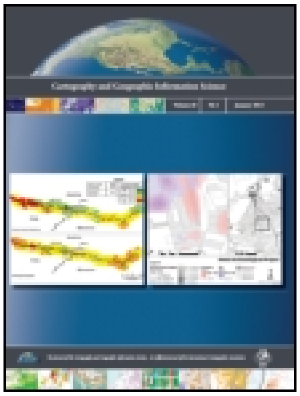

Cartography and Geographic Information Science
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tcag20

Demarcating new boundaries: mapping virtual polycentric communities through social media content

Anthony Stefanidis [a b] , Amy Cotnoir [a] , Arie Croitoru [a b] , Andrew Crooks [c] , Matthew Rice [b] & Jacek Radzikowski [a]

[a] Center for Geospatial Intelligence, George Mason University, Fairfax, VA, 22030, USA

[b] Department of Geography and Geoinformation Science, George Mason University, Fairfax, VA, 22030, USA

[c] Department of Computational Social Science, George Mason University

Version of record first published: 19 Apr 2013.

To cite this article: Anthony Stefanidis , Amy Cotnoir , Arie Croitoru , Andrew Crooks , Matthew Rice & Jacek Radzikowski (2013): Demarcating new boundaries: mapping virtual polycentric communities through social media content, Cartography and Geographic Information Science, 40:2, 116-129

To link to this article: http://dx.doi.org/10.1080/15230406.2013.776211

Cartography and Geographic Information Science, 2013
Vol. 40, No. 2, 116–129, http://dx.doi.org/10.1080/15230406.2013.776211

Taylor & Francis
Taylor & Francis Group

Demarcating new boundaries: mapping virtual polycentric communities through social media content

Anthony Stefanidis*[a,b], Amy Cotnoir[a], Arie Croitoru[a,b], Andrew Crooks[c], Matthew Rice[b] and Jacek Radzikowski[a]

[a]*Center for Geospatial Intelligence, George Mason University, Fairfax, VA 22030, USA;* [b]*Department of Geography and Geoinformation Science, George Mason University, Fairfax, VA 22030, USA;* [c]*Department of Computational Social Science, George Mason University, Fairfax, VA 22030, USA*

(*Received 2 September 2012; accepted 26 January 2013*)

The proliferation of social media has led to the emergence of a new type of geospatial information that defies the conventions of authoritative or volunteered geographic information, yet can be harvested to reveal unique and dynamic information about people and their activities. In this paper we address the identification and mapping of global virtual communities formed around issues of specific *national* interest. We refer to these connected virtual communities formed around issues related to a specific state as the *polycentric virtual* equivalent of that state. Identifying, mapping, and analyzing these virtual communities is a novel challenge for our community, and this is the subject we pursue in this paper. We present these communities relative to established conventions of statehood, address the harvesting of relevant geographical information from social media feeds, and discuss the challenge of visualizing such information. In order to do so we use the current geopolitical situation in Syria as a demonstrative example.

Keywords: social media; Twitter; community detection; community mapping; statehood; Syria

1. Introduction

By substituting physical with virtual interaction, social media have introduced a novel avenue for community building, transcending established administrative boundaries (Quercia, Capra, and Crowcroft 2012; Takhteyev, Gruzd, and Wellman 2012) to diffuse ideas and information across space (Allen-Robertson and Beer 2010). As the topics that bring together these communities vary widely, covering the broad spectrum from trivial to substantial issues (de Choudhury, Counts, and Czerwinski 2011), individuals may participate in multiple virtual communities, exploring various dimensions of their identities (Greenhow and Robelia 2009; Bennett 2012). Of particular interest for this paper are the global virtual communities formed around issues of specific *national* interest, just like the global activist communities that emerged as response to the recent Arab Spring events (Khondker 2011). During such events, a state's core community (namely, its own citizens within its boundaries) is expanded through the addition of satellite communities from abroad who are involved with this state's issues and become connected to its citizens. We refer to these connected virtual communities formed around issues related to a specific state as the *polycentric virtual* equivalent of that state. These virtual polycentric communities can play an important role in shaping global opinion about, and response to, events affecting a specific state, ranging from natural disasters to political crises (Goolsby 2010). Identifying, mapping, and analyzing these virtual communities is a novel challenge for our community, and this is the subject we pursue in this paper. We present these communities relative to established conventions of statehood, address the harvesting of relevant geographical information from social media feeds, and discuss the challenge of visualizing such information. In order to do so we use the current geopolitical situation in Syria as a demonstrative example.

The motivation for our work stems from the massive adoption of social media by the general public. In the spring of 2011, just five years after its 2006 launch, *Twitter* announced that it had over 200 million accounts, distributed all over the world. Among these accounts, it was estimated in September 2011 that *Twitter* had at least 100 million active users, logging in at least once a month, and 50 million users who do so daily. Less than a year later, in March of 2012 *Twitter* announced that its active users were up to 140 million[1], a 40% increase in only 6 months. As a measure of reference, a population of 140 million would make *Twitter* the 10th most populous country in the world, just behind Russia and Bangladesh. Furthermore, *Twitter*'s #numbers entry[2] indicates that a record 572,000 new accounts were created on a single day (12 March 2011, the day after the Sendai earthquake and resulting Fukushima nuclear disaster in Japan). An average of 140 million tweets are sent daily, resulting in an estimated billion tweets sent every week. However, *Twitter* is not even the largest social media user

*Corresponding author. Email: astefani@gmu.edu

community; Mark Zuckerberg announced on 4 October 2012 that *facebook* had reached a billion users, with more than half of them being active participants, using it daily. This would make *facebook* the third most populous nation in the world, behind only China and India. Extending beyond the English speaking world, *QQ* is a Chinese service for instant messaging, with over 700 million accounts, while *Renren* and *Kaixin001* are dominant Chinese social networking applications. This massive participation in social media makes the international virtual communities formed through them both populous and influential, thus rendering their study important. The identification and mapping of these communities by exploiting the geospatial content of social media feeds is a representative example of the novel types of geographical analysis that are enabled through the proliferation of social media.

The remainder of our paper is organized as follows. In section 2 we compare social media communities to established concepts of statehood, followed in section 3 by a discussion on the harvesting of geospatial information from social media and relevant cartographical considerations. Section 4 presents a test case, namely the participation of the international community in the developments in Syria, and we conclude with an outlook in section 5.

2. Connected virtual communities as novel hybrid states

It has now almost been a century since early seminal work in political geography addressed the notion of states and their borders (Holdich 1916; Fawcett 1918). During that time, the political boundaries of the world have changed numerous times: since its inception in 1945 membership in the United Nations has grown fourfold, from 51 to 193 states. Supported by this flux in statehood statistics, the analysis of states and boundaries has remained a popular topic in the geographic community (e.g., Newman and Paasi 1998; Agnew 2003). More recently our community has been re-examining the concept of borders and the identification of *effective* rather than *administrative* divisions of space, driven by the effects of globalization (Flint 2002; Sparke 2006; McCann and Acs 2011), new security challenges (Andreas 2003), and the push by geographically-agnostic communication means towards a 'borderless' world (Diener and Hagen 2009; Paasi 2009). Newman (2006) discussed the increased importance of the *process* of bordering, and the function of borders as zones of contact rather than barriers. He also discussed border scale, with borders refocused away from the state level, and down to internal regions, municipalities, and neighborhoods. Towards this same direction, Thiemann et al. (2010) analyzed human mobility networks and connectivity patterns to identify effective geographic subdivisions within the US by using the circulation of banknotes as a proxy for human mobility and interaction. In their approach they assessed subdivisions as internally similar and externally dissimilar when compared to their neighbors, and showed effective internal borders and subdivisions in the US, thus extending earlier work of Newman and Paasi (1998) on sociospatial identities.

While the above concepts had been distilled through physical interactions, the emergence of the Internet and cyber space presented challenges and alternatives to them. Negroponte (1996) argued early on that "the role of the nation-state will change dramatically" by the Internet. Web 1.0 presented a serious challenge to their domestic legal authority due to the nebulous nature of cyber activities. States struggled in the 1990s to impose their local laws against Internet activity, attempting to regain sovereign control over the virtual activities of their citizens, by regulating Internet access (Herrera 2007; Demchak and Dombrowski 2011). The Chinese government went as far as publishing a white paper outlining its arguments for Internet sovereignty (Jiang 2010).

However, while these early challenges were related to information *access* in the Web 1.0 era, a more interesting situation emerged through the introduction of Web 2.0, which altered the process through which this information is *contributed.* With its support for openness, interactive information sharing, interoperability, and on-line collaboration (see e.g., O'Reilly 2005), Web 2.0 ushered in a new era of user-generated content and information dissemination. At the same time, it expanded social connections through the emergence of social media applications. Social media enable the general public to communicate with their peers, sharing information with them instantly and constantly in an effortless and intuitive way. In doing so, social media enabled the formation of virtual communities that shared many of the characteristics of communities formed in the physical space. Accordingly, whereas Web 1.0 challenged the states to project their physical authority onto the cyberspace, Web 2.0 enabled the cyberspace (and virtual communities formed in it) to project itself into the real world, and reshape it accordingly.

As these virtual communities are formed across state boundaries, their projection onto the physical space also spans many states. When the topic that brings together such a virtual community is a national issue of one state (e.g., a natural disaster affecting it), this is equivalent to expanding that reference state's citizenry to include the involved overseas communities. Accordingly, a state's core community (namely, its own citizens within its boundaries) is expanded through the addition of satellite communities from abroad who are involved with this state's issues and become connected to its citizens. This process generates a *polycentric state structure*, where a state comprises not only its own citizens, within its established borders and under its jurisdiction, but also satellite communities from abroad, sharing an interest and/or active involvement with this state's affairs, as visualized in Figure 1.

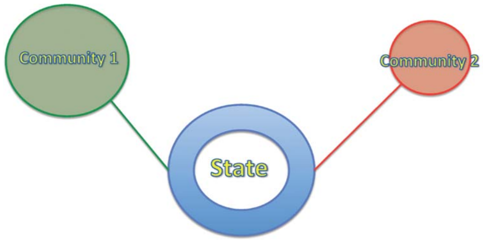

Figure 1. A polycentric state structure, comprising the core nation-state and satellite communities with active involvement in this state's affairs.

This participation-driven (as opposed to administrative) view of statehood can be considered as an evolution of the concept of identity-driven cultural citizenship and its relation to public opinion formation as a means to support state operations (Everard 2001; Hermes 2006). It is also consistent with the observation that on-line communities formed through social networks tend to display characteristics typically observed in nation-states (The Economist 2010). Lastly, it is also consistent with relevant observations in the field of political science regarding the evolution of the territorially consolidated sovereign nation-states through the internationalization of their societies (Goldmann 2001; Anderson 2002). It is worth mentioning here that the term 'polycentric' has been used in that community to refer primarily to the emergence of third parties within a state that complement its regulatory government structures (Axtmann 2004), rather than the context in which we use it here.

Thus, the traditional state and its borders is substituted by its virtual equivalent, in the form of connected communities of stakeholders, and the aggregate of their spatial footprints. The State itself clearly remains as the central unit in this new union, and satellite communities may vary in their membership size: the green community in Figure 1 has more members than the red community. The links among these communities may stem from various aspects of life, ranging for example from cultural and economical to political, thus creating multiple levels of such connections. This is comparable to the manner in which different urban realms are forming large metropolitan areas (Vance 1977), or to the polycentric nature of the modern megalopolis (Lang and Knox 2009).

There are however some notable variations when comparing a polycentric state to its megalopolis equivalent. The first difference is geographical: the connected communities forming a polycentric state do not have to be contiguous, as spatial proximity is no longer necessary for the establishment of social connections. The second is temporal: polycentric state connections tend to have an ephemeral component too. Communities that are traditionally connected to a particular state (e.g., ex-patriot communities abroad) are occasionally joined temporarily by additional communities, who may be drawn, for example, by particular events that affect a state, as part of an outpour of sympathy following a natural disaster, or as an expression of concern during a political upheaval. The third difference has to do with the nature of the relations linking these communities and which, in a polycentric state, may also be adversarial. For example, the red community in Figure 1 may be opposing the ideas of the green community, or even of the State itself (ranging from simple political disagreement to nationalistic disputes).

The last difference concerns their formation process. The polycentric statehood concept reflects a process of *aggregation* from different units to form a connected cluster, rather than the *dispersal* that is commonly leading to the generation of urban realms out of a growing central unit in a modern megalopolis. This aggregation is comparable to Foucault's argument that societal configurations (such as states) are products of bottom-up processes, and are built through aggregations of smaller units to form power structures out of micro-power relations (Foucault 1980). We, in turn, argue that the societal connections observed through social media interaction are an example of this bottom-up approach: this interaction supports the formation of social networks that become *de facto* members of the states to which their members belong. Thus, these connections link states and in doing so expand each one individually, to incorporate in it communities from abroad.

As we use social media content for our analysis in the next section we will address the harvesting of relevant information from social media feeds in order to support our analysis.

3. Harvesting and mapping geospatial information from social media feeds

Harvesting information from social media feeds entails, in general, three operations: extracting data from the data providers (various social media servers) via APIs; parsing, integrating, and storing these data in a resident database; and then analyzing these data to extract information of interest (Croitoru et al. 2012).

There exist a number of tools that perform parts of these processes, such as 140kit (http://140kit.com/), or twapperkeeper (http://twapperkeeper.com/), but these are limited in their scalability with respect to large datasets. Sites such as ushahidi (http://www.ushahidi.com/) also provide a means to collect and disseminate information over the Web. However, currently available tools offer limited capabilities to add context to content, or to support detailed analysis, thus forcing the development of custom systems to perform the above-mentioned three operations.

Original social media feeds can be retrieved from source data providers through queries. This entails submitting a query in the form of an http request and receiving, in response, data in XML format (e.g., *Atom* or *RSS*). The query parameters may be, for example, based on location (e.g., specifying an area of interest to which the feed is related), time (e.g., specifying a period of interest), content (e.g., specifying keywords), or even by user handle/ID. In response to these queries, and depending on the characteristics of the information provided by the service, we can receive from the server just *metadata* or *metadata and actual data*. A representative example of the first case is *flickr*, where the query result contains exclusively metadata information (e.g., author, time, and geolocation when available), and information on how to access the actual image itself. A representative example of the second is *Twitter*, where the data received in response to a query are actual tweets and associated metadata (e.g., user information, time of tweet publication, geolocation when available, and information on whether this particular tweet is in response to or retweet of an earlier message).

Once this information is harvested from the social media server it can be parsed to become part of a *local database* (e.g., implemented using ProstgresSQL) for subsequent querying and analysis. Depending on the subject, queries may be submitted infrequently, with their frequency intensified during episodes of crisis. While the information harvested from social media in this manner is not explicitly geospatial, it does include implicit geospatial content, thus rendering it suitable for novel types of geospatial analysis as we show in the following section.

The geolocation component of these feeds is of particular importance for our community. Considering tweets as an example, this geolocation information can be provided directly by the contributors, if they decide to make this information available. Alternatively, the *Twitter* service itself can deduce geolocation information from a contributor's IP address (see Eriksson et al. 2010) and append a corresponding placename to the resulting tweet. The accuracy of this IP-derived Twitter-provided geolocation information may range from building level all the way to broader neighborhood (see Poese et al. 2011). We are focusing primarily on geolocation information that is contributed either directly by the user or provided through the client application.

This geolocation information may be available at two different levels of granularity: either in the form of precise coordinates as shown in Figure 2, or in a descriptive manner (e.g., listing only a city name). It is typically


```
{
  ...
  "place":{
    "name":"Rochester",
    "attributes":{},
    "full_name":"Rochester, NY",
    "place_type":"city",
    "url":"http:\/\/api.twitter.com\/1\/geo\/id\/2f1fc0d72969452b.json",
    "country_code":"US",
    "bounding_box":{
      "type":"Polygon",
      "coordinates":[[[-77.701632,43.103318],[-77.531166,43.103318],[-77.531166,43.269045],[-77.701632,43.269045]]]},
      "country":"United States",
      "id":"2f1fc0d72969452b"
    },
  },
  "coordinates":{
                        "type":"Point",
                        "coordinates":[-77.60793686,43.16239715]
                },
  "geo":{
                "type":"Point",
                "coordinates":[43.16239715,-77.60793686]
                },
  ...
  "user":{
    ...
    "location":"NY x 2",
    ...
  },
  ...
  "text":"..."
}
```



```
"geo":{
      "type":"Point",
      "coordinates":[43.16239715,-77.60793686]
      },
```


Figure 2. An example of the metadata of a tweet, provided by the *Twitter* API. Among a variety of geolocation information we have zoomed to the actual user coordinates.

harvested from *Twitter* using the capabilities provided by the communication protocol linking the client to it. For example, the World Wide Web Consortium (W3C) geolocation Application Programming Interface (API) enables scripting code to access device information from Web browsers of any Web-capable device (e.g., a mobile phone or a laptop). In this way information is collected in a dynamic mode, reflecting the actual location from where a tweet was sent. In addition to this, geolocation information can also be harvested from the content of users' profiles, but this is less reliable as it is static and does not necessarily reflect user location at the moment that the tweet was sent.

Reports on the percentage of geolocated tweets vary substantially, from as high as two thirds (Hecht et al. 2011) or half of the tweets having some location information, in the form of coordinates or description (Java et al. 2009), to as low as 5% of the users listing actual coordinates with another 21% listing descriptive geolocation information (Cheng, Caverlee, and Lee 2010). These variations in the availability of geolocation information tend to be geographic and thematic in nature. Precise geolocational information is more frequently available when a mobile device has been used to post a tweet. Accordingly, one can reasonably argue that precise geolocation information will be more frequently available in areas where the latest mobile technology is adopted easily and rapidly. The variation is also thematic in nature, as it appears to relate to the event under study. Users are more likely to use their mobile devices to tweet about significant events as they are unfolding. For example, in an earlier study where we used twitter data from Japan during the Fukushima disaster, we harvested a sample with a very large percentage of tweets (16%) having precise coordinate information. During our studies over the past two years we have observed that on average the percentage of precisely geolocated tweets tends to be much lower, often in the range of 0.5–3%.

Considering the particular topic pursued in this paper, our understanding of political and social structures is influenced by the methods we use to analyze, represent and display information harvested from social media feeds. We assert that the geospatial content harvested from social media feeds is important in analyzing social and political events, and in fact, can provide evidence of geographically dispersed communities operating as a polycentric virtual state. Kitchin and Dodge (2011), in their discussion of software, society, and the transduction of space, note that many social scientists ignore the role that space plays in understanding how software and technology shape society. They assert that social relations are inherently spatial and temporal, and that spatial context is not incidental to understanding social space, but rather that "the *where* makes a difference to the *what* that unfolds" (Kitchin and Dodge 2011). When considering the issue of national boundaries, Wood (2010) asserted that mapping has been indispensible in shaping the notion of statehood, in defining borders, consolidating frontier territory, and giving the state shape. He notes that maps 'helped bring the state *into being* – maps helped *construct* the state' (Wood 2010).

Despite the emergence of a variety of methodologies to visualize Web-accessible geospatial information (see Elwood 2010), traditional choropleth maps are still a popular choice to map datasets that are comparable to ones that we harvest from social media. However, their inherent aggregation of information within pre-set geographical boundaries (e.g., political) introduces well-known limits in the ability to visualize trends that transcend traditional boundaries. Dodge and Kitchin (2001) extensively explored the mapping of cyberspace, including synchronous social spaces where end-users interact with each other. They include examples of maps and diagrams of the connections in virtual communities, most of which are highly schematic and lack a correspondence or tie to actual geographic space due to the abstract nature of the virtual-interaction space. Abrams and Hall (2006) show many other examples of schematic maps where connections between individuals are depicted in schematic fashion.

In a rapidly changing human landscape, where social media plays a central role in shaping events, what role should maps play and what cartographic considerations should come into play? In the particular case we are addressing here it is important to identify both the spatial distribution of the global community that is involved in a specific cause, as well as the links among states as they are established through the links of their subjects. We will use as a test case the issue of Syria in the next section, in order to show how mapping relevant social media content can reveal the structure of the corresponding polycentric virtual community.

4. A polycentric Syria

4.1. Dataset

As a test case for this paper we are using Syria, with its strong geopolitical importance given the prolonged civil war that started in March 2011 as part of Arab Spring and is currently at what is expected to be its last stages. Using a system prototype that we have developed to harvest such information (Croitoru et al. 2011) we collected *Twitter* feeds through keyword queries to *Twitter*'s streaming API.

In order to demonstrate how the structure of a virtual polycentric Syria can be gleaned through the analysis of social media content we use for this publication tweets that we collected over a period of one week (10 July through 17 July 2012) and included mentions to Syria or its hashtag equivalent #Syria. As per standard *Twitter* terms of use, these tweets are a random 1% worldwide sample of *Twitter* traffic on this topic. We harvested a total

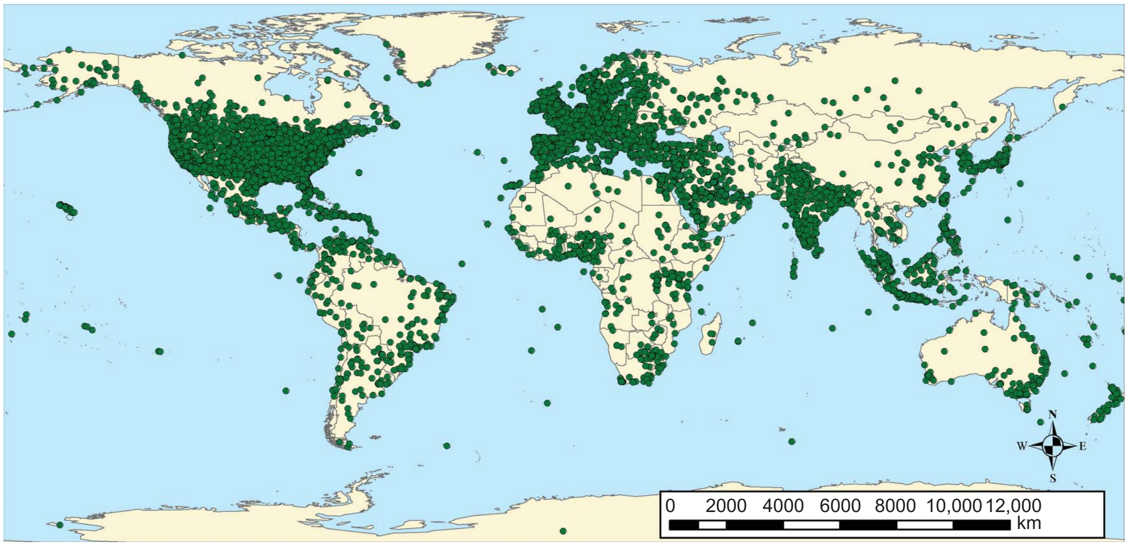

Figure 3. The originating locations of geolocated tweets referring to Syria in our sample dataset (period 10 July 10–17 July 2012). The locations are marked by green dots.

of over 700 thousand tweets about Syria (for a rate of approximately 4 thousand tweets per hour). From among these tweets 325,612 (approximately 45% of the total) had some sort of geolocation information associated with them, ranging from state and city level down to precise coordinates. In Figure 3 we show the global distribution of these tweets, with green dots marking originating locations of these contributions. It is easily understood that, given the large number of geolocated tweets in our study, a single dot in Figure 3 may correspond to multiple tweets originating from that location.

4.2. Analysis

In Figure 4 we show a temporal histogram depicting the flow of this information over our study period. We can observe a peak on 12 July from 21:00:00 to 23:00:00 EDT. This coincides to news breaking out of Syria reporting the massacre of over 200 activists in the city of Tremseh (Hama region). We analyzed the word content of the tweets over that period, and in Figure 6 we show a frequency word cloud capturing the key words communicated by the tweets. We can see that even in a busy news day, with Nawaf al-Fares, the Syrian ambassador to Iraq, defecting[3], and a Russian ship attempting to deliver refurbished helicopters to the Syrian army[4], the massacre news broke into *Twitter* very quickly and prominently. An interesting observation to be made from the content of Figure 5 relates to toponym usage within these messages. We can observe that the public uses two different variants of the latin spelling of the village (Traymseh, Tremseh) as well as the formal arabic spelling (التريمسة). For a more detailed analysis of trend detection over streaming feeds the reader is referred to Mathioudakis and Koudas (2010) or Lerman and Ghosh (2010).

An alternative technique to visualize the stream of data is through the data clock feature of ArcGIS. As we see in Figure 6, the data are plotted as concentric circles representing days of the month, with radial lines representing each hour in a day. The peak in the sample data of Figure 4 is clearly recognizable in figure 6 (as a dark blue patch), and is pointed by the red arrow.

In addition to easily visualizing outliers, this approach is also revealing persistent patterns of contributions. We

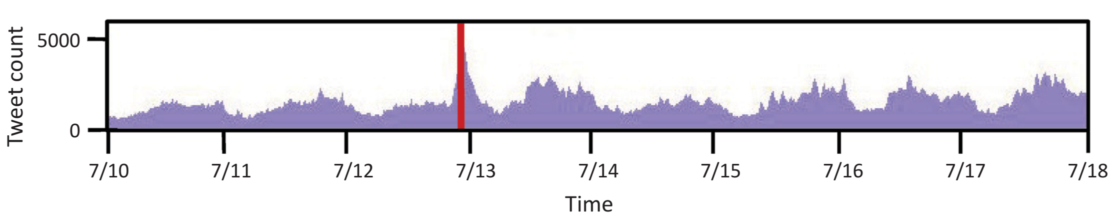

Figure 4. A histogram of tweet samples from 10 July 2012 (12:00:00) through 17 July 2012 (22:46:59) with a peak in activity on 12 July from 21:00:00 to 23:00:00 EDT.

Figure 5. A frequency word cloud summarizing *Twitter* content during the peak hour identified in Figure 5. It shows the 150 most frequently used words in that data corpus (excluding stop words), and the size is proportional to usage frequency. The word cloud was generated using wordle (Viegas, Wattenberg, and Feinberg 2009) and its default list of English language stop words (such as frequently used articles, conjunctions etc.)

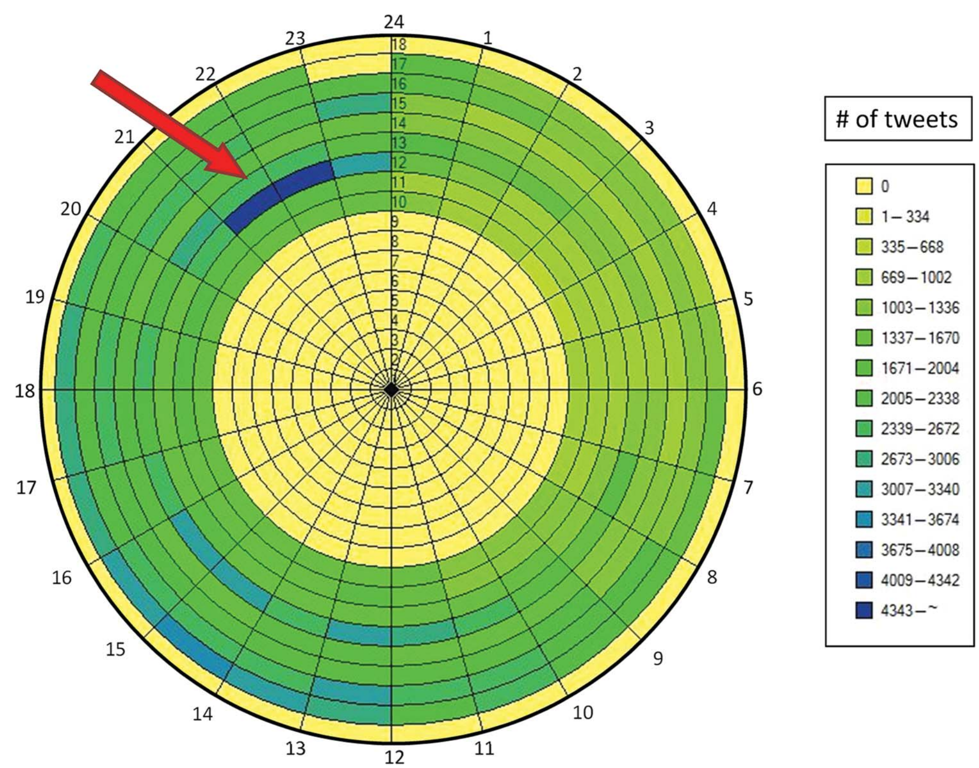

Figure 6. A data clock showing tweet counts per day of the month by hour of the day. As our data are for the period 10–17 July 2012, the interior circles, corresponding to days 1–9 have zero values and are shown in the corresponding yellow color.

can see for example that afternoons and evenings show higher traffic than mornings. Based on that we can also identify deviations from such patterns (e.g., lower than expected traffic around 15:00 of 14 July, or in the evening of 10 July). As the above-mentioned example, with the peak associated with the breaking of news reporting the Tremseh massacre demonstrated, these observations allow us to link data trends to real-life events, and are therefore extremely valuable.

In order to identify the spatial distribution of these contributions we generated a density surface of the origin locations of geolocated tweets in our data corpus. The map in Figure 7 is visualizing the global epicenters of the discussion about Syria in Twitterdom, as the hotspots from which these contributions originate. We can recognize that in addition to the major epicenter in Syria (and across its neighboring states) the secondary regions of concentration comprise the Gulf States (spanning Kuwait, Saudi Arabia, and the United Arab Emirates), Western Europe, and the United States (primarily the two coasts, and Southeast).

The data displayed in Figure 7 are aggregates over our period of study, and offer a reliable representation of the spatial clusters in our data. A similar analysis may also be performed over streaming data. In Figure 8 we show epicenters using two 120-second-long slices of streaming data from our dataset. These two datasets are captured 5 minutes apart from each other. We can see that in such brief time windows there is significant temporal variation in hotspots, but nevertheless the key actors (Syria, Gulf States, and the east coast of the US) remain recognizable. All these data were collected using a system prototype developed by our group (*self citation withheld*), and in Figure 9 we show a screenshot of the system user interface as it harvests, analyzes, and displays geolocated tweets.

In order to assess the structure of the polycentric virtual Syria we clustered our data by states and identified the key international communities that are involved with the topic of Syria. Table 1 is summarizing the top 11 communities (USA plus the top ten contributors worldwide). In this table we have ranked these communities by the absolute number of tweets contributed by their members during our period of

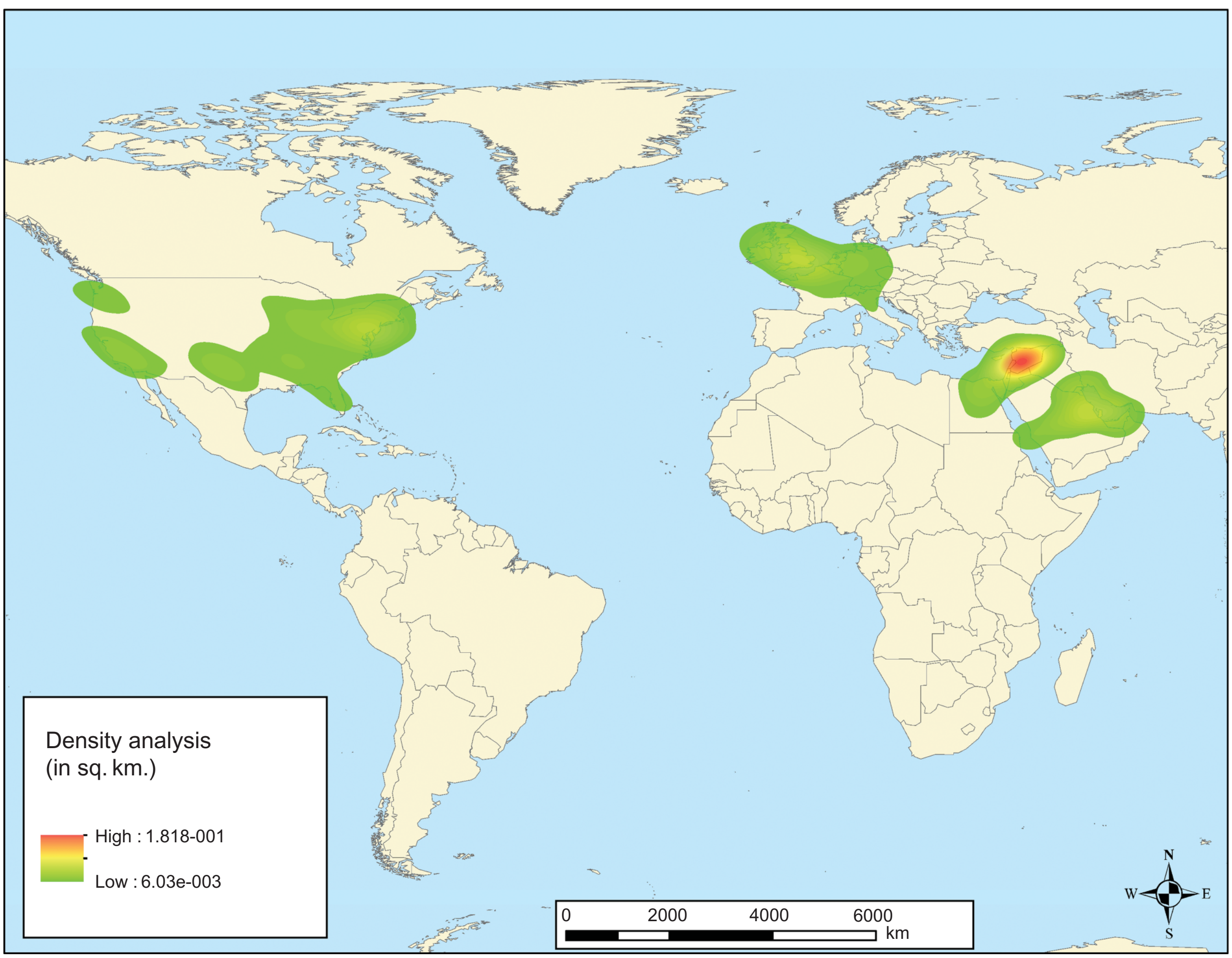

Figure 7. A point density analysis revealing hotspots in geolocated tweets, generated using ArcGIS 10 Spatial Analysis.

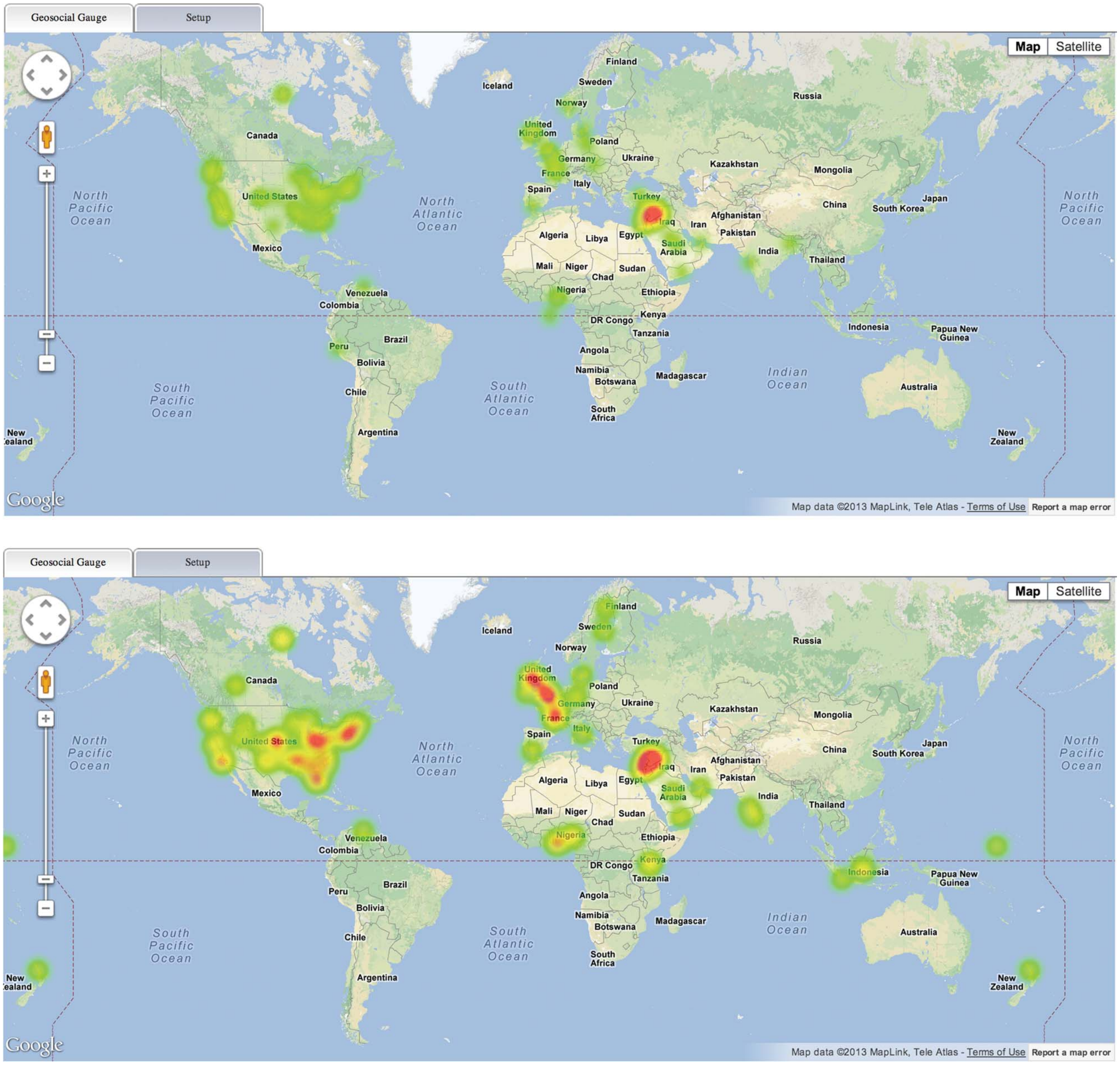

Figure 8. Temporal variations of hotspots as they are captured from 2-minute data streams that are 5 minutes apart from each other. Top is the earlier snapshot.

study (second column). While the absolute value is a valuable metric of participation it has an obvious population bias: large nations are expected to contribute more feeds than small. In order to remove this bias we introduced the concept of a *participation index* (fourth column of Table 1), defined as the number of tweets per thousand citizens for each of these countries. This allows us to recognize for example that while the United States are responsible for more *Twitter* traffic on the subject, Syria (as expected) and Kuwait (an interesting observation) are much more involved with the problem (with their participation indexes being tenfold that of the US). The normalized index serves as a more suitable metric (compared to absolute number of tweets) to express the level of involvement of various states in this particular topic. This information is plotted in Figure 10, showing the spatial distribution of these key actors.

Using the above information we can generate the network of these key international communities as we had conceptually introduced in section 2 and Figure 1. Figure 11 shows the polycentric model of virtual Syria as it is constructed using the content of tweets we harvested for our test case. We are showing the structure of this network using both raw traffic values and population-

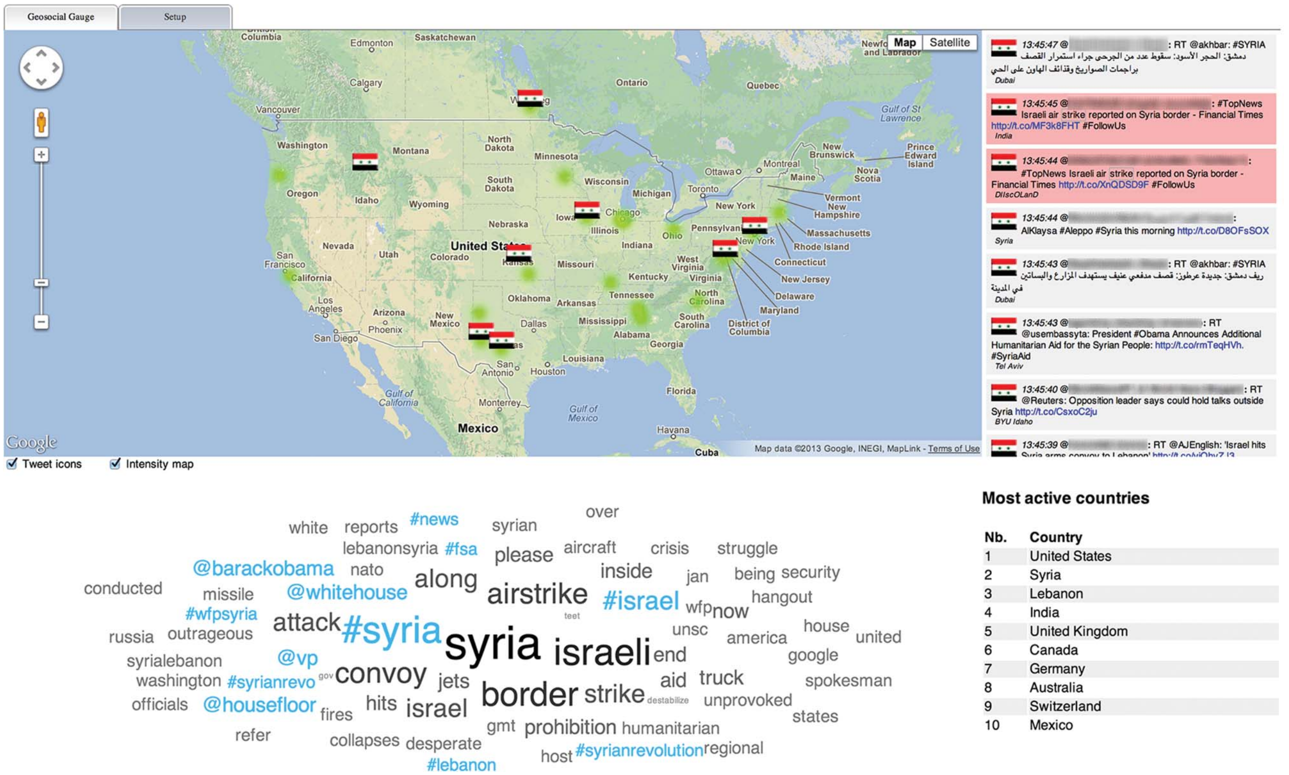

Figure 9. The user interface of our system prototype to harvest, map, and analyze streaming *Twitter* feeds.

Table 1. Top tweet-generating countries for our dataset.

State	Twitter traffic (number of tweets)	Population	Participation index
USA	82,361	313,847,465	0.262
Syria	59,717	22,530,746	2.650
Saudi Arabia	19,315	26,534,504	0.728
United Kingdom	17,242	63,047,162	0.273
Kuwait	8808	2,646,314	3.328
Canada	8204	34,300,083	0.239
Egypt	8012	83,688,164	0.096
India	4546	1,205,073,612	0.004
Germany	4149	81,305,856	0.051
United Arab Emirates (UAE)	3838	5,314,317	0.722
Australia	3531	22,015,576	0.160

normalized participation indices because they both provide valuable insight. The raw data network (Figure 11a) shows very well the community that dominates in the discussion in Twitterdom (US in this case). The population-normalized network (Figure 11b) shows the importance of the topic for the corresponding communities. We can see that despite traffic data it remains a rather marginal issue for the US society, while it is a major topic for the Kuwait community for example. These network representations complement the map of originating locations as shown in Figure 3, and the map of hotspots as shown in Figure 7 to capture the polycentric virtual presence of Syria.

In the above process we took into account country populations in order to normalize our social media data, deriving population-normalized participation indices. An alternative option would be to normalize by the number of social media users per country, in order to have a better understanding of the impact of a specific event on the corresponding user community. In order to do so we collected information on the number of active *Twitter* users in the various countries of our study (using records for the Arab nations in particular[5] and the rest of the world[6]), and generated a *Twitter-user-normalized* representation of the polycentric virtual Syria, defined as the number of tweets per thousand active *Twitter* users for

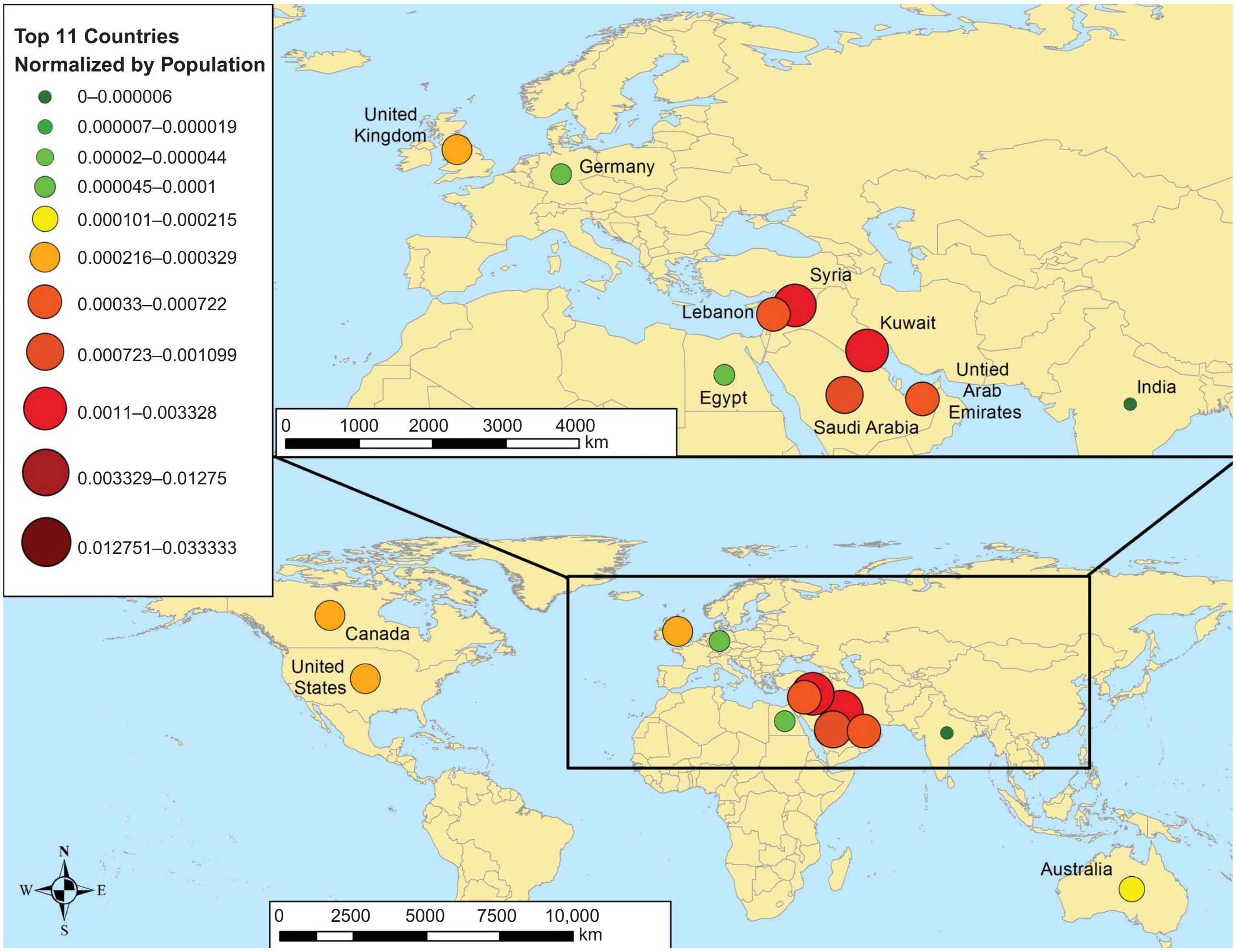

Figure 10. The top 11 international communities of Table 1 plotted with their normalized participation index metrics.

each of these countries. This alternately normalized dataset is shown in Figure 11(c). Compared to its counterpart of Figure 11(b) it emphasizes more the connection between Syria and the Arab nations of the group (Kuwait, Egypt, Saudi Arabia, UAE).

While the graph of Figure 11(c) expresses better the impact of the studied event on various *Twitter* user communities, it does not communicate well the relative importance of these user communities within their respective states. An event may impact massively the *Twitter* community of a nation, but that community may be a miniscule part of the population, thereby diminishing its state impact. In contrast, the population-normalized graph of Figure 11(b) better represents the impact of the event on the general population of a state.

5. Outlook

The proliferation of social media has led to the emergence of a new type of geospatial information that defies the conventions of authoritative or volunteered geographic information, yet can be harvested to reveal unique and dynamic information about people and their activities. In this paper we addressed the formation of virtual communities around national issues and the identification of *polycentric virtual* equivalents of states. Stemming from the community-building nature of social media, a state's core community (namely, its own citizens within its boundaries) is expanded through the addition of satellite communities from abroad who are involved with this state's issues and become connected to its citizens. In this paper we presented key concepts behind this new emerging paradigm of *polycentric virtual statehood*. We discussed in particular how *Twitter* content can be harvested to collect relevant information, and discussed relevant cartographic considerations, using Syria as a test case. We showed how we can map globally-distributed communities that are involved with the situation in Syria. This offers us a new perspective to geopolitical boundaries, showing how the world is structured and connected despite its political boundaries rather than because of them.

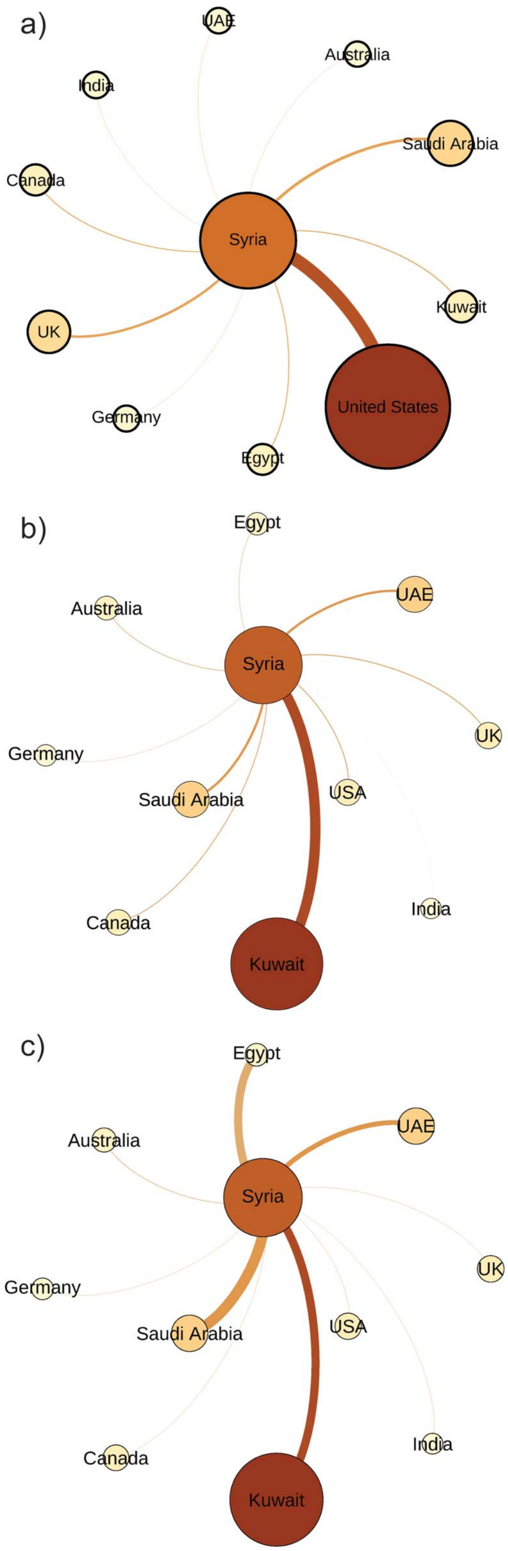

Figure 11. A network representation of the polycentric virtual Syria as it is captured using *Twitter* feeds for the period 10–17 July 2012. (a): using raw *Twitter* traffic values. (b): using the population-normalized participation indices. The size of the nodes indicates level of participation in the corresponding model (c): using Twitter-user-normalized metrics.

Our objective was to highlight these emerging geographic opportunities as a result of the proliferation of social media, rather than to create a definitive map of virtual Syria. Accordingly, we used for our analysis *Twitter* user references to Syria over a period of 8 days in order to demonstrate the involved processes and cartographic issues. In addition to simple references to Syria we can also analyze the structure of the social network that comprises the contributors of these feeds (Stefanidis, Crooks, and Radzikowski 2012) in order to identify and quantify the links among the various satellite communities that comprise virtual Syria. This was beyond the scope of this publication, but nevertheless, the concepts and issues we presented here would apply to this additional information as well. Furthermore, it is easily understood that social media participation is just one of the ways through which these types of connections are established. Various aspects of life, ranging for example from cultural and economical to political activities, are creating comparable links, but this type of analysis was beyond the scope of this publication. Nevertheless, relevant data can be mined to establish and quantify these additional connections in a manner comparable to the ones captured in the network representations of Figure 11, and this would be a suitable future extension of the work we presented here.

Harvesting this type of information brings forward novel challenges to the issue of privacy, as analysis can reveal information that the contributor did not explicitly communicate (see Friedland and Sommer 2010). This presents an additional issue of concern to the rather long list of challenges associated with privacy protection in the presence of pervasive location acquisition technologies (Xie, Kulik, and Tanin 2011; Lu and Liu 2012). Overcoming these privacy issues will help us take full advantage of the tremendous opportunities offered by social media to gain a unique understanding of the human landscape, its structure, connections, and its variations over time.

Notes

1. http://www.prnewswire.com/news-releases/twitter-users-have-quadrupled-over-past-two-years-158401225.html
2. http://blog.twitter.com/2011/03/numbers.html
3. http://www.guardian.co.uk/world/2012/jul/11/syria-ambassador-iraq-defected-opposition
4. http://www.nytimes.com/2012/07/13/world/europe/russian-ship-with-syrian-helicopters-embarks-on-renewed-voyage.html
5. http://www.arabist.net/blog/2012/8/24/twitter-active-users-in-arab-world.html
6. http://semiocast.com/publications/2012_07_30_Twitter_reaches_half_a_billion_accounts_140m_in_the_US

References

Abrams, J. and P. Hall, eds. 2006. *Else/Where: Mapping New Cartographies of Networks and Territories*. Minneapolis, MN: University of Minnesota Press.

Agnew, J. 2003. "Contemporary Political Geography: Intellectual Heterodoxy and its Dilemmas." *Political Geography* 22 (6): 603–606.

Allen-Robertson, J., and D. Beer. 2010. "Mobile Ideas: Tracking a Concept through Time and Space." *Mobilities* 5 (4): 529–545.

Anderson, J. 2002. "Questions of Democracy, Territoriality and Globalisation", In *Transnational Democracy: Political Spaces and Border Crossings*, edited by J. Anderson. London: Routledge.

Andreas, P. 2003. "Redrawing the Line – Borders and Security in the Twenty-first Century." *International Security* 28 (2): 78–111.

Axtmann, R. 2007. "The State of the State: The Model of the Modern State and Its Contemporary Transformation." *International Political Science Review* 25 (3): 259–279.

Bennett, W. L. 2012. "The Personalization of Politics Political Identity, Social Media, and Changing Patterns of Participation." *The Annals of the American Academy of Political and Social Science* 644 (1): 20–39.

Cartwright, W. 2005. "Linking Geographical Facts with Cartographic Facts", In *Cybercartography: Theory and Practice*, Modern Cartography Series. Vol. 4, edited by D. R. Fraser Taylor. Amsterdam: Elsevier

Cheng, Z., J. Caverlee, and K. Lee. 2010. "You Are Where You Tweet: A Content-based Approach to Geolocating Twitter Users", *Proceedings of the ACM Conference on Information and Knowledge Management*, Toronto, 759–768. New York: ACM.

Croitoru, A., A. Stefanidis, J. Radzikowski, A. T. Crooks, J. Stahl, and N. Wayant. 2012. "Towards a Collaborative GeoSocial Analysis Workbench." *COM-Geo*, Washington, DC: ACM.

De Choudhury, M., S. Counts, and M. Czerwinski. 2011. "Find Me the Right Content! Diversity-Based Sampling of Social Media Spaces for Topic-Centric Search." *Proceedings of the 5th international AAAI Conference on Weblogs and Social Media*, Barcelona, July 17–21.

Demchak, C., and P. Dombrowski. 2011. "Rise of a Cybered Westphalian Age." *Strategic Studies Quarterly* 5 (2): 32–61.

Diener, A. C., and J. Hagen. 2009. "Theorizing Borders in a 'Borderless World': Globalization, Territory and Identity." *Geography Compass* 3 (3): 1196–1216.

Dodge, M., and R. Kitchin. 2001. *Mapping Cyberspace*. London: Routledge.

The Economist. 2010. Social Networks and Statehood: The Future is Another Country. Accessed August 30, 2012. http://www.economist.com/node/16646000

Elwood, S. 2010. "Geographic Information Science: Visualization, Visual Methods, And The Geoweb." *Progress in Human Geography* 35 (3): 401–408.

Eriksson, B., P. Barford, J. Sommers, and R. Nowak. 2010. "A Learning-Based Approach for IP Geolocation", In *Passive and Active Measurement, Lecture Notes in Computer Science, Volume 6032*, edited by A. Krishnamurthy and B. Plattner, 171–180. Berlin: Springer.

Everard, J. 2001. *Virtual States: The Internet and the Boundaries of the Nation-State*. London: Routledge.

Fawcett, C. 1918. *Frontiers: A Study in Political Geography*. Oxford: Oxford University Press.

Flint, C. 2002. "Political Geography: Globalization, Metapolitical Geographies and Everyday Life." *Progress in Human Geography* 26 (3): 391–400.

Foucault, M. 1980. *Power/Knowledge. Selected Interviews and Other Writings 1972–1977*. New York: Pantheon.

Friedland, G., and R. Sommer. 2010. "Cybercasing the Joint: On the Privacy Implications of Geotagging." *Proceedings of the Fifth USENIX Workshop on Hot Topics in Security (HotSec 10)*, Washington, DC, August 10.

Goldmann, K. 2001. *Transforming the European Nation-State: Dynamics of Internationalization*. London: Sage.

Goolsby, R. 2010. "Social Media as Crisis Platform: The Future of Community Maps/Crisis Maps." *ACM Transactions on Intelligent Systems and Technology (TIST)* 1 (1): 7.

Greenhow, C., and B. Robelia. 2009. "Informal Learning and Identity Formation in Online Social Networks." *Learning, Media and Technology* 34 (2): 119–140.

Hecht, B., L. Hong, B. Suh, and E. Chi. 2011. "Tweets From Justin Bieber's Heart: The Dynamics of the 'Location' Field in User Profiles." *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*, Vancouver, BC, May 7–12.

Hermes, J. 2006. "Citizenship in the Age of the Internet." *European Journal of Communication* 21 (2): 295–309.

Herrera, G. J. 2007. "Cyberspace and Sovereignty: Thoughts on Physical Space and Digital Space", In *Power and Security in the Information Age: Investigating the Role of the State in Cyberspace*, edited by M. D. Gavelty, V. Mauer, and S. F. Krishna-Hensel, 67–94. Aldershot: Ashgate.

Holdich, T. H. 1916. *Political Frontiers and Boundary Making*. London: Macmillan.

Java, A., X. Song, T. Finin, and B. Tseng. 2009. "Why We Twitter: An Analysis of a Microblogging Community." In *Advances in Web Mining and Web Usage Analysis, Lecture Notes in Computer Science*, Vol. 5439, edited by R. Goebel, J. Siekmann, and W. Wahlster, 118–138. Berlin: Springer.

Jiang, M. 2010. "Authoritarian Informationalism: China's Approach to Internet Sovereignty." *SAIS Review* 30 (2): 71–89.

Khondker, H. H. 2011. "Role of the New Media in the Arab Spring." *Globalizations* 8 (5): 675–679.

Kitchin, R., and M. Dodge. 2011. *Code/Space: Software and Everyday Life*. Cambridge, MA: MIT Press.

Lang, R., and P. Knox. 2009. "The New Metropolis: Rethinking Megalopolis." *Regional Studies* 43 (6): 789–802.

Lerman, K., and R. Ghosh. 2010. "Information Contagion: An Empirical Study of the Spread of News on Digg and Twitter Social Networks." *Proceedings of the 4th International AAAI Conference on Weblogs and Social Media*, Washington, DC, 90–97. Menlo Park, CA: AAAI Press.

Lu, Y., and Y. Liu. 2012. "Pervasive Location Acquisition Technologies: Opportunities and Challenges for Geospatial Studies." *Computers, Environment and Urban Systems* 36 (2): 105–108.

Mathioudakis, M., and N. Koudas. 2010. "TwitterMonitor: Trend Detection Over the Twitter Stream." *Proceedings of the International Conference on Management of Data – SIGMOD'10*, Indianapolis, IN, 1155–1158. New York: ACM.

McCann, P., and Z. Acs. 2011. "Globalization: Countries, Cities and Multinationals." *Regional Studies* 45 (1): 17–32.

Negroponte, N. 1996. *Being Digital*. New York: Vintage Press.

Newman, D. 2006. "The Lines That Continue to Separate Us: Borders in Our 'Borderless' World." *Progress in Human Geography* 30 (2): 142–161.

Newman, D., and A. Paasi. 1998. "Fences and Neighbourhoods in the Postmodern World: Boundary Narratives in Political Geography." *Progress in Human Geography* 22 (2): 186–207.

O'Reilly, T. 2005. What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. Accessed February 20, 2009. http://www.oreillynet.com/lpt/a/6228

Paasi, A. 2009. "Bounded Spaces in a 'Borderless World': Border Studies, Power and the Anatomy of Territory." *Journal of Power* 2 (2): 213–234.
Poese, I., S. Uhlig, M. A. Kaafar, B. Donnet, and B. Gueye. 2011. "IP Geolocation Databases: Unreliable?." *Computer Communication Review* 4 (2): 53–56.
Quercia, D., L. Capra, and J. Crowcroft. 2012. "The Social World of Twitter: Topics, Geography, and Emotions." *Proceedings of the Sixth International Conference on Weblogs and Social Media*, Dublin, 298–305. Palo Alto, CA: AAAI Press.
Sparke, M. 2006. "Political Geography: Political Geographies of Globalization (2) – Governance." *Progress in Human Geography* 30 (2): 1–16.
Stefanidis, T., A. T. Crooks, and J. Radzikowski. 2012. "Harvesting Ambient Geospatial Information from Social Media Feeds." *GeoJournal* doi: 10.1007/s10708-011-9438-2.
Takhteyev, Y., A. Gruzd, and B. Wellman. 2012. "Geography of Twitter Networks." *Social Networks* 34 (1): 73–81.
Thiemann, C., F. Theis, D. Grady, R. Brune, and D. Brockmann. 2010. "The Structure of Borders in a Small World." *PLoS ONE* 5 (11): e15422.
Vance, J. E. 1977. *This Scene of Man: Role and Structure of the City in the Geography of Western Civilization*. Palatine, IL: Harper's College Press.
Viegas, F. B., M. Wattenberg, and J. Feinberg. 2009. "Participatory Visualization with Wordle." *IEEE Transactions on Visualization and Computer Graphics* 15 (6): 1137–1144.
Wood, D. 2010. *Rethinking the Power of Maps*. New York: Guilford Press.
Xie, H., L. Kulik, and E. Tanin. 2011. "Privacy-aware Collection of Aggregate Spatial Data." *Data and Knowledge Engineering* 70 (6): 576–595.