A Data Driven Network Analysis of Coffee Consumer Brand Identities and Sentiments*

Anam Khan, Haley Roberts, Nuoya Wu †

June 2022

Abstract

Coffee is one of the most-consumed products worldwide and is provided by both well-known large companies and small boutique stores. Of the many popular beverages on the market, coffee appears unique in the sense that people often find identity and feel strong loyalty toward their coffee brand of choice, claiming to be a "small coffee shop person", a "Starbucks person", an "artisanal coffee person", etc. To understand the sentiments and overlap of coffee identities and identify other key characteristics of brand loyalties, this study utilizes tweets about two major coffee brands: "Starbucks" and "Dunkin Donuts". Ten thousand tweets are collected regarding each coffee brand and textually broken down into significant words. For each tweet, metadata is collected regarding the tweet user's follower counts, the tweet's retweet counts, and more. Then, text analysis is used to score each tweet's sentiment (negative, neutral, positive). Finally, a network is created connecting tweets by their textual commonalities. Network analysis is performed to understand each brand's clientele and investigate similarities and differences between the brands and their consumers' sentiments and attributes. The results provide insights into these brands' consumers and a template for similar analyses of other brands.

Key words: Coffee, Consumer, Brand Identity, Brand Loyalty, Network, Twitter, Culture, Sentiment, Text Analysis, Community Detection, Social Network.

*We are grateful to Tilburg Science Hub for providing this template.

†Anam Khan: Georgetown University; ak1963@georgetown.edu. Haley Roberts: Georgetown University;hne7@georgetown.edu: hne7@georgetown.edu. Nuoya Wu: Georgetown University; nw388@georgetown.edu.

1 Introduction

Coffee is one of the most-consumed products worldwide and has developed into a major international market. Over the past few decades, the consumption of coffee has changed from a purely practical activity into a cultural and personal experience. Both large-scale coffee houses such as Starbucks and local single-store coffee shops continue to boom in popularity and patron loyalty. World consumption of coffee is currently projected to increase from roughly 9.9 billion kilograms in the 2020/21 coffee year to 10.2 billion kilograms in the 2021/22 coffee year (ICO [4]). With this growing market and increase in popularity, consumers have begun to find identity in their chosen coffee brands. A sense of pride is established in a person's claim to being a Starbucks loyalist or a small-shop artisanal coffee connoisseur. Hence, several distinct groups of coffee-drinkers have naturally formed. And with the advent of third wave coffee, a higher emphasis is being placed on higher quality coffee. As marketers now try to bring this higher quality coffee to the masses, opinions toward tried and true establishments with often cult-like followings might be subject to change.

The mass consumption of coffee has led to many studies regarding coffee in terms of its consumption and marketplace. A 2018 investigation into the existing research done regarding coffee consumption found that most researchers looked into the sustainability and ethical aspects of coffee consumption (Samoggia and Riedel [9]). Following that topic, the next most researched area revolved around "consumer behavior towards coffeehouse brands, mainly Starbucks" (Samoggia and Riedel [9]). This area of research aligns with the focus of this study. The investigation herein uses text and network analytics of Twitter data to understand consumer identities and sentiments toward two of the largest coffee brands in the United States: Starbucks and Dunkin' Donuts. Additionally, through tweets posted about each of these brands, this study aims to identify common attributes of the critics and proponents of the two brands, respectively and comparatively, and how these consumers interact and relate.

Before diving into the details of the project objective, a deep literature review is performed to understand the applications and background of the topic and what methodology could be considered. Furthermore, researching previous studies aid in understanding limitations to the study and provide motivation for why this research is important. The literature review provides use cases for the study and a foundation of previous work to grow off of.

One application of this study is to understand marketing directions that a company (namely, Starbucks or Dunkin' Donuts) could apply to better reach out to both their loyal fans and their critics. By targeting these two groups in unique ways, marketing efforts can be optimized. Understanding consumer perceptions of brands is an essential component of marketing strategy that marketing managers have long relied on to inform marketing strategy (John et al. [5], Lehmann et al. [6]). Because of this, developing improvements to perceptual mapping techniques, a method of visualizing how customers rank a company holistically and comparatively with other companies, have been a priority for marketing researchers (Bijmolt and van de Velden [2]). However, manually collecting data about consumer perceptions has historically been very difficult and expensive due to the need for recruiting study participants and maintaining participant attention and cooperation throughout the study (Steenkamp and Van Trijp [10]). Considering these challenges, the rise of social media has laid a pathway for new methods of data collection.

Social media has redefined the marketing landscape by changing how information about consumer perceptions is collected. Now, marketers can examine consumer opinions by collecting textual evidence about thoughts toward and interactions with the product. Not only can marketers absorb information about their consumers, but they can also use social media to distribute messages to their consumers about their products and services. To investigate this, a 2021 study analyzed the Twitter networks of two multi-brand retailers to investigate network activities and network patterns based on centrality and betweenness (Watanabe et al. [14]). The study utilized network statistics to examine the ego networks (brand-centered) and whole networks (hashtag networks) produced via social media communications (Watanabe et al. [14]). The findings revealed that although brands are able to deliver messages through digital platforms, they have limited control over the communication within networks, and thus cannot govern the spread of information (Watanabe et al. [14]). This suggests that marketing teams need to dive deeper into optimizing their methodology for effectively spreading information on social media. One way to do this would be through the collection of consumer thoughts regarding their product and producing material targeted at certain attributes of groups. Thus, understanding this idea is critical to the ways in which this study could be applied. One way this study plans to apply the findings of the aforementioned literature, is to analyze sentiments towards Starbucks and Dunkin' Donuts consumers and whether negative or positive sentiments are clustered together based on group attributes. This could aid and validate the use of market segmentation strategies in favor against a "one size fits all" approach. For instance, when determining which characteristics individuals in different groups should have, and when curating unique messages to advertise to these different groups of consumers.

While marketers may not be able to control the inter-communications of their consumers regarding their product, it is clear that social media has revolutionized how information in a networked environment is received and disseminated. In particular, Twitter has generated a great deal of attention for its ability to broadly propagate information to a large audience. Information diffusion, the process of how information is spread and interacted, has been widely studied in the field of network science. Information is often diffused in conjunction with a major event such as an earthquake or a political demonstration (Sakaki et al. [8], Beguerisse-Díaz et al. [1]). Studies following the dissemination of information after a major event map the network of tweets, mentions, and retweets to identify the popularity and measure influence. While this may seem straightforward, tweets are not always indicative of a user's connection with a topic or their community identity due to the noise produced by many personal messages, jokes, and fanfares. Therefore, there is a need to transform into a more structured network in terms of "community" (Myers et al. [7]). This idea is important to keep in mind as a limiting reagent to the study here, as analyzing Twitter networks must be done carefully and very analytically to filter through the noise and gain true insights.

Although the detection of communities may be challenging, understanding the sentiments of consumers has the potential to provide valuable insights about the consumer perspectives on products. Sentiment analysis has long been used in the commercial space, especially on product reviews and predicting future buying behaviors. One study on this topic compared tweets containing two different hashtags related to weight loss (#thinspiration and #fitspiration) in order to understand how users' views on health and fitness differed for each group (Tiggemann and Zaccardo [12], Tiggemann et al. [13]). Analysis of the sentiments of each set of tweets showed that on average, both #thinspiration and

#fitspiration tweets were mildly positive in sentiment but that #fitspiration tweets were significantly more positive than #thinspiration tweets (Tiggmann & Zac-cardo 2018). This is consistent with the findings that much of the text on #fitspiration Instagram imagery is positive, suggesting that the text may be the source of inspiration that people feel (Tiggemann and Zaccardo [12], Tiggemann et al. [13]). Considering this study, it is established that sentiments can be evaluated and useful in comparing tweets on two different yet similar topics. This validates the basis of this research topic.

Sentiment analysis also plays a role in building quantitative measures of users' attitudes towards brands as well as underlying business components such as product, website, support and customer service. Another study analyzed the opinion of 19 million Twitter users towards 62 popular industries, encompassing 12,898 enterprise and consumer brands via sentiment analysis of 330 million tweets over a period of one month (Hu et al. [3]). The study found that users tended to be more positive or negative when interacting with brands than in general on Twitter but that the sentiment towards brands between and within industries varied greatly (Hu et al. [3]). For example, airline industries produce very high negative sentiments and Automotive industries provide high positive sentiment (Hu et al. [3]). By analyzing sentiment towards topics associated with users' brand interaction tweets, the authors found that highly negative topics included "Fox News" while highly positive topics include "Video Games" and "Music" (Hu et al. [3]). The results of this study suggest that evaluating sentiments about products on Twitter produce valuable insights and it is important to compare brands within the same industry. Again, this solidifies the direction of this research project's ideation.

Before diving more into the basis of this project, the world-wide nature of coffee as an industry was investigated to confirm the international nature of this consumer product. One study looked at the dynamics of the international coffee trade network using data from the World Trade Organization from 1996 to 2017 (Sujaritpong et al. [11]) . They found that more than 82% of countries participated in the coffee trade network but only 3% of countries controlled more than 90% of international trade (Sujaritpong et al. [11]). Understanding the vast nature of the coffee industry is important to validate that this topic is a good representation of how tweets can be used to understand consumer sentiments toward products. Thus, the methodology of this study could be replicated with other major consumer products to gain insights about them.

With a review of literature validating the ideation of this project, a firm objective can be stated. This research study aims to understand and compare consumer sentiments and identities surrounding two popular coffee brands: Starucks and Dunkin' Donuts. Several questions are sought to be answered. Are there distinguishable communities in a coffee-based social network? If so, what characteristics do they share? Are Starbucks/Dunkin' drinkers more/less positive/negative? Do people prefer Starbucks or Dunkin' Donuts? Is there an "influencer" effect on coffee brand preference? These questions will be answered through a network analysis of ten thousand tweets about each of the two brands. For each tweet, metadata is collected regarding the tweet user's follower counts, the tweet's retweet counts, and more. Then, text analysis is used to score each tweet's sentiment (negative, neutral, positive). Finally, a network is created connecting tweets by their commonalities. Network analysis is performed to understand each brand's clientele and investigate commonalities and differences between the brands and their consumers' sentiments and attributes. The results provide insights into these brands and a template for similar analysis of other brands.

References

[1] Mariano Beguerisse-Díaz, Guillermo Garduño-Hernández, Borislav Vangelov, Sophia N Yaliraki, and Mauricio Barahona. *Interest communities and flow roles in directed networks: the Twitter network of the UK riots.* Journal of The Royal Society Interface, 11(101):20140940, 2014.

[2] Tammo HA Bijmolt and Michel van de Velden. *Multiattribute perceptual mapping with idiosyncratic brand and attribute sets.* Marketing Letters, 23(3):585–601, 2012.

[3] Guoning Hu, Preeti Bhargava, Saul Fuhrmann, Sarah Ellinger, and Nemanja Spasojevic. *Analyzing users' sentiment towards popular consumer industries and brands on twitter.* IEEE International Conference on Data Mining Workshops (ICDMW), September 2017.

[4] ICO. *Coffee Market Report*, June 2022.

[5] Deborah Roedder John, Barbara Loken, Kyeongheui Kim, and Alokparna Basu Monga. *Brand concept maps: A methodology for identifying brand association networks.* Journal of marketing research, 43(4):549–563, 2006.

[6] Donald R Lehmann, Kevin Lane Keller, and John U Farley. *The structure of survey-based brand metrics.* Journal of International Marketing, 16(4):29–56, 2008.

[7] Seth A Myers, Aneesh Sharma, Pankaj Gupta, and Jimmy Lin. *Information network or social network? The structure of the Twitter follow graph.* In Proceedings of the 23rd International Conference on World Wide Web, pages 493–498, 2014.

[8] Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo. *Earthquake shakes twitter users: real-time event detection by social sensors.* In Proceedings of the 19th international conference on World wide web, pages 851–860, 2010.

[9] Antonella Samoggia and Bettina Riedel. *Coffee consumption and purchasing behavior review: Insights for further research.* Appetite, 129:70–81, 2018.

[10] Jan-Benedict Steenkamp and Hans Van Trijp. *Attribute elicitation in marketing research: A comparison of three procedures.* Marketing Letters, 8(2):153–165, 1997.

[11] Ookrit Sujaritpong, Sikarin Yoo-Kong, and Pradeep Bhadola. *Analysis and dynamics of the international coffee trade network.* Journal of Physics: Conference Series, 1719 (1):012106, 2021.

[12] Marika Tiggemann and Mia Zaccardo. *Strong is the new skinny: A content analysis of# fitspiration images on Instagram.* Journal of health psychology, 23(8):1003–1011, 2018.

[13] Marika Tiggemann, Owen Churches, Lewis Mitchell, and Zoe Brown. *Tweeting weight loss: A comparison of# thinspiration and# fitspiration communities on Twitter.* Body image, 25:133–138, March 2018.

[14] Nicholas M. Watanabe, Jiyeon Kim, and Joohyung Park. *Social network analysis and domestic and international retailers: An investigation of social media networks of cosmetic brands.* Journal of Retailing and Consumer Services, 58:102301, 2021.