Online Appendix

Research Note: The Importance of Interactions betweenContent Characteristics and Creator Characteristics forStudying Virality in Social Media

Content

Table 1: Variable Description

Category	Variable	Description
Content Features[1]	Length	Length of the tweet. Number of characters.
	Hashtag	Whether the tweet includes hashtag. Binary.
	Image	Whether the tweet includes image. Binary.
	Video	Whether the tweet includes video. Binary.
	QuoteRT	Whether the tweet includes quote. Binary.
	Mention	Whether the tweet includes mention. Binary.
	Link	Whether the tweet includes link. Binary.
Content Message[2]	M_Positive	Number of positive words in the tweet (based on the NRC Emotion Lexicon).
	M_Negative	Number of negative words in the tweet (based on the NRC Emotion Lexicon.
	M_Profane	% of tweet that is about profanity.
	M_Humor	% of tweet that is about humor.
	M_Family	% of tweet that is about family.
	M_Pop	% of tweet that is about pop culture.
	M_Politic	% of tweet that is about politics.
	M_Sports	% of tweet that is about sports.
	M_Relation	% of tweet that is about relationships.
	M_Animal	% of tweet that is about animals.
	M_Emotion	% of tweet that is emotional.
Creator Features[3]	Age	Age of the user. Categorical. Young: 65.34% Middle Aged: 10.47% Old: 2.25% Other/Unknown: 21.94%
	TweetVolume	Number of tweets by a user previous to this tweet. Logged.
	Followers	Number of followers. Logged.

	Following	Number of following. Logged.
	Race	Race of the user. Categorical. Asian: 2.79% Black: 11.43% White: 74.84% Other: 10.95%
Creator History[4]	H_Length	Average length of tweets created by a user.
	H_Viral	Mean number of retweets a user received for previous tweets (within a month before this tweet). Logged.
	H_RT	Percentage of previous tweets that are retweets.
	H_Original	Percentage of previous tweets that are original tweets.
	H_Profane	User's propensity to include profanity in their tweets.
	H_Humor	User's propensity to include humor in their tweets.
	H_Family	User's propensity to include family in their tweets.
	H_Pop	User's propensity to include pop culture in their tweets.
	H_Politic	User's propensity to include politics in their tweets.
	H_Sports	User's propensity to include sports in their tweets.
	H_Relation	User's propensity to include relationships in their tweets.
	H_Animal	User's propensity to include animals in their tweets.
	H_Emotion	User's propensity to include emotion in their tweets.
Control Variables	Year	Year of the tweet.
	Month	Month of the tweet. 1-12.
	Day	Day of the tweet. 1-31.
	Laglog	Lag (Seconds) between this tweet and the user's previous tweet. Logged.
	Weekday	Weekday of the tweet. Categorical: 1-7.

Table 2: Full Results for the Logistic Regression Model on Virality

		Model 1		Model 2	
		Coef	Std Errors	Coef	Std Errors
	Intercept	-2438.457***	133.371	-2554.943***	-136.999
Content Features[1]	Length	0.009***	0.001	0.009***	0.002
	Hashtag	-1.053***	0.177	-1.481***	0.273
	Image	1.405***	0.059	1.231***	0.073
	Video	1.014***	0.098	0.498***	0.153
	QuoteRT	0.270	0.189	0.210	0.203
	Mention	-1.974***	0.219	-2.016***	0.280
	Link	-1.261***	0.172	-1.315***	0.182
Content Message[2]	M_Positive	-0.086***	0.027	-0.061**	0.031
	M_Negative	-0.072**	0.034	-0.063	0.040
	M_Profane	3.836***	0.259	2.179	1.191
	M_Humor	7.652***	0.165	6.205***	0.610
	M_Family	4.335***	0.400	3.814***	0.601
	M_Pop	-1.099***	0.272	-0.21	0.982
	M_Politic	-2.608***	0.477	-1.707	1.905
	M_Sports	-5.836***	0.818	-5.712***	1.199
	M_Relation	2.419***	0.385	1.327**	0.589
	M_Animal	3.255***	0.305	2.806***	0.421
	M_Emotion	1.212***	0.262	1.102***	0.274
Creator Features[3]	Age: Middle age	0.333***	0.126	0.323**	0.158
	Age: Old	0.237	0.267	0.087	0.310

	Age: Young	0.227***	0.075	0.056	0.106
	TweetVolume	-0.736***	0.053	-0.65***	0.067
	Followers	-0.138***	0.020	-0.009	0.036
	Following	-0.042	0.023	-0.022	0.025
	Race: Asia	-0.282	0.177	-0.437	0.289
	Race: Black	0.295***	0.099	0.418***	0.156
	Race: White	0.004	0.080	-0.059	0.129
Creator History[4]	H_Length	-0.012***	0.001	-0.004	0.003
	H_Viral	0.175***	0.026	0.240***	0.051
	H_RT	1.012***	0.222	0.017	0.448
	H_Original	2.997***	0.203	1.34***	0.440
	H_Profane	-2.425***	0.409	-5.106***	0.882
	H_Humor	-4.035***	0.408	-1.609**	0.784
	H_Family	-2.347**	1.072	-2.177**	1.095
	H_Pop	0.555***	0.156	-4.42E-04	0.289
	H_Politic	0.392	0.221	-0.291	0.432
	H_Sports	2.415***	0.314	1.099	0.644
	H_Relation	-2.097***	0.785	-2.087***	0.798
	H_Animal	-3.567***	0.564	-3.083***	0.609
	H_Emotion	-0.382	0.342	-0.160	0.342
Control Variables	Year	1.204***	0.066	1.262***	0.068
	Month	0.002	0.003	0.003	0.003
	Day	0.194***	0.010	0.198***	0.011
	Laglog	0.196***	0.014	0.200***	0.014

	Weekday: 1	-0.135	0.092	-0.159	0.094
	Weekday: 2	-0.286***	0.096	-0.284***	0.098
	Weekday: 3	-0.294***	0.097	-0.312***	0.099
	Weekday: 4	-0.123	0.095	-0.110	0.097
	Weekday: 5	-0.035	0.095	-0.046	0.098
	Weekday: 6	0.071	0.090	0.053	0.092
Interactions	H_Length * Length			-7.83E-05***	1.97E-05
	H_Length * Hashtag			0.011	0.006
	H_Length * Image			-0.006**	0.003
	H_Length * Video			-0.013***	0.005
	H_Length * Mention			0.004	0.008
	H_Length * M_Profane			-0.063***	0.015
	H_Length * M_Humor			-0.021**	0.008
	H_Length * M_Family			0.009	0.020
	H_Length * M_Pop			0.006	0.012
	H_Length * M_Politic			-0.044**	0.021
	H_Length * M_Sports			-0.023	0.043
	H_Length * M_Relation			-0.063***	0.020
	H_Length * M_Animal			0.007	0.011
	TweetVolume * M_Profane			-0.991**	0.526
	TweetVolume * M_Humor			-0.138	0.303
	Followers * Length			-6.68E-04**	3.04E-04
	Followers * Hashtag			-0.146	0.100
	Followers * Image			-0.067	0.035

	Followers * Video			-0.213***	0.065
	Followers * Mention			0.253**	0.121
	Followers * M_Positive			0.012	0.011
	Followers * M_Negative			0.006	0.015
	Followers * M_Profane			-0.52***	0.199
	Followers * M_Humor			-0.751***	0.118
	Followers * M_Family			-0.002	0.276
	Followers * M_Pop			-0.225	0.143
	Followers * M_Politic			0.088	0.264
	Followers * M_Sports			0.937**	0.510
	Followers * M_Relation			-0.488	0.255
	Followers * M_Animal			-0.775***	0.171
	H_Viral * Length			-2.08E-04	3.62E-04
	H_Viral * Hashtag			0.099	0.140
	H_Viral * QuoteRT			0.197***	0.057
	H_Viral * Image			-0.217***	0.049
	H_Viral * Video			0.009	0.087
	H_Viral * Mention			-0.141	0.189
	H_Viral * M_Profane			-0.737***	0.231
	H_Viral * M_Humor			0.297	0.159
	H_Viral * M_Family			-0.369	0.373
	H_Viral * M_Pop			-0.22	0.211
	H_Viral * M_Politic			-0.616	0.365
	H_Viral * M_Sports			-1.966***	0.659

	H_Viral * M_Relation			0.588	0.320
	H_Viral * M_Animal			-0.145	0.200
	Following * Mention			-0.319**	0.141
	H_RT * Length			-5.65E-04	0.004
	H_RT * Hashtag			1.226	1.569
	H_RT * Image			1.095**	0.464
	H_RT * Video			1.047	0.804
	H_RT * Mention			-1.195	1.889
	H_RT * M_Profane			-5.551**	2.418
	H_RT * M_Humor			-1.169	1.324
	H_RT * M_Family			-5.648	4.161
	H_RT * M_Pop			-4.592**	1.870
	H_RT * M_Politic			-7.122**	3.774
	H_RT * M_Sports			1.188	6.978
	H_RT * M_Relation			-1.088	3.164
	H_RT * M_Animal			-10.591***	2.147
	H_Original * Length			0.002	0.004
	H_Original * Hashtag			1.505	1.382
	H_Original * Image			1.227***	0.434
	H_Original * Video			1.438**	0.703
	H_Original * Mention			-0.646	1.796
	H_Original * M_Positive			-0.31	0.174
	H_Original * M_Negative			0.251	0.216
	H_Original * M_Profane			-1.249	1.782

	H_Original * M_Humor			2.56**	1.289
	H_Original * M_Family			4.83	2.959
	H_Original * M_Pop			-2.06	1.835
	H_Original * M_Politic			-4.736	3.640
	H_Original * M_Sports			0.297	6.908
	H_Original * M_Relation			1.495	2.897
	H_Original * M_Animal			-4.393***	1.652
	Race: Asia * Length			-0.002	0.003
	Race: Black * Length			-0.002	0.002
	Race: White * Length			-0.001	0.001
	Race: Asia * M_Profane			-0.754	1.557
	Race: Black * M_Profane			0.107	0.700
	Race: White * M_Profane			-0.917	0.747
	Race: Asia * M_Humor			0.821	0.956
	Race: Black * M_Humor			-0.675	0.593
	Race: White * M_Humor			0.206	0.461
	Race: Asia * M_Pop			-1.227	1.963
	Race: Black * M_Pop			0.1	0.950
	Race: White * M_Pop			-0.059	0.809
	Race: Asia * M_Politic			0.247	3.149
	Race: Black * M_Politic			0.372	1.883
	Race: White * M_Politic			-1.631	1.554
	Age: Middle age * Length			-0.003	0.002
	Age: Old * Length			-0.003	0.004

	Age: Young * Length			0.001	0.001
	Age: Middle age * M_Profane			0.616	1.559
	Age: Old * M_Profane			0.696	3.345
	Age: Young * M_Profane			-0.621	0.808
	Age: Middle age * M_Humor			-0.646	0.753
	Age: Old * M_Humor			0.144	2.096
	Age: Young * M_Humor			0.389	0.398
	Age: Middle age * M_Pop			-0.546	0.914
	Age: Old * M_Pop			0.276	1.815
	Age: Young * M_Pop			-1.098	0.631
	Age: Middle age * M_Politic			0.074	1.711
	Age: Old * M_Politic			-2.406	2.897
	Age: Young * M_Politic			-0.733	1.269
	H_Profane * Length			-0.001	0.006
	H_Profane * Hashtag			2.399	2.387
	H_Profane * QuoteRT			-5.372***	1.493
	H_Profane * Image			3.552***	0.797
	H_Profane * Video			2.587**	1.131
	H_Profane * Mention			1.889	2.784
	H_Profane * Link			5.707***	1.291
	H_Profane * M_Profane			7.113**	2.807
	H_Profane * M_Humor			3.439	2.431
	H_Profane * M_Family			12.153**	4.707

	H_Profane * M_Pop			6.946	3.750
	H_Profane * M_Politic			6.032	7.468
	H_Profane * M_Sports			6.734	11.791
	H_Profane * M_Relation			-3.891	6.198
	H_Humor * Length			-0.017**	0.007
	H_Humor * Hashtag			-5.171	3.663
	H_Humor * Image			0.364	0.811
	H_Humor * Video			-4.923***	1.757
	H_Humor * Mention			4.178**	1.877
	H_Humor * M_Humor			-12.851***	2.281
	H_Humor * M_Family			8.704	6.932
	H_Humor * M_Pop			-1.296	3.732
	H_Humor * M_Politic			-10.921	6.933
	H_Humor * M_Sports			10.269	7.634
	H_Humor * M_Relation			14.105***	5.263
	H_Humor * M_Animal			7.93	4.295
	H_Humor * M_Emotion			-4.907	3.541
	H_Pop * Length			0.004**	0.002
	H_Pop * Hashtag			-0.25	0.778
	H_Pop * Image			0.036	0.287
	H_Pop * Video			-0.456	0.532
	H_Pop * Mention			-0.536	0.740
	H_Pop * M_Profane			-1.63	1.894
	H_Pop * M_Humor			1.528	0.920

	H_Pop * M_Family			-3.208	2.786
	H_Pop * M_Pop			0.847	1.069
	H_Pop * M_Politic			-9.853***	3.097
	H_Pop * M_Sports			-1.937	4.234
	H_Pop * M_Relation			0.273	2.104
	H_Pop * M_Animal			3.627**	1.592
	H_Pop * M_Emotion			-1.812	1.066
	H_Politic * Length			0.005	0.003
	H_Politic * Hashtag			0.183	0.772
	H_Politic * Image			0.897**	0.416
	H_Politic * Video			0.465	0.890
	H_Politic * Mention			-0.235	0.883
	H_Politic * M_Positive			-0.199	0.158
	H_Politic * M_Negative			-0.164	0.203
	H_Politic * Link			0.071	0.470
	H_Politic * M_Profane			1.35	4.043
	H_Politic * M_Humor			-3.653**	1.557
	H_Politic * M_Family			-1.452	3.223
	H_Politic * M_Pop			2.034	1.905
	H_Politic * M_Politic			-1.078	1.903
	H_Politic * M_Sports			-4.747	9.251
	H_Politic * M_Relation			-2.371	3.751
	H_Politic * M_Animal			-1.495	2.018
	H_Politic * M_Emotion			1.949**	1.218

	H_Sports * Length			0.006	0.005
	H_Sports * Hashtag			0.906	1.124
	H_Sports * Image			0.257	0.628
	H_Sports * Video			0.554	0.923
	H_Sports * Mention			0.451	1.030
	H_Sports * M_Profane			7.207**	3.060
	H_Sports * M_Humor			3.884**	1.932
	H_Sports * M_Family			0.316	4.892
	H_Sports * M_Pop			2.331	2.761
	H_Sports * M_Politic			-1.19	2.704
	H_Sports * M_Sports			-0.89	3.619
	H_Sports * M_Relation			0.009	5.202
	H_Sports * M_Animal			3.16	2.910
McFadden's R2		0.268		0.330	
Cragg &Uhler's R2		0.274		0.338	

Table 3: Multicollinearity (VIF)

Category	Variable	VIF	Tolerance
Dependent Variable	IsViral	1.01	0.9878
Content Features[1]	Length	1.95	0.5128
	Hashtag	1.17	0.8553
	Image	1.11	0.9015
	Video	1.08	0.9269
	QuoteRT	1.91	0.5232
	Mention	1.14	0.8787
	Link	2.04	0.4890
Content Message[2]	M_Positive	1.36	0.7367
	M_Negative	1.23	0.8116
	M_Profane	1.17	0.8558
	M_Humor	1.10	0.9085
	M_Family	1.01	0.9896
	M_Pop	1.72	0.5806
	M_Politic	1.59	0.6300
	M_Sports	1.89	0.5300
	M_Relation	1.03	0.9740
	M_Animal	1.18	0.8504
	M_Emotion	1.26	0.7906
Creator Features[3]	Age	1.18	0.8480
	TweetVolume	1.20	0.8303
	Followers	3.97	0.2521
	Following	1.13	0.8857
	Race	1.17	0.8571

Creator History[4]	H_Length	1.88	0.5327
	H_Viral	3.00	0.3328
	H_RT	1.43	0.6993
	H_Original	1.44	0.6947
	H_Profane	1.58	0.6344
	H_Humor	1.23	0.8163
	H_Family	1.05	0.9494
	H_Pop	2.02	0.4947
	H_Politic	2.08	0.4818
	H_Sports	1.95	0.5134
	H_Relation	1.48	0.6735
	H_Animal	1.70	0.5877
	H_Emotion	2.45	0.4085
Control Variables	Year	1.59	0.6305
	Month	1.17	0.8541
	Day	1.00	0.9984
	Laglog	1.08	0.9284
	Weekday	1.00	0.9968

Table 4: Results for Non-linear Classification Algorithms on Virality

Model	With Interactions		Without Interactions	
	Accuracy	Standard Deviation	Accuracy	Standard Deviation
Random Forest	86%	2%	73%	5%
SVM	83%	3%	74%	74%
Multinomial NB	77%	6%	65%	11%
KNN	75%	5%	66%	8%
Logit	83%	2%	71%	6%

- The parameters of all prediction algorithms where tuned via grid search on a validation sample.
- In order to provide additional evidence, we use the best-first method(Dash and Liu 1997)[1] for selecting a subset of highly predictive features for the Random Forest Classifier, which achieved the best results among all considered algorithms. A 10-fold cross validation revealed that between 41 and 47 interactions were consistently included in the set of selected attributes for all 10 folds. The included interactions were consistently among those reported as significant by our logit model.

[1]Dash M, Liu H (1997) Feature selection for classification. *Intell. Data Anal.* 1(1–4):131–156.

Figure 1: Distribution of User Average RT

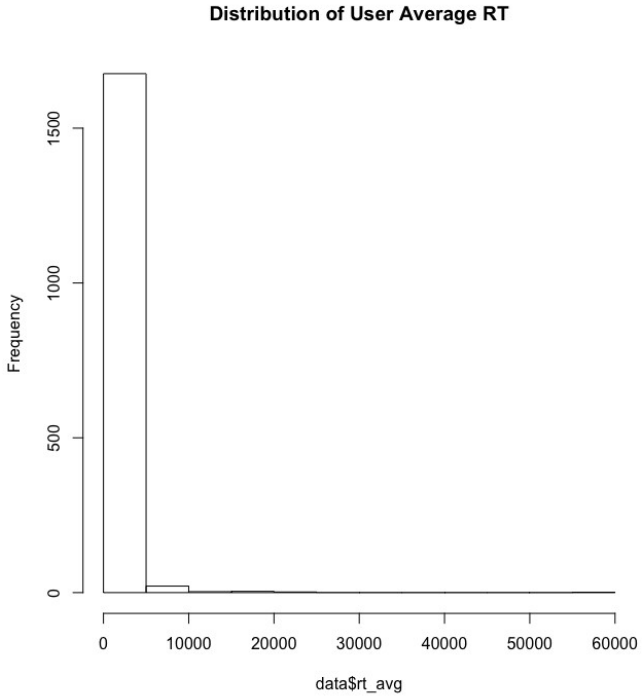

Figure 2: Distribution of User Average RT (log)