REMOD: Relation Extraction for Modeling Online Discourse

Matthew Sumpter
mjsumpter@usf.edu
University of South Florida

Giovanni Luca Ciampaglia
glc3@mail.usf.edu
University of South Florida

ABSTRACT

The enormous amount of discourse taking place online poses challenges to the functioning of a civil and informed public sphere. Efforts to standardize online discourse data, such as ClaimReview, are making available a wealth of new data about potentially inaccurate claims, reviewed by third-party fact-checkers. These data could help shed light on the nature of online discourse, the role of political elites in amplifying it, and its implications for the integrity of the online information ecosystem. Unfortunately, the semi-structured nature of much of this data presents significant challenges when it comes to modeling and reasoning about online discourse. A key challenge is relation extraction, which is the task of determining the semantic relationships between named entities in a claim. Here we develop a novel supervised learning method for relation extraction that combines graph embedding techniques with path traversal on semantic dependency graphs. Our approach is based on the intuitive observation that knowledge of the entities along the path between the subject and object of a triple (e.g. `Washington,_D.C.`, and `United_States_of_America`) provides useful information that can be leveraged for extracting its semantic relation (i.e. `capitalOf`). As an example of a potential application of this technique for modeling online discourse, we show that our method can be integrated into a pipeline to reason about potential misinformation claims.

CCS CONCEPTS

• **Information systems** → *Web mining*; *Semantic web description languages*; **Information extraction**.

KEYWORDS

relation extraction, semi-structured data, semantic ontology, claim matching, fact-checking

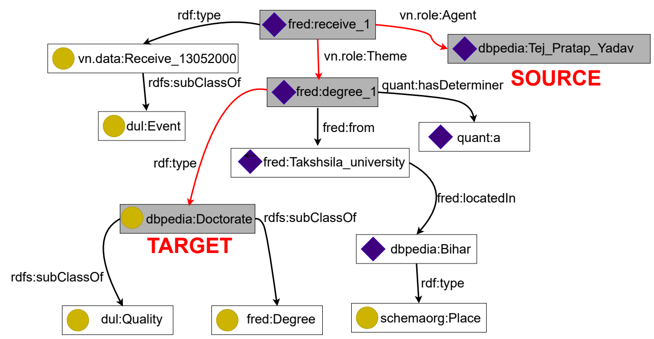

Figure 1: Schematic example of our approach. The RDF graphlet generated by a machine-reading tool (FRED) for the claim "*Tej Pratap Yadav receives a doctorate degree from Takshsila University in Bihar*" (a known misinformation claim [26]). The shortest undirected path between the source (`dbpedia:Tej_Pratap_Yadav`) and target (`dbpedia:Doctorate`) is shown in red. The nodes along the path are highlighted in gray.

1 INTRODUCTION

The prevalence of false and inaccurate information in its myriad of forms — a persistent and dangerous societal problem — is still a poorly understood phenomenon [1, 7, 30], especially in the context of political communication [21]. Even though strong exposure to so-called "fake news" is limited to the segment of most active news consumers [19], individual claims echoing the false or misleading content shared by these audiences can spread rapidly through social media [57, 69], amplified by bots [46] or other malicious actors [60], who often target elites, like celebrities, pundits, or politicians. From there, false claims rebroadcast by these elites enjoy further dissemination, reaching even wider audiences.

Misinformation has become an emerging focus of computational social scientists seeking to understand and combat it [10, 56]. Network analysis and natural language processing (NLP) provide insight into the community organization and stylistic patterns that are indicative of misinformation, respectively, however they often fail to engage with the ideological content being shared. Online discourse typically takes the form of unorganized and unstructured data which is a significant limiting factor to performing content analysis. Existing work on semantic ontologies and knowledge base development has proved to be a guiding method in structuring online information. A knowledge base most commonly structures knowledge in the shape of semantic triples; a semantic triple is composed of two entities (e.g. a person, place, or thing) and a predicate relation between them. An example of a semantic triple is <Washington_D.C, capitalOf, United_States_of_America>. This structure allows for concepts to be reduced to machine-readable data which can be compiled into traversable (and understandable) networks of information. The result is a data structure that can be used to provide quantitative analysis of online discourse.

An example of knowledge bases application in combating misinformation regards computational fact-checking. Fact-checking is recognized as an antidote to misinformation [32], especially with respects to claims spread by political elites. For example, Nyhan and Reifler [36] show that alerting politicians to the risk of being fact-checked leads to less inaccuracy and better ratings. Unfortunately, fact-checking claims at the scale of the web is a hard task. A fact-checker must first identify claims that are worthy of being checked, then they must research the claim [6, 51], and finally write, publish, and circulate their conclusion on the web. In general, there is a lag of approximately 15 hours between the consumption of misinformation and the appearance of corrections [45]. The time investment required of human fact-checkers leads to an open opportunity for the development of many automated fact-checking [11, 63] or verification [34] strategies. One approach is based on identifying missing relations in structured knowledge bases [11, 29, 47, 48]. This approach takes a claim in the form of a semantic triple and checks its validity against the sets of triples in the knowledge base that connect the subject and object. When the knowledge base is viewed as a network, this task is equivalent to link prediction [33].

This approach has proven very promising, but its main restriction is that of its input. Modeling a claim using semantic triples is a nontrivial task, and has limited the application of such an approach. It requires choosing a semantic ontology (or developing a new one) which is able to model claims in a consistent and non-redundant manner. Once an ontology has been established, the next step is relation extraction — the task of reducing a text into a semantic triple that both captures the meaning and fits within the ontology. This task is challenging when addressing a compound factual claim with many subjects and relations; this challenge is amplified when considering a claim that may contain sarcasm, opinion, humor, or any other nuance of language that can be present in online discourse.

In this paper, we present a novel relation extraction method built upon semantic dependency trees, see Figure 1 for a schematic example. Our approach to the problem is based on the intuition that knowledge of the nodes and relations along the path between the subject and object of a triple (e.g. `Washington,_D.C.`, and `United_States_of_America`) provides useful information that can be leveraged for extracting its relation (i.e. `capitalOf`). This well-established phenomenon was first observed by Richards and Mooney [41]. Later, Bunescu and Mooney [9] used it in the context of a kernel-based approach. Here, we take advantage of recent advances in graph representation learning to overcome the above challenges posed by online discourse in applying such an approach. Specifically, we parse a large corpus of Wikipedia snippets, annotated with information about one of 5 relations from the DBpedia ontology, combine the resulting dependency trees into a larger semantic network, and finally use node embedding techniques to obtain a high-dimensional representation of this corpus-level network. We find that graph traversal in this learned representation provides a strong signal to discriminate between multiple possible relations.

This approach allowed us to effectively extract these relations in natural language (extraction accuracy measured as the area under the ROC curve, AUC = 0.976). We then tested this model's ability to generalize to a set of real-world claims (reviewed by professional fact-checkers and annotated using the ClaimReview [22] schema), obtaining again a very good signal (extraction AUC = 0.958).

As an example of a potential application of this technique, we show that, thanks to our method, a wider range of online discourse samples is amenable to analysis than before. In particular, we integrate our approach into a pipeline (see Figure 2) that uses off-the-shelf fact-checking algorithms to analyze a subset of ClaimReview-annotated online discourse samples. Using this pipeline, we obtain very encouraging results on two separate tasks: First, on samples of 'simple' online discourse claims, which can be effectively summarized (and thus fact-checked) by extracting a single RDF triple, we outperform a claim-matching baseline based on state-of-the-art representation learning (verification AUC = 0.833). Second, on more complex claims, from which one can extract multiple relevant relations, and therefore cannot be fact-checked directly, the fact-checker can still identify evidence in support or against the claim with good accuracy (verification AUC = 0.773).

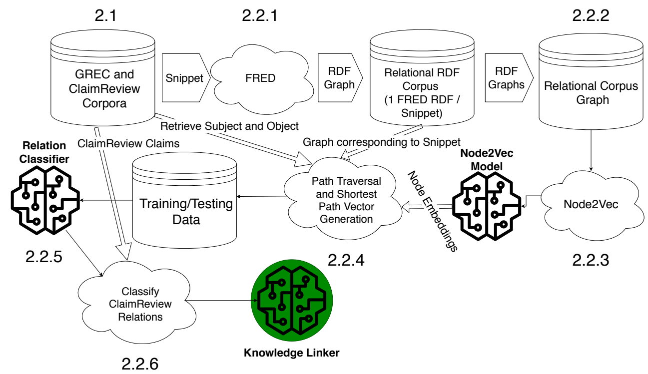

Figure 2: Schematic illustration of an integrated extraction and verification pipeline using our relation extraction tool REMOD. The white components correspond to the various steps needed to perform relation extraction. Numbered labels correspond to section headings in the manuscript. To show the potential for integration with external tools, as an additional step in the pipeline the green node shows the use of an off-the-shelf fact-checking algorithm [11].

The rest of this paper is structured as follows: Section 2 details the datasets used, as well as the methods used in the various steps of the pipeline. Section 3 shows the results of both the relation classification task and the fact checking tasks. Section 4 goes into detail on relevant prior work from the literature on relation classification, misinformation detection, and computational fact-checking. Finally, Section 5 discusses the impact and importance of our results, as well as addresses methods that may be used to improve upon this work in the future.

2 METHODS

Our relation extraction pipeline is described in Figure 2. Roughly speaking, the main task of our pipeline is a supervised relation extraction task (white nodes), but since later we show how this task can be integrated to perform an additional unsupervised fact-checking, in the figure we show also this final step (green node). Collectively these two tasks leverage a number of different data sources, so we start by describing the various datasets used in building the pipeline. We then describe the various components of the pipeline proper.

2.1 Datasets

For the main relation extraction task, we use two main corpora, both compiled by Google: the Google Relation Extraction Corpus (GREC) and the Google Fact Check Explorer corpus, described below.

2.1.1 Google Relation Extraction Corpus (GREC). The dataset of relations used was the Google Relation Extraction Corpus (GREC) [37].

Table 1: Number of snippets per relation before and after filtering the GREC corpus.

	Total	Retained	% Retained
Institution	42,628	19,900	46.7
Education	1,850	806	43.6
Date of Birth	2,490	1,010	40.6
Place of Birth	9,566	4,005	41.9
Place of Death	3,042	1,307	43.0

This dataset contains text snippets extracted from Wikipedia articles that represent a subject/object relation, which can be described by the following defining questions:

Institution "What educational institution did the subject attend?"
Education "What academic degree did the subject receive?"
Date of Birth (DOB) "On what date was the subject born?"
Place of Birth (POB) "Where was the subject born?"
Place of Death (POD) "Where did the subject die?"

Each entry in the dataset consists of a natural language snippet of text, the URL of the Wikipedia entry from which the text was pulled, the Freebase predicate, a Freebase ID for subject and object, and the judgements of five human annotators on whether the snippet does or does not contain the relation (some annotators also voted to "skip", representing no decision either way). Freebase has been replaced with the Google Knowledge Graph since this dataset was generated, which limited the use of this dataset in its original form. We made a set of addenda[1] to the GREC to update it to be more machine-ready for current relation extraction tasks and knowledge bases. The addenda include the following for each entry: text strings for both subject and object, DBpedia URI for both subject and object, Wikidata QID for both subject and object, a unique identifier, and the majority annotator vote.

The snippets varied considerably in length. The distribution of word lengths can be found in Figure 3. Because we relied on a third-party API to parse the snippets, to reduce potential bias due to snippet length, and to ensure only the most characteristic relations were modeled, snippets were removed if they were not within ±0.5 standard deviations of the mean snippet length (measured in words), per relation. Table 1 shows the number of snippets retained, per relation.

2.1.2 Google Fact Check Explorer Corpus. Researchers at Duke University and Google have developed an annotation standard named ClaimReview [44] to help annotate structured fact-checks on the web. It allows fact-checkers to add structured markup to their fact-checks with info that identifies distinct properties of a claim (i.e. claim reviewed, the rating decision, the source, etc.). This semi-structured data allows fact-checks to be catalogued and queried by search engines. The Google Fact Check Explorer tool[2] collects all the ClaimReview fragments published by fact-checking organizations that meet a set of established guidelines[3], which are the same standards for accountability, transparency, and accuracy used by Google News to select publishers. We collected claims from the Google Fact Check Explorer tool up until 04/2020. From this corpus, we produced a dataset of 49,770 ClaimReview-annotated claims. Of the 20,817 English claims in the dataset, we searched for claims that contained one of the relations represented in the GREC, using WordNet [14] synonyms to select search terms (see Table 2). This procedure yielded a subset of 28 claims that met this criteria.

[1]https://github.com/mjsumpter/google-relation-extraction-corpus-augmented
[2]https://toolbox.google.com/factcheck/explorer
[3]https://developers.google.com/search/docs/data-types/factcheck#guidelines

Table 2: The set of WordNet synonyms used to extract relevant claims from the ClaimReview database

	WordNet synonyms per relation
Institution	*attend, university, college, graduate*
Education	*graduate, degree*
Date of Birth	*born, born on*
Place of Birth	*born, birthplace, place of birth, place of origin*
Place of Death	*deceased, died, perished, passed away, expired*

2.2 REMOD

The main contribution of this work is REMOD (which stands for Relation Extraction for Modeling of Online Discourse), a novel tool for relation extraction that extract RDF triples from semi-structured samples of online discourse. To do so, our tool leverages an annotated corpus of past claims and relations. In the example pipeline shown in Figure 2, the various steps of REMOD correspond to the white nodes, which we describe in more detail below. (The figure is labeled with numbers corresponding to the following section numbers, which elaborate on each step of the process.) To facilitate the replication of our results, the source code of REMOD is freely available online at https://github.com/mjsumpter/remod.

2.2.1 Semantic Parsing. Our workflow begins with natural language snippets. To parse these snippets we used FRED, a machine reading tool based on Discourse Representation Theory and linguistic frames [17], described by the authors as "semantic middleware". FRED is an NLP tool that combines frame detection, type induction, named-entity recognition, semantic parsing, and ontology alignment, all into a single tool. The authors provide a RESTful API to access it. When provided with a text string as input, it returns a Resource Description Framework (RDF) graphlet of the semantic parse tree of the input. (In practice, FRED produces DAGs instead of trees due to entity linking to external ontologies, hence our referring to them as 'graphlets'.) An example of these RDF graphlets is shown in Figure 1 for the ClaimReview snippet of a known misinformation claim [26].

2.2.2 Corpus Graph Composition. In a realistic environment, many claims of different relations will exist in the same corpus. To mimic this environment, we composed a single 'corpus' graph, which was composed of every FRED RDF graphlet generated from the corpus snippets. For named entities, FRED defaults to generating nodes for its own namespace (e.g. `fred:Doctorate`), then if it finds that the same entity is present in an existing ontology, it links to that ontology (e.g. `dbpedia:Doctorate`). Since these equivalent entities were redundant, we contracted the two nodes into a single

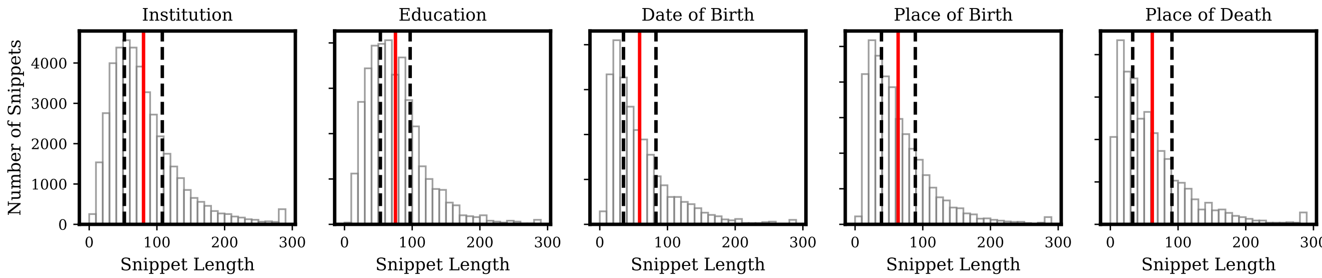

Figure 3: Distribution of snippet lengths found in the GREC. The red solid line corresponds to the average snippet length (in words) and the dashed lines to $\pm 0.5\sigma$ of the average. Snippets were kept if they were within this interval.

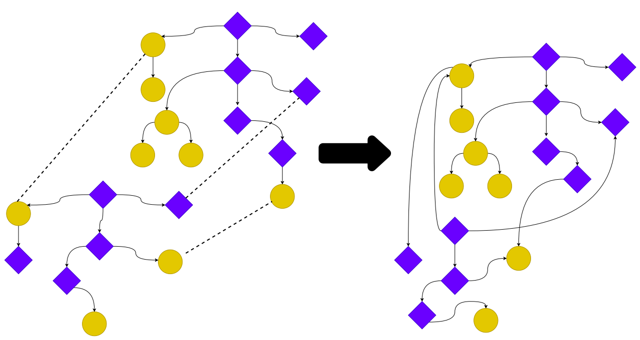

Figure 4: A visualization of how two separate RDF graphlets were stitched together along identical nodes.

vertex, and use the URI from the linked ontology (i.e. DBpedia in this example) as its new URI. The corpus graph was than created by stitching together all the contracted RDF graphlets: if two graphlets share one or more nodes (i.e. two or more nodes have the same URI), then we consider the union of the two graphlets, and contract any pair of such nodes into a single node. This new node is incident to the union of all incident edges in the two original graphlets. An example of this is shown in Figure 4. The resulting corpus graph consists of $212,976$ nodes and $832,367$ edges.

2.2.3 Node Embedding. The corpus graph is effectively a combined semantic parse tree of the selected snippets from the corpus. To better exploit this structure in machine learning tasks, we generated node embeddings using the Node2Vec algorithm [20]. Node2Vec generates sets of random walks for each node, which are then substituted in place of natural language sentences as input into the Word2Vec model. There are two important parameters which will influence the nature of the embeddings: the return parameter p and the in–out parameter q. For $p > 1$ there is a higher likelihood of returning to a visited node in the random walks, whereas for $q > 1$ there is an increased likelihood of exploring unvisited nodes. We performed a grid search of p and q parameters (see Section 3.2), and determined the best choice for these parameters to be $p = 2$ and $q = 3$; this configuration captures what the authors of Node2Vec call the 'global' topological structure of the graph. The other parameters of Node2Vec were chosen as follows: the dimension of the vector space was set to 256; the number of walks to 200; the walk length to 200; and, finally, the context window to 50.

2.2.4 Path Traversal for Finding Relations. Our approach is inspired by the well-known idea that finding paths over structured knowledge representations can help learning new concepts [41]. More recently, Bunescu and Mooney [9] confirmed the intuitive conclusion that the shortest path between entities in a dependency tree captures the significant information contained between them. Therefore, we sought to develop a classifier that could distinguish between the shortest paths of different semantic relationships. To do so, for each snippet in the corpus, the subject and object were retrieved, along with the original (i.e., non-stitched) RDF graphlet of that specific snippet. The nodes corresponding to the subject and object were identified in the RDF graphlet. With the terminal nodes identified, the shortest path in the original RDF graphlet was calculated (Figure 1). Finally, we generated a final embedding by averaging along the path:

$$\frac{1}{n}\sum_{i=1}^{n}\vec{v_i}$$

where $v_1, \ldots, v_n$ is a path and $\vec{v} \in \mathbb{R}^d$ is the vector associated to $v \in V$. This resulted in a final vector representing the aggregated sequence of nodes along the shortest path between subject and object.

This process resulted in a 256-dimensional vector for each snippet in the corpus. All results shown in the next section were obtained from these vectors. We projected the vectors into a lower-dimensional space using t-SNE. The visualization of these vectors is shown in Figure 5, where each color corresponds to a different relation. The projection reveals a good separation of vectors based on the relation they represent.

2.2.5 Relation Classification. We trained several classification models on the resulting set of shortest path vectors. The selected classifiers were Logistic Regression, k-NN, SVM, Random Forest, Decision Tree, and a Wide Neural Net. Samples that were rated by the annotators to not contain a specified relation were removed, and then the dataset was balanced to the lowest frequency class (Education, $N = 598$ samples). Readers will note this is a decrease from the 806 reported in Table 1; FRED was not always accurate at identifying entities and occasionally returned corrupted RDF graphs, resulting in a small loss of data. To effectively compare

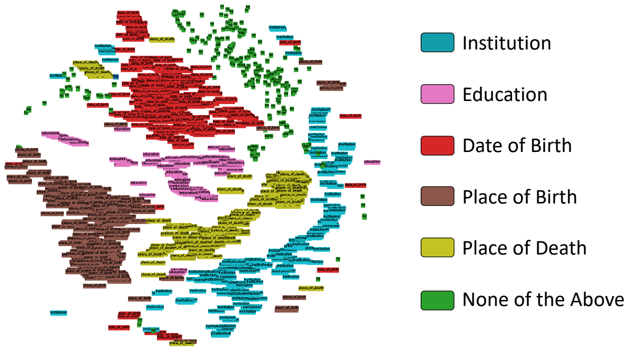

Figure 5: The shortest path vectors of GREC relations projected into 2D using t-SNE. Each color represents a different semantic relation, with a sixth color to mark snippets for which a majority of annotators voted 'No (relation)'.

different classifiers, training was done using a 64%/16%/20% training/validation/testing split. This resulted in a final training dataset of $1,913$ samples (5 classes, $N \approx 382$ samples/class), with a validation set of 479 samples, and an additional 598 samples held for testing. The 28 selected ClaimReview claims were held as an additional test set, which is elaborated on in Section 2.2.6.

2.2.6 Fact-Checking. To demonstrate the usefulness of our method, we show that REMOD can be integrated as the first step of a fact-checking pipeline using existing, off-the-shelf tools to verify online discourse claims annotated using the ClaimReview standard. To perform fact-checking, we rely on the work of Shiralkar et al. [48], who provide open-source implementations of several fact-checking algorithms[4]. These algorithms can be used to assess the truthfulness of a statement, but of course any tool that takes RDF triple in input could be used as well. To extract relations from ClaimReview snippets, we used the deep neural network classifier, which was the most successful classifier from the prior step, and feed the extracted triples into the fact-checker.

Of course, when integrating two distinct tools one has to make sure that any error originating in the first tool does not affect the performance of the second tool. Therefore, to avoid cascading errors we removed some claims from our dataset. We removed two types of errors. First, we removed any claim where the relation was misclassified, to avoid feeding inaccurate inputs into the fact-checker. Second, FRED is not always able to link both the subject and object entities to DBpedia, which is a requirement for using the fact-checking algorithms of Shiralkar et al. [48]. Thus we also removed claims that did not have both the subject or object linked to the DBpedia ontology. Of the original 28 claims, this filtering resulted in 13 remaining ClaimReview claims used in our evaluation.

Additionally, we also manually checked whether the overall claim reduces to the extracted triple (in the sense that verifying the triple also verifies the overall claim). This distinction is important since it allows us to gauge the ability of our system to check entire claims automatically, in a purely end-to-end fashion. Finally, these remaining claims were passed as input to three fact-checking algorithms: Knowledge Stream, Knowledge Linker, and Relational Knowledge Linker [48].

As a baseline, we trained a Doc2Vec model [31] on the entirety of the ClaimReview corpus, and used this model to fact-check statements by matching them with other similar claims. In particular, given an input claim, to produce a truth score with the baseline model we ranked all claims in the ClaimReview corpus by their similarity and averaged the truth scores of the top k most similar matching claims. We removed fact-checking organizations that used scaleless fact-check verdicts (i.e. factcheck.org); for those that had scales, we assigned truth scores to every claim, setting "False" to a baseline of 0, unless a scale explicitly stated a different baseline (i.e. PolitiFact ranks "Pants on Fire" lower than "False").

[4]https://github.com/shiralkarprashant/knowledgestream

Table 3: AUC of Wide DNN on the relation classification task using different types of graph to represent the corpus graph.

AUC	Unweighted	Weighted
Undirected	**0.976**	0.964
Directed	0.966	0.967

3 RESULTS

3.1 Graph Representation

The corpus graph is composed of dependency trees, and so the corpus graph is naturally a directed graph; edges are also all weighted equally. This design has a strong influence on path traversal, since directed edges reduce the number of available paths and the cost of taking an edge (or its absence) influences the choice of one path over another. For completeness, we considered all four combinations of taking either a directed or undirected graph, and of having edge weights or not. Let $v_i, v_j \in V$ represent two nodes in the dependency graph that are incident on the same edge. The weight w_{ij} between them is the angular distance between the respective node embeddings:

$$w_{ij} = \frac{1}{\pi} \arccos\left(\frac{\vec{v}_i \cdot \vec{v}_j}{\|\vec{v}_i\| \cdot \|\vec{v}_j\|}\right)$$

Where $\vec{v}$ is the vector associated to $v \in V$.

Table 3 shows that the undirected, unweighted graph yields the best classification results, which prompts two observations. The first is that directed edges reduce the number of available pathways to connect two nodes. Second, and perhaps a bit surprisingly, we observe that the unweighted network performs better than the weighted one. Because node embeddings were the same in the two variants, the final feature vector used for relation classification would be different only if a different shortest path was found. This could be possible if edges that are more relevant to discriminating the relation were assigned large weights, compared to other, less relevant edges.

3.2 Classification for Relation Extraction

The results of the relation classification task are shown in Table 4. The outcome of these various tests reveal that the node embeddings do contain information regarding the semantic nature of the GREC

Table 4: Results of the relation classification task using different ML models, on an unweighted, undirected corpus graph, as compared to training with Word2Vec embeddings.

	Precision	Recall	F1	AUC
Decision Tree	0.64	0.64	0.64	0.773
Random Forest	0.81	0.67	0.61	0.793
k-NN	0.78	0.74	0.74	0.841
SVM	0.81	0.77	0.77	0.855
Log. Regr.	0.80	0.71	0.71	0.827
Wide DNN	**0.85**	**0.85**	**0.85**	**0.976**
Word2Vec+Log. Regr.	0.66	0.47	0.44	0.658
Word2Vec+Wide DNN	0.61	0.63	0.61	0.883

relations, however they are not neatly separable by decision planes. It is notable that models we tested are often more successful in precision than in recall. This suggests that the more complex model, such as a deep neural network (DNN), is necessary to identify the less characteristic samples of a relation. To improve these results, we performed a grid search on the Node2Vec p and q parameters (with values of 0.25, 0.5, 1, 2, 3, and 4). The best overall results were a product of a 'global' configuration, using $p = 2$ and $q = 3$, which achieved an AUC of 0.976 on the test set. To evaluate our method, as a baseline we generated 300-dimension vectors for each snippet from a Word2Vec model, pre-trained on Wikipedia [66]. This is the same source of the GREC corpus, which provided training data for model. These embeddings were then used as features to train a DNN and a logistic regression model for relation extraction. REMOD shows a marked improvement in both instances, indicating an effective approach to relation extraction.

3.3 Extraction of ClaimReview Relations

Table 7 in the appendix shows the claims selected from the ClaimReview corpus, in addition to the relation they contain ("Actual"), the relation predicted by REMOD ("Predicted"), the truth rating as determined by a fact-checker ("Rating"), and whether verifying the relation is equivalent to verifying the claim ("Claim ≡ Triple"). The AUC of the predicted relations is 0.958. Inspecting the misclassified samples, we see that REMOD made mistakes between similar relations (e.g. place of birth and date of birth), which often occur in similar sentences.

3.4 Fact-Checking

We next test the integration with fact-checking algorithms. In particular, we use the fact-checker for two similar, but conceptually distinct tasks: 1) fact-checking an entire claim (*fact-checking*), and 2) identifying evidence in support or against a claim (*fact verification*). For example, for claim #1 (see Table 7), Penny Wong was indeed born in Malaysia, even though the assertion that she is ineligible for being elected into the Australian parliament is false. Thus, in this case the extracted triple is only additional evidence, but is not able in itself to capture the entire claim. We manually fact-checked all the extracted relations, and compared their truth rating with the one provided by the human fact-checker for the whole claim.

Table 5: The performance of the fact-checking algorithms on predicting the validity of the relations.

Method	AUC
Knowledge Linker	0.636
Relational Knowledge Linker	**0.773**
Knowledge Stream	**0.773**

Table 7 lists this information under the column "Claim ≡ Triple", which is true (indicated by a checkmark) when the extracted relation summarizes the whole claim (e.g. claim #3). This distinction is important: as mentioned before, although our relation extraction pipeline is capable of predicting a relation for all the entries in Table 7, not all triples that are correctly predicted can be fed to the fact-checking algorithms, due to incomplete entity linking. For the task of identifying supporting evidence, we find a total of 13 ClaimReview claims that are amenable to fact-checking. For the task of checking an entire claim, this number is further reduced to 7 claims.

3.4.1 Fact Verification. Table 5 shows the results of verifying individual pieces of evidence in support or against any of the 13 ClaimReview claims identified by REMOD, using any of the three algorithms for fact-checking RDF triples. Relational Knowledge Linker and Knowledge Stream were the best performers. Note that since our baseline is intended to emulate a true fact-checking task, in this case we do not run the baseline since the similarity is based on the whole claim, and thus would not be a meaningful comparison with our method, which focuses only on a specific relation within a larger claim.

3.4.2 Fact-Checking. We test here the subset of claims for which checking the triple is equivalent to checking the entire claim. In this case, REMOD yields 7 claims that can be used as inputs to the fact-checking algorithms. Table 6 shows the results of our 7 ClaimReview claims, on the three fact-checking algorithms, along with the baseline. Here, the baseline emulates fact-checking by claim matching.

Since we are using claim-matching to perform fact-checking, we consider three different scenarios to make the task more realistic. In particular, we match the claim against three different corpora by higher degree of realism: 1) the full ClaimReview corpus ('All'), 2) all ClaimReview entries by PolitiFact only ('PolitiFact'), and 3) all ClaimReview entries from the same fact-checker of the claim of interest ('Same'). The first case ('All') is meant to give an upper bound on the performance of claim matching but is not realistic, since it makes use of knowledge of the truth score of potentially future claims, as well as of ratings for the same claim but by different fact-checkers. The second case ('PolitiFact') partially addresses this second unrealistic assumption by using only claims from a single source. Thus, it does not have access to truth scores by different organizations for the same claim, but it does still have access to future information. Both 1) and 2) can be thus regarded as gold standard measures of performance. The last one ('Same') is the more realistic one, since it emulates the scenario of a fact-checker who may check a claim for the first time, and who thus cannot

Table 6: Results of the fact-checking algorithms. (CM = Claim Matching; KL = Knowledge Linker; Rel. KL = Relational Knowledge Linker; KS = Knowledge Stream.)

	$k = 1$	$k = 3$	$k = 5$	$k = 10$
CM (All)	0.417	**0.625**	0.500	**0.625**
CM (PolitiFact)	0.666	0.625	**0.833**	0.750
CM (Same)	0.500	**0.583**	0.25	0.25
KL	0.500			
Rel. KL	**0.833**			
KS	**0.833**			

have access to claims fact-checked afterwards nor by ratings of the same claim by different fact-checkers. In all three cases, the claim being matched was removed from the corpus, to prevent trivially perfect predictions. Relational Knowledge Linker and Knowledge Stream are still the best performing of the fact-checking algorithms and manages to reach, if not exceed, the performance of the gold standard (Claim Matching–All, or –PolitiFact).

4 RELATED WORK

4.1 Relation Extraction and Classification

Relation extraction and classification is the task of extracting semantic relationships between two entities in natural language text and matching them to semantically equivalent or similar relations. This task is at the core of information extraction and knowledge base construction, as it effectively reduces statements to their core meaning; this is typically modeled as a semantic triple, (*s,p,o*), where two entities (*s* and *o*) are connected with a predicate, *p*. There are several distinct nuances and open challenges to effective relation extraction. Identifying attributes that discriminate between two objects provides a descriptive explanation to supplement word embeddings (i.e. lime is separated from lemon by the attribute 'green'), and is currently most successful with SVM classifiers [27]. Multi-way classification attempts to distinguish the direction of one-way relations (the `sonOf` relation is not bidirectional between two people), and has seen similar levels of success from solutions built with language models [3], convolutional neural networks [58], and recurrent neural networks [64]. Distantly supervised relation extraction is a two-way approach whereby semantic triples are generated from natural language by aligning them with information already present in knowledge graphs [65]. Relation extraction performance is often assessed on the TACRED dataset [68]. This is a large-scale dataset of $106,264$ examples used in the annual TAC Knowledge Base Population challenges, and covers 41 relation types. The most successful solution to date is from Baldini Soares et al. [3], who achieved a micro-averaged F1 score of 71.5%. Despite increasing availability of state-of-the-art machine learning architectures, relation extraction continues to be an open problem with much room for improvement.

4.2 Knowledge Base Augmentation

Knowledge base augmentation is a task that aims to add new relations to existing knowledge bases in an automated fashion [61]. This task takes one of two approaches; the first infers new relations from existing triples in a knowledge base [8, 53] — this is essentially a link-prediction task that builds upon patterns found between entities in knowledge bases. The second approach mines data found on the web for knowledge discovery [12, 67]. This approach relies on redundant relations found among the selected source materials, which may be as restrictive as Wikipedia articles [39] or as extensive as the entire web [12]. Due to the potential for error based on the sources, Dong et al. [13] developed a Knowledge-Based Trust (KBT) score for measuring the trustworthiness of selected sources. Yu et al. [67] expand upon this by combining KBT scores with other entity/relation-based features to assign a unique score to each individual triple.

4.3 Detecting Information Disorder

Information disorder is a catch-all term for the many kinds of unreliable information that one may encounter online or in the real-world [59], which includes disinformation, misinformation, fake news, rumor, spam, etc. Information disorder can also take on several modalities, including text, video, and images. The many varieties of information disorder make it challenging to develop any one approach for detection. This leads to a multi-model approach to detection based on three main modalities: the content of the information, the users who shared it, and the patterns of information dissemination on a network. Often bad content is generated by bots; this suggests that features captured from user profiles can be useful for distinguishing bots from humans [50]. Content detection is dependent on the medium; lexical features, sentiment, and readability metrics are used for text, while neural visual features are extracted from other content [40, 42, 43]. Network detection methods model social media networks as propagation networks, measuring the flow of information [49]. There has also been promising work into crowd-sourcing the task by allowing users to flag questionable content [55]. This task, while likely to remain imperfect, provides the important supplement of human supervision to all of the aforementioned tasks.

4.4 ClaimBuster

Hassan et al. [24] released the first-ever end-to-end fact-checking system in 2017, called ClaimBuster. ClaimBuster is composed of several distinct components that work in sequence to accomplish the task of automated fact-checking. The first, *claim monitor*, continuously monitors text published as broadcast television closed-captions, Twitter accounts, and as content on a selected set of websites. This text is passed to the *claim spotter*, which scores every sentence by its likelihood to contain a claim that is worthy of fact-checking — subjective and opinionated sentences receive a low score in this task. Once it has identified a set of check-worthy sentences, it uses a *claim matcher* to search through fact-check repositories to return existing fact-checks that match the selected sentences. *Claim checker* generates questions from the selected sentences and uses those questions to query Wolfram Alpha and Google to fetch supporting or debunking evidence as a supplement to the findings of *claim matcher*. Finally, the *fact-check reporter* builds a report from all of the gathered evidence that summarizes the findings of the ClaimBuster pipeline, and disseminates these findings through social media.

4.5 Claim Verification

Claim verification is arguably the key task of fact-checking — to check a claim against existing evidence. It is related to the matching and checking subtasks of ClaimBuster, in that it is the task of checking whether a natural language sentence selected as evidence supports or debunks the correlated claim. To build out computational solutions to this task, datasets containing claims and their corresponding evidence are needed. There have been some datasets [2, 15, 56] relevant to this task, however they are either not machine-readable or lacking in size. Thorne et al. [54] recognized this gap, and has since released a large-scale dataset to address these concerns, called FEVER. This dataset contains 185,445 claims with corresponding evidence that were manually classified as `SUPPORTED`, `REFUTED`, or `NOTENOUGHINFO`. This has been followed up with annual workshops that encourage participants to improve upon both the dataset and the claim verification task. The CLEF CheckThat! [4] series of workshops and conferences also seek to bring researchers together to improve claim verification, along with identifying and extracting checkworthy claims.

4.6 Other Fact-Checking Methods

Besides claim-matching approaches, there are a handful of existing algorithms for fact-checking, mostly based on exploiting content or characteristics of existing knowledge bases. Embedding approaches, such as TransE [5], seek to generate vector embeddings of knowledge bases, a task which is conceptually related to our approach. By generating these embeddings, they can perform link-prediction based on structural patterns of (*s, p, o*) triples. In terms of a knowledge base, this amounts to adding new facts without any needed source material. For fact-checking, this approach can be used to test whether a triple extracted from a claim is a predicted link in the knowledge base; the pitfall of these methods, as with all embedding techniques, is they lack both interpretability and scalability. Other algorithms similarly consider paths within knowledge bases, but seek to address the interpretability problem. PRA [28], SFE [18], PredPath [47], and AMIE [16] all take the approach of mining possible pathways between two entities within a knowledge base. From these mined pathways, they generate sets of features to be used in supervised learning models for link-prediction. These have shown promise in their success at predicting the validity of a claim, however this also suffers from scalability. Knowledge bases that contain enough relevant information to be useful are very large, and path mining and feature generation becomes necessarily time-consuming. There are a few rule-based [38] methods for fact-checking, which rely on logical constraints of a knowledge graph and are naturally explainable. General, large-scale knowledge graphs do not have these logical constraints from which to build rules from, leaving this approach to fact-checking an open problem [25].

4.7 Threats to Validity

No method is perfect and our approach suffers from a number of limitations, which we briefly describe here. The main limitation of our pipeline lies in its discrete structure, which is prone to cascading failures. Our main NLP tool, FRED, is a powerhouse of a tool and performed many important NLP tasks at once; however, it was not always completely accurate and many of our samples were returned as corrupted RDF graphs. Additionally, it was not always able to link the nodes to DBpedia, which limited the number of triples we could feed into our fact-checking algorithms. Cascading failures are common to many machine reading pipelines [35]. One way to overcome this issue would be to rely on a joint inference approaches [52]. Another limitation of our methodology has to do with our use of distributed representations. For the task of fact-checking, the corpus is always growing; Node2Vec cannot generalize to unseen data and requires retraining. An inductive learning framework, such as GraphSAGE [23], can generate embeddings for unseen nodes, and is therefore a more practical algorithm for extending this pipeline. For the classification task, our machine learning models were relatively simple, and optimizing both the parameters and architecture of the neural network would likely see an increase in the accuracy and effectiveness of this method. Finally, a full evaluation of our method against transformer language models for relevant relation classification tasks [62] is left as future work.

5 DISCUSSION

In this paper, we have presented a novel relation extraction algorithm and previewed its application when used to classify relations present in online discourse and automatically fact-check them against the information present in a general knowledge graph. We developed a pipeline to facilitate the linkage of these two tasks. Our relation classification method leverages graph representation learning on the shortest paths between entities in semantic dependency trees; it was shown to be comparable to state-of-the-art methods based on a corpus of labeled relations ($\text{AUC} = 97.6\%$). This classifier was then used to reduce claims from online discourse to semantic triples with an AUC of 95.8%; these were used as input to fact-checking algorithms to predict the accuracy of the claim. We achieved an AUC of 83% on our selected claims, which is at the least comparable to claim matching, but without the need for the corpus of existing claims that claim matching relies on.

Our relation extraction method is a promising approach to distinguishing relations present in large online discourse corpora; scaling up this algorithm could provide an outlet for modeling online discourse within an established ontology. Additionally, our pipeline may serve as a proof-of-concept for future research into automated fact-checking. While it is a challenge to model all possible relations in a generalistic ontology like DBpedia, this pipeline could form the basis of tools for reducing the time needed to research an online discourse claim.

Acknowledgements

The authors would like to thank Google for making publicly available both the GREC dataset and the Fact Check Explorer tool, and Alexios Mantzarlis for feedback on the manuscript.

REFERENCES

[1] Hunt Allcott and Matthew Gentzkow. 2017. Social media and fake news in the 2016 election. *Journal of economic perspectives* 31, 2 (2017), 211–36.

[2] Gabor Angeli and Christopher D. Manning. 2014. NaturalLI: Natural Logic Inference for Common Sense Reasoning. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, Doha, Qatar, 534–545. https://doi.org/10.3115/v1/D14-1059

[3] Livio Baldini Soares, Nicholas FitzGerald, Jeffrey Ling, and Tom Kwiatkowski. 2019. Matching the Blanks: Distributional Similarity for Relation Learning. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Florence, Italy, 2895–2905. https://doi.org/10.18653/v1/P19-1279

[4] Alberto Barron-Cedeno, Tamer Elsayed, Preslav Nakov, Giovanni Da San Martino, Maram Hasanain, Reem Suwaileh, and Fatima Haouari. 2020. CheckThat! at CLEF 2020: Enabling the Automatic Identification and Verification of Claims in Social Media. arXiv:2001.08546 [cs.CL]

[5] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating Embeddings for Modeling Multi-relational Data. In *Advances in Neural Information Processing Systems 26*, C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger (Eds.). Curran Associates, Inc., Red Hook, NY, United States, 2787–2795.

[6] B. Borel. 2016. *The Chicago Guide to Fact-Checking*. University of Chicago Press, Chicago, IL, USA.

[7] Alexandre Bovet and Hernán A. Makse. 2019. Influence of fake news in Twitter during the 2016 US presidential election. *Nature Communications* 10, 1 (Jan. 2019), 7. https://doi.org/10.1038/s41467-018-07761-2

[8] Lorenz Bühmann and Jens Lehmann. 2013. Pattern Based Knowledge Base Enrichment. In *The Semantic Web – ISWC 2013*, Harith Alani, Lalana Kagal, Achille Fokoue, Paul Groth, Chris Biemann, Josiane Xavier Parreira, Lora Aroyo, Natasha Noy, Chris Welty, and Krzysztof Janowicz (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 33–48.

[9] Razvan C. Bunescu and Raymond J. Mooney. 2005. A Shortest Path Dependency Kernel for Relation Extraction. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing* (Vancouver, British Columbia, Canada) *(HLT '05)*. Association for Computational Linguistics, USA, 724–731. https://doi.org/10.3115/1220575.1220666

[10] Giovanni Luca Ciampaglia. 2018. Fighting fake news: a role for computational social science in the fight against digital misinformation. *Journal of Computational Social Science* 1, 1 (29 Jan. 2018), 147–153. https://doi.org/10.1007/s42001-017-0005-6

[11] Giovanni Luca Ciampaglia, Prashant Shiralkar, Luis M. Rocha, Johan Bollen, Filippo Menczer, and Alessandro Flammini. 2015. Computational Fact Checking from Knowledge Networks. *PLOS ONE* 10, 6 (06 2015), 1–13. https://doi.org/10.1371/journal.pone.0128193

[12] Xin Dong, Evgeniy Gabrilovich, Geremy Heitz, Wilko Horn, Ni Lao, Kevin Murphy, Thomas Strohmann, Shaohua Sun, and Wei Zhang. 2014. Knowledge vault: A web-scale approach to probabilistic knowledge fusion. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Association for Computing Machinery, New York, New York, USA, 601–610. https://doi.org/10.1145/2623330.2623623

[13] Xin Luna Dong, Evgeniy Gabrilovich, Kevin Murphy, Van Dang, Wilko Horn, Camillo Lugaresi, Shaohua Sun, and Wei Zhang. 2015. Knowledge-Based Trust: Estimating the Trustworthiness of Web Sources. *Proc. VLDB Endow.* 8, 9 (May 2015), 938–949. https://doi.org/10.14778/2777598.2777603

[14] C. Fellbaum and G.A. Miller. 1998. *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, MA, USA.

[15] Kim Fridkin, Patrick J. Kenney, and Amanda Wintersieck. 2015. Liar, Liar, Pants on Fire: How Fact-Checking Influences Citizens' Reactions to Negative Advertising. *Political Communication* 32, 1 (Jan 2015), 127–151. https://doi.org/10.1080/10584609.2014.914613

[16] Luis Antonio Galárraga, Christina Teflioudi, Katja Hose, and Fabian Suchanek. 2013. AMIE: association rule mining under incomplete evidence in ontological knowledge bases. In *Proceedings of the 22nd international conference on World Wide Web - WWW '13*. ACM Press, New York, New York, USA, 413–422. https://doi.org/10.1145/2488388.2488425

[17] Aldo Gangemi, Valentina Presutti, Diego Reforgiato Recupero, Andrea Giovanni Nuzzolese, Francesco Draicchio, and Misael Mongiovì. 2017. Semantic Web Machine Reading with FRED. *Semantic Web* 8, 6 (2017), 873–893. https://doi.org/10.3233/SW-160240

[18] Matt Gardner and Tom Mitchell. 2015. Efficient and Expressive Knowledge Base Completion Using Subgraph Feature Extraction. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Lisbon, Portugal, 1488–1498. https://doi.org/10.18653/v1/D15-1173

[19] Nir Grinberg, Kenneth Joseph, Lisa Friedland, Briony Swire-Thompson, and David Lazer. 2019. Fake news on Twitter during the 2016 US presidential election. *Science* 363, 6425 (2019), 374–378.

[20] Aditya Grover and Jure Leskovec. 2016. Node2vec: Scalable feature learning for networks. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 13-17-Augu (2016), 855–864. https://doi.org/10.1145/2939672.2939754 arXiv:1607.00653

[21] Andrew M Guess, Brendan Nyhan, and Jason Reifler. 2020. Exposure to untrustworthy websites in the 2016 US election. *Nature human behaviour* 4, 5 (2020), 472–480.

[22] R. V. Guha, Dan Brickley, and Steve Macbeth. 2016. Schema.Org: Evolution of Structured Data on the Web. *Commun. ACM* 59, 2 (Jan. 2016), 44–51. https://doi.org/10.1145/2844544

[23] William L. Hamilton, Rex Ying, and Jure Leskovec. 2017. Inductive Representation Learning on Large Graphs. In *Proceedings of the 31st International Conference on Neural Information Processing Systems* (Long Beach, California, USA) *(NIPS'17)*. Curran Associates Inc., Red Hook, NY, USA, 1025–1035.

[24] Naeemul Hassan, Gensheng Zhang, Fatma Arslan, Josue Caraballo, Damian Jimenez, Siddhant Gawsane, Shohedul Hasan, Minumol Joseph, Aaditya Kulkarni, Anil Kumar Nayak, Vikas Sable, Chengkai Li, and Mark Tremayne. 2017. Claim buster: The firstever endtoend factchecking system. *Proceedings of the VLDB Endowment* 10, 12 (2017), 1945–1948. https://doi.org/10.14778/3137765.3137815

[25] Viet Phi Huynh and Paolo Papotti. 2019. A benchmark for fact checking algorithms built on knowledge bases. In *International Conference on Information and Knowledge Management, Proceedings*, Vol. 10. Association for Computing Machinery, New York, NY, USA, 689–698. https://doi.org/10.1145/3357384.3358036

[26] Krutika Kale. 2018. No, Tej Pratap Yadav Did Not Receive A Doctorate From Takshashila University. Online at https://www.boomlive.in/no-lalus-son-tej-pratap-did-not-receive-a-doctorate-from-takshashila-university/. Last accessed 2021-02-21.

[27] Sunny Lai, Kwong Sak Leung, and Yee Leung. 2018. SUNNYNLP at SemEval-2018 Task 10: A Support-Vector-Machine-Based Method for Detecting Semantic Difference using Taxonomy and Word Embedding Features. In *Proceedings of The 12th International Workshop on Semantic Evaluation*. Association for Computational Linguistics, New Orleans, Louisiana, 741–746. https://doi.org/10.18653/v1/S18-1118

[28] Ni Lao and William W. Cohen. 2010. Relational retrieval using a combination of path-constrained random walks. *Machine Learning* 81, 1 (2010), 53–67. https://doi.org/10.1007/s10994-010-5205-8

[29] Ni Lao, Tom Mitchell, and William W. Cohen. 2011. Random Walk Inference and Learning in A Large Scale Knowledge Base. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Edinburgh, Scotland, UK., 529–539.

[30] David MJ Lazer, Matthew A Baum, Yochai Benkler, Adam J Berinsky, Kelly M Greenhill, Filippo Menczer, Miriam J Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, et al. 2018. The science of fake news. *Science* 359, 6380 (2018), 1094–1096.

[31] Quoc Le and Tomas Mikolov. 2014. Distributed Representations of Sentences and Documents. In *Proceedings of the 31st International Conference on International Conference on Machine Learning - Volume 32 (ICML'14)*. JMLR.org, Beijing, China, II–1188–II–1196.

[32] Stephan Lewandowsky, Ullrich K. H. Ecker, Colleen M. Seifert, Norbert Schwarz, and John Cook. 2012. Misinformation and Its Correction: Continued Influence and Successful Debiasing. *Psychological Science in the Public Interest* 13, 3 (2012), 106–131. https://doi.org/10.1177/1529100612451018

[33] David Liben-Nowell and Jon Kleinberg. 2007. The Link-Prediction Problem for Social Networks. *Journal of the American society for Information Science and Technology* 58, 7 (2007), 1019–1031.

[34] Xiaomo Liu, Armineh Nourbakhsh, Quanzhi Li, Rui Fang, and Sameena Shah. 2015. Real-Time Rumor Debunking on Twitter. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management* (Melbourne, Australia) *(CIKM '15)*. Association for Computing Machinery, New York, NY, USA, 1867–1870. https://doi.org/10.1145/2806416.2806651

[35] T. Mitchell, W. Cohen, E. Hruschka, P. Talukdar, B. Yang, J. Betteridge, A. Carlson, B. Dalvi, M. Gardner, B. Kisiel, J. Krishnamurthy, N. Lao, K. Mazaitis, T. Mohamed, N. Nakashole, E. Platanios, A. Ritter, M. Samadi, B. Settles, R. Wang, D. Wijaya, A. Gupta, X. Chen, A. Saparov, M. Greaves, and J. Welling. 2018. Never-Ending Learning. *Commun. ACM* 61, 5 (April 2018), 103–115. https://doi.org/10.1145/3191513

[36] Brendan Nyhan and Jason Reifler. 2015. The Effect of Fact-Checking on Elites: A Field Experiment on U.S. State Legislators. *American Journal of Political Science* 59, 3 (2015), 628–640. https://doi.org/10.1111/ajps.12162

[37] Dave Orr. 2013. 50,000 Lessons on How to Read: a Relation Extraction Corpus. https://ai.googleblog.com/2013/04/50000-lessons-on-how-to-read-relation.html

[38] Stefano Ortona, Venkata Vamsikrishna Meduri, and Paolo Papotti. 2018. Robust discovery of positive and negative rules in knowledge bases. In *2018 IEEE 34th International Conference on Data Engineering (ICDE)* (Paris, France). IEEE, IEEE, Piscataway, NJ, USA, 1168–1179.

[39] Heiko Paulheim and Simone Paolo Ponzetto. 2013. Extending DBpedia with Wikipedia List Pages. In *Proceedings of the 2013th International Conference on NLP & DBpedia - Volume 1064* (Sydney, Australia) *(NLP-DBPEDIA'13)*. CEUR-WS.org, Aachen, DEU, 85–90.

[40] Hannah Rashkin, Eunsol Choi, Jin Yea Jang, Svitlana Volkova, and Yejin Choi. 2017. Truth of Varying Shades: Analyzing Language in Fake News and Political Fact-Checking. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Copenhagen, Denmark, 2931–2937. https://doi.org/10.18653/v1/D17-1317

[41] Bradley L. Richards and Raymond J. Mooney. 1992. Learning Relations by Pathfinding. In *Proceedings of the Tenth National Conference on Artificial Intelligence* (San Jose, California) *(AAAI'92)*. AAAI Press, Palo Alto, CA, USA, 50–55.

[42] Victoria L. Rubin, Yimin Chen, and Nadia K. Conroy. 2015. Deception detection for news: Three types of fakes. *Proceedings of the Association for Information Science and Technology* 52, 1 (2015), 1–4. https://doi.org/10.1002/pra2.2015.145052010083

[43] Victoria L. Rubin and Tatiana Vashchilko. 2012. Identification of Truth and Deception in Text: Application of Vector Space Model to Rhetorical Structure Theory. In *Proceedings of the Workshop on Computational Approaches to Deception Detection*. Association for Computational Linguistics, Avignon, France, 97–106.

[44] schema.org. 2020. ClaimReview schema. https://schema.org/ClaimReview

[45] Chengcheng Shao, Giovanni Luca Ciampaglia, Alessandro Flammini, and Filippo Menczer. 2016. Hoaxy: A Platform for Tracking Online Misinformation. In *Proceedings of the 25th International Conference Companion on World Wide Web* (Montréal, Québec, Canada) *(WWW '16 Companion)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 745–750. https://doi.org/10.1145/2872518.2890098

[46] Chengcheng Shao, Giovanni Luca Ciampaglia, Onur Varol, Kai-Cheng Yang, Alessandro Flammini, and Filippo Menczer. 2018. The spread of low-credibility content by social bots. *Nature communications* 9, 1 (2018), 1–9.

[47] Baoxu Shi and Tim Weninger. 2016. Discriminative predicate path mining for fact checking in knowledge graphs. *Knowledge-Based Systems* 104 (Jul 2016), 123–133. https://doi.org/10.1016/j.knosys.2016.04.015 arXiv:1510.05911

[48] Prashant Shiralkar, Alessandro Flammini, Filippo Menczer, and Giovanni Luca Ciampaglia. 2017. Finding Streams in Knowledge Graphs to Support Fact Checking. In *2017 IEEE International Conference on Data Mining (ICDM)* (New Orleans, Louisiana, USA). IEEE, Piscataway, NJ, 859–864. https://doi.org/10.1109/ICDM.2017.105 arXiv:1708.07239 [cs.AI] Extended Version.

[49] Kai Shu, Deepak Mahudeswaran, Suhang Wang, and Huan Liu. 2020. Hierarchical propagation networks for fake news detection: Investigation and exploitation. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 14. AAAI, Palo Alto, CA, USA, 626–637.

[50] Kai Shu, Suhang Wang, and Huan Liu. 2018. Understanding User Profiles on Social Media for Fake News Detection. In *IEEE 1st Conference on Multimedia Information Processing and Retrieval (MIPR 2018)*. IEEE, Piscataway, NJ, USA, 430–435. https://doi.org/10.1109/MIPR.2018.00092

[51] Craig Silverman (Ed.). 2014. *Verification Handbook*. European Journalism Center, Maastricht, the Netherlands.

[52] Sameer Singh, Sebastian Riedel, Brian Martin, Jiaping Zheng, and Andrew McCallum. 2013. Joint Inference of Entities, Relations, and Coreference. In *Proceedings of the 2013 Workshop on Automated Knowledge Base Construction* (San Francisco, California, USA) *(AKBC '13)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/2509558.2509559

[53] Richard Socher, Danqi Chen, Christopher D Manning, and Andrew Ng. 2013. Reasoning With Neural Tensor Networks for Knowledge Base Completion. In *Advances in Neural Information Processing Systems*, C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger (Eds.), Vol. 26. Curran Associates, Inc., 57 Morehouse Lane, Red Hook, NY, United States, 926–934. https://proceedings.neurips.cc/paper/2013/file/b337e84de8752b27eda3a12363109e80-Paper.pdf

[54] James Thorne, Andreas Vlachos, Christos Christodoulopoulos, and Arpit Mittal. 2018. FEVER: a Large-scale Dataset for Fact Extraction and VERification. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*. Association for Computational Linguistics, Stroudsburg, PA, USA, 809–819. https://doi.org/10.18653/v1/N18-1074

[55] Sebastian Tschiatschek, Adish Singla, Manuel Gomez Rodriguez, Arpit Merchant, and Andreas Krause. 2018. Fake News Detection in Social Networks via Crowd Signals. In *Companion Proceedings of the The Web Conference 2018* (Lyon, France) *(WWW '18)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 517–524. https://doi.org/10.1145/3184558.3188722

[56] Andreas Vlachos and Sebastian Riedel. 2014. Fact Checking: Task definition and dataset construction. In *Proceedings of the ACL 2014 Workshop on Language Technologies and Computational Social Science*. Association for Computational Linguistics, Baltimore, MD, USA, 18–22. https://doi.org/10.3115/v1/W14-2508

[57] Soroush Vosoughi, Deb Roy, and Sinan Aral. 2018. The spread of true and false news online. *Science* 359, 6380 (2018), 1146–1151. https://doi.org/10.1126/science.aap9559

[58] Linlin Wang, Zhu Cao, Gerard de Melo, and Zhiyuan Liu. 2016. Relation Classification via Multi-Level Attention CNNs. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Berlin, Germany, 1298–1307. https://doi.org/10.18653/v1/P16-1123

[59] Claire Wardle and Hossein Derakhshan. 2017. *Information disorder: Toward an interdisciplinary framework for research and policy making*. Technical Report. Council of Europe Report.

[60] Jen Weedon, William Nuland, and Alex Stamos. 2017. *Information Operations and Facebook*. Technical Report. Facebook, Inc.

[61] Gerhard Weikum and Martin Theobald. 2010. From Information to Knowledge: Harvesting Entities and Relationships from Web Sources. In *Proceedings of the Twenty-Ninth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems* (Indianapolis, Indiana, USA) *(PODS '10)*. Association for Computing Machinery, New York, NY, USA, 65–76. https://doi.org/10.1145/1807085.1807097

[62] Shanchan Wu and Yifan He. 2019. Enriching pre-trained language model with entity information for relation classification. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. 2361–2364.

[63] You Wu, Pankaj K. Agarwal, Chengkai Li, Jun Yang, and Cong Yu. 2014. Toward Computational Fact-Checking. *Proc. VLDB Endow.* 7, 7 (March 2014), 589–600. https://doi.org/10.14778/2732286.2732295

[64] Minguang Xiao and Cong Liu. 2016. Semantic Relation Classification via Hierarchical Recurrent Neural Network with Attention. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*. The COLING 2016 Organizing Committee, Osaka, Japan, 1254–1263.

[65] Peng Xu and Denilson Barbosa. 2019. Connecting Language and Knowledge with Heterogeneous Representations for Neural Relation Extraction. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Association for Computational Linguistics, Minneapolis, Minnesota, 3201–3206. https://doi.org/10.18653/v1/N19-1323

[66] Ikuya Yamada, Hiroyuki Shindo, Hideaki Takeda, and Yoshiyasu Takefuji. 2016. Joint Learning of the Embedding of Words and Entities for Named Entity Disambiguation. In *Proceedings of The 20th SIGNLL Conference on Computational Natural Language Learning*. Association for Computational Linguistics, 209 N. Eighth Street, Stroudsburg PA 18360, USA, 250–259.

[67] Ran Yu, Ujwal Gadiraju, Besnik Fetahu, Oliver Lehmberg, Dominique Ritze, and Stefan DIetze. 2018. KnowMore - Knowledge base augmentation with structured web markup. , 159–180 pages. https://doi.org/10.3233/SW-180304

[68] Yuhao Zhang, Victor Zhong, Danqi Chen, Gabor Angeli, and Christopher D. Manning. 2017. Position-aware Attention and Supervised Data Improve Slot Filling. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Copenhagen, Denmark, 35–45. https://doi.org/10.18653/v1/D17-1004

[69] Zhao, Zilong, Zhao, Jichang, Sano, Yukie, Levy, Orr, Takayasu, Hideki, Takayasu, Misako, Li, Daqing, Wu, Junjie, and Havlin, Shlomo. 2020. Fake news propagates differently from real news even at early stages of spreading. *EPJ Data Sci.* 9, 1 (2020), 7. https://doi.org/10.1140/epjds/s13688-020-00224-z

A SELECTED CLAIMREVIEW CLAIMS

Table 7: Selected ClaimReview claims, the relation they contain, and the relation predicted by the model. The text bold indicates the entities participating in the relation. The AUC of the relation classification task is 0.958.

ID	Claim	Actual*	Predicted*	Rating	Claim ≡ Triple
1	**Malaysian**-born Senator **Penny Wong** ineligible for Australian parliament	POB	DOB	False	
2	Donald Trump says **President Obama**'s grandmother in Kenya said he was born in **Kenya** and she was there and witnessed the birth.	POB	Institution	False	✓
3	Donald Trump says his father, **Fred Trump**, was born in a very wonderful place in **Germany**.	POB	POB	False	✓
4	**Barack Obama** was born in the **United States**.	POB	POB	True	✓
5	**Barron Trump** was born in **March 2006** and Melania wasn't a legal citizen until July 2006. So under this executive order, his own son wouldn't be an American citizen.	DOB	POB	False	
6	**Isabelle Duterte** was born on **January 26, 2002**, which makes her only 15 years old today.	DOB	DOB	False	
7	**Tej Pratap Yadav** receives a **doctorate degree** from Takshsila University in Bihar	education	education	False	✓
8	**Smriti Irani** has a **MA degree**.	education	institution	False	✓
9	**Melania Trump** lied under oath in 2013 about graduating from college with a **bachelor's degree** in architecture.	education	institution	False	
10	Did **Michelle Obama** recently earn a **doctorate degree** in law?	education	education	False	✓
11	**Pravin Gordhan** does not have a **degree**.	education	education	False	✓
12	**Alexandria Ocasio-Cortez**'s **economics degree** recalled.	education	institution	False	✓
13	Ilocos Norte Governor **Imee Marcos** claimed on January 16 that she earned a **degree** from Princeton University.	education	education	False	
14	Ilocos Norte Governor **Imee Marcos** claimed on January 16 that she earned a degree from **Princeton University**.	institution	institution	False	✓
15	**Tej Pratap Yadav** receives a doctorate degree from **Takshsila University** in Bihar.	institution	education	False	
16	**Patrick Murphy** embellished, according to reports, his **University of Miami** academic achievement.	institution	institution	True	
17	Mahmoud Abbas, Ali Khamenei, and **Vladimir Putin** met each other in the class of 1968 at **Patrice Lumumba University** in Moscow	institution	institution	False	
18	**Mahmoud Abbas**, Ali Khamenei, and Vladimir Putin met each other in the class of 1968 at **Patrice Lumumba University** in Moscow	institution	institution	False	
19	Mahmoud Abbas, **Ali Khamenei**, and Vladimir Putin met each other in the class of 1968 at **Patrice Lumumba University** in Moscow	institution	institution	False	
20	**Maria Butina** is a human rights activist, a student of the **American University**, and the most relevant is that she is a person who did not work (collaborate) with the Russian state bodies.	institution	institution	False	
21	Ilocos Norte Governor **Imee Marcos** graduated cum laude from the **University of the Philippines** (UP) College of Law.	institution	institution	False	
22	**David Hogg** graduated from **Redondo Shores High School** in 2015.	institution	institution	False	✓
23	Sadhvi Pragya Singh Thakur said **Manohar Parrikar** died of cancer because he allowed the consumption of beef in **Goa**.	POD	POD	False	
24	Fox star **Tucker Carlson** in critical condition (then died) after head on collision driving home in **Washington D.C.**	POD	POD	False	✓
25	**Nasser Al Kharafi** died in **Kuwait**.	POD	POD	False	✓
26	DCP **Amit Sharma** passed away in **Delhi** riots	POD	institution	False	✓
27	It is being claimed that **Jason Statham** was murdered at his home in **New York** by assailants who broke into his mansion.	POD	POD	False	
28	Actor **Robert Downey Jr.** died in a car crash stunt in **Hollywood** on July 8.	POD	POD	False	

* DOB = Date of Birth, POB = Place of Birth, POD = Place of Death