Multi-domain Semantic Similarity in Biomedical Research

João D Ferreira[*] and Francisco M Couto

Abstract

Background: Given the increasing amount of biomedical resources that are being annotated with concepts from more than one ontology and covering multiple domains of knowledge, it is important to devise mechanisms to compare these resources that take into account the various domains of annotation. For example, metabolic pathways are annotated with their enzymes and their metabolites, and thus similarity measures should compare them with respect to both of those domains simultaneously.

Results: In this paper, we propose two approaches to lift existing single-ontology semantic similarity measures into multi-domain measures. The *aggregative* approach compares domains independently and averages the various similarity values into a final score. The *integrative* approach integrates all the relevant ontologies into a single one, calculating similarity in the resulting multi-domain ontology using the single-ontology measure.

Conclusions: We evaluated the two approaches in an multidisciplinary epidemiology dataset by evaluating the capacity of the similarity measures to predict new annotations based on the existing ones. The results show a promising increase in performance of the multi-domain measures over the single-ontology ones in the vast majority of the cases. These results show that multi-domain measures outperform single-domain ones, and should be considered by the community as a starting point to study more efficient multi-domain semantic similarity measures.

Keywords: semantic similarity; epidemiology; multiple-domain; prediction

1 Introduction

Ontology-based semantic similarity uses the machine-readable definitions of concepts provided by ontologies to compare annotated entities based on their *meaning*. Contrast this with other similarity measures that use *structural* and/or *physical* properties of the entities: *e.g.* proteins have traditionally been compared based on their aminoacid sequence, chemical compounds on the graph representing their molecular structure, *etc*. While non-semantic measures are effective to a certain degree, they fail in some edge cases, such as proteins with similar functions having different sequences, or chemical compounds with similar molecular structure having disparate biological roles.

Semantic similarity between annotated biomedical resources has been a topic of research since Lord et al. [1] applied this technique to GO-annotated proteins, as a search tool within a protein database. With the increase in the amount of biomedical domains being represented in formal ontologies, the de-

sire to use ontologies to annotate biomedical entities increases, which resulted in multiple ontologies being used to that effect: metabolic pathways [2], mathematical models of biological processes [3], functional tissue units [4], epidemiological resources [5], *etc*. These multidisciplinary entities, along with their multi-ontology annotations, can be regarded as biomedical digital resources that describe complex real-world phenomena.

Given the success of single-ontology semantic similarity measures in the past, for example, to assist text-mining [6–8], machine-learning [9–11], differential diagnosis [12], visualization [13], *etc.*, we argue that semantic similarity measures need to be developed to handle the multidisciplinary of these types of resources; nevertheless, research in this field is still stalled in the single-domain world. For example, to compare metabolic pathways, Clemente et al. [14] used semantic similarity between its enzymes, and Grego et.al [15] used semantic similarity between its metabolites; a more accurate approach, however, would be to take into consideration both the enzymatic and chemical domains: the simultaneous use of both types of information should, in theory, provide a more accurate

[*]Correspondence: jdferreira@fc.ul.pt
LASIGE, Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal
Full list of author information is available at the end of the article

insight into what the pathways represent in the real world and, ultimately, contribute to a similarity measure more aligned with the scientific knowledge that surrounds the pathways.

However, given the lack of proper evaluation testbeds for the purpose of validating multi-domain semantic similarity, to the best of our knowledge, this type of algorithm has yet to be proposed. In this paper, we describe and apply a methodology to create multi-domain semantic similarity measures from single-ontology ones.

The case study is a dataset of epidemiology resources, an inherently multidisciplinary field of research. The results obtained in this dataset are meant to achieve two goals: (a) we show that the proposed approaches to the multi-domain similarity problem are effective, at least in comparison with the single-ontology counterpart; and (b) we hope to stimulate the community to think about the problematic of multi-disciplinary similarity surrounding the ideas of knowledge representation and ontologies in the biomedical domain.

Instead of a new measure of semantic similarity designed from scratch to handle multidisciplinarity, we propose two approaches that can *lift* single-ontology measures into multi-domain measures. The "aggregative" approach compares each of the domains of relevance independently using existing single-ontology measures and then aggregates the several calculated values; the "integrative" approach integrates all the ontologies under the same common root and then applies single-ontology measures on it.

2 Multi-domain semantic similarity

Multi-ontology semantic similarity comes in two flavours: "single-domain" and "multi-domain". "Single-domain multi-ontology" semantic similarity is a technique that takes into account multiple ontologies that try to represent the same domain of knowledge, *i.e.* the ontologies have common concepts that represent the same real-world ideas, for example two ontologies of anatomy. The existence of these various ontologies that represent the same domain can result from the ontologies offering complementary views of the reality. Some previous work has been carried out with respect to this type of semantic similarity [16–18]. Contrast this with "multi-domain multi-ontology" measures, which use ontologies representing different domains of reality. In this case, the ontologies are orthogonal, *i.e.* they represent different domains of reality, and thus rarely have concepts in common, and when they do, the overlapping concepts are very general. This type of measure is able to compare resources annotated with concepts

from multiple domains of knowledge, as the biomedical entities mentioned above.

Notice that we are considering the multidisciplinarity of biomedical resources from the point of view of knowledge representation (KR): we propose a means to explore the ontology-provided definition of concepts to compare multidisciplinary entities annotated with concepts from more than one ontology (non-KR measures exist that are agnostic to the issue of multiple domains; *e.g.* Pederson et al. [19] compare concepts by comparing the textual neighbourhood of the concepts—the set of words that often appear *near* the concept in scientific literature).

Instead of creating a multi-domain measure from scratch, our methodology is to leverage on existing single-ontology measures, which have already been validated in a variety of scenarios, and *lift* them into multi-domain measures. As such, both the "aggregative" and "integrative" approaches take as input a single-ontology semantic similarity measure able to compare a set of concepts with another set of concepts (often called *groupwise* measures [20]).

The "aggregative" approach is depicted in Figure 1. In this approach we independently compare each domain using a single-ontology measure, *i.e.* the concepts from one domain in the first resource are compared to the concepts from the same domain in the second resource. We do this for all the domains used to annotate the resources and aggregate these single-ontology results into a single value by using an aggregating function such as the raw average, where all domains weight the same, or the weighted average, where each domain is weighted proportionally to the number of concepts used to annotate the resources in that domain. Schliker & Albrecht [21] propose a similar methodology, where GO-based semantic similarity is calculated independently for each of the three branches and then aggregated into a final similarity score.

The "integrative" approach consists in merging the relevant domain-specific ontologies into a single multi-domain ontology. In case the ontologies share a common upper ontology (as is common in the biomedical domain, where reference ontologies are expected to de derived from BFO [22]), this merging means that concepts from different ontologies have now common superclasses, even though they are from different domains. In the absence of a shared upper ontology, this merging is done by creating a root concept that subsumes all the root concepts of all the ontologies. We then use the single-ontology measure on top of this multi-domain ontology (see Figure 2).

The integrative approach has the advantage of being easy to implement and to straightforwardly enable the application of existing measures that have been
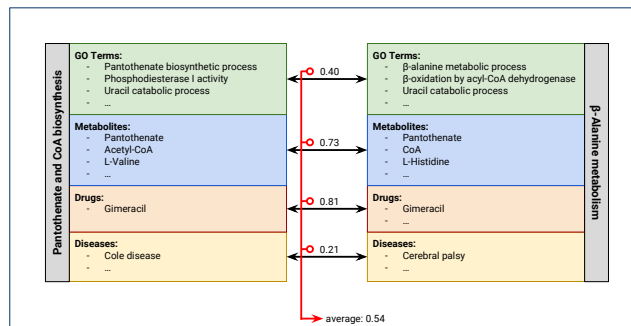
Figure 1 The aggregative approach. For each annotation domain in the entities being compared, the concepts in the first resource are compared with the concepts in the second. All the similarity values are aggregated into a final similarity score between, for example by using the average.
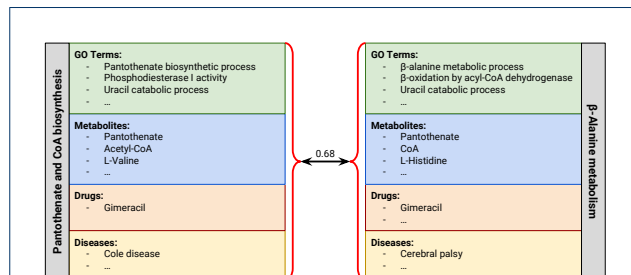
Figure 2 The integrative approach. All the concepts, irrespective of domain, are used to perform semantic similarity, which is done not with the individual ontologies but using a multi-domain ontology that consists of all the various ontologies merged under the same root. Only one similarity measure is used, resulting in a single final value.

proved useful in other endeavours. Additionally, it does not make use of arbitrary parameters for the domain weights. However, sometimes the ontologies are not as interoperable as expected. For example, both GO and CL contain a concept that represents "Cell", and this approach does not allow the measure to be aware of the fact that both represent the same thing and are, therefore, equivalent classes. On the plus side, these collisions are rare, and their number is decreasing, as the biomedical informatics community strives to create their ontologies in the most orthogonal way, with as much re-usability of concepts as possible [23].

In a multi-domain context, therefore, we can separate our measures of semantic similarity in four different settings:

Baseline This is a collection of measures, each corresponding to the single-ontology measure carried out in one of the domains used to annotate the entities. These measures serve as a baseline to determine whether the multi-domain approaches outperform single-ontology ones.

Aggregative (raw) All the single-ontology values obtained with the baseline setting are averaged with equal weights.

Aggregative (weighted) This is the same as last setting, except that the average of the values obtained for each domain are weighted in proportion to the number of annotations in that domain.

Integrative All the ontologies relevant for the similarity calculation are merged into one ontology and then the single-ontology measure is applied to it.

3 Multi-domain case study

Epidemiology is an inherently multidisciplinary subject, relying on areas of knowledge as diverse as medicine, biology, statistics, sociology and geography [24]. Even under the scope of medicine and biology, epidemiology deals with chemistry concepts, diseases, symptoms, environmental conditions, methods of transmission, vaccines *etc*. A multi-domain semantic similarity measure would enhance information retrieval mechanisms on a repository of epidemiology data. To support a meaningful search functionality, the repository has to show to the user a set of resources similar to their query, which requires a means to compare resources based not only on one domain of interest (such as diseases), but on all the domains of annotations of the resources. It is conceivable, for example, to imagine a user in need of data related to "flu" in "Europe" with "fever" and "sneezing" symptoms. A search engine needs to be able to deal with these domains in order to properly return to the user the set of resources they are requesting, in an order that meaningfully reflects the their relevance to the query.

In fact, the multidisciplinarity of epidemiology has been previously explored and a network of epidemiology-related ontologies has been created, which contains ontologies that represent most of the epidemiology domains mentioned above [25]. This network has been developed within the scope of an European project that developed the Epidemic Marketplace, a repository of epidemiology information [5] that used it to assist users annotate their resources, using ontology concepts as metadata.

A set of 204 resources were extracted from the Epidemic Marketplace, each corresponding to a scientific paper published in an epidemiology journal and annotated with concepts from the aforementioned network of ontologies.

Among the annotations for these resources, some use concepts from the NCI Thesaurus and MeSH, which are on the less formal end of the ontology spectrum,
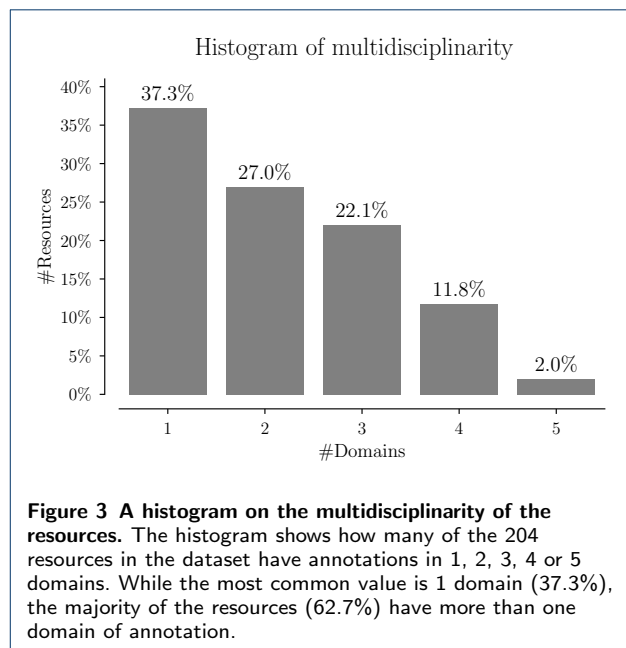
i.e. they resemble *ad-hoc* vocabularies more than formal ontologies, where the relationships between class and subclass do not always reflect subsumption (for example, in MeSH, "Population" is classified under "Population Characteristics", and in NCI Thesaurus "Inactivity" under "Physical activity" but no true hypernymy exist in these cases). Additionally, they are used in this dataset mainly to provide non-biomedical-specific concepts, such as "Family characteristics", which belong to the socio-economic sub-domain of epidemiology. For these reasons, these annotations were not included in our analysis.

A summary of the relevant annotations for these resources is given in Table 1 and Figure 3. The table shows that the resources are annotated with concepts from seven ontologies. These ontologies represent the domains of chemistry (CHEBI), diseases (DOID), environmental conditions (ENVO), phenotypic qualities (PATO), symptoms (SYMP), modes of disease transmission (TRANS) and vaccines (CO). In the table, **Coverage** is the fraction of resources that have at least one annotation in the specified domain, **Volume** is the average number of annotations from that domain within those resources, **Diversity** is the number of distinct concepts in that domain used in those annotations, and **Isolation** is the fraction of those resources that have annotations only in that domain. The figure shows that while a lot of resources are annotated with concepts from a single domain, the majority contain concepts from multiple domains. It also shows that the maximum number of domains is 5.

Table 1 Annotation statistics for the multi-domain resources extracted from the epidemic marketplace. "Coverage" is the fraction of resources with annotations in each domain, "Volume" is the average number of annotations in these resources for that domain, "Diversity" is the number of distinct concepts of that domain used to annotate the resources, and "Isolation" is the fraction of resources that have only annotations in that domain.

Ontology	Statistics			
	Coverage	Volume	Diversity	Isolation
CHEBI	0.01	1.00	1	0.000
DOID	0.67	1.76	70	0.215
ENVO	0.24	1.00	9	0.020
PATO	0.01	1.00	1	0.000
SYMP	0.51	3.55	79	0.118
TRANS	0.48	1.00	9	0.010
VO	0.23	1.06	16	0.010

As can be seen from this results, each domain contributes with a partial description of the resources: there is a sparseness in the annotation profile, with many resources having annotations in only a few domains, and not always on the same domains. No domain covers the whole dataset, and most resources are annotated with more than one domain. Additionally, even though 37.3% of the resources are annotated in a

Figure 3 A histogram on the multidisciplinarity of the resources. The histogram shows how many of the 204 resources in the dataset have annotations in 1, 2, 3, 4 or 5 domains. While the most common value is 1 domain (37.3%), the majority of the resources (62.7%) have more than one domain of annotation.

single domain (*e.g.* 21.5% of the resources have DOID annotations only), it is not the same domain that covers those resources. As such, to compare the resources in this dataset using the classical single-ontology semantic similarity, it would be necessary to select one domain, which means that several of the resources would need to be left out of the analysis and that a high volume of information would be disregarded, as it belongs to some other domain. Multi-domain semantic similarity seems to be essential in this case study to enable a proper comparison of the resources.

4 Evaluation

To assess the validity of semantic similarity in the gathered dataset, we determined the degree to which it is possible to predict the DOID annotations from the other annotations. The rational behind this method is that performing a clinical diagnosis is equivalent to predicting the diseases based on the other known factors (most notably symptoms) and is, therefore, one of the most important problems in biomedical informatics. In other words, we tried to predict which diseases are related to a resource based on the chemical compounds, the environmental conditions, the phenotype qualities, the symptoms, the modes of transmission and the vaccines that are related to it.

For this purpose, we used a multi-label machine learning algorithm, described by Zhang & Zhou [26]. This algorithm is known as ML-KNN, and uses a k-nearest neighbours (k-NN) approach to assign, to each resource, a set of DOID concepts. Using a k-NN-based algorithm is appropriate, since its performance highly

depends on the performance of the similarity measure used to find the neighbours. The following steps describe ML-KNN:

1. Compare each resource r to the other resources, and determine the k most similar ones (this is the neighbourhood of r);
2. With these k resources, build a Bayesian model to calculate the probability that each DOID concept (from the set of all concepts in the DOID ontology) is also one of the annotations, based on the frequency with which each distinct possible concept appears in the k neighbours.
3. Compute a metric of performance based on the probabilities derived in the previous step. The ML-KNN paper suggests five different metrics, which we use here.

We executed these steps with each of the settings delineated in the previous section (the baselines, raw aggregative, weighted aggregative and integrative settings); we also ran the calculations using several different groupwise single-ontology measures (Resnik + BMA [27, 28], Lin + BMA [29, 28], simUI [30] and simGIC [28]).

Figure 4 depicts the evaluation measure with respect to the several settings defined in the previous section using various values of k. These results were obtained using only the Resnik + BMA as the groupwise single-ontology similarity measure, because (i) the overall behaviour of the other groupwise measures does not differ significantly from the results that we are about to show, and (ii) this shows the best performance on this dataset. As can be observed, the integrative approach almost always outperforms the other settings, irrespective of evaluation metric and the value of k, which suggests that this measure is indeed superior to single-ontology measures in this dataset.

We can observe that the single-ontology measure performed on the "Symptoms" domain is the most successful baseline. This is justified by taking into consideration the annotation profile shown in Table 1. In fact, except for "Diseases", this is the domain with the highest coverage, volume and diversity. Additionally, from the set of domains used to annotate these resources, symptoms are the most closely related to diseases. For small values of k, the performance of this baseline is either on par or above the performance of the aggregative approaches.

However, 88% of the resources have annotations to concepts from ontologies other than SYMP, which means that using only SYMP to measure similarity leaves out information; the results suggest that, in fact, incorporating the other domains into the comparison algorithm increases the accuracy of the measure, as the evaluation metrics increase when we go from the SYMP baseline to the multi-domain measures.

It is interesting to notice the following overall trend: as k increases, the performance of the ML-KNN algorithm in the baseline settings decreases, especially for the SYMP baseline. This may be related to the fact that as we keep incorporating more and more neighbouring resources to predict DOID labels, we start including resources annotated with irrelevant symptoms, to the point where the extraneous information leads to a decrease in the algorithm's performance. However, the performance for the TRANS baseline increases with k, and it appears that its incorporation in the multi-domain approaches helps keep the multi-domain performance either at a plateau or even to increase (for the aggregative measures). This further consolidates the idea that taking into account multiple domains of annotation is necessary for a useful similarity measure.
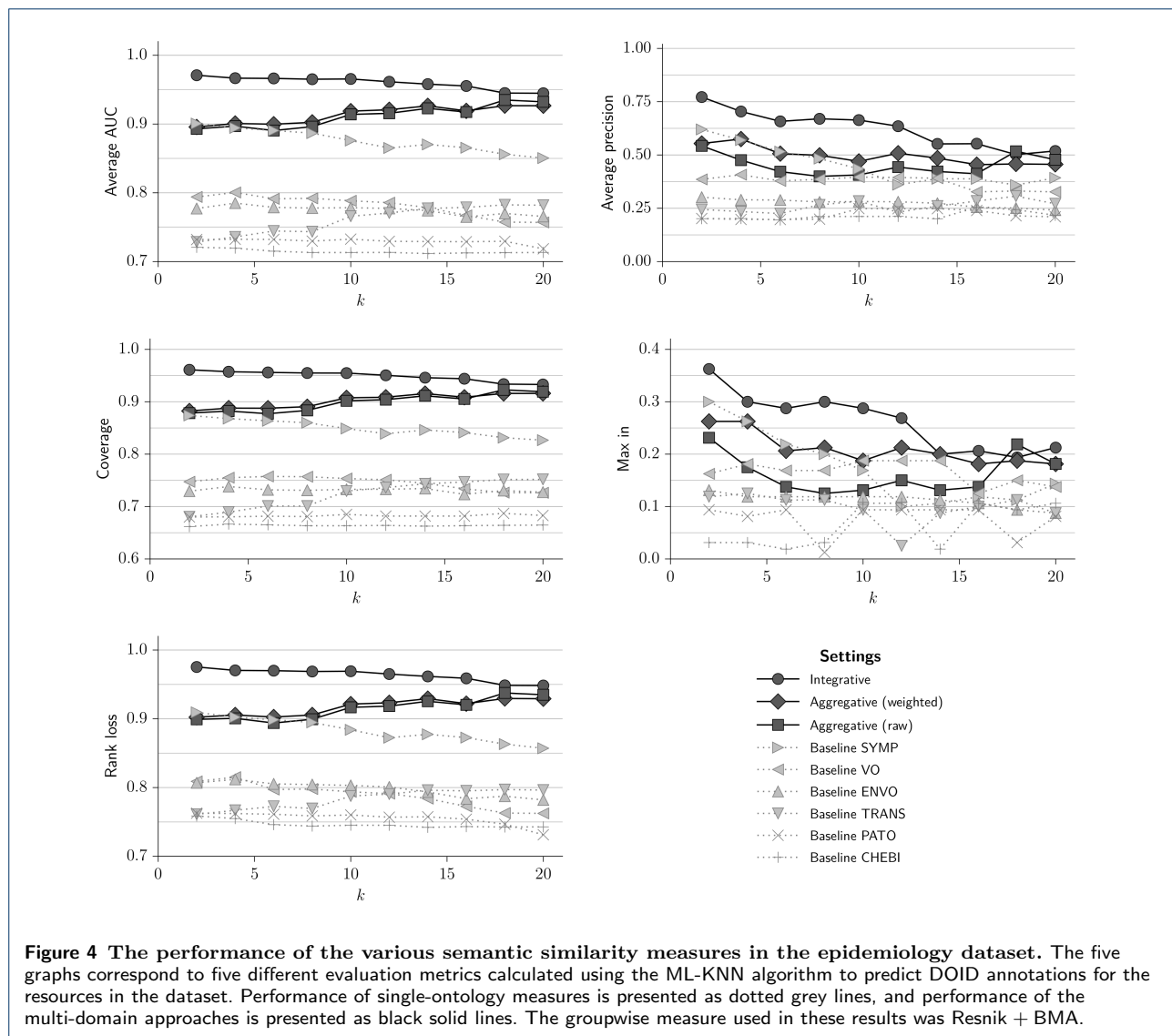
5 Discussion and conclusions

In this work, we demonstrate both the necessity and the feasibility of applying multi-domain semantic similarity measures in a dataset of epidemiology resources. Namely, we found that multi-domain semantic similarity measures can outperform single-ontology measures. This seems to be true especially when the annotations are sparsely distributed among the various domains. In these cases, the domains contribute equally to the overall similarity and increase the accuracy of the measure. For example, if "SYMP" annotations are not enough to make a prediction, annotations to concepts from the "TRANS" ontology can lead to an increase in accuracy.

The second fact extracted from the results is that the integrative approach almost always outperforms the single-ontology baselines and the other multi-domain approaches.

A third conclusion is that the "weighted aggregative" approach seems to be slightly more accurate than the "raw aggregative" approach (see the Max_in and Average precision metrics in Figure 4). We conjecture that this happens because the weighted approach uses more information to calculate similarity. However, the two approaches present almost indistinguishable performance in the other evaluation metrics.

To the best of our knowledge, no other previous work has tried to tackle this multi-domain problem. Thus, we feel necessary to provide a few proposals for future work. First, while our results show the superiority of the integrative approach over the baselines, they were obtained by a validation method that cannot be directly applied to create new knowledge or to perform information retrieval. We would like to test these multi-domain approaches with actual data repository users (*e.g.* by validating whether a "Related resources" section actually provides related resources, or

Figure 4 The performance of the various semantic similarity measures in the epidemiology dataset. The five graphs correspond to five different evaluation metrics calculated using the ML-KNN algorithm to predict DOID annotations for the resources in the dataset. Performance of single-ontology measures is presented as dotted grey lines, and performance of the multi-domain approaches is presented as black solid lines. The groupwise measure used in these results was Resnik + BMA.

by testing whether the multi-domain semantic similarity can suggest data owners new annotations based on the ones already used to annotate a resource). Human similarity gold-standards are also a possible avenue for future assessment.

Other possible avenues to pursue include (i) trying new groupwise single-ontology measures in evaluating the behaviour of the aggregative and integrative approaches, and (ii) defining new aggregation methods for the aggregative approach (*e.g.* weighting the average on the information content of concepts rather than the amount of concepts in each domain).

Furthermore, ontologies are starting to make cross-references and to reuse concepts from one another. For example, some GO concepts make explicit references to CHEBI concepts: the formal definition of "carbo-hydrate binding" says that this is a subclass of the GO concept "binding" with an explicit relationship ("has_input") to the ChEBI concept "carbohydrate". Given that current single ontology measures are not able to exploit this inter-domain relationships directly, the "aggregative" and "integrative" approaches are also unable to use the cross-references. To solve this issue, we need to create new measures that properly explore such relationships. While such measures have still not been developed, we think that a measure proposed by us in the past could be a starting point to tackle this problem [31]. This measure builds a semantic neighbourhood of concepts based on the relationships between the concepts, and then compares two concepts based on the overlap of their neighbourhoods. By incorporating cross-references in the semantic neighbourhood, we can indeed include inter-domain knowledge in the multi-domain measure.

In conclusion, we present evidence to support the hypothesis that multi-domain semantic similarity is both necessary and feasible, and propose two approaches to lift single-ontology measures (which have been actively developed throughout the last two decades) into multi-domain measures. Therefore, this paper presents two main contributions: we show that, in multidisciplinary contexts, we should not limit ourselves to single-ontology measures, as that has negative implications on the overall performance of semantic similarity; and, by extension, we provide a baseline for future multi-domain measures.

References
1. Lord, P.W., Stevens, R.D., Brass, A., Goble, C.A.: Investigating semantic similarity measures across the Gene Ontology: the relationship between sequence and annotation. Bioinformatics **19**(10), 1275–1283 (2003). doi:10.1093/bioinformatics/btg153
2. Croft, D., O'Kelly, G., Wu, G., Haw, R., Gillespie, M., Matthews, L., Caudy, M., Garapati, P., Gopinath, G., Jassal, B., Jupe, S., Kalatskaya, I., MayMahajan, S., May, B., Ndegwa, N., Schmidt, E., Shamovsky, V., Yung, C., Birney, E., Hermjakob, H., D'Eustachio, P., Stein, L.: Reactome: A database of reactions, pathways and biological processes. Nucleic Acids Research **39**(SUPPL. 1), 691–697 (2011). doi:10.1093/nar/gkq1018
3. Juty, N., Ali, R., Glont, M., Keating, S., Rodriguez, N., Swat, M.J., Wimalaratne, S.M., Hermjakob, H., le Novère, N., Laibe, C., Chelliah, V.: BioModels: Content, Features, Functionality, and Use. CPT: Pharmacometrics & Systems Pharmacology **4**(2), 3 (2015). doi:10.1002/psp4.3
4. de Bono, B., Grenon, P., Baldock, R., Hunter, P.: Functional tissue units and their primary tissue motifs in multi-scale physiology. Journal of biomedical semantics **4**(1), 22 (2013). doi:10.1186/2041-1480-4-22
5. Lopes, L.F., Silva, F.A.B., Couto, F.M., Zamite, J., Ferreira, H., Sousa, C., Silva, M.J.: Epidemic Marketplace: an information management system for epidemiological data. In: Information Technology in Bio- and Medical-Informatics, pp. 31–44 (2010)
6. Spasic, I., Ananiadou, S., McNaught, J., Kumar, A.: Text mining and ontologies in biomedicine: making sense of raw text. Briefings in Bioinformatics **6**(3), 239–51 (2005)
7. Varelas, G., Voutsakis, E., Raftopoulou, P., Petrakis, E.G.M., Milios, E.E.: Semantic similarity methods in wordNet and their application to information retrieval on the web. In: International Workshop on Web Information and Data Management, pp. 10–16 (2005). doi:10.1145/1097047.1097051
8. Lamurias, A., Ferreira, J.D., Couto, F.M.: Improving chemical entity recognition through h-index based semantic similarity. Journal of Cheminformatics **7**(Suppl 1 Text mining for chemistry and the CHEMDNER track), 13 (2015). doi:10.1186/1758-2946-7-S1-S13
9. Ferreira, J.D., Couto, F.M.: Semantic Similarity for Automatic Classification of Chemical Compounds. PLoS Computational Biology **6**(9), 1000937 (2010). doi:10.1371/journal.pcbi.1000937
10. Azuaje, F., Bodenreider, O.: Incorporating ontology-driven similarity knowledge into functional genomics: an exploratory study. In: IEEE Symposium on Bioinformatics and Bioengineering, pp. 317–324. IEEE, ??? (2004). doi:10.1109/BIBE.2004.1317360
11. Guo, X., Liu, R., Shriver, C.D., Hu, H., Liebman, M.N.: Assessing semantic similarity measures for the characterization of human regulatory pathways. Bioinformatics **22**(8), 967–73 (2006). doi:10.1093/bioinformatics/btl042
12. Köhler, S., Schulz, M.H., Krawitz, P., Bauer, S., Dölken, S., Ott, C.E., Mundlos, C., Horn, D., Mundlos, S., Robinson, P.N.: Clinical diagnostics in human genetics with semantic similarity searches in ontologies. American Journal of Human Genetics **85**(4), 457–64 (2009). doi:10.1016/j.ajhg.2009.09.003
13. Supek, F., Bošnjak, M., Škunca, N., Šmuc, T.: REVIGO summarizes and visualizes long lists of gene ontology terms. PloS One **6**(7), 21800 (2011). doi:10.1371/journal.pone.0021800
14. Clemente, J., Satou, K., Valiente, G.: Reconstruction of phylogenetic relationships from metabolic pathways based on the enzyme hierarchy and the gene ontology. Genome Informatics **16**(2), 45–55 (2005)
15. Grego, T., Ferreira, J.D., Pesquita, C., Bastos, H., Vila Viçosa, D., Freire, J.M., Couto, F.M.: Chemical and Metabolic Pathway Semantic Similarity. Technical report, Department of Informatics, Faculty of Sciences, University of Lisbon (2010). doi:DOI:10455/3335
16. Rodríguez, M.A., Egenhofer, M.J.: Determining Semantic Similarity among Entity Classes from Different Ontologies. Knowledge Creation Diffusion Utilization **15**(2), 442–456 (2003). doi:10.1109/TKDE.2003.1185844
17. Al-Mubaid, H., Nguyen, H.A.: Measuring Semantic Similarity Between Biomedical Concepts Within Multiple Ontologies. IEEE Transactions on Systems, Man, and Cybernetics **39**(4), 389–398 (2009). doi:10.1109/TSMCC.2009.2020689
18. Sánchez, D., Batet, M.: A semantic similarity method based on information content exploiting multiple ontologies. Expert Systems with Applications **40**(4), 1393–1399 (2013). doi:10.1016/j.eswa.2012.08.049
19. Pedersen, T., Pakhomov, S.V.S., Patwardhan, S., Chute, C.G.: Measures of semantic similarity and relatedness in the biomedical domain. Journal of Biomedical Informatics **40**(3), 288–99 (2007). doi:10.1016/j.jbi.2006.06.004
20. Pesquita, C., Faria, D., Falcão, A.O., Lord, P.W., Couto, F.M.: Semantic similarity in biomedical ontologies. PLoS computational biology **5**(7), 1000443 (2009). doi:10.1371/journal.pcbi.1000443
21. Schlicker, A., Albrecht, M.: FunSimMat: A comprehensive functional similarity database. Nucleic Acids Research **36**(SUPPL. 1) (2008). doi:10.1093/nar/gkm806
22. Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., Goldberg, L.J., Eilbeck, K., Ireland, A., Mungall, C.J., Leontis, N., Rocca-Serra, P., Ruttenberg, A., Sansone, S.-A., Scheuermann, R.H., Shah, N.H., Whetzel, P.L., Lewis, S.: The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nature Biotechnology **25**(11), 1251–5 (2007). doi:10.1038/nbt1346
23. Kamdar, M.R., Tudorache, T., Musen, M.A.: Investigating Term Reuse and Overlap in Biomedical Ontologies. In: Proceedings of the 6th International Conference on Biomedical Ontology, ICBO, pp. 27–30 (2015)
24. Porta, M. (ed.): A Dictionary of Epidemiology, 5th edn. Oxford University Press, 2008, ??? (2008)
25. Ferreira, J.D., Pesquita, C., Couto, F.M., Silva, M.J.: Bringing epidemiology into the Semantic Web. In: International Conference on Biomedical Ontologies (2012)
26. Zhang, M.L., Zhou, Z.H.: ML-KNN: A lazy learning approach to multi-label learning. Pattern Recognition Letters **40**(7), 2038–2048 (2007). doi:10.1016/j.patcog.2006.12.019
27. Resnik, P.: Using information content to evaluate semantic similarity in a taxonomy. In: International Joint Conference on Artificial Intelligence, vol. 1 (1995). 9511007v1
28. Pesquita, C., Faria, D., Bastos, H., Ferreira, A.E.N., Falcão, A.O., Couto, F.M.: Metrics for GO based protein semantic similarity: a systematic evaluation. BMC Bioinformatics **9 Suppl 5**, 4 (2008). doi:10.1186/1471-2105-9-S5-S4
29. Lin, D.: An information-theoretic definition of similarity. In: International Conference on Machine Learning, vol. 1, pp. 296–304 (1998)
30. Gentleman, R.: Visualizing and distances using GO. Technical report (2007)

31. Ferreira, J.D., Couto, F.M.: Generic semantic relatedness measure for
    biomedical ontologies. In: International Conference on Biomedical
    Ontologies (2011)