

Graph Machine Learning to Map and Predict Customer Journeys and Churn

Yashika Shankheshwaria

Washington University of Science and Technology, Virginia United States of America

Abstract

Understanding customer behavior and predicting customer churn remain vital and central strategic objectives for modern digital organizations. Customer interactions are becoming multi-touch, they are cross-channel, and dynamically interconnected. The traditional linear modeling approach does not meet these demands. The traditional model does not capture the complexity that is vital in digital customer journeys. These challenges should be addressed, and graph machine learning (GraphML) and graph neural networks (GNNs) have become effective frameworks for addressing these challenges. The frameworks are capable of modeling the behavior of the customers as a network of entities, transitions, and interactions. The paper explores how graph machine learning can be applied to map the journeys of the customer and predict churn with higher accuracy, as well as deeper behavioral insight than conventional models.

This paper will draw insights from peer-reviewed sources. It develops a conceptual, training graph model to detect churn signals, a multi-layered framework for constructing customer journey graphs, and operationalizing predictions into retention strategies that are actionable.

The findings depict that graph-based modeling captures network dependencies. It captures cross-channel transitions and the behavior patterns that traditional machine learning does not incorporate. The findings indicate that GraphML offers a better method for describing customers as interconnected nodes within a relational structure. It enables more accurate churn prediction and a comprehensive mapping of the real-world customer journeys. The paper will conclude with a managerial implication, suggestions for future research, and limitations.

Keywords: Graph Machine Learning, Customer Journey, Customer Churn, Graph Neural Networks, Customer Experience Analytics, Retention Modeling

A. INTRODUCTION

Businesses today operate in a hyper-connected digital environment. The interactions of customers unfold across a wide range of channels. This includes social media, mobile apps, emails, websites, chatbots, and offline experiences. Important to note that the interactions no longer resemble linear funnels. The interactions are complex and nonlinear customer journeys. These are characterized by cross-channel transitions, varied sequences of engagement, and loops.

Pantouvakis and Kottika (2022) indicate that contemporary customer journeys entail psychological, behavioral, and emotional components that emerge dynamically as the customer interacts with brands. Customer churn is defined as a customer leaving, ceasing, or cancelling a purchase, and this poses significant financial risks. Important to note that churn directly impacts the revenue for subscription businesses. It impacts monthly recurring revenue. For the e-commerce business, churn impacts lifetime

value and ROI acquisition. Traditional churn prediction models rely on tabular customer features and sequence models that treat interactions as events that are isolated or linear timelines. The emerging work depicts that churn behavior is influenced by social ties, sequential dependencies, peer behaviors, and multi-channel activity. These are all relational and best modeled as networks.

GraphML provides the ideal paradigm that can address this complexity. It presents customers, devices, touchpoints, channels, and nodes that are connected through edges that encode interactions or relationships. Graph modeling captures the structured and contextual richness of customer behavior. GNNs are vital in enabling learning from these structures as they propagate messages across nodes. They allow churn predictions to incorporate neighborhood information and higher-order connections (Gao et al., 2023; Khemani et al., 2024).

The paper explores how graph machine learning can be applied to map and predict the journeys of customers and churn. This paper synthesizes insights from the provided sources. The goal is to produce an academic analysis that is detailed and provides insights that can be implemented by organizations that seek to modernize customer analytics.

B. LITERATURE REVIEW

1. Customer Journey Analytics in the Digital Era

The journey of the customer describes the end-to-end experiences that customers have when they interact with the brand. Pantouvakis and Kottika (2022) depict that modern customer journey research has evolved from static linear models to systems that are dynamic and encompass multi-channel engagement and psychological processes. The authors depict that the complexity of contemporary journeys demands frameworks that are capable of capturing the interconnected touchpoints as well as emotional states.

Digital commerce is on the rise, and this has driven the need for quantitative customer journey mapping. According to Pantouvakis and Kottika (2022), customer journey analysis must integrate large-scale behavioral data. This includes transactions, clickstreams, and cross-device interactions. This is important as it helps uncover patterns that support personalization and friction reduction. The research emphasizes that traditional journey-mapping tools lack the analytical power that is required to reveal deeper structural relationships in the journeys.

Aluwala (2022) indicates that AI-enhanced journey analytics are vital as they provide automated identification of journey patterns. They can segment the cluster behaviors and detect macro-moments where interventions can minimize churn or increase conversions. The existing AI approaches treat the journey as sequences or feature sets, and this limits their ability to capture relational complexity.

This gap sets the stage for graph-based customer journey representation. This is where interactions form edges between the customers, devices, and touchpoints. This is vital in enabling a holistic and interconnected view.

2. Machine Learning and Churn Prediction

The prediction of churn has traditionally relied on machine learning techniques like random forests, logistic regression, decision trees, deep neural networks, and SVMs. AbdelAziz and El-Khatib (2025) conducted a review of machine learning and deep learning models for the prediction of churn. Their analysis depicts that while advanced models outperform order techniques, some of them still operate on flat, tabular data. They ignore social, relational, or cross-channel influences.

Matuszelański and Kopczewska (2022) conducted a study that revealed that churn in retail e-commerce is influenced by peer purchasing, geographical clustering, and spatial independencies. Churn decisions are

not independent, but they reflect structural relationships. The author's insights align with the principles of graph machine learning.

Important to note that traditional churn models continue to improve in accuracy. They, however, fall short when modeling complex or relational behavior. They fall short when modeling multi-touch customer journeys. This limit motivates the shift towards graph-aware churn modeling.

3. Graph Neural Networks and Graph Machine Learning

GNNs represent a significant breakthrough in machine learning for non-Euclidean data. According to Khemani et al. (2024), GNNs allow each node to aggregate information from its neighbors. This makes them well-suited for the applications that involve interconnected entities. This includes customers navigating multi-channel journeys. The authors provide a comprehensive review of GNN architectures like GAT, GCN, heterogeneous graph networks, and GraphSAGE.

Gao et al. (2023) further depict that GNNs outperform traditional deep learning approaches. They are effective in recommender systems as they capture user-item relationships, higher-order interactions, and collaborative filtering structures. The recommender systems share structural similarities with customer journey graphs. The insights from this domain are effective as they translate well to churn prediction and mapping of the journey.

The advantages of GNN include:

- Modeling influence between the related customers.
- Capturing the structural patterns in customer behavior.
- Incorporating temporal and sequential dependencies through dynamic graphs.
- Handling heterogeneous entities like products, channels, and devices.

4. Big Data-Driven Customer Journey Mapping

Machine learning is effective as it can be used to enhance marketing strategies. Singh and Yousuf (2024) indicate that through big-data-driven customer journey mapping, machine learning can boost marketing. Their research depicts the role of machine learning in analyzing large-scale behavioral data to define and identify actionable insights that can enhance segmentation. The authors recommend a journey-centric modeling where behavior is understood as part of an interconnected process and not as an isolated event. Their analysis supports the argument that modeling customer journey requires techniques that can synthesize complex, large datasets into coherent and interpretable structures. This is an objective that graph modeling can uniquely fulfill.

C. METHODOLOGY

Research Design

The study adopts a qualitative and conceptual research design based on the analysis of the resources provided. The approach is effective as it synthesizes the findings from customer journey analytics, graph neural networks, machine learning, and churn prediction. This is vital in producing an integrated conceptual framework. It aligns with academic standards for a research paper. It follows methodological principles that are used in conceptual information systems and machine learning research. The conclusions are drawn by analyzing the patterns and also building a framework from prior scholarly work.

Data Sources

The study draws insights from scholarly sources. The sources represent the major pillars that are required for conceptual investigation. Churn prediction research is highlighted by AbdelAziz & El-Khatib (2025) and Matuszelański & Kopczevska (2022). We also have sources that offer insights into graph neural

networks and machine learning frameworks. Other sources provide details on customer journey analytics and big data techniques.

The sources were selected as they are reputable academic sources. They provide direct insight into churn prediction, GNN methodologies, and journey mapping. The sources included quantitative, review-based, and conceptual contributions that are relevant to the objective of the study. They offer methodological detail that could be synthesized into a holistic framework. They also represent recent developments in the field as they are the latest sources.

Search Strategy

The methodology mirrors standard literature review protocols. The analysis followed a thematic synthesis process and helped in extracting vital information. Categorization of the sources allowed for structural mapping across different fields. This enabled the development of a multidisciplinary conceptual model. Although this research is conceptual, there are various methodological techniques that were employed to strengthen validity. For instance, triangulation across big data, machine learning, marketing, and recommender system literature. We have cross-verification where the findings about the complexity of the journey align with what other authors depicted. The sources are recent, which is vital in ensuring that the information is reliable and valuable.

Limitations of the Methodology

The methodology has various constraints. First, no empirical dataset was analyzed in the research. The insights are subject to the conceptual focus of the literature that is selected. There are real-world implementation challenges, and these include data sparsity, model explainability, and computing cost. All these are inferred, meaning they are not tested and confirmed.

D. FINDINGS

1. Customer Behavior Naturally Forms Graph Structures

The reviewed literature supports the idea that customer behavior contains underlying relational structures. The behavior can be physical, digital or spatial. The central finding across the literature indicates that customer interactions in the digital environment reflect a graph-like structure. Important to note that customer journeys are no longer sequences. These journeys are complex networks of interactions that happen across channels, platforms, and devices. They form nodes and edges that represent behavioral transitions.

Pantouvakis and Kottika (2022) indicate that modern customer journeys involve multi-dimensional, nonlinear, and interconnected behavioral paths. The paths include emotional, experimental, and cognitive components. These components evolve dynamically over time. They analyzed contemporary journey theories, and this aligns with the notion that traditional linear modeling always fails to capture the relational and interconnected nature of the experiences of customers in the real world.

Ismail and Abdulkareem (2024) depict that customer behavior in digital commerce produces huge amounts of data. It produces large-scale, multivariate datasets that are deeply connected. They include clickstreams, cart movement, browsing sequences, and cross-device interactions. This resembles a graph structure. Their quantitative framework depicts how customer interactions create natural relationships between touchpoints and events.

Matuszelański and Kopczevska (2022) demonstrate this relational complexity. They identify spatial and contextual dependencies in customer churn. Important to note that customers who share behavioral or geographic similarities often exhibit churn tendencies that are similar. Such dependencies are best captured

by models that recognize the influence of the neighborhood. This is an idea that is aligned with graph representation learning.

The conclusion of these findings indicates that customer behavior is relational and best represented as a graph and not as isolated rows in a dataset.

2. Traditional Churn Prediction Models Fail to Capture Relational Dependencies

This is a major finding. Conventional machine learning approaches fail to model the complex dependencies that influence churn despite being able to improve.

AbdelAziz and El-Khatib (2025) confirm this limitation by demonstrating that churn behavior clusters spatially and behaviorally. Clustering indicates that customer churn decisions are never independent. The decisions are influenced by peer groups, co-occurring behavioral patterns, and geographic proximity. Traditional ML tools like random forests or logistic regression cannot capture these relational and spatial dependencies. These findings depict that churn cannot be accurately modeled through isolated customer-level features. Important to note that relational, multi-hop behavioral and contextual dependencies require graph-based models for predictions to be accurate.

3. Graph Neural Networks Provide Superior Modeling for Multi-Touch Customer Journeys

Graph learning literature offers a theoretical and methodological finding. Gao et al. (2023) and Khemani et al. (2024) depict that graph neural networks outperform traditional Deep and ML learning approaches when working with relational and multi-entity behavioral data.

According to Gao et al. (2023), GNNs achieve state-of-the-art results in recommender systems as they learn from user-item interaction graphs. They learn from collaborative filtering structures and higher-order relationships. Important to note that customer journeys share these characteristics as they involve sequences of interactions across multiple entities. The similarity strengthens the argument that GNNs are suitable for journey modeling.

Khemani et al. (2024) provide detailed information about GNN architectures like Graph Attention Networks, GCN, and GraphSAGE. They show that these models excel in tasks that require message passing, neighbor aggregation, and relational reasoning. The research's findings depict that GNNs are capable of capturing long-range dependencies across multiple graph layers. This capability is crucial for the analysis of the customer journey.

When this is applied to a journey, it means that GNNs can analyze multi-touch transitions across channels. It means they can incorporate interaction context like device switching, time graphs or channel switching. It means they can model influence propagation. They can be employed to account for customer communities and segment the impact of their behaviors.

4. AI and Big Data Strengthens Customer Journey Mapping and Make GraphML Feasible.

The findings depict that the volume, velocity, and variety of modern customer data make large-scale graph modeling feasible. AI-driven journey analytics enable organizations to detect hidden journey patterns. It is easy to identify friction points. Organizations can personalize experiences at scale, and it is now easy to predict future customer behaviors.

Singh and Yousuf (2024) show that leveraging big data and ML helps improve marketing strategies as it enables real-time and granular analysis of customer interactions. They highlight the importance of large-scale behavioral data in generating insights that are actionable. They reinforce the need for models that are capable of handling complex datasets.

Big data is vital when integrated with graph modeling as it enriches node and edge attributes. It increases predictive accuracy as it provides more relational signals. Big data supports dynamic graph evolution, wh-

ere the journeys of customers change hourly or daily.

Quantitative analysis of journey behavior demands computational methods that are capable of synthesizing high-volume and high-dimensional data. These conditions are well-suited for GraphML.

5. GraphML Bridges Behavioral Complexity with Predictive Accuracy

The most integrative finding of the study is that graph machine learning creates a bridge between customer journey complexity and churn prediction accuracy. Important to note that this is a result of various factors, such as journey complexity, which is inherently relational. GNNs excel at modeling relational data. Churn behavior exhibits spatial and contextual clustering.

Synthesizing these insights depicts that machine learning provides a deeper behavioral understanding. It improves churn risk scoring. It improves the prediction of high-risk transitions. Helps in the identification of journey paths leading to churn. It guarantees better segmentation through graph clustering. It also improves targeted retention interventions. This means that GraphML is positioned to unify the fields of behavioral analytics, churn predictions, and customer journey mapping.

E. DISCUSSION

This integrates the findings from the literature to illustrate why machine learning is suited for customer journey analytics and churn prediction. It draws together key insights from the findings and interprets the significance.

1. Customer Journey Should Be Modeled as Dynamic Graphs

Customer journeys are multi-channel, nonlinear, and interconnected, and they are accurate when they are represented using graph structures. Customer experiences do not resemble funnel-shaped paths. They form complex webs of emotional, psychological, and behavioral interactions. Graph modeling is effective as it captures this complexity.

Quantitative approaches and AI-driven insights have been proven to be effective in interpreting modern journeys as they generate structured relational data across touchpoints and channels. Translating the patterns into graph form is vital as it allows analysts to see the journey of customers as an ecosystem that is interconnected.

2. Churn Prediction Improves When Relational Dependencies Are Modeled

Prediction has been limited by assumptions that customers act independently. Churn behavior is influenced by spatial dependencies, community patterns, and peer effects. Important to note that GraphML addresses this by modeling similarities between journeys and modeling customer-to-customer influence. Multi-step behavioral dependencies and spatial factors or shared contextual factors.

3. GNNs Provide the Analytical Backbone for Advanced Journey and Churn Modeling

Graph neural networks like Graph Attention Networks, GCN, GraphSAGE, and heterogeneous GNNs are effective as they provide computational power that is needed to process large-scale data.

4. Operational Benefits

Integration of GraphML into the journey of customer analytics provides various benefits that support retention, marketing, and customer experience. The benefits include identifying churn-inducing paths, as GNNs can detect patterns, and this facilitates better understanding. RNNs are in a position to rank-touchpoints, as graph attention layers are important in identifying the needs that have a great influence on churn.

GNNs are effective in allowing one to detect customer micro-communities. Clustering algorithms are able to reveal groups of customers that share similar patterns or risk profiles. It supports the implementation of

real-time retention strategies as it can update predictions.

5. Challenges

Literature supports the potential of GraphML; however, there are limitations that must be addressed. For instance, data integration is a challenge. Combining session, CRM, clickstream, and transition data into a unified graph demands robust governance.

We have computational complexity as GNNs are computationally intensive. This is common in large-scale digital commerce environments. Graph neural networks are complex than traditional ML models. GNN explainability tools are still maturing, and that presents a challenge. We have an ethical use of behavioral data challenge. The use of detailed journey graphs raises many concerns, and this includes data consent, fairness, and surveillance.

CONCLUSION

The research revealed that GraphML represents a transformative step forward in customer analytics. It is evident from the findings that modern customer behavior forms relational and complex networks rather than linear consequences. Important to note that customers interact with brands across multiple touchpoints, devices, and channels. This creates behavioral structures that are best modeled as graphs. Traditional machine learning is improving, but it is not able to handle the needs of modern technology and demand. They continue to treat customers as isolated entities, and this means they ignore dependencies that influence churn and journey patterns.

The graph neural networks address this limitation. GNNs excel at learning from relational and multi-entity data structures. They can capture higher-order dependencies. They can integrate heterogeneous data types and dynamically update predictions. This positions GNNs as ideal tools for understanding the complexity of the journey and predicting churn.

Graph machine learning provides a comprehensive and powerful framework for modeling customer journeys. It helps predict churn in ways that reflect the complexity of customer behavior. Digital interactions continue to grow in connectivity and volume. GraphML and GNNs are considered to be foundational components of modern marketing analytics. They are vital in improving customer experience management.

REFERENCES

1. AbdelAziz, N. M., & El-Khatib, N. (2025). A comprehensive evaluation of machine learning and deep learning models for customer churn prediction. *Information*, 16(7), 537. <https://doi.org/10.3390/info16070537>
2. Aluwala, A. (2022). Leveraging AI to enhance customer journey mapping and insights. *Science Data & Learning Machine Artificial Intelligence Journal*, 1(1), 1317–1322.
3. Gao, C., Zheng, Y., Li, N., Li, Y., Qin, Y., Piao, J., ... Li, Y. (2023). A survey of graph neural networks for recommender systems: Challenges, methods, and directions. *ACM Transactions on Recommender Systems*
4. Ismail, A. A. H. E., & Abdulkareem, A. M. (2024). Data-driven techniques for quantitative analysis of customer journey mapping in digital commerce. *Emerging Trends in Machine Intelligence and Big Data*, 16(4), 1-8.
5. Khemani, B., Patil, S., Kotecha, K., & Tanwar, S. (2024). A review of graph neural networks: Concepts, architectures, techniques, challenges, datasets, applications, and future directions. *Journal*

of Big Data, 11, 18. <https://doi.org/10.1186/s40537-023-00876-4>

6. Matuszelański, K., & Kopczewska, K. (2022). Customer churn in retail e-commerce business: Spatial and machine learning approach. *Journal of Theoretical and Applied Electronic Commerce Research*, 17(1), 165-198. <https://www.mdpi.com/0718-1876/17/1/9>
7. Pantouvakis, A., & Kottika, E. (2022). The theoretical and practical evolution of customer journey research. *Sustainability*, 14(15), 9610. <https://doi.org/10.3390/su14159610>
8. Singh, C. P., & Yousuf, R. (2024, May). Enhancing Marketing Strategies Through Big Data-Driven Customer Journey Mapping: An Analysis Using Machine Learning Algorithms. In *2024 International Conference on Emerging Innovations and Advanced Computing (INNOCOMP)* (pp. 479-484). IEEE.