THE IMPACT OF THE 2015 REFUGEE CRISIS ON THE INTERNATIONAL PLACE BRAND OF HUNGARY

A Thesis

presented to

the Faculty of the Graduate School

at the University of Missouri-Columbia

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

---

by

ANNA MAIKOVA

Dr. Harsh Taneja, Thesis Committee Chair

MAY 2016

The undersigned, appointed by the dean of the Graduate School, have examined the thesis entitled

THE IMPACT OF THE 2015 REFUGEE CRISIS ON THE INTERNATIONAL PLACE BRAND OF HUNGARY

presented by Anna Maikova,

a candidate for the degree of Master of Arts,

and hereby certify that, in their opinion, it is worthy of acceptance

_____

Professor Harsh Taneja

_____

Professor Glen Cameron

_____

Professor Mitchell McKinney

_____

Professor Bradley Curs

ACKNOWLEDGEMENTS

I would like to thank my adviser Harsh Taneja and Andrey Simonov together with Pradeep Chintagunta who kindly provided the dataset.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

The Impact of the 2015 Refugee Crisis on the

International Place Brand of Hungary

Anna Maikova

Dr. Harsh Taneja, Thesis Committee Chair

Abstract

This quantitative study examines how the 2015 European refugee crisis events affected the international place brand of Hungary on Twitter in both short and long terms. The study supports the application of quantitative methods and dictionary-based sentiment analysis of tweets to the discipline of place branding. There is a significant increase in the amount of negative sentiment in tweets during the crisis (July-October, 2015) due to a high number of tweets about the refugee crisis. However, this effect has not persisted after November 2015. To conduct the sentiment analysis, we apply the lexicon-based polarity dictionary SentiStrength; to divide tweets into specific topics, we use Latent Dirichlet Allocation (Blei et al, 2003). The tweets that are likely published by media organizations are excluded from the analysis.

There is no significant increase in the amount of negative sentiment in tweets after the crisis, which suggests no persistent effect of the crisis on the place brand of Hungary in the long term because that negative sentiment about the crisis comes only from the tweets about the refugee crisis. The contribution of this study is the establishment of a research framework for the social media analysis of place brands in a crisis as well as in forming a solid basis for the application of this framework to studying other place brands in a crisis.

THE IMPACT OF THE 2015 REFUGEE CRISIS ON THE

INTERNATIONAL PLACE BRAND OF HUNGARY

Anna Maikova

Keywords: Dictionary-based Sentiment Analysis, Place Branding, Twitter, LDA, Crisis

INTRODUCTION

The United Nations Refugee Agency reported that the world is currently experiencing the highest inflow of forcibly displaced people since the World War II also referred to in major news media as 'the European refugee crisis' (UNHCR, 2014). Although non-EU countries (Turkey, Lebanon, and Jordan) host more refugees than any other countries combined, it is the European Union (EU) that primarily attracted international media attention in 2015 (UNHCR, 2015; Fry, 2015).

The EU border control reports that roughly half million refugees and migrants entered the EU since the beginning of 2015. The refugee inflow intensified in the spring of 2015 and spiked in August 2015 when about 160,000 migrants crossed the EU borders (Fry, 2015). Hungary, Romania, Czech Republic, and Slovakia closed their borders in order to stop the flow of refugees. One of the biggest migration streams ended up going through Hungary due to geographic reasons, and Hungary answered in a reactive way that resulted in a large media outbreak. This decision to close the borders for refugees became a controversial humanitarian issue for Hungary. The country faced a reputation 'exam' that motivated this research in the area of public diplomacy, place branding, and crisis communication. By the events of the 2015 European refugee crisis, we consider the increased inflow of refugees in the summer 2015 and the accumulative response of Hungary to it (building a fence with Serbia, adverse treatment of refugees by Hungarian police and reporters, and other events that received an intensive coverage by international media).

The research framework is the theories of place branding and crisis communications on social media. The place brand of Hungary is compared with the place brand of a similar country, Romania (control group), which did not experience the intensive international media coverage at the same time. However, a major protest happened in Romania after a fire broke out in a nightclub in Bucharest on October 30, 2015, and killed 30 people. It led to a massive protest accusing the incumbent government of corruption until November 8, 2015, when the Romanian Prime Minister Victor Ponta stepped down with his cabinet (Tran, 2015). The research controls for this event, adding an extra variable in the regression. Both national economies of Hungary and Romania are dependent on services, including tourism revenues, and consider their reputation in international news and social media, where prospective tourists could seek information about the countries (Findthedata, 2015; Indexmundi, 2015).

Our assumption that the Hungarian government values its international image in the eyes of tourists and news media is supported by the authorities' reaction to the recent increase in a tourist flow. In the first half of 2015, the number of international tourists to Hungary has increased by 17.5 percent, which is four times the global average, according to the World Tourism Organization (Varga, 2015). However, the Deputy State Secretary for Tourism Ádám Ruszinkó pointed out that despite the impressive growth rate since 2010, the tourism industry is a "fragile genre" and "a couple of bad news may cause massive damage" (Varga, 2015).

Twitter is an appropriate data source for this study because it is known to react instantly to crisis events with real-time conversations. In addition, this research suggests that destination management conversations about places on Twitter can be considered

equivalent to real word-of-mouth (Tham et al, 2013). In this study, the international reputation of Hungary is the dependent variable that is measured through a sentiment analysis of randomly selected tweets that contain the word 'Hungary' in the period of one year from March 2015 to March 2016. We study tweets collected in English because the project measures the *international* reputation of Hungary. The sentiment analysis is conducted before, during, and after the crisis to spot the changes happening to the place brand of Hungary.

Place brand is an image based on geographical and psychological specifications of the place that is reflected in the minds of public (Sevin, 2011). The majority of place branding studies are qualitative and they rarely explore the reaction of social media users to a place brand crisis (Lucarelli & Brorstrom, 2013). This study contributes to both place branding and crisis communication theories by evaluating the reputation of place brand on social media at different stages of a crisis (Benoit, 1995; Lyon & Cameron, 2004).

Specifically, the study investigates how the international reputation of Hungary on Twitter changed after the inflow of refugees and the consequent events that happened during the 2015 European refugee crisis. The research assumes that the events of the 2015 European refugee crisis and inflow of migrants had a negative influence on the reputation of Hungary on Twitter. Analytically, the study uses a difference-in-differences design based on a natural experiment.

The 2015 Refugee Crisis Events in Hungary

In the summer of 2015, more than 200,000 people primarily from Syria entered Hungary with a key goal of getting to Germany that welcomed refugees. These migrants preferred to get to Hungary from Greece, which, contrary to the Schengen Agreement,

had not registered those migrants and allowed them to further move to other countries such as Hungary. On July 13, 2015, Hungary started building a temporary metal fence along its 109-miles-long border with Serbia, a non-EU and non-Schengen country, in order to stop the flow of refugees (Gorondi, 2015). The Foreign Minister of Hungary Peter Szijjarto announced that the country "will apply temporary border seals on every border section where there is no other effective way to impede illegal immigration" (Jahn, & Cook, 2015, p.1). Two months later, on September 9, right before the fence was completed, the record number of 6,000 refugees entered Hungary in one day. Simultaneously, the country had issued a law prosecuting anyone who illegally entered Hungary. Migrants who reached the border of Hungary but could not proceed further went on a hunger strike asking the Hungarian government to open the border and let them reach Germany and other European countries as a final destination.

On September 9, major international news media such as CNN and the Guardian aired a video, in which a Hungarian TV reporter had tripped a refugee who was carrying a child and running away from the police near the border with Serbia. The video went viral on social media after several major international news outlets broadcasted it. For example, the Guardian's article with the video received more than 80,000 shares, and it was one of a few news outlets reporting about the events at the Hungarian border (Nolan, 2015). Consequently, that reporter was fired and faced a fine (Mackey, 2015). In a continuation of the crisis events, that Hungarian reporter again received media attention when she intended to sue Facebook for "facilitating campaign against her" (Buchanan, 2015).

During the refugee crisis events, the Hungarian police had also attracted adverse media coverage, when in September several international and human rights groups claimed being harassed by the Hungarian police. The Organization for Security and Cooperation in Europe received four registered cases of the police harassment of journalists. For example, while covering the events near the country's border for the Associated Press, a reporter was detained and a cameraman was forced by the police to delete the footage about the refugees. Another journalist from Poland had received head injuries struck by the Hungarian police. The Hungarian officials denied any involvement of the country's police with these incidents. The human rights organizations explicitly demanded that Hungary stopped harassing journalists and obstructing their work (Gera, 2015).

On September 18, 2015, Hungary started building a new fence on its border with another EU member and a non-Schengen Area country, Croatia. On October 16, Hungary closed its border with Croatia, moving the flow of migrants to Slovenia. Hungary received a broad international criticism from international media and human rights groups such as Amnesty International. For example, on September 17, Amnesty International tweeted: "*#Refugees at border are suffering the extremely dire consequences of authorities in #Hungary. http://amn.st/6010BJIHj",* which was 'liked' 203 times and retweeted 591 times. The ones who questioned the Hungary's inclusion in the European Union shared a hashtag #Hungxit. For example, a Twitter user Marcus Engert texted on September 3: *" #Hungary breaks elementary European rules, Again. Tie to discuss a #Hungxit, hm? #refugeecrisis #Budapest #Keleti #magyarorszag #refugees*" (Engert, 2015).

On the contrary, the internal approval rating of the Hungarian Prime Minister Mr. Orban has significantly increased in the period from April to October 2015. The increased approval rating was likely connected with the support of the political course of dealing with the refugee crisis (Trust.org, 2015).

On September 23, Romania along with Hungary and three other EU countries voted against asylum-seekers' relocation among other EU members (Die Welt, 2015). Romania had not received the intensive media coverage of the same strength until a fire broke out in a Bucharest nightclub on October 30, which led to a 20,000-people protest and a consequent resignation of the Romanian Prime Minister Victor Ponta on November 4, 2015 (Tran, 2015).

On September 28, a very popular TV show, "the Last Week Tonight Show with John Oliver", aired an episode titled "Migrants and refugees" about the misrepresentation of these groups by media and a hostile treatment by the Hungarian reporter, politicians, and police. The episode showed the video created by the mayor's office of the Hungarian city of Ásotthalom threatening migrants crossing the country's border. The video reached over 5,4 million views on YouTube only, not counting the entire audience of this HBO show (Last Week Tonight, 2015).

This study investigated how the inflow of refugees and the consequent events of the 2015 European refugee crisis affected the reputation of Hungary on Twitter. The study hypothesizes that the 2015 European refugee crisis negatively affected the reputation of Hungary on Twitter. The following literature review shows that several researchers used Twitter in place branding studies but not for estimating changes in place brand's reputation before and after a crisis happened at the place.

LITERATURE REVIEW

Global urbanization forces countries, regions, cities, and other entities (generalized here under the umbrella term 'places') to compete with each other for better images in order to attract investments, tourists, talents, and signature events. With technological advancements, many places recognize the necessity of being promoted on social media and developing positive images in the digital environment. Social media has become a useful tool for launching and maintaining place marketing campaigns, monitoring opinions about a certain place, and for communicating with the existing and future customers (residents, tourists, investors). Social media, therefore, provides a unique opportunity to analyze place brands through the eyes of consumers, by conducting sentiment analysis or opinion mining.

This literature review first defines a place brand, its components, and ways of measuring it, and then describes how a crisis could affect place brand. The posed question is whether place brand's attributes change according to the events that are happening in that location.

Defining Place Brand

To begin exploring place brand, it is necessary to study its constituents such as an image. The image is a "structured and dynamic representation" of something that is reflected in people's minds and opinions (Velcin et al, 2012, p.818). That is why it is

crucial to explore the image on social media, a virtual space, where the conversation is ongoing, inclusive, and reactive to events. Social media gives people with computer skills and Internet access an opportunity to express their opinions and participate in a conversation about a certain issue.

Place brand is "the perception of a place by external and internal audiences" based on geographical, physical, and psychological specifications of the place that is reflected in the minds of public (Sevin, 2011, p.4). To explicate a concept of the image more broadly, it is important to know that it is a summary of perceptions in a consumer's mind, which is difficult to influence directly. (Anholt, 2009; European Travel Commission and World Tourism Organization, 2014, p.34). That is why advertising about a place is not a primary source where people perceive places: word-of-mouth (WOM) and news media coverage play a significant part in the creation of place brand images. Thus, social media provides a relevant platform for not only learning about people's perception of place brands, but also an opportunity to study how these perceptions evolve over time.

One of the characteristics of place branding that particularly limits the development of it as a discipline and complicates research on place branding is its connection with a number of other research areas. Lucarelli and Brorstrom (2013) analyzed 292 articles on place branding and found out that only 36 percent of the studied literature is primarily related to the discipline, while the majority of the articles correspond to other disciplines such as urban studies (15 percent), tourism (12), planning (7), geography (6), marketing (6), and branding (4). Other disciplines connected with the topic to a smaller extent are management, economics, social science and sociology, retail,

and regional studies. The findings of several scholars acknowledge the interdisciplinary nature of place branding.

Since place branding is interdisciplinary, it is important to look at it from the perspective of the most related area of research - urban studies. A classical work by Lynch defined the image of the city [place] as the series of physical elements that transform into public images: paths that one routinely takes (roads, walks, streets), edges or boundaries between the paths (waterfronts, road cuts, walls), districts, nodes (moments of shift from one place to another, which Lynch called "strategic spots in a city"), and landmarks (Lynch, 1960, p.47).

Perhaps one of the most interesting urban studies observations of place branding is the influence of physical appearance on the entire image of the place. That is why many places tend to use their most famous landmarks for place advertisements as well as hire renowned architects to design and build attractive buildings, such as Burj Khalifa in the city of Dubai. In the article for the Guardian, the Brooklyn College professor and the author of books on urban life and culture, Sharon Zukin, described a case when these tactics and place marketing campaigns cause criticism by both citizens and scholars because these efforts are often expensive and their long-term outcomes are not always clear (Zukin, 2014). For instance, the government of Rio de Janeiro put a lot of efforts in hosting both the 2014 World Cup and 2016 Olympics, but the local population protested against tremendous city spending on these events.

Place Branding Campaign

How is a place branding campaign created? A governing body of a place, such as city, region, or national government, commissions an external or internal agency that

creates a new or builds upon an existing visual identity of a place, and then launches a marketing campaign in accordance with the campaign's goals. The objective of a place branding campaign could be in attracting tourists, drawing investments, repositioning the existing image, etc. However, place branding is more than just creating attractive logos and slogans. Place branding is about managing reputation and brand equity of a place.

Robert Govers, one of the most influential scholars in the area of place branding, argues that logotypes normally serve as creators of awareness for goods that do not have a wide recognition. On the contrary, cities and places usually have a vast reputation and a distinct image in people's minds. Unlike consumer brands, places do not fall into specific product categories, i.e. food or clothes, because they represent various aspects, such as travel, study, relocation destinations, investment opportunities, spaces for cultural events, producers of goods and many others. That is why a place branding campaign could turn into a complicated task for marketers who develop these campaigns (Govers, 2013).

In addition, it is harder to control the image of a place than the image of a product because more "image formation agents" are involved in this process. From the most to the least important, these are organic (personal experience), social (word-of-mouth, opinions of friends and relatives), autonomous (mass media), and induced agents (advertisements) (Govers, 2013, p.2). Advertising and marketing messages are called induced agents and considered to be the least valuable and perceivable by the audience (Govers & Go, 2009). That is why social media could become the source for researching, creating, and evaluating place brand because it provides a constant access to people's perception of places. This paper further discusses the examples of using social media in place branding research.

Scholars urge marketing professionals to develop a consistent 'strategy and substance' behind the visual identity of their subjects (logos and slogans). In order to create place brands, many campaign managers orient on very developed examples such as of Amsterdam or New York City, which have the "luxury of historically built global awareness and reputation", unlike the majority of other places (Govers, 2013, p.4). Therefore, it is important for governments to implement place brand strategies and analysis in the process of creating their visual identities. This study contributes to the place branding literature by applying both content and sentiment analyses of place brands on social media.

In 1991, Aaker, one of the most renowned brand strategists in the world, introduced the model of brand equity and described the way marketing and advertising specialists conducted the research on brand associations. One of these ways is a survey or a qualitative study of the associations that a given brand reveals. The potential downsides of these methods are the required time and financial resources, which lead market researchers to find new ways of conducting research for branding campaigns.

Moving forward to the present day, with the development of digital tools and social media, it is easier for market researchers to explore brand associations constantly while employing fewer resources. Currently, there is an increasing number of software providers that help track brand associations, sentiment analysis, and media presence of a single word, brand, or hashtag. Examples include Twibuzz, Tweetreach, Hashtracking, SocialMention, etc. These tools and the application of a social media analysis help place branding specialists who adopted these technologies research, monitor, and manage place brands.

Place Brand Equity

The reputation and value of a brand are often estimated through its brand equity. The concept of brand equity was introduced by Aaker as a set of "assets or liabilities linked to a brand's name and symbol that add or subtract from the value provided by a product or service." These assets are divided into five categories: brand awareness, brand associations, perceived quality, brand loyalty and other proprietary brand assets such as trademarks or patents (Aaker, 1991, p.15-16; Aaker & Joachimsthaler, 2000).

There are several ways for a place to enhance its brand equity. One of them is to affiliate that place with a strong brand or initiative. It could be a sports event or a culture initiative with a broad recognition. For example, in 2011, the Finnish city of Turku was awarded the status of the European Capital of Culture (ECoC), a prestigious title that is given to one or two economically and culturally recognized cities in Europe on a yearly basis (Hakala & Lemmetyinen, 2013).

This kind of events contributes to the development of a place brand and its equity in both short and long terms. In the case of Turku, this process started with the application in 2004 and continued until 2007, when the city was chosen for 2011 ECoC (Hakala & Lemmetyinen, 2013). A careful and long-term enhancement of a place brand through co-branding could lead to several tangible outcomes. In the short and long terms, those are an increased number of domestic and international travelers, a consequent hotel and tourism revenue growth, augmented visibility in local and international media outlets, activation of projects and initiatives, and increased cash flow from foundations and other entities. In addition, one of the key factors in boosting place brand equity through co-branding is the active engaging and maintaining of relationships with several

important stakeholder groups: project and event organizers, local people, visitors, and, most importantly - media.

Methods of Place Brand Evaluation

Evaluating place brand is not an easy task because there are multiple stakeholders and events involved in the creation of it. However, researchers identified several approaches to doing it effectively. There are two main ways of measuring the success of place branding activities. The first type is the *customer-oriented* analysis that aims to measure place brand through the eyes of a customer and analyze either customer equity or customer satisfaction (Zenker, 2014, p.152). The second type is a *brand-oriented approach*, a brand value driver that studies a consumer's reaction to a brand, and brand equity that measures "consumer response to the marketing of the brand" (Zenker, 2014, p.152).

One of the tools that helps researchers measure *brand image*, a part of brand equity, is a Brand Concept Map (aBCM) that is constructed in two stages. During the first step, *the elicitation*, major brand associations are identified through interviews; then, at the second stage, *the mapping*, respondents create their own maps based on the discovered associations (Zenker, 2014, p.160). The a- component to the BCM model is the additional process during the mapping - individual evaluation of judgments for every association and their ranking in accordance with "a purchase decision", which is a willingness to consume the place. The final step is necessary because the associations are not equally important when a consumer decides to travel or to invest in a certain place (Zenker, 2014, p.162).

There is also a conceptualized three-dimensional framework for the place brand analysis based on its elements, measurement, and brand impact. There are several groups of elements that participate in the construction of a place brand: history and heritage, artifacts and spatial plan, events and activities, process and institutions, and symbols such as logos and slogans (Lucarelli, 2012, p.236). This appropriation of terms from urban studies shows a strong connection between those two fields, urban studies and place branding.

Place brand measurement includes all research methods that evaluate the impact of place brand. According to the 2013 Lucarelli and Brorstrom's analysis of 292 articles on place branding, most of the published works were conducted by using case studies and qualitative methods. The rest of the studies used quantitative methods, which were primarily based on questionnaires and surveys, and few works used mix methods, including analysis of primary and secondary data. A place brand impact measures the impact of place branding on different dimensions: employment rates, tourist flows, or other economic or socio-political data. Overall, this three-dimensional framework explains why the equity of a place brand is different from a consumer brand: places are very dynamic and harder to control than consumer brands.

Traditionally, place branding researchers measured public mood and attitude towards places by applying qualitative methods such as in-depth interviews or surveys, which required time, financial resources and needed to be conducted frequently to follow constant changes happening there (Lucarelli & Brorstrom, 2013). However, in the past years, a quick emergence of new media and social networks as well as overall technological development increased the popularity of quantitative methods in the field.

More than 200 companies sell social media data and its analysis and many organizations developed their own methods of crawling and monitoring customer opinion and satisfaction (Schreiner & Go, 2012).

Place branding researchers only recently started approaching social media such as Twitter and Facebook as their study field, so the number of place brand studies on social media is scarce at this point. Twitter is a real-time social network that allows users to create profiles, share 140-character messages as well as a visual content and external links (Kawash, 2014; Shrivatava, Mayor, and Pant, 2014). Hashtags are the words or their combinations started with a #, which are created by users to refer to a specific topic or phenomenon; retweet is the re-posting of a message of another user started with "RT" and the user's name (Seargeant & Tagg, 2014, p.139). This study counts retweets in the randomly selected pool of tweets because a retweet can have a bigger effect on the development of a topic and its discourse compared with the individual tweet that is not retweeted (Hawthorne et al, 2013, p.554).

What follows is a brief review of two studies, in which tweets are used as the data source in place branding research.

Unlike traditional news sources that are sometimes considered to play a role of agenda-setters and gatekeepers of information, social media allows a wider range of people to express their opinions about certain issues, including their attitudes towards places. There are two relevant research works that use Twitter to crawl the image of a place. The first one is the analysis of people's attitude towards the city of Johannesburg as an organization (Schreiner & Go, 2012). For one month all tweets that contained the name of the city in different variations were collected, excluding the irrelevant to the

research topic tweets. Then, all words in tweets were divided into two groups based on their proximity to the studied term, the name of the city. The maximum length of one tweet is 140 characters, so the closest ten words to the term were considered to be close-proximity, and the rest of the words – of extended proximity. The references were calculated and put into two tables: "top-of-mind" and extended (Schreiner, & Go, p.126). By using this method, the researchers were able to identify what issues were of the most importance and concern to the users of Twitter, who tend to represent economically active residents of the city; as well as what topics directly affected how people felt about the place. Given the access to a social media analytics tool, this type of research usually requires fewer resources in comparison with a traditional qualitative study.

Another study is a research of user-generated place brand equity of the city of Stockholm through the analysis of more than 70,000 tweets that contained the name of the city in the period of three months. Thematic and semantic analyses were conducted on the content of tweets to identify five major themes in Stockholm's brand equity. Interestingly, the most mentioned association with the city name was the Stockholm syndrome, followed by events in the city, reflections on trips to Sweden, and explanations of the country's culture. The study concluded that Twitter is a new platform for the formation of a city brand *meaning* (Andéhn, Kazeminia, Lucarelli, Sevin, 2014, p.106).

Reputation and Crisis

The projected image and visual identity are constructed and projected towards the audience. On the contrary, "reputation is *owned* by publics" (Lyon & Cameron, 2004, p.215). It is an aggregate evaluation of how stakeholders perceive organization fulfilling its promise (Coombs, 2007, p.164). A good reputation is considered important for several

reasons, not only for financial benefits. It often expresses in beneficial relationships with internal and external publics, stakeholders, and decision makers. However, it is hard to measure the reputation because many factors intervene in the process of reputation formation.

Public perception of reputation is constructed on the basis of its financial, institutional (i.e. social responsibility), and strategy signals (Frombrun & Shanley, 1990). The researchers evaluated reputation based on its quality of products and management, innovativeness, attractiveness, community and environmental responsibility (Bovet, 1994). There are several ratings that evaluate reputation. For corporations, one of the reliable ratings is the Forbes Most Reputable Rating, and for countries - the Good Country Index. The countries are rated based on their development of science and technology, culture, international peace and security, world order, planet and climate, prosperity and equality, health and wellbeing. According to the 2015 rating, the U.S.A. takes the 23rd position, Hungary – 61st, Romania – 65th, Russia – 95th (Goodcountry.org, 2015). The data is collected through the United Nations and other international nongovernmental organizations. The rating has got more popularity in the last couple of years.

There are many studies that described how a company should address its stakeholders during and after a crisis (Coombs, 2007; Coombs & Holladay, 2008; Park & Cameron, 2014; Yang, Kang & Johnson, 2010). However, every crisis is different, which complicates drawing conclusions based on a single case. The reputation will most likely deteriorate if the organization is perceived and found responsible for the crisis. The stakeholders could ‘break up’ with that organization or create negative word-of-mouth. It

is the responsibility of the company's management to prevent this situation (Coombs, 2007, p. 136).

A crisis can characterize the company's behavior by taking into account its history of dealing with crises and the treatment of its stakeholders. There are two intensifying factors that affect the post-crisis communications. Consistency, or crisis history, describes whether a company was involved in similar situations before. High consistency means that a company experienced several crises prior to this particular one. Distinctiveness characterizes the relationships of a company with its stakeholders before a crisis. The higher the distinctiveness, the better the relationships of a company with its important audiences (Coombs, 2007, p. 137).

Many crisis communication scholars refer to the *Attribution Theory*, which implies that people tend to look for a reason why a negative event or crisis happened and consequently search for a responsible entity. It is important to note that during and after a crisis the management has to deal not only with the reputational damage, but also with the legal and financial consequences of a crisis, such as sales decline, decreased purchase intention, and failing market share.

The Situational Crisis Communication Theory (SCCT) assesses how a crisis could impend the existing reputation. There are three clusters of response: victim, accidental, and preventable, in which organizations intentionally put people at risk (Coombs, 2007, p.168). All of them have different strong attributions of crisis responsibility. (Coombs & Holladay, 2008). The perception of a crisis might differ based on the source, from which people get their news. The data from the Pew Research Center showed that TV still remains a primary news source for the majority of the U.S. population as of 2009

(Coombs & Holladay, 2009, p.1). However, when the researchers compared TV news with newspaper articles, the perception of the crisis did not differ significantly between these two sources (Coombs, Holladay, 2009, p.4). In addition, there was no significant difference between applying sympathy and compensation strategies. Therefore, it would be beneficial for the future studies to compare other news sources, especially social media, to spot the peculiarities in the audience's perception of a crisis.

There are several theories and guidebooks of how to manage crisis communications and organization's reputation in different situations. It is important to note that crises have dynamic nature; therefore, communications during the crisis also do not take only one position, they could move from advocating the organization to accommodating the crisis, its victims, and publics (Pang et al 2010, p.533).

Any crisis communications process is very dependent on how close and autonomous public relations practitioners to the organization's top management or other decision makers that are collectively called dominant coalition in research works (Shin et al, 2006, p.286). Crisis communications scholars stressed the importance of PR practitioners' participation in the "dominant coalition" (a decision-making group) and having an influence upon the important decisions during the crisis (Shin et al, 2002).

Research Hypothesis

Both place branding and social media analysis are emerging areas of reseach and are not vastly developed disciplines. Nonetheless, place branding interests many researchers across various disciplines, including media studies, marketing, urban planning, and computer science. Social media and Twitter, in particular, provides an opportunity to dive into people's opinions about certain places in order to identify brand

associations with a place, conduct research about branding campaigns, or even check the performance of a specific area's governance. Unlike traditional media, online social networks allow users to express their opinion in an open way, which adds an external validity to the research as it analyzes the expressed real-time opinions of users and not in artificial laboratory conditions or surveys. However, there are some constraints in researching place brands on social media, such as the absence of a solid body of prior academic work, the dominance of qualitative studies in the field, restricted access to sentiment analysis software for researchers, and other peculiarities typical to any developing area of study.

According to a report by the United Nations, in 2014, 54 percent of the world's population resided in urban areas, and the number is expected to grow up to 66 percent of the world's population by 2050 (Department of Economic and Social Affairs, 2014). This global trend of increasing urban population suggests the intensification of competition between places for resources, which, in turn, will facilitate the field of place branding and make it an even more interesting area for the researchers who are interested in interdisciplinary studies.

Few studies have investigated and measured a place brand's reputation at the different stages of a crisis. In addition, even fewer scholars used quantitative methods for researching place brands in crisis. Our research draws on a recent crisis with the investigation of how a crisis and consequent negative media coverage affect a place brand. By doing it, the research contributes to the fields of place branding and crisis communications to investigate how place brands are changed due to the sudden events and negative coverage and by drawing implications for the field of crisis communication.

Specifically, the study investigates whether the place brand of Hungary was impacted by the 2015 European refugee crisis and posits the following hypothesis.

H1: It is hypothesized that the 2015 European refugee crisis negatively affected the place brand of Hungary on Twitter.

Additionally, the thesis poses the following question:

RQ: Has the hypothesized negative effect (posited in H1) persisted after the immediate upshot of the crisis?

METHODOLOGY

This paper employs several methods of conducting the analysis of tweets. First, we conduct content analysis of media sources to establish time periods of the crisis, and then we conduct sentiment analysis, which measures the sentiment of tweets – in a range from positive to negative valence. To calculate this valence, we apply two lexicon-based approaches that use dictionaries with identified sentiment and polarity. Second, since all tweets are related to one or another topic about the country, we apply a Latent Dirichlet Allocation, a classification model that fits the discrete data such as text (Blei at el. 2003). This model divides each tweet into separate words or phrases and creates the corpora representing word count in each tweet.

In addition, the events in Hungary associated with the 2015 refugee crisis fit a difference-in-differences method in a natural experiment, which happens when an unexpected event, such as a sudden migrant flow to Europe and a decision of refugees to get to Germany through Hungary, assigned Hungary as a treatment group and Romania as a control group. What follows is a more detailed description of the methods implemented in this paper.

Content Analysis

As mentioned above, in this research, we are doing two content analyses: first, the analysis of media to identify the time periods of the crisis, the second – on tweets.

Content analysis is one of the increasingly popular research methods in mass media. One of the first practical examples of modern content analysis was devised during the World War II. The Allied intelligence units researched the content of the European radio stations to identify an increase of a number of German songs played, in order to locate and identify the movements of the German troops on occupied territories (Wimmer & Dominick, 2014, p. 159).

There are several definitions of the content analysis, and the current paper used one defined by Kerlinger (1986) as a method of analyzing communication and measuring variables in a *systematic*, *objective*, and *quantitative* manner (Wimmer & Dominick, 2014, p. 160). Systematic manner implies devising special rules for approaching and analyzing data. Objective manner means explicating operational definitions and rules and providing a set of criteria. Quantitative manner implies that the collected data will be represented precisely and the research findings presented succinctly (Kerlinger, 1986; Wimmer & Dominick, 2014, p. 160).

What follows is the example of using content analysis in place branding research as well as a sample analysis of the place brand of Hungary. Some governments establish companies or organize departments that purposefully promote their place brands. One of them is Brand USA, a private-public entity that aims to promote the United States as a premier travel destination (Thebrandusa.com). Sevin (2013) analyzed the content and purposes of more than 5,000 tweets created by five most mentioned by Brand USA destination marketing accounts (Illinois, San Francisco, Idaho, Texas, and Milwaukee Area) in the period of one year starting from October 2011. The content of the tweets was divided by their usage pattern: “outside link, hashtag, mention, and retweet” (Sevin,

2013, p.3). The study showed that 78 percent of the place brand tweets mentioned other accounts, and 65 percent of the studied tweets had outside links, compared with 25 percent in an average Twitter account. The content analysis showed that these accounts mostly described events and attractions, rather than conveyed a local identity of the place or its residents. Twitter usage by these destination accounts was limited as it primarily aimed at promoting events in places to the mass audience. Place marketers need to better execute a social media strategy in destination branding campaigns to influence their target audiences.

As in any research method, there are some limitations inherent to a content analysis. For example, the implications of the research alone cannot measure the impact of messages on audience or media effects. The results of the study may also differ based on the working operational definitions and categories of that research. The stage of selecting a unit of analysis is crucial to content analysis and it means defining "the smallest element" but also "the most important" studied unit (Wimmer & Dominick, 2014, p. 167). In this research, the unit of analysis is a single tweet that contains the word "Hungary" or "Romania."

Here are the results of a sample content analysis of two newspapers, the Guardian and the New York Times, which helped identifying periods before, during, and after the 2015 refugee crisis in Hungary. Table 1 provides a total number of articles about Hungary and Romania that demonstrate the increased coverage of the refugee crisis in Hungary from June to October 2015:

Table 1.
The Number of News Articles about Hungary and Romania in both the New York Times and the Guardian

Period / Country	Pre-crisis (March-May)		Start of Crisis (June-Oct.)			Post-crisis (Nov.- Feb.)	
	# of articles	Refugee crisis	# of articles	Refugee crisis	Negative sentiment (refugee crisis)	# of articles	Refugee crisis
Hungary	16	0	206	187	160	12	6
Romania	12	0	34	3	1	13*	0

**Note.* The increase in the number of articles about Romania in the post-crisis period is associated with the massive protest that led to the government's resignation.

Sentiment Analysis

Sentiment analysis, which is sometimes called opinion mining, is one of the most widely used functions in social media research, which stands for the measurement of a "conversation's valence" from a positive to negative range, determined by a researcher (Shrivatava et al, 2014). Other important dimensions include buzz or comment analysis, mindshare (currently trending news and comments), a meme (rising trends and ideas), and a share of voice (a visibility of a studied phenomenon) (Wilcox et al, 2015, p.145).

In recent years, there has been an increase in a number of companies that provide 'crawl' software with sentiment analysis – a service that detects what people talk about in regards to a specific issue, by analyzing social media content (Pang and Lee, 2008; Velcin et al, 2012, p.819). In addition, there are some classification tools that help researchers automatically assign tweets based on their sentiments (Shrivatava, Mayor, & Pant, 2014). The examples of these tools are Topsy, Simply Measured, Lexalytics, etc.

To evaluate place brands, some researchers employed the *signaling theory,* according to which place branding is considered to be a sign that transmits "place's

quality" and ability to "satisfy residents' needs" (Shafranskaya & Potapov, 2014, p.117). This example study conducted a consumer-oriented analysis of place brand equity on Twitter of the city of Perm, Russia. The researchers identified 16 attributes of the city that represented the highest importance to citizens due to both functional and symbolic utilities. Then, the researchers assigned these attributed to four groups: city diversity, city safety and comfort, professional and job chances, and city facilities (Shafranskaya & Potapov, 2014, p.123). In the example study, two pools of respondents assessed these attributes: a group of highly educated people with above-average income and a group of people with a lower level of education and salary. The responses were compared against a perfect score of 100 and revealed a 45 percent satisfaction rate of the city brand (Shafranskaya & Potapov, 2014, p.126). This kind of place branding studies help governments identify which attributes should be improved in the studied place. In addition, this study represents an example of how social media could be employed for opinion mining and place brand's evaluation.

Not many place branding researchers employ quantitative methodology, and there is a lack of existing research about the changes in place on social media during a crisis. Our research contributes both to the quantitative study of place brands as well as crisis communications research. In addition, there are several sentiment analysis tools that differ in their sentiment categories, coding schemes, propriety, and applicability to specific disciplines (Young & Soroka, 2012, p.206). This study has chosen a specific tool, SentiStrength, and acknowledges limitations of using the automated sentiment analysis, such as the accuracy and specifications of each tool in particular as well as a possible necessity to check the results with the help of human coders.

Lexicon-based Sentiment Analysis

The automatic sentiment analysis of texts is still an evolving field of study. There are several prominent scholars in the field, such as Qian Liu, Zhiquian Gao, Bing Liu, Yuanlin Zhang, etc. Sentiment analysis deals with the computational treatment of languages, opinions, sentiment, comparison, and other speculations in a studied text (Liu, 2012; Pang & Lee, 2008). Liu (2012) defines two types of the sentiment analysis: an unsupervised learning that employs predefined dictionaries to divide texts based on used positive, neutral or negative words and phrases; and a supervised learning that involves training of the machine to match the texts with known responses, such as restaurants reviews and the corresponding ratings.

There are three steps in a typical sentiment analysis study: depending on the initial text, there is a stage for dividing the dataset into separate sections, such as sentences (Pang & Lee, 2004). Then, there is study of data's polarity, and division by the topic (Gamon et al, 2005). The majority of sentiment analysis studies deal with the negative-positive polarity, whereas some of them also study emotions such as anger or surprise expressed in the text (Thelwall et al, 2010; Soroka et al, 2015).

Twitter in particular has attracted sentiment analysts and scholars' attention (Jiang et al., 2011; Hu et al., 2013). Twitter as a communication tool includes messages on various topics and public interests; the discussion is not moderated, which happens in case of a newspaper or blog with an editor. Twitter and other social media have been used to conduct sentiment analysis of brand names. However, this application is not apparent in the field of place branding.

Lexicon-based approach employs dictionaries with identified sentiment and associated polarity, such as Lexicoder or SentiStrength (Thelwall et al, 2012). It is possible to study the negative sentiment further by identifying the emotions in tweets using Lexicoder, which is commonly used for the analysis of speeches in the field of political communications (Soroka et al, 2015). Other applied algorithms, such as SentiStrength, are oftentimes used for the social media analysis, and on average recognize 60.6% of negative sentiment and 72.8% of positive sentiment, according to multiple tests conducted by the researchers (Thelwall et al, 2010).

Latent Dirichlet Allocation

Tweets about one topic, such as a country, can still correspond to a variety of issues. Once the sentiment of the tweets is established, it is useful to examine the identities and topics of negative tweets more closely. Classification techniques such as Latent Dirichlet Allocation are designed to help the researchers with this task. We use LDA to divide the massive of tweets into several topics, such as tourism or a refugee crisis, which further helps us analyze the sentiment of these tweets.

Latent Dirichlet Allocation is a classification model that fits the discrete data like text (Blei at el. 2003). The model divides each tweet into separate words or phrases, and creates the corpora representing word count in each tweet. Model treats each occurrence of word in the tweet as an observation. Each tweet and word has a probability to belong to each topic k, with probability distribution being Dirichlet (Blei at el. 2003). A researcher provides the model with tweets, number of topics K, and parameters of the Dirichlet distribution. Having these inputs, the model runs a Bayesian hierarchical model

and produces posterior distribution of probabilities for each word and document to belong to each topic.

To choose the optimal number of topics K for classification, researchers can examine the estimated maximum likelihood and choose K, which produces the lowest values of the Akaike Information Criterion (Akaike, 1973) or the Bayesian Information Criterion (Schwarz, 1978; Tang et al, 2014).

Difference-in-differences in a Natural Experiment

A natural experiment happens when an unexpected event or external reason "assigns participants randomly to potential treatment or control group" (Murnane, 2011, p.136). Internal validity is crucial to experiments. The conditions of natural experiments are called *exogenous* because the participants do not control assignment to these conditions (Murnane, 2011, p.136). This research explores the natural experiment that happened to the reputation of Hungary due to the unexpected refugee crisis events in the summer and fall of 2015.

In a canonical example of the natural experiment, Dynarski (2003) investigated how the change in the Social Security Survivors Benefits (SSSB) federal policy affected the attendance and completion of college by the children of the program beneficiaries. From 1965 to 1981, the policy provided financial aid for the children of deceased, disabled, or retired SSBN grantees. At the program's paramount, the grantees' children comprised up to 12 percent of all students enrolled in U.S. colleges.

Three years after the program's elimination, which happened to become the biggest change in the U.S. education policy, the college enrollment had dropped sharply. Therefore, two groups of high-school seniors within two years (a year before the

program's elimination and a year after) provided a basis for the natural experiment with difference-indifference discontinuity design. Both groups were eligible for the grant, but one had a chance to receive that educational grant before 1981, and the other group had not because the policy changed. The study found out that the program's elimination decreased the probability of attending college by third (Dynarski, 2003). In addition, some researchers argued that the increase of a sample size through the extension of the time period could have increased the internal validity of this natural experiment. However, in many cases, external factors could influence the validity as well because other conditions change over time (Murnane, 2011).

The difference-in-differences method has become popular among quantitative studies after 1985, when Ashenfelter and Card pioneered this method for estimating the relationships between the CETA training and the consequent earnings of trainees. The study compared two groups of data (predicted and actual earnings) in two time periods (prior and after the training). The researchers concluded that if the comparison is the appropriate criterion, then the effect on salary after the training is significantly stronger for female participants (Ashenfelter & Card, 1985).

As the example above demonstrates, in the difference-in-differences method, the outcomes are observed for two groups in two time periods. The first group is exposed to treatment only in the second period of time. The second one, which is called a control group, is not exposed to the treatment in either period (Imbens &Wooldridge, 2007; Murnane, 2011).

In this study, the treatment group is a set of tweets with the word 'Hungary', and the control group is a set of tweets with the word 'Romania' in them. The period before

the crisis events is March – May 2015, the crisis period is July – October 2015, the post-crisis time is November 2015 – February 2016, as the preliminary content analysis of newspaper articles showed.

Hungary and Romania joined the European Union in 2004 and 2007 respectively and have relatively similar geographical location and demographics (Indexmundi, 2015). Both countries have transitioned from centrally planned to the market economies in the 1990s and sought financial aid from IMF after the 2008 economic crisis. The countries' GDP is dependent on services (Romania - 59 percent, Hungary – 69 percent) such as tourism revenues (Indexmundi, 2015). Interestingly, according to the 2012 data, the number of refugees and internally displaced people composed 248 people in Romania (0.00001% of the country's population) and 111 in Hungary (0.00001% of the country's population) (Indexmundi, 2015). Additionally, before we completed the difference-in-differences analysis, we checked the common trend assumption. Figure 1 plots the common trend of the change of negative sentiment over time – before the crisis, both trends for Romania and Hungary looked similar.

Figure 1. Trend in overall negative Sentiment of tweets

RESULTS

Data Description

The studied dataset contains a sample of about 50,000 randomly selected tweets that include words “Romania” and “Hungary” in the period from March 19, 2015, to March 22, 2016. The data is collected using Twitter API and represents approximately one percent of all tweets that had contained the words ‘Romania’ or ‘Hungary’ in the period of one year. Twitter API provides an opportunity to retrieve one percent of all tweets about a specific topic: Morstatter et al (2013) compare the tweets sampled from Twitter API and Twitter’s Firehose (an entire universe of tweets) and argue that while Twitter API has some caveats, it represents a valid research sample of tweets. Certainly, the possibility of having an entire dataset of tweets about Hungary would bring an utmost benefit to this study; however, a major downside of Twitter Firehose is its “restrictive cost” (Morstatter et al, 2013, p.1). The resulting dataset represents 47,572 randomly selected tweets, with 26,355 tweets containing the word “Hungary”, and 21,217 tweets with the word “Romania”.

Selecting the Sentiment Analysis Dictionary

We have compared and applied two dictionaries for the sentiment analysis. The first dictionary is a lexicon-based sentiment dictionary, which had been developed by Liu and Hu from 2004 to 2015 (Liu, 2015). The dictionary developed by Liu and Hu is a

popular method for allocating the sentiment of consumer product reviews. One can think about tweets about countries as some form of a review, so we picked this dictionary as a candidate for the analysis. The second dictionary is the SentiStrength dictionary, developed by Thelwall and Buckley (Thelwall et al 2010; Thelwall et al, 2012; Thelwall & Buckley, 2013). SentiStrength is a polarity-based sentiment dictionary designed for the analysis of short informal text and social media (Sentistrength.com, 2016).

Both SentiStrength and the dictionary of Liu and Hu are reported to be the state-of-the-art methods for conducting sentiment analysis. We start with comparing both methods and find an 84.7% match in the identities of negative and positive sentiments. Tables 2 and 3 present the results of the comparison.

Table 2.
The comparison of two sentiment analysis methods in the number of tweets

Number of tweets		*SentiStrength polarity classification*		
		Negative	Neutral and Positive	Total
Liu and Hu classification	Negative	5,538	3,928	9,466
	Neutral and Positive	3,342	34,764	38,106
	Total	8,880	38,692	47,572

Table 3.
The comparison of two sentiment analysis methods in the percentage of tweets

Percentage of tweets		*SentiStrength polarity classification, %*		
		Negative	Neutral and Positive	Total
Liu and Hu classification	Negative	11.64	8.26	19.90
	Neutral and Positive	7.03	73.08	80.10
	Total	18.67	81.33	100

We then perform the validity test of our measures by comparing the sentiment classifications based on SentiStrength and Liu and Hu dictionaries to the benchmark of human coding. A random sample of 200 tweets was coded by hand to test the accuracy of the applied methods. The hand coding divided tweets into two groups: "0" when the tweets were non-negative, and "1" when the meaning of the tweets was negative toward Hungary as a country. We treated a tweet as negative if it had negative words in it (See Appendix for the human-coded table).

Tables 4 and 5 contain the results of the comparisons. Table 4 compares human coding to the Liu and Hu sentiment classification. 77 percent of tweets are classified correctly, with 16 percent of tweets being incorrectly labeled as neutral or positive, and 7 percent of tweets being incorrectly labeled negative by the machine.

Table 4.

The comparison of human coding to the Liu and Hu classification

Percent of tweets		*Liu and Hu classification*		
		Negative	Neutral and Positive	Total
Hand-coded classification	Negative	14.5	16	30.5
	Neutral and Positive	7	62.5	69.5
	Total	21.5	78.5	100

Table 5 compares human coding to the SentiStrength polarity classification, where 77.5 percent of tweets are classified correctly, with 11.5 percent of tweets being incorrectly labeled as neutral or positive, and 11 percent of tweets being incorrectly labeled negative by the machine. We conclude that both methods perform reasonably well. Given that SentiStrength is designed for the analysis of social media data, we

proceed with this method. In the dataset of 200 randomly selected tweets, 23 tweets that are likely to be published by news media agencies are not removed from the analysis. The results do not change significantly if the analysis of 177 tweets is conducted. However, these 23 media tweets are more likely to be negative than the rest of the tweets.

As we can see in Table 5, a lower number of 11.5 percent of negative tweets that was recognized as neutral and positive justifies the usage of SentiStrength in this research.

Table 5.

The comparison of human coding to the SentiStrength polarity classification

Percent of tweets		*SentiStrength polarity classification*		
		Negative	Neutral and Positive	Total
Hand-coded classification	Negative	19.5	11.5	30.5
	Neutral and Positive	11	58	69.5
	Total	30.5	69.5	100

Twitter Sentiment Word Clouds

We begin with the initial description of the data. Figure 1 presents the clouds of words, which are labeled as negative in the SentiStrength database for the periods of before, during and after the crisis for Hungary and Romania. In the resulting clouds, the words ‘refuge’ and ‘crisis’ have dominated the amount of negative sentiment and have taken the majority of space in the second and third periods, so we deleted both words in the analysis and added the word ‘refuge’ for the sake of demonstration in Figure 2.

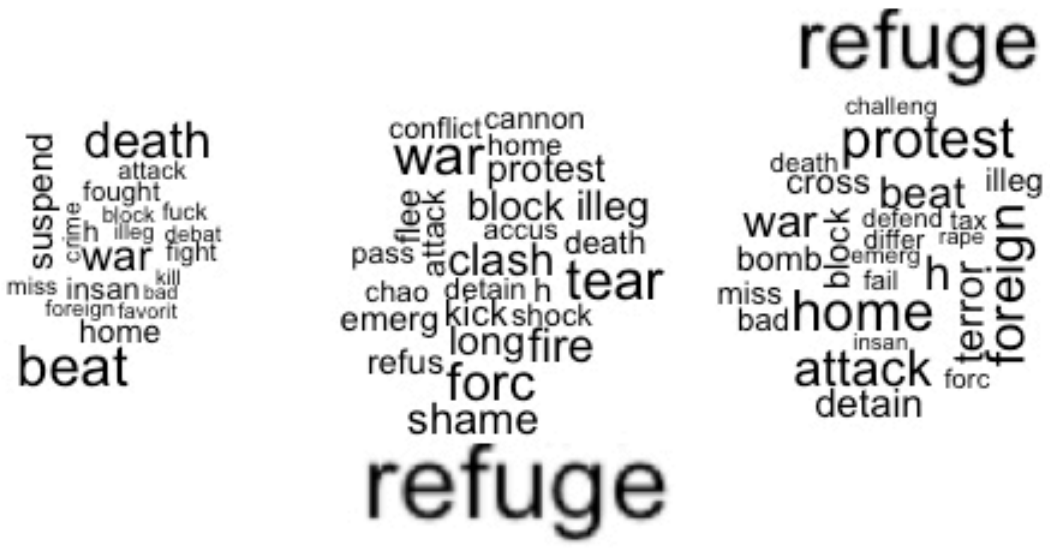

Figure 2. A word cloud from the Twitter dataset about Hungary in three periods (ordered as before, during, and after the crisis)

As we can see from a representation of the most popular words in tweets about Hungary in Figure 2, the vocabulary has changed significantly throughout the period from March 2015 to March 2016. Before the crisis, the most popular words were ‘suspend’, ‘death’, ‘beat’, etc. Then, starting from June 2015, the lexicon shifted toward the words associated with the 2015 refugee crisis (‘forc’, ‘tear’, ‘clash’) with the word ‘refuge’ as the most used and prominent during the period of the crisis. After that, in November 2015, the lexicon of tweets about Hungary shifted toward a mixture of two previous periods: the original vocabulary (‘death’, ‘war’, ‘beat’, ‘terror’), and the vocabulary about the refugee crisis period (‘refuge’, ‘detain).

Figure 3. A word cloud from the Twitter dataset about Romania in three periods (ordered as before, during, and after the crisis)

On the other hand, the lexicon of tweets about Romania remained more constant compared to Hungary. For example, the topic of 'corrupt' and 'home' remained prominent during the whole period of the study (See Figure 3). Additionally, during the crisis period the vocabulary in tweets about Romania has not shifted significantly towards expressing opinions about the refugees or migrant crisis. This finding suggests that the conversations on Twitter about refugees and migrants were more salient in Hungary compared to Romania.

Removal of Media Tweets

Scholars agree that there is a tendency of a higher negative sentiment in news content (Soroka et al, 2015, p.108). To calculate the amount of tweets posted by news agencies and media organizations, the sample dataset of tweets with metadata (usernames, user bios, and a number of followers) was collected in the period from March 14 to March 17, 2016. In the sample dataset, the estimate of the percentage of media tweets is 11.6 percent. In the bigger dataset, 12.3 percent of media tweets were identified as media-related and removed for further analysis.

Hashtag Identification

Below is the list of the twenty most popular hashtags about Hungary and Romania in the period from March 2015 to March 2016, with the respective number of tweets.

Table 6.
Top 20 hashtags about Hungary and Romania

#Hungary	#Romania	#romania	#hungary	#refugees	#Greece	#EU	#travel	#stiri
4565	3033	1760	1036	562	476	471	459	432
#Budapest	#Poland	#Russia	#Bahrain	#1	#Slovakia	#Ukraine	#Finland	#budapest
429	428	404	399	385	368	365	352	343
#Serbia	#Czech							
340	330							

Regression Description

To measure the effect of the 2015 European refugee crisis on the reputation of Hungary, which is the dependent variable y measured by sentiment analysis of the tweets, the research compares Hungary and Romania, with Hungary facing negative media coverage in the crisis period. We distinguish between the short-term and long-term effect of the crisis. We define the short-term effect as the difference in a country image before and during the crisis, and the long-term effect as the difference in a country image before and after the crisis. To estimate both effects of the refugee crisis on the image of Hungary, the research uses a difference-in-differences model with the logistic specification:

$$Y^{*}{}_{ijt} = \beta_0 + \beta_1{}^{*}Country_j + \beta_2{}^{*}during_{crisis_t} + \beta_3{}^{*}Country_j{}^{*}during_{crisis_t} + \beta_4{}^{*}after_{crisis_t} + \beta_5{}^{*}Country_j{}^{*}after_{crisis_t} + \beta_6{}^{*}\ Romania\_protests_{jt} + \epsilon_{ijt}$$

$$\epsilon_{ijt} \sim Logistic(0,1)$$

$$\mathrm{Y_{ijt}} = 1 \text{ if } \quad Y^{*}_{ijt} > 0$$

$$\mathrm{Y_{ijt}} = 0 \text{ if } \quad Y^{*}_{ijt} \leq 0$$

In this formula:

- Y_i - the indicator variable for the individual tweet i about country j on day t to be negative;
- $during\_crisis_t$ - the indicator variable representing crisis;
- $after\_crisis_t$ is an indicator variable denoting post-crisis;
- $Country_j$ is a dummy variable (Hungary = 1 and Romania =0) for the country; $Romania\_protests_{jt}$ is an indicator variable denoting protests in Romania in 2015,
- β_2 - the coefficient that estimates the differences between the treatment and control groups before the crisis.

There are two coefficients of interest in the model:

- β_3 is the coefficient on the interaction term, $Country_j{*}during\_crisis_t$ captures the short-term effect of the crisis on Hungary's image;
- β_5 is the coefficient on the interaction term $Country_j{*}after\_crisis_t$, captures the long-term effect of the crisis.

Regression Estimates

The regression estimates are below in the table, in which:

- Hungary – a logical variable, which is true when a tweet is about Hungary;
- During_crisis – a logical variable, which is TRUE during the crisis;
- After_crisis – a logical variable, which is TRUE after the crisis;
- Romania:protests – a logical variable, which is TRUE for the protest period in Romania from October 30 to November 9, 2015.

Table 7. The results of regression using SentiStrength by a single tweet		
Predictor	*B*	*SE B*
Intercent (constant)	-1.5808***	0.0346
Hungary	0.1674**	0.0529
during_crisis	0.0152	0.0481
after_crisis	0.0502	0.0473
Hungary:during_crisis	1.0801***	0.0651
Hungary:after_crisis	0.1402*	0.0689
Romania:protests	1.1235***	0.0792
Signif. Codes: *p<.05, **p<.01, ***p<.001		

We used logistic regression to estimate the difference-in-differences model. Table 7 presents the estimates. Regression reveals that: 1) tweets about Hungary are more negative than tweets about Romania; 2) tweets about Hungary have become more negative during the crisis compared to the tweets about Romania in the same period; 3) tweets about Hungary remained slightly more negative after the crisis compared to the tweets about Romania in the same period; 4) tweets about Romania during the protest of November 2015 are significantly more negative than other tweets about Romania after the crisis. Results 2 and 3 are the most important: these are the difference-in-differences coefficients, which show the increase in the level of negative sentiment in tweets about Hungary compared to the level of negative sentiment in tweets about Romania during and after the crisis period.

To interpret the magnitude of these results, we compute a predicted probability based on the model estimates. We can compute the probability of negative sentiment in tweets by plugging in the regression estimates to the logistic model:

$$\widehat{\Pr}(tweet\ is\ negative) = \frac{\exp\left(\sum_l \widehat{\beta_l} X_l\right)}{1 + \exp\left(\sum_l \widehat{\beta_l} X_l\right)}$$

In this formula, β_l is the regression estimate, and X_l is the independent variable l. We can compute the difference in predicted probability during the period of crisis caused by the events associated with the 2015 refugee crisis:

$$\frac{\exp(-1.58 + 0.17 + 0.02 + 1.07)}{1 + \exp(-1.58 + 0.17 + 0.02 + 1.07)} - \frac{\exp(-1.58 + 0.17 + 0.02)}{1 + \exp(-1.58 + 0.17 + 0.02)}$$
$$= 0.4207 - 0.1994 = 0.2213$$

This implies that the refugee crisis has increased the average probability of a negative tweet about Hungary in the period of crisis by 22.13 percentage points.

We now examine the effect of the refugee crisis on the average probability of a negative tweet about Hungary in the after crisis period:

$$\frac{\exp(-1.58 + 0.17 + 0.05 + 0.14)}{1 + \exp(-1.58 + 0.17 + 0.05 + 0.14)} - \frac{\exp(-1.58 + 0.17 + 0.05)}{1 + \exp(-1.58 + 0.17 + 0.05)}$$
$$= 0.2279 - 0.2042 = 0.0237$$

This implies that the refugee crisis has increased the average probability of a negative tweet about Hungary in the after crisis period by 2.37 percentage points.

Figure 4. Overall negative sentiment in tweets

To aid interpretation of the regression results, we plot the average weekly share of tweets classified as negative for Hungary and Romania in Figure 4. We can see a large spike in the number of negative tweets about Hungary during the period of the crisis. The magnitude of the spike suggests a larger effect of the crisis on the probability of tweets to be negative, with a probability of negative tweets increased up to 50 percent during the crisis weeks. The model somewhat underestimated the increase in negative sentiment because the time frame of the crisis is elongated in the difference-in-differences specification.

We find a significant increase in the number of negative tweets for Hungary after the crisis. It suggests that the effect of the refugee crisis lasted for the studied period from November 2015 to March 2016.

Sentiment of Tweets in Topics Unrelated to the Refugee Crisis

We have shown that there was an increase in the amount of negative sentiment in tweets about Hungary during the refugee crisis. At the same time, we do not know if this increase in the negative sentiment was only due to the tweets about the refugee crisis, or there was a spillover effect on the sentiment in tweets that are not related to the refugee crisis. The following analysis aims to isolate tweets related to the crisis alone.

Conducting Latent Dirichlet Allocation

To divide the massive amount of tweets into groups on specific topics, the Latent Dirichlet Analysis (LDA) was applied. We estimate LDA with the hierarchical Bayesian procedure. The posterior distribution is approximated with the draws from the collapsed Gibbs sampler. Since we do not have users' metadata in the main dataset, we wanted to calculate the relative frequency of words to identify the distribution of topics in tweets.

To divide tweets by topic, we first deleted the words that matched with the dictionary of negative and positive words, and then proceeded to LDA. We did this so that negative and positive tweets are not allocated together just by the construction.

To decide what number of topics is optimal for the analysis, we need to compare the model fit. We estimate the model with 2 to 10 groups and compute the average posterior likelihood. We use the standard parameters for prior distributions suggested by Blei et al (2003). By construction, the average posterior likelihood increases as we increase the number of model components. To penalize the model for additional variables, we compute Akaike information criterion (AIC) and Bayesian Information Criterion (BIC), which are the information criteria that adjust the likelihood.

AIC helps minimize the estimated information loss. The lowest AIC is for the model with 3 segments, and the lowest BIC is for the model with 2 segments – it is because BIC penalizes the model more for extra parameters. Given that the number of parameters in the LDA model rapidly increases with the increase in the number of segments used, BIC may overpenalize the model. Thus, we proceed with AIC criteria. The estimates of log-likelihood (ln(L)) across models with the different number of segments are in the Figure 5.

AIC = 2 * k – 2 * ln(L), where k is the number of parameters in the model. LDA estimates n_topics * (n_words+1) parameters.

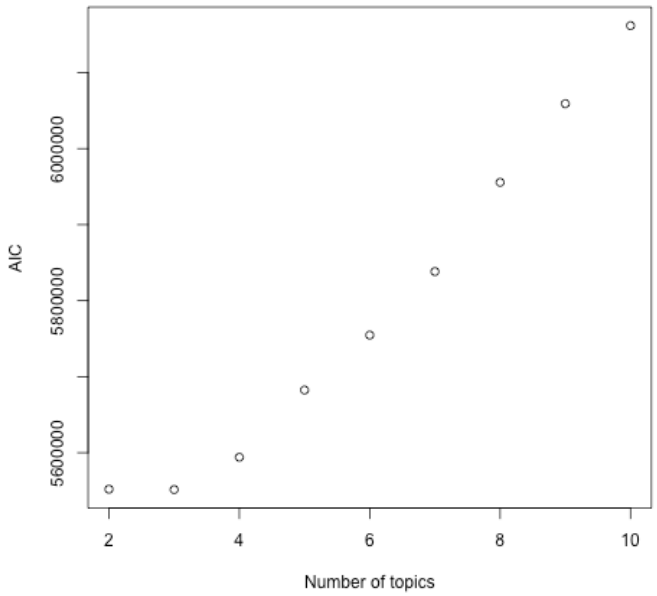

Figure 5. Results of AIC

Figure 4 shows BIC = -2ln(L) + log(n_observations) * k, where k again are parameters.

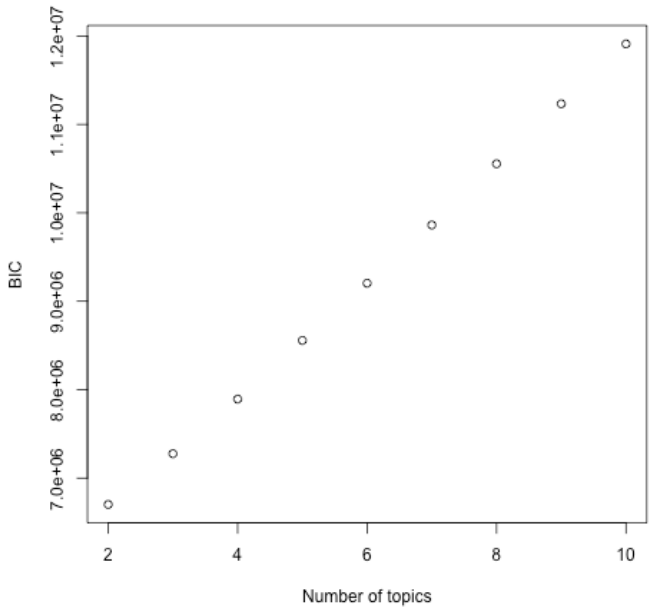

Figure 6. Results of BIC

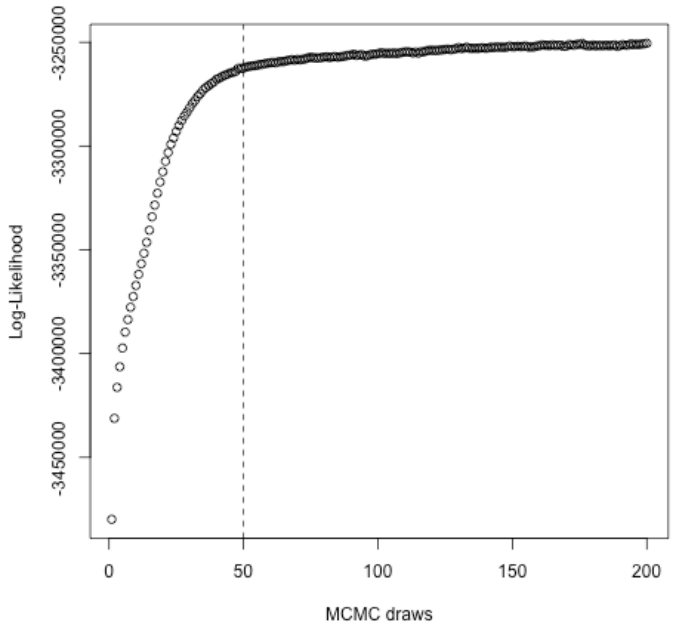

Figure 7. Convergence of log likelihood by MCMC draws

The optimal percentage of mutually exclusive topics discovered by LDA for the given dataset is three.

Table 8.
The percentage of tweets per topic

Tweet group number and name	Percentage of tweets
1. Tourism	23%
2. Refugee crisis	41%
3. Miscellaneous	36%

Table 9.
The list of most popular words in three identified topics

```
     [,1]       [,2]        [,3]
 [1,] "de"       "refuge"    "budapest"
 [2,] "la"       "migrant"   "v"
 [3,] "un"       "border"    "travel"
 [4,] "si"       "eu"        "world"
 [5,] "finland"  "austria"   "budapesthungari"
 [6,] "now"      "polic"     "go"
 [7,] "slovakia" "europ"     "f"
 [8,] "ukrain"   "train"     "ireland"
 [9,] "czech"    "croatia"   "euro"
[10,] "e"        "sai"       "bucharest"
[11,] "russia"   "fenc"      "bahrain"
[12,] "belaru"   "http"      "pleas"
[13,] "sa"       "close"     "photo"
[14,] "poland"   "crisi"     "get"
[15,] "pe"       "serbia"    "dai"
[16,] "plu"      "hungarian" "just"
[17,] "mai"      "germani"   "last"
[18,] "cu"       "pm"        "first"
[19,] "canada"   "countri"   "land"
[20,] "pentru"   "syrian"    "kleur"
```

Table 9 presents the most common words characterizing the resulting topics. The first group of words is related to miscellaneous issues, sports championships and cultural contests such as the Eurovision Song Contest. The second group of words is about refugees and the migrant crisis. The third group is primarily about tourism in Hungary and Romania.

Figures 8 to 11 plot the volume and relative shares of tweets of these three topics over time for Hungary and Romania. Figure 9 shows that during the time of the crisis, tweets about the refugees corresponded to around 80-85 percent of all tweets about Hungary. There is no such spike in the reporting about Romania, with tweets about refugees having been the least tweeted topic.

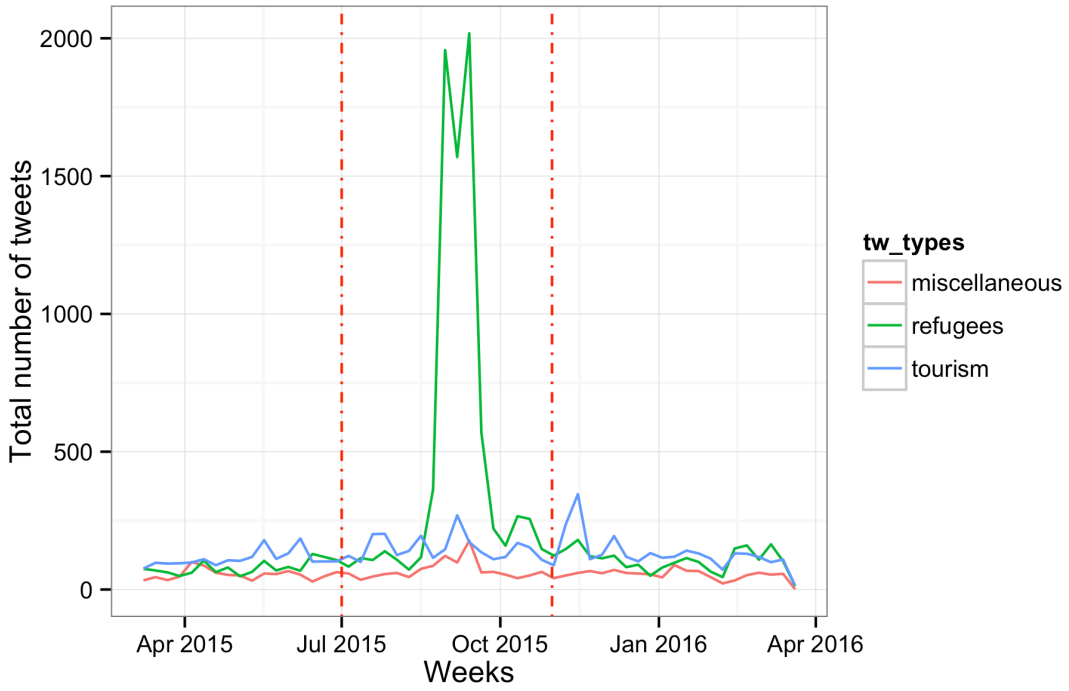

Figure 8. Volume of tweets across different topics, Hungary

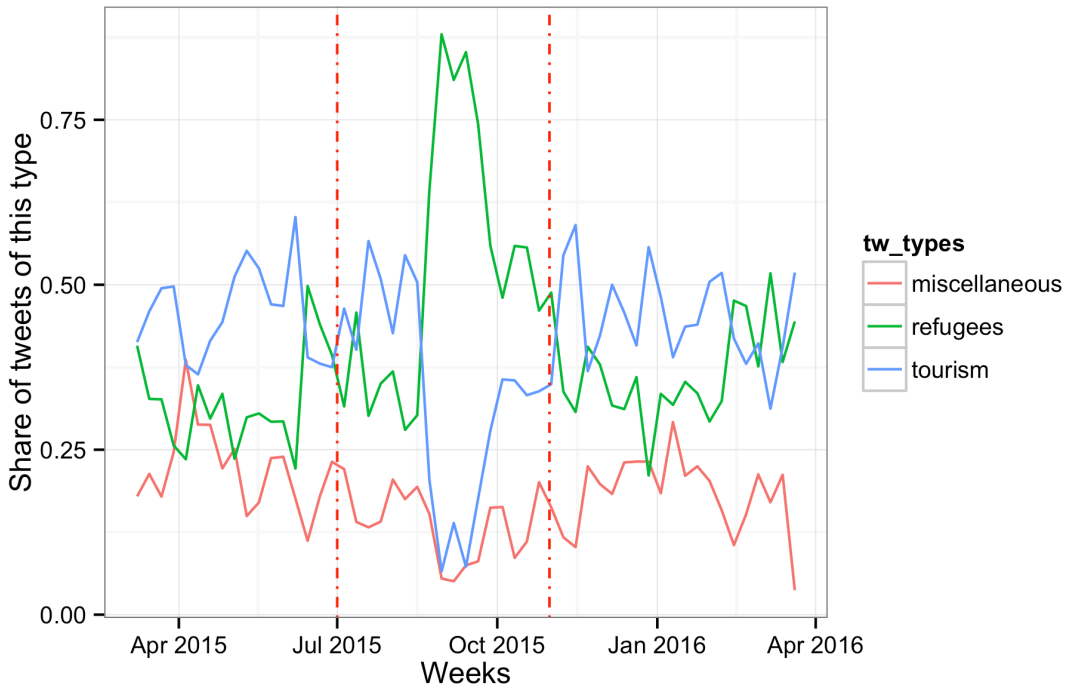

Figure 9. Share of tweets across all 3 groups in Hungary

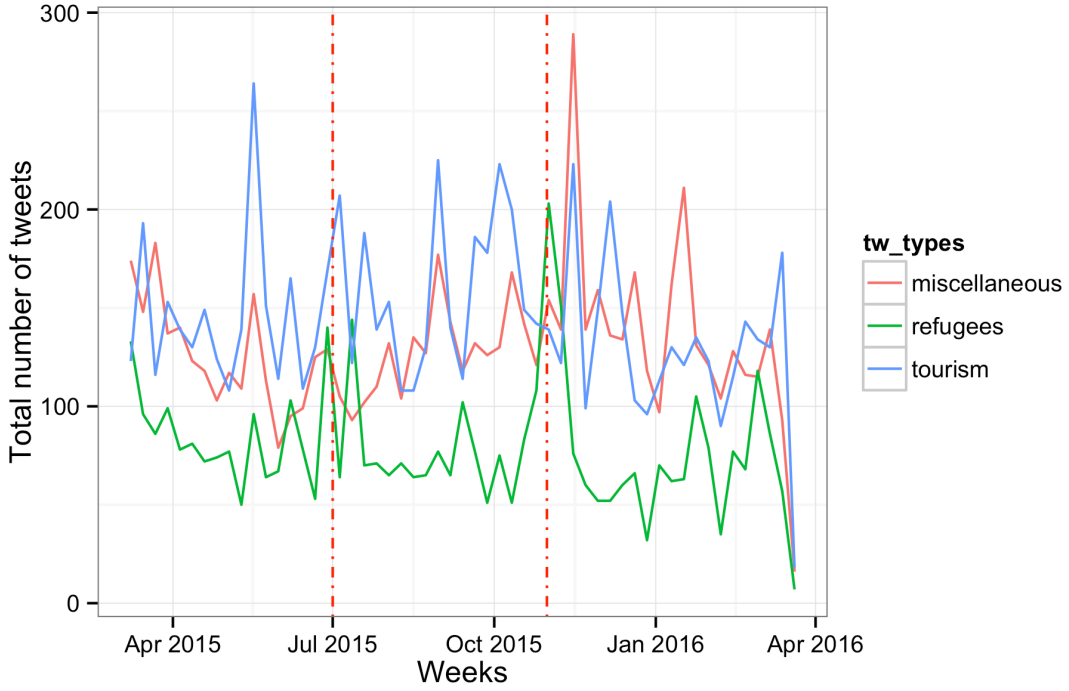

Figure 10. Volume of tweets across different topics, Romania

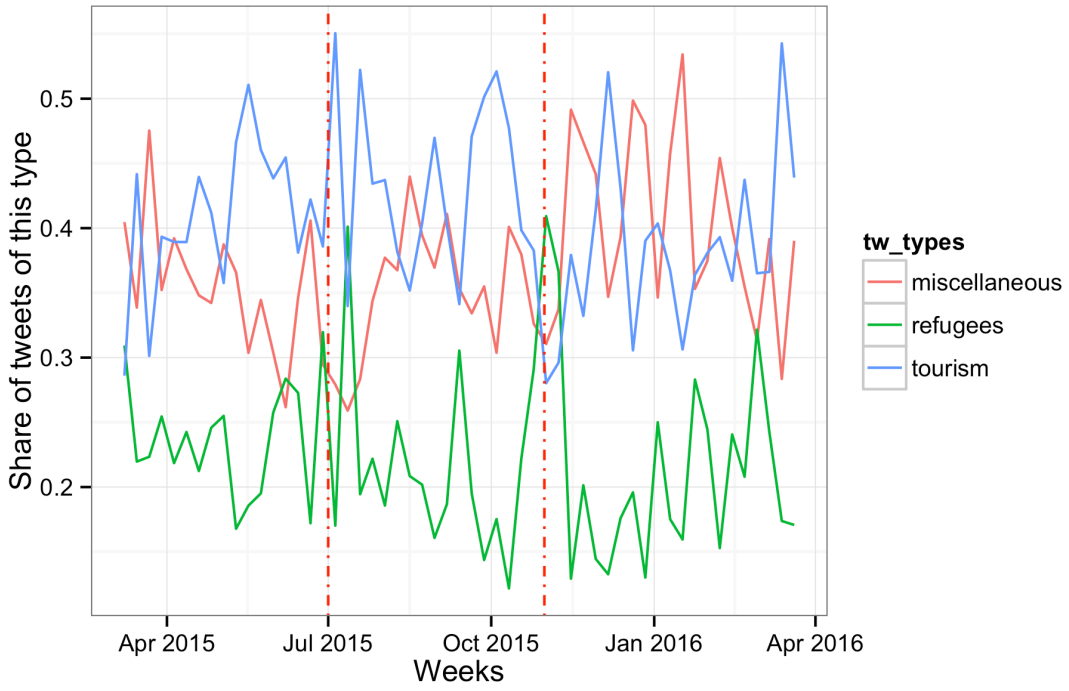

Figure 11. Share of tweets in all 3 topics in Romania

We now examine the changes in the level of negative sentiments for tweets classified as 'tourism' (*topic* 1) and 'miscellaneous' (*topic* 3).

Negative Sentiment by LDA Groups

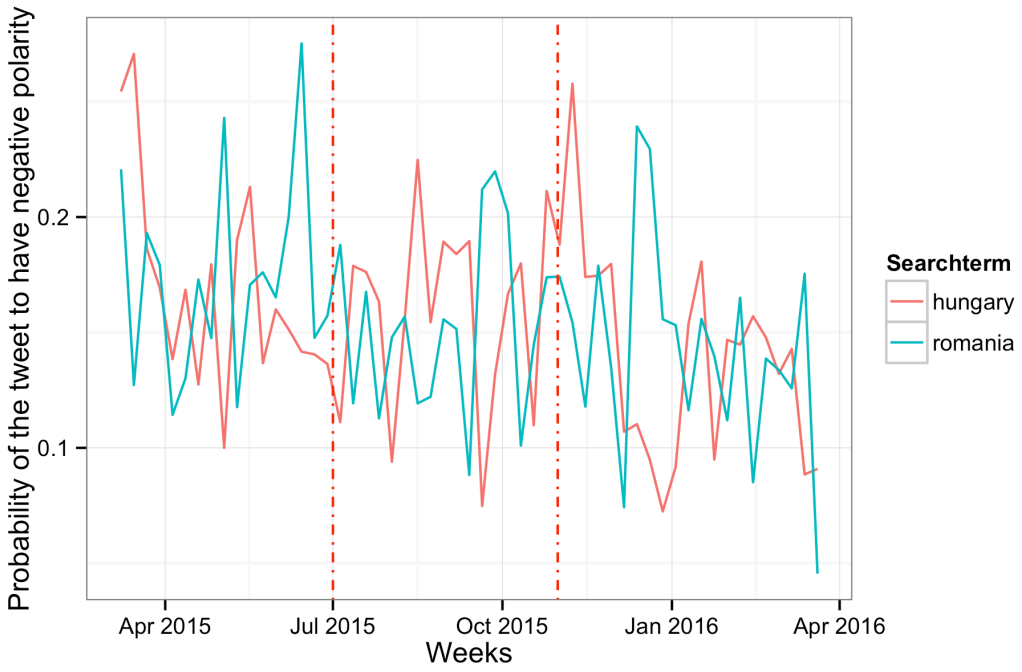

Figure 12. Negative sentiment in 1st topic ('tourism')

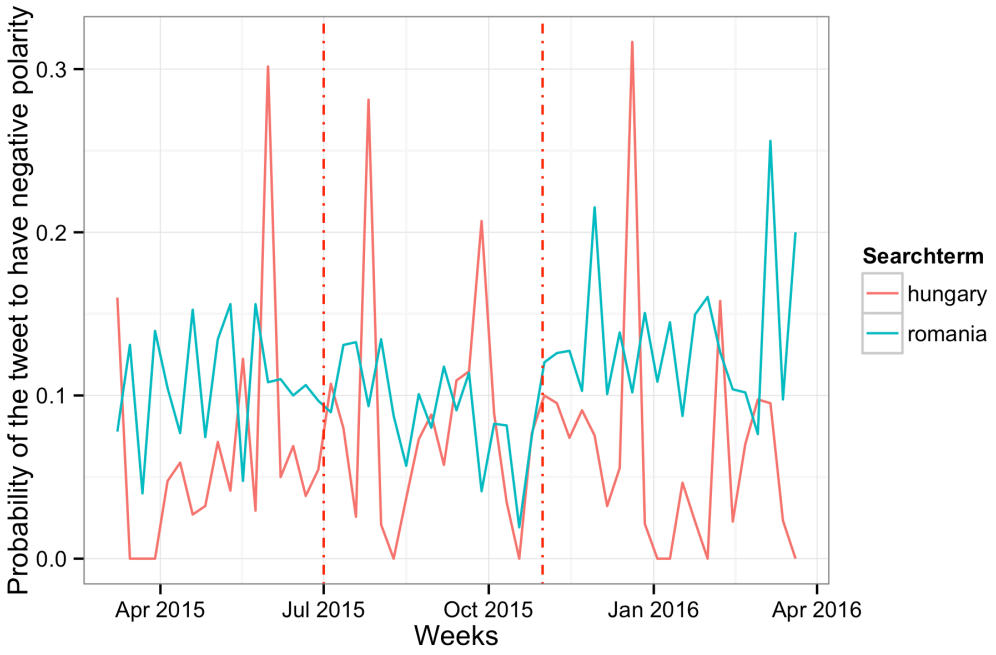

Figure 13. Negative sentiment in 3rd topic ('miscellaneous')

We do not see a spike in negative sentiment during and after the crisis in the 1st and 3st topics. Once again, we ran separate difference-in-differences regressions based on tweets about topics 1 and 3; we do not find any significant increase in negative sentiment during or after the crisis. We take this as evidence that there was no increase in the negative sentiments in topics other than the refugee crisis (topic 2).

Finally, we examine the change in the sentiment in the refugee crisis topic separately. Figure 14 plots the average sentiment for this topic over time.

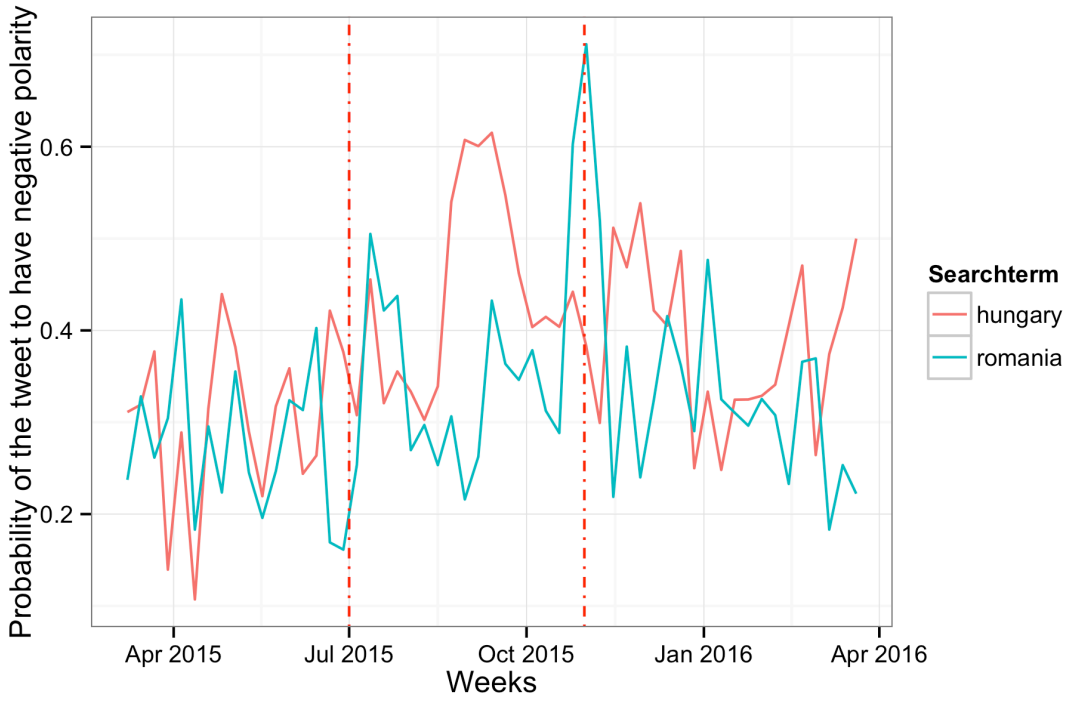

Figure 14. Negative sentiment in 2nd topic ('refugee crisis')

As expected, there is a spike in negative sentiment during the crisis in the 2nd group and a smaller increase in the post-crisis period. We ran a difference-in-difference regression based only on tweets about topic 2. Results are presented in Table 11 below.

Table 10. The results of regression using SentiStrength by a single tweet (the refugee crisis topic)		
Predictor	*B*	*SE B*
Intercent (constant)	-0.9569***	0.0611
Hungary	0.1247	0.0871
during_crisis	0.2368**	0.0825
after_crisis	0.2040*	0.0836
Hungary:during_crisis	0.8190***	0.1056
Hungary:after_crisis	0.1719	0.1126
Romania:protests	1.8843***	0.1410
Signif. Codes: *p<.05, **p<.01, ***p<.001		

From the regression estimates, we can confirm a significant increase in the amount of negative sentiment for Hungary during the crisis. However, we do not find a significant increase in the amount of negative sentiment for Hungary after the crisis.

DISCUSSION

Place brands change because the reality in which they exist evolves and places "project their images to their audiences" (Kavoura, 2013, p.29). We have examined the sentiments in the tweets about Hungary to estimate the effect of the 2015 refugee crisis on the place brand of Hungary on Twitter. We have started with computing the overall sentiment in the tweets across time. In this case, we see a large increase in the amount of negative sentiment in tweets in the period of the 2015 refugee crisis. There is a slight increase in the amount of negative sentiment in tweets right after the crisis, which quickly dissolves after one month. This suggests that the refugee crisis had a small carry-over effect on the conversations on Twitter that followed the crisis, yet not for long.

We classified the tweets into topics by applying a Latent Dirichlet Allocation method (LDA), which is the classification algorithm based on unsupervised learning. Using AIC and BIC criteria, we specify 3 groups of topics: tourism, refugee crisis, and miscellaneous (such as sports and events). We examined the change in the amount of negative sentiment for each of these topics. We do not find any increase in the amount of negative sentiment for tourism and miscellaneous topics, which supports our hypothesis that negative sentiment in the tweets has to do with the migration crisis. We do find an increase in the amount of negative sentiment in the tweets about the refugee crisis in Hungary compared to Romania in the same time period.

All in all, our analysis suggests that the refugee crisis had a limited effect on the place brand of Hungary as measured by sentiments in the tweets. We find a statistically significant effect of the refugee crisis on the sentiment of tweets in the period after the crisis. Although statistically significant, the magnitude of the effect is only 2 percentage

points, which suggests a limited role of the refugee crisis in the place brand of Hungary. The plot of the average probability of the negative sentiment also reveals that this increase is more visible in the month following the crisis. All of these suggest that there is a small long-term effect on the place brand of Hungary. However, future research is required to measure this effect more precisely.

Implications

What happens after a country undergoes intensive negative coverage due to a crisis? Research suggests the consequences depend on how the country has dealt with the crisis (Govers et al, 2012). In the case of international place brand of Hungary, we can see that the spike in negative sentiment on Twitter has not persisted after the crisis events past November 2015. We might further see the consequences by comparing the number of tourists in the summer of 2016 with the previous year's records. In 2015, a record number of 10 million international tourists visited Hungary, constituting a 5 percent total contribution to the country's GDP (World Travel Tourism Council, 2015, p.11). However, it is hard to argue that the change in the number of tourists is caused by the events of the refugee crisis and the consequent reaction of the Hungarian government to it. There could be multiple reasons why tourists decided not to travel to Hungary, such as a budget spent on the promotion of tourism activities in Hungary, overall economic

situation in the region, or people's choices to visit another country this time.

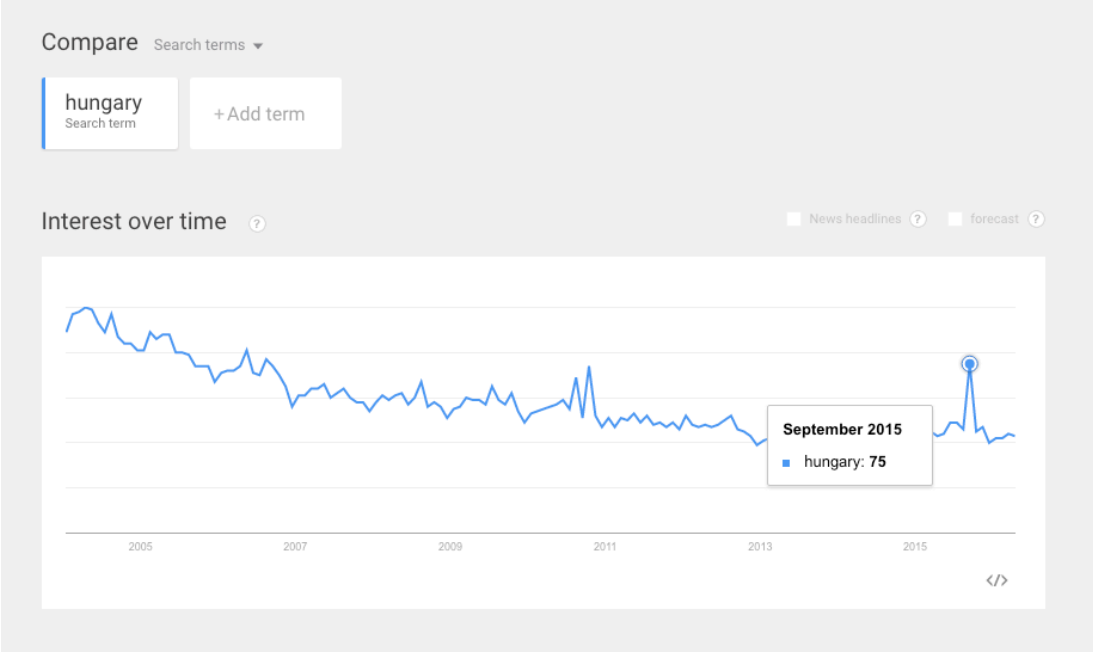

Figure 15. Overall trend in search of 'Hungary' on Google as of April 2016.
Source: Google Trends

In addition, in Figure 15, a search trend of 'Hungary' on Google shows a spike in the search during September 2015, with the latest spike in September 2010, when another crisis, the Ajkai alumina plant toxic sludge, hit Hungary last time (Tran, 2010). The graph suggests a relatively stable and even decreasing search of 'Hungary' on the Google search engine, besides the spikes during these two crisis occasions.

Limitations

The computational analysis of sentiment in social media is still an evolving discipline. One of the limitations of social media analysis is a pre-existing bias in the population sample of social media users through their socioeconomic factors, such as age, education, computing skills, and income (Hargittai, 2015, p.63). The users of social media, such as Facebook and Twitter, tend to be of a younger age (Hargittai, 2015, p.64). In addition, Twitter users are more socially connected; therefore, a use of Twitter is

limited to a number of people who actively engage with this social medium (Hawthorne et al, 2013).

To support the results of this study and to address its shortcomings, a further data triangulation is needed by finding a bigger dataset that includes user bios, a number of followers, locations; data from other social networks, such as Facebook; and surveys of Twitter users.

Another limitation is the timeline of this research – from March 2015 to March 2016. Although the paper addresses the 2015 refugee crisis, this crisis is still ongoing as of April 2016. In the beginning of April 2016, the Hungarian government rejected the European Commission's draft plan to allocate refugees among all 28 countries of the European Union (Simon, 2016). The Prime Minister of Hungary Victor Orban advocates for the plan to "defend Europe"; the outline is to be presented in the second half of 2016. Mr. Orban's refusal to accommodate primarily Muslim refugees is delivered as a protection of "Europe's Christian identity and heritage" (Dnaindia.com, 2016). It would be beneficial for the fields of place branding and crisis communication to address these limitations in further research.

Future Research

In addition to the statement above, it would be interesting to apply this research to other countries that are actively involved in the ongoing refugee crisis, such as Germany and Greece. However, there is a risk to face many externalities because media outlets usually cover Greece and Germany more actively than either Romania or Hungary – so it would be somewhat intricate to estimate the exact effect on their place brands and reputation.

Also, it would be valuable to compare the effect of the 2015 refugee crisis on changes in the image of the country among local population of Hungary and Romania as well as a sample of international audience. It could be done through interviews with these group of people.

REFERENCES

Aaker, D. (1991). Managing brand equity: capitalizing on the value of a brand name. Toronto: Maxwell Macmillan Canada.

Aaker, D., & Joachimsthaler, E. (2000). Brand leadership. The Free Press, New York, p.17.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov and F. Csaki, editors, Proceedings of 2nd International Symposium on Information Theory, 267–281. Akademiai Kiado.

Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control 19 (6): 716–723. *doi*:*10.1109/TAC.1974.1100705*.

Andéhn, M., Kazeminia, A., Lucarelli, A., & Sevin, E. (2014). User-generated place brand equity on twitter: The dynamics of brand associations in social media. *Place Branding and Public Diplomacy, 10*(2), 132-144. doi:http://dx.doi.org/10.1057/pb.2014.8

Angrist, J. D. (2004). American Education Research Changes Tack. *Oxford Review of Economic Policy*. 20 (2), 198-212. doi: 10.1093/oxrep/grh011

Anholt, S. (2009). The media and national image. *Place Branding and Public Diplomacy, 5*(3), 169-179. doi:http://dx.doi.org/10.1057/pb.2009.11

Benoit, W. (1997). Image repair discourse and crisis communication. *Public Relations Review*, 23(2), 177-186.

Blei, D.M., Ng, A. Y., Jordan, M. I. (2003). Latent Dirichlet Allocation. ". *Journal of Machine Learning Research* 3 (4–5): *pp.* 993–1022. doi:10.1162/jmlr.2003.3.4-5.993.

Buchanan, R.T. (2015, October 22). Hungarian camerawoman plans to sue Facebook and refugee she was filmed tripping on Serbian border. *The Independent*. Retrieved from http://www.independent.co.uk/news/world/europe/hungarian-camerawoman-plans-to-sue-facebook-and-refugee-she-was-filmed-tripping-on-serbian-border-a6702336.html

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. doi:10.1371/journal.pone.0014118

Coombs, T.W. (2007). Attribution Theory as a guide for post-crisis communication research. *Public Relations Review,* 33, 135–139.

Coombs, T.W. (2007). Protecting organization reputations during a crisis: the development and application of situational crisis communication theory. *Corporate Reputation Review*, 10 (3), 163-176.

Cresswell, J. W. (2014). Research design: qualitative, quantitative, and mixed methods approaches. Thousand Oaks, Calif.: Sage Publications, 4 ed.

Die Welt (2015, September 22). EU immigration quota plan map. Wikipedia. Retrieved from https://en.wikipedia.org/wiki/European_migrant_crisis#/media/File:2015-09-22_EU_JHA_Council_majority_vote_to_relocate_120,000_refugees.svg

Dnaindia.com. (2016, February 29). Hungarian PM Orban vows to resist EU's 'misguided' refugee policy. *DNA India.* Retrieved from http://www.dnaindia.com/world/report-hungarian-pm-orban-vows-to-resist-eu-s-misguided-refugee-policy-2183626

Dynarski, S. M. (2003). Does Aid Matter? Measuring the Effect of Student Aid on College Attendance and Completion. *The American Economic Review*, *93*(1), 279–288. Retrieved from http://www.jstor.org/stable/3132174

Engert, M. [marcusengert]. (2015, September 3). #Hungary breaks elementary European rules, Again. Tie to discuss a #Hungxit, hm? #refugeecrisis #Budapest #Keleti #magyarorszag #refugees [Tweet]. Retrieved from https://twitter.com/marcusengert/status/639504300889231360

European Travel Commission and World Tourism Organization. (2014). Handbook on E-marketing for tourism destination – Fully revised and extended version 3.0, UNWTO, Madrid.

Frombrun, C., & Shanley, M. (1990). What's in name? Reputation, building and corporate strategy. *Academy of Management Journal*, 33 (1), 233-258.

Fry, L. (2015, September 15). Refugee crisis timeline: How the crisis has grown. *The Independent.* Retrieved on October 26, 2015 from http://www.independent.co.uk/news/world/europe/refugee-crisis-timeline-how-the-crisis-has-grown-10502690.html

Gera, V. (2015, September 17). Human rights group demands Hungary stop harassing reporters. *The Associated Press.* Retrieved from http://bigstory.ap.org/article/fdb7a8b6e205418fbe5cd7eae1093da2/human-rights-group-demands-hungary-stop-harassing-reporters

Go, F.M., & Govers, R. (Eds). (2012). International Place Branding Yearbook 2012. Palgrave Macmillan.

Goodcountry.org. (2015). The Good Country Index. Retrieved from http://www.goodcountry.org/overall

Gorondi, P. (2015, June 19). Hungary's leader says Hungary will protect Europe's borders. *The Associated Press.* Retrieved on October 26, 2015 from https://web.archive.org/web/20150621171828/http://hosted.ap.org/dynamic/stories/E/EU_HUNGARY_BORDER_FENCE

Govers, R. and Go, F.M. (2009). Place Branding: Glocal, Virtual and Physical, Identities Constructed, Imagined and Experienced. Basingstoke, UK: Palgrave Macmillan.

Hargittai, E. (2015). Is bigger always better? Potential Biases of Big Data Derived from Social Network Sites. ANNALS, *AAPSS,* 659, pp. 63-76. DOI: 10.1177/0002716215570866

Hawthorne, J., Houston, J. B., & McKinney, M. S. (2013). Live-tweeting a presidential primary debate: Exploring new political conversations. *Social Science Computer Review*, *31*, 552–562. http://doi.org/10.1177/0894439313490643

Hu, X., Tang. J., Gao, H., & Liu, H. (2013). Unsupervised sentiment analysis with emotional signals. In Proceedings of the International World Wide Web Conference, pp. 607–618.

Indexmundi. (2015). Romania vs. Hungary. Retrieved on October 28, 2015 from http://www.indexmundi.com/factbook/compare/romania.hungary

Jahn, G., & Cook, L. (2015, June 24). Hungary to build fence on Serbian border as fast as possible. *The Associated Press.* Retrieved from http://www.timesfreepress.com/news/national/international/story/2015/jun/24/hungary-build-fence-serbiborder-fast-possible/311245/

Jiang, L., Yu, M., Zhou, M., Liu, X., & Zhao, T. (2011). Target-dependent twitter sentiment classification. The Proceeding of Annual Meeting of the Association for Computational Linguistics, 1:151–160.

Kavoura, A. (2013). Contemporary Approaches of the Scientific Theory of Place Marketing-Place Branding in Globalized Conditions and Economic Crisis: Place Branding in Globalized Conditions & Economic Crisis. Nova Science Publishers, Incorporated.

Kawash, J. (2014). Online Social Media Analysis and Visualization. *Springer International Publishing*. doi: 10.1007/978-3-319-13590-8

Kerlinger, F. N. (2000). Foundations of behavioral research. 4th ed. New York, NY: Holt, Rinehart & Winston.

Last Week Tonight. (2015, September 28). Last Week Tonight with John Oliver: Migrants and Refugees (HBO) [Video file]. *Youtube.com.* Retrieved from https://www.youtube.com/watch?v=umqvYhb3wf4

Liu, B. (2012). *Sentiment analysis and opinion mining*. Chicago, IL: Morgan & Claypool.

Liu, B. (2015). Sentiment analysis. Cambridge University Press.

Lucarelli, A. (2012). Unraveling the complexity of "city brand equity": A three-dimensional framework. *Journal of Place Management and Development, 5*(3), 231-252. doi:http://dx.doi.org/10.1108/17538331211269648

Lucarelli, A., Brorstrom, S. (2013). Problematising place branding research: A meta-theoretical analysis of the literature. *Marketing Review*, 13(1), 65-81. doi: 10.1362/146934713X13590250137826

Lynch, K. (1960). The image of the city. Cambridge [Mass.]: Technology Press.

Lyon, L., Cameron, G.T. (2004). A relational approach examining the interplay of prior reputation and immediate response to a crisis. *Journal of Public Relations Research*, 2004, Volume 16, Issue 3.

Mackey, R. (2015, September 8). Hungarian Journalist, Petra Laszlo, Fired for Kicking Migrants on Camera. *The New York Times.* Retrieved from http://www.nytimes.com/2015/09/09/world/europe/hungarian-journalist-fired-after-abusing-migrants-on-camera.html

Morstatter, F., Pfeffer, J., Liu, H., & Carley, K. M. (2013). Is the sample good enough? comparing data from twitter's streaming API with twitter's firehose. *arXiv preprint arXiv:1306.5204*.

Murnane, R. (2011). Methods matter: improving causal inference in educational and social science research / Richard J. Murnane, John B. Willett. Oxford; New York: Oxford University Press, 2011.

Nolan, D. (2015, September 8). Hungarian nationalist TV camera operator filmed kicking refugee children. *The Guardian.* Retrieved from http://www.theguardian.com/world/2015/sep/08/hungarian-nationalist-tv-camera-operator-filmed-kicking-refugee-children

Paltoglou, G. & and Thelwall, M. (2012). Twitter, MySpace, Digg: unsupervised sentiment analysis in social media. *ACM Transactions on Intelligent Systems and Technology (TIST)* 3 (4): 1–19.

Pang, A., Jin, Y., Cameron, G.T. (2010). Contingency theory of Strategic Conflict Management: Directions for the practice of Crisis Communication from a Decade of Theory Development, Discovery, and Dialogue. *The Handbook of Crisis Communication*. Wiley, 768. DOI: 10.1002/9781444314885.ch26

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

Pang, B., and Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and trends in Information Retrieval. Vol. 2.

Schreiner, W., & Go, F. (2012). Social media: an insight into "public mood" of places? A case study of the city of Johanessburg. *International Place Branding Yearbook 2012, 119-130*. Palgrave Macmillan.

Seargeant, P., & Tagg, C. (Eds). (2014). The language of social media: identity and community on the Internet. Houndmills, Basingstoke, Hampshire: Palgrave Macmillan.

Sentistrength.com. (2016, April 10). Sentistrength. Retrieved from http://sentistrength.wlv.ac.uk/

Sevin, E. (2011). Thinking about place branding: Ethics of concept. *Place Branding and Public Diplomacy, 7*(3), 155-164. doi:http://dx.doi.org/10.1057/pb.2011.15

Sevin, E. (2013). Places going viral: Twitter usage patterns in destination marketing and place branding. *Journal of Place Management and Development, 6*(3), 227-239. doi:http://dx.doi.org/10.1108/JPMD-10-2012-0037

Shafranskaya, I., & Potapov, D. (2014). An empirical study of consumer-based city brand equity from signalling theory perspective. *Place Branding and Public Diplomacy, 10*(2), 117-131. doi:http://dx.doi.org/10.1057/pb.2014.7

Shin, J., Cameron, G.T., & Cropp, F. (2002). Asking what matters most: A national survey of public relations professional response to the contingency model. Miami, AEJMC.

Shin, J., Cameron, G.T., & Cropp, F. (2006). Occam's razor in the contingency theory: A national survey on 86 contingency variables. *Public Relations Review*, 32: 282-286. doi:10.1016/j.pubrev.2006.05.005

Shrivatava, A., Mayor, S., & Pant, B. (2014). Opinion mining of real time twitter tweets. *International Journal of Computer Applications, 100*(19). doi:http://dx.doi.org/10.5120/17630-0691

Simon, Z. (2016, April 8). Hungary Rejects EU Draft Plans on Refugee Policy, Orban Says. *Bloomberg News*. Retrieved from http://www.bloomberg.com/news/articles/2016-04-08/hungary-rejects-eu-draft-plans-on-refugee-policy-orban-says

Soroka, S., Young, L., & Balmas, M. (2015). Bad News or Mad News? Sentiment Scoring of Negativity, Fear, and Anger in News Content. *The ANNALS of the American Academy of Political and Social Science*, *659*(1), 108-121.

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, *6*(2), 461-464.

Tham, A., Croy G., & Mair, J. (2013). Social Media in Destination Choice: Distinctive Electronic Word-of-Mouth Dimensions. *Journal of Travel & Tourism Marketing*. Vol. 30, Issue 1-2. DOI: 10.1080/10548408.2013.751272

Tang, J., Zhang, M., & Mei, Q. (2014). " Look Ma, No Hands!" A Parameter-Free Topic Model. *arXiv preprint arXiv:1409.2993*.

The Guardian (2015, December 5). Hungary. *The Guardian.* Retrieved from http://www.theguardian.com/world/hungary

The Guardian (2015, December 5). Romania. *The Guardian.* Retrieved from http://www.theguardian.com/world/romania

The New York Times (2015, December 5). Romania. Chronology of coverage. Retrieved from http://topics.nytimes.com/top/news/international/countriesandterritories/hungary/index.html

The New York Times (2015, December 5). Romania. Retrieved from http://topics.nytimes.com/top/news/international/countriesandterritories/romania/index.html?8qa

Thelwall, M., Buckley, K., Paltoglou G., Cai D., & Kappas, A. (2010). Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology* 61 (12): 2544–58.

Thelwall, M., Buckley, K., & Paltoglou, G. (2012). Sentiment strength detection for the social web. *Journal of the American Society for Information Science and Technology*, *63*(1), 163-173.

Thelwall, M., & Buckley, K. (2013). Topic-based sentiment analysis for the social web: The role of mood and issue-related words. *Journal of the American Society for Information Science and Technology*, *64*(8), 1608-1617.

Tran, M. (2010, October 5). Hungary toxic sludge spill an 'ecological catastrophe' says government. *The Guardian*. Retrieved from http://www.theguardian.com/world/2015/nov/04/romanian-government-resigns-nightclub-fire-victor-ponta

Tran, M. (2015, November 4). Bucharest nightclub fire: PM and government resign after protests. *The Guardian.* Retrieved from http://www.theguardian.com/world/2015/nov/04/romanian-government-resigns-nightclub-fire-victor-ponta

Trust.org. (2015, October 28). Hungary PM rating jumps on handling of refugee crisis – Nezopont. *Trust.org.* Retrieved from http://www.trust.org/item/20151028201757-hj66d

UNHCR. (2014, June 20). World Refugee Day: Global forced displacement tops 50 million for first time in post-World War II era. *The UN Refugee Agency.* Retrieved on October 26, 2015 from http://www.unhcr.org/53a155bc6.html

UNHCR. (2015). Syria Regional Refugee Response. Inter-agency Informataion Sharing Portal. *The UN Refugee Agency.* Retrieved from http://data.unhcr.org/syrianrefugees/regional.php

United Nations, Department of Economic and Social Affairs, Population Division. (2014). World Urbanization Prospects: The 2014 Revision, Highlights (ST/ESA/SER.A/352).

Uysal, N., Schoeder, J., & Taylor, M. (2012). Social Media and Soft Power: Positioning Turkey's Image on Twitter. *Middle East Journal of Culture & Communication, 5(3), 338-359*. doi:10.1163/18739865-00503013

Varga, M. (2015, September 28). Hungarian tourism sector set to hit new highs in 2015. *Website of the Hungarian Government.* Retrieved from http://www.kormany.hu/en/ministry-for-national-economy/news/hungarian-tourism-sector-set-to-hit-new-highs-in-2015

Velcin, J., Brun, C., Dormagen, J., Kim, Y., Roux, C., Boyadjian, J., Bonnevay, J., Neihouser, M. (2012). Investigating the Image of Entities in Social Media: Dataset Design and First Results. ImagiWeb ANR-2012-CORD-002-01. 818-822.

World Travel Tourism Council. (2015). Travel & Tourism. Economic Impact 2015 Hungary. Retrieved from https://www.wttc.org/-/media/files/reports/economic%20impact%20research/countries%202015/hungary2015.pdf

Wilcox, D., Cameron, G., & Reber, B. (2015). Social Media Monitoring Tools. *Public Relations: Strategies and Tactics* (11th Ed.). Pearson Education, Inc. ISBN-10: 0-133-80937-4

Wimmer, R., & Dominick, J. (2014). Mass media research: an introduction, Tenth Edition. Australia: Cengage-Wadsworth.

Young, L., & Soroka, S. (2012). Affective News: The Automated Coding of Sentiment in Political Texts. *Political Communication*, *29*(2), 205–231. http://doi.org/10.1080/10584609.2012.671234

Zenker, S. (2014). Measuring place brand equity with the advanced brand concept map (aBCM) method. *Place Branding and Public Diplomacy, 10*(2), 158-166. doi:http://dx.doi.org/10.1057/pb.2014.2

Zukin, S. (2014, May 6). Postcard-perfect: The big business of city branding. *The Guardian.* Retrieved on October 10, 2015 from: http://www.theguardian.com/cities/2014/may/06/postcard-perfect-the-big-business-of-city-branding

APPENDIX

1. Sample of randomly selected tweets for hand coding.

Text	Created At	Not-negative	Negative	Related to Crisis	Topic
@onedirection The #DragMeDown is the third in Hungary at #class40 :) Love it!	2015-09-21T14:03:33.000Z	0			Culture/Event
- WHEN ARE YOU GUYS COMING TO HUNGARY? #JANOFest x16	2015-03-14T21:08:52.000Z	0			Conversation
Bones of Thomas Becket to return to Canterbury viaHungary: https://t.co/t5A1KZg7Kg https://t.co/kfa5EHmvs6	2016-01-22T20:50:26.000Z	0			Other
Unioffice Hungary built TelVir app on #@ibmmobile #ibmcloud to help monitor with 1 click water levels across Hungary:http://t.co/GyVCM5geWa	2015-10-09T06:25:26.000Z	0			Other
Latest 6 hour News Hungary shuts border crossings triggering standoff with stranded migrants - Los Angeles Times http://t.co/IEXLwni7je	2015-09-16T00:34:35.000Z		1	1	
HUNGARY CLEARANCE - 33 GOOD USED STAMPS - ALL DIFFERENT http://t.co/OkfGSlVesm http://t.co/dyvtYLWQeQ	2015-08-16T09:33:17.000Z	0			Other
Croatia 'forces' Hungary to open borders to refugees as crisis deepens http://t.co/VaoxHro6tr	2015-09-19T13:31:29.000Z		1	1	
RT @nytimesworld: Hungary premier calls migrant crisis a German problem http://t.co/RqTHCAggwd http://t.co/ng7GVp2IR3	2015-09-04T05:05:36.000Z		1	1	
RT @VineyardKapolna: Check out these 8 hot new #wine regions included #Hungary @winewankers @winetrackerco https://t.co/IdKkq7RV78 https	2016-01-14T10:25:55.000Z	0			Tourism
Germany and Hungary http://t.co/B95X9rvbY7 #travel #vacation	2015-04-28T16:53:16.000Z	0			Tourism
They use our streets as a toilette and makes the train transport in Hungary impossible. They're migrants but all of them have iPhone 6 :DD	2015-09-02T13:22:01.000Z		1	1	
RT @Conflicts: Left - Hungarian refugees walking to Austria in 1956 Right - Syrian refugees walking from #Hungary to #Austria 2015 http://	2015-09-05T02:44:42.000Z	0		1	Photo
RT @RedHotSquirrel: Sweden Poland Denmark Hungary Czech Rep Romania Lithuania and Croatia are NOT yet in the urozone. Finland is con	2015-12-31T21:14:18.000Z	0			Politics
RT BBCBreaking: Hungary asks Serbia to act on migrants ""attacking"" police at border Hungarian media reports http://t.co/K7ipRI9rSg	2015-09-16T15:23:39.000Z		1	1	
Bcs/Adis. Budapest @ Budapest Hungary https://t.co/Iw45nft0oL	2015-07-17T19:07:31.000Z	0			Other
Euro qualifiers relevant to RBL: Hungary lost 4:3 to Greece. Zsolt Kalmr was substituted on and assisted the 3rd goal.	2015-10-12T00:33:01.000Z	0			Sport
#EurovisionSongContest #HUN well hungary now..... GOODNIGHT	2015-05-23T20:50:53.000Z	0			Culture/Event
RT @RealVIXX: [VIXX_TV2] I'm ROVIX. #VIXXTV2 Ep.15. If you wonder about #VIXX agents' solo concert #HEX_SIGN in Hungary (cont) http://t.co/	2015-04-17T23:13:30.000Z	0			Culture/Event
RT @Skype: Japan Hungary Kenya where will your class go for the #SkypeaThon on Dec. 3 & 4? https://t.co/aVJaXfGhYu https://t.co/C1QV3Jy	2015-11-24T21:52:38.000Z	0			Conversation
Hungary army gets powers to curb illegal migration Europe struggles with refugee crisis http://t.co/Ye8omOvXJB #worldnews #news #breakin	2015-09-22T05:27:49.000Z		1	1	
Bloomberg: Hungary mandates pro-government ads for banks in loan program https://t.co/KwBUr2z9Gf #hungary	2016-03-07T16:57:27.000Z	0			Economy
Budapest city council approves venues for 2024 Olympic bid - BUDAPEST Hungary (AP) Budapest's city council h... https://t.co/Q1kzmyehwz	2016-01-27T12:32:43.000Z	0			Tourism
@officialR5 You come to Hungary!?	2015-05-30T18:18:23.000Z	0			Conversation
The footage of all the refugees in Hungary is heartbreaking	2015-09-17T06:03:59.000Z		1	1	
RT @ajplus: This 17-year-old Syrian girl went on the run in Hungary after she was separated from her parents in a refugee camp. https://t.c	2015-09-10T05:50:05.000Z		1	1	
Free! https://t.co/UOSACQra34 #Greece #Hungary #India #Italy #Kenya #Kuwait 3	2015-12-22T17:57:43.000Z	0			Conversation
Celebrating 120 Years of Olympic History in Hungary @xtwintergames	2015-12-15T15:42:35.000Z	0			Tourism
RT @ThatsEarth: Chain Bridge Budapest Hungary https://t.co/Kk1pZXlzRh	2015-10-23T03:15:51.000Z	0			Other
RT @LindaSuhler: PICTURES: Migrants Riot At Hungary Border #SecureTheBorders #NoAmnesty http://t.co/rbNfr80Joa #tcot http://t.co/kOp0eTcnu5	2015-09-16T18:55:10.000Z		1	1	
RT @YourHistoryPics: Gypsy boy with cello Hungary 1931 by Eva Besny. https://t.co/OizdaSJJ4W	2016-01-21T01:54:21.000Z	0			Photo
VIDEO: Hungary's fence to keep the migrants out: Ed Thomas reports from Hungary where thousands of refugees ha... http://t.co/Bbh45We3N4	2015-08-24T22:27:36.000Z		1	1	
Hungary authorizes govt. to use army in migration crisis http://t.co/xkyZ3mkn1B	2015-09-21T16:25:04.000Z		1	1	
Hungary accused of treating refugees inhumanely as migration crisis worsens - via @Telegraph http://t.co/m4BVGQuEAT	2015-08-29T09:44:43.000Z		1	1	
RT @Adel__Almalki: #news by #almalki: Austria and Germany open borders to migrants offloaded by Hungary http://t.co/OvKk17XV19	2015-09-06T08:36:17.000Z		1	1	
RT @BBCSport: Watch Inside F1 from Hungary with @TomClarksonF1 and @allanmcnish. Live now on the @BBCNews channel http://t.co/U942GwApR3	2015-07-25T18:54:26.000Z	0			Sport
RT @FOXSoccer: Notable Euro 2016 QF scores (FT): Gibraltar 0-4 Ireland Denmark 0-0 Albania Serbia 2-0 Armenia Hungary 0-0 Romania	2015-09-04T20:39:56.000Z	0			Sport
RT @dorrkaaa: I hope you recognize yourself I had been working on this for hours @JackJackJohnson lots of love from Hungary! xx https://t.	2015-10-28T14:21:49.000Z	0			Conversation
RT @DrawingPencil: by Eniko Szabo (Hungary) https://t.co/YzbHgsdWdd	2016-02-01T14:11:33.000Z	0			Other
Syrian refugees had tense exchanges with police in Hungary after being told to return to camps.	2015-09-13T17:58:01.000Z		1	1	
RT @EnglishBasNews: Hungary Pledges Extension to #Peshmerga Training - #BasNews http://t.co/VclezyvYrN via @EnglishBasNews	2015-10-04T08:44:30.000Z	0			News
RT @UK__News: Poland Slovakia Hungary and Czech Republic have rejected Cameron's proposal to reduce EU migrant benefits. https://t.co/5C	2016-02-17T21:15:18.000Z		1	1	
@_14LC me: Belgium Germany Denmark Switzerland Austria Czech Rep Hungary Croatia Italy Spain France Italy Portugal USA UK	2015-03-29T13:33:50.000Z	0			Other
RT @EseJapan: Reaching the Sky: Monument to the Uprising (56-os emlkm) in city park #budapest #Hungary http://t.co/OiTrDBZ0Hs	2015-05-29T05:45:10.000Z	0			Tourism
RT @WRNicholls: This buzzard was sat in a tree above a bird hide in Hungary. It made a few swoops down but failed to catch anything! http:	2015-04-29T00:34:48.000Z	0			Other
RT gofondue RT hungarianbc RT itsmedalma I was hungry #foodporn #food #hungary #christmasfair #christmas #hotdo https://t.co/Z0J5N2ySO6	2015-12-11T22:02:54.000Z	0			Conversation
RT CountryResearch #Hungary - Linseed Oil : 2015 #Market #Forecast : https://t.co/kelrzrz0pU #MarketResearchReports	2015-11-10T23:47:58.000Z	0			Other
RT @MetalBulletinZn: * free album of death metal: Amorite old school death metal from Hungary http://t.co/lm7vfvd4dT http://t.co/QZzAYl8mzg	2015-09-04T16:49:07.000Z	0			Culture/Event

2. Code in R

```
#####
# (0) libraries
#####
library("stringr")
library('data.table')
library('ggplot2')
library("multiwayvcov")
library("lmtest")
library("sandwich")
library('wordcloud')
library('tm')
library('SnowballC')
library('parallel')
library('lda')
library("textcat")

setwd('~/Dropbox/TwitterSentiment/')
rm(list = ls())

######
### (0) functions
######
# to remove punct etc
clean_words <- function(x) {
  x <- gsub("[[:punct:]]", "", x)
  x <- gsub("[[:digit:]]", "", x)
  x <- gsub("http\\w+", "", x)
  x <- gsub("[ \t]{2,}", "", x)
  x <- gsub("^\\s+|\\s+$", "", x)
  try.error = function(x)
  {
    y = NA
    try_error = tryCatch(tolower(x), error=function(e) e)
    if (!inherits(try_error, "error"))
      y = tolower(x)
    return(y)
  }
  x <- sapply(x, try.error)
  return(x)
}

# stem the text
stem_text<- function(text, language = "porter", mc.cores = 1) {
  # stem each word in a block of text
  stem_string <- function(str, language) {
    str <- strsplit(x = str, split = " ")
    str <- wordStem(unlist(str), language = language)
    str <- paste(str, collapse = " ")
    return(str)
  }

  # stem each text block in turn
  x <- mclapply(X = text, FUN = stem_string, language, mc.cores = mc.cores)

  # return stemed text blocks
  return(unlist(x))
}

# remove words in the list
remove_matched <- function(x, w_remove) return(x[is.na(match(x, w_remove))])

# find tweets with certain words
find_tweets <- function(x, words_to_search) return(sum(!is.na(match(x, words_to_search))) > 0)

######
#### (0) Load the supporting data sources
######
```

```
#### polarity measure (Thelwall and Buckley)
polarity_data <- fread('~/Dropbox/ TwitterSentiment /data/SentStrength_Data_Sept2011/polar.csv')
polarity_data <- polarity_data[, list(V1, V2, V3)]
polarity_data <- polarity_data[!is.na(V2)]
polarity_data[, V1 := stem_text(clean_words(V1))]
# words already stemmed

# vocabulary of sentiments
neg.words <- fread('data/opinion-lexicon-English/neg-words.txt', header = F)
pos.words <- fread('data/opinion-lexicon-English/pos-words.txt', header = F)
# delete specific negative words (to get conservative estimates)
# neg.words <- neg.words[V1 != 'crisis'][V1 != 'breaking']
neg.words[, V1 := stem_text(V1)]
pos.words[, V1 := stem_text(V1)]

# stop words
stopwords <- fread('data/stopwords_ranksnl.txt', header = F)
stopwords <- as.vector(clean_words(stopwords$V1))
stopwords <- c(stopwords, "rt")

# other words to remove for the lda allocation
words_remove <- c("rt", "hungary", "romania", "slovenia", neg.words$V1, pos.words$V1, stopwords)
words_remove <- stem_text(words_remove)

# load news agecies names (@newsagency)
news_agencies <- fread('data/anna_news_agencies.txt', header = F)
news_agencies[, V1 := clean_words(V1)]
news_agencies <- news_agencies[V1 != 'business'][V1 != 'guardian'][V1 != 'ap'][V1 != 'ft'][
  V1 != 'life'][V1 != 'times'][V1 != 'time'][V1 != 'io'][V1 != 'abc'][V1 != 'good'][V1 != 'waitwait']
news_agencies <- as.vector(news_agencies$V1)
news_agencies <- c(news_agencies, 'bbc')

#######
### (1) examine how well the hand-coding worked
#######
handcoded <- fread('~/Dropbox/ TwitterSentiment /data/handcoded_tweets.csv')
handcoded[is.na(Negative)]
handcoded[, textdata := stem_text(clean_words(Text))]

# remove the new agencies tweets from handcoded data
y <- rep(F, dim(handcoded)[1])
for (i in 1:(length(news_agencies)+1)) {
  x <- handcoded[, grepl(news_agencies[i], textdata)]
  if (i == length(news_agencies) + 1) x <- handcoded[, grepl("new", textdata)]
  y[y==F] <- x[y==F]
}
handcoded <- handcoded[!y]

## run the sentiment analysis
n <- dim(handcoded)[1] # number of tweets
word.list <- str_split(handcoded[, textdata], ' ') # words
l_index <- c(1, cumsum(sapply(word.list, length))[-n] + 1)
u_index <- cumsum(sapply(word.list, length))
words <- unlist(word.list)
pos.matches <- match(words, pos.words$V1)
neg.matches <- match(words, neg.words$V1)
pos.matches_data <- array(0, n)
neg.matches_data <- array(0, n)
for (i in 1:n) {
  pos.matches_data[i] <- sum(!is.na(pos.matches[l_index[i]:u_index[i]]))
  neg.matches_data[i] <- sum(!is.na(neg.matches[l_index[i]:u_index[i]]))
}

handcoded[, senti_pos := pos.matches_data] # number of positive words
handcoded[, senti_neg := neg.matches_data] # negative words
handcoded[, senti_dif := senti_pos - senti_neg] # difference between sentiment

### compute the polarity
polar_matches <- match(words, polarity_data$V1)
```

```
polar_matches <- polarity_data[polar_matches, V2]
polar.matches_data.pos <- array(0, n)
polar.matches_data.neg <- array(0, n)
for (i in 1:n) {
  poltmp <- polar_matches[l_index[i]:u_index[i]]
  polar.matches_data.pos[i] <- sum(poltmp[!is.na(poltmp) & poltmp > 0])
  polar.matches_data.neg[i] <- sum(poltmp[!is.na(poltmp) & poltmp < 0])
}
handcoded[, polar_pos := polar.matches_data.pos]
handcoded[, polar_neg := -polar.matches_data.neg]
handcoded[, polar_dif := polar_pos - polar_neg]
## test the matching of senti dict:
dim(handcoded[!is.na(Negative) & senti_dif < 0])[1]/n # negative match
dim(handcoded[is.na(Negative) & senti_dif >= 0])[1]/n # positive match
dim(handcoded[is.na(Negative) & senti_dif < 0])[1]/n # senti neg, handcode pos
dim(handcoded[!is.na(Negative) & senti_dif >= 0])[1]/n # senti pos, handcode neg
## test the matching of polarity:
dim(handcoded[!is.na(Negative) & polar_dif < 0])[1]/n # negative match
dim(handcoded[is.na(Negative) & polar_dif >= 0])[1]/n # positive match
dim(handcoded[is.na(Negative) & polar_dif < 0])[1]/n # polar neg, handcode pos
dim(handcoded[!is.na(Negative) & polar_dif >= 0])[1]/n # polar pos, handcode neg

### described -- pick SentiStrength

#######
### (2) Load the main data
#######

#### old data
data0 <- fread('data/Data.csv')
data0[, ind := 1]
data0[, date := substr(Tweet_Creation,5,10)]
data0[, sum(ind), by='Searchterm,date']
data0[, rat_agent :=- User_Followers + User_Friend_Count]
setkey(data0, rat_agent)
data0 <- data0[Searchterm != "slovenia"]

# distribution of followers - following
hist(data0[, rat_agent], xlim = c(-10000,10000), breaks = 10000)
data0[, news_agency3 := (grepl("News", User_Description) | grepl("news", User_Description))]
data0[, textdata := clean_words(Tweet)]
data0[, textdata := stem_text(textdata)]
y <- rep(F, dim(data0)[1])
for (i in 1:(length(news_agencies)+1)) {
  x <- data0[, grepl(news_agencies[i], textdata)]
  if (i == length(news_agencies) + 1) x <- data0[, grepl("new", textdata)]
  y[y==F] <- x[y==F]
}
## percent of media tweets, data0
sum(y)/dim(data0)[1] # 11.6 percent

### new data -- over time

data <- fread('data/Data(1).csv')
# check the sentiment method
#data <- fread('data/random_tweets_hungary.csv')
colnames(data)[3] <- 'stamp'
data <- data[Searchterm != "slovenia"]
data[, date := substr(stamp, 1, 10)]
data[, date := as.Date(date, "%Y-%m-%d")]
setkey(data, Searchterm, date)
data[, ind := 1]
data[, week := as.numeric(date - min(date))%/%7]
data[, week := week*7 + min(date)]
data[, sum(ind), by="Searchterm,week"]
data[, sum(ind), by="Searchterm"]

n <- dim(data)[1] # number of tweets

## clean the text
```

```
data[, textdata := clean_words(Text)]
data[, textdata := stem_text(textdata)]

## remove the news agencies tweets
y <- rep(F, dim(data)[1])
for (i in 1:(length(news_agencies)+1)) {
  x <- data[, grepl(news_agencies[i], textdata)]
  if (i == length(news_agencies) + 1) x <- data[, grepl("new", textdata)]
  y[y==F] <- x[y==F]
}

sum(y)/dim(data)[1] # 12.3 percent of tweets with media names
# tweets that are done by the news agencies
data[, news_agencies_tweets := y] # mention the news agencies
# remove the 'news agencies' tweets
data <- data[news_agencies_tweets == F]
n <- dim(data)[1]


######
## create a list of hashtags
######

z <- c()
for (i in 1:n) {
  x <- unlist(strsplit(data[i, Text], " "))
  for (j in 1:length(x)) if (grepl("#", x[j])) z <- c(z, x[j])
}
z1 <- table(z)
z1 <- z1[order(z1, decreasing = T)]
z1[c(1:3,5:21)]
#ggplot(data[, sum(ind), by="Searchterm,week"], aes(x=week, y=V1, colour = Searchterm, group = Searchterm)) +
#  geom_line() + theme_bw() # volume of tweets


######
### LDA
######

# remove words for lda and find news agencies tweets

yy <- array(NA, n)
yy2 <- array(NA, n)
for (i in 1:n) {
  y <- strsplit(data[i, textdata], " ")
  y <- unlist(y)
  yy[i] <- paste(remove_matched(y, words_remove), sep = " ", collapse = " ")
  yy2[i] <- find_tweets(y, c(news_agencies, 'new'))
}

# text with removed words
data[, textdata_lda := yy]

#### classify the text by language
#data[,  lang_tweet := textcat(Text)]
#data_english <- data[lang_tweet == "english" | lang_tweet == "scots" | lang_tweet == "irish"]
#n_english <- dim(data_english)[1] # number of tweets

### run the LDA analysis

# other options:
# library('stm')
# library('topicmodels')

lda_prepare <- lexicalize(data[, textdata_lda])
# number of words
n_words <- length(lda_prepare$vocab)
# number of documents in which each word appears
words_doc_count <- array(0, n_words)
```

```
for (i in 1:n) words_doc_count[lda_prepare$documents[[i]][1,] + 1] <- words_doc_count[lda_prepare$documents[[i]][1,]+ 1] +
lda_prepare$documents[[i]][2,]
# filter out the words which occur only once
#lda_prepare$documents <- filter.words(lda_prepare$documents, which(words_doc_count == 1))
#n_words_left <- n_words - length(which(words_doc_count == 1))
# examine the word counts, from top words to bottom
x <- word.counts(lda_prepare$documents, lda_prepare$vocab); x <- x[order(x, decreasing = T)]
# run the lda allocation exercise via collapsed gibbs
n_tries <- 9
choose_nmodels <- array(0, n_tries)
store_doc_alloc <- list()

for (i in 1:n_tries+1) {
reslda <- lda.collapsed.gibbs.sampler(lda_prepare$documents, i, lda_prepare$vocab, burnin = 50,
                              num.iterations = 500, alpha = .1, eta = .1, compute.log.likelihood = T)
# convergence of the chain
plot(reslda$log.likelihoods[2,])
#plot(reslda$log.likelihoods[2,]) # burn-in around 100 first draws
# based on the number of words in the document allocated to each topic (after throwing away the burn-in)!
# allocate the documents to groups based on this!
#reslda$document_expects
choose_nmodels[i-1] <- reslda$log.likelihoods[2, 100]
print(reslda$topic_sums)
print(top.topic.words(reslda$topics))
store_doc_alloc[[i-1]] <- reslda
}
bic <- -2 * choose_nmodels + (1:length(choose_nmodels) + 1) * (1 + n_words) * log(n)
aic <- -2 * choose_nmodels + (1:length(choose_nmodels) + 1) * (1 + n_words) * 2
png(file = 'figs/ll.png')
plot(y = choose_nmodels, x = c(2:10), ylab = "Log-Likelihood", xlab = "Number of topics") # choose 3
dev.off()
png(file = 'figs/aic.png')
plot(y = aic, x = c(2:10), ylab = "AIC", xlab = "Number of topics") # choose 3
dev.off()
png(file = 'figs/bic.png')
plot(y = bic, x = c(2:10), ylab = "BIC", xlab = "Number of topics") # choose 2
dev.off()

#str(reslda)
#reslda$document_sums[1:10,1:12]
#word.counts(lda_prepare$documents)

# allocate tweets to the groups - 3 clusters, re-run the code
png(file = 'figs/ll_evol.png')
plot(store_doc_alloc[[2]]$log.likelihoods[2,], ylab = "Log-Likelihood", xlab = "MCMC draws") # plot of the likelihood evolution
abline(v = 50, lty = 2)
dev.off()
print(store_doc_alloc[[2]]$topic_sums)
print(top.topic.words(store_doc_alloc[[2]]$topics))
data[, group := apply(store_doc_alloc[[2]]$document_expects, 2, which.max)]

######
#### sentiment analysis on the entire data
######


# compute the sentiment for each tweet
word.list <- str_split(data[, textdata], ' ')
l_index <- c(1, cumsum(sapply(word.list, length))[-n] + 1)
u_index <- cumsum(sapply(word.list, length))
words <- unlist(word.list)
pos.matches <- match(words, pos.words$V1)
neg.matches <- match(words, neg.words$V1)
pos.matches_data <- array(0, n)
neg.matches_data <- array(0, n)
for (i in 1:n) {
  pos.matches_data[i] <- sum(!is.na(pos.matches[l_index[i]:u_index[i]]))
  neg.matches_data[i] <- sum(!is.na(neg.matches[l_index[i]:u_index[i]]))
}
```

```
data[, senti_pos := pos.matches_data] # number of positive words
data[, senti_neg := neg.matches_data] # negative words
data[, senti_dif := senti_pos - senti_neg] # difference between sentiment
#write.csv(data, file = 'sample_senti.csv')

### try computing the polarity
polar_matches <- match(words, polarity_data$V1)
polar_matches <- polarity_data[polar_matches, V2]
polar.matches_data.pos <- array(0, n)
polar.matches_data.neg <- array(0, n)
for (i in 1:n) {
  poltmp <- polar_matches[l_index[i]:u_index[i]]
  polar.matches_data.pos[i] <- sum(poltmp[!is.na(poltmp) & poltmp > 0])
  polar.matches_data.neg[i] <- sum(poltmp[!is.na(poltmp) & poltmp < 0])
}
data[, polar_pos := polar.matches_data.pos]
data[, polar_neg := -polar.matches_data.neg]
data[, polar_dif := polar_pos - polar_neg]

## compare polar and senti
# senti neg: 9729 out of 47572
# polar neg: 8880 out of 47572
# both neg: 5538

# number of tweets
ggp <- ggplot(data[group == 3, sum(ind), by="Searchterm,week"], aes(x=week, y=V1, colour = Searchterm, group = Searchterm)) +
  geom_line() + theme_bw() +
  geom_vline(xintercept = as.numeric(as.Date(c('2015-07-01'))), colour = "red", linetype = 4) +
  geom_vline(xintercept = as.numeric(as.Date(c('2015-10-31'))), colour = "red", linetype = 4) +
  ylab("Volume of tweets") +
  xlab("Weeks")
ggp
ggsave(filename="figs/volume_trend.png", plot=ggp)

## 2015-10-29 - 2015-11-09 - something goes on in romania

# plot of negative sentiments
ggp <- ggplot(data[group ==1, mean(senti_dif), by="Searchterm,week"], aes(x=week, y=V1, colour = Searchterm, group =
Searchterm)) +
  geom_line() + theme_bw() +
  geom_vline(xintercept = as.numeric(as.Date(c('2015-07-01'))), colour = "red", linetype = 4) +
  geom_vline(xintercept = as.numeric(as.Date(c('2015-10-31'))), colour = "red", linetype = 4) +
  ylab("Average sentiment of the tweet") +
  xlab("Weeks")
ggp
ggsave(filename="figs/neg_senti_trend.png", plot=ggp)

# plot of polarity
ggp <- ggplot(data[group == 1, mean(polar_dif), by="Searchterm,week"], aes(x=week, y=V1, colour = Searchterm, group =
Searchterm)) +
  geom_line() + theme_bw() +
  geom_vline(xintercept = as.numeric(as.Date(c('2015-07-01'))), colour = "red", linetype = 4) +
  geom_vline(xintercept = as.numeric(as.Date(c('2015-10-31'))), colour = "red", linetype = 4) +
  ylab("Average polarity of the tweet") +
  xlab("Weeks")
ggp
ggsave(filename="figs/polarity_trend.png", plot=ggp)

# plot of polarity for logit
ggp <- ggplot(data[, mean(polar_dif < 0), by="Searchterm,week"], aes(x=week, y=V1, colour = Searchterm, group = Searchterm)) +
  geom_line() + theme_bw() +
  geom_vline(xintercept = as.numeric(as.Date(c('2015-07-01'))), colour = "red", linetype = 4) +
  geom_vline(xintercept = as.numeric(as.Date(c('2015-10-31'))), colour = "red", linetype = 4) +
  ylab("Probability of the tweet to have negative polarity") +
  xlab("Weeks")
ggp
ggsave(filename="figs/polarity_logit_trend.png", plot=ggp)
```

```
crisis_begin <- as.Date(c('2015-07-01'))
crisis_end <- as.Date(c('2015-10-31'))

#2015-10-29 - 2015-11-09

# regression of negative tweets - on the day level
# neg_sent <- data[, list(polar = mean(polar), neg = mean(neg)), by="Searchterm,date"]
# neg_sent[, Hungary := Searchterm == "hungary"]
# neg_sent[, during_crisis := date >= crisis_begin & date < crisis_end]
# neg_sent[, after_crisis := date >= crisis_end]
# res <- lm(polar ~ Hungary + during_crisis + after_crisis + Hungary*during_crisis + Hungary*after_crisis, neg_sent)
# summary(res)

# on the tweet level
data[, Hungary := Searchterm == "hungary"]
data[, during_crisis := date >= crisis_begin & date < crisis_end]
data[, after_crisis := date >= crisis_end]
data[, ind_neg := polar_dif < 0]
data[, romania_protests := Hungary == F & (date >= as.Date("2015-10-30") & date < as.Date("2015-11-10"))]
res <- glm(ind_neg ~ romania_protests + Hungary + during_crisis + after_crisis + Hungary*during_crisis + Hungary*after_crisis,
data[group == 2], family = binomial)
summary(res)

exp(sum(res$coefficients[c(1,2,3,4)]))/(1 + exp(sum(res$coefficients[c(1,2,3,4)]))) -
exp(sum(res$coefficients[c(1,2,3)]))/(1 + exp(sum(res$coefficients[c(1,2,3)])))
# during the crisis - negative is significant

# miscellaneous

## composition of tweets
data[group ==1 , tw_types := "tourism"]
data[group ==2 , tw_types := "refugees"]
data[group ==3 , tw_types := "miscellaneous"]
share_groups <- data[, list(tw = sum(ind)), by='tw_types,Searchterm,week']
share_groups[, total := sum(tw), by="week,Searchterm"]
share_groups[, share := tw/total, by="week,Searchterm"]

ggp <- ggplot(share_groups[Searchterm == "hungary"], aes(x=week, y=share, colour = tw_types, group = tw_types)) +
  geom_line() + theme_bw() +
  geom_vline(xintercept = as.numeric(as.Date(c('2015-07-01'))), colour = "red", linetype = 4) +
  geom_vline(xintercept = as.numeric(as.Date(c('2015-10-31'))), colour = "red", linetype = 4) +
  ylab("Share of tweets of this type") +
  xlab("Weeks")
ggp
ggsave(filename="figs/tw_share_romania.png", plot=ggp)


######
### apply the results of LDA to the regression
######

### pictures are above (in the senti analysis)

# regression by groups of tweets
neg_sent_group <- data[, list(neg = mean(ind_neg), vol = sum(ind)), by="Searchterm,date,group,week"]
neg_sent_group[, Hungary := Searchterm == "hungary"]
neg_sent_group[, during_crisis := date >= crisis_begin & date < crisis_end]
neg_sent_group[, after_crisis := date >= crisis_end]
res <- lm(neg ~ Hungary + during_crisis + after_crisis + Hungary*during_crisis + Hungary*after_crisis, neg_sent_group[group == 3])
#summary(res)
coeftest(res, vcov. = vcovHAC)
# by groups - on the crisis topic marginally significant, otherwise -- nothing

ggplot(neg_sent_group[group == 3, list(neg = mean(neg), vol = sum(vol)), by="Searchterm,week"], aes(x=week, y=neg, colour =
Searchterm, group = Searchterm)) +
  geom_line() + theme_bw() +
  geom_vline(xintercept = as.numeric(as.Date(c('2015-07-01'))), colour = "red", linetype = 4) +
  geom_vline(xintercept = as.numeric(as.Date(c('2015-10-31'))), colour = "red", linetype = 4) +
  ylab("Number of negative words per tweet") +
  xlab("Weeks")
```

```
#data[Searchterm == "hungary" & neg > 0 & date >= crisis_begin & date < crisis_end]

######
## wordl for periods
######

data[, c_begin := date >= crisis_begin & date < crisis_end]
data[, c_end := date >= crisis_end]
data[Searchterm == "hungary" & c_begin == F & c_end == F]

make_wordcl <- function(x, maxw, minf, remw, flag_negsenti, flag_possenti) {
  y <- unlist(str_split(x, ' '))
  y <- y[is.na(match(y, stopwords(kind = "en")))]
  y <- y[is.na(match(y, remw))]
  #if (flag_negsenti) y <- y[!is.na(match(y, neg.words$V1))]
  #if (flag_possenti) y <- y[!is.na(match(y, pos.words$V1))]
  if (flag_negsenti) y <- y[!is.na(match(y, polarity_data[V2 < 0, V1]))]
  if (flag_possenti) y <- y[!is.na(match(y, polarity_data[V2 > 0, V1]))]
  un_words <- names(table(y))
  un_words_numbers <- as.numeric(table(y))
  wordcloud(un_words, un_words_numbers, scale=c(3,.5), min.freq = minf, max.words = maxw)
}


png(file = "figs/hungary_world.png")
#layout(matrix(c(1, 2, 1, 2, 1, 2), nrow=2), heights=c(1, 1))
#par(mfrow=c(2,3))
#plot.new()
#text(x=0.5, y=0.5, "Hungary: Before Crisis")
#plot.new()
#text(x=0.5, y=0.5, "Hungary: Crisis")
#plot.new()
#text(x=0.5, y=0.5, "Hungary: After Crisis")
par(mfrow=c(1,3))
make_wordcl(data[Searchterm == "hungary" & c_begin == F & c_end == F, textdata], 25, 10,
          c('hungary', 'rt', 'budapest', 'crisi', 'break', 'refuge'), T, F)
make_wordcl(data[Searchterm == "hungary" & c_begin == T & c_end == F, textdata], 25, 10,
          c('hungary', 'rt', 'budapest', 'crisi', 'break', 'refuge', 'cross'), T, F)
make_wordcl(data[Searchterm == "hungary" & c_begin == F & c_end == T, textdata], 25, 10,
          c('hungary', 'rt', 'budapest', 'crisi', 'break', 'refuge'), T, F)
dev.off()

png(file = "figs/romania_world.png")
par(mfrow=c(1,3))
#layout(matrix(c(1, 2, 1, 2, 1, 2), nrow=2), heights=c(1, 1))
#par(mfrow=c(2,3))
#plot.new()
#text(x=0.5, y=0.5, "Romania: Before Crisis")
#plot.new()
#text(x=0.5, y=0.5, "Romania: Crisis")
#plot.new()
#text(x=0.5, y=0.5, "Romania: After Crisis")
make_wordcl(data[Searchterm == "romania" & c_begin == F & c_end == F, textdata], 25, 10,
          c('romania', 'rt', 'din', 'di', 'dog', 'b', 'strai', 'fire', 'h'), T, F)
make_wordcl(data[Searchterm == "romania" & c_begin == T & c_end == F, textdata], 25, 10,
          c('romania', 'rt', 'din', 'di', 'dog', 'b', 'strai', 'fire', 'h'), T, F)
make_wordcl(data[Searchterm == "romania" & c_begin == F & c_end == T, textdata], 25, 10,
          c('romania', 'rt', 'din', 'di', 'dog', 'b', 'strai', 'fire', 'h'), T, F)
dev.off()

######
### additional to do:
######
## do a pseudo-lda: just get tweets which contain words like 'tourism' of 'refugee', and check the sentiment there

## find the tourism words :
## remove the news agencies tweets
```

```
tourism_words <- c('travel')
refugees_words <- c('refuge', 'migrant', 'border', 'polic', 'close')

y <- rep(F, dim(data)[1])
for (i in 1:(length(tourism_words))) {
  x <- data[, grepl(tourism_words[i], textdata)]
  y[y==F] <- x[y==F]
}
data[, group_tourism := y]

y <- rep(F, dim(data)[1])
for (i in 1:(length(refugees_words))) {
  x <- data[, grepl(refugees_words[i], textdata)]
  y[y==F] <- x[y==F]
}
data[, refugees_words := y]

######
### save the file
######

#save(list = ls(), file = 'data/output/working.RData')
#load(file = 'data/output/working.RData')
```