The Promise and Challenge of ML on Graphs in Financial Services

3/4/21 3:30 PM

Speaker: Bayan Bruss (CapitalOne)
Date: 03/04/2021

Part 1: Graphs in Finance

Graphs are everywhere in financial services

- E.g. customers can be connected to lenders by loans
- E.g. In banking the edge can be directed representing ACH, wire or other transfers

Customer graphs can be used in a variety of use-cases across customer lifecycle (marketing, fraud detection, acquisition…)

Part 2: Machine Learning on Graphs

Challenges

- **High dimensional**: millions of nodes and possibly billions of connections
- **Sparse**: each node interacts with a small number of other nodes

Recent advances in Graph ML enables models to overcome these challenges when applying ML directly to the graph

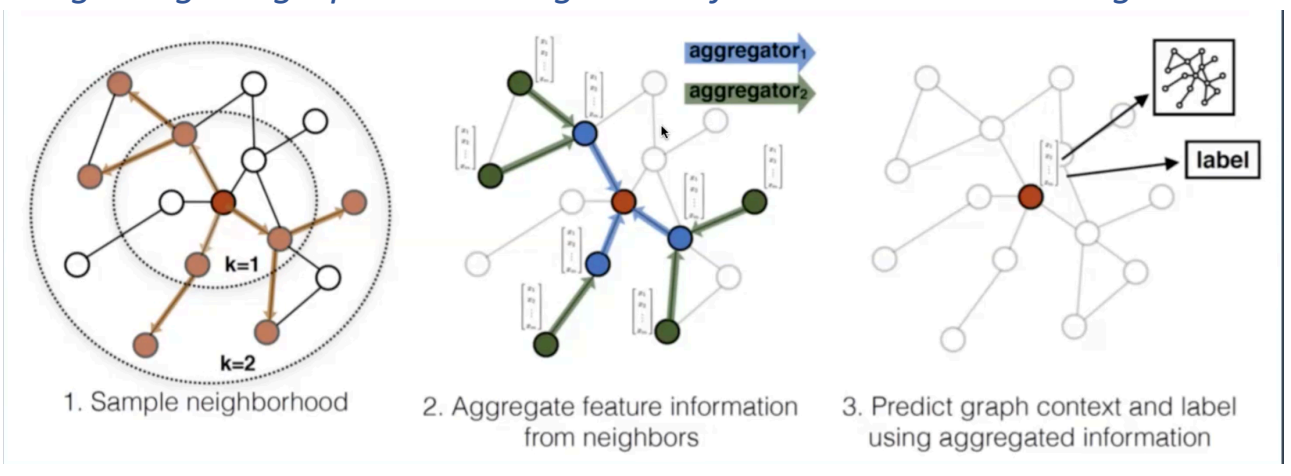
Graph ML has a variety of approaches depending on the context and data

Methods	Description
Graph mining	Analyze structural properties and identify groups with similar patterns
Graph Representation Learning (Unsupervised)	Directly encode the topology of the network as an objective
Semi-supervised	Propagate information from a small subset of labelled nodes to the rest of the graph
Supervised	Combine structural positioning & information propagation with a node classification task

The global structure an provide valuable customer insights

- Community detection
- Similarity definition through shortest path, node importance, etc.
- Information propagation from neighbors

Integrating the graph with a target label for end-to-end modeling

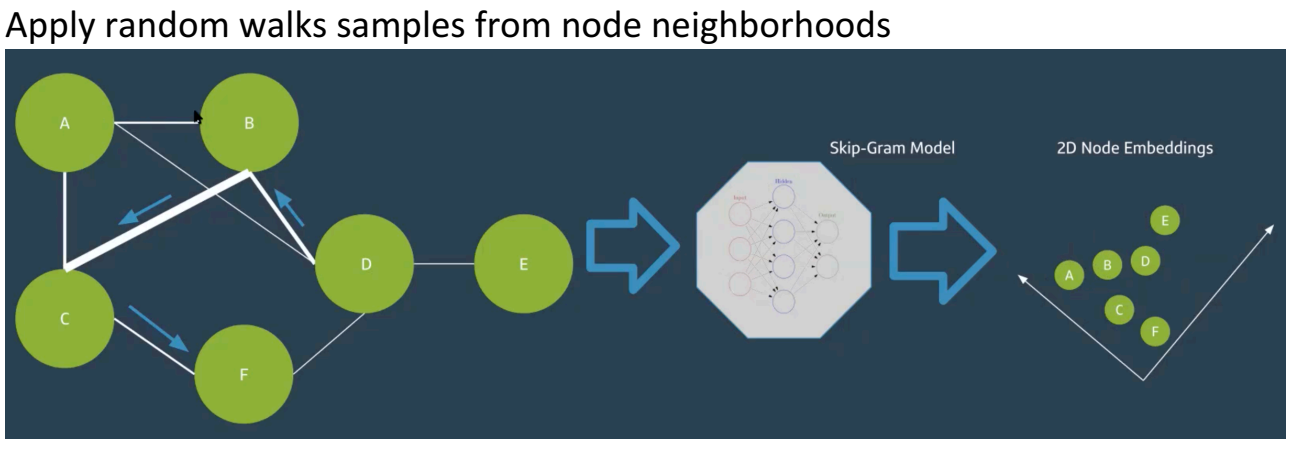


1. Sample neighborhood
2. Aggregate feature information from neighbors
3. Predict graph context and label using aggregated information

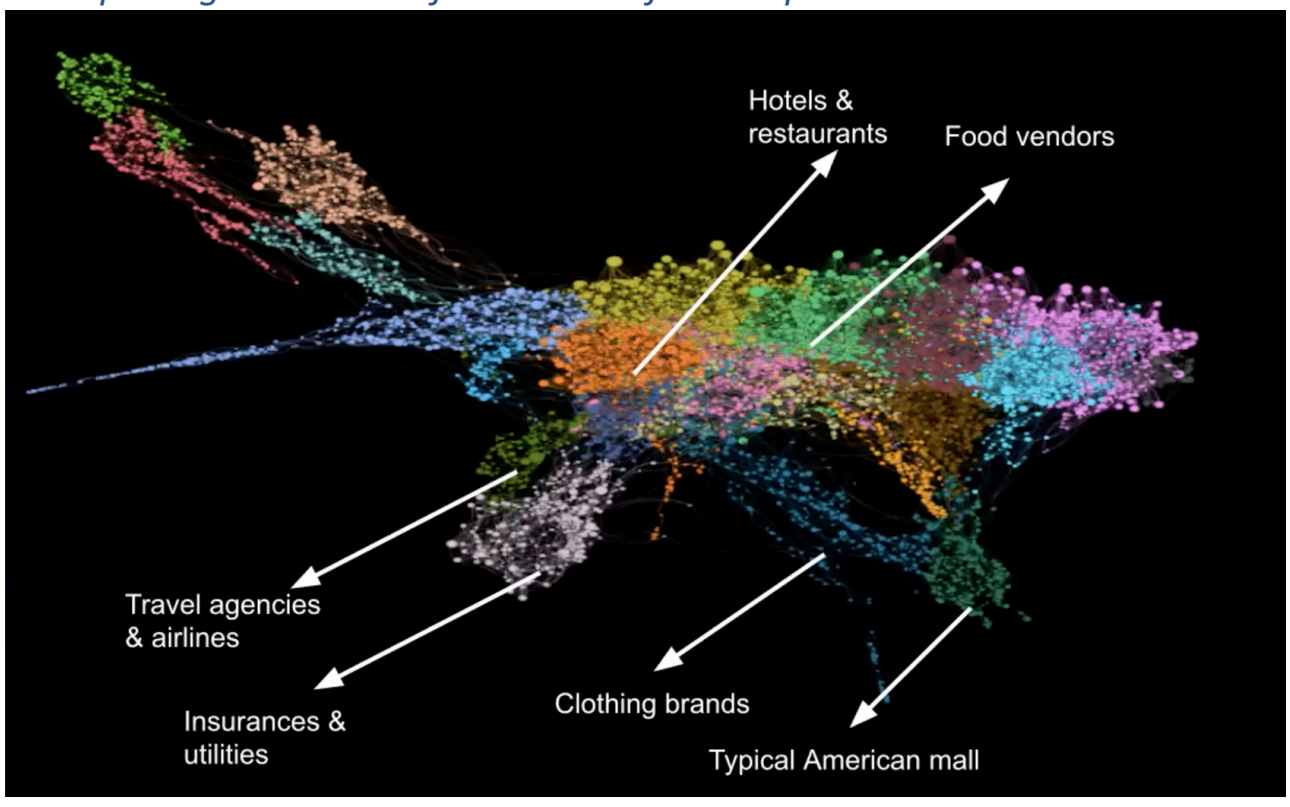
Part 3: What's been done so far

Transaction Graph Embeddings

Applied unsupervised graph machine learning on the graph of accounts & merchants based on credit card transactions

- Operate on a bipartite graph (two different nodes only) of accounts and merchants weighted by transaction frequency
- Graph projection creates two homogeneous graphs to be individually embedded into 2 separate spaces
  - One graph for merchants, one for accounts
  - Edges connect merchants or accounts with same activity engagement
- Apply random walks samples from node neighborhoods
- 

Example segmentation of merchants from Capital One transaction data



- Graph regenerated based on top-k nearest neighbors in the embedding space
- Colors based on modularity community detection

Conclusion

1. Natural Graphs abound in financial services
2. They provide a unique and high value look at customer behavior
3. They present unique challenges when seeking to apply machine learning