Graph Databases in Football Analysis

A PROJECT PROPOSAL

FOR THE DATA MANAGEMENT COURSE OF THE MASTER'S DEGREE

IN

Engineering in Computer Science

Submitted by:

Federico Carmignani (1845479)

20th MAY 2022

Project

[Option 2.2]: Pick a NoSQL database system and a dataset showing a relevant usage of them over the selected dataset.

Content

The project consists in showing the utility of graph databases in **football analysis**, how they can be exploited and their points of strength in this field from match analysis to transfermarket.

Tool

The goal is to deepen the use of a graph-based and RDF-based DBMS: **Neo4j**. The reason is the need of working with events and temporal sequences, therefore the graph is a suitable data model for these tasks.

Dataset

The dataset to be used is composed of open source JSON or CSV files taken from Transfermarkt.com or StatsBomb.com, like the one in this repo: https://github.com/statsbomb/open-data.

Tasks

The idea is to basically use Neo4j in two tasks: the **first task** related to a dataset of transfermarkt.com and, using the graph as a data model, to perhaps represent the transfers that took place in a certain year and similar things; then it is about to show queries that can be let's say advantageous in a representation of this type of data, following a bit the false line of the demo that presented to us in class, but applied to this specific field. The **second task** is instead to show how a graph can be exploited in match analysis: in this, take a dataset from statsbomb.com relating to the events of a single match and show how the graph can be structured and then some useful queries.

Presentation and demo

Short demo on a graph properly created from the dataset and showing relevant queries on it, underlining the efficiency of the data model for the specific task.
Slideshow presentation organized approximately in this way:
1) graph-databases in sport analysis;
2) dataset: how it is composed and its content;
3) tool: how to use it and main features;
4) experiments: application to football analysis: transfers and events.