Tracking comet ISON through the twittersphere: visualizing science communication in social media

Citation:

DOI: [10.4018/IJVCSN.2015100104](http://10.4018/IJVCSN.2015100104)

© 2015, IGI Global

Reproduced with permission of the copyright owner.

Available from Deakin Research Online:

[http://hdl.handle.net/10536/DRO/DU:30083762](http://hdl.handle.net/10536/DRO/DU:30083762)
Table of Contents

Research Articles

1 Impact of Social Media towards Improving Productivity at AADC
   Zoheir Ezziane, Al Ain Women’s College, Higher Colleges of Technology, Al Ain, UAE
   Abdurahman Abdulla Al Kaabi, Al Ain Women’s College, Higher Colleges of Technology, Al Ain, UAE

23 Pattern-Based Requirements Engineering of New Online Social Networks
   Sarah Bouraga, University of Namur, Namur, Belgium
   Ivan Jureta, University of Namur, Namur, Belgium
   Stéphane Faulkner, University of Namur, Namur, Belgium

45 Trends in Social Media Usage: An Investigation of its Growth in the Arab World
   Sulaiman Ibrahim Alreyaeec, Al-Imam Mohammed Ibn Saud Islamic University, Riyadh, Saudi Arabia
   Aquil Ahmed, Aljouf University, Sakaka, Saudi Arabia

57 Tracking Comet ISON through the Twittersphere: Visualizing Science Communication in Social Media
   Stuart Palmer, Deakin University, Waurn Ponds, Australia
Tracking Comet ISON through the Twittersphere: Visualizing Science Communication in Social Media

Stuart Palmer, Deakin University, Waurn Ponds, Australia

ABSTRACT

People are increasingly turning to the Internet and social media for science-related information and communication. Online social media systems are a convenient source of timely and rapid information, including science-related information; however it is argued that the emergence of social media systems is a fundamentally new development in science communication. A distinguishing characteristic of social media systems is their support for 'democratic,' user-generated content, and for the public debate and contestation of scientific ideas. The emergence of the online environment as an important source of science information is still comparatively new, as is research to understand how people find and use science information in this setting. This study investigates how information about comet ISON was communicated on the Twitter social media platform. Tweet time sequence visualization and network visualization were found to offer complementary insights into the Twitter data, and to offer a productive methodological approach for future research work.

Keywords: Comet ISON, Science Communication, Social Media, Twitter, Visualization

INTRODUCTION

The Internet has revolutionized the way people communicate and access information. Internationally, large numbers of people now get their news and other information from online sources (Anderson, Brossard, & Scheufele, 2010). In particular, in the USA, a number of surveys have indicated that the Internet is a very important source of public information about science (Sugimoto & Thelwall, 2013; Wilcox, 2012); with it reported as being the first place that most people would search to find out more about a specific science issue or question that they were interested in (Anderson et al., 2010; de Semir, 2010; Segev & Baram-Tsabari, 2012). Further, social media communications systems that operate via the Internet are becoming increasingly important as sources of science information for both the general public and science communication professionals (Pinholster & Ham, 2013; Runge et al., 2013; Veltri, 2013). Online social media systems may be a convenient source of timely and rapid information, including science-related informa-
tion (Shan et al., 2014); however, it is argued that the emergence of social media systems is a fundamentally new development in science communication.

Social media systems include a wide range of generic and specific applications, including weblogs (blogs), Twitter, Facebook, Tumblr, Pinterest and others (Bik & Goldstein, 2013). The most fundamental aspects of these systems are that they support the rapid creation and sharing of user-generated content online. The emergence of these systems has meant that the Internet and World Wide Web (the web) have moved from being an online repository of knowledge to an environment where users of all kinds can interact with, and generate their own, science-related content (Watermeyer, 2010). It is suggested that social media have created a democratic online discussion environment (Himelboim, McCreery, & Smith, 2013), and that they offer the possibility of bypassing traditional ways of doing science communication (de Semir, 2010). They provide a space where interest groups communicate directly to wider audiences, unmediated by conventional information gatekeepers such as journalists, but often using similar communication formats to those previously used by ‘official’ news and information sources (Fahy & Nisbet, 2011). The new role that non-specialists in science and science communication have taken on has created a new conceptualization of the public understanding of science (Watermeyer, 2010), and changed how, and indeed what, science issues are communicated (Anderson et al., 2010).

There is evidence that compared to traditional media channels such as television, the proportion of science-related content in social media channels is significantly higher (Sugimoto & Thelwall, 2013). It is also suggested that social media provide important channels for both interpersonal influence about, and the propagation of, popular topics in science (Segev & Baram-Tsabari, 2012). All in all, social media offer a “spectacular opportunity” for (Pinholster & Ham, 2013), and are now an “integral part” of (Wilcox, 2012), the modern communication of science. In social media, “… scientists, journalists, advocates, and the people formerly known as audiences are all content contributors, each with varying knowledge, background and perspectives.” (Fahy & Nisbet, 2011, p. 782) However, the democratization of the production and communication of scientific knowledge raises questions about the reliability and validity of such knowledge (Watermeyer, 2010). The social media process can collect and propagate information very rapidly, but users may not be able to separate fact from fiction (Castillo, Mendoza, & Poblete, 2011; de Semir, 2010).

Twitter (twitter.com) is one of the most widely-used social media systems, it is a ‘microblogging’ service where users can post quick and frequent short messages (up to 140 characters in length) called ‘tweets’ (Castillo et al., 2011). Tweets can be directed specifically to other named user accounts, referred to as ‘mentioning’ those users, or broadcast generally to all ‘followers’ of the sending account (Veltri, 2013). Except for the content of tweets from protected (private) accounts, all tweets are effectively broadcast to ‘the world’ and also discoverable publicly via the search function available in the Twitter platform (Runge et al., 2013). A user can ‘retweet’ to all of their followers a tweet that they receive from another user, and the 140-character restriction on tweet length means that the use of hyperlinks to link to additional material (including images, videos and websites) is common (Himelboim et al., 2013). Tweets can be identified with a ‘hashtag’ (a keyword prepended by a ‘#’ symbol, for example ‘#nanotech’), and a Twitter search using the hashtag as a keyword will aggregate recent Tweets containing that tag (Bik & Goldstein, 2013).

Twitter facilitates the rapid transmission of information to a large number of users, and has become a popular communication platform for the dissemination of breaking news stories and information, and has found particular utility in emergency events such as fires, floods and earthquakes (Castillo et al., 2011) and international political events (Runge et al., 2013). In 2012, when rumours surfaced of an impending announcement about the discovery of the Higgs Boson,
this topic was very popular on Twitter, and Tweets tagged with ‘#HiggsRumors’ appeared in large numbers (Drake, 2012). In 2009, the Newbury Astronomical Society (NAS) in the UK encouraged the public to observe the Perseid meteor shower and share their experiences via Twitter with the hashtag ‘#Meteorwatch’. Although the NAS is a relatively small organization, it was estimated that nearly 300,000 people participated in some way (Sandu & Christensen, 2011). On Twitter, users often follow persons and organizations based on their interests, including science interests. For example, a science topic that has emerged as topical in the media in recent years is nanotechnology (Anderson et al., 2010), and Twitter has become an important source of information on nanotechnology (Runge et al., 2013; Veltri, 2013). However, the popularity of Twitter has also attracted spammers seeking to attract visitors to their websites selling products and services, and also those seeking an audience for propaganda and false claims (Castillo et al., 2011).

Some observations have been made about Twitter as a medium for science information and communication. NASA has used Twitter to build a network of amateur astronomers (twitter.com/asteroidwatch) to track asteroids and other near-Earth objects (NEOs) that might potentially impact the Earth sometime in the future, and, interestingly, to help debunk false information surrounding NEOs (Revkin, 2010). In the context of investigating how the topic of nanotechnology is communicated via the Twitter platform, it was observed that the most active users were individuals rather than official organizations or their representatives, however the latter generally have more followers than the former, and a low observed proportion of mentions and retweets suggests that Twitter is generally being used to broadcast information about nanotechnology, rather than supporting a social media conversation about it (Veltri, 2013). The enormous range of users, with their consequent wide range of interests and perspectives, gives users of social media, including Twitter, the ability to follow, and participate in, only those online fora that reflect their own pre-existing world views, potentially giving rise to an online ‘echo chamber’ of self-reinforcing ideas (Fahy & Nisbet, 2011).

The emergence of the online environment as an important source of science information is still comparatively new, as is research to understand how people find and use science information in this setting (Anderson et al., 2010). Likewise, the associated development of online social media systems creates a need to understand how they can be used effectively for science communication and education (Mahrt & Puschmann, 2014). A distinguishing, and perhaps the central, characteristic of social media systems is their support for ‘democratic’, user-generated content, and for the public debate and contestation of scientific ideas (Jaspal, Nerlich, & Kotevyko, 2013). The enormous general popularity and growth of social media means that research is required to understand the types of scientific discourse that are occurring in these settings, and there has been only limited work published in this area to date (Runge et al., 2013). Certainly, for those employing social media in their education and outreach efforts, it is essential to monitor and evaluate the effectiveness of their science communication activities in these arenas (Sandu & Christensen, 2011).

**OBJECTIVE**

In between the breaking story (earthquakes, the Higgs Boson, etc.) and an on-going pervasive theme (i.e., nanotechnology), there are science-related stories that have an extended but finite duration – one such recent story was comet ISON. Discovered in Russia in September 2012 by the International Scientific Optical Network (from which it draws its common name), comet ISON was first observed very early in its approach to the inner solar system, allowing for an extended period of media reporting and public interest, including reports that it would be the ‘Comet of
the Century’ (Chang, 2013). Social media, in this case Facebook, were mobilized to facilitate an international community of amateur astronomers to track comet ISON (Yanamandra-Fisher & Warner, 2013).

This study investigates how information about comet ISON was communicated on social media, in this case Twitter, over a nine-month period, including a significant part of the period between its initial discovery and close approach to the Earth and Sun, the period that it was visible from the Earth, and the period immediately after the point at which it was officially declared to have disintegrated. More than 96,000 tweets were collected, and while some basic text analysis was performed on the Twitter content to broadly classify the types of information being circulated, the principal focus of this investigation was the visualization of the ‘bulk’ quantitative characteristics of the comet ISON-related Twitter data. The purpose of this approach was two-fold—firstly to identify key temporal and network features of the Twitter data for closer investigation; and secondly to test and develop appropriate data analysis and visualization methods to support this form of investigation. This study describes in detail the analysis methods developed, and discusses the large-scale and fine-scale characteristics of the comet ISON Twitter data thus observed. It provides both findings specific to this case study, and offers methodological approaches for those interested in similar future research work.

**METHODS**

A ruling was obtained from the relevant institutional human research ethics committee that the collection and use of publically accessible Twitter data did not require formal ethics approval for research purposes. In the work presented here, popular public Twitter accounts are identified by name where relevant, but no accounts of private individuals are identified unless they expressly agreed to be named. A public application programming interface (API) provided by the Twitter platform allows data to be directly collected from the system (Miller, 2011; Veltri, 2013). However, the Twitter system quickly archives data, such that there is limit how far back in time a search or other data request will reach (Bik & Goldstein, 2013), and there may be other limits applied to the results of popular searches that are not predictable. By accessing the Twitter API, the NCapture program (QSR International, 2012a) is able to capture publicly available Twitter data at that point in time, including from a specific user account, arising from a hashtag search, arising from a keyword search, etc. Over the period 15 April 2013 to 10 January 2014, tweets containing both of the keywords ‘comet’ and ‘ison’ were collected every 2-3 days, and with increasing frequency during periods when large number of results were being returned by the Twitter API. It is acknowledged that the Twitter data collected does not represent all Comet ISON related tweets arising from the Twitter API. It is acknowledged that the Twitter data collected does not represent all tweets mentioning Comet ISON—the limitations on the depth of the publicly accessible Twitter data means that there are gaps in the data set, and there are likely to be other tweets related to Comet ISON not captured by the basic keyword search strategy used. The NVIVO program (QSR International, 2012b) was used to convert the captured Twitter data into Microsoft Excel (Microsoft, 2010) spreadsheets for further processing and analysis.

Visualizing the frequency of tweets related to a theme over time can provide useful insights into the ways that users are interacting with that theme via social media (Letierce, Passant, Breslin, & Decker, 2010; Miller, 2011; Shan et al., 2014). Daily collected tweet totals were compiled. Given that the data do not represent all Comet ISON Twitter content, the median daily tweet totals for each week of the period under investigation were plotted to characterize the frequency of tweets related to Comet ISON over time.

Network visualization of Twitter data can be a useful method to reveal the communication structures embodied in the data (Himelboim et al., 2013; Letierce et al., 2010; Miller, 2011). The
spreadsheet Twitter data were also exported in comma separated values (CSV) format, and then imported into the Gephi program (The Gephi Consortium, 2012) for network visualization. As outlined in Figure 1, Gephi can be used to represent Twitter user accounts as ‘nodes’, and the communication path (representing one or more tweets) between two nodes as an ‘edge’. In the Twitter network diagrams used in this study, edges are presented as curved lines, the direction of tweets is clockwise around the edge, and the width of an edge is proportional to the total number of tweets recorded between the two nodes in that direction. Twitter data will contain undirected tweets – those from a user not mentioning any other account, hence implicitly directed to the followers of the user, but also to the word at large.

Because undirected tweets may represent a significant proportion of all tweets (Veltri, 2013), a meaningful way must be found for dealing with them in analyses (Honeycutt & Herring, 2009). Not being explicitly directed to a named user, undirected tweets cannot be formed into a network using the schema in Figure 1. In the analyses presented here, all undirected tweets are allocated as directed to a notional Twitter user identified as @the_world. While there is a single topological arrangement of the data for a given network, it can be visualized in many ways. The Gephi program provides a range of algorithms for laying out networks. The Fruchterman-Reingold (F-R) layout algorithm (Fruchterman & Reingold, 1991) has a number of desirable characteristics (good node distribution, minimization of edge crossings, uniform edge lengths, reflection of inherent symmetries, etc.), and was chosen for use here.

Bulk properties of Twitter data can be informative, such as the proportions of retweets and mentions (Veltri, 2013), the proportion of tweets containing URLs (Himelboim et al., 2013) and the proportions of tweets categorized by theme/content (Runge et al., 2013; Shan et al., 2014). The captured Twitter data indicate whether a post is a tweet or a retweet – the proportions of each were calculated. The captured Twitter data indicate whether a tweet mentions any other user; if yes it is a directed or mention tweet; if not it is an undirected tweet – the proportions of each were calculated. The captured Twitter data indicate whether a tweet contains a URL or not – the proportions of each were calculated. Science communication-related Twitter data is known to typically contain both anti-scientific and spam content (Castillo et al., 2011). Using an initial set of keywords based on conspiracy, paranormal, alien, apocalyptic and spam concepts, searches were performed to identify tweets with anti-scientific and spam content, as well as to identify additional new search keywords. This process was repeated until no new results were found. Based on this, the proportions of tweets containing anti-scientific and spam content were...
estimated. The remaining tweets contained a small number of tweets in languages other than English. Four groups of 500 tweets were selected from the remaining tweets at approximately equidistant points in the tweet time sequence. These 2000 tweets were inspected manually to identify any non-English content, and based on this the proportion of tweets in languages other than English was estimated. For science communication and education, evaluation of reach and impact is essential (Sandu & Christensen, 2011). Measures of reach and impact can be derived from the account statistics and activity of Twitter users (Veltri, 2013). The Twitter data were inspected to identify the most frequently tweeting users, the most retweeted users and the most mentioned users. In addition, for every uniquely identified user in the data set, their total number of tweets multiplied by their average number of followers, during the period under investigation, was computed as an empirical measure of potential Twitter influence.

RESULTS

For the period under investigation a total of 96084 tweets were collected. These originated from 51803 unique Twitter user accounts, and connected 54542 unique user accounts (nodes) via 72998 unique pathways (edges). Figure 2 presents the median daily total tweet count recorded for each week during the period under investigation – week 1 commenced on 15 April 2013. Points of interest are labelled and discussed further below. Based on the schema presented in Figure 1 and using the F-R layout algorithm, Figure 3 presents a visualization of the Twitter network data recorded during the period under investigation. Points of interest are labelled are discussed further below. Figure 4 and Figure 5 present selected details from Figure 3. Points of interest are labelled are discussed further below. Table 1 presents summary statistics of the Comet ISON Twitter data collected. Table 2 presents the ‘Top three’ Twitter users for various measures of impact.

DISCUSSION

Assuming that a complete Twitter data set was not obtained, and given that it is based on median daily tweet count by week, Figure 2 does not attempt to represent all Twitter activity related to Comet ISON, rather it provides a measure of the relative intensity of tweets over time. For most of the first half of the time period covered, the median daily level of activity for each week.
Figure 3. Overall comet ISON Twitter network visualization

Figure 4. Selected comet ISON Twitter network visualization – lower left region
recorded was around 75-200 tweets per day. However, there are a number of exceptions when this level rose to approximately 400 tweets per day, such as for the week 6 (labelled as A) and week 9 (labelled as B). An inspection of the sequence of tweets recorded during week 6 revealed that a Twitter account popular (nearly 700,000 followers) for tweeting interesting images posted

---

Table 1. Summary statistics of Comet ISON Twitter data collected

<table>
<thead>
<tr>
<th>Variable</th>
<th>Count</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total tweets</td>
<td>96084</td>
<td>100%</td>
</tr>
<tr>
<td>Tweets</td>
<td>53400</td>
<td>55.6%</td>
</tr>
<tr>
<td>Retweets</td>
<td>42684</td>
<td>44.4%</td>
</tr>
<tr>
<td>Directed/Mention</td>
<td>58683</td>
<td>61.1%</td>
</tr>
<tr>
<td>Undirected</td>
<td>37401</td>
<td>38.9%</td>
</tr>
<tr>
<td>Including URL</td>
<td>54670</td>
<td>56.9%</td>
</tr>
<tr>
<td>No URL</td>
<td>41414</td>
<td>43.1%</td>
</tr>
<tr>
<td>General content</td>
<td>80847</td>
<td>84.1%</td>
</tr>
<tr>
<td>Anti-scientific</td>
<td>10941</td>
<td>11.4%</td>
</tr>
<tr>
<td>Spam</td>
<td>752</td>
<td>0.8%</td>
</tr>
<tr>
<td>Non-English content</td>
<td>3544</td>
<td>3.7%</td>
</tr>
</tbody>
</table>
an image about comet ISON and explicitly called for that message to be shared. Subsequently, many users did retweet that message, including a number of other popular Twitter accounts. One of those subsequent sharers was the Twitter account @Earth_Pics (1.1 million followers). This secondary sharing of the original message stimulated a further large cascade of 2750 retweets, and is the source of the plume feature in Figure 3 labelled A, and is discussed further below. Similarly, an inspection of the sequence of tweets recorded during week 9 revealed that another popular Twitter account @TheWeirdWorld (1.2 million followers) had posted a reworking of the original comet ISON image, again including a request for the image to be shared, resulting in 2803 retweets, which is the source of the plume feature in Figure 3 labelled B.

Following the discovery of comet ISON in November 2012 there was on-going low-level general media coverage through the period represented by the first half of Figure 2, during which a series of space telescopes and then large ground-based telescopes imaged the comet as it approached the Sun. In early October (week 25 / point C in Figure 2) comet ISON passed close to Mars and became visible to small ground-based telescopes, sparking renewed media and public interest. At the end of October (week 28 / point D in Figure 2) comet ISON was visible from earth in some locations using binoculars. In mid-November comet ISON was visible to the naked eye in locations with low light pollution. During week 31 (point E in Figure 2) searches using the Twitter public API returned very shallow data sets, with 100 tweets or less. Presumably, this was a time of very high public interest in comet ISON, including on social media channels, and this apparent ‘throttling’ of Twitter search results continued for a number of days, resulting in a significantly incomplete tweet collection during week 31 and the very low tweet count at point E in Figure 2. In late November, during week 33, comet ISON reached its closest approach to the Sun (perihelion), and an increase in tweet collection was observed at point F in Figure 2. In early December, comet ISON had faded dramatically and was eventually declared to have disintegrated – this time is represented by point G in Figure 2, and it can be seen that Twitter activity related to comet ISON continued to decline relatively sharply after that point.

Figure 3 is the overall network visualization showing the large-scale structure of the comet ISON Twitter data. The F-R layout algorithm produces a complicated network rich in features. The plume-like structures represent the Twitter mentions of a single user, located at the focus of the plume, by a relatively large number of other users, whom appear as the nodes in the plume. The mentions are represented by clockwise edges connecting inward to the user at the focus. These mentions include any Tweets directed to the user at the focus point; but the majority of

| Most Prolific (Number of Tweets) | @ISONupdates | 539 | @CometISONnews | 451 | Private account | 224 |
| Most Retweeted (Number of Retweets) | @TheWeirdWorld | 2602 | @Earth_Pics | 2424 | @ScienceWTF | 2306 |
| Most Mentioned (Number of Mentions) | @YouTube | 5073 | @TheWeirdWorld | 2756 | @Earth_Pics | 2624 |
| Most Influential (Number of Tweets x Average Number of Followers) | @NatGeo | 13676637 | @mashable | 13356784 | @AsteroidWatch | 9489042 |
| No. of tweets | 3 | No. of tweets | 4 | No. of tweets | 9 |
| Av. No. followers | 4558879 | Av. No. followers | 3339196 | Av. No. followers | 1054338 |
these mentions are typically retweets of an initial tweet originating from the user at the focus point. As noted above, the plume at point B in Figure 3 represents approximately 2800 retweets of a single post from the Twitter account of @TheWeirdWorld – this feature is shown in more detail as point B in Figure 4. The largest plume in the upper right quadrant of Figure 3 is centered on the notional user @the_world, i.e., it represents the accumulation of all of the 37401 undirected tweets recorded. The symmetry-emphasizing feature of the F-R layout algorithm is apparent in Figure 3. The overall layout is approximately circular and balanced. There is a heavily linked/interconnected region in the centre; @the_world with its large plume of nearly 40% of all network connections appears on one side of the centre; and the balance of the network connections are positioned ‘opposite’ to @the_world. The nodes/users other than @the_world with the largest number of connections (including the ‘plumes’ A and B) are positioned furthest from the @the_world, near the opposite edge of the network. On the outside of the main plume associated with @the_world (at the upper-most right) is a ‘crown’ that represents those nodes with the simplest connections to @the_world – the more complex connections are made through the central region of Figure 3. Finally, there is a halo of nodes around the outer edge of the network that have no connection to the main network proper.

Figure 4 is an expanded view of the lower left region around point B in Figure 3, showing some of the fine-scale structure of the comet ISON Twitter data. The plume structure in Figure 4 labelled as I is similar to the many other plumes visible in the network visualization, including the one labelled B in Figure 3 and Figure 4. Point I represents 705 retweets of a single tweet (initially posted by the node at the focus point of I) about comet ISON (as before, including an image of a comet and an invitation to retweet) from the Twitter account of @MostSecretFacts (which claims to tweet about interesting and lesser known facts). The lower number of retweets, compared to the plume at point B, means that the retweet plume at point I does not completely encircle the focus point, as it does for point B. Point J in Figure 4 represents a group of 14 Twitter users who retweeted both the tweet from @TheWeirdWorld at point B and the tweet from @MostSecretFacts at point I – edges can be seen connecting in a clockwise direction from the group of 14 nodes to both @TheWeirdWorld and @MostSecretFacts.

Point K in Figure 4 captures a Twitter ‘conversation’ – indicated by the closed loops of edges. It represents a series of 13 tweets between 11 accounts. It was initiated in week 2 of the period shown in Figure 2 by the user/node at the upper right of the edge loops around point K. The initial tweet was directed to six other users suggesting that they arrange to observe comet ISON at a later time. This initial tweet was retweeted by three of the recipients. In week 12 the initiator sent another comet ISON-related tweet to multiple recipients, this time drawing in some new recipients. In week 25, which corresponds to the beginning of the peek Twitter activity recorded in Figure 2, one of the original recipients tweeted a link to a comet ISON-related article on a popular space science website to seven recipients, including some additional new recipients. Over the next three weeks this spawned a series of retweets and related tweets between various combinations of the users previously involved in the conversation. The sequence concludes with an undirected tweet from initiator in week 38, announcing that comet ISON had officially been declared as disintegrated. The final feature noted in Figure 4 is some of the detail of outer halo of nodes around the edge of the network, labelled as point L. Various isolated small groups of nodes (generally two to five) can be seen, connected only to each other and not to the main network. The typical form of these clusters is a single tweet referencing comet ISON sent from one user to one (or at most a few) other user(s).

Figure 5 is an expanded view of the upper central region around point H in Figure 3. The point labelled M in Figure 5 indicates the beginning of a very wide edge that connects clockwise in to @the_world (can be seen more in Figure 3). This edge represents 155 undirected (hence...
connecting to @the_world) tweets from a single user. Inspection of the Twitter data shows that these 155 tweets occurred daily for nearly the entire period from mid-April to mid-September in 2013, and were all of the identical form – reporting the current apparent visual magnitude of, constellation being traversed by, and spherical equatorial coordinates (right ascension and declination) for, comet ISON. Although publicly discoverable (as in the case of the data collection for this research project), being undirected, these tweets were essentially only appearing in the Twitter timelines of those users following the sender, which was on average 3718 accounts during the period recorded.

Inspection of Figure 3 indicates that most (though not all) of the plume-like structures in the network are composed of edges that curve clockwise inward to the plume focus point – including the one labelled N in Figure 5, and those labelled B and I in Figure 4. This configuration typically represents a relatively large number of users (represented by the plume) retweeting or mentioning a single user (the node at the focus point). Two plumes with the opposite structure can be seen above and below the point labelled O in Figure 5 – here the plumes are composed of edges that curve clockwise outwards from the focus point to the nodes that make up the plume. This counter-configuration typically represents the user at the focus undertaking a ‘bulk mailing’ exercise via directed tweets to, or mentions of, a large number of other users. Here, the plume to the lower right of point O represents 119 separate directed tweets from a user to individually named accounts. These tweets all have the same message (referring the recipient to a YouTube video purporting to link comet ISON to crop circles and other anti-science ideas) and they were all sent on a single day during week 35 of the period shown in Figure 2 – after comet ISON had been officially declared as disintegrated. The plume to the upper left of point O represents (coincidentally) 119 ‘Follow Friday’ mentions of separate accounts. It is customary, on a Friday, for Twitter users to recommend to others who they should follow by tweeting the username of their recommendation with a ‘#ff’ hashtag (Sheets, 2013). The 119 #ff mentions here were sent using 16 separate tweets posted on the Friday of week 30 of the period shown in Figure 2, with each tweet containing seven or eight account names as follow recommendations, based on those named accounts as having an interest in comet ISON.

Point P in Figure 5 (shaped somewhat like a wrapped sweet) shows the most complicated structure in the outer halo of nodes not connected to the main network. The unique cluster shape arises from what would otherwise be two separate plume-like structures – the two ‘arms’ on either side are small separate plumes of nodes with edges connecting clockwise inward to two separate focus points inside the area of the main circular body, representing separate mentions of the nodes at the focus points. What is different in this case is that there is a third larger group of nodes that connect to/mention both of the focus point nodes. While the symmetry feature of the F-R layout algorithm leads to the observed shape of this small, disconnected sub-network, it does seem somewhat unnatural. Inspection of the underlying Twitter data related to point P does indeed reveal an ‘artificial’ origin. Aside from the two nodes at focus points, all of the other 229 nodes associated with point P represent Twitter accounts with names composed from random combinations of 8-9 characters, which follow no other users and which indicate their location as one of six large UK cities. All of the 357 edges associated with Point P occur in a 2-minute period during week 32 of the period shown in Figure 2, all represent tweets not retweets, and these tweets occur at a rate of up to nine per second. The content of all of these tweets is identical – text referring to comet ISON and a mention of one or both of the Twitter accounts at the focus points of the sub-network at point P in Figure 5. It is clear that the network activity embodied in the feature labelled point P is some sort of automated Twitter ‘attack’ on the two users at the focus points; what is not clear is the purpose of this activity. The identical tweets contain no URL that would indicate a spam message, a check at random of ten of the 229 sending accounts revealed
that all were now deleted, and the two mentioned recipients are unlikely to follow nearly 180 unique accounts each mentioning them once. The final feature noted in Figure 5 is more of the detail of outer halo of nodes around the edge of the network, labelled as point Q – this is similar to point L in Figure 4.

Table 1 shows that 55.6% of the tweets collected were ‘new’ content from a user, while 44.4% were retweets of a post from someone else. The level of retweets in general Twitter data has been reported as relatively low (3%) (Boyd, Golder, & Lotan, 2010); but higher levels have been reported in science communication contexts – 7% in tweets relating to nanotechnology (Veltri, 2013), and 15%-20% at science-related conferences (Letierce et al., 2010). A low level of retweets has been taken to indicate largely one-way communication rather than conversation (Veltri, 2013); whereas higher levels of retweeting have been seen as indicators of a more active engagement and interaction in the Twitter environment (Himelboim et al., 2013; Letierce et al., 2010). Space science events have previously been reported as being popular (as measured by number of retweets) on Twitter (Pinholster & Ham, 2013), and the apparently very high proportion of retweeting observed here suggests that Twitter users were highly engaged around the topic of comet ISON. Table 1 shows that 38.9% of the tweets collected were undirected (appearing in the Twitter timeline of those users following the sender, and discoverable in searches by other users), while 61.1% were directed or otherwise specifically mentioned another user. The proportion of mentions in general Twitter data has been reported as 36% (Boyd et al., 2010). High levels of undirected tweets could be seen as one-way communication, whereas the relatively high levels of directed/mentioning tweets observed here could again be taken as an indicator of a more interactive form of Twitter communication on the topic of comet ISON.

Table 1 shows that 56.9% of the tweets contained a hyperlink to a web page, while 43.1% did not. The proportion of tweets containing hyperlinks in general Twitter data has been reported as 22% (Boyd et al., 2010). In a study of political views on Twitter, 62.9% of tweets were observed to contain a URL, and it was suggested that the restricted length of tweets (140 characters) is the reason for the relatively common practice of including hyperlinks to more extensive content hosted elsewhere on the web (Himelboim et al., 2013). These findings suggest that there is a higher level of inclusion of hyperlinks in Twitter content that is associated with an issue or theme (such as comet ISON), compared to general Twitter content.

Table 1 shows that approximately 11.4% of tweets were anti-scientific in nature, approximately 0.8% were spam and approximately 3.7% were composed in a language other than English. Based on the methods described above for identifying these categories of tweets, these figures should be considered as lower estimates, with the actual proportions likely to be somewhat higher. It is known that social media systems provide a convenient platform for those with an anti-science message to reach a large and potentially impressionable audience (Castillo et al., 2011; Mahrt & Puschmann, 2014). However, the finding that at least 11.4% of the collected comet ISON tweets contained messages relating to conspiracy, paranormal, alien, or apocalyptic themes was surprising, and is deserving of further investigation that is beyond the scope of this research project.

Spam is a common occurrence on Twitter, typically in the form of tweets designed to lure readers to a web site (Castillo et al., 2011), and one study found 8% of Twitter traffic was spam content (Grier, Thomas, Paxson, & Zhang, 2010). Here the estimated spam content of 0.8% is much lower. As observed for the proportion of retweets and mentions related to comet ISON being different to that observed in general Twitter content, it may be that the strongly themed nature of the comet ISON Twitter content is also an influence on the proportion of spam observed – an investigation of politically themed content on Twitter observed no spam at all (Himelboim et al., 2013). All 752 identified spam tweets had the same form – a combination of two or more popular/topical phrases (in this case including some mention of comet ISON) and a hyperlink.
to a website. The estimated proportion of tweets composed in a language other than English of 3.7% means that a significant number of tweets were not directly available for classification as anti-scientific or spam content (Veltri, 2013). It is possible to perform basic machine translation into English of tweet content for analysis, but that was not done in this case because tweet content analysis was not the principal purpose of this investigation.

Table 2 identifies the most prolific tweeters observed in the data set. All ‘top three’ accounts belonged to private individuals; however, the account names of the top two tweeters indicate a strong association/interest with comet ISON. Interestingly, following the demise of comet ISON, both of these account names have subsequently been changed to have a more general space science theme. Table 2 identifies the most retweeted accounts – here all three were popular Twitter accounts that post quirky facts and images, and all having large numbers of followers. Note that the top two retweeted accounts are the focus point nodes for point B and point A in Figure 3 respectively, and the third account is the focus point for the unlabelled third largest plume structure above the centre left in Figure 3. Table 2 identifies the most mentioned accounts – Twitter mentions include retweets, so it is not surprising that the results for these two categories overlap. The large number of mentions for @YouTube arise from the social media sharing functions built into the YouTube website. In addition to the ability to directly ‘share’ a link to a YouTube video on Twitter via a one button link, whenever a YouTube user ‘likes’, ‘favorites’ or adds a video to a playlist, they also have the opportunity to share this (and a link to the video in question) on Twitter. This result highlights the importance of providing functions in the online environment that make it easy for your science-related content to be shared via social media and thereby increase its reach and influence. Based on the empirical measure of their number of tweets multiplied by their average number of followers, Table 2 identifies the most influential tweeters observed in the comet ISON Twitter data set. @NatGeo, @mashable and @AsteroidWatch are the official Twitter accounts of National Geographic (one of the largest non-profit scientific and educational institutions), Mashable (a British-American technology news website) and Asteroid Watch (the public face of the NASA near earth object observation program) respectively. As shown in Table 2, while these three accounts made relatively low numbers of tweets relating to comet ISON, they all had a very large number of Twitter followers, so the potential reach of their messages was very wide. In a previous investigation of the nanotechnology themes on Twitter, it was observed that the most active users were individuals, but the users with the most followers were ‘official organizations’ (Veltri, 2013), and the same general findings were observed here.

CONCLUSION

This study investigates how information about comet ISON was communicated on the Twitter social media platform, based on visualization of the bulk quantitative properties of more than 96,000 tweets collected over 39 weeks. Analysis of the time sequence of tweets showed a general low level of Twitter activity until the point at which comet ISON became visible to small ground-based telescopes, after which the median daily level of activity for each week rose significantly. The level of Twitter activity remained relatively high until comet ISON passed its perihelion, faded dramatically and was declared to have disintegrated, after which Twitter activity declined sharply. The initial period of low-level Twitter activity was punctuated by a number of peaks caused by large numbers of retweets in response to popular Twitter accounts posting tweets related to comet ISON and specifically calling for readers to share them. A network visualization of the Twitter data using the Fruchterman-Reingold layout algorithm also revealed the presence of these peaks in activity in the form of plume-like features. This network visualization also
revealed a range of other interesting features including network loops representing Twitter ‘conversations’ taking place over time between groups of users, complementary plume structure forms representing both many-to-one and one-to-many Twitter mention events, and the unique wrapped sweet-shaped sub-network capturing an automated Twitter attack in progress. As well as providing confirmation of some of the features observed in the time sequence visualization, the network visualization revealed additional features in the Twitter data (such as the ‘attack’) that would be unlikely to have been discovered via any other means, and in this way adds a valuable tool for the analysis and characterization of Twitter data.

Bulk statistics for the observed proportion of both retweets and mentions were significantly higher than that reported for general Twitter traffic, suggesting higher levels of user engagement and interaction around the specific science-related theme of comet ISON. In line with other investigations of Twitter content, more than 50% of tweets contained a hyperlink to an external web page – a common strategy for engagement via Twitter that overcomes the 140-character limitation for postings. A relatively high level of anti-scientific content was observed. As has been observed in other studies of science information and communication on Twitter, the most prolific tweeters were individuals, but the most influential were official organizational accounts with very large numbers of followers. The most mentioned user was @YouTube, courtesy of the multiple opportunities built into the YouTube website to easily share their video content via social media, including Twitter.

People are increasingly turning to the Internet and social media, including Twitter, for science-related information and communication. Almost all of the authors cited above urge scientists, science educators and science communicators to understand and engage with social media, to both stay connected with their existing audiences and to reach new ones. Overall, it was found that Twitter users were strongly engaged in interactive communication about comet ISON, but that a significant proportion of this social media communication was of an anti-science nature. Scientists, science educators and science communicators have a vital role to play in establishing points of reference and authority for science information in the social media environment. The methodological approaches described in detail here provide a framework for those interested in future research into how Twitter is being used in science communication generally.
REFERENCES


Stuart Palmer is Associate Professor of Integrated Learning in the Faculty of Science, Engineering and Built Environment at Deakin University. He completed his undergraduate degree with distinction in electronics engineering. During nearly a decade of professional practice in consulting engineering he completed an MBA in technology management. In 1995 he joined the School of Engineering at Deakin University and lectured in the management of technology for 12 years. During that time, he was awarded the Australasian Association for Engineering Education McGraw-Hill New Engineering Educator Award, completed his doctoral studies in engineering management education, and completed a Graduate Certificate in Higher Education. In 2011 he was awarded an Australian Learning and Teaching Council Citation for Outstanding Contribution to Student Learning. More recently he has taken a role in the Faculty of Science, Engineering and Built Environment at Deakin University. He maintains a strong interest, and active involvement, in engineering education. His research interests include frequency domain image analysis, social network data analysis and visualisation, and the effective use of online technologies in learning and teaching.