Applications of Graph Theory in Data Science

Exploring Real-World Implementations and Insights

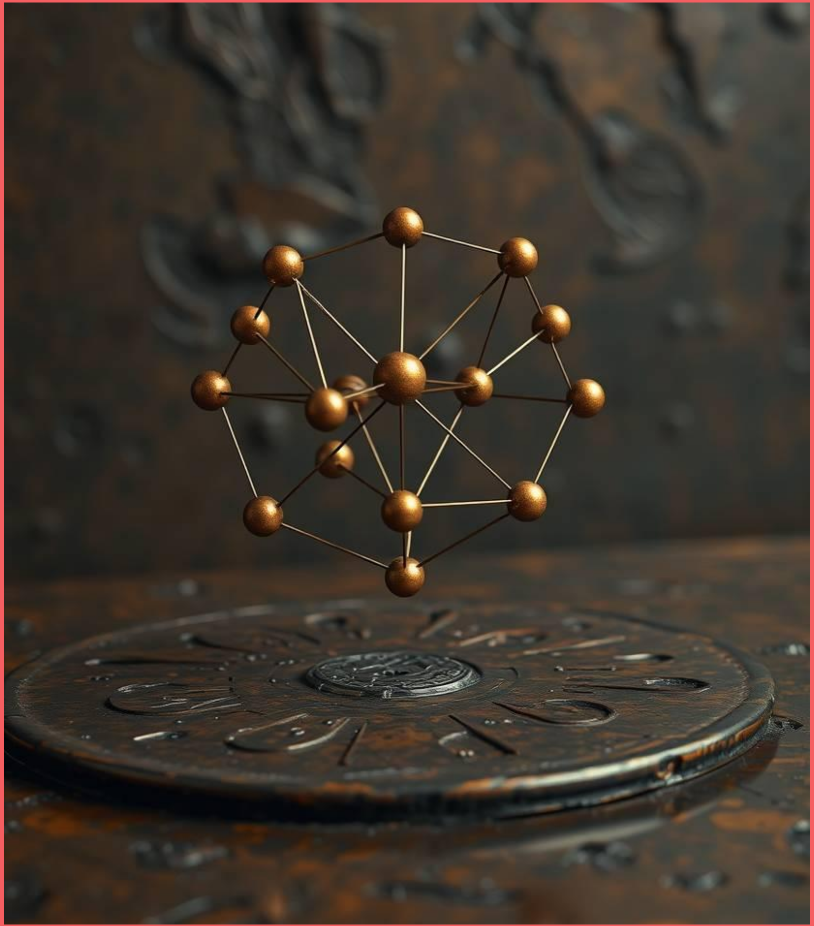
Presented by	Babji Manohar Erle
Date	23rd May, 2025

Introduction to Graph Theory

Definition: Graph theory studies structures made up of nodes (vertices) and connections (edges).

Relevance: Essential for modeling relationships in complex datasets.

Image Reference:
AI Generated



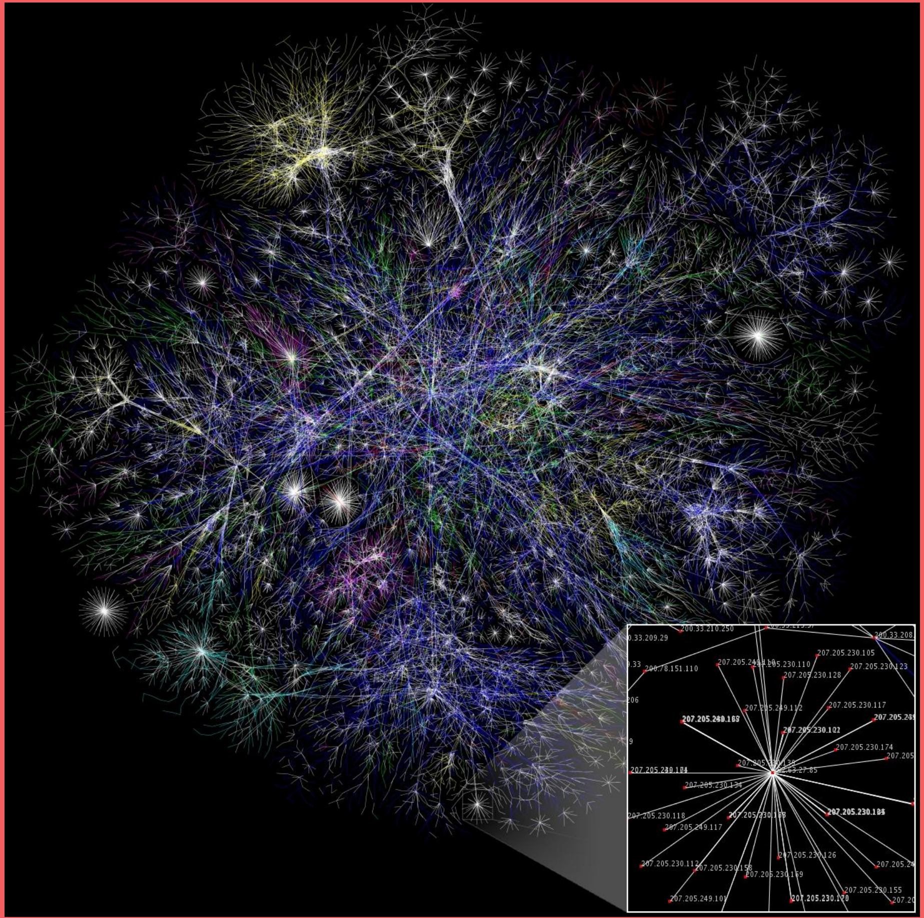
A 3D diagram with nodes and edges

Importance in Data Science

Data Representation: Graphs model relationships in data naturally.

Applications: Used in social networks, recommendation systems, fraud detection, etc.

Image Reference: https://en.wikipedia.org/wiki/Social_network_analysis



Partial map of the Internet

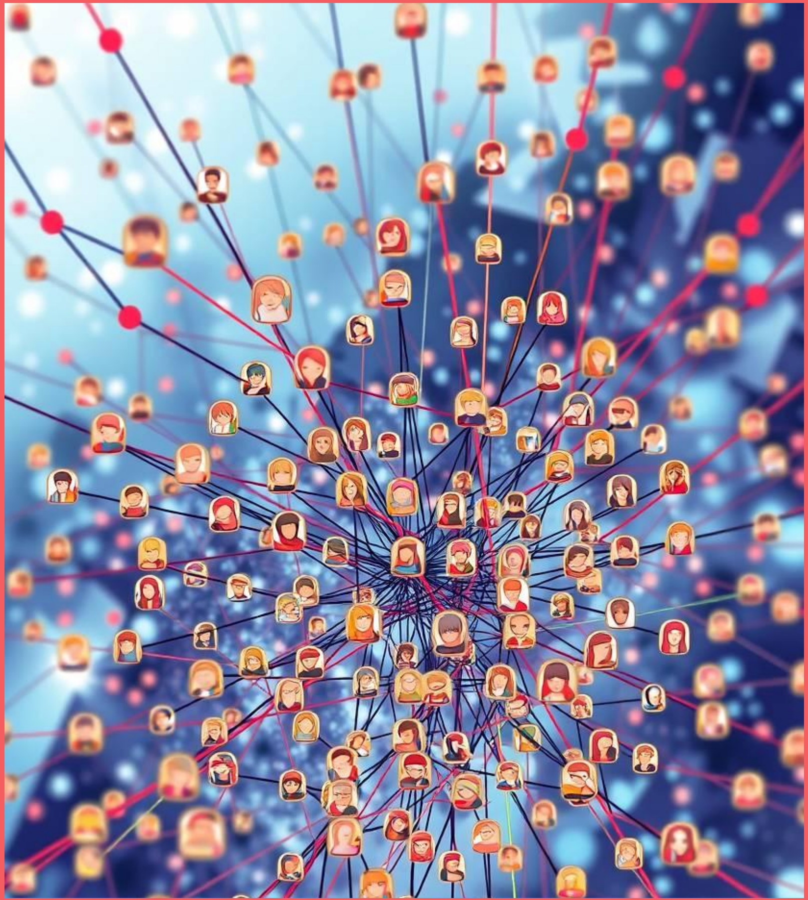
Social Network Analysis

Use Case: Analyzing connections between individuals or entities.

Example: Facebook's friend suggestions based on mutual connections.

Tools: NetworkX, Gephi.

Image Reference:
AI Generated



Graph showing interconnected users

Recommendation Systems

Use Case: Suggesting products or content to users.

Example: Netflix recommending movies based on viewing history.

Method: Collaborative filtering using user-item graphs.

Image Reference: AI generated

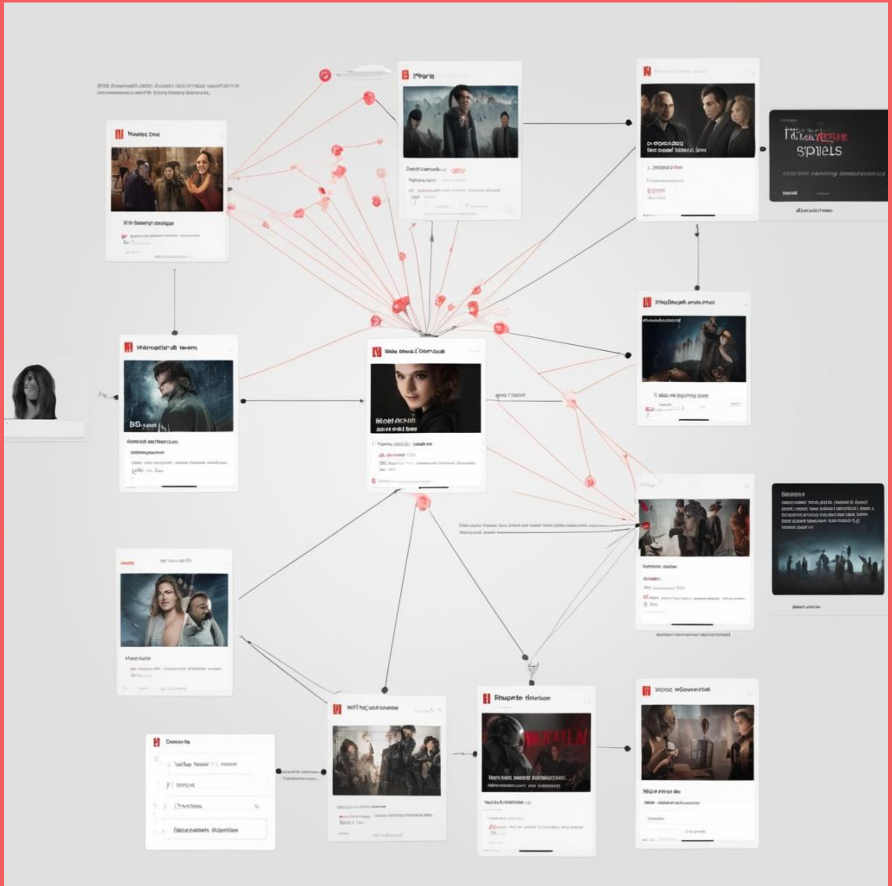


Illustration of movie recommendations

Fraud Detection

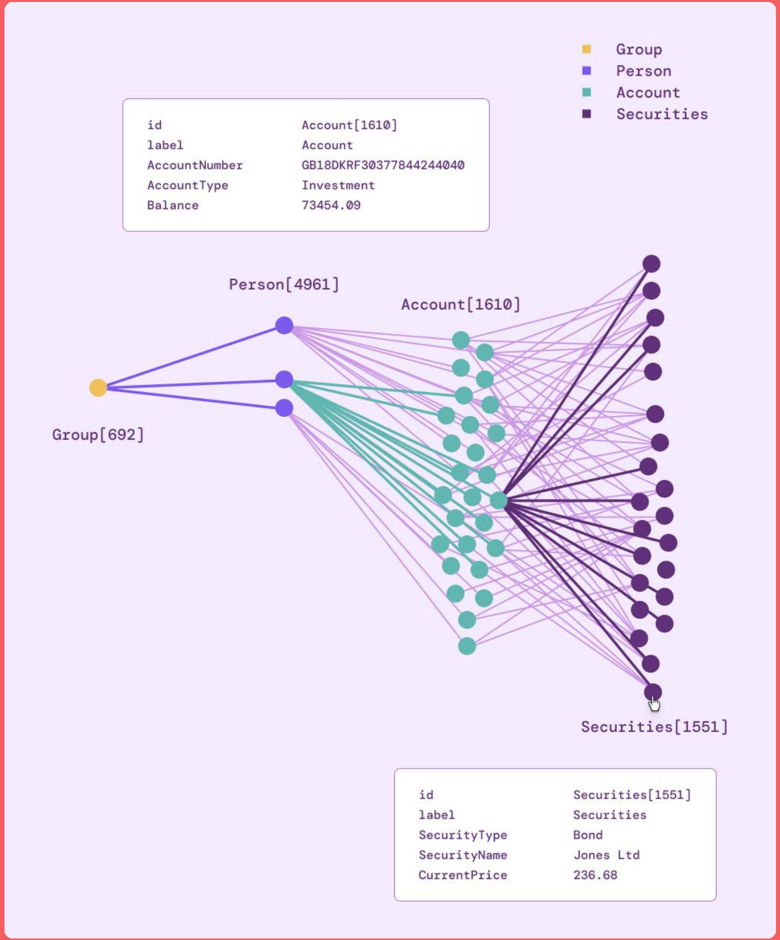
Use Case: Identifying suspicious activities in transactions.

Example: Banks detecting fraudulent credit card transactions.

Method: Analyzing transaction graphs for anomalies.

Visualization: Graph highlighting abnormal transaction patterns.

Reference: https://www.puppygraph.com/blog/fraud-graph



An example fraud graph

Knowledge Graphs

Definition: A Knowledge Graph is a flexible, reusable data layer that connects and contextualizes data across various silos, enabling complex queries and insights.

Key Features:

- **Contextual Understanding:** Captures the situational, layered, and evolving nature of real-world knowledge.
- **Flexible Data Integration:** Links diverse data sources without altering the underlying data structures.
- **Ontology Implementation:** Utilizes ontologies to define domain knowledge, relationships, and rules.
- **Virtualization:** Accesses and integrates data from various sources without the need for duplication.

Reference:
https://www.stardog.com/knowledge-graph/

What is Knowledge Graph? by Stardog

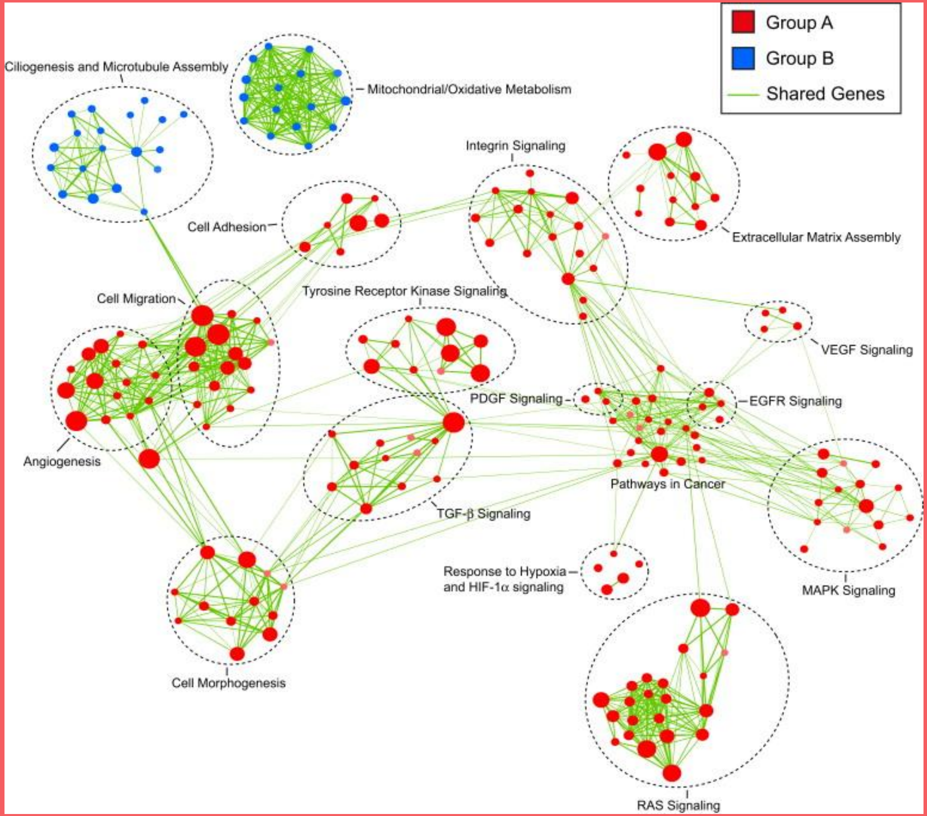
Biological Network Analysis

Definition: Graph theory models complex biological systems, such as gene regulatory networks and protein-protein interactions, to understand their structure and function.

Use Case: Identifying functional modules within protein-protein interaction networks to understand cellular processes.

Method: Applying clustering algorithms to detect densely connected subgraphs representing protein complexes

Image Source: https://cytoscape.org/cytoscape-tutorials/presentations/ppi-tools1-2017-mpi.html#/16



Functional annotation of a network of gene sets.

Graph Neural Networks (GNNs)

Definition: GNNs are deep learning models designed to operate on graph-structured data. They capture dependencies between nodes by aggregating and transforming information from their neighbors through message-passing mechanisms.

Use Case: Predicting molecular properties in chemistry by modeling molecules as graphs of atoms and bonds.

Method: Utilizing architectures like Graph Convolutional Networks (GCNs) to learn node representations by aggregating features from neighboring nodes. This approach allows the model to capture both local and global graph structures.

Image Source: https://arxiv.org/abs/1812.08434

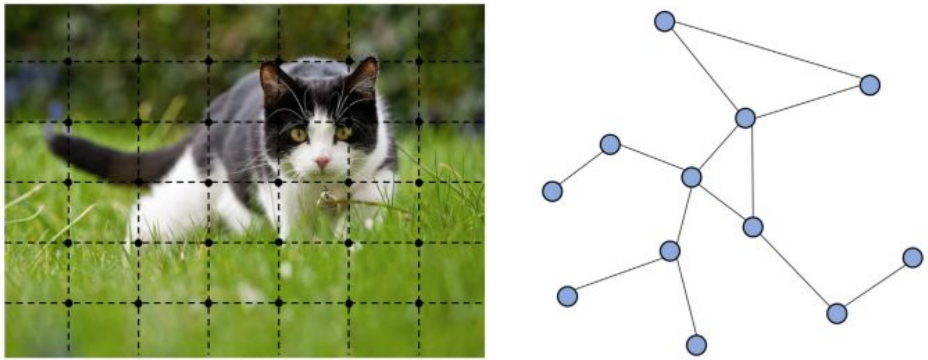


Fig. 1. Left: image in Euclidean space. Right: graph in non-Euclidean space.

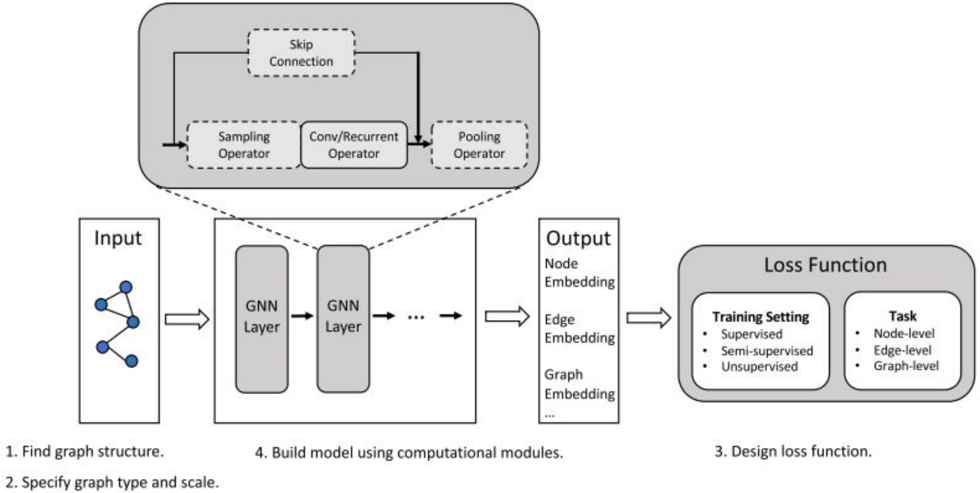


Fig. 2. The general design pipeline for a GNN model.

Transportation and Logistics

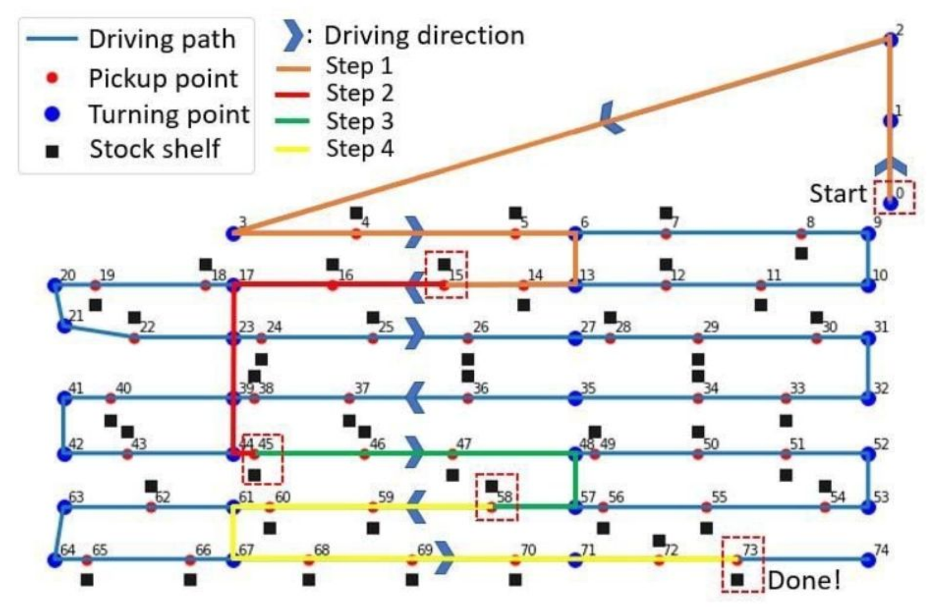
Description: Graph theory optimizes routing and logistics by modeling transportation networks as graphs, facilitating efficient pathfinding.

Use Case: Determining the shortest delivery routes for logistics companies to minimize time and cost.

Method: Applying algorithms like Dijkstra's or A* to find the shortest paths between nodes representing locations.

Visualization: Map-based graph with nodes as delivery points and edges as possible routes, highlighting optimal paths.

Reference:
https://builtin.com/machine-learning/graph-theory



Optimized driving route from picking list. | Image: Vegard Flovik

Application of Graph Theory in Transportation

Recent Advances in GNNs for AI Applications (2023–2025)

Integration with Large Language Models (LLMs):
- Approach: Combining GNNs with LLMs to enhance reasoning over structured data. GNNs provide topology-aware embeddings, which LLMs utilize for improved contextual understanding.
- Application: Enhancing knowledge graph completion and question-answering systems.
- Reference: "Injecting Knowledge Graphs into Large Language Models" (arXiv:2505.07554) .

Advancements in Temporal Graph Learning:
- Approach: Developing models like FLASH that adaptively sample temporal neighborhoods, improving learning efficiency on dynamic graphs.
- Application: Real-time recommendation systems and social network analysis.
- Reference: "FLASH: Flexible Learning of Adaptive Sampling from History in Temporal Graph Neural Networks" (arXiv:2504.07337)

GNNs in Materials Science:
- Approach: Employing GNNs to predict properties of novel materials by modeling atomic structures.
- Application: Accelerating the discovery of stable crystalline structures and new compounds.
- Reference: "Scaling Laws of Graph Neural Networks for Atomistic Materials Modeling" (arXiv:2504.08112)

Recent Advances in GNNs for AI Applications (2023–2025)

Applications in Intelligent Transportation Systems (ITS):

- Approach: Utilizing GNNs for traffic forecasting, vehicle control, and demand prediction by modeling transportation networks as graphs.
- Application: Optimizing traffic flow and enhancing autonomous vehicle navigation.
- Reference: "Graph Neural Networks in Intelligent Transportation Systems: Advances, Applications and Trends" (arXiv:2401.00713) .

Explainability in GNNs:

- Approach: Introducing frameworks like L2XGNN that identify subgraphs responsible for specific predictions, enhancing model interpretability.
- Application: Critical in domains requiring transparency, such as healthcare and finance.
- Reference: "L2XGNN: Learning to Explain Graph Neural Networks" (arXiv:2209.14402)

Thank You!