Using Twitter network to detect market segments in the airline industry

Aymeric Punel[a], Alireza Ermagun[a,b,*]

[a] Department of Civil and Environmental Engineering, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208, United States
[b] Department of Civil and Environmental Engineering, Mississippi State University, 501 Hardy Road, Mississippi State, MS 39762, United States

ABSTRACT

The use of social media is increasingly considered a state-of-the-art practice for marketing campaigns. Researchers have investigated various social networks and media platforms to capture the behavior and characteristics of customers. This paper analyzes the official Twitter account of an airline company, Air New Zealand, to explore its market segments. To detect the communities of customers, we develop a network clustering method, which reveals the community classes of the network, along with a text mining analysis on each community detected by the cluster analysis. The results of the network analysis demonstrate that the social network of customers in the airline industry follows Pareto's principle that is similar to scale free networks. The findings of network clustering indicate Air New Zealand is essentially followed by New Zealand citizens. The local accounts are categorized into four communities: (1) lambda New Zealand citizens, (2) management, marketing, and digital media companies, (3) tourism and dining sectors, and (4) New Zealand sport players; while the global accounts fall into two communities: (1) worldwide celebrities and (2) the travel and aviation industry. The community detection method developed in this research is beneficial for marketing and customer strategy purposes as it enables airline companies to detect the categories of passengers interested into the brand. It also allows them to identify the potential sources for advertising by seeking out exceptionally connected customers who have high degrees of centrality.

1. Introduction

The airline industry is evolving in a highly competitive world, with each airline attempting to captivate and retain new customers. To remain competitive in the market, airlines are required to identify their customers properly and respond to their needs. The emergence of big data has spawned new solutions for airlines to respond to the needs of customers, and thereby improve their operations. Creating profiles on customers allows airlines to assess specific marketing campaigns, and consequently provides personalized service (Airlines International, 2013). Text mining and sentiment analysis on thousands of reviews help to gain useful insights on feedback and opinions of passengers and to identify specific traveler segments (Liau and Tan, 2014). In addition, search engine query data has the potential to accurately identify and reflect short-term fluctuations for airline's operations (Kim and Shin, 2016).

Big data sources are numerous, ranging from online reviews on websites to customers' information obtained through search engines or frequent flyer programs. They provide relevant details on passengers' behavior. Recently, big data has relied heavily on social media webpages, such as Facebook and Twitter, to generate large amounts of data.

For instance, Facebook users share about 4.75 billion in content each day, which accounts for nearly 2 billion users. Twitter users post a total of 500 million tweets each day viewable by more than 300 million individuals (Coeffe, 2017). This information is extremely valuable for marketing purposes (Mostafa, 2013).

Different techniques exist to analyze big data obtained from social media (Gandomi and Haider, 2015; Bello-Orgaz et al., 2016). Community detection through social network analysis is one of the techniques attracting attention among transportation analysts. This technique clusters customers using their similarities and dissimilarities and enables airline companies to:

- Gather information that traditional techniques cannot capture. Grouping customers into communities highlights the significant similarities that are shared between members of a same cluster. Airlines then would be able to identify the most significant interests of their passengers.
- Explore the potential sources for advertising campaigns. Understanding the network and deriving its statistics informs airlines of key actors in a cluster, who can influence the sharing of an idea among other members of a cluster.

---

* Corresponding author. Department of Civil and Environmental Engineering, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208, United States.
E-mail addresses: aymericpunel2014@u.northwestern.edu (A. Punel), aermagun@cee.msstate.edu (A. Ermagun).

- Provide incentives. Grouping users into communities facilitates us to spot similarities that are not obvious. Consequently, using these similarities allows airlines to propose incentive packages.

We contribute to the current knowledge on market segmentation in the airline industry by analyzing the official Twitter account of one airline company, Air New Zealand, and exploring its market segments. To detect the community of customers, we develop a network clustering method along with a text mining analysis on each modularity class detected by the cluster analysis. We describe the content of each community by using both quantitative and qualitative analyses and create a list of network statistics for each modularity class. In addition, we perform a text mining analysis on the Twitter description of each user in each community to highlight the most frequent words in each class. The results allow us to examine the characteristics and similarities of the users within a community. The findings are of interest to airlines for marketing purposes, and to understand which categories of customers to target, what features the airline should highlight, and who to employ to promote the brand.

The remainder of the paper is organized as follows. The first section explores the literature related to the use of social media in the airline industry. We continue by presenting the data used for the analysis, explaining the data collection process, and displaying a visualization of the network. We then define and describe the statistics of the network, and follow this by analyzing the clusters defined by the modularity scores. We provide a quantitative description as well as a qualitative analysis of each detected community. This analysis identifies the content of each class and highlights the similarities between its members. Finally, we summarize our findings and discuss future research avenues.

2. Airline companies and social media

The use of social media is increasingly considered a state-of-the-art practice in marketing activities. Researchers have investigated various social networks and media platforms to capture the behavior and characteristics of customers (Wieneke and Lehrer, 2016). Like any other sector of activity, airlines have attempted to have a significant presence on social media (Mostafa, 2013), which has become a common tool for communication (Paniagua and Sapena, 2014) and presents several advantages over traditional channels (Mostafa, 2013; Enginkaya and Yilmaz, 2014; Carnein et al., 2017; Dimitriou and Antoniou, 2017). For instance, airlines can discern customers' opinion and the pros and cons of their service by reading the reviews and posts on micro-blogs, and can directly communicate with users to alleviate their concerns.

Fig. 1 depicts the number of followers for the Facebook and Twitter accounts of the top 20 air carriers extracted from Skytrax Top 100 Airlines of 2017 (Skytrax, 2017). As shown in Fig. 1, some companies are managing social media accounts comparable to some worldwide celebrities. It is found that 16 out of the 20 airlines have more than 1 million of followers on Facebook. The figure also shows some great disparities in social media presence of the airlines as the number of followers can vary significantly. An airline's size seems to play a role in the number of followers. Grančay (2014) explained that the larger the airline, the more likely it is to have active operations on a social network.

In conjunction with the growth of social media accounts among airline companies, research is growing in three distinct streams: (1) building strategies to engage customers on social media (Bygstad and Presthus, 2012; Hvass and Munar, 2012; Leung et al., 2013; Grančay, 2014; Carnein et al., 2017); (2) investigating the potential of social media in obtaining opinions (Sreenivasan et al., 2012; Mostafa, 2013; Liau and Tan, 2014; Misopoulos et al., 2014; Wan and Gao, 2015; Dimitriou and Antoniou, 2017), and (3) measuring customers perceptions toward social media (Dijkmans et al., 2015; Seo and Park, 2018). Table 1 summarizes studies falling into these three streams.

2.1. Stream 1: building strategies to engage customers on social media

Hvass and Munar (2012) showed most of the contents published on airlines social media are related to advertising. Through a 6-month analysis of the official Facebook and Twitter accounts of different commercial airlines, they found the content of the posts is essentially marketing oriented. In another study, Grančay (2014) explored how airlines use the Facebook platform. He collected information available on the 250 largest airlines (distinguished by the number of passengers in 2010) during the second week of November 2012 and ran content analysis. His results showed the level of social media service significantly differs between air carriers. Only two third of the studied airlines maintained an active Facebook page, and only half of them provided contact phone numbers on their page.

Leung et al. (2013) performed a content analysis on the Facebook pages of three budget airlines (i.e., Air Asia, Easy Jet, and Jet Blue) to understand the use of social media by low-cost air carriers. They collected 13 different types of information available on their Facebook pages including the airlines' posts, their likes, their comments, and their shares between June 15th, 2012 and August 10th, 2012. They highlighted three different usages of social media by airlines: (1) advertising and distributing promotional airfare, (2) engaging with customers, and (3) sharing airline company information.

Two studies (Bygstad and Presthus, 2012; Carnein et al., 2017) investigated the actual performances of airline companies on social media. Carnein et al. (2017), for example, evaluated the performance of airlines' customer service by analyzing their response rate and average response time to customers' posts on social media platforms. They collected millions of posts written by customers for 48 airline companies on Facebook (58 accounts) and Twitter (66 accounts), as well as the answers provided by the airlines' officials between January 2016 and November 2016. They tested their analysis on top 10 airlines with the largest social media volume (equivalent to about 1.5 million of posts). The results indicated high variations in the level of service provided by airlines on social media. Most of the full-service air carriers tended to offer a great service, while results were more mitigated for low-cost airlines.

2.2. Stream 2: potential of social media in obtaining opinions

Several studies (Sreenivasan et al., 2012; Mostafa, 2013; Liau and Tan, 2014; Misopoulos et al., 2014; Wan and Gao, 2015) investigated Twitter and its potential of obtaining customer knowledge relevant to airlines. By collecting tweets and analyzing the content of the messages, Sreenivasan et al. (2012) focused on only a five, a four, and a three-star airline, based on Skytrax's official world airline star ranking (www.airlinequality.com). For a two-week period in September 2009, they collected 8978 tweets regarding the three airlines, as well as 260 tweets posted from the official accounts of one of the studied airlines. They analyzed the tweets using content analysis and found most of the tweets were compliments rather than grievances. However, when messages were directly pointed at the airlines, they essentially expressed an issue or a concern. They suggested airlines should track such messages to evaluate consumers' sentiments and address in a near real-time issues they might have.

Mostafa (2013) used a random sample of 2105 tweets related to one of the sixteen airlines selected from the top 100 of major airlines, collected between July and August 2012. His goal was to evaluate consumers' sentiments toward airline services provided by running a sentiment analysis on the collected posts using an expert pre-defined lexicon. Contrary to Sreenivasan et al. (2012), Mostafa's findings suggested an asymmetry in the sentiment scores between the studied airlines. His results indicated a generally negative consumer sentiment toward airline services.

Liau and Tan (2014) used text mining techniques on 10,895 tweets mentioning one of the 5 Malaysian companies studied by hashtag (#) or