Coordination Detection

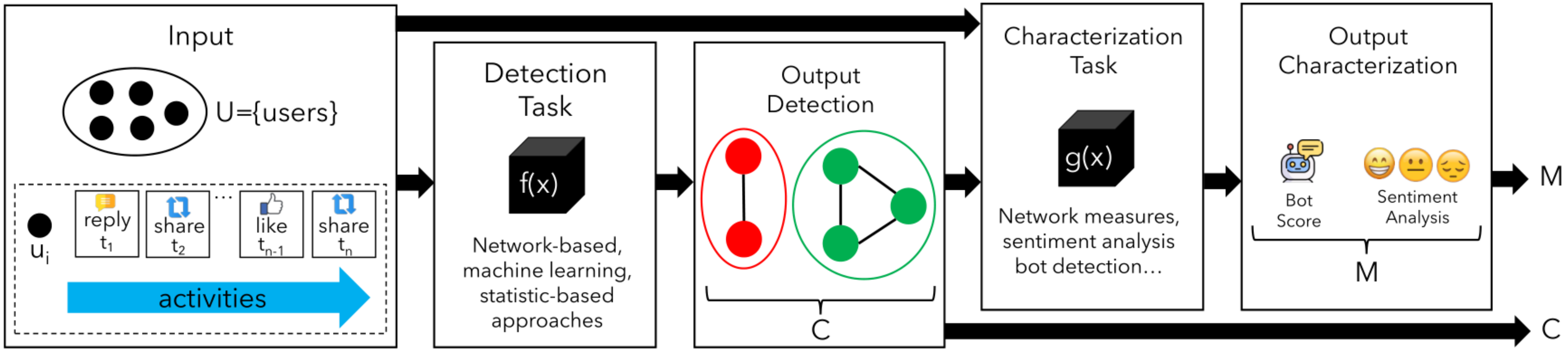
Lorenzo Cima

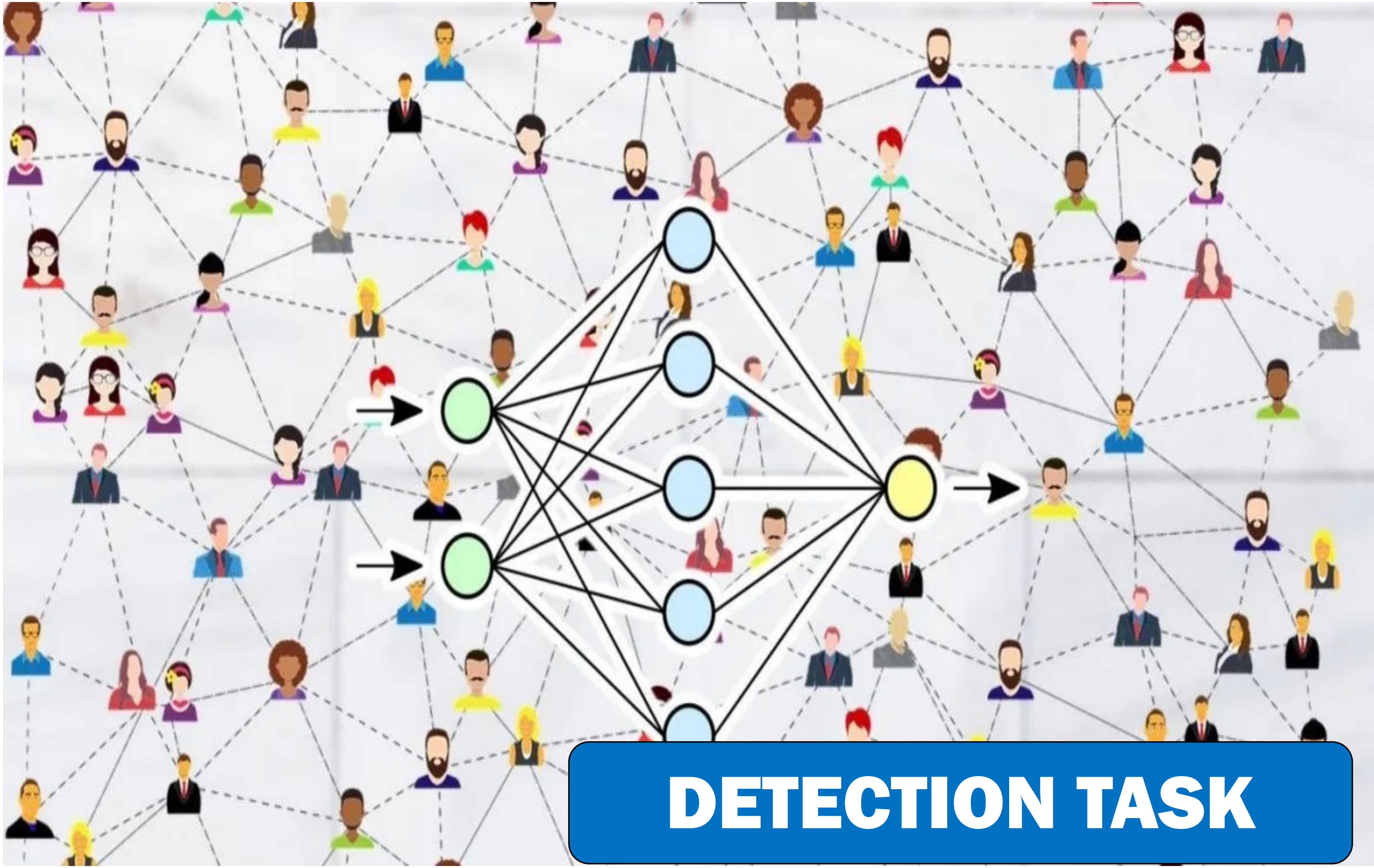
Università di Pisa

Istituto di Informatica e Telematica

lorenzo.cima@phd.unipi.it; lorenzo.cima@iit.cnr.it

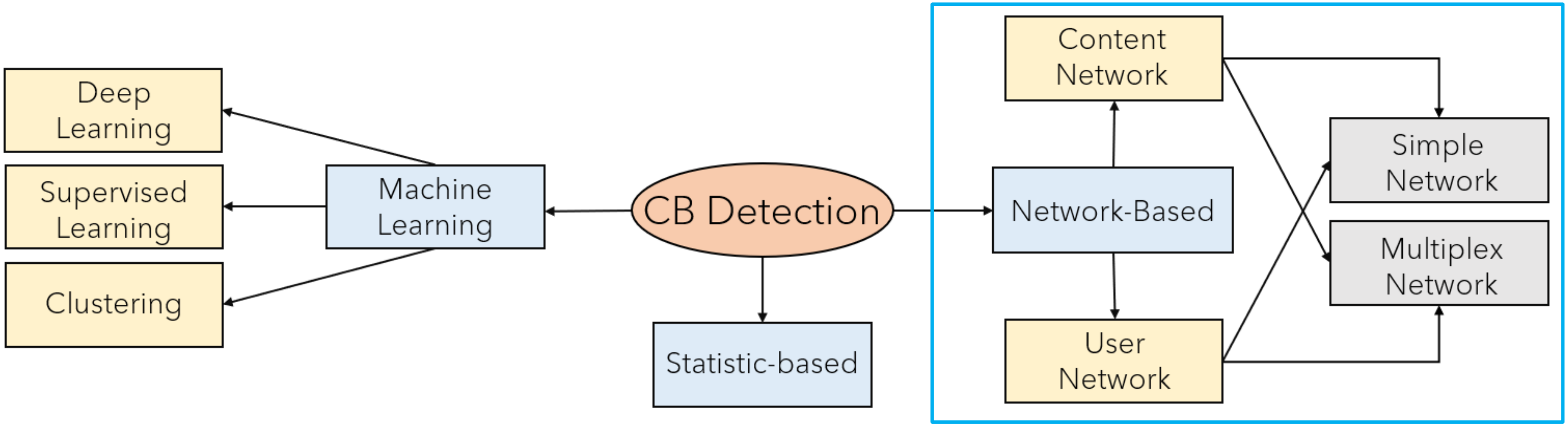
Problem Definition





DETECTION TASK

Detection Approaches



Some Definitions

- **Direct interactions:** voluntary interactions done with a direct action (message, follower…)

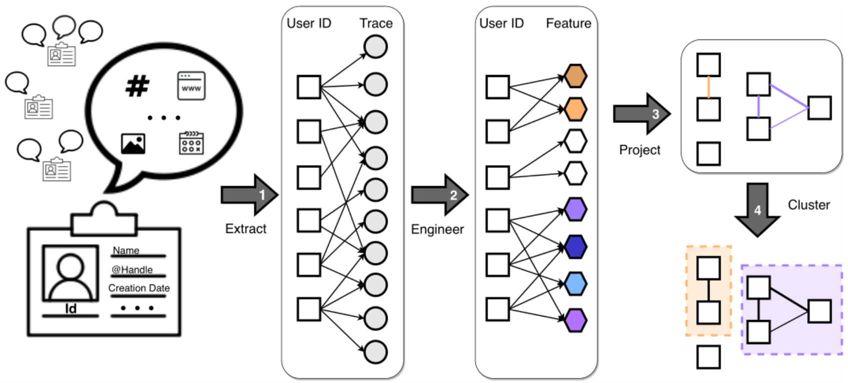
- **Indirect interactions:** interactions between users based on similar behaviours (co-actions)

- **Co-actions:** two actions of the same type performed by two different users, such as re-sharing a post, or sharing a post with the same URL or mention.

Detection Frameworks

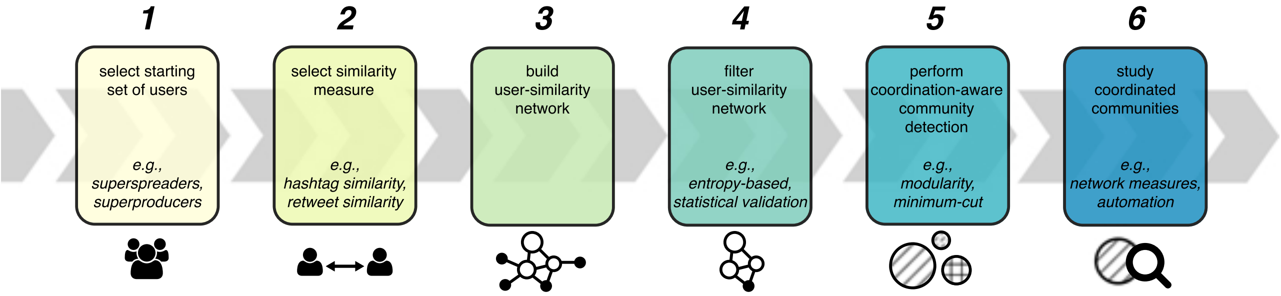
- **Pacheco** method
  - Bipartite graph

Pacheco, Diogo, et al. "Uncovering coordinated networks on social media: methods and case studies." Proceedings of the international AAAI conference on web and social media. Vol. 15. 2021.



- **Nizzoli** method
  - Complete weighted network

Nizzoli, Leonardo, et al. "Coordinated behavior on social media in 2019 UK general election." Proceedings of the International AAAI Conference on Web and Social Media. Vol. 15. 2021.



1 select starting set of users — *e.g., superspreaders, superproducers*

2 select similarity measure — *e.g., hashtag similarity, retweet similarity*

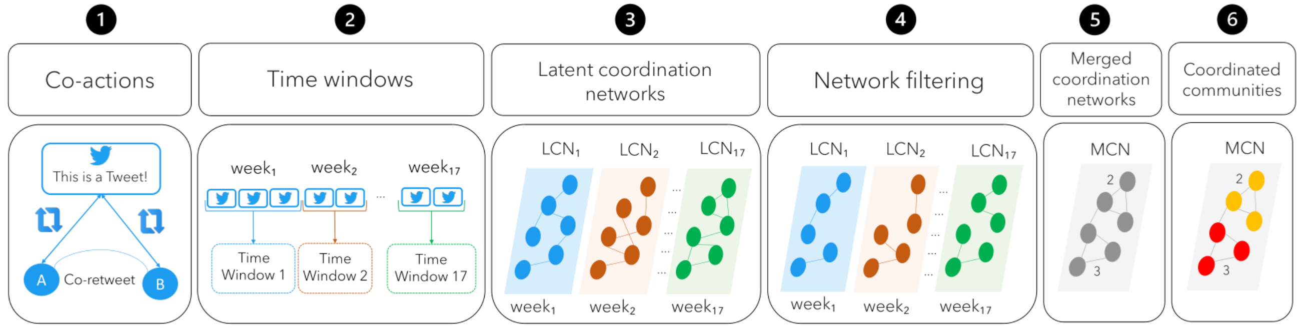
3 build user-similarity network

4 filter user-similarity network — *e.g., entropy-based, statistical validation*

5 perform coordination-aware community detection — *e.g., modularity, minimum-cut*

6 study coordinated communities — *e.g., network measures, automation*

- **Weber** framework



1 Co-actions

2 Time windows

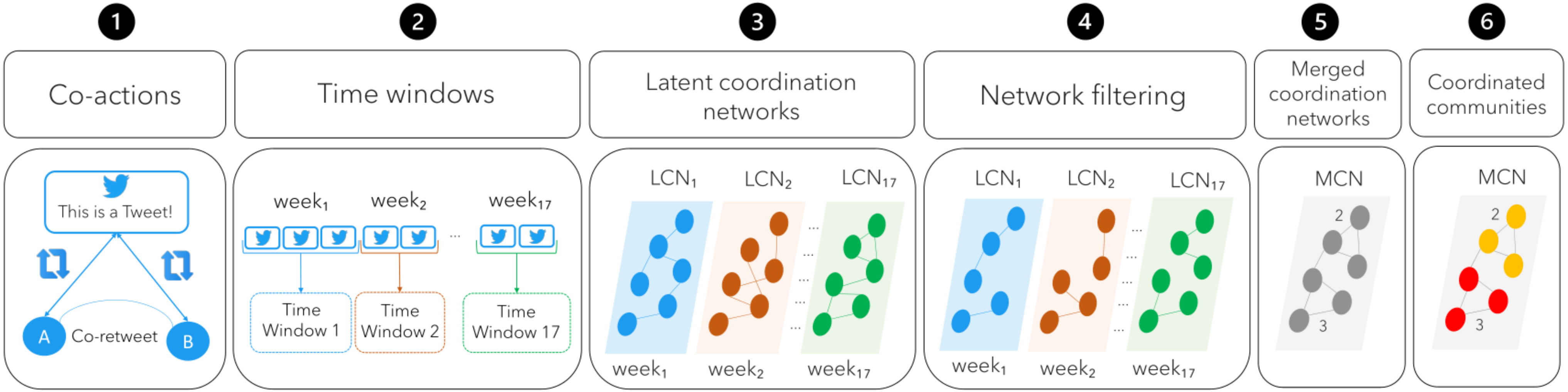
3 Latent coordination networks

4 Network filtering

5 Merged coordination networks

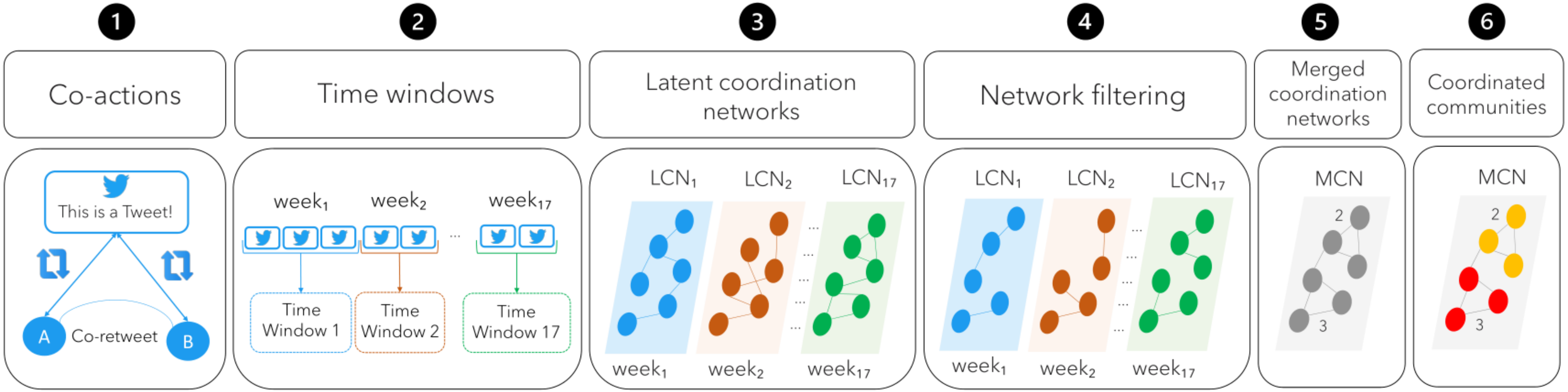
6 Coordinated communities

Weber Coordination Detection Method



Weber, Derek, and Frank Neumann. "Amplifying influence through coordinated behaviour in social networks." Social Network Analysis and Mining 11.1 (2021).

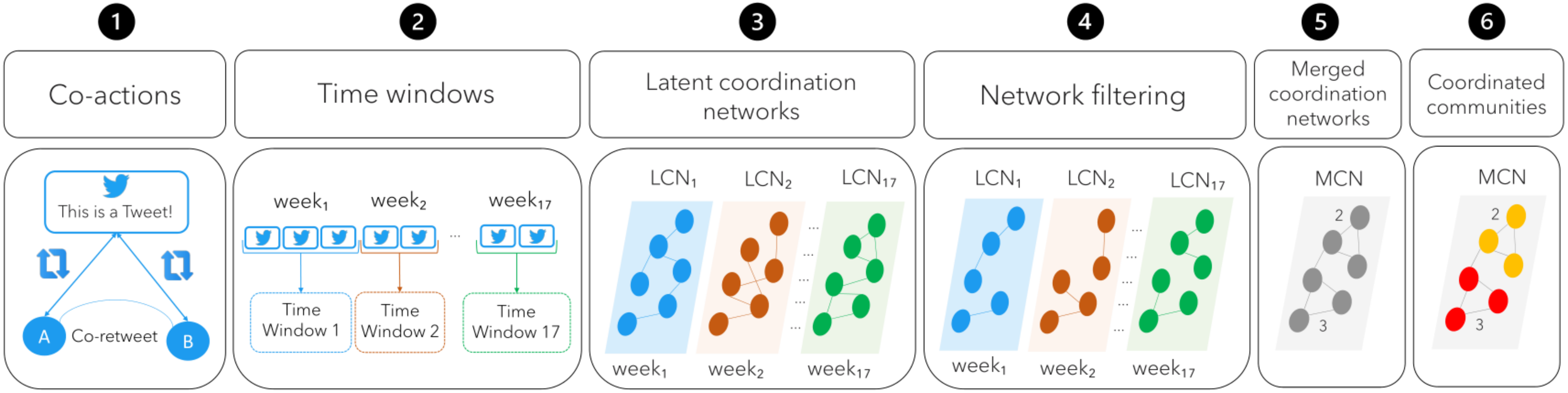
Weber Coordination Detection Method



① Extract co-actions

Convert social media posts to a set of one or more interesting actions and define the corresponding set of co-actions

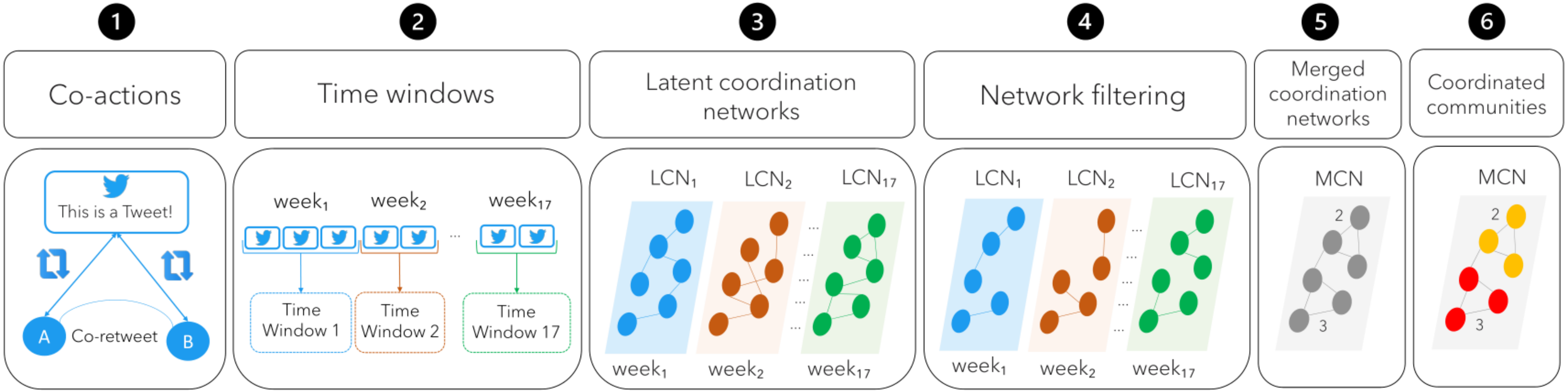
Weber Coordination Detection Method



② Define time windows

Why? Temporal Analysis, Orchestration,
Decrease Network Size (computational limits)

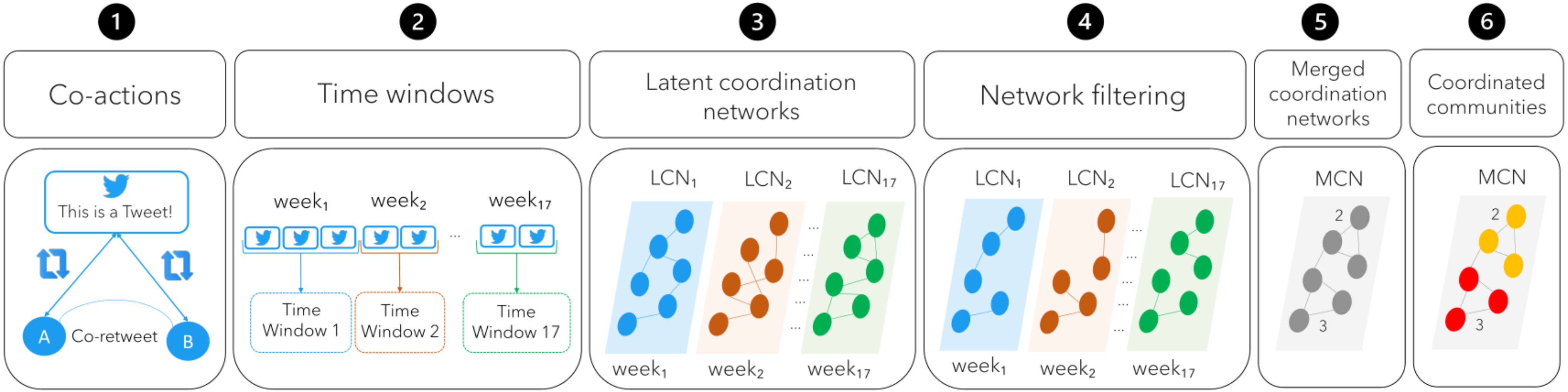
Weber Coordination Detection Method



2 Define time windows

The time windows may overlap (**sliding window)** or be **adjacent**
We validate a co-action only if composed of two actions
occurring within the time window

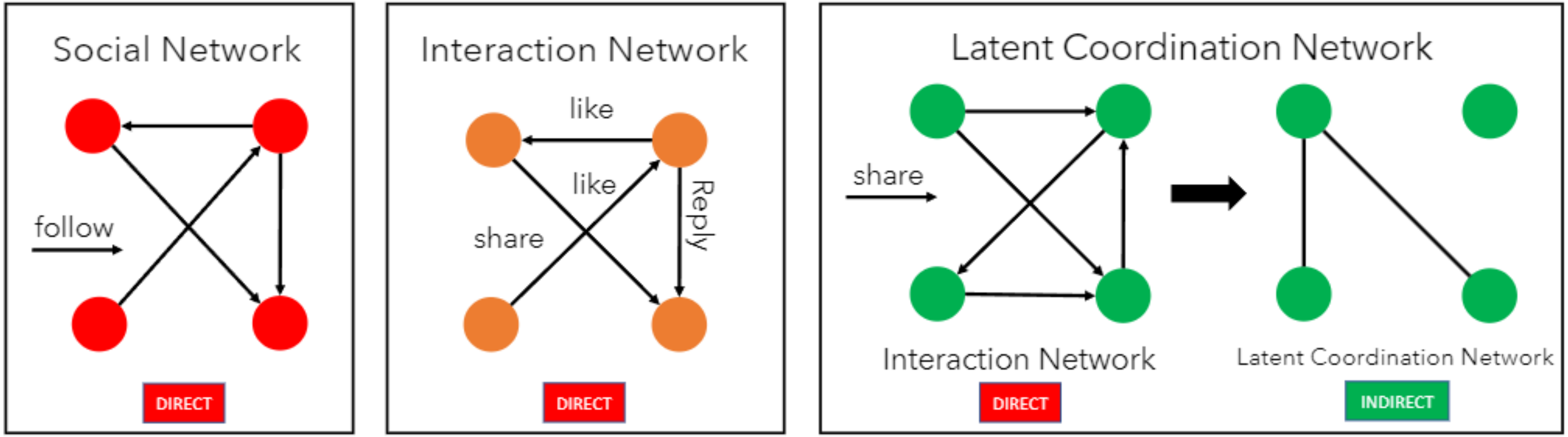
Weber Coordination Detection Method



3 Build Latent Coordination Networks (LCNs)

On each time window an LCN is extracted

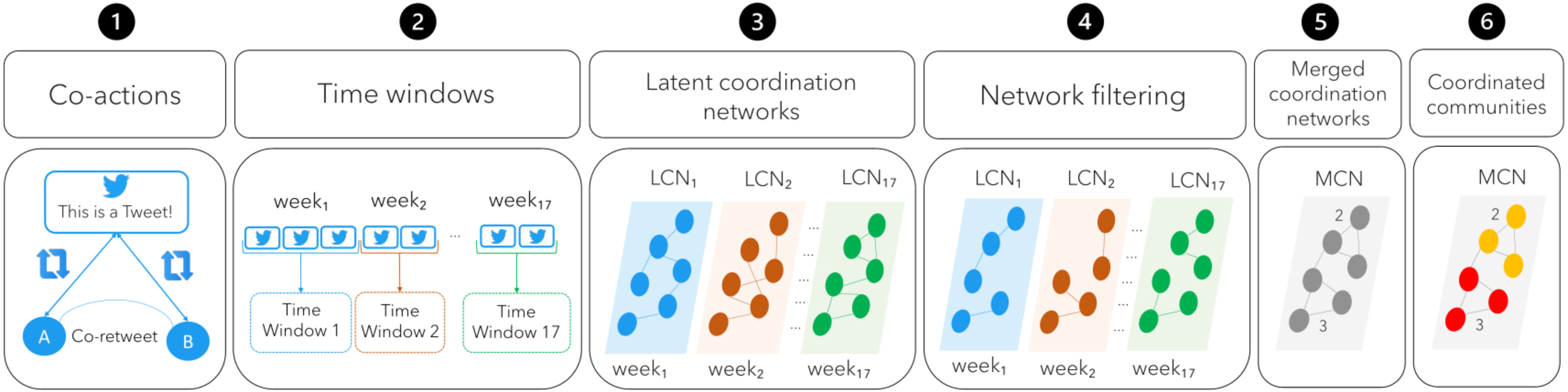
Latent Coordination Network



The **Latent Coordination Network (LCN)** is a weighted network built in this way: when two users do a co-action, they become nodes and an edge connects them.

The higher the number of co-actions, the higher the edge weight

Weber Coordination Detection Method

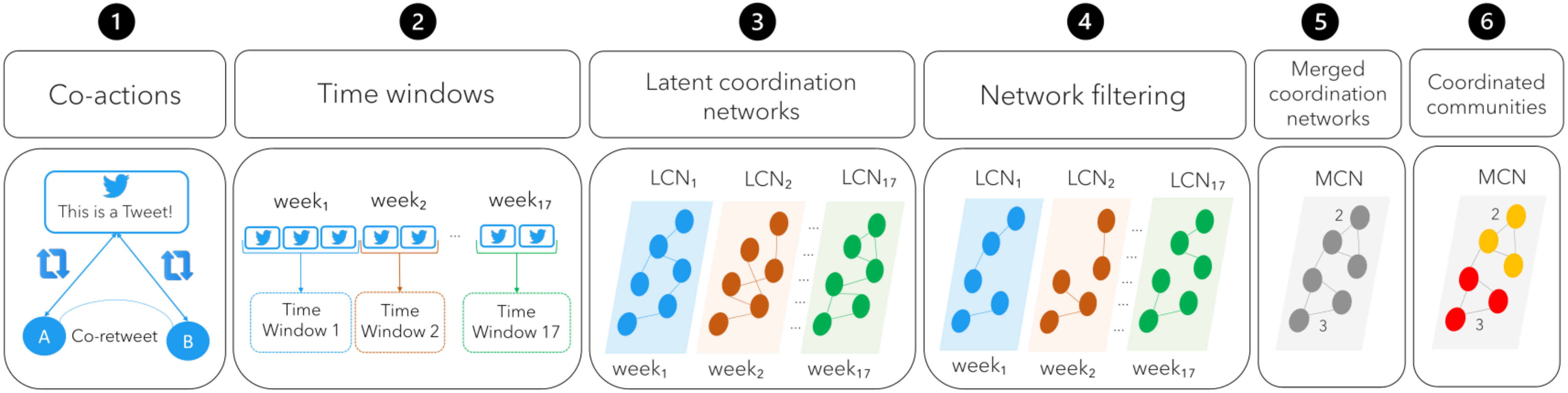


4 Network filtering

The network is filtered using a variant of the Focal Structural Analysis (FSA) algorithm provided by the authors

In the end, on each time window we obtain Highly Coordinated Communities (HCCs)

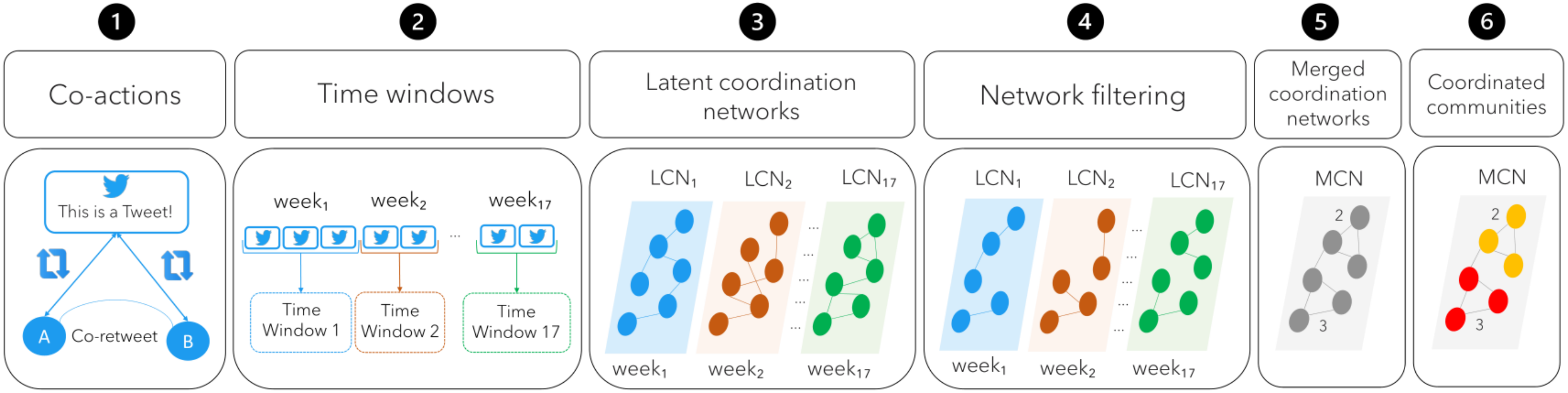
Weber Coordination Detection Method



⑤ Compute the Merged Coordination Network (MCN)

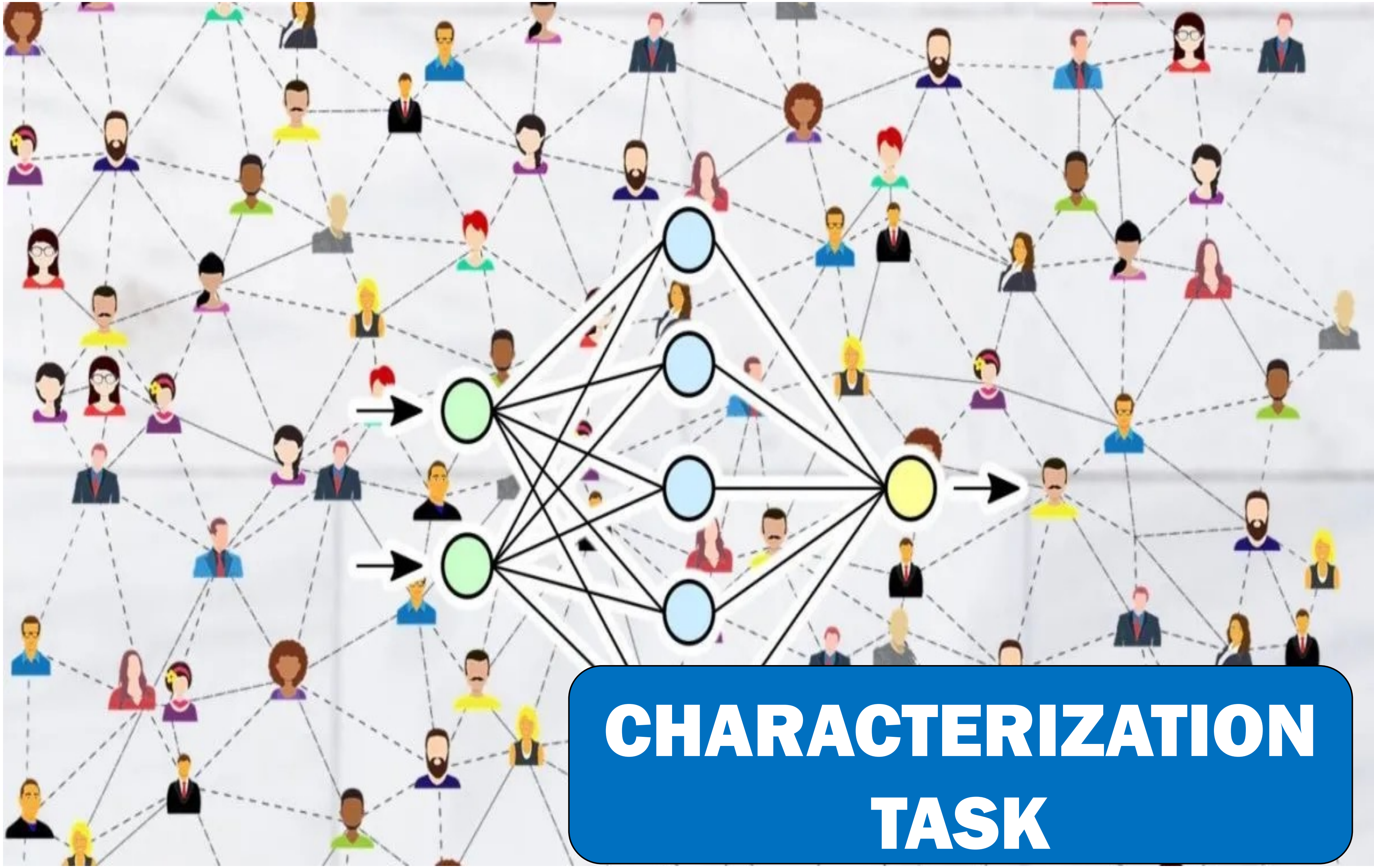
All the HCCs are combined to obtain a unique merged network, which accounts for all the time windows

Weber Coordination Detection Method



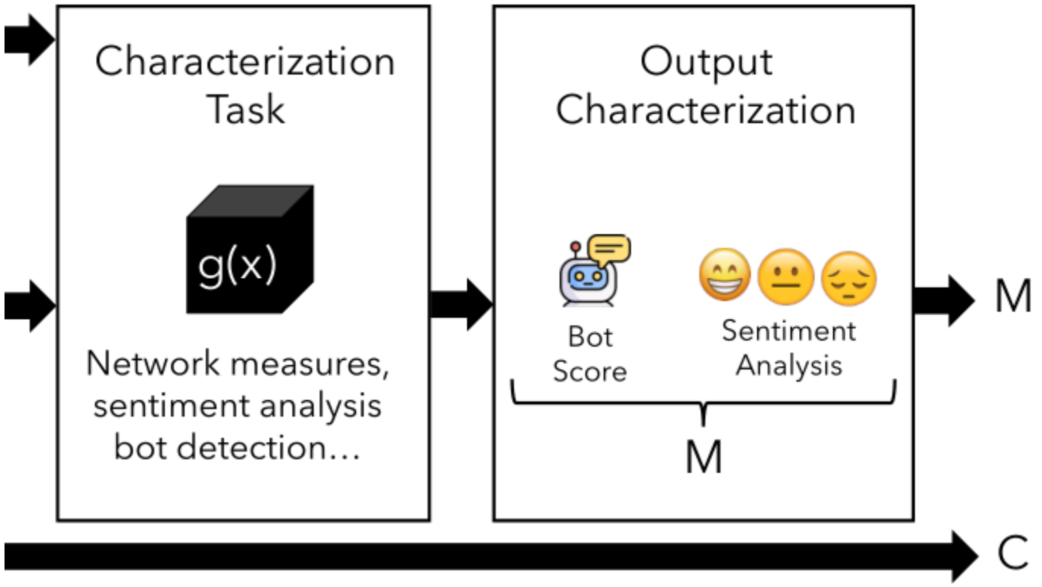
⑥ Find coordinated communities

The authors use the Louvain algorithm to find coordinated communities



CHARACTERIZATION TASK

Charachterization Task

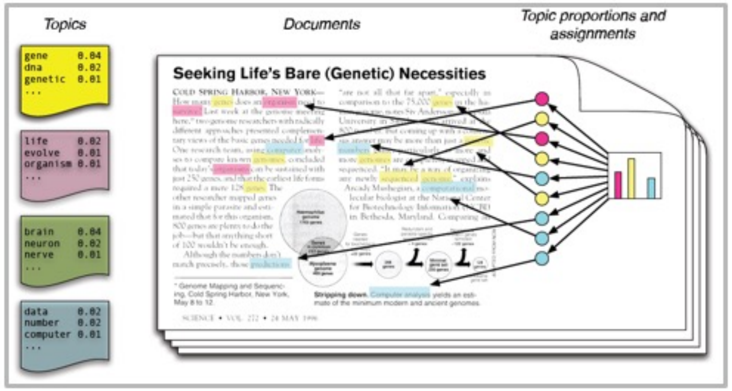


Lack of Ground Truth

- Harmfulness
- Authenticity
- Orchestration

The content is not the focus anymore, but it is a proxy for insights into the type of coordination and the intent of an orchestrated campaign

Harmfulness Proxies

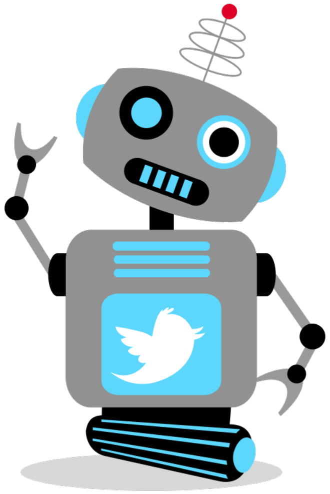


Topic-modelling

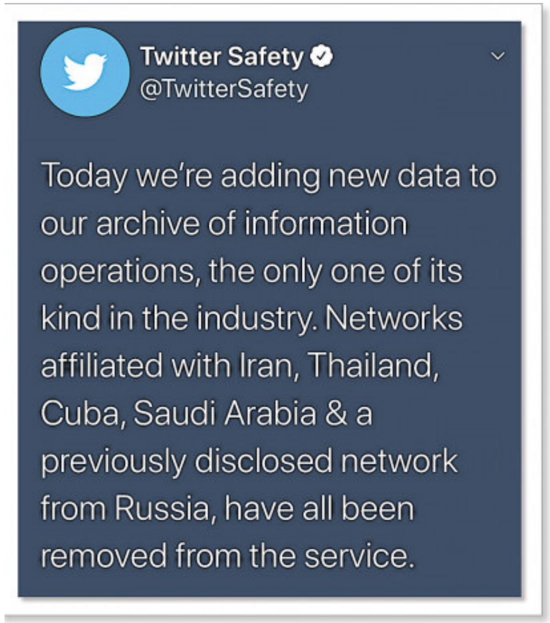
News URLs Shared

Propaganda

Inauthenticity Proxies



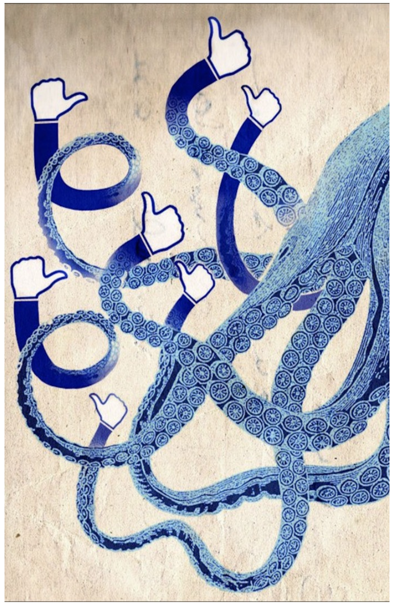
Bot Detection



Twitter Safety
@TwitterSafety

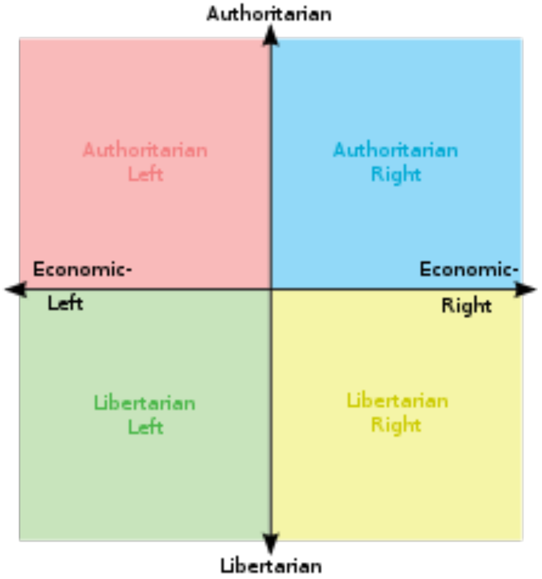
Today we're adding new data to our archive of information operations, the only one of its kind in the industry. Networks affiliated with Iran, Thailand, Cuba, Saudi Arabia & a previously disclosed network from Russia, have all been removed from the service.

Data Collection

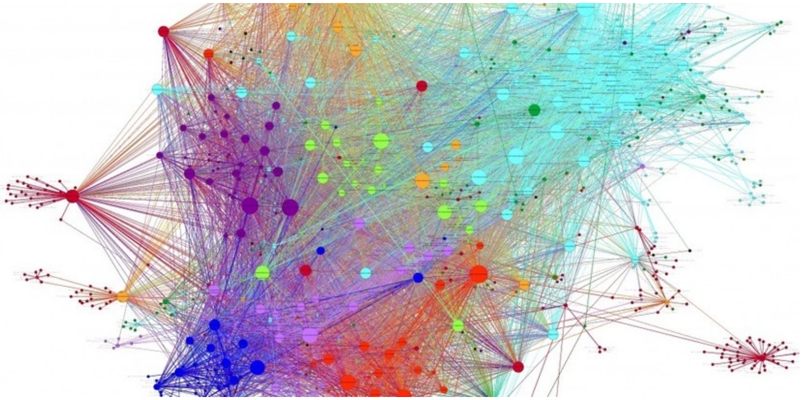


Troll Detection

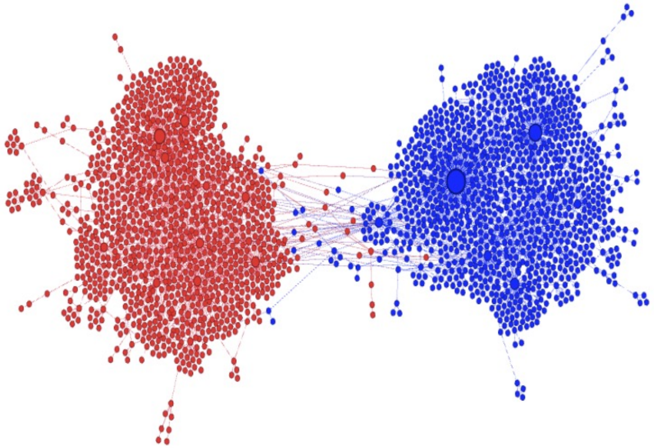
Other Characterizations



Political Leaning

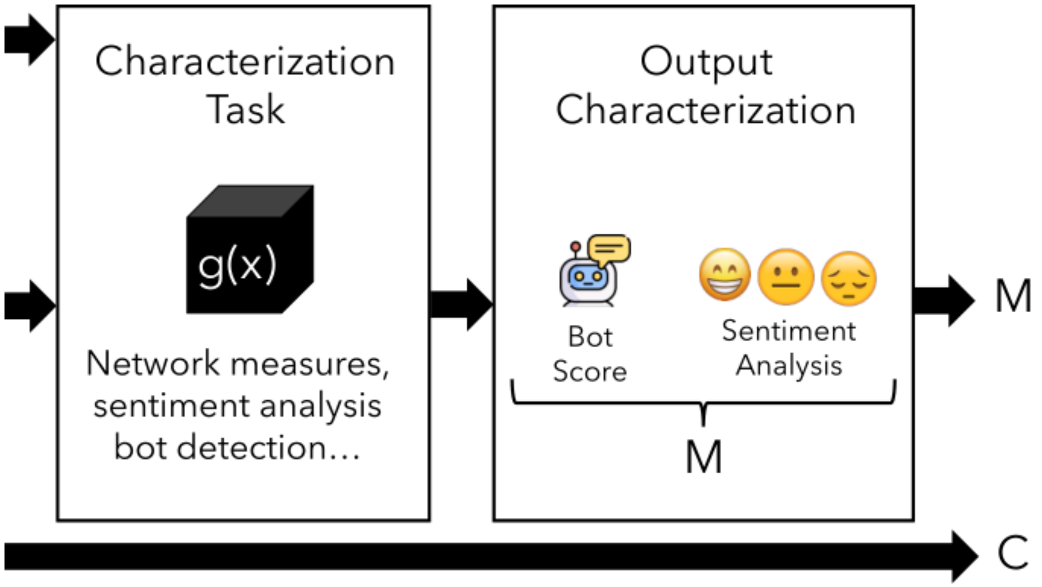


Network Measures



Polarization

Ground-Truth Characterization

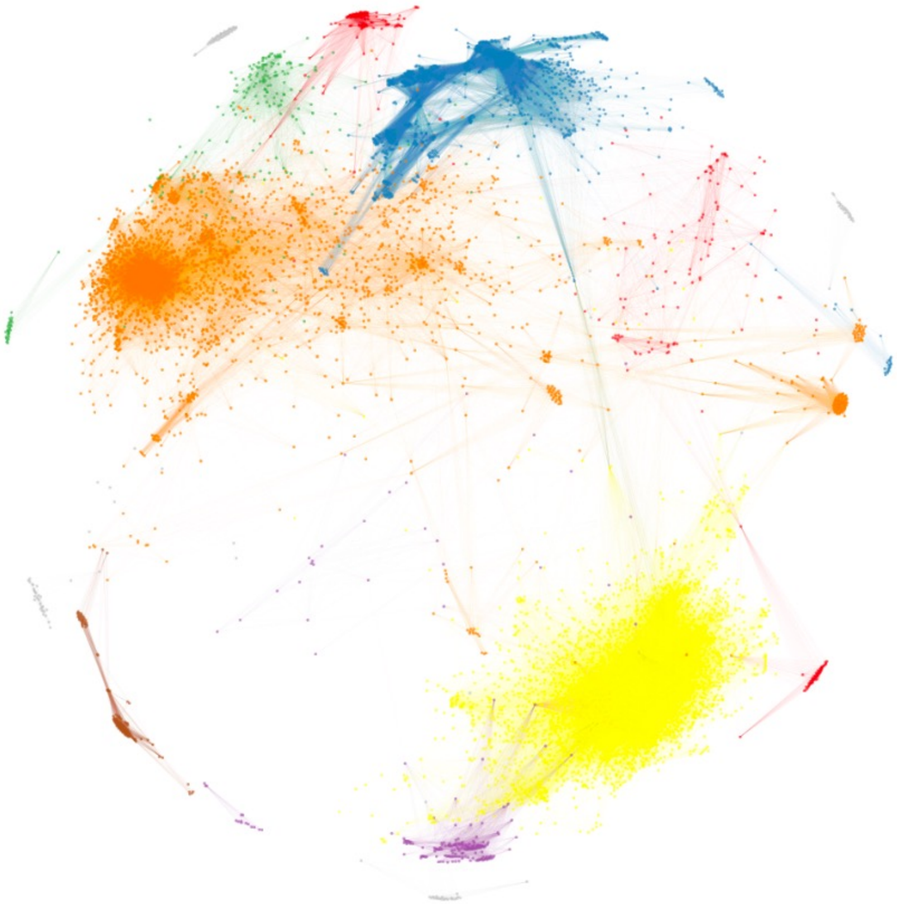


Presence of a Ground Truth

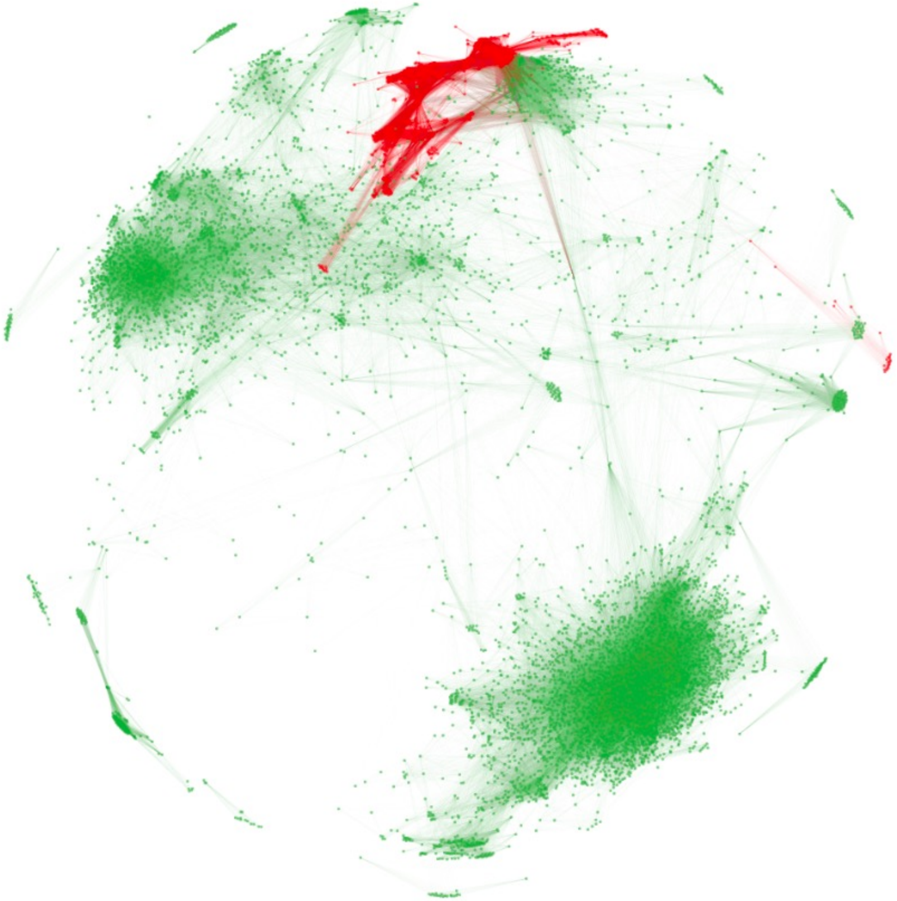
- Clustering quality for labelled data

The results should be compared with the ground truth, to evaluate clustering quality

Ground-Truth Characterization



Coordinated Communities
(*CLUSTERS*)



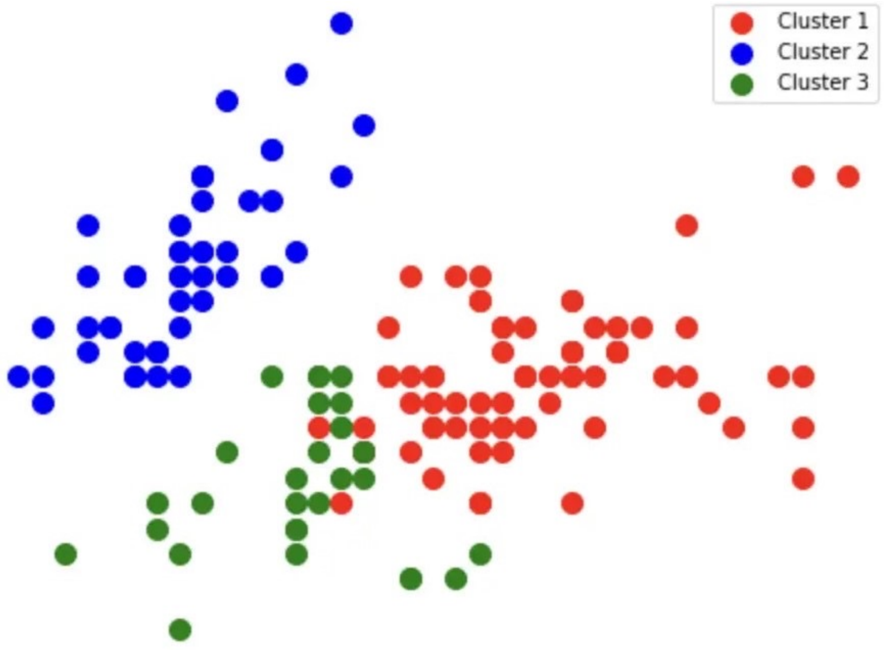
Ground-Truth
(*TYPES*)

Cima, Lorenzo, et al. "Coordinated Behaviors in Information Operations on Twitter." IEEE Access. 2024.

Clustering quality

How can we define a good clustering result for labelled data?

Perfect clustering: one cluster for each type, composed only of elements of that type



Example: we have 3 clusters and the division (blue, green, red) perfectly match the type labelling (type 1, type 2, type 3)

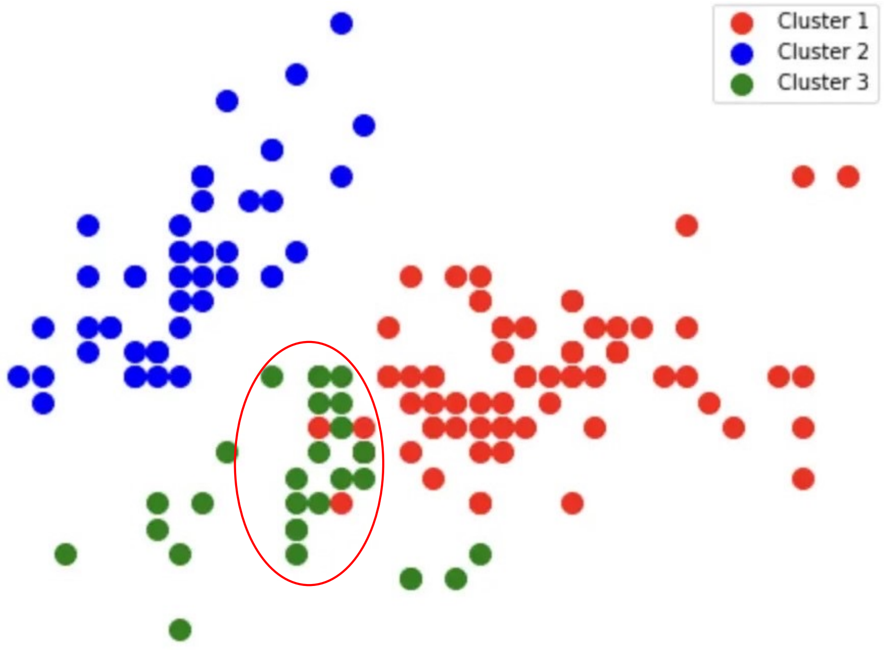
Perfect clustering!

Clustering quality

We have three red nodes graphically into the green cluster. Is that a problem for clustering quality?

The visual position of the nodes doesn't matter, if we have labelled data!

We only need to check that red nodes have really the red label



Clustering quality

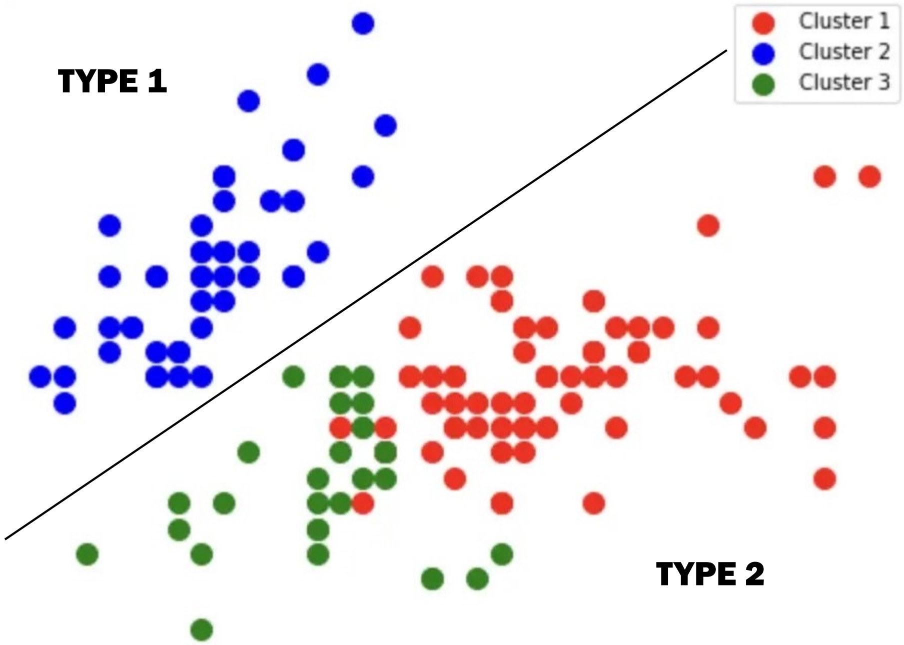
How can we define a good clustering result for labelled data?

Perfect clustering: one cluster for each type, composed only by elements of that type

Example: we have 3 clusters, but only two different labels (type 1 and type 2)

No perfect clustering!

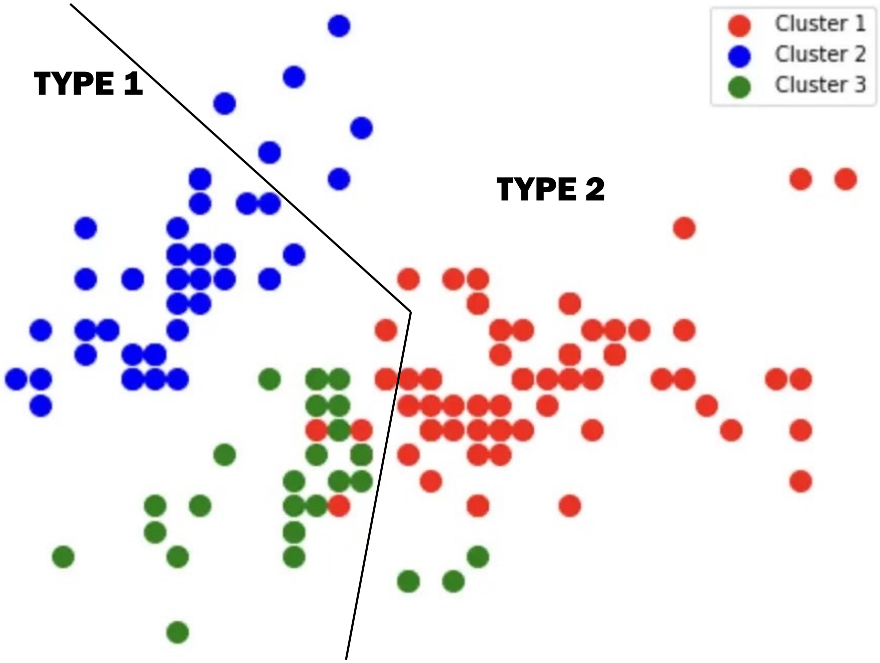
Good quality because type 1 nodes and type 2 nodes are not mixed



Clustering quality

How can we define a good clustering result for labelled data?

Perfect clustering: one cluster for each type, composed only of elements of that type



Example: we have 3 clusters, but only two different labels (type 1 and type 2)

No perfect clustering!

Low quality because type 1 nodes and type 2 nodes are mixed, we have elements of the three clusters for each of them

Clustering quality

- RAND INDEX

C: ground truth assignment

K: clustering assignment

a: pairs of elements that are in the same set in C and in the same set in K

b: pairs of elements that are in different sets in C and in different sets in K

$C_2^{n_{samples}}$: number of possible unordered pairs in the dataset

$$\text{RI} = \frac{a+b}{C_2^{n_{samples}}}$$

Clustering quality

- Homogeneity and Completeness

Homogeneity: a measure that identifies if each cluster contains only members of a single type

Completeness: a measure that identifies if all members of a given type are assigned to the same cluster

A good clustering is homogeneous and complete

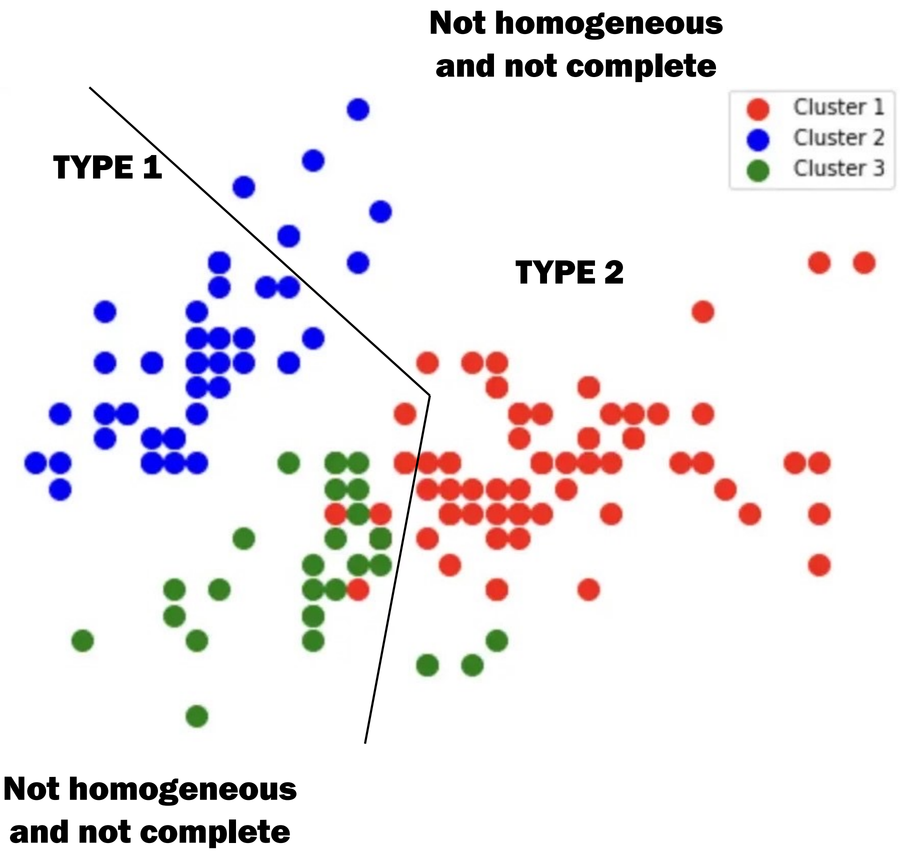
V_measure: harmonic mean between homogeneity and completeness, weighted by β factor

$$v = \frac{(1+\beta) \times \textbf{homogeneity} \times \textbf{completeness}}{(\beta \times \textbf{homogeneity} + \textbf{completeness})}$$

Clustering quality

Homogeneity: each cluster contains only members of a single type?

Completeness: all members of a given type are assigned to the same cluster?

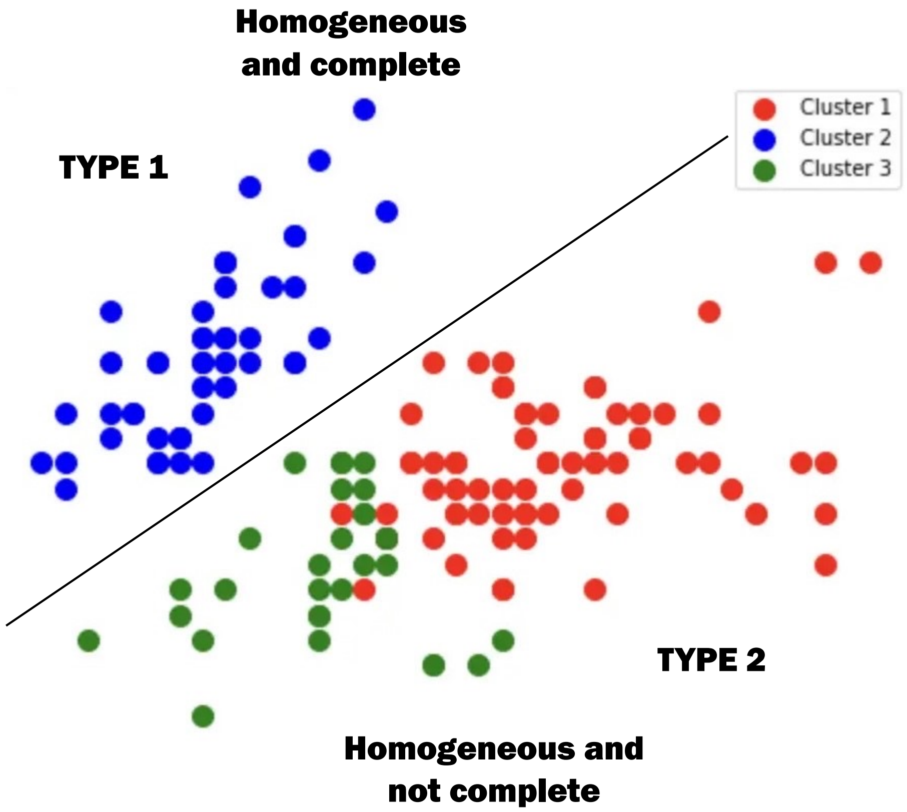


Not homogeneous and not complete

TYPE 1

TYPE 2

Not homogeneous and not complete



Homogeneous and complete

TYPE 1

TYPE 2

Homogeneous and not complete

Clustering quality

- Fowlkes-Mallows index

TP: pairs that belong to the same clusters in both the ground-truth labels and the predicted labels

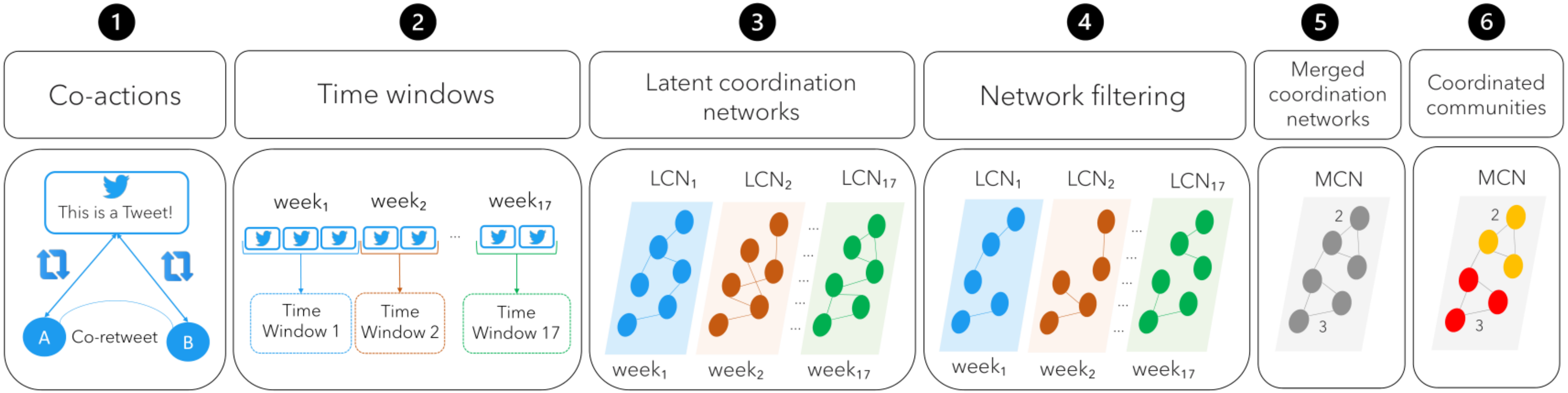
FP: pairs that belong to the same clusters in the ground-truth labels and not in the predicted labels (incomplete)

FN: pairs that belong in the same clusters in the predicted labels and not in the ground-truth labels (not homogeneous)

$$\mathrm{FMI} = \frac{\mathrm{TP}}{\sqrt{(\mathrm{TP}+\mathrm{FP})(\mathrm{TP}+\mathrm{FN})}}$$

CASE STUDY

Detection Task



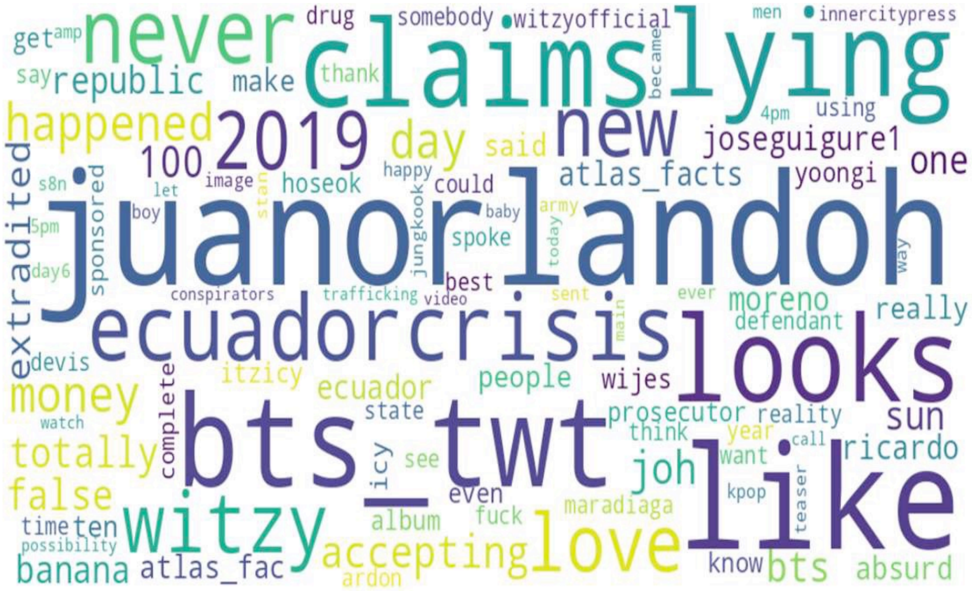
Case study: **Inauthentic information operations provided by Twitter**

Usage of **Weber's** detection framework

Cima, L., Mannocci, L., Avvenuti, M., Tesconi, M., Cresci S. (2024). Coordinated Behavior in Information Operations on Twitter. IEEE Access. 2024

Malicious Dataset

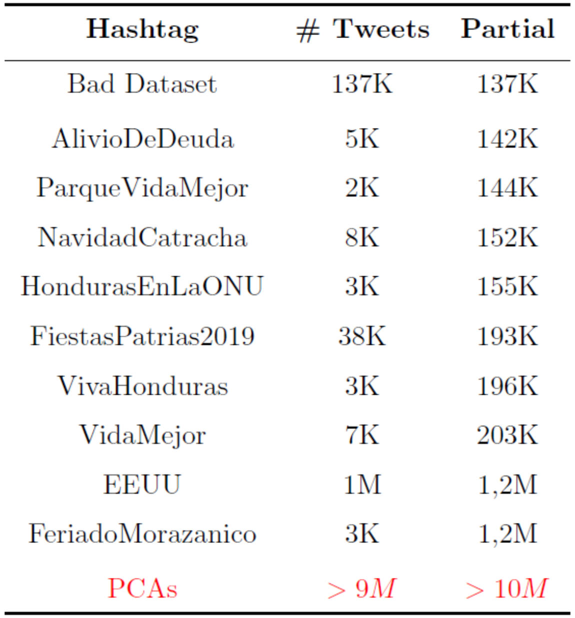
- Two kinds of datasets: “malicious” (from inauthentic IOs) and “genuine” (from harmless conversations about the same topics)
- Twitter provides malicious datasets about discovered inauthentic IOs, to be used as a ground truth for analyses
- Honduras campaign: compulsive retweeting of president’s communications (@JuanOrlandoH)



Honduras

Genuine Dataset

- Twitter provides only “malicious” datasets, used as a ground truth.
  The counterpart has to be built using search APIs
- TOP-N used hashtag strategy
- The number N of considered hashtags depends on a virtual machine's capacity for analysis (about 3M tweets)
- **Policy “all or nothing”** when the limit of tweets is reached

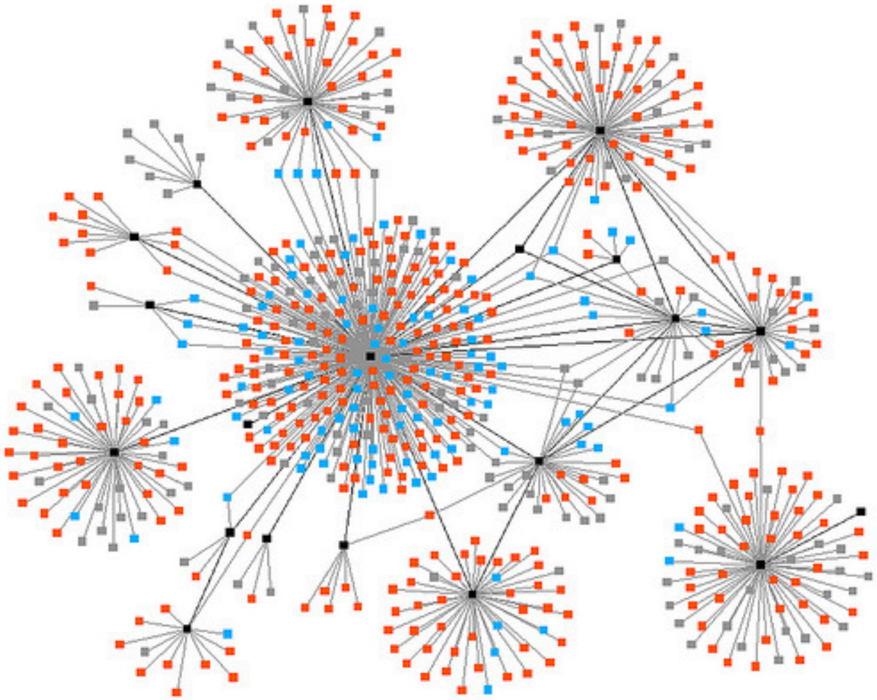


Hashtag	# Tweets	Partial
Bad Dataset	137K	137K
AlivioDeDeuda	5K	142K
ParqueVidaMejor	2K	144K
NavidadCatracha	8K	152K
HondurasEnLaONU	3K	155K
FiestasPatrias2019	38K	193K
VivaHonduras	3K	196K
VidaMejor	7K	203K
EEUU	1M	1,2M
FeriadoMorazanico	3K	1,2M
PCAs	$> 9M$	$> 10M$

Honduras

Purpose

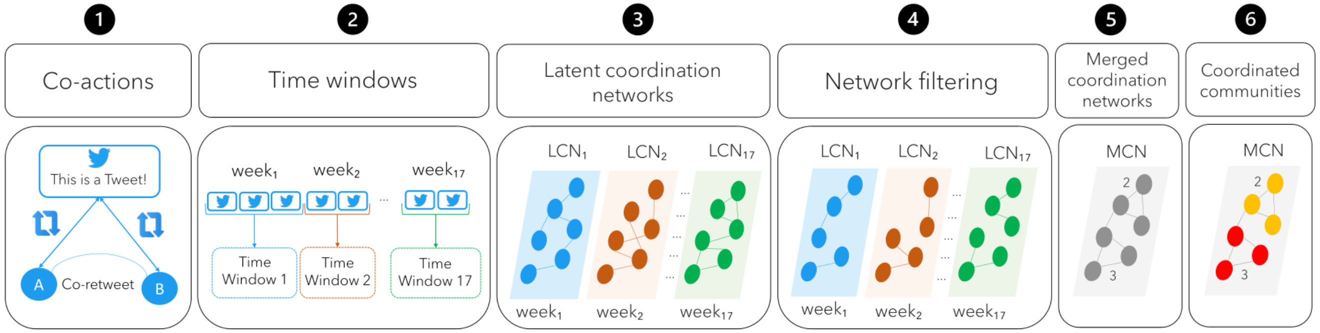
Isolate inauthentic coordinated communities (malicious) from the authentic ones (genuine) using network science



Computational Limits

- **Computational Time**
- **Memory**
- **Disk**
- **Connectivity**
- **…**

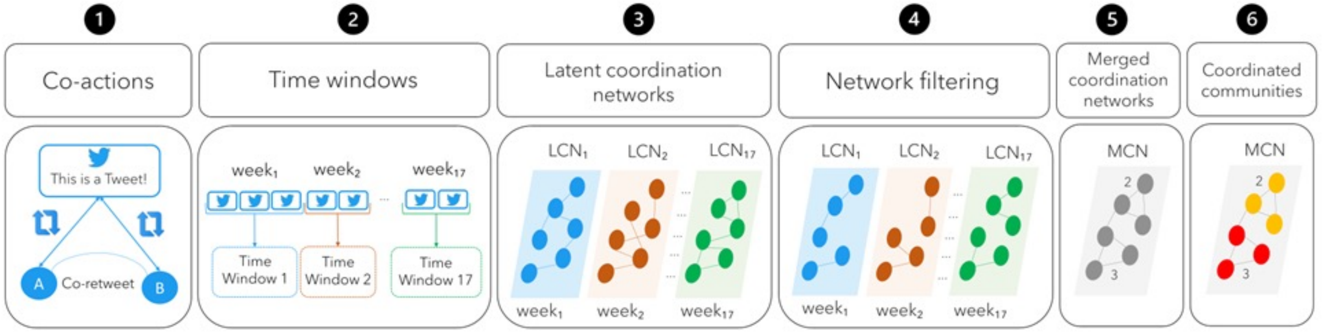
Co-Actions



- Retweets used as interaction primitive

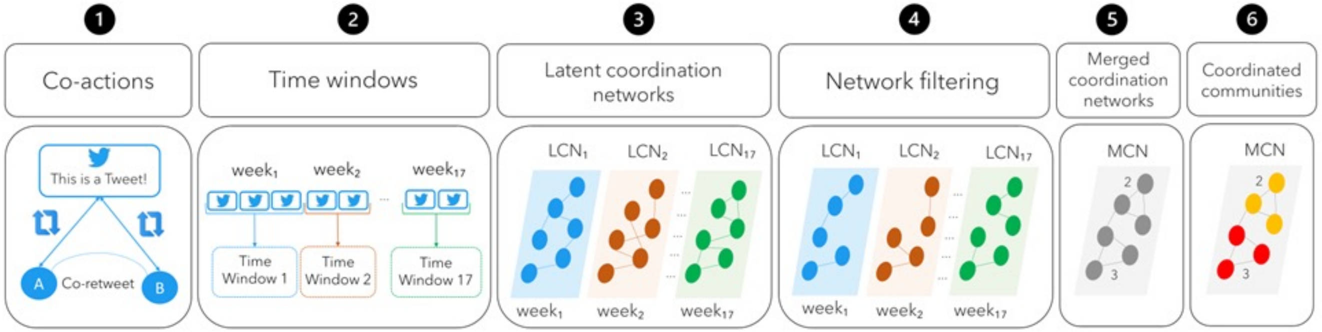
- 980K (77% of the whole dataset) retweets

Time-Windows

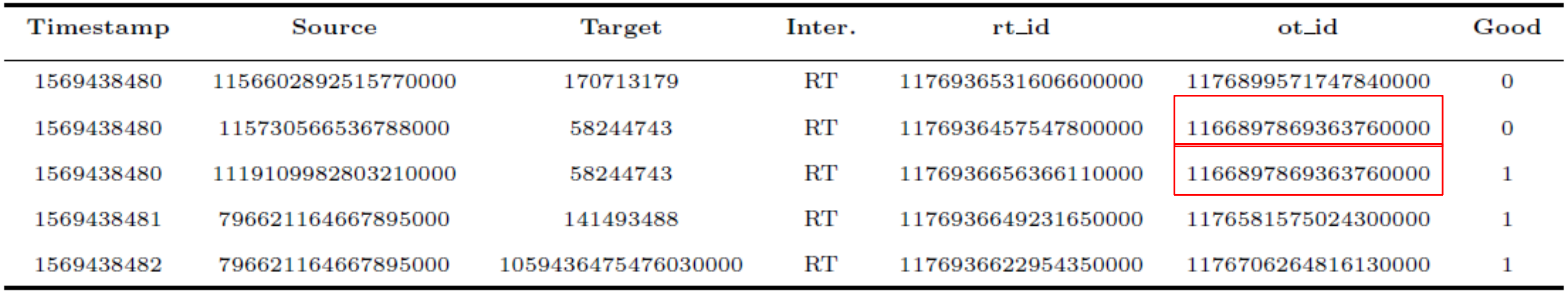


- The datasets are divided into non-overlapping time windows
- Time window: a week

Latent Coordination Networks (LCNs)



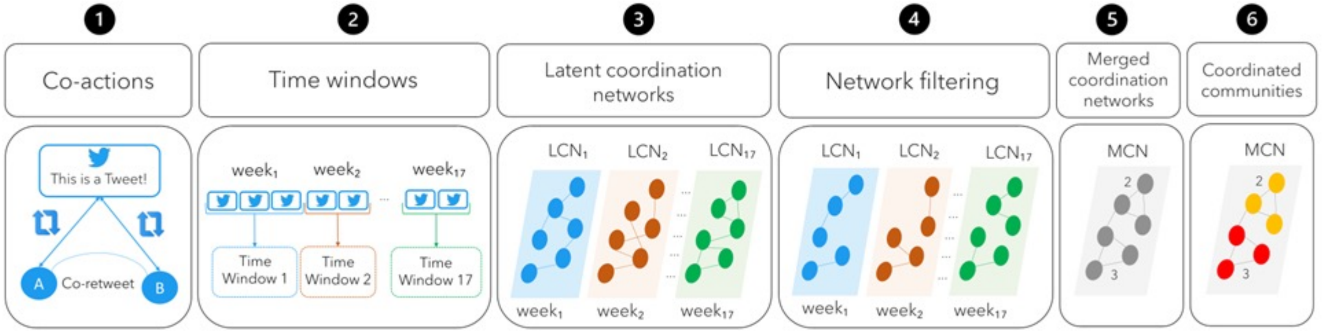
Good indicates the provenience dataset (malicious or genuine)



Timestamp	Source	Target	Inter.	rt_id	ot_id	Good
1569438480	1156602892515770000	170713179	RT	1176936531606600000	1176899571747840000	0
1569438480	115730566536788000	58244743	RT	1176936457547800000	1166897869363760000	0
1569438480	1119109982803210000	58244743	RT	1176936656366110000	1166897869363760000	1
1569438481	796621164667895000	141493488	RT	1176936649231650000	1176581575024300000	1
1569438482	796621164667895000	1059436475476030000	RT	1176936622954350000	1176706264816130000	1

- If two different sources retweet the same tweet (same ot_id), they become nodes on the network (co-retweet)
- An edge connects the two nodes
- The higher the number of co-retweets, the higher the edge weight
- One LCN graph for each time window

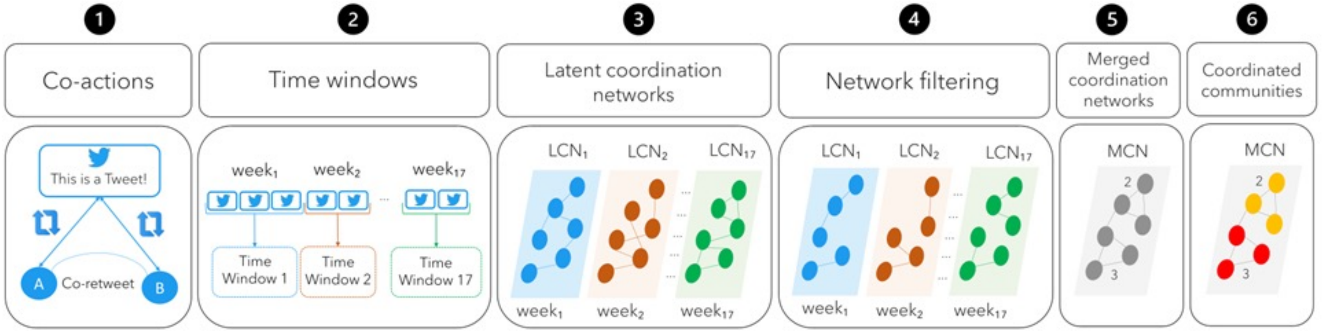
Network filtering



- First, a threshold filtering to delete edges with low weights (random connections)

- Then a complex filtering step, based on the FSA_V algorithm, to obtain highly coordinated communities (HCCs) from all the 17 LCN graphs
- One HCC graph for each time window

Merged Coordination Network

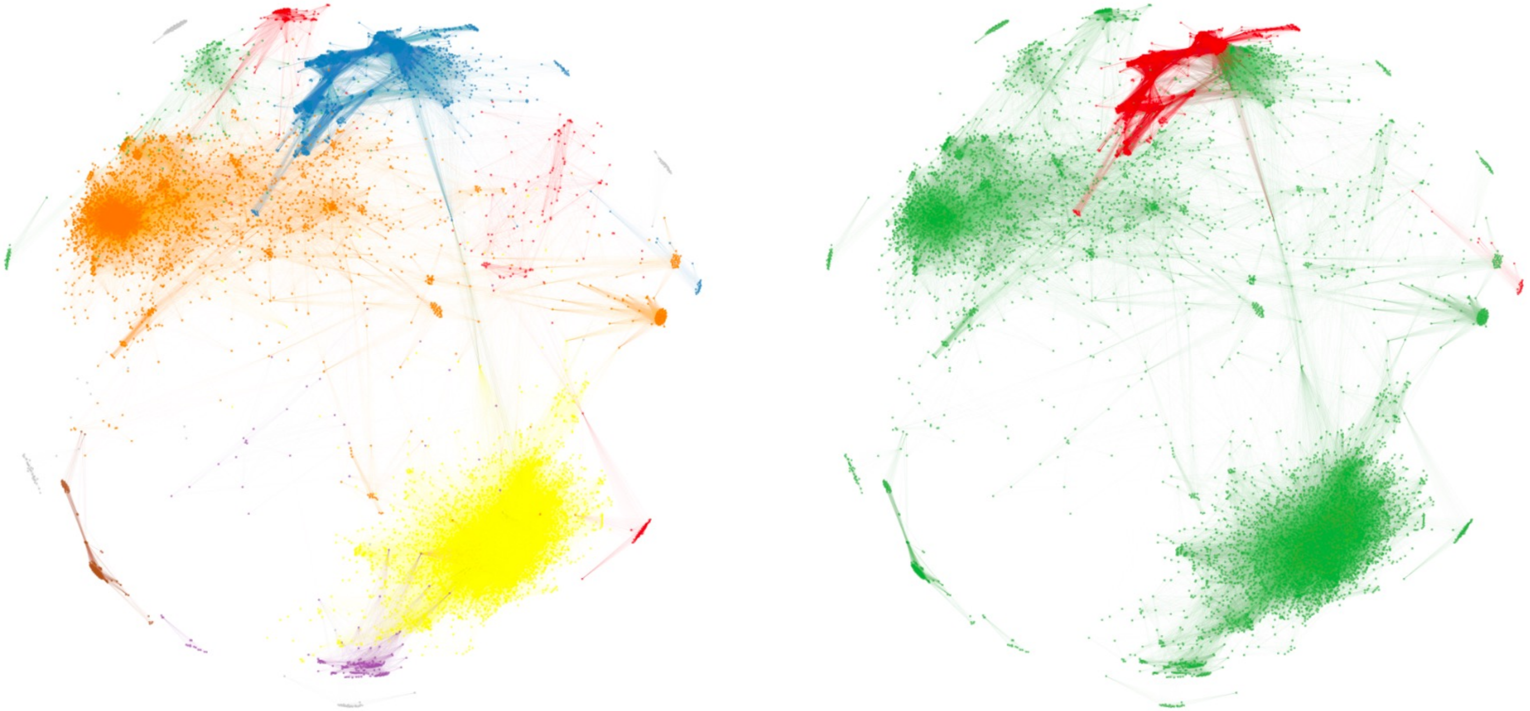


- Final merge of the HCC graphs, to obtain a single HCCs graph, which covers the whole analyzed period

- Merge done using inference rules

Characterization Task

Perfect clustering: one cluster for each class, composed only by elements of that class



In our binary example, perfect clustering is obtained if we have only two communities: one contains only malicious (red) nodes, and the other only genuine (green) nodes

Time For Exercises!

Why can’t we do in our exercises the previous detection steps?

https://github.com/weberdc/find_hccs