DEVELOPING RESPONSIBLE AND RELIABLE ENTITY AND RELATION EXTRACTION METHODS: A HUMAN-IN-THE-LOOP PERSPECTIVE

December 15, 2023

BY

PINGJING YANG

Email: py2@illinois.edu
University of Illinois Urbana-Champaign

Research Proposal for Preliminary Exam

Doctoral Committee:

Professor Jana Diesner, School of Information Sciences, Chair
Professor Jennifer Cromley, College of Education
Professor Halil Kilicoglu, School of Information Sciences
Professor Unnati Narang, Gies College of Business

December 15, 2023

ABSTRACT

In this digital age, information extraction is still a fundamental task as the basis of more complex and user-oriented tasks, such as chatbots and search engines. Although efforts have been made in this area, adopting information extraction for a broad audience is still an underdeveloped problem. Both algorithm practitioners and end users still lack easy-to-access guidelines for using information extraction tools. This proposal aims to advance information extraction application research by developing generic and domain-specific (e.g., education science) information extraction tools to enable practitioners to apply relation extraction on their own corpus. This proposal starts with analyzing needs expressed by online social media users during an urgent social event, the Ukraine-Russia Conflict, using Twitter (X.com) data. Then, I will systematically evaluate state-of-the-art relation extraction approaches on public relation extraction datasets, including newswire, wiki, and web text. Finally, I will apply optimized models developed on public datasets to domain-specific text data in education science and involve education experts in the development life cycle. I expect to develop guidelines for practitioners to leverage better relation extraction tools on different types of corpus and insights on involving human experts in the loop of algorithm method development.

December 15, 2023

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1 Background

Relation extraction is a classic task within the realm of Information Extraction. It involves the process of transforming valuable information from text data into structured formats that information products can readily consume. A standard relation extraction task examines related text spans, often referred to as entities, and determines how they meaningfully relate to each other.

To date, the widespread adoption of relation extraction continues to face challenges regarding algorithms and datasets across different domains. This makes democratization of A.I. challenging, specifically the goal of enabling everyone to create artificial intelligence systems using state-of-the-art algorithms [39]. For example, Bassignana et al. [6] contend that relation extraction (RE) datasets are not subjected to fair comparisons because many researchers fail to present datasets and tasks in a manner that facilitates accurate comparisons. This affects the accuracy of method evaluations. Additionally, there is still a lack of robust scientific evaluations to support the selection of suitable algorithms for specific domains to achieve best performance [90]. The reported performance inconsistencies of these algorithms mean that developers face the challenge of making informed choices without extensive research.

In this dissertation, I intend to explore theories and applications of relation extraction algorithms, particularly focusing on their use in social network analysis, which utilizes relational data to build networks. I aim to unearth method guidelines for relation extraction across domains. I plan to address subsequent research questions to achieve the goal.

1.2 Research questions

RQ1: How can we reliably extract relations from social media platforms to detect the needs expressed during crisis events?

Social media serve as public arenas where people can swiftly respond to unfolding events [79, 69]. I will delve into the use of relation extraction for crisis response and management. I will use Twitter data from various events, such as the COVID-19 pandemic, Ukraine-Russian Conflicts of 2022, and the Haiti earthquake of 2010, to develop and test relation extraction methods. Initial findings suggest that certain empirical methods combining word2vec and syntax parsing perform notably well on Twitter data. I plan to concentrate on detecting users' needs and empower stakeholders, including policymakers, social media analysts, and the general public, to gain awareness of crises.

RQ2: Which factors contribute to the reliability of relation extraction for constructing social networks? What additional factors influence the performance of relation extraction?

This question focuses on the core application of relation extraction within the broader context of social network analysis. I will scrutinize various information extraction datasets, such as SemEval [11] and ACE [41], and the social relations between standard social entities like individuals and organizations. I will use control experiment methods to systematically evaluate relation extraction methods on datasets from various sources, ranging from rule-based to learning-based approaches. Preliminary findings suggest that the choice of models and parameters significantly impacts the results of relation extraction for different types of corpus.

RQ3: How can we extract information from think-aloud and textbook data to improve the evaluation of study outcomes in the area of education science?

Transitioning from the general domain, I will work on employing relation extraction to tackle the challenges of reliably constructing semantic networks in the field of education science. Given that the input data come from textbooks and interviews, I will develop relation extraction algorithms to discern relations from both types of sources. I will then compare the identified relations and devise semi-automated means to augment human evaluations of learning outcomes.

CHAPTER 2

RELIABLE INFORMATION EXTRACTION FOR SOCIAL MEDIA

In this chapter, I will discuss a social media analysis project that focuses on extracting relational information for disaster-related events. This project explores the research area on how to extract relation information from social media in a more accurate and transparent way [85]. We aim to provide insights for better information extraction from social media platforms. At the end of this chapter, I will report the current progress and future plan.

2.1 Disclaimer

Part of the work in this chapter is a follow-up work of papers accepted or submitted to conferences with the approval of authors and copyright.

- M. Janina Sarol, Ly Dinh, Rezvaneh Rezapour, Chieh-Li Chin, Pingjing Yang, and Jana Diesner. 2020. An Empirical Methodology for Detecting and Prioritizing Needs during Crisis Events. In Findings of the Association for Computational Linguistics: EMNLP 2020, pages 4102–4107, Online. Association for Computational Linguistics. [72]

- Pingjing Yang, Ly Dinh, Alex Stratton, and Jana Diesner. 2024. Detection, Categorization, and Comparison of Needs Expressed during Crises on Twitter. ICWSM 2024 (in submission).

For work [72], I contributed to the development of ideas, the coding of data, and the writing and revision of the manuscript. For the other work, I innovate and lead the project, design and conduct data analysis, and the writing and revision of the manuscipt.

2.2 Introduction

As modern civic squares, social media platforms such as Twitter and Facebook have evolved to be opinion centers for people with different backgrounds to exchange and share information, especially for emergency events [18, 77]. For almost every single public emergency event, social media often emerges as the first source to collect and assess public opinions, such as infectious diseases [47], typhoons [50], hurricanes [70], explosions [15], earthquakes [96], and floods [77, 16]. As a result, studying social media becomes an invaluable probe to inspect and understand social opinions in a broad sense.

Extracting information regarding emergency events requires both efficiency and performance considerations. As events often take place suddenly, related information may stream into into social media in a short time. Training specific complex models to uncover useful information is challenging as the data volume becomes extremely large. Meanwhile, emergency events also demand higher accuracy and precision as the extracted information may be critical for timely responses. As a result, extracting information both efficiently and accurately become a crucial requirement for extracting emergency event related information.

Additionally, extracting information for emergency events also requires domain knowledge in the process of algorithm development. For example, when developing a algorithm for detecting resource needs for disaster response, such as food and shelter, categories of needs will be used for grouping information. However, disaster response guidelines proposed by Federal Emergency Management Agency (FEMA) may have different focus compared to United Nations Crisis Relief, which may lead the grouping of needs varying significantly. Thus, adopting a human-in-the-loop strategy to guide developing useful information extraction approaches is necessary.

In this project, I will use **needs detection** method as an example to study approaches to extract needs expressed by Twitter users reliably from social media with respect to different emergency events, including earthquake (Haiti), pandemic (COVID-19), and war (Ukraine-Russian Conflict). Through identifying and categorizing types of needs, subjects asking for resources, and specific needs mentioned from social media platform – Twitter. I aim to deepen understanding of information extraction for needs detection on social media platforms.

2.3 Aims and Objectives

The goal of this project is to:

- Develop and evaluate needs detection datasets and methods for social media information extraction.
- Compare and understand need differences with respect to emergency event types.

2.4 Potential Contribution

By studying and evaluating needs detection approaches, I aim to provide practical suggestions on applying information extraction on social media data, uncover the difference of rule-based and learning-based approaches, and impact of algorithms on extracted information. This project will contribute to the development of more reliable information extraction applications on Social Media data by demonstrating effective and explainable approaches.

2.5 Related Work

Prior literature in crisis informatics has shown how NLP methods can help to detect crisis-related information such as information about affected individuals and their locations [59], infrastructural damages [3], and immediate needs expressed by vulnerable populations [71, 65, 8]. Such information can help crisis responders in their work and to facilitate the review of response effectiveness.

A common theme across these studies is the focus on detecting reliable information about affected individuals and/ or communities so that responders can prioritize relief efforts with respect to the urgency of a need [87, 71]. For instance, [71] in their examination of needs during the 2010 Haiti earthquake found that priorities related to livelihoods such as food, water, and medical equipment are salient in tweets. Another analysis of tweets about eight natural disaster events by [48] also found that 22% of the expressed urgent needs pertained to essentials such as food, water, clothing, and medical aid. Other studies have matched urgent needs to resource availabilities as uttered in social media content, which helps to match needs to supplies [65, 7]. Purohit et al. [65] built a binary classifier that separates tweets into either requests for or offers of help with respect to resources such as money, clothing, food, medical aid, shelter, volunteer work, and achieved accuracy of 82% and 75% for request and offer, respectively. [7] identified top needs and availabilities of resources to support humanitarian assistance and disaster relief work during the 2015 Nepal Earthquake using word embeddings and extracted words closest to the terms "needs" and "supplies". Their proposed methodology for matching needs and availabilities achieved an accuracy of 86%. Overall, these studies have demonstrated the feasibility and relevance of detecting needs as well as resources available to meet these needs from social media content generated during crisis events. In addition to these studies, I will expand the list of terms used for extracting needs and update the word embedding approaches to incorporate more cutting-edge, large language models.

Existing literature on needs detection has primarily focused on natural crises such as earthquakes and floods, with limited research examining the efficacy of the method in other types of crises. Given that each crisis type requires a unique sets of resources and response activities [37, 19], there is a reason to suspect that needs would widely vary across different crisis types. For example, a large-scale natural disaster like a hurricane or earthquake may require more resources such as medical supplies, food, and shelter [71, 72], while a man-made disaster such as a terrorist attack may require more safety and security resources [64, 62]. To test whether needs expressed on social media differ across crisis types, I will examine how the tweets about 2022 Ukraine-Russia conflict differ from tweets about the 2010 Haiti earthquake, and tweets about the COVID-19 pandemic using the same needs classifications, including OCHA and FEMA to identify any missing classifications in existing frameworks.

The ongoing 2022 Ukraine-Russia conflict is a prime example of a crisis event where social media plays a major role for citizens and netizens to share and get information about the situation on-site in the Ukraine, [82, 43], for family members to connect with each other [75], and for political groups to launch campaigns (e.g. *Ghost of Kyiv*), with the latter possibly contributing to online disinformation [80]. The negative impact of the war, such as lack of adequate housing and healthcare for Ukrainian refugees [56] and disruptions in the global supply chains for food [9] and energy [60] has continued since February 2022. According to prior literature, timely detection and prioritization of resource needed is crucial to reduce harmful effects of a crisis [45, 3]. I aim to contribute to a better understanding of this and possibly future crisis by using NLP and classification mapping to detect needs with high urgency. I will further expand needs detection by tracing changes in needs over time, and determine if there are certain needs which might be missing from existing disaster response guidelines (e.g., financial).

I will examine the utility of our extended needs detection method by comparing the top needs detected for the Ukraine-Russia conflict to 2010 Haiti earthquake and COVID-19 pandemic. Based on prior literature, needs during the 2010 Haiti earthquake response were centered around medical resources for the affected communities, as well as immediate humanitarian assistance. As for COVID-19 pandemic, the needs were primarily medical supplies and personal protective equipment. I expect that the needs would significantly vary across these three events, but there may also be some overlap in terms of needs relating to safety of the affected population.

2.6 Research Design and Method

In this project, I will focus on the Ukraine-Russian war event as an example to study needs expressed on social media. I choose this setting for the following reasons: 1. As an ongoing emergency public event, Ukraine-Russian conflict has the complexity of large-scale multi-stakeholder events, which represents an impactful public social media event. 2. Needs detection is a novel information extraction task that needs more exploration and development. Furthermore, I will compare the results from Ukraine-Russian war with previous events, such as COVID-19 pandemic and 2010 Haiti earthquake with respect to the types of needs and stakeholders. More specifically, I attempt to answer the following research questions:

- RQ1: What needs do Twitter users express in the context of the Ukraine-Russia conflict? And which approaches perform better in terms of accuracy and recall?

- RQ2: How do these needs change over time and across different regions?

- RQ3: How do needs expressed in a warfare situation compare to needs in other types of disaster, namely natural disasters and a biological crisis?

2.6.1 Data

I will use a dataset from Chen et al.[17]. Their dataset contains 454,488,445 tweet Ids collected from February 22, 2022, through October 01, 2022. The tweets were pulled based on the existence of a keyword related to the Russia-Ukraine conflict. Examples of these keywords include, but are not limited to: *Ukraine, Russia, Zelensky.* 70.65% or roughly 321 million tweets were in English. Tweet Ids are a numeric code that can be used to request available information tied to a specific Tweet.

To obtain the tweet text and available meta data, I will use Twarc for rehydration; a process in which the Twitter API is used to pull a tweet given its unique key identifier. An Id is passed in and a Tweet is returned in a nested json format. Though Twitter has enforced a new API policy for users, I pulled a subset of the twitter dataset before the change. To preserve the distribution of tweets, I will the number of Ids for each day, multiply that number by 10 million, and divide the result by the total size of the data set. I then randomly sample that percentage of Ids from the corresponding day. This procedure will ensure that we maintain the original distribution of Reddit posts in a small sample, which is still representative.

The collection of the subset of Twitter data took four days. After removing non-English and deleted tweets, 5,822,234 tweets remained. The dataset includes tweets posted by 2,310,347 users from 2/25/2022 (the second day after the outbreak of the war) to 9/30/2022.

To estimate the percentage of Tweets posted by bots, I further sampled 1,000 Twitter accounts and used their corresponding metadata and tweets to predict the possibility of being automated accounts. Botometer API [73] was used and found that 82% accounts are not bots (as the distribution of scores indicates a cut-off with a threshold of 0.7 [93]). As the Botometer has a limit number of API call per account, I can not verify all the Twitter accounts used in this study, which is a limitation of our study.

To compare needs expressed and identified across different disaster contexts, I will implement needs detection on two existing datasets: COVID-19 tweets, and 2010 Haiti earthquake tweets. The COVID-19 tweets dataset contains 665,667 English tweets that were posted from February 28 to May 8, 2020. The authors collected tweets that contained at least one of the hashtags, such as #COVID19, #COVID-19, #coronavirusoutbreak.

The 2010 Haiti earthquake dataset contains 54,660 English tweets that were posted between January 12 and June 1, 2010, which corresponded to the immediate response to early recovery phases of earthquake response. The authors collected tweets that contained the keyword "haiti earthquake", or hashtags #haiti, #haitiearthquake during the specified timeframe of data collection.

2.6.2 Extraction of Needs

Word-embedding Approach. To determine what needs were expressed, I will apply a word embedding-based approach to extract a list of needs using the seed words "needs" and "supplies", following [72, 70]. The approach involves four steps:

1. Applying phrase detection models to annotate the dataset using AutoPhrase [76] method with a threshold of 0.8 to ensure models extract not only words but also phrases.

2. Removing @RT_username, replacing full url links with "URL", splitting tweets into sentences, and tokenizing sentences.

3. Applying word2vec model on the sentences from Step 2.

4. Retrieving and list-ranking the top nouns closest to the word embeddings of needs and supplies based on cosine similarity. I manually checked the top 500 nouns and noticed after the top 100 nouns, the accuracy of nouns decreased significantly. Thus, I chose the top 100 nouns for follow-up analysis.

For step three and four of the approach, I also tested recent state-of-the-art pre-trained language models in creating word embeddings and calculating cosine similarities. Two models are selected for comparison: BERT from Google [49] and TwHIN-BERT [98]. BERT is the first Transformers-based [49] bidirectional encoder representation model widely adopted by NLP practitioners. It is trained on the BooksCorpus (800M words) and English Wikipedia text (2,500M words) and have achieved state-of-the-art results on eleven natural langauge processing tasks, including Semantic Textual Similarity Benchmark. TwHIN-Bert is a BERT model using the same Transformer architecture as BERT [49] with different training data from Twitter. Since the experimental dataset is from Twitter platform, using a BERT model pre-trained with Twitter data source may improve the accuracy of semantics. However, after extracted top need nouns, I found BERT models do not show advantages over word2vec model that I trained for this study. I validated top 50 most similar needs from the generated list, BERT from Google generated 50%, and TwHIN-BERT generated 26%, and the word2vec method generated 92% valid priority needs. This observation shows small model trained for our task can achieve a better performance than the state-of-art large language models.

Need Categorization. I will further apply open and axial coding to categorize expressions of needs within their context of use. First, two coders independently labeled our top 100 need words into categories defined by each coder, along with notes on their rationale for each (category and) assignment. Both coders leveraged the same two existing categorizations of needs, namely ([58]'s seven categories of needs and [38]'s seven community lifelines) as reference points to develop their own self-generated categories. The categories from OCHA are: *Education, Protection, Food Security and Livelihoods, Shelter and Non-Food Items, Health, Water,* and *Sanitation and Hygiene (WASH)*. The seven categories from FEMA are: *Safety and Security, Food and Water and Shelter, Health and Medical, Energy, Communications, Transportation,* and *Hazardous Materials*. These two schemas were chosen as reference points as FEMA's community lifelines has been the standard for U.S. governmental agencies to prioritize response efforts, and OCHA's needs schema was created based on events that happened (and are happening) in Ukraine as a result of the conflict. The two coders then cross-validated their annotations and reconciled their disagreements until all annotations aligned.

Extraction of Who-Need-What. In addition to extracting and categorizing needs, I will also apply a rule-based approach to extract {who, need, what} triples [72], where who represents requesters and what represents resources and/ or targets. I will use the syntactic structure of sentences to extract these triples. The method involves the following steps:

1. Apply dependency parser from spaCy [46] to tokenized sentences.
2. Find "need" terms in sentences, which include word forms like need, needs, needed.
3. Extract subject (who) and direct object (what) from sentences where the "need" term is in the form of a verb.
4. If the "need" term is a noun, extract the descendant of the need term (what), which links through a preposition, and the copular verb of the need term. We use the left child of the copular verb term as who term. For example, "Ukraine is in need of weapons."

We saw that different from other types of disasters, such as natural disasters or pandemics, warfare includes a more complex set of stakeholders with diverse needs. We used the extracted who-need-what triples to study who the stakeholders are and what needs they express. We expand other terms that also indicate needs, such as "demand" and "request". We report and compare the outputs based on these terms in addition to the "need" term.

Additionally, since generative AI solutions such as ChatGPT have recently achieved impressive performance in information extraction [42], I also tested the who-needs-what extraction task in the most recent version of ChatGPT-4. Results were pulled on 11 September 2023 on a paid ChatGPT account. The prompts designed in this study consist of elements including task instruction, demonstration examples, and input text, following the prompt design in [42]. *Prompt: Do not show code; extract who-need-what triple information from the following text data. Each row represents a sentence. For example, for the text "We need UNESCO's help!", the triple should be "we-need-help".*

The ChatGPT output who-needs-what triples achieved a comparative result with our rule-based approach. After inspecting randomly sampled 50 who-needs-what triples, I identified three cases that ChatGPT cannot correctly extract. For example, for text "Hungary, please we need a total EmbargoRussianOilandGas!" ChatGPT only identified a total as an object, while our approach correctly identified EmbargoRussianOilandGas. This error implies the lack of ability for ChatGPT to extract nonstandard English content in Twitter context. Therefore, we use our rule-based approach for the following analysis.

2.7 Progress

For now, I have finished using traditional word2vec+syntax parsing models to analyze social media data from various disaster events. Our recent paper **Detection, Categorization, and Comparison of Needs Expressed during Crises on Twitter.** submitted to ICWSM-2024 received a 'Revise and Resubmit' decision, with most reviewers providing positive feedback. I will follow the main review feedback to include more validation of the proposed approaches and resubmit the paper to ICWSM-2024 January round. Furthermore, I will replace word2vec with more cutting-edge language models and conduct a more comprehensive comparison of approaches.

CHAPTER 3

RELIABLE RELATION EXTRACTION FOR TEXT NETWORK ANALYSIS

In this chapter, I will discuss an empirical study which systematically evaluates relation extraction approaches on various types of corpora and relations. This project explores the research area of text network constructions by applying different types of relation extraction algorithms, including distance-based, rule-based, and learning-based approaches. The insights of the study will be used to advance more reliable text network construction and analysis. At the end of this chapter, I will report current progress and submission plan.

3.1 Disclaimer

This chapter is based on several published work or presentations in conferences. The use of these content will not violate any copyright requirements and get approval from authors.

- Ly Dinh, Sumeet Kulkarni, Ping-Jing Yang, Jana Diesner. Reliability of Methods for Extracting Collaboration Networks from Crisis-related Situational Reports and Tweets. Proceedings of Information Systems for Crisis Response and Management (ISCRAM2022). [35]

- Ping-Jing Yang, Janina Sarol, Ly Dinh, & Jana Diesner. (2021). Reliable relation extraction for social network construction. Presentation at the North American Regional Social Networks Conference 2021 (NASN2021) (held online). [95]

- Ping-Jing Yang, Janina Sarol, Ly Dinh, & Jana Diesner. (2020, 2021). Annotation guidelines for entity tagging and semantic role labeling of disaster-related text documents. Critical Infrastructure Resilience Institute 11th Maritime Risk Symposium (held online) AND at US Department of Homeland Security Centers of Excellence (COE) Summit. Fairfax, Virginia (held online). [94]

I contributed to the development of automatic solution for information extraction for the first work. For the submissions to conferences, I lead these projects and was in charge of ideation, implementations and presentations of the work.

3.2 Introduction

Text network analysis is one approach to encoding relationships between concepts in a text and building a network of linked concepts [13]. The assumption behind this analysis is that words in a text are interconnected [12], and the structure of networks can provide insight into themes [81]. Thus, the structure and change of networks provide people lenses to study underlying social cognition, such as team situation awareness [92, 21], mental models of specific topic [33], and impacts of documents [32, 34].

Existing text network construction often uses distance or proximity-based approaches as well as semantic, syntactic, and statistical features to find co-occurrence of text units from the corpus [22, 29, 74]. For example, in constructing a Person-to-Person network from news articles, a common method involves creating an edge between any two Person entities if they appear in the same sentence. However, practitioners often use arbitrary distance windows to decide whether two text units are related, which may be problematic in practice [20]. With the development of relation extraction (RE) algorithms, adapting RE algorithms to the construction of text networks seems promising but still lack evaluations in reliable ways.

From an NLP perspective, the construction of text network includes the following steps: 1. identification of entities (NER); 2. detection of relations (RE) [14, 86]. NER tasks refer to the process of detecting the boundary and class of text units that vary from simple words to phrases, which refer to self-consistent text span [26]. RE tasks refer to the extraction of relations between text spans identified in NER tasks. These relations can be either directed or undirected, and they may have a specific type (e.g., employment) or simply a boolean value (e.g., related/not related). More sophisticated networks even define various properties such as temporal or spatial information for relations [28].

The construction of such networks requires significant effort to analyze text, resources to verify networks, and improvement in terms of reliability and accuracy [24]. Subtle change of parameter choices in subtasks in network construction may unexpectedly influence research outcome, such as entity recognition [24, 23, 83], and co-reference resolution [27]. Addressing these issues requires a systematic investigation of datasets and application domains [24].

Furthermore, as social networks become abundant sources for text data, the integration of text analysis into online social network analysis can reveal and verify classic social science theories on a large scale [31]. An immediate challenge is to verify and test the accuracy of these networks derived from online social media data, which is impractical if manually checked [29]. To approach this challenge, a systematic review of factors affecting network constructions will potentially enable practitioners to construct networks more responsibly.

In this work, I will conduct an empirical study on the construction of text networks from a single document or sets of documents. The goal of this study is to recommend empirically-tested practices in constructing text networks from text data depending on precision and recall rates. Notably, this study does not intend to create new models, but the insights gained may guide the selection of features for network construction processes.

3.3 Aims and Objectives

The goal of this project is to:

- Implement and evaluate relation extraction approaches for the construction of text networks, in terms of precision, accuracy and recall.
- Compare and understand how factors, such as sources of text (e.g., newswire or wiki), types of relations (e.g., relation with specific definition or simply related/not related), and nodes of networks (e.g., location and person), affect the performance of relation extraction approaches.

- Develop a guideline to suggest model and parameter selection for relation extractions for text network constructions depending on the features tested.

3.4 Potential Contribution

By studying and evaluating relation extraction approaches, I plan to provide a systematic review of existing relation extraction approaches on various ground-truth relation annotation corpus. The results will contribute to the selection and use of relation extraction approaches for text network construction tasks and make future Text network analysis more reliable and reproducible. The direct outcome of this project includes a guideline for practitioners to select models and parameter settings based on properties of text and networks. To our knowledge, this project is the first effort to systematically study the correlations between relation extraction with text network constructions.

3.4.1 Literature Review

3.4.2 Using Text data to Construct Networks

Social and political scientists often develop and apply codebooks to extract meaningful information from text data to construct nodes in networks [25, 5]. These codebooks include detailed manual rules for creating nodes from text data. Typical nodes can be specific instance or categorizations representing groups of concepts [30] depending on the area of study. Typically, codebooks are created in a manual or semi-automated process [26], allowing for integrating expert knowledge, and human verification at the cost of scalability and efficiency [22]. On the other hand, automated approaches, including natural langauge processing, also serve as a solution for generating such nodes for network [26]. For this project, we will reuse ground-truth nodes from relation extraction datasets instead of generating our own nodes.

Among all the network extraction methods, Co-occurrence network is considered to be a reliable method to be widely used. [61], and [66] both used

Cooccurrence methods to extract network, which leverages the distance between terms to decide edge existence. Grayson et al. [40] developed novel variation rules of window size to capture more reliable information from co-occurred words. However, the reliability of co-occurrence network is still insufficiently investigated [29]. Park et al. [63] criticized the uncertainty of the selection of sliding window for identifying edges. The choice of appropriate window size is pretty random and may affect the constructed network. Grayson et al. [40] found for character text networks extracted from novels, the selection of sliding windows methods has considerable impacts on the constructed text networks.

Comparing to traditional distance based methods, NLP methods leverage richer information from text to construct networks. This information includes but not limit to Part-of-speech tagging, sub-phrases, or dependency tree [4]. NLP methods use both rule-based and machine learning methods to process linguistic features. Simple rule-based models such as Snowball method, infer the relations between two entities heuristically based on the phrases around entities. Supervised learning methods rely on annotated text to train models to extract relations. [100] and [97] trained SVM classifiers using dependency parse trees to classify different types of entity relations.

To date, there is a need for a systematic study to understand how the quality of extracted network may vary differently with regarding to the change of extraction methods. This motivates us to conduct this study to uncover the correlation between corpus types and network construction methods.

3.4.3 Analysis of Text Network

In this work, my focus will be solely on constructing and analyzing text networks in the social domain, which consist of sets of social actors, including individuals and organizations. Social actors connected by edges which represent interactions in social networks. The selection of edges vary according to scenarios, such as emailing [51], communication [54], collaborations [55]. A more formal definition of edges in a text network needs to consider directionality, temporality, and labels. Based on these properties, directed/undirected, temporal/static, or non-labelled/labelled networks can be further generated.

Among all the text networks, the most basic one-undirected, non-label, static network, only considers the existence of connections between entities. For this type of network, its edges represent the existence of the connection between entities without annotations of temporal information and other properties. In this study, I will focus on this type of simple network. It is reasonable to start thinking guidelines for creating reliable text networks from the most basic one instead of more complicated networks.

3.5 Research design and Method

To understand the reliability of relation extraction approaches, I ask the following research questions.

- RQ1: How do corpus genres influence the performance of relation extraction approaches? Corpus genres include newswire, Wikidata, and social media data.
- RQ2: How do relation types influence the performance of relation extraction approaches? Relation types include binary relation types, entity-type-based relations, and typed relations. Binary relation types only consider if two entities are related or not. entity-type-based relations consider combinations of entities as entity types, like ORG-ORG. Typed relations are these relation types defined in corpora, such as employment relation.

As prior works has studies how nodes construction affect networks [22, 26, 29], studying how relations affect network will mitigate the gaps in literature and improve the overall reliability of networks. Research question one focuses on evaluating the performance of each relation extraction method, namely, distance/syntax and learning-based relation extraction approaches. I choose to evaluate these methods because they are the most widely used approaches in text network construction which do not require a complex training process to develop models. For each research question, I will evaluate the ratio of relations extracted with respect to corpus genre, relation types, and parameter settings of the approaches. I choose these specific factors because they are the main decisions that developers need to make when constructing a text network from corpus. To make a fair comparison, I will also include several cutting-edge transformer-based approaches with pre-trained models [2].

RQ2 wants to answer in real scenarios when developers want to develop different text networks, how do these methods perform differently. I am interested in knowing under the social text network, which include social entities which occur most in the corpus being selected.

3.5.1 Data

After an extensive investigation of benchmark datasets for relation extraction, I select following datasets, table 3.1 exhibits a descriptive statistics of these datasets. Though we find these datasets all have relation annotations, only a few have negative annotations, such as TACRED [99]. Here, negative annotations refer to examples of non-related entities, such as 'a' and 'b' not being related. For datasets lacking annotated negative samples, I will assume that all entity combinations appearing in the same sentence without a relation label are negative. Table 3.1 shows the differences in annotations, and the additional task(s) in which each data set is used, if any. These RE datasets contain non-social relations (e.g. X was born on Y(date)), but I will limit this study to social relations present in these datasets. Here social relations refer to relations among Organizations (ORG), Person (PER), Geopolitical Entities (GPE), Nationalities, religions or political groups (NORP), and Location (LOC). On the other hand, we also select two domain-specific datasets Hurricane Matthew Twitter and Hurricane Matthew Sitreps datasets to test my findings in specific areas. These two datasets were collected from online documents and social media. The ground-truth annotations considering how ORG, LOC and PER interacted with each other. Though there are other relation extraction datasets, for example, BioNLP [57], I will exclude them as they focus on domain-specific relations that do not align with the social domain settings.

I also notice a significant amount of relations do not align with the definition of relations: relations between two entities. For example, from file 20000715_AFP_ARB.0054.eng in ACE2004 dataset, annotated relation "{head (entity 1)} of the {Ministry of Finance and Economic Development (entity 2)}" refers to a complete entity instead of a relation between "head" and "the

Table 3.1: Summary of the datasets used in the experiment settings.

Dataset	Genre	#Social Entities	#Social Relations (negative cases)	Annotations		
				Level	Relation Type	Entity Type
ACE2004	Newswire	8119	73189 (67336)	Document	Yes	Yes
ACE2005	Newswire	54798	113797 (105209)	Document	Yes	Yes
TACRED	Newswire Web Text	212528	101395 (80219)	Sentence	Yes	Yes
KBP37	Wikipedia	41800	19180 (1871)	Sentence	No	No
ORE	Newswire Transcription	1000	888 (597)	Sentence	Yes	Yes
Hurricane Matthew Twitter dataset	Social Media	1835	1306 (469)	Sentence	Yes	Yes
Hurricane Matthew Sitreps dataset	Government Document	587	1304 (304)	Sentence	Yes	Yes
DocRED	-	-	-	Document	-	-
re3d	-	-	-	Sentence	-	-
hlt-naacl08	-	-	-	Sentence	-	-
FewRel	-	-	-	Sentence	-	-

ministry". Similarly, "{Zimbabwe(entity 1)} 's {President(entity 2)}" should also be classified as an entity rather than a relation. Additionally, pronouns such as "he", "they" are also common in benchmark datasets. Though these pronouns implicitly link to meaningful entities, people can not always tell which entities they point by observing single sentence. For example, some pronouns refer to mentioned entities appearing in an early sentence which violate the experiment settings that I will only focus on same-sentence relation extraction. Thus, I will remove these "of", apostrophe, and pronoun relations from the benchmark datasets. Unlike traditional benchmarking studies that compare results with the original dataset, we have deliberately selected and limited our experiment to settings of text network construction that anticipate relationships between entities within the same sentence. Further work could remove these constraints and make the study more generalizable.

Further, I have designed a control experiment to evaluate different relation extraction methods on different network relations. Considering there are more sophisticated deep neural network (DNN) based methods (CNN, RNN), I will choose distance and syntax-based rule-based models for their explainability for the computational social science community, and also include bench-mark results using CNN, RNN, and transformer-based models.

3.6 Progress

I have completed the preliminary results of this study and presented this work at multiple conferences (COE, NASN). The experiment results of rule-based methods on situational reports and tweets have been published in ISCRAM. I have also completed the application of feature-based supervised approaches, including boosting and SVM models. The results require an indepth investigation to determine which language features affect performance, including precision and recall. I plan to finish a paper draft by the end of 2023 and submit it to Social Networks Journal.

CHAPTER 4

RELIABLE RELATION EXTRACTION FROM THINK-ALOUD IN EDUCATION SCIENCE

This project studies approaches to adopt information extraction techniques for the analysis of educational outcomes and materials. As educational materials become increasingly abundant and digitized, recorded during-learning data, including data created during learning processes, is more accessible recently. Meanwhile, large-scale online education, such as massive open online courses (MOOC) have also become accessible to hundreds of millions of students, efficiently evaluating these students' performance becomes a crucial task for education. Thus, I will focus on using NLP techniques integrated with semantic network analysis to analyze students' during-learning data and advance the development of reliable semi-automatic solutions for students' performance evaluations. I will describe the progress of integrating social network analysis into the analysis of education materials, the research questions that I intend to answer, and the plan to answer these research questions.

4.1 Introduction

Educational materials refer to materials including textbooks, and slides used in the procedures of teaching and learning. The analysis of educational materials using computing techniques has been widely adopted in education practice to evaluate and improve student performance. For example, ETS uses the e-rater engine to score students' writing ability on their standard English tests [1]. [91] leveraged Bert models to score student essays and achieved a performance comparable to that of human graders. With the rapid development of computing technologies, education evaluation has become more automatic than ever before.

In the context of education science, verbal data, including transcripts of classroom discourse, small-group dialogues, talk-aloud protocols from reasoning and problem-solving tasks are increasingly used by researchers in their studies. Analyzing these verbal data is not a simple task such as counting the number of key concepts, but a complicated process that includes capturing linguistic features [67], finding structures from text [52], and applying statistical analysis [53]. Semantic Network Analysis emerges to be an approach that systematically integrates various verbal analyses. Verbal text is first converted into nodes representing key concepts defined by researchers, and edges representing connections between nodes. Furthermore, researchers can apply network metrics, such as betweenness centrality [89] and PageRank centrality [10] or network algorithms such as community detections [78], to understand patterns in verbal text, and correlating them with standard performance.

Conducting semantic network analysis requires efforts to construct faithful networks that represent verbal text data. Classic approaches often use manual coding or simple heuristics [44] to extract nodes and edges, which is not applicable to large volumes of text data or information-rich corpus. For example, two-page textbook may generate more than hundreds of nodes and edges. Also, as heuristics do not always capture the subtle change and diverse expression of human language, inaccurate networks can be anticipated. Thus, reliable and semi-automatic approaches to construct semantic networks from educational verbal text are highly valuable to education science researchers to test hypotheses and develop studies.

In this project, I plan to use during-learning verbal data from students who had completed a college-level biology class as an example to test and improve methods to construct text networks, which include key concepts and their relations from class. I will develop a domain-specific annotated corpus with ground-truth semantic networks, apply and test models from Chapter 3 to construct semantic networks, and develop an optimized approach to construct semantic networks for our intended corpus with the collaborations with domain experts.

4.2 Research Questions

I plan to answer the following research questions in the context of education.

1. What are the expected semantic networks of students' during-study

think-aloud data?

2. How can we achieve accurate semantic networks by applying general relation extraction models tested on public ground-truth datasets?

3. How can we achieve accurate semantic networks by integrating domain expertise in the development process?

4. How do semantic relations extracted from students' verbal data correlate with students' learning outcomes and knowledge structure?

To answer RQ1, I will involve experts in the education science field to develop such ground-truth semantics. I will use an iterative bottom-up method to develop a systematic guideline for generating such ground-truth datasets (see more detail in the Methodology section). For RQ2, I will compare the students' verbal data networks with the textbook data networks in terms of network and educational metrics. This comparison will test the accuracy of using verbal data semantic networks to evaluate student learning performance. To address RQ3, rule-based and statistical models will be adopted to extract relations from text data. I will apply error analysis to understand what challenges and errors we may encounter if we directly apply models optimized with public ground-truth datasets to education materials. Finally, I will optimize relation extraction models by integrating domain experts' knowledge in a human-in-the-loop fashion.

4.3 Methodology

This project includes two major sub-tasks: 1. develop a domain-specific corpus for semantic networks from learning materials and students' verbal data. 2. Develop and test models to extract relations from education text data.

4.3.1 Data

We will use the textbook from an introductory biology course as the source of a textbook semantic networks [68]. Additionally, we will reuse verbal data collected from a in-lab think-aloud study conducted by one of the collaborators from educational psychology department. The think-aloud study asked students to learn provided textbook passages in 40 minutes and then thought aloud while learning and were given paper and pen to take notes. After the study, these students moved to another room to type in everything they remembered from the text. All the verbal and typed text were collected as the source of individual students' semantic networks. Figure 4.1 shows what a participant read and spoke during the study. All the italicized text refers to texbook content, and all the regular text refers to what students spoke during the study, like "So, it's implying that white blood cells are larger than red blood cells.".

Participant Number:

See Selected sections from Chapter 42 Immunology_Blackboard.doc and the corresponding JPEGS in Box for support in recording the text.

Passage 3

White blood cells play many defensive roles
One milliliter of human blood typically contains about 5 billion red blood cells .
I'm writing it down, *red blood cells and 7 million (OMITTED: of the) larger white blood cells.*
So, it's implying that white blood cells are larger than red blood cells.
All of these cells originate from multipotent stem cells (constantly dividing undifferentiated cells that can form several different cell types).
Yes, I am aware of that.
in the bone marrow.
Examine Figure 42.2 and you will see that there are two major families of white blood cells (also called leukocytes): phagocytes and lymphocytes.
Lymphocytes, which include B cells and T cells, are smaller than phagocytes and are not phagocytic.
Each family contains different types of cells with specialized functions.
Natural killer cells and some kinds of phagocytes are also referred to collectively as granulocytes because they contain numerous granules .
So let's go back again.

Figure 4.1: In-lab Think-aloud corpus.

4.3.2 Develop Ground-Truth Annotations

To create a ground-truth dataset of verbal data based semantic network, we will use a bottom-up method to code semantic triples, namely *subject-verb-object*, from text data. We will first conduct a open-coding approach to find all possible triples from text data. Then, we will work with an education science researcher to verify these extracted triples. Based on extracted triples, we will summarize several guidelines and standards to generate semantic triples, such as how to handle content in parentheses. Through iterative test and verifications of these guidelines, we will finalize our codebook when achieving a satisfying inter-coder agreement.

With the verified coding guidelines, we will recruit undergraduates to help with coding more materials to develop a sizable ground-truth corpus. We will first train these coders with the coding guidelines and verify their coding accuracy rate, ensuring raters have a high agreements. We plan to deploy a Label Studio service [84] on a lab server for multiple users to code separately on the same corpus. Coding results will be consolidated to form ground-truth semantic network corpus.

4.3.3 Test existing models

Chapter 3 will give us insights into model selections for different corpora. We will apply selected best models to the domain datasets that we collected for this project. We hypothesize that these models will perform poorly on our ground-truth datasets as the corpus we tested in Chapter 3 were from Wikipedia, newswire, and web. We will treat these results as a baseline to integrate human experts in the development of a more robust relation extraction approach.

4.3.4 Improve Relation Extraction using Human-in-the-loop fashions

Analyzing baseline models will enable us to identify errors and flaws of existing models in constructing semantic networks. Education science researchers will be involved in grouping and analyzing these errors. Further, we will adopt pre-trained large language models [88, 36] and fine-tune these models using the data annotations we collected for this project. Optimized models will be released to researchers to analyze verbal data to advance the research of education psychology.

December 15, 2023

4.4 Progress

I am at the stage of developing an annotated ground-truth corpus. I have developed and iterated a codebook and have enlisted the help of two other students to develop an annotated corpus. We plan to finish the annotations by December and complete a draft paper targeting AIED 2024, with a due date of January 29.

CHAPTER 5

TIMELINE AND CONTRIBUTIONS

5.1 Timeline

Based on this proposal, I have an expected timeline for completing three projects. Figure 5.1 shows the corresponding subtasks and the expected completion time. The timeline takes into account the dependencies of subtasks and depicts the development and publication plan.

	Before Prelim			Past Prelim								
	September, 2023	October, 2023	November, 2023	December, 2023	January, 2024	February, 2024	March, 2024	April, 2024	May, 2024	June, 2024	July, 2024	August, 2024
Social Media												
- Survey Recent LLM models												
- Fine-tune LLM models using Twitter Data												
- Optimize who-need-what models												
- Submit Revised Paper												
Textual Network Analysis												
- Collect and clean ground-truth datasets												
- Run classic relation extraction models on gt datasets												
- Analyze Errors and Optimize Model Settings												
- Submit Paper to Social Networks Journal												
Education Data Semantic Analysis												
- Develop Codebook and Recruit annotators												
- Finish Coding Education Materials												
- Applying Existing Models												
- Optimize Relation Extraction Models												
- Write and Submit												
Dissertation												

Figure 5.1: Timeline of research

5.2 Overall Expected Contributions

This proposal presents a comprehensive plan to advance information extraction from a human-in-the-loop perspective, in which we actively involve human experts in the process. The goal of this proposal is to depict a three-stage development of information extraction technologies, from empirical use cases to methodological contributions. I will start with an urgent task, identifying needs from social media data during Ukraine-Russian conflicts. I hope to contribute to the research community by demonstrating how reliable relation extraction can help the public perceive the needs expressed by social media users during crises. Then, I will systematically evaluate how different factors, ranging from feature selections to relation types, affect the performance of information extraction on public benchmark datasets. I hope to contribute to the selection of methodologies for conducting relation extractions and provide suggestions for practitioners to reliably extract relations. Finally, I will test relation extraction approaches for semantic network construction in the context of education science and involve domain experts in the loop to create insights for practitioners to construct more reliable semantic networks for education science studies.

5.3 Expected Contributions

This proposal depicts a comprehensive plan to advance information extraction from a human-in-the-loop perspective. I expect to contribute to the research of relation extraction by making contributions from the following aspects:

5.3.1 Methodological Contributions

- Contribute to a systematic review of traditional relation extraction approaches. (Chapter 3)
- Contribute to understanding influential factors, such as language features, for developing and evaluating relation extraction approaches. (Chapter 3)

5.3.2 Empirical Contributions

- Contribute to the understanding of the needs expressed by users on social media platforms. (Chapter 2)
- Contribute to developing a disaster relief guideline using social media data to satisfy front-line needs better. (Chapter 2)

- Contribute to developing relation extraction approaches for constructing semantic networks for education science. (Chapter 4)

5.3.3 Dataset Contributions

- Contribute to developing a ground-truth annotation corpus for relation extraction in the area of education science. (Chapter 4)
- Contribute to developing a dataset of user needs on social media. (Chapter 2)

ACKNOWLEDGMENTS

This material is based upon work partially supported by the National Science Foundation under Grant No. 2225298. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

[1] e-rater® scoring engine. https://www.ets.org/erater/about.html: :text=The Accessed: 2023-08-20.

[2] Christoph Alt, Marc Hübner, and Leonhard Hennig. Fine-tuning pre-trained transformer language models to distantly supervised relation extraction. In Anna Korhonen, David Traum, and Lluís Màrquez, editors, *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1388–1398, Florence, Italy, July 2019. Association for Computational Linguistics.

[3] Zahra Ashktorab, Christopher Brown, Manojit Nandi, and Aron Culotta. Tweedr: Mining twitter to inform disaster response. In *ISCRAM*, pages 269–272, 2014.

[4] Nguyen Bach and Sameer Badaskar. A review of relation extraction. *Literature review for Language and Statistics II*, 2:1–15, 2007.

[5] George A Barnett. *Encyclopedia of social networks*. Sage Publications, 2011.

[6] Elisa Bassignana and Barbara Plank. What do you mean by relation extraction? a survey on datasets and study on scientific relation classification. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, pages 67–83, Dublin, Ireland, May 2022. Association for Computational Linguistics.

[7] Moumita Basu, Anurag Shandilya, Kripabandhu Ghosh, and Saptarshi Ghosh. Automatic matching of resource needs and availabilities in microblogs for post-disaster relief. In *Companion Proceedings of the The Web Conference 2018*, pages 25–26, 2018.

[8] Moumita Basu, Anurag Shandilya, Prannay Khosla, Kripabandhu Ghosh, and Saptarshi Ghosh. Extracting resource needs and availabilities from microblogs for aiding post-disaster relief operations. *IEEE Transactions on Computational Social Systems*, 6(3):604–618, 2019.

[9] Mohamed Behnassi and Mahjoub El Haiba. Implications of the russia–ukraine war for global food security. *Nature Human Behaviour*, pages 1–2, 2022.

[10] Madelen Bodin. Mapping university students' epistemic framing of computational physics using network analysis. *Physical Review Special Topics-Physics Education Research*, 8(1):010115, 2012.

[11] Davide Buscaldi, Anne-Kathrin Schumann, Behrang Qasemizadeh, Haïfa Zargayouna, and Thierry Charnois. Semeval-2018 task 7: Semantic relation extraction and classification in scientific papers. In *International Workshop on Semantic Evaluation (SemEval-2018)*, pages 679–688, 2017.

[12] Ramon Ferrer I Cancho and Richard V Solé. The small world of human language. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 268(1482):2261–2265, 2001.

[13] Kathleen M Carley. Network text analysis: The network position of concepts. In *Text analysis for the social sciences*, pages 79–100. Routledge, 2020.

[14] Kathleen M Carley, Jana Diesner, Jeffrey Reminga, and Maksim Tsvetovat. Toward an interoperable dynamic network analysis toolkit. *Decision Support Systems*, 43(4):1324–1347, 2007.

[15] Christopher A Cassa, Rumi Chunara, Kenneth Mandl, and John S Brownstein. Twitter as a sentinel in emergency situations: lessons from the boston marathon explosions. *PLoS currents*, 5, 2013.

[16] Guido Cervone, Elena Sava, Qunying Huang, Emily Schnebele, Jeff Harrison, and Nigel Waters. Using twitter for tasking remote-sensing data collection and damage assessment: 2013 boulder flood case study. *International Journal of Remote Sensing*, 37(1):100–124, 2016.

[17] Emily Chen and Emilio Ferrara. Tweets in time of conflict: A public dataset tracking the twitter discourse on the war between ukraine and russia. *Proceedings of the International AAAI Conference on Web and Social Media*, 17(1):1006–1013, Jun. 2023.

[18] Seong Eun Cho, Kyujin Jung, and Han Woo Park. Social media use during japan's 2011 earthquake: how twitter transforms the locus of crisis communication. *Media International Australia*, 149(1):28–40, 2013.

[19] Camille Cobb, Ted McCarthy, Annuska Perkins, Ankitha Bharadwaj, Jared Comis, Brian Do, and Kate Starbird. Designing for the deluge: understanding & supporting the distributed, collaborative work of crisis

volunteers. In *Proceedings of the 17th Conference on Computer Supported Cooperative Work & Social Computing (CSCW)*, pages 888–899. ACM, 2014.

[20] Steven R Corman, Timothy Kuhn, Robert D McPhee, and Kevin J Dooley. Studying complex discursive systems. centering resonance analysis of communication. *Human communication research*, 28(2):157–206, 2002.

[21] Laura Dabbish, Ben Towne, Jana Diesner, and James Herbsleb. Construction of association networks from communication in teams working on complex projects. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 4(5):547–563, 2011.

[22] Jana Diesner. Uncovering and managing the impact of methodological choices for the computational construction of socio-technical networks from texts. Technical report, CARNEGIE-MELLON UNIV PITTSBURGH PA INST FOR SOFTWARE RESEARCH, 2012.

[23] Jana Diesner. Small decisions with big impact on data analytics. *Big Data & Society*, 2(2):1–6, 2015.

[24] Jana Diesner. Words and networks: How reliable are network data constructed from text data? In *Roles, Trust, and Reputation in Social Media Knowledge Markets*, pages 81–89. Springer, 2015.

[25] Jana Diesner and Kathleen M Carley. Using network text analysis to detect the organizational structure of covert networks. In *Proceedings of the North American Association for Computational Social and Organizational Science (NAACSOS) Conference*, volume 3. NAACSOS Pittsburgh, 2004.

[26] Jana Diesner and Kathleen M Carley. Conditional random fields for entity extraction and ontological text coding. *Computational and Mathematical Organization Theory*, 14(3):248–262, 2008.

[27] Jana Diesner and Kathleen M Carley. He says, she says. pat says, tricia says. how much reference resolution matters for entity extraction, relation extraction, and social network analysis. In *2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications*, pages 1–8. IEEE, 2009.

[28] Jana Diesner and Kathleen M Carley. Semantic networks. *Encyclopedia of social networking*, pages 766–769, 2011.

[29] Jana Diesner and Kathleen M Carley. Impact of relation extraction methods from text data on network data and analysis results. In *Short*

paper at ACM Web Science Conference, Words and Networks Workshop (WON 2012), Evanston, IL, 2012.

[30] Jana Diesner, Kathleen M Carley, and Laurent Tambayong. Extracting socio-cultural networks of the sudan from open-source, large-scale text data. *Computational and Mathematical Organization Theory*, 18(3):328–339, 2012.

[31] Jana Diesner, Chieh-Li Chin, and Marc A. Smith. *Combining Online Social Networks with Text Analysis*, pages 312–321. Springer New York, New York, NY, 2018.

[32] Jana Diesner, Jinseok Kim, and Susie Pak. Computational impact assessment of social justice documentaries. *Journal of Electronic Publishing*, 17(3), 2014.

[33] Jana Diesner, Ponnurangam Kumaraguru, and Kathleen M Carley. Mental models of data privacy and security extracted from interviews with indians. In *55th Annual Conference of the International Communication Association (ICA), New York, NY*, 2005.

[34] Jana Diesner and Rezvaneh Rezapour. Social computing for impact assessment of social change projects. In *Social Computing, Behavioral-Cultural Modeling, and Prediction: 8th International Conference, SBP 2015, Washington, DC, USA, March 31-April 3, 2015. Proceedings 8*, pages 34–43. Springer, 2015.

[35] Ly Dinh, Sumeet Kulkarni, Ping-Jing Yang, and Jana Diesner. Reliability of methods for extracting collaboration networks from crisis-related situational reports and tweets. In *Proceedings of Information Systems for Crisis Response and Management (ISCRAM2022)*, 2022.

[36] Alexander Dunn, John Dagdelen, Nicholas Walker, Sanghoon Lee, Andrew S Rosen, Gerbrand Ceder, Kristin Persson, and Anubhav Jain. Structured information extraction from complex scientific text with fine-tuned large language models. *arXiv preprint arXiv:2212.05238*, 2022.

[37] Russell Rowe Dynes. *Organized behavior in disaster*. Heath Lexington Books, 1970.

[38] FEMA. Community lifelines, 2019.

[39] Colin Garvey. A framework for evaluating barriers to the democratization of artificial intelligence. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1), Apr. 2018.

[40] Siobhán Grayson, Karen Wade, Gerardine Meaney, and Derek Greene. The sense and sensibility of different sliding windows in constructing co-occurrence networks from literature. In *International Workshop on Computational History and Data-Driven Humanities*, pages 65–77. Springer, 2016.

[41] Ben Hachey, Claire Grover, and Richard Tobin. Datasets for generic relation extraction. *Natural Language Engineering*, 18(1):21–59, 2012.

[42] Ridong Han, Tao Peng, Chaohao Yang, Benyou Wang, Lu Liu, and Xiang Wan. Is information extraction solved by chatgpt? an analysis of performance, evaluation criteria, robustness and errors. *arXiv preprint arXiv:2305.14450*, 2023.

[43] Hans W. A. Hanley, Deepak Kumar, and Zakir Durumeric. quot;a special operationquot;: A quantitative approach to dissecting and comparing different media ecosystems' coverage of the russo-ukrainian war. *Proceedings of the International AAAI Conference on Web and Social Media*, 17(1):339–350, Jun. 2023.

[44] Trazíbulo Henrique, Inácio de Sousa Fadigas, Marcos Grilo Rosa, and Hernane Borges de Barros Pereira. Mathematics education semantic networks. *Social Network Analysis and Mining*, 4:1–9, 2014.

[45] Starr Roxanne Hiltz and Linda Plotnick. Dealing with information overload when using social media for emergency management: Emerging solutions. In *ISCRAM*, 2013.

[46] Matthew Honnibal and Ines Montani. spacy 2: Natural language understanding with bloom embeddings, convolutional neural networks and incremental parsing. *To appear*, 7, 2017.

[47] Liaquat Hossain, D Kam, F Kong, RT Wigand, and Terence Bossomaier. Social media in ebola outbreak. *Epidemiology & Infection*, 144(10):2136–2143, 2016.

[48] Muhammad Imran, Prasenjit Mitra, and Carlos Castillo. Twitter as a lifeline: Human-annotated twitter corpora for nlp of crisis-related messages. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, pages 1638–1643, Paris, France, may 2016. European Language Resources Association (ELRA).

[49] Jacob Devlin Ming-Wei Chang Kenton and Lee Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of naacL-HLT*, volume 1, page 2, 2019.

[50] Katsushige Kitazawa and Scott A Hale. Social media and early warning systems for natural disasters: A case study of typhoon etau in japan. *International Journal of Disaster Risk Reduction*, 52:101926, 2021.

[51] Bryan Klimt and Yiming Yang. Introducing the enron corpus. In *CEAS*, 2004.

[52] Jay L Lemke. Intertextuality and text semantics. *Advances in Discourse Processes*, 50:85–114, 1995.

[53] Jay L Lemke. Analyzing verbal data: Principles, methods, and problems. *Second international handbook of science education*, pages 1471–1484, 2012.

[54] Jure Leskovec, Daniel Huttenlocher, and Jon Kleinberg. Signed networks in social media. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1361–1370. ACM, 2010.

[55] Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. Graph evolution: Densification and shrinking diameters. *ACM transactions on Knowledge Discovery from Data (TKDD)*, 1(1):2–es, 2007.

[56] Andrea Maggioni, Jose A Gonzales-Zamora, Alessandra Maggioni, Lori Peek, Samantha A McLaughlin, Ulrich von Both, Marieke Emonts, Zelde Espinel, and James M Shultz. Cascading risks for preventable infectious diseases in children and adolescents during the 2022 invasion of ukraine. *International Journal of Environmental Research and Public Health*, 19(12):7005, 2022.

[57] Claire Nédellec, Robert Bossy, Jin-Dong Kim, Jung-Jae Kim, Tomoko Ohta, Sampo Pyysalo, and Pierre Zweigenbaum. Overview of bionlp shared task 2013. In *Proceedings of the BioNLP shared task 2013 workshop*, pages 1–7, 2013.

[58] OCHA. Humanitarian needs overview: Ukraine, 2022.

[59] Alexandra Olteanu, Sarah Vieweg, and Carlos Castillo. What to expect when the unexpected happens: Social media communications across crises. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*, pages 994–1009. ACM, 2015.

[60] Ebru Orhan. The effects of the russia-ukraine war on global trade. *Journal of International Trade, Logistics and Law*, 8(1):141–146, 2022.

[61] Arzucan Özgür and Haluk Bingol. Social network of co-occurrence in news articles. In *International Symposium on Computer and Information Sciences*, pages 688–695. Springer, 2004.

[62] Leysia Palen, Sarah Vieweg, Sophia B Liu, and Amanda Lee Hughes. Crisis in a networked world: Features of computer-mediated communication in the april 16, 2007, virginia tech event. *Social Science Computer Review*, 27(4):467–480, 2009.

[63] Chisung Park, Junseok Lee, and Doosan Paik. *Identifying Policy Frames Using Semantic Network Analysis*. SAGE Publications Ltd, 2019.

[64] Miruna Petrescu-Prahova and Carter T Butts. Emergent coordination in the world trade center disaster. *Institute for mathematical behavioral sciences*, 26(3):1–23, 2005.

[65] Hemant Purohit, Carlos Castillo, Fernando Diaz, Amit Sheth, and Patrick Meier. Emergency-relief coordination on social media: Automatically matching resource requests and offers. *First Monday*, 19(1), 2014.

[66] Srinivasan Radhakrishnan, Serkan Erbis, Jacqueline A Isaacs, and Sagar Kamarthi. Novel keyword co-occurrence network-based methods to foster systematic reviews of scientific literature. *PloS one*, 12(3):e0172778, 2017.

[67] Rebecca Rogers. An introduction to critical discourse analysis in education. In *An introduction to critical discourse analysis in education*, pages 31–48. Routledge, 2004.

[68] David E Sadava, David M Hillis, and H Craig Heller. *Life: the science of biology*, volume 2. Macmillan, 2009.

[69] Anita Saroj and Sukomal Pal. Use of social media in crisis management: A survey. *International Journal of Disaster Risk Reduction*, 48:101584, 2020.

[70] M Janina Sarol, Ly Dinh, and Jana Diesner. Variation in situational awareness information due to selection of data source, summarization method, and method implementation. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 15, pages 597–608, 2021.

[71] M Janina Sarol, Ly Dinh, and Jana Diesner. Variation in situational awareness information due to selection of data source, summarization method, and method implementation. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 15, pages 597–608, 2021.

[72] M. Janina Sarol, Ly Dinh, Rezvaneh Rezapour, Chieh-Li Chin, Pingjing Yang, and Jana Diesner. An empirical methodology for detecting and prioritizing needs during crisis events. In *Findings of the Association for Computational Linguistics: EMNLP*, pages 4102–4107, 2020. 10.18653/v1/2020.findings-emnlp.366.

[73] Mohsen Sayyadiharikandeh, Onur Varol, Kai-Cheng Yang, Alessandro Flammini, and Filippo Menczer. Detection of novel social bots by ensembles of specialized classifiers. In *Proceedings of the 29th ACM international conference on information & knowledge management*, pages 2725–2732, 2020.

[74] Elad Segev. Textual network analysis: Detecting prevailing themes and biases in international news and social media. *Sociology Compass*, 14(4):e12779, 2020.

[75] Tatiana Serafin. Ukraine's president zelensky takes the russia/ukraine war viral. *Orbis*, 66(4):460–476, 2022.

[76] Jingbo Shang, Jialu Liu, Meng Jiang, Xiang Ren, Clare R Voss, and Jiawei Han. Autophrase: Automated phrase mining from massive text corpora. Technical report, Retrieved 2019-08-22, from https://github.com/shangjingbo1226/AutoPhrase ..., 2019.

[77] Frances Shaw, Jean Burgess, Kate Crawford, and Axel Bruns. Sharing news, making sense, saying thanks: Patterns of talk on twitter during the queensland floods. *Australian Journal of Communication*, 40(1):23–39, 2013.

[78] Cynthia SQ Siew, Dirk U Wulff, Nicole M Beckage, and Yoed N Kenett. Cognitive network science: A review of research on cognition through the lens of network representations, processes, and dynamics. *Complexity*, 2019, 2019.

[79] Tomer Simon, Avishay Goldberg, and Bruria Adini. Socializing in emergencies—a review of the use of social media in emergency situations. *International journal of information management*, 35(5):609–619, 2015.

[80] Bridget Smart, Joshua Watt, Sara Benedetti, Lewis Mitchell, and Matthew Roughan. # istandwithputin versus# istandwithukraine: The interaction of bots and humans in discussion of the russia/ukraine war. In *International Conference on Social Informatics*, pages 34–53. Springer, 2022.

[81] John F Sowa. *Conceptual structures: information processing in mind and machine*. Addison-Wesley Longman Publishing Co., Inc., 1984.

[82] Felix Olajide Talabi, Ayodeji Boluwatife Aiyesimoju, Ishola Kamorudeen Lamidi, Samson Adedapo Bello, Joshua Kayode Okunade, Chinedu Joel Ugwuoke, and Verlumun Celestine Gever. The use of social media storytelling for help-seeking and help-receiving among nigerian refugees of the ukraine–russia war. *Telematics and Informatics*, 71:101836, 2022.

[83] Lynda Tamine, Laure Soulier, Lamjed Ben Jabeur, Frederic Amblard, Chihab Hanachi, Gilles Hubert, and Camille Roth. Social media-based collaborative information access: Analysis of online crisis-related twitter conversations. In *Proceedings of the 27th ACM conference on hypertext and social media*, pages 159–168, 2016.

[84] Maxim Tkachenko, Mikhail Malyuk, Andrey Holmanyuk, and Nikolai Liubimov. Label Studio: Data labeling software, 2020-2022. Open source software available from https://github.com/heartexlabs/label-studio.

[85] Wil MP van der Aalst, Martin Bichler, and Armin Heinzl. Responsible data science, 2017.

[86] Tracy Van Holt, Jeffrey C Johnson, Kathleen M Carley, James Brinkley, and Jana Diesner. Rapid ethnographic assessment for cultural mapping. *Poetics*, 41(4):366–383, 2013.

[87] István Varga, Motoki Sano, Kentaro Torisawa, Chikara Hashimoto, Kiyonori Ohtake, Takao Kawai, Jong-Hoon Oh, and Stijn De Saeger. Aid is out there: Looking for help from tweets during a large scale disaster. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, pages 1619–1629, Sofia, Bulgaria, August 2013. Association for Computational Linguistics.

[88] Somin Wadhwa, Silvio Amir, and Byron C. Wallace. Revisiting relation extraction in the era of large language models. *Proceedings of the conference. Association for Computational Linguistics. Meeting*, 2023:15566–15589, 2023.

[89] S Wagner and B Priemer. Assessing the quality of scientific explanations with networks. *International Journal of Science Education*, pages 1–25, 2023.

[90] Hailin Wang, Ke Qin, Rufai Yusuf Zakari, Guoming Lu, and Jin Yin. Deep neural network-based relation extraction: an overview. *Neural Computing and Applications*, pages 1–21, 2022.

[91] Yongjie Wang, Chuang Wang, Ruobing Li, and Hui Lin. On the use of bert for automated essay scoring: Joint learning of multi-scale essay

representation. In Marine Carpuat, Marie-Catherine de Marneffe, and Ivan Vladimir Meza Ruiz, editors, *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 3416–3425, Seattle, United States, July 2022. Association for Computational Linguistics.

[92] Shawn A Weil, Pacey Foster, Jared Freeman, Kathleen Carley, Jana Diesner, Terrill Franz, Nancy J Cooke, Steve Shope, and Jamie C Gorman. Converging approaches to automated communications-based assessment of team situation awareness. In *Macrocognition in Teams*, pages 276–304. CRC Press, 2017.

[93] Kai-Cheng Yang, Emilio Ferrara, and Filippo Menczer. Botometer 101: Social bot practicum for computational social scientists. *CoRR*, abs/2201.01608, 2022.

[94] Ping-Jing Yang, Janina Sarol, Ly Dinh, and Jana Diesner. Annotation guidelines for entity tagging and semantic role labeling of disaster-related text documents, 2020. Presented at the Critical Infrastructure Resilience Institute 11th Maritime Risk Symposium AND at US Department of Homeland Security Centers of Excellence (COE) Summit, Fairfax, Virginia (held online).

[95] Ping-Jing Yang, Janina Sarol, Ly Dinh, and Jana Diesner. Reliable relation extraction for social network construction. Presentation at the North American Regional Social Networks Conference 2021 (NASN2021), 2021. Held online.

[96] Dave Yates and Scott Paquette. Emergency knowledge management and social media technologies: A case study of the 2010 haitian earthquake. *International journal of information management*, 31(1):6–13, 2011.

[97] Min Zhang, Jie Zhang, Jian Su, and Guodong Zhou. A composite kernel to extract relations between entities with both flat and structured features. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, pages 825–832, 2006.

[98] Xinyang Zhang, Yury Malkov, Omar Florez, Serim Park, Brian McWilliams, Jiawei Han, and Ahmed El-Kishky. Twhin-bert: A socially-enriched pre-trained language model for multilingual tweet representations at twitter. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 5597–5607, 2023.

[99] Yuhao Zhang, Victor Zhong, Danqi Chen, Gabor Angeli, and Christopher D. Manning. Position-aware attention and supervised data improve slot filling. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP 2017)*, pages 35–45, 2017.

[100] Shubin Zhao and Ralph Grishman. Extracting relations with integrated information using kernel methods. In *Proceedings of the 43rd annual meeting of the association for computational linguistics (acl'05)*, pages 419–426, 2005.