What Four Million Mappings Can Tell You about Two Hundred Ontologies

Amir Ghazvinian, Natasha Noy, Clement Jonquet, Nigam Shah, Mark Musen

What Four Million Mappings Can Tell You About Two Hundred Ontologies

Amir Ghazvinian, Natalya F. Noy, Clement Jonquet, Nigam Shah, Mark A. Musen

Stanford University, Stanford, CA 94305, US
{amirg,noy,jonquet,nigam,musen}@stanford.edu

Abstract. The field of biomedicine has embraced the Semantic Web probably more than any other field. As a result, there is a large number of biomedical ontologies covering overlapping areas of the field. We have developed BioPortal—an open community-based repository of biomedical ontologies. We analyzed ontologies and terminologies in BioPortal and the Unified Medical Language System (UMLS), creating more than 4 million mappings between concepts in these ontologies and terminologies based on the lexical similarity of concept names and synonyms. We then analyzed the mappings and what they tell us about the ontologies themselves, the structure of the ontology repository, and the ways in which the mappings can help in the process of ontology design and evaluation. For example, we can use the mappings to guide users who are new to a field to the most pertinent ontologies in that field, to identify areas of the domain that are not covered sufficiently by the ontologies in the repository, and to identify which ontologies will serve well as background knowledge in domain-specific tools. While we used a specific (but large) ontology repository for the study, we believe that the lessons we learned about the value of a large-scale set of mappings to ontology users and developers are general and apply in many other domains.

1 Why Create the Mappings?

The field of biomedicine has embraced the Semantic Web probably more than any other field. Ontologies in biomedicine facilitate information integration, data exchange, search and query of heterogeneous biomedical data, and other critical knowledge-intensive tasks [12]. As a result, there is a large number of biomedical ontologies covering overlapping areas of the field [3]. Creating mappings among ontologies by identifying similar concepts is a critical step in integrating data and applications that use different ontologies. With these mappings, for example, we can link resources annotated with terms in one ontology to resources annotated with related terms in another ontology, discovering new relations among the resources themselves (e.g., linking drugs and diseases).

As part of our work for the National Center for Biomedical Ontology (NCBO), we have developed BioPortal—an open community-based repository of biomedical ontologies [10].[1] This repository contains 140 ontologies with more than one million concepts among them. We view mappings between concepts in different ontologies as an

[1] http://bioportal.bioontology.org

essential part of the ontology repository. Users can browse the mappings, create new mappings, upload mappings created with other tools, download the mappings stored in BioPortal, or comment on the mappings and discuss them [11]. Other NCBO tools, such as a service for automatic creation of ontology-based text annotations [7], rely on the mappings to annotate biomedical resources with terms from different ontologies and for linking these resources to one another.

Over the past year, our team and our collaborators have uploaded more than 30,000 mappings to BioPortal. However, these mappings constitute only a tiny subset of the mappings between concepts in BioPortal ontologies. Thus, one of our goals was to use simple methods to generate quickly a large number of high-precision mappings to include in the repository. Our earlier studies have shown that in the case of biomedical ontologies simple lexical techniques, such as comparing preferred names of concepts and their synonyms, are extremely effective in creating mappings [6]. We have reported elsewhere [6] that this simple method for generating mappings achieves extremely high precision for biomedical ontologies—where preferred names of concepts and synonyms are used extensively and represent a rich source of information. In addition, we compared our results with the gold standard produced by the Ontology Alignment Evaluation Initiative (OAEI) [4]. For the use case of biomedical ontologies, our method achieved levels of recall and precision comparable to the best tools in the competition.

We have applied this simple lexical matching of preferred names and synonyms to generate mappings between concepts in BioPortal ontologies. In addition, the Unified Medical Language System (UMLS) [9] is a large collection of biomedical ontologies and terminologies that the researchers at the US National Library of Medicine have integrated. We plan to include UMLS terminologies in BioPortal in the near future and thus, when creating the mappings, we included the UMLS terminologies as well.

We generated mappings across all concepts in 140 BioPortal ontologies and 67 UMLS terminologies with more than 4 million concepts among them. We had two goals in generating the mappings: (1) we wanted to create a large set of mappings for the BioPortal resource that other applications can access and use; and (2) we wanted to learn more about the characteristics of the ontologies and the relationships between them. Using a set of more than 4 million mappings generated over our ontology set, we would like to answer several practical questions with implications for ontology reuse and development. These questions include, but are not limited to, the following:

To what degree are the domains covered by different ontologies connected? Mapping out the connections between ontologies will help us understand which domains are closely related, which ontologies may serve as a bridge between domains, and so on.

If you are new to a domain, what are the important or representative ontologies with good coverage? Ontology developers seeking to develop or reuse knowledge from a particular domain will be interested to know which ontologies have good coverage within that domain.

If you want to build domain-specific tools for creating ontology mappings, what are good ontologies to use for background knowledge? Many tools that seek to provide mappings for the purpose of ontology alignment use background knowledge to improve

their ability to produce valid and accurate mappings (e.g., [15, 1]). Such background knowledge allows these tools to use information about the representation of the domain to identify equivalent concepts. Thus knowing which ontologies are optimal for use as a source of background knowledge may greatly improve these tools.

What can we learn about the characteristics of the ontologies themselves and the ontology repository from the mappings between them? A set of mappings can provide insight about the ontologies themselves, their importance to their respective domains, or their coverage.

Researchers have previously successfully applied network analysis to gain insights into the structure and connectedness of large data sets [8]. In this paper, we apply network-analysis methods to analyze the ontologies and their mappings, to answer the questions posed above, and to reason about the distribution of mappings among the ontologies.

Note that our analysis does not depend on the specific methods that was used to generate the mappings and relies only on the fact that we have large set of high-precision mappings.

This paper makes the following contributions:

- We demonstrate that large-scale mapping sets can be useful in understanding the structure of an ontology repository, by identifying the most pertinent ontologies, the domains of overlap among ontologies, and the missing parts in an ontology repository.
- We propose network-based analysis metrics of ontologies based on mappings between them.
- We produce a set of more than 4 million mappings for the repository of ontologies and terminologies in BioPortal and UMLS.

2 Materials and Methods: What's in a Link?

We will now describe how we created the mappings used in this study and what data we analyzed.

We define a **mapping** as a relationship between two classes from different ontologies. The mappings that we discuss in this paper are **similarity mappings**: we declare that two classes from different ontologies are **similar** if the meaning that one class represents is similar or identical to the meaning of the other. We use the term "mapping" throughout this paper to refer to similarity mapping.

2.1 The NCBO Ontology Set

To develop our mappings, we used a set of 207 ontologies in the domain of biomedicine. These ontologies include 140 ontologies in BioPortal and 67 terminologies in the UMLS. The ontologies in BioPortal come from two sources: (1) 70 ontologies are downloaded nightly from the OBO Foundry repository;[2] (2) 70 ontologies are submitted by their developers directly to BioPortal. Among them, the 207 ontologies and terminologies that we used in this study contained 4,021,662 concepts.

[2] `http://obofoundry.org`

2.2 Creating Lexical Mappings Between Concepts

We created the mappings using the following steps, which we describe in detail in the rest of this section:

1. generate a database of *terms* used for preferred names and synonyms of ontology concepts;
2. *normalize* the strings in the database;
3. find pairs of *matching* terms;
4. create *mappings* between concepts based on the matching terms identified in the previous step.

In the first step—creating a database of preferred names and synonyms of all concepts in all ontologies—we needed to identify for each ontology which properties contained the strings representing these preferred names and synonyms. All UMLS terminologies have preferred names and synonyms clearly identified for all the concepts. Many of the BioPortal ontologies are represented in the OBO format, which also has designated properties to define preferred name and synonyms of a class. The OWL language does not itself provide any special annotation properties to store preferred names and synonyms (although many ontologies use `rdfs:label` for the former).[3] Thus, when users submit an OWL ontology to BioPortal, we ask them to indicate which OWL properties their ontology uses for preferred names and synonyms. We store the names of these properties as part of ontology metadata.

Extracting all terms for preferred names and synonyms resulted in a database of 7,637,125 terms. We then normalized all the strings, by converting them to lower case and removing all delimiters (e.g., spaces, underscores, parentheses, etc.). We used a mySQL database to store each term along with the ID for the ontology and the concept that it came from.

We used an SQL query to find pairs of matched terms among the normalized terms. From the database table of normalized terms, we utilized an SQL query to identify pairs of terms that matched exactly. To improve precision, we compared only strings with at least three characters and ignored the strings with three characters or less.

Since each term refers to a preferred name or synonym of a specific concept from a particular ontology, we used matching terms to connect concepts from different ontologies.

Consider the following example. The class "Myocardium" in Foundational Model of Anatomy (FMA) has the preferred name "Myocardium." The class "Heart myocardium" in the ontology of Mouse adult gross anatomy has a synonym "myocardium."[4] We will match the two normalized terms "myocardium" and therefore create a mapping between the two classes, "Myocardium" in FMA and "Heart myocardium" in Mouse adult gross anatomy ontology. Such mapping between the mouse myocardium represented in Mouse adult gross anatomy ontology and the human myocardium represented in FMA

[3] The Simple Knowledge Organization System (SKOS) provides the RDF-based vocabulary for defining preferred names and synonyms for concepts, but so far none of the BioPortal ontologies use SKOS.

[4] `http://bioportal.bioontology/org/ontologies`

can facilitate cross-species data exploration and integration. Having created this mapping, we could then integrate data annotated with the concept "Heart myocardium" in a database describing mouse experiments and data annotated with "Myocardium" describing human-related data.

This process resulted in a set of 4,001,775 mappings, where each each mapping represents a class from one ontology that is similar to a class from another ontology. The mapping is bi-directional as the similarity relationship generated this way is symmetric.

Note that UMLS itself contains a large set of manually created mappings between terms in different ontologies and terminologies. In the work described in this paper, we did not include those mappings. Rather, we used only the lexical mappings that we have generated. In future work, we plan to include the mappings that UMLS provides for additional information.

2.3 Identifying Links Between Ontologies

Because our goals include analysis of relations between *ontologies* and not individual concepts, we define a link between two ontologies based on a set of mappings between concepts from those ontologies. We use the notation $mapping(c_1, c_2)$ to describe a mapping between two concepts from different ontologies, such as the mappings that we described in Section 2.2. We denote a set of all concept-to-concept mappings between two ontologies S and T as $M(S,T)$, where $M(S,T) = \{mapping(c_s, c_t), c_s \in S, c_t \in T\}$.

Definition 1 (Mapping-Based Link Between Ontologies). *Given two ontologies, the source ontology S and the target ontology T, and a set of mappings between them $M(S,T)$, we say that there exists a mapping-based link L between ontologies S and T iff M is not an empty set: $M \neq \emptyset$.*

If two ontologies have at least one pair of concepts with similar names or synonyms between them, there will be a mapping-based link between the two ontologies. However, a more meaningful measure is the *number* of links between two ontologies or, more precisely, the *fraction* of one ontology that is mapped to another. For instance, if two ontologies each have 1,000 concepts, then, intuitively, the two ontologies are much closer to each other if 700 of these concepts match than if 5 concepts do. Thus, we define the notion of a percent-normalized link between ontologies which reflects not only how many mappings one ontology has to another, but also normalizes this measure with respect to the ontology size.

Definition 2 (Percent-Normalized Link Between Ontologies). *Given two ontologies, the source ontology S and the target ontology T, and a set of mappings $M(S,T)$ between them, we say that there is a percent-normalized link between S and T, $L_p(S,T)$ where $p \geq 0$ and $p \leq 100$, iff at least p% of the concepts in the ontology S are sources for the mappings in $M(S,T)$. We say that $L_0(S,T)$ holds if there is at least one mapping between concepts in S and T*

For instance, if an ontology S has 1,000 concepts, and 500 of these concepts are mapped to concepts in an ontology T, then $L_p(S,T)$ is true for all values of p from 0% to 50%.

Note that $L_p(S,T)$ is directional and it is entirely possible (and, in fact, common) for $L_p(S,T)$ to be true and for $L_p(T,S)$ to be false at the same time. If one ontology is much larger than another, a large fraction of the smaller ontology may be mapped to the larger ontology, but the set of mappings still constitutes a small portion of the larger ontology.

Intuitively, the percent-normalized link reflects how significant a set of mappings between ontologies is in the context of those ontologies. We evaluated the distribution of these links for several different values of p. We determined what percent of all ontology links were present at different values of p. Additionally, we implemented a graphical visualization of the links at each of these thresholds to analyze the clustering patterns and the link distribution for different values of p. Finally, we counted the number of links for each ontology at different values. We used this data to analyze the frequency with which an ontology has exactly k links.

3 Results

We used the data that we collected from the mappings to plot and analyze several metrics: First, we analyzed the number of links between ontologies at different values of p (i.e., at varying sizes of the mapped portion of the ontology, normalized by the ontology size) and the distribution of ontologies based on the number of links (Section 3.1). Second, we treated ontologies and the links between them as nodes and edges in a graph, again for several values of p. We analyzed the properties of these graphs as networks, using metrics such as the number of hubs and clusters (Section 3.2). Finally, we examined the overall similarity of the ontologies (Section 3.3). We discuss and analyze our results in Section 4.

3.1 How many links do ontologies have?

Figure 1 shows a distribution of the number of links that ontologies have for two values of p: 20% (Figure 1a) and 1% (Figure 1b). Recall that when $p = 20\%$, we create a link between source ontology S and target ontology T iff at least 20% of concepts from S are mapped to concepts from T. In other words, ontologies that we count in the graph on the right have a looser connection to each other than the ontologies in the graph on the left. Due to lack of space, we do not present the graphs for values of $p > 20\%$.[5] The distribution for larger values of p is very similar to the distribution for $p = 20\%$.

The graphs in Figure 1 show that the links between ontologies follow a power-law distribution for $p = 20\%$ (and larger values of p): There is a small number of ontologies that have large number of links and a large number of ontologies with just a few links. For smaller values of p, however, such as $p = 1\%$, where we include ontologies with very little overlap, our network becomes essentially random.

We analyzed the average distance between two nodes in the graph for some values of p. We found that for small values of p, the network is quite well connected and

[5] This data is available at `http://www.bioontology.org/wiki/index.php/Mapping_Set`

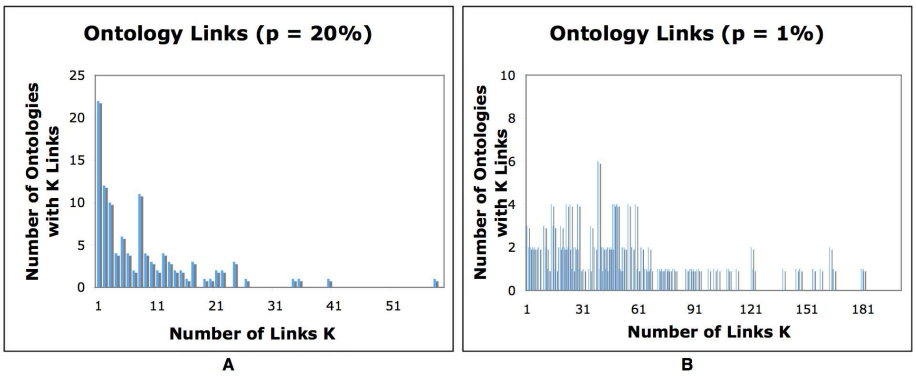

Fig. 1. Number of links between ontologies for(a) $p = 20\%$ **and (b)** $p = 1\%$**:** The x-axis represents a number of links to other ontologies that each ontology has. The y-axis represents the number of ontologies with that number of links. The graph demonstrates the power-law distribution for $p = 20\%$: there is a small number of ontologies that have a large number of links (the hubs) and a large number of ontologies with just a few links. If we use $p = 1\%$ (there is a link from one ontology to another if at least 1% of its concepts are mapped), the distribution becomes essentially random.

represents a small world: A network exhibits the small world property if any one node in the network is never more than a few hops away from any other [2]. For $p = 10\%$, the average distance between any two nodes is only 1.1 hops.

3.2 Hubs and clusters

We constructed directed graphs of ontologies at several different thresholds of p (Figure 2). Nodes in the graphs are ontologies. There is a directed edge from a node corresponding to the ontology S to a node corresponding to an ontology T iff there is a link $L_p(S, T)$ between these ontologies for this values of p. We say that a particular node is a **hub** in this graph if it has more than twice the number of links (incoming and outgoing) than the average number of links for nodes in the graph. A set of connected ontologies forms a **cluster**.

We used these graphs to identify connections between different ontologies, based on varying levels of overlap. The graphs identified clear hubs, ontologies to which many other ontologies link, with the Systematized Nomenclature of Medicine - Clinical Terms (SNOMED-CT) ontology being the most prominent hub: We found that SNOMED-CT had the most links of any ontology, with 58 links (57 as "target", 1 as "source") at p = 20%. In other words, 58 out of 207 had at least 20% of their concepts mapped to concepts to SNOMED-CT. On further analysis, we found that more than 85% of SNOMED-CT concepts are mapped to at least one other concept in our repository.

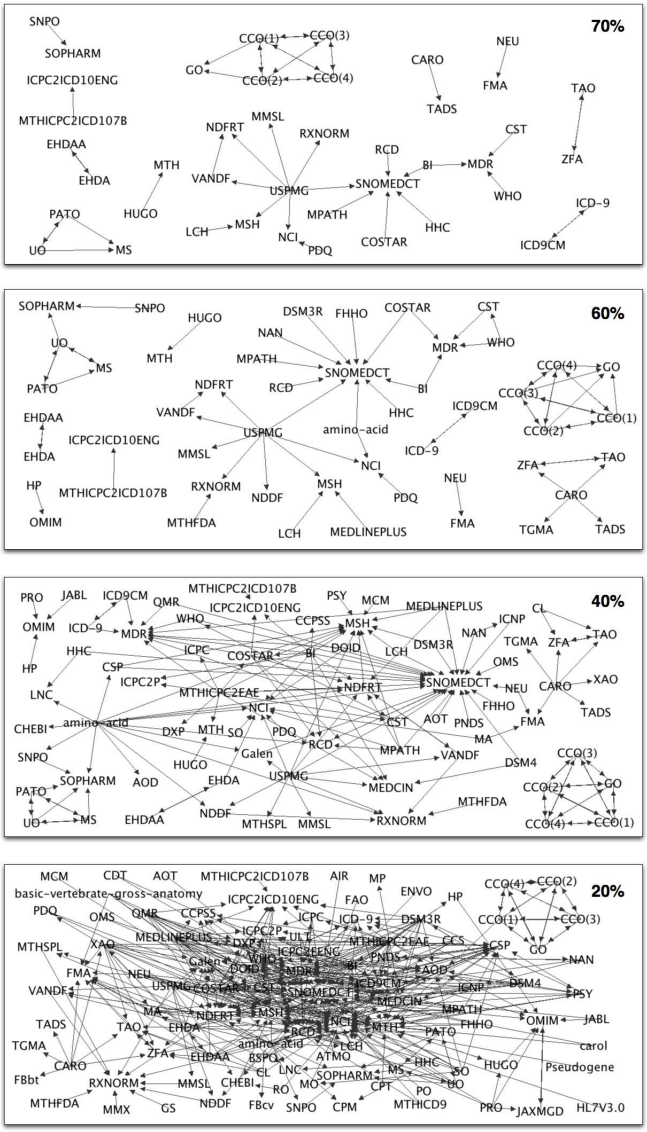

Fig. 2. The graphs show percent-normalized links between ontologies that are true for p = 20%, 40%, 60%, and 70%. Nodes represent ontologies in the repositories. Please refer to http://bioportal.bioontology.org/ontologies for the list of full ontology names corresponding to the acronyms that we use as labels in the graph. An edge from a node representing on ontology O_1 to a node representing an ontology O_2 means that at least p% of concepts from O_1 map to some concept in O_2.

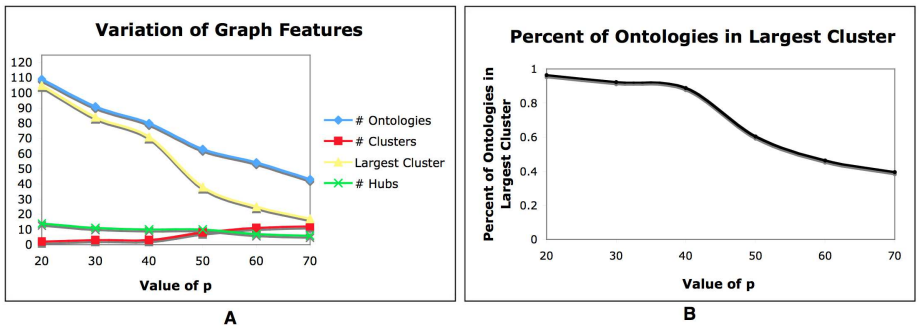

Fig. 3. Variation of graph features as p changes: (a) The graph shows how the number of ontologies, clusters, hubs, and the size of the largest cluster, all on the y-axis, change as p changes. As p decreases (from right to left), the number of clusters decreases slowly as clusters merge and the number of hubs increases slowly. Additionally, the largest cluster increases to include a larger fraction of the ontologies, which also increase in number because more ontologies are connected as the threshold for p decreases. (b) The graph shows the percent of ontologies that are in the largest cluster for different thresholds of p. As p decreases (right to left), almost all ontologies become connected very quickly.

Figure 3a shows variation of some key graph features (number of hubs, number of ontologies, number of clusters, and size of the largest cluster) as we change the value of p. Additionally, the plot in Figure 3b displays the percent of ontologies in the largest cluster.

Note that the graphs have two distinct types of hubs: (1) hubs in which almost all of the links (directed edges) were incoming and (2) hubs in which almost all of the directed edges were outgoing. That is, some, usually small, ontologies have a large fraction of their concepts covered by other ontologies. Other, usually larger, ontologies include concepts that are similar to a large fraction of concepts in many smaller ontologies in the repository.

3.3 How similar are the ontologies?

Finally, Figure 4 shows a graph of the percentage of overall links that are included at each threshold value of p. This graph demonstrates that:

- 96% of ontology links are between ontologies that are less than 20% similar
- 91% between ontologies less than 10% similar
- 65% between ontologies less than 1% similar

And our final observation: Out of 207 ontologies, 68 ontologies (or 33%) have at least 50% of their terms mapped to terms in some other ontology.

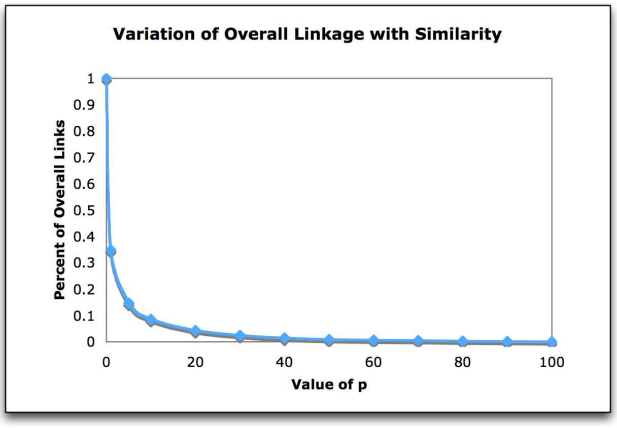

Fig. 4. Percentage of overall links: The x-axis is the value of p. The y-axis shows the fraction of all links between ontologies (L_0) that are present for a given value of p (i.e., L_p / L_0) as p changes

4 Discussion and Analysis

The figures and data that we presented in Section 3 allow us to make several observations and answer several of the questions that we posed at the beginning of the paper.

To what degree are the domains covered by different ontologies connected?

- If we use lexical mappings as the basis for determining how connected the ontologies are, then the biomedical ontologies in our repository are very closely connected, with 33% of them having at least half of their concepts mapped to concepts in other ontologies.
- With such a large overlap among the ontologies, attempts to find "canonical" representations for concepts may be doomed: a large number of concepts in biomedicine are already represented in many different ways. One could argue that our data shows that given the state of the biomedical ontologies today, the most feasible way of integrating ontologies is by creating mappings, possibly complex ones, between them rather than trying to eliminate overlap (as the OBO Foundry initiative is trying to do [14]).
- Our study found that the small world property holds on our set of ontologies for low values of p, with the average distance between the nodes being as low as 1.1 hops. Other research has found the small world property to be true for other large data sets as well[8]. Further analysis on the properties of small world networks, such as strength of ties and k-core among others, may provide additional useful insight about the connectedness of ontologies.

If you are new to a domain, what are the important or representative ontologies with good coverage?

- The lexical mappings identified SNOMED-CT as the most prominent hub, and, indeed, SNOMED-CT is the largest and one of the most prominent and popular biomedical ontologies. Thus if we use mappings as a way of identifying prominent ontologies in a domain (i.e., an ontology with lots of mappings to other ontologies is an "important" one), then at least in this case, this approach would have identified correctly the ontology that a newcomer to the domain of biomedical ontologies must become familiar with.
- Hubs with many outgoing links show shared domains, particularly at high threshold values for p. For these hub ontologies, a large portion of their concepts is mapped to several different ontologies. Thus, ontologies that are linked through such a hub likely share the content that is represented in the hub ontology. For example, at p= 50%, the Common Anatomy Reference Ontology (CARO) is a hub with outgoing links to Foundational Model of Anatomy, Zebrafish anatomy and development, Tick gross anatomy, Teleost anatomy and development, and Mosquito gross anatomy—all ontologies in the anatomy domain. At p= 70%, the United States Pharmacopeia Model Guidelines ontology (USPMG) has outgoing links to Multum MediSource Lexicon, RxNorm Vocabulary, Veterans Health Administration National Drug File, National Drug File Reference Terminology, Medical Subject Headings, National Cancer Institute Thesaurus, and SNOMED-CT—all ontologies that describe drugs, among other things.

If you want to build domain-specific tools for creating ontology mappings, what are good ontologies to use for background knowledge?

- The two previous points lead to several practical uses of hubs identified through mappings: First, for ontology-mapping algorithms that require domain-specific background knowledge, hubs with many incoming links (such as SNOMED-CT) can serve as useful sources of such background knowledge. Second, these hubs are also good candidates for being representative ontologies for a domain.

What can we learn about the characteristics of the ontologies themselves and the ontology repository from the mappings between them?

- Links at a low value of p (1%) (i.e., when less than 1% of the concepts from the source ontology have a mapping to the target) do not say much about connectedness of ontologies. The domain of biomedicine is such that there is a little bit of overlap in everything, resulting in the extremely connected model we see at 1% mark. At 20%, however, we see a meaningful power-law distribution. At even higher thresholds, we can see ontologies that are very closely related. For example, we see that the Gene Ontology (GO) is very closely related to the cell cycle ontologies (CCO). 65% of links fall in a range lower than 1% similarity, which indicates that links below the p = 1% threshold are not as informative of connections between ontologies.
- If we were to use mappings between terms (identified in any way) as an indication of distance or similarity between ontologies in a repository, then the previous observation leads to the following practical implication: These links at low values of

p are not very meaningful and should probably not be used as an indication of any relation between the ontologies.

5 Conclusions, Limitations, and Future Work

Our analysis does not depend on the specific method used to identify mappings between concepts. We used a simple method because it worked well and was very scalable (cf Section 2.2). Our earlier research has found that most of the openly available advanced mapping algorithms are simply not scalable to the size of biomedical ontologies in our repository [6]. One of the interesting directions for future work, when more scalable advanced algorithms become available, would be to perform similar analysis of relationships between ontologies taking a more advanced set of mappings as input.

Our main contribution in this paper is not the method for generating mappings between ontologies, but rather the analysis of these mappings. We believe that network analysis serves as a powerful tool for analyzing the structure of an ontology repository by providing insights into the characteristics of the ontologies and the structure of the repository. As the Semantic Web grows in popularity and use of ontologies expands, these methods may play an important role in understanding the connections among ontologies.

Our approach has certain critical limitations. Because we use a simple lexical matching method, our results are limited to the domain of biomedicine and other domains where such mapping method works well. In other domains, where concept definitions do not contain rich lexical information in the form of preferred names and synonyms, one will need to find scalable tools that would produce a large number of mappings that enable statistically significant analysis. Also, because of the way we generate the mappings, we do not account for the ontologies that have alternate lexical structures to represent the same concepts. Thus, we may miss a connection between two ontologies that actually have a significant amount of overlap in terms of the actual concepts they represent simply because these concepts have different lexical structures in the two ontologies. Our methods would work best for sets of ontologies that use similar naming conventions [13].

Another limitation of our work is that it gives no information as to the nature of the mappings or the actual specific relationship between linked ontologies. We cannot know whether one ontology simply has the same concept names as another ontology or if it imports terms from that ontology directly. Many BioPortal ontologies use OBO format and often simply copy the ontology that they import rather than use the import mechanism that recently has become available in OBO. As a result, a number of the mappings that we created are not actually mappings in the traditional sense. We plan to develop heuristics to identify this overlap and exclude it from the mappings set.

As part of our future work, we plan to compare our lexical mappings on the set of UMLS terminologies to the set of mappings provided by UMLS. The UMLS has a large number of mappings between the terminologies that were created manually. While the purpose of those mappings was slightly different from ours, comparing the results of the lexical mapping to the manual one will likely produce useful insights.

We also plan to implement automatic maintenance and updates on the set of mappings generated with this method. Not only does the number of BioPortal ontologies increase regularly, but also new versions of some of the ontologies are uploaded every night. These frequent updates makes manual creation of mappings among ontologies in the repository a daunting, if not impossible, task. By contrast, UMLS terminologies update twice a year.

In addition to the work outlined above, we plan to perform similar analysis on the mappings from other sources if the number of mappings is significant enough to make such analysis valid. For instance, the Alignment Server [5] in the NeON project could prove to be one such source of mappings in the future.

Finally, we uploaded the set of mappings between BioPortal ontologies to BioPortal. Users can browse the mappings through the BioPortal user interface and access them programmatically through REST services. We believe that the mappings should prove useful to developers of tools dependent on these mappings or for ontology developers looking at specific domain ontologies. All the data that we used for the analysis in this paper is available in raw spreadsheet form at `http://www.bioontology.org/wiki/index.php/Mapping_Set`.

The study that we reported in this paper offered the first glimpse at the possibilities and challenges that large numbers of related ontologies bring to the fore. Our results show that using network analysis on a network defined by mappings between ontology terms helps us understand and navigate a world with a large number of ontologies. As more ontologies become available on the Semantic Web, such analysis will become more interesting, more useful, and more challenging.

Acknowledgements

This work was supported by the National Center for Biomedical Ontology, under roadmap-initiative grant U54 HG004028 from the National Institutes of Health. Nick Griffith has implemented the support for mappings in BioPortal.

References

1. Z. Aleksovski, M. Klein, W. ten Kate, and F. van Harmelen. Matching unstructured vocabularies using a background ontology. In *15th International Conference on Knowledge Engineering and Knowledge Management (EKAW'06)*, 2006.
2. A.-L. Barabsi. *Linked: How Everything Is Connected to Everything Else and What It Means for Business, Science, and Everyday Life*. Basic Books, 2003.
3. O. Bodenreider and R. Stevens. Bio-ontologies: current trends and future directions. *Briefings in Bioinformatics*, 7:256–274, 2006.
4. C. Caracciolo, J. Euzenat, L. Hollink, R. Ichise, A. Isaac, V. Malaise, C. Meilicke, J. Pane, P. Shvaiko, H. Stuckenschmidt, O. Svab Zamazal, and V. Svatek. Results of the ontology alignment evaluation initiative 2008. In *3d International Workshop on Ontology Matching (OM-2008) at ISWC 2008*, Karlsruhe, Germany, 2008.
5. J. Euzenat. Alignment infrastructure for ontology mediation and other applications. In *Workshop on Mediation in Semantic Web Services*, 2005.

6. A. Ghazvinian, N. F. Noy, and M. A. Musen. Creating mappings for ontologies in biomedicine: Simple methods work. In *AMIA Annual Symposium (AMIA 2009)*, San Francisco, CA, 2009.
7. C. Jonquet, N. H. Shah, and M. A. Musen. The open biomedical annotator. In *AMIA Summit on Translational Bioinformatics*, pages 56–60, San Francisco, CA, USA, 2009.
8. J. Leskovec and E. Horvitz. Planetary-scale views on a large instant-messaging network. In *17th International World Wide Web Conference (WWW2008)*, Beijing, China, 2008.
9. D. Lindberg, B. Humphreys, and A. McCray. The unified medical language system. *Methods of Information in Medicine*, 32(4):281, 1993.
10. N. Noy, N. Shah, B. Dai, M. Dorf, N. Griffith, C. Jonquet, M. Montegut, D. Rubin, C. Youn, and M. Musen. Bioportal: A web repository for biomedical ontologies and data resources. In *Demo session at 7th International Semantic Web Conference (ISWC 2008)*, Karlsruhe, Germany, 2008. Springer.
11. N. F. Noy, N. Griffith, and M. A. Musen. Collecting community-based mappings in an ontology repository. In *7th International Semantic Web Conference (ISWC 2008)*, Karlsruhe, Germany, 2008.
12. D. L. Rubin, N. H. Shah, and N. F. Noy. Biomedical ontologies: a functional perspective. *Briefings in Bioinformatics*, 9(1):75–90, 2008.
13. D. Schober, B. Smith, S. E. Lewis, W. Kusnierczyk, J. Lomax, C. Mungall, C. F. Taylor, P. Rocca-Serra, and S.-A. Sansone. Survey-based naming conventions for use in obo foundry ontology development. *BMC Bioinformatics*, 2009.
14. B. Smith, M. Ashburner, C. Rosse, J. Bard, W. Bug, W. Ceusters, L. J. Goldberg, K. Eilbeck, A. Ireland, C. J. Mungall, N. Leontis, P. Rocca-Serra, A. Ruttenberg, S. A. Sansone, R. H. Scheuermann, N. Shah, P. L. Whetzel, and S. Lewis. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology*, 25(11):1251–5, 2007.
15. S. Zhang and O. Bodenreider. Alignment of multiple ontologies of anatomy: Deriving indirect mappings from direct mappings to a reference. In *AMIA Annual Symposium*, pages 864–868, 2005.