You Shall Know a User by the Company It Keeps: Dynamic Representations for Social Media Users in NLP

Marco Del Tredici
University of Amsterdam
`m.deltredici@uva.nl`

Diego Marcheggiani[*]
Amazon
`marchegg@amazon.es`

Sabine Schulte im Walde
University of Stuttgart
`schulte@ims.uni-stuttgart.de`

Raquel Fernández
University of Amsterdam
`raquel.fernandez@uva.nl`

1 Extended Results Table

Tables 1, 2 and 3 include, for each task, the results for LING, LING+PV, LING+N2V and LING+GAT as reported in the main paper, together with the Standard Deviation values computed on the ten runs of each model. Additionally, the precision, recall and F-score for each class are reported.

---
[*]Research conducted when the author was at the University of Amsterdam.

Model	Av. Rec.	Negative			Neutral			Positive		
		P	R	F1	P	R	F1	P	R	F1
LING	0.676 ±0.005	0.585	0.656	0.618	0.684	0.678	0.680	0.737	0.694	0.712
LING+PV	0.671 ±0.004	0.584	0.639	0.609	0.677	0.679	0.678	0.734	0.693	0.713
LING+N2V	0.672 ±0.004	0.584	0.639	0.609	0.681	0.679	0.680	0.734	0.699	0.715
LING+GAT	0.666 ±0.01	0.599	0.597	0.596	0.666	0.691	0.677	0.730	0.691	0.709

Table 1: **Sentiment Analysis**: Average Recall across the three classes, plus precision, recall and F1 per class.

Model	Av. Ag./Fa.	Against			Neutral			Favor		
		P	R	F1	P	R	F1	P	R	F1
LING	0.569 ±0.01	0.730	0.625	0.672	0.355	0.462	0.399	0.446	0.490	0.466
LING+PV	0.601* ±0.02	0.739	0.673	0.701	0.353	0.380	0.362	0.479	0.536	0.501
LING+N2V	0.629*◇ ±0.01	0.761	0.697	0.727	0.380	0.369	0.370	0.488	0.588	0.531
LING+GAT	0.640*◇† ±0.01	0.749	0.725	0.734	0.380	0.316	0.330	0.507	0.600	0.545

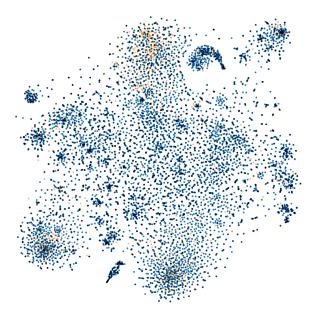Table 2: **Stance Detection**: Average F1 of the Against and Favor classes, plus precision, recall and F1 per class.

Model	F1 Hateful	Normal			Hateful		
		P	R	F1	P	R	F1
LING	0.624 ±0.01	0.968	0.989	0.978	0.773	0.526	0.624
LING+PV	0.667* ±0.02	0.974	0.983	0.979	0.730	0.621	0.667
LING+N2V	0.656* ±0.008	0.972	0.986	0.979	0.742	0.589	0.656
LING+GAT	0.674*◇† ±0.005	0.973	0.989	0.980	0.765	0.605	0.674

Table 3: **Hate Speech Detection**: F1 for the Hateful class, plus precision, recall and F1 per class.

2 Additional Plots

The plot representing the N2V representations for the Hate Speech dataset. Orange dots are authors of tweets labelled as `HATEFUL`, blue dots of `NORMAL` tweets.

Below, the same plots in Figure (2) in the main paper showed in a larger size.
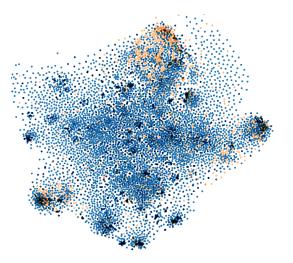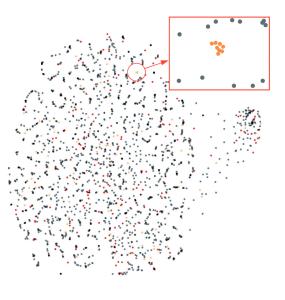

Figure 2: PV user representations for the Stance dataset. Orange dots are authors of tweets `AGAINST` atheism, red dots authors in `FAVOR` of 'climate change is a real concern'. All other users are represented as grey dots.

Figure 1: PV user representations for the Hate Speech dataset. Orange dots are authors of `HATEFUL` tweets, blue dots of `NORMAL` tweets.
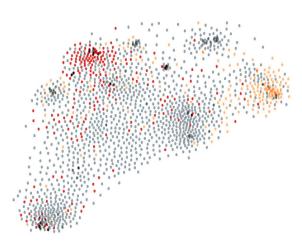

Figure 3: N2V user representations for the Stance dataset. Orange dots are authors of tweets `AGAINST` atheism, red dots authors in `FAVOR` of 'climate change is a real concern'. All other users are represented as grey dots.