Appendix

Table s1. Multilevel Logistic Regression Coefficients Predicting Retweeting in Diffusion Cascades (Similarity Measured by Hashtag versus FastText, English Only)

In addition to measuring interest similarity based on hashtags, it is also possible to calculate a more general semantic similarity between users using word embedding models. In doing so, the words in a tweet were represented by vectors in a semantic space. And the mean of the vectors could be used to represent the original tweet. Eventually, users' timelines could be represented as different vectors in the same semantic space. And then cosine similarity between the vectors was a measure of semantic similarity between the users. In particular, this study obtained the word vectors from Facebook's pre-trained fastText model (Grave et al., 2018). Only users tweeting in English were included in the multilevel model (967 cascades / 5,022 parents / 84,909 observations).

Although hashtag similarity and semantic similarity may refer to different concepts, the results as presented below are very consistent in general. Moreover, the coefficients of interaction frequency and number of exposures did not change substantively, indicating the robustness of the proposed model in the present study.

	Hashtag	fastText
Intercept	-3.45(0.05)**	-4.80(0.06)**
Interaction frequency	0.94(0.02)**	0.90(0.03)**
Number of exposures	-1.40(0.05)**	-1.42(0.05)**
Similarity	0.72(0.03)**	3.64(0.06)**
Cascade depth	-0.27(0.03)**	-1.15(0.05)**
Time elapsed	0.16 (0.02)**	0.17(0.03)**
Interaction frequency × depth	-0.26(0.02)**	-0.32(0.02)**
Number of exposures × depth	0.87(0.04)**	0.92(0.04)**
Similarity × depth	0.17(0.03)**	1.63(0.05)**
Structural Factors		
Reciprocity	-0.12(0.04)*	-0.25(0.04)**
Structural redundancy	0.05(0.02)*	0.03(0.03)
Retweeter Attributes		
Retweeting inertia	0.83(0.02)**	0.62(0.02)**
Number of followees	-0.74(0.03)**	-0.75(0.04)**
Number of followers	-0.02 (0.04)	-0.00 (0.04)
Number of statuses	1.53(0.03)**	1.42 (0.03)**
Account age	-0.28(0.02)**	-0.27(0.02)**
Parent Attributes		
Number of followees	-0.08(0.03)*	-0.02(0.03)
Number of followers	0.86(0.03)**	0.99(0.04)**
Number of statuses	-0.53(0.03)**	-0.66(0.03)**
Account age	0.11(0.02)**	0.07(0.03)**
Group Means		
Interaction (parent level)	-0.38(0.06)**	-0.16(0.06)*
Exposure (parent level)	0.33(0.06)**	0.40(0.08)**
Similarity (parent level)	-0.65(0.07)**	0.144(0.08)
Interaction (cascade level)	-0.25(0.10)*	-0.54(0.11)**
Exposure (cascade level)	0.48(0.09)**	0.51(0.10)**
Similarity (cascade level)	-0.20(0.09)*	-1.87(0.06)*
σ^2 – Residual variance	3.29	3.29

τ – Cascade level/parent level		
Intercept	0.22/0.05	0.27/0.24
Interaction frequency	0.07/0.12	0.11/0.12
Number of exposures	0.30/0.51	0.30/0.64
Interest similarity	0.22/0.05	0.30/0.00
Marginal R^2	73.8%	89.4%
Conditional R^2	75.8%	NA
AIC	39,147.85	32,139.10
Sample size	967 cascades / 5,022 parents / 84,909 observations	

Note. All predictors except reciprocity were standardized to have a mean of 0 and a standard deviation of 1. All count variables were log-transformed before standardization. **p* < .01, ***p* < .001.

Reference:

Grave, E., Bojanowski, P., Gupta, P., Joulin, A., & Mikolov, T. (2018). Learning word vectors for 157 languages. *arXiv preprint arXiv:1802.06893.*

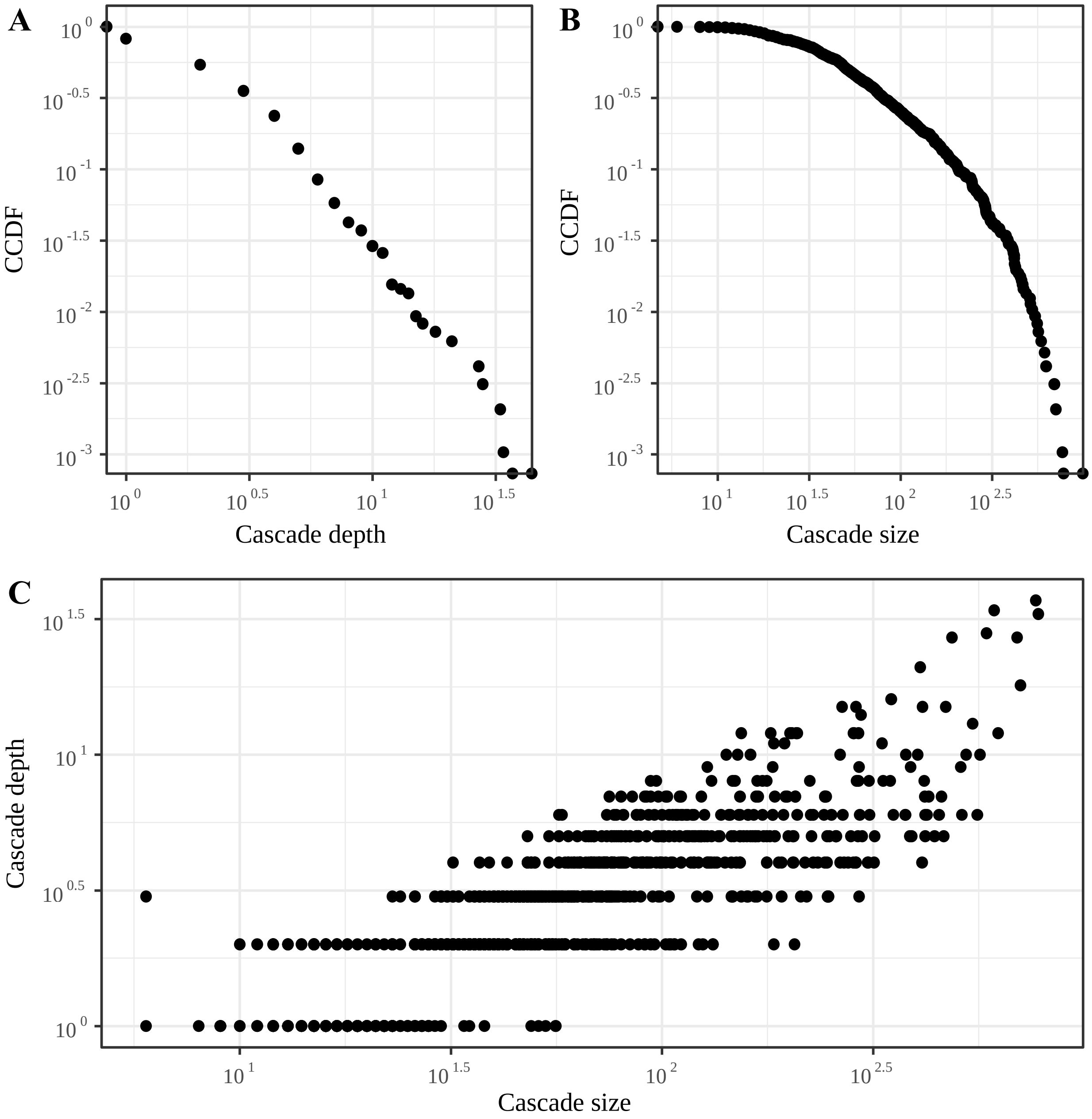

Figure s1. Complementary cumulative distribution functions (CCDF) of cascade depth (A) and cascade size (B) and the correlation between cascade size and cascade depth (C).

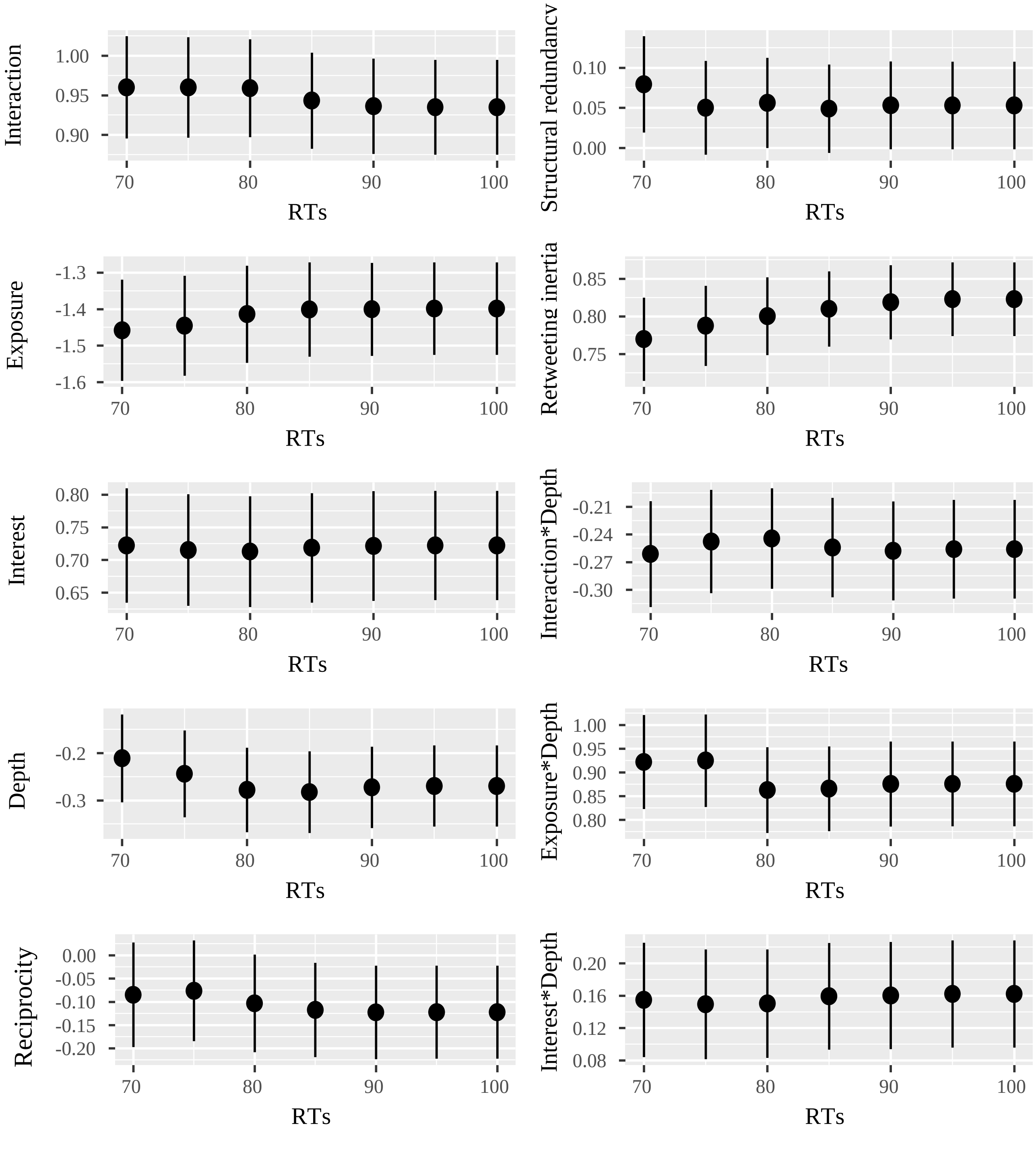

Figure s2. Coefficients of the major variables in Table 1 against the number of retweets. RTs: the number of retweets.

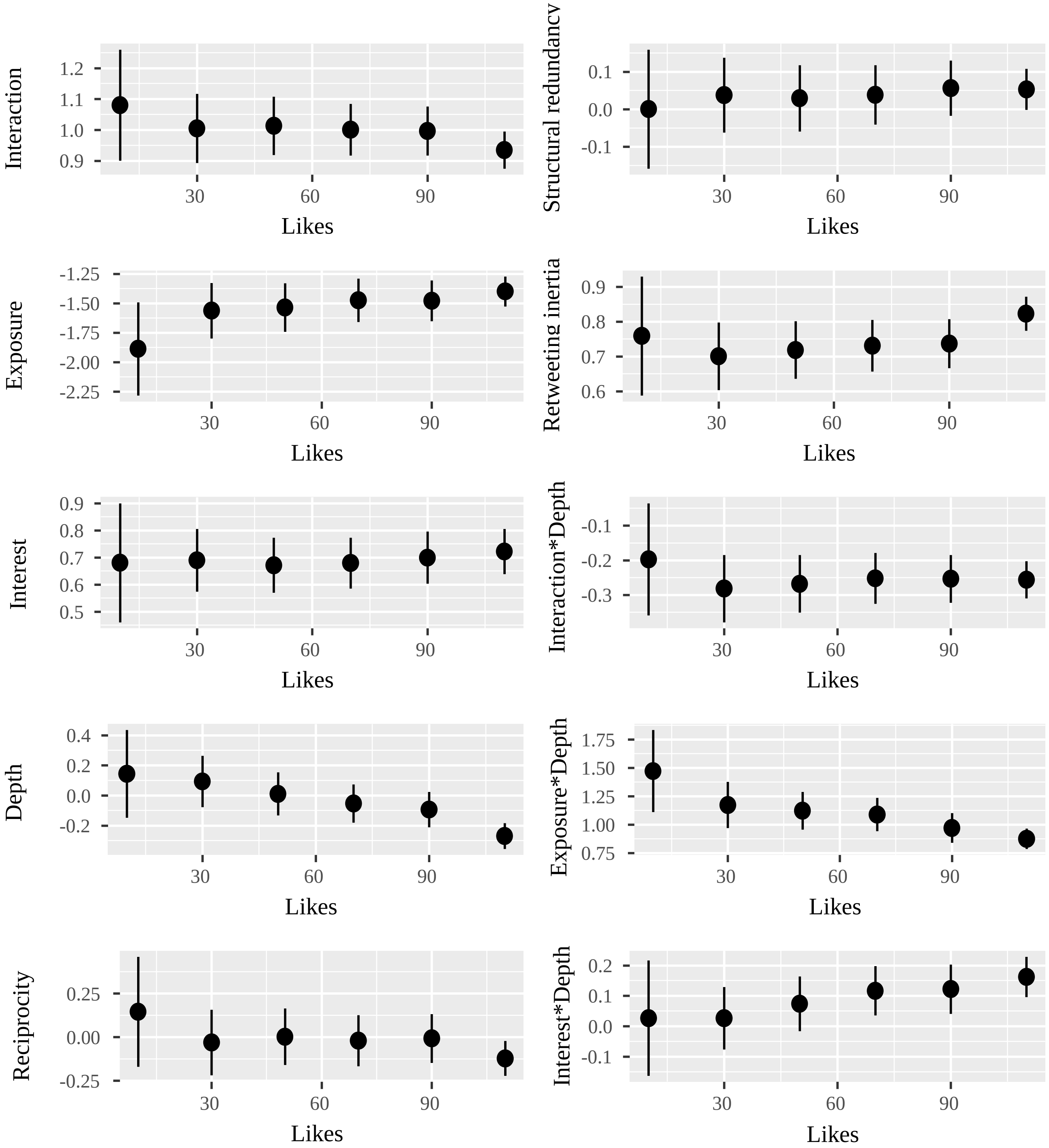

Figure s3. Coefficients of the major variables in Table 1 against the number of likes of tweets. Likes: the number of likes.