
Published online at: http://paee.dps.uminho.pt/past-events/PAEE2016/PAEE_ALE_2016_PROCEEDINGS.pdf#page=63

This is the published version.

©2016, The Author

Reproduced by Deakin University with the kind permission of the copyright owner and the PAEE conference organisers.

Available from Deakin Research Online:

http://hdl.handle.net/10536/DRO/DU:30084976
The Representation of Engineering Education as a Social Media Topic on Twitter

Stuart Palmer

Faculty of Science, Engineering and Built Environment, Deakin University, Geelong, Australia
Email: stuart.palmer@deakin.edu.au

Abstract

Online social media systems have created new ways for individuals to communicate, share information and interact with a wide audience. For organisations, social media provide new avenues for communication and collaboration with their stakeholders. The potential value of social media tools to assist in the successful communication and marketing inside and outside of engineering organisations has been identified. In the context of engineering education, the potential of social media to open new modes of communication, interaction and experimentation between students and teachers has also been identified, and a limited number of examples can be found documented in the literature. One of the most widely-used social media tools is the ‘microblogging’ service Twitter. This research presents an analysis of nearly 19,000 tweets relating to ‘engineering education’ collected over a period of almost a year. Social network analysis is used to visualise the Twitter data. The Twitter social media communication is examined to identify who is active on this topic, who is influential, and what is the structure of the online conversations relating to engineering education. This work provides insights regarding how engineering education is currently represented in social media internationally, and offers a methodology to those interested in related future research.

Keywords: Engineering Education; Social Media; Twitter; Social Network Analysis.

1 Introduction

In the context of engineering education, the potential of social media to open new modes of communication, interaction and experimentation between students and teachers has been identified (Kamthan, 2010). Examples in the literature include: social media tools being used to link students with practicing industry professionals (Morgado et al., 2012); the use of Twitter to engage a large information literacy class (Morrow, 2010); the use of Twitter by engineering students on work integrated learning placements (Paku & Lay, 2011); the use of Twitter to send remote commands to a numerical computing environment (Judd & Graves, 2012); and students collaborating at two universities autonomously adopting Facebook for group communications (Charlton, Devlin, Marshall, & Drummond, 2010). Research on the use of social media by higher education institutions is still limited, and evaluation of the impact of social media activities is not straightforward, as few benchmarks exist. One approach to evaluation is social network analysis (SNA). The network data inherently created by social media tools represent the connections between participants as they interact, and can be used to make visible the social processes at play, to identify strategically important components and participants in the social network, and to show the development of the communication links over time (Smith et al., 2009).

One of the most widely-used social media tools is Twitter (twitter.com) (Bik & Goldstein, 2013; Himelboim, McCreey, & Smith, 2013; Naaman, Becker, & Gravano, 2011; Xu, Ru, Xiang, & Yang, 2011). Twitter is a popular ‘microblogging’ service where users can post quick and frequent short messages (up to 140 characters) called ‘tweets’, which may contain links to other online material such as photos and websites, to their ‘followers’ who have subscribed to their Twitter account. Tweets can be tagged with a searchable ‘hashtag’, and a user can ‘retweet’ to all of their followers a tweet that they receive from another user. Tweets can be directed specifically to other named user accounts, or broadcast generally to all followers of the sending account. Except for the content of tweets from protected (private) accounts, all tweets are effectively broadcast to the world and are publicly discoverable via a search. A growing number of academic units involved in engineering education internationally now advertise a link to a Twitter account on their Internet home page and/or use Twitter as part of their communication and marketing strategy. This research presents an analysis of nearly 19,000 tweets relating to ‘engineering education’ collected over a period of almost a year. Rather than describing educational
uses of social media, SNA is used to visualise the Twitter data. The Twitter social media communication is examined to identify who is active on this topic, who is influential, and what is the structure of the online conversations relating to engineering education.

2 Methodology

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. The public application programming interface (API) provided by the Twitter platform allows data to be directly collected from the system (Miller, 2011). However, the Twitter system quickly archives data, such that there is a 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, 2013a) is able to capture publicly available Twitter data at that point in time, including that arising from a keyword search. Over the period 30 March 2015 to 4 March 2016, tweets containing both of the keywords ‘engineering’ and ‘education’ were collected weekly. It is acknowledged that the Twitter data collected do not represent all tweets mentioning engineering education – the limitation 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 ‘engineering education’ not captured by the basic keyword search strategy used. The NVivo program (QSR International, 2013b) was used to convert the captured Twitter data into Microsoft Excel (Microsoft, 2010) spreadsheets for further processing and analysis.

Bulk properties of Twitter data can be informative, such as the proportions of retweets and mentions, most active users, most followed users, etc. (Veltri, 2013). 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. 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.

Network visualisation of Twitter data can be a useful method to reveal the communication structures embodied in the data (Himelboim et al., 2013; 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 visualisation. 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. The size of a node is proportional to the total number of edges.
connecting to it (referred to as the node ‘degree’). 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, 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 automatically 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 @undirected. While there is a single topological arrangement of the data for a given network, it can be visualised (laid out) 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.

3 Results and Discussion

For the period under investigation a total of 18973 tweets were collected. These originated from 7785 unique Twitter user accounts, and connected 8975 unique user accounts (nodes) via 13376 unique pathways (edges). Table 1 presents summary statistics for engineering education Twitter data.

Table 1. Summary statistics of engineering education Twitter data collected.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Count</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of tweets</td>
<td>18973</td>
<td>100.0 per cent</td>
</tr>
<tr>
<td>Tweets</td>
<td>10676</td>
<td>56.3 per cent</td>
</tr>
<tr>
<td>Retweets</td>
<td>8297</td>
<td>43.7 per cent</td>
</tr>
<tr>
<td>Directed / Mention tweets</td>
<td>9954</td>
<td>52.5 per cent</td>
</tr>
<tr>
<td>Undirected tweets</td>
<td>9019</td>
<td>47.5 per cent</td>
</tr>
</tbody>
</table>

Table 1 shows that 56.3 per cent of the tweets collected were ‘new’ content from a user, while 43.7 per cent were retweets of a post from someone else. The level of retweets in general Twitter data has been reported as relatively low (3 per cent) (Boyd, Golder, & Lotan, 2010). However, higher proportions (29.3 per cent) have been reported in studies looking at the tweeting characteristics of individuals (rather than organisations) (Xu et al., 2011). Trends occur on Twitter when a topic of discussion becomes popular, and a trend is typically detected by the presence of a commonly occurring keyword/phrase. Naaman et al. (2011) propose two types of Twitter trend: i) exogenous – arising from an external event, such as an earthquake; and ii) endogenous – arising from groups of users deliberating sharing information on a topic. They found the proportion of retweets associated with trends to be much higher than general Twitter traffic – 32 per cent for exogenous trends and 47 per cent for endogenous trends. 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). The tweets collected via the keyword search strategy used here probably results in a Twitter data set that synthetically exhibits characteristics much more like an endogenous trend than general Twitter traffic. The common, if largely asynchronous, interest in the topic of ‘engineering education’ embodied in the data collected may explain the high proportion of retweets observed. Table 1 shows that 47.5 per cent 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 52.5 per cent were directed or otherwise specifically mentioned another user. The proportion of mentions in general Twitter data has been reported as 36 per cent (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 engineering education.

Based on the schema presented in Figure 1 and using the F-R layout algorithm, Figure 2 presents an overall network visualisation of the large-scale structure of the engineering education Twitter data during the period under investigation.
The F-R layout algorithm produces a complicated network rich in features. The plume-like structures (such as point E in Figure 2) 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 within the plume. These 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; however, the majority of these mentions are typically retweets of an initial tweet originating from the user at the focus point. There is a large whirlpool-like structure in the middle of Figure 2 that centres on a relatively large node. This structure arises from the method chosen to represent undirected tweets in the network – all 9019 connecting to the large ‘undirected’ node near the centre of the network. Outside of the central whirlpool there is a ring of more complicated connections – including edges representing retweets of undirected tweets, and/or interactions (Twitter conversations) between users. Finally, there is a thin halo of nodes around the outer edge of the network that have no connection to the main network proper. This outer region represents small groups of two or more users sharing tweets about engineering education – as exemplified by point H in Figure 3. The symmetry-emphasising characteristic of the F-R layout
algorithm is apparent in Figure 2 – the overall layout is approximately circular and balanced, with the principal regions in concentric rings.

A notable feature in Figure 2 is the spiral structure observed at point D – more detail can be seen at point G in Figure 3. Inspection of the tweets underlying this feature reveals that it is a series of ‘robot’ Twitter accounts systematically retweeting posts from the account at the centre of the spiral. Tweets from this central account are largely advertising engineering text books, and the whole structure is effectively an attempt at large-scale spam via Twitter. The structure at point D in Figure 2 accounts for 6679 tweets, which is 35.2 per cent of all tweets collected in this work. Spam is a common occurrence on Twitter, typically in the form of tweets designed to lure readers to a web site, and one study found that 8 per cent of general Twitter traffic was spam content (Grier, Thomas, Paxson, & Zhang, 2010). Spam on Twitter is often targeted at a theme or trend related to the item being advertised, so it is perhaps not surprising that the proportion of spam in tweets associated with a specific topic can be significantly higher than general Twitter content.

Figure 3 shows expanded details associated with point D in Figure 2. Figure 4 shows expanded details associated with point C in Figure 2. Additional features present in these Figures are discussed below.
Table 2 presents the ‘top’ Twitter accounts based on a range of measures of impact.

Table 2. ‘Top’ engineering education Twitter accounts for various measures of impact.

<table>
<thead>
<tr>
<th>Most prolific (number of tweets)</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>@deg511</td>
<td>243</td>
<td>@BigBeacon</td>
<td>211</td>
</tr>
<tr>
<td>Spam account 1</td>
<td>179</td>
<td>Spam account 2</td>
<td>179</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Most retweeted (number of retweets)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Spam account 4</td>
<td>6679</td>
<td>@ScotGovFM</td>
</tr>
<tr>
<td>@HRDMinistry</td>
<td>83</td>
<td>@careersingov</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Most mentioned (number of mentions)</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Spam account 4</td>
<td>204</td>
<td>@chronicle</td>
<td>107</td>
</tr>
<tr>
<td>@NSF</td>
<td>94</td>
<td>@OlinCollege</td>
<td>89</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Most influential (number of tweets x average number of followers)</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>@NSF</td>
<td>60153720</td>
<td>@careersingov</td>
<td>9948136</td>
</tr>
<tr>
<td>Spam account 1</td>
<td>4492721</td>
<td>@BigBeacon</td>
<td>2650055</td>
</tr>
</tbody>
</table>

Table 2 identifies the most prolific tweeters observed in the data set. The account with the most tweets is @deg511. This account belongs to David E. Goldberg, an emeritus engineering professor and founder of the organisation Big Beacon, which promotes engineering education reform. The second ranked account is the @BigBeacon account. @deg511 can be seen at point I in Figure 4, and @BigBeacon can be seen nearby at
point K. Both of these accounts have wide edges connecting them to the central @undirected node, which represent the relatively large number of tweets from these accounts. Edges connecting @deg511 and @BigBeacon can be seen, highlighting the ‘real-world’ links between these two accounts. The third most prolific tweeter is @campusmckvie, which can be seen at point B in Figure 2, again with a wide edge connecting to @undirected. @campusmckvie is an Indian engineering college that was relatively active on Twitter during the period under consideration. The next three highest tweeting accounts were all members of the outer spiral arms of the spam robot array at point G in Figure 3. Table 2 identifies the most retweeted accounts. The most retweeted account by nearly two orders of magnitude is the one located at the centre of the spiral at point G in figure 3, due the robotic spam operation. The second most retweeted account, @ScotGovFM, is the First Minister of Scotland promoting a new engineering program at City of Glasgow College, visible at point A in Figure 2 – further notes below. The third most retweeted account is @deg511, noted above. The fourth most retweeted account, @HRDMinistry, is the Indian Ministry of Human Resource Development promoting an engineering education symposium, visible at point A in Figure 2. The next most retweeted account, @careersingov, is an online recruitment site specialising in US government jobs, and can been seen at point F in Figure 3. The final most retweeted account is @BigBeacon – as noted above, this account is associated with @deg511, and it is likely that there is some level of mutual retweeting between these accounts.

Table 2 identifies the most mentioned accounts – Twitter mentions include retweets, so there can be some overlap with the previous category. As with retweets, the most mentioned account is the one located at the centre of the spiral at point G in figure 3. The second and third most mentioned accounts, @chronicle and @Forbes, are two specialised news publications with associated online presences. The fourth most mentioned account, @NSF, belongs to the US National Science Foundation, who might be expected to tweet messages about engineering education. The next most mentioned account, @OlinCollege, is an engineering school and can be seen at point J in Figure 4. Note that @deg511 (point I in Figure 4) is a former academic associate of Olin College, and strong network links can be seen in Figure 4 between @deg511, @OlinCollege and @BigBeacon (point K). The final most mentioned account, @CofGcollege, is the City of Glasgow College. As noted above, @ScotGovFM tweeted about a new engineering program at City of Glasgow College (including the account name @CofGcollege), and the relatively large number of retweets of that led to a relatively large number of mentions of @CofGcollege. This interaction can be seen at point A in Figure 2 – where the plume structure has two foci – one for retweets of @ScotGovFM, and one for mentions of @CofGcollege. 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 engineering education Twitter data set. @NSF and @careersingov appear as first and second most influential due to their large respective follower bases. @deg511 and the associated account @BigBeacon appear as third and fifth most influential due to a combination of their relatively large follower bases and their relatively high level of tweeting. Two spam accounts round out the top six most influential accounts, again testament to the high proportion of spam activity observed here in the engineering education data set.

4 Conclusion

This paper presents an analysis of nearly 19,000 tweets relating to ‘engineering education’ collected over a period of almost a year. Descriptive statistic were compiled and social network analysis was used to visualise the Twitter data. Compared to studies of general Twitter traffic, relatively high proportions of retweets and mentions were observed. This suggests that engineering education is an active and interactive topic on Twitter, and this avenue for stakeholder communication should be of interest to all engineering academic units. The most prolific, most retweeted, most mentioned and most influential (based on the empirical measure of number of tweets multiplied by average number of followers) ‘engineering education’ Twitter accounts were identified. While some engineering colleges are apparent as active tweeters or as being frequently mentioned, the most influential accounts on the topic of engineering education are government institutions and individuals who are particularly active on Twitter and have a significant number of followers. This suggests that engineering academic units wishing to have influence on Twitter need to cultivate an audience of followers through both content and interaction relevant to the audience(s) targeted. A very high proportion of spam tweets were
observed. Spam in social media is a constant presence, and needs to be accounted for in institutional social media strategies, even in engineering education. A limitation of this study is that the data collection strategy and the operation of the Twitter public API mean that the data obtained are only a subset of all tweets related to engineering education. Additionally, space limitations meant that only the top half-dozen accounts in each category could be presented in Table 2. Finally, this study looked only at the structure of the network embodied by the Twitter communication, and not at the content of the tweets comprising the network. The tweet text content is a rich and valuable data set in its own right, and deserves a separate analysis. This work provides insights regarding how engineering education is currently represented in social media (specifically Twitter) internationally, and offers a detailed methodology to those interested in future research in this area.

5 References


