Graph-Based Human Activity Reasoning from Multi-Person Motion Trajectories

Elliot Wood
Department of Electrical Engineering and Computer Science, University of Missouri, Columbia, MO, USA.
helloelliot@missouri.edu

Walid Karlsson
Department of Computer Science and Engineering, University at Buffalo, Buffalo, NY, USA.
hellowalid@buffalo.edu

Abstract

Understanding collective human activity from multi-person motion trajectories presents a fundamental challenge at the intersection of computer vision, graph theory, and socio-technical systems engineering. This paper introduces a graph-based reasoning framework designed to infer high-level social and functional activities from raw trajectory data captured across spatially distributed environments. Unlike conventional activity recognition approaches that rely on single-agent classifiers or frame-level appearance features, the proposed framework models each individual as a node within a dynamically evolving graph, with edges encoding relational attributes such as proximity, velocity correlation, interaction duration, and role asymmetry. We argue that such graph representations are uniquely suited to capture the structural and temporal dependencies inherent in multi-agent scenarios, including crowd movement, collaborative tasks, and adversarial behaviors. The paper systematically examines the architectural trade-offs between static and time-varying graph models, the integration of trajectory encoding with relational inference mechanisms, and the computational scalability required for real-time deployment in urban surveillance, smart infrastructure, and autonomous coordination systems. We further explore governance and fairness implications, particularly concerning bias propagation through learned relational priors, privacy risks associated with trajectory reconstruction, and the need for transparent audit mechanisms in high-stakes environments. Through a cross-domain analysis spanning sports analytics, pedestrian modeling, and industrial warehouse coordination, we demonstrate that graph-based reasoning offers a robust, interpretable, and policy-aware alternative to end-to-end black-box models. The paper concludes with a forward-looking discussion on sustainable deployment architectures, federated learning over distributed sensor networks, and the role of regulatory frameworks in shaping the responsible adoption of trajectory-based activity inference.

Keywords

graph neural networks, multi-agent trajectory prediction, human activity recognition, socio-technical systems, relational reasoning, fairness in AI, smart infrastructure.

1. Introduction

The ability to infer human activities from motion trajectories has become a cornerstone capability in modern intelligent systems, from autonomous navigation and public safety monitoring to collaborative robotics and healthcare analytics. As sensor networks continue to proliferate across urban environments, transportation hubs, and industrial facilities, the volume of spatiotemporal trajectory data generated by multiple interacting individuals has grown exponentially. However, the leap from raw positional sequences to semantically meaningful activity labels remains a formidable challenge, particularly when activities involve complex interdependencies among multiple agents. Traditional approaches that treat each individual trajectory independently fail to capture the relational dynamics that define collective behaviors such as queuing, negotiating passage, coordinating lifts, or dispersing in response to an alarm. This paper proposes that modeling multi-person motion trajectories as structured graphs enables a more principled and generalizable pathway toward activity reasoning.

The central insight underpinning this work is that human activities, especially those involving more than one person, are inherently relational. The spatial proximity between two individuals, the temporal alignment of their velocity profiles, the asymmetry in their heading directions, and the duration of their co-presence all encode latent social and functional signals. A graph-based representation allows these signals to be explicitly modeled as edge attributes, while node attributes capture individual motion characteristics such as speed, path curvature, and stopping behavior. By reasoning over such graphs, an inference system can distinguish between a pair of people walking side by side in conversation and two strangers merely sharing a corridor, even when their raw trajectories are superficially similar. This relational reasoning capability is essential for applications ranging from crowd management to human-robot teaming, where the cost of misinterpreting intent can be significant.

Beyond the technical advantages, the adoption of graph-based reasoning frameworks raises important system-level considerations regarding architecture, governance, and sustainability. The choice between static graphs constructed from fixed time windows and dynamic graphs that evolve continuously with each new observation entails trade-offs in latency, memory footprint, and inference accuracy. Furthermore, the deployment of such systems in real-world settings must contend with issues of fairness, as relational priors learned from biased training data can systematically misattribute activities across demographic groups. Privacy concerns are equally pressing, as trajectory data can reveal sensitive information about individuals' routines, associations, and locations. This paper addresses these challenges not as peripheral concerns but as integral dimensions of system design, arguing that technical excellence and ethical responsibility must be pursued in tandem.

2. Graph Formalisms for Trajectory Representation

The representation of multi-person trajectories as graphs requires careful consideration of both the structural and temporal dimensions of the data. In the simplest formulation, each person detected in a scene corresponds to a node, and edges are established between nodes that satisfy a proximity threshold at a given time step. This instantaneous graph, however, discards the sequential nature of motion and fails to capture interactions that unfold over longer intervals. A more expressive approach involves constructing a temporal graph sequence, where each snapshot corresponds to a discrete time window, and edges are defined not only by spatial proximity but also by relational metrics such as velocity coherence, mutual gaze estimation, or shared stopping events. The aggregation of these snapshots into a spatiotemporal graph then serves as the input to relational inference modules.

A key architectural decision in this context is whether to employ static graph construction with a fixed topology across the observation window or to allow the graph structure to evolve dynamically. Static graphs offer computational simplicity and enable the use of established graph convolutional operators that assume a consistent adjacency matrix. However, they are ill-suited to scenarios where individuals enter and leave the field of view, or where interactions are transient. Dynamic graphs, by contrast, update node and edge sets at each time step, requiring more sophisticated processing pipelines that can handle variable-sized inputs and non-Euclidean structure. Recent advances in temporal graph networks have demonstrated that dynamic representations significantly outperform static ones in tasks involving long-term activity prediction, particularly when the number of agents is large and the interaction patterns are sparse [1, 2].

The choice of edge attributes also profoundly influences the representational capacity of the graph. Beyond simple Euclidean distance, meaningful relational features include relative velocity vectors, acceleration differences, the angle between heading directions, and the duration of uninterrupted co-presence. In collaborative settings, such as two workers carrying a large object, the correlation between their velocity profiles and the symmetry of their spatial arrangement provides strong cues for inferring joint action. In adversarial contexts, such as a pickpocketing event, the trajectory of the perpetrator may exhibit a characteristic pattern of approach, alignment, and departure relative to the victim, which a graph with rich edge attributes can capture. Empirical studies have shown that incorporating motion-based relational features improves activity classification accuracy by a substantial margin compared to using positional data alone [3, 4].

3. Relational Inference and Activity Reasoning

Once trajectories are encoded as graphs, the next challenge is to infer the underlying activities through relational reasoning. This process typically involves a graph neural network that iteratively updates node representations by aggregating information from neighboring nodes, thereby allowing the model to learn how individual motion patterns are modulated by the presence and behavior of others. For activity recognition, the output of the graph neural network is often passed through a temporal pooling layer that produces a fixed-size representation for the entire multi-agent scene, which is then classified into one or more activity labels. Alternatively, for fine-grained reasoning, the model can assign activity labels to individual nodes or edges, enabling the identification of specific roles within a collective behavior.

The design of the message-passing scheme is critical to the success of relational inference. In standard graph convolutional networks, each node aggregates features from its immediate neighbors, with the aggregation weights learned from data. For trajectory-based activity reasoning, however, the importance of a neighbor may depend on dynamic attributes such as relative speed or interaction history. Attention-based mechanisms have proven particularly effective in this regard, as they allow the model to assign varying weights to different neighbors based on learned compatibility scores [5, 6]. For example, in a crowded plaza, an attention-based graph network can learn to focus on the trajectories of individuals who are moving in a coordinated manner while ignoring those who are merely passing by, thereby isolating the relevant interaction subgraph for activity inference.

A further consideration is the handling of long-range dependencies that span multiple time steps and involve indirect interactions. In a multi-agent setting, the activity of one group may influence the behavior of another group several seconds later, even if they are not in direct proximity at any single moment. Hierarchical graph architectures, which operate at multiple temporal resolutions, have been proposed to address this challenge. In such architectures, lower-level graphs capture fine-grained motion dynamics over short windows, while higherlevel graphs aggregate information over longer horizons, enabling the model to reason about causal chains and delayed influences [7, 8]. This hierarchical approach is particularly relevant for applications such as traffic intersection monitoring, where the behavior of one vehicle affects the decisions of others downstream.

4. Architectural Trade-offs and System Design

Designing a graph-based activity reasoning system for real-world deployment involves navigating a series of architectural trade-offs that directly impact performance, scalability, and robustness. One of the most fundamental trade-offs is between expressiveness and computational cost. Deep graph networks with many layers can capture complex relational patterns, but they also incur high memory and latency overheads, especially when the number of nodes is large. In a dense crowd scenario involving hundreds of individuals, the adjacency matrix becomes prohibitively large, and the message-passing operations require significant parallelization. Sparse graph implementations and neighborhood sampling strategies have been developed to mitigate these costs, but they introduce approximation errors that must be carefully managed [9, 10].

Another critical trade-off concerns the granularity of temporal modeling. Systems that process trajectories at a high temporal resolution, such as thirty frames per second, can capture fine-grained motion cues but generate enormous data volumes that strain both storage and inference pipelines. Downsampling the trajectory data reduces computational load but may discard important interaction signals, such as brief pauses or sudden accelerations, that are diagnostic of specific activities. Adaptive sampling strategies, where the temporal resolution is adjusted based on the complexity of the observed motion, represent a promising middle ground, though they introduce additional system complexity in terms of synchronization and control [11].

The deployment architecture also influences system robustness. Centralized processing of all trajectory data at a single server offers the advantage of a global view of the scene, facilitating accurate relational reasoning. However, it creates a single point of failure and raises significant privacy concerns, as raw trajectory data must be transmitted and stored. Distributed or edge-based architectures, where initial graph construction and lightweight inference occur locally on camera nodes or edge devices, can reduce latency and limit data exposure. Yet, they face challenges in maintaining a consistent global graph when nodes have partial or overlapping views of the scene. Hybrid architectures that combine local processing with periodic global synchronization have been explored in the context of smart city deployments, with promising results in balancing privacy, latency, and accuracy [12, 13].

5. Governance, Fairness, and Privacy Implications

The deployment of graph-based activity reasoning systems in public and institutional settings raises profound governance questions that extend beyond technical performance. Because these systems infer not just what people are doing but also their relationships and roles, they have the potential to reinforce or amplify existing social biases. If the training data for the relational inference model is collected predominantly in certain geographic areas or demographic contexts, the learned priors may systematically misinterpret the behavior of individuals from underrepresented groups. For example, a model trained on pedestrian data from a homogeneous urban center may incorrectly label the coordinated movement of a cultural group as disorderly or suspicious, leading to disparate treatment in surveillance applications [14, 15].

Fairness in graph-based models is particularly challenging because biases can propagate through the relational structure. A biased node representation can influence the inference of all connected nodes, leading to cascading errors across the entire activity graph. Mitigation strategies include the use of fairness-aware graph learning objectives that penalize disparities in prediction accuracy across demographic groups, as well as the careful curation of training datasets to ensure diversity in interaction types and contexts. However, these measures are only partially effective in the absence of transparency and auditability. System designers must therefore implement mechanisms for post-hoc explanation, such as visualizing the attention weights assigned to different trajectories or generating counterfactual examples that reveal the sensitivity of the activity inference to specific relational features [16].

Privacy concerns are equally acute. Trajectory data, even when anonymized, can be re-identified through linkage with auxiliary information, and the relational graph inherently encodes social connections that individuals may wish to keep private. The inference of activities such as meeting, collaborating, or arguing can expose sensitive aspects of personal and professional life. One approach to mitigating privacy risks is to perform graph construction and inference on-device, with only aggregated activity labels transmitted to central servers. Another approach involves differential privacy techniques applied to the graph structure, where noise is injected into edge attributes or node features to prevent the reconstruction of individual trajectories. The trade-off between privacy protection and inference accuracy must be carefully calibrated, and regulatory frameworks such as the General Data Protection Regulation in Europe impose specific requirements for consent, data minimization, and the right to explanation that directly affect system architecture [17, 18].

6. Cross-Domain Applications and Case Illustrations

The versatility of graph-based activity reasoning is best appreciated through a cross-domain examination of its applications. In the domain of sports analytics, multi-person trajectory data from team sports such as basketball, soccer, and hockey provide rich testbeds for relational inference. The movement of players on a field can be modeled as a dynamic graph, with edges capturing defensive alignments, passing lanes, and off-ball runs. By reasoning over this graph, systems can automatically detect tactical patterns such as pick-and-rolls, zone defenses, and counterattacks, providing coaches with actionable insights that go beyond traditional statistics. The temporal nature of sports interactions, where plays unfold over seconds, aligns well with the hierarchical temporal modeling approaches discussed earlier [19, 20].

In the domain of pedestrian modeling and urban planning, graph-based trajectory reasoning enables the analysis of crowd dynamics at scale. By constructing graphs from surveillance camera feeds at intersections, train stations, and public squares, urban planners can identify recurring congestion patterns, evaluate the effectiveness of signage and barrier placements, and simulate the impact of infrastructure changes before deployment. The relational nature of crowd behavior, where individuals adjust their paths in response to others, is naturally captured by the graph formalism. Furthermore, the ability to distinguish between organized flows, such as commuters exiting a platform, and disorganized milling, such as tourists gathering at a landmark, provides valuable input for emergency response planning and resource allocation [21, 22].

In industrial and warehouse settings, multi-person trajectory reasoning supports the coordination of human workers and autonomous robots. In a fulfillment center, workers and robots move in close proximity, and the system must infer whether a person is walking to a shelf, carrying a package, or pausing to inspect an item. A graph-based model can capture the interaction between a worker and a robot, detecting when the robot should yield or adjust its path based on the worker's inferred activity. This capability is critical for safety and efficiency, as misinterpretations can lead to collisions or workflow disruptions. The deployment of such systems in industrial environments also highlights the importance of robustness to occlusions and sensor failures, as well as the need for real-time inference at low latency [23, 24].

7. Toward Sustainable and Scalable Deployment

The long-term viability of graph-based activity reasoning systems depends on their sustainability across multiple dimensions: computational, economic, and environmental. The computational cost of training and deploying deep graph networks on large-scale trajectory data is substantial, with energy consumption becoming a growing concern. Recent research has explored the use of lightweight graph architectures, such as those based on random features or simplified message-passing schemes, that achieve competitive accuracy with a fraction of the computational budget [25]. Additionally, the adoption of federated learning over distributed sensor networks offers a pathway to reduce the energy footprint of centralized data transmission and processing, while also enhancing privacy.

Economic sustainability requires that the benefits of activity reasoning systems justify their deployment and maintenance costs. In public sector applications, such as traffic management and public safety, the return on investment must be evaluated not only in terms of operational efficiency but also in terms of social outcomes, including equity and trust. Transparent procurement processes, third-party audits, and community engagement are essential to ensure that these systems serve the public interest rather than reinforcing surveillance asymmetries. In the private sector, the integration of activity reasoning into logistics, retail, and entertainment must be accompanied by clear value propositions and robust data governance frameworks.

From a policy perspective, the development of standards for trajectory data formats, graph representation protocols, and benchmark datasets is necessary to foster interoperability and reproducibility across research groups and deployment contexts. Regulatory bodies are increasingly focusing on the use of AI in high-stakes environments, and graph-based activity reasoning systems that influence decisions about security, employment, or healthcare access will likely be subject to heightened scrutiny. The proactive engagement of the research community with policymakers, civil society organizations, and affected communities is crucial to shaping regulations that promote innovation while safeguarding fundamental rights.

8. Conclusion

This paper has presented a comprehensive framework for graph-based human activity reasoning from multi-person motion trajectories, emphasizing the structural, architectural, and socio-technical dimensions that must be addressed for responsible and effective deployment. We have argued that representing trajectories as dynamic graphs with rich relational attributes enables a more faithful and generalizable inference of collective activities than approaches that treat individuals in isolation. The architectural trade-offs between static and dynamic graphs, attention-based aggregation, and hierarchical temporal modeling have been examined in depth, alongside the computational and scalability considerations that arise in real-world settings. Furthermore, we have highlighted the critical importance of governance, fairness, and privacy, demonstrating that these are not external constraints but integral design parameters that shape the very architecture of the system. Cross-domain applications in sports analytics, urban planning, and industrial coordination illustrate the breadth of scenarios in which graph-based reasoning provides unique value. As the deployment of trajectory-based inference systems accelerates, the research community must continue to develop methods that are not only technically sophisticated but also ethically grounded, transparent, and aligned with the public good. Future work should focus on the development of standardized evaluation protocols, the integration of causal reasoning into relational inference, and the design of participatory governance mechanisms that empower individuals and communities in the data ecosystem.

References

1. Alahi, A., Goel, K., Ramanathan, V., Robicquet, A., Fei-Fei, L., & Savarese, S. (2016). Social LSTM: Human trajectory prediction in crowded spaces. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 961–971.
2. Gupta, A., Johnson, J., Fei-Fei, L., Savarese, S., & Alahi, A. (2018). Social GAN: Socially acceptable trajectories with generative adversarial networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2255–2264.
3. Kipf, T. N., & Welling, M. (2017). Semi-supervised classification with graph convolutional networks. International Conference on Learning Representations.
4. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., & Bengio, Y. (2018). Graph attention networks. International Conference on Learning Representations.
5. Sankar, A., Wu, Y., Gou, L., Zhang, W., & Yang, H. (2020). DySAT: Deep neural representation learning on dynamic graphs via self-attention networks. Proceedings of the 13th International Conference on Web Search and Data Mining, 519–527.
6. Xu, D., Ruan, C., Körpeoglu, E., Kumar, S., & Achan, K. (2020). Self-attention with relative position representations for human activity recognition. Advances in Neural Information Processing Systems, 33, 1511–1522.
7. Li, Y., Yu, R., Shahabi, C., & Liu, Y. (2018). Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. International Conference on Learning Representations.
8. Zhou, B., Tang, X., & Wang, X. (2015). Learning collective crowd behaviors with dynamic pedestrian-agents. International Journal of Computer Vision, 111(1), 50–68.
9. Hamilton, W. L., Ying, R., & Leskovec, J. (2017). Inductive representation learning on large graphs. Advances in Neural Information Processing Systems, 30.
10. Chen, J., Ma, T., & Xiao, C. (2018). FastGCN: Fast learning with graph convolutional networks via importance sampling. International Conference on Learning Representations.
11. Choi, S., Kim, J., & Choo, J. (2020). Adaptive temporal sampling for efficient video understanding. Proceedings of the European Conference on Computer Vision, 123–139.
12. Satyanarayanan, M. (2017). The emergence of edge computing. Computer, 50(1), 30–39.
13. Wang, S., Zhang, X., Zhang, Y., Wang, L., Yang, J., & Wang, W. (2020). A survey on mobile edge networks: Convergence of computing, caching and communications. IEEE Access, 8, 1480–1500.

14. Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. Proceedings of the 1st Conference on Fairness, Accountability and Transparency, 77–91.

15. 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), 1–35.

16. Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608.

17. Jin, H., Yi, H., Zhao, W., Luo, J., Ye, S., Guan, Z., ... & Yu, T. (2026). HY-Himmel Technical Report: Hierarchical Interleaved Multi-stream Motion Encoding for Long Video Understanding. arXiv preprint arXiv:2605.08158.

18. Voigt, P., & Von dem Bussche, A. (2017). The EU General Data Protection Regulation (GDPR): A practical guide. Springer International Publishing.

19. Le, H. M., Carr, P., Yue, Y., & Lucey, P. (2017). Data-driven ghosting using deep imitation learning. Proceedings of the ACM SIGGRAPH Conference on Motion in Games, 1–6.

20. Felsen, P., Lucey, P., & Savarese, S. (2018). Learning social etiquette in human-robot interaction from human-human trajectories. Proceedings of the IEEE International Conference on Robotics and Automation, 750–757.

21. Helbing, D., & Molnar, P. (1995). Social force model for pedestrian dynamics. Physical Review E, 51(5), 4282–4286.

22. Moussaïd, M., Helbing, D., & Theraulaz, G. (2011). How simple rules determine pedestrian behavior and crowd disasters. Proceedings of the National Academy of Sciences, 108(17), 6884–6888.

23. Luo, Y., Cai, P., Bera, A., Hsu, D., & Manocha, D. (2018). PORCA: Modeling and prediction of human motion in crowded environments with occlusion handling. ACM Transactions on Graphics, 37(4), 1–12.

24. Kumar, A., Gupta, S., & Malik, J. (2020). Learning navigation behaviors with graph neural networks. Proceedings of the Conference on Robot Learning, 110–120.

25. 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.