

Social big data: A Twitter text mining approach to the communication of universities during the Lebanese protests

RAYA Katia
CELSA - ESIB
Sorbonne University – Saint Joseph
University of Beirut
Beirut, Lebanon
katia.raya@usj.edu.lb

D'ALMEIDA Nicole
CELSA
Sorbonne University
Paris, France
nicole.d_almeida@sorbonne-
universite.fr

CHAMOUN Maroun
ESIB
Saint Joseph University of Beirut
Beirut, Lebanon
Maroun Chamoun
maroun.chamoun@usj.edu.lb

Abstract— Since October 17, 2019, Lebanon has experienced unprecedented popular protests, demanding the departure of the entire political class, accused of being gangrened by corruption. Country paralyzed, institutions closed for more than two weeks, the eyes are turned to universities that have closed their doors but whose community (teachers and students) actively participate in the national jump. This study explores the use of social media by universities in Lebanon during the national revolution using social big data technology on Twitter in comparison to the national usage of twitter. Important information was collected, analyzed and visualized using the R language.

Keywords— *Social Media Mining, R language, Higher Education, Text mining, Visualization*

I. INTRODUCTION

The role and the importance of social media platforms rose dramatically the last few years within the realm of business, political and public affairs, as well as individuals' everyday lives [10]. Their growing popularity raises questions about their role in the deliberative and participatory democracies [3].

Across the world, demonstrators are using social media platforms to organize and spread their messages. These revolts are usually leaderless like in France, Lebanon and Chile, and inspired by hashtags and slogans. These platforms are playing an important role in the protests worldwide especially when the mainstream media is absent, manipulated or restrained. It has become the norm that connected eyewitnesses are the first to report and share what is happening on social media. Voluntary citizens reports are often used by official news agency as a verified information. Users are not only able to get news and information, but they also are able to express themselves openly and freely and to cover the real scenes to the word. Informing users of the ever-shifting security situation, reporting violence and human rights violations, requesting reinforcement and showing the different angles of the picture are the main practical uses of social media platforms. Furthermore, social media play an increasingly important role in shaping political communication in the US and around the world (e.g., [20];[22]). In Lebanon, politicians and parties embrace Twitter actively. The number one Lebanese politician with the biggest number of followers is the Prime Minister Saad Hariri, on Twitter since 5 Oct 2009 with 1.7M followers on Dec. 2019, followed by the parliament and leader of Druze community, Walid Jumblatt, on Twitter since 26 Oct 2014, with 882.9k followers on December 2019.

On the night of the 17th of October protests erupted in Lebanon, triggered by government plan for the 2020 budget,

to place a new \$6-a-month tax on Voice over Internet Protocols (VoIP) for those using WhatsApp calls or other VoIP apps. This act came on top of a deteriorating economic situation and several embarrassing failures by the government, including its incompetence in putting out fires that scorched the mountains through the country few days before the announcement and the panic buying at petroleum stations and fears over the banking systems. Across the country, the protests gained a lot of popularity and became larger. People come out to demand a new secular and corruption-free order and took massively to social media to express their opinions.

Long time ago, universities are no longer simply dedicated teaching and research institutions; they are now more involved in society than before and seen as the engines of the knowledge economy [19]. The Lebanese ex-minister of Education and the Arts, Charles Malek [12], qualify universities as having a dominant influence in the world in his book “A Christian Critique of the University” and this influence is much bigger than that of the church, the government and all other institutions. Since 1972, universities are adopting a marketing approach whose initiator is Krachenberg [7]. Their usage of social media has gradually become institutionalized, resulting in practices that have also progressively evolved [17]. In Lebanon, all universities have official account on Social networks like Facebook, Twitter and the others. Their involvement in political and social life is old and recurrent in several circumstances.

The objective of this study is to explore how universities in Lebanon used Twitter from the beginning of the protests on 17 October till the Independence Day on 22 November while the expectations from them are high and the usage of twitter by citizens is massive.

The rest of the paper is organized as follows: Section II presents brief overview about social media mining, introducing few studies that shed light on the importance of Twitter analysis to better understand public opinion and the ethical issues related to social big data mining; Section III discusses our methodology to collect, analyze and visualize data. The experimental results are presented in section IV; In section V, we analyze and discuss our findings and we conclude the paper in section VI.

II. LITTERATURE REVIEW

A. Social Media Mining overview

Social media Mining has become increasingly a source of information for decision making in the context of marketing

and business (e.g., [4]; [8]) but also for politics, crisis and government (e.g., [15]; [20]). It is not surprising to find out that a number of studies were conducted in many fields including computer science, business, linguistics, data science and computational studies to show the potential of big data analysis to solve problems [21], more precisely, the big data collected from social media, especially twitter platform [1]; [22].

These studies allowed scholars to analyze vast textual corpora and networks on a vast scale with very limited manual intervention. For example, Tweets were collected lately, to analyze public sentiments about Syrian refugee crisis and compare sentiments in Turkish and English languages [14]. The study found out that sentiments were significantly different between these two languages with slightly more positive sentiments towards Syrians and refugees in Turkish. A content-analysis study was conducted of over 100,000 messages in the context of the German federal election [22] showed that Twitter is used extensively for political deliberation and reflects the election result. Another text mining and sentiment analysis on collected tweets during a storm in Lebanon, revealed the value of social media analysis enhancing the capability to identify weak signals, threats and needs, and to manage emergencies and safety [16]. Some studies have focused on information accuracy of crisis reports contributed by online people [13] and the ways in which social media can be used to spread rumor, misinformation, gossip and, in extreme cases, propaganda [9].

There are two major strategies for social media data mining: one is linkage-based or structure-based, very used in Social Network Analysis to study communities and links between nodes (actors, persons), and the other is content-based. The data instance collected from twitter is times-series oriented [11]. Analyzing the data pass by a process that consists of identifying, extracting, cleaning, analyzing, visualizing and interpreting data [6]. In his book about "Social Big data mining", Hiroshi Ishikawa[5], explains the bidirectional interaction that exists between data on social media, also called "social data (i.e., social media data such as Twitter articles and Flickr photos)", and "physical real-world data (i.e., heterogeneous data such as science data, event data, and transportation data)". He suggests analyzing social media data and physical world data by relating them to each other, in order to reveal mutual interdependencies that in turn would produce bigger values.

B. Ethical issues in social media mining

The debate generates many concerns and controversies about Social media mining [2]. It is becoming very attractive and benefic for business, scholars and politics but to make sure that this technique is used properly, we need to focus on the possible objections to it. Dangers lie in the different ways in which privacy is threatened. According to Van Wel & Royakkers [23] "*the privacy can be violated when information concerning an individual is obtained, used, or disseminated, especially if this occurs without their knowledge or consent.*" Effectively, in addition to technical data on phones, clothes or food, social media mining involves personal data. Around the world, the data is collected and resolved to get information about consumers and voters [18]. It could come to manipulating behaviors by addressing highly targeted messages according to humans' preferences and psychology. Another problem of data mining

is that, even if the information is taken from public pages, they are taking it out of context and people are judged and treated as group members rather than individuals [23].

Since the introduction of the DPD (Data Protection Directive) in 1995, EU is working on public awareness and concern over data protection. DPD regulation aims to protect personal information from being misused or unnecessarily collected. After 20 years, the world has changed dramatically, and new data protection laws were needed, especially after the Cambridge Analytica's scandal - exploiting Facebook user data during the 2016 US presidential election. In early 1996, the EU adopted the General Data Protection Regulation (GDPR) to get more control over EU citizen's personal data that applies automatically to all the 28-member states. This regulation sets a new global baseline for privacy that require businesses to prioritize the protection and privacy of personal data belonging to European Union (EU) residents. GDPR protects individual's privacy and gives them ability to delete negative information about themselves.

Social media mining has obviously many benefits when ethical considerations are considered as an integrated and essential part of the process. It's the responsibility of web miners and users at the same time who should discern what to publish and what to keep private. In our study, we are fully aware of this legal and ethical issue, that's why we tried to use data responsibly and respect user privacy.

III. RESEARCH METHODOLOGY

Before Our analysis draws on quantitative data from Twitter official accounts of six universities in Lebanon that appears in QS ranking results in 2019. We wanted to add the Lebanese University, since it is the one and only university of the government, but we found out that they didn't post any tweet since May 2019, (5 months before the protests).

To collect, analyze and visualize data, we used the free and open source statistical programming language R for its excellent capabilities in statistics, machine learning, and graphics and its vast set of libraries categorized by discipline and developed by a large and active community. Tweets were harvested from Twitter using an in-house script created for the purpose using Rstudio - the integrated development environment (IDE) for R. Two types of tweets were collected: the first set was taken from universities official pages timeline and the second one is a set of tweets that includes the constantly trending hashtag #لبنان_ينتفض (Lebanon_Protests), not related to universities. In addition to the base packages, we used specified packages such as ArabicStemR for reading Arabic words and ggplot2 for visualizing data.

We downloaded tweets (and their meta-data) for the period 17 October to 22 November 2019. 331 tweets were founded on universities official timeline during these 36 days, and around 250000 tweets containing the hashtag mentioned above. The data was stored in two separate tables of a database: one for universities and one for global tweets. Since Twitter allow only 350 requests per hour for registered users, and since Twitter only makes the last 6-day tweets available to the public, we collected our tweets day by day during the period of our study. In order to protect user's identity, we removed from our data every information about tweets' owners.

a protester hugging tight a soldier while in tears were circulating massively on Twitter and other social media platforms.

- **Protesting activities:** Most trending hashtags show the main action taken by the demonstrators during the first month as: #سبت_الساحات (Saturday of squares), #بوسطه_الثوره (The revolution bus) that traveled from north to south to unite protesters, #احد_الاصرار (Sunday of determination), #قرع_الطناجر (cooker percussion), #ثوره_الطلاب (students revolution), #احد_الوحده (Sunday of union), #السلسله_البشريه the “Lebanese Human Chain” formed across the country along coastal roads from north to south on 171 kilometer to manifest national unity.

Many trending hashtags were about closing roads like #انتين_السيارات (Monday of cars), aiming to close highways with cars while other trending hashtags were against taking this actions by protesters: #حقك_تتظاهر_والطريق_مفتوح (it’s your right to protest while the road is open), and #لا_لقطع_الطرق (no for closing the roads).

- **Victims, journalists and cities:** The tragic death of the protester #علاء_بو_فخر Alaa Bou Fakher, shocked the country and the hashtag of his name flooded the social media after he has been shot in front of his wife and kids. In addition to names of reporters from different TV channels being subject to verbal attacks and insults during the coverage of the popular demonstrations, especially #جويس_عقيقي (Joyce Akiki) insulted by the musical composer Samir Sfeir outside Baabda justice palace and #ديما_الواطيه (the whore Dima) massively insulted by Hezbollah followers.

The main cities of the protests were often trending like #طرابلس_عروس_الثوره (Tripoli, the bride of the revolution), the main roads blocked by protesters: #جل_الديب (Jal al-Dib), #ذوق_مصيح (Zouk Mosbeh), #جسر_الرينغ (Ring bridge) and #نهر_الكلب (Naher El Kaleb), in addition to the presidential city #بعدا (Baabda).

B. Twitter usage by Lebanese universities during the protests

The number of tweets is slightly different between universities and somewhat low during the first 36 first days of the protests. Note that the first day, the number of Tweets by AUB is high. The reason is that posts was not related to the demonstrations that began at night.

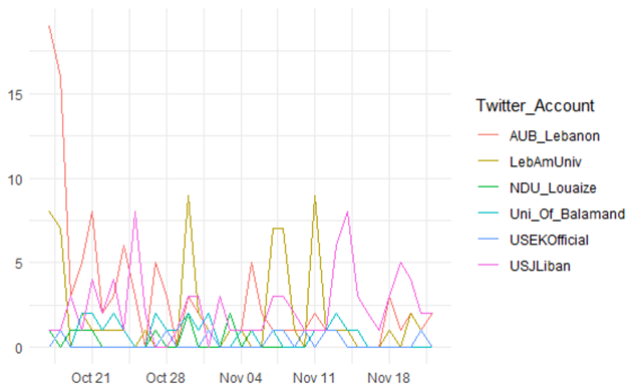


Fig. 5. Frequency of Tweets aggregated by day

A large number of tweets were in Arabic while in normal days, universities prefer English or French depending on their language speaking and teaching.

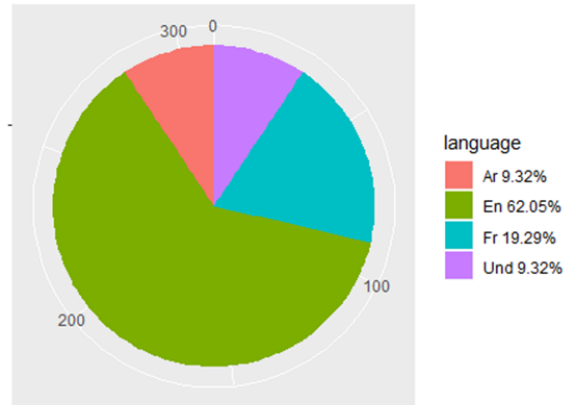


Fig. 6. Languages used in Tweets

We divided these tweets into three categories: organic (posted by universities), replies (answering questions or demands) and retweets (share information from other pages and especially from president’s personal accounts). We found that the number of replies and retweets is important comparing to the organic tweets (around half of total tweets). Which means that many persons were commenting and reacting.

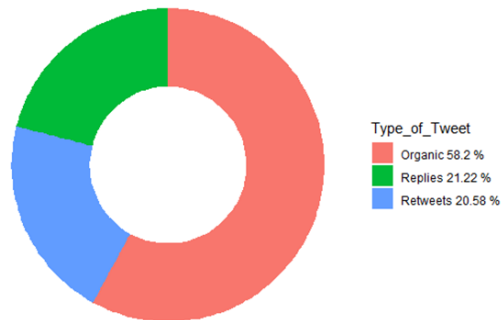


Fig. 7. Type of tweets

We noticed that the word Lebanon in Arabic لبنان is the most used hashtag by universities. Many of the hashtags in graph 8 are related to the situation like the words: democracy, Lebanon uprising, and Lebanese protests in Arabic.



Fig. 8. Word cloud of the most used hashtags

In order to better understand this content, we gave a type for each tweet. We found out that tweets were either: replies on comments, academic events related to the situation (e.g. a cycle of conferences held by a group of teachers from Saint Joseph University of Beirut in a tent on martyr’s square and

sharing knowledge about political, cultural, legal and psychological questions), prayers (one of these universities, shared only two posts about the situation, in the form of prayers for peace and good health for all), messages from the president, posts related to the Independence Day (e.g. coverage of celebrations on campus or messages of happy holidays), Events related to the protests (e.g. providing 1st aid services and sharing safety instructions), declarations to announce closing, resuming classes or taking measures to help students complete the academic semester and announcement or coverage of normal events like conferences and finally academic activities.

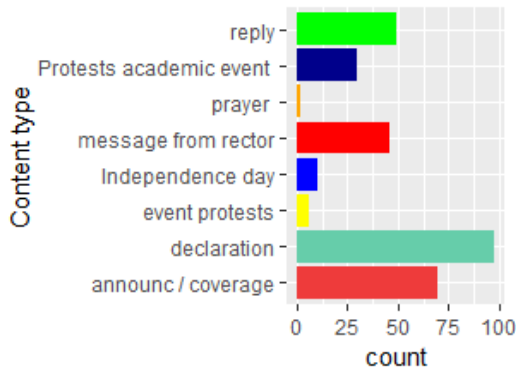


Fig. 9. Count of tweets by content type

Using the mean and the boxplot, we notice that tweets in accordance with the situation in the country like events related to protests, messages from presidents, and announcements related to the Independence Day got the biggest number of engagements (favorites and retweets) than announcements and events related to normal academic year. Messages from presidents were the most retweeted posts.

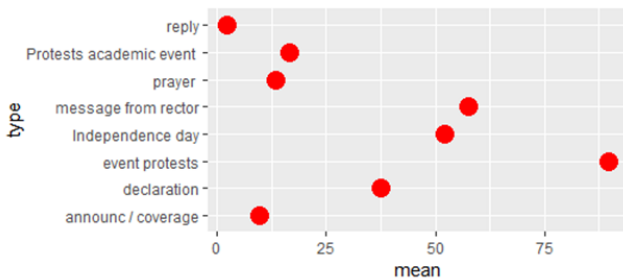


Fig. 10. Favorite content by type (point graph)

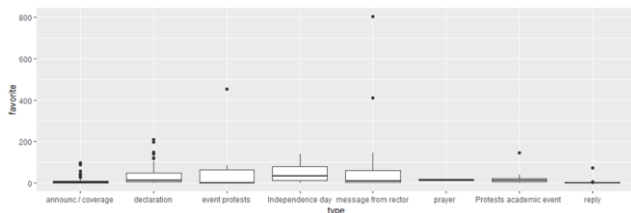


Fig. 11. Favorite content by type (boxplot)

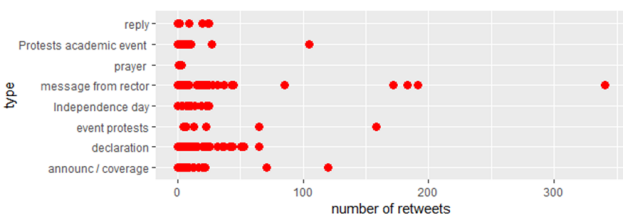


Fig. 12. Number of retweets by type

V. DISCUSSION

University tweets didn't match with trending subjects and hashtags on Twitter. They did not bounce back on the most trending questions, although the academic activities under the tents did so, as in the case of the conferences organized by the teachers of the Saint-Joseph University of Beirut. However, with the exception of the Lebanese university which has stopped its activity on twitter since May 17 and one university whose activity was limited to praying for the security of the country and highlighting the civility, courtesy and responsibility of their students protesting inside their campus, Universities in Lebanon had a clear position supporting the protests, especially through the messages of their respective presidents that collected the highest number of engagement. Some of them were more present than others. On twitter, we collected 24 president messages from five universities during the first month of the demonstrations, three TV interviews with American University of Beirut (AUB) and Saint-Joseph University of Beirut (USJ) presidents on CNN, Sky News Arabia and Annahar TV and one article in THE_Atlantic.

Graphs in the previous section shows that universities were mainly tweeting to inform their communities about administrative decisions. Despite the low number of posts showing a social and/or political position related to the demonstrations, the engagement was the highest by the number of favorites, retweets and comments, which explains the high number of replies by community managers. It's important to emphasize that universities position was at a general level, so they didn't mention any trending topic related to the daily events of the protests except the most trending hashtag for that period [#لبنان_ينتفض](#) (Lebanon_Protests).

The biggest and oldest two universities AUB and USJ issued an unprecedented joint statement. In five points, the two presidents express their encouragement and their support to the demonstrations against all kinds of suppress to the freedom of speaking. They also charge the army to protect the protesters and the authorities to respond to people's "aspirational hopes" and to find solutions to the crisis.

These presidents take the floor to express their pride and to encourage their communities to continue their right movements. "I am in awe of your passionate commitment to our country. I admire your noble aspirations for a better Lebanon. I am proud of your dreams ... I applaud your determination to bring about peacefully much needed change... I salute you all for courageously demanding respect for the rights of all Lebanese to live honorably in their country, ..." (cf. message of Lebanese American University - LAU president on Twitter, 26 October). They also made announcement of measures taken like suspending classes and administrative activities, postponing deadlines, creating possibilities for remote teaching, assisting students unable to attend their classes, etc. (cf. message of University Of Balamand - UOB president on Twitter, 5 and 7 Nov.).

Presidents explain their supporting position as part of their values and mission. "The American University of Beirut has done everything possible to meet the high expectations of our community in upholding our values and fulfilling our mission during these best and worst of times" (Khuri, AUB). AUB's president activities were highly criticized on social media which prompted them to issue a statement declaring

that “AUB does not have a political stance in this turbulent time”.

VI. CONCLUSIONS AND FUTURE RESEARCH

Our analysis is unique in that it is the first documentation of the Lebanese Oct 17th protests and the first study of Lebanese universities communication on social media during a national crisis. Our data collected from Twitter is an important source of information for further studies and our study complemented assessment tools to execute social big data mining.

Our study has some limitations. We examined the official universities accounts while unofficial accounts were very engaged. We can mention for example the public Facebook group “تجمع أساتذة من USJ” and twitter account “@ProfsUSJ” created by the USJ professors supporting the protests and the Lebanese University Students Association tweeter account “@cP7xmLQKi0Gc4OZ” and Facebook page “Lustudents1951”. Furthermore, many hashtags related to universities went viral and were not used by official accounts like the one created by the union of university teachers, #جامعات_لبنان_تنتفض (# (Lebanese universities are rising up) and the hashtag created by the Lebanese University Students Association #Aal_Jamaa_Mesh_Rayheen (#WeAreNotGoingToTheUniversity) to reject the call of the president of the Lebanese University to resume studies on October 22.

Finally, our study opens multitude of avenues for research on this topic. Most importantly, we suggest that future studies be conducted to verify the source of each tweet and if there are any “shill” accounts that tweeted to support or oppose the protests and certain parties. According to a study made by Euronews on accounts tweeting both pro- and anti-government hashtags, an automated behavior was found on accounts tweeting to support one Lebanese party related to government. Furthermore, from the first day, social media has been a vital tool for rumors, fake news and propaganda. The number of these fake news going from fabricated images, videos and voice notes is very important. Lebanese army, banks and internet companies released many times, statements to deny and give clarifications according this fake news on their official websites and twitter accounts. This phenomenon deserves to be studied in depth.

REFERENCES

- [1] Boyadjian, J. (2019). *Analyser les opinions politiques sur internet Enjeux théoriques et défis méthodologiques*, Paris : Dalloz, Nouvelle Bibliothèque de Thèses, 382 p.
- [2] D’Almeida, N. (2011). *L’opinion publique*, CNRS
- [3] Gil de Zúñiga, H., & Valenzuela, S. (2011). The mediating path to a stronger citizenship: Online and offline networks, weak ties, and civic engagement. *Communication Research*, 38(3), 397-421.
- [4] Gruhl, D., Nagarajan, M., Pieper, J., Robson, C., & Sheth, A. (2010). Multimodal social intelligence in a real-time dashboard system. *The VLDB Journal—The International Journal on Very Large Data Bases*, 19(6), 825-848.
- [5] Ishikawa, H. (2015). *Social big data mining*. CRC Press.
- [6] Khan, G. F. (2015). *Seven Layers of Social Media Analytics: Mining Business Insights from Social Media Text, Actions, Networks, Hyperlinks, Apps, Search Engines, and Location Data*. Gohar Feroz Khan.
- [7] Krachenberg, A. R. (1972). Bringing the concept of marketing to higher education. *Journal of Higher Education*, 43(5): 369-380.
- [8] Larson, K., & Watson, R. (2011). The value of social media: toward measuring social media strategies, in international conference on information systems (ICIS), Shanghai, China.
- [9] Leberecht, T. (2010). Twitter grows up in aftermath of Haiti earthquake. *CNET News*, 19.
- [10] Lenhart, A. (2009). *Adults and social network websites*. Washington, DC: Pew Internet & American Life Project.
- [11] Makhabel, B., Mishra, P., Danneman, N., & Heimann, R. (2017). *R: Mining Spatial, Text, Web, and Social Media Data*. Birmingham, Packt Publishing Ltd.
- [12] Malik, C. H. (1982). *A Christian critique of the university*. Downers Grove, Illinois: InterVarsity Press.
- [13] Oh, O., Agrawal, M., & Rao, H. R. (2013). Community intelligence and social media services: A rumor theoretic analysis of tweets during social crises. *Mis Quarterly*, 407-426.
- [14] Öztürk, N., & Ayvaz, S. (2018). Sentiment analysis on Twitter: A text mining approach to the Syrian refugee crisis. *Telematics and Informatics*, 35(1), 136-147.
- [15] Paris, C., & Wan, S. (2011, May). Listening to the community: social media monitoring tasks for improving government services. In *CHI’11 Extended Abstracts on Human Factors in Computing Systems* (pp. 2095-2100). ACM.
- [16] Raya, K., Chamoun, M., & D’Almeida, N. (2019). The role of social media analytics in defence and security. In *International Conference AISD 2019*.
- [17] Roginsky, S., & Perrier, V. J. (2014). La fabrique de la communication des parlementaires européens. *Politiques de communication*, (2), 85-124.
- [18] Russell, M. A., & Klassen, M. (2018). *Mining the social web: data mining Facebook, Twitter, LinkedIn, Instagram, GitHub, and more*. O’Reilly Media.
- [19] Stensaker, B. (2007). Quality as fashion: Exploring the translation of a management idea into higher education. In *Quality assurance in higher education* (pp. 99-118). Springer, Dordrecht.
- [20] Stieglitz, S., & Dang-Xuan, L. (2013). Social media and political communication: a social media analytics framework. *Social network analysis and mining*, 3(4), 1277-1291.
- [21] Tapscoff, D., & Williams, A. D. (2006). *Wikinomics: How Mass Collaboration Changes Everything*, 2006. New York: Portfolio.
- [22] Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010, May). Predicting elections with twitter: What 140 characters reveal about political sentiment. In *Fourth international AAAI conference on weblogs and social media*.
- [23] Van Wel, L., & Royakkers, L. (2004). Ethical issues in web data mining. *Ethics and Information Technology*, 6(2), 129-140.