University of Amsterdam

Amsterdam School of Communication Research

NEWS FLOWS

European Research Council

Mapping Dark Platforms: Modelling Information Flows on Telegram

Rupert Kiddle[1], Mónika Simon[1], Kasper Welbers[2], Anne Kroon[1], Damian Trilling[1]

1: The University of Amsterdam (UvA)
2: The Vrije University (Amsterdam)

Telegram

Private: end-to-end encrypted messaging.

Public: *chats* (many-to-many) and *channels* (few-to-many).

A home for the **deplatformed**: permissively governed, low-moderated space.

Relatively free of **algorithmic interference**.

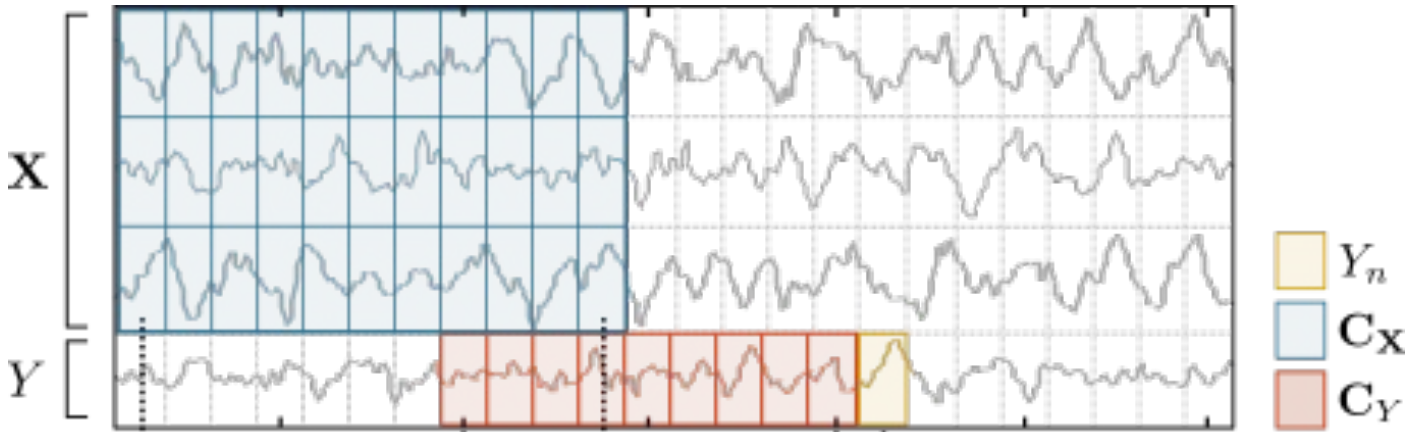
Information about network structure is very limited!

What we have: full messaging text-streams of chats/channels.

What we lack: relational data describing user, chat/channel associations.

from agents to flows -

Multivariate Transfer Entropy (mTE) -



IDTxl package for Python: P. Wollstadt, J. T. Lizier, R. Vicente, C. Finn, M. Martinez-Zarzuela, P. Mediano, L. Novelli, M. Wibral (2018). *IDTxl: The Information Dynamics Toolkit xl: a Python package for the efficient analysis of multivariate information dynamics in networks.* Journal of Open Source Software, 4(34), 1081. https://doi.org/10.21105/joss.01081.

mTE applied -
‘Network Toxicity’

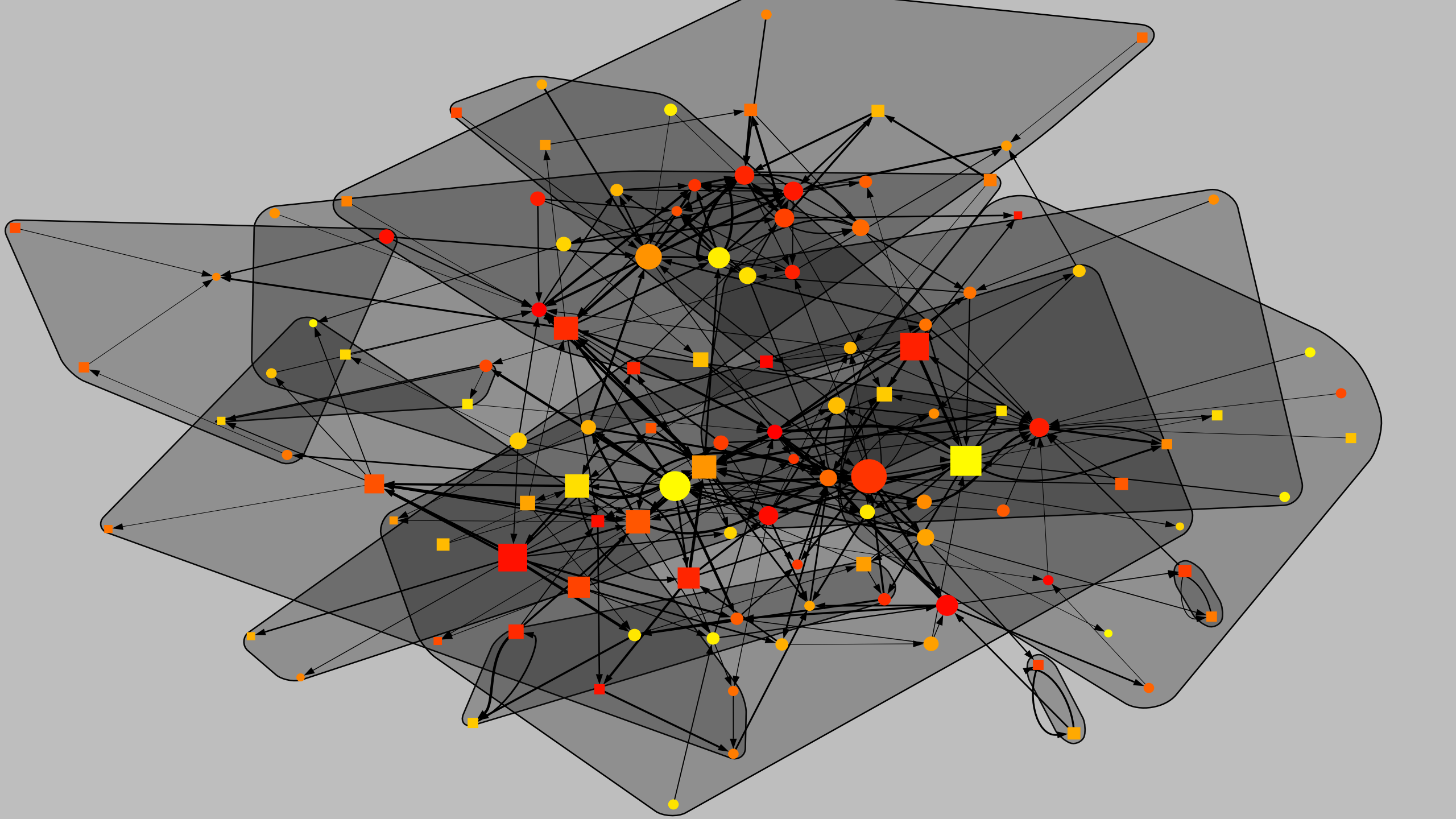
- **Aim**: to infer an *effective network* that describes the flows of toxicity between chats/channels on Telegram.

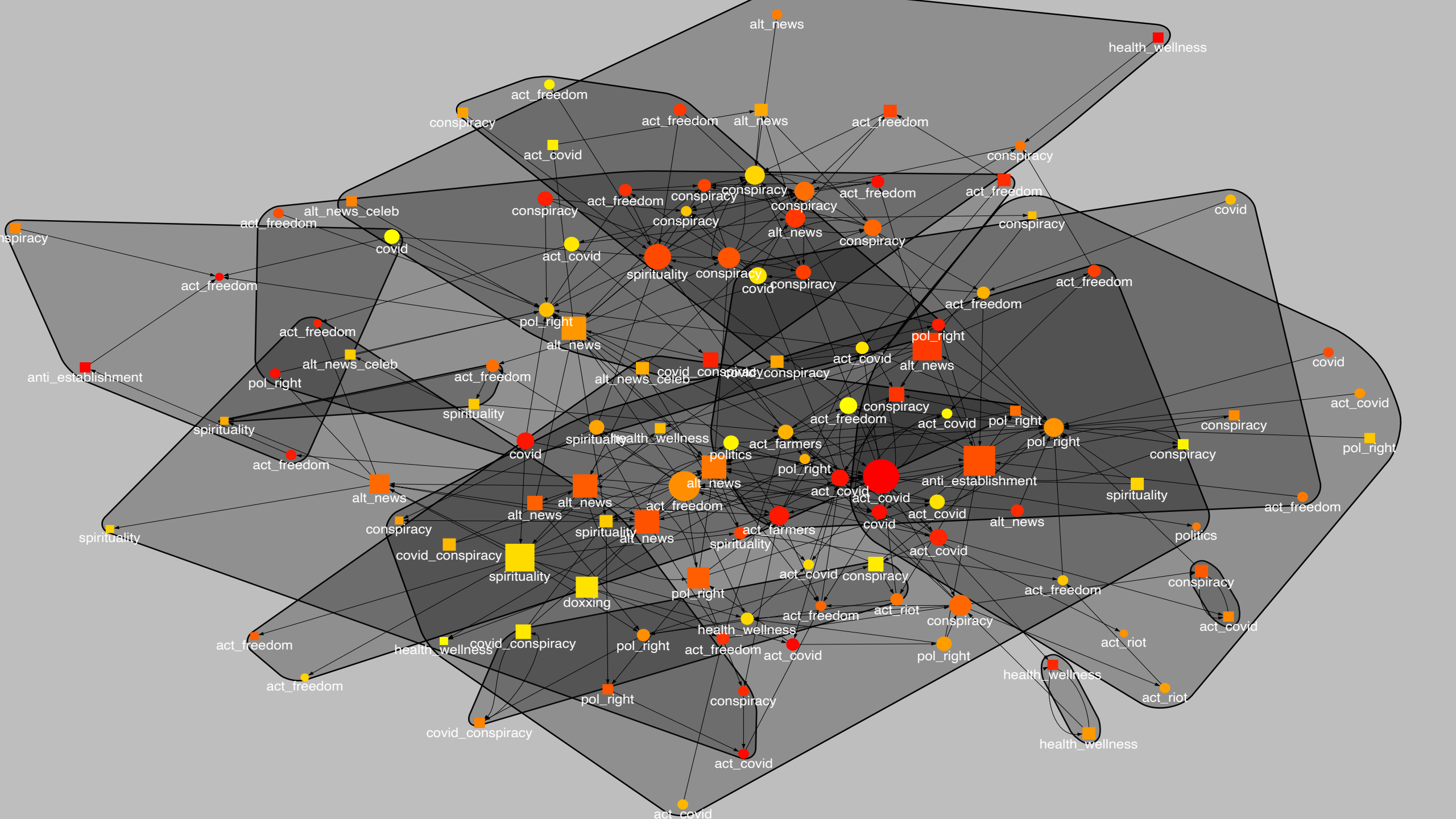
- **Toxicity**: *hateful, hurtful or marginalizing language that makes a person want to leave a conversation*.

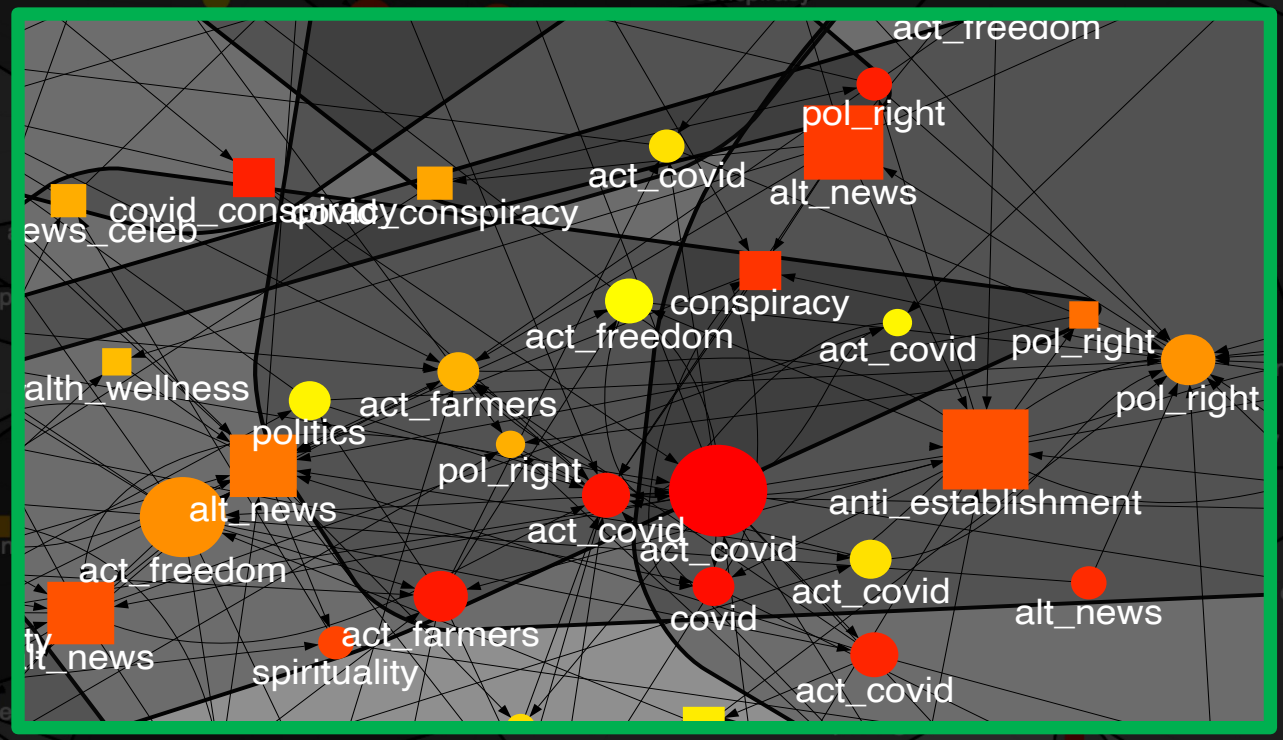
> A growing threat to individuals & platforms.
> Telegram users are particularly at-risk.
> Existing approaches would struggle here.

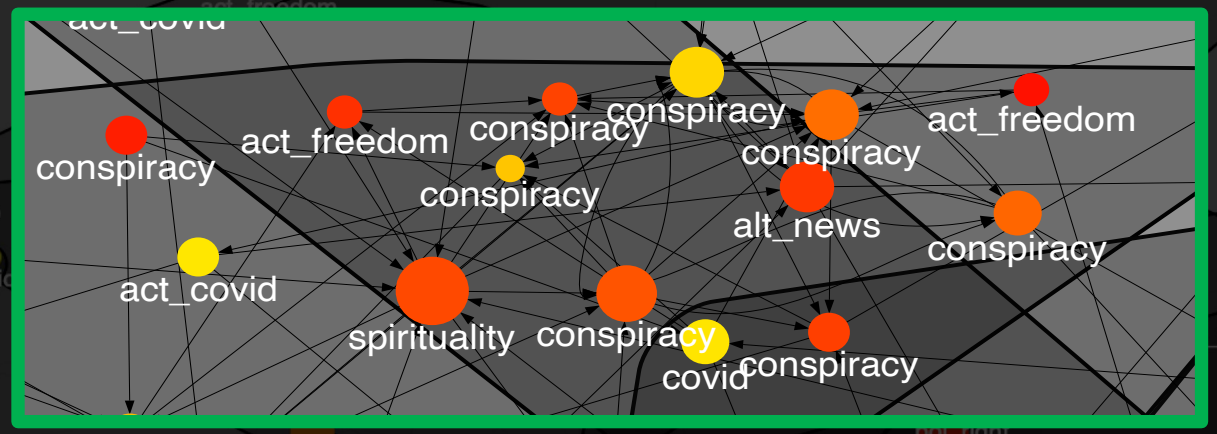
Method -

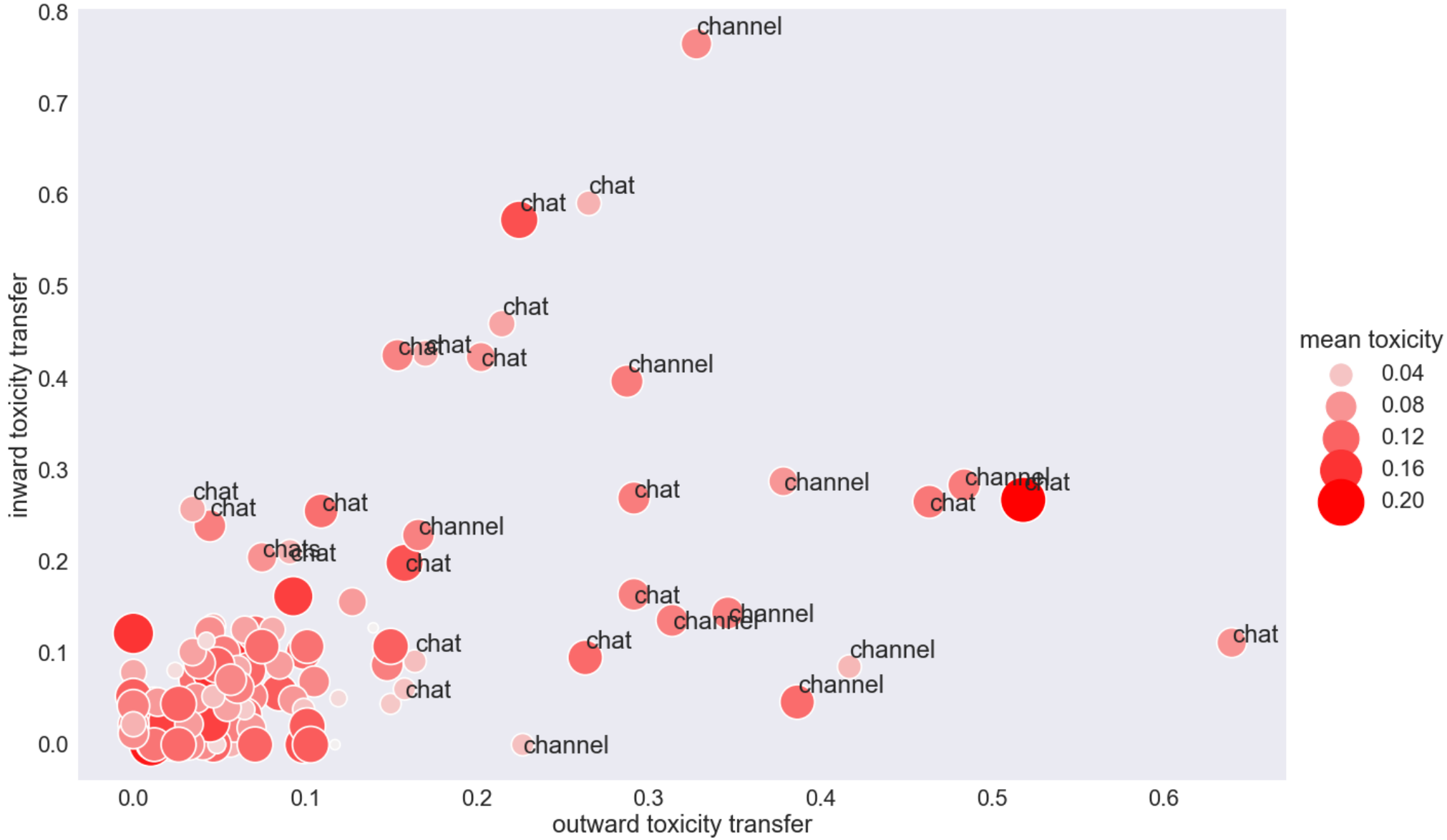
1. **Data**: 6m snapshot (1M messages) of Dutch ‘Telegramsphere’ ---- *Jan ‘21 to Jun ’21.*
2. **Classification***:* messages → *toxicity scores* with the *Perspective API (Google).*
3. **Resampling**: 5-minute windows; average of scores; 24-hour leeway.
4. **Estimation:** linear estimator using **IDTxl,** 30-minute maximum lag.
5. **Aggregation:** *n for each i,j / max(n) for any i,j = ‘toxicity transfer’.*
6. **Pruning:** disparity filter: 3% of edges retained (p=0.05).




alt_news
health_wellness
act_freedom
conspiracy
act_freedom
alt_news
act_freedom
act_covid
conspiracy
conspiracy
act_freedom
act_freedom
conspiracy
act_freedom
alt_news_celeb
conspiracy
conspiracy
conspiracy
conspiracy
covid
act_freedom
alt_news
conspiracy
conspiracy
covid
act_covid
spirituality
conspiracy
covid
conspiracy
act_freedom
act_freedom
act_freedom
pol_right
alt_news
act_freedom
pol_right
alt_news_celeb
anti_establishment
pol_right
act_freedom
alt_news_celeb
covid_conspiracy
conspiracy
act_covid
alt_news
covid
act_freedom
spirituality
spirituality
conspiracy
act_freedom
act_covid
pol_right
covid
spirituality
health_wellness
act_farmers
politics
pol_right
pol_right
conspiracy
act_freedom
pol_right
conspiracy
alt_news
act_covid
act_covid
anti_establishment
spirituality
alt_news
alt_news
act_freedom
act_covid
act_covid
alt_news
act_freedom
spirituality
conspiracy
spirituality
alt_news
act_farmers
covid
politics
covid_conspiracy
spirituality
spirituality
act_covid
act_covid
conspiracy
act_freedom
conspiracy
doxxing
pol_right
act_freedom
act_riot
conspiracy
act_covid
health_wellness
act_freedom
covid_conspiracy
pol_right
act_freedom
act_covid
pol_right
act_riot
health_wellness
act_riot
act_freedom
pol_right
conspiracy
covid_conspiracy
act_covid
health_wellness
act_covid


act_freedom
pol_right
act_covid
alt_news
covid_conspiracy
conspiracy
act_freedom
conspiracy
act_covid
pol_right
alth_wellness
act_farmers
politics
pol_right
pol_right
alt_news
act_covid
anti_establishment
act_covid
act_freedom
act_covid
alt_news
act_farmers
covid
act_covid
spirituality
act_covid


conspiracy
act_freedom
conspiracy
conspiracy
act_freedom
conspiracy
conspiracy
alt_news
conspiracy
act_covid
spirituality
conspiracy
covid
conspiracy


inward toxicity transfer
outward toxicity transfer
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
channel
chat
mean toxicity
0.04
0.08
0.12
0.16
0.20

Concluding Remarks -

- A '***model-free***' approach to understand the **dynamic consequences of toxicity.**
  + non-linear estimators (more complex relationships)
  + different timescales (beyond 30 minutes)
  + validation methods ('unpacking')

- **mTE framework** for **information flows & influence** more broadly.
  + different representations (e.g. word embeddings)
  + beyond Telegram..

Thank you very much!

@rptkiddel

@newsflows_erc