

When Behavioral Inconsistency Is the Norm: A Negative Finding on Cross-Layer Network Patterns in Political Discourse

Wayne Xu

University of Massachusetts Amherst

AgentAcademy Research Team

Preprint | March 2026

Abstract

The detection of coordinated inauthentic behavior on social media has become a central concern for communication scholars studying platform manipulation and democratic discourse. This study investigated whether behavioral inconsistency across interaction modalities—what we term “cross-layer behavioral discordance”—could serve as a structural indicator of coordination. Drawing on multi-layer network theory, we hypothesized that authentic users would demonstrate coherent patterns across retweet, reply, and quote networks, while coordinated actors might exhibit discordant behavior reflecting role specialization within influence campaigns.

Analysis of 266,242 Ukraine-related tweets revealed a finding that fundamentally challenged our theoretical expectations: behavioral discordance is not an anomaly but the norm. Fully 80% of users who engage across multiple interaction layers show zero overlap between their amplification targets and conversation partners. More striking still, this pattern intensifies with account maturity—established accounts exhibit significantly higher discordance than newer ones. This inverse relationship disqualifies cross-layer discordance as a coordination signal while illuminating how platform affordances shape the differentiation of communicative practices over time.

We also document how multi-model artificial intelligence review, conducted through three large language models with distinct analytical perspectives, successfully identified the flawed foundational assumption underlying our initial approach—a methodological contribution with implications for how computational communication research might leverage AI collaboration while maintaining empirical rigor.

Keywords: coordinated inauthentic behavior, multi-layer networks, platform affordances, negative results, computational methods, AI-assisted research

Introduction

The integrity of public discourse on social media platforms has emerged as one of the defining challenges of contemporary communication scholarship. From the Internet Research Agency’s

interference in the 2016 U.S. presidential election to coordinated campaigns surrounding Brexit, COVID-19, and numerous national elections worldwide, the manipulation of online information environments through networks of inauthentic accounts has demonstrated both its prevalence and its consequences (Benkler, Faris, & Roberts, 2018; Freelon & Wells, 2020). For communication researchers, the detection and analysis of such coordination represents not merely a technical problem but a fundamental question about the conditions under which authentic public deliberation remains possible in networked media environments.

Yet coordination detection remains methodologically challenging. The most widely adopted approaches rely on content-based signals: tools like CooRnet identify accounts that share identical URLs within narrow temporal windows, revealing coordinated amplification campaigns through their synchronized distribution of specific content (Giglietto, Righetti, Rossi, & Marino, 2020). These methods have proven effective for detecting certain forms of manipulation, but they require content matching and operate at the group level, flagging clusters of accounts rather than identifying individual coordinated actors. The search for complementary detection methods—particularly those that might operate independently of content analysis—remains an active area of inquiry.

This study began with what seemed a theoretically grounded intuition: that the structure of users' engagement patterns across different interaction modalities might reveal something about the authenticity of their participation. Social media platforms like Twitter offer multiple ways of engaging with content and other users—retweeting to amplify, replying to converse, quoting to comment. We reasoned that authentic users, who develop genuine interests and relationships over time, would demonstrate some consistency across these modalities. An account that finds a journalist worth retweeting might also engage that journalist through replies; a user invested in a political community might both amplify and converse with fellow community members. Coordinated accounts, by contrast, designed for single-purpose amplification without genuine social investment, might exhibit what we termed “cross-layer behavioral discordance”—retweeting one set of accounts while conversing (if at all) with an entirely different set.

The findings we report here did not confirm this hypothesis. They refuted it in terms that forced us to reconsider fundamental assumptions about what “normal” social media behavior looks like. But negative findings, though undervalued in publication hierarchies that favor novel positive results, serve essential scientific functions (Fanelli, 2012). They prevent the duplication of failed approaches, establish empirical boundaries for theoretical claims, and sometimes—as in this case—reveal patterns of human behavior that prove more interesting than the anomalies we initially sought to detect.

Theoretical Framework: Multi-Layer Networks and Platform Affordances

The conceptual foundation for this study draws on two streams of scholarship that have increasingly informed computational approaches to communication research. The first is the multi-layer network paradigm, which reconceptualizes social media platforms not as singular networks but as overlapping systems of interaction operating through distinct relational types (Kivelä et al., 2014; Magnani, Rossi, & Vega, 2021). On Twitter, a retweet network captures amplification and endorsement patterns; a reply network maps conversational engagement; a quote network reveals practices

of commentary and reframing. These layers may share nodes—the same user can retweet, reply, and quote—but the relationships they map are not equivalent, and the extent to which users’ positions and behaviors align across layers is an empirical question with theoretical significance.

The second framework concerns platform affordances and their role in shaping communicative practice (boyd, 2010; Bucher & Helmond, 2018). Different features of social media platforms enable and constrain different kinds of communication. Retweeting requires minimal cognitive investment and serves primarily amplification functions; it asks users to endorse and redistribute without elaboration. Replying demands greater engagement and establishes conversational relationship; it positions users as interlocutors rather than merely distributors. Quote-tweeting combines amplification with commentary, allowing users to frame and contextualize others’ content. These affordances do not determine behavior, but they structure the range of communicative possibilities in ways that may encourage differentiation of practice across interaction types.

The intersection of these frameworks suggested our initial hypothesis. If authentic users develop differentiated practices across modalities—retweeting for one set of communicative purposes with one set of partners, conversing for other purposes with other partners—then behavioral consistency across layers might itself be anomalous, potentially indicating the kind of single-purpose automation characteristic of coordinated campaigns. Alternatively, if authentic engagement involves sustained relationships that manifest across multiple interaction types, then discordance—engaging entirely different accounts through different modalities—might signal the superficiality of inauthentic participation.

We operationalized this through a measure we termed Cross-Layer Behavioral Discordance (CLBD): the Jaccard similarity between a user’s retweet targets and conversation targets (combining replies and quotes). Low similarity indicated high discordance; we initially hypothesized this as a potential marker of coordination.

Research Questions

Rather than advancing directional hypotheses, we framed our inquiry through research questions that allowed the data to adjudicate between competing theoretical possibilities:

RQ1: What is the distribution of cross-layer behavioral discordance among Twitter users engaged in political discourse?

RQ2: Does discordance vary systematically with account characteristics that might distinguish authentic from potentially inauthentic users?

RQ3: Can multi-model AI review identify problematic assumptions in computational research design before empirical testing?

Method

Data and Context

We analyzed a corpus of 266,242 tweets discussing the war in Ukraine, collected during a 72-hour window from June 7-9, 2023. This context was selected for several reasons relevant to the study’s

theoretical concerns. The Ukraine conflict has been characterized by substantial information warfare, with documented coordination campaigns by multiple state and non-state actors, making it a context where coordination detection methods would be particularly valuable if effective. The geopolitical salience of the topic ensures substantial participation from diverse user populations, providing variance in user characteristics necessary for baseline analysis. The temporal density of the crisis period allowed construction of meaningful interaction networks within the collection window.

The dataset comprised activity from 102,706 unique users, including 191,566 retweets (72% of total activity), 74,676 original tweets and replies (28%), and 27,657 quote tweets. User metadata included account creation date, follower count, and cumulative tweet count—characteristics we employed to distinguish established from newer accounts in baseline analysis.

Network Construction

Following established protocols for network analysis of political communication (Barberá et al., 2015; Freelon, McIlwain, & Clark, 2020), we constructed three directed networks representing distinct interaction modalities on the platform:

The retweet network (87,267 nodes, 150,581 edges) captured amplification relationships, with directed edges from retweeting users to the original authors whose content they amplified. The reply network (11,153 nodes, 11,504 edges) mapped conversational engagement, connecting replying users to those they addressed. The quote network (9,362 nodes, 10,387 edges) represented commentary relationships, linking users who quoted content to its original authors.

The substantial size differences across layers reflect known patterns in platform affordance usage: retweeting requires minimal effort and serves broadcast functions, attracting far more activity than the more effortful conversational modes.

Measuring Cross-Layer Behavioral Discordance

For the 3,211 users who appeared in multiple network layers and thus had calculable cross-layer patterns, we computed CLBD as the Jaccard coefficient between their retweet targets and their conversation targets (the union of reply and quote targets). This measure ranges from zero—indicating complete discordance, with no overlap between amplification and conversation partners—to one, indicating perfect consistency across modalities.

Baseline Analysis Strategy

The critical methodological move in this study—one prompted by AI review, as we discuss below—was the establishment of baselines stratified by user characteristics. If cross-layer discordance indicates coordination rather than normal behavioral differentiation, we would expect established accounts (older, more followed, more active) to exhibit lower discordance than newer accounts. The logic follows from the assumption that established accounts are more likely to represent authentic, long-term participants while newer accounts may include coordinated actors entering the platform for specific campaigns.

We stratified accounts by age (under 3 months, 3 months to 1 year, 1-3 years, over 3 years), follower count (under 100, 100-1,000, 1,000-10,000, over 10,000), and cumulative activity (under 1,000 tweets,

1,000-10,000, 10,000-100,000, over 100,000). Mann-Whitney U tests compared CLBD distributions across groups.

Multi-Model AI Review Protocol

A distinctive methodological feature of this study was the incorporation of review by multiple large language models prior to finalizing the research design. Following protocols developed through the CommDAAF (Communication Data Analysis Augmentation Framework) initiative, we subjected our methodology to critique by three AI systems: Claude (Anthropic), GLM-4 (Zhipu AI), and Kimi (Moonshot AI). Each was instructed to identify methodological flaws, challenge foundational assumptions, and assess whether findings would warrant publication.

This approach reflects emerging interest in AI-assisted research workflows while raising questions about how such collaboration should be structured to maximize methodological benefit without introducing new sources of bias or error.

Findings

The Prevalence of Behavioral Discordance

Our first finding concerns the sheer prevalence of cross-layer discordance in political discourse on Twitter. Among the 3,211 users with calculable CLBD scores, fully 80.3% demonstrated zero overlap between their retweet targets and conversation partners. The mean CLBD was 0.074, with a median of zero—indicating that the modal Twitter user engaged in Ukraine discourse retweeted one set of accounts and conversed with an entirely different set.

This finding alone gave us pause. If four out of five users exhibit complete behavioral discordance, the pattern cannot meaningfully distinguish coordinated from authentic actors; it simply describes how most people use the platform.

The Inverse Relationship with Account Maturity

The baseline analysis produced findings that directly contradicted our theoretical expectations. When we stratified CLBD by account age, we found that discordance was not merely prevalent but intensified with account maturity. New accounts (under 3 months old) showed a mean CLBD of 0.283, with 53.1% exhibiting zero cross-layer overlap. Accounts between 3 months and one year old showed mean CLBD of 0.078 and 78.3% zero overlap. For accounts over three years old—those most likely to represent established, authentic participants—mean CLBD dropped to 0.060, with 83.5% showing complete behavioral discordance.

This pattern was statistically robust. Mann-Whitney U tests comparing old accounts (over 3 years) to new accounts (under 3 months) confirmed that established accounts exhibit significantly higher discordance, with the test for whether old accounts would show higher consistency (lower discordance) returning $p = 1.000$ —the data ran entirely opposite to hypothesis.

Parallel patterns emerged for other indicators of account establishment. Higher-activity accounts (over 10,000 cumulative tweets) showed greater discordance than lower-activity accounts. The

relationship with follower count was more complex, with both the lowest and highest follower brackets showing somewhat elevated CLBD, but the overall pattern confirmed that the markers we might associate with authentic, invested participation predicted more, not less, behavioral differentiation across interaction layers.

Interpretation: Why Established Users Are More Discordant

These findings demand interpretation. Why would long-term, active, invested Twitter users show greater behavioral inconsistency across interaction modalities than newer arrivals?

We propose that cross-layer discordance reflects the differentiation of communicative practice that develops through sustained platform engagement. Established users learn, over time, that different interaction types serve different purposes and involve different relationships. They retweet journalists, news organizations, and public figures whose content they wish to amplify to their followers. They reply to friends, colleagues, and community members with whom they maintain conversational relationships. They quote-tweet to comment on content for their audiences. The accounts they retweet and the accounts they converse with need not—and typically do not—overlap, because these activities serve distinct functions in their communicative repertoire.

Newer users, by contrast, have not yet developed this differentiation. With smaller networks and less experience with platform affordances, they may engage the same limited set of accounts across modalities. Their higher consistency reflects not authenticity but the undifferentiated engagement characteristic of users still learning how to use the platform.

This interpretation aligns with affordance theory's emphasis on how features shape practice over time (Bucher & Helmond, 2018). It also recalls research on how users develop distinct audiences and self-presentation strategies across platforms and features (Marwick & boyd, 2011). The discordance we initially suspected as anomalous turns out to be a marker of communicative sophistication.

The Contribution of Multi-Model AI Review

The multi-model review process contributed substantially to the study's development, though in ways that illuminate both the potential and limits of AI-assisted research. When we presented our initial methodology—before conducting baseline analysis—to the three AI reviewers, their assessments diverged instructively.

Claude, conducting self-critique, identified the need for baseline validation but assessed it as one concern among several, rating confidence in the foundational assumption at approximately 60%. GLM-4, instructed to adopt a skeptical stance, was considerably harsher. It identified the core premise as “empirically unsupported and likely incorrect,” noting that the assumption of cross-layer consistency for authentic users “is asserted without evidence.” GLM-4 articulated the key insight: “Users RT sources—news outlets, journalists, experts. Users reply to community members, friends, people in conversations. These sets are fundamentally different in organic behavior.”

Kimi, tasked with synthesis, offered a reframing that proved prescient: “CLBD may be detecting user types (broadcasters vs. conversationalists) rather than authenticity (organic

vs. coordinated).”

The divergence across models proved methodologically productive. Had we relied on a single AI review—particularly a more sympathetic one—we might have proceeded without the baseline analysis that ultimately revealed the fatal flaw in our approach. The incorporation of an explicitly skeptical reviewer, primed to challenge assumptions, surfaced concerns that might otherwise have remained implicit until peer review or, worse, post-publication critique.

Discussion

Implications for Coordination Detection Research

The negative finding reported here has direct implications for the broader project of detecting coordinated inauthentic behavior on social media. It demonstrates that intuitive structural signals—patterns that seem, on theoretical grounds, like they should distinguish authentic from inauthentic actors—may simply capture normal behavioral variation. Cross-layer discordance seemed like a reasonable coordination indicator. It was not.

This suggests caution in developing new detection methods. Content-based approaches like CooRnet have demonstrated validity through extensive empirical testing against known coordination campaigns (Giglietto et al., 2020). Novel structural approaches require similar validation against ground truth before deployment. The intuition that “inconsistency equals suspicious” proved wrong for cross-layer behavior; it may prove wrong for other structural patterns as well.

More broadly, coordination detection research must grapple with what we might call the baseline problem. Any claim that a pattern indicates anomalous behavior presupposes knowledge of what normal behavior looks like. Establishing these baselines requires the kind of stratified analysis we conducted here—examining how proposed indicators vary across user types we can independently assess. Without such baselines, we risk flagging normal behavioral variation as suspicious.

Platform Affordances and Communicative Differentiation

While our coordination detection hypothesis failed, the findings illuminate patterns of platform use with independent theoretical interest. The intensification of cross-layer discordance with account maturity suggests that sustained engagement with social media platforms involves a process of communicative differentiation—the development of distinct practices for distinct features, with distinct relational targets.

This pattern resonates with research on context collapse and audience management on social media (Marwick & boyd, 2011; Vitak, 2012). Users learn, over time, that different features reach different audiences and serve different purposes. They develop what amounts to a repertoire of communicative practices adapted to platform affordances. The retweet is not the reply is not the quote-tweet, and experienced users treat them accordingly.

Whether this differentiation should be understood as a positive development—the cultivation of communicative sophistication—or a concerning fragmentation of public discourse depends on normative

frameworks beyond the scope of this study. But the empirical pattern is clear: the longer users engage with Twitter, the more their interaction patterns diverge across modalities.

AI-Assisted Research: Promise and Caution

The multi-model review process that contributed to this study offers a template for AI-assisted methodology validation while also illustrating its limits. The divergence across models—with the skeptical GLM-4 reviewer identifying concerns that the more sympathetic Claude underweighted—suggests that incorporating multiple AI perspectives with distinct analytical orientations can surface blind spots in research design.

Yet AI review cannot replace empirical validation. The models could identify that our foundational assumption was unsupported; they could not determine whether it was correct or incorrect. Only the baseline analysis could adjudicate that question. AI-assisted research workflows should be understood as supplements to, not substitutes for, the empirical testing that remains the foundation of scientific inference.

There is also a question of appropriate disclosure. Studies that incorporate AI assistance in design, analysis, or writing face evolving norms about transparency. We have chosen to document the AI review process in detail, treating the models as collaborators whose contributions merit acknowledgment. This may or may not become standard practice as the role of AI in research continues to evolve.

Limitations

Several limitations constrain the generalizability of these findings. The analysis focused on a single platform (Twitter/X) during a specific crisis context (Ukraine war discourse), and patterns may differ on other platforms or in other topical domains. The 72-hour collection window, while sufficient for network construction, may not capture stable long-term behavioral patterns. Data collection through hashtags may oversample broadcast-oriented behavior, potentially inflating baseline discordance rates. And the study examined only one specific operationalization of cross-layer behavior; alternative measures might yield different results.

Most significantly, the invalidation of CLBD as a coordination signal does not preclude its potential contribution to ensemble detection models that incorporate multiple signals. A pattern that fails as a standalone indicator might still contribute predictive value in combination with other features.

Future Directions

Several extensions of this work merit pursuit. Cross-platform analysis could examine whether discordance patterns hold on other social media platforms with different affordance structures. Longitudinal analysis could track how individual users' cross-layer patterns evolve over time, potentially identifying developmental trajectories. Community-specific baselines could examine whether certain discourse communities—academic Twitter, political activist networks, professional communities—exhibit different normative patterns that would affect interpretation of discordance. And integration with other signals could test whether CLBD contributes to ensemble coordination detection models even when not discriminative alone.

Conclusion

We proposed cross-layer behavioral discordance as a potential indicator of coordinated inauthentic behavior, reasoning that authentic users would demonstrate consistency across interaction modalities while coordinated actors might exhibit discordance reflecting role specialization. Baseline analysis stratified by account characteristics revealed that this hypothesis was wrong: discordance is the norm in political discourse on Twitter, and it intensifies with the markers of authentic, established participation.

This negative finding contributes to coordination detection research by eliminating a plausible but incorrect approach, establishing empirical baselines for cross-layer behavior in political discourse, and demonstrating methodological practices—particularly baseline validation and multi-model review—that can identify flawed assumptions before they propagate into accepted methods. In a field where the stakes of misidentification are high—where false accusations of coordination can silence legitimate voices while missed detection enables manipulation—the careful establishment of what does and does not indicate inauthenticity serves the broader project of preserving spaces for genuine public deliberation.

The finding also reveals something about how people use social media platforms over time. The differentiation of communicative practice across interaction modalities—retweeting one set of accounts, conversing with another—appears not as fragmentation or inconsistency but as a feature of sustained engagement, the development of a repertoire of practices adapted to platform affordances. Whether this differentiation serves or undermines the communicative functions we hope social media might fulfill remains a question for ongoing inquiry.

References

- Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A., & Bonneau, R. (2015). Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological Science*, 26(10), 1531-1542.
- Benkler, Y., Faris, R., & Roberts, H. (2018). *Network propaganda: Manipulation, disinformation, and radicalization in American politics*. Oxford University Press.
- boyd, d. (2010). Social network sites as networked publics: Affordances, dynamics, and implications. In Z. Papacharissi (Ed.), *A networked self: Identity, community, and culture on social network sites* (pp. 39-58). Routledge.
- Bucher, T., & Helmond, A. (2018). The affordances of social media platforms. In J. Burgess, A. Marwick, & T. Poell (Eds.), *The SAGE handbook of social media* (pp. 233-253). SAGE.
- Fanelli, D. (2012). Negative results are disappearing from most disciplines and countries. *Scientometrics*, 90(3), 891-904.
- Freelon, D., McIlwain, C. D., & Clark, M. D. (2020). *Beyond the hashtags: #Ferguson, #Blacklivesmatter, and the online struggle for offline justice*. Center for Media & Social Impact.
- Freelon, D., & Wells, C. (2020). Disinformation as political communication. *Political Communication*, 37(2), 145-156.

Giglietto, F., Righetti, N., Rossi, L., & Marino, G. (2020). It takes a village to manipulate the media: Coordinated link sharing behavior during 2018 and 2019 Italian elections. *Information, Communication & Society*, 23(6), 867-891.

Kivelä, M., Arenas, A., Barthelemy, M., Gleeson, J. P., Moreno, Y., & Porter, M. A. (2014). Multilayer networks. *Journal of Complex Networks*, 2(3), 203-271.

Magnani, M., Rossi, L., & Vega, D. (2021). Analysis of multiplex social networks with R. *Journal of Statistical Software*, 98(8), 1-30.

Marwick, A. E., & boyd, d. (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*, 13(1), 114-133.

Vitak, J. (2012). The impact of context collapse and privacy on social network site disclosures. *Journal of Broadcasting & Electronic Media*, 56(4), 451-470.