7-2017

The role of different tie strength in disseminating different topics on a microblog

Felicia NATALI
Singapore Management University, felician.2013@smu.edu.sg

Kathleen M. CARLEY
Carnegie Mellon University

Feida ZHU
Singapore Management University, fdzhu@smu.edu.sg

Binxuan HUANG
Carnegie Mellon University

DOI: https://doi.org/10.1145/3110025.3110130

Citation

The Role of Different Tie Strength in Disseminating Different Topics on a Microblog

Felicia Natali*, Kathleen M. Carley†, Feida Zhu*, and Binxuan Huang†
*School of Information Systems
Singapore Management University
Email: felician.2013@smu.edu.sg, fdzhu@smu.edu.sg
†Center for Computational Analysis of Social and Organizational Systems
Carnegie Mellon University
Email: kathleen.carley@cs.cmu.edu, binxuanh@andrew.cmu.edu

***Abstract*—The study of information flow typically does not distinguish the choices of tie strength on which the information flows. All receivers of the information are assumed to have the same potential to pass on the information. Modifying the SEIZ (susceptible, exposed, infected, skeptic) model, we discover that people choose to retweet strong or weak ties based on the topic. We made two modifications in the model. In the first modification (Model I), we assume that the contact rates of agents in different compartment and the probability of an agent transitioning from one compartment to another are different for strong ties and weak ties. In the second modification (Model II), we assume that only the probability of transitioning is different for strong ties and weak ties. We discover that people do not discriminate strong ties and weak ties when retweeting controversial topic, perhaps because this topic can both be personal and breaking news. On the other hand, people discriminate strong ties and weak ties when retweeting non-controversial topic. They prefer to retweet strong ties when the topic is donation, and kids, and weak ties when the topic is news on hurricane and music. Meanwhile, SEIZ model and its modifications are found to be inadequate to model tweets on event promotion.**

I. Introduction

The recent proliferation of fake news around the world through social media has made the study of information spread on Twitter becoming more relevant. Information spread on Twitter is performed through a mechanism called retweet. For decades, researchers have spent effort to model the retweet process. Some of these models use epidemiological modeling [1][2][3][4][5], and the others try to recreate the retweet process based on the characteristics that they have observed on real data [6][7][8]. Recently, Shi et al. found that weak, one-way relationships rather than a two-way relationship between Twitter users made it more likely that a follower would re-tweet or re-post [9]. All these investigations neglect the interdependence between topic and tie strength in a retweet.

However, previous studies have shown that there is a preference for strong ties in passing social influence and important information. Interdepartmental information is more likely to flow through weak ties than intradepartmental information [10]. Information related to a major change that may challenge the status quo and standard routines propagates mostly through strong ties [11]. Many studies have indicated that political influence and discussion happen mainly among strong ties [12][13][14]. In a 61-million experiment on Facebook, strong ties are shown to be instrumental to influence others to vote [15].

As such, it is also possible for certain topics on Twitter to spread more widely among strong ties. Because not all topics on Twitter are important and personally influential, we expect not all topics are likely to spread among strong ties. In sum, our research question is: "Does the spread of information on Twitter among strong ties and weak ties differ given the topic?" In this study, we contribute by being the first to investigate the interdependence between tie strength and topic on the retweet process.

II. Method

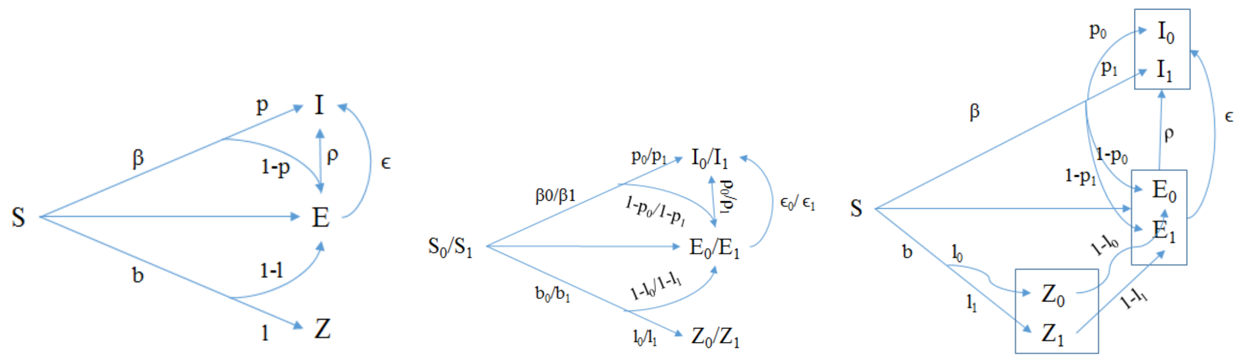
We modify the SEIZ (susceptible, exposed, infected, skeptic) model on several Twitter topics. In the first modification, we separate all the parameters into two, for strong ties and for weak ties. In the second modification, we separate the probability of transitioning into two, for strong ties and weak ties. The SEIZ model has been proven to fit well several news and rumors on Twitter[1][16].

A. SEIZ Model

In the SEIZ model [1][16] there is a state called exposed (E) where an agent is exposed with the information, but has yet to spread the information.

In the SEIZ model (Figure 1a), S is a susceptible person who has not tweeted the tweet. E is an exposed person who has received the tweet but has yet to retweet. I is an infected person who retweets the tweet. Z is a skeptic who decides not to spread the tweet even upon hearing it. A susceptible person meets an infected person at the rate β and meets a skeptic at the rate b. After meeting the infected, some of the susceptible

Fig. 1: Basic Fit



(a) SEIZ model framework (b) Model 1 Framework (c) Model 2 Framework

TABLE I: Parameter definitions in SEIZ model

Parameter	**Definition**
β	S-I contact rate
b	S-Z contact rate
ρ	E-I contact rate
ϵ	Incubation rate (self-arising or due to outside influence)
l	S->Z probability given contact with skeptics
$1-l$	S->E probability given contact with skeptics
p	S->I probability given contact with adopters
$1-p$	S->E probability given contact with adopters

people are also becoming infected at the probability p, and are becoming exposed at the probability $1-p$. The exposed people may get in touch with an infected people again at the rate ρ and thus, are becoming infected. They may also, after a certain time, becoming infected at the rate ϵ by themselves or by the outside influence. Meanwhile, the susceptible persons who meet the skeptic persons transform into skeptic at the probability l. Unintended, by meeting the skeptics, the susceptible people can get exposed to the news and enter the exposed compartment at the probability $1-l$.

The parameter definitions in the SEIZ model can be seen in the Table I.

We modify the SEIZ model into two. The first model is a liberal assumption where we differentiate all parameters given strong and weak ties. In this study, **strong** ties are **reciprocal** ties on Twitter, and **weak** ties are **non-reciprocal** ties. The second model takes a conservative approach where we change as few parameters as possible based on reasonable assumptions. In the first model, we assume that both the contact rates and the probabilities of transition are different. Contact rates indicate the rate at which a user receives tweets. Twitter creates a mechanism such that the tweets that you are likely to care about most will show up first in your timeline. These tweets are chosen based on accounts you interact with most, tweets you engage with, and much more. As such, the intuition behind the model is, people interact differently among strong ties and weak ties depending on the tweet.

The second model assumes that only the probabilities of transition are different. Probabilities of transition indicate the probability of retweeting and not retweeting. The intuition behind the model is people interact blindly regardless of topic. However, they discriminate strong ties and weak ties when they retweet depending on the topic. Meanwhile, we do not discriminate the parameters after second-hand exposure to the information (ϵ, ρ) because we assume that the popularity of the message and repeated exposure have created equanimity regarding who passes the message.

B. Model I

Figure 1b represents Model 1. Subscript 0 represents weak ties and subscript 1 represents strong ties. We first divide the datasets into two, the number of retweets that come from the strong ties and the number of retweets that come from the weak ties. We then run the model separately on these two datasets.

The ODEs of the first model does not change from the ODEs of the original model but the original parameters search is performed twice, one for the retweet chains among strong ties, and one for the retweet chain among weak ties.

C. Model II

Figure 1c shows Model 2 framework.

Model 2 is mathematically represented by the following system of ODEs. In the equations, t represent total, 0 represents weak ties, and 1 represents strong ties.

$$\frac{d[S]}{dt} = -\beta S\frac{I_t}{N} - bS\frac{Z}{N} \tag{1a}$$

$$\frac{d[E_0]}{dt} = (1-p_0)\beta S\frac{I_0}{N} + (1-l_0)bS\frac{Z_0}{N} - \rho E_0\frac{I_0}{N} - \epsilon E_0 \tag{1b}$$

$$\frac{d[E_1]}{dt} = (1-p_1)\beta S\frac{I_1}{N} + (1-l_1)bS\frac{Z_1}{N} - \rho E_1\frac{I_1}{N} - \epsilon E_1 \quad (1c)$$

$$\frac{d[I_0]}{dt} = p_0\beta S\frac{I_0}{N} + \rho E_0\frac{I_0}{N} + \epsilon E_0 \quad (1d)$$

$$\frac{d[I_1]}{dt} = p_1\beta S\frac{I_1}{N} + \rho E_1\frac{I_1}{N} + \epsilon E_1 \quad (1e)$$

$$\frac{d[Z_0]}{dt} = l_0 bS\frac{Z_0}{N} \quad (1f)$$

$$\frac{d[Z_1]}{dt} = l_1 bS\frac{Z_1}{N} \quad (1g)$$

The ODE shows that the compartment S meets the compartment I at the rate β, and meets the compartment Z at the rate b. Meanwhile, after second-hand exposure, the compartment E can self-transition to the compartment I at the rate ϵ and non-self-transition into the compartment I at the rate ρ. However, after first-hand exposure there are several choices of transition that compartment S, E and Z can go to depending on the tie. The compartment S can transition into the compartment I at the probability $(1-p_0)$ if he meets a weak tie and $(1-p_1)$ if he meets a strong tie. The compartment S can also transition into the compartment Z at the rate $(1-l_0)$ if he meets a weak tie and $(1-l_1)$ if he meets a strong tie.

D. Parameter Identification

We identify the parameters exactly in the same way that Fang et al. did them on their paper[1]. The set of parameter values chosen are those that minimize $|I_t(t) - tweets_t(t)|$ for the first model, and those that minimize $|I_0(t) - tweets_0(t)| + |I_1(t) - tweets_1(t)|$ for the second model. $I_t(t)$ is the total number of predicted tweets at time t, $I_0(t)$ is the total number of predicted tweets retweeted by weak ties at time t, and $I_1(t)$ is the total number of predicted tweets retweeted by strong ties at time t. Total Population $S(t_0)$, $E(t_0)$, $I(t_0)$, $Z(t_0)$, and N are considered as unknowns and treated as parameters. The **lsqnonlin** function performed the least squares fit, while the ODE systems were solved with a forward Euler function. All the parameters are initialized.

The following explanation takes as an example Model 1 to describe the *lsqnonlin* and the forward Euler method. *Lsqnonlin* is a matlab function that tries to find the values of x (in this study, all the transition probabilities and contact rates) so that the sum of squares of the values of $y(x)$, i.e. $|I_t(t) - tweets_t(t)|$, is minimum. At each iteration of *lsqnonlin*, the forward Euler method estimates the other parameter values (the number of population in each compartment) through solving the ODE system. The forward Euler method states that the value of $I_1(t+1)$ is equal to the value of $I_1(t)$, added with a constant h times $f(I_1(t))$. In our case, $f()$ is the ODE system we have derived. We set h to be 0.1. So, the total step would be (end timestep - start timestep)/0.1. The original code by Fangjin et al. can be found on her homepage. We modify this original code for our studies.

III. Dataset

We collected retweet data from October 12th 2016 to December 2nd 2016 on several topics that you can see on Table II.

IV. Results

We compare our results with the simulation result of the original SEIZ model. Our results are displayed on Table III.

The results can be grouped into three: controversial topic, non-controversial topic, and event promotion. Controversial topic includes clinton, trump, brexit. Non-controversial topic includes hurricane, donation, music, and kids. Meanwhile, concert and sports represent event promotion.

In controversial topic, except for the topic trump, the SEIZ basic model prevails (See error values on Table III). The results show that people do not discriminate strong ties and weak ties in spreading information on this topic, perhaps because controversial topic is both personal and breaking news. It is personal because it may involve personal values. It is breaking news because it often affects the lives of many. An exception happens for the topic trump. Model 1 is the best fit for the topic. All the parameters are larger for weak ties, indicating the heavy preference and utiliization of weak ties to spread the tweet on Twitter. As there are more weak ties than strong ties on Twitter, the heavy utilization of weak ties could be the cause to the popularity of Trump's tweets compared to Clinton's even before the election.

In non-controversial topic, either Model 1 or Model 2 performs better than the basic model for all tweets. As such, people discriminate strong ties and weak ties when retweeting. However, the results for both Model 1 and Model 2 are generally not contradictory. On the topic hurricane, weak ties are retweeted more on the first exposure (Model 1: $p_0 > p_1$ Model 2: $p_0 > p_1$), but strong ties are retweeted more after the second exposure (Model 1: $\rho_1 > \rho_0$). On the topic donation, strong ties are used more to retweet (Model 1: $p_1 > p_0$, $l_0 > l_1$, $\rho_1 > \rho_0$ Model 2: $p_1 > p_0$). It is unsurprising as donation depends on personal ties. On the topic music, weak ties are used more (Model 1: $\beta_0 > \beta_1$, $p_0 > p_1$ Model 2: $p_0 > p_1$), but strong ties provide a stronger influence outside Twitter (Model 1: $\epsilon_1 > \epsilon_0$). Lastly, on the topic kids, strong ties are more important (Model 1: $\beta_1 > \beta_0, \epsilon_1 > \epsilon_0, \rho_1 > \rho_0$ Model 2: $p_1 > p_0, l_0 > l_1$).

The third group is event promotion that promotes an offline event on Twitter. The first tweet promoted Justin Biebers purpose tour that just happened in order to market Calvin Klein product and promote a similar offline event in the future (purpose tour in other cities/countries). The second tweet invited users to watch a baseball game at the Wrigley field the next day. All models perform poorly in fitting this topic, and as such we will not even try to interpret the parameters. On the topic sports, the fit overestimates the spread of tweets among strong ties and underestimates the spread of tweets among weak ties. Meanwhile, on the topic concert, the fit underestimates the spread of tweets among strong ties and overestimates the spread of tweets among weak ties. The

TABLE II: Tweets

Topic	#Retweet	Text
clinton	21953	Well there you have it. A highly intelligent experienced woman just debated a giant orange Twitter egg. Your move America. #debate
trump	20299	Time to #DrainTheSwamp in Washington D.C. and VOTE #TrumpPence16 on 11/8/2016. Together we will MAKE AMERICA SAFE https://t.co/rVcjXdWxzp
brexit	77062	BRITAIN: Brexit is the stupidest most self-destructive act a country could undertake. USA: Hold my beer.
hurricane	28870	WE WERE OUT HERE PRAYING FOR FLORIDA TO STAY SAFE FROM HURRICANE MATTHEW. LITTLE DID WE KNOW. HURRICANE MATTHEW WAS https://t.co/DRbKFRbkhv
donation	13155	Florida just got hit by a category 5 Hurricane! Please donate. Me: https://t.co/xYjALm72Gw
music	63662	All Weekend Long: Official Music Video https://t.co/VRvN60NU1v #AWLMusicVideo
kids	8482	Check out my newest science advisors! These kids are fearless in using science to tackle our toughest problems. Tha https://t.co/pLBZQWDFin
concert	36076	#PurposeTour in #mycalvins https://t.co/FahXxb3JsL
sports	23413	Wrigley Field will be loud tomorrow. RT this for your chance to win two tickets to #NLCS Game 6! #FlyTheW https://t.co/L0mwAGmNSV

results show that offline events do not spread according to the SEIZ framework on Twitter. Estimating on which framework they spread is beyond the scope of this study. As Twitter is a promotion platform for these offline events, information spread is going to depend very much on the nature of the event. Depending on the nature of the event, strong ties and weak ties may react unpredictably. Similarly, other means of circulation may play an unexpectedly greater or smaller role in information spread that may overestimate or underestimate the tweets spread on Twitter.

References

[1] F. Jin, E. Dougherty, P. Saraf, P. Mi, Y. Cao, and N. Ramakrishnan, "Epidemiological modeling of news and rumors on twitter," in *Proceedings of the 7th Workshop on Social Network Mining and Analysis*. ACM, 2013, p. 8.

[2] S. Abdullah and X. Wu, "An epidemic model for news spreading on twitter," in *2011 23rd IEEE International Conference on Tools with Artificial Intelligence*. IEEE, August 2011.

[3] D. J. Daley and D. G. Kendall, "Epidemics and rumours," *Nature*, vol. 204, no. 4963, pp. 1118–1118, 1964.

[4] D. Maki and M. Thompson, *Mathematical Models and Applications: With Emphasis on the Social, Life, and Management Sciences*. NJ: Prentice-Hall, 1973.

[5] D. Meng, L. Wan, and L. Zhang, "A study of rumor spreading with epidemic model based on network topology," in *PAKDD 2014: The 18th Pacific-Asia Conference on Knowledge Discovery and Data Mining*, vol. 8643. LNAI, May 2014, p. 381392.

[6] J. Weng, E. P. Lim, and Q. He, "Twitterrank: Finding topic-sensitive influential twitterers," in *WSDM 2010: Proceedings of the Third ACM International Conference on Web Search and Data Mining*. ACM, February 2010, pp. 261–270.

[7] S. A. Macskassy and M. Michelson, "Why do people retweet? anti-homophily wins the day!" in *ICWSM 2011: Proceedings of the Fifth Annual Conference on Weblogs and Social Media*. ACM, July 2011, pp. 209–216.

[8] V. Arnaboldi, M. Conti, M. L. Gala, A. Passarella, and F. Pezzoni, "Information diffusion in osns: the impact of nodes sociality," in *SAC 2014: Proceedings of the 29th Annual ACM Symposium on Applied Computing*, 2014, pp. 616–621.

[9] Z. Shi, H. Rui, and A. B. Whinston, "Content sharing in a social broadcasting environment: Evidence from twitter," *MIS Quarterly*, vol. 38, no. 1, pp. 123–142, 2014.

[10] N. E. Friedkin, "Information flow through strong and weak ties in intraorganizational social networks," *Social Networks*, vol. 3, pp. 273–285, 1982.

[11] D. Krackhardt, "The strength of strong ties: The importance of philos in organizations," in *Organizations and Networks: Structure, Form, and Action*, N. Nohria and R. G. Eccles, Eds. Boston: Harvard Business School Press, 1992, pp. 216–239.

[12] B. C. Straits, "Bringing strong ties back in interpersonal gateways to political information and influence," *Oxford Journals*, vol. 55, no. 3, pp. 432–448, 1991.

[13] K. Beyerlein and K. Bergstrand, "It takes two: A dyadic model of recruitment to civic activity," *Social Science Research*, vol. 60, pp. 163–180, JUly 2016.

[14] D. Lazer, B. Rubineau, C. Chetkovich, N. Katz, and M. Neblo, "The co-evolution of networks and political attitudes," *Political Communication*, vol. 27, pp. 248–274, 2010.

[15] R. M. Bond, C. J. Fariss, J. J. Jones, A. D. I. Kramer, C. Marlow, J. E. Settle, and J. H. Fowler, "A 61-million-person experiment in social influence and political mobilization," *Nature*, vol. 489, pp. 295–298, 2012.

[16] F. Jin, W. Wang, L. Zhao, E. Dougherty, Y. Cao, C.-T. Lu, and N. Ramakrishnan, "Misinformation propagation in the age of twitter," *Computer*, vol. 47, no. 12, pp. 90–94, 2014.

TABLE III: Parameters Estimation Results

Basic SEIZ								
Topic	**Error**	**Avg Dev**	β	ϵ	**p**	**l**	ρ	**b**
clinton	0.005	40.338	0.000	0.056	0.000	0.000	0.226	0.000
trump	0.031	221.655	0.000	0.010	0.070	0.706	0.249	0.000
brexit	0.002	64.603	0.000	0.072	1.000	0.499	0.331	0.012
hurricane	0.023	193.465	0.000	0.048	1.000	0.892	1.153	0.000
donation	0.072	343.268	0.000	0.000	0.033	0.714	1.711	0.000
music	0.002	121.077	0.000	0.198	0.000	0.787	0.000	0.000
kids	0.027	109.572	0.025	0.090	0.000	0.942	0.000	0.249
concert	0.694	19635.874	1.173	0.033	0.059	0.489	0.000	1.305
sports	0.108	1135.689	0.000	0.000	0.074	0.870	5.556	0.000
Model 1								
clinton (S)	0.055	70.898	3.75E-08	0.029	1.000	0.848	4.00E-18	0.987
clinton (W)	0.137	829.812	3.30E-04	0.003	1.000	0.843	0.00E+00	0.000
trump (S)	0.004	6.677	0.004	0.000	0.000	0.938	0.204	0.033
trump (W)	0.002	12.706	0.053	0.003	1.000	1.000	1.164	0.228
brexit (S)	0.029	457.770	0.000	0.103	1.000	0.000	0.044	0.000
brexit (W)	0.003	56.298	0.001	0.043	1.000	0.999	3.155	0.011
hurricane (S)	0.012	51.773	1.04E-05	2.00E-05	0.001	0.679	3.389	5.800
hurricane (W)	0.013	33.058	0.00E+00	2.00E-05	0.462	1.000	2.923	10.000
donation (S)	0.004	11.789	1.98E-18	2.00E-05	0.025	0.000	10.000	4.01E-06
donation (W)	0.073	120.410	0.00E+00	2.00E-05	0.020	0.743	1.873	2.98E-15
music (S)	0.002	62.540	2.84E-06	0.164	0.000	0.813	0.00E+00	0.000
music (W)	0.002	27.157	2.06E-04	0.136	0.674	1.000	6.28E-17	2.246
kids (S)	0.021	15.921	0.154	0.148	1.04E-16	0.950	0.207	9.87E-17
kids (W)	0.024	76.738	0.000	0.028	0.00E+00	0.881	0.000	2.45E-04
concert (S)	0.984	11053.427	7.14E-16	0.000	0.000	0.742	1.361	1.488
concert (W)	0.156	2605.591	2.49E-16	0.001	0.099	1.000	3.174	0.000
sports (S)	0.313	1975.682	0.00E+00	0.000	0.424	0.883	10.000	0.257
sports (W)	0.038	153.766	0.00E+00	0.451	1.000	0.584	10.000	0.000
Model 2								
clinton (S)	0.008	26.901	0.003	0.065	3.61E-15	0.786	0.085	2.351
clinton (W)	0.007	35.835			8.88E-16	0.994		
trump (S)	0.270	1010.042	0.004	0.000	0.000	0.831	8.692	0.788
trump (W)	0.332	4576.826			0.853	1.000		
brexit (S)	1.508	26893.553	0.589	0.000	0.000	0.962	2.998	2.084
brexit (W)	0.203	6351.668			0.177	0.717		
hurricane (S)	0.010	39.965	0.180	0.000	0.000E+00	1.000	0.983	0.512
hurricane (W)	0.012	47.890			2.711E-20	1.000		
donation (S)	0.004	16.628	0.083	0.086	1.129E-15	0.987	1.314	0.602
donation (W)	0.007	12.902			0.000E+00	0.988		
music (S)	0.001	34.251	0.026	0.058	0.812	1.000	0.000	1.255
music (W)	0.001	29.093			1.000	0.991		
kids (S)	0.026	56.784	0.677	0.004	0.291	0.949	1.726	1.919
kids (W)	0.018	68.807			0.211	0.990		
concert (S)	0.586	6386.223	0.397	0.000	1.000	1.000	3.507	1.716
concert (W)	10.064	175682.619			1.000	0.727		
sports (S)	9.397	62209.439	0.189	0.002	0.000	0.873	7.416	1.363
sports (W)	0.145	629.074			1.000	1.000		