TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM
TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG
KHOA CÔNG NGHỆ THÔNG TIN

THÔNG QUỐC HƯNG - 52000763
TRỊNH DUY KHOA - 52000772

A STUDY ON GRAPH NEURAL NETWORK APPLIED TO THE FAKE NEWS DETECTION

DỰ ÁN CÔNG NGHỆ THÔNG TIN

KHOA HỌC MÁY TÍNH

THÀNH PHỐ HỒ CHÍ MINH, NĂM 2024

TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM
TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG
KHOA CÔNG NGHỆ THÔNG TIN

THÔNG QUỐC HƯNG - 52000763
TRỊNH DUY KHOA - 52000772

A STUDY ON GRAPH NEURAL NETWORK APPLIED TO THE FAKE NEWS DETECTION

DỰ ÁN CÔNG NGHỆ THÔNG TIN

KHOA HỌC MÁY TÍNH

Người hướng dẫn
ThS. Dung Cẩm Quang

THÀNH PHỐ HỒ CHÍ MINH, NĂM 2024

LỜI CẢM ƠN

Chúng tôi xin gửi lời cảm ơn chân thành nhất đến ThS. Dung Cẩm Quang về sự hướng dẫn tận tâm trong suốt quá trình làm bài báo cáo này. Báo cáo này không chỉ là kết quả nổ lực của nhóm chúng tôi mà còn là sự hướng dẫn, hỗ trợ tận tình từ ThS. Dung Cẩm Quang. Trong suốt quá trình làm bài chúng tôi nhận thấy bài làm còn nhiều chỗ thiếu sót, mong thầy/cô có thể bỏ qua cũng như góp ý để chúng tôi có thể hoàn thành tốt hơn ở những bài báo cáo sau. Chúng tôi xin chân thành cảm ơn!

TP. Hồ Chí Minh, ngày 23 tháng 02 năm 2024.

Tác giả

(Ký tên và ghi rõ họ tên)

CÔNG TRÌNH ĐƯỢC HOÀN THÀNH
TẠI TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG

Tôi xin cam đoan đây là công trình nghiên cứu của riêng tôi và được sự hướng dẫn khoa học của ThS. Dung Cẩm Quang Các nội dung nghiên cứu, kết quả trong đề tài này là trung thực và chưa công bố dưới bất kỳ hình thức nào trước đây. Những số liệu trong các bảng biểu phục vụ cho việc phân tích, nhận xét, đánh giá được chính tác giả thu thập từ các nguồn khác nhau có ghi rõ trong phần tài liệu tham khảo.

Ngoài ra, trong Dự án còn sử dụng một số nhận xét, đánh giá cũng như số liệu của các tác giả khác, cơ quan tổ chức khác đều có trích dẫn và chú thích nguồn gốc.

Nếu phát hiện có bất kỳ sự gian lận nào tôi xin hoàn toàn chịu trách nhiệm về nội dung Dự án của mình. Trường Đại học Tôn Đức Thắng không liên quan đến những vi phạm tác quyền, bản quyền do tôi gây ra trong quá trình thực hiện (nếu có).

TP. Hồ Chí Minh, ngày 23 tháng 02 năm 2024

Tác giả

(Ký tên và ghi rõ họ tên)

TÌM HIỂU GRAPH NEURAL NETWORK VÀ ÁP DỤNG VÀO BÀI TOÁN PHÁT HIỆN TIN GIẢ

TÓM TẮT

Tin xuyên tạc và tin giả đã gây ra những tác động bất lợi cho cá nhân và xã hội trong những năm gần đây, thu hút sự chú ý rộng rãi đến phát hiện tin tức giả mạo. Do đó, các bài toán phương pháp tự động phát triển tin giả đang rất được quan tâm hiện nay. Trong bài báo cáo này, chúng tôi sẽ tập trung tìm hiểu bài toán phát hiện tin giả và phân tích một số bài báo về lĩnh vực này. Sau đó, tiến hành tìm hiểu 2 tập dữ liệu Politifact và Gossipcop để áp dụng vào bài toán phát hiện tin giả. Chúng tôi sẽ tìm hiểu các mô hình có sẵn áp dụng cho bài toán Fake News Detection. Sau đó, thiết kế và hiện thực lại các mô hình dựa trên kiến thức đã nghiên cứu. Dựa trên kết quả hiện thực mô hình sẽ tiến hành phân tích và đánh giá một cách chi tiết, bao gồm việc so sánh với các phương pháp hiện có và phân tích các điểm mạnh, điểm yếu của mô hình. Trong những phương pháp mà chúng tôi đã nghiên cứu trong bài báo này thì phương pháp tận dụng siêu đồ thị (hypergraph) để thể hiện sự tương tác giữa các tin tức theo nhóm đang thể hiện kết quả rất tốt trên cả 2 bộ dữ liệu. Điều đó thấy rằng phương pháp tiếp cận này mang lại hiệu suất vượt trội và nó vẫn duy trì kết quả tốt ngay cả với một tập dữ liệu tin tức nhỏ được gắn nhãn.

A STUDY ON GRAPH NEURAL NETWORK APPLIED TO THE FAKE NEWS DETECTION

ABSTRACT

Transparent news and fake news have caused adverse personal and societal benefits in recent years, attracting widespread attention to fake news detection. Therefore, problems and methods for automatically developing fake news are of great interest today. In this report, we will focus on understanding the problem of fake detection and analyze some articles in this field. Then, proceed to explore two data sets Politifact and Gossipcop to apply to the problem detection hypothesis. We will learn about available models applied to the problem of detecting fake news. Then, design and reimplement the models based on the researched knowledge. Based on the results of model implementation, detailed analysis and evaluation will be conducted, including comparison of existing methods and analysis of the strengths and weaknesses of the model. Among the methods we have researched in this article, the method that takes advantage of hypergraphs to show interactions between news in groups shows very good results on both datasets. It is shown that this approach provides outstanding performance and it maintains good results even with a small subset of labeled news data.

MỤC LỤC

DANH MỤC HÌNH VẼ

DANH MỤC BẢNG BIỂU

DANH MỤC CÁC CHỮ VIẾT TẮT

CNN	Convolutional Neural Networks
FND	Fake News Detection
GCN	Graph Convolutional Network
GNN	Graph Neural Network
HGFND	Hyper Graph for Fake News Detection

CHƯƠNG 1. MỞ ĐẦU VÀ TỔNG QUAN ĐỀ TÀI

1.1 Lý do chọn đề tài

Internet và các nền tảng truyền thông xã hội đã trở thành trung tâm không thể thiếu để mọi người cập nhật tin tức hàng ngày. Nhờ vào tính nhanh chóng và sự tự do trong truyền thông, mọi người có thể dễ dàng tiếp cận và tìm hiểu ý kiến mọi lúc, mọi nơi. Theo một nghiên cứu đến tháng 8 năm 2018, hơn 68% người Mỹ lấy tin từ mạng xã hội. Tuy nhiên, việc thiếu sự kiểm soát đã làm giảm chất lượng tin tức so với các phương tiện truyền thông truyền thống. Internet đang trở thành một môi trường nhiễu loạn, nơi thông tin sai lệch và tin tức giả lan tràn.

Tin tức giả là những thông tin được tạo ra với mục đích đánh lừa, gây hậu quả tiêu cực đối với cả cá nhân và cộng đồng. Những thông tin này thường mang tính chất thiên vị hoặc không chính xác, ảnh hưởng nghiêm trọng đến ý kiến công cộng và ổn định xã hội. Ví dụ, trong đại dịch COVID-19, sự pha trộn giữa thông tin thật và tin giả đã tạo ra một làn sóng thông tin không chính xác, được Tổ chức Y tế Thế giới gọi là "đại dịch thông tin". Trong ba tháng đầu năm 2020, khoảng 6000 người đã nhập viện vì thông tin sai lệch về virus corona, và ít nhất 800 người có thể đã chết do những thông tin không chính xác liên quan đến COVID-19. Có thể thấy, tin tức sai lệch đang gây ra những hậu quả vô cùng nghiêm trọng. Do đó, vấn đề phát hiện tin giả một cách tự động đang rất được quan tâm.

1.2 Mục tiêu thực hiện đề tài

Mục tiêu đề tài là nghiên cứu và phát triển một phương pháp phát hiện tin giả dựa vào mạng nơ-ron đồ thị (GNN) và mạng nơ-ron đồ thị tích chập (GCN). Nghiên cứu tập trung vào việc áp dụng các kỹ thuật học máy và phân tích đồ thị để xây dựng một hệ thống thông minh có khả năng phát hiện và phân loại các tin tức giả một cách hiệu quả, từ đó cung cấp công cụ hữu ích trong việc chống lại thông tin sai lệch trên mạng.

Nghiên cứu tập trung vào những mô hình, bộ dữ liệu dùng để giải quyết bài toán Fake News Detection đã có sẵn. Tiếp theo đó, tiến hành thực nghiệm lại các mô

hình. Sau đó, so sánh các mô hình dựa trên kiến trúc, kết quả đạt được hiểu rõ điểm mạnh, điểm yếu của từng mô hình áp dụng và rút ra kết luận.

1.3 Phương pháp nghiên cứu

Nghiên cứu về Mạng nơ-ron đồ thị (GNN) và Mạng nơ-ron đồ thị tích chập (GCN). Trong giai đoạn này, sẽ thực hiện tìm hiểu sâu về cơ sở lý thuyết và cách thức hoạt động của GNN và GCN. Sau đó tìm hiểu các mô hình có sẵn áp dụng cho bài toán Fake News Detection. Tìm hiểu 2 tập dữ liệu Politifact và Gossipcop. Thiết kế mô hình: dựa trên kiến thức đã nghiên cứu và các mô hình có sẵn để hiện thực lại các mô hình. Huấn luyện và đánh giá mô hình: mô hình được thiết kế sẽ được huấn luyện trên dữ liệu huấn luyện và đánh giá thông qua các phương pháp đánh giá hiệu suất như accuracy và F1-score. Các thử nghiệm sẽ được thực hiện để đánh giá tính hiệu quả và độ tin cậy của mô hình.

Phân tích và đánh giá kết quả: Kết quả của mô hình sẽ được phân tích và đánh giá một cách chi tiết, bao gồm việc so sánh với các phương pháp hiện có và phân tích các điểm mạnh, điểm yếu của mô hình.

Cuối cùng, sẽ tổng kết lại kết quả của nghiên cứu, đưa ra những nhận xét và đề xuất về hướng phát triển tiếp theo.

1.4 Ý nghĩa khoa học

Nâng cao hiểu biết về phương pháp phân tích đồ thị trong phát hiện tin giả: báo cáo sẽ đóng góp vào việc tăng cường kiến thức về cách mà mạng nơ-ron đồ thị (GNN) và mạng nơ-ron đồ thị tích chập (GCN) có thể được áp dụng vào việc phát hiện tin giả. Bằng cách này, nó mở ra cánh cửa cho những phương pháp mới trong lĩnh vực này.

Cung cấp giải pháp mới cho vấn đề ngày càng nghiêm trọng của tin giả: với sự lan truyền nhanh chóng của thông tin trên mạng Internet, việc phát hiện tin giả trở thành một thách thức lớn đối với cộng đồng. Bài báo cáo này cung cấp một giải pháp mới, có tiềm năng để giúp làm giảm tác động của thông tin sai lệch và tin giả đối với cộng đồng trực tuyến.

Khuyến khích sự phát triển và ứng dụng của học máy trong lĩnh vực xã hội và nhân văn: Việc áp dụng các kỹ thuật học máy và phân tích đồ thị vào việc giải quyết vấn đề xã hội như phát hiện tin giả không chỉ mở ra một lĩnh vực nghiên cứu mới mà còn thúc đẩy sự phát triển của các ứng dụng thực tế có ích cho cộng đồng.

Tiềm năng ứng dụng trong các lĩnh vực khác: phương pháp phát hiện tin giả dựa vào GNN và GCN có thể được áp dụng không chỉ trong lĩnh vực truyền thông mà còn trong các lĩnh vực khác như y tế, tài chính, và giáo dục để phát hiện và ngăn chặn các hình thức lừa đảo và thông tin sai lệch.

1.5 Ý nghĩa thực tiễn

Chống lại thông tin sai lệch trên mạng: với sự bùng nổ của thông tin trên mạng, việc phát hiện và ngăn chặn tin tức giả trở thành một ưu tiên quan trọng. Phương pháp dựa vào GNN và GCN cung cấp công cụ mạnh mẽ để tự động phát hiện các tin giả, từ đó giúp làm sạch không gian thông tin trực tuyến.

Bảo vệ người dùng trực tuyến: cung cấp cho người dùng các công cụ và giải pháp để tự bảo vệ mình trước các thông tin giả mạo và thông tin sai lệch trên mạng xã hội và các nền tảng truyền thông.

Tăng cường đáng tin cậy trong nguồn tin: phát triển các phương pháp phát hiện tin giả giúp tăng cường sự đáng tin cậy của nguồn tin trên mạng, giúp người đọc có thể tin cậy hơn vào các thông tin mà họ tiếp nhận.

Hỗ trợ cho các cơ quan quản lý và truyền thông: cung cấp cho các cơ quan quản lý và truyền thông công cụ để nhanh chóng và hiệu quả xác minh tính chân thực của các thông tin trước khi lan truyền, giúp giảm thiểu tác động tiêu cực từ thông tin sai lệch.

Đóng góp vào nghiên cứu và phát triển công nghệ: việc áp dụng GNN và GCN vào phát hiện tin giả mở ra cánh cửa cho nghiên cứu và phát triển công nghệ trong lĩnh vực trí tuệ nhân tạo và học máy, đồng thời tạo ra những cơ hội mới để khám phá và ứng dụng trong các lĩnh vực khác của xã hội và kinh tế.

CHƯƠNG 2. CƠ SỞ LÝ THUYẾT

2.1 Tìm hiểu về bài toán phát hiện tin giả

2.1.1 Giới thiệu vấn đề

Trong thời đại bùng nổ thông tin, cũng như sự phát triển của mạng xã hội như hiện nay thì việc gặp phải các tin giả tràn lan là điều không thể tránh khỏi. Thậm chí nó đang được nhìn nhận như một trong những mối đe dọa lớn nhất đến sự tiếp cận, nền kinh tế tri thức và tranh luận tự do.

Trong thời gian gần đây, trên internet, đặc biệt là trên các mạng xã hội, đã xuất hiện một số tài khoản giả mạo, chia sẻ thông tin không được kiểm chứng liên quan đến nhiều lĩnh vực như chính trị, y tế, thời tiết, tin mê tín, quảng cáo không trung thực... Hiện tượng này đã gây ra sự lo lắng, làm rối tung và tác động lớn đến cuộc sống hàng ngày của cộng đồng.

Thuật ngữ "tin giả" là một khái niệm khá mới mẻ và cho đến nay vẫn chưa có một định nghĩa chính xác và phổ quát về nó. Theo Oxford thì "Tin giả là thông tin sai sự thật được phát sóng hoặc xuất bản dưới dạng tin tức nhằm mục đích lừa đảo hoặc có động cơ chính trị. Tin giả tạo ra sự nhầm lẫn đáng kể của công chúng về các sự kiện hiện tại. Tin giả bùng nổ trên phương tiện truyền thông xã hội, đang xâm nhập vào các kênh truyền thông chính".

Tin giả, hoặc tin tức giả mạo, đã trở thành một vấn đề nghiêm trọng đối với xã hội hiện đại. Tác hại của tin giả không chỉ làm mất lòng tin của công chúng vào các phương tiện truyền thông và nguồn thông tin trực tuyến, mà còn gây ra những hậu quả nghiêm trọng đối với sức khỏe, an ninh và mối quan hệ xã hội.

Một trong những tác hại chính của tin giả là tạo ra sự hoang mang và lo ngại trong cộng đồng. Khi thông tin không chính xác lan truyền nhanh chóng qua mạng xã hội và các phương tiện truyền thông, người dân dễ bị hoảng loạn và cảm thấy không biết tin ai. Điều này có thể dẫn đến sự phân hóa và xung đột trong xã hội, khi mà mỗi người tin vào một nguồn thông tin khác nhau. Ngoài ra, tin giả cũng có thể gây ra những hậu quả tiêu cực đối với sức khỏe cộng đồng. Ví dụ, trong mùa dịch

COVID-19, thông tin sai lệch về phòng ngừa và điều trị có thể khiến người dân bỏ qua các biện pháp an toàn, gây ra sự lan truyền nhanh chóng của virus và tăng nguy cơ lây nhiễm.

Hơn nữa, tin giả còn có thể được sử dụng như một công cụ để tạo ra sự phân biệt và kích động, thúc đẩy mối quan hệ xã hội trở nên căng thẳng và bất ổn. Bằng cách lan truyền thông tin sai lệch và kích động cảm xúc, các nhóm có chủ đích có thể tạo ra một môi trường căng thẳng, làm suy yếu mối giao tiếp và tin tưởng giữa các cộng đồng.

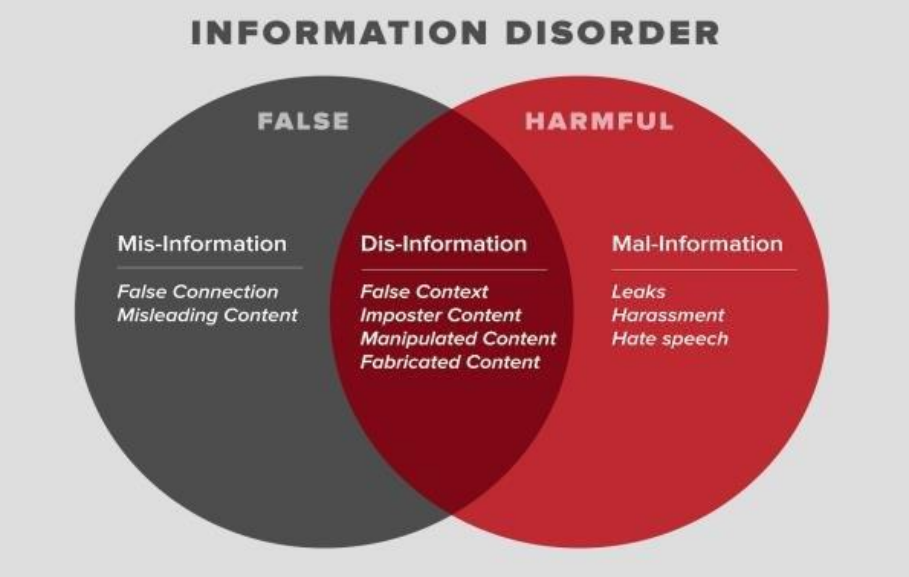
Trong tổng thể, tác hại của tin giả đối với xã hội là rất lớn nhưng lại chưa có một biện pháp nhận biết tin giả nào một cách chính xác. Tin giả có thể được nhận biết bằng đánh giá chủ quan của người xem tin tức nhưng với số lượng vô cùng lớn các tin tức giả hiện nay thì việc nhận biết tin tức giả dựa vào con người là không khả thi. Do đó, các phương pháp nhận biết tin tức giả tự động dựa vào các phương pháp Machine Learning hay Deep Learning đang được phát triển một cách rộng rãi để kịp thời giải quyết các vấn đề về tin giả của xã hội.

2.1.2 Phân loại tin giả

Theo một bài báo nghiên cứu về tin giả (Wardle & Derakhshan, 2017) thì tin giả được chia làm 3 nhóm chính đó là: thông tin sai lệch (Mis-information), thông tin giả mạo (Dis-information) và thông tin độc hại (Mal-information).

- Thông tin sai lệch (Mis-information): đây là loại thông tin không chính xác hoặc thiếu thông tin, thường được phổ biến mà không có ý định gây tổn hại. Dù không có ý đồ tiêu cực, thông tin sai lệch có thể dẫn đến sự hiểu lầm và nhầm lẫn trong cộng đồng.
- Thông tin giả mạo (Dis-information): đây là thông tin được tạo ra và lan truyền bởi những cá nhân hoặc tổ chức với ý định gây hại. Thông tin giả mạo thường nhằm mục đích lừa đảo, xuyên tạc sự thật hoặc tạo ra sự hỗn loạn trong cộng đồng.
- Thông tin độc hại (Mal-information): đây là loại thông tin chính xác nhưng được chia sẻ với mục đích gây hại. Thông tin độc hại thường bao

gồm việc phổ biến tin tức nhạy cảm, cá nhân hoá, hoặc sử dụng thông tin để đe dọa, xúc phạm hoặc tấn công người khác.



Hình 2.1 Phân loại tin giả
(Nguồn: (Wardle & Derakhshan, 2017))

a. Thông tin sai lệch (Mis-information)

Kết nối sai (False connection): đây là trường hợp khi hình ảnh, tiêu đề không phù hợp với nội dung chính. Ví dụ, các tiêu đề gây chú ý nhưng không phản ánh đúng nội dung bài viết, hoặc việc sử dụng hình ảnh không liên quan để thu hút sự chú ý của người đọc.

Nội dung gây hiểu lầm (Misleading content): đây là trường hợp khi thông tin được sử dụng một cách không chính xác và tạo ra sự hiểu lầm cho đọc giả. Ví dụ, các trang web quảng cáo có thể lừa dối người xem để truy cập vào các trang web không uy tín.

b. Thông tin giả mạo (Dis-information)

Bối cảnh sai (False context): khác với các loại thông tin khác, loại này là thông tin có thật nhưng được biến tấu với mục đích gây ra hậu quả nguy hiểm cho người đọc.

Nội dung mạo danh (Imposter content): đây là những thông tin giả mạo sự thật hoặc gây hiểu lầm bằng cách sử dụng danh tính hay tin tức của các cá nhân, tổ chức

nổi tiếng. Điều này tận dụng sự tin cậy của những cá nhân hoặc tổ chức này để lan truyền thông tin không đúng.

Nội dung bị thao túng (Manipulated content): loại này xảy ra khi một phần nào đó của nội dung gốc bị chỉnh sửa hoặc thay đổi, thường là qua các hình ảnh hoặc video. Ví dụ, việc chỉnh sửa hình ảnh để tạo ra sự hiểu lầm về một vụ scandal của một người nổi tiếng.

Nội dung bịa đặt (Fabricated content): đây là loại thông tin không chứa một chút nào thông tin chính xác.

c. Thông tin độc hại (Mal-information)

Rò rỉ (Leaks): đây là hành động tiết lộ thông tin mà không có sự cho phép từ những người hoặc tổ chức có thẩm quyền. Ví dụ, trong các cuộc bầu cử tổng thống tại Hoa Kỳ hoặc trước các cuộc họp quan trọng của Đảng ở Việt Nam, thông tin được cho là bị rò rỉ từ các tài liệu mật thường gây ra sự hoang mang và tạo ra nhiều ý kiến trái chiều trong cộng đồng.

Quấy rối (Harassment): đây là hành động nhằm vào việc làm tổn thương hoặc xúc phạm cá nhân hoặc tổ chức, thông qua lời nói, hình ảnh, văn bản hoặc các phương tiện khác. Trên mạng xã hội, các hành vi quấy rối ngày càng trở nên phổ biến và tinh vi hơn, ví dụ như các trang fanpage của những người nổi tiếng sử dụng thông tin để đánh bóng hình ảnh của họ và làm hạ thấp đối thủ.

Gây chia rẻ, thù hận (Hate speech): đây là các nội dung biểu hiện sự kỳ thị, châm biếm đối với một cá nhân hoặc một nhóm xã hội dựa trên các đặc điểm như chủng tộc, dân tộc, giới tính, tôn giáo, tuổi tác, khuyết tật về thể chất hoặc tinh thần.

2.1.3 Các phương pháp phát hiện tin giả

Theo một nghiên cứu về các phương pháp phát hiện tin giả (Fallis, 2015) theo cách thủ công thì có các cách sau:

1. Kiểm tra nguồn tin: xác minh địa chỉ website của trang hoặc nguồn phát tán các nội dung. Các trang web tin tức không uy tín thường có URL chứa lỗi chính tả hoặc có phần mở rộng tên miền lạ.

2. Kiểm tra tác giả: tìm hiểu về tác giả để xem họ có đáng tin hay không và xem xét động cơ của họ trong việc viết nội dung.
3. Kiểm tra các nguồn khác: so sánh thông tin với các cơ quan truyền thông uy tín và kiểm tra nguồn tài nguyên được trích dẫn trong tin tức.
4. Duy trì tư duy phản biện: hỏi về mục đích của câu chuyện và duy trì sự cảnh giác với những nội dung kích động cảm xúc mạnh mẽ.
5. Kiểm tra sự thật: kiểm tra các dữ kiện, thống kê, và nguồn tin được trích dẫn trong bài báo và xác minh tính logic của nội dung.
6. Kiểm tra nhận xét của người dùng: đọc các nhận xét dưới bài báo để xem liệu có thông tin phản hồi hợp lý không.
7. Kiểm tra thành kiến cá nhân: không để thành kiến chi phối lý trí khi đọc tin tức, phải đọc nhiều nguồn và quan điểm khác nhau.
8. Kiểm tra xem có phải là trò đùa hay không: xác minh nếu câu chuyện được đăng trên trang web châm biếm hoặc không chứa thông tin châm biếm.
9. Kiểm tra tính xác thực của hình ảnh: sử dụng công cụ để kiểm tra nguồn gốc và xem xét xem hình ảnh có bị chỉnh sửa hay không.

Các phương pháp trên chỉ là các phương pháp phát hiện tin tức một cách thủ công, chỉ áp dụng với một số lượng tin tức ít. Do đó, các phương pháp phát hiện tin giả một cách tự động dựa vào Machine Learning hay Deep Learning đang được phát triển một cách rộng rãi để giải quyết các vấn đề trên.

2.2 Lý thuyết bài toán

Các phương pháp FND dựa trên ML truyền thống hiện tại yêu cầu kỹ thuật tính năng. Theo các đặc điểm mà các mô hình sử dụng, các phương pháp đó có thể được chia thành ba loại: đặc điểm ngôn ngữ (linguistic features), đặc điểm cấu trúc thời gian (temporal-structural features) và đặc điểm lai (hybrid features).

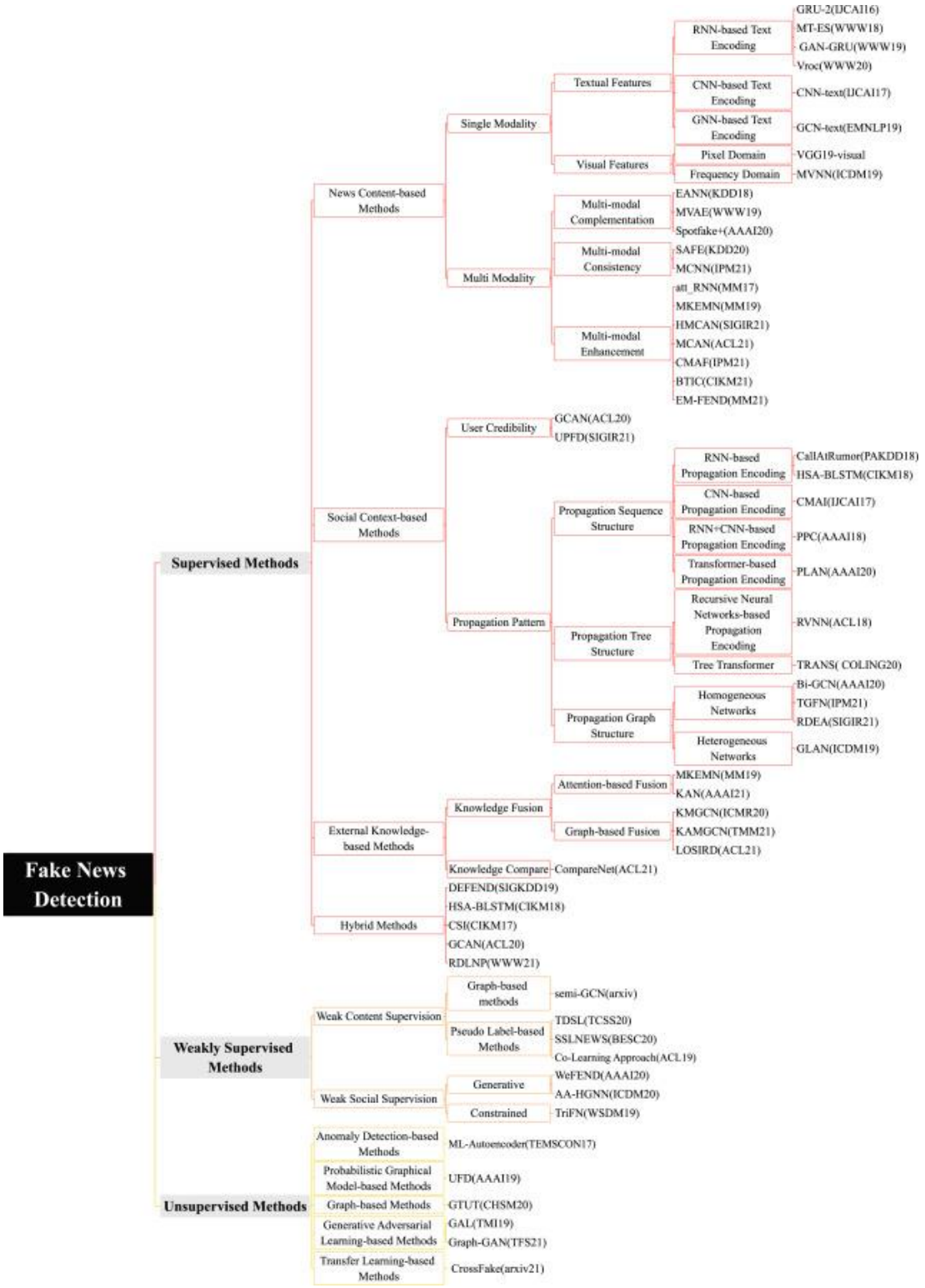
Theo (Zhou & Zafarani, 2021), bài toán Fake News Detection được chia làm 4 loại: phương pháp phát hiện dựa trên kiến thức bên ngoài (external knowledge-based), phương pháp phát hiện dựa trên kiểu (knowledge-based), phương pháp phát

hiện dựa trên sự lan truyền (propagation-based) và phương pháp phát hiện dựa trên độ tin cậy (credibility-based).

Theo (Shu et al., 2017), bài toán Fake News Detection được chia làm: phương pháp dựa trên nội dung tin tức và các phương pháp dựa trên bối cảnh xã hội.

Theo (Hu et al., 2022), Fake News Detection dựa trên Deep Learning được chia thành 3 loại: giám sát (supervised), giám sát yếu (weakly supervised) và không giám sát (unsupervised).

- Phương pháp được giám sát (supervised): mô hình được huấn luyện trên một tập dữ liệu đã được gán nhãn. Mục tiêu là để mô hình có thể dự đoán nhãn cho dữ liệu mới dựa trên những học được từ tập dữ liệu huấn luyện. Trong quá trình huấn luyện, mô hình sẽ cố gắng giảm thiểu sai số giữa các dự đoán của nó và nhãn thực sự.. Có thể chia các phương pháp này thành: dựa trên nội dung tin tức, dựa trên bối cảnh xã hội và dựa trên kiến thức bên ngoài.
- Phương pháp yếu (weakly supervised): mô hình được huấn luyện sử dụng một tập dữ liệu gồm cả dữ liệu đã gán nhãn và dữ liệu chưa gán nhãn. Mục tiêu là để mô hình có thể cải thiện hiệu suất học của mình bằng cách học từ cả dữ liệu có nhãn lẫn không có nhãn. Theo cách thức nhận được sự giám sát yếu, có thể chia phương pháp bán giám sát thành giám sát nội dung yếu và giám sát xã hội yếu.
- Phương pháp không giám sát (unsupervised): học không có giám sát là một phương pháp nơi mô hình được huấn luyện trên dữ liệu không gán nhãn. Mục tiêu của học không có giám sát không phải là dự đoán nhãn, mà là phát hiện ra các mẫu, cấu trúc, hoặc kiến thức tiềm ẩn từ dữ liệu đó. Học không có giám sát bao gồm các kỹ thuật như phân cụm và giảm kích thước.



Hình 2.2 Phân loại bài toán dựa trên Deep Learning

(Nguồn: (Hu et al., 2022))

2.3 Phân tích một số bài báo liên quan

Đặc điểm cơ bản nhất của tin tức là nội dung văn bản của nó. Các nghiên cứu ban đầu về phát hiện tin giả chủ yếu dựa vào nội dung hoặc tiêu đề của tin tức, phương pháp này cần hiểu đặc điểm ngôn ngữ của nó và so sánh nó với các bằng chứng liên quan lấy từ Wikipedia (Hanselowski et al., 2018). Các bài báo đã cho thấy kết quả đầy hứa hẹn trong việc mã hóa tin tức thông tin văn bản như (Riedel et al., 2018) và khác các mô hình dựa trên BERT (Kaliyar et al., 2021). Mặt khác, các cách tiếp cận dựa trên bối cảnh xã hội bao gồm các cam kết xã hội do người dùng điều khiển tiêu thụ tin tức trên nền tảng mạng xã hội. Ví dụ, CSI (Ruchansky et al., 2017) kết hợp hành vi của cả hai bên, người dùng và bài viết và hành vi nhóm của những người dùng truyền bá tin tức.

Trong khi đó, dEFEND (Shu, Cui, et al., 2019) sử dụng nhận xét của người dùng và tin tức dựa trên mạng lưới chú ý có thứ bậc (hierarchical attention network) (Z. Yang et al., 2016). (Han et al., 2020) sử dụng các tính năng như số lượng người theo dõi và số lượng bạn bè mà không cần có bất kỳ thông tin văn bản nào. (Cheng et al., 2021) thể hiện mối quan hệ nhân quả giữa các thuộc tính của người dùng và sự nhạy cảm của người sử dụng. Các bài báo sử dụng GCN (Kipf & Welling, 2017) đã được sử dụng để mô hình hóa cây truyền tin như GCNFN (Monti et al., 2019), BiGCN (Bian et al., 2020) và GNN-CL (Han et al., 2020). UPFD (Dou et al., 2021) điểm chuẩn FakeNewsNet (Shu, Mahudeswaran, et al., 2019) bằng cách cùng nhau nắm bắt tin tức nội dung và bối cảnh ngoại sinh xung quanh của tin tức thông qua propagation tree. GTN (Matsumoto et al., 2021) và TGNF (Song et al., 2021) xem xét thời gian sự khác biệt giữa các tweet/retweet trong propagation trees bằng cách sử dụng tập hợp tin nhắn nhận biết thời gian và mạng lưới chú ý. (Silva et al., 2021) tập trung vào việc dự đoán mô hình lan truyền của giả tin tức ở giai đoạn đầu bằng cách xây dựng lại việc truyền bá tin tức cây. HGAT (Ren & Zhang, 2021) tận dụng biểu đồ mới không đồng nhất (heterogeneous graph) để mô hình hóa mối quan hệ theo cặp giữa tin tức, người sáng tạo và chủ đề thông qua nút và cơ chế cấp độ ngữ nghĩa (semantic-level) attention . Siêu đồ thị (hypergraph) gần đây đã thu hút sự chú

ý giả mạo phát hiện tin tức bằng cách phân cụm các mẫu chia sẻ nội dung tin tức của người dùng (Muhuri & Mukhopadhyay, 2021) hoặc cùng mô hình hóa các mối quan hệ chủ đề giữa tin tức (Borse & Kharate, 2022) với hierarchical attention network.

Các mô hình Mạng nơ-ron đồ thị (GNN) nhằm mục đích đạt được kết quả tốt hơn biểu diễn nút thông qua thông điệp truyền giữa các lân cận cục bộ trong biểu đồ bằng cách sử dụng tổng hợp vùng lân cận (Ding et al., 2020). GCN (Kipf & Welling, 2017) có thể được giải thích như một lân cận tổng hợp mean-pooling. GraphSAGE (Hamilton et al., 2018) sử dụng các tính năng của nút với max pooling hoặc tổng hợp dựa trên LSTM. GAT (Veličković et al., 2018) kết hợp các trọng số chú ý có khả năng đào tạo để tìm hiểu các trọng số trên các nút lân cận khi tổng hợp thông tin lân cận về một nút. Heterogeneous graph đã thể hiện nhiều các loại nút và cạnh (X. Wang et al., 2021) để nâng cao khả năng biểu đạt của đồ thị. Tuy nhiên, các mạng nơ-ron đồ thị hiện tại chỉ xem xét mối quan hệ cặp đôi giữa các đối tượng và không thể áp dụng cho việc học quan hệ không theo cặp.

Gần đây, các phương pháp dựa trên hypergraph như Convolutional hypergraph neural networks (Bai et al., 2020) đã cho thấy kết quả đầy hứa hẹn bằng cách áp dụng phép toán tích chập trên đồ thị cho siêu đồ thị trong môi trường học semi-supervised. Ví dụ: HGNN (Feng et al., 2019) đề xuất xây dựng các siêu cạnh bằng cách kết nối các nút tương tự về mặt ngữ nghĩa dựa trên thước đo khoảng cách. Trong khi Convolutional hypergraph neural networks mở rộng tích chập thao tác trên đồ thị tới siêu đồ thị, chú ý đến siêu đồ thị nâng cao hơn nữa năng lực học tập biểu diễn bằng cách tận dụng mô-đun chú ý (Bai et al., 2020). Hyper GAT (Ding et al., 2020) sử dụng cơ chế chú ý cấp độ kép để nắm bắt thông tin bậc cao giữa các từ và câu cho phân loại tài liệu. Ngoài ra, siêu đồ thị Mạng nơ-ron đã được ứng dụng rộng rãi trong nhiều lĩnh vực khác các ứng dụng, chẳng hạn như hệ thống khuyến nghị (J. Wang et al., 2020), trực quan phát hiện đối tượng (Feng et al., 2019) và phân loại quảng cáo (Jeong et al., 2022).

2.4 Các bộ dữ liệu sử dụng trong bài toán phát hiện tin giả

Dưới đây là tổng hợp một số bộ dữ liệu phổ biến đã được ứng dụng vào bài toán phát hiện tin giả:

BuzzFace (Santia & Williams, 2018) được tổ chức bằng cách thêm các bình luận liên quan đến tin tức trên Facebook vào tập dữ liệu BuzzFeed. Bộ dữ liệu bao gồm 2263 tin tức và 1,6 triệu bình luận.

LIAR (W. Y. Wang, 2017) bao gồm 12.836 tin tức thực tế được thu thập từ PolitiFact. Mỗi tin tức được dán nhãn mức độ trung thực sáu cấp: đúng, sai, nửa đúng, một phần đúng, hầu như không đúng và gần như đúng.

CREDBANK (Mitra & Gilbert, 2015) là một tập dữ liệu lớn có nguồn gốc từ cộng đồng bao gồm 60 triệu tweet trong 96 ngày kể từ tháng 10 năm 2015. Các tweet liên quan đến gần một nghìn sự kiện tin tức.

FacebookHoax (Tacchini et al., 2017) chứa thông tin về các bài đăng từ các trang Facebook liên quan đến tin tức khoa học (non-hoax) và các trang âm mưu (hoax), được thu thập bằng API Facebook.

Twitter15 và Twitter16 (Ma et al., 2017) lần lượt chứa 1381 và 1181 propagation trees. Trong mỗi tập dữ liệu, cấu trúc cây chứa một tập hợp các tweet nguồn cũng như các luồng lan truyền của chúng, chẳng hạn như phản hồi và tweet lại. Mỗi cây được phân loại thành tin đồn không tin đồn, tin đồn sai sự thật, tin đồn thực sự hoặc tin đồn chưa được xác minh.

Media-Twitter (Boididou et al., 2014) có hai phần: bộ phát triển chứa khoảng 9000 tweet giả và 6000 tweet thực từ 17 sự kiện và bộ thử nghiệm chứa khoảng 2000 tweet từ 35 sự kiện khác.

Weibo-20 (Zhang et al., 2021) là tập dữ liệu phát hiện tin tức giả của Trung Quốc. Nó giữ cài đặt hai lớp (tức là giả hoặc thật cho mỗi mẩu tin). Đối với tin giả, Nó giữ lại 1355 mẩu tin giả trên Media-Weibo và thu thập các mẩu tin được Trung tâm Quản lý Cộng đồng Weibo chính thức xác minh. Đối với tin thật, nó giữ lại 2351 mẩu tin thật của Media-Weibo và tập hợp 850 mẩu tin mới độc đáo trong cùng thời điểm với tin giả. Weibo-20 chứa 3161 mẩu tin giả và 3201 mẩu tin thật.

Weibo21 (Nan et al., 2021) là tập dữ liệu tin tức giả đa miền bằng tiếng Trung với nhãn tên miền được chú thích. Về dữ liệu giả, Weibo21 thu thập các mẩu tin được Trung tâm quản lý cộng đồng Weibo chính thức đánh giá là thông tin sai lệch. Đối với dữ liệu thật, Weibo21 tập hợp các mẩu tin thật trong cùng khoảng thời gian với tin giả, đã được NewsVerify (một nền tảng phát hiện và xác minh những tin tức đáng ngờ trên Weibo) xác minh.

MCG-FNeWS (Cao et al., 2019) là bộ dữ liệu phát hiện tin tức giả đa phương thức. Mỗi mục tin tức trong tập dữ liệu bao gồm cả văn bản và hình ảnh đi kèm. Bộ dữ liệu bao gồm tin tức Sina Weibo từ tháng 5 năm 2012 đến tháng 11 năm 2018.

Ti-CNN (Y. Yang et al., 2023) là một bộ dữ liệu đa phương thức để phát hiện tin tức giả. Bộ dữ liệu chứa tổng cộng 20.015 mục tin tức, trong đó 11.941 mục là giả và 8074 mục là sự thật.

PolitiFact: tin tức trong tập dữ liệu PolitiFact được xuất bản từ tháng 5 năm 2002 đến tháng 7 năm 2018. Các chuyên gia về miền đưa ra các nhãn sự thật cơ bản (sai hoặc thật) cho các mục tin tức trong tập dữ liệu. Kết quả xác minh tính xác thực có thể chứng minh độ tin cậy của các bài báo tương ứng True (2149), Mostly True (2676), Half True (2765), Mostly False (2539), False (2601), Pants on Fire (1322).

GossipCop: là một trang web kiểm tra sự thật. Tin tức trong tập dữ liệu GossipCop được xuất bản từ tháng 7 năm 2000 đến tháng 12 năm 2018. Các chuyên gia về miền đưa ra nhãn sự thật cơ bản cho các mục tin tức trong tập dữ liệu, đảm bảo chất lượng của thẻ tin tức.

FakeNewsNet (Shu, Mahudeswaran, et al., 2019) chứa tin tức từ các trang web xác minh tính xác thực BuzzFeed 18 và PolitiFact. Tập dữ liệu chứa nội dung tin tức, thông tin người dùng và tin nhắn lại. Bộ dữ liệu chứa tổng cộng 23.196 bài báo và 69.733 lượt tweet lại.

PHEME (Zubiaga et al., 2017) được cấu thành bởi các tweet từ nền tảng Twitter. Ngoài ra, nó còn thu thập từ năm nguồn tin nóng, mỗi nguồn đều có một bộ sưu tập các dòng tweet. Mỗi đoạn tweet chứa văn bản cũng như hình ảnh.

WeChat (Y. Wang et al., 2020) là một tập dữ liệu phát hiện tin tức giả bán giám sát. Cơ sở dữ liệu chứa cả các bài báo và bình luận của người dùng. Tập dữ liệu bao gồm tin tức từ nền tảng truyền thông xã hội WeChat 19 trong khoảng thời gian từ tháng 3 năm 2018 đến tháng 10 năm 2018.

Fakeddit (Nakamura et al., 2020) có nguồn gốc từ nhiều subreddits của nền tảng Reddit. Sáu nhãn phân loại được mô tả là: True, Satire/Parody, Deceptive Content, Fake Content, False Connection, Manipulated Content..

FakeHealth (Dai et al., 2020) là bộ dữ liệu dùng để phát hiện tin giả liên quan đến sức khỏe. Dữ liệu trong tập dữ liệu có nguồn gốc từ trang web của Health News Review. Tập dữ liệu chứa nội dung tin tức, bình luận của người dùng và mạng xã hội dành cho mọi người.

CoAID (Cui & Lee, 2020) là một bộ dữ liệu để phát hiện tin tức giả mạo có liên quan đến COVID-19. Tập dữ liệu bao gồm các bài viết tin tức, nhận xét của người dùng và dữ liệu người dùng. Bộ dữ liệu bao gồm 4251 bài báo, 296.000 bình luận của người dùng và các nhãn tin tức thực tế.

FakeCovid (Shahi & Nandini, 2020) là một tập dữ liệu đa miền đa ngôn ngữ thu thập 5182 bài báo được lưu hành ở 105 quốc gia từ 92 người xác minh tính xác thực.

ReCOvery (Zhou et al., 2020) là tập dữ liệu phát hiện tin tức giả đa phương thức có liên quan đến COVID-19 được thu thập từ trang web NewsGaurd. Tập dữ liệu này chứa 2029 bài báo và 1.40820 tweet.

MM-Covid (Li et al., 2020) là tập dữ liệu phát hiện tin tức giả đa ngôn ngữ liên quan đến COVID-19. Tập dữ liệu chứa nội dung tin tức, bài đăng trên mạng xã hội và thông tin không gian về tin tức. Bộ dữ liệu chứa 3981 mục tin giả và 7192 mục tin thật.

Cross-lingual COVID-19 (Du et al., 2021) là tập dữ liệu COVID-19 đa ngôn ngữ chứa cả tin tức tiếng Anh và tiếng Trung về COVID-19. Tập huấn luyện chứa 2840 báo cáo bằng tiếng Anh và 49,43% tin tức là tin giả. Bộ thử nghiệm chứa 200 tin tức Trung Quốc, 43% trong số đó là tin giả.

LUN (Rashkin et al., 2017). Tập dữ liệu chứa tin tức từ PolitiFact và các trang web chị em của nó (PunditFact,…). Tin tức trong tập dữ liệu có thể được phân thành sáu loại: True, Mostly True, Half True, Mostly False, False, and Pants-on-Fire.

GermanFakeNC (Vogel & Jiang, 2019) là bộ dữ liệu phát hiện tin tức giả bằng tiếng Đức. Bộ dữ liệu bao gồm 490 bài báo.

CHƯƠNG 3. MÔ HÌNH ỨNG DỤNG

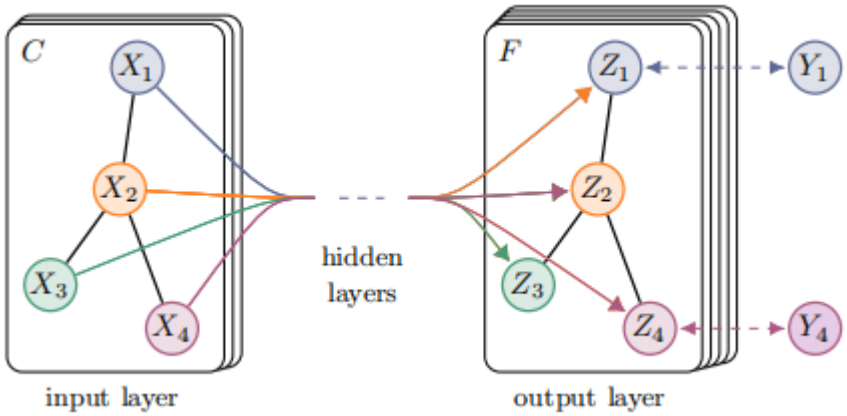
3.1 Graph convolutional networks (GCN)

Xem xét GCN hai lớp để phân loại nút bán giám sát trên biểu đồ có ma trận kề đối xứng A (nhị phân hoặc có trọng số). Đầu tiên chúng ta tính:

$$\hat{A} = \tilde{D}^{-1/2}\tilde{A}\tilde{D}^{-1/2} \tag{3.1}$$

Mô hình chuyển tiếp sau đó có dạng:

$$Z = f(X, A) = softmax(\hat{A}ReLU(\hat{A}XW^{(0)})W^{1}) \tag{3.2}$$



Hình 3.1 Graph Convolutional Network

(Nguồn: (Kipf & Welling, 2017))

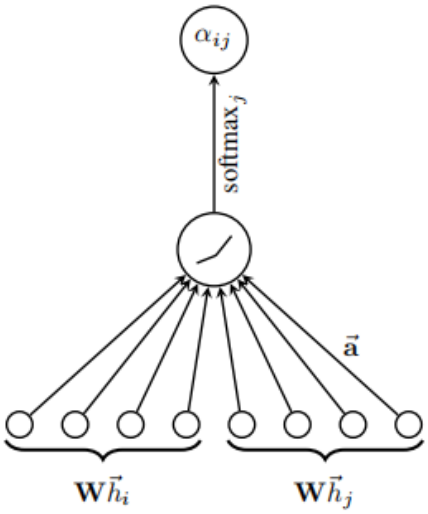
Sơ đồ miêu tả mạng GCN để học bán giám sát (semi-supervised) với các kênh đầu vào C và bản đồ tính năng F ở lớp đầu ra. Cấu trúc đồ thị (cạnh hiển thị là đường màu đen) được chia sẻ trên các lớp, nhãn được ký hiệu là Yi.

3.2 Graph attention networks (GAT)

Trong các thử nghiệm, cơ chế chú ý a là mạng nơ-ron truyền thẳng một lớp, được tham số hóa bởi vectơ trọng số $\vec{a} \in R^{2F}$ và áp dụng tính phi tuyến tính LeakyReLU (α = 0,2). Các hệ số được tính toán bằng cơ chế chú ý, khi đó có thể được biểu diễn dưới dạng:

$$\alpha_{ij} = \frac{exp(LeakyReLU(\overrightarrow{a}^T(W\overrightarrow{h_i} \parallel W\overrightarrow{h_j}))}{\sum k \in N_i exp(LeakyReLU(\overrightarrow{a}^T(W\overrightarrow{h_i} \parallel W\overrightarrow{h_k}))} \quad (3.3)$$

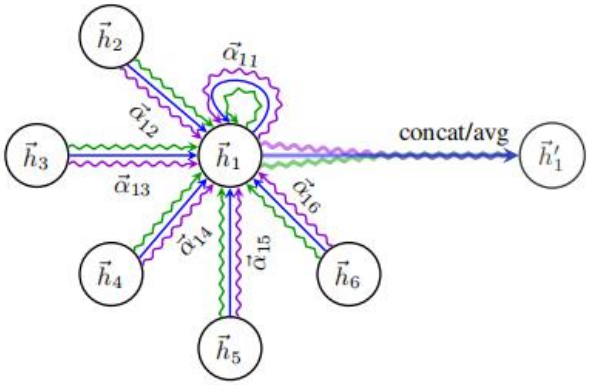
Cơ chế chú ý ($W\overrightarrow{h_i}$,$W\overrightarrow{h_j}$) được mô hình sử dụng, được tham số hóa bởi vectơ trọng số $\vec{a} \in R^{2F}$, áp dụng hàm kích hoạt LeakyReLU.



Hình 3.2 Cơ chế chú ý

(Nguồn: (Veličković et al., 2018))

Hình minh họa về multihead attention của node 1 trên vùng lân cận của nó. Các đặc điểm tổng hợp từ mỗi đầu được ghép nối hoặc lấy trung bình để thu được $\overrightarrow{h_1}$.

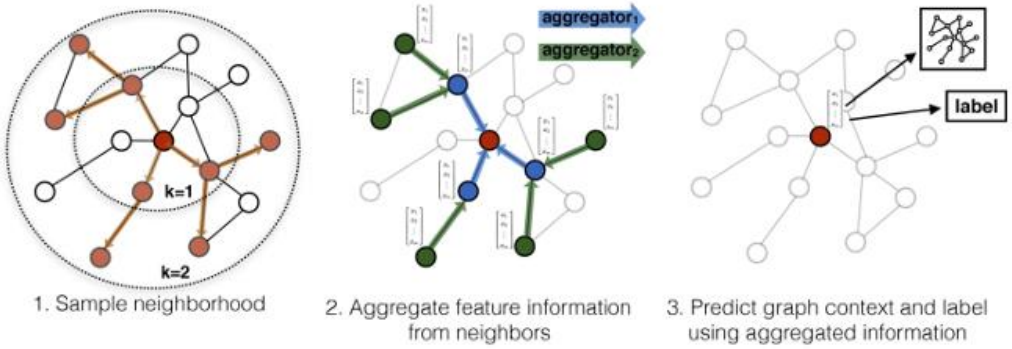


Hình 3.3 Minh họa Multihead attention

(Nguồn: (Veličković et al., 2018))

3.3 GraphSAGE

GraphSAGE (SAmple và aggreGatE). Không giống như các phương pháp nhúng dựa trên yếu tố ma trận, GraphSAGE tận dụng các tính năng của nút (ví dụ: thuộc tính văn bản, thông tin cấu hình nút, node degree) để tìm hiểu một hàm nhúng tổng quát hóa cho các nút không nhìn thấy. Bằng cách kết hợp các tính năng nút trong thuật toán học tập, chúng ta đồng thời tìm hiểu cấu trúc tôpô của vùng lân cận của mỗi nút cũng như sự phân bố các tính năng nút trong vùng lân cận. Mặc dù chỉ tập trung vào các biểu đồ giàu tính năng (ví dụ: dữ liệu trích dẫn với các thuộc tính văn bản), cách tiếp cận của GraphSAGE cũng có thể sử dụng các đặc điểm cấu trúc có trong tất cả các biểu đồ (ví dụ: node degree). Do đó, thuật toán cũng có thể được áp dụng cho các biểu đồ không có tính năng nút.



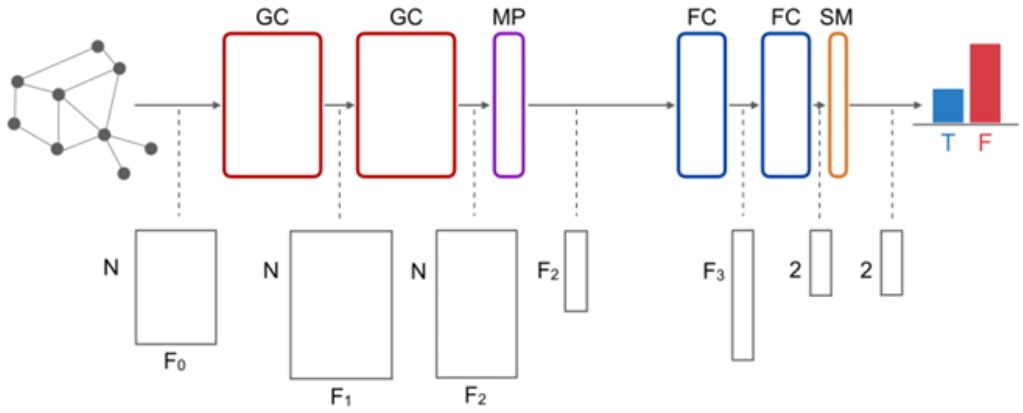
Hình 3.4 Minh họa trực quan GraphSAGE
(Nguồn: (Hamilton et al., 2018))

Thay vì đào tạo một vectơ nhúng riêng biệt cho mỗi nút, nó đào tạo một tập hợp aggregator functions để học cách tổng hợp thông tin tính năng từ vùng lân cận cục bộ của nút (Hình 3.4). Mỗi hàm tổng hợp tổng hợp thông tin từ một số bước nhảy khác nhau hoặc độ sâu tìm kiếm cách xa một nút nhất định. Sau đó, tạo nhúng cho các nút hoàn toàn không nhìn thấy bằng cách áp dụng các hàm tổng hợp (aggregator functions) đã học.

3.4 Graph convolutional networks sử dụng Geometric deep learning (GCNFN)

Là mô hình sử dụng Geometric deep learning để giải quyết bài toán Fake News Detection. Về kiến trúc, mô hình sử dụng bốn lớp Graph CNN với hai lớp tích chập (64 chiều trong mỗi lớp) và hai lớp fully connected (tạo ra các tính năng đầu ra 32 và 2 chiều) để dự đoán xác suất lớp giả/thật.

Hình 3.5 mô tả sơ đồ khối của mô hình. Một graph attention đã được sử dụng trong mọi lớp tích chập để thực hiện các bộ lọc cùng với mean-pooling để giảm kích thước. Sử dụng Scaled Exponential Linear Unit (SELU) như là một lớp phi tuyến tính trên toàn bộ mạng.

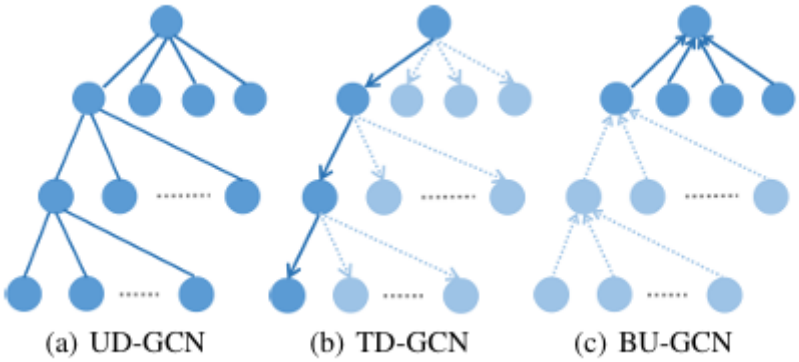


Hình 3.5 Kiến trúc mô hình sử dụng Geometric deep learning
(Nguồn: (Monti et al., 2019))

Trong đó: GC = Graph Convolution, MP = Mean Pooling, FC = Fully Connected, SM = SoftMax layer

3.5 Bi-Directional Graph Convolutional Networks (BiGCN)

Bi-directional GCN (Bi GCN), hoạt động trên cả tuyên truyền tin đồn từ trên xuống và từ dưới lên. Phương pháp được đề xuất có được các tính năng lan truyền và phân tán thông qua hai phần, Mạng tích chập đồ thị từ trên xuống (TD-GCN) và Mạng tích chập đồ thị từ dưới lên (BU-GCN).



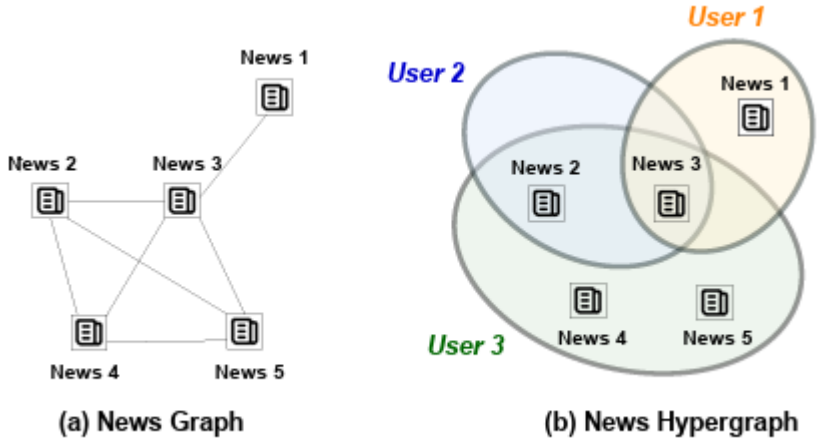
Hình 3.6 UD-GCN, TD-GCN và BU-GCN
(Nguồn: (Bian et al., 2020))

Như thể hiện trong Hình 3.6, TD-GCN cho thông tin từ parent node của một nút trong cây tin đồn để xây dựng sự lan truyền tin đồn trong khi BU-GCN tổng hợp thông tin từ các children node của một nút trong cây tin đồn để đại diện cho sự phân tán tin đồn. Sau đó, các đại diện của sự lan truyền và phân tán được merged lại từ việc nhúng TD-GCN và BU-GCN thông qua fully connected để tạo ra kết quả cuối cùng. Ngoài ra, nó còn ghép nối các features của roots trong cây tin đồn với các tính năng ẩn ở mỗi lớp GCN để tăng cường ảnh hưởng từ gốc rễ của tin đồn.

3.6 Hypergraph Neural Networks (HGFND)

Gần đây, các mạng nơ-ron đồ thị (GNN) đã được áp dụng để tận dụng thông tin quan hệ phong phú hơn giữa cả các trường hợp được gắn nhãn và không gắn nhãn. Nhưng GNN tập trung vào mối quan hệ cặp giữa các tin tức, điều này có thể hạn chế sức mạnh biểu đạt để nắm bắt tin tức giả mạo lan truyền ở cấp độ nhóm. Để giải quyết vấn đề, chúng ta có thể tận dụng một siêu đồ thị (hypergraph) để thể hiện sự tương tác theo nhóm giữa các tin tức.

Trong một hypergraph, mỗi cạnh (được gọi là hyperedge) có thể kết nối ba đỉnh trở lên, cho phép biểu diễn các quan hệ phức tạp hơn so với mạng lưới thông thường (graph), nơi mỗi cạnh chỉ kết nối hai đỉnh. Các mô hình hyper graph nhằm mục đích khai thác cấu trúc phức tạp và mối quan hệ đầy đủ này để cải thiện hiệu suất trong các mô hình.



Hình 3.7 Minh họa Graph và Hypergraph

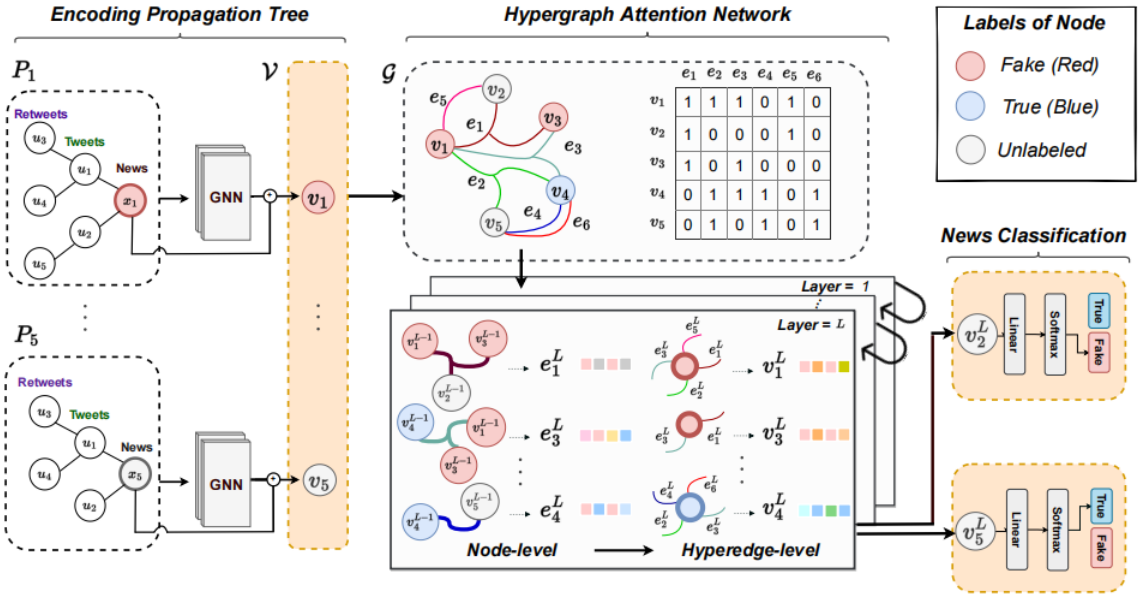
(Nguồn: (Jeong et al., 2022))

Phương pháp xây dựng siêu đồ thị để phát hiện tin tức giả mạo bằng cách sử dụng ba loại siêu cạnh, như sau:

1) User (ID người dùng được chia sẻ giữa các mẩu tin tức): loại siêu cạnh này ngụ ý hành vi chia sẻ tin tức của người dùng. Điều này có nghĩa là một hyperedge tương ứng với User ID. Truy xuất ID người dùng duy nhất từ tweet / retweet trong cây lan truyền của tin tức, nếu một mẩu tin tức chia sẻ cùng ID người dùng với các mẩu tin tức khác, chúng sẽ được ghép lại thành một siêu cạnh.

2) Time (Tin tức được chia sẻ trong phạm vi thời gian gần): loại siêu cạnh này giả định rằng tin tức tương tự có thể xuất hiện và được chia sẻ trong cùng một thời gian. Điều này có nghĩa là một siêu cạnh tương ứng với dấu thời gian tạo tweet / retweet trong cây lan truyền tin tức.

3) Entity (Thực thể được chia sẻ giữa các nội dung tin tức): loại siêu cạnh này nhằm mục đích kết nối tin tức xử lý các nội dung tương tự bằng cách kiểm tra thực thể (loại tin) được chia sẻ giữa các mẩu tin tức.



Hình 3.8 Quy trình của mô hình HGFND

(Nguồn: (Jeong et al., 2022))

Quy trình của mô hình HGFND gồm 3 giai đoạn. Encoding Propagation Tree: mã hóa cây truyền tin để biểu diễn nút trong hypergraph, Hypergraph Attention Network: tìm hiểu thông tin quan hệ giữa các nút và siêu cạnh sử dụng cơ chế chú ý cấp độ kép, lớp cuối cùng để phân loại tin tức.

CHƯƠNG 4. THỰC NGHIỆM

4.1 Dữ liệu thực nghiệm

Tập dữ liệu bao gồm các tin tức giả và tin thật trên Twitter được xây dựng dựa trên thông tin xác minh tính xác thực từ 2 trang web Politifact và Gossipcop. Các biểu đồ retweet ban đầu được trích xuất bởi FakeNewsNet.

Mỗi biểu đồ là một biểu đồ có cấu trúc cây phân cấp trong đó nút gốc biểu thị tin tức; các nút lá là những người dùng Twitter đã tweet lại tin tức gốc. Nút người dùng có cạnh nối với nút tin tức nếu họ đăng lại tweet tin tức. Hai nút người dùng có cạnh nối nếu một người dùng đăng lại tweet tin tức từ người dùng kia.

Thu thập dữ liệu gần 20 triệu tweet lịch sử từ những người dùng tham gia truyền bá tin tức giả mạo trên FakeNewsNet. Nó là kết hợp bốn loại tính năng nút trong tập dữ liệu.

- Feature bert (768 chiều) và spacy (300 chiều) được mã hóa lần lượt bằng BERT và spaCy word2vec đã được huấn luyện trước.
- Feature profile (10 chiều) được lấy từ hồ sơ của tài khoản Twitter.
- Feature content (310 chiều) bao gồm nhúng nhận xét người dùng word2vec (spaCy) embedding 300 chiều cộng với tính năng profile(10 chiều).

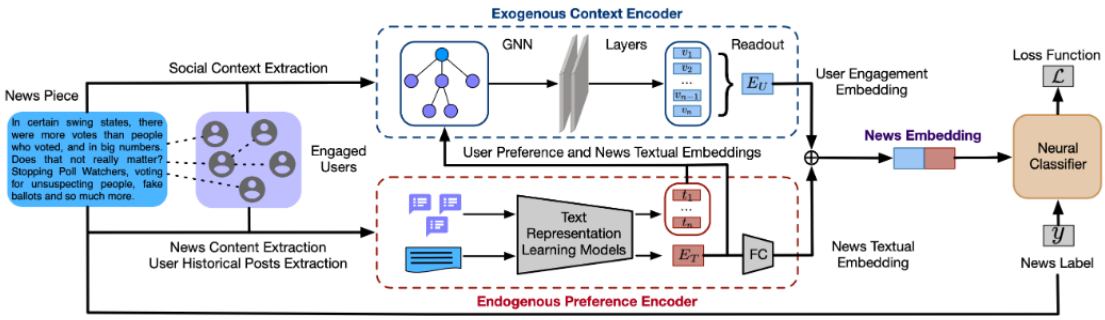
Bảng 4.1 Thống kê 2 tập dữ liệu Politifact và Gossipcop

	Graphs	Fake News	Total Nodes	Total Edges	Avg. Nodes per Graph
Politifact	314	157	41,054	40,740	131
Gossipcop	5464	2732	314,262	308,798	58

(Nguồn: (Dou et al., 2021))

Sử dụng khung phát hiện tin tức giả end-to-end có tên User Preference-aware Fake Detection (UPFD) để lập mô hình nội sinh sở thích và bối cảnh ngoại sinh cùng nhau. Cụ thể, UPFD bao gồm các thành phần chính sau: (1) Để mô hình hóa sở thích

nội sinh của người dùng, họ mã hóa nội dung tin tức và các bài đăng lịch sử của người dùng bằng cách sử dụng các cách học biểu diễn văn bản khác nhau cách tiếp cận. (2) Để có được bối cảnh ngoại sinh của người dùng, họ xây dựng một biểu đồ lan truyền có cấu trúc cây cho mỗi tin tức dựa trên chia sẻ tầng trên phương tiện truyền thông xã hội. Các bài viết tin tức được coi là nút gốc và các nút khác đại diện cho những người dùng đã chia sẻ các bài viết tin tức tương tự. (3) Để tích hợp thông tin nội sinh và ngoại sinh, họ sử dụng các biểu diễn vectơ của tin tức và người dùng làm tính năng nút của họ và sử dụng Mạng thần kinh đồ thị (GNNs) để tìm hiểu cách nhúng tương tác của người dùng chung. Các nhúng tương tác của người dùng và nhúng văn bản tin tức được sử dụng để đào tạo một bộ phân loại thần kinh để phát hiện tin tức giả.



Hình 4.1 UPFD framework

(Nguồn: (Dou et al., 2021))

UPFD có ba thành phần chính. Đầu tiên, với một mẩu tin tức, họ thu thập thông tin lịch sử bài đăng của người dùng tham gia vào tin tức để tìm hiểu nội sinh của người dùng sự ưa thích. Sau đó, họ ngầm trích xuất sở thích của người dùng đã tương tác bằng cách mã hóa các bài đăng lịch sử bằng cách sử dụng các kỹ thuật học biểu diễn văn bản (ví dụ: word2vec, BERT). Dữ liệu văn bản tin tức là được mã hóa bằng cách sử dụng cùng một cách tiếp cận. Thứ hai, để tận dụng bối cảnh ngoại lai của người dùng, họ xây dựng biểu đồ lan truyền tin tức theo thông tin tương tác trên các nền tảng truyền thông xã hội (ví dụ: tin nhắn lại trên Twitter). Thứ ba, họ nghĩ ra một quy trình tổng hợp thông tin có thứ bậc để hợp nhất sở thích nội sinh của người dùng và bối cảnh ngoại sinh.

Cụ thể, họ có được sự tham gia của người dùng bằng cách sử dụng GNN làm bộ mã hóa biểu đồ, trong đó tin tức và phần nhúng của người dùng được mã hóa bởi bộ mã hóa văn bản được sử dụng làm nút tương ứng của chúng các tính năng trong biểu đồ truyền bá tin tức. Các tin tức nhúng cuối cùng được cấu thành bởi sự kết hợp của việc nhúng tương tác của người dùng và nhúng văn bản tin tức.

4.2 Cài đặt chạy thực nghiệm

Trong bài báo cáo này, chúng tôi đã tiến hành thực nghiệm lại các mô hình với tham số được công bố trong các bài báo. Các tham số của mỗi mô hình với mỗi bộ dữ liệu khác nhau được điều chỉnh khác nhau.

Đối với bộ dữ liệu Politifact:

Bảng 4.2 Tham số của các mô hình với bộ dữ liệu Politifact

Model	**Feature**	**Epoch**	**Learning rate**	**Emb_size**	**Batch_num**
BiGCN	bert	50	0.001	128	128
GAT	bert	50	0.001	128	128
GCN	spacy, profile	60	0.001	128	128
GCN	bert	100	0.001	128	128
SAGE	profile	70	0.01	128	128
SAGE	spacy	45	0.01	128	128
SAGE	bert	30	0.01	128	128
GCNFN	profile	80	0.001	128	128
GCNFN	spacy	50	0.001	128	128
GCNFN	bert	60	0.001	128	128
HGFND	bert	200	0.001	128	128

Đối với bộ dữ liệu Gossipcop:

Bảng 4.3 Tham số của các mô hình với bộ dữ liệu Gossipcop

Model	**Feature**	**Epoch**	**Learning rate**	**Emb_size**	**Batch_num**
BiGCN	bert	35	0.001	128	128
GAT	bert	30	0.001	128	128
GCN	profile	50	0.01	128	128
GCN	spacy, bert	50	0.001	128	128
SAGE	profile	50	0.01	128	128
SAGE	spacy bert	80	0.001	128	128
GCNFN	profile, spacy, bert	50	0.001	128	128
HGFND	bert	200	0.001	128	128

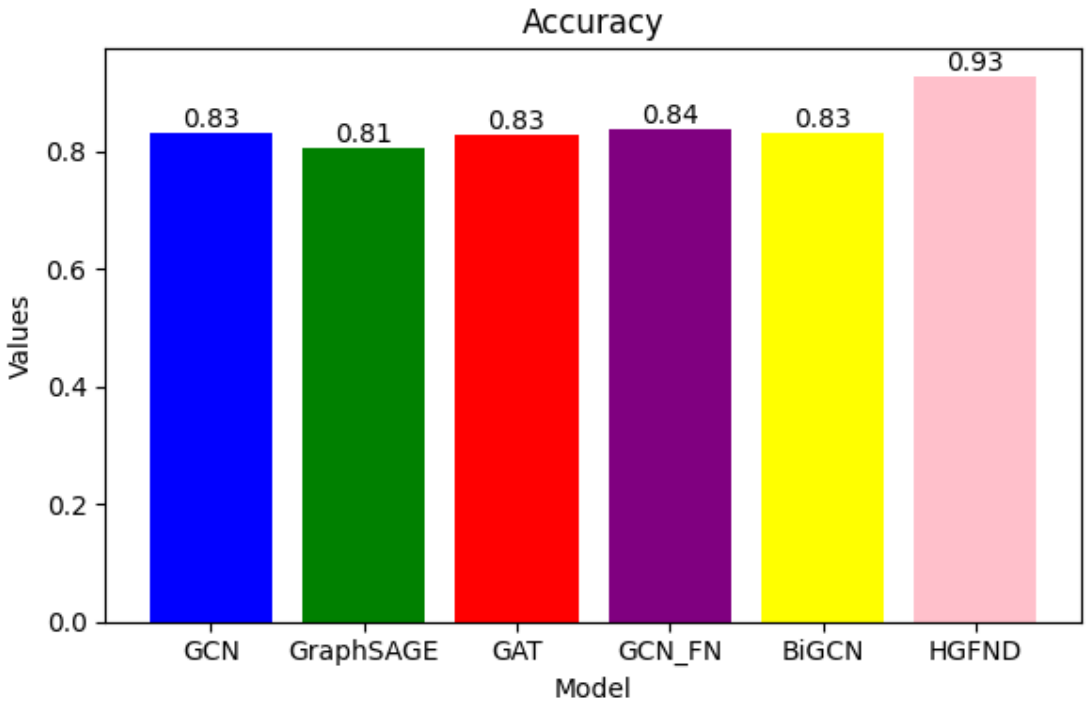
4.3 Kết quả chạy thực nghiệm

Trong phần này, chúng tôi đã chạy và thực nghiệm lại các mô hình đã được công bố bởi các bài báo để so sánh kết quả và rút ra các kết luận. Cụ thể, đó là các mô hình đã được trình bày trong chương 3: Bi-GCN, GCNFN, UPFD-GCN, UPFD-GAT, UPFD-SAGE và HGFND.

Các tham số của mô hình như: số epoch, learning rate, batch size, hidden dimension, dropout… được điều chỉnh theo hyperparameters đã được công bố kèm theo các bài báo.

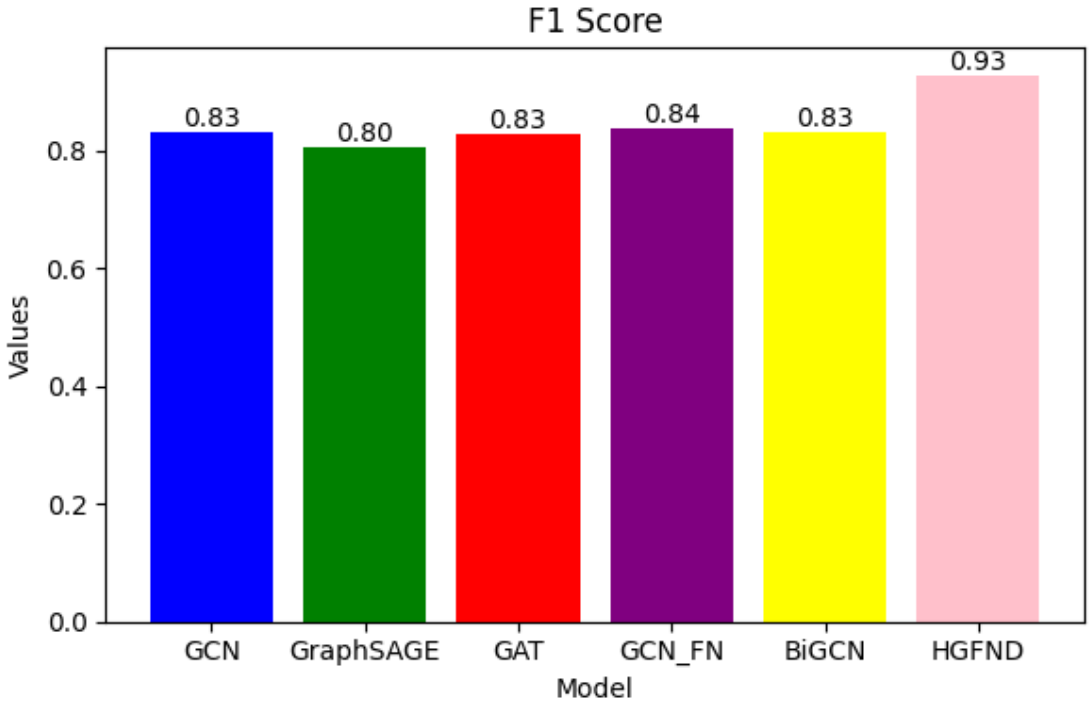
Kết quả chạy thực nghiệm cho thấy hai độ đo accuracy và F1-score không quá chênh lệch so với các bài báo đã công bố.

Đây là kết quả thực nghiệm accuracy của các mô hình trên bộ dữ liệu Politifact.



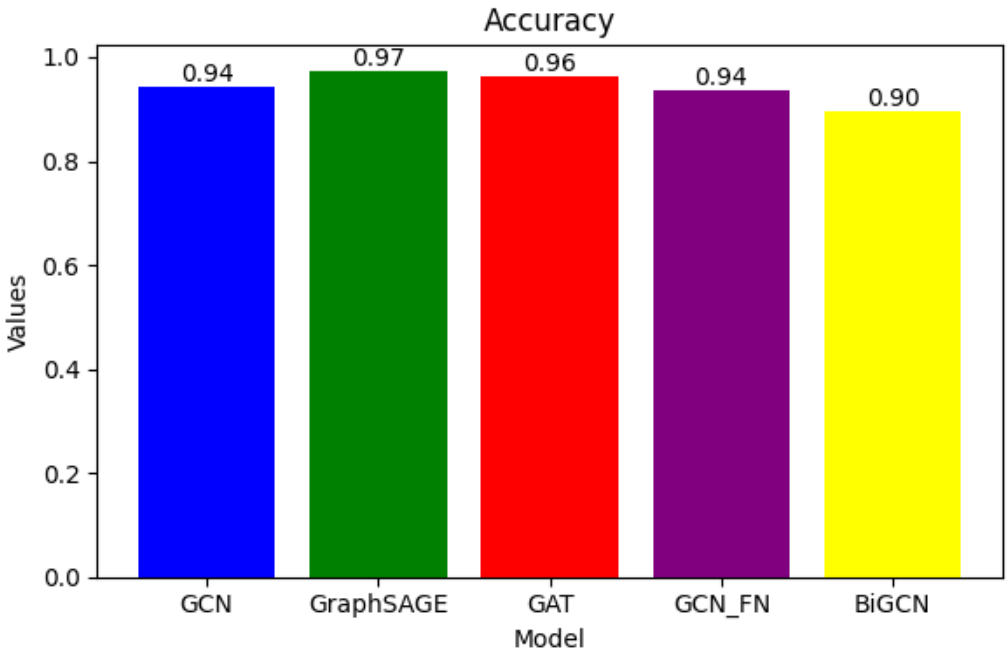
Hình 4.2 Accuracy trên bộ dữ liệu Politifact

Đây là kết quả thực nghiệm F1-score của các mô hình trên bộ dữ liệu Politifact.



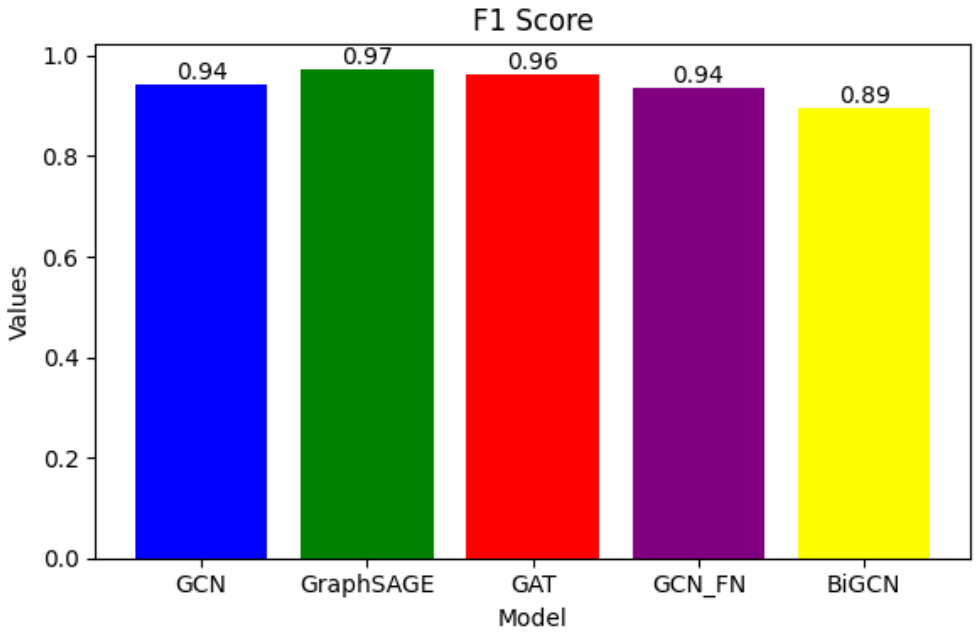
Hình 4.3 F1-score trên bộ dữ liệu Politifact

Đây là kết quả thực nghiệm accuracy của các mô hình trên bộ dữ liệu Gossispcop.



Hình 4.4 Accuracy trên bộ dữ liệu Gossipcop

Đây là kết quả thực nghiệm F1-score của các mô hình trên bộ dữ liệu Gossipcop.



Hình 4.5 F1-score trên bộ dữ liệu Gossipcop

Accuracy và F1-score của các mô hình trên cả 2 tập dữ liệu được thống kê trong bảng 4.4. Do thiếu tài nguyên phần cứng nên chưa thể thực nghiệm mô hình HGFND trên bộ dữ liệu Gossispcop, các kết quả mô hình HGFND trên bộ dữ liệu Gossipcop được lấy trực tiếp từ bài báo về HGFND (Jeong et al., 2022) đã được công bố).

Bảng 4.4 Thống kê accuracy và F1-score trên 2 tập dữ liệu thử nghiệm

Model	**Politifact**		**Gossispcop**	
	Accuracy	*F1-score*	*Accuracy*	*F1-score*
GCN	83.26	83.19	94.28	94.24
SAGE	80.54	80.41	97.36	97.35
GAT	82.81	82.69	96.37	96.35
GCNFN	83.81	83.71	93.57	93.57
BiGCN	83.26	83.18	89.57	89.49
HGFND	**92.76**	**92.76**	**97.76**	**97.76**

4.4 Phân tích kết quả

Về hiệu suất tổng quan: trái ngược với các phương pháp dựa trên sự lan truyền, HGFND cho thấy hiệu suất tốt hơn hẳn. Mô hình HGFND đạt hiệu suất cao nhất trên cả hai tập dữ liệu, với Accuracy và F1-score đều ở mức 92.76% cho Politifact và 97.76% cho Gossipcop. Điều này cho thấy HGFND có khả năng tổng quát hóa và xử lý tốt trên cả hai tập dữ liệu. Điều này có thể giải thích được bằng cơ chế của HGFND, nó cho thấy sự khác biệt giữa việc sử dụng graph thường và hypergraph. Hypergraph giảm sự mất mát dữ liệu so với graph thường bằng phương pháp gán nhóm cho các node.

Xét mô hình GraphSAGE trên tập dữ liệu Politifact, accuracy và F1-score chỉ đạt ở mức 80% (chênh lệch so với mô hình cao nhất trên tập Politifact khoảng 12%). Mặt khác, mô hình GraphSAGE trên tập dữ liệu Gossipcop có accuracy và F1-score đạt đến mức 97.36% (gần bằng với mô hình cao nhất trên tập Gossipcop). Điều này

cho thấy cơ chế lấy mẫu và tổng hợp thông tin từ hàng xóm (sample và aggregate) của GraphSAGE có ưu thế hơn trên tập dữ liệu lớn.

Xét trên tập Politifact, kết quả accuracy của các mô hình: GCN, GAT, GraphSAGE, GCNFN và Bi-GCN gần như chỉ ở mức từ 80-84 còn mô hình HGFND thì có kết quả cao hơn hẳn các mô hình còn lại, cụ thể là 92.76. Nếu xét trên tập Gossipcop, kết quả accuracy của các mô hình có chênh lệch nhưng không quá nhiều (chỉ riêng Bi-GCN là thấp dưới 90), còn lại các mô hình có kết quả khá tốt từ 93-97. Kết quả của mô hình HGFND khá cao nhưng cũng không qua chênh lệch so với nhưng mô hình khác. Sự khác biệt hiệu xuất của mô hình HGFND trên 2 bộ dữ liệu chứng minh mô hình HGFND mạnh mẽ đối với số lượng nhãn đào tạo hạn chế.

CHƯƠNG 5. KẾT LUẬN

5.1 Kết luận

Vấn đề phát hiện tin giả vẫn đang là vấn đề vô cùng cấp thiết cần được giải quyết trong thời đại tin tức tràn lan ngày nay. Trong bài báo cáo này chúng tôi đã tập trung vào việc nghiên cứu, phân tích và so sánh các mô hình hiện có để giải quyết vấn đề Fake News Detection. Dựa trên việc thực nghiệm lại các mô hình cho thấy, HGFND đang là mô hình tốt nhất trên cả 2 tập dữ liệu nhờ sử dụng HyperGraph thay vì graph như các mô hình khác. Điều đó có được là nhờ cơ chế giảm sự mất mát dữ liệu so với graph thường bằng phương pháp gán nhóm cho các node của HyperGraph. Ngoài ra, từ việc so sánh kết quả thực nghiệm của mô hình HyperGraph cũng cho thấy mô hình này mạnh mẽ đối với số lượng nhãn đào tạo hạn chế.

5.2 Hướng phát triển

Trong tương lai, vấn đề phát hiện tin giả sẽ càng khó khăn hơn do càng ngày lượng tin tức sinh ra càng nhiều, các thủ đoạn phát tán tin giả ngày càng tinh vi hơn. Để tiếp tục phát triển bài toán tự động phát hiện tin giả trong tương lai, chúng ta cần phải làm rộng tập thông tin ra, cụ thể chúng ta cần tìm kiếm nhiều loại thông tin hơn, tìm ra mối quan hệ giữa chúng để ứng dụng vào graph và các biến thể của chúng. Có thể tiếp tục khai thác cấu trúc của hypergraph vì nó đã thể hiện khá tốt trong vấn đề phát hiện tin giả gần đây bằng cách hợp nhất hoặc tách các siêu cạnh để tăng khả năng biểu diễn các mối quan hệ. Cuộc chiến chống tin giả là một cuộc chiến lâu dài và phát triển theo thời gian, do đó nó cần sự phát triển không ngừng từ các mô hình, không chỉ phát hiện tin giả dựa vào nội dung, hay bối cảnh mà còn dựa vào nhiều thông tin khác có thể được phát hiện trong tương lai.

TÀI LIỆU THAM KHẢO

Bai, S., Zhang, F., & Torr, P. H. S. (2020). *Hypergraph Convolution and Hypergraph Attention* (arXiv:1901.08150). arXiv. https://doi.org/10.48550/arXiv.1901.08150

Bian, T., Xiao, X., Xu, T., Zhao, P., Huang, W., Rong, Y., & Huang, J. (2020). *Rumor Detection on Social Media with Bi-Directional Graph Convolutional Networks* (arXiv:2001.06362). arXiv. https://doi.org/10.48550/arXiv.2001.06362

Boididou, C., Papadopoulos, S., Kompatsiaris, Y., Schifferes, S., & Newman, N. (2014). Challenges of computational verification in social multimedia. *Proceedings of the 23rd International Conference on World Wide Web*, 743–748. https://doi.org/10.1145/2567948.2579323

Borse, A., & Kharate, D. G. (2022). *Fake News Prediction using Hierarchical Attention Network and Hypergraph* (SSRN Scholarly Paper 4043857). https://doi.org/10.2139/ssrn.4043857

Cao, J., Sheng, Q., Qi, P., Zhong, L., Wang, Y., & Zhang, X. (2019). *False News Detection on Social Media* (arXiv:1908.10818). arXiv. https://doi.org/10.48550/arXiv.1908.10818

Cheng, L., Guo, R., Shu, K., & Liu, H. (2021). Causal Understanding of Fake News Dissemination on Social Media. *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 148–157. https://doi.org/10.1145/3447548.3467321

Cui, L., & Lee, D. (2020). *CoAID: COVID-19 Healthcare Misinformation Dataset* (arXiv:2006.00885). arXiv. https://doi.org/10.48550/arXiv.2006.00885

Dai, E., Sun, Y., & Wang, S. (2020). Ginger Cannot Cure Cancer: Battling Fake Health News with a Comprehensive Data Repository. *Proceedings of the International AAAI Conference on Web and Social Media*, *14*, 853–862. https://doi.org/10.1609/icwsm.v14i1.7350

Ding, K., Wang, J., Li, J., Shu, K., Liu, C., & Liu, H. (2020). *Graph Prototypical Networks for Few-shot Learning on Attributed Networks* (arXiv:2006.12739). arXiv. https://doi.org/10.48550/arXiv.2006.12739

Dou, Y., Shu, K., Xia, C., Yu, P. S., & Sun, L. (2021). *User Preference-aware Fake News Detection* (arXiv:2104.12259). arXiv. https://doi.org/10.48550/arXiv.2104.12259

Du, J., Dou, Y., Xia, C., Cui, L., Ma, J., & Yu, P. S. (2021). *Cross-lingual COVID-19 Fake News Detection* (arXiv:2110.06495). arXiv. https://doi.org/10.48550/arXiv.2110.06495

Fallis, D. (2015). What Is Disinformation? *Library Trends*, *63*, 401–426. https://doi.org/10.1353/lib.2015.0014

Feng, Y., You, H., Zhang, Z., Ji, R., & Gao, Y. (2019). *Hypergraph Neural Networks* (arXiv:1809.09401). arXiv. https://doi.org/10.48550/arXiv.1809.09401

Hamilton, W. L., Ying, R., & Leskovec, J. (2018). *Inductive Representation Learning on Large Graphs* (arXiv:1706.02216). arXiv. https://doi.org/10.48550/arXiv.1706.02216

Han, Y., Karunasekera, S., & Leckie, C. (2020). *Graph Neural Networks with Continual Learning for Fake News Detection from Social Media* (arXiv:2007.03316). arXiv. https://doi.org/10.48550/arXiv.2007.03316

Hanselowski, A., PVS, A., Schiller, B., Caspelherr, F., Chaudhuri, D., Meyer, C. M., & Gurevych, I. (2018). *A Retrospective Analysis of the Fake News Challenge Stance Detection Task* (arXiv:1806.05180). arXiv. https://doi.org/10.48550/arXiv.1806.05180

Hu, L., Wei, S., Zhao, Z., & Wu, B. (2022). Deep learning for fake news detection: A comprehensive survey. *AI Open*, *3*, 133–155. https://doi.org/10.1016/j.aiopen.2022.09.001

Jeong, U., Alghamdi, Z., Ding, K., Cheng, L., Li, B., & Liu, H. (2022). *Classifying COVID-19 Related Meta Ads Using Discourse Representation Through a Hypergraph*. 35–45. https://doi.org/10.1007/978-3-031-17114-7_4

Kaliyar, R. K., Goswami, A., & Narang, P. (2021). FakeBERT: Fake news detection in social media with a BERT-based deep learning approach. *Multimedia Tools and Applications*, *80*(8), 11765–11788. https://doi.org/10.1007/s11042-020-10183-2

Kipf, T. N., & Welling, M. (2017). *Semi-Supervised Classification with Graph Convolutional Networks* (arXiv:1609.02907). arXiv. https://doi.org/10.48550/arXiv.1609.02907

Li, Y., Jiang, B., Shu, K., & Liu, H. (2020). *MM-COVID: A Multilingual and Multimodal Data Repository for Combating COVID-19 Disinformation* (arXiv:2011.04088). arXiv. https://doi.org/10.48550/arXiv.2011.04088

Ma, J., Gao, W., & Wong, K.-F. (2017). *Detect rumors in microblog posts using propagation structure via kernel learning*. 708. https://doi.org/10.18653/v1/P17-1066

Matsumoto, H., Yoshida, S., & Muneyasu, M. (2021). Propagation-Based Fake News Detection Using Graph Neural Networks with Transformer. *2021 IEEE 10th Global Conference on Consumer Electronics (GCCE)*, 19–20. https://doi.org/10.1109/GCCE53005.2021.9621803

Mitra, T., & Gilbert, E. (2015). CREDBANK: A Large-Scale Social Media Corpus With Associated Credibility Annotations. *Proceedings of the International AAAI Conference on Web and Social Media*, *9*(1), Article 1. https://doi.org/10.1609/icwsm.v9i1.14625

Monti, F., Frasca, F., Eynard, D., Mannion, D., & Bronstein, M. M. (2019). *Fake News Detection on Social Media using Geometric Deep Learning* (arXiv:1902.06673). arXiv. https://doi.org/10.48550/arXiv.1902.06673

Muhuri, S., & Mukhopadhyay, D. (2021). A Hypergraph Clustering-based Technique for Detecting Fake News from Broadcasting Network. *2021 Asian Conference on Innovation in Technology (ASIANCON)*, 1–5. https://doi.org/10.1109/ASIANCON51346.2021.9544803

Nakamura, K., Levy, S., & Wang, W. Y. (2020). *r/Fakeddit: A New Multimodal Benchmark Dataset for Fine-grained Fake News Detection* (arXiv:1911.03854). arXiv. https://doi.org/10.48550/arXiv.1911.03854

Nan, Q., Cao, J., Zhu, Y., Wang, Y., & Li, J. (2021). MDFEND: Multi-domain Fake News Detection. *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, 3343–3347. https://doi.org/10.1145/3459637.3482139

Rashkin H., Choi E., Jang J. Y., Volkova S., & Choi Y. (2017). *Truth of Varying Shades: Analyzing Language in Fake News and Political Fact-Checking*. 2931–2937. https://doi.org/10.18653/v1/D17-1317

Ren, Y., & Zhang, J. (2021). *Fake News Detection on News-Oriented Heterogeneous Information Networks through Hierarchical Graph Attention* (arXiv:2002.04397). arXiv. https://doi.org/10.48550/arXiv.2002.04397

Riedel, B., Augenstein, I., Spithourakis, G. P., & Riedel, S. (2018). *A simple but tough-to-beat baseline for the Fake News Challenge stance detection task* (arXiv:1707.03264). arXiv. https://doi.org/10.48550/arXiv.1707.03264

Ruchansky, N., Seo, S., & Liu, Y. (2017). CSI: A Hybrid Deep Model for Fake News Detection. *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 797–806. https://doi.org/10.1145/3132847.3132877

Santia, G., & Williams, J. (2018). BuzzFace: A News Veracity Dataset with Facebook User Commentary and Egos. *Proceedings of the International AAAI Conference on Web and Social Media*, *12*(1), Article 1. https://doi.org/10.1609/icwsm.v12i1.14985

Shahi, G. K., & Nandini, D. (2020). *FakeCovid—A Multilingual Cross-domain Fact Check News Dataset for COVID-19*. https://doi.org/10.36190/2020.14

Shu, K., Cui, L., Wang, S., Lee, D., & Liu, H. (2019). dEFEND: Explainable Fake News Detection. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 395–405. https://doi.org/10.1145/3292500.3330935

Shu, K., Mahudeswaran, D., Wang, S., Lee, D., & Liu, H. (2019). *FakeNewsNet: A Data Repository with News Content, Social Context and Spatialtemporal Information for Studying Fake News on Social Media* (arXiv:1809.01286). arXiv. https://doi.org/10.48550/arXiv.1809.01286

Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake News Detection on Social Media: A Data Mining Perspective. *ACM SIGKDD Explorations Newsletter*, *19*(1), 22–36. https://doi.org/10.1145/3137597.3137600

Silva, A., Han, Y., Luo, L., Karunasekera, S., & Leckie, C. (2021). Propagation2Vec: Embedding partial propagation networks for explainable fake news early detection. *Information Processing & Management*, *58*(5), 102618. https://doi.org/10.1016/j.ipm.2021.102618

Song, C., Shu, K., & Wu, B. (2021). Temporally evolving graph neural network for fake news detection. *Information Processing & Management*, *58*(6), 102712. https://doi.org/10.1016/j.ipm.2021.102712

Tacchini, E., Ballarin, G., Della Vedova, M. L., Moret, S., & de Alfaro, L. (2017). *Some Like it Hoax: Automated Fake News Detection in Social Networks* (arXiv:1704.07506). arXiv. https://doi.org/10.48550/arXiv.1704.07506

Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., & Bengio, Y. (2018). *Graph Attention Networks* (arXiv:1710.10903). arXiv. https://doi.org/10.48550/arXiv.1710.10903

Vogel, I., & Jiang, P. (2019). Fake News Detection with the New German Dataset "GermanFakeNC." In A. Doucet, A. Isaac, K. Golub, T. Aalberg, & A. Jatowt (Eds.), *Digital Libraries for Open Knowledge* (pp. 288–295). Springer International Publishing. https://doi.org/10.1007/978-3-030-30760-8_25

Wang, J., Ding, K., Hong, L., Liu, H., & Caverlee, J. (2020). Next-item Recommendation with Sequential Hypergraphs. *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1101–1110. https://doi.org/10.1145/3397271.3401133

Wang, W. Y. (2017). *"Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection* (arXiv:1705.00648). arXiv. https://doi.org/10.48550/arXiv.1705.00648

Wang, X., Ji, H., Shi, C., Wang, B., Cui, P., Yu, P., & Ye, Y. (2021). *Heterogeneous Graph Attention Network* (arXiv:1903.07293). arXiv. https://doi.org/10.48550/arXiv.1903.07293

Wang, Y., Yang, W., Ma, F., Xu, J., Zhong, B., Deng, Q., & Gao, J. (2020). Weak Supervision for Fake News Detection via Reinforcement Learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, *34*(01), Article 01. https://doi.org/10.1609/aaai.v34i01.5389

Wardle, C., & Derakhshan, H. (2017). *INFORMATION DISORDER: Toward an interdisciplinary framework for research and policy making Information Disorder Toward an interdisciplinary framework for research and policymaking*.

Yang, Y., Zheng, L., Zhang, J., Cui, Q., Li, Z., & Yu, P. S. (2023). *TI-CNN: Convolutional Neural Networks for Fake News Detection* (arXiv:1806.00749). arXiv. https://doi.org/10.48550/arXiv.1806.00749

Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., & Hovy, E. (2016). Hierarchical Attention Networks for Document Classification. In K. Knight, A. Nenkova, & O. Rambow (Eds.), *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 1480–1489). Association for Computational Linguistics. https://doi.org/10.18653/v1/N16-1174

Zhang, X., Cao, J., Li, X., Sheng, Q., Zhong, L., & Shu, K. (2021). Mining Dual Emotion for Fake News Detection. *Proceedings of the Web Conference 2021*, 3465–3476. https://doi.org/10.1145/3442381.3450004

Zhou, X., Mulay, A., Ferrara, E., & Zafarani, R. (2020). ReCOVery: A Multimodal Repository for COVID-19 News Credibility Research. *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, 3205–3212. https://doi.org/10.1145/3340531.3412880

Zhou, X., & Zafarani, R. (2021). A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities. *ACM Computing Surveys*, *53*(5), 1–40. https://doi.org/10.1145/3395046

Zubiaga, A., Liakata, M., & Procter, R. (2017). Exploiting Context for Rumour Detection in Social Media. In G. L. Ciampaglia, A. Mashhadi, & T. Yasseri (Eds.), *Social Informatics* (pp. 109–123). Springer International Publishing. https://doi.org/10.1007/978-3-319-67217-5_8