Corporate Reputation in the Online Universe based on Social Media Content Analysis

VASILEIOS BASDRAS 1, KLIMIS NTALIANIS 2
[1]Faculty of Humanities and Social Sciences,
Open University of Cyprus
3, Giannou Kranidioti Ave., 2220, Latsia, Nicosia
CYPRUS
billbasdras@yahoo.gr
[2]Department of Business Administration
University of West Attica
Agiou Spiridonos 28, Egaleo 122 43, Attica
GREECE
kdal75@gmail.com, http://www.uniwa.gr

Abstract: - The current paper is a part of a master dissertation carried out at the Open University of Cyprus. Starting from an individual to a multinational corporation level we are looking to find, understand and define the purpose, the need and the impact of managing the impressions made. We argue that strategies may vary regarding the type of business. We suggest that anybody has an opinion as soon as they identify the brand. This opinion is disseminated and collectively can be seen as reputation. The progress in communication is implying bigger social influence between the people, trends are changing fast. We are presenting different tools to find these insights, which perform Online Reputation Management (ORM), free, paid or open source. Using theory from social science and technological facts, we discuss possible caveats. Finally we aim to use python scripts in order to collect, preprocess and analyze real data to see it in practice through the prism of a real business entity.

Key-Words: - corporate reputation, ORM, online reputation management, brand, sentiment analysis, customer insights, social media analytics

1 Introduction

Business reputation is probably dating since the first day of business. Reputation though is not something concerning exclusively the business world. From the individual level that we care to feed our social media profile to multinational corporations we try to manage the impressions we made [1]. So what is reputation, why is important and who is addressed to are all questions that people have been bothered since the very beginning. Our approach and mission is to juxtapose to the reputation in the on-line universe.

People are living in highly organized structures. On our daily lives we depend on each other in many different levels. We are social creatures. We exchange information and we influence each other. Big recent advances in technology add even more power in our communication needs. Therefore we can disseminate information with an unprecedented rate. Unfortunately communicating f2f in the physical world is not the same with CMC (Computer Mediated Communication). The technology we use at the moment cannot transmit in all cases the nuances of the face, the body language, the prosody etc. [2].

In a generic way someone has to build, maintain or recover the reputation of his/her brand. How we can utilize technology to make something viral, address scandals, find insights, make a competitive analysis and most importantly how we deal with this plethora of data.

2 In Search for Definitions

2.1 Impression Management

As expected the social complicatedness that has been inherited to us is exerted in our online behavior as well. In CMC (Computer Mediated Communication) the users are trying to manage the impression to facilitate the desirable relationships [3]. This communication offers different ways of deception though. It can be identity-based or message-based and of course a combinations of both [4]. An embarrassing event is something we usually try to avoid disclosing.

2.2 Restrictions

In the 'pathetic dot theory' there four constrains that limit our behavior and ultimately our individuality. These are: architecture, markets, law and norms, while the pathetic dot is us, helpless to avoid their forces. People have to behave under certain ways [5]. Laws are written but norms are not, so how to know how to behave?

2.3 Homophily

People are grouping themselves with people with similar attributes, in social science this concept is called Homophily. There are two ways that this is facilitated, selection and social influence or socialization. The first one, selection, is related more with immutable characteristics, such as race or ethnicity. On the contrary social influence serves to shape the characteristics of people as discussed in the work presented in [6].

2.4 Corporate Reputation

People can be confused between corporate ***Identity,*** corporate ***image*** and corporate ***reputation.*** In the literature review, of the mentioned ones, those are certainly distinguishable. The corporate identity can be seen as the underlying core or basic character of the firm [7]. Image is a general impression of a corporation's distinct collection of symbols, whether that observer is internal or external to the firm. Image is 'what comes to mind when one hears the name or sees the logo' of a particular firm (Gray and Balmer, 1998: 696). The transition from identity to image is a function of public relations, marketing and other organizational processes that attempt to shape the impression people have of the firm. But image can also be shaped but not controlled by an organization because of factors such as media coverage, governmental regulations and surveillance, industry dynamics and other external forces also influence impressions of the firm. This brings us to the following definition:

Corporate Reputation: Observers' Collective judgments of a corporation based on assessments of the financial, social, and environmental impacts attributed to the corporation over time [8].

2.5 Online Reputation Management

Although not viewed as such, PageRank may be thought of as a way of rating the "reputation" of web sites [9]. In that sense it was one of the first online reputation algorithms. ORM stands for Online Reputation Management which in fact is a complete way of dealing with online reputation. Obviously dealing with data from different sources structured or not, is not an easy task. There are plenty of tools, paid, free and open-source, presenting different opportunities to approach the problem. Our study is done with python programming language, dealing with data from Amazon, Twitter and Youtube. This method offers the most complexity and customization. Since is not a single button solution requires a minimum of programming knowledge and time.

3 Type of Company/Product

There are all kind of different products and services out there. Lately there are P2P (peer to peer) producing billions of dollars annually. It is also called the sharing Economy which is crowd-based capitalism.

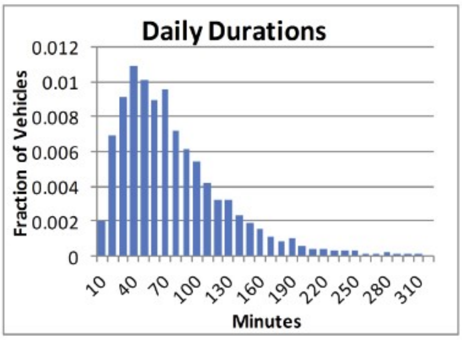

Fig 1: vehicle usage, USA. NHTS 2009

As it can be seen in Fig.1, the average American car is underutilized. This surplus of car time is responsible for the creation of Uber. Similarly Airbnb is a big property provider without owning any of the properties.

In the dawn of sharing economy Lisa Gansky in her book The Mesh, is proposing two dimensions that might evaluate a product before determining whether a peer-to-peer rental platform for it might emerge [10]. The value of the product (cost) and the time of use (how intensively is used by the owner, frequency). A product that is not used very intensively by an owner like a car is a good prospect for renting. On the other hand if the product is of minimal value the coordination costs would not made sense to be paid as the person (renter) could easily afford to buy a new one. This is a generally good idea. Following that logic renting out a luxury watch, a Rolex for instance makes a lot of sense. Since it cost a lot of money to buy a new one and it can often be used for only a few hours per week of month by their owners. As a matter of fact a company called ElevenJames was launched in 2013 in New York. Their premise was to borrow luxury watches to their clients based on a subscription program. Today this company strives to stay alive (https://www.bloomberg.com/news/articles/2018-08-23/**what-s-happening-to-subscription-watch-club-eleven-james**). It might have been the subscription model or bad management but we believe that there is an ownership effect to these products. In other words the important thing in wearing a Rolex watch is that you can afford it. It provides a certain prestige or credit to the owner. Similarly a wedding ring cannot be borrowed. Even though is a valuable item, most of it is value is attributed to a person buying for you. Therefore the type of the product or service that a company is offering plays a decisive role.

4 Experimental Results

We chose Rolex as our example company. Being a luxury product, wearing a Rolex watch is absolutely no necessity. Therefore is an imperative need for the company to provide an image of superiority. As a matter of fact Rolex would probably be in the first words that come to mind when we hear 'Luxury Watches'. A good indicator is that these watches keep their value in time. Also if James Bond wears it, it's probably something.

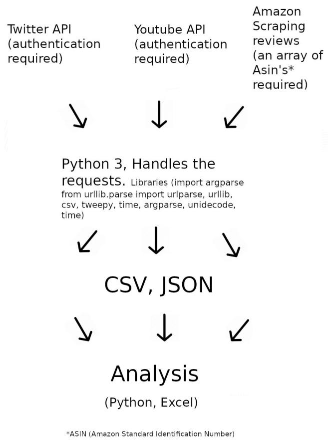

Fig 2: Schema

4.1 Amazon

Please mind that Amazon is not the best channel to explore since people usually would buy a Rolex from another vendor. Some of the products there might be even imitating a real Rolex. Nevertheless is an excellent way to scrape the reviews and make an analysis for a big variety of products. The script was inspired by on-line work [11].

We managed to scrape 767 reviews. We cleaned up the data frame, cutting off stop-words etc. We start by presenting a Histogram of Reviews' Lengths.

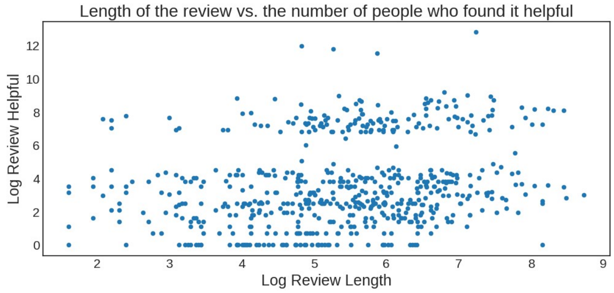

Fig 3: Length of Review vs People who found it helpful

The next one is a amazon review style histogram. It averages these 767 reviews we had. Once we tried some single product reviews that were some like popular, the histogram was matching exactly amazon's histogram for the product. Here we don't have one to compare one from amazon. A nice way to think about it is like a Unified Rolex Product in amazon's listings. Mind that there were a lot of listings without review that are not part of this average. This average is between the listings that had a review.

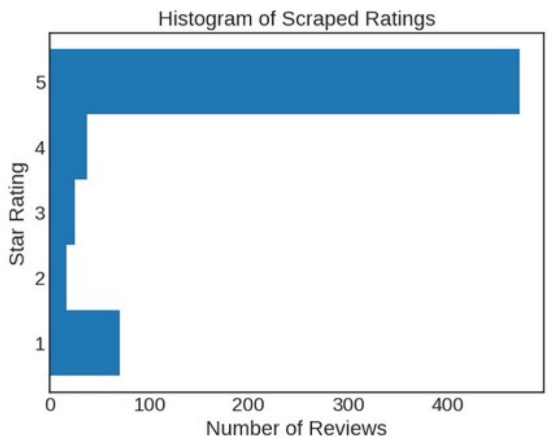

Fig 4: Average Rating

We extended our stop-words additionally appending 'Rolex', 'Rolex watches', 'watches', etc. We did not want to see them in our wordcloud and therefore we excluded them.

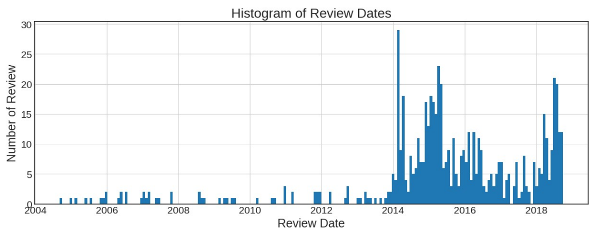

Fig 5: Review Date Histogram

Fig 6: Wordclouds, grouped by stars

4.2 Twitter

We obtained 3,200 tweets. That is regarding 07 and 08 of October 2018. We uploaded the data-frame in to the memory together with all the libraries needed. So let us move on to a time series analysis for each of the days. Keyword was “Rolex”.

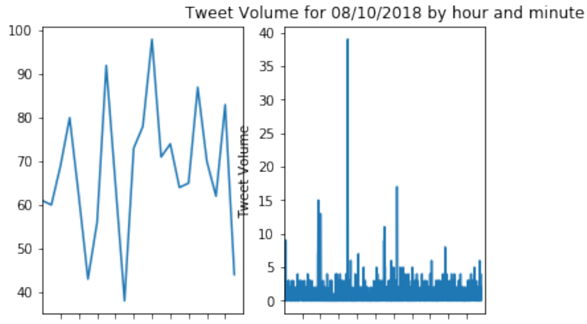

Fig 7: Tweet Volume for 8 Oct, 2018

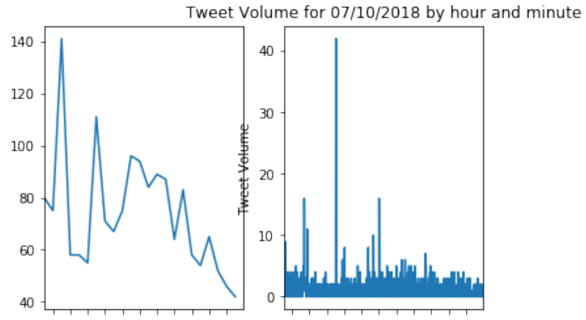

Fig 8: Tweet Volume for 7 Oct, 2018

There are a few minutes throughout the day that have around 40 tweets containing “Rolex”, in both days.

We also tried to locate the most influential users. So in Figure 9 a bar chart is presented containing the 25 most popular users (in terms of influence), in descending order.

Here it should be mentioned that all 3,200 tweets that were kept, were made in English. Other languages could also been used in future experiments.

There is another field that derives from our implementation in the tw_search() function (a search function to search for data posted on Twitter). It returns a Boolean value if the text of the tweet contains a link. Another important thing to point out is that the purpose of the post was to direct the bystanders in a landing page.

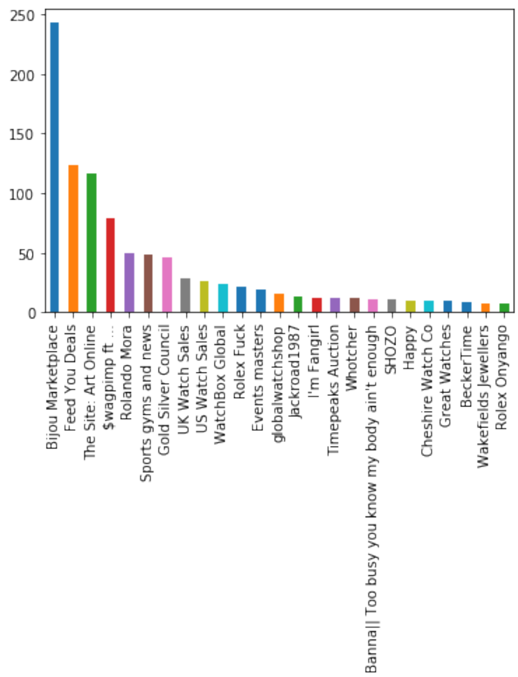

Fig 9: Top contributors

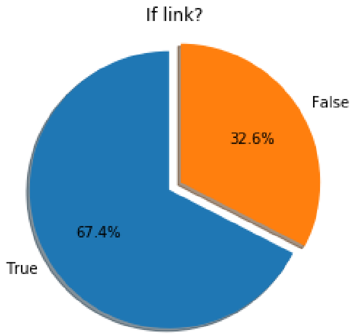

Fig 10: Link containment Pie-Chart

The sources of the users creating tweets with ‘Rolex’. Please mind that the part missing is other sources.

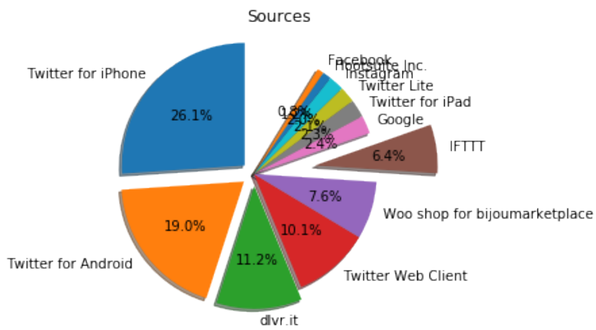

Fig 11: Tweet Sources

4.3 Twitter Sentiment Analysis

Subjectivity with polarity was estimated to be 0.384942 positively correlated.

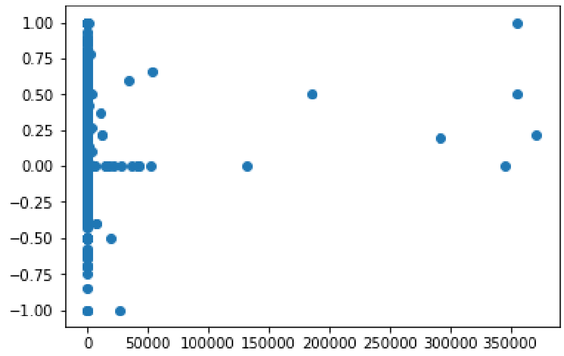

Fig 12: Polarity and number of re-tweets

Then the following pie is provided regarding the polarity: negative, neutral, positive.

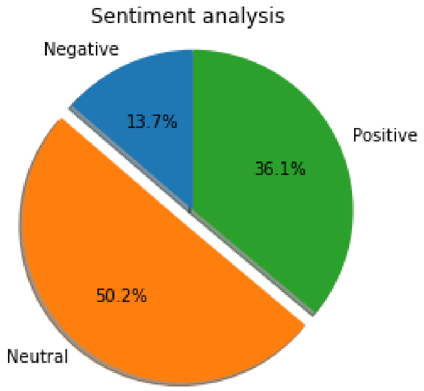

Fig 13: Neutral, Negative, Positive

4.4 Youtube

Here the sample is definitely poor. We only have information about 50 videos regarding Rolex. Unfortunately there are not many corners to cut when it comes to download data from Youtube API unless you own the data (a.k.a. originally uploaded the videos). Also the information that comes structured together with a video is poor in comparison to Amazon or Twitter. Nevertheless let us see a wordcloud made out of Youtube titles.

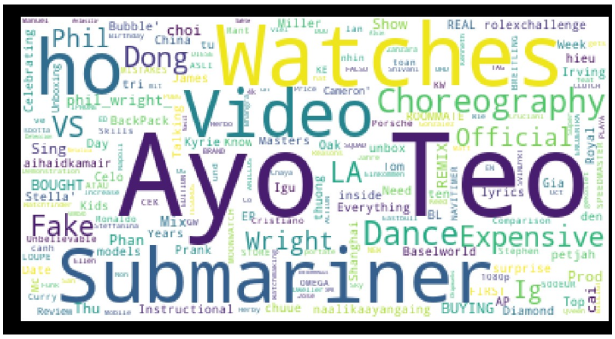

Fig 14: Youtube's WordClouds

We also made a linear regression model. Based on Ordinary Least Squares and by plotting the linear model together with the scatter plot (Like vs Dislike), we take Figure 15.

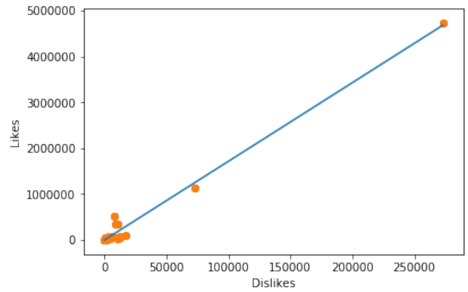

Fig 15: Likes vs Dislikes Linear Regression Model

As it can be observed for every 20 likes we also have 1 dislike.

5 Caveats and Disclaimers

5.1 CMC is Lacking

As already mentioned CMC cannot include all the possible information that a person broadcasts in physical space communication. Additionally it has been implied that by choosing a specific channel over another one can add to the communication. A table (Figure 16) has been proposed to identify the affordances of the media [12].

5.2 Sentiment Analysis

Sentiment Analysis is not a perfect method. It is a good way of dealing with a big amount of data, one that we could not read and conclude. The NLTK Natural Language Tool Kit we are using is based on **naive Bayes classifiers** that they are a family of simple "probabilistic classifiers" based on applying Bayes' theorem with strong (naive) independence assumptions between the features. Same technology has been used in spam filter in our mail boxes but even though they promise 99.5% accuracy we still receive spam e-mails. And even worst sometimes we miss mail that is not spam. It is a heuristic rather than an algorithm.

5.3 Structured Fraudulence

It is in a person's, who owns an on-line business, best interest to have positive reviews, recommendations and testimonials. Nobody more than personal ethics can prohibit to an individual to create an account and write a dithyrambic testimonial. Likewise someone can create a prestigious review site to outbound link to the original business. Here is a very recent example of it. There is a rehabilitation center in Malibu, USA called Cliffside Malibu (cliffsidemalibu.com). There is evidence that people who own the business also own two review websites for rehabilitation services. Unfortunately these results appear in Google's first page for the query 'Cliffside Malibu'. The first one is thefix.com and the second one is

Affordances of Media	*Face-to-face*	*Video conference*	*Telephone*	*Instant messaging or chat*	*E-mail*
(1) *Physical co-presence*: Participants share a physical environment, including a view of what each is doing and looking at.	++	??	−−	−−	−−
(2) *Visibility*: One participant sees another, but not necessarily what the other is doing or looking at.	++	+	−−	−−	−−
(3) *Audibility*: One participant can hear another.	++	++	++	−−	−−
(4) *Cotemporality*: Messages are received without delay (close to the time that they are produced and directed at addressees), permitting fine-grained interactivity.	++	??	++	??	−−
(5) *Simultaneity*: Participants can send and receive messages at the same time, allowing communication in parallel.	++	??	++	??	−−
(6) *Sequentiality*: Participants take turns in an orderly fashion in a single conversation at a time; one turn's relevance to another is signaled by adjacency.	++	++	++	−−	−−
(7) *Reviewability*: Messages do not fade over time.	−−	−−	−−	++	++
(8) *Revisability*: Messages can be revised before being sent.	−−	−−	−−	++	++

[a]Present in a particular medium: ++; present to a limited extent: +; present in some systems: ??; absent: −−. Physical co-presence (1), the hallmark of face-to-face communication, nearly always includes affordances (2) through (5).
Adapted from Clark & Brennan, 1991.

Fig 16: Affordances of Media

www.rehabreviews.com. Someone who is considering to use their services and leave there a loved one wants to make a double check on the facilities, the staff and the business reputation in general. You can easily understand how this is misleading, scummy and unfair.

5.4 Reviews and the Rating Game

In March 2013, Uber drivers started a protest. The reason was that they had been dropped from the platform because of low rating. It would be useful to separate the 'bad' or dangerous drivers but could that be just bad draw of customers? Josh Dziera in his 2015 article argued 'The proliferation of online feedback systems has simply turned us customers into really bad bosses'. In social science there are much research in the topics on information cascades, conformism and network effects. The citations could be endless, from Milgram to very recent work. Even phenomena like suicide and obesity can be transmitted through the network [13].

The "anchoring effect" names our tendency to be influenced by irrelevant numbers. Shown higher/lower numbers, experimental subjects gave higher/lower responses. As an example, most people, when asked whether Gandhi was more than 114 years old when he died, will provide a much larger estimate of his age at death than others who were asked whether Gandhi was more or less than 35 years old. Experiments show that our behavior is influenced, much more than we know or want, by the environment of the moment [14]. There is strong evidence that people have a propensity to be biased by the ratings they have already seen, and rate a highly rated restaurant (in Yelp) based simply on the fact that the establishment has a higher rating to begin with [15].

That kind of bias should be taken seriously. Could that be that an African-American man or a woman paid less? In the last work has also been found that a one-star increase in Yelp rating leads to a 5-9 percent increase in revenue. In TripAdvisor (shown below) you can actually choose which of the five stars review you wish to see.

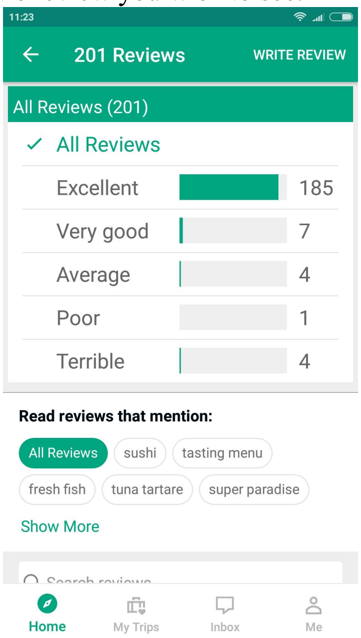

Fig 19: TripAdvisor reviews

For someone who is rating, if the rating is not obligatory, the incentives to submit an opinion could be either the user is not happy or the user is satisfied with the service. It can be argued that one of the two cases offers more incentive to be proactive. It can also be the case that people are kind with their reviews, even though they did not find the service satisfactory, because they do not want to affect negatively the revenue of the business. Lastly at has been stated as social loafing or the free ride problem [16]. People want to harness the power of reviews to define their choices but after they are done they are being lazy to contribute. We are aware of this since a long time and there are efforts to motivate people, usually with emails, but with other ways as well [17].

Nearly 95% of Airbnb properties boast an average user-generated rating of either 4.5 or 5 stars (the maximum); virtually none has less than a 3.5 star rating. With a juxtaposition with the ratings of approximately half a million hotels worldwide that have been collected on TripAdvisor, where there is a much lower average rating of 3.8 stars, and more variance across reviews . Moreover, there is only weak correlation in the ratings of individual cross-listed properties across the two platforms. It could be the case that TripAdvisor users prefer higher-priced accommodations, while Airbnb users are more price-conscious. Yet, when comparing properties within each price segment, their relative preferences are the same [18].

For the immunization of online reputation reporting systems against unfair ratings and discriminatory behavior two mechanisms have been proposed: controlled anonymity and cluster filtering [19].

5.5 API Limitations

In Twitter we were allowed to pull 180 tweets per 15 min, up to 3,200 tweets, for a keyword (aka “Rolex”), which is exactly the amount of tweets we collected. Youtube allows again only authorized users, 50 videos per keyword. For videos that are publicly available and do not belong to the user. For the case that the account requesting with the account of the uploader is the same then different limits apply. For amazon we did not use the API they are providing. Instead we scrape their product review webpages. We used a workaround to find the ASIN we were interested for, gave the list to the script to scrape the webpages of the corresponding ASIN’s. Amazon’s server however has a limit of request that a user is able to do. So we set a small delay for that purpose. Amazon’s server like any other server likes to know from what browser and device we are accessing the server. We have the power to overwrite this information, we declared we are : *'User-Agent': 'Mozilla/5.0 (X11; Ubuntu; Linux x86_64; rv:62.0) Gecko/20100101 Firefox/62.0'}.*

5.6 Segmentation and Competitive Analysis

We are not using segmentation methods in our analysis. Segmentation is very useful in strategizing. Knowing how products are doing to specific crowds is very important. However we are not dealing with future strategies of the company and therefore we have not include segmentation in our paper. We might want to group the information with filters like age, race, religion, gender, family size, ethnicity, income, and education. That would be only demographic segmentation; it could also be behavioral, psychographic or geographic. Competitive analysis is a crucial part in creating a strategy. Since we are not doing this analysis for the company we skip that but we still feel it is important enough to mention.

5.7 Big Tech Algorithmic non-Transparency

Google is responsible for delivering results that bring people closer to the information they are looking for. As a result having a website that is SEO (Search Engine Optimization/optimized) is imperative. Though they are actions and tricks that we are absolutely should not do (black hat SEO), on the other hand there are not actions we should do but rather good practices. Google does not to release publicly these parameters that can bring you in the top of the page. That is because: 1) they make money selling keywords 2) they don’t want their competition to produce similar results 3) they do not want people to create content for the search engine robot but rather content for human users. That personalized result page can restrict us in a bubble. [20]

6 Conclusion

As soon as a person is aware for a brand or a product the very same moment starts to develop an opinion about it. Reputation is not addressed only to the end consumer but everybody. Therefore we are treating reputation as multidimensional.

It is an imperative need to use a interdisciplinary approach. We explored a lot of restrictions and limitations, either due to the human nature and the physical world but also because of the IoT. In the dawn of the Internet, we were excited about this new democratic medium that could change the world. It’s true that the end-to-end principal that the TCP/IP is made allows a lot of freedom and functionality. Another opportunity was arising, one that it did not exist in Television. That was the click-stream analysis. Everything we do in our computers can be monitored and analyzed. This is great because from the standpoint of the company, we want to adapt to our users habits, not the other way around. For instance we set our social media to publish in the popular hours and not try to attract users the hours that are convenient for us. Big part of the traditional marketing is turning to customer insights for that very reason.

Academic institutions that carry out research with human subjects fall into a lot of restrictions and have to be approved by a special committee. On the other hand big corporations run experiments to us all the time and the results are not even published.

A new opportunity for on-line marketing is now available. A rich getting richer phenomenon has been proposed that works on-line. The problem is that is falling in to the hand of a few. Facebook, Google, Linkedin and Instagram are responsible for a big part of it.

This field is new; it is not yet an applied science. We are using techniques we hope they will work and we cannot know for sure since the code that promotes our content is not an open source. Not only in paid adverting but in organic results too. Needless to say both can shape reputation.

With good practices, updated laws and regulations, NGOs and academic institutions, we expect they can change the future on-line, protecting us from the giant Internet companies but the human nature as well.

References:

[1] Goffman, E. (1956). The presentation of self in everyday life. Edinburgh, University of Edinburgh.

[2] Ackerman, M. (2000). The intellectual challenge of computer-supported cooperative work: the gap between social requirements and technical feasibility. *Human-Computer Interaction* 15: 179-203.

[3] Walther, J.B. (2007). Selective self-presentation in computer-mediated communication: Hyperpersonal dimensions of technology, language and cognition. Computers in Human Behavior 23: 2538-2557

[4] Hancock, J.T. (2007). Digital deception: Why, when and how people lie online. In: Joinson, A., McKenna, K., Postmes, T. and Reips, U.-D. (eds.) The Oxford Handbook of Internet Psychology. Oxford: Oxford University Press.

[5] Lawrence Lessig, (1999). Code 2.0

[6] Easley D. And Kleinberg J. (2010) Networks, crowds and markets: Reasoning about a highly connected world.

[7] Melewar, T.C. and Jenkins, E. (2002) 'Defining the corporate identity construct', Corporate Reputation Review, 5(1), 76-90

[8] Barnett, Michael L. and Jermier, John and Lafferty, Barbara A., Corporate Reputation: The Definitional Landscape. Corporate Reputation Review, Vol. 9, No. 1, 2006.

[9] Zhang H., Goel A., Govindan R., Mason K., Van Roy B. (2004) Making Eigenvector-Based Reputation Systems Robust to Collusion. In: Leonardi S. (eds) Algorithms and Models for the Web-Graph. WAW 2004. Lecture Notes in Computer Science, vol 3243. Springer, Berlin, Heidelberg

[10] Lisa Gansky 2010. The Mesh: Why the Future of Business Is Sharing

[11] https://www.knoyd.com/blog/amazon-review-scraper

[12] Brennan, S. E., and Lockridge, C.B. (2006). Computer-mediated communication: A cognitive science approach."*ELL2, Encyclopedia of Language and Linguistics*: 775-780

[13] Christakis, N. A., & Fowler, J. H. (2009). *Connected: The surprising power of our social networks* and how they shape our lives.

[14] Kahneman, Daniel. (2011) Thinking, Fast and Slow. New York: Farrar, Straus and Giroux.

[15] M Luca (2011) - Reviews , Reputation, and Revenue: The Case of Yelp .com

[16] Hardin, R. (2003) The Free Rider Problem. In E.N. Zalta (Ed.),*The Stanford Encyclopedia of Philosophy* (Summer 2003 Edition)

[17] Ling, K., Beenen, G., Ludford, P., Wang, X., Chang, K. Li, X., Cosley, D., Grankowski, D., Terveen, L. Rashid, A.M., Resnick, P. and Kraut, R. (2005). Using social psychology to motivate contributions to online communities. Journal of Computer-Mediated Communication 10(4).

[18] Zervas, Georgios and Proserpio, Davide and Byers, John, (2015) A First Look at Online Reputation on Airbnb, Where Every Stay is Above Average.

[19] C Dellarocas (2000) - Immunizing online reputation reporting systems against unfair ratings and discriminatory behavior

[20] Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. London: Viking/Penguin Press.