# Research Proposal

M. Callaghan and V. Nieberg

October 23, 2015

## Research Proposal for M. Callaghan and V. Nieberg

### Introduction

Twitter is an online social network that allows users to broadcast short posts known as tweets. Since its launch in 2006, the platform has increasingly been used for everyday communication as well as for political debates, crisis communication, marketing, and cultural participation (Weller et al. 2013). The public-debt crisis in Europe is widely discussed across Europe and presents an interesting point in time to investigate whether European issues are discussed in a common European public sphere. This project will do so by looking at data from the communication platform Twitter. It will specifically look at the reaction in the Twittersphere to the negotiation between the Troika and Greece leading up to the signing of the three memorandums.

### Research Question

In our research project we will investigate the following questions:

What can twitter tell us about pan-European reactions to the European governance of the public-debt crisis in Greece?

- What can variation across time and space in the volume of tweets regarding the euro crisis tell us about popular engagement with the issues?

- What can the content of tweets related to the crisis tell us about the spread of public opinion on the handling of the crisis in Greece between and within countries?

By investigating European public discourse on the Euro crisis the answer to these questions could potentially add to the literature on the emergence of a European public sphere.

### Literature Review

#### On Twitter Research

The body of twitter research has constantly grown over the past years (for a comprehensive analysis and typology of twitter research up to 2013, see Zimmer and Proferes 2014). Some of the findings with relevance to our research design are discussed below.

Twitter is a Source for Meaningful Information about Engagement with and Opinions about Political Topics. Twitter is used as a platform for political deliberation. In a recent study on Tweets mentioning parties or politicians before the 2009 German federal election, Tumasjan et al. found that "Twitter is not just used to spread

political opinions, but also to discuss these opinions with other users" (Tumasjan et al. 2010, 183). Furthermore, specific patterns of twitter usage have been identfed that correspond with high-profile political events. Hughes and Palen found that, compared to general Twitter usage, more broadcast-based information sharing activities happen (Hughes and Palen 2009, 259). Moreover, Tumasjan et al. found that it was possible to extact meaningful information about political opinions from both the volumes and the content of these tweets: "the mere number of tweets reflects voter preferences and came close to traditional election polls" (Tumasjan et al. 2010, 183).

Twitter gives information on location of tweets and users, which must be carefully interpreted. Devin Gaffney points out methodological problems with using the given location of twitter users - "in many cases user-entered profile locations differ from the physical locations users are actually tweeting from" (Graham, Hale, and Gaffney,Devin 2014, 1) which must be considered when interepreteing results.

Though the field of Sentiment Analysis (SA) is perhaps most developed in the business world (Zimmer and Proferes 2014, 250), an increasing body of literature has developed, focused on retrieving information about political opinions from the Twittersphere. Though Tumasjan's results have come under scrutiny (see Jungherr, Jürgens, and Schoen 2012), the authors found that "the sentiment of Twiter messages closely corresponded to political programs, candidate profiles, and evidence from the media coverage of the campaign trail" (Tumasjan et al. 2010, 183).

Grimmer provides an overview of recent developments in SA in political science, noting how "automated content methods can make possible the previously impossible in political science: the systematic analysis of large-scale text collections without massive funding support" (Grimmer and Stewart 2013, 2). He advises caution, however, about the utility of SA in predictive models: "The goal of building text models is therefore different than model building to make causal inferences. [...] Emphasis in evaluations should be placed on helping researchers to assign documents into predetermined categories, discover new and useful categorizytion schemes for texts, or in measuring theoretically relevant quantities from large collections of text." (Grimmer and Stewart 2013, 4).

Due to the enourmous amount of text available, Pak and Paroubek identify that "microblogging web-sites are rich sources of data for opinion mining and sentiment analysis" (Pak and Paroubek 2010, 1320). The multilingual nature of tweets across Europe presents some difficulties, but is the subject of a growing body of research: "Noisy social media, such as Twitter, are especially interesting for sentiment analysis (SA) [...] given the amount of data and their popularity in different countries, where users simultaneously publish opinions about the same topic in different languages" (Vilares, Alonso, and Gómez-Rodriguez 2015, 2). Balahur and Turchi are confident about the ability of Statistical Machine Translation (SMT) to provide a basis for consistently applied SA across languages (Balahur and Turchi 2012, 58). Other approaches include using emoticons to train models that assign sentiment to a multilingual text corpus (Narr, Hulfenhaus, and Albayrak 2012).

Finally, some studies discuss ethical aspects of twitter research. For example, concerns about creating a permanent archive of tweets have been voiced. These concerns included whether "such archive was aligned with users' privacy expectations" (Zimmer and Proferes 2014, 258; Zimmer 2010).

**On Awareness and Public Opinion across Europe on the Governance of the Public-Debt Crisis in Greece**

Academic research on the emergence of a European public sphere is not a recent phenomenon (Risse 2003, 1). Hitherto, however, research has been characterized as rather normative, as the "research community has been [...] interested in producing policy recommendations for public sphere-building" (Trenz 2015, 234). Recent studies, on the other hand, seem to put emphasis on an empirical grounding of the debate (Trenz 2015; Drewski, Gerhards, and others 2015). This development is being mirrored in research on the public debate across Europe on the euro crisis. It has been suggested that "there is an emerging demos in the European polity and it has been strengthened during the euro crisis" (Risse 2014, 1213). When testing this hypothesis empirically, though, by looking at newspaper editorials in Spain and Germany, Drewski found that there were significant differences along national instead of ideological lines in the discussion of the Euro crisis (Drewski, Gerhards, and others 2015, 5).

Max Hänska and Stefan Bauchowitz in a recent LSE blog entry track twitter activity during the negotiations leading up to the third Greek bailout agreement. (Haenska and Bauchowitz 2015) According to their findings,

tweets synchronised around key mini-events throughout the negotiations, with peaks and troughs mirrored across national twitter-spheres. These results suggest that popular engagement with the issue converges across Europe.
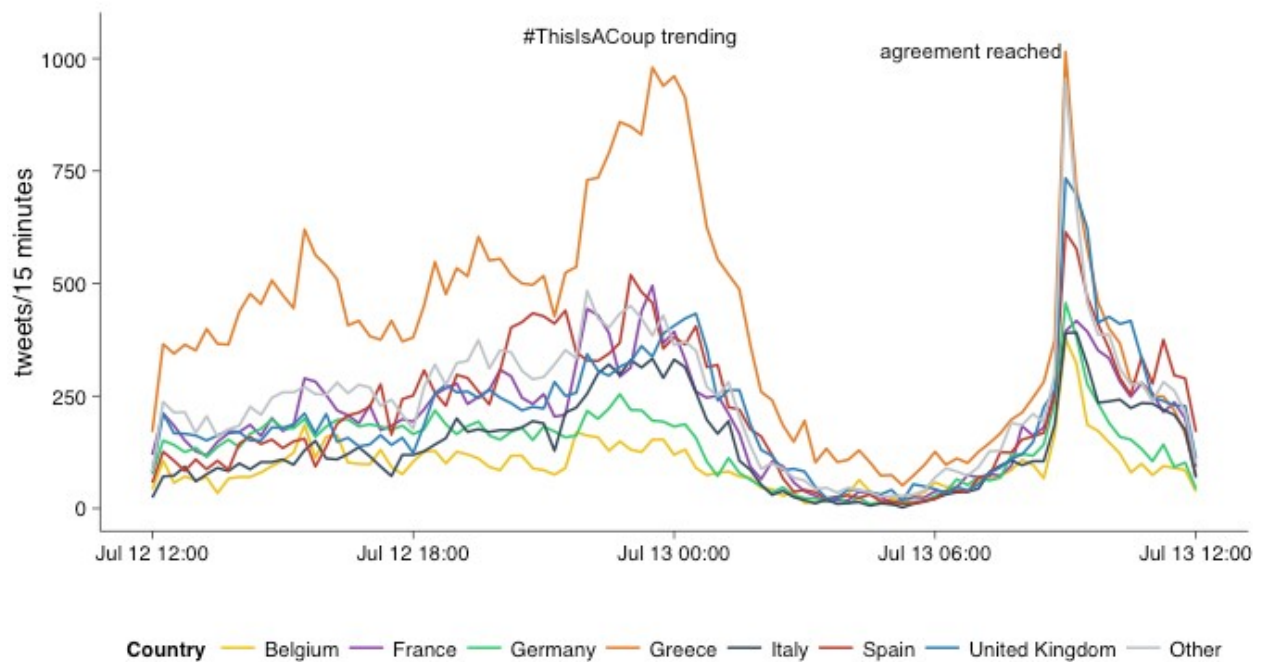


They further looked at instances of tweets containing #ThisIsACoup, representing a particular opinion on the agreement. The then showed that the spread of #ThisIsACoup was not reflected in the studied countries equally. This indicated a divergence of public opinion along national lines.
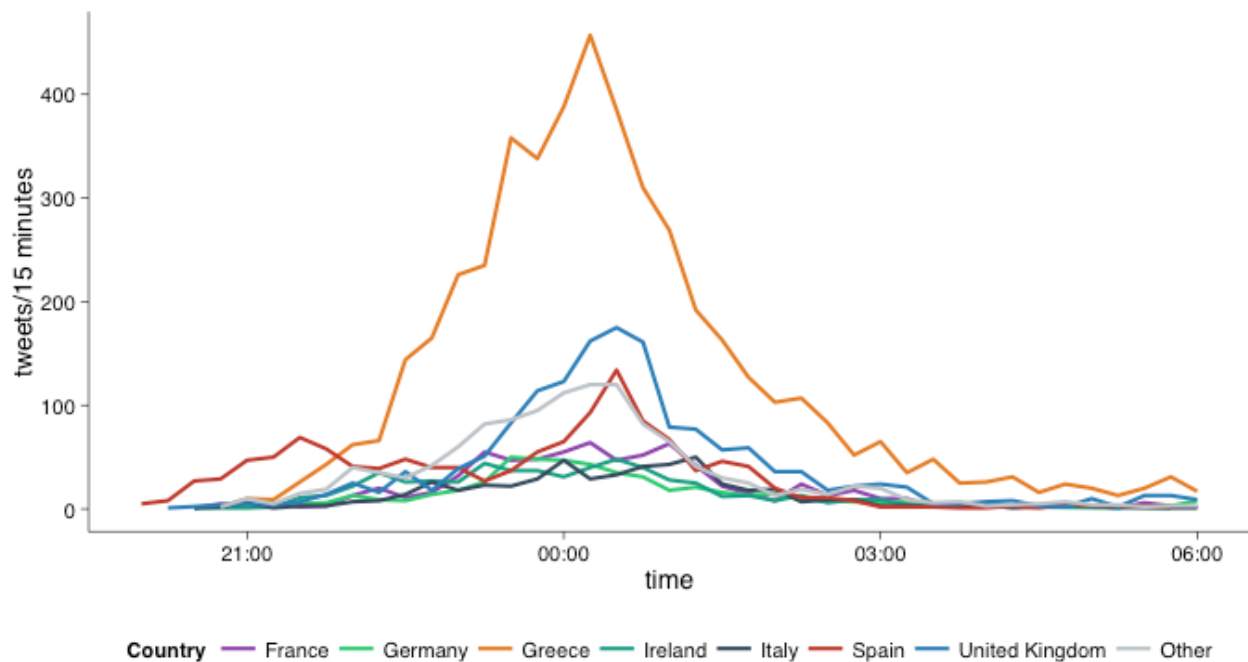
## Data Sources

This decrease in the research output and size of datasets analyzed is possibly related to changes Twitter made to its API and terms of service in early 2011 (Melanson, 2011; Ramji, 2011) that limited researchers' access to Twitter data and effectively shut down popular services used by researchers to track and archive Twitter activity, such as TwapperKeeper and 140kit (Watters, 2011; Sample, 2011). (Zimmer and Proferes 2014, 257)

Two datasets are required for this project. The first is a corpus of tweets relating to the Greek debt crisis and the measures taken to manage the crisis by European institutions. The second is information about the users whose tweets form the body of that corpus.

Zimmer and Proferes identify the Library of Congress' decision to place every tweet since Twitter's inception in 2006 into an archive as validating "the research importance of twitter" (Zimmer and Proferes 2014, 251). Despite this announcement occuring in 2010, five years later, the archive is still not open to researchers (Scola).

Since late 2014, the whole corpus of twitter data has been searchable online (Metz). Programmatic access to this archive is, however, more restricted. Twitter's public search API "is not complete index of all Tweets, but instead an index of recent Tweets. At the moment that index includes between 6-9 days of Tweets." ("The Search API"). Twitter sells access to historical tweets through an API provided by its "enterprise API platform" GNIP (Tornes 2015). This paper will adapt a publicly available program written in Java which scrapes results from Twitter's online search page (Henrique 2015). A list of queries involving combinations of keywords to do with the Greek debt crisis will be drawn up, and we will programmatically run through the list, using the GetOldTweets software to scrape the tweets returned by Twitter's comprehensive online search function that are given by each query.

We can use the R package 'TwitteR' (Gentry 2015) to retrieve more information about the unique users in our corpus dataset. The two datasets can then be merged so that tweets can be mapped by location.

## Methodology

- Volumes of topic-relevant tweets will be mapped across space and time, to analyse the distribution of topic-awareness and its relation to political developments in responses to the crisis.

- The distribution of hashtags that clearly represent an opinion on the response to the crisis (e.g. '#ThisIsaCoup', '#ThisIsNotaCoup' *inter alia*) will be similarly mapped in order to approximate the distribution of opinion within and between countries over time.

- The twittersphere's validity in representing political opinion will be tested by comparing volumes of tweets containing the hashtags '#ναί' and '#όχι' with the results of the referendum.

- The paper will attempt a sentiment analysis of tweets expressing opinions about the agreed bailout deals using either/or

  - A multilingual EU funded open-source sentiment analysis tool (opeNER)
  - Machine translation to translate all texts into English before performing sentiment analysis
  - Sentiment analysis based on emoticons

Option two will be preferred, due to simplicity, and analysis will then be carried out using the unigrams to indicate polarity through comparison with a lexicon. Following Grimmer, "We will assume documents are a *bag of words*, where order does not inform our analyses" as "In practice, for common tasks like measuring sentiment, topic modeling, or search, *n-grams* (combinations of words rather than individual words) do little to enhance performance" (Grimmer and Stewart 2013, 6).

# References

Balahur, Alexandra, and Marco Turchi. 2012. "Multilingual Sentiment Analysis Using Machine Translation?" *Proceedings of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis P. 52-60*. http://publications.jrc.ec.europa.eu/repository/handle/JRC73293.

Drewski, Daniel, Jürgen Gerhards, and others. 2015. "Has There Been a European Public Discourse on the Euro Crisis?"

Gentry, Jeff. 2015. *TwitteR: R Based Twitter Client*. http://CRAN.R-project.org/package=twitteR.

Graham, Mark, Scott a. Hale, and Gaffney,Devin. 2014. "Where in the world are you ? Geolocation and language identification in Twitter." *The Professional Geographer* 00 (2013): 00. doi:10.1080/00330124.2014.907699.

Grimmer, Justin, and Brandon M Stewart. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis*. SPM-PMSAPSA, mps028.

Haenska, Max, and Stefan Bauchowitz. 2015. "A European Twitter Sphere? What Tweets on the Greek Bailout Say About How Europeans Interact Online." Accessed October 23. http://blogs.lse.ac.uk/impactofsocialsciences/2015/10/18/a-european-twitter-sphere-what-tweets-on-the-greek-bailout-say-about-how-europeans-i

Henrique, Jefferson. 2015. "Get Old Tweets Programmatically." Accessed October 21. https://github.com/Jefferson-Henrique/GetOldTweets.

Hughes, Amanda Lee, and Leysia Palen. 2009. "Twitter adoption and use in mass convergence and emergency events." *International Journal of Emergency Management* 6 (May): 248. doi:10.1504/IJEM.2009.031564.

Jungherr, Andreas, Pascal Jürgens, and Harald Schoen. 2012. "Why the Pirate Party Won the German Election of 2009 or the Trouble with Predictions: A Response to Tumasjan, a., Sprenger, tO, Sander, PG, & Welpe, IM 'Predicting Elections with Twitter: What 140 Characters Reveal About Political Sentiment'." *Social Science Computer Review* 30 (2). SAGE Publications: 229–34.

Metz, Cade. "Twitter Now Lets You Search for Any Tweet Ever Sent." http://www.wired.com/2014/11/twitter-now-lets-search-tweet-ever-sent.

Narr, Sascha, Michael Hulfenhaus, and Sahin Albayrak. 2012. "Language-Independent Twitter Sentiment Analysis." *Knowledge Discovery and Machine Learning (KDML), LWA*, 12–14.

Pak, Alexander, and Patrick Paroubek. 2010. "Twitter as a Corpus for Sentiment Analysis and Opinion Mining." *In Proceedings of the Seventh Conference on International Language Resources and Evaluation*, 1320–26. doi:10.1371/journal.pone.0026624.

Risse, Thomas. 2003. "An Emerging European Public Sphere? Theoretical Clarifications and Empirical Indicators."

———. 2014. "No Demos? Identities and Public Spheres in the Euro Crisis." *JCMS: Journal of Common Market Studies* 52 (6): 1207–15. doi:10.1111/jcms.12189.

Scola, Nancy. "Library of Congress' Twitter Archive Is a Huge #FAIL." http://www.politico.com/story/2015/07/library-of-congress-twitter-archive-119698.html.

"The Search API." https://dev.twitter.com/rest/public/search.

Tornes, Adam. 2015. "Instant and Complete Access to Every Historical Public Tweet." https://blog.twitter.com/2015/full-archive-search-api.

Trenz, Hans-Jörg. 2015. "Europeanising the Public Sphere – Meaning, Mechanisms, Effects." In *Interdisziplinäre Europastudien*, edited by Ulrike Liebert and Janna Wolff, 233–51. Springer Fachmedien Wiesbaden. doi:10.1007/978-3-658-03620-1_11.

Tumasjan, Andranik, Timm Oliver Sprenger, Philipp G Sandner, and Isabell M Welpe. 2010. "Predicting Elections with Twitter: What 140 Characters Reveal About Political Sentiment." *ICWSM* 10: 178–85.

Vilares, David, Miguel A Alonso, and Carlos Gómez-Rodriguez. 2015. "Sentiment Analysis on Monolingual, Multilingual and Code-Switching Twitter Corpora." In *6TH WORKSHOP oN COMPUTATIONAL APPROACHES tO SUBJECTIVITY, SENTIMENT aND SOCIAL MEDIA ANALYSIS WASSA 2015*, 2.

Weller, Katrin, Axel Bruns, Jean Burgess, Merja Mahrt, and Cornelius Puschmann. 2013. *Twitter and Society*. Vol. 89. Peter Lang New York.

Zimmer, Michael. 2010. "Is It Ethical to Harvest Public Twitter Accounts Without Consent?" http://www.michaelzimmer.org/2010/02/12/is-it-ethical-to-harvest-public-twitter-accounts-without-consent/.

Zimmer, Michael, and Nicholas John Proferes. 2014. "A topology of Twitter research: disciplines, methods, and ethics." *Aslib Journal of Information Management* 66 (3): 250–61. doi:10.1108/AJIM-09-2013-0083.