MultiLayerET: A Unified Representation of Entities and Topics using Multilayer Graphs

Jumanah Alshehri ✉, Marija Stanojevic, Parisa Khan, Benjamin Rapp, Eduard Dragut, and Zoran Obradovic

Center for Data Analytics and Biomedical Informatics, Temple University, Philadelphia, PA, USA
{shehri.j,marija.stanojevic,edragut,zoran.obradovic}@temple.edu

Abstract. Many online news outlets, forums, and blogs provide a rich stream of publications and user comments. This rich body of data is a valuable source of information for researchers, journalists, and policymakers. However, the ever-increasing production and user engagement rate make it difficult to analyze this data without automated tools. This work presents MultiLayerET, a method to unify the representation of entities and topics in articles and comments. In MultiLayerET, articles' content and associated comments are parsed into a multilayer graph consisting of heterogeneous nodes representing named entities and news topics. The nodes within this graph have attributed edges denoting weight, i.e., the strength of the connection between the two nodes, time, i.e., the co-occurrence contemporaneity of two nodes, and sentiment, i.e., the opinion (in aggregate) of an entity toward a topic. Such information helps in analyzing articles and their comments. We infer the edges connecting two nodes using information mined from the textual data. The multilayer representation gives an advantage over a single-layer representation since it integrates articles and comments via shared topics and entities, providing richer signal points about emerging events. MultiLayerET can be applied to different downstream tasks, such as detecting media bias and misinformation. To explore the efficacy of the proposed method, we apply MultiLayerET to a body of data gathered from six representative online news outlets. We show that with MultiLayerET, the classification F1 score of a media bias prediction model improves by 36%, and that of a state-of-the-art fake news detection model improves by 4%.

Keywords: News Mining · Multilayer Graphs · Text Mining · Social Network Analysis

1 INTRODUCTION

The amount of published articles is steadily increasing, and readers are shifting toward online platforms because of the affordable technology costs and the ability to share their opinions. News articles are conveniently accessed either via news outlets' websites or news aggregator platforms, like Google News, that collect articles and recommend a subset of them to readers according to their interests.

The current news ecosystem escalates the competition between platforms and motivates them to scale and enhance their systems. Such growth makes it harder for readers and analysts to get a complete picture of a particular event or entity without falling into the news source bias or the contrasting opinions hidden between the lines of articles across sources. Therefore, building an automated system able to represent and model semantic data is essential to help readers, researchers, journalists, and decision-makers understand emerging events and their associated entities. Such a system will benefit downstream applications, such as news popularity, media bias, news recommendations, and fake news detection. We propose a unified representation of news and comments in our MultiLayerET system based on shared entities and topics.

To our knowledge, most news research focuses on extracting topics and entities from the news articles and omits the user-generated content associated with news items in comment sections. In this study, we formulate the problem as follows, *having a large set of documents in the form of articles and their associated comments. We aim to extract a rich graph representation of emerging events or topics*. To achieve this objective, we introduce our system MultiLayerET. The output of our system is a heterogeneous *attributed multilayer graph*. A multilayer graph is a graph with a set of nodes, each assigned a type from a set of types [36]; our graph has two types of edges: intra-layer edges, which connect nodes within the same layer, and inter-layer edges, which connect nodes across layers. In MultiLayerET, nodes are entities and topics, and edges are attributed. We consider three types of attributes: 1) co-occurrence, which is the co-occurrence frequency of a pair of nodes, 2) contemporaneity, which is the published times of the documents where the two nodes co-occur (note that articles and comments have temporal information, published and posted times, respectively), and 3) sign, which denotes the aggregated sentiment of the text where two nodes co-occur. The extracted graph consists of two layers, the G_a layer where nodes and edges are extracted from articles, and the G_c layer where the nodes and edges are obtained from comment sections related to the articles in G_a.

The multilayer approach provides a unique representation of emerging events. To illustrate, Figure 1 shows both layers with their attributes on a set of documents from Washington Post (WP) during the 2016 US election, we unfold the graph based on time to show the contemporaneity attribute over three months period from May - July 2016. Comparing G_a and G_c, we can see that G_c complements G_a greatly including entities that are not mentioned in G_a. For example, the comment layer mentions other candidates such as *Bernie Sanders* during the 2016 election that are from the same party and their relationship with the topic *Hillary Clinton's email story*. Moreover, *James Comey* the director of the FBI's investigation, appears in the comment layer around July when the investigation started, and *Colin Powell* and his relationship with Clinton email controversy appeared in early stages in G_c. Related to topics, we see that the topic *election debates* and its relationship with the *Clinton's email story* appeared much earlier in G_c compared to G_a. This indicates that information mined from comments

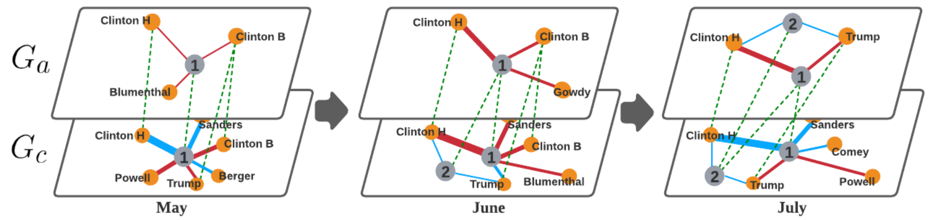

Fig. 1: Monthly representation of a two-layer graph of Washington Post articles and comments during the 2016 US election over three months (May - July 2016). Gray node 1 represents the topic “Hillary Clinton’s email story," and gray node 2 represents the topic “2016 elections debates". Orange nodes represent associated entities in articles and comments layers $(G_a, G_c) \in WP$. Blue and red edges represent the intra-layer connections, while green edges represent the inter-layer connections. The width of the link represents the weight, red edges represent a negative sentiment, while blue edges represent positive sentiment.

complements that mined from articles; we capture even richer (latent) information if we mine concomitantly articles and comments from multiple sources.

MultiLayerET can be applied to other areas of study besides news like blogs and their comments and research papers discussed on social media platforms. In this work, our focus is on online news and its application. We apply our system to six English online news sources, Washington Post (WP), Cable News Network (CNN), Wall Street Journal (WSJ), British Broadcasting Corporation (BBC), Fox News (FN), and Daily Mail (DM). The number of articles and comments varies across sources, resulting in graphs with a different number of nodes per source. The number of nodes varies between $4K$ and $200K$, producing small to extensive graphs. For this study, we focus on entities and topics related to politics, and we consider people entities and associated topics. We make the following contribution in this work:

1. We introduce MultiLayerET, a system that represents entities and topics in articles and comments as a heterogeneous attributed multilayer graph. Such representation assists in highlighting significant events and their associated entities; it enhances the analysis of emerging events reported in news streams.
2. We analyze the topological graph structure to show the unique information encoded within the network.
3. We show that MultiLayerET improves downstream applications, e.g., media bias classification by 36% and fake news detection by 4%.
4. We build a dictionary of over $1M$ people entities; the dictionary contains useful data such as aliases crucial to downstream applications.

2 RELATED WORK

Here we give an overview of the related work to properly position MultiLayerET within the literature on topic modeling and entity extraction from articles and their comments.

Comments carry valuable information concerning public opinion [47,48,49,53,54]. Utilizing user comments can enhance the performance of many models for downstream tasks, such as, fake news detection [26,32,50,51], news popularity prediction [21,23,52], media bias [22,25,46], news recommendations [37,38,39], and news summarization [41].

Latent Dirichlet Allocation (LDA) [1] is a hierarchical Bayesian model that generates probabilities of corpora for a given document. LDA is the Bayesian version of the Probabilistic Latent Semantic Analysis (PLSA) model [13]. It is considered the foundation of many other models such as Topic-link LDA [12], Labeled LDA [11], and Spatial LDA [14]. LDA is utilized in many applications (e.g., information retrieval, topics overlapping, and visualization) where the extraction of topics is needed. For a given set of news sources, a line of work aims to link entities, and topics using statistical [5,10] and other techniques [29,30,31,33,34]. Other studies, construct an entity-centric graph and the topic associated with it [6,7,8,9]. [16] uses an attribute proximity graph to mine events reported in the news.

Most works in news mining either work with news articles and ignore comments or with comments and ignore articles. There are a few exceptions. For example, [40,42] shows that comments combined with articles improve topic discovery and [18] shows that comments improve explainability in fake news detection. Our MultiLayerET approach creates a unified representation of articles and comments, giving a richer graph representation of topics discussed in the news and their associated entities. It is argued that entities are an essential component significantly affecting the comprehension level of a given document [15]. Therefore, entities are first-class citizens in MultiLayerET as the candidate documents to be mined are added to the graph according to an input set of entities. We extract topics from a set of documents, which assists in producing a coherent list of terms for each extracted topic. To our knowledge, our work is the first to propose a heterogeneous attributed multilayer graph to represent information in news streams.

3 MultiLayerET

This section defines the problem and the notations used in this paper. We also introduce our system MultiLayerET to represent entities and topics in articles and comments. The system pipeline is shown in Figure 2.

3.1 Problem Formulation

Having a source $s \in S$ that consists of several articles and their comments $s_i = \{\langle a_1, (c_{11}, .., c_{1n})\rangle, .., \langle a_n, (c_{n1}, .., c_{nm})\rangle\}$, MultiLayerET's objective is to extract

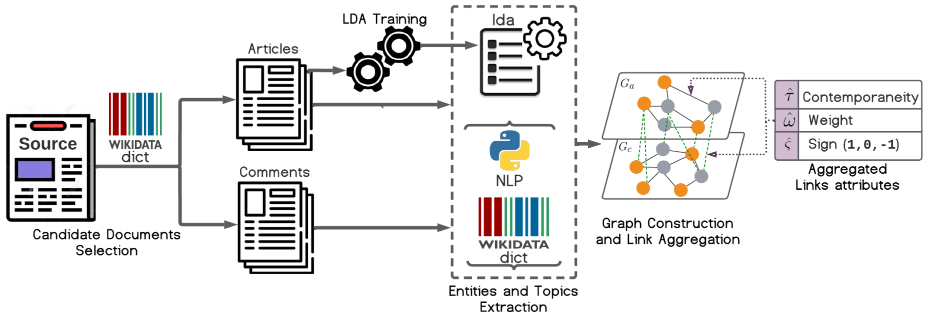

Fig. 2: MultiLayerET pipeline to represent entities and topics in articles and comments

(major) topics and entities mentioned in articles and comments. MultiLayerET produces an undirected heterogeneous attributed multilayer graph $G = \{G_a, G_c\}$ where G_a is the article graph and G_c is the comment graph. The nodes in the graph are sets of entities E and topics T, respectively, along with attributed edges contemporaneity $\hat{\tau}$, co-occurrence count $\hat{\omega}$, and sign $\hat{\varsigma}$. Table1 describes the notations used in this paper.

Table 1: Notations used in this paper.

Symbol	Description	Symbol	Description
S	Set of sources	T	Set of topics $\in \{D_a, D_c\}$
s_i	The $i^{th} \in S$ where $i = 1..n$	t	topic node $\in T$
A	Set of articles $\in s$	G_a	Articles graph
a_i	The i^{th} article $\in A$	G_c	Comments graph
C	Set of comments section	G	Multilayer graph $\{G_a, G_c\}$
c_{ij}	The j^{th} comment for an article a_i	τ	Contemporaneity, link attribute
D_a	Set of candidate articles $\in A$	ω	Weight, link attribute
D_c	Set of candidate comments $\in C$	ς	Sign, link attribute
$dict$	Entities dictionary	$\hat{\tau}$	Aggregated contemporaneity
lda	Best LDA model	$\hat{\omega}$	Aggregated weights
E	Set of entities in $D \in \{D_a, D_c\}$	$\hat{\varsigma}$	Aggregated signs
e	entity node $\in E$		

3.2 Selection of Candidate Documents

Importing a large set of articles and comments is a challenge in itself. In this work, we assume one can access an extensive collection of documents. We focus on the challenge of gleaning knowledge from such data. An essential step in the MultiLayerET pipeline is collecting representative documents; this preliminary curation process helps eliminate noise, enabling gaining meaningful and interpretable information.

In order to fulfil this objective, we build an entity dictionary *dict* from *wikidata*[1]. The dictionary contains more than one million person entities along with their names, alias names, affiliations, descriptions, and URLs to their Wikidata pages. The list of aliases contains the most commonly misspelled names to capture better the varied ways entities are mentioned in articles and comments. In *dict*, entities are divided into six categories according to their current or most recent affiliation, such as politicians and officeholders, military figures, sports figures, musicians and actors, writers, and social media personalities.

For a specific source s, article a and comment c are added to D_a, the set of candidate articles, and D_c, the set of candidate comments, respectively, if any entity token of e is present in the entity dictionary *dict*. For example, we add the article "Bernie Sanders FINALLY unloaded on Hillary Clinton for not being 'qualified.' Here's why."[2] from WP and its subset of comments since the entities *Hillary Clinton* and *Bernie Sander* in our *dict* are present in that article. In this work, we focus on politics-related documents; therefore, we use the politicians and office holders *dict* to extract coherent topic terms.

3.3 Entity Extraction and Topic Mining

Once candidate documents D_a and D_c are obtained, we preprocess the data to construct the graph where $nodes = \{E, T\}$. In this section we describe how to extract entity nodes E and topic nodes T from D_a and D_c.

Entity Nodes: To obtain meaningful information from the selected articles and comments, we focus on entities present in our entity dictionary *dict*. We utilize it along with NLTK[45] and TextBlob[3] to extract entity name phrases. Entities may appear multiple times in the same D_a and D_c; therefore, we keep track of their frequencies.

Topic Nodes: To extract topics, we train an LDA [1] model for each D_a in s. The LDA model looks into the bag of words from D_a and returns a set of terms with their probabilities. Each probability represents how much each word contributes to that cluster. We evaluate the quality of the *lda* model by measuring 1) Coherence score [3], which is the semantic degree of similarity between high-scoring words in a given cluster (topic), and 2) Perplexity score; normalized log-likelihood. MultiLayerET chooses the best *lda* model according to the coherence value; a higher coherence value returns interpretable topics. If two LDA models have the same coherence score, we choose the one with the lower perplexity score. We only train *lda* on D_a, assuming that D_c will align with at least one of the extracted topics. We select the most relevant topic for a given article in D_a and comment D_c based on the topic probability given by the trained *lda* model.

[1] https://www.wikidata.org/
[2] Full article: https://wapo.st/3yOMYdO
[3] https://textblob.readthedocs.io/en/dev/

Algorithm 1 Single-layer graph construction

1: **procedure** GC(D, lda, dict)
`/* Nodes Extraction */`
2: $E = \{token \in D \wedge token \in dict\}$
3: $T = \{lda(d) \ \ where \ \ d \in D\}$
`/* Intra-layer Edges */`
4: $Initialize\ G \leftarrow [\ ]$
for $(e, t) \in \langle E, T \rangle$ **do**
  for $d \in D$ **do**
    if (e, t) *co-occur ind* **then**
      $\varsigma_d \leftarrow Majority(\varsigma_i)$
      $G[e, t] \leftarrow \langle \tau_d, \omega_d, \varsigma_d \rangle$
`/* Link aggregation */`
5: **for** $(e, t) \in G$ **do**
  $\hat{\tau}_d \leftarrow \cup_{\langle \tau_d, \omega_d, \varsigma_d \rangle \in G[e,t]} \tau_d$
  $\hat{\omega}_d \leftarrow \sum_{\langle \tau_d, \omega_d, \varsigma_d \rangle \in G[e,t]} \omega_d$
  $\hat{\varsigma}_d \leftarrow Majority(\varsigma_d)$
  $G[e, t] \leftarrow \langle \hat{\tau}_d, \hat{\omega}_d, \hat{\varsigma}_d \rangle$
6: **return** Single layer $\in G$

Algorithm 2 Multilayer graph construction

1: **procedure** MGC(D_a, D_c, lda, dict)
`/* Graph construction */`
2: $G_a \leftarrow$ GC(D_a, lda, dict)
3: $G_c \leftarrow$ GC(D_c, lda, dict)
`/* Inter-layer Edges */`
4: $Initialize\ G \leftarrow [\ ]$
`/* Entity-entity edges */`
5: **for** $(e_a, e_c) \in \{G_a, G_c\}$ **do**
  if $(e_a = e_c)||(e_a, e_c) \in \{a_i, c_{ij}\}$ **then**
    $G[e_a, e_c] \leftarrow \langle \hat{\tau_d}, \hat{\omega_d} \rangle$
`/* Topic-topic edges */`
6: **for** $(t_a, t_c) \in \{G_a, G_c\}$ **do**
  if $(t_a = t_c)||(t_a, t_c) \in \{a_i, c_{ij}\})$ **then**
    $G[t_a, t_c] \langle \hat{\tau_d}, \hat{\omega_d} \rangle$
7: **return** G

3.4 Graph Construction

In our setting, G is a multilayer, undirected, weighted, attributed graph. The first layer in the graph G is G_a, which represents the article graph, and the second layer G_c is the comment graph. The nodes in our graph are heterogeneous; they consist of 1) a set of entities E and 2) a set of topics T that maps to one of the extracted topics. Some nodes may appear in G_c but not in G_a and vice-versa. This phenomenon depends on the commenters' behavior; they tend to discuss or leave out entities and topics that might be mentioned in the article. Algorithm 1 summarizes the graph construction for a single layer in G, and Algorithm2 shows the process of constructing the multilayer graph G.

Edges Construction: Since we are dealing with a multilayer graph, we have two types of edges: intra-layer edges, which represent edges within the same layer, and inter-layer edges, which represent edges between the layers. **Intra-layer edges** are (e, t), which links $e \in E$ with $t \in T$. This link captures the relationship between (e, t) pairs in news articles and comments. The **inter-layer**

edges are *entity-entity links* (e_a, e_c), which link entities together; this link gives an intuition of how entities are connected between articles and their comments. Another type of inter-layer edge is *topic-topic* (t_a, t_c), which links topics together; this helps in projecting topics in G and analyzing the level of relevancy between events in news articles and comments. Edges are formed between nodes v and w if they co-occur in a single document d.

Edge Attributes and Aggregation: Once a link is formed between a pair of nodes, we compute its attributes ω, τ, and ς. The first attribute is ω, representing the link weight. For two nodes v and w, we compute the co-occurrence frequency in a single document D, where D can be an article or comment. This attribute represents the connection strength. Second, τ contemporaneity, is a concatenated list of all publish times where node v and w co-occurred. This attribute assists in understanding the temporal evolution of pair of nodes. Finally, ς is the sign attribute representing the text sentiment between a pair of nodes co-occurring in d. This attribute is different as it is only found in intra-layer edges *(e,t)*. The value of ς is 1 if the sentiment of d is positive, -1 if it is negative, and 0 if it is neutral. We calculate the aggregated set $\langle \hat{\tau}, \hat{\omega}, \hat{\varsigma} \rangle$ of edges as follows, $\hat{\tau}$ is a concatenation of all τ for a given pair of nodes. $\hat{\omega}$ is the sum of all ω for a given pair of nodes, and finally, $\hat{\varsigma}$ is the majority vote of all ς for a given pair of nodes.

3.5 Graph Construction Complexity

The graph construction complexity of Algorithm 2 is $\mathcal{O}(|S| \cdot |T| \cdot (|D_a|^2 \cdot max_{len}(D_a)^2 + |D_c|^2 \cdot max_{len}(D_c)^2)$. Here, $|S|$ is the number of sources, $|T|$ is the number of unique topics, $|D_a|$ is the total number of the selected candidate articles, $|D_c|$ is the total number of selected candidate comments, $max_{len}(D_a)$ is the maximum number of tokens calculated over all selected candidate articles, and $max_{len}(D_c)$ is the maximum number of tokens calculated over all selected candidate comment sections. The graph construction is quadratic, where $|D_a|$ and $|D_c|$ play larger roles in controlling complexity compared to $|S|$ and $|T|$. This indicates that construction runtime grows gracefully with the number of articles and comments.

4 Graph Analysis

Here, we describe the dataset and analyze the topological structure of the multilayer graph G.

4.1 Data

News articles and comments were collected from Google News [2] between 2015 and 2017; the database contains over one million articles and 33 million comments from 22 thousand difference English and Spanish news sources. For this study, we draw six English news sources, Washington Post (WP), Cable News

Table 2: Candidate documents D statistics for each source, showing the total number of articles and comments used in the study.

Dataset	WP	DM	FN	CNN	WSJ	BBC
no. articles	$36K$	$14K$	$9K$	$4K$	$2K$	101
no. comments	$290K$	$90K$	$410K$	$14K$	$69K$	$105K$

Network (CNN), Wall Street Journal (WSJ), British Broadcasting Corporation (BBC), Fox News (FN), and Daily Mail (DM). We selected all articles and comments for this study between January 2016 and July 2016. Table 2 shows the datasets statics.

4.2 Topological Graph Structure Analysis

We selected sources where the number of nodes and edges varies across sources, leading to the construction of small to extensive graphs as shown in Table 3. In addition, the size of the graph varies between layers; for example, CNN and WSJ have a similar number of nodes in G_a. However, WSJ has a 3 times larger number of nodes in G_c compared to CNN. This phenomenon will aid in better understating the structural differences across sources.

Table 3: Topological structure properties for each layer G_a and G_c, and multilayer graph G. N_a = Number of nodes in article layer, N_c = Number of nodes in comments layer, E_a= Number of edges in article layer, E_c= Number of edges in comments layer, avg N_d = Average node degree for each layer G_a and G_c, $Diameter$= Diameter of the layer largest component in each layer G_a and G_c, $Inter$= The number of multilayer graph inter-layer edges, avg C = Multilayer average clustering coefficient, and r = Multilayer assortativity coefficient.

Dataset	N_a	N_c	E_a	E_c	avg N_d		Diameter		Inter	avg C	r
					G_a	G_c	G_a	G_c			
WP	1.2K	2K	3.6K	18K	3.50	5.01	7	6	200K	0.82	-0.34
DM	1K	1.3K	1.3K	2.8K	2.53	4.15	6	5	56K	0.81	-0.33
FN	515	1.7K	3K	10K	2.96	4.81	6	4	115K	0.80	-0.35
WSJ	308	2K	475	6K	3.08	4.98	6	4	149K	0.83	-0.34
CNN	297	534	447	10K	3.01	3.76	6	4	16K	0.82	-0.35
BBC	59	176	75	436	2.54	4.95	7	5	4K	0.81	-0.19

We analyze each layer, the multilayer graph structure, and edge sentiment to understand the topological graph structure. In terms of topological structure, we consider the following properties: *1) Average Node Degree*: that helps understand the connectivity differences between layers in G. *2) Diameter*: that shows the graph connectivity, which indicates how many steps we need to take to traverse the graph; we calculated the diameter of the largest component in each layer. *3) avg C*: which measures how well the nodes tend to form clusters [4]; a value close to 1 means that nodes have a high tendency to form clusters, while a value near to 0 means otherwise. *4) Assortativity Coefficient*: this property indicates the tendency of nodes to be connected, whether they have the same degree magnitude, large, or low-degree. Assortativity is calculated as the Pearson correlation coefficient of nodes at either side of an edge. The assortativity value ranges from -1 to +1; positive values mean that nodes of similar degrees connect, while negative values mean that large-degree nodes tend to attach to low-degree nodes. The degree sequence of the graph heavily influences the measure. Finally, *5) Edge Sentiment Distribution* which is the edge sign $\hat{\varsigma}$, gives the stance of an entity toward a topic.

Across Sources Analysis: Table 3 shows that the largest graph is *WP* with more than $1K$ nodes and $3K$ intra-layer edges in G_a and $2K$ nodes and $18K$ intra-layer edges in G_c, respectively. The smallest graph is *BBC* with 59 nodes and around 176 intra-layer edges in G_a, and around 75 nodes and 436 intra-layer edges in G_c. The average N_d indicates that the nodes tend to be connected similarly across sources, which suggests that the size of the graph does not affect the connectivity. The diameter for the largest component across sources is similar, and we can see that the diameter of G_a is larger than that of G_c.

Multilayer Graph Analysis: Comparing G_a and G_c together, we can see that G_c is always larger than G_a. This indicates that users tend to mention entities and discuss topics that are not present in articles. In other words, users tend to drift into topics unrelated to that of the article, but still mentioning entities present in the article along with new ones. The number of inter-layer edges is much greater than the number of intra-layer edges resulting of high average clustering coefficient across all outlets. Although *BCC* is the smallest graph, it has the highest density among all sources, which suggests that *BCC*'s topics and entities are highly connected compared to other sources. The analysis of assortativity coefficient shows that nodes with high degree tend to be connected with nodes with smaller degree; this is a sign of the existence of hubs.

Signed Edge Analysis: Signed edges give an intuition of the sentiment difference between articles and comments. In G_a, we observe that the number of positive and neutral edges represent $78\% - 85\%$ of the total signed edges in all sources. The percentages indicate that most entities have positive or neutral sentiments towards a particular topic. In G_c, neutral edges represent between $6\% - 10\%$, while the percentages of both positive and negative edges are between $90\% - 94\%$. This phenomenon indicates that even though sources may have some inherent (political) bias, the articles are written in a way that their text conveys positive or neutral sentiment. Users, however, tend to express more polarized opinions, which explains the low percentage of neutral edges.

5 System Evaluation

In this section, we present the added benefit of MultiLayerET on two downstream applications, media bias classification, and fake news detection. We compare methods that only utilize the textual representation against the same methods when combined with our MultiLayerET system.

5.1 Experimental Setup

We pre-process the text by removing stop words, punctuation, and digits. To obtain the base of the words, we utilize NLTK [45] to perform lemmatization, which removes the conjugation ending of the word. Our comparison models are: 1) Doc2Vec [43], which is trained to predict words in the text; it produces a dense vector for a given text. We train Doc2Vec for 100 epochs to produce 300 dimension vectors for articles and comments. 2) BERT[27], which generates an expressive feature embedding for a given text using a self-attention mechanism and bidirectional cross attention. We utilize BERT$_{base}$ to get the text feature representation. Then we feed the feature vectors to a feed-forward network with a softmax function to get the predictions.

In MultilayerET, we use Node2Vec [17] to obtain the graph feature representation. Node2Vec maps the graph nodes to low-dimensional vectors while preserving the graph structure. We train Node2Vec and produce a vector feature representation of 100 dimensions. Then we concatenate the graph features with text features obtained with BERT. Once we obtain the feature vectors, we feed them to a feed-forward network with a softmax function to get the predictions. We investigate the performance of MultiLayerET by analyzing the performance of each layer of MultiLayerET. We run experiments 1) using only the graph representation of articles layer G_a along with the representation of the article obtained by BERT, 2) using the graph representation of comments layer G_c along with the comments representation obtained by BERT, and 3) using the multilayer representation of MultiLayerET along with the articles and comments representation obtained by BERT.

In both applications, we repeat experiments five times and test on a different fold not used for training. The dataset split is 80:20 ratios for training, and testing, respectively. We report the average accuracy, precision, recall, F1 score, and standard deviation.

5.2 Media Bias Classification

To perform this experiment, we selected articles and comments from six news sources mentioned in Section 4.1 that discuss well-known politicians from different political parties, such as Donald Trump and Hillary Clinton. We label all examples from each source as left, right, or center based on their media bias rank given by AllSides[4]. To get a balanced dataset, we randomly selected around $1K$

[4] https://www.allsides.com/media-bias/media-bias-ratings

Table 4: Performance on MediaBias. The first row is the input representation; G_a and G_c use only the articles or the comments layer, respectively. MultiLayerET means using the representation of the multilayer graph. We report the average scores for each metric and standard deviation

Representation	Doc2Vec	BERT	G_a	G_c	MultiLayerET
Accuracy	0.547(0.004)	0.687(0.006)	0.785(0.009)	0.806(0.009)	**0.912 (0.005)**
Precision	0.525(0.009)	0.656(0.009)	0.736(0.008)	0.791(0.009)	**0.880 (0.007)**
Recall	0.538(0.008)	0.669(0.007)	0.739(0.008)	0.795(0.006)	**0.894 (0.003)**
F1 score	0.545(0.008)	0.686(0.007)	0.774(0.007)	0.797(0.006)	**0.901 (0.003)**

article and $5K$ comments from each outlet except for BBC, where we have only 101 articles. This result in having a smaller sample of the center class. The total number of articles and comments is about $6K$ and $30K$ respectively, and the proportion of classes are: left = 34%, center = 10%, and right=56%.

The worst performance was obtained using only Doc2Vec representation, where the accuracy is 54% (Table 4). Using any part of MultiLayerET enhanced the model performance from 21% to 37%. In sign edge analysis (Section 4.2) we show that journalists tend to write articles mostly with positive or neutral sentiments, which makes it harder to understand the hidden bias in online news from the textural representation itself. We also observe that using the G_a or G_c already enhances the prediction results across all metrics. From the last column of Table 4, one notices that the best accuracy and F1 score are obtained when the full MultiLayerET system is used. This supports our hypothesis that the comment section is important in increasing the prediction accuracy in downstream tasks.

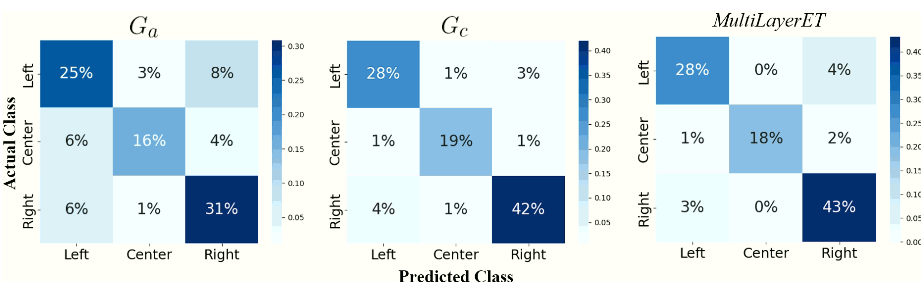

Fig. 3: Confusion Matrix on Media Bias Classification for proposed methods

To better understand the advantage of using MultiLayerET compared to separate layers G_a and G_c, we plot the confusion matrix for each of these experiments as shown in Figure 3. MultiLayerET performs the best in predicting each class. The left class with the highest error rate in all cases is the most challenging class to predict. MultiLayerET and G_c have a lower error rate in predicting the center class than G_a. Looking at the misclassified examples, when MultiLayerET is not sure about the prediction, it consistently predicts the right class and avoids predicting the center class. On the contrary, G_a and G_c randomly assign a class to miss classified examples. This observation indicated that MultiLayerET learns better in imbalanced data cases compared to G_a and G_c.

5.3 Fake News Detection

To evaluate MultiLayerET on Fake News Detection we utilize the benchmark dataset FakeNewsNet [19,20]. We focus on political news retrieved from *PolitiFact* articles and comments (tweets) and use it to perform a binary classification, where texts are labeled as fake and real. *PolitiFact* consists of 415 news articles with around $89K$ comments, where 35% are real news and 65% are fake. We compare MultiLayerET to dEFEND [18] since it utilizes both articles and comments. It reports the best results in FakeNewsNet compared to alternative models for fake news detection.

Table 5 shows that our MultiLayerET-based method for fake news detection outperforms dEFEND's reported accuracy between 1% - 6% in F1 score. The results highlight the importance of utilizing the interaction between entities and topics in this classification task. We note that when we use G_a alone, our prediction performs similarly to that of dEFEND; dEFEND outperforms our prediction model when we use G_c alone.

Table 5: Performance on fake news detection. The first row is the input representation; G_a and G_c mean that only articles and comment layers are utilized, respectively. MultiLayerET means that the entire multilayer graph is used. We report the average scores for each metric and standard deviation. dEFEND results are reported as in [18]; we did not report dEFEND standard deviation since we do not have access to the results.

Model	dEFEND	G_a	G_c	MultiLayerET
Accuracy	0.904	0.939(0.007)	0.895(0.008)	**0.972 (0.003)**
Precision	0.902	0.919(0.009)	0.878(0.007)	**0.942 (0.003)**
Recall	0.956	0.919(0.009)	0.880(0.007)	**0.959 (0.003)**
F1 score	0.928	0.929(0.009)	0.892(0.008)	**0.960 (0.004)**

The problem appears to be complex; the articles carry latent information that is useful to distinguish between real and fake news compared to comments. We should mention that dEFEND is an explainable fake news detection model that indicates the article sentences and comments that lead to a specific prediction. However, in this work, we focus on the performance of the models; we believe that appending MultiLayerET to dEFEND will enhances dEFEND's performance while maintaining its explainability power. We leave this for future work.

6 CONCLUSION

We propose a novel system, MultiLayerET, to create a unified representation of entities and topics in online news using multilayer graphs. The layers of the graph are the articles and comments, respectively. This study is the first to consider the comments representing topics and entities and analyze them from a multilayer perspective. Our proposed system encodes novel interactions between articles and comments, which proves beneficial to downstream tasks. MultiLayerET is not limited to online news articles and their comments; it can be applied to many areas such as blogs and their comments, research papers, and discussion such as in Twitter. To characterize the capabilities of our proposed system on real applications, we provided a detailed analysis of MultiLayerET on six representative online news sources. We showed how MultiLayerET assisted in highlighting significant events and their associated entities to better understand and extract information from large-scale online news. We applied MultiLayerET to two downstream tasks. The results obtained on the media bias classification showed that MultiLayerET enhanced the textual representation and helped in better understanding the bias across sources. We also showed that MultiLayerET outperforms a state-of-the-art fake news detection model that considers both articles and comments. In the future, we will focus on expanding the dictionary of entities to include organizations, locations, subjects, materials, and other entities. We also plan to study the multilayered graph from a temporal aspect.

Acknowledgements

This research was supported in part by the U.S. NSF awards 2026513 and 1838145, and the ARL subaward 555080-78055 under Prime Contract No. W911NF2220001 and Temple University office of the Vice President for Research 2022 Catalytic Collaborative Research Initiative Program. AI & ML Focus Area. In addition, this research includes calculations carried out on HPC resources supported in part by the U.S. NSF through major research instrumentation grant number 1625061 and by the U.S. Army Research Laboratory under contract number W911NF-16-2-0189.

References

1. Blei, D., Ng, A. & Jordan, M. Latent Dirichlet Allocation. *Journal Of Machine Learning Research.* **3** (2003)
2. He, L., Han, C., Mukherjee, A., Obradovic, Z. & Dragut, E. On the Dynamics of User Engagement in News Comment Media. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov..* **10** (2020)
3. Röder, M., Both, A. & Hinneburg, A. Exploring the Space of Topic Coherence Measures. *WSDM.* (2015)
4. Watts, D. & Strogatz, S. Collective Dynamics of 'Small-World' Networks. *Nature.* **393** (1998)

5. Newman, D., Chemudugunta, C., Smyth, P. & Steyvers, M. Analyzing Entities and Topics in News Articles using Statistical Topic Models. *ISI*. (2006)
6. Spitz, A. & Gertz, M. Exploring Entity-Centric Networks in Entangled News Streams. *TheWebConf*. (2018)
7. Spitz, A. & Gertz, M. Entity-Centric Topic Extraction and Exploration: A Network-Based Approach. *ECIR*. (2018)
8. Spitz, A., Almasian, S. & Gertz, M. TopExNet: Entity-Centric Network Topic Exploration in News Streams. *WSDM*. (2019)
9. Wu, C., Kanoulas, E. & Rijke, M. Learning Entity-Centric Document Representations using an Entity Facet Topic Model. *Inf. Process. Manage.*. **57** (2020)
10. Kim, H., Sun, Y., Hockenmaier, J. & Han, J. ETM: Entity Topic Models for Mining Documents Associated with Entities. *ICDM*. (2012)
11. Ramage, D., Hall, D., Nallapati, R. & Manning, C. Labeled LDA: A Supervised Topic Model for Credit Attribution in Multi-labeled Corpora. *EMNLP*. (2009)
12. Liu, Y., Niculescu-Mizil, A. & Gryc, W. Topic-Link LDA: Joint Models of Topic and Author Community. *ICML*. (2009)
13. Hofmann, T. Probabilistic Latent Semantic Analysis. *UAI*. (1999)
14. Wang, X. & Grimson, E. Spatial Latent Dirichlet Allocation. *NeurIPS*. **20** (2008)
15. Wu, C., Kanoulas, E. & Rijke, M. It All Starts with Entities: A Salient Entity Topic Model. *Nat. Lang. Eng.*. **26** (2020)
16. Kim, H., El-Kishky, A., Ren, X. & Han, J. Mining News Events from Comparable News Corpora: A Multi-Attribute Proximity Network Modeling Approach. *IEEE BigData*. (2019)
17. Grover, A. & Leskovec, J. Node2vec: Scalable Feature Learning for Networks. *SIGKDD*. (2016)
18. Shu, K., Cui, L., Wang, S., Lee, D. & Liu, H. DEFEND: Explainable Fake News Detection. *SIGKDD*. (2019)
19. Shu, K., Sliva, A., Wang, S., Tang, J. & Liu, H. Fake News Detection on Social Media: A Data Mining Perspective. *ACM SIGKDD*. **19** (2017)
20. Shu, K., Mahudeswaran, D., Wang, S., Lee, D. & Liu, H. Fakenewsnet: A Data Repository with News Content, Social Context, and Spatiotemporal Information for Studying Fake News on Social Media. *Big Data*. **8** (2020)
21. Tatar, A., Leguay, J., Antoniadis, P., Limbourg, A., Amorim, M. & Fdida, S. Predicting the Popularity of Online Articles Based on User Comments. *WIMS*. (2011)
22. Yigit-Sert, S., Altingovde, I. & Ulusoy, Ö. Towards Detecting Media Bias by Utilizing User Comments. *WebSci*. (2016)
23. Rizos, G., Papadopoulos, S. & Kompatsiaris, Y. Predicting News Popularity by Mining Online Discussions. *TheWebConf*. (2016)
24. Tsagkias, M., Weerkamp, W. & Rijke, M. News Comments:Exploring, Modeling, and Online Prediction. *ECIR*. (2010)
25. Lee, E. That's Not the Way It Is: How User-Generated Comments on the News Affect Perceived Media Bias. *Journal Of Computer-Mediated Comm.*. **18** (2012)
26. Yanagi, Y., Orihara, R., Sei, Y., Tahara, Y. & Ohsuga, A. Fake News Detection with Generated Comments for News Articles. *INES*. (2020)
27. Devlin, J., Chang, M., Lee, K. & Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *NAACL*. (2019)
28. Reimers, N. & Gurevych, I. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. *EMNLP*. (2019)
29. Leban, G., Fortuna, B., Brank, J. & Grobelnik, M. Event Registry: Learning about World Events from News. *TheWebConf*. (2014)

30. Watanabe, K., Ochi, M., Okabe, M. & Onai, R. Jasmine: A Real-Time Local-Event Detection System Based on Geolocation Information Propagated to Microblogs. *CIKM*. (2011)
31. Sankaranarayanan, J., Samet, H., Teitler, B., Lieberman, M. & Sperling, J. TwitterStand: News in Tweets. *GIS*. (2009)
32. Panagiotou, N., Saravanou, A. & Gunopulos, D. News Monitor: A Framework for Exploring News in Real-Time. *Data*. **7** (2022)
33. Saravanou, A., Stefanoni, G. & Meij, E. Identifying Notable News Stories. *ECIR*. (2020)
34. Mathioudakis, M. & Koudas, N. TwitterMonitor: Trend Detection over the Twitter Stream. *SIGMOD*. (2010)
35. Syed, M., Wang, D., Jiang, M., Conway, O., Juneja, V., Subramanian, S. & Chawla, N. Unified Representation of Twitter and Online News using Graph and Entities. *Frontiers In Big Data*. **4** (2021)
36. Barabási, A. Network science. *Philosophical Transactions Of The Royal Society A: Mathematical, Physical And Engineering Sciences*. **371** (2013)
37. Trevisiol, M., Aiello, L., Schifanella, R. & Jaimes, A. Cold-Start News Recommendation with Domain-Dependent Browse Graph. *RecSys*. (2014)
38. Bach, N., Hai, N. & Phuong, T. Personalized recommendation of stories for commenting in forum-based social media. *Information Sciences*. **352-353** (2016)
39. Li, Q., Wang, J., Chen, Y. & Lin, Z. User Comments for News Recommendation in Forum-Based Social Media. *Inf. Sci.*. **180** (2010)
40. Guo, W., Li, H., Ji, H. & Diab, M. Linking Tweets to News: A Framework to Enrich Short Text Data in Social Media. *ACL*. (2013)
41. Wei, Z. & Gao, W. Gibberish, Assistant, or Master? using Tweets Linking to News for Extractive Single-Document Summarization. *SIGIR*. (2015)
42. Li, M., Wang, J., Tong, W., Yu, H., Ma, X., Chen, Y., Cai, H. & Han, J. EKNOT: Event Knowledge from News and Opinions in Twitter. *AAAI*. (2016)
43. Le, Q. & Mikolov, T. Distributed Representations of Sentences and Documents. *ICML*. **32** (2014)
44. Reimers, N. & Gurevych, I. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. *EMNLP*. (2019)
45. Loper, E. & Bird, S. NLTK: The Natural Language Toolkit. *ACL*. (2004)
46. Stanojevic, M., Alshehri, J., Dragut, E. & Obradovic, Z. Biased News Data Influence on Classifying Social Media Posts. *NewsIR@ SIGIR*. (2019)
47. Stanojevic, M., Alshehri, J. & Obradovic, Z. Surveying Public Opinion using Label Prediction on Social Media Data. *ASONAM*. (2019)
48. Alshehri, J., Stanojevic, M., Dragut, E. & Obradovic, Z. Stay on Topic, Please: Aligning User Comments to the Content of a News Article. *ECIR*. (2021)
49. Yang, F., Dragut, E., & Mukherjee, A. Predicting Personal Opinion on Future Events with Fingerprints. *COLING*. (2020)
50. Yang, F., Dragut, E., & Mukherjee, A. Claim Verification under Positive Unlabeled Learning. *ASONAM*. (2020)
51. Yang, F., Dragut, E., & Mukherjee, A. Improving Evidence Retrieval with Claim-Evidence Entailment. *RANLP*. (2021)
52. He, L., Shen, C., Mukherjee, A., Vucetic, S., & Dragut, E. Cannot Predict Comment Volume of a News Article before (a few) Users Read It. *ICWSM*. (2021)
53. Hosseinia, M., Dragut, E., Boumber, D., Mukherjee, A. On the Usefulness of Personality Traits in Opinion-oriented Tasks. *RANLP*. (2021)
54. Tumarada, K., Zhang, Y., Yang, F., Dragut, E., Gnawali, O., Mukherjee, A. Opinion Prediction with User Fingerprinting. *arXiv*. (2021)