Applying Social Network Extraction With Named Entity Recognition to the Examination of Political Bias Within Online News Articles

Chih Ming Tsai
102051@mail.hwu.edu.tw
Hsing Wu University
Taipei, Taiwan

Kuan Heng Lin
jordanlin@xischina.com.cn
Xiamen International School
Xiamen, Fujian, China

ABSTRACT

We aim to expand the application of social network extraction with NER tools, which to date is largely limited to fiction. With the premise that news articles resemble mini-stories, this study explores the extraction of social networks from online United States news articles to examine relationships between political bias and network features. We find statistical significance with most trends, and find no substantial differences between Liberal and Conservative bias, but bias and neutrality. Furthermore, this study identifies several issues with social network analysis, proposing a more rigorous examination of textual characteristics that affect network features.

CCS CONCEPTS

• **Computing methodologies → Information extraction**.

KEYWORDS

social network extraction, network features, named entity recognition, political bias, online news articles

ACM Reference Format:
Chih Ming Tsai and Kuan Heng Lin. 2020. Applying Social Network Extraction With Named Entity Recognition to the Examination of Political Bias Within Online News Articles. In *2020 3rd International Conference on Artificial Intelligence and Pattern Recognition (AIPR 2020), June 26–28, 2020, Xiamen, China.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3430199.3430219

1 INTRODUCTION

Political objectivity in news media is becoming increasingly scarce, but online news consumers' awareness of media bias is growing faster than ever. This created a demand for quantitative measures to determine political bias of news outlets [1], giving birth to sites like *Media Bias Fact Check*[1], *Ad Fontes Media*[2], and *AllSides* [2]. Although bias rating methodologies do include statistical analysis, the primary process requires human-based, laborious analyses from experts of "different political affiliations" [3] to examine articles in the vast, rapidly-changing online political landscape [4].

[1]https://mediabiasfactcheck.com
[2]https://www.adfontesmedia.com

AIPR 2020, June 26–28, 2020, Xiamen, China

ACM ISBN 978-1-4503-7551-1/20/06…$15.00
https://doi.org/10.1145/3430199.3430219

Named Entity Recognition (NER) is a major branch of information extraction (IE), which aims to classify and tokenize texts for natural language processing (NLP) and related fields. First proposed by [5], NER operates by annotating named entities with tags, ranging from manual to computational methods. The rudimentary expression types include names, times, and numbers [6]. To date, NER finds wide applications, such as machine translation [7], biomedical text IE [8], and social network extraction [9]. This study focuses on the last application and its aid in examining political biases of United States online news media.

Therefore, we define the following research questions:

- How does political bias in United States news articles manifest in their social network features?
- To what extent is social network extraction from pre-existing NER tools suitable for analyzing the political bias of online mainstream media articles?

The rest of the study is structured as follows. Section 2 summarizes relevant works. Section 3 outlines our method for extracting social networks with NER tools. Section 4 presents the statistical results and analysis. Section 5 draws our conclusion and evaluation.

2 RELATED WORKS

The fundamental structure of social networks relies on entities, which NER tools identify. Although stricter rule-based NER systems relying on English grammar [10] favor the formal register of news articles, state-of-the-art systems based on convolutional neural networks can be beneficial for entity-heavy political journalism. More recently, neural network-based NER tools with pretrained language models have been increasingly off-the-shelf, such as `AllenNLP`, `NLTK`, `OpenNLP`, `Stanford NER`, and `spaCy` [11]. This study employs `spaCy` for its wide selection of entity types, sufficient claimed accuracy (92.6%), and compatibility with other libraries used in this study.

Anaphora resolution is crucial in network extraction to link pronouns and names. Although methods like linear word distance with gender tagging is sufficient for fiction [9], the subject-rich style of political articles lends this task confusing even to humans [12]. Thus, neural network-based methods are more suitable for our corpus.

To ensure that all named entities are consistently identified throughout, [13] proposed a Conditional Random Field-based NER system that simulates a non-local IE system which allows ambiguous tokens to be consistently labeled. As social network extraction values entity label consistency over the accuracy, the implementation of this method in modern off-the-shelf NER tools is beneficial.

Name resolution is similar to anaphora resolution but is instead concerned with linking names. All occurrences of first names must be linked with their respective full names to maintain accuracy of the social network. Unlike fiction, news articles enable context-based name resolution. For instance, [14] proposed using *Wikipedia*'s vast database to resolve differing entity names, either with category labels or treating *Wikipedia* as a large dictionary of names.

Social network extraction with NER tools sees two major approaches [9]. One is conversational networks, which, although effective in dialogue-heavy fiction, is impractical in non-fiction. Conversely, co-occurrence networks are more suitable for news articles with the latter's similarity to text types shown to be effective, like biomedical texts [15] (frequent cross-referencing and concise register).

The examination of how textual characteristics affect network features has been explored by [9], where classical and modern novels are compared. The work found no statistical significance between the two groups of novels, but it proposed an area of application for network extraction on which this study expands on.

3 METHODOLOGY

Since this study focuses on online news articles, the categorical independent variable [9] explores – classic versus modern novels – is limiting. However, the political nature of news articles presents another obvious and numerical variable: political bias.

3.1 Bias Score Calculation

To quantify political bias, we selected three bias rating databases – *Media Bias Fact Check* (*MBFC*), *Ad Fontes Media* (*AFM*), and *AllSides* (*AS*) – chosen for their wide range of rated sites to ensure sufficient diversity and size for the corpus. The provided ratings, after normalization between -1 (Liberal bias) and 1 (Conservative bias), are weighted and averaged to produce the final bias score.

Table 1: News rating databases and weighting

	Media Bias Fact Check	*Ad Fontes Media*	*AllSides*
Bias rating range	$[0, 600]$[3]	$[-42, 42]$	$[1, 5]$[4]
Weighting	1.0	1.5	0.5

To compensate for the differing practical ranges of these databases – roughly, *MBFC* is $[125, 475]$, *AFM* is $[-24, 24]$, and *AS* is $[1, 5]$ – we used the above (arbitrary) weightings. Additionally, we weighted *AFM* heaviest for its detailed white paper [3].

3.2 Corpus Selection

To cater to the characteristics of social networks, we imposed a rough 'criteria' when selecting the topics and news articles:

- *The topics must be politically significant*, ensuring a diverse and sufficiently-sized corpus.
- *The topics must have a short time-span (e.g. announcements or decisions)*, eliminating investigations (e.g. Muller Report), large elections (e.g. midterms), protests, or speeches (e.g. State of the Union), as articles of those topics lack a strong focus. This ensures sufficient similarity across articles for network feature differences to reflect biased reporting than content disparities – a small test on random-topic articles revealed no statistical significance.
- *The topics are recent*, as bias ratings are time-sensitive.
- *The articles must be immediate initial reports with the overarching purpose of informing*, eliminating opinion columns, analyses, and follow-up articles, which provides obvious article choices and minimizes content & bias disparities – *AS* even distinguishes their ratings for informational articles and opinion columns.
- *The articles must at least contain 400 words*, an arbitrary amount to create social networks of sufficient sizes, while also eliminating short 'update' articles meant divert readers to opinion columns.

Thus, we selected articles of the following three topics:

- **Topic 1:** Senate acquits Donald Trump after impeachment trial (February 5, 2020)
- **Topic 2:** Donald Trump declares national emergency over COVID-19 (March 13, 2020)
- **Topic 3:** Donald Trump declares national emergency over the U.S. Southern border (February 15, 2019)

Each topic contains 27 separate articles from different U.S. online news media. Although we prioritized mainstream sites, the bias scores are kept statistically random for each topic.

Since national emergencies are great exemplars of topics satisfying the criteria, we selected two of them. From the list of American politics events in the last two years listed on *Wikipedia*[5], few completely satisfy the criteria and include diverse coverage. (For instance, although 'House elects Nancy Pelosi as speaker' meets the criteria, the topic lacks Conservative non-opinion articles.)

3.3 NER Processing

3.3.1 Text Processing. Before processing, we formatted the articles into text files, leaving only titles, subtitles, headings, and texts. Line breaks are deleted, with periods added after lines without them at the end. We applied anaphora resolution with `neurocoref`. Applying this before entity extraction allows `neurocoref` to parse texts with context (e.g. grammar) to improve accuracy.

3.3.2 Entity Extraction. To extract entities, we employed `spaCy` [11]. `spaCy` supports 18 entity types, but due to our focus on political bias, we selected four entity types: `PERSON`, `NORP` (nationalities, religions, or political groups), `ORG` (organizations), and `GPE` (geopolitical regions). We included `ORG` and `GPE` to address `spaCy`'s occasional mislabeling (e.g. 'Trump' is consistently identified as `ORG`) due to its lack of context using pre-trained models for political articles. However, as the specific tags of entities are irrelevant to network construction, `spaCy`'s tagging consistency justifies its inaccurate labeling.

3.3.3 Name Resolution. Name resolution matches first and last names, which requires a dictionary of all full names. Rather than manually creating one, we employed *Wikipedia* as a makeshift look-up table for full names of public figures. Since *Wikipedia*'s person pages follow a strict format for its first sentence, appearing in the format of '[full name] ([years lived, phonetics]) is …' (a left parenthesis always follows the full name), names can be

[3] *MBFC* displays their bias rating on spectrum images (600 pixels wide) scaled numerically, which we used as the precise bias scores.
[4] *AS* categorizes bias as left, lean left, center, lean right, and right, which we numericized to 1, 2, 3, 4, and 5, respectively.

[5] https://w.wiki/Wmu and https://w.wiki/Wmv

easily extracted with string indexing [14]. This simplifies the name resolution algorithm into two steps:

- *Creation of a mini name-dictionary.* For all multi-word `PERSON` (prevents ambiguation), employ *Wikipedia* to obtain the full name. Disregard strings longer than seven words (indicates non-person pages). Add the name to a local dictionary after using `python-nameparser`.
- *Name replacement.* All entities with matching dictionary words are replaced with their full names. People with identical last names are rare enough to be ignored. The brevity of news articles often lacks enough context for gender assignation and title recognition methods more common in fiction or long-form articles [9, 14, 16].

3.3.4 Network Construction. For extracting social networks, we chose the co-occurrence network approach, where connections between nodes from coexistence in the same 'window of text' [16]. Although sensible for fiction, paragraph-length co-occurence can be heavily influenced by authors' line break habits, which are amplified in news articles due to its short length. The alternative of sentence-level co-occurence from [9], conversely, often produces networks with little edges. Thus, we define the window as a moving range of five sentences, a typical length for short paragraphs.

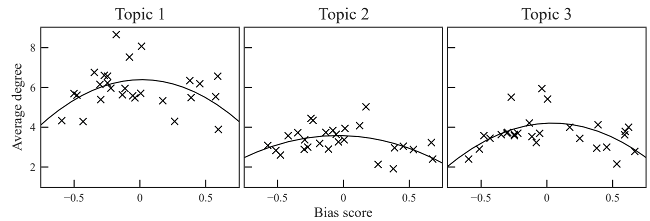
3.3.5 Network Features. The method from [9] examines eight network features. Out of the eight, we selected features with lower p-values in the results of [9] and avoided conceptually similar features (e.g. network diameter and graph density), leaving the following three:

(1) **Average degree** measures the mean degree (number of edges) of all nodes in a network. In news articles, average degree represents the frequency of entity connection, indicating characteristics like cross-referencing and average importance of entities.
(2) **Average clustering coefficient** is the mean of clustering coefficients of all nodes in a network, intuitively as the 'unique-path' indicator or "all-my-neighbors-know-each-other" coefficient [9]. To visualize, star networks have clustering coefficients of 0 (minimum), and mesh networks have 1 (maximum). In practice, this value indicates the 'closeness' of political entities, in relevancy to topic or political orientation.
(3) **Average path length** measures the mean shortest path between all nodes in a network. (With subgraphs, it is the weighted average of the average path lengths of all subgraphs.) In news articles, the value indicates the range of political entities and the presence (or lack thereof) of distinct entity groups.

4 ANALYSIS AND RESULTS

4.1 Statistical Results

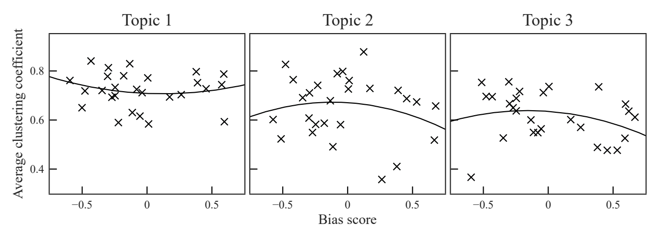
To answer the first research question, we examine the statistical relationships between the independent variables of bias score and word count with the dependent variables of average degree, average clustering coefficient, and average path length. When bias score is the independent variable, we define the trend as parabolic based on the data points, suggesting that Liberal and Conservative bias are similar, and what differs is bias and neutrality. When word count is the independent variable, we define the trend as linear. Word count is examined for its logical connection with the network features, especially in shorter texts, and it acts as an evaluation tool for answering the second research question.



$\overline{x} = 5.9189$ $p = 0.0303$ $\overline{x} = 3.3116$ $p = 0.0190$ $\overline{x} = 3.6928$ $p = 0.0085$

Figure 1: Bias score to average degree

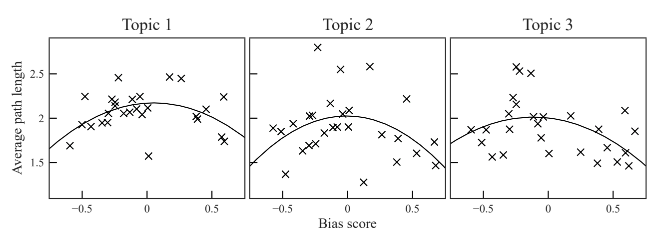
Figure 1 displays statistically significant negative parabolic trends across all three topics. There are two interpretations of this result. One interpretation is that politically biased articles tend to report on entities separately, reporting the actions of people; instead, neutral articles do so with frequent cross-referencing, focusing on 'reactions' or quotes (places entities in closer proximity). The former's inclination for explanations instead of quotes or more factual information is thus one of the sources of its bias. Another interpretation is that biased articles tend to mention entities in separate clusters (Democrats and Republicans), whereas neutral ones do so more dispersed (chronologically or by relevancy). Consequently, the former can shift tones to favor one 'cluster,' presenting more 'story-like' articles with protagonists and antagonists which lack objectivity.



$\overline{x} = 0.7201$ $p = 0.5387$ $\overline{x} = 0.6534$ $p = 0.3498$ $\overline{x} = 0.6167$ $p = 0.2399$

Figure 2: Bias score to average clustering coefficient

Figure 2 displays no statistical significance, suggesting that political bias does not necessarily affect entity 'closeness'. However, there are still significant variances of average clustering coefficients, indicating that it is not necessarily limited by text types.



$\overline{x} = 2.0703$ $p = 0.0396$ $\overline{x} = 1.8987$ $p = 0.0441$ $\overline{x} = 1.8916$ $p = 0.0349$

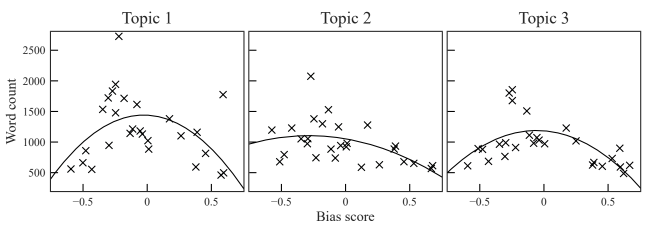
Figure 3: Bias score to average path length

Figure 3 displays statistically significant negative parabolic trends across all three topics. This suggests that politically biased articles

Table 2: Network feature statistics

Statistic	Topic	Bias score				Word count		
		Average degree	*Clustering coefficient*	*Average path length*	*Word count*	*Average degree*	*Clustering coefficient*	*Average path length*
Mean	**1**	5.9189	0.7201	2.0703	1203.8	5.9189	0.7201	2.0703
	2	3.3116	0.6534	1.8987	983.2	3.3116	0.6534	1.8987
	3	3.6928	0.6167	1.8916	969.8	3.6928	0.6167	1.8916
P-value	**1**	0.0303	0.5387	0.0396	0.0159	0.0153	0.4053	0.0017
	2	0.0190	0.3498	0.0441	−0.6052	0.1615	0.6913	0.1047
	3	0.0085	0.2399	0.0349	0.0008	0.0361	0.3986	0.0000
Vertex	**1**	0.0146	0.1200	0.0518	−0.0272			
	2	−0.0420	−0.1154	−0.0083	−0.2863			
	3	0.0426	−0.1578	−0.1145	−0.0578			

tend to report entities more politically clustered, whereas neutral ones tend to report with more dispersion, including more diverse opinions which decreases entity proximity. This is largely consistent with the second interpretation of Figure 1.

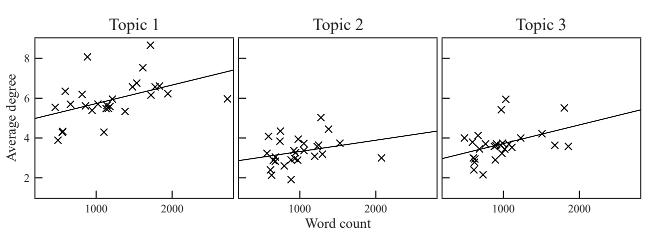


$\overline{x} = 1203.8$ $p = 0.0159$ $\overline{x} = 983.2$ $p = 0.0593$ $\overline{x} = 969.9$ $p = 0.0008$

Figure 4: Bias score to word count

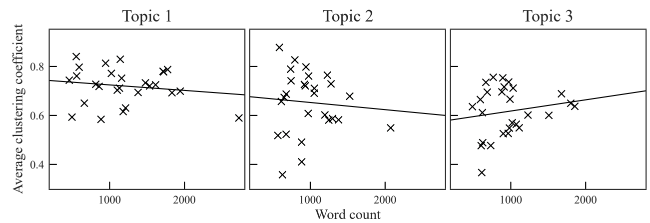
Figure 4 displays negative parabolic trends across all three topics, with Topic 1 and 3 showing statistical significance and Topic 2 being close, with the latter trend displaying a strongly negative vertex. The trends from Topic 1 and 3 suggest that politically biased news sites tend to focus on longer opinion columns than initial informative reporting, shortening the latter. Also, this indicates that more words are required for neutrality, as political bias in informative articles tend to manifest in deliberate omissions than overt language, and reporting diverse opinions naturally occupies longer lengths. This is largely consistent with the findings of Figure 1 and 3.

Note how these negative parabolic trends resemble those in Figure 1 and 3. Thus, to examine how word count affects network features, independent or not from bias score, we employ Figure 5 to 7.



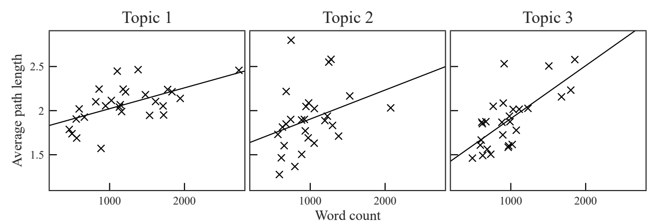
$\overline{x} = 5.9189$ $p = 0.0153$ $\overline{x} = 3.3116$ $p = 0.1615$ $\overline{x} = 3.693$ $p = 0.0361$

Figure 5: Word count to average degree



$\overline{x} = 0.7201$ $p = 0.4053$ $\overline{x} = 0.6534$ $p = 0.6913$ $\overline{x} = 0.6167$ $p = 0.3986$

Figure 6: Word count to average clustering coefficient



$\overline{x} = 2.0703$ $p = 0.0017$ $\overline{x} = 1.8987$ $p = 0.1047$ $\overline{x} = 1.8916$ $p = 0.0000$

Figure 7: Word count to average path length

Figure 5, we find consistent increasing linear trends, statistically significant for Topic 1 and 3. This trend is expected with small-world social networks like news articles, as each topic only has a small set of relevant entities, higher lengths increase the appearance frequency of entities, which generally raises the average degree. Conversely, with large-world networks, especially fictional ones, length generally correlates with node count as more entities are introduced, thus not affecting average degree. Figure 6 displays no statistical significance, which is expected from the statistical insignificance of Figure 2. Figure 7 displays consistent increasing linear trends, statistically significant for Topic 1 and 3, which is expected from the trend of Figure 4, as longer articles tend to be more neutral, thus manifesting in higher average path lengths.

4.2 Topic Comparisons

Comparing the topics horizontally shows how much more the contents of news articles affect network features than their political biases. From Table 3, we observe that Topic 1 is distinctly different than Topic 2 and Topic 3 for all three network features, shown in its statistical significance ($p \leq 0.0174$) when compared to the latter two ($p = 0.4055$). This is consistent with the topics themselves, as Topic 1 is an impeachment trial, and Topic 2 and 3 are national emergencies – the articles are structurally different, which manifests in their network features. Furthermore, examining Table 2, we find that Topic 1 has higher mean values than Topic 2 and 3, which is expected, as the former's purer political nature and relevance to the Senate allows articles to include more entities with cross-referencing, increasing all three network features and word count.

Furthermore, we observe less statistical significance with Topic 2 with Figure 5 to 7, as Topic 2 has a left-transformed negative parabolic trend in Figure 4, shifting the 'neutral' center of Topic 2 articles to more Liberal – the articles are generally more Conservative than those of Topic 1 and 3. While the latter two are controversial events for any President, with article positivity/negativity depending on the President's political party, it is difficult for the former – declaration of a national emergency over a global pandemic – to be portrayed negatively: the action seems reasonable. As Donald Trump is Republican, articles would naturally lean Conservative for their more positive nature.

Table 3: Topic difference significance comparisons

Comparison	**P-value**		
	Average degree	*Clustering coefficient*	*Average path length*
Topic 1 & 2	0.0000	0.0174	0.0073
Topic 1 & 3	0.0000	0.0002	0.0034
Topic 2 & 3	0.0334	0.4055	0.9608

5 CONCLUSION

In this study, we set out to expand the applications of utilizing NER for social network extraction and analysis, applying it to online news articles while examining the relationship between political bias and network features.

We found that, in terms of network features, there are little differences between Liberal and Conservative-bias, but what differs is neutrality and bias itself. This manifested in the statistical significance of several relationships: bias score to average degree, average path length, & word count (negative parabolic trends), and word count to average degree & average path length (positive linear trends). Furthermore, biased news sites tend to cluster entities by political orientation, and neutral articles tend to employ more cross-referencing of politically different entities.

Social network extraction and the analysis of network features can be over-dependent on textual characteristics like topic and text length, despite its ability to provide insights on political bias in news articles. This was demonstrated in the greater statistical significance of horizontal topic comparisons compared with bias score to network features, suggesting that the content of articles have much larger effects on its social network than political bias. More importantly, the strong correlations in Figure 5 and 7, along with the similarities of Figure 4 to Figure 1 and 3, indicate the possibility that political bias and network features have little causation, but that text length affects social networks more significantly – bias may only have causation with length.

Due to the dependence of network features on topic content, the limitation of articles severely restricted our corpus size, allowing outliers to drastically alter the statistical results. Thus, future work is to strive for larger, randomly selected corpora to evaluate the generalizability of our conclusion to any news articles, including more complex statistical analyses of the correlation of political bias to network features. Regarding the prominent effects of word count, it questions most studies exploring the correlation of textual characteristics to social network features, especially those that disregard text length. For instance, perhaps the statistically insignificant results of [9] are due to the varying lengths of the novels. Therefore, future work is to explore the relationships of text length and network features, while examining text processing methods which may circumvent this problem, filling in the gap of a vital topic absent in literature.

REFERENCES

[1] Erick Elejalde, Leo Ferres, and Eelco Herder. 2018. On the nature of real and perceived bias in the mainstream media. 13, 3 (2018), 1–28. DOI: http://dx.doi.org/10.1371/journal.pone.0193765

[2] AllSides.com. 2020. Media Bias Ratings. (2020). https://www.allsides.com/media-bias/media-bias-ratings

[3] Vanessa Otero. 2019. White Paper. (Aug 2019). https://www.adfontesmedia.com/white-paper-multi-analyst-ratings-project-august-2019

[4] Media Bias Fact Check. 2020. Methodology. (Jan 2020). https://mediabiasfactcheck.com/methodology

[5] National Institute of Standards and Technology. 1995. Message Understanding Conferences (MUC). (Oct 1995). https://www-nlpir.nist.gov/related_projects/tipster/muc.htm

[6] Nancy A. Chinchor. 1998. Overview of MUC-7/MET-2. In *Proceedings of the Seventh Message Understanding Conference (MUC-7)*. 1–4. https://www.aclweb.org/anthology/M98-1001

[7] Stephen Mayhew, Chen-Tse Tsai, and Dan Roth. 2017. Cheap Translation for Cross-Lingual Named Entity Recognition. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. 2536–2545. DOI: http://dx.doi.org/10.18653/v1/D17-1269

[8] Anna Śniegula, Aneta Poniszewska-Marańda, and Łukasz Chomątek. 2019. Study of Named Entity Recognition methods in biomedical field. 160, 2019 (2019), 260–265. DOI: http://dx.doi.org/10.1016/j.procs.2019.09.466

[9] Niels Dekker, Tobias Kuhn, and Marieke van Erp. 2019. Evaluating named entity recognition tools for extracting social networks from novels. (2019), 1–20. DOI: http://dx.doi.org/10.7717/peerj-cs.189

[10] Indra Budi and Stéphane Bressan. 2003. Association Rules Mining for Name Entity Recognition. In *Proceedings of the Fourth International Conference on Web information Systems Engineering*. 325–328. DOI: http://dx.doi.org/10.1109/WISE.2003.1254504

[11] Matthew Honnibal and Ines Montani. 2017. spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. (2017). https://github.com/explosion/spaCy

[12] David Bamman, Ted Underwood, and Noah A. Smith. 2014. A Bayesian Mixed Effects Model of Literary Character. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*. 370–379. DOI: http://dx.doi.org/10.3115/v1/P14-1035

[13] Jenny Rose Finkel, Trond Grenager, and Christopher Manning. 2005. Incorporating Non-local Information into Information Extraction Systems by Gibbs Sampling. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*. 363–370. https://nlp.stanford.edu/~manning/papers/gibbscrf3.pdf

[14] Jun'ichi Kazama and Kentaro Torisawa. 2007. Exploiting Wikipedia as External Knowledge for Named Entity Recognition. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*. 698–707. http://www.jaist.ac.jp/~kazama/papers/kazama_emnlp07b_final.pdf

[15] Alexander Junge and Lars Juhl Jensen. 2019. CoCoScore: context-aware co-occurrence scoring for text mining applications using distant supervision. 36, 1 (2019), 264–271. DOI: http://dx.doi.org/10.1093/bioinformatics/btz490

[16] Mariona Coll Ardanuy and Caroline Sporleder. 2014. Structure-based Clustering of Novels. In *Proceedings of the 3rd Workshop on Computational Linguistics for Literature (CLfL)*. 31–39. DOI: http://dx.doi.org/10.3115/v1/W14-0905

A DATA AVALIABILITY

All statistical results can be found at *GitHub*: https://github.com/kuan-heng-lin/ner-social-network-bias.