Team KSU at PAN 2025: Multi-Author Writing Style Analysis in English Texts Based on Graph Convolutional Networks

Notebook for PAN at CLEF 2025

Abeer Saad Alsheddi[1,2,*], Mohamed El Bachir Menai[1]

[1]Computer Science Department, King Saud University, Riyadh, Saudi Arabia
[2]Computer Science Department, Imam Mohammad Ibn Saud Islamic University, Riyadh, Saudi Arabia

Abstract

The Multi-Author Writing Style Analysis (MAWSA) task asks to find the locations of writing style changes at different text levels. This task can assist in other applications such as plagiarism, security, and commerce. 2020, existing MAWSA models have commonly represented any boundary between two consecutive segments by joining them. The representations of these joined segments then serve as the input for these models. This join may lose style features within each segment. In this paper, the proposed method exploits relationships between segments using Graph Convolutional Networks (GCNs). Boundaries and segment representations are depicted independently. The PAN 2025 dataset is provided at three different levels of topic distributions: easy, medium, and hard, while changes appear on the sentence level. The trained model, named STAR-GCN-MAWSA, achieved an F_1-score of 0.857, 0.764, and 0.662 for easy, medium, and hard MAWSA instances on validation sets, respectively.

Keywords

Style change detection, Multi-author Analysis, Graph convolutional networks, Pretrained models

1. Introduction

PAN[1] organizes a series of scientific competitions to promote research on stylometry and digital text forensics. It has provided a Multi-Author Writing Style Analysis (MAWSA) task since 2017. This task focuses on differentiating author styles within multi-authored text documents without providing comparison documents. It segments a text document. It then examines the boundary between each segment to determine whether it separates two segments written by the same author. For example, if a document is segmented into five segments S_1 to S_5, there are four boundaries located between these five segments B_1 to B_4. Developing MAWSA models can assist other practical applications such as plagiarism, security, and commerce. In plagiarism, MAWSA solutions can suggest potential plagiarism cases by identifying changes in writing style without comparing the suspected and source documents. In security measures, unauthorized modifications to sensitive documents can be identified to fortify the security systems. In commerce, the coherence of writing style can be improved for proofreaders and institutions by minimizing variations in writing style to adhere to a single style in their documents.

The previous PAN editions aimed to tackle the MAWSA task from different aspects by proposing different levels of subtasks, which can be categorized into four subtasks. The first subtask determined single/multi-author document, which was provided in MAWSA 2017 - MAWSA 2022 [1, 2, 3, 4, 5, 6]. The second subtask detected change positions (boundaries) on the sentence level in MAWSA 2017 and MAWSA 2022 [1] or on the paragraph level in MAWSA 2020 - MAWSA 2024 [4, 5, 6]. The third subtask identified the actual number of authors who wrote a given document in MAWSA 2019 [3]. The last subtask considered the attribution that assigned all segments uniquely to their respective authors in

CLEF 2025 Working Notes, 9 – 12 September 2025, Madrid, Spain
*Corresponding author.
asalsheddi@imamu.edu.sa (A. S. Alsheddi); menai@ksu.edu.sa (M. E. B. Menai)
0009-0000-0446-3262 (A. S. Alsheddi); 0000-0001-5981-6299 (M. E. B. Menai)

[1]https://pan.webis.de

MAWSA 2021 and MAWSA 2022 [5, 6]. In this year, MAWSA 2025 [7, 8] is related to the second subtask, detecting change positions on the sentence level, while focusing on topic diversity in datasets.

Since 2020, most existing models represent boundaries between segments by concatenating these segments. For example, the boundary B_i concatenates two segments S_i and S_{i+1} to form one pair S_iS_{i+1}. This input pair is then represented by using the representations of the pair S_iS_{i+1} as input in most existing models. This concatenation can eliminate the need to explicitly work on boundary features. However, it loses segment representations. In other words, this concatenation does not preserve the representation of each segment alone through the processing within models. Thus, segments cannot be retrieved at the end of the processing. For example, author attribution and author counting were studied in MAWSA 2019, MAWSA 2021, and MAWSA 2022, which are based on segment representations themselves. This motivates us to close this gap.

In addition, comprehending relationships between textual segments, such as words and sentences, would enhance the detection of writing styles. Graph-based solutions take a graph as input, trying to involve structural properties within the data. Graph Neural Networks (GNNs) extend existing neural networks to operate directly on graph-structured data [9]. Recently, GNN models have achieved promising results for some Natural Language Processing (NLP) tasks, such as an authorship verification task that determines whether an unknown text was written by a specific author [10, 11, 12] and semantic relationship tasks that analyze semantic relations between textual segments [13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25]. To the best of our knowledge, no existing GNN-based solution is available for the MAWSA task.

In this study, we participated in the MAWSA 2025 task by representing boundaries as standalone matrices while preserving the segment representations surrounding these boundaries. The boundary representations are learnable and updated within layers. This work also explores Graph Convolutional Networks (GCNs) [26] for MAWSA. It is considered the first to leverage the characteristics of graph neural networks to detect style changes. In this work, graphs indicate documents, nodes represent sentences, and edges state boundaries. Moreover, sentence features were extracted using STAR [27]. It is a recent pre-trained model trained on authorship representations, which is more related to MAWSA.

The remaining parts of this paper are organized as follows. Section 2 describes the task and the provided dataset in MAWSA 2025. Section 3 investigates related work for the task. Sections 4 and 5 describe the proposed approach and present its results. Conclusions are raised in Section 6.

2. Task Description and Dataset

The MAWSA 2025 task asks to determine whether the writing style changes on the sentence level in a given document. This edition pays more attention to topic diversity in datasets. Therefore, it provides the datasets with three different levels of topic diversity to decrease the use of topic information in identifying style changes [8]:

Dataset 1 (Easy): The sentences in a document cover various topics.

Dataset 2 (Medium): The sentences in a document cover fewer topics than the easy level.

Dataset 3 (Hard) All the sentences in a document cover the same topic.

The results of simple statistical analysis are shown in Table 1. Each dataset was split into training, validation, and test sets. The training and validation sets are available with ground truth labels to train and optimize proposed models. The test set is hidden until the end of the competition, and it is not publicly available so far. The average length of documents is measured as the average number of sentences per document. The average length of sentences is measured as the average number of words per sentence. The average number of writing style changes and the percentage of these changes are measured per document.

The input files in the dataset contain a row of English text *.txt. The expected output is a list of binary values representing the change of writing style within a document. Value ’0’ indicates that the same

Table 1
MAWSA 2025 dataset statistics

Dataset	Split	# Doc.	Avg. Doc. length	Avg. Sent. length	Avg. # changes
Dataset 1	train	4200	12.524	16.859	2.434 (21.123%)
(Easy)	validation	900	12.386	16.935	2.446 (21.479%)
Dataset 2	train	4200	15.004	17.653	2.977 (21.257%)
(Medium)	validation	900	15.177	17.751	3.056 (21.553%)
Dataset 3	train	4200	13.157	18.545	2.099 (17.264%)
(Hard)	validation	900	12.831	18.895	2.118 (17.900%)

author writes the two consecutive sentences and then has the same style. While Value '1' suggests that their authors are different and have unique writing styles.

3. Related Works

This section provides an overview of the methods incorporated in previous MAWSA works. According to the conducted review, three methods were adopted: statistical, classical machine learning, and deep neural networks [28].

In statistical methods, models were developed by selecting stylistic features, followed by applying statistical methods without training their models [29, 30, 31]. Khan [29, 30] defined a measure for each type of handcrafted feature, assuming a style has changed if the score is less than a threshold. Karas et al. [31] adopted a distribution test called the Wilcoxon Signed Rank Test [32] to predict the style changes.

In machine learning methods, models are obtained based on either supervised or unsupervised learning. Most of supervised-based works relied on the logistic regression and the random forest algorithms [33, 34, 35], whereas Support Vector Machine outperformed them in other works proposed MAWSA 2018 [36]. The unsupervised learning-based works mostly utilized clustering documents based on the similarity of their writing styles using the K-means clustering algorithm [34, 37, 38, 39, 40] with the cosine similarity function that outperformed Jaccard and Dice functions [37].

In deep neural network methods, features represent whole documents instead of selecting specific features. Documents in these works were fed into a CNN model [41], a Siamese neural network of one or two BiLSTM [36, 42] layers. Other works used pretrained models, such as ELECTRA [43] and BERT with a CNN layer [44, 45], an MLM head [46], or feedforward neural networks [47, 48].

4. Proposed Method

The proposed method is illustrated by Figure 1. It is based on a deep neural network architecture comprising four layers that extract more non-local features. The following paragraphs provide a more detailed description of them.

Features: The input documents were segmented into sentences. This research uses the pretrained model Style Transformer for Authorship Representations (STAR) [27] to represent sentences. STAR characterizes writing style in social media and trained for learning authorship representations. A pretrained model is a transformer with trained parameters and saved values from a large dataset. These models can adapt their parameters to better suit a particular task by retraining some or all. Pretrained models have provided high results in SOTA models. $X^{(0)}_{N\times F^{(0)}}$ in in Figure 1 indicates to the initial node representations. All the trainable parameters of STAR were frozen. Thus, no fine-tuning was performed on their parameters for extracting embeddings. This freezing allowed us to assess the models' capabilities rigorously within the constraints of our experimental setup.

Graphs: After that, the representations are structured as graphs. Each document is treated as a disjoint subgraph because the task does not look for a common writing style between two documents.

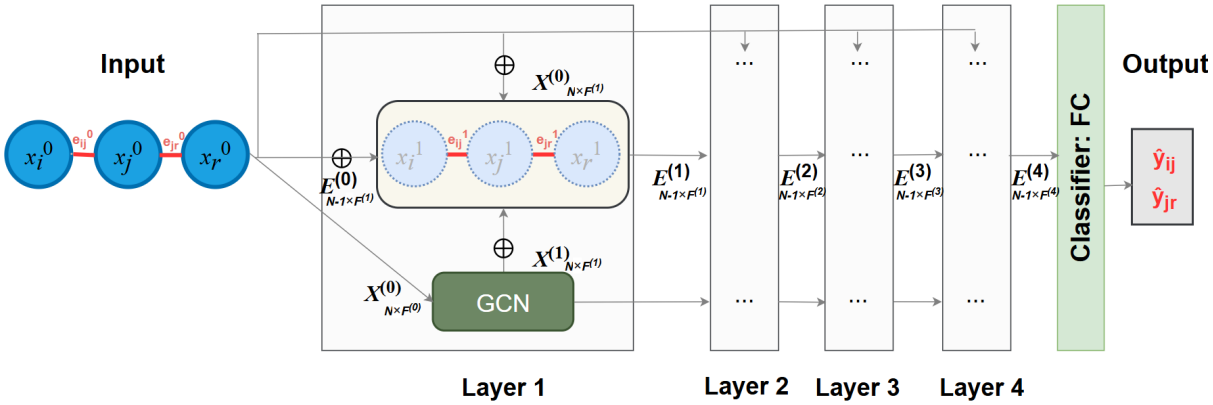

Figure 1: The proposed approach architecture

Thus, there is no direct relationship between any two documents, so they can be treated as a disjoint subgraph. Each node represents a sentence. Edges connect the preceding to the succeeding sentences. Thus, edges concern the boundaries between sentences. The edge representations are the writing styles of consecutive sentences. The input graph $\mathcal{G}^{(0)}$ has an empty set of edge representations $E^{(0)}_{N-1\times F^{(0)}}$. These representations are updated across layers and then classified for the MAWSA task. As shown in Figure 1, the graph is represented as the path graph because the task focuses on boundaries between only consecutive paragraphs.

GCNs was introduced by Kipf and Welling [26] as one of the early GNN models. GCNs are suitable for learning the representation of nodes in the input graph. GCNs help capture the relationships between nodes, in our case, between the sentence styles Equation 1 shows GCN's message-passing schema. The messages from all neighbors $\mathcal{N}(i)$ are normalized by the degrees of the neighbor $j \in \mathcal{N}(i)$ and the target i nodes. These messages are then summed to aggregate them. The aggregated messages are combined with the current target node representations to update the latter. Although GCNs are widely used for diverse NLP tasks [16, 17, 13, 24, 18, 21, 14, 25, 15], GCNs focus on node representations and can handle edge weights within the adjacency matrices as shallow edge representations.

$$\mathbf{x}_i^{(k)} = \sum_{j \in \mathcal{N}(i) \cup \{i\}} \frac{1}{\sqrt{\deg(i)} \cdot \sqrt{\deg(j)}} \cdot \left(\mathbf{W}^{\top} \cdot \mathbf{x}_j^{(k-1)} \right) \tag{1}$$

Edge representation: Every layer k outputs new edge representations. These representations aggregate three values. First, the edge representations from the previous layer $E^{(k-1)}$, whereas the first layer receives an empty set. Second, the node representations $X^{(k)}$ are generated from GCNs$^{(k)}$, where GCNs$^{(k)}$ receive the previous node representations $X^{(k-1)}$. Third, the initial node representations, $X^{(0)}$, are added to alleviate the over-smoothing issue. The symbol $\oplus$ in Figure 1 indicates to sum these three values to generate edge representations in the current layer, $E^{(k)}$. Equation (2) shows how the edge representations $\mathbf{e}_{ij}^{(k)}$ can be updated, where $\sigma(\cdot)$ represents a nonlinear activation function, *Sum* represents the summation operation as the aggregation function, x_i and x_j indicate the representations of end nodes i and j of edge e_{ij} ($j \in \mathcal{N}(i)$), and W^k is a learnable parameter in the layer k that adjusts the dimension of output representation vectors. The edge representations in each layer are updated according to the new node representation in the same layer $X^{(k)}$.

$$\mathbf{e}_{ij}^{(k)} = \mathbf{Sum}(\mathbf{W}_e^k \mathbf{e}_{ij}^{(k-1)}, \sigma(EdgeConv_k(\mathbf{x}_i^{k-1}, \mathbf{x}_j^{k-1})), \mathbf{W}_0^k \mathbf{x}_i^0). \tag{2}$$

Classification: The output graph $\mathcal{G}^{(4)}$ contains edge representations $E^{(4)}_{N-1\times F^{(4)}}$. These representations are classified by a single fully connected (FC) layer. It is followed by the activation function Sigmoid, and then are classified using a threshold of 0.5 to round outputs to 0 and 1.

5. Evaluation

5.1. Experiment Settings

Setup: Each input document is used as one batch. Dropout rates of 0.5, warmup rates of 0.1, learning rate of 2e-5, and 20 epochs were used during the training. The experiments were conducted using a personal computer with the following specifications:

- CPU: Intel(R) i7 processor up to 5.60 GHz, 64-bit
- GPU: ASUS TUF RTX 4090 24GB OC GAMING.
- RAM: 64 GB (2x32 GB) DDR5 5600 Mhz
- Programming language: Python with the seed of 42 and the PyTorch framework.

Encoder: Three candidates were selected. They are three pretrained models used with their default configuration. BERT[2] had the top usage from 2020 to 2022 [44, 46, 45, 47, 48, 49, 50], RoBERTa [3] had the top usage in the latest two editions 2023 and 2024 [51, 52, 53, 54, 55, 56, 56, 57, 51, 58], and STAR[4] is used as the third pretrained model. This study used 256 tokens as the maximum length. Any sentence exceeding this length will be truncated. The special token [CLS] was used to represent the entire sentence. All the models were developed under the same setting.

Evaluation metric: The performance of each model is measured using the F_1-score[5] metric. The macro-averaged computes each class's average separately, *change* or *not change*, and returns the average without considering the proportion of each class in the dataset.

5.2. Result and Discussion

Table 2
Experimental results on the validation sets

Encoder	Easy	Medium	Hard
BERT-GCN-MAWSA	0.85388	0.75813	0.64265
RoBERTa-GCN-MAWSA	0.78799	0.72829	0.59855
STAR-GCN-MAWSA	**0.85676**	**0.76354**	**0.66175**

Table 2 shows the results of the proposed models based on the three encoders: BERT, RoBERTa, and STAR. The results indicate that STAR significantly outperformed the others at all three levels (this model was submitted to the competition by the TIRA platform [59]). BERT achieved high F_1-score values, which are closer to STAR's results. The results indicate that STAR is better suited for the MAWSA 2025 task, as it was trained on authorship representations.

Regarding the three datasets, the performance of the proposed models decreased as the diversity level of the topics increased. At the same time, the most challenging level was Dataset 3, the absence of diversity. Several possible explanations for this result. In Dataset 3, all sentences in a document cover the same topic. This homogeneity decreased the chance of relying on topic information to detect changes in writing style. In contrast, the model trained on the easy level, Dataset 1, achieved the highest F_1-score, as it can combine topic information with style features.

Table 1 shows that the average lengths of the documents and sentences are almost similar between the training and validation sets at the same level. However, these lengths are shorter than the maximum length of the selected pretrained models. Specifically, the sentences in Dataset 1 are the shortest. This length probably decreases the feature embeddings of each sentence and then increases the difficulty of style discrimination. Besides, the proportion of change positions to the total number of boundaries is

[2]https://huggingface.co/google-bert/bert-base-uncased
[3]https://huggingface.co/FacebookAI/roberta-base
[4]https://huggingface.co/AIDA-UPM/star
[5]https://scikit-learn.org/1.5/modules/generated/sklearn.metrics.f1_score.html

less than a quarter. This small percentage leads models to train on almost the *not change* case, making it harder to detect *change* in the validation set. In particular, the number of changes has diminished as the datasets become more complicated. This bias also appears in some previous editions of MAWSA datasets [28].

It is important to bear in mind the possible bias, specifically in Datasets 1 and 2. This case may be related to the topic distributions in the training and the validation sets. The difference in the distributions may guide models to train on specific styles more than others, especially with a small size of the validation set. Further research is needed to understand the relationship between the distribution of topics and writing styles.

Table 3
Results of the ablation experiments on the validation sets

Components	Easy	Medium	Hard
Basic GCN model	0.43984	0.57777	0.45085
Adding warmup	0.47415	0.57917	0.45085
Adding prev-X	0.78490	0.72623	0.58484
Adding initial-x (Submitted)	**0.85676**	**0.76354**	**0.66175**

Ablation experiments were conducted to evaluate the components of the edge features, which were added cumulatively in the experiments. Table 3 shows the results obtained from the three datasets. First, the basic four GCN layers were developed as the baseline models, and their edge representations were measured by summing the representations of the two end nodes extracted from the fourth GCN layer. Second, a warmup mechanism optimized the model performance. Third, adding edge representations obtained from previous layers helped adjust them across layers. Fourth, initial node information was aggregated into edge representations. Table 3 shows that the fusion of both initial node and edge representations enhances the learning of edge representations for MAWSA and can mitigate the over-smoothing issue.

Table 4
Experimental results on the test sets

Encoder	Easy	Medium	Hard
STAR-GCN-MAWSA	0.507	0.747	0.467

Table 4 shows the results on the test sets shown on the Tira platform. The results obtained from the validation and test sets have revealed some intriguing disparities, despite both sets being withheld during the training process. While the validation results suggest that STAR-GCN-MAWSA performs acceptably, especially on easy instances, the test results have not mirrored this stability. This inconsistency between the results may be due to various factors. One plausible explanation could be the presence of data distribution differences between the validation and test sets, which leads to an increase in the model's sensitivity during evaluation. Investigating these discrepancies further in the future is essential through an analysis of the data distribution to ensure consistency between the sets and help achieve more stability.

Table 5
Experimental results on the validation sets based on EdgeConv (not submitted)

Encoder	Easy	Medium	Hard
BERT-EdgeConv-MAWSA	0.90346	0.76172	0.69549
RoBERTa-EdgeConv-MAWSA	0.90317	0.76753	0.70181
STAR-EdgeConv-MAWSA	**0.90670**	**0.77678**	**0.72124**

Beyond the conclusion of the official competition, our efforts to enhance the GNN-based solution for MAWSA continued. The latest advancement in this ongoing work has been achieved through the integration of an alternative GNN module: EdgeConv [60]. This specific architecture was chosen for its ability to incorporate edge representations directly into the node message-passing mechanism, thereby enabling a richer understanding of local graph structures and relationships. This EdgeConv-based solution yielded improved performance on the validation set, as detailed in Table 5. A more comprehensive description of this advanced EdgeConv-based approach and its implementation can be found in our recent work [61].

6. Conclusion

This paper set out to address the MAWSA 2025 task, which focuses on the topic diversity in datasets on the sentence level. The proposed solution mainly considers to integrate GCNs to address boundary style features independently. This method updates edge features across layers while preserving sentence representations. One of the findings from this study is that the STAR-GCN-MAWSA model outperformed the two models, BERT-GCN-MAWSA and RoBERTa-GCN-MAWSA. The other major finding is that aggregating three components for representing boundary styles achieved high results for the MAWSA task. As future work, we plan to investigate the static bias that may occur during model training and its impact on the model's performance.

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

[1] M. Tschuggnall, E. Stamatatos, B. Verhoeven, W. Daelemans, G. Specht, B. Stein, M. Potthast, Overview of the Author Identi□cation Task at PAN-2017: Style Breach Detection and Author Clustering, in: Working Notes of CLEF 2017 - Conference and Labs of the Evaluation Forum, volume 1866 of *CEUR Workshop Proceedings*, CEUR-WS.org, Dublin, Ireland, 2017, p. 22.

[2] M. Kestemont, M. Tschuggnall, E. Stamatatos, W. Daelemans, G. Specht, B. Stein, M. Potthast, Overview of the Author Identi□cation Task at PAN-2018, in: Working Notes of CLEF 2018 - Conference and Labs of the Evaluation Forum, volume 2125 of *CEUR Workshop Proceedings*, CEUR-WS.org, Avignon, France, 2018, p. 25.

[3] E. Zangerle, M. Tschuggnall, G. Specht, B. Stein, M. Potthast, Overview of the Style Change Detection Task at PAN 2019, in: Working Notes of CLEF 2019 - Conference and Labs of the Evaluation Forum, volume 2380 of *CEUR Workshop Proceedings*, CEUR-WS.org, Lugano, Switzerland, 2019, p. 11.

[4] E. Zangerle, M. Mayerl, G. Specht, M. Potthast, B. Stein, Overview of the Style Change Detection Task at PAN 2020, in: Working Notes of CLEF 2020 - Conference and Labs of the Evaluation Forum, volume 2696 of *CEUR Workshop Proceedings*, CEUR-WS.org, Thessaloniki, Greece, 2020, p. 11.

[5] E. Zangerle, M. Mayerl, M. Potthast, B. Stein, Overview of the Style Change Detection Task at PAN 2021, in: Working Notes of CLEF 2021 - Conference and Labs of the Evaluation Forum, volume 2936 of *CEUR Workshop Proceedings*, CEUR-WS.org, Bucharest, Romania, 2021, pp. 1760–1771.

[6] E. Zangerle, M. Mayerl, M. Potthast, B. Stein, Overview of the Style Change Detection Task at PAN 2022, in: Working Notes of CLEF 2022 - Conference and Labs of the Evaluation Forum, volume 3180 of *CEUR Workshop Proceedings*, CEUR-WS.org, Bologna, Italy, 2022, pp. 2344–2356.

[7] J. Bevendorff, D. Dementieva, M. Fröbe, B. Gipp, A. Greiner-Petter, J. Karlgren, M. Mayerl, P. Nakov, A. Panchenko, M. Potthast, A. Shelmanov, E. Stamatatos, B. Stein, Y. Wang, M. Wiegmann, E. Zangerle, Overview of PAN 2025: Voight-Kampff Generative AI Detection, Multilingual Text

Detoxification, Multi-Author Writing Style Analysis, and Generative Plagiarism Detection, in: J. C. de Albornoz, J. Gonzalo, L. Plaza, A. G. S. de Herrera, J. Mothe, F. Piroi, P. Rosso, D. Spina, G. Faggioli, N. Ferro (Eds.), Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the Sixteenth International Conference of the CLEF Association (CLEF 2025), Lecture Notes in Computer Science, Springer, Berlin Heidelberg New York, 2025.

[8] E. Zangerle, M. Mayerl, M. Potthast, B. Stein, Overview of the Multi-Author Writing Style Analysis Task at PAN 2025, in: G. Faggioli, N. Ferro, P. Rosso, D. Spina (Eds.), Working Notes of CLEF 2025 – Conference and Labs of the Evaluation Forum, CEUR Workshop Proceedings, CEUR-WS.org, 2025.

[9] F. Scarselli, M. Gori, Ah Chung Tsoi, M. Hagenbuchner, G. Monfardini, The Graph Neural Network Model, IEEE Transactions on Neural Networks 20 (2009) 61–80.

[10] D. Embarcadero-Ruiz, H. Gómez-Adorno, A. Embarcadero-Ruiz, G. Sierra, Graph-Based Siamese Network for Authorship Verification, in: Working Notes of CLEF 21- Conference and Labs of the Evaluation Forum, volume 2936 of *CEUR Workshop Proceedings*, CEUR-WS.org, Bucharest, Romania, 2021, p. 11.

[11] D. Embarcadero-Ruiz, H. Gómez-Adorno, A. Embarcadero-Ruiz, G. Sierra, Graph-Based Siamese Network for Authorship Verification, in: Working Notes of CLEF 22- Conference and Labs of the Evaluation Forum, volume 3180 of *CEUR Workshop Proceedings*, CEUR-WS.org, Bologna, Italy, 2022, p. 277.

[12] A. Valdez-Valenzuela, J. A. Martinez-Galicia, H. Gomez-Adorno, Heterogeneous-Graph Convolutional Network for Authorship Verification, in: Working Notes of CLEF 2023 - Conference and Labs of the Evaluation Forum, CEUR Workshop Proceedings, CEUR-WS.org, Thessaloniki, Greece, 2023, p. 8.

[13] Y. Zhang, P. Qi, C. D. Manning, Graph Convolution over Pruned Dependency Trees Improves Relation Extraction, in: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Brussels, Belgium, 2018, pp. 2205–2215.

[14] B. Yu, X. Mengge, Z. Zhang, T. Liu, W. Yubin, B. Wang, Learning to Prune Dependency Trees with Rethinking for Neural Relation Extraction, in: Proceedings of the 28th International Conference on Computational Linguistics, International Committee on Computational Linguistics, Barcelona, Spain (Online), 2020, pp. 3842–3852.

[15] B. Li, Y. Fan, Y. Sataer, Z. Gao, Y. Gui, Improving Semantic Dependency Parsing with Higher-Order Information Encoded by Graph Neural Networks, Applied Sciences 12 (2022) 16.

[16] Y. Hu, H. Shen, W. Liu, F. Min, X. Qiao, K. Jin, A Graph Convolutional Network With Multiple Dependency Representations for Relation Extraction, IEEE Access 9 (2021) 81575–81587.

[17] K. Sun, R. Zhang, Y. Mao, S. Mensah, X. Liu, Relation Extraction with Convolutional Network over Learnable Syntax-Transport Graph, Proceedings of the AAAI Conference on Artificial Intelligence 34 (2020) 8928–8935.

[18] L. Zhou, T. Wang, H. Qu, L. Huang, Y. Liu, A Weighted GCN with Logical Adjacency Matrix for Relation Extraction, Santiago de Compostela (2020) 8.

[19] A. Mandya, D. Bollegala, F. Coenen, Graph Convolution over Multiple Dependency Sub-graphs for Relation Extraction, in: Proceedings of the 28th International Conference on Computational Linguistics, International Committee on Computational Linguistics, Barcelona, Spain (Online), 2020, pp. 6424–6435.

[20] Z. Jin, Y. Yang, X. Qiu, Z. Zhang, Relation of the Relations: A New Paradigm of the Relation Extraction Problem, CoRR abs/2006.03719 (2020) 11.

[21] Y. Tian, G. Chen, Y. Song, X. Wan, Dependency-driven Relation Extraction with Attentive Graph Convolutional Networks, in: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Association for Computational Linguistics, Online, 2021, pp. 4458–4471.

[22] K. Zhao, H. Xu, Y. Cheng, X. Li, K. Gao, Representation Iterative Fusion Based on Heterogeneous

Graph Neural Network for Joint Entity and Relation Extraction, Knowledge-Based Systems 219 (2021) 9.
[23] P. Liu, L. Wang, Q. Zhao, H. Chen, Y. Feng, X. Lin, L. He, ECNU_ica_1 SemEval-2021 Task 4: Leveraging Knowledge-enhanced Graph Attention Networks for Reading Comprehension of Abstract Meaning, in: Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021), Association for Computational Linguistics, Online, 2021, pp. 183–188.
[24] Z. Guo, Y. Zhang, W. Lu, Attention Guided Graph Convolutional Networks for Relation Extraction, in: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Florence, Italy, 2019, pp. 241–251.
[25] Z. Li, Y. Sun, J. Zhu, S. Tang, C. Zhang, H. Ma, Improve Relation Extraction with Dual Attention-Guided Graph Convolutional Networks, Neural Computing and Applications 33 (2021) 1773–1784.
[26] T. N. Kipf, M. Welling, Semi-Supervised Classification with Graph Convolutional Networks, Technical Report arXiv:1609.02907, arXiv, 2017.
[27] J. Huertas-Tato, A. Martín, D. Camacho, Understanding writing style in social media with a supervised contrastively pre-trained transformer, Knowledge-Based Systems 296 (2024) 12.
[28] A. S. Alsheddi, M. E. B. Menai, Writing Style Change Detection: State of the Art, Challenges, and Research Opportunities, Artificial Intelligence Review (in press) (2025) 60.
[29] J. A. Khan, Style Breach Detection: An Unsupervised Detection Model, in: CLEF 2017 Labs and Workshops, Notebook Papers, volume 1866 of *CEUR Workshop Proceedings*, CEUR-WS.org, Dublin, Ireland, 2017, p. 10.
[30] J. A. Khan, A Model for Style Change Detection at a Glance, in: CLEF 2018 Labs and Workshops, Notebook Papers, volume 2125 of *CEUR Workshop Proceedings*, CEUR-WS.org, Avignon, France, 2018, p. 8.
[31] D. Karas, M. Spiewak, S. Piotr, OPI-JSA at CLEF 2017: Author Clustering and Style Breach Detection, in: CLEF 2017 Labs and Workshops, Notebook Papers, volume 1866 of *CEUR Workshop Proceedings*, CEUR-WS.org, Dublin, Ireland, 2017, p. 12.
[32] K. M. Ramachandran, C. P. Tsokos, Mathematical statistics with applications in R, third ed., Elsevier, Philadelphia, 2020.
[33] R. Singh, J. Weerasinghe, R. Greenstadt, Writing Style Change Detection on Multi-Author Documents, in: CLEF 2021 Labs and Workshops, Notebook Papers, volume 2936 of *CEUR Workshop Proceedings*, CEUR-WS.org, Bucharest, Romania, 2021, p. 9.
[34] F. Alvi, H. Algafri, N. Alqahtani, Style Change Detection using Discourse Markers, in: CLEF 2022 Labs and Workshops, Notebook Papers, volume 3180 of *CEUR Workshop Proceedings*, CEUR-WS.org, Bologna, Italy, 2022, p. 6.
[35] K. Sa□n, A. Ogaltsov, Detecting a Change of Style Using Text Statistics, in: CLEF 2018 Labs and Workshops, Notebook Papers, volume 2125 of *CEUR Workshop Proceedings*, CEUR-WS.org, Avignon, France, 2018, p. 6.
[36] S. Nath, Style Change Detection using Siamese Neural Networks, in: CLEF 2021 Labs and Workshops, Notebook Papers, volume 2936 of *CEUR Workshop Proceedings*, CEUR-WS.org, Bucharest, Romania, 2021, p. 11.
[37] M. Elamine, S. Mechti, L. H. Belguith, An Unsupervised Method for Detecting Style Breaches in a Document, in: 2019 IEEE/ACS 16th International Conference on Computer Systems and Applications (AICCSA), IEEE, Abu Dhabi, United Arab Emirates, 2019, pp. 1–6.
[38] S. Alshamasi, M. B. Menai, Ensemble-Based Clustering for Writing Style Change Detection in Multi-Authored Textual Documents, in: CLEF 2022 Labs and Workshops, Notebook Papers, volume 3180 of *CEUR Workshop Proceedings*, CEUR-WS.org, Bologna, Italy, 2022, p. 18.
[39] L. Mandic, F. Milkovic, S. Doria, Combining the Powers of Clustering Affinities in Style Change Detection, Course Project Reports, University of Zagreb, Zagreb, Croatia, 2019.
[40] C. Zuo, Y. Zhao, R. Banerjee, Style Change Detection with Feed-forward Neural Networks, in: CLEF 2019 Labs and Workshops, Notebook Papers, volume 2380 of *CEUR Workshop Proceedings*, CEUR-WS.org, Lugano, Switzerland, 2019, p. 9.
[41] N. Schaetti, Character-based Convolutional Neural Network for Style Change Detection, in:

CLEF 2018 Labs and Workshops, Notebook Papers, volume 2125 of *CEUR Workshop Proceedings*, CEUR-WS.org, Avignon, France, 2018, p. 6.

[42] M. Hosseinia, A. Mukherjee, A Parallel Hierarchical Attention Network for Style Change Detection, in: CLEF 2018 Labs and Workshops, Notebook Papers, volume 2125 of *CEUR Workshop Proceedings*, CEUR-WS.org, Avignon, France, 2018, p. 7.

[43] X. Jiang, H. Qi, Z. Zhang, M. Huang, Style Change Detection: Method Based On Pre-trained Model And Similarity Recognition, in: CLEF 2022 Labs and Workshops, Notebook Papers, volume 3180 of *CEUR Workshop Proceedings*, CEUR-WS.org, Bologna, Italy, 2022, p. 6.

[44] Q. Lao, L. Ma, W. Yang, Z. Yang, D. Yuan, Z. Tan, L. Liang, Style Change Detection Based On Bert And Conv1d, in: CLEF 2022 Labs and Workshops, Notebook Papers, volume 3180 of *CEUR Workshop Proceedings*, CEUR-WS.org, Bologna, Italy, 2022, p. 6.

[45] J. Zi, L. Zhou, Style Change Detection Based On Bi-LSTM And Bert, in: CLEF 2022 Labs and Workshops, Notebook Papers, volume 3180 of *CEUR Workshop Proceedings*, CEUR-WS.org, Bologna, Italy, 2022, p. 5.

[46] Z. Zhang, Z. Han, L. Kong, Style Change Detection based on Prompt, in: CLEF 2022 Labs and Workshops, Notebook Papers, volume 3180 of *CEUR Workshop Proceedings*, CEUR-WS.org, Bologna, Italy, 2022, p. 4.

[47] Z. Zhang, Z. Han, L. Kong, X. Miao, Z. Peng, J. Zeng, H. Cao, J. Zhang, Z. Xiao, X. Peng, Style Change Detection Based On Writing Style Similarity, in: CLEF 2021 Labs and Workshops, Notebook Papers, volume 2936 of *CEUR Workshop Proceedings*, CEUR-WS.org, Bucharest, Romania, 2021, p. 4.

[48] T.-M. Lin, C.-Y. Chen, Y.-W. Tzeng, L.-H. Lee, Ensemble Pre-trained Transformer Models for Writing Style Change Detection, in: CLEF 2022 Labs and Workshops, Notebook Papers, volume 3180 of *CEUR Workshop Proceedings*, CEUR-WS.org, Bologna, Italy, 2022, p. 9.

[49] X. Liu, H. Chen, J. Lv, Team foshan-university-of-guangdong at PAN: Adaptive Entropy-Based Stability-Plasticity for Multi-Author Writing Style Analysis, in: Working Notes of CLEF 2024 - Conference and Labs of the Evaluation Forum, volume 3740 of *CEUR Workshop Proceedings*, CEUR-WS.org, Grenoble, France, 2024, pp. 2750–2754.

[50] T. M. Mohan, T. V. S. Sheela, BERT-Based Similarity Measures Oriented Approach for Style Change Detection, in: Accelerating Discoveries in Data Science and Artificial Intelligence II, volume 438, Springer Nature Switzerland, Cham, 2024, pp. 83–94.

[51] T.-M. Lin, Y.-H. Wu, L.-H. Lee, Team NYCU-NLP at PAN 2024: Integrating Transformers with Similarity Adjustments for Multi-Author Writing Style Analysis, in: Working Notes of CLEF 2024 - Conference and Labs of the Evaluation Forum, volume 3740 of *CEUR Workshop Proceedings*, CEUR-WS.org, Grenoble, France, 2024, pp. 2716–2721.

[52] Y. Huang, L. Kong, Team Text Understanding and Analysis at PAN: Utilizing BERT Series Pre-training Model for Multi-Author Writing Style Analysis, in: Working Notes of CLEF 2024 - Conference and Labs of the Evaluation Forum, volume 3740 of *CEUR Workshop Proceedings*, CEUR-WS.org, Grenoble, France, 2024, pp. 2653–2657.

[53] Q. Wu, L. Kong, Z. Ye, Team bingezzzleep at PAN: A Writing Style Change Analysis Model Based on RoBERTa Encoding and Contrastive Learning for Multi-Author Writing Style Analysis, in: Working Notes of CLEF 2024 - Conference and Labs of the Evaluation Forum, volume 3740 of *CEUR Workshop Proceedings*, CEUR-WS.org, Grenoble, France, 2024, pp. 2963–2968.

[54] Z. Chen, Y. Han, Y. Yi, Team Chen at PAN: Integrating R-Drop and Pre-trained Language Model for Multi-author Writing Style Analysis, in: Working Notes of CLEF 2024 - Conference and Labs of the Evaluation Forum, volume 3740 of *CEUR Workshop Proceedings*, CEUR-WS.org, Grenoble, France, 2024, pp. 2547–2553.

[55] B. Wu, Y. Han, K. Yan, H. Qi, Team baker at PAN: Enhancing Writing Style Change Detection with Virtual Softmax, in: Working Notes of CLEF 2024 - Conference and Labs of the Evaluation Forum, volume 3740 of *CEUR Workshop Proceedings*, CEUR-WS.org, Grenoble, France, 2024, pp. 2951–2955.

[56] M. K. Sheykhlan, S. K. Abdoljabbar, M. N. Mahmoudabad, Team karami-sh at PAN: Transformer-based Ensemble Learning for Multi-Author Writing Style Analysis, in: Working Notes of CLEF

2024 - Conference and Labs of the Evaluation Forum, volume 3740 of *CEUR Workshop Proceedings*, CEUR-WS.org, Grenoble, France, 2024, pp. 2676–2681.

[57] C. Liu, Z. Han, H. Chen, Q. Hu, Team Liuc0757 at PAN: A Writing Style Embedding Method Based on Contrastive Learning for Multi-Author Writing Style Analysis, in: Working Notes of CLEF 2024 - Conference and Labs of the Evaluation Forum, volume 3740 of *CEUR Workshop Proceedings*, CEUR-WS.org, Grenoble, France, 2024, pp. 2716–2721.

[58] M. T. Zamir, M. A. Ayub, A. Gul, N. Ahmad, K. Ahmad, Stylometry Analysis of Multi-authored Documents for Authorship and Author Style Change Detection, 2024.

[59] M. Fröbe, M. Wiegmann, N. Kolyada, B. Grahm, T. Elstner, F. Loebe, M. Hagen, B. Stein, M. Potthast, Continuous Integration for Reproducible Shared Tasks with TIRA.io, in: Advances in Information Retrieval. 45th European Conference on IR Research (ECIR 2023), Lecture Notes in Computer Science, Springer, Berlin Heidelberg New York, 2023, pp. 236–241.

[60] Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, J. M. Solomon, Dynamic Graph CNN for Learning on Point Clouds, ACM Transactions on Graphics 38 (2019) 1–12.

[61] A. S. Alsheddi, M. E. B. Menai, Edge Convolutional Networks for Style Change Detection in Arabic Multi-Authored Text, Applied Sciences 15 (2025) 6633. URL: https://www.mdpi.com/2076-3417/15/12/6633. doi:10.3390/app15126633.