A Background

A.1 Hyperbolic Space Models

Hyperbolic space has been studied under five isometric models [18]. In this section, we discuss the *Poincaré ball* and *Lorentz* models, which are utilized in the source-post encoder. Hyphen relies on the *Poincaré ball* model.

Poincaré ball model. The Poincaré ball $\left(\mathbb{B}_c^d, g^{\mathbb{B}}\right)$ of radius $1/\sqrt{|c|}$, equipped with Riemannian metric $g^{\mathcal{B}}$ and constant negative curvature c $(c < 0)$, is a d-dimensional manifold $\mathbb{B}_c^d = \{\mathbf{x} \in \mathbb{R}^d : c\|\mathbf{x}\|^2 < -1\}$, where $g^{\mathcal{B}}$ is *conformal* to the Euclidean metric $g^{\mathcal{E}} = \mathbf{I}_d$ with *conformal* factor $\lambda_{\mathbf{x}}^c = 2/(1 + c\|\mathbf{x}\|^2)$. The distance between two points $\mathbf{x}, \mathbf{y} \in \mathbb{B}_c^d$ is measured along a *geodesic* and is given by $d_{\mathcal{B}}^c(\mathbf{x}, \mathbf{y}) = (2/\sqrt{|c|}) \tanh^{-1}(\sqrt{|c|}\|-\mathbf{x} \oplus_c \mathbf{y}\|)$.

Lorentz model. With constant negative curvature c $(c < 0)$, and equipped with Riemannian metric $g^{\mathcal{L}}$, the Lorentz model $\left(\mathbb{L}_c^d, g^{\mathbb{L}}\right)$ is the Riemannian manifold $\mathbb{L}_c^d = \{\mathbf{x} \in \mathbb{R}^{d+1} : \langle \mathbf{x}, \mathbf{x}\rangle_{\mathcal{L}} = 1/c\}$, where $g^{\mathcal{L}} = \text{diag}([-1, 1, ..., 1])_n$ .The distance between two points $\mathbf{x}, \mathbf{y} \in \mathbb{L}_c^d$ is given by $d_{\mathcal{L}}^c(\mathbf{x}, \mathbf{y}) = (1/\sqrt{|c|}) \cosh^{-1}(c\langle \mathbf{x}, \mathbf{y}\rangle_{\mathcal{L}})$, where $\langle \mathbf{x}, \mathbf{y}\rangle_{\mathcal{L}}$ is the Lorentzian inner product.

Klein model. With constant negative curvature c $(c < 0)$, the Klein model is also the Riemannian manifold $\mathbb{K}_c^d = \{\mathbf{x} \in \mathbb{R}^d : c\|\mathbf{x}\|^2 < -1\}$. The isomorphism between the Klein model and Poincaré ball can be defined through a projection on or from the hemisphere model.

A.2 Hyperbolic Graph Convolutional Network (HGCN)

Given a graph $\mathcal{G} = (\mathcal{V}, E)$ and Euclidean input features $(\mathbf{x}_i^{\mathcal{E}})_{i\in\mathcal{V}}$, HGCN [40] can be interpreted as transforming and aggregating neighbours' embeddings in the tangent space of the center node and projecting the result to a hyperbolic space with a different curvature, at each stacked layer. Suppose that $\mathbf{x}_i$ aggregates information from its neighbors $(\mathbf{x}_j)_{j\in\mathcal{N}(i)}$. Further, we use $\mathcal{H}$ in superscript to denote the hyperbolic manifold. Then, the aggregation operation AGG can be formulated as, $\text{AGG}^K(\mathbf{x}^{\mathcal{H}})_i = \exp_{\mathbf{x}_i^{\mathcal{H}}}^K\left(\sum_{j\in\mathcal{N}(i)} w_{ij} \log_{\mathbf{x}_i^{\mathcal{H}}}^K(\mathbf{x}_j^{\mathcal{H}})\right)$, where, $w_{ij} = \text{SOFTMAX}_{j\in\mathcal{N}(i)}(\text{MLP}(\log_{\mathbf{o}}^K(\mathbf{x}_i^{\mathcal{H}})||\log_{\mathbf{o}}^K(\mathbf{x}_j^{\mathcal{H}})))$. More precisely, HGCN applies the Euclidean non-linear activation in $\mathcal{T}_{\mathbf{o}}\mathbb{H}^{d,K_{\ell-1}}$ and then maps back to $\mathbb{H}^{d,K_\ell}$, as $\sigma^{\otimes^{K_{\ell-1},K_\ell}}(\mathbf{x}^{\mathcal{H}}) = \exp_{\mathbf{o}}^{K_\ell}(\sigma(\log_{\mathbf{o}}^{K_{\ell-1}}(\mathbf{x}^{\mathcal{H}})))$. Therefore, the message passing in a HGCN layer can be shown as, $\mathbf{x}_i^{\ell,\mathcal{H}} = \sigma^{\otimes_{K_{\ell-1},K_\ell}}(\text{AGG}^{K_{\ell-1}}((W^\ell \otimes_{K_{\ell-1}} \mathbf{x}_i^{\ell-1,\mathcal{H}}) \oplus_{K_{\ell-1}} \mathbf{b}^\ell))$, where $-1/K_{\ell-1}$ and $-1/K_\ell$ are the hyperbolic curvatures at layer $\ell - 1$ and ℓ, respectively. As a final step, we can use the hyperbolic node embeddings at the last layer $(\mathbf{x}_i^{L,\mathcal{H}})_{i\in\mathcal{V}}$ for downstream tasks.

A.3 Hyperbolic Hierarchical Attention Network (HyperHAN)

HyperHAN [41] learns the source document representation through a hierarchical attention network in the hyperbolic space. Consider the input embedding of the t^{th} word appearing in the i^{th} sentence as $\mathbf{x}_{it}$, in the candidate document. The Euclidean hidden state of $\mathbf{x}_{it}$ within the sentence can be constructed using forward and backward Euclidean-GRU layers as: $\mathbf{h}_{it}^{\mathcal{E}} = [\overrightarrow{GRU}(\mathbf{x}_{it}), \overleftarrow{GRU}(\mathbf{x}_{it})]$. We denote the Klein and Lorentz models using $\mathcal{K}$ and $\mathcal{L}$ in superscript, respectively. Zhang and Gao [41] aim to jointly learn a hyperbolic word centroid $\mathbf{c}_w^{\mathcal{L}}$ from all the training documents. $\mathbf{c}_w^{\mathcal{L}}$ can be considered as a baseline for measuring the importance of hyperbolic words based on their mutual distance. To learn $\mathbf{c}_w^{\mathcal{L}}$, they consider another layer upon hidden state $\mathbf{h}_{it}^{\mathcal{E}}$ as: $\mathbf{h}_{it}^{\mathcal{E}'} = tanh(\mathbf{W}_w \mathbf{h}_{it}^{\mathcal{E}} + \mathbf{b}_w)$. The next step is activating $\mathbf{h}_{it}^{\mathcal{E}'}$ as $\mathbf{h}_{it}^{\mathcal{L}'}$. The word-level attention weights are then computed as α_{it} by: $\alpha_{it} = \exp(-\beta_w d_{\mathcal{L}}(\mathbf{c}_w^{\mathcal{L}}, \mathbf{h}_{it}^{\mathcal{L}'}) - c_w)$. After capturing the hyperbolic attention weights, the semantic meaning of words appearing in the same sentences is aggregated via Einstein midpoint: $\mathbf{s}_i^{\mathcal{K}w} = \sum_t \left[\frac{\alpha_{it}\gamma(\mathbf{h}_{it}^{\mathcal{K}})}{\sum_l \alpha_{il}\gamma(\mathbf{h}_{it}^{\mathcal{K}})}\right]$, where, $\gamma\left(\mathbf{h}_{it}^{\mathcal{K}}\right) = \frac{1}{\sqrt{1-||\mathbf{h}_{it}^{\mathcal{K}}||^2}} = \frac{1}{\sqrt{1-\frac{\sinh^2(r_{it})}{\cosh^2(r_{it})}}}$, $\gamma\left(\mathbf{h}_{it}^{\mathcal{K}}\right)$ is the so-called Lorentz factor, and $\mathbf{s}_i^{\mathcal{K}w}$ is the learned representation for the ith sentence. Similar to the word-level encoder, *Mobiüs*-GRU units are utilized with aggregation using Einstein midpoint to encode each sentence in the source post, yielding the final document level representation.

B Experiments

Dataset preparation. In this section, we list out the dataset collection and augmentation procedure. Since we need both source-post text and public discourse information, we augment all the datasets to yield sufficient comments per social media post. **Politifact** and **Gossipcop** [43] were collected from two fact-verification platforms PolitiFact and GossipCop, and contain news content with two labels (fake or real) and social context information. After scraping the tweets corresponding to the news articles in the datasets, we get 837 and 19266 news articles (Politifact and Gossipcop respectively) which have atleast 1 tweet available. Finally, we filter the news articles with atleast 3 comments which gives us datasets with 415 and 2813 news articles (source posts) for Politifact and Gossipcop respectively. **AntiVax** [44] is a novel Twitter dataset for COVID-19 vaccine misinformation detection, with more than 15,000 tweets annotated as fake or not. We manually scrape the user comments corresponding to the tweets present in the dataset, which resulted in 3797 tweets (2865 real and 932 fake) with atleast one user comment. **HASOC** [45] was taken from the HASOC 2019 sub-task B with over 3000 tweets (comments and replies) from 82 conversation threads labelled as hate speech or not. Due to the available labels, we consider the top level comments on the 82 conversation threads as separate tweets and the corresponding replies as the public discourse. This yields a dataset with 712 tweets with public discourse (in contrast to the original 82 conversation threads). **Pheme** [46] is a collection of 6425 Twitter rumours and non-rumours conversation threads related to 9 events and each of the samples is annotated as either True, False or Unverified. **RumourEval** was introduced in SemEval-2019 Task 7, and has 446 twitter and reddit posts belonging to three categories: *real*, *fake* and *unverified* rumour. [48]. **Twitter15** and **Twitter16** [47] consist of source tweets (1490 and 818 resp.) along with the sequence of re-tweet users. We choose only *true* and *fake* rumour labels as the ground truth. Since there is no discourse available, we scrape the user comments corresponding to the posts and filter the tweets with atleast one comment giving us 543 and 362 source-tweets respectively. **FigLang (Twitter)** and **FigLang (Reddit)** [49] are FigLang 2020 shared task datasets with 4400 samples each labelled as either sarcasm or not.

Experimentation details. We adopt a pre-trained AMR parser from the AMRLib[3] library and use the `parse_xfm_bart_base` model to generate the comment-level AMRs. We resolve co-references on the comment-level AMRs using an off-the-shelf model AMRCoref[4], which yields the various co-reference clusters. Finally, we convert all the merged AMRs to the Deep Graph Library[5] (DGL) format. All node embeddings for AMR are initialised using 100D Glove embeddings[6]. Data-specific hyperparameters have been laid out in Table B.

Data-specific baselines. We experiment with three different baselines per dataset. We compare the model performance using F1 Score, Precision and Recall. These models are known to have reported representative results on the benchmark datasets. For *Fake news detection*, (v) **TCNN-URG** [27] utilises a CNN-based network for encoding news content, and a variation auto-encoder (VAE) for modelling the user comments (vi) **CSI** [28] is a hybrid deep learning model that utilizes subtle clues from text, responses, and source post, while modelling the news representation using an LSTM-based network, for fake news detection, and lastly (vii) **HPA-BLSTM** [50] learns news representations through a word-level, post-level, and event-level user engagements on social media. These turned out to have representative performance for fake news detection and therefore, we consider them as baselines for Politifact, Gossipcop, and AntiVax datasets (which are related to the task of fake news detection). In addition to CSI and HPA-BLSTM

Dataset	Euclidean		Hyperbolic		Max sents	Max coms
	lr	Batch size	lr	Batch size		
Politifact	1e-3	16	1e-2	16	30	10
Gossipcop	2e-3	64	2e-3	64	50	10
ANTiVax	1e-4	64	1e-2	32	2	8
HASOC	1e-4	16	1e-3	32	2	9
Pheme	1e-3	64	1e-2	32	2	17
Twitter15	1e-4	32	1e-2	32	2	8
Twitter16	1e-4	32	1e-3	32	2	20
RumourEval	1e-4	16	1e-2	32	2	3
FigLang Twitter	1e-3	32	1e-2	32	2	3
FigLang Reddit	1e-4	64	1e-2	32	2	2

Table 5: Data-specific hyperparameters for `Hyphen`. *lr*: learning rate, *max sents*: Max. sentences considered in a source post while training, *max coms*: Max. comments on post considered while training.

[3] `https://amrlib.readthedocs.io/en/latest/`

[4] `https://github.com/bjascob/amr_coref`

[5] `https://www.dgl.ai/`

[6] `https://nlp.stanford.edu/projects/glove/`

as baselines for *Hate speech detection* on the HASOC dataset, we use **CRNN** as a baseline due to the ability of CNNs to capture the sequential correlation in text. In *rumour detection*, for Twitter15 and Twitter16 datasets, (v) **AARD** [52] uses a weighted-edge transformer-graph network and position-aware adversarial response generator to capture the malicious user attacks while spreading rumours, (vi) **GCAN** [9] employs a dual co-attention mechanism between source social media post and the underlying propagation patterns, and (vii) **BiGCN** [51], utilizes the original graph structure information and the latent correlation between features assisted by bidirectional-filtering. Further, for Pheme dataset we use (v) **RumourGAN** [54] which adheres to a GAN-based approach, where the generator is designed to produce uncertain or conflicting opinions (voices), complicating the original conversational threads in order to penalise the discriminator to learn better, (vi) **DDGCN** [53], which models spatial and temporal features of human actions from their skeletal representations, and (vii) **STS-NN** [55]. On RumourEval, (v) **DeClarE** [56] provides a strong baseline. Moreover simple yet effective models like (vi) **CNN** and (vi) **MTL-LSTM** show comparable performance, and hence are included in our set of baselines. For the task of (d) *Sarcasm detection* on Figlang (Twitter) and Figlang (Reddit) datasets, we use (v) **CNN + LSTM** [58], (vi) an ensemble of CNN, LSTM, SVM and MLP [59], and lastly (vii) **C-Net** [60] for efficient sarcasm classification.

C Explainability

Data annotation. To evaluate the efficacy of `Hyphen` in producing suitable explanations, we fact-check and annotate the Politifact dataset on a sentence-level. Each sentence has the following possible labels – *true*, *false*, *quote*, *unverified*, *non_check_worthy* or *noise*. The annotators were further supposed to arrange the fact-checked sentences in the order of their check-worthiness. We take the help of four expert annotators in the age-group of 25-30 years. The final labels for a sentence were decided on the basis of majority voting amongst the four annotators. To decide the final rank-list (since different annotators might have different opinions about the level of check-worthiness of the sentences), the fourth annotator compiled the final rank-list by referring to the fact-checked rank-lists by the first three annotators using Kendall's τ and Spearman's ρ rank correlation coefficients, and manually observing the similarities between the three rank-lists. The compiled list is then cross-checked and re-evaluated by the first three annotators for consistency.

Explainability evaluation. To evaluate the performance of `Hyphen` against the annotated rank-list, we measure the rank-correlation between the two. If `Hyphen` predicts a news article in Politifact to be fake, we filter the sentences in the ground-truth annotation with the label *fake* (in the order of their check-worthiness). We adopt a similar procedure in case `Hyphen` predicts a news article to be true. This is done because if a news article is fake, we aim to identify the sentences in the article which are misinformation and thus most relevant to the final prediction. Finally, we compare the filtered ground-truth rank-list with the rank-list produced by `Hyphen` using Kendall's τ and Spearman's ρ coefficients. Figure 3 shows sample rank-lists produced by `Hyphen-hyperbolic` and dEFEND [33].

Sentence
Federal Judge Peter J. Messitte has just ruled in favor of two attorney generals seeking to subpoena the Trump organization relating to President Trump unlawfully receiving emoluments from foreign and domestic governments.D.C.
Attorney General Karl A. Racine and Maryland Attorney General Brian E. Frosh can now subpoena the Trump organization, thereby forcing them to preserve documents in relation to President Trump's alleged indiscretions.
"The Justice Department had sought to squash the subpoena earlier in September, but Judge Messitte wasn't convinced with their argument.
The case advances a very high-profile attempt to see if President Trump is violating the emoluments clause of the U.S. Constitution, which precludes him from receiving gifts from foreign or state governments.
Per the Post:Because Trump continues to benefit financially from his hotel, resort and golf properties — in some cases from clients affiliated with foreign governments — Frosh and Racine alleged in their June complaint that Trump had committed "unprecedented constitutional violations."
State spending that benefits the president may be considered a violation of the domestic emolument clause, which says the president "shall not receive" any emolument, other than fixed compensation, from "the United States, or any of them.
President Trump has been accused of profiting from the presidency, and this case will seek to prove that assertion.
The Trump Organization will be compelled to comply with the court's ruling."This ruling is an important first step in our litigation against President Trump for unlawfully receiving emoluments from foreign and domestic governments," Racine said in a statement.

(a) Fact-checked and sentence-level annotated rank-list for `politifact14810`

dEFEND
Per the Post:Because Trump continues to benefit financially from his hotel, resort and golf properties — in some cases from clients affiliated with foreign governments — Frosh and Racine alleged in their June complaint that Trump had committed "unprecedented constitutional violations."
The Trump Organization will be compelled to comply with the court's ruling."This ruling is an important first step in our litigation against President Trump for unlawfully receiving emoluments from foreign and domestic governments," Racine said in a statement.
"The Justice Department had sought to squash the subpoena earlier in September, but Judge Messitte wasn't convinced with their argument.
State spending that benefits the president may be considered a violation of the domestic emolument clause, which says the president "shall not receive" any emolument, other than fixed compensation, from "the United States, or any of them.
The case advances a very high-profile attempt to see if President Trump is violating the emoluments clause of the U.S. Constitution, which precludes him from receiving gifts from foreign or state governments.
Attorney General Karl A. Racine and Maryland Attorney General Brian E. Frosh can now subpoena the Trump organization, thereby forcing them to preserve documents in relation to President Trump's alleged indiscretions.
Federal Judge Peter J. Messitte has just ruled in favor of two attorney generals seeking to subpoena the Trump organization relating to President Trump unlawfully receiving emoluments from foreign and domestic governments.D.C.
President Trump has been accused of profiting from the presidency, and this case will seek to prove that assertion.

(b) Rank-list generated by dEFEND

`Hyphen`
Federal Judge Peter J. Messitte has just ruled in favor of two attorney generals seeking to subpoena the Trump organization relating to President Trump unlawfully receiving emoluments from foreign and domestic governments.D.C.
Attorney General Karl A. Racine and Maryland Attorney General Brian E. Frosh can now subpoena the Trump organization, thereby forcing them to preserve documents in relation to President Trump's alleged indiscretions.
The case advances a very high-profile attempt to see if President Trump is violating the emoluments clause of the U.S. Constitution, which precludes him from receiving gifts from foreign or state governments.
The Trump Organization will be compelled to comply with the court's ruling."This ruling is an important first step in our litigation against President Trump for unlawfully receiving emoluments from foreign and domestic governments," Racine said in a statement.
Per the Post:Because Trump continues to benefit financially from his hotel, resort and golf properties — in some cases from clients affiliated with foreign governments — Frosh and Racine alleged in their June complaint that Trump had committed "unprecedented constitutional violations."
State spending that benefits the president may be considered a violation of the domestic emolument clause, which says the president "shall not receive" any emolument, other than fixed compensation, from "the United States, or any of them.
"The Justice Department had sought to squash the subpoena earlier in September, but Judge Messitte wasn't convinced with their argument.
President Trump has been accused of profiting from the presidency, and this case will seek to prove that assertion.

(c) Rank-list generated by `Hyphen`

Figure 3: Sample rank-lists generated by `Hyphen-hyperbolic` and dEFEND. (a) Ground-truth annotation for `politifact14810` sample. *Red*: fake sentences, *Green*: true sentences, and *Yellow*: quote. It can be observed that there is almost no correlation between the dEFEND rank-list and the ground-truth. The rank-list produced by `Hyphen` is observably quite similar to the annotated list.