

FUNDAÇÃO GETÚLIO VARGAS
ESCOLA DE ADMINISTRAÇÃO DE EMPRESAS DE SÃO PAULO

LAURA KATHLEEN NELSON

**RISKY BUSINESS: SOCIAL MEDIA METRICS AND POLITICAL RISK
ANALYSIS**

**SÃO PAULO
2015**

FUNDAÇÃO GETÚLIO VARGAS
ESCOLA DE ADMINISTRAÇÃO DE EMPRESAS DE SÃO PAULO

LAURA KATHLEEN NELSON

**RISKY BUSINESS: SOCIAL MEDIA METRICS AND POLITICAL RISK
ANALYSIS**

Dissertação apresentada à Escola de Administração de Empresas de São Paulo da Fundação Getúlio Vargas, como requisito para obtenção do título de Mestre Profissional em Gestão Internacional.

Campo do Conhecimento:
Economia e Finanças Internacionais

Orientador Prof. Dr. Julia Alice
Sophia von Maltzan Pacheco

**SÃO PAULO
2015**

Nelson, Laura Kathleen.

Risky Business: Social Media Metrics and Political Risk Analysis / Laura Kathleen Nelson. - 2015.

96 f.

Orientador: Julia von Maltzan Pacheco

Dissertação (MPGI) - Escola de Administração de Empresas de São Paulo.

1. Risco país. 2. Administração de risco. 3. Mercados emergentes. 4. Mídia social. 5. Twitter. I. Pacheco, Julia von Maltzan. II. Dissertação (MPGI) - Escola de Administração de Empresas de São Paulo. III. Título.

CDU 330.131.7

LAURA KATHLEEN NELSON

**RISKY BUSINESS: SOCIAL MEDIA METRICS AND POLITICAL RISK
ANALYSIS**

Dissertação apresentada à Escola de Administração de Empresas de São Paulo da Fundação Getúlio Vargas, como requisito para obtenção do título de Mestre Profissional em Gestão Internacional.

Campo do Conhecimento:
Economia e Finanças Internacionais

Data de Aprovação:

___/___/___.

Banca Examinadora:

Prof. Dr. Julia Alice Sophia von
Maltzan Pacheco (ORIENTADOR)

Prof. Dr. Frederico Araujo Turolla

Prof. Dr. Benjamin Rosenthal

ABSTRACT

Quantifying country risk – and in particular, political risk – poses great difficulties for business, institutions, and investors alike. As economic indicators are updated far less frequently than Facebook feeds, it can be challenging for political risk analysts to understand, and more importantly measure, what is taking place in real time on the ground. However, with the growing availability of big data from social media platforms such as Twitter, now is an opportune moment to examine the types of social media metrics that are available and the limitations to applying them to country risk analysis, particularly during episodes of political upheaval. This study, using the qualitative method of bibliographical research, identifies the current landscape of data available from Twitter, analyzes the current and potential methods of analysis, and discusses their possible application to the field of political risk analysis. After a thorough review of the field to date, and given the expected near- to medium-term technological advancements, this study concludes that despite obstacles like the cost of data storage, limitations of real-time analysis, and the potential for data manipulation, the potential benefits of the application of social media metrics to the field of political risk analysis, particularly for structured-qualitative and quantitative models, outweigh the challenges.

KEY WORDS: political risk analysis, country risk, emerging markets, social media, big data, Twitter.

RESUMO

A quantificação do risco país – e do risco político em particular – levanta várias dificuldades às empresas, instituições, e investidores. Como os indicadores econômicos são atualizados com muito menos frequência do que o Facebook, compreender, e mais precisamente, medir – o que está ocorrendo no terreno em tempo real pode constituir um desafio para os analistas de risco político. No entanto, com a crescente disponibilidade de “big data” de ferramentas sociais como o Twitter, agora é o momento oportuno para examinar os tipos de métricas das ferramentas sociais que estão disponíveis e as limitações da sua aplicação para a análise de risco país, especialmente durante episódios de violência política. Utilizando o método qualitativo de pesquisa bibliográfica, este estudo identifica a paisagem atual de dados disponíveis a partir do Twitter, analisa os métodos atuais e potenciais de análise, e discute a sua possível aplicação no campo da análise de risco político. Depois de uma revisão completa do campo até hoje, e tendo em conta os avanços tecnológicos esperados a curto e médio prazo, este estudo conclui que, apesar de obstáculos como o custo de armazenamento de informação, as limitações da análise em tempo real, e o potencial para a manipulação de dados, os benefícios potenciais da aplicação de métricas de ferramentas sociais para o campo da análise de risco político, particularmente para os modelos qualitativos-estruturados e quantitativos, claramente superam os desafios.

PALAVRAS CHAVE: análise de risco político, risco país, mercados emergentes, mídias sociais, ferramentas sociais, big data, Twitter.

TABLE OF CONTENTS

1. Introduction	1
<i>Figure 1: Globalization of Facebook.....</i>	<i>3</i>
2. Literature Review	5
2.1 Country risk analysis: Theory & practice	5
<i>Figure 2: Four models of country risk analysis.....</i>	<i>11</i>
<i>Figure 3: ICRG Composite Rating System.....</i>	<i>14</i>
<i>Figure 4: BERI Political Risk Index.....</i>	<i>14</i>
2.2 Sociology: Network theory & information studies	15
<i>Figure 5: Strength of Weak Ties.....</i>	<i>17</i>
<i>Figure 6: Interpersonal bridges.....</i>	<i>24</i>
2.3 Political science: Social movement theory, the Arab spring and Gezi park.....	26
<i>Figure 7: Gezi park & Taksim square.....</i>	<i>31</i>
<i>Figure 8: Proposed redesign.....</i>	<i>31</i>
3. Research Question.....	33
4. Methodology.....	35
5. Data Description	37
5.1 Twitter data.....	37
<i>Figure 9: Twitter metadata map</i>	<i>38</i>
5.2 Sources of twitter data.....	40
5.2a The planet of the APIs	42
5.2b Full stream ahead.....	42
<i>Figure 10: Topsy infographic.....</i>	<i>45</i>
5.3 Other players	47
5.4 Cost of Twitter data.....	49
5.5 Relevant financial data	51
6. Interpretation & Analysis	53
6.1 Limitations of traditional political risk indicators.....	54
6.2 Gezi protests in tweets	54
<i>Figure 11: Geolocated tweets</i>	<i>56</i>
<i>Figure 12: Distribution of tweets by language</i>	<i>57</i>
<i>Figure 13: Active MENA Twitter users</i>	<i>58</i>
<i>Figure 14: CNN-Turk vs CNN International coverage.....</i>	<i>59</i>
<i>Figure 15: User distribution.....</i>	<i>60</i>
6.3 Financial impact	60
<i>Figure 16: BIST 100 Index closing values.....</i>	<i>61</i>
<i>Figure 17: USD:TL foreign exchange rate.....</i>	<i>61</i>
6.4 Additional methods of tweet analysis.....	63
6.5 Smarter spam.....	67
<i>Figure 18: Crowdturfer behavior mapping</i>	<i>70</i>
<i>Figure 19: AP tweet & DJIA movement.....</i>	<i>73</i>
6.6 Ethical questions	74
7. Conclusions	76

1. Introduction

In the wake of political uprisings like the Arab spring, the Gezi park protests, Euromaidan, and others, the fields of both research and media are increasingly focused on the use of social media as a tool for collective action and social movement mobilization. Adding to that, revelations about the extent of data collection by the National Security Agency (NSA) of the US has raised awareness about the proverbial “digital breadcrumbs” or “footprint” that users leave behind when they go online. Short of using Tor (software that purports to grant users anonymity while surfing the web – though research is mixed on its effectiveness), users are invariably having their data harvested and put to use by other parties – especially free social media sites. As the adage goes: if you are not paying for something, you are probably the product.

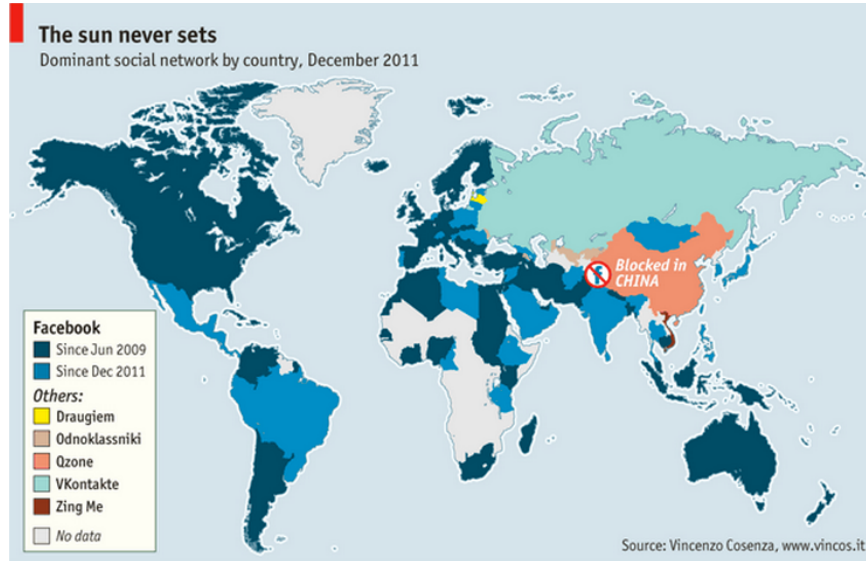
These data trends have accelerated in recent years, with many firms, including IBM, citing the shocking statistic that 90% of the data in the history of the world was created in the past two years alone, estimating that 2.5 quintillion bytes of data are being created every day – equivalent to 25 billion copies of *The Economist* (IBM, n.d.; The Economist, 2010). So what distinguishes this “Big Data” from the information collected from consumers in the past? Many point to the so-called 3Vs – volume, velocity, and variety – as one of the defining characteristic of big data, identified by Gartner analyst Doug Laney back in 2001 (Press, 2013). IBM added a fourth V – veracity – to the equation, emphasizing the importance of the reliability of data (IBM, 2013). Laney defined the term more fully, as “high-volume, -velocity and -variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making” (Sicular, 2013). In short, the nature of the four Vs of big

data must lead to different approaches and tools for analysis. Given social media's anonymous nature, the fourth V – veracity – is of particular importance, and will be discussed in greater detail in this study.

The exponential growth in data capture and analysis is necessarily being driven by revenue. Consumer goods companies were some of the earliest adopters of big data analytical tools, with McKinsey finding 6% higher profits for big data leaders relative to other companies from 1999-2009 (McKinsey, 2013). Indeed, the concept of return on investment for social media (as measured by click-through rates, lead generation, customer acquisition, revenue, etc.) has become commonplace in the modern marketing world (Eckerling, 2014). Governments, too, have recognized the value of data collection, with the NSA building a USD1.5 billion, one million square-foot facility for data storage in Utah, nicknamed Bumblehive (NSA, n.d.). According to the International Data Corporation (IDC), the big data market would be worth USD16.1 billion in 2014, at a growth rate six times higher than that of the IT market as a whole (Dwoskin, 2014a).

Furthermore, as prominent actors like the Dow Jones Industrial Average and the New York Stock Exchange begin to get onboard the big data train, social media metrics may well be the next frontier for financial analysts looking to tap into market sentiment.

The increasingly global reach of social media platforms, such as Facebook and Twitter, offers greater opportunity for analysis. Indeed, according to research firm comScore, one out of every seven minutes spent online in the world is spent on Facebook (The Economist, 2012). Furthermore, much of the network's recent growth has come from emerging markets (see Figure 1).

Figure 1: *Globalization of Facebook*¹

This globalization of social media growth parallels that of American multinational companies, many of which now see around half their revenues generated abroad. In this setting of truly international business and increasing cross-border investments, country risk continues to be a challenge. Indeed, quantifying country risk – and in particular, political risk – poses great difficulties for business, institutions, and investors alike.

Traditionally, political risk has been a field dominated by subjective analysis and regional experts rather than hard data – and what hard data was available was so infrequently updated (monthly, quarterly, yearly, etc.) as to be of limited use during rapidly evolving political risk events. Indeed, the Egyptian uprising of 2011 took a mere 18 days to unseat a government. Kobrin (1982) hit the nail on the head when he described the “inherent subjectivity of the task environment” this way: “the integration of political assessments into investment decision making is subjective in the sense that the translation

¹ <http://www.economist.com/node/21546020>

of qualitative summaries of environments into impacts on operations, into forecasts of risk and return, takes place intuitively through mental processes that are difficult if not impossible to make explicit, much less replicate.” And while political risk assessment – some methods more than others – does make also use of quantitative analysis, the field is still considered more art than science. Indeed, a political risk analyst can assign a numerical ranking to a country’s political stability or the strength of its institutions, but these are necessarily subjective assessments, as strength and stability in a political setting are not inherently quantifiable variables in the same way as GDP or return on equity. Furthermore, the secrecy that surrounds different firms’ approaches, treated in the manner of intellectual property, serves to further obscure the quantitative-qualitative mix.

With the growing availability of big data from social media platforms such as Twitter, particularly in emerging markets, now is an opportune moment to examine the prudence of instrumentalizing these social media metrics, especially during episodes of political upheaval. In the absence of comprehensive Twitter data sets (see Chapter 4), this study uses the qualitative approach of biographical research *to explore the type of social media data that is available from Twitter, the current and potential methods of analysis, and the limitations of their application in the field of political risk analysis.*

After a thorough review of the field to date, and given the expected near- to medium-term technological advancements, this study concludes that despite obstacles like the cost of data storage, limitations of real-time analysis, and the potential for data manipulation, the potential benefits of the application of social media metrics to the field of political risk analysis, particularly for structured-qualitative and quantitative models, outweigh the challenges.

2. Literature Review

In looking at the possible channels of interaction between social media and country risk, it is crucial to first examine the various disciplines that inform such an exploratory study. As a starting point, country risk has been richly developed in prior academic literature, with different approaches for measuring and quantifying country risk applied in practice by for-profit risk management firms, as well as by regulators and other institutions. To understand how social media metrics could potentially be incorporated into country risk analysis, it is important to examine both the theoretical underpinnings of country risk, as well as some of the prevailing ways in which various stakeholders currently perceive and analyze country risk.

The study of social media draws from a variety of fields, ranging from pre-internet network theory to more contemporary areas such as information and communications technology (ICT). Lastly, political science offers a crucial link between social media and country risk, as seen through foundational texts on collective action, as well as newer political science forays examining the role of social media in recent protest movements like the Arab spring and Occupy demonstrations.

2.1 Country risk analysis: Theory and practice

Country risk analysis has a long theoretical history and diverse real world applications that serve as background for this study. A review of the working definitions of the term, as well as the pertinent academic literature – from both a theoretical and practical perspective – offers insight into the current understanding of country risk in the field.

As is true of both the literature and practice of country risk analysis, there is no real consensus on the definition of country risk. For *The Financial Times*, country risk and sovereign risk are interchangeable, defined as the “risk associated with investing in a particular country or providing funds to its government” (Financial Times Lexicon, n.d.). For its part, the IMF asserts that the traditional definition of the term is “broadly, the probability that a country may not pay its debts” (IMF, 2011). However, in the same breath, the IMF has pushed for a wider definition of the term, due to the “very complex interactions between fiscal balances, public and private debt, and the financial sector” exposed by the 2008 global financial crisis.

Others still view country risk from the perspective of a corporation or an investor. Roy and Roy (1994), who similarly equate country risk with sovereign debt risk, define it as “the potential financial loss that could arise from macroeconomic or political events in a country” and “the minimum risk that arises from holding a claim on a public or private equity within a foreign nation”. Bouchet, Clark and Gros Lambert (2003) perhaps put it most simply, defining country risk as “all the additional risks induced by doing business abroad, as opposed to domestic transactions.”

Some authors focus on a more operational definition of the term, focusing instead on the analysis or assessment of risk, as opposed to the strict definition of the risk itself. According to Simon (1984), it concerns the “identification, analysis, management, and reduction of sociopolitical risks to foreign investment”. Meldrum (2000) too focuses on the operational context, defining the field by its “attempts to identify imbalances that increase the risk of a shortfall in the expected return of a cross-border investment.”

According to Fitzpatrick (1983), there are four groups of operational definitions of political risk. The first concerns government intervention or sovereign action, such as embargos or expropriation; the second focuses on political occurrences, including changes in government and violence; and the third views political risk in terms of an environment as opposed to an event. This would encompass evaluations of broad factors like the general level of political stability or strength of institutions in a given country. The fourth shares the same environmental approach as the third, but does not seek to assign a strict definition to it. However, upon closer examination of the four categories, Fitzpatrick (1983) concludes that:

“the literature is found to define political event risk rather than political risk. Such a definitional situation is unsatisfactory. The character of politics is a continuous process rather than a discrete event series. The definition of political risk would be improved if it were evolved in terms of process variables rather than event variables.”

Clearly, the field is far from achieving a consensus on the precise definition, let alone the best approach to analysis, as will be seen later in Section 2.1. More interestingly still, the foundational authors on the subject each bemoan the lack of standardization of terms or approaches – and yet, half a decade later there appears to be no further consensus or consolidation. As Simon (1984) highlighted, “with each additional international crisis, the demand for risk analyses proliferates, even though no one as of yet has established a proven track record for forecasting such developments.” In Fitzpatrick’s (1983) words, “the literature contains an ad hoc proliferation of modeled attempts to explain the impact of the political environment on the firm, attempts that suffer from a lack of rigorous testing.” Kobrin (1979) likewise acknowledges the lack of

consensus: “while [political risk’s] usage almost universally implies a possibility of unwanted consequences arising from political activity, there is certainly no agreement on its precise meaning.”

In terms of theory, the field of modern country risk analysis began in earnest the 1960s, as more firms were pursuing growth outside their national borders, and in the wake of the widespread corporate expropriations that took place during the Cuban revolution (Bouchet et al., 2003). Through the oil crisis of the 1970s and the debt crises of the 1980s, the body of literature expanded to encompass more viewpoints. Some of the earlier literature – from the likes of Kobrin (1979) and others – views political risk as stemming, in particular, from government policy or interference.

Interestingly, much of the early literature did not achieve consensus in terms of whether political risk had an inherently negative or positive impact (Clark and Tunaru, 2003). Indeed, Kobrin (1979) criticizes the viewpoint that political risk has inherently negative consequences as “an implicit normative assumption that may not be universally valid.” Robock (1971) also believes that “political risk can result in gains as well as losses”. However, as it is currently conceived of, political risk is usually used to refer to the probability of negative impacts.

To return to the notion of probability of default, Eaton, Gersovitz and Stiglitz (1986) provide a more detailed definition of the term in two-period model as “whenever the borrower gives resources to the lender that are less than the fixed amount that he is committed to pay the lender,” and in a multi-period model as “whenever the lender formally declares that the borrower has violated a certain condition of the loan.” Importantly, the authors highlight the fact that “in most situations, a default is a result of

a set of decision, not the mechanical realization of some outcome” (Eaton et al., 1986, pp. 3). This makes a necessary distinction between the *inability* to pay, and the *unwillingness* to pay.

To understand the ways in which this theoretical background is applied in practice – and therefore how social media metric could be incorporated going forward – it is also useful to examine the country risk measures employed by leading financial institutions and analysts. It is important to note that many of these firms, as a logical consequence of their business model, do not fully disclose all of their metrics – particularly with respect to the more qualitative facets of their analysis, which are often at play in political risk analysis. Nonetheless, understanding the standard operating approaches within the industry is a valuable starting point for this study, which seeks to address a potential quantitative gap in the political risk analysis status quo.

Iranzo’s (2008) report on country risk analysis is a comprehensive discussion of prevailing country risk approaches and industry best practices. As the former head of the Country Risk Service for the International Affairs Office of the Bank of Spain, Iranzo is well versed in strategies and applications of country risk and her work really serves as a literature review in itself. As Iranzo (2008) is quick to point out, country risk is a broad concept that covers many complex and overlapping subject areas. Through many economic transmission channels, what begins as political risk can quickly spread into transfer risk to affect private and sovereign debt alike. However, in an effort to understand the leading industry approaches to political risk analysis, this portion of the literature review tries to keep the fields in discrete parts, to the extent possible.

As Iranzo (2008) defines it, political risk “derives from specific political actions or decisions, and it is therefore one of the types of risk included in the more general concept of country risk.” Iranzo goes on to acknowledge that even within the field, political risk is sometimes used interchangeably with country risk. This study uses to the more specific definition above, in addition to other relevant definitions or concepts provided in this section. Iranzo (2008) also acknowledges that country risk – and political risk in particular – “inevitably rests on human judgment”, also referring to the study as “an unsystematic discipline”. The field as a whole has sought overtime to standardize and quantify as much of the analysis as process, which is the primary reason that this study seeks to examine the possible uses of social media metrics in one of the remaining qualitative-heavy facets of country risk – i.e., political risk. As Iranzo (2008) reports, the Export-Import Bank classifies country risk analysis into four models, which serve as categories to discuss industry practices (see Figure 2):

- Purely qualitative;
- Structured qualitative, with some statistical data;
- Structured qualitative with scoring techniques (or checklists); and
- Quantitative (econometric and highly structured)

Not surprisingly, Iranzo (2008) points to qualitative and structured qualitative as two categories for political risk analysis. In terms of the latter, Iranzo (2008) makes special mention of the Business Environment Risk Intelligence (BERI) political risk index. Given their semi-structured and structured approaches, the structured-qualitative and quantitative models in particular stand to benefit from the integration of social media data analytics, though purely qualitative assessments could also make use of this data in painting a picture of a country’s general risk environment.

Figure 2: *Four models of country risk analysis*

	Purely qualitative	Structured qualitative (some stats)	Structured qualitative (scoring)	Quantitative
<i>Type of analysis</i>	Descriptive, unstructured; lacks fixed structure	Qualitative system built on a standard structure/format	Qualitative with checklist or scoring method	Econometric, highly structured
<i>Description</i>	Description of a political or economic situation; prospects for the external sector	Defined set of variables, but qualitative evaluation Enables more comparisons than purely qualitative analysis	Defined series of political, macroeconomic, historical indicators; allocate score within a set range to each; weights assigned	Quantitative assessment of set variables; often a scoring method Alert ceilings (debt-to-GDP, import cover ratio, etc)
<i>Where used</i>	Often within multinational companies	Multinationals, banks	Banks, ratings agencies	Banks, international institutions
<i>Examples</i>	“Old hands”: get advice from experts (e.g., journalists, diplomats) “Grand tour”: send representatives to country to gain insight	BERI Political Risk Index	Weighted checklist approach PRS International Country Risk Guide	OECD Country Risk Assessment Model Probability of default Early crisis warning systems

Within country risk, Iranzo (2008) highlights six core variables: political, macroeconomy, economic structure, the banking sector, the external sector, and market assessments. In terms of the political category, Iranzo (2008) lists some of the indicators to be considered: political stability, government effectiveness, strength of institutions, risk of internal/external conflict, proximity or inclusion in geography that is conflictive, degree of corruption, and debt payment culture – citing the World Bank’s Governance Indicators (WGI) as a chief measurement for political risk as they cover government

policy and succession, among others. However, many of the models cited are confidential – notably, the OECD Country Risk Assessment Model – which limits the ability to question the variables used and relative weights assigned.

To demonstrate how political risk can transfer to other areas of the economy, Iranzo (2008) outlines a typically transmission:

“Political events may also trigger a confidence crisis in the country, giving risk to the withdrawal of foreign funds and the flight of domestic capital, causing the subsequent devaluation of the currency. Examples of such events are a radical party electoral victory, a coup d’état, or the outbreak of a war with a neighboring country. In all these cases, domestic and foreign capital flight in search of safe havens prior to the crisis, or once it has broken out, is an aggravating factor, as it represents an additional source of foreign currency demand, with the resulting deepening of the foreign exchange crises.”

Understanding the fragmented and varied approaches to country risk across regulators and private ratings agencies is key to the study of country risk – and underscores the potential to improve upon these processes with newly available measures, such as social media metrics. Iranzo (2008) emphasizes that there is no common regulation at the EU or eurozone level. Within the US, however, the Interagency Country Exposure Review Committee (ICERC) meets annually to coordinate studies and assessments of country risk in exposure areas. Iranzo (2008) concludes her study with an acknowledgement of the shortcomings of country risk forecasts, and calling for improvement in knowledge and in analytical tools to assess risk. Clearly, at least from the perspective of an industry expert like Iranzo, there is openness in the field to new approaches.

Erb et al. (1996) also focus on industry approaches to political risk analysis, albeit about 20 years prior to Iranzo's (2008) writing. However, many of the approaches detailed by the authors are still leading methodologies, and are worthy of review. Erb et al. (1996) cover five measures of country risk: political, economic, financial, composite and Institutional Investor country credit ratings. In their study of these five categories, the authors found that political risk analysis offered the least amount of information, on average, about the future (as opposed to financial risk, which offered the most).

The authors cite Political Risk Services' International Country Risk Guide (ICRG) as a key resource for risk analysis, as the ICRG is updated monthly with political, economic, and financial score, as well as a composite index. As Erb et al. (1996) explain, the ICRG analyzes political risk according to "subjective staff analysis of available information" – the key word being *subjective*. The guide includes 13 political risk factors (see below list), for a total of 100 points and a 50% weight in the composite index – double that of both financial and economic, which demonstrates the importance of political risk, even if it remains a largely subjective field. The 13 political risk factors are: economic expectations vs. reality; economic planning failures; political leadership; external conflict; corruption in government; military in politics; organized religion in politics; law and order tradition; racial and national tensions; political terrorism; civil war; political party development; and quality of the bureaucracy.

Figure 3: ICRG Composite Rating System (Nath, 2013)

CHART A.II — *Example of Checklist Method*

The ICRG Composite Rating System

<i>Political</i>	<i>Weight</i>	<i>Financial</i>	<i>Weight</i>
Economic expectations versus reality	6%	Loan default or unfavorable loan restructuring	5%
Economic planning failures	6%	Delayed payment of suppliers' credits	5%
Political leadership	6%	Repudiation of contracts by government	5%
External conflict	5%	Losses from exchange controls	5%
Corruption in government	3%	Expropriation of private investments	5%
Military in politics	3%	Total Financial Points	25%
Organized religion in politics	3%	Economic	
Law and order tradition	3%	Inflation	5%
Racial and nationality tension	3%	Debt service as a % of exports of goods and services	5%
Political terrorism	3%	International liquidity ratios	3%
Civil War	3%	Foreign trade collection experience	3%
Political party development	3%	Current account balance as % of goods and services	8%
Quality of bureaucracy	3%	Parallel foreign exchange rate market indicators	3%
Total Political Points	50%	Total Economic Points	25%
Overall Points			100%

Source: Erb *et al.* (1996) and Harvey (1996)

The structured-qualitative BERI political risk index evaluates some of the concepts seen in the ICRG checklist.

Figure 4: BERI Political Risk Index (Nath, 2013)

CHART A.I — *Example of Structured Qualitative Method*

The BERI Political Risk Index

<i>Components</i>
Political Factionalization
Linguistic/Ethnic/Religious Tension
Coercive Measure to Maintain Regime
Mentality : Nationalism, Corruption, Nepotism
Social Conditions : Population, Income Distribution
Radical Left Strength
Dependence on Outside Major Power
Regional Political Forces
Social Conflict
History of Regime Instability

Source: Harvey (1996), Appendices

These lists overlap in some areas with the World Bank's WGI's cited by Iranzo (2008). Specifically, the WGI's focus on voice and accountability; political stability and absence of violence; government effectiveness; regulatory quality; rule of law; and control of corruption (World Bank, n.d.). Again, perhaps the most important thing to emphasize across these different political risk measures is their highly qualitative nature, and limitations therein, which will be discussed in further detail in Chapter 6. This underscores the potential for finding other sources of quantitative data, such as social media metrics.

2.2 Sociology: Network theory and information studies

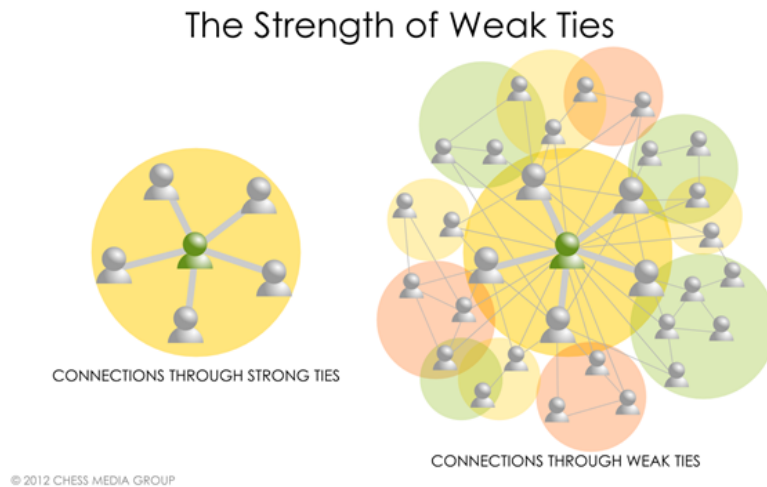
To understand the dynamics of social media, it is crucial to look at the underlying sociological theories – in particular, those focused on network theory and information studies, and how they help explain the role of social media in collective action. Network theory is a particularly helpful lens through which to view social media, as it focuses on the flow of information through a network, as well as the links between different actors within said networks. Granovetter (1973) and Skinner (2011) provide different but useful lenses for the study of social media networks – both of which help inform the analysis of these platforms.

Although published well in advance of the invention of the internet, Granovetter's (1973) seminal work on the propagation of information within social networks remains a central and frequently cited text for understanding the proliferation of information in social media. Information flow is central to any study of social media and political risk,

as the spread of information – both as an alternative to conventional media, and as a means of mass mobilization – is a core mechanism of social media in this context.

In his piece, Granovetter (1973) argues for the strength of so-called “weak ties” in the spread of information across a network. He defines the strength of ties as a function of a “combination of the amount of time, the emotional intensity, the intimacy (mutual confiding), and the reciprocal services which characterize the tie.” In this sense, weak ties would be those that are characterized by the previous factors to a lesser degree – that is to say, those that “entail more limited investments of time and intimacy, subsuming an array of social acquaintances” (Ruef, 2002).

In his view, strong ties typically develop between parties with a great degree of homogeneity and social overlap. In this sense, when information spreads across this type of tie, it does not tend to travel far beyond the central strong-tie nucleus of the network. Conversely, weak ties allow for information to spread between less similar groups, thereby propagating the information further out in the network. Figure 4 below shows the basic premise of strong and weak ties in a network structure.

Figure 5: *Strength of Weak Ties*²

While Granovetter’s (1973) study is often cited for its work on the spread of information regarding employment opportunities, it also has widespread applications for social media network theory as well as collective action mobilization strategies, as weak ties are a prime dimension of social media networks. Indeed, Granovetter argues that a lack of weak ties in networks can help explain failures of mobilization. While he uses instances of failed community action as example, the theory can also be applied to present-day social movements and their mobilization structures.

The field of information studies (IS) is also useful for understanding social media, as it “is concerned with the origination, representation, organization, retrieval, and utilization of information” – all of which is central to the ways in which social media is created, recorded, and interpreted (University at Albany, n.d.). Skinner (2011) is a key resource for understanding social media research from this perspective. She presents a

² Morgan, J. (2012) The DARPA Collaboration Experiment and What it Means for Your Organization. *The Future Organization*. Retrieved from: <http://www.thefutureorganization.com/darpa-collaboration-experiment/>.

helpful review of three IS paradigms that are useful for studying social media in conjunction with political uprisings.

The three paradigms explored – physical, cognitive, and social informatics – also mirror the development of the IS field from that of a study of information as a “tangible and unchanging” object, to one influenced by context and perception. The physical paradigm is primarily useful from an information propagation and retrieval standpoint, while the cognitive paradigm focuses more on micro-level interpretation of information. Lastly, the social informatics paradigm tries to reconcile the two by aggregating individual interpretation to the group level in order to achieve a better understanding of various cultural and institutional contexts. In social media, when thoughts and perspectives are transcribed and interpreted by a wide variety of individual users – from difference nationalities, backgrounds, mother tongues, etc. – perception is crucial to understanding the mechanisms at work.

While elements of all three paradigms are pertinent for understanding social media, the social informatics paradigm seems the most useful for this study, as the influence of social media in political movements naturally is a function of individual interpretation aggregated at the group level. Interestingly, Skinner (2011) also briefly touches upon the repression paradox of social media and internet censorship during the Egyptian uprisings, advocating the further study of the use of alternative technologies (fax machines, band radios, etc.) and their impact on information and behavior.

While the previous two authors provide different but useful lens for the study of social media networks, they do not stand in opposition to one another – rather, they both inform the analysis of social media. However, the literature focusing more precisely on

social media's role in collective action shows two relatively distinct schools of thought that can be roughly summarized as Shirky vs. Gladwell and Morozov.

Clay Shirky, NYU professor and prominent techno-optimist, focuses on the enabling power of social media for collective action. While he does not believe that social media inherently guarantees either widespread mobilization or successful protest movements, he sees it as the latest in a line of cost-reducing communication tools ranging from Gutenberg's printing press to the *samizdat* use of photocopiers in Soviet bloc countries.

This view stands in stark contrast to that of both Malcolm Gladwell and Evgeny Morozov – stalwarts of the techno-pessimist school of social media. In response to Shirky's work, Gladwell wrote a piece in *The New Yorker* in October 2010 in which he vilifies the classic successful social media mobilization examples of uprisings in Moldova and Iran as false. This “outsized enthusiasm for social media” or “digital evangelism,” in his words, undermines the historical importance of strong ties and high-risk activism. In Gladwell's (2010) view, comparing “liking” the Save Darfur Facebook page with black youths sitting at the counter of a segregated restaurant in Greensboro in the 1960s is to drastically overstate the power of social media to transform low-risk participation into high-risk activism. His argument also suggests that low-cost online activism, also known as “slactivism,” tends to replace rather than supplement actual participation in social movements.

Gladwell (2010) also underscores other weaknesses in the social media mobilization theory, such as its lack of hierarchy. While he sees networks as strong in terms of low-risk activism or the spreading of information, he views them as too weak

and messy to think strategically, reach consensus, or achieve discipline. Gladwell puts it eloquently when he states that social media “makes it easier for activists to express themselves, and harder for that expression to have any impact.”

In response to Gladwell’s critique, and fortuitously concurrent to the start of the Arab spring, Shirky (2011) wrote a piece in *Foreign Affairs* detailing numerous examples of the use of social media and text messaging to coordinate “massive and rapid responses.” Shirky (2011) is; however, quick to acknowledge that the use of social media does not guarantee a successful outcome, and that academic debate on the issue too often degenerates into “dueling anecdotes.” Rather, Shirky sees social media as having the same strength as communication technologies before it (e.g., photocopiers, telephones, telegraphs, the postal service) – in particular, thanks to their wide usage in society. To that end, the more widely these technologies are adopted, the greater the “dictator’s dilemma” and potential for backlash if they are banned.

Morozov is another high profile techno-pessimist that writes about social media and society. He is known for his litany of opinion pieces criticizing the narrative of internet as mankind’s savior, and is a frequently cited thinker when it comes to social media and revolution. In his book *The Net Delusion: How Not to Liberate the World*, Morozov looks at the ways in which the internet is actually supporting authoritarian regimes rather than undermining them. Morozov (2011) is a key resource for the discussion of the dark side of social media data collection in the wrong hands.

Understanding the viewpoints of these three authors is important, as the relative value of social media data for political risk analysis necessarily depends upon the power of these platforms to affect change. The ongoing discussion in the press and academic

journals between the Shirky and Gladwell/Morozov camps is a fascinating dialectic that highlights just how far from consensus the academic community is on this issue.

However, whether social media is viewed as a savior of mankind, just another mode of communication that facilitates mobilization, or even an instrument for state repression, all three authors implicitly accept that social media is a tool to be reckoned with. As such, for the purposes of this study, the arguments of both camps are given credence.

Furthermore, in an effort to bypass Shirky's (2011) caveat of "dueling anecdotes," this research focuses primarily on the utilitarian aspects of social media, foregoing the Thomas Friedman-esque naïve triumphalism too often seen in opinion pieces on the subject.

Any study of big data must question the validity and authenticity of said data before attempting to apply it. While this will be discussed in greater detail in Chapter 6, it is worth briefly mentioning some of the leading voices in this sub-field of ICT. To best understand the potential corruptibility of sources of social media metrics in the digital age, this study heavily references Dr. Lutz Finger's work on the subject of social media spam. In a speech at the 2013 Social Media Measurement and Monitoring Conference, Finger details the scope of spam bots – computer programs tasked with propagating information of seeking to influence members in a social network – in the internet today. According to Finger (2013), much of the activity on the internet is actually not generated by human users. Fingers (2013) explains that an estimated 51% of web traffic, 38% of online reviews, and 7% of Twitter users are not the result of human users, but rather computer programs known as "bots." While the evolution of these programs and potential applications in social media will be discussed in greater detail later in this study, Finger

(2013) is a crucial resource for highlighting the potentially coercive influence of these programs, which are often excluded from current academic discussions of social media movements.

Cybenko et al.'s (2002) paper on covert cognitive hacking – i.e. the “provision of misinformation, the intentional distribution or insertion of false or misleading information intended to influence readers’ decisions and/or activities” – provides a foundation for understanding the basic response options to covert hacks. While their paper focuses primarily on single source attempts to gain access to user information systems, as opposed to the sophisticated multi-source hacking attempts being run by Finger’s (2013) spam bots, the basic premise is useful for understanding the present limitations of prevention and detection in social media networks – which is central to evaluating the integrity of social media metrics.

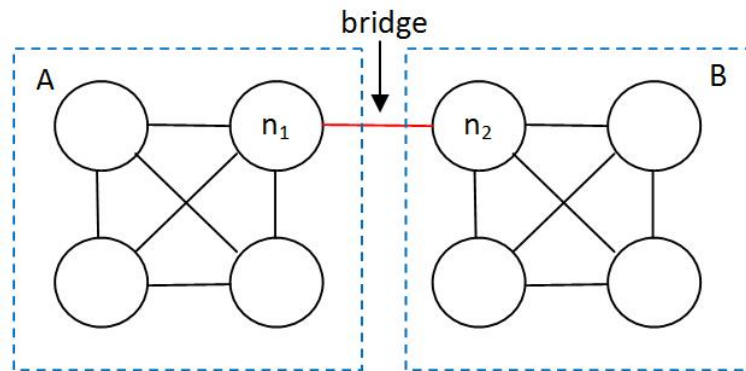
Lee et al. (2013) likewise discuss the disruptive potential of hacking or spam, while introducing some helpful vocabulary into the discussion. Terms such as “astroturfing” and “crowdturfing” – plays on word from the mass mobilizations terms “grassroots” and “crowdsourcing,” respectively – highlight how spam bots attempt to influence mass user behavior through social networking platforms. Lee et al. (2013) define crowdturfing as “leverag[ing] human-powered crowdsourcing platforms to spread malicious URLs in social media, form ‘astroturf’ campaigns, and manipulative search engines, ultimately degrading the quality of online information and threatening the usefulness of these systems.” Unlike Cybenko et al.’s (2002) single source cognitive hacking, social media allows cognitive hacking to occur on a much wider scale. Particularly useful for the context of this study is the authors’ example of crowdturfing by

Vietnamese propaganda officials who hired 1,000 crowd-turfers to hold online discussions and post comments supportive of the Communist party, demonstrating how states as well as political movements have the ability harness the power of social media to try and influence thought and propagate information advantageous to their cause.

Gonzalez-Bailon and Wang (2013) combine network and brokerage theory to discuss the key role played by “bridges” in filling structural gaps in online networks. Networks in their mind are “both a structure that reflects organic forms of organization and the channels through which information flows”, continuing on to explain that “when analyzed as conduits for information flow, online networks facilitate information diffusion only under very specific circumstances” (Gonzalez-Bailon and Wang, 2013).

By that logic, networks will naturally “exhibit holes that hamper diffusion and information flow”, requiring bridges – which they define as “a minority of ties spanning structural holes” – and information brokers – “minority of users that make use of those bridges to disseminate information” – to perform the necessary role of filling said gaps (Gonzalez-Bailon and Wang, 2013).

The concepts of bridges and information brokers are not particularly well illustrated by the authors, or by the field of network theory more generally. However, the following graphic demonstrates the function of a bridge, defined as the link without which two groups would not be connected. An information broker, in turn, is also a bridge – set apart by the fact that it makes use of this connection to spread information. Simply put, the difference between a bridge and an information broker is akin to the difference between having two friends from different social circles, and setting those two friends up on a blind date.

Figure 6: *Interpersonal bridges*³

According to Gonzalez-Bailon and Wang, these information brokers are characterized by their “larger-than average networks and their higher-than average levels of participation”. In layman’s terms, they are the social butterflies, or gossips, of a network. In network theory terms, Gonzalez-Bailon and Wang describe brokers as nodes with low structural constraint and high betweenness centrality – which is a technical way of discussing Granovetter (1973) strength of weak ties uniting otherwise distant groups in a network. The concept of bridges versus information brokers is echoed in political science, which argues that bridges may or may not be successfully activated (i.e., transformed into information brokers) depending on the situation. This will be reviewed in greater depth in Section 2.3 of the literature review.

Gonzalez-Bailon and Wang (2013) examine the social media links, primarily via Twitter, between the Occupy movements and the *indignados* in Spain to test their hypotheses. After identifying the top-five hashtags related to the indignados movement,

³ [http://en.wikipedia.org/wiki/Bridge_\(interpersonal\)](http://en.wikipedia.org/wiki/Bridge_(interpersonal))

the researchers collected a total of nearly 450,000 tweets spanning from April 30 to May 30, 2012, which was a equivalent to a network of over 38 million users. By using retweets (RT) and mentions (@), they were able to visualize the flow of information and interaction between network participants. Their analysis revealed a highly uneven distribution of centrality, with “a minority of users concentrat[ing] most of the connections, and again a minority attract[ing] and send[ing] most of the messages” (Gonzalez-Bailon and Wang, 2013). This typifies influential user models that are often seen from popular demonstrations, such as the Gezi park protests (see Chapter 6), whereby most social network participants are passive consumers of the information, while an elite group of influencers propagates the majority of tweets. This is akin to the difference between scrolling through the Facebook news feed to see what others are saying or doing, versus actively posting one’s own content.

Although this thesis focuses on the type of data collected via social media and the ways in which it could be utilized for political risk analysis, a clear understanding of the sociological theories at work – in particular, network theory and information studies – is a necessary starting point for any discussion of social media. Indeed, the flow of information, and the network actors that make it possible, are indispensable to social media. Likewise, it is important to understand the potential for these ties and networks to be used as channels for coercive behavior by both humans and programs. Last but not least, there is still no consensus in the field on the importance of social media relative to previous modes of mass communication, as exemplified by the Shirky and Gladwell/Morozov debate.

2.3 Political science: Social movement theory, the Arab spring and the Gezi park protests

While this thesis seeks to focus on the business applications of social media metrics, political science – and in particular, collective action and social movement theory – plays a central role in linking social media and country risk. Understanding the underlying theory of collective action, as well as what has been written in the field to date about social media’s role in the Gezi park protests and the Arab spring in particular, is crucial for this study.

At the heart of these foundational theories are Tarrow (1994) and Tilly (2003), as well as Brockett (1993). Indeed, any academic work on collective action and social movements inevitably makes reference to Tarrow (1994) and Tilly (2003). In particular, these authors are incredibly useful for understanding the necessary conditions for social movements to emerge, as well as the strategies they may apply. Despite writing in advance of the widespread adoption of internet as a social movement tool, Tarrow and Tilly’s theories are directly applicable to internet-age protests, as well. Brockett (1993) adds to this foundation, expanding on the repression paradox and its relationship to the protest cycle.

Tarrow (1994) focuses on the prerequisites to social movements, which he views as necessary but not sufficient to ensure sustained collective action. Tarrow (1994) highlights political opportunity structures, social networks, forms of collective action, and cultural framing as crucial variables. Within the category of political opportunities, Tarrow (1994) identifies five structures that can facilitate social movements: 1) increasing access, 2) shifting alliances, 3) divided elites, 4) influential partners, and either 5) repression or facilitation. This last category in particular speaks to the “repression

paradox” that is an important theme of this study and is further discussed in the work of Brockett (1993).

Tilly (2003) expands on Tarrow’s (1994) “forms of collective action,” which he refers to as “repertoires of contention.” These can include anything from violent riots to peaceful sit-ins. In collective action theory, scholars often see the same repertoires repeated across very different types and settings of protests. In the Arab spring for example, similar Arabic protest chants spread from Tunisia to Egypt and internet memes were widely propagated via Twitter and Facebook.

Brockett (1993) focuses on the repression paradox, which is an important concept for any discussion of social media and collective action that challenges the state. The repression paradox essentially states that state repression – such as internet censorship, violence, etc. – will result in either the successful repression of the social movement or the radicalization and spread of public discontent and mobilization. Essentially, it is a high-risk/high-reward state response to unwanted public behavior. As we have seen in the case of Egypt, the repression paradox, sparked by the shutdown of mobile phones and the internet, was an important catalyst for the eventual overthrow of the government. Likewise, with Turkey’s growing state opposition to and blocking of social media platforms such as YouTube and Twitter, the repression paradox remains an integral theme connecting social media and country risk.

Brockett (1993) goes beyond a description of the repression paradox and seeks to identify patterns between the repression and the outcome. He found that the most important determinant was the temporal location in the protest cycle at the time that state repression was initiated. If the protests were in their infancy, state repression was more

likely to successfully deter mobilization. However, if the repression is targeted during the ascendant phase of protests, it is much more likely to provoke mass oppositional activities, as was the case in Egypt. Although Brockett (1993) comes to this conclusion long before the internet and social media became central tools for modern collective action, his conclusions nonetheless remain important for any study of today's protest movements and state responses, as well as the ethical dilemma surrounding the instrumentalization of social media metrics (see Chapter 6).

Scholars, researchers and pundits alike have emphasized the crucial role played by social media in the Arab spring, citing the example of the Egyptian man who named his daughter "Facebook" as homage to the site (CNN, 2011). As the dust has cleared in the aftermath of the regional upheaval, soaring rhetoric extolling the liberating power of social networking sites (SNS) has given way to more scholarly, sometimes quantitative, analysis on the social media dynamics of a given protest movement.

Starbird and Palen's (2012) paper on the 2011 Egyptian uprising examines, in particular, Twitter's retweet mechanism – "a mechanism by which Twitter users pass on or forward information to other users" – as a vehicle for collective framing and information propagation. The study also focuses on the interplay between those actually "on the ground" in Cairo and those supporting the uprising from their computers, via the roles that each group played in disseminating and spreading information. Starbird and Palen (2012) found that 30% of the 1000 most highly retweeted users during the protests were actually in Cairo at the time, and were often writing about injuries, troop movements, the supplies that were needed, etc.

However, they are also quick to point out that the remaining 70% were not necessarily purposeless “slacktivists” as Gladwell or Morozov might perceive them. Even if they were not on the streets or even in Egypt, they played a key role as filters in processing and recommending information to the wider Twittersphere. This mechanism is closely related to Granovetter’s (1973) strength of weak ties, whereby weak tie relationships in networks further the spread of information, as well as Skinner’s (2011) social informatics paradigm and the bridges and information brokers discussed by Gonzalez-Bailon and Wang (2013).

Unlike a formal calling trees or other mobilization strategies of strong-tie social movements, the connections established through social networks are frequently weak ties. Facebook, as an SNS that relies on individuals actively requesting to link to other individuals, arguably is set up to facilitate stronger ties than blogs or Twitter, where users can more passively consume information. Indeed, these latter SNS often do not require the same level of user-to-user approval to view information as Facebook. Furthermore, according to a Topsy official, Facebook does not have a comparable firehose of data that can be accessed, as in the case of Twitter, which limits Facebook’s usefulness for outside analysis (Bertolucci, 2012).

That said, the use of Facebook groups – which can be set up as open for all interested parties to follow without permission required – played a large role in mobilization and the propagation of information in the Arab spring. In the case of the 2011 Egyptian uprising, the Kolenah Khaled Said Facebook group, which was one of the first to call for the January 25th protests, was founded in 2010 in the wake of the brutal murder of Khaled Said, a young Alexandrian, by Egyptian police. The group – whose

name translates to “We are all Khaled Said” – argued that what happened to Khaled could have happened to any Egyptian, so vast were the powers of the state security apparatus. Indeed, the author of this study, who was living in Cairo at the time of the 2011 uprising, found out about the Police Day (Jan 25) protests from the Khaled Said page, which actually created an event for the revolution and asked users to RSVP (Trottier and Fuchs, 2014).

While sometimes lumped into the Arab spring category, the Gezi park protests in Istanbul, Turkey were a distinct event, occurring more than two and a half years after the unrest began in Tunisia, and – like the Turks themselves – not Arab. Beginning in the spring of 2014, the protests were sparked by urban planning policy aimed at pedestrianizing Taksim square’s Gezi park in an effort to build, among other things, a multi-purpose building that would include a ground-floor mall, luxury housing, and a mosque.

The plans also entailed rebuilding Ottoman-era military barracks and demolishing the Ataturk (founder of the Republic of Turkey and icon of the secular zeitgeist) Cultural Center. Beyond the opposition to radically altering one of the city center’s last green spaces, the selection of the Kaylon Group – known to have close ties to the Erdoğan administration – as the project’s main contractor helped to spark public outrage. The existing and proposed layouts of the park and the square can be seen in Figures 6 and 7 below.

Figure 7: Gezi park and Taksim square⁴

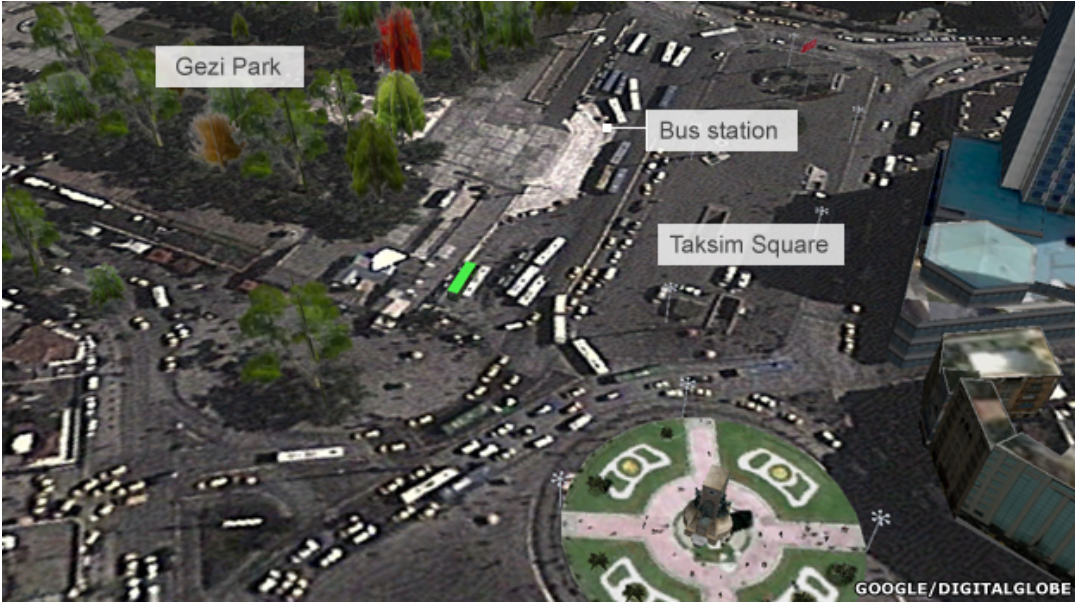


Figure 8: Proposed redesign⁵



⁴ <http://www.bbc.com/news/world-europe-22753752>

⁵ Ibid.

The events of these protests, which included millions of people and continued over several weeks, will be examine in depth in Chapter 6. In particular, the research of the NYU Social Media and Political Participation (SMaPP) lab chronicles the development of the protests from a social media perspective in an effort to demonstrate the ways in which social media data could potentially be applied to the field of political risk analysis.

3. Research Question

Once social media data is collected, the question remains: what can be done with it? Companies have targeted users with online advertising based on their preferences and interests, and the government has worked to ascertain terrorist links based on online behavior. Even the US Library of Congress, despite lacking the technological infrastructure to effectively carry out keyword searches, is collecting every tweet ever posted, in the hopes that one day the data set can be put to use (see Section 5.3).

Our ability to collect data continues to outpace our capacity to analyze it – particularly in real time. This trend, labeled as Data as a Service (DaaS), has seen data go from being a basic good to a value-added service (Wagner, 2014). As the sophistication of analytics inevitably evolves, what other applications could there be for the veritable treasure trove of social media data out there? For example, could the January 2011 Egyptian uprising and its financial impact have been better understood, or even predicted, by looking not just at the pro-Western narrative of Facebook or Twitter usage, but at the actual data behind it – perhaps even in real time?

There is certainly financial incentive to better understand these complex political events. In addition to lower economic growth and higher inflation, the 2011 Egyptian uprising engendered a number of other costly impacts on the country's economy. Unemployment figures saw a 3.2 percentage point increase from 4Q2010 to 4Q2012. From a financial standpoint – of particular interest to this study – the Egyptian Central Bank reported USD481.1 million and USD3.3 billion in FDI and portfolio investment outflows, respectively, in the second half of 2011 alone (Singh, 2013). The carry-through effect on the foreign exchange led the Egyptian pound to fall to an historic low of 1:0.16

to the dollar, despite heavy central bank intervention (RT, 2012). Efforts to prop up the pound saw foreign exchange reserves tumble to about one third of their pre-revolution value, degrading the nation's import coverage ratio (excluding gold) to just over one month of imports – a critical mass for a nation reliant on dollar-denominated wheat imports to feed its growing population (Singh, 2013). These transmission channels of political risk to financial markets will be discussed in greater detail for the Gezi park protests in Chapter 6.

The domino effect of political risk takes time to transmit through the economic system to impact key measures such as GDP forecasts or FDI stock, which are at best revised on a monthly basis. Indeed, as economic indicators are updated far less frequently than Twitter feeds, it can be challenging for political risk analysts to understand, and more importantly measure, what is taking place on the ground. To that end, this study seeks to answer the following research question: *What type of social media data is available from Twitter, and what are the limitations to its application in the field of political risk analysis?*

4. Methodology

As this is a new area of research, the study seeks to paint a picture of the current state of the field and explore some of the possible limitations of applying social media data to country risk analysis. In the absence of comprehensive Twitter data⁶, a qualitative approach has been used, with the hopes that further research will be done to advance the subject in the future. The qualitative method employed in the study centers on bibliographical research, with a closer look at some case studies of historical social media analysis of the Egyptian uprising of 2011 and the Gezi park protests of 2013.

The study breaks the research question into its two main parts: availability and limitations. Firstly, in Chapter 5 the study details the data points that Twitter collects from users, to establish a basis for what information is actually available from this platform. Next, it discusses some of the leading providers of Twitter data, other players in the field, and the costs associated with acquiring data. The chapter also looks at the type of financial data available that could, in a future study, be examined for correlation with Twitter data to demonstrate the transmission of risk to financial markets. As macroeconomic indicators are updated too infrequently to be of use for real-time political risk analysis of an event, more reactive information, such as movements in the stock market, the foreign exchange market, interest rates, default insurance premiums, bank deposits, and tourism figures, could be more useful.

To address the secondary question of the limitations of using social media metrics for political risk analysis, Chapter 6 first looks at the shortcomings of the current

⁶ Twitter Search and related, free platforms only selectively return results. The only comprehensive (100% of results) sources of Twitter data are controlled by a handful of firms – one of which was recently acquired by Twitter – that sell data on a business-to-business basis for exorbitant prices. See Chapter 5 for greater detail.

measures of political risk being used in the industry. Next, the study examines some of the ways in which Twitter data has been analyzed by researchers to date and how these can add value for political risk analysts. This includes the work of the NYU Social Media and Political Participation (SMaPP) lab, as well as the growing use of microblog sentiment analysis. Then, the chapter details some of the possible activities that can undermine the veracity of Twitter data, before concluding with a brief discussion of the ethical dilemma of advancing the application of social media metrics to political risk analysis. Lastly, Chapter 7 provides conclusions of the discussion.

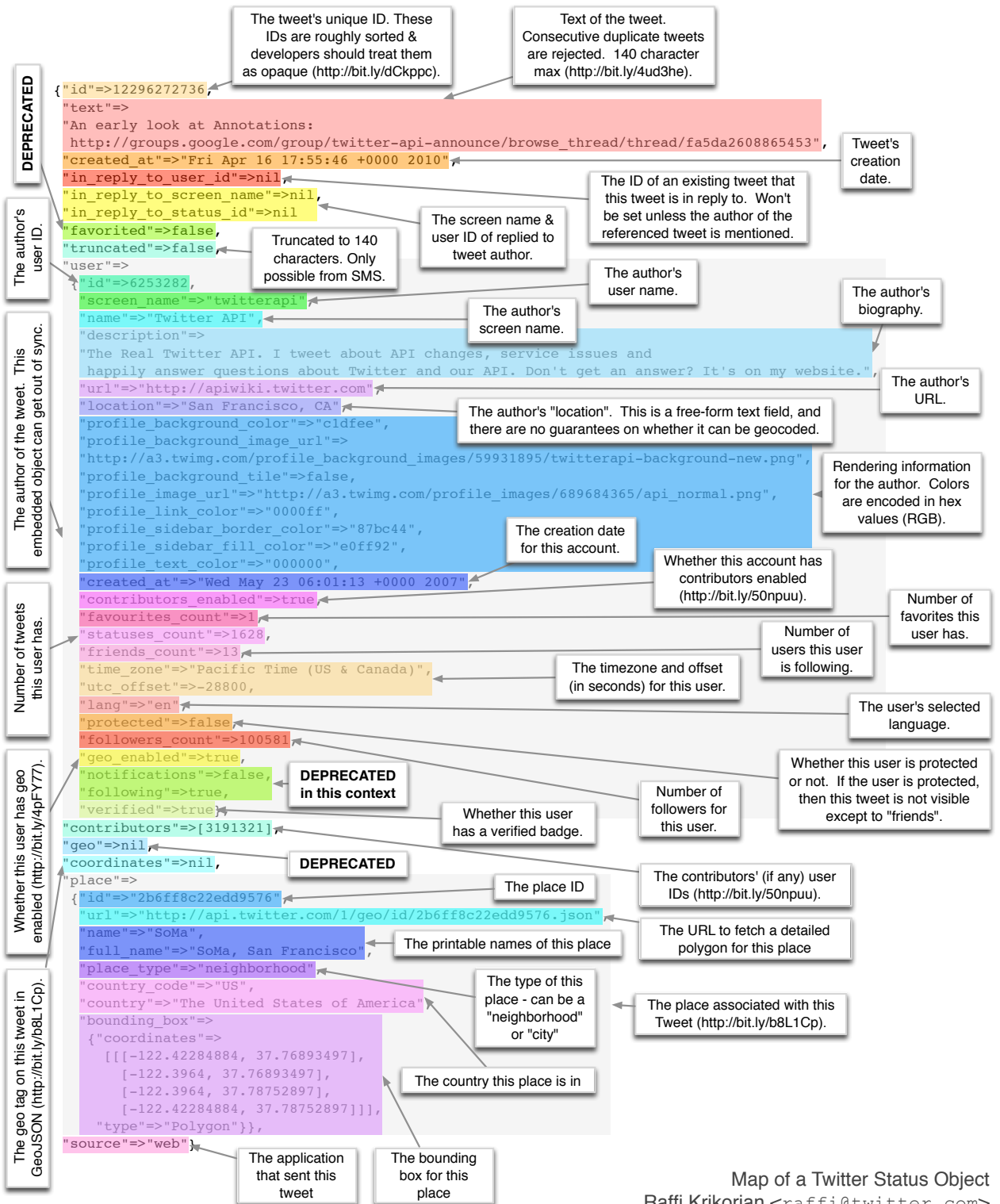
5. Data Description

This section describes in detail the type of data that can be collected from Twitter, with a focus on the metadata of tweets. Next, it discusses the leading sources of Twitter information today, some other players in the data collection field, and the pricing for access to these data streams. Lastly, the chapter highlights what types of financial data could potentially demonstrate the transmission of political risk from social movements to capital markets.

5.1 Twitter data

To understand what kinds of data Twitter collects about users and their activity, it is important to first distinguish between *content* and *metadata*. While content is the 140 characters of a tweet; metadata is defined as “data that describes other data” and “provides information about a certain item’s content” (Tech Terms, n.d.). In the case of Twitter, there are numerous points of metadata that are collected for each tweet. In 2010 a Twitter employee created a map explaining the different sets of metadata collected (see Figure 8 on the next page, which, along with the information from Twitter’s most recent privacy terms, has been summarized in this section (Krikorian, 2010; Twitter Privacy Policy, n.d.).

Figure 9: Twitter metadata map



For each tweet, there are a variety of identification numbers – or IDs – that are collected. The tweet itself has a unique ID number, as does the author of the tweet. Likewise, each reply, favorite, and retweet is assigned its own ID number. In terms of information about the author, each tweet contains the date the account was created, the URL given by the author (if any), the author’s Twitter biography (if given), as well as the number of followers they have, the number of users they are following, and the number of favorites they have. Twitter metadata also indicates whether other contributors are allowed on the author’s account.

From a content standpoint, Twitter also obviously collects the text of the tweet itself – a maximum of 140 characters – as well as any links or photos included. The text of the tweet often includes hashtags, which Twitter explains are “used to mark keywords or topics in a Tweet. It was created organically by Twitter users as a way to categorize messages” (Using Hashtags on Twitter, n.d.). Social media researchers and analysts are also able to use hashtags as content proxies. The language indicated by the user’s profile is recorded, as is the language of the tweet itself, which can differ.

A variety of other metadata is collected from the tweet pertaining to location. For starters, many authors indicate their geographic location in their profile, which is recorded. Furthermore, users may indicate a location in the tweet itself, which is also collected. The metadata further records what type of place the location is – e.g., a neighborhood, a city, etc. If the user has selected a trend location, which can differ from their actual location, this will also be collected. If location services are enabled, a geocode location will also be collected from the tweet. Additionally, Twitter is able to determine a user’s actual location based on the cell towers or wireless networks in

proximity or the IP address. A bounding box is also collected for the location given, which includes GPS longitude and latitude coordinates.

The timestamp of the tweet, in seconds, is another data point that Twitter collects, along with the application it is sent from, or the internet browser used. Log data can also be collected by Twitter, including operating systems, mobile carriers, device information (device and application IDs), search terms, web pages visited, cookies, and the referring web page. Many users are also unaware that the Twitter buttons and widgets integrated on many third-party websites also track user movements on the internet, regardless of whether users engage them. Last but not least, if a user has used Twitter for any commercial transaction, the user's payment information will also be collected (credit/debit card number, expiration date, CVV code, billing address, and shipping address); however, Twitter does not make this public.

Clearly, there is a tremendous amount of data that can be collected from even a single tweet, well beyond the content the tweet itself. Indeed, according to Elasticsearch, a software start-up focused on social media data mining, there are actually as many as 150 points of metadata available from a single tweet (Dwoskin, 2014b).

5.2 Sources of Twitter data

In terms of accessing this data, there are a limited number of sources that can provide comprehensive search functions and full data results. There are other entities that collect and store historical Twitter data, notably the US Library of Congress, but they either do not have access to the full Twitter firehose, or do not have the technological infrastructure to support searches. Similarly, Twitter has other data partners that operate

in specific areas or countries, but also do not have access to the full stream of historical and real-time tweets.

As previously mentioned, hashtags function as a content proxy that categorizes tweets, allowing researchers to search for relevant tweets by narrowing results down to those that mention a particular hashtag or hashtags. For example, during the Egyptian uprising in 2011, the hashtag #jan25 was used by many Twitter users to facilitate information sharing through the search tool. Though the search function on Twitter's website allows users to pull up current and/or historical tweets and retweets by hashtag and narrow results according to a specific time window, it is not necessarily comprehensive as search results are pre-filtered by Twitter (Twitter Search Rules and Restrictions, n.d.). However, the results do not include all of the metadata described in the previous section, and are not exportable to an excel file for more rigorous data analysis. To get that detailed level of data – which would be necessary for the possible application of these metrics to country risk analysis – one must look to other providers.

So how do entities get access to Twitter's full stream of tweets, which comprises about 500 million per day from its base of 140m-170m active users (Koh, 2014; Honan, 2012)? Many applications and developers collect tweets via Twitter's open API (application programming interface), which allows them to create applications, websites, widgets, products and platforms that can interact with Twitter data (Developers FAQ, n.d.; Bright Planet, 2013). Twitter is particularly generous with its data relative to other social media platforms, such as Facebook, which are far more zealous and protective. That said, Twitter's willingness to share segments of its data is not entirely altruistic – the firm of course has a financial interest in widening the applications of its product. This is

particularly evident in light of the series of acquisitions that Twitter has carried out to date – over 30 firms in total, more than half of which have occurred since 2012 (CB Insights, 2013).

5.2a The Planet of the APIs

Within its APIs, Twitter distinguishes between “pull” and “push” data functions. Through the Search API – or “pull” – a developer can aggregate historical Twitter data based on criteria such as keywords (words, phrases or hashtags), geographical boundary boxes, and user IDs (Morstatter et al., 2013). The Streaming API, on the other hand, allows tweets to be “pushed” directly in real time, also by the previously mentioned criteria. In both cases, developers must obtain authorization from Twitter to make authorized calls to its APIs (Developers Overview, n.d.).

In terms of looking at social media data in real time during a political risk event, the Streaming API would be the obvious choice for risk analysts. However, this necessarily raises questions about the comprehensiveness or sampling validity of Streaming API results. According to a recent study, Twitter’s Streaming API retrieves “at most a 1% sample of all the data” (Morstatter et al., 2013). Furthermore, these APIs are increasingly rate limited, with the number of data requests limited per user or per application per 15-minute window (Rate Limits: Chart, n.d.).

5.2b Full stream ahead

The Twitter Firehose, on the other hand, is a feed that provides 100% of the social media platform’s tweets in real time, with a combined total of over 425 billion points of

data to date (Vincent, 2013). However, few entities have access to this stream of data. While Twitter is less transparent in this regard, limited firehose access is thought to be a combination of control on the part of Twitter and high barriers to entry for data firms. Indeed, the bandwidth and server infrastructure required to handle the full stream of tweets, let alone the cost of acquiring access from Twitter, is incredibly expensive. In 2013, Twitter received USD70 million in data licensing payments, up 48% year-on-year, with *Business Insider* estimating that the firm's data revenue could hit USD100 million by the end of 2014 (Edwards, 2014).

While other firms receive segments of the Twitter firehose, such as NTT Data, which has access to the full Japanese firehose, as of mid-2014 only four firms had full access: Gnip, Topsy, Dataminr, and DataSift. These firms then resell the data to other corporations or analytics firms that are hired by other parties. Indeed, according to Gnip's website, the company has customers in more than 40 companies, which in turn cover over 95% of Fortune 500 companies (Gnip, n.d.). For example, one of Gnip's biggest clients is Klout, which analyzes social media influence for users and firms through its well-known "Klout score". According to a blog post by Gnip CEO Chris Moody, Gnip had delivered more than 2.3 trillion tweets to its customers as of April 2014 (Moody, 2014).

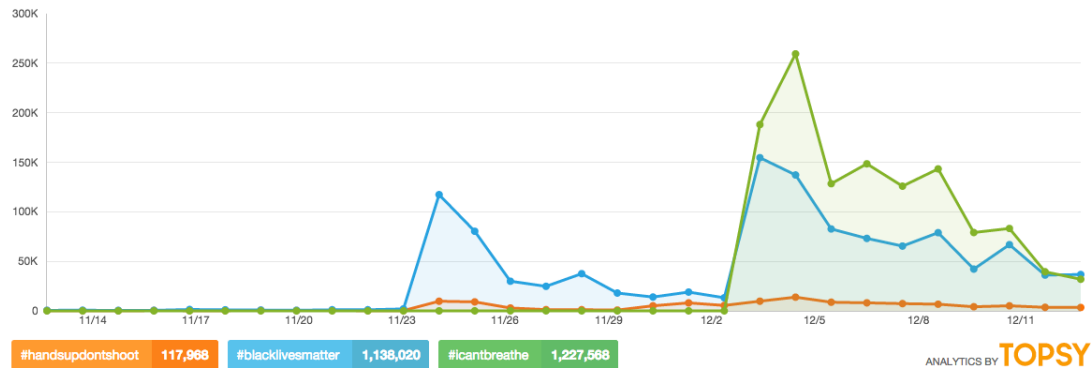
Gnip, a US-based firm founded in 2008 and Twitter's first data partner, was the first to be granted access to Twitter's complete data archive. A social media API aggregation company, Gnip has been called the "Grand Central Station of social media", capturing real-time data from social media sites and then reselling that data to third parties. After partnering with Twitter in 2010, Gnip was bought by the social media giant in April 2014 for an estimated USD134.1 million. Twitter also bought MoPub, a mobile

advertising network, for USD350 million in 2013, in addition to a couple dozen other firms specializing in advertizing and analytics. Many expect these acquisitions to change some of the products that Twitter will offer, and could also potentially have repercussions on the availability of Twitter data through competing platforms (Kepes, 2014). Gnip is also a key player in the US Library of Congress' efforts to collect an exhaustive archive of historical and current tweets (see Section 5.3).

Topsy, also based in the US, was founded in 2007 by a former Napster engineer, and has been run by a former Cisco executive since 2011 (Wakabayashi and MacMillan, 2013). Having raised more than USD32 million from investors and venture capital firms since its inception, Topsy was acquired by Apple in December 2013 for an estimated USD200 million (Koh, 2014). Like Gnip, Topsy has access to the full Twitter firehose and uses its own algorithms and software to resell and analyze tweets for other companies, particularly brands and advertisers (Vincent, 2013). A feature on its website allows users to input up to three hashtags and Topsy generates an analytic graph of recent usage in the past month (see Figure 9). As the Gezi park protests are long past, this study chose a more timely collective action theme, in the form of outrage in the US over two grand jury refusals to indict white police officers accused of murdering unarmed black men, using some of the key hashtags for this movement: #ICantBreathe, in reference to Eric Garner's on-video exclamation while being held in a chokehold by a New York City police officer; #BlackLivesMatter, a rally cry of protesters; and #HandsUpDontShoot, another trending hashtag.

Figure 10: Topsy infographic⁷

Tweets per day: #handsupdontshoot, #blacklivesmatter, and #icantbreathe
November 13th — December 13th



Topsy was the first firm to offer full search capability for the eight year Twitter archive, in September 2013 (Hern, 2014). However, in the wake of Topsy's purchase by Apple, there are some doubts about continued data provision (see Section 5.4). Prior to the acquisition, Topsy and Twitter partnered, in conjunction with two polling groups, the Mellman Group and NorthStar Opinion Research, to create the Twitter Political Index – or Twindex – tracking the political sentiment of tweets in real-time in the run-up to the 2012 US presidential elections (Twitter Election, n.d.; Honan, 2012; Bertolucci, 2012). This type of social media data analysis, which examines the underlying sentiment of tweets, will be discussed in greater detail in Chapter 6.

DataSift, an independent enterprise social data platform, was founded in the UK in 2007. One of its earliest services was TweetMeme – an API that aggregated popular subjects on Twitter and created the first-ever embeddable retweet button, which was later adopted by Twitter; however, the service ceased operations in October 2012 (McCarty,

⁷<http://topsy.com/analytics?q1=%23handsupdontshoot&q2=%23blacklivesmatter&q3=%23icantbreathe&via=Topsy>

2012). Thanks to its TweetMeme experience, DataSift has expanded operations in recent years to provide more sophisticated analytical tools to over 1000 customers in 40 countries, including heavy-hitters such as Bloomberg, Dow Jones, Dell and CBS Interactive (DataSift Customers, n.d.; DataSift News, 2013). Since it began fundraising in early 2010, DataSift has raised over USD70 million in venture capital funding, including USD42 million in Series C financing in December 2013 (DataSift News, 2013).

In addition to its access to the full Twitter firehose, DataSift announced the inclusion of Sina Weibo, the Chinese micro-blogging and social networking site, in December 2013, helping to bridge the social media data gap in China, where Facebook and Twitter are not widely available (Wagner, 2014; DataSift Sina Weiba, n.d.). The company also recently launched an innovative machine-learning-based product for tweet sentiment analysis, which will be discussed in greater detail in Chapter 6 (Wagner, 2014).

Last but not least, Dataminr likewise benefits from access to the Twitter firehose. Founded in the US in 2009, the firm focused on early detection of important events and news not yet covered by mainstream media, with a focus on market-moving activity (The Economist, 2010). Indeed, the company is credited with detected news of the killing of Osama bin Laden a full 23 minutes before conventional media. On average, the firm estimates that it has a five-to-ten-minute edge over news outlets. With USD50 million in venture capital raised so far, in 2014 Dataminr launched a new service that provides alerts to leading news services (see Section 5.3). The company also counts the city of Boston as one of its estimated ten clients, which used its tracking service in an effort to avoid another terrorist attack during the 2014 marathon (The Economist, 2014).

However, in contrast to DataSift and Gnip, Dataminr is not a wholesale data reseller, and has a much greater focus on real-time news alerts than its competitors, sending out about 100 alerts per day. The firm has a decent track record of predicting market-moving events, such as when it noticed a pattern of negative words in tweets about Apple prior to a downturn in its share price. Dataminr was even able to correctly warn its financial clients about a fraudulent Associated Press tweet about an attack on the White House in April 2013 a full five minutes in advance of an AP correction – the uncertainty caused by the fraudulent tweet had led to a 145-point drop in the Dow Jones Industrial Average in mere seconds (see Chapter 6) (Gayomali, 2013; Popper, 2014).

5.3 Other players

Twitter has also entered into data partnerships with other companies and research entities. In late-October 2014 Twitter and IBM announced a partnership whereby IBM will use Twitter data to “help businesses predict trends in the marketplace and consumer sentiment about products and brands...train[ing] 10,000 employees to consult businesses on the best use of Twitter data” (Lopes, 2014). Furthermore, IBM will offer Twitter data as part of its analytics services, in addition to enabling software developers to use Twitter data in applications built using IBM’s Bluemix and Watson Developer Cloud.

In January 2014 CNN, the US-based news agency, also announced a partnership with Twitter – in tandem with Dataminr (Ha, 2014). The new tool, called Dataminr For News, will provide CNN with news alerts as well as data analysis about who broke the news and the pattern in which it is spreading on social media. In the intervening months, this service has been expanded to the BBC, The Weather Channel, and USA Today,

among others, before commercially launching the product in wider North America in September 2014 (Ha, 2014; Dataminr Announces Launch, 2014).

As previously mentioned, Twitter has an ongoing relationship with the US Library of Congress (LoC). The LoC signed an agreement with Twitter in April 2010 to receive a historical archive of tweets dating back to 2006 – or 21 billion tweets – as well as the promise to continue providing tweets on an ongoing basis (Library of Congress, 2013). The LoC had accumulated two copies of around 170 billion tweets as of January 2013, for a total of 133.2 terabytes of data. Twitter selected Gnip to be the delivery agent for the transfer of tweets in a move that, despite being several years ahead of Twitter's acquisition of the company, underscored the close partnership of the two firms from an early stage. From a logistical standpoint, the LoC reports that tweets are transferred from Gnip in hour-long segments, which are uploaded to a secure server for retrieval by the library. Starting from an initial volume of 140 million tweets per day, the daily data transfer has expanded to around 500 million tweets at end-2012, according to a LoC press release about the program.

The agreement with the LoC laid out stipulations surrounding the sharing of tweets with researchers and the public. While the library was granted permission to make portions of the data available to “bona fide researchers”, it could only do at a delay of a minimum of six months after the tweet was published – hampering the ability of researchers to evaluate the tweets in real time. Furthermore, the LoC was forbidden from making a “substantial portion of the collection [available] on its web site in a form that can be easily downloaded,” underscoring Twitter concerns about the free flow of data (Library of Congress, 2013). Perhaps most importantly for any potential application in

the field of country risk analysis – researchers working with the LoC collection were prohibited from the commercial use or redistribution of any tweet collection.

The LoC's search capabilities, however, remain limited. In addition to the data sharing limitations imposed by the agreement with Twitter, a single keyword search of the LoC's 2006-2010 fixed archive of tweets currently takes around 24 hours. This obviously limits the ability of the library to offer any kind of useful access to researchers. The LoC's January 2013 press release highlighted the difficulty of addressing this going forward: "to achieve a significant reduction of search time, however, would require an extensive infrastructure of hundreds if not thousands of servers. This is cost-prohibitive and impractical for a public institution.... It is clear that technology to allow for scholarship access to large data sets is not nearly as advanced as the technology for creating and distributing that data" (Library of Congress, 2013). While the library highlighted the possibility of developing research- and scholarship-based access to historical tweets in cooperation with Gnip, for the time being, the capital-intensive nature of maintaining a tweet archive that is searchable seems to necessitate a paid access structure.

5.4 Cost of Twitter data

In terms of establishing the cost of accessing this data, price inquiries were made to each of the four Twitter firehose data resellers mentioned in the previous section, with varying results. The inquiries sought to determine the cost of the data set described in the NYU SMaPP report on the Gezi park protests, as a benchmark. According to the information available from the report, this data set was collected from May 31 to June 11,

2013, with at least four hashtags (#direngeziparkı, #occupygezi, #geziparkı, and #BugünTelevizyonlarıKapat), and totaled over 22 million tweets (SMaPP, 2013).

Secondly, the quotes would paint a picture of the cost of real-time access to Twitter data, to get a better idea of the cost structure for companies, institutions, or analysts looking to apply social media metric to their country risk analysis practices. While the SMaPP report provides a useful historical example of social media metrics during a political risk event, the type of future application discussed in this study would necessarily be in real-time, and therefore said pricing is an important variable to know.

Dataminr was unable to provide a quote for a dataset equivalent to the SMaPP study, as they do not do single sales. However, to “provide contextual analytics and signals based on the entire Twitter data stream”, Dataminr clients pay an annual subscription (Pradeep Menon, personal communication, December 2014). Depending on the particular package selected, this could cost anywhere from USD50,000 to over USD1 million. For DataSift, pricing is also dependent on volumes. For a platform license and training for six months, the company charges USD9,000 – plus an additional USD100 for every million tweets accessed (Paul Intrieri, personal communication, December 2014). At this rate, fees could add up very quickly. Indeed, at these rates, the SMaPP data set would cost close to USD12,000, while access to all 500 million tweets each day would cost the base USD9,000 plus USD50,000 per diem. Gnip did not return a quote after learning it was for an academic inquiry, though informal discussions with their sales team suggest a price well into the thousands for the kind of data loads that would be required – comparable to its peers. Topsy also never returned my request for a bid (lending some credence to unconfirmed reports that the firm is no longer delivering Twitter data –

indeed the Gnip website seems to poke fun at this with the sales pitch “Not getting data from Topsy anymore?”).

5.5 Relevant financial data

As previously discussed, the infrequent updating of macroeconomic country risk indicators such as GDP or inflation makes it difficult to link political risk events like protests with their carry-through effects on the economic and financial soundness of the country. To that end, any future study of the prudence of applying real-time social media metrics to political risk analysis in a quantifiable manner requires using economic and financial data that responds more immediately to events on the ground.

Stock markets and foreign exchange markets – thanks to daily trading – could paint a better picture the transmission of political risk to the financial sector on a day-to-day basis during a disruption similar to the Gezi park protests. To demonstrate this, information from Turkey’s stock and foreign exchange markets during the period of the SMaPP report (May 31 to June 11, 2013) was accumulated as a means of hypothesizing about any potential relationship between the political risk generated by the protests and the performance of Turkey’s capital markets – which can be seen as a measure of investor confidence and expectations.

The Borsa Istanbul (BIST), which includes both equities and debt trading, was founded in December 2012, bringing the previous Istanbul Stock Exchange and the Istanbul Gold Exchange together under one roof (Oxford Business Group, 2013, pp. 58). The BIST later merged with the Turkish Derivatives Exchange (TurkDex) trading platform in August 2013 (Borsa Istanbul, n.d.). The BIST’s broadest index – BIST 100 –

is a good measure of the market's performance, and is tracked by a wide variety of financial market news outlets.

The Wall Street Journal's historical price tool provides information on the BIST 100 Index's performance for the days in question, such as highs and lows, as well as opening and closing levels (Wall Street Journal Historical Prices, n.d.). In terms of foreign exchange movements, Oanda provides comprehensive historical foreign exchange data, which can be specified for a specific window of time (Oanda Historical Exchange Rates, n.d.). Movements in interest rates and default insurance premiums, which are revised on a more frequent basis, could also demonstrate the financial impact of political risk generated by the protests. Other potential data of interest could be the behavior of domestic actors – for example, bank account withdrawals or changes in consumer behavior, such as hotel reservations.

6. Interpretation & Analysis

This chapter briefly focuses on some of the limitations of current political risk measures, underscoring the possible usefulness of alternate flows of information, such as social media. Now that the types and sources of Twitter data have been laid out, it is time to look at some of the possible ways in which this data is being analyzed by researchers. To understand this within the context of the Gezi park protests, we look to the work of researchers at NYU's Social Media and Political Participation (SMaPP) lab, which conducted an in-depth study of social media metrics collected from the Gezi park protests. Their work demonstrates some of the ways in which Twitter metadata can be analyzed to extract information about the progression of social movements, which this study contextualizes in terms of useful application for political risk analysis.

Next, we look at some of the possible market data that could be examined in tandem with tweets to better understand the financial impact of social movements in emerging markets. Using the same timeframe as the SMaPP study, we look at some of the movements in Turkey's stock market, foreign exchange rate, interest rates, default premium insurance, bank deposits, and tourism industry. The study then discusses advances in and applications of sentiment analysis, before turning its attention to the challenges posed by the growing online trends of "astroturfing" and "crowdsourcing". Lastly, the chapter concludes with a brief discussion of the ethical perils of advancing the understanding of political risk events through social media data analysis.

6.1 Limitations of traditional political risk indicators

As discussed in Chapter 2, the industry best practices suggest a highly qualitative approach to political risk analysis, which focuses largely on the conditions of a country generally. However, these structural characteristics (corruption, military involvement, religion in politics, turnover in leadership, etc.) are unlikely to change alongside a protest movement. While they may indicate the general likelihood of unrest breaking out in a given country, they provide much less predictive power once said unrest is starting.

In this sense, much like the infrequently updated macroeconomic indicators that fail to demonstrate the negative financial impact of an uprising, so do could traditional political risk variables fail to account for growing masses of protesters in a square. Social media metrics, then, could provide a real-time indication of the development of a protest movements – how it is growing, how information is spreading, whether it is actually a grassroots phenomenon or something being promulgating from outside the country, etc.

So much of business and sociology can now be expressed and studied numerically. While the qualitative added value of regional experts should not be underestimated, social media metrics have the potential to add quantitative weight to forecasts and provide companies with actionable information as events progress.

6.2 Gezi protests in tweets

Given the difficulty and cost of accessing historical, comprehensive Twitter data on an independent research level (see Chapter 5 for greater detail), this study focuses on the findings of researchers with access to full Twitter data sets to illustrate the both the types and possible applications of information collected by Twitter. Perhaps the most

fascinating and rigorous research to date on the protests has been conducted by the New York University Social Media and Political Participation (SMaPP) lab.

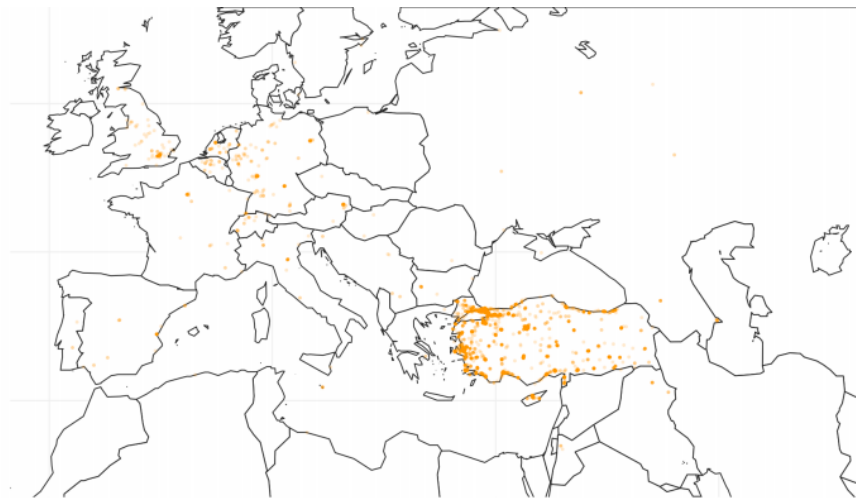
Using a data set of over 22 million tweets from the early days of the 2013 Gezi park protests in Istanbul, Turkey, the SMaPP team analyzed tweet volumes over time, the geolocation of tweets, hashtag mention volumes, the language of the tweets, and the network position of the users in an effort to understand the role played by social media in “initiating, promoting, and spreading political participation” (SMaPP, 2013). The SMaPP researchers narrowed their search field by focusing on tweets using the most common hashtags associated with the Gezi protests.

The results of their analysis offered some key insight into the nature of the Gezi protests relative to other recent movements similarly associated with social media. In contrast with Starbird and Palen’s (2012) finding that only 30% of the most highly retweeted Twitter users were physically in Cairo during the 2011 uprising, the SMaPP lab noted that 90% of the tweets surveyed came from inside Turkey and 50% from Istanbul itself. The geolocation of tweets is an important variable in understanding an uprising. Indeed, in the aftermath of the 2009 protests in Iran, much of the Western media optimism surrounding Twitter use turned out to be misguided. Many if not most of the tweets were coming from Iranian bloggers in the diaspora, who were understandably prone to be more staunchly anti-regime than those within modern day Iran.

Here too, the language the tweet is written in can be instructive. These same Iranian bloggers in exile were largely writing in English, which points to the fact that they were more likely trying to reach international media (i.e., the *Atlantic*’s Andrew Sullivan) than coordinate activism on the ground in Tehran – a task for which Farsi

would make far more logistical sense (Morozov, 2009). In Turkey, by contrast, 88% of the tweets in the initial days were written in Turkish – a statistic that speaks to the more domestically targeted nature of protest-related social media (SMaPP, 2013). While there were pockets of tweets coming from outside Turkey, they appeared to be located in areas with large numbers of Turkish citizens or descendants, such as Germany (see Figure 10 below).

Figure 11: *Geolocated tweets*⁸

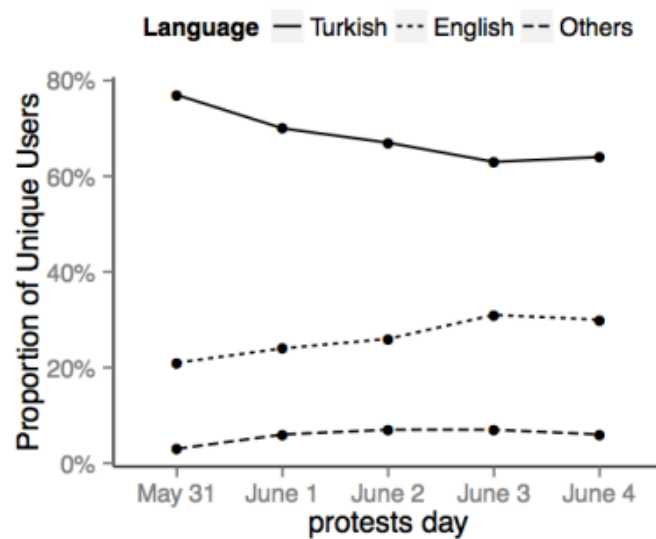


Interestingly, the pattern of the language of tweets would change as the days went on. According to the SMaPP researchers, by the fourth day of protests, 30% of unique users using the most popular hashtags were English speakers, with the network of “hacktivists”, Anonymous, and its related Twitter accounts becoming the chief objects of retweets. This transition, seen in Figure 11 below, demonstrates the internationalization of the Gezi park protests and the shift from local, grassroots focus to a larger anti-

⁸ SMaPP, 2013.

government crusade (SMaPP, 2013). The language of tweets can also impact market confidence on an international scale, as foreign news outlets increasingly covered the Gezi park protests and English-language tweet activity, while Turkish broadcasters remained largely quiet (see Section 6.3 for discussion of financial impact).

Figure 12: *Distribution of tweets by language*⁹

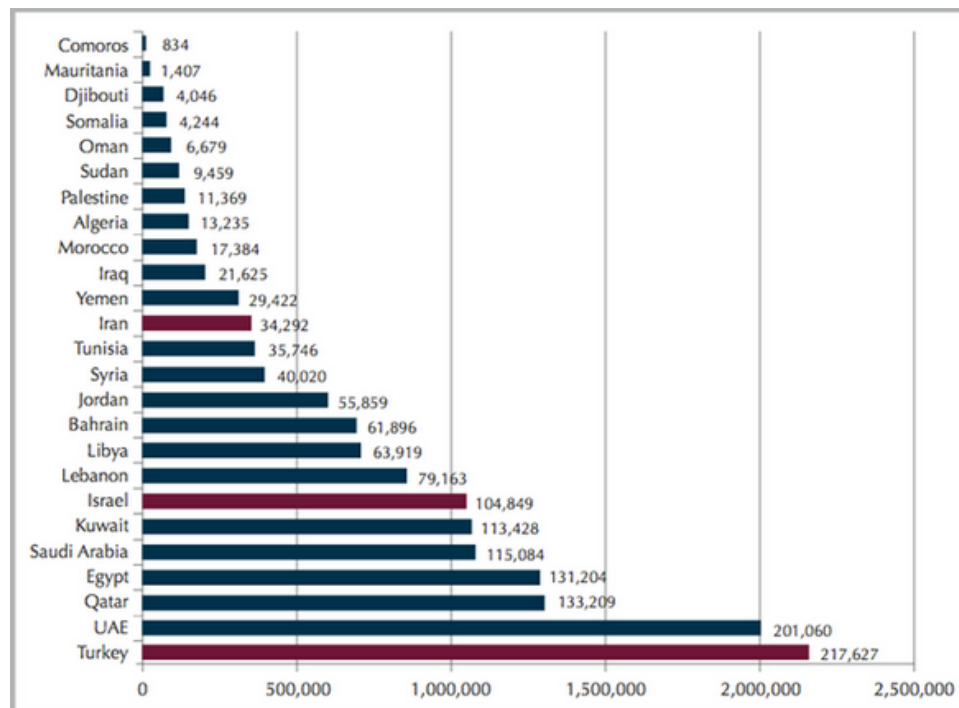


The volume of tweets in a given span of time also can speak to the intensity and size of a protest. The SMaPP researchers found that in the first 18 hours of the protests, there were over 2 million tweets referencing the top-three hashtags: #direngeziparkı, #geziparkı, and #occupygezi. By June 3, 2013, the hashtag #direngeziparkı alone had received 1.8 million tweets (SMaPP, 2013) – well above the fewer than 700,000 tweets with the hashtag #jan25 seen during the entire 18 days of the 2011 Egyptian uprising (Wilson and Dunn, 2011). The relative lack of social media volume in Egypt must be viewed in the context of lower social media penetration and the internet shut down by the

⁹ SMaPP, 2013.

government. Indeed, as of January-March 2011, Egypt had 131,204 active Twitter users to Turkey's 217,627 – with active users seen as a better indicator of social media user than the total number of accounts (see Figure 12) (Interactive Middle East, 2011). By 2013, when the Gezi park protests occurred, Turkey's active Twitter population had grown to an estimated 4.3 million, according to social media monitoring group Monitera (Kesen, 2014). Understanding the context of data received from different countries is therefore crucial to understanding the on-the-ground dynamics of a political risk event.

Figure 13: Active MENA Twitter users¹⁰



SMaPP researchers also point to the lack of traditional media coverage in Turkey as a catalyst for prolific social media use. Twitter and Facebook effectively became

¹⁰ <http://interactiveme.com/2011/06/twitter-usage-in-the-mena-middle-east/>

replacements for mainstream media that failed to cover the protests – indeed, as SMaPP researchers point out, CNN-Turk was actually airing a special on penguins as tear gas canisters were being launched in Gezi on June 2, 2013 (SMaPP). A photo highlighting the absurd juxtaposition of CNN-Turk and CNN International coverage quickly surfaced on an occupy Gezi Tumblr account (see Figure 13 below).

Figure 14: *CNN-Turk vs. CNN International coverage (June 2, 2013)*¹¹

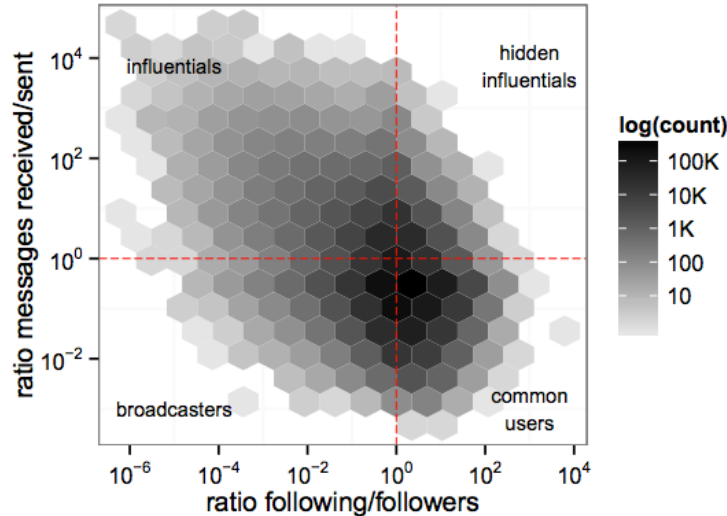


SMaPP researchers also examined the network structure of the early days of tweets about Gezi, and found an interesting pattern of information that seems to confirm Gonzalez-Bailon and Wang's (2013) demonstration of highly uneven centrality in the Occupy/indignados global network. In the instance of the Gezi park protests, 1% of users concentrated roughly 80% of all retweets received, and three-quarters of users were not retweeted at all, revealing a significant asymmetry of sources of information and authority (SMaPP, 2013). Their network mapping also revealed that while these

¹¹ <http://www.dailydot.com/news/cnn-turk-istanbul-riots-penguin-doc-social-media/> from <http://occupygezi.tumblr.com/>

“authorities” had large audiences (i.e., a large follower base), they were not particularly central to the network. In fact, 37% of authorities had average-sized networks, even following more users than followed back. The SMaPP researchers have dubbed these actors as “hidden influentials” due to their low-visibility in the network but high visibility in the information stream (SMaPP, 2013). In Figure 14 below, SMaPP replicates a graph from Gonzalez-Bailon et al, 2012 to demonstrate the distribution of users in the Gezi park protests.

Figure 15: User distribution¹²



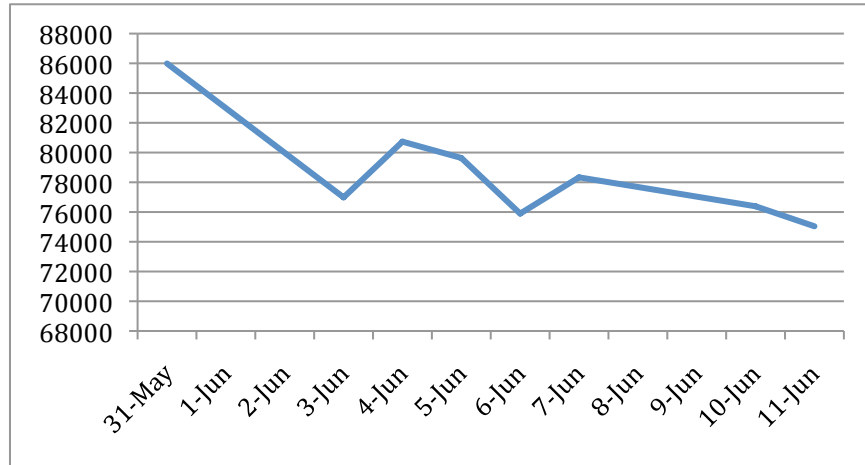
6.3 Financial impact

The results of the SMaPP study are useful when examined in relation to the stock market and foreign exchange market movements over the same period. From May 31 to June 11, 2013 – the days covered by the SMaPP study – the Istanbul stock market’s BIST 100 Index saw a nearly 13% decline, equivalent to a loss of billions of dollars in market

¹² SMaPP, 2013.

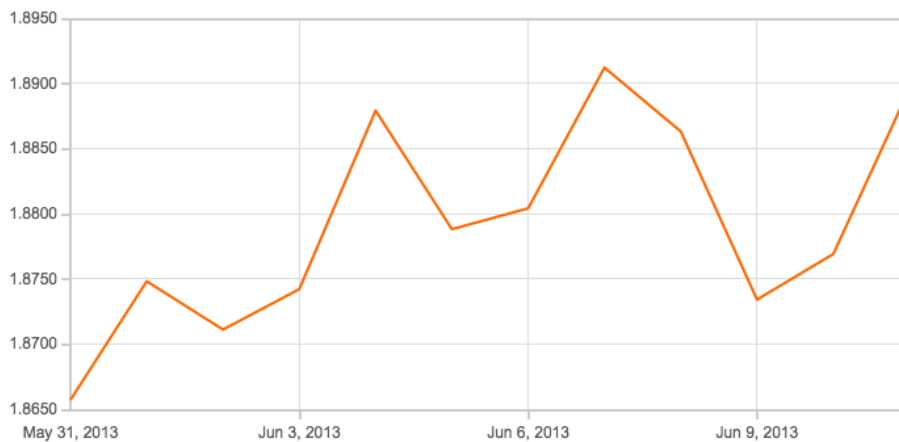
capitalization, which points to the protests' effect on market confidence (see Figure 15 below).

Figure 16: *BIST 100 Index closing values*



Over the same period the USD:TL foreign exchange rate also deteriorated, albeit less dramatically – moving from 1.8657 to 1.8896, a 1.3% decline (see Figure 16 below).

Figure 17: *USD:TL foreign exchange rate*



Interest rate movements also demonstrate the financial impact of political risk from the Gezi park protests. In May 2013, before the protests began, Turkey's interest rate sat at 4.5%. Within one week this had jumped to around 7%, equivalent to a USD12-billion increase in three-month borrowing costs, according to Ünsal Ban, rector at Türk Hava Kurumu University (Delta Stock, n.d.; Today's Zaman, 2013). In terms of the impact on the private sector, Nevzat Saygılıoğlu, former head of the Ministry of Finance's General Directorate of Revenues, estimated the cost of the stock market decline, higher interest rates, and a lower exchange rate on the sector's USD250 billion in foreign-currency-denominated debt to be in the neighborhood of USD20 billion. Default insurance premiums, another measure of the perception of risk, increased by more than 60 basis points to 183 as a result of the protests (Gölay and Xypolia, 2013).

Another interest metric for the financial impact of the protests is the behavior of domestic actors. While no official figures were released, Al Monitor reported information from the banking circles that roughly TL50 million was withdrawn from banks on June 3rd, the Monday after the protests began (Ozay and Celik, 2013). As also seen in the Arab spring, tourist revenues can be drastically affected by the perception of political instability. In Turkey, the Touristic Hotels and Investors Association (TUROB) reported nearly 216,000 cancelled reservations in Istanbul between May 31 and June 17, 2013, adding to a financial loss of EUR54.7 million for the hotel industry alone (Hurriyet, 2013). Similarly, future bookings were down 55%. While these figures admittedly pale in comparison to the financial costs of the Tahrir protests, for example, they are nonetheless considerable sums of money and had a serious impact on businesses operating in the country, even if on a smaller scale.

6.4 Additional methods of tweet analysis

Another method by which tweets can be analyzed is sentiment analysis. Part of the field of Natural Language Processing (NLP), sentiment analysis seeks to classify the sentiments expressed in written content – in this case, tweets (Kouloumpis et al., 2011). However, unlike sentiment analysis for product reviews or other longer, more formal online writing, “the character limitations on tweets...and informal and specialized language used in tweets, as well as the very nature of the microblogging domain make Twitter sentiment analysis a very different task” (Kouloumpis et al., 2011).

Indeed, the very nature of using hashtags as a content proxy raises questions about the limitations of deriving the meaning users intent from tweets. Furthermore, abbreviations, misspelling, and slang can throw wrenches into the process of tokenization and normalization of the text of tweets. Indeed, as Go et al. (2009) acknowledge, as “Twitter users post messages from many different mediums, including their cell phones...the frequency of misspellings and slang in tweets is much higher than other domains”.

In terms of data availability, not all of the Twitter data resellers support all languages: for example, while DataSift support over 140 languages, as of February 2014 Gnip only covers 24 (Wagner, 2014). Also, as was seen in the Tahrir and Gezi park protests, domestic actors were increasingly tweeting in English to garner international media coverage. Here, too, there is greater room for spelling errors or non-native phrasing of ideas.

Arabic presents a particular challenge, as Arabizi – or “Arabic text that is written using Latin characters...used to present both Modern Standard Arabic (MSA) or Arabic

dialects” – is quite popular on social networking sites, particularly given the Latin alphabet emphasis of many leading smartphones, computers, and applications (Darwish, 2014). As Arabizi is a transliteration of Arabic sounds to the Latin alphabet, some sounds do not have direct equivalents and must be spelled using a combination of letters or even different letters depending on the language being transliterated into: French bilingual speakers often use “ch” for the Arabic letter “ش”, while English bilingual speakers typically use “sh”, for example. The use of numerals to mimic letters that do not exist in the Latin alphabet is also commonplace. To replicate the letter “ع”, Arabizi frequently makes use of the number “3”, as it somewhat resembles the Arabic letter. The same can be said for “ح”, which often becomes a “7” due to the resemblance. However, some users may express the “ع” as “aa” or “a”, while others would transliterate “ح” as a regular “h”.

The lack of standardization can understandably lead to difficulties for proper data analysis. In terms of automating conversion from Arabizi to Arabic for language processing and sentiment analysis, a recent study demonstrated some difficulties with the process (Darwish, 2014). While identification accuracy of Arabizi was very high, at 98.5%, conversion accuracy was demonstrably lower, at 88.7%.

Furthermore, simply classifying tweets broadly as positive or negative also leads to oversimplification of communication, and perhaps fails to understand the intention of messages. For example, in the case of protests, determining whether messages are simply expressions of solidarity or efforts to organize, mobilize, and disseminate information is crucial to understanding the scope of collective action. Indeed, as Kouloumpis et al. (2011) found in their study, some of the conventional methods of sentiment analysis,

such as part-of-speech features, “may not be useful for sentiment analysis in the microblogging domain.

Network analysis is another important method for deriving meaning from the onslaught of tweets. As seen in the SMaPP analysis of the Gezi park data set, understanding which actors are most influential in a network lends context to the pattern of tweets surrounding a particular subject. As explained by IHS, a US-based analytics and software company, a retrospective look at Twitter shows the signs were clear for the ramp-up of the Gezi park protests (Hansen, 2013). Indeed, an IHS blog post explains that there were 10,000 tweets containing the word “eylem” (Turkish for “protest”) in the 24 hours leading up to the demonstrations. Furthermore, network analysis “revealed that the number of tweets from people with influence within the protest movement was increasing...these signals added up to a clear indication that the unrest would deepen”. IHS even goes so far as to suggest that Twitter patterns foretold of the uprising a full 48 hours before Western media coverage ensued. In a statement that echoes the very premise of this study, “such advance notice could...enable a company to divert shipments or arrange for alternative sourcing before an outbreak of civil unrest disrupts a key link in its supply chain” (Hansen, 2013).

However, as the saying goes, hindsight is 20/20. The SMaPP lab and IHS’s analysis of the Gezi park protests, while masterful, can be seen as fairly useless on a practical level, as companies were in fact not able to divert shipments or arrange for alternate sourcing before the demonstrations flared up. For social media sentiment or network analysis to be truly useful for political risk analysis, it has to give analysts an edge, a head start on events.

This is certainly not to say that this will always be the case – indeed, our ability to analyze big data is constantly improving. DataSift, for example, now employs machine learning to analyze tweets, which it says allows it to “identify and extract subtle concepts such as ‘emotion’, ‘intention’, and ‘topic discussed’” – a serious improvement over the traditional “positive/neutral/negative” classifications of traditional sentiment analysis (DataSift Machine Learning, n.d.).

In the 2012 US presidential election, Topsy and Twitter partnered with a couple of polling firms to create a political index, dubbed Twindex, to analyze the sentiments of tweets surrounding the two candidates – Barack Obama and Mitt Romney (Bilton, 2012). While the Twindex centers on the binary positive/negative results, it demonstrates one very interesting analytical advance: the ability to interpret colloquialisms. Colloquial expressions can be understandably difficult for computer programs to correctly interpret. For example, in the Massachusetts, where Romney previously held office as governor, “wicked” is used to mean “very” or “a lot”, not necessarily “evil” as it could mean in Minnesota.

Sentiment analysis is also slated to play a greater role in the financial markets going forward. Bollen et al. (2011) used large-scale Twitter feeds to determine “collective mood states”, working to correlate these with movements in the value of the Dow Jones Industrial Average (DJIA) over time. The authors were able to correctly predict daily movements of the index with 87.6% accuracy.

Stock market indexes seem to be buying into the premise of sentiment analysis as well. In 2013 the New York Stock Exchange Technologies partnered with Social Market Analytics to provide customers with social media sentiment analysis tools (NYSE, 2013).

Likewise, the DJIA has been working with Columbia University and the University of Notre Dame to put together a dictionary of 3,700 key words to signal changes in market sentiment. The DJIA is using this to provide services to traders and analysts to allow sentiment analysis in print media, though it has yet to be applied on a larger scale to social media (Horwitt, 2011).

While the Twitter data evaluated by SMaPP and others would likely add the most value to the structured-qualitative and quantitative risk assessment models described in Chapter 2 (see Figure 2), given their semi-structured and structured approaches, respectively, social media metrics also help to paint a more general picture of the evolution of social movements that would be useful background for purely qualitative analysis.

6.5 Smarter spam

Another growing trend on the internet has the potential to undermine the veracity of Twitter data. As discussed in Chapter 2, computer programs known as bots are created to mimic human user behavior and generate content, often with the intention of influence human user views on anything from products to politics. As Dr. Lutz Finger highlighted in his speech at the 2013 Social Media Measurement and Monitoring Conference, the scope of these spam bots is massive, and not well understood by everyday users of social media sites.

Finger's (2013) statistics are important, particularly on the Twitter level, given the potentially coercive influence of these programs, which are largely not discussed in academic literature on social media data analysis to date. Part of the difficulty of

discussing bot behavior, according to Finger (2013), is that the only bots that can be proven, are those that fail in their goals. Indeed, if bots are completing these tasks correctly, they will not be detected as fraudulent users.

Finger (2013), however, did notice a pattern of behavior on Twitter that *appeared* to be the behavior of bots. In the early days of the Syrian uprising, the hashtag #syria was often used as a content proxy for sharing information amongst anti-government actors. However, as this hashtag grew in popularity, Finger (2013) noticed a proliferation of seemingly unrelated tweets, pertaining to tourism or other non-sequiturs, using the hashtag #syria. For those unfamiliar with social media streams, a Twitter search for #syria would yield the most recent posts first. As such, a deluge of unrelated content would, in effect, drown out the relevant information and dilute the effectiveness of that hashtag for information sharing and coordination. Whether or not this can be verified, it still points to the potential for social media platforms to be exploited by the state, or others that stand in opposition to forms of collective action.

The concept of machines behaving as humans is hardly new. As early as 1950, British scientist Alan Turing published a paper on artificial intelligence, establishing the concept of a Turing test by which the identity of a user (man or machine) could be determined by a human user. According to Finger (2013), many bots succeed in these efforts, with about one third of social media users failing to recognize that they are speaking with a computer program when chatting online. Indeed, as recently as June 2014, a Russian computer program passed the test, convincing 33% of human questioners that they were talking to another human – echoing Finger’s (2013) statistics (McCoy, 2014).

Cognitive hacking can also come in the form of humans. Indeed, the outsourcing of spam is big business in some circles. While Finger (2013) refers to astroturfing, Lee et al. (2013) define “crowdturfing” as human-powered campaigns. The authors use the example of Vietnamese propaganda officials, who hired 1,000 users to hold fictitious online discussions and post supportive comments about the Communist party, which demonstrates how states can use crowdturfing to their advantage. Similar behavior has taken place in China, with the so-called Internet water army (because they “flood” sites with posts), which are paid to post specific content en masse (Holmes, 2011). Crowdturfing is increasingly organized on websites such as Amazon Mechanical Turk and Crowdfunder, leading to what Lee et al. (2013) term “the commoditization of large-scale turfing campaigns).

The recent example of ZunZuneo – the US government’s attempt to create a Twitter-like network in Cuba – also comes to mind as an example of state manipulation of social media sites (Hill, 2014). Channeling funding through offshore accounts, the US Agency for International Development (USAID) launched a test-message based service in Cuba in 2010, collecting information on users on everything from gender and age, to “receptiveness” and “political tendencies”, “in the hope that the information might be used someday for political purposes” (AP, 2014). During the more than the two years that it operated, ZunZuneo achieved some 40,000 subscribers, mostly young Cubans, and cost an estimated \$1.6 million.

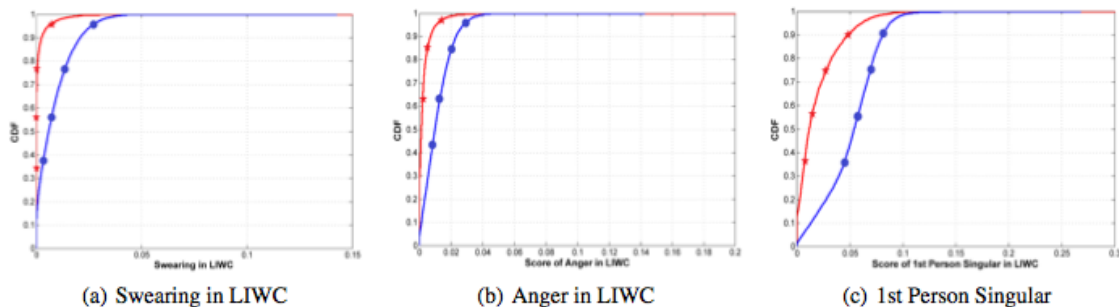
According to an investigation by the Associated Press in 2014, two years after the program had ended, “the plan was to push [users] toward dissent” (Politico, 2014). Documents and interviews showed that they hoped the “network would reach critical

mass so that dissidents could organize ‘smart mobs’ – mass gatherings called a moment’s notice – that could trigger political demonstrations, or ‘renegotiate the balance of power between the state and society’”. The program’s organizers wanted the network to grow slowly, so as to avoid government detection. Nonetheless, the Cuban administration “tried to trace the text messages and break into the ZunZuneo system” (Politico, 2014).

ZunZuneo is an important demonstration of how external manipulation has the potential to create false streams of information that – when social metrics are being applied to political risk analysis – have the potential to create false forecasts or bubbles. Finds tools to differentiate between actual dissent and/or mass movements and fraudulent influence peddling – whether by humans or computer programs – will be crucial for social media metrics to be useful to the field.

So, how then to identify these paid users or computer programs? Researchers have identified certain patterns of behavior from both types of users, ranging from word choice to network position. As shown in Figure 17 below, Lee et al. (2013) found lower rates of swearing, anger, and use of the first-person singular amongst crowdturfers (red line) versus regular users (blue line).

Figure 18: *Crowdturfer behavior mapping*



Similarly, the network positions of crowdturfers or bots are noticeably different than regular users. As Finger (2013) explains, bots tend to link to one another in social media networks, forming an observable pattern that distinguishes them from other users. Lee et al. (2013) also found more a more close-knit network structure among human crowdturfers, with the “average graph density of [normal] users on Twitter to be...many orders of magnitude less dense”.

However, there are many online services, such as Twiends, that increase a user’s follower count, which can mistakenly give the impression of greater influence in the network, or even credibility. An example of the real-world applications of this comes from a contact of the author of this study, who worked for an e-lobbying firm that specialized in using social media to advance the agendas of its clients – self-promotion, or so-called “black PR” about their rivals.¹³ The firm would create profiles for fictional journalists, and then write articles attributed to them on low-profile news websites. To increase the reach of these articles, the firm would use TweetAdder, a social promotion software, to create a large following on social media. This in turn would catch the attention of more prominent news outlets, which would, if not repost their stories, reference aspects of them as background for a story. This reach even extended to well-respected global news outlets. This is a stark example of the ways in which social media data – such as the size of a user’s following – can be manipulated to lend undue credibility to some users, even for commercial purposes.

If the pattern of behavior of bots and crowdturfers is theoretically distinguishable, why are these users continuing to operate? According to Finger (2013), social networking sites are partly to blame for this. Indeed, particularly in the context Twitter and

¹³ The name and company could not be disclosed.

Facebook's initial public offerings, figures on the amount of users are closely related to the financial prospects of these sites. As such, they could be understandably reluctant to crack down as harshly on empty users.

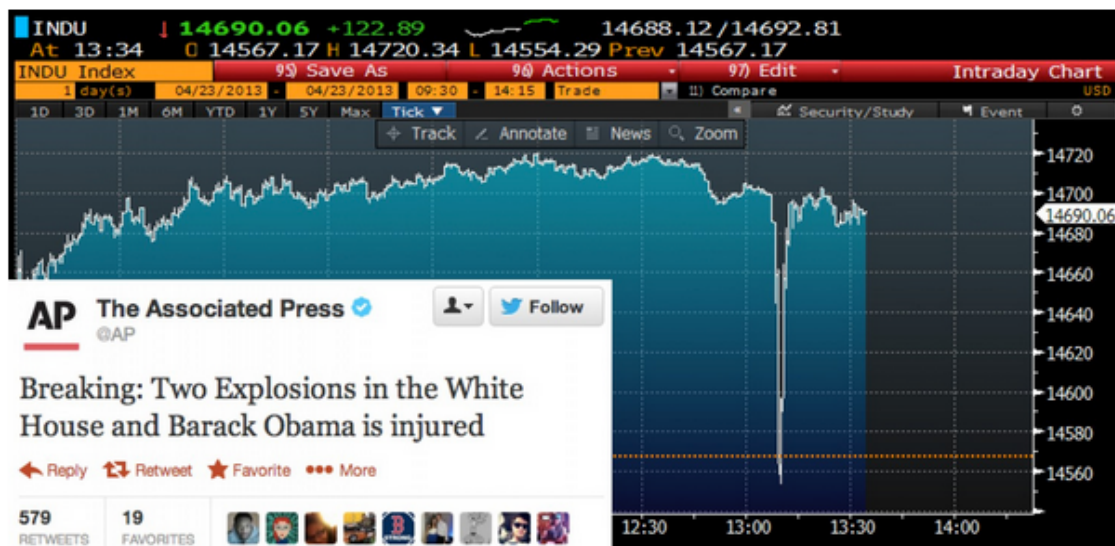
That said, they can be seen taking steps to crack down on social media spam. Twitter has a team dedicated to spam detection, and has filed court cases in the US against five large-scale spammers (Twitter, 2012). However, according to Finger (2013), there is a darker side to more effective detection. When bots are caught and their accounts deactivated, this sends a signal to the programmer that certain behavior is more likely to get the program flagged. Much like antibiotic-resistant bacteria, the next batch of bots programmed will be less likely to demonstrate the behavior that led to its detection. This leads to an arms race of sorts in the computer world, with both detection and programming ramping up over time.

In an effort to avoid this, some websites are taking a more light-handed approach. OK Cupid, a US online dating service, noticed a large number of spam bots on its site. Rather than deleting their accounts and spawning a more sophisticated iteration of the programs, OK Cupid created bots of its own, which it uses to lure the spam bots into a shadow dimension of the site, allowing the spammers to still think they are achieving their programmed task by interacting with other users, and keeping the bots away from actual humans on the site (Finger, 2013).

The stakes are much higher than simply buying the wrong shampoo because a computer program recommended it on Amazon. Cognitive hacking can have serious financial repercussions, as well. Indeed, the hack of the Associated Press Twitter account and posting of a fake terrorist attack on the US White House caused a USD136 billion

drop in three minutes of trading on the DJIA (see Figure 18) (Fisher, 2013). While the market quickly recovered as the fraudulent tweet was revealed (allegedly the work of pro-Assad Syrian hackers, the Syrian Electronic Army), this nonetheless demonstrates the awesome corruptibility of tweets.

Figure 19: AP tweet and DJIA movement¹⁴



This chart shows the Dow Jones Industrial Average during Tuesday afternoon's drop, caused by a fake A.P. tweet, inset at left.

So what is to stop this crowdturfers and bots from influence social media data? The unfortunately answer: at present, not much. According to Finger (2013), there are three necessary conditions for bots to successfully create mass movements: reach, ease of action, and intention. The first two conditions are easily satisfied by the reach and low costs afforded by social media sites; however, the third is not as simple. Research suggests that users are best influenced by information that comes from both on- and

¹⁴ <http://www.washingtonpost.com/blogs/worldviews/wp/2013/04/23/syrian-hackers-claim-ap-hack-that-tipped-stock-market-by-136-billion-is-it-terrorism/>

offline. However, the divide between these two sources of information is shrinking, as 55% of journalists routinely source information from social media (Finger, 2013).

As such, it is imperative to consider the possible influence that these spammers – both human and computers – could have on social media metrics, and determine solutions for removing their corruptive influence from data sets in order to track the actual progression of social movements.

6.6 Ethical questions

The application of social metric data to political risk analysis also raises some ethical questions. In our efforts to harness the predictive power of tweet volumes, language, network influence, etc., it is also possible that repressive states will better understand these movements and target citizens accordingly. Indeed, this is not hard to imagine. During the Euromaidan protests in the Ukraine, for example, the government used the locator function on cell phones to send a text message to those in the vicinity of the protests that read “Dear subscriber, you are registered as a participant in a mass disturbance”, in an effort to intimidate protesters (Lopez, 2014).

Techno-pessimist Evgeny Morozov (2011) decries the naïveté of “cyber-utopians” in failing to foresee the ways in which oppressive regimes would use the internet for their own purposes. As he writes in *The Net Delusion: How Not to Liberate the World*, “cyber-utopians did not predict how useful [the internet] would prove for propaganda purposes, how masterfully dictators would learn to use it for surveillance, and how sophisticated modern systems of internet censorship would become.”

Here Brockett's (1993) findings about the repression paradox come into play. As discussed in Chapter 2, Brockett (1993) found that the outcome of the repression paradox – i.e., whether state repression would end a movement or create an even bigger backlash from the populous – was dependent on the temporal location of the state's response in the protest cycle. If our real-time analysis of social media data advances to the point where we can pinpoint where exactly in a protest cycle we are – albeit in the spirit of more effective political risk analysis – this could also grant tremendous predictive power to repressive regimes looking to suppress dissent.

If the Egyptian government had a better idea of where in the protest cycle the Tahrir demonstrations were in January 2011, might they have waited to shut off the internet until a time when it was less likely to engage the repression paradox? These questions are not easy to ask, and even less easy to answer; however, they are an important element of this initial discussion about using social media metrics to better understand political risk events, and must not be overlooked as this field moves forward.

7. Conclusions

The body of tweets is growing by the day, with more than 20 million messages posted every hour, 330,000 per minute, and 5,600 per second (Hansen, 2013). As our ability to analyze big data catches up to our data storage capabilities, firms and individuals alike will be looking for more ways to harness the predictive power of social media for financial gain. As a field traditionally dominated by the softer-science of qualitative analysis and recommendations by regional experts, political risk analysis is a fascinating frontier for the application of social media metrics. While these traditional methods of analysis remain important, it is also important to be open to advancements in the field. This study has laid out the current state of affairs in the field, in the hopes that further research will be conducted going forward as our real-time analytical sophistication progresses and data prices decline on the back of increased competition. In particular, the structured-qualitative and quantitative models highlighted by Iranzo (2008) stand to benefit from the integration of social media data analytics, given their semi-structured and structured approaches, though purely qualitative assessments could also make use of this data in crafting their broader narratives of a country's risk environment.

Historical analysis of Twitter data, ranging from tweet volumes to network structures, demonstrates patterns that reveal the inner workings of social movements, including the ways in which information is spread – such as through the retweet mechanism, or via information brokers and bridges – how collective framing takes place online, and how actors influence one another. With growing analytical sophistication, it is not a stretch to imagine the day when this can be done in real time and with greater accuracy, thereby alerting companies to changes in political risk conditions on the ground

and giving them time to adjust their risk exposure where possible – be it through currency hedging or supply chain adjustments, among others.

However, this is far from a perfect science. As the preponderance of bots and crowdturfers demonstrates, the anonymous nature of social media sites also lends itself to manipulation. As these programs and users also become more sophisticated, detection and prevention methods will need to keep pace in order to eliminate the potential corruption of data these users can cause – particularly if governments continue their attempts to use social media to counter popular uprisings.

The costs associated with accessing and analyzing Twitter are also substantial. As Twitter limits the access to its full firehose, there is less price competition in the data resale market at present, and firms can charge steep fees for their value-added analytics and algorithms. Indeed, the costs described in Section 5.4 greatly exceed the funding of this study, and could well exceed the budget of many companies evaluating political risk in emerging markets. In this sense, if social media metrics are useful for political risk analysis to such an extent that it can mitigate risk and save companies money – the cost of these data services would need to be viewed as something akin to established instruments for hedging against other types of risk, such as currency or credit risk.

Last but not least, the ethical concerns at stake are also important. Given the importance of timing in the outcomes of state intervention (Brockett, 1993), greater sophistication in predictive social media analytics also has the potential to equip repressive governments with stronger tools for suppressing dissent. This is not to say that companies should not develop social media metrics for their political risk analysis, but rather to highlight that this subject is more nuanced than a balance sheet. Indeed, at the

end of the day, the tweets in question are more than metadata or trends – they also represent the hopes and beliefs of those on the ground, putting themselves in harms way for something they believe in.

An analysis of the field and the available literature demonstrates the pros and cons of applying social media metrics. On the positive side, the capacity to analyze social media data and glean relevant, actionable information about social movements is clearly there. However, this has been conducted largely on a historical basis thus far. With increasing technological sophistication – in terms of both data storage and analysis – it should become possible to do so in real-time and on a wider scale. Furthermore, a future study to examine the correlation between these social media data points and the financial indicators discussed in Sections 5.3 and 6.6 would be helpful.

The biggest cons to the application of social media metrics to political risk analysis today are the cost of data storage; the present limitations of real-time, multi-variable analysis of such large data streams; the growing attempts at manipulation by humans and computer programs alike; and the ethical questions surrounding its use.

Some of these challenges are likely resolvable. The price of data storage should continue to fall, and more sophisticated processing technology will eventually permit the scale of analysis required – as well as the flexibility needed to overcome some of the complications of natural language processing. In terms of the risks posed by bots and cognitive hackers, more intelligent detection and prevention mechanisms – such as the behavioral markers identified by Lee et al. (2013) and the bot solution employed by dating website Ok Cupid – will be integral to halting the arms race of bot programming and ensuring the validity of social media metrics. More proactive involvement of social

networking sites like Facebook and Twitter, which is starting to take place, would also help to curb this trend, as the organizations themselves already have access to the full range of data and can more easily identify and pursue social media spammers.

Ultimately, the obstacles to the applied analysis of social media data are outweighed by its potential benefits to the field of political risk analysis. Social movements will continue to be a fixture of political life – particularly in emerging markets – and in the absence of more frequently updated indicators, social media metrics can provide a useful barometer of political risk events like protests, from everything from their stage in Brockett’s (1993) protest cycle, to their location, scope, and network structure.

REFERENCES

- Associated Press. (2014) Senate committee probes 'Cuban Twitter' USAID ZunZuneo programme. *The Guardian*. Retrieved from <http://www.theguardian.com/world/2014/apr/10/senate-committee-cuban-twitter-usaid-zunzuneo>
- Bertolucci, J. (2012) How Topsy Tames Twitter's Big Data Fire Hose. *Information Week*. Retrieved from <http://www.informationweek.com/healthcare/analytics/how-topsy-tames-twitters-big-data-fire-hose/d/d-id/1105668?>
- Bilton, N. (2012) Twitter Unveils the Twindex, a New Political Index. *The New York Times*. Retrieved from <http://bits.blogs.nytimes.com/2012/08/01/twitter-unveils-the-twindex-a-new-political-index/>
- Bollen, J., Mao, H., Zeng, X.J. (2011) Twitter mood predicts the stock market. Retrieved from <http://arxiv.org/pdf/1010.3003v1.pdf>
- Borsa Istanbul. (n.d.) Milestones in Borsa Istanbul History. Retrieved from <http://borsaistanbul.com/en/corporate/about-borsa-istanbul/milestones-in-borsa-istanbul-history>
- Bouchet, M.H., E. Clark, and B. Gros Lambert (2003) *Country Risk Assessment: A Guide to Global Investment Strategy*. London: Wiley.
- Bright Planet. (2013) API: What's the Difference and Why Should You Care? Retrieved from <http://www.brightplanet.com/2013/06/twitter-firehose-vs-twitter-api-whats-the-difference-and-why-should-you-care/>
- Brockett, C.D. (1993) A Protest-Cycle Resolution of the Repression/Popular-Protest Paradox. *Social Science History*, 17(3), 457-484. Retrieved from <http://www.jstor.org/stable/1171433>
- CB Insights. (2013) Twitter M&A Trends: 31 Acquisitions To-Date with the Pace of Acquisitions Accelerating. Retrieved from <https://www.cbinsights.com/blog/twitter-acquisition-history/>
- Clark, E. and R. Tunaru, (2003), Quantification of political risk with multiple dependent sources, *Journal of Economics and Finance*, 27(2), 125-135. Retrieved from <http://repec.org/mmfc05/paper37.pdf>
- CNN. (2011) Report: Egyptian dad names child 'Facebook'. Retrieved from <http://edition.cnn.com/2011/WORLD/meast/02/21/egypt.child.facebook/>

- Cybenko, G., Giani, A., Thompson, P. (2002) Cognitive Hacking and the Value of Information. *Workshop on Economics and Information Security*, 16-17. Retrieved From <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.306.4277&rep=rep1&type=pdf>
- Dataminr. (2014) Dataminr Announces the North American Launch of Dataminr for News. Retrieved from <http://www.dataminr.com/dataminr-announces-the-north-american-launch-of-dataminr-for-news/>
- DataSift. (n.d.) Customers. Retrieved from <http://datasift.com/customers/>
- DataSift. (n.d.) Machine Learning. Retrieved from <http://datasift.com/vedo/machine-learning/>
- DataSift. (2013) DataSift Raises \$42 Million to Fuel Continued Growth & Platform Innovation. Retrieved from <http://datasift.com/press-releases/series-c-funding-announcement/>
- DataSift. (n.d.) Sina Weibo. Retrieved from <http://datasift.com/platform/datasources/sina-weibo/>
- Delta Stock. (n.d.) Central Bank Interest Rates. Retrieved from http://www.deltastock.com/english/resources/interest_rate_history.asp?curr=TRY
- Darwish, K. (2014) Arabizi Detection and Conversion to Arabic. *Proceedings of the EMNLP 2014 Workshop on Arabic Natural Language Processing*, 217-224. Retrieved from <http://www.aclweb.org/anthology/W14-3629>
- Dwoskin, E. (2014a) Elasticsearch Raises \$70 Million in Big Data Push. *The Wall Street Journal Blogs*. Retrieved from <http://blogs.wsj.com/digits/2014/06/05/elasticsearch-raises-70-million-in-big-data-push/>
- Dwoskin, E. (2014b) In a Single Tweet, as Many Pieces of Metadata as There Are Characters. *The Wall Street Journal Blogs*. Retrieved from <http://blogs.wsj.com/digits/2014/06/06/in-a-single-tweet-as-many-pieces-of-metadata-as-there-are-characters/>
- Eaton, J., Gersovitz, M., Stiglitz, J.E. (1986) The Pure Theory of Country Risk. *The National Bureau of Economic Research*. Retrieved from <http://www.nber.org/papers/w1894>
- Eckerling, D. (2014) How to Measure Social Media ROI. *Social Media Examiner*. Retrieved from <http://www.socialmediaexaminer.com/measure-social-media-roi/>

- Edwards, J. (2014) Twitter is Quietly Building a \$100 Million Business in Big Data. *Business Insider*. Retrieved from <http://www.businessinsider.com/twitter-100-million-big-data-business-2014-4>
- Erb, C.B., Harvey, C.R., Viskanta, T.E. (1996) Political Risk, Economic Risk, and Financial Risk. *Financial Analysts Journal*, 52(6), 29-46. Retrieved from https://faculty.fuqua.duke.edu/~charvey/Research/Published_Papers/P38_Political_risk_economic.pdf
- Financial Times Lexicon. (n.d.) Definition of country risk. Retrieved from <http://lexicon.ft.com/Term?term=country-risk>
- Finger, L. (2013) 7% of Twitter users are not human. *Our Social Times*. Retrieved from <http://oursocialtimes.com/7-of-twitter-users-are-not-human/>
- Fisher, N. (2013) Syrian hackers claim AP hack that tipped stock market by \$136 billion. Is it terrorism? *Washington Post*. Retrieved from <http://www.washingtonpost.com/blogs/worldviews/wp/2013/04/23/syrian-hackers-claim-ap-hack-that-tipped-stock-market-by-136-billion-is-it-terrorism/>
- Fitzpatrick, M. (1983) The Definition and Assessment of Political Risk in International Business: A Review of the Literature. *The Academy of Management Review*, 8(2), 249-254. Retrieved from <http://www.jstor.org/stable/257752>
- Gayomali, C. (2013) The AP Twitter hack: Wall Street's tweet-scraping technology that sparked a 145-point nosedive. *The Week*. Retrieved from <http://theweek.com/article/index/243291/the-ap-twitter-hack-wall-streets-tweet-scraping-technology-that-sparked-a-145-point-nosedive>
- Gladwell, M. (2010) Does Egypt Need Twitter? *The New Yorker*. Retrieved from <http://www.newyorker.com/news/news-desk/does-egypt-need-twitter>
- Gnip. (n.d.) Retrieved from <http://gnip.com/>
- Go, A., Huang, L., Bhayani, R. (2009) Twitter Sentiment Analysis. *Natural Language Processing Stanford*. Retrieved from <http://www-nlp.stanford.edu/courses/cs224n/2009/fp/3.pdf>
- Gonzalez-Bailon, S., Wang, N. (2013) Networked Discontent: The Anatomy of Protest Campaigns in Social Media. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2268165.

- Gölay, B., Xypolia, I. (2013) Reflections on Taksim – Gezi Park Protests in Turkey. *Journal of Global Faultlines*. Retrieved from <http://www.keele.ac.uk/journal-globalfaultlines/publications/geziReflections.pdf>
- Granovetter, M.S. (1973) The Strength of Weak Ties. *American Journal of Sociology*, 78(6), 1360-1380. Retrieved from https://sociology.stanford.edu/sites/default/files/publications/the_strength_of_weak_ties_and_exch_w-gans.pdf+&cd=1&hl=en&ct=clnk
- Ha, A. (2014) CNN and Twitter Partner with Dataminr to Create News Tool for Journalists. *Tech Crunch*. Retrieved from <http://techcrunch.com/2014/01/29/dataminr-for-news/>
- Hansen, C. (2013) Sentiment Intelligence: Converting social media babble into business advantage. *IHS*. Retrieved from <https://www.ihs.com/articles/features/sentiment-intelligence.html>
- Hern, A. (2014) Twitter buys Gnip, one of only four companies with ‘firehose’ access. *The Guardian*. Retrieved from <http://www.theguardian.com/technology/2014/apr/16/twitter-buys-gnip-firehose-analytics-apple-toppsy>
- Hill, K. (2014) From Tinder Bots to ‘Cuban Twitter’, Welcome to ‘Cognitive Hacking’. *Forbes*. Retrieved from <http://www.forbes.com/sites/kashmirhill/2014/04/17/from-tinder-bots-to-covert-social-networks-welcome-to-cognitive-hacking/>
- Holmes, B. (2011) Chinese ‘water army’ hijacks online product review. *New Scientist*. Retrieved from <http://www.newscientist.com/blogs/onepercent/2011/12/chinese-water-army-hijacks-onl.html>
- Honan, M. (2012) Twitter’s New Political Index Proves Big Data Knows What You’re Thinking. *Wired*. Retrieved from http://www.wired.com/2012/08/twindex_twitter_politicalindex/all/
- Horwitt, E. (2011) Sentiment analysis comes of age. *Computer World*. Retrieved from <http://www.computerworld.com/article/2513129/enterprise-applications/sentiment-analysis-comes-of-age.html>
- Hurriyet. (2013) Gezi protests cost \$55 mln euros to hotels. Retrieved from <http://www.hurriyetaidailynews.com/gezi-protests-cost-55-mln-euros-to-hotels.aspx?pageID=238&nID=49170&NewsCatID=344>
- IBM. (2013) 4 Vs of Big Data. Retrieved from http://www.ibmbigdatahub.com/sites/default/files/infographic_file/4-Vs-of-big-data.jpg

- IBM. (n.d.) What is big data. Retrieved from <http://www-01.ibm.com/software/data/bigdata/what-is-big-data.html>
- IMF. (2011) Definition and Measurement of Sovereign Risk Need to be Broadened, IMF Roundtable Concludes. Retrieved from <https://www.imf.org/external/np/sec/pr/2011/pr1191.htm>
- Interactive Middle East. (2011) Twitter Usage in the MENA region. Retrieved from <http://interactiveme.com/2011/06/twitter-usage-in-the-mena-middle-east/>
- Iranzo, S. (2008) Delving into country risk. *Banco de España Occasional Paper No. 0802*. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1120723
- Kepes, B. (2014) Twitter Buys Gnip – Its All About the Data. *Forbes*. Retrieved from <http://www.forbes.com/sites/benkepes/2014/04/15/twitter-buys-gnip-its-all-about-the-data/>
- Kesen, S. (2014) Turkey tweeting at a rapid pace. *Daily Sabah*. Retrieved from <http://www.dailysabah.com/features/2014/04/16/turkey-tweeting-at-a-rapid-pace>
- Kobrin, S. J. (1979) Political Risk: a Review and Reconsideration. *Journal of International Business Studies*, 10 (1), 67-80.
- Kobrin, S.J. (1982) *Managing Political Risk Assessment: Strategic Response to Environmental Change*. University of California Press: California.
- Koh, Y. (2014) Twitter Paid \$134 Million for Data Partner Gnip. *The Wall Street Journal Blogs*. Retrieved from <http://blogs.wsj.com/digits/2014/08/11/twitter-paid-134-million-for-data-partner-gnip/>
- Kouloumpis, E., Wilson, T., Moore, J. (2011) Twitter Sentiment Analysis: The Good the Bad and the OMG! *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*. Retrieved from <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/viewFile/2857/3251>
- Krikorian, R. (2010) Tweet Metadata. *The Wall Street Journal*. Retrieved from <http://online.wsj.com/public/resources/documents/TweetMetadata.pdf>
- Lee, K., Tamilarasan, P., Caverlee, J. (2013) Crowdturfers, Campaigns, and Social Media: Tracking and Revealing Crowdsourced Manipulation of Social Media. *Association for the Advancement of Artificial Intelligence*. Retrieved from <http://faculty.cse.tamu.edu/caverlee/pubs/lee13icwsm-crowdturfing.pdf>

- Library of Congress. (2013) Update on the Twitter Archive at the Library of Congress. Retrieved from http://www.loc.gov/today/pr/2013/files/twitter_report_2013jan.pdf
- Lopes, M. (2014) IBM, Twitter to partner on business data analytics. *Reuters*. Retrieved from <http://www.reuters.com/article/2014/10/29/us-twitter-ibm-analytics-idUSKBN0I1T720141029>
- Lopez, T. (2014) How Did Ukraine's Government Text Threats to Kiev's EuroMaidan Protesters? *Slate*. Retrieved from http://www.slate.com/blogs/future_tense/2014/01/24/Ukraine_texting_euromaidan_protesters_kiev_demonstrators_receive_threats.html
- McCarty, B. (2012) TweetMeme, the precursor to DataSift, closes its doors after 500,000 installs. *The Next Web*. Retrieved from <http://thenextweb.com/insider/2012/09/26/tweetmeme-precursor-datasift-closes-doors-500k-installs/>
- McCoy, T. (2014) A computer just passed the Turing Test in landmark trial. *The Washington Post*. Retrieved from <http://www.washingtonpost.com/news/morning-mix/wp/2014/06/09/a-computer-just-passed-the-turing-test-in-landmark-trial/>
- McKinsey. (2013) Infographic: Big data, big profits. Retrieved from <http://mckinseyonmarketingandsales.com/infographic-big-data-big-profits>
- Meldrum, D.H. (2000) Country Risk and Foreign Direct Investment. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.195.9987&rep=rep1&type=pdf>
- Moody, C. (2014) Twitter and IBM Partner to Transform Decision Making. *Gnip Company Blog*. Retrieved from <https://blog.gnip.com/author/chris/>
- Morgan, J. (2012) The DARPA Collaboration Experiment and What it Means for Your Organization. *The Future Organization*. Retrieved from: <http://www.thefutureorganization.com/darpa-collaboration-experiment/>.
- Morozov, E. (2011) *The Net Delusion: How Not to Liberate the World*. Allen Lane: London.
- 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. *Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media*. Retrieved from <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM13/paper/viewFile/6071/6379>

- Nath, H.K. (2008) Country Risk Analysis: A Survey of the Quantitative Methods. Retrieved from http://www.shsu.edu/~tcq001/paper_files/wp08-04_paper.pdf
- NSA. (n.d.) Utah Data Center. Retrieved from <https://nsa.gov1.info/utah-data-center/>
- NYSE. (2013) NYSE Technologies Enters Sentiment Analysis Market with SMA Alliance. Retrieved from <https://exchanges.nyx.com/fr/us-trading-news/nyse-technologies-enters-sentiment-analysis-market-sma-alliance>
- Oanda. (n.d.) Historical Exchange Rates. Retrieved from <http://www.oanda.com/currency/historical-rates/>
- Ozay, H., Celik, O. (2013) Turkey's Economy Shaken by Protests. *Al Monitor*. Retrieved from <http://www.al-monitor.com/pulse/business/2013/06/turkey-business-affected-protests-gezi-taksim.html#>
- Oxford Business Group. (2013) The Report: Turkey 2013, 58.
- Politico. (2014) U.S. secretly built 'Cuban Twitter' to stir unrest. Retrieved from <http://www.politico.com/story/2014/04/united-states-cuba-twitter-105333.html>
- Popper, B. (2014) CNN announces partnerships with Twitter to 'revolutionize' news gathering. *The Verge*. Retrieved from <http://www.theverge.com/2014/1/29/5357826/cnn-announces-partnership-with-twitter>
- Press, G. (2000) A Very Short History of Big Data. *Forbes*. Retrieved from <http://www.forbes.com/sites/gilpress/2013/05/09/a-very-short-history-of-big-data/2/>
- Robock, S.H. (1971) Political Risk: Identification and Assessment. *Columbia Journal of World Business, July/August, 6(4)*, 6. Retrieved from <http://connection.ebscohost.com/c/articles/5553733/political-risk-identification-assessment>
- Roy, A. and Roy, P.G. (1994) Despite Past Debacles, Predicting Sovereign Risk Still Presents Problems, *Commercial Lending Review, 9 (3)*, 92-95.
- RT. (2012) Egypt's costly revolution. Retrieved from <http://rt.com/business/egypt-revolution-economy-impact-657/>
- Ruef, M. (2002) Strong ties, weak ties and islands: structural and cultural predictors of organizational innovation. *Industrial and Corporate Change, 11(3)*, 427-449. Retrieved from <http://www.cs.princeton.edu/~sjalbert/SOC/Ruef.pdf>

- Shirky, C. (2011) Gladwell vs. Shirky: A Year Later, Scoring the Debate Over Social-Media Revolutions. *Wired*. Retrieved from <http://www.wired.com/2011/12/gladwell-vs-shirky/all/>
- Sicular, S. (2013) Gartner's Big Data Definition Consists of three Parts, Not to Be Confused with Three "V"s. *Forbes*. Retrieved from <http://www.forbes.com/sites/gartnergroup/2013/03/27/gartners-big-data-definition-consists-of-three-parts-not-to-be-confused-with-three-vs/>
- Simon, J.D. (1984) A Theoretical Perspective on Political Risk. *Journal of International Business Studies*, 15(3), 123-143. Retrieved from <http://www.jstor.org/stable/154279>
- Singh, M. (2013) Egypt's Economic Crisis: How to Help Cairo Help Itself. *The Washington Institute*. Retrieved from <http://www.washingtoninstitute.org/policy-analysis/view/egypts-economic-crisis-how-to-help-cairo-help-itself>
- Skinner, J. (2011) Social Media and Revolution: The Arab Spring and the Occupy Movement as Seen through Three Information Studies Paradigms, *Sprouts: Working Papers on Information Systems*, 11(169). Retrieved from <http://sprouts.aisnet.org/11-169>
- SMaPP. (2013) A Breakout Role for Twitter? The Role of Social Media in the Turkish Protests. *Social Media and Political Participation Lab, New York University*. Retrieved from http://smapp.nyu.edu/reports/turkey_data_report.pdf
- Starbird, K., Palen, L. (2012) (How) Will the Revolution be Retweeted? Information Diffusion and the 2011 Egyptian Uprising. Retrieved from https://www.cs.colorado.edu/~palen/StarbirdPalen_RevolutionRetweeted.pdf
- Today's Zaman. (2013) International media increases economic costs of Gezi protests. Retrieved from http://www.todayzaman.com/business_intl-media-increases-economic-costs-of-gezi-protests_318554.html
- Tarrow, S.G. (1994) *Power in Movement: Social Movements, Collective Action and Politics*. Cambridge University Press.
- Tech Terms. (n.d.) Metadata. Retrieved from <http://www.techterms.com/definition/metadata>
- The Economist. (2010) All too much. Retrieved from <http://www.economist.com/node/15557421>
- The Economist. (2012) The value of friendship. Retrieved from <http://www.economist.com/node/21546020>

- The Economist. (2014) Mining for tweets of gold. Retrieved from <http://www.economist.com/news/business/21603468-startup-finding-valuable-information-tittersphere-mining-tweets-gold>
- Tilly, C. (2003) *The Politics of Collective Violence*. Cambridge University press.
- Trottier and Fuchs. (2014) Social Media, Politics and the State: Protests, Revolutions, Riots, Crime and Policing in the Age of Facebook, Twitter and YouTube. *Routledge Research in Information Technology and Society*, 178-179
- Twitter (2012) Shutting down spammers. Retrieved from <https://blog.twitter.com/2012/shutting-down-spammers>
- Twitter (n.d.) Election. Retrieved from <https://election.twitter.com/>
- Twitter (n.d.) Privacy Policy. Retrieved from <https://twitter.com/privacy>
- Twitter (n.d.) Twitter search rules and restrictions. Retrieved from <https://support.twitter.com/articles/42646-twitter-search-rules-and-restrictions>
- Twitter. (n.d.) Using hashtags on Twitter. Retrieved from <https://support.twitter.com/articles/49309-using-hashtags-on-twitter>
- Twitter Developers. (n.d.) FAQ. Retrieved from <https://dev.twitter.com/faq>
- Twitter Developers. (n.d.) Overview. Retrieved from <https://dev.twitter.com/oauth/overview>
- Twitter Developers (n.d.) Rate Limits: Chart. Retrieved from <https://dev.twitter.com/rest/public/rate-limits>
- University at Albany. (n.d.) Information Studies FAQ. Retrieved from http://www.albany.edu/informationstudies/ist_faq.php
- Vincent, J. (2013) Topsy: who are they are why did Apple buy them? *The Independent*. Retrieved from <http://www.independent.co.uk/life-style/gadgets-and-tech/topsy-who-are-they-and-why-did-apple-buy-them-8979882.html>
- Wagner, J. (2014) Two Great Social Data Platforms: How DataSift and Gnip Stack Up. *Programmable Web*. Retrieved from <http://www.programmableweb.com/news/two-great-social-data-platforms-how-datasift-and-gnip-stack/brief/2014/02/10>
- Wakabayashi, D., MacMillan, D. (2013) Apple Taps into Twitter, Buying Social Analytics Firm Topsy. *The Wall Street Journal Blogs*. Retrieved from <http://www.wsj.com/articles/SB10001424052702304854804579234450633315742>

Wall Street Journal. (n.d.) Historical Prices. Retrieved from <http://quotes.wsj.com/TR/XIST/XU100/index-historical-prices#>

World Bank (n.d.) Worldwide Governance Indicators. Retrieved from <http://info.worldbank.org/governance/wgi/index.aspx#home>