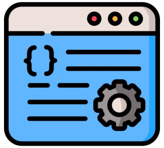
Welcome! Before we start...

Let’s go to the code!



shorturl.at/cgwB3: shorturl.at/cgwB3

Graph Machine Learning: The Next Frontier in Artificial Intelligence

Alejandro Correa Bahnsen, PhD
Chief AI Officer
RappiBank

Jaime D. Acevedo-Viloria
Machine Learning Scientist
Neural Design

Luisa Roa
Data Scientist

Graph Machine Learning

1	2	3
Value of Graphs	Graph Basics	ML with Graphs

4 Graph Neural Networks

Value of graphs

1

1 What is a graph

Graphs are a powerful tool in data representation, taking into account not only the data point themselves but also their interactions

This allow us to leverage not only the data points characteristics, but also **the context those interactions provide**

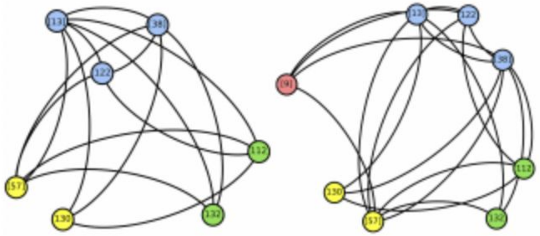
1 Why are graphs useful?

Because they provide us:

Context

Dynamic

Practicality



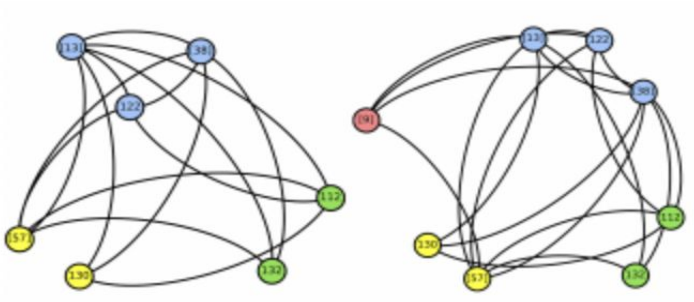
1 Why are graphs useful?

Context

Leveraging the environment of the User through the interactions and neighborhoods

1 Why are graphs useful?

Dynamic



Allowing us to leverage **changes to the user context**

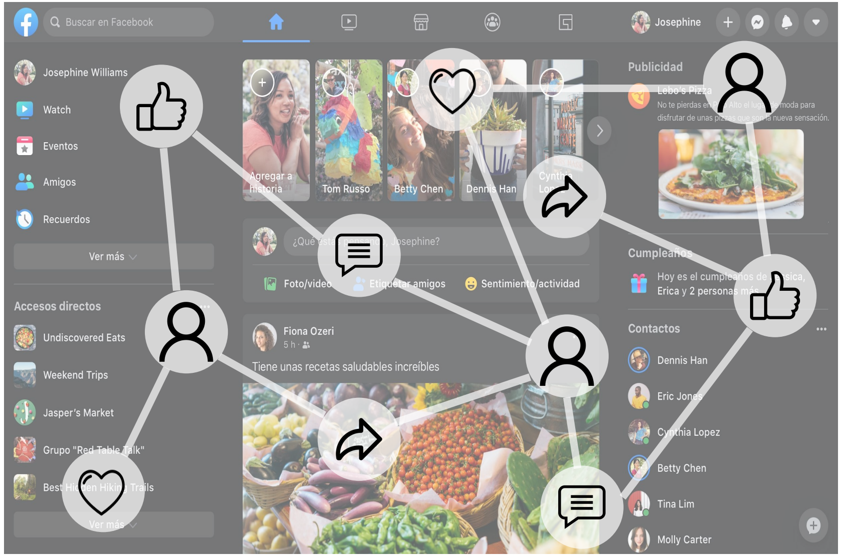
1 Why are graphs useful?

Practicality

They allow us to **visually explain** our logic and solutions

1 When can we use Graphs?

They are practically everywhere, so there are plenty of possibilities…

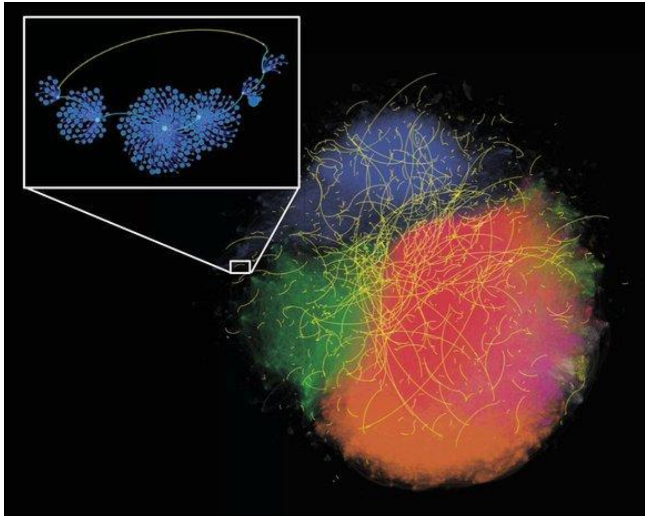


1 When can we use Graphs?

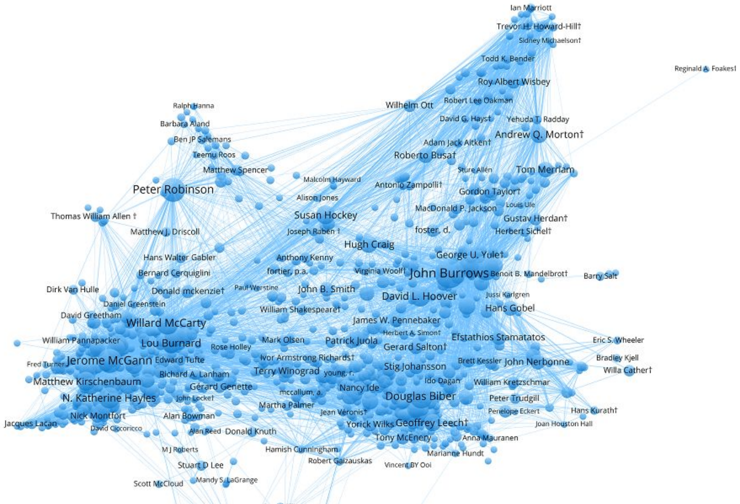
They are practically everywhere, so there are plenty of possibilities...

1 When can we use Graphs?

They are practically everywhere, so there are plenty of possibilities...



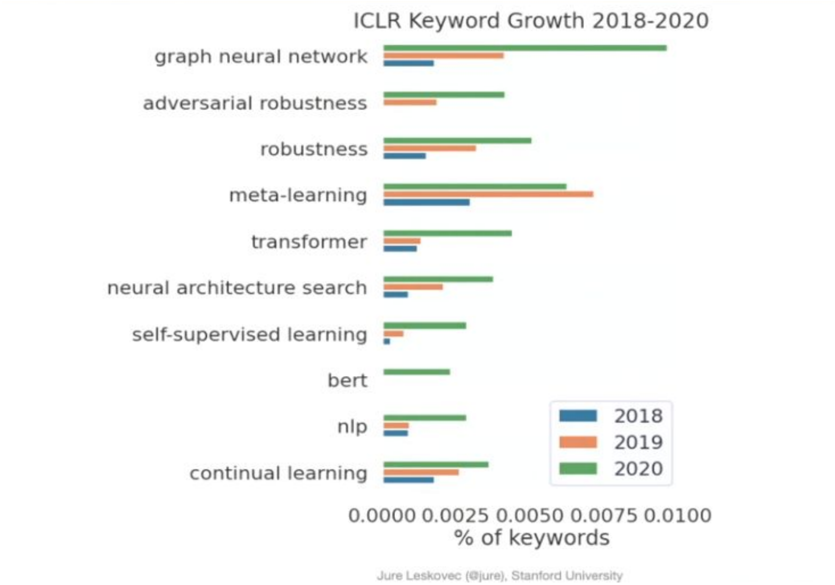
Singapore's Twitter network
https://science.sciencemag.org/content/362/6421/1410



Digital Humanities Citation Network
https://dh2018.adho.org/en/visualising-the-digital-humanities-community-a-comparison-study-between-citation-network-and-social-network/

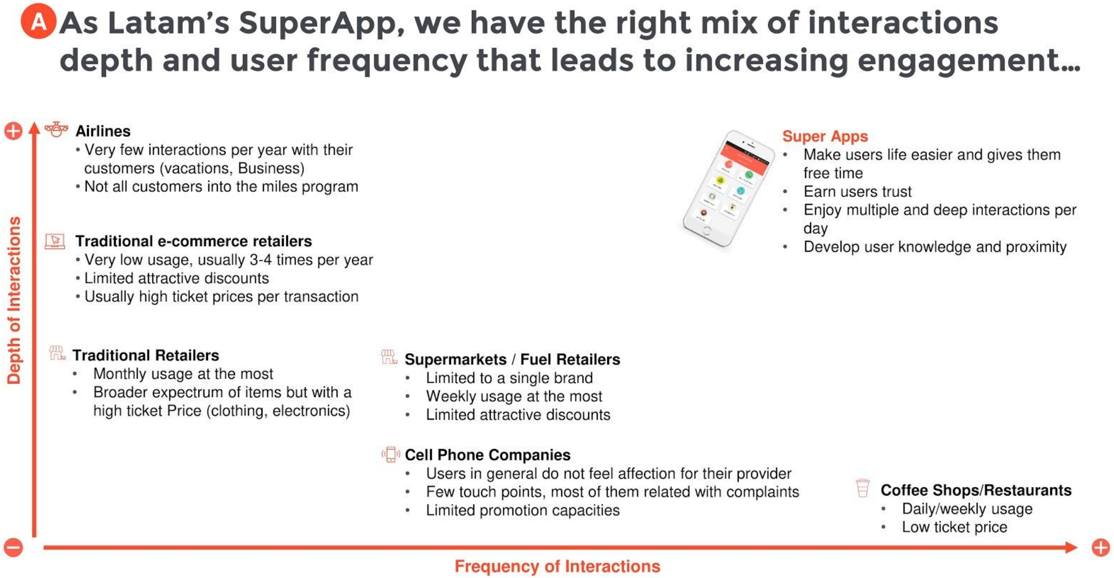
1 When can we use Graphs?

This has led to GML subjects becoming one of the most popular research subjects, and the tool for powerful engines in top companies



1 When can we use Graphs?

This subject becomes especially relevant in a Super-App, where the high frequency of use creates many interactions



All these aspects lead to...

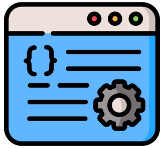
Connections in data are as valuable as the data itself

Graph Basics

2

2 Different types of Graphs

Let’s go to the code!



URL: shorturl.at/cgwB3

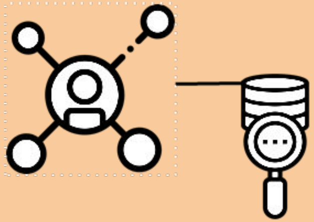
ML with Graphs

3

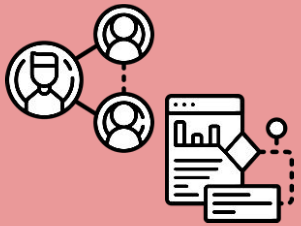
3 How to leverage Graphs in ML?

Therefore we can use different tactics to gain insight into complex systems through graphs

Graph Querying



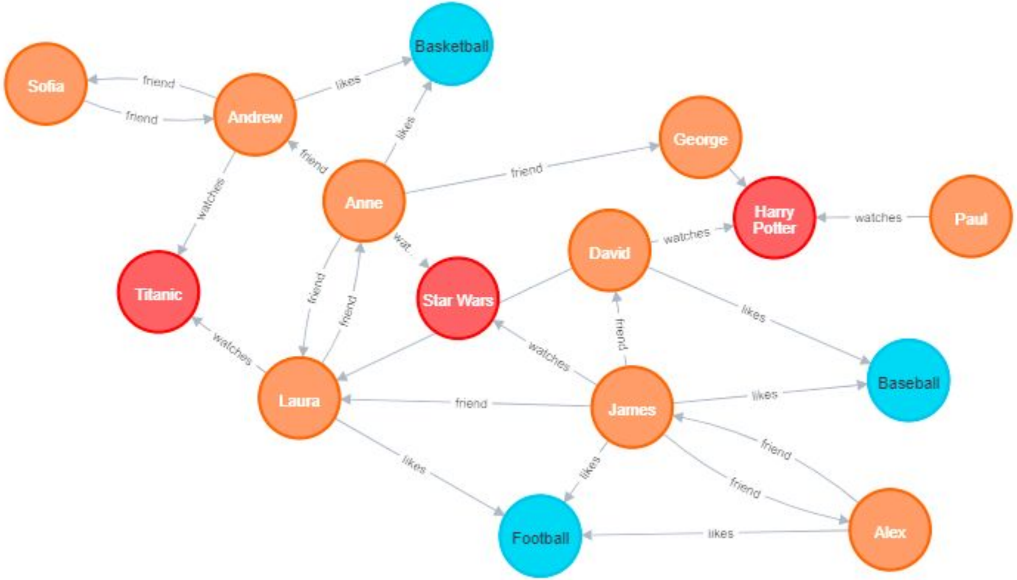
Graph Algorithms



Graph-Based Features

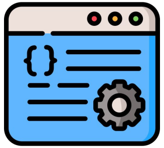
3 Graph Querying

Exploring graphs locally to gain insight into the Users context and take better decisions



3 Graph Algorithms and Graph-based Features

Let's go to the code!



URL: shorturl.at/cgwB3

3 Graph Algorithms and Graph-based Features

Let 's review!

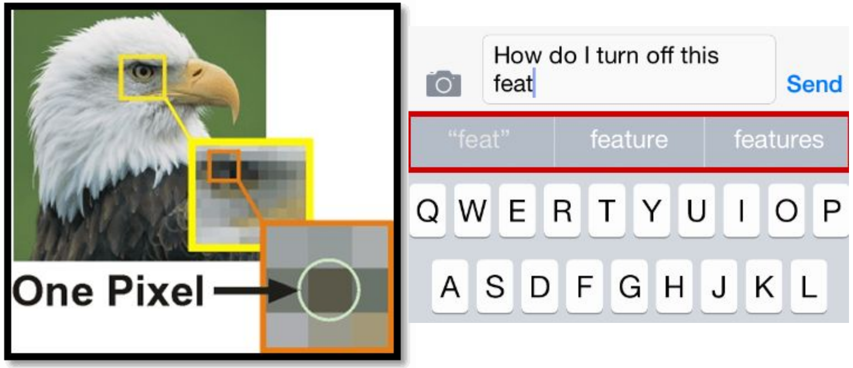
URL: shorturl.at/gikE3

Graph Neural Networks

4

4 What are GNN's?

They are Neural Network adaptations to non-structured data

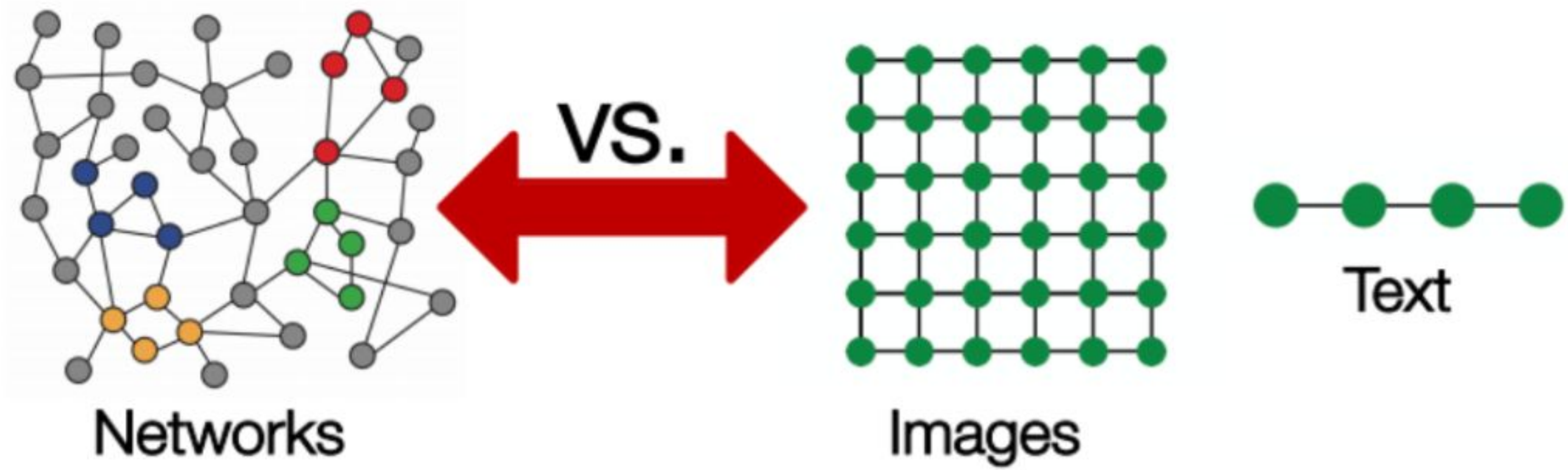


Non-structured Data

Structured Data

4 What are GNN's?

They are Neural Network adaptations to non-structured data

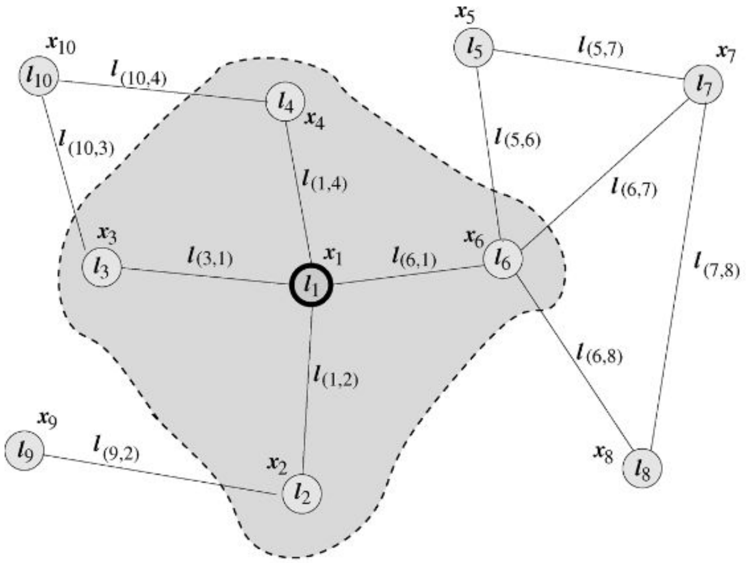


Non-structured Data

Structured Data

4 What are GNN’s?

GNN’s learn a representation of a node or interactions given their own properties and the properties of their neghborhood

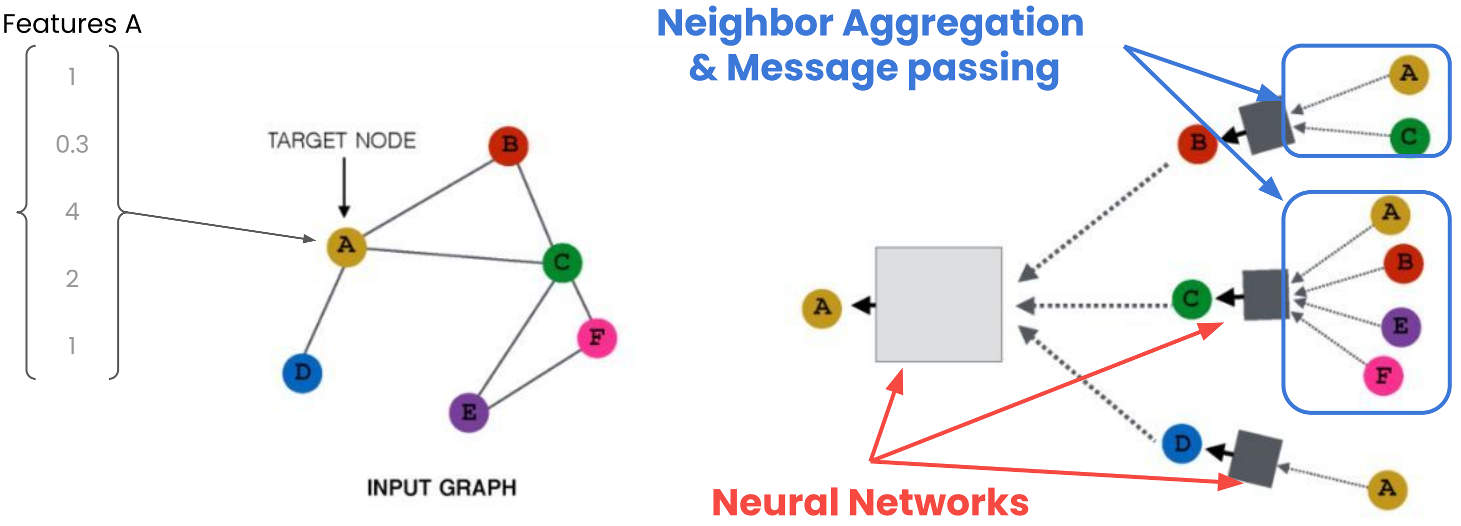


4 What are GNN's?

And this representation is done through a set of operations known in the literature as Message Passing

4 What are GNN's?

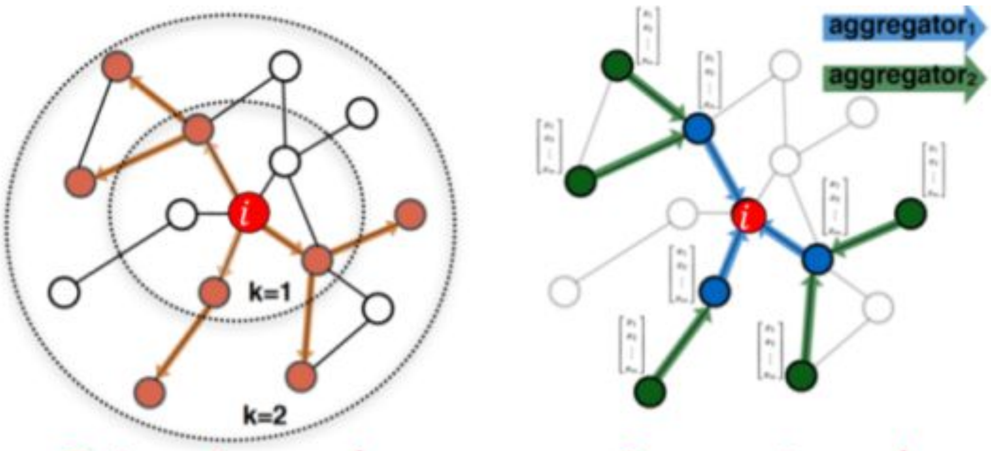
And this representation is done through a set of operations known in the literature as Message Passing



Representation Learning on Networks, snap.stanford.edu/proj/embeddings-www, WWW 2018

4 What are GNN's?

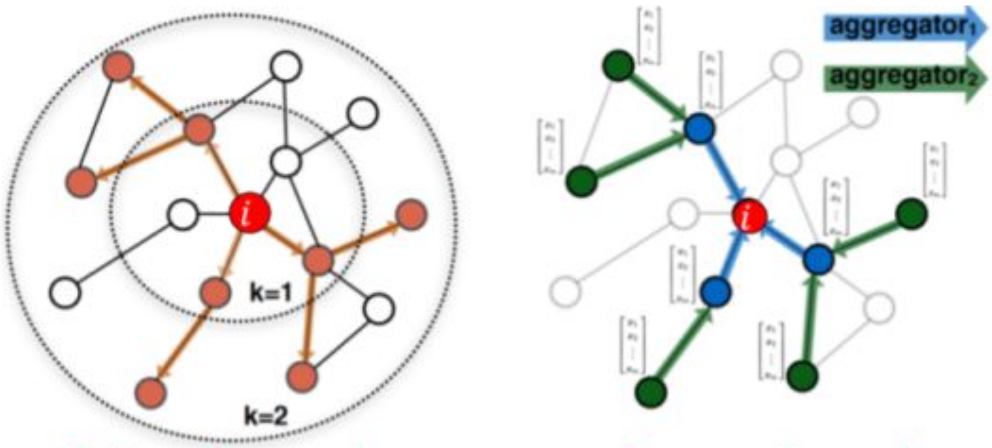
This generates a peculiarity when comparing to other neural networks:



As we add more layers to the GNN we are also adding more context

4 What are GNN’s?

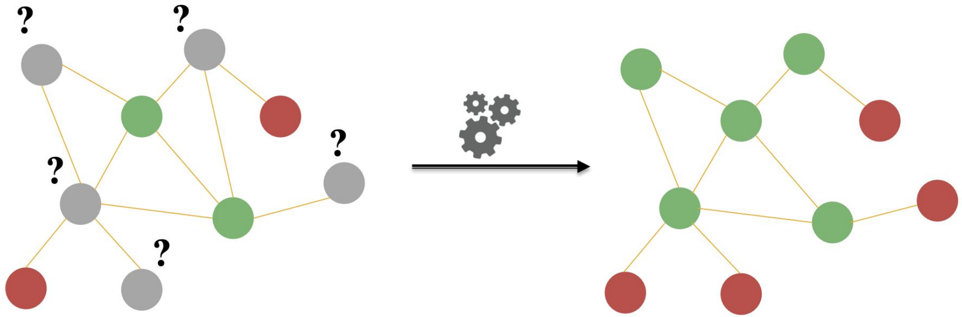
This generates a peculiarity when comparing to other neural networks:



As we add more layers to the GNN we are also adding more context

4 GNN’s – ML Tasks

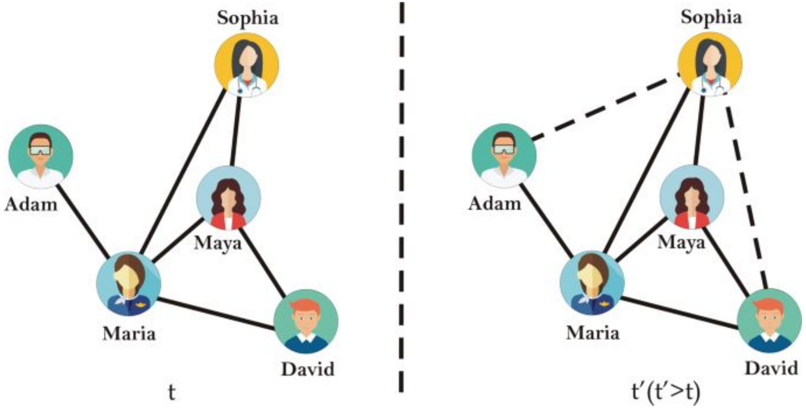
Node Classification



First academic publication of GNN’s!

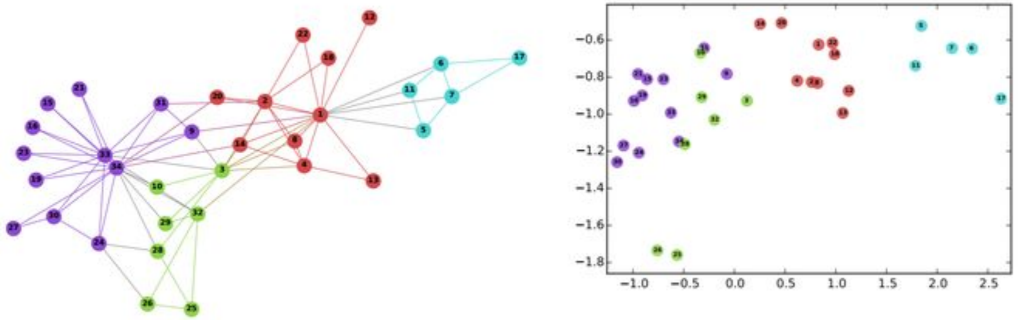
4 GNN’s – ML Tasks

Link Prediction



4 GNN's – ML Tasks

Graph Representation



(a) Input: Karate Graph (b) Output: Representation

4 GNN’s – Implementation

Let’s go to the code!

URL: shorturl.at/gikE3

Closing Notes

Relevant Papers we've developed

Supporting Financial Inclusion with Graph Machine Learning and Super-App Alternative Data

Luisa Roa[1], Andrés Rodríguez-Rey[2], Alejandro Correa-Bahnsen[1], and Carlos Valencia Arboleda[3]

[1] Rappi, Bogotá, Colombia
[2] University of California, San Diego, La Jolla, CA
[3] Universidad de los Andes, Bogotá, Colombia

Abstract. The presence of Super-Apps have changed the way we think about the interactions between users and commerce. It then comes as no surprise that it is also redefining the way banking is done. The paper investigates how different interactions between users within a Super-App provide a new source of information to predict borrower behavior. To this end, two experiments with different graph-based methodologies are proposed, the first uses graph based features as input in a classification model and the second uses graph neural networks. Our results show that variables of centrality, behavior of neighboring users and transactionality of a user constituted new forms of knowledge that enhance statistical and financial performance of credit risk models. Furthermore, opportunities are identified for Super-Apps to redefine the definition of credit risk by contemplating all the environment that their platforms entail, leading to a more inclusive financial system.

Keywords: Credit Score, Graph Machine Learning, Alternative Data, Super-App

https://arxiv.org/abs/2102.09974

Relational Graph Neural Networks for Fraud Detection in a Super-App environment

JAIME D. ACEVEDO-VILORIA*, Rappi, Colombia
LUISA ROA, Rappi, Colombia
SOJI ADESHINA, University of California Berkeley, USA
CESAR CHARALLA OLAZO, Rappi, Perú
ANDRÉS RODRÍGUEZ-REY, University of California, San Diego, USA
JOSE ALBERTO RAMOS, Rappi, México
ALEJANDRO CORREA-BAHNSEN, Rappi, Colombia

Large digital platforms create environments where different types of user interactions are captured, these relationships offer a novel source of information for fraud detection problems. In this paper we propose a framework of relational graph convolutional networks methods for fraudulent behaviour prevention in the financial services of a Super-App. To this end, we apply the framework on different heterogeneous graphs of users, devices, and credit cards; and finally use an interpretability algorithm for graph neural networks to determine the most important relations to the classification task of the users. Our results show that there is an added value when considering models that take advantage of the alternative data of the Super-App and the interactions found in their high connectivity, further proofing how they can leverage that into better decisions and fraud detection strategies.

CCS Concepts: • **Applied computing → Secure online transactions; Online banking**; • **Computing methodologies** → *Neural networks; Artificial intelligence; Machine learning*; • **Mathematics of computing** → Graph theory.

Additional Key Words and Phrases: Fraud Detection, Graph Neural Networks, Super-App, Geometric Deep Learning

ACM Reference Format:
Jaime D. Acevedo-Viloria, Luisa Roa, Soji Adeshina, Cesar Charalla Olazo, Andrés Rodriguez-Rey, Jose Alberto Ramos, and Alejandro Correa-Bahnsen. 2021. Relational Graph Neural Networks for Fraud Detection in a Super-App environment. In *KDD-MLF '21, August 14–18, 2021, Virtual Workshop*. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/1122445.1122456

https://arxiv.org/abs/2107.13673

Takeaways

- Graphs are a novel source of information that improve decision-making.
- Graph-based features enhance machine learning models performance.
- Companies should not limit their analysis/prediction to traditional methods.

Additional Resources

Get started

- Paper - A Comprehensive Survey on Graph Neural Networks (https://arxiv.org/abs/1901.00596)
- Paper repository - Must read papers on GNN (https://github.com/thunlp/GNNPapers#survey-papers)
- Stanford course - Machine Learning with Graphs (http://web.stanford.edu/class/cs224w/)
- Book - Graph Algorithms: Practical Examples in Apache Spark and Neo4j (https://neo4j.com/graph-algorithms-book/)
- Library - Deep Graph Library (https://www.dgl.ai/)

Contact us

Contact us for more information, questions about this subject or research ideas!

Jaime D. Acevedo-Viloria

james@neuraldesign.co

https://www.linkedin.com/in/jd-acevedoviloria/

Alejandro Correa Bahnsen

alejandro.correa@rappi.com

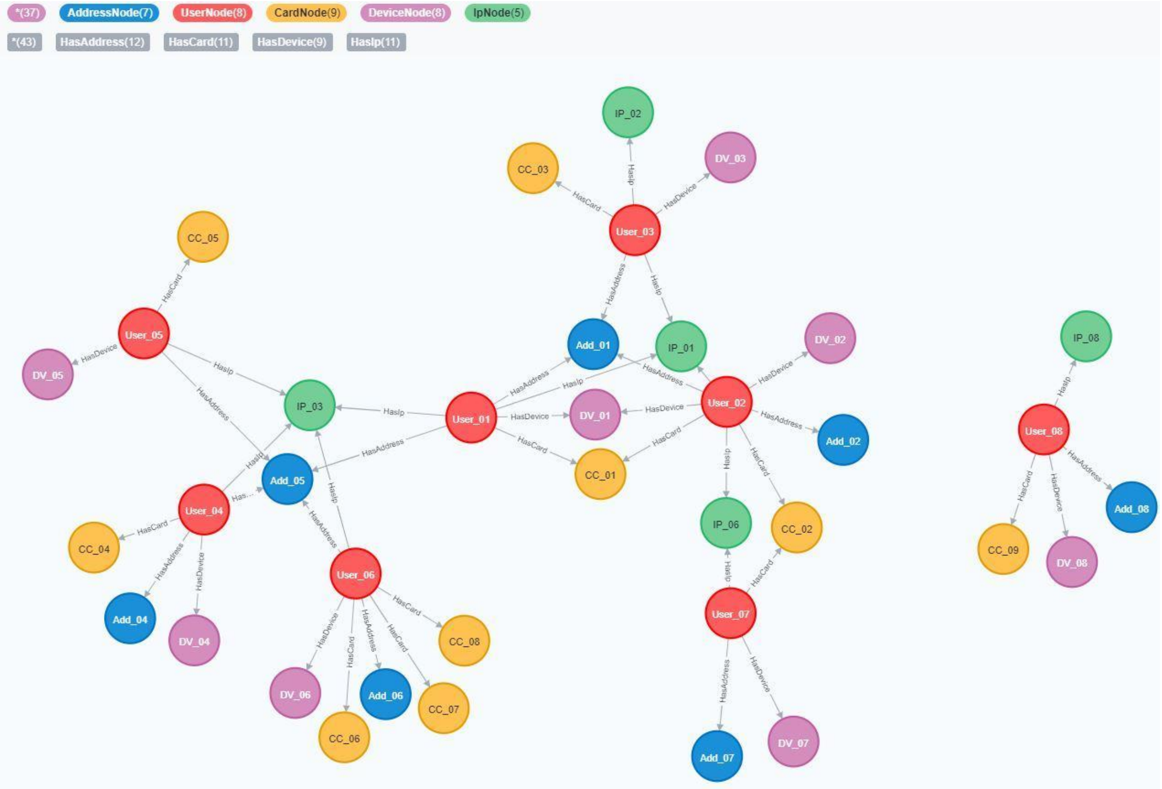
https://www.linkedin.com/in/albahnsen/

¡Thanks!

Real World use cases

Rappi's **Graph-Based** Features

Analyzing users as part of their community

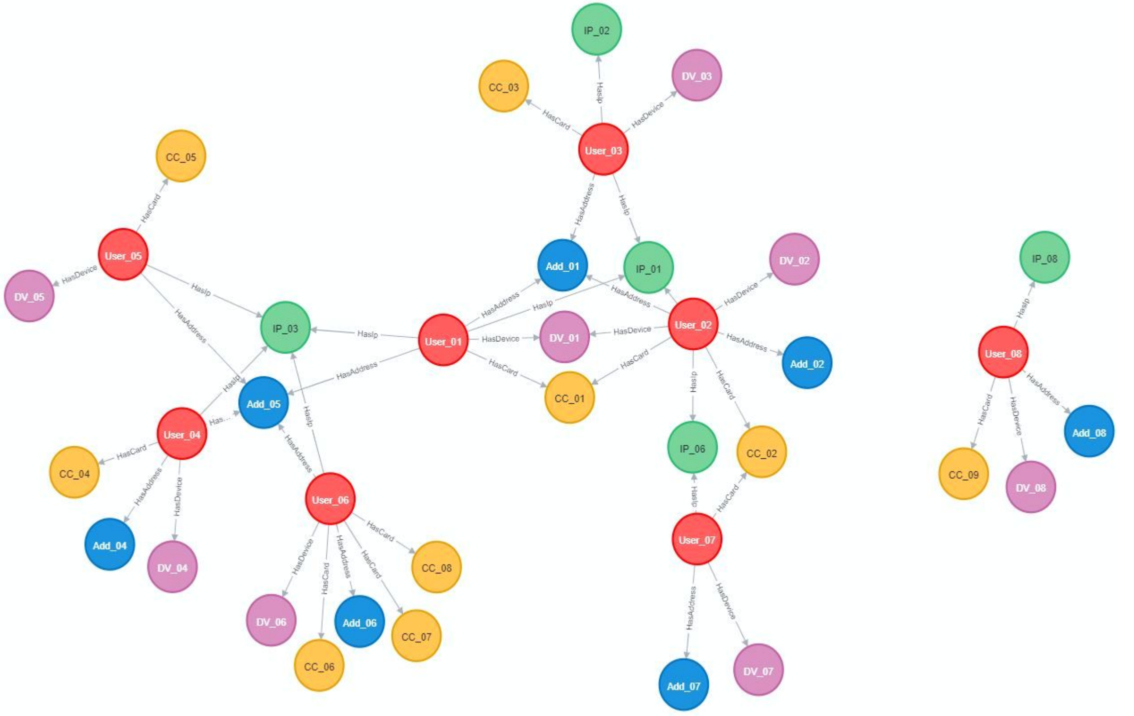


Possible Relations:

- IP Addresses
- P2P transfers
- Cell Phone Contacts
- Addresses
- Users Referrals
- Mobile Devices
- Credit cards
- Stores

Real world use cases - Community Detection

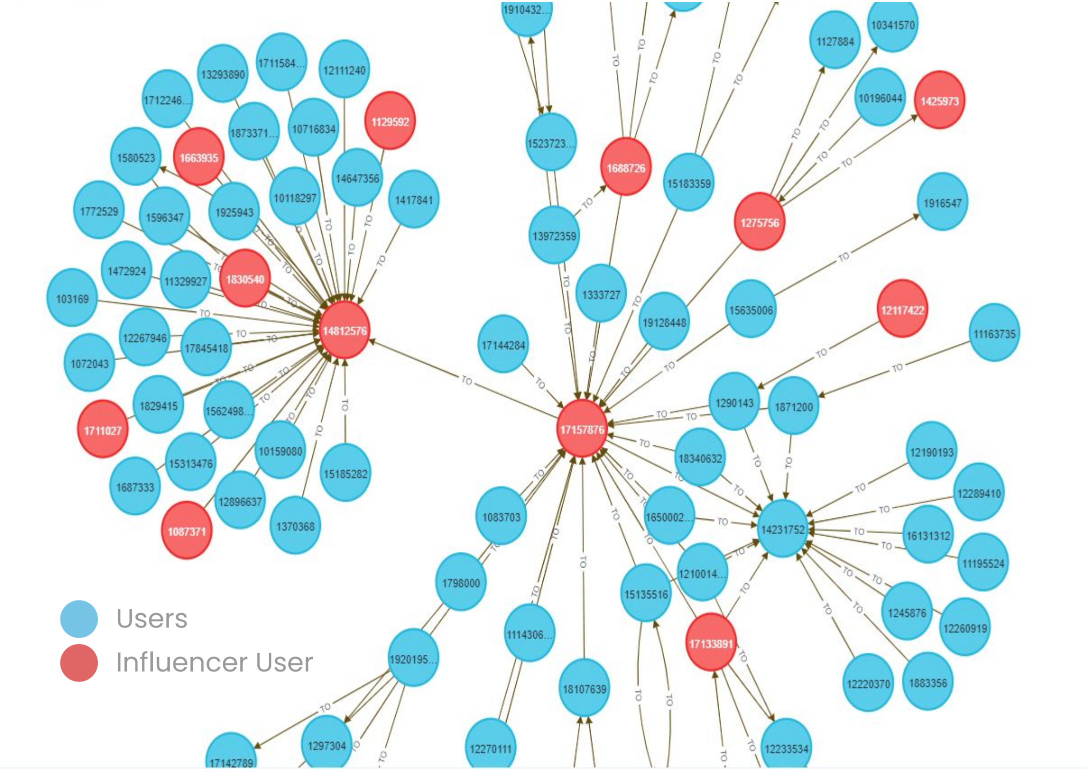
Identify families, friends or user groups that share experiences in Rappi. This allows for segmentation and targeted marketing.



Real world use cases - Influencer Detection

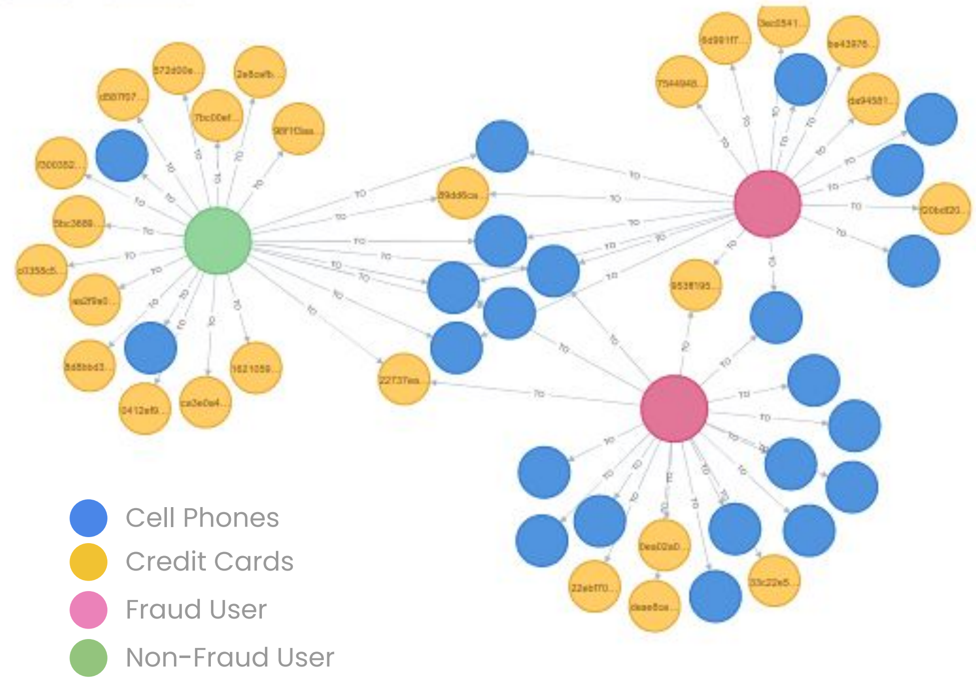
–

Identify users who are influencers in Rappi and RappiPay, that is, users who bring in new users and promote the use of the application.



Real world use cases - Fraud prevention

Risk score with graph-based features to get **insights of fraudulent users behavior** depending on the people/objects the user is connected to.



Real world use cases - Fraud prevention

We created a framework for the comparison of of the effectiveness of RGCN different graph structures:

- RGCN 1: Graph with users, credit cards and devices where no node has features.
- RGCN2: Graph with users, neighbors, credit cards and devices. Only users have features.
- RGCN3: Graph with users, neighbors, credit cards and devices. Users have features and neighbors have another set of features.
- RGCN4: Graph with users, neighbors, credit cards and devices. Users have features and neighbors have a partial set of features and also part of a randomly initialized embedding.

Table 3. AUC mean and standard deviation

Dataset	RGCN1	RGCN2	RGCN3	RGCN4
Dataset 1	0,6261±0,0071	0.7866±0.0154	0.7682±0.0152	0.7697±0.0109
Dataset 2	0,5952±0,0059	0.7284±0.0202	0.7392±0.0168	0.7213±0.0068
Dataset 3	0,5554±0,0064	0.7137±0.0189	0.7224±0.0180	0.7178±0.0254

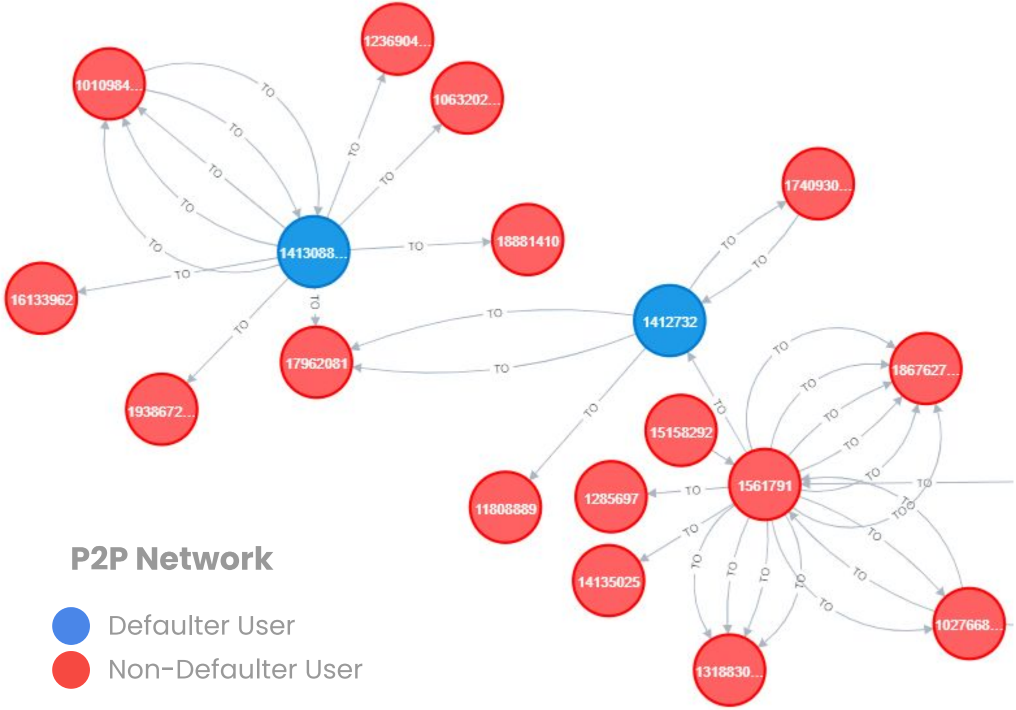
AUC Comparison of models on different datasets

Real world use cases - Credit risk assessment

–

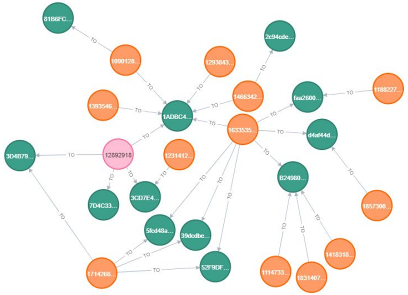
Enhance credit risk prediction by analyzing 5 different networks within the App.

The **graph allows the evaluation of users with no historical financial information**.

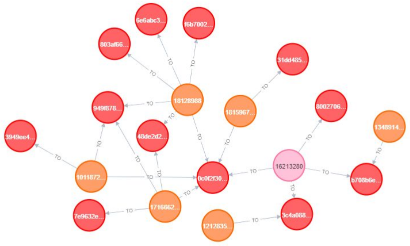


Real world use cases - Credit risk assessment

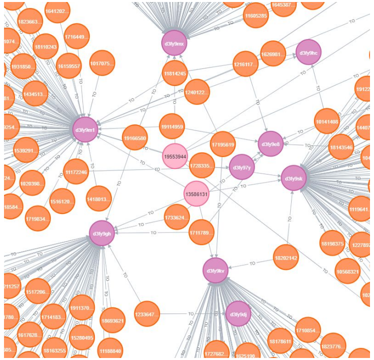
Device Network



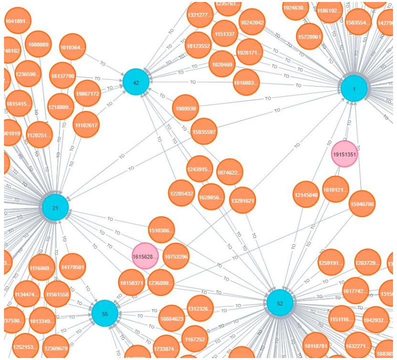
Credit Card Network



BIN Network



Geohash Network



Graph Querying:

- **Count** the number of **neighbors who are defaulter**
- **Average features** (orders paid with credit card, orders with payment error) of neighbors.

Graph-Based Features:

- PageRank
- Eigenvector centrality
- Louvain

Real world use cases - Credit risk assessment –

Variables derived from the BINs graph centrality seem to have a high effect in credit evaluation as well as centrality from credit card and geohashes and variables of the super-app related to the use of the credit card, use of discounts,and amount spent.