RDSM2020@Coling

COLING 2020 The 28th International Conference on Computational Linguistics

Covid or not Covid? Topic Shift in Information Cascades on Twitter

13 December 2020

Liana Ermakova, Diana Nurbakova, Irina Ovchinnikova

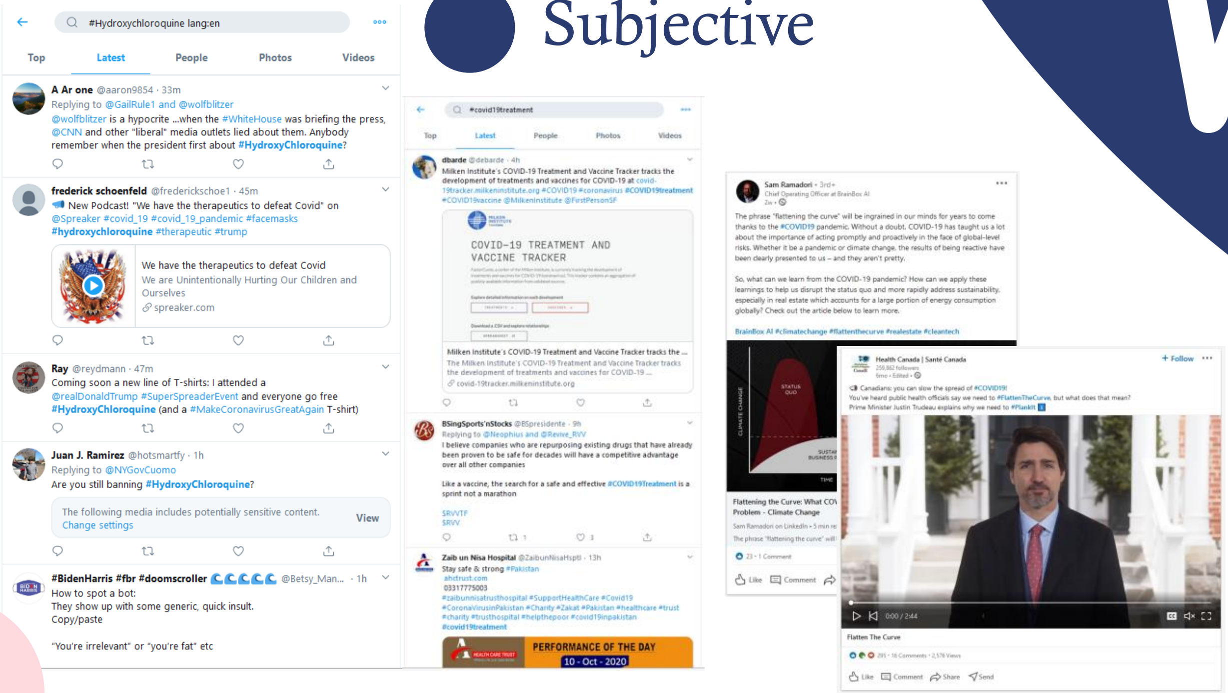
Sechenov University
HCTI
UBO Université de Bretagne Occidentale
LIRIS
INSA Institut National des Sciences Appliquées Lyon

Social Media

- Easy to understand
- Powered by Influencers
- Providing anecdotal evidence
- Subjective

Research Literature

- Complicated
- Full of terminology
- Objective



Information Cascades in Social Media

quote

Fiona Mills 🇬🇧
@Fiona_Mills

'The majority of people who actually get this infection do not die' - Chris #Whitty, Chief Medical Officer.

Not such a deadly virus then is it?

We don't need a #vaccine. We need to build up our immune systems with vitamins & minerals & take hydroxychloroquine if required.

Katie Pavlich @KatiePavlich · May 12
"The overwhelming majority of people recover from this virus." - Dr. Fauci

9:56 PM · May 12, 2020 · Twitter Web App

24 Retweets **1** Quote Tweet **38** Likes

reply

Katie Pavlich
@KatiePavlich

"The overwhelming majority of people recover from this virus." - Dr. Fauci

6:54 PM · May 12, 2020 · Twitter Web App

22.2K Retweets **2.2K** Quote Tweets **64.8K** Likes

quote

Katie Pavlich @KatiePavlich · May 12
"The overwhelming majority of people recover from this virus." - Dr. Fauci
5.3K 24.4K 64.8K

This Tweet is from a suspended account. Learn more

the smocking gun
@KalaikiMele

Replying to @WillCumberland1 @KatiePavlich and @hollandcourtney

The reason the right loves #Hydroxychloroquine so much is because it's what morons think a big sciencey-word sounds like 😂

9:21 PM · May 12, 2020 · Twitter for iPhone

Original user generated content ≠ retweet

Godgivenrights
@Godgivenrights2

Replying to @thegordonkerr @USAmbUK and 2 others

What a moronic statement. I can keep myself alive and I don't need you or any State to do it for me. Thanks for your concern and with the greatest of respect, mind your own bloody business

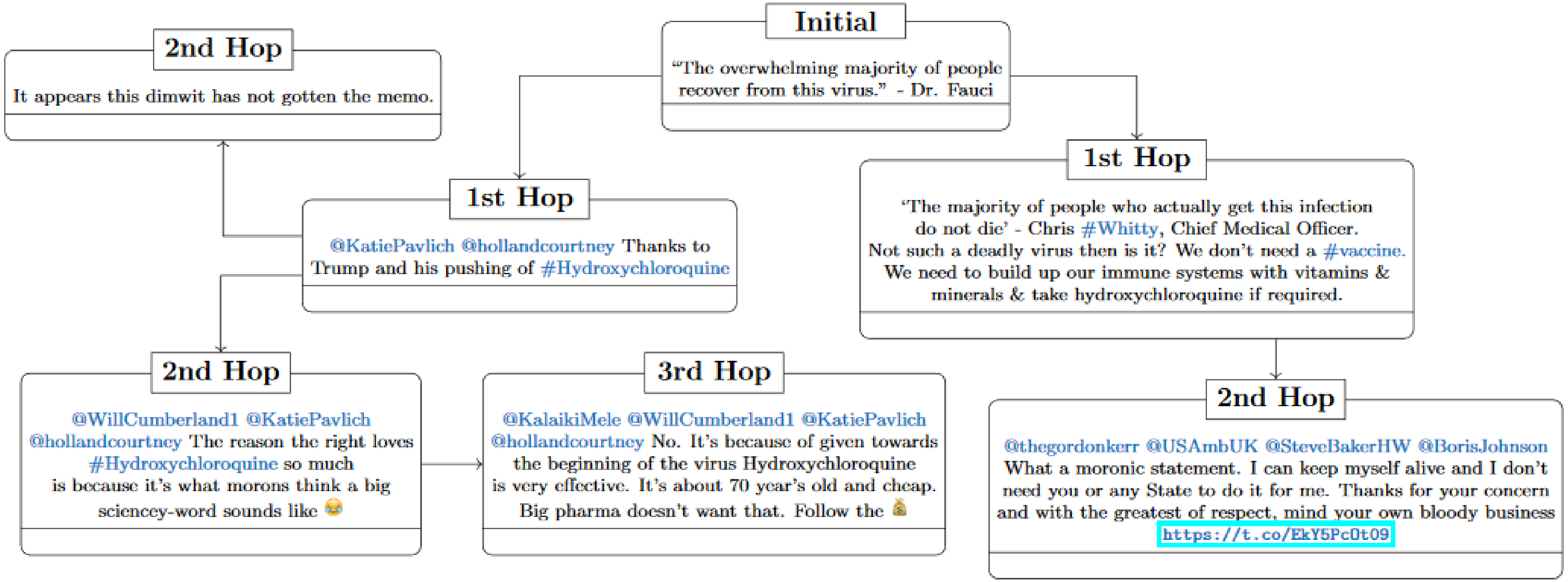
Fiona Mills 🇬🇧 @Fiona_Mills · May 12
'The majority of people who actually get this infection do not die' - Chris #Whitty, Chief Medical Officer.

Not such a deadly virus then is it?

We don't need a #vaccine. We need to build up our immune systems with vitamins & minerals & take hydroxychloroquine if required.
twitter.com/KatiePavlich/s...

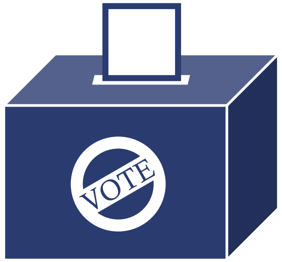
5:57 PM · May 13, 2020 · Twitter Web App

Information Cascades in Social Media

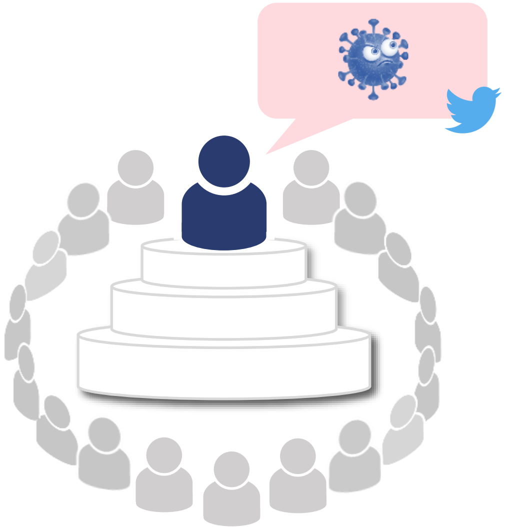


▸ D.M. Romero, W. Galuba, S. Asur, and B.A. Huberman. 2011. Influence and Passivity in Social Media. In Dimitrios Gunopulos, Thomas Hofmann, Donato Malerba, and Michalis Vazirgiannis, editors, *Machine Learning and Knowledge Discovery in Databases*, pages 18–33, Berlin, Heidelberg. Springer Berlin Heidelberg.

Public Figures on Social Media



politics



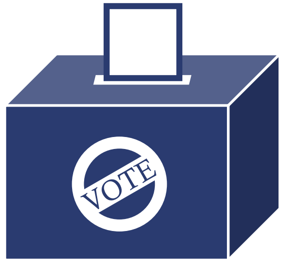
journalism

healthcare

art & show business

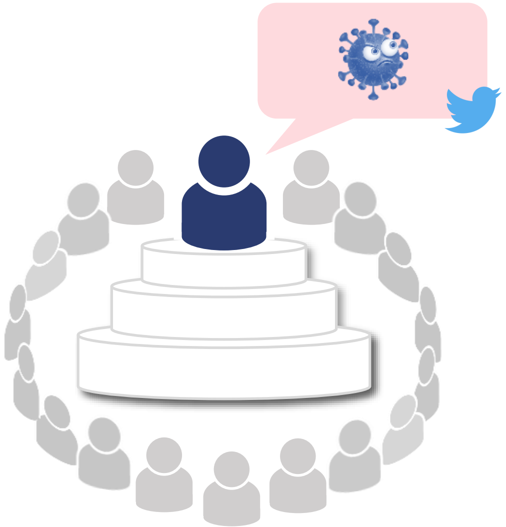
business

Public Figures on Social Media



politics

Donald Trump (US President)



journalism

Nail Cavuto (Fox News)

healthcare

Dr. Anthony Fauci

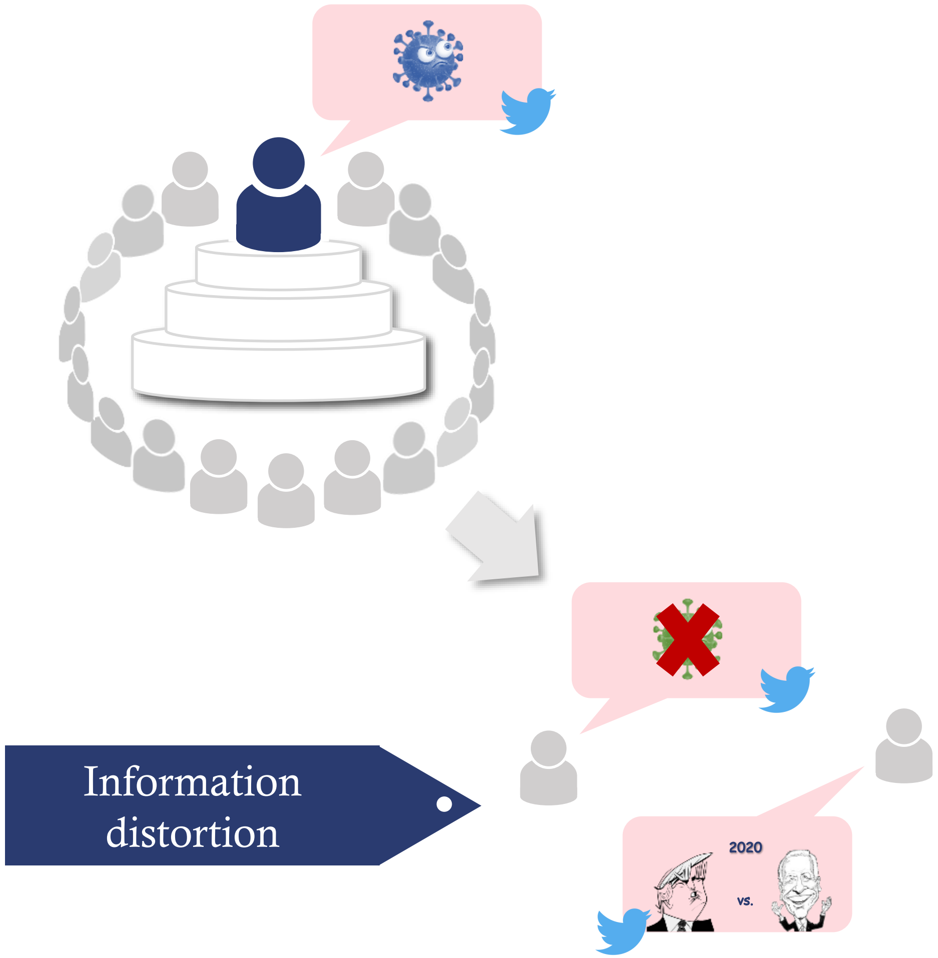
art & show business

Scott Adams (Dilbert comics)

business

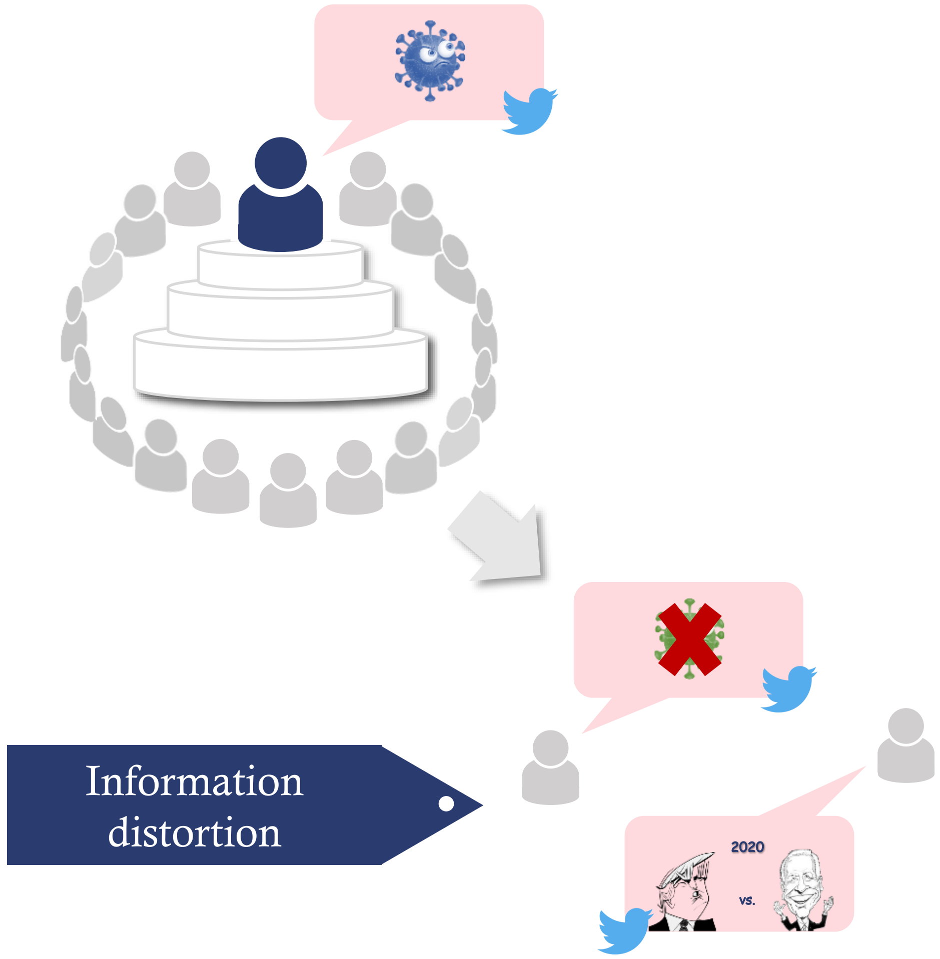
Bill Gates

Information Distortion on Social Media



- J.S. Brennen, F.M. Simon, P.N. Howard, and R.K. Nielsen. 2020. Types, Sources, and Claims of COVID-19 Misinformation. page 13.
- J. Shin, L. Jian, K. Driscoll, and F. Bar. 2018. The diffusion of misinformation on social media: Temporal pattern, message, and source. *Computers in Human Behavior*, 83:278–287, June.
- G. Pennycook, J. McPhetres, Y. Zhang, J.G. Lu, and D.G. Rand. 2020. Fighting COVID-19 Misinformation on Social Media: Experimental Evidence for a Scalable Accuracy-Nudge Intervention. *Psychological Science*, 31(7):770–780, July. Publisher: SAGE Publications Inc.
- Horta Manoel Ribeiro, Kristina Gligoric, and Robert West. 2019. Message distortion in information cascades. page 681–692.

Objectives



What are the Public Figures (PF) **tweets** on healthcare topics that **generate information cascades**?

How does a **transformation of the initial tweet** involve misinformation (topic shift)?

Plan

PART 1

Data Collection

☑ Collection of information cascades from the tweets about controversial COVID-19 treatments

PART 2

Topic Shift

☑ Topic shifts within information cascades

PART 3

Semantic Analysis of Topic Shifts

☑ Case study of distortion mechanisms within cascades

☑ Disease related terms and their occurrences

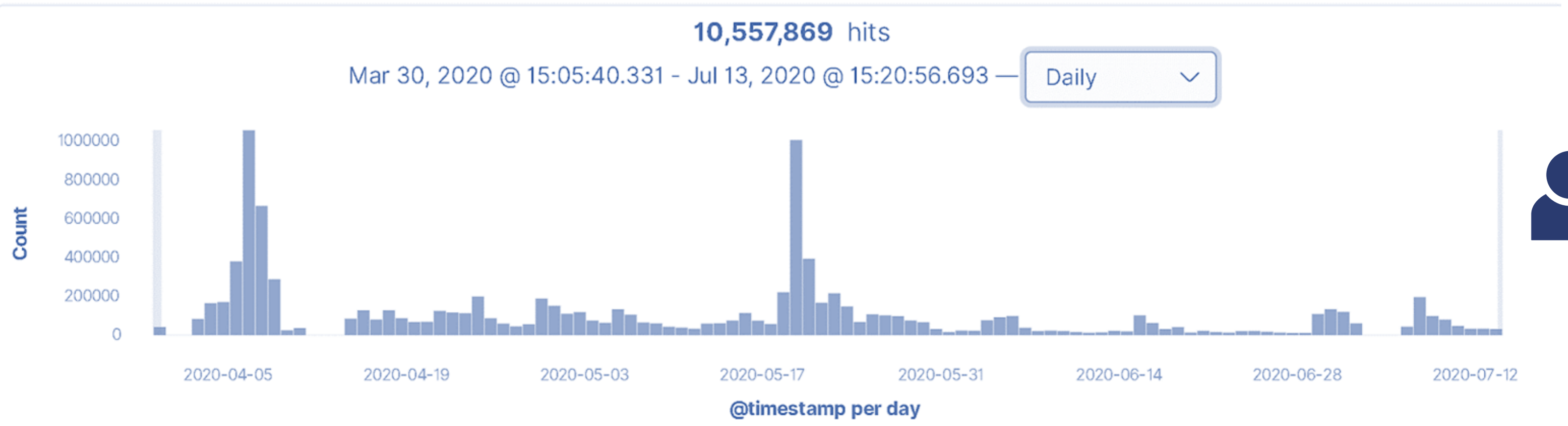
Data Collection: Controversial COVID-19 Treatments

Twitter Streaming

API

Query

science retraction, chloroquine, hydroxychloroquine, HCQ, Raoult, remdesivir, tocilizumab, favipiravir, Avigan, azithromicyn, azithromicyne, Axemal, Dolquine, Quensyl, Hydroxychloroquinum, Hydroxychloroquin, Hidroxicloroquina, Montagnier, Hydroquin, Quinoric



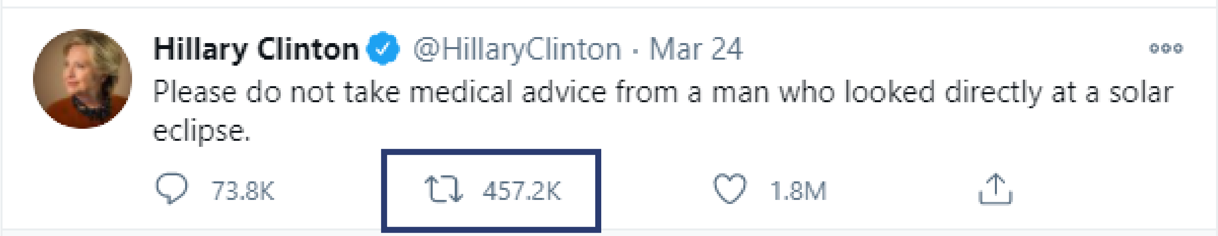
2,159,932

Marianne Noel, Liana Ermakova, Pedro Rammaciotti, Alexis Perrier, and Bilel Benbouzid. 2020. Controverse scientifique: https://dpa.hypotheses.org/controverse-scientifique

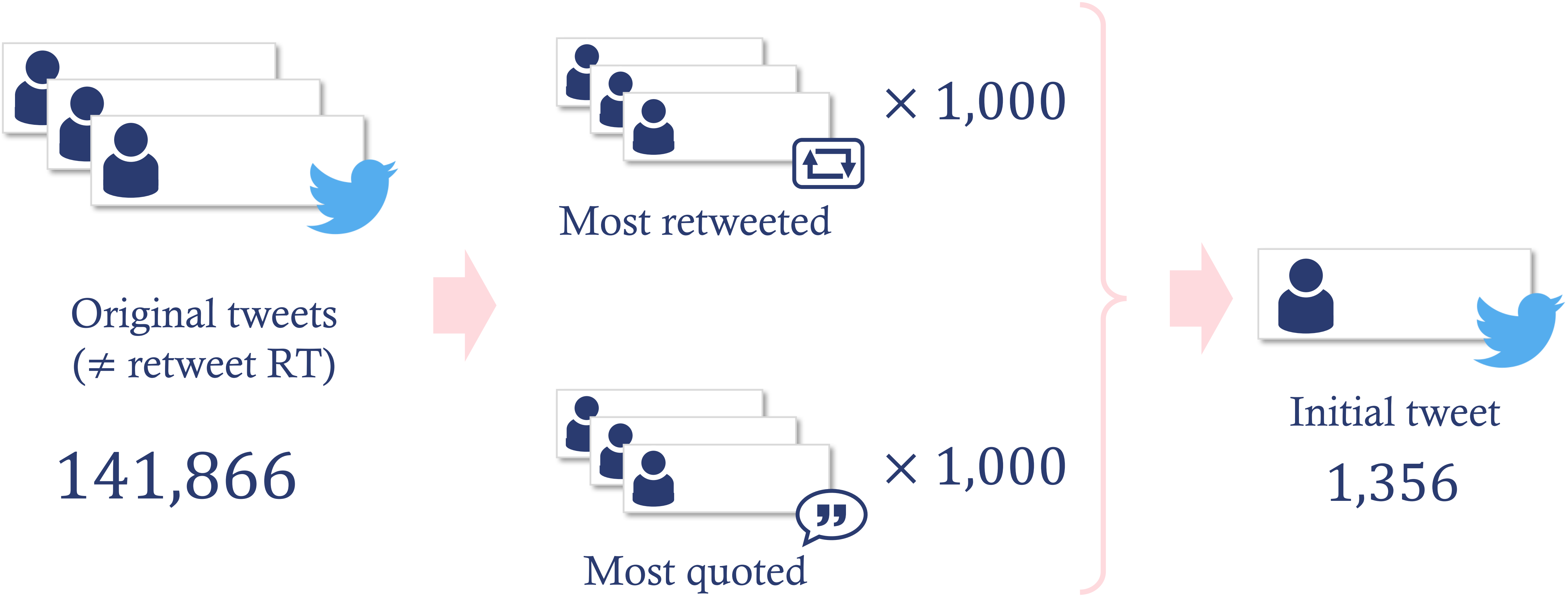
Data Collection: Controversial COVID-19 Treatments → Cascades

Original tweets
(≠ retweet RT)

141,866



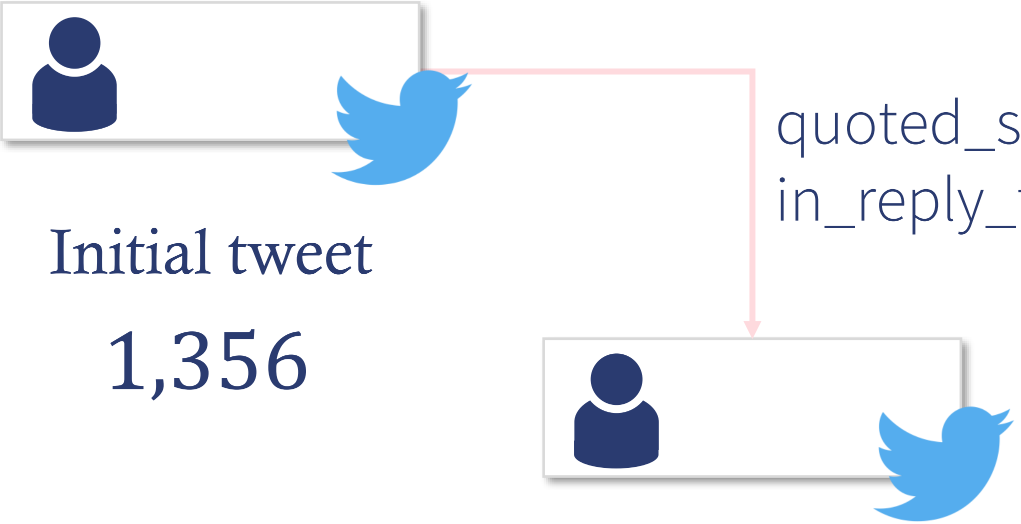
Data Collection: Controversial COVID-19 Treatments → Cascades



Data Collection: Controversial COVID-19 Treatments → Cascades

Original tweets
(≠ retweet RT)

141,866



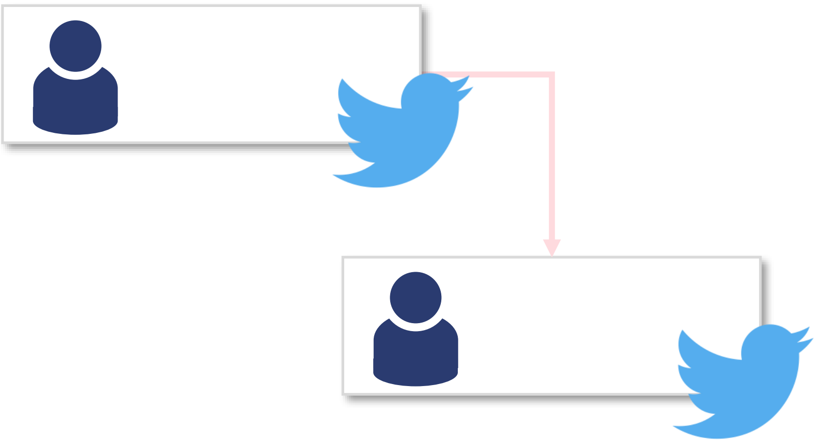
Initial tweet

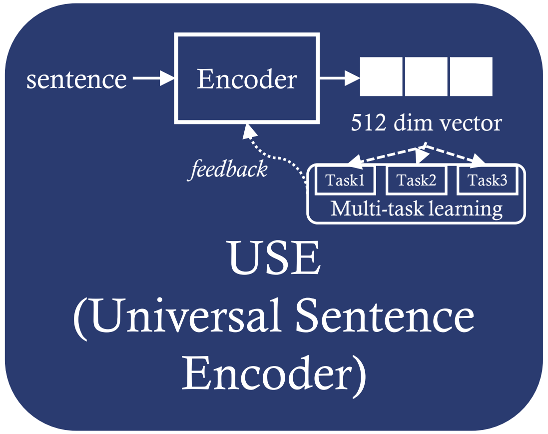
1,356

Hops

Max depth = 10 (9 hops)

Topic Shift between Hops and Initial Tweets





between neighbouring hops

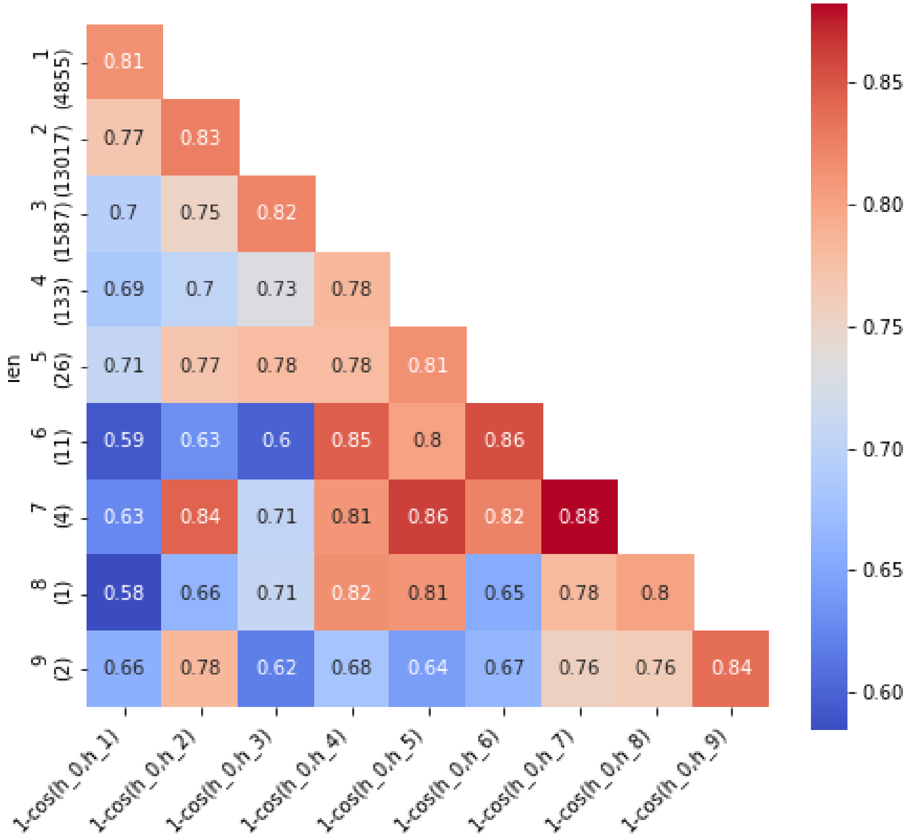
$$\Delta^{(i-1)} = 1 - \cos\left(\overrightarrow{hop_i}, \overrightarrow{hop_{i-1}}\right) = 1 - \frac{\overrightarrow{hop_i} \cdot \overrightarrow{hop_{i-1}}}{\left\|\overrightarrow{hop_i}\right\| \left\|\overrightarrow{hop_{i-1}}\right\|}$$

between a hop and initial tweet

$$\Delta^{(0)} = 1 - \cos\left(\overrightarrow{hop_i}, \overrightarrow{hop_0}\right) = 1 - \frac{\overrightarrow{hop_i} \cdot \overrightarrow{hop_0}}{\left\|\overrightarrow{hop_i}\right\| \left\|\overrightarrow{hop_0}\right\|}$$

Daniel Cer, Yinfei Yang, Sheng-yi Kong, Nan Hua, Nicole Limtiaco, Rhomni St. John, Noah Constant, Mario Guajardo-Cespedes, Steve Yuan, Chris Tar, Brian Strope, and Ray Kurzweil. 2018. Universal Sentence En-coder for English. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 169–174, Brussels, Belgium, November. Association for Computational Linguistics.

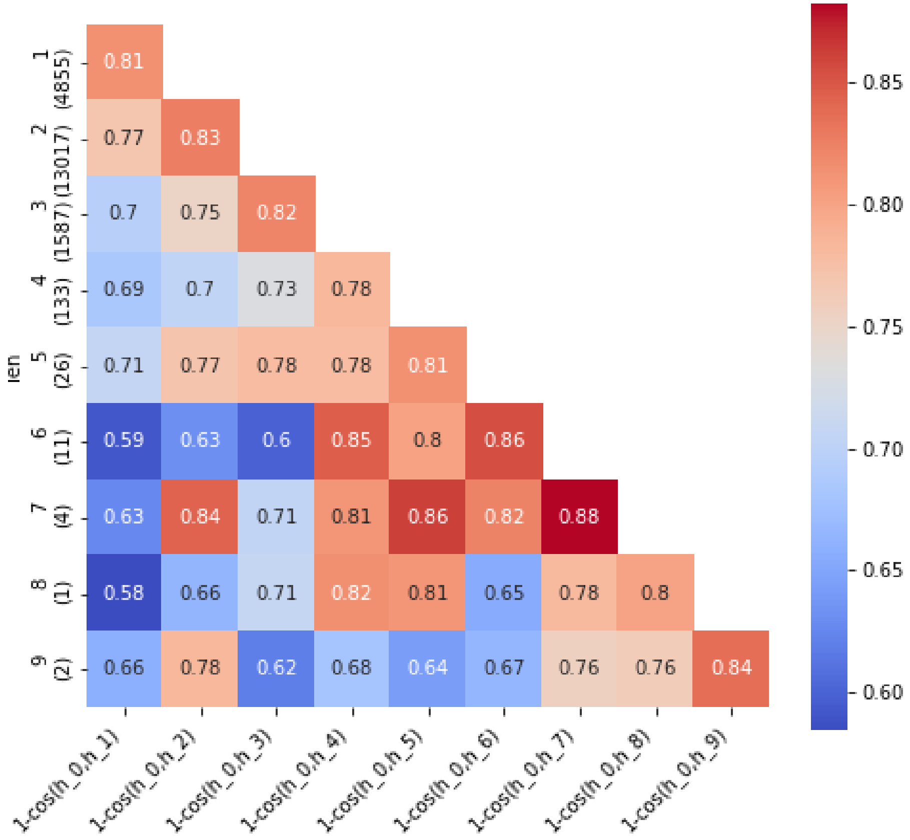
Topic Shift between Hops and Initial Tweets: Results



Current vs. Initial Hops

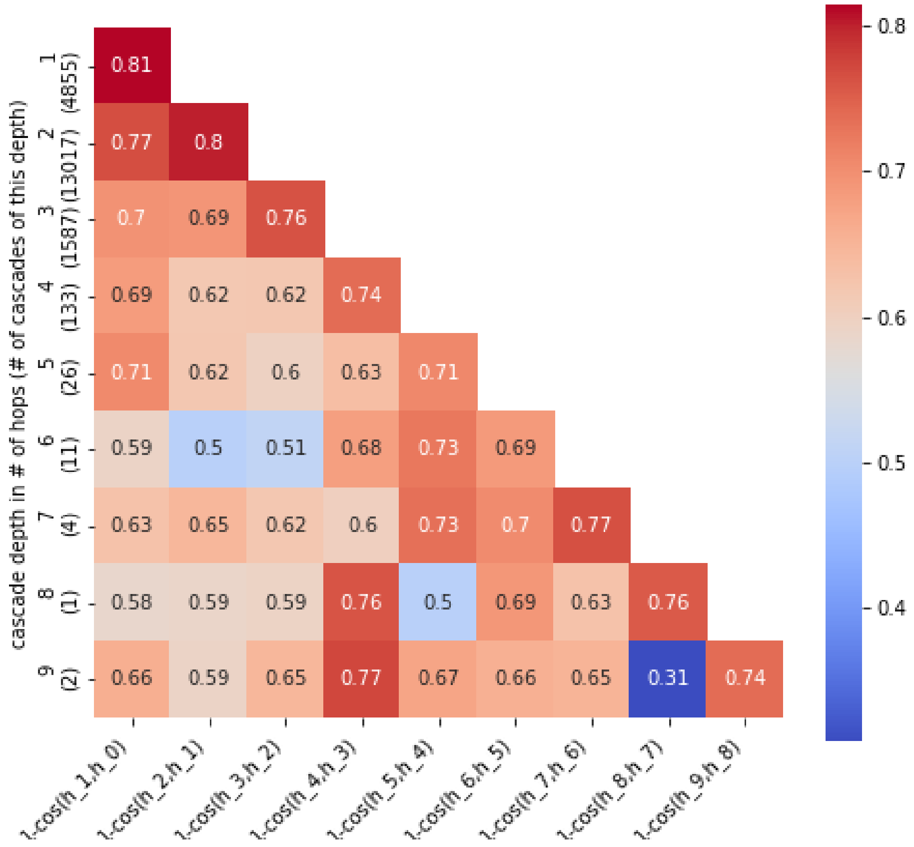
Topic Shift between Hops and Initial Tweets: Results

Accumulation of semantic difference



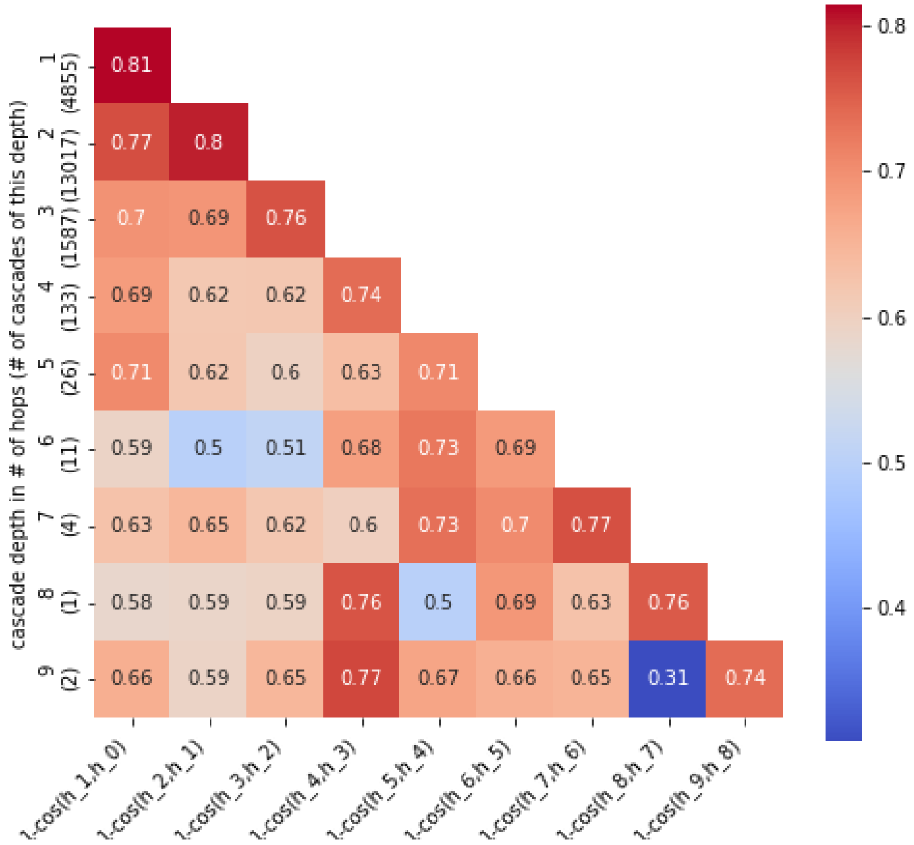
Current vs. Initial Hops

Topic Shift between Hops and Initial Tweets: Results

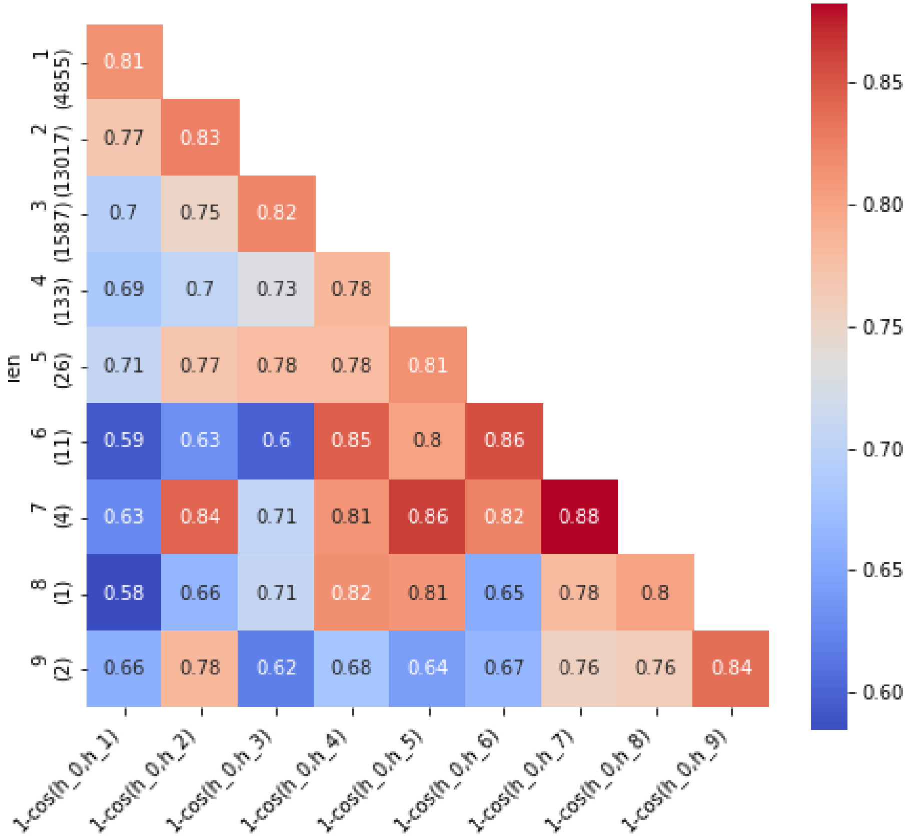


Neighbouring Hops

Topic Shift between Hops and Initial Tweets: Results

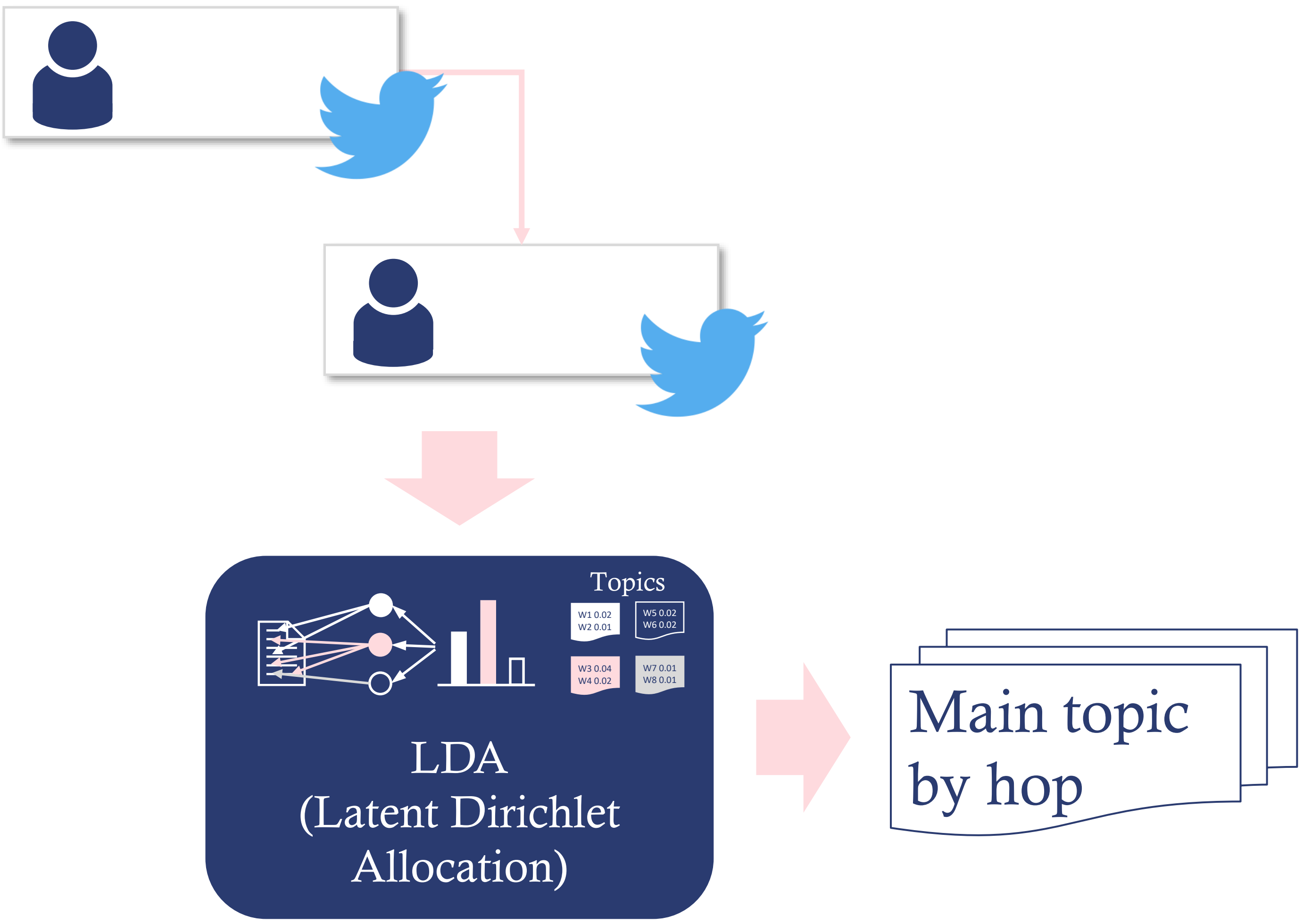


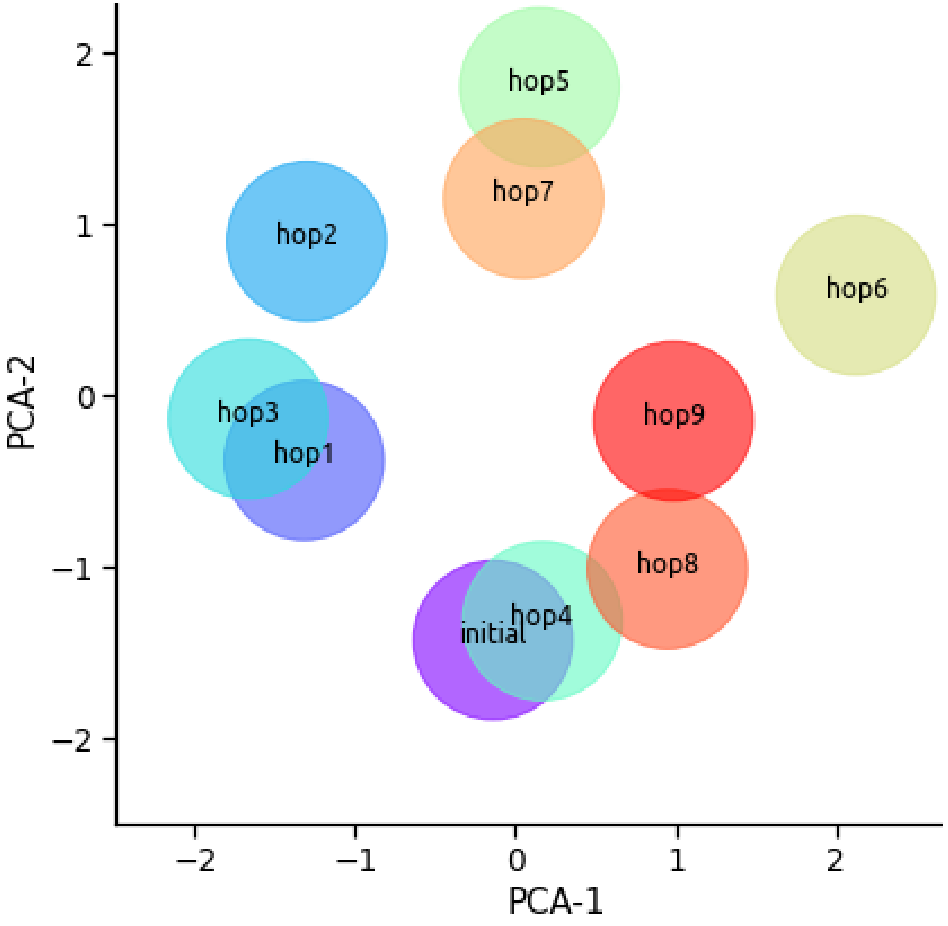
Neighbouring Hops



Current vs. Initial Hops

Topic Modelling of Hops





LDA

Matthew D. Hoffman, David M. Blei, and Francis R. Bach. 2010. Online learning for latent dirichlet allocation. In *Advances in Neural Information Processing Systems 23: 24th Annual Conference on Neural Information Processing Systems 2010. Proceedings of a meeting held 6-9 December 2010, Vancouver, British Columbia, Canada*, pages 856–864. Curran Associates, Inc.

Topic Modelling of Hops

Initial

hydroxychloroquine, **trump**, coronavirus, president, say, **people**, news, good, state, **covid19**

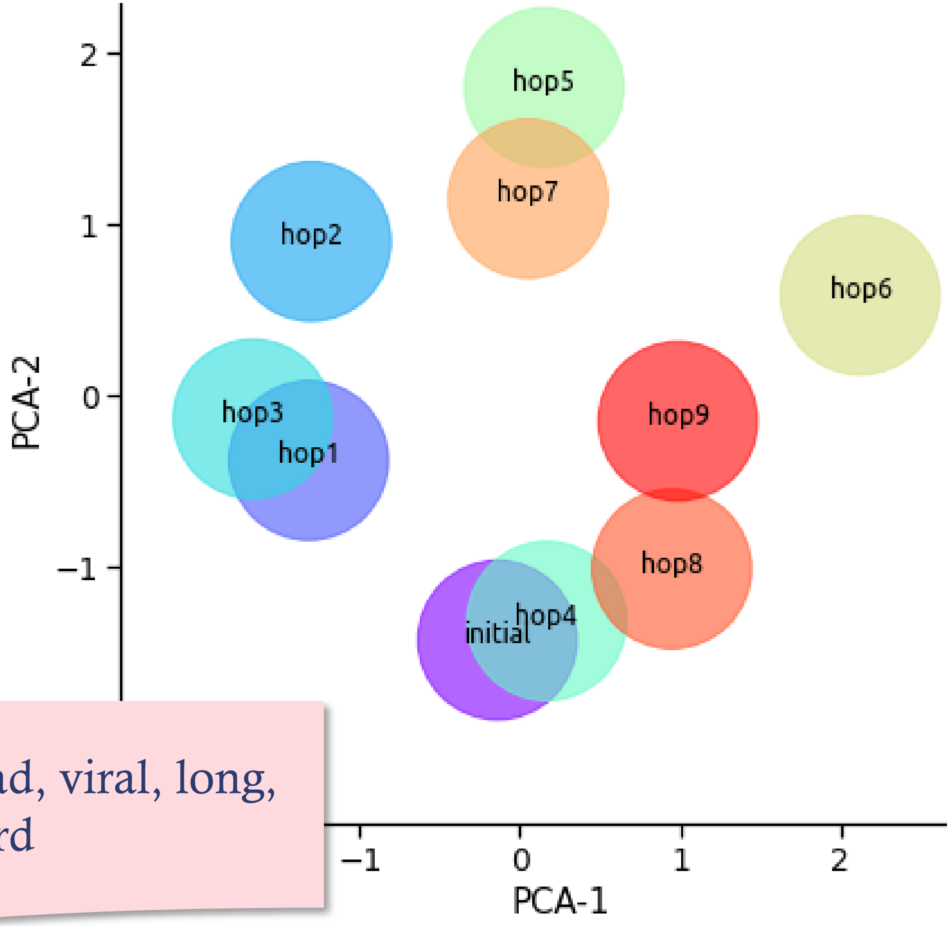
Hop 4

hydroxychloroquine, use, **covid19**, **trump**, **people**, know, drug, make, fda, effect

Hop 6

hydroxychloroquine, texas, diagnosis, load, viral, long, patient, reduce, **covid19**, board

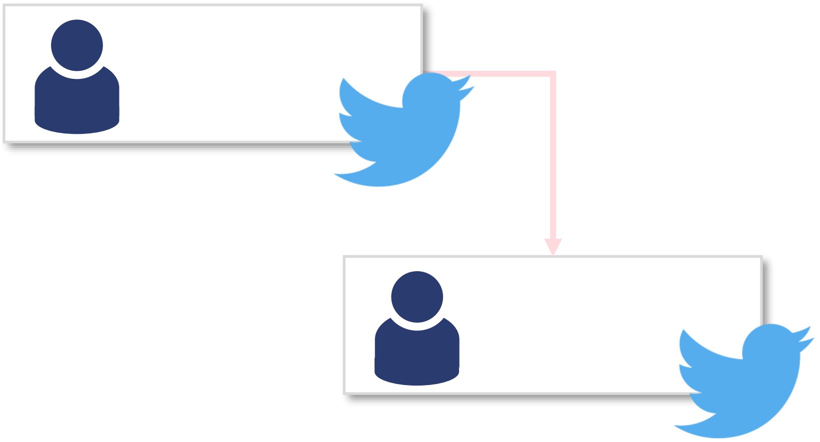
The word common to topics of all hops



LDA

Matthew D. Hoffman, David M. Blei, and Francis R. Bach. 2010. Online learning for latent dirichlet allocation. In *Advances in Neural Information Processing Systems 23: 24th Annual Conference on Neural Information Processing Systems 2010. Proceedings of a meeting held 6-9 December 2010, Vancouver, British Columbia, Canada*, pages 856–864. Curran Associates, Inc.

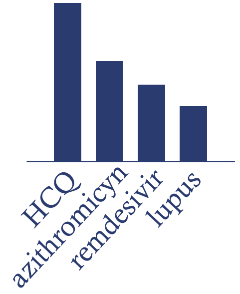
Semantic Analysis



Manual semantic analysis

+

Key term distribution in cascades



+

Context and substitution analysis

Term substitutions

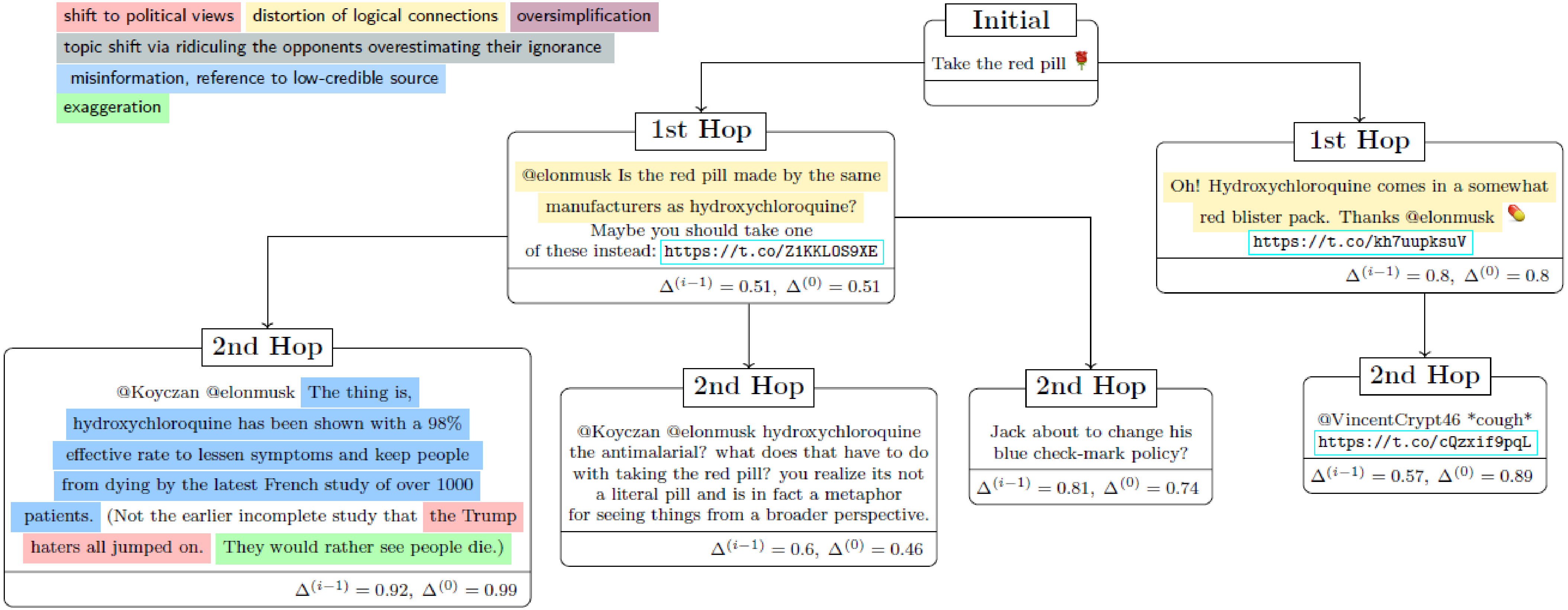
Verification of logical relationships among hops

Distortions w.r.t. the initial tweets

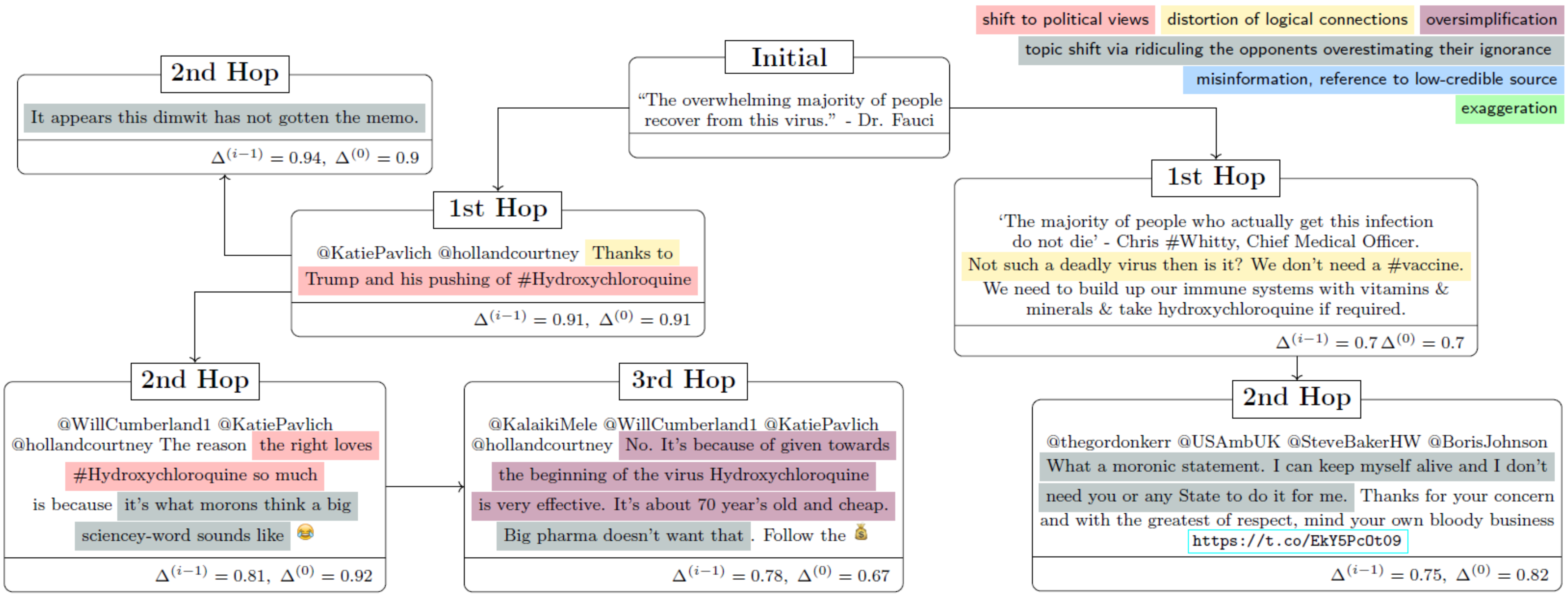
Semantic Analysis: Information Distortion within Information Cascades

- Erroneous logical connections
- Misuse of medical concepts
- Distortion of connections between medical concepts
- Insertion of erroneous conclusions
- Substitution
- Oversimplification
- Omission of facts
- Overgeneralisation
- Exaggeration

Semantic Analysis: Information Distortion within Information Cascades



Semantic Analysis: Information Distortion within Information Cascades



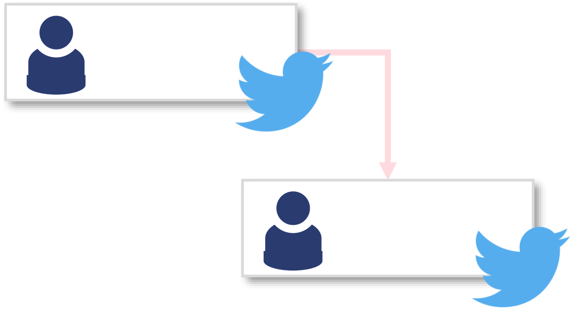
Disease Related Terms (DRTs) and Their Occurrence

'plague', 'swine.flu', 'bird.flu', 'avian.flu', 'avian.influenza', 'zika', 'zika.fever', 'ebola', 'dengue', 'dengue.fever', 'meningitis', 'yellow.fever', 'jaundice', 'cholera', 'chicken.pox', 'encephalitis', 'measles', 'aids', 'koh', 'nipah', 'bunyavirus', 'lyme.disease', 'heart.disease', '(?<!in|mi|an|wu|na|ne)flu(?!ence|nked|sh|ke)', 'rheumatological', 'hiv', 'whipple.disease', 'whipple.s.disease', 'malaria', 'infection(?!.disease)', 'infection.disease', 'parasite', 'headache', 'cough', 'body.aches', 'auto.immune.disease', 'wuhanflu', 'wuhan.flu', 'wuflu’, '(?<!auto)immune.disease', '(?<!dis)cours(?!e)', 'sars', 'cardiac.disease', 'china.flu', 'covid', 'coronovirus', 'sars.cov.2', 'cancer', 'obesity', 'diabete’, 'china.virus', 'sinus.infection', 'rhinitis', 'illness', 'heart.attack', 'herpes’, 'corona', 'erythematosus', '\spolio\s'

“disease” → WordNet → hyponyms + Custom hyponyms → Occurrences in the cascades

6,493 lemmas

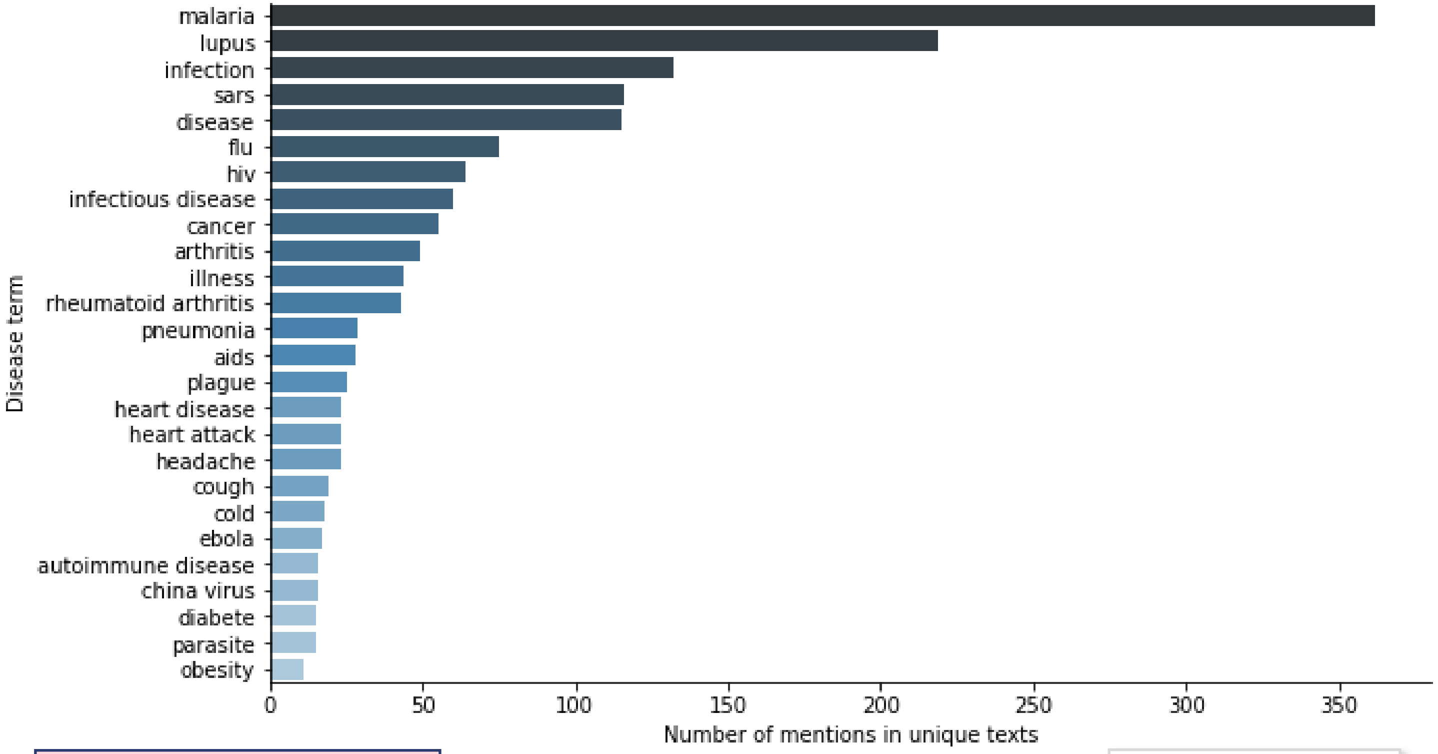
75 unique DRTs



3,939/21,585

- Princeton University "About WordNet." WordNet. Princeton University. 2010.
- George A. Miller (1995). WordNet: A Lexical Database for English. Communications of the ACM Vol. 38, No. 11: 39-41.

Disease Related Terms (DRTs) and Their Occurrence



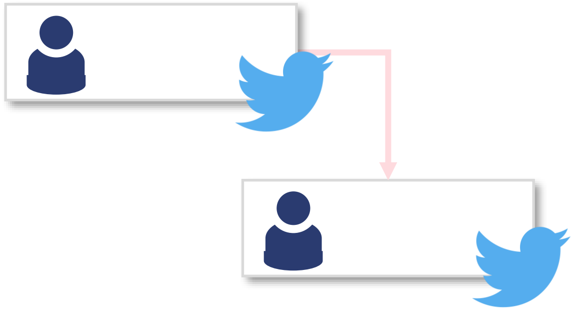
"disease"

WordNet

hyponyms + Custom hyponyms

6,493 lemmas

Occurrences in the cascades



75 unique DRTs

- Princeton University "About WordNet." WordNet. Princeton University. 2010.
- George A. Miller (1995). WordNet: A Lexical Database for English. Communications of the ACM Vol. 38, No. 11: 39-41.

Disease Related Terms (DRTs) and Their Occurrence

75 unique DRTs

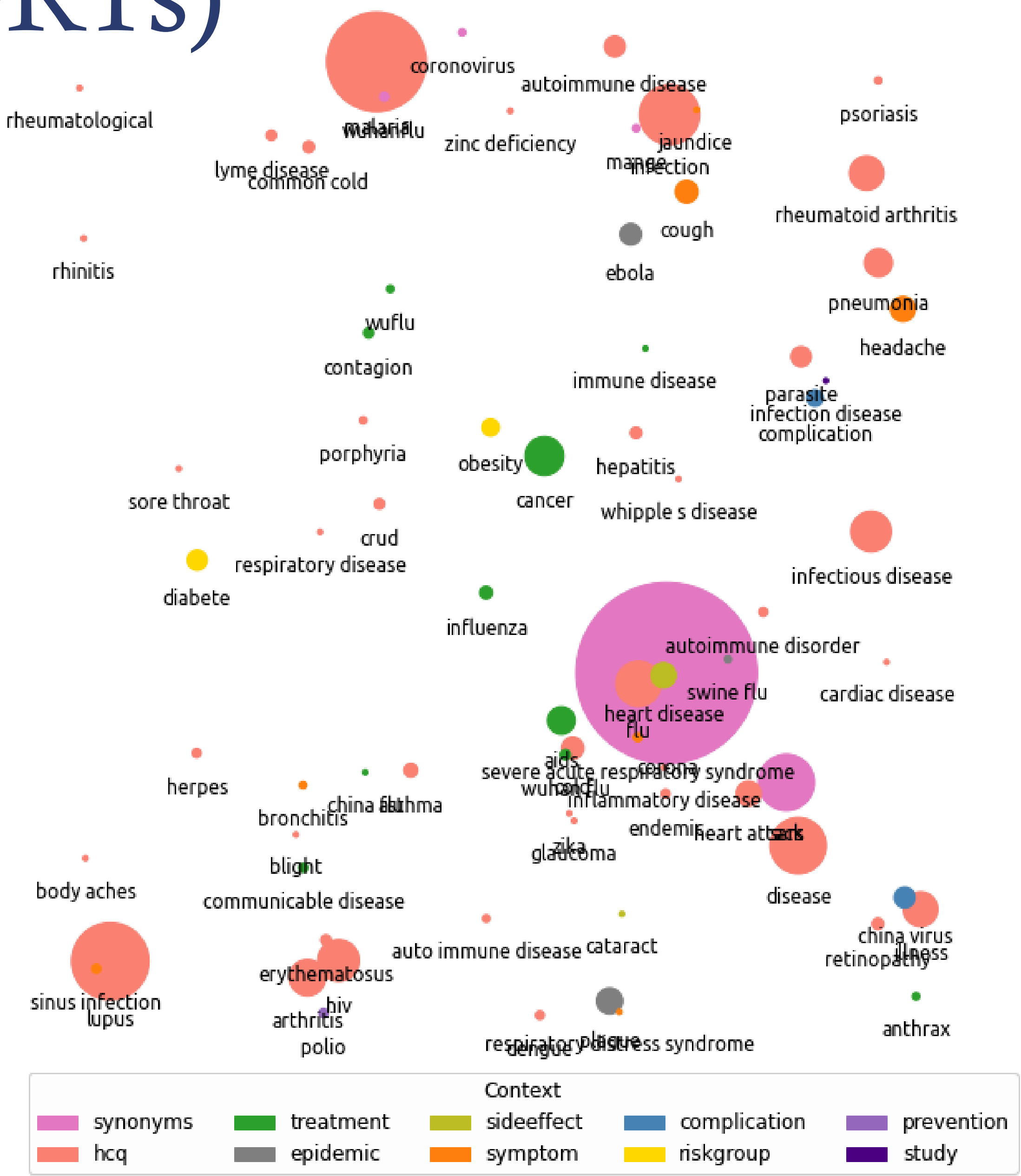
14 contexts

hcq = {hydroxychloroquine, hcq, hc, azithromycin, chloroquine, zpack, z.pac, antimalarial, zinc, sulfate, zithromax}
symptom = {symptom, congestion, blood, cough, aches, lungs, fever, antibody, headache, mucus, signs, asymptomatic, respiratory, shortness.of.breath, symptom.free, back.pain, diarrhea, nausea}
treatment = {treatment, cure, curing, treat, pill, medicament, remedy, therapy, drug, acetaminophen, prescribe, prescription, breathlessness, medications, diagnos, recovery}
prevention = {vaccin, mask, hand.wash, distanc, prevention, detection, test, cover.*mouth, self.isol, prophylaxis, immunity, stayhome, staying.home, stay.home, prophylactic, serum.test, preventative}
alternatives = {antibiotic, anti.hiv.drug, analgesics, remdesivir, favipiravir, antiviral.medication, anti.biotic}
propagation = {contaminate, contamination, propagation, viral.load, (?<!dead.zone.)\splague(?!.*dead.zone)}
study ={study, control.group, randomi.ed, research, treatment.group, trial, expert, scientific.evidence, success.rate, science, protocol, effective.rate, placebo}
complication = {ventilator, complication, transfusion, coma, hospitalization, death, severe.case, critically, critical.condition, severe, urgent.care.center, emergency, icu, intensive.care.unit}
epidemic = {epidemic, pandemic, plague, zika, ebola, lockdown, locked.down, outbreak, swine.flu}
sideeffect ={side.effect, heart.disease, cardiac.problem, hallucination, psychiatric.symptom, vision.loss, vomiting, loss.of.appetite, dizziness, slow.heartbeat, heart.failure, swelling.ankles}
riskgroup = {elderly, diabete, obesity, obese, asthma, comorbidity, 60.plus, 60.year}
synonyms = {corona, wuhan.virus, wuhan.disease, sars.cov.2, covid19, covid, c19, coronavirus, chinese.flu, china.flu, cv.19, sars.cov, sars, chinese.plague, coronahoax, wuhanflu}
plaguereference = {dead.zone.*plague, 2003.*plague, plague.*dead.zone, plague.*2003, dead.zone}
otherissues = {shortage, economic.crisis, economy.crisis}

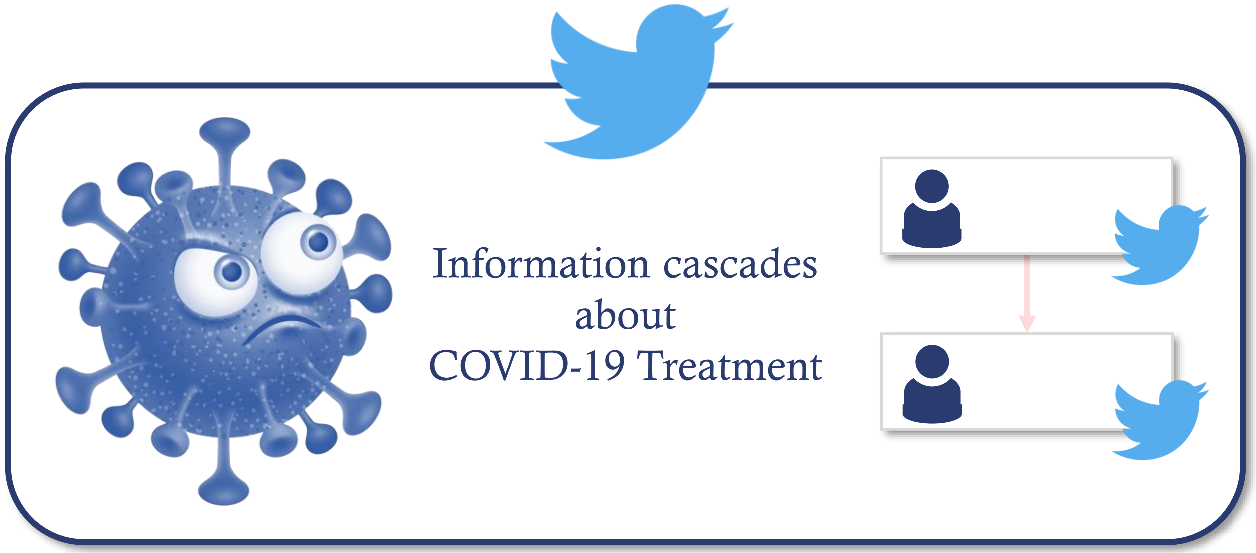
Disease Related Terms (DRTs) and Their Occurrence

75 unique DRTs

14 contexts



Conclusions



Tweets posted by Public Figures generated multiple information cascades

Topic Shift as a common phenomenon of information cascades about Covid-19 treatments

Substitution & confusion of medical terms

Politisation of the discussions

Oversimplification & distortion of logical links between fact