

Graph Neural Networks for Social Context Modeling and User Preference Prediction in Conversational Agents

Vishal Cook

Department of Computer Science, Colorado State University, Fort Collins, CO, USA.

vishal476@colostate.edu

Otis Castro

School of Computing, Clemson University, Clemson, SC, USA.

contactotis@clemson.edu

Alessandro D. Franklin

Department of Electrical Engineering and Computer Science, University of Kansas, Lawrence, KS, USA.

alessandro421@ku.edu

Abstract

Conversational agents increasingly operate within dense social environments where user preferences are not isolated preferences but are shaped by interpersonal influence, group memberships, historical interactions, and shifting contextual expectations. Modeling these dependencies requires representational frameworks capable of capturing relational structure beyond sequential dialogue content. Graph neural networks have emerged as a suitable class of architectures because they learn from irregular relational topologies, propagate contextual signals across connected entities, and support heterogeneous interaction graphs. This paper presents a system-level examination of graph neural network-based social context modeling and user preference prediction in conversational agents. It addresses the structural design of graph-based user and interaction representations, the architectural trade-offs among message passing, attention, temporal modeling, and pairwise preference prediction, and the requirements of production-scale data infrastructure. The discussion further considers deployment constraints such as latency, incremental graph updates, model drift, and integration with existing dialogue management stacks. Attention is given to fairness, robustness, privacy, explainability, and the governance mechanisms needed to prevent socially learned preferences from amplifying bias or undermining user autonomy. Rather than focusing narrowly on model accuracy, the paper emphasizes infrastructure sustainability, organizational accountability, and policy alignment. The analysis integrates perspectives from graph representation learning, recommender systems, conversational AI, sociotechnical design, and operational machine learning to provide a multidimensional research agenda for building preference-aware conversational systems that are both technically effective and institutionally responsible.

Keywords

graph neural networks, conversational agents, social context modeling, preference prediction, sociotechnical systems, fairness, governance.

1. Introduction

Conversational agents have moved from rule-based response systems toward adaptive platforms that must infer user goals, anticipate preferences, and negotiate contextual ambiguity across extended interactions. The central challenge of preference prediction in dialogue is not merely mapping language to intent but understanding how users change their preferences through social exposure, interaction histories, and platform-specific incentive structures. Relational information such as shared communities, interaction partners, frequently co-selected items, and feedback patterns provides a useful substrate for learning these dependencies. Graph neural networks have become influential in this context because they operate directly over graph-structured data and propagate learned representations across neighboring nodes. Foundational advances in inductive graph learning [1], graph convolutional networks [2], and attention-based graph architectures [3] have enabled practitioners to model large, evolving graphs without relying on fixed transductive embeddings. Comprehensive reviews of graph neural networks [4] and graph classification methods [5] have documented the architectural variety now available for such tasks.

The appeal of graph-based preference modeling in conversational agents lies in its capacity to integrate multiple social signals within a unified representational space. A user can be represented as a node connected to dialogue sessions, product mentions, peer users, topic entities, and historical feedback signals. The graph neural network learns how these relations jointly shape preference trajectories, while dialogue encoders continue to capture the immediate linguistic context. From a systems perspective, this integration introduces substantial engineering and governance complexity. It requires careful decisions about graph schema design, feature extraction, update frequency, privacy boundaries, and inference latency. The present paper therefore adopts an interdisciplinary stance, treating graph neural network models not as isolated algorithmic artifacts but as components embedded within large-scale conversational infrastructures. The discussion emphasizes structural trade-offs, deployment feasibility, fairness, robustness, and sustainability, offering a publication-ready analysis intended for researchers and system architects working at the intersection of machine learning, social computing, and dialogue systems.

2. Conversational Agents and Preference Modeling as a Socio-Technical Problem

Preference modeling in conversational agents has traditionally relied on sequence encoders, memory networks, and collaborative filtering signals derived from explicit or implicit feedback. Survey work on graph neural network applications has highlighted the shift toward relational inductive biases that are particularly relevant for social systems [6]. In dialogue settings, users often formulate preferences conditionally, referring to previous interactions, social comparisons, or contextual constraints. Graph neural networks can represent these conditional dependencies as edges with semantic types, allowing the model to distinguish between influence from a close peer, a co-purchased item, a shared community, or a conversational mention of a product attribute. Social recommendation research has demonstrated that graph-based propagation over user-item graphs improves recommendation accuracy when social ties are available [7]. Neural graph collaborative filtering has further shown that high-order connectivity among users and items encodes preference signals beyond direct interaction pairs [8].

The sociotechnical nature of conversational preference prediction also means that system designers must account for the origin and quality of social context. Graph edges may be derived from explicit friendship relations, follower structures, group memberships, co-interaction patterns, or inferred similarity. Each edge type carries different interpretative

weight and introduces distinct risks. Relational graph convolutional approaches offer a natural mechanism for handling heterogeneous edges by assigning type-specific transformations [9]. Knowledge graph embedding methods, originally developed for multi-relational data [10], have been adapted to model semantic relationships among items, attributes, and topics. These advances enable conversational agents to ground user preferences in external knowledge while also preserving social influence pathways. Hierarchical dialogue models [11] and survey treatments of neural conversational AI [12] reinforce the observation that dialogue systems benefit from multiple levels of abstraction, from utterance-level semantics to conversation-level goals and user-level preference shifts. Graph neural architectures extend this hierarchy outward to community-level and platform-level structures, making them well suited to the social embedding of conversation.

3. Graph Neural Network Architectures for Social Context Representation

Architectural choices for social context modeling in conversational agents range from relatively simple homogeneous graph convolutions to complex heterogeneous temporal models. In many deployments, the first structural decision concerns whether the graph should be treated as homogeneous, with all edges sharing the same transformation, or heterogeneous, with distinct message functions for different relation types. Heterogeneous designs are more expressive and prevent the conflation of social influence, item similarity, and conversational co-occurrence. They also introduce additional parameterization and require schema stability. Knowledge-aware conversational recommendation systems have shown that fusing dialogue context with structured knowledge graphs can improve both relevance and explainability [13]. Temporal graph models extend this reasoning by representing the evolution of preferences over time, which is particularly important in dialogue because user interests may shift rapidly during a single session or across monthly usage cycles [14].

Pairwise preference prediction within dialogue systems has also been addressed through dual-stage transformer designs that incorporate gated fusion, as examined in recent work [15]. Such architectures separate the processing of user representations and item or candidate representations before integrating them through a gated mechanism, allowing the model to learn which part of the social and conversational context should dominate a particular decision. In production contexts, the choice between pairwise and listwise preference objectives has direct consequences for training data construction, inference cost, and ranking stability. Multi-view attentive approaches in adjacent domains illustrate how different modalities can be fused for preference prediction [16], while earlier work on visual and stylistic recommendation demonstrates the value of learning complementary representations across heterogeneous signals [17]. For conversational agents, the analogous challenge is to combine dialogue semantics, social graph context, and platform behavioral signals without allowing any single modality to overwhelm the representation.

Architectural depth is another important system-level consideration. Deeper graph networks can capture longer-range social dependencies, but they also risk over-smoothing, where node representations become increasingly indistinguishable. Attention mechanisms mitigate this problem by weighting neighbors differently according to learned importance. However, attention over dense social graphs can be computationally expensive at inference time. Sparse sampling strategies, neighborhood subsampling, and precomputed influence summaries are therefore necessary in large-scale conversational platforms. The structural trade-off between expressive relational learning and operational feasibility frequently determines whether a graph neural network can be deployed behind a real-time dialogue interface. In practice, many

systems combine a fixed graph encoder with a lighter online preference predictor, using the graph network to maintain rich user context representations that are refreshed asynchronously.

4. Data Infrastructure and Graph Construction in Production Systems

The effectiveness of graph neural networks in conversational agents depends heavily on the quality and stability of the underlying graph infrastructure. Social context graphs must be constructed from heterogeneous data sources, including account relationships, conversation logs, item catalogs, feedback events, and sometimes external knowledge bases. Each source poses distinct extraction, deduplication, and linkage challenges. Edge confidence may vary widely, and inferred social ties can introduce privacy-sensitive information even when explicit social network data are not collected. Production systems must therefore maintain clear provenance for every edge type and support automated audits of graph construction pipelines. Hidden technical debt in machine learning systems is a well-documented risk in exactly this kind of multi-component architecture [18]. The entanglement of graph feature extraction, training data generation, and real-time inference can create feedback loops that are difficult to diagnose and correct.

From an operational perspective, graph construction is never a one-time process. User preferences evolve, conversational topics shift, and social connections churn. Incremental updates, temporal decay, and versioned graph snapshots are needed to keep learned representations aligned with current behavior. Energy and policy considerations for large-scale neural systems also apply directly to graph-based conversational infrastructure [19]. Graph neural network training is computationally intensive, particularly when the graph contains hundreds of millions of nodes and billions of edges. Distributed training, mixed-precision computation, and sampling-based message passing can reduce the environmental footprint, but they introduce additional complexity in experiment reproducibility and model evaluation. Sustainable deployment therefore requires not only algorithmic efficiency but also organizational discipline around capacity planning, retraining frequency, and the decommissioning of stale models.

5. Deployment Architectures and System-Level Trade-offs

Deploying graph neural network-based preference prediction in conversational agents requires careful coordination between offline representation learning and online inference. A common pattern is to maintain user and item embeddings in a graph-based vector store that is updated asynchronously, while the conversational agent queries these embeddings during dialogue turns. This design reduces latency but can create training-serving skew if the graph encoder and the online serving pipeline use different feature normalization or aggregation logic. The challenge is compounded when conversational agents themselves are built on large language models with their own internal representations. Risks associated with large generative models, including encoded social biases and the production of plausible but unsupported statements, have been widely discussed [20]. Graph-based social context models can either amplify or mitigate these risks depending on how their outputs are integrated into the dialogue policy.

Fairness concerns are particularly acute in social preference prediction because graph propagation can distribute sensitive attributes to users who have not explicitly disclosed them. Neighborhood inference, homophily, and structural bias can cause predictions to reflect demographic concentrations even when demographic features are not included in the model [21]. Measuring discrimination in algorithmic decision making requires careful definition of

reference classes and outcome distributions [22]. In conversational systems, the relevant outcome may not be a binary classification but a ranked list of suggested items, a predicted preference strength, or a dialogue response that confirms or denies a user request. Each of these outputs requires tailored fairness evaluation. System architects should also consider whether graph edges encode historical inequities, such as unequal access to products or unequal visibility in recommendation surfaces, and whether graph-based propagation may entrench those patterns over time.

The deployment architecture must support real-time governance interventions. For example, if a fairness audit detects that a particular social relation type is associated with discriminatory outcomes, the system may need to downgrade or remove that edge type without retraining the entire model. This requires modular graph encoders, relation-level configuration, and runtime switches. It also requires a strong model versioning system so that interventions are traceable. The interplay between graph-based preference prediction and large language model dialogue management is a further governance concern. If the graph model produces a preference score and the language model generates a persuasive explanation, the system may present socially learned preferences as if they originated from the user, undermining user autonomy. Policy mechanisms should therefore govern not only what predictions are made but also how they are expressed in interaction.

6. Robustness, Fairness, and Policy Implications

Robustness in graph-based conversational preference prediction involves both technical reliability and social legitimacy. Technical robustness includes resistance to noisy edges, adversarial influence, distribution shift, and cascading errors caused by graph propagation. Because graph neural networks aggregate information from neighboring nodes, a small number of compromised or heavily weighted neighbors can distort predictions across a broader social region. Robustness testing should therefore include stress tests that inject noisy edges, remove influential nodes, and simulate adversarial attempts to manipulate social context. Social legitimacy requires that preference predictions align with user expectations, support informed consent, and do not reproduce unjust social hierarchies. Fairness in recommender systems is an established research area that offers useful frameworks for multistakeholder evaluation, where users, providers, and platform operators may have conflicting interests [23]. In conversational settings, these interests become visible through dialogue and can affect trust, satisfaction, and long-term engagement.

Policy implications extend beyond legal compliance to include organizational norms, ethical review, and participatory design. A unified framework of AI principles emphasizes beneficence, non-maleficence, autonomy, justice, and explicability [24]. Applying these principles to graph-based social context modeling requires that the social graph itself be subject to governance. Users should be able to understand why certain preferences were predicted, what social information contributed to the prediction, and how to modify or delete those influences. Because graph neural networks create distributed representations, explaining an individual prediction often requires tracing multiple propagation paths. This creates new demands for explanation interfaces that are interpretable to non-experts without oversimplifying the underlying relational logic. Policymakers and platform designers must also address the possibility that social context models will be used to nudge users toward choices that benefit platform partners rather than users themselves. Transparency in edge provenance and influence weighting is therefore an ethical requirement, not merely a technical feature.

7. Sustainability and Operational Governance

Sustainable operation of graph-based conversational preference systems depends on aligning model lifecycles with institutional governance. Retraining schedules should balance predictive freshness against computational cost and environmental impact. Because social graphs change continuously, full retraining from scratch is rarely necessary; incremental fine-tuning and selective recalculation of affected neighborhoods can reduce resource consumption. However, incremental learning introduces its own risks, including catastrophic forgetting, uneven representation updates, and drift in fairness guarantees. Operational governance should therefore define when full retraining is required, when incremental updates are acceptable, and how model performance is monitored across user segments. The hidden technical debt framework is again relevant because graph feature pipelines, dialogue systems, and preference predictors often interact in ways that are not captured by standard unit tests [18].

Organizational structures also need clear accountability for social context modeling. Data engineers, applied scientists, product managers, legal reviewers, and user advocacy teams must collaborate to define acceptable relational features, edge retention policies, and intervention mechanisms. Energy reporting can be integrated into model development workflows so that sustainability is treated as a first-order metric rather than an afterthought [19]. Documentation should include the intended use of social context, limitations of the graph representation, known biases, and procedures for user correction. Model cards, fairness audits, and adversarial reviews can provide external visibility and build institutional trust. Long-term sustainability therefore requires not only technical efficiency but also a governance culture that treats social graph models as sociotechnical artifacts with ongoing responsibilities.

8. Conclusion

Graph neural networks offer a powerful framework for representing social context and predicting user preferences in conversational agents. Their capacity to learn from heterogeneous relational structures enables dialogue systems to capture interpersonal influence, historical feedback, knowledge associations, and temporal preference shifts within a unified graph-based representation. At the same time, the adoption of graph neural architectures introduces structural challenges that extend far beyond predictive accuracy. System designers must make careful trade-offs among expressiveness, latency, update frequency, explainability, and fairness. Data infrastructure must support dynamic graph construction, edge provenance, privacy protection, and sustainable retraining. Deployment architectures must coordinate asynchronous graph updates with real-time dialogue inference while avoiding training-serving skew and undesirable interaction with large language models.

From a policy perspective, graph-based social preference prediction is not a neutral technical optimization. It embeds assumptions about how social influence operates, whose relationships matter, and how past behavior should constrain future choices. Without robust governance, such systems risk amplifying platform biases, reducing user autonomy, and entrenching structural inequities. The integration of graph neural networks into conversational agents therefore requires interdisciplinary collaboration across machine learning, human-computer interaction, data engineering, ethics, and policy. Future research should continue to develop architectures that are both relationally expressive and operationally manageable, while also establishing evaluation standards that account for fairness, robustness, sustainability, and user control. The long-term success of these systems will be measured not only by their ability to

predict preferences but by their capacity to support socially legitimate, transparent, and accountable conversational experiences.

References

1. Hamilton, W. L., Ying, R., & Leskovec, J. (2017). Inductive representation learning on large graphs. In *Advances in Neural Information Processing Systems* 30 (NeurIPS 2017).
2. Kipf, T. N., & Welling, M. (2017). Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations (ICLR)*.
3. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., & Bengio, Y. (2018). Graph attention networks. In *International Conference on Learning Representations*.
4. Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Yu, P. S. (2021). A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 32(1), 4-24.
5. Zhang, M., Cui, Z., Neumann, M., & Chen, Y. (2018). An end-to-end deep learning architecture for graph classification. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1), 4438-4445.
6. Zhou, J., Cui, G., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C., & Sun, M. (2020). Graph neural networks: A review of methods and applications. *AI Open*, 1, 57-81.
7. Fan, W., Ma, Y., Li, Q., He, Y., Zhao, E., Tang, J., & Yin, D. (2019). Graph neural networks for social recommendation. In *The World Wide Web Conference (WWW)* (pp. 417-426).
8. Wang, X., He, X., Wang, M., Feng, F., & Chua, T.-S. (2019). Neural graph collaborative filtering. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 165-174).
9. Schlichtkrull, M., Kipf, T. N., Bloem, P., van den Berg, R., Titov, I., & Welling, M. (2018). Modeling relational data with graph convolutional networks. In *European Semantic Web Conference* (pp. 593-607). Springer.
10. Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., & Yakhnenko, O. (2013). Translating embeddings for modeling multi-relational data. In *Advances in Neural Information Processing Systems* 26.
11. Serban, I. V., Sordoni, A., Bengio, Y., Courville, A., & Pineau, J. (2016). Building end-to-end dialogue systems using generative hierarchical neural network models. *Proceedings of the AAAI Conference on Artificial Intelligence*, 30(1), 3776-3783.
12. Gao, J., Galley, M., & Li, L. (2019). Neural approaches to conversational AI. *Foundations and Trends in Information Retrieval*, 13(2-3), 127-298.
13. Zhou, K., Zhao, W. X., Bian, S., Zhou, Y., Wen, J.-R., & Yu, J. (2020). Improving conversational recommender systems via knowledge graph based semantic fusion. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1006-1014).
14. Chen, H., Li, Y., Sun, X., Xu, G., & Yin, H. (2021). Temporal meta-path guided explainable recommendation. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining* (pp. 1056-1064).

15. Shih, K., Deng, Z., Chen, X., Zhang, Y., & Zhang, L. (2025, May). DST-GFN: A Dual-Stage Transformer Network with Gated Fusion for Pairwise User Preference Prediction in Dialogue Systems. In 2025 8th International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE) (pp. 715-719). IEEE.
16. Wu, C., Wu, F., Qi, T., Huang, Y., & Xie, X. (2022). Neural news recommendation with attentive multi-view learning. *IEEE Transactions on Knowledge and Data Engineering*, 34(7), 3110-3122.
17. McAuley, J., Targett, C., Shi, Q., & van den Hengel, A. (2015). Image-based recommendations on styles and substitutes. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 43-52).
18. Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., Chaudhary, V., Young, M., Crespo, J.-F., & Dennison, D. (2015). Hidden technical debt in machine learning systems. In *Advances in Neural Information Processing Systems* 28.
19. Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 3645-3650).
20. Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610-623).
21. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys*, 54(6), Article 115.
22. Zliobaite, I. (2017). Measuring discrimination in algorithmic decision making. *Data Mining and Knowledge Discovery*, 31(4), 1060-1089.
23. Ekstrand, M. D., Das, A., Burke, R., & Diaz, F. (2022). Fairness in recommender systems. In *Recommender Systems Handbook* (pp. 679-707). Springer.
24. Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review*, 1(1).