DATA+AI SUMMIT 2022

Powering Geospatial Data Science with Graph Machine Learning

Anirudh Shah
Research Engineer, Iggy

ORGANIZED BY databricks

Agenda

- Geospatial Data Overview
- Graph Representations of Geospatial Data
- Graph Machine Learning on Geospatial Data
- Use cases
- Final thoughts

What is geospatial data?

Maps?

What is Geospatial Data?

Object Characteristics + Location

Characteristics

- Some or all of:
  - What?
  - When?
  - Who?
  - How?

Location

- Where?
  - Coordinates
  - Address
  - Zipcode
  - City
  - Census Block Group

Geospatial Data Sources

- Satellite Imagery
- Location Data
- Government
- Census
- Open Street Map (OSM)

Example

Object Characteristics

Characteristics

- Some or all of:
  - **What?**
  - When?
  - Who?
  - How?

Coffee Shop

Example

Object Characteristics

Characteristics

- Some or all of:
  - What?
  - **When?**
  - Who?
  - How?

06/02/2022

Example

Object Characteristics

Characteristics

- Some or all of:
  - What?
  - When?
  - **Who?**
  - How?

N/A

Example

Object Characteristics

Characteristics

- Some or all of:
  - What?
  - When?
  - Who?
  - **How?**

N/A

Example

Location

New York

Location

- **Where?**
  - Coordinates
  - Address
  - Zipcode
  - **City**
  - Census Block Group

Example

Object Characteristics

Characteristics

- Some or all of:
  - **What?**
  - When?
  - Who?
  - How?

Enters Stadium

Example

Object Characteristics

Characteristics

- Some or all of:
  - What?
  - **When?**
  - Who?
  - How?

03/22/2022

Example

Object Characteristics

Characteristics

- Some or all of:
  - What?
  - When?
  - **Who?**
  - How?

Spike Lee

Example

Object Characteristics

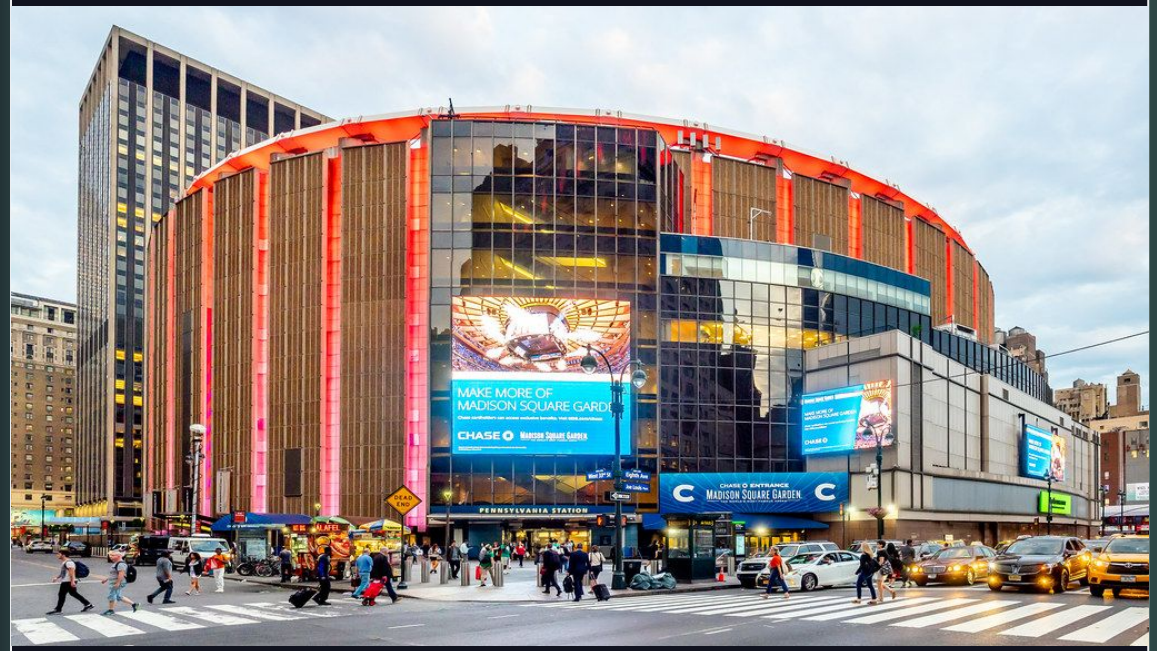
Characteristics

- Some or all of:
  - What?
  - When?
  - Who?
  - **How?**

On Foot

Example

Location

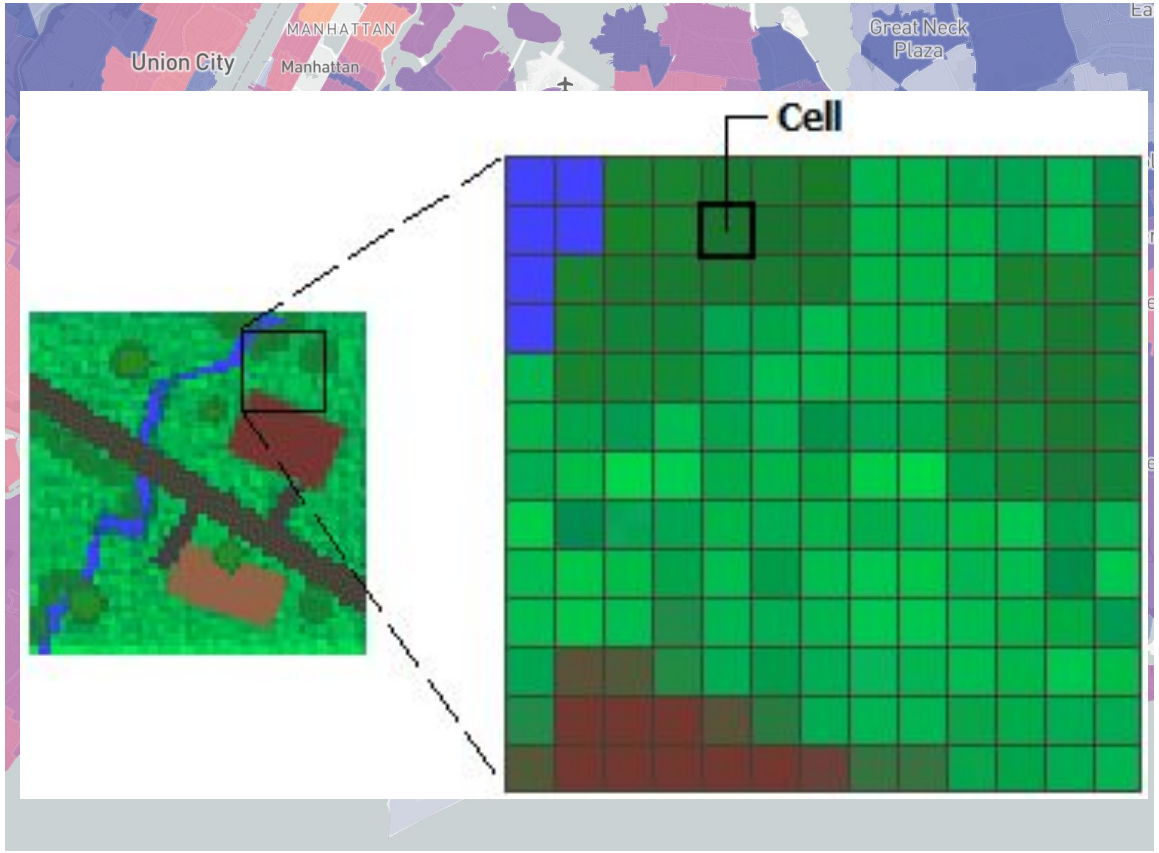


Location

- **Where?**
  - Coordinates
  - **Address**
  - Zipcode
  - City
  - Census Block Group

Geospatial Data Formats

- Raster[1] – Continuous
- Vector[2] – Discrete



[1] *ArcMap*. What is raster data?-ArcMap | Documentation. (n.d.). Retrieved June 2, 2022, from https://desktop.arcgis.com/en/arcmap/latest/manage-data/raster-and-images/what-is-raster-data.htm
[2] www.askiggy.com

Geospatial Data at Iggy

What is nearby?

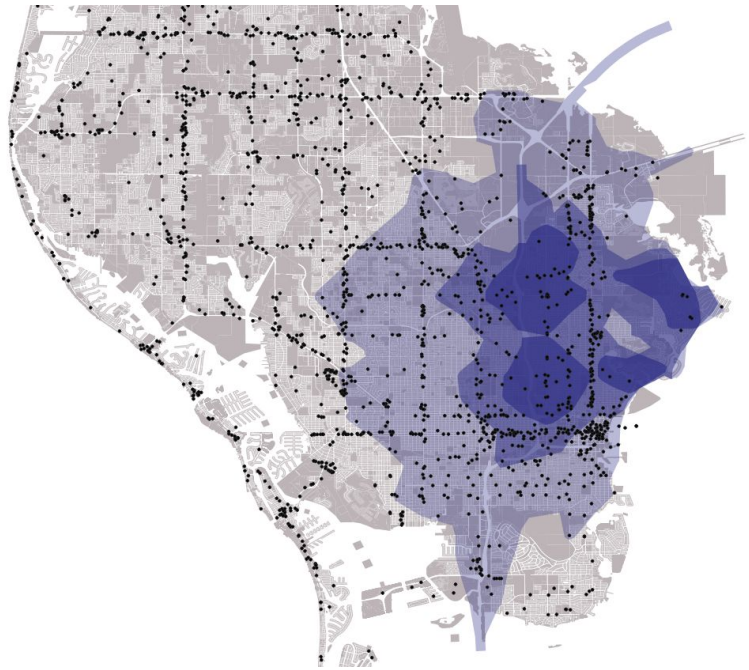
- Help you understand ‘place’
- Characteristics of a neighborhood
- ‘Nearest’
- Aggregate to geographic boundaries

- Types of data:
  - Points of Interest (POIs)
    - Bookstore
    - Restaurant
    - Coffee Shop
  - Water bodies
  - Socio-economic data
  - And much more!

Geospatial Data at Iggy

Boundaries

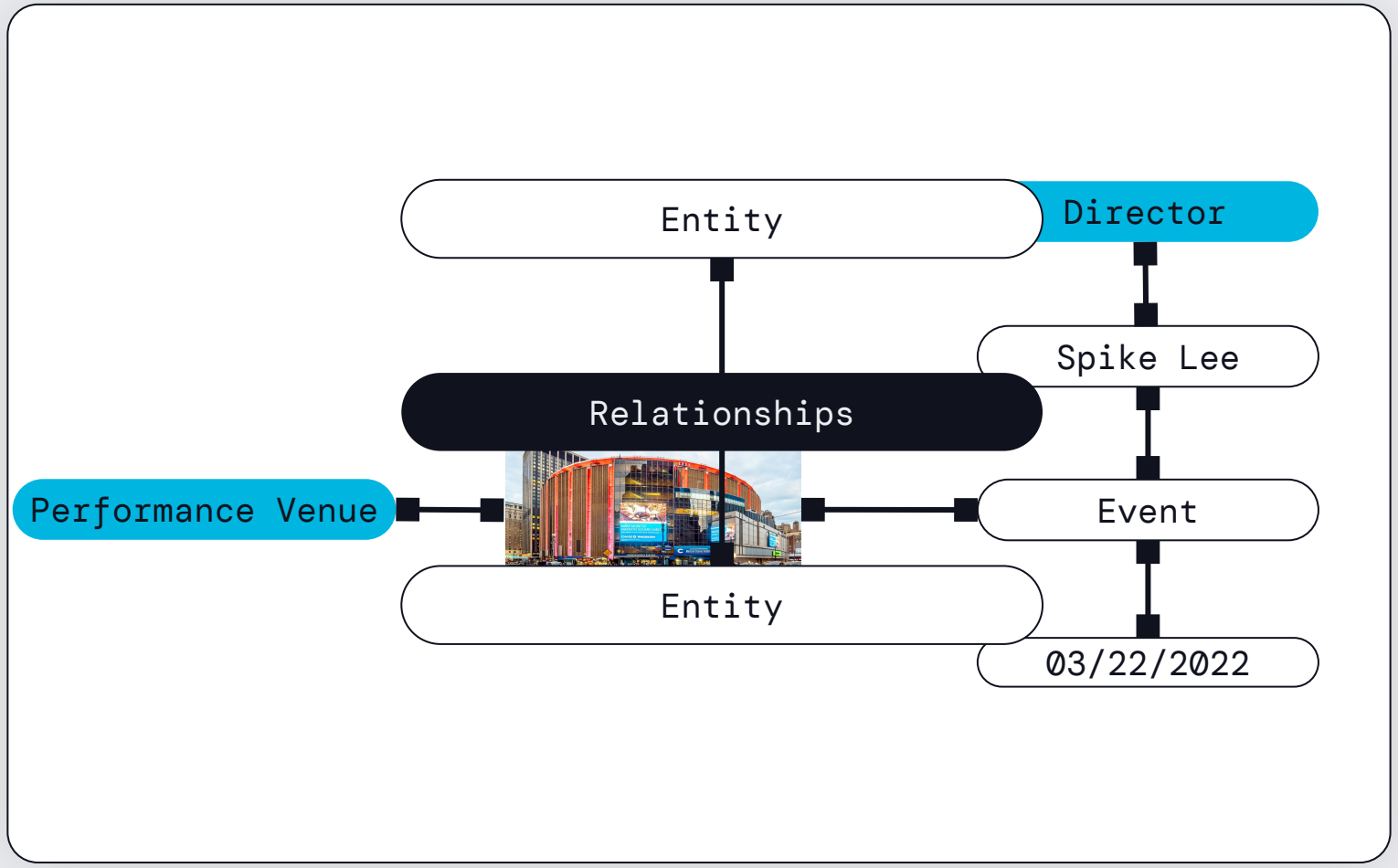
- Geographic area over which data is aggregated
- Administrative boundaries:
  - Zipcode
  - Census Block Group (CBG)
- Derived boundaries:
  - Isochrone 10 minute walk



Graph Representations

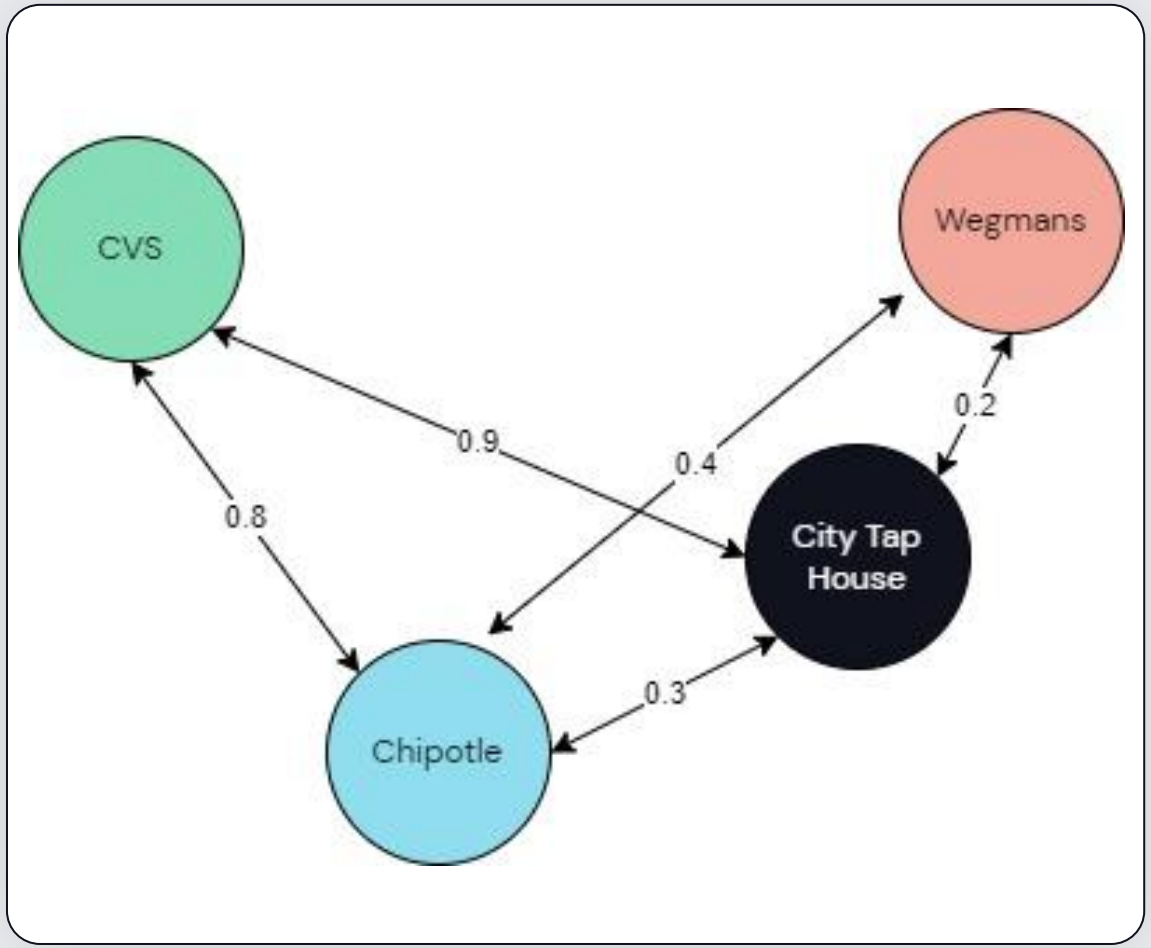
Power of Graphs

- Entities + Relationships
- Flexible – multimodal data
- Hierarchical



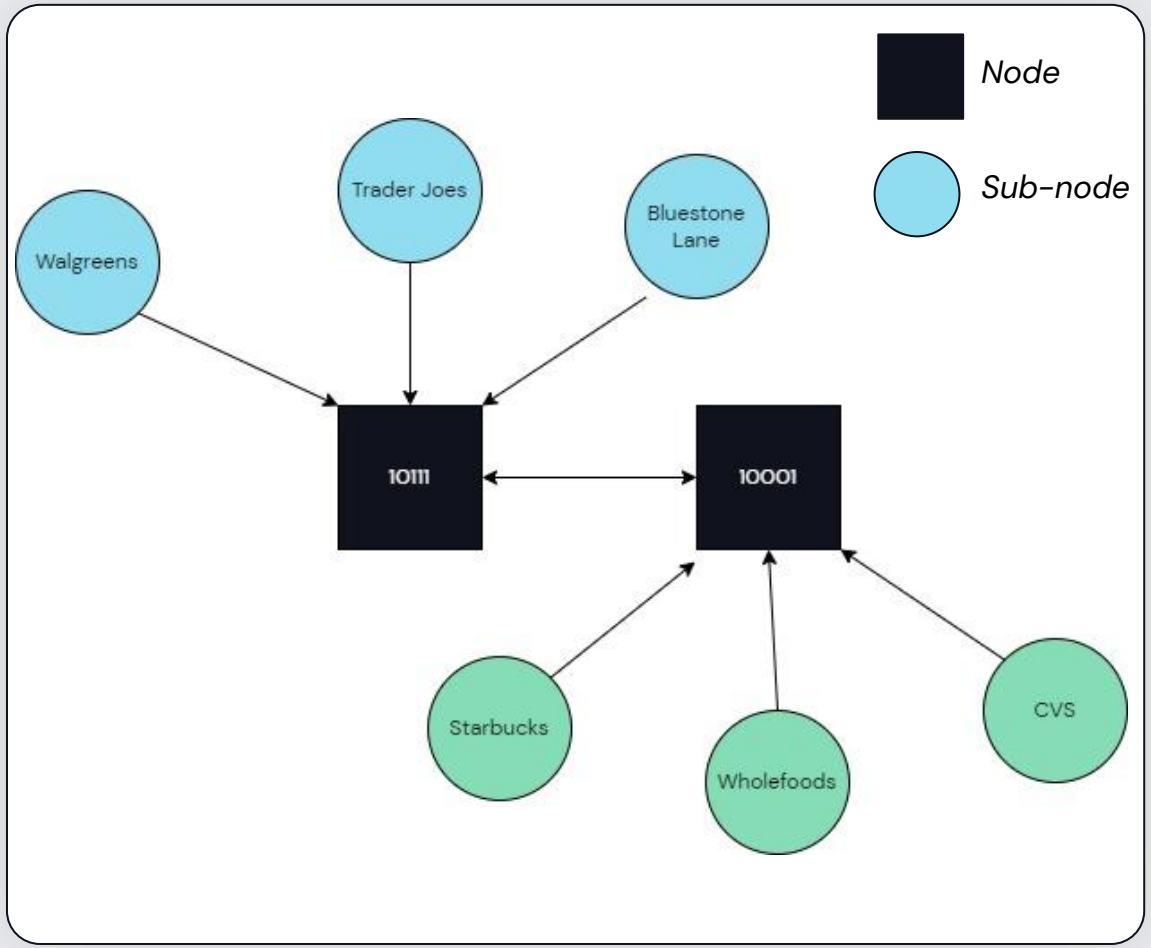
Representation 1

- Individual places
- Node features:
  - Word embedding of place name
  - Type of place
- Edge features:
  - Distance between each pair of places
  - Within a 1km radius



Representation 2

- Individual geographic boundaries
- Node features:
  - Mean property price
- Sub nodes are places within those boundaries
- Sub Node features:
  - Word embedding of place name
  - Type of place



Graph Machine Learning on Geospatial Data

Goal = Generate embeddings for the graph using an inductive model

Two potential tasks

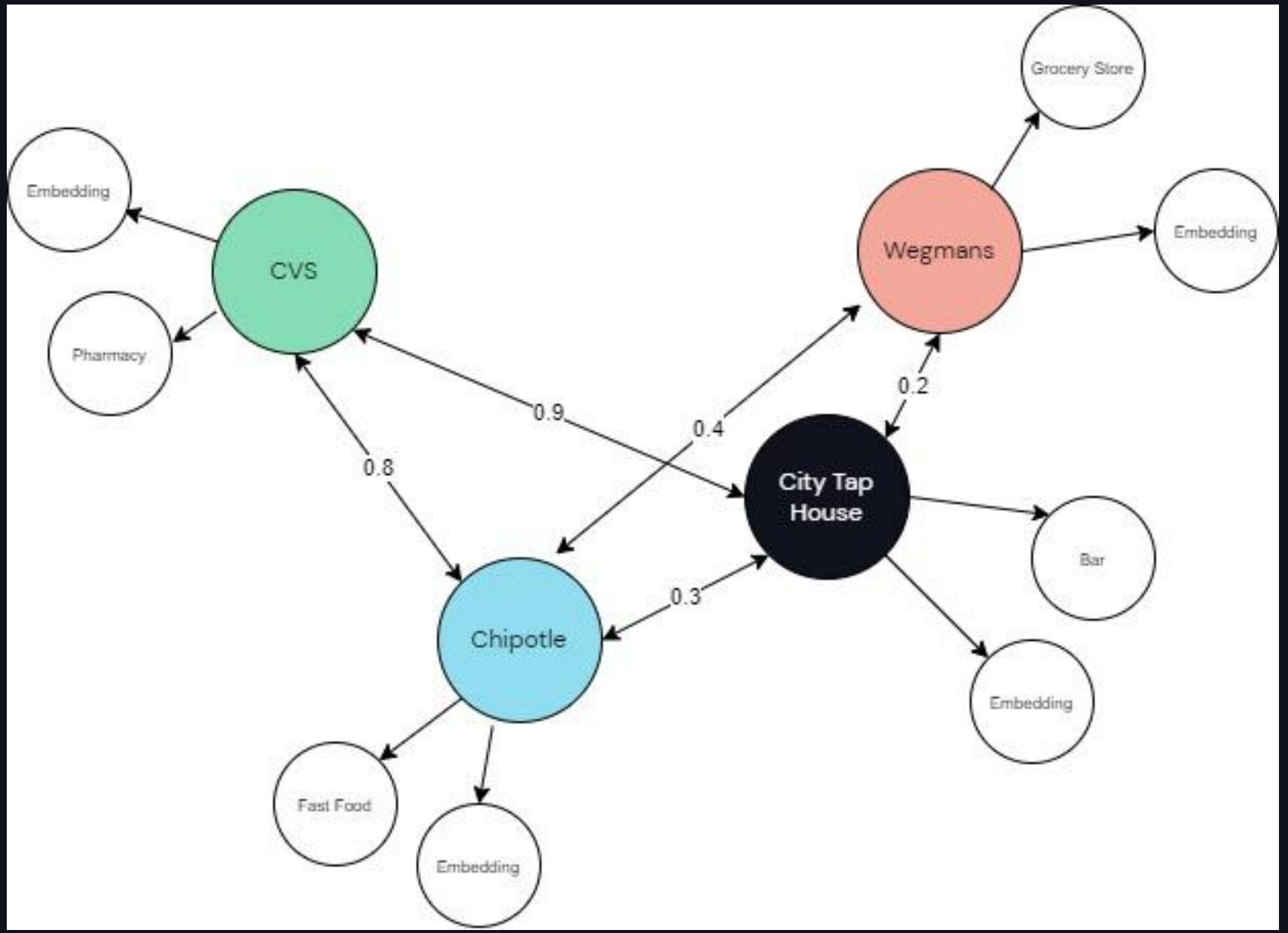
Task 1 with Representation 1

- Predict the type of place for nodes in graph
- 70 types

Easier place to start

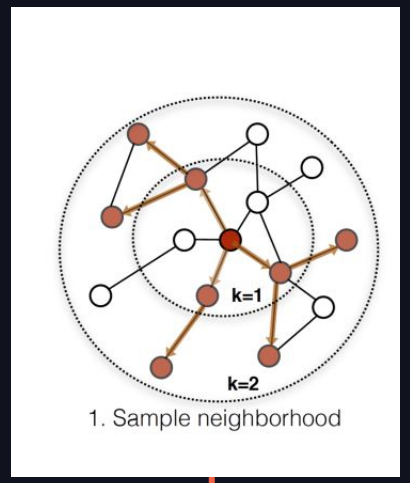
Task 2 with Representation 2

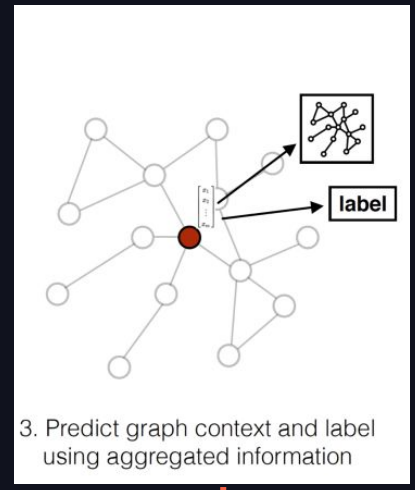
- Predict median house price for each zipcode
- In progress

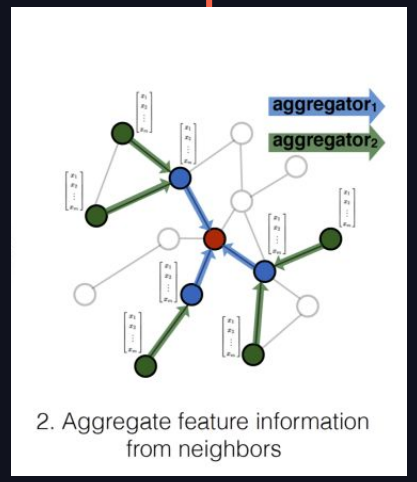

Grocery Store
Embedding
CVS
Wegmans
Embedding
Pharmacy
0.2
0.9
0.4
0.8
City Tap House
Bar
0.3
Chipotle
Embedding
Fast Food
Embedding

Model Architecture

GraphSAGE







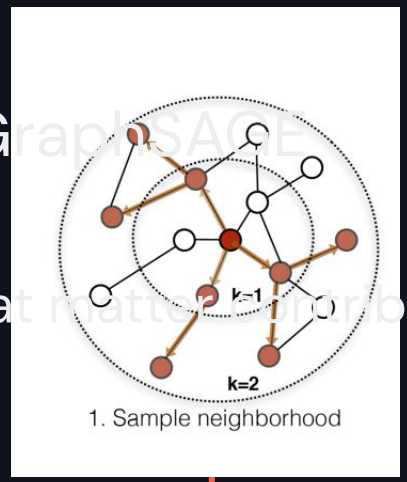
83% Accuracy in Location Type prediction task

Hamilton, William L., et al. “Inductive Representation Learning on Large Graphs.” ArXiv:1706.02216 [Cs, Stat], Sept. 2018. arXiv.org, http://arxiv.org/abs/1706.02216.

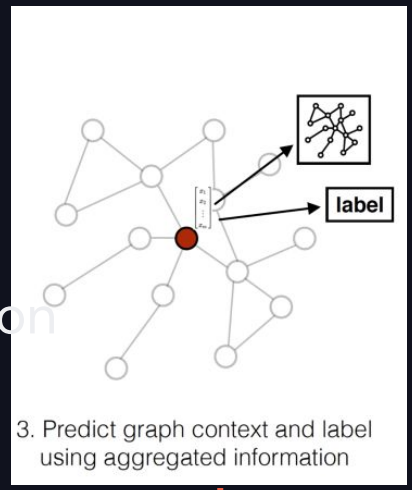
Model Architecture

Graph Attention Network

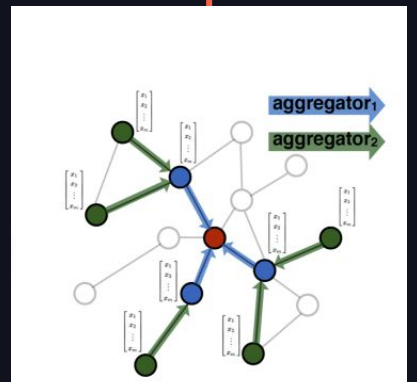
- Similar to GraphSAGE
- Attention
  - nodes that matter contribute more to aggregation



1. Sample neighborhood



3. Predict graph context and label using aggregated information



2. Aggregate feature information from neighbors

86% Accuracy in Location Type prediction task

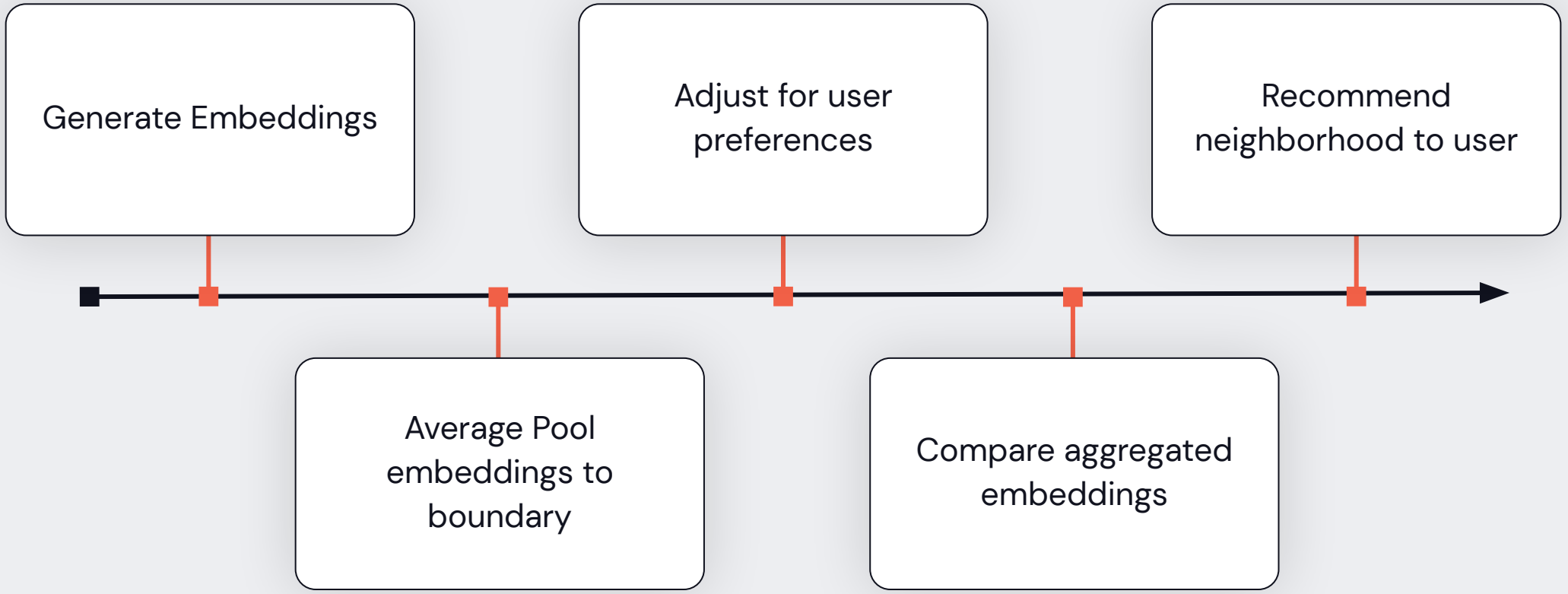
Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., & Bengio, Y. (2017). Graph attention networks. arXiv preprint arXiv:1710.10903.

Use Cases

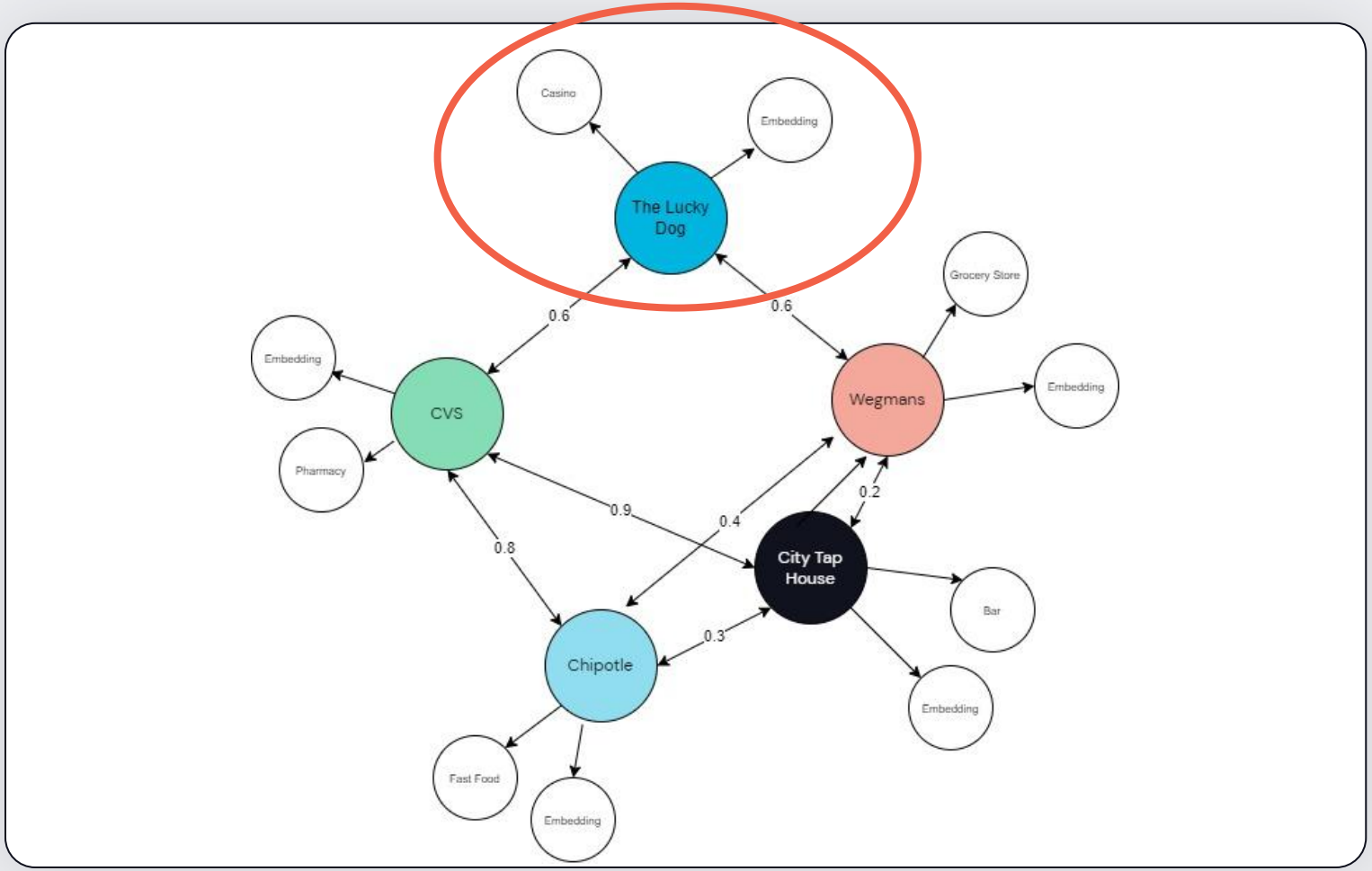
What is Iggy?

- Help our customers understand ‘place’
- Data and data-derived products
- We take on a lot of the data engineering and data science workload
- Unlocking geospatial insights

Neighborhood recommendation



Quality Control



- Validate labels in ingested dataset
- Add node to existing graph & predict its type
- Predicted type = ingested label?
- Flag for human review

Challenges

Data preparation and scale

Computational efficiency

Evaluation

Our Tech

Data

- BigQuery
- VertexAI
- Compute Engine

Machine Learning

DeepSNAP

MLOps

What's next?

- Account for spatial correlations
- Add more data features
  - Multimodal
- Use more SOTA modeling techniques
- Build query-able knowledge graph

DATA+AI
SUMMIT 2022

Thank you

Anirudh Shah
Research Engineer
iggy