



GRAPH EMBEDDINGS

By Millis Sahar
May 2020

[LinkedIn](#) [Medium](#) [GitHub](#)

STORY TIME

When I was a kid...

01

THE GOAL

What are we trying to achieve

02

GRAPHS

How does it work

03

EMBEDDINGS

Technique, Word2Vec

04

05

DEEPWALKS

Novel Approach

06

GRAPH EMBEDDINGS

Embed an entire graph

07

TEST YOURSELF

Q&A

08

CONCLUSIONS

What now...

01

Story Time

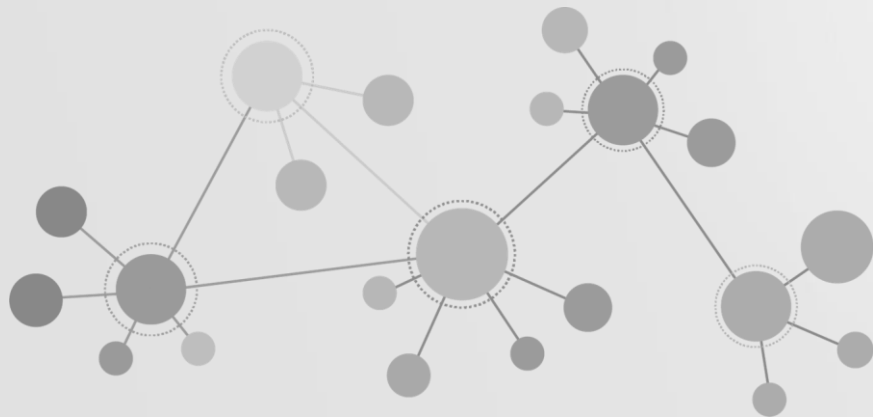
AI, ML, DL & GRAPHS





IT'S NOT WHAT YOU KNOW... IT'S WHO YOU KNOW!

Tabular | Text | Images | Audio | Graphs





02

The Goal

APPLICATIONS

APPLICATIONS



**FINANCIAL
SERVICES**



**DRUG
DISCOVERY**



**COSTUMER
SEGMENTATIONS**



CYBER



SEARCH

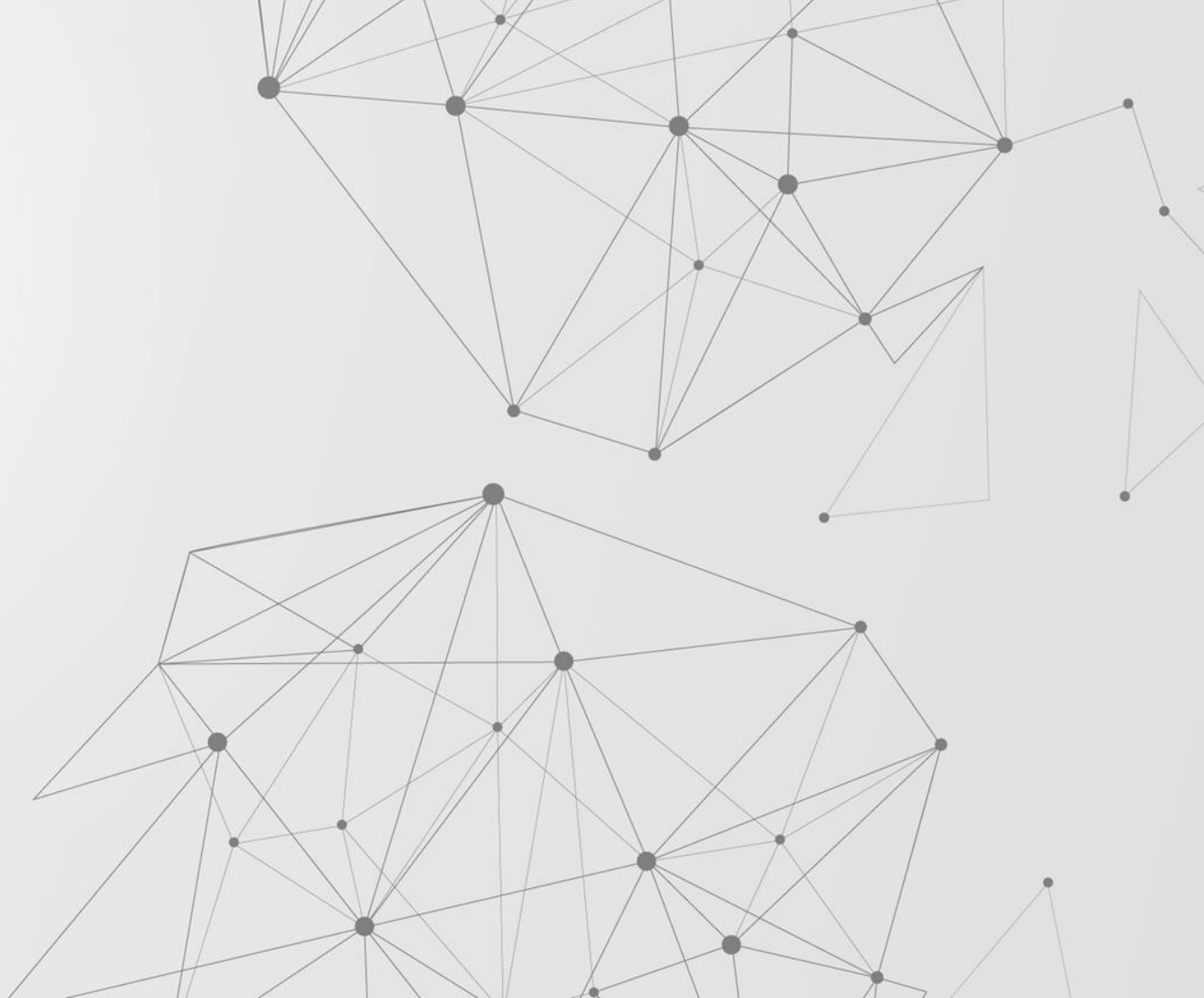


CHURN

03

GRAPHS

DEEP LEARNING EVOLUTION
DEEP GRAPHS



DEEP LEARNING EVOLUTION



CNNs

Images, Video and more



RNNs

Text, Signals, and more



GNNs

Interactions, Links, Patterns,
Pathfinding, Centrality, Similarity,
Community Detection,
and more...

Scale, Parallel
Visualize, Interpretability

FRUAD DETECTION EXAMPLE

Financial institutions have an existing ML pipeline for identifying fraud
and graph-based features improve accuracy

CONNECTED COMPONENTS

Identify disjoint graphs sharing identifiers

PAGERANK

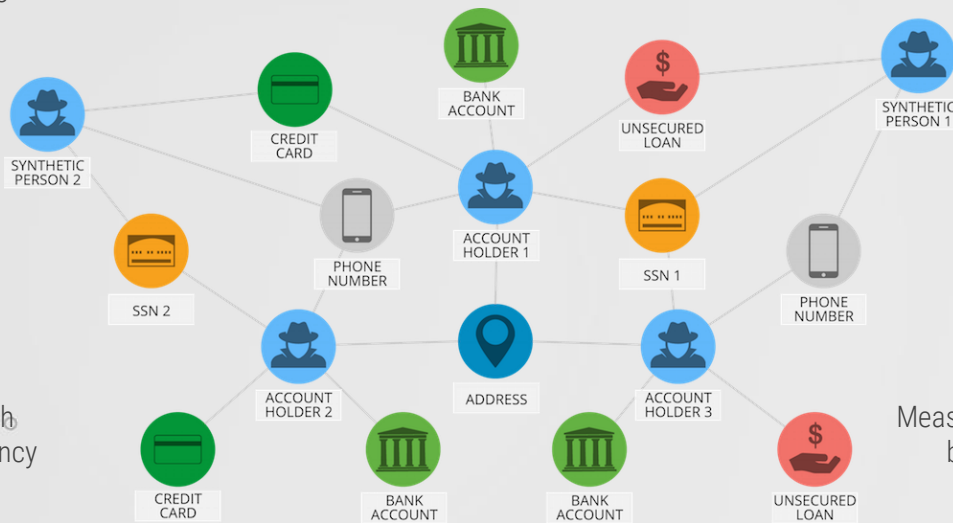
Measure influence and transactions volumes

LOUVAIN

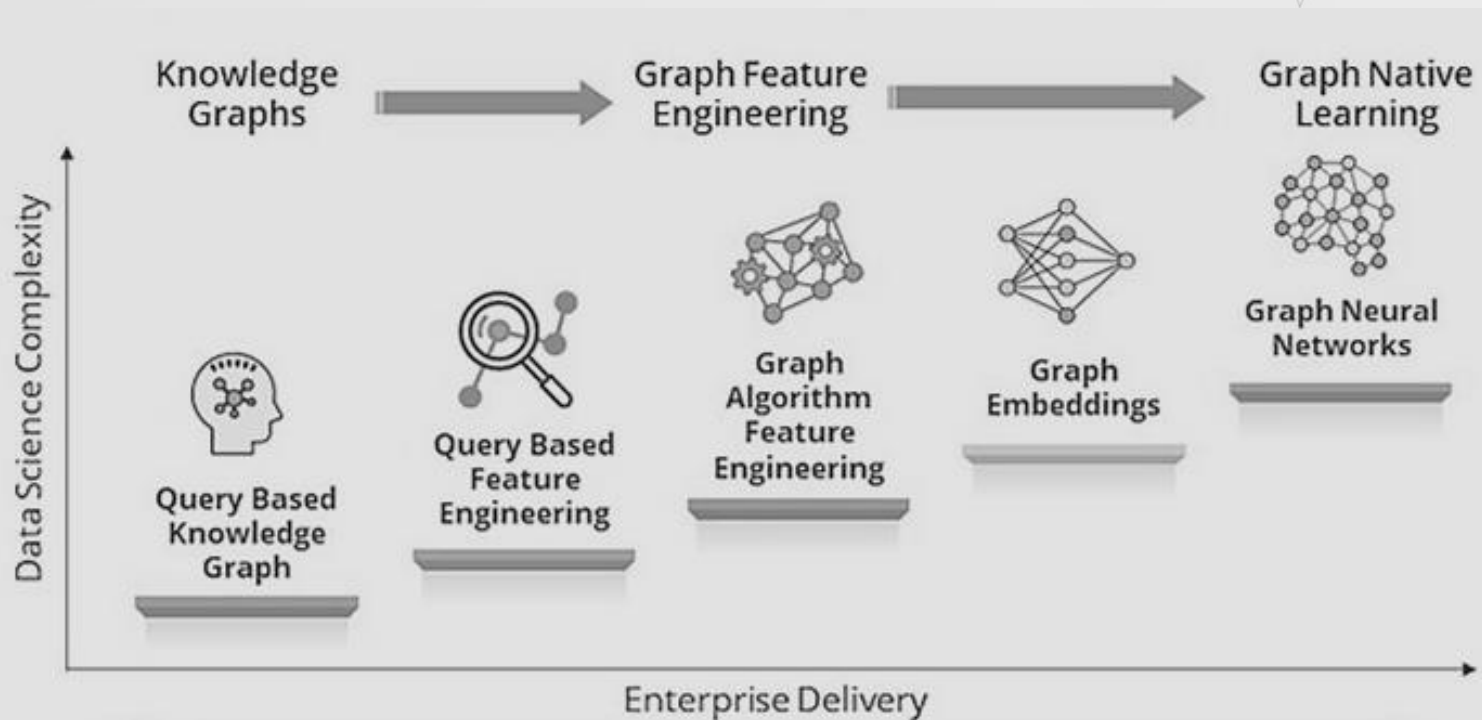
Identify communities with high interactional frequency

JACCARD

Measure account similarity based on relationships



GRAPH DATA SCIENCE





04

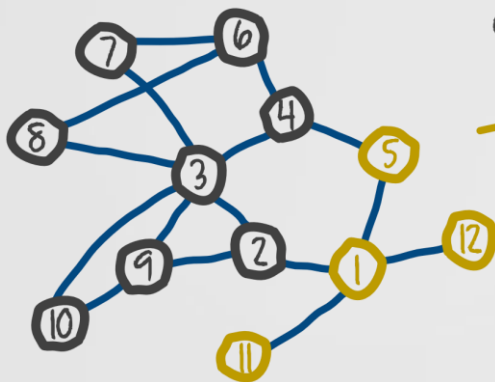
EMBEDDINGS

Feature Learning Technique

WHAT IS EMBEDDINGS?

A way of mapping something into a fixed length vector that captures key features while reducing the dimensionality

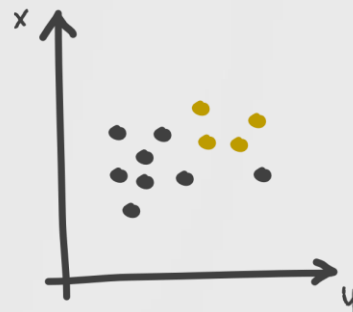
from a graph representation ...



embedding
algorithm

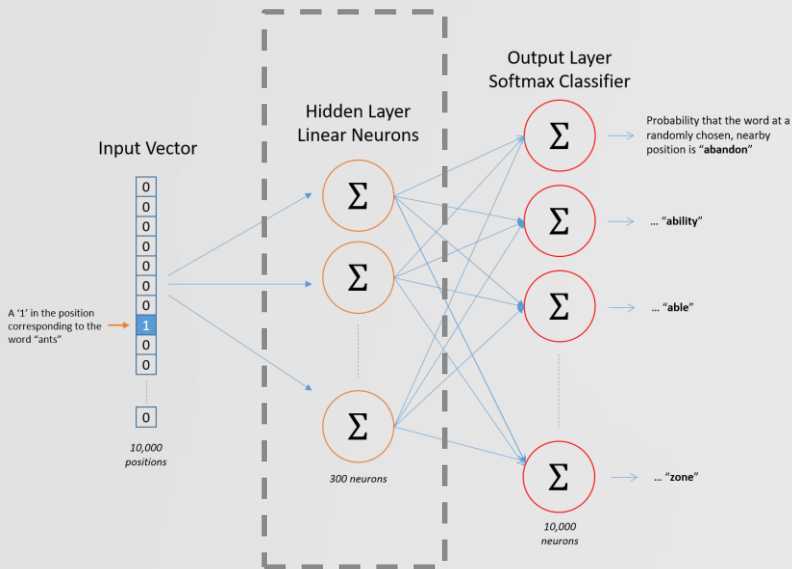


to real vector representation



WORD EMBEDDING MOTIVATION

How to represent words in a mathematical way?
How similar two words are?
Similar meaning?



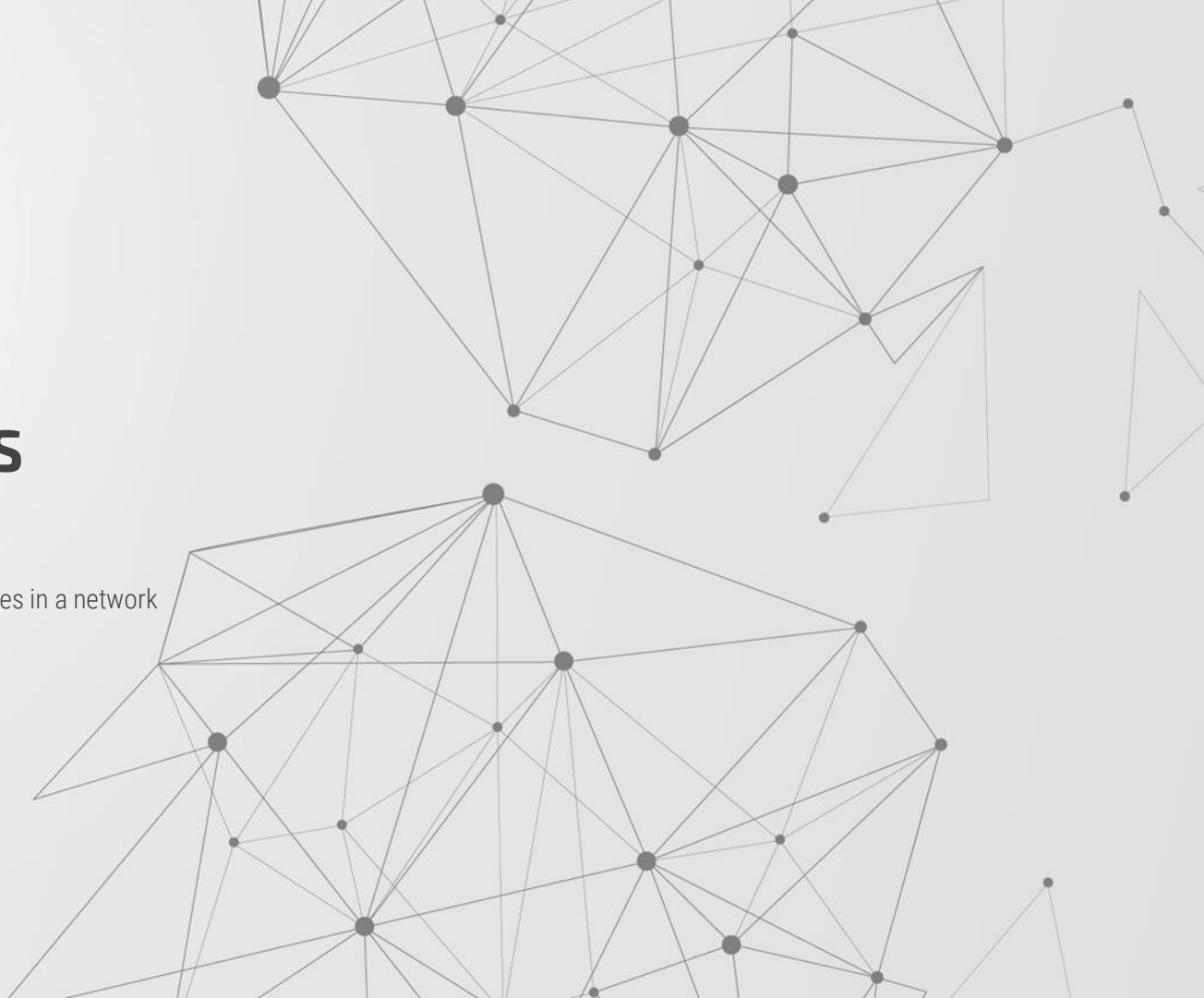
Hand-drawn table showing word embeddings for four words: King, Queen, Woman, and Princess. The embeddings are represented as vertical columns of values for various features: Royalty, Masculinity, Femininity, Age, and ... (vertical ellipsis).

	King	Queen	Woman	Princess
Royalty	0.99	0.99	0.02	0.98
Masculinity	0.99	0.05	0.01	0.02
Femininity	0.05	0.93	0.999	0.94
Age	0.7	0.6	0.5	0.1
...

05

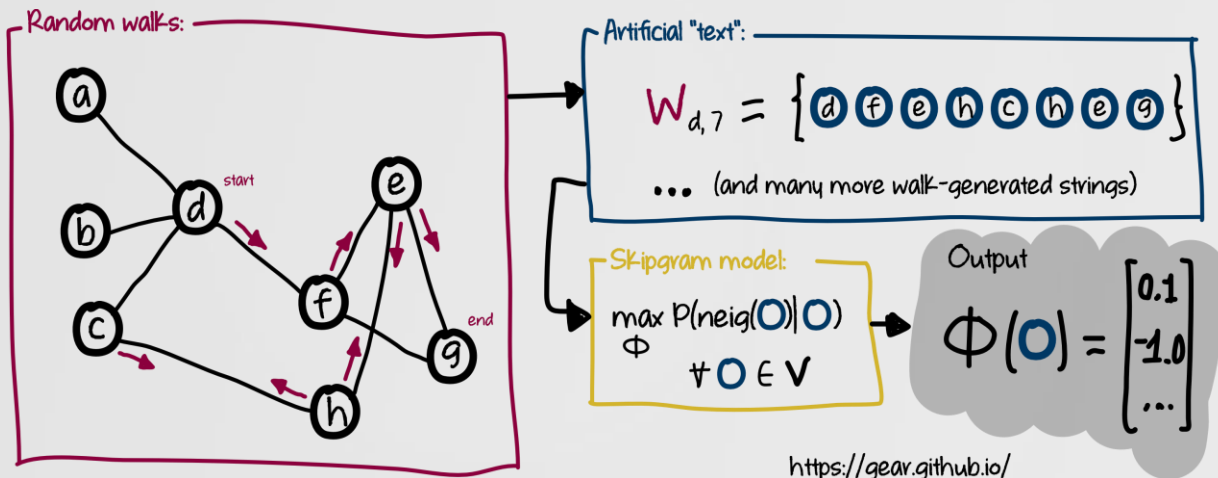
DEEPWALKS

A novel approach for learning
latent representations of vertices in a network



FROM NODES TO EMBEDDINGS

Run short fixed-length random walks starting from each node
Optimize embedding using SGD



<https://gear.github.io/>

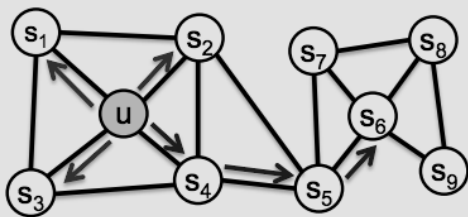
BIAS 2-ORDER RANDOM WALKS

ATTRIBUTES WEIGHTS AND MORE

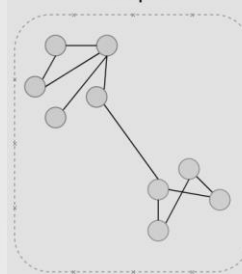
Node2Vec

Random walks Algorithm to generate vector representations of nodes on a graph, and learns low-dimensional representations for nodes

BFS DFS

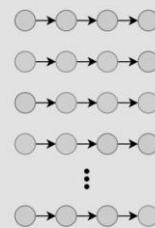


Graph



sampling
strategy

Input data






Q: WHY NOT STOP HERE?

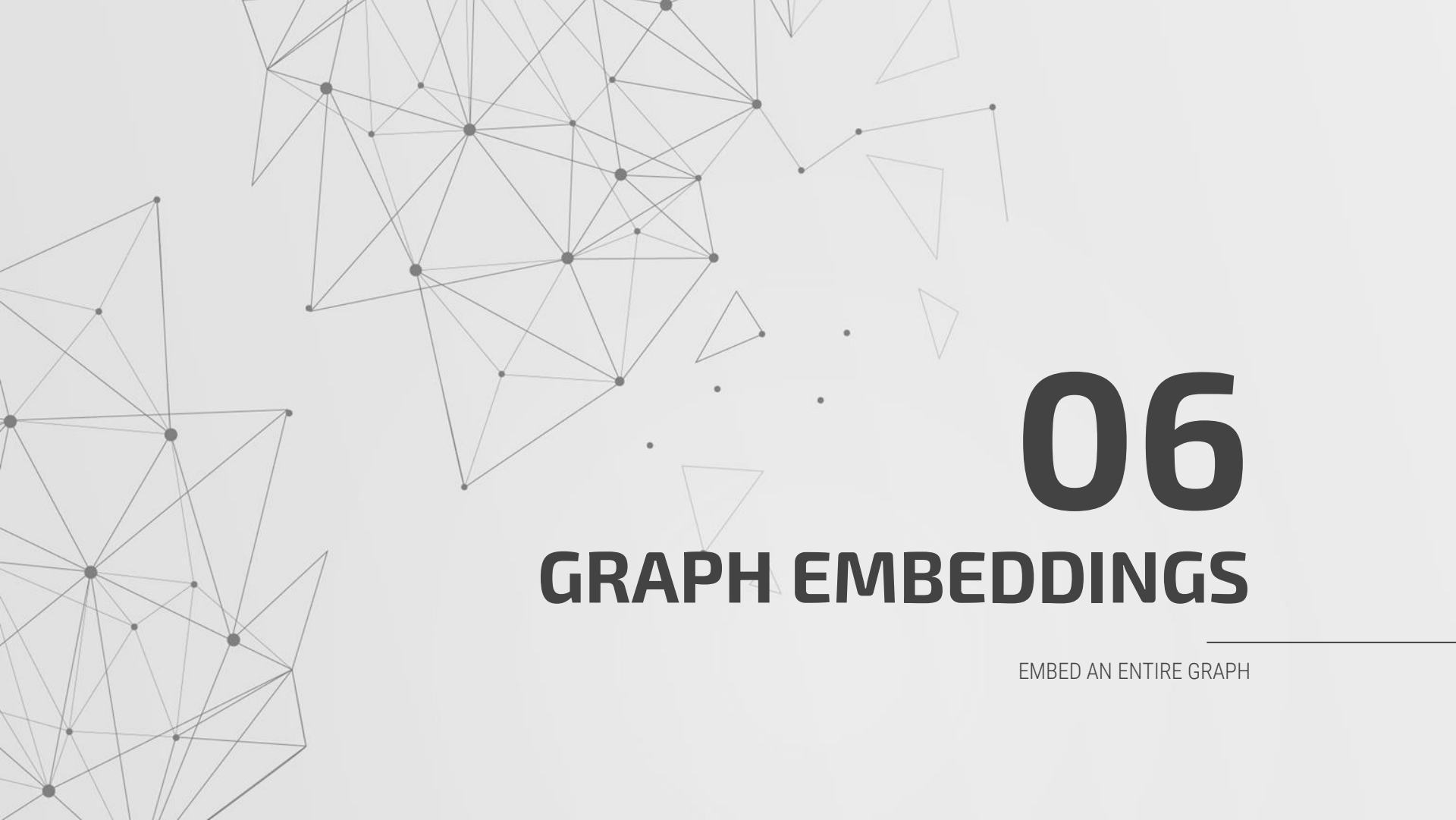
MATRIX FACTOTIZATION

Massive memory footprint
Computationally intense

RANDOM WALKS

Local-Perspective only
Assume similarity by closeness



The background features a complex network of thin, light gray lines and dots, resembling a graph or a molecular structure. These elements are scattered across the slide, with some forming dense clusters and others appearing as isolated points or small triangles. The overall aesthetic is clean and technical.

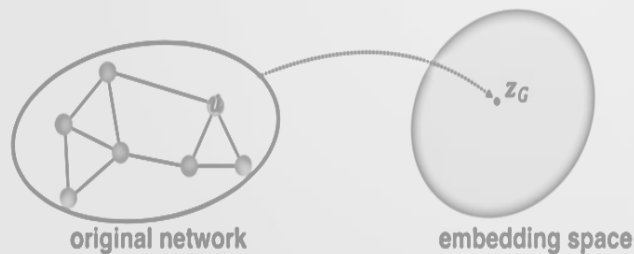
06

GRAPH EMBEDDINGS

EMBED AN ENTIRE GRAPH

WHY TO EMBED AN ENTIRE GRAPH?

Toxic vs non-toxic molecules
Identify anomalies
And more

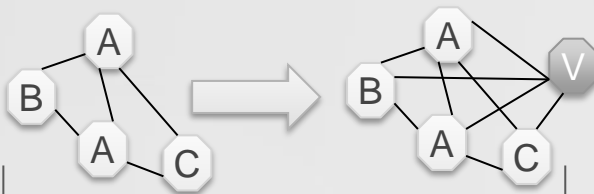


TECHNIQUES



EMBEDDINGS SUMATION

$$Z_G = \sum_{v \in G} Z_v$$



Connect a new node
to all the nodes

VIRTUAL NODE



ANONYMOUS WALKS

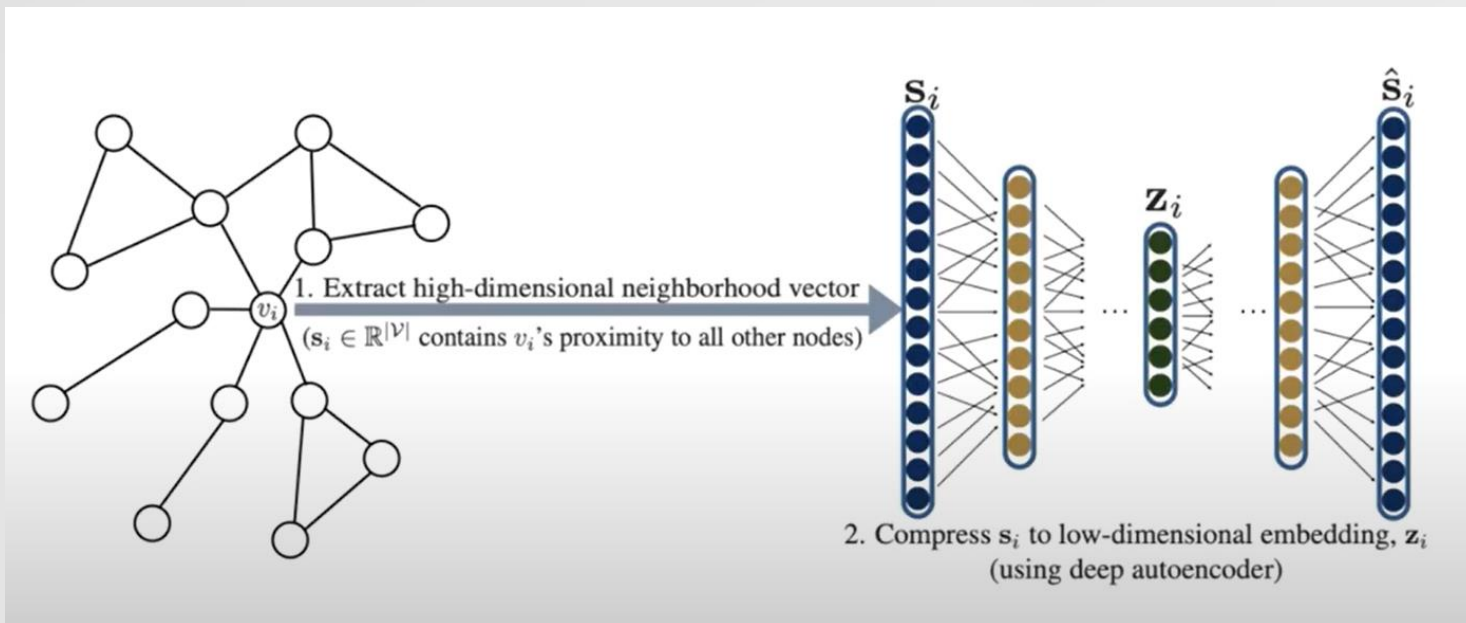
Index the nodes in each walk

Represent by
a sample distribution or
all the distribution



SOTA?

Using Autoencoders



The background features a complex network of thin grey lines connecting various-sized dark grey circular nodes. These nodes are scattered across the page, with a higher concentration on the right side, creating a web-like or molecular structure. The overall aesthetic is minimalist and technical.

07

TEST YOURSELF

Q & A

Q: Which of the following is not a step in graph data science?

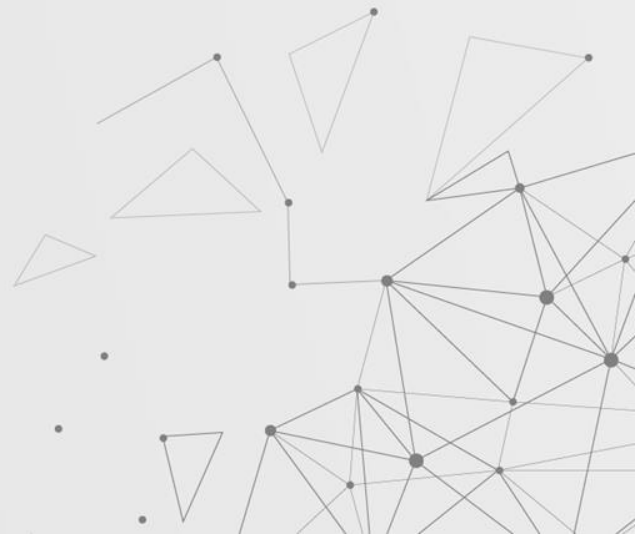
- Building a knowledge graph
- Using graph algorithm for feature engineering
- Using Kafka for transactional messaging

Q: Louvain is an example of...?

- Centrality
- Pathfinding
- Community Detection

Q: In which task Random Walks are “weaker”?

- Link Prediction
- Node Classification



Q: A graph embeddings is a fixed length vector of...?

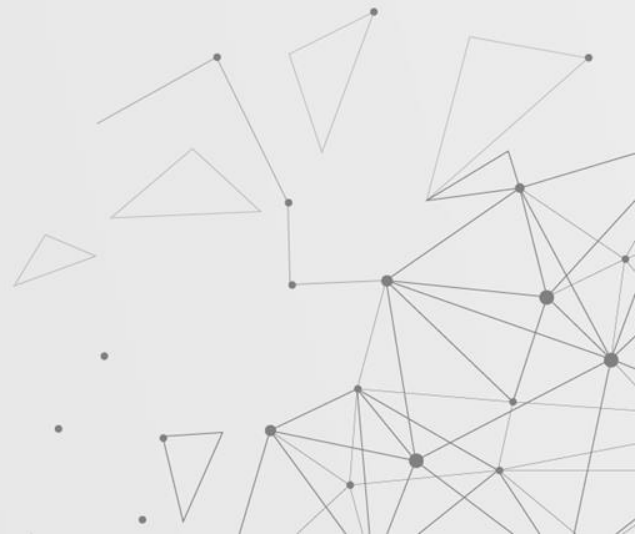
- Numbers
- Letters
- Nodes

Q: An embedding is a ____ representation of your data.

- Binary
- Human Readable
- Lower Dimensional

Q: What is a way to embed an entire graph?

- GNNs
- BFS/DFS
- Virtual Node





08

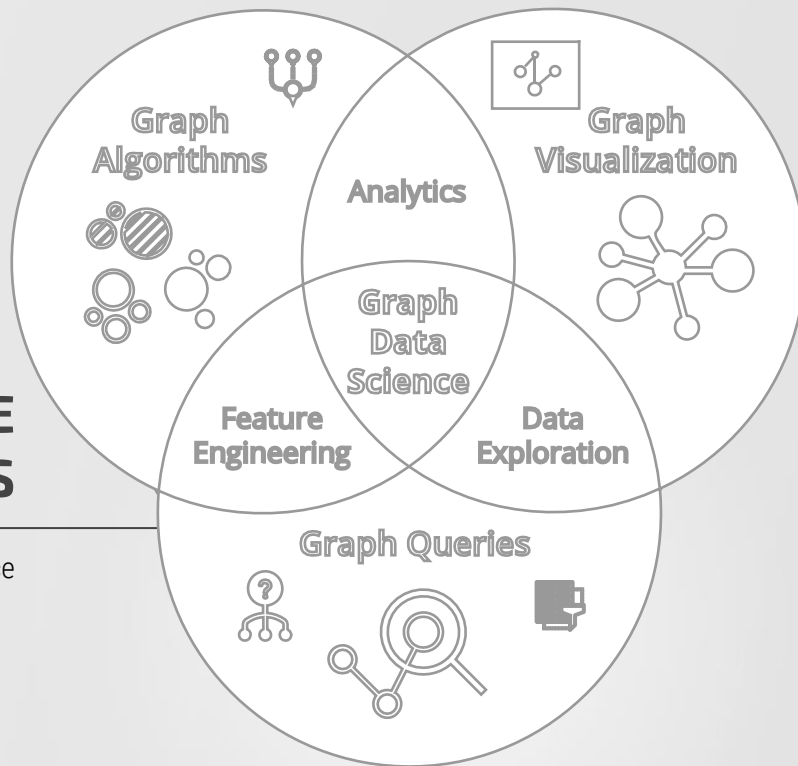
CONCLUSIONS

REFERENCES & RELATED WORK



DATA SCIENCE WITH GRAPHS

There's an unique balance



RESOURCES

Relational inductive biases, deep learning, and graph networks

By DeepMind

<https://arxiv.org/pdf/1806.01261>

DeepWalk: Online Learning of Social Representations

By Bryan Perozzi, Rami Al-Rfou, Steven Skiena

<https://arxiv.org/pdf/1403.6652>

Anonymous Walk Embeddings

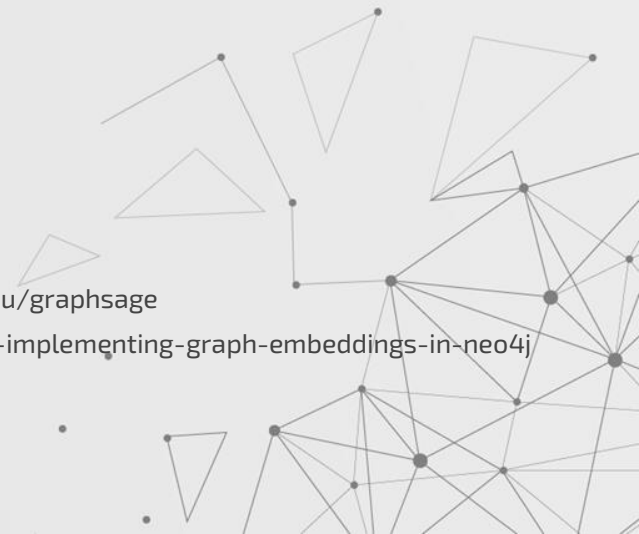
By Sergey Ivanov, Evgeny Burnaev

<https://arxiv.org/pdf/1805.11921>

Graph Nets library | https://github.com/deepmind/graph_nets

GraphSAGE: Inductive Representation Learning on Large Graphs | <http://snap.stanford.edu/graphsage>

DeepWalk: Implementing Graph Embeddings in Neo4j | <https://neo4j.com/blog/deepwalk-implementing-graph-embeddings-in-neo4j>





THANKS

Does anyone have any questions?

SAHAR.MILIS@GMAIL.COM

<https://www.linkedin.com/in/sahar-millis>

<https://medium.com/@sahar.millis>

<https://github.com/saharmilis>

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**.

Please keep this slide for attribution.