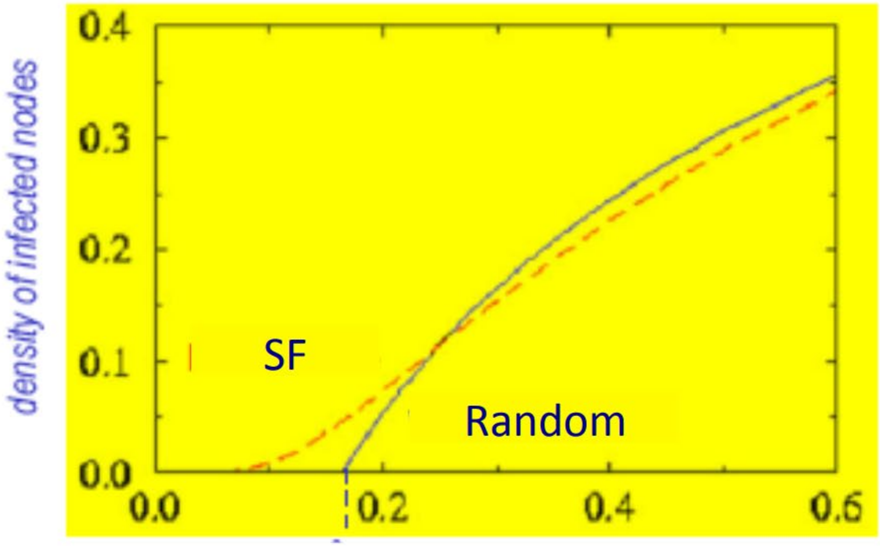
[2]*The Abdus Salam International Centre for Theoretical Physics (ICTP), P.O. Box 586, 34100 Trieste, Italy*

(Received 20 October 2000)

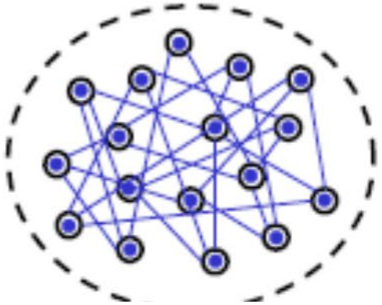
The Internet has a very complex connectivity recently modeled by the class of scale-free networks. This feature, which appears to be very efficient for a communications network, favors at the same time the spreading of computer viruses. We analyze real data from computer virus infections and find the average lifetime and persistence of viral strains on the Internet. We define a dynamical model for the spreading of infections on scale-free networks, finding the absence of an epidemic threshold and its associated critical behavior. This new epidemiological framework rationalizes data of computer viruses and could help in the understanding of other spreading phenomena on communication and social networks.

Preventing an outbreak is tough in Scale Free Networks!

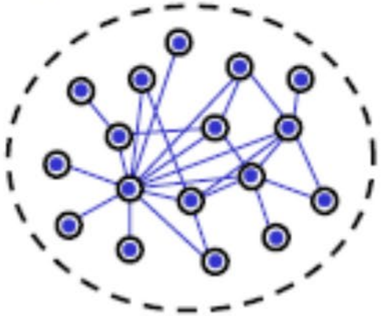
No Epidemic Threshold for SF!



Infections proliferate in SF networks independently of their spreading rates!



Random Network



Scale Free Network

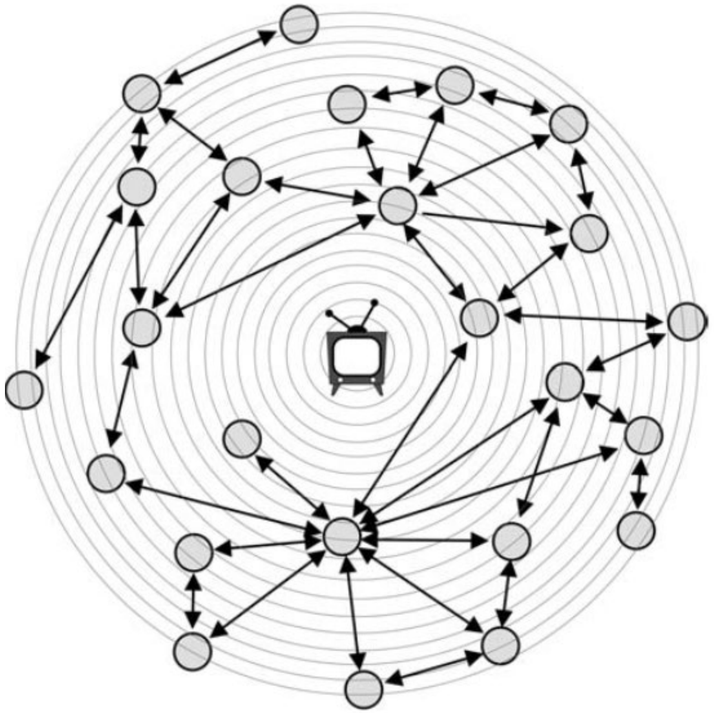
[Slide by Dr. Cecilia Mascolo Social and Technological Network Analysis]()

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"

Superspreaders and Micro-influencers

Everyone's an Influencer: Quantifying Influence on Twitter

Eytan Bakshy*
University of Michigan, USA
ebakshy@umich.edu

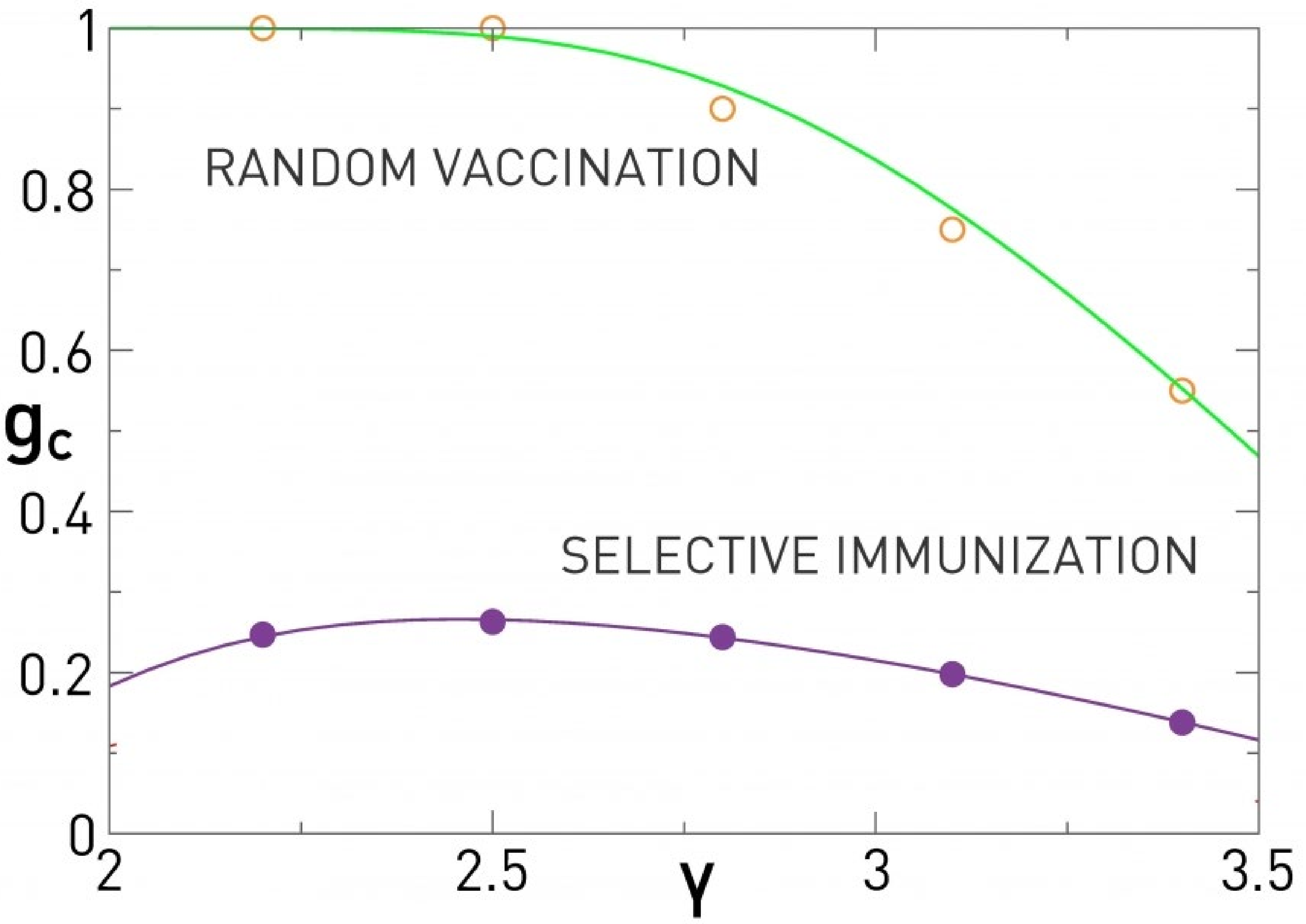
Jake M. Hofman
Yahoo! Research, NY, USA
hofman@yahoo-inc.com

Winter A. Mason
Yahoo! Research, NY, USA
winteram@yahoo-inc.com

Duncan J. Watts
Yahoo! Research, NY, USA
djw@yahoo-inc.com

We find that although under some circumstances, the most influential users are also the most cost-effective, under a wide range of plausible assumptions the most cost-effective performance can be realized using "ordinary influencers"---individuals who exert average or even less-than-average influence.

Immunization strategy intervention



[*Network Science* textbook](#)
by Albert-László Barabási

Network Science

by Albert-László Barabási

Personal Introduction
1. Introduction
2. Graph Theory
3. Random Networks
4. The Scale-Free Property
5. The Barabási-Albert Model
6. Evolving Networks
7. Degree Correlations
8. Network Robustness
9. Communities
10. Spreading Phenomena
Preface

Network Analysis to Discover Emerging Influencers & Superspreaders

- Information is continuously diffusing over the network

- We don’t have much visibility into the **network topology** (only the internal staff has access to the full graph; they don’t track/moderate all the content)

- We can collect data on small communities that form around specific topics to reconstruct **network dynamics** by monitoring how influencers emerge and spread ideas over time

Hands-on Exercise

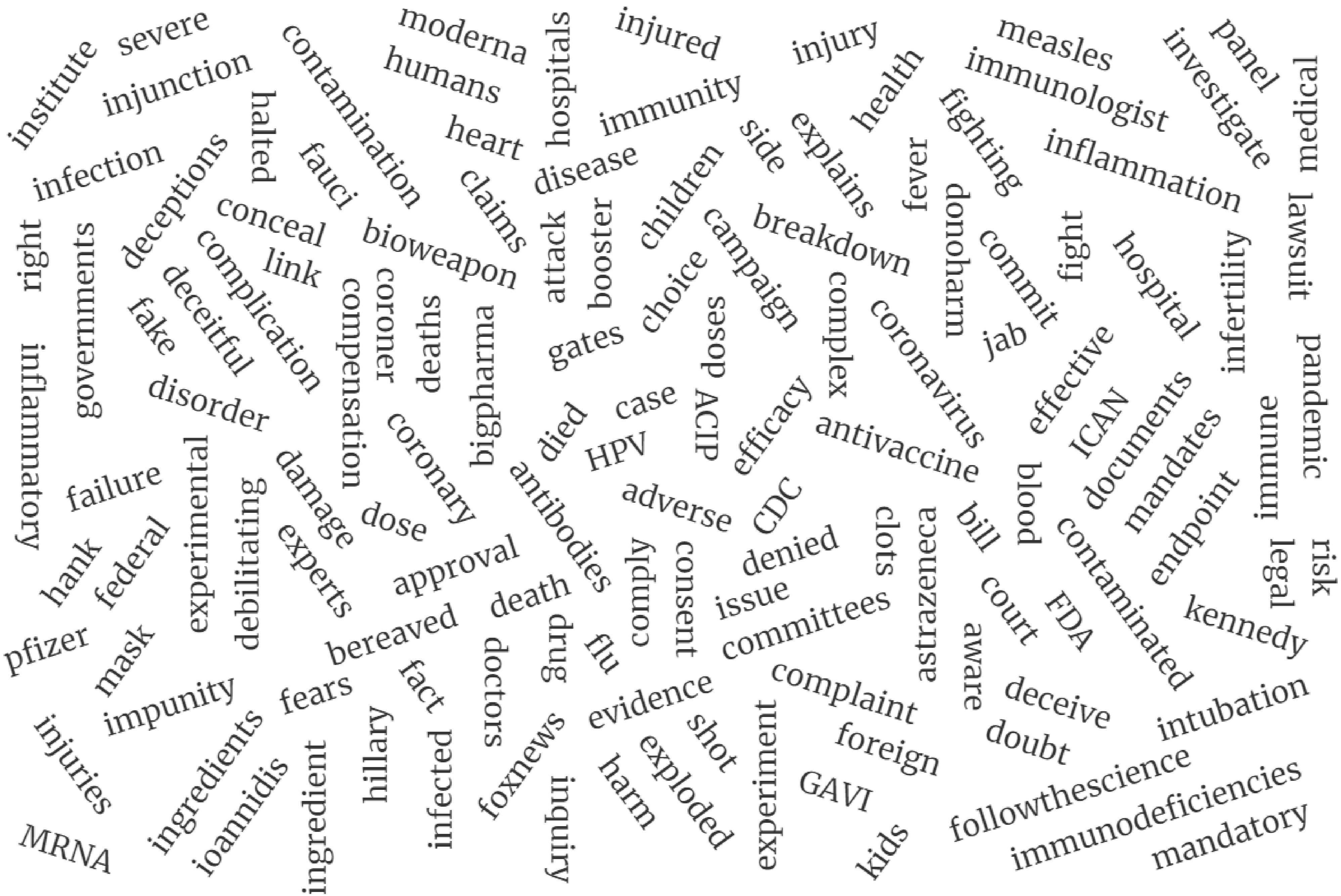
Anti-vaccine tweets network analysis

Please download data files here:

https://tinyurl.com/trustcondatafiles

Search keywords used to filter for antivax tweets

Can be expanded iteratively



Snowball sampling – Collect tweets over 4 time periods

Generation 1 (Sept 2021 – Jan 2022)
https://x.com/i/lists/1333771036274331654

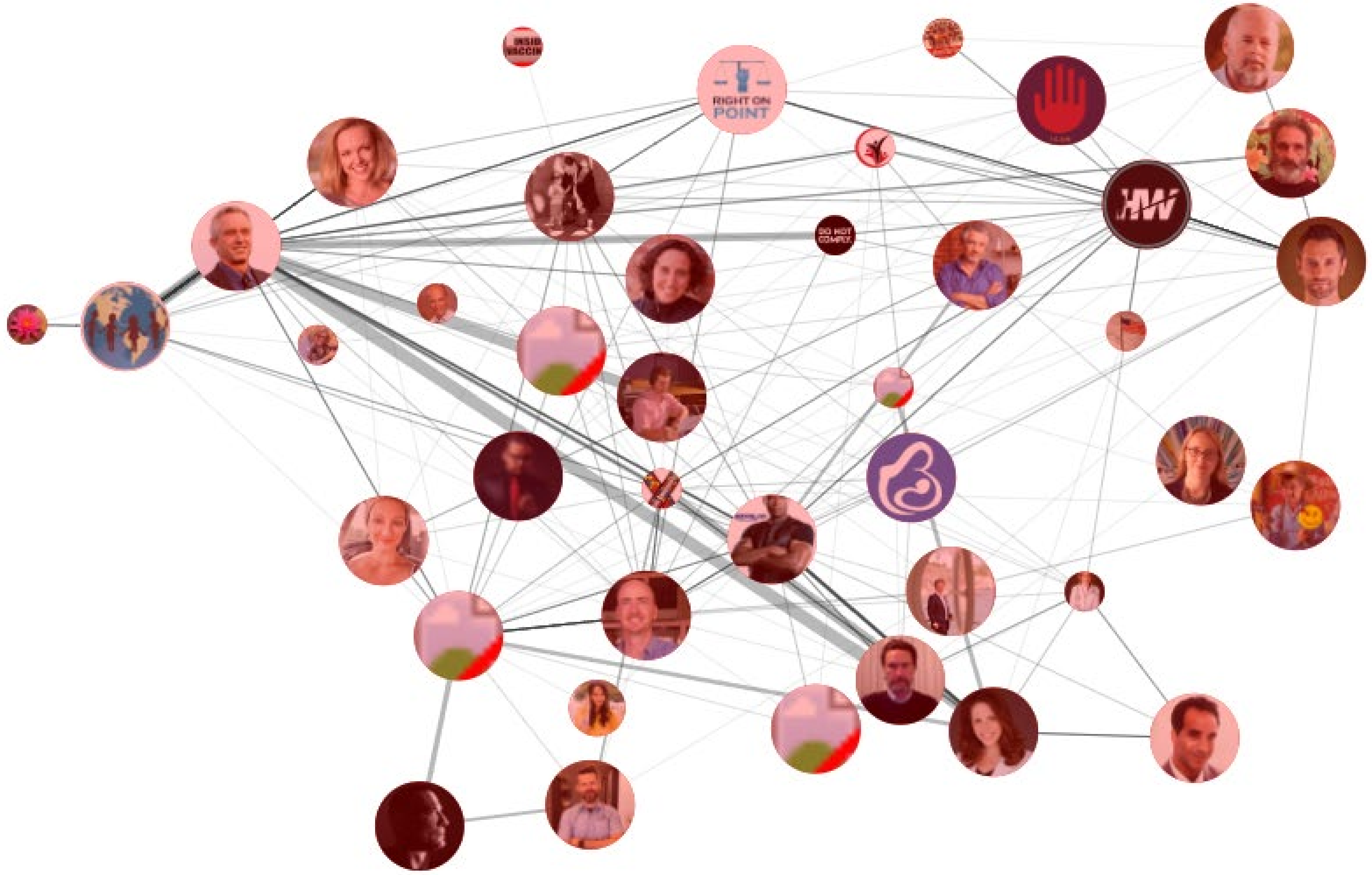
Generation 2 (Jan 2022 – March 2022)
https://x.com/i/lists/1507739466647146498

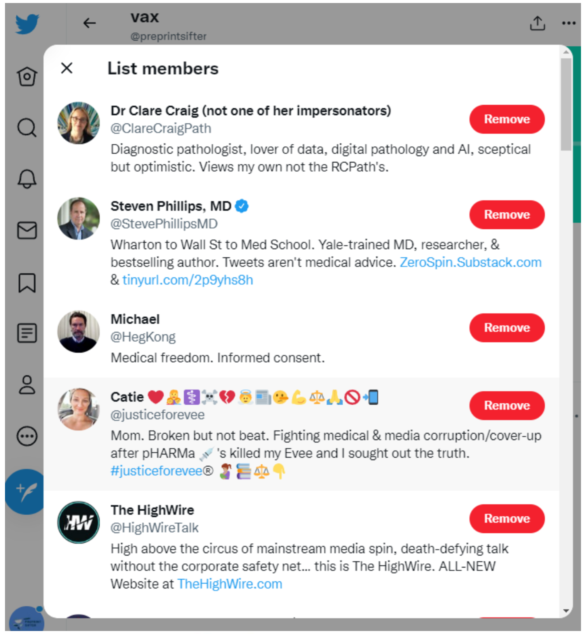
Generation 3 (March 2022 – April 2022)
https://x.com/i/lists/1511377412411506688

Generation 4 (April 2022 – May 2022)
https://x.com/i/lists/1513731136958210049

List of all Twitter Users in a Google Sheet

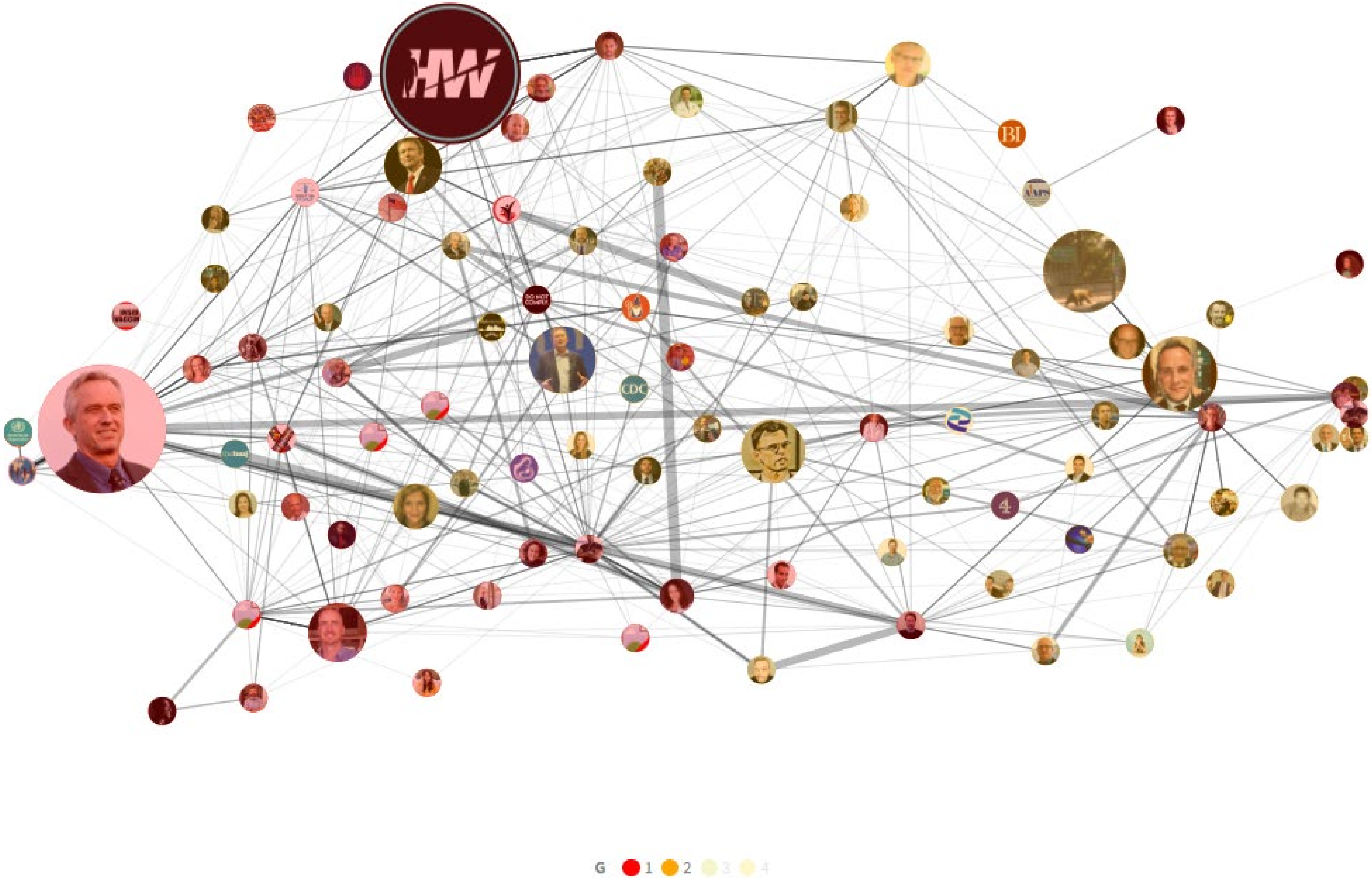
Generation 1: Seeded with 30 anti-vax accounts compiled by public health experts from Harvard Global Health Institute

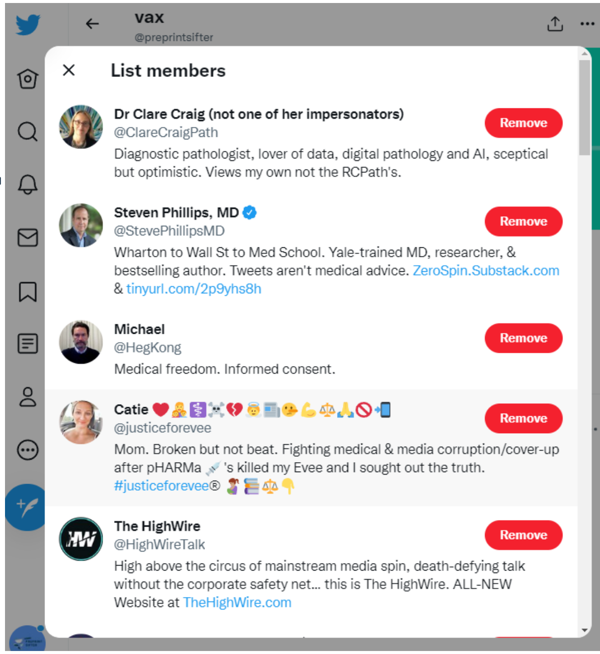


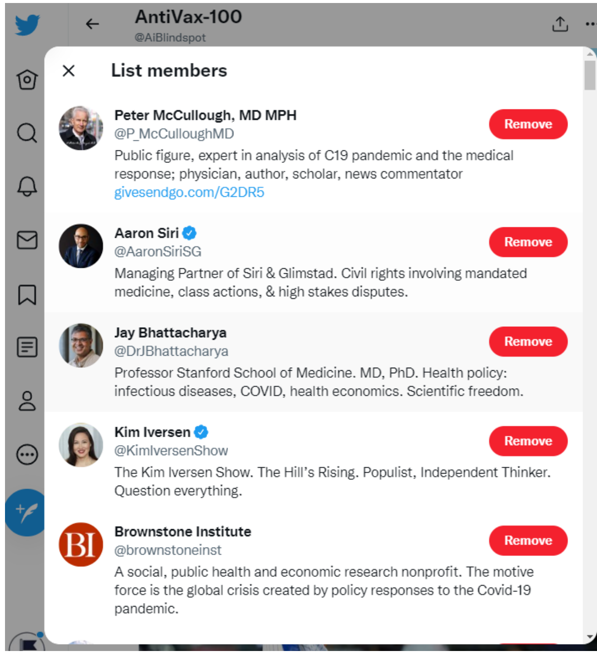


https://twitter.com/i/lists/1333771036274331654/members

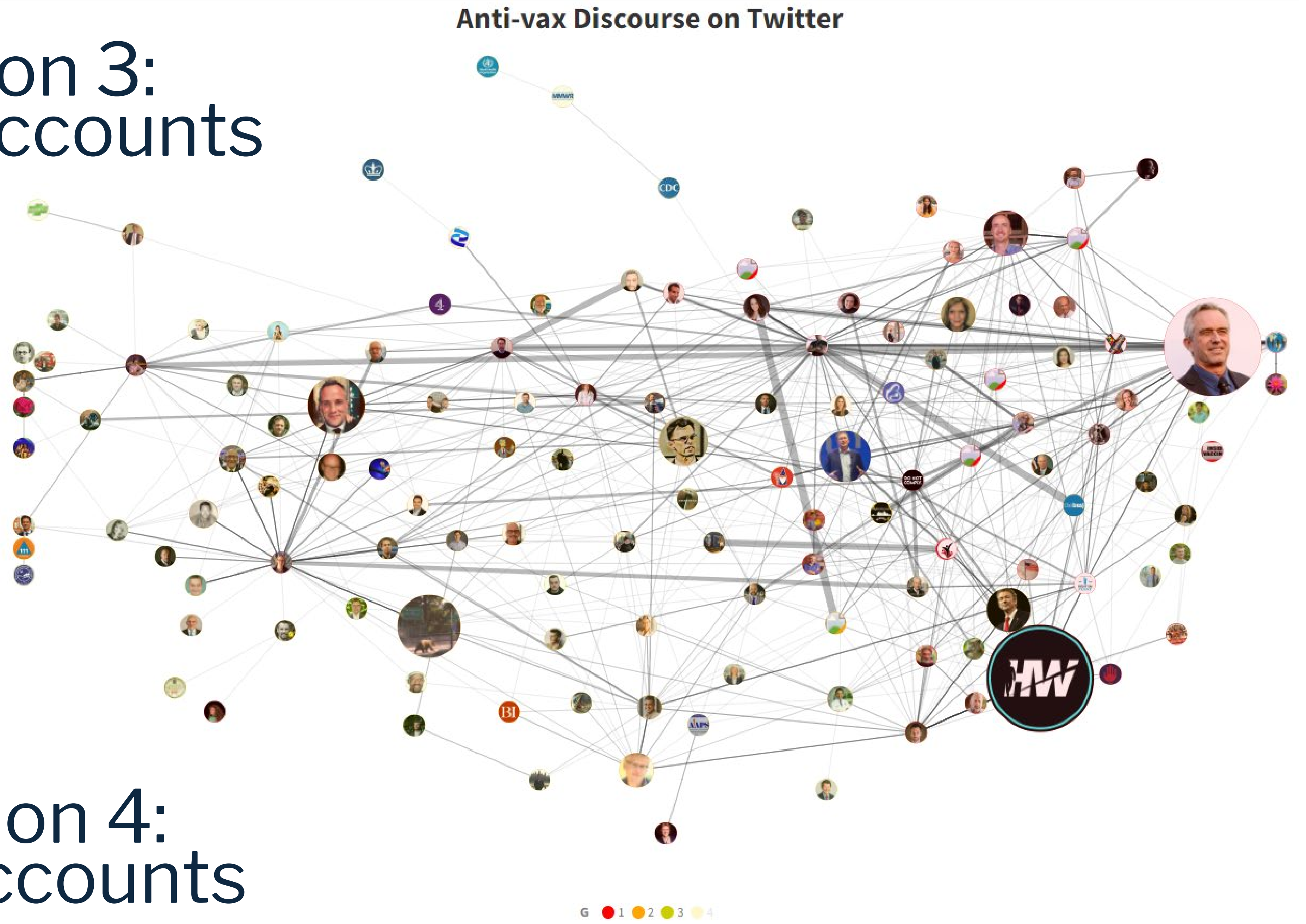
Generation 2: add 76 Twitter users who are retweeted frequently by generation 1







Generation 3:
add 38 accounts



Generation 4:
add 11 accounts

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

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

Retweet by Not Your Mom

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

Retweet by Dr Clare Craig

Dr Aseem Malhotra
@DrAseemMalhotra

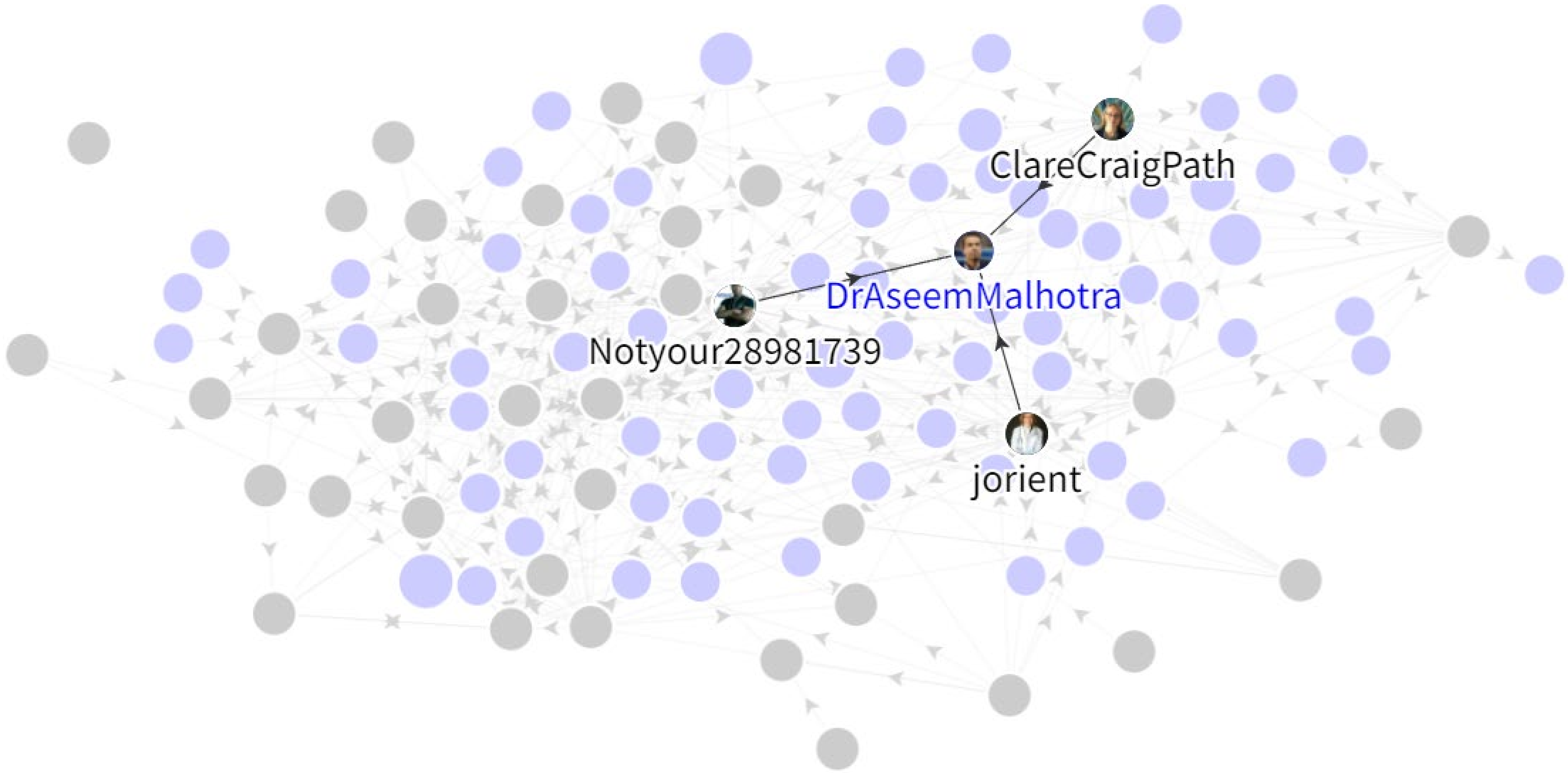
Brilliant. @sajidjavid PLEASE U TURN on vaccine mandate for #NHS staff. It is not ethical, will result in catastrophic staff shortages, and is not backed by the BMA nor the Academy of Medical Royal Colleges

#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

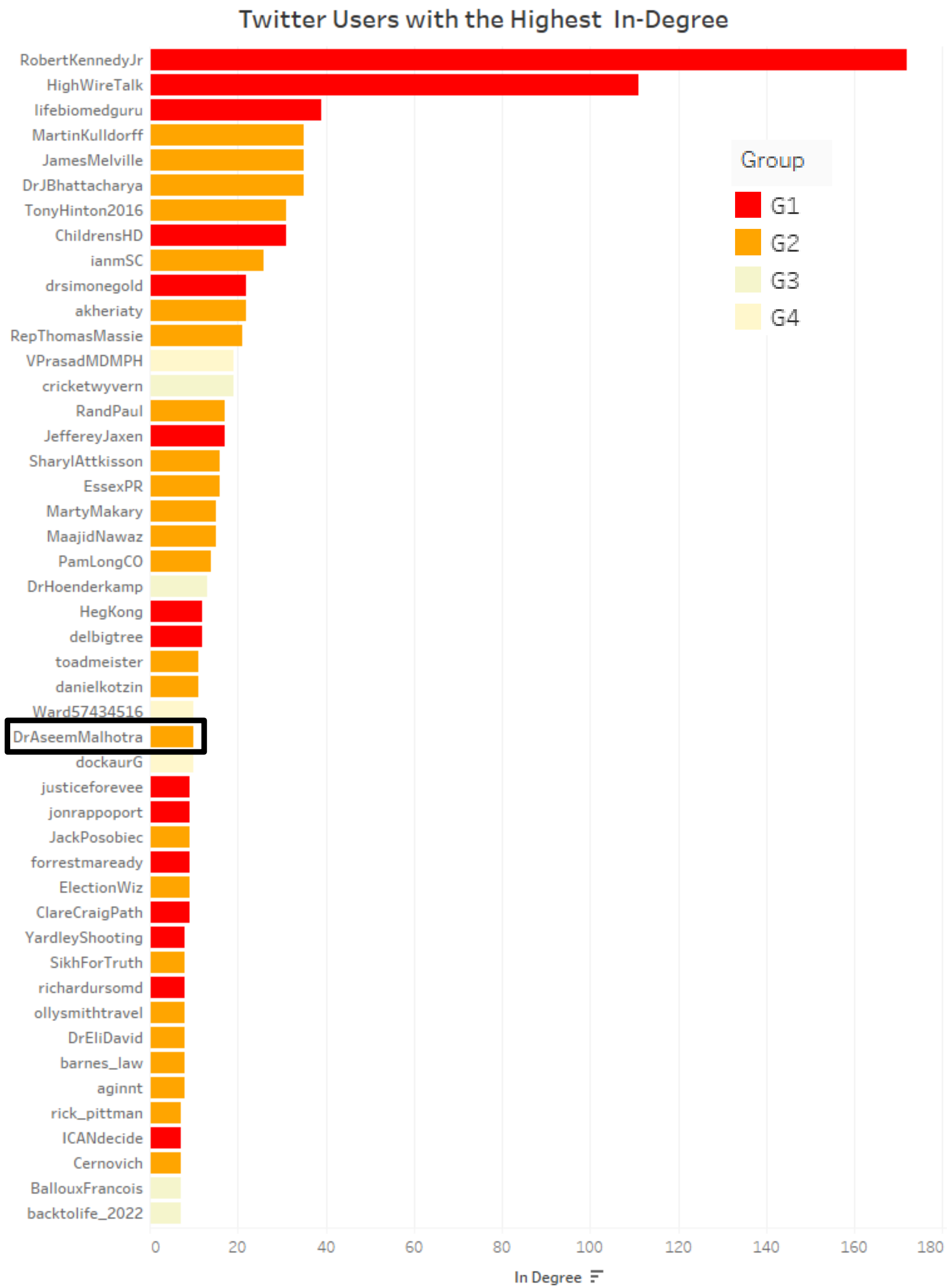

ClareCraigPath
DrAseemMalhotra
Notyour28981739
jorient

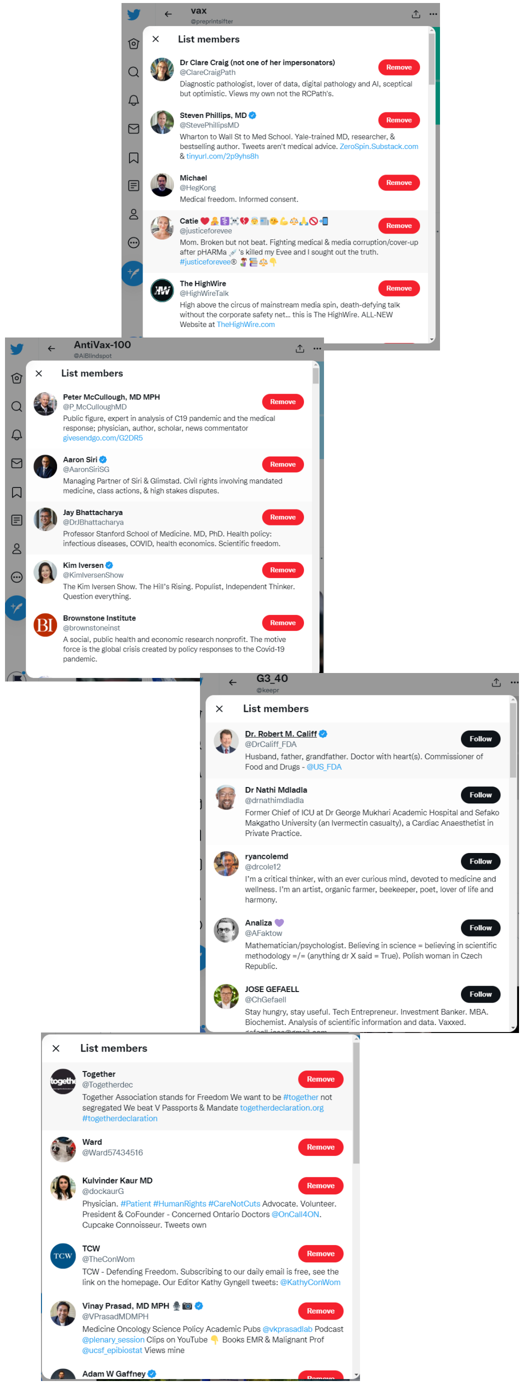
https://tinyurl.com/findingsuperspreaders

```
# prompt: From the Mention column in df_network_g2, create a frequency table df_frequency_g2 that contains M

df_frequency_g2 = df_network_g2['Mention'].value_counts().rename_axis('Mention').reset_index(name='Frequency
df_frequency_g2_sorted = df_frequency_g2.sort_values(by='Frequency', ascending=False)
df_frequency_g2_sorted[:100]
```

	Mention	Frequency
0	YardleyShooting	240
1	DrAseemMalhotra	235
2	RobertKennedyJr	166
3	DrJohnB2	154
4	P_McCulloughMD	109
...	...	...
92	stayingfreepod,\nSikhForTruth, TruthTalkUK	14
93	AviDascalu	14

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	bucksccgs	
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		





https://tinyurl.com/rankingsuperspreaders

Generate

calculate the degree centrality of all the nodes and save it in a new table df_nodes with the node name in the first column and centrality in the second column sorted in descending order

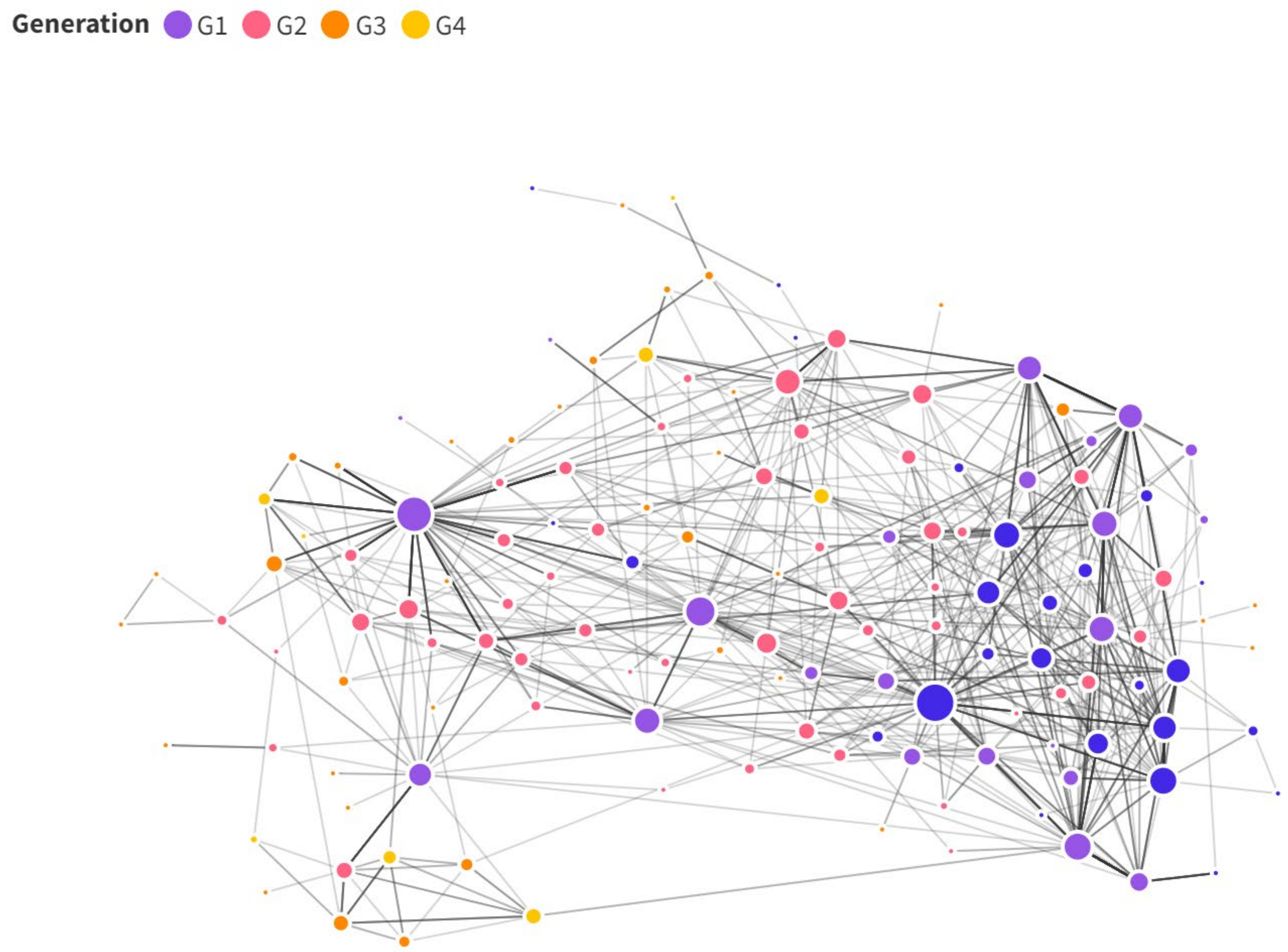
Close

1 of 1

Use code with caution

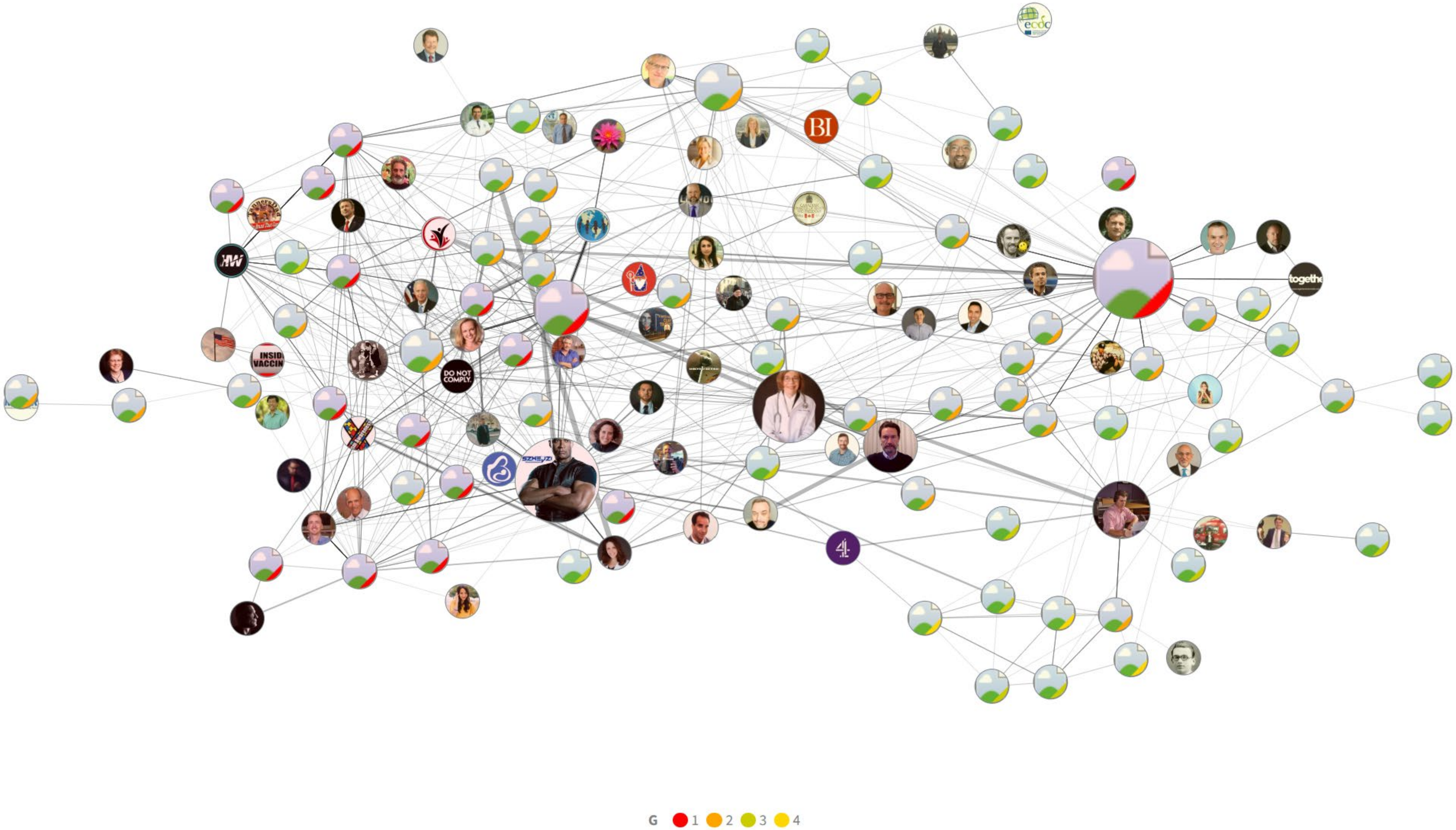
```
# prompt: calculate the degree centrality of all the nodes and save it in a new table df_nodes with the node

# Calculate degree centrality
degree_centrality = nx.degree_centrality(G)

# Create a list of tuples (node, centrality)
node_centrality = [(node, centrality) for node, centrality in degree_centrality.items()]

# Sort the list in descending order of centrality
node_centrality.sort(key=lambda x: x[1], reverse=True)

# Create a DataFrame
df_nodes = pd.DataFrame(node_centrality, columns=['Node', 'Centrality'])

print(df_nodes)
```

Network Visualization



Network Visualization with Twitter profile icons

Anti-vax Discourse on Twitter - Bridges



Discussion

Brainstorm interventions for misinformation social media

- Remove super-spreaders (deplatforming)
- Cut links between nodes
- Identify and isolate coordinated inauthentic disinformation campaigns
    - Quarantine clusters of malicious actors
- Slow down the speed and spread of strong emotions and sensational outbursts
- Decrease demand for motivated reason and confirmation bias
- Decrease affective polarization (otherizing and demonizing outgroups)
- Incentivize constructive civic discourse and deliberation
    - Prosocial bridging algorithms amplifying trusted voices/messengers
- Many, many other strategies and approaches—but will we run the experiment and ship features???
    - Virulence of narrative variants (memes) drive by algorithmic feedback loop
    - Determinants of people’s susceptibility – radicalization pathway

Summary

- It is difficult to investigate how misinformation spreading on social media using the real-world data due to lack of access to high volume, high velocity data
  a. Select an appropriate time scale for data collection
- Learn how to do network analysis and epidemiological modeling to track down and "contact trace" superspreaders
- Use an iterative method for **finding** and **ranking** influencers
  a. Start with a seed list of influencers and snowball sample to iteratively grow the visible network based on retweet frequency
  b. Rank superspreaders by node centrality (Python)
     - Degree centrality
     - Betweenness centrality
  c. Visualize the relationships between influencers (Flourish)