Patterns of Social Media Conversations Using Second Screens

Partha Mukherjee[1], Jian-Syuan Wong[2], Bernard J. Jansen[3]
College of Information Science and Engineering
Pennsylvania State University, University Park, PA, USA
{pom5109,jxw472}@ist.psu.edu[1,2], jjansen@acm.org[3]

Abstract

In this research, we analyze the pattern of conversation resulting from second screen interactions. Second screen refers to the phenomenon of people simultaneously engaged with more than one computer technologies. Specifically, we examine people's social media conversations while watching TV shows, both live and previously recorded. In our work, we analyze the user's social media conversation postings concerning three TV show, categorize the postings into five different classifications, and investigate the predominant categories for both live and previously recorded broadcasts of the TV shows. Our objective is to discern the conversation patterns within different aspects of the second screen conversations. The classifications are 1) *questions*, 2) *response*, 3) *referral*, 4) *broadcast* and 5) *retweet*. The user interactions in form of tweets are collected using Twitter as the second screen. We collect more than 418,000 tweets for three different TV programs. Using One Way Analysis of Variance, we examine the five tweet categories collected during live broadcast of the program and when the show is not aired. Findings imply that viewers post mainly personal opinion during live broadcasts, but they engage more in directing/redirecting information or recommendations with URLs when the show is not live. There are many implications for those interested in understanding social conversation around mass media in the emerging second screen environment.

Keywords: Second screen; Social media; TV shows, Twitter; ANOVA; Games-Howell test; Interaction pattern.

1. Introduction

The phenomenon of simultaneously engaging with more than one computer technology is referred to as second screen. When combined with social media, this phenomenon has the potential to be an important social soundtrack, especially as a mode of communication interactivity around TV shows, both live and previously recorded. The integration of Twitter (or other online social network) as the interactive medium with televised broadcasts marks the emergence of a new occurrence augmenting the social possibilities of TV or other mass communication [14]. This new usage phenomenon is an instantiation of second screen (e.g., TV and a computing device), although there may be multiple screens involved (TV and several computing devices). The second screen allows the social soundtrack to be a conversation with others regarding TV programing.

There has been some academic research concerning the second screen interaction, but analysis of the conversation patterns associated with TV shows is scarce. The advent of mobile technology and emergence of social media changes the TV viewing habit of the audience to more active from strictly passive and expands the social possibility of TV, as the merging of technologies now allow a number of social activities and conversation concerning TV content via social networks (e.g., Twitter, Facebook, Weibo, etc.). The second screen phenomenon has embedded itself within the modern TV culture and it acts as a social soundtrack for TV content with a variety of social implications for mass communication.
In this research, we investigate the characteristics of second screen interaction during the telecast of three popular U.S. TV programs, specifically examining the patterns of discussion that are present in users' second screen interactions around live and pre-recorded TV program. This research is important as fruitful analysis of the leading characteristics of users' social conversation can facilitate the personalization of TV content and advertising, along with implications for many other areas. Findings can assist both the channel owners and advertisers to formulate new strategies for TV airing, launching product ads to engage more viewers, promote sales, and earn revenues.

2. Related Work

There are previous studies on Twitter content classification framework to focus on macro–level public timeline at the expense of the richness of depth from individual histories. Java, Song, Finin and Tseng [11] examined miscellaneous tweets and presented four categories of content: a) daily chatter b) conversation, c) information sharing and d) reporting. Krishnamurthy, Gill and Arlitt [13] studied the social infrastructure by user classification based on follower/following counts, means for using the service and volume of posts. Dann [5] proposed a Twitter content classification framework as a tool for personal, professional, commercial and phatic communications happen in real world application based on grounded theory. Honeycutt and Herring [9] examined the tweets to find specific purposes of interlocution (i.e., '@' symbol) in directed communication and referencing. boyd, Golder, and Lotan [4] studied the conversational aspects of retweet and investigated the reasons of retweeting in Twitter, while Naaman, Boase and Lai [15] introduced an item list of broadcast statements including information sharing, personal opinion along with random thoughts and observations in an undirected manner [10].

ISBN: 978-1-62561-000-3

Regarding research of participation using second screens on content analysis of TV shows, Benton and Hill [2] investigated the resulting buzz of specific American reality show related tweets on the TV screen during the show. The content analysis of tweets during live telecast of a talk show indicated different forms of participations (i.e., audience and political) [8].

Though the aforementioned research talked about the analysis of TV show content by investigating tweets collected via second screen, the studies regarding finding significance of specific categories of second screen interaction (real time and non-real time) about TV shows are scarce. As such, there are several unanswered questions concerning the second screen interaction. What are the interaction points between TV and social media? What are the discussion patterns of second screen usage during live telecast? What are the discussion patterns of second screen usage after the live telecasts? These are some of the questions that motivate our present research.

3. Research Question

Our research question is: Is there any significant difference in patterns of social interaction among viewers regarding TV shows using second screen?

To investigate our research question we have segregated the tweets from three TV shows into five categories such as: 1) Question (Q), 2) Response (RS), 3) Referral (RF), 4) Retweet (RT) and 5) Broadcast (BC). We categories the queries based on the prior literature [4,9,15]. The effects of these five categories are evaluated on TV show based second screen interaction collected in form of tweets. Table 1 describes the communication patterns for the categories. We inquire the existence of such patterns as described in Table 1 in the tweets posted by viewers to classify the collected tweets into five categories. The effects of these five categories are evaluated on TV shows from second screen interactions collected in form of tweets.

As it is observed that conversation among the users in form of mentions ('@') increases after the show [17], we believe that the tweets belonging to the category of *Response* (RS) or *Referral* (RF) will result in more volume than other categories when the show is not televised. We believe that during live transmission of TV show, viewers tweet their momentary feeling in an undirected fashion and don't engage in reciprocation of messages, as it may divert their attention from the TV screen. Therefore, it leads us to assume that the undirected *broadcast* (BC) category will prevail during the live transmission of the TV shows. Based on the research question and the above assumptions, we form two research hypotheses to evaluate real time and non-real time interaction around TV shows.

Hypothesis 01: There is a significant difference in patterns of social interaction among viewers using second screens during live telecast of a TV show.

Table 1. Categories of second screen social interactions

Category	Description
Question (QN):	The tweets that uses @statement to address another user with questions '?'.
Referral (RF):	Any full length or shortened URL directed at another user. It does not contain any '?' symbol.
Response (RS):	Tweets intentionally engaging another user by means of '@' symbol which does not meet the other requirements of containing queries or referrals.
ReTweet (RT):	Any retweet as recognized by "'RT: @', 'retweeting @', 'retweet @', '(via @)', 'RT (via @)', 'thx @', 'HT @' or 'r @' ".
Broadcast (BC):	Undirected statements (i.e., does not contain any addressing) which allow for opinion, statements and random thoughts to be sent to the author's followers. Any undirected statement followed by questions '?' belongs to **Question (QN)** category instead of **Broadcast (BC)**.

Hypothesis 02: There is a significant difference in patterns of social interaction among viewers using second screens during a not live telecast of a TV show.

The underlying theoretical understanding of our research question is based on the social cognitive theory of mass communication [1] that analyzes the media influence on participants of the social network in terms of supporting potential diffusion of TV watching behavior across the virtual community.

4. Data Collection

We selected three popular TV shows from the U.S. and collected users' interactions in form of tweets from Twitter. The TV shows selected for this research are: 1) Dancing with the Stars, 2) Mad Men, and 3) True Blood. In order to increase the generalizability of our research, we collected data about TV shows that represent different genres. The tweets for Dancing with the Stars were collected for two consecutive weeks starting from 13th May to 25th May 2013. These two weeks account for selection of finalists and champion for season six respectively.

Regarding Mad Men and True Blood, we collected tweets for three successive weeks in the month of June. For both shows, it spans from 9th June to 29th June 2013. As 23rd June was the date for the season finale for Mad Men, we stopped collecting tweets for both Mad Men and True Blood the following week. For each show, the numbers of tweets collected in English texts are displayed in Table 2, where the queries are the TV show names. The number of tweets for

ISBN: 978-1-62561-000-3

Dancing with the Stars is less than that for other two TV shows as the version of Twitter API for Dancing with the Stars tweets was older (API 1.0) compared to that used (API 1.1) in tweet collection for other two shows.

The tweets are pulled into MySQL database by running three different PHP scripts each taking one TV show as the search query for twitter API. Once a set of tweets were collected and pulled into the database by the scripts, each script waits for 60 seconds before they become active again to search for new tweets. The tweets are stored in the database based on the unique tweet id (i.e., primary key of the tweet tables in the database).

Once the tweets were collected, we segregated the count of tweets collected in 24x7 hours across the weeks for all three TV shows into fifteen minutes intervals. The count of tweets during fifteen minute- time interval is considered as the unit of analysis. We annotated the timings of the tweets generated and categorize them as “real time second screen” (rtSS) (i.e., live) and “non-real time second screen” (nrtSS) tweets w.r.t Eastern Daylight Time (EDT). The annotation of tweet timings and categorization into rtSS and nrtSS groups is done manually. We monitor the show timings each week and the tweets that appear within show timings are marked as rtSS tweets. The nrtSS counterpart corresponds to that collected in rest of the days other than show timings. The rtSS tweets indicate that the tweets are posted during live broadcasts. The nrtSS counterparts are the ones posted by the users while the TV shows are not live. We need to focus on the tweets as rtSS tweets collected in hours shown in Table 3 combining the show timings of all six different US time zones (i.e., Eastern, Pacific, Central, Mountain, Alaska and Hawaii) considering the time differences w.r.t EDT. The airing time for all three TV shows is about 60 minutes each day except the week for champion selection for Dancing with the Stars. The airing time of Dancing with the Stars in final week is about two hours each day.

Table 2. Number of tweets collected for each TV show

Dancing With the Stars	Mad Men	True Blood
46,269	152,259	220,390

Table 3. Time in hour w.r.t EDT focusing collection of rtSS tweets per week for three TV shows

	Sun	Mon	Tue	Wed
Dancing with the Stars		8 PM, 9 PM, 11 PM	9 PM, 10 PM	12 AM
Mad Men	10 PM	1 AM		
True Blood	9 PM	12 AM		

5. Methodology

With the five categories of interaction patterns constructed, we import both the rtSS and nrtSS data into SPSS. The data contains the count of tweets for each of the five categories in fifteen minute time interval within 24x7 hours across the weeks for all three TV shows. The rtSS data is counts of tweets in fifteen minutes time interval for all five categories when the show is transmitted live across weeks. The nrtSS data is counts of tweets in fifteen minutes time interval for all five categories across weeks when the show is not in the air. The tweet counts in fifteen minute time interval for both rtSS and nrtSS data are considered the units of analysis in our research. As we use ANOVA procedure, the clumping of data within a specific incremental time interval is necessary. The choice of fifteen minute as the clumping interval is purely subjective.

In SPSS, we test our hypothesis using one way analysis of variance (ANOVA) procedure among five groups to test the differences between the means of both rtSS and nrtSS tweets (i.e., the average of the tweet count in fifteen minutes time interval) among the five categories. However, our data follows the power law distribution and hence is not multivariate normal. To perform ANOVA over five categories of rtSS and nrtSS tweets, we need to normalize the data by means of Box-Cox transformation [3]. We transform the data via the Box-Cox transformation using log transformation function log(variable + 1.0) before conducting the ANOVA test. The data was successfully normalized by means of log transformation.

6. Result

To test the hypotheses, we carry out one way ANOVA test over fifteen minute time interval counts of tweets across five categories for both rtSS and nrtSS interaction patterns. In one way ANOVA, the conversation pattern categories are used as the independent variable. ANOVA test identifies that means of the tweet counts in fifteen minute time interval of at least one category is significantly different from others. The critical value of the F-statistic is 2.214 at the 95% confidence interval.

Table 4. The result of ANOVA test over categories for rtSS tweets regarding TV shows

TV Show	F statistic	df	Sig.
Dancing with the stars	122.36	4	0.00
Mad Men	65.92	4	0.00
True Blood	323.99	4	0.00

We use Games–Howell test for post hoc analysis across the groups with unequal sizes as the assumption of homogeneity of variances is not satisfied (the significance level of Levene statistic should be greater than 0.05). The Games- Howell test takes both unequal variances and the unbalanced sample sizes into account by suggesting a critical difference between means, separately for every pair of means with Gaussian-q distribution [16]. The modification is derived from Tukey–

Table 5. T values between Broadcast and other categories when TV shows are not in the air

TV Show	QN	RS	RF	RT
Dancing with the stars	24.82*	4.61*	3.27*	2.97*
Mad Men	21.71*	6.54*	2.98*	4.31*
True Blood	38.45*	14.95*	8.18*	12.89*

*Denotes significance

-Kramer test and is recommended for sample sizes greater than five. The test is significantly more powerful than other tests in terms of confidence interval and rejection rates [6, 12]. In our conversation-pattern data, we observe that the larger group sizes have relatively smaller variances. We adopt the Games-Howell test as the most suitable method for post hoc analysis of the data with unequal group sizes and unequal variances where the sample size and sample variance are inversely paired. The Games-Howell modification always remains close to the level of significance and maintained control over Type-1 error under such a condition [12].

As the assumption of homogeneity of variances does not hold and the group sizes are unbalanced, we resort to Welch statistic to test the equality of group means assumption. We observe that our data follows the equality of means assumption (i.e. the value of Welch statistic was always < 0.05). The satisfaction of equality of means assumption is the precondition before carrying out Games–Howell test in post hoc analysis.

From the result of the post hoc analysis the t-tests are performed to find out the differences between categories. Since there are multiple chances to find a difference between the two groups (i.e., multiple tasks), the probabilities of getting at least one significant difference by chance were inflated. Some correction for that is needed. If the correction is not done then the risk that some of the repeated t tests would provide seemingly significant results just out of pure chance, may be increased.

To reduce such risk, we therefore introduce Bonferroni correction for the comparisons between conversation categories. Though traditional Bonferroni correction is a bit conservative and tends to lack power due to several reasons [7], the risk of getting inflated significant difference will be reduced. We are benefitted here from assuming that all tests are independent of each other. In our research as there are five categories of conversation patterns, the number of comparisons is 10. In our research the Bonferroni correction set the cutoff of significance level at 0.005 (i.e., the p value of significance is dropped).

6.1 Testing of Hypothesis 01

While testing hypothesis 01, the result of the ANOVA test for rtSS tweets shows that there is a significant difference of means of tweet counts between the communication pattern categories for three TV shows when the shows are broadcast live, as shown in Table 4. We observe that there is at least one category that is significantly different from other categories in terms of pattern of interaction. So, Hypothesis 01 is fully supported.

The Games–Howell test for pairwise comparison between the means of rtSS tweet counts in fifteen minute time intervals for five categories is reported in Table 5. It is seen from the magnitude of reported t-values that *Broadcast* (*BC*) category has a significant difference of means of tweet counts within fifteen minute time intervals over the rest four categories for all three TV shows when the TV show is in the air. The significance of the difference of means is measured w.r.t $\alpha = 0.005$ taking Bonferroni correction into account. This is because the viewers do not want to lose the attention from TV screen and hence avoid engaging in communication.

6.2 Testing of Hypothesis 02

While testing hypothesis 02, the result of the ANOVA test for nrtSS tweets shows that there is a significant difference of means of tweet counts between the communication pattern categories for three TV shows when the shows are broadcast live, as shown in Table 6. We observe that there is at least one category that is significantly different from other categories in terms of pattern of interaction. So, Hypothesis 02 is fully supported.

The Games–Howell test for pairwise comparison between the means of nrtSS tweet counts in fifteen minute time intervals for five categories is reported in Table 7. It is seen from the magnitude of reported t-values that Referral (RF) category becomes dominant in terms of difference of means of tweet counts within fifteen minute time intervals over the rest four categories for all three TV shows when the TV show is in not the air. The significance of the difference of means is measured w.r.t $\alpha = 0.005$ taking Bonferroni correction into account. This is because the viewers do not want to lose the attention from TV screen and hence avoid engaging in communication.

Table 6. The result of ANOVA test over categories for nrtSS tweets regarding TV shows

TV Show	F statistic	df	Sig.
Dancing with the stars	1680.43	4	0.00
Mad Men	4529.27	4	0.00
True Blood	6542.97	4	0.00

Table 7. T values between Referral and other categories when TV shows are not in the air

TV Show	QN	RS	RT	BC
Dancing with the stars	111.41*	35.12*	4.03*	35.15*
Mad Men	193.71*	58.56*	17.4*	20.73*
True Blood	323.41*	87.51*	77.69*	23.35*

*Denotes significance

7. Discussion and Implications

While investigating the effect of interaction pattern on Twitter based on TV programs, from Table 4 and Table 6, this research shows that there is significant difference in patterns of social interaction among viewers using second screens. Viewers prefer posting undirected messages (Broadcast) most on social media while the TV show is telecast live as observed in Table 5. Moreover from Table 7, the study finds out that the directed conversation with URLs (Referrals) will appear as the most significant category of communication when the TV shows are not transmitted.

The result implies that during live TV show users do not want to be distracted and intend to maintain their focus on TV show content. This leads to increased display of undirected opinions in the discussion forum. In pre-recorded interaction, users engage in responding to other viewers via directed communication or recommendations to other viewers using URLs. Regarding practical significance, analyzing sentiments of undirected communication and URL based directed recommendation will help cable providers and advertisers to identify the positive and negative effects of the televised shows and ads respectively, which results in better personalization of ads and TV shows.

8. Conclusion

The results regarding evaluating the significant second screen interaction pattern regarding TV shows in this research indicates that during live telecast the viewers are more inclined towards undirected messages while the directed communication with recommendation via URL seems most significant when the TV programs are not televised live. Access and evaluation of the sentiments of undirected broadcast and directed recommendations will benefit channel owners and retailers to personalize TV show and leveraging brand image by creating ad recommendation.

For future work, we will evaluate the significance of interaction patterns on larger amounts data collected over lengthier periods with a broad range of TV genres to reinforce the underlying theoretical framework [1]. We will carry out the sentiment analysis of both directed and undirected tweets with a view of improved personalization from the perspective of cable providers and retailers. Clumping of data in 15 minutes time interval may not detect the significant activity of commercials. So in future we will extend our research into detecting the commercial activity by mining the patterns of second screen interactions.

Acknowledgment

We thank Adan Ortiz Cordova and Sagnik Ray Choudhury for their assistance with Twitter data collection.

Reference

[1] A. Bandura, "Social cognitive theory of mass communication," Media effects: Advances in theory and research, 2002, Vol.2, pp. 121-153.

[2] A. Benton and S. Hill, "The Spoiler Effect?: Designing Social TV Content That Promotes Ongoing WOM", in Informs conference on Information Systems and Technology, 2012.

[3] G.E. Box and D.R. Cox, "An analysis of transformations," Journal of the Royal Statistical Society, Series B, 1964. Vol. 26(2), pp. 211-252.

[4] D. Boyd et.al, "Tweet, tweet, retweet: Conversational aspects of retweeting on twitter," in 43rd IEEE Hawaii International Conference on System Sciences, 2010, pp. 1-10.

[5] S. Dann, Twitter content classification. First Monday, 2010, Vol. 5(12).

[6] J.E. De Muth, Basic statistics and pharmaceutical statistical applications, second edition, CRC Press, 2006.

[7] O.J. Dunn, "Multiple comparisons among means," Journal of the American Statistical Association, 1961. 56(293): pp. 52-64.

[8] F. Giglietto and D. Selva, "Second Screen and Participation: A Content Analysis of a Full Season Dataset of Tweets," Available at Social Science Research Network, 2013, pp. 1-24.

[9] C. Honeycutt and S.C. Herring, "Beyond microblogging: Conversation and collaboration via Twitter," in 42nd IEEE Hawaii International Conference on System Sciences, 2009, pp. 1-10.

[10] B.J. Jansen et. al., "Twitter power: Tweets as electronic word of mouth," Journal of the American society for information science and technology, 2009, Vol. 60(11), pp. 2169-2188.

[11] A. Java et. al., "Why we twitter: understanding microblogging usage and communities," in ACM workshop on Web mining and social network analysis. 2007, pp. 56-65.

[12] H.J. Keselman and J.C. Rogan, "A comparison of the modified-Tukey and Scheffe methods of multiple comparisons for pairwise contrasts," Journal of the American Statistical Association, 1978. Vol. 73(361), pp. 47-52.

[13] B. Krishnamurthy et. al., "A few chirps about twitter," in ACM workshop on Online social networks. 2008, pp. 19-24.

[14] G.D.F. Morales and A. Shekhawat, "The Future of Second Screen Experience," in ACM conference on Computer-Human Interaction (Extended Abstracts). 2013, pp. 1-10.

[15] M. Naaman et. al., "Is it really about me?: message content in social awareness streams," in ACM conference on Computer supported cooperative work. 2010, pp. 189-192.

[16] J.W. Osborne, Best practices in quantitative methods, Sage publications, 2008.

[17] D.A. Shamma et. al., "Tweet the debates: understanding community annotation of uncollected sources," in ACM workshop on Social media. 2009, pp. 3-10.

 ISBN: 978-1-62561-000-3