

“Warren Buffet is my cousin”: Shaping Public Understanding of Big Data
Biotechnology, Direct-to-Consumer Genomics, and 23andMe on Twitter

*Peter A. Chow-White

*Stephan Struve

*Alberto Lusoli

*Frederik Lesage

**Nilesh Saraf

*Amanda Oldring

*School of Communication

**Beedie School of Business

Simon Fraser University

8888 University Drive

Burnaby, BC V5A 1S6

Corresponding author:

Peter Chow-White

petercw@sfu.ca

Abstract: Scholars, educators, regulators, pundits and other observers are advocating for regulation and oversight of direct-to-consumer (DTC) genomic testing. As a result, the technology has been subject of highly visible public and regulatory controversy. In this article, we explore the nature and the shape of the sentiment of public discourse about the DTC company, 23andMe. We conduct a quantitative content analysis and qualitative framing analysis on Tweets. We find the discourse surrounding DTC genomics and 23andMe is largely positive. To describe the structure of this discourse our analysis reviews particular frames that seem to underlie social media exchanges related to the company and its technology. We argue that these frames create meaning around this emerging technology for its users.

Keywords: Social media, Twitter, direct-to-consumer (DTC) genomics, 23andMe,
content analysis, framing analysis

The scientific success of the Human Genome Project (HGP) and its successor the International Hapmap Project paved the way for personalized medicine and direct-to-consumer (DTC) genomics (Collins, Green, Guttmacher & Guyer, 2003; Hunter, Khoury & Drazen, 2008; McGuire, Diaz, Wang & Hilsenbeck, 2009; Harris, Wyatt and Kelly, 2013; Wiechers, Perin, & Cook-Deegan, 2013). The DTC genomics industry emerged in the early 2000s and aimed to improve and personalize healthcare by leveraging knowledge about a person's DNA (Caulfield & McGuire, 2012; Borry, Cornel & Howard, 2010; Heshka, Palleschi, Howley, Wilson and Wells, 2008). Stakeholders had high hopes for personalized medicine and the role of DTC genomics in these early days (Collins, Green, Guttmacher & Guyer, 2003). Some claimed DTC genomics created a new "developmental stage in human medicine" (Gane, 2008, p. 799). Time Magazine called DTC "revolutionary" and its "Innovation of the year" in 2008 (Hamilton, 2008). Since then, over 30 companies appeared. For example, 23andMe, Pathway Genomics, and Gene By Gene deliver personal genomic information to their customers for under one hundred dollars (Borry, Cornel & Howard, 2010). This cost represents a sharp decline over a decade prior when genomic tests cost thousands of dollars (U.S. National Library of Medicine, 2015). Retail genomics allows end-users to access data and genomic information that was once the preserve of trained researchers and physicians. The reactions from doctors (Kuehn, 2008) and from users (Wilde, Meiser, Mitchell & Schofield, 2009) to this newfound access has been both positive and negative.

Proponents describe retail DNA tests as tools for democratizing access to health information (Green & Farahany, 2014; Howard & Borry, 2009). The genomic information from the tests enable individuals to take control of their health (Caulfield,

Ries, Ray, Shuman, & Wilson, 2010; Caulfield & McGuire, 2012). Others claim genomic devices have limited clinical value (Baudhuin, 2014; Hunter et al., 2008; McGuire et al., 2009; Kuehn 2008). They also argue DTC services may create misperceptions among consumers about their health conditions (Offit, 2008) and displace clinicians' primary role as health information providers and managers (Foster & Sharp, 2008).

Scholars are beginning to describe and understand the public's opinion on personalized health care because of the advancement of genomic technologies and new national programs like President Obama's Precision Medicine Initiative (Kaufman et al., 2010; McGowan, Fishman, Lambrix, 2010; Su, Howard & Borry, 2011). Studies highlight the willingness of the public to contribute to large scale genomic testing (Kaufman et al., 2016) or show the public support for direct-to-consumer genetic testing (Neumann et al., 2001, Kaufman et al., 2010, Bollinger et al., 2013). A Canadian survey finds the public is eager to participate in the advancement of personalized health care (Pushpangadan & Seckman, 2015). Research also shows that end users' voice is often excluded from debates concerning medical technologies, despite their willingness to participate in personalized health care (e.g., P. Su 2013, Y. Su et al. 2011). Social media is a public communicative space where researchers can hear potentially these voices.

Social media platforms enable a greater number of people and a larger set of stakeholders to engage in discussions and debates about the perceptions and social value of DTC genomics. Scholars like Harris et al. (2012) for instance used Twitter data as anecdotal evidence to examine the restructuring of clinical labor through DTC genomics. These sites play a role in the social shaping of emerging technologies where actors deploy competing discourses about the affordances and constraints of a

technology (Bijker, Hughes & Pinch, 2012). Observing Twitter allows researchers to see how a technological artifact is culturally constructed and interpreted (Bijker, Hughes, & Pinch, 2012).

In this paper, we investigate how actors shape sentiment about DTC genomic technology and personalized medicine on social media. We explore social media as a space for observing public discourse about DTC genomics. We conduct a quantitative content analysis and a qualitative frame analysis on an ‘average’ week in the daily life of 23andMe on Twitter. We argue Twitter and social media more generally are spaces where actors use discourse to socially shape new technologies.

We also have an empirical methodological goal. Like many researchers these days, we wanted to understand how to conduct social media research on Twitter. But we did not know how gather and analyze efficiently large Twitter data. There are computing science methods that require a certain level of expertise not normally found in the field of communication. At least not by the average researcher. We describe this process so other communication and social science researchers who are also interested can benefit from our experiences. We are not claiming to be unique or novel by using two off the shelf data collection programs. We want to share our experiences so others who are already using other techniques or may want to get into automated data collection can benefit and carry on the conversation. To begin, we describe DTC genomics and the significance of 23andMe. We follow with a discussion about the affordances and constraints of Twitter as a site of communication research.

Direct-to-Consumer Genomic Testing

Welcome to 23andMe, a web-based service that helps you read and understand your DNA. After providing a saliva sample using an at-

home kit, you can use our interactive tools to shed new light on your distant ancestors, your close family and most of all, yourself.

(23andMe.com)

Initially, DTC genomic testing was a commercial novelty and an expensive one at USD\$1,000 per kit (Offit, 2008). DTC companies advertised they could reveal a person's DNA ancestry, such as potential ties to Marie Antoinette or Genghis Kahn, racial identity, and health information. Tech magazines celebrated the new services and companies sold their kits as stocking stuffers. 23andMe has emerged as the biggest player (Kaye, 2008). 23andMe launched its services in November 2007. 23andMe referred to itself as the first personal genome service and offered to “unlock the secrets of your own DNA” (23andMe, 2015). Google and Genentech partnered to create the company. This partnership was significant because they were the largest information company and largest biotechnology company in the world at the time (Roche bought Genentech in 2009). Customers, investors and media hailed the company as the harbinger of the personal genomic revolution (e.g., Harris et al., 2012) ushering in the democratizing of health (e.g., Adams, 2013; Green & Farahany, 2014) by “empowering” individual customers (Kaye, 2008). 23andMe received regular mention across the mainstream and technology media (e.g. Goetz, 2010; Lee & Crawley, 2009).

The public exposure also attracted the attention of the U.S. Food and Drug Agency (FDA) that launched an investigation in the spring of 2010. The FDA claimed the test produced health information, and therefore had to be considered a medical device. 23andMe test should have obtained the safety, efficacy, and security approval prior to its commercialization (Baudhuin, 2014). 23andMe argued their proprietary technology and the science supporting it were safe and effective in estimating their

customers' risk factors. The FDA did not agree and banned the DTC test kit in December, 2013. The company continued to provide its ancestry service but was forced to remove the health report from its website, thus limiting the information provided to customers only to genealogical data¹. 23andMe and the FDA engaged in a two-year negotiation that led to a partial rehabilitation of the health report in 2015.

23andMe's regulatory battle with the FDA triggered an energetic debate about DTC genomic tests. News outlets widely reported on the latest scuffle between a biotech start-up and government regulator. The story also gained traction on social media. Many individuals and science professionals came to 23andMe's defense and criticized the FDA for getting in the way of innovation. In the week following the ban, the FDA went from being a regulator doing its job to a government agency getting in the way of people's access to emerging health services. By the end of the week, some users on the twittersphere called the FDA "anti-genetics." The debate can be read at different levels: not just a discussion between supporters and detractors of 23andMe, but also a public exchange about the legitimacy of the science and technology behind the tests and about the freedom of individuals to access their genomic information (for an overview of the debates, see Green & Farahany, 2014; Baudhuin, 2014). It also highlights the significance of social media platforms as sites for public discussion: not only did 23andMe use its blog to publicly respond to the FDA, but also thousands of users joined the discussion on Twitter and other social media.

Twitter as a research site

Twitter is a social network that allows anyone with an account to connect with

¹ The restriction only applies to US customers. UK and Canadian customers have always had unrestricted access to the health report on the 23andMe website.

other users and post short messages of up to 140 characters (Park, Rodgers and Stemmler, 2013; Huberman, Romero & Wu, 2008). Twitter messages in this social network are a form of “talk” or everyday speech that reproduces and affirms cultural values through the interaction between users and the aggregation of public tweets, or conversations (Murthy, 2013). The social media platform has been covered extensively for its role in politics (e.g., Murthy, 2015), disasters (Chew & Eysenbach, 2010), and delivering health-related messages (Park et al., 2013). The immense popularity and the social nature of Twitter has caught the attention of scholars in the field of communication and other social science disciplines (Hargittai & Litt, 2012).

Some scholars argue social media sites like Twitter present challenges for traditional social science research methods and requires new techniques of investigation (Karpf, 2012) to deal with the inherent fast-paced challenges of ‘Internet time’. Researchers also argue observing Twitter messages gives access to naturally occurring language, free from the engineering that occurs in staged research interviews or surveys (Jungherr, 2014). Some scholars have described this mining of Twitter data and analysis of textual data in real-time as *infoveillance* (Chew & Eysenbach, 2010). Others suggest Twitter discussion can also be viewed as a large focus group (Emery, Szczypka, Abril, Kim, & Vera, 2014). Instead of a dozen person panel, Twitter would allow access to a ‘focus group’ of a million people. These are interesting and provocative claims about the nature of social media discourse. Twitter does enable access to potentially big data sets. Typical limitations exist such as self-selection and access based on socio-economic status.

Twitter is an efficient opportunity for sampling the public discourse and the meanings constructed around the emerging DTC genomic technology. It provides large

quantities of publicly available data from a wide variety of accounts over time and space. Furthermore, the short form simplifies the qualitative human coding in text analysis. Studying tweets produced about a topic in real time can be an efficient and reliable means for understanding public sentiment and the formal and informal production of frames of meaning about a topic.

We examine the meanings associated with an emerging technology, and their potential discursive effects on its development (Bijker, Hughes, & Pinch, 1984). We examine public sentiment about genomic biotechnology and 23andMe in order to understand what frames of meaning are constructed about this new technology in social media discourse. Through this analysis, we explore the implications this discourse can have in the rhetorical closure of a technological artifact (Bijker, Hughes, & Pinch, 2012). Specifically, we ask the following research questions:

- RQ1: What are the user sentiments about 23andMe?
- RQ2: What are the positive and negative frames users construct about 23andMe?

We investigate these questions through a combination of content analysis and framing analysis. The sentiment content analysis measures the amount of discourse that is positive and negative. The framing analysis further explores the ways through which the positive and negative sentiments about 23andMe are articulated.

Data and Methods

The structured nature and vast size of Twitter data is ideal for quantitative analysis. Big data studies excel at identifying macroscopic network structures as well as identifying communication and behavior patterns within and across large, amorphous data sets (Bruns & Burgess, 2012). Yet too often they leave the social facets underlying these patterns in the first place untouched. In contrast, we opted for a small

data study that combines basic quantitative metrics with close readings of Twitter data to hone in on the microcosmic details of a single community within a much larger platform (Stephanson & Couldry, 2014). Doing so enriches our understanding of Twitter as a social phenomenon and circumvents many of the critical questions surrounding the social shaping of big data research.

Research protocol

Two researchers independently collected tweets from February 19 to February 25th, 2014, using NVivo, a content analysis software, and NodeXL, a network analysis software. Since both NodeXL and NCapture make use of the same application programming interface, the sampling frame of this research was the Twitter Search API. We identified the tweet population using the search keyword “23andMe” without a @ (mentions sign) or # (hashtag) appended. Often, community samples are drawn using an @ sign to gather tweets around a specific user, or a # sign to collect tweets around a given topic. 23andMe is a unique term and company name. We excluded all tweets from the official @23andMe account in order to reduce company-generated *branding noise* and an obvious source of positive sentiment. Also, we wanted to emphasize an often ignored component of the DTC genomic testing debate: the people’s voice (Su et al. 2011). In the period of our analysis (February 19th - February 25th 2014), the company published 32 tweets (1.54% of total): 23 original messages, 7 retweets and 2 responses. Of the 23 original tweets, 7 promoted contents coming from the 23andMe official blog, 2 advertised contests organized through Facebook page and the remaining 14 linked to generic news about medicine, health and technology. In terms of diffusion, each tweet received on average 4.8 retweets and 4.1 likes. The sentiment of the 23andMe account was, in the period of analysis, overwhelmingly positive: 16 tweets

were coded as positive, none as negative, and 16 as unassigned.

Table 1 goes about here

We gathered tweets on a twice-daily schedule, morning and evening, to account for differences in site activity over time that might bias the volume extracted to a particular time zone or demographics. Although the twice-daily sampling method helped mitigate sampling bias, it also resulted in a large number of duplicate tweets. To overcome the problem, we compared NCapture and NodeXL datasets and removed duplicates, resulting in a total of 2,075 unique tweets.

We conducted a content analysis on the tweets combining two methods: sentiment analysis and frame analysis. These methods are non-invasive, unobtrusive, and grounded in empirical observation. Moreover, they afford access to a population without introducing a survey scenario, which mitigates the social desirability bias or the Hawthorne Effect that often occurs in interview and survey techniques (Cheng, Fleischmann, Wang, Ishita, & Oard, 2010; Fleischmann, Oard, D.W., Cheng, A.S., Wang, & Ishita, 2009; Morris, 1994; Weber, 1990). We measured the sentiment of twitter statements about 23andMe and consumer genomics by classifying tweets as positive or negative. Subsequently, by means of interpretative frame analysis, we inductively identified the main patterns and themes of both the positive and the negative sentiment.

We used the qualitative analysis software Nvivo 10 for the framing analysis. We created a coding protocol as a research team and reached consensus on tweet and URL codes. We resolved disagreements through consensus. The researchers classified tweets as either positive or negative sentiment. All URLs included in tweets were also checked and title, first, and last paragraphs were coded for positive or negative sentiment. The

“unassigned” category was used for all those tweets lacking signals denoting the presence of one of the two sentiments. We tested intercoder reliability by comparing coding by four researchers on a sample of tweets (10%, n=200) and achieved 91% agreement². The researchers discussed areas of disagreement and once again established consensus across the group and re-coded wherever necessary.

The coding protocol guided the independent content analysis. It outlined specific rules and guidelines for identifying value-laden wording in order to determine whether or not attitude was expressed in each tweet. We used the codebook for deductive coding to assess each tweet for an expression of one or none of the attitudes. Positive tweets were determined to promote goodwill or “boost” 23andMe as a beneficial entity. Negative tweets were determined to denounce the company, or to indicate mistrust or skepticism. Tweets coded as Unassigned failed to specify any attitude towards the company. Coders were instructed to apply the Unassigned category if no attitude could be determined. Mixed attitude tweets were weighted by number of positive or negative connotation, and an ultimate value assigned, including both tweet and URL content. When sentiment in a tweet was unclear, the group would convene and discuss the potential interpretations of a tweet until consensus was reached. The protocols taken from that discussion were then applied to future analysis of tweets.

We analyzed each attitude independently and established larger frames to describe the overarching themes. We used the open coding approach (Babbie & Benaquisto, 2014) to develop these frames in recursive fashion. Through this procedure, the corpus of collected texts was broken down into discrete parts and analyzed to find

² We calculated intercoder agreement for each attitude (94% positive, 89% unassigned, 91% negative) and averaged all values. Kappa coefficient ranged from 0.76 to 0.9.

commonalities and differences amongst tweets. We grouped tweets according to their conceptual similarity or meaning relatedness (Strauss & Corbin, 1998). The researchers identified seven conceptual frames about 23andMe and DTC genomic test in general.

Results

The content analysis of positive and negative sentiment shows a predominance of the positive attitude in discourse pertaining to 23andMe (Figure 1). The vast majority of tweets (86%) with a sentiment connotation (thus excluding those categorized as Unassigned) have in fact been coded as positive. 14% of the tweets show a negative connotation. A majority of the discourse is positive while the negative discourse represents a much smaller fraction of the total discourse. In terms of Twitter accounts, positive and negative discourse feature particularly active accounts, strongly expressing their support, or critique, towards 23andMe.

Figure 1 goes about here

Figure 2 shows the most active account in the positive side (User7) posted 23 positive tweets, which account for 25.5% of the total tweets published by the top eight positive sentiment accounts. Figure 3 shows the most vocal critic to 23andMe (User11) published 15 negative tweets. This account produced more than a quarter (26.8%) of the total tweets published by the top eight negative sentiment accounts. While Figure 4 emphasizes how both accounts (User7 and User11) not only published positive and negative tweets respectively, but also tweets classified as unassigned. In detail, User7 published 5 unassigned tweets while User11 published 8 unassigned tweets alongside 15 negative tweets. While it is understandable to find positive or negative tweets coexisting along with unassigned ones within the same account, it is unclear what might have been the role of the latter in the discourse.

Figures 2-4 go about here

Interpretive Frames

We explored the qualitative dimensions of our tweets dataset using frame analysis. The results of this investigation helped us to understand which themes (frames) animated the discussions within the sentiment. The positive frames are: POS1) General excitement, POS2) Genetic reductionism, POS3) Ancestry, and POS4) Libertarianism/Paternalism. The negative frames are: NEG1) Scientific validity, NEG2) Genomic literacy, and NEG3) Risk (see Figure 5 for complete overview).

POS1. General Excitement

The public expressed general excitement about the potential of DTC genomic tests and they like to share their results. Curiosity-driven search for better lifestyles (Su, 2013; Roberts & Ostergren, 2013) and fascination (Su et. al, 2011) seem to motivate many citizens. In what can be described as a sense of emotional attachment and verbal hyperbole, users are eager to express such overly positive experiences.

A number of Tweets expressed a sense of excitement and positive predisposition towards the 23andMe test. They were often Tweets from actors who claimed to be 23andMe customers who shared the experience of undertaking a DTC genomic test. These tweets referred to technology as “unlocking” information from DNA. For instance, *“got my #23andme data today! pretty excited. I am looking forward to the health data being unlocked,”* *“in essence, high throughput. awesome. virtual high five,”* or *“just got my @23andMe results. Connecting with cousins and sharing my genome! @23andMe is going to replace @facebook - Mark my words!”*.

Tweets also described the test kit as empowering patients. For example, *“@User @23andMe I was thinking the same thing. Tools to empower patient to support their*

own diagnosis process #pgen #abcDRBcha". The same user also tweeted about the genetic test kit enabling diagnostic insights that they did not have access to before. A response to a tweet from the Mayo Clinic Center For Individual Medicine states: "*@MayoClinicCIM That's how I got my diagnosis finally - 23andme results! Yes, despite FDA concerns, they help #abcDRBchat*". We also found a recurring comments referring to the results as a 'gift of knowledge', figuratively and literally. "*@User bf got me a kit from [URL-REMOVED] for Xmas and I sent them my SPIT*".

POS2. Genetic reductionism

Are humans simply the sum and result of their DNA? Scholars like Sarkar (1998) discuss a certain lack of modesty that seems to surround personal genomics. Often, genomic code is treated as the essential answer to everything, from complex human behaviour, to intelligence, to social problems. This perspective favours genetic explanations over environmental ones, conceives human beings as the sum of their genetic sequence and sees DNA as over-determinants of most human traits such as behavior and intelligence, for example. Some studies have demonstrated that the discourse around DTC genomics revolves around such simplified descriptions (Singleton et al., 2012). This perspective forms a foundation for decision making about behavior changes. Some tweets refer to making lifestyle changes based on genetic test results. For example, "*Amazing taking the results of my 23andme and tweaking my diet. Dropped 4lbs this week alone.*" Actors argue the genetic information can provide information they can use to change their health behavior.

POS3. Ancestry

Some customers choose genetic testing to learn about their ancestry and identity. Su (2013) describes this function as identity-seeking for the "purpose of identifying

ancestry, paternity, and ethnicity” (p.360). The “ancestry” frame included all those tweets referring to the concepts of family, roots and belonging. It represented the positive attitudes toward the possibility to rediscover the past and reconnect with relatives. For example, *“According to 23andMe, my ancestry is 99.7% European, mainly British and Irish descent. Neatest Christmas present I’ve ever received #23andme”*, *“Just shared mother’s (76 yo) results with her. The 23andMe thing was a Xmas gift. Neat to confirm Eastern Eur/Solvenian roots”*, and *“This is all about me as told via 23andMe [URL-REMOVED] - A truly global mix. #knowyourhistory #beproud [URL-REMOVED]”*.

Users expressed excitement about their DNA results helping find relatives. One particular tweet, that can be considered a micro narrative, stated: *“My dad is standing in my kitchen taking a phone call from his new found brother right now. Thank you @23andMe!”*. This particular narrative seemed to resonate among contributors. It was retweeted 35 times. Ancestry results also stretched the definition of *cousins*. Several accounts reported about finding distant cousins (or a total 997 as in case of one account). Or, in the case of a Princeton employee, results were used to reach out to newly extended family: *“@WarrenBuffett Hi there...I just did 23andMe and found out we are distant cousins. I work at Princeton University”*. Interestingly, the question of ancestry was occasionally wrapped into notions of ‘good’: *“Nice results @23andMe: our Saharian daughter as a control of North-African ancestry and a good European profile for the rest of us.”*

POS4. Libertarian Vs Paternalism:

DTC genomic testing has been described as an act of empowerment, allowing citizens to “recuperate power away from medical and scientific institutions through

bottom-up activism and the political economy of hope” (Levina, 2010, p. 5). Some users saw the FDA’s intervention as a manipulation of this sense of agency. This frame collected all the exchanges about the FDA-23andMe dispute, in which 23andMe was often identified as an agent of change, of liberation against the paternalistic power of the FDA. This view in turn represented a conservative government unable to cope with new technologies and keep up with the speed of innovation. One statement indicated that the FDA was possibly overstepping their bounds: “*@FDA_Drug_Info I want my health info from @23andMe - why don't you go after food coloring or something.*” Along the same line, these accounts denounced the FDA as repressive: “*Having FDA impede anything clinical genomics not a full consumer win. As a patient glad to have my 23andMe results!*” and “*Why Policymakers Should Set 23andMe Free [URL-REMOVED] via @thedailybeast Agree and why does the 'FoodDrugA' have a say anyway?*” Others called on the FDA to stop all actions. Under the libertarian hashtag #libertarian, and citing independent science correspondent Ronald Bailey, one tweet stated: “*Leave 23andMe Alone [URL-REMOVED] #JeffsList , #libertarian.*” Similarly, another user denounced the FDA as anti-innovation: “*Free the data meets quantified self in 23andMe co-founder's startup, Curious [URL-REMOVED].*”

Some tweets were sympathetic to the cause of the FDA, but considered their judgement paternalistic and the ban excessive. Triggered by a post of a radiologist on a social media page for physicians, several tweets described the ban as “putting a libertarian in a dilemma” (Saurabh, 2014). In a blog post, written by a Twitter user, and whose link was retweeted numerous times, the radiologist stated he was “*libertarian seeped in the Austrian school of economics*” and that he was “*generally disposed against regulations*” (Saurabh, 2014). The debate of paternalistic influence also took on

an explicit political dimension. One account accused critics of the FDA of having a conservative orientation. One tweet states: *“From the right: @SenAlexander calls for clarity on citizen access to health info, citing FDA vs. @23andMe [URL-REMOVED]”* The mentioned senator has stated on his own senate homepage that *“this is the information age. Individuals should have direct access to personal health information, and federal policies should clearly support that goal”* (Alexander, 2014).

NEG1. Scientific Validity:

Criticism about DTC genomic testing has included concerns around the clinical value and science behind the tests. Su (2011) summarized some of the literature that deals with the discussions around the technological and scientific validity of these tests. Here, we re-identified some of these concerns. Scientific Validity was a negative frame that encompassed all tweets about scientific accuracy of the test. Tweets often expressed doubt and brought scientific research as counterarguments to the supposed validity of the 23andMe test. An account who identified as a biotechnologist in her bio and had a network of 1453 followers responded to another tweet’s mention of FDA impeding knowledge: *“@User @23andMe I found article interesting b/c it tested accuracy of sev companies (an FDA concern)& found patient could get dif result”* (sic). Further statements about the negative aspects of the test kits include: *“Very interesting stuff; FDA thought so too MT @User Here is a big reason @23andMe test should be regulated”*. This particular tweet seemed to resonate with other in her network as it was retweeted 14 times. Tweets also expressed concerns about the 23andMe algorithm: *“100 new relatives in the last 30 days? @23andme's algorithms must be a little off this week”*.

NEG2. Genomic Literacy:

Knowledge about the technology and the significance of the results is critical for understanding affordances and constraints. Critics argue the public may not necessarily be able to understand or evaluate the test results. Wade and Wilfond (2006) describe challenges of limited genomic literacy. If confronted with genetic counselors, who are experts in their domain “clients who do not have a strong grasp of genetics themselves may perceive themselves as being caught between two information sources that both claim expertise in genetics, yet offer different conclusions about the value of the test” (p. 284).

This negative frame refers to all those tweets that express concern about the 23andMe test because of unfamiliarity with the information provided or lack of clarity of the results. A number of tweets expressed a sense of information overload. For example, “*My 23andMe results came in this week. I've spent several days going over them and trying to understand it all. It's a lot of info*”. Users also found the genomic information difficult to understand: “*just got most of my data from @23andMe ..really complicated stuff*”. Sometimes users tried to help each other by answering each other's questions. However, we found a major roadblock to these discussions is varying levels of understanding genomics: “*@User @23andMe Nobody can understand your insider jargon. Breach the divide @User!*”.

NEG3. Risk:

Finding out about health risks motivates some DTC genetic tests consumers to purchase the kit (Gollust et al., 2012), but only a small part of the debate revolves around the actual risks of taking the test, such as the worry or anxiety (Singleton et al., 2012). Tweets that present individuals with information about the potential risks of

DTC genetic tests may affect motivations or intentions to get tested. For instance, Gray et al. (2012) found that individuals who are presented with information on the risks of DTC testing are less supportive. Individuals also display a particularly negative attitude towards the company and appeal to negative emotions in what has been described as emotional cost of negative results (Su, 2013).

This frame also includes comments that reject 23andMe or the general idea of using medical tests to learn about themselves. For example, *“Next time I stumble onto money... throw it away on a 23andMe DNA test, take the official Mensa test or buy more sessy cheap suits?”* or *“mmmm this [stuff] keeps on popping up on the internet 23andme it tellin me to buy it imma learn bout my self because some people aint [sh*t]”*. We found a number of tweets that discussed risks of DTC and 23andMe such as, *“23andMe: the pitfalls of genetic testing <http://t.co/JVMdm8V77i> #Bioethics”* (links to blog post about a false positive disease result), *“23andMe Is Terrifying, but Not for the Reasons the FDA Thinks <http://t.co/8Qj0ZAQWbr>”* and *“99 dollars to sell your DNA soul”*. Finally, tweets also expressed worry about the risks for their individual privacy and surveillance: *“@23andMe Do your searches remain stateside?”* (23andMe never responded to this question) and *“the retail genetics test is meant to be a front end for a massive information-gathering operation against an unwitting public.”*

Discussion and Conclusion

The quantitative sentiment content analysis found strong support for 23andMe and genomic health technology (RQ1). Only 10% of all collected Tweets expressed criticism. Both positive (58%) and unassigned tweets (32%) well outnumbered negative tweets. Overall, the discourse on Twitter was positive (86% excluding the unassigned tweets) which points to the continued adoption of this DTC genomic technology. People

tended to be enthusiastic to learn about their genome and curious about the science behind it. They also pushed back against government intervention in the regulation of 23andMe.

The unassigned tweets carry some ambiguity in their role in shaping the sentiment about 23andMe. They could play a role in directing users to supportive or critical information. As shown in the analysis of the most active users (Figure 4), unassigned tweets coexisted, at the level of individual user, along with positive and negative ones. Within our dataset, unassigned tweets often included links to peer-reviewed articles from *Nature News*, or announced the publication of a post, or to simply interesting posts with fellow users. Some tweets point to a high order summary of the topic and do not take position. One user states simply “*Personal #genomics an issue of debate #geneticstesting @23andme via the @thetartan <http://t.co/YxuwFu4P4G>*”. In the same fashion, some tweets link to educational resources, such as “*Everything You Ever Needed to Know About What To Do with Your 23andMe Raw Data Results... <http://t.co/7UPVoWwNQb>*”. Other tweets took advantage of the open discussion space and ask generally “*Has anyone used the 23andme genetic testing service before?*”. While seemingly neutral at the denotative level, these tweets serve a vital function for twitter users by linking to crucial resources. However, unassigned tweets could have assumed a different connotation depending on their use in support or in opposition to positive or negative tweets. For example, a tweet mentioning scientific research could be used to bolster a critique expressed in a previous negative tweet. If proven, this ambivalence of unassigned tweets would impact the same definition of positive and negative sentiment, which would be defined at the level of discussion and not of single tweet. Likewise, it would also impose a methodological

shift in the unit of observation that would be no longer the single tweet only, but also the context of a discussion created by connected tweets. The qualitative framing analysis helps us understand the nuances and contours of stakeholders different discursive moves.

The qualitative framing analysis (RQ2) identified some of the themes which also characterize the scientific debate, and its translation into the mass media, such as patients' empowerment (frame POS1), unmediated access to personal health information (frame POS2), entrepreneurial innovation Vs governmental conservatism (frame POS4) and skepticism towards the scientific validity of DTC genomic tests (frame NEG1). There was a curious finding in the qualitative data.

Sharing information about one's genome seemed to be a normal occurrence for Twitter users. We were surprised at how many people shared this very private information by posting either screenshots or the numerical result of their genetic test. According to Murthy (2013), it should not be a surprise that Twitter has become a familiar stage to talk about personal health outcomes, share screenshots of results, and opinions (See also Nelson, 2012). Health-related posts on Twitter "illustrate ways in which our bodies have become subject of Twitter's update culture" and that "tweeting about one's body marks an even greater shift" (Murthy, 2013 p.115). This open disclosure of health information also marked a shift in the public sphere, shrinking areas of life that were once considered private. In this regard, Murthy's concept provided an explanation for the relevance, in our sample, of the positive frames generic excitement (POS1), genetic reductionism (POS2), and ancestry (POS3). In fact, in these frames the 23andMe test results were often the starting point for larger reflections on ancestry, freedoms and empowerment.

This study faced a couple limitations that are typical to traditional media analysis and novel to Twitter. The short time span of tweet collection of seven days limits generalizability. Certain events, such as the FDA intervention, can heavily influence the nature of the discussion for a short period of time. A study with a larger scope could collect weekly data at different parts of the month or year. Researchers could also collect data over a longer time frame by ‘hand,’ build an automated data collection system, or purchase a tweet dataset from GNIP (now owned by Twitter) or another commercial site. Future research should emphasize the effect that closed and privately owned data can have on reliability and validity of established social scientific methods (González-Bailón et al., 2014).

The 140 character limit was a challenge for detecting sentiment. While we coded carefully the positive and negative sentiment in each tweet, it is not always possible to find a positive or negative signal. This could inflate the amount of ‘unassigned’ tweets because the intent of the user sending the information cannot be discerned. There is no clear solution in sentiment analysis for coding text with no sentiment. This is an element of sentiment analysis that needs to be addressed by future research. Despite these limitations, we found Twitter to be a rich communication space and a highly accessible site for data collection with available commercial and free research software.

Through the analysis of the twittersphere we have been able to identify which values are associated to a technology, and to what extent these values can influence the meaning of the technology and its adoption and use (Bijker, Hughes, & Pinch, 2012). In the case of 23andMe, the frames in the sample demonstrate how people interpret DTC genomic technology in several, generally positive, ways. Interestingly, we note how DTC genomic testing has become, within our sample, the means through which larger

interests and beliefs (e.g. freedom, government paternalism, health data ownership and scientific validity) are brought to the table for discussion. For this reason, applying a social construction perspective to social media can help communication researchers understand if these rhetorical moves can actually play a role in the closure of a technological artifact (Bijker, Hughes, & Pinch, 2012). Future studies on technological closure could examine the role of companies influencing social media discourse.

Bibliography

- 23andMe. (2008). *Haplogroups of the Rich and Famous*. *23AndMe Blog*. Retrieved October 29, 2016, from <https://blog.23andme.com/23andme-and-you/haplogroups-of-the-rich-and-famous/>
- 23andMe. (2012). *23andMe Takes First Step Toward FDA Clearance*. *23AndMe Blog*. Retrieved October 26, 2015, from <http://blog.23andme.com/news/23andme-takes-first-step-toward-fda-clearance/>
- 23AndMe. (2015). *What The FDA Decision Means For 23andMe Customers*. *23AndMe Blog*. Retrieved from <http://blog.23andme.com/news/what-the-fda-decision-means-for-23andme-customers/>
- Adams, S. A. (2013). Maintaining the collision of accounts: crowdsourcing sites in health care as brokers in the co-production of pharmaceutical knowledge. *Information, Communication and Society*, *17*(6), 657–669.
- Alexander, L. (2014). *Alexander Calls for Administration to Clarify Position on Patient Access to Personal Health Information*. *Senate Committee on Health, Education, Labor, and Pensions*. Retrieved October 29, 2015, from <http://www.help.senate.gov/chair/newsroom/press/alexander-calls-for-administration-to-clarify-position-on-patient-access-to-personal-health-information>
- Babbie, E., & Benaquisto, L. (2014). *Fundamentals of Social Research* (Third Cana.). Toronto, Ontario: Nelson Education Limited.
- Baudhuin, L. M. (2014). The FDA and 23andMe: Violating the first amendment or protecting the rights of consumers? *Clinical Chemistry*, *60*(6), 835–837.

- Berg, B. L. (2001). *Qualitative Research Methods for the Social Sciences (4th ed.)*. Boston, MA: Allyn and Bacon.
- Bijker, W. E., Hughes, T. P., & Pinch, T. F. (2012). *The Social Construction of Technological Systems. New Directions in the Sociology and History of Technology (Fourth Edi)*. Cambridge, Massachusetts: The MIT Press.
- Bollinger, J. M., Green, R. C., & Kaufman, D. (2013). Attitudes about regulation among direct-to-consumer genetic testing customers. *Genetic testing and molecular biomarkers, 17(5)*, 424-428.
- Borry, P., Cornel, M. C., & Howard, H. C. (2010). Where are you going, where have you been: a recent history of the direct-to-consumer genetic testing market. *Journal of Community Genetics, 1(3)*, 101–106.
- Bruns, A., & Burgess, J. (2012). Researching News Discussion on Twitter. *Journalism Studies, 13(5-6)*, 801–814.
- Cheng, A.-S., Fleischmann, K. R., Wang, P., Ishita, E., & Oard, D. W. (2010). Values of Stakeholders in the Net Neutrality Debate: Applying Content Analysis to Telecommunications Policy. *2010 43rd Hawaii International Conference on System Sciences*, 1–10.
- Chew, C., & Eysenbach, G. (2010). Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. *PloS one, 5(11)*, e14118.
- Caulfield, T., & McGuire, A. L. (2012). Direct-to-Consumer Genetic Testing: Perceptions, Problems, and Policy Responses. *Annual Review of Medicine, 63(1)*, 23–33.
- Caulfield, T., Ries, N., Ray, P., Shuman, C., & Wilson, B. (2010). Direct-to-consumer genetic testing: good, bad or benign? *Clinical Genetics, 77(2)*, 101–105.
- Collins, F. S., Green, E. D., Guttmacher, A. E., & Guyer, M. S. (2003). A vision for the future of genomics research. *Nature, 422(6934)*, 835–847.
- Emery, S. L., Szczypka, G., Abril, E. P., Kim, Y., & Vera, L. (2014). Are You Scared Yet? Evaluating Fear Appeal Messages in Tweets About the Tips Campaign. *Journal of Communication, 64(2)*, 278–295.
- Fleischmann, K. R., Oard, D.W., Cheng, A.S., Wang, P., & Ishita, E. (2009). Automatic classification of human values: Applying computational thinking to information

- ethics. *Proceedings of the American Society for Information Science and Technology*, 46(1), 1–4.
- Foster MW, Sharp RR (2008) Out of sequence: how consumer genomics could displace clinical genetics. *Nat Rev Genet* 9:419
- Gane, M. (2008). Conflicting Visions of Code-Work in Recent Social Science Fiction. *Information, Communication & Society*, 11(6), 799-815.
- Goetz, T. (2010). *Sergey Brin's Search for a Parkinson's Cure*. *Wired*. Retrieved December 14, 2015, from http://www.wired.com/2010/06/ff_sergeys_search/all/1
- Goff, D. A., Jones, C., Toney, B., Nwomeh, B. C., Bauer, K., & Ellison, E. C. (2016). Use of Twitter to Educate and Engage Surgeons in Infectious Diseases and Antimicrobial Stewardship. *Infectious Diseases in Clinical Practice*.
- Gollust, S. E., Gordon, E. S., Zayac, C., Griffin, G., Christman, M. F., Pyeritz, R. E., Wawac, L. & Bernhardt, B. A. (2012). Motivations and perceptions of early adopters of personalized genomics: perspectives from research participants. *Public health genomics*, 15(1), 22-30.
- González-Bailón, S., Wang, N., Rivero, A., Borge-Holthoefer, J., & Moreno, Y. (2014). Assessing the bias in samples of large online networks. *Social Networks*, 38(1), 16–27.
- Gray SW, Hornik RC, Schwartz JS, Armstrong K. (2012). The impact of risk information exposure on women's beliefs about direct-to-consumer genetic testing for BRCA mutations. *Clin Genet*. 2012;81:29–37.
- Green, R. C., & Farahany, N. a. (2014). Regulation: The FDA is overcautious on consumer genomics. *Nature*, 505(7483), 286–7.
- Hamilton, A. (2008). *Best Inventions of 2008*. *Time Magazine*. Retrieved from http://content.time.com/time/specials/packages/article/0,28804,1852747_1854493_1854113,00.html
- Hargittai, E., & Litt, E. (2012). Becoming a tweep: How prior online experiences influence Twitter use. *Information, Communication & Society*, 15(5), 680-702.
- Harris, A., Wyatt, S., & Kelly, S. E. (2012). The Gift of Spit (And the Obligation to Return it) How consumers of online genetic testing services participate in research. *Information, Communication & Society*, 16(2), 236-257.

- Heshka, J. T., Palleschi, C., Howley, H., Wilson, B., & Wells, P. S. (2008). A systematic review of perceived risks, psychological and behavioral impacts of genetic testing. *Genetics in Medicine : Official Journal of the American College of Medical Genetics*, *10*(1), 19–32.
- Hogarth, S., Javitt, G., & Melzer, D. (2008). The Current Landscape for Direct-to-Consumer Genetic Testing: Legal, Ethical, and Policy Issues. *Annual Review of Genomics and Human Genetics*, *9*(1), 161–182.
- Howard HC, Borry P (2009). Personal genome testing: do you know what you are buying? *Am J Bioeth* 9:11–13
- Huberman, B., Romero, D. M., & Wu, F. (2008). Social networks that matter: Twitter under the microscope. *First Monday*, *14*(1).
- Hunter DJ, Khoury MJ, Drazen JM (2008). Letting the genome out of the bottle—will we get our wish? *N Engl J Med* 358:105–107
- Jungherr, A. (2014). The Logic of Political Coverage on Twitter: Temporal Dynamics and Content. *Journal of Communication*, *64*(2), 239–259.
- Karpf, D. (2012). Social science research methods in Internet time. *Information, Communication & Society*, *15*(5), 639–661.
- Kaufman, D., Murphy Bollinger, J., Devaney, S., & Scott, J. (2010). Direct from consumers: a survey of 1,048 customers of three direct-to-consumer personal genomic testing companies about motivations, attitudes, and responses to testing. *In Annual Meeting of the American Society of Human Genetics in Washington, DC*.
- Kaufman, D. J., Baker, R., Milner, L. C., Devaney, S., & Hudson, K. L. (2016). A Survey of US Adults' Opinions about Conduct of a Nationwide Precision Medicine Initiative® Cohort Study of Genes and Environment. *PloS one*, *11*(8), e0160461.
- Kaye, J. (2008). The regulation of direct-to-consumer genetic tests. *Human Molecular Genetics*, *17*(R2), R180–R183.
- Kuehn BM. (2008). RIsks and benefits of direct-to-consumer genetic testing remain unclear. *JAMA*, *300*(13), 1503–1505.
- Lee, S. S.-J., & Crawley, L. (2009). Research 2.0: Social Networking and Direct-To-Consumer (DTC) Genomics. *The American Journal of Bioethics*, *9*(6-7), 35–44.

- Lenhart, A. (2009). *The Democratization of Online Social Networks*. Pew Research Center. Retrieved December 14, 2015, from <http://www.pewinternet.org/2009/10/08/the-democratization-of-online-social-networks/>
- Levina, M. (2010). Googling your genes: personal genomics and the discourse of citizen bioscience in the network age. *Jcom*, 9(1).
- McGiboney, M. (2009). *Twitter's Tweet Smell Of Success*. Nielsen Online. Retrieved December 14, 2015, from <http://www.nielsen.com/us/en/insights/news/2009/twitters-tweet-smell-of-success.html>
- McGowan ML, Fishman JR, Lambrix MA (2010). Personal genomics and individual identities: motivations and moral imperatives of early users. *New Genet Soc* 29:261–290
- McGuire, A. L., Diaz, C. M., Wang, T., & Hilsenbeck, S. G. (2009). Social networkers' attitudes toward direct-to-consumer personal genome testing. *The American journal of bioethics*, 9(6-7), 3-10.
- Morris, R. (1994). Computerized content analysis in management research: A demonstration of advantages & limitations. *Journal of Management*, 20(4), 903–931.
- Murthy, D. (2013). *Twitter: Social Communication in the Twitter Age*. Hoboken, NJ: John Wiley & Sons, Inc.
- Murthy, D. (2015). Twitter and elections: are tweets, predictive, reactive, or a form of buzz?. *Information, Communication & Society*, 18(7), 816-831.
- Nelson, A., & Win Hwang, J. (2012). Roots and Revelation: Genetic Ancestry Testing and the Youtube Generation. In L. Nakamura & P. Chow-White (Eds.), *Race After the Internet* (pp. 271–290). New York, London: Routledge.
- Neumann, P. J., Hammitt, J. K., Mueller, C., Fillit, H. M., Hill, J., Tetteh, N. A., & Kosik, K. S. (2001). Public attitudes about genetic testing for Alzheimer's disease. *Health Affairs*, 20(5), 252-264.
- Offit K. (2008). Genomic profiles for disease risk: Predictive or premature? *JAMA*, 299(11), 1353–1355.

- Pak, A., & Paroubek, P. (2010). Twitter as a Corpus for Sentiment Analysis and Opinion Mining. *LREc 10*. 1320-1326.
- Park, H., Rodgers, S., & Stemmler, J. (2013). Analyzing health organizations' use of twitter for promoting health literacy. *Journal of health communication, 18*(4), 410-425.
- Park, M.R. (2014). Big data in communication research: Its contents and discontents. *Journal of Communication 64*(2), 355-360.
- Pushpangadan, S. & Seckman, C. (2015). Consumer perspective on personal health records: A review of the literature. *Online Journal of Nursing Informatics (OJNI), 19* (1)
- Saurabh, J. (2014). *The FDA and 23andMe puts a libertarian in a dilemma. KevinMD*. Retrieved October 29, 2015, from <http://www.kevinmd.com/blog/2014/01/fda-23andme-puts-libertarian-dilemma.html>
- Shah, D. V., Cappella, J. N., & Neuman, W. R. (2015). Big Data, Digital Media, and Computational Social Science: Possibilities and Perils. *The ANNALS of the American Academy of Political and Social Science, 659*(1), 6–13.
- Singleton A, Erby LH, Foisie KV, Kaphingst KA. (2012) Informed choice in direct-to-consumer genetic testing (DTCGT) websites: a content analysis of benefits, risks, and limitations. *J Genet Couns. 2012*(21). 433–9
- Stephansen, H. C., & Couldry, N. (2014). Understanding micro-processes of community building and mutual learning on Twitter: a “small data” approach. *Information, Communication & Society, 17*(10), 1212–1227.
- Strauss, A., & Corbin, J. M. (1998). *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. Thousand Oaks, CA: SAGE Publications.
- Su, Pascal. (2013). “Direct-to-Consumer Genetic Testing: A Comprehensive View.” *The Yale Journal of Biology and Medicine 86* (3): 359–65.
- Su, Y., Howard, H. C., & Borry, P. (2011). Users’ motivations to purchase direct-to-consumer genome-wide testing: an exploratory study of personal stories. *Journal of Community Genetics, 2*(3), 135–146.
- U.S. National Library of Medicine, National Institutes of Health, Department of Health & Human Services. (2015). *What is the cost of genetic testing, and how long*

- does it take to get the results? Help Me Understand Genetics*. Retrieved from <http://ghr.nlm.nih.gov/handbook/testing/costresults>
- Weber, R. (1990). *Basic Content Analysis*. Thousand Oaks, CA: SAGE Publications, Inc.
- Wiechers, I. R., Perin, N. C., & Cook-Deegan, R. (2013). The emergence of commercial genomics: analysis of the rise of a biotechnology subsector during the Human Genome Project, 1990 to 2004. *Genome Medicine*, 5(9), 83.
- Wilde, A., Meiser, B., Mitchell, P. B., & Schofield, P. R. (2010). Public interest in predictive genetic testing, including direct-to-consumer testing, for susceptibility to major depression: preliminary findings. *European Journal of Human Genetics*, 18(1), 47-51.
- Wright, C. F., & Gregory-Jones, S. (2010). Size of the direct-to-consumer genomic testing market. *Genetics in Medicine*, 12(9), 594–594.
- Roberts, J. S., & Ostergren, J. (2013). Direct-to-Consumer Genetic Testing and Personal Genomics Services: A Review of Recent Empirical Studies. *Current Genetic Medicine Reports*, 1(3), 182–200.
- Sarkar, Sahotra. (1998). *Genetics and Reductionism*. Cambridge University Press.
- Wade, Christopher H., and Benjamin S. Wilfond. 2006. "Ethical and Clinical Practice Considerations for Genetic Counselors Related to Direct-to-Consumer Marketing of Genetic Tests." *American Journal of Medical Genetics. Part C, Seminars in Medical Genetics* 142C (4): 284–92, discussion 293.

Figures

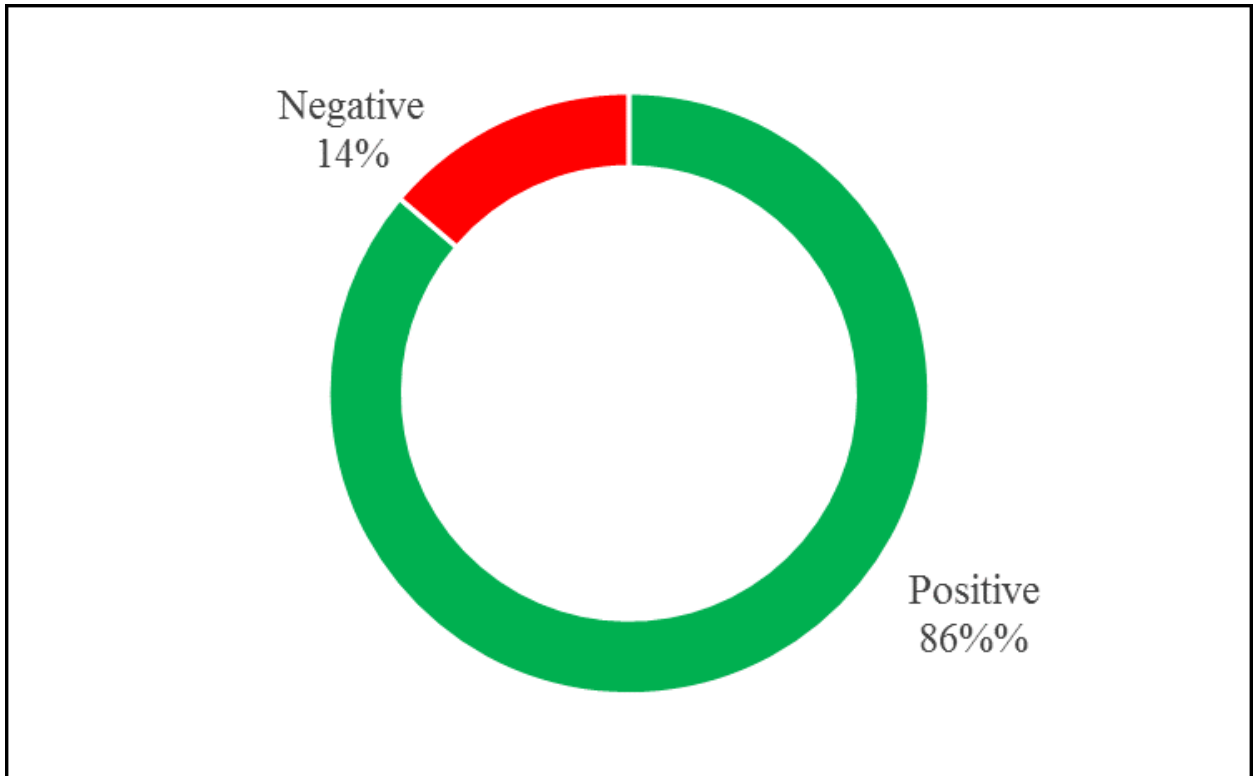


Figure 1. Positive and negative, sentiments of tweets (%).

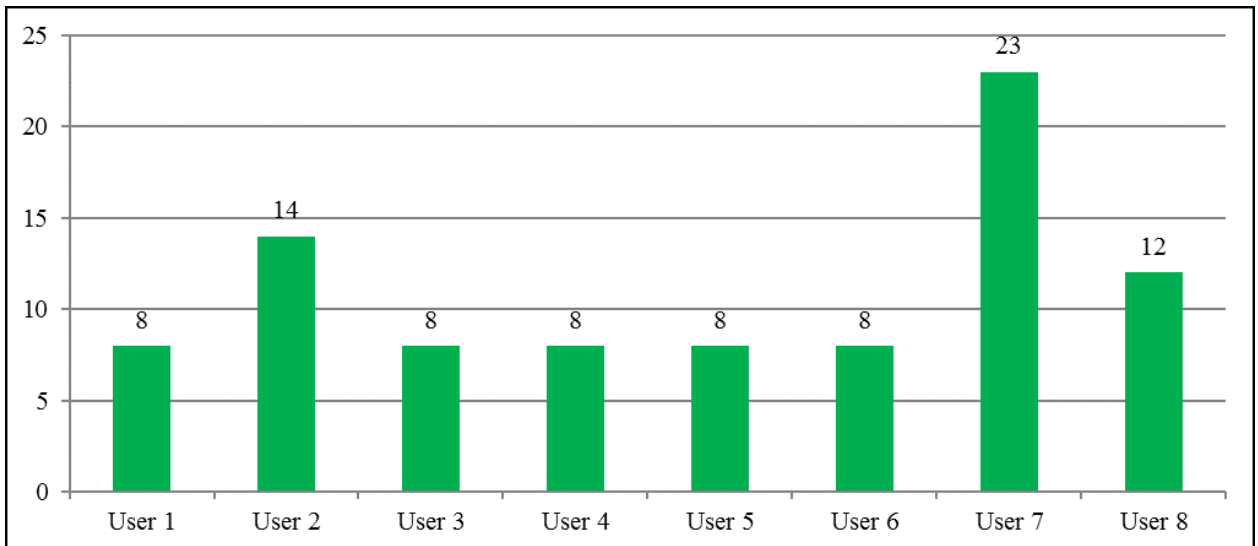


Figure 2. Top positive sentiment accounts based on tweet volume

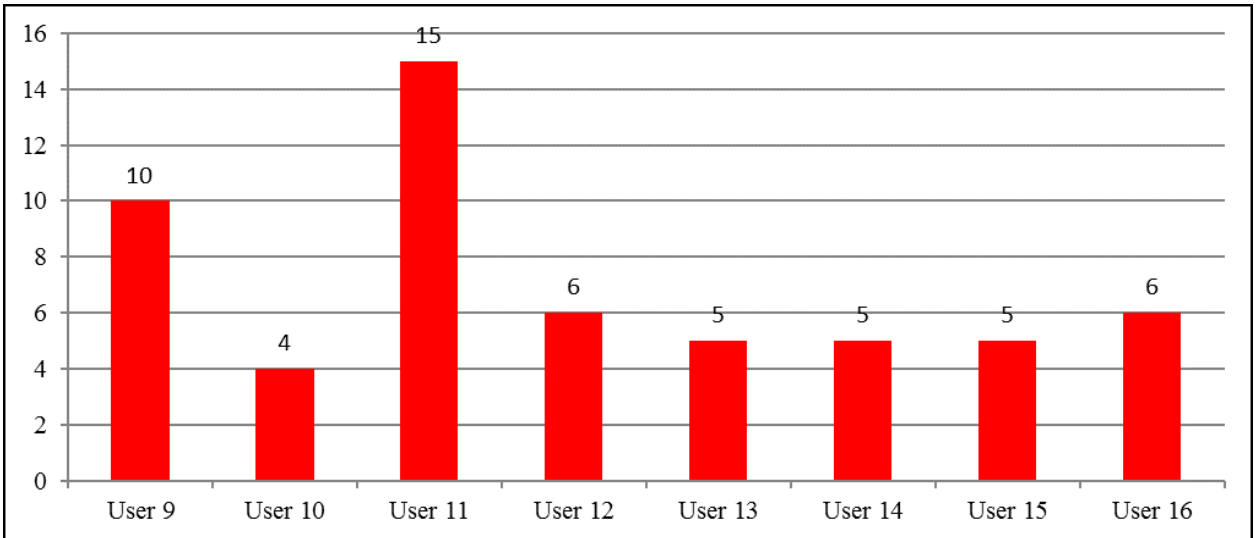


Figure 3. Top negative sentiment accounts based on tweet volume.

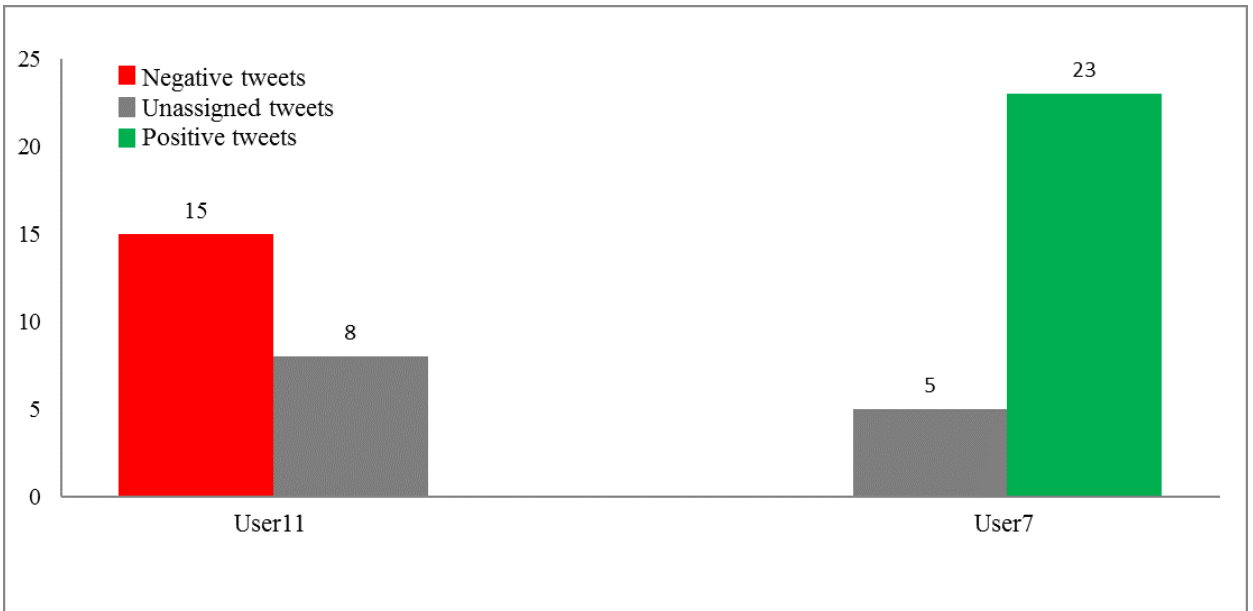


Figure 4. Unassigned tweets alongside positive and negative ones within the most active accounts.

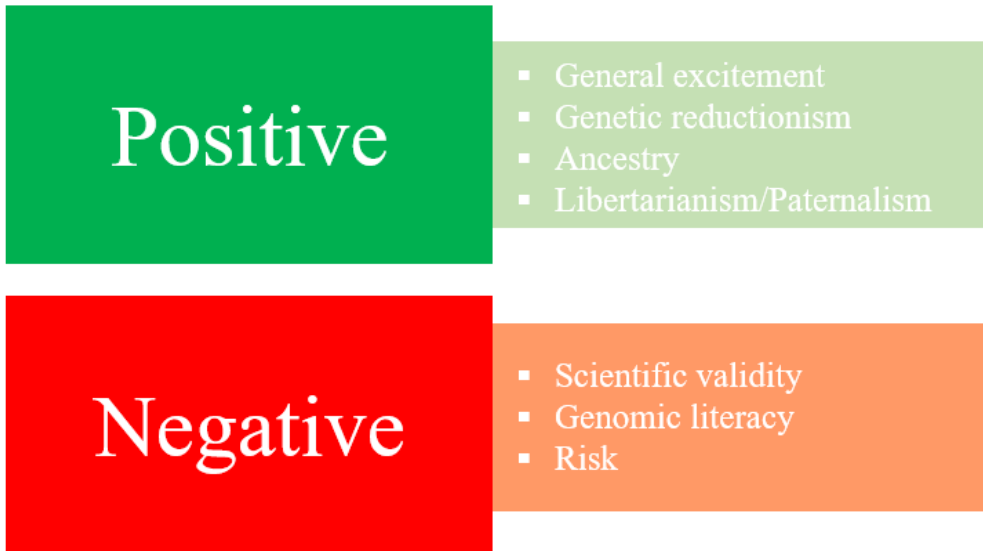


Figure 5. Overview of the interpretative frames for each sentiment.

Tables

		Of which positive	Of which unassigned	Of which negative
Tweets	2 3	14	9	0
Re-tweets	7	2	5	0
Responses	2	0	2	0
TOTAL MESSAGES	3 2	16	16	0

Table 1. 23andMe official Twitter account activity (Feb.19th 2014 – Feb.25th 2014).