TSGAN: Temporal Social Graph Attention Network for Aggressive Behavior Forecasting

Swapnil Mane[1], **Suman Kundu**[1], **Rajesh Sharma**[2,3]
[1]IIT Jodhpur, India
[2]University of Tartu, Estonia
[3]Plaksha University, India
{mane.1, suman}@iitj.ac.in, rajesh.sharma@ut.ee

Abstract

The propagation of aggressive behavior in online social networks presents a growing threat to digital well-being and social harmony. While existing research focuses on modeling aggression diffusion or detecting aggressive content, forecasting individual user aggression remains an open challenge. This work fills this gap by introducing **T**emporal **S**ocial **G**raph **A**ttention **N**etwork (TSGAN), a social-aware sequence-to-sequence architecture designed to forecast aggressive behavior in dynamic social networks. The core of TSGAN is an adaptive socio-temporal attention module that dynamically models social influence and temporal dynamics. To capture global social influence, TSGAN employs a graph contrastive learning approach to generate global network context embeddings. TSGAN utilizes an aggression intensity metric derived from a proposed hybrid aggression content detection model (92.87% F1), combining a fine-tuned transformer with a large language model to quantify user aggression over time. TSGAN uniquely addresses user inactivity, models dynamic follower relationship impacts, and accounts for temporal behavioral decay while scaling to large networks. Experiments on real-world datasets (X for aggression forecasting and Flickr for popularity prediction) demonstrate TSGAN's versatility and effectiveness. TSGAN outperforms baselines in forecasting across hourly, daily, and weekly temporal intervals, showing up to 24.8% improvement in daily aggression predictions.

Introduction

Aggressive behavior on social media has emerged as a critical concern, with substantial implications for both individual well-being and societal harmony. These aggressive interactions can trigger a cascade of negative outcomes, ranging from psychological distress to, in extreme cases, real-world violence (Mishna et al. 2018; Quang-Loc 2021; Vladimirou, House, and Kádár 2021). As the digital landscape continues to evolve, the urgency to develop effective measures to mitigate this challenge has become increasingly apparent (Wong et al. 2022). Predicting future aggressive behavior on social media platforms is crucial for implementing preventive measures against potential escalations of not only online but also offline violence. However, the inherently temporal and relational nature of user behavior on these platforms, where past behavior and social connections strongly influence future conduct (Henneberger, Coffman, and Gest 2017), presents unique challenges in forecasting user aggressive behavior.

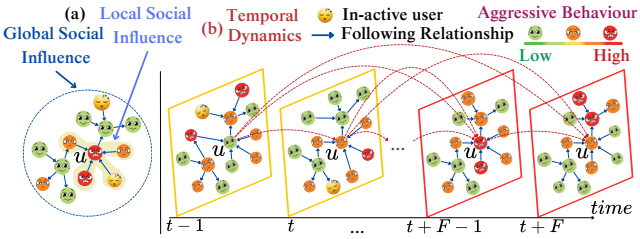

Figure 1: Aggression dynamics in temporal social networks: (a) Users are influenced by both direct (local) and indirect (global) peers. (b) Aggressive behavior evolves over time, with actions linked to both recent and distant past interactions.

The study of temporal graphs has gained significant attention, especially for forecasting tasks in which node and edge features evolve over discrete time steps, often represented as multivariate time series patterns. Several models have been proposed to study these patterns (Yu, Yin, and Zhu 2018; Li et al. 2018; Lai et al. 2018), with further developments focusing on learning node-specific dependency matrices (Bai et al. 2020), long-term forecasting (Wu et al. 2019), and attention-based learning (Guo et al. 2019; Zheng et al. 2020). However, these models, evaluated on smaller networks (325–3848 nodes), failed to scale effectively even for a moderate-size social network. Moreover, these models fail to incorporate social-behavioral dimensions, which are essential for capturing the dynamics of aggressive behavior.

Aggressive behavior often propagates through social circles, influenced by direct (friends/peers) and indirect (friends-of-friends/network) connections (Henneberger, Coffman, and Gest 2017; Vladimirou, House, and Kádár 2021; Quang-Loc 2021; Mane, Kundu, and Sharma 2025; Shankaran and Sharma 2024). Sociological investigations have revealed that individuals with a history of hostility are more likely to engage in future aggressive acts (nonlinear temporal correlation) (Birkley and Eckhardt 2015). Considerable research has focused on the detection of aggressive content (Chen, Yan, and Wong 2020; Srivastava

and Khurana 2019; Pareek et al. 2022; Bansal et al. 2022; Samghabadi et al. 2020; Risch and Krestel 2020) and its diffusion in static networks (Poiitis, Vakali, and Kourtellis 2021; Terizi et al. 2021; Chatzakou et al. 2017). These approaches fail to capture the dynamic nature of social interactions and the temporal evolution of user behavior as illustrated in Figure 1. On the other side, some sociological studies have analyzed aggressive behavior on social media (Vladimirou, House, and Kádár 2021; Quang-Loc 2021), but their findings have not been effectively integrated into predictive computational models (Mane, Kundu, and Sharma 2023). To our knowledge, no existing models are specifically designed for forecasting aggressive behavior on social networks.

The proposed work bridges this gap by introducing the Temporal Social Graph Attention Network (TSGAN), a social-aware seq2seq architecture designed to forecast user aggressive behavior. At its core, we designed an adaptive socio-temporal attention module (ASTAM) that dynamically models local social influence (via active incoming peer interactions) and temporal dynamics (via decay-aware attention to historical behaviors). ASTAM independently processes both historical and future socio-temporal behavioral patterns. For global social context, TSGAN leverages graph contrastive learning to encode static network topology and user attributes. TSGAN utilizes an aggression intensity metric derived from a proposed hybrid aggression content detection model, combining a fine-tuned transformer with a large language model to quantify user aggression over time. This hybrid aggression detection model achieves a 92.87% F1.

TSGAN handles the user inactivity periods in social interactions and the scalability issues inherent in social networks by focusing on sub-graphs and the most common active peers/friends while still considering the global social network context. The model also accounts for the strength of the relationship due to temporal change in context, as well as the temporal decay in user's behaviors (Asur et al. 2011) in its attention mechanism. TSGAN achieves 18%, 24.8%, and 3.2% forecasting improvements over baselines on X (Twitter) for hourly, daily, and weekly aggression predictions, respectively. It demonstrates robustness to label noise (e.g., perturbed aggression scores) and generalizes effectively to non-aggression tasks, achieving state-of-the-art user popularity prediction on Flickr. This underscores TSGAN's cross-platform adaptability and socio-temporal modeling efficacy.

Notations and Problem Formulation

A directed social network is denoted as $G = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ represents the set of $N_{\mathcal{V}} = |\mathcal{V}|$ users and directed edge $\mathcal{E}$ denotes the set of $N_{\mathcal{E}} = |\mathcal{E}|$ following relationships. The relationship captures the directional nature of information flow in social networks, crucial for understanding the propagation of aggressive behavior. Each node and edge is associated with a set of attributes referred to as user and relationship features, respectively. These features are categorized into static and temporal attributes, as described below.

The static user features $X_S^{\mathcal{V}} = (VF, PT, Q)$ encompass attributes that do not change over time. Here $VF_i \in \{0, 1\}$ indicates whether user i's account is verified or not. Verified accounts are generally considered more trustworthy, which can possibly influence user behavior and interactions. Protected $PT_i \in \{0, 1\}$ indicates whether user i's account is protected or not. Protected accounts typically have restricted visibility, affecting the spread and reception of content. Finally, popularity (Q_i), which reflects a user's influence within the network, is the normalized popularity score of a user i, calculated as $Q_i = \frac{F_i - minF}{maxF - minF}$; where F_i is the total number of followers of user i and $minF$ and $maxF$ are the minimum and maximum follower counts.

The temporal user features $X_T^{\mathcal{V}} = (V, Ac, AGI)$ capture attributes that vary over time. Activity $(Ac_i^{t_k}) \in \{0, 1\}$ is a binary indicator, which represents whether a user i is active at time t_k or not. User activity levels are crucial for understanding engagement patterns and their impact on behavior propagation. Virality $(V_i^{t_k})$ for user i at time t_k is calculated as:

$$V_i^{t_k} = \frac{L_i^{t_k} + RT_i^{t_k} + R_i^{t_k} + Q_i^{t_k}}{TP_i^{t_k}} \times \frac{minTP^{t_k} - TP_i^{t_k}}{maxTP^{t_k} - minTP^{t_k}}$$

where $L_i^{t_k}, RT_i^{t_k}, R_i^{t_k}, Q_i^{t_k}$, and, $TP_i^{t_k}$ are likes, retweets, replies, quotes, and total posts, respectively. This metric indicates the reach and impact of the users' posts. Aggression Intensity $AGI_i^{t_k}$ (Mane, Kundu, and Sharma 2025) measures the aggressive activity of a user i at time t_k. This is calculated as:

$$AGI_i^{t_k} = \frac{AG_i^{t_k}}{TP_i^{t_k}} \times \frac{minTP^{t_k} - TP_i^{t_k}}{maxTP^{t_k} - minTP^{t_k}}$$

where $AG_i^{t_k}$ is the total number of aggressive posts. This metric helps profile the level of aggression in user behavior over time. The aggression detection algorithm used in our methodology is described in RQ3.

A static relationship feature $(X_S^{\mathcal{E}})$ constitutes two features. The first is Social Profile Dominance, defined as: $SPD_{ij} = \frac{F_i}{F_i + F_j}$ where F is the number of followers. This metric indicates the relative influence between users based on their follower counts. The other static feature is Network Power Dominance (Poiitis, Vakali, and Kourtellis 2021): $NPD_{ij} = \frac{NP_i}{NP_j}$, where $NP_i = \frac{inDegree_i}{outDegree_i}$ capturing the dominance of user i over user j in terms of influence from and on their neighbors. Topic Similarity $(TS_{ij}^{t_k})$ measure is considered as a temporal relationship feature $(X_T^{\mathcal{E}})$. This is defined by the common topics discussed by users i and j at time t_k, i.e., $TS_{ij}^{t_k} = \frac{n(T_i^{t_k} \cap T_j^{t_k})}{n(T_j^{t_k})}$ where $T_i^{t_k}$ and $T_j^{t_k}$ are the sets of #tags related to topics discussed by users i and j at time t_k, respectively. This metric captures the contextual similarity in relationships between users based on their shared interests over time. We refer to this as a temporal contextual relationship.

Problem Definition. *We formulate the forecasting aggressive behavior problem as a temporal graph prediction problem. Given a social network with $N_{\mathcal{V}}$ users (nodes) and $N_{\mathcal{E}}$ following relationships (directed edges), we consider their historical interactions over H time steps. For each time*

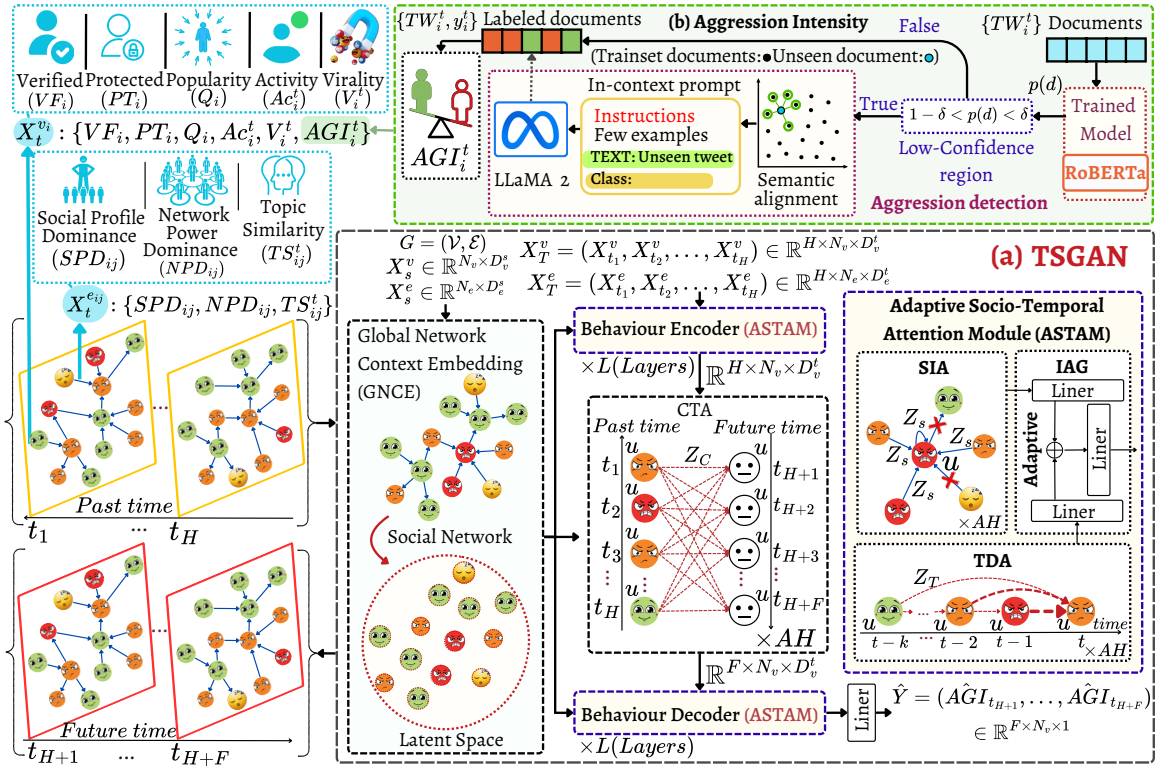

Figure 2: The proposed methodology for forecasting user aggressive behavior includes (a) the architecture of the Temporal Social Graph Attention Network and (b) a hybrid approach for aggression detection.

step, we have vertex features $X_T^{\mathcal{V}} \in \mathbb{R}^{N_{\mathcal{V}} \times D_T^{\mathcal{V}}}$ *representing user behaviors and edge features* $X_T^{\mathcal{E}} \in \mathbb{R}^{N_{\mathcal{E}} \times D_T^{\mathcal{E}}}$ *capturing interaction dynamics. Additionally, we incorporate static features* $X_S^{\mathcal{V}}$ *and* $X_S^{\mathcal{E}}$ *for nodes and edges, respectively, to encapsulate time-invariant user and relationship characteristics. Our objective is to forecast the Aggression Intensity (AGI) for all users over the next timesteps* F*, as* $\hat{Y} \in \mathbb{R}^{F \times N_{\mathcal{V}} \times 1}$.

TSGAN: Temporal Social Graph Attention Network

The proposed TSGAN is a social-aware seq2seq architecture designed to transform historical user behavior sequences to future aggression predictions. As illustrated in Figure 2, historical observations ($X_T^{\mathcal{V}}$, $X_T^{\mathcal{E}}$) are first transformed into latent representations (R, $R_{\mathcal{E}}$) via fully-connected layers. These representations are processed by a behavior encoder with L ASTAM layers, which model peer interactions and decaying historical behaviors. The CTA layer establishes attention-based dependencies between historical representations and future contexts. The behavior decoder refines these future behavior representations (o/p of CTA) through L ASTAM layers, which generate stepwise predictions $\hat{Y} \in \mathbb{R}^{F \times N_{\mathcal{V}} \times 1}$. TSGAN integrates pre-trained global network context (GNCE) via contrastive learning as a static node representation. Training minimizes the mean absolute error as, $L(\Theta) = \frac{1}{F} \sum_{t=t_{H+1}}^{t_{H+F}} \mid Y_t - \hat{Y}_t \mid$. The GNCE, ASTAM, and CTA are explained in detail below.

Global Network Context Embedding (GNCE) GNCE captures the network topology, which is crucial for understanding the dynamics of aggressive behavior. We employ Bootstrapped Graph Latents (BGRL) (Thakoor et al. 2022) to encode nodes into vector representations, preserving the local and global structure information. Given the static features ($X_S^{\mathcal{V}}$, $X_S^{\mathcal{E}}$), BGRL learns node embeddings $\lambda_i \in \mathbb{R}^{D_\lambda}$ for each user $i \in \mathcal{V}$ through a self-supervised contrastive learning approach, where D_λ is the embedding dimensionality. We precomputed these embeddings and integrated them into TSGAN using a ReLU non-linear function to enable co-training with other components. This enables TSGAN to incorporate a broader network context while processing local subgraphs temporally.

Adaptive Socio-Temporal Attention Module (ASTAM) ASTAM jointly models dynamic social influence and temporal decay to capture the evolving aggression patterns in social networks. At layer $(l+1)$, the input consists of latent representations $R^{(l)} \in \mathbb{R}^{N_{\mathcal{V}} \times D}$, along with GNCE $\lambda_i \in \mathbb{R}^{D_\lambda}$ for user i and temporal edge features $\mathbf{r}_{\mathcal{E}_{(i,j)}} \in \mathbb{R}^{D_T^{\mathcal{E}}}$ for relationship (i,j). ASTAM models through two parallel attention mechanisms, Social Influence Attention (SIA) and Temporal Dynamics Attention (TDA). SIA captures peer influence by dynamically prioritizing active users with

strong incoming social ties while incorporating relationship features. TDA captures historical behavioral dependencies through decay-aware attention. To balance the influence of social and temporal factors per user and per time step, the Integration Attention Gateway (IAG) adaptively joins the outputs of SIA ($r^{(l+1)}_{S_{(i,t_h)}}$) and TDA ($r^{(l+1)}_{T_{(i,t_h)}}$).

Social Influence Attention (SIA) Scientific evidence shows that peers' behavior significantly impacts aggressive behavior (Henneberger, Coffman, and Gest 2017; Mane, Kundu, and Sharma 2025). SIA addresses the dynamic and sparse nature of social influence in aggression propagation by adaptively focusing on active peer interactions. Unlike conventional graph attention (e.g., GAT (Velickovic et al. 2017)) that assumes static or uniformly active neighborhoods, SIA prioritizes users who are actively engaging at each timestep. Given temporal user features $\alpha_i \in \mathbb{R}^{D^{\mathcal{V}}_T}$ and relationship features $\beta_{i,j} \in \mathbb{R}^{D^{\mathcal{E}}_T}$, SIA first transforms them into latent representations $r^{(0)}_i$ and $r_{\mathcal{E}_{i,j}}$ using a non-linear function $f(x)$. SIA utilizes $G_{sg} = (\mathcal{V}_{sg}, \mathcal{E}_{sg})$, where $G_{sg} \subseteq G$, to process users with their most common active neighbors in parallel, resulting in computational effectiveness. For user i at timestep t_h, SIA computes the influence of active peers using Eq. 1.

$$r^{(l+1)}_{S_{(i,t_h)}} = \sum_{j \in \mathcal{V}_{sg}} Z^{(l)}_{S_{(i,j),t_h}} \cdot \left(r^{(l)}_{S_{(j,t_h)}} \| \lambda_j \right) \tag{1}$$

This formulation considers only active users, meaning inactive users (those not posting on social media) have limited immediate influence on others' behaviors. The term λ_j represents the GNCE of user j, ensuring that the model incorporates persistent social hierarchies into its predictions. The attention scores are computed by applying scaled dot-product attention using Eq. 2. The $Z^{(l)}_{S_{(i,j),t_h}}$ quantifies the influence of user j on user i.

$$Z^{(l)}_{S_{(i,j),t_h}} = \underbrace{\frac{\exp\left(\langle \mathbf{q}^{(l)}_i, \mathbf{k}^{(l)}_j \rangle / \sqrt{d_{AH}}\right) \cdot \exp\left(\mathbf{r}^{(l)}_{\mathcal{E}_{(i,j),t_h}}\right)}{\sum_{r \in \mathcal{V}_{sg}} \exp\left(\langle \mathbf{q}^{(l)}_i, \mathbf{k}^{(l)}_r \rangle / \sqrt{d_{AH}}\right) \cdot \exp\left(\mathbf{r}^{(l)}_{\mathcal{E}_{(i,r),t_h}}\right)}}_{\text{Social Influence Score}} \tag{2}$$

Here, $\mathbf{q}_i = f^{(l)}_Q(r^{(l)}_{S_{(i,t_h)}} \| \lambda_i)$ and $\mathbf{k}_j = f^{(l)}_K(r^{(l)}_{S_{(j,t_h)}} \| \lambda_j)$ represent the query and key projections for user i and neighbor j, respectively. The term $\mathbf{r}^{(l)}_{\mathcal{E}_{(i,j),t_h}}$ encodes the temporal contextual relationship. The SIA utilizes a multi-head attention mechanism to enhance stability and capture multiple aspects of social influence simultaneously. The attention-based user representation at layer $(l+1)$ is computed using Eq. 3.

$$r^{(l+1)}_{S_{(i,t_h)}} = \Big\|_{ah=1}^{AH} \left(\sum_{j \in \mathcal{V}_{sg}} Z^{(ah,l)}_{S_{(i,j),t_h}} \cdot \mathbf{V}^{(ah,l)}_j \right) W_O \tag{3}$$

Here, $r^{(l+1)}_{S_{(i,t_h)}} \in \mathbb{R}^D$ represents the updated social influence latent representation for user i at time t_h. AH denotes the number of attention heads and $W_O \in \mathbb{R}^{D \times D}$ is a learnable projection matrix that maps the concatenated multi-head outputs into a unified space. The attention weight $Z^{(ah,l)}_{S_{(i,j),t_h}}$ for each head ah is computed using Eq. 4.

$$Z^{(ah,l)}_{S_{(i,j),t_h}} = \frac{\exp\left(\langle \mathbf{Q}^{(ah,l)}_i, \mathbf{K}^{(ah,l)}_j \rangle / \sqrt{d_{AH}}\right) \cdot \exp\left(\mathbf{r}^{(ah,l)}_{\mathcal{E}_{(i,j),t_h}}\right)}{\sum_{r \in \mathcal{V}_{sg}} \exp\left(\langle \mathbf{Q}^{(ah,l)}_i, \mathbf{K}^{(ah,l)}_r \rangle / \sqrt{d_{AH}}\right) \cdot \exp\left(\mathbf{r}^{(ah,l)}_{\mathcal{E}_{(i,r),t_h}}\right)} \tag{4}$$

Here, $\mathbf{Q}^{(ah,l)}_i = W^{(ah,l)}_Q(r^{(l)}_{S_{(i,t_h)}} \| \lambda_i)$, $\mathbf{K}^{(ah,l)}_j = W^{(ah,l)}_K(r^{(l)}_{S_{(j,t_h)}} \| \lambda_j)$, and $\mathbf{V}^{(ah,l)}_j = W^{(ah,l)}_V(r^{(l)}_{S_{(j,t_h)}} \| \lambda_{v_j})$ represent the query, key, and value projections in the multi-head attention. The matrices $W^{(ah,l)}_Q, W^{(ah,l)}_K, W^{(ah,l)}_V \in \mathbb{R}^{d_{AH} \times D}$ are the learnable projection matrices for queries, keys, and values. The scaled dot-product attention normalizes the attention scores using $\sqrt{d_{AH}}$ to ensure stable gradient updates, where the per-head dimensionality is given by $d_{AH} = \frac{D}{AH}$. The outputs from all AH attention heads are then concatenated and projected through a final learnable matrix W_O to produce the updated social influence representation $r^{(l+1)}_{S_{(i,t_h)}}$.

Temporal Dynamics Attention (TDA) Sociological investigations have revealed that individuals with a history of hostility are more likely to engage in future aggressive acts (Birkley and Eckhardt 2015). TDA handles the non-linear decay of behavioral influence in aggression forecasting, where recent actions usually have more influence than older ones, but sometimes past events can unexpectedly become important again (Mane et al. 2025). Unlike standard temporal attention (like in Transformers (Vaswani et al. 2017)), which gives equal importance to all past events, TDA uses an exponential decay factor to focus more on recent behaviors. However, it can also bring back older events if they are important in the current context. For user i at timestep t_h, TDA computes the influence of past behaviors using Eq. 5.

$$r^{(l+1)}_{T_{(i,t_h)}} = \sum_{t_f \in \mathcal{V}_{t_h}} Z^{(l)}_{T_{i,(t_h,t_f)}} \cdot \left(r^{(l)}_{T_{(i,t_f)}} \Big\| \lambda_i \right) \tag{5}$$

Here, $r^{(l+1)}_{T_{(i,t_h)}} \in \mathbb{R}^D$ represents the updated temporal attention representation for user i at time t_h. The set $\mathcal{V}_{t_h}$ includes all preceding time steps, ensuring causality in prediction. A key component of TDA is the exponential decay factor, defined as $\gamma_{(t_h,t_f)} = e^{-\delta \cdot (t_h - t_f)}$, where δ is a learnable decay rate parameter that controls how quickly attention to past behaviors diminishes. The term $t_h - t_f$ represents the time difference between the current time step t_h and a past step t_f. The exponential function ensures that attention weights decrease as the time gap increases, prioritizing recent behaviors while still allowing long-term dependencies to be maintained. In Eq. 6, we integrate temporal decay into attention.

$$Z^{(l)}_{T_{i,(t_h,t_f)}} = \underbrace{\frac{\exp\left(\langle \mathbf{q}^{(l)}_{t_h}, \mathbf{k}^{(l)}_{t_f} \rangle / \sqrt{d_{AH}}\right) \cdot \exp\left(\gamma_{(t_h,t_f)}\right)}{\sum_{t_r \in \mathcal{V}_{t_h}} \exp\left(\langle \mathbf{q}^{(l)}_{t_h}, \mathbf{k}^{(l)}_{t_r} \rangle / \sqrt{d_{AH}}\right) \cdot \exp\left(\gamma_{(t_h,t_r)}\right)}}_{\text{Decay-Modulated Relevance}} \tag{6}$$

Here, query/key projections for current/past steps. TDA employs multi-head attention to capture diverse temporal patterns while integrating behavioral decay. For user i at timestep t_h, the multi-head formulation is defined in Eq. 7.

$$r^{(l+1)}_{T_{(i,t_h)}} = \Big\|_{ah=1}^{AH} \left(\sum_{t_f \in \mathcal{V}_{t_h}} Z^{(ah,l)}_{T_{i,(t_h,t_f)}} \cdot \mathbf{V}^{(ah,l)}_{t_f} \right) W_O, \tag{7}$$

The per-head attention weight incorporates the decay factor (Eq. 8). Here, all parameters are defined as in SIA.

$$Z_{T_{i,(t_h,t_f)}}^{(ah,l)} = \frac{\exp\left(\langle \mathbf{Q}_{t_h}^{(ah,l)}, \mathbf{K}_{t_f}^{(ah,l)} \rangle / \sqrt{d_{AH}}\right) \cdot \exp\left(\gamma_{(t_h,t_f)}\right)}{\sum_{t_r \in \mathcal{V}_{t_h}} \exp\left(\langle \mathbf{Q}_{t_h}^{(ah,l)}, \mathbf{K}_{t_r}^{(ah,l)} \rangle / \sqrt{d_{AH}}\right) \cdot \exp\left(\gamma_{(t_h,t_r)}\right)} \tag{8}$$

Further, the Integration Attention Gateway (IAG) dynamically resolves the socio-temporal duality in aggression forecasting by adaptively integrating localized social influence and historical behavioral patterns. The mechanism utilized a learnable gating vector $g = \sigma(W_{g,1} R_S^{(l+1)} + W_{g,2} R_T^{(l+1)} + b_g)$, where $g \in [0,1]^{N_\mathcal{V}}$ balances peer-driven social signals ($R_S^{(l+1)}$) and decay-aware temporal dynamics ($R_T^{(l+1)}$) using $R^{(l+1)} = g \odot R_S^{(l+1)} + (1-g) \odot R_T^{(l+1)}$. Unlike static aggregation, IAG's context-sensitive fusion enables granular adaptation: for socially active users (e.g., those engaged in hostile interactions), $g \to 1$ prioritizes peer influence, while isolated users (e.g., with recurrent aggression) rely on historical patterns ($g \to 0$). This flexibility is critical given sociological evidence that the cause of aggression varies across individuals—some driven by real-time social triggers, others by latent behavioral histories (Terizi et al. 2021).

Cross-temporal attention (CTA) addresses the challenges of long-term aggressive behavior prediction by capturing non-linear and time-decoupled dependencies. CTA integrates past user behavior into predictions by directly linking historical and future time steps. This allows the model to adaptively prioritize distant yet influential past behaviors. This is valuable in aggressive behavior forecasting, where behavior patterns may not follow simple linear progressions and distant past events could suddenly become relevant (Birkley and Eckhardt 2015; Zheng et al. 2020). For user i, CTA converts encoded past behavior representations $r_{C_{(i,t_h)}}^{(l)}$ into future representations $r_{C_{(i,t_f)}}^{(l+1)}$ through Eq. 9, which are subsequently utilized as input for the behavior decoder.

$$r_{C_{(i,t_f)}}^{(l+1)} = \sum_{t_h=t_1}^{t_H} \underbrace{\frac{\exp\left(\langle \mathbf{q}_{t_f}, \mathbf{k}_{t_h} \rangle / \sqrt{d_{AH}}\right)}{\sum_{t_r=t_1}^{t_H} \exp\left(\langle \mathbf{q}_{t_f}, \mathbf{k}_{t_r} \rangle / \sqrt{d_{AH}}\right)}}_{\text{Adaptive Temporal Relevance}} \cdot r_{C_{(i,t_h)}}^{(l)} \tag{9}$$

To capture diverse aggression trajectories, CTA uses multi-head attention (Eq. 10). $Z_{C(i,(t_f,t_h))}^{(l)}$ represents the normalized attention weight.

$$r_{C_{(i,t_f)}}^{(l+1)} = \Big\Vert_{ah=1}^{AH} \left(\sum_{t_h=t_1}^{t_H} Z_{C_{i,(t_f,t_h)}}^{(ah,l)} \cdot f_{C_V}^{(ah,l)} \left(r_{C_{(i,t_h)}}^{(l)} \right) \right) \tag{10}$$

Here, the attention utilizes pre-computed GNCE for future time steps to maintain a realistic forecasting approach. This acknowledges that the future states of users are unknown.

Experiments

We conducted experiments on two real-world datasets in order to evaluate the proposed TSGAN. Our investigation focused on three key research questions: (RQ1) How does TSGAN's performance compare to baseline methods in behavior forecasting?, (RQ2) What is the contribution of each TSGAN submodule to capturing temporal social influence correlations, and how is TSGAN robust for aggression label noise?, and (RQ3) How effective is our hybrid aggression content detection model compared to existing approaches?

Experimental Setup

We configured TSGAN to ensure optimal performance and reproducibility. The model's weights were initialized using the Xavier uniform method. After extensive tuning, we determined that 8 AH with a dimensionality (d_{AH}) of 8, resulting in a total dimension $D = 64$, produced the best results. Trained the model for 200 epochs with a weight decay of 0.18 and implemented early stopping with the patience of 100 epochs. Both the encoder and decoder utilized the ASTAM layer ($L = 1$). The GNCE was $D_\lambda = 512$, and temporal decay rate was $\delta = 0.0001$. For optimization, we employed the Adam optimizer with a learning rate of 1e-3. Followed a chronological data split (80% train/10% test/10% validation) to maintain temporal consistency. We used Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) as our performance metrics, averaging results over 5 randomized trials with seeds 100 to 500 by 100 steps. We present our results with means and standard deviations.

Aggressive Behavior Forecasting Performance Evaluation (RQ1)

Dataset We evaluated the performance and versatility of TSGAN on two complementary datasets. The first is an X dataset (Mane, Kundu, and Sharma 2025), containing 30,307 user nodes and 505,938 edges, which captures English-language interactions over a six-month period. This dataset is used to forecast user aggressive behavior. The second is the Social Media Prediction Dataset (SMPD) (Wu et al. 2016), derived from Flickr, consisting of 38,312 nodes and 312,063 edges based on shared content categories over sixteen months. SMPD is used to forecast user popularity.

Baselines We conducted comparisons against a diverse array of baseline models to rigorously evaluate TSGAN's performance. These ranged from traditional statistical methods to advanced deep learning approaches, including both time series and graph-based models. Our baselines included: ARIMA (Makridakis and Hibon 1997) for classical time series forecasting; RNN (Werbos 1990), GRU (Cho et al. 2014), and LSTM (Hochreiter and Schmidhuber 1997) as fundamental deep learning architectures; LSTNet (Lai et al. 2018), which combines CNNs and RNNs for multi-scale temporal pattern modeling; STGCN (Yu, Yin, and Zhu 2018), integrating graph convolutions with temporal convolutions; and DCRNN (Li et al. 2018), utilizing diffusion convolutions in a recurrent framework. While our task presents unique challenges, these models have demonstrated efficacy in related multivariate time series and graph-based prediction tasks. Notably, we attempted to include more graph-based models such as ASTGCN (Guo et al. 2019), AGCRN (Bai et al. 2020), Graph WaveNet (Wu et al. 2019), GMAN (Zheng et al. 2020), and MTGNN (Wu et al. 2020). However, these models, typically evaluated on smaller networks (325-3848 nodes), failed to scale effectively to our

		Models	1H → 1F		2H → 2F		3H → 3F		4H → 4F	
			MAE ± s.d.	RMSE ± s.d.	MAE ± s.d.	RMSE ± s.d.	MAE ± s.d.	RMSE ± s.d.	MAE ± s.d.	RMSE ± s.d.
X Dataset	Hourly	ARIMA	0.928 ± 0.000	0.928 ± 0.000	0.931 ± 0.000	0.950 ± 0.000	0.922 ± 0.000	0.950 ± 0.000	0.926 ± 0.000	0.957 ± 0.000
		GRU	0.922 ± 0.000	0.959 ± 7.055	0.921 ± 0.000	0.959 ± 6.788	0.921 ± 0.000	0.959 ± 8.322	0.921 ± 0.000	0.959 ± 8.041
		RNN	0.911 ± 0.001	0.947 ± 0.001	0.911 ± 0.002	0.947 ± 0.001	0.913 ± 0.002	0.948 ± 0.001	0.914 ± 0.002	0.949 ± 0.001
		LSTM	0.922 ± 0.000	0.960 ± 3.019	0.922 ± 0.000	0.959 ± 5.371	0.922 ± 0.000	0.959 ± 6.829	0.922 ± 3.435	0.959 ± 3.007
		LSTNet	0.951 ± 0.077	1.006 ± 0.120	0.936 ± 0.018	0.970 ± 0.021	0.902 ± 0.028	0.933 ± 0.034	0.886 ± 0.027	0.914 ± 0.028
		STGCN	0.929 ± 0.004	0.961 ± 0.001	0.922 ± 0.001	0.959 ± 0.001	0.926 ± 0.004	0.960 ± 0.001	0.926 ± 0.003	0.960 ± 0.001
		DCRNN	0.999 ± 0.003	0.999 ± 0.001	0.998 ± 0.023	0.998 ± 0.038	0.995 ± 0.004	0.995 ± 0.004	0.969 ± 0.008	0.971 ± 0.007
		TSGAN	**0.029 ± 0.001**	**0.112 ± 0.001**	**0.011 ± 0.004**	**0.106 ± 0.001**	**0.025 ± 0.001**	**0.106 ± 0.002**	**0.020 ± 0.002**	**0.107 ± 0.001**
	Daily	ARIMA	0.939 ± 0.000	0.939 ± 0.000	0.921 ± 0.000	0.924 ± 0.000	0.920 ± 0.000	0.924 ± 0.000	0.914 ± 0.000	0.919 ± 0.000
		GRU	0.445 ± 0.000	0.652 ± 0.000	0.433 ± 0.001	0.646 ± 0.000	0.427 ± 0.001	0.642 ± 0.000	0.430 ± 0.001	0.643 ± 0.001
		RNN	0.378 ± 0.006	0.594 ± 0.006	0.383 ± 0.005	0.600 ± 0.004	0.378 ± 0.004	0.597 ± 0.004	0.378 ± 0.002	0.597 ± 0.002
		LSTM	0.445 ± 0.000	0.653 ± 0.000	0.434 ± 0.000	0.646 ± 0.000	0.428 ± 0.001	0.642 ± 0.000	0.430 ± 0.001	0.644 ± 0.001
		LSTNet	0.540 ± 0.204	0.649 ± 0.162	0.495 ± 0.160	0.663 ± 0.090	0.619 ± 0.210	0.739 ± 0.129	0.641 ± 0.086	0.715 ± 0.060
		STGCN	0.476 ± 0.128	0.659 ± 0.053	0.422 ± 0.021	0.632 ± 0.004	0.436 ± 0.047	0.632 ± 0.009	0.422 ± 0.019	0.630 ± 0.004
		DCRNN	0.991 ± 0.009	0.991 ± 0.009	0.993 ± 0.007	0.994 ± 0.006	0.976 ± 0.032	0.980 ± 0.025	0.980 ± 0.019	0.983 ± 0.016
		TSGAN	**0.138 ± 0.003**	**0.295 ± 0.001**	**0.128 ± 0.005**	**0.284 ± 0.001**	**0.126 ± 0.002**	**0.278 ± 0.001**	**0.133 ± 0.002**	**0.277 ± 0.000**
	Weekly	ARIMA	0.973 ± 0.000	0.973 ± 0.000	0.970 ± 0.000	0.971 ± 0.000	0.967 ± 0.000	0.970 ± 0.000	0.962 ± 0.000	0.965 ± 0.000
		GRU	0.786 ± 0.001	0.797 ± 0.001	0.772 ± 0.001	0.784 ± 0.001	0.724 ± 0.089	0.746 ± 0.073	0.772 ± 0.003	0.786 ± 0.003
		RNN	0.245 ± 0.006	0.456 ± 0.007	0.239 ± 0.010	0.450 ± 0.012	0.250 ± 0.010	0.455 ± 0.012	0.260 ± 0.002	0.469 ± 0.002
		LSTM	0.786 ± 0.001	0.798 ± 0.001	0.773 ± 0.001	0.784 ± 0.001	0.770 ± 0.001	0.782 ± 0.001	0.776 ± 0.002	0.778 ± 0.002
		LSTNet	0.454 ± 0.265	0.530 ± 0.235	0.409 ± 0.188	0.539 ± 0.151	0.449 ± 0.174	0.557 ± 0.144	0.405 ± 0.259	0.526 ± 0.204
		STGCN	0.239 ± 0.007	0.440 ± 0.001	0.216 ± 0.005	0.420 ± 0.001	0.233 ± 0.069	0.428 ± 0.021	0.229 ± 0.032	0.432 ± 0.001
		DCRNN	0.981 ± 0.023	0.984 ± 0.019	0.985 ± 0.006	0.986 ± 0.005	0.984 ± 0.012	0.985 ± 0.010	0.971 ± 0.014	0.975 ± 0.011
		TSGAN	**0.218 ± 0.023**	**0.355 ± 0.002**	**0.199 ± 0.019**	**0.347 ± 0.001**	**0.201 ± 0.014**	**0.351 ± 0.003**	**0.213 ± 0.017**	**0.352 ± 0.002**
SMPD	Weekly	ARIMA	1.013 ± 0.000	1.013 ± 0.000	1.006 ± 0.000	1.012 ± 0.000	1.001 ± 0.000	1.008 ± 0.000	0.996 ± 0.000	1.005 ± 0.000
		GRU	0.437 ± 0.001	0.467 ± 0.001	0.441 ± 0.001	0.471 ± 0.001	0.441 ± 0.001	0.470 ± 0.001	0.439 ± 0.002	0.469 ± 0.001
		RNN	0.367 ± 0.002	0.413 ± 0.002	0.371 ± 0.002	0.415 ± 0.001	0.373 ± 0.005	0.417 ± 0.003	0.378 ± 0.005	0.421 ± 0.003
		LSTM	0.444 ± 0.004	0.473 ± 0.003	0.445 ± 0.001	0.474 ± 0.001	0.440 ± 0.002	0.469 ± 0.002	0.439 ± 0.002	0.469 ± 0.002
		LSTNet	0.402 ± 0.024	0.441 ± 0.017	0.445 ± 0.001	0.475 ± 0.001	0.445 ± 0.001	0.475 ± 0.001	0.446 ± 0.001	0.475 ± 0.006
		STGCN	0.448 ± 0.001	0.476 ± 0.001	0.449 ± 0.001	0.478 ± 0.000	0.449 ± 0.001	0.477 ± 0.001	0.449 ± 0.001	0.477 ± 0.001
		DCRNN	0.450 ± 5.433	0.478 ± 5.126	0.451 ± 2.27	0.479 ± 2.146	0.452 ± 5.735	0.480 ± 5.411	0.451 ± 7.042	0.480 ± 6.624
		TSGAN	**0.114 ± 0.043**	**0.161 ± 0.031**	**0.115 ± 0.048**	**0.163 ± 0.035**	**0.100 ± 0.040**	**0.149 ± 0.026**	**0.102 ± 0.040**	**0.151 ± 0.027**

Table 1: Performance comparison of TSGAN with baselines. Reports average performance (mean ± standard deviation) over five runs for different historical time steps ('H') to predict future time steps ('F').

moderate-size social network of 30,307 nodes.

Forecasting Performance Table 1 presents comparative results for timesteps F = 1, 2, 3, and 4 weeks, days, and hours ahead predictions. Due to space constraints, we included only the weekly results for the SMPD dataset. The results reveal an interesting pattern in modeling aggressive behavior. Traditional deep learning approaches (GRU, RNN, LSTM, LSTNet), while designed for capturing complex patterns, struggle with this high sparse temporal data as they are typically optimized for and evaluated on dense time series. Graph-based models, particularly STGCN, show improved performance over sequence models by incorporating network structure, though they still face challenges with temporal sparsity. TSGAN effectively addresses these challenges through (i) ASTAM's focus on active users, which handles temporal sparsity by dynamically attending to relevant social signals, and (ii) IAG's adaptive weighting mechanism that balances temporal and social influences. This design enables TSGAN to achieve state-of-the-art performance on both datasets, with significant improvements over baselines across all experiments. The model shows pronounced results in both short-term and long-term predictions. To validate our findings statistically, we conducted a t-test comparing TSGAN to STGCN (the next DL best performer) across all time intervals, with a p-value < 0.01 confirming TSGAN's significant performance edge.

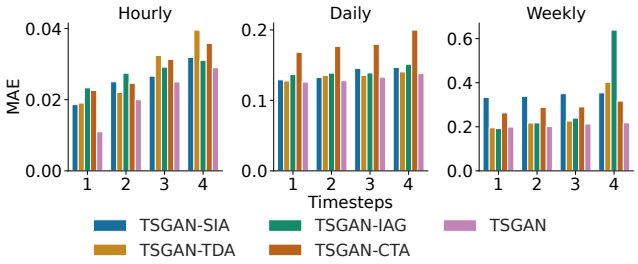

Figure 3: TSGAN submodule contribution analysis.

Ablation Studies (RQ2)

TSGAN Submodule Contribution Analysis We conducted an ablation study to evaluate the contribution of each submodule in TSGAN. We created four variants by removing key submodules: TSGAN-SIA (without SIA), TSGAN-TDA, TSGAN-IAG, and TSGAN-CTA. These variants were tested against the full TSGAN model for predictions ranging from 1 to 4 steps ahead, across hourly, daily, and weekly time horizons. The results presented in Figure 3 clearly demonstrate each module's significance. The full TSGAN model consistently outperformed all ablated variants across all prediction steps and time horizons. These findings validate model design choices and demonstrate that each submodule contributes to TSGAN's forecast performance.

Robustness to Aggression Detection Noise To evaluate TSGAN's resilience against label noise inherent in real-world

aggression detection, we introduced synthetic perturbations to historical AGI values. For noise ratios $\vartheta \in [0.1, 0.6]$, we replaced $\vartheta \times 100\%$ of AGI values with uniform noise $\mathcal{U}(0, 1)$, simulating misclassifications from imperfect detectors. As shown in Figure 4, TSGAN maintained superior forecasting accuracy across all horizons (aggregated results), even at 60% noise.

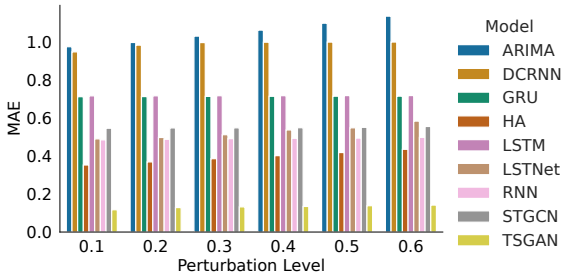

Figure 4: Model robustness for aggression detection noise.

Aggression Content Detection & Evaluation (RQ3)

We developed a hybrid approach for aggression content detection that combines fine-tuned transformer and LLM. Our method builds upon a transformer-based model as recent advancements in aggression detection using transformers. We considered more recent models as baselines, including the BERT-base (Akter et al. 2022), the RoBERTa-base (Rawat et al. 2023), the RoBERTa-large (Mane, Kundu, and Sharma 2025) and (Ghosh et al. 2023) XLMRoBERTa model for the identification of aggressive content. Our hybrid approach utilizes RoBERTa-large, fine-tuned on the dataset from (Mane, Kundu, and Sharma 2025), and enhances it with the advanced reasoning capabilities of LLaMa-2 for ambiguous cases. The approach works in two stages, as detailed in Algorithm 1. First, the fine-tuned RoBERTa-large model namely AG-BERT, processes all documents. For a document d, the model outputs a probability score $p(d) \in [0, 1]$ for the aggressive class. If $p(d) \geq 1 - \delta$ or $p(d) \leq \delta$, where δ is a confidence threshold, the document is classified as aggressive or non-aggressive, respectively. For low-confidence predictions where $1 - \delta < p(d) < \delta$, we employ LLaMa-2 with in-context prompting, utilizing semantic alignment to construct relevant prompts. The algorithm obtains the document embedding $\mathbf{v}_d$ using AG-BERT and selects the top k most contextually similar documents from the training set ($\mathcal{D}_{train}$) along with their labels $\mathcal{S}$. It then constructs a prompt P using d and the selected similar documents, and LLaMa-2 predicts the final aggression label y by maximizing the probability $P(c|P; \theta_L)$ for class c. This approach balances computational efficiency with accuracy, which is crucial for real-time processing of large-scale social network data.

Algorithm 1: Aggression Content Detection

Require: Document d, Fine-tuned RoBERTa model R, LLaMa-2 model L, Confidence threshold δ, Number of contextual similar documents k
Ensure: Aggression label $y \in \{0, 1\}$

$p(d) \leftarrow R(d)$ ▷ RoBERTa prediction
if $p(d) \geq 1 - \delta$ **then**
  return $y \leftarrow 1$ ▷ Aggressive
else if $p(d) \leq \delta$ **then**
  return $y \leftarrow 0$ ▷ Non-aggressive
else
  $\mathbf{v}_d \leftarrow R_{embed}(d) \in \mathbb{R}^m$ ▷ Get document embedding
  $\mathcal{S} \leftarrow \text{TopK}(\{\frac{\mathbf{v}_d \cdot \mathbf{v}_i}{\|\mathbf{v}_d\| \|\mathbf{v}_i\|} : d_i \in \mathcal{D}_{train}\}, k)$
  $P \leftarrow \text{ConstructPrompt}(d, \mathcal{S})$
  $y \leftarrow \arg\max_{c \in \{0,1\}} P(c|P; \theta_L)$ ▷ LLaMa-2 prediction
  return y
end if

Models	Baseline Models		
	Precision	Recall	F1
Akter et al. (2022)	0.8628	0.8613	0.862
Ghosh et al. (2023)	0.8708	0.8748	0.862
Rawat et al. (2023)	0.9004	0.9035	0.9016
Mane et al. (2023)	0.9172	0.9199	0.9185
	LLM Ablations		
AG-BERT + Falcon	0.9267	0.9282	0.9274
AG-BERT + Gemma	0.9200	0.9222	0.921
AG-BERT + Llama-3	0.9238	0.9263	0.9249
Ours	**0.9277**	**0.9300**	**0.9287**

Table 2: Performance evaluation of aggression content detection.

Evaluation Our evaluation demonstrates the effectiveness of this hybrid approach. As shown in Table 2, we achieved the highest performance compared to baseline models. In the ablation study, we evaluated our AG-BERT model against several LLMs, including Llama-2-7B, Llama-3-8B, Falcon-7B, and Gemma-1.1-7B, using both semantically enriched few-shot (in-context) prompting and zero-shot prompting. Our findings reveal that in-context prompting consistently outperformed zero-shot settings across all models. Notably, the combination of Llama-2 with in-context prompting and AG-BERT achieved the highest performance among the tested LLMs.

Conclusion

We introduced TSGAN, a novel Temporal Social Graph Attention Network designed to forecast aggressive behavior in social media. Through extensive experiments on real-world datasets, TSGAN consistently outperformed state-of-the-art baselines in both short-term and long-term predictions. Moreover, TSGAN's consistent excellence across different datasets and prediction tasks not only validates its effectiveness in addressing specific behavioral challenges but also underscores its potential for broader applications in social network analysis. Furthermore, our hybrid aggression content detection model highlights the potential for more nuanced content analysis in behavioral prediction tasks. This superior performance is particularly valuable for practical applications, as it provides social media platforms with extended lead time to implement preventive measures against potential escalations of online aggression.

Acknowledgments

This research was partially supported by the Prime Minister Research Fellowship funded by the Ministry of Education (MoE), India. Suman Kundu would like to acknowledge grant no. 4(2)/2024-ITEA of MeitY, GoI; and Srijan: Center for Generative AI (grant no. ET/23/2024-ET) of MeitY under the IndiaAI mission with the support of Meta for partial support. Rajesh Sharma is supported by EU H2020 program under the SoBigData++ project (grant agreement No. 871042), and partially funded by CHIST-ERA project HAMISON.

References

Akter, M.; Shahriar, H.; Ahmed, N.; and Cuzzocrea, A. 2022. Deep Learning Approach for Classifying the Aggressive Comments on Social Media: Machine Translated Data Vs Real Life Data. In *2022 IEEE International Conference on Big Data (Big Data)*, 5646–5655. Los Alamitos, CA, USA: IEEE Computer Society.

Asur, S.; Huberman, B. A.; Szabo, G.; and Wang, C. 2011. Trends in social media: Persistence and decay. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 5, 434–437.

Bai, L.; Yao, L.; Li, C.; Wang, X.; and Wang, C. 2020. Adaptive graph convolutional recurrent network for traffic forecasting. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS '20. Red Hook, NY, USA: Curran Associates Inc. ISBN 9781713829546.

Bansal, V.; Tyagi, M.; Sharma, R.; Gupta, V.; and Xin, Q. 2022. A Transformer Based Approach for Abuse Detection in Code Mixed Indic Languages. *ACM Transactions on Asian and low-resource language information processing*.

Birkley, E. L.; and Eckhardt, C. I. 2015. Anger, hostility, internalizing negative emotions, and intimate partner violence perpetration: A meta-analytic review. *Clinical psychology review*, 37: 40–56.

Chatzakou, D.; Kourtellis, N.; Blackburn, J.; Cristofaro, E. D.; Stringhini, G.; and Vakali, A. 2017. Mean Birds: Detecting Aggression and Bullying on Twitter. *Proceedings of the 2017 ACM on Web Science Conference*.

Chen, J.; Yan, S.; and Wong, K.-C. 2020. Verbal aggression detection on Twitter comments: convolutional neural network for short-text sentiment analysis. *Neural Computing and Applications*, 32: 10809–10818.

Cho, K.; van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; and Bengio, Y. 2014. Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. In Moschitti, A.; Pang, B.; and Daelemans, W., eds., *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1724–1734. Doha, Qatar: Association for Computational Linguistics.

Ghosh, S.; Priyankar, A.; Ekbal, A.; and Bhattacharyya, P. 2023. A transformer-based multi-task framework for joint detection of aggression and hate on social media data. *Natural Language Engineering*, 1–21.

Guo, S.; Lin, Y.; Feng, N.; Song, C.; and Wan, H. 2019. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 922–929.

Henneberger, A. K.; Coffman, D. L.; and Gest, S. D. 2017. The effect of having aggressive friends on aggressive behavior in childhood: Using propensity scores to strengthen causal inference. *Social Development*, 26(2): 295–309.

Hochreiter, S.; and Schmidhuber, J. 1997. Long Short-Term Memory. *Neural Comput.*, 9(8): 1735–1780.

Lai, G.; Chang, W.-C.; Yang, Y.; and Liu, H. 2018. Modeling Long- and Short-Term Temporal Patterns with Deep Neural Networks. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, SIGIR '18, 95–104. New York, NY, USA: Association for Computing Machinery. ISBN 9781450356572.

Li, Y.; Yu, R.; Shahabi, C.; and Liu, Y. 2018. Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting. In *International Conference on Learning Representations (ICLR '18)*.

Makridakis, S.; and Hibon, M. 1997. ARMA models and the Box–Jenkins methodology. *Journal of forecasting*, 16(3): 147–163.

Mane, S.; Kundu, S.; and Sharma, R. 2023. A Survey on Online Aggression: Content Detection and Behavioural Analysis on Social Media Platforms. *ACM Computing Surveys*.

Mane, S.; Kundu, S.; and Sharma, R. 2025. You are what your feeds make you: A study of user aggressive behavior on Twitter. *Applied Intelligence*, 55(6): 385.

Mishna, F.; Regehr, C.; Lacombe-Duncan, A.; Daciuk, J.; Fearing, G.; and Van Wert, M. 2018. Social media, cyberaggression and student mental health on a university campus. *Journal of mental health*, 27(3): 222–229.

Pareek, K.; Choudhary, A.; Tripathi, A.; Mishra, K.; and Mittal, N. 2022. Hate and Aggression Detection in Social Media Over Hindi English Language. *International Journal of Software Science and Computational Intelligence (IJSSCI)*, 14(1): 1–20.

Poiitis, M.; Vakali, A.; and Kourtellis, N. 2021. On the Aggression Diffusion Modeling and Minimization in Twitter. *ACM Trans. Web*, 16(1).

Quang-Loc, N. 2021. Some thoughts on Vietnamese aggression and violence due to social media effects.

Rawat, A.; Nafis, N.; Bhadane, D.; Kanojia, D.; and Murthy, R. 2023. Modelling Political Aggression on Social Media Platforms. In *Proceedings of the 13th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis*, 497–510.

Risch, J.; and Krestel, R. 2020. Bagging BERT models for robust aggression identification. In *Proceedings of the Second Workshop on Trolling, Aggression and Cyberbullying*, 55–61.

Samghabadi, N. S.; Patwa, P.; Pykl, S.; Mukherjee, P.; Das, A.; and Solorio, T. 2020. Aggression and misogyny detection using BERT: A multi-task approach. In *Proceedings of the second workshop on trolling, aggression and cyberbullying*, 126–131.

Shankaran, V.; and Sharma, R. 2024. Analyzing Toxicity in Deep Conversations: A Reddit Case Study. *arXiv preprint arXiv:2404.07879*.

Srivastava, S.; and Khurana, P. 2019. Detecting Aggression and Toxicity using a Multi Dimension Capsule Network. 157–162. Association for Computational Linguistics (ACL).

Terizi, C.; Chatzakou, D.; Pitoura, E.; Tsaparas, P.; and Kourtellis, N. 2021. Modeling aggression propagation on social media. *Online Social Networks and Media*, 24: 100137.

Thakoor, S.; Tallec, C.; Azar, M. G.; Azabou, M.; Dyer, E. L.; Munos, R.; Veličković, P.; and Valko, M. 2022. Large-scale representation learning on graphs via bootstrapping. *International Conference on Learning Representations (ICLR)*.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L.; and Polosukhin, I. 2017. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, 6000–6010. Red Hook, NY, USA: Curran Associates Inc. ISBN 9781510860964.

Velickovic, P.; Cucurull, G.; Casanova, A.; Romero, A.; Lio, P.; Bengio, Y.; et al. 2017. Graph attention networks. *stat*, 1050(20): 10–48550.

Vladimirou, D.; House, J.; and Kádár, D. Z. 2021. Aggressive complaining on social media: the case of# MuckyMerton. *Journal of Pragmatics*, 177: 51–64.

Werbos, P. J. 1990. Backpropagation through time: what it does and how to do it. *Proceedings of the IEEE*, 78(10): 1550–1560.

Wong, N.; Yanagida, T.; Spiel, C.; and Graf, D. 2022. The association between appetitive aggression and social media addiction mediated by cyberbullying: the moderating role of inclusive norms. *International journal of environmental research and public health*, 19(16): 9956.

Wu, B.; Cheng, W.-H.; Zhang, Y.; and Mei, T. 2016. Time matters: Multi-scale temporalization of social media popularity. In *Proceedings of the 24th ACM international conference on Multimedia*.

Wu, Z.; Pan, S.; Long, G.; Jiang, J.; Chang, X.; and Zhang, C. 2020. Connecting the Dots: Multivariate Time Series Forecasting with Graph Neural Networks. *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*.

Wu, Z.; Pan, S.; Long, G.; Jiang, J.; and Zhang, C. 2019. Graph WaveNet for Deep Spatial-Temporal Graph Modeling. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, 1907–1913. International Joint Conferences on Artificial Intelligence Organization.

Yu, B.; Yin, H.; and Zhu, Z. 2018. Spatio-temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI)*.

Zheng, C.; Fan, X.; Wang, C.; and Qi, J. 2020. GMAN: A Graph Multi-Attention Network for Traffic Prediction. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(01): 1234–1241.