FAN PAGE MANAGEMENT VIA CONTENT GENERATION AND FEEDBACK STRATEGIES

Hamidreza Shahbaznezhad

University of Auckland Business School, h.shahbaznezhad@auckland.ac.nz

Follow this and additional works at: http://aisel.aisnet.org/pacis2016

Recommended Citation
http://aisel.aisnet.org/pacis2016/324
FAN PAGE MANAGEMENT VIA CONTENT GENERATION AND FEEDBACK STRATEGIES

Hamidreza Shahbaznezhad, Information Systems and Operations Management, University of Auckland Business School, Auckland, New Zealand, h.shahbaznezhad@auckland.ac.nz

Abstract

The airline industry is known as one of the most active industries on Twitter. Twitter fan pages are used by airlines for different purposes, e.g., customer service, complaint management, branding, and promotions. Firms try to motivate users to enhance fan page engagement. Thus, social computing techniques have become more important in recent years as these enable firms to improve their impact on fan page activities and to analyze user behavior. Apart from operational factors and reputation concerns, different fan page management strategies adopted by the airlines lead to different numbers of followers and followings, and, consequently, different rates of user engagement. This research explores how airlines develop their fan page strategies on Twitter, and how the similarity of generated content leads to higher engagement rates. We examine different aspects of firms’ tactics in reaction to users’ activities on fan pages, and common methods in managing content of fan pages. Our empirical study consists of a sample of 36 prominent airlines all around the world. We analyze the data obtained from a set of almost 153,992 postings (including tweets, retweets, and replies) that are published on the companies’ fan pages. Our results show that the topics of the company tweets can be categorized based on some characteristics of the fan page. Moreover, there is a relationship between a company’s activities on fan pages and number of followers.

Keywords: fan page management, content strategy, social computing, clustering, text mining.
INTRODUCTION

Airlines are considered as large scale companies which engage a huge part of society with their services, thus, they are recognized as one of the critical players in the travel industry. Also, from the customer communications and public relation point of view, airlines have exclusive touch points all around the globe. Hence developing high quality campaigns and strong presence on social media is a means of creating excellent engagement outcomes. The formal presence of firms in social media is by fan pages. Therefore, in this paper, we explore how airlines are harnessing fan pages.

There are many studies in marketing and information systems (IS) literature which address different perspectives of social media and specifically fan page management. There are quite a few number of papers which discuss strategies of listening or talking in social media from a user’s or company’s perspective. Some studies address users’ activities on social media, like the study of Shahbaznezhad and Tripathi (2016). Some others study companies’ strategies in this environment, like the study of Miller and Tucker (2013). But it is still unclear which parties are interested in what kind of content and topics in this open environment. For instance, do companies just follow whatever users’ interests are and try to concentrate on a feedback strategy? Or do they generate their own content? Or even a combination of these actions. If they are more inclined to the user contents, do they try to react more towards negative content or to positive ones? Perhaps customer tweets are related to customer service, while factors like timeliness, food and entertainment, and luggage handling might make up most of the company side tweets. Also, do different companies generate similar content, or does each company have its own fan page content strategy? Accordingly, if we suppose that some companies have similar patterns in attracting users’ attention, like the same rate of following new users, do these companies have similar content strategy in social media, or do they develop different content types?

As Aral et al. (2013) mentioned, there is not much understanding about the best methods that firms should employ and manage, given their presence on social media. The research gap is that little attention has been paid to a firm’s role as a dynamic phenomenon in absorbing users’ attention in competition with similar firms on social media. There lacks comprehensive work which we are aware of regarding how a firm’s role (active and passive) relates to user engagement. Thus, we want to address this issue by answering the following question from the firm’s point of view. How do companies’ actions and reactions on their fan pages lead to different content and to user’s following action? As Aral et al. (2013) suggested, an important question in this domain is how firms should communicate with different stakeholders on social media. Moreover, Miller and Tucker (2013) argued that there is a huge amount of complexity and subtlety in the way that firms should manage various constituencies in social media. In this research, we try to analyze firms’ behavior from two different perspectives. Firstly, we characterize their fan page behavior based on some active and passive variables. For example, when a company reacts to a user tweet in terms of a reply we call it an active behavior, but when a company retweets a user tweet, it is deemed a passive reaction. We then categorize firms into clusters. These clusters explain characteristic similarities of the firms’ fan pages. Secondly, we run the topic modelling algorithm to explore most recent common topics of the fan pages that are generated by the companies within a cluster. This will help us to study the relationship of three aspects of a firm’s fan page: feedback handling, content generation (by the company), and users’ attention to the fan page. To answer this question, the following research objective will be pursued: studying the companies’ strategies in terms of content generation and feedback management to see what type of firm communication provokes which level of response (Aral et al. 2013).

LITERATURE REVIEW

In this section, firstly, we review the existing studies on content generation strategies in social media environment, i.e., fan pages. Particularly, the necessity of applying different content management
strategies for interacting with users is discussed. Secondly, we review the literature for investigating the ecosystem between firms and users to see how companies feedback effect user behavior. Finally, we present the appropriate theories to support our arguments and results.

**Content management and Fan page engagement**

Content refers to the resources available in a network such as information, gossip, etc. (Kane et al. 2014). It has been argued that the value of a social network is determined by the nature of content flow (e.g., Lin 1999). The real content that users exchange makes a web of cooperative relationships that reproduce norms, trust, common purpose, and harmonization, this is called social capital (e.g., Brehm and Rahn 1997; Coleman 1988). The contents of the firms’ fan page is important in relation to the customers’ engagement and most of this content is supplied by the firms. Businesses make advertising content to directly promote their products and services (Singh et al. 2010; Wen et al. 2009). Also, they generate non-product-related content in order to increase the richness of their dialogue with users (Senadheera et al. 2011; Yu et al. 2010), and to promote future purchases indirectly (Karson and Fisher 2005). On fan pages, not only firms but also users generate content by conversing about companies’ products and services or by contributing in non-product related discussions (Kim and Miranda 2011; Senadheera et al. 2011). By this means, users share their sentiments and understandings and haven’t any concern to express their positive or negative reactions honestly, reflecting their overall attitude towards debated products, services, or non-product-related subjects (Lin and Goh 2011).

As Rishika et al. (2013) argued in their research, each message posted to the firm’s fan page has its own specifications. These elements may aim to build store equity, improve customer relationship, inform customers about events or activities, and spread information about products and services. Also Aral et al. (2013) demonstrates the importance of carefully specifying the nature of the social communications and considering the multidimensional effects they may face. These models project that the company’s optimal product policy responses to the growth in social connections depend on both the content and the structure of the subordinate conversations. The firm strategy for engaging in generating content with itself or customers is another aspect that should be considered. The amount of a firm’s engagement can show to the firm’s current and potential customers the firm’s involvement with its clients and its commitment for delivering valuable information.

With respect to a recent market report, engaging and valuable content allows businesses to socially engage prospects, and differentiate themselves from competitors and enhance customer loyalty and consciousness (Emerson 2012). Basically useful content has social exchange value (Homans 1958), and individuals might share it to generate reciprocity (Fehr et al. 1998). Social capital (Coleman 1988) produced in the customer and user communication would develop the form of mutual trust, goodwill, and reciprocity (Adler and Kwon 2002). When the interaction between firm and customers, through social media, increases, in order to earn significant respect and expect corresponding returns in the future, both partners have a strong motivation to exchange information with one another (Kankanhalli et al. 2005; Wasko and Faraj 2005). Scott (2007) and Sterne (2010) specified that, before publishing in social media, organizations should assume a user’s perspective and present only those posts that bring value-added information for the reader. The works of Heymann-Reder (2011) validated this report. In the aforementioned studies, they discussed those posts manifesting funny things about the working environment, news affecting the business or information that may report direct financial benefits to the reader are more disposed to capture a user’s consideration. These results indicate that post category has a substantial consequence over user interaction and it can be used for planning of the communications strategy (Cvijikj and Michahelles 2011). Miller and Tucker (2013) argue that companies have a tendency to post general substance rather than client focused content. These posts usually put emphasis on recent organizational achievements or recent issues. According to their data gathering sample, 25 percent of a
firm’s postings are simply client-focused. This result indicates that there is a gap in firm’s content management. Consequently, they cannot plan appropriately to generate proper contents.

Contents can be positive or negative. But generally in firm fan pages companies avoid to disseminate negative contents. When members repeatedly receive positive information and feedback from an online community, they have more tendency to exhibit confidence and trust in the brand (Deighton 1992). Furthermore, feelings of trust can improve members’ psychological attachment to a brand (Mattila and Wirtz 2002), and consequently creates brand commitment (Ha and Perks 2005). If an online community is not successful in delivering consistent benefits to community members, its prosperity may be endangered (Wang et al. 2002). As soon as users perceive the benefits they receive is valuable, they would like to become more active contributors (Morgan and Hunt 1994). Berger and Milkman (2012) studied the content characteristics and their finding to support the reasons that people share content and provide insight about the design of the content in the marketing area. Their psychological approach designates that positive content is more viral than negative, also high-arousal positive or negative contents are more viral. Much research has addressed designing virality through word-of-mouth in the product or content of the media. It’s essential to explore how WOM pushes consumer demand, public opinion, and product diffusion (Aral et al. 2009; Brown and Reingen 1987) and how companies can make broad, systematic propagation of WOM through clients (Godes and Mayzlin 2009; Phelps et al. 2004).

For answering this question that which type of content is more appropriate for fan pages, we need to know why users engage online. Different studies addressed various motives. Information needs, entertainment needs, social interaction and reward/incentives are considered the main drivers for users to engage online (Cvijikj and Michahelles 2011; Cvijikj and Michahelles 2013; De Vries et al. 2012; Fuller 2006; Katz and Blumler 1974; Ko et al. 2005; Park et al. 2009; Smock et al. 2011; Taylor et al. 2011). Scholars have also employed Uses and Gratifications Theory (UGT) to distinguish how specific motives for social media use can be applied to investigate a range of outcomes including fan page engagement. Firstly for satisfying information needs, previous studies, classify informative content as details about product deals, price and product-related aspects and availability. Secondly for entertainment needs, it has been shown that persuasive/entertaining content stimulates users engagement, and informative content stimulates engagement significantly if combined with persuasive content such as emotional and philanthropic content. Thirdly for social interaction, Ko et al. (2005) explain that users with high social interaction motivations tend to engage in human-to-human interaction. This interaction refers to some online social behaviour such as: providing comments, feedback, personal information to an advertiser, participating in on-line discussion or forums and so on. Finally, rewarding content can include monetary incentives, giveaways, prize drawings or monetary compensations. Fuller (2006) argued that factors such as the ability to learn something new, the possibility to get exclusive content and the ability to gain acknowledgement and support from the community have a far greater impact on community members’ rather monetary incentives.

Following the literature above, in this research, it has been tried to use sentiment and type of the content as the basis for social media analysis. I use sentiment analysis tool for understanding the Positiveness or Negativeness of the users’ content. I also use topic modelling algorithm for finding the type of the content. These methods help the researcher to understand the factors which stimulate engagement.

Feedback management and fan page engagement

Historically, businesses planned to reach their goals through one-sided relationships with customers through one-to-many channels like print, radio, television, and more recently the Internet, broadcasting carefully-controlled messages of encouragement with restricted chances for reciprocity
Social media technologies changed this dynamic by empowering a high level of two-way discussion between the organization and its clienteles by bringing new mechanisms for customers to cooperate amongst themselves (Larson and Watson 2011). Furthermore, by popularization of social media and smart phones, the power of the consumers is changed. This transformation in the digital age alters the role of customers from passive to active users. These active users can raise complaints or compliments publicly to a large audience in real time. In the traditional definition of customer relationship management (CRM) it was designed to build and maintain profitable customer relationship by providing higher customer value and satisfaction (Sen and Sinha 2011). With the emergence of SM this fundamental concept of the CRM process is changed and customers do not hang on to a passive role in their relationship with a company (Malthouse et al. 2013). Moreover, customers have the opportunity to spread their thoughts and feelings to the vast amount of users, and organizations have many difficulties to cope with the information that their customers may receive about the product or services (Schultz et al. 2012).

On the other hand, social media brought up new occasions for business to listen and engage with customers and inspire them to become supporters of their products (Malthouse et al. 2013). In this relationship between company and online community members, Jarvenpaa et al. (2013) specified two types of socialization tactics that are employed by firms. They suggest that companies are using institutionalized tactics to push structured and collective communication about the organization heterogeneous internet users. In this posture companies are cautious about the information that is presented. In a different way, pull tactics are informal and contemplate the exceptional desires of heterogeneous individuals. So they make the most advantage of social listening and for example by retweeting the good stuff and show they listen, and that they have fans. Using these audience interaction tactics would enable firms to create multidirectional, productive and open service co-creative dialogue among users of social media. Most of the literature capturing the social interactions do occur and have been shown to influence consumer behavior, little attention has been paid to how the firms can manage and strategically influence these interactions (Godes et al. 2005). The reaction of the firm in a public domain, where every user is considering the details, is very important. For each user behavior category, a company should react concisely. Jones.S. (2013) discussed companies’ reaction for different types of customers in social media is different.

Some of the variables that is defined in this research are based on company reaction to online community. These variables show us what is the companies' follow back strategy. By investigating the proportion of positive or negative responded tweets by the company to total users' tweets, the company behavior in fan page can be investigated. All these efforts demonstrate the amount of company interest to engage with users in fan page to give them constructive feedback or not.

**Practice Theory**

There are so many theories that assume different roles for the firms in regards to community. For instance Garriga and Melé (2013) provide a taxonomy for the prevailing theories in corporate social responsibility. These 4 different categories are instrumental theories, political theories, integrative theories and ethical theories. Also McWilliams and Siegel (2001) presented a framework to study the level of firm’s corporate social responsibility and other factors like advertisement. In another study Schau et al. (2009) take the advantages of practice theory to contribute to forms of collective value creation in fan pages. Practices consist of behaviors, performances, and representations through procedures, understandings and engagements. To engage in practices, people must develop shared understandings and demonstrate competencies to distinguish themselves through adroit performances. They have introduced twelve different practices through four thematic categories as social networking, impression management, community engagement and brand use. Kjellberg and Helgesson (2007) define practices as any activity that contributes to shaping markets. Also Reckwitz (2002) define practices as the routinized behaviors and repeated actions that may provide shared meaning between users, create consumption opportunities, and/or be used to (co-) create value. Yet little attention has been paid to
engagement practices in fan pages. In a related study, Schau et al. (2009) address engagement practices in hybrid online/offline communities. Also Brodie et al. (2013) study the consumer engagement in virtual brand communities, although specific practices not identified within this context. In this study, I have detected ten different practices that can be measured in fan page environment. These practices comes from the platform mechanisms that the fan page is established in that environment. The theoretical contribution of this research is that some practices are identified which are not introduced in other studies. Based on fan page mechanism usage, I group companies that have similar practices and find the most frequent topics. So following the similarity of fan page mechanism’s usage, we want to see what kind of engagement practices are common between different airlines and bring about which levels of user engagement.

DATA AND MEASUREMENT

Data Collection

I have selected airlines based on two different but related factors, service quality and fan page activity. Firstly, I tried to find a verified ranking of airline service quality and real performance in the transportation market. I considered the latest Top 100 airlines awards list¹ at the time of this study. Secondly, I selected airlines that were nominated for their fan page activity. After a comprehensive search, there were many reports and articles that addressed top airlines in social media and specifically Twitter². So I listed all of the available and reliable rankings and removed the redundancies. Merging the two aforementioned factors, I extracted airlines that had both high service quality and good fan page activity. I also considered two additional criteria. The language of the fan pages of the selected airlines should be in English and English should be the first and main language on the airlines’ fan pages in Twitter. Therefore, the result of sifting provided me with 36 airlines across the world. The top 36 Twitter accounts had a combined follower base of 22,413,524 Twitter users who follow these airlines and 755,204 users who are followed by those airlines.

I considered each company’s activity based on the latest 3200 tweets, and collected data using Twitter APIs. One of the most important characteristics of data gathering in this research is that the period of data collection varies for different companies. This depended on the amount of company actions on its fan page. So one company could reach the limit of 3200 tweets in two weeks and another in 6 months. This was another criterion that I considered for clustering that shows firm effort in terms of fan page management.

Method

In order to answer the research questions of this research, I mix three different but related quantitative methods. I use sentiment analysis for measuring the sentiments of the users’ posts that were retweeted or replied to by the company. This helps us to see what each company is interested in regarding posts to be disseminated on their fan page. This answers the question whether all companies are interested in positive posts or they also considered negative ones. After that, I employ k-means clustering for finding different available clusters amongst the companies. Following the results of sentiment analysis and clustering, I continue the research by applying natural language processing technology to analyze more than 16361 tweets that are generated by these top airlines. Topic modelling is the main tool that assists me to analyze the tweets.

¹ http://www.worldairlineawards.com/awards/world_airline_rating.html
Sentiment Analysis

Opinion mining or sentiment analysis is a systematic approach for understanding an author’s opinions, emotions, evaluations, attitudes, and behavior through a specific subject or its characteristics (Liu 2012). Considering several general-purpose sentiment analysis algorithms have been developed, interpreting the context and domain specific sentiments is still a big challenge (Liu 2012). Since the interaction of users as potential customers and companies on microblogging services like Twitter represents a unique context, I decided to use machine learning algorithms and distant supervision for classifying the sentiment of messages (Go et al. 2009). The tweets were analyzed using the “Sentiment 140” package in R. After running the algorithm, the result of the sentiment analysis for each tweet would be positive, negative, or neutral.

Following the results of sentiment analysis I generate two variables for each company. Positiveness is the percentage of positive posts that is generated by users and replied to by a company compared to the total number of posts that are replied to by the company on its fan page. Negativeness is the percentage of negative posts that is generated by users compared to the total number of posts that are replied to by the company on its fan page. In this regard we measure how companies react to different posts. The amount of Positiveness or Negativeness is an important part of a company’s feedback strategy.

Spearman Correlation

I performed a correlation test to measure the potential relationship between different variables. This will help us to remove the variables that generate more noise in our clustering. I rely on the Spearman test because it is flexible enough for testing non-normal data like ours, which is right skewed (Shahbaznezhad and Tripathi 2015). Further, Spearman is a non-parametric test, which assesses the relationship between two variables with a monotonic function (Conover and Iman 1981).

K-means Clustering

I performed clustering to categorize different firms based on common characteristics and strategies employed in their fan page to see their similarities in terms of content generation. I used k-means as it is one of the best unsupervised learning algorithms. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters. The main idea is to define k-centers, one for each cluster. This will help us to position the clusters correctly.

Topic Modeling

Topic modeling algorithms are statistical methods to uncover latent topics that are inherent in documents to help researchers interpret documents with topic labels (Blei 2012). Topic modelling techniques present semantically coherent and interpretable topics by calculating the most probable words for each topic. Additionally, topic models can help us to discover the interrelation between different topics (Blei and Lafferty 2007) and how authors are related to topics (Rosen-Zvi et al. 2004). Also this approach does not require human intervention or prior labeling of documents, which allows an unbiased and repeatable analysis of documents (Vakulenko et al. 2014). This study exploits the Latent Dirichlet Allocation (LDA) (Blei 2012; Blei et al. 2003) technique to identify the most relevant topics in each cluster of users on a company’s fan page. The goal of LDA is to infer hidden distributions given the observed words per document (Blei et al. 2003). Topic modeling via LDA, or its predecessor Latent Semantic Analysis (LSA), is considered a popular research method for the quantitative analysis of qualitative data (Vakulenko et al. 2014). LDA has many advantages over other topic modelling techniques (Uys et al. 2008).

The R package “tm” and “RTextTools”, which is available in the R library, was used for creating corpus and corresponding matrix. Also, the “topicmodels” package was used for fitting topics (Hornik and Grün 2011). Regarding our data set, LDA considers each tweet as a single document. Since the number of tweets in each cluster is not the same and varies from 100 postings to 2100 postings, and considering that I wanted to compare the topics of different clusters with one another, I decided to select 8 topics per
cluster. This number was the maximum number of topics and in some cases I could not even find 8 different topics based on the output of the LDA. Other researchers have experienced the same issues with a Twitter data set as well (e.g., Uys et al. 2008). Moreover, the 8 topics were highly discriminant and were easy to interpret. Subsequently, the words in each topic were thoroughly interpreted and the topics labelled.

**ANALYSIS AND RESULTS**

In the first step, I captured ten different variables, which are listed and briefly explained in Table 1.

<table>
<thead>
<tr>
<th>Variable No.</th>
<th>Variable name</th>
<th>Variable explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>V1</td>
<td>Followers per day</td>
<td>Total number of each airline’s followers divided by the total number of days since they established their fan page</td>
</tr>
<tr>
<td>V2</td>
<td>Followings per day</td>
<td>Total number of each airline’s followings divided by the total number of days since they established their fan page</td>
</tr>
<tr>
<td>V3</td>
<td>Average number of airline original tweets per day</td>
<td>Total number of tweets that is generated by each airline divided by the total number of days that data was captured for the company</td>
</tr>
<tr>
<td>V4</td>
<td>Average number of airline retweets per day</td>
<td>Total number of tweets that is retweeted by each airline divided by the total number of days that data was captured for the company</td>
</tr>
<tr>
<td>V5</td>
<td>Average number of airline replies per day</td>
<td>Total number of tweets made in reply to a user’s tweet divided by the total number of days that data was captured for the company</td>
</tr>
<tr>
<td>V6</td>
<td>Proportion of tweets to total activity</td>
<td>The number of tweets that is generated by each company divided by total number of tweets that I could capture for the company (for most companies this number is 3200)</td>
</tr>
<tr>
<td>V7</td>
<td>Proportion of re-tweets to total activity</td>
<td>The number of tweets that is retweeted by each airline divided by the total number of tweets that I could capture for the company (for most of the companies this number is 3200)</td>
</tr>
<tr>
<td>V8</td>
<td>Proportion of replies to total activity</td>
<td>The number of tweets made in reply to a user’s tweet divided by the total number of tweets that I could capture for the company (for most companies this number is 3200)</td>
</tr>
<tr>
<td>V9</td>
<td>Positiveness of original tweets replied to</td>
<td>Percentage of positive tweets that is generated by users and replied to by each company to the total number of tweets that is generated by users and replied to by the company</td>
</tr>
<tr>
<td>V10</td>
<td>Negativeness of original tweets replied to</td>
<td>Percentage of negative tweets that is generated by users and replied to by each company to the total number of tweets that is generated by users and replied to by the company</td>
</tr>
</tbody>
</table>

*Table 1. Variable definition*

After calculating each variable for each airline, the correlation between these variables is measured by performing a Spearman correlation test in R. Table 2 shows the test results that reflect a meaningful correlation between V8 and other variables.

<table>
<thead>
<tr>
<th></th>
<th>V1</th>
<th>V2</th>
<th>V3</th>
<th>V4</th>
<th>V5</th>
<th>V6</th>
<th>V7</th>
<th>V8</th>
<th>V9</th>
<th>V10</th>
</tr>
</thead>
<tbody>
<tr>
<td>V1</td>
<td></td>
<td>0.531165</td>
<td>0.103924</td>
<td>-0.141</td>
<td>0.38134</td>
<td>-0.0295</td>
<td>-0.1544</td>
<td>0.04972</td>
<td>-0.16633</td>
<td>-0.16925</td>
</tr>
<tr>
<td>V2</td>
<td>0.531165</td>
<td></td>
<td>-0.24502</td>
<td>-0.1617</td>
<td>0.53076</td>
<td>-0.34844</td>
<td>-0.29849</td>
<td>0.36084</td>
<td>-0.17507</td>
<td>0.15048</td>
</tr>
<tr>
<td>V3</td>
<td>0.103924</td>
<td>-0.24502</td>
<td></td>
<td>0.37968</td>
<td>0.05877</td>
<td>0.41378</td>
<td>0.37967</td>
<td>0.4273</td>
<td>-0.27649</td>
<td>0.01594</td>
</tr>
<tr>
<td>V4</td>
<td>-0.14906</td>
<td>-0.16169</td>
<td>0.379767</td>
<td></td>
<td>0.01124</td>
<td>0.010142</td>
<td>0.599043</td>
<td>-0.1054</td>
<td>-0.07524</td>
<td>-0.07251</td>
</tr>
<tr>
<td>V5</td>
<td>0.381341</td>
<td>0.530762</td>
<td>0.058773</td>
<td>-0.1124</td>
<td></td>
<td>-0.33264</td>
<td>-0.30575</td>
<td>0.3439</td>
<td>0.30972</td>
<td>0.308814</td>
</tr>
<tr>
<td>V6</td>
<td>-0.0295</td>
<td>-0.34844</td>
<td>0.413776</td>
<td>0.010142</td>
<td>-0.3326</td>
<td></td>
<td>0.667988</td>
<td>-0.9926</td>
<td>0.045356</td>
<td>-0.25017</td>
</tr>
<tr>
<td>V7</td>
<td>-0.1544</td>
<td>-0.29849</td>
<td>0.379767</td>
<td>0.599043</td>
<td>-0.3057</td>
<td>0.667988</td>
<td></td>
<td>0.7521</td>
<td>0.034111</td>
<td>-0.3011</td>
</tr>
<tr>
<td>V8</td>
<td>0.049723</td>
<td>0.360844</td>
<td>-0.42734</td>
<td>-0.1054</td>
<td>0.3439</td>
<td>-0.99262</td>
<td>-0.7521</td>
<td></td>
<td>-0.05334</td>
<td>0.279188</td>
</tr>
<tr>
<td>V9</td>
<td>-0.16633</td>
<td>-0.17507</td>
<td>-0.27649</td>
<td>0.010142</td>
<td>-0.3326</td>
<td>0.667988</td>
<td>-0.9926</td>
<td>0.045356</td>
<td></td>
<td>-0.52943</td>
</tr>
<tr>
<td>V10</td>
<td>-0.16925</td>
<td>0.15048</td>
<td>0.01594</td>
<td>-0.0725</td>
<td>0.308814</td>
<td>-0.25017</td>
<td>0.3011</td>
<td>0.27919</td>
<td>-0.52943</td>
<td></td>
</tr>
</tbody>
</table>

*Table 2. Spearman test matrix result*
This shows the proportion of replies is explained by other existing variables. So I decided to remove V8 from the list of variables for clustering. To find out the optimum number of clusters, I followed the Elbow criterion, also known as the F-test (Thorndike 1953). The percentage of variance would be attained by the division of the between-cluster sum of squares to the total within-cluster sum of squares. This estimates the goodness of the classification (Shahbaznezhad and Tripathi 2015). As displayed in figure 1, the optimum number of clusters is 6.

![Figure 1. The Elbow criterion for finding the optimum number of clusters](image)

This number increases the ratio between the cluster sum of squares to the total cluster sum of squares. Based on this result, I tried to find which variables still have meaningful diversions between different clusters. We need to keep variables that create a high variance to find meaningful differences in each cluster. After running the k-means cluster algorithm in R, and finding which company belongs to which cluster, a boxplot is used for each variable to see if there are any meaningful difference between clusters. Figure 2 shows these plots.

![Figure 2- Boxplot of each cluster for each variable](image)
Based on the results of box plot analysis, V3, V4, V6 and V7 are removed because there was not enough variance between different clusters for those variables. So the value of the elbow criterion is recalculated and the value is again 6 which allows for running the clustering algorithm with k = 6. Table 3 shows the results of clustering for different airlines. I also tried to change the quantitative demonstration of data to a qualitative one to have a better understanding of data behavior. So for each variable, I used the quartile points for specifying airline activities for each variable. In this order for each column I find the first, second, and third quartile. Those companies whose activity for a given variable is between zero and the first quartile, the label “low value” is assigned. Accordingly, for 25-50 percent “medium”, for 50 to 75 “high”, and from 75 percent to the maximum value “very high” is assigned. This will help us to compare different airlines based on their values (labels) for each variable. Table 3 shows this qualitative version of the data for each cluster.

<table>
<thead>
<tr>
<th>Airline</th>
<th>Followers Rate (V1)</th>
<th>Followings Rate (V2)</th>
<th>Replies Rate (V5)</th>
<th>Org user post Positiveness (V9)</th>
<th>Org user post Negativeness (V10)</th>
<th>Cluster Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>EasyJet</td>
<td>Medium</td>
<td>High</td>
<td>Very High</td>
<td>Low</td>
<td>Very High</td>
<td>1</td>
</tr>
<tr>
<td>Air Canada</td>
<td>Medium</td>
<td>High</td>
<td>Very High</td>
<td>Medium</td>
<td>Very High</td>
<td>1</td>
</tr>
<tr>
<td>JetBlue Airways</td>
<td>Very High</td>
<td>Very High</td>
<td>Very High</td>
<td>Medium</td>
<td>Very High</td>
<td>2</td>
</tr>
<tr>
<td>British Airways</td>
<td>High</td>
<td>Very High</td>
<td>Very High</td>
<td>Medium</td>
<td>Very High</td>
<td>2</td>
</tr>
<tr>
<td>Royal Dutch Airlines (KLM)</td>
<td>Very High</td>
<td>Very High</td>
<td>Very High</td>
<td>Low</td>
<td>Low</td>
<td>2</td>
</tr>
<tr>
<td>United</td>
<td>Very High</td>
<td>Very High</td>
<td>Very High</td>
<td>Low</td>
<td>Very High</td>
<td>2</td>
</tr>
<tr>
<td>Alaska Airlines</td>
<td>Low</td>
<td>Medium</td>
<td>Very High</td>
<td>Very High</td>
<td>Medium</td>
<td>3</td>
</tr>
<tr>
<td>Virgin Australia</td>
<td>Medium</td>
<td>High</td>
<td>Medium</td>
<td>High</td>
<td>High</td>
<td>3</td>
</tr>
<tr>
<td>Finnair</td>
<td>Low</td>
<td>Medium</td>
<td>Medium</td>
<td>High</td>
<td>High</td>
<td>3</td>
</tr>
<tr>
<td>Asiana Airlines</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>3</td>
</tr>
<tr>
<td>Jetstar Airways</td>
<td>Low</td>
<td>Very High</td>
<td>High</td>
<td>Medium</td>
<td>Very High</td>
<td>3</td>
</tr>
<tr>
<td>Swiss Intl Air Lines</td>
<td>Medium</td>
<td>Low</td>
<td>High</td>
<td>Medium</td>
<td>Medium</td>
<td>3</td>
</tr>
<tr>
<td>Porter Airlines</td>
<td>Low</td>
<td>High</td>
<td>Medium</td>
<td>Very High</td>
<td>High</td>
<td>3</td>
</tr>
<tr>
<td>Hawaiian Airlines</td>
<td>Low</td>
<td>Medium</td>
<td>Low</td>
<td>High</td>
<td>Very High</td>
<td>3</td>
</tr>
<tr>
<td>LAN Airlines USA</td>
<td>Low</td>
<td>Medium</td>
<td>Low</td>
<td>Low</td>
<td>Very High</td>
<td>3</td>
</tr>
<tr>
<td>SAA - South Africa</td>
<td>Low</td>
<td>Medium</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>3</td>
</tr>
<tr>
<td>All Nippon Airways</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Very High</td>
<td>Low</td>
<td>3</td>
</tr>
<tr>
<td>Southwest Airlines</td>
<td>Very High</td>
<td>Very High</td>
<td>Medium</td>
<td>High</td>
<td>High</td>
<td>4</td>
</tr>
<tr>
<td>American Airlines</td>
<td>Very High</td>
<td>Very High</td>
<td>Very High</td>
<td>High</td>
<td>Low</td>
<td>4</td>
</tr>
<tr>
<td>AirAsia</td>
<td>Very High</td>
<td>Low</td>
<td>Medium</td>
<td>Low</td>
<td>Low</td>
<td>4</td>
</tr>
<tr>
<td>Philippine Airlines</td>
<td>Very High</td>
<td>Very High</td>
<td>High</td>
<td>Medium</td>
<td>Medium</td>
<td>4</td>
</tr>
<tr>
<td>Emirates Airline</td>
<td>Very High</td>
<td>Low</td>
<td>Low</td>
<td>Very High</td>
<td>Low</td>
<td>4</td>
</tr>
<tr>
<td>Delta</td>
<td>High</td>
<td>Medium</td>
<td>Low</td>
<td>Very High</td>
<td>Low</td>
<td>5</td>
</tr>
</tbody>
</table>
The first cluster represents the airlines for whom most of their activities are dedicated to replying negative tweets. The rate of following users in this cluster is high but they couldn’t grab an outstanding rate of followers’ absorption. The second cluster also has a high level of answering negative tweets as well as a high level of content generation. The rate of following new users is also very high and it was the main reason that we have these 4 airlines in this cluster. The main difference between the first and second clusters is that both airlines have the same strategy of dedicating their fan page activities to replying to negative posts and following lots of users. However the first cluster hasn’t reached a high level of engagement compared to airlines in the second cluster. Since companies of both clusters are similar in terms of tweet generation, we can conclude that they might have different page content, which causes different user behavior. Companies in the third cluster concentrate more on tweet generation and retweet propagation instead of replying. The firms in this cluster do not have any specific pattern for replying to positive or negative tweets but most of them consider “replying” as an important strategy, and focus more on positive tweets. They have medium activity in terms of following new users and it shows they do not have any specific strategy for gaining users’ attentions by following them. All in all the common characteristic in this cluster is that they have a low rate of gaining new followers in their fan page. In the fourth cluster, the rate of following new users by the companies is high. They do not have a considerable amount of activity in terms of content generation or propagating users’ tweets. Most of their activities are dedicated to replies. They tend to focus more on positive tweets than negative ones. But they have a high level of user fan page engagement and the rate of followers that they obtain is among the best. It means that focusing on a reply strategy, even on positive tweets, might be good. Also, we can conclude that the airlines in this cluster might generate fancy contents and this was the cause of user attention to the fan page. So this emphasize the importance of the topic modeling.

Most of the companies in the fifth cluster do not follow new users very much and they are not looking for a following strategy. Their reply rate is not high and they also do not have a replying strategy. They tend to generate new content rather than reply to customer comments. So when we compare the high amount of user engagement (high rate of followers) with paying less attention to users (low rates of followings and reply), we might challenge this statement that a reply strategy is the best strategy for gaining user attention. Generating eye-catching tweets and propagating users’ tweets might also effect user behavior to follow company.

The last cluster includes airlines for whom the greatest proportion of their activities are dedicated to tweet generation and propagation. They also have a moderate activity level for replying to users’ posts and most of their replying activity is about replying to positive posts. The special thing about this cluster is that all the airlines in this category are full service airlines. Their usage behavior of fan page

<table>
<thead>
<tr>
<th>Airlines</th>
<th>High</th>
<th>Very High</th>
<th>Low</th>
<th>High</th>
<th>High</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Virgin America</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>Qatar Airways</td>
<td>High</td>
<td>Low</td>
<td>Medium</td>
<td>Medium</td>
<td>Medium</td>
<td>5</td>
</tr>
<tr>
<td>Turkish Airlines</td>
<td>Very High</td>
<td>Low</td>
<td>Low</td>
<td>Medium</td>
<td>Medium</td>
<td>5</td>
</tr>
<tr>
<td>Malaysia Airlines</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>5</td>
</tr>
<tr>
<td>Cathay Pacific</td>
<td>Medium</td>
<td>Medium</td>
<td>Medium</td>
<td>Very High</td>
<td>Medium</td>
<td>6</td>
</tr>
<tr>
<td>Kenya Airways</td>
<td>Medium</td>
<td>High</td>
<td>High</td>
<td>Medium</td>
<td>Medium</td>
<td>6</td>
</tr>
<tr>
<td>Air New Zealand</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>Very High</td>
<td>Medium</td>
<td>6</td>
</tr>
<tr>
<td>Virgin Atlantic</td>
<td>High</td>
<td>Medium</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>6</td>
</tr>
<tr>
<td>Etihad Airways</td>
<td>Medium</td>
<td>Medium</td>
<td>High</td>
<td>Very High</td>
<td>Low</td>
<td>6</td>
</tr>
<tr>
<td>Air France</td>
<td>Medium</td>
<td>High</td>
<td>Very High</td>
<td>Low</td>
<td>Low</td>
<td>6</td>
</tr>
<tr>
<td>Lufthansa</td>
<td>Medium</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>Very High</td>
<td>6</td>
</tr>
<tr>
<td>Qantas</td>
<td>High</td>
<td>High</td>
<td>Medium</td>
<td>Very High</td>
<td>Medium</td>
<td>6</td>
</tr>
<tr>
<td>Singapore Airlines</td>
<td>High</td>
<td>Low</td>
<td>Medium</td>
<td>High</td>
<td>High</td>
<td>6</td>
</tr>
</tbody>
</table>

Table 3. The results of clustering
mechanisms shows they pay moderate attention to managing their page in competing in a social media market with other airlines. As a result of this fan page strategy, they attract users’ attention moderately.

Due to the diversity of companies’ fan page strategies, we recommend content analysis at the cluster level. We employed a topic modelling algorithm for each cluster to examine the differences across clusters (Shahbaznezhad and Tripathi 2016). By comparing these topics, I explore the dissimilarities in the content created and disseminated by firms from different clusters thereby affecting users’ decisions to follow a company fan page. Table 4 lists the topics for each cluster.

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Most specific topics</th>
<th>Topic category</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Travel alert and asking for rebooking, reservation modification, encouraging users to use new services, advertising new destinations, discussing about on-board options like Wi-Fi, suggesting weekend plans</td>
<td>Customer service, Product promotion</td>
</tr>
<tr>
<td>2</td>
<td>Informing about weather conditions of airports or cities, talking about attractions like cinemas and destinations, presenting some fancy and popular topics like drinking coffee after sleeping, acknowledging inflight gifts like suitcases, informing about luggage missing, talking about social plans like aviation introduction for students, seat quality and seat options, using eye-catching words like naked aircraft,</td>
<td>Customer service, Timeliness, Social conversations, Entertainment, Product promotion, Comfort and security, Luggage handling</td>
</tr>
<tr>
<td>3</td>
<td>Airport condition, flight time, wishing a happy trip for passengers, motivating users to participate in contests by offering flights, talking about new booking system, sharing flight photos, congratulating events like valentine, presenting new destinations</td>
<td>Customer service, Timeliness, Social conversations, Product promotion</td>
</tr>
<tr>
<td>4</td>
<td>Booking and checking in, new flight schedule, lounge specification, motivating to plan for holiday, encouraging users to have new experiences or asking to share their own flight or travel experience, focusing on new airplane (airbus) specifications, encouraging users to participate in contests to win movie tickets/airfares/hotels/tour packages, introducing best cities to fly in, inspiring users to visit specific links, announcing celebrities who visited the airline</td>
<td>Customer service, Timeliness, Social conversations, Product promotion, Comfort and security, Branding by celebrities</td>
</tr>
<tr>
<td>5</td>
<td>Energizing users to share tweets and photos, boosting new destinations and flights, cheering users for winning flights/accommodations/great prizes, informing followers about flight cancellations, guiding users to some links to check flight statuses online, wishing to have a great holiday or weekend, sharing company success in the competing market with customers, inspiring users to share their moments with airline on fan page</td>
<td>Social conversations, Product promotion, Customer service</td>
</tr>
<tr>
<td>6</td>
<td>Inviting users to watch company video clips, offering special discounts via coupons, introducing and advertising new destinations, informing users about arrival and departure times, inspire fans to share their photos, new year congratulations and new offers, appreciation of customers for selecting company services or participating in company fan page events, providing reasons for flight delays that was out of their control like airport issues or air traffic control</td>
<td>Social conversations, Product promotion, Customer service</td>
</tr>
</tbody>
</table>

Table 4. Clusters and topics

Since there are different topics for each cluster, I categorize the topics to be able to compare them. For this purpose, I use a report from social times³. This report presents an infographic for analyzing

airline twitter activity to study which one was doing best in terms of twitter activity. The main categories of this study are customer service, timeliness, social conversations, food and entertainment, comfort and security, luggage handling.

I found those airlines that have the highest rate of followers’ absorption also cover most of the topic categories in their fan pages. Looking at the topics of clusters 2 and 4 as successful companies in terms of high rate of followers’ absorption, we can see that they tweet about a wide range of topics. Comparing the topics that are covered by these two clusters and the rate of followers’ absorption rate on their fan pages shows the amount of successfulness in social conversations. They could cover social plans like aviation tours for students, or running small contests, and presenting movie tickets to winners. It shows that they have permanent plans for engaging more with users.

DISCUSSION AND CONCLUSION

Fan pages on social media platforms are becoming an important part of firms’ social media strategy. Firms are becoming strategic in how they manage their fan pages on social media platforms and how they generate more followers. Though a firm fan page has become an integral part of any firm’s social media strategy, we are yet to understand how firms’ fan page strategies affect user engagement on social media platforms. The diversity of firms’ business, user’s motivation for engagement, and functionalities of social media platforms have made it challenging to manage fan pages. The main purpose of this study is to explore company fan page behavior in interaction with users on twitter, which is one of the most popular social media platforms. In this study, we investigate if companies’ activities in terms of generating tweets, retweeting users’ posts or replying to users’ posts have any effect on users’ decisions to follow the company on Twitter. To this end, I explored sentiments of the original tweets that are posted by users and are replied to by companies. I have also created other variables to capture company fan page activity. The airline industry is selected because the level and sensitivity of fan page activities are higher than other industries. The result of this research can be expanded in the other context as well. The proposed quantitative method is industry independent and defined following platform characteristics, not industry specification.

Topics of the content that is generated by the company on their fan page is another important aspect that grabbed our attention. I used different approaches to examine how users react to diverse companies to follow them. I discuss results obtained from three different methods: clustering companies based on their activities, sentiment analysis of users’ posts, and topic modelling. I employed Latent Dirichlet Allocation (LDA) to analyze tweets of six clusters to identify recurrent topics. Our results demonstrate that by relying on new quantitative analytical methods to structure, analyze and manage the content, companies can get better insight of how to codify an appropriate strategy. One of the limitations of this study is that dataset used is a snapshot of user activities. Since companies which are selected in this study may change over time, a longitudinal data set of user behavior and activities would be better to understand the firms’ strategies and evolving user behavior in response to firm strategies. I aim to continue data collection for each active follower and monitor them over a period of time. Our results show that nature of the content generated by firms and how to react to positive and negative sentiments of the contents that is generated by users play a significant role in gaining new followers.

ACKNOWLEDGMENT

This study was funded by the University of Auckland Business School Department of Information Systems and Operations Management. The author would like to thank Associate Professor Arvind Tripathi for his continuous support. The author would also like to thank Ron Tiong and Assistant Professor Amir Karami for their valuable comments to improve this paper.
References


Vakulenko, S., Müller, O., and Brocke, J.v. 2014. "Enriching Itunes App Store Categories Via Topic Modeling.")

