基於寫作風格的圖神經網路假新聞偵測模型
(A Fake News Detection Model Utilizing Graph Neural Networks to Capture Writing Styles)

Yen-Tsang Wu **Lawrence Y. H. Low** **Jenq-Haur Wang**

Department of Computer Science and Information Engineering
National Taipei University of Technology
Web Information Retrieval Lab
t107599005@ntut.edu.tw t113999402@ntut.org.tw jhwang@ntut.edu.tw

摘要

本文提出 CWSMN（Capture Writing Style Multi-Graph Network），一個以圖神經網路為基礎的早期假新聞偵測方法，透過捕捉寫作風格克服傳統語意內容與傳播特徵方法在標註稀缺與跨域泛化不足下的限制。CWSMN 結合文體分析、語意嵌入與多圖融合：以 Bi-GRU 進行上下文初始化，採用 GAT 進行注意力導向的圖聚合，並以 LDA 建構主題圖，最終以輕量級前饋分類器輸出。於多個資料集之實驗顯示，CWSMN 對比 BERT、ALBERT 與 GraphSAINT 等強基準皆有優勢；在未知來源的 Source-CV 場景尤為顯著，證明其於低資源與跨領域環境之穩健泛化能力，並實現不依賴傳播的早期偵測，實驗結果證實本方法在樣本稀缺與未知來源條件下，仍能達成有效的早期偵測。

Abstract

We present CWSMN, a graph neural network for early fake-news detection that foregrounds writing style to address the fragility of purely semantic- or propagation-based approaches under label scarcity and domain shift. CWSMN fuses stylistic cues with semantics through a multi-graph design: Bi-GRU initializes contextual token representations; GAT performs attention-driven aggregation over style- and relation-aware graphs; LDA induces a topic graph, and a lightweight feed-forward head produces predictions. Across multiple datasets, CWSMN consistently surpasses strong baselines (BERT, ALBERT, GraphSAINT), with the largest margins under source-level cross-validation (Source-CV) on unseen sources. These results demonstrate robust generalization in low-resource, cross-domain scenarios and support propagation-agnostic early decisions, underscoring practical value for timely mitigation across platforms and domains.

關鍵字：假新聞偵測、早期偵測、圖神經網路、寫作風格、多圖融合

Keywords: Fake news detection, Graph neural networks, Multi-graph fusion, Early detection, Writing style

1 Introduction

隨著社群媒體與線上新聞平台的蓬勃發展，資訊的傳播速度與規模遠超過以往。然而，這樣的資訊環境同時為假新聞的快速擴散提供了溫床。假新聞所引發的負面影響不僅包括公共輿論的操縱與社會信任的侵蝕，更可能導致經濟損失與公共安全危機(Yang & Pan, 2021)。在近年的重大事件中，如疫情與選舉期間，假新聞所造成的廣泛誤導與社會混亂，進一步凸顯了有效偵測與早期抑制其傳播的迫切需求(Shahid et al., 2022)。傳統的假新聞偵測方法主要可分為兩類：其一為**語意內容導向方法**(Przybyla, 2020)，透過自然語言處理技術抽取文本特徵以進行分類；其二為**傳播結構導向方法**(Cheng et al., 2024)，利用社群媒體互動與訊息擴散模式來辨識真假資訊。雖然這些方法已取得一定成效，但仍面臨幾項挑戰：（1）跨領域的新聞資料存在語言風格與主題差異，使得預訓練語言模型難以有效泛化(Abdali & Krishnamachari, 2022)；（2）標註資料的取得成本高昂且數量有限，導致模型訓練受到限制(Deng & Wang, 2022; Gao et al., 2023)；（3）依賴傳播路徑的模型需等待樣

Proceedings of the 37th Conference on Computational Linguistics and Speech Processing (ROCLING 2025), pages 286–295
November 20-22, 2025

本累積，無法滿足假新聞「早期偵測」的需求(Deng & Wang, 2022; Shahid et al., 2022)。值得注意的是，雖然假新聞的內容會因領域或事件而有所不同，其**寫作風格**卻往往具有一致性，例如情緒化的詞彙使用、誇張的語氣與特定的句法結構。因此，若能設計能夠有效捕捉寫作風格的模型，即可避免對傳播資訊的依賴，並在訊息發佈初期即進行偵測，達到更即時且跨領域的假新聞辨識效果。基於此，本研究提出一種**多圖寫作風格捕捉網路（Capture Writing Style Multi-Graph Network, CWSMN）**，其核心概念為透過圖神經網路對文本進行多層次風格特徵萃取與融合，進而訓練分類器進行判斷。本研究的主要貢獻可歸納如下：1. **提出寫作風格驅動的假新聞偵測框架**：不同於依賴內容或傳播的傳統方法，本研究專注於捕捉風格特徵，以提升早期偵測能力。2. **設計多圖嵌入與融合策略**：結合 GRU、GAT 與 LDA 生成多種異質子圖，並提出文件層級與詞層級的圖融合方法，有效整合多樣化的風格訊號。3. **驗證跨領域泛化能力**：實驗結果顯示，CWSMN 在未見過來源中顯著優於現有方法，展現了強大的跨領域遷移能力。綜上所述，本研究以寫作風格為核心設計假新聞偵測模型，不僅提升了模型在資料稀缺與跨領域場景下的效能，更具備實務應用價值，能有效支援即時的假新聞擴散。

2 Literature Review

本章回顧與本研究直接相關之理論與方法，包括 Graph Neural Networks 於假新聞偵測的應用、多模態與混合式偵測路徑、寫作風格與文字特徵的假新聞偵測。近年來，GNNs 憑藉其建模結構關係與多模態的能力，逐步成為假新聞偵測的重要技術路線；其中圖注意力網路（Graph Attention Network, GAT）(Veličković et al., 2017)、GraphSAINT(Zeng et al., 2019) 與主題模型（LDA）(Blei et al., 2003)等模型，提供了從詞彙、主題、文件多層次的特徵提取與融合訊的方法，亦為以寫作風格為核心的早期偵測框架奠定基礎。

2.1 圖神經網路於假新聞偵測之基礎與應用

圖神經網路透過訊息傳遞聚合鄰域特徵，能以結構化的方式同時處理文本、主題、來源等異質關係。其中， GAT 以可學習的注意力權重為不同鄰接節點賦予差異化重要度，提升跨語言與結構之表徵能力；結合專為社群和內容而設計的資料集與建模框架，能在訓練資料與測試來源不一致時維持穩健性。其中，多圖架構與注意力機制雖能提升特徵擷取能力，但亦帶來訓練與計算成本。而圖採樣與叢集切分等方法(Chen et al., 2018; Chiang et al., 2019; Dhawan et al., 2024; Goldani et al., 2021; Golovin et al., 2025)使模型在維持辨識能力的同時兼顧記憶體占用與延遲，更能貼近早期偵測與實務部署(Alghamdi et al., 2024; Chang et al., 2024; Phan et al., 2023; Y. Zhang et al., 2024)。

2.2 多模態與混合式假新聞偵測方法

多模態模型能整合文本、視覺與傳播路徑的特徵；混合式方法則結合語意表徵與結構表徵。深度語言模型如 BERT 或 ALBERT 提供強語意表示，但在跨主題或跨來源時，常面臨域偏移問題；與此同時，圖取樣與子圖學習可在大規模圖上有效訓練並保留關鍵關係訊號；兩者互補可提升跨域泛化(Alghamdi et al., 2022; Galli et al., 2022; Mahmoudi et al., 2024; H. Zhang et al., 2024).

2.3 基於寫作風格與文字特徵的假新聞偵測

相較於主題內容，寫作風格在不同領域間更具穩定性，如情緒化詞彙、誇飾語氣、句法節奏與詞彙多樣性。近年工作結合 CoreNLP 的 POS/NER 與 General Inquirer（GI）之心理語意類別以形成細粒度風格訊號，並在圖上建模不同的方式上，以增進來源未知情境下的辨識力（Horne & Adalı, 2018; Potthast et al., 2018; Lima et al., 2020; Stone et al., 1966; Gehrmann et al., 2019）。而寫作風格導向研究顯示，僅依賴語意內容或傳播路徑的偵測器對新興來源與主題漂移較為敏感；反之，將文體作為重要特徵並結合主題資訊與圖結構，可以在跨來源情境取得更穩健表現；相關資料集與基準研究提供來源多樣性與可比性脈絡（Przybyła, 2020; Shu et al., 2018; Galli et al., 2022）。主題建模（如 LDA）能提供文件層級潛在語意結構，適合用於建立 Topic node 與主題感知劃分策略；在模型評估上，除 Doc-CV 外，Topic-CV 與 Source-CV 更能反映真實部署環境，其中 Source-CV 模擬未知來源的泛化能力，常見語意導向模型在此退化，而風格導向加圖融合有較好的效能(Masciari et al., 2020; Nadeem et al.,

2023; Nan et al., 2024; Sharma et al., 2023; Tsai, 2023; Wu et al., 2024; Yang et al., 2024)

3 Methodology

為了解決傳統假新聞偵測方法過度依賴人工特徵與語意預訓練模型的限制，我們提出一個名為「基於捕捉寫作風格多圖神經網路的假新聞偵測（CWSMN）」的新穎框架。該模型是基於圖神經網路對於寫作風格的捕捉，旨在將常用的寫作特徵進行提取與融合，應用於假新聞偵測。我們的方法著重於跨領域寫作風格的一致性運用，使得在無需等待使用者傳播模式累積資料的情況下，即可進行早期偵測。CWSMN 包含四個主要階段：風格分析、嵌入、跨模態圖融合與分類。詳細架構如圖 1 所示。

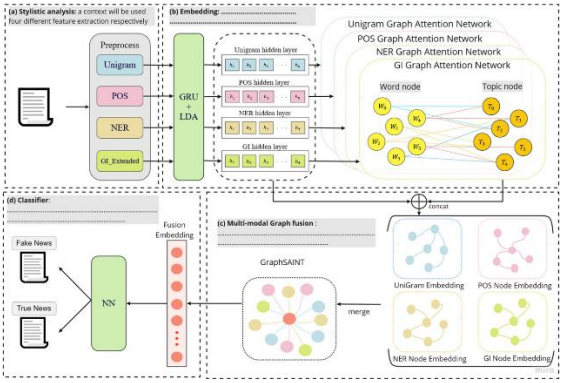

圖 1: CWSMN 架構圖

3.1 Stylistic analysis

在寫作風格分析的整體架構方面，我們採用了一系列文體特徵，進而使用圖注意力機制捕捉寫作風格。在本研究中，我們採用 NER(Named Entity Recognition), POS tagging, and Unigram 以及 General Inquirer dictionaries 作為文件特徵。為了避免分類器過於擬合於特定來源或主題的特徵，在 *N*-gram 的特徵擷取中，我們僅使用詞性單詞（unigrams），不使用雙詞（bigrams）和三詞（trigrams）。我們首先使用 Stanford CoreNLP(Manning et al., 2014) 對輸入文件進行預處理，處理步驟包括 tokenization、NER 以及 POS tagging。除此之外，我們也使用 General Inquirer dictionaries(Stone et al., 1962) 進行正規化並且將所有 word 歸入 182 類別。General Inquirer dictionaries 是一種辭典工具，經常被用於極端黨派新聞識別(Potthast et al., 2017)，字典一共包含 8640 個詞彙。有別於先前的研究 (Przybyla, 2020)，我們並未使用詞向量例如 Word2Vec 來擴充 GI 詞典，因為這類擴充可能會引入語意上的雜訊。

3.2 Embedding and Multi-Graph Construction

我們的詞嵌入過程整合了 Bi-GRU、GAT 和 LDA，以提取上下文表示和多關係圖結構。嵌入的過程主要分為兩個步驟：嵌入初始化和多圖構建。**Step 1. Embedding Initialization:** 由於 Stylistic analysis 例如 *unigram* 以及 NER 之後的 token，我們將其視為一個新單詞，透過使用雙向 Bi-GRU 根據前後文的關係計算出每個 word vector 作為新 node 的 initialization vector. 這樣的好處是可以賦予每個 node 更多前後文的關係，捕捉更多寫作風格上的特徵。**Step 2: Multi-Graph Construction:** 為了捕捉多樣化的寫作風格關係，我們建構四種以 GAT 為主的圖結構：

- Topic-based Graphs：先以 LDA 模型為每一個 token 計算主題分布，據此為各主題建立主題節點（topic nodes），並將每個 token 與其對應的主題節點連結。藉由這些主題節點，相同主題的 tokens 得以間接關聯。對於同一份文件，我們依不同建構觀點可得到三張圖，分別為 Unigram 圖、NER 圖與 POS 圖。
- GI-based Graphs：依據 GI 字典所定義的語義類別，建立 182 個類別節點；凡屬於同一 GI 類別的詞彙，皆透過對應的類別節點彼此連結，從而強化風格與語用傾向的特徵。

為了捕捉節點之間不同程度的關聯性，我們使用 GAT 來捕捉寫作風格中潛藏的語意與結構。GAT 在進行圖中節點資訊聚合時，引入了注意力機制，使模型能夠根據鄰近節點的重要性，動態學習節點之間的重要性權重。透過注意力機制，模型能自動關注於對風格判別較具代表性的詞彙與其關聯，有效強化跨主題、跨領域的風格一致性建模能力。因此，我們提出的方法透過 Bi-GRU 捕捉到前後文序列的向量，採用 LDA model 與 GI 產生不同結構的 graph，有效強化跨主題、跨領域的風格一致性建模能力。

3.3 Multi-graph Fusion

為了整合來自不同圖結構的資訊，我們採用了兩種融合策略。

3.3.1 Token-level Fusion

為建構全域層級（global graph）的語意關係圖，我們將每個 token 視為節點，並為每份 document 建立一個 document node；該文件內之所有 token-nodes 與其 document node 連邊。考量圖結構規模龐大，訓練成本較高，我們引入 GraphSAINT 演算法進行子圖抽樣，以提升模型的訓練效率與可擴展性。 GraphSAINT 透過基於節點、邊或隨機游走的子圖抽樣技術，僅從原始大圖中選取部分子圖進行訓練，既保留了全圖的結構資訊，又有效降低了運算成本。如下圖所示。

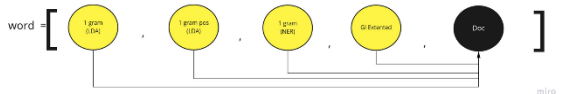

圖 2: Token-level 融合示意圖。

GraphSAINT 是一種針對大規模圖神經網路訓練所提出的高效子圖抽樣方法。GraphSAINT 透過基於節點、邊或隨機游走的子圖抽樣技術，僅從原始大圖中選取部分子圖進行訓練，既保留了全圖的結構資訊，又有效降低了運算成本。與其他 mini-batch 訓練方法相比，GraphSAINT 在保留圖結構統計特性方面表現更佳，並能在不犧牲精度的前提下，大幅提升訓練效率。因此，由於實驗的 tokens 的數量十分龐大，因此我們使用 GraphSAINT 作為我們全域層級的融合模型。經由 GraphSAINT 後，我們可以獲得每一個 document 的 embedding。

3.3.2 Document-level Fusion

另外一種融合方式為 Document level fusion。我們分成 2 個步驟。

- Embedding aggregation：對四張子圖的詞嵌入採平均池化以得其子圖表示（subgraph representation）。
- GAT-based fusion: 使用 GAT 模型進行 fusion。 我們創立一個 document node，然後 document node 與四個子圖 node 連接，進行運算以融合這些資訊。GAT 通過注意力機制動態計算文件節點與各子圖表示節點之間的權重（注意力分數），以決定每個子圖表示對最終文件表示的貢獻。最終更新文件節點的表示，融合來自四個子圖的特徵，形成局部層級(Local Graph)的文件表示。如下所示:

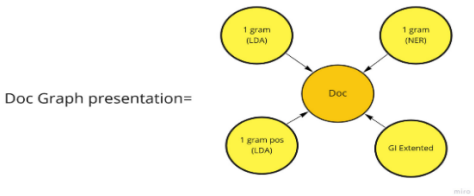

圖 3: Document-level 融合示意圖。

由於 document-level fusion 只使用 5 個 node。因此，我們使用 GAT 模型，通過注意力機制動態地為圖中的節點分配不同的權重，從而捕捉節點間的關係和重要性。最後，我們獲得每一個 document node 的 embedding 作為融合後 document 的特徵向量。

3.4 Classification

通過圖融合（graph fusion，使用 GAT 融合四個子圖的表示，生成反映局部層級的文件向量）後，將得到的文件向量輸入到一個簡單的前饋神經網路（feed-forward neural network, FNN）分類器，進行二元分類，判斷新聞是真（real）還是假（fake）。Pseudo-code 如下:

Algorithm 1 Training CWSMN

Require: corpus $\mathcal{D}$, graphs $\{G^{(g)}\}$, parameters θ
Ensure: trained parameters θ

$\mathcal{D} \leftarrow \text{PREPROCESSWITHCORENLP}(\mathcal{D})$ ▷ tokenize, GI/POS/NER; LDA with K=100
$E_{\text{word}} \leftarrow \text{INITWORDEMBEDDINGSWITHGRU}(\mathcal{D})$; $E_{\text{nonword}} \leftarrow \text{INITLEARNABLENONWORDVECTORS}()$
$\{G^{(g)}\} \leftarrow \text{BUILDGRAPHS}(\mathcal{D}, \{\text{TW}, \text{CO}, \text{SP}, \text{EN}, \text{GI}, \text{DW}\})$
for epoch = 1 **to** E **do** ▷ Word-level path on DW with GRAPHSAINT
  for sampled $G_{\text{sub}} \subset G^{(\text{DW})}$ via GRAPHSAINT **do**
    $H^{(\text{DW})} \leftarrow \text{GNN\_FORWARD}(G_{\text{sub}})$
    $\tilde{z}_d^{(w)} \leftarrow \text{READOUT}(H^{(\text{DW})}; \text{center} = d)$
  end for
  ▷ Graph-level path on $\{\text{TW}, \text{CO}, \text{SP}, \text{EN}, \text{GI}\}$
  for $g \in \{\text{TW}, \text{CO}, \text{SP}, \text{EN}, \text{GI}\}$ **do**
    $H^{(g)} \leftarrow \text{GAT\_2LAYERS}(G^{(g)})$
    $\tilde{z}_d^{(g)} \leftarrow \text{ATTENTIONAVG}(\{h_w^{(g)} \mid w \in W(d)\})$
  end for
  $\{z_d^{(g)}\}_g \leftarrow \text{SOFTMAXGATE}(\{\tilde{z}_d^{(g)}\}_g)$ ▷ gated cross-graph fusion
  $z_d \leftarrow \text{PROJ}\left(\text{CONCAT}\left(z_d^{(w)}, z_d^{(g)}\right)\right)$
  $\hat{p} \leftarrow \text{SOFTMAX}(\text{FFN}(z_d))$
  $L \leftarrow L_{\text{cls}}(\hat{p}, y_d) + \alpha L_{\text{align}} + \lambda L_{\text{sparse}}$
  $\theta \leftarrow \text{ADAMW\_UPDATE}(\theta, \nabla_\theta L)$
end for

4 Experiment

4.1 Dataset description

我們使用一個公開的 dataset 來評估我們提出的 CWSMN 模型。

4.1.1 資料特徵

實驗用的資料集來自(Przybyla, 2020)從網路爬取的一個語料庫，內容包含52,790 篇的假新聞與 50,429 篇的真新聞，總計 103,219 篇文本文件、超過 1.17 億個詞元（tokens），一共 205 個不可信網站與 18 個可信網站作為資料來源。為提升語料品質，我們排除了重複內容與那些平均每行少於 15 個詞彙的文本，以確保文本具有足夠語意密度。我們觀察資料集特性

可以發現，假新聞來源較常涉及的主題包括穆斯林與移民、健康與營養、總統競選對手比較等；而真新聞則較常涵蓋的主題則為電影與體育等。

4.1.2 資料集劃分方式

為了評估模型在不同情境下的表現，實驗設計了三種場景進行交叉驗證 Document-based CV，Topic-based CV 和 Source-based CV。

- Document-based CV: 將所有 103,219 篇文件隨機分成 5 個 fold，確保每個 fold 包含來自不同來源和主題的文件。模擬測試文件來自已知來源和已知主題的情況，作為基準評估。
- Topic-based CV: 使用潛在狄利克雷分配生成 100-topic LDA 模型，並將每篇文件分配到與其關聯最強的主題。將這些主題隨機分成 5 個 fold，確保每個 fold 包含與特定主題相關的所有文件。目的是為了模擬真實世界裡文件屬於訓練資料中未見過的主題的情況，測試模型對新主題的泛化能力。這種劃分方式確保訓練和測試資料的主題分離，防止模型依賴特定主題的詞彙或內容進行分類。
- Source-based CV: 將所有 223 個來源分成 5 個 fold，確保每個折包含來自特定來源的所有文件。每個 fold 的測試集包含來自訓練資料中未見過的來源的文件。目的是模擬真實世界假新聞來自全新來源（例如新興新聞網站）的情況，測試模型對新來源的泛化能力。

實驗的資料集劃分策略旨在模擬現實世界中假新聞檢測的挑戰，特別是新主題和新來源的情境。通過隨機文件劃分、主題分離和來源分離，實驗確保模型不僅在已知資料上表現良好，還能泛化到未知情境。

4.2 Implementation Details

GAT 模型架構所有參數都遵循 Graph Attention Networks 論文裡的設定，Number of Attention Heads 為 8. Number of Layers 為 2. 激活函數分別為 LeakyReLU；輸出層使用 softmax。GraphSAINT 根據相同的文件路徑構建圖結構，每個 node 的維度設為 100。Activation function 採用 LeakyReLU。使用 Adam 作為優化器。在圖抽樣過程中，我們使用隨機游走抽樣（random walk sampling）方法。分類損失使用 cross-entropy loss. 為了捕捉不同新聞來源間的主題差異，我們利用潛在狄利克雷分配訓練一個包含 100 個主題的模型。此後，將每個詞彙根據其最大主題關聯度，指派至對應的主題。

4.3 Comparison of Methods

我們將本模型與當前最先進的方法進行比較以評估其效能，並同時測試多個模型變體：

- BERT(Devlin et al., 2019): BERT 是一種基於 Transformer 的雙向語言模型，廣泛用於自然語言處理任務。受限於最大序列長度，在本研究中，僅使用文件的前 512 個 token。
- ALBERT(Lan et al., 2019): 是一種相較於 BERT 參數量更少的模型，並且專注於句子間的連貫性。通過參數縮減技術和自監督學習目標，顯著降低模型大小，同時保持高性能。兩者皆屬於預訓練的 Transformer 模型，透過擷取文字語意來進行分類。
- Word2Vec +GraphSAINT: 透過 Word2Vec 計算出 word 的 embedding 後，相加取平均作為 Document 的向量。 然後依據傳播路徑建圖進行分類。
- ALBERT+GraphSAINT: 先透過 ALBERT 計算出所有 token 的 embedding 後相加取平均作為 Document 的向量。然後依據傳播路徑建圖進行分類。
- GI+GraphSAINT: 僅使用 GI 字典裡的字向量相加取平均作為 Document 的向量.然後依據傳播路徑建圖進行分類。 以上三種方法都是根據傳播路徑進行假新聞偵測，屬於圖神經網路。
- Stylometric: Stylometric 分類器是一種基於文體特徵的模型，專注於捕捉文本的寫作風格，避免依賴特定來源或主題特徵。該方法通過提取文體相關特徵並使用線性模型進行分類。
- BiLSTMAvg: 是一種基於雙向長短期記憶網絡的神經網絡模型，通過對文件中所有句子的表示進行平均，生成文件級別的預測。該模型旨在捕捉句子的上下文和文體特徵，用於假新聞檢測的文體分析。

- Bag-of-Words: 是一種簡單的基線分類器，基於詞彙頻率表示文件，用於假新聞檢測。以上三種方法都是假新聞偵測常用的方法，採用傳統文字語意作為特徵-包含統計詞彙頻率與文體分析。

4.4 Evaluation metric

由於真和假新聞的數量大致平衡，因此我們實驗使用準確率（accuracy）作為評估指標。此外，為了確保每個來源和主題的文件都被用於測試，我們採用 5-fold cross-validation。

4.5 Experimental Results

本節呈現了我們在假新聞檢測任務中的實驗結果，比較了多種模型在三種交叉驗證情境下的表現：文件級交叉驗證、主題級交叉驗證和來源級交叉驗證。實驗結果如下表:

Model	Doc-CV	Topic-CV	Source-CV	Average
CWSMN_GAT	0.9870	0.9760	0.9784	0.9805
CWSMN_GraphSAINT	0.9930	0.9910	0.9746	0.9862
ALBERT	0.9815	0.9754	0.7165	0.8911
ALBERT+GraphSAINT	0.9985	0.9961	0.7385	0.9110
Word2Vec+GraphSAINT	0.9993	0.9853	0.9095	0.9647
GI+GraphSAINT	0.9983	0.9995	0.7096	0.9025
Stylometric	0.9274	0.9173	0.8097	0.8848
BiLSTMAvg	0.8994	0.8921	0.8250	0.8722
Bag-of-Words	0.9913	0.9886	0.7078	0.8959
BERT	0.9976	0.9965	0.7960	0.9300

表 1：各模型在三種場景下的準確率。

Document CV scenario: Document cv 模擬已知來源和主題的場景，測試的文件來自已知來源和主題。結果顯示：Word2Vec+GraphSAINT 表現最佳，達到 0.9993 的準確率，這表示複合類型的假新聞偵測模型-結合語意和傳播路徑有最強的效能。而預訓練模型 BERT 和 ALBERT+GraphSAINT、GI+GraphSAINT 達到 0.9993 準確率，優於其餘方法，顯示在已知資料上的優勢。而我們提出的模型效能雖然輸給最好的模型，分別為 0.0115 跟 0.0063，但是也能高達 0.987 與 0.993，遠勝過於傳統的方法。傳統的方法中 Stylometric 和 BiLSTMAvg 表現相對較弱，可能是因為文體特徵提取的簡單性限制了其在已知資料上的表現。**Topic CV scenario:** Topic-CV 模擬新事件場景，測試文件來自未見過的主題（基於 LDA 分配的 100 個主題）。結果顯示：GI+GraphSAINT 達到最高準確率 0.9995，顯示其對新主題的強適應性，歸功於 GI 詞典的情感特徵和 GraphSAINT 的結構化建模。我們提出的模型分別達到 0.9910 與 0.9760 略輸最好的模型 0.0085 跟 0.0235。值得一提的是 Bag-of-Words (0.9886) 僅下降 0.27%（相較於 Doc-CV 的 0.9913），顯示詞彙的頻率對主題變化的有效性。而傳統的方法表現略低，表明文體特徵對主題變化的適應能力較為不足。

Model	Doc-CV	Topic-CV	Source-CV	Average
BERT	0	0	0	0
ALBERT fine-tune	-0.0161	-0.0211	-0.0795	-0.0389
CWSMN_GAT	-0.0106	-0.0205	0.1824	0.0504
CWSMN_GraphSAINT	-0.0046	-0.0055	0.1786	0.0562

Source CV scenario: Source-CV 模擬新興假新聞網站場景，測試文件來自未見過的來源，是最嚴苛的泛化測試。結果顯示：我們所提出的兩個模型表現最佳，分別為 0.9784 跟 0.9746，這顯示結合寫作風格特徵和圖神經網絡的方法具有卓越泛化能力。在這個 scenario 裡，我們的方法領先了基於傳播路徑與語意的多模態模型 Word2Vec+GraphSAINT (0.9095)分別為 0.0689 與 0.0651.而領先 pretrained Transformer models such as BERT (0.7960)、ALBERT (0.7165)分別達到 0.2619 與 0.1824.這證明了比起大成本的預訓練模型，我們的方法能夠以較小的計算量達到更好的效果。此外，我們觀察到，BiLSTMAvg (0.8250) 和 Stylometric (0.8097) 在文體特徵模型中表現最佳，這也更加證明，使用以寫作風格作為特徵的方法，更符合真實世界中，對新興假新聞的有效性。

平均準確率: 平均準確率綜合評估模型在三種情境下的整體性能。我們提出的兩種方法都達到最佳的效果(0.9862 與 0.9805)，顯示其在不同 scenario 下的均衡表現，擁有優異的性能。而多模態模型 Word2Vec+GraphSAINT (0.9647) 表現出色，僅輸給我們的模型 0.0751 pretrained Transformer models 分別落後我們 0.056 與 0.095，受 Source-CV 準確率較低的影響。整體而言，實驗結果顯示我們所提出 CWSMN 模型在未曾學習過的資料上情境下展現出優異的假新聞偵測能力。，在已有資料可以學習的情境下，也有不俗的效能，相比於預訓練模型或是圖神經網路，我們的平均準確率最高，最高可達 0.9862，進一步證明了本模型的強大效能。

5 Evaluation and Analysis

5.1 與預訓練的深度語意模型的比較

表 2：不同預訓練的深度語意模型的比較結果

BERT 與 ALBERT fine-tune 都是使用雙向 Transformer 編碼器的預訓練模型，我們 fine-tune 之後進行預測，模型性能以 BERT 為基準，差值表示相對於 BERT 的準確率變化。Doc-CV 場景下，ALBERT fine-tune 相較於 BERT 下降 0.01606，可能是由於參數縮減（12M-235M 相較於 BERT 的 110M-340M）影響了模型的效能。CWSMN_GAT 則下降 0.0106，表明基於特徵組合的模型在已知資料上與 BERT 接近，但效能略遜。CWSMN_GraphSAINT 下降最少，顯示 GraphSAINT 的採樣技術增強了特徵表達能力，接近 BERT 的性能。Topic-CV 場景下，ALBERT fine-tune 相較於 BERT 下降 0.0211，表明其對新主題的適應性略弱，可能是 SOP 目標未完全補償 NSP 的影響。CWSMN_GAT 則下降 0.0205，與 ALBERT 接近，CWSMN_GraphSAINT 下降最少，顯示 GraphSAINT 的結構化建模有效提升了對新主題的泛化能力。Source-CV 模擬新興假新聞網站場景，測試資料來自未見過的來源，這在早期偵測中，是最重要的一點。ALBERT fine-tune 相較於 BERT 下降 0.0795，而我們提出的 CWSMN_GAT 和 CWSMN_GraphSAINT 相較於 BERT 分別提升 0.1824 跟 0.1786，表明基於寫作風格在捕捉文體特徵方面優於預訓練模型。BERT 和 ALBERT fine-tune 作為基準模型，在已知樣本中效能優秀，但在 Source-CV 由於未曾學習相關的深度語意，因此限制了泛化能力，準確率只有 0.796。而我們提出的方法，主要依靠寫作風格作為特徵，不受未知主題與語意的限制，對新來源的顯著提升，顯示結合單詞、命名實體、詞性標記和 GI 詞典的情感特徵能有效捕捉文體模式，超越預訓練的深度語意模型，這證明了對於現今不斷新增類型與主題的動態假新聞偵測上的重要性。

5.2 Comparison of Graph Models with Different Edge Construction Methods

Model	Doc-CV	Topic-CV	Source-CV	Average
Word2Vec_GraphSAINT	0	0	0	0
ALBERT+GraphSAINT	-0.0008	0.0108	-0.171	-0.0537
GI_GraphSAINT	-0.001	0.0142	-0.1999	-0.0622
CWSMN_GAT	-0.0123	-0.0093	0.0689	0.0158
CWSMN_GraphSAINT	-0.0063	0.0057	0.0651	0.0215

表 3：不同建邊策略之比較

本節比較了採用不同建邊方法的圖模型（Graph Models）在假新聞檢測任務中的性能，特別聚焦於不同特徵的比較。表 3 列出了五種圖神經網路模型在三種場景下的準確率變化。在 document-CV 與 Topic-CV 下，我們的方法的準確率僅些微落後於最強基準；然而在最具挑戰性的 Source-CV 情境中，本方法取得最佳表現。此結果顯示：當測試分布出現來源轉換時，「以寫作風格為特徵並透過多圖融合」的設計能帶來更好的跨來源泛化能力。同樣基於圖神經網路，但 GraphSAINT 的建邊機制核心在於：若兩個文件來自同一網站，則在圖上形成較強連結。此設計等同於將來源特徵視為主要特徵。然在現實中，虛假資訊網站往往快速出現又迅速消失，且有大量短命或一次性域名；因此，即時偵測時常無法依賴「過去的資料」來推斷真偽。結果是：在 Source-CV 情境下，過度依賴來源同質性的建邊策略容易失效，並導致對新來源的遷移能力不足。我們亦嘗試將來源關係納入本模型（將來源感知的建邊與多圖風格圖結合，類似於與 GraphSAINT 的混合）。實驗顯示，整體效能反而下降。可能原因包括：**overfitting**：風格圖已提供可區辨的文體訊號；再疊加來源同質性，容易學到「來源＝標籤」的虛假關聯，降低對新來源的魯棒性。**訊號冗餘與過度同質（over-smoothing）**：來源邊會把同站點文件過度拉近，使得節點表示在圖上趨於同質，削弱內容/風格細節。**分布偏移（distribution shift）**：訓練時的來源分布與測試差異過大，來源邊成為不穩定的遷移支點。

5.3 與 Baseline 的比較

Model	Doc-CV	Topic-CV	Source-CV
Stylometric	0	0	0
BiLSTMAvg	-0.028	-0.0252	0.0153
Bag-of-Words	0.0639	0.0713	-0.1019
BERT	0.0702	0.0792	-0.0137
CWSMN_GAT	0.0596	0.0587	0.1687
CWSMN_GraphSAINT	0.0656	0.0737	0.1649

表 4：Baseline 的比較結果

首先，相較於傳統文本方法（如 BoW、Stylometric、統計式分類器等），本研究方法在三種場景設定上皆取得全面領先，顯示以寫作風格為特徵的多圖融合能夠在已知與未知分

布下同時維持高準確率與穩健性。其次，相較於同為神經網路的基準模型，本研究方法仍呈現顯著優勢。這證明：單純提升網路性能不足以克服跨來源的分布轉移；反之，結合多視角建邊與跨子圖注意力融合，更能提取對假新聞具有域不變性的風格線索。第三，就詞袋模型而言，本研究方法在 document-CV 與 Topic-CV 皆呈小幅領先，而在 source-CV 上的優勢更為明顯，領先幅度達 0.2706。此一差距反映：當測試來源為未曾出現時，依賴來源或主題近鄰的建邊策略容易失效；相對地，以文體／語用為核心的多圖表示能更有效對抗來源漂移。第四，與目前最具代表性的深度語意模型 BERT 相比，雖然本研究方法在 Document-CV 與 Topic-CV 略遜一籌，但在 Source-CV 卻展現顯著領先，優勢高達 0.1824。此結果凸顯：僅依賴語意預訓練的表徵在面對新來源時較易退化；相較之下，風格驅動＋多圖融合提供了更具泛化性的決策訊號。

5.4 Ablation Study

我們以下列四類特徵分別建圖並以 GAT 進行聚合：Unigram， NER， POS， GI，並且測試在上述四特徵融合下，再疊加 GraphSAINT 的來源建邊之影響。其結果如下表。

Model	Doc-CV	Topic-CV	Source-CV
GI	0.9728	0.9671	0.7179
POS+GI	0.9733	0.966	0.7449
NER+GI	0.9726	0.9645	0.7317
Unigram+NER	0.9864	0.98	0.9456
Unigram+GI+POS	0.9862	0.986	0.9624
Unigram+NER+POS	0.9872	0.984	0.958
Unigram+NER+POS+GI (Average)	0.987	0.976	0.9784
Unigram+NER+POS+GI+GraphSAINT	0.993	0.991	0.9746

表 5： Ablation Study 結果。

僅使用 GI 建圖能在 Doc-CV 與 Topic-CV 達到合理表現，但在 Source-CV 顯著落後，說明僅靠心理語意類別不足以支撐面對新來源的泛化。而 POS+GI 與 NER+GI 在三種 CV 均優於單獨 GI，代表結合句法或實體脈絡能補足 GI 的粒度；但在 Source-CV 的提升仍有限，顯示僅雙特徵尚不足以克服來源漂移。再加入 Unigram 的三種特徵組合後在三種 CV 均有明顯增益，尤其 Source-CV 提升最為突出，顯示 Unigram 對未知來源的區辨尤為關鍵。將 4 種特徵以平均方式融合，在 Doc-CV 與 Topic-CV 維持高準確率，同時在 Source-CV 取得全表最佳或近最佳；此設定提供了「已知分布準確」與「未知來源泛化」間的最優特徵。但是，值得注意的，4 種特徵融合基礎上加入來源同站建邊後，Doc-CV 與 Topic-CV 小幅上升，但 Source-CV 略為降低，這證明來源邊容易讓模型學到來源＝標籤的誤差，對未見來源產生過擬合。由結果顯示，在三種 CV 下，四訊號融合取得最佳整體表現，特別是在 Source-CV 顯示最強泛化；而在此基礎上再加入來源建邊，雖能微幅提升 Doc/Topic-CV，卻不利於 Source-CV，印證來源特徵對未知來源的過擬合風險。

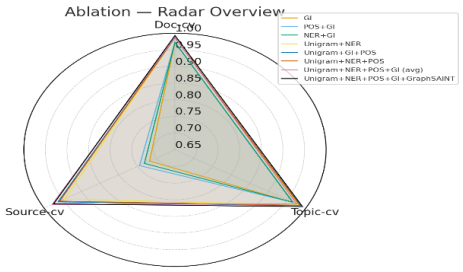

圖 4: Ablation 效果總覽

6 Conclusion

本研究提出 CWSMN 模型，以寫作風格作為偵測核心，透過寫作風格等多視角建邊，採用兩種融合策略：Token-level（以 GraphSAINT 子圖抽樣訓練）與 Document-level（以 GAT 進行跨子圖注意力融合），在三種場景進行測試與消融分析，得到以下結論：1.跨來源泛化的優勢：在 Source-CV 的嚴苛情境，CWSMN 顯著優於各組對照，對目前最強深度語意模型亦具明確領先，驗證「風格驅動結合多圖融合」能有效對抗來源漂移並支撐早期偵測。2. 已知分布下的穩健性：在 Doc 以及 Topic-CV 中，CWSMN 與最佳比較模型效能相當或僅小幅落後；顯示在不犧牲已知分布準確的情況下，仍能換取對未知來源的泛化能力。3.早期偵測的可能性: 由於不依賴傳播軌跡，因此可以達到早期假新聞偵測的目的。

Acknowledgments

作者感謝國家科學及技術委員會（NSTC）之經費支持（計畫編號：NSTC 114-2221-E-027-068、NSTC 114-2634-F-027-001-MBK），使本研究得以順利進行。亦感謝匿名審查者提供的中肯建議，對於論文品質之提升助益良多；同時感謝參與資料蒐集、系統建置與實驗驗證

之人員。本文內容僅代表作者個人觀點與責任，與補助機關立場無涉。

References

Abdali, S., & Krishnamachari, B. (2022). Multi-modal misinformation detection: Approaches, challenges and opportunities. *arXiv preprint arXiv:2203.13883*.

Alghamdi, J., Lin, Y., & Luo, S. (2022). Modeling fake news detection using bert-cnn-bilstm architecture. 2022 IEEE 5th international conference on multimedia information processing and retrieval (MIPR),

Alghamdi, J., Luo, S., & Lin, Y. (2024). A comprehensive survey on machine learning approaches for fake news detection. *Multimedia Tools and Applications*, *83*(17), 51009-51067.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, *3*(Jan), 993-1022.

Chang, W., Liu, K., Yu, P. S., & Yu, J. (2024). Enhancing Fairness in Unsupervised Graph Anomaly Detection through Disentanglement. *arXiv preprint arXiv:2406.00987*.

Chen, J., Ma, T., & Xiao, C. (2018). Fastgcn: fast learning with graph convolutional networks via importance sampling. *arXiv preprint arXiv:1801.10247*.

Cheng, L.-C., Wu, Y. T., Chao, C.-T., & Wang, J.-H. (2024). Detecting fake reviewers from the social context with a graph neural network method. *Decision Support Systems*, *179*, 114150.

Chiang, W.-L., Liu, X., Si, S., Li, Y., Bengio, S., & Hsieh, C.-J. (2019). Cluster-gcn: An efficient algorithm for training deep and large graph convolutional networks. Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining,

Deng, Y., & Wang, S.-W. (2022). Detecting Fake News on Social Media by CSIBERT. Proceedings of the 2022 6th International Conference on Deep Learning Technologies,

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. 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),

Dhawan, M., Sharma, S., Kadam, A., Sharma, R., & Kumaraguru, P. (2024). Game-on: Graph attention network based multimodal fusion for fake news detection. *Social Network Analysis and Mining*, *14*(1), 114.

Galli, A., Masciari, E., Moscato, V., & Sperlí, G. (2022). A comprehensive Benchmark for fake news detection. *Journal of Intelligent Information Systems*, *59*(1), 237-261.

Gao, W., Ni, M., Deng, H., Zhu, X., Zeng, P., & Hu, X. (2023). Few-shot fake news detection via prompt-based tuning. *Journal of Intelligent & Fuzzy Systems*, *44*(6), 9933-9942.

Goldani, M. H., Momtazi, S., & Safabakhsh, R. (2021). Detecting fake news with capsule neural networks. *Applied Soft Computing*, *101*, 106991.

Golovin, A., Zhukova, N., Delhibabu, R., & Subbotin, A. (2025). Improving Recommender Systems for Fake News Detection in Social Networks with Knowledge Graphs and Graph Attention Networks. *Mathematics*, *13*(6), 1011.

Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., & Soricut, R. (2019). Albert: A lite bert for self-supervised learning of language representations. *arXiv preprint arXiv:1909.11942*.

Mahmoudi, G., Behkamkia, B., & Eetemadi, S. (2024). Zero-Shot Stance Detection using Contextual Data Generation with LLMs. *arXiv preprint arXiv:2405.11637*.

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J. R., Bethard, S., & McClosky, D. (2014). The Stanford CoreNLP natural language processing toolkit. Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations,

Masciari, E., Moscato, V., Picariello, A., & Sperli, G. (2020). A deep learning approach to fake news detection. International Symposium on Methodologies for Intelligent Systems,

Nadeem, M. I., Ahmed, K., Zheng, Z., Li, D., Assam, M., Ghadi, Y. Y., Alghamedy, F. H., & Eldin, E. T. (2023). SSM: Stylometric and semantic similarity oriented multimodal fake news detection.

Journal of King Saud University-Computer and Information Sciences, *35*(5), 101559.

Nan, Q., Sheng, Q., Cao, J., Hu, B., Wang, D., & Li, J. (2024). Let Silence Speak: Enhancing Fake News Detection with Generated Comments from Large Language Models. *arXiv preprint arXiv:2405.16631*.

Phan, H. T., Nguyen, N. T., & Hwang, D. (2023). Fake news detection: A survey of graph neural network methods. *Applied Soft Computing*, *139*, 110235.

Potthast, M., Kiesel, J., Reinartz, K., Bevendorff, J., & Stein, B. (2017). A stylometric inquiry into hyperpartisan and fake news. *arXiv preprint arXiv:1702.05638*.

Przybyla, P. (2020). Capturing the style of fake news. Proceedings of the AAAI conference on artificial intelligence,

Shahid, W., Jamshidi, B., Hakak, S., Isah, H., Khan, W. Z., Khan, M. K., & Choo, K.-K. R. (2022). Detecting and mitigating the dissemination of fake news: Challenges and future research opportunities. *IEEE Transactions on Computational Social Systems*.

Sharma, A., Sharma, M., & Dwivedi, R. K. (2023). Exploratory data analysis and deception detection in news articles on social media using machine learning classifiers. *Ain Shams Engineering Journal*, *14*(10), 102166.

Stone, P. J., Bales, R. F., Namenwirth, J. Z., & Ogilvie, D. M. (1962). The general inquirer: A computer system for content analysis and retrieval based on the sentence as a unit of information. *Behavioral Science*, *7*(4), 484.

Tsai, C.-M. (2023). Stylometric fake news detection based on natural language processing using named entity recognition: In-domain and cross-domain analysis. *Electronics*, *12*(17), 3676.

Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., & Bengio, Y. (2017). Graph attention networks. *arXiv preprint arXiv:1710.10903*.

Wu, J., Guo, J., & Hooi, B. (2024). Fake news in sheep's clothing: Robust fake news detection against LLM-empowered style attacks. Proceedings of the 30th ACM SIGKDD conference on knowledge discovery and data mining,

Yang, H.-C., Hung, Y.-L., & Wang, L.-C. (2024). Stylometry-based Fake News Classification Using Text Mining Techniques. Proceedings of the 2024 11th Multidisciplinary International Social Networks Conference,

Yang, J., & Pan, Y. (2021). COVID-19 Rumor Detection on Social Networks Based on Content Information and User Response. *Frontiers in Physics*, *9*, 763081.

Zeng, H., Zhou, H., Srivastava, A., Kannan, R., & Prasanna, V. (2019). Graphsaint: Graph sampling based inductive learning method. *arXiv preprint arXiv:1907.04931*.

Zhang, H., Liu, X., Yang, Q., Yang, Y., Qi, F., Qian, S., & Xu, C. (2024). T3RD: Test-Time Training for Rumor Detection on Social Media. Proceedings of the ACM on Web Conference 2024,

Zhang, Y., Ma, X., Wu, J., Yang, J., & Fan, H. (2024). Heterogeneous Subgraph Transformer for Fake News Detection. Proceedings of the ACM on Web Conference 2024,