

Generalizing User Networks Across Platforms Through Shared Discourse

Narratives don't respect platform boundaries — our models shouldn't either.

Patrick Gerard — USC/ISI



patrickgerard_



patrikgerard.bsky.social

Who Am I?

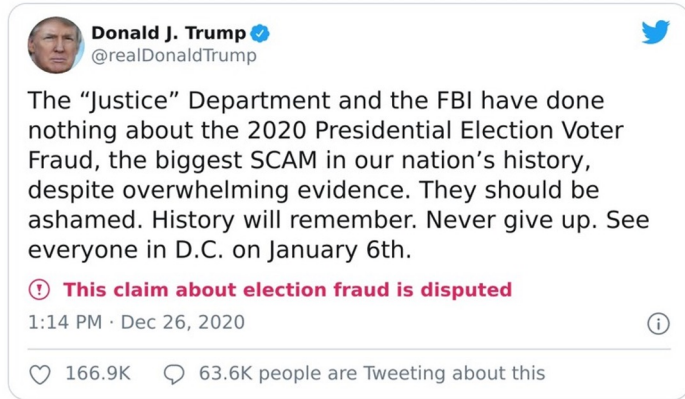
Kevin (handsome)

Me

I'm Patrick Gerard, a PhD student working under Prs. **Kristina Lerman** and **Emilio Ferrara**. I'm interested in the intersection of **machine learning**, **natural language processing**, and **network science** and how they can be utilized to uncover the mechanisms of **misinformation diffusion** and narrative evolution across media.

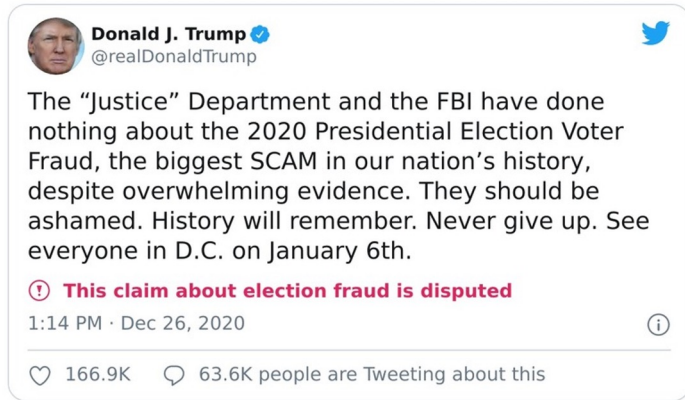


January 6th: Same Narratives, Different Ecosystems



(Donovan et al.)

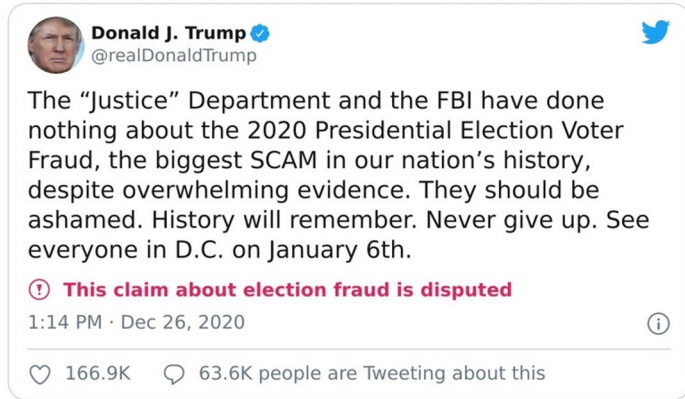
January 6th: Same Narratives, Different Ecosystems



President Trump is calling us to FIGHT...He knows this is the only way to save our great country, show up @January6th

(Donovan et al.)

January 6th: Same Narratives, Different Ecosystems



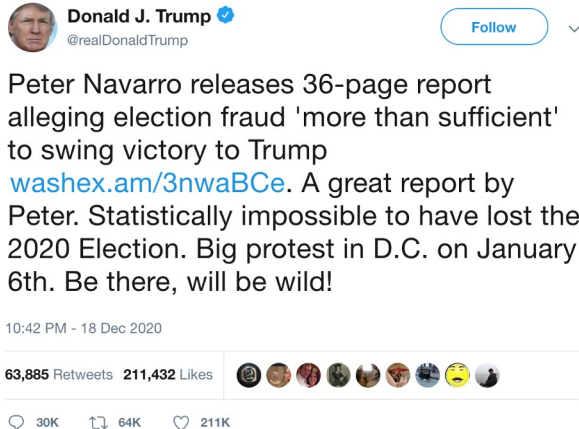
President Trump is calling us to FIGHT...He knows this is the only way to save our great country, show up @January6th



This is the third time he's tweeted about it. This isn't a joke, this is where and when we make our stand. #January6th. #LIVEFREEORDIE #FIGHTFORTRUMP

(Donovan et al.)

January 6th: Same Narratives, Different Ecosystems



(Donovan et al.)

January 6th: Same Narratives, Different Ecosystems



Follow

Peter Navarro releases 36-page report alleging election fraud 'more than sufficient' to swing victory to Trump washex.am/3nwaBCe. A great report by Peter. Statistically impossible to have lost the 2020 Election. Big protest in D.C. on January 6th. Be there, will be wild!

10:42 PM - 18 Dec 2020

63,885 Retweets 211,432 Likes



30K 64K 211K



Trump said It's gonna be wild!!!!!! It's gonna be wild!!!!!! He wants us to make it WILD that's what he's saying. He called us all to the Capitol and wants us to make it wild!!! Sir Yes Sir!!! Gentlemen we are heading to DC pack your shit!!

(Donovan et al.)

January 6th: Same Narratives, Different Ecosystems



Follow

Peter Navarro releases 36-page report alleging election fraud 'more than sufficient' to swing victory to Trump washex.am/3nwaBCe. A great report by Peter. Statistically impossible to have lost the 2020 Election. Big protest in D.C. on January 6th. Be there, will be wild!

10:42 PM - 18 Dec 2020

63,885 Retweets 211,432 Likes



30K 64K 211K



Trump said It's gonna be wild!!!!!! It's gonna be wild!!!!!! He wants us to make it WILD that's what he's saying. He called us all to the Capitol and wants us to make it wild!!! Sir Yes Sir!!! Gentlemen we are heading to DC pack your shit!!



It's important that millions of Americans show up in DC on January 6th to support the legitimate President, Donald Trump, and show Democrats what they will be facing if they continue to try to steal the Presidency.! Gentlemen we are heading to DC pack your shit!!

(Donovan et al.)

January 6th: Same Narratives, Different Ecosystems



Source: BBC



President Trump has been right on just about everything he said. So when he says the election was rigged I'll bet my life on him being right.

(Donovan et al.)

January 6th: Same Narratives, Different Ecosystems



Source: BBC



President Trump has been right on just about everything he said. So when he says the election was rigged I'll bet my life on him being right.



We have witnessed the destruction of the very fiber of our freedom, the election process.

(Donovan et al.)

January 6th: Same Narratives, Different Ecosystems

January 6th was fundamentally a **cross-platform event**—coordination happened across platforms, not just within them.

Ng et al. 2022 showed that users on Parler and Twitter consumed **completely different information sources**.

Yet, **narrative themes converged** across platforms.



Source: NPR

Springfield: Same Narratives, Different Ecosystems

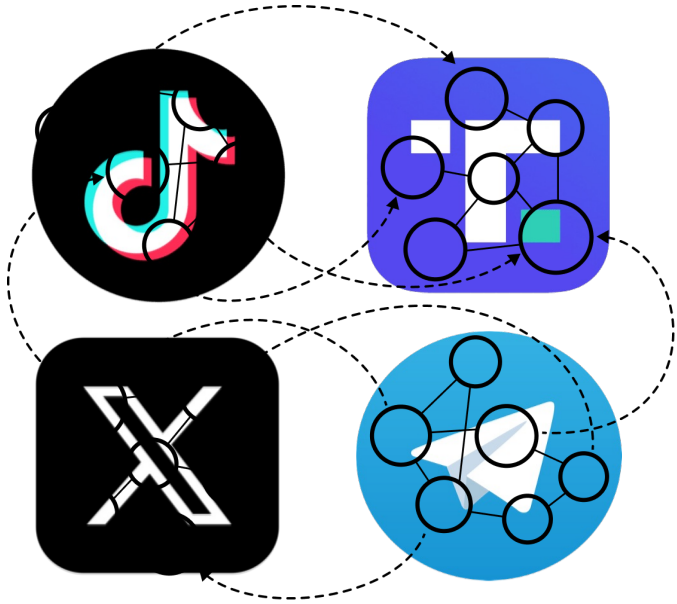
When a small local rumor emerged, **fragmented communities told different versions:**

- **Facebook groups** shared park and duck anecdotes
- **Telegram and Gab** reframed it into racist conspiracies
- **X influencers** amplified screenshots and memes to millions



Despite these differences, **a unified narrative form emerged across platforms**, eventually reaching the U.S. Presidential debate.

Modern Narratives Take Shape Across Platforms

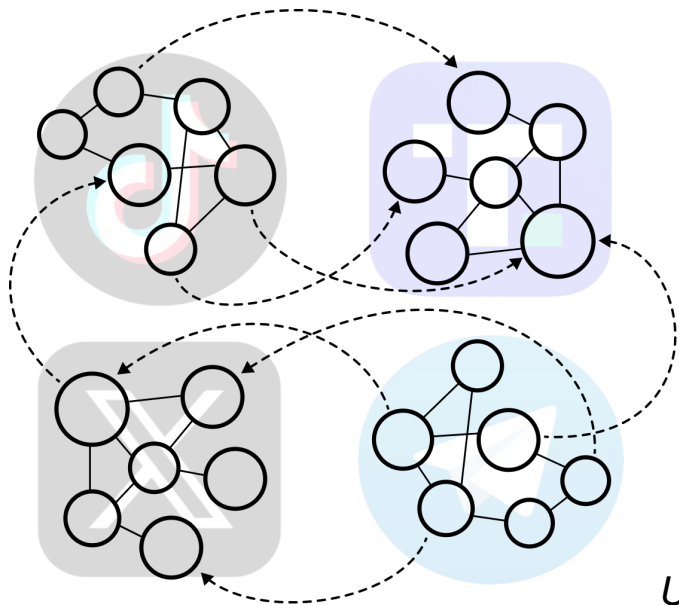


The online information ecosystem is fragmented.

- Users consume and share content across many platforms.
- The **borders between platforms are arbitrary** — **ideas and narratives do not respect them.**
- Without cross-platform visibility, we fail to anticipate real-world impacts.

If we treat platforms as islands, we will always be surprised when ideas jump between them

Why Traditional Methods Fail

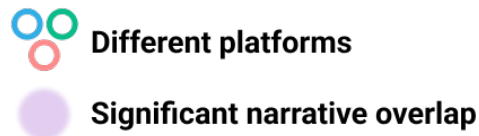
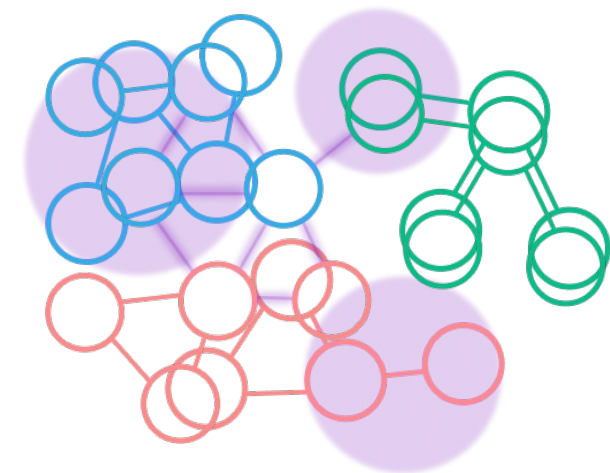


Traditional networks break at the platform boundary.
Ideas don't.

- Follower graphs, reposts, hashtags → **platform-locked**
- Semantic similarity → brittle under platform-specific language
- APIs → disappearing, incomplete, inconsistent

Understanding modern information flow requires understanding how users connect across platforms, not just within them.

Rethinking User Representation Through Shared Discourse



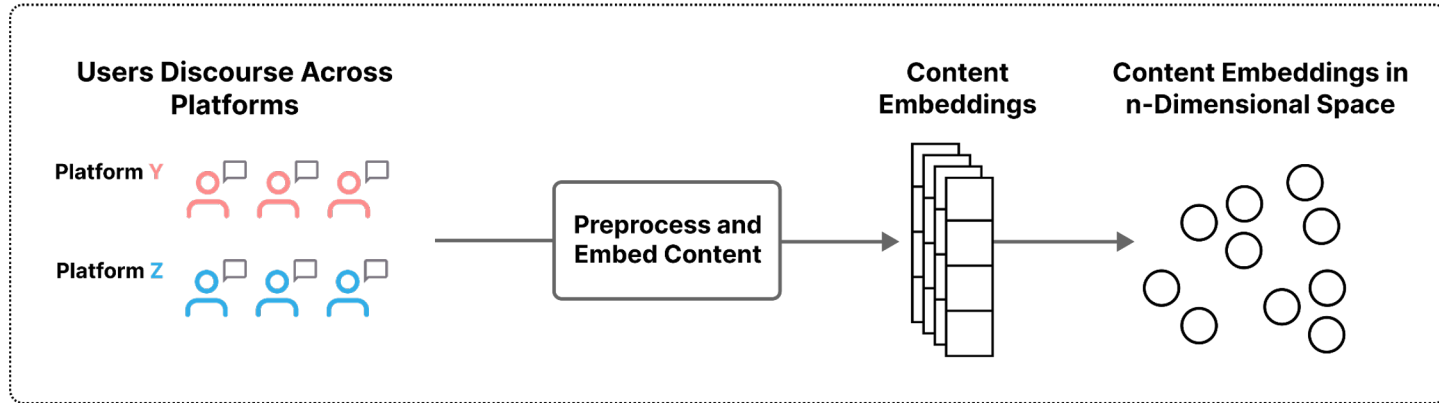
Represent users through shared narrative participation, not platform behaviors.

- Each user = **distribution over latent narratives**
- Users connect via **shared discourse**, not interactions
- Works across platforms simultaneously

Content → Embeddings → Narratives → User Distribution → Unified Network

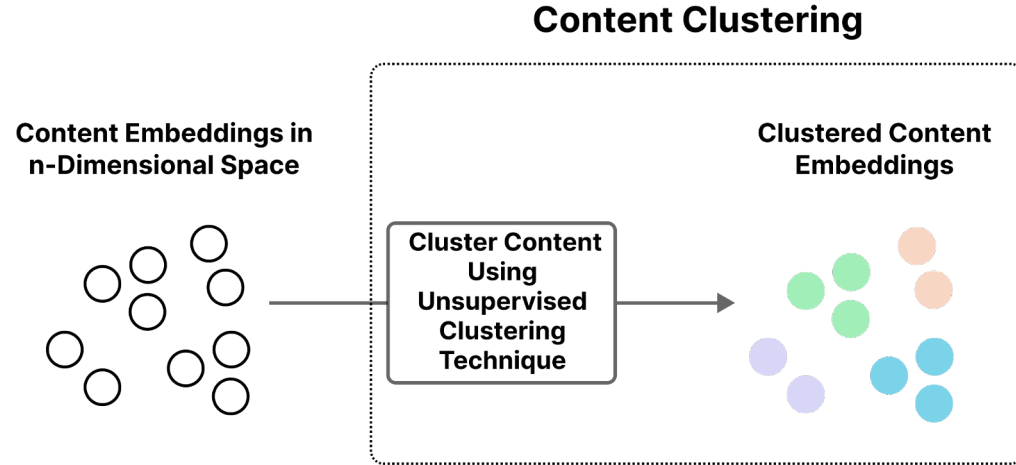
Our Approach: Modeling Discourse as Network

Content Aggregation and Processing



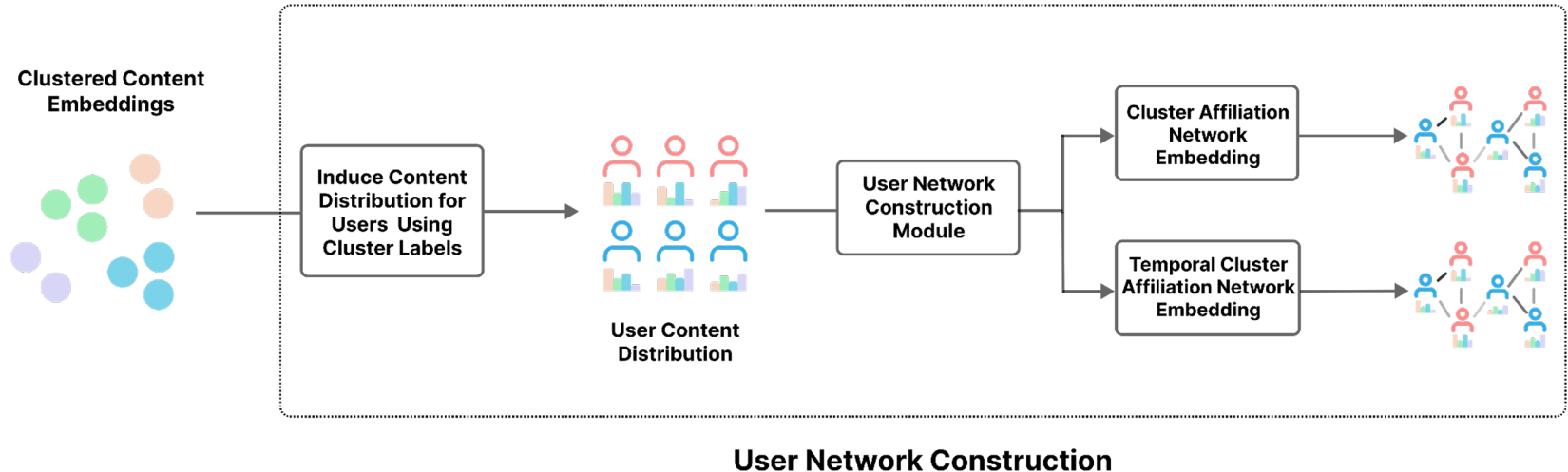
Step 1: Content Aggregation and Processing *Collect cross-platform discourse and transform text into comparable embeddings*

Our Approach: Modeling Discourse as Network



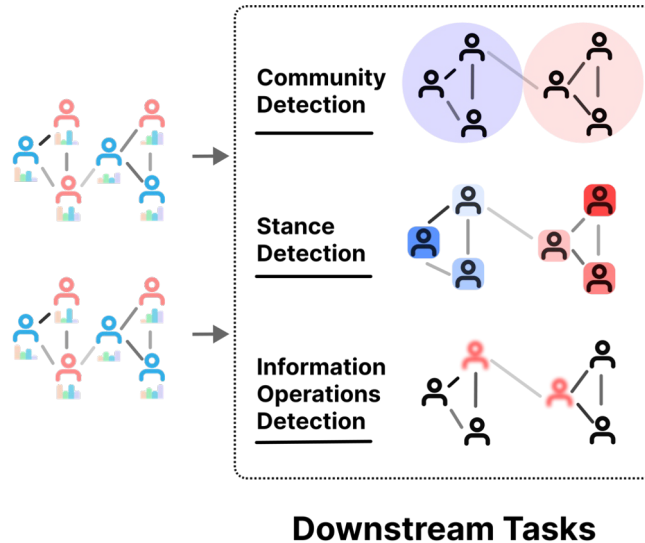
Step 2: Content Clustering *Group similar discourse into thematic clusters using unsupervised learning*

Our Approach: Modeling Discourse as Network



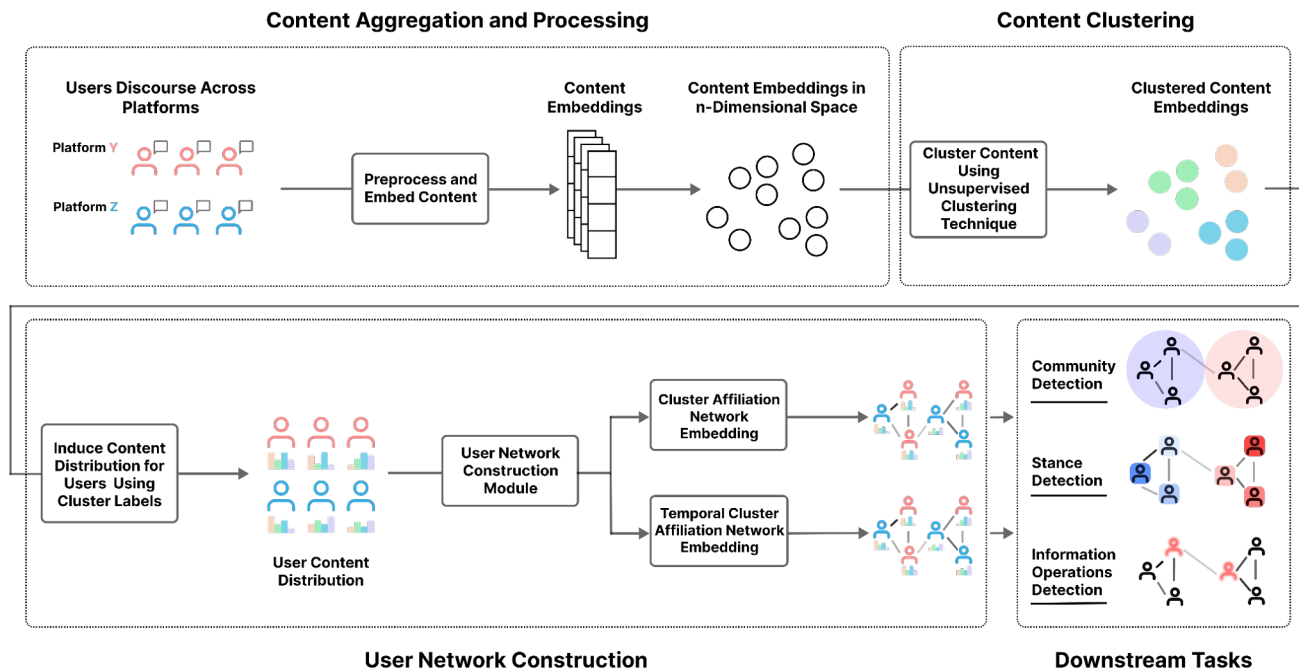
Step 3: User Network Construction *Transform content clusters into user networks based on shared discourse patterns*

Our Approach: Modeling Discourse as Network



Step 4: Network Applications *Use the constructed networks for traditional network analysis and detection tasks*

Our Approach: Modeling Discourse as Network



Full Pipeline: From Cross-Platform Discourse to User Networks

Building behavioral networks from shared discourse patterns across platform boundaries

Our Approach: Modeling Discourse as Network

Modern discourse spans platforms, modalities, and linguistic norms.

- Each platform produces its own linguistic fingerprints
- Even when discussing the same event, posts diverge in form, phrasing, and structure



Our Approach: Modeling Discourse as Network

Two example narratives appearing in cross-platform discourse during the 2024 Election:

FEMA Interference During Hurricane Helene

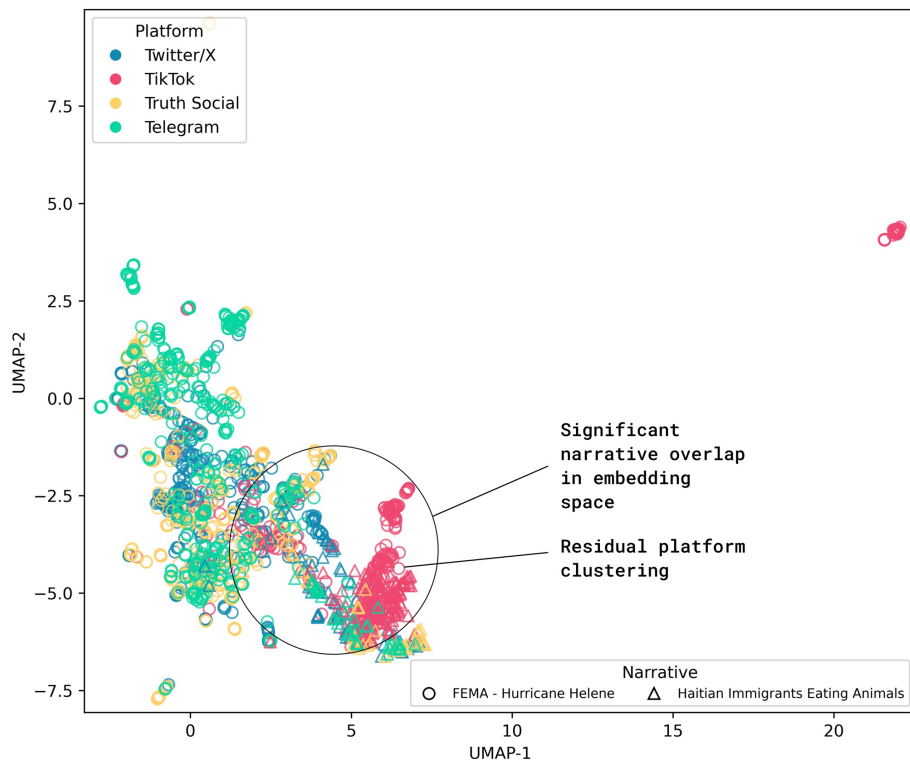
Posts alleging FEMA prevented volunteer rescue operations, restricted access, or intentionally mismanaged relief efforts.

Haitian Immigrants 'Eating Pets' in Springfield

Claims that Haitian migrants were killing or consuming neighborhood animals, fueling local panic and national amplification



UMAP of Original Posts



Haitian Immigrants:

Twitter/X

"People in Springfield say Haitian migrants are killing neighborhood pets. Local shelters warning families to keep animals inside."

TikTok (video transcript)

"Residents are claiming their cats went missing after new arrivals moved in. I'm not saying it's true but something is going on..."

Hurricane Helene:

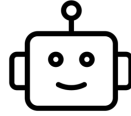
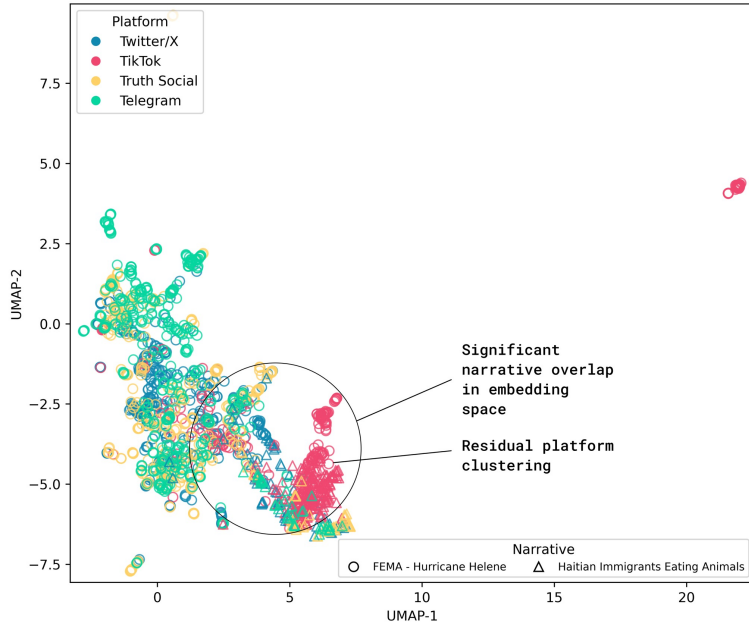
Twitter/X

"Hearing reports FEMA is blocking volunteer rescue boats in NC. Why are they stopping people from helping? #HurricaneHelene"

TikTok (video transcript)

"So apparently FEMA told these volunteers they *can't* go in and help. Like... how does that make sense when people are stranded?"

UMAP of Original Posts



Normalized Claim
Extraction



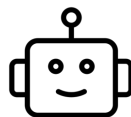
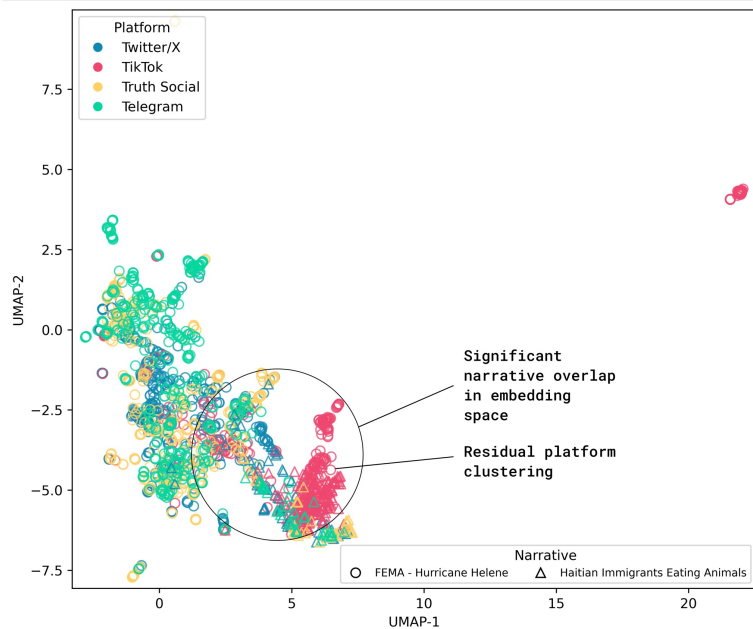
Hurricane Helene Narrative — Core Claims

- FEMA blocked volunteer rescue efforts.
- FEMA told volunteers they could not help victims.
- Federal authorities interfered with private relief.
- FEMA restricted access to disaster zones.

Haitian Immigrant Narrative — Core Claims

- Haitian immigrants killed neighborhood pets.
- Pets disappeared after immigrants arrived.
- Immigrants slaughtered animals for food.
- Media suppressed reports of the incidents.

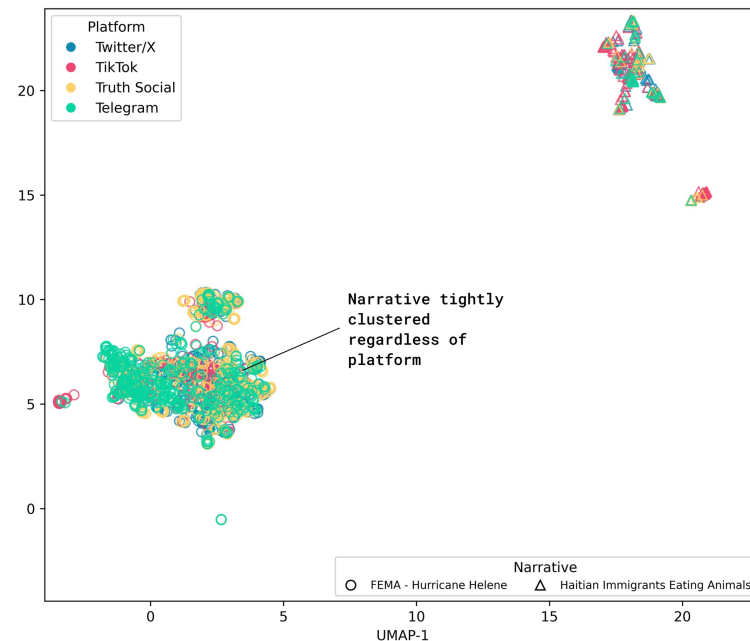
UMAP of Original Posts



Normalized Claim
Extraction

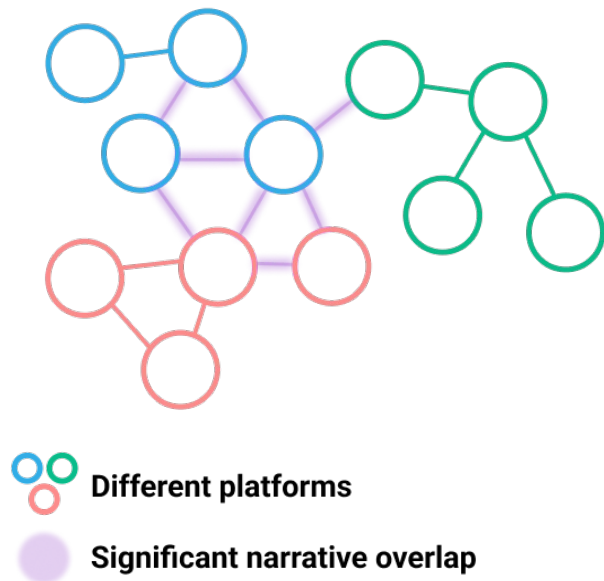


UMAP After Normalized Claim Extraction



Within Platforms: Discourse Networks

Perform Better, Need Less Data, and Cover More Users



Discourse Networks:

Match or beat behavioral networks

On classic CSS tasks like stance or IO detection.

Use a fraction of the data

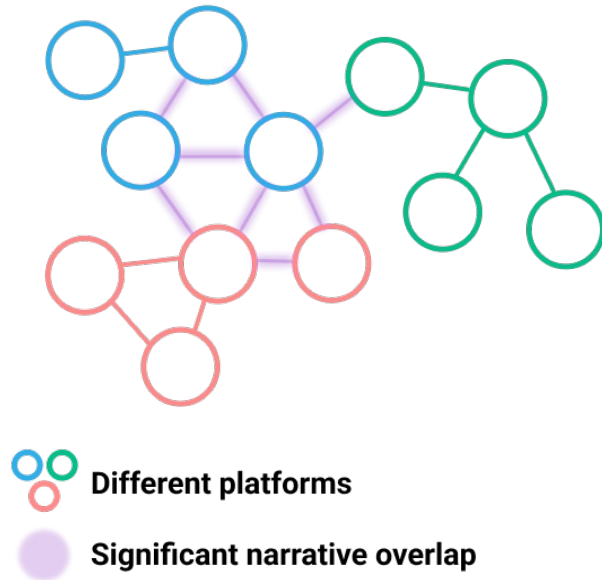
Behavioral networks need 3-5x more to get close.

Cover nearly all users

Significantly more user coverage than traditional methods.

Across Platforms: Discourse Networks

Extend Discourse into a Unified Ecosystem



Discourse Networks:

Extend social proximity beyond platform boundaries

Provide a foundation for predicting cross-platform information spread.

Capture cross-platform “bridge zones”

Invisible to traditional traditional methods.

Integrate platforms into a single actionable view

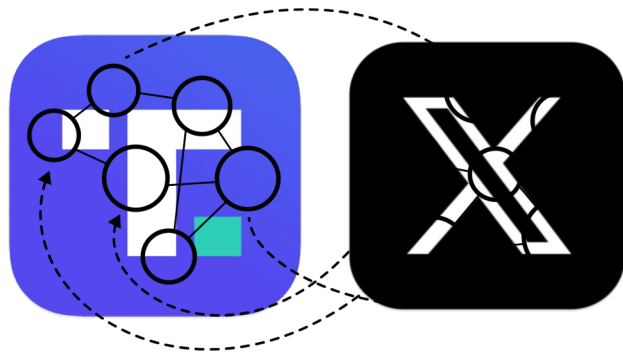
Allow operators to monitor narrative movement across the information ecosystem.

Case Study 1: A Tiny Bridge Zone Spans Both Worlds

Do shared narrative communities emerge despite different ecosystems?

Studying discourse on X and Truth Social leading up to the 2024 U.S. Presidential Election:

- **0.33% of users** form one mixed-platform community associated with **>68% of cross-platform narrative flow**
- Only visible through **discourse modeling**, not traditional methods



There's a hidden bridge between these platforms, but you need the right lens to see it.

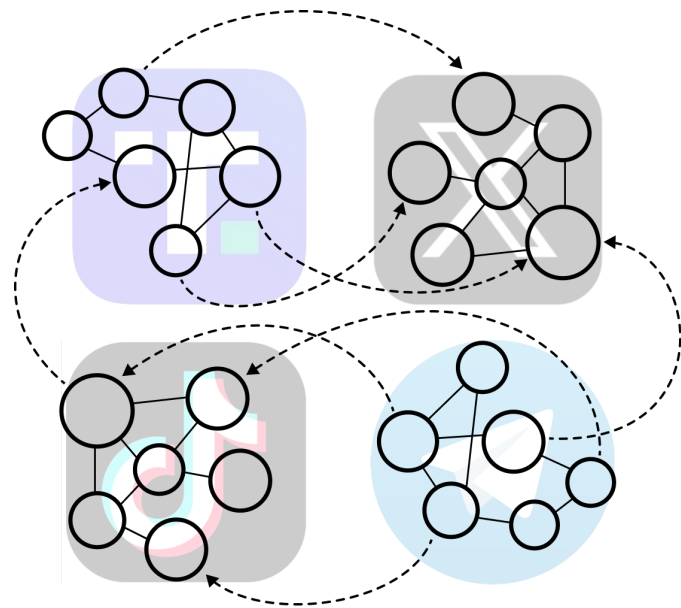
Case Study 2: Cross-Platform Information Diffusion

How can we rethink cross-platform information diffusion tasks?

Representing users by their shared narratives reveals a **predictive social proximity signal** that spans platforms

This structure lets us recast multi-platform information diffusion as a **predictive social proximity task**, rather than a causal cascade

The result: **performance doubles** Hawkes processes, independent cascades, and other classical diffusion models—using just a **2-feature RF model**.

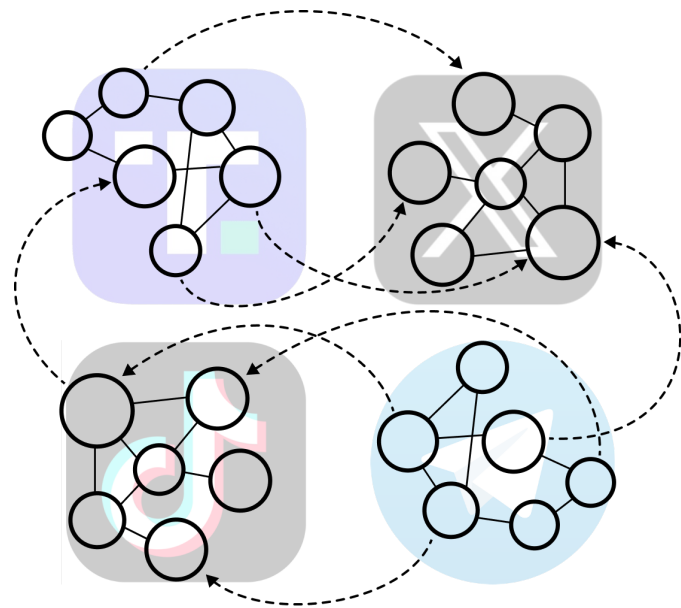


Case Study 3: Visualizing Cross-Platform Narratives

How can we rethink cross-platform information diffusion tasks?

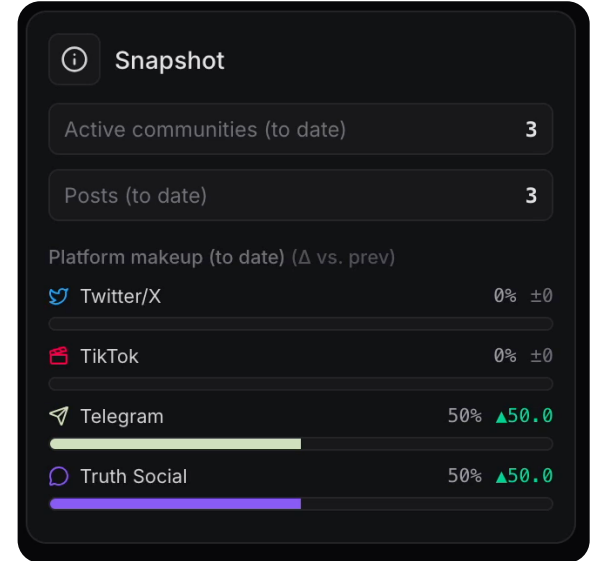
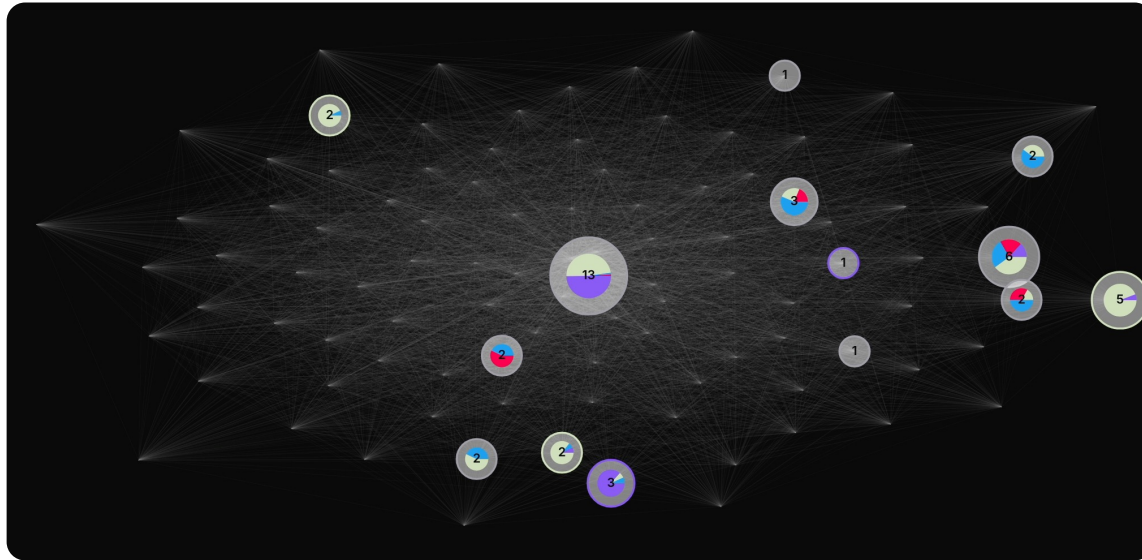
To make cross-platform dynamics interpretable to journalists and investigators, we collaborated with **Der Spiegel** to develop interactive visualizations of narratives moving through the online ecosystem.

Unified discourse embeddings allow us to watch a narrative form on one platform, surge on another, and fracture into competing storylines — all in one space.



Future Directions

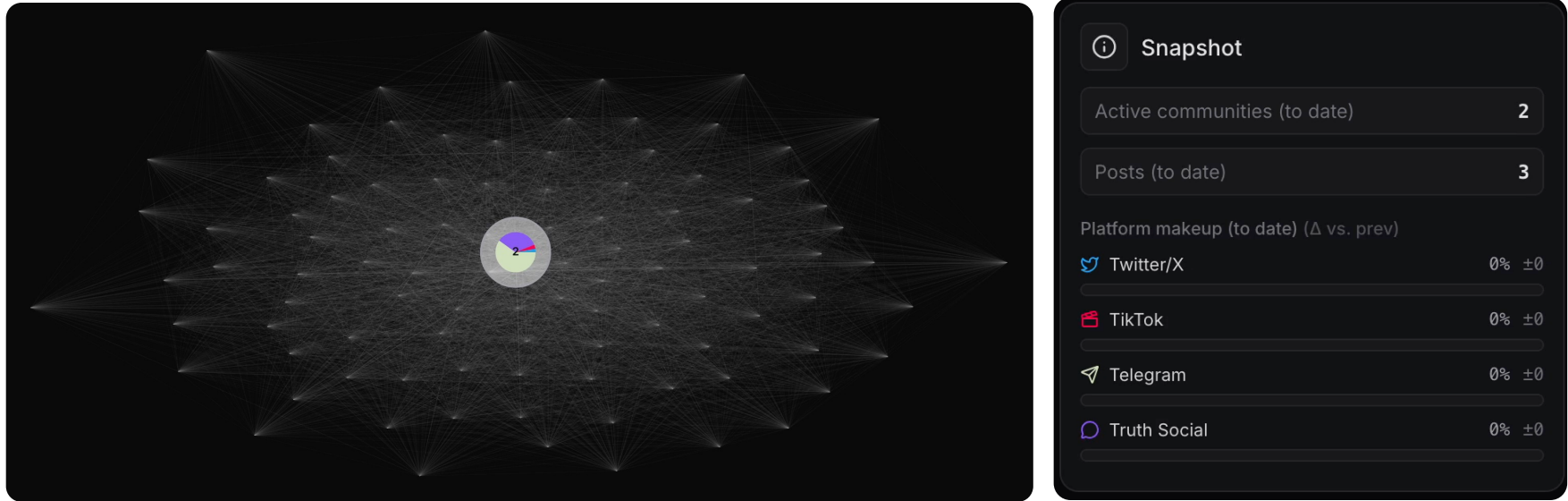
Narrative activation mapped across the online ecosystem



Narrative: Fauci and NIH involved with releasing COVID-19 from lab.

Future Directions

Narrative activation mapped across the online ecosystem



U.S. Energy independence debate – push for drilling vs. green transition.

Takeaways: Birds of a Feather *Think* Together

Discourse-based graphs:

Outperform traditional ones

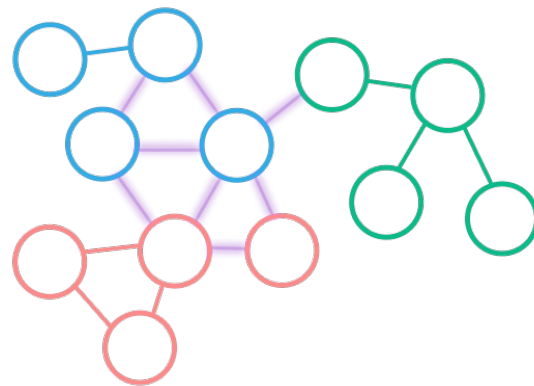
Discourse-based networks capture **more users** with **less data**, while matching or exceeding traditional methods.

Cross-platform by design

The approach **works across fragmented ecosystems**, without needing reposts, URLs, or user overlap.

Uncover the hidden architecture of the online information ecosystem

Discourse networks surface alignment and migration patterns invisible to traditional graphs.



pgerard@isi.edu



patrickgerard.co



patrickgerard_



patrikgerard.bsky.social

Takeaways: Birds of a Feather *Think* Together



Case Study 1



Case Study 2



Case Study 3

Appendix

Time Complexity

Method	Embedding	Clustering	Graph Construction
<i>Our Methods</i>			
CANE (FAISS)	$O(Nd)$	$O(NKT)$	$O(U \log U + Uk)$
t-CANE (FAISS)	$O(Nd)$	$O(NKT)$	$O(T \cdot (U \log U + Uk))$
<i>Baselines</i>			
Embedding Averaging	$O(Nd)$	—	$O(U^2d)$
k-NN Embedding Graph	$O(Nd)$	—	$O(N \log N + kN)$

Table 10: Breakdown of computational costs for different network construction methods. All methods require embedding N posts into d -dimensional space. CANE and t-CANE include clustering via DP-Means (K clusters), and use approximate user-user graph construction via FAISS. Baseline methods compute similarity at the post or user level without intermediate clustering. **Notation:** N - posts, U - users, d - embedding dim, K - clusters, T - time bins, k - neighbors.

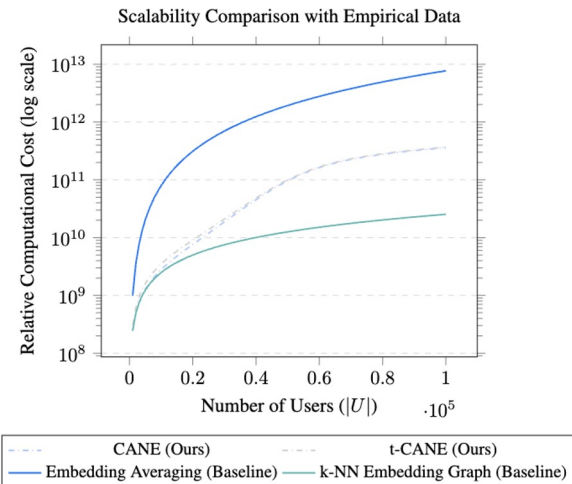


Figure 3: Comparison of computational complexity across network construction methods using empirically-informed scaling trends. CANE and t-CANE scale more efficiently with users due to FAISS and clustering, while baselines exhibit steep cost increases tied to post or user volume.

t-Cane Component Ablation

Model Variant	ROC-AUC	Precision	Recall	F ₁
Full t-CANE	0.98 \pm 0.01	1.00 \pm 0.00	0.75 \pm 0.01	0.83 \pm 0.01
<i>w/o Memory ($\alpha=0$)</i>	0.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00
<i>w/o Decay ($\beta=0$)</i>	0.98 \pm 0.01	1.00 \pm 0.00	0.73 \pm 0.02	0.81 \pm 0.02
<i>No Memory or Decay (Average Similarity)</i>	0.96 \pm 0.01	1.00 \pm 0.00	0.72 \pm 0.02	0.79 \pm 0.02
<i>No Memory or Decay (Naive Aggregation)</i>	0.97 \pm 0.00	0.99 \pm 0.00	0.71 \pm 0.02	0.80 \pm 0.01

Table 13: Ablation study on key components of **t-CANE** based on results from the China Information Operations detection task. All metrics are macro-averaged. The full model incorporates temporal memory (α) and decay smoothing (β). Removing memory ($\alpha=0$) leads to a collapse in performance, as user similarity cannot meaningfully propagate across time. The final two rows reflect ablations of both components, where user similarity is computed independently at each timestep and aggregated via averaging or summation.

Baseline Network Construction Methods

Network Type	Construction Description
Co-Repost	Connect users who repost the same piece of content.
Co-URL	Connect users who share the same URLs in their posts.
Fast Repost	Connect users who repost identical content within a short time window.
Hashtag Sequence	Connect users based on ordered sequences of shared hashtags.
Text Similarity	Connect users if they post at least one highly similar post; edge weight reflects average text similarity across matches (Pacheco et al. 2021).
k-NN Embedding Graph	Build a full text-to-text kNN graph first, then induce a user-to-user graph reflecting overall proximity across all posts (Ng, Cruickshank, and Carley 2023).
Fused Graph	Construct a unified network where users are linked if they are connected in any underlying similarity network (Co-Repost, Co-URL, Fast Repost, Hashtag Sequence, or Text Similarity) (Luceri et al. 2024a).

Table 1: Baseline similarity network construction methods.

Information Operations Results

Method	Macro-F ₁	AUC
Co-Repost	0.71 ± 0.02	0.77 ± 0.03
Co-URL	0.50 ± 0.02	0.56 ± 0.06
Fast Repost	0.46 ± 0.01	0.54 ± 0.02
Hashtag Sequence	0.51 ± 0.04	0.49 ± 0.03
Text Similarity	0.53 ± 0.01	0.33 ± 0.00
k-NN Embedding Graph	0.58 ± 0.02	0.39 ± 0.01
Fused Graph	0.74 ± 0.02	0.81 ± 0.02
CANE	0.72 ± 0.02	0.91 ± 0.01
t-CANE	0.83 ± 0.01	0.98 ± 0.01

Table 2: Performance on the China IO dataset.

Method	Macro-F ₁	AUC
Co-Repost	0.80 ± 0.01	0.73 ± 0.01
Co-URL	0.49 ± 0.00	0.52 ± 0.01
Fast Repost	0.59 ± 0.01	0.60 ± 0.01
Hashtag Sequence	0.54 ± 0.03	0.48 ± 0.01
Text Similarity	0.62 ± 0.03	0.52 ± 0.02
k-NN Embedding Graph	0.65 ± 0.01	0.57 ± 0.00
Fused Graph	0.81 ± 0.01	0.76 ± 0.01
CANE	0.82 ± 0.01	0.94 ± 0.00
t-CANE	0.90 ± 0.01	0.94 ± 0.02

Table 3: Performance on the Iran IO dataset.

Ideology Prediction Results

Method	Similarity Network	Macro-F ₁	AUC
Baseline	Co-Repost	0.43 ± 0.30	0.56 ± 0.22
	Co-URL	0.51 ± 0.26	0.53 ± 0.05
	Fast Repost	0.48 ± 0.04	0.51 ± 0.03
	Hashtag Sequence	0.65 ± 0.22	0.63 ± 0.14
	Text Similarity	0.39 ± 0.12	0.32 ± 0.09
	k-NN Embedding Graph	0.41 ± 0.03	0.34 ± 0.11
	Fused Graph	0.69 ± 0.15	0.70 ± 0.05
Ours	CANE	0.81 ± 0.01	0.73 ± 0.02
	t-CANE	0.83 ± 0.01	0.76 ± 0.02

Table 4: Ideological classification on X.

Method	Similarity Network	Macro-F ₁	AUC
Baseline	Hashtag Sequence	0.71 ± 0.03	0.62 ± 0.01
	Text Similarity	0.38 ± 0.14	0.42 ± 0.07
	k-NN Embedding Graph	0.45 ± 0.10	0.45 ± 0.08
	(Partial) Fused Graph	0.71 ± 0.03	0.64 ± 0.02
Ours	CANE	0.83 ± 0.02	0.83 ± 0.01
	t-CANE	0.82 ± 0.02	0.83 ± 0.03

Table 5: Ideological classification on TikTok.

Cross-Platform Narrative Prediction Results

Method	Similarity Network	Macro-F ₁	AUC
Baseline	Random GCN	0.00 ± 0.00	0.56 ± 0.03
	Co-URL	0.01 ± 0.00	0.43 ± 0.04
	Hashtag Sequence	0.11 ± 0.04	0.48 ± 0.08
	Text Similarity	0.02 ± 0.00	0.58 ± 0.05
	k-NN Embedding Graph	0.02 ± 0.01	0.61 ± 0.06
	(Partial) Fused Graph	0.05 ± 0.01	0.64 ± 0.06
Ours	CANE	0.30 ± 0.02	0.89 ± 0.02
	t-CANE	0.35 ± 0.06	0.94 ± 0.02

Table 6: Performance of similarity networks for cross-platform narrative engagement prediction (t=7).

Data Efficiency Comparisons

Category	Method	China	Iran
Baseline	Co-Repost	15%	20%
	Co-URL	20%	25%
	Fast Repost	20%	20%
	Hashtag Sequence	80%	60%
	Text Similarity	5%	10%
	k-NN Embedding Graph	10%	5%
	Fused Graph	15%	30%
Ours	CANE	5%	10%
	t-CANE	5%	5%

Table 15: Percentage of training data required to reach 95% of peak AUC for each method on the China and Iran IO campaigns. Lower values reflect higher data efficiency.

Category	Method	X	TikTok
Baseline	Co-Repost	15%	–
	Co-URL	20%	–
	Fast Repost	20%	–
	Hashtag Sequence	80%	60%
	Text Similarity	5%	10%
	k-NN Embedding Graph	10%	5%
	Fused Graph	15%	30%
Ours	CANE	5%	10%
	t-CANE	5%	5%

Table 21: Percentage of training data required to reach 95% of peak AUC for each method on the X and TikTok datasets. Lower values indicate faster convergence toward near-optimal performance.

Migration Statistics

Cluster Type	Top 1	Top 2	Top 3	Top 4	Top 5
Significant Migration	69.33%	77.73%	79.41%	79.83%	80.67%
Simple Migration	67.72%	76.61%	78.29%	78.99%	79.45%

Table 28: Percentage of cross-platform narrative clusters where a bridge user appeared within the first n posts on the receiving platform.

Bridge User Engagement Metrics

Metric	Median	Bridge Percentile (Platform-Normalized)
Total Posts	14.00	58.49%
Reply Count	0.16	47.06%
Like Count	0.58	53.26%
Repost Count	0.02	55.19%

Table 26: Median engagement metrics and corresponding platform-normalized percentile ranks for bridge users. Percentiles are calculated within each platform and then averaged.

Bridge User Detection Across Network Construction Methods

Graph Type	% Users	% Posts	% Narratives Introduced
Co-URL	0.00%	0.00%	0.00%
Hashtag Sequence	0.04%	0.37%	4.21%
Text Similarity	1.42%	2.21%	15.70%
kNN Embedding Graph	1.73%	2.41%	25.90%
Fused Graph	2.68%	3.19%	27.00%
Ours (Discourse)	0.33%	2.14%	67.72%

Table 27: Comparison of bridge user detection and narrative introduction across graph construction methods.

Table 2: Multi-horizon emergence detection performance. Reported are AUC, F1, and Precision scores (mean \pm std) across prediction windows.

Method	3d	AUC 7d	14d	3d	F1 7d	14d	3d	Precision 7d	14d
<i>No Network Baselines</i>									
Popularity Baseline	0.50 \pm 0.00	0.50 \pm 0.00	0.50 \pm 0.00	0.24 \pm 0.00	0.28 \pm 0.00	0.31 \pm 0.00	0.13 \pm 0.00	0.16 \pm 0.00	0.18 \pm 0.00
Platform Transitions	0.62 \pm 0.00	0.79 \pm 0.00	0.76 \pm 0.00	0.45 \pm 0.00	0.53 \pm 0.00	0.54 \pm 0.00	0.36 \pm 0.00	0.43 \pm 0.00	0.43 \pm 0.00
LSTM (engagement)	0.50 \pm 0.00	0.50 \pm 0.00	0.50 \pm 0.00	0.24 \pm 0.02	0.29 \pm 0.00	0.31 \pm 0.01	0.14 \pm 0.00	0.17 \pm 0.01	0.19 \pm 0.01
<i>Diffusion Models</i>									
Hawkes Process	0.57 \pm 0.00	0.56 \pm 0.01	0.55 \pm 0.001	0.26 \pm 0.01	0.28 \pm 0.00	0.32 \pm 0.00	0.18 \pm 0.00	0.21 \pm 0.00	0.19 \pm 0.00
Independent Cascade	0.58 \pm 0.00	0.57 \pm 0.00	0.56 \pm 0.00	0.26 \pm 0.00	0.28 \pm 0.00	0.32 \pm 0.0	0.17 \pm 0.00	0.19 \pm 0.00	0.19 \pm 0.00
<i>Other Networks</i>									
Co-URL	0.50 \pm 0.01	0.50 \pm 0.00	0.50 \pm 0.00	0.25 \pm 0.01	0.24 \pm 0.01	0.23 \pm 0.01	0.15 \pm 0.03	0.13 \pm 0.00	0.13 \pm 0.01
Hashtag Sequence	0.51 \pm 0.03	0.50 \pm 0.00	0.50 \pm 0.00	0.24 \pm 0.03	0.26 \pm 0.01	0.23 \pm 0.00	0.15 \pm 0.02	0.14 \pm 0.00	0.13 \pm 0.01
Text Similarity	0.61 \pm 0.02	0.72 \pm 0.01	0.76 \pm 0.01	0.40 \pm 0.01	0.50 \pm 0.01	0.47 \pm 0.00	0.32 \pm 0.01	0.36 \pm 0.00	0.35 \pm 0.00
k-NN Embedding	0.65 \pm 0.01	0.71 \pm 0.00	0.76 \pm 0.01	0.41 \pm 0.02	0.52 \pm 0.02	0.47 \pm 0.01	0.32 \pm 0.02	0.37 \pm 0.01	0.36 \pm 0.01
Fused Network	0.69 \pm 0.01	0.74 \pm 0.01	0.79 \pm 0.02	0.44 \pm 0.04	0.52 \pm 0.01	0.49 \pm 0.01	0.34 \pm 0.01	0.39 \pm 0.00	0.35 \pm 0.02
<i>Discourse Network</i>									
No Claim Extraction	0.79 \pm 0.01	0.80 \pm 0.01	0.80 \pm 0.00	0.49 \pm 0.01	0.58 \pm 0.01	0.59 \pm 0.02	0.49 \pm 0.00	0.45 \pm 0.01	0.48 \pm 0.02
Claim Extraction	0.88 \pm 0.01	0.87 \pm 0.01	0.86 \pm 0.02	0.58 \pm 0.02	0.61 \pm 0.02	0.62 \pm 0.01	0.56 \pm 0.02	0.48 \pm 0.01	0.54 \pm 0.01
+ Platform Transitions	0.94 \pm 0.02	0.94 \pm 0.01	0.92 \pm 0.01	0.66 \pm 0.01	0.72 \pm 0.03	0.68 \pm 0.02	0.65 \pm 0.03	0.63 \pm 0.01	0.74 \pm 0.01
Random Baseline	0.49 \pm 0.03	0.50 \pm 0.02	0.53 \pm 0.02	0.24 \pm 0.03	0.28 \pm 0.01	0.31 \pm 0.03	0.14 \pm 0.03	0.24 \pm 0.04	0.19 \pm 0.03