May 2015

Three Essays on Retransmission of Brand Messages in Social Media

Nima Yahyapour Jalali
University of Wisconsin-Milwaukee

THREE ESSAYS ON RETRANSMISSION OF BRAND MESSAGES IN SOCIAL MEDIA

by

Nima Yahyapour Jalali

A Dissertation Submitted in Partial Fulfillment

of the Requirement for the Degree of

Doctoral of Philosophy

in

Management Science

at

The University of Wisconsin-Milwaukee

May 2015

ABASTRACT

THREE ESSAYS ON RETRANSMISSION OF BRAND MESSAGES IN SOCIAL MEDIA

by

Nima Yahyapour Jalali

The University of Wisconsin-Milwaukee, 2015

Under supervision of Professor Purushottam Papatla

Social networks have emerged as an important channel for brands to communicate with customers both directly as well as secondarily through customers who share the communications with others. A value of this channel to brands, therefore, depends on how effective they are in increasing the retransmission of their messages by customers. This is the issue that we investigate through three essays using the social media site Twitter as our research setting. Our investigation is based on the theory that retweeting is a choice made by consumers who rely on constructive preferences (Bettman, Luce and Payne 1998) while seeking three intangible benefits: altruism, self-enhancement, and social interaction. They also have two overarching metagoals of accuracy maximization and effort minimization (Bettman et al 1998) as they seek the benefits. Within this theoretical context, we examine the design attributes of tweets that increase retweeting. Specifically, we investigate more than 14000 tweets by 62 brands, across four product categories, over periods ranging from 18 to 400 days. Our empirical results are consistent with the theory and suggest that brand

and tweet characteristics that increase the recipients' ability to maximize the benefits of retweeting, and minimize the cognitive effort required to decide whether to retweet or not, increase retweets. The key managerial implication of our findings therefore is that brands should design tweets carefully to increase altruism, self-enhancement, and social interaction benefits while reducing the amount of effort that recipients need to undertake to assess whether the tweet offers these benefits. We replicate and extend these findings in essay two in the context of celebrities as brands by investigating the volume and duration of retweets of more than 2900 tweets by 65 celebrities across seven categories of the entertainment industry. Our results from this essay suggest that traits of the sources, i.e., the celebrities, also play a role in how recipients assess whether a tweet can deliver the three benefits while realizing the two metagoals. The third essay focuses on brands' desire to generate retweets at a rapid rate before the tweet loses its relevance. In addition to volume and duration, therefore, we also investigate the rate at which a tweet is retweeted in this essay. Our investigation examines the retweet rates, in fifteen minutes intervals over a 24-hour period, of more than 2400 tweets posted by 62 celebrities using a Modulate Poisson Process model (Soyer and Tarimcilar 2008). Our results suggest that tweets that do not need recipients to interact with them and are related to significant cultural events are retweeted at a faster rate than others.

Acknowledgements

First and foremost, I would like to express my appreciation to my PhD advisors, Professor Purushottam Papatla and Professor Ehsan S. Soofi, for supporting me during my PhD studies. Professor Papatla has been a tremendous mentor for me, and provided me with the scientific advice and knowledge and many insightful discussions and suggestions. I would like to thank you for encouraging my research and help me to grow as a professional in my career. Professor Soofi has been supportive, and his advice on both research as well as on my career have been priceless. He has been helpful in providing advice many times during my graduate studies. He will be one of my best role models for a professional scholar. I hope that I could be as lively, enthusiastic, and energetic as he is.

I also want to thank the members of my PhD committee, Professors V. Kanti Prasad, Amit Bhatnagar, and Janice Miller for their helpful career advice, brilliant comments, and suggestions in general. Thanks for letting my defense to be an enjoyable moment. I am indebted to the University of Wisconsin-Milwaukee for their generous funding, which removed financial concerns from my decision to embark on this journey. Lastly and most importantly, my deepest gratitude goes to my family for their love and support. My wife, Sepideh Maleki, who has always supported, encouraged and believed in me throughout my life. My mother, Jaleh Bandeganian, and my father, Mohammad Jalali, for all of the sacrifices that they have made on my behalf. Words cannot express how grateful I am to them. My little sister, Niloufar Jalali, who always inspired me with hope and faith.

Nima Y. Jalali

Table of Contents

Essay 3: Velocity of Retweeting: Insights from Celebrity Tweets

List of Tables

List of Figures

Essay 1: Constructive Preferences and the Retransmission of Tweets

1.1. Introduction

Over the past few years, social media have evolved into a medium where brands and consumers interact. Most brands therefore maintain a presence and build their social networks on popular social media like Twitter and Facebook. Media reports (Huffington Post 2014) suggest that almost all of the Fortune 500 firms have accounts on Twitter. In fact, most of these firms have about six accounts each (Malhotra et al 2012).

While the primary goal of brands is to interact with customers, an equally important objective is to stimulate conversations about their products. Such consumer-to-consumer conversations can be more effective and less expensive than traditional advertising campaigns (Fournier and Avery 2011) since they serve as electronic word of mouth that implicitly endorses the brand (Malhotra et al 2012). Brands attempt to stimulate conversations by seeding (Deighton and Kornfeld 2009, Schau, Muñiz and Arnould 2009, Singh and Sonnenburg 2012) the network with announcements, promotional offers, contests and events (Muniz and Schau 2011) or any other topic that might be of interest to participants. For instance, the fast-food brand Arby's sent out a tweet (Figure 1.1) during the Grammy awards ceremony connecting the hat worn by the popular singer Pharrell to the hat in its logo (Time 2014). The tweet was retweeted over eighty thousand times attracting in the process a retweet from the singer which was again retweeted over sixteen thousand times (AdWeek 2014a). Arby's subsequently bought the hat from the singer who donated the proceeds to charity, which stimulated an additional forty thousand references to the brand on Twitter (AdWeek 2014b).

Figure 1.1 – Tweet posted by Arby's about Grammys

The enormous success of Arby's tweet, however, is an exception. Typically, tweets from brands attract far fewer retweets and the average number of retweets is only about 25 (Malhotra et al 2012). Brands, therefore, need insights into what motivates their followers on social media to retransmit their communications. Such insights can be useful in crafting messages that customers will share and, thus, help brands take advantage of social media for communications and promotions. Surprisingly, although there is substantial research in the literature on virality (Berger and Milkman 2012, Schulze, Schöler and Skiera 2014) and discussions of brands and categories by consumers on social media (Lovett, Peres, and Shachar 2013), there is little research as yet on why messages from brands to customers are shared with others. In particular, questions related to characteristics of messages from brands that lend themselves to retransmission by consumers in social media have not yet been researched. This is the issue that we investigate using the social media platform Twitter and tweets by brands to customers as our research setting.

Our investigation is based on the theory that retweeting is a choice made by consumers based on whether doing so allows them to reach specific goals (Bettman et al 1998). Multiple characteristics of this choice task, however, distinguish it from product or brand choice decisions investigated in the literature (e.g., Guadagni and Little 1983, Hoyer and Brown 1990, McFadden 1990) and make it a goal-based rather than a utility-based decision. Specifically, since tweets are neither products nor brands, consumers will neither have well-defined preferences among tweets nor the ability to identify tweets that would give them greater utility from retweeting than not doing so. Consequently, their preferences would be constructive (Bettman, Luce and Payne 1998) and retweeting choices are likely to be made to realize one or more goals. First, consumers, seek several intangible benefits such as self-enhancement (De Angelis et al 2012), altruism (Hennig-Thurau et al 2004), and social interaction (Lovett, Peres and Shachar 2013) from retweeting. The goal in this case would therefore be to maximize the accuracy of their retweeting decision in realizing those benefits. Second, consumers receive a large number of tweets from brands[1] but, typically, do not realize any tangible benefits from retweeting. Their involvement and, hence, the willingness to put in much cognitive effort into the retweeting decision is therefore likely to be low. Thus, the goal would be to minimize the cognitive effort expended. The retweeting decision will therefore be taken to realize the two metagoals of accuracy maximization and effort minimization (Bettman, Luce and Payne 1998), rather than based on utility. Consequently, consumers are more likely to retweet those tweets that allow them to realize these metagoals.

We test our theory on the role of metagoals by empirically investigating the retweeting response to tweets in four product categories: automotive, food and beverage, dining and airline. In each category, we consider the top twenty brands, in terms of the number of consumers that follow them on Twitter. For each brand, we investigate

[1] For instance, Starbucks sends about ten tweets on average per day (Hassan Zadeh and Sharda 2014).

retweeting behavior with regard to 500 of its most recent tweets as of the date on which we collected the data[2]. The ability of a tweet to help the recipient maximize accuracy, while minimizing effort, in realizing the intangible benefits of social interaction, altruism, and self-enhancement is operationalized in terms of brand and tweet characteristics. For instance, the ability of a brand's tweet to provide social interaction benefits is operationalized through the number of followers of the brand on Twitter and on other social media sites like Facebook. Similarly, a tweet's ability to provide altruistic benefits to the recipient is operationalized through the presence of characters related to monetary benefits such as a dollar sign, references to a promotion or to taking advantage of one in the near future. The characteristics of tweets that help recipients take a decision on retweeting without undertaking extensive effort are captured through the presence of characters like hash tags and exclamation marks and the length in terms of number of characters included.

Our empirical results are consistent with the theory and suggest that the retweeting decision is based on whether a tweet meets the two metagoals of accuracy maximization and effort minimization. Specifically, brand and tweet characteristics that increase the recipients' ability to maximize the benefits of retweeting and minimize the cognitive effort required to take the decision on whether or not to retweet increase the total number of retweets. The key managerial implication of our findings therefore is that, in order to increase the re-transmission of their tweets, brands should design tweets carefully such that they increase social interaction, altruism and self-enhancement benefits while reducing the amount of effort that recipients need to undertake to assess tweet's benefits.

[2] We collected our data on April 1, 2012.

We next provide an overview of the theory and follow with a description of our data. Following this, we present our model and empirical results. The chapter concludes with a discussion of the managerial implications and directions for future research.

1.2. Background and Theory

1.2.1. Background

Findings in the literature on choice over many decades (e.g., Guadagni and Little 1983, McFadden 1973, 2001) confirm that individuals make choices based on the perceived utilities of the options, and that choice behavior "can be characterized by a decision process which is informed by perceptions and beliefs based on available information, and is influenced by affect, attitudes, motives and preferences" (Ben-Akiva et al 1999, p. 188).

Retweeting is a choice made by individuals who receive tweets. Since individuals make choices regarding what actions to take based on the utility they realize from those actions, whether members of a brand's network retransmit or propagate messages depends on whether they conclude that retweeting provides a higher utility than being a passive recipient. Importantly, the utility of retransmission is intangible since the transmitter is neither paid nor realizes any other material benefit by transmitting a message from a brand. For example, Toubia and Stephen (2013) demonstrate that individuals are primarily motivated by two types of intangible utilities to contribute to Twitter: intrinsic and image.

The literature suggests several additional intangible benefits that motivate transmission: self-enhancement (De Angelis et al 2012), altruism (Hennig-Thurau et al 2004), and social interaction (Lovett, Peres and Shachar 2013). Further the magnitude of these benefits is greater when the number of others who may be interested in the message that the recipient can broadcast it to. For instance, Berger and Milkman (2012) (the narrowcasting vs broadcasting paper) suggest that "self-presentation motives, identity signaling (e.g., Berger and Heath 2007), or affiliation goals may play a stronger role in shaping what people share with larger audiences."

Our main thesis is that recipients of a message from a brand decide on whether or not to transmit it based on whether or not it provides any intangible utility and how big that utility is. As suggested by Berger and Milkman (2012), the benefits that they are most interested in should be self-enhancement and altruism since, being altruistic, increases their affiliation with the recipients who would feel obligated. For instance, a message that is more general about a category (e.g., autos) has greater utility of retransmission than one that is specifically about a brand in terms of how many recipients it can be forwarded to. Thus, categories to which brands belong to may play a bigger role than brands since the number of people interested in a category is always larger than the number interested in a brand of the category.

1.2.2. Theory

To assess the utility of retweeting, individuals need to evaluate the benefits that can be derived from a tweet. Tweets, however, are not well defined choice options due to many reasons. First, they are only 140 characters in length and short. Hence, their emotional content and, hence, their potential contribution to self-enhancement or altruism is difficult to evaluate. Further, the content of tweets is not predefined and no two tweets may be alike since, if they were, tweeters would not be sending multiple tweets. Hence, individuals cannot have a well-defined and rehearsed judgment process to evaluate each tweet. Consequently, individuals would have to rely on a constructive choice process (Bettman et al 1998) to assess the utility of retweeting a tweet.

The key difference between the classic and constructive view of preferences is that the former assumes that consumers have a "master list of preferences in memory when making a choice" (Bettman et al 1998, p.188) and also that "they apply some invariant algorithm such as a weighted adding model" (Bettman et al 1998, p.188) to evaluate choice options and select one. In contrast, the constructive view of preferences argues that consumers may not always have well-defined and rehearsed preferences for choice options. Instead, they may construct their preferences on the spot when confronted with the options. This is particularly likely when either the number of options is large or the knowledge of the options and the choice task is low. In addition to constructing preferences "on the fly" (Bettman et al 1998, p.188), decision-makers may also construct their approaches to evaluation on the fly. Thus, rather than carefully considering all the features of all the presented options, they may follow simple "algorithms" (Bettman et al 1998) such as choosing an option based on whether or not it has the best value on the most important criterion or a lexicographic decision-strategy. Alternatively, they may arrive at a choice based on some metagoals, which focus on realizing objectives other than utility maximization.

Bettman et al (1998) suggest four metagoals:

1. Maximizing the accuracy of a decision
2. Minimizing the cognitive effort required for the decision
3. Minimizing the experience of negative emotion while making the decision

4. Maximizing the ease with which a decision can be justified to others or to one's self

Bettman et al (1998) also discuss how different goals may become more prominent in different situations. In the case of irreversible actions of high importance, for instance, maximizing the accuracy of the decision, minimizing the experience of negative emotion, and maximizing the ease with which a decision can be justified to others or one's self may dominate. On the other hand, in situations where there is little involvement or extensive need to justify, maximizing the accuracy of the decision and minimizing the cognitive effort required may become the dominant goals (Beach and Mitchell 1978, Bettman et al 1998, Hogarth 1987, Payne, Bettman, and Johnson 1993, Shugan 1980)

We suggest that retweeting falls into the class of decisions where there is little involvement on the part of the consumer or extensive need to justify for multiple reasons. One, individuals receive multiple tweets per day. Hence, evaluating each tweet to decide whether to retweet would require substantial cognitive effort and time. Two, social media and the benefits from being active there is only one of the many tasks that individuals have during the day. Further, spending time on social media is a discretionary rather than a required activity like work. Three, retweeting a brand's tweet is neither a critical nor an irreversible decision with major consequences. Therefore, recipients are unlikely to want to put in extensive effort into the process of deciding which ones to transmit and which ones to passively react to. Overall, therefore, consumers would wish to minimize their effort but maximize the returns on that investment in terms of the accuracy in realizing the three benefits mentioned previously, i.e., self-enhancement, altruism, and social interaction. We briefly discuss below how specific characteristics of the brand or the tweet could help consumers increase the accuracy in realizing these benefits while minimizing their effort.

Accuracy in realizing the benefits can be achieved by relying on cues in the message to judge whether further transmission is likely to provide utility. The message cues are:

1. The brand itself - more popular brands may attract more attention from recipients and, hence, result in higher *social interaction*. This can be operationalized as the number of members in the social media networks of the brand, and how the network is growing.
2. The category itself – categories that are of interest to a greater number of people may attract more attention and hence increase the *social interaction*
3. The presence of characters or words that may indicate tangible benefits to recipients of the retweet can increase the benefits of *altruism* by retweeting. These include:
    a. Words related to promotions
    b. Presence of a $ sign
    c. Time related – sense of urgency indicating a promotion
    d. Action oriented words – “act now”, “take advantage” – again indicating a promotion or a potential benefit
4. The presence of characters or words that may indicate intangible benefits to recipients of the retweet can increase the benefits of *self-enhancement* by retweeting. These include:
    a. Presence of words indicating an event
    b. Presence of a link to additional content that the recipients may benefit from
    c. Presence of words related to the brand

The goal of effort minimization can be achieved by looking for the presence of specific characters in the tweet that serve as cues to the effort required, or effort that can avoided, to evaluate them. Characters that reduce the effort required to evaluate include:

1. The hashtag which makes it easier to decide whether or not the subject of the tweet would be of interest to recipients
2. Explicit request by the brand to retweet as indicated by the presence of the phrase “RT_if”
3. Presence of an exclamation mark suggesting something interesting

Effort minimization can also be achieved by relying on cues that suggest increased effort to retweet as a means of deciding not to retweet. These cues include

1. Presence of blanks that need to be filled in thus requiring more effort
2. Tweets that are long requiring more cognitive effort to process
3. Tweets that are from a brand that tweets with high frequency indicating that each tweet may not have much value
4. Presence of a question mark requiring the recipient to put in more effort into thinking about an answer to the question
5. How long ago the tweet was first sent – the longer it has been the less likely that it is current and, hence, less likely that it will be of interest when retweeted

We next discuss our data collection approach and operationalization of the variables.

1.3. Data

In this research, we consider data from a well-known social network, Twitter. We considered tweets originated from 62 brands across four product categories of "Auto", "Food/Beverages", "Dinning", and "Airline", and collected data through twitter API (Application Programming Interface). We chose these categories due to data availability and less fragmentation in terms of brands within each category. Initially, we considered top 20 brands in terms of number of followers in each category, where we collected 500 recent tweets as of 1 April 2012 for the selected 80 pages. However, not all the 500 tweets from brands have general audience, and generally, the tweets posted by brands have two types of audience: all followers or specific individuals. We only consider tweets posted to general audience, because mainstream of this research is to study retransmission of the tweets targeted to general audience. The exclusion of individual-specific tweets (reply tweets) from the sample resulted in some pages having very few tweets. After dropping the reply tweets, we did not consider brands with less than 50 tweets in our sample. In addition, we did not include pages posting tweets other than English language. Given all these conditions, we have 14163 tweets across 62 brands. Table 1.1 provides aggregate summary statistics across four categories.

Table 1.1 – Category Breakdowns

Categories	# of RT	# of tweets	Ratio	# of Followers	# of FB Fans
Auto	64400	5262	12.24	1,236,568	37,900,199
Food/Beverage	68254	3905	17.48	2,343,389	147,194,921
Dinning	67624	3529	19.16	1,683,801	63,086,630
Airline	17490	1467	11.92	2,661,346	5,799,490

Table 1.2 provides the brand names and number of tweets for each in our sample. The brands that are crossed out are not included in the final sample.

Table 1.2 – List of brand names for each category with the number of tweets considered for each brand (crossed brands have been dropped)

#	Auto Category	# of tweets	#	Food/Beverages Category	# of tweets
	~~Ferrari (Italian)~~	~~390~~		~~Starbucks Coffee~~	~~37~~
1	Audi	184	20	Pepsi	163
2	Ford	241	21	Red Bull	146
3	Chevrolet	122		~~Coca-Cola (Coke)~~	~~8~~
4	Toyota	362	22	Monster Energy	443
5	Nissan	309	23	Dunkin' Donuts	113
6	VW	181	24	Domino's Pizza	86
7	Harley-Davidson	448	25	Dr Pepper	98
8	Jeep	307	26	PepsiCo	62
9	BMW	53	27	Gatorade	286
10	Porsche	369	28	Mountain Dew	84
11	General-Motors	253	29	Ben & Jerry's	272
12	Honda	306	30	Wheat Thins	347
13	Dodge	294	31	Arizona Iced Tea	270
14	Chrysler	306	32	Oreo	129
15	Tesla	396	33	Sierra Nevada	203
16	Aston-Martin	416	34	Kraft	286
17	Cadillac	297	35	Bacardi	450
18	Mazda	267		~~Lipton Brisk Iced Tea~~	~~25~~
19	Hyundai	151	36	Skittles	467
	Dinning Category	**# of tweets**		**Airline Category**	**# of tweets**
37	Subway	197		~~JetBlue Airways~~	~~6~~
38	McDonald's	338	54	Southwest Airlines	151
	~~Taco Bell~~	~~10~~	55	Air Asia	194
39	Hard Rock Cafe	261		~~American Airlines~~	~~3~~
40	Chick-fil-A	137		~~TAM Airlines~~	~~13~~
	~~Pizza Hut~~	~~6~~	56	Virgin America	100
41	KFC	154	57	Delta Air Lines	275
	~~Wendy's Restaurant~~	~~26~~		~~KLM~~	~~5~~
42	Chili's Grill & Bar	371	58	British Airways	82
43	Papa John's Pizza	279		~~US Airways~~	~~2~~
44	Arby's	96		~~WestJet~~	~~42~~
45	P.F. Chang's	53		~~Lufthansa (German)~~	~~467~~
46	The Cheesecake Factory	96		~~Virgin Atlantic~~	~~15~~
47	Dairy Queen	302	59	Air New Zealand	65
48	Hooters	354		~~Alaska Airlines~~	~~10~~
49	Sonic Drive-In	103	60	Hawaiian Airlines	256
50	Outback Steakhouse	110		~~easyJet~~	~~21~~
51	California Pizza Kitchen	98	61	Lufthansa - USA	108
52	Popeyes Chicken	352	62	Air Canada	236
53	T.G.I. Friday's	228		~~Cathay Pacific~~	~~27~~

1.3.1. Dependent Variable

The dependent variable is the number of times the tweets posted by brands have been retransmitted. Our data collection is a one-time process, occurred 1 April 2012, and it is cross sectional. We considered the recent 500 tweets for each page as of 1 April 2012. Therefore, the posting date of the tweets could date back differently with respect to 1 April 2012, depending on the activity level of the pages. For example, a brand that posts frequently (i.e. 50 tweets per day), its 500 tweets will probably range a short period (i.e. 10 days), whereas for infrequent senders (i.e. 5 tweets per day), that range could go back several months (i.e. 100 days). This data collection mechanism results in unbalanced data with respect of number of days over which a tweet could have been retweeted. One potential problem with this data collection procedure is that tweets posted right before the data collection day might not get their ultimate retweet count; therefore, considering their retweet count as of observation day might bias the results. In order to overcome this issue, we re-observed the dependent variable, retweet count, a year after the initial data collection (1 April 2013) to make sure that all the tweets in our sample received their ultimate retweet count. Table 1.3 and Figure 1.2 provide summary statistics and histogram of dependent variable (for better visualization, histogram is shown to 100).

Table 1.3 – Summary statistics of dependent variable

Range of Retweet Count	**Frequency**
0	1332
(0,10]	8669
(10,100]	3904
(100,1000]	248
(1000,max]	10

Summary Statistics	**Retweet count**
Mean	15.38
Std	67.58
Min	0
1st Q	2
Median	5
3rd Q	12
Max	3131

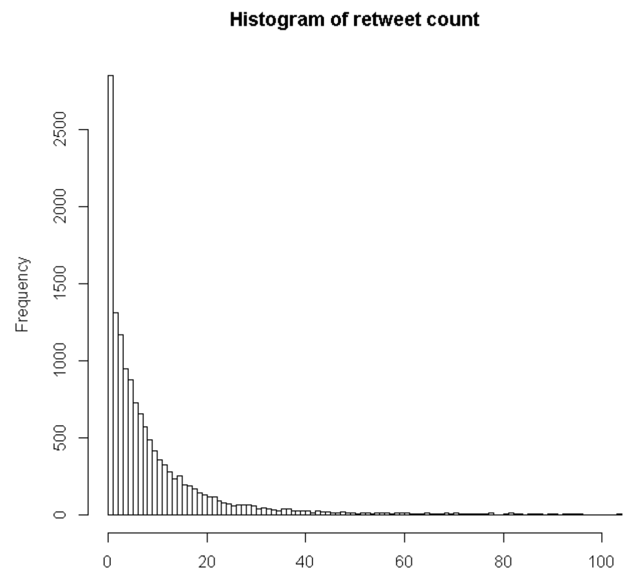

Figure 1.2 – Histogram of retweet count (up to 100 for better visualization)

1.3.2. Text Related Covariates

In this section, we discuss the construction of text-based covariates. We used two sets of covariates, each extracted differently. The first set consists of seven indicator variables constructed based on the presence of specific symbols or characters. Table 1.4 shows these variables as well as their description and summary statistics.

Table 1.4 – List of text related covariates based on presence of symbols with their description and their summary statistics in the data

Variable	**Description**	**Yes**	**No**
Http	If the tweet includes any link to a website	8810	5353
HashTag	If text contains any hashtag "#"; Hash tags are popular in twitter to make the word coming after hashtag traceable through search, and also gives the reader a general idea of the tweet topic	5764	8399
DollarSign	If the tweet includes any pricing information, measured by presence of "$" sign	529	13634
Blank	If the tweet contains a blank as " _______ ". Pages usually ask the audience to fill in the blank to generate engagement.	72	14091
Exclamation	Presence of "!" sign	6457	7706
Question	Presence of "?" sign	3123	11040
RT_if	If the twee specifically asks for retweet, and contains the "RT if …." expression	145	14018

The second set of constructed covariates has the objective of categorizing the tweet content. Methods such as bag of words and text classifiers have been popular. In the bag of words method, the procedure is to find the frequently used words, extract the factors based on the co-occurrence of those words, and subsequently cluster the factors to categorize the content. This method is not effective in this study since twitter only allows 140 characters; hence, the co-occurrence of related words is less probable, leading to inaccurate factors and clusters. In the classifying method, the task is to label a class based on the content. It usually starts with a training set, where the classifier will learn and subsequently it will be able to classify the data.

The approach we used has the advantage of being simple to implement from the brand perspective, and interpretable from the researcher side. We construct indicator variables based on presence of some category of words in the tweet text. We took the following preparation steps before the procedure for constructing the text categories.

1- Removing “@” mentions and punctuation
2- Removing http link in the content
3- Removing stop words
4- Converting all words to lower case
5- Stemming all words since there might be several versions of the same word

Upon completion of these tasks, we extracted the most frequent unique words, as well as the combination of two and three words. Based on the list of words, we identified specific categories where presence of some frequently used words identifies those categories. A potential problem with this approach is that two categories could happen at the same time, whereas using a classifier provides the probability that a tweet belongs to a category, and based on the highest probability, it categorizes the tweet. Table 1.5 shows the words used for constructing indicator variables for each category.

Table 1.5 – List of extracted categories based on the used words

Category	**Words used as the category identifier**
Promotion	free, chance, sale, commercial, giveaway, promo, chance win, give away, gift card, enter chance, enter win, win trip, last chance, chance 2, enter chance win, no purchase necessary, chance 2 win, send gift card
Brand Mentioned	Any explicit mention of the brand on the tweet
Time Mentioned	day, today, now, year, week, weekend, tomorrow, tonight, Friday, night, lunch, Monday, month, morning, April, breakfast, Saturday, Sunday, Tuesday, winter, yesterday, last night, new year, right now, last week, next week, every day
Action Requested	check, watch, join, vote, tell us, tweet us, let us, check new, check video, click here, vote favorite, let us know
Event Oriented	game, birthday, Event, holiday, NCAA, valentine, super bowl, big game, spring break, St Patrick' day

1.3.3. Additional Tweet Characteristics

In addition to text covariates, we controlled for the posting weekday and daytime of the tweet, the time elapsed after the posting time of the tweet to the observation date, and the number of characters in the tweet. The posting weekday and daytime has an effect on the visibility of tweets. If the posting time is such that less people are likely to see the tweet, then the overall retransmission of the tweet decreases. Since the posting time varies based on time zone, we unified the posting time based on U.S. Central time zone, and constructed the weekday that tweet originally posted as well as the time of the day that tweet posted. We categorized daytime variables into four hours intervals, hence six categories. Table 1.6 shows the frequency of observations for weekdays and daytime. In addition, the time elapsed after posting the tweet is measure based on the secondary observation day, 1 April 2013. Table 1.7 have the summary statistics of these variables.

Table 1.6 – Frequency of weekday and daytime variables across their levels

Weekday	Frequency
Friday	2361
Saturday	1274
Sunday	1111
Monday	2155
Tuesday	2394
Wednesday	2531
Thursday	2337

Daytime	Frequency
time 0-4	248
time 4-8	487
time 8-12	4721
time 12-16	5119
time16-20	2618
time 20-24	970

Table 1.7 – Summary statistics of additional tweet level covariates

	Mean	Std	Min	1st Q	Median	3rd Q	Max
NumCharacter	109.40	28.29	9.00	91.00	118.00	134.00	147.00
LOG(DaysElapsed)	6.06	0.17	5.86	5.93	6.01	6.13	6.69

1.3.4. Brand Related Variables

We observed the network characteristics of the brands on a daily basis. For a given tweet, we observed the following variables (acquired from fanpagelist.com) and their summaries are in table Table 1.8.

- Number of followers (*NumFollower*): total number of individuals signed up to receive tweets from the page as of the day tweet has been posted
- Number of followers gain (*NumFollowerGain*): number of individuals signed up to receive tweets from the page on each day
- Number of Facebook fans (*NumFacebookFan*): total number of individuals who liked the page on Facebook as of the day tweet has been posted
- Number of Facebook fans gain (*NumFacebookFanGain*): number of individuals signed up to receive Facebook updates from the page on each day
- Number of following (*NumFollowing*): the total number of pages that the brand is following on twitter

Table 1.8 – Summary statistics of brand network size

	Mean	**Std**	**Min**	**1st Q**	**Median**	**3rd Q**	**Max**
LOG(NumFollower)	10.76	0.95	8.94	10.11	10.54	11.12	14.07
LOG(NumFollowerGain+200)	5.75	0.49	3.76	5.48	5.60	5.82	9.29
LOG(NumFaceFan)	14.07	1.49	10.19	13.14	14.22	15.02	17.15
LOG(NumFaceFanGain+2000)	8.45	0.75	6.70	7.83	8.25	8.85	11.83
LOG(NumFollowing)	8.03	1.95	1.79	6.56	8.50	9.65	10.89

1.3.5. Additional Brand Variables

We considered social media presence of the brands on few other well-known social networks, Google+, YouTube, Instagram, Pinterest, and incorporated this data as four indicator variables in the analysis. In addition, we included a dummy variable defined as the presence of brand on the Interbrand list of 2012. This well-known list includes rankings of the top 100 global brands based on their estimated brand equity. In general, brand equity is the sum of unique attributes of brands that affect the marketing effectiveness of a product. There are multiple ways to estimate the brand equity. The estimated brand equity by Interbrand consists of several internal factors (clarity, commitment, protection, and responsiveness) and external factors (authenticity, relevance, differentiation, consistency, presence, and understanding). Another characteristic of brand in our data is the tweeting frequency of the brand page, defined as the number of tweets posted on each day. It is day level data and identical for the tweets posted on the same day (*TweetFreq*).

Table 1.9 – Summary Statistics of additional brand variables and tweeting frequency

Social Network	Google+	YouTube	Instagram	Pinterest	Interbrand
Frequency	27	44	15	18	13

	Mean	**SD**	**Min.**	**1st Q**	**Median**	**3rd Q**	**Max**
LOG(TweetFreq)	0.739	0.688	0.000	0.000	0.693	1.099	3.497

1.4. Modeling Approach

Since the dependent variable is the number of retweet count, therefore, we employed an empirical Poisson model for our modeling approach. However, due to possible over-dispersion, we accounted for extra variation in the model through a lognormal mixture, which affects variance but not mean. In addition, we used a hierarchical model, since we have covariates in multiple levels. In specific, let the $NumRT_{ijkt}$ to be observed retweet count of tweet i belongs to brand j in category k posted on day t, then we have

$$NumRT_{ijkt} \sim Poisson(\lambda_{ijkt}\,\epsilon_i)$$

The parameter ϵ_i is the mixing component having a lognormal distribution with as $\log(\epsilon_i) \sim Normal(0, \tau_\epsilon)$. This term will capture the over-dispersion. The mean of the Poisson distribution, λ_{ijkt} relates to covariates through a log link.

$$\log(\lambda_{ijkt}) = Category.effect_k + Brand.effect_{jt} + \overline{tweet.charactristics}_{ijk}\bar{\beta}$$

The first component is the category specific intercept; the second component is the tweet characteristics, which includes all the tweet level variables described in section 1.3.2 and 1.3.3. The vector $\bar{\beta}$ is the effect of these covariates. The third component is the effect of brand j on day t that relates to network covariates as below,

$$Brand.effect_{jt} = \overline{Network.effect}_{jt}.\bar{\gamma} + Tweeting.Frequecny_{jt}\,\eta + Brand_j$$

The vector $\overline{Network.effect}_{jt}$ includes the variables in section 1.3.4, and the vector $\bar{\gamma}$ is its respective parameters. The component $Brand_j$ is the role of brand's social media presence and Inter-brand as below,

$$Brand_j = \overline{Social.Media.Presence_j}\,\bar{\eta} + Interbrand_j\theta + \delta_j$$

The vector $\overline{Social.Media.Presence_j}$ includes the four indicator variables described in 1.3.5. The variable $Interbrand_j$ is the presence of brand on the Interbrand list, and θ is its respective effect. In addition, we included a brand specific random effect δ_j, which captures unobserved characteristics of brand j. Since the assumed distribution for the random effect affects estimated brand effects, we assume a nonparametric Dirichlet Process Prior (DPP) for the brand random effect. Under the DPP assumption, the distribution of δ_j is an unknown distribution G with its average equals to G_0, called the initial guess. A precision parameter α, represent our confidence about the initial guess, such that as $\alpha \to \infty$ then $G_0 \to G$. In specific,

$$\delta_j \sim G$$

$$G \sim DPP(\alpha\,, G_0)$$

A formal definition of DPP by Ferguson (1974) is as follows: For finite k and any measurable partition $(A_1, A_2, \dots, A_k)$ of R, the distribution of $G(A_1), \dots, G(A_k)$ is $Dirichlet\big(\alpha G_0(A_1), \dots, \alpha G_0(A_k)\big)$. The value k represents the maximum number of partitions. In a general DPP model, the values of α and G_0 are unknown to researcher, and will be estimated in the model by assigning priors. In general, the choice of α is related to the estimated number of clusters k, in a way that a smaller value for α leads to a smaller number of clusters and vice versa. We adopted a gamma distribution for α with the mean of two and variance of 20. This prior accounts both for large and small values of α, and is weakly informative. The choice of G_0 is not critical since both large and small values of α has been accounted for. We employed a normal distribution with an unknown precision parameter for the baseline distribution, where a weakly informative prior has been adapted for the precision parameter of the baseline distribution. By choosing these specifications, the induced upper value for the number of clusters is equals 62, number of brands.

1.4.1. Model Selection

We considered models with different specifications as competing models, and compared their performance through several measures. The competing models are below.

M1: Poisson – Lognormal with DPP brand random effect

M2: Poisson – Gamma Model – $Gamma(r, r)$ for the distribution of ϵ_i

M3: Poisson Log-normal without brand random effect

M4: Poisson Model – No over-dispersion, $\epsilon_i = 1$ for all i

We compare the models based on the DIC measure of fit as well as predictive assessment of the models, measured by mean square error of prediction and posterior predictive density check. The DIC in a model with parameters θ and the random effect u is calculated as follows (Ntzoufras 2011).

$$DIC = 2\overline{D(\theta, u)} - D(\bar{\theta}, \bar{u})$$

The first term is the posterior mean of the deviance (-2×Log Likelihood), and the second term is the deviance, evaluated at the posterior mean of the parameters. The mean square error is

$$MSE(\hat{y}, y) = \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n}$$

The posterior predictive test is a model-checking procedure based on the posterior predictive density. Model checks assess whether the model satisfactorily reproduces certain important aspects of the actual data. In specific and for the over-dispersed count data, the model should reproduce such features in the replicates $\hat{y}_i$, sampled from the model. Suppose $C(Y;\theta)$ is the observed criterion of the actual data (the ratio of observed variance to mean in this case) and let the same criterion based on new data be denoted as $C(\hat{Y};\theta)$. The comparison of these features in MCMC will give a probabilistic value for the discrepancy of model and data. In specific, in each MCMC iteration, the average of the calculated probability (6) over all iterations gives the measure of density check.

$$\hat{P}_c = \frac{\sum_{t=1}^{T} 1\{C(\hat{Y};\theta) > C(Y;\theta)\}}{T}$$

The function $1\{\,\}$ is an indicator function and T is the length of MCMC samples. Values of $\hat{P}_c$ near zero or one (above 0.9 or below 0.1) indicate discrepancy between the observations and the model. Values relatively close to 0.5 mean that actual data and new data sampled from the model are comparable in terms of the above feature

1.4.2. Model Estimation

The proposed model has been estimated using Bayesian approach and by using BUGS. The prior distributions for the coefficients are proper but not-informative (Normal with mean zero and large variance), and for the two precision terms, Gamma distribution with mean one and large variance has been considered. The estimation is based two chains with 5000 burin, 5000 samples with thinning of 10. Convergence assured graphically.

1.5. Results

Table 1.10 shows the three criteria we use to compare fit and predictive performance of our models. Based on these measures, model 1 (M1) performs better in terms of all three measures. The second model, Poisson Gamma with brand random effects has a poor performance on $\hat{P}_c$, which means the model cannot fully capture over-dispersion. The third model is Poisson Lognormal with brand random effects. This model also performs very well, but in terms of DIC, it is not performing as well as M1. The fourth model is the regular Poisson model without any mixture, and as expected, it is performing very poorly. In sum, we chose model 1, and next, we present results from this model.

Table 1.10 – Model Comparison Results

	DIC	MSE	$\hat{P}_c$
M1	73444	30.69	0.366
M2	73653	32.56	0.026
M3	73660	30.73	0.379
M4	272300	3850.12	0.000

Based on the estimated mean and 95% probability intervals of the parameters presented on Table 1.11, the "Auto" category has the highest retweet volume among the four categories, and subsequent to that, "Food/Beverage" and "Dining" are the second and third categories. Interestingly, "Airline" is the last category in terms of retweet volume. Extent of brand's popularity on social media also has a significant role on retweet count. Almost all of the five social media presence variables have a significant positive effect on retweet volume, except number of following which is not significant. Social media presence variables have some diverse effects. Presence on Google+ has a positive effect on retweet volume, whereas presence on YouTube and Instagram has negative effects. Brands mostly promote their YouTube videos and Instagram photos on Twitter, therefore, the results of such negative effect could be due to irrelevance of tweets to their audience, assuming that they include links to YouTube video or Instagram photos. In other words, individual may find such links to have less benefit for them to retweet, although we cannot explain this effect with great confidence, since we do not observe cross social network performances. Presence of brand on Interbrand list also does not have a significant effect of Twitter performance. Among the text-related variables that signal altruistic benefits, as expected, the effect of “Promotion”, “Dollar Sign”, and “Time Mentioned” are positive and significant. However, “Action Oriented” variables does not have a significant effect on retweet count. Among the content-related variables that signal self-enhancement benefits, brand mentions and presence of a link have significant positive effects. In addition, frequency of tweeting have a negative effect, since more tweets per day make a particular one to be irrelevant to the audience, hence it reduces the overall benefits individuals might gain by retweeting. Presence of event related content is not significant.

The effect of effort on retweeting has been examined in two categories of effort minimization and effort avoidance. Presence of a hashtag has a positive effect on retweets. Hashtag are usually used to highlight a keyword in a tweet, and its presence also makes the hashtag phrase to stand out against other words, hence it grabs the attention and reduces the effort to process information. Presence of “RT_if” in a tweet has a positive effect on retweeting, since it reduces the information processing required to make retweet decision. Presence of an exclamation mark does not have a significant effect on retweeting, but its excessive use as captured by number of exclamation marks has a negative effect. The three variables of effort avoidance, all have negative and significant effect on retweeting. Presence of brank and question mark require additional effort to make the tweet ready for retweeting; therefore, receivers may avoid such efforts and do not retransmit the tweet. In addition, higher tweet length leads to less retweets, since a longer text needs more effort for processing, and individuals might avoid reading a long tweet. Although a longer tweet means more information, but considering huge amount of tweets individuals receiving every day, they might want to avoid longer tweets. Regarding the control variables on time of the day and weekday, based on the estimated coefficients and base level of Friday, it seems almost every weekday will work better than Friday and Saturday, and Sunday is the best time to get more retweets. That might be due to less content competition on Sundays. In terms of time of the day, it seems only 4 to 8 in the morning is the best time to tweet.

Figure 1.3 shows density of the estimated posterior mean of random effects across the brands. In addition, Figure 1.4 shows boxplots of the random effects' posterior distributions for all brands. It is interesting the there are few brands that are on the positive and negative side of the distribution, and most of the brands do not have significant differences in terms of their unobservable characteristics, and their differences are captures by the included variables in our model.

Table 1.11 – Estimated coefficients of the model

Variable	**Mean**	**2.50%**	**97.50%**
Category – Auto *	1.75	1.44	2.03
Category – Food/Beverage *	1.56	1.37	1.74
Category – Dinning *	1.33	1.16	1.53
Category – Airline *	1.20	0.94	1.44
LOG(NumFollower) *	0.37	0.30	0.46
LOG(NumFaceFan) *	0.13	0.05	0.21
LOG(NumFaceFanGain) *	0.04	0.01	0.07
LOG(NumFollowerGain) *	0.18	0.15	0.20
LOG(NumFollowing)	-0.06	-0.12	0.01
Google+ *	0.37	0.24	0.48
YouTube *	-0.25	-0.42	-0.08
Instagram *	-0.44	-0.60	-0.25
Pinterest	0.11	-0.05	0.27
Interbrand	0.08	-0.11	0.31
Promotion *	0.36	0.28	0.43
Dollar Sign *	0.69	0.59	0.80
Time mentioned*	0.20	0.16	0.24
Action	-0.01	-0.07	0.06
Event	-0.08	-0.17	0.01
Brand *	0.22	0.17	0.26
Http *	0.12	0.08	0.16
LOG(NumTweet) *	-0.23	-0.25	-0.21
Hash tag *	0.05	0.01	0.09
RT_if *	1.63	1.45	1.81
Exclamation	0.02	-0.06	0.12
# of Exclamation *	-0.09	-0.17	-0.02
Blank *	-0.37	-0.63	-0.09
Question *	-0.13	-0.18	-0.08
NumCharacter *	-0.16	-0.18	-0.13
Saturday	-0.01	-0.09	0.07
Sunday *	0.17	0.09	0.25
Monday *	0.15	0.09	0.22
Tuesday *	0.11	0.04	0.18
Wednesday *	0.13	0.07	0.19
Thursday *	0.07	0.01	0.14
Days Elapsed *	-0.07	-0.10	-0.04
Time – [0,4)	0.00	-0.16	0.15
Time – [4,8) *	0.15	0.03	0.28
Time – [8,12)	0.03	-0.05	0.13
Time – [12,16)	-0.03	-0.11	0.07
Time – [16,20)	0.02	-0.07	0.12

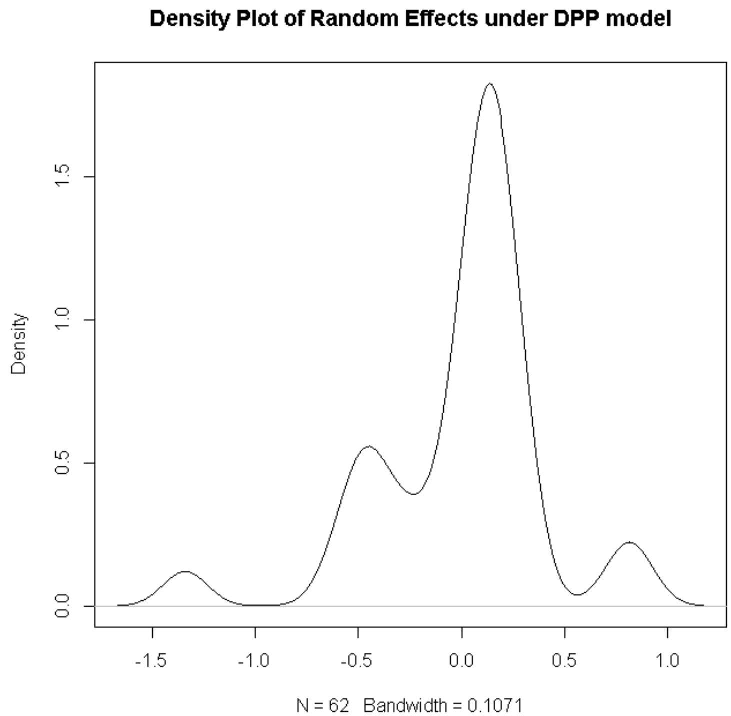

Figure 1.3 – Density ploy of the estimated random effects for the distribution of brand random effects

1.5.1. Other parameters of Model

In addition to the model coefficients, Table 1.12 shows the estimated parameters of the DPP model as well as precision parameters of the lognormal mixture. The estimated DDP parameters suggest that with a high degree of belief (9.13) distribution of random effect is a normal distribution with precision parameter of 3.4.

Table 1.12 – Other Parameters of the model

Variable	Mean	2.5%	97.5%
α	9.136	1.812	28.88
τ_G	3.404	1.005	7.237
τ_ϵ	1.003	0.976	1.033
K	17.390	8	31

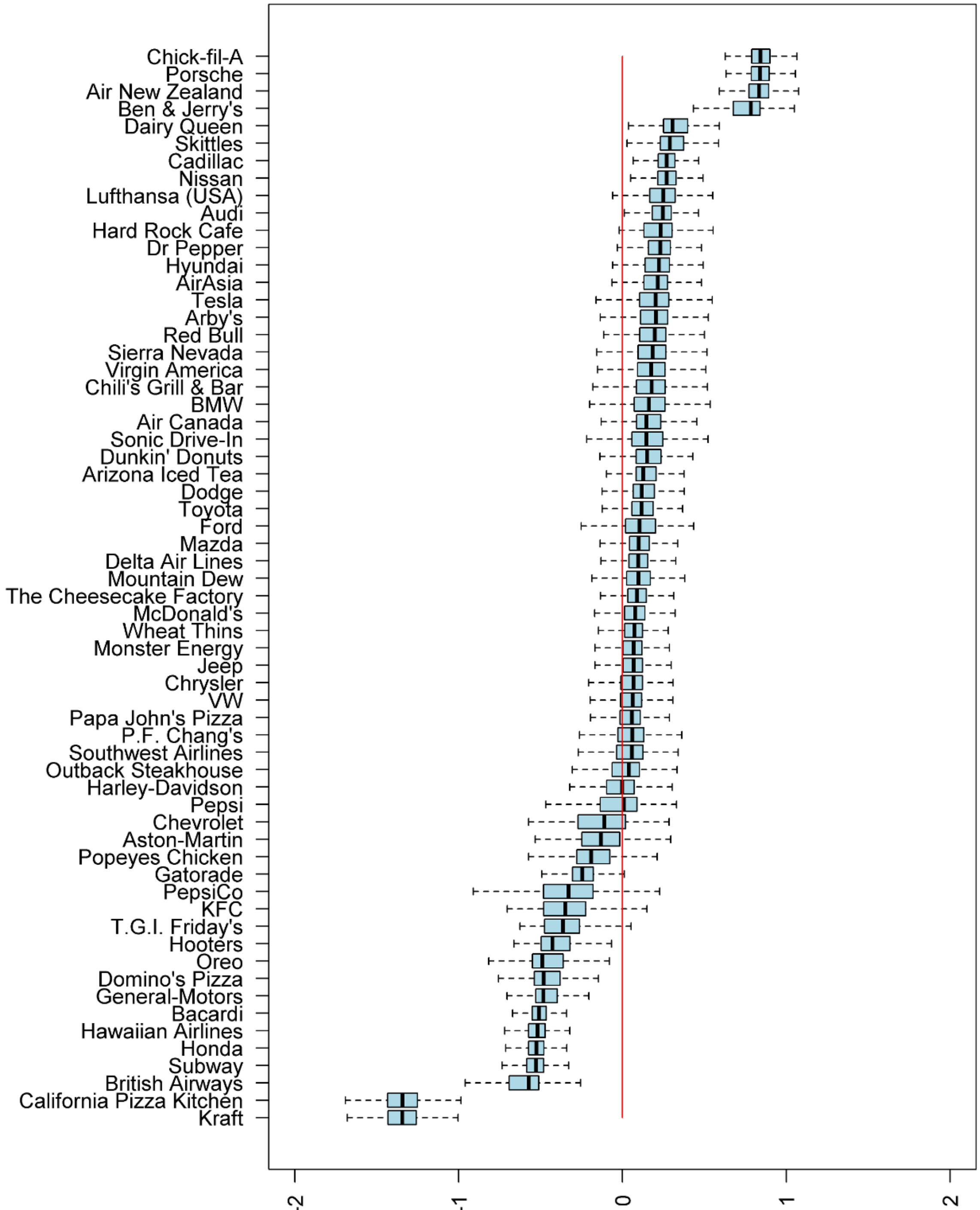

Figure 1.4 – Boxplot estimated random effects across brands

1.6. Discussion

In this research, we argued that individuals construct their retransmission choice in Twitter based on two metagoals of benefit maximization and utility minimization. In addition, by employing more than 14000 tweets posted by 62 brands, we articulated satisfying two metagoals of benefit maximization and effort minimization based on brand and tweet characteristics. We demonstrated that tweet characteristics that signal benefit to the receivers in the form of social interaction, self-enhancement, and altruism would have a positive effect on retransmission of tweets posted by brands. In addition, we showed that message cues that reduce the effort required for processing tweet content would have positive effects on retransmission.

1.6.1. Managerial Implication

Our research provides important managerial implications for brands. By building a larger network and having a stronger presence on social media, brands could potentially increase retransmission of their messages. In addition, brands in popular and relevant (to everyone) product categories that provide social interaction benefits will have a higher chance for their messages to be retransmitted, hence using social media could benefit them more than other, less popular product categories. Our results will also have implications for content strategies on social media. By identifying content categories that provide tangible and intangible benefits for receivers to share them with their followers, brands can increase effectiveness of their social media strategies for retweeting. Furthermore, by incorporating strategies that reduce the effort require to process information will also enhance their social media efforts in terms of retransmission of their messages.

1.6.2. Limitation and Future Research

In this research, we are limited to our methods and data. First, we did not fully captured context of tweets in which they have been posted. We only accounted for presence of specific words that signal benefits to the individuals, while the overall context of the tweet might also has a role on retransmission. In addition, social media world as well as Twitter is evolving, therefore, individual's behaviors and habits change, which subsequently affects how firms used these medium for marketing purposes. Therefore, verifying such effects through several studies will also help to understand the change and evolving aspect of social media. Furthermore, our results are bounded by product categories we considered, and considering multiple product categories and their properties that could lead to more retransmission is an interesting venue for further research.

References

AdWeek. 2014a. "Arby's Slayed the Grammys With This Tweet About Pharrell Williams' Hat." Accessed April 16. http://www.adweek.com/adfreak/arbys-slayed-grammys-tweet-about-pharrell-williams-hat-155237.

AdWeek. 2014b. "Arby's Dishes on Awesome Pharrell Williams Tweet.". Accessed April 16. http://www.adweek.com/news/technology/arbys-dishes-awesome-pharrell-williams-tweets-156149.

Beach, Lee Roy, and Terence R. Mitchell. 1978. "A Contingency Model for the Selection of Decision Strategies." Academy of Management Review 3 (3): 439–49.

Ben-Akiva, Moshe, Daniel Mcfadden, Kenneth Train, Joan Walker, Chandra Bhat, Michel Bierlaire, Denis Bolduc, et al. 2002. "Hybrid Choice Models: Progress and Challenges." Marketing Letters 13 (3): 163–75.

Berger, Jonah, and Chip Heath. 2007. "Where Consumers Diverge from Others: Identity Signaling and Product Domains." Journal of Consumer Research 34 (2): 121–34.

Berger, Jonah, and Katherine L Milkman. 2012. "What Makes Online Content Viral?" Journal of Marketing Research 49 (2): 192–205.

Bettman, James R., Mary Frances Luce, and John W. Payne. 1998. "Constructive Consumer Choice Processes." Journal of Consumer Research 25 (3): 187–217.

De Angelis, Matteo De, Andrea Bonezzi, Alessandro M Peluso, Derek D Rucker, and Michele Costabile. 2012. "On Braggarts and Gossips: A Self-Enhancement Account of Word-of-Mouth Generation and Transmission." Journal of Marketing Research 49 (4): 551–63.

Deighton, John, and Leora Kornfeld. 2009. "Interactivity's Unanticipated Consequences for Marketers and Marketing." Journal of Interactive Marketing, Anniversary Issue, 23 (1): 4–10.

Ferguson, Thomas S. 1974. "Prior Distributions on Spaces of Probability Measures." The Annals of Statistics 2 (4): 615–29.

Fournier, Susan, and Jill Avery. 2011. "The Uninvited Brand." Business Horizons, SPECIAL ISSUE: SOCIAL MEDIA, 54 (3): 193–207.

Guadagni, Peter M., and John D. C. Little. 1983. "A Logit Model of Brand Choice Calibrated on Scanner Data." Marketing Science 2 (3): 203–38.

Hassan Zadeh, Amir, and Ramesh Sharda. 2014. "Modeling Brand Post Popularity Dynamics in Online Social Networks." Decision Support Systems, Crowdsourcing and Social Networks Analysis, 65 (September): 59–68.

Hennig-Thurau, Thorsten, Kevin P. Gwinner, Gianfranco Walsh, and Dwayne D. Gremler. 2004. “Electronic Word-of-Mouth via Consumer-Opinion Platforms: What Motivates Consumers to Articulate Themselves on the Internet?” Journal of Interactive Marketing 18 (1): 38–52.

Hogarth, Robin M. 1987. Judgement and Choice: The Psychology of Decision (2nd Ed.). Vol. xii. Oxford, England: John Wiley & Sons.

Hoyer, Wayne D., and Steven P. Brown. 1990. “Effects of Brand Awareness on Choice for a Common, Repeat-Purchase Product.” Journal of Consumer Research 17 (2): 141–48.

Lovett, Mitchell J., Renana Peres, and Ron Shachar. 2013. “On Brands and Word of Mouth.” Journal of Marketing Research 50 (4): 427–44.

Malhotra, Arvind, Claudia Kubowicz Malhotra, and Alan See. 2012. “How to Get Your Messages Retweeted.” MIT Sloan Management Review 53 (2): 61–66.

McFadden, Daniel. 1973. “Conditional Logit Analysis of Qualitative Choice Behavior.” Frontiers in Econometrics, 105–42.

———. 1980. “Econometric Models for Probabilistic Choice among Products.” The Journal of Business 53 (3): S13–29.

———. 2001. “Economic Choices.” The American Economic Review 91 (3): 351–78.

Muñiz Jr., Albert M., and Hope Jensen Schau. 2011. “How to Inspire Value-Laden Collaborative Consumer-Generated Content.” Business Horizons, SPECIAL ISSUE: SOCIAL MEDIA, 54 (3): 209–17.

Ntzoufras, Ioannis. 2011. Bayesian Modeling Using WinBUGS. John Wiley & Sons.

Payne, John W., James R. Bettman, and Eric J. Johnson. 1993. The Adaptive Decision Maker. Cambridge University Press.

Schau, Hope Jensen, Albert M Muñiz, and Eric J Arnould. 2009. “How Brand Community Practices Create Value.” Journal of Marketing 73 (5): 30–51.

Schulze, Christian, Lisa Schöler, and Bernd Skiera. 2013. “Not All Fun and Games: Viral Marketing for Utilitarian Products.” Journal of Marketing 78 (1): 1–19.

Shugan, Steven M. 1980. “The Cost of Thinking.” Journal of Consumer Research 7 (2): 99–111.

Singh, Sangeeta, and Stephan Sonnenburg. 2012. “Brand Performances in Social Media.” Journal of Interactive Marketing 26 (4): 189–97.

Toubia, Olivier, and Andrew T. Stephen. 2013. “Intrinsic vs. Image-Related Utility in Social Media: Why Do People Contribute Content to Twitter?” Marketing Science 32 (3): 368–92.

Huffington Post. 2014 “Twitter Triumph: 100% of Fortune 500 Will Soon Tweet.” Accessed April 16. http://www.huffingtonpost.com/adam-wexler/twitter-triumph-100-of-fo_b_4235213.html.

Essay 2: Source Effects In Social Media: How Celebrities Affect Lives of Their Messages

2.1. Introduction

Social media evolved into an important communication channel for brands by enabling marketers to communicate messages directly to customers. (Mangold and Faulds 2009). However, due to customer-to-customer relationships on social networks, marketers gain addition reach for their messages through their followers as receivers of the messages, sharing them with their followers. This sharing mechanism spreads brand’s message beyond its audience, and brands gain social advocacy through individuals sharing their messages (Malhotra et al 2012).

Rather than sending messages through brand’s social media accounts, marketers also employ celebrities in order to promote messages about their brands. The huge follower base of celebrities provides significant exposure for brands and their messages. For instance, celebrity-sponsored tweets where celebrities tweet about brands to their followers on Twitter are quite common today. In addition to reaching celebrities’ followers, brands will get additional reach for their messages, through retransmission of message by celebrities’ followers. For instance, on Mother’s day in May 2013, a tweet by singer Justin Bieber about the flower delivery service, 1-800-FLOWERS, was retweeted over 75 thousand times over multiple days. Given the widespread use of social media and celebrities by brands, there are few insights regarding the role of celebrities on how messages sent by them are shared on social media. Specifically, how does celebrity attributes such as number of followers affect volume and duration of sharing? If such an effect exists, given popularity of celebrities and their role as source of the message, the question of whether celebrity attributes have any role on how message characteristics are perceived by receivers, which subsequently affects volume and duration of sharing, is critical. Clearly, understanding role of celebrities and message characteristics on how messages are retransmitted leads to a development of better celebrity sponsored campaigns in the case of promoting messages on social media. In this research, we examine effects of tweets characteristics and celebrity attributes on how often and how long the tweets posted by them are retweets. In addition, we investigate how celebrities as source of message moderate effects of content characteristics on extent of the contents being shared.

In order to develop our conceptual framework, we consider retweeting process in underlying behavior of individuals who decide to retransmit tweets on social media. Those individuals involve with a choice-making situation, where they decide to share a content based on benefits of two options, retweeting or being passive (not retweeting). In general, choice characteristics reinforce how individuals process information in order to make a decision. Choice characteristics in the context of retweeting are tweet attributes. However, tweets' contents do not follow a standard format, and their attributes are unknown to individuals prior to observing them. Thus, the process of making retransmission decision in such a situation is constructive (Bettman et al. 1998). Based on this choice situation, we argue that two meta-goals of maximizing benefits and minimizing effort are desired. In this choice-making process (retweeting), individuals rely on message cues to evaluate content (tweet) and construct their choice criteria, in order to evaluate benefits of retransmitting a message, and minimize their efforts to process information. Therefore, tweet characteristics that signal utility for retweeting and require less effort to process will have a positive effect on retweeting and vice versa.

Context of celebrities posting tweets is a source-dominated communication, where source of a message has an impact on how individuals react to the message. The reason for such a source domination is their popularity and attractiveness to their followers. Source effect theory has been an extensive area of research in marketing (Wilson and Sherrell 1993). Several studies examined effect of source characteristics such as trustworthiness, expertise, and attractiveness on persuasion, attitude, purchase intention, and likeability (Petty, Cacioppo, and Goldman 1981, Harmon and Coney 1982, Harkins and Petty 1987). For instance, Kang and Herr (2006) showed that in situation where the cognitive resources to process information is low, because of either low involvement or low ability, the positive source characteristics would positively influence product attitudes. By relying on source effect theory, we develop our theoretical framework, where we argue that in the context of twitter, where involvement with a tweet is low, source characteristics affect how a tweet is perceived. In the case of tweets posted by celebrities, we argue that celebrity's popularity affects how individuals process tweet content, which subsequently affects their decision to retweet. Specifically, we argue that celebrity's popularity, measured by number of followers, will moderate the effect of tweet attributes (message cues) on the extent the tweet is retransmitted.

Our empirical analysis is based on 2935 tweets posted during March 2013 by 65 celebrities across seven categories of entertainment industry: actors, musicians, TV hosts, journalists, bloggers, athletes, and models. We considered top 10 celebrities in each category in terms of number of individuals who follow them. In addition to volume of retweets, we also considered duration over which the retweets occurred. Both measures are important for marketers. The volume of retweet is a measure of how many individuals shared a tweet, while duration of retweet measures the lifetime of the message on social media, and is measured as the time between posting a tweet and its last retweet. Clearly, the more a tweet lives on social media, the more attention it gets from prospective receivers, hence, there is a higher chance it will be retransmitted again, which subsequently increases the duration of retweets. However, in theory, tweets could be retweeted at any time, and since duration of retweets is calculated based on the last retweet, therefore, a retweet at any time will change the duration of retweets. As a result, observing the "true" duration might be challenging. In order to observe duration of retweets, we employed a data collection algorithm with two fixed intervals, where the tweets are tracked in the first interval to observe their last retweet, and their last retweet is confirmed in the second interval, by not getting any addition retweet in the second interval. If retweets occurred both in the first and second interval, we consider the duration as censored with respect to the end of first interval. The volume of retweet is observed along with the duration through this censoring mechanism. The primary celebrity attribute that we considered is the number of followers, but we also controlled for other, unobserved, attributes via random effects. Our tweet attributes include tweet type (Retweet and Reply), length of tweet in characters, presence of a hyperlink, hashtag, exclamation mark, and question mark. We use a Poisson-Lognormal model for the volume of retweets and a Weibull proportional hazard model for duration of retweets, and the two models are estimated jointly in a Bayesian framework.

Our results suggest that popularity of a celebrity measured by number of followers affects both volume and duration of retweets. However, there are additional celebrity characteristics that are not captured by number of followers, which subsequently affects volume and duration of retweets. Tweet types of reply and retweet will significantly reduce the volume and duration of retweets. Among the variables related to tweet content, question mark and presence of link decreases the volume and duration of retweet. Presence of exclamation mark in a tweet only affects volume of retweet, but not duration of retweet. The effect of hashtag is significant both through its main effect and its interaction effect with the number of followers. That is, celebrities with more number of followers will benefit more by including hashtag on their tweets. In addition, the association of volume and duration is relatively strong.

Our research has several managerial implications for firms. It provides insight to brands on role of celebrities, their number of followers, and their unobservable characteristics on how tweets posted by them will be retransmitted on social media. Firms could use these insights for sponsoring celebrities for marketing purposes. In addition, by considering the general implications of source effect on how message attributes are perceived, firms could investigate message types and attributes that fit with the brand image and firm's overall communication strategies, and subsequently employ them in their content strategies. For example, a specific hashtag might perform better with a brand than others since it aligns better with the brand's image.

In the next section, we discuss our theoretical framework, and next, we discuss our data and modeling approach. Finally, we discuss results and managerial implications, and conclude by limitation and direction for future studies.

2.2. Theoretical Background

Volume and duration of retweets on social media underlie in behavior of individuals who decide to share contents. As a result, understanding decision-making process of individuals who involve in this process is critical. In choice making situations, in general, individuals take actions based on utility of options (Guadagni and Little 1983). In the case of retweeting, such action is sharing the content, utilities of sharing are self-enhancement (De Angelis et al 2012), social interaction (Lovett, Peres and Shachar 2013), and altruism (Hennig-Thurau et al 2004), and the options are either retweeting, or staying passive (not retweeting). Thus, in the case of a retweet, we argue that utility of retweeting is higher than being passive. Therefore, evaluating utility of retweeting is a critical component of this choice-making situation for individuals who are involved with this process. However, there are several ways that individuals could evaluate utility of actions, and it primary depends on choice attributes. The choice attributes that individuals have to evaluate in case of retweeting are tweet attributes. Although a tweet has a maximum of 140 characters, but its content varies significantly based on intention of sender, therefore, its features are not expectable by receivers prior to observing the tweet. As a result, in order to evaluate tweet content, individual will construct their choice criteria on the fly, and evaluate utility of retweeting based on their constructed criteria. In sum, individuals rely on constructive choice process to evaluate contents, and not well-rehearsed decision criteria. In the next section, we discuss the constructive choice theory and its application to the case of retweeting.

2.2.1. Constructive Choice Process

Decision making under the constructive choice process involves evaluation on the fly. Instead of evaluating all options carefully, individuals rely on metagoals in order to evaluate and arrive at a choice. Bettman et al (1998) suggest four metagoals of constructing choice criteria 1) maximizing the accuracy of a decision 2) minimizing the cognitive effort required for the decision 3) minimizing the experience of negative emotion while making the decision, and 4) maximizing the ease with which a decision can be justified to others or to one's self. In addition, in different situations, different metagoals will be active and desired. However, in the situation where there is little involvement and there is no need for justification, the two metagoals of maximizing the accuracy of the decision and minimizing the cognitive effort required may become the dominant goals (Beach and Mitchell 1978, Bettman et al 1998, Hogarth 1987, Payne, Bettman, and Johnson 1993, Shugan 1980). Twitter is a social media where individuals receive multiple tweets per day, have a low involvement with a particular tweet, and are active due to entertaining purposes. Therefore, we argue that choice of tweet retransmission is such a situation, where choice criteria are constructed based on satisfying two metagoals of benefit maximization and effort minimization. The benefits individuals might gain by retransmission of celebrities' tweets, individuals rooted in self-identity and presenting themselves to other, which translates to self-enhancement benefits. In sum, in the case of celebrity-originated tweets, individuals will rely on message cues to evaluate their utility in terms of self-enhancement, but they also wish to minimize their effort as well.

2.2.2. Source Effect

There has been extensive literature in marketing on effect of message source on persuasiveness (Hovland and Weiss 1951, Fuchs 1964, Harmon and Coney 1982, Harkins and Petty 1987, Wilson and Sherrell 1993, Kang and Herr 2006). The theoretical foundation of these studies rooted in Elaboration Likelihood model (ELM) and the heuristic-systematic information processing theory of information processing (Chaiken 1980, Petty, Cacioppo, and Goldman 1981). According to these theories, there are two routes for information processing, systematic (central), and heuristic (peripheral). In the systematic information processing route, which needs significant amount of cognitive resources in order to evaluate message's arguments, individuals use their cognitive resources to process the message. In the heuristic information processing route, which individuals employ less effort to process the arguments, "recipients may rely on more accessible information such as source identity or non-content cues in deciding to accept a message's conclusions" (Chaiken 1980). In such a situation, where personal involvement is low, peripheral cues become more important and their effect will be more dominant on persuasion (Petty, Cacioppo, and Goldman 1981).

Individuals receive several tweets every day. Due to exposure to this huge amount of tweets, in general, involvement with a particular tweet is low. In the process of constructing choice criteria to retweet a message, since there is little involvement with a particular tweet, we argue that individuals rely on heuristic information processing in order to process information. In such a situation, peripheral cues are an important element of the message. Source of a message has been known as an important heuristic cue in the literature. For example, Chaiken (1980) showed that when there is little involvement, opinion changes are significantly greater given a likeable communicator. Harmon and Coney (1982) considered the effect of source credibility on persuasion. They found that source credibility is important when individuals do not favor an argument or have low involvement with the argument. Wilson and Sherrell (1993) considered a meta-analysis of 745 studies that considered the source effect on persuasion. They found that source expertise has highest effect on persuasion. Kang and Herr (2006) provided a unified framework to explain the counter-arguments in the source effect results. They hypothesized that "when the level of cognitive resources available for information processing is relatively low, positive source characteristics, either affectively or heuristically, will positively influence product attitudes, irrespective of product category". Such findings suggest that celebrity's popularity affects how contents originated by them are perceived.

In the context of celebrities communicating a message, Rossiter and Smidts (2012) examined the role of celebrities in print advertising, and considered the pairing of presenter and product, but they found that, generally, inclusion of celebrity presenter does not significantly increase the persuasiveness of the message. However, they found that expertise is the most important attribute that could affect persuasiveness. In the context of social media, in a series of experiments, Jin and Phua (2014) showed that message promoted by celebrities with more followers affects product involvement, buying intention, and the intention to spread eWOM. The also showed that celebrities with high vs. low number of followers are perceived as a more credible source in terms of attraction, trustworthiness, and competence. These findings suggest that celebrities' popularity influence receiver's response to the message. In other words, their popularity could act as a source effect on subsequent actions, where sharing in one among many others.

In the context of celebrities posting tweets on social media, based on the findings in the literature, we argue that content characteristics that signal positive (negative) benefit for the receiver on retransmitting of the content will have a positive (negative) effect on the volume and duration of retweet. In addition, any tweet characteristic that decreases (increases) the effort of the receiver to process the content will have a positive (negative) effect on the extent the tweet is shared. Such expectations root in the constructive choice process theory, and realization of the two meta-goals of maximizing utility and minimizing effort are desired. In addition, by relying on source effect theory, we argue that celebrities' number of followers, as a measure of popularity, affects individuals' decision to retweet directly, as well as indirectly and through affecting how individuals evaluate a message. For example, in the case of a REPLY tweet, which compared to a regular tweet has less benefit for the receiver to retweet; the sender's popularity will have a positive effect on how REPLY tweets are perceived in terms of its benefit, hence, the REPLY tweet will have a less negative effect for a more popular sender. We argue that such an effect exists for message cues that signal effort minimization and effort avoidance. For example, in the presence of QUESTION mark, which increases the effort to retweet, a more popular celebrity who sends a tweet with a question mark will have a less negative effect on retweeting, compared to a less popular sender in terms on number of followers.

2.3. Data

The data of this research consists of tweets from a popular social network, Twitter. We considered tweets posted by multiple celebrities during month of March 2013. We chose celebrities within seven categories of the entertainment industry: Actor, Musician, TV host, Journalist, Blogger, Athlete, and Model. Within each category, we chose top ten celebrities in terms of number of followers. There have been few overlaps across categories such that, for example, a celebrity was both an actor and a musician. In order to make categories homogenous, we consider only those celebrities that are active in one category. In addition, there have been few celebrities posting tweets in languages other than English, which we did not consider them in our final sample. Table 2.1 shows categories, number of pages, average number of followers, and number of observed tweets for each category.

Table 2.1 – Number of tweets in each category

Category	**# of pages**	**Mean # of Followers**	**# of observations**
Actor	9	7.47E+06	350
Music	10	2.31E+07	416
TV host	10	8.46E+06	434
Journalist	9	1.63E+06	248
Blogger	7	1.93E+06	663
Athlete	10	6.35E+06	571
Model	10	9.26E+05	253

As it can be seen from Table 2.1, the seven categories have significant difference in terms of number of followers, especially, the “Music” category has significant larger number of followers, and as much as three times of second largest category “TV host” in terms of number of followers. Table 2.2 shows celebrity names within each category, their respective number of followers, and number of tweets for each of them in our sample. In the next section, we discuss description of variables.

Table 2.2 – Pages in each category and number of tweets within each page

Celebrity Name	#Follower	#tweet
Actor		
Ashton Kutcher	1.40E+07	7
Paris Hilton	1.00E+07	84
Charlie Sheen	9.10E+06	12
Russell Brand	6.20E+06	64
Tom Hanks	5.90E+06	2
Emma Watson	5.70E+06	17
Stephen Fry	5.60E+06	121
Neil Patrick Harris	5.40E+06	15
Eva Longoria	5.30E+06	28
Musician		
Justin Bieber	3.60E+07	165
Katy Perry	3.30E+07	13
Rihanna	2.90E+07	70
Taylor Swift	2.50E+07	15
Britney Spears	2.50E+07	4
Shakira	2.00E+07	31
Justin Timberlake	1.70E+07	30
Nicki Minaj	1.60E+07	37
Bruno Mars	1.60E+07	49
Eminem	1.40E+07	2
TV-host		
Ellen DeGneres	1.70E+07	92
Oprah Winfrey	1.70E+07	58
Ryan Seacrest	9.20E+06	36
Jimmy Fallon	8.10E+06	35
Conan O' Brien	7.90E+06	20
Daniel Tosh	7.80E+06	114
Chelsea Handler	5.40E+06	10
Stephen Colbert	4.60E+06	14
Fearne Cotton	4.30E+06	47
Holly Willoughby	3.30E+06	8

Celebrity Name	# Follower	#tweet
Journalist		
Maria Shriver	2.10E+06	30
Bill Simmons	2.00E+06	67
George Stephanopoulos	1.80E+06	3
Dr. Sanjay Gupta	1.60E+06	12
David Gregory	1.60E+06	15
David Pogue	1.50E+06	24
Nicholas Kristof	1.40E+06	75
John Dickerson	1.40E+06	11
Ann Curry	1.30E+06	11
Blogger		
Perez Hilton	6.30E+06	476
Heather Armstrong	1.60E+06	11
iJustine	1.50E+06	71
Stefanie Michaels	1.40E+06	5
Agent M (Ryan Penagos)	1.30E+06	62
Jason Sweeney	1.10E+06	14
Robert Scoble	3.20E+05	24
Sport		
Cristiano Ronaldo	1.70E+07	32
LeBron James	7.60E+06	47
Shaquille O'Neal	6.80E+06	19
Wayne Rooney	5.90E+06	13
Chad Ochocinco	5.00E+06	226
Cesc Fàbregas	4.90E+06	13
Dwayne Johnson	4.20E+06	115
Floyd Mayweather, Jr.	4.10E+06	17
Rio Ferdinand	4.00E+06	58
Lance Armstrong	4.00E+06	31
Model		
Katie Price	1.80E+06	52
Dita Von Teese	1.40E+06	20
Gisele Bündchen	1.20E+06	6
Holly Madison	1.20E+06	52
Cindy Crawford	1.10E+06	14
Barbie Blank	7.00E+05	36
Alessandra Ambrosio	6.30E+05	35
Brooklyn Decker	4.70E+05	6
Padma Lakshmi	3.80E+05	13
Elizabeth Hurley	3.80E+05	19

2.3.1. Dependent Variables

The variables of interest in this research are duration and volume of retweets. Duration of retweets is a measure of how long a tweet has been shared, whereas volume of retweet is a measure of how many times it has been shared. The longer a tweet lives on social media (higher duration of retweet), the higher chance it has to be seen (and shared). Therefore, a longer duration of retweets increases volume of retweets, and more generally, it increases its effectiveness and reach. On the other hand, if a tweet has been shared more frequently (higher volume of retweet), it has been diffused deeper and wider to social networks. Therefore, it will have a higher chance to be seen and retweeted, which subsequently increases duration of retweets. As a result, duration and volume of retweet are two inter-dependent variables. Hence, our data collection algorithm and modeling approach should consider their dependency.

Duration of Retweet

Duration of retweets is simply defined as a period, over which a tweet has been retweeted. More specifically, it is the time elapsed after posting a tweet to its last retweet. Since contents on twitter exist forever, theoretically, a tweet has a chance of being retweeted at any time, hence, duration of retweet changes. However, twitter is an immediate social network, and tweets receive attention from individuals after their posting, until the point that the tweet will not be retweeted as frequently as before. Therefore, we define duration of retweet as the period that most of the retweets occur. More formally, we observe duration through a data collection algorithm that includes two identical periods,

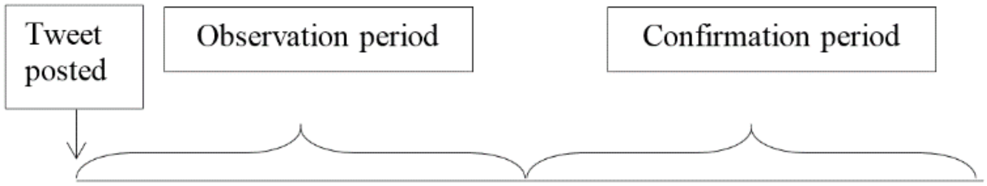

Figure 2.1 – Observation and Confirmation Period Relative to Tweet Posting Time

observation period, and *confirmation period*. Figure 2.1 shows the two periods with respect to posting time of a tweet. The tweets are tracked within the observation period in order to measure (observe) duration of retweets, and confirmation period is used to confirm that they have not been retweeted. More specifically, we have two scenarios across all tweets:

1- Last retweet of a tweet happened during the observation period, and there is no retweet in the confirmation period. In this case, we observe the duration of retweet by taking the difference between the time of last retweet and the time of tweet posted.

2- The tweet has been retweeted during both observation period and confirmation period. In this case, we considered the tweet to be censored with respect to the end of observation period, and the duration is equal to the length of observation period.

It can be understood that the length of observation and confirmation periods will have an impact of what proportion of the observations are censored and what proportion are not censored. In other words, if we take a short interval for the two periods, most of the observed durations will be censored, which is not favorable in order to make inference. In order to make objective selection of the length of periods, we chose the length such that 80% of the observations are not censored and 20% of the observations are censored. Given this criterion, the length of the two periods is 28.7 days. In all, we have 2935 observations in our dataset, where 563 observations are censored. Figure 2.2 shows distribution of duration, and the black column shows the proportion of the censored observations.

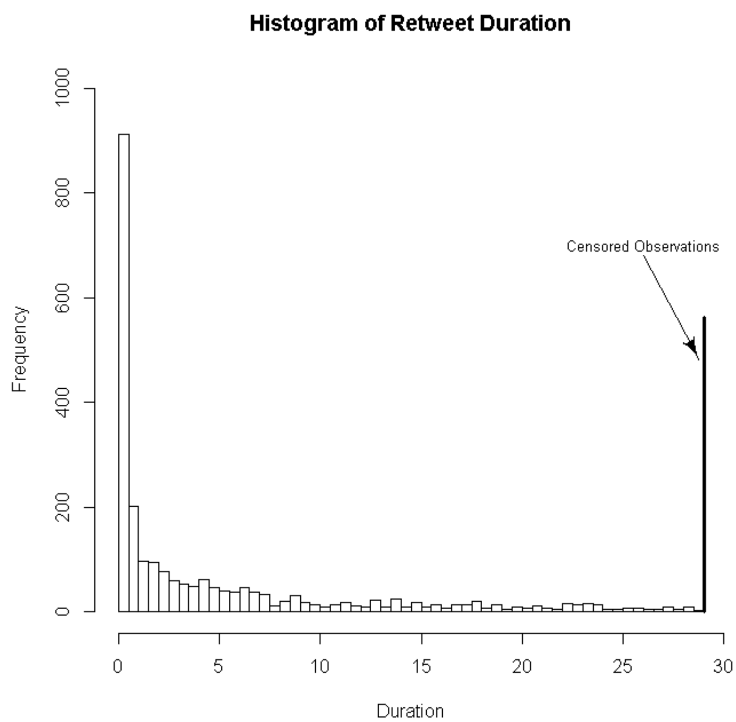

Figure 2.2 – Histogram of Retweet Duration (in days)

Volume of Retweet

The retweet volume is a measure of how many times a tweet has been shared. Since we observed the duration in a censoring algorithm that has been described above, the volume of retweet variable is also affected by how we observed duration. In specific, if the tweet has been retweeted only during the observation, and not in the confirmation period, we will take the retweet count at its last retweet as the volume of retweet. In the case that the tweet has been retweeted both in the observation and confirmation period, we observed the retweet count at the end of observation period, but we consider such observations as censored, as in the case of the duration. In other words, we do not observe the final retweet volume for the censored observations, and we only observe volume, up to the censored time, and we assume that the final retweet count will be more than what we observed. Figure 2.3 shows the histogram of log(retweet volume) for the censored and non-censored observations.

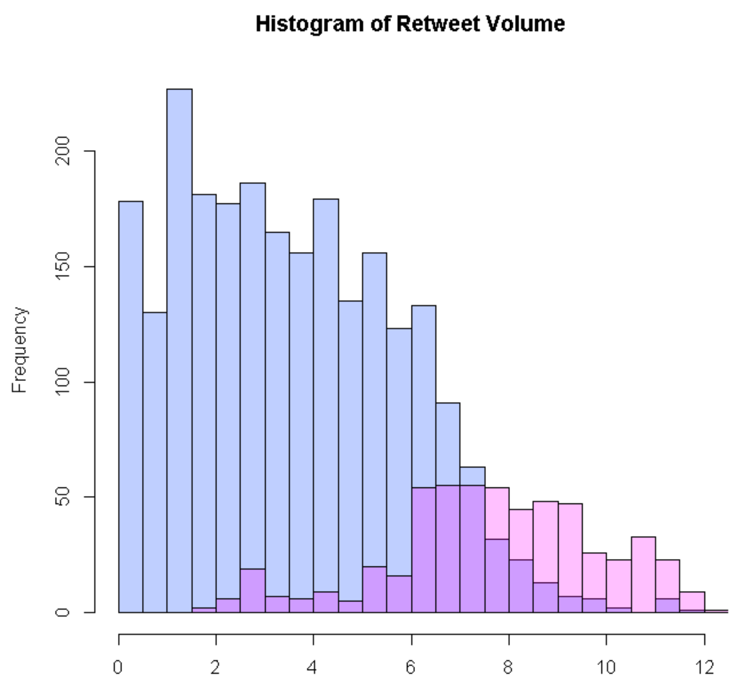

Figure 2.3 – Histogram of LOG(Retweet Volume), the blue bars (on left) are non-censored tweets, the pink bars (on right) are censored, and the overlapping area is in violet (in the middle).

2.3.2. Tweet Characteristics

In this research, we incorporate two indicator variables for tweet type. In addition to tweeting their own status, sometimes, celebrities reply to the follower's tweet. Also, rather than composing a tweet themselves, celebrities share (retweet) a tweet originally posted by other users. In both situations, receivers of the tweet could spot that the tweet neither is meant for general audience (reply tweet) nor is directly from the celebrity (retweet). In specific, we included these two variables as below.

- **REPLY:** an indicator variable, which takes the value of one, if the tweet is a reply to another user
- **RETWEET:** an indicator variable, which takes the value of one, if the tweet is a retweet of another person's tweet

In addition to tweet types, we also incorporated five indicator variables based on tweet content. In specific, we incorporated the following variables.

- **HASH:** an indicator variable for the presence of hashtag (#) in the tweet
- **QUES:** an indicator variable for the presence of question mark (?) in the tweet
- **EXCL:** an indicator variable for the presence of exclamation mark (!) in the tweet
- **LINK:** an indicator variable for the presence of any link in the tweet
- **LEN:** this variable measures length of tweet in terms on number of characters

Table 2.3 provides summary statistics for the covariates.

Table 2.3 – Summary statistics of covariates

Variable					**Yes**	**No**	
REPLY					998	1937	
RETWEET					252	2683	
HASH					647	2288	
QUES					341	2594	
EXCL					1128	1807	
LINK					1275	1660	
	Mean	**Std**	**Min**	**25%**	**Median**	**75%**	**Max**
LEN	92.72	35.82	2	65	99	124	149

2.3.3. Celebrity Characteristics

We included celebrities' number of followers as a proxy for their popularity in our study. It can be understood that as a celebrity becomes more popular, the chance that he/she will get followers on social media will be higher; hence, number of followers is a good proxy for celebrity popularity. This variable in observed at a celebrity level, hence for all tweets posted by a celebrity, it will be identical. The number of followers for all the pages has been shown in Table 2.2.

2.4. Modeling Approach

In this section, first, we present modeling approach for duration and volume, then, we discuss issues of censoring, covariates, unobserved heterogeneity, and joint model.

2.4.1. Model for Duration

We used a proportional hazard model for the duration of retweets. We assumed a Weibull baseline hazard function for the proportional hazard model. In specific, let t_i be the duration of retweets for tweet i, then the baseline hazard and the respective probability density function are,

$$h\left(t_i|\alpha, \lambda_i^{DU}\right) = \alpha\ \lambda_i^{DU}\ t_i^{\alpha-1}$$

$$t_i|\alpha, \lambda_i^{DU} \sim Weibull\left(\alpha, \lambda_i^{DU}\right) = \lambda_i^{DU}\ \alpha\ t_i^{\alpha-1}\ exp\left(-\lambda_i^{DU} t_i^{\alpha}\right) \qquad (2.1)$$

Where λ_i^{DU} is the tweet specific scale of Weibull distribution (the superscript DU stands for duration) and α is the shape parameter of the distribution. In this representation of the Weibull proportional hazard model, the survival probability is $S\left(t_i\middle|\alpha, \lambda_i^{DU}\right) = \exp\left(-\ \lambda_i^{DU} t_i^{\alpha}\right)$. In order to verify suitability of the Weibull model, one would look to the following relationship: $\log\left(-\log\left(S\left(t_i\middle|\alpha, \lambda_i^{Du}\right)\right)\right) = \log\left(\ \lambda_i^{DU}\right) + \alpha\log(t_i)$, derived from the survival probability. As a result, a plot of $\log\left(-\log\left(S\left(t_i\middle|\alpha, \lambda_i^{DU}\right)\right)\right)$, estimated non-parametrically, against $\log(t_i)$ should be approximately linear. We used the Kaplan-Meier non-parametric estimate of the survival probabilities and Figure 2.4 shows the graph of the two above terms. Based on the figure Figure 2.4, the relationship is close to linear, which shows that Weibull specification of the hazard model is suitable.

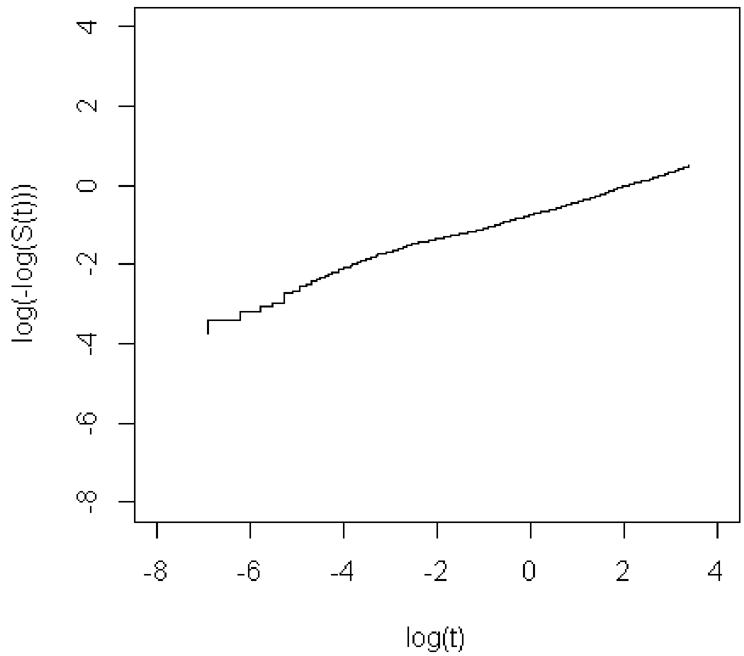

Figure 2.4 – Plot of $log(t)$ *vs.* $log\left(-log\left(S(t)\right)\right)$ *showing suitability of Weibull Model*

2.4.2. Model for Volume

We employed an empirical count model to represent the observed volume of the retweets. Based on the overall shape of the retweet count distribution on Figure 2.3, it can be seen that the observed retweet count has larger variance than mean; hence, it is over-dispersed, which makes the Poisson model unsuitable. Therefore, we adopted a mixture of Poisson distribution to overcome the over-dispersion. In specific, let N_i be the observed retweet count of the tweet i.

$$N_i \sim Poisson\left(\lambda_i^{VO}\right) \propto \frac{\lambda_i^{VO^{N_i}}}{N_i!} \tag{2.2}$$

The λ_i^{VO} is the mean of distribution (the superscript VO stands for the model of retweet volume). We will specify a normal distributed additive random term in sequel (unobserved heterogeneity section), which captures over-dispersion. In the next sections, we specify the censoring, covariates, unobserved heterogeneity, and joint model.

2.4.3. Censoring

The data collection mechanism is such that duration and volume of retweets are censored for some tweets. For such tweets, we specify the model through the survival function of their respective distribution, since we only observe that their "true" observed value for duration and volume is larger than what we observed. In specific, we specify the survival function of the duration model and count model as follows:

$$S(N_i) = 1 - \frac{\Gamma(\lfloor N_i + 1 \rfloor, \lambda_i^{VO})}{\lfloor N_i \rfloor !}$$

$$S(t_i) = \ exp\left(-\lambda_i^{DU} t_i^{\alpha}\right)$$

Where $\Gamma(\ )$ is the incomplete gamma function and $\lfloor\ \rfloor$ is the floor function.

2.4.4. Covariates, Unobserved Heterogeneity, and Joint Model

In order to include tweet specific covariates into our model, as well as brand effects, we write the scale parameter of the hazard model (in logarithm scale), and the mean of the count model (in logarithm scale) with respect to the covariates. In specific, let tweet i posted by celebrity $j, j = 1 \dots J$ with the scale parameter of λ_{ij}.

$$log\left(\lambda_{ij}^{L}\right) = \mu_{ij}^{L}, \quad L = DU \ or \ VO$$

We specify μ_{ij}^{L} with respect to a celebrity specific covariate, discussed in section 2.3.3, and tweet specific covariates discussed in section 2.3.2 as below,

$$\mu_{ij}^{L} = Intercept + NUM.FOL_j\,\gamma^{L} + REPLY_i\,\beta_1^{L} + RETWEET_i\,\beta_2^{L} + HASH_i\,\beta_3^{L} + QUES_i\,\beta_4^{L} + EXCL_i\,\beta_5^{L} + LINK_i\,\beta_6^{L} + LEN_i\,\beta_7^{L}, \qquad L = DU \text{ or } VO$$

As we discussed in the theoretical background section, we want to examine the effect of celebrity popularity on how the message characteristics are perceived. Therefore, we include the interaction term of celebrity's number of followers with the tweet specific covariates. The main effects of tweet characteristics are captured through β_1^{L} ... β_7^{L}, and the interaction effect of celebrity popularity with the tweet content are captured through α_1^{L} ... α_7^{L}. In specific, we have the following specification for μ_{ij}^{L}.

$$\begin{aligned}\mu_{ij}^{L} = {} & Intercept^{L} + NUM.FOL_j\,\gamma^{L} + REPLY_i\left(\beta_1^{L} + \alpha_1^{L}\,NUM.FOL_j\right)\\ & + RETWEET_i\left(\beta_2^{L} + \alpha_2^{L}\,NUM.FOL_j\right) + HASH_i\left(\beta_3^{L} + \alpha_3^{L}\,NUM.FOL_j\right)\\ & + QUES_i\left(\beta_4^{L} + \alpha_4^{L}\,NUM.FOL_j\right) + EXCL_i\left(\beta_5^{L} + \alpha_5^{L}\,NUM.FOL_j\right)\\ & + LINK_i\left(\beta_6^{L} + \alpha_6^{L}\,NUM.FOL_j\right)\\ & + LEN_i\left(\beta_7^{L} + \alpha_7^{L}\,NUM.FOL_j\right), \qquad L = DU \text{ or } VO\end{aligned}$$

We observed few tweet attributes, but since tweet attributes cannot be captured fully through the variables that we included in our model, we included an additive frailty, δ_i^{DU} into the duration model, and a random term, δ_i^{VO} in the volume model. We also specify their joint distribution, in order to capture the relationship between volume and duration of the retweets, through a bivariate normal distribution between the two random terms. In order to guarantee identification of the intercepts in both models, the frailty and random term should have mean of zero. In specific,

$$\log\left(\lambda_{ij}^{L}\right) = \mu_{ij}^{L} + \delta_{i}^{L},\ L = DU\ or\ VO$$

$$\delta_i^{DU}, \delta_i^{VO} \sim BVN(\begin{bmatrix}0\\0\end{bmatrix}, \Sigma_{2\times 2})$$

In addition, there will be unobserved characteristics of the celebrities such as attractiveness, which may not be captured by number of followers. In order to capture such effects, we included celebrity specific random effect, η_j^{DU} and η_j^{VO}, for both models. Therefore, the full specification of mean and scale parameters are,

$$\log\left(\lambda_{ij}^{L}\right) = \mu_{ij}^{L} + \delta_{i}^{L} + \eta_j^L, \quad L = DU\ or\ VO$$

$$\eta_j^L \sim Normal\left(0, \tau_\eta^L\right), \quad L = DU\ or\ VO$$

2.4.5. The Likelihood Function

For non-censored tweets, the likelihood is $\left(\lambda_i^{DU}\ \alpha\ t_i^{\alpha-1}\ exp\left(-\lambda_i^{DU} t_i^{\alpha}\right)\right).\left(\frac{{\lambda_i^{VO}}^{N_i}}{N_i!}\right)$, where t_i is the observed duration of retweets, N_i is the observed retweet volume, λ_i^{DU} and λ_i^{VO} are specified above. However, for the censored observations, the likelihood is the product of the two survival functions, $\left(\exp\left(-\lambda_i^{DU} t_i^{\alpha}\right)\right).\left(1 - \frac{\Gamma(\lfloor N_i+1 \rfloor, \lambda_i^{VO})}{\lfloor N_i \rfloor!}\right)$. The full likelihood function of tweet i by considering censoring variable, C_i (=1 if censored) is,

$$\left(\left(\lambda_i^{DU}\ \alpha\ t_i^{\alpha-1}\ exp\left(-\lambda_i^{DU} t_i^{\alpha}\right)\right).\left(\frac{{\lambda_i^{VO}}^{N_i}}{N_i!}\right)\right)^{1-C_i}\left(\left(\exp\left(-\lambda_i^{DU} t_i^{\alpha}\right)\right).\left(1 - \frac{\Gamma\left(\lfloor N_i + 1 \rfloor, \lambda_i^{VO}\right)}{\lfloor N_i \rfloor!}\right)\right)^{C_i}$$

2.4.6. Model Estimation

We estimated the proposed model in a Bayesian framework through BUGS software. We specified prior distribution of model parameters, and employed MCMC methods to simulate from posterior distribution of the model parameters, and make inference based on the 95% posterior intervals. The prior specifications for the coefficients of the covariates (all β's and α's) and intercepts of the model, $intercept_k^L$, are normal with zero mean and large variance. The prior for the precision parameters of the celebrity random effects, η^{DU} and η^{CO}, and shape parameter of the Weibull model are $Gamma(0.001{,}0.001)$. Finally, the prior distribution for precision matrix of the joint random effects, $\Sigma_{2\times 2}$, is Wishart with scale of $\begin{bmatrix} 0.001 & 0 \\ 0 & 0.001 \end{bmatrix}$ and degree of freedoms of two. We estimated the model with burn-in of 100K, thinning of 100 with the final sample from 1000 iterations. Convergence has been assured graphically across two chains.

2.4.7. Model Fit

In order to ensure models performance in terms of its fit and predictive performance, for the duration model, we calculated the predictive survival probability for the tweets in our sample, which is the probability that predicted duration of retweets is equal of greater than observed duration. In specific, we calculated $\frac{\sum_i exp\left(-\hat{\lambda}_i^{DU} t_i^{\hat{\alpha}}\right)}{N}$, where $\hat{\lambda}_i^{DU}$ and $\hat{\alpha}$ are the estimated values of these parameters in each MCMC iteration, and t_i is the observed duration, and N is sample size. We also calculated a discrepancy measure between the predictive distribution and data for the count model, which has been discussed in section 1.4.1.

2.5. Results

In the following two sections, we discuss results of duration and volume model.

2.5.1. Duration of Retweet

Table 2.4 presents estimated coefficients of duration model. It is noteworthy to mention that since we modeled hazard rate, therefore, estimated coefficients demonstrate effects of included variables on the hazard rate, and this has to be taken into consideration in order to interpret signs of the coefficients. In the case of negative coefficients, the variable will reduce the hazard rate, or decrease probability of failure, hence it increases the duration, and vice versa for positive coefficients. We present the results of tweet type, tweet content, and other parameters of the model.

Tweet Type

Tweet type of REPLY has a significant positive effect on hazard, which means the duration of retweet for these tweets is less than regular tweets. Since REPLY tweets are posted directly to an individual, and are not posted for general audience, therefore, they will be less relevant to all followers. As a result, retweeting the REPLY tweets will provide less benefit to the individuals in terms of self-enhancements purposes. However, the interaction of REPLY with NUM.FOL does not have a significant effect on hazard. The second variable of tweet type, RETWEET, has a significant positive effect on hazard, which means these tweet will have less duration of retweets. The RETWEET tweets are originally posted by another user, and a celebrity retweeted that tweet to his/her followers. Clearly, it is not an original tweet by celebrities, and although it might be relevant to general audience, since the original poster of the tweet is not the celebrity, hence the self-enhancement purposes that individuals might want to signal by retweeting, will not be satisfied. However, its interaction with NUM.FOL is not significant.

Table 2.4 – Estimated parameters of duration model

		Mean	**2.5%**	**Median**	**97.5%**
Tweet Type					
β_1^{DU}	**REPLY ****	2.05	1.61	2.04	2.64
α_1^{DU}	REPLY × NUM.FOL	-0.25	-0.62	-0.25	0.11
β_2^{DU}	**RETWEET ****	1.65	1.16	1.63	2.25
α_2^{DU}	RETWEET × NUM.FOL	-0.44	-1.11	-0.44	0.21
Tweet Content					
β_3^{DU}	**HASH ****	-0.35	-0.64	-0.34	-0.05
α_3^{DU}	**HASH × NUM.FOL ****	-0.60	-0.91	-0.60	-0.31
β_4^{DU}	EXCL	0.20	-0.03	0.20	0.47
α_4^{DU}	EXCL × NUM.FOL	0.10	-0.18	0.10	0.37
β_5^{DU}	**QUES ****	0.33	0.01	0.32	0.65
α_5^{DU}	QUES × NUM.FOL	-0.28	0.13	0.39	-0.72
β_6^{DU}	**LINK ****	0.62	0.33	0.62	0.97
α_6^{DU}	**LINK × NUM.FOL ****	0.38	0.08	0.38	0.69
β_7^{DU}	LEN	-0.01	-0.12	-0.01	0.09
α_7^{DU}	**LEN × NUM.FOL ****	0.27	0.14	0.27	0.41
Celebrity Attribute					
β_8^{DU}	**NUM.FOL ****	-2.02	-2.76	-2.01	-1.37
Other Parameters of Model					
α		1.37	1.18	1.63	1.37
τ_η^{DU}		0.16	0.09	0.26	0.16
$\Sigma[2,2]$		0.39	0.22	0.62	0.39
ρ		-0.73	-0.76	-0.69	-0.73
$Intercept^{DU}$		-3.91	-4.96	-3.01	-3.91

** 95% probability interval does not include zero

Tweet Content

Presence of hashtag in a tweet, HASH, has a significant negative effect on hazard. Presence of a hashtag in a tweet makes that phrase to stand out against other words. As a result, the required effort to process a hashtag is less, which due to satisfying effort minimization metagoals, ultimately makes the message to be retweeted more. The interaction effect of NUM.FOL and HASH has a positive effect on duration, which means a hashtag used by celebrities that are more popular will be more effective compared to less popular ones. This is consistent with the two metagoals of benefit maximization, and effort minimization, since, in the case of a more popular celebrity tweeting using a hashtag, both metagoals are present, more self-enhancement, and less effort to process, hence increase retransmission of tweets. The effects of EXCL and its interaction with NUM.FOL on duration of retweets are not significant. However, QUES has a negative effect on duration of retweet. Presence of question mark signals that a tweet contains a question, which requires additional effort by receivers to be ready for retweeting. Therefore, individuals might avoid such effort and avoid retweeting that. The interaction of QUES with number of followers is not significant. The presence of LINK on a tweet has a positive effect on hazard or a negative effect on duration. As discussed, individuals rely on message source to make their sharing decision, and with the case of no link on the tweet, they could easily make their decision either to retweet or not, since there is no additional content to evaluate or check. However, in the case of presence of LINK in a tweet, individuals may want to check out the link before actually share that. Therefore, there is an additional action to be taken by individuals before making their decision. In the case of checking the LINK, they may be exposed to the LINK content, and possibly change their mind in terms of retweeting. In other words, presence of LINK in a tweet makes the decision of retweeting to be less relied only on the source and more on the additional content through the link. Its interaction with NUM.FOL is significant, which means for a more popular celebrity, this effect is even more negative. The effect of tweet length, LEN, on duration is not significant, but its interaction with NUM.FOL has a positive effect on hazard. That is, for a celebrity with more number of followers, a shorter message will have more retweet duration.

Celebrity Attribute

The effect of number of followers on hazard is negative, which means more number of followers lead to higher duration. This meets our expectations, since a celebrity with more number of followers is possibly more popular among his/her followers; hence as a source of message, it affects how individuals respond to the tweets posted by him/her. In addition to the effect of NUM.FOL, we controlled for unobserved characteristics of the celebrities. Figure 2.5 shows the estimated random effects of these random effects.

Other Parameters of Duration Model

Table 2.4 shows estimates of other model parameters. The shape parameter of Weibull model, α, is 1.34. Since it is larger than one, it means the baseline hazard rate will increase as time goes by, which is an expected shape. The precision parameter of brand's random effect is 0.18, which is relatively small, and it signals, other than number of followers, there are other attributes, which affect the duration of retweets. The precision of frailty is 0.41, which means there are also some unobserved attributes for the tweets. The correlation between the frailties is -0.73, which is relatively a very strong relationship. The negative sign means higher volume of retweets will reduce the probability of failure (hazard) or increase the duration, and these two are associated to each other.

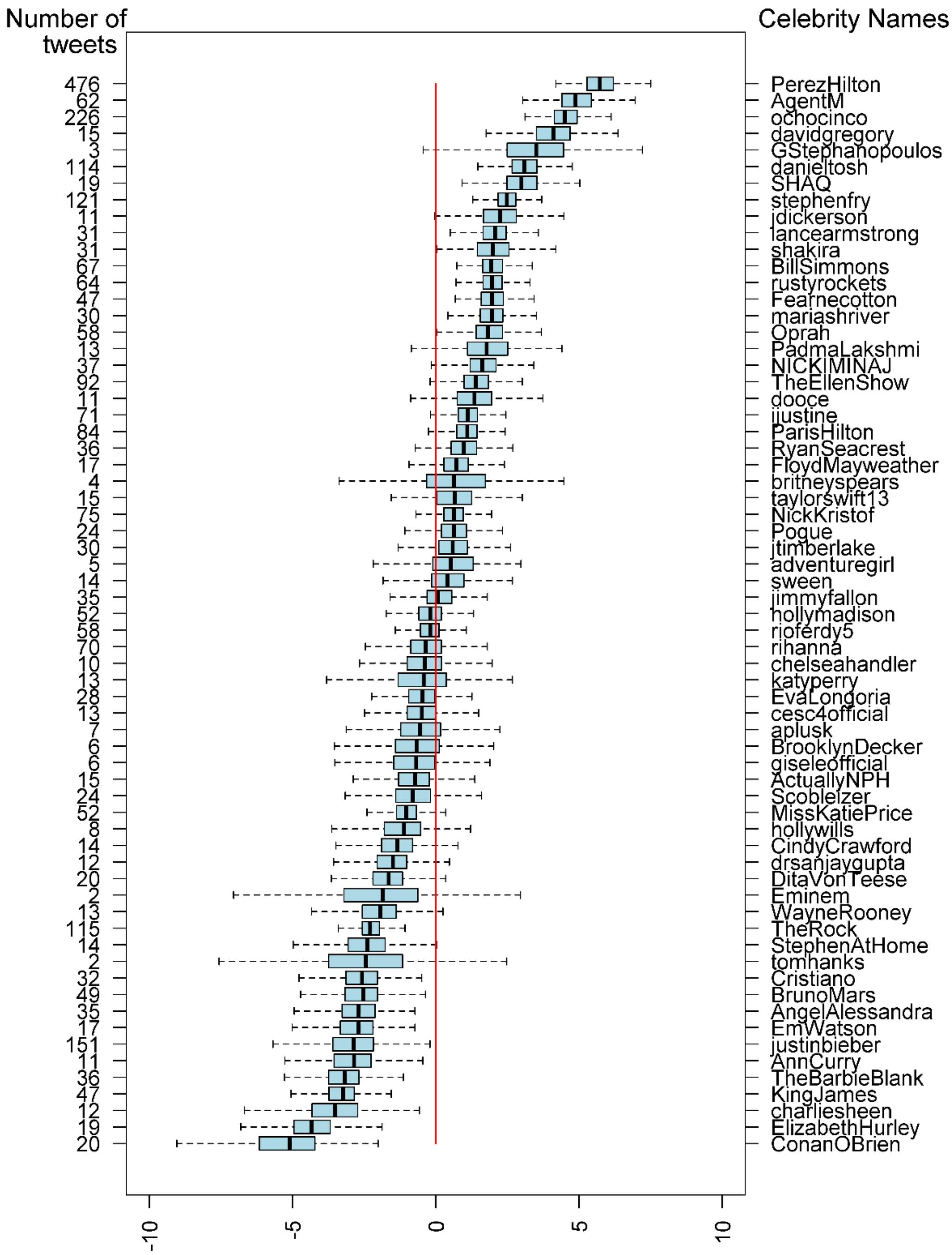

Figure 2.5 – Boxplot of the estimated celebrity random effects in the duration of retweets model

2.5.2. Volume of Retweet

Table 2.5 provides estimated parameters of the volume model. We present results of variables in each segment as below.

Tweet type

Based on the 95% probability intervals, the effect of REPLY tweet on retweet volume is significant and negative. This effect is consistent with the results of duration model. Since REPLY tweets are meant for a specific individual, therefore, they will be less relevant to all followers, and the utility of retweeting for other followers will be less compared to a regular tweet. However, the interaction effect of REPLY and NUM.FOL is not significant. The effect of RETWEET on volume of retweet is also negative, which is consistent with our expectations. Since it is not an "original" tweet by celebrities, and it is a retweet of a tweet posted by someone else, therefore, the benefit of sharing will be less for their followers. Interestingly, its interaction effect does not have a significant effect on volume of retweet.

Table 2.5 – Estimated parameters of volume model

		Mean	2.5%	Median	97.5%
Tweet Type					
β_1^{VO}	**REPLY ****	-2.25	-2.45	-2.25	-1.94
α_1^{VO}	REPLY × NUM.FOL	0.20	-0.15	0.20	0.64
β_2^{VO}	**RETWEET ****	-1.04	-1.24	-1.04	-0.84
α_2^{VO}	RETWEET × NUM.FOL	0.03	-0.29	0.03	0.34
Tweet Content					
β_3^{VO}	**HASH ****	0.16	0.05	0.16	0.33
α_3^{VO}	**HASH × NUM.FOL ****	0.28	0.17	0.25	0.55
β_4^{VO}	**EXCL ****	-0.18	-0.34	-0.17	-0.07
α_4^{VO}	EXCL × NUM.FOL	0.01	-0.15	0.02	0.12
β_5^{VO}	**QUES ****	-0.22	-0.38	-0.23	-0.03
α_5^{VO}	QUES × NUM.FOL	0.13	-0.19	0.15	0.31
β_6^{VO}	**LINK ****	-0.28	-0.40	-0.28	-0.16
α_6^{VO}	LINK × NUM.FOL	0.03	-0.11	0.05	0.14
β_7^{VO}	LEN	-0.01	-0.05	-0.01	0.03
α_7^{VO}	**LEN × NUM.FOL ****	-0.22	-0.34	-0.22	-0.11
Celebrity Attribute					
β_8^{VO}	**NUM.FOL ****	1.86	1.83	1.86	1.88
Other Parameters of Model					
τ_η^{VO}		0.27	0.18	0.27	0.37
$\Sigma[1,1]$		1.36	1.20	1.36	1.55
$Intercept^{VO}$		6.82	6.79	6.82	6.85

Tweet Content

Tweet content consists of five variables in our analysis. Presence of hashtag in a tweet increases volume of retweet. This is consistent with our expectations, since presence of hashtag, makes the word or phrase to stand out, and gets the attention of receivers. Therefore, it reduces the effort to process information; hence, it has a positive effect on retweet volume. The effect of its interaction with NUM.FOL is positive, which means tweets with hashtags posted by celebrities that are more popular will have higher volume of retweets. This effect is consistent with the interaction effect of HASH and NUM.FOL on duration model. Since the keywords and important words in a text are used as hashtag, therefore, the retransmission decision is very easy for the case of a more popular celebrity tweeting about an important topic highlighted through hashtag. The effect of EXCL is negative on volume of retweet, whereas its interaction does not have a significant effect. The effect of QUES on retweet volume is negative. Since presence of question mark signals a question, therefore, it requires additional effort, such as adding an answer, for the recipient in order to make it ready for sharing. Therefore, due to the additional effort required, its negative effect is expected. Its interaction effect with NUM.FOL is positive, but it is not significant. The presence of LINK in a tweet is negative and significant. This is the same effect as the duration model. Presence of a LINK makes the receiver to check it out before retransmitting, which requires additional effort by them, but by considering the LINK and its content, individuals might think the benefit they want to gain by sharing will be lower due to the content of the link. In addition, since we did not characterize the link type, so there might be a mixed effect here. The interaction effect of LINK and NUM.FOL is not significant. The effect of tweet length, LEN, is not significant; however, its interaction effect is significant and negative. That means a shorter message will have more volume of retweets in the case of more popular celebrities.

Celebrity Attribute

The effect of number of followers on volume of retweet is significant and positive. That means a larger network size for a celebrity leads to a higher volume of retweet. However, the number of followers on social media for celebrities is mainly due to their attractiveness, competence, and influence in the real world, therefore, we could conclude that those attributes correlates with volume of retweet for the tweets posted by them. Figure 2.6 provides boxplots of celebrities' random effects, which captures unobserved characteristics of them.

Other Parameters of Volume Model

The precision parameter of celebrity random effect is 0.3, which means there are some unobserved characteristics of the celebrities that were not captured by number of followers. The precision parameter of frailty is 1.36, which is not large, but is served its purpose. The intercept of the model 6.6, which is a large value compared to previous studies, which means response to tweets posted by celebrities are fairly more than other categories. That could be due to average number of followers of celebrities versus other brands.

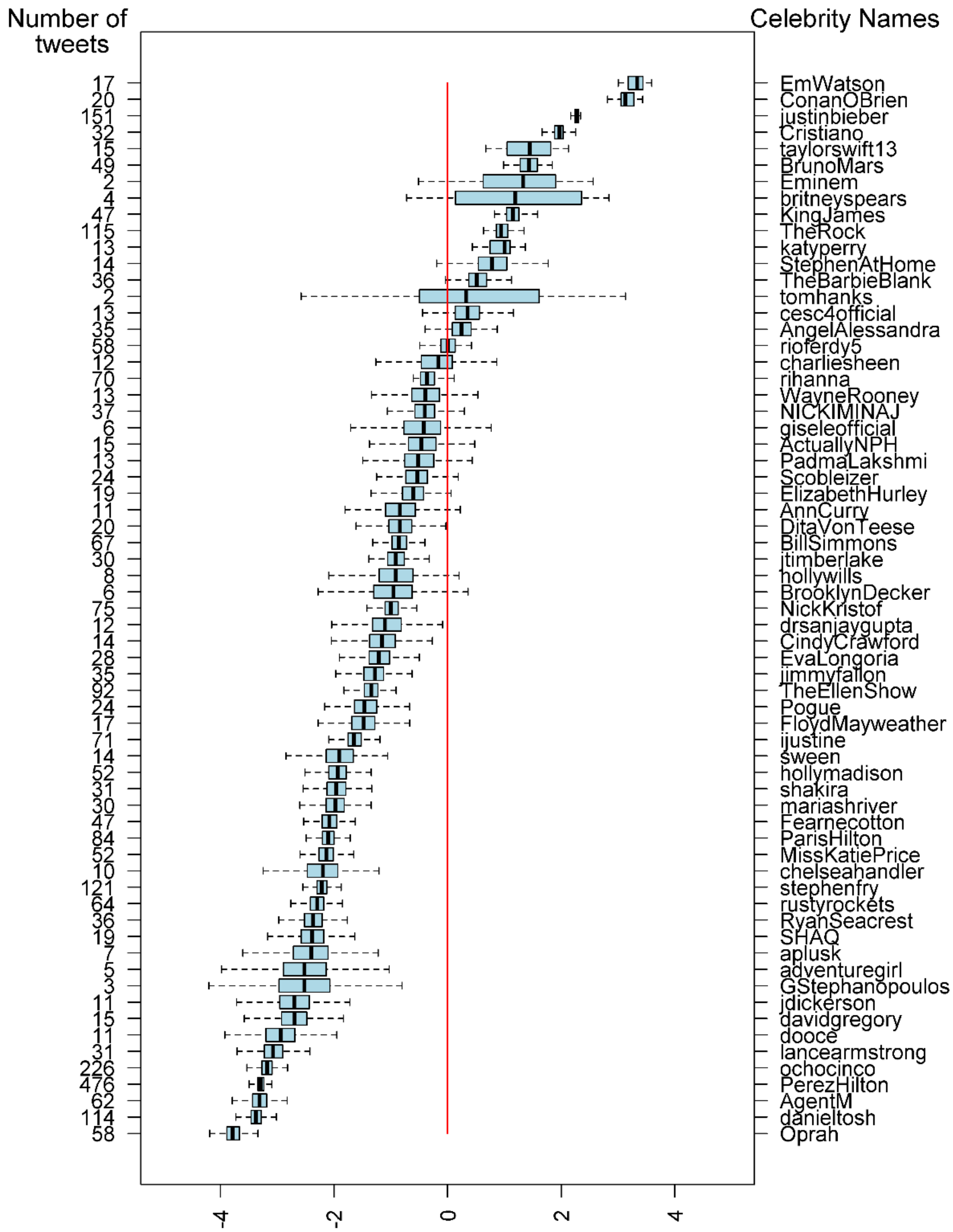

Figure 2.6 – Boxplot of the estimated celebrity random effects in the retweet volume model

2.6. Discussion

In this research, we examined retweet volume and duration of tweets posted by 65 celebrities across seven subcategories of entertainment industry. By considering source effect theory, we considered the effect of source popularity measured by number of followers on how tweet attributes affect volume and duration of retweet. We showed that source characteristics moderate effects of few tweet characteristics on volume and duration of retweets. However, most of the tweet characteristics effects on volume and duration were identical in terms of significance and direction. In addition, as expected, we found that volume and duration of retweet are highly associated.

2.6.1. Managerial Implications

In this research, we found that tweet source acts as a message cue and affects how receivers evaluate content, in order to make a decision to retransmit a tweet. Our findings have several implications for brands. It could help brands to choose among social media celebrities, through which, they could send out a tweet about brands. We also showed that few tweet attributes might not be suitable in the case of tweets by celebrities. Brands should consider those aspects in order to increase tweets' effectiveness in social media. In addition, our findings could have implications beyond the case of celebrity-sponsored tweets. By considering effect of tweet source on the response to the posted tweet, brands could investigate on content types or attributes that work better with their overall image and strategy, which subsequently can be incorporated into their social media content strategy, and makes their social media activities more effective.

Our findings have implications beyond Twitter, and apply to other social networks as well. Twitter is a social network that generally involvement is lower, compared to more information rich social networks such as Facebook, Instagram, or YouTube. It also weights immediacy and quickness, which is different from nature of other social networks. We argued that since involvement is low in twitter, message source acts as cue in the decision making process to retransmit a message. We believe that source effect will be less prominent on information rich and involving social media channels, since the content will have a more important role on sharing decision. Nevertheless, we believe that there will be differences between celebrities and ordinary individuals, and in the case of celebrities, message attributes will be less important, due to the influence of celebrities on their followers.

2.6.2. Limitation, and Future Studies

One limitation of our research lies on the celebrities we considered in this research. We investigated top celebrities in each category, and our results may not be generalizable to social media celebrities, also known as "micro-celebrities". These groups of celebrities acquired substantial amount of followers, only through their appealing contents on social media. They are also becoming a venue for marketing purposes, but they are mostly known in a particular area such as beauty, fashion, and home design. Our dataset included the well-known celebrities, and our results may not be applicable to micro-celebrities, because they acquired their followers only through their social media activities, and compared to most-well-known celebrities, they might even have a higher influence on their followers.

Our research showed the effect of source on how retransmission of tweets varies. This could be a starting point on investigating several aspects of source characteristics on social media responses. Utilizing celebrities and their effects as message source have been investigated in marketing literature extensively. Potential research questions such as match between celebrity and brand could have implications for promoting tweets. In addition, understanding celebrity attributes such as attractiveness, credibility and trustworthiness that could lead to likeability and persuasiveness of their tweets on social media is an avenue for further research.

References

Beach, Lee Roy, and Terence R. Mitchell. 1978. “A Contingency Model for the Selection of Decision Strategies.” Academy of Management Review 3 (3): 439–49.

Bettman, James R., Mary Frances Luce, and John W. Payne. 1998. “Constructive Consumer Choice Processes.” *Journal of Consumer Research* 25 (3): 187–217.

Chaiken, Shelly. 1980. “Heuristic versus Systematic Information Processing and the Use of Source versus Message Cues in Persuasion.” *Journal of Personality and Social Psychology* 39 (5): 752–66.

De Angelis, Matteo De, Andrea Bonezzi, Alessandro M Peluso, Derek D Rucker, and Michele Costabile. 2012. “On Braggarts and Gossips: A Self-Enhancement Account of Word-of-Mouth Generation and Transmission.” Journal of Marketing Research 49 (4): 551–63.

Fuchs, Douglas A. 1964. “Two Source Effects in Magazine Advertising.” *Journal of Marketing Research* 1 (3): 59–62.

Guadagni, Peter M., and John D. C. Little. 1983. “A Logit Model of Brand Choice Calibrated on Scanner Data.” Marketing Science 2 (3): 203–38.

Harkins, Stephen G., and Richard E. Petty. 1987. “Information Utility and the Multiple Source Effect.” *Journal of Personality and Social Psychology* 52 (2): 260–68.

Harmon, Robert R., and Kenneth A. Coney. 1982. “The Persuasive Effects of Source Credibility in Buy and Lease Situations.” *Journal of Marketing Research* 19 (2): 255–60. doi:10.2307/3151625.

Hennig-Thurau, Thorsten, Kevin P. Gwinner, Gianfranco Walsh, and Dwayne D. Gremler. 2004. “Electronic Word-of-Mouth via Consumer-Opinion Platforms: What Motivates Consumers to Articulate Themselves on the Internet?” Journal of Interactive Marketing 18 (1): 38–52.

Hogarth, Robin M. 1987. Judgement and Choice: The Psychology of Decision (2nd Ed.). Vol. xii. Oxford, England: John Wiley & Sons.

Hovland, Carl I., and Walter Weiss. 1951. “The Influence of Source Credibility on Communication Effectiveness.” *Public Opinion Quarterly* 15 (4): 635–50.

Jin, Seung-A. Annie, and Joe Phua. 2014. “Following Celebrities’ Tweets About Brands: The Impact of Twitter-Based Electronic Word-of-Mouth on Consumers’ Source Credibility Perception, Buying Intention, and Social Identification With Celebrities.” *Journal of Advertising* 43 (2): 181–95.

Kang, Yong-Soon, and Paul M. Herr. 2006. “Beauty and the Beholder: Toward an Integrative Model of Communication Source Effects.” *Journal of Consumer Research* 33 (1): 123–30.

Lovett, Mitchell J., Renana Peres, and Ron Shachar. 2013. “On Brands and Word of Mouth.” Journal of Marketing Research 50 (4): 427–44.

Malhotra, Arvind, Claudia Kubowicz Malhotra, and Alan See. 2012. “How to Get Your Messages Retweeted.” MIT Sloan Management Review 53 (2): 61–66.

Mangold, W. Glynn, and David J. Faulds. 2009. “Social Media: The New Hybrid Element of the Promotion Mix.” *Business Horizons* 52 (4): 357–65.

Payne, John W., James R. Bettman, and Eric J. Johnson. 1993. The Adaptive Decision Maker. Cambridge University Press.

Petty, Richard E., John T. Cacioppo, and Rachel Goldman. 1981. “Personal Involvement as a Determinant of Argument-Based Persuasion.” *Journal of Personality and Social Psychology* 41 (5): 847–55.

Rossiter, John R., and Ale Smidts. 2012. “Print Advertising: Celebrity Presenters.” *Journal of Business Research*, (1) Scientific Muddling (2) Executional Elements In Advertising, 65 (6): 874–79.

Shugan, Steven M. 1980. “The Cost of Thinking.” Journal of Consumer Research 7 (2): 99–111.

Wilson, Elizabeth J., and Daniel L. Sherrell. 1993. “Source Effects in Communication and Persuasion Research: A Meta-Analysis of Effect Size.” *Journal of the Academy of Marketing Science* 21 (2): 101–12.

Essay 3: Velocity of Retweeting: Insights from Celebrity Tweets

3.1. Introduction

With the growing use of social media by individuals and marketers, it became an environment, where enormous amount of contents are generated and consumed every day. For instance, more than 500 million tweets are posted each day, and this number is increasing as well (Internet Stats 2015). This high pace of content generation leads to competition between contents in order to get attention of individuals. In addition, generation of new contents pushes older ones out of the competition, which subsequently makes the older ones to lose their relevance and effectiveness, and fade out into social media crowd. As a result, in context of communicating messages on a social network, brands not only look for a high volume of response to their messages, but also strive for a higher response rate for them. In other words, contents posted by brands on social media should generate higher response (of any kind) in a short period, and before new-coming contents crowd them out. Therefore, understanding what leads to a higher response rate of contents posted on social media is an important issue for marketers.

In addition to sending messages directly to their followers through their official social media accounts, marketers also employ celebrities, where celebrities send out messages (i.e. tweets) about brands on social media. For instance, T-Mobile sponsored a celebrity for their TV ads mainly because of the celebrity's strong presence on social media (Ad Age 2015). She communicated about T-Mobile through her Twitter account as well. Since celebrities are among the most followed users on social media, marketers not only reach celebrities' followers by sending messages through them, but also gain additional reach for their messages through receivers of the messages sharing them with their followers. While in this first scenario, marketers gain substantial exposure for their brands through reaching celebrities' followers, but in the latter case, marketers will advantage through spreading their messages through ordinary individuals, where they advocate the celebrity and brand. While both purposes are favorable, however, we focus on the second scenario in this research. In the face of content competition on social media, and in the case of celebrities' followers sharing messages posted by celebrities, a higher retransmission rate for the message is desired. Such a faster rate leads to a faster spread of the message on social media, which makes the message to stand out against the competition, and ultimately increases its effectiveness. Therefore, in the case of promoting message through celebrities on social media, understanding celebrity attributes and content characteristics that lead to a higher spread rate of the content posted by them on social media is critical.

Despite widespread use of celebrities on social media for marketing purposes, and importance of spread rate in the case of celebrity-sourced tweets, there are few insights regarding role of celebrity attributes and message characteristics on how fast messages posted by celebrities will spread on Twitter. Existing literature on this issue focuses on diffusion over social networks, and in context of online games (Schulze et al. 2014), adoption of social networks (Katona et al. 2011), and UGC (user-generated contents) (Liu-Thompkins and Rogerson 2012). Findings highlight effects of network size (Yoganarasimhan 2012), network structure (Watts and Dodds 2007), influencers (Aral and Walker 2012), content (Berger and Milkman 2012), and seeding strategies (Hinz et al 2011) on diffusion over social networks. However, these findings were in a different context, where retransmission mechanism and incentives to adopt are different from case of sharing a message on Twitter. In addition, current research examines spread rate of content posted through official account of a brand (celebrities), whereas previous researches mainly focused on UGC.

In this research, we consider tweets posted by celebrities, and examine effects of celebrity attributes and content characteristics on retweet rate of the tweets posted by them. Our empirical analysis is based on 2486 tweets posted by 60 celebrities across five categories of "Musician", "Television Actors/Actress", "Actor/Actress", "Athletes", and "Personalities" (TV and Radio celebrities). We considered the celebrities within the "2014 The World's Most Powerful Celebrities" published by Forbes. We considered "original" tweets posted by these 60 celebrities during month of February 2015. We observed number of retweets within approximate 15 minutes intervals during first 24 hours of the tweets' life. We examined role of celebrity's popularity measured by number of followers on how fast the tweets are retweeted. We also examined effects of several tweet attributes such as presence of photo, link to additional content, hashtag, question mark, exclamation mark on retweet rate. In addition, since Super Bowl, Grammy's, and Oscars happened during the month of February 2015, we identified tweets that contained words related to these events and examined role of such event related content on retweeting rate. Furthermore, by considering category of the celebrities posting about an event-related content, we examine effect of "match" between category of tweet sender and tweet content on retweet rate. We also considered match between the event-related content and event timing. For our modeling approach, we employed a Modulated Poisson Process model (Soyer and Tarimcilar 2008), where the observed count in every observed time-window follows a

Poisson model with a rate that varies based on time, tweet characteristics, and celebrity attributes. In addition to the observed celebrity and tweet characteristics, we also controlled for their unobserved characteristics through random effects.

Our results suggest that presence of a link in a tweet reduces the retweet rate, and type of the link affects the retweet rate as well. In addition, we found that presence of hashtag reduces the retweet rate, but presence of a photo in a tweet increases the retweet rate. Among the event related variables, only presence of Oscar related content had a higher retweet rate. However, an interesting finding is about timing of the event and event-related content. For all three cases of Super Bowl, Grammy, and Oscar, we found that contents related to these events have a higher retweet rate if they are posted when the event is in progress. Regarding the effect of celebrities, we found that celebrities' number of followers has a positive effect on retweet rate, but there are unobserved characteristics of celebrities, which result in higher retweet rate. Our findings could help brands and marketers in several ways. First, insights from effect of celebrities' number of followers and their unobserved characteristics help brands to choose celebrities for their marketing purposes on social media. Second, our findings on the effect of content characteristic can be used for designing branded-contents that generate higher response rate. Third, we empirically showed the effectiveness of event-related content during the event, which suggests that brands could increase their social media effectiveness by carefully planned campaigns for events.

In the next section, we discuss related literature. We follow that with a description of our data. Subsequently, we present our modeling approach, and next, we discuss our results. Finally, we conclude by discussion of limitation and opportunities for future research.

3.2. Related Literature

3.2.1. Content Diffusion on Social Networks

In general, the issue of how contents spread on social networks has been studied as a diffusion problem (Yoganarasimhan 2012, Liu-Thompkins and Rogerson 2012, Katona, Zubcsek, and Sarvary 2011, Peters et. al. 2013). In addition, other studies considered diffusion pattern over social networks in different contexts such as online games. Main objectives of these studies were to examine effects of content, network structure, influencers, and seeding strategies on "adoption" decision of individuals in a social network. Examples of such adoption decisions are viewing a YouTube video, signing up for a social network, and gaming decisions. The main application of these studies is to design viral contents and campaigns. Considering the context of current research, where celebrities posting tweets through their social media accounts, we discuss existing literature on the effect of network structure, influencers, and content attributes on diffusion of content within social networks, and highlight the differences of current research with previous studies.

The effect of network on diffusion of content has been examined in several dimensions. Yoganarasimhan (2012) studied size and structure of network around a node, and showed its causal effect on the diffusion of content distributed through that node in case of YouTube videos. She examined viewership of YouTube videos posted by different individuals over a month. By controlling video characteristics and network endogeneity in data structure, she concluded that first and second network around a node affect diffusion of content, where the first level accounts for initial growth, and the second level accounts for further growth. She also found that strong relationships around a node reduce the overall diffusion, but increase the diffusion within the network. Susarla et al (2012) also considered diffusion of YouTube videos over a two-month period, and showed the effect of network size on diffusion rate through a modified Bass model. They also showed the effect of social influence and social interaction on diffusion of YouTube videos. In a different context, Katona, Zubcsek, and Sarvary (2011) made the same conclusion about the effect of network size and density by considering adoption decision of a social network in relation to individual's network structure. They concluded that size of a network (number of connections) and density of the connections (how strong the connections are) have effects on individual's decision. They also found a counterintuitive result about the effect of network size on influence, where they showed as the network size around nodes increases, their power to influence their network decreases, and it negatively affects diffusion of content from such nodes. Liu-Thompkins and Rogerson (2012) also made the same conclusion regarding the effect of network size and network density by considering YouTube data. In addition, they confirmed the inverted-U relationship between the effect of network density and diffusion. Contents will not spread too far in a dense network. Hence, diffusion is low for such networks. On the other hand, in a less dense network, the connections are weak, which results in less diffusion of content within these networks. However, there is an optimum for the network connectivity, which could lead to a higher diffusion.

The role of influence of networks has been studies more extensively. Trusov, Bodapati, and Bucklin (2010) examined the login activity of individuals on a social network, and showed that there is heterogeneity across individuals in terms of influence on their networks. They highlighted that a segment of “influencers” exists, which influences the behavior of individuals connected to them. Aral and Walker (2012) in a randomized experiment of a 1.3 million Facebook users examined the influence and susceptibility of individuals with respect to their demographics. They also found that influential users in a network are less susceptible, and using such cluster of influential users is effective to spread content in a network. Watts and Dodds (2007) challenged the idea of opinion leadership and influence in the sense that the spread of information to the opinion leader and from them to the general audience might not be as simple as discussed; therefore, a closer look is important on studying the effect of influencers. They suggest that besides the effect of influential people, there might exist “a critical mass of easily influenced individuals”, which subsequently affect another group of susceptible people.

Clearly, contents posted on social media networks impact how individuals respond to them. de Vries et al (2012) considered the effect of content characteristics posted by brands on their fan pages on the number of “Likes” and “Comments”. They considered several aspects of content such as vividness, interactivity, and informational vs. entertainment on the two metrics. The found that a post with video increases number of likes, but inclusion of photo in a post does not have a significant effect. In addition, they found that a high level of interactivity such as a question affects number of likes negatively, but has a positive effect on number of comments. In the context of YouTube videos, Liu-Thompkins and Rogerson (2012) showed that entertainment and educational content affects popularity and ratings of the videos.

In the context of content diffusion on Twitter, Zaman et al. (2014) developed a model for predicting popularity of tweets in terms of retweet count and lifetime. Their model was able to predict the popularity of tweets within few minutes of posting. However, they did not consider tweet content, and took into account very few tweets (52 tweets) for their research setting. In addition, they only considered tweets with less than 1800 overall retweet count due to data collection issues.

Several new aspects of spread rate of content on social media have been addressed in this research. First, the previous research mainly focuses on UGC. In this research, we examine branded content, defined as contents originated from official social media accounts of celebrities. Clearly, there will be differences between content generated by ordinary individuals and contents posted by celebrities with millions of followers. Celebrities mainly post on social media in order to communicate with their followers, therefore, compared to UGC, celebrity originated content will be crafted more carefully, and with more advertising and branding purposes. Second, the size of celebrities' networks is relatively large compared to average social media users, and they have strong relationships with their followers, which are rooted in their popularity and credibility on a different domain (offline world). Therefore, the inverted-U shaped effect of network size and structure that have been discussed in the literature may not hold for celebrities. Third, most of the previous studies considered YouTube videos and diffusion of "view" count over time. The diffusion of YouTube videos' viewership is based on several factors such as recommended videos, friendship, subscribers, and traffic sources. However, the diffusion mechanism on Twitter is through individuals sharing tweets.

3.2.2. Celebrity Endorsement

The effect of using celebrities for advertising purposes has been an extensive area of research in marketing and communication (for an extensive literature review see Erdogan 1999). Generally, the effects of communicating a message through celebrities on persuasiveness (Petty, Cacioppo, and Schumann 1983, Petty, Cacioppo and Goldman 1981), brand recall (Kahle and Homer 1985), attitude (Kahle and Homer 1985), and purchase intention (Ohanian 1991) have been examined. In addition, the effect of celebrity on financial outcomes has been considered (Agrawal and Kamakura 1995). Several aspects of celebrities such as expertise, trustworthiness, and attractiveness have been put into examination. Expertise refers to knowledge and experience of celebrities in his/her field, while trustworthiness refers to credibility of the source in the subject of matter in the viewpoint of his/her audience in terms of having honesty in the matter. The attractiveness of celebrities consists of both physical attractiveness as well as non-physical attractiveness such as personality, lifestyle, or skills (Choi and Rifon 2012, Erdogan 1999). It has been shown that celebrity endorser's attractiveness, trustworthiness, and expertise affect advertising effectiveness (Dholakia and Sternthal 1977), but they may not lead to purchase intention. For instance, Ohanian (1991) found that while trustworthiness and attractiveness are important constructs for effective communication, they do not affect purchase intention, and only expertise has a role. That could be due to the reason that individuals expect attractiveness from celebrities, and due to nature of endorsement being a paid commercial, they do not see the celebrity as trustworthy source to the product they are endorsing.

In addition to the main effect of celebrity on communication effectiveness, many studies examined the "match-up" effect, where they examine a fit between celebrity and the product he/she is endorsing is considered (Kahle and Homer 1985, Kamins 1990). Although there is no clear definition of “match” concluded in the literature, but the argument is that if such a fit exists between the celebrity and the product, the message will be more persuasiveness, and it enhances likeability and attractiveness of the celebrity as well (Kamins and Gupta 1994). In addition, other findings of the “fit” effect suggest that attractive celebrities will be more advantageous when the product category is meant for becoming attractive such as beauty products (Kahle and Homer 1985, Kamins 1990). Such findings lead to several research studies in order to characterize celebrity attributes such that their match with the product category could lead to higher ad effectiveness. In conclusion, the findings across several studies on the occasions where the fit of product and celebrity are appropriate, and what aspects of celebrities are leading to higher effectiveness are sparse.

In the context of celebrities communicating message on social media, Jin and Phua (2014) showed that message promoted by celebrities with more followers affects product involvement, buying intention, and intention to spread eWOM. They also showed that celebrities with high number of followers are perceived as a more credible source in terms of attractiveness, trustworthiness, and competence.

3.3. Data

Data of this research consist of the tweets posted by the celebrities in the "2014 The World's Most Powerful Celebrities" published by Forbes, who have a twitter account and tweet in English (Forbes 2014). We only considered major celebrity categories and those related to the entertainment industry. Therefore, we did not consider "Director/Producer" and "Authors" celebrities in our sample. Table 3.1 provides number of celebrities, average, and median number of followers in each category.

Table 3.1 – Average and median number of followers in the categories

Category	# of Celebrities	Average # of Followers	Median # of Followers	# of Tweets
Musicians	23	19,777,258	11,184,645	874
Television actors/actress	8	6,621,265	5,057,503	257
Actor/Actress	9	4,034,975	2,126,076	128
Athletes	12	8,461,620	5,084,299	272
Personalities	13	11,517,057	3,155,550	955

Table 3.2 provides list of the celebrities in each category, their number of followers on twitter, number of tweets in our sample for each celebrity, as well as their rank in the published list by Forbes. We considered tweets posted by these celebrities in Table 3.2 during the month of February 2015. We only considered original tweets by the celebrities (tweets originally posted by celebrities) and did not consider the "Reply" and "Retweet" tweets, since the reply tweets is not targeted to all followers, and retweets are not posted by celebrities, and sender of the tweet is someone else. Our final sample includes 2486 tweets from 60 celebrities. Although our initial sample of celebrities include 65 celebrities, but due to having no activity during our observation period by few celebrities, our final sample includes 60 celebrities.

Table 3.2 – Celebrity names, number of followers and their other characteristics

Rank	Name	# of Followers	#twt	Rank	Name	# of Followers	#twt
	Musicians				**Actors/Actress**		
~~#1~~	~~Beyonce Knowles~~	~~1.38E+07~~		#10	Robert Downey Jr	3.51E+06	4
#3	Dr. Dre	2.37E+06	1	#23	Dwayne Johnson	8.14E+06	72
~~#6~~	~~Jay-Z~~	~~3.09E+06~~		#52	Leonardo DiCaprio	1.20E+07	3
#8	Rihanna	4.03E+07	38	#52	Matthew McConaughey	1.29E+06	2
#9	Katy Perry	6.45E+07	38	#60	Mark Wahlberg	2.13E+06	4
#13	Bon Jovi	1.60E+06	7	#63	Hugh Jackman	5.05E+06	42
#13	Bruno Mars	1.99E+07	12	~~#67~~	~~Ben Affleck~~	~~1.64E+06~~	
#17	Miley Cyrus	1.91E+07	8	#89	Gwyneth Paltrow	2.06E+06	1
#18	Taylor Swift	5.19E+07	45	~~#97~~	~~Cameron Diaz~~	~~5.30E+05~~	
#19	Lady Gaga	4.40E+07	61		**Athletes**		
#20	Kanye West	1.12E+07	38	#2	LeBron James	1.85E+07	53
#21	Calvin Harris	5.39E+06	67	#7	Floyd Mayweather	5.56E+06	18
#25	Bruce Springsteen	6.68E+05	1	#15	Kobe Bryant	6.19E+06	13
#26	Justin Timberlake	4.12E+07	12	#16	Roger Federer	2.62E+06	9
#28	One Direction	2.24E+07	78	#21	Tiger Woods	4.23E+06	1
#29	Paul McCartney	2.22E+06	34	#24	Rafael Nadal	7.42E+06	19
#31	Sean "Diddy" Combs	1.02E+07	244	#30	Cristiano Ronaldo	3.33E+07	8
#33	Justin Bieber	6.00E+07	42	#33	Kevin Durant	9.19E+06	25
#33	Jennifer Lopez	3.08E+07	102	#43	Novak Djokovic	3.85E+06	7
#38	Pharrell Williams	5.59E+06	14	#55	Dwyane Wade	4.61E+06	59
#47	Avicii	1.52E+06	12	#63	Maria Sharapova	1.48E+06	27
#51	Toby Keith	8.41E+05	12	#69	Serena Williams	4.56E+06	33
#70	Kenny Chesney	2.11E+06	8		**Personalities**		
	Television actors/actress			#4	Oprah Winfrey	2.66E+07	39
#63	Ashton Kutcher	1.66E+07	9	#5	Ellen DeGeneres	3.87E+07	177
#72	Neil Patrick Harris	1.33E+07	54	#31	Ryan Seacrest	1.32E+07	35
#74	Kevin Spacey	3.83E+06	5	#39	Glenn Beck	9.38E+05	74
#88	Bryan Cranston	1.72E+06	3	#42	Simon Cowell	1.13E+07	7
#54	Sofia Vergara	7.17E+06	28	#45	Jimmy Fallon	2.08E+07	79
#93	Kerry Washington	1.98E+06	112	#58	Gordon Ramsay	2.11E+06	52
#96	Zooey Deschanel	6.29E+06	3	~~#59~~	~~Rush Limbaugh~~	~~4.38E+05~~	
#97	Lena Dunham	2.03E+06	43	#60	Jon Stewart	3.16E+06	75
				#62	Howard Stern	1.60E+06	4
				#80	Kim Kardashian	2.84E+07	131
				#83	Dr. Phil McGraw	1.36E+06	101
				#86	Sean Hannity	1.08E+06	181

3.3.1. Dependent Variable

The variable of interest in this research is the retweet rate. In order to construct this variable, we observed the numbers of retweets for the tweets posted by the celebrities listed in Table 3.2, approximately every 15 minutes during the first 24 hours of the tweet life. The only exception from this data collection pattern is right after posting the tweet. Since the API requests for the retweet count for all tweets are made every 15 minutes (approximately), the first observation for every new tweet posted in between API requests will be made in proceeding API call, hence making the first period of observation to be less than 15 minutes. One possible issue with this observation pattern is that the time window of the first period might be significantly less than 15 minutes; hence, its retweet rate may not be comparable to the other periods, which are approximately 15 minutes. In order to overcome this issue, we combined the first and the second periods of observation. Therefore, the time window for the first period is approximately between 15 minutes and 30 minutes. All other time intervals will be approximately 15 minutes. However, we observed the exact timestamp of API calls, therefore, we take the exact length of the interval into account in our modeling approach. Figure 3.1 shows a visual perspective of

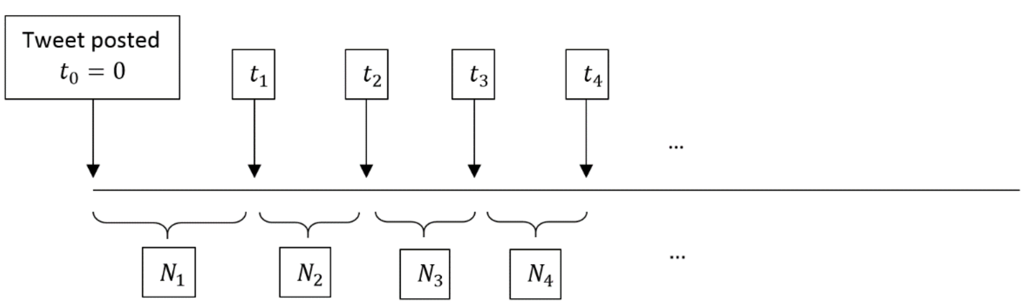

Figure 3.1 – A graphical representation of data collection algorithm

the data collection algorithm. Tweet i has been observed across K_i intervals, and at every observation points, t_k, $k = 1 \ldots K_i$, the cumulative retweet count has been observed, hence the retweet count in each interval has been constructed by subtracting retweet count in two consecutive observations points. Table 3.3 shows the number of intervals across all tweets. Figure 3.2 presents histogram of $N_i(t_k)$ across all tweets and all intervals, where $N_i(t_k)$ is retweet count of tweet $i = 1 \ldots S$ at interval t_k, $k = 1 \ldots K_i$, where S is the sample size, and K_i is the number of observed intervals for tweet i.

Table 3.3 – Frequency of number of intervals observed for the tweets

Number of Intervals	82	83	84	85	86	87	88	89	90	91
Frequency	56	26	3	3	2	2	55	647	1676	16

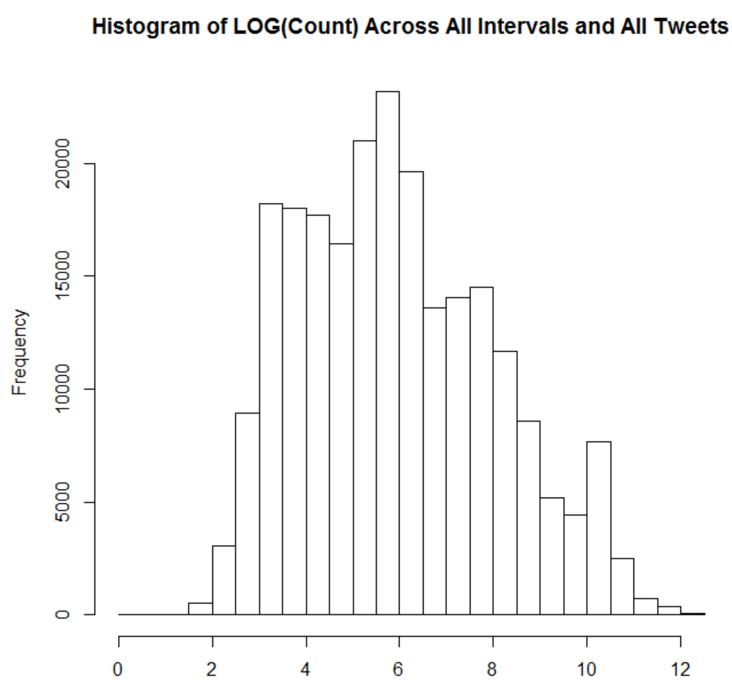

Figure 3.2 – Histogram of Log(observed count) across all intervals and all tweets

3.3.2. Tweet Characteristics

The variables that we want to examine its effect on the retweet rate are observed across different levels. We have tweet level covariates as well as celebrity level covariates. Below, we provide description of the covariates. Table 3.4 provides summary statistics of the covariates discussed below.

- **HTTP:** this is a dummy variable indicating presence of a link in a tweet
- **HTTP_FB:** if the link was directed to a Facebook page
- **HTTP_YT:** if the link was directed to a YouTube page
- **HTTP_INS:** if the link was directed to an Instagram image
- **TWITTER_PHOTO:** this is an indicator variable, which will take the value of one, if the tweet has one or more twitter images attached to it
- **HASH**: an indicator variable for the presence of hashtag in a tweet
- **QUEST:** an indicator variable for the presence of question mark in a tweet
- **EXCL:** an indicator variable for presence of exclamation mark in a tweet
- **NUM_EXCL:** the number of exclamation marks that are used in a tweet
- **SB:** an indicator variable if the content includes the words “SB49” or “Super Bowl” either directly or through a hashtag
- **CATG_ATH:** an indicator variable for the tweets that have been posted by an “athlete”
- **SB_TIME:** an indicator variable for the tweets that have been posted while Super Bowl 49 was in progress
- **GR:** an indicator variable if the content includes the word “Grammy” either directly or through a hashtag

- **CATG_MUS:** an indicator variable for the tweets that have been posted by a "musician"
- **GR_TIME:** an indicator variable for the tweets that have been posted while Grammy's award was in progress
- **OS:** an indicator variable if the content includes the word "Oscar" either directly or through a hashtag
- **CATG_ACT:** an indicator variable if the tweet is posted by either of the "Actor/Actress", "TV Actor/Actress", or "Personalities"
- **OS_TIME:** an indicator variable for the tweets that has been posted while Oscar's was in progress

3.3.3. Celebrity Characteristics

In addition to tweet level covariates, we included a celebrity specific variable into our model, where we measured celebrities' number of followers (NUM.FOL). In addition, we controlled for their unobserved characteristics through random effects.

Table 3.4 – Summary statistics of the covariates

		Yes		**No**
HTTP		1130		1356
	No		**Yes**	
HTTP_FB	1120		10	
HTTP_INS	575		555	
HTTP_YT	1032		98	
TWITTER_PHOTO		594		1892
HASH		1380		1106
SB		25		2461
CATG_ATH	22		3	
SB_TIME	19		6	
GR		61		2425
CATG_MUS	28		33	
GR_TIME	53		8	
OS		97		2389
CATG_ACT	23		74	
OS_TIME	88		9	
QUEST		262		2224
EXCL		870		1616

NUM.EXCL	Frequency
1	609
2	83
3	106
4	32
5	21
6	6
7	6
8	3
9	1
14	1
18	1
20	1

3.4. Modeling approach

Since we observed retweet count across multiple intervals during the first 24 hours of tweet's lifetime, we considered a Modulated Poisson Process model (Soyer and Tarimcilar 2008) that accounts for tweet characteristics and the nature of the process. Specifically, let $N_{ij}(t_k)$ to be observed number of retweets for tweet $i = 1 \dots N_j$ posted by celebrity $j = 1 \dots J$, occurred in the time interval $[t_{k-1}, t_k], k = 1 \dots K_i$. Since the number of retweets occurred during each interval across the lifetime of a tweet decreases, hence, the counting process is a non-homogeneous Poisson process with rate of $\lambda_{ij}(t_k)$. The $N_{ij}(t_k)$ follows a Poisson distribution with mean of $E\left(N_{ij}(t_k)\right)$ as below, where $E(.)$ is expectation function.

$$E\left(N_{ij}(t_k)\right) = \int_{t_{k-1}}^{t_k} \lambda_{ij}(u)\,\mathrm{du} = \Lambda_{ij}(t_k) - \Lambda_{ij}(t_{k-1})$$

where $\frac{\mathrm{d}\Lambda}{\mathrm{du}} = \lambda$. In order to incorporate effects of both tweet characteristics and time, we specify $\lambda_{ij}(t_k)$, rate of the Poisson process, as a function of time and tweet characteristics, as,

$$\lambda_{ij}(t_k) = \lambda^0(t_k)\, e^{\mu_{ij}}.$$

The first component, $\lambda^0(t_k)$, captures baseline behavior of retweet rate at k^{th} interval, and the second component, $e^{\mu_{ij}}$, captures the effect of tweet i characteristics from celebrity j, which will be linked to covariates subsequently. This specification captures effects of time elapsed after tweet has been posted, tweet and celebrity characteristics. By considering this specification, we can write $\Lambda_{ij}(t_k)$ as follows.

$$\Lambda_{ij}(t_k) = \Lambda^0(t_k)\, e^{\mu_{ij}}$$

The specification of $\Lambda^0(t_k)$ is critical, since it captures the underlying behavior of retweet rate across time. Figure 3.3 provides average of observed retweet count of tweets posted by a celebrity (Kanye West) in our sample across the first 24 hours after posting.

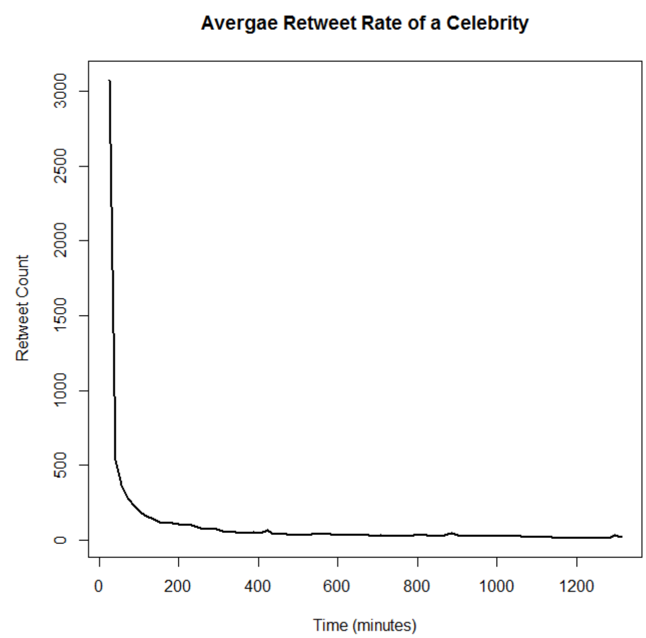

Figure 3.3 – Average retweet rate of a celebrity in our sample

Clearly, it can be seen that retweet rate has a decreasing pattern, such that after a tweet has been posted, it generates substantive amount of retweet, and subsequently will decline to a smaller rate. In order to capture such behavior, we choose a power law function for the baseline of retweet rate as below. This function has two parameters and captures both shape and scale of such decreasing function.

$$\Lambda^0(t_k) = \gamma\, t^\alpha$$

Based on the above specification, we can examine this assumption and its suitability. By taking the logarithm of both sides, we will get $\log \Lambda^0(t_k) = \log(\gamma) + \alpha \log(t)$. By making the scatter plot of log(cumulative number of retweets) and log(time), we could verify this linear relationship. Figure 3.4 shows scatterplot of cumulative retweet count and time (both in log scale) for a celebrity (Kayne West) with a linear fit in solid red.

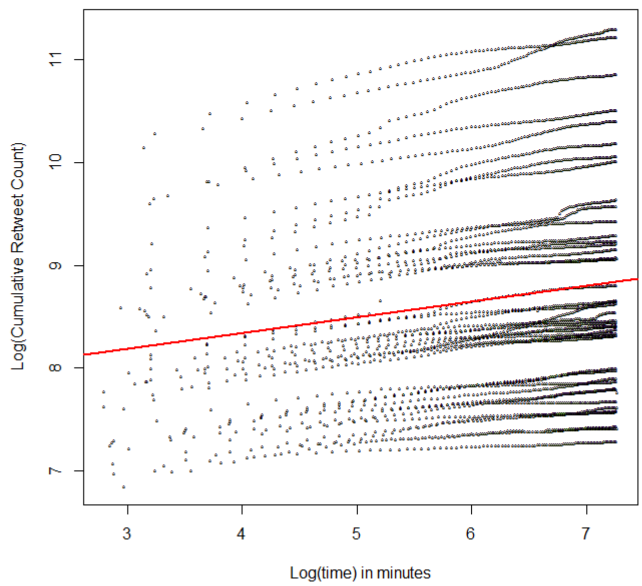

Figure 3.4 – Scatterplot of Log(Cumulative Count) vs. Log(time)

Clearly, it can be seen that there is a linear relationship between the two and least square line captures this relationship. Hence, the power law assumption for the baseline retweet rate is suitable. It should be note that slope of the fitted line (in red) is the shape of power law function, α. However, there might be differences between celebrities in terms of how their tweets spread on social network. Therefore, we make the shape parameter of power law function to be celebrity specific. Therefore, the baseline will be,

$$\Lambda_j^0(t_k) = \gamma\, t^{\alpha_j} \text{ where } \alpha_j \sim Gamma(a, b)$$

In addition, as it can be seen from Figure 3.4, there are significant differences between the tweets in terms of their retweet rate, which will be captured by including covariates into the model. However, there might be unobserved tweet characteristics that have not been captured by the observed tweet characteristics. In order to incorporate such effects, we specify the scale parameters of the baseline retweet rate, γ with respect to tweet specific covariates, hence we have γ_{ij} for tweet i posted by celebrity j, which will be specified as follows,

$$\log(\gamma_{ij}) = \phi + \epsilon_{ij}$$

The first component, ϕ, acts as the baseline scale (intercept) of the model, and the second component, ϵ_{ij}, is the random effect, which captures tweet heterogeneities. We assume a normal distribution with a zero mean and an unknown precision parameter, τ_ϵ for the random terms as below,

$$\epsilon_{ij} \sim N(0, \tau_\epsilon)$$

In order to include tweet characteristics into our model to examine what affects the retweet rate, we will write the tweet specific part of $\Lambda_{ij}(t_k)$ with respect to covariates described in section 3.3.2. In specific, we write the μ_{ij} with respect to observed characteristics of the tweets as follows.

$$\mu_{ij} = HTTP_i(\beta_1 + \beta_2\, HTTP\_FB_i + \beta_3 HTTP\_INS_i + \beta_4 HTTP\_YT_i)$$
$$+ \beta_5 TWITTER\_PHOTO_i + \beta_6 HASH_i + \beta_7 QUEST_i$$
$$+ EXCL_i(\beta_8 + \beta_9 NUM\_EXLC_i)$$
$$+ SB_i\,(\beta_{10} + \beta_{11} CATG\_ATH_i + \beta_{12} TIME\_SB_i)$$
$$+ GR_i\,(\beta_{13} + \beta_{14} CATG\_MUS_i + \beta_{15} TIME\_GR_i)$$
$$+ OS_i\,(\beta_{16} + \beta_{17} CATG\_ACT_i + \beta_{18} TIME\_OS_i) + CELEBRITY_j$$

In this specification, the main effect of including a link in a tweet, $HTTP$, is captured by β_1, while the effect of including specific links to Facebook, Instagram or YouTube are captured through β_2, β_3, and β_4, which should be interpreted as a shift from β_1. The effect of including exclamation mark is captured through β_8, while β_9 captures its interaction effect with the number of exclamation marks to control for the case of using excessive exclamation marks. For the case of Super Bowl, Grammy, and Oscars, we included a main effect of using content related to each of these events. In addition, we included an interaction term to examine the effect of using such content with the celebrity posting that who is in the same category, and using such content during the time that the related event is in progress. The $CELEBRITY_j$ is the effect of the celebrity who posted tweet i. We specify the celebrity effect with respect to the celebrity's number of followers as below. We also included a celebrity specific random effect, $\delta_j \sim Normal(0, \tau_\delta)$ in our model to capture the unobserved characteristics of the celebrities. Therefore, we have,

$$CELEBRITY_j = NUM.FOL_j\, \beta_{19} + \delta_j\,, \qquad j = 1 \ldots 60$$

3.4.1. Likelihood Function

The likelihood function is as below, where μ_i, $\Lambda_0(t_k)$ are specified as above.

$$L(N_i(t_k)|\mu_i, \gamma_i, \alpha) \propto \frac{\left(\Lambda_j^0(t_k)\, e^{\mu_i} - \Lambda_j^0(t_{k-1})\, e^{\mu_i}\right)^{N_{ij}(t_k)}}{N_{ij}(t_k)!}$$

3.4.2. Model Estimation

We estimated our model in Bayesian framework through BUGS software. We employed a Normal distribution with zero mean and large variance for the coefficients of the covariates, and scale of the baseline, ϕ. We used $Gamma(0.001{,}0.001)$ for a and b, random effect distribution of shape in the power law specification, and precision parameters of brand random effect (τ_δ), and tweet random effect (τ_ϵ).

3.4.3. Model Comparison

In addition to the model we discuss above, we also examined several modifications of the above model in order to select a model that represents our data better. Therefore, we evaluate each model in terms of DIC and select the model with the lowest DIC.

3.5. Results

Based on estimated DIC, the proposed model does well in terms of fit compared to the two reduced versions of the proposed model. Next, we present estimates of this model.

Table 3.5 – Model Comparison Results

Model	DIC
Proposed model	2433740
Proposed model with constant α for all celebrities	2775450
Proposed model without celebrity random effects (δ_j)	2433789

Table 3.6 shows estimated mean, median and 95% probability intervals of the model coefficients. Based on 95% credible intervals, presence of link in a tweet, HTTP, has a significant negative effect on retweet rate. However, the interaction effect of HTTP with a Facebook link (HTTP_FB) and YouTube link (HTTP_YT) is significant and positive. We should note that links types are included as interaction terms, and their estimated coefficients should be interpreted as shift from the main effect of HTTP. Therefore, since the mean main effect of HTTP is negative (-0.60), their positive effects (HTTP_FB 0.49 and HTTP_YT 0.27) will enhance the negative effect of link, but the overall effect still will be insignificant or negative. On the other hand, the interaction of HTTP with HTTP_INS is negative and significant, which means presence of Instagram link will make the effect even more negative. In general, the estimated coefficients of HTTP and their types suggest that presence of link will have a negative effect on retweet rate, and the link type makes the effect to be more or less negative. This result could be due to receivers of tweet interacting with the link (Noort et al. 2012, Liu and Shrum 2002), which makes the tweet not to be the focal point of the communication. In other words, individuals interact with the tweet by clicking on the link, which subsequently displays a webpage, an Instagram photo, or a YouTube video, and as a result, it reduces engagement with the tweet content and message, hence reduces the retweet rate. In addition, content of the link might affect individual's decision to share the content. Presence of photo in a tweet (TWITTER_PHOTO) has a significant positive effect on retweeting rate. Obviously, including a photo increases the engagement with the tweet, and draws the attention to the tweet and its text. Hence, it makes the tweet to stand out and increases the retweet rate.

Table 3.6 – Estimated coefficients of the covariates

	Variable	Mean	2.5%	Median	97.5%
β_1	HTTP **	-0.60	-0.72	-0.60	-0.50
β_2	HTTP × HTTP_FB **	0.48	0.04	0.49	0.88
β_3	HTTP × HTTP_INS **	-0.16	-0.27	-0.16	-0.01
β_4	HTTP × HTTP_YT **	0.27	0.10	0.27	0.43
β_5	TWITTER_PHOTO **	0.21	0.13	0.20	0.32
β_6	HASH **	-0.15	-0.20	-0.15	-0.11
β_7	QUEST **	-0.11	-0.22	-0.11	-0.01
β_8	EXCL **	-0.23	-0.30	-0.23	-0.10
β_9	EXCL × NUM_EXCL	0.03	-0.01	0.03	0.05
β_{10}	SB	-0.05	-0.50	-0.03	0.32
β_{11}	SB × CATG_ATH **	0.97	0.01	0.93	2.12
β_{12}	SB × TIME_SB **	0.79	0.09	0.80	1.63
β_{13}	GR	0.16	-0.24	0.18	0.48
β_{14}	GR × CATG_MUS	-0.20	-0.58	-0.19	0.26
β_{15}	GR × TIME_GR **	0.57	0.01	0.58	1.25
β_{16}	OS	0.36	-0.03	0.37	0.71
β_{17}	OS × CATG_ACT	-0.07	-0.48	-0.05	0.25
β_{18}	OS × TIME_OS **	0.87	0.36	0.84	1.61
β_{19}	NUM.FOL **	1.30	1.22	1.30	1.36

Presence of hashtag in a tweet (HASH) has a significant negative effect on retweet rate. Hashtags are another tweet component that individuals could interact with them. Receivers of the tweets could click on a hashtag and check out other tweets related to that hashtag. Therefore, presence of hashtag reduces the engagement with the tweet through drawing attention away from the tweet. However, by comparing the effects of hashtag and link, we could see that the effect of hashtag (-0.15) is less negative compared to the effect of link (-0.60). Presence of question mark, QUES, also has a negative significant effect on retweet rate, possibly due to the effort required to make it ready for retransmission. Presence of exclamation mark has a significant negative effect on retweet rate, and its interaction with number of exclamation has a positive effect, but it is not significant. This effect might be due to attractiveness and attention grabbing of the excessive use of exclamation mark, which signals sense of urgency and attention. Among the event related variables, none have a significant effect, but the effect of Oscar and Grammy related content (OS and GR) is positive. The interaction effect of celebrities' categories and the event is only significant in the case of Super Bowl. However, an interesting result is for the case of interaction of event related content and event timing. For all three cases, the interaction effect is significant and positive, which means a tweet related to these event will have a higher retweet rate when it is posted during the time that event is in progress. The reason behind this pattern might be that the tweet content is specifically about what is going on during the event; therefore, it will have a higher benefit for receivers to share that tweet. The effect of number of followers is also positive and significant, which means more number of followers will lead to a higher retweet rate. This can be due to the reason that more individuals are receiving the tweet, so it will have a higher retweet rate.

Table 3.7 – Other parameters of the model

	Mean	**2.5%**	**Median**	**97.5%**
$\bar{\alpha}$ – Mean Shape of Power Law Function	0.19	0.18	0.19	0.20
Var_{α}– Variance of Shape in Power Law Function	0.00	0.00	0.00	0.00
ϕ – Scale of Power Law Function	5.05	5.01	5.04	5.11
τ_{ϵ} – Precision of Tweet Random Effect	1.45	1.37	1.45	1.53
τ_{δ} – Precision of Celebrity Random Effect	0.63	0.41	0.62	0.87

Table 3.7 includes parameters of the model other than the coefficients of the regression. Mean of the power-law shape parameters across celebrities in the Poisson rate is 0.19, and its variance is less than 0.01. Figure 3.5 shows the boxplots of shape parameters across all celebrities. A larger value of shape parameter means that tweet will have a higher growth rate of retransmission count across its lifetime. The precision parameter of tweet random effects is 1.46. However, the precision parameter of celebrity random effects, τ_{δ},

is 0.65, which is relatively small, and it signals there are unobserved characteristics other than their number of followers that affect retweet rate. Figure 3.6 shows the boxplot of posterior estimates for celebrity random effects. It is interesting that for some celebrities in the positive side of the boxplot, such as “One Direction” or “Justin Bieber”, the number of followers is also large. For example, for the mentioned celebrities, numbers of followers are 22 and 60 million respectively, which is a relatively large number of followers compared to other celebrities.

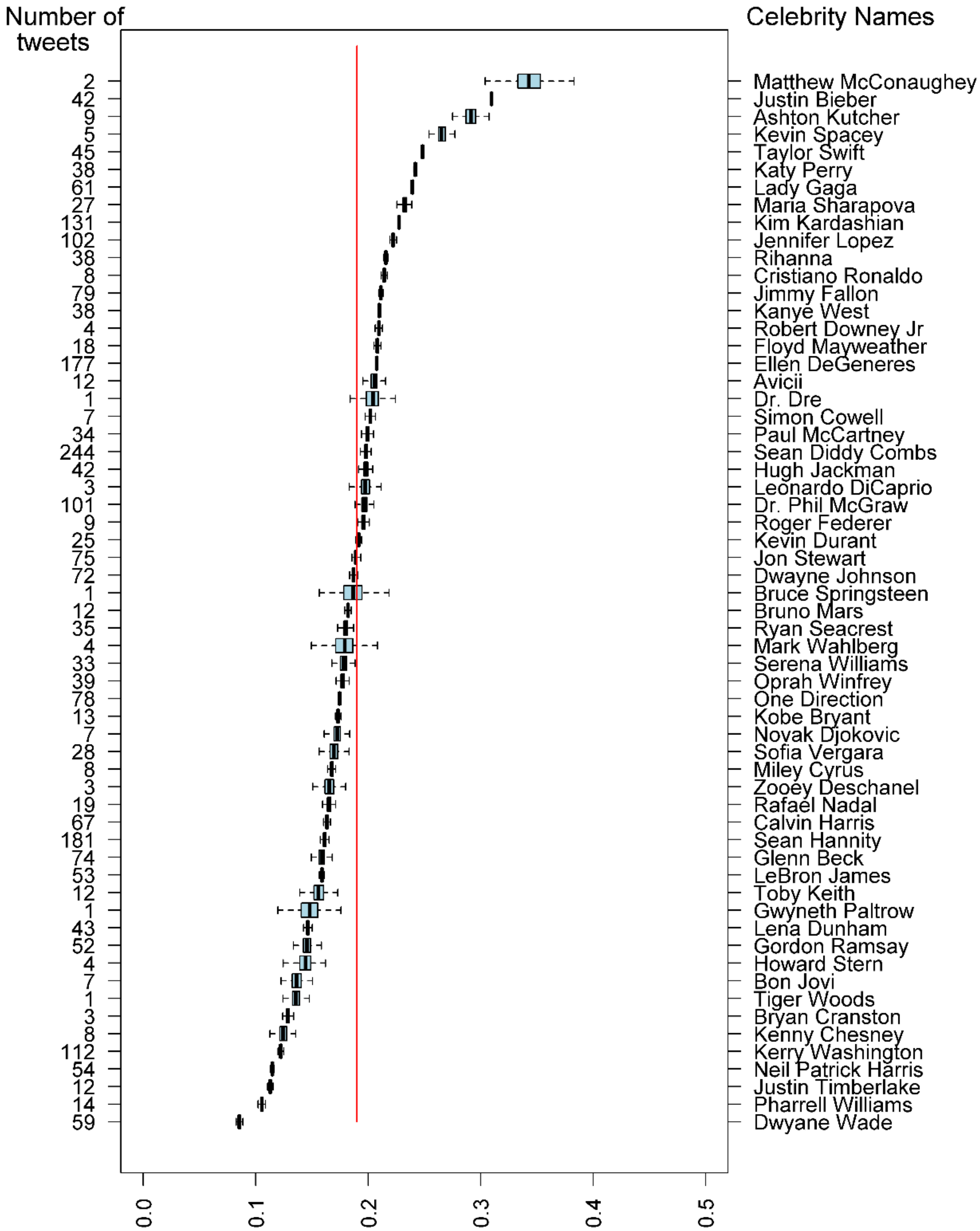

Figure 3.5 – Boxplot of the estimated power-law shape parameter across celebrities

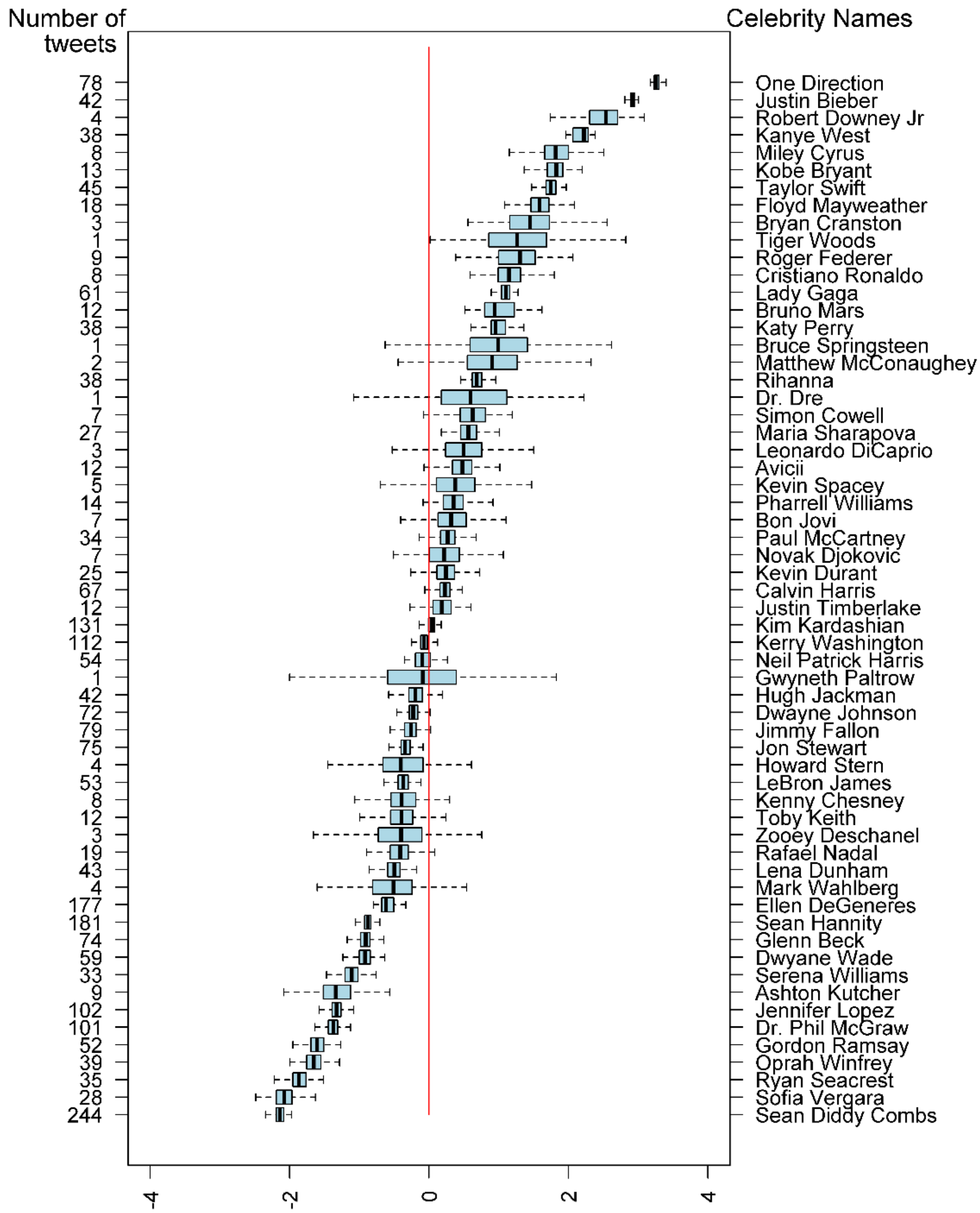

Figure 3.6 – Boxplot of the estimated celebrity random effects in the scale parameter of power law function

3.6. Discussion

In this research, we considered tweets posted by multiple celebrities, and observed number of retweets in several fifteen minutes intervals across first 24 hours of tweets life. We investigate the effect of tweet and celebrities characteristics on retweeting rate through a non-homogeneous Poisson process model. Findings suggest that presence of link and hashtag reduces retweet rate, while presence of photo increases the retweet rate. Links and hashtags are two components of a tweet that receivers can interact with and explore further contents through this interaction. Therefore, in the Twitter context, this result suggests that interactive contents could potentially draw attention away from the content, and therefore reduce the rate that individuals respond to them, specifically, in the case of retweeting. However, previous findings in the context of web browsing concluded that interactive content could increase satisfaction and effectiveness, and generally, interactivity leads to a better flow, which enhances the overall experience (Noort et al. 2012, Liu and Shrum 2002). Despite these findings, our findings suggest that interactivity in the case of Twitter might have a negative effect on retweet rate. This result could be due to nature of Twitter, since Twitter emphasizes short text (up to 140 characters), and it is meant for fast transmission of messages. It is not an information rich medium, and level of involvement with a tweet is relatively low. This result may not hold in the case of other social networks such as Facebook or Instagram, since they are content rich, and engagement with a particular content is higher that a tweet. In addition, we examined the role of event related content and its match with celebrity type as well as event timing. Our results suggest that event related content would have a higher retweet rate if it were posted during that event. The reason behind such effect may root in individual's presence on social media, once the event is in progress, hence, more people will see the tweet, and it increases the retweet rate. In addition, this effect could be due to relevance of the tweet content with the event, and the benefit that individuals might gain by retweeting that, whereas after the event and once everything is finished, the utility through sharing the content will not be gained.

3.6.1. Managerial Implications

Findings from this research have implication across several dimensions: content strategy on Twitter, celebrity selection for marketing purposes, and event related marketing. We demonstrated that while more interactive content such as a link or hashtag may increase the information that a tweet carries, but it might have a negative effect on the retweet rate, due to drawing attention away from the tweet to the hashtag or link. This could have important implications for marketing purposes and specifically content marketing. In the case of promoting a tweet with the goal of higher retweet rate, the focus should be the message content and even an image could be attached since it increase engagement but does not reinforce interaction. The implication of content engagement and interaction could be beyond Twitter, and it applies to other social networks such as Facebook or Instagram, although involvement level is higher for those social networks, and these effects may not be prominent. Our research also provides insight for firms about the effect of celebrities' number of followers and their unobservable characteristics, on how fast tweets posted by them are retransmitted. These insights can be employed for celebrity selection for marketing purposes. In addition to insights on celebrity and content, our research also shows the effect of event related content and event timing. It suggests that chance of retweeting will be higher if an event related content is posted during the events. This insight suggests that carefully planned campaigns for events could increase the spread rate of the content posted about that event, which subsequently affects brands reach and exposure.

3.6.2. Limitation, and Future Research

The limitations of this research are data collection pattern and its context. We only considered first 24 hours of tweets life, and tweets might show significant changes in their retweet pattern after a day, since tweets will diffuse deeper and wider into social networks, and that affects subsequent pattern of retweets. In addition, we only considered well-known celebrities, which bounds out inferences to these types of celebrities. For instance, a celebrity, who is famous to his/her niche followers, and is not a widely well-known celebrity, might have less number of followers compared to well-known ones, but due to their popularity within their niche, he/she might have a higher influence on their followers, and be more effective on sponsoring tweets for brands. Therefore, future research could investigate differences between well-known celebrities and "micro-celebrities" in terms of their influence on their followers. In addition, investigating the retweet rate across multiple categories and brands could provide more insight for branding purposes. One more area that could be a venue for future research is to consider the effect of interactive content and flow on social media, and investigate how it increase or reduces effectiveness of social media contents. In addition, generalizing our results to other social networks could be also insightful.

References

Ad Age, 2015, retrieved from http://adage.com/article/special-report-super-bowl/kim-kardashian-debuts-t-mobile-super-bowl-ad-conan/296808/

Agrawal, Jagdish, and Wagner A. Kamakura. 1995. "The Economic Worth of Celebrity Endorsers: An Event Study Analysis." *Journal of Marketing* 59 (3): 56–62.

Aral, Sinan, and Dylan Walker. 2012. "Identifying Influential and Susceptible Members of Social Networks." *Science* 337 (6092): 337–41.

Berger, Jonah, and Katherine L Milkman. 2012. "What Makes Online Content Viral?" *Journal of Marketing Research* 49 (2): 192–205.

Choi, Sejung Marina, and Nora J. Rifon. 2012. "It Is a Match: The Impact of Congruence between Celebrity Image and Consumer Ideal Self on Endorsement Effectiveness." *Psychology & Marketing* 29 (9): 639–50.

De Vries, Lisette, Sonja Gensler, and Peter S. H. Leeflang. 2012. "Popularity of Brand Posts on Brand Fan Pages: An Investigation of the Effects of Social Media Marketing." *Journal of Interactive Marketing* 26 (2): 83–91.

Dholakia, Ruby Roy, and Brian Sternthal. 1977. "Highly Credible Sources: Persuasive Facilitators or Persuasive Liabilities?" *Journal of Consumer Research* 3 (4): 223–32.

Erdogan, B. Zafer. 1999. "Celebrity Endorsement: A Literature Review." *Journal of Marketing Management* 15 (4): 291–314.

Forbes, 2014, retrieved from http://www.forbes.com/celebrities/list/

Hinz, Oliver, Bernd Skiera, Christian Barrot, and Jan U Becker. 2011. "Seeding Strategies for Viral Marketing: An Empirical Comparison." *Journal of Marketing* 75 (6): 55–71.

Internet Stats, 2015, retrieved from http://www.internetlivestats.com/twitter-statistics/

Jin, Seung-A. Annie, and Joe Phua. 2014. "Following Celebrities' Tweets About Brands: The Impact of Twitter-Based Electronic Word-of-Mouth on Consumers' Source Credibility Perception, Buying Intention, and Social Identification With Celebrities." *Journal of Advertising* 43 (2): 181–95.

Kahle, Lynn R., and Pamela M. Homer. 1985. "Physical Attractiveness of the Celebrity Endorser: A Social Adaptation Perspective." *Journal of Consumer Research* 11 (4): 954–61.

Kamins, Michael A. 1990. "An Investigation into the 'Match-Up' Hypothesis in Celebrity Advertising: When Beauty May Be Only Skin Deep." *Journal of Advertising* 19 (1): 4–13.

Kamins, Michael A., and Kamal Gupta. 1994. "Congruence between Spokesperson and Product Type: A Matchup Hypothesis Perspective." *Psychology and Marketing* 11 (6): 569–86.

Katona, Zsolt, Peter Pal Zubcsek, and Miklos Sarvary. 2011. "Network Effects and Personal Influences: The Diffusion of an Online Social Network." *Journal of Marketing Research* 48 (3): 425–43.

Liu-Thompkins, Yuping, and Michelle Rogerson. 2012. "Rising to Stardom: An Empirical Investigation of the Diffusion of User-Generated Content." *Journal of Interactive Marketing* 26 (2): 71–82.

Liu, Yuping, and L. J. Shrum. 2002. "What Is Interactivity and Is It Always Such a Good Thing? Implications of Definition, Person, and Situation for the Influence of Interactivity on Advertising Effectiveness." *Journal of Advertising* 31 (4): 53–64.

Ohanian, Roobina. 1991. "The Impact of Celebrity Spokespersons' Perceived Image on Consumers' Intention to Purchase." *Journal of Advertising Research* 31 (1): 46–54.

Peters, Kay, Yubo Chen, Andreas M. Kaplan, Björn Ognibeni, and Koen Pauwels. 2013. "Social Media Metrics — A Framework and Guidelines for Managing Social Media." *Journal of Interactive Marketing*, Social Media and Marketing, 27 (4): 281–98.

Petty, Richard E., John T. Cacioppo, and Rachel Goldman. 1981. "Personal Involvement as a Determinant of Argument-Based Persuasion." *Journal of Personality and Social Psychology* 41 (5): 847–55.

Petty, Richard E., John T. Cacioppo, and David Schumann. 1983. "Central and Peripheral Routes to Advertising Effectiveness: The Moderating Role of Involvement." *Journal of Consumer Research* 10 (2): 135–46.

Schulze, Christian, Lisa Schöler, and Bernd Skiera. 2013. "Not All Fun and Games: Viral Marketing for Utilitarian Products." *Journal of Marketing* 78 (1): 1–19.

Soyer, Refik, and M. Murat Tarimcilar. 2007. "Modeling and Analysis of Call Center Arrival Data: A Bayesian Approach." *Management Science* 54 (2): 266–78.

Susarla, Anjana, Jeong-Ha Oh, and Yong Tan. 2011. "Social Networks and the Diffusion of User-Generated Content: Evidence from YouTube." *Information Systems Research* 23 (1): 23–41.

Trusov, Michael, Anand V Bodapati, and Randolph E Bucklin. 2010. "Determining Influential Users in Internet Social Networks." *Journal of Marketing Research* 47 (4): 643–58.

Van Noort, Guda, Hilde A. M. Voorveld, and Eva A. van Reijmersdal. 2012. "Interactivity in Brand Web Sites: Cognitive, Affective, and Behavioral Responses Explained by Consumers' Online Flow Experience." *Journal of Interactive Marketing* 26 (4): 223–34.

Watts, Duncan J., and Peter Sheridan Dodds. 2007. "Influentials, Networks, and Public Opinion Formation." *Journal of Consumer Research* 34 (4): 441–58.

Yoganarasimhan, Hema. 2011. "Impact of Social Network Structure on Content Propagation: A Study Using YouTube Data." *Quantitative Marketing and Economics* 10 (1): 111–50.

Zaman, Tauhid, Emily B. Fox, and Eric T. Bradlow. 2014. "A Bayesian Approach for Predicting the Popularity of Tweets." *The Annals of Applied Statistics* 8 (3): 1583–1611.

Nima Y. Jalali, Ph.D.

nima@uwm.edu

EDUCATION

- Ph.D., University of Wisconsin-Milwaukee, May 2015
    - Major: Marketing
    - Minor: Quantitative Methods/Econometrics
- MBA, Marketing, University of Tehran, Iran, 2009
- B.Sc., Engineering, University of Isfahan, Iran, 2006

RESEARCH INTERESTS

- Social Media
- Digital Marketing
- Consumer Word of Mouth
- Quantitative Marketing
- Bayesian Modeling

JOURNAL PUBLICATIONS

- Ebrahimi, N., Jalali, N.Y., Soofi, E.S. (2014), "Comparison, Utility, and Partition of Dependence under Absolutely Continuous and Singular Distributions", *Journal of Multivariate Analysis*, 131, October 2014, pp 32–50.
- Ebrahimi, N., Jalali, N.Y., Soofi, E.S., Soyer, R. (2014), "Importance of Components for a System", *Econometric Reviews*, 33 (1-4) pp 395-420.

PRESENTATIONS

- "The Role of Brand and Category Intangibles in the Virality of Tweets", INFORMS Marketing Science Conference, Atlanta, GA, June 2014.
- "Uncertainty Reduction and Departure from Independence", Section on Statistics in Marketing, Joint Statistical Meetings, San Diego, CA, August 2012.
- "Why We All Should Be Bayesians: Illustrated by Examples", Lubar School of Business, UWM, Joint presentation with Ehsan Soofi, September 2011.
- "Dynamics of Satisfaction: A Regime Switching Ordinal Model for Affective and Cognitive factors", INFORMS Marketing Science Conference, Houston, TX, June 2011.

CONFERENCE ATTENDANCE

- INFORMS Marketing Science, Atlanta, June 2014
- Seminar on Bayesian Inference in Econometrics and Statistics (SBIES), Chicago, May 2014
- Joint Statistical Meeting, San Diego, August 2012
- Seminar on Bayesian Inference in Econometrics and Statistics, St. Louis, April 2011
- INFORMS Marketing Science, Houston, June 2011
- Quantitative Marketing and Economics, Los Angeles, October 2010