

ON KNOWLEDGE GRAPHS & PLAYER SCOUTING REPORTS

ARNAV PRASAD

UNIVERSITY OF CHICAGO – B.S. 2022

2020 SOUNDERS FC SOCCER ANALYTICS
CONFERENCE

arnavprasad@uchicago.edu

3 truths and the elephant in the room (or stadium)...

- **Truth #1:** Scouting has—and will always be—an essential component of soccer recruiting and player evaluation
- **Truth #2:** Traditional analytics cannot capture certain relationships and non-physical qualities of players that scouts/experts can identify
- **Truth #3:** Scouting ops. can be more integrated with data analytics
- **Question:** Why is there still a disconnect between scouting ops. and analytics?



It's time...

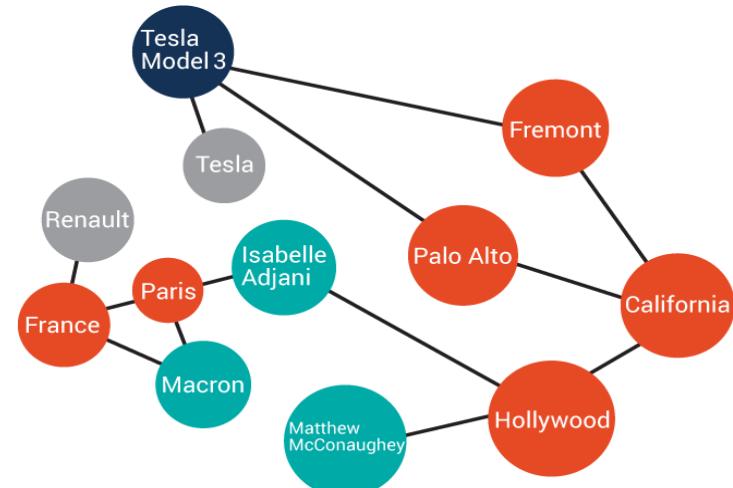


Iconic scene from Moneyball. But was the issue resolved?

A Problem & Proposal & Another Problem

- **Problem #1:** We need to use our scouts and scouting reports more intelligently. Teams need the context that scouts provide on players, but also the scalability and precision of analytics
- **Proposal:** Apply cutting edge knowledge graph and NLP technology to bring meaning to data
- **Problem #2:** Text data is difficult—(1) Teams need access to “digitalized” scouting reports (2) It is unstructured and thus hard to model (3) We are just scratching the surface
- Knowledge graph visualizations help teams add structure to their data without sacrificing at the expense of domain-expertise or technology
- The aesthetic presentation of the graphs can be fit to each group’s preference and need

So, What is a Knowledge Graph?



KGs represent a collection of interlinked “entities” (e.g., real-world objects, people, events, situations) and their relationships

Key Characteristics – Database (can be queried), Graph (can be analyzed as network data structure), Knowledge Base (“formal semantics” – to interpret data and infer new facts)

We can create KGs automatically using preexisting NLP libraries and visualize them by using and building visual graphics

KG image provided by Ontotext

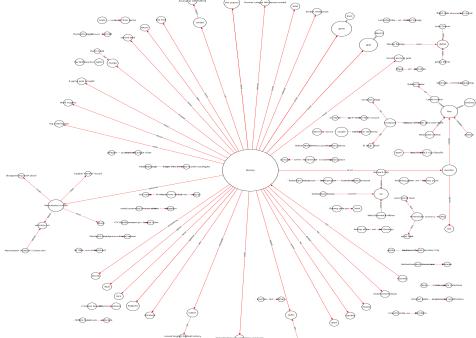
Approach & Results

- **Data:** Publicly-available Wikipedia articles that contained a certain player's name were scraped (in lieu of scouting reports)
 - NLP pipeline (in Python) →
 - (1) Collect/generate text data
 - (2) Automatic entity-pair extraction & other text preprocessing
 - (3) Visualization & knowledge graph creation
 - **Approach:** Text analytics visualizations, topic visualizations, interactive player knowledge graph, "queried" graph to specific entity → complete, visual scouting profile
 - **Results:** *What scouting-specific relationships can we capture?* Player's club, transfers, coaches, defining game moments, opponents, injuries, awards...

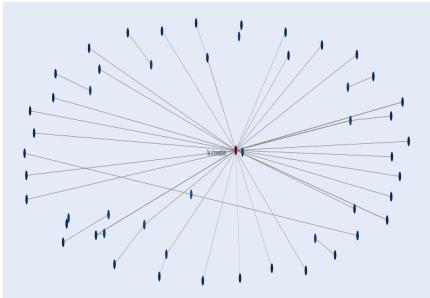
Note: Non-soccer topics were filtered out in KGs (Wikipedia has personal information that scouting reports do not!)

Example for Wayne Rooney

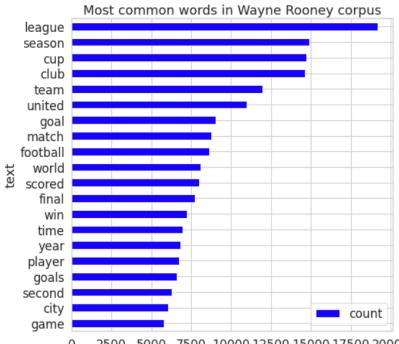
KG for 5 pages



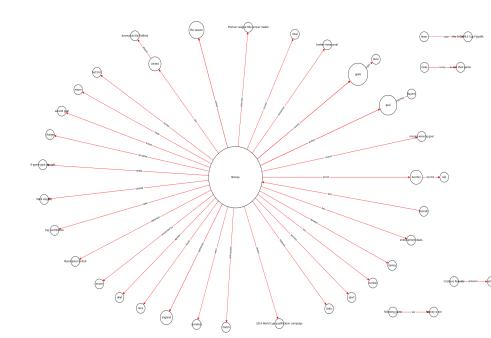
Interactive web module



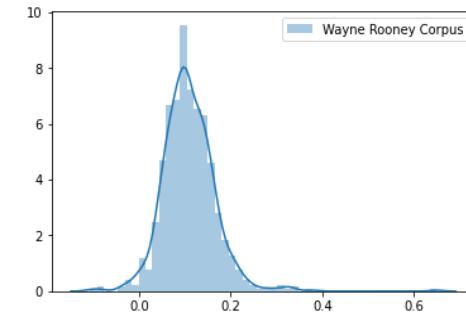
Most common words in entire corpus of ~800 documents



KG for 1 page (see slide 5)



Distribution of word polarity



Word-cloud for the corpus



To Infinity & Beyond...

- The potential for knowledge graphs and NLP in soccer and player scouting is limitless. We can:
 - Auto-generate visual profiles for scouts based on language in their reports
 - Visualize similar players based on their knowledge graphs (not just on entities themselves, but on the underlying relationships!)
 - Build knowledge graph visualizations to capture contextual information in-real time during games
 - Knowledge graph visualizations offer valuable contextual information on prospects and players for coaches, scouts, data analysts, and other key decision-makers
 - We need to close the gap between scouting and performance analytics!

Images & code are available upon request.

Thank You!

Questions, Comments, & Other Inquiries:
arnavprasad@uchicago.edu

