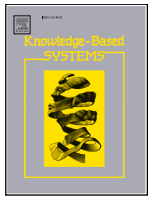
Knowledge-Based Systems 334 (2026) 115103

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys



Integrated multilayer reinforcement model: Explaining the dynamics of online radicalization

Enrico Corradini [a,1,*], Francesco Cauteruccio [b,1]

[a] DII, Polytechnic University of Marche, Italy
[b] DIEM, University of Salerno, Italy

ARTICLE INFO

Keywords:
Online radicalization
Multilayer networks
Composite reinforced centrality
Temporal burst influence
Sentiment synchronization

ABSTRACT

Online social platforms have become a fertile ground for the rapid spread of extremist narratives, yet traditional single-layer network analyses overlook the interplay of content, timing, and polarization that fuels radicalization. We introduce the Integrated Multilayer Reinforcement Model (IMRM), a unified framework that represents each user as a node in four interdependent layers (interaction, content similarity, temporal dynamics, and sentiment) and explicitly encodes feedback loops via uniform interlayer coupling. We define three novel measures: Composite Reinforced Centrality (CRC), which multiplicatively aggregates a user's normalized influence across layers; Temporal Burst Influence (TBI), which captures episodic surges in activity; and Sentiment Synchronization Coefficient (SSC), which quantifies emotional alignment with peers. We derive four theoretical propositions linking these measures to radicalization processes and validate them on Reddit data surrounding five major United States socio-political events. Our experiments reveal that *(i)* high CRC users have a high probability of being radicalized, *(ii)* radicalized users have higher values of both TBI and SSC, *(iii)* bridging nodes linking different communities exhibit elevated CRC, and *(iv)* CRC remains a robust, context-invariant predictor of radical engagement. The findings aim at highlighting how online radicalization emerges from the synergistic fusion of who you interact with, what you share, when you act, and how you feel.

1. Introduction

The proliferation of social media platforms over the past decade has fundamentally reshaped how individuals form communities, exchange ideas, and mobilize around shared beliefs. While these online environments can foster healthy debate and collective action, they have also become fertile ground for the spread of extremist narratives and radicalization. In particular, the phenomenon of clusters, in which users predominantly interact with like-minded peers, has been shown to amplify ideological content and reinforce preexisting biases [1–5]. Traditional Social Network Analysis (SNA) methods, however, have largely relied on single-layer representations of these interactions, modeling each user as a node and each reply, mention, or friendship as an edge [6–12]. Although such approaches have proven valuable for mapping broad patterns of connectivity and influence, they fall short when tasked with explaining the complex, multi-faceted dynamics that drive online radicalization.

First, single-layer SNA treats all edges as interchangeable, regardless of the content or context of the interaction. In the case of radical movements, the substance of messages, ranging from provocative slogans to detailed conspiratorial narratives, plays a decisive role in persuading and mobilizing individuals. By collapsing relationships into undifferentiated links, traditional models overlook how thematic alignment catalyzes cohesion and facilitates the diffusion of extremist ideologies. Moreover, radicalization is rarely a smooth, continuous process; rather, it typically unfolds in intense bursts of activity around salient political events or triggering incidents [13–15]. Static or heavily aggregated network snapshots obscure these temporal surges, preventing researchers from detecting the synchronization of mobilizing calls to action and the rapid consolidation of radical communities.

Equally important is the emotional dimension of online discourse [16–18]. Anger, fear, and solidarity, conveyed through sentiment-laden language, can galvanize users far more effectively than neutral exchanges. Single-layer frameworks, which ignore affective tone, miss the reinforcing feedback loops whereby emotional contagion deepens commitment and intensifies group cohesion. Finally, individuals often participate across multiple behavioral dimensions simultaneously [19–21]: they comment, they share thematic content, they react strongly during

* Corresponding author.
E-mail addresses: e.corradini@univpm.it (E. Corradini), fcauteruccio@unisa.it (F. Cauteruccio).
[1] Enrico Corradini and Francesco Cauteruccio have contributed equally to this work.

https://doi.org/10.1016/j.knosys.2025.115103
Received 11 June 2025; Received in revised form 9 September 2025; Accepted 6 December 2025
Available online 13 December 2025

heated moments, and they express emotions that resonate with peers. Capturing all these interactions demands a multilayer perspective, one that preserves content, timing, and sentiment as distinct but interconnected layers of influence.

Despite the promise of multilayer approaches, existing studies of online radicalization have generally focused on only one or two behavioral dimensions at a time, whether interaction patterns or sentiment flows, and have treated them in isolation [22,23]. This fragmented treatment obscures the holistic effects by which content, timing, and emotion jointly propel individuals toward extremist viewpoints. Our central problem is therefore to develop a unified network framework that can simultaneously capture all relevant dimensions of online behavior and quantify their combined influence on radicalization processes.

The main contributions of our paper are fourfold. First, we introduce the Integrated Multilayer Reinforcement Model (IMRM), a comprehensive formalism that represents each user as a node in four interdependent layers and that explicitly encodes the feedback loops by which activity in one layer both drives and is driven by activity in the others. Second, we define three novel multilayer SNA measures, namely, *(i)* Composite Reinforced Centrality (CRC), *(ii)* Temporal Burst Influence (TBI), and *(iii)* Sentiment Synchronization Coefficient (SSC), which enact the concept of multi-dimensional reinforcement and yield interpretable, normalized scores reflecting a user's propensity toward radicalization. Third, we formulate four theoretical propositions that link these measures to the emergence, propagation, and potential intervention of extremist behavior. Finally, we validate the entire framework through experiments on Reddit data around five major U.S. socio-political events (2008 Elections, 2011 Occupy Wall Street, 2016 Elections, 2017 Charlottesville Rally, 2021 Capitol Riot), demonstrating that each of our four propositions holds consistently across these diverse contexts. Together, these contributions advance both the theoretical understanding and empirical practice of modeling online radicalization via a truly holistic, multilayer approach.

The paper is structured as follows. In Section 2 we present an overview of related literature. In Section 3 we formalize the IMRM model, the novel SNA measures and the four theoritical propositions and hypotheses. Moreover, in Section 4 we formulate four different theoretical propositions and hypotheses. In Section 5 we present our experimental campaign on Reddit data. Subsequently, in Section 6, we provide a detailed discussion of our framework and its limitations. Finally, in Section 7 we discuss the results obtained and draw some conclusions and possible future works.

2. Related literature

Social Network Analysis (SNA) has long provided the methodological backbone for studying online platforms, with classic concepts such as centrality and modularity continuing to underpin the analysis of social media today [24–27]. Modern research regularly employs these foundational tools to identify influential actors (using degree, PageRank, or activity-based influence scores) and to detect cohesive communities via clustering or modularity-based algorithms [24,28–30]. For instance, the authors of Logan et al. [31] applied PageRank and modularity-based community detection to a multilayer Twitter network, demonstrating that classic SNA metrics can effectively reveal both key influencers and group structure at scale. On Reddit, the authors of Sawicki et al. [32] built a large-scale network of subreddits connected by crossposts, measuring degree distributions and clustering, and employing community detection algorithms to uncover both expected and novel relationships among groups. The search for influential actors is further advanced in Hasan et al. [33], which proposed a behavioral influence ranking (BRI) method for Twitter that uses posting frequency and interaction patterns, outperforming traditional centrality measures. Community detection research has also embraced machine learning: the authors of Nooribakhsh et al. [34] provided a systematic review showing that unsupervised methods and deep learning models are increasingly used for community discovery in large social networks, though often evaluated with classic modularity and clustering benchmarks. Finally, the authors of Choi et al. [35] analyzed YouTube comment networks to identify polarized communities, combining SNA with content analysis to flag coordinated or suspicious behavior. Together, these recent studies illustrate that the core tools of SNA, centrality for influence and clustering for community, remain highly relevant for making sense of modern, large-scale, and dynamic online social networks.

Online radicalization and extremist behavior have been intensively studied in recent years, with researchers examining how extremist ideas diffuse and how online platforms may facilitate radicalization [22,23,36–38]. A consistent finding is that social media can create closed or insular communities that reinforce users' pre-existing beliefs [39,40]. For example, Cinelli et al. [39] compared user interactions across major platforms and showed that each platform's design influences the formation of echo chambers and polarization differently, with significant differences in homophily and information diffusion bias between Twitter, Facebook, Gab, and Reddit. However, empirical studies increasingly reveal that radicalization is rarely confined to a single platform. Instead, today's extremist communities often form multi-platform ecosystems, and interventions targeting one platform can have unintended consequences elsewhere. The authors of Ribeiro et al. [40] analyzed extremist groups banned from Reddit that reconstituted on standalone websites, finding that while bans decreased activity, the remaining users on fringe sites became more toxic and radicalized. Similarly, Russo et al. [41] tracked users migrating from banned subreddits to fringe platforms, demonstrating that cross-platform participation increased users' toxicity even on mainstream Reddit, revealing "spillover" effects. Complementing this, Buntain et al. [42] studied the aftermath of the January 6th deplatforming, showing that while hate speech decreased on mainstream sites, toxic content surged on alternative platforms like Gab, amplifying extremism in new venues. These findings highlight a critical limitation of earlier single-layer studies: focusing on one dimension of user behavior in isolation can miss how content, timing, sentiment, and interactions jointly fuel radicalization. This underscores the need for integrated, multi-layer approaches in computational social science.

A growing body of recent research leverages multilayer network models to capture the complex dynamics of online extremism and radicalization, recognizing that single-layer or single-platform approaches miss critical interdependencies [15,43–45]. For instance, the authors of Wippell and Haynie [46] employed a multiplex network approach to study the Proud Boys, integrating layers for chapter co-affiliation, social-media connections, and co-membership in other extremist groups. Their work revealed that cross-layer bridges can significantly influence group mobilization and participation in real-world extremist actions. However, their focus was primarily on organizational ties and structural overlap, whereas our IMRM incorporates not just multiple relational layers but also content, sentiment, and temporal dynamics, thus capturing more behavioral nuance. In [47], the authors advanced the field with a socio-semantic multilayer model that links actors both by their digital communication and by ideological content similarity. This framework enabled mapping of far-right ecosystems in greater semantic detail, but it is largely descriptive and did not provide mechanisms to model dynamic reinforcement or feedback effects across layers, a gap addressed by our reinforcement-driven IMRM. In a related vein, Baele et al. [48] showed that the online incelosphere is a "highly dynamic and multi-layered environment", tracing the migration and escalation of violent rhetoric across platforms and over time. While their cross-platform temporal lens demonstrated the necessity of multilayer analysis, it did not formally model the mutual reinforcement of radicalization factors as our framework does. The authors of Peralta et al. [49] constructed a two-layer Twitter network connecting politicians and ordinary users, analyzing how polarization and echo chambers form across interaction types; however, they restricted their model to structural and role-based layers, not integrating content or sentiment as behavioral drivers. In [50], the

authors compared structural interaction networks with language-based opinion networks, showing that semantic and social ties can reinforce each other to entrench echo chambers. Yet, their approach kept structural and semantic networks analytically separate, while IMRM unifies such information in a single multilayer architecture with explicit interlayer coupling. Collectively, these studies illustrated the trend toward richer, multilayer modeling, but they typically address either structural or content dimensions in isolation or lack mechanisms to represent cross-layer reinforcement and feedback. The proposed IMRM advances this research frontier by tightly integrating structural, content, temporal, and affective dimensions, with explicit modeling of how engagement in one layer feeds back to reinforce radicalization processes in others. This holistic integration provides new theoretical and empirical leverage to analyze not only how extremist communities are structured but also how they grow and persist through dynamic, cross-layer reinforcement.

3. The integrated multilayer reinforcement model (IMRM)

We devote this section to presenting our Integrated Multilayer Reinforcement Model (IMRM). Particularly, the structure of this section is as follows. We first define the model and its core components, along with the rationale behind them. Then, we define the mechanisms of reinforcement the model encompasses and their utility in the study of different phenomena in Online Social Networks (OSNs from now on). Finally, we present some novel measures formally defined by the combination of the model and the reinforcement mechanisms.

3.1. Core components of IMRM

The main aim of IMRM is to comprehensively capture the different dynamics typically of OSNs, such as online radicalization and echo chamber formation. To do so, we resort to the expressive power of multilayer networks, i.e., network structures that allow the modeling of multiple types of interactions or relationships among nodes. Unlike to what happens when using traditional single-layer network models, IMRM represents the online ecosystem induced by a OSN as a multilayer network with four layers, namely: *(i)* User Interaction Layer (UIL), *(ii)* Content Similarity Layer (CSL), *(iii)* Temporal Dynamics Layer (TDL), and *(iv)* Affective and Sentiment Layer (ASL). Here, each layer captures and represents, via its edges, a distinct dimension of *user behavior*, briefly described as follows:

1. UIL, which models social interactions between users of the network, such as replies and mentions;
2. CSL, which captures thematic convergence and shared narratives between users, extracted and inferred from the content;
3. TDL, which models the timing, bursts, and event-triggered nature of interactions;
4. ASL, which models the emotional tones and sentiment flows embedded in the interactions.

Being represented by a multilayer network, IMRM also includes interlayer edges, that is, links that connect the same user across different layers, thereby effectively integrating the diverse behavioral dimensions into a unified framework.

We are now able to formally define the model. Let $\mathcal{U}$ be the set of all nodes in the network, where each node $u_i \in \mathcal{U}$ is associated with a user in the OSN. The OSN consists of a number of interactions $T > 0$ between users, with these interactions being organized in timestamps. Given two users $u_i, u_j \in \mathcal{U}$, we denote with $1 \leq t_{ij}^{(k)} \leq T$ the kth interaction between u_i and u_j. Here, we deliberately give a general definition of interaction since its inherent nature depends on the OSN of study; in presenting our model, we assume an interaction between two users may produce textual content, e.g., replying to a user with text, or not, e.g., mentioning or retweeting a user. Let $\mathcal{L} = \{\mathsf{UIL}, \mathsf{CSL}, \mathsf{TDL}, \mathsf{ASL}\}$ be the set of behavioral dimension. In Eq. (3.1), we represent each dimension $L \in \mathcal{L}$ as a (possibly weighted) network

$$G^{(L)} = (\mathcal{U}, E^{(L)}, \omega^{(L)}) \tag{3.1}$$

where $\mathcal{U}$ is the set of nodes (users) of the network, $E^{(L)}$ is the set of edges representing the relationships captured within the layer L, and $\omega^{(L)}$ is a weight function associating a numerical value to each edge of the network. For instance, $G^{(\mathsf{UIL})} = (\mathcal{U}, E^{(\mathsf{UIL})}, \omega^{(\mathsf{UIL})})$ is the network associated with the UIL: particularly, each edge $e_{ij} \in E^{(\mathsf{UIL})}$ indicates a social interaction between users $u_i, u_j \in \mathcal{U}$, and $\omega^{(\mathsf{UIL})}(e_{ij})$, simply denoted as $\omega_{ij}^{(\mathsf{UIL})}$, quantifies such interaction. Note that a relationship is specific to its corresponding dimension, that is, if $e_{ij} \in E^{(L_1)}$, this does not imply that $e_{ij} \in E^{(L_2)}$, where $L_1, L_2 \in \mathcal{L}$ and $L_1 \neq L_2$.

Given all layers, in Eq. (3.2) we denote and define the complete multilayer network representing our model as

$$\mathfrak{M} = \left\{ G^{(L)} = \left(\mathcal{U}, E^{(L)}, \omega^{(L)}\right) \;\middle|\; L \in \mathcal{L} \right\} \cup \left\{ E^{(\text{inter})} \right\}. \tag{3.2}$$

Here, $E^{(\text{inter})}$ denotes the set of interlayer edges that connect the same user across different layers. Formally, an interlayer edge is defined for every node $u_i \in \mathcal{U}$ (associated to a single user) and for every pair of distinct layers $L_1, L_2 \in \mathcal{L}$, $L_1 \neq L_2$, connecting node u_i in layer L_1 with node u_i in layer L_2. This ensures that the representation of the same user, for each user, is consistent and interconnected across all behavioral dimensions. Notably, $\mathfrak{M}$ is also called a multiplex network [51]. A graphical depiction of the IMRM is reported in Fig. 1.

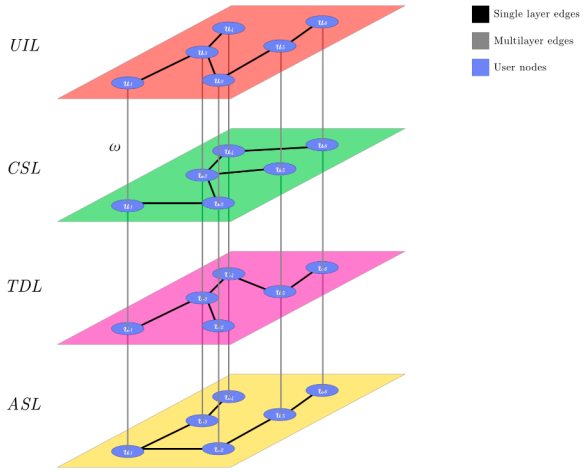


Fig. 1. A graphical depiction of IMRM.

As it is clear from the previous definition, both $E^{(L)}$ and $\omega^{(L)}$ are strictly determined by the definition of the corresponding layer L. This design makes the overall model flexible and expressive enough to capture the structure of different online ecosystems, thus enabling various types of analysis. In what follows, we formally describe for each layer L its corresponding $G^{(L)}$.

3.1.1. User interaction layer - UIL

The UIL is a network capturing the interactions among users. Here, we deliberately maintain a general definition of interactions, which can accommodate different paradigms. For instance, in an OSN such as X (formerly known as Twitter), an interaction could represent a mention, while in Reddit an interaction could represent a reply to a comment [52]. This network is defined in Eq. (3.3) as

$$G^{(\mathsf{UIL})} = \left(\mathcal{U}, E^{(\mathsf{UIL})}, \omega^{(\mathsf{UIL})}\right), \tag{3.3}$$

where $E^{(\mathsf{UIL})} \subseteq \{\{u_i, u_j\} \mid u_i, u_j \in \mathcal{U}, u_i \neq u_j\}\}$, and $\omega^{(\mathsf{UIL})} : E^{\mathsf{UIL}} \to \mathbb{N}^+$ quantifies the interaction intensity. In our case, given an edge $e_{ij} \in E^{(\mathsf{UIL})}$, we define $\omega_{ij}^{(\mathsf{UIL})}$ as the number of interactions from user u_i to user u_j. Note that, in our case, we assume that $G^{(\mathsf{UIL})}$ is undirected, that

is, we are not interested in the possible direction of the interactions. Nevertheless, the model is general enough to encompass this information when needed.

3.1.2. Content similarity layer - CSL

The CSL is a network encompassing the similarity between the content, e.g., posts and comments, users interact to and with within the overall ecosystem. In particular, it is an undirected network defined in Eq. (3.4) as

$$G^{(\mathsf{CSL})} = \left(\mathcal{U}, E^{(\mathsf{CSL})}, \omega^{(\mathsf{CSL})}\right). \tag{3.4}$$

To enable a more detailed representation of the content and the narrative induced by it, we assume that each user $u_i \in \mathcal{U}$ is associated with a content attribute vector $\mathbf{c}_i \in \mathbb{R}^d, d > 0$. Here, $\mathbf{c}_i$ is a d-dimensional vector derived from the aggregated data the user shared and interacted with within the OSN; generally, this vector can be obtained by aggregating all textual contributions of user u_i, such as posts and comments, into a single representation. Conceptually, $\mathbf{c}_i$ encodes the semantic profile of the user's discourse, intending to capture recurrent topics, linguistic patterns, and narrative alignments. In practice, the aggregation can be derived by exploiting different data mining and text extraction techniques, such as TF-IDF, word embeddings, and topic modelling, as commonly adopted in literature [13,53]. In our experimental campaign, as we will illustrate in Section 5.5, we adopt a state-of-the-art pre-trained sentence embedding model, which provides semantically enriched embeddings that capture contextual meaning beyond surface-level keywords, thus allowing the CSL layer to reflect thematic convergence rather than lexical overlap between users' contents. Eventually, given two users $u_i, u_j \in \mathcal{U}$, we can simply quantify the similarity between them using a similarity measure on the associated content attribute vectors $\mathbf{c}_i$ and $\mathbf{c}_j$. In our setting, we resort to the classical cosine similarity measure, defined in Eq. (3.5) as

$$s_{ij} = \frac{\mathbf{c}_i \cdot \mathbf{c}_j}{\|\mathbf{c}_i\| \, \|\mathbf{c}_j\|}. \tag{3.5}$$

Note that, in principle, the similarity score s_{ij} can be computed for every pair of users $u_i, u_j \in \mathcal{U}$. However, many of these scores might be very low, reflecting incidental or non-meaningful similarity, rather than a substantive shared narrative or thematic convergence. To enable a more fine-grained analysis, we filter this noise by introducing a threshold parameter τ_c which serves two main purposes, that is, noise and network sparsification reductions. First off, it avoids considering weak similarities that do not represent significant thematic overlap. At the same time, it reduces the density of the network by retaining only the more meaningful edges, enhancing both interpretability and computational efficiency. τ_c can be computed via different methods such as quantile-based selection and statistical significance [54]. In the first case, one common method is to set τ_c at a specific quantile, e.g., the 75th or 90th percentile of the distribution of all similarity scores $\{s_{ij}\}$. This approach ensures that only the top fraction of pairwise similarities is retained. In the second case, τ_c can be determined through statistical testing by comparing the observed similarity scores against a null model of random similarity, retaining only those scores that are significantly higher than what would be expected by chance. Nonetheless, τ_c can be defined to accommodate different kinds of scenarios according to the analyses of interest.

Given τ_c, the edge set of $G^{(\mathsf{CSL})}$ is then defined in Eq. (3.6) as

$$E^{(\mathsf{CSL})} = \{\{u_i, u_j\} \mid u_i, u_j \in \mathcal{U}, u_i \neq u_j, s_{ij} > \tau_c\}, \tag{3.6}$$

that is, the edges in CSL are only the ones connecting a pair of users u_i and u_j such that their similarity is greater than τ_c. Consequently, to each of these edges we assign the similarity value, that is defined in Eq. (3.7) as

$$\omega_{ij}^{(\mathsf{CSL})} = s_{ij}. \tag{3.7}$$

3.1.3. Temporal dynamics layer - TDL

The TDL incorporates the time aspect of interactions. Indeed, it has been shown that different phenomena in social networks, such as radicalization, often do not unfold gradually, but they accelerate during key events, e.g., elections, pandemics, and protests [55,56]. These moments are usually characterized by episodic surges in user activity, which we aim to capture with the TDL by identifying and quantifying temporal bursts in order to understand whether users get drawn into extremist discourse or interactions. This is also in line with studies on synchrony between groups that tend to radicalize, where the detection of patterns of burst synchronization is of utmost importance.

Also in this case, the TDL is represented as the undirected network defined in Eq. (3.8) as

$$G^{(\mathsf{TDL})} = \left(\mathcal{U}, E^{(\mathsf{TDL})}, \omega^{(\mathsf{TDL})}\right). \tag{3.8}$$

Our aim is to define the TDL by taking into account some desiderata, namely, *(i)* the possibility of quantifying activity bursts between users, *(ii)* the possibility of enabling time-sensitive analyses, and *(iii)* the possibility of coupling this with other behavioral layers to simulate feedback-driven escalation. To do so, we propose the use of timestamped interactions between two users $u_i, u_j \in \mathcal{U}$ and their aggregation over time windows, and in particular, in Eq. (3.9) we introduce the time-dependent interaction function

$$f_{ij}(t) = \sum_k \delta(t - t_{ij}^{(k)}) \tag{3.9}$$

where $0 < t \leq T$ is a given timestamp, $t_{ij}^{(k)}$ is the timestamp denoting the kth interaction between u_i and u_j, and $\delta(\cdot)$ is the Dirac delta function [57]. Intuitively, $f_{ij}(t)$ models all interactions between u_i and u_j, and encodes when these interactions happened. In particular, $f_{ij}(t)$ is defined to be 0 at all times except for the specific timestamps when u_i and u_j interact. To fully capture how often and when users interact, we define the edge weights of TDL upon $f_{ij}(t)$. More formally, given a timestamp t and an offset Δt, the weight of the edge e_{ij} quantifies how many interactions occurred within the time window $[t, t + \Delta t]$. Formally, we define it in Eq. (3.10) as

$$\omega_{ij}^{(\mathsf{TDL})} = \int_t^{t+\Delta t} f_{ij}(t') \, dt'. \tag{3.10}$$

Essentially, with $\omega_{ij}^{(\mathsf{TDL})}$ we count interactions in a sliding time window, and by doing so we aim to provide a reflection regarding burstiness or temporal clustering of interactions. Indeed, by integrating $f_{ij}(\cdot)$ in $\omega_{ij}^{(\mathsf{TDL})}$, we try to give an answer to the question *"do users interact in bursts?"*, which is crucial for detecting temporal signs of different phenomena, such as online radicalization.

3.1.4. Affective and sentiment layer - ASL

The last layer of $\mathfrak{M}$ is ASL. The aim of this layer is to capture the emotional dynamics within the OSN. Similarly to the CSL, which identifies thematic similarity, this layer enables the identification of emotional alignment, that is, it helps identify users expressing similar sentiments or affect. We define it in Eq. (3.11) as an undirected network

$$G^{(\mathsf{ASL})} = \left(\mathcal{U}, E^{(\mathsf{ASL})}, \omega^{(\mathsf{ASL})}\right). \tag{3.11}$$

Here, similarly to the CSL, we assume that each user $u_i \in \mathcal{U}$ is additionally associated with an aggregated affective vector $\mathbf{a}_i \in \mathbb{R}^g, g > 0$. This vector encompasses the affective signals, dimensions such as valence, arousal, dominance, and the sentimental signals, such as joy, fear, and happiness, extracted from the content u_i has shared and interacted with. The aggregated affective vector can be extracted via different techniques and methods, following the plethora of studies on the matter [55,58,59]. In Section 5, we will detail the methods selected for our experiments. Moreover, we are interested in studying the similarity between two users u_i and u_j in terms of $\mathbf{a}_i$ and $\mathbf{a}_j$, respectively. Also in this case, we resort to the classical cosine similarity measure, which we refer here to as a_{ij}. Indeed, similarly to the rationale described for the

CSL, to keep the network informative, we rely on a threshold to reduce noise, that is, to maintain meaningful similarities only. To do so, we define the threshold parameter τ_a which mimics the behavior of τ_c but for the layer ASL. Also in this case, τ_a can be defined to accommodate different kinds of scenarios, as we will show in our Experiments.

Given τ_a, the edge set of $G^{(\mathsf{ASL})}$ is then defined in Eq. (3.12) as

$$E^{(\mathsf{ASL})} = \{\{u_i, u_j\} \mid u_i, u_j \in \mathcal{U}, u_i \neq u_j, a_{ij} > \tau_a\}, \tag{3.12}$$

that is, the intralayer edges of ASL are only the ones connecting a pair of distinct users u_i and u_j such that their similarity, w.r.t. $\mathbf{a}_i$ and $\mathbf{a}_j$, respectively, is greater than τ_a. Eventually, for each edge $e_{ij} \in E^{(\mathsf{CSL})}$, we assign the similarity value via the edge weight function, defined in Eq. (3.13) as

$$\omega_{ij}^{(\mathsf{ASL})} = a_{ij}. \tag{3.13}$$

3.1.5. Interlayer edges

Being represented by a multilayer network, our model $\mathfrak{M}$ also includes interlayer edges, that is, edges connecting nodes from different layers. Specifically, in our context, interlayer edges connect the same user across different layers. Formally, we define the set of interlayer edges in Eq. (3.14) as

$$E^{(\text{inter})} = \{\{(u_i, L), (u_i, L')\} \mid u_i \in \mathcal{U} \text{ and } L, L' \in \mathcal{L}, L \neq L'\}. \tag{3.14}$$

Pragmatically speaking, there is an interlayer edge connecting the same user u_i across every pair of distinct layers $L, L' \in \mathcal{L}$. To each edge in E^{inter} we also assign a uniform coupling weight $\gamma \in \mathbb{R}_{>0}$, which quantifies the strength of the connection between a user's representations in different behavioral dimensions. By defining γ as the same for each interlayer edge, we intend that every connection linking a user across different dimensions is equally weighted. In this case, γ is not computed individually for each edge, but can be treated as a parameter of the model, that is, the value of γ can be determined based on prior domain knowledge or by employing empirical methods such as grid search or sensitivity analysis. The rationale behind γ is as follows. Suppose that in our study intralayer interactions, such as the ones in UIL or CSL, have weights in the range $[0, 10]$. If γ assigns a value of 2.0, then the interlayer coupling is relatively strong compared to some intralayer interactions, enforcing a tight integration across dimensions. Conversely, if γ assigns a value of 0.5, the interlayer connections are weaker, allowing each layer to retain more of its independent characteristics. Consequently, adjusting γ thus provides a mechanism to balance the influence of cross-layer integration versus layer-specific dynamics.

3.2. Mechanisms of multilayer reinforcement

The IMRM posits that phenomena such as online radicalization emerge not from isolated influences in a single dimension, but rather from the simultaneous and interdependent interactions across multiple behavioral dimensions. In this section, we elucidate two primary mechanisms of reinforcement enabled by our model, namely, *(i)* amplification through simultaneous influences across layers, and *(ii)* feedback loops and interdependencies among layers. In the following, with the term reinforcement we intend the mechanisms by which activity in one layer boosts or amplifies activity in others.

3.2.1. Amplification through simultaneous influences

In our model, we claim that a user's propensity toward radicalization, or other similar phenomena, is driven by their engagement in four distinct dimensions, namely, *(i)* direct interactions, represented by UIL; *(ii)* thematic similarity, represented by CSL; *(iii)* temporal dynamics, represented by TDL; *(iv)* and affective signals, represented by ASL. Rather than combining these influences additively, we model their interaction as a multiplicative process, which in turn leads to an amplification effect.

We now formally define this process. Given a user $u_i \in \mathcal{U}$ and a layer $L \in \mathcal{L}$, we define with $\phi_i^{(L)}$ a quantitative measure of user u_i's activity in layer L. Here, with this measure, we indicate a quantitative signal of the membership of the user in that particular layer. For instance, $\phi_i^{(L)}$ could represent the degree or the betweenness centrality of the node u_i in the network $G^{(L)}$ [54]. In our setting, we define $\phi_i^{(L)}$ as the sum of edge weights of u_i in layer L. However, the original distribution of edge weights could exhibit a heavy-tailed shape, thus being unsuitable for different analyses. In light of this, for each $e_{ij} \in E^{(L)}$, we define the transformation in Eq. (3.15) as

$$\tilde{\omega}_{ij} = \log(1 + \omega_{ij}). \tag{3.15}$$

At this point, we can define the effective activity in layer L for a node u_i in Eq. (3.16) as

$$\phi_i^{(L)} = \sum_{e_{ij} \in E_i^{(L)}} \tilde{\omega}_{ij} \tag{3.16}$$

where $E_i^{(L)} \subseteq E^{(L)}$ is the subset of edges in $G^{(L)}$ connecting u_i. Note that such a transformation penalizes extremely high weights by inducing sublinear growth, which helps us defining the overall reinforcement factor for user u_i in Eq. (3.17) as

$$R(u_i) = \prod_{L \in \mathcal{L}} \left(1 + \beta_L \, \phi_i^{(L)}\right), \tag{3.17}$$

where $\beta_L > 0$ is a weighting user-defined parameter reflecting the relative contribution of layer L to the radicalization process. It is interesting to observe that although the trend remains increasing with higher activity, the transformation moderates the magnitude of the increase, ensuring that $R(u_i)$ does not grow excessively.

Furthermore, note that the definition of $R(u_i)$ takes into account all layers in $\mathcal{L}$. Thus, we believe it is important to better understand the rationale behind it. First off, we note that the quantities $\phi_i^{(L)}$ capture different behavioral dimensions depending on the layer L. For instance, when $L = \mathsf{UIL}$, it captures interaction intensity, while when $L = \mathsf{TDL}$, then it captures temporal burstiness. Therefore, these quantities have inherently heterogeneous raw meaning. To render them comparable, we apply a log-based transformation (see Eq. (3.15)), which allows us to work with bounded values across layers, preventing high edge weights from dominating. Interestingly, we also point out that the multiplicative formulation of $R(u_i)$ is intended to capture the synergistic nature of reinforcement, rather than representing a product between physical features. Indeed, we believe that activity across multiple layers jointly amplifies radicalization risk more strongly than a simple additive combination would. Therefore, in this sense, the multiplicative formulation in $R(u_i)$ serves as a method to capture cross-layer synergy, where the simultaneous presence of high values in several dimensions could yield a disproportionately large reinforcement effect.

3.2.2. Feedback loops and interdependencies among layers

Another key feature of IMRM lies in its multilayer structure, in which interlayer edges, especially when a uniform coupling weight γ is chosen, connect the same user across different layers. The design of such a structure facilitates dynamic feedback loops, that is, an increase in activity in one layer propagates via interlayer coupling to reinforce activity in the other layers. For example, if user u_i experiences a surge in activity in the Temporal Dynamics Layer (TDL) due to an external event, e.g., a temporal burst of interactions induced by some content shared within the network, then this increase may lead to different dynamics, such as:

- more frequent direct interactions in the User Interaction Layer (UIL);
- enhanced thematic convergence in the Content Similarity Layer (CSL);
- and intensified affective signals in the Affective and Sentiment Layer (ASL).

Indeed, to capture these cross-layer effects would be useful to extrapolate different insights from the dynamics occurring within the network. Therefore, to do so, we define an effective influence measure for each

layer that incorporates both the direct activity in that layer and the contributions from the other layers. We denote and define it in Eq. (3.18) as

$$\tilde{\phi}_i^{(L)} = \phi_i^{(L)} + \gamma \sum_{L' \in \mathcal{L}, L' \neq L} \phi_i^{(L')}. \tag{3.18}$$

Essentially, $\tilde{\phi}_i^{(L)}$ reflects how activity in one dimension is amplified by simultaneous engagement in the others, thus constituting an effective way of capturing feedback loops and interdependencies among layers.

3.3. Deriving novel measures from IMRM

While the aforementioned mechanisms are generally useful even when used *as it is*, we can specifically leverage them to define novel measures to capture complex dynamics of different phenomena on OSN. In our setting, we are interested in online radicalization, and to effectively capture it we propose three novel measures based on our proposed framework:

1. **Composite Reinforced Centrality** (CRC) - This measure aggregates a user's effective influence across all layers via a multiplicative formulation;
2. **Temporal Burst Influence** (TBI) - This measure captures the impact of temporally clustered activity by evaluating the ratio of peak to total interactions over a specified time window.
3. **Sentiment Synchronization Coefficient** (SSC) - This coefficient quantifies emotional resonance by computing the average similarity, e.g., cosine similarity, between a user's aggregated affective vector and those of its neighbors.

In what follows, we formally introduce and define these three novel measures.

3.3.1. Composite reinforced centrality (CRC)

We define the CRC as a normalized measure that aggregates a user's influence across multiple behavioral layers.

Let $u_i \in \mathcal{U}$ be a user, then we define in Eq. (3.19) the normalized effective influence in layer L as

$$\bar{\phi}_i^{(L)} = \frac{\tilde{\phi}_i^{(L)}}{|\mathcal{U}| - 1}. \tag{3.19}$$

Then, in Eq. (3.20) we denote and define the CRC as

$$\mathrm{CRC}(u_i) = \prod_{L \in \mathcal{L}} \left(1 + \beta_L \, \bar{\phi}_i^{(L)}\right), \tag{3.20}$$

where $\beta_L > 0$ serves as a weighting parameter for layer L, similarly to the one introduced in Section 3.2.1. The formulation of CRC provides a composite measure of effective influence that is inherently normalized for the number of nodes, thus allowing for direct comparisons across networks of varying sizes. Moreover, we note that the multiplicative formulation of CRC builds directly on the same rationale as $R(u_i)$ presented in Section 3.2.1, namely that radicalization risk is amplified when a user is simultaneously active across multiple behavioral dimensions. At the same time, CRC differs from $R(u_i)$ in its purpose. Indeed, CRC is designed as a metric that allows us to compare users within and across networks. For this reason, in the definition of $\mathrm{CRC}(u_i)$ we take into account the normalization given by $\bar{\phi}_i^{(L)}$, which ensures that obtained values are dimensionless and not biased by properties of the network, thus making CRC an interpretable measure for empirical analysis. It is also worth noting that the normalization exploited in the definition of CRC offers a convenient way to deal with the analysis of datasets containing a large number of users, as we will illustrate in our experiments.

3.3.2. Temporal burst influence (TBI)

Another interesting aspect to capture is the irregularity and intensity of user activity, which usually play a key role in phenomena such as online radicalization [37,38]. In order to capture it, we introduce the TBI. Let $u_i \in \mathcal{U}$ be a user. We denote with $\{w_i(t)\}$ the sequence of interaction counts observed in sliding time windows of duration Δt over the observation period, that is, the sequence of occurrences of $t_{ij}^{(k)}$, for all $u_j \in \mathcal{U}, u_i \neq u_j$. Pragmatically speaking, we count the interactions of u_i with the rest of users in $\mathcal{U}$ during the period Δt. Then, we compute the mean μ_i and standard deviation σ_i of this sequence. The Temporal Burst Influence (TBI) is defined in Eq. (3.21) as

$$\mathrm{TBI}(u_i) = \frac{\sigma_i - \mu_i}{\sigma_i + \mu_i}. \tag{3.21}$$

This metric yields values in the range $[-1, 1]$; a value close to 1 indicates highly bursty behavior, i.e., occasional intense bursts of activity with a low baseline, whereas values near -1 suggest very regular activity. Users with high burstiness (values near 1) are hypothesized to drive rapid information diffusion, thereby facilitating the conditions under which radical ideas can spread.

3.3.3. Sentiment synchronization coefficient (SSC)

Finally, another key aspect of different OSN phenomena is given by the sentiment and affective dynamics. Here, to robustly capture strong emotional alignment among users, we propose the SSC measure. Let $u \in \mathcal{U}$ be a user. We recall that we assume having its aggregate affective vector $\mathbf{a_i} \in \mathbb{R}^g$, as specified in Section 3.1.4. Then, let us denote with $N^{(\mathsf{ASL})}(u_i)$ the neighbor nodes $u_j \in \mathcal{U}$ of u_i in ASL, that is, $N^{(\mathsf{ASL})}(u_i) = \{u_j \in \mathcal{U} \mid \{u_i, u_j\} \in E^{(\mathsf{ASL})}\}$. The SSC is defined in Eq. (3.22) as

$$\mathrm{SSC}(u_i) = \frac{1}{|N^{(\mathsf{ASL})}(u_i)|} \sum_{u_j \in N^{(\mathsf{ASL})}(u_i)} (a_{ij})^2. \tag{3.22}$$

Note that, instead of averaging the raw cosine similarity scores, we square these values to accentuate cases of near-perfect alignment and downweight moderate similarities. Here, higher values of SSC imply that the user's affective expression is highly synchronized with that of their neighbors, a condition that could reinforce and stabilize radical or centered sentiment dynamics within the group.

4. Theoretical propositions and hypotheses

In the previous section, we introduced and illustrated our proposed IMRM model and its representation via the multilayer network $\mathfrak{M}$. Building on it and on the novel measures introduced above, we are now able to propose the following theoretical propositions and associated hypotheses regarding the dynamics of phenomena on OSN, and in particular of online radicalization. These propositions and related hypotheses form the backbone of our study; in turn, in the subsequent experimental campaign we aim to validate these claims by investigating the relationships between the proposed measures, e.g., CRC, TBI, SSC, and the observable patterns of online radicalization.

Proposition 1. *Multi-dimensional reinforcement increases radicalization probability.*

Hypothesis 1. We hypothesize that users exhibiting high effective influence across multiple layers are more likely to adopt and propagate radical ideologies, defined in Eq. (4.1) as:

$$P(u_i \text{ is radicalized}) = f(\mathrm{CRC}(u_i)), \tag{4.1}$$

where $f : \mathbb{R}_{>0} \to [0, 1]$ is a monotonically increasing function. This means that as the CRC score increases, the probability that user u_i becomes radicalized also increases. For example, Eq. (4.2) models f as a logistic function:

$$f(x) = \frac{1}{1 + e^{-\alpha(x-\theta)}}, \tag{4.2}$$

where $\alpha > 0$ controls the steepness of the increase, and θ is the CRC value at which the probability of radicalization is 0.5. This formulation captures the intuition that low CRC scores correspond to low probabilities of radicalization, while high CRC scores correspond to high probabilities.

Proposition 2. *Temporal clustering and affective alignment catalyze radicalization.*

Hypothesis 2. The combination of episodic surges in activity and strong emotional concordance among users creates a potent mechanism for the rapid emergence and spread of radical ideologies. Specifically, we propose that users who exhibit high TBI, i.e., indicating concentrated bursts of activity, in conjunction with high SSC, i.e., reflecting strong affective alignment with their peers, are more likely to participate in, and catalyze, the formation of extremist behaviors.

Proposition 3. *Bridging nodes exhibit higher cross-layer reinforcement.*

Hypothesis 3. Users that serve as bridges between communities, i.e., those connecting disparate clusters, tend to have higher effective reinforcement across layers. Formally, if bridging nodes show higher values of $\vec{\phi}_i^{(L)}$ compared to core radicalized members, then their Composite Reinforced Centrality (CRC) will be higher. Consequently, such nodes are hypothesized to be more entrenched in the radicalization process and may represent strategic targets for de-radicalization interventions. We denote this claim as

Bridging node $\Rightarrow$ Higher CRC(u_i) than non-bridging nodes,

implying increased susceptibility to intervention.

Proposition 4. *Context-invariant multi-dimensional reinforcement predicts radicalization.*

Hypothesis 4. Our model is general enough to accommodate different contexts of analysis. By Proposition 1, we hypothesize that users exhibiting high effective influence are more likely to adopt and propagate radical ideologies, and we represent this through a classical logistic function f. Given an instance $\mathfrak{M}_k$ of our model, we hypothesize in Eq. (4.3) that the logit coefficient $\alpha_{\mathfrak{M}_e}$ in

$$P(u_i \text{ is radicalized}) = \frac{1}{1 + e^{-(\alpha_{\mathfrak{M}_k} \mathrm{CRC}(u_i) + \theta_k)}} \tag{4.3}$$

remains significantly positive, demonstrating that CRC is a robust, context-invariant predictor. Here, θ_k is the CRC value at which the probability of radicalization is 0.5 for the instance $\mathfrak{M}_k$.

5. Experiments

In this section, we present our experimental campaign to evaluate the Integrated Multilayer Reinforcement Model (IMRM) and to test the theoretical propositions and hypotheses introduced earlier. Full code and data used for this paper can be found at https://github.com/ecorradini/reddit-ecochambers.

5.1. Dataset description and event selection

The dataset under study comprises Reddit posts and comments collected by pushshift and curated by some users of the platform (u/RaiderBDev, stuck_in_the_matrix, and Watchful1[1]). This comprehensive dataset spans nearly two decades of online activity, providing a rich source for understanding the evolution of online radicalization. Given the sheer size of our dataset, consisting of Reddit submissions and comments from June 2005 to December 2024, amounting to 3.12 TB in compressed form, we adopt a targeted approach to ensure the analysis is both computationally feasible and theoretically robust. Indeed, building and analyzing a complete multilayer network over the full 19-year period is computationally prohibitive. To address this challenge, we focus our analysis on five major events in the United States that have significantly shaped online discourse and are known to have catalyzed radical narratives. The selected events are:

1. 2008 US presidential election and financial crisis, a period marked by economic uncertainty and intense political debate, which contributed to the emergence of radical viewpoints [60];
2. 2011 Occupy Wall Street movement, a grassroots protest against economic inequality and corporate influence, which polarized public opinion and spurred extensive online debates on wealth disparity and systemic corruption [61];
3. 2016 US presidential election, an election cycle that witnessed the rise of extremist rhetoric and a surge in online polarizing content [62, 63];
4. 2017 Charlottesville Rally (Unite the Right Rally), an event that brought far-right ideologies to the forefront [64];
5. January 6th, 2021 Capitol Riot, a culmination of radicalized online discourse leading to a major political event, providing a critical case for studying radicalization dynamics [65].

For each event, we will extract a temporally bounded subset of the dataset. In doing so, we apply additional filters, such as relevant keywords and targeted subreddit selection, to isolate the discussions most pertinent to our study of online radicalization.

5.2. Experimental strategy

For each of the five selected events, we construct a separate IMRM that captures the radicalization dynamics specific to that event. Our experimental workflow for each event proceeds as follows:

1. identify and extract Reddit submissions and comments corresponding to a defined time window around the event, which involves keyword filtering and subreddit selection to ensure that the resulting dataset captures the radicalization-related discourse pertinent to that event;
2. for each extracted data, build an event-specific IMRM by constructing four intralayer networks and integrating the interlayer edges as detailed in Section 3.1;
3. compute the measures defined in Section 3.3 for each event-specific network;
4. evaluate the theoretical propositions and hypotheses defined in Section 4 by analyzing the computed measures in relation to the network dynamics observed during the event.

In summary, our experimental approach leverages the vast Reddit dataset by concentrating on key event-driven snapshots of online activity. By constructing focused IMRM networks for five major US events, we aim to validate the proposed theoretical framework and measures. This strategy not only makes the analysis computationally manageable but also enables us to explore the dynamic interplay of multiple behavioral dimensions in fostering online radicalization.

5.3. Data selection and filtering

Given the vast scale of our Reddit dataset, we concentrate our analysis on five major US events that are hypothesized to catalyze radicalization dynamics. For each event, we define temporal extension, apply tailored keyword filters, and select a broad range of subreddits known to host politically charged and radical discussions.

5.3.1. Temporal extensions

For each event, we delineate an extended time window to capture not only the immediate peak of activity but also the build-up and gradual de-escalation phases of radical discourse. The proposed time windows are as follows:

- October, November, and December 2008 for the 2008 US presidential election and financial crisis;
- August, September, and October 2011 for the Occupy Wall Street movement;

[1] https://academictorrents.com/details/ba051999301b109eab37d16f027b3f49ade2de13

- October, November, and December 2016 for the 2016 US presidential election;
- July, August, and September 2017 for the 2017 Charlottesville Rally;
- December 2020, January, and February 2021 for the 2021 Capitol Riot.

These extended windows are chosen to ensure that our analysis encompasses the lead-up, peak, and aftermath of each event, thereby providing a comprehensive view of the dynamics that drive online radicalization.

5.3.2. Keyword filtering

For each event, we define an extended set of keywords to capture not only the immediate surge of radical discourse but also the underlying narratives that emerge during the build-up and persist into the aftermath. These keyword filters are designed based on prior literature on online radicalization and initial exploratory analyses of the dataset [37,38]. The proposed keyword sets for each event are as follows:

- for the 2008 US presidential election and financial crisis, keywords include "financial crisis", "bailout", "bankruptcy", "election", "radical", "extremist", "conspiracy", and related variants;
- for the 2011 Occupy Wall Street movement, keywords include "Occupy Wall Street", "OWS", "corporate greed", "economic inequality", "wealth disparity", "social justice", "protest", and related variants;
- for the 2016 US presidential election, keywords include "Trump", "Clinton", "election fraud", "alt-right", "fake news", "radical", "hate", along with other emerging terms associated with political polarization;
- for the 2017 Charlottesville Rally, keywords include "Charlottesville", "Unite the Right", "neo-Nazi", "alt-right", "white supremacist", "racist", and similar expressions;
- for the 2021 Capitol Riot, keywords include "Capitol", "January 6", "riot", "insurrection", "election fraud", "radical", "extremist", and similar terms.

These extended keyword filters are applied to the submission dataset within the corresponding temporal windows to extract the segments most likely to contain discussions relevant to our study of online radicalization. By doing so, we aim to capture both the peak and the evolution of radical discourse associated with each event.

5.3.3. Subreddit selection

In addition to keyword filtering, we focus on a broad range of subreddits where politically charged and radical discussions are prevalent. Our subreddit selection is based on existing research on online radicalization as well as exploratory analysis of the dataset [56,66,67]. Selected subreddits include:

- `r/politics`, `r/news`, `r/worldnews`, `r/Ask_Politics`, for general politics and news;
- `r/conspiracy`, `r/The_Donald` (for data prior to its ban), `r/AltRight`, `r/Conservative`, `r/liberal`, for ideological and conspiracy disourse;
- additional subreddits that become active in relation to the events under study, which may include community-specific forums or subreddits dedicated to discussing particular incidents.

By including a diverse set of subreddits, we ensure that our analysis captures a wide spectrum of online discourse, from mainstream political discussion to more fringe or radical ideologies.

5.4. Exploratory data analysis

In this section, we provide an overview of our data selection and cleaning process, which serves as a precursor to constructing the IMRM networks for each event. Our analysis is based on two Comma Separated Values (CSV) files for each event, one for submissions and one for comments. The comments in the dataset are replying to a submission or another comment. For each selected event, we obtained two CSV files, centered on the temporal boundaries defined above, extracted from the whole Reddit dataset spanning June 2005 to December 2024. The datasets include the following fields:

- `submissions.csv`
  - `id`: unique identifier for the submission;
  - `author`: username of the submission's creator;
  - `created_utc`: timestamp of creation in Coordinated Universal Time;
  - `title`: title of the submission;
  - `selftext`: body text of the submission (if any);
  - `score`: net score (upvotes minus downvotes) of the submission;
  - `num_comments`: number of comments on the submission;
  - `subreddit`: subreddit where the submission was posted;
  - `permalink`: URL slug for accessing the submission;
  - `author_flair_text`: user-provided flair text (if any).
- `comments.csv`
  - `id`: unique identifier for the comment;
  - `author`: username of the comment's creator;
  - `created_utc`: timestamp of creation in Coordinated Universal Time;
  - `body`: text content of the comment;
  - `score`: net score of the comment;
  - `parent_id`: identifier of the parent submission or comment;
  - `subreddit`: subreddit where the comment was posted;
  - `permalink`: URL slug for accessing the comment;
  - `author_flair_text`: user-provided flair text (if any).

To ensure the quality and reliability of our data, we performed several cleaning steps. First, we removed all entries where the `author` field is set to "[deleted]". When a user is deleted, this field no longer allows unique identification, which would hinder our network construction and introduce potential biases. Additionally, we filtered out entries associated with bot accounts. Specifically, we removed all entries where the `author` field contained the substring "bot" (in any case) or matched a list of known bots, such as "AutoModerator". Finally, we removed all comments whose `parent_id` does not correspond to an existing submission or another valid comment. This step ensures the integrity of the reply structure, preventing orphaned comments from distorting the network.

5.4.1. 2008 US presidential election and financial crisis

For the 2008 U.S. presidential elections and the concurrent financial crisis, we have selected a comprehensive set of keywords intended to capture both the political dynamics of the 2008 elections and the economic turmoil of the financial crisis. The keywords include:

> "election", "Obama", "McCain", "campaign", "vote", "crisis", "recession", "bailout", "bankruptcy", "subprime", "housing bubble"

These keywords aim at isolating discussions directly related to the political contest, and target discourse around the economic crisis that coincided with the election. For this event, the subreddits chosen are *(i)* `politics`, and `Ask_Politics` that represent general political and news-related discourse; *(ii)* `conspiracy`, `Conservative`, and `liberal` that capture ideologically driven conversations and discussions where radical viewpoints may emerge.

Table 1 summarizes some of the basic statistics extracted from the filtered dataset for the 2008 Elections event.

5.4.2. 2011 occupy wall street movement

For the 2011 Occupy Wall Street Movement, we selected a comprehensive set of keywords intended to capture both the immediate surge

Table 1
Basic statistics for the 2008 Elections event dataset.

Statistic	Value
Number of posts	26,500
Number of comments	65,511
Average number of comments replying to posts	3.30
Average number of comments replying to other comments	1.36
Average number of submissions per author	5.84
Average number of comments per author	5.89
Percentage of authors that write both posts and comments	54.72 %
Average post score	17.47
Average comment score	3.58

Table 2
Basic statistics for the 2011 Occupy Wall Street Movement event dataset.

Statistic	Value
Number of posts	9477
Number of comments	105,798
Average number of comments replying to posts	8.70
Average number of comments replying to other comments	1.47
Average number of submissions per author	1.78
Average number of comments per author	3.64
Percentage of authors contributing both posts and comments	51.88 %
Average post score	34.54
Average comment score	4.38

of protest-related discourse and the broader socio-economic narratives that fueled the movement. The keywords include:

> "Occupy Wall Street", "OWS", "corporate greed", "economic inequality", "wealth disparity", "social justice", "protest"

These keywords aim at isolating discussions directly related to the protest movement and its associated activities, and target discourse surrounding the underlying economic and social grievances that motivated the protests. For this event, the subreddits chosen are *(i)* `politics`, and `Ask_Politics` that represent general political and news-related discourse; *(ii)* `OccupyWallStreet`, `socialjustice`, and `conspiracy` that capture grassroots protest dynamics and alternative narratives.

Table 2 summarizes some of the basic statistics extracted from the filtered dataset for the 2011 Occupy Wall Street Movement event.

5.4.3. 2016 US presidential election

For the 2016 U.S. presidential election we selected a set of keywords intended to capture both the intense political contest and the ideological polarization that characterized this election cycle. The keywords include:

> "Trump", "Clinton", "election fraud", "populism", "media bias", "alt-right", "fake news"

These keywords aim at isolating discussions directly related to the electoral contest and its controversies, along with the ideological debates and extreme viewpoints that emerged during this period. For this event, the subreddits chosen are *(i)* `Ask_Politics` that represents general political and news-related discourse; *(ii)* `conspiracy`, `The_Donald` (for data prior to its ban), `AltRight`, `Conservative`, and `liberal` that captures ideological debates and discussions where radical viewpoints may emerge.

Table 3 summarizes some of the basic statistics extracted from the filtered dataset for the 2016 US Presidential Election event.

5.4.4. 2017 charlottesville rally

For the 2017 Charlottesville Rally, widely known as the Unite the Right Rally, we selected a set of keywords intended to isolate discussions directly related to the rally and to capture the extremist and ideological narratives that surfaced during this period. The keywords include:

Table 3
Basic statistics for the 2016 US Presidential Election event dataset.

Statistic	Value
Number of posts	238,459
Number of comments	4,986,149
Average number of comments replying to posts	9.70
Average number of comments replying to other comments	1.66
Average number of submissions per author	4.66
Average number of comments per author	17.74
Percentage of authors contributing both posts and comments	71.25 %
Average post score	226.17
Average comment score	9.64

Table 4
Basic statistics for the 2017 Charlottesville Rally event dataset.

Statistic	Value
Number of posts	22,964
Number of comments	603,973
Average number of comments replying to posts	10.59
Average number of comments replying to other comments	1.74
Average number of submissions per author	2.56
Average number of comments per author	5.95
Percentage of authors contributing both posts and comments	71.67 %
Average post score	372.27
Average comment score	13.48

> "Charlottesville", "Unite the Right", "nazi", "alt-right", "white supremacist", "racist", "white nationalist", "fascist", "kkk", "rally"

These keywords help us pinpoint conversations that are focused on the event itself as well as the extremist ideologies that were prominently featured in the ensuing discourse. For this event, the subreddits chosen are *(i)* `politics`, and `Ask_Politics` that represent general political and news-related discourse; *(ii)* `The_Donald` (for data prior to its ban), `AltRight`, `Conservative`, and `liberal` that captures ideological debates and discussions where radical and extremist viewpoints are likely to be discussed.

Table 4 summarizes some of the basic statistics extracted from the filtered dataset for the 2017 Charlottesville Rally event.

5.4.5. 2021 capitol riot

For the 2021 Capitol Riot, we selected a set of keywords intended to isolate discussions directly related to the riot, insurrection, and the associated claims of election fraud, as well as broader extremist narratives. The keywords include:

> "Capitol", "January 6", "riot", "insurrection", "election fraud", "stop the steal".

These keywords are designed to capture both the immediate references to the event and the wider ideological narratives that fueled the mobilization of extremist sentiment. For this event, the subreddits chosen are *(i)* `politics`, and `Ask_Politics` that represents general political and news-related discourse; *(ii)* `conspiracy`, `QAnon`, `Conservative`, and `liberal` that captures ideological debates and extremist narratives.

Table 5 summarizes some of the basic statistics extracted from the filtered dataset for the 2021 Capitol Riot event.

5.5. Multilayer network construction based on IMRM

Our Integrated Multilayer Reinforcement Model (IMRM) is implemented in Python using several key libraries: `pandas` for data manipulation, `scikit-learn` for auxiliary processing, `NetworkX` and `MultinetX` for constructing multilayer networks, and `Torch` and `Transformers` for sentiment analysis. We construct four distinct layers that capture the different dimensions of online behavior, and we integrate these layers into a unified multilayer network. For the User Interaction Layer (UIL),

Table 5
Basic statistics for the 2021 Capitol Riot event dataset.

Statistic	Value
Number of posts	14,685
Number of comments	1,429,786
Average number of comments replying to posts	50.41
Average number of comments replying to other comments	1.91
Average number of submissions per author	2.41
Average number of comments per author	5.66
Percentage of authors contributing both posts and comments	66.57 %
Average post score	1545.56
Average comment score	14.35

Table 6
Values of τ_c and τ_a for each event.

Event	τ_c	τ_a
2008 US Presidential election and financial crisis	0.569	0.999
2011 Occupy Wall Street movement	0.579	0.999
2016 US Presidential election	0.549	0.999
2017 Charlottesville Rally	0.559	0.999
2021 Capitol Riot	0.559	0.999

we extract directed interactions (such as replies to posts or comments) from the cleaned Reddit dataset. Each directed edge from user u_i to u_j is weighted by the frequency of their interactions, and the associated timestamps are retained for subsequent temporal analysis. In the Content Similarity Layer (CSL), we aggregate textual content from both submissions and comments for each user. We leverage a pre-trained sentence embedding model, e.g., Sentence Transformers' `all-mpnet-base-v2`[2], to obtain semantically enriched embeddings that capture the deeper contextual meaning of the text. This specific model defines the dimensionality of the content attribute vector $\mathbf{c}_i$ as $d = 768$. We then compute the cosine similarity between these semantic embeddings for all user pairs, and retain an edge between two users if their similarity exceeds an empirically determined threshold τ_c. In our experiments, we set τ_c to the 90th percentile of the similarity distribution. This approach ensures that only significant thematic overlaps contribute to the network, thereby reducing noise and controlling network density. For the Temporal Dynamics Layer (TDL), we leverage the timestamps from the UIL to capture the burstiness of user interactions. For each pair of users, we calculate the maximum number of interactions occurring within any sliding window of length ΔT (set to 3600 s, i.e., 1 h). This maximum burst value is used as the edge weight, reflecting the synchronization of their activity over time. In the Affective and Sentiment Layer (ASL), we capture the emotional tone of user-generated content. Rather than relying solely on a single compound sentiment score, we construct a four-dimensional sentiment vector for each user by incorporating sentiment vectors computed by RoBERTA model. We compute the cosine similarity between these sentiment vectors and establish an edge if the similarity exceeds a threshold τ_a. Based on the empirical distribution of these similarity scores, we set τ_a to the 90th percentile, ensuring that only strong emotional alignments are preserved. In Table 6, we report the determined values of τ_c and τ_a for each event. Finally, to integrate these four layers into a cohesive multilayer network, we add interlayer edges that connect the same user across different layers. Each interlayer edge is assigned a uniform coupling weight ω (set to 1.0), ensuring that a user's multifaceted behavior is consistently linked across dimensions.

For each of the five events in our study, we construct a separate instance of IMRM. These instances are denoted as follows: $\mathfrak{M}_{2008}$ for the 2008 US presidential election and financial crisis, $\mathfrak{M}_{2011}$ for the 2011 Occupy Wall Street movement, $\mathfrak{M}_{2016}$ for the 2016 US presidential election, $\mathfrak{M}_{2017}$ for the 2017 Charlottesville Rally, and $\mathfrak{M}_{2021}$ for the 2021 Capitol Riot. Table 7 reports overall statistics for these multilayer networks. Interestingly, note that the number of interlayer edges exceeds the number of nodes. As defined in Section 3.1, interlayer edges connect each user to its representation in every distinct pair of layers. In our model, we consider four layers, and thus each user contributes six such interlayer edges. Therefore, the total number of interlayer edges for an instance of IMRM is $|E^{(\text{inter})}| = |U| \cdot \binom{4}{2} = |U| \cdot 6$.

[2] https://huggingface.co/sentence-transformers/all-mpnet-base-v2

5.6. Experiments on theoretical propositions and hypotheses

In this section, we describe our experiments designed to test the theoretical propositions outlined in Section 4 using five event-specific IMRM instances constructed in the previous section. We use the measures defined in Section 3.3 to empirically test our theoretical propositions presented in Section 4.

5.6.1. Testing Proposition 1: multi-dimensional reinforcement increases radicalization probability

To evaluate Proposition 1, which posits that higher multi-dimensional reinforcement increases the probability of radicalization, we focus on the Composite Reinforced Centrality (CRC) metric as our primary indicator, as defined in Eq. (3.20). CRC aggregates a user's effective influence across all four behavioral dimensions (UIL, CSL, TDL, and ASL), thereby capturing the compounded impact of their engagement in the network. It is worth pointing out that in this case, as well as for evaluating other propositions, we could also exploit the reinforcement factor $R(u_i)$, as defined in Eq. (3.17). Nevertheless, $\text{CRC}(u_i)$ builds directly on $R(u_i)$, and offers the possibility of comparing users within and across networks, especially when large networks are involved. Note that this is exactly our experimental case; therefore, CRC results in a valuable measure for evaluating our propositions.

Our experimental procedure is as follows:

1. For each user $u_i \in \mathcal{U}$ in the multilayer network, we compute $\text{CRC}(u_i)$.
2. We define a radicalization indicator for each user based on extremist keywords present in their contents. Users are then classified into radicalized and non-radicalized groups.
3. We test the hypothesis that higher CRC values are associated with increased radicalization probability by performing logistic regression with radicalization status as the dependent variable and CRC as the predictor, and by computing correlation coefficients between CRC and the radicalization measure to quantify the strength of their association.

A consistent, statistically significant positive relationship between CRC and the probability of radicalization across these tests would confirm Proposition 1, while a lack of such a relationship would call it into question. The results presented in Table 8 provide strong empirical support for Proposition 1. In all event-specific instances, users with CRC values at the 90th percentile exhibit markedly higher probabilities of radicalization compared to those at the 25th percentile. Specifically, while the probability of radicalization at the lower end ranges from 20 % to 27 %, it rises sharply to between 71 % and 81 % at the higher end. Moreover, the logistic regression analyses yield highly significant p-values ($p < 0.0001$) across all events, and the Pearson correlation coefficients (ranging from 0.39 to 0.47) indicate a positive relationship between CRC and radicalization status.

These findings confirm the hypothesis that multi-dimensional reinforcement, captured by the aggregated and normalized effective influences across the UIL, CSL, TDL, and ASL layers, substantially increases the likelihood of a user adopting radical ideologies. The consistency of these trends across diverse socio-political events (from the 2008 elections and financial crisis to the 2021 Capitol Riot) underscores the robustness and generalizability of the IMRM framework. This empirical evidence not only validates our theoretical proposition but also highlights the potential of CRC as a predictive tool for identifying radicalization in online social networks.

Table 7
Overall statistics of the IMRM instances for each event. Note that each node contributes to six interlayer edges, one for each distinct pair of layers.

Statistic	$\mathfrak{M}_{2008}$	$\mathfrak{M}_{2011}$	$\mathfrak{M}_{2016}$	$\mathfrak{M}_{2017}$	$\mathfrak{M}_{2021}$
Number of nodes	13,184	31,627	295,829	104,125	254,480
Number of interlayer edges	79,104	189,762	1,774,974	624,750	1,526,880
UIL					
Number of edges	56,209	87,695	4,068,992	516,521	1,144,819
Average weight	1.1655	1.2064	1.2254	1.1810	1.2489
Density	0.00032	0.00009	0.00005	0.00005	0.00002
CSL					
Number of edges	782,725	4,109,852	325,550,958	36,622,452	142,582,076
Average weight	0.6189	0.6245	0.5888	0.6031	0.6018
Density	0.00450	0.00411	0.00372	0.00338	0.00220
TDL					
Number of edges	56,209	87,695	4,068,992	516,521	1,144,819
Average weight	1.0498	1.0952	1.1153	1.1045	1.0915
Density	0.00032	0.00009	0.00005	0.00005	0.00002
ASL					
Number of edges	3,170,370	17,264,842	1,903,483,968	285,070,967	1,494,069,039
Average weight	0.9995	0.9995	0.9996	0.9996	0.9996
Density	0.01824	0.01726	0.02175	0.02629	0.02307

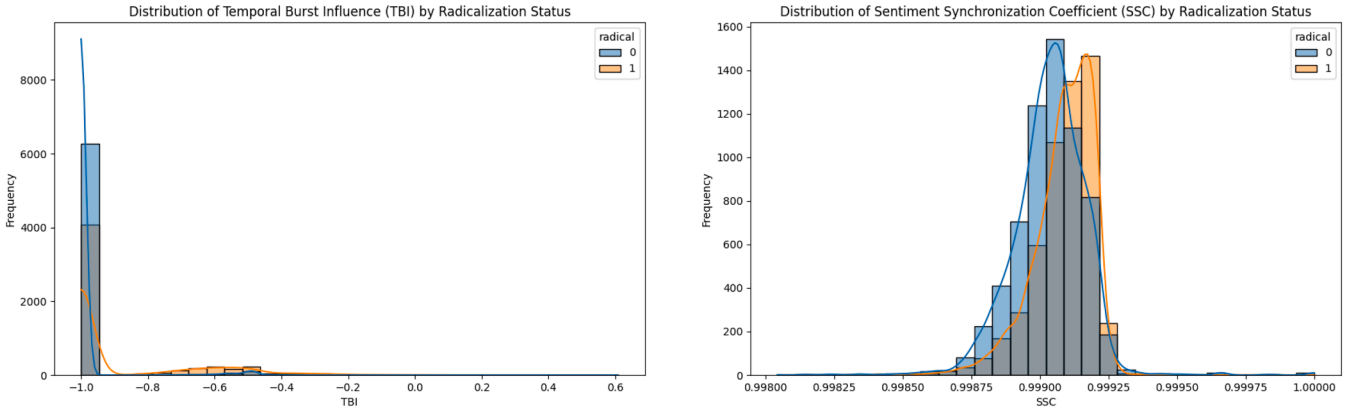


Fig. 2. Distribution of TBI (left) and SSC (right) by radicalization status for the 2008 US Elections event.

Table 8
Logistic regression results and Pearson correlation coefficients for each event-specific IMRM instance.

Instance	Prob. of radicalization (25 % of CRC)	Prob. of radicalization (90 % of CRC)	p-value	Pearson Corr.
$\mathfrak{M}_{2008}$	20 %	72 %	< 0.0001	0.45
$\mathfrak{M}_{2011}$	23 %	81 %	< 0.0001	0.45
$\mathfrak{M}_{2016}$	21 %	81 %	< 0.0001	0.47
$\mathfrak{M}_{2017}$	27 %	78 %	< 0.0001	0.43
$\mathfrak{M}_{2021}$	23 %	71 %	< 0.0001	0.39

5.6.2. *Testing Proposition 2: temporal clustering and affective alignment catalyze radicalization*

To explore Proposition 2, we examine how the revised Temporal Burst Influence (TBI) and Sentiment Synchronization Coefficient (SSC) measures distribute across radicalized and non-radicalized users in each of our five events. Figs. 2–6 present histograms (and kernel density estimates) of TBI and SSC, illustrating how episodic surges in user activity and strong affective alignment may correlate with radicalization status.

In each of these events, TBI is computed as the degree of burstiness in user activity, while SSC measures how strongly a user's affective signals align with those of their neighbors. The visualizations reveal how these measures vary between users labeled as radicalized and non-radicalized, thereby providing a basis for evaluating whether episodic surges in activity and strong affective alignment are indeed associated with the emergence and spread of radical ideologies.

For the 2008 US Elections event, as shown in Fig. 2, the TBI distribution for radicalized users is visibly shifted toward higher values compared to that of non-radicalized users. This suggests that those who experience more pronounced episodic surges in activity are more likely to adopt or propagate radical ideologies. Similarly, the SSC distribution in this event indicates that radicalized users tend to have higher levels of affective alignment with their peers, underscoring the role of emotional contagion in fostering radical views. In the 2011 Occupy Wall Street movement (Fig. 3), we observe a comparable pattern. The TBI histogram reveals that the radicalized group exhibits greater burstiness in their interaction patterns relative to non-radicalized users. The SSC distribution further supports this observation, as the radicalized users display a pronounced rightward shift, implying stronger emotional synchronization within their social circles. The figures corresponding to the 2016 US Elections event (Fig. 4) continue this trend. The TBI values for radicalized users are generally higher, suggesting that bursts of activity are a characteristic of the radicalized cohort. Concurrently, the SSC distribution for these users shows elevated values, indicating that strong affective alignment is a common feature among users who engage in radical discourse during this polarized electoral cycle. Fig. 5 shows that, while radicalized users do tend to occupy slightly higher ranges of TBI values compared to non-radicalized users, the overall

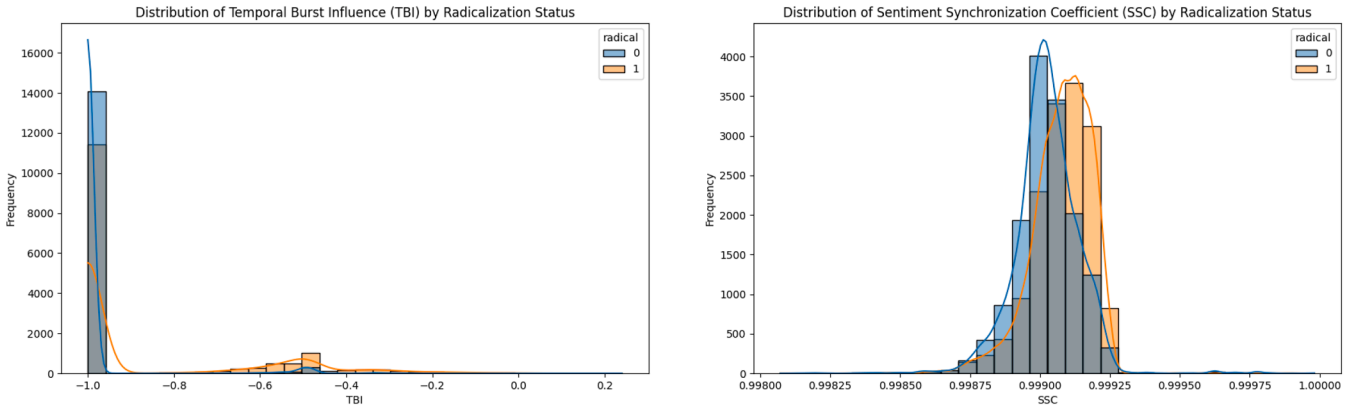


Fig. 3. Distribution of TBI (left) and SSC (right) by radicalization status for the 2011 Occupy Wall Street movement.

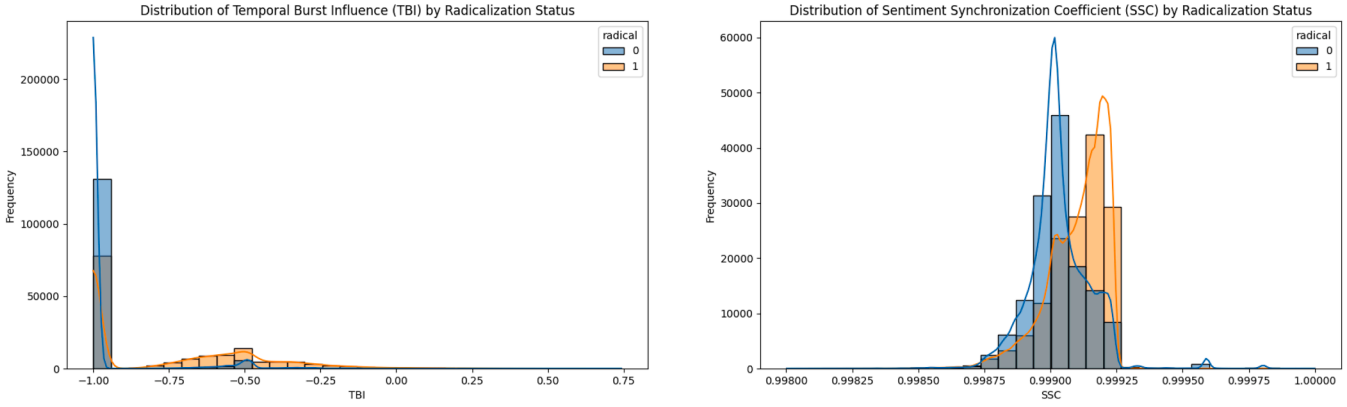


Fig. 4. Distribution of TBI (left) and SSC (right) by radicalization status for the 2016 US Elections.

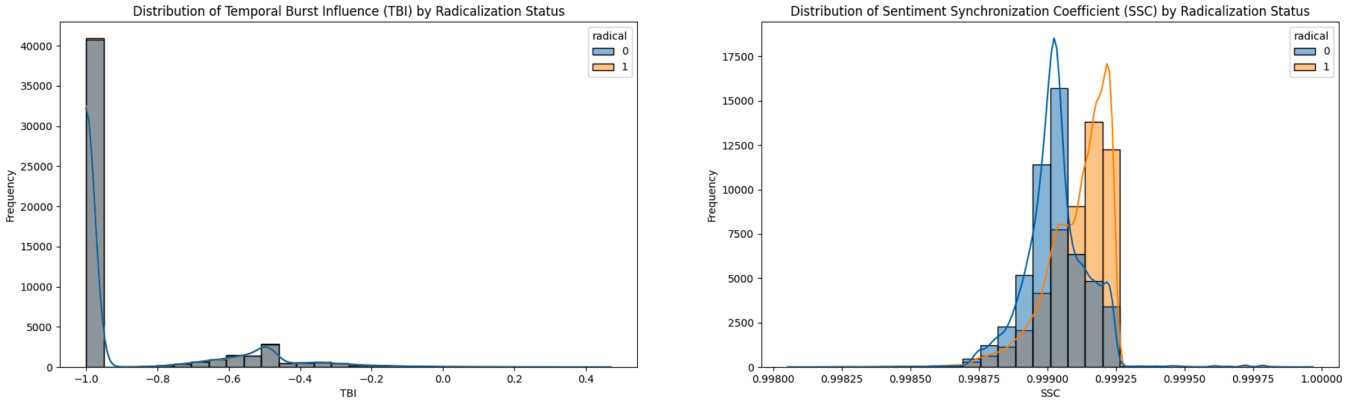


Fig. 5. Distribution of TBI (left) and SSC (right) by radicalization status for the 2017 Charlottesville Rally event.

separation between the two groups is less pronounced than in previous events. Both distributions exhibit considerable overlap, indicating that episodic surges in user activity (as captured by TBI) are not exclusively the domain of radicalized users. Likewise, the SSC distribution shows a moderate rightward shift for the radicalized group but again, the distinction is subtler than in other events. These patterns imply that temporal clustering and affective synchronization may still play a role in catalyzing extremist behavior, albeit in a more modest or context-dependent fashion for the Charlottesville Rally. Finally, the results for the 2021 Capitol Riot event, as depicted in Fig. 6, mirror the patterns observed in previous events. Radicalized users in this context are characterized by both higher TBI and SSC values compared to their non-radicalized counterparts. This consistency across events strengthens the hypothesis that the synergistic effect of temporal clustering and affective alignment plays a crucial role in catalyzing radicalization.

Collectively, these figures offer a visual evidence that users exhibiting high burstiness in their interactions and strong emotional alignment with their peers are more prone to radicalization. The observed patterns across different socio-political contexts suggest that the combined effects of episodic activity surges and synchronized affective expression may indeed be a driving force behind the emergence and propagation of radical ideologies.

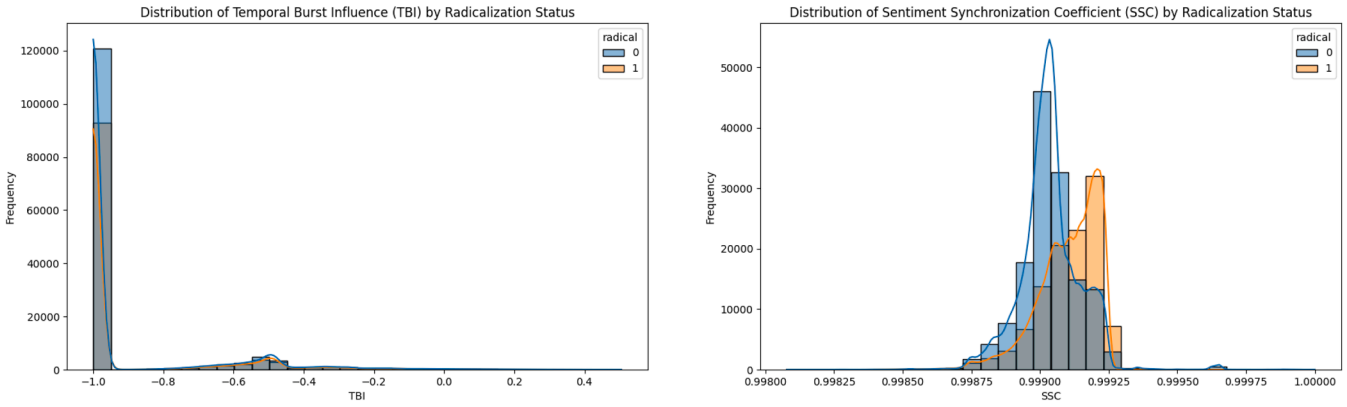


Fig. 6. Distribution of TBI (left) and SSC (right) by radicalization status for the 2021 Capitol Riot event.

Table 9
CRC by Bridging Status for 2008 US elections and financial crisis.

	count	mean	std	min	25 %	50 %	75 %	max
Bridging	1749	1.20	0.13	1.00	1.09	1.20	1.27	1.77
Non-Bridging	11,435	1.12	0.11	1.00	1.02	1.10	1.21	2.00

Table 10
CRC by Bridging Status for 2011 Occupy Wall Street.

	count	mean	std	min	25 %	50 %	75 %	max
Bridging	459	1.18	0.11	1.01	1.10	1.19	1.25	1.52
Non-Bridging	31,168	1.12	0.10	1.00	1.02	1.10	1.21	1.68

Table 11
CRC by Bridging Status for 2016 US elections.

	count	mean	std	min	25 %	50 %	75 %	max
Bridging	37,849	1.21	0.13	1.00	1.10	1.23	1.30	2.02
Non-Bridging	257,980	1.14	0.12	1.00	1.02	1.11	1.25	1.66

Table 12
CRC by Bridging Status for 2017 Charlottesville Rally.

	count	mean	std	min	25 %	50 %	75 %	max
Bridging	3984	1.21	0.15	1.00	1.07	1.21	1.32	1.65
Non-Bridging	100,141	1.17	0.13	1.00	1.04	1.17	1.30	2.50

Table 13
CRC by Bridging Status for 2021 Capitol Riot.

	count	mean	std	min	25 %	50 %	75 %	max
Bridging	6958	1.16	0.12	1.00	1.05	1.15	1.27	1.55
Non-Bridging	247,522	1.15	0.11	1.00	1.03	1.14	1.26	4.21

Table 14
Summary of Mean CRC Values by Bridging Status Across Events.

Event	Bridging	Non-Bridging	Difference
2008 US elections and financial crisis	1.20	1.12	0.08
2011 Occupy Wall Street	1.18	1.12	0.06
2016 US elections	1.21	1.14	0.07
2017 Charlottesville Rally	1.21	1.17	0.04
2021 Capitol Riot	1.16	1.15	0.01

Table 15
Logistic regression CRC coefficients (α) predicting radicalization across five events.

Event	CRC Coefficient α	p-value
2008 US elections and financial crisis	9.013	< 0.001
2011 Occupy Wall Street	9.443	< 0.001
2016 US elections	8.720	< 0.001
2017 Charlottesville Rally	6.363	< 0.001
2021 Capitol Riot	6.930	< 0.001

5.6.3. Testing Proposition 3: bridging nodes exhibit higher cross-layer reinforcement, offering intervention points

To evaluate Proposition 3, we compare the Composite Reinforced Centrality (CRC) values of users classified as bridging nodes, those who post or comment in more than one subreddit, with non-bridging nodes. Tables 9–13 report the CRC statistics for each event.
Finally, Table 14 summarizes the mean CRC values and their differences across events.

These results indicate that, across all events, users identified as bridging nodes exhibit slightly higher CRC values compared to non-bridging users. For instance, in the 2008 US Elections event, the mean CRC for bridging nodes is 1.20 versus 1.12 for non-bridging nodes. Similar differences are observed in the 2011, 2016, and 2017 events, although the difference is less pronounced in the 2021 Capitol Riot data. This pattern suggests that users who participate in multiple subreddits, and thus serve as connectors between communities, tend to accumulate greater cross-layer reinforcement. Such a higher CRC indicates that these bridging nodes are more deeply embedded within the network's cross-layer dynamics, making them potentially critical targets for de-radicalization interventions.

The observed differences consistently support the hypothesis that bridging nodes exhibit higher cross-layer reinforcement than core non-bridging nodes. This enhanced cross-layer influence may both contribute to and sustain radicalized behavior, highlighting the strategic importance of addressing bridging nodes in intervention efforts.

5.6.4. Testing Proposition 4: context-invariant multi-dimensional reinforcement predicts radicalization

To evaluate whether Composite Reinforced Centrality (CRC) consistently predicts radicalization across different socio-political contexts, we fit, for each event, a logistic regression of the form

$$\Pr(\text{radicalized} = 1) = \frac{1}{1 + e^{-(\alpha\,\text{CRC}+\beta)}}. \tag{5.1}$$

Table 15 reports the estimated CRC coefficient α and its p-value for each of the five events.

Across all five events, the CRC coefficient is large, positive, and highly significant, demonstrating that multi-dimensional reinforcement, as captured by CRC, is a powerful and context-invariant predictor of radicalized behavior.

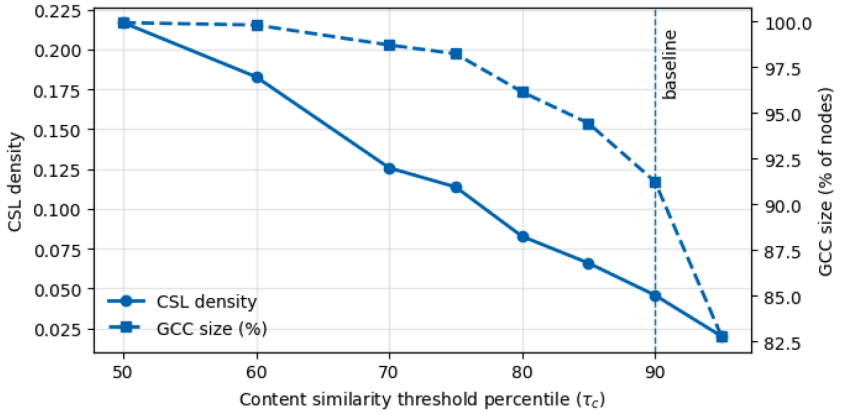


Fig. 7. CSL density (left axis) and GCC size (right axis) across $\tau\_c \in \{50\,\%, 60\,\%, 70\,\%, 75\,\%, 80\,\%, 85\,\%, 90\,\%, 95\,\%\}$.

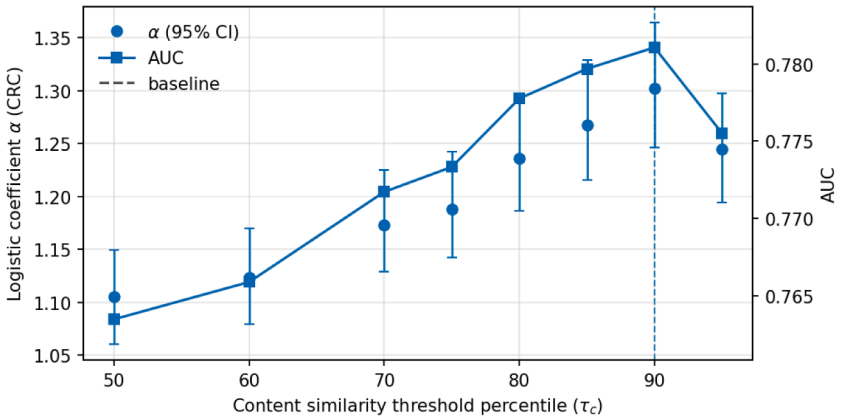


Fig. 8. Logistic coefficient α for CRC→radicalization (points with 95 % CIs, left axis) and AUC (line, right axis) as a function of τ_c.

5.7. Robustness and sensitivity analyses

In this section, we test the robustness of our framework, through two different sensitivity analyses. First of all, in Section 5.7.1 we show how IMRM behaves on different threshold choices. We focus in particular on τ_c. Then, in Section 5.7.2 we prove how the multiplicative formulation of CRC(u_i) better capture user radicalization.

5.7.1. Threshold sensitivity to similarity quantiles

We assess the sensitivity of our results to the choice of similarity cutoffs by varying the content threshold τ_c while keeping all other steps fixed (preprocessing, time binning ΔT, interlayer coupling γ, labeling and modeling pipeline). For computational tractability we use the smallest event (2008 Elections). At each value of $\tau_c \in \{50\,\%, 60\,\%, 70\,\%, 75\,\%, 80\,\%, 85\,\%, 90\,\%, 95\,\%\}$ we rebuild the CSL, reconstruct the multilayer network, recompute IMRM measures (including CRC), and refit the logistic model for radicalization using CRC as predictor.

In Fig. 7, we show some graph properties for CSL (density and size of the giant connected component, GCC). It shows a monotonic reduction in CSL density as τ_c increases, as expected from stricter sparsification. Despite the pruning, connectivity remains high over the entire sweep: the GCC covers essentially all nodes up to $\tau_c = 85\,\%$, stays above 90 % at $\tau_c = 90\,\%$, and remains above 80 % even at $\tau_c = 95\,\%$. This indicates that increasing τ_c primarily removes weak similarities without fragmenting the graph in the neighborhood of the baseline.

In Fig. 8 we show the logistic coefficient α for CRC→radicalization with 95 % confidence intervals (CIs) and the model AUC. CIs are computed from the estimated standard errors of the logistic regression (Wald CIs [68–70]); intervals not crossing zero indicate a statistically positive association. It reports the logistic coefficient α (left axis) and AUC (right axis) across τ_c. The coefficient is positive throughout and generally increases with the threshold, reaching its peak around $\tau_c = 90\,\%$, with confidence intervals that remain well above zero across the sweep. In parallel, AUC varies only marginally (on the order of a few 10^{-3}), indicating that predictive performance is essentially invariant to reasonable choices of τ_c.

Finally, in Fig. 9 we show the stability of high-CRC rankings via the Jaccard overlap between the top-k sets at each τ_c and the baseline ($\tau_c = 90\,\%$), for $k \in \{0.5\,\%, 1\,\%, 5\,\%\}$. It summarizes the overlap between the top-k CRC users at each τ_c and the baseline ranking ($\tau_c = 90\,\%$). Overlaps are high for all k in the range $\tau_c \in [80\,\%, 90\,\%]$ (e.g., ≥ 0.8 for top-1 % and ≥ 0.9 for top-5 %), with only a mild decline at $\tau_c = 95\,\%$ consistent with increased sparsity.

As a last experiment, in Table 16 we replicated the threshold sweep at the baseline pipeline across all five events. Discriminative performance is stable: AUC@90 lies between 0.737 and 0.781 (median 0.767), and the largest relative deterioration over $\tau_c \in [70\,\%, 95\,\%]$ is at most 2.57 %. High-CRC identities are likewise robust: the minimum Jaccard overlap of the top-1 % set with the $\tau_c = 90\,\%$ baseline (computed over $\tau_c \in [80\,\%, 95\,\%]$) is ≥ 0.81 across events (median 0.85) and equals 1.00 in two cases, indicating virtually identical heads of the ranking. On the structural side, CSL connectivity responds heterogeneously to aggressive pruning: election datasets maintain large components throughout (minimum GCC% $\geq 82.8\,\%$ in 2008 and 86.6 % in 2016), Occupy Wall Street dips to 59.3 %, and short-horizon mobilization events can become extremely sparse at the highest quantiles (near-zero minimum GCC%). Notably, even in these sparse regimes predictive metrics and top-1 % overlaps remain essentially unchanged, placing our baseline choice ($\tau_c = 90\,\%$) at the center of a broad stability plateau.

Fig. 9. Jaccard overlap between the top-k CRC sets at each τ_c and the baseline ranking at $\tau\_c = 90\,\%$, for $k \in \{0.5\,\%, 1\,\%, 5\,\%\}$.

Table 16
Cross-event robustness summary for the threshold sweep.

Metric	Result
AUC@90 (range; median)	0.737–0.781 (median 0.767)
Max relative AUC drop over $\tau_c \in [70\,\%, 95\,\%]$	$\leq 2.57\,\%$
Min top-1 % Jaccard vs. baseline over $\tau_c \in [80\,\%, 95\,\%]$	≥ 0.81 (median 0.85); equals 1.00 in two events
Min GCC% (event notes)	2008: 82.8 %; 2016: 86.6 %; 2011: 59.3 %; 2017/2021: ≈ 0 at 95 %

Table 17
Five-fold cross-validation comparison of CRC_+ (additive) vs. $\mathrm{CRC}_\times$ (multiplicative).

Event	Δlog-loss	Wilcoxon p	ΔAUC
2008 Elections	+0.0000	0.594	+0.0000
2011 Occupy Wall Street	−0.0012	0.031	+0.0000
2016 Elections	−0.0009	0.031	+0.0000
2017 Charlottesville Rally	−0.0067	0.031	−0.0003
2021 Capitol Riot	−0.0063	0.031	+0.0000

Taken together, these results place the baseline choice at the center of a stability scenario: the network remains well connected, the CRC effect size is strong and significant, and the identity of high-CRC users is largely preserved across thresholds.

5.7.2. Additive vs. multiplicative aggregation

As a final comparison at the baseline thresholds ($\tau_c = 90\,\%$ for CSL and baseline τ_a for ASL), we benchmark the multiplicative CRC against additive formulations while reusing the same layer-specific signals as in Section 3.3. Let $\mathcal{L} = \{\mathsf{UIL}, \mathsf{CSL}, \mathsf{TDL}, \mathsf{ASL}\}$ and let $\bar{\phi}_u^{(L)}$ denote the normalized effective influence of user u in layer L (Eq. (3.19)). With non-negative layer weights β_L (set to 1 unless noted), we consider

$$\mathrm{CRC}_\times(u) = \prod_{L\in\mathcal{L}} \left(1 + \beta_L\, \bar{\phi}_u^{(L)}\right), \qquad \mathrm{CRC}_+(u) = \sum_{L\in\mathcal{L}} \beta_L\, \bar{\phi}_u^{(L)}.$$

In addition to CRC_+, we also use a Z-sum, which standardizes each layer within the event and then averages them; this yields a scale-robust composite where all dimensions contribute on a comparable base [71,72]. Also, we exploit a learned linear index, obtained by fitting a logistic model on the standardized layer signals to estimate data-driven weights; this serves as a strong additive baseline that optimally combines layers for prediction and remains interpretable via per-layer coefficients [68,73].

As we can see from Table 17, across all five events, additive baselines are indistinguishable from the multiplicative CRC in discrimination: the 5-fold cross-validated ΔAUC relative to $\mathrm{CRC}_\times$ is effectively zero for CRC_+, Z-sum, and the learned linear index, and top-1 % rankings show near-complete overlap. This means that, for separating radicalized from non-radicalized users, both multiplicative and additive formulations rank individuals similarly and arise the same high-risk head of the distribution. At the same time, a model that augments the additive specification with pairwise interactions among layer-wise signals (L2-regularized logistic) consistently improves calibration. In practice, the multiplicative aggregation yields lower 5-fold cross-validation log-loss than purely additive baselines with the same AUC and near-identical top-1 % selections, meaning its probability estimates match observed frequencies more closely while preserving ranking performance. This means that modeling cross-layer synergy through a multiplicative formulation produces probability estimates that match observed frequencies, e.g., groups predicted at 70 % risk are positive about 70 % of the time, while leaving ranking performance (AUC) unchanged. These findings align with the theoretical rationale of CRC as a non-compensatory, cross-layer reinforcement measure, multiplicative on the original scale and additive on the log scale, providing the intended synergy semantics without any loss in predictive power.

6. Discussion and limitations

Our Integrated Multilayer Reinforcement Model (IMRM) provides a comprehensive framework for understanding the multifaceted dynamics of online radicalization, revealing how interactions, content similarity, temporal bursts, and sentiment alignment synergistically contribute to the emergence and propagation of extremist behaviors. Our findings support a multilayer view of radical engagement, yet our study is not without limitations, which we address here to contextualize the results and guide future research.

First of all, in our evaluation we rely on a content-based heuristic that flags users as radicalized when their posts or comments contain extremist-related keywords. This choice favors transparency and reproducibility, but it is inevitably context-sensitive. Quotation, critique, satire, and stance-taking can trigger keywords without signaling endorsement; conversely, dog whistles, euphemisms, and multimodal cues may elude lexical rules. As a result, both false positives, e.g., users denouncing or rebutting extremist narratives, and false negatives, e.g., users adopting oblique language, are possible. More generally, keyword lists drift over time and across communities, and they may confuse radicalization with adjacent constructs such as toxicity or incivility. While IMRM framework itself is label agnostic (our measures are computed independently of how labels are assigned) the empirical tests do depend on the labeling strategy. To mitigate such issues in future work,

alternative labeling strategies could be explored, including the integration of advanced natural language processing techniques like contextual embeddings or Large Language Models (LLMs) to discern supportive versus oppositional stances [74–78]. Additionally, incorporating manual annotation by domain experts on a subset of data for ground-truth validation, or hybrid approaches combining keyword filtering with supervised machine learning models trained on annotated corpora, could enhance accuracy. These enhancements would not only reduce misclassification but also allow for more nuanced detection of radicalization trajectories over time.

Secondly, we analyze five high-salience U.S. events using temporally bounded subsets extracted from a multi-terabyte Reddit corpus. This design is methodologically coherent with our interest in bursty, event-driven dynamics, and it makes the analysis tractable on commodity hardware; however, it also introduces limitations. Event windows emphasize acute mobilization and may over-represent actors who surface during peaks while under-representing chronic, low-level radical engagement. Subreddit and keyword filters can also induce topical spectrum bias. Consequently, external validity to continuous, platform-wide monitoring should be claimed with caution. From a framework perspective, nothing in IMRM requires event gating: the four layers can be maintained in a streaming fashion with incremental updates. In practice, scaling to always-on monitoring calls for *(i)* approximate nearest-neighbor indexing for CSL/ASL embeddings (to avoid $O(|U|^2)$ similarity scans), *(ii)* time-decayed edge weights and rolling windows for TDL/UIL to control memory and capture drift, *(iii)* adaptive sparsification (percentile- or null-model-based thresholds) to stabilize density while preserving strong ties, and *(iv)* incremental or sketch-based centralities for CRC. These choices convert the present "snapshot" evaluation into a near-real-time service while preserving the semantics of our measures. In short, our multilayer formalism is robust to alternative supervision (labels can be swapped without redesigning the network) and is architecturally compatible with continuous operation. The main risks lie not in the measures themselves but in *(i)* how radicalization is operationalized from text and *(ii)* how we subsample the platform for tractability. We have foregrounded these constraints and outlined concrete upgrades that would reduce misclassification while enabling deployment at scale, thereby strengthening the generalizability of the conclusions.

7. Conclusion

In this paper, we have introduced the Integrated Multilayer Reinforcement Model (IMRM), a novel framework that brings together interaction patterns, thematic content, temporal burstiness, and sentiment flows into a single, cohesive multilayer network. By explicitly modeling interlayer coupling and feedback loops, IMRM goes beyond descriptive multilayer snapshots to capture how activity in one dimension amplifies, and is amplified by, activity in the others. This reinforcement perspective not only aligns closely with sociological theories of social learning and emotional contagion, but also provides a transparent, quantitative apparatus for measuring a user's compounded influence and vulnerability to radicalization. Our extensive empirical evaluation on five high-profile US socio-political events (from the 2008 election and financial crisis to the 2021 Capitol Riot) offers strong support for the model's core premises. First, Composite Reinforced Centrality (CRC), our multiplicative aggregation of normalized layer influences, consistently exhibits a large, positive, and highly significant relationship with the probability of extremist engagement. Users in the top decile of CRC are three to four times more likely to produce radical content than those in the bottom quartile, across all events and network sizes. Second, the joint analysis of Temporal Burst Influence (TBI) and Sentiment Synchronization Coefficient (SSC) confirms that episodic surges in activity, when paired with high emotional alignment, act as powerful catalysts for rapid community consolidation. In every context, radicalized users cluster in the upper ranges of both measures, highlighting the interplay between "when" people speak and "how" they feel. Compared to traditional single-layer SNA or even descriptive multilayer approaches, IMRM makes two critical advances. On the one hand, classic centrality or modularity measures treat all ties as monolithic and ignore content or affect; on the other, many existing multilayer studies examine layers in isolation or impose only weak interlayer coupling. IMRM bridges these gaps by enforcing equal-strength coupling across layers and by embedding concave transformations and multiplicative interactions that temper extreme values while preserving interpretability. The result is a set of normalized, comparable measures (CRC, TBI, and SSC) that can be computed on networks of different scales and platforms, offering both theoretical insight and practical utility.

These findings have key implications for both theory and practice. Theoretically, they underscore that radicalization is neither purely structural nor solely narrative-driven but emerges from the interplay of who you interact with, what you share, when you act, and how you feel. Practically, CRC, TBI and SSC offer interpretable, normalized indicators that platforms or policymakers could use to flag emergent extremist cores or to tailor timely, dimension-specific interventions, e.g., disrupting bursts of coordinated outreach or countering emotion-laden messaging. As with any modeling effort, certain simplifications were necessary. For instance, we rely on keyword-based labels and fixed thresholds for content similarity and burst detection, choices informed by the literature and bolstered by robustness checks that may not reflect every subtlety of coded language or platform quirks. Similarly, although the robust associations we uncover are encouraging, complementary experimental or longitudinal work will be needed to unpack the underlying causal pathways. Finally, our focus on Reddit provides a valuable testbed but invites further exploration on platforms with different norms, user bases, or moderation policies.

Looking ahead, several promising directions could extend and enrich the IMRM framework. First, integrating additional behavioral dimensions, such as multimedia sharing or network multiplexity across entirely different platforms, would deepen our understanding of cross-venue radicalization. Second, adapting the model to operate in an online or streaming setting could support real-time monitoring and early-warning systems. Third, embedding more sophisticated content representations, e.g., transformer-based embeddings or discourse structure, and adaptive thresholding mechanisms may capture subtler thematic shifts and reduce reliance on static parameters. Fourth, coupling IMRM with causal inference methods or field experiments would help disentangle reinforcement mechanisms and validate intervention strategies. Finally, translating our measures into actionable tools, for instance, by designing de-radicalization simulations or targeted moderation policies, could bridge the gap between descriptive analysis and practical impact.

CRediT authorship contribution statement

Enrico Corradini: Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization; **Francesco Cauteruccio:** Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Data availability

Data and code are available at https://github.com/ecorradini/reddit-ecochambers.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research has received funding from the project Vitality - Project Code ECS00000041, CUP I33C22001330007 - funded under the National Recovery and Resilience Plan (NRRP), Mission 4 Component 2 Investment 1.5 - "Creation and strengthening of innovation ecosystems," construction of "territorial leaders in R&D" - Innovation Ecosystems - Project "Innovation, digitalization and sustainability for the diffused economy in Central Italy - VITALITY" Call for tender No. 3277 of 30/12/2021, and Concession Decree No. 0001057.23-06-2022 of Italian Ministry of University funded by the European Union - NextGenerationEU. This research is also supported by the PNRR project FAIR–Future AI Research (PE00000013) under the NRRP MUR program funded by NextGenerationEU.

References

[1] M. Minici, F. Cinus, C. Monti, F. Bonchi, G. Manco, Cascade-based echo chamber detection, in: Proceedings of the 31st ACM International Conference on Information & Knowledge Management, 2022, pp. 1511–1520.
[2] R. Interian, R.G. Marzo, I. Mendoza, C.C. Ribeiro, Network polarization, filter bubbles, and echo chambers: an annotated review of measures and reduction methods, Int. Trans. Oper. Res. 30 (6) (2023) 3122–3158. Wiley.
[3] S.D.G. Putri, E.P. Purnomo, T. Khairunissa, Echo chambers and algorithmic bias: the homogenization of online culture in a smart society, in: SHS Web of Conferences, 202, EDP Sciences, 2024, p. 05001.
[4] A.F. Peralta, M. Neri, J. Kertész, G. Iñiguez, Effect of algorithmic bias and network structure on coexistence, consensus, and polarization of opinions, Phys. Rev. E 104 (4) (2021) 044312.
[5] F. Alatawi, L. Cheng, A. Tahir, M. Karami, B. Jiang, T. Black, H. Liu, A survey on echo chambers on social media: Description, detection and mitigation, arXiv:2112.05084 (2021).
[6] J. Chen, G. Kou, H. Wang, Y. Zhao, Influence identification of opinion leaders in social networks: an agent-based simulation on competing advertisements, Inf. Fusion 76 (2021) 227–242. Elsevier.
[7] N. Dakiche, F.B. Tayeb, Y. Slimani, K. Benatchba, Tracking community evolution in social networks: a survey, Inf. Process. Manag. 56 (3) (2019) 1084–1102. Elsevier.
[8] Z. Xu, X. Rui, J. He, Z. Wang, T. Hadzibeganovic, Superspreaders and superblockers based community evolution tracking in dynamic social networks, Knowledge-Based Syst. 192 (2020) 105377. Elsevier.
[9] M. Lai, M. Tambuscio, V. Patti, G. Ruffo, P. Rosso, Stance polarity in political debates: a diachronic perspective of network homophily and conversations on Twitter, Data Knowl. Eng. 124 (2019) 101738. Elsevier.
[10] B. Zhang, L. Zhang, C. Mu, Q. Zhao, Q. Song, X. Hong, A most influential node group discovery method for influence maximization in social networks: a trust-based perspective, Data Knowl. Eng. 121 (2019) 71–87. Elsevier.
[11] A. Almars, X. Li, X. Zhao, Modelling user attitudes using hierarchical sentiment-topic model, Data Knowl. Eng. 119 (2019) 139–149. Elsevier.
[12] G. Bonifazi, C. Buratti, E. Corradini, M. Marchetti, F. Parlapiano, D. Ursino, L. Virgili, Defining, detecting, and characterizing power users in threads, Big Data Cogn. Comput. 9 (3) (2025) 69. MDPI.
[13] F. Cauteruccio, E. Corradini, M. Marchetti, D. Ursino, L. Virgili, A framework for investigating discording communities on social platforms, Electronics 14 (3) (2025) 609. MDPI.
[14] A. Bermingham, M. Conway, L. McInerney, N. O'Hare, A.F. Smeaton, Combining social network analysis and sentiment analysis to explore the potential for online radicalisation, in: 2009 International Conference on Advances in Social Network Analysis and Mining, IEEE, 2009, pp. 231–236.
[15] E. Ferrara, Contagion dynamics of extremist propaganda in social networks, Inf. Sci. 418 (2017) 1–12. Elsevier.
[16] P. Gerbaudo, C.C. De Falco, G. Giorgi, S. Keeling, A. Murolo, F. Nunziata, Angry posts mobilize: emotional communication and online mobilization in the Facebook pages of Western European right-wing populist leaders, Social Media+ Soc. 9 (1) (2023) 20563051231163327. SAGE.
[17] G. Persson, Love, affiliation, and emotional recognition in# kämpamalmö:—The social role of emotional language in Twitter discourse, Social Media+ Soc. 3 (1) (2017) 2056305117696522.
[18] H. Sakariassen, Women's emotion work on Facebook: strategic use of emotions in public discourse, Comput. Hum. Behav. Rep. 4 (2021). Elsevier, https://api.semanticscholar.org/CorpusID:240200249.
[19] A.A. Kane, L.M.V. Swol, I.G. Sarmiento-Lawrence, Emotional contagion in online groups as a function of valence and status, Comput. Hum. Behav. 139 (2023) 107543. Elsevier.
[20] M.D. Vicario, G. Vivaldo, A. Bessi, F. Zollo, A. Scala, G. Caldarelli, W. Quattrociocchi, Echo chambers: emotional contagion and group polarization on facebook, Sci. Rep. 6 (1) (2016) 37825. Nature Publishing Group.
[21] K. Lerman, D. Feldman, Z. He, A. Rao, Affective polarization and dynamics of information spread in online networks, npj Complex. 1 (1) (2024) 8. Nature Publishing Group.
[22] E.L. Wang, L. Luceri, F. Pierri, E. Ferrara, Identifying and characterizing behavioral classes of radicalization within the qanon conspiracy on Twitter, in: Proceedings of the International AAAI Conference on Web and Social Media, 17, 2023, pp. 890–901.
[23] H. Habib, P. Srinivasan, R. Nithyanand, Making a radical misogynist: how online social engagement with the manosphere influences traits of radicalization, Proc. ACM Human-Computer Interact. 6 (CSCW2) (2022) 1–28.
[24] Z. Ghalmane, M.E. Hassouni, C. Cherifi, H. Cherifi, Centrality in modular networks, EPJ Data Sci. 8 (1) (2019) 15. Springer.
[25] A. Mislove, M. Marcon, K.P. Gummadi, P. Druschel, B. Bhattacharjee, Measurement and analysis of online social networks, in: Proc. of the ACM SIGCOMM International Conference on Internet Measurement (IMC'07), San Diego, CA, USA, 2007, pp. 29–42. ACM.
[26] C. Buntain, J. Golbeck, Identifying social roles in reddit using network structure, in: Proc. of the International Conference on World Wide Web (WWW'14), Seoul, Korea, 2014, p. 615–620. ACM.
[27] A.K. Yadav, R. Johari, R. Dahiya, Identification of centrality measures in social network using network science, in: International Conference on Computing, Communication, and Intelligent Systems (ICCCIS'19), Greater Noida, India, 2019, pp. 229–234. IEEE.
[28] P. Howlader, K.S. Sudeep, Degree centrality, eigenvector centrality and the relation between them in Twitter, in: International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT'16), Bangalore, India, 2016, pp. 678–682. IEEE.
[29] D. Zhuang, J.M. Chang, M. Li, DynaMo: dynamic community detection by incrementally maximizing modularity, IEEE Trans. Knowl. Data Eng. 33 (5) (2019) 1934–1945. IEEE.
[30] P. De Meo, E. Ferrara, G. Fiumara, A. Provetti, Mixing local and global information for community detection in large networks, J. Comput. Syst. Sci. 80 (1) (2014) 72–87. Elsevier.
[31] A.P. Logan, P.M. LaCasse, B.J. Lunday, Social network analysis of Twitter interactions: a directed multilayer network approach, Soc. Netw. Anal. Min. 13 (1) (2023) 65. Springer.
[32] J. Sawicki, M. Ganzha, M. Paprzycki, Y. Watanobe, Reddit CrosspostNet—Studying reddit communities with large-scale crosspost graph networks, Algorithms 16 (9) (2023) 424. MDPI. https://doi.org/10.3390/a16090424
[33] M.A.U. Hasan, A.A. Bakar, M.R. Yaakub, Measuring user influence in real-time on Twitter using behavioural features, Physica A 639 (2024) 129662. Elsevier. https://doi.org/10.1016/j.physa.2023.129662
[34] M. Nooribakhsh, M. Fernández-Diego, F. González-Ladrón-De-Guevara, Community detection in social networks using machine learning: a systematic mapping study, Knowl. Inf. Syst. 66 (2024) 7205–7259. Springer. https://doi.org/10.1007/s10115-023-02031-2
[35] J. Choi, R. Fernandez, A. Park, A multidimensional analysis of YouTube communities in the Indo-Pacific region, in: Proc. of SOTICS 2022, the Twelfth International Conference on Social Media Technologies, Communication, and Informatics, 2022, pp. 452–460.
[36] M. Youngblood, Extremist ideology as a complex contagion: the spread of far-right radicalization in the United States between 2005 and 2017, Humanit. Social Sci. Commun. 7 (1) (2020) 1–10. Nature Publishing Group.
[37] D. Shin, K. Jitkajornwanich, How algorithms promote self-radicalization: audit of Tiktok's algorithm using a reverse engineering method, Soc. Sci. Comput. Rev. 42 (4) (2024) 1020–1040. SAGE.
[38] A. Fahad, S.E. Mustafa, Locked in echoes: unveiling the dynamics of social media echo chambers and Hindu radicalization targeting Muslim youth in Delhi, Humanit. Social Sci. Commun. 12 (1) (2025) Article 324. Springer Science and Business Media LLC.
[39] M. Cinelli, A. Pelicon, I. Mozetič, W. Quattrociocchi, P.K. Novak, F. Zollo, Dynamics of online hate and misinformation, Sci. Rep. 11 (1) (2021) 1–12. Nature Publishing Group.
[40] M.H. Ribeiro, P.H. Calais, Y.A. Santos, V.A.F. Almeida, R. West, Do platform migrations compromise content moderation? Evidence from r/the_donald and r/incels, in: Proceedings of the ACM on Human-Computer Interaction (CSCW), vol. 5, 2021, pp. 1–27. https://doi.org/10.1145/3449102
[41] G. Russo, L. Luceri, E. Ferrara, Spillover of antisocial behavior from fringe platforms: the unintended consequences of community banning, in: Proceedings of the International AAAI Conference on Web and Social Media (ICWSM), 17(1), 2023, pp. 881–892.
[42] C. Buntain, M. Innes, T. Mitts, J.N. Shapiro, Cross-platform reactions to the post-January 6 deplatforming, J. Quant. Descr. 1 (2023) 1–44. HBZ Open Publishing Environment. https://doi.org/10.51685/jqd.2023.004
[43] S. Phadke, T. Mitra, Educators, solicitors, flamers, motivators, sympathizers: characterizing roles in online extremist movements, Proc. ACM Human-computer Interact. 5 (CSCW2) (2021) 1–35.
[44] U. Kursuncu, M. Gaur, C. Castillo, A. Alambo, K. Thirunarayan, V. Shalin, D. Achilov, I.B. Arpinar, A. Sheth, Modeling islamist extremist communications on social media using contextual dimensions: religion, ideology, and hate, Proc. ACM Human-Computer Interact. 3 (CSCW) (2019) 1–22.
[45] E. Corradini, Deconstructing cultural appropriation in online communities: a multilayer network analysis approach, Inf. Process. Manag. 61 (3) (2024) 103662. Elsevier.
[46] S. Wippell, D.L. Haynie, Multiplex networks and the structure of the proud boys: organizational and social media ties in far-right mobilization, Soc. Netw. 75 (2025) 45–58. Elsevier. https://doi.org/10.1016/j.socnet.2024.11.003
[47] S.J. Baele, L. Brace, Mapping the far-right online ecosystem with socio-semantic networks: an integrative multilayer approach, Inf., Commun. Soc. 28 (2) (2025) 189–212. Taylor & Francis. https://doi.org/10.1080/1369118X.2024.2015657

[48] S.J. Baele, L. Brace, D. Ging, A diachronic cross-platforms analysis of violent extremist language in the incel online ecosystem, Terror. Polit. Violence 36 (1) (2024) 145–164. Taylor & Francis. https://doi.org/10.1080/09546553.2022.2161373
[49] E. Peralta, J.G. Cabañas, A. Cuevas, R. Cuevas, N. Ducheneaut, Two-layer networks of politicians and users on Twitter: mapping polarization and community structure, J. Complex. Netw. 12 (1) (2024) cnad073. Oxford Academic. https://doi.org/10.1093/comnet/cnad073
[50] K. Durrheim, J. Schuld, Echo chambers and the multilayer structure of online opinion: comparing interaction and semantic networks, New Media Soc. 27 (3) (2025) 399–418. SAGE. https://doi.org/10.1177/14614448231234567
[51] S. Boccaletti, G. Bianconi, R. Criado, C.I.D. Genio, J. Gómez-Gardenes, M. Romance, I. Sendina-Nadal, Z. Wang, M. Zanin, The structure and dynamics of multilayer networks, Phys. Rep. 544 (1) (2014) 1–122. Elsevier.
[52] A.N. Medvedev, R. Lambiotte, J.C. Delvenne, The anatomy of reddit: an overview of academic research, in: Dynamics on and of Complex Networks, Indianapolis, IN, USA, 2017, pp. 183–204.
[53] F. Cauteruccio, Y. Kou, Investigating the emotional experiences in eSports spectatorship: the case of league of legends, Inf. Process. Manag. 60 (6) (2023) 103516. Elsevier.
[54] M. Newman, Networks, 2018. Oxford University Press.
[55] V. Basile, F. Cauteruccio, G. Terracina, How dramatic events can affect emotionality in social posting: the impact of COVID-19 on reddit, Future Internet 13 (2) (2021) 29. MDPI.
[56] J. Massachs, C. Monti, G. De Francisci Morales, F. Bonchi, Roots of trumpism: homophily and social feedback in donald trump support on reddit, in: Proceedings of the 12th ACM Conference on Web Science, 2020, pp. 49–58.
[57] J.J. Duistermaat, J.A.C. Kolk, Distributions: Theory and Applications, Springer, 2010.
[58] M. Birjali, M. Kasri, A. Beni-Hssane, A comprehensive survey on sentiment analysis: approaches, challenges and trends, Knowledge-Based Syst. 226 (2021) 107134. Elsevier.
[59] L. Yue, W. Chen, X. Li, W. Zuo, M. Yin, A survey of sentiment analysis in social media, Knowl. Inf. Syst. 60 (2) (2019) 617–663. Springer.
[60] T. Nam, J. Stromer-Galley, The democratic divide in the 2008 US presidential election, J. Inf. Technol. Polit. 9 (2) (2012) 133–149. Taylor & Francis.
[61] S. Van Gelder, This Changes Everything: Occupy Wall Street and the 99% Movement, Berrett-Koehler Publishers, 2011.
[62] A. Bessi, E. Ferrara, Social bots distort the 2016 US presidential election online discussion, First Monday 21 (11) (2016) n.p.. doi: 10.5210/fm.v21i11.7090. Published 7 Nov 2016.
[63] N. Grinberg, K. Joseph, L. Friedland, B. Swire-Thompson, D. Lazer, Fake news on Twitter during the 2016 US presidential election, Science 363 (6425) (2019) 374–378. American Association for the Advancement of Science.
[64] D.C. Atkinson, Charlottesville and the alt-right: a turning point?, Polit., Groups, Ident. 6 (2) (2018) 309–315. Taylor & Francis.
[65] J. Jakubik, M. Vössing, N. Pröllochs, D. Bär, S. Feuerriegel, Online emotions during the storming of the US capitol: evidence from the social media network Parler, in: Proc. of the International AAAI Conference on Web and Social Media, 17, 2023, pp. 423–434.
[66] G. Weld, M. Glenski, T. Althoff, Political bias and factualness in news sharing across more than 100,000 online communities, in: Proceedings of the International AAAI Conference on Web and Social Media, 15, 2021, pp. 796–807.
[67] C. Monti, J. D'Ignazi, M. Starnini, G. De Francisci Morales, Evidence of demographic rather than ideological segregation in news discussion on reddit, in: Proceedings of the ACM Web Conference 2023, 2023, pp. 2777–2786.
[68] D.W. Hosmer, S. Lemeshow, R.X. Sturdivant, Applied Logistic Regression, Wiley, third ed., 2013.
[69] A. Agresti, Categorical Data Analysis, Wiley, third ed., 2013.
[70] SAS/STAT User's Guide: The LOGISTIC Procedure, SAS Institute Inc., 2004. Wald CIs based on asymptotic normality, https://www.math.wpi.edu/saspdf/stat/chap39.pdf.
[71] M. Nardo, M. Saisana, A. Saltelli, S. Tarantola, A. Hoffman, E. Giovannini, Handbook on Constructing Composite Indicators: Methodology and User Guide, OECD Publishing, Paris, 2008. https://doi.org/10.1787/9789264043466-en
[72] M.L. Anderson, Multiple inference and gender differences in the effects of early intervention: a reevaluation of the abecedarian, perry preschool, and early training projects, J. Am. Stat. Assoc. 103 (484) (2008) 1481–1495. https://doi.org/10.1198/016214508000000841
[73] F.E. Harrell, Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis, Springer, Cham, second ed., 2015. https://doi.org/10.1007/978-3-319-19425-7
[74] M.E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, L. Zettlemoyer, Deep contextualized word representations, in: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), Association for Computational Linguistics, New Orleans, Louisiana, 2018, pp. 2227–2237. https://doi.org/10.18653/v1/N18-1202
[75] J. Devlin, M.W. Chang, K. Lee, K. Toutanova, BERT: pre-training of deep bidirectional transformers for language understanding, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), Volume 1 (Long and Short Papers), Association for Computational Linguistics, Minneapolis, Minnesota, 2019, pp. 4171–4186. https://doi.org/10.18653/v1/N19-1423
[76] N. Reimers, I. Gurevych, Sentence-BERT: sentence embeddings using siamese BERT-networks, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Association for Computational Linguistics, Hong Kong, China, 2019, pp. 3982–3992. https://doi.org/10.18653/v1/D19-1410
[77] T.B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D.M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, D. Amodei, Language models are few-shot learners, in: Advances in Neural Information Processing Systems 33 (NeurIPS 2020), 2020. https://arxiv.org/abs/2005.14165.
[78] F. Gilardi, M. Alizadeh, M. Kubli, ChatGPT outperforms crowd workers for text-annotation tasks, Proc. Natl. Acad. Sci. 120 (30) (2023) e2305016120. https://doi.org/10.1073/pnas.2305016120