Into the Dynamics of Interpersonal Relationships from Diachronic Documents: Text-based Temporal Social Network Construction

Chieh-Yu Lee*, Christian Henriot†, Hua-Yuan Hsueh‡, Jyi-Shane Liu* and Hen-Hsen Huang§
*Department of Computer Science, National Chengchi University, Taipei City, Taiwan
Emails: 109753133@nccu.edu.tw, liujs@nccu.edu.tw
†Institute of Asian Studies, Aix-Marseille University, Aix-en-Provence, France
Email: christian.henriot@univ-amu.fr
‡Graduate Institute of Taiwan History, National Chengchi University, Taipei City, Taiwan
Email: hy5595@nccu.edu.tw
§Institute of Information Science, Academia Sinica, Taipei City, Taiwan
Email: hhhuang@iis.sinica.edu.tw

***Abstract*—In this work, we propose an alternative approach for constructing temporal social networks from diachronic documents by leveraging both graph and textual information. Our framework utilizes a randomized relation extraction model to extract interpersonal relationships among people from long-form documents and then enhances the extracted social network through inference on a temporal knowledge graph. The effectiveness of our approach is demonstrated through experiments on two datasets, including a politician's diary and a newspaper archive, and it has potential applications in various interdisciplinary fields such as computational politics and computational history.**

***Index Terms*—interpersonal relation extraction, sentence compression, temporal graph neural network, diachronic document processing, dynamics of social networks**

I. Introduction

The study of interpersonal relationships plays a critical role in various fields including epidemiology, politics, and history [1], [2]. Representing relationships computationally as a social network enables the use of computational methods for analyzing the network [3]. However, there are two key challenges in this field: constructing social networks for large groups is challenging without expert help and past research has ignored the fact that relationships change over time by treating them as static.

This paper introduces a new text-based method for analyzing interpersonal relationships dynamics. Our method derives social networks from textual sources like newspapers, diaries, and government documents without manual intervention. The core of our approach is an information extraction model that recognizes individuals and predicts their relationships at specific time points. We enhance the BERT-based model for relation extraction and achieve superior results compared to state-of-the-art models.

Our extraction model is applied to diachronic data to construct a temporal social network showing the evolution of relationships at scale. However, there may be inaccuracies due to misinformation in source documents or errors in the extraction model, resulting in an incomplete network. Thus, we further enhance the temporal social network by predicting the missing links on a temporal graph network (TGN) [4]. By analyzing the graph, we can fill in missing information and forecast future relationship changes.

Our framework allows for the analysis of interpersonal relationship dynamics using a TGN. It models changes in graph structures and provides a new approach for discovering insights from large amounts of unstructured textual data. Previous works on temporal knowledge inference used structured data from knowledge bases, not unstructured textual data [5]–[10], while textual-based relation extraction methods ignore the dynamics of relationships over time [11], [12]. To the best of our knowledge, this work is the first to model dynamic relationships directly from unstructured textual data. The contributions of this work are summarized as follows.

1) We introduce a new computational approach for analyzing the dynamics of interpersonal relationships using unstructured textual data, such as newspapers, diaries, and government documents. Our approach automatically derives a social network from these sources and refines it using temporal graph network analysis.
2) We propose a novel randomized relation extraction model that is effective in identifying relationships between individuals in long text.
3) Our framework not only recovers missing relationships but also forecasts future changes in the social network, providing valuable insights for researchers in fields such as epidemiology, politics, and history.
4) We have created two datasets, one based on a politician's diary and the other on a newspaper archive, to demonstrate the effectiveness of our approach. These datasets will be made available to the research community.

II. Related Work

A. Relation Extraction

Main approaches to relation extraction could be roughly divided into distant supervision and fully supervision [11]. More recent work on distant supervised relation extraction has explored how to minimize the error when generating training triplets. Beside considering external knowledge [12], previous study also considers temporal information. The study of [13] focuses on the influence of time expression on relations, the work of [14] employs temporal rule-based sieve, the study of [15] explores temporal reasoning, and the work [16] applies graph neural network to do relation reasoning.

Although distant supervision can improve the efficiency of building datasets by extracting relations automatically, the noise and errors are very significant. Especially in historical data, many people have aliases (e.g., Wang Shijie, a Chinese politician who styled himself Xue Ting), which would cause errors when existing relationship pairs align with the sentences. For application domains where the data are mostly private (e.g., the internal records of non-public individuals), external data such as Wikidata can only provide limited information or background knowledge. Therefore, a small scale of annotations are necessary in the specific domain.

Graph neural network is shown capable of addressing the problem that the existing model cannot reason about relational information, and dealing with a group of entity pairs and their relations. In our case, our model benefits from exploitation of multi-hop inference on the temporal knowledge graph, and finds the appropriate number of hops for the relations of different attributes.

B. Temporal Knowledge Graph

As temporal knowledge graph attracts the attention in recent years, various approaches have been made such as Translational distance based TTransE [5] and HyTE [6], and decomposition based ConT [17] and TNTComplEX [7]. However the conversion from traditional knowledge graph to temporal knowledge graph has the two disadvantages. (1) Each entity is discontinuous at different times because it learns knowledge on the graph independently. (2) A limitation of the model is that it is difficult to explain whether an inference is based on existing knowledge.

The graph attention network with attention as the learning structure was proposed [18]. In addition to using neighbor node sampling, it will also weight the attention value according to the importance of different neighbor nodes to the target node. Aiming at adding temporal information into the graph attention network, the attention weights can be adjusted by learning from a series of knowledge graph snapshots in every timestamp like EvolveGCN, ySAT [19], RE-NET [10] based on GCN combined with RNN architecture, and CyGNet [8]. Another study cooperates with message passing to propagate attention weights to achieve node feature update [9]. In our case, the social network at a time can be regarded as a snapshot of the temporal knowledge graph. We focus on the change of edges (i.e., interpersonal relationships) and the node representation due to the change in the overall structure of the graph, but message passing method puts more emphasis on the nodes/edges changes in the graph that impact on the task or graph.

Lei Chen's Diary	Training	Validation	Test
Time span	1956-1958	1959	1960
# People	1,069	390	338
# Relation Pairs	2,326	802	723
Shen Pao	**Training**	**Validation**	**Test**
Time Span	1947/1-6	1947/7-9	1947/10-12
# People	462	259	210
# Relation Pairs	2,415	1,242	576

TABLE I
Statistics of Our Datasets

III. Dataset Construction

In order to demonstrate the applicability of our approach to real-world scenarios, we aim to validate our method on various types of time-stamped texts such as diaries, letters, and news articles. To this end, we have created two datasets based on the diary of Lei Chen, a Chinese politician, and the archive of Shen Pao, a newspaper published from 1872 to 1949 in Shanghai, China.

The Lei Chen's Diary dataset contains records of the founding of the "Free China" magazine from 1950 to 1960, and the interactions of the author with the people of the time. On the other hand, our Shen Pao dataset comprises of news articles from 1947, which provide insights into the interactions between political figures of the time. The time granularity of the Lei Chen dataset is one year, while that of the Shen Pao dataset is a quarter (i.e., three months). These datasets have been chosen because they vary greatly in terms of writing style and time granularity, thus allowing us to evaluate the generalizability of our approach on two very different datasets.

The construction of the Shen Pao and Lei Chen's diary personal relationship datasets involves annotation of people's names and their interpersonal relationships. We first extract the names of individuals from the documents $X_1, X2, ..., X_t$ using a named entity recognition (NER) model provided by the CKIP toolkit.[1] We then manually review and correct the extracted named entities. For every pair of individuals mentioned in the same document, we label their interpersonal relationship as revealed by the text. As a result, Table I shows the statistics of the two datasets.

IV. Methodology

As shown in Fig. 1, our framework consists of three components, including named entity recognition, relation extraction, and temporal relation inference. The overview of our model is shown in Figure 2. Given a collection of diachronic data $\mathbf{x} = (x_1, x_2, ..., x_T)$ over a time span from 1 to T, we attempt to construct a temporal social network G that consists of

[1] https://github.com/ckiplab/ckiptagger

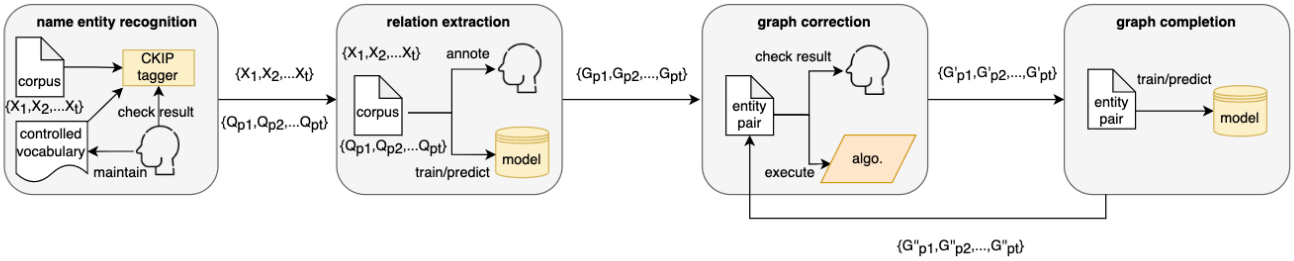

Fig. 1. The Dataflow of Our System for Text-based Temporal Social Network Construction

T snapshots as $(g_1, g_2, ..., g_T)$ corresponding to the T time points in $\mathbf{x}$. Each data entry x_t in $\mathbf{x}$ is a set of documents in which the events at the time point t are recorded in natural language. From a document at t, we aim to identify all the people from the text and extract the interpersonal relationship between every pair of individuals. An extracted relationship can be denoted as (s, r, o, t), where r indicates the relationship between two individuals s and o as the subject and the object, respectively, at the timestamp t. All the extracted relationships will be added to G to form the final temporal social network. As a result, additional interpersonal relations can be automatically deduced by performing link prediction on G. In other words, we can predict not only the missing relations at an existing time point $t \in [1, T]$ but also the forthcoming relations at $t > T$. The details are given in the following subsections.

A. Named Entity Recognition

NER is a widely-used natural language processing (NLP) technique that identifies people, organizations, locations, time expressions, and other named entities from textual data. Our focus is on identifying people, as our goal is to construct their social network. As NER is a well-established technology, it is not the primary focus of this study. Instead, we use the CKIP toolkit and tailor it to our needs by performing a specialized NER model that identifies all people mentioned in the text. We further supplement the model with a controlled vocabulary of authorized terms for entity linking.

B. Relation Extraction with Randomized Sentence Compression

Given a piece of contextual information x, the relation extraction model aims to determine whether an interpersonal relationship r exists between a subject s and an object o. Because more than one relationship possibly exists between a pair of s and o, the model can be defined as a probabilistic multi-label classifier as follows.

$$Pr(r|s, o, x) > \tau \tag{1}$$

where $r \in R$ is an interpersonal relationship from of a pre-defined set R, and τ is the threshold of the probability. In this

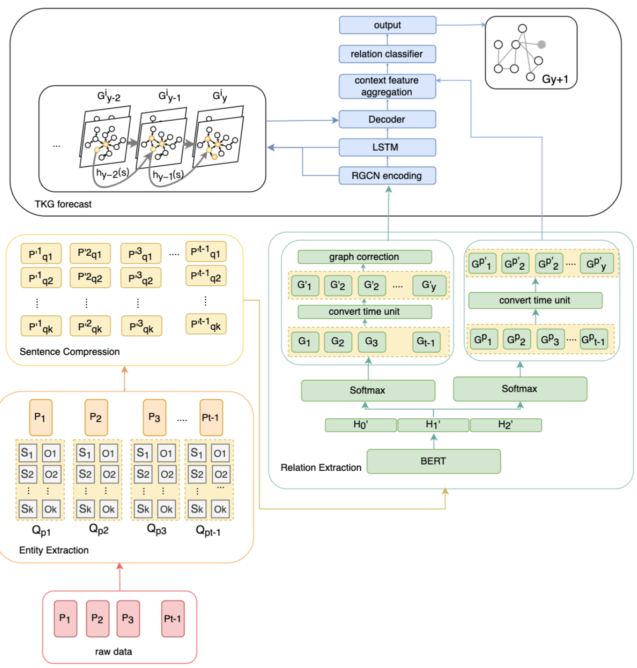

Fig. 2. Overview of Our Framework

work, R = {Work-with, Identify with, Conflict-with, Other, No-Relation} indicates three interpersonal relations and no relation at all. More complex situations with multiple relation extraction were explored in previous work [20]. Transformer-based models such as BERT achieve promising performances in many NLP tasks including relation extraction [21]. Our model for relation extraction is also built on the top of BERT to determine the relation between s and o given the context x.

One limitation of current Transformer-based models is their restriction on input size. For example, the pre-trained BERT model has a maximum input size of 512 tokens. Even the recent large language model (LLM), ChatGPT, also has an input limitation up to 4,096 tokens. In our scenario, the context x can be a very long document that exceeds this limit, with the subject s and object o appearing in different parts of the document and separated by a significant distance. Recent studies have shown that handling long inputs can result in performance degradation and increased computation cost for Transformer-based models [22]. To address this limitation, we aim to propose a sentence compression algorithm that can shorten the input for Transformer-based NLP models. This paper presents a novel randomized relation extraction model that extracts relationships between entities in a long document by randomly sampling segments of the document, enabling general NLP model to capture relationships between entities that may be far apart in the text.

To reduce the length of x, we employ a tree-based extractive sentence compression model to each sentence in x, obtaining x', the compressed version of x. Sentence compression is a task that aims to condense a sentence by reducing its length. Typically, extractive sentence compression prunes a sentence by removing less important constituents [23]. By utilizing information from constituency parsing, we carefully select a number of parts-of-speech and syntactic constituents that are less informative for relation extraction. The details of our compression algorithm are presented in Algorithm 1.

First, the original input x is parsed into T using the CKIP tagger and the Stanford parser.[2] From the parse tree T, the Word-Pruning step removes all leaf nodes (i.e. words) belonging to the following categories.

1) Adverbs of quantity, degree, place, manner, and frequency
2) Adjectives
3) Attributive adjective classifiers
4) Particles

Then, the step of Short-Phrase-Pruning removes the following non-leaf nodes in T.

1) Determinate noun phrases (DNPs)
2) Determinate verb phrases (DVPs)
3) Temporal clauses which denote exact time (e.g., "it's four o'clock")

Note that the phrases that contain either the subject s or the object o will be preserved. The step of Parenthesis-Pruning further removes all the words in parentheses or brackets. The parenthesis and bracket marks themselves will also be removed. When the resultant T forms a sentence that is still longer than the maximum length L, the step of Complementizer-Phrase-Pruning performs to remove all complementizer phrases that contain neither s nor o in T. Furthermore, the step of Random-Phrase-Prune, which picks up a random sentence/clause/phrase that contains neither s nor o and removes it from T, performs to prune T iteratively until the length criterion satisfies.

Afterward, we perform the relation extraction model on the shortened x' instead of the original x to find the relation r with the highest probability $P(r|x', s, o)$. As x' is still a valid and fluent natural language sentence that retains important information in the original x, x' can be handled by general NLP models, including LLMs.

[2]https://nlp.stanford.edu/software/lex-parser.shtml

Algorithm 1: Randomized Sentence Compression

```
Data: the original input x and the pair of (s, o) that
      we aim to extract their relations
Result: x', the compressed version of x
T ← Dependency-Parsing(x);
T ← Word-Pruning(T);
T ← Short-Phrase-Pruning(T, s, o);
T ← Parenthesis-Pruning(T);
if |T| > L then
    T ← Complementizer-Phrase-Pruning(T, s, o);
    while |T| > L do
        T ← Random-Phrase-Pruning(T, s, o);
    end
end
x' ← Preorder-Traversal(T);
```

Given a collection of documents or passages written over a specific time period, we can construct a temporal social network by utilizing named entity recognition and relation extraction models to extract all individuals and their interpersonal relationships at different points in time. However, the interpersonal relationships extracted from the text may be incomplete or inaccurate. To improve the accuracy of the temporal social network, we model it as a temporal graph network and apply inference techniques to the graph.

C. Inference over the Temporal Social Network

Inspired by the task of link prediction for knowledge graph completion, we fine-tune the temporal social network to predict future interpersonal relationships. In our scenario, link prediction on the social network helps researchers uncover relations between individuals that were not recorded in the documentation.

Given g_t, the graph at time point t, we aim to predict whether a missing edge with type r between vertices s and o should be added to g_t for completeness. The prediction can be made by estimating the likelihood that g_t contains the edge with interpersonal relationship r between s and o. Temporal graph networks facilitate link prediction by introducing the dimension of time [24]. To predict whether a missing edge should be added to g_t, information from previous states $(g_1, g_2, ..., g_t)$ can be taken into account. Initially, the temporal social network $G = (g_1, g_2, ..., g_T)$ can be obtained by applying our relation extraction model to diachronic data over T time points. We add an edge of the interpersonal relationship r between s and o at time point t by measuring the probability recursively:

$$Pr(r|g_t, s, o) = \alpha Pr(r|g_{t-1}, s, o) + (1-\alpha) Pr(r|x, s, o, t)$$

where $Pr(r|x, s, o, t)$ is the probability estimated by our relation extraction model given the contextual information at the time point t, $Pr(r|g_{t-1}, s, o)$ is determined by the temporal graph network within the time span from 1 to $t-1$ recursively, and α is a hyper-parameter determining the weight of the two factors. The loss is calculated as:

$$\mathcal{L} = -\sum_{(s,r,o,t)\in G} \log Pr(r_t|s_t,o_t)+\log Pr(o_t|s_t)+\log Pr(s_t) \tag{2}$$

In the temporal graph G, we learn the representation of each node by using the information from its multi-hop neighboring nodes and from its preceding/succeeding states. Based on multi-relational graph (R-GCN) aggregator [25], the embedding of a node s at a timestamp t is aggregated from the neighbors of s at t by considering different relation types as follows.

$$h_s^{(l+1)} = \sigma\left(\sum_{r\in R}\sum_{v\in N_t^{(s,r)}} C_s W_r^{(l)} h_o^{(l)} + W_0^{(l)} h_s^l\right) \tag{3}$$

where l indicates l-th layer of the neural networks, $N_t^{(s,r)}$ are the neighbors of s which are connected by a relation type r at the timestamp t. Every relation type r has its own weight matrix W_r. The diachronic information of the graph is handled by the recurrent event network (RE-NET) [10] to encode recurrent events over time. In order to capture different time axes, a bidirectional LSTM (BiLSTM) [26] is employed to encode the temporal events.

V. Experiments

A. Experimental Setup

Due to the nature of diachronic data, we divide the temporal data into training, validation, and test sets based on their chronological order. In the experiments, we train our model using the oldest training set and evaluate it on the newest test set. The set in between is used for validation. It is worth noting that the time unit granularity in the Lei Chen dataset is one year, while in the Shen Pao dataset it's one quarter (i.e. three months). The division of the datasets is illustrated in Table I. For the hyper-parameter α, we set it to 0.9, and for the variable k (which represents the number of randomized sentence compression), we set it to 5.

	Lei Chen's Diary			Shen Pao		
Method	P	R	F	P	R	F
BERT	0.5389	0.5472	0.5421	0.7057	0.7518	0.7278
BERT w/ cmp	0.5596	0.6094	0.5834	0.7178	0.7334	0.7250
Our Model	0.6245	0.6612	0.6423	0.7295	0.7445	0.7367

TABLE II
Results of relation extraction, reported in Micro-averaged Precision (P), Recall (R), and F-score (F)

B. Results of Relation Extraction

Table II shows the performances of our method for relation extraction on the Lei Chen and the Shen Pao datasets. The results of the experiments demonstrate that incorporating our randomized sentence compression method leads to an improvement in the performance of the vanilla BERT model.

Original:
... 上午八時三十分雷震參加東南軍政長官公署團拜，九時參加中央黨部團拜，儀式均簡單而莊嚴，十時本在中山堂舉行中央政府各機關團拜，因已參加二次故未去... 上午同希孔去拜訪，並至辭修、至柔處賀年，並訪 PER1[尹葆宇]，悉已告 PER2[朱世明] 請其二日以前必須復電...   ... At 8:30 in the morning, [Lei Chen] participated in the worship of the Southeastern Military and Political Office, and at 9:00, he participated in the worship of the Central Party Committee. The ceremonies were simple and solemn. .. I went to visit with Hsi-Kong in the morning, and went to Tsz-Shiou and Chi-Rou to celebrate the New Year, and also visited PER1 [Yin, Pao-Yu] and told that PER2 [Chu, Shi-Ming] want him to call back in two days...
Compressed:
... 參加東南軍政長官團拜，參加中央黨部團拜，儀式簡單莊嚴，中山堂舉行中央政府團拜，參加二次未去... 同希孔去拜訪至辭修、至柔處賀，訪 PER1[尹葆宇] 告 PER2[朱世明] 請二日必須覆電...   ... [Lei Chen] participated in the worship of the Southeastern Military and Political Office and the worship of the Central Party Committee. The ceremonies were simple and solemn... I went to visit with Hsi-Kong in the morning, and went to Tsz-Shiou and Chi-Rou, and also visited PER1 [Yin, Pao-Yu] and told that PER2 [Chu, Shi-Ming] want him to call back in two days...

Fig. 3. Example of our sentence compression. In the ground-truth, the relation "Other" exists between PER1 (i.e., 尹葆宇 "Yin, Pao-Yu") and PER2 (i.e., 朱世明 "Chu, Shi-Ming").

Additionally, our relation extraction model, which is designed to work with the compressed input, outperforms baseline models. These results support the effectiveness of our approach in pruning irrelevant phrases in the input, resulting in a more refined input for the Transformer model to process.

As illustrated in Figure 3, the use of our randomized sentence compression method results in the truncation of prepositional phrases and adverbs from the original input. This allows the relation extraction model to focus on the interaction between the entities 尹葆宇 "Yin, Pao-Yu" and 朱世明 "Chu, Shi-Ming", who are the subject and object, respectively.

Because of the nature of extractive-based sentence compression, our model for relation extraction performing on the shortened context x' is precision-oriented. That is, the model is less likely to suggest a wrong relation, but tends to conclude no relation between s and o. For this reason, we repeat the randomized sentence compression k times for obtain k different shortened context x' and take the union of all relations extracted from the k inputs.

Method	Lei Chen's Diary			Shen Pao		
	P	R	F	P	R	F
TTransE	0.5115	0.6785	0.5833	0.4802	0.5682	0.5205
TA-TransE	0.4330	0.6744	0.5274	0.4941	0.6207	0.5502
TA-DisMult	0.5502	0.6378	0.5908	0.5000	0.6620	0.5697
RE-NET	0.5700	0.7330	0.6416	0.4713	0.8573	0.6082
Our Approach	0.6066	0.7598	0.6746	0.6109	0.8383	0.7068

TABLE III
Results of temporal social network forecast, reported in Micro-averaged Precision (P), Recall (R), and F-score (F)

C. Results of Temporal Social Network Inference

We evaluate our model's ability to predict future interpersonal relationships within a temporal social network by considering all information up to a given timestamp. Specifically, our goal is to predict all edges at timestamp t, using information available prior to t. We compare our model to several baseline graph models, including TTransE [5], TA-TransE [27], TA-DisMult [27], and RE-NET [10].

As outlined in Section I, making direct comparisons between our system and prior works is impractical as our model utilizes unstructured textual data as a source of information, while existing TGN models utilize structured, graph-like data. Instead of focusing on performance comparisons between the models, we aim to highlight the advantages of using a textual-based approach for constructing temporal social networks when raw, textual information is available at time t. For the baseline TGN models, we provide the correct relations from timestamps $1, 2, ..., t-1$ to initialize the graph and predict the relations at timestamp t. In contrast, our temporal graph model initializes the graph using the output from our randomized relation extraction model and does not require any prior knowledge such as the correct relations used by the baseline models.

Experimental results are presented in Table III. Our model outperforms all baseline TGN models on both datasets, with RE-NET performing second best. These results suggest the advantages of our approach, which directly builds the graph from textual data without the need for prior knowledge such as golden relations used by the baseline models. Furthermore, our model not only achieves the best performance compared to the baseline models, but also shows its ease of use in real-world applications.

VI. Conclusions

This work presents a novel approach for analyzing the dynamics of interpersonal relationships among people in a society by constructing temporal social networks from unstructured textual data. By introducing randomized sentence compression and temporal graph networks, we address the challenges in relation extraction and improve the temporal social network, creating a systematic methodology for modeling the complexities of interpersonal relationships.

Acknowledgements

This work is partially supported by National Science and Technology Council, Taiwan under grants 109-2222-E-001-004-MY3, 109-2410-H-004-150-MY2, and 112-2221-E-001-016-MY3, by Academia Sinica under grants 3006-37C4527 and 3006-37C4386, and by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 788476).

References

[1] E. U. Haq, T. Braud, Y. D. Kwon, and P. Hui, "A survey on computational politics," *IEEE Access*, vol. 8, pp. 197 379–197 406, 2020.

[2] J. Pal *et al.*, "Studying political communication on twitter: the case for small data," *Current opinion in behavioral sciences*, vol. 18, pp. 97–102, 2017.

[3] J. Perl, C. Wagner, J. Kunegis, and S. Staab, "Twitter as a political network: Predicting the following and unfollowing behavior of german politicians," in *Proceedings of the ACM Web Science Conference*, 2015, pp. 1–2.

[4] E. Rossi, B. Chamberlain, F. Frasca, D. Eynard, F. Monti, and M. Bronstein, "Temporal graph networks for deep learning on dynamic graphs," 2020.

[5] T. Jiang, T. Liu, T. Ge, L. Sha, B. Chang, S. Li, and Z. Sui, "Towards time-aware knowledge graph completion," in *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, Dec. 2016, pp. 1715–1724.

[6] S. S. Dasgupta, S. N. Ray, and P. Talukdar, "HyTE: Hyperplane-based temporally aware knowledge graph embedding," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, Oct.-Nov. 2018, pp. 2001–2011.

[7] T. Lacroix, G. Obozinski, and N. Usunier, "Tensor decompositions for temporal knowledge base completion," *arXiv preprint arXiv:2004.04926*, 2020.

[8] C. Zhu, M. Chen, C. Fan, G. Cheng, and Y. Zhan, "Learning from history: Modeling temporal knowledge graphs with sequential copy-generation networks," *arXiv preprint arXiv:2012.08492*, 2020.

[9] J. Jung, J. Jung, and U. Kang, "T-gap: Learning to walk across time for temporal knowledge graph completion," *arXiv preprint arXiv:2012.10595*, 2020.

[10] W. Jin, M. Qu, X. Jin, and X. Ren, "Recurrent event network: Autoregressive structure inferenceover temporal knowledge graphs," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Nov. 2020, pp. 6669–6683.

[11] M. Mintz, S. Bills, R. Snow, and D. Jurafsky, "Distant supervision for relation extraction without labeled data," in *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, 2009, pp. 1003–1011.

[12] A. Bastos, A. Nadgeri, K. Singh, I. O. Mulang, S. Shekarpour, J. Hoffart, and M. Kaul, "Recon: relation extraction using knowledge graph context in a graph neural network," in *Proceedings of the Web Conference 2021*, 2021, pp. 1673–1685.

[13] J. Yan, L. He, R. Huang, J. Li, and Y. Liu, "Relation extraction with temporal reasoning based on memory augmented distant supervision," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, Jun. 2019, pp. 1019–1030.

[14] P. Mirza and S. Tonelli, "Catena: Causal and temporal relation extraction from natural language texts," in *The 26th international conference on computational linguistics*. ACL, 2016, pp. 64–75.

[15] B. Zhou, K. Richardson, Q. Ning, T. Khot, A. Sabharwal, and D. Roth, "Temporal reasoning on implicit events from distant supervision," *arXiv preprint arXiv:2010.12753*, 2020.

[16] A. Nadgeri, A. Bastos, K. Singh, I. O. Mulang, J. Hoffart, S. Shekarpour, and V. Saraswat, "Kgpool: Dynamic knowledge graph context selection for relation extraction," *arXiv preprint arXiv:2106.00459*, 2021.

[17] Y. Ma, V. Tresp, and E. A. Daxberger, "Embedding models for episodic knowledge graphs," *J. Web Semant.*, vol. 59, 2019.

[18] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," *stat*, vol. 1050, p. 20, 2017.

[19] A. Sankar, Y. Wu, L. Gou, W. Zhang, and H. Yang, "Dysat: Deep neural representation learning on dynamic graphs via self-attention networks," in *Proceedings of the 13th International Conference on Web Search and Data Mining*, 2020, pp. 519–527.

[20] J. Liu, H. Ren, M. Wu, J. Wang, and H.-j. Kim, "Multiple relations extraction among multiple entities in unstructured text," *Soft Computing*, vol. 22, pp. 4295–4305, 2018.

[21] S. Wu and Y. He, "Enriching pre-trained language model with entity information for relation classification," in *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, ser. CIKM '19, 2019, p. 2361–2364.

[22] I. Beltagy, M. E. Peters, and A. Cohan, "Longformer: The long-document transformer," *arXiv preprint arXiv:2004.05150*, 2020.

[23] W. Xu and R. Grishman, "A parse-and-trim approach with information significance for Chinese sentence compression," in *Proceedings of the 2009 Workshop on Language Generation and Summarisation (UCNLG+Sum 2009)*, Aug. 2009, pp. 48–55.

[24] R. Goel, S. M. Kazemi, M. Brubaker, and P. Poupart, "Diachronic embedding for temporal knowledge graph completion," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 04, pp. 3988–3995, Apr. 2020. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/5815

[25] M. Schlichtkrull, T. N. Kipf, P. Bloem, R. Van Den Berg, I. Titov, and M. Welling, "Modeling relational data with graph convolutional networks," in *European Semantic Web Conference*. Springer, 2018, pp. 593–607.

[26] A. Graves, S. Fernández, and J. Schmidhuber, "Bidirectional lstm networks for improved phoneme classification and recognition," in *Proceedings of the 15th International Conference on Artificial Neural Networks: Formal Models and Their Applications - Volume Part II*, ser. ICANN'05. Berlin, Heidelberg: Springer-Verlag, 2005, p. 799–804.

[27] A. García-Durán, S. Dumancic, and M. Niepert, "Learning sequence encoders for temporal knowledge graph completion," in *EMNLP*, 2018.