TGN-SMM: A Temporal Graph-Based Shared Mental Model for Modelling Evolving Team Dynamics in Multi-Party Meetings

Moinam Chatterjee[1], Behnam Ensan[1], Ayoub Al-Hamadi[2], Andreas Wendemuth[1]

[1]Chair of Cognitive Systems, Otto-von-Guericke University, Germany
[2]Neuro-Information Technology Group, Otto-von-Guericke University, Germany

moinam.chatterjee@ovgu.de, behnam.ensan@ovgu.de, ayoub.al-hamadi@ovgu.de, andreas.wendemuth@ovgu.de

1. Motivation and Objectives

Motivation

- **AI teammates need more than context; they need cognition**. Effective proactive assistance in Human-Agent Teams (HATs) requires agents to map not just what is said, but the underlying shared cognitive states driving team decisions.
- **Shared Mental Models (SMMs) are the missing link**. SMMs capture the common understanding of tasks, roles, and interpersonal dynamics that underlie effective teamwork, but current computational models treat them as static end-states, not living processes.
- **Team dynamics are inherently temporal**. Real teams converge and diverge through moments of support, objection, and uncertainty, signals that shift continuously across a meeting, not just at its conclusion.
- **Existing approaches miss the fine-grained evolution**. Prior frameworks infer SMMs from static snapshots or global behavioural summaries, overlooking how interpersonal alignment actually builds, breaks, and rebuilds over time.

Research Goal

We aim to develop a temporal-relational framework that models the temporal evolution of Shared Mental Models (SMMs) in multi-party human teams, moving beyond static, end-state representations of team cognition. The goal is to capture how team members' cognitive states and interpersonal alignments dynamically emerge, shift, and stabilise as collaboration unfolds, enabling AI agents to better understand and support real-time team decision-making in Human-Agent Teaming contexts.

Research Objectives

1. **Model team cognition as a temporal process** to represent multi-party meetings as an evolving speaker-interaction graph rather than a static snapshot of communication.
2. **Characterise team members' cognitive states** by encoding each speaker node using the context of their utterances, team role, and speech intention to reflect their evolving cognitive state.
3. **Validate on real-world meeting data** by operationalising annotated adjacency pairs from the AMI Meeting Corpus as temporal events to train and evaluate the framework end-to-end.
4. **Benchmark against static and temporal baselines** by assessing whether temporal graph reasoning improves the detection of interpersonal alignment, particularly for under-represented alignment classes.

2. Dataset Setup

Dataset Summary

- AMI Meeting Corpus: benchmark for multi-party conversational analysis, 100 hours of recordings containing both natural and scenario-driven meetings capturing naturalistic group interactions.
- In this study, only scenario-driven meetings are used, where the meetings have been decomposed into ordered adjacency pairs annotated with speaker role, dialogue act, speech text, and alignment classes.
- Corpus split into four independent sessions: *a*, *b*, *c*, *d*. Each session is taken as an independent meeting in this study.
- An overlap analysis between Support/Positive Assessment and Partial agreement/support showed substantial feature-space overlap, as seen in Figure 2, indicating that Partial agreement/support was neither a stable nor a distinct category and was frequently absorbed into the support class. As a result, the Partial agreement/support label was removed before data splitting.

Alignment Classes

There are four alignment classes in the AMI Meeting Corpus:

- **Support/Positive Assessment**: clear agreement or endorsement of another speaker's contribution.
- **Partial agreement/support**: qualified or conditional agreement, reflecting incomplete alignment.
- **Objection/Negative Assessment**: disagreement, pushback, or critical evaluation of a prior contribution.
- **Uncertain**: ambiguous or non-committal responses that signal hesitation or lack of clear stance.

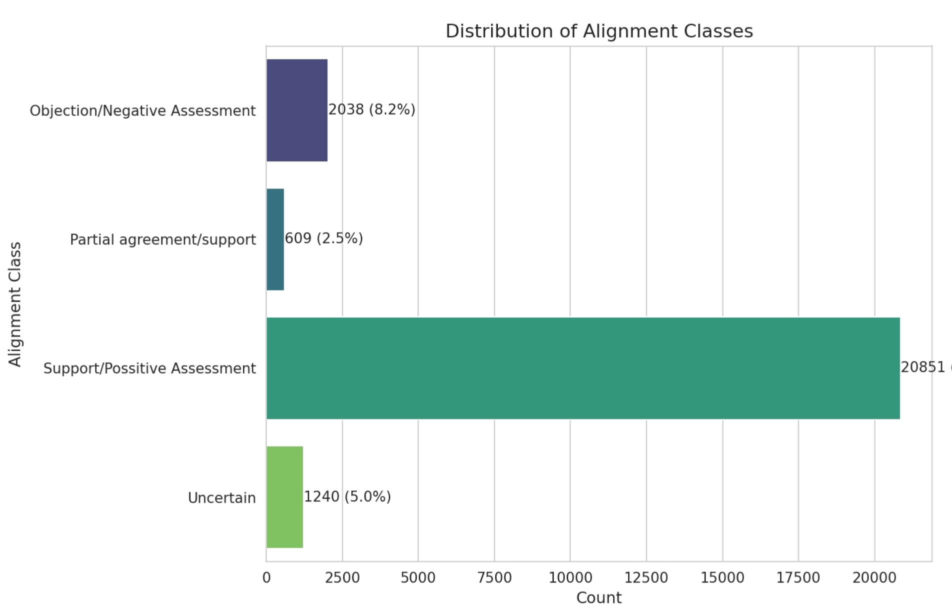

Figure 1: *Alignment Class Distribution in AMI Meeting Corpus dataset.*

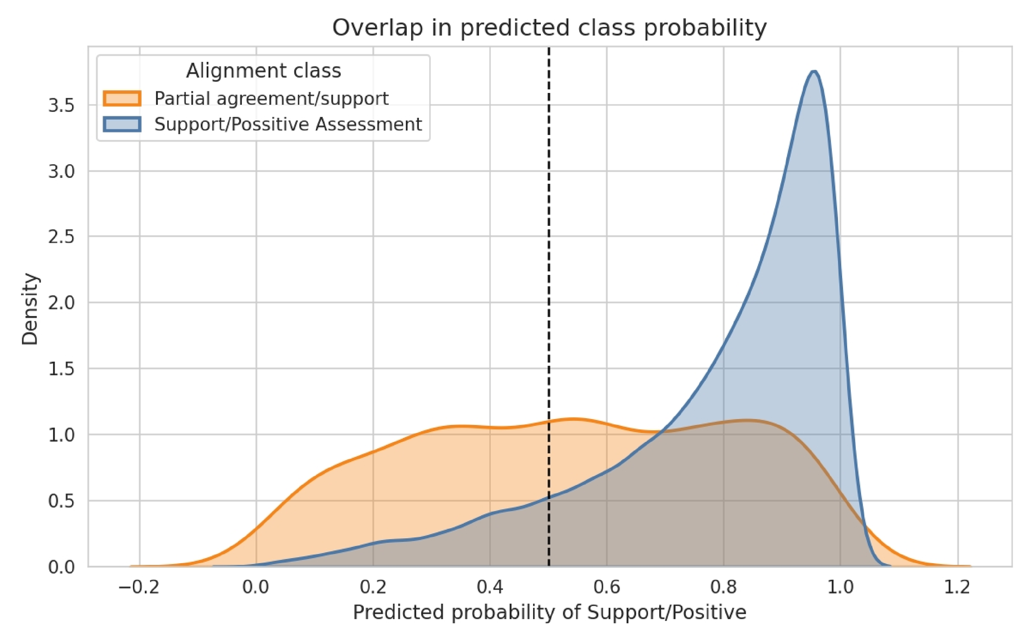

Figure 2: *Predicted probability distributions for Partial agreement/support and Support/Positive Assessment show substantial overlap, indicating limited separability between the two classes.*

3. Methodology and Evaluation

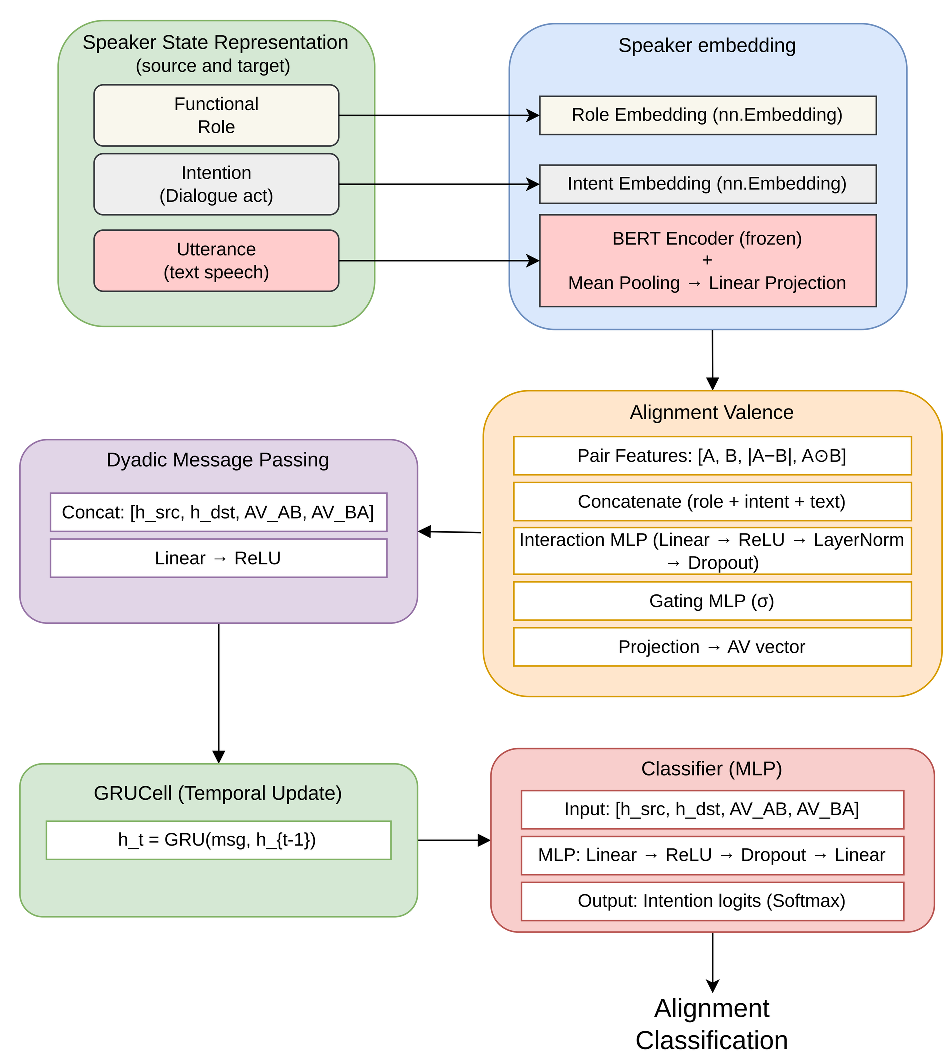

Figure 3: *Temporal Graph Neural Network Shared Mental Model (TGN-SMM) Model Architecture*

Temporal Graph Neural Network Shared Mental Model (TGN-SMM)

We propose **TGN-SMM**, a Temporal Graph Neural Network framework for modelling Shared Mental Models in multi-party meetings. The framework represents each meeting as an evolving speaker-interaction graph, where team members' cognitive states are updated incrementally as conversational events unfold over time.

- **Speaker embeddings:** Each speaker is represented by a concatenated vector encoding their team role, dialogue act, and speech utterance semantics.
- **Alignment valence:** Directional relational alignment between speakers is computed via pairwise feature interactions and a gated MLP, capturing speaker-to-speaker similarity.
- **Dyadic message passing:** Speaker cognitive states and alignment valences are fused through an MLP to generate interaction-specific messages.
- **Temporal update (GRU):** Each speaker's hidden state is updated via a GRU that integrates the incoming message with their previous cognitive state.
- **Alignment prediction:** The updated speaker states and their edge representation are used to predict the interpersonal alignment class of each event.

Baselines

We compare TGN-SMM against two ablated variants. *Static SMM* uses only multimodal speaker embeddings, with no message passing or temporal updates. *Speaker-GRU* adds GRU-based temporal updates to speaker states, but excludes alignment valence and dyadic message passing.

Evaluation Criteria

We evaluate TGN-SMM using a combination of class-level and overall performance metrics to capture both aggregate accuracy and per-class alignment quality, given the class imbalance across alignment categories. Macro-averaged F1 is treated as the primary metric, as it best reflects the model's ability to identify minority alignment classes such as objection and uncertainty, which are central to capturing nuanced team dynamics.

- **Accuracy:** Overall proportion of correctly predicted alignment labels.
- **F1 (per class):** Assesses per-class alignment quality, highlighting performance on minority classes.
- **Macro F1:** Class-balanced average, unaffected by the dominant Support/Positive class.
- **Weighted F1:** Class-frequency-weighted average, reflecting natural class distribution.

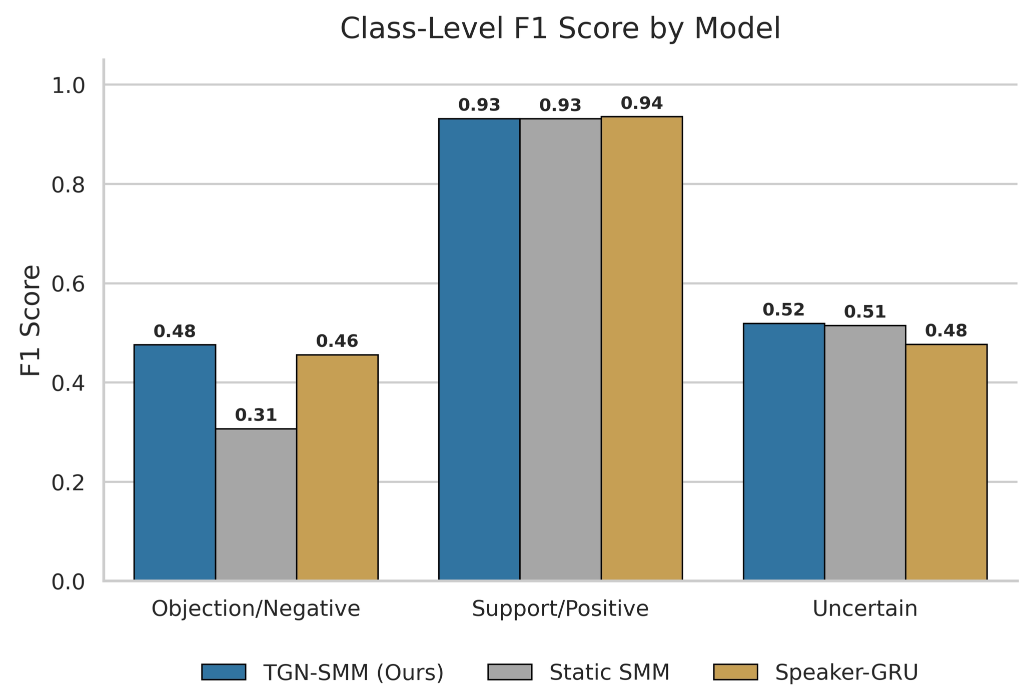

Figure 4: *Class-level F1 scores for TGN-SMM and baseline models on the AMI Corpus. TGN-SMM shows the largest gains on minority alignment classes (Objection/Negative, Uncertain), while performance on the majority class (Support/Positive) remains comparable across all models.*

Class level Performance on Alignment Recognition

TGN-SMM outperforms both baselines on the minority classes, improving F1 on Objection/Negative by +17 points over Static SMM (0.4757 vs 0.3066) and achieving the highest Uncertain F1 (0.5191). While also remaining competitive on Support/Positive (F1: 0.9317), where all models perform similarly due to its dominance in the dataset.

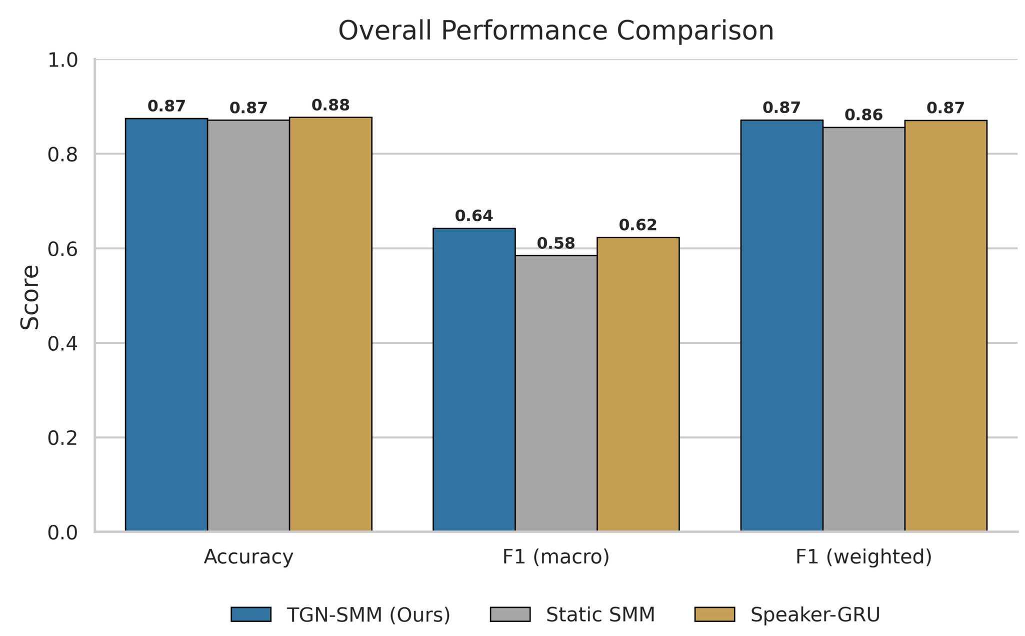

Figure 5: *Overall performance comparison across models on the AMI Corpus. While accuracy is comparable across all models, TGN-SMM achieves the highest macro-averaged F1, reflecting improved balance across alignment classes rather than a bias toward the majority class.*

Overall Performance on Alignment Recognition

TGN-SMM achieves the highest macro-averaged F1 (0.6422) and weighted F1 (0.8710) among all models, despite Speaker-GRU attaining marginally higher accuracy (0.8767 vs. 0.8746). This indicates that TGN-SMM delivers more balanced performance across alignment classes rather than accuracy driven primarily by the majority class.

4. Key Findings

Main Results

- TGN-SMM achieves the highest macro-averaged and weighted F1 scores, outperforming both static and temporal baselines.
- Temporal message passing and GRU-based updates substantially improve the model's ability to capture conversational dynamics over time.
- Alignment-aware valence learning yields better class balance rather than merely boosting majority-class performance.
- All models perform well on the majority class (*Support/Positive*), but TGN-SMM shows clear gains on the minority *Objection/Negative* and *Uncertain* classes.
- Temporal modelling alone (Speaker-GRU) was unable to explicitly model inter-speaker alignment, providing additional gains and suggesting dialogue is relational, not just sequential.
- Combining temporal dynamics with structured interpersonal interaction modelling is especially beneficial for under-represented, context-dependent dialogue acts.

5. Conclusion and Outlook

Conclusion

In this work, we proposed TGN-SMM for modelling interpersonal team dynamics and shared cognitive states in multi-party conversations. Experimental results showed that TGN-SMM improved performance across both minority and majority classes over its non-temporal and static baselines. It highlighted the importance of explicitly modelling temporal interaction alignment while capturing conversational dynamics. We showcased an approach that models evolving shared mental representations in group interactions, providing dynamic multi-party dialogue understanding and collaborative behaviour analysis in socially aware HCAI systems.

Future Work

- Extend the speaker embeddings with speaker social roles and a persona model.
- Explore multimodal interaction by integrating visual input.
- Explore and evaluate a human-agent deployment scenario that leverages representations from the TGN-SMM to make decisions to proactively assist in reaching shared team goals.

Presented at "3rd Workshop on Nonverbal Cues for Human-Robot Cooperative Intelligence at IEEE RO-MAN 2026", Fukoka, Japan