Social TV:

Linking TV Content to Buzz and Sales

Research-in-Progress

Shawndra Hill
The Wharton School
University of Pennsylvania
Philadelphia, PA 19104
shawndra@wharton.upenn.edu

Adrian Benton
The Wharton School
University of Pennsylvania
Philadelphia, PA 19104
adrian.donald.benton@gmail.com

Abstract

"Social TV" is a term that broadly describes the online social interactions occurring between viewers while watching television. In this paper, we show that TV networks can derive value from social media content placed in shows because it leads to increased word of mouth via online posts, and it highly correlates with TV show related sales. In short, we show that TV event triggers change the online behavior of viewers. In this paper, we first show that using social media content on the televised American reality singing competition, *The Voice*, led to increased social media engagement during the TV broadcast. We then illustrate that social media buzz about a contestant after a performance is highly correlated with song sales from that contestant's performance. We believe this to be the first study linking TV content to buzz and sales in real time.

Introduction

"Social TV" is the term used to describe the current integration of social media interaction with television programming. Social television has sought to recapture the early days of television when families gathered in their homes to share the experience of watching television together (Dumenco 2011). Over the past several years, online social media communities, such as message boards, Twitter, and Facebook, have become a virtual water cooler for today's tech-savvy television viewers. With the proliferation of social media applications and Smartphone technology, social interaction around television programming can be shared amongst millions of viewers simultaneously. Worldwide, it is estimated that, on average, 10 million public online comments are made each day related to television content (Talbot 2011). Twitter and other social media platforms have "become an integral outlet for TV viewers who are looking to express themselves while watching broadcasts of their favorite television programs." This "backchannel" of communication during TV shows also has led to a resurgence in people's interest in watching live shows (Proulx and Shepatin 2012).

In this paper, we present work aimed at quantifying the effect of Social TV exposure on viewership and engagement in the program, as evidenced by both the amount of discussion and sales associated with content on the show. We focus on a popular American reality singing talent show, which made headlines last season (2011) due to its use of social media content within the show. We show that the content of the television program has a strong correlation with the online community's response, sentiment and purchasing behavior. In the future, we will link qualities of the specific content to the online buzz, including emotion, controversy, and sentiment, all shown in prior research to drive word of mouth. To our knowledge, this is the first academic study linking Social TV strategy to buzz and sales outcomes. We believe it is an important first step to indicating that TV show trigger events can be designed to drive online word of mouth.

Dataset

The Voice, an American reality singing talent show, made headlines last season due to its use of social media content during the show. The singing contestants on *The Voice* are mentored by one of four coaches: Adam Levine, Cee Lo, Kristina Aguilera, or Blake Shelton. Throughout the season, viewers widely discuss the show, contestants, and coaches. *The Voice* frequently displays Tweets, Hashtags and Facebook messages on the show during the broadcast in an effort to keep viewers engaged. The audience engages using social media outlets, such as Twitter and Facebook. The audience also participates in the show by voting for contestants. Viewers vote, in part, by purchasing songs from the online music store, iTunes, providing data that can be tracked over time. In this study, we will combine three types of data: time-stamped TV dialogue, social media buzz, and sales.

We primarily use Twitter as our test bed. Twitter makes a subset of its data available to researchers through a portal-supplied application programming interface (API). The Twitter Streaming API was used to collect real-time Twitter statuses that contain pre-specified terms and tags related to the show as they are posted. All data are publicly available. In addition to the time-stamped tweets, we collected the exact times that specific tweets appeared on the TV as part of the viewing. The data was made anonymous for research purposes by mapping all users in our database to our own set of anonymous IDs. Metadata was replaced by corresponding anonymous IDs whenever statuses made reference to usernames, Twitter IDs, "@" and reply-to mentions of other usernames. Twitter status updates are particularly amenable to anonymizing because sensitive fields, such as usernames and personal names are encoded in separate fields in the JSON object returned by the Twitter API, and other users' tweets are prefaced by an "@" character within the text body of the status. In addition, we sampled the Twitter networks of the judges and contestants appearing on the show continuously during the course of the season, at a rate of once per fifteen minutes. We also collected iTunes sales data, and were able to track the songs that were performed on the show in the top 1000 and top 100 of iTunes rankings.

In total, we collected over 5.6 million tweets from February 5th to April 28th. Of those, 3.3 million were contributed during the show. We restricted the set of tweets to those made by accounts that are public and associated with the show. Of those tweets, 3.0% were displayed on the TV during the show and 97.0% were not. Each tweet is a candidate for being retweeted. Some are retweeted in great numbers while others are not.

Research Questions and Approach

In this work, we aim at addressing two questions. The first question is the following: Are the social TV strategies employed by *The Voice* correlated with increased social network buzz? We attempt to answer this question by evaluating two strategies used by the show as part of viewing. The first strategy is the placement of specific tweets on the screen. These tweets often come from contestants on the show, judges, or the MC. The second strategy is the placement of the hashtag (#thevoice) on the show, which serves as a reminder to viewers to continue to tweet and get involved with the conversation about the voice. We also compare the two strategies. The second question is as follows: Is online buzz correlated with sales outcomes? We answer this question by using panel data analysis to correlate online buzz for contestants and iTunes sales for their songs.

Placement of Specific Tweets on Screen

While placing show-related tweets on screen has become increasingly popular on TV programs, the effectiveness of this strategy in promoting users' online activities has yet to be empirically verified. As a first attempt toward these ends, we test the following: *Does a tweet displayed on the show cause a much higher engagement level with viewers, as measured by the number of retweets*? We use propensity score matching over the control variables in order to account for potential endogeneity effects (e.g., tweets that are more likely to engage viewers are more likely to be displayed). We then use linear regression to correlate tweet features with retweeting. We include a number of control variables, the attribute of interest to test our hypothesis, and the dependent variable, which is the log-transformed number of retweets. Table 1 provides a list of definitions of the variables used.

Table 1: Displayed tweet variable description	
	Description
Control Variables	
log(1+number of followers)	Number of followers of message poster
log(general buzz)	Total number of tweets occurring within an hour around tweet
time from start	Number of seconds since beginning of show
log(1 + num mention followers)	Number of followers for a judge/contestant mentioned in the tweet
num hashtags	Number of hashtags present in the tweet
num exclamation points	Number of exclamation points in the tweet, captures arousal
Independent Variable	
was displayed	Whether or not the tweet was displayed on TV
Dependent Variable: log(1 + num of retweets)	Number of retweets elicited by tweet

Placement of General Hashtags on Screen

A hashtag is simply a token preceded by a # sign, which can be inserted into tweets in order to index them under a particular topic. The hashtag can pull people interested in the same TV program together and facilitate communication between them. Given that viewers' attention might be attracted largely by the show, we posit that *the buzz during commercial breaks should be greater when they are preceded by a hashtag than when they are not*. In this work, we compare the change in buzz during a commercial break when #thevoice is shown prior to commercial break buzz and when #thevoice is not shown. We do this by fitting a linear model to the data to predict the proportion change in buzz during a commercial break, controlling for other possible factors that would lead to buzz. The proportional change in buzz is calculated by counting the number of program-related tweets that occurred during the first three minutes of the commercial break, then observing the total activity that occurred in the three minutes prior to the commercial. We define a commercial break as being preceded by a hashtag if a hashtag is displayed at most two minutes prior to the start of that break. A description of the variables used is included in Table 2. We found a total of 180 commercial breaks in our data, 53 of which were preceded by a hashtag. The timestamps for display of hashtags and tweets on screen were labeled manually using recorded episodes.

Table 2: Commercial break variable description	
	Description
Control Variables	
episode number	Episode the commercial occurred in, controls for episode-level effects.
episode type	The format of the current episode (e.g., audition, performance).
time from start	Number of seconds since beginning of show
time from performance	Number of seconds since last performance
log(1 + num performer followers)	Number of followers for the last performer/s
time from performance * log(1 + num performer followers)	Interaction between time from performance and log(1 + num performer followers)
time from tweet	Number of seconds since a tweet was displayed on the show
Independent Variable	
preceded by hashtag	Whether or not the commercial break was preceded by a hashtag in the previous two minutes.
Dependent Variable: log(1 + prop buzz change)	Change in the proportion of buzz before commercial to during. Log-transformed to account for skewed distribution.

Comparing General Hashtags to Specific Tweet Placement on Screen

We attempted to compare the two different strategies of placing a hashtag on screen to placing a tweet on screen in terms of their effects on Twitter responses. To do so, we considered all instances where a tweet/hashtag was displayed on screen, and noted the proportional change in buzz from three minutes before the display compared to three minutes afterward. The variables we controlled for are included in Table 3.

Table 3: Comparison of hashtag to tweet placement variable description

Factor type	Factor	Description
Episode feature	IS_SECONDHALF	Is in the second half of the season, after April 9 showing. All "specific" tweets occurred during the social media room after this date.
	EPISODE_NUM	The episode number, from 0 to 20.
Episode context	SECS_SINCE_PERFORMANCE	Seconds since the last performance.
	LAST_PERFORMERS_POPULARITY	The sum of number of followers for each of the previous performers.
	SECS_SINCE_COMMERCIAL	Seconds since the last commercial break.
	SECS_SINCE_START	Seconds since the start of the show.

Linking Buzz to Sales

	Round1						Round2						Round3			Round4		
	Buzz	Prop. Buzz	Rank	Buzz	Prop. Buzz	Rank	Buzz	Prop. Buzz	Rank	Buzz	Prop. Buzz	Rank	Buzz	Prop. Buzz	Rank	Buzz	Prop. Buzz	Rank
	Kim			Karla			Pip			Mathai			Katrina			Tony		
Adam Levine	953	0.05	454	1026	0.05	495	5017	0.25	182	4739	0.24	92	1471	0.07	413	6903	0.34	101
							10400	0.42	160	3031	0.12	215	3438	0.14	49	8077	0.32	20
													3374	0.33	MISS	6734	0.67	MISS
																24835	1.00	4
	Tony			Erin			James			Cheesa			Jamar			Juliet		
Cee Lo Green	1844	0.06	276	2645	0.09	NA	8030	0.28	100	1216	0.04	597	5870	0.20	91	9093	0.32	22
							7853	0.30	252	2598	0.10	108	8561	0.33	53	6983	0.27	34
													8583	0.47	MISS	9800	0.53	MISS
																13793	1.00	10
	Sera			Moses			Jesse			Ashley			Lindsey			Chris		
Christina Aguilera	1709	0.11	530	2562	0.16	369	2491	0.16	78	1353	0.09	229	3235	0.20	26	4478	0.28	79
							11665	0.59	123	2455	0.12	90	2676	0.14	95	3018	0.15	99
													3484	0.4	MISS	5141	0.6	MISS
																17744	1.00	12
	Naia			Charlotte			Jordis			RaeLynn			Erin			Jermaine		
Blake Shelton	3072	0.18	95	2236	0.13	62	2628	0.16	96	4849	0.29	NA	1568	0.09	139	2354	0.14	124
							5863	0.34	126	6756	0.39	NA	1028	0.06	233	3690	0.21	38
													1454	0.31	MISS	3301	0.69	MISS
																21738	1.00	14

Figure 1: Buzz Proportions and Sales Rank. For each round of viewer voting, we report the total amount of buzz (tweets), the proportion of buzz for the contestant within their coach's group, and the highest iTunes sales rank for each contestant. Contestants further to the left were voted off in earlier rounds. Contestants in pink were removed from the competition by coach decision, not as a result of viewer voting. Round 3 sales rank is omitted due to missing data. This table shows that proportion of buzz within a team was a perfect predictor of which contestants were voted off the show by viewers.

For each round of shows, during the live performances, viewers were given the opportunity to vote for which contestants they would like to continue. Votes were cast by calling a toll-free number with their choice of contestant. Viewers could also vote directly on *The Voice* website, Facebook page, or a mobile phone app. Finally, they could send phone text messages to the show, or purchase the contestant's song on iTunes. The contestants who received the fewest “votes” during a voting period within their coach's current set of contestants were removed from the competition. Figure 1 displays the amount of buzz on Twitter about each contestant, the proportion of buzz they received from within their coach's group, and the highest sales rank they achieved during each voting period. From the figure, it is evident that mentions of a contestant on Twitter were strong predictors of whether that person would advance to the next round for any given round. We use the online buzz and sales data over time to better understand the relationship between content of the show and sales. For the time series data, time was divided into windows of 20 minute granularity. Fields that were collected at a granularity higher than 20 minutes were populated by linear interpolation between the times they were sampled. The features that were considered are described in Table 4.

Table 4: Features to correlate online buzz about contestants with their song's sales rank

Feature name	Feature description	Sample granularity
ITunes top 100 song sales	For each song performed on The Voice (and made available on ITunes) the number of sales it generated at this point of time, if it was in the top 100 highest selling songs on ITunes at this time.	1 sample/20 minutes
ITunes top 1024 song rank	For each song performed on The Voice (and made available on ITunes) the rank of this song, if it was in the top 1024 best-selling songs at this time.	1 sample/20 minutes
Total buzz	Total number of Voice-related tweets within this 20-minute window.	Continuously
Average tweet sentiment	Average sentiment of Voice-related tweets in this 20 minute window. Automatically generated by sentiment classifier.	Continuously
Performer-related buzz	Total number of tweets containing a mention of a specific performer within this 20 minute window.	Continuously
Performer-generated buzz	Total number of tweets generated from this performer's Twitter account in this 20 minute window.	Continuously
Performer network size	Number of Twitter followers for this performer at this time.	1 sample/20 minutes

To control for song quality, we also collected metadata for each song performed on *The Voice*. The features collected were as follows: Year of original song release, the artist and album of the original song release, song genre (pop, opera, country, jazz, alternative rock, R&B, rock, soul, Christian rock, indie, soft rock, folk rock, dance, power ballad, grunge, disco, or hip hop), number of original song sales, number of original album sales, and song's highest billboard rank. Sales data was gathered from the music sales database at http://riaa.com/ at the granularity of 500,000 (gold), 1,000,000 (platinum), and X * 1,000,000 (multi-platinum) total album sales. Top Billboard ranking was gathered from http://www.billboard.com/. The song genre, year of original release, album, and artist also was gathered

Results

We first test whether displaying a tweet on screen increases the number of retweets it receives, by first matching tweets in each group via propensity score matching, then using a linear regression model to estimate the now-controlled weight. In Table 5, we show that displaying the tweet on the screen is correlated with more retweets, even after controlling for the control variables mentioned in Table 1. After propensity score matching, we matched 100 of the 103 tweets that were displayed on screen to a control with a suitably close propensity score (within two standard deviations of the observed difference of closest matched pairs). After matching, we note that the means of the two groups, displayed and not displayed on television, are significantly different by t-test ($p = 6.88*10^{-7}$, $mean_1 = 1.65$, $mean_2 = 3.45$).

Table 5: Total proportion change in buzz as a function of whether a tweet or hashtag was displayed on screen, likely controls, and which showing this episode was.

Factor type	Factor	Total Tweet Proportion Change
	Intercept	1.683***
Episode feature	Is season second half	1.418*10^-2
	Episode number	-1.971*10^-2***
Episode context	Seconds since performance	3.15*10^4***
	Sum of last performers network sizes	-8.648*10^-9**
	Seconds since performance X Sum of last performers network sizes	-8.765*10^-11***
	Seconds since commercial	-2.887*10^-4***
	Seconds since start	2.088*10^-5***
Online context	Total number of tweets around (60 min/10 min)	-2.81*10^-6***
Hypothesis	Is general social media mention	-7.515*10^-2* (p=0.049)
	Is west coast showing	-3.432*10^-1*** (p<2*10^-16)
	Is general social media mention X Is season second half	-3.5*10^-2
	Is general social media mention X Is west coast showing	8.849*10^-2 (p=0.015)
	Adjusted R^2	0.339

Next, we test whether displaying a general #thevoice hashtag before a commercial leads to more online viewer engagement during the commercial, after matching the controls listed in Table 2. We find that engagement stays at significantly higher levels during commercials when the hashtag for the show is displayed on the TV. We omit the detailed results of this particular model in the interest of space.

We next compare specific tweets to general hashtags. Table 6 displays the weights and significance levels for the proportion change in buzz given either a tweet or a hashtag was posted on screen, which showing it was posted on, along with control variables. From the learned model, it is clear that specific tweets displayed on screen seem to generate more buzz than displayed hashtags displayed on screen. However, hashtags displayed on screen seem to increase west-coast (delayed showing) buzz more than tweets, perhaps because the audience is already aware that those tweets have been made and are "old news."

Finally, we link buzz to sales using a random effects model. We control for the variables discussed in Table 4 as well as prior song quality indicators. We find that contestant level buzz during an episode is significantly correlated with the contestant's sales for the week (see Figure 2 for an illustration). In addition, we find that the network size of the contestant, the song quality attributes, the song's genre, and the contestant's gender all have a strong significant relationship with song sales, indicating that buzz alone does not explain sales rank; rather, the value of the song to users as well as buzz also impact sales.

Table 6: Total proportion change in buzz as a function of whether a tweet or hashtag was displayed on screen, likely controls, and which showing this episode was.		
Factor type	**Factor**	**Total Tweet Proportion Change**
	Intercept	1.683***
Episode feature	Is season second half	1.418*10^-2
	Episode number	-1.971*10^-2***
Episode context	Seconds since performance	3.15*10^4***
	Sum of last performers network sizes	-8.648*10^-9**
	Seconds since performance X Sum of last performers network sizes	-8.765*10^-11***
	Seconds since commercial	-2.887*10^-4***
	Seconds since start	2.088*10^-5***
Online context	Total number of tweets around (60 min/10 min)	-2.81*10^-6***
Hypothesis	Is general social media mention	-7.515*10^-2* (p=0.049)
	Is west coast showing	-3.432*10^-1*** (p<2*10^-16)
	Is general social media mention X Is season second half	-3.5*10^-2
	Is general social media mention X Is west coast showing	8.849*10^-2 (p=0.015)
	Adjusted R^2	0.339

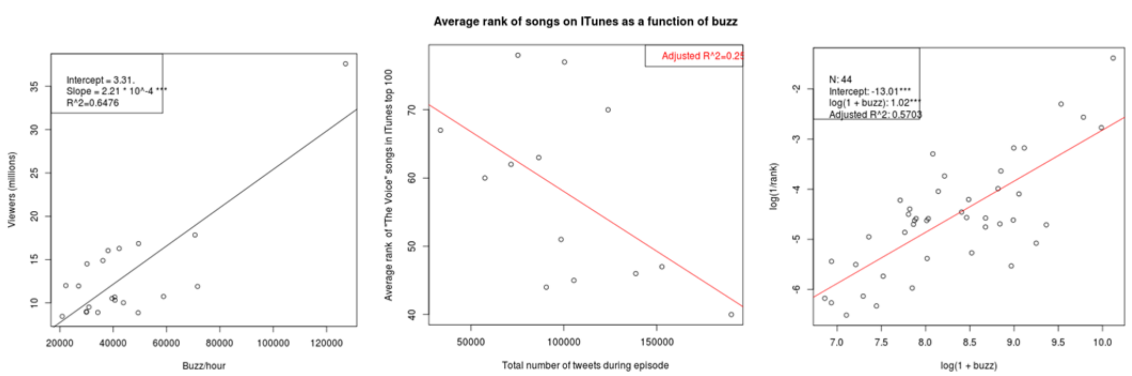

Figure 2. left Number of households viewing The Voice per episode as a function of buzz/hour during that episode. **center** Average sales rank of all contestants and total number of tweets during that episode. **right** 1/Sales rank for each contestant and their corresponding buzz for an episode

Limitations

While *The Voice* might be one of the most salient TV programs using the social TV strategy, we still plan to test our hypotheses on other programs. A possible concern is that some TV programs could be more engaging than others. We hope to control for this potential factor in our future studies. Users might have different levels of activity on Twitter at different times of the day/week. Thus far, we have not considered the impact of the overall level of activities on Twitter on our results. While we do believe that a large proportion of the bumps on Twitter result from social TV strategies, controlling the effect of the overall temporal dynamics on Twitter can increase the accuracy of our estimate. We have used shallow metrics of tweet content, by simply counting the occurrence of popular users, hashtags, and exclamation marks. This strategy can be a good start for understanding the effect on content on popularity, but we still need to analyze the content of the tweets more comprehensively to capture the underlying reasons for a high volume of buzz. Some useful features for describing tweet content are the sentiment of tweets and their emotional arousal capability, as analyzed by trained classifiers. Given that the dataset we analyze is observational, there are many events that are difficult to control. For example, it could be the case that #thevoice was only displayed after a very popular contestant performed. The fact that they performed could thus drive the increase in tweets. We did not notice such an effect, but it could nevertheless exist. Despite these limitations, we believe that this research provides the first evidence that social TV strategies do actually create significant engagement and generate sales.

Discussion, Conclusions, and Next Steps

In this research, we show three results related to TV show trigger events. The first is that displaying a tweet during a program will increase its retweet rate. Even controlling for the popularity of the tweeter, the content of the tweet also affects the expected number of retweets. Second, we show that displaying hashtags during a program seems to increase the number of program-related tweets, in this case by a relatively high proportion, 18.8%. This may be surprising because this hashtag is displayed very frequently, and one might expect that the viewers would become desensitized to it. We also show that specific tweets are more effective at sustaining engagement than general hashtags. Finally, we provide the first evidence that TV show content can be linked to sales for TV-show related items.

From these preliminary results, it seems as though displaying hashtags causes a tangible increase in viewers' Twitter activity overall; it also increases their engagement with the program during commercial breaks. Although we did not notice systematic placement of hashtags during the show, there could be a correlation between noteworthy content during the show and the hashtag display. In this case, much of the additional buzz that we observe could be attributed to the content of the episode. There are many other questions we would like to address to this dataset. For example, how does sentiment on Twitter relate to the Nielsen ratings of a television program? How does this Twitter activity vary over different regions in the U.S.?

References

Dumenco, S. 2011. "7 Things You Need to Know About 'Social Tv' Right Now the State of the Art -- and Science -- of Monitoring Social-Media Chatter About Tv." from http://adage.com/article/the-media-guy/social-tv/229845/

Harboe, G. 2009. "In Search of Social Television," in *Social Interactive Television: Immersive Shared Experiences and Perspectives,* D.G. P. Cesar, and K. Chorianopoulos (ed.). IGI Global, pp. 1-13.

Helweh, A. 2011. "Twitter, Time Shifting, Technology & Television. Social Media Explorer." from http://www.socialmediaexplorer.com/social-mediamarketing/twitter-time-shifting-technology-television/

Nathan, M., Harrison, C., Yarosh, S., Terveen, L., Stead, L., and Amento, B. 2008. "Collaboratv: Making Television Viewing Social Again," in: *Proceedings of the 1st international conference on Designing interactive user experiences for TV and video*. Silicon Valley, California, USA: ACM.

Proulx, M., and Shepatin, S. 2012. *Social Tv: How Marketers Can Reach and Engage Audiences by Connecting Television to the Web, Social Media, and Mobile*. Wiley.

Talbot, D. 2011. "A Social Media Decoder." from http://static.bluefinlabs.com/website/bluefin_mit-tech-review.pdf

Torrez-Riley, J. 2011. "The Social Tv Phenomenon: New Technologies Look to Enhance Television's Role as an Enabler of Social Interaction ". Elon, North Carolina: Elon University, p. 23.

Wohn, D., and Na, E.-K. 2011. "Tweeting About Tv: Sharing Television Viewing Experiences Via Social Media Message Streams," *First Monday [Online]* (16:3).