Commonsense knowledge bases and network analysis

Tanya Berger-Wolf and **Dimitrios I. Diochnos**
University of Illinois at Chicago

András London and **András Pluhár**
University of Szeged

Robert H. Sloan
University of Illinois at Chicago

György Turán
University of Illinois at Chicago,
University of Szeged

Abstract

We explore the ConceptNet 4 common-sense knowledge base using the tools of *network analysis*. This exploration is interesting in its own right, but a primary goal is developing methods that are potentially useful for improving the performance of the knowledge base on various common-sense reasoning tasks. ConceptNet has been used by its creators and others for various tasks, and ConceptNet 4 has recently been shown to perform well on the verbal portion of an IQ test for young children, using simple procedures to answer test questions.

The network analysis of ConceptNet 4 is a task of interest in itself, as the knowledge base is an example of a large network with ground truth provided by the meaning of concepts. We give several examples of the results of community finding algorithms, spreading activation and rule mining. These results can be useful for finding missing and incorrect links, and for building a background knowledge that could be used to provide additional forms of reasoning in the knowledge base.

Introduction

Enabling computers to do common-sense reasoning is one of the basic challenges in AI, implicit in the work of Turing, and made explicit by McCarthy (McCarthy 1959). The availability of large amounts of common-sense knowledge is widely accepted to be a necessary condition for such reasoning. It is an important and relatively recent development that large common-sense knowledge bases such as Cyc and ConceptNet are publicly available. This availability, combined with potential new applications in web search, robotics, human-computer interaction and other areas, has led to an increased interest in common-sense reasoning.

ConceptNet, on which we focus in this paper, is a semantic net of triples of the form (`concept`$_1$, `relation`, `concept`$_2$), with every `relation` coming from a fixed set of about two dozen relations, such as `IsA` and `Causes` (Havasi, Speer, and Alonso 2007; Havasi et al. 2009). Its data was initially collected via the web in the form of sentences, and turned into statements using NLP tools. The triples are represented as a sparse matrix with concepts as rows and relation-concept pairs as columns. Low-rank approximations of a condensed version of the matrix, called AnalogySpace, are also available (Speer, Havasi, and Lieberman 2008). Thus, ConceptNet allows for *both symbolic and statistical reasoning*.

ConceptNet has been used for various applications (e.g., (Shen, Lieberman, and Lam 2007)), including query answering (Kotov and Zhai 2012). Moreover, ConceptNet was recently evaluated for its ability in *answering IQ-tests for children*, which was proposed as a general evaluation method for common-sense knowledge bases. ConceptNet's verbal IQ corresponded to an average 4-year old (Ohlsson et al. 2012; 2013). The algorithms used for answering the IQ-test items were quite simple; that work was also intended to evaluate the ease of use of the system.

Thus, the ConceptNet system is an example of a large knowledge base that has had at least some success in some common-sense tasks, and is a suitable target for studying properties of large common-sense knowledge bases.

Network Analysis of Knowledge Bases

Here we explore the properties of ConceptNet using the tools of *network analysis* (Brandes and Erlebach 2005; Easley and Kleinberg 2010; Newman 2003). Social, collaboration, and information networks are major well-studied classes of networks. Network analysis has been applied also to biological, infrastructure, and transportation networks. *Knowledge networks*, such as word association networks and WordNet have also been studied to some extent (De Deyne and Storms 2008; Steyvers and Tenenbaum 2005). The recent work of (Harrington and Clark 2007) on automated generation is a potential source of many new networks.

To date, however, there appears to be no general understanding of the characteristics of large knowledge networks. The availability of *ground truth* in common-sense knowledge networks, provided by the meaning of concepts, is an unusual, useful feature for network analysis. Ground truth, for example, allows one to evaluate and compare the quality of groupings found by various community inference algorithms. For example, the reader can immediately see that the community of concepts in Figure 4 (inferred by the clique percolation algorithm (Derényi, Palla, and Vicsek 2005)) is meaningful, while in a social network it is typically difficult to tell whether a group of people is, in fact, a community.

A detailed network analysis of ConceptNet is given in (Diochnos 2013). Here we give a small sample of the results. There are many different ConceptNet-based networks to consider (directed/undirected edges, multiple edges and loops allowed or not, all relations are considered or just a specific subset of them). In general, those networks have a highly skewed degree distribution and the small world property[1]. Cores form a nested structure of increasing density, similar to other large networks (Leskovec et al. 2009). It is likely that an inner core contains more important concepts, and those could perhaps be given closer attention for additional processing. Among the many community finding algorithms, clique percolation (Derényi, Palla, and Vicsek 2005) for appropriate clique sizes seems to give interesting communities. These communities can be useful for finding missing entries and identifying new concepts. The seminal paper (Steyvers and Tenenbaum 2005) proposed cognitive science applications of network information for semantic networks, such as relevance for the age of acquisition of concepts. Cores, communities and other specific structures found by network analysis could be of interest in this context as well. There is theoretical computer science work on exploiting structural properties to get faster algorithms for problems which are hard for large networks (see, e.g., (Gonen et al. 2008)). Such problems include versions of centrality which may also be relevant for cognitive science (De Deyne and Storms 2008).

Potential Benefits for Common-Sense Reasoning

We believe the network analysis of ConceptNet and other such networks has potential benefits for common-sense reasoning applications of ConceptNet. The most difficult question type for ConceptNet in the IQ testing was Comprehension, which tests the comprehension of concepts using *why*-questions like Why do we put on sunscreen in summer? Answering WPPSI-III Comprehension questions has overlap with the area of open-domain question answering, of *Jeopardy!* fame, which involves information retrieval, natural language processing and human-computer interaction (Maybury 2004). In general, however, knowledge representation and reasoning are often weak spots for question answering (Balduccini, Baral, and Lierler 2008, p. 780), and those abilities seem to be absolutely necessary to answer questions like the one about sunscreen. Incidentally, answering specifically *why*-questions is considered a difficult task (Verberne et al. 2010).

Answering *why*-questions with ConceptNet remains an important and interesting challenge and it serves as one motivation for the explorations described in this paper. Improving the results and being able to answer test questions for older children is likely to require using *more involved test-answering algorithms* and *improving and enhancing the knowledge base*, for example, by adding missing entries, correcting incorrect entries and providing additional knowledge. Additional knowledge could include additional facts, but also new general knowledge, and capabilities for doing different forms of common-sense reasoning. These issues are also discussed in the papers on ConceptNet (Havasi, Speer, and Alonso 2007; Havasi et al. 2009; Speer, Havasi, and Lieberman 2008); here we propose some further approaches.

ConceptNet provides spreading activation as a tool to find semantically related concepts. This, in turn, can be used as a tool for question answering. The answers obtained using spreading activation are often meaningful, especially if one considers not only the highest ranked answer, but also the best answer among the highest ranked ones. This suggests refined search procedures, where one analyzes the detailed results of spreading activation to rank candidate answers and to identify errors.

Moreover, ConceptNet provides a rich body of knowledge about similarity, ontologies, causality and other notions. It would be useful to enhance this body of knowledge adding further reasoning tools. As a first step, one could build a 'microtheory' of the relations used, for example, that IsA is transitive: (a, IsA, b), (b, IsA, c) $\rightarrow$ (a, IsA, c). As there are a large number of possible rules, one could try to mine ConceptNet for such rules. We mention some results of rule mining and give observations for possible applications.

ConceptNet 4

In this document by "ConceptNet" (Havasi, Speer, and Alonso 2007; Havasi et al. 2009; Speer, Havasi, and Lieberman 2008) we mean specifically the version of ConceptNet 4 released in March 2012. In fact, there are two versions of ConceptNet. One, which we call the *large graph*, contains roughly 280,000 English-language concepts, and is released in SQLite database format. The other, which we call the *small graph*, contains roughly 22,000 concepts, and is released as part of a Python package called Divisi, which in general is "a general-purpose tool for reasoning over semantic networks" (`http://csc.media.mit.edu/analogyspace`) and working with large sparse matrices (Speer, Arnold, and Havasi 2010). When unspecified, in this paper we refer to the large graph. Divisi also contains tools for creating truncated singular value decomposition (SVD) forms of the small graph, which its authors refer to as AnalogySpace. The work on IQ-testing ConceptNet (Ohlsson et al. 2012; 2013) was done primarily with AnalogySpace and made no use of the large graph.

Both the large graph and the small graph are sparse, with the large graph being considerably sparser. The small graph was formed from the large graph by dropping some combination of triples that had relatively few users supporting them and concepts that had very little connectivity to the rest of the graph. (The AnalogySpace graph, which we do not discuss much in this paper, is dense.)

ConceptNet triples are called *assertions*. Each assertion also has a *score, frequency*, and *polarity*. The score measures the reliability of an assertion, based on the amount of user support it received. Frequency expresses how often the assertion is true, in the range of "never" to "always". Polarity is a coarse-grained version of the frequency and is positive or negative. For example the statement Penguins are not capable of flying has negative polarity. Roughly 3.5% of assertions have negative polarity. Associated with each assertion there is at most one *sentence* and *raw assertion*. The sentence is actual user input that generated or supported the

[1]A graph has the *small-world property* if the distance between two randomly chosen nodes is small, typically logarithmic in the number of nodes.

assertion, and the raw assertion is a lightly processed sentence put into one of a large number of standard frames.

Network analysis of ConceptNet 4

The prevalence of large networks such as the Web, the internet and online social networks, has led to the explosive growth of research and the development of computational approaches for network analysis and algorithms for large networks, with central concepts such as highly skewed node degree distribution, small world property (Watts and Strogatz 1998) and algorithms like PageRank (Page et al. 1999). The main insight gained is that, perhaps surprisingly, networks coming from completely different disciplines have quite similar structural properties. In this section we apply this methodology to ConceptNet with an eye toward exploiting its properties for knowledge base algorithms. We expect similar properties to hold for future versions of ConceptNet and other knowledge bases as well. We use `igraph` (Csardi and Nepusz 2006) for most of the network analysis tasks, `CFinder` (Palla et al. 2005) for computing communities by percolating cliques, and the software that is available online (`http://tuvalu.santafe.edu/~aaronc/powerlaws/`) for the *maximum likelihood estimate (MLE)* for power law fitting described in (Clauset, Shalizi, and Newman 2009).

Among the many possibilities for viewing ConceptNet as a network, for *degree distribution* we consider the directed multigraph with self-loops formed by assertions with a positive score (and arbitrary polarity). There are $279,497$ concepts appearing in such assertions in the English language version. Figure 1 presents the degree distributions of both the large and small graphs in a log-log plot. The network has a highly skewed node degree distribution, as is the case in pretty much all other networks. Applying the MLE for power law fit we obtain 1.82572 and 1.90602 respectively for the exponents. However, the quality of the fit is poor; see (Diochnos 2013, Chapter 4) for details. The average degree of the large graph is about 3.5. The induced directed and undirected graphs in this case have both average degrees about 3.0. The average degree of the small graph is about 40.7 while the induced directed and undirected graphs have respectively average degrees of about 16.0 and 15.1.

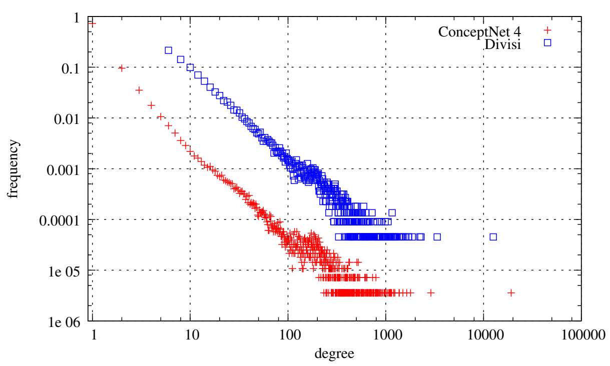

Figure 1: Total degree distribution in ConceptNet 4 and Divisi.

Large networks typically have a *giant component*. For the directed graph induced by assertions with any polarity, there is a giant component with $228,784$ vertices, and the remaining $32,701$ connected components have size at most 55, including $16,922$ singletons. For *strongly connected components*, there is a giant component with $14,025$ vertices, and the remaining $265,373$ components have size at most 3, including $265,276$ singletons.

The maximal distance in the undirected graph induced by assertions with both polarities is 16. The pair returned by `igraph` with distance 16 is `anti-charm quark` and `double-breasted de fursac jacket`. The average distance in the giant component is 4.28. Thus the graph exhibits a small-world property. Details for the distances are given in (Diochnos 2013).

The k-*core* of a graph is obtained by iteratively removing vertices of degree less than k while such vertices exist (Seidman 1983). In each step there may be several choices, but it turns out that the final result is independent of those choices. The *maximum coreness* of a graph is the largest k for which the k-core is nonempty. The maximum coreness of the graph induced by the assertions with positive polarity is 26 and there are 869 concepts belonging to that core. Table 1 gives data on the core structure.

Now we turn to cliques, i.e., complete subgraphs, which are the strongest possible form of community. There are $107,100$ cliques with positive polarity, out of which there is only one clique of size 12, composed of the concepts `person`, `build`, `house`, `home`, `apartment`, `room`, `live room`, `couch`, `table`, `chair`, `cat`, and `dog`. [2] There are also 23 cliques of size 11. It turns out that all these 24 cliques are created from 36 concepts. Table 2 shows some of those cliques. Examining the *overlap* of cliques is also interesting as it can uncover different meanings of a concept (see also (Palla et al. 2005)) and other useful relationships. Figure 2 gives an example of two overlapping communities corresponding to different meanings of the concept `cut`.

For *community-finding*, the *clique-percolation* algorithm (Derényi, Palla, and Vicsek 2005; Palla et al. 2005) produced interesting results. Let S be a k-clique. Clique percolation with parameter k builds a community starting from clique S and taking the union of all cliques reachable by k-chains from S, where a k-chain is a sequence of k-cliques such that each clique has $k-1$ vertices in common with the previous one. We found the following communities: 362 using triangles, 290 using K_4's (cliques of size 4), 287 using K_5's, 209 using K_6's, 120 using K_7's, 84 using K_8's, 16 using K_9's, 12 using K_{10}'s, 6 using K_{11}'s, and of course one community by percolating K_{12}'s. Figures 3 and 4 present communities that occur by percolating cliques of various sizes. Communities could be presented to the user for suggesting a new concept or link. For example, the concept `dishonesty` could be suggested for the community shown in Figure 3, resulting in the addition of new assertions. Figure 4 already contains `religion`, but

[2] The interpretation (*surface form*) of ConceptNet's `live room` is `living room`, or `in a living room`, etc., and of `build` is `a building`.

Table 1: Number of vertices and average degree of undirected subgraphs; positive polarity only, self-loops are neglected.

coreness	$\geqslant 0$	$\geqslant 2$	$\geqslant 5$	$\geqslant 8$	$\geqslant 11$	$\geqslant 14$	$\geqslant 17$	$\geqslant 20$	$\geqslant 23$	$\geqslant 26$
vertices	$279,497$	$41,659$	$11,483$	$6,750$	$4,634$	$3,407$	$2,617$	$2,007$	$1,514$	869
avg. degree	2.872	9.682	22.421	30.093	35.839	40.278	43.515	45.984	47.384	47.241

Table 2: Concepts participating in maximal cliques with positive polarity and frequency in the range $\{5, \ldots, 10\}$. The cliques are obtained from English-language assertions with positive score. The first clique has size 12. Among all cliques of size 11 or 12 we show those where the concept `apartment` appears.

	clique								
concept	1	2	3	4	5	6	7	8	✓
apartment	✓	✓	✓	✓	✓	✓	✓	✓	8
bed		✓							1
bedroom			✓	✓	✓				3
build	✓	✓	✓	✓	✓	✓	✓	✓	8
cat	✓	✓	✓	✓		✓			5
chair	✓		✓	✓	✓	✓	✓		6
city						✓	✓	✓	3
couch	✓	✓							2
dog	✓		✓	✓	✓	✓	✓	✓	7
home	✓	✓	✓	✓	✓	✓	✓	✓	8
house	✓	✓	✓	✓	✓	✓	✓	✓	8
human					✓		✓	✓	3
live room	✓	✓							2
person	✓	✓	✓	✓	✓	✓	✓	✓	8
room	✓	✓	✓	✓	✓	✓	✓	✓	8
table	✓	✓	✓						3
town				✓	✓	✓	✓	✓	5

the user might suggest the addition of an assertion involving `belief` and `prayer`.

Spreading activation

Spreading activation is a technique inspired by neural models, for identifying related nodes (Collins and Loftus 1975), used for example, in information retrieval (Crestani 1997). It is related to PageRank (Page et al. 1999) and other similar algorithms. Recently, it has also been used in knowledge network acquisition (Harrington and Clark 2007). We illustrate the application of spreading activation for query answering in the case of Comprehension queries, using a variant of Harrington's approach. Refined versions of such algorithms may be useful for improving the quality of the answers obtained.

Spreading activation can be started by activating several concepts which simultaneously spread activation values in rounds to their neighbors. The nodes also propagate their labels to neighboring nodes. The firing thresholds of the nodes and the decay factors are parameters of the process. We implemented different versions depending on the type of the underlying graph, the firing regime and the termination criterion. After the activation process is terminated, we find paths with a significant amount of activation; again, we implemented different algorithms for finding such paths. One may then start a second round of spreading activation from

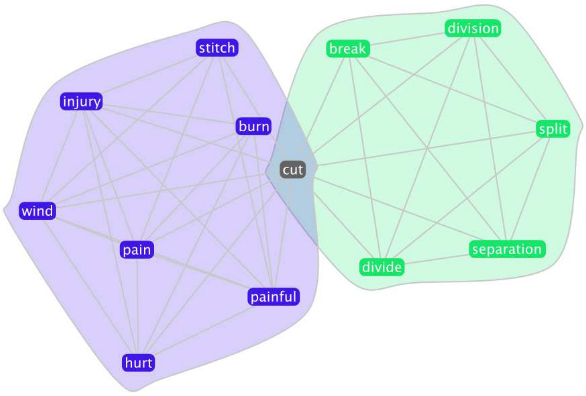

Figure 2: Overlapping communities for different meanings of `cut`.

nodes on the significant paths, and finally look for assertions with the highest levels of activation along those paths.

We illustrate the application of spreading activation on the question Why do we put on sunscreen is summer? This turned out to be a difficult question for ConceptNet. More precisely, it turned out to be a difficult question *for the AnalogySpace-based algorithm* used in (Ohlsson et al. 2012). The answers received included `UsedFor/cook` and `Causes/strike match` with large weight. As we will see below, ConceptNet in fact contains sufficient information to answer the question correctly, the problem is 'only' how to find it. We also get an answer to the unexpected appearance of `cook`.

Running spreading activation from the concepts `put sunscreen` and `summer`, the first phase activates $6,700$ concepts and the three intermediate nodes where both labels appear are `heat`, `hot`, and `fall`. The undirected primary paths, involving 7 different nodes, are:

- `put sunscreen` — `go swim` — `heat` — `summer`,
- `put sunscreen` — `go swim` — `hot` — `summer`,
- `put sunscreen` — `go fish` — `fall` — `summer`.

The top ten most activated nodes in the network in the first round are, in that order, `summer`, `put sunscreen`, `heat`, `season`, `hot`, `winter`, `hot weather`, `after spring`, `spring`, and `camp`. The top ten most activated nodes in the network after the second round are, in that order, `summer`, `heat`, `put sunscreen`, `hot`, `fall`, `go swim`, `go fish`, `fire`, `person`, and `winter`.

The top ten activated pairs of concepts[3] after the first round involve the concepts `summer`, `heat`, `season`, `hot`,

[3]A pair of concepts typically involves several assertions due to different relations connecting the concepts.

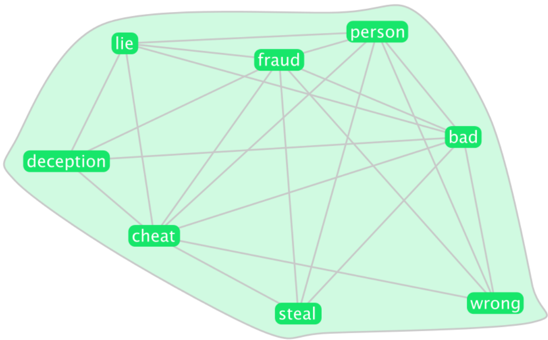

Figure 3: Community 'dishonest/dishonesty'. Eight nodes by percolating cliques of size 5; dishonest/dishonesty itself is missing from the community.

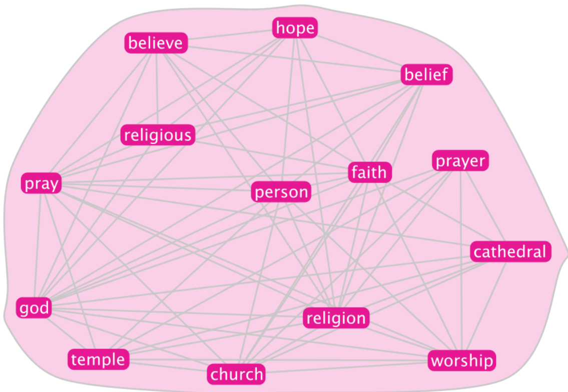

Figure 4: Community 'religion'. Fourteen nodes by percolating cliques of size 7; missing link between belief and prayer.

winter, hot weather, after spring, spring, camp, warm season, and hot month. After the second round, the concepts go swim, go fish, and put sunscreen appear in the corresponding list. In particular, the assertion (go swim, HasPrerequisite, put sunscreen) ranks in the 30th place.

Now we make some more observations on specific relations, which are relevant in the context of 'why' questions. Let us start with the relation HasPrerequisite. In both rounds the top two most activated assertions with this relation connect the pairs of concepts (go swim, put sunscreen) and (go fish, put sunscreen).

Regarding the relation CausesDesire, at the end of the first round the top three assertions in this relation connect the concepts (summer, play baseball), (summer, fish), and (summer, go walk). After the second round, the assertions (heat, CausesDesire, go swim) and (hot, CausesDesire, go swim) move to the top two positions, from positions four and five for this relation. So, we can build a slightly better justification since heat causes desire to go swimming, and going for swimming has as prerequisite to put on sunscreen.

Finally, for the Causes relation, in both rounds the top three assertions for this relation connect the pair of concepts (heat, fire), (fire, heat), and (sun, heat).

Examining other relations shows how some of the incorrect answers received in (Ohlsson et al. 2012) are caused by multiple meanings. For example, in the first round of the spreading activation process, the top assertion for the relation HasLastSubevent connects cook meal and season, while in the end of the second round it connects climb to fall. Apparently, the problem arises due to a lack of disambiguation for the concepts season and fall, both of which should be used here as three-month periods. However, season is used as the act of putting seasonings while cooking meals and fall is used as the verb *"to fall"*. Moreover, the idea of sunscreen in the summer typically activates nodes that are related to heat and water, which in combination with seasonings further justifies why cooking meals appears. Finally, the problem remains in the second round but due to fall that appears along a primary path.

Rule mining

In this section we discuss the application of data mining towards the automated construction of a background theory for the relations used in the knowledge base. We consider rules of the simplest form, mainly for computational considerations.

A *rule* is given by an ordered triple of relations (X, Y, Z), where X, Y are the *premisses* and Z is the *conclusion*. For such a triple we consider triples of concepts (a, b, c) such that the assertions

(a, X, b) and (b, Y, c)

are in the knowledge base. Such triples form the *support* of the rule. If (a, Z, c) is also in the knowledge base then (a, b, c) is a *success* for the rule (X, Y, Z), otherwise it is a *failure*. The *success rate* of a rule is the percentage of successes in the support. Consider, for example, the rule (Desires, LocatedNear, AtLocation) and the triple of concepts (human, drink, bar). The assertions (human, Desires, drink) and (drink, LocatedNear, bar) are both in the knowledge base. Therefore, we check whether the assertion (human, AtLocation, bar) is in the knowledge base. It is, so (human, drink, bar) is a success for the rule (Desires, LocatedNear, AtLocation).

A triple of concepts (a, b, c) is *valid* for a rule (X, Y, Z) if the claim

(a, X, b) and (b, Y, c) therefore (a, Z, c)

makes sense as a reasoning step. Otherwise (a, b, c) is *invalid. Making sense is a subjective judgement* and its intended meaning is up for discussion. In what follows we use the sense "given that the premisses hold it is reasonable to assume that the conclusion holds". For example, (human, drink, bar) is valid for the rule (Desires, LocatedNear, AtLocation). Note that by the nature of its definition, deciding about validity requires an (often ambiguous) decision by a human and so computing precise statistics about it is difficult.

We performed an exhaustive test for all possible rules involving relations that have at least 300 assertions with positive score regardless of their polarity. We searched for *frequent* rules, with support at least 300 and success rate at least 5% [4]. Success rates are expected to be low even for correct rules due to the sparsity of the network. There are 76 such triples of relations. We give examples of some such relations, plus an interesting one with low success rate, and comment on issues raised by these examples.

Our first example is the rule (`Desires`, `LocatedNear`, `AtLocation`). This is the highest scoring rule with 251 successes and support 2050 (12% success rate). The triples (`human`, `drink`, `bar`) and (`bird`, `seed`, `garden`) are successful and valid. The triple (`human`, `love`, `heart`) is successful but invalid. The triple (`bird`, `seed`, `plant garden`) is a failure but it is valid. The reason for the failure is that the assertion (`bird`, `AtLocation`, `plant garden`) is missing from the knowledge base. This is an example of using the mined rules to identify missing entries.

The rule (`AtLocation`, `PartOf`, `AtLocation`) has 2,394 successes and support 27,917 (8.5% success rate). The triple (`text book`, `classroom`, `school`) is successful and valid. On the other hand, (`text book`, `classroom`, `school system`) is a failure. In contrast to the failure discussed for the first rule above, this is not due to a missing assertion, because the triple is *invalid*. This points to a general problem with this rule: it is only expected to hold if the third concept is a physical object, like `school` and unlike `school system`. Thus examining this example suggests a weakening of the rule.

The rule (`PartOf`, `AtLocation`, `AtLocation`) is similar to the previous one. However, its success rate is much smaller, only 1.4% (with support 78,804, but only 1,112 successes). A possible explanation of the discrepancy can be illustrated by the triple (`engine oil`, `car`, `town`). It is a failure as the assertion (`engine oil`, `AtLocation`, `town`) is not in the knowledge base. Its validity depends on the status of (`engine oil`, `AtLocation`, `town`). This assertion is not to be expected as input from a user (or from a text). On the other hand, it is reasonable as a factual statement about the world.

Let us elaborate on the difference between the two rules. For (`AtLocation`, `PartOf`, `AtLocation`), the combined facts that `a` is an appropriate[5] left argument for `AtLocation`, `b` is an appropriate right argument for `AtLocation`, and (`b`, `PartOf`, `c`) mean that if `c` is an appropriate right argument for `AtLocation` (like `school` but unlike `school system`) then the assertion (`a`, `AtLocation`, `c`) *makes sense both as a factual statement about the world and in terms of natural language usage*. By way of contrast, for (`PartOf`, `AtLocation`, `AtLocation`), things that are appropriate as left arguments for `PartOf` are normally not thought of as appropriate left arguments for `AtLocation`; if they do occur as such a left argument then they occur as being `AtLocation` of the thing they are part of. Thus, in this case (`a`, `AtLocation`, `c`) *may make sense as a factual statement about the world but not in terms of natural language usage*. Thus, the observed difference between the success rates of two similar rules points to a *possible mismatch between natural language usage and intended question answering applications*. This may be an issue to consider for further knowledge base development.

The rule (`LocatedNear`, `PartOf`, `IsA`) does not make much sense even if it has 253 successes and support 4252 (6% success rate). Most successes we examined are false or nonsensical. This is an example of a rule with high success rate but with many successful, invalid triples. An example is the triple (`desk`, `classroom`, `school`). The wrong assertion (`desk`, `IsA`, `school`) comes from the sentence `Schools have desks` through the intermediate form `Desk is a type of school`. Thus the problem presumably comes from a programming error and fixing it might eliminate many wrong assertions. Hence this in an example where rule mining can be used to correct mistakes.

[4] For rules involving more than three concepts such an exhaustive search is not feasible, and it will be necessary to use more advanced data mining techniques.

[5] By *appropriate* we mean "makes common sense for users asked to give natural language statements".

Conclusion

We considered the ConceptNet knowledge base from the point of view of network analysis. We discussed degree distribution, small world property, cores, cliques and communities. We also discussed spreading activation and rule mining for the relations used in ConceptNet. Possible applications to improved question answering include using communities to find missing assertions, using spreading activation to find answers and explanations for *why*-questions and using rule mining to find missing assertions and correct errors.

The mined rules, such as the transitivity of IsA, could be used to add many new assertions. However, adding all these assertions is neither feasible, nor desirable, as it would make the knowledge base denser. The rules appear to be more useful as a background theory, to be used in deriving and refining answers. This could be one instance of building additional knowledge into the system.

ConceptNet provides a possibility to combine statistical and logic-based approaches to commonsense reasoning, exemplified by SVD and spreading activation. Exploring ways of combining the two approaches to enhance performance is an interesting research direction.

How much commonsense reasoning capability is implicit in a large commonsense knowledge base like ConceptNet? Of course it is too early to even guess an answer, but we hope that the explorations outlined in this paper might prove to be useful towards answering this fundamental question.

References

Balduccini, M.; Baral, C.; and Lierler, Y. 2008. Knowledge representation and question answering. In *Handbook of Knowledge Representation*. Elsevier. 779–819.

Brandes, U., and Erlebach, T. 2005. *Network Analysis: Methodological Foundations*, volume 3418 of *LNCS*. Springer.

Clauset, A.; Shalizi, C. R.; and Newman, M. E. J. 2009. Power-Law Distributions in Empirical Data. *SIAM Review* 51(4):661–703.

Collins, A. M., and Loftus, E. F. 1975. A spreading-activation theory of semantic processing. *Psychological Review* 82(6):407.

Crestani, F. 1997. Application of spreading activation techniques in information retrieval. *Artificial Intelligence Review* 11(6):453–482.

Csardi, G., and Nepusz, T. 2006. The igraph software package for complex network research. *InterJournal* Complex Systems(1695).

De Deyne, S., and Storms, G. 2008. Word associations: Network and semantic properties. *Behavior Research Methods* 40(1):213–231.

Derényi, I.; Palla, G.; and Vicsek, T. 2005. Clique percolation in random networks. *Physical review letters* 94(16):160202.

Diochnos, D. I. 2013. Commonsense Reasoning and Large Network Analysis: A Computational Study of ConceptNet 4. *arXiv:abs/1304.5863 [cs/AI]*.

Easley, D., and Kleinberg, J. 2010. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press.

Gonen, M.; Ron, D.; Weinsberg, U.; and Wool, A. 2008. Finding a dense-core in jellyfish graphs. *Computer Networks* 52(15):2831–2841.

Harrington, B., and Clark, S. 2007. Asknet: Automated semantic knowledge network. In *Proc. AAAI Conf. on Artificial Intelligence (AAAI)*.

Havasi, C.; Pustejovsky, J.; Speer, R.; and Lieberman, H. 2009. Digital intuition: Applying common sense using dimensionality reduction. *Intelligent Systems, IEEE* 24(4):24–35.

Havasi, C.; Speer, R.; and Alonso, J. 2007. ConceptNet 3: a flexible, multilingual semantic network for common sense knowledge. In *Recent Advances in Natural Language Processing*, 27–29.

Kotov, A., and Zhai, C. 2012. Tapping into knowledge base for concept feedback: leveraging conceptnet to improve search results for difficult queries. In *Proceedings of the fifth ACM international conference on Web search and data mining*, 403–412.

Leskovec, J.; Lang, K. J.; Dasgupta, A.; and Mahoney, M. W. 2009. Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters. *Internet Mathematics* 6(1):29–123.

Maybury, M. T. 2004. Question answering: An introduction. In *New directions in question answering*. AAAI Press. 3–8.

McCarthy, J. 1959. Programs with common sense. In *Proceedings of the Teddington Conference on the Mechanization of Thought Processes*, 756–91.

Newman, M. E. J. 2003. The structure and function of complex networks. *SIAM Review* 45:167–256.

Ohlsson, S.; Sloan, R. H.; Turán, G.; Uber, D.; and Urasky, A. 2012. An approach to evaluate AI commonsense reasoning systems. In *Proc. 25th Int. Florida Artificial Intelligence Research Society Conference*, 371–374.

Ohlsson, S.; Sloan, R. H.; Turán, G.; and Urasky, A. 2013. Verbal IQ of a Four-Year Old Achieved by an AI System. Submitted for publication.

Page, L.; Brin, S.; Motwani, R.; and Winograd, T. 1999. The PageRank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab.

Palla, G.; Derényi, I.; Farkas, I.; and Vicsek, T. 2005. Uncovering the overlapping community structure of complex networks in nature and society. *Nature* 435(7043):814–818.

Seidman, S. B. 1983. Network structure and minimum degree. *Social networks* 5(3):269–287.

Shen, E.; Lieberman, H.; and Lam, F. 2007. What am I gonna wear?: Scenario-oriented recommendation. In *Proceedings of the 12th international conference on Intelligent user interfaces*, 365–368.

Speer, R.; Arnold, K.; and Havasi, C. 2010. Divisi: Learning from semantic networks and sparse SVD. In *Proc. 9th Python in Science Conf. (SCIPY 2010)*.

Speer, R.; Havasi, C.; and Lieberman, H. 2008. AnalogySpace: reducing the dimensionality of common sense knowledge. In *Proc. AAAI Conf. on Artificial Intelligence (AAAI)*.

Steyvers, M., and Tenenbaum, J. B. 2005. The Large-Scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science* 29(1):41–78.

Verberne, S.; Boves, L.; Oostdijk, N.; and Coppen, P.-A. 2010. What is not in the bag of words for why-QA? *Computational Linguistics* 36(2):229–245.

Watts, D., and Strogatz, S. 1998. Collective dynamics of small-world networks. *Nature* 393:440–442.