FreebaseViz: Interactive Exploration of Freebase Schema Using Query-driven Visualisation

Mahmoud Elbattah
College of Engineering and Informatics
National University of Ireland
mahmoud.elbattah@nuigalway.ie
Outline

• What is Visualisation?, and why it is important?
• Related Challenges in the Context of Big Data
• Dataset of Interest: Freebase
• Architecture of FreebaseViz Tool
• Live Demo and Visualisation Scenarios
• Conclusions
Introduction
What is Visualisation?

• The transformation of the symbolic into the geometric (McCormick et al., 1987).

• The use of computer-generated, interactive, visual representations of data to amplify cognition (S. K. Card et al., 1999).

Sources:
What is Different about Visualisation?

- The interpretation of visual formats happens immediately in a “pre-attentive” manner.


- The pictorial representation of data can help answer or discover questions.

- A particular significance in the era of Big Data.
Challenge: Visualisation of Large-Scale Data

- Query-ability
- Scalability
- Data Model Understanding

Visualisation at Large Scale
Dataset of Interest: Freebase
• Freebase: A huge structured entity database.
• Entities (Topics) about people, places, and things.
• ≈ 57 million Topics
• ≈ 650 Domains

Source: Kirrily Robert, Freebase 101 Presentation, August 2010
What we are trying to visualise?

Freebase Schema

- Categories (n=9)
- Domains (n=82)
- Types (n=2,120)
- Properties (n=6,851)
Example: What we are trying to visualise?
Proposed Approach
Architecture Overview

VivaGraph JS

ASP.NET Web Application

REST API (Neo4jClient)

JSON Response

Cypher Query

Neo4j Graph Database
Why Graph Database?

- High Reasoning Capabilities: Supported by complex graph traversals
- Graph-based Queries: Optimized for graph operations
- Flexibility: Schema-less data model
- NoSQL: Can Scale horizontally
Live Demo

http://freebaseviz.apphb.com/
Visualisation Scenarios
Scenario 1: Finding Dominant Types
Scenario 2: Category-Filtered Schema Graph
Scenario 3: Type-Filtered Schema Graph
Conclusions and Observations

- The Freebase schema resembled the structure of a scale-free network.

- The degree distribution followed a power law distribution.

- A few super-connected nodes dominated the schema graph connections.

- In contrast, a considerable proportion of the schema Types seemed isolated with no connections in the schema graph.
Conclusions and Observations (cont’d)

Graph databases can present promising potentials for visualisation environments as follows:

• Flexible schema-less modeling.

• Powerful query potentials.

• Complex graph traversal can answer queries requiring extensive navigation around a graph.

• Advantageous scalability compared to traditional relational models.
Original Paper

The original paper can be accessed from:

• https://books.google.ie/books?id=0YSKDQAAQBAJ&pg=PT130

Thank You!

Mahmoud Elbattah
mahmoud.elbattah@nuigalway.ie