The Refugee Crisis in Europe: The (Weak) Connection Between Political Media and Twitter Activity

Aleksandar Stojkov, Ss. Cyril and Methodius University (Macedonia)

Figure 1: Syrian refugees on the Macedonian-Greek border attempting to reach Germany

Source: Photo taken by the authors.

Research inspiration

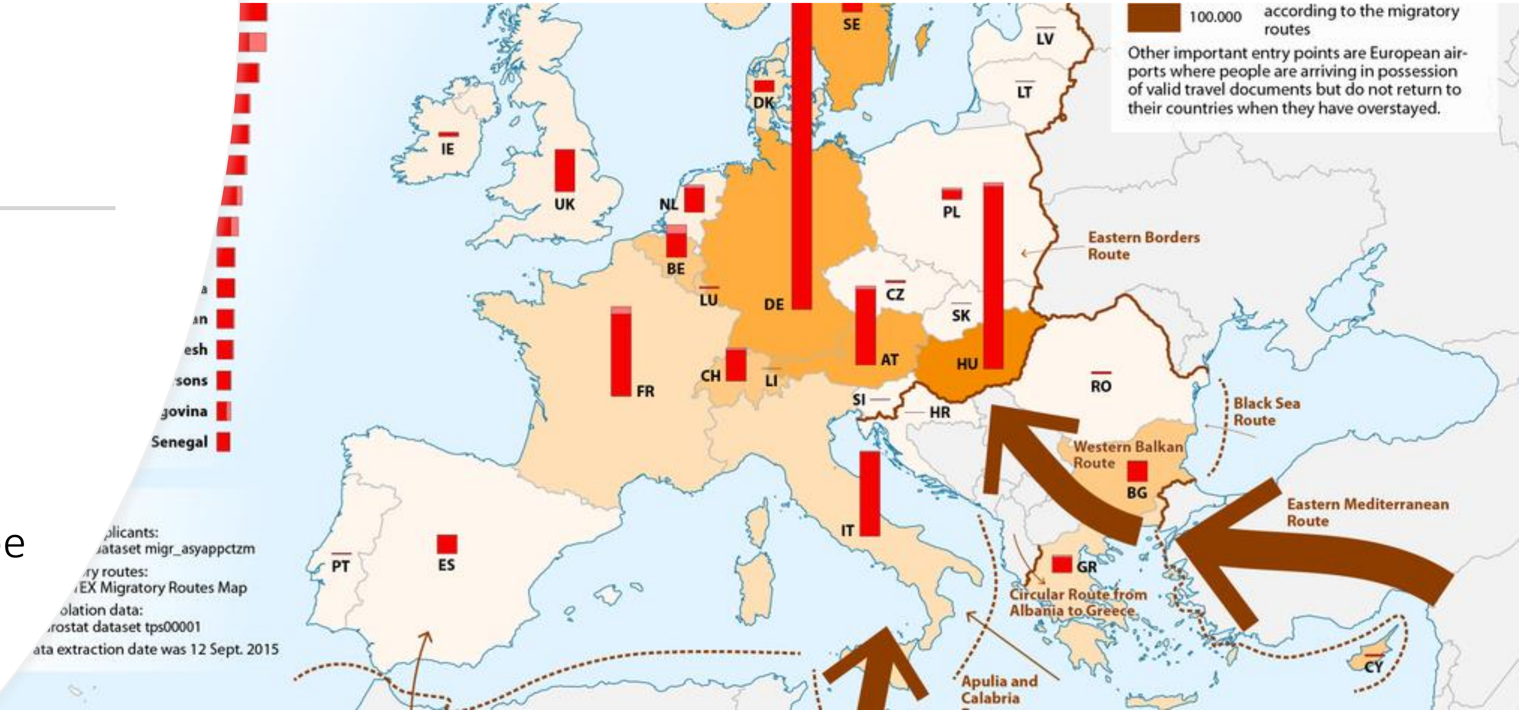


Figure 2: Migratory routes to Europe

Source: Syria's refugee crisis in numbers, Amnesty International, September 4th, 2015.

Outline of the Presentation

INTRODUCTION

RESEARCH CONTEXT

LITERATURE REVIEW

METHODOLOGY

RESULTS

1. Introduction

- A well-documented power of the news media to set a nation's political agenda and to focus public attention on a few key public issues
- The news media are a primary source of those "pictures in our heads" (Lipmann, 1922) about the larger world of public affairs, a world that for most citizens is "out of reach, out of sight, out of mind."
- What we know about the world is largely based on what the media decide to tell us. More specifically, the result of this mediated view of the world is that the priorities of the media strongly influence the priorities of the public.
- **Public agenda-setting:** The agenda-setting theory and its hypotheses inform political communications studies of media influence
- **Policy agenda-setting:** Examines the policymaking process

Our contribution

- Our contribution builds on the agenda-setting theory.

- Using a Bayesian way of thinking, we focus on one **issue** – the refugee crisis conversation, one **territory** – Europe – and we look at the different actors' **agendas**.

- Issues are often proxied in the literature by conversations pushed forward in the news media. Thus, we will use news media as a proxy, but we will also consider conversations directly generated by other actors, namely political groups and politically-motivated individuals

Research objectives

- **Central research objective**:
  - To examine the opinion formation process and the driving forces of influential thinking on Twitter activity regarding the European refugee crisis 2014/15.
  - This is a methodological contribution (not a search for a potential causal relationship)

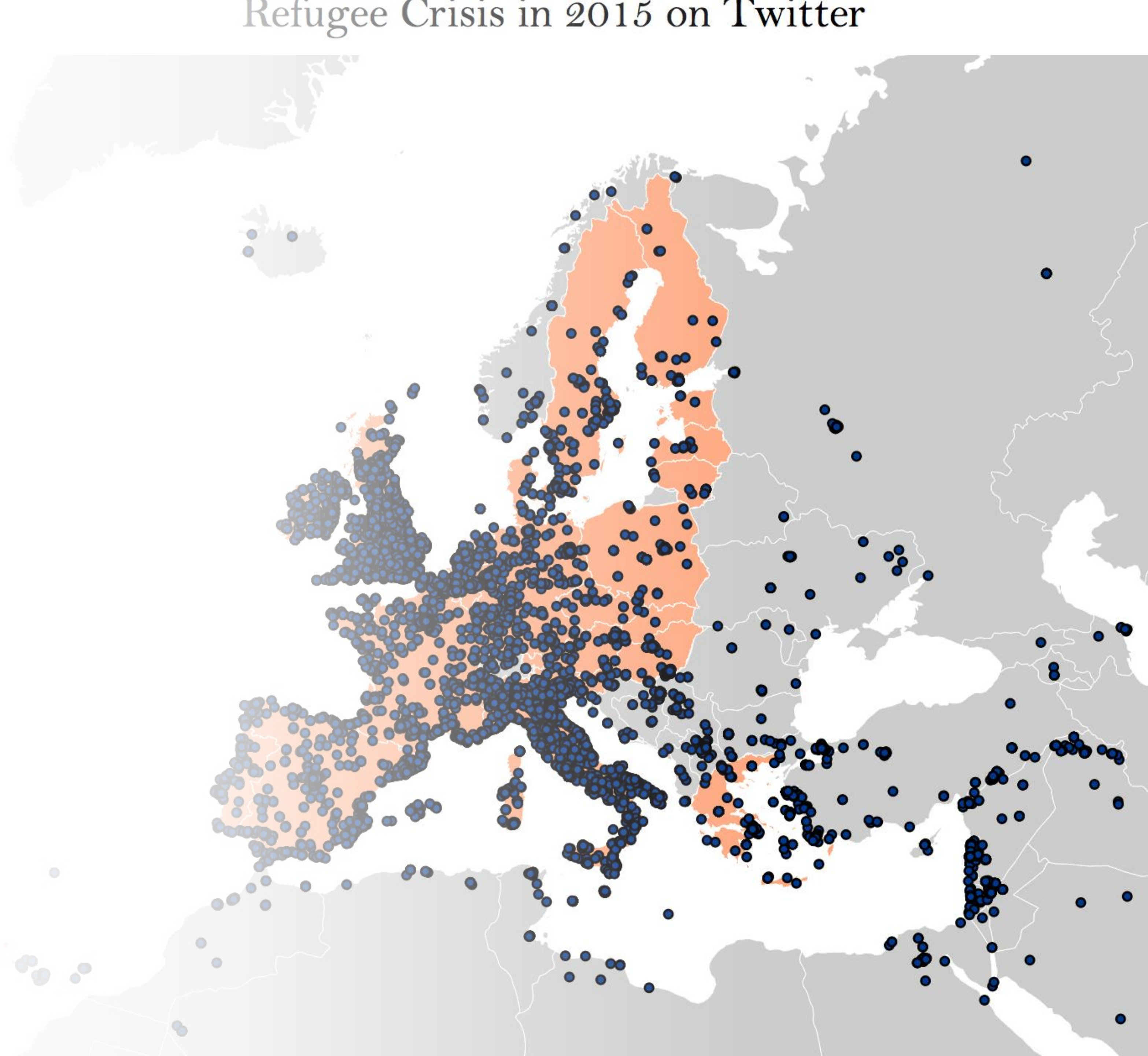
- **More specific research questions**:
  - (1) what content starts a conversation?;
  - (2) which kind of user (from the first layer) triggers a conversation, news agencies or individuals?, and
  - (3) what are the features of an influencer?

2. Research Context

- **Increase in asylum applicants.**
- The number of persons seeking asylum from non-EU countries in the EU-28 during the third quarter of 2015 reached 413,800. This was 250,400 more than in the same quarter of 2014. Out of the 430,600 total asylum applicants, 413,800 (96%) were first time applicants.

- **Countries of origin.**
- Citizens of 149 countries sought asylum for the first time in the EU in the third quarter of 2015. Syrians, Afghans, and Iraqis were the top three citizenships of asylum seekers, lodging around 138,000, 56,700 and 44,400 applications, respectively.

- **Destination countries.** The highest number of first-time asylum applicants in the third quarter of 2015 was registered in Germany and Hungary (both with slightly over 108 000 applicants, or 26% each of total applicants in the EU Member States), followed by Sweden (42,500, or 10%), Italy (28,400, or 7%) and Austria (27,600, or 7%).



Refugee Crisis in 2015 on Twitter

The most intensive Twitter activity appears in UK (pre-Brexit), Italy, Germany, and France

3. Literature Review

- Twitter
- Agenda-setting theory
- Agenda-setting and political topics
- Media bias
- Agendas and social media

3.1. Twitter

- **Value of Twitter and Crisis Responses**
  - Location-based social networks like Twitter hold significant research potential
    - Evacuations (An et al., 2013);
    - Gentrification (Gibbons et al., 2018);
    - Happiness (Mitchell et al., 2013);
    - Hurricanes (Madireddy et al., 2015; Sadri et al., 2018);
    - Politics (Jung and Shin, 2018; Stefanidis et al., 2013);
    - Traffic (Gu et al., 2016; Zhang et al., 2016);
    - Transit (Schweitzer, 2014; Nisar and Prabhakar, 2018), etc.
  - Researchers can analyze the sentiments in tweets
  - “The assumption is that the aggregated judgment of several people is often better or more precise than the judgment of experts or the smartest forecaster” (Hogarth, 1978).

Twitter (Cont'd)

- **Use of Twitter in Politics through Content Generation**

  - Twitter has become an important medium for debating about politics, organizing collective action, and showing support for, or critique of politicians and political issues.
    - Prominent campaigning tool
    - A political communication space
    - Opportunity for other Twitter users to contact them
    - Marketing stunt, to generate positive press
    - Interaction and socialization platform.

Twitter (Cont'd)

- Increasingly, researchers turn to digital trace data in the analysis of social phenomena. We can group these approaches into two categories.

- In one, Twitter is treated as a sensor in documenting the reactions of users to their direct or mediated experiences through data traces produced by interactions of users with the service (Jungherr, Schoen, et Jürgens 2015; Shamma et al. 2010).

- In the second category of studies, researchers go even as far as to draw inferences on attitudes, affiliations, and opinions of Twitter users based on the data traces of their behavior on the platform (Barberá et Steinert-Threlkeld 2020).

Twitter (Cont'd)

- In general, these findings illustrate the importance of political coverage by traditional media for Twitter activity related to politics.
- However, this relationship is not deterministic as only a selection of mediated events creates significant volume spikes on Twitter and the intensity of Twitter coverage of political actors follows different patterns from its coverage in traditional media.
- Prominent messages posted during mediated events also show that Twitter is used as a space for contextualizing and contesting the events presented by traditional media, thereby potentially opening the political communication space to new actors (Jungherr 2015c).

3.2. Agenda-setting Theory

- Starting with the seminal contribution by McCombs et Shaw (1972) on Chapel Hill voters during the 1968 presidential campaign, a host of empirical studies have investigated the presence and extent of agenda-setting effects, i.e., of a causal relationship that goes from the coverage of issues on the mass media to the priorities entertained by the public.

- Experimental evidence, such as that provided by Iyengar, Peters and Kinder (1982) lends the strongest support to this hypothesis. The choice itself of the topics being covered by the news media could produce electoral effects, to the extent that most citizens consistently perceives one party to be more competent than the other at handling a given policy issue. This is the notion of issue ownership (Petrocik, 1996).

- mass media outlets can influence the agenda of the public (Erbring, Goldenberg, et Miller 1980; Iyengar et Simon 2000).

3.5. Agendas and Social Media

- When we talk about media bias and agenda-setting nowadays, what comes in mind is the “fake news” concept. Fake news enter into the issue aspect of the agenda-setting theory. The dynamics of fake news on social media are quite interesting (Guess, Nagler, et Tucker 2019), notably to look at the notion of influence (Alizadeh et al. 2020), but our article is more about the agenda-side.

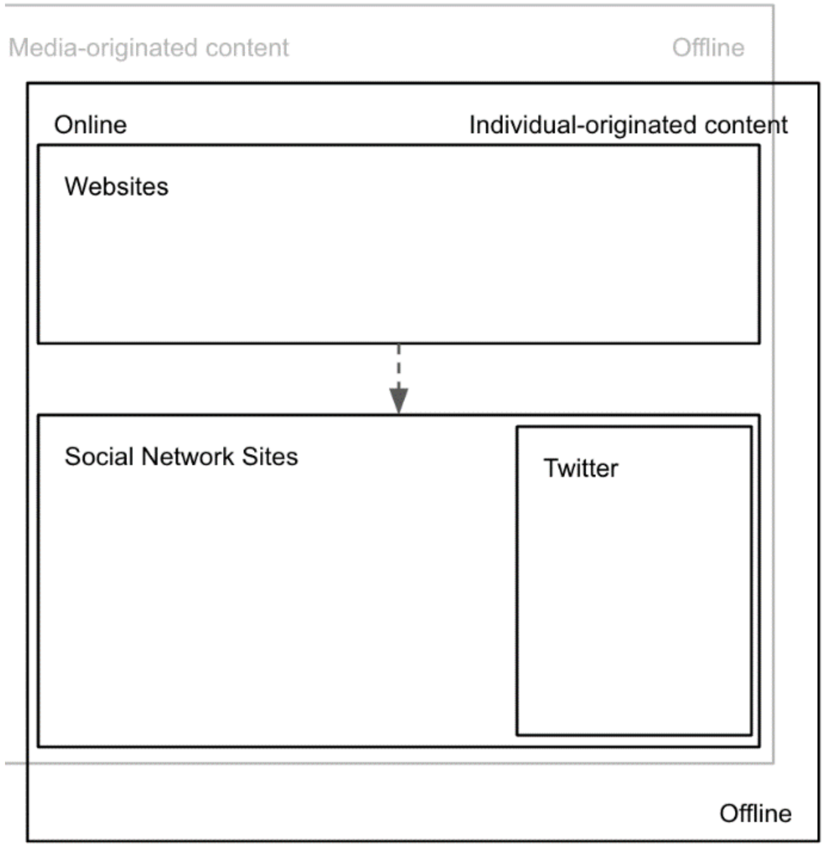
- The literature has also looked at the media slant from the network’s content to the local news outlets. How does it work?

- In an analog world, all the conversations are, by definition, offline. With the Internet, the first stage was for news media outlets to create a website and push their conversations on the newly appeared social media. Rapidly, social media were taken over by individuals and groups. A stage 2 was to push content generated by individuals to social media. People would decide the story to push and would build up their case by using other sources (see Figure 1).

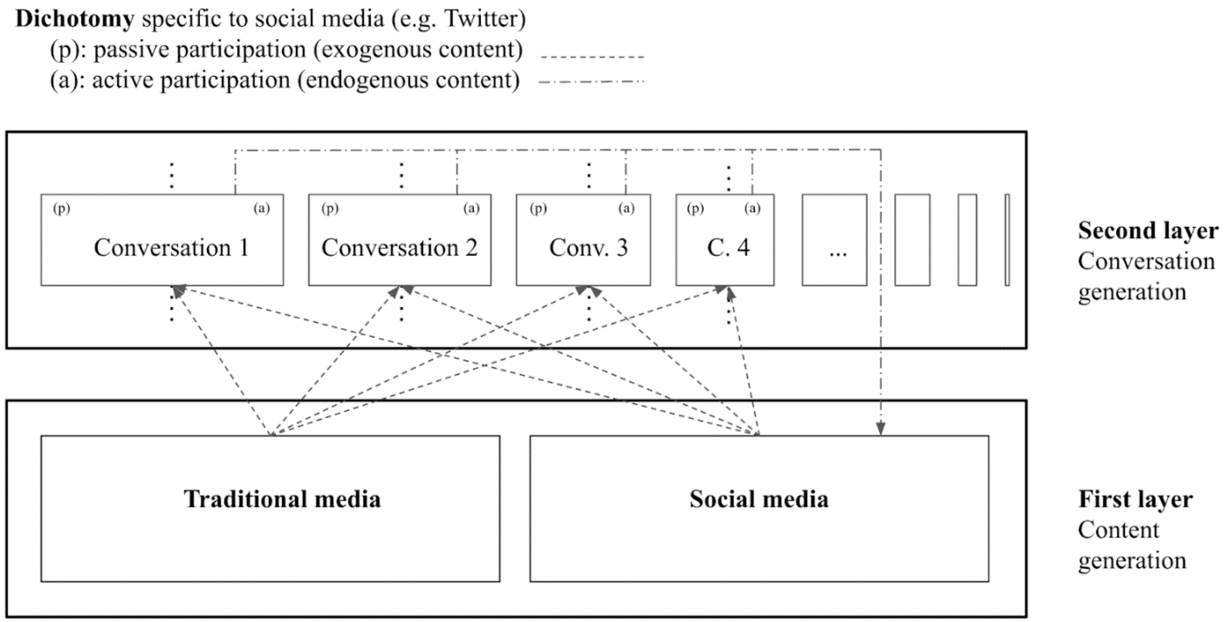
3.5. Agendas and social media (cont'd)

- Our research here takes place at a time when almost all newspapers, TV channels, radio stations and magazines publish their relevant content to Twitter.

- There is also on Twitter another source of information: the crowdsourced information. It is information generated by people with topics that may not have been covered (yet) by traditional media.



Traditional Media vs. Social Media



- **Traditional media:**
  - Almost all newspapers, TV channels, radio stations and magazines publish their relevant content to Twitter.
- **Crowdsourced information:**
  - Information generated by people with topics that may not have been covered (yet) by traditional media

Methodology: Deconstruction of Tweets: The endogenous versus the exogenous dimension

- The content generation layer (the first layer) is first composed of exogenous information coming from traditional and social media.
- Then, users generate the conversations in both an active (editorialized) and passive way (retweets, for instance).
- As such, the editorialized content will feed the first layer in our taxonomy, as it corresponds to some new - though related - content generation.
- This pattern illustrates the endogenous dimension on Twitter.

Methodology
Content analysis: Natural Language Processing

- Feature Selection Strategy in Text Classification:
  - We formulate the feature selection process as a dual objective optimization problem and identify the best number of features for each document, rather than determining a fixed threshold which optimizes the overall classification accuracy for different categories.
  - We provide a documented framework for conducting text preprocessing in text classification in order to optimize the classifier performances, regardless of which classification model one intends to use.
- Refining Content Analysis
  - We estimate four attributes (gender, age, occupation, and interests) of a Twitter user from the contents (a profile document and tweets) generated by the user and the user’s social neighbors, i.e. those with whom the user has conversed (mentioned)

Top Hashtags

- europeanmigrantcrisis
- migrants
- refugees
- refugeeswelcome
- refugeecrisis
- réfugiés
- MigrationEU
- migrazione
- rifugiati
- migranti
- fluchtlinge
- asylmissbrauch
- refugiados
- norefugees

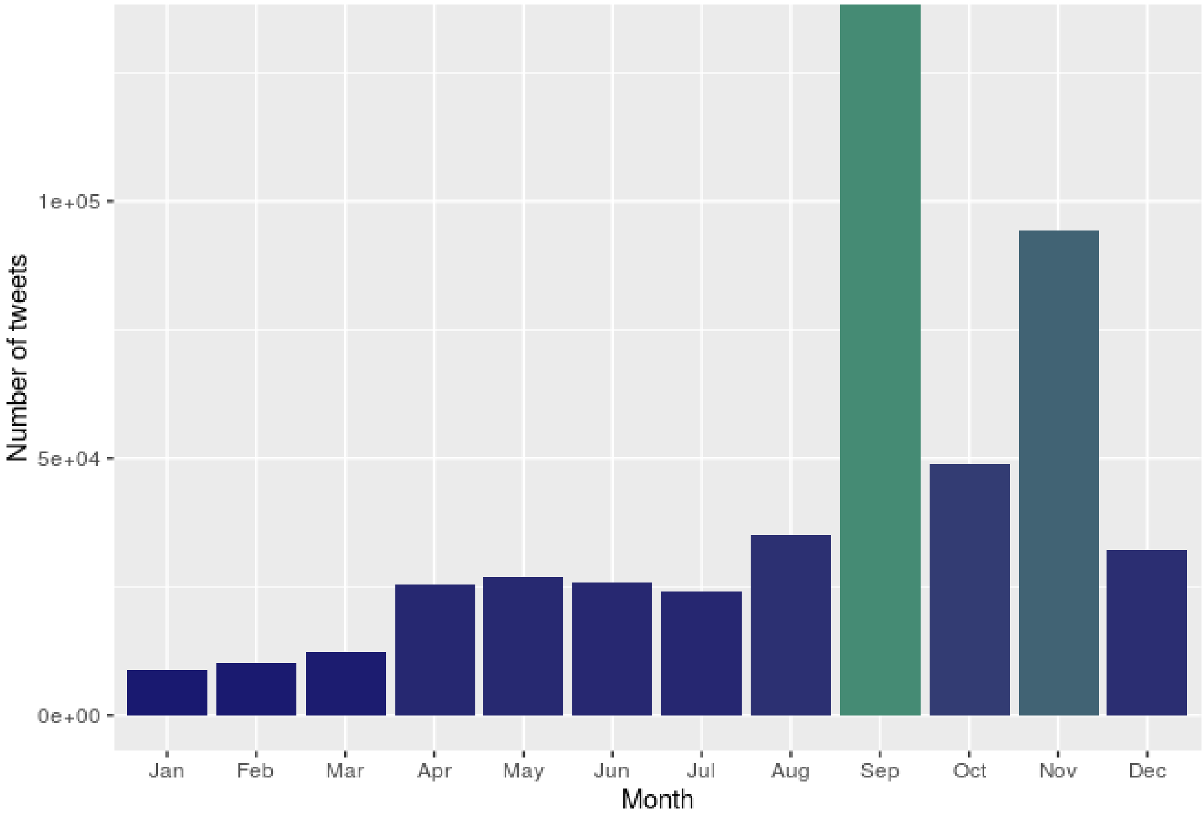
Results

- We collected tweets and their metadata (latitude, longitude, retweets, hashtags, etc.) over a period of 3 years.

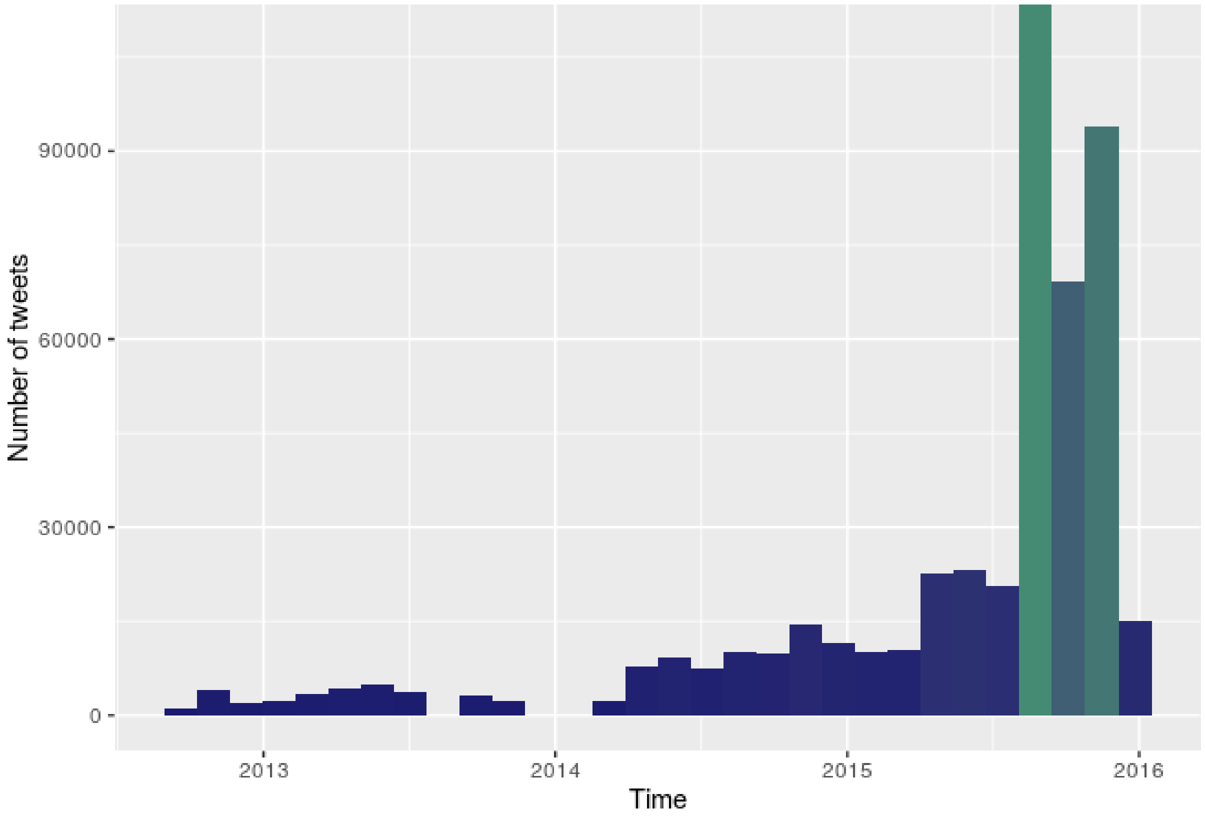
- We collected a total of 482,869 messages from September 9th, 2012 to December 16th, 2015.

Results:
Distribution of Tweets

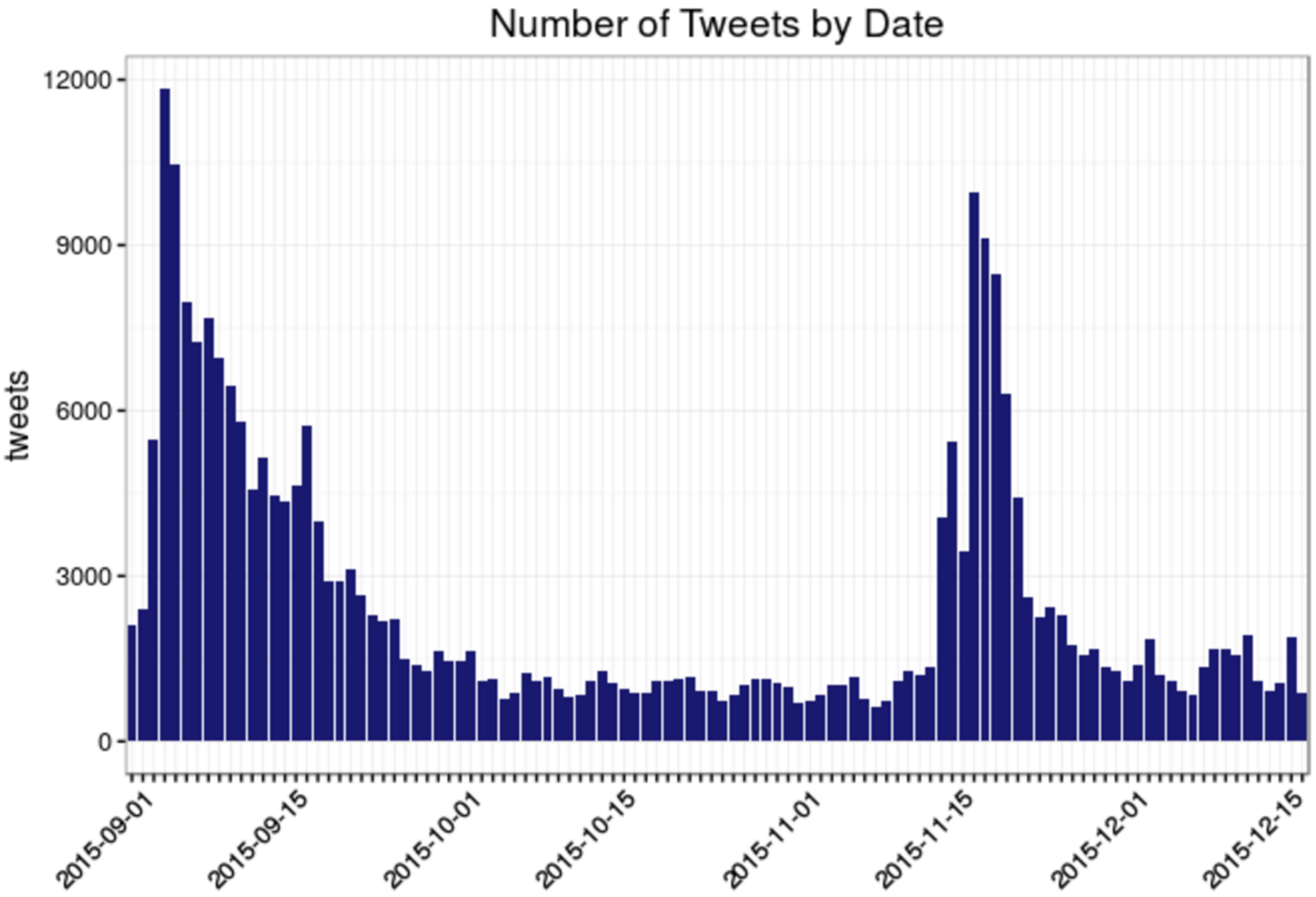
Overall distribution of tweets



Distribution of tweets by month

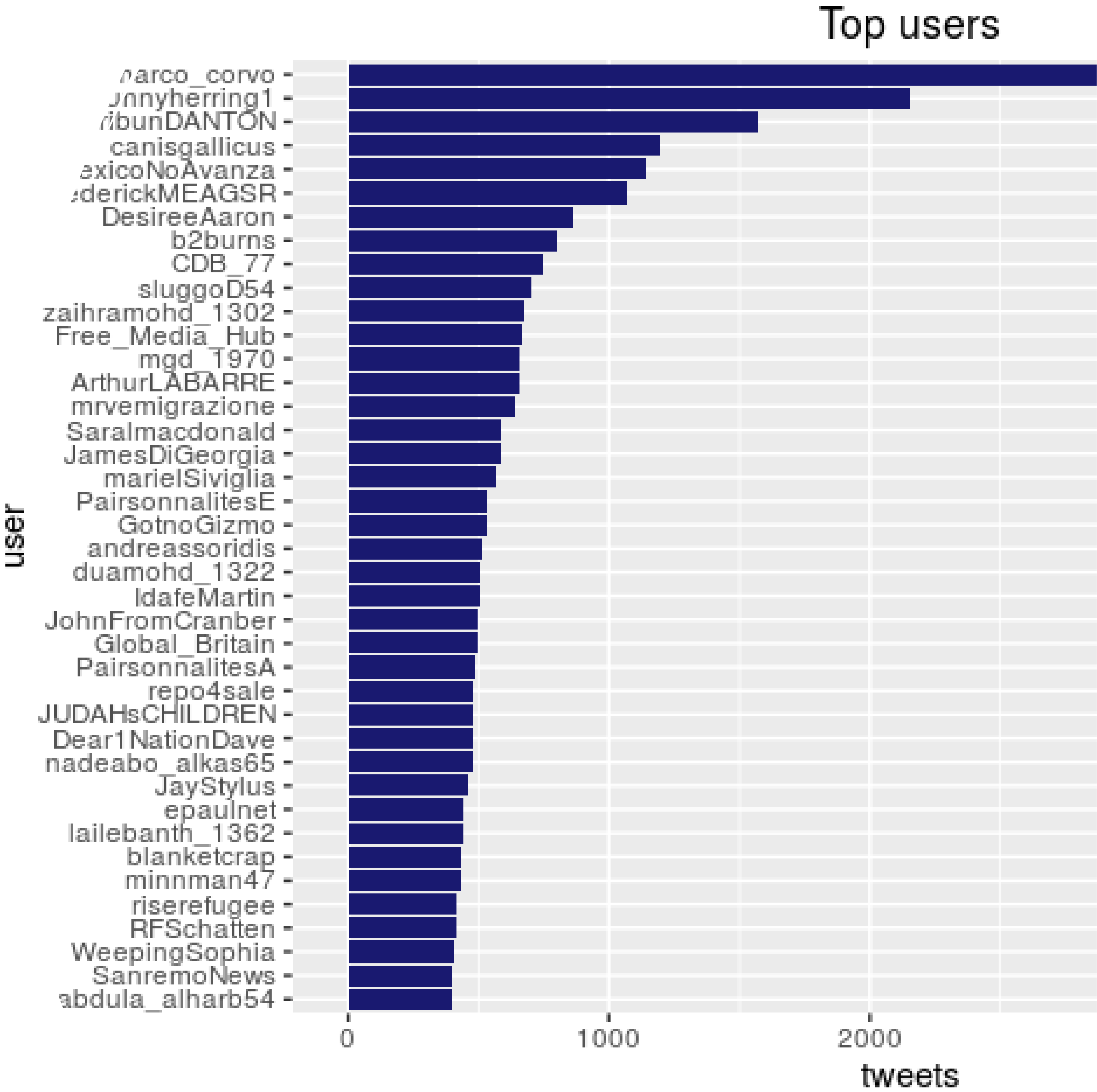


Results



Day with the highest amount of tweets: September 3rd, 2015 with 11,842

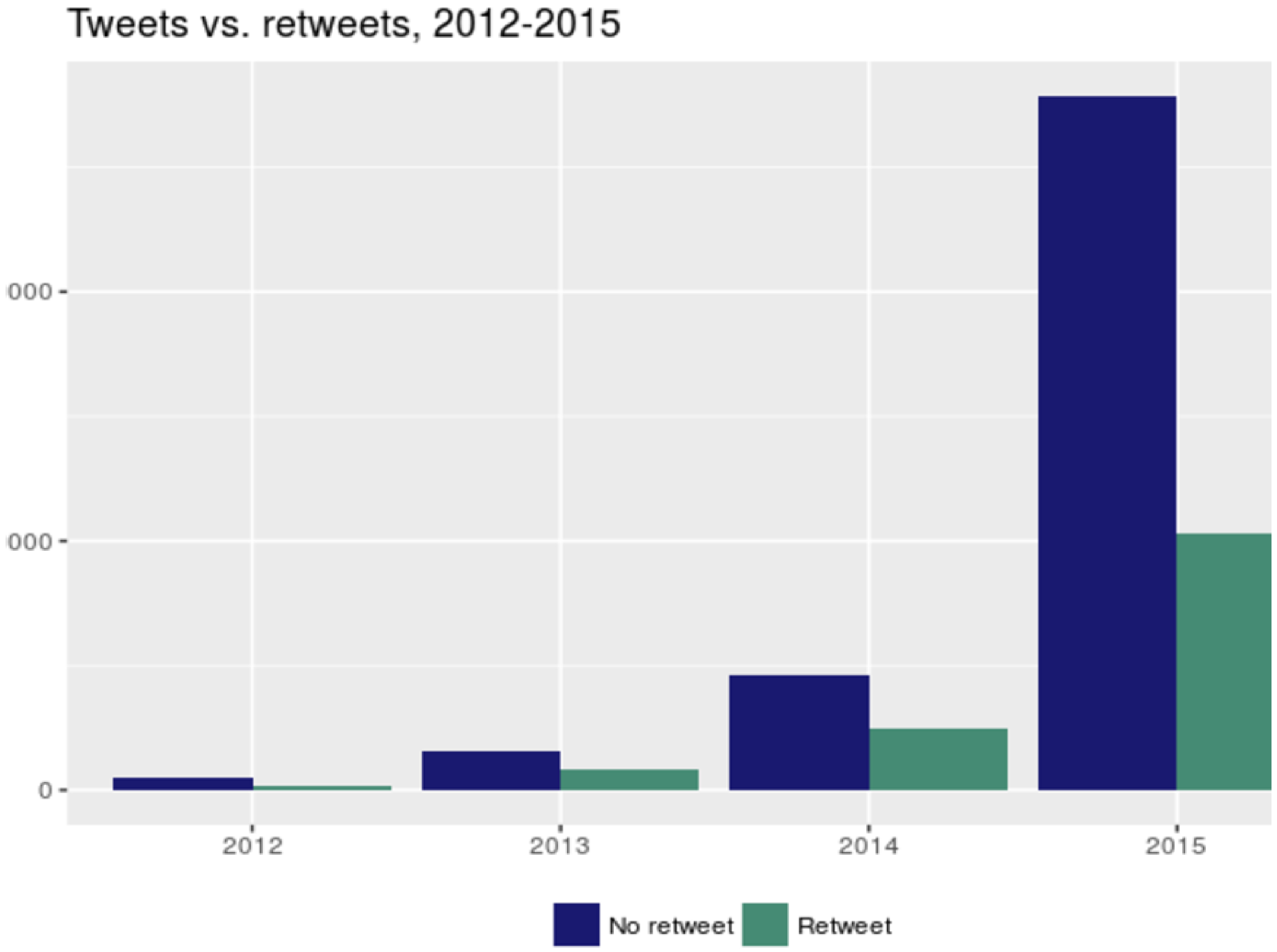
Results: Top users



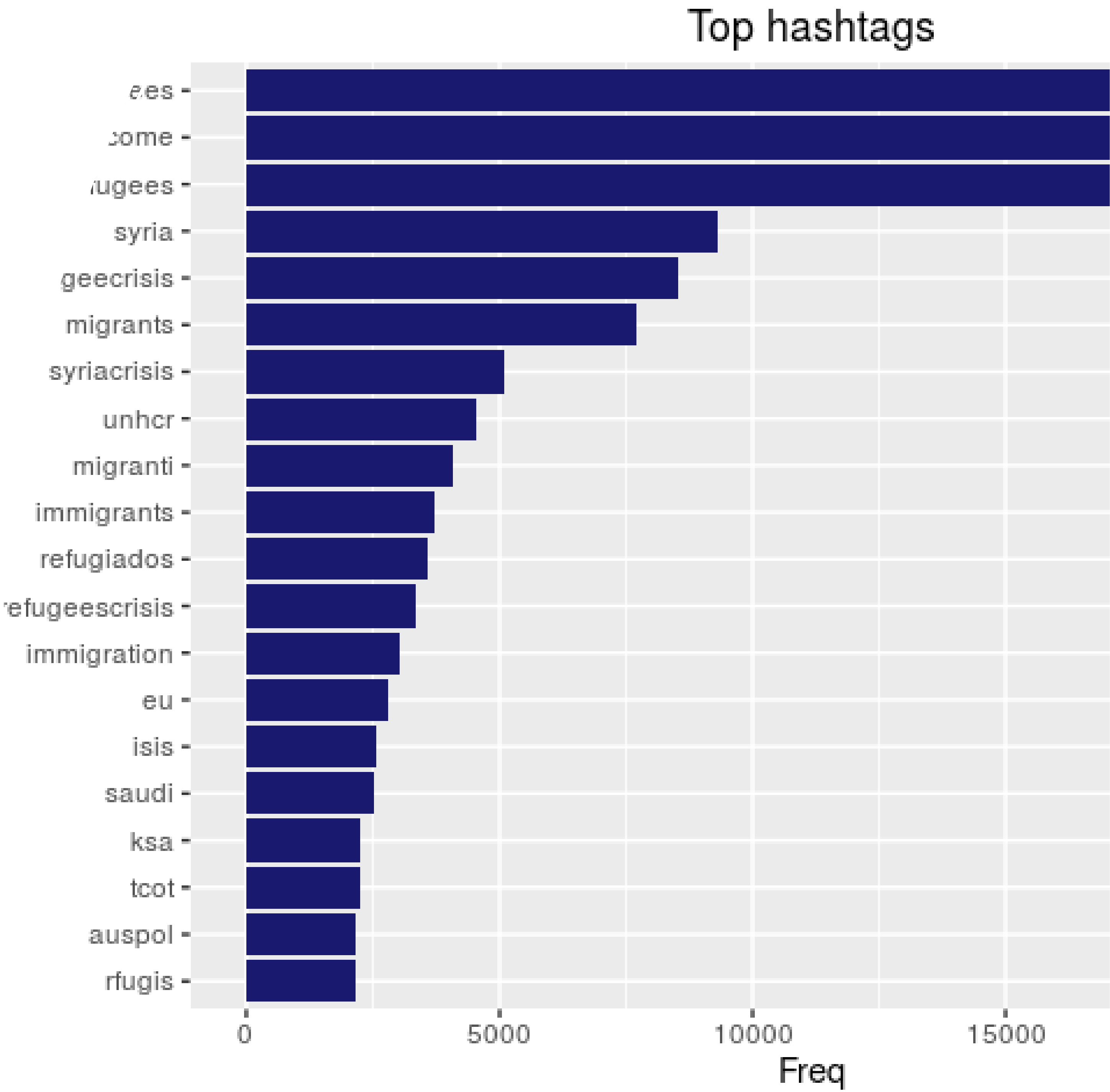
Traditional media: Newspapers

France	Germany	Italy	Spain	United Kingdom	United States	Other
@lemondefr	@SPIEGELONLINE	@SkyTG24	@muyinteresante	@BBCBreaking	@washingtonpost	@AJELive
@TF1	@ProSieben	@repubblicait	@elmundoes	@BBCWorld	@cnnbrk	@AP
@M6	@tagesschau	@Corriere	@voguespain	@Reuters	@WSJbreakingnews	@breakingnews
@FRANCE24	@zeitonline	@fattoquotidiano	@diarioas	@BBCNews	@CBSTopNews	
@canalplus	@BILD	@sole24ore	@actualidadrt	@guardian	@ABCNewsLive	
@lefigaro	@SPIEGEL_EIL	@Internationale	@el_pais	@FinancialTimes	@nytimes	
	@SZ	@ilmessaggero		@SkyNews	@WSJ	
	@sternde			@SkyNewsBreak	@FoxNews	
	@welt			@ReutersLive	@ABC	

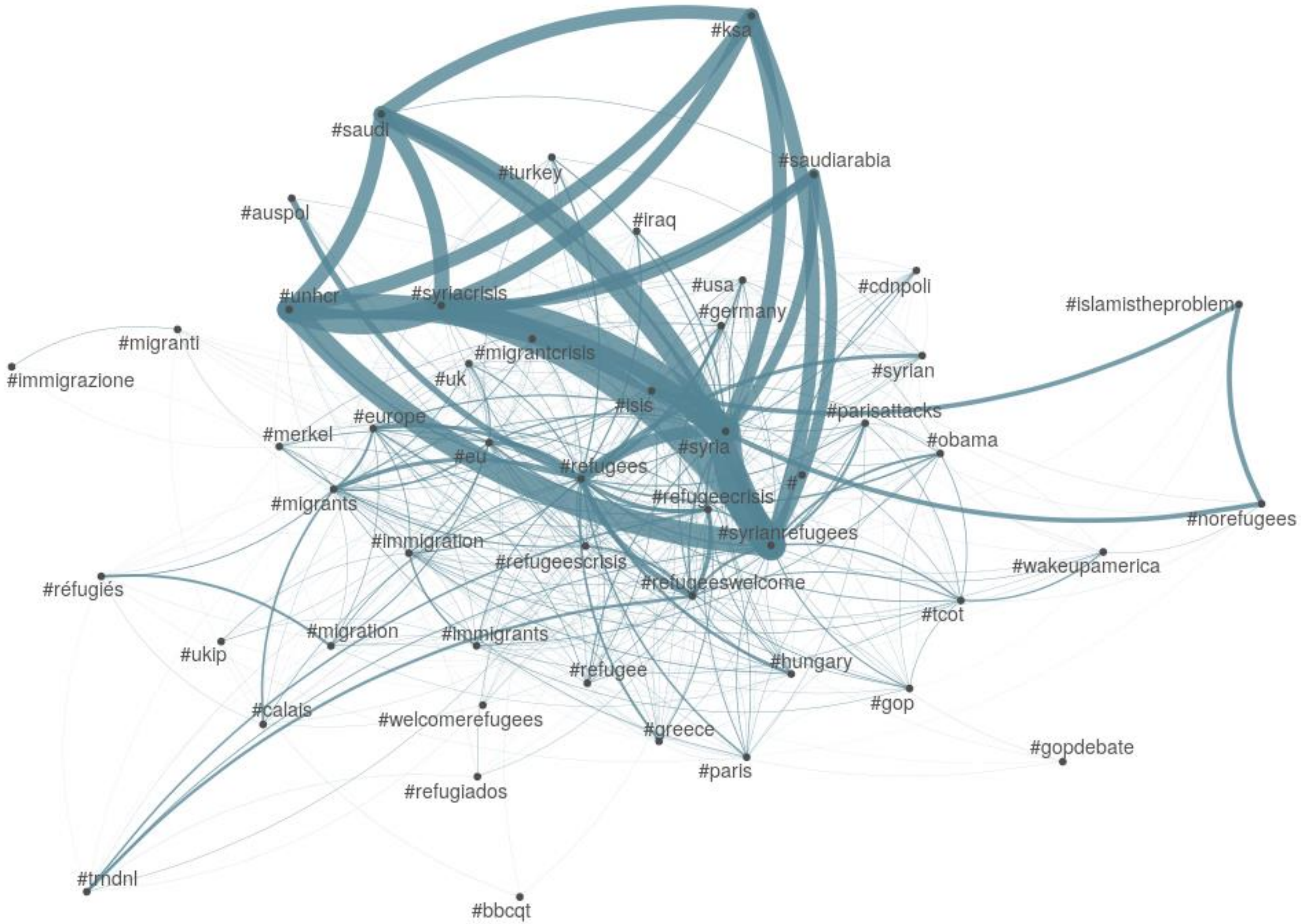
Results:
Tweets vs. retweets



Results: Top hashtags



The Hashtags Network



Results:
Sentiment and Polarization Analysis

Computer-assisted linguistic analysis

We describe the most common tools for pre-processing textual data, including stop word removal, stemming, lemmatization, compounding, decompounding, and segmentation.

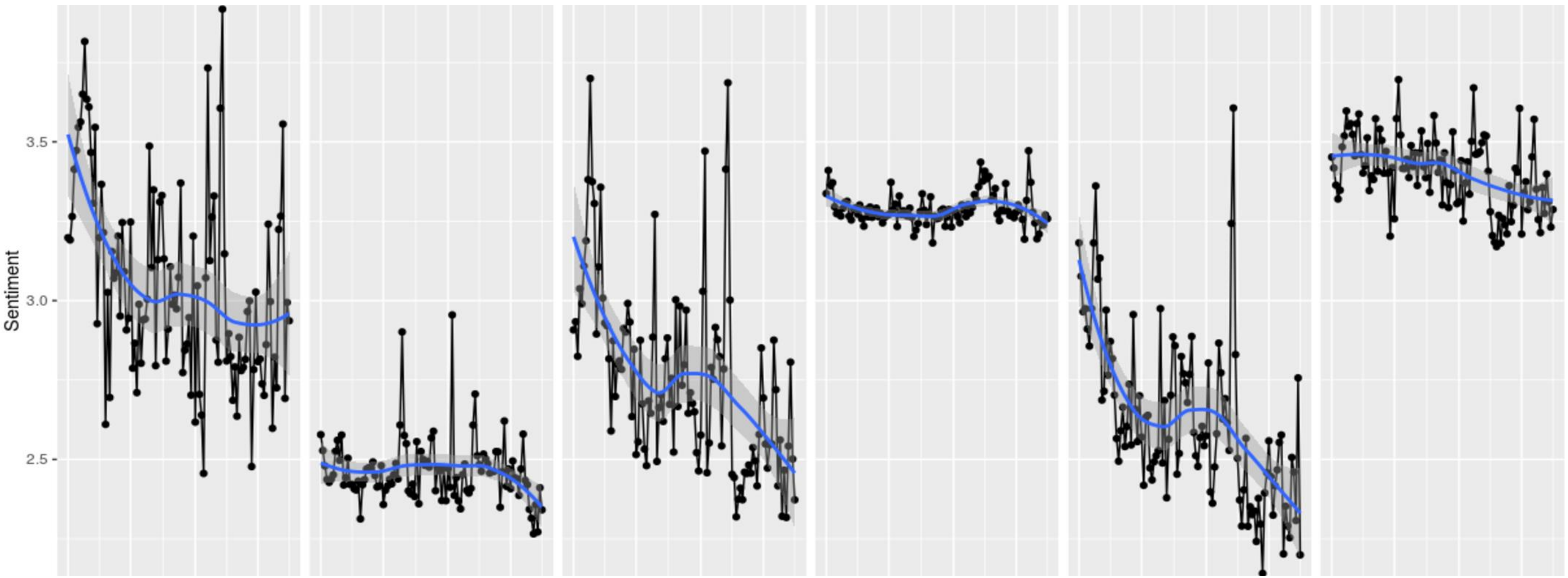
In each case, the goal is to reduce the scale of the problem by treating words with very similar properties identically and removing words that are unnecessary to our interpretation and our model.

Along with disregarding word order, the so-called “bag-of-words” assumption, these procedures are common preprocessing steps but can differ across languages.

Sentiment and Polarization Analysis: **Pre-processing Textual Data**

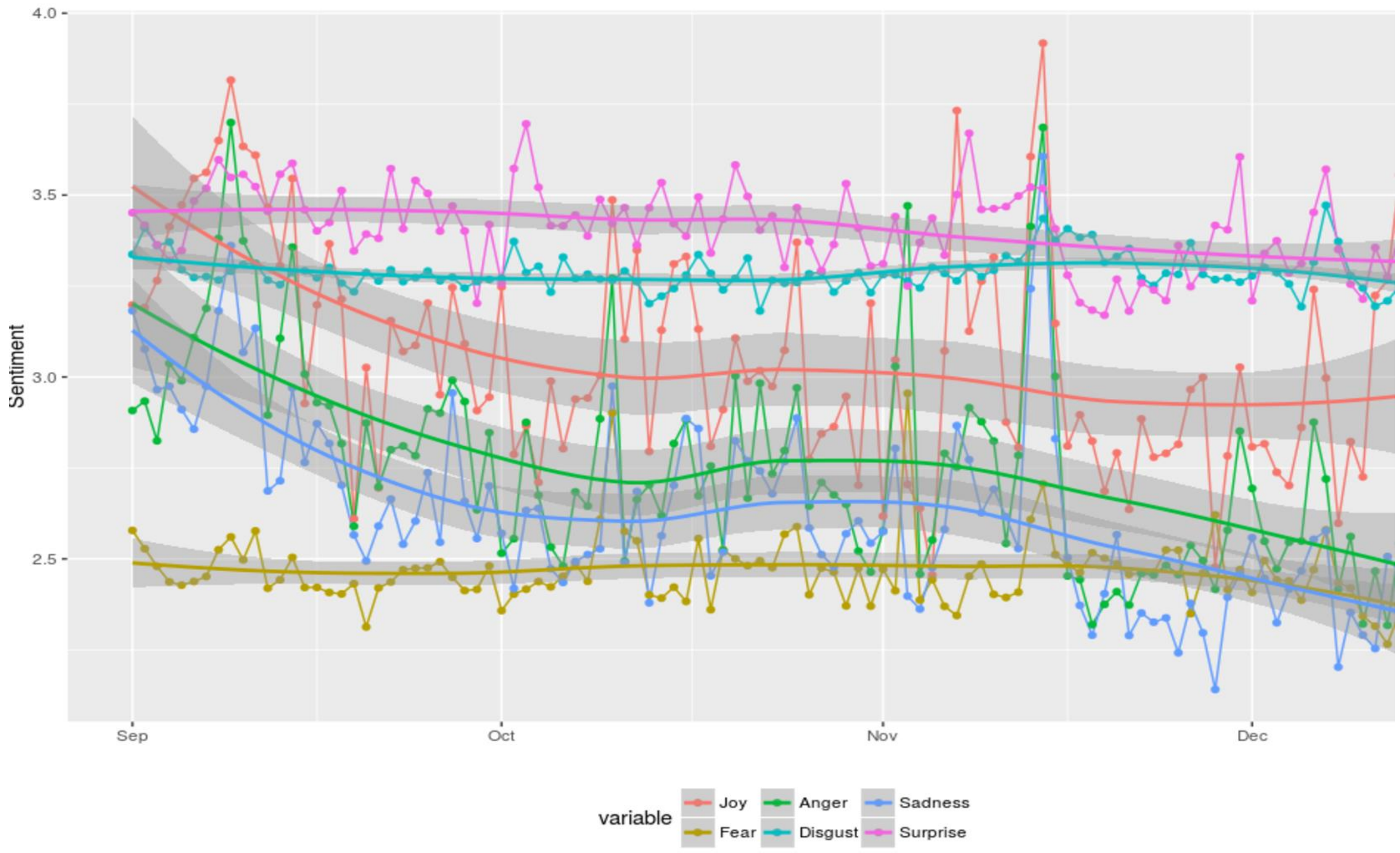
Common tools	Examples
Stop word removal	“~~and~~” and “~~the~~”
Stemming	accounts -> account accounting -> account accountants -> account
Lemmatization	The word "better" has "good“ as its lemma
(De)compounding	The German word “Kirche,” or church, can be appended to “rat,” forming “Kirchenrat,” one who is a member of the church council, or “pfleger” to form “Kirchenpfleger,” or church warden.
Segmentation	
Linguistic inquiry and word count (LIWC)	Words have specific psychological meanings (e.g., Weintraub 1989)

Sentiment and Polarization Analysis: **Mean per day**

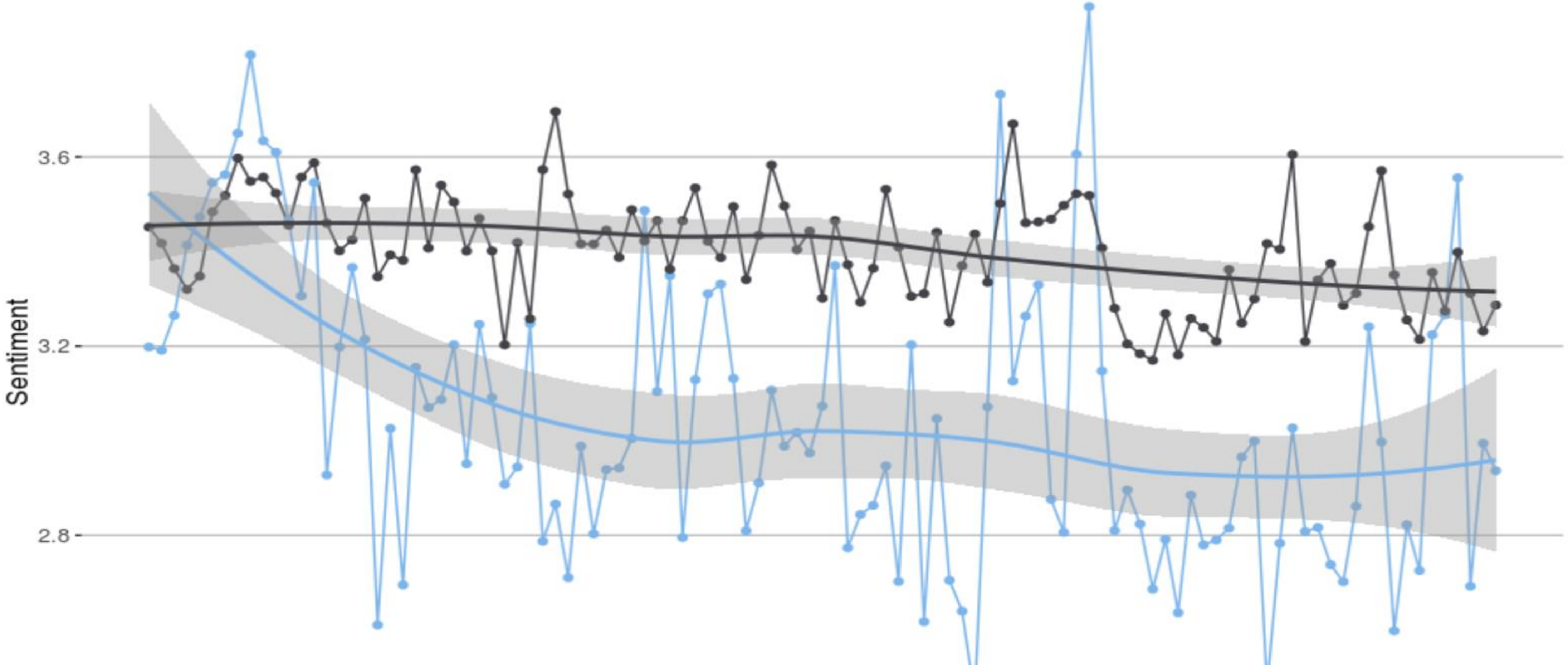


Sentiment Analysis

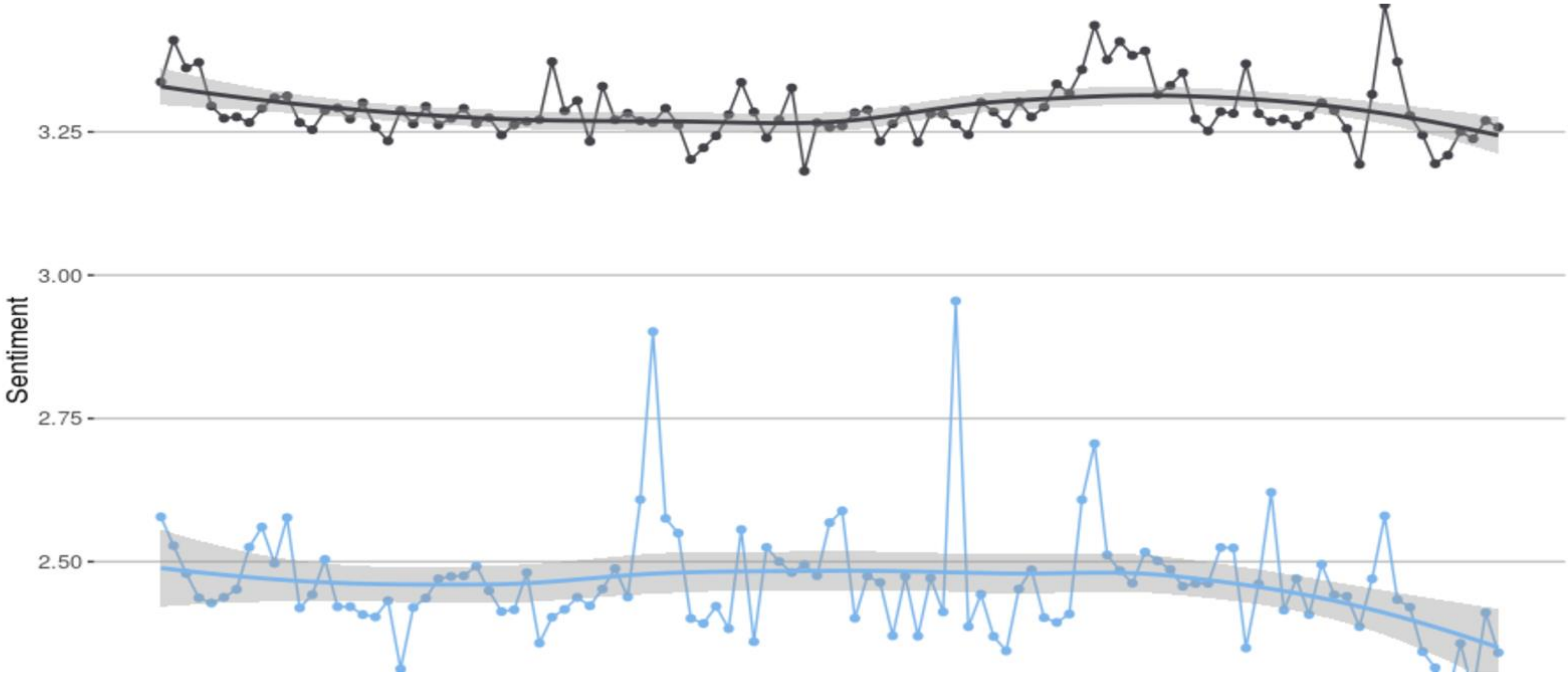
- Joy
- Fear
- Anger
- Disgust
- Sadness
- Surprise



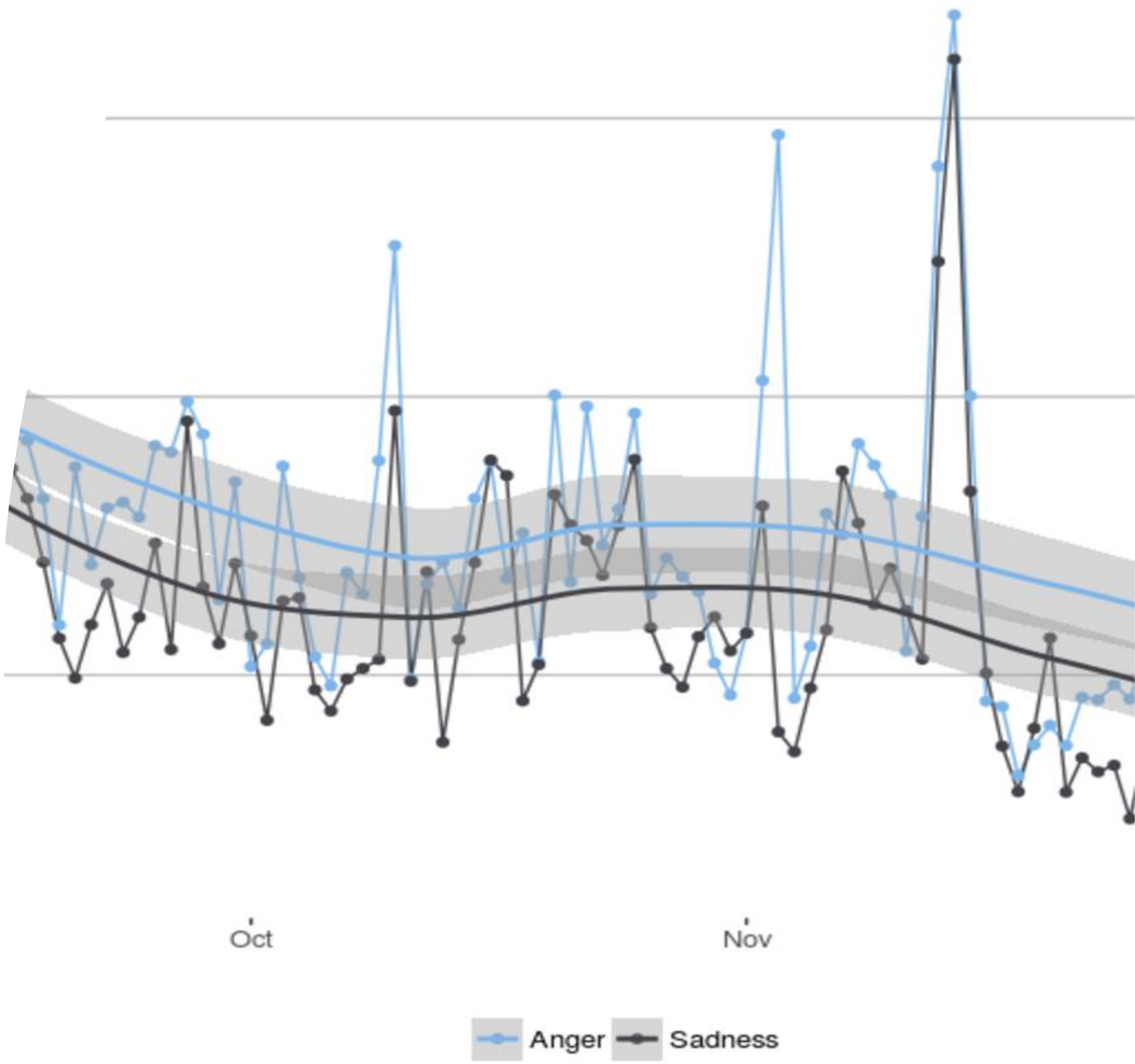
Sentiment analysis: Joy + Surprise



Sentiment analysis:
Disgust plus fear

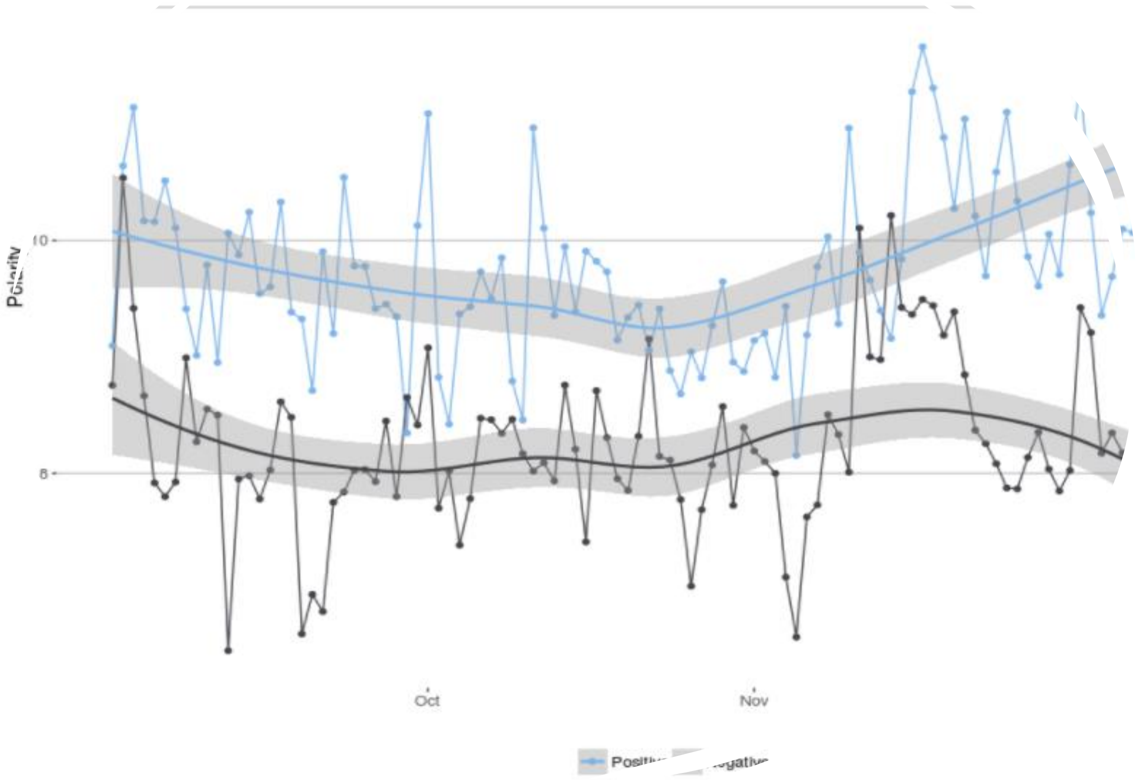


Sentiment analysis: Sadness plus fear



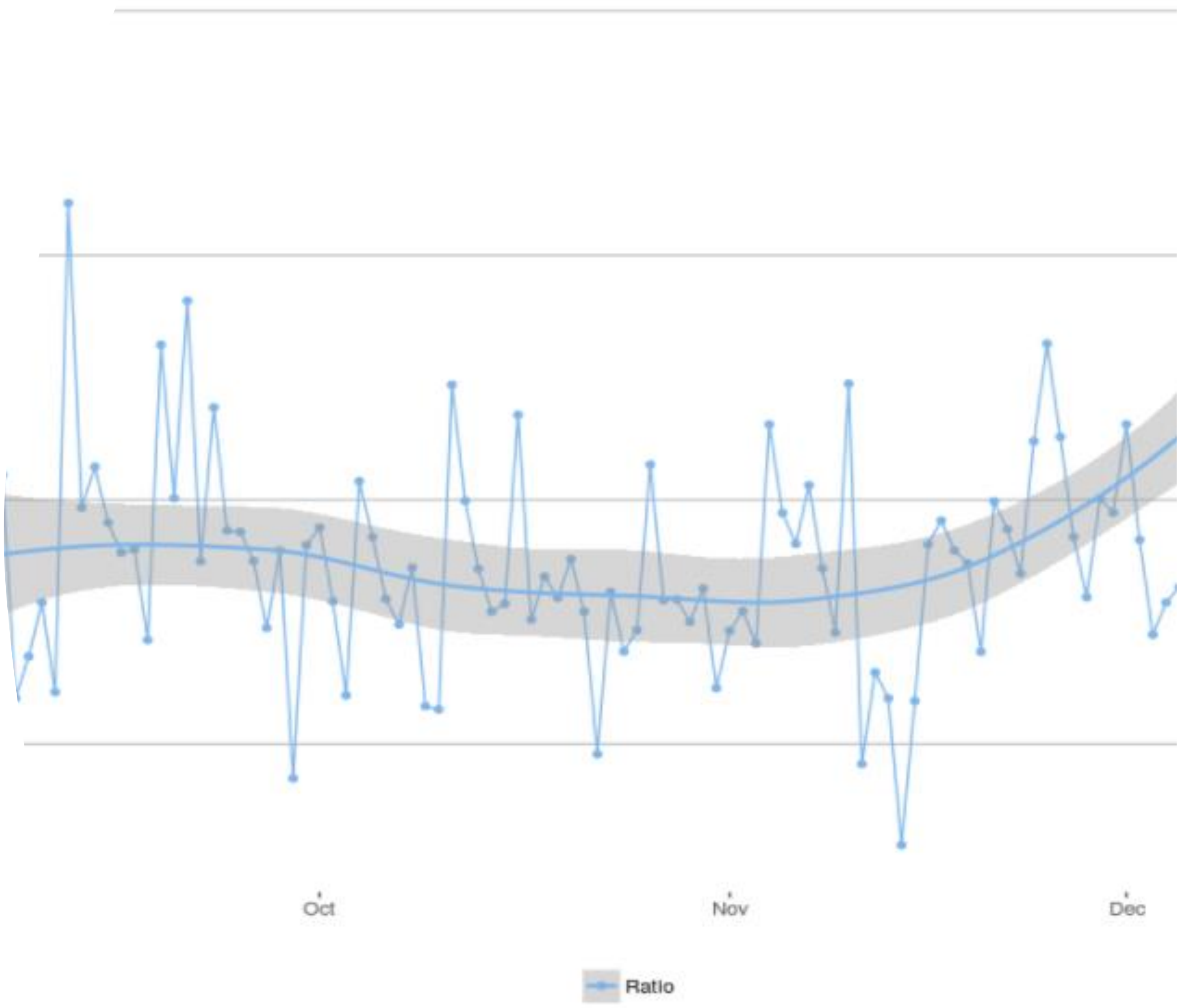
Results:
Polarity Analysis

- The second step of our sentiment detection approach is **polarity classification**, i.e., predicting positive or negative sentiment on subjective tweets.
- In this section, first we analyze the quality of the polarity labels provided by the three sources, and whether their combination has the potential to bring improvement.
- Two questions we investigate regarding these sources are: (1) how useful are these polarity labels? and (2) does combining them bring improvement in accuracy?
- We take the following aspects into consideration:
    - Labeler quality
    - Number of labels provided by the labelers
    - Labeler bias
    - Different labeler bias



Polarity Analysis: **Positive + Negative**

Polarity Analysis: **Ratio**



Conclusion

- The conversations are mostly from individuals in volume, but at the core of the network, we find traditional news media.

- The next step in the development of this paper is about creating a dataset of sentiment and polarity indices and test whether they could have some exploratory power for certain general election results.

References

- Warin Th. “European Central Bank’s Monetary Policy Decisions: A Dataset of Two Decades of Press Conferences” (with Sanger, W.), *Data in Brief*, pp. 794-798, 2018, [DOI: 10.1016/j.dib.2018.08.061] [Data]
- Sanger W., de Marcellis-Warin N., Warin Th. "Text-as-Data Analysis of Political Parties versus Government Parties: To Blend or not to Blend? The Appendix" DOI: 10.6084/m9.figshare.7781051.v2, pp.1-63, February 2019.
- Sanger W., Warin Th. “Jaccard Similarity of 1517 European Political Manifestos across 27 Countries (1945-2017)” *Data in Brief*, DIS-S-18-02150. 2019 [DOI: 10.1016/j.dib.2019.103907]
- Sanger W., Warin Th. “How Data Science can (also) help central bankers: An NLP study of the European Central Bank presidents’ speeches” *Global Economy Journal*, Vol. 20, No. 02, 2050009, 2020 [DOI:10.1142/S2194565920500098]