Group Dynamics in Online Social Networks:

Communities and High-order Interactions in Time

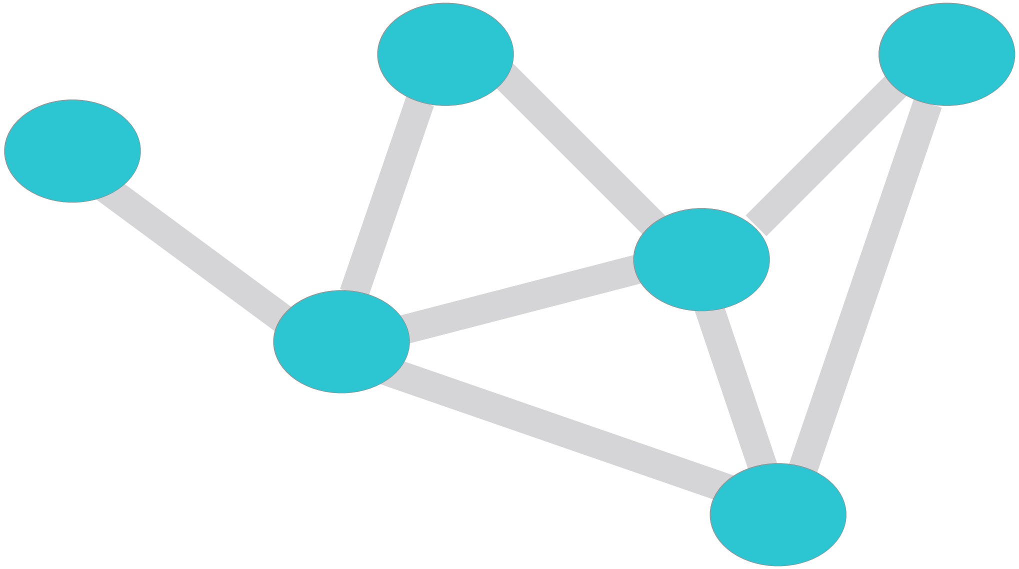
Andrea **Failla**

1. Department of Computer Science, University of Pisa, Pisa, Itay
2. Institute of Science and Information Technologies (ISTI), National Research Council (CNR), Pisa, Italy

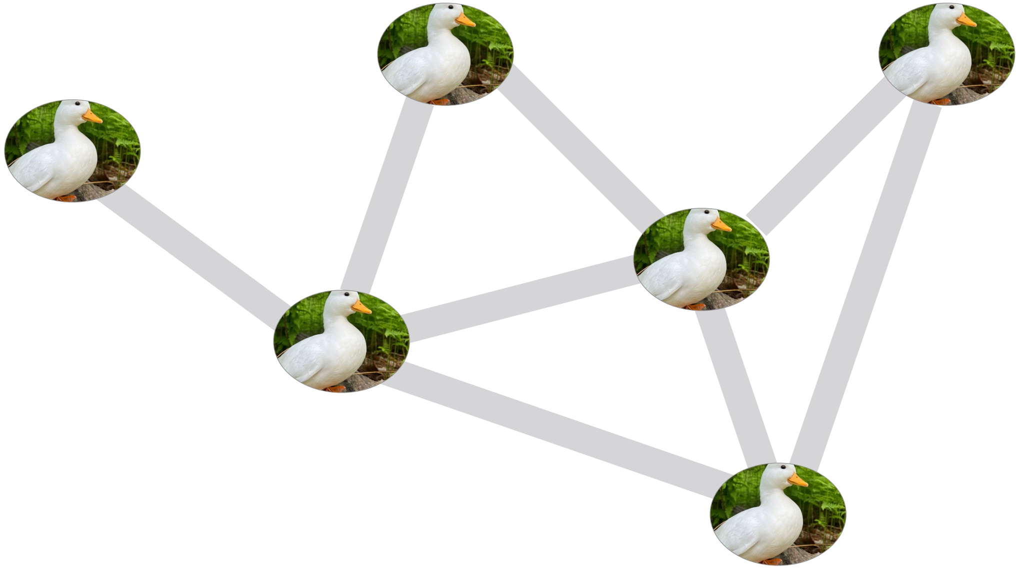
Università di Pisa

Consiglio Nazionale delle Ricerche

Let’s start!

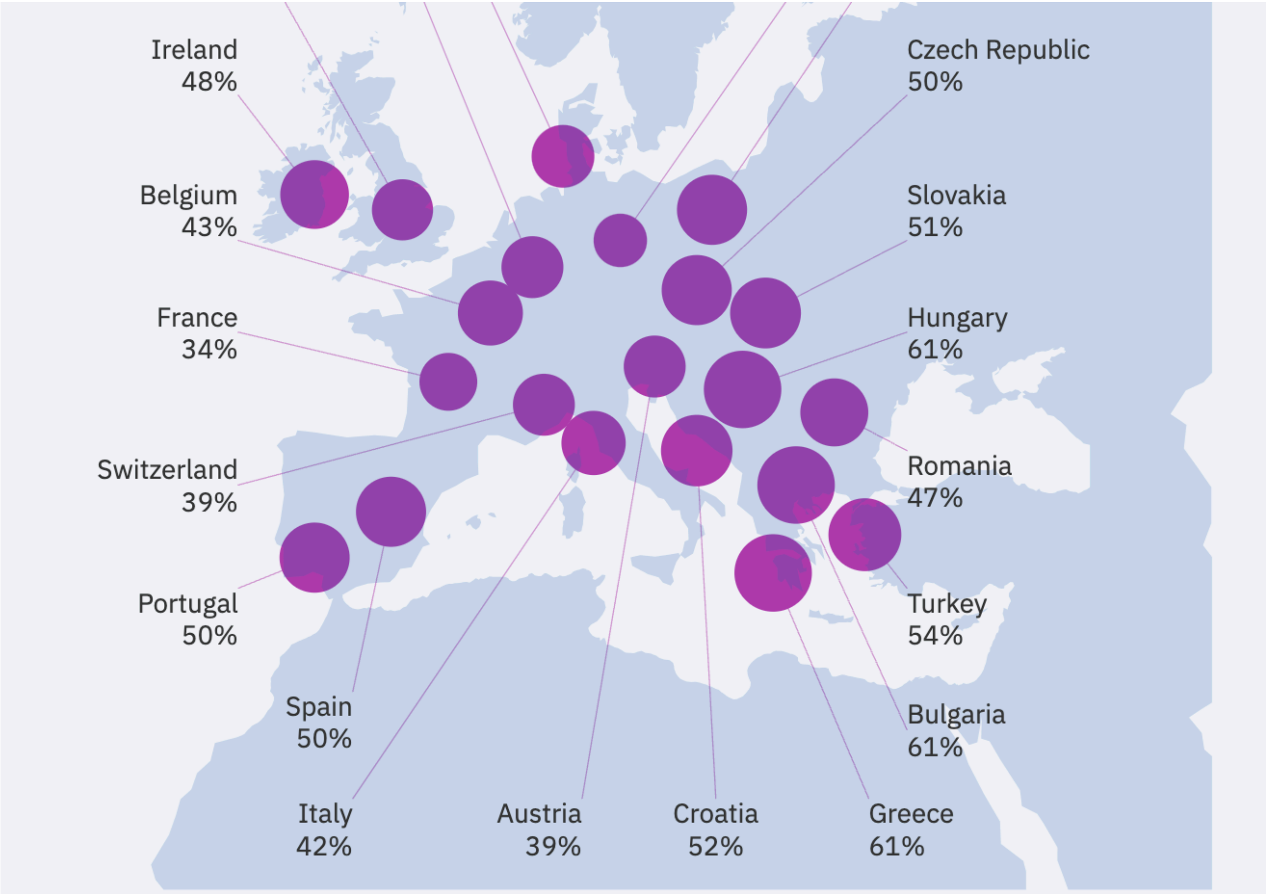


Let’s start!



Ok, let’s get serious

% of people who use social media primarily for news




Birds are always right

The Echo Chamber Effect

Environments where an opinion is reinforced, and opposing views are actively discredited

"Please estimate the likely impact (severity) of the following risks over a 2-year period"

2 years

1st Misinformation and disinformation

2nd Extreme weather events

3rd Societal polarization

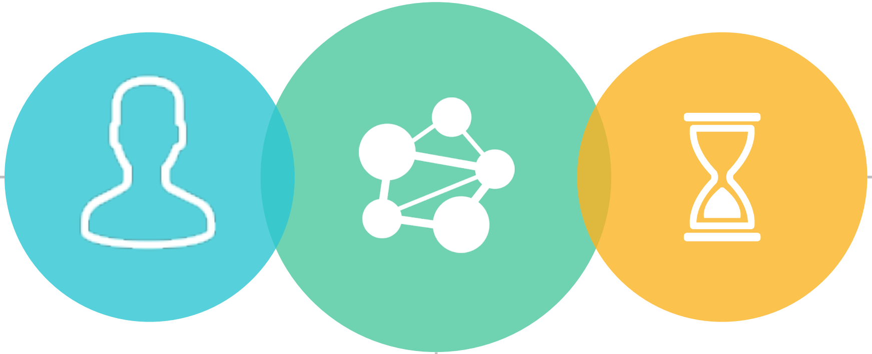
4th Cyber insecurity

5th Interstate armed conflict

…

World Economic Forum, **Global Risk Perception Survey** (2023-2024)

A general framework



User-Generated Content

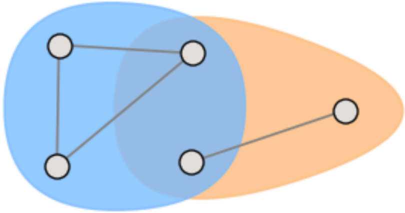
Build user *personae* from textual data

Temporal Information

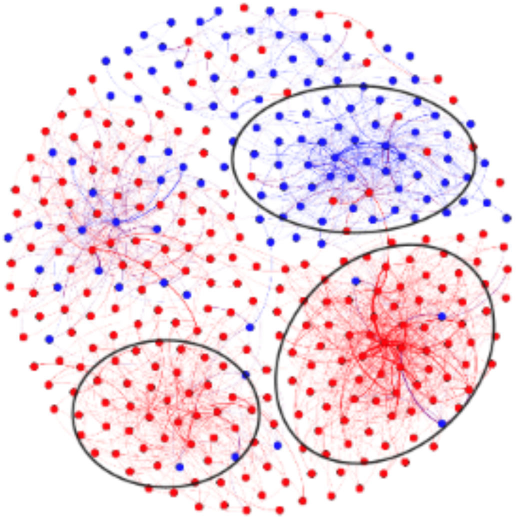
Changes in social ties and user preferences

Social Structure

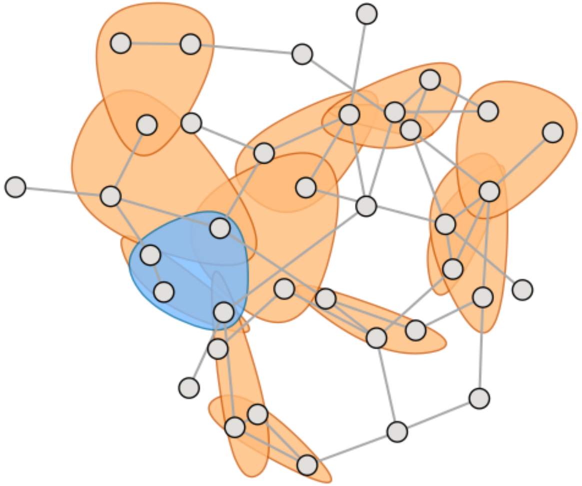
Pairwise and/or high-order network modeling



Modeling groups: Graph Communities and Hypergraphs



e.g., Echo Chambers,
fandoms, etc.



e.g., Conversations/debates

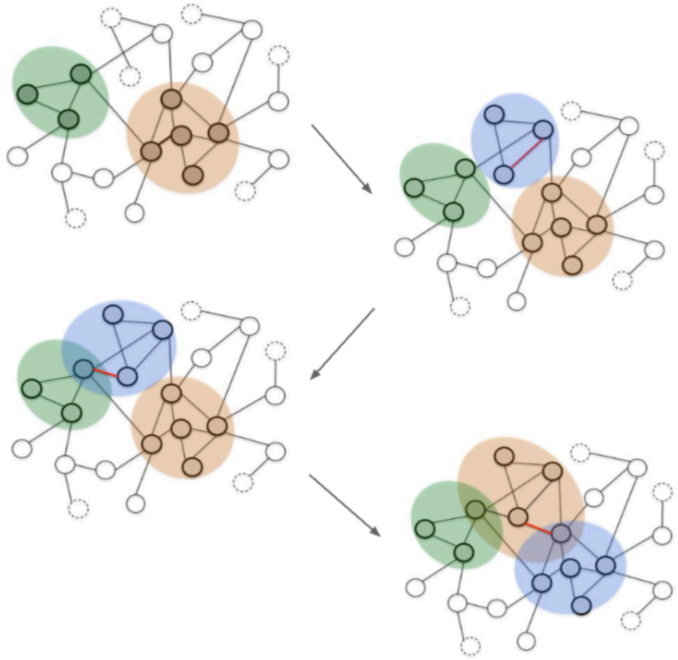
A Continuous Framework to Understand Group Evolution

Università di Pisa
Consiglio Nazionale delle Ricerche

Dynamic Communities I

Groups/Communities change over time.

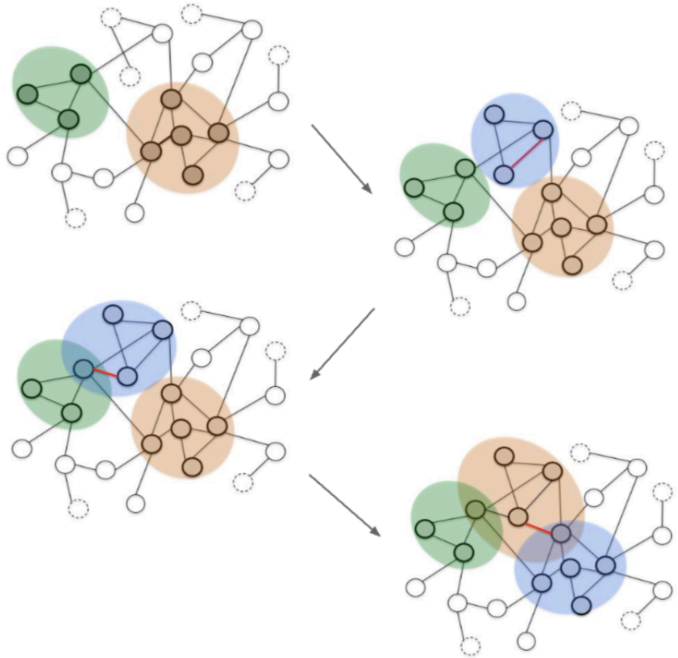
Dynamic Community Detection attempts at identifying/tracking changes in the community structure



Cazabet, R., Rossetti, G. **Challenges in community discovery on temporal networks.** Temporal Network Theory. Springer, Cham, 2019. 181-197.

Dynamic Communities II

1. **Identify communities** at each time step
2. **Match communities** across adjacent time steps
3. **Look for events** depending on some criteria

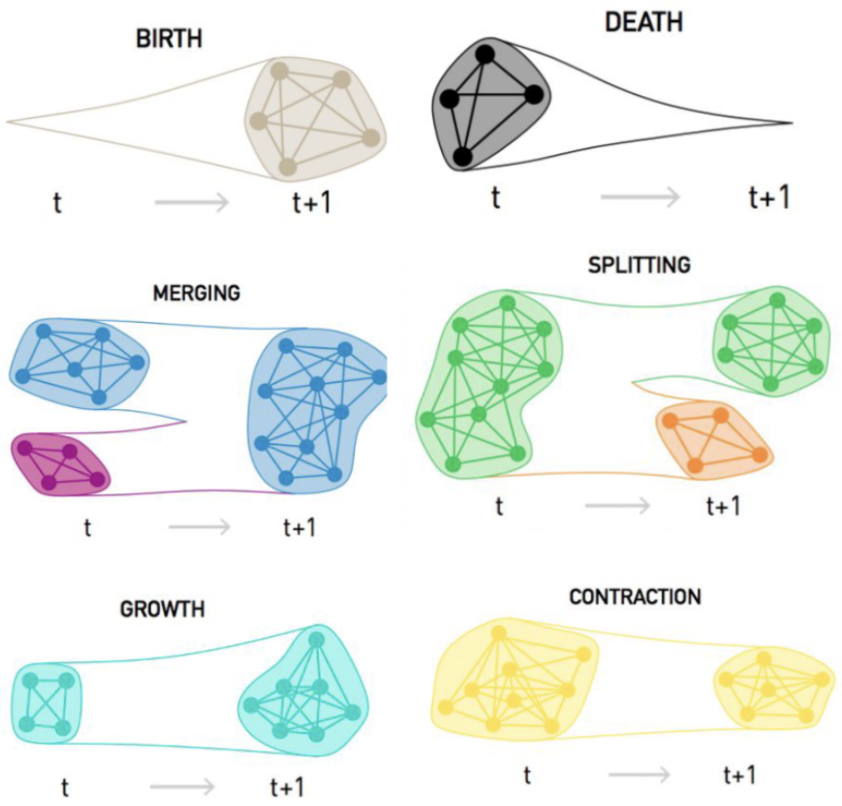


Cazabet, R., Rossetti, G. **Challenges in community discovery on temporal networks.** Temporal Network Theory. Springer, Cham, 2019. 181-197.

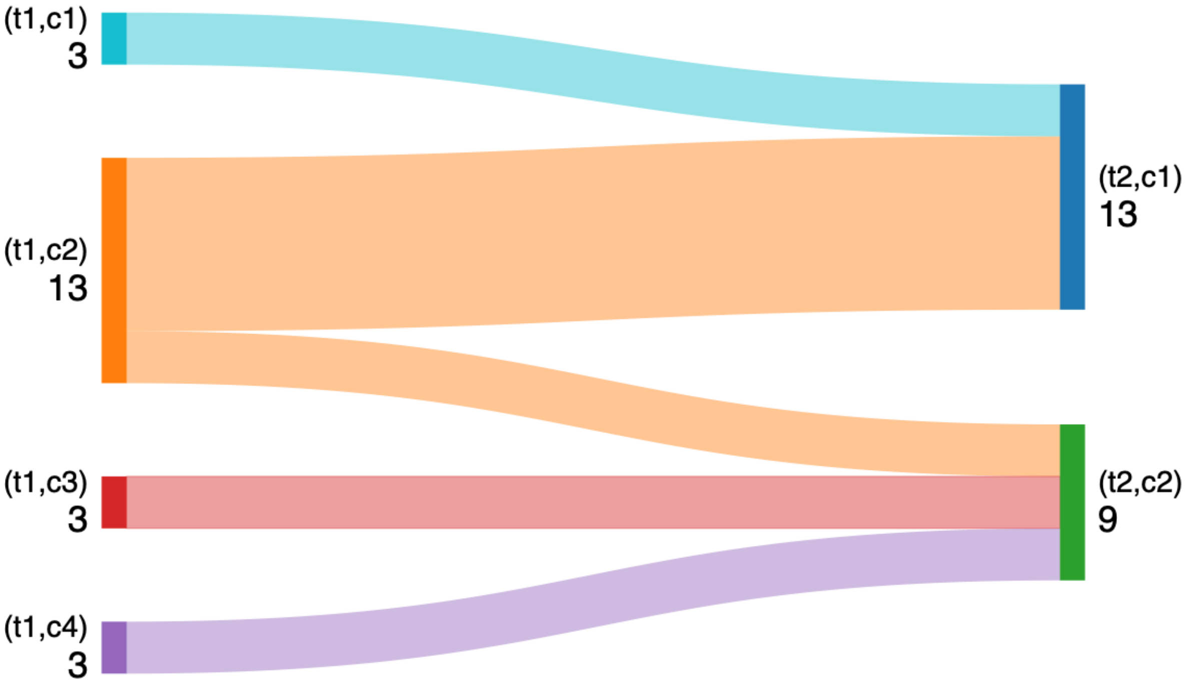
Community Events

Group evolution has traditionally been described with **strict categories**

Several event taxonomies exist, as well as methods to detect events

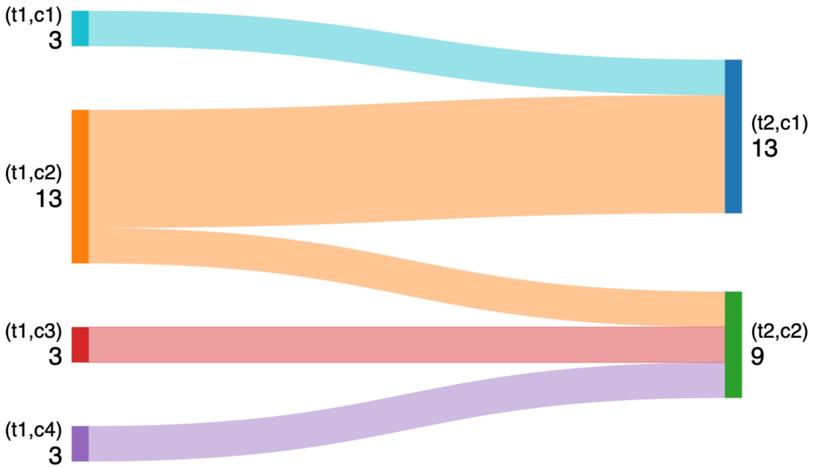


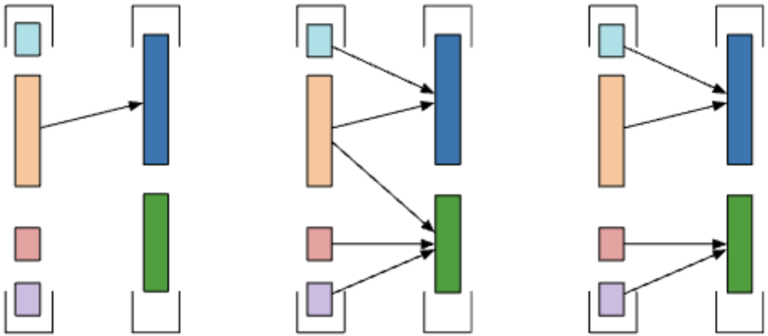
Real temporal data is *more complex*!



Failla A., Cazabet R., Rossetti G., Citraro S., **Redefining Event Types and Group Evolution in Temporal Data**, 2024, arXiv preprint, arXiv:2403.06771

Real temporal data is *more complex*!





Which one
should I choose?

Failla A., Cazabet R., Rossetti G., Citraro S., **Redefining Event Types and Group Evolution in Temporal Data**, 2024, arXiv preprint, arXiv:2403.06771

The Facets of Group Evolution

Unicity

One vs. Many sources

Identity

Part vs. Whole

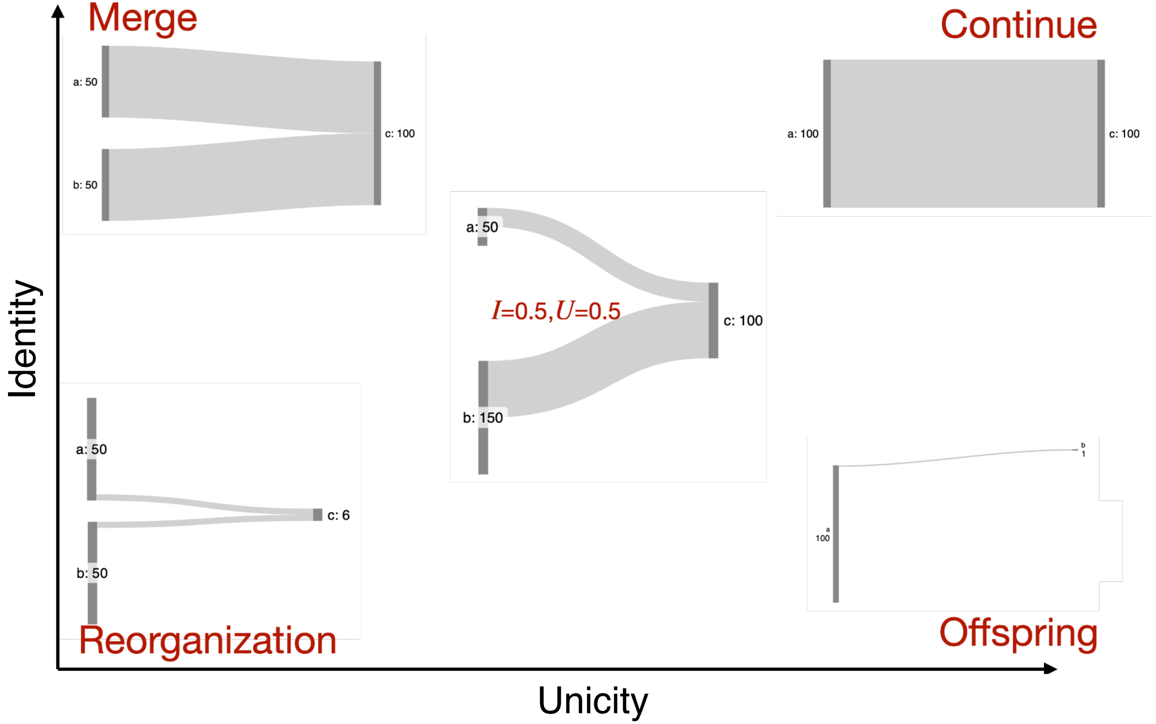
Outflow

New vs. Old nodes

Let’s focus on Unicity and Identity…

Unicity: one vs. many

Identity: part vs. whole



Failla A., Cazabet R., Rossetti G., Citraro S., **Redefining Event Types and Group Evolution in Temporal Data**, 2024, arXiv preprint, arXiv:2403.06771

Event Taxonomy

Quantify how “close” a community is from undergoing an **archetypal event**

EVENT	Measure
Birth	U (1-I) O
Accumulation	(1-U) (1-I) O
Continue	U I (1-O)
Merge	(1-U) I (1-O)
Offspring	U (1-I) (1-O)
Reorganization	(1-U)(1-I)(1-O)
Growth	U I O
Expansion	(1-U) I O

Failla A., Cazabet R., Rossetti G., Citraro S., **Redefining Event Types and Group Evolution in Temporal Data**, 2024, arXiv preprint, arXiv:2403.06771

Event Taxonomy

Events change their semantics depending on the temporal direction

Intuition: a merge looks like a split, when reversing the flow of time.

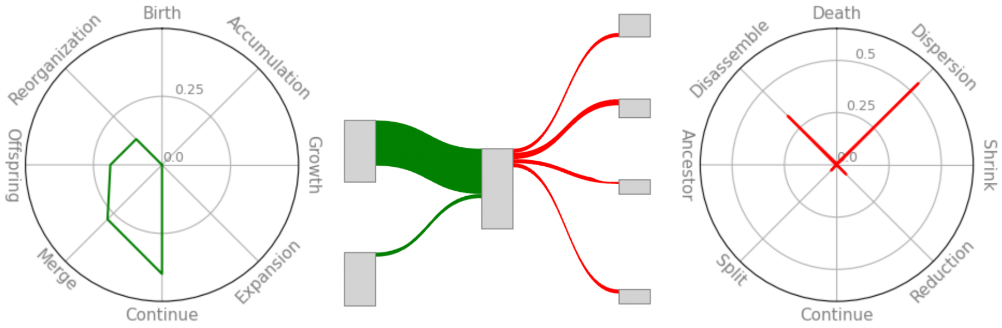
BACKWARD	FORWARD	Measure
Birth	Death	U (1-I) O
Accumulation	Dispersion	(1-U) (1-I) O
Continue	Continue	U I (1-O)
Merge	Split	(1-U) I (1-O)
Offspring	Ancestor	U (1-I) (1-O)
Reorganization	Disassemble	(1-U)(1-I)(1-O)
Growth	Shrink	U I O
Expansion	Reduction	(1-U) I O

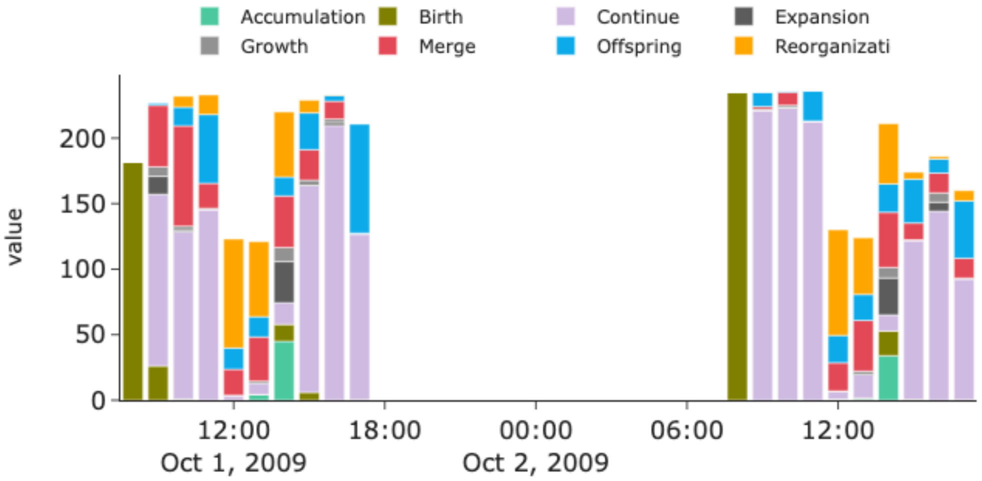
Failla A., Cazabet R., Rossetti G., Citraro S., **Redefining Event Types and Group Evolution in Temporal Data**, 2024, arXiv preprint, arXiv:2403.06771

Event Analysis

- Specific Events
- Global Dynamics
- Event Quality
- Temporal Stability

On OSNs: Echo Chambers' Dynamics





Failla A., Cazabet R., Rossetti G., Citraro S., **Redefining Event Types and Group Evolution in Temporal Data**, 2024, arXiv preprint, arXiv:2403.06771

On Github and PyPI

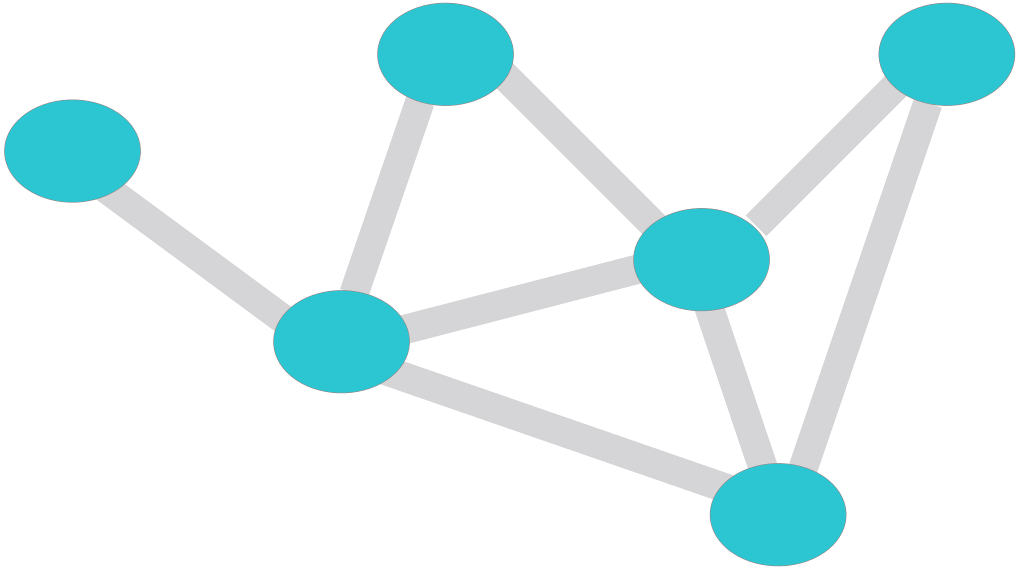
LifeCycles

CDlib

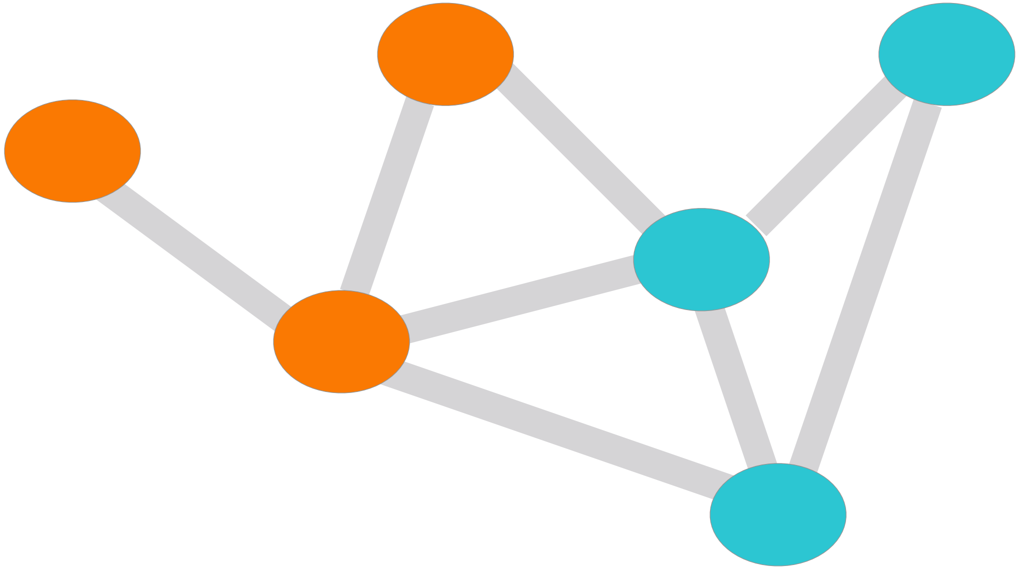
G. Rossetti, L. Milli, R. Cazabet. **CDlib: a Python Library to Extract, Compare and Evaluate Communities from Complex Networks.**
Applied Network Science Journal. 2019. DOI:10.1007/s41109-019-0165-9

Beyond Plain Graph Representations

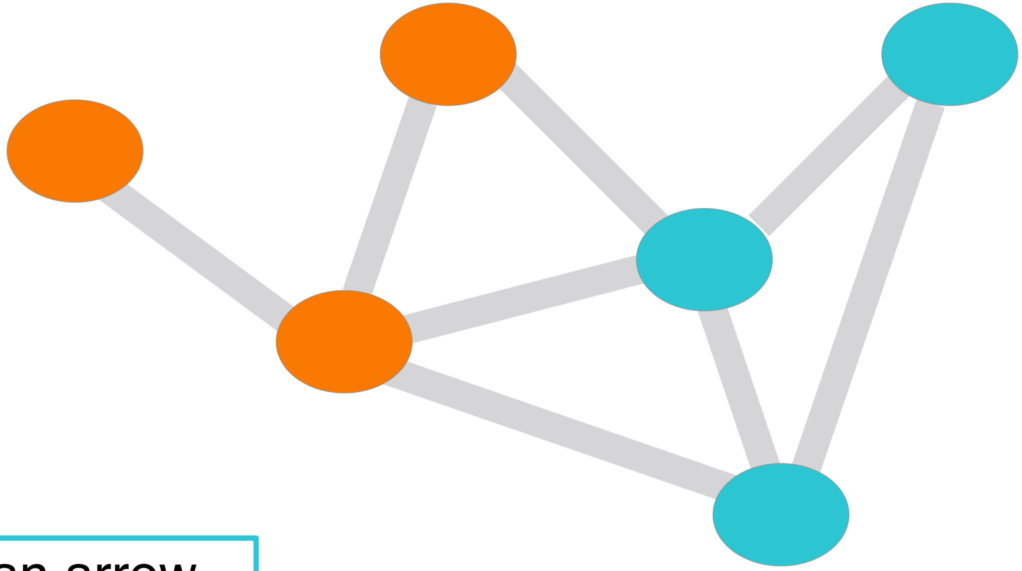
The Graph



The (node-)<u>Attributed</u> Graph



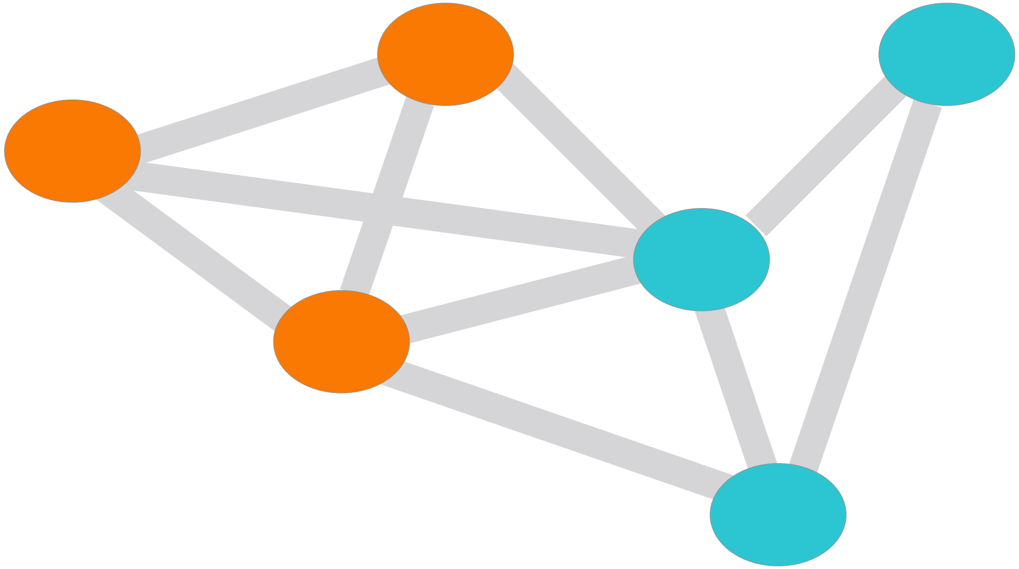
The Attributed Stream Graph



t=0

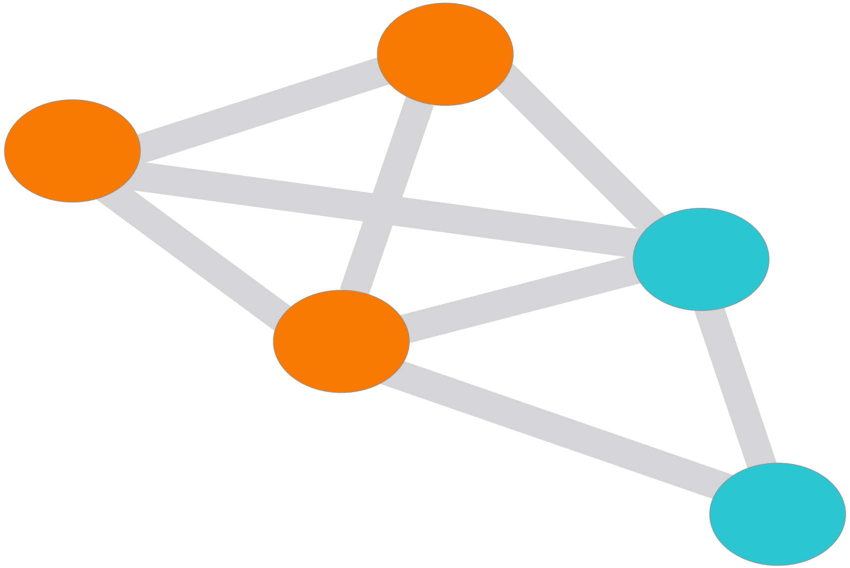
Time flies like an arrow…

The Attributed Stream Graph



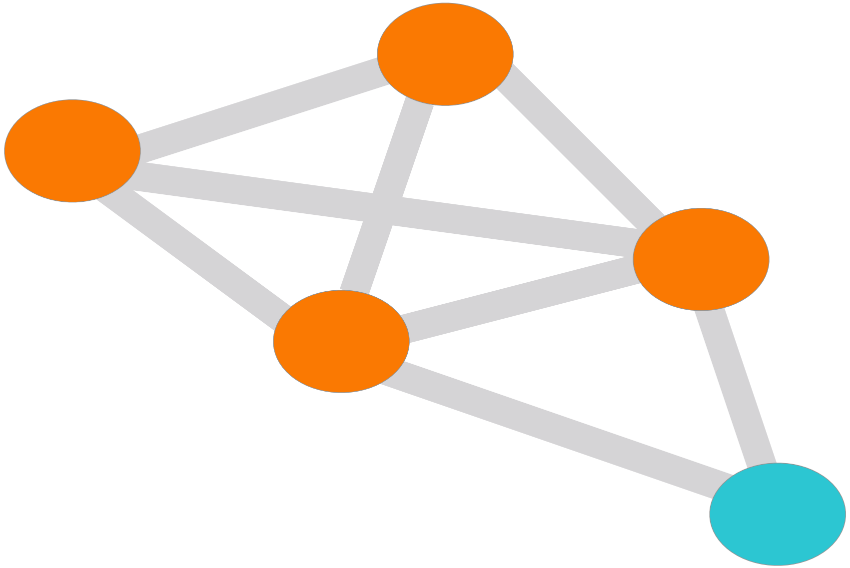
t=1

The Attributed Stream Graph



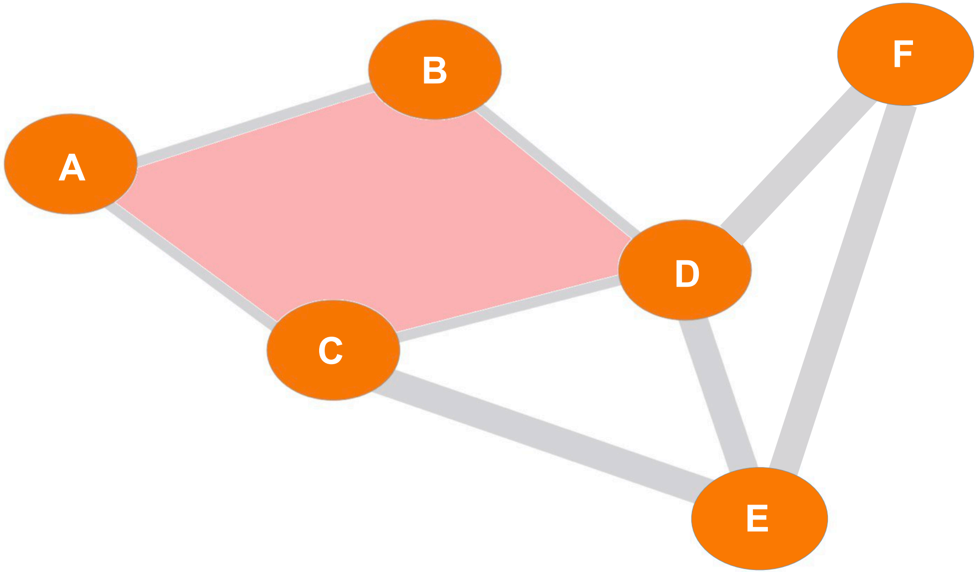
t=2

The Attributed Stream Graph



t=3

The Attributed Stream Hypergraph



t=4

Failla A, Citraro S, Rossetti G. **Attributed Stream Hypergraphs: temporal modeling of node-attributed high-order interactions.**
Applied Network Science. 2023 Jun 9;8(1):31.

The Attributed Stream Hypergraph

Node Semantics

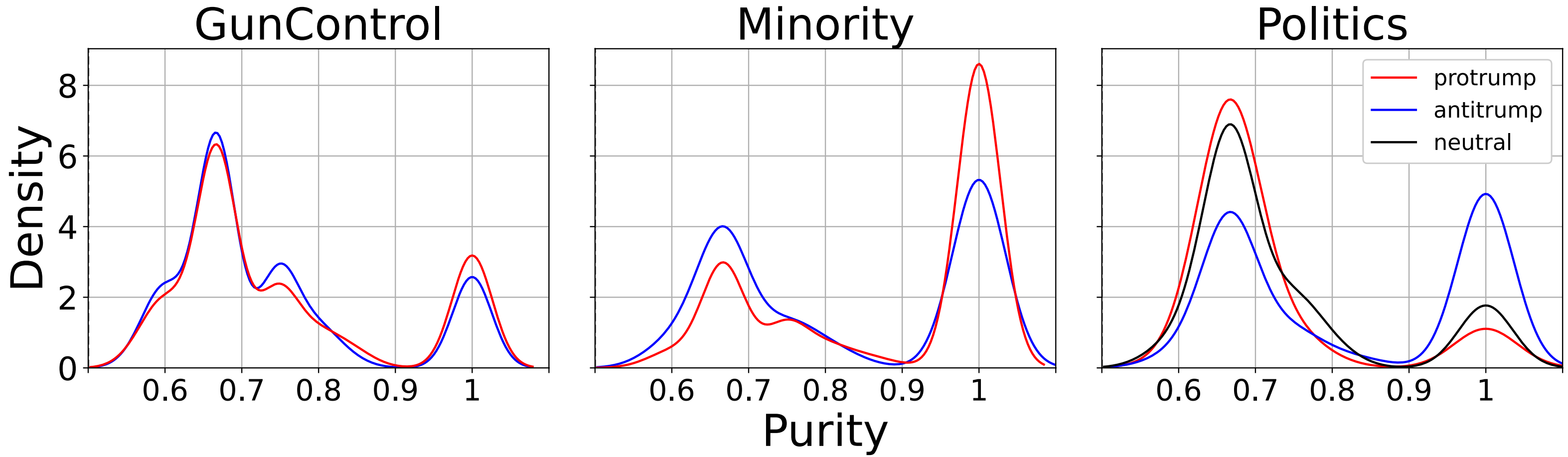
Temporal Dynamics

Higher-order interactions

Failla A, Citraro S, Rossetti G. **Attributed Stream Hypergraphs: temporal modeling of node-attributed high-order interactions.** Applied Network Science. 2023 Jun 9;8(1):31.

Hyperedge-level homophily: Purity

$$Purity(t, N, l) = \frac{\max_{l \in L}(\sum_{n \in N} l_{(t,n)})}{|(t, N)|},$$

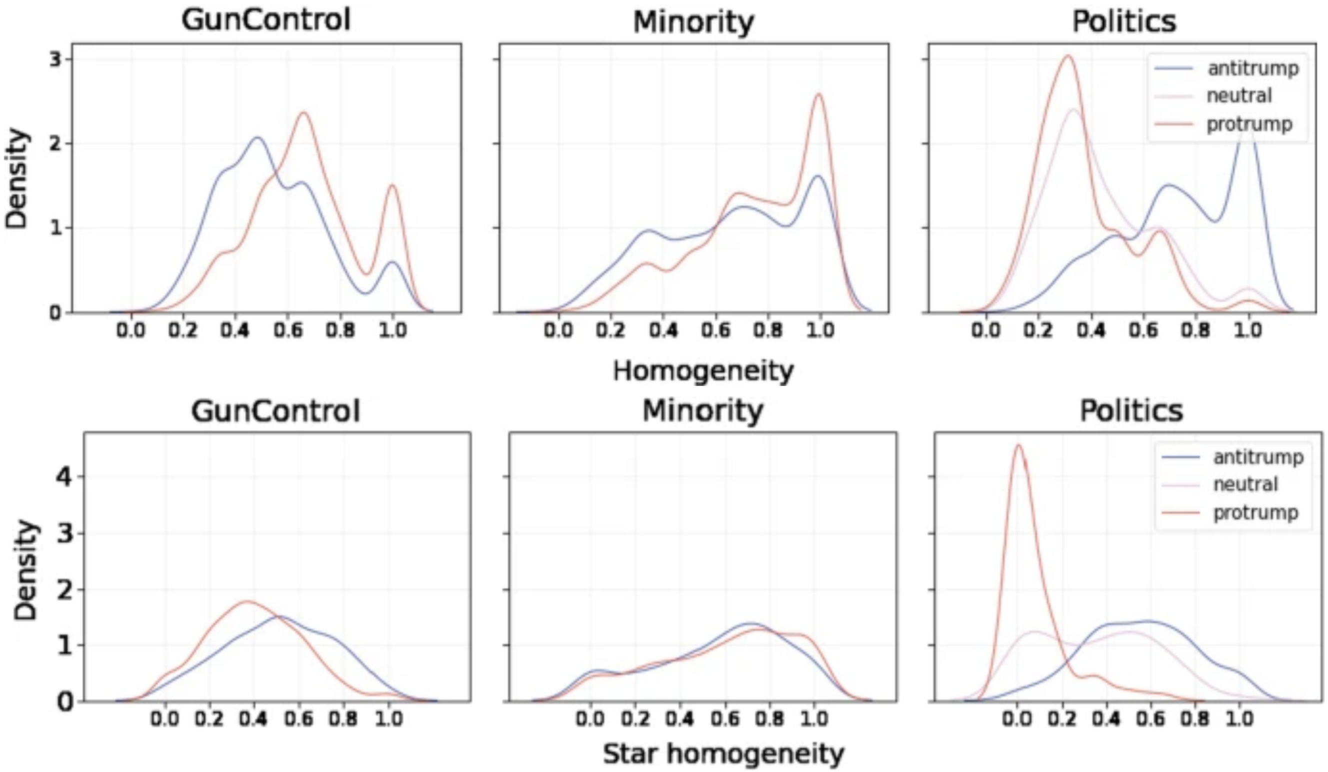


Failla A, Citraro S, Rossetti G. **Attributed Stream Hypergraphs: temporal modeling of node-attributed high-order interactions.** Applied Network Science. 2023 Jun 9;8(1):31.

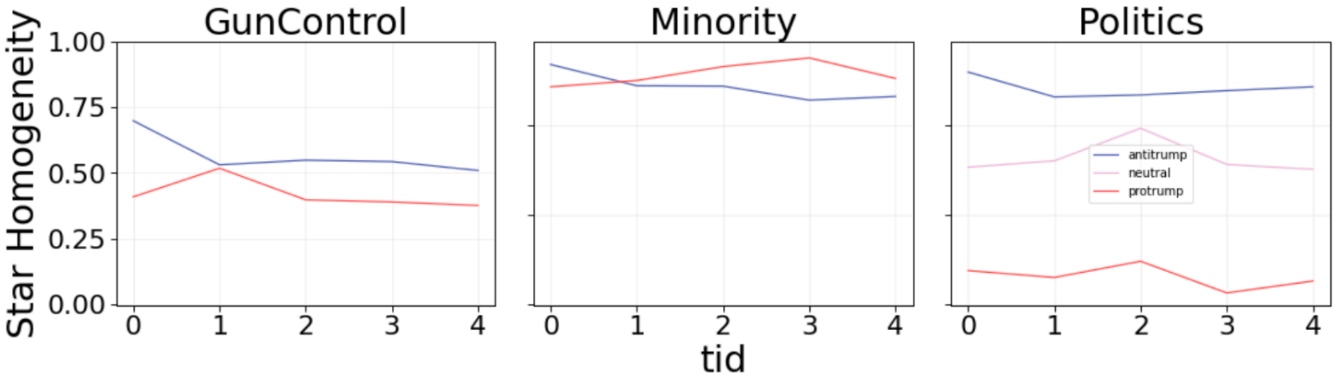
Pairwise and Higher-order Homophily on Reddit

Fraction of neighbors that share the same opinion as the ego

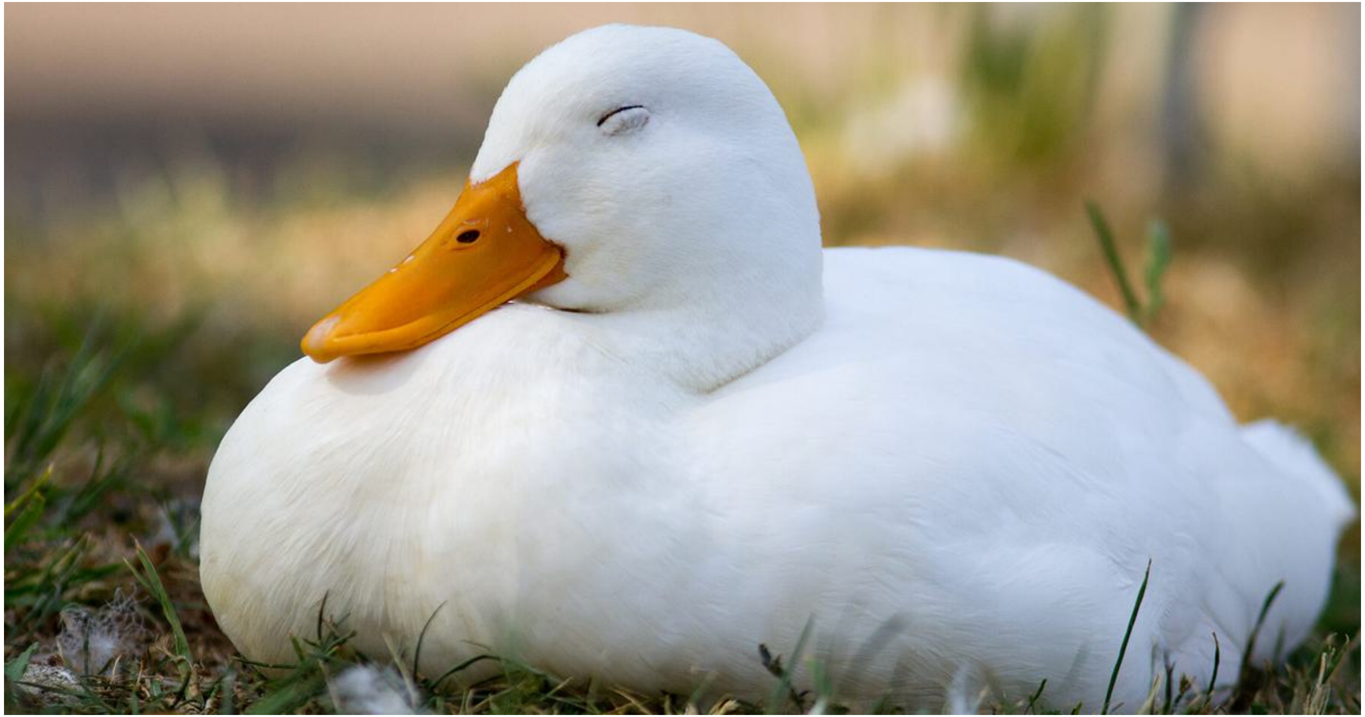
Fraction of discussions where ego is involved and where her opinion is predominant



Temporal Trends of Higher-order Homophily



Failla A, Citraro S, Rossetti G. **Attributed Stream Hypergraphs: temporal modeling of node-attributed high-order interactions.** Applied Network Science. 2023 Jun 9;8(1):31.

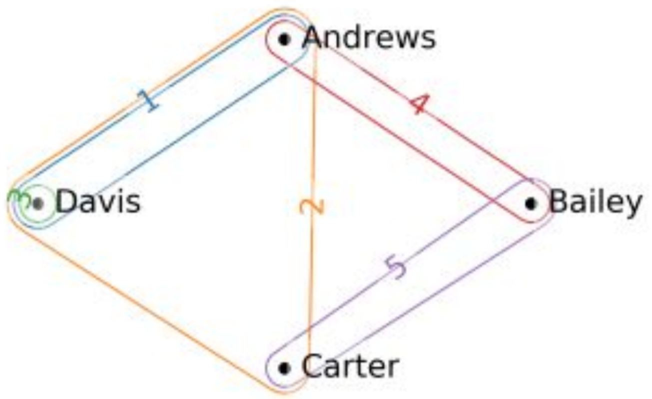


From Graphs to Hypergraphs

Pairwise to Higher-order I

S-Analysis Framework

Intuition: Leverage the hypergraph incidence matrix. Extend graph concepts and measures via hyperedge intersection size (the **s parameter**)



Example:
- 1 and 2 are 2-incident (they share {Davis, Andrews})
- 5 and 4 are 1-incident (they share {Bailey})

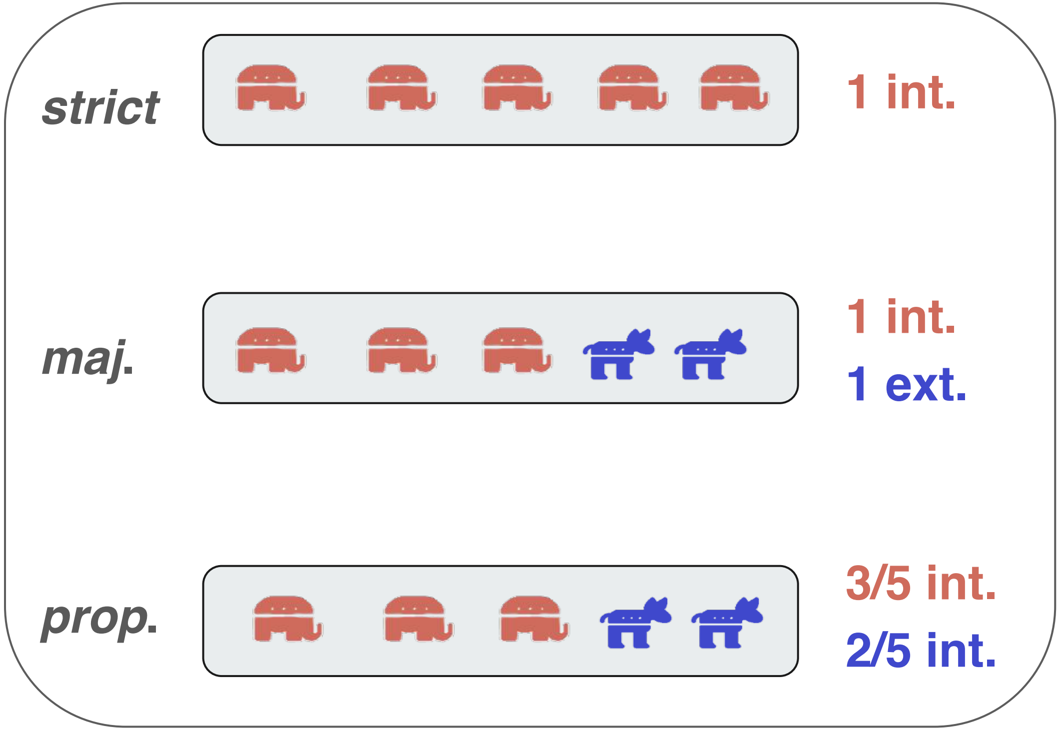
Aksoy SG, Joslyn C, Marrero CO, Praggastis B, Purvine E. **Hypernetwork science via high-order hypergraph walks.** EPJ Data Science. 2020 Dec 1;9(1):16.

Pairwise to Higher-order II

Sometimes extensions are non-trivial.

EI index: external/internal edge ratio in a node-attributed network

How to define **internal hyperedges**?

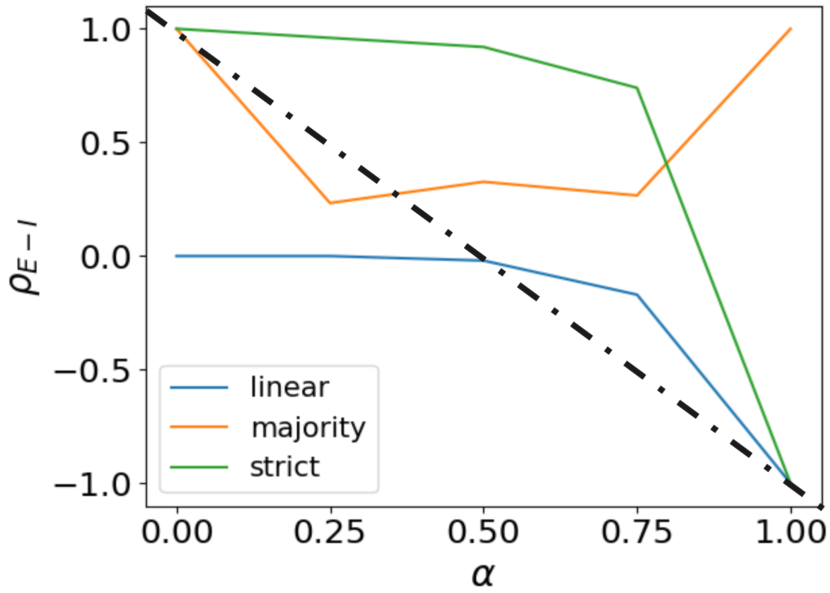


Failla A, Rossetti G., Cauteruccio F., **Beyond Boundaries: Capturing Social Segregation on Hypernetworks,**
Advances on Social Network Analysis and Mining (ASONAM 2024)

Pairwise to Higher-order III

Sometimes, it doesn't work as you'd think

Linear does not capture heterogeneity. *Majority* captures heterogeneity when the network is segregated



Alpha: **planted** tendency to be segregated.
The lower the EI, the more segregated.

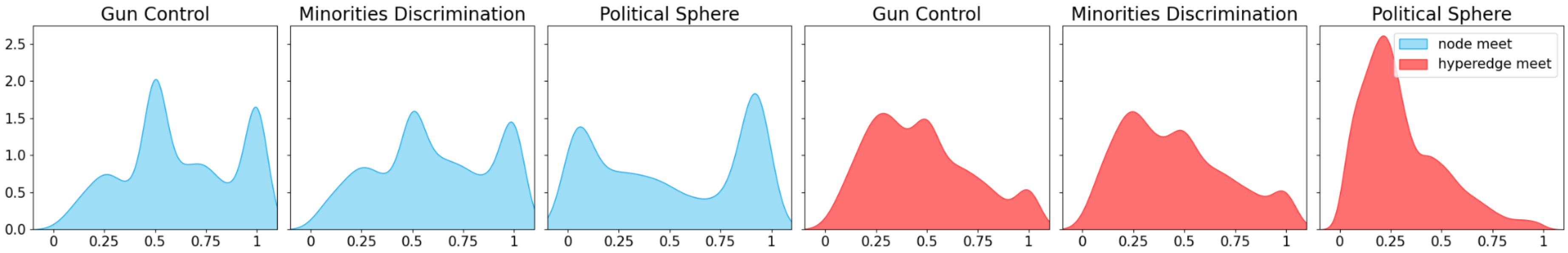
Failla A, Rossetti G., Cauteruccio F., **Beyond Boundaries: Capturing Social Segregation on Hypernetworks,**
Advances on Social Network Analysis and Mining (ASONAM 2024)

Hypernetwork Segregation via Random Walks I

Using random walks to estimate segregation

$$\phi_m^{t,k}(v_i) = \frac{1}{k}\sum_{r=1}^{k}\frac{|\{v_j \in \Omega^r : \gamma(v_j) = \gamma(v_i)\}|}{t}$$

Walk on nodes or edges



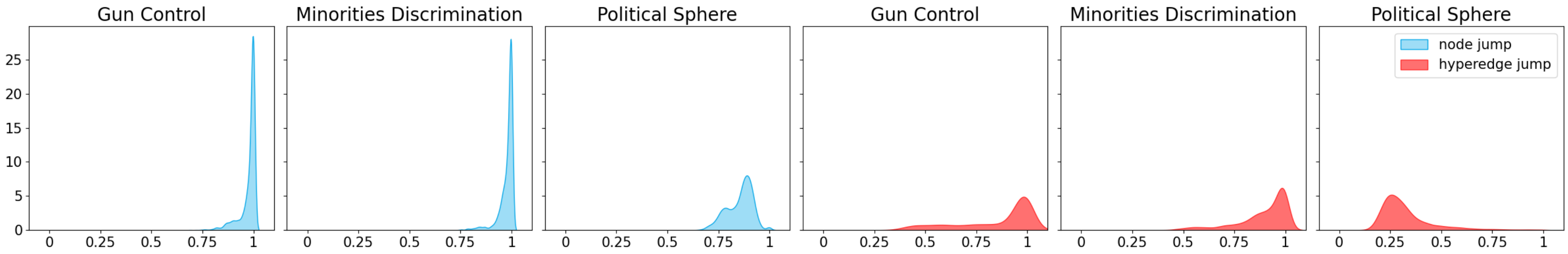
Failla A, Rossetti G., Cauteruccio F., **Beyond Boundaries: Capturing Social Segregation on Hypernetworks,**
Advances on Social Network Analysis and Mining (ASONAM 2024)

Hypernetwork Segregation via Random Walks II

Using random walks to estimate segregation

Walk on nodes or edges

$$\phi_j^{t,k}(v_i) = \frac{1}{k}\sum_{r=1}^{k}\frac{|\{(v_q \in \Omega^r, v_{q+1} \in \Omega^r) : \gamma(v_q) = \gamma(v_{q+1}))\}|}{t-1}$$



Failla A, Rossetti G., Cauteruccio F., **Beyond Boundaries: Capturing Social Segregation on Hypernetworks,** Advances on Social Network Analysis and Mining (ASONAM 2024)

Conclusions

Concluding Remarks

- Communities and Hypergraphs are **effective representations** to describe and analyze social groups

Concluding Remarks

- Communities and Hypergraphs are **effective representations** to describe and analyze social groups

- Working with them is not always straightforward, might require **special care**

Concluding Remarks

- Communities and Hypergraphs are **effective representations** to describe and analyze social groups

- Working with them is not always straightforward, might require **special care**

- Ducks are **cool**!

Thank you!
Any questions?

Università di Pisa
Consiglio Nazionale delle Ricerche

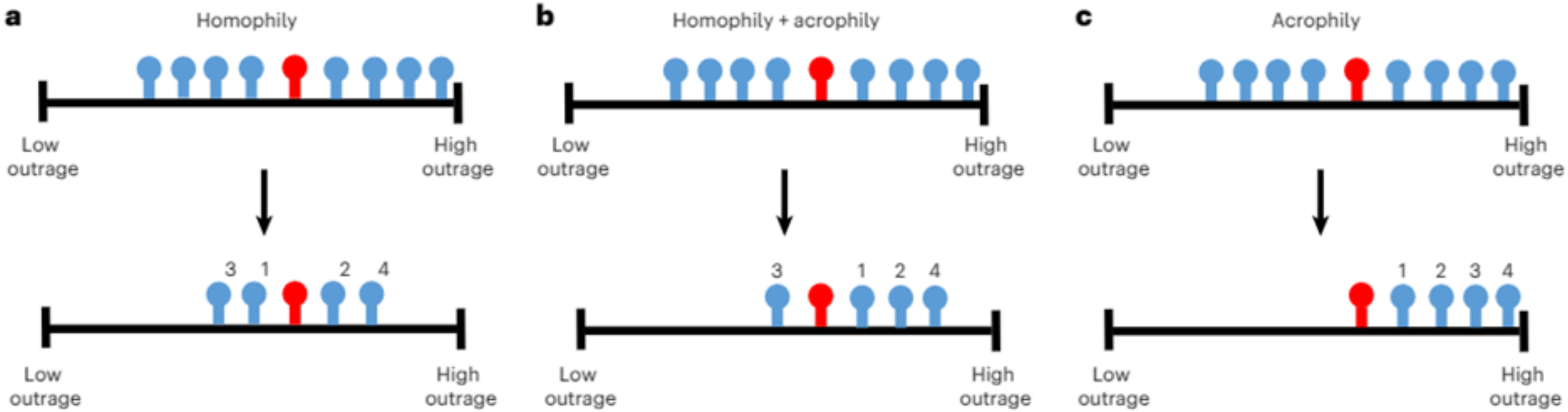
Post-Credit Scene

Attraction to Extreme Opinions

Università di Pisa

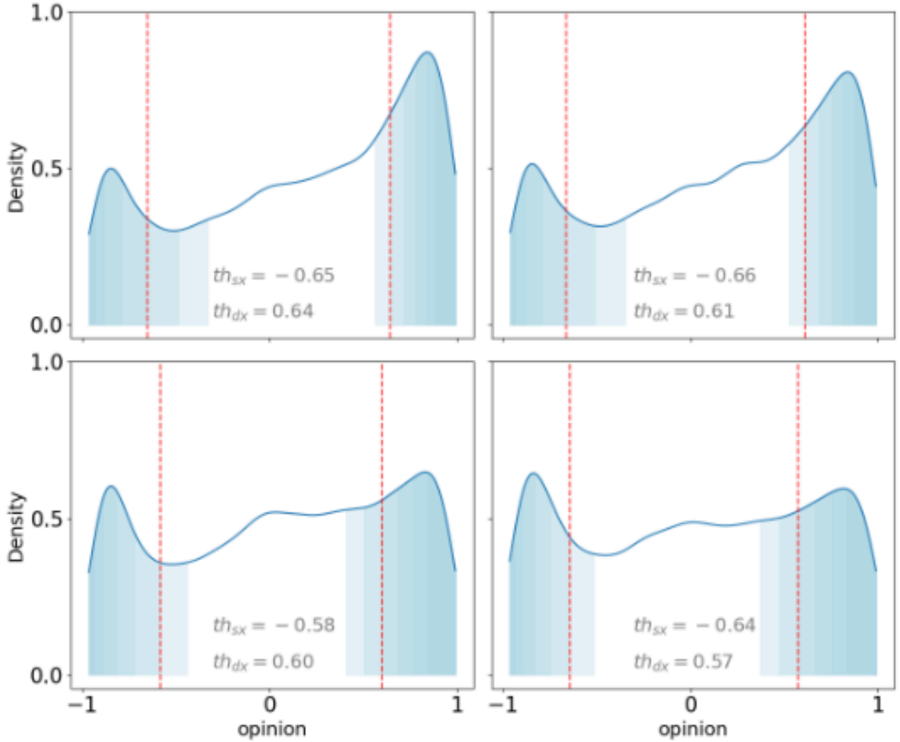
Consiglio Nazionale delle Ricerche

Acrophily in Online Discussions

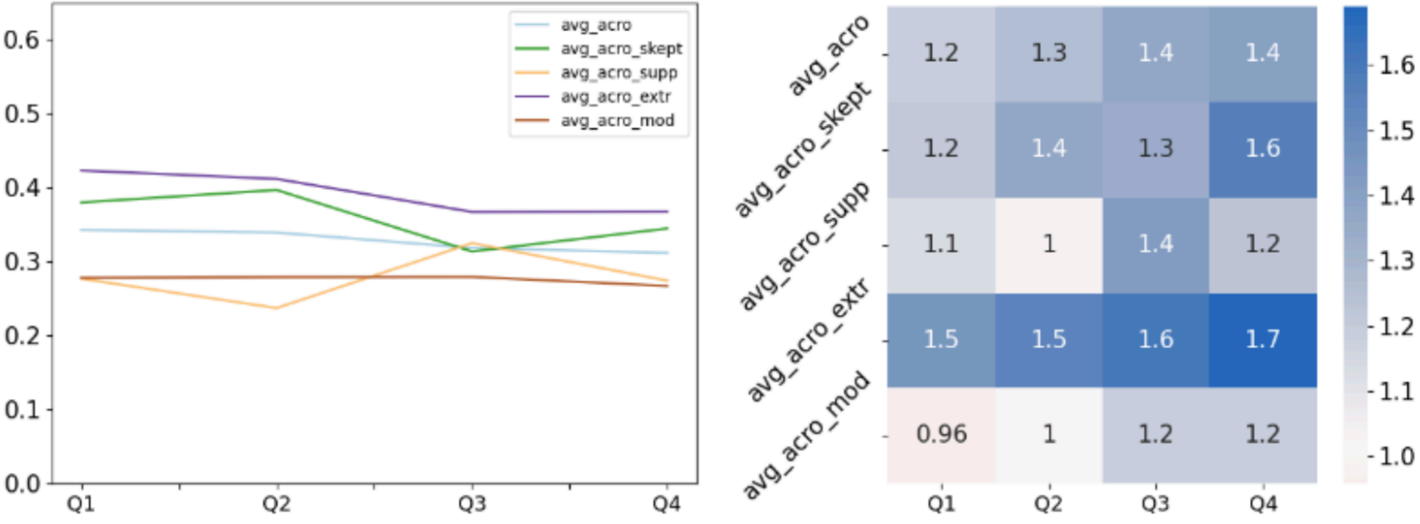


Opinion Distributions

Debates about climate change on Reddit, 2022



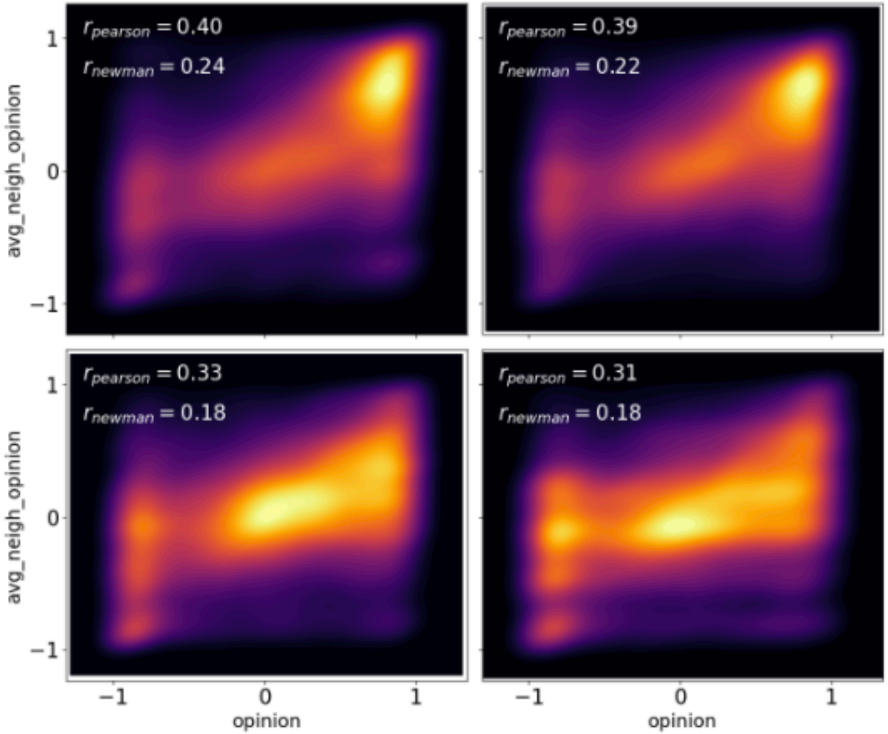
Acrophily over time



(a) (b)

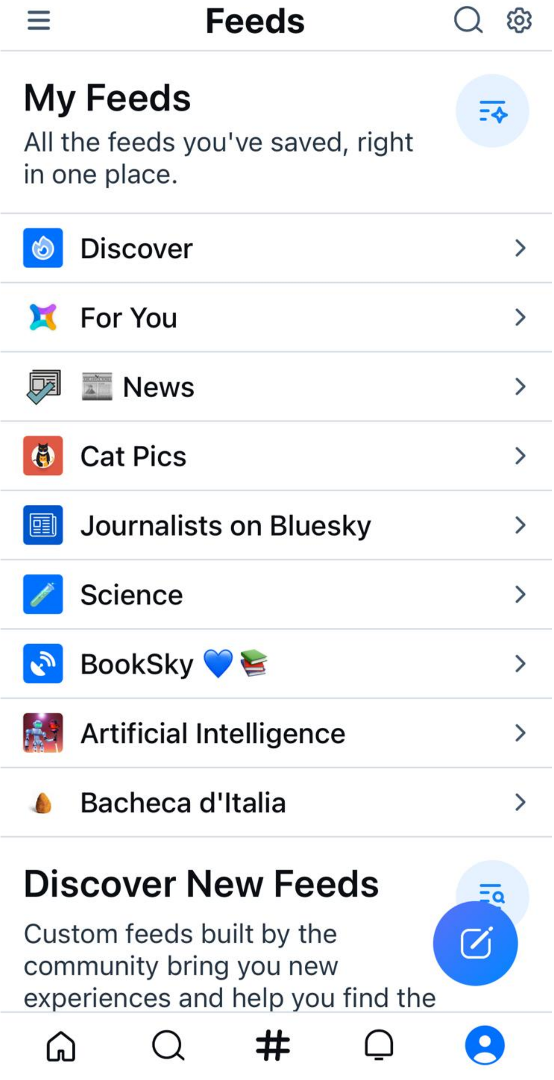
Polarization

Comparing user opinion with its neighbors' opinions



Hungry for Data?


Discover
For You
News
Cat Pics
Bodega Cats @bodegac... · 10h
12
167
1447
Ed Zitron @zitron.bsky.s... · 13h
I've got my best people on it


Feeds
My Feeds
All the feeds you've saved, right in one place.
Discover
For You
News
Cat Pics
Journalists on Bluesky
Science
BookSky
Artificial Intelligence
Bacheca d'Italia
Discover New Feeds
Custom feeds built by the community bring you new experiences and help you find the


Edit Profile
andreajpg.bsky.social
@andreajpg.bsky.social
2 followers 4 following 1 post
Posts
Replies
Media
Likes
Feeds
Reposted by andreajpg.bsky.social
Cats of Yore @catsofyor... · 20d
Porch guardian. Photo from my collection, ca. 1918 - 1922.

The Pipeline

Collect followers

Start at @bsky.app. Once 1M users are discovered, distribute subsequent requests over 10 machines.

Feb 25 – Mar 2

Collect users' timelines

Distribute requests over 10 machines

Mar 20 – Mar 21

Collect followings

A second pass to improve coverage. If new user, also collect their timeline.

Mar 21 – Mar 22

Collect feeds

Most popular feeds on various topics & social issues

Mar 23

Clean, Anonymize & Augment

Solve inconsistencies in post metadata, assign unique ids, sentiment analysis, etc.

Mar 23 – Mar 29

Analyze

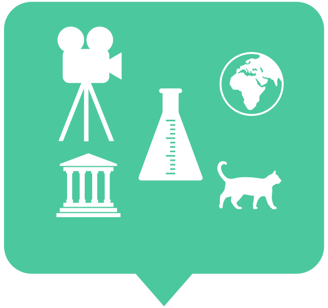
More on that in a few slides…

Apr

Results

Follow Network

Post Collection



Feed Collections

81%

Registered Users

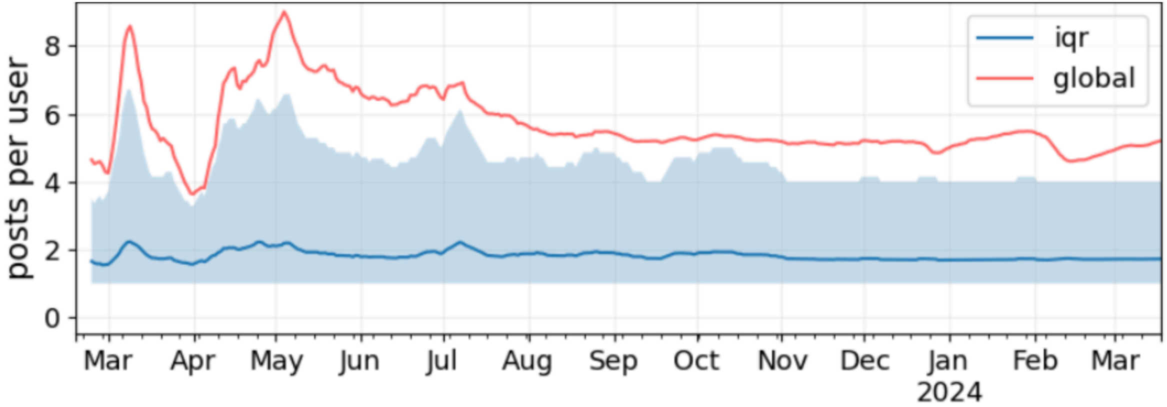
235M

Posts

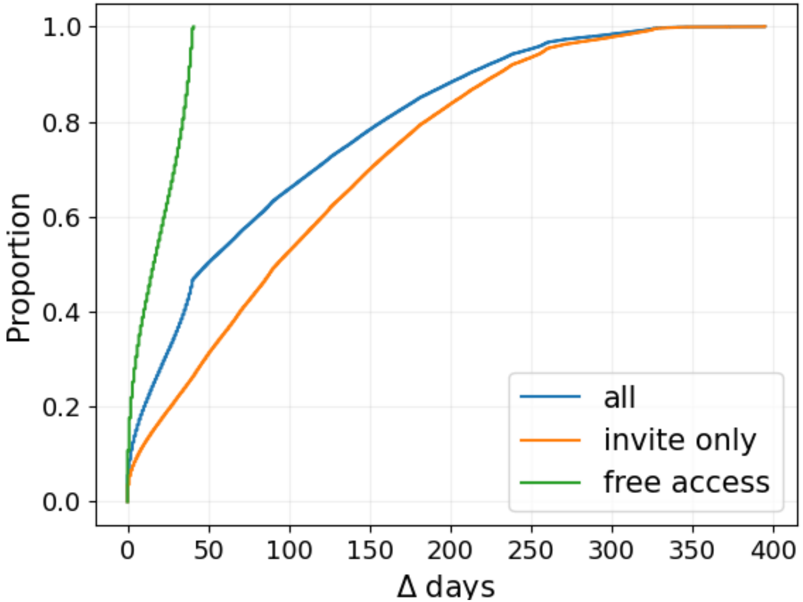
Category	Field	type	#non-null	description
User	user_id	int	235,567,116	an identifier univocally associated with each author/user.
	instance	str	235,567,116	the name of the instance that the user is registered to
Content	post_id	int	235,567,116	an identifier univocally associated with each post
	date	int	235,567,116	the post date and time formatted as YYYYmmddhhMM.
	text	str	235,567,116	the post's text content
	langs	list	220,628,598	the language(s) associated with each post, standardized to ISO 639-2
	labels	list	4,027,096	the content warning label(s) that the post is tagged with
	like_count	int	235,567,116	the number of likes as per the post metadata
	reply_count	int	235,567,116	the number of replies as per the post metadata
	repost_count	int	235,567,116	the combined number of reposts and quotes as per the post metadata
	sent_label	int	128,664,788	the text's sentiment
	sent_score	float	128,664,788	the sentiment model's confidence
Relational	reply_to	int	87,704,964	the ID of the post to which the current post replies to
	replied_author	int	87,704,964	the ID of the replied post's author
	thread_root	int	87,704,964	the ID of the post that initiated the discussion thread
	thread_root_author	int	87,704,964	the ID of the root post's author
	repost_from	int	63,549,643	the ID of the reposted post
	reposted_author	int	63,549,643	the ID of the reposted post's author
	quotes	int	12,110,474	the ID of the quoted post
	quoted_author	int	12,110,474	the ID of the quoted post's author

Table 1

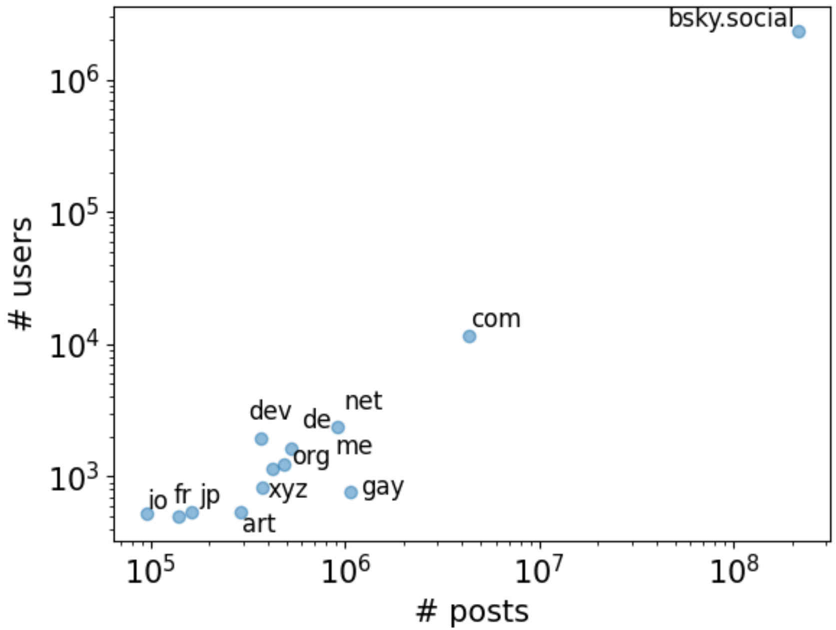
Posting Activity I



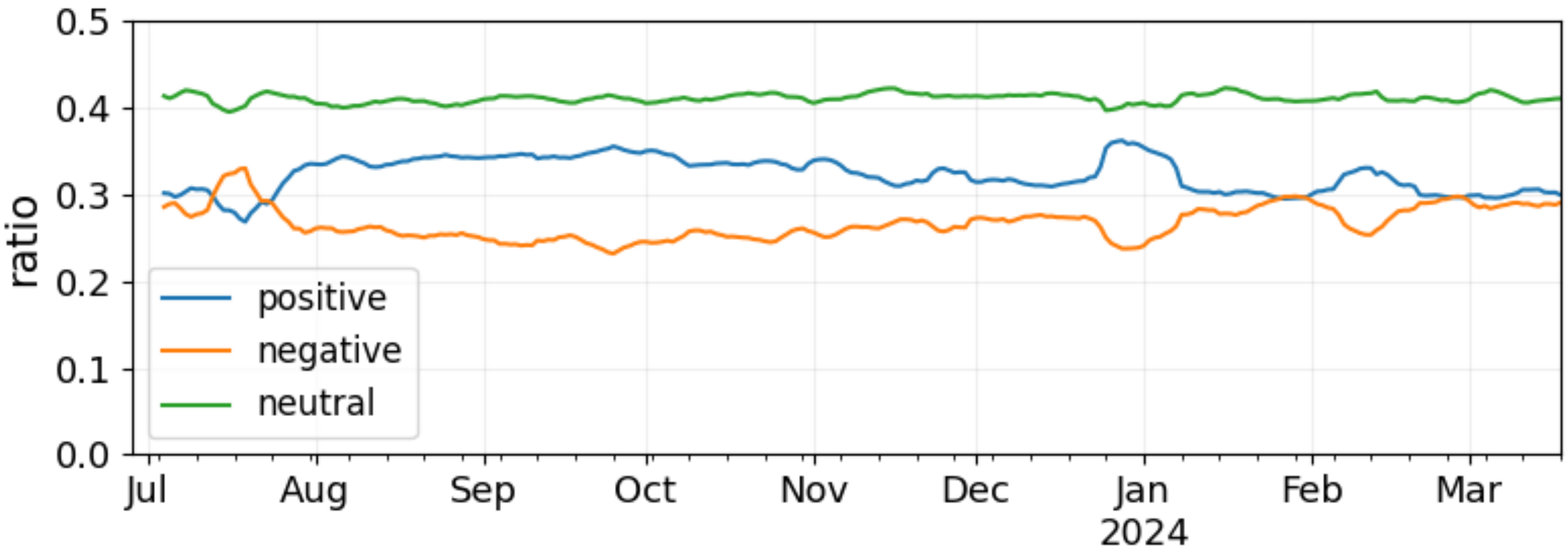
Posting Activity II



Instances


bsky.social
com
net
dev
de
me
org
xyz
gay
io
fr
jp
art
10^6
10^5
10^4
10^3
users
10^5
10^6
10^7
10^8
posts

Sentiment


0.5
0.4
0.3
0.2
0.1
0.0
ratio
positive
negative
neutral
Jul
Aug
Sep
Oct
Nov
Dec
Jan
2024
Feb
Mar

Thanks!