

Measuring Political Mobilization: Insights from Massive Machine-Coded Datasets

Andrew Halterman
Caerus Associates
Washington, DC

Jill Irvine
University of Oklahoma
Norman, OK

1 Introduction

This paper explores the ways in which event data extracted by computers from news articles can be used to investigate broad patterns of political mobilization and government responses. It seeks to gain greater comparative understanding of the scope and focus of political activism and government responses during two waves of social and political mobilization in Europe: the color revolutions (electoral breakthroughs) in postcommunist countries in the 1990s and 2000s, and the more recent waves of strikes and protests in response to the economic crisis in the European Union. We hope to illustrate how massive machine-coded event data can answer questions that would be infeasible with traditional methods and can work alongside in-depth, qualitative studies by area specialists. To do so, we rely on two new machine coded datasets, the Global Dataset of Events, Language, and Tone and the associated Global Knowledge Graph.

The Global Database of Events, Language, and Tone (GDEL) is a relatively new machine-coded event dataset of geo-referenced political and social activity drawn from English language news reporting around the world from 1979 to present (Leetaru and Schrod, 2013). Updated daily, it includes around 250 million events coded in the CAMEO taxonomy (Gerner et al., 2002) with the TABARI engine (Schrod, 2011), with the form “*Source Actor does Action to Target Actor.*” GDEL also includes an experimental release of the “Global Knowledge Graph” (GKG), a companion dataset that tracks the relationships (via co-mentions in news articles) of organizations, people, themes, counts, and locations around the world daily and the emotional tone of their coverage (Leetaru, 2013). While we use GDEL for this paper, our approaches could be implemented with some modifications in a similar dataset (including the U.S. government ICEWS dataset and other datasets coded with TABARI or its successor, PETRARCH).

We wished to explore GDEL’s utility in tracing and measuring civil society influence, political mobilization, and government responses. Political mobilization in Europe provides good test cases for exploring event datasets such as GDEL and GKG because the data collected are relatively complete and consistent across countries in this region. Moreover, by using this approach to investigate a wave of mobilization during the color revolutions that has been extensively researched, we have a “gold standard” or “ground truth” to use in assessing the overall reliability of the data. We can then apply it to gain a better understanding of more recent events. Using this data, we ask three main questions:

1. What is the extent and timing of political mobilization during these two periods of political unrest?

Paper presented at the International Studies Association 55th Annual Convention, Toronto, Canada, March 26-29, 2014.

This paper and replication code are available at <http://github.com/ahalterman/ISA-2014>

Thank you to John Beiler for helpful comments.

Version 1.0

2. What are the themes and potential frames associated with political mobilization?
3. What are the patterns of government responses to political mobilization?

We begin by using GDELT to measure civil society activities before, during, and after the electoral breakthroughs in postcommunist Europe. We next examine the relationship of civil society activities to USAID democracy and governance funding. USAID spent billions of dollars attempting to strengthen civil society as a means to encourage electoral breakthroughs. We argue that while this form of event data can tell us little about the impact of aid, it can provide a useful tool for assessing levels of aid in relation to levels of civil society activity over time. We then analyze political mobilization and government responses, tracing patterns of cooperation/confrontation and accommodation/repression. Here GDELT proves useful, we argue, in allowing us to identify broad patterns of popular mobilization and regime response that characterized the different dynamics and trajectories of electoral revolutions. It may also help generate hypotheses about the impact of repression on mobilization more broadly. The final section of the paper, drawing upon GKG, focuses on themes and issues associated with strikes, protests, and social movements in EU countries in 2013. We highlight clusters of issues and levels of violence associated with these events, paying particular attention to the association of extreme right themes with protests. The Global Knowledge Graph provides a useful tool, we argue, for gauging the resonance of extreme right appeals with particular populations.

While some causal and associational work has been done with GDELT (including Gallop and Ward (2013) and ongoing work by Morgan and Reiter at Emory University on road construction in response to insurgency in India), most work has been in conflict forecasting applications (Arva et al., 2013; Brandt et al., 2013; Yonamine, 2013), as has work using event data more broadly (Ward et al., 2013b; O'Brien, 2010). We also build on previous work on using newspaper articles to study protest events (see Earl et al. (2004) for a useful overview). Earl et al. (2004) identify a number of advantages to machine-coded data on protests, including complete consistency between re-codings and stable coding schema over time. They recommend including multiple news sources to balance out potential biases, which GDELT meets with its with hundreds of sources (via Google News).

Methodological Challenges of Working with GDELT and GKG¹

GDELT and GKG's size, scope, and temporal coverage create a number of difficulties, especially for cross-national comparison. Essentially, each country has its own data generating process, which is affected by the amount of English language coverage in each country, propensities for the international media to cover topics differently in different countries (the media is very aware of protests in Egypt now), the geographic resolution of news articles (do they say "Syria," "Aleppo," or "Sheik Maqsood"?), the quality of the geographic gazetteer relating place names to coordinates, and the vast and uneven increase over the past three decades of articles available online. These variations make comparative inference very difficult and are common to all global machine-coded datasets. GDELT's faces more difficulties than other event datasets as a result of its philosophy of including as much information as possible, rather than aggressively filtering articles as they are ingested (see Ward et al. (2013a) for a discussion of GDELT vs. ICEWS).

Two main approaches have emerged for reducing GDELT's variation in coverage by country and by time: scaling all events of interest as a percentage of all events in the country (Brandt et al., 2013) and building multilevel models that generate estimates for the number of a type of event, conditioned on the year, the country, and the type of event (D'Orazio and Masad, 2013). The residuals from the model indicate whether an event-type of interest, say protests, is higher in a given month than would be expected. While total levels of events cannot be compared across cases with this method, increases or decreases are easily identifiable. In other cases (including the ones we discuss below) the monthly totals of events appear stationary over time.

¹In January 2014, serious concerns were raised about the permissions to the news articles used in generating the historical backfiles for GDELT. We hope that the ongoing legal dispute around GDELT's sources will be resolved soon. We believe that the techniques we employ will transfer easily to any other similar dataset and look forward to a greater diversity of datasets in the future.

While GDELT is in a well-understood event data format, the Global Knowledge Graph includes aggregated “namespaces”: unique combinations of actors, themes, and locations occurring in a day’s news, presented in a network structure. GKG aims to contextualize news events and to allow social network analysis of how actors, locations, and themes are linked in the news. From a purely technical standpoint, GDELT is easily stored in tabular form in a relational database, whereas GKG, with its wildly varying number of fields for each event, is not.

GKG’s components include counts of people involved in various events, “themes,” article sentiment scores, and lists of organizations, people, and locations that are mentioned together in each set of articles. These components vary in how transparently they are assigned and how easily they are interpreted. Themes are perhaps the most transparently generated: they are each triggered by the presence of keywords in an article (for example, if the term “collective bargaining” appears in an article, GKG will attach the “unions” theme).² Themes are also one of the more useful components, adding easily interpreted data about events that cannot be captured in a simple Actor 1–Event–Actor 2 form.

Organizations, people, and locations are added to GKG through a much more opaque process and are less useful than themes both because of high levels of miscoding and the black box nature of the generation process. The (unknown) algorithm for extracting organizations and people misses many organizations and people, and misclassifies some people as organizations and vice versa. Location co-mentions, while perhaps useful in some cases for network analysis, have been less useful for us than simple geo-located event data. Counts, while theoretically able to convey very important information like the number of people at a protest or killed in a fight, seems to pick up too many historical cases and are also not transparently determined. Finally, article tone is the least transparent and, in our experience, the least useful. For instance, does negative sentiment expressed in domestic (state owned) Egyptian media in articles about women’s rights indicate condemnation of violence against women, or opposition to women’s activists? Analysis of sentiment and tone, while clearly an important field for computational social science, need to be done at a more precise level. We return to the issues of transparency and interpretability in the conclusion.

2 Measuring Civil/Political Society Activity With Event Data

We begin our attempt to broaden GDELT’s application to a wider set of questions concerning social and political mobilization by exploring its utility in measuring civil society. Our primary aim is to assess GDELT’s usefulness in exploring the impact or influence of civil society and its possible relationship to other factors such as levels of external funding. The American government spent millions of dollars in aid to strengthen civil society in postcommunist countries, and this civil society funding figured prominently in strategies to promote electoral revolutions. The impact of this aid in these and other cases of nation building has been extremely difficult to evaluate, in part due to the lack of clarity about what constitutes civil society. Academics and practitioners have spawned a vast literature on civil society, attempting to gain greater conceptual, analytical and operational clarity. Most definitions of civil society focus on its characteristic as “an arena outside the family, the state, and the market where people associate to advance common interests.” (Heinrich, 2004) While in practice this has usually meant focusing on civil society actors, scholars such as Edwards have emphasized the context or space in which their activity takes place (Edwards, 2009). Similarly, while some approaches to analyzing and measuring civil society focus on its contributions to the economy and service sector, others emphasize its essentially political character (Malena and Heinrich, 2007). Using a single dataset of political and social activity enables systemic comparative studies across multiple countries in different years.

Two large studies have produced indexes that attempt to capture the essential dimensions of civil society, the Civil Society Index and the Global Civil Society Index. These comprehensive measures involve multiple dimensions and dozens of indicators about such characteristics of civil society as structure, environment,

²The complete list of themes and corresponding keywords is available from <https://github.com/ahalterman/ISA-2014>.

culture, and values. Both include an impact dimension, which attempts to measure the influence of civil society actors and activity on broader political and policy making processes. The CSI employs the broadest measure of impact, measuring level of influence on public policy, responsiveness to social needs and empowerment of citizens (Malena and Heinrich, 2007). Neither of these indexes captures, however, the important aspects of what civil society actors do. What kinds of behavior are they engaged in and when? Focusing on actions, rather than organizational structure, allows us to better measure political mobilization. Moreover, these indices are compiled annually, while event data can reflect day-to-day changes (although we aggregate to the month in most cases).

GDELT provides just such a measure by capturing the number of civil society actions and the levels of political activity in a country. The data set provides numbers of mentions of civil society actors, movements, opposition parties, media, and labor, in the local and international press. To measure civil society and political mobilization using GDELT, we count domestically occurring events for each country of interest where the source actors (Actor 1) match one of the following codes: “civilian” (CVL), “opposition” (OPP), “labor” (LAB), “media” (MED), “NGO” (NGO), or “human rights organizations” (HRI). Filtering the data using only actor codes is very effective at capturing the events we are interested in analyzing. We explored whether also including events that fall into relevant-seeming event categories (e.g. “appeal for change in leadership”) would allow us to pick up more events, but it does not seem to do so (and results in many additional false positives). There is considerable overlap in this approach to measuring civil society activity and measuring political mobilization, understood as “the deliberate activity of an individual or group of individuals for the realization of political objectives” (Political Mobilization of Minorities and Anti-Diversity Discourse in the Context of the European Crisis of Multiculturalism).

The results in Figure 1 show varying patterns of civil society activity across a number of postcommunist cases. Patterns of high activity correspond roughly to what we know about levels and timing of political mobilization and citizen-led opposition in particular countries (Bunce and Wolchik, 2011; Lane, 2009). Thus, we see spikes in activity, particularly in those cases of electoral breakthrough, like Serbia and Ukraine, in which citizens “took to the streets” in protest of regime actions and on behalf of democratic election outcomes (See Figure 1). In two cases, Albania and Macedonia, where electoral revolutions did not take place, our measurement of civil society activity corresponds to high levels of political mobilization in response to government corruption, in the Albanian case, and ethnic demands, in the Macedonian case. Thus the data appear to have a fairly high degree of face validity in terms of measuring the extent and timing of civil society activity in the political sphere in the twenty postcommunist cases we examined here.

An important question we tackle next is the potential impact of external aid on civil society activity. The US government spent millions of dollars promoting democratization in Eastern Europe. The funding model for promoting electoral breakthroughs was developed in the early cases of Romania and Bulgaria and applied in roughly similar fashion to the later breakthroughs in Croatia, Serbia, Georgia, Ukraine, Kyrgyzstan, and Azerbaijan (Carothers, 2004; Kalandadze and Orenstein, 2009). Scholars have debated the impact of this funding, with some stressing its essential role in building a viable opposition (Bunce and Wolchik, 2011) and others finding a potentially demobilizing effect (Gagnon, 2006). We do not believe that we can use GDELT to make a strong causal claim about foreign aid. But by pairing civil society activity with the level of USAID funding, it may allow us to discern broad patterns in the relationship between the two.

Several observations can be made from Figure 1 concerning the relationship between US government aid and civil society activity. In calculating levels of aid, we use data from Finkel et al. (2007) on USAID’s democracy and governance funding, combined with data from Lawson (2009) on money spent by USAID’s Office of Transition Initiatives (OTI) for countries where it is missing from Finkel’s data (specifically, Kosovo). Combining monthly data on civil society activity in the political sphere from GDELT with datasets like the Finkel data on USAID democracy and governance (D&G) funding allows us to quickly and visually assess the relationship of different interventions and civil society activity.

When it comes to general levels of aid, all countries in the region received democracy and governance assistance during this period after the collapse of state socialism. Countries that experienced successful electoral

breakthroughs do not seem systematically different from non-breakthrough countries in the amount of funding they received. Rather, the highest levels of spending are in the strategically important countries of Serbia, Bosnia, and Ukraine. Serbia and Bosnia received particularly high levels of spending due to American and NATO military presence in the area and the related imperative to stabilize the region politically. Moreover, while the relationship between USAID funding and levels of civil society activity vary, in most cases of electoral breakthroughs funding increases significantly after the breakthrough rather than before. The exceptions to this pattern are Azerbaijan and Ukraine, where funding levels climb before the electoral breakthroughs. Finally, in the two cases of Croatia and Romania, funding and levels of mobilization remain steady through the periods before, during, and after the electoral revolutions. Thus, we cannot discern a broad pattern of significantly increased funding preceding, and therefore potentially contributing to, significant increases in civil society activity.

A more interesting use for GDELT than tallying civil society events is to identify and trace patterns of citizens' mobilization and government/regime response. Aid programs were more or less the same in the electoral breakthrough countries. The US government aid focused on strengthening civil society, political parties, the media and the judiciary, and forging links among them. In short, aid was directed toward creating and bolstering opposition to the semi-authoritarian regimes. Yet, while the funding approach was similar, local citizens adopted different strategies of opposition based on location conditions.. What different patterns of citizen-led opposition and regime responses in the electoral breakthroughs in post communist countries can we discern, and what might this tell us about the impact of governmental repression or accommodation more broadly?

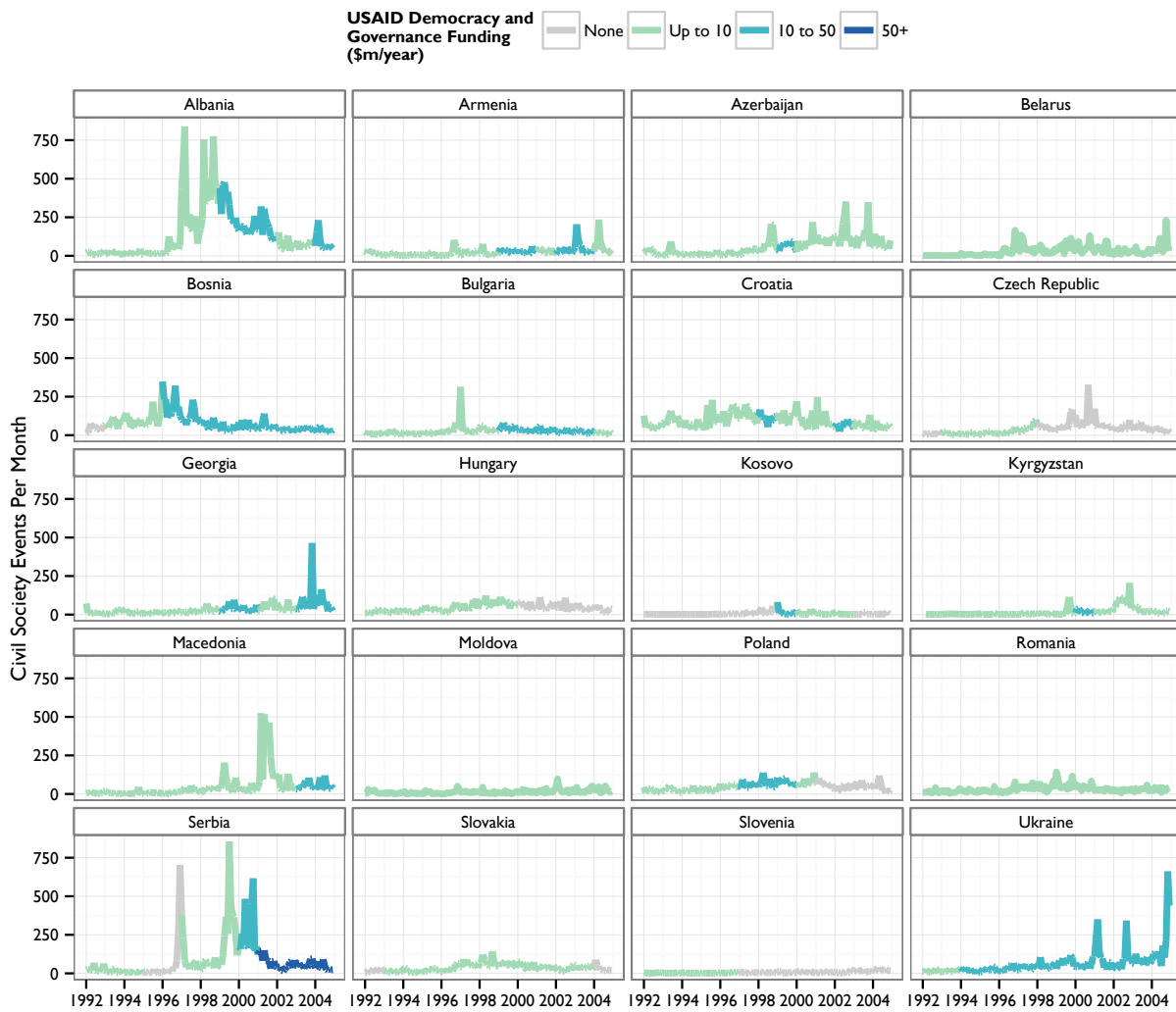


Figure 1: Civil Society Activity and Democracy and Governance Funding In Eastern Europe

GDELT can give us disaggregated information about what types of actions civil society actors are directing at government actors, and how governments act toward civil society. GDELT’s CAMEO action codes can be aggregated into four broad categories (“QuadClass”): verbal cooperation, material cooperation, verbal conflict, and material conflict (See Table 1). Disaggregating by `QuadClass` reveals the changing relations between government and civil society and the strategy of each across the eight cases of electoral breakthroughs (See Figure 2). In the case of Serbia, for example, government “material conflict events” directed at civil society increased in 1999, indicating repressive measures directed by the government against civil society. At the same time, civil society actors engaged in high levels of verbal conflict with the government, but almost no material conflict (CAMEO classifies protests as verbal, not material, conflicts). This is very much in keeping with civil society’s oppositional/confrontational strategy and with the high level of repression employed by the Serbian government. After Milošević’s extradition in June 2001, events involving civil society, either as the source or target actor, fell off sharply. The Serbian opposition’s confrontational strategy of political action “in the streets” against a highly repressive regime contrasts with the pattern of regime/opposition relations in Croatia (Irvine, 2013). In the Croatian case, the citizen-led opposition pursued a strategy “in the government” against a regime that engaged in much lower levels of direct repression. In other words, GDELT can tell us something about political opportunity structures as well as state-society relations during the electoral breakthroughs.

While we do not have the space here to examine closely the remaining cases of electoral breakthroughs as highlighted by the event data here, we can confirm that the patterns of mobilization and repression follow the general developments established in the secondary literature. Thus in Serbia, Kyrgyzstan, and Azerbaijan, the government adopted a repressive strategy at various points in relation to citizen protests, whereas the Croatian, Bulgarian, and Romanian cases were relatively free of such repressive responses. The opposition’s strategy in Ukraine, Georgia, and Kyrgyzstan seem to be mostly in the streets, while Bulgaria’s opposition seems much more verbally engaged with the government. In these cases, it appears that the electoral revolutions resulted in a higher level of verbal cooperation between government and citizens, whereas in Kyrgyzstan and Azerbaijan, very low levels of verbal cooperation existed in the aftermath of citizen protests. Event data with source and target actors allows for careful tracing of government and opposition strategies, week by week, without hindsight bias. Is a higher level of verbal cooperation after protests associated with better outcomes for politically mobilized citizens pressing for democratic changes? When levels of material conflict are higher from civil society to government than the other way around, is breakthrough and reform the likely outcome? Drawing upon a large number of cases using GDELT might allow us to answer such questions.

Table 1: CAMEO Code to Quad Classification Mapping

Verbal Cooperation (1)	Material Cooperation (2)
Make Public Statement	Engage in Diplomatic Cooperation
Appeal	Engage in Material Cooperation
Express Intent To Cooperate	Provide Aid
Consult	Yield
Engage in Diplomatic Cooperation	Investigate
Verbal Conflict (3)	Material Conflict (4)
Demand	Exhibit Force Posture
Disapprove	Reduce Relations
Reject	Coerce
Threaten	Assault
Protest	Fight
	Use Unconventional Mass Violence

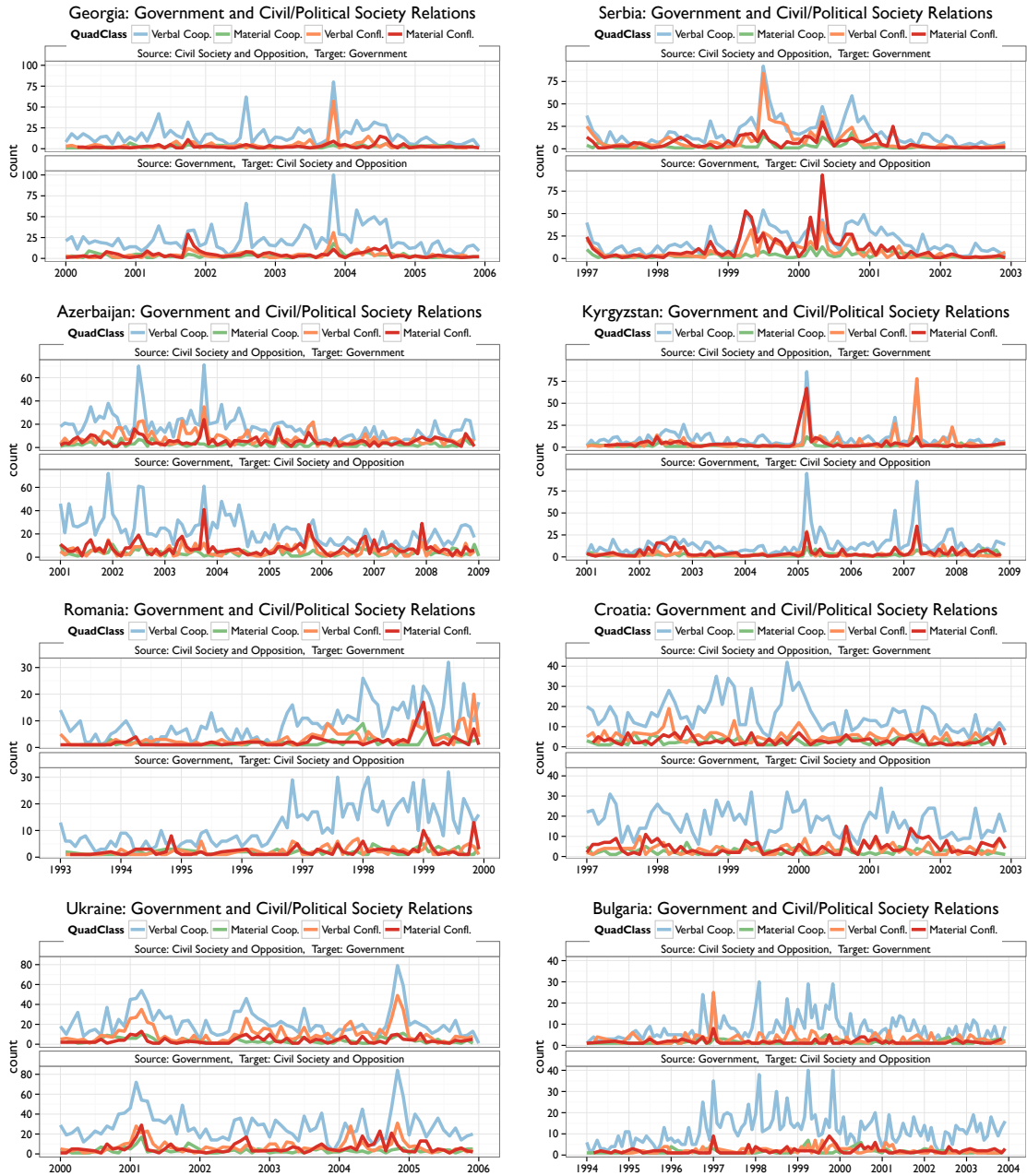


Figure 2: Government to Civil Society/Opposition Events, By QuadClass

3 Mapping Mobilizing Themes with “Theme” Data

In this second part of our study, we explore the utility of the Global Knowledge Graph for expanding our understanding of political mobilization. The core component of GDELT is the stream of “source actor–action–target actor” events going back to 1979. GDELT also includes an experimental dataset called the “Global Knowledge Graph,” which records when organizations, people, themes, counts, and actors involved in events with global media coverage are mentioned together in news articles. While GKG is currently experimental and exists only back to April 2013 at present, it allows for more contextualization of the events that appear in GDELT and allows us to make comparisons between countries and over time about which themes are linked to events of interest.

In this part of the paper we focus on protests, strikes, and general social movement activity in various EU countries during the summer and fall of 2013 in response to the on-going economic crisis in Europe.³ While commentators have tried to make some generalizations about what issues the protesters are concerned about, there has not been a systematic way of comparing the media discussion around all protests in Europe. GKG includes a “themes” column, attaching tags to events based on keywords used in the article. This allows us to attach themes to protests, strikes, and general movement activity across all countries, in this case our subset of EU members, during this period.

GKG includes around 150 themes, which range from the very broad (“General Government”) to the very specific (“Locusts”). In tracking themes with mobilizing events in Europe during this period of time, we decided to focus on a cluster of themes that might be associated with economic concerns and extreme right political movements. Concern has grown about a potential rise in extreme right appeals in response to economic hardship and the ability of extreme right political parties to make significant gains at the polls (Economist, 2013; Bartlett et al., 2012). The literature on extreme right movements and parties has long considered economic hardship and crisis to be associated with support for far right extremism, though this theory of right-wing activism has also been contested (Mudde, 2013). To what extent has political mobilization in Europe over the past year been associated with extreme right issues and appeals? By looking at the percentage of extreme right themes associated with protests, strikes, and general social movement activity in Europe, we can begin to explore GKG’s usefulness as a tool in answering this and related questions.

We begin by examining the overall level of political mobilization (as manifested in strikes, protests, and events involving social or political movements) across the 28 EU countries (see Figures 3 & 4). Note that in the overall count, the largest four countries by event numbers (Germany, France, UK, and Italy) are displayed on a separate scale so as not to obscure the counts in smaller or less extensively covered countries. As we can see, the United Kingdom has the highest overall level of political mobilization as measured by protests, strikes, and other activity linked to social movements, followed by France, Germany, and Italy. The usual GDELT caveats apply to comparing numbers of protests across countries: the UK may indeed have greater number of protests than Greece, but it is perhaps more likely that the international English-language news media cover protests in the UK in greater detail and with a greater number of specific person and place names, inflating the number of observed events. In any case, this result is not surprising, since these countries also have among the highest levels of population in Europe. When it comes to violent unrest, the UK also displays the highest levels of violence, followed by France, Ireland, and Germany.

Measuring per capita levels of mobilization can provide a better relative picture of overall levels of mobilization, which we can see in Figure 5. Taking population size into account, Belgium, Bulgaria, Greece, Ireland, Lithuania, the Netherlands, and Sweden have the highest level of protests, with Ireland a clear first. These countries also have relatively high levels of strikes, general social movement activity, and violence. Overall,

³Because GKG is not in event data format, we needed to reverse-engineer event counts from GKG’s namespaces. To do so, we treated each mention of a separate location as a unique event. For example, a protest with the locations “Tahrir Square” and “Cairo (general)” would count as two separate protests. We believe that this will somewhat over count events, but is the best approach for counting events in GKG’s data. Ideally, of course, we would be able to use event data with themes.



Figure 3: Political Mobilization in Europe (Counts)

Belgium, Netherlands, Ireland, and Spain have the highest numbers of per capita political mobilization.

While GKG can give us a snapshot view of how reported political mobilization in Europe manifests itself, a better use of this dataset is to explore the themes with which it is associated. We examine two themes that are associated with the economic crisis: general economic themes and unemployment⁴. We also identify three themes that could be associated with extreme right views and appeals: immigration, Muslim, and LGBT issues. Anti-immigration themes are often linked with extreme right views and have long been a central position of the major extreme right parties in Europe. Anti-Muslim sentiment is also central to extreme right rhetoric and views in many European countries. While GKG’s “Muslim” tag includes all discussion of “Muslim,” “Islam,” etc., we believe that in the context of protests and strikes, Muslim codes will be picking up largely anti-Muslim sentiment. Similarly, we assume that immigration theme codes (when linked with protests as they are here) will be picking up anti-immigrant sentiment. Finally, we include LGBT issues because high levels of extreme right violence and protest have been directed at gay rights activists and pride parades, particularly in the newest, postcommunist members of the EU. This theme is difficult to interpret because GKG could be picking up pro-gay rights themes as well as anti-gay rights themes. Nevertheless, it may provide a useful measure of the extent of political contestation over extending full citizenship rights to

⁴See the online appendix for a complete listing of all search terms used for each theme tag.

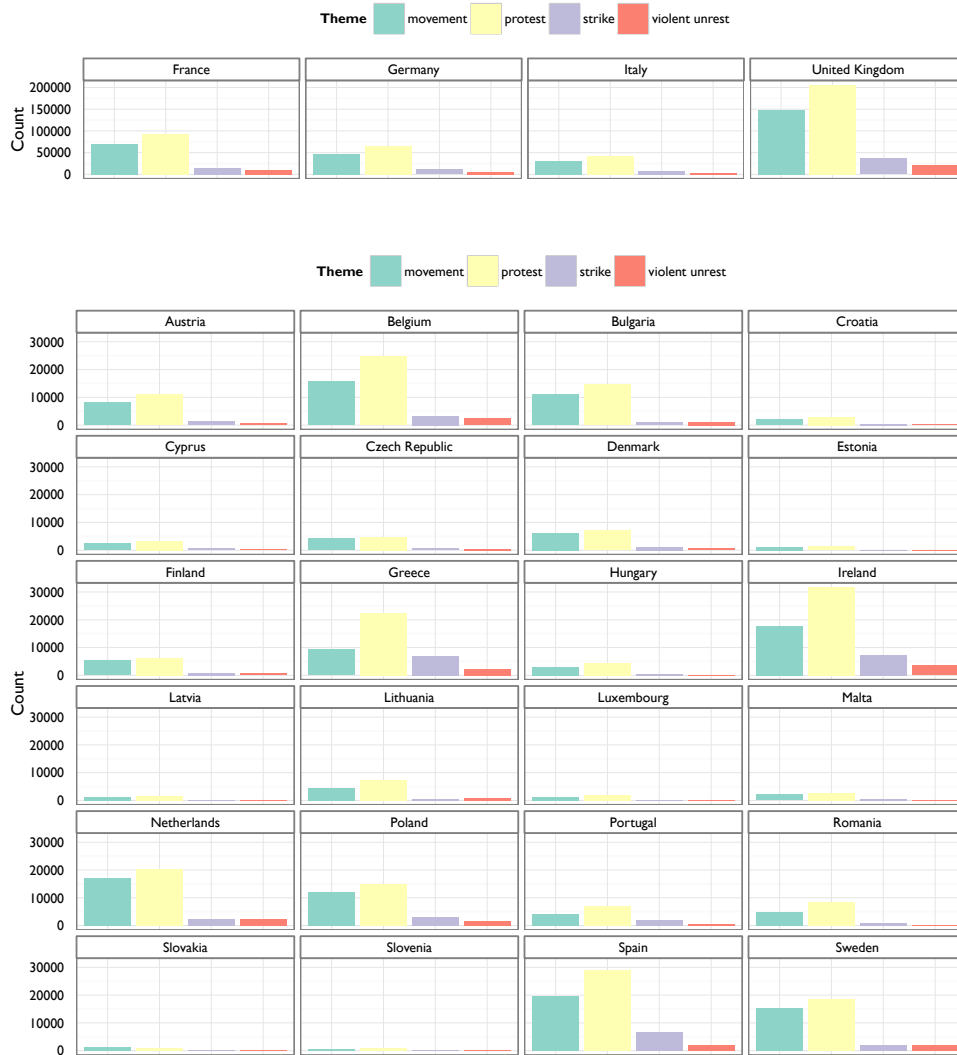


Figure 4: Types of Political Mobilization in Europe (Counts), April–December 2013

Note the difference in scaling between the top four countries with many counts and the lower grid, containing the countries with less reported activity.

this segment of the population, and therefore the extent of right wing opinions among the populace.

What can GKG tell us about the themes associated with protests in Europe in 2013? Examining both the number of protests and percentage of protests reported alongside certain issues can help us answer this question. In every case (with the exception of France) general economic themes are the most discussed (see Figure 6). In several countries, Croatia, Portugal, Greece, and Slovenia, these are followed by the second theme of unemployment. In at least three countries, Austria, the United Kingdom, and Sweden, however, Muslim themes are a close second. Indeed, the Muslim theme is, overall, the second most important theme after the general economic theme. The countries of Austria, Belgium, Croatia, Cyprus, Denmark, France, Germany, and the United Kingdom show a relatively high percentage (10-20%) of protests associated with Muslim themes. In Austria, the largest percentage of protests is associated with Muslim themes. While we

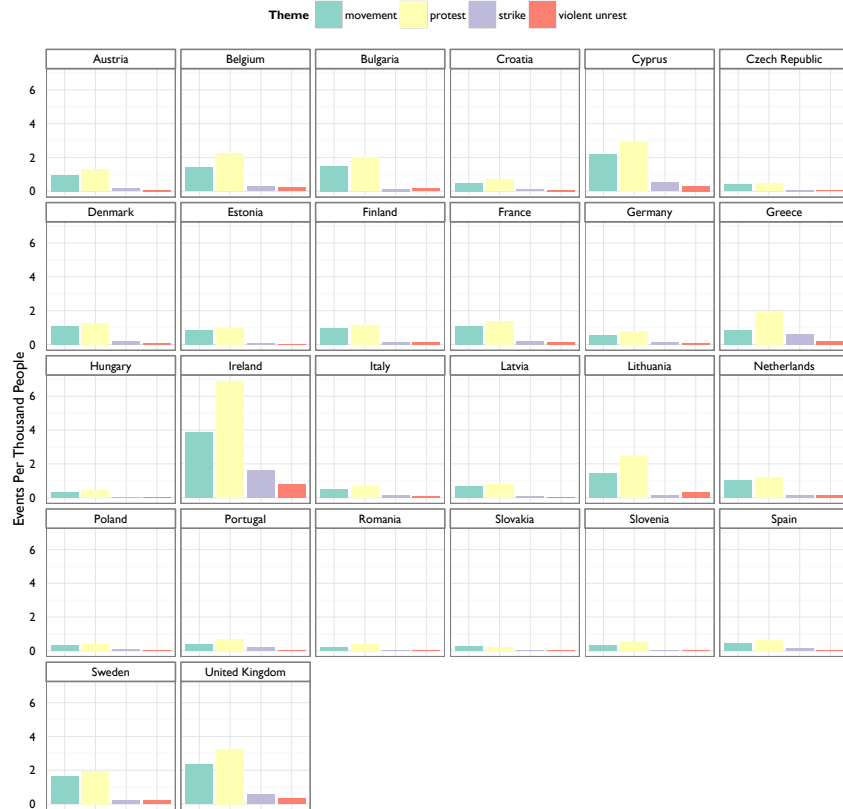


Figure 5: Types of Political Mobilization in Europe (Per Capita), April–December 2013

expected that those countries with a high percentage of Muslim themes might also embrace anti-immigration themes, this does not appear to be the case. Countries with a relatively high percentage of protests associated with immigration themes are Luxembourg, Bulgaria, Germany, Greece, Hungary, Italy, Malta, and Portugal.

Countries with a relatively high percentage of protests associated with LGBT issues may also reveal a high level of contestation over extending rights to this portion of the population and therefore relatively high levels of extreme right views. Opposition to LGBT citizenship rights including marriage rights is a central position of numerous extreme right parties. The relatively high percentage of protests associated with this theme in Croatia, Latvia, Luxembourg, and the Netherlands may signify significant activity of extreme right or conservative groups around this issue. In Croatia, for example, the Catholic Church and affiliated groups as well as extreme right political parties such as the Croatian Party of Rights, have launched an effort to pass a constitutional amendment banning same sex marriage and the “homosexual agenda” in public school health and sex education (Irvine and Sutlovic, (forthcoming); Kuhar, (forthcoming)). Moreover, significant incidents of violence surrounded the Pride Parade in the Croatian coastal city of Split in June in 2013, as in years past.

Nevertheless, it is difficult to measure the extent to which LGBT themes signify the presence of extreme right activity and views. For example, the very high level of protests in Latvia related to LGBT themes most likely reflects a mixed picture of support and opposition. Riga hosted the fourth-annual Baltic Pride march in June 2013. This march included nearly 600 participants, police were cooperative with pride activists, and members of Parliament as well as the Minister of Foreign Affairs attended the march (Amnesty International,

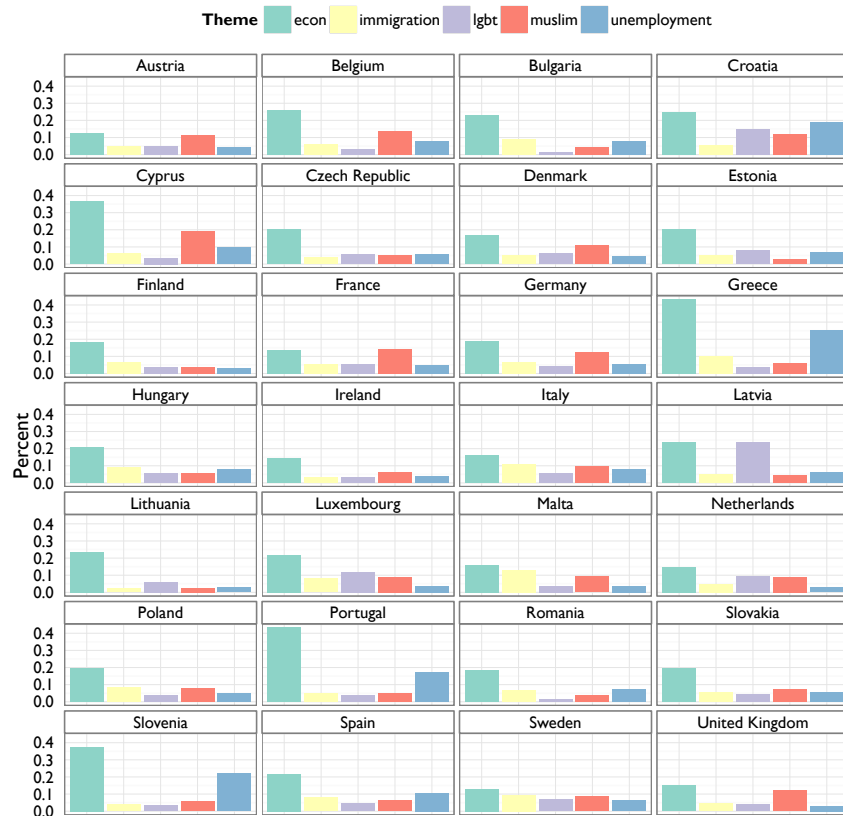


Figure 6: Protests and Associated Themes in Europe (Percent).
 Note that percentages may sum to more than 100 since events can have multiple themes.

2013) Nevertheless, considerable resistance to gay rights has continued in Latvia. Riga was chosen to host EuroPride 2015, which has resulted in popular efforts to pass a constitutional amendment banning gay marriage (IronClosetBlog, 2013).

Two groups of countries merit close attention when it comes to the percentage of protests associated with extreme right themes. The first is the group of bailout countries, Cyprus, Ireland, Portugal, Spain, and Greece, where the economic crisis reached particularly high levels and the bailout imposed harsh austerity measures. To what extent are extreme right themes associated with protests in these countries? A quick glance at Figure 6 reveals that immigration, LGBT, and Muslim issues are not obviously greater than in other countries. Rather, economic issues and unemployment are the most important themes in all bailout countries, suggesting that concern over these issues has not been directed at immigrant and other marginalized populations to any large extent. A possible exception is Greece where support for the main extreme right political party has increased significantly.

A second group of countries merits particular attention due to the increase in popular support for extreme right political parties. Extreme right parties in Austria, Finland, France, Greece, Latvia, and Hungary have all experienced substantial gains in the round of elections after 2008 (Mudde, 2013). The extent to which extreme right themes are associated with protests in these countries may tell us something about the extent to which these protests are associated with the political power of extreme right parties and predictive value of GKG when it comes to extreme right electoral successes. Austria and France show a relatively

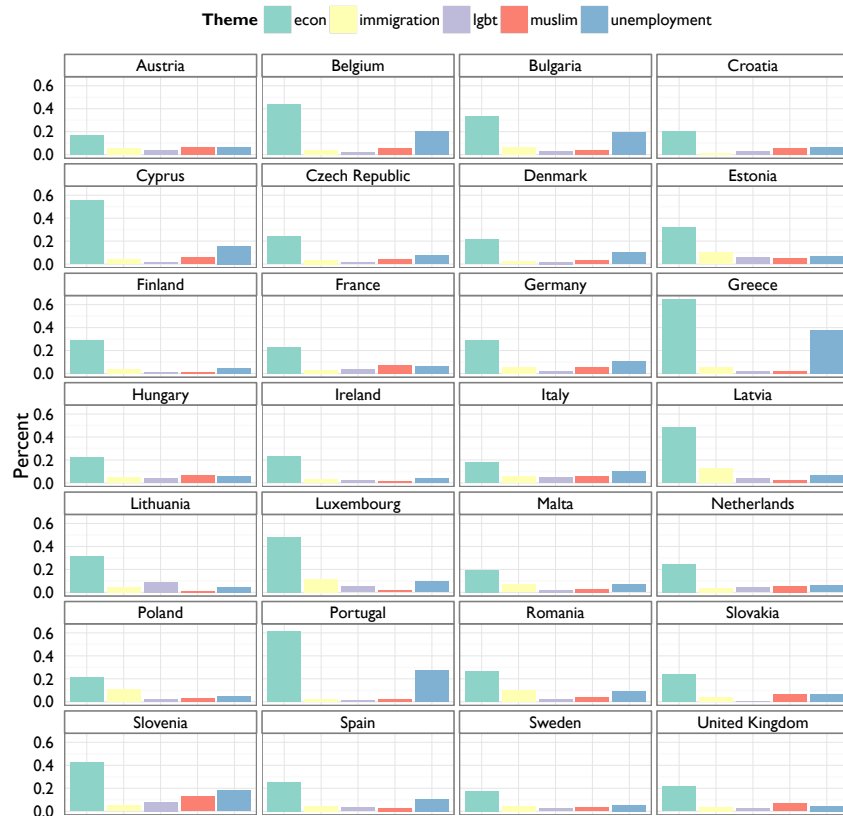


Figure 7: Strikes and Associated Themes in Europe (Percent)

high proportion of protests associated with (anti-)Muslim sentiment, and this has been a central aspect of appeals from extreme right parties and politicians in these two countries. Hungary, Finland, and Greece show moderate levels of protests associated with (anti-)immigration themes, while in Latvia a strong association exists between LGBT issues and protests. Thus, in every case one of the three extreme right themes is associated to a significant extent with protests, though in no case are all three themes present to a large degree.

In Figure 7 we consider the extent to which strikes are associated with the themes we believe would be linked to extreme right movements. Are discussions of Muslims and immigrants linked to strikes over working and other economic conditions? The data displayed in Figure 7 suggest that Muslim themes are somewhat more associated with strikes than immigration or LGBT themes, though economic issues and unemployment tend to be dominant, as we would expect. Slovenia is a curious case when it comes to strikes, showing the highest level of Muslim themes, the second highest number of LGBT issues, and in the top half of countries with immigration-linked strikes (by percent).

When it comes to themes associated with social and political movements, we see a much higher proportion of the three extreme right themes across most countries. While general economic concerns are the highest, several countries show levels of 20% or more of movement activity associated with a particular theme. For example, while strikes and protests in Croatia had relatively low associations with immigration, Muslim, and LGBT themes, when we look at general movement activity, these levels rise significantly to 8%, 20%, and 14% respectively. Nevertheless, the bailout countries of Cyprus, Ireland, Portugal, Spain, and Greece do not

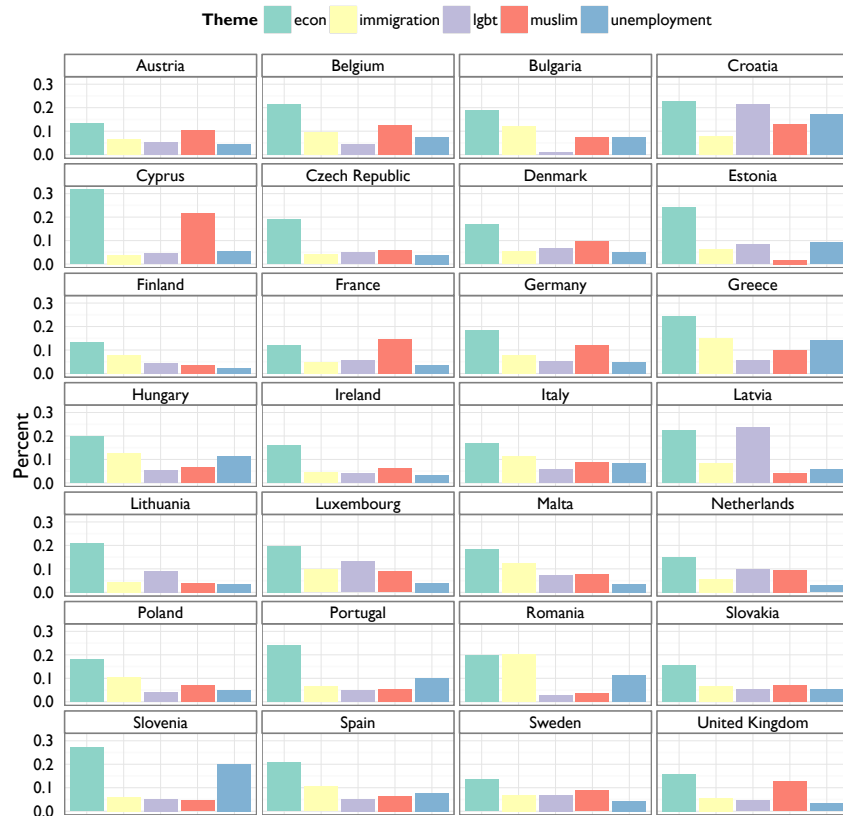


Figure 8: Social/Political Movement Events and Associated Themes in Europe (Percent), April–December 2013

show higher levels of extreme right themes associated with general movement activity; in all cases they are under 10 percent of movement activity. Of the countries where extreme right parties have done well (Greece, Latvia, France, Austria, Hungary, and Finland) only Latvia shows high levels of association between these three extreme right themes and general movement activity, with LGBT themes reaching 24 percent.

Finally, the last portion of the paper graphs themes related to protest over the time period for which we have data (see Figure 9). These longitudinal data allow us to present visually the appearance of particular themes during particular time periods. They also allow us to trace the movement of particular themes in relation to one another. One immediately obvious insight from the charts is that issues and countries tend to have stable levels of mobilization, while others have periods of higher and lower levels of mobilization. Spain and the UK have very consistent levels of coverage and the numbers of events around each issue are very stable, with only several cases of issues jumping or dropping in prominence. Contrast this with Italy, which has periods where economic, Muslim, and immigration themes rise from 3rd or 4th place to obvious prominence. We can imagine three potential explanations for this behaviour:

1. Some countries have more stable civil society organizations and institutions that consistently mobilize around a defined set of issues.
2. Some countries experience a series of mobilization-inducing events, such as austerity measures, new legislation, or financial problems.

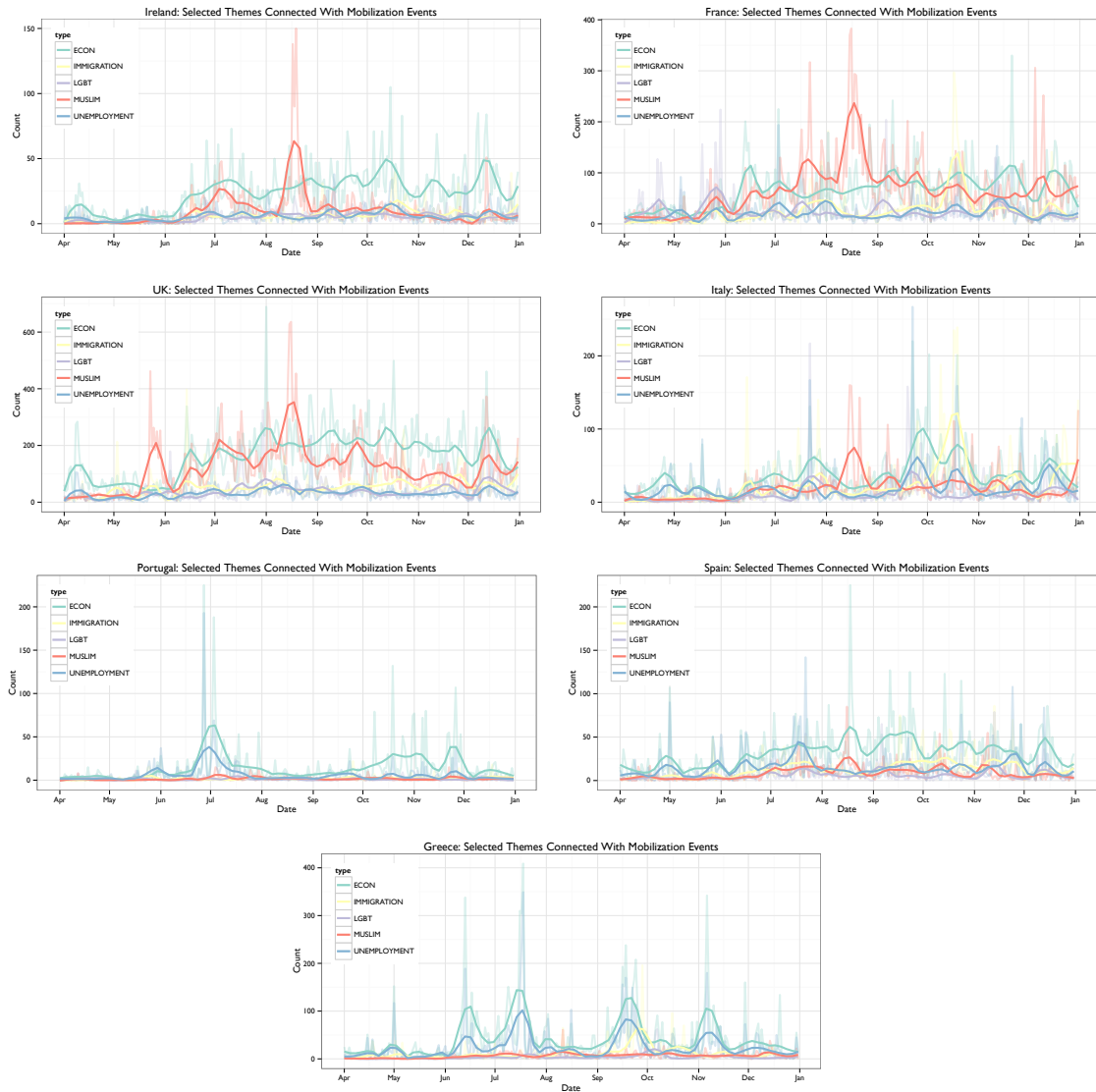


Figure 9: Selected Themes Related to Protest in Europe, April–December 2013. Fainter, spikier lines are raw counts; darker, smoother lines are less smoothing functions (see source code for details).

3. Countries' smoothness or spikiness is an artifact of media coverage, not entirely the result of actual mobilization. Certain issues tend to fluctuate more than others. Immigration, for example, tends to receive large amounts of coverage in brief periods only, in contrast to economic issues, which tend to receive relatively stable coverage and in periods of high mobilization are the most important issue.

In sum, although we are working with a limited time frame in using this experimental data set, GKG's theme data does appear to produce useful insights about the relative importance of particular themes over time. As we've seen here, in looking at our test themes we believe are related to the extreme right, we were able to make several observations that merit further study. First, the themes we designate as extreme right themes (immigration, LGBT, and Muslim) tend to vary in importance across countries, time, and types of events. Second, extreme right themes do not appear strongly associated with one another—the appearance of a relatively strong Muslim theme is not linked with potentially associated immigration themes in the

time series graph. Third, GKG’s theme data does not appear to support the argument that poor economic conditions in the bailout countries created favorable ground for extreme right appeals and political support. Rather, association between extreme right themes and political mobilization is relatively low in these cases.

4 Conclusion

We would like to conclude by offering lessons we learned from using massive event data to study political mobilization. Several observations can be made about the utility of mass event data to study political mobilization and government actions. First, large event data sets allow rapid comparison of a large number of cases over time, providing a powerful birds’ eye view. Visual presentation of the data allow us to discern broad patterns and identify outlier cases that can aid in generating new hypotheses and identifying cases that merit further study. For example, by mapping verbal and material conflict and cooperation, we can generate and explore hypotheses about mobilizing strategies and the effectiveness of government repression. Used in this way, event data provides a useful overview of an issue or a country because an in-depth qualitative research project.

Second, supplementing event data with theme codes offers a promising though perhaps limited research tool. Pairing extreme right themes with political mobilization events highlighted broad patterns of association and suggested fruitful areas for further research. For example, the data suggested a surprisingly low association of the extreme right themes with countries often assumed to be most susceptible as well as a general lack of extreme right theme clustering. Further multi-level, mixed method study of these questions is necessary, but GKG offers a valuable “first pass” and broad overview of patterns. Nevertheless, as we mentioned in our discussion of the methodological challenges of working with these and similar data sources, GKG appears to suffer from serious shortcomings which limit its usefulness for other purposes including tone and network analysis. Future event datasets, we believe, should incorporate theme information for its use both in post-hoc explanations of political dynamics, but also for forecasting applications.

This leads to our third conclusion, which is that knowledge of the data generating process is very important for event data machine coded from news reporting. With GDELT, the data generating process was not always transparent and working with it required a large degree of learned informal knowledge about its quirks and weaknesses, mostly learned from the large and active community of GDELT users. Future datasets should be much more transparent about how their components are generated, especially for innovations beyond the traditional actor and verb dictionaries. While machine coding will not capture the full understanding of area specialists, it can be improved by scholars adding to dictionaries and coding schema, which will shape future datasets into even more useful sources for studying political mobilization. Ideally, datasets should include URLs for all events and easily replicated coding systems to allow end users to understand how the data was generated and the effects of changes in dictionaries, coding algorithms, and sources. In sum, while the use of massive, machine-coded data to study such phenomena as civil society, political mobilization and government repression should be approached with caution we suggest that it can provide a useful research tool. Indeed, the predominant use in of machine-coded event data up to this point for conflict forecasting should be expanded to include a much wider range of comparative questions and methods, which will be made possible by more transparency and replicability.

References

- Amnesty International. Annual Report 2013: Latvia, 2013. URL <http://www.amnesty.org/en/region/latvia/report-2013>.
- Bryan Arva, John Beielser, Benjamin Fisher, Gustavo Lara, Philip A. Schrodt, Wonjun Song, Marsha Sowell, and Sam Stehle. Improving Forecasts of International Events of Interest. In *EPSA 2013 Annual General Conference Paper*, volume 78, 2013.
- Jamie Bartlett, Jonathn Birdwell, and Sarah de Lange. Populism in Europe: Netherlands, 2012.
- Patrick T. Brandt, John R. Freeman, Tse-min Lin, and Philip A. Schrodt. Forecasting Conflict in the Cross-Straits: Long Term and Short Term Predictions. In *Annual Meeting of the American Political Science Association, August 28-September 1, 2013*.
- Valerie J Bunce and Sharon L Wolchik. *Defeating authoritarian leaders in postcommunist countries*. Cambridge University Press, 2011.
- Thomas Carothers. *Critical mission: Essays on democracy promotion*. Carnegie Endowment, 2004.
- Jennifer Earl, Andrew Martin, John D. McCarthy, and Sarah A. Soule. The Use of Newspaper Data in the Study of Collective Action. *Annual Review of Sociology*, 30:65–80, 2004.
- The Economist. This monster called europe, 2013.
- Michael Edwards. *Civil society*. Polity, 2009.
- Steven E. Finkel, Anbal Prez-Lin, and Mitchell A. Seligson. The Effects of U.S. Foreign Assistance on Democracy Building, 1990–2003. *World Politics*, 59(3):404–438, 2007.
- Valère Philip Gagnon. *The myth of ethnic war: Serbia and Croatia in the 1990s*. Cornell University Press, 2006.
- Max Gallop and Michael D. Ward. Modeling, Computing and Evaluating Endogeneous Networks: Understanding International Cooperation. Duke University Working Paper, 2013.
- Deborah J. Gerner, Philip A. Schrodt, Omür Yilmaz, and Rajaa Abu-Jabr. Conflict and Mediation Event Observations (CAMEO): A new event data framework for the analysis of foreign policy interactions. *International Studies Association, New Orleans*, 2002.
- Volkart Finn Heinrich. *Assessing and Strengthening Civil Society Worldwide: A Project Description of the CIVICUS Civil Society Index: a participatory needs assessment & action-planning tool for civil society*. Civicus, 2004.
- IronClosetBlog. Latvian Group Seeks to Limit EuroPride 2015, 2013. URL <http://ironcloset.wordpress.com/2013/10/09/latvian-group-seeks-to-limit-europride-2015/>.
- Jill Irvine and Leda Sutlovic. Gender Equality in Croatia: Closing the Compliance Gap. In *Paper presented at the Conference on Gender (In)equality in Southeastern Europe, Trondheim Norway 6-8 September, 2013*, (forthcoming).
- Jill A Irvine. Electoral Breakthroughs in Croatia and Serbia: Women’s Organizing and International Assistance. *Communist and Post-Communist Studies*, 46, 2013.
- Katya Kalandadze and Mitchell A Orenstein. Electoral protests and democratization beyond the color revolutions. *Comparative Political Studies*, 42(11):1403–1425, 2009.
- Roman Kuhar. Playing with Science: Intimate/Sexual Citizenship and the Roman Catholic Church Counter Narratives. In *Paper presented at the Conference on Gender (In)equality in Southeastern Europe, Trondheim Norway 6-8 September, 2013*, (forthcoming).

- David Lane. 'Coloured Revolution' as a Political Phenomenon. *Journal of Communist Studies and Transition Politics*, 25(2-3):113–135, 2009.
- Marian Leonardo Lawson. USAID's Office of Transition Initiatives After 15 Years: Issues for Congress, 2009.
- Kalev Leetaru. Announcing the debut of the GDELT Global Knowledge Graph. <http://gdeltblog.wordpress.com/2013/10/27/announcing-the-debut-of-the-gdelt-global-knowledge-graph/>, 2013.
- Kalev Leetaru and Philip A. Schrodt. GDELT: Global data on events, location, and tone, 1979–2012. In *Paper presented at the ISA Annual Convention*, volume 2, page 4, 2013.
- Carmen Malena and Volkhart Finn Heinrich. Can we measure civil society? a proposed methodology for international comparative research. *Development in Practice*, 17(3):338–352, 2007.
- Cas Mudde. Three decades of populist radical right parties in western europe: So what? *European Journal of Political Research*, 52(1):1–19, 2013. ISSN 1475-6765.
- Sean P. O'Brien. Crisis Early Warning and Decision Support: Contemporary Approaches and Thoughts on Future Research. *International Studies Review*, 12(1):87–104, 2010.
- Political Mobilization of Minorities and Anti-Diversity Discourse in the Context of the European Crisis of Multiculturalism. Research report on the measurement of the political mobilization of minorities, CNCS-UEFISCDI: PN-II-ID-PCE-2011-3-0394.
- Philip A Schrodt. Tabari: Textual analysis by augmented replacement instructions, 2011.
- Michael D. Ward, Andreas Beger, Josh Cutler, Matt Dickenson, Cassy Dorff, and Ben Radford. Comparing GDELT and ICEWS Event Data. 2013a.
- Michael D. Ward, Nils W. Metternich, Cassy L. Dorff, Max Gallop, Florian M. Hollenbach, Anna Schultz, and Simon Weschle. Learning from the Past and Stepping into the Future: Toward a New Generation of Conflict Prediction. *International Studies Review*, 15(4):473–490, 2013b.
- James E. Yonamine. Predicting Future Levels of Violence in Afghanistan Districts Using GDELT. Chapter in PhD Thesis, Pennsylvania State University, 2013.