

Twitter content classification

by Stephen Dann

Abstract

This paper delivers a new Twitter content classification framework based 16 existing Twitter studies and a grounded theory analysis of a personal Twitter history. It expands the existing understanding of Twitter as a multifunction tool for personal, profession, commercial and phatic communications with a split level classification scheme that offers broad categorization and specific sub categories for deeper insight into the real world application of the service.

Contents

[Introduction](#)

[Background](#)

[Public timeline 'Twiterature'](#)

[Specific use 'Twiterature'](#)

[A new category structure](#)

[Method](#)

[Twitter categories](#)

[Results, limitations and conclusions](#)

Introduction

Current Twitter classifications have focused on the macro-level public timeline at the expense of the richness of depth from individual histories. This paper delivers a new classification framework that offers a deeper insight into Twitter content through six broad categories, and 23 detailed subcategories for analyzing individual timelines based on ground theory analysis and an extensive review of the existing Twitter literature.

Background

Twitter is a popular short messaging service available through Web page, desktop and mobile software. Built on a restricted set of features including public timeline messages and private direct messages, it has evolved rapidly through user innovation with the retweet (RT) reply (@), and hashtag (#) makes being introduced by consensus and community behavior. Prior research on Twitter has clustered around criticism, attempts to define "best practice", end-user motivations, and content classifications.

Criticism of Twitter have been prevalent in regards to its perceived limited value as a platform for public relations (Eyrich, *et al.*, 2009), the accuracy and value of unfiltered live information from crisis events (Thussu, 2009) and aspects of its security against malicious attacks, social engineering and trust-based exploitation (*Infosecurity*, 2009).

Best practice focuses on applications of Twitter in health community (Berger, 2009), public libraries (Cahill, 2009; Cuddy, 2009), political campaigns (Cetina, 2009; Henneburg, *et al.*, 2009), business (Dudley, 2009; Power and Forte, 2008), journalism (Ettama, 2009), civil unrest and protests (Fahmi, 2009), social activism (Galer-Unti, 2009), live coverage of events (Gay, *et al.*, 2009), and

eyewitness accounts of news stories (Lariscy, *et al.*, 2009), government (Macintosh, 2009), education and learning (Parslow, 2009).

End-user motive research likened Twitter to radio as a casual listening platform (Crawford, 2009), a means for creating an illusion of physicality (Hohl, 2009), a sense of connectedness with others (Henneburg, *et al.*, 2009), and as a venue for conversation (Steiner, 2009).

Content analysis has ranged between the "Twitter as a serious business" and "Twitter as conversations apparently about nothing" with content analysis tending towards the serious end expressing concerns over the volume of personal and social content (Pear Analytics, 2009; Zhao and Rosson, 2009; Java, *et al.*, 2007).

Public timeline 'Twiterature'

Four papers have used samples of the public timeline data in content analysis to develop insight into the real life application of Twitter. Java, *et al.* (2007) examined 1,348,543 tweets from 76,177 users from 1 April 2007 to 30 May 2007. The analysis resulted in two metrics — style of use (followers/following numbers), and user intention (manually coded Twitter content). Three user categories were identified as information sharing (high follower, low following), information seeking (low follower, high following), and friendship-relationship (rough equivalency in follower/following score). The four meta-categories of content include "daily chatter" which covered the daily routine of the individual user, "conversations" which included replies to other users (use of the @) protocol, "information or URL sharing" which were classified according to the presence of full length or shortened URL and "news reporting" which included sport, weather and commentary on current affairs.

Krishnamurthy, *et al.* (2008) removed the content of the tweets from consideration in favor of studying the social and technical infrastructure by classifying users based on follower/following counts, means for using the service and volume of posts. The most common methods for posting updates was the Twitter Web site (61.7 percent), custom applications (22.4 percent), mobile/txt (7.5 percent) and instant messenger (7.2 percent). Volume of posts and follower/following counts were connected with accounts followed by more than 250 posting more than accounts with less than 250 followers.

Pear Analytics (2009) six content categories based on a sample of 2,000 tweets. News (3.6 percent) represented mainstream media content, spam (3.75 percent), self-promotion (5.85 percent) covered corporate messages about products, services or special offers. Pointless babble (40.55 percent) included personal communications, observations and general chat, conversational (37.55 percent) covered questions, polls and @replies and pass-along (8.70 percent) covered retweeted (RT) content.

Jansen, *et al.* (2009) looked at branded Twitter accounts and the appearance of brand names in the public timeline. Two classification schemes were developed during the research — a sentiment scale (No Sentiment, Wretched, Bad, So-so, Swell, Great) to classify tweets on a negative to positive spectrum; and an action-object pair approach for classifying brand related tweets which describe actions towards a specific object which resulted in four categories of tweets: sentiment as the expression of positive or negative opinions; information seeking asking about a brand, information providing answering other brand questions and comment as the use of a brand in a tweet where the brand was secondary to the purpose of the tweet.

Specific use 'Twiterature'

Three papers looked at specific purpose uses of Twitter in terms of replies (Honeycutt and Herring, 2009), retweeting (boyd, *et al.*, 2010) and personal broadcasts (Naaman, *et al.*, 2010).

Honeycutt and Herring (2009) examined two uses for the @symbol — "addressivity" in directed conversations (91 percent) and referencing (5 percent) which indicates the presence of another user. Use of the @ statement was broken down into reports of personal experience (17 percent), direct conversation (33 percent), information (16 percent), and encouragement (11 percent).

boyd, *et al.* (2010) focused on a non-exhaustive list of reasons for the retweeting including relaying valuable content (news, information), endorsing a particular user or topic, creating a conversation about an existing tweet and for personal reasons of friendship, loyalty or karma. Four categories of retweet were recognized — retweets containing URLs (52 percent), hashtags (18 percent),

encapsulated retweets which contain an existing retweet (11 percent) and reply retweets where an @reply to the person retweeting the tweet is present (9 percent).

Naaman, *et al.* (2010) produced a nine item list of broadcast statements including information sharing, opinions/complaints; statements including undirected random thoughts and observation; self-promotion which includes links to blog posts or other user generated content; "Me now" answers to the status update question; "Anecdote (me)" which is a past event; "Anecdote (others)" which is a story of other people; "Question to followers" is a directive seeking tweet and Presence Maintenance represents messages covering the user's location and movements.

A new category structure

All prior studies seeking to classify Twitter have been limited to broad categories of content such as "Information", "Conversation", "Broadcast", or "Other". These six top-level categories can be divided into distinct subcategories for in-depth analysis of Twitter using a grounded theory content analysis that combines the existing research categories with a content analysis of a Twitter timeline.

Method

A self-reflective analysis of the author's own Twitter timeline was chosen as the primary data set. The Twitter stream is licensed under a Creative Commons Public Domain Dedication which allows for the captured dataset to be made available for future replication studies. The author's account adheres to the Krishnamurthy, *et al.* (2008) 250 follower rule (274 Following /355 Follower).

The captured Twitter data set consists of 2,841 comments from the origin of the account in 2007 (Tue 13 Mar 2007 11:53:01) to the point of data collection (Tue 18 Aug 2009 07:29:30). Archival Twitter data was captured from the timeline using the Sujathan (2009) "Twitter to PDF" software. Data from the text files was then imported into individual Microsoft Excel for manual coding based on the existing theoretical categories.

For the purpose of the exercise, any category which did not have matching tweet content has remained in the full conceptual framework, and is treated as a null result, rather than being excluded from the results as not all Twitter accounts are expected to produce content to fit all categories.

Twitter categories

Table 1 outlines the respective content categories from the prior literature and was sourced from the works of Java, *et al.* (2007); Jansen, *et al.* (2009); Pear Analytics (2009); Honeycutt and Herring (2009); and, Naaman, *et al.* (2010).

Table 1: Current and Prior Categorizations

	<i>Java et al (2007)</i>	<i>Jansen et al (2009)</i>	<i>Pear Analytics (2009)</i>	<i>Honeycutt and Herring (2009)</i>	<i>Naaman et al (2010)</i>
<i>Conversational</i>	“conversations”	information seeking	conversational	Addressivity About addressee Exhort Solicit information	Question to followers
<i>Pass along</i>	“information or URL sharing”	information providing	pass-along value self-promotion	Information for others Announce/ advertise Other’s experience	Information Sharing Self Promotion
<i>News</i>	“news reporting”	information providing	News	Information for others	
<i>Status</i>	“daily chatter”	Comment / Sentiment	pointless babble	Self experience Announce/ advertise Opinion	Opinions/Complaints Me now Anecdote (me) Self Promotion
<i>Phatic</i>	“daily chatter”	Comment / Sentiment	pointless babble	Self experience Exhort Information for self Metacommentary Media use Other	Presence Maintenance Anecdote (others) Statements & Random Thoughts
<i>Spam</i>			Spam		

Each primary domain was divided in a series of subcategories based on data from the author’s account, and the detailed descriptions of the prior research paper’s different content categories. Priority orders were placed on the ranking of the content to determine which category should be used for a tweet which may match multiple criteria. The final subcategory in each section acts as a “catch-all” classification for tweets which matched the primary domain, but did not fit one of the other subcategories. Table 2 represents the summary of the subcategories and the supporting literature for the category.

Category	Definition	Reference
<i>Conversational</i>	<i>Uses an @statement to address another user</i>	
1. Query	Questions, question marks or polls	Pear Analytics, 2009; Java, et al., 2007; Jansen, et al., 2009; Naaman, et al., 2010; Honeycutt and Herring, 2009
2. Referral	An @response which contains URLs or recommendation of other Twitter users. (Excludes RT @user)	Pear Analytics, 2009; Naaman, et al., 2010; Honeycutt and Herring, 2009
3. Action	Activities involving other Twitter users	Jansen, et al., 2009; Hohl, 2009; Honeycutt and Herring, 2009
	Catch-all classification for	Jansen, et al., 2009; Java, et al., 2007;

4. Response	conversation @tweets	Steiner, 2009; Honeycutt and Herring, 2009
<i>Status</i>	<i>An answer to "What are you doing now?"</i>	
1. Personal	Positive or negative sentiment in the form of personal opinion or emotional status	Jansen, <i>et al.</i> , 2009; Naaman, <i>et al.</i> , 2010; Honeycutt and Herring, 2009
2. Temporal	Content referencing specific dates, times, statements of temporal nature (waiting) and temporal action ("Time to")	Naaman, <i>et al.</i> , 2010
3. Location	Geographic references and location statements, including statements of traveling, location change	Makice, 2009; Naaman, <i>et al.</i> , 2010
4. Mechanical	Statements relating to any form of technology or mechanical systems (cars, phones, printers and photocopiers)	
5. Physical	Sensory experiences of a physical nature	Naaman, <i>et al.</i> , 2010
6. Work	Reference to work related activity	Heany and McClurg, 2009; DiMicco, <i>et al.</i> , 2008; Zhao and Rosen, 2009
7. Automated	Status announcements triggered by third party applications such as media players, games or software	Honeycutt and Herring, 2009
8. Activity	Activity statements answering "What are you doing now?"	Naaman, <i>et al.</i> , 2010; Honeycutt and Herring, 2009
<i>Pass along</i>	<i>Tweets of endorsement of content</i>	
1. RT	Any statement reproducing another Twitter status using the via @ or RT protocol	Java, <i>et al.</i> , 2007; Pear Analytics, 2009; Naaman, <i>et al.</i> , 2010; boyd, <i>et al.</i> , 2010
2. UGC	Links to content created by the user (blog/video/picture)	Pear Analytics, 2009; Naaman, <i>et al.</i> , 2010; Honeycutt and Herring, 2009
3. Endorsement	Links to Web content not created by the sender	Zhao and Rosson, 2009; Naaman, <i>et al.</i> , 2010; Honeycutt and Herring, 2009; boyd, <i>et al.</i> , 2010
	<i>Identifiable news</i>	

<i>News</i>	<i>content which is not UGC</i>	
1. Headlines	Coverage of breaking news and personal eyewitness accounts of news events	Fahmi, 2009; Lariscy, et al., 2009; Gay, et al., 2009; Pear Analytics, 2009; Java, et al., 2007
2. Sport	Identifiable results of sporting events	Java, et al., 2007; Pear Analytics, 2009
3. Event	Any tweet which represents the live discussion of an identified or identifiable event	Lariscy, et al., 2009; Gay, et al., 2009
4. Weather	Report of weather conditions without commentary	Honeycutt and Herring, 2009
<i>Phatic</i>	<i>Content independent connected presence</i>	
1. Greeting	Statements of greetings to the broader Twitter community	Naaman, et al., 2010, Hohl, 2009; Honeycutt and Herring, 2009; Miller, 2008
2. Fourth wall	Textual equivalent of comments made directly to camera in television or cinema	Miller, 2008; Honeycutt and Herring, 2009
3. Broadcast	Textual soliloquy, monologue and undirected statements of opinion	Zhao and Rosson, 2009; Crawford, 2009; Honeycutt and Herring, 2009; Naaman, et al., 2010
4. Unclassifiable	Errors, cat-on-keyboard input and unclassifiable strings of text	Honeycutt and Herring, 2009
<i>Spam</i>	Junk traffic, unsolicited automated posts, and other tweets generated without user consent due to malware	Pear Analytics, 2009

Conversational

Conversational posts provide the building blocks of the social interaction between users which leads to the development of community, creation of interpersonal relationships, and the perception of reciprocity between Twitter users and their followers.

Table 3: Conversational content.			
Category	N	Percentage	Exemplar
<i>Conversational</i>			
1. Query	480	17	Invading Germany from France. Who's with me?
2. Referral	66	2	@USERNAME Items under \$1000 are exempt. http://is.gd/AV7K

3. Action	77	3	*waves at @USERNAME*
4. Response	850	30	@USERNAME Beware the polar bears.

Conversational categories use the @ symbol to indicate a directed message to another Twitter user, and account for significant volumes of traffic as users host conversations across Twitter. Four classifications have been created within the conversational category — action, query, referral, and response.

1. *Conversational — Query*: Any @tweet with both an @ symbol and a question mark, any direct question (Naaman, *et al.*, 2010; Java, *et al.*, 2007; Jansen, *et al.*, 2009; Honeycutt and Herring, 2009) and any tweet link to a poll (Pear Analytics, 2009).

2. *Conversational — Referral*: Any full length or shortened URL directed at another user, statement such as "@User1 meet @User2" and the #followfriday (#FF) tradition of recommending other Twitter users (Pear Analytics, 2009; Naaman, *et al.*, 2010; Honeycutt and Herring, 2009).

3. *Conversational — Action*: Comments involving other Twitter users such as "at event with @user1", or "watching @user2 performing" based on Honeycutt and Herring (2009) 'reference' use of the @symbol, descriptions of activity involving other users through the use of self-referential asterisks (*undertakes activity with @user*) or mock-IRC commands (/me engages with @user) which create the illusion of physicality between twitter users (Hohl, 2009; Jansen, *et al.*, 2009).

4. *Conversational — Response*: Tweets intentionally engaging another user by means of the @ symbol which does not meet the other requirements of containing action, questions or referrals (Honeycutt and Herring, 2009; Jansen, *et al.*, 2009; Java, *et al.*, 2007; Steiner, 2009).

Status

Status messages take the form of an answer the question Twitter poses on the update page on their Web site — "What are you doing now?"

Category	N	Percentage	Exemplar
<i>Status</i>			
1. Personal	221	8	I liked Modest Mouse after they became famous.
2. Temporal	170	6	Waiting for my 2pm performance review to start.
3. Location	69	2	Standing in a lecture theatre talking about Marketing Management.
4. Mechanical	106	4	Well ... I'm in trouble. Used 3829.060MB (62.322%) of your 6GB. You have 22 days remaining.
5. Physical	37	1	It's freezing out there this morning.
6. Work	196	7	Firing off e-mail after e-mail to clear my to do list (knowing that's a great way to regenerate to do list items doesn't stop me or help me).
7. Automated	—	—	—
8. Activity	35	1	Playing with the Internet in the name of science.

Eight categories of status have been identified from the literature and data set covering personal, temporal, location, mechanical, physical, work, automated and activity statements of what activity or action the user is currently undertaking.

1. *Status — Personal*: Any tweets using personal pronouns, statements of positive or negative sentiment, personal opinions or emotional status (Jansen, *et al.*, 2009; Naaman, *et al.*, 2010; Honeycutt and Herring, 2009).

2. *Status — Temporal*: Any content referencing specific dates, times, temporal activity (or inactivity such as waiting) present activity or recent action with an emphasis on the time of the event (Naaman, *et al.*, 2010).

3. *Status — Location*: Comments regarding travel, flights, cab rides and other transport issues and social network 'ping' commands notifying followers of changes in the author's locations manually or through automated announcements such as Foursquare checkins (Makice, 2009; Naaman, *et al.*, 2010).

4. *Status — Mechanical*: Any tweets relating to technology or mechanical systems such as computers, cars, phones or equipment, data, and related technical issues of these devices functioning or malfunctioning.

5. *Status — Physical*: Physical or sensory experiences such as heat, cold, tiredness, hunger or related content as a subset of the "Me Now" responses (Naaman, *et al.*, 2010).

6. *Status — Work*: Any reference to work related activity including getting things done (GTD), to do lists, employment, jobs, bosses, co-workers, colleagues and similar concepts. This incorporates both the office water cooler function of work-related commentary (Heany and McClurg, 2009; DiMicco, *et al.*, 2008) and the use of Twitter as status update for coworkers (Zhao and Rosson, 2009).

7. *Status — Automated*: Status update triggered automatically by a third party application such as Foursquare mayorship announcements, Live.fm (Honeycutt and Herring, 2009).

8. *Status — Activity*: Non-work activities includes any verb-based update describing an activity in progress including the use of instant messenger chat or IRC action protocols (/me or *undertakes activity*) excluding those actions captured elsewhere in the status or *Conversational — Action categories*. This catch-all integrates the Anecdote Me, Me Now and Anecdote Other categories from Naaman, *et al.* (2010) with the self-experience aspects from Honeycutt and Herring (2009).

Pass along

Twitter authors have a level of trust, credibility and perceived expertise in their role as a content provider for their followers' Twitter streams, and the inclusion of another user's message, URL or other content counts as an act of endorsement by the tweet author (Java, *et al.*, 2007; Pear Analytics, 2009).

Table 5: Pass along content.			
Category	N	Percentage	Exemplar
<i>Pass along</i>			
1. RT	48	2	L4D Survivors in Rockband2 singing L7 Pretend We're Dead. http://is.gd/BsVE (HT to @USERNAME). It's seriously amazing.
2. UGC	122	4	http://twitpic.com/2o1c1 — Bus Slogan Generator Time — http://is.gd/hU2Q .
3. Endorsement	108	4	I'm looking myself up on Publish or Perish (http://rurl.org/iw4) to find a reference to a paper that cited me because I want to cite them.

Three categories of tweet exist in this area — retweeting other user comments, passing along the author’s own materials and the implicit endorsement of a link by publication in the Twitter timeline.

1. *Pass along — ReTweet*: Any retweet as recognized by boyd, *et al.* (2010) including “RT: @’, ‘retweeting @’, ‘retweet @’, ‘(via @)’, ‘RT (via @)’, ‘thx @’, ‘HT @’, ‘r @’” or other acknowledgement of the original source tweet (Java, *et al.*, 2007; Pear Analytics, 2009; Naaman, *et al.*, 2010).

2. *Pass along — UGC*: Any full-length or shortened URL which can be identified as the user’s blog, Twitter photo hosting or service such as Ping.fm. This category incorporates Pear Analytics (2009) and Naaman, *et al.*’s (2010) “Self Promotion” with Honeycutt and Herring’s (2009) announce/advertise cluster to recognize Twitter as a distribution channel for personally created content.

3. *Pass along — Endorsement*: Any other content containing full length or shortened URLs which were not captured in *Conversational — Referral, Conversational — Query, Pass along — ReTweet* or *Pass along — UGC*. This has been described by Zhao and Rosson (2009) as a “people based RSS feed” whereby Twitter users act as editors and filters of content for their follows by the selective posting of content to their Twitter stream (Zhao and Rosson, 2009; Naaman, *et al.*, 2010; Honeycutt and Herring, 2009; boyd, *et al.*, 2010).

News

News tweets incorporate coverage of mainstream media issues, events, liveblog coverage, social media news, and other identifiable news content based on the use of event #hashtags.

Category	N	Percentage	Exemplar
<i>News</i>			
1. Headlines	—	—	—
2. Sport	—	—	—
3. Event	13	0	Between NASA’s satellite and autoanalysis of imagery, and Google Map data, scientific proof where there’s smoke, there are fires #bcc2
4. Weather	—	—	—

This category systematically excludes retweet or rebroadcasted news and excludes identifiable user generated content as these have been classified under the Pass along category.

1. *News — Headlines*: Tweets which resemble mainstream media (print, radio, TV) news coverage of breaking events such as the Hudson River Plane crash (Krums, 2009), or personal eyewitness accounts of live events (Fahmi, 2009; Lariscy, *et al.*, 2009; Gay, *et al.*, 2009), breaking news stories (Pear Analytics, 2009; Java, *et al.*, 2007).

2. *News — Sport*: Tweets which contain identifiable results of sporting events including live scores, on-field/in-game events, or announcements of event outcomes (Java, *et al.*, 2007; Pear Analytics, 2009).

3. *News — Live Event coverage*: Any tweet which represents the live discussion of an identified or identifiable event through the use of a hashtag (#INSM09), live opinion and discussion of a television show (#qanda) or other opinionated commentary on a live broadcast (Lariscy, *et al.*, 2009; Gay, *et al.*, 2009).

4. *News — Weather*: Any tweet reporting temperature without any accompanying commentary (Honeycutt and Herring, 2009). These should be fairly rare, although they may be represented within the *Status — Automated* category once thermostats start tweeting for themselves.

Phatic communications

Phatic communications represent the connected presence between members of a social network through sheer existence of a tweet rather than any specific content (Miller, 2008). The communications cover the use of tweet to broadcast opinion, statements, actions and fourth wall

breaking messages about Twitter, the timeline or the author's Twitter account.

Table 7: Phatic communications.			
Category	N	Percentage	Exemplar
<i>Phatic</i>			
1. Greeting	17	1	Good morning Twitterverse. How's the world outside?
2. Fourth wall	49	2	Note to self: Just because you're carrying tiny vials of hypercaffeine is no reason to start calculating remote delivery systems for them.
3. Broadcast	140	5	Diplomacy is the art of saying "Nice doggy" until you find a big enough rock. Captaincy is the timely provision of large enough rocks.
4. Unclassifiable	7	0	AAAAAAAAAAAAAAAAARGH

Broadcast communications are distinguished from the status updates by their third person nature, and/or use of the platform for undirected (unidirectional) opinion stating (rather than conversations, news, or status updates). These messages may be perceived as "pointless babble" to outsiders to the broadcaster's social network (Pear Analytics, 2009) or simple daily chatter (Java, *et al.*, 2007).

1. *Phatic — Greetings*: Generic statements of time, place and greetings to the broader community to create a sense of telepresence (Hohl, 2009) and virtual community (Miller, 2008) by addressing followers indirectly (Honeycutt and Herring, 2009; Naaman, *et al.*, 2010). Excludes simple statements of time (*Status — Temporal*) or direct remarks to other Twitter users (*Conversational — Response*).

2. *Phatic — Fourth Wall*: Text equivalent of comments made directly to camera such as "Dear [Brand/Person]", and the Honeycutt and Herring (2009) "information for self category" through "Note to self", "FYI" or "Just for the record" "thought bubble" style comments. This category also incorporates meta-commentary tweets which discuss Twitter, Twittering, and the Twitter account itself.

3. *Phatic — Broadcast*: Undirected statements which allow for opinion, statements and random thoughts to be sent to the author's followers (Zhao and Rosson, 2009; Honeycutt and Herring, 2009; Naaman, *et al.*, 2010).

4. *Phatic — Unclassifiable*: Any undecipherable tweet due to errors, half-posted sentences, garbled text or genuine cat-on-the-keyboard input (Honeycutt and Herring, 2009; Miller, 2008).

Spam

Spam is included in the framework to classify junk traffic, unsolicited automated posts, and other tweets generated without user consent due to malware or unethical sites (Pear Analysis, 2009).

Table 8: Spam.			
Category	N	Percentage	Exemplar
Spam	—	—	—

The author admits that their level of caution when it comes to avoiding Mafia Family Twitter games, URLs in money-making direct messages, or any shortened URL from an account with 0 followers/1 tweet/n+1 following means this category is currently empty.

Results, limitations and conclusions

The paper presents a proof of concept build of a multi-dimensional Twitter content classification scheme that can classify individual timelines into six broad categories, and which provides the option to then further refine the content classification into one of 23 subcategories. The paper draws together existing classification frameworks of (REF) into a larger framework. However, this framework was designed and tested on an individual timeline, and as such, the results from this study cannot be generalized to a broader population. With that in mind, further research using the timeline on a larger sample of timelines is needed to test the reliability of the categories across different Twitter use styles, and the robustness of the different category definitions. **EM**

About the author

Stephen Dann is Senior Lecturer in the School of Management, Marketing & International Business, ANU College of Business & Economics (<http://www.cbe.anu.edu.au>), Australian National University, Canberra.

E-mail: stephen [dot] dann [at] anu [dot] edu [dot] au or stephen [at] stephendann [dot] net or [at] stephendann

References

- E. Berger, 2009. "This sentence easily would fit on Twitter: Emergency physicians are learning to 'Tweet'," *Annals of Emergency Medicine*, volume 54, number 2, pp. 23A–25A. <http://dx.doi.org/10.1016/j.annemergmed.2009.06.002>
- d. boyd, S. Golder, and G. Lotan, 2010. "Tweet, tweet, retweet: Conversational aspects of retweeting on Twitter," *Proceedings of HICSS-43 IEEE* (6 January), at <http://www.danah.org/papers/TweetTweetRetweet.pdf>, accessed 23 November 2010.
- K. Cahill, 2009. "Building a virtual branch at Vancouver Public Library using Web 2.0 tools," *Program: Electronic library & information systems*, volume 43, number 2, pp. 140–155.
- K. Cetina, 2009. "What is a pipe? Obama and the sociological imagination," *Theory, Culture & Society*, volume 26, number 5, pp. 129–140. <http://dx.doi.org/10.1177/0263276409106354>
- K. Crawford, 2009. "Following you: Disciplines of listening in social media," *Continuum*, volume 23, number 4, pp. 525–535. <http://dx.doi.org/10.1080/10304310903003270>
- C. Cuddy, 2009. "Twittering in health sciences libraries," *Journal of Electronic Resources in Medical Libraries*, volume 6, number 2, pp. 169–173. <http://dx.doi.org/10.1080/15424060902932284>
- J. DiMicco, D. Millen, W. Geyer, C. Dugan, B. Brownholtz, and M. Muller, 2008. "Motivations for social networking at work," *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work (CSCW '08)*, pp. 711–720.
- E. Dudley, 2009. "Editorial: Lines of communication," *Journal of Librarianship and Information Science*, volume 41, number 3, pp. 131–134. <http://dx.doi.org/10.1177/0961000609337093>
- J. Ettama, 2009. "New media and new mechanisms of public accountability," *Journalism*, volume 10, number 3, pp. 319–321. <http://dx.doi.org/10.1177/1464884909102591>
- N. Eyrich, M. Padman, and K. Sweetser, 2009. "PR practitioners' use of social media tools and communication technology," *Public Relations Review*, volume 34, number 4, pp. 412–414. <http://dx.doi.org/10.1016/j.pubrev.2008.09.010>
- W. Fahmi, 2009. "Bloggers' street movement and the right to the city: (Re)claiming Cairo's real and virtual 'spaces of freedom'," *Environment and Urbanization*, volume 21, number 1, pp. 89–107. <http://dx.doi.org/10.1177/0956247809103006>
- R. Galer-Unti, 2009. "Guerilla advocacy: Using aggressive marketing techniques for health policy change," *Health Promotion Practice*, volume 10, number 3, pp. 325–327. <http://dx.doi.org/10.1177/1524839909334513>
- P. Gay, P. Plait, J. Raddick, F. Cain, and E. Lakdawalla, 2009. "Live casting: Bringing astronomy to the masses in real time," *CAP Journal*, number 6, pp. 26–29, at

http://www.capijournal.org/issues/06/06_26.pdf, accessed 23 November 2010.

M. Heany and S. McClurg, 2009. "Social networks and American politics: Introduction to the special issue," *American Politics Research*, volume 37, number 5, pp. 727–741. <http://dx.doi.org/10.1177/1532673X09337771>

S. Henneburg, M. Scammell, and N. O'Shaughnessy, 2009. "Political marketing management and theories of democracy," *Marketing Theory*, volume 9, number 2, pp. 165–188. <http://dx.doi.org/10.1177/1470593109103060>

M. Hohl, 2009. "Beyond the screen: Visualizing visits to a website as an experience in physical space," *Visual Communication*, volume 8, number 3, pp. 273–284. <http://dx.doi.org/10.1177/1470357209106469>

C. Honeycutt and S. Herring, 2009. "Beyond microblogging: Conversation and collaboration via Twitter," *Proceedings of the Forty-Second Hawai'i International Conference on System Sciences (HICSS-42)*, pp. 1–10, <http://ella.slis.indiana.edu/~herring/honeycutt.herring.2009.pdf>, accessed 23 November 2010.

Infosecurity, 2009. "Twitter worm steals user details" (1 June), p. 6, and at <http://www.infosecurity-us.com/view/1990/twitter-worm-steals-user-details/>, accessed 23 November 2010.

B. 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*, volume 60, number 11, pp. 2,169–2,188, and at http://ist.psu.edu/faculty_pages/jjansen/academic/jansen_twitter_electronic_word_of_mouth.pdf, accessed 23 November 2010.

A. Java, X. Song, T. Finin, and B. Tseng, 2007. "Why we Twitter: Understanding microblogging usage and communities," *Proceedings of the Ninth WEBKDD and First SNA-KDD Workshop on Web Mining and Social Network Analysis*, pp. 56–65.

B. Krishnamurthy, P. Gill, and M. Arlitt, 2008 "A few chirps about Twitter," *Proceedings of the First Workshop on Online Social Networks*, pp. 19–24.

R. Lariscy, E. Avery, K. Sweetser, and P. Howes, 2009. "An examination of the role of online social media in journalists' source mix," *Public Relations Review* volume 35, number 3, pp. 314–316. <http://dx.doi.org/10.1016/j.pubrev.2009.05.008>

A. Macintosh, 2009. "The emergence of digital governance," *Significance*, volume 5, number 4, pp. 176–178. <http://dx.doi.org/10.1111/j.1740-9713.2008.00325.x>

K. Makice, 2009. "Phatics and the design of community," *Proceedings of the 27th International Conference on Human Factors in Computing Systems (Boston)*, pp. 3,133–3,136.

V. Miller, 2008. "New media, networking and phatic culture," *Convergence*, volume 14, number 4, pp. 387–400.

M. Naaman, J. Boase, and C.--H. Lai, 2010. "Is it really about me? Message content in social awareness streams," *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work*, pp. 189–192.

G. Parslow, 2009. "Commentary: Twitter for educational networking," *Biochemistry and Molecular Biology Education*, volume 37, number 4, pp. 255–256. <http://dx.doi.org/10.1002/bmb.20307>

Pear Analytics, 2009. "Twitter study" (August), at <http://www.pearanalytics.com/wp-content/uploads/2009/08/Twitter-Study-August-2009.pdf>, accessed 23 November 2010.

R. Power and D. Forte, 2008. War & peace in cyberspace: Don't twitter away your organisation's secrets," *Computer Fraud & Security*, volume 2008, number 8, pp. 18–20.

H. Steiner, 2009. "Reference utility of social networking sites: Options and functionality," *Library Hi Tech News*, volume 26, numbers 5–6, pp. 4–6.

H. Sujathan, 2009. "Archive tweets to pdf + offline browsing," *SourceForge*, at <http://sourceforge.net/projects/twitter-to-pdf/>, accessed 23 November 2010.

D. Thussu, 2009. "Turning terrorism into a soap opera," *British Journalism Review*, volume 20, number 1, pp. 13–18. <http://dx.doi.org/10.1177/0956474809104198>

D. Zhao and M. Rosson, 2009. "How and why people Twitter: The role that micro-blogging Plays in informal communication at work," *Proceedings of the ACM 2009 International Conference on Supporting Group Work*, pp. 243–252.

Editorial history

Received 3 November 2009; accepted 21 October 2010.



"Twitter content classification" by Dr. Stephen Dann is licensed under a [Creative Commons Attribution-ShareAlike 3.0 Unported License](https://creativecommons.org/licenses/by-sa/3.0/).

Twitter content classification
by Stephen Dann.

First Monday, Volume 15, Number 12 - 6 December 2010

[http://firstmonday.org/ojs/index.php/fm/rt/printerFriendly/2745/2681,](http://firstmonday.org/ojs/index.php/fm/rt/printerFriendly/2745/2681)