Master in Computational Science and Engineering
Mathematics Section, EPFL

SUPERVISORS

Dr. Erik Hemberg
Prof. Una-May O'Reilly
Prof. Robert West

Describing information influence in social media with coupling inference methods

Master Thesis Defense

March 6th, 2023

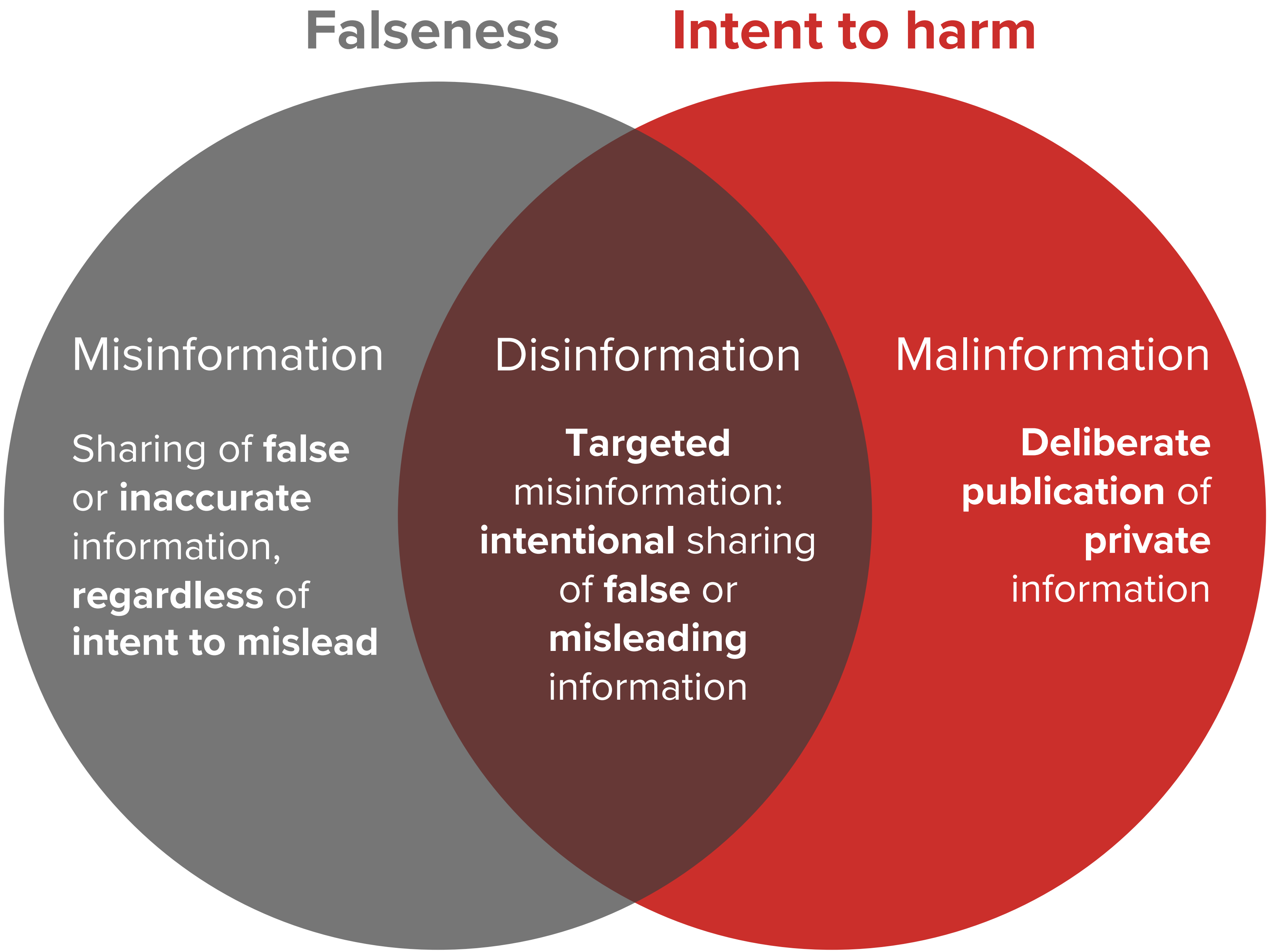
Ecole Polytechnique Fédérale de Lausanne,
Massachusetts Institute of Technology

STUDENT

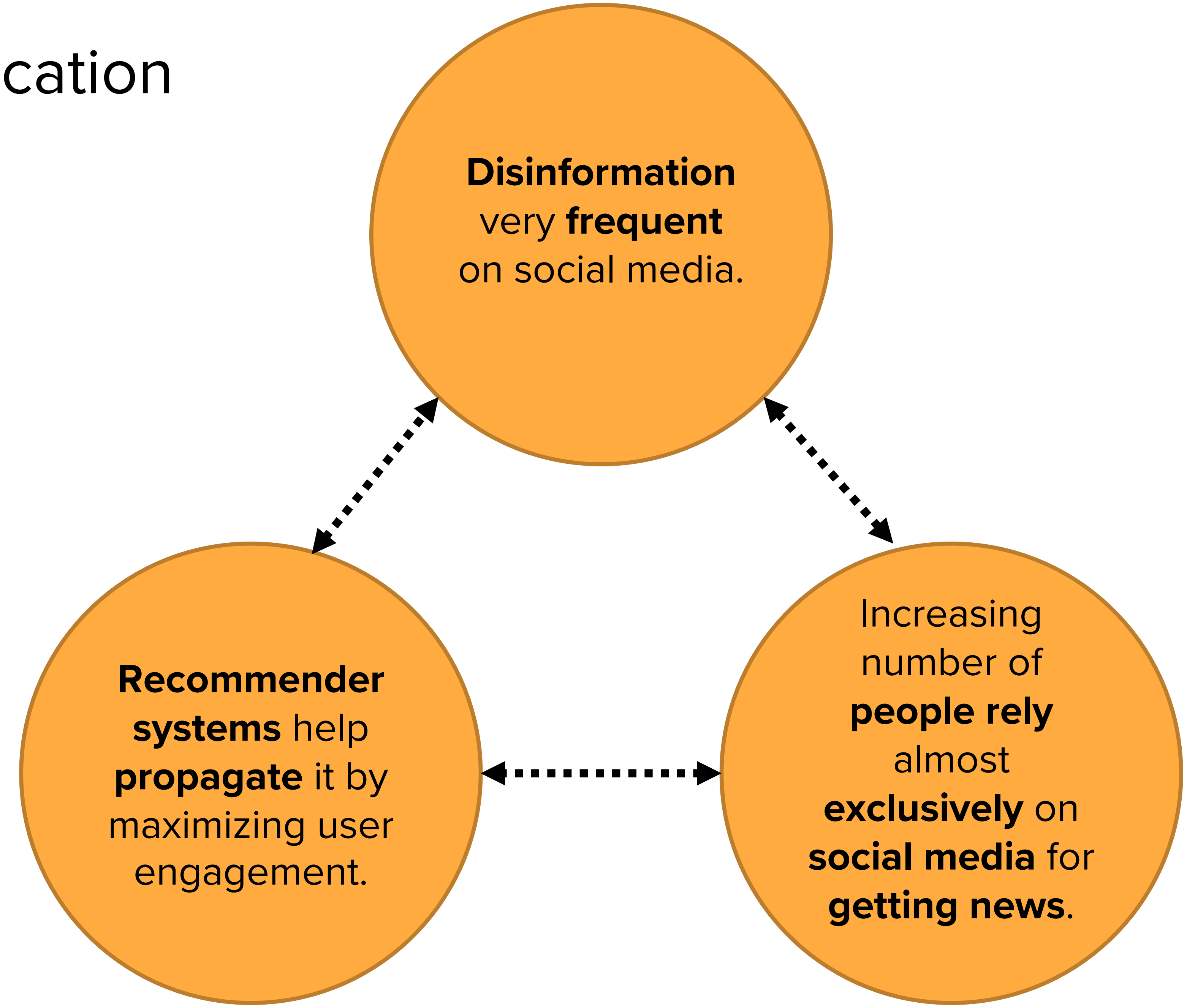
Cyril Vallez

Introduction

Terminology



Circle of amplification



Impact of disinformation

- Confused the public, led to **political inaction**, and to the **rejection** of mitigation **policies**

- **Lost of trust** in government, scientific community, medicine...

- **Polarization** and **social fracture**

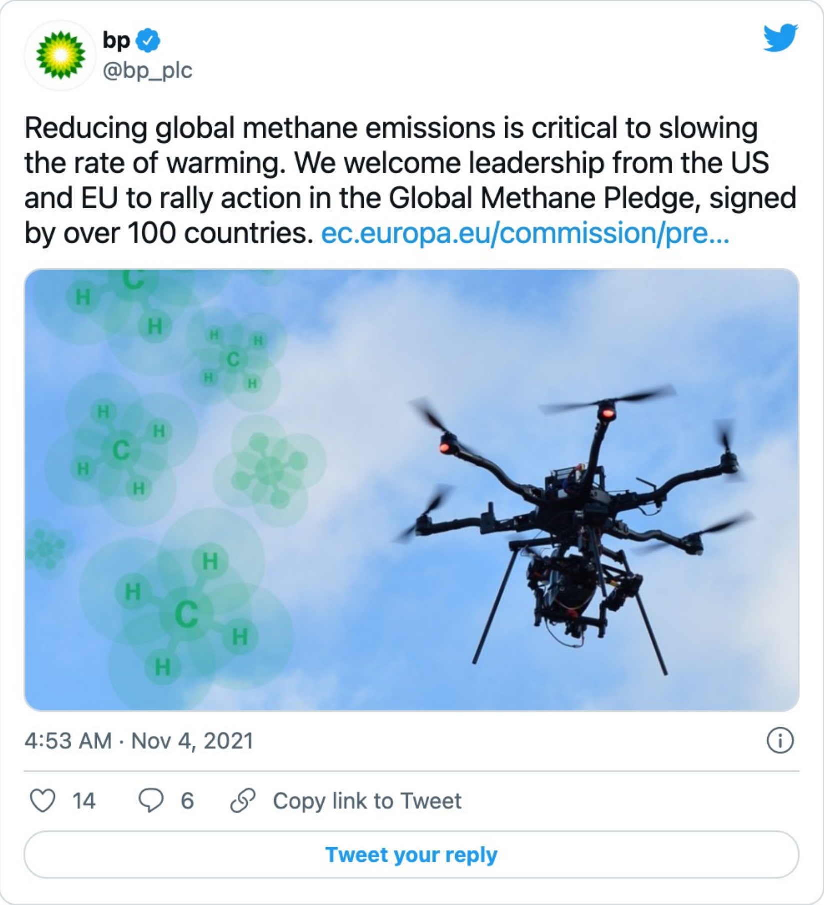
Image from nrdc.org

From Whom?

- Individuals

From Whom?

- Individuals
- Companies



From Whom?

- Individuals
- Companies
- Institutions

American Petroleum Institute
@APIenergy

We will continue to focus on industry actions and innovations that will help reduce emissions, while producing reliable, affordable natural gas and oil.

3:10 PM · Nov 8, 2021

8 3 Copy link to Tweet

Tweet your reply

From Whom?

- Individuals
- Companies
- Institutions
- States

Global Times

OPINION / EDITORIAL

China keeps its promises on climate, so should the US: Global Times editorial

By Global Times

Published: Nov 11, 2021 10:26 PM

CGTN See the difference

Home China World Politics Business Opinions Tech & Sci Culture Sports Travel Nature Picture Video Live TV Specials Learn Chinese

China 22:35, 03-Nov-2021 Translate

China tells U.S. to take actions, not empty words on climate change

CGTN

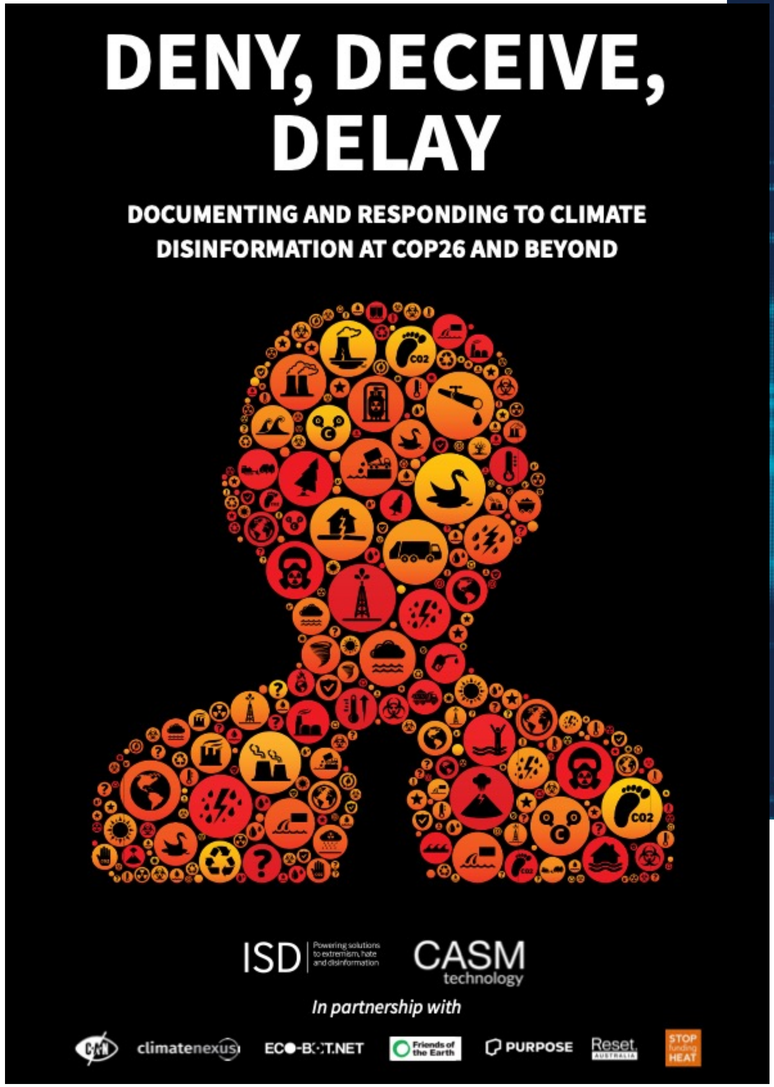
Share

China-controlled media

Goal & Motivation

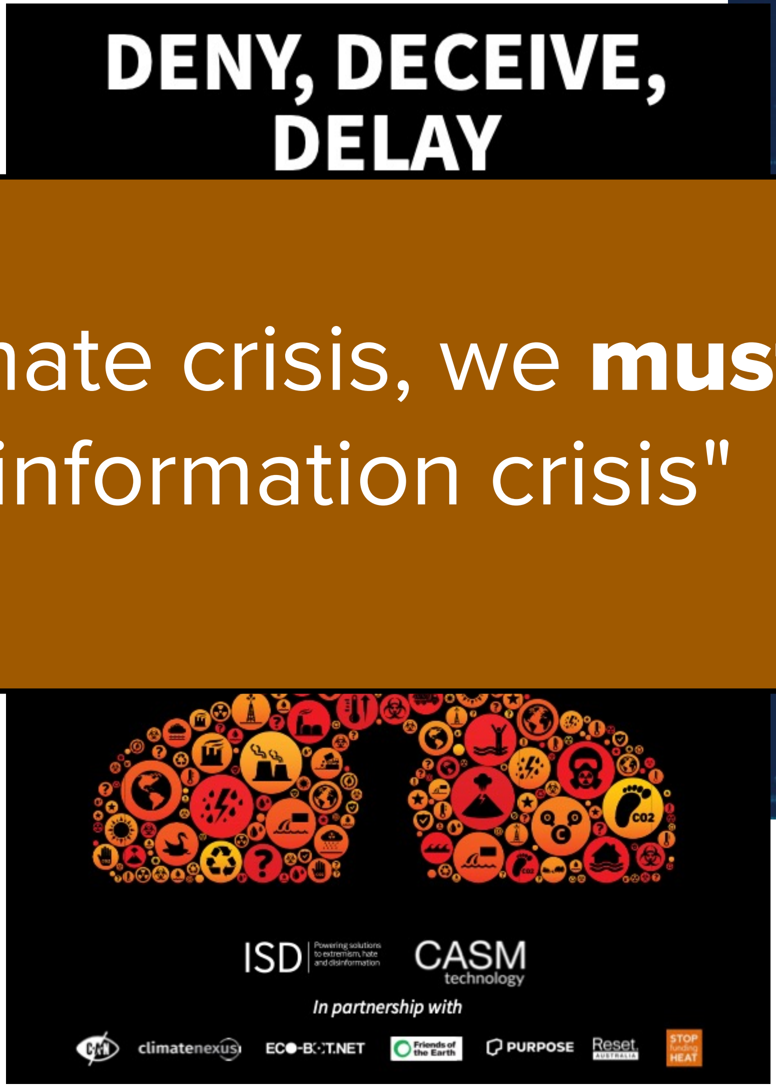
Disinformation campaigns

- **Virulent, organized, large scale disinformation operation** by coordinated actors
- **Flooding the zone**
- **Deny responsibility** in assassination, **opposition** to climate **regulations**



Disinformation campaigns

- **Virulent, organized, large scale disinformation operation** by
-
- **Deny responsibility** in assassination, **opposition** to climate **regulations**



"To solve the climate crisis, we **must** also tackle the information crisis"

Modeling influence

- Create a **system** able to **capture influence attempts** relating to (climate change) **disinformation** between users on **Twitter**.

- **Quantify** the **influence** between users on **Twitter**

- Derive **new influence measures** describing the amount of **disinformation** shared by the users on **Twitter**.

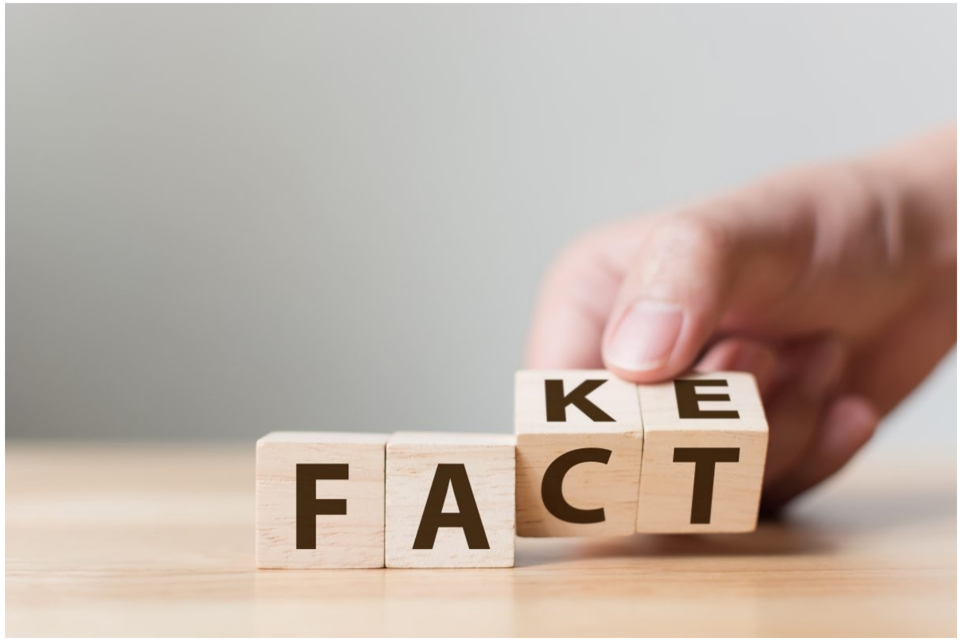


Image from commonexperience.edu

Contribution

Contribution

- **Create** and curate **2 new datasets** of tweets around **COP26** and **COP27**

Contribution

- **Create** and curate **2 new datasets** of tweets around **COP26** and **COP27**
- **Derive** new **influence measures**

Contribution

- **Create** and curate **2 new datasets** of tweets around **COP26** and **COP27**
- **Derive** new **influence measures**
- **Study** existing and new **influence measures** and how they are **correlated**

Contribution

- **Create** and curate **2 new datasets** of tweets around **COP26** and **COP27**
- **Derive** new **influence measures**
- **Study** existing and new **influence measures** and how they are **correlated**
- **Sensitivity** analysis of **key-components** of the framework

Contribution

- **Create** and curate **2 new datasets** of tweets around **COP26** and **COP27**
- **Derive** new **influence measures**
- **Study** existing and new **influence measures** and how they are **correlated**
- **Sensitivity** analysis of **key-components** of the framework
- **Generalize** on different **types** of **events**

Methodology

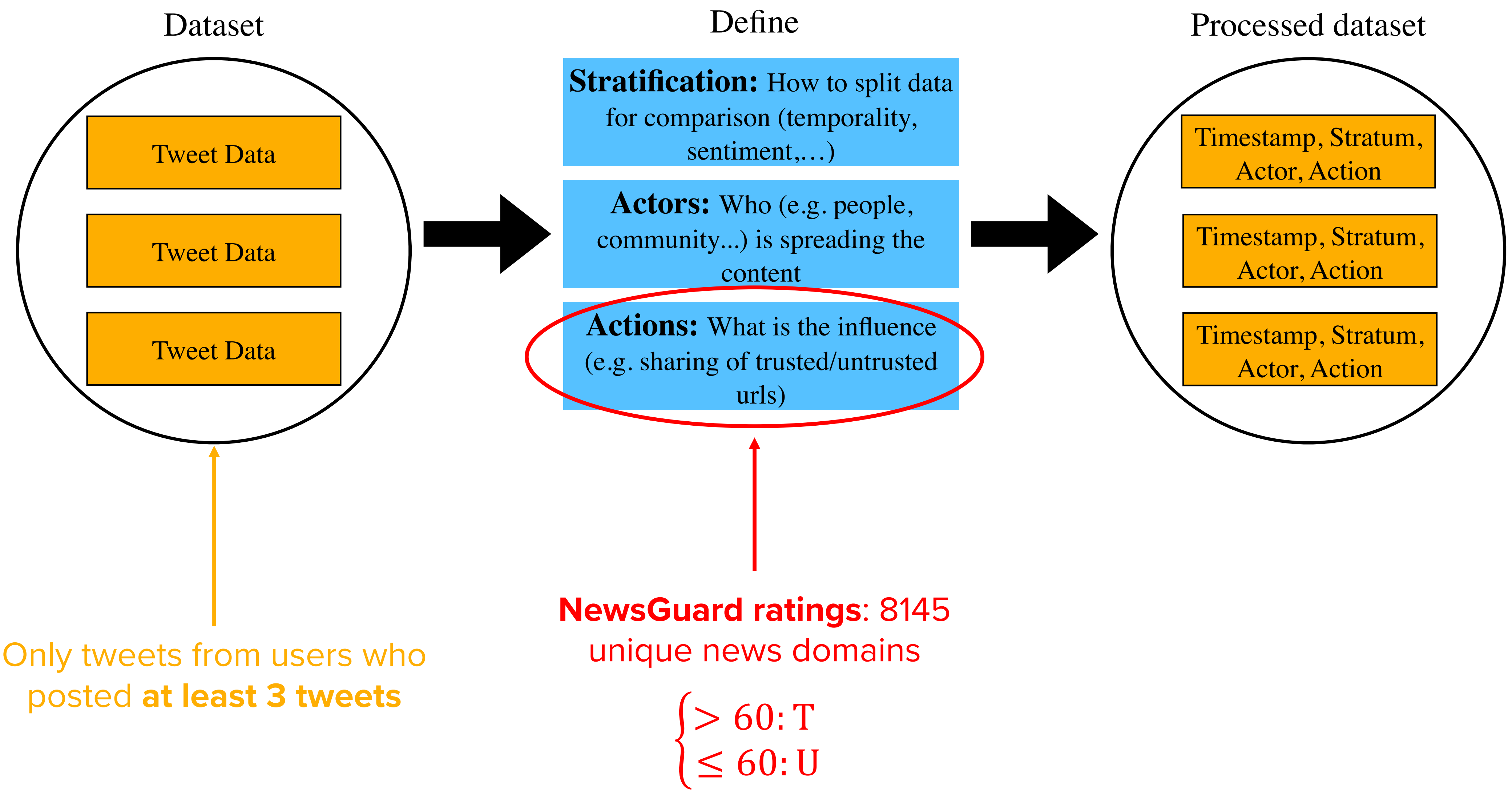
Datasets

- **COP26:** all **climate-related** tweets containing a **URL** around COP26 (2021).
- **COP27:** all **climate-related** tweets containing a **URL** around COP27 (2022).
- **Control:** all **climate-related** tweets containing a **URL** at random dates.
- **Skripal:** all tweets related to the **Skripal poisoning** in 2018 and containing a **URL**.

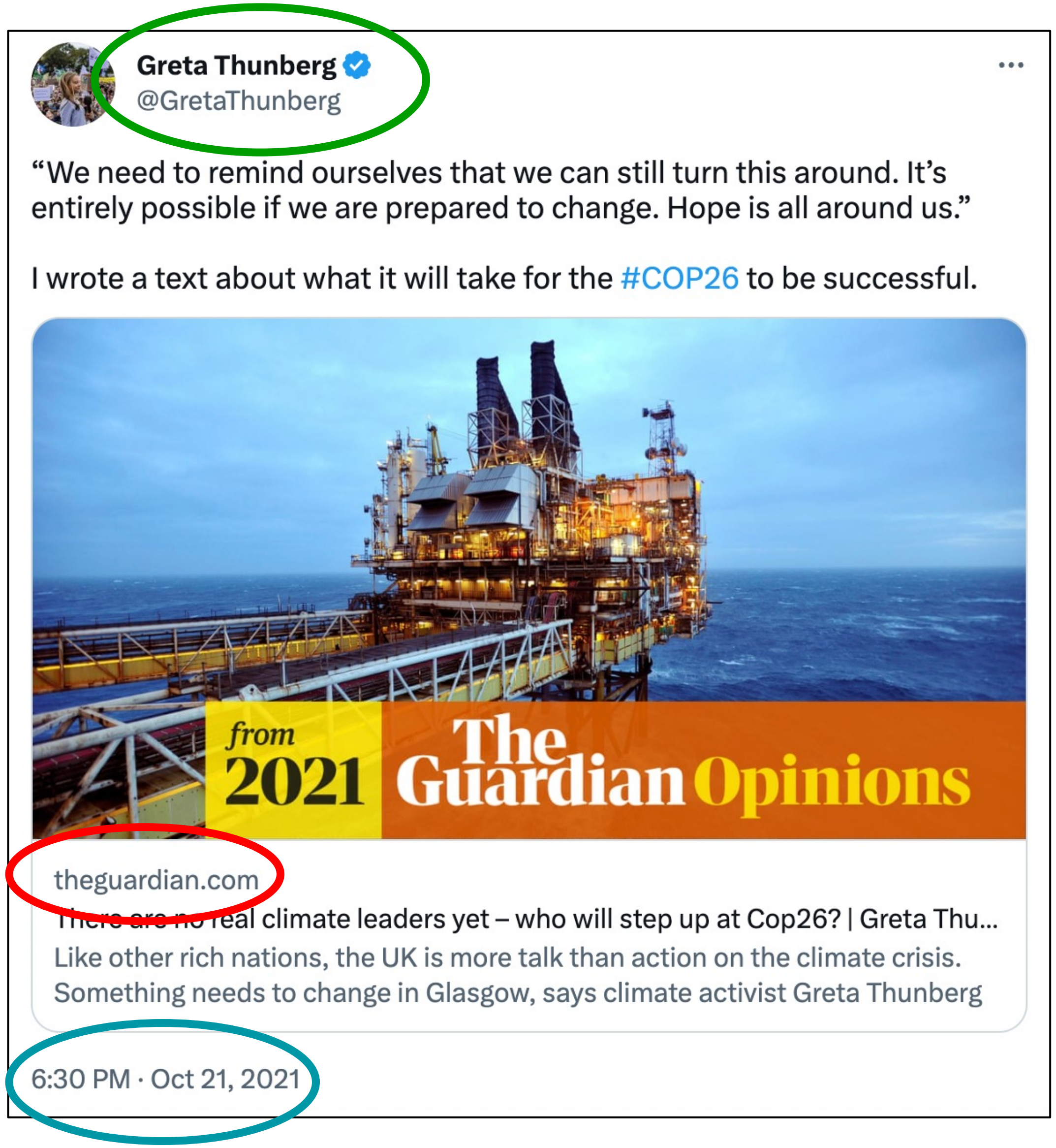
```
("climate change" OR #climatechange OR #climate_change OR "climate crisis" OR #climatecrisis OR
#climate_crisis OR "climate emergency" OR #climateemergency OR #climate_emergency OR "global warming"
OR #globalwarming OR #global_warming OR "climate action" OR #climateaction OR #climate_action)
has:links lang:en
```

Twitter API query to obtain climate-related tweets

Data processing

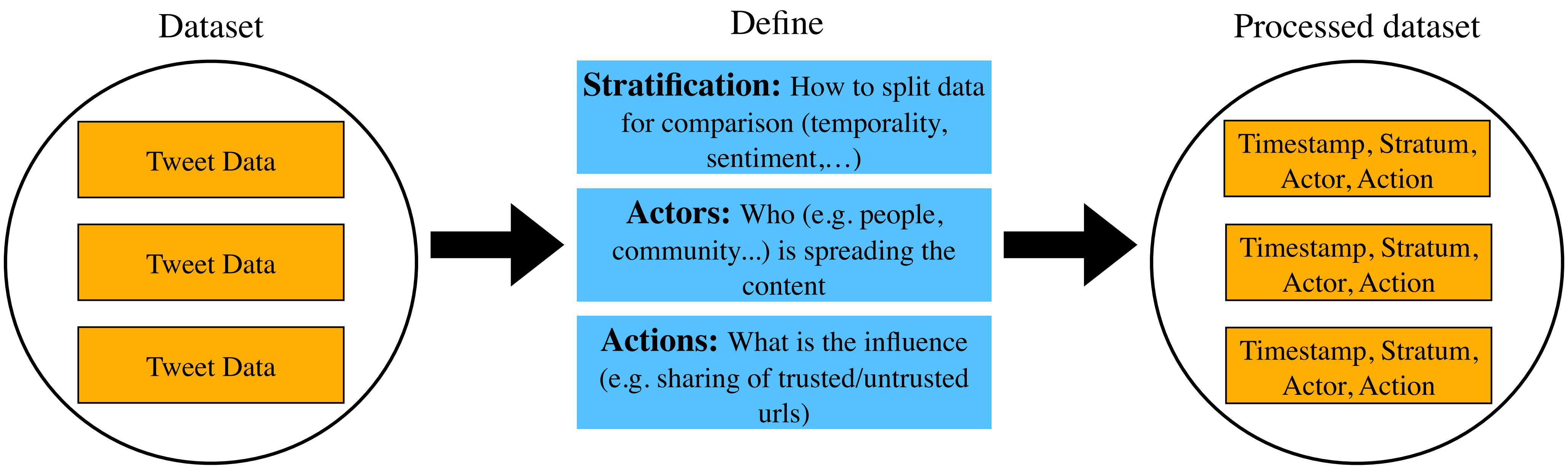


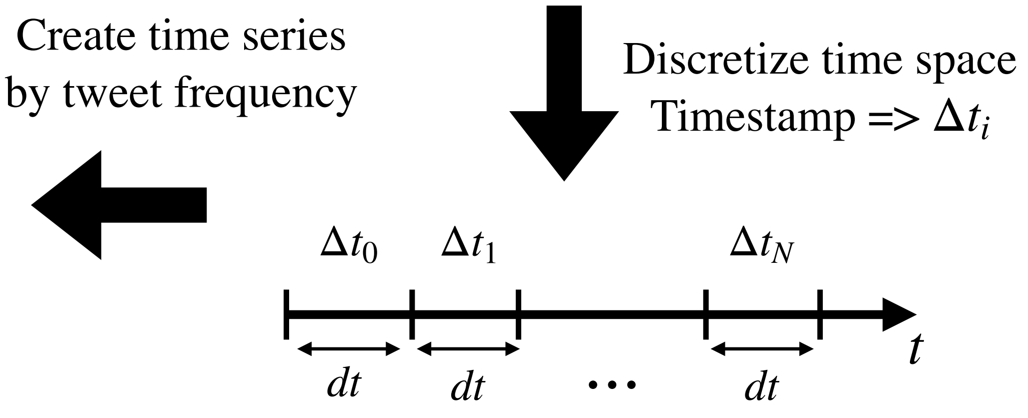
Data processing (example)



- **Timestamp:** 2021-10-21 at 18:30
- **Stratification:** Before COP26 (based on timestamp)
- **Actor:** Greta Thunberg
- **Action:** sharing of **T**rustworthy (**T**) news (NewsGuard score for *theguardian.com* is 100)

Time series creation





	Bill Gates	
	U	T
Δt_0	2	2
Δt_1	1	5
Δt_2	0	1
$\vdots$	$\vdots$	$\vdots$
Δt_N	0	1

	Greta Thunberg	
	U	T
Δt_0	0	4
Δt_1	0	2
Δt_2	0	1
$\vdots$	$\vdots$	$\vdots$
Δt_N	1	3

Influence graphs

	A : Bill Gates	
	U	T
Δt_0	2	2
Δt_1	1	5
Δt_2	0	1
$\vdots$	$\vdots$	$\vdots$
Δt_N	0	1
	X_U^A	X_T^A

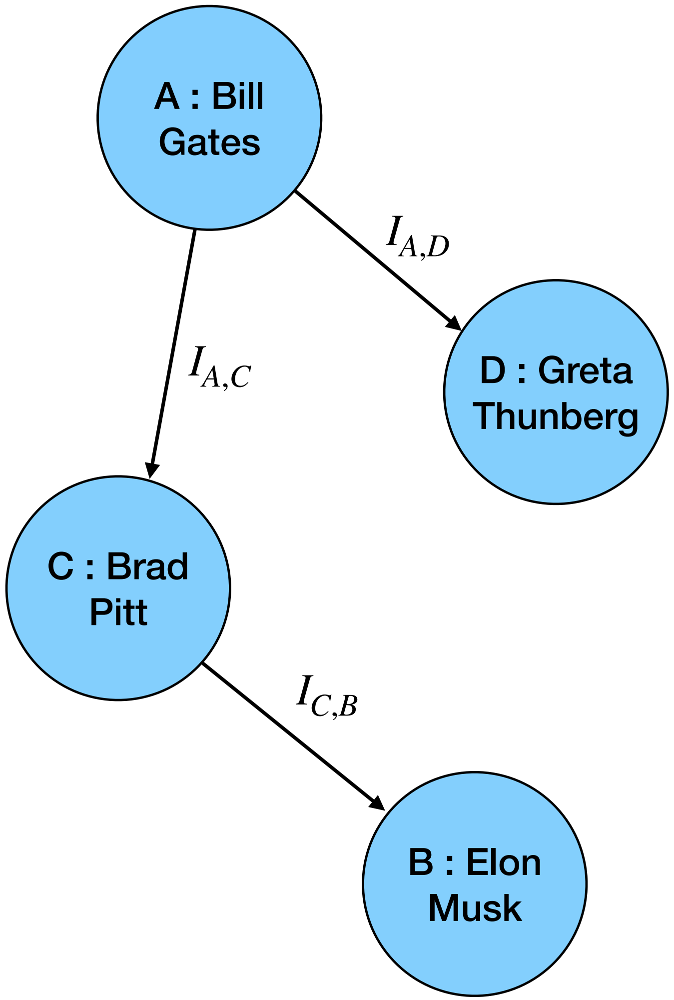
$\cdots$

	D : Greta Thunberg	
	U	T
Δt_0	0	4
Δt_1	0	2
Δt_2	0	1
$\vdots$	$\vdots$	$\vdots$
Δt_N	1	3
	X_U^D	X_T^D

Compute influence matrix between actors

$$I_{A,B} = \begin{bmatrix} d(X_U^A, X_U^B) & d(X_U^A, X_T^B) \\ d(X_T^A, X_U^B) & d(X_T^A, X_T^B) \end{bmatrix}$$

Create influence graph



Coupling Inference Methods

Key component: Coupling Inference Methods

Transfer Entropy (TE)

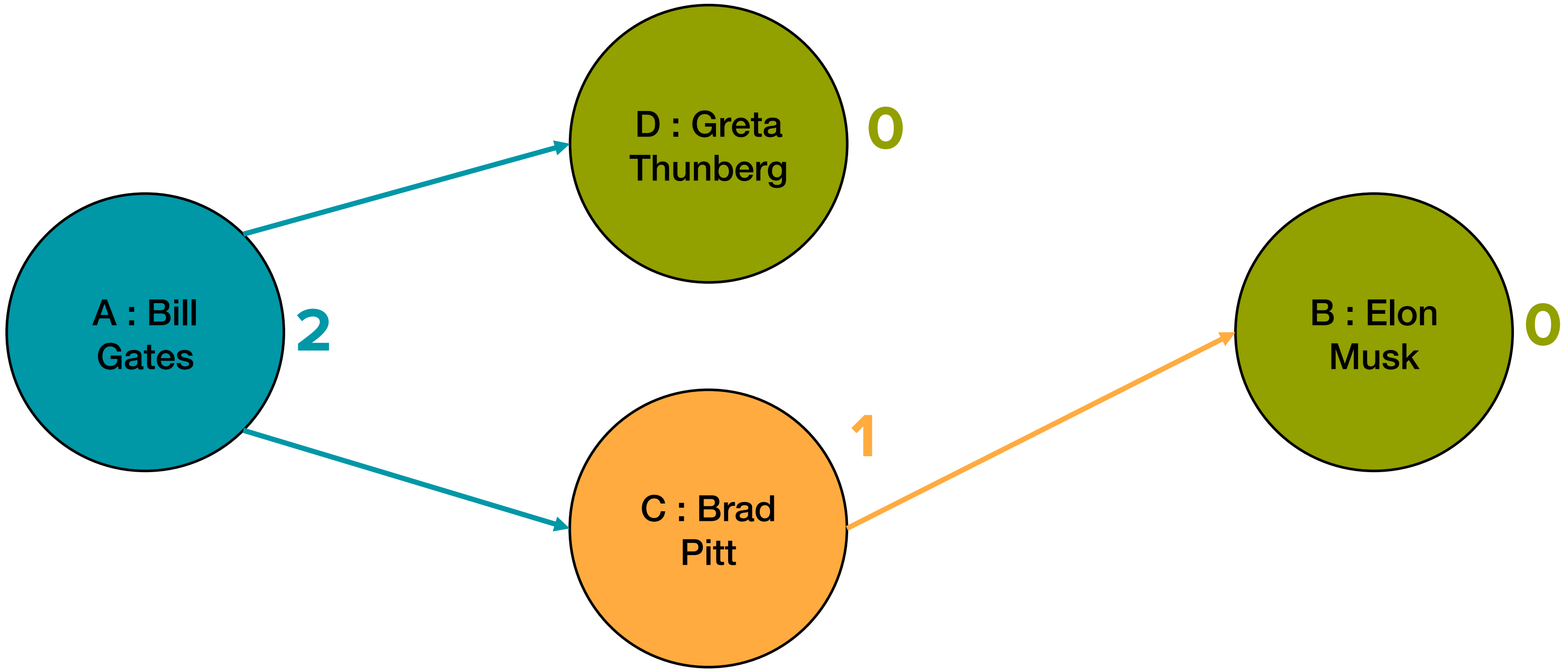
- Need to **binarize** the time series
- **Naive estimation** of state probabilities
- Need an **arbitrary threshold**
- Well-known and **less** subject to **noise**

Joint Distance Distribution (JDD)

- Use **original** (standardized) **values**
- Based on **distances**
- Formulated as **hypothesis testing**
- **More** subject to **noise**

Influence measures

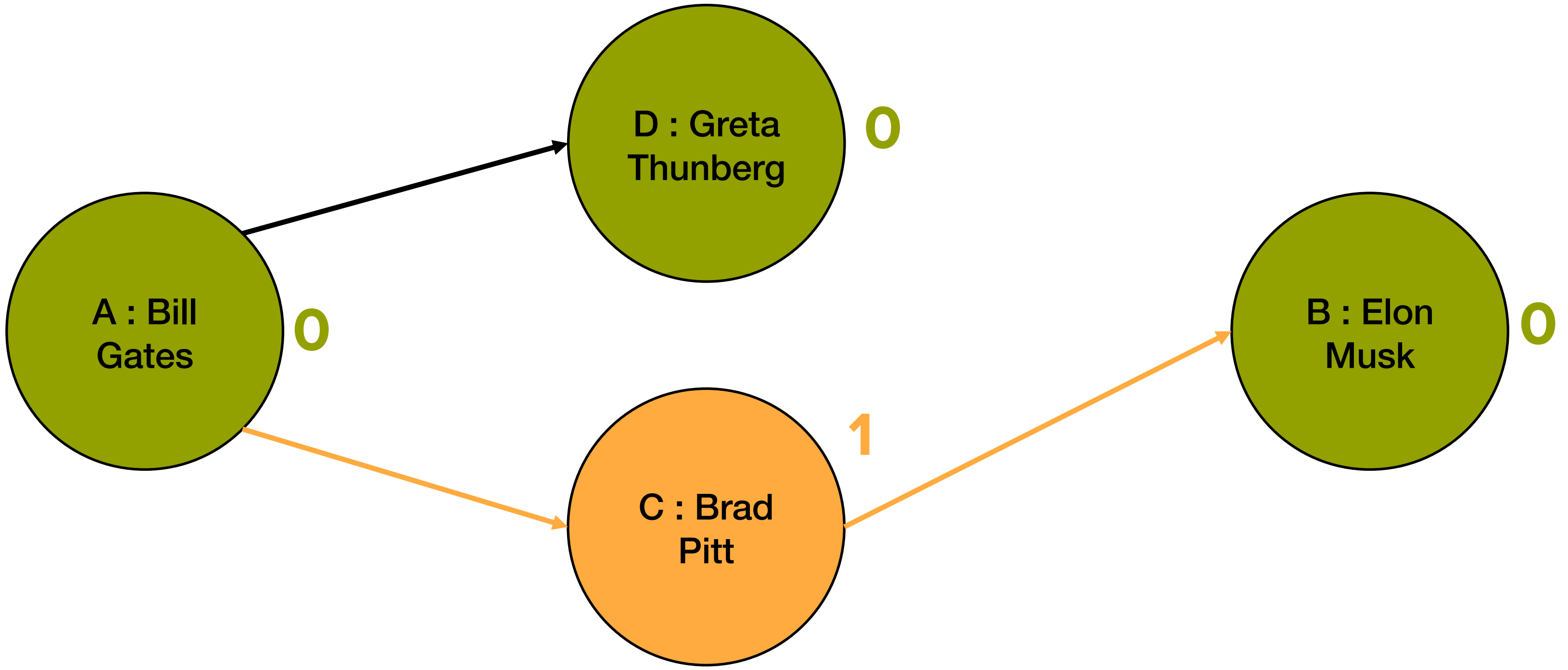
- ***Outdegree:*** for each node (actor) in the influence graph, outdegree directly **indicates how many other actors this actor influenced.**



Influence measures

- ***Outdegree:*** for each node (actor) in the influence graph, outdegree directly **indicates how many other actors this actor influenced.**

- ***Betweenness:*** indication of **how well one actor serves as a bridge to influence many other users.**



Influence measures

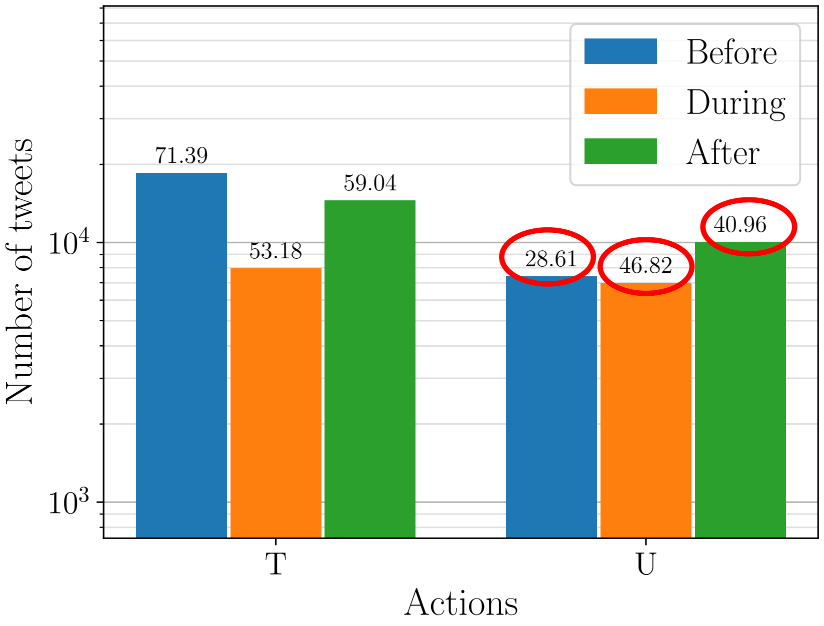
These measures can be fine-grained and derived for each edge type in the graph:

- T-T: **echo chamber** of trustworthy news sharing.
- T-U: **credibility cross-over** of trustworthy to untrustworthy news sharing
- U-T: **credibility cross-over** of untrustworthy to trustworthy news sharing
- U-U: **echo chamber** of untrustworthy news sharing.

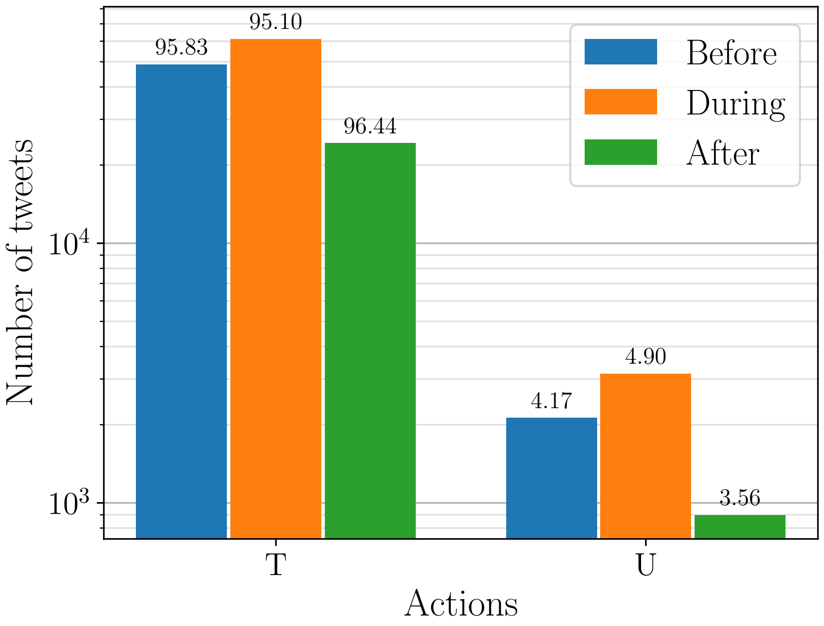
Results

Action distributions

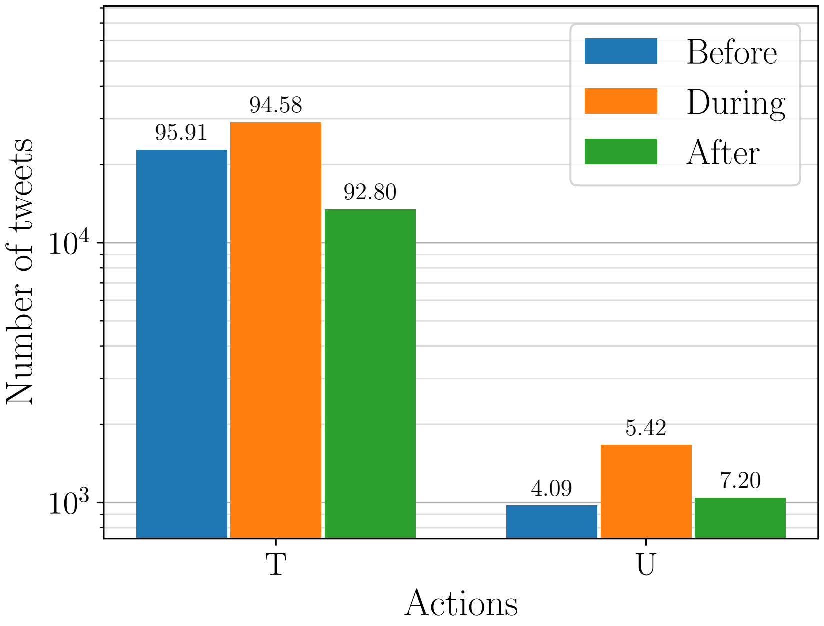
- Proportion of **untrustworthy news** sharing **larger** during **Skripal**
- May be a **hint** of **underlying disinformation campaign**



Skripal dataset



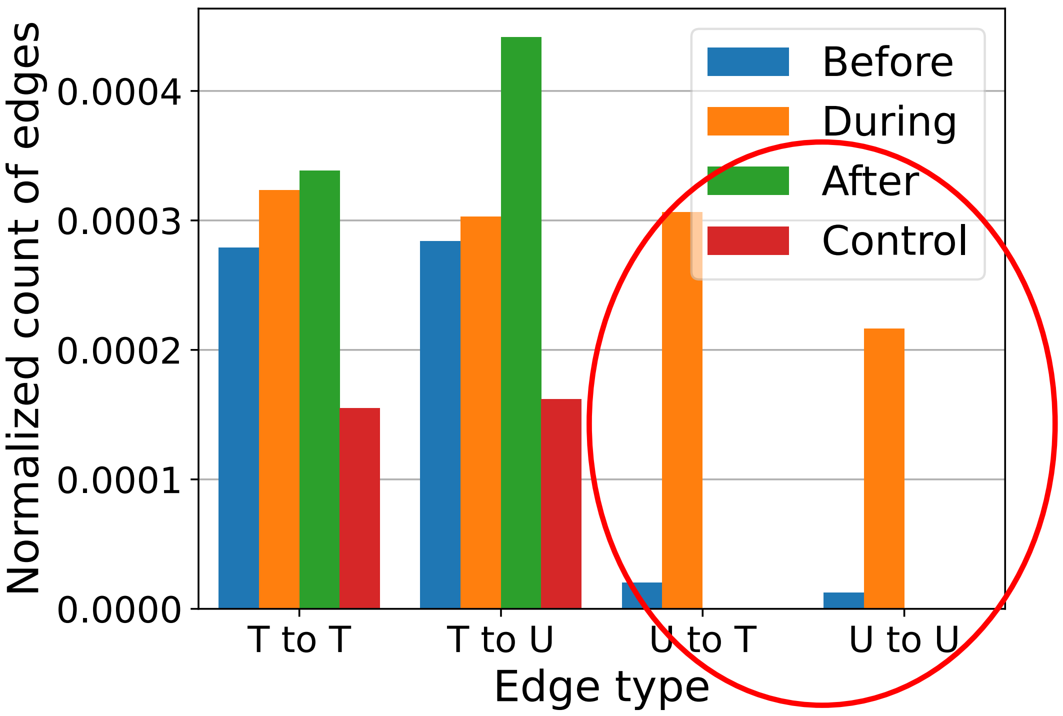
COP26 dataset



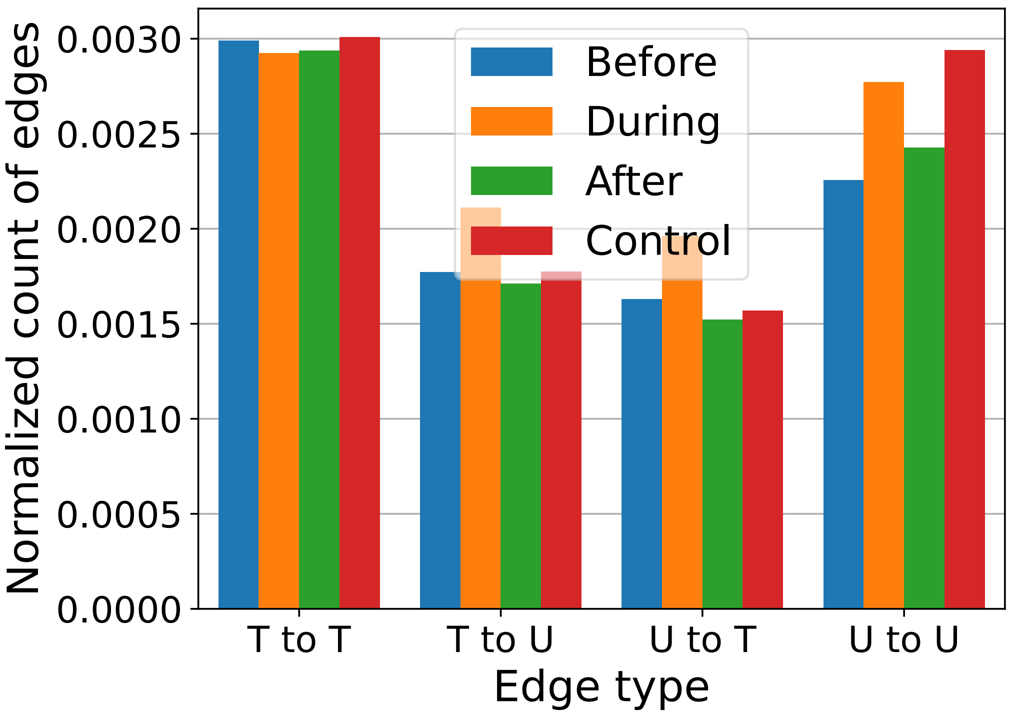
COP27 dataset

Comparison JDD-TE influence graphs

- JDD detects **large spikes** of influence of the **untrustworthy** types during COP26.
- **No difference** with the **control** for **TE**.



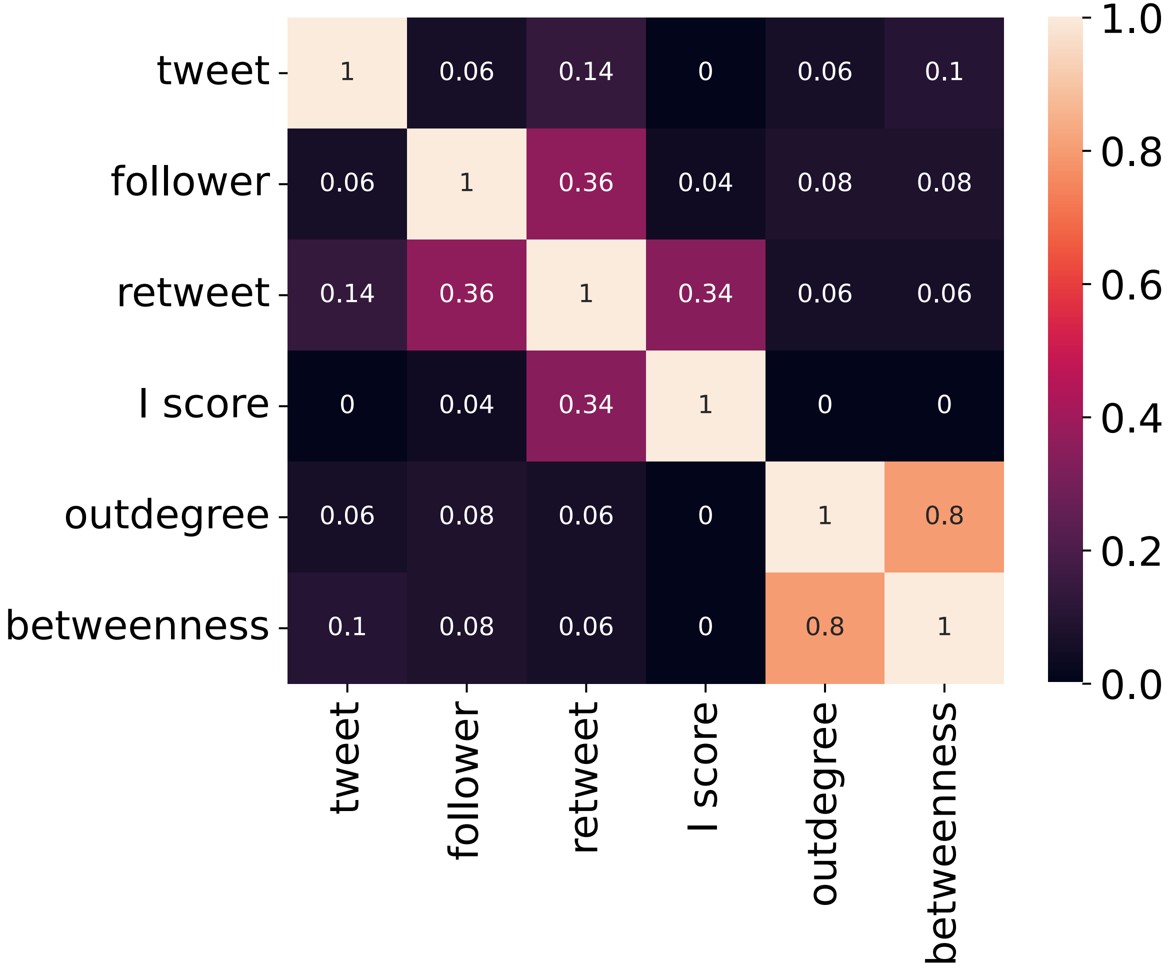
JDD on the COP26 dataset



TE on the COP26 dataset

One influence measure to rule them all?

- **Weak** overall **correlations**
- Only **outdegree** and **betweenness** exhibit relatively **high correlation**.
- **Each measure** is in fact capturing **different types of information**, and influence definitions.



Correlation matrix between influence measures for JDD on the COP26 dataset

What do influential people share? -- Skripal

	Skripal dataset	Top 10 JDD	Top 10 TE
Before	4.8	-	32.9
During	3.9	33.6	15.1
After	6.3	20.7	10.9

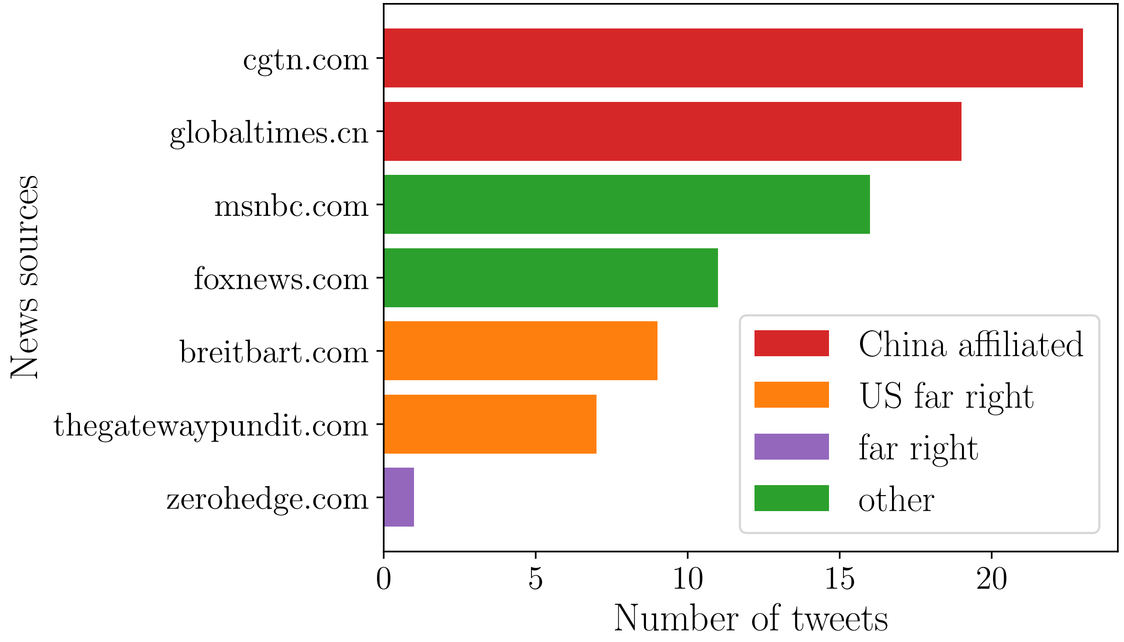
Mean number of mentions of “rt.com”

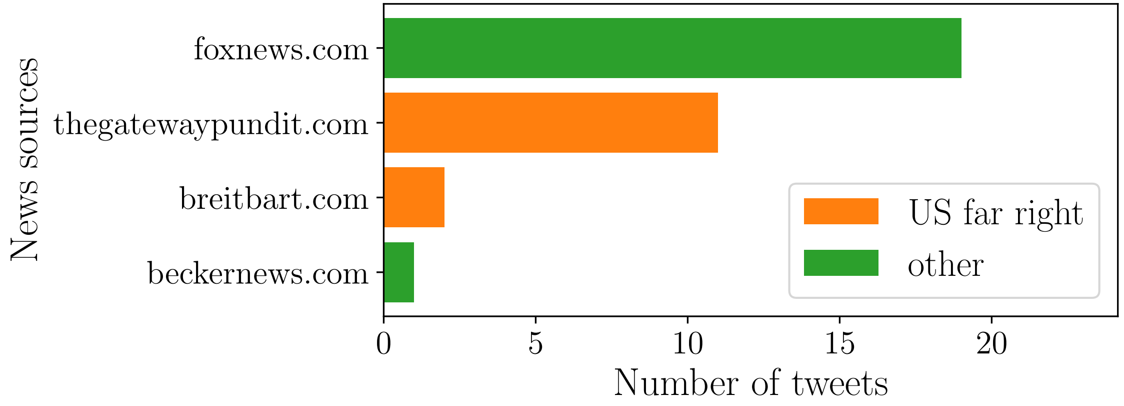
	Skripal dataset	Top 10 JDD	Top 10 TE
Before	1.1	-	4.0
During	1.9	28.7	19.5
After	1.4	11.5	0

Mean number of mentions of “sputniknews.com”

- **RT** and **Sputnik** vectors of Russian **disinformation** campaign during Skripal
- **Significantly more sharing** of these websites by the most **influential** users **we detected**
- Even **more sharings** by users detected by **JDD compared to TE**

What do influential people share? -- COP26





- **Most influential users** share articles from **China-controlled** or **far-right-affiliated** media
- **JDD** detects **more** content directly related to **disinformation than TE**
- **China-affiliated** news outlets share **anti-US** and **anti-Western** climate **narratives**.
- **Far-right US** outlets share **counter narrative** and narrative **against Biden**'s climate awareness

Conclusion

Conclusion

- **Quantifying influence** between users on social media is **difficult**, but **essential** to detect and **counter disinformation**.
- **Derived** and **explored** a **novel method** to **measure influence** and **detect disinformation** on social media.
- **Compared** two **coupling inference methods** to derive the influence graphs.
- **Influence measures** have relatively **low correlation** between themselves, showcasing that they **do not capture the same types of influence**.
- **JDD** results in **sparser influence graphs**, but finds individuals who seem **more active** in spreading **disinformation** than when using **TE**.

Future work

- **Explore** other **action** definitions
- Incorporate **content analysis**
- Further **validate** on other **known cases** of **disinformation campaigns**

Thank you!