SubInterNM: Optimizing the Matching of Networks of Ontologies

Fabio Santos[1], Kate Revoredo[2], and Fernanda Baião[3]

[1] Northern Arizona University, United States
[2] Vienna University of Economics and Business, Vienna, Austria
[3] Department of Industrial Engineering, Pontifical Catholic University of Rio de Janeiro, Brazil
fd252@nau.edu, kate.revoredo@wu.ac.at, fbaiao@puc-rio.br

Abstract. System of systems (SoS) are interconnected systems that bring value to different domains like health, emergency and crisis management systems. The integration of these SoS creates opportunities to change, validate the information, and add more value to information systems. SoS may have ontologies in their background to support knowledge description and semantic integration. Consequently, the integration of SoSs may benefit from the integration of the network of ontologies behind. However, the task of integrating networks of ontologies, especially the ones describing real-world SoS can be infeasible due to the size of the networks. In this work, we propose an approach, SubInterNM, based on algebraic operations that reduces the number of comparisons needed to match the networks behind the SoSs. We validated our approach using networks of ontologies created from the OAEI ontologies. The SubInterNM combined with Alin and LogMap can overcome these matchers, when running alone in some cases. [*]

1 Introduction

A System-of-Systems (SoS) is defined as a set of independent systems, providing functionalities derived from the interoperability among them [2]. Examples of SoS scenarios are smart cities, health, and emergency response systems, and crisis management systems [5]. Current SoS are increasingly supported by networks of ontologies, which provide a semantic backbone for modeling and reasoning over data.

A network of ontologies (NO) is a set of two or more aligned ontologies. The network can represent a set of domains of their compound ontologies. Each ontology describes the knowledge of a domain of interest [9]. Many ontologies networks may be created inside organizations as a response to the demand for a semantic interoperability layer among their information systems (IS).

In current scenarios demanding data integration and systems interoperability (such as company acquisitions) within the same business domain, different SoS must be integrated. Because of the intrinsically multiple possible relationships inside the SoS that includes IS from distinct companies, the integration can be challenging and may be viewed as a matching of ontologies networks, requiring mapping concepts between both SoSs. This research addresses how to match a network of ontologies.

This study proposes an approach to optimize internetwork matching in the context of networks of ontologies, by systematically examining the characteristics of the NO and avoiding computing all possible matchings between entities. More specifically, we implement the subsumed internetwork matching (SubInterNM) [8], which reduces the number of pairs to be evaluated in the matching process. We evaluated the proposed approach in a preliminary experiment using an OAEI dataset.

This work addresses the following research question: *"Is it possible to align two ontology networks without computing all the possible alignments, with viable computational effort, time and precision, recall and f-measure?"*. To answer this question, this paper contributes by proposing an approach to match network of ontologies and its implementation into a prototype tool, as well as an empirical study showing the viability of the approach and its performance gains over state-of-the-art matching tools.

This work is organized as follows: in section 2, we summarize background. Evaluation results and a discussion are presented in section 3. Section 4 describes the limitations, and finally, section 5 concludes and points to future work.

2 Background

Current approaches for ontology matching include pairwise matching [9] and holistic matching [7]. Both may be adapted to match Networks of Ontologies; however, they perform an exhaustive checking of every single possible pair of entities for each ontology that composes the networks. They also have limited scalability, since the required number of steps for computing all the alignments grows exponentially to the number and sizes of the ontologies composing each network. Indeed, both pairwise or holistic approaches are not prepared to match Networks of Ontologies [8] since they do not take into account the structure of the networks and cannot limit the number of comparisons.

The ontology matching problem is not new and has been researched for a decade. However, to our knowledge, the matching of Network of Ontologies has not received the same attention. Although there are few studies dealing with networks of ontologies and, consequently, matching networks, there are still open challenges [1].

3 Results and Discussion

To assess SubInterNM we conducted an experiment using ontologies from the OAEI conference domain, so as to limit the size and help checking the results manually.

We selected Alin [4] and LogMap [6] as the matchers for the experiment. Alin obtained one of the best metrics in the OAEI initiative, and LogMap is one of the best to handle large ontologies. The SubInterNM approach uses the definitions in [3] and is available online [b]. For the sake of reprodutibility, the results are also available online [c].

We first selected ontologies to compose two networks of ontologies. Since LogMap and Alin are not able to natively process networks, the baseline submitted as input to

[b]https://github.com/fabiojavamarcos/interNetworkOntologyMatching

[c]zenodo.org (10.5281/zenodo.3977855)

Table 1. Precision Recall and F-Measure subInterNM+LogMap(1) subInterNM+Alin(2) LogMap(3) Alin(4) - *Did not calculate the metrics

Exp	**P(1)**	**R(1)**	**F(1)**	**P(2)**	**R(2)**	**F(2)**	**P(3)**	**R(3)**	**F(3)**	**P(4)**	**R(4)**	**F(4)**
2x2	0.818	0.600	0.692	**0.90**	**0.667**	**0.769**	0.842	0.432	0.571	0.348	0.405	0.375
4x4	0.818	0.600	0.692	**0.90**	**0.667**	**0.769**	0.750	0.141	0.237	0.197	0.192	0.195
5x5	0.818	0.600	0.692	**0.90**	**0.667**	**0.769**	0.583	0.126	0.207	0.096	0.102	0.099
5x1	0.690	0.233	0.348	**0.904**	0.221	0.355	0.778	**0.244**	**0.371**	0.0*	0.0*	0.0*
5x2	0.655	**0.268**	**0.380**	**0.888**	0.225	0.359	0.388	0.164	0.265	0.537	0.132	0.212
5x3	0.833	**0.435**	**0.572**	**0.882**	0.326	0.476	0.674	0.179	0.283	0.3	0.162	0.21

these matchers consisted of the union of all ontologies from each network. The alternative scenario to be compared consisted of executing SubInterNM and then submitting its partial results to the LogMap and Alin matchers. Following, each possible pair of ontologies was submitted to LogMap and Alin alone, with duplication and without duplication (when pairs of the same ontologies were manually eliminated, i.e. Edas x Ekaw and Ekaw x Edas; Edas x Edas). For each scenario we collected the following metrics: processing time, average precision, average recall, and average f-measure.

- 2x2: Ω = {sigkdd, confof} and Ω' = {conference, confof};
- 4x4: Ω = {sigkdd, confof, ekaw, edas } and Ω' = {conference, confof, ekaw, edas};
- 5x5: Ω = {sigkdd, confof, ekaw, edas, iasted } and Ω' = {conference, confof, ekaw, edas, iasted};
- 5x1: Ω = {sigkdd, confof, ekaw, edas, iasted} and Ω' = {conference};
- 5x2: Ω = {sigkdd, confof, ekaw, edas, iasted } and Ω' = {conference, confof};
- 5x3: Ω = {sigkdd, confof, ekaw, edas, iasted} and Ω' = {conference, confof, ekaw};

The results show higher precision, recall, and f-measure using the SubInterNM combined with the Alin compared with the matcher alone and when combined with LogMap (Table 1). LogMap combined with SubInterNM or alone was faster than SubInterNm+Alin or Alin alone, even when the network size grew (Table 2).

The metrics (Table 1) showed a decrease in the values of the Alin and LogMap approach as the network grew. It can be explained by the lack of flexibility of the solutions in understanding a reference alignment that contains concepts coming from different ontologies. It occurs because when finishing the union operation, the resulting temporary ontology is composed of the union of concepts from all the ontologies together. LogMap handled better than complexity and loosed significantly less precision than Alin. The experiments using the matcher alone started aligning the structure after the union to standardize the input for all approaches.

In Table 3, we ran the matcher in all possible combinations, O_i x O_j, (column ”All”) or without duplications (column ”Time”). We computed the sum of processing time and the average of the quality metrics. Alin obtained higher quality metrics again, while LogMap continued to be faster. Looking at the 5x5 and 5x1 cases, we observe that Alin used significantly more time than when combined with the SubInterNM, but delivered better quality metrics. LogMap outperformed all the options wrt processing time and obtained similar metrics in the 5x5 case and better ones in the 5x1 case.

Table 2. Processing Time (seg)

Experimemt	subInterNM+LogMap	subInterNM+Alin	LogMap	Alin
2x2	4.345	8.587	**3.766**	8.71
4x4	29.736	34.931	**20.541**	22.516
5x5	65.65	71.494	**19.999**	32.137
5x1	16.548	21.442	**11.668**	20.537
5x2	20.041	25.151	**11.4**	15.989
5x3	23.496	29.445	**11.784**	19.445

Examining the column "All" (Table 3), we observe that LogMap processing times were comparable to SubInterNM + LogMap, but the latter approach produced a result without duplicated alignments. Alin alone had better results than SubInterNM + Alin but used significantly more time and delivered solutions with duplications and may cost more $O(nlogn)$ effort and more time. In networks with many isomorphisms, the SubInterNM + Alin delivered more balanced results combining metrics and time processing. On the other hand, classic pairwise approaches have better outcomes when the networks had few isomorphisms.

Finally, it is possible to avoid the Cartesian product and keep processing time and resulting metrics depending on the characteristics of the networks being aligned.

Table 3. Time (seg) and average (Precision, Recall and F-Measure) - Pairwise Individually

Exp	Alin Time	Alin All	P	R	F	LogMap Time	LogMap All	P	R	F
2x2	18.581	24.292	**0.950**	**0.889**	**0.918**	**8.323**	**10.274**	0.897	0.727	0.801
4x4	63.892	88.148	**0.942**	**0.756**	**0.836**	**25.447**	**33.615**	0.809	0.604	0.688
5x5	96.768	163.343	**0.979**	**0.721**	**0.823**	**38.521**	**63.556**	0.853	0.592	0.674
5x1	32.570	32.570	**0.934**	**0.74**	**0.820**	**15.256**	**15.256**	0,753	0.595	0.660
5x2	57.076	62.787	**0.972**	**0.768**	**0.853**	**24.076**	**26.027**	0.891	0.596	0.707
5x3	82.548	94.621	**0.959**	**0.711**	**0.815**	**33.368**	**37.238**	0.845	0.631	0.719

4 Limitations

When pruning the Network of Ontologies to reduce the posterior matcher computations, some entities can be missing. This may impact on how the similarity algorithms find the alignments, which may lead to different results. The use of a dataset from the same domain is not a real scenario, as discussed in Section 1. Yet, this enabled us to manually verify the algebraic operations and the computed metrics, since we needed many customized reference alignments.

Because of the many possibilities in the experiments, we needed to create some new reference alignments based on the existing from OAEI. These were validated by the research group but are not error-proof.

Finally, the intrinsic characteristics of the ontologies considered in the experiment generated sparse graphs, which may have helped the algebraic algorithms. In scenarios with more dense ontologies (i.e., with more connections among their concepts), we could have strongly connected graphs and, consequently, worse time processing when using our proposed SubInterNM.

5 Future Work and Conclusion

This paper contributes by presenting a novel concept of a Network of Ontologies Matcher approach.The proposal was implemented in a prototype to show its feasibility and confirm (our research question) that it is possible to align two ontology networks without computing all the possible alignments, with viable computational effort, time and precision, recall and f-measure. As predicted, the experiment results confirmed that SubInterNM computed the matching among distinct networks of ontologies more efficiently than using the traditional pairwise approach in specific cases, due to avoiding unnecessary comparisons without losing information.

For future work, we aim to run experiments with larger ontologies and networks, discover the optimal point where the SubInterNM can be used and add a strategy to retracting/forgetting axioms/entities [10] while preserving entailment.

References

1. de Abreu Santos, F.M., Revoredo, K., Baião, F.A.: Paving a research roadmap on network of ontologies. In: OM@ ISWC. pp. 221–222 (2017)
2. Boehm, B.: A view of 20th and 21st century software engineering. In: Proceedings of the 28th international conference on Software engineering. pp. 12–29. ACM (2006)
3. Casanova, M.A., Magalhães, R.C.: Operations over lightweight ontologies and their implementation. In: Implicit and Explicit Semantics Integration in Proof-Based Developments of Discrete Systems, pp. 61–82. Springer (2020)
4. Da Silva, J., Revoredo, K., Baião, F., Euzenat, J.: A lin: improving interactive ontology matching by interactively revising mapping suggestions. The Knowledge Engineering Review **35** (2020)
5. Fitzgerald, J., Foster, S., Ingram, C., Larsen, P.G., Woodcock, J.: Model-based engineering for systems of systems: the compass manifesto. COMPASS Interest Group, Tech. Rep. Manifesto Version **1** (2013)
6. Jiménez-Ruiz, E., Cuenca Grau, B.: Logmap: Logic-based and scalable ontology matching. In: International Semantic Web Conference. pp. 273–288. Springer (2011)
7. Megdiche, I., Teste, O., Trojahn, C.: An extensible linear approach for holistic ontology matching. In: International Semantic Web Conference. pp. 393–410. Springer (2016)
8. Santos, F., Revoredo, K., Baiao, F.: A proposal for optimizing internetwork matching of ontologies. In: Ontology Matching: OM-2018: Proceedings of the ISWC Workshop. p. 71 (2018)
9. Shvaiko, P., Euzenat, J.: Ontology matching: state of the art and future challenges. IEEE Transactions on knowledge and data engineering **25**(1), 158–176 (2013)
10. Wang, K., Wang, Z., Topor, R.W., Pan, J.Z., Antoniou, G.: Concept and role forgetting in ALC\{\mathcal \{ALC\}\} ontologies. In: The Semantic Web - ISWC 2009, 8th International Semantic Web Conference. pp. 666–681