Master Thesis

How Fintech Has Affected the Public Discourse About Banking: Evidence from Twitter

Authors:

Alan Joonatan Rebane

Perur Ramasanthosh Kumar Reddy

Supervisor:

Mark Alexander Conley

May 2021

Stockholm

Abstract

The purpose of this paper is to understand how fintech has affected the public discourse about banking. The focus area of scholars has mostly been on technical and practical aspects of fintech and largely the public discourse around it has been overlooked. However, many prominent economists have argued that markets are social constructions and the dynamics of the social discourse is reflected in real business activity. By analysing over four million tweets on the Twitter social media platform we try to explain what dominating discourse the fintech movement brings about and how it has changed the public discourse. First, we conclude that since 2016 the number one topic in banking-related discussions on Twitter has been fintech; second, fintech tweets are generally associated with a significantly higher positive sentiment; and third, the positive sentiment of fintech has increased over the years. We add that the public discourse about banking is now mostly controlled by users that are promoting and advancing fintech. We argue that the immense popularity and positive sentiment of fintech enables new financial technology firms to gain power over the incumbents due to the shifting power relations in the market. This in turn could lead to structural changes in the banking market. However, we caution that even though the social stance is supportive towards fintech, the regulation regarding the new technology is largely in development and once enforced, it may favour incumbent banks to maintain control.

Acknowledgements

First and foremost, we would like to thank our supervisor assistant professor Mark Alexander Conley for guiding us in our research. There were times when we faced difficulties in providing statistically and academically sound results and arguments that were in line with the research question that we set out to discover, but Professor Conley assisted us in discovering the methods that can provide these rigorous and robust results. We are also grateful for his encouragement and interest regarding our work and that he was quick to respond to our inquiries.

We would also like to thank Lea Vatsel for supporting us in our process, providing critical feedback, and proofreading our thesis. Comments and critique from her were especially valuable, because they helped us break free from the vacuum of our own ideas and vision.

Lastly, we would also like to acknowledge Steffen Hartwig for taking the time to provide critique and proofread the thesis.

TABLE OF CONTENTS

1 Introduction

New information technology provides many opportunities to reimagine how we use financial services. Due to large databases and intensely frequent information exchange, it is intuitive to think why finance and banking are subject to transformation by the new emerging technology known as fintech. The attention fintech has received can be justified by the ever-increasing investments flowing into fintech companies (Statista, 2020) and its potential to change the whole financial industry (Gomber et al., 2018; Schneider et al., 2016).

A vast research body has tried to understand how exactly financial technology will trigger changes. For instance, Gomber et al. (2018) studied how the operations management in finance will change; Blakstad and Allen (2018) examined the potential of fintech to serve the customers who are underserved by banks and Romanova and Kudinska (2016) outlined suggestions for commercial banks in coping with the immense innovation. Overall, the focus area of scholars has mostly been on technical and practical aspects of fintech.

On the other hand, many prominent economists have argued that markets are social constructions and the dynamics of the social discourse produces much of what occurs in real business activity. For example, Fligstein (1996) argues that when markets are in a phase of transformation, invaders can introduce social movements that change the existing conception of control. Furthermore, Rindova et al. (2006) describe how firms take nonconforming actions and actively manage their public image in order to achieve a celebrity firm status in a market. Loasby (2000) adds that markets are products of evolutionary processes and unintended consequences of human action. Following these insights, we propose to study how the banking market has been shaped by analysing the changes in social discourse about banking and fintech.

Today, online social media platforms have become hugely important in our society and some even argue that they have become the avenue where public opinion is formed (Gayo-Avello, 2013, Anstead and O'Loughlin, 2014). The history of social interactions and online behaviour that social media giants such as Twitter and Facebook possess "can open up a new era in social and behavioural sciences" (Golder and Macy, 2012). Moreover, social media enables researchers to observe expressions and interactions on a large scale and in real-time (McCormick et al., 2017). Building on these foundations, we will also use Twitter to extract intelligence and study the changes in public opinion and discourse. In this paper, we will focus on how the social discourse regarding fintech and banking is changing and the potential effect it has on the banking market.

To group tweets (public posts) on Twitter, hashtags are often used (e.g., #MondayMotivation, #Wimbledon, #crypto). In this paper, we use the hashtags #banking, #banks, and #bank to group all banking-related tweets from the period of 2009-2021, in total a little more than 4 million tweets. We attempt to gauge the differences between banking- and fintech-related tweets that have used either of the three hashtags. As fintech is an umbrella term for all the technologies employed in finance, we are specifically interested in fintech that is associated with banks. Thus, we chose to focus on these three hashtags and specifically on the subset of tweets that are about fintech.

1.1 Purpose and Research Question

The purpose of this paper is to understand how fintech has affected the public discourse about banking. Relying on evidence from academic papers and actual business activity, we argue that the banking market is going through a transformation. Furthermore, we intend to analyse the transformation through changes in public discourse by using the Twitter social media platform. As we will later show, public discourse can be an indicator of changes in the conception of control, which in turn implies changes in the power and market structure. Our research will focus on discovering what new dominating discourse the fintech movement is bringing about through social movements. We are going to use several natural language processing methods to process the tweets and answer the following research question:

How has the emergence of fintech affected the public discourse about banking among Twitter users?

1.2 Expected Contributions

First, we hope to contribute to the academic discussion around how markets are shaped by providing a methodological framework to study the changes in social discourse. Second, since public discourse around banking is documented historically and on a large scale by Twitter, we hope to contribute to the theory of a market as a social construct by providing empirical evidence for the changes that have taken place in the public discourse as the banking market is going through a transformation.

2 Literature Review

In this section, we will discuss relevant concepts and literature concerning fintech and banking. We provide evidence for how the banking market has been historically shaped and how it is happening right now. After that, we will explore what role social movements play in shaping markets and how can we study them using the Twitter social media platform. Lastly, we will pose two hypotheses that address the theory.

2.1 Providing a Definition for Fintech

To start with, fintech is a popular term and rightfully so, because fintech or financial technology has had a significant impact on how consumers, companies, and organizations use and interact with financial services (Schueffel, 2016). Schueffel (2016) attempted to give a single definition for fintech by exploring and reviewing more than 200 scholarly articles that refer to the term. Schueffel (2016) concluded: “fintech is a new financial industry that applies technology to improve financial activities.” When referring to fintech in this paper, we refer to fintech companies that are creating this new financial industry rather than a specific technology that is implemented.

Broadly, fintech has influenced all kinds of financial services through innovations such as peer-to-peer lending, online verification, robo-advisory, crowdfunding, and internet banking (Schueffel, 2016). However, in this paper, we will specifically focus on the relationship between fintech and banks. Chishti and Barberis (2016, p. 7) and Blakstad and Allen (2018, p. 2) point out that big banks, such as the major retail banks, are vulnerable to disruptive innovation partly due to degrading trust resulting from the most recent financial crises. Also, Baba et al. (2020) mention that competition in traditional banks is pushing banks to adopt financial technology or acquire such companies that provide it, which in turn causes these banks to look more like fintech companies. Thus, many interesting dynamics are at play in the banking market, which has a long history and has been subject to significant regulation and frequent state intervention.

One of the findings of Sangwan et al. (2019), is that fintech is experiencing a rapid and uninterrupted development through product as well as process innovation, transformation and disruption, and fintech academic research is still in its early stages. Therefore, performing more research with a focus on fintech is very important given its potential to disrupt the financial industry. In a review by Das (2019), the explosion of fintech has been attributed to a combined effect of the advancements in the following fields: big data, computing technology, cryptography, psychology, computer interfaces, linguistics, statistics, mathematics, and econometrics. In all the fintech-related literature reviews by Sangwan et al. (2019), Das (2019), and Allen, Gu and Jagtiani (2020), the focus areas of scholars were on technical and practical aspects of fintech, and largely the public discourse around fintech has been overlooked.

2.1.1 Evolution of financial technologies

Physical media such as paper and coins were used as storage of value and information, and they were the initial forms of technology used by banks and trading companies. These markets were limited by geography, as they required the physical movements of documents for the transfer of value and information (Alt et al., 2018).

Later, technological inventions and advancements, such as the electrical telegraph, enabled the transmission of information without actual physical movement of objects. This usage of analogue technologies lasted until the mid-twentieth century that could be termed as the second phase of financial technologies (Alt et al., 2018).

The age of digital financial technologies, the third phase, started with the introduction of digital information and communication technologies (Gomber et al., 2018). Since the 1960s, banks have created IT departments with thousands of employees for the in-house development of IT. The functions of these departments included the creation of proprietary applications, maintenance of internal networks, consumer-facing interfaces like ATMs, and online banking. In-house IT departments were expensive to maintain, among other inherent inefficiencies. To alleviate these inefficiencies with the proprietary systems, many IT vendors like SAP, Finastra, Temenos, etc., came up with packaged software systems (Alt et al., 2018). By the late 1980s, financial institutions, customers, and market participants started transacting digitally and the financial services transformed majorly into the digital industry. According to Lee and Shin (2018), the internet revolution in the early 1990’s drastically affected the financial markets all around the world. One significant outcome of this was a lowering in financial transaction costs.

According to Alt et al. (2018), it was during the global financial crisis of 2008 when the current (fourth) fintech phase started. This phase was built on the evolution of technologies such as web technologies, mobile devices, and wireless networks. Their convergence was accompanied by the customer-oriented innovation and entrepreneurial spirit, which did not have the same presence previously. In their paper, Alt and Puschmann (2012) summarized that this phase of financial technologies was built on four driving forces.

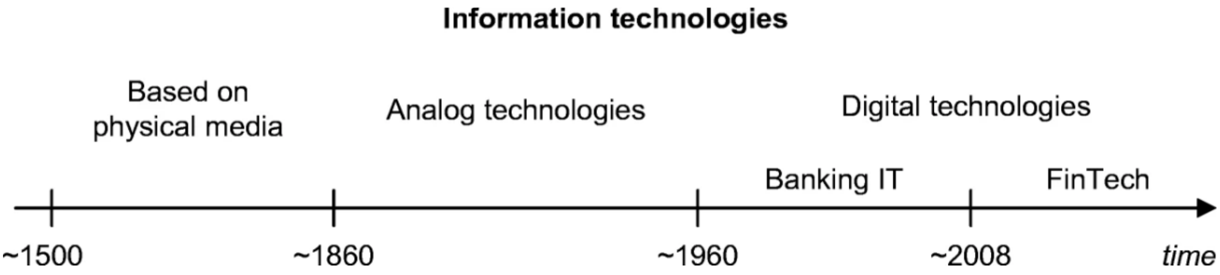
First among these driving forces was the rapid innovation in downstream IT solutions. The second was the emergence of start-up companies and non-banks that are providing financial services. To compete with these newcomers, established financial services providers set up their own new internal departments to resemble start-ups, (e.g., spin-offs and innovation labs). As Alt and Puschmann (2012) conclude, even though these units were established, they could not emulate the similar creativity and dynamism which were the key characteristics of many successful start-up companies.

The proliferation of digital financial services and mobile devices has empowered customers to access their financial information at the time and place of their convenience. This increase in digital adoption was accompanied by the decrease in the customers' loyalty towards their primary bank and they became more comfortable with using multiple financial service providers. This shift in the customers' behaviour towards being at ease with online banking and favouring multiple banking relations was observed as the third driving force (Alt and Puschmann, 2012).

The fourth driving force was identified as the new regulations, which were enforced because of the financial crisis of 2008 (Alt and Puschmann, 2012). These new regulations included the separation of investment banking and retail banking. (e.g., Dodd-Frank Act), higher capital coverage requirements (e.g., Basel agreements), protections for markets and consumers (e.g., MiFID), and fraudulent behaviour reporting schemes (e.g., FACTA, AIA). These new regulations increased pressure, mainly on traditional financial service providers.

Bons et al. (2012) and Alt and Puschmann (2012) conclude that the potential and possibilities offered by internet technologies and mobile devices are just the beginning to be identified by the banks. Figure 1 from Alt et al. (2018) explains the evolution of financial technologies.

Figure 1. Evolution of Information Technologies



Note. Figure courtesy of Alt et al. (2018).

2.2 Characteristics of the Banking Market

From the 1933 Glass-Steagall Act (Federal Reserve Bank of St. Louis., 1933), banking activities in the U.S. can be broadly divided into two: commercial banking, which is related to consumers, and investment banking, which typically deals with advising large corporations and states. Due to the implementation of the single-market economy, the adoption of the euro, and country-wide differences in Europe, the European banking market cannot be classified as easily and banks can often pursue both commercial and investment banking activities. Moreover, the definition of the banking markets can be different in different parts of the world. However, Gual (1999) describes the three main functions of the banking system:

1. To decrease transaction costs in payments by eliminating the need to verify the solvency of the parties transacting.
2. To collect liquid deposits and give out long-term credit.
3. To transform the quality of assets by diversifying their investments and loans in different assets.

Chishti and Barberis (2016, p. 8) offer a similar explanation for the functions provided by retail banks. The importance of the main functions has varied over time and often shifted in the level of relevance when changes in the banking market have taken place. Nevertheless, these are the three main characteristics that we will rely on when defining the banking market.

Also, though this thesis focuses mainly on financial technology, Baba et al. (2020) and the Financial Stability Board [FSB] (2019) draw attention to the fact that fintech has had a minuscule effect on the provision of long-term credit, which after all is one of the main characteristics of the banking market. Overall, the relationship between incumbent banks and new fintech firms is currently mainly cooperative and complementary in nature (FSB, 2019).

2.2.1 Operationalizing market shaping in the banking market

Changes in the banking market are fairly frequent and historically most often brought about by state institutions (as in, for example, Goddard and Molyneux, 2007; Gual, 1999; and Rajan, 1998). That is, the banking market can be characterized by those institutions imposing changes and banks having to adapt to these new conditions. Loasby (2000) explains that, in general, the evolution of markets can be explained by “both evolution within institutional constraints and the evolution of institutions.” Earlier, the development of the banking market was mostly determined by the evolution of state institutions rather than the evolution within the institutional constraints since most of the changes were brought about by state actors. Due to the role of banks in modern economies in providing stability, it is easy to understand why these financial establishments have been subject to constant regulatory pressures from the state (Bofondi and Gobbi, 2017).

Furthermore, DiMaggio (1989) also explains that a market is the result of an institutionalization project which is equivalent to realizing a shared conception of control. Therefore, the changes in the conception of control in the banking market can give insights into how the main functions have changed and how the market has been shaped over time. Defining the conception of control – which is the shared understanding of how a market must function – is inherently a political undertaking where the most powerful firms have the most say (Fligstein, 1996). Chishti and Barberis (2016, p. 7) argue that many innovative solutions that have influenced the overall service provision have indeed been brought about by new fintech companies and later been followed by incumbent banks. This suggests that the conception of control might be shifting and changes are now taking place within the institutional constraints as opposed to the evolution of institutions.

However, caution must be made when rushing to the conclusion that the status quo is being changed in the banking market. After all, the hype cycle of new technology can lead one to make substantial claims at the peak of inflated expectations (Dedehayir and Steinert, 2016). Due to the access to longitudinal data on the public discourse about banking and fintech, we can hopefully overcome this as we can study the changes rather than a fixed point in time. What is more, to date, fintech firms have not had access to the same low-cost funding or the customer base which incumbent banks enjoy. Thus, fintech firms are currently not posing a serious competitive threat to incumbent financial institutions in mature markets (FSB, 2019).

To sum up, the changes in the banking market have historically mainly resulted from the evolution of state institutions. By redefining the conception of control, new firms can change the overall power structure and shape the banking market within the current legal landscape. Also, currently fintech firms do not pose serious competition to incumbent banks and when claiming that the conception of control is being changed, one must critically consider whether it is hype that leads to these conclusions. Next, we will provide some more evidence on how the banking market has been shaped historically and how it is different this time.

2.2.2 Previous examples of market shaping in the banking industry

Economic historians can point to the events in the past when significant market transformation in the banking sector took place. Most notably, in the US, the Glass Steagall act in 1933 ended much of the financial innovation that took place before the Great Depression (Bofondi and Gobbi, 2017). Interestingly, the regulators were trying to limit innovation and competition in favour of stability, but in the 1980s, the stability mindset began to weaken and a series of deregulations followed (Bofondi and Gobbi, 2017). Fabocci and Modigliani (1992) explain that in the 1970s and the 1980s, the US government passed many laws that facilitated the securitization of mortgages and enabled banks to deal with complicated derivatives which later escalated beyond control and lead to the worldwide economic recession in 2008.

These observations refer to the U.S. banking market, but Fligstein (2013) explained that at that time, the conception of control in the banking market internationally was to invest into U.S. mortgages by buying them and issuing U.S. mortgage-backed securities. Fligstein (2013) explained further that not only American banks, but also other international banks were pursuing the same practices in similar volumes of issuing mortgage-backed securities and financing them from short-term loans. In a sense, the practice of dealing with mortgage-backed securities was a common understanding of how the banking market operated in the early 2000s and what a bank must do to be successful in that market.

In retrospect, the motivation behind imposing the Glass Steagall act in 1933 and the deregulation of money-market funds in the 1980s can be explained by governments wanting to influence the economy. Also, these observations strengthen the argument that historically, the evolution of institutions has guided the evolution of the banking markets, such as in Loasby (2000).

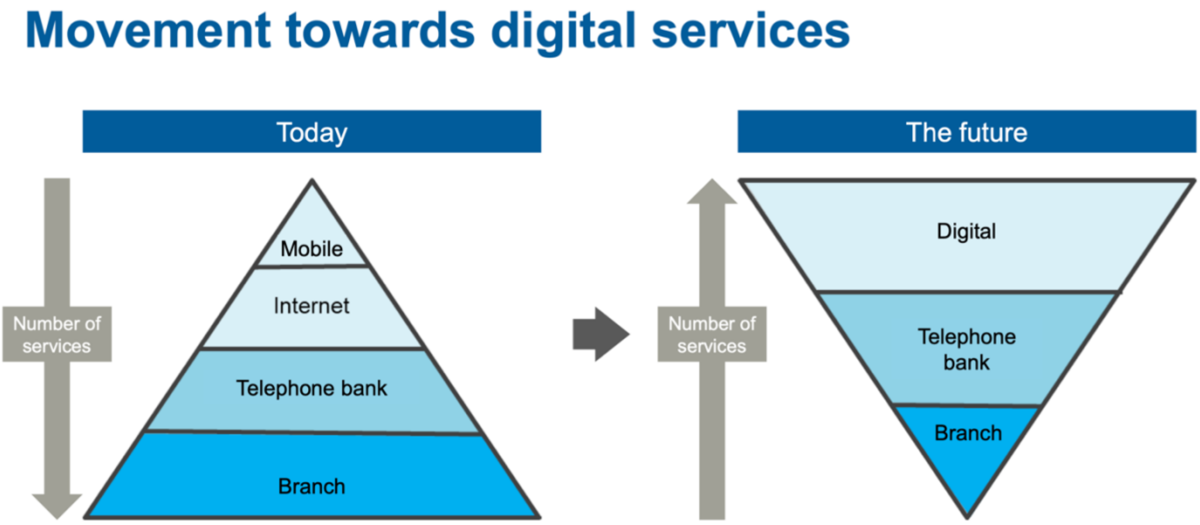
The current transformation phase, however, is different from the ones elaborated on above. For instance, Romanova and Kudinska (2016) report that the current state of the market can be explained by new fintech companies invading the banking market. The “fintech revolution” is also a popular term that is being used to describe the current state of the industry, implying that the change is brought about by new financial technology rather than a state trying to impose a change (Blakstad and Allen, 2018; and Gomber et al., 2018). Thus, the way the banking market is being transformed currently is different from the historical observations in that the evolution of the market happens inside the institutional constraints. Next, we will provide empirical evidence from actual business activity to strengthen the case that the banking market transformation is brought about by new fintech companies.

2.2.3 Empirical evidence of transformation in the banking market

Several signs from actual banking activity point to the fact that the banking market is indeed in a transformation phase. For instance, banks in Europe and the U.S. are closing branches in bulk. Handelsbanken in Sweden is set to close nearly 50% of its branches in 2021 and the bank justifies its actions by stating that digitalisation among customers has now reached a point where physical branches have become redundant (Handelsbanken, 2020). Handelsbanken (2020) adds that the customers show increasing demand to perform transactions digitally and thus the bank is going to invest one billion SEK in IT in the near future. Rowan (2020) reports that Wells Fargo, one of the largest banks in the U.S., has plans to lower the number of branches from a total of 5400 to 4000. The closing of branches points to the overall trend of banking activities going digital and opening opportunities for technological advancements to reimagine the banking activities. Alt et al. (2018) explain that fintech companies have used this opportunity to transform the focus of banking activities from an internal business processes perspective to a more customer-centric perspective.

In line with this, the CEO of Handelsbanken has also shared about their change in strategy from physical banking to mobile banking, adding, “We have reached a tipping point where we see that our customers are opting more to meet us in a digital space or other forums” (Reuters, 2020). Figure 2 represents the new strategy of Handelsbanken.

Figure 2. Strategy change at Handelsbanken



Note. Figure courtesy of Handelsbanken (2020)

As of writing this paper, public companies specializing in financial technology have received very high valuations for their equity stake. For instance, the equity of Square Inc., a fintech company that facilitates mobile payments and offers financial services for SMEs and consumers in the U.S., is valued at around 113 billion USD (as of April 30, 2021). In contrast, the equity of Deutsche Bank, one of the largest banks in Europe, is valued at 29 billion USD (as of April 30, 2021). At the same time, Square Inc. reports 10 billion dollars of assets on its balance sheet, whereas Deutsche Bank possesses 1.55 trillion dollars of assets. Granted, the business models are not the same for the two companies as Deutsche Bank has a much more comprehensive range of operations and a lot more liabilities, but this stark contrast in valuation is evident in many other fintech and bank comparisons as well. The evidence from the stock market, which is inherently forward-looking, hints that investors are much more optimistic for the outlook of fintech companies than they are for traditional banks. Also, Statista (2020) reported that the total value of investments into fintech companies worldwide from 2010 to 2019 increased from 9 billion USD to 135.7 billion USD. A stark increase of investments into fintech was also observed in 2018 when the total investments into fintech grew 120% (Statista, 2020).

Whereas these facts point to the changes in the banking market in Europe and the U.S., where financial development is high, the emergence of fintech in Asia and Africa also supports the argument that fintech is transforming the banking industry internationally. Baba et al. (2020) report that big tech lending, which refers to lending activities by large technology companies, is the largest in terms of volume in South Korea, Japan, and China. Baba et al. (2020) also show that in Africa, where bank credit to the private sector is the lowest in the world, mobile money transactions made up over 25% of GDP in 2018. Overall, the impact of fintech in different parts of the world is quite dissimilar and it is more pronounced in regions where the overall financial development is lower (Baba et al., 2020).

Alt et al. (2018) explain explicitly how fintech is transforming the banking market on the internal organization level, the business network level, and at the external organization level. During a market transformation phase, new firms try to redefine the market as they have the opportunity to redefine the existing conception of control (Fligstein, 1996). Fligstein (1996) exemplifies that during market transformation, when the internal power struggles are most intense, challenger firms can introduce more fluid social-movement-like conditions to change the existing understanding of the market. Interestingly, when incumbents begin to fail and lose power, they start to reconstruct themselves to resemble the invaders (Fligstein, 1996). McAdam (1982), Snow et al. (1986), and Tarrow (1994) add that the success of these social movements during market transformation depends on multiple factors:

1. The size of the social groups
2. The resources they have access to
3. Whether they have the political opening
4. State actors willing to comply with the groups' complaints
5. The capacity to build a social coalition around a new collective identity

To conclude, though there is plenty of evidence that fintech has altered the banking market, we have yet to understand whether it has gone through fundamental changes where the power has shifted and the conception of control changed. In the following few chapters, we discuss how social movements can be studied by analysing the public discourse and possibly shed light on how the public discourse on social media is reflected in actual activity.

2.3 Background: Public Discourse, Social Movements, and Twitter

Ultimately, considering markets as politics, such as in Fligstein (1996), one can make sense of the market as a social structure, where this structure can only be changed through social movements. Furthermore, this also implies that by changing public discourse through these movements, the power relations in the social structure are also changed, which in the end alters the politics of power (Woodly, 2015). Woodly (2015) notes that “a movement that effectively alters the terms of discourse can overcome considerable opposition and structural disadvantages to achieve sustained, meaningful change.” Considering markets as social structures is therefore also the reason why Fligstein (1996) considers internal power struggles as one of the main forces of market shaping as these internal struggles reveal the overall shifting of power, which in the end gives an opportunity to a few chosen actors to impose their will on the market. Thus, studying public discourse could lead to new insights into what kind of conception of control is the new fintech movement advocating and where potentially the banking market is heading. However, as Woodly (2015) notes, changing the public discourse is an ongoing struggle and should not be thought of as a series of victories that quickly lead to the overall change in the discourse. Therefore, the change in the public discourse about banking should also not be seen as a binary process.

2.3.1 Using social media and Twitter to study public discourse

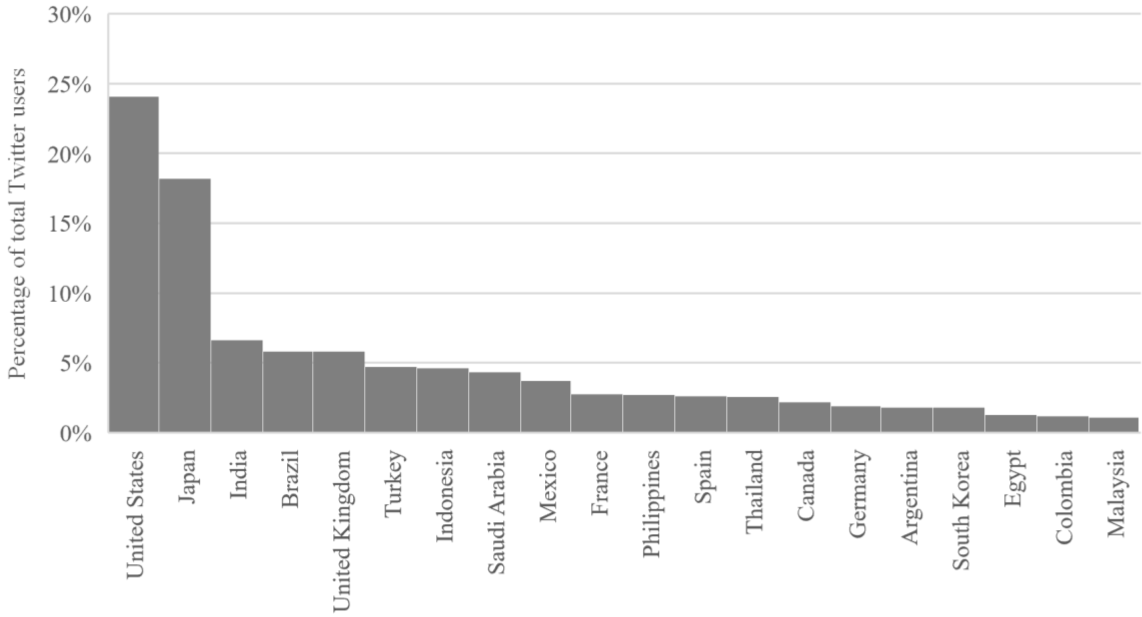
Social media has evolved into a place where people engage in discourses about various issues ranging from political uprisings (Hamdy and Gomaa 2012; Starbird and Palen 2012), sports, and natural disaster response (Vieweg et al. 2010), among others. Social media is considered by sociologists and political scientists as new avenues where public opinions are being formed (Gayo-Avello, 2013, Anstead and O’Loughlin, 2014).

In recent years there has been an increase in social movements, which consisted of both offline actions in the form of protest and online actions such as the dissemination of information and serious discussions. Also, Kou et al. (2017) found that there is coexistence and co-development of public discourses on social media and offline actions. Public discourse does not necessarily mean prominent views and statements; it is an assemblage of distinct human actors and it may be at times conflicting with ideologies, perspectives, and values (Kou et al., 2017).

In this paper, we propose to use the Twitter social media platform to analyse how the content and emotions in public discourse have changed. How exactly we will use Twitter will be elaborated further in the method section of the paper.

Twitter is one of the most prominent microblogging platforms, with 192 million active users (Twitter, 2021). Twitter is being used in over 150 countries, with three countries, the U.S, Japan, and India, contributing to nearly 50% of total users (Figure 3). Of the total userbase, 50% of the total users belong to the age group of 18-34 years, with approximately 70% of the total user base being male and 30% female (Hootsuite, 2020).

Figure 3. Twitter users by countries



Note. Figure created by the authors using data from Hootsuite (2020).

On Twitter, users can post messages of up to 280 characters, which used to be 140 until October 2018. These messages of limited size are called tweets. A message (tweet) is written by one person and read by a number ranging from zero to millions of people, called followers. Twitter messages can deal with various topics, ranging from personal news to interesting information related to various fields. The content of the tweets can be textual, visual, or links to other websites. A retweet is a feature that helps users to quickly reshare any tweet to their followers, used generally to amplify a message, and heart is a feature used to acknowledge/appreciate any tweet. The amount of information flowing through Twitter is enormous. The latest official information comes from 2014, which revealed that people sent more than 500 million tweets per day in that year (Twitter, 2014).

Furthermore, hashtags (#topic name) are a way to channel and group this information. Overall, Twitter has become a repository of large volumes of information, which opens the possibility for data mining, sentiment analysis, opinions analysis, and information retrieval (Martínez-Cámara et al., 2014).

Twitter is being used in academic research to study public discourse, perceptions, and attitudes. Also, for some time now, there is an exponentially growing interest in sentiment analysis (SA) among the research community (Martínez-Cámara et al., 2014; and Soleymani et al., 2017).

There are at least two prominent meanings for the word sentiment in conversational language. It is employed to describe (1) a feeling or to signify something emotional and (2) a specific point of view (Puschmann and Powell, 2018). Sentiment analysis, also known as opinion mining, defines a group of approaches that are used to measure sentiment, opinion, and subjectivity in texts (Liu, 2010; Pang and Lee, 2008).

One of the prominent meanings for the words sentiment in conversational language as defined by Puschmann and Powell (2018) is a feeling or an emotion. In their paper Bagozzi, R. P.(1999) defined emotion as a mental state of readiness that arises from cognitive appraisals of events and thoughts. For the same event or incident, different emotional reactions can be expected from different people. The extent to which an emotion is negative or positive is described as Emotional Valence. (Citron, F. M. et.al., 2014). In Social interactions, Emotional valence is conveyed majorly by words and facial expression (Kauschke, C. et.al., 2019). In online social interactions, which happen sans facial expressions, emotional valence is majorly conveyed by words. The emotional valence, when people write a tweet is what we will be gauging by carrying out the sentiment analysis on those words in the tweets.

Although sentiment analysis is considered a subfield belonging to computational linguistics, it is treated as a method in social science (Pang and Lee, 2008; Puschmann and Powell, 2018). Owing to the ability of sentiment scores to objectively predict the emotions of social media users and consumers, sentiment analysis is gaining a lot of popularity as a tool for analysing the discourse on social media (Puschmann and Powell, 2018).

Puschmann and Powell (2018) argue that sentiment analysis is quicker and more consistent than a human judge in being able to segregate a million posts by predefined criteria, unlike human judge who may take weeks to do the same task or develop fatigue in the process, leading to inaccuracies. For the purpose of this study, we use the definition of Go et al. (2009) to define sentiment as “a personal positive or negative feeling.”

Next, we will discuss the role of Twitter in social movements. As argued above, during the phase of market transformation, invaders can introduce social movement like conditions to alter the existing status quo. Therefore, it is crucial to examine the role of Twitter in social movements.

2.4 Social Movements and Industry Development

Using a social media platform to study social movements can reveal how these movements are scaled up (Mundt et al., 2018) and intensified (Shirazi, 2013). Also, these platforms are very easy to join and they can very effectively mobilize people to participate in such movements (Lopes, 2014; and Hwang and Kim, 2015). By using the definition of social movements by Della Porta and Mattoni (2016), there are multiple benefits of using Twitter to study social movements:

First, social movements are **mostly informal networks of interaction** (Della Porta and Mattoni, 2016). Among Facebook, Twitter, Instagram, and Snapchat, Twitter has the highest bridging social capital (Phua et al., 2017), which is associated with weak and distant relationships that facilitate information sharing and the creation of informal interactions. Informal interactions enable social movements to develop and mobilize people (Della Porta and Mattoni, 2016).

Second, social movements are **based on shared beliefs and solidarity** (Della Porta and Mattoni, 2016). Due to being able to form one's own network of people sharing the same beliefs without any complexity, users can surround themselves with others who share the same beliefs. Therefore, the formation of groups around shared beliefs is swift, sometimes dangerously so, since this can result in only a certain kind of information flowing among those user groups and causing social bubbles. Twitter has often been used to study political bubbles, for example, in Eady et al. (2019) and Bozdag et al. (2014).

Third, social movements are **mobilized around contentious themes** (Della Porta and Mattoni, 2016). Mostly due to the use of hashtags, Twitter is a good place to discover how many people and how actively do these people mobilize around contentious themes. For example, with the #MeToo movement, women united against sexual harassment and made public their own horrible experiences. Gilbert (2017) reported that the movement started on Twitter when the actor Alyssa Milano urged everyone to write "Me too" in their status if they had been sexually harassed or assaulted. Twitter confirmed that #MeToo had been tweeted nearly half a million times in the first 24 hours (Gilbert, 2017).

Fourth, social movements are characterized by the **frequent use of various forms of protest** (Della Porta and Mattoni, 2016). Using Twitter to study the fourth pillar of social movements can be difficult since 280 characters might not be a good proxy for various forms of live protests. However, Twitter has been the source and accelerator of live protests in the past. For instance, Bruns et al. (2013) show that the role of Twitter in Arab Spring is often exaggerated, but its role was still significant.

In the context of this thesis, one might argue that the emergence of financial technology is simply superior to the existing banking services and the resulting efficiency will outweigh all other factors (such as social movements) in determining the development of the banking market. Efficiency gains are certainly a prerequisite for any changes in a market as there must be a strong reason (such as national security or trade tariff) for any market participant to adopt less efficient practices in a market economy. However, we have also argued that big leaps in market structure or formation of markets are dependent on what kind of conception of control are market participants agreeing upon (Fligstein, 1996).

Consider a prominent article by Granovetter and McGuire (1998), where they outlined how the electricity industry was formed in the United States. More specifically, they described how the electric utility industry was not born of maximizing efficiency but rather from existing relationships and the active formation of new social structures. It is important to note, however, that the formation of the electric utility industry depended on a relatively small social group (Granovetter and McGuire, 1998). In this larger social structure that is considered in this thesis – the Twitter users – we can study who are the most visible users that influence the discussion around banking the most.

2.5 Hypotheses

For answering the main research question, "*How has the emergence of fintech affected the public discourse about banking among Twitter users?*", we will also pose 2 relevant hypotheses that follow the theoretical framework and aid us in the discussion.

Firstly, we noted that historically the banking market had been shaped by state actors. However, the changes that are occurring now are brought about by new fintech companies within the institutional constraints. Generally, the real effect of financial technology on banking activities is becoming increasingly more influential all over the world (Baba et al., 2020; Statista, 2020).

In addition, we noted that social media is considered by sociologists and political scientists as new avenues where public opinions are being formed (Gayo-Avello, 2013, Anstead and O'Loughlin, 2014). We have established Sentiment Analysis as the study of subjective information and emotions in text using computational methods. Relying on Kou et al. (2017) and previous studies on sentiment analysis, we also claimed that there is coexistence and co-development of public discourses on social media and offline actions. More specifically, we found that positive sentiment on Twitter was correlated with the stock market and election outcomes.

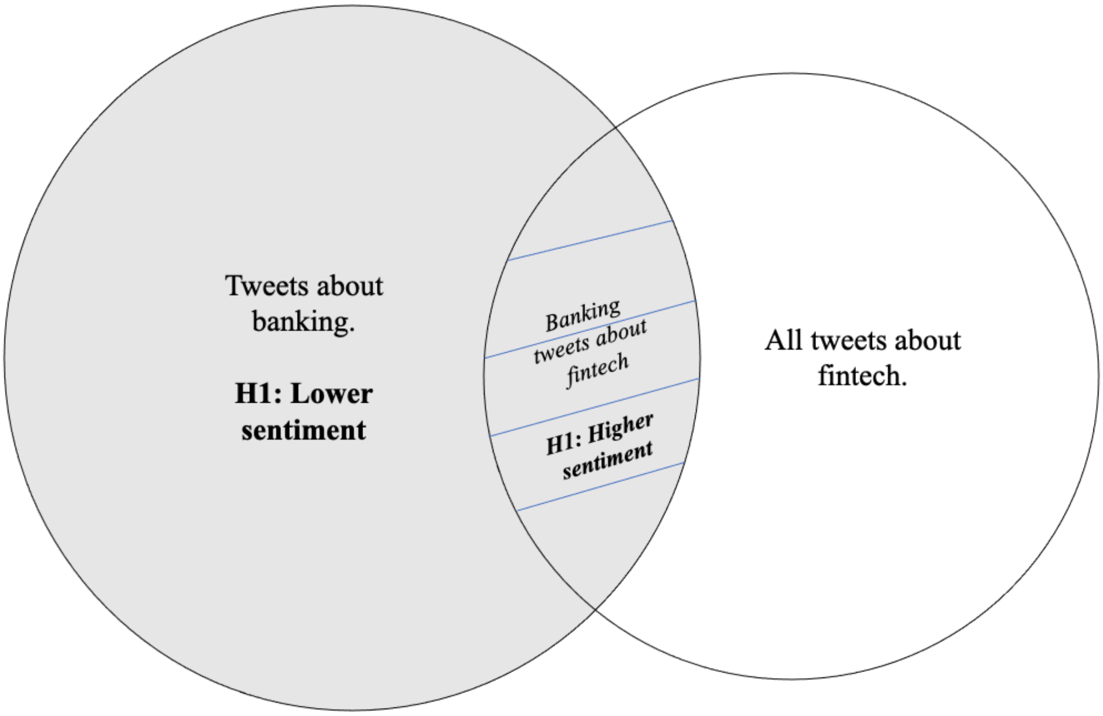
Lastly, we argued that when markets are in a phase of transformation, invaders can steer the public discourse through social movements and establish a new conception of control (Fligstein, 1996). Hence, we hypothesize that the changes in the banking market are also reflected in the public discourse about banking and a new conception of control is enforced. We hypothesize that fintech-related tweets are higher in positive emotion or stated differently, public opinion on Twitter regarding fintech is positive (H1c). Figure 4 explains the hypothesis graphically.

H1a: Compared to other banking-related tweets, there is no difference in sentiment in fintech-related tweets.

H1b: Compared to other banking-related tweets, fintech tweets are higher in lower sentiment.

H1c: Compared to other banking-related tweets, fintech tweets are higher in positive sentiment.

Figure 4. Defining the area of interest



Notes. The grey area corresponds to banking-related tweets, the white area corresponds to fintech-related tweets, and the shaded area corresponds to the overlap of banking- and fintech-related tweets. The shaded area represents fintech-related discussion in the context of banking and is the area of interest in this research. Figure created by the authors.

Furthermore, we argued that the dominance of fintech in public discourse is not a binary process, where it suddenly emerged and changed the landscape, but rather that fintech has slowly ingrained itself into the discourse and became more relevant and supported over the years. After all, Woodly (2015) also pointed out that changes in the public discourse take time and are "an ongoing struggle". Another reason why we are interested in sentiment over time is to support the first hypothesis that the sentiment of fintech-related tweets is also positive at the present time. Thus, the second hypothesis that we pose includes the time dimension and argues that over time, the sentiment about fintech has become more positive (H2c).

H2a: Among banking tweets, the sentiment of fintech has remained the same over the years.

H2b: Among banking tweets, the sentiment of fintech has become more negative over the years.

H2c: Among banking tweets, the sentiment of fintech has become more positive over the years.

To conclude, we hypothesize that among banking tweets, fintech-related tweets contain more positive sentiment, and this positive sentiment has increased over the years. Table 1 provides the summary of the hypotheses.

Table 1. Summary of Hypotheses

	Hypothesis
H1a	Compared to other banking-related tweets, there is no difference in sentiment in fintech-related tweets.
H1b	Compared to other banking-related tweets, fintech tweets are higher in lower sentiment
H1c	Compared to other banking-related tweets, fintech tweets are higher in positive sentiment.
H2a	Among banking tweets, the sentiment of fintech has remained the same over the years.
H2b	Among banking tweets, the sentiment of fintech has become more negative over the years.
H2c	Among banking tweets, the sentiment of fintech has become more positive over the years.

Note. Table created by the authors.

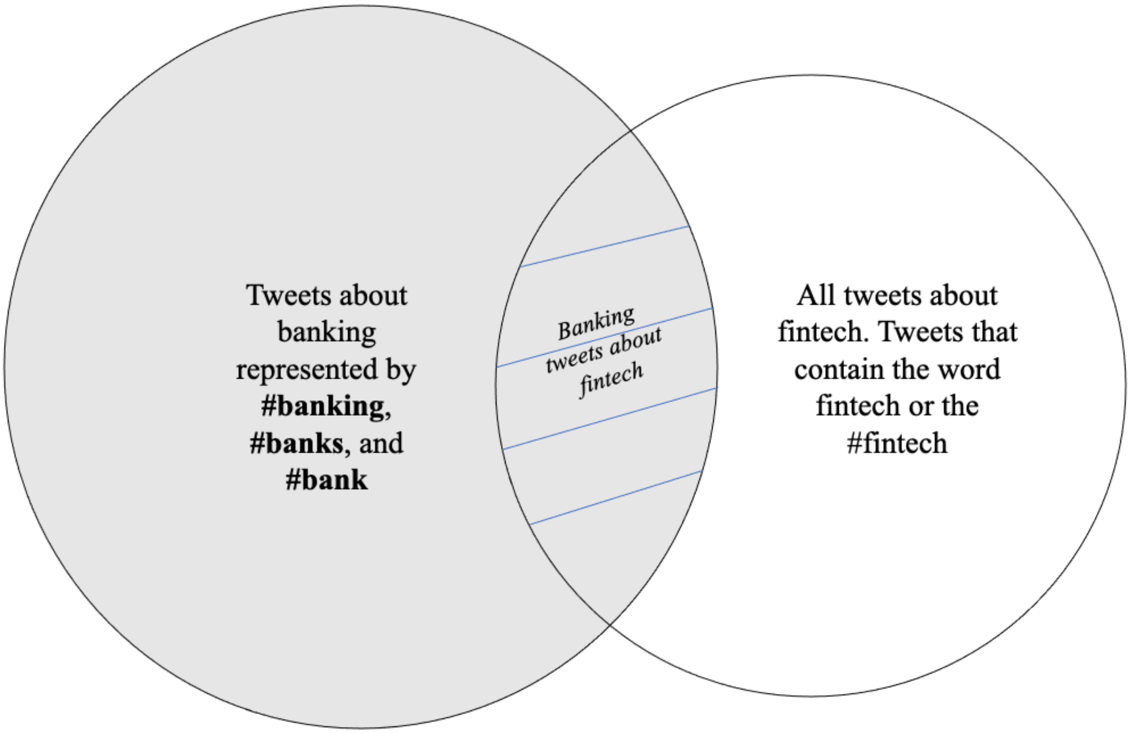
3 Methodology

The method that we use aims to answer the research question: *"How has the emergence of fintech affected the public discourse about banking among Twitter users?"* Also, with the method proposed, we attempt to find evidence to support or reject the two hypotheses.

In short, our method relies on extracting intelligence from tweets related to banking by offering visual evidence and using regression to model the relationship between the sentiment of a tweet and whether it is about fintech.

We started by looking at which banking hashtags are the most popular and extracted all tweets that contain at least one of them. We found that #banking, #banks, and #bank are the most used hashtags to represent banking activities. Going forward, we refer to these tweets as banking tweets. Furthermore, in this paper, we attempt to divide the banking tweets into two: those that contain the keyword fintech and those that do not; this is done in order to study the differences between the tweets about traditional banking and financial technology in the context of banking. Therefore, we are interested in comparing two sets of tweets: banking tweets without the word fintech and banking tweets that contain the word fintech. Figure 5, with the already familiar Venn diagram, highlights the two areas that we are interested in (coloured in grey). Notice that a tweet does not necessarily have to contain the hashtag #fintech, but it can only contain the word in order to be classified as a banking tweet about fintech.

Figure 5. Defining the research focus



Notes. The grey area corresponds to banking-related tweets, the white area corresponds to fintech-related tweets, and the shaded area corresponds to the overlap of banking- and fintech-related tweets. Figure created by the authors.

3.1 Description of Data

One of the main motivations for using Twitter data to study the public discourse is that in February 2021, Twitter made available the new academic research product track, where researchers can query the complete history of public tweets. As Twitter is one of the most popular social media platforms, the new academic product track can become a very important source of data for social scientists because it enables to study longitudinal trends and different effects of historical events. However, Twitter is also increasingly falling victim to public discourse manipulation (see for example Broniatowski et al., 2018; and Monsted et al., 2017), which makes it more important for researchers to clean their data appropriately in order to avoid noise and deliberately manipulated information.

Throughout the analysis, we have used python and its different packages to conduct the analysis of tweets. All python scripts are provided on GitHub[1]. We collected the tweets using the Twitter application programming interface (API) and queried the database by the following conditions:

- It must contain #banking, #banks, or #bank (not case sensitive).
- It must be in the English language.
- It must not be a retweet.
- It must not contain *cashtags* (*cashtags* are $ signs that are associated with tickers, e.g., $AAPL or $BTC)

The response of the API included the tweet and user attributes provided in Table 2.

Table 2. Tweet and user attributes

Tweet Attributes	User Attributes
The full content of the tweet	The username and the description
The date when the tweet was created at	The verification status
The user associated with the tweets	The date when the user was created
	Total number of followers and following
	Total number of tweets over time

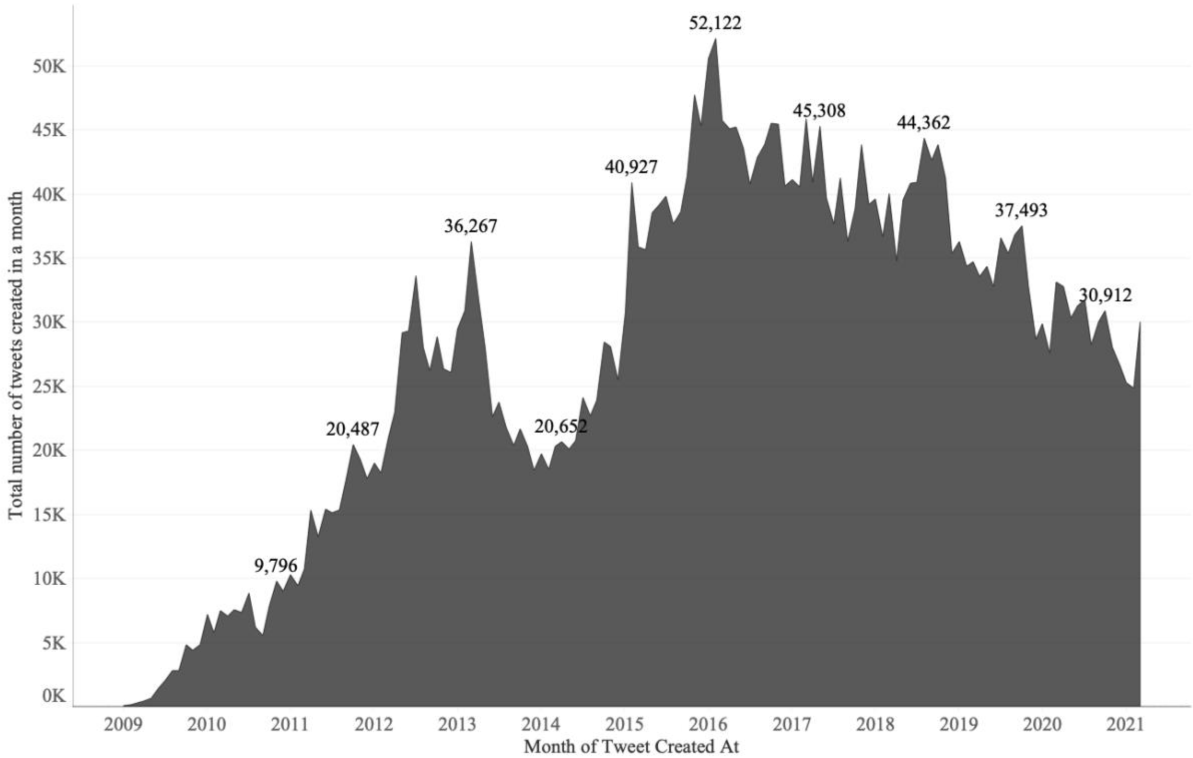
Note. Table created by the authors.

After going through some samples of the dataset manually, we found that suspicious activity, which corresponded to standardized tweets at fixed time intervals, was usually carried out by users with very few followers. For this reason, we applied a very rough exclusion parameter by excluding all users with less than 100 followers from our dataset. This reduced the number of tweets by 483,768. The final dataset consisted of 4,018,801 tweets with #banking,

[1] All the code can be accessed from this GitHub repository: https://github.com/AlanRebane/TwitterAnalytics

#banks, or #bank in the period of 1st of January 2009 till 31st of March 2021. Figure 6 shows the sum of monthly banking tweets over time.

Figure 6. Banking tweets over time



Note. Figure created by the authors.

3.2 Research Framework

Our method relies on descriptive analysis and regression modelling. First, to give relevant descriptive insights, we use the framework outlined by Chae (2014), where he used three dimensions to analyse tweets. In this paper, we divide descriptive analysis to two: descriptive analytics and network analytics. Though the developments in natural language processing have advanced since the paper by Chae (2014) was published, the fundamentals of Twitter and the API response objects have not changed significantly, therefore allowing us to use a similar approach for the descriptive analysis.

Secondly, to arrive at a statistically significant and robust conclusion regarding the sentiment of tweets and provide evidence for the hypotheses, we employ regression analysis. We will be using the generalized estimating equation (GEE) approach, developed by Liang and Zeger (1986), to model the relationship between the sentiment of a tweet and whether the tweet was about fintech or not. More specifically, we are interested in finding out if the tweets that contain the word fintech are associated with higher positive sentiment. GEE is used because in situations where a researcher faces clustered data, it is more efficient than ordinary least squares (OLS), as it enables to specify within-subject correlation and thus group together clusters, which are implicitly correlated. In the context of this research framework, we consider the clusters as Twitter users. GEE enables to control for the correlation of sentiment within the users and provide coefficients that represent the general population-level effect that fintech has on the sentiment of banking-related tweets.

Mustafaraj et al. (2011) emphasize that the difference between the content generated by the vocal minority and the silent majority is significantly different on Twitter and Facebook. They insist that when designing the research framework, scholars take good care of controlling for those differences. This also supports the choice of GEE as it not only enables control for two different groups, but for each cluster separately.

However, Ballinger (2004) noted that the drawback of using a GEE model is that they do not have a goodness-of-fit measure that is equivalent to the magnitude of squared differences estimated by OLS. Regardless of that limitation, we are still able to estimate unbiased and robust coefficients.

3.2.1 Descriptive analysis

As elaborated above, the descriptive analysis will focus on descriptive analytics and content analytics.

Descriptive analytics gives a broad overview of the tweet metrics. Tweet metrics include tweet statistics, such as how many tweets were posted over time, whether the tweet was about fintech or not, and how many likes and retweets the tweet received. This information can reveal how the topics in banking have shifted more towards fintech. Also, by looking at the public metrics such as likes and retweets, we can get a sense of how users on average react to different content. These twitter metrics can be compared with actual banking data, such as the number of investments into fintech companies to study the relationship between Twitter activity and actual banking activity.

Furthermore, we will also employ term frequency and sentiment analysis since they can reveal what the prominent topics are and how people react towards them. However, it is also important to note that extracting information from a large corpus of tweets is complex and may result in some information loss due to the need to use quantitative methods with non-numeric data (George et al., 2016; Zeng et al., 2010).

One method to extract intelligence from the content of the tweets is to simply analyse the frequency of words. Word frequency analysis has long been a method to analyse large corpora (Baron et al., 2009). The method relies on counting the frequency of words that occur in the corpus of tweets. After determining the word frequencies for all the years through 2009 till 2020, it can show how the most frequent words have changed through time. Chae (2014) adds that term frequencies can help identify topics of discussion.

What is more, to conduct sentiment analysis in this paper (both in the descriptive and regression part), we rely on Hutto and Gilbert's (2014) VADER (Valence Aware Dictionary and Sentiment Reasoner) analysis tool. VADER is a rule-based tool that is specifically developed to work with social media posts (Hutto and Gilbert's, 2014) and it has been used in research papers before (e.g., Elbagir and Yang, 2019 and Borg and Boldt, 2020). In short, it includes a lexicon of words, emojis, abbreviations, and phrases that each has been assigned a sentiment score by 10 different independent raters. These terms are then used to calculate the positivity, negativity, and neutrality scores of the social media post. These scores are "ratios for proportions of text that fall in each category" (Hutto and Gilbert, 2014). We evaluate the sentiment of a tweet with the compound measure, which is simply the sum of the valence scores of the words present in the tweet and then normalized to be between –1 and 1. Hutto and Gilbert (2014) add that it is the most useful metric if the purpose is to get a single unidimensional measure.

There are also many other sentiment analysis tools developed, and it is important to find the right one that fits the context of the research. We tried 3 different tools: the Stanford Core NLP (Manning et al., 2014), TextBlob (Loria, 2018), and VADER (Hutto and Gilbert, 2014). The Stanford Core NLP was the most advanced NLP tool that we experimented with. Its key advantage is that it provides a pipeline for manipulating raw text by performing such actions as tokenization, lemmatization, and generating annotations. However, after reviewing a random sample of around 500 tweets and determining the sentiment ourselves, it was evident that VADER was the most accurate in assigning the sentiment level of the tweets. Since VADER is a rule-based sentiment analysis tool, it can be scaled more easily, which becomes important as we are dealing with a relatively large sample of tweets. On the other hand, the

Stanford Core NLP first needs to break the sentence apart and then determine the overall sentiment, which is computationally more intensive.

To give a sense of what sentiment score is generated by VADER, we will provide three different tweets that generated positive, neutral and negative scores.

- *"Fintechs are celebrating the success in expense of the retail banks"* –sentiment score 0.81
- *"Citi announces the acquisition of a new fintech company #banking"* –sentiment score 0.0
- *"Legacy bank system has reached its end. Without new technology we cannot build a better tomorrow."* – sentiment score -0.25

As can be seen from these examples, tweets that include words with positive connotation (*celebrating* and *success*) carry a high overall sentiment score. Tweets that are generally related to an announcement or a piece of news have neutral sentiment and tweets that include words that signal negativity or deficiency (such as *without* and *cannot*) often lead to an overall negative sentiment score.

Network Analytics will focus on user metrics such as the number of tweets by users, visibility, and activity. Here, visibility is of focal interest, because it indicates which users potentially have the most impact on the public discourse about banking.

Twitter users create two different networks: (1) the *friendship* network through followers and following and (2) the @reply (mention) network, which creates more interpersonal relationships (Chae, 2014). Motivated by Chae (2014), we calculate the visibility by summing up all the retweets and @mentions the user received. This visibility metric will give a rough measure of which users have potentially been seen the most by other users. Activity is calculated by simply summing up all the tweets about banking that the user has made.

Another way we can calculate a user's visibility is to use the in-degree centrality metric. The goal of centrality analysis is to find out which users are the *central* users in the banking network. In other words, the centrality measure will show which users have the most connections and are the most influential. Everett and Borgatti (2006) emphasize that centrality analysis is one of the most important tools that is used to make sense of social networks. As we have a directed network (we know who mentioned who), we can calculate the fraction of users a single user is connected to through mentions which gives us the in-degree centrality metric. Therefore, a high in-degree centrality indicates that the user is mentioned by many other users and consequently is highly visible.

Another centrality metric that researchers often focus on is the out-degree centrality, which measures the fraction of users that a specific user has mentioned. However, in the case of Twitter, this metric's relevance is not as pronounced because users can send out a lot of @mentions, but essentially no real connection may never be established.

By developing the NetworkX network analysis tool, Hagberg et al. (2008) have made it very easy to draw network graphs and calculate the centrality measures with Python. We will also rely on NetworkX in this research.

3.2.2 Modelling the relationship between sentiment and fintech-related tweets

The descriptive analysis gives a holistic view of the public discourse; however, the focus of this paper is to find out whether the sentiment of fintech tweets is significantly different from all other banking tweets. This will also enable us to tackle the research question with more robust evidence since as was argued above, the polarity of discourse on social media can be an important indicator in determining the actual outcomes in an economy or the society as a whole.

Sentiment analysis is being used as a method to analyse Twitter data in many fields, including cancer research (Clark et al., 2018), forecasting election results (Agrawal and Hamlig, 2017), predicting stock price moments (Bing et al., 2014 and Pagolu et al., 2016) and predicting the popularity of literature (Hassan et al., 2020) among many other topics.

Bing et al. (2014) and Pagolu et al. (2016) showed that indeed the movement of stock prices and sentiment on Twitter are strongly correlated, where positive sentiment is associated with growth and negative sentiment corresponds to a decline. Thus, the polarity of the discussions on Twitter can reveal the overall value of a company or even an index. This is encouraging for researchers using Twitter, because it strengthens the argument of Kou et al. (2017), who claimed that there is coexistence and co-development of public discourses on social media and offline actions.

Also, there are many papers providing evidence that the sentiment on Twitter can predict election outcomes. For instance, Agrawal and Hamling (2017) discovered that Twitter sentiments corresponded with 66.7% of the actual electoral college outcome and concluded that overall twitter sentiments leaned positively towards Donald Trump relative to Hillary Clinton in the 2016 presidential election. Also, Budiharto and Meiliana (2018) generated reliable prediction results by mining the sentiments of tweets associated with presidential candidates and they found that sentiment analysis produced reliable predictions. Overall, using Twitter to analyse political affiliations and election outcomes is very popular. One of the reasons for this could be that a platform with many users and frequent social interactions is a good avenue for studying how social structure is formed or how it is changing.

In addition, interesting research has been conducted in a variety of fields with Twitter and sentiment analysis. In one study, Hassan et al. (2020) measured the early impact of research articles with the sentiment analysis of tweets about the articles, concluding that there is an encouraging positive correlation between citation counts and sentiment of tweets, also adding the utility of twitter-based opinions as an additional predictor of the early impact of literature. This left us wondering if they could predict the early impact of their article as well.

Overall, there is evidence that sentiment analysis, especially the polarity of it, can provide insights into real behaviour.

The econometric model that we provide now will attempt to provide evidence for the first hypothesis. To reiterate, the first hypothesis was posed as: *Compared to other banking-related tweets, fintech tweets are higher in positive sentiment.* Thus, we are interested in the difference of sentiment of tweets that contain the word fintech hence labelled as fintech tweets, and tweets that do not contain the word fintech, hence labelled as banking tweets. Though we have data from 2009, we will be using the period from 1st of January 2015 till 31st of March 2021 to run the regressions, because as it will be later shown, among the banking tweets, fintech tweets became popular only in 2015.

To model sentiment, we will again use Hutto and Gilbert's (2014) VADER sentiment analysis tool to calculate the normalized sentiment score of a tweet, which will be the dependent variable. Also, we will need to label the tweets into two: (a) those that contain the word fintech and (b) those that do not; this will be the independent variable of interest. Equation 1 illustrates this relationship:

$$SENTIMENT_{t,i} = \alpha + \beta(FINTECH\ DUMMY)_{t,i} + \mathrm{g}CONTROL_{t,i} + \mathrm{e}_{t,i} \qquad (1)$$

Where the dependent variable *SENTIMENT* is a continuous number from –1 to 1 representing the normalized sentiment score from VADER. Appendix 1 plots the distribution of the dependent variable, where it becomes evident that many of the tweets have a sentiment score of 0. The *FINTECH DUMMY* is a dummy variable corresponding to 1 if the tweet contains the word fintech (or #fintech) and 0 if it does not contain the word fintech. *CONTROL* is a vector of four control variables that include: the number of replies, likes, retweets, and quotes (quotes are retweets with a comment). The subscript *t* corresponds to a tweet and subscript *i* corresponds to a user.

Furthermore, our coefficient of interest is β, which measures the difference that fintech-related tweets have on sentiment *versus* the base case, where the fintech dummy is equal to zero and which corresponds to other banking tweets. Our hypothesized estimate for β is that it is positive (as in hypothesis 1).

The set of control variables are chosen in order to obtain unbiased estimates for the effect that fintech has on sentiment. Here we utilize all the information that Twitter provides and include all these variables into our regression model. Table 3 provides descriptive statistics with mean and standard deviations for all variables. From the Table 3 we can see that the mean sentiment score for fintech-related tweets is higher than it is for banking-related tweets.

Table 3. Descriptive statistics

	Fintech-related tweets	Banking-related tweets	All tweets
Sentiment	0.18 (0.34)	0.12 (0.38)	0.13 (0.38)
Retweets	2.1 (12.8)	0.7 (12.7)	1.0 (12.8)
Replies	0.1 (1.0)	0.1 (1.7)	0.1 (1.6)
Likes	2.5 (10.3)	1.0 (17.1)	1.3 (16.0)
Quotes	0.1 (1.1)	0.1 (1.0)	0.1 (1.0)
Number of tweets	561,511	2,275,932	2,837,443
Number of users	35,445	242,482	217,812

Notes. Mean and standard deviation (in parentheses) of the variables. Table created by the authors.

Lastly, as elaborated above, we will be using the generalized estimating equations (GEE) approach by Liang and Zeger (1986) to solve for the coefficients and allow clustering. The clustering is done at the user level. Also, we will use the identity link function, which corresponds to the same model as the general linear model. Another important decision that must be made is choosing the correlation structure of tweets for users. To start with, we assume constant correlation (exchangeable correlation structure) of sentiment among users. However, we will also use the Quasi-Information Criterion (QIC), which is used for GEE models, to test whether the independent correlation structure (zero correlation within-subjects) performs better. It is intuitive to think that constant correlation fits best, because one can assume that individual users are somewhat consistent in their sentiment towards banking and fintech.

3.2.3 Modelling the relationship between sentiment and fintech-related tweets over time

The model that is provided in this section is focused on the second hypothesis. To reiterate, the second hypothesis is: *Among banking tweets, the sentiment of fintech has become more positive over the years*. The relationship between sentiment of a tweet and different time periods can be explained with the following linear equation (2):

$$SENTIMENT_{t,i} = \alpha + \beta(FINTECH\ DUMMY) * (YEAR)_{t,i} + \mathrm{g}CONTROL_{t,i} + \mathrm{e}_{t,i} \quad (2)$$

Where the dependent variable *SENTIMENT* is again a continuous number from –1 to 1 representing the normalized sentiment score from VADER. However, now we have an interaction term (*FINTECH DUMMY) * (YEAR DUMMY)* which represents the vector of interactions between the fintech dummy and years. The years span from 2015 to 2021 and the *YEAR DUMMY* is equal to one for each specific year. *CONTROL* is again a vector of five control variables: the number of replies, likes, retweets, and quotes. Also, as was in the previous model, the subscript *t* corresponds to a tweet and subscript *i* corresponds to a user.

Now the coefficients of interest are βs that represent the effect that fintech tweets have on the sentiment of a tweet in a given year. Here we are not so much focused on whether the coefficients are positive, but rather we care about the change in coefficients for different years. The GEE specifications are the same as they were in the first model.

The method in this paper is somewhat novel, because the analysis of sentiment over such a long time period as in this paper has not been done before according to the authors. The main reasons for it are firstly that the data generated on Twitter (or any online forum for that matter) has taken place only in the last 10-15 years and secondly in February 2021 Twitter enabled full access to academic researchers to use their whole database for research. In order to take a critical stance towards the new kind of dataset (i.e., a long period of Twitter data) and address the issues mentioned by Benjamin et al. (2018) relating to reproducibility, P-hacking, and underpowered studies, we will set a high threshold for p value at < 0.005.

4 Analysis of Results

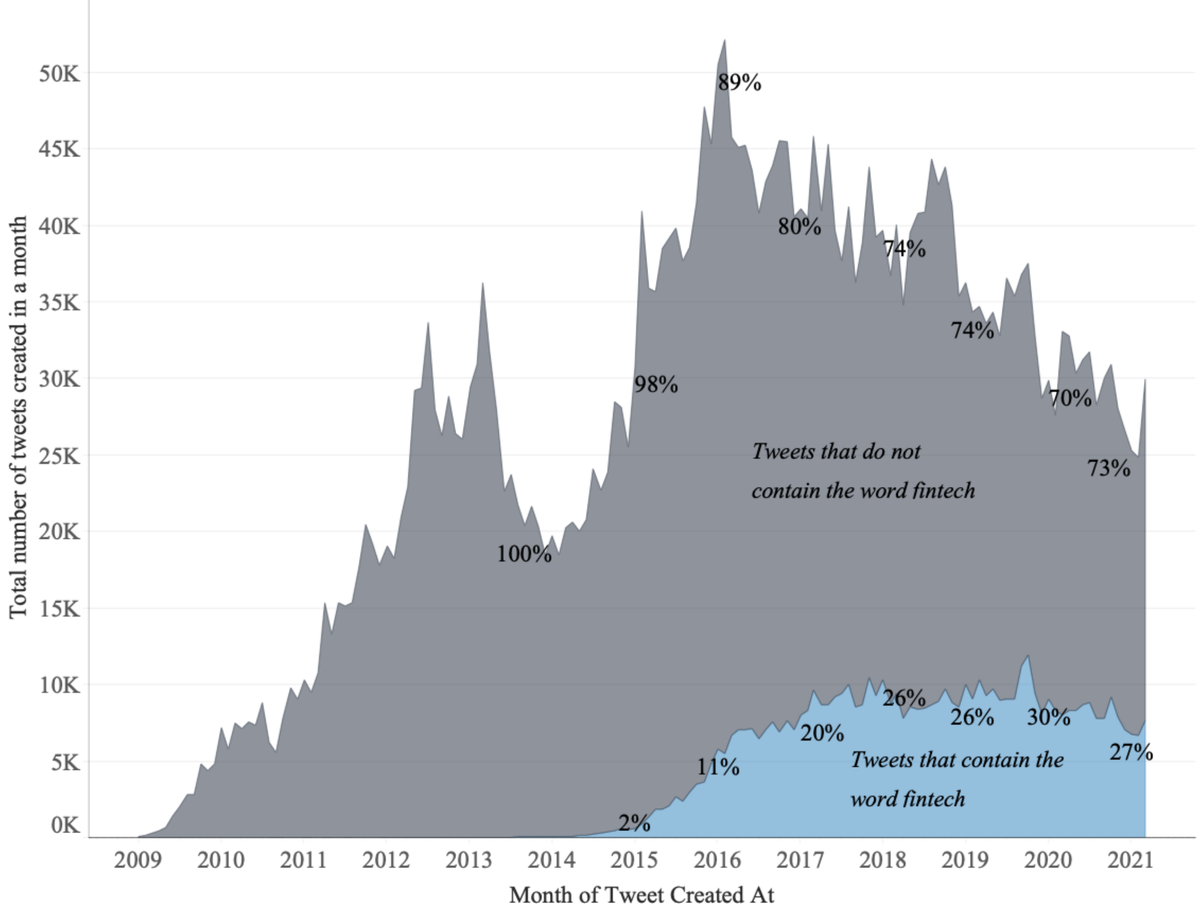
4.1 Descriptive Analytics

To start with, we will provide the findings from descriptive analytics. Keeping in mind that we are mostly interested in the intersection of banking- and fintech-related tweets, Figure 7 shows how the portion of fintech tweets in banking has changed over time.

In this thesis the testing of the two hypotheses that we set out earlier are most important in drawing statistically sound conclusions from the data. The accompanying descriptive analysis in this section gives a holistic overview of what has happened in the public discourse around banking over the last 12 years. In this section we use different measures: term frequency analysis, likes and retweets, real banking activity, and user statistics to give an overview of how banking and fintech are represented on Twitter.

The descriptive analysis is later combined with the results from the hypotheses tests in the discussion section to explain a) whether fintech is actually a topic that is frequently discussed in relation to banking related content and b) who are the most popular Twitter users that engage in the discussion around banking. Though the changes in sentiment, which are the focus of the hypotheses, reveal the general stance of the public and the changing attitudes, the additional knowledge regarding the popularity of fintech and Twitter users support these results by providing information on how big of an impact the change in sentiment carries with it. We use an approach similar to Chae (2014) to provide descriptive analysis on Twitter data.

Figure 7. Banking and fintech tweets over time

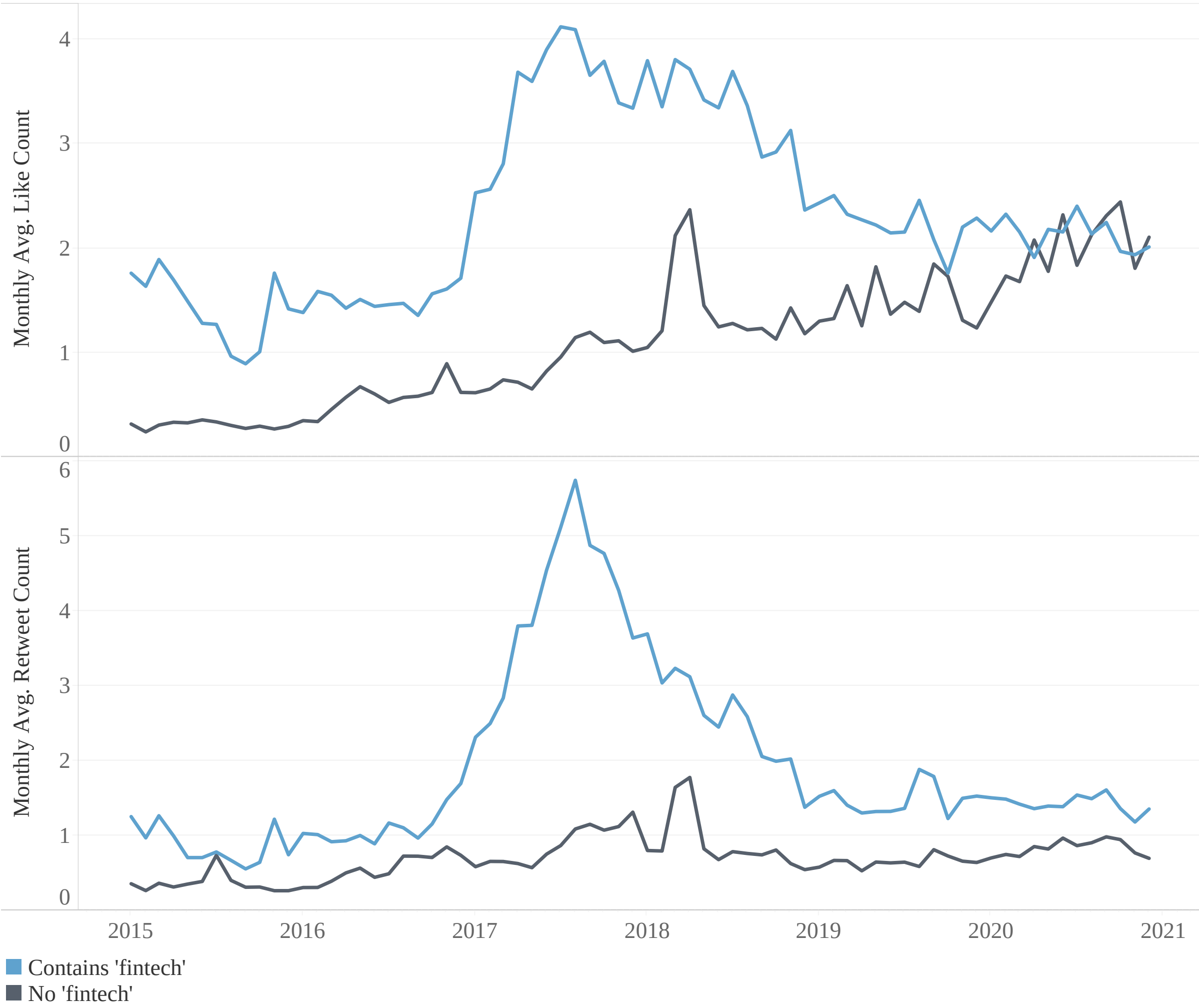


Note. Figure created by the authors.

Tweeting about fintech became popular in 2015 and starting from 2017, tweets about fintech account for around a fourth of all the banking tweets with a high of 30% in 2020. It is also notable that the volume of tweets in banking has reached its zenith in around 2016-17 and has been in decline ever since whilst the decline in the number of fintech-related tweets is not as prominent. This should also be viewed keeping in mind that Twitter as a platform has increased the characters limit of its tweet from 140 to 280 in the year 2018.

Another interesting observation comes from likes and retweets in the period where fintech tweets became popular. Namely, when looking at the monthly average of likes and retweets per individual tweets, we can observe that the averages are higher for fintech-related tweets (Figure 8). The monthly averages of likes and retweets for banking tweets that do not contain the keyword fintech are lower. Another observation from Figure 8 is that the gap between fintech-related and not related tweets achieved its climax in 2017 and has since been decreasing. The average like count and retweet count of fintech tweets has been showing a downward trend since it's 2017.

Figure 8. Monthly averages of likes and retweets per individual tweet over time



Note. Figure created by the authors.

We also observed the relationship between fintech-related tweets and the number of investments made in the respective quarters by three leading US banks: Goldman Sachs, Citigroup, and JPMorgan Chase in fintech companies over the period of 2010-2020 (Figure 9). CB Insights (2021) provides details about the deals, but the dollar value of those investments is not disclosed.

Figure 9. The number of fintech deals and fintech tweets in a quarter

Notes. The number of deals with fintech firms by Goldman Sachs, Citigroup, JPMorgan Chase and the number of banking tweets relating to fintech in a quarter. Graph created by the authors relying on the data from CB Insights (2021).

A key observation is that the number of investments made by leading US banks has taken off in 2015, at the point where also fintech-related banking tweets became popular. Of the 258 investments made collectively in 2010-2020, 226 were made after 2015. However, Figure 9 should be only considered as a visual aid.

Further, to provide a descriptive overview of the content of the tweets, we focused on term frequency and sentiment analysis.

4.1.1 Content analysis

To start with, term frequency analysis can shed light on what have been the most discussed topics on Twitter. Table 4 shows the five most frequent terms in a year while disregarding *banking*, *banks*, and *bank* because naturally, all observed tweets contain at least one of those words. The analysis shows that in the early years of the discussion, banking-related content was attuned towards job seekers as *jobs* and *job* were the most frequent words. Interestingly, Chae (2014) also found that one key application of Twitter for supply chain-related discussions revolved around job applications.

However, in 2020, Table 4 shows that fintech-related content is by far the most relevant in the banking discussion. In fact, the top 5 most frequent terms in 2020 besides finance are *fintech*, *digital*, *payments*, and *blockchain*, which are all technological terms, unlike the most frequent words in 2010 and 2015, which are terms generally not related directly to technology. It is also interesting to note that already in 2015, the fourth most frequent word was *fintech.* It is also important to recollect that it was the year 2015 when fintech had just started to appear among the banking tweets. Thus, fintech gained its popularity extremely quickly.

Table 4. Most frequent terms in 2010, 2015 and 2020

2010	Frequency	2015	Frequency	2020	Frequency
jobs	14 145	jobs	68 712	fintech	101 856
risk	9 666	finance	43 853	finance	51 946
new	8 727	business	42 065	digital	38 401
finance	8 322	fintech	29 655	payments	34 770
credit	7 311	job	24 744	blockchain	33 270

Notes. The top 3 words: banking, banks, and bank have been removed as they do not provide any insights since all tweets contains at least one of those words. Table created by the authors.

Figure 10. Word clouds

Notes. Word cloud on the left is for the year 2010 and on the right is for the year 2020. The top 3 words: banking, banks, and bank have been removed. Figures created by the authors.

A word cloud is a cluster or collection of words visualized in different sizes where the size of the word is proportional to the number of times it is mentioned within a given text. The word clouds (Figure 10), which are a decade apart from 2010 and 2020, respectively, show a clear contrast between the important words/topics during those two periods. It is interesting to note the terms that remained very prominent between these two are *money*, *business*, and *finance*. Besides these, the emerging new words *fintech*, *digital*, *blockchain*, *bitcoin*, and *AI* are among the most visible words in the discourse around banking in 2020 (see Appendix 2 for the 30 most visible words).

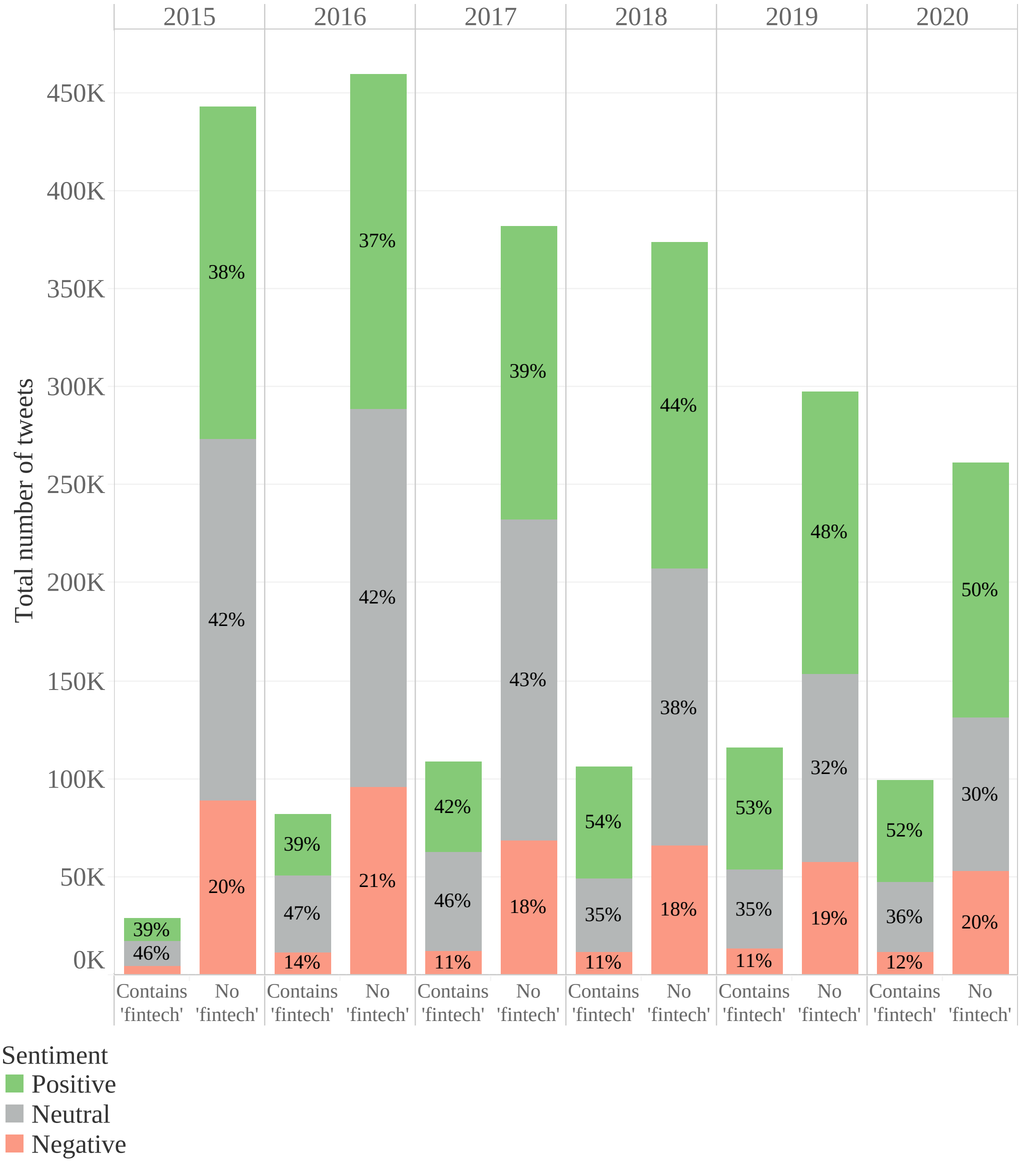
As we already elaborated above, sentiment analysis is a powerful tool to study large unstructured text data as it can quantify emotion and polarity and give relevant insights.

Furthermore, Hutto and Gilbert (2014) explain that when the purpose is to classify the sentences into positive, negative, and neutral, researchers often use these standardized thresholds:

1. Positive sentiment: *compound* score >= 0.05
2. Neutral sentiment: *compound* score > -0.05 and < 0.05
3. Negative sentiment: *compound* score <= -0.05

Figure 11 shows the sentiment for banking tweets that contain the keyword fintech and *vice versa* over time and are grouped into positive, neutral, and negative categories. Appendix 1 shows the whole distribution of tweets with their compound scores and without grouping the tweets into three categories. The key finding here is that fintech-related tweets contain significantly less content with negative sentiment. Also, the portion of tweets with positive sentiment is slightly higher for fintech-related tweets. Another interesting observation is that over time the portion of neutral tweets has decreased significantly for both groups, while the portion of positive tweets has increased. Also, it is notable to add that the portion of negative tweets has remained more or less the same.

Figure 11. Yearly sentiment for banking- and fintech-related tweets



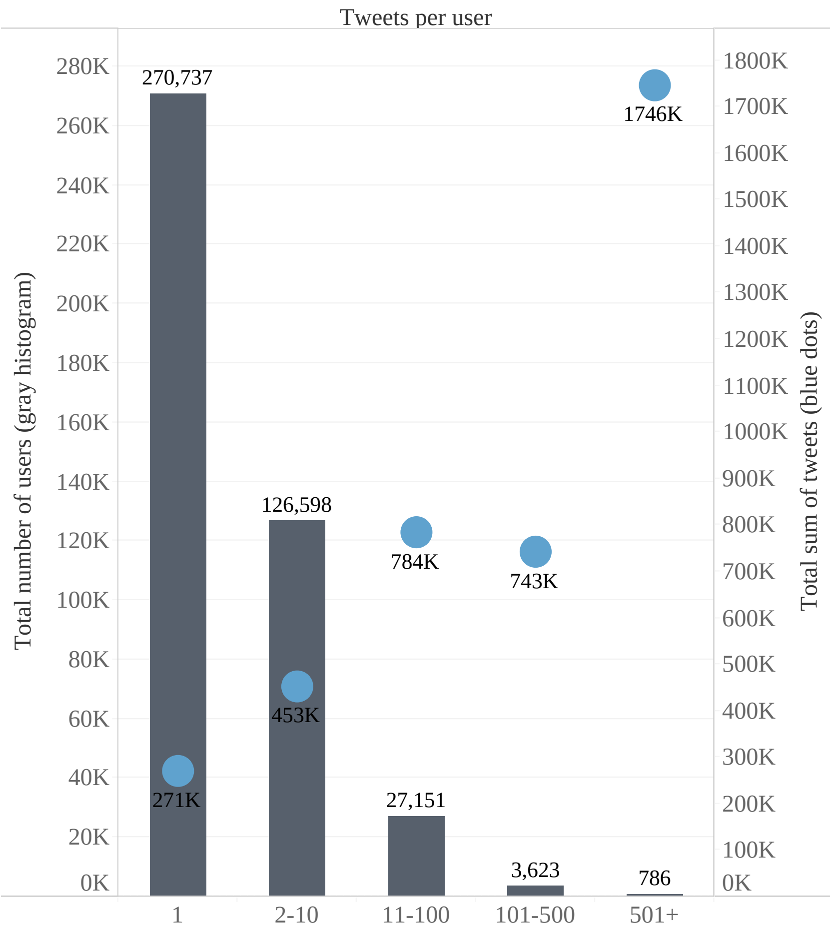
Note. Figure created by the authors.

4.2 Network analytics

As was elaborated above, the networks that Twitter users form can be analysed in two ways: with user metrics and centrality measures. Furthermore, network analytics gives an overview of the most influential users and how they are related to fintech.

First of all, the user analysis shows that from 2009 till 2021, there have been 428,895 distinct users that have tweeted using the three hashtags: #banking, #banks, and #bank. Of those users, 35,792 have tweeted about fintech at least once. Most of the time, Twitter users used one of the three hashtags only once, but there were also many active users that often tweeted about banking. Furthermore, though many users have tweeted about banking, most of the tweets are still generated by the users who have tweeted about banking more than once. Below is a graph showing the number of banking tweets per user group and how many tweets the specific user group has created (Figure 12).

Figure 12. Number of tweets per users

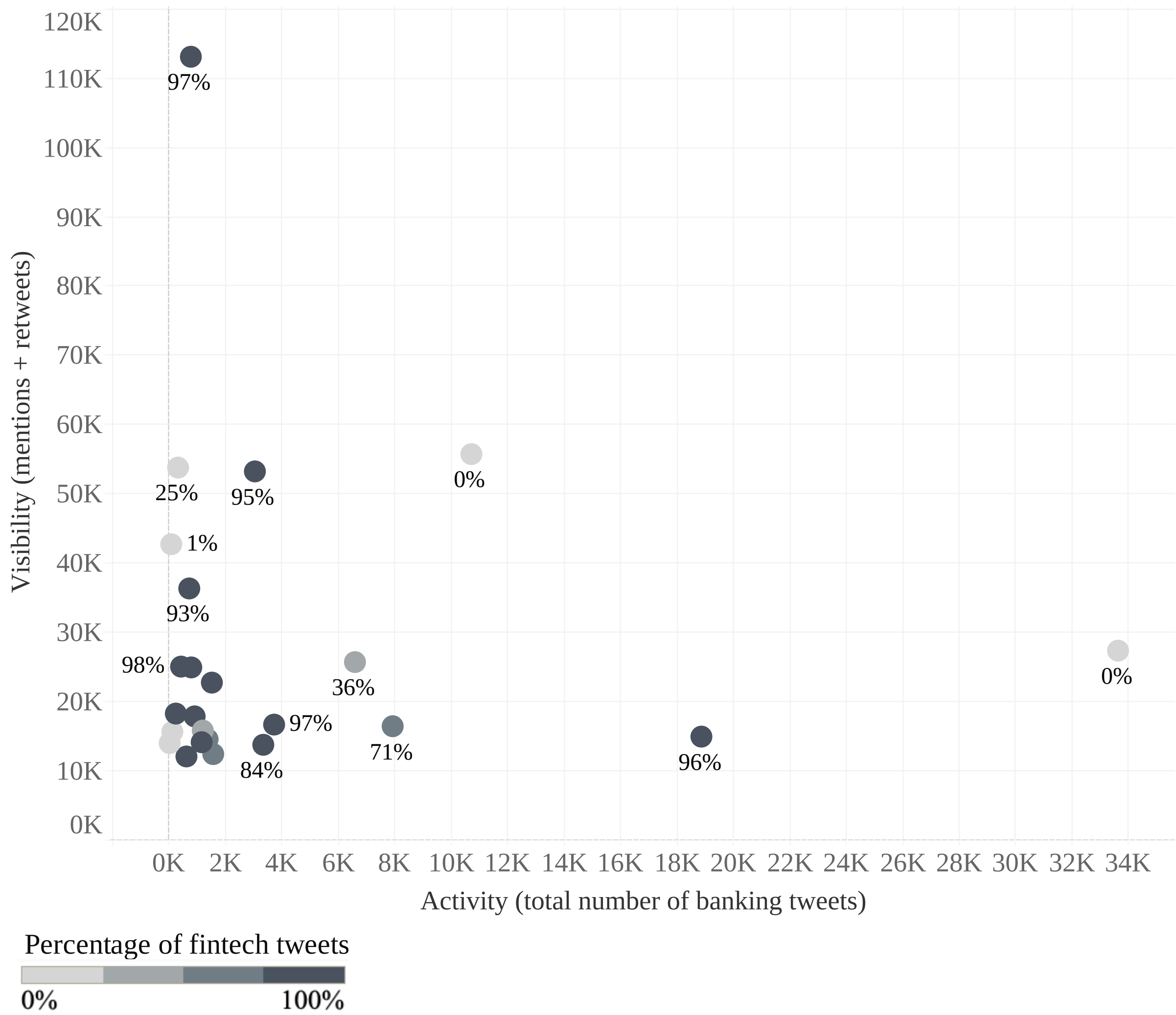


Note. Figure created by the authors.

It is also noteworthy that 0.2% of the users contribute to 44% of tweets. This could be a possible sign that a vocal minority group is steering the public discourse.

The other relevant user metrics are visibility (@mentions + retweets), which helps to determine which users' tweets are the most engaging, and activity (sum of banking tweets). Figure 13, where we display the 24 most visible users that have used the banking-related hashtags, shows that visible users are users that mostly tweet about fintech. Of the 24 most visible users, 13 users have tweeted 80% or more times about fintech, and 16 users have tweeted 50% or more times about fintech, when using the banking hashtags.

Figure 13. User visibility and activity



Notes. The dots on the graph are accompanied by the percentage representing how many tweets were about fintech (the darker the dot the higher the fraction). Graph created by the authors.

It is important to note that Figure 13 provides visibility and activity measures for the whole period (2009-2021), which goes to show that in most cases users that tweet about fintech are more visible than those users that do not. Or if there were such users, these accounts have now been deleted. Among the 24 highly visible users, the follower count ranges from 4,300 to 243,000, with the most visible user having around 100 thousand followers. As was in Chae (2014), the most active users are not necessarily the most visible. Of those 24 users in Figure 13 only a third represents an organization or a firm. The other 18 users are mostly individual speakers, thought leaders, and podcast hosts.

Another way to assess the visibility of the users is to calculate the in-degree centrality metrics. The centrality metric gives a different overview of the visibility since it also considers the diversity of users that are connected and not the sum of mentions and retweets, which can be made frequently by small groups of people. To calculate the in-degree centrality metric, we decided to focus only on the period of 2015-2021 since this includes the period when fintech was also discussed in the banking context.

The in-degree-centrality simply shows the fraction of all the users that have mentioned the specific user. The nodes are the users and the edges are the links between the users (the mentions). Here the main finding is that the network around #banking, #banks, and #bank is very dispersed and the highest in-degree centrality metric for a single user was only 0.0065, meaning that only 0.65% of all the users that have used either of the three hashtags have mentioned this specific user.

This suggests that the most central users are connected to only a small fraction of all the users. Furthermore, the most central users tended to be major news outlets, but there were also some individual fintech promoters and a few major banks. Therefore, even though the most visible and active users tended to be users that often tweet about fintech (Figure 13), the whole network is actually very broad and their influence might not reach the majority of users interested in the public discourse about banking. The individual users and their in-degree centrality metrics are not presented in this paper for the sake of retaining their privacy and maintaining the analysis on an aggregate level.

4.3 Hypothesis test results

Hypothesis 1c: Compared to other banking-related tweets, fintech tweets are higher in positive sentiment.

Table 6 provides evidence for the first hypothesis. The results from model 1 indicate that across all banking-related tweets, if the tweet contains the word fintech, then the sentiment of the tweet is higher by 0.0417 points, holding all other variables constant. This finding is statistically significant at the 0.005 level. Thus, we have evidence to reject H1a, which stated there is no difference in sentiment, and H1b, which stated that fintech tweets have lower sentiment.

Table 5. Results from model 1

	Sentiment
Intercept	*0.144*
	*(194)****
Fintech	*0.0367*
	*(11.684)****
Replies	*-0.001*
	*(-2.117)**
Likes	*0.0002*
	*(3.474)****
Retweets	*-8.129e-05*
	*(-2.838)****
Quotes	*0.0009*
	(1.667)
Dependence structure	*Exchangeable*
Scale parameter	*0.141*
Observations (total tweets)	*2837443*
Clusters (total users)	*253259*
Mean cluster size	*11.2*
Min. cluster size	*1*
Max. cluster size	*54499*

Notes. t-statistics for robust standard errors clustered at the user level in parentheses.
*Statistical significance: *<.05, **<.01, ***<.005.*
Table created by the authors.

What is more, the exchangeable correlation structure justifies itself as the QIC score for the model with an exchangeable dependence structure was less than that of the model with the independent dependence structure. This implies that when considering users having constant correlation within the sentiment of the tweets, then the model performs better.

In an attempt to proxy sentiment with likes and retweets, we also constructed models where the dependent variable was either likes or retweets. However, as these two variables are zero-inflated count variables, one cannot use an ordinary model where the outcome variable resembles a normal distribution.

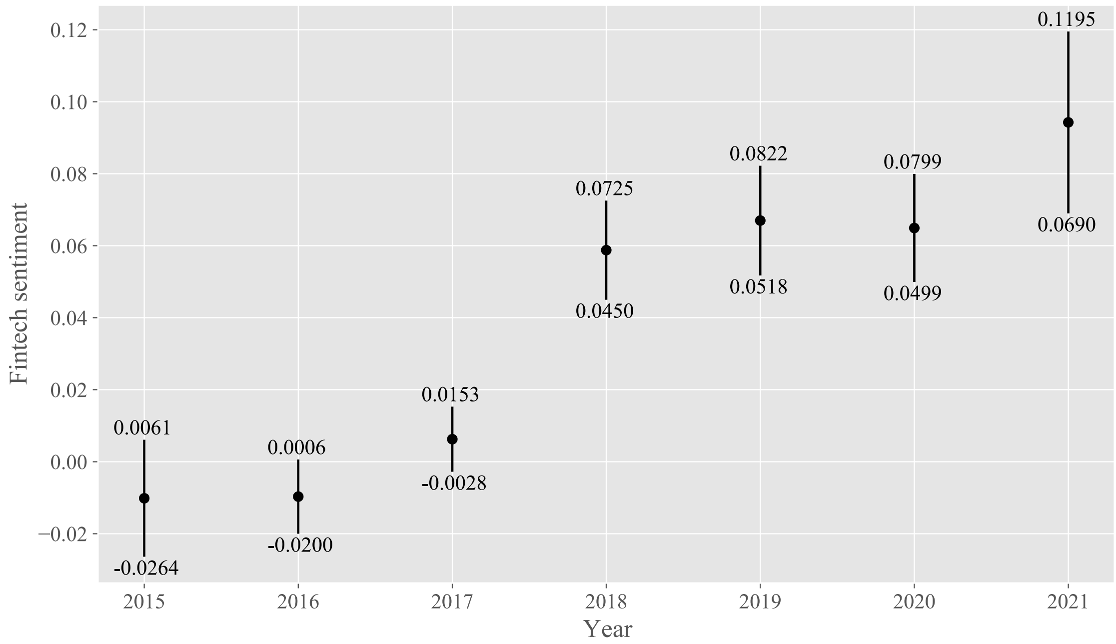
Furthermore, researchers often log-transform the dependent variable, but O'Hara and Kotze (2010), in their paper titled "Do not log transform count data" implicitly argued that by adding a value 1 before transformation and with large dispersion, the models perform poorly. The Poisson regression, which is a better option, in this case, has a strict assumption that the mean must equal the variance in the dependent variable. This, of course, is not the case with this dataset since the dispersion of likes and retweets is large and the mean is small. In this case, the best option would be to use the negative binomial regression. However, the negative binomial model converged only for small subsamples but not for the whole sample of tweets. Kong et al. (2015) acknowledged this issue and developed a GEE type zero-inflated negative binomial regression to fit clustered counts with excessive zeros. Though the developments by Kong et al. (2015) could offer a solution, the implementation of their approach is out of the scope of this research project.

Hypothesis 2c: Among banking tweets, the sentiment of fintech has become more positive over the years.

Figure 14 shows that indeed the sentiment of fintech-related tweets has become more positive over the years. Appendix 3 provides the full results from the regression. As the 99.5% confidence intervals show, we have sufficient evidence to reject H2a, which states that the sentiment of fintech-related tweets has remained the same, and H2b, which states that the sentiment of fintech-related tweets has become more negative over the years.

From Figure 14 we observe that there has been a significant spike in positive sentiment fintech-related tweets in 2018. As Twitter increased its character length in 2018, one can argue that the probability that a tweet contains a word associated with positive sentiment became higher, because users now use more words to express their thoughts. However, it also applies to words related to negative sentiment. Either way, by having more characters to express oneself, the possibility that the user uses words that represent sentiment is higher.

Figure 14. The fintech coefficient over the years



Notes. The effect of fintech on sentiment over the years. The error bars represent the 99.5% confidence interval. Note that fintech sentiment represents the additional positive sentiment that fintech-related tweets have versus other banking-related tweets. Table created by the authors.

Table 6 shows that, indeed in 2018, the mean number of characters increased by almost 60%. However, the increase in character length for tweets associated with fintech increased even before Twitter made the changes in October 2018. In 2018 and up until the end of September, the mean character length was 125.

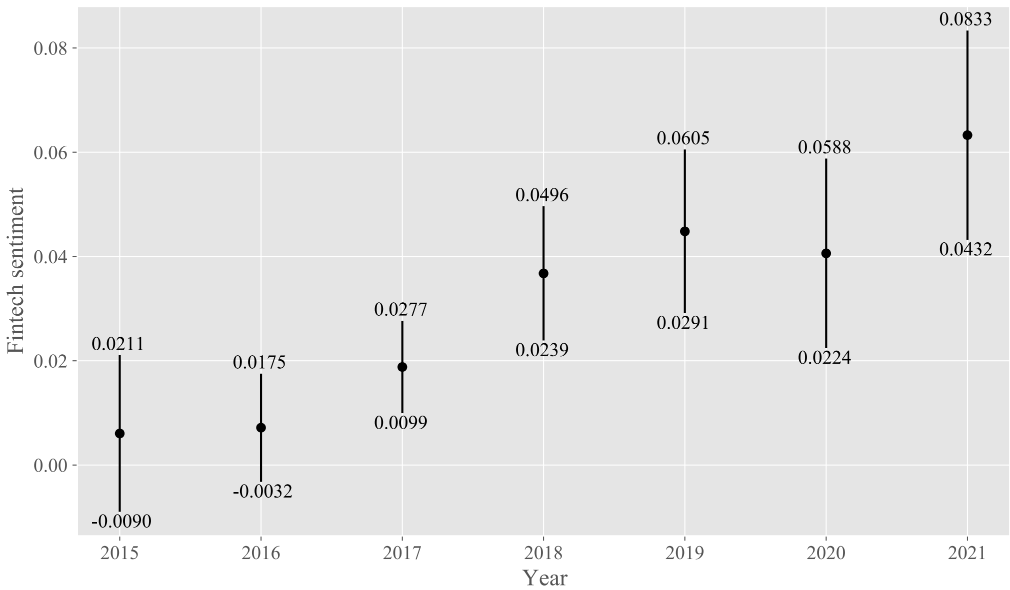
Table 6. Mean number of characters in a tweet per year

Year	*2015*	*2016*	*2017*	*2018*	*2019*	*2020*	*2021*
Mean number of characters	82	81	91	144	143	150	157

Notes. Blanks are also considered as characters. Table created by the authors.

To validate the robustness of the result that fintech-related tweets have indeed become more positive over the years (Table 7), we introduced a new independent variable to explicitly control for the number of characters and provide unbiased estimates for the fintech coefficient. After controlling for the number of characters in a tweet, we found that the sentiment of fintech-related tweets has still increased (Figure 15). However, in this case, the increase is not as pronounced and the spike in 2018 is not as intensive. We also now cannot claim that the positive sentiment associated with fintech-related tweets in 2020 was statistically significantly different from 2017 since the 99.5% confidence intervals are overlapping for the two years. However, the general trend for the positive sentiment is still upward.

Figure 15. The coefficient of fintech over the years after controlling for tweet length



Notes. The error bars represent the 99.5% confidence interval. Note that fintech sentiment represents the additional positive sentiment that fintech-related tweets have versus other banking-related tweets. Table created by the authors.

To conclude, the results suggest that across all the banking tweets, fintech-related tweets are associated with higher positivity and the positivity of the tweets has increased gradually over the years. Table 7 provides the summary of findings regarding the hypotheses.

Table 7. Summary of findings

	Hypothesis	
H1a	Compared to other banking-related tweets, there is no difference in sentiment in fintech-related tweets.	Rejected at p = 0.005
H1b	Compared to other banking-related tweets, fintech tweets are higher in lower sentiment	Rejected at p = 0.005
H1c	Compared to other banking-related tweets, fintech tweets are higher in positive sentiment.	Not rejected
H2a	Among banking tweets, the sentiment of Fintech has remained the same over the years.	Rejected at p = 0.005
H2b	Among banking tweets, the sentiment of Fintech has become more negative over the years.	Rejected at p = 0.005
H2c	Among banking tweets, the sentiment of Fintech has become more positive over the years.	Not rejected

Note. Table created by the authors.

4.4 Assessing validity and reliability

To evaluate the quality of research, we also considered the validity and reliability of the method we chose. Firstly, we argued that the polarity of sentiment on Twitter can provide insights into real behavior. That is, relying on studies on politics and the stock market, we found that the sentiment measured on Twitter is reflected in the outcome of elections and stock market movement. However, measuring sentiment is a complicated task since sentiment analysis tries to operationalize human emotion (positive or negative) and quantify it. In this paper, to assess the extent to which the sentiment score really measures the emotion component in text we first used our own interpretation of emotion in the tweets and evaluated how accurately the sentiment score represented our interpretation. In fact, we started with The Stanford Core NLP and TextBlob sentiment analysis tools and found that they did not do particularly well on longer Tweets. For this reason, we turned to VADER, which was much more accurate in outputting a sentiment score that was in line with our own interpretation of emotion in the tweet. In assessing whether VADER is a valid method to assess the emotion in text, we also reviewed several other articles that used this specific tool in quantifying sentiment (e.g., Elbagir and Yang, 2019 and Borg and Boldt, 2020) and found that VADER is a common tool that is used in large scale studies that involve analysing the sentiment of tweets. VADER is a popular choice, because it also captures other elements in text such as exclamation marks and emojis that are frequently used by Twitter users to express their emotions. The argument for why the sentiment is accurately measured using the method and the tool in this paper is also supported by how the VADER analysis tool is developed. Namely, Hutto and Gilbert (2014) employed 10 different independent raters to assign a sentiment score to words, emoticons, and emojis, therefore essentially VADER represents the understanding of emotion by 10 different people (*please see the VADER lexicon on GitHub*).

However, when interpreting the results, we must note that operationalizing human emotions from text is currently not perfectly solved by computer algorithms, therefore the overall validity also suffers to some extent. Though sentiment analysis using complex algorithms has been around for some time now and it also has improved over time, capturing human emotions, sarcasm, and irony from text is a subtle capability and the artificial intelligence tools that are freely available to all researchers do not offer a flawless way to analyse them yet. All in all, we believe that the method and the sentiment analysis tool that we chose generally measures accurately the emotion component in text, though for some tweets it may output an inaccurate sentiment score due to the complexity of human language.

In arguing for reliability, we claim that the method chosen in this paper is highly reliable. The results that we derived can be easily reproduced. Since we are using a universe of banking tweets (that is tweets that contain either #banking, #banks, or #bank), other researchers that download a sufficient number of banking related tweets will potentially end up with the tweets that were also in our sample of tweets with the same proportions. Furthermore, we believe that by providing all the scripts on GitHub, it is transparent how we arrived at the results and if a researcher wishes to reproduce these results, they will need to apply for a Twitter API key (academic track) and then by running all the scripts they will arrive at the same results.

5 Discussion of results

5.1 Descriptive analytics

People in today's world have several avenues for expressing their views and opinions regarding a wide range of topics. Microblogging platforms are usually popular owing to their crisp format and to character limits set by the platforms. Among them, Twitter is one of the most famous microblogging platforms and it has existed uninterruptedly since 2006, acquiring more users and increasing activity. This long uninterrupted usage allowed us to conduct longitudinal research and study the change in public discourse over time.

To start with, we have argued that the real effect of financial technology on banking activities is increasingly more influential all over the world (Baba et al., 2020; Statista, 2020). Relying on the findings from the descriptive analysis, we can also claim that the changes in public discourse have been substantial as now approximately a fourth of the banking tweets are about fintech. Furthermore, it is important to note that fintech appeared in the banking-related tweets in 2014, but it very quickly became the number one word, as it was the fourth most used word in 2015, and starting from 2016, it has been the first. There are also many banks that have reacted quickly to the fintech revolution as leading investment banks like Goldman Sachs, Citigroup, and JPMorgan Chase started to acquire many fintech companies starting from 2015 (CB Insights, 2021). However, as was explained in the literature review, other banks such as Wells Fargo and Handelsbanken, for example, have started to change their strategy from 2019. Certainly, more work needs to be done on the timeline when banks have internally changed their strategy because this paper relies only on limited public information. We can, however, conclude with confidence that the changes in discourse have been swift, and they took place in 2015 and 2016.

One might argue that financial technology came with a lot of hype. After all, as Figure 8 shows, fintech appeared in the public discourse as something that was liked and retweeted a lot, but later the likes and retweets that fintech-related tweets received tailed off. The simplest explanation is that fintech represented the classical Gartner's hype cycle as explained in Dedhayir and Steinert (2016), where a new technology first brings about a lot of intensive publicity. Though the difference in the number of likes and retweets became marginal starting from 2018, the finding that the tweets themselves became more positive and the number of fintech tweets has not decreased supports the argument that fintech-related discussion is a threat to the existing conception of control in the banking industry.

5.2 Network analytics

Furthermore, as was presented in the network analytics, the most central users were mostly either major news outlets or digital media platforms. Referring to the most visible users as those who received the most retweets and mentions, we can claim that the discussion around fintech is not lead by incumbent banks but rather by individual users who are devoted to fintech or major news outlets. This provides some support to Fligstein (1996) that the power structure is changing through the social movement around fintech and users related to fintech are now the most visible and consequently have more power over the discourse in banking-related discussions. However, based on user visibility and centrality analysis we also found that the social discourse on Twitter about banking is not driven by challenger firms, as was suggested by Fligstein (1996), but rather by users not associated with a particular company.

One interesting observation was that only 0.2% of users contributed to 44% of tweets, and 64% of users contributed to only 7% of all the banking-related tweets. Of the users that tweeted the most, the most visible were the ones who tweeted mostly about fintech. This offers some support for Mustafaraj et al. (2011) that the active and inactive groups tweet differently since the active group is mostly associated with fintech-related content and, as was shown fintech-related content is in turn associated with higher positive sentiment. Furthermore, this stark contrast in activity reminds the importance of controlling for user effects when conducting regression analysis.

As was elaborated in the literature review, the success of the social movement towards redefining the banking market depends on the size of the groups, their resources, the availability of a political opening, state actors' willingness to comply with the groups' complaints, and the capacity to build a social coalition around a new collective identity (McAdam, 1982; Snow et al., 1986; and Tarrow 1994). The network analysis showed that the size of the groups is large as the users who tweet about fintech are much more visible and often have a significant following. Also, as fintech is such a popular concept in banking and major retail banks are facing degrading trust (Blakstad and Allen, 2018 p. 2), the groups have the capacity to build a social coalition around it. However, state actors' actions regarding the regulation of fintech are largely in development (FSB, 2017). Therefore currently, we can conclude that the prerequisites for a successful transformation regarding the size of the groups that represent fintech and the capacity to build a social coalition are certainly present.

5.3 Hypothesis results

As the descriptive and network analytics showed, fintech has truly ingrained itself into banking-related discussions and there are concentrated and visible groups on Twitter that represent fintech. The results from the regression analysis showed that the discourse around fintech is associated with much higher positive sentiment and the positivity has increased over time.

Naturally, a question arises why fintech-related tweets are associated with higher positive emotion? Firstly, the investments that have been pouring into fintech companies (Statista, 2020) and the increasing valuations of public fintech companies represent a general positive belief in the future of financial technology and this belief is also reflected in the positive public opinion.

Furthermore, even though FSB (2019) and Baba et al. (2020) concluded that the effect of fintech firms on banking activities is quite small in that they contribute little to credit provision, fintech is disproportionately more popular than other banking-related topics on Twitter and it inhibits significantly more positivity. We conclude that this is a strong sign that the conception of control in the banking market is changing. This is also supported by the observations that many innovative solutions that have influenced banking services have come from new fintech companies (Chishti and Barberis, 2016 p.7). In addition, the sentiment regarding fintech is not only positive, but it has become more positive over the years, which is also a sign that over the years, fintech has gained more support.

What is more, we explained that before the economic recession in 2008, the shared understanding of the international banking market was to invest in U.S. mortgage-backed securities (Fligstein, 2013). Similarly, currently the banking market is largely influenced by financial technology and since the public has a positive stance towards it, we can expect more banks to follow the invading fintech companies and adjust their services accordingly. Therefore, we argue that the power to decide how and which banking services are provided will be increasingly more decided by small fintech companies rather than incumbent banks.

Lastly, as the public sentiment is becoming more positive towards fintech, we argue that the social movement is being scaled up (as in Mundt et al., 2018) and intensified (as in Shirazi, 2013). We argue that this implies that the power shift from incumbent banks towards fintech firms is becoming increasingly more pronounced and the “fintech revolution” termed by Blakstad and Allen (2018) and Gomber et al. (2018) will continue.

6 Ethical Considerations

It is asserted in Bryman and Bell (2017) that when conducting research, there are certain rules and ethical aspects that need to be respected and followed. When people sign up to Twitter, they agree with the Terms of Service, Privacy Policy, Twitter Rules and Policies, and all the other incorporated policies, which enable academic researchers to use their public output for research. However, Taylor and Pagliari (2017) point to the deficit in ethical guidelines and the overall low-level of awareness when using the social media data for research purposes.

Some of the most important ethical aspects to consider are interference with the public discourse and using sensitive personal information (such as health) (Taylor and Pagliari, 2017). Even though we would not have a significant effect in steering the public discourse around the three hashtags that were used, we did not post any public tweets regarding banking and fintech. Also, we restrained from using any sensitive and personal user information intentionally or specifically, though it may be possible that among the four million tweets there were some that were about personal financial condition which is considered as sensitive personal information.

Moreover, we have knowingly avoided naming any individual username, whether an organization or an individual, since we do not have the explicit consent to display their tweets or usernames. Even though naming the individuals could have benefitted the presentation of results (i.e., in network analytics), we have intentionally kept the research on an aggregate level to maintain the privacy of users.

We have taken caution to work in line with the terms and conditions provided by Twitter and to work within the boundaries of the application we have submitted and not stray away from it and exploit the API access. Data provided by Twitter consisted of the publicly available usernames (which are not all necessarily real identities) and the content of the tweets. Though there could have been important findings from conducting research on the users and their real-life profession to better understand where their opinions are coming from and how strongly they can potentially influence the public discourse, we have made a choice not to identify their occupation and personality outside Twitter.

The decision to use the Twitter platform for the research work was also made because it is an open public platform and presents lesser ethical issues than data available on closed or private online platforms (Townsend, and Wallace, 2016). Also, as we are using hashtags as a method to gather data to work on, the use of hashtags from users can be understood as their keenness to take part in the discussion and expect their views to reach a wider audience.

In line with GDPR guidelines (GDPR, Principle (c): Data minimization, Information Commissioner's Office, n.d.), we have only collected sufficient data to fulfil our stated purpose: we have collected relevant data and deleted that data which we didn't need any longer. We have sought access to Twitter data by filling an application that required us to state our project and its description and how we planned to use the Twitter data and their API's, methodology that will be employed for analysing the data, and mode of outcome sharing of the research conducted. We were also required by twitter not to use it for any commercial purposes. We have not employed the datasets either for our own benefit or commercial interests.

We have also taken steps to maintain the security of the data downloaded for analysis. For instance, good care has been taken to ensure that the API keys that Twitter provided did not leak (e.g., to GitHub). Also, as we queried the Twitter database multiple times, the datasets that were not included in this research or were considered redundant were immediately deleted.

7 Limitations

Some factors limit us in generalizing the results to the whole population that takes part in the public discourse about banking and how impactful the discourse is in determining real behaviours.

Firstly, we rely on the theory that explains the relationship between social discourse and market shaping. Though this theory has some empirical evidence from other industries, such as the electricity utility (Granovetter and McGuire, 1998), consumer discretionary, and consumer staples industries (Rindova et al., 2006); the banking industry can exhibit different characteristics where the social discourse and opinions do not influence industry-level change. Certainly, one determining factor in producing industry-level change in the banking industry is state regulation, which has also been the main force that has historically determined the course of the banking industry. Also, even though we argued that the most recent change taking place in the banking market is taking place within institutional constraints (as in Loasby, 2000), states can step in and impose harsh regulation, such as the Glass-Steagall Act that was imposed in the US in the 1930s, which can potentially constrain innovation (Bofondi and Gobbi, 2017). Therefore, although the social movement is strongly agitating change within the banking market and within the institutional constraints, state interests can hinder large structural changes, especially since banks are central in providing economic and social stability (Bofondi and Gobbi, 2017).

Another limitation concerns the general understanding and use of the banking and fintech constructs. We have defined the main functions of the banking market as: a) to decrease transaction costs, b) to give out long-term credit, and c) to transform the quality of assets by diversifying the investments (as in Gual, 1999; Chishti and Barberis, 2016, p. 8). However, when Twitter users discuss banking or fintech-related topics, they might discuss it in a different context or use the terms loosely. Although this limitation does not affect the results, since we can expect that in general there is a shared theme of the two important constructs, it limits us to conclude why exactly are users on Twitter more positive towards fintech. The purpose of this research was to give a holistic overview of the banking- and fintech-related discussion on Twitter, but to mitigate the limitation regarding the shared understanding of constructs, we encourage researchers to conduct a qualitative analysis with fintech and banking experts to understand better the underlying reasons for the movement towards fintech.

Furthermore, in view of the GDPR Guidelines and Twitter policies, we have constrained ourselves and did not associate Twitter accounts with real individuals as we would need their explicit consent. Therefore, we did not have access to the data about users' gender, geographical location, age, education, and occupation. Twitter has a gender imbalance, with 70% of users being male and 30% being female (Hootsuite, 2020). Similarly, only three countries (the U.S., Japan, and India) contribute to 50% of Twitter users. The same limitation applies to the age group, education, and occupation level data, which could not have been accessed. Thus, the results of this thesis are not representative of the population that has an impact on the social discourse on banking.

In addition, the form of content we wanted to analyse for the study was delimited to text format only. We decided against analysing other form of content such as photos, videos and URLs, and any other forms of media that could be used to communicate on Twitter. The method for analysing other formats, unlike for text analysis, are not as well developed and not used as extensively in research yet. The URL links which were used in Chae (2014), for example, contained valuable information regarding what kind of content users directed other users to.

Lastly, we have observed and discussed that 0.2% of total users contributed to 44% of tweets while 63% of users contribute to 7% of the total tweets. One needs to be cognizant of this distribution and this skewness could be a possible limitation since many people might be vocal and have strong influence in real-life decisions but passive in online discourse. The comparison to similar studies regarding the activity of users is limited owing to the uniqueness of the theoretical lens adopted, methodology, and the area of interest. As we have discussed, the language of tweets is strictly limited to English, so the views of any other language of tweets have been excluded, which limits the research to Anglophonic Twitter.

8 Conclusion

We have established with evidence from real-world business activities and academic papers that the banking industry is going through a fast transformation, fuelled by new technological advancements, increased adoption of technology by consumers, and changing consumer behaviours. In this paper, we developed a framework to study the change in social movements and we used the Twitter social media platform to study these changes.

We conclude that across all banking tweets, fintech-related tweets are associated with higher positivity and the positivity of the tweets has increased gradually over the years. We argued that the reason for these observations is that there is a social movement towards fintech in the banking industry and this social movement represents the change in the conception of control in the banking industry. To quote Woodly (2015): “A movement that effectively alters the terms of discourse can overcome considerable opposition and structural disadvantages to achieve sustained, meaningful change.” We argue that the immense popularity and positive sentiment of fintech on Twitter enables structural changes in the banking market and new financial technology firms to gain power over the incumbents. We also found that the most visible Twitter users in the public discourse around banking are mostly individuals that promoted fintech and concluded that these users have the most power to steer the conversation.

To answer the research question: "*How has the emergence of fintech affected the public discourse about banking among Twitter users?*" we conclude that since 2016 the number one topic of interest in banking-related discussions has been financial technology and fintech tweets are generally associated with a higher positive sentiment that has increased over the years. We add that the public discourse is now mostly controlled by users that are devoted to advancing fintech.

Furthermore, we consider the main limitations of this study to be the lack of information on state regulation regarding the new financial technology; evidence on real banking activity which would give us insights how banks are internally changing their strategy; and the limited sample of Twitter users, which may not necessarily be representative of the people who have a strong influence on the public discourse about banking.

One of the aims of this paper was to contribute to the academic discussion around how markets are shaped by providing a methodological framework to study the changes in social discourse. In this paper, we used the generalized estimating equation method to model the difference in sentiment between banking-related tweets that contain the keyword fintech and tweets that do not and we emphasize the importance of controlling for user fixed effects to study sentiment on Twitter on a large scale.

Lastly, researchers interested in studying the field of fintech can consider our findings that fintech first appeared in public discourse in 2015 and since 2016 it has been the most popular topic, which also exhibits significantly higher positive emotion. The research on fintech is still in its early stages and the focus of the research that has been conducted has been majorly on tangible aspects (Sangwan et al., 2019). In this thesis, we have contributed to the research of intangible aspects of fintech by choosing an intangible context such as public discourse.

9 Future Research

As we have found, the size of the group representing fintech is much more visible and organized than other groups on Twitter that discuss banking-related topics. Also, these groups are rather successful in building a new collective identity since the positive sentiment of fintech has been increasing over the years. However, McAdam (1982), Snow et al. (1986), and Tarrow (1994) elaborated on other conditions that need to be met in order to successfully transform a market. Most importantly, the availability of a political opening and state actors' willingness to comply are the other key factors that will determine whether the banking market will have a shift in the status quo. Thus, future research regarding the regulation of financial technology, specifically on how states will attempt to protect national banks from international fintech companies, is welcomed. This will provide additional insights into whether the social movement proves to be successful.

Furthermore, as we have suggested in the limitations section, we encourage future research to address the limitation regarding the shared understanding of constructs. By conducting a quantitative analysis with banking and fintech experts, researchers can better understand the underlying reason for the movement towards fintech. Chishti and Barberis (2016, p. 7) and Blakstad and Allen (2018, p. 2) suggested that degrading trust resulting from the most recent financial crises has created a negative image of commercial banks and increased hope for the new financial technology.

Lastly, with the help of visual aid in Figure 9, we have observed a relationship between the number of fintech tweets made and the investments made in fintech by three large US banks. We encourage researchers to collect more data and investigate the investments made by banks in fintech. By using time-series analysis with a better dataset on banks' investment into fintech, one could control for autocorrelation and model the relationship between investments made by banks and the sentiment of fintech-related tweets.

10 References

Agrawal, A., & Hamling, T. (2017). Sentiment analysis of tweets to gain insights into the 2016 US election. Columbia Undergraduate Science Journal, 11.

Allen, F., Gu, X., & Jagtiani, J. (2020). A Survey of Fintech Research and Policy Discussion. In Working Paper (Federal Reserve Bank of Philadelphia). Federal Reserve Bank of Philadelphia. https://doi.org/10.21799/frbp.wp.2020.21

Alt, R., & Puschmann, T. (2012). The rise of customer-oriented banking-electronic markets are paving the way for change in the financial industry. Electronic Markets, 22(4), 203-215.

Alt, R., Beck, R. & Smits, M.T. (2018, August) FinTech and the transformation of the financial industry. Electronic Markets 28, 235–243. https://doi.org/10.1007/s12525-018-0310-9

Anstead, N., & O'Loughlin, B. (2014). Social Media Analysis and Public Opinion: The 2010 UK General Election. Journal of Computer-Mediated Communication, 20(2), 204–220. https://doi.org/10.1111/jcc4.12102

Baba, C., Batog, C., Flores, E., Gracia, B., Karpowicz, I., Kopyrski, P., Roaf, J., Shabunina, A., Van Elkan, R., Xu, X.C. (2020). Fintech in Europe: Promises and Threats. IMF Working Paper, European Department, WP/20/241.

Bagozzi, R. P., Gopinath, M., & Nyer, P. U. (1999). The role of emotions in marketing. Journal of the academy of marketing science, 27(2), 184-206.

Ballinger, G. A. (2004). Using Generalized Estimating Equations for Longitudinal Data Analysis. Organizational Research Methods, 7(2), 127–150. https://doi.org/10.1177/1094428104263672

Baron, A., Rayson, P., & Archer, D. (2009). Word frequency and key word statistics in corpus linguistics. Anglistik, 20(1), 41-67.

Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E. J., Berk, R., & Johnson, V. E. (2018). Redefine statistical significance. *Nature human behaviour*, *2*(1), 6-10.

Bing, L., Chan, K. C. C., & Ou, C. (2014, November). Public Sentiment Analysis in Twitter Data for Prediction of a Company's Stock Price Movements. 2014 IEEE 11th International Conference on E-Business Engineering. 2014 IEEE 11th International Conference on e-Business Engineering (ICEBE). https://doi.org/10.1109/icebe.2014.47

Blakstad, S., & Allen, R. (2018). FinTech Revolution. Cham, Switzerland: Springer, 121-32.

Bofondi, M. & Gobbi, G. (2017). The Big Promise of Fintech. European Economy: Banks, Regulation, and the Real Sector, Issue 2017.2 FinTech and Banks: Friends or Foes? Retrieved February 15, from https://european-economy.eu/2017-2/the-big-promise-of-fintech/

Bons, R. W., Alt, R., Lee, H. G., & Weber, B. (2012). Banking in the Internet and mobile era. Electronic markets, 22(4), 197-202.

Borg, A., & Boldt, M. (2020). Using VADER sentiment and SVM for predicting customer response sentiment. Expert Systems with Applications, 162, 113746.

Borgatti, S. P., & Everett, M. G. (2006). A graph-theoretic perspective on centrality. Social networks, 28(4), 466-484.

Bozdag, E., Gao, Q., Houben, G.-J., & Warnier, M. (2014). Does offline political segregation affect the filter bubble? An empirical analysis of information diversity for Dutch and Turkish Twitter users. Computers in Human Behavior, 41, 405-415. https://doi.org/10.1016/j.chb.2014.05.028

Broniatowski, D. A., Jamison, A. M., Qi, S., AlKulaib, L., Chen, T., Benton, A., & Dredze, M. (2018). Weaponized health communication: Twitter bots and Russian trolls amplify the vaccine debate. American journal of public health, 108(10), 1378-1384.

Bruns, A., Highfield, T., & Burgess, J. (2013). The Arab Spring and Social Media Audiences. American Behavioral Scientist, 57(7), 871–898. https://doi.org/10.1177/0002764213479374

Budiharto, W., & Meiliana, M. (2018). Prediction and analysis of Indonesia Presidential election from Twitter using sentiment analysis. Journal of Big data, 5(1), 1-10.

CB Insights. (2021). Where Top US Banks Are Betting on Fintech. CB Insights. https://www.cbinsights.com/research/report/fintech-investments-top-us-banks/

Handelsbanken. (2020 September 16). CEO of Handelsbanken Carina Åkerström at press conference [press release]. Retrieved May 5, 2021 from https://vp292.alertir.com/sites/default/files/report/20200915_presskonf_16_sep_eng.pdf

Chishti, S., & Barberis, J. (2016). The Fintech book: The financial technology handbook for investors, entrepreneurs and visionaries. John Wiley & Sons.

Citron, F. M., Gray, M. A., Critchley, H. D., Weekes, B. S., & Ferstl, E. C. (2014). Emotional valence and arousal affect reading in an interactive way: neuroimaging evidence for an approach-withdrawal framework. Neuropsychologia, 56, 79-89.

Clark, E. M., James, T., Jones, C. A., Alapati, A., Ukandu, P., Danforth, C. M., & Dodds, P. S. (2018). A sentiment analysis of breast cancer treatment experiences and healthcare perceptions across twitter. arXiv preprint arXiv:1805.09959.

Das, S. R. (2019). The future of fintech. Financial Management, 48(4), 981-1007.

Dedehayir, O., & Steinert, M. (2016). The hype cycle model: A review and future directions. Technological Forecasting and Social Change, 108, 28-41.

della Porta, D. and Mattoni, A. (2016). Social Movements. In the International Encyclopaedia of Political Communication, G. Mazzoleni (Ed.). Retrieved from https://doi.org/10.1002/9781118541555.wbiepc010

Eady, G., Nagler, J., Guess, A., Zilinsky, J., & Tucker, J. A. (2019). How Many People Live in Political Bubbles on Social Media? Evidence From Linked Survey and Twitter Data. SAGE Open, 9(1), 215824401983270. https://doi.org/10.1177/2158244019832705

Elbagir, S., & Yang, J. (2019). Twitter sentiment analysis using natural language toolkit and VADER sentiment. In Proceedings of the International Multi Conference of Engineers and Computer Scientists (Vol. 122, p. 16).

Fabocci, F. & Modigliani, F. (1992). Mortgage and Mortgage Backed Securities Markets. Harvard Business School Press.

Federal Reserve Bank of St. Louis. (1933). “Banking Act of 1933” June 16, 1933, https://fraser.stlouisfed.org/title/466/item/15952

Financial Stability Board [FSB] (2017). Financial stability implications from Fintech: supervisory and regulatory issues that merit authorities’ attentions. Retrieved from https://www.fsb.org/wp-content/uploads/R270617.pdf

Financial Stability Board [FSB] (2019). FinTech and market structure in financial services: Market developments and potential financial stability implications. Retrieved from https://www.fsb.org/wp-content/uploads/P140219.pdf

Fligstein, N. (1996). Markets as Politics: A Political-Cultural Approach to Market Institutions. American Sociological Review, 61(4), 656-673. Retrieved February 14, 2021, from http://www.jstor.org/stable/2096398

Fligstein, N. (2013). The Spread of the Worldwide Financial Crisis. Lecture presented at WZB Distinguished Lecture in Social Sciences event. Retrieved February 16, 2021, from https://www.youtube.com/watch?v=-BNR218UFX4&t=2087s

Gayo-Avello, D. (2013). A meta-analysis of state-of-the-art electoral prediction from Twitter data. Social Science Computer Review, 31(6), 649-679.

George, G., Osinga, E. C., Lavie, D., & Scott, B. A. (2016). Big data and data science methods for management research.

Gilbert, S. (2017, October 16). The Movement of #MeToo, How a Hashtag Got Its Power. THE ATLANTIC. Retrieved 15 March 2021 from https://www.theatlantic.com/entertainment/archive/2017/10/the-movement-of-metoo/542979/

Go, A., Huang, L., & Bhayani, R. (2009). Twitter sentiment analysis. Entropy, 17, 252.

Goddard, J. A., Goddard, J., Molyneux, P., Wilson, J. O., & Tavakoli, M. (2007). European banking: An overview. Journal of Banking and Finance, 31(7), 1911-1935. https://doi.org/10.1016/j.jbankfin.2007.01.002

Golder, S. A. (2017). Social Science with Social Media. Cornell University Library. https://doi.org/10.7298/X4NV9GCX

Gomber, P., Kauffman, R. J., Parker, C., & Weber, B. W. (2018). On the fintech revolution: Interpreting the forces of innovation, disruption, and transformation in financial services. Journal of Management Information Systems, 35(1), 220-265.

Granovetter, M., & McGuire, P. (1998). The Making of an Industry: Electricity in the United States. The Sociological Review, 46(1_suppl), 147–173. https://doi.org/10.1111/j.1467-954X.1998.tb03473.x

Gual, J. (1999). Deregulation, Integration and Market Structure in European Banking. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.183868

Hagberg, A., Swart, P., & S Chult, D. (2008). Exploring network structure, dynamics, and function using NetworkX (No. LA-UR-08-05495; LA-UR-08-5495). Los Alamos National Lab. (LANL), Los Alamos, NM (United States).

Hamdy, N., & Gomaa, E. H. (2012). Framing the Egyptian uprising in Arabic language newspapers and social media. Journal of Communication, 62(2), 195-211.

Handelsbanken. (2020, September 16). Handelsbanken gathering its forces at branches, accelerating the pace of digital development and cutting costs [Press release]. Retrieved March 5, 2021 from https://www.handelsbanken.com/en/press-and-media/news-and-reports

Hassan, S. U., Aljohani, N. R., Idrees, N., Sarwar, R., Nawaz, R., Martínez-Cámara, E. & Herrera, F. (2020). Predicting literature's early impact with sentiment analysis in Twitter. Knowledge-Based Systems, 192, 105383.

Hutto, C.J. & Gilbert, E.E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014.

Information Commissioner's Office. (n.d.). GDPR (General Data Protection Regulation), Principle (c): Data minimization. Retrieved May 5, 2021 from https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/data-minimisation/

Kauschke, C., Bahn, D., Vesker, M., & Schwarzer, G. (2019). The role of emotional valence for the processing of facial and verbal stimuli—positivity or negativity bias?. Frontiers in psychology, 10, 1654

Kong, M., Xu, S., Levy, S. M., & Datta, S. (2015). GEE type inference for clustered zero-inflated negative binomial regression with application to dental caries. Computational statistics & data analysis, 85, 54–66. https://doi.org/10.1016/j.csda.2014.11.014

Kou, Y., Kow, Y. M., Gui, X., & Cheng, W. (2017). One social movement, two social media sites: A comparative study of public discourses. Computer Supported Cooperative Work (CSCW), 26(4), 807-836.

Lee, I., & Shin, Y. J. (2018). Fintech: Ecosystem, business models, investment decisions, and challenges. Business Horizons, 61(1), 35-46.

Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. Biometrika, 73(1), 13-22.

Liu, B. (2010). Sentiment analysis and subjectivity. In Indurkhya, N., Damerau, F. J. (Eds.), Handbook of natural language processing (2nd ed., pp. 627–666). New York, NY: Chapman & Hall.

Loasby, B. (2000). Market institutions and economic evolution. Journal of Evolutionary Economics (10), 297–309. Retrieved February 14, 2021, from https://doi.org/10.1007/s001910050016

Lopes, A. R. (2014). The impact of social media on social movements: The new opportunity and mobilizing structure. Journal of Political Science Research, 4(1), 1-23.

Loria, S. (2018). textblob Documentation. Release 0.15, 2.

Manning, Christopher D., Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. (2014). The Stanford CoreNLP Natural Language Processing Toolkit In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, pp. 55-60.

Martinez-Camara, E., Martin-Valdivia, M. T., Urena-Lopez, L. A., & Montejo-Raez, A. R. (2012). Sentiment analysis in Twitter. Natural Language Engineering, 20(1), 1–28. https://doi.org/10.1017/s1351324912000332

McAdam, D. (1982). Political Process and the Development of Black Insurgency. Chicago, IL: University of Chicago Press.

McCormick, T. H., Lee, H., Cesare, N., Shojaie, A., & Spiro, E. S. (2017). Using Twitter for demographic and social science research: Tools for data collection and processing. Sociological methods & research, 46(3), 390-421.

Mønsted, B., Sapieżyński, P., Ferrara, E., & Lehmann, S. (2017). Evidence of complex contagion of information in social media: An experiment using Twitter bots. PLOS ONE, 12(9), e0184148. https://doi.org/10.1371/journal.pone.0184148

Mundt, M., Ross, K., & Burnett, C. M. (2018). Scaling social movements through social media: The case of Black Lives Matter. Social Media+ Society, 4(4), 2056305118807911.

Mustafaraj, E., Finn, S., Whitlock, C., & Metaxas, P. T. (2011, October). Vocal minority versus silent majority: Discovering the opinions of the long tail. In 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing (pp. 103-110). IEEE.

National Information Center (n.d.). A repository of financial data and institution characteristics collected by the Federal Reserve System. Retrieved February 19 from https://www.ffiec.gov/nicpubweb/Content/HELP/Institution%20Type%20Description.htm

O'Hara, R., & Kotze, J. (2010). Do not log-transform count data. Nature Precedings, 1-1.

Pagolu, V. S., Reddy, K. N., Panda, G., & Majhi, B. (2016, October). Sentiment analysis of Twitter data for predicting stock market movements. In 2016 international conference on signal processing, communication, power and embedded system (SCOPES) (pp. 1345-1350). IEEE.

Pang, B., Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2, 1–135. doi:10.1561/1500000011

Phua, J., Jin, S. V., & Kim, J. (Jay). (2017). Uses and gratifications of social networking sites for bridging and bonding social capital: A comparison of Facebook, Twitter, Instagram, and Snapchat. Computers in Human Behavior, 72, 115–122. https://doi.org/10.1016/j.chb.2017.02.041

Puschmann, C., & Powell, A. (2018). Turning Words into Consumer Preferences: How Sentiment Analysis Is Framed in Research and the News Media. Social Media + Society. https://doi.org/10.1177/2056305118797724

Rajan, R. G. (1998). The Past and Future of Commercial Banking Viewed Through an Incomplete Contract Lens. Journal of Money, Credit and Banking, Vol. 30, No. 3, Part 2 (August 1998). Retrieved from https://www.jstor.org/stable/2601255?seq=1

Reuters (2020). Handelsbanken cuts jobs, closes Swedish branches to invest in IT. https://www.reuters.com/article/handelsbanken-costs-idINKBN2670U1

Rindova, V. P., Pollock, T. G., & Hayward, M. L. (2006). Celebrity firms: The social construction of market popularity. Academy of management review, 31(1), 50-71.

Romānova, I. and Kudinska, M. (2016), "Banking and Fintech: A Challenge or Opportunity?", Contemporary Issues in Finance: Current Challenges from Across Europe (Contemporary Studies in Economic and Financial Analysis, Vol. 98), Emerald Group Publishing Limited, pp. 21-35. https://doi.org/10.1108/S1569-375920160000098002

Rowan, L. (2020, October 01). No, banks aren't closing tons of branches during the pandemic. Forbes. Retrieved March 5, 2021, from https://www.forbes.com/sites/advisor/2020/09/30/no-banks-arent-closing-tons-of-branches-during-the-pandemic/?sh=1dd071640421

Sangwan, V., Prakash, P., & Singh, S. (2019). Financial technology: a review of extant literature. Studies in Economics and Finance.

Schneider, L. A., Shaul, M. A. X., & Lascelles, C. K. (2016). Regulatory priorities for FinTech firms--and investors--in the coming year. Journal of Taxation & Regulation of Financial Institutions, 29(4).

Schueffel, P. (2016). Taming the beast: a scientific definition of fintech. Journal of Innovation Management, 4(4), 32-54

Shirazi, F. (2013). Social media and the social movements in the Middle East and North Africa. Information Technology & People.

Snow, D.A., Rochford E.B., Worden, S., and Benford R.D. (1986). Frame Alignment and Mobilization. American Sociological Review 51 (4), pp 464-481.

Soleymani, M., Garcia, D., Jou, B., Schuller, B., Chang, S. F., & Pantic, M. (2017). A survey of multimodal sentiment analysis. Image and Vision Computing, 65, 3-14.

Starbird, K., & Palen, L. (2012, February). (How) will the revolution be retweeted? Information diffusion and the 2011 Egyptian uprising. In Proceedings of the acm 2012 conference on computer supported cooperative work (pp. 7-16).

Statista. (2020, February). Total value of investments into Fintech companies worldwide from 2010 to 2019. Statista Research Department. Retrieved March 5, 2021 from https://www.statista.com/statistics/719385/investments-into-fintech-companies-globally/

Tarrow, S. (1994). Power in Movement. Cambridge, England: Cambridge University Press.

Taylor, J., & Pagliari, C. (2017). Mining social media data: how are research sponsors and researchers addressing the ethical challenges? Research Ethics, 14(2), 1-39.

Townsend, L., & Wallace, C. (2016). Social media research: A guide to ethics. University of Aberdeen, 1, 16.

Twitter. (2014, December 10). “The 2014 #YearOnTwitter”. Twitter Blog. https://blog.twitter.com/official/en_us/a/2014/the-2014-yearontwitter.html

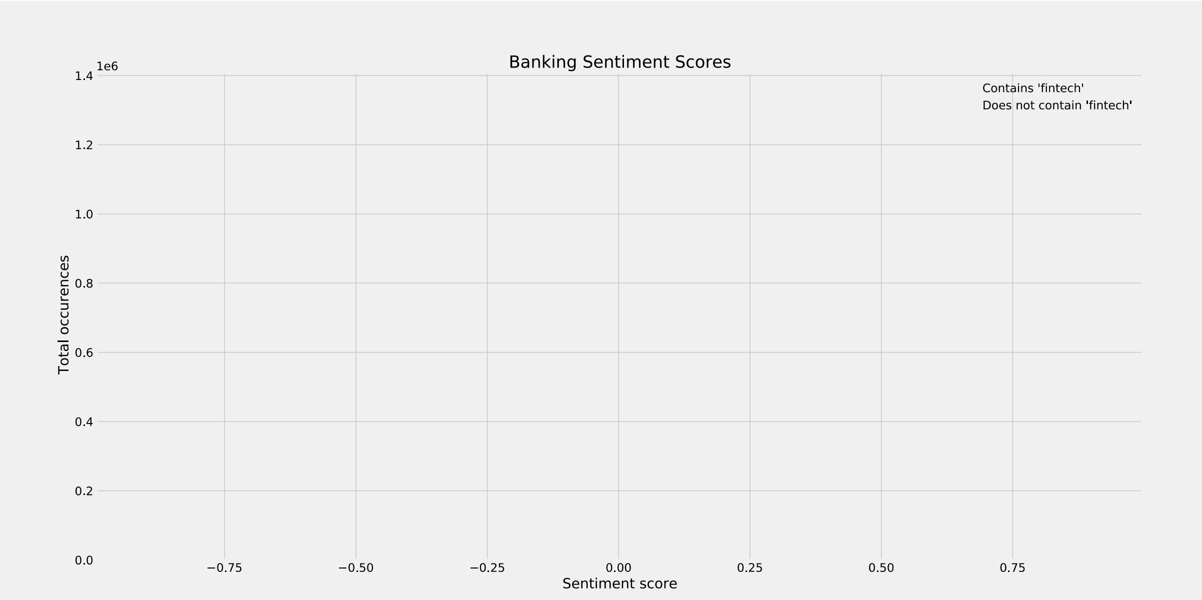
Twitter. (2021, February 9). Selected Company Metrics and Financials – Monetizable Daily Active Usage (mDAU): Worldwide. Twitter. Retrieved March 5 from https://s22.q4cdn.com/826641620/files/doc_financials/2020/q4/FINAL-Q4'20-TWTR-Selected-Metrics-and-Financials.pdf

Woodly, D. R. (2015). The politics of common sense: How social movements use public discourse to change politics and win acceptance. Oxford University Press.

Zeng, D., Chen, H., Lusch, R., & Li, S. H. (2010). Social media analytics and intelligence. IEEE Intelligent Systems, 25(6), 13-16.

11 Appendices

Appendix 1. Sentiment scores for banking- and fintech-related tweets



Note. Figure created by the authors.

Appendix 2. Word Frequencies of banking-related tweets

	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
fintech	3	18	18	118	445	3093	29655	84604	107432	107377	122972	101856	21417
finance	1875	8322	11634	25133	27575	27791	43853	51945	43944	45477	48249	51946	12537
digital	17	82	331	1074	1970	5220	17444	22043	28130	35607	42451	38401	8126
payments	77	693	820	2507	2573	3526	10377	13680	16770	25218	37435	34770	6408
blockchain	0	0	0	0	0	122	4808	19709	25864	42773	41713	33270	3631
money	1149	4708	7909	12273	15350	13438	20804	19511	18273	27133	42255	33037	9708
business	2189	6478	11009	14146	11282	12247	42065	46814	21737	31502	35828	28732	9009
financial	3329	7180	9223	14223	11695	10120	16177	18505	20012	30415	30090	27738	6254
technology	110	534	1083	1824	2035	3226	7602	10554	14670	20845	22267	24152	3783
new	1227	8727	9976	13217	14911	14446	24278	26571	23671	26687	29730	23211	5446
ai	1	1	9	11	16	41	299	3256	30149	31162	27080	19969	4638
bitcoin	0	2	42	175	1751	4534	9090	8176	13973	19155	24669	19640	3479
read	214	396	887	1321	1896	2614	4178	4825	5457	10404	9235	18167	2397
crypto	0	4	1	35	11	135	580	613	2841	19509	28924	17965	3303
coronavirus	0	0	0	0	0	0	0	0	0	0	0	16753	568
credit	1556	7311	6365	8892	6238	5180	8105	9295	7140	10574	12293	16592	3501
finserv	0	9	40	542	734	769	1752	3938	9664	20419	19158	16104	4445
latest	90	182	442	1005	1056	1253	2597	8229	13193	21966	18785	15692	2515
innovation	92	195	726	1452	1861	3087	7763	11414	14763	21044	18754	14734	4302
online	267	1636	2625	3054	3209	3871	8093	6564	5429	5999	7832	14570	6454

Notes. Table sorted by 2020-word frequencies. Table created by the authors.

Appendix 3. Model 2 Regression Results

	Sentiment
Intercept	*0.144*
	*(199)****
*Fintech * 2015*	*-0.0102*
	(-1.755)
*Fintech * 2016*	*-0.0097*
	*(-2.640)***
*Fintech * 2017*	*0.0063*
	(1.941)
*Fintech * 2018*	*0.0588*
	*(11.972)****
*Fintech * 2019*	*0.0670*
	*(12.340)****
*Fintech * 2020*	*0.0649*
	*(12.132)****
*Fintech * 2021*	*0.0943*
	*(10.471)****
Dependence structure	*Exchangeable*
Scale parameter	*0.141*
Observations (total tweets)	*2837443*
Clusters (total users)	*253259*
Mean cluster size	*11.2*
Min. cluster size	*1*
Max. cluster size	*54499*

Notes. t-statistics for robust standard errors clustered at the user level in parentheses.
*Statistical significance: *<.05, **<.01, ***<.005.*
Table created by the authors.