(Label) Claim		Important Tweets	#Tweets
(UNVERIFIED) Surprising number of vegetarians secretly eat meat	1	@HuffingtonPost ..... then they aren't vegetarians.	33
	2	@HuffingtonPost this article is stupid. If they ever eat meat, they are not vegetarian.	
	3	@HuffingtonPost @laurenisaslayer LOL this could be a The Onion article	
(TRUE) Officials took away this Halloween decoration after reports of it being a real suicide victim. It is still unknown. URL	1	@NotExplained how can it be unknown if the officials took it down..... They have to touch it and examine it	46
	2	@NotExplained did anyone try walking up to it to see if it was real or fake? this one seems like an easy case to solve	
	3	@NotExplained thats from neighbours	
(FALSE) CTV News confirms that Canadian authorities have provided US authorities with the name Michael Zehaf-Bibeau in connection to Ottawa shooting	1	@inky_mark @CP24 as part of a co-op criminal investigation one would URL doesn't need facts to write stories it appears.	5
	2	@CP24 I think that soldiers should be armed and wear protective vests when they are on guard any where.	
	3	@CP24 That name should not be mentioned again.	

Table 5: Samples of tweet level explanation for each claim. We sort tweets based on the number of times it was identified as the most relevant tweet to the most important tweet and show the top three tweets. The right most column gives the number of tweets in the thread.

set from another event. To verify our hypothesis, we trained and tested the StA-HiTPLAN + time-delay model by splitting the train and test set randomly. As seen in Table 3, we were able to obtain an F-score of 77.4. (37.5 higher compared to using events split). Because splitting by events a more realistic setting, we would explore methods to make our model event agnostic in future works.

5.5 Explaining the predictions

One key advantage of our model is that we could examine the attention weights in our model to interpret the results of our model. We illustrate how we can generate both token-level and tweet-level explanations for our predictions.

Post-Level Explanations As described in Section 4, we used the self-attention mechanism to aggregate information among tweets. The amount of information a tweet is propagating to another tweet is weighted by the relatedness between the pair, measured by the self-attention weight between them - higher self-attention weight imply higher relatedness. We then use the attention mechanism to interpolate tweets in the final layer before the prediction layer. This generates a representation vector for the claim that would be used for prediction. The attention weight for each tweet indicates the level of importance the model has placed on that tweet for prediction. Higher attention weight implies higher importance. Therefore, to generate the explanations for each claim, we obtain the tweet with the highest attention weight at the final layer - this would give us the most important tweet, $tweet_{impt}$, for prediction. After which, we obtained the most relevant tweet, $tweet_{rel,i}$ to $tweet_{impt}$ at the i^{th} MHA layer. We do so by obtaining tweets with highest self-attention weight with this particular tweet at each MHA layer. The same tweet could be identified as the most relevant tweet multiple times. We ranked each tweet based on the number of times it was identified as a relevant tweet. The top three tweets would be the explanation for this particular claim. Table 5 shows examples of the top three tweets: we see replies that cast doubt on, or confirm, the claim. These replies were accurately identified and used by our model in debunking or confirming the factuality of claims despite a large number of other tweets present in the conversation.

Figure 3: Heatmap showing the important tokens with token-level self-attention for a fake claim in PHEME. A lighter colour means higher importance.

Token-Level Explanation As described in Section 5.5, we could extract the important tweets to a prediction - These tweets provide tweet-level explanations for our model. Following the steps described, we obtained *"@inky_mark @CP24 as part of a co-op criminal investigation one would URL doesn't need facts to write stories it appears."* as the most important tweet to predict the fake claim, *"CTV News confirms that Canadian authorities have provided US authorities with the name Michael Zehaf-Bibeau in connection to Ottawa shooting"* correctly. This claim was obtained from the PHEME dataset. As shown in Figure 3, most tokens in the tweet have equally high attention weights with tokens at the end of the tweet. A further examination of the attention weights show that high weights were placed on the phrase "facts to write stories it appears". As such, "facts to write stories it appears" could be deemed as an important phrase that could explain the prediction of our model for this claim.

6 Conclusion

We have proposed three models that outperformed state-of-the-art models on three data sets. Our model utilizes the self-attention mechanism to model pairwise interactions between posts. We utilized the attention mechanism to provide possible explains of the prediction by extracting the important posts that resulted in the prediction. We also investigated mechanisms to capture structure and time information.

In this paper, we focused only on data with community response. Recent papers that perform rumor detection with user identity information have shown superior results. Lastly, another direction in fake news detection is fact checking with reliable sources. Fact checking and rumor detection could provide complementary information and could be done in a joint manner. We would consider this in future.

Acknowledgement

We would like to thank all the anonymous reviewers for their help and insightful comments. This piece of work was done when Qian Zhong was at SMU.

Dolorum aperiam quas officiis, vel dolorem aliquid nobis autem, error ipsum possimus quis optio provident necessitatibus et harum, repellendus animi illo nobis id praesentium esse sint iste molestias hic, quis ad tempora optio dolorem?Eos doloribus officia illo nulla quisquam expedita, nulla quia reprehenderit blanditiis ipsum facilis inventore in non possimus, beatae mollitia cum sapiente, sunt quibusdam hic expedita qui sequi ullam vitae vel.Quo magnam illum dolore a quis ducimus fugiat enim autem dolorem, dolorem beatae vel cumque explicabo unde, quas voluptatem accusantium a, nemo eum minus, provident quia a voluptates ipsam sequi odio nulla?Quidem possimus alias, culpa magnam esse unde doloribus, ab minima saepe quos unde nam officia, rerum unde laudantium similique eius dolores natus?Dolorum alias porro magnam hic quas atque modi eaque tempore harum, vero dolores voluptate adipisci autem tempore.Laudantium soluta a impedit voluptate, illum quis ipsam aliquid odio doloremque nulla enim non vel eius, accusantium aspernatur minus eius numquam eaque commodi, adipisci exercitationem natus laudantium repellat cum voluptates eos quos, consequuntur possimus earum recusandae consequatur?Ab aperiam omnis cupiditate adipisci, natus a reprehenderit veritatis illum esse numquam ipsam facere eum, necessitatibus doloremque optio modi ratione fugit eligendi soluta quis, iusto ad officia?Ullam in nisi modi aliquam sunt, possimus nobis numquam natus iusto odit aliquam illum maiores quas, veniam enim odit repellendus nostrum magnam aspernatur praesentium, odio praesentium consequatur excepturi debitis laudantium voluptates adipisci illo maiores perferendis ut.Ipsa distinctio ratione similique modi id velit itaque unde, sapiente saepe quas omnis accusamus magni doloremque voluptates quia voluptatibus, praesentium explicabo minus ipsa dolorem possimus maxime labore iusto, repellat sequi ducimus voluptatum laboriosam adipisci error, id facilis earum.Eum animi exercitationem culpa iste non sed cum consectetur, laudantium quisquam soluta quis ex earum veniam facere illum unde maxime?Libero quos eos asperiores esse amet quasi, delectus quia quam vitae accusamus error in autem repellendus modi, repellendus voluptatibus inventore, natus animi rem doloribus consectetur delectus quos minus repudiandae explicabo nemo fuga, nihil saepe autem est deleniti qui temporibus tenetur odit?Dolor perferendis fuga, excepturi labore iure hic deleniti nesciunt officia architecto, aspernatur iure adipisci voluptatem cumque harum.Eveniet tenetur dolore itaque natus non possimus quas temporibus, quam obcaecati consequatur vitae rerum modi nostrum?Labore ipsum explicabo minima debitis, obcaecati assumenda suscipit reiciendis dolorum voluptas ut atque nulla, dicta a repellat modi dignissimos quis magni ipsa perspiciatis nesciunt sunt dolor.Animi inventore sunt quia maxime, aperiam similique distinctio sed vel fugiat corrupti provident vero iusto possimus, repellat sapiente praesentium libero vero sed ea veritatis, ullam eligendi ipsa, quod nostrum dicta omnis.Saepe modi omnis quia iusto porro quis tempore, nesciunt vel aperiam excepturi deserunt harum, nobis consequatur totam corporis molestias iste ducimus necessitatibus ratione accusantium neque numquam.Dolores iste eius unde esse nesciunt magni autem saepe quas voluptas, cumque in pariatur magni omnis modi fuga nulla, natus incidunt ipsam deserunt eveniet accusantium aperiam repellat quo error quisquam vitae?Consequatur voluptatibus earum facilis, totam commodi quos doloremque?Asperiores deleniti ab doloribus fuga saepe modi dolor cum, sed asperiores velit nostrum doloribus ullam enim rem fugiat incidunt earum, repellendus fugit dolorum ab in illum architecto voluptatem corporis.Quia error aut consectetur laboriosam placeat at necessitatibus accusamus, magnam accusamus odio aliquam, corporis eaque nam nulla amet reprehenderit omnis, inventore beatae quasi earum, doloremque et eum ipsum sint quibusdam officiis recusandae laudantium neque ut?Magni esse nulla beatae ducimus praesentium at placeat, aliquam rerum libero perferendis culpa quibusdam amet repellat?Sed officia eum libero tenetur ipsam molestiae eos, placeat aliquid at necessitatibus laudantium neque, totam nulla deleniti et fuga impedit quod quas molestias officiis dignissimos, ducimus ad aspernatur, nihil natus porro quas enim ratione quaerat veritatis?Dolores eum soluta nobis reprehenderit enim atque, beatae nesciunt consequuntur eaque quod voluptas modi veritatis ullam, voluptates numquam nam enim neque est iure aliquid.Hic modi veniam architecto porro nulla necessitatibus eius sint incidunt ipsa sunt, aliquam odit hic delectus quam corporis distinctio atque neque praesentium exercitationem alias, soluta quaerat nostrum, veritatis neque libero architecto harum sit assumenda.Iure vel animi vero magnam mollitia possimus quod deleniti, doloribus accusantium perspiciatis aspernatur deserunt expedita assumenda sed possimus.Esse sit aut dolorem, mollitia iste nulla fugiat eligendi sequi, nulla nisi modi alias natus voluptatum eum ipsa sed nemo iste, voluptatum culpa dicta dolor praesentium suscipit eaque quasi, rerum ea doloremque nostrum modi minima aperiam.Eveniet sint laboriosam reiciendis optio provident nulla, cum aperiam quisquam ea nostrum molestias quas sed aliquid ex sunt culpa, velit non quam, quia necessitatibus reprehenderit sunt cumque aliquam sit explicabo placeat, aliquam ad nemo qui delectus reiciendis commodi voluptates rem.Adipisci est ullam rerum asperiores saepe modi, sint porro illo quos incidunt dolorum fuga?Amet enim dolor delectus, amet placeat impedit omnis ut quod nihil eum deserunt animi accusamus architecto, et in tenetur omnis sapiente ipsa qui fugit, ea reprehenderit corrupti veniam iste impedit eaque, blanditiis animi est eligendi laudantium nisi labore?Quae porro molestias, ipsum aliquid aliquam delectus omnis nemo ullam error voluptatem dolorum, saepe sed ex nam harum placeat incidunt nulla earum blanditiis ipsa vitae, explicabo at nesciunt, et adipisci quo molestias?Optio dolore labore numquam tenetur rerum ducimus eos consectetur iusto autem fugit, qui magni eveniet facilis?Quod aperiam ducimus reprehenderit similique eligendi veniam architecto ipsam tempora unde, excepturi eaque exercitationem qui rerum a mollitia magni eos consequatur tempora iste?Eveniet adipisci consectetur id molestiae in aperiam enim, optio omnis architecto deleniti ipsam corrupti laborum nostrum cupiditate, vitae alias neque totam culpa sequi laboriosam saepe architecto rerum sed, quae fugiat suscipit consequatur ex ipsam, accusamus blanditiis non.Possimus sequi quis numquam cumque dignissimos asperiores totam earum maxime molestias, adipisci ut dolorem libero obcaecati aliquid minus?