

# Organization or conversation in Twitter: A case study of chatterboxing

**Heather Lea Moulaison**

University of Missouri  
303 Townsend Hall  
Columbia, MO 65211  
+1 573.882-8323  
moulaisonhe@missouri.edu

**C. Sean Burns**

University of Missouri  
111 London Hall  
Columbia, MO 65211  
+1 210 704 7326  
csbc74@mail.missouri.edu

## ABSTRACT

This paper reports on a case study of Twitter posts (tweets) by chatterboxers to study whether theories of organization of information are applicable to the study of user-supplied labels in Twitter. Chatterboxing is the act of watching a televised event such as the Super Bowl and using a second screen to engage with others, primarily in real time. Researchers have used communication theory as a framework for study of Twitter, considering both #hashtags and @mentions to be primarily communicative. To ascertain whether #hashtags may be fundamentally different and amenable to study as organizational conventions as well, we first compared differences between usage of #hashtags and @mentions during the Super Bowl by taking tweets from three locations identified as heavily invested in the event (hometowns of the teams and the location of the game: Boston, NYC, Indianapolis) and tweets from locations that were not invested (Dallas, Miami, Seattle). Non-parametric statistical comparisons were made between tweets from the three invested and non-invested groups to ascertain whether the uses of labeling conventions were identical. Next a qualitative analysis of a subset of non-location specific tweets supplied information about the content of tweets, the *aboutness* of #hashtags, and the placement of #hashtags in the tweets. Our findings indicate that #hashtags and @mentions do have two separate functions but that location has a positive influence on their statistical dependency. We also find that #hashtags are used as organizational mechanisms and can reflect *aboutness*. Specifically, #hashtags are used to describe in order to categorize and to retrieve in order to follow or join a conversation, and future studies should be able to use theories of organization of information to analyze these labels as a way of complementing their otherwise communicative nature.

## Keywords

Twitter, hashtags, mentions, *aboutness*, organization,

conversation, real-time communication, chatterboxing, geographic location

## INTRODUCTION

This study seeks to answer the fundamental question: is it possible to say that Twitter #hashtags are organizational in nature or are they purely conversational? To explore this question, this case study uses chatterboxing during the Super Bowl to identify quantitative differences between #hashtags and @mentions. In a second step, a qualitative analysis of a sub-section of tweets identified topics of tweets and characteristics of #hashtags as labeling conventions in Twitter. Analysis reveals that the two are different. Qualitative investigations lead us to believe that #hashtags are organizational and therefore in the domain of study of Library and Information Science (LIS).

Uses of Twitter have been evolving since its introduction (Huang, Thornton, Efthimiadis, 2010, p. 5). #Hashtags have emerged as a primary labeling convention, but researchers are divided as to the character of that convention. One way to test the uses of labeling conventions in Twitter is through a shared-context environment like chatterboxing. Chatterboxing is the discussion about a real-time televised event through the use of a second screen connected to a social media outlet such as Twitter or Facebook or to other electronic means of communication such as email (TV Licensing, 2012). Labeling posts correctly in Twitter is essential for chatterboxers if they want to participate in the conversation. We find that differences in labeling conventions confirm different approaches to organization and conversation in Twitter among chatterboxers.

The contributions of this paper include the first large-scale study of chatterboxing thereby expanding the literature on real-time social uses of Twitter. This paper is also a systematic examination of the function of #hashtags in tweets by a sophisticated user group focusing on a shared topic of interest in a geographic area. The effect of geography on use is also analyzed as part of the study, and further qualitative analysis of tweets enriches the literature on real-time uses of social media.

## REVIEW OF THE LITERATURE

### Twitter as a Social Medium for Communication

Twitter is a microblogging platform allowing users to post short, 140-character messages. Users are able to follow other users' accounts and to follow the Twitter stream of posts from all users concurrently and in real-time. Tweets by users in geographic proximity may also be viewed. Users generally connect to their accounts on Twitter through web browsers on computers and tablets and through apps on smartphones and tablets. Twitter's API has been made available to the developer community allowing for the creation of other interfaces to interact with Twitter and its contents. Some examples of web-based third-party access are through HootSuite (hootsuite.com) and TweetDeck (tweetdeck.com) for which a desktop client also exists. Smartphone applications for the two services already mentioned along with Twitterrific (<http://twitterrific.com/>), and others also exist.

### Labeling Conventions within Tweets

Twitter is a noisy medium (Jackoway, Samet, & Sankaranarayanan, 2011; Honeycutt & Herring, 2009), with a broad range of topics being tweeted currently at any given time. The noisiness of the platform creates difficulty for users both in following conversations and in being heard. Early use of Twitter by a user-base already familiar with technology such as the ones in San Francisco and those at South by Southwest (Levy, 2007) paved the way for importing labeling conventions already in use on Internet Relay Chat (IRC) channels such as the use of the *at sign* (Honeycutt & Herring, 2009). The three main conventions for labeling tweets include the use of the *hash mark* or pound sign (#), the use of the *at sign* (@), and the use of the characters "RT" to indicate that a post has been retweeted.

**#Hashtags.** #Hashtags in Twitter are short labels with a hash mark or pound sign preceding the characters of the text string. #Hashtags either serve as a label like social tags in sites like Flickr, or they function as a prompt for comment (Huang, Thornton, & Efthimiadis, 2010, p. 3). Because there are no official Twitter #hashtags nor is there a #hashtag registry, users are free to create and use #hashtags. Not all #hashtags have meaning or are meant to be interpreted seriously (Weng, Lim, He, & Leung, 2010). Yang and Leskovec (2010) found that Twitter users with very high numbers of followers were least influential in the social realm of #hashtag adoption by followers, though they were good at informing these followers by sharing news. Users with an intermediate number of followers have a higher degree of influence (p. 9). Uses of #hashtags can overlap between users and groups, and community members may not all use the same #hashtag to label their posts. Organizers of groups such as conferences or events can decide on and announce a recommended #hashtag; the consistent use of the suggested #hashtag allows conference-goers and non-attendees alike to follow and interact with

the conversation surrounding the event. #Hashtags may not be unique, and disambiguating threads posted by different user groups may prove to be challenging for users.

In the larger context of Twitter, #hashtag use populates lists of trending topics featured on the Twitter website and in many third party software (Huang, Thornton, & Efthimiadis, 2010), although not all trending topics are #hashtagged; the authors have the impression that some are paid advertisements. The adoption of new #hashtags on Twitter was investigated by Huang, Thornton, and Efthimiadis (2010) who, in comparing their use to tags in Delicious, found #hashtags to be fundamentally different because they can prompt conversation.

**@Mentions.** @Mentions on Twitter add the *at sign* immediately preceding a user's Twitter handle (user name). The use of @mentions provides coherence in an otherwise seemingly disorganized system where posts are displayed in chronological order (Honeycutt & Herring, 2009). In their 2009 study, Honeycutt and Herring found that roughly 30% of all tweets in all languages included @mentions, and they note that this percentage is double the reported rate from 2006 (p. 8).

**RT Retweets.** A tweet that is retweeted (reposted) from another user conventionally begins with the letters *RT* followed by the @mention of the original author's handle and the original message, which may be edited for space.

### Conversation and Organization on Twitter

When attempting to decide on a theoretical framework to use in the study of #hashtags, it became apparent that conflicting schools of thought exist as to the nature of #hashtags and their function in tweets. Conventions for labeling in tweets can be seen either as conversation or of organization. For the purpose of this study, conversation implies information sharing between and among interlocutors while organization implies curation, analysis, and assumed retrieval. Weng, Lim, He, and Leung (2010) see #hashtags as keywords or phrases that may or may not be interesting or adopted (p. 1121). Because of the similarity of the mechanics to social tagging, the #hashtag convention can be considered an organizing mechanism. Huang, Thornton, and Efthimiadis (2010) raised the question of conversation versus organization in relation to #hashtags, concluding that *a priori* #hashtags are not organizational in nature because users are not indexing for retrieval at a later date. They feel that the goal of #hashtag use is different from *a posteriori* tagging in social media sites (p. 1); instead, Twitter users use #hashtags to join and participate in a discussion (p. 5). Conversely, Efron (2010 & 2011) assumes #hashtags to be organizational, lending themselves to the study of retrieval of tweets.

The study of @mentions has been less fraught. Honeycutt and Herring (2009) interpret any use of @mentions as promoting conversation and potentially collaboration.

Within an hour of being tweeted, 31.2% of tweets with @mentions received a public response indicating that mentions promote conversation (Honeycutt & Herring, 2009). Retweets, as studied by boyd, Golder, and Lotan (2010), are also considered to be conversational. If @mentions and #hashtags are both conversational as some researchers suggest, then we theorize they will be used in a way that is consistent within tweets.

### **#Hashtags as *Aboutness*, or Tagging Talk: Theoretical Considerations**

Although, as noted, there is some belief that #hashtags are less an organizational device and more a communication device, when #hashtags are compared to the kind of tagging done on social bookmarking sites, the #hashtag as primarily an organizational device seems to offer a better explanation of their use in tweets. For example, #hashtags apparently describe or interpret the content of a tweet and in this performance, they express *aboutness*, as indexers understand the term (Maron, 1977) and as folksonomies might be understood (Peters and Stock, 2007). They also have the ability to index a conversation based on their content, and therefore they function as pointers to these conversations. Furthermore, #hashtags seem to share the same benefits and problems as Peters and Stock (2007) note that folksonomies do, where benefits include a type of low cost indexing, identification of communities, and so forth and while problems include an uncontrolled vocabulary, language merging, and so forth.

However, unlike professional index terms and folksonomies, #hashtag use is far more transient since tweets quickly come and go. Thus, unlike folksonomies, such as on social bookmarking sites (Kipp, 2010), and controlled index terms, as used in bibliographic description (Saracevic, 2007), conceptualizing them as expressions of *aboutness* and as having an indexing function is much more constrained by their temporality. This mostly likely explains why #hashtags are not, at present, used in any formal retrieval system, although, as Huang, Thornton, and Efthimiadis (2010) indicate, the acquisition of the Twitter archives by the Library of Congress implies that tweets will at a future point be searchable and that user behavior may change. Further, as natural language terms not unlike like tags in social media sites, #hashtags cannot suffer from an "indexer effect" (De Bellis, 2009). Overall, what separates #hashtags from professional indexing is the lack of control, and what separates them from tagging is their impermanence, given that older tweets are not easily retrievable.

It is true there is a sense that #hashtags have a stronger communicative function than folksonomies have, but this seems to be related to what the #hashtag describes --- the tweet, which itself must be taken into consideration as something that is part of a real time conversation, and to the purpose of tweeting. Thus, the communicative nature of a

tweet as the object of description differs from other objects of description where tagging is used, such as social bookmarking sites and image sites (Ransom and Rafferty, 2011). Furthermore, just as the nature of tags work differently for textual items and non-textual items (Peters and Stock, 2007), they also seem to work differently depending on the first order purpose of the item, such as whether the item is a fixed blog post or a fast moving tweet. This should not be surprising, as Hjørland (2001) notes, *aboutness* may very well depend on the role it performs, and any theory should therefore take this into consideration. It follows that since #hashtags possess a strong communicative function, any theory of #hashtags as *aboutness* should probably incorporate a relevant communication theory

### **Mobility and Geographic Location**

Tweets can emanate from users with apps installed on mobile devices or tablets; they can be posted through the Twitter web interface on a web browser; or they can be made through a third-party client installed on a user's personal computer. We theorize that users coordinating events in real-time will direct their tweets to particular users through the @mention and that users describing and organizing their own personal tweets will tend to use #hashtags. Location-based aspects of Twitter can contribute to the richness of the data (Jackoway, Samet, & Sankaranarayanan, 2011). It is possible to know who in proximity is tweeting and to see the content of those tweets. Location and mobility contribute to the overall quality of the conversation on Twitter.

### **Chatterboxing**

Chatterboxing, a term coined in the TV Licensing report (2012), is the participation in a social media platform discussion about a television show (the *box*) in real-time and is identified as a "widespread phenomenon" in the UK (TV Licensing, 2012, p. 10). Although chatterboxing may be a new term, dividing one's attention by engaging in emerging communication technologies and attending to another event is not new. For example, Mueller (2009) discusses the classroom complications of a "digital underlife" as it plays out in a type of digital "backchannel." Marwick and boyd (2010) write about how content producers imagine their audiences when they use social media technologies such as Twitter. (2010).

Research about engaging in social computing while consuming "traditional media" is growing. Much of this research is motivated by the market; Twitter (2011) itself encourages content creators to engage their audiences on the platform during televised events. BlueFin labs (bluefinlabs.com) and TrendrrTV (<http://trendrr.com>) have been monitoring social media use during real-time television events as well, but their goals are to produce data for marketers to be used in ad creation (Kafka, 2012b). The

potential for “massively-multi-person conversation” through the use of #hashtags has been previously acknowledged (Huang, Thornton, & Efthimiadis, 2010, p. 3); it can be an efficient way for users of the *second screen* model to follow each other’s posts, and for researchers to follow the conversation as well.

Chatterboxing is prevalent among those who watch television. Young British television viewers are chatterboxing more than older ones, but 26% of those surveyed reported commenting on a television program “on a second screen” (TV Licensing, 2012, p. 10). Chatterboxing has modified some viewing habits of respondents under the age of 35, with nearly one-fifth (19%) watching important programs in real time to avoid “social media spoilers” and 17% of them watching a missed show as a result of chatterboxing (p. 11). Table 1, reproduced from the data in the TV Licensing/BBC report, shows the percentages of users who chatterbox, by age.

Age group	%
18-24	46%
25-34	43%
35-44	31%
45-54	21%
55-64	14%
Over 65	11%

**Table 1: Percentage of people chatterboxing (TV Licensing, 2012, p. 10)**

Chatterboxers who worry about social media spoilers can be considered avid consumers of social media. We assume they are also at ease with the Twitter labeling conventions mentioned above, as these are the mechanisms that allow for participation in the online conversations. As they follow both screens, chatterboxers have as a basis for their conversations the shared content on the television; they adhere to the shared conventions for discussion on Twitter when they choose to use the trending #hashtag promoted by other chatterboxers.

### Real-Time Social Viewing

Already in 2009, Twitter was being described as the “new pulsating heart of the real-time Internet” (Malik, 2009). Chatterboxing may be seen as a logical progression for viewers who are already using social apps like GetGlue (getglue.com) to check in to events in real-time and to advertise potentially their activity to friends and followers on Facebook, Twitter, and Tumblr. Nationally televised real-time events like Super Bowl XLVI can be topics of interest in the social media sphere. Teams from two major East Coast cities (New England Patriots from Boston area, Massachusetts and Giants from New York City area) played

in the Midwest (Indianapolis, Indiana) on February 5, 2012, with the team from New York ultimately winning in the final moments. Ticket holders watched the game at the Lucas Oil Stadium and on-site fans participated in events in Super Bowl Village<sup>1</sup>. Around the country, fans watched at home and in public venues.

At the time of writing, 173,753 users had checked in to Super Bowl XLVI on GetGlue<sup>2</sup>. BlueFin reported the Super Bowl as being the largest social TV event recorded at the time, with a total of 12.2 million social media comments during the game. This number represented a growth of nearly 600% over social media comments in the previous year’s game (BlueFin, 2012). TrendrrTV reported a higher number and one that focused on Twitter instead of on social media in general: Mark Ghuneim of TrendrrTV reported 15.8 million tweets during the Super Bowl in 2012, up from 3.01 million in 2011 (Kafka, 2012a). Twitter as a medium provides a venue where the “public interplay of voices [...] give[s] rise to an emotional sense of shared conversational context” (boyd, Golder, & Lotan, 2010, p. 1); this shared context can be a major sporting event such as the Super Bowl. The instantaneous aspects of Twitter along with its mobility have helped it evolve into a medium to support the viewing of real-time televised events and shows and as a *second screen*.

### Qualitative Analyses of Tweets

Inferring meaning based on tweets is an uncertain task. Topics of tweets are difficult to ascertain because tweets themselves are ungrammatical and internally noisy (Michelson & Macskassy, 2010). Content analysis and other qualitative analyses of tweets are nonetheless undertaken in studies of Twitter usage, often on subsets of tweets as a way of gathering a richer data pool. Huang, Thornton, and Efthimiadis (2010) used qualitative methods to analyze the content of a subset of 224 tweets but did not report on the results in their paper. Honeycutt and Herring (2009) analyzed the functions of @mentions in a subset of 200 tweets, to address questions of language of the tweet. Another subset of 50 tweets in English was analyzed to count specific instances of the @mention in tweets (Honeycutt & Herring, 2009). In another Twitter study, user topics of interest were analyzed, with a pilot of four users’ tweets being compared against Wikipedia to test the topicality of the content (Michelson & Macskassy, 2010). As with other studies where an online user-created artifact is analyzed, Twitter studies that are qualitative in nature are unable to state with certainty the topic of a given tweet or the intention of the user.

<sup>1</sup>[http://www.indianapolissuperbowl.com/files/SuperBowlVillage\\_Map\\_FINAL.pdf](http://www.indianapolissuperbowl.com/files/SuperBowlVillage_Map_FINAL.pdf)

<sup>2</sup> [http://getglue.com/tv\\_shows/super\\_bowl\\_xlvi](http://getglue.com/tv_shows/super_bowl_xlvi)

## RESEARCH QUESTIONS

The primary research questions being investigated in this study are the following:

RQ1: Are @mentions and #hashtags in posts (*tweets*) different enough to represent two different labeling conventions?

To address the question of differences, tweets were taken from specific locations during a chatterboxing event where half of the locations implied more active involvement on the part of Twitter users.

RQ2: To what extent do #hashtags indicate *aboutness* in tweets?

A related question helps us understand the appropriateness of theories of organization of information in future analysis and study of tweets. In order to address these questions, we analyzed and studied labeling conventions in Twitter-based chatterboxing during a real-time nationally televised event as a case study.

## METHODOLOGY

The 2012 Super Bowl, held on February 5, 2012, was a nationally televised event with a large anticipated chatterboxing activity. Tweets from three locations with an invested interest (Boston, New York City metro area, and Indianapolis) in the event and three locations with a non-invested interest (Dallas, Miami, and Seattle) in the event were selected for study; an additional set of location-independent tweets was also collected. These latter tweets are used for the location-independent qualitative study.

Given that this chatterboxing event concerns a game played between two teams geographically based in two locations and that the game is played in a third location, we use these locations of Twitter users engaged in chatterboxing as a device to examine the conceptual differences of #hashtags and @mentions. The first two invested locations were selected because their teams (geographically-speaking) played in the event and the last because the place hosted the event. Three locations with a non-invested interest were selected only on two conditions: each location had a National Football League team and each location lie in a separate region of the United States. Tweets were analyzed for statistical differences.

Qualitative aspects of location-independent tweets were examined using a subset of tweets. A true random set of 100 numbers between 1 to 1490 was generated using random.org. Of these, 96 numbers were unique. The location-independent tweets corresponding to the selected random numbers were retained for study. Three of these tweets were empty of content and were discarded, leaving a set of 93.

## Data Collection

We used the *twitterR* (Gentry, 2012) package for the R

statistical language (R Development Core Team, 2012) to search and retrieve the respective samples of *tweets*. We retrieved seven sets of data based on seven separate searches using the #hashtag '#superbowl'. Per the GET search (Twitter, 2012) Twitter API, we attempted to retrieve a maximum of  $n=1500$  *tweets* per location. The first search was for the '#superbowl' hashtag and was limited only by a date range from 2012-02-05 to 2012-02-06. Despite the broad date range, timestamps indicate all retrieved tweets were posted during the event. The six location-based searches used in this analysis were additionally constrained and bounded by a twenty-mile buffer zone around points defined by the locations' latitudinal and longitudinal coordinates. These coordinates were identified through *toolserver.org* via the Wikipedia entry for each of the locations.

To aid our study, we used the R programming language to quantify the number of #hashtags and @mentions identified by their respective symbols within each location-based sample. These six location-based samples were divided into separate samples based on these counts. This resulted in six location-based samples representing counts of #hashtags and six other location-based samples representing counts of @mentions. Quantifying and dividing the *tweets* in this way allowed us to conduct statistical comparisons. After a preliminary analysis, an overlap between Twitter posts in the first hundred Boston and NY Metro cases (*tweets*) was observed. To correct for this, we reduced the sample size for each location to 1,300 cases each, taking the 101:1400 range from each sample. This insured that all tweets were unique and that each sample had the same total count of cases. Furthermore, since the tweets were retrieved using the '#superbowl' #hashtag, each tweet had this #hashtag. The #superbowl tags were thus subtracted from statistical counts, only counting the number of additional #hashtags. For example, in a tweet containing two hashtags, such as #superbowl and #giants, only the #giants hashtag is counted.

## Analysis

In the first part of the analysis, the overall goal was to compare the invested-interest group against the non-invested-interest group to test for differences in the uses of the two labeling conventions. A series of preliminary but necessary steps in the analysis addresses this step. The correlations between @mentions and #hashtags are first examined. The purpose is twofold: to understand the extent of the relationship between the two conventions and to determine whether there is justification for testing each separately in the succeeding group tests. If there is no statistical or substantial relationship between the two, then we assume that @mentioning and #hashtagging are two separate actions and should be examined separately.

Next we examine and compare the three locations within the invested-interest group (Boston ~ NY Metro area ~

Indianapolis). Our goal is to understand whether there is reason to believe that these three locations are similar or come from similar populations. If, based on the relationship test between #hashtag use and @mentioning, we have reason to believe that the two have separate functions, we test each separately. We repeat this process for the three locations within the non-invested-interest group (Dallas ~ Miami ~ Seattle) in order to determine whether these locations are similar.

If based on the previous two tests there is a reason to believe that members in the invested-interest group are similar and members in the non-invested group are also similar, we examine and compare the two groups to each other (between the two conversation groups and between the two organization groups, separately). This final statistical test will also say something about the influence of location on Twitter users in their posts during a real time nationally televised event. An analysis of the platform used to post the tweets, when possible, will also be analyzed in a small random sample of tweets to assess the potential degree of mobility of users.

To complement the quantitative methods, a qualitative analysis of tweets informs our understanding of conventions used in Twitter while chatterboxing. By studying a small random sample of tweets (n=100) we can develop a richer understanding of the character of content in Twitter while addressing one of our research questions. Of the 100 random numbers selected from among a set of non-location-dependent #superbowl tags, 93 complete and unique tweets were studied. Content of individual tweets was first analyzed through content analysis. The following four categories were identified, with the first two being

drawn from Honeycutt and Herring (2009)'s list of 12 *content themes*. Due to the limited context of chatterboxing, 12 content themes were deemed unnecessary; two similar *content theme* categories identified by Honeycutt and Herring (2009) were collapsed into one and one additional *content theme* used by them was selected. The second two categories were identified during the process of analysis. Each tweet was assigned one category:

- 1) Announce/advertise; Information for others: Announces information to the general readership of Twitter; posted information apparently intended for others – may be solicited or volunteered.
- 2) Opinion: Asserts a subject or evaluative position
- 3) Game events: Names specific content, players, events that pertain to the real-time game
- 4) Commercials: Names specific content, products, advertisers that pertain to the real-time commercials being shown

Other information about the tweets is also retained, including language of the tweet and posting platform.

To investigate the *aboutness* of #hashtags in relation to the content of the tweets in which they were located, #hashtags (other than #superbowl) are analyzed in the context of the tweet. These #hashtags are then coded according to their reflection of the tweet's *aboutness* and their relative location in the tweet (*in situ* or as extra-grammatical tags). Finally, to develop a richer understanding of chatterboxers' use of labeling conventions, we tallied the number of RTs in posts from all the locations.

	#hashtags								@mentions						
	0	1	2	3	4	5	>5		0	1	2	3	4	5	>5
<b>Invested</b>															
Boston	684	391	116	90	14	3	2		844	392	46	10	1	4	3
NY Metro	742	299	223	20	11	2	3		545	663	68	22	2	–	–
Indy	661	349	175	72	32	5	6		669	433	115	32	23	12	16
<b>Non-Invested</b>															
Dallas	824	312	104	28	26	4	2		880	308	75	13	12	3	9
Miami	845	313	103	25	7	4	3		842	360	73	17	5	2	1
Seattle	878	279	94	39	5	1	4		789	429	58	12	5	2	5

**Table 2: Frequencies of #hashtags (minus #superbowl hashtag) and @mentions, by location.**

	Boston @	NY Metro @	Indy @	Dallas @	Miami @	Seattle @
Boston #	.1133***	-.0030	-.0054	.0349	.0234	-.0337
NY Metro #	.0400	.0917***	-.0444^	-.0185	.0189	.0306
Indy #	.0180	.0266	.0492*	.0490*	-.0138	.0039
Dallas #	.0319	-.0094	-.0147	.0478^	-.0262	-.0476^
Miami #	.0237	-.0573	.0014	-.0085	.0799**	-.0193
Seattle #	.0124	.0104	.0057	.0273	.0004	.0344

\*\*\* p < 0.001, \*\* p < 0.01, \* p < 0.05, ^ p < 0.10

**Table 3: Kendall's Rank Correlation *tau* Matrix, #hashtags to @mentions.**

## RESULTS

### Descriptive statistics

Descriptive statistics indicate that the counts of @mentions and counts of #hashtags in each sample are skewed. Except for the NY Metro area's use of @mentions, for each of the eleven other samples, the means are greater than the medians and the standard deviations are greater than the means. Therefore, we chose various nonparametric rank-based statistical methods of analysis. We consider all samples independent and we use the Wilcoxon Rank Sum Test (sometimes called Mann-Whitney) to test two groups and the Kruskal-Wallis for three groups (Conover, 1980; Sprent, 1993). In both cases the null hypothesis is that all populations are identical with respect to their locations (here 'location' refers to a distribution's location along the x-axis), and the alternate hypothesis is that the locations are not identical. All tests are two-tailed. Table 2 shows the frequencies for the use of #hashtags and for the use of @mentions for all locations.

### Relationship Between #hashtags and @mentions

Our first task involves examining the relationship between #hashtags and @mentions to satisfy the first research question. Table 3 comprises a Kendall's *tau* (Sprent, 1993) correlation matrix between #hashtags and @mentions in all the samples. Kendall's *tau* is a correlation coefficient based on a ranked set of counts in ascending order. It tests whether two variables are statistically dependent. There are five statistically significant correlations at the  $p < .05$ , all of them are positive, and the two highest correlations involve locations with an invested interest in the event.

We continued the investigation of organization and conversation use in each location by conducting Chi-Square tests of independence. For this test, we composed 2x2 frequency tables representing the counts of events or non-events, expressed as 0 indicating no additional use of #hashtags or @mentions and 1 indicating any additional use of the #hashtags or @mentions. Here we test the null hypothesis that the outcomes, as expressed in the relation between organization and conversation, of the event are

statistically independent. As seen in Table 4, we reject the null hypothesis in the case of the Boston, NY Metro area, as well as in the Miami area. This suggests that for these three areas, and especially for the Boston and NY Metro area (highest Chi-square values), watching the nationally televised event in

invested interest areas increases the statistical dependence between #hashtags and @mentions within tweets. This is likely the case because users in Boston and users in the NY Metro area will be tweeting to specific other users in their respective areas as their teams play the game.

	$\chi^2$	df	p
Boston	9.4616	1	0.002**
NY Metro	7.0798	1	0.008**
Indy	0.6637	1	0.415
Dallas	2.4503	1	0.118
Miami	5.4624	1	0.019*
Seattle	0.0057	1	0.9399

**Table 4: Chi-Square test of independence for frequency of #hashtags and @mentions, within each sample.**

This test demonstrates that dependence increases between conversation and organization. The correlations and the Chi-Square values, however, are low enough to let us conclude that #hashtaging and @mentioning are two separate actions and so require separate tests.

### Analysis: Invested-Interest Groups

#### @Mentions

The Kruskal-Wallis rank sum test was used to compare @mentions in each of the invested-interest groups: Boston, the NY Metro area, and Indianapolis. The null hypothesis is that these three samples are from the same population. The relationship was significant ( $H = 130.0359$ ,  $df = 2$ ,  $p < 0.001$ ). A post hoc pairwise Wilcoxon rank sum test, with  $p$  adjusted by the Holm method (Holm, 1979), was used to test the null hypothesis in pairwise fashion. The test showed a highly significant relationship between the NY Metro area and Boston ( $p < 0.001$ ) and Indianapolis and Boston ( $p < 0.001$ ) and a somewhat significant relationship between the NY Metro area and Indianapolis ( $p = 0.058$ ). This suggests that the NY Metro area and Boston as well as Indianapolis

and Boston have different distribution locations and only the NY Metro area and Indianapolis probably have the same distribution locations.

**#Hashtags**

The Kruskal-Wallis rank sum test was used to compare organization use among the three invested-interest locations: Boston, NY Metro area, Indianapolis. The test was statistically significant ( $H = 12.1263, df = 2, p = 0.0023$ ). A post hoc Wilcoxon rank sum test was used to test the null hypotheses in pairwise fashion. The test only showed a significant result between the NY Metro area and Indianapolis ( $p = 0.0015$ ). This suggests that these two populations do not have identical distribution locations and that the other pairwise comparisons (NY Metro area and Boston; Indianapolis and Boston) probably have identical distribution locations.

**Analysis: Non-Invested-Interest Groups**

**@Mentions**

We used the Kruskal-Wallis rank sum test to examine conversations among the three non-invested-interest locations: Dallas, Miami, and Seattle. The relationship was significant ( $H = 8.7802, df = 2, p = 0.0124$ ). The post hoc pairwise Wilcoxon rank sum test only showed a significant relationship between the Seattle and Dallas populations ( $p = 0.0098$ ). This suggests that only these two samples have different distribution locations and that the pairwise comparisons between Miami and Dallas as well as Seattle and Miami probably have identical distribution locations.

**#Hashtags**

The Kruskal-Wallis rank sum test was used to compare organization use among the three non-invested interest locations: Dallas, Miami, and Seattle. The relationship was not significant ( $H = 5.1268, df = 2, p = 0.07704$ ). In terms of #hashtags, we fail to reject the null hypothesis that these three samples are from the same population. That is, the populations are probably identical with respect to their distribution locations.

**Invested-Interest Groups and Non-Invested Interest Groups**

The tests make it complicated to generalize and justify our *a priori* grouping: invested-interest and non-invested interest. For example, in the @mention use in the *a priori* invested-interest group, we only have strict justification to group the NY Metro area and the Indianapolis locations. Likewise, in #hashtag use, we only have strict justification for two separate groups: the NY Metro area and Boston as well as Indianapolis and Boston.

We see similar difficulties in the non-invested interest group. In @mention use, we only have strict justification that Miami and Dallas form one group and Seattle and Miami form another group. In #hashtag use, however, we

have justification to group together all three areas: Dallas, Miami, and Seattle.

The tests show difficulties with broadly distinguishing between an invested-interest group and a non-invested interest group. However, we still find it worthwhile to compare *a priori* groupings such that we compare the two @mention groups with each other and the two #hashtag groups with each other. The two null hypotheses are that (1) the invested-interest @mention group is the same as the non-invested-interest @mention group and (2) the invested-interest #hashtag group is the same as the non-invested-interest #hashtag group. The Wilcoxon rank sum test is used to test these hypotheses. For @mention use, the test was significant with  $Z = 9.9411, p < 0.001, 95\% \text{ CIs } [0.0005, 0.00011]$ . For #hashtag use, the test was also significant with  $Z = 11.5831, p < 0.001, 95\% \text{ CIs: } [0.00004, 0.00009]$ . We reject both null hypotheses that the two set groups are each from the same population.

Although the pairwise comparisons stress that there are *a priori* difficulties with specific groupings, the broad comparison between the two groups supports a positive response to our main research question. Further, we suggest that location influences Twitter users use of @mention and #hashtag conventions in their posts (*tweets*) during a real time nationally televised event and that location serves as a good way to conceptually differentiate between #hashtags and @mentions.

**Table 5. Frequencies of labeling conventions in the**

Location	#	% #	@	% @	RT	% RT
Boston	980	75%	556	43%	321	25%
NY Metro	879	68%	873	67%	659	50%
Indianapolis	1116	86%	1013	78%	338	26%
Dallas	743	57%	627	48%	217	17%
Miami	661	51%	594	46%	297	23%
Seattle	637	49%	648	50%	372	29%
Location-Independent	417	32%	569	44	350	27%

samples, by location.

**Mobileness of #superbowl Users**

To gauge the mobileness of users, a small sample of #superbowl tweets emanating from any geographic location were analyzed for the mobileness of the platform used. Of the sample of 93 complete and unique tweets, information about the platform was recorded for only 64 tweets. Of them, 47 (73%) came from mobile devices, 1 (2%) from a personal computer, and 16 (25%) from third party software that operates both on phones and in web browsers. When



considering the notion of interaction by location, it is useful to revisit the kinds of activity in terms of percentages of the three primary labeling conventions. Table 5 (above) revisits the content indicated in Table 2, focusing on the total percentages of tweets with labeling conventions. Relative to previous studies, these percentages indicate more aggressive use of all three labeling conventions by this sophisticated group of mobile chatterboxing users.

### Qualitative Analysis

We begin the qualitative analysis of the randomly selected set of location-independent #superbowl tweets by assessing their topical content. Beginning with Honyecutt and Herring (2009)'s coding scheme and simplifying it greatly based on the limited content produced by chatterboxers, we analyzed the content of the set of tweets (see Table 6). Slightly over one-third of tweets (n=32, 35%) were about the game itself, with over one-quarter (n=26, 28%) expressing general opinions, right under one-quarter (n=22, 24%) providing some kind of announcement for others, and 14% (n=13) commenting specifically on commercials. Of the 93 tweets, 87 were in English, 5 in Spanish, and 1 in German; all of these tweets were included in the analysis.

Content of tweets	No.	%
Announce/advertise; Info for others	22	24%
Express opinion (general likes, dislikes)	26	28%
Commercials (explicit mention of content)	13	14%
Game events (explicit mention of content)	32	35%

**Table 6. Content of random sample of #superbowl tweets independent of location.**

Of the 93 tweets containing the #superbowl #hashtag selected for study, 27 (29%) had additional #hashtags (66 tweets (68%) contained only the #superbowl #hashtag); a total of 138 #hashtags were retained for study. All of the #superbowl hashtags were deemed to have been about the venue except one, which was embedded in a sentence about a singing competition. #Hashtags described the contents of the tweet in which they were located by either reinforcing the words or substituting directly for words. In two cases, #hashtags were complete sentences. Preliminary observations reveal that #hashtags about venue, when accompanied by another #hashtag, were generally accompanied by a hashtag indicating the *aboutness* of the tweet. When three #hashtags were indicated, the third generally indicated affective content or reiterated the venue. See Table 7 (following page) for more details about *aboutness* of the 46 non-superbowl #hashtags.

Placement of all 138 #hashtags within the tweets was also observed: 102 (75%) #hashtags functioned in a way similar to tags or descriptors, being placed outside of the text of the

<i>Aboutness</i> of non-#superbowl #hashtags (n=46)	No.	%
Directly concerning the contents of the tweet	23	50%
Teams (e.g. #Patriots)	8	17%
Affective or conversational (e.g. #Imfao, #JustSaying)	6	13%
Venue other than #superbowl (e.g. #SB46)	3	7%
Network (e.g. #NBC)	2	4%
Other (e.g. #IndianapolisChildrensChior)	2	4%
Complete sentences (e.g. #DaddysDrinkinCabernet)	2	4%

**Table 7. *Aboutness* of non-#superbowl #hashtags in the random sample independent of location.**

tweets at the end (n=94, 68%), in the beginning (n=4, 3%), or in the middle of two sentences or complete thoughts (n=4, 3%). The remaining #hashtags (n=36, 26%) were placed within the text of tweets, functioning as words or as complete sentences.

### DISCUSSION

Based on the results of the analysis, similar to Efron (2010), we view #hashtags on Twitter as first-order organizational acts rather than first-order communicative acts. In this way, library and information science has much to offer to the study of #hashtag use. Specifically, #hashtags serve to categorize a tweet by what the tweet is *about*. A #hashtag on Twitter functions as an index term that descriptively points to a tweet's topic or its *aboutness* in a process similar to how Maron (1977) describes the work of a professional indexer. #Hashtags therefore seem to be a unique form of topical relevance that is uncontrolled, chaotic, and impermanent. Users apply #hashtags to perform two functions: to describe in order to categorize, and, in their purely relevance form, to retrieve in order both to follow or join a conversation.

To further investigate the question of whether Twitter #hashtags, within this context, enhance organization or enable conversation, we again consider the role of organization and conversation. Organization implies that content is valuable (curated), and should be retrievable. Conversation implies a communicative act with an implied interlocutor. Because the tweets selected for this study have #hashtags, they are not solely meant to be broadcast to the followers of the users posting them; they are meant to be either included as parts of a larger conversation. In this way, we concede, as Weng et al (2010) do, that there is a facilitation of conversation through the use of #hashtags;

However, in a real-time environment like a nationally

televised event, #hashtags use promotes organization. #Hashtags function in a way similar to social tags in traditional social tagging sites like CiteULike, Flickr, and Delicious due to their spontaneous use and creation and their grouping function for *aboutness* and affective topics.

When deciding on a theoretical framework for analyzing and studying #hashtagged tweets in Twitter, it is necessary to distinguish between the field of organization and communication for choosing a theory. Results in this study imply that the greater the number of Twitter conventions used, the more the tweet should be considered conversation.

Our work continues the discussion of how users label the content of their posts for organization, discovery, and consumption by other users. As a case study, it raised questions for further investigation, while answering some basic questions about the functioning of labeling conventions, location, and *aboutness* in #hashtags.

### LIMITATIONS, AND FURTHER STUDY

This study includes a number of limitations. The Super Bowl is a popular event and therefore the locational, organizational, and conversational aspects of chatterboxing during this event may be more pronounced than in other Twitter studies where certain groups have a non-invested interest in a much less dramatic event and its outcome. Also, while we show that there is a difference between users' symbolic conventions to organize and communicate on Twitter, and that the relationship between these conventions becomes statistically dependent for invested-interest locations in our chatterboxing context, we make no generalizable, qualitative claims about the use of these two conventions. More understanding about the qualitative differences and uses is needed.

Despite these limitations, this study offers the first large-scale investigation testing the similarity of Twitter labeling conventions and the potential to consider #hashtags as organizational mechanisms, doing so from within the novel context of chatterboxing. This study therefore lays the groundwork for the investigations mentioned above. Additionally, it claims that location is an important variable within the chatterboxing context. How location may be an important variable in other social computing contexts and activities, as well as chatterboxing, also needs to be examined in further study.

### CONCLUSION

Based on the quantitative aspects of the study, we rejected the notion that #hashtags are identical to @mentions, and based on the qualitative aspects, we reject the notion that #hashtags do not provide for the description of tweet *aboutness* in a way that may be similar to social tagging. Accordingly, we feel that it is appropriate for researchers to use organization of information theory when analyzing #hashtags in tweets.

Within the context of chatterboxing during a real-time event, we drew several conclusions from the statistical tests. First, empirical evidence suggested the use of #hashtags and @mentions in Twitter posts are different actions and result in different entities. Location has a slight positive influence on their statistical dependency.

A more minor conclusion was that location seemed to play an important role in implicit group formation when frequencies are used to measure the respective symbols. The implication is that location may be an important dependent variable in chatterboxing studies. We argued later that this is an important research area to be investigated in future research.

### REFERENCES

- BlueFin. (2012, February 6). Super Bowl XLVI: A social TV phenomenon. Retrieved from <http://bluefinlabs.com/blog/2012/02/06/super-bowl-xlvi-a-social-tv-phenomenon/>
- boyd, d., Golder, S., and Lotan, G. (2010). Tweet, tweet, retweet: Conversational aspects of retweeting on Twitter. In *System Sciences (HICSS), 2010 43rd Hawaii International Conference on*, pages 1-10. IEEE.
- Conover, W.J. (1980). *Practical nonparametric statistics*. Second edition. New York: John Wiley & Sons.
- De Bellis, N. (2009). *Bibliometrics and citation analysis: From the Science Citation Index to cybermetrics*. Lanham, MD: Scarecrow Press.
- Dougall, J. (2012, March 15). Social web makes TV viewers 'chatterboxers'. Sky News. Retrieved from <http://news.sky.com/home/technology/article/16189017>
- Efron, M. (2010). Hashtag retrieval in a microblogging environment. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, pages 787-788. doi: 10.1145/1835449.1835616
- Efron, M. (2011). Information search and retrieval in microblogs. *Journal of the American Society for Information Science and Technology*, 62(6), 996-1008.
- Gentry, Jeff. (2012). twitteR: R based Twitter client. R package version 0.99.19. <http://CRAN.R-project.org/package=twitteR>.
- Hjørland, B. (2001). Towards a theory of aboutness, subject, topicality, theme, domain, field, content . . . and relevance. *Journal of the American Society for Information Science and Technology*, 52(9), 774-778. doi: 10.1002/asi.1131
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2), 65-70. <http://www.jstor.org/stable/4615733>
- Honeycutt, C. and Herring, S. C. (2009). Beyond microblogging: Conversation and collaboration via

- twitter. In *System Sciences, 2009. HICSS'09. 42nd Hawaii International Conference on*, pages 1-10. IEEE.
- Huang, J., Thornton, K. M., and Efthimiadis, E. N. (2010). Conversational tagging in twitter. *Proceedings of the 21st ACM conference on Hypertext and hypermedia HT 10*, page 173.
- Jackoway, A., Samet, H., and Sankaranarayanan, J. (2011). Identification of live news events using Twitter. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks*, LBSN '11, New York, NY, USA. ACM.
- Kafka, P. (2012a). A super social bowl. *All Things Digital*. Retrieved from <http://allthingsd.com/20120205/a-super-social-bowl>
- Kafka, P. (2012b). Twitter while you watch TV? BlueFin labs is watching. *All Things Digital*. Retrieved from <http://allthingsd.com/20120124/twitter-while-you-watch-tv-bluefin-labs-is-watching/>
- Levy, S. (2007, April 29). Twitter: Is brevity the next big thing? *Newsweek*. Retrieved from <http://www.thedailybeast.com/newsweek/2007/04/29/twitter-is-brevity-the-next-big-thing.html>
- Kipp, M.E.I. (2010). Searching with tags: Do tags help users find things? *Proceedings of the American Society for Information Science and Technology*, 46(1). doi:10.1002/meet.2009.14504603406
- Malik, O. (2009). A brief history of Twitter. *GigaOM.com*. Retrieved from <http://gigaom.com/2009/02/01/a-brief-history-of-twitter/>
- Maron, M. E. (1977). On indexing, retrieval and the meaning of about. *Journal of the American Society for Information Science*, 28(1), 38-43.
- Marwick, A.E., & boyd, d. (2010). I tweet honestly, I tweet passionately: Twitter uses, context collapse, and the imagined audience. *New Media & Society*, 13(1), 114-133. doi:10.1177/1461444810365313
- Michelson, M. and Macskassy, S.A. (2010). Discovering users' topics of interest on Twitter. *AND '10 Proceedings of the fourth workshop on Analytics for noisy unstructured text data*. pages 73-79. ACM Press.
- Mueller, D. N. (2009). Digital underlife in the networked writing classroom. *Computers and Composition*, 26(4), 240-250. doi:10.1016/j.compcom.2009.08.001
- Peters, I. and Stock, W. G. (2007), Folksonomy and information retrieval. *Proceedings of the American Society of Information Science and Technology*, 44, 1-28. doi:10.1002/meet.1450440226
- R Development Core Team. (2012). R: A language and environment for statistical computing. The R Foundation for Statistical Computing. <http://www.R-project.org/>.
- Ransom, N., & Rafferty, P. (2011). Facets of user-assigned tags and their effectiveness in image retrieval. *Journal of Documentation*, 67(6), 1038-1066. doi: 10.1108/00220411111183582
- Sprent, P. (1993). *Applied nonparametric statistical methods*. Second edition. London: Chapman & Hall.
- Saracevic, T. (2007). Relevance: A review of the literature and a framework for thinking on the notion in information science. Part II. Nature and manifestations of relevance. *Journal of the American Society for Information Science and Technology*, 58(13), 1915-1933. DOI:10.1002/asi.20682
- TV Licensing. (2012, March). TeleScope: A look at the nation's changing viewing habits from TV Licensing. Retrieved from [http://www.tvlicensing.co.uk/resources/library/BB\\_C/MEDIA\\_CENTRE/TV\\_Licensing\\_Telescope\\_Report\\_2012.pdf](http://www.tvlicensing.co.uk/resources/library/BB_C/MEDIA_CENTRE/TV_Licensing_Telescope_Report_2012.pdf)
- Twitter. (2011, May 04). Watching together: Twitter and TV [Web log post]. Retrieved from <http://blog.twitter.com/2011/05/watching-together-twitter-and-tv.html>.
- Twitter Developers. (2012). GET search. Retrieved on April 15, 2012 from <https://dev.twitter.com/docs/api/1/get/search>.
- Weng, J., Lim, E.-P., He, Q., and Leung, C. W. K. (2010). What do people want in microblogs?: Measuring interestingness of hashtags in twitter. In *Data Mining (ICDM), 2010 IEEE 10th International Conference on*, ICDM '10, pages 1121-1126, Washington, DC, USA. IEEE.
- Yang, J. and Leskovec, J. (2010). Modeling information diffusion in implicit networks. [http://ieeexplore.ieee.org/xpl/freeabs\\_all.jsp?arnumber=5694014](http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=5694014). Data Mining (ICDM), 2010 IEEE 10th International Conference on.