Evaluating Pattern for Group Interactions using Second Screens

Partha Mukherjee
College of Information Science and Technology
Pennsylvania State University
University Park, PA, USA
pom5109@ist.psu.edu

Bernard J. Jansen
College of Information Science and Technology
Pennsylvania State University
University Park, PA, USA
jjansen@acm.org

Abstract—**In our research, we analyze the group conversation in Twitter via. second screen concerning two popular US based TV shows, categorize the postings into four different classifications, and investigate the predominant categories for group interactions. The group conversations are identified by hashtags present in tweets where the number of members in the group interacting is at least two. The classifications are 1) *response*, 2) *referral*, 3) *broadcast* and 4) *retweet*. We collect nearly 354,000 tweets for these two TV programs. Using one-way analysis of variance, we examine the four tweet categories collected during second screen based group interactions. Our findings indicate that most group-based interaction occurs when the TV show is not being transmitted. The prevalent conversational pattern observed is *referral* tweets. There are many implications for those interested in understanding social conversation around mass media in the emerging second screen environment.**

Keywords— ANOVA; Games-Howell test, hypothesis testing; group interaction; second screen; social TV

I. Introduction

The phenomenon of simultaneously engaging with more than one computer technology is referred to as second screen. When combined with social network and social media (e.g., Facebook, Twitter, Weibo etc.), this phenomenon has the potential to be an important social soundtrack, especially as a mode of communication interactivity around TV shows.

The viewers do exchange information related to the TV show via second screen in terms of posting of tweets [1]. The exchange of information can happen live during show time or when the show is not transmitted live. The information in the social interaction may relate to different aspects of the show content (e.g., the actors, directors, costumes, characters, themes, etc.) or aspects of the products advertised (e.g., brand, sale, customer preference etc.). These attributes surrounding to a TV show in the discussion are identified by means of the hashtags (#) before a relevant keyword or phrase in the tweets to categorize the tweets as part of the popular topics marked by those keywords or phrases. Audiences reciprocating the tweet mentioning a particular hashtag to the commentary form a group engaging in a conversation. The group size is determined by the number of members interact within that group. The greater the interaction within a group, the more informed group it will be. Greater interactions within groups promote diffusion of more information about the aspects concerning a TV show.

In this research, we investigate the characteristics of second screen interaction during the telecast of two popular U.S. TV programs, specifically examining the patterns of discussion that are present in second screen interactions among groups of viewers around the TV shows. This research is important as fruitful analysis of the leading characteristics of viewers' group based 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.

II. Related Work

A. Content classification of social network

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 [2] examined miscellaneous tweets and presented four categories of content: a) daily chatter b) conversation, c) information sharing and d) reporting. Krishnamurthy, Gill and Arlitt [3] studied the social infrastructure by user classification based on follower/following counts, means for using the service and volume of posts. Dann [4] 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 [5] examined the tweets to find specific purposes of interlocution (i.e., '@' symbol) in directed communication and referencing. boyd, Golder, and Lotan [6] studied the conversational aspects of retweet and investigated the reasons of re-tweeting in Twitter, while Naaman, Boase and Lai [7] introduced an item list of broadcast statements including information sharing, personal opinion along with random thoughts and observations in an undirected manner [8].

978-1-4799-6959-3/15/$31.00 ©2015 IEEE

TABLE 1

CATEGORIES OF GROUP BASED SOCIAL INTERACTIONS

Category	Description
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 Broadcast (BC) category.

B. Integration of prior research

Regarding research of participation using second screens on content analysis of TV shows, Benton and Hill [9] 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) [10].

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 group interactions? Whether group conversations occurs when the show is transmitting live? These are some of the questions that motivate our present research.

III. RESEARCH QUESTIONS

In our research, we identify groups by means of hashtags (#) that can be present anywhere in the comments posted by the TV viewers in Twitter tweets. The collection of viewers with communications messages sharing the same hashtag forms a group. The groups may be small or large depending on the number of interactants within the groups. The probability of information sharing is higher if there are more members within a group. The theoretical understanding of our research is characterized by social cognitive theory of mass communication [11] that analyzes the media influence on participants of the social network where the structured interconnectedness in terms of proximity of interactive mediums (text, video, etc.) support the potential diffusion of TV watching behavior across the virtual community. In our study, we express the information shared within a group as a function of social and temporal features. The social features deal with the viewers' patterns of conversation within groups and group size. The temporal features identify the volume of real time or non-real time communication. Information sharing in active discussion via second screen by forming groups among viewers lead to our first research question:-

1) Are there significant differences in patterns of social interaction among viewers engage within group conversation regarding TV shows using second screen?

To investigate our research question, we segregated tweets from three TV shows into four categories based on prior research [5, 6, 7] on patterns of social media interactions: Referral (RF), Retweet (RT), Broadcast (BC) and Response (RS). 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 group members to classify the collected tweets into four categories. Based on our first research question, we can form our first research hypothesis:

Hypothesis 01: There is a significant difference in social interaction in terms of communication patterns used in second screen based group conversation about TV shows.

The first research question analyzes the media influence on participants of the social network in terms of supporting potential diffusion of TV watching behavior within groups across the virtual community.

The next research question examines the prevalent temporal feature in interactions within a group. We categorize the group interaction as real time and non-real time interaction respectively based on whether the TV show is being televised or the show is not in the air. To identify whether group conversations are primarily real time or non-real time is important as it may create commercial opportunities by tracking the information at the intersection of the social networks, the TV, and the second screen.

2) Is there any significant difference of secondary screen usage in group conversation based on transmission of the TV shows?

This research question is evaluated from second screen interactions collected in form of tweets about a TV show. As it is observed that conversation among the users in form of mentions ('@') increases after the show [12], we believe that conversation in groups will occur more when the show is not in the air. It is our intuition that viewers will engage in group communication once the show is over, as they may not want to be distracted from TV screen when the show is televised live. Based on research question 2, we form the following hypothesis:

Hypothesis 02: There is a significant difference between real time and non-real time social interaction among viewers engage in group conversation based on a TV show w.r.t usage of secondary screen.

Research question 1 and research question 2 refer to the possible development of the ecosystem of the service providers to serve the shift in user's viewing behavior and the emerging business opportunities.

IV. Data Collection and Research Design

We select "*Game of Thrones*" and *"Arrested Development"* as the TV shows to evaluate our research questions and collect audience interactions in form of tweets from Twitter. *Game of Thrones* is a popular TV show that is broadcast by channel HBO while *Arrested Development* is broadcasted by FOX television channel. The tweet collection is continued with a span from 5th February 2014 to 4th March 2014 and we amassed a total of 353,547 tweets. We use PHP scripts to access the Twitter API to collect data concerning the TV shows by using the TV show name as the Twitter API query. The tweets retrieved as JSON objects are pulled into a MySQL database. As Twitter is one of the most popular micro-blogging sites, we use it as the platform for the second screen based social interaction. Most micro-blogging services share commonalities such as: a) short text messages, b) instantaneous message delivery, and c) subscription to message updates [8]. So, although we use Twitter as the platform of interaction, we believe that our findings are applicable to other micro-blogging applications.

Once the tweets were collected, we retrieve those tweets filtered by the presence of hashtags (#). In our research, we identify the potential groups by means of these hashtags. In an informed group, there should be at least two or more users interacting with each other and mentioning the same hashtag within tweets. A user can be a member of multiple groups. From our data collection, 4973 informed groups were identified for *"Game of Thrones"* where the range of unique group members is between 2 to 7435. For *"Arrested Development"* there are 710 groups with the range of unique group members lies between 2 to 216.

A. Social features

In this study, we extract the social features in terms of count of tweets corresponding to (a) pattern of viewers' conversation within groups, (b) number of unique words present in the tweets, and (c) number of members in a group. The identifiers for categories of patterns for group conversation are described by Table 1.

B. Temporal features

We extract temporal features from the perspective of show timings of the episodes. The episodes of the TV show are broadcasted at 9:00 PM EST every Sunday with duration of one hour, including commercials. The broadcast time for aired episodes of *Arrested Development* is 9:00 PM EST every Tuesday for duration of 40 minutes (with commercials).There are two such temporal features of the group conversations for a particular informed group: 1) Real-time second screen tweet (rtSS) count: number of interactions recorded when the show is being aired, and 2) Non-real time second screen tweet (nrtSS) count: number of interactions recorded when the show is not televised live.

V. Methodology

To investigate our first two research questions, we import the data into SPSS. In SPSS, we test our hypothesis 01 using one-way analysis of variance (ANOVA) procedure among four categories of communication pattern to test the differences between the means of tweets posted as group interaction. Here the unit of analysis is the tweet count across all four categories of conversation patterns over 4973 and 710 groups for *"Game of Thrones"* and *"Arrested Development"* respectively.

To test hypothesis 02, we resort to statistical t-test to evaluate the difference between means of real time and non-real time tweet counts during group interaction. However, our data follows the power law distribution and hence is not multivariate normal. To perform ANOVA over four categories of tweets, we need to normalize the data by means of Box-Cox transformation [13]. We transform the data via the Box-Cox transformation using log transformation function *log (variable + 2.0)* before conducting the ANOVA test. The similar Box-Cox transformations over real time and non-real time counts are carried out to perform t-test. The Box-Cox transformation of the non-normalized data increases the efficiency of univariate t-tests [14]. The data is successfully normalized by means of log transformation.

VI. Results

We examine the patterns of virtual communication viewers use to engage via second screen as group members. In addition to that, the significance of temporal features of group interaction is investigated. The study of research question 1 and research question 2 evaluates hypothesis 01 and hypothesis 02 respectively.

A. Testing of Hypothesis 01

Hypothesis 01: There is a significant difference in patterns of social interaction among viewers engage within group conversation regarding TV show using second screen.

Evaluating hypothesis 01, we use one-way ANOVA to test the differences between means of conversation patterns for all informed groups. The conversation pattern categories are used as the independent variable. ANOVA test identifies that mean of log transformed tweet counts over 4973 and 710 groups (i.e., the respective shows) for at least one pattern category is significantly different from that of others. The critical value of the F-statistic is 2.214 at the 95% confidence interval. While evaluating hypothesis 01 using ANOVA, it is found that $F(3) = 110 > 2.214$ with p-value = 0.00 for "Game of Thrones". For "Arrested Development", it is: $F(3) = 74.681 > 2.214$ with p-value 0.0 The result shows that there is at least one category that is significantly different from other categories in terms of communication pattern of group interaction. So, hypothesis 01 is fully supported.

We use Games–Howell test for post hoc analysis across the groups with unequal sizes as the assumption of homogeneity of variances is violated (the significance level of Levene statistic should be greater than 0.05). As Games-Howell test takes both unequal variances and the unbalanced sample sizes into account by suggesting a critical difference between means, we adopt the test as the most suitable method for post hoc analysis of our data with unequal group sizes and unequal variances.

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 means assumption. We observe that our data follows the equality of means assumption (i.e. the value of Welch statistic was always < 0.05).

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, otherwise the risk of providing seemingly significant results just out of pure chance by t tests, may be increased.

To reduce such risk, we therefore introduce Bonferroni correction for the comparisons between conversation categories. We were benefitted here from assuming that all tests are independent of each other. In our research, as there are four categories of conversation patterns, the number of comparisons is six. So the Bonferroni correction set the cutoff of our present significance level (α) at 0.008 (i.e., 0.05 / 6).

The Games–Howell tests for pairwise comparison between the means of tweet counts across the informed groups for four categories are reported in Table 2 and Table T3 for "Game of Thrones" and "Arrested Development" respectively. It is seen from the reported t-values that Referral (RF) category becomes dominant in terms of difference of means of tweet counts across the 4,973 and 710 groups over the remaining three categories for the selected TV shows. The significance of the difference of means is measured w.r.t α = 0.008 taking Bonferroni correction into account.

TABLE 2

T VALUES ('*' DENOTES SIGNIFICANCE) BETWEEN FOUR GROUP CONVERSATION PATTERN CATEGORIES FOR GAME OF THRONES

	RS	RF	RT	BC
RS		-17.81*	-12.75*	-10.43*
RF	17.81*		5.34*	7.63*
RT	12.75*	-5.34*		2.19
BC	10.43*	-7.63*	-2.19	

B. Testing of Hypothesis 02

Hypothesis 02: There is a significant difference between real time and non-real time social interaction among viewers engage in group conversation based on a TV show w.r.t usage of secondary screen.

To evaluate hypothesis 02, we perform two tailed t test at 95% confidence interval over two temporal features of the TV show tweets collected for 4973 and 710 groups. Here, we are investigating the difference between means of log transformed real-time (rtSS) and non-real time tweet (nrtSS) counts collected for all informed groups. The t statistic is 1.96 at significance level α = 0.05 for large degrees of freedom.

TABLE 3

T VALUES ('*' DENOTES SIGNIFICANCE) BETWEEN FOUR GROUP CONVERSATION PATTERN CATEGORIES FOR ARRESTED DEVELOPMENT

	RS	RF	RT	BC
RS		-13.69*	-2.65	-2.03
RF	13.69*		11.01*	10.68*
RT	2.65	-11.01*		0.282
BC	2.03	-10.68*	-0.282	

The two-tailed t test result indicates that there is a significant difference between the rtSS and nrtSS temporal categories of tweets for the selected TV show. For *"Game of Thrones"* the magnitude of t statistic is 81.49 at large degrees of freedom (i.e., p-value = 0.00) while t-value for *"Arrested Development"* is 39.24 with p-value = 0.00. So, hypothesis 02 is fully supported.

The statistical significance of hypothesis 02 led us to explore the direction of the second screen effect on group interaction. For one tailed two sample t test, we have $t_{critical}$ = 1.645 at α = 0.05. The one tailed t-test for the selected TV show infer that the mean of non-real time tweet counts across 4,973 and 710 informed groups are greater than that of real time interactions as t = 81.49 > 1.645 (*"Game of Thrones"*) and t = 39.24 > 1.645 (*"Arrested Development"*) with p-values for both the cases = 0.00. So, we conclude that information exchange within groups focuses mainly on virtual communication when the show is not aired.

VII. Discussion and Implications

While evaluating the research hypotheses related to research question 1 and research question 2, we notice that audience make URL based recommendations while interacting in groups. The emergence of referral (RF) as the predominant pattern satisfies hypothesis 01. The assessment of hypothesis 02 identifies that viewers usually form groups to initiate second screen interaction using hashtags when the TV shows are not televised. They don't engage within group communication when the shows are being aired.

A. Theoretical Implication

Concerning the theoretical implications of the findings, The observed change in pattern in group communication is characterized by social cognitive theory of mass communication [11]. From the user's point of view, the increased recommendation based second screen interaction in groups when the TV show is over leads to a change in viewing habits of the audience in terms of interaction and information sharing via URLs (viewers of TVs using second screen are also technology users). The finding sheds greater insight on the growing phenomenon of second screen interactions.

B. Practical Implication

We observe from Table 2 and Table 3 that within group communication using second screens; directed conversation with URLs and recommendation appears to be the dominant modes of interaction. Recommendation as a form of interaction may influence other viewers to interact with the brand enhancing word-of-mouth advertising if a brand is advertised on a TV show. From a retailers' point of view, the information about the brand might be diffused to the viewer's larger social community via recommendation-based interaction. Broadcasters may reorganize the theme or improvise the aspects of the show by analyzing the reactions corroborated by viewers' recommendations. Both broadcasters and retailers should monitor the group communication when the show is not in the air as the group interaction is mainly non real time.

VIII. CONCLUSION

In our research, we explore the prevalent pattern of social media communication related to two TV shows in groups of viewers that are formed by hashtags. We observe that the groups are mainly created when TV shows are not in the air. It indicates that viewers use to spend time by engaging in directed conversation with recommendations when the shows are not being aired. This in turn points out that users may not want to lose the focus from the TV screen when the show is being televised.

In our research, we don't evaluate the primary second screen for group interaction in terms of device usage. We will evaluate the effect of secondary screen usage during group based information exchange in future research. In present study the TV shows considered for research belong to a single cultural bias (US). We will extend our research over several TV shows to further investigate our findings for different cultural biases.

REFERENCES

[1] P. Mukherjee and B. J. Jansen. (2014). Social TV and the social soundtrack: significance of second screen interaction during television viewing. Presented at SBP, Springer, pp. 317-324

[2] A. Java, X. Song, T. Finin and B. Tseng (2007). Why we twitter: understanding microblogging usage and communities. Presented at SNA-KDD workshop on Web mining and social network analysis, ACM, pp. 56-65.

[3] B. Krishnamurthy, P. Gill and M. Arlitt. (2008). A few chirps about twitter. Presented at of the first workshop on Online social networks, pp. 19-24

[4] S. Dann. (2010). Twitter content classification. *First Monday*, *15(12)*.

[5] C. Honeycutt and S. C. Herring. (2009). Beyond microblogging: Conversation and collaboration via Twitter. Presented at HICSS'09, pp. 1-10.

[6] D. boyd, S. Golder and G. Lotan. (2010). Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. Presented at HICSS'10, pp. 1-10.

[7] M. Naaman, J. Boase and C. H. Lai. (2010). Is it really about me?: message content in social awareness streams, Presneted at ACM CSCW, pp. 189-192.

[8] B. J. Jansen, M. Zhang, K. Sobel and A. Chowdury. (2009). Twitter power: Tweets as electronic word of mouth. *Journal of the American Society for Information Science and Technology, 60(11)*, pp. 2169-2188.

[9] A. Benton and S. Hill. (2012). The Spoiler Effect?: Designing Social TV Content That Promotes Ongoing WOM, Presented at INFORMS on Information Systems and Technology.

[10] F. Giglietto and D. Selva. (2013). Second Screen and Participation: A Content Analysis of a Full Season Dataset of Tweets. Available at SSRN, pp. 1-24.

[11] A. Bandura. (2002). Social cognitive theory of mass communication. *Media effects: Advances in theory and research, 2*, pp. 121-153.

[12] D. A. Shamma, L. Kennedy and E. F. Churchill. (2009). Tweet the debates: understanding community annotation of uncollected sources. Presented at SIGMM workshop on Social media, pp. 3-10.

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

[14] J. Freeman and R. Modarres. (2006). Efficiency of t-Test and Hotelling's T 2-Test After Box-Cox Transformation. *Communications in Statistics—Theory and Methods, 35(6)*, pp. 1109-1122.