WHITE PAPER: By Dmitri Kaznachey and Kelly Dakin

Predicting New-to-Bank Customer Attrition Using Graph Neural Networks

Executive Summary

Banks and credit unions continue to struggle with elevated new-to-bank (NTB) customer attrition, particularly within the first 12 months of account opening. Traditional analytics approaches rely heavily on tabular, account-level features and often fail to capture the relational, behavioral, and engagement patterns that determine whether a newly acquired customer will remain active or quietly disengage.

This white paper documents a production-oriented proof of concept (POC) developed to test the feasibility and performance of using graph-based machine learning, including Graph Neural Networks (GNNs) and related advanced analytics techniques, to predict NTB customer attrition. Using one year of real customer data from a financial institution, the team constructed a multi-entity graph representation of customer behavior, trained multiple models of increasing sophistication, and evaluated their ability to predict attrition for a future cohort of newly onboarded customers.

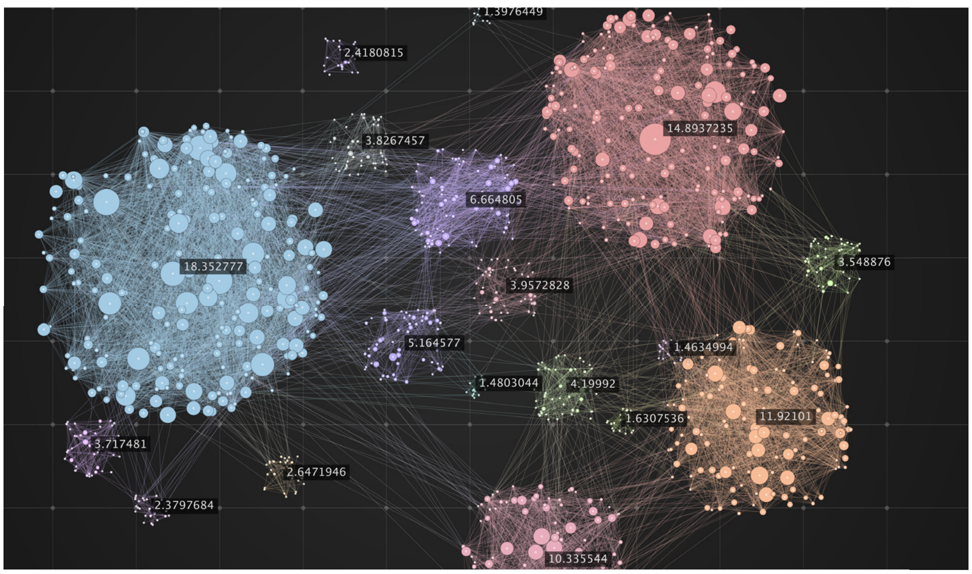
The results demonstrate that graph-aware models materially improve predictive performance over traditional approaches by explicitly modeling how customers interact with accounts, products, channels, and services over time.

Business Problem

The Cost of Early New-to-Bank Attrition

New-to-bank customers represent one of the most fragile—and expensive—segments in retail banking. Industry benchmarks consistently show that first year churn among newly acquired customers ranges from 20% to 25%, meaning that as many as one in four new customers disengages within 12 months of opening an account. Roughly half of new customer attrition occurs within the first 90 days, during and immediately after onboarding—when customers are forming their initial perceptions of trust, ease, and relevance.

This early concentration of attrition is critical. It indicates that new-to-bank churn is rarely driven by long term lifecycle factors such as rates or product maturity. Instead, it is most often the result of unresolved onboarding friction, broken journeys, unmet expectations, and early negative experiences that are not identified or corrected quickly enough.



Why Early Churn Is So Economically Damaging

The financial implications of early attrition are severe. While exact figures vary by institution, industry estimates put customer acquisition cost (CAC) for retail banking customers in the range of several hundred dollars per account, with many analyses citing approximately $400–$600 per new customer, depending on channel, product mix, and incentives.

At the same time, most retail banking relationships do not become profitable until the second year or later, after onboarding costs are absorbed and customers begin to consolidate balances, adopt additional products, or establish primary bank behavior. When 20%–25% of new customers leave before reaching that breakeven point, the institution effectively funds acquisition and onboarding expenses without ever realizing the revenue or lifetime value those customers were expected to generate.

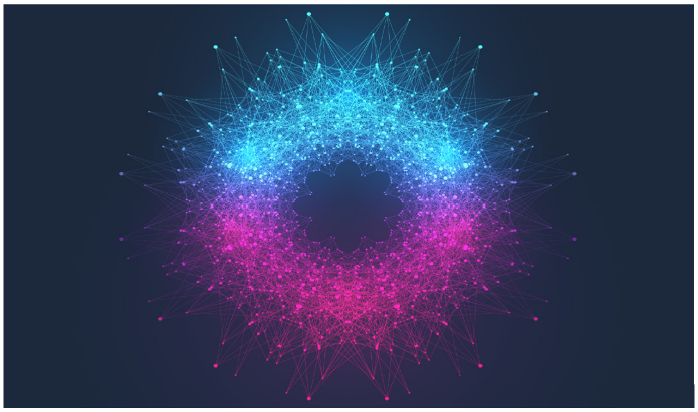
The economics are straightforward. If the average acquisition cost is $500 per customer and 20% of a new customer cohort churns in the first year, that equates to $125 in lost acquisition investment per customer, on average, before accounting for the far larger opportunity cost of lost cross-sell, deposits, lending, and long-term relationship value. Scaled across thousands—or tens of thousands—of new accounts, early attrition quickly translates into millions of dollars in preventable annual losses.

Why Early Prediction and Root Cause Insight Is the Only Sustainable Fix

These figures highlight a central reality: new-to-bank attrition is not a volume problem—it is a visibility problem. Banks are often aware that customers are leaving, but lack the insight needed to pinpoint why they disengage so early, where friction accumulates across onboarding and early journeys, and which customers are worth proactive remediation.

Without uncovering the root causes of early disengagement—whether they stem from account setup failures, digital access issues, service breakdowns, fraud anxiety, or unmet expectations—institutions are left reacting after the fact. By the time churn appears in traditional reporting, the opportunity to repair trust and reset the relationship has already passed.

Understanding how early experiences compound into disengagement, and intervening before customers silently exit, is therefore essential—not only to reduce attrition, but to ensure that acquisition investments translate into durable, profitable relationships.



New-to-bank customers represent a significant acquisition investment. When these customers disengage within their first year, institutions face:

- Lost deposits and fee revenue
- Increased funding pressure as deposits roll off
- Higher customer acquisition costs to replace lost balances

Most existing attrition models treat customers as isolated records. Customer behavior is inherently relational. Customers transact with other customers, interact with multiple accounts, products, channels, and service touchpoints, and it is the structure and evolution of these interactions that often signal future inactivity.

The objective of this POC was to determine whether graph-based modeling could surface these signals earlier and more accurately than traditional methods.

Data and Cohort Definition

Cohort Construction

The team obtained a full year of historical customer data covering a cohort of newly onboarded customers. Each customer was observed for their first 12 months following account opening.

At the end of the 12-month observation window, each customer was assigned an attrition indicator:

1 = Customer became inactive

0 = Customer remained active

This labeled cohort served as the foundation for model training.

Holdout and Forecasting Window

To simulate real-world deployment:

- Models were trained exclusively on historical data from the initial cohort
- A fresh cohort of NTB customers in month 13 was used for prediction
- Predictions were back-tested against observed outcomes to evaluate performance

Graph Data Model

Core Design Principles

The graph was designed to reflect how banks experience customer behavior operationally, rather than forcing interactions into flattened tables.

Key principles included:

- Explicit separation of customers, accounts, products, transactions, channels, and interactions
- Preservation of temporal signals through transactional and service events
- Flexibility to support both traditional graph analytics and deep learning models

Edge Definitions

Edges capture the relationships and interactions between nodes, including:

- Member-to-Account ownership
- Account-to-Product classification
- Transaction-to-Account monetary transfers between accounts
- Member-to-Service interactions capturing engagement intensity
- Service Interaction-to-Channel indicating preferred engagement paths

This structure allowed the graph to encode both behavioral frequency and relationship topology.

Node Types

The graph model consisted of the following primary node types:

1. Member Nodes

- Represent individual customers

2. Account Nodes

- Represent financial accounts associated with members
- Examples include checking, savings, and certificates of deposit

3. Product Nodes

- Represent product categories
- Accounts link to specific product types

4. Transaction Nodes

- Represent funds moved between accounts
- Examples include direct deposits, intra-bank transfers, external transfers

5. Channel Nodes

- Represent engagement channels such as branch, website, and mobile

6. Service Interaction Nodes

- Represent service events such as inquiries, issue resolution, or technical support

Feature Engineering and Graph Representations

Rather than relying solely on handcrafted features, the POC leveraged multiple graph-aware representations:

Structural Metrics

- Centrality measures to capture customer importance and connectivity
- Measures reflecting engagement breadth across products and channels

End-to-End Graph Learning

- A graph attention network leveraging neighborhood aggregation
- Learned representations updated through message passing across the graph

Embedding Techniques

- Node embeddings generated using metapath2vec to capture higher-order relationship patterns
- Embeddings preserved similarity between customers with comparable behavioral paths

Modeling Approach

Experimental Design

Six experiments were conducted, each increasing in modeling complexity while holding the evaluation workflow constant.

Across all experiments:

- A stratified sample was created with an equal split of active and inactive customers
- Models were trained on historical cohort data
- Predictions were generated for a future NTB cohort
- Back-testing was performed to assess predictive accuracy and stability

Evaluation Methodology

Model performance was evaluated using:

- Back tested predictions on unseen NTB customers using standard classification metrics.
- Stability analysis across multiple sampling runs

Special attention was paid to:

- Early warning capability
- False positives that could trigger unnecessary retention costs
- Interpretability of model drivers

Explainability and Business Insights

Although GNNs are often viewed as black boxes, the POC incorporated explainability techniques to surface:

- The most influential relationship patterns
- Critical interaction drop-off points
- Channel and service behaviors correlated with disengagement

These insights provide a foundation for downstream action frameworks without hardcoding tactics into the predictive layer.

Models Evaluated

1. Baseline Random Forest
- Trained on traditional tabular features
- Served as a performance benchmark

2. Random Forest with Graph Derived Features
- Incorporated centrality metrics

3. Metapath2vec Embedding Models
- Used learned node embeddings as input features

4. Graph Neural Network
- Trained directly on the graph
- Learned representations through message passing across nodes and edges

Key Findings

The experiments demonstrated several consistent outcomes:

- Graph aware models outperformed traditional models in predicting NTB attrition
- Relationship structure and engagement pathways were strong leading indicators of inactivity
- Customers with narrow, shallow engagement graphs were significantly more likely to attrite
- GNN models captured nonlinear interaction effects that were invisible to tabular approaches

Importantly, these gains were achieved without embedding any post hoc retention tactics into the model. The focus remained solely on prediction and identifying root-cause signals driving retention outcomes.

Conclusion

This proof of concept demonstrates that Graph Neural Networks and related graph-based analytics are not only feasible for predicting new-to-bank customer attrition, but meaningfully superior to traditional approaches when customer behavior is inherently relational.

By modeling customers as part of an interconnected system of accounts, products, channels, and interactions, banks can detect early warning signals of disengagement with greater accuracy, confidence, and business relevance.

This work establishes a strong technical and conceptual foundation for production grade customer analytics platforms designed to address the most complex behavioral challenges in modern financial institutions.

Concrete Case Anecdote from POC: Overdraft as an Attrition Trigger

One cluster identified during the analysis exhibited an 88 percent attrition rate despite appearing, by traditional measures, to be highly valuable. Members in this group had higher-than-average income levels and transaction volumes, signaling strong engagement and financial capacity. However, graph-based clustering revealed a critical and previously obscured pattern: 94 percent of these members experienced a checking account overdraft shortly before becoming inactive.

This finding fundamentally reframed the attrition narrative.

These were not financially distressed customers gradually disengaging due to insufficient funds. Instead, the overdraft fee acted as a discrete trust-breaking trigger: customers had overdrawn temporarily, but their accounts and balances were otherwise healthy. In tabular models, they appeared fully engaged until the moment they disappeared. In the graph, however, their engagement pathways narrowed, and their exposure to overdraft events surfaced weeks earlier, providing clear early-warning signals that traditional analytics failed to detect.

Armed with this insight, the bank could intervene with precision. Rather than deploying broad fee waivers or generic retention campaigns, the institution could proactively protect this specific segment through targeted overdraft buffers, real-time balance alerts tied to spending velocity, or temporary fee forgiveness during the first year of membership. The result was not only preventable attrition, but a direct line of sight between customer experience, trust erosion, and balance sheet impact. This case illustrates how graph-based analytics shift attrition prediction from reactive risk scoring to root-cause detection, enabling banks to retain high-value customers before a single negative experience ends the relationship.

Limitations and Future Work

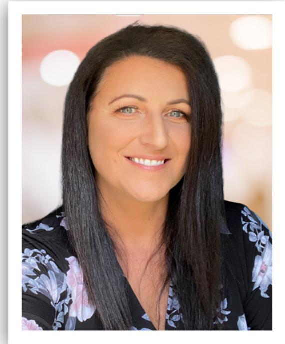
This POC was intentionally scoped to feasibility and signal discovery. Areas for future expansion include:

- Longer observation windows beyond the first year
- Dynamic graph updates in near real time
- Integration with household modeling
- Extension to retention, growth, and cross selling

Author: Dmitri Kaznachey

As Chief Technology Officer, Dmitri Kaznachey works across project teams to lead the design and delivery of complex software solutions for FI Consulting's clients. Dmitri has 40 years of experience in software engineering, including senior leadership roles in delivering complex software systems in the finance and healthcare sectors. He is also an adjunct professor at George Mason University, where he teaches graduate-level courses in computer science, including mathematics, algorithms, human-computer interaction, artificial intelligence, and more. He earned his PhD from the University of Memphis on applying neural networks to graph problems. Dmitri is a holder of the Chartered Financial AnalystTM designation.

Author: Kelly Dakin



Kelly is a seasoned financial services executive with over 25 years of leadership across digital banking, customer analytics, and product innovation. Kelly is currently Director, Banking Analytics at FI Consulting. In this role, she helps banks and credit unions solve their most complex data challenges ranging from customer retention and product propensity to fraud detection and financial householding through advanced graph analytics.

Her career spans private wealth management, large national banks, regional banks, credit unions and fintech companies, where she has consistently led initiatives that modernize customer experience, unify fragmented data, and increase lifetime value. Kelly is known for driving innovation, building high-performing teams, and delivering measurable results in high-stake environments. She brings a unique blend of business acumen, digital strategy, technical fluency, and a relentless focus on improving financial outcomes for both institutions and customers.

If you are interested in learning more about how FI Consulting can support your organization, please email Kelly Dakin at kdakin@ficonsulting.com or call us at 571.255.6889.

FI·CONSULTING

2107 Wilson Boulevard | Suite 400 | Arlington, VA 22201 | 571.255.6900 | ficonsulting.com | info@ficonsulting.com