Bridging the Narrative Divide:
Cross-Platform Discourse Networks in Fragmented Ecosystems

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

pkgerard@

patrickgerard_

patrikgerard.bsky.social

Patrick Gerard — University of Southern California

USC INFORMATION SCIENCES INSTITUTE

Who Am I?

Patrick Gerard — PhD Candidate, USC Information Sciences Institute

I study how information, misinformation, and influence operations operate **across the online information ecosystem.**

Advised by Emilio Ferrara and Kristina Lerman (USC ISI); close collaborations with Mohsen Mosleh (Oxford), David Rand (MIT/Cornell), Luca Luceri (USC ISI), and Hans Hanley (Stanford).

Works in this talk: ICWSM 2026, WWW 2026, work under review at PNAS.

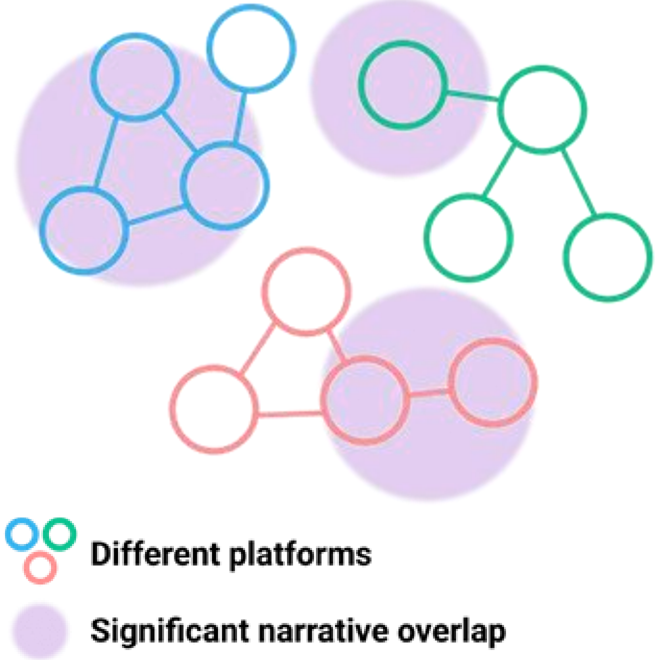
This Talk

A More Realistic Way to Represent Users, Channels, and Creators Across the Information Ecosystem

Current methods for **representing users and their networks tend to be platform-siloed** — networks built from retweets, follows, forwards, each confined to one platform.

That made sense when platforms *were* separate worlds.

The result is a **huge blind spot.**



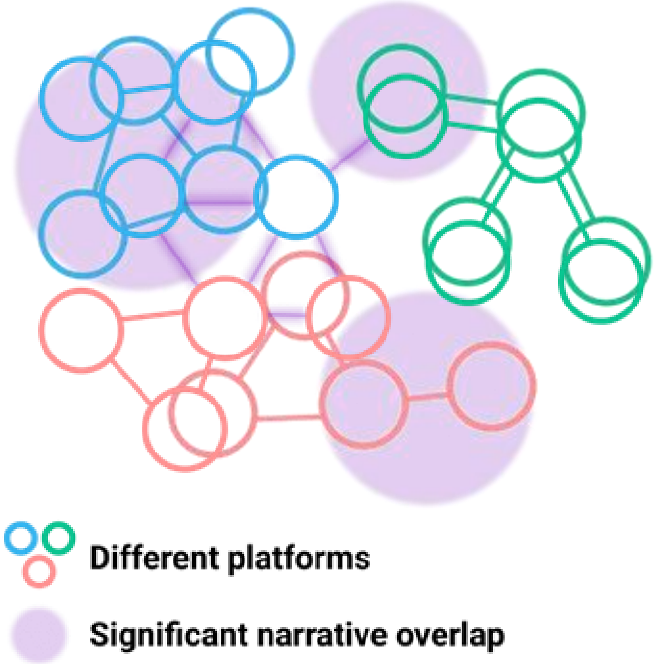
This Talk

A More Realistic Way to Represent Users, Channels, and Creators Across the Information Ecosystem

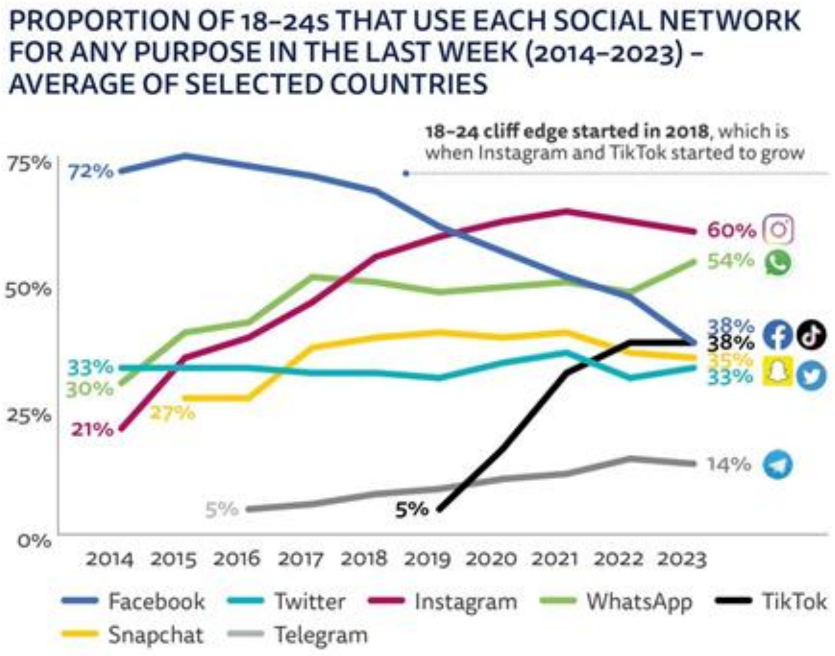
A new way to **model the full ecosystem:**

Three parts:

(1) How we build that representation;
(2) What it lets us understand and predict that we couldn't before;
(3) Where we're pushing it next.



An Era of Social Media Fragmentation



Source: Reuters

The **information ecosystem has been fracturing** since 2018.

User **networks** have stayed largely **platform-bound**.

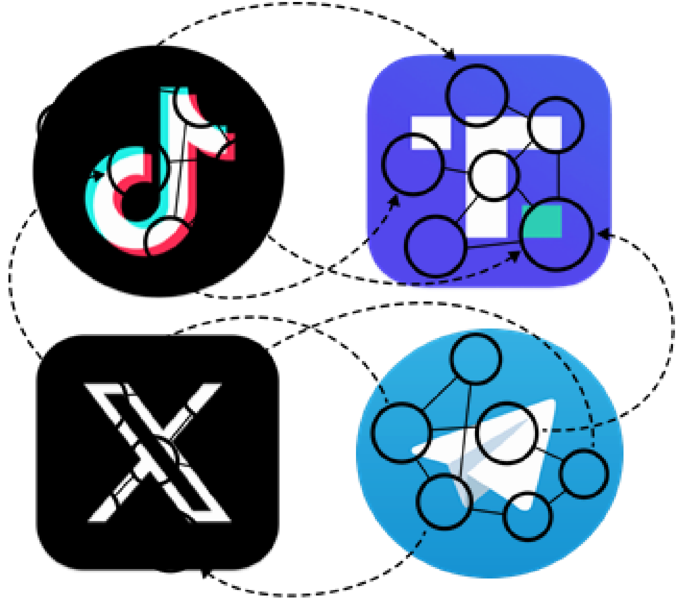
USC INFORMATION SCIENCES INSTITUTE

The Same Narrative, Told Across Many Platforms At Once

Springfield, OH, 2024: “They’re eating the dogs, they’re eating the cats.”

Source: ABC

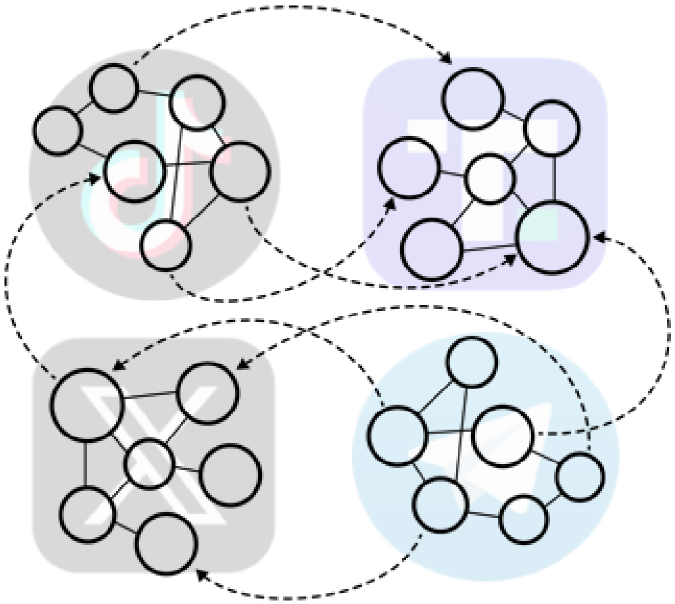
Modern Narratives Take Shape *Across* Platforms



The information ecosystem **defies platform boundaries.**

But current methods for **modeling users and information spread** are platform-locked.

Modern Narratives Take Shape *Across* Platforms



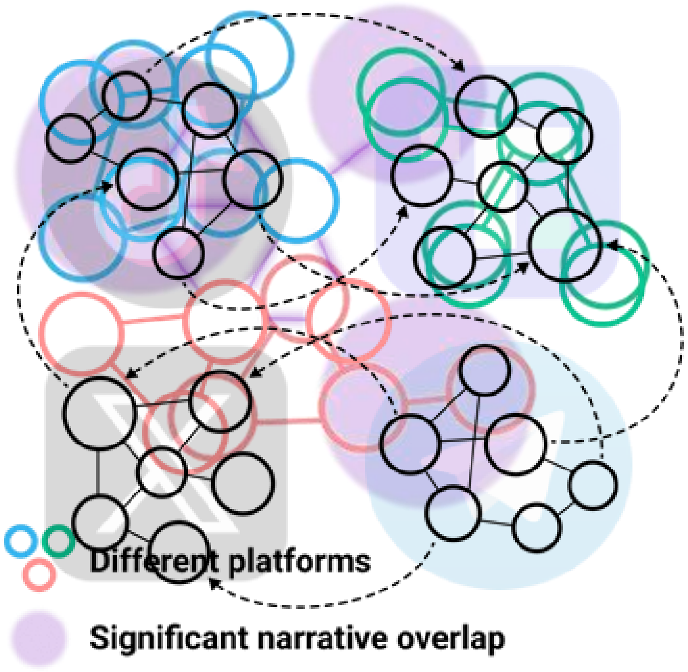
*So how do we **represent users** and **predict information spread** in a world without platform boundaries?*

Section 1:
A Platform-Agnostic, Content-Based Node Representation

AAAI ICWSM 2026: Bridging the Narrative Divide

USC INFORMATION SCIENCES INSTITUTE

Rethinking User Representation Through Shared Discourse

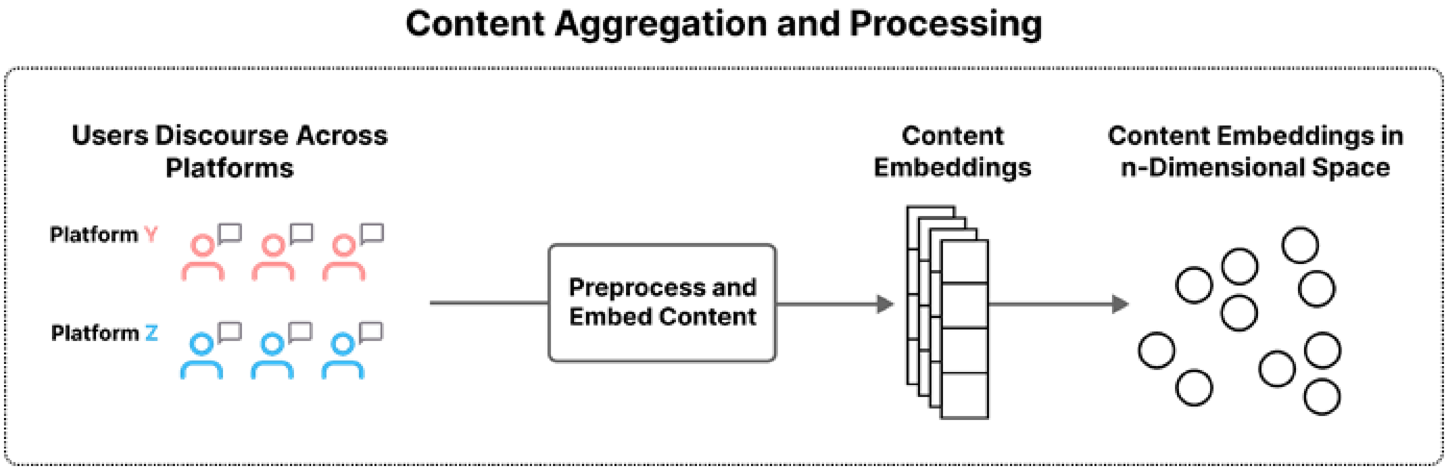


Instead of **"whom did this user repost?"**
→ **"what ideas does this user inhabit?"**

- Each user = **distribution over narratives** they engage with
- Users who engage with the same narratives connect — regardless of platform.
- Content is the only input: no follows, reposts, or platform metadata required

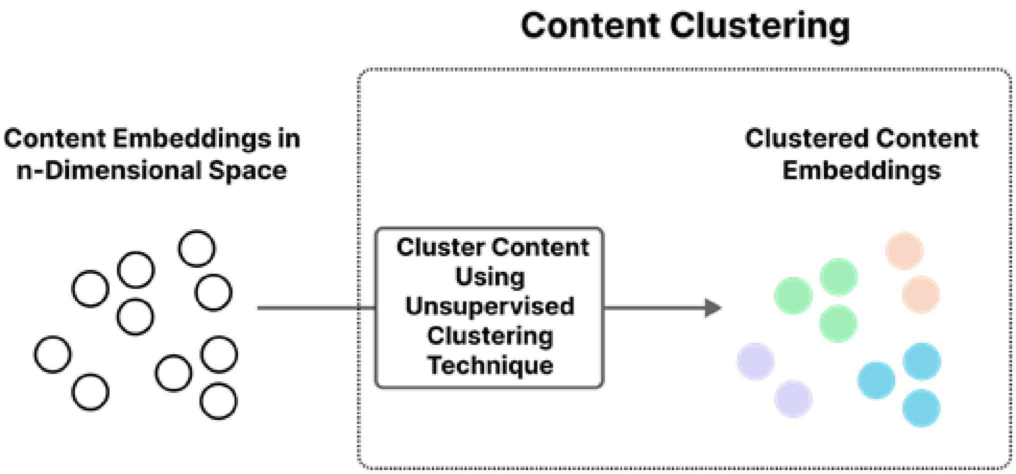
USC INFORMATION SCIENCES INSTITUTE

Our Approach: Modeling Discourse as Network



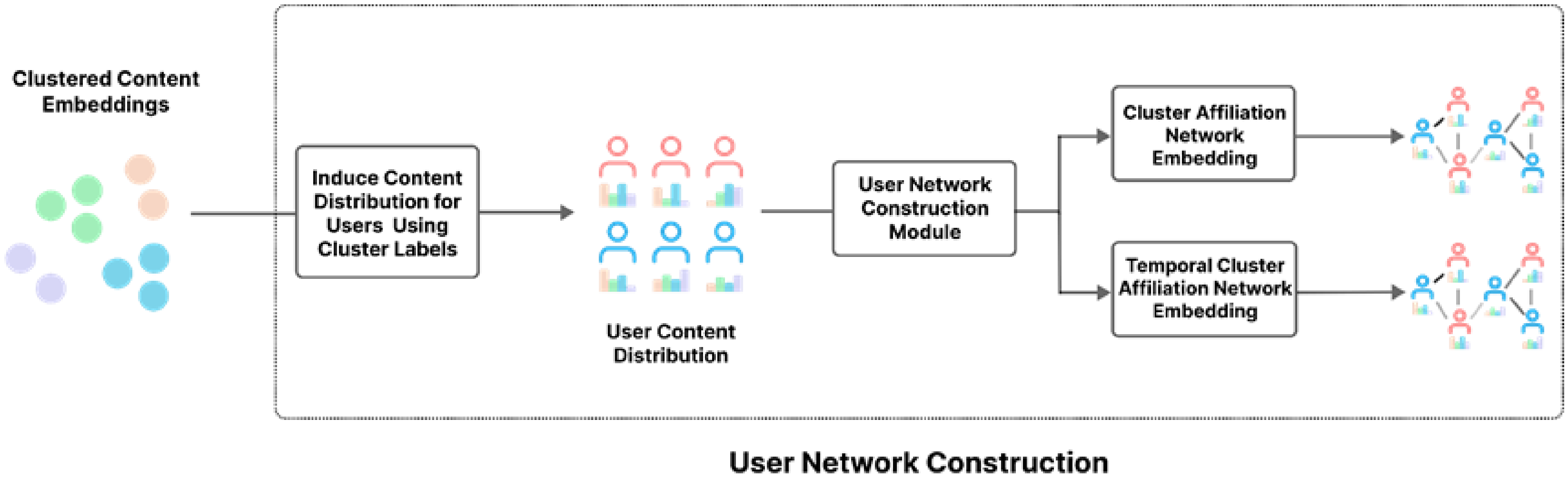
Step 1: Content Aggregation and Processing *Collect cross-platform discourse and transform text into comparable embeddings*

Our Approach: Modeling Discourse as Network



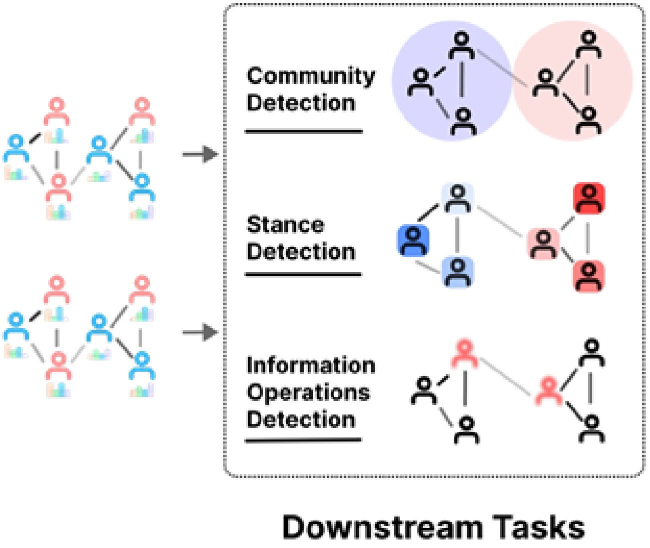
Step 2: Content Clustering *Group similar discourse into thematic clusters using unsupervised learning*

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*

Does it Work? Let's Test

We **benchmark** our approach against **classic network construction methods** on **established and novel CSS tasks** using a consistent set of classifiers.

Network Type	How Users Are Connected
Co-Repost	Repost the same content
Co-URL	Share the same URLs
Fast Repost	Repost identical content within a short window
Hashtag Sequence	Use the same ordered hashtag sequences
Text Similarity	Post highly similar text
k-NN Embedding Graph	Close in overall text embedding space
Fused Graph	Linked if connected in any of the above networks

Influence Operation Detection
Detecting covert accounts that promote coordinated influence campaigns (**X**: Seckin et al. 2024).

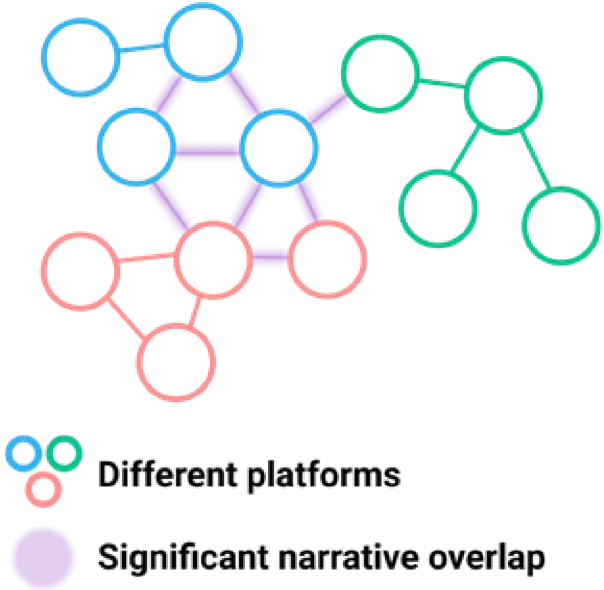
Ideological Stance Prediction
Inferring a user's political leaning based on their network position (**X**: Balasubramanian et al. 2024; **Tiktok**: Pinto et al. 2024).

Cross-Platform Engagement Prediction
Predicting which users will engage with content as it moves across platforms (**X/Truth Social**).

USC INFORMATION SCIENCES INSTITUTE

Results: Leaner, broader, and consistently best-in-class

Discourse Networks perform better, need less data, and cover more users



IO Detection

11% F1, 24% AUC improvement over baseline

Ideological Classification

15-20% F1 improvement over best baseline

Cross-Platform Engagement:

47% AUC improvement over best baseline

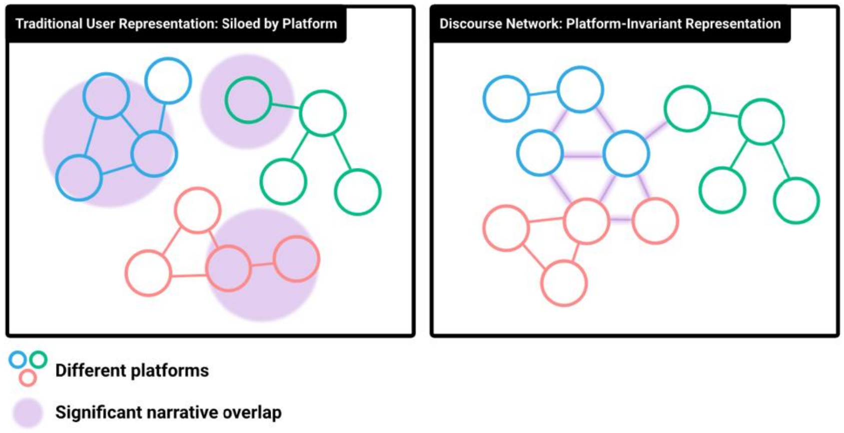
Reaches peak performance with a fraction of users content; covers nearly all users.

Takeaways – Section 1

A Node Doesn't Have to be Tied to a Platform

Birds of a feather talk together, regardless of platform

Building networks from that signal gives you **one graph of the whole ecosystem**, instead of a separate graph per platform.



Section 2:

Moving From Structure to Forecasting:

Predicting When Narratives Route Across Platforms

ACM WWW 2026: Cross-Platform Narrative Prediction

USC INFORMATION SCIENCES INSTITUTE

Hurricane Helene – September, 2024

Category 4 landfall in Florida's Big Bend, then catastrophic inland flooding across western North Carolina.

One of the **deadliest U.S. hurricanes in decades**: hundreds killed, entire towns cut off, massive federal relief response.

USC INFORMATION SCIENCES INSTITUTE

Could we have predicted this?

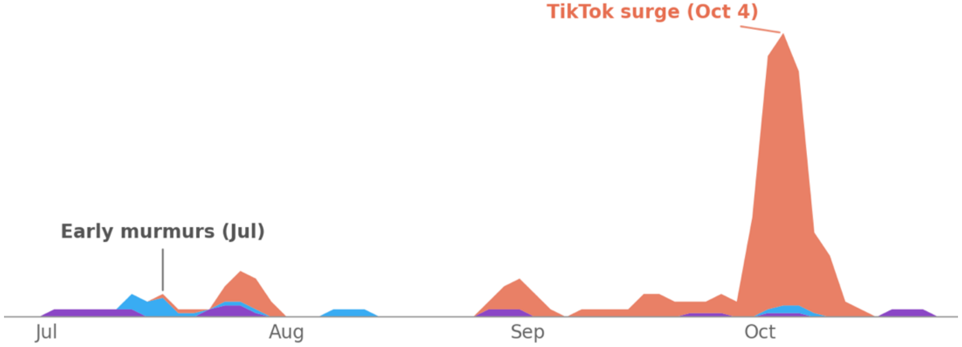
“…an armed group harassing Helene aid workers.” – AP News

Hurricane Helene – September, 2024

“FEMA is surveilling US citizens and [later] stealing from hurricane victims” — post volume per platform, Apr–Nov 2024

Current models for Cross-Platform Prediction fail across 2024 narratives:

- **LSTMs**: 0.50 AUC
- **Popularity Features**: 0.50.
- **Hawkes Processes:** 0.55–0.58
- **Cascade Models**: 0.55–0.58



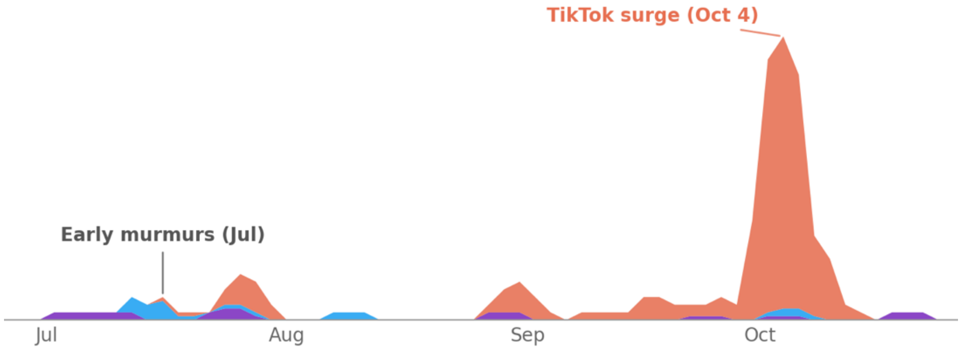
Post volume across fringe, Twitter/X, and TikTok, Jul–Oct 2024.

Hurricane Helene – September, 2024

“FEMA is surveilling US citizens and [later] stealing from hurricane victims” — post volume per platform, Apr–Nov 2024

A seed was sown months earlier; the ground became fertile; the event lit it.

How do you put “the ground was ready” into a time-series model's training data?



Post volume across fringe, Twitter/X, and TikTok, Jul–Oct 2024.

USC INFORMATION SCIENCES INSTITUTE

Hurricane Helene – September, 2024

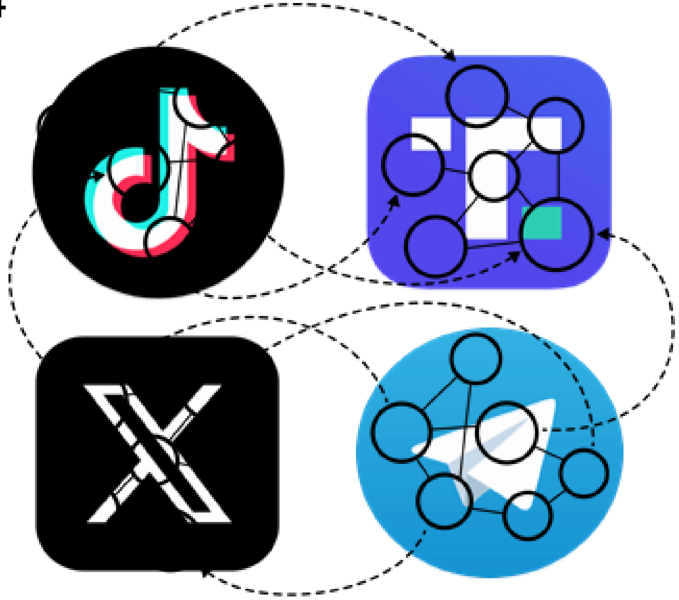
“FEMA is surveilling US citizens and [later] stealing from hurricane victims” — post volume per platform, Apr–Nov 2024

What if we used **social proximity** instead?

“~~What does the time-series do next?~~”

“Are this user's cross-platform neighbors engaging?”

Understand each platform's *role in the wider ecosystem*



Can We Use Discourse Networks for Prediction?

Can we predict when a narrative will travel from one platform to another?

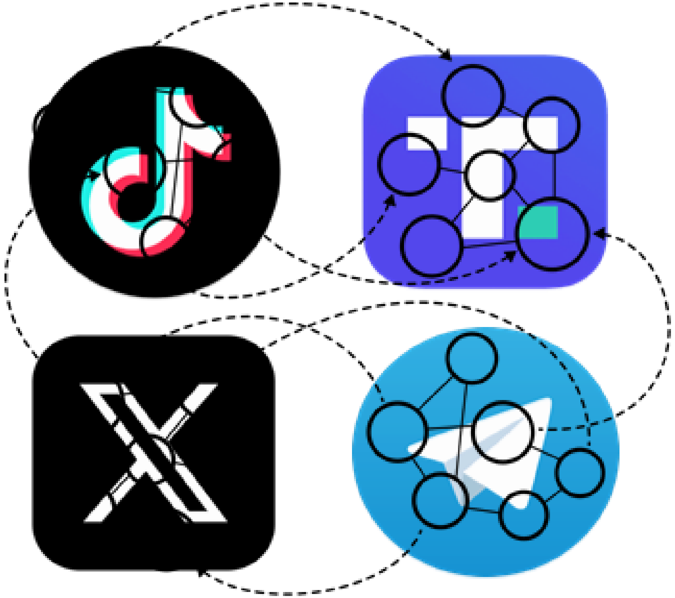
Given a narrative active on a source platform at time t, **predict whether it emerges on a target platform** within $\Delta \in \{3, 7, 14\}$ days.

Approach: for each user on the target platform, ask if this user's cross-platform neighbors are engaging.

Trained and evaluated on 5.7M posts across **X, TikTok, Truth Social, and Telegram**, with strict train-up-to-time-t / predict-forward evaluation.

It works: the intractable becomes a random forest

Cross-platform social proximity carries the prediction



0.94 AUC across all three horizons — **+27%** over the next-best baseline, **+54%** over diffusion simulation.

Run retrospectively over the 2024 election: flagged the Springfield rumor **days before its jump to X,** and the FEMA narrative **before relief crews faced guns.**

Takeaways – Section 2

Cross-Platform Representation Makes the Ecosystem Predictable

People move through the information ecosystem in **predictable** ways.

If someone's neighbors on other platforms are talking about something, they probably will be soon.

Our **prediction systems** should be built around that

Article About This!

USC INFORMATION SCIENCES INSTITUTE

Section 3:
What Are We Doing With It Now?

Moving from four platforms to the “whole” ecosystem

What Are We Doing With It Now?

Same representation, Bigger Lens

Mapping the "full" ecosystem
~34M posts across 15 platforms from the 2024 election

Going past social media
Podcasts as nodes in information ecosystem

What if we could see the whole 2024 election as one graph?

Scaling the representation to 15 platforms – with Mosleh, Rand

Initial Findings: The ecosystem behaves as a single connected object, and platforms occupy structurally distinct positions in the graph.

What this means: Now, “where does this narrative sit in the ecosystem, and where is it headed” is a single well-defined graph query.

What if we could see the whole 2024 election as one graph?

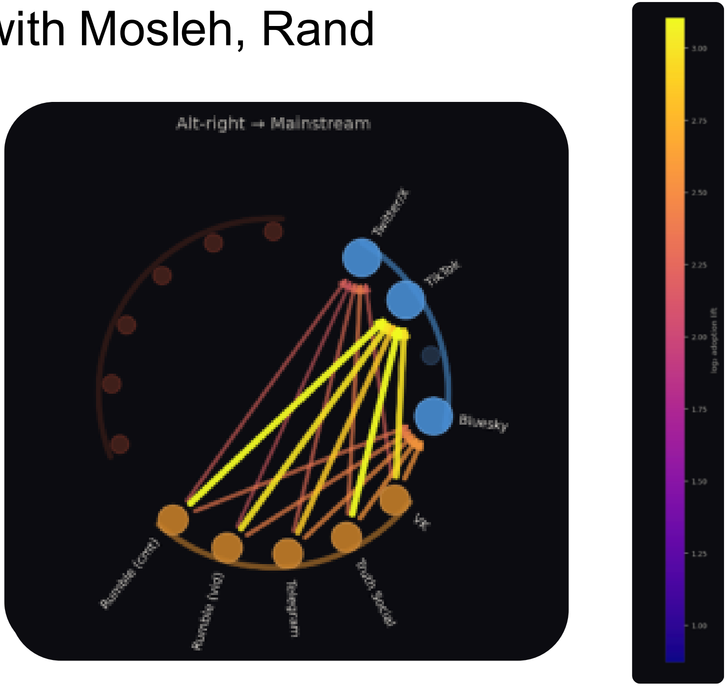
Scaling the representation to 15 platforms – with Mosleh, Rand

Example Finding: Meaningful cross-platform adoption lift (2–6×) across virtually every platform pair.

TikTok shows **anomalously high adoption lift** as a target.

Tightly clustered ideological communities show comparatively lower lift as targets.

Cross-platform lift = adoption rate when user has active neighbor on source platform vs. within-platform. Full corpus, all platform pairs.



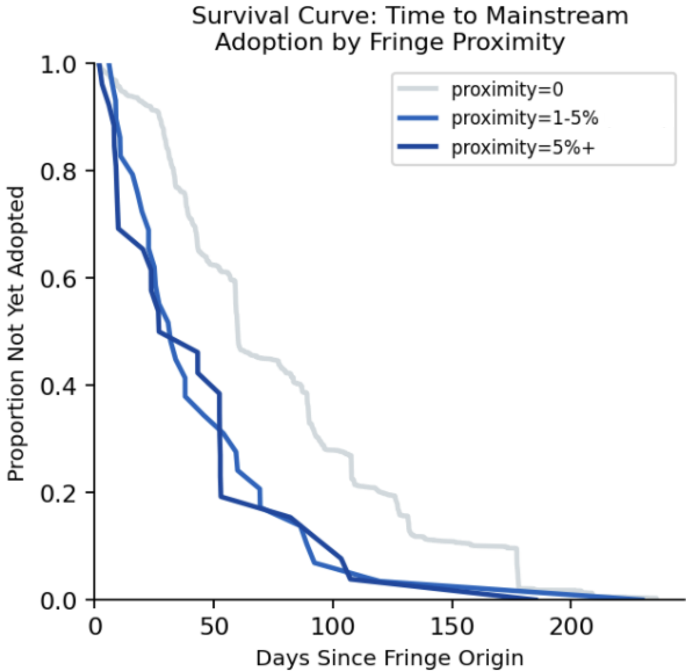
What if we could see the whole 2024 election as one graph?

Scaling the representation to 15 platforms – with Mosleh, Rand

Example Finding: Cross-platform proximity is highly predictive of a narrative's leap from fringe to mainstream platforms.

Users **with cross-platform exposure reliably adopt fringe narratives**.

Users **without it largely never adopt**, no matter how long the narrative persists.



What About Podcast Propaganda?

TENET Media — detecting foreign influence with zero behavioral anomalies

Case: TENET Media, a podcast network covertly funded by Russian state actors in 2024, using real domestic hosts throughout.

Initial Findings: The operation is detectable, purely as a content-space narrative concentration.

What this means: Once podcast hosts are nodes in the information ecosystem, we can look for anomalous positions in narrative space.

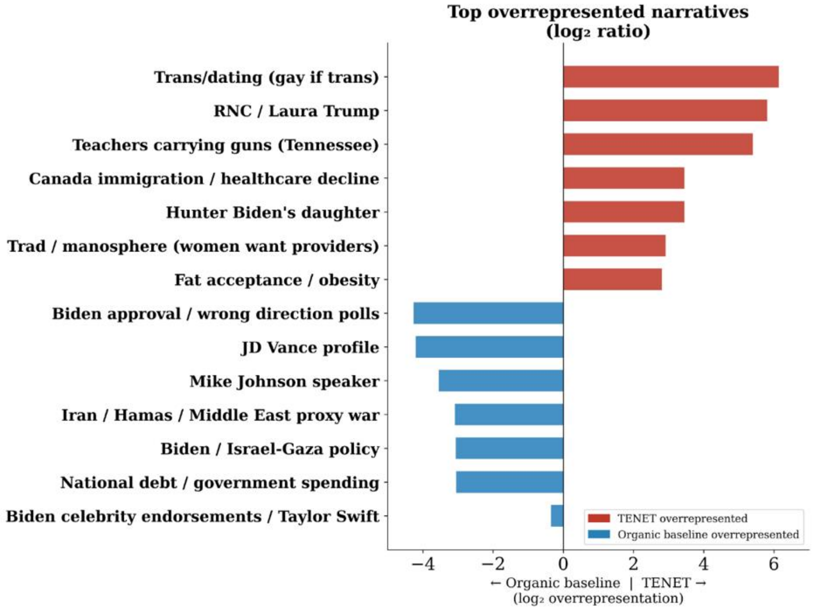
USC INFORMATION SCIENCES INSTITUTE

What About Podcast Propaganda?

TENET Media — detecting foreign influence with zero behavioral anomalies

Example Finding: TENET occupies an anomalous position in the right-wing podcast narrative space, overinvesting in a distinctive culture-war niche .

The same fingerprint holds up across independent methods (density ratios, lift scores, within-host comparison of the same podcaster's TENET vs. non-TENET episodes)

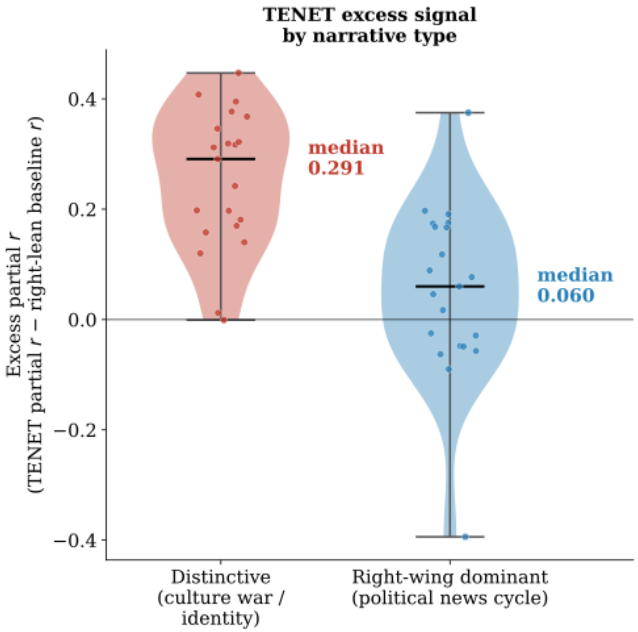


What About Podcast Propaganda?

TENET Media — detecting foreign influence with zero behavioral anomalies

Example Finding: TENET's narrative volume predicts subsequent social media volume on exactly those distinctive narratives.

Controlling for the organic right-wing baseline, TENET predicts downstream social media volume only on its distinctive narratives, not the shared news cycle.



Final Takeaways

Representing and predicting information as it moves across platforms and formats

Any **discourse-producing node** — a user, a channel, a podcast host — **can live in one shared graph**, regardless of platform or format.

People **move through that ecosystem predictably**: cross-platform neighbors reveal what someone engages with next, days ahead.

So, **the unit of analysis should be the entire information ecosystem** — for detection, for prediction, and for whatever comes next.

ICWSM Paper **WWW Paper**

```
pip install discourse-networks
```

pkgerard@

patrickgerard_

Appendix

What Do We Mean by "Narrative"?

A coherent unit of discourse, not a keyword or a hashtag

Following prior work on narrative tracking, we define a **narrative** as a collection of posts addressing the same issue or event: a coherent unit of discourse that evolves across time and platforms.

Example manifestations of the same underlying narrative.

Platform	Expression of Narrative
X/Twitter	FEMA blocking private rescue ops in NC #HurricaneHelene #FEMA
TikTok	*[Video transcript]* "...apparently FEMA is telling volunteers they can't help hurricane victims? [inaudible] is actually happening here..."
Truth Social	More evidence of federal incompetence: FEMA actively interfering with private relief efforts in North Carolina.
Telegram	FEMA interference with volunteer rescue operations confirmed by multiple sources. Patterns consistent with a deliberate resource control strategy.

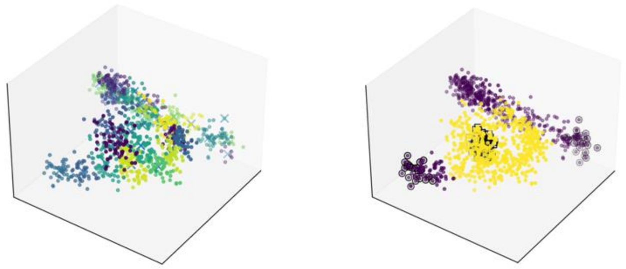
Why HNSW + FAISS?

The problem: Cross-platform narrative matching across 8.5M posts requires approximate nearest neighbor search that is both semantically precise and computationally tractable.

Why FAISS HNSW:

- Builds a multi-layer proximity graph — queries navigate greedily from coarse to fine, finding approximate nearest neighbors in log time
- Inner product on L2-normalized, mean-centered vectors = cosine similarity
- K=64 neighbors per document, edges filtered at τ=0.65 (validated via human annotation)
- Out-degree capped at 20 strongest edges to prevent hubs dominating graph structure

Why not exact search: Quadratic scaling makes exhaustive pairwise comparison intractable at corpus size. HNSW trades negligible recall loss for near-log query time.
Why not DP-Means: Assumes spherical, isotropic clusters. Scales quadratically. Global distance threshold acts identically across all clusters regardless of local density.



Left: DP-Means (K=41) fragments the narrative manifold into radial tiles.

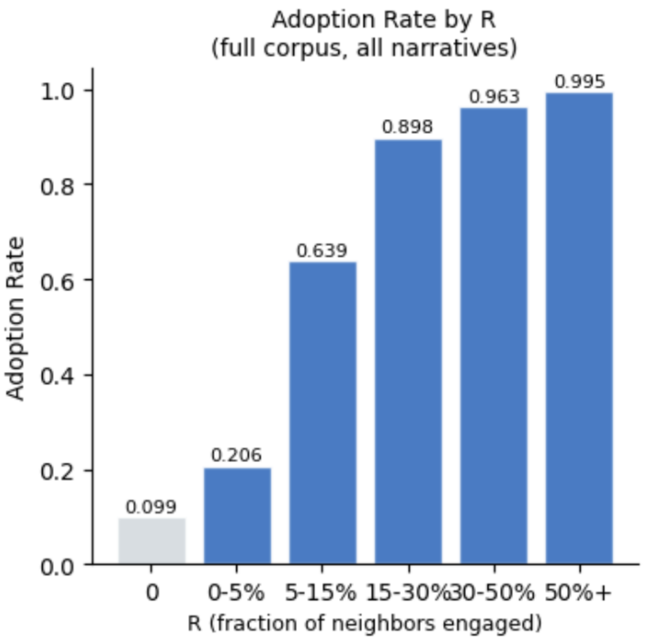
Finding 2: Discourse Proximity Predicts Adoption

Using a logistic regression controlling for platform, we predict whether a user engages with a narrative based only on how many neighbors have engaged.

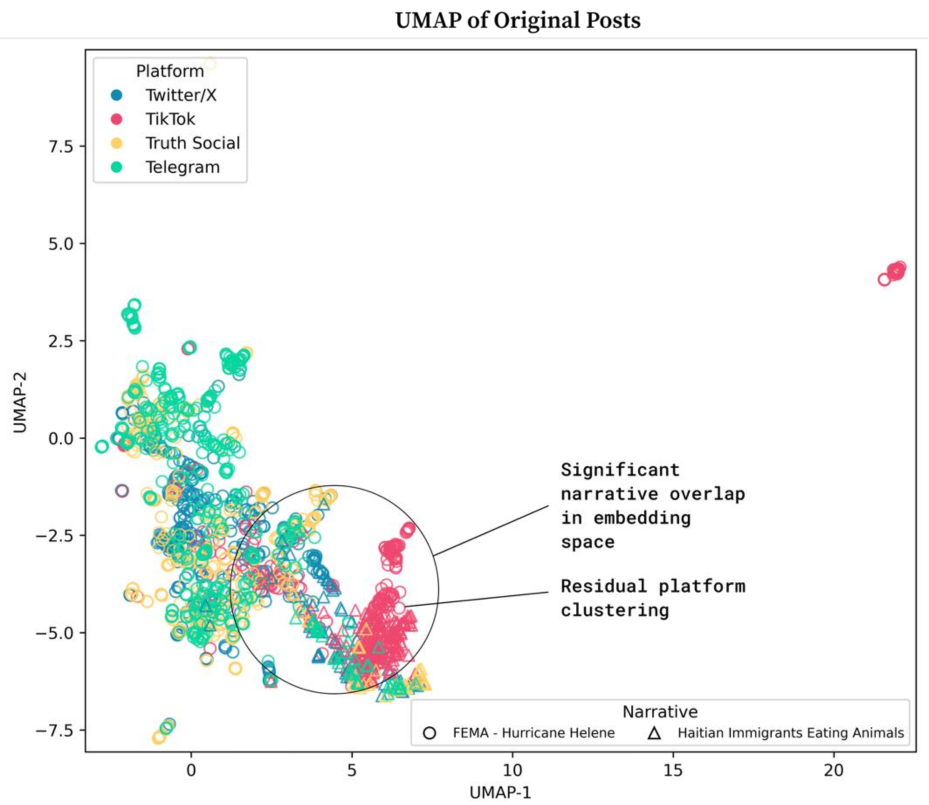
Adoption rate: 9.9% (no active neighbors)
→ Monotonically rises from there
→ 99.5% (R > 50%)

OR = 4.08, pseudo-R^2 = 0.57

Holds within AND across platform boundaries



Based on 5.1M user × narrative observations across 7,043 narratives. Logistic regression predicting whether a user engages with a narrative, controlling for their platform and the narrative's origin platform.



<u>Haitian Immigrants</u>:

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

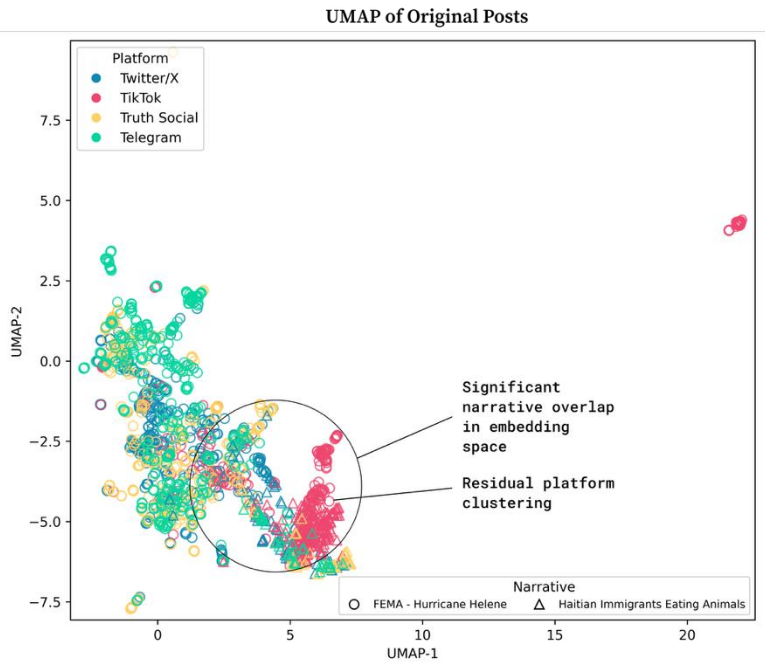
TikTok (video transcript)
"Residents are claiming their cats went missing after new arrivals moved in. I'm not saying it's true but something is going on…"

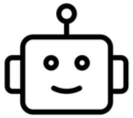
<u>Hurricane Helene</u>:

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?"

USC INFORMATION SCIENCES INSTITUTE





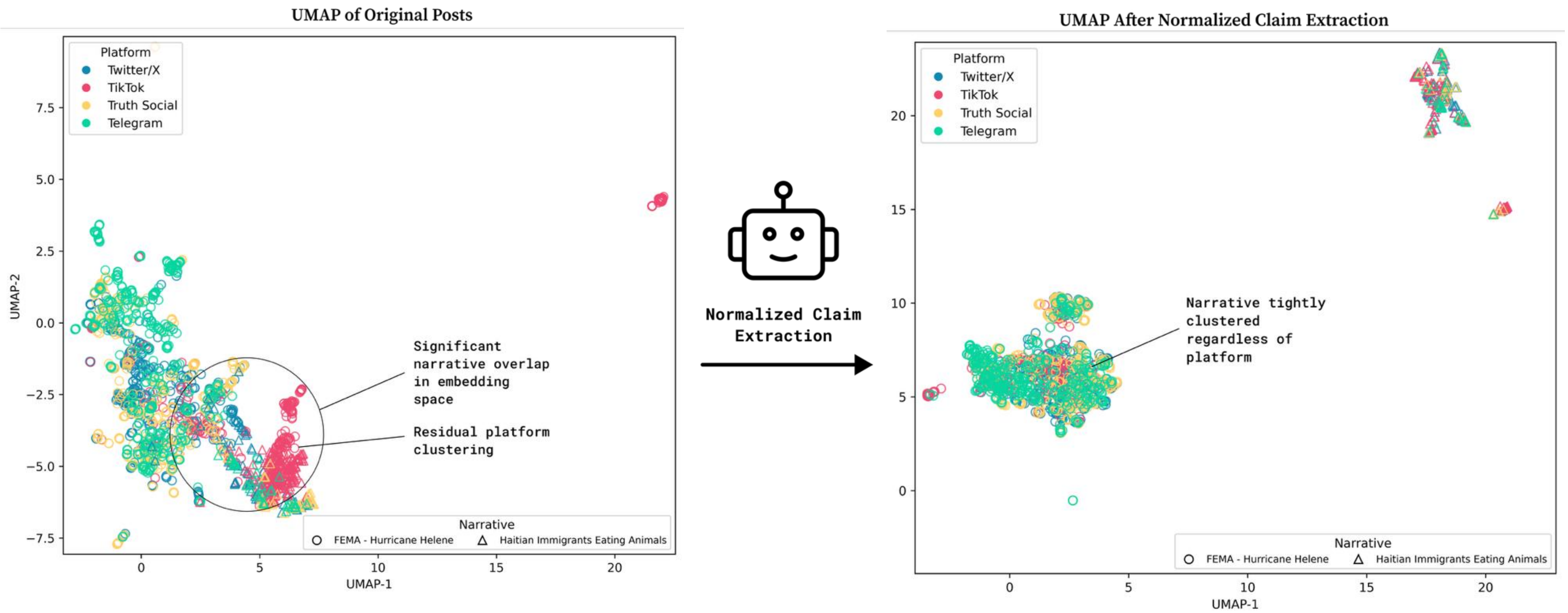
Normalized Claim Extraction

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
Platform
Twitter/X
TikTok
Truth Social
Telegram
UMAP-2
UMAP-1
Significant narrative overlap in embedding space
Residual platform clustering
Narrative
FEMA - Hurricane Helene
Haitian Immigrants Eating Animals
Normalized Claim Extraction
UMAP After Normalized Claim Extraction
Narrative tightly clustered regardless of platform

The Simulation

We model each narrative as a contagion spreading through the discourse network.

A user “gets infected” — adopts the narrative — based on how many of their discourse neighbors have already engaged.

Transmission probability = f(R) — the fraction of a user's neighbors already engaged

Outcome variable: final narrative reach: how many users ultimately adopt across the network

Validated against 1000 held-out narratives: r = 0.76, median ratio 0.61×. Appropriate for counterfactual analysis.

USC INFORMATION SCIENCES INSTITUTE

Section 3

The Simulation: Misinformation-Conditioned Counterfactuals

To test whether platform roles differ for misinformation specifically, we classified each narrative as misinformation or general political content.

Method: Taxonomy-grounded semantic matching against expert-constructed queries spanning 10 misinformation categories salient to the 2024 election: election integrity, candidate attacks, immigration, health misinformation, and others.

All matches human-validated.

Final sample: 326 misinformation narratives · 1,174 general political content narratives

Details and validation in appendix.

The Simulation

We model each narrative as a contagion spreading through the discourse network.

For each narrative we run:

- **15 single-platform removals** — remove all users of platform X, re-run simulation
- **105 pairwise joint removals** — remove all users of platforms X and Y simultaneously

Computed separately for **all narratives**, **misinfo narratives**, and **non-misinfo narratives**.

Outcome: % change in final narrative reach vs. baseline.

Validated against 1000 held-out narratives: r = 0.76, median ratio 0.61×. Appropriate for counterfactual analysis.

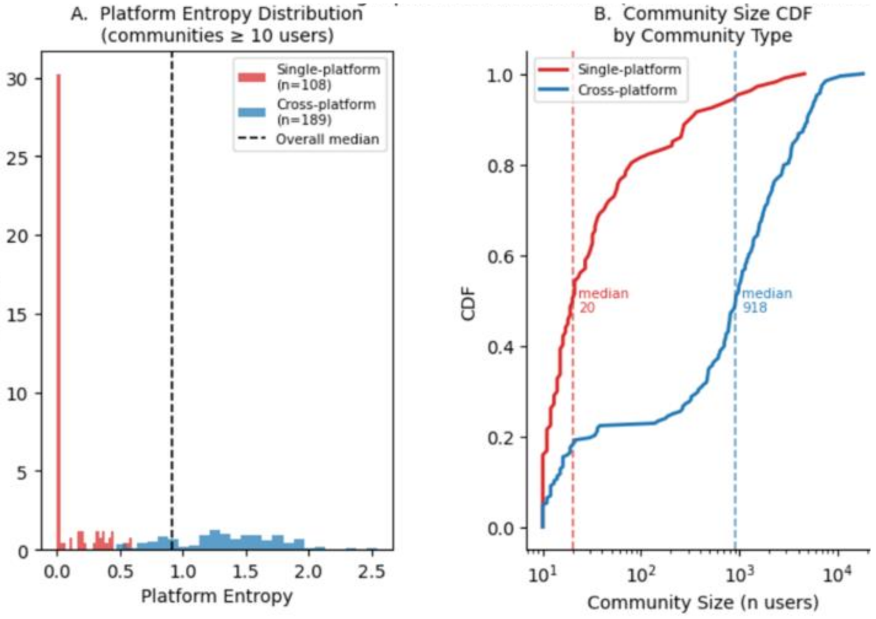
Finding: Communities are Cross-Platform Objects

Using Leiden community detection on the user network, we recover latent community structure across all platforms simultaneously.

99% of active users belong to communities that **span multiple platforms**.

Discourse network: Leiden community detection on narrative co-engagement edges. Active users = 2+ accepted narratives.

Platform entropy: $H = -\Sigma\, p_i \log p_i$ over platform membership within a community. Score of 0 = all users from one platform; higher = more evenly mixed across platforms.



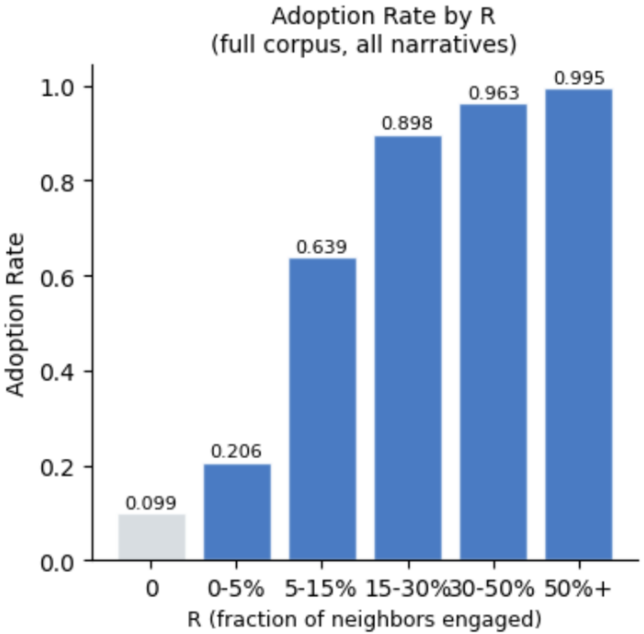
Finding: Discourse Proximity Predicts Adoption

Using a logistic regression controlling for platform, we predict whether a user engages with a narrative based only on how many neighbors have engaged.

Adoption rate: 9.9% (no active neighbors)
→ Monotonically rises from there
→ 99.5% (R > 50%)

OR = 4.08, pseudo-R^2 = 0.57

Holds within AND across platform boundaries



Based on 5.1M user × narrative observations across 7,043 narratives. Logistic regression predicting whether a user engages with a narrative, controlling for their platform and the narrative's origin platform.

USC INFORMATION SCIENCES INSTITUTE

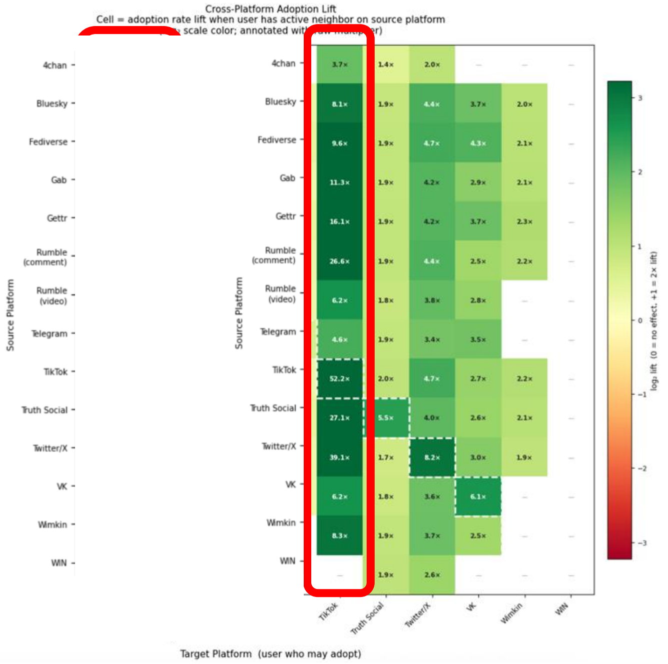
Section 1

Finding: Cross-Platform Lift is Pervasive

Cross-platform adoption rates are 2–6× higher when a user has an active neighbor on another platform.

TikTok shows anomalously high cross-platform adoption rates as a destination — users on TikTok with neighbors active on other platforms are far more likely to engage with the same narrative than on any other platform.

Tight ideological communities — Gettr, Gab, Truth Social — show lower cross-platform adoption rates but are not isolated from the pattern



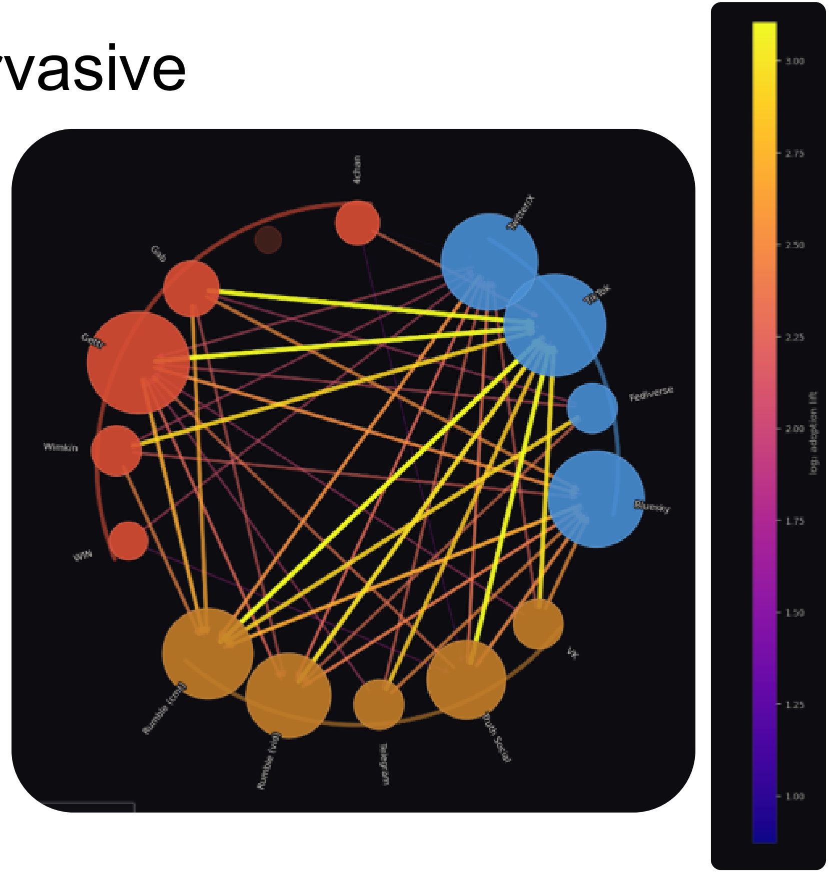
Cross-platform lift = adoption rate when user has active neighbor on source platform vs. within-platform. Full corpus, all platform pairs.

Finding: Cross-Platform Lift is Pervasive

Cross-platform adoption rates are 2–6× higher when a user has an active neighbor on another platform.

TikTok shows anomalously high cross-platform adoption rates as a destination — users on TikTok with neighbors active on other platforms are far more likely to engage with the same narrative than on any other platform.

Tight ideological communities — Gettr, Gab, Truth Social — show lower cross-platform adoption rates but are not isolated from the pattern.



Cross-platform lift = adoption rate when user has active neighbor on source platform vs. within-platform. Full corpus, all platform pairs.