

On the Analysis of Large Integrated Knowledge Graphs for Economics, Banking, and Finance

EcoFinKG 2022

Shuai Wang (<https://shuai.ai>)

The Knowledge Representation & Reasoning Group
Vrije Universiteit Amsterdam

29th March, 2022

A decorative geometric pattern in the bottom right corner, consisting of a grid of triangles in various shades of blue, arranged in a larger triangular shape pointing towards the bottom right.

Content

1. Introduction

2. Integration

3. Analysis

4. Discussion

5. Conclusion

Related Work

Introduction

Integration

Analysis

Discussion

Conclusion

- 2007-10 US subprime mortgage crisis / 2007-08 global financial crisis
- company ownership & money laundering (Deloitte)
- the impact of the COVID-19 outbreak on Italian companies (Bellomarini et al.)
- customer banking (Shao et al.)
- public health policy, social resilience,

More projects that require bigger and and inter-disciplinary knowledge bases.



Introduction

Integration

Analysis

Discussion

Conclusion

- **Finance:** Financial Industry Business Ontology (FIBO), Financial Regulation Ontology (FRO), Insurance Regulation Ontology (IRO), Fund Ontology (FO), etc.
- **Economics:** STW (Standard Thesaurus Wirtschaft) Thesaurus, JEL classification system, etc.
- **Banking:** The WBG Taxonomy, Bank Regulation Ontology (BRO), etc. **Many are not open source projects.**



Our focus

Introduction

Integration

Analysis

Discussion

Conclusion

- Economics, Finance, Banking are three closely related domains.
- Most ontologies are developed for domain-specific tasks.
- We have more projects that requires inter-disciplinary knowledge.

We study the integration and the resulting graph:

- 1 statistical analysis
- 2 identity links
- 3 (pseudo-)transitive relations



The KGs (10 graphs + 1 mapping)

Introduction

Integration

Analysis

Discussion

Conclusion

- 1 the Financial Industry Business Ontology (we collected the FIBO ontology using OWL and FIBO vocabulary using SKOS)
- 2 the Financial Regulation Ontology (FRO)
- 3 the Hedge Fund Regulation (HFR) ontology
- 4 the Legal Knowledge Interchange Format (LKIF) ontology
- 5 the Bank Regulation Ontology (BRO)
- 6 the Financial Instrument Global Identifier (FIGI)
- 7 the STW Thesaurus for Economics (and its mappings)
- 8 the Journal of Economic Literature (JEL) classification system
- 9 the Fund Ontology



LogMap (web interface) without reasoner.

Table 1: Alignment of knowledge graphs

| | FIBO- vD | FIBO- OWL | LKIF | FIGI | STW | JEL | Fund |
|----------|-------------|--------------|------|------|-----|-----|------|
| FIBO-vD | - | 599 | 1 | 147 | 12 | 204 | 11 |
| FIBO-OWL | - | - | 24 | 516 | 5 | 57 | 70 |
| LKIF | - | - | - | 1 | 0 | 0 | 23 |
| FIGI | - | - | - | - | 0 | 34 | 2 |
| STW | - | - | - | - | - | 2 | 0 |
| JEL | - | - | - | - | - | - | 1 |
| Fund | - | - | - | - | - | - | - |



The Integrated Graph

RDFpro: <http://rdfpro.fbk.eu/>

Table 2: General statistics of knowledge graphs

| Name | V | E | Size |
|-------------------|----------------|------------------|-------------|
| FIBO-vD | 17,547 | 28,128 | 3.1MB |
| FIBO-OWL | 103,288 | 250,002 | 16MB |
| FRO | 94,215 | 283,976 | 16MB |
| HFR | 14,235 | 34,771 | 2.6MB |
| LKIF-Core | 1,005 | 2,363 | 141KB |
| BRO | 259,074 | 838,007 | 43MB |
| FIGI | 12,180 | 16,434 | 822KB |
| STW | 51,128 | 113,276 | 3.4MB |
| JEL | 12,109 | 177,57 | 1.1MB |
| Fund | 10,119 | 35,005 | 3.2MB |
| STW-mappings | 78,398 | 177,603 | 11MB |
| alignment | 2,327 | 1,698 | 255KB |
| integrated | 610,866 | 1,778,755 | 93MB |

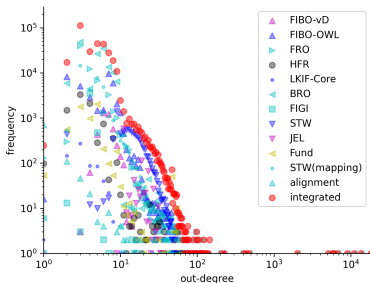
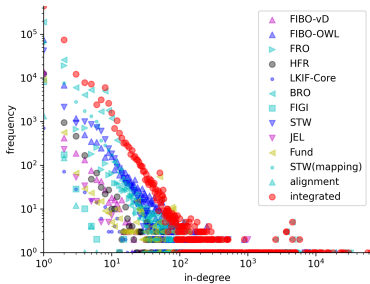


Table 3: Graph-theoretical statistics of knowledge graphs

| Name | maxSCC | p_S (%) | maxWCC | p_W (%) |
|-------------------|---------------|-------------|----------------|--------------|
| FIBO-vD | 1 | 0.01 | 17,535 | 99.93 |
| FIBO-OWL | 297 | 0.29 | 103,208 | 100 |
| FRO | 17 | 0.02 | 94,015 | 99.79 |
| HFR | 849 | 5.96 | 14,230 | 99.96 |
| LKIF-Core | 88 | 8.76 | 963 | 95.82 |
| BRO | 13 | 0.01 | 258,982 | 99.96 |
| FIGI | 13 | 0.11 | 12,180 | 100 |
| STW | 6777 | 13.25 | 51,128 | 100 |
| JEL | 1 | 0.01 | 12,099 | 99.92 |
| Fund | 109 | 1.08 | 10,111 | 99.92 |
| STW-mappings | 617 | 0.79 | 78,398 | 100 |
| alignment | 3 | 0.13 | 119 | 5.11 |
| integrated | 36,853 | 6.03 | 610,792 | 99.98 |



Analysis: Degree



Identity Links: CCs

Introduction

Integration

Analysis

Discussion

Conclusion

- 5,253 triples about `owl:sameAs`
- 29,556 triples about `skos:exactMatch`
- 8,172 triples about `skos:relatedMatch`
- 6,418 triples about `skos:closeMatch`



Identity Links: biggest WCCs

Introduction

Integration

Analysis

Discussion

Conclusion

- owl:sameAs: 8 and 6
- skos:exactMatch: 119 and 45
- skos:relatedMatch: 21
- skos:closeMatch: 52

Not big at all. Can be manually refined.



Compare against LOD-a-lot

- LOD-a-lot is a huge integrated graph of 650k knowledge graphs.
- LOD-a-lot has much more `owl:sameAs` than `skos:broaderMatch`.
- LOD-a-lot has much more nested cycles for (pseudo-)transitive relations.
- LOD-a-lot needs (semi-)automatic tools for the refinement.
- They show different graph-theoretical features.



Conclusion

Introduction

Integration

Analysis

Discussion

Conclusion

- produced a large integrated KG.
 - 1 studied how inter-connected the subgraphs are
 - 2 studied how integration enriches information of entities
 - 3 examined the quality of the integrated graph
- our graph can be used for future inter-disciplinary work

Contact: shuai.wang@vu.nl



Future Work

Introduction

Integration

Analysis

Discussion

Conclusion

- interoperability
- use cases: social resilience, public health, etc.

