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

Shuai Wang (shuai.wang@vu.nl)

The Knowledge Representation and Reasoning Group, Vrije Universiteit Amsterdam, 1081 HV Amsterdam, the Netherlands

Abstract

Knowledge graphs are being used for the detection of money laundering, insurance fraud, and other suspicious activities. Some recent work demonstrated how knowledge graphs are being used to study the impact of the COVID-19 outbreak on the economy. The fact that knowledge graphs are being used in more and more interdisciplinary problems calls for a reliable source of interdisciplinary knowledge. In this paper, we study the integration of knowledge graphs in the domains of economics, banking, and finance. Our integrated knowledge graph has over 610K nodes and 1.7 million edges. By performing statistical and graph-theoretical analysis, we demonstrate how the integration results in more entities with richer information. Its quality was examined by analyzing the subgraphs of the identity links and (pseudo-)transitive relations. Finally, we study the sources of error, and their refinement and discuss the benefit of our integrated graph.

Introduction

In this paper, we study properties of the integration of knowledge graphs by analyzing the statistical and graph-theoretical properties. More specifically, we study properties of integrated knowledge graphs by combining existing knowledge graphs in the domains of economics, banking, and finance.

Finance

The Financial Industry Business Ontology (FIBO) includes formal models that are intended to define unambiguous shared meaning for financial industry concepts. Another popular ontology is the Financial Regulation Ontology (FRO), which has been used as a higher level, core ontology for ontologies such as the Insurance Regulation Ontology (IRO), the Fund Ontology, etc.

Economics

The STW (Standard Thesaurus Wirtschaft) Thesaurus for Economics was developed by the German National Library of Economics (ZBW) and gained popularity in scientific institutes, libraries and documentation centers, as well as business information providers. The JEL classification system was initially developed for use in the Journal of Economic Literature (JEL) and is now a standard method of classifying scholarly literature in the field of economics.

Banking

Knowledge graphs have attracted increasing attention in the banking industry over the past decade. The WBG Taxonomy includes 3,882 concepts. In comparison, the Bank Regulation Ontology (BRO) is much bigger and uses two industrial standards, namely FIBO and LKIF, as its upper ontology. It was built on top of the FRO ontology, as mentioned above. Unfortunately, many knowledge graphs are developed by banks and are not open source.

Integration

By integrating knowledge graphs of various domains, we expect more entities and richer information for entities. The following is a list of 11 knowledge graphs we collected from 9 projects in the domains of economics, banking, and finance.

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

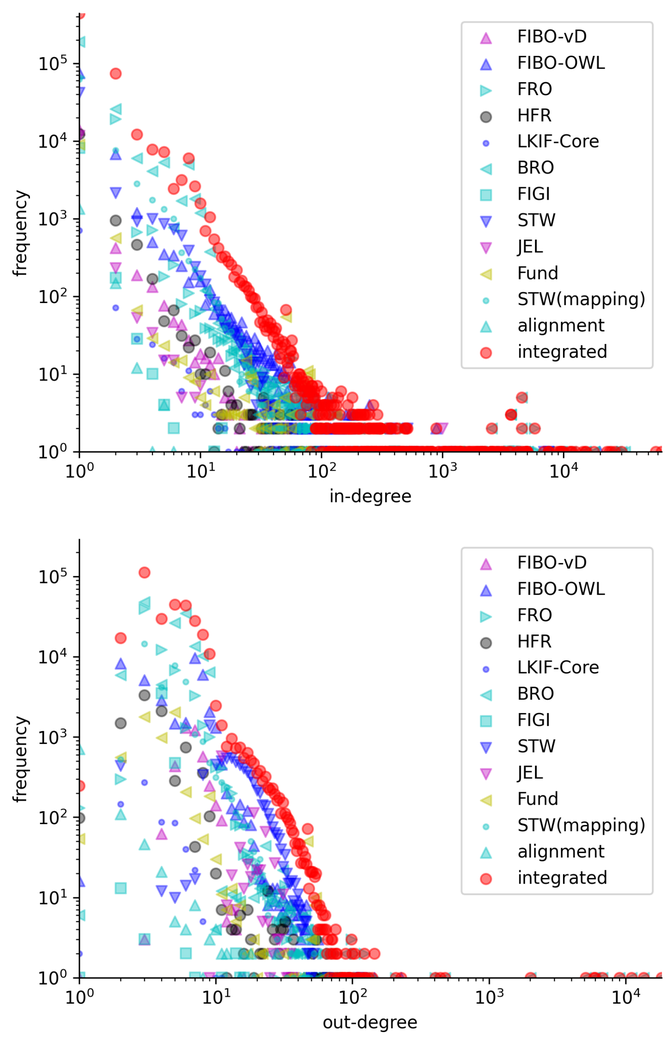
We used LogMap for the alignment between knowledge graphs. Overall, 1,698 unique identity links of `skos:exactMatch` were added to the integrated graph. All the knowledge graphs were first converted to Turtle format and then used the RDFpro for the integration process with duplicated triples removed.

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	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

Analysis of the integrated graph

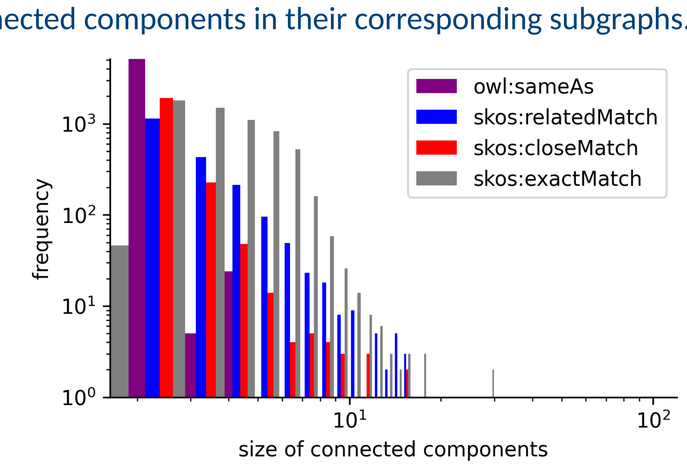
Statistical Analysis

We study how the information of entities can be enriched when combining different resources. When an entity is described in different domains, its in- and out-degree are expected to increase. The figure below illustrates the in-/out-degree of the knowledge graphs and the integrated knowledge graph. Both the in- and out-degrees of the integrated graph show a power-law distribution.



Analysis of identity links

Our analysis shows that only 5,253 triples about `owl:sameAs` are in the integrated graph against 31,254 triples about `skos:exactMatch`. In addition, there are 8,172 triples about `skos:relatedMatch`, and 6,418 triples about `skos:closeMatch`. The figure below shows the frequency distribution of the weakly connected components in their corresponding subgraphs.



Analysis of transitive and pseudo-transitive relations

There are in total 20 relations typed `owl:TransitiveProperty`. Our analysis shows that `rdfs:subClassOf` is a popular relation with 47,597 triples but without cycles. We found some small cycles when examining the subgraph of `skos:broader`, `skos:broaderMatch`, `skos:narrower`, and `skos:narrowerMatch`.

Links

Source code & paper: https://github.com/shuaiwangvu/EcoFin-integrated.