1: Introduction

The power of social media to serve as an effective and irrefutable information dissemination mechanism was demonstrated quite vividly during the Arab Spring events across North Africa and the Middle East in early 2011. Platforms like Twitter, Facebook, and YouTube were instrumental not only in reporting news from these events, but also in supporting the organization and coordination of activities that were part of these events.1

While this is widely considered to be a watershed moment for the use of social media in geopolitical events, it was not the first time that this happened. Twenty months before Arab Spring, in June 2009, social media platforms were used to broadcast to the world real-time information from the clashes in the streets of Tehran following the rigged Iranian presidential election, bypassing the state-imposed crackdown on crisis coverage.2

The Tehran unrest eventually fizzled out, but this experience served notice to the general public in the Arab world. It made them fully aware of the power of social media to bypass state-controlled news channels in order to bring their message to the attention of the rest of their countrymen, and to engage the world community, leading to the organized and purposeful use of social media during Arab Spring. The user-generated and Web-delivered content of social media is complementing established information sources to support open-source intelligence analysis.

The emergence of social media feeds as an intelligence source is presenting interesting challenges and opportunities to the geospatial intelligence community in particular. These feeds more often than not have some sort of geographic content, for example communicating the location from which a particular report is contributed, the geolocation of an image, or making a reference to a specific sociocultural hotspot. It is very interesting to observe for example that Fukushima and Cairo were among the top 10 trending topics of discussion in Twitter during the first six months of 2011, while events were unfolding at the Japanese nuclear plant and the Egyptian uprising.3

The geographic content of social media feeds represents a new type of geographic information. It does not fall under the established geospatial community definitions of crowdsourcing.

or volunteered geographic information, as it is not the product of a process through which citizens explicitly and purposefully contribute geographic information to update or expand geographic databases. Instead, the type of geographic information that can be harvested from social media feeds can be referred to as Ambient Geographic Information: it is embedded in the content of these feeds, often across the content of numerous entries rather than within a single one, and has to be somehow extracted. Nevertheless, it is of great importance as it communicates in real time information about emerging issues, and also provides an unparalleled view of the complex social networking and cultural dynamics within a society, capturing the temporal evolution of the human landscape. In this paper we present emerging analysis techniques to harvest geospatial intelligence from social media feeds, focusing particularly on Twitter, as a representative data source.

The article is organized as follows. In Section 2 we trace the emergence of social media, followed in Section 3 by a discussion on their geolocation content, and discuss harvesting social media feeds in Section 4. We follow with a discussion of case studies in Section 5, showcasing novel types of geospatial analysis that can be performed using ambient geospatial information. Finally, in Section 6 we offer our outlook assessment.

2: The emergence of social media

During the last few years and aided by the growth and evolution of Web 2.0 technologies, social media applications have emerged to facilitate interactive information sharing, interoperability, user-centered design, and collaboration on the World Wide Web. The term social media is typically used to refer to services like Facebook, Twitter, Flickr and YouTube, which enable the general public to communicate with their peers, sharing information with them instantly and constantly in an effortless and intuitive way. By bypassing the need for advanced computing skills to participate, and by enabling practically everybody to contribute, social media has revolutionized information contribution and dissemination through the Internet. This allowed the public at large, who were primarily information consumers in the Web 1.0 world, to become active contributors and disseminators of information, thus bypassing traditional media outlets (such as news organizations).

Today, social media applications thrive. 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 that Twitter has at least 100 million active users, logging in at least once a month, and 50 million users who do so daily. A year later, in October 2012, Twitter CEO Dick Costolo announced that these numbers had doubled, with the number of accounts reaching 400-500 millions, with 200 million among them considered regular users. As a measure of reference, if the online Twitter community were to be viewed as a country, a population of 200 million would make it the 5th most populous country in the world, on par with Brazil. However, it is not just Twitter that has a large user community: Facebook had reached 1.1 billion users in 2013, with 665 million of them using it daily. A population of 1.1 billion would have Facebook in a close race with China and India to be the world's most populous community. Extending beyond the English speaking world, QQ is a Chinese service for instant messaging, with over 800 million accounts, while Sina Weibo (a Chinese microblogging service) reported over 500 million users in 2012.

In addition to constantly increasing user communities, the amount of data released through social media applications is also increasing at very impressive rates.

---

Seven years after its 2004 launch, Flickr was hosting 6 billion photos uploaded by its user community in 2011, with over 3,000 photos uploaded to Flickr every minute and a 20 percent annual rate increase over the past few years. And while 6 billion photos is a very impressive number, it only reflects the estimated number of photos uploaded monthly to Facebook by its user community, bringing the total number of photos hosted by Facebook to nearly 100 billion. At the same time, 100 million active users are uploading daily an estimated 40 million images in Instagram, while every minute, Flickr users upload in excess of 3,000 images, and YouTube users upload approximately 72 hours of video.

While their content ranges in format from SMS-like messages limited to 140 characters (Twitter) to images (Flickr) and video (YouTube), these social media feeds share a common nature: they are real-time published expressions of a society's cultural and societal interests. Thus harvesting and analyzing their content can offer unparalleled insight on social-cultural dynamics. For example, they allow us to:

- map the manner in which ideas and information propagate in a society—information that can be used, for example, to identify appropriate strategies for information dissemination in crisis response;
- map people's opinions and reaction on specific topics and current events, thus improving our ability to collect precise cultural, political, economic and health data, and to do so at near-real-time rates; and
- identify emerging sociocultural hotspots.

This represents an evolution of the manner in which human landscape data can be collected and analyzed. While traditional human geography approaches are rather static (primarily due to the cumbersome strategies used to collect the necessary data, for example censuses), this emerging opportunity allows us to treat the human landscape as the living, breathing organism that it is: we can witness the explosion-like dissemination of information within a society, or the clusters of individuals who share common opinions or attitudes, and map the locations of these clusters. This is an unprecedented development that broadens drastically our current geointelligence capabilities.

In the remainder of this paper we focus on Twitter as a representative social media feed and discuss its potential use for geointelligence applications.

3: Geolocating Twitter feeds

Geolocation information in tweets can be provided directly by the contributing bloggers, if they decide to make this information available, or it can be deduced from IP addresses using any of the IP geolocation solutions. In this paper we are focusing 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 Figures 1 and 2, or in a descriptive manner (e.g., listing only a city name). It is typically 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, see). 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

---

11. http://www.allfacebook.com/infographic-facebook-will-have-100-billion-photos-this-summer-2011-02
it is static and does not necessarily reflect user location at the moment that the tweet was sent.

The reported percentage of geolocated tweets varies, from as high as two thirds or half of the tweets having some location information in the form of coordinates or description to as low as 5 percent of the users listing actual coordinates with another 21 percent listing descriptive geolocation information. In our own study earlier we had approximately 16 percent of our feeds with detailed coordinate location information, with another substantial percentage of the tweets having some geolocation information at coarser granularity (e.g., the city level). This variation in the rate of geolocated tweets can be attributed to a range of factors, such as geographic area, time, and theme.

For example, the seemingly high rate of 16 percent, which was observed following the 2011 tsunami and Fukushima disaster in Japan, can be attributed to the response of the Japanese population to a major event (e.g., evacuation), the increased use of mobile devices after the event, and the high penetration rate of Twitter in Japan (26.6 percent).

A recent study by Leetaru et al. using Twitter’s Decahose data stream has reported that even in the absence of a significant disruptive event...

---


(e.g., a nuclear disaster) the rate of geolocated tweets may vary depending on the time of day (from 2.3 percent at 1 p.m. PST to 1.7 percent at 6 a.m. PST) or the geographic location (from 2.86 percent in Jakarta to 0.77 percent in Moscow). Furthermore, it is worth mentioning that in the absence of other information, geolocation data can be obtained from IP addresses using any of the IP geolocation solutions. The accuracy of this geolocated information may range from building level all the way to broader neighborhood.

4: Harvesting social media feeds

Harvesting information from social media feeds involves 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.

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 PostgreSQL), thus creating a local mirror of the content of the original server for the entries specified by our queries. Depending on the subject, the queries may be periodic, or may be 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.

5. Case studies: turning Twitter content into geospatial intelligence
Tweet content is a prototypical description of human landscape dynamics, as it includes references to geographic entities that are sociocultural hotspots at the time of the reference. It can also be analyzed to identify social dynamics within the Twitter community, in the form of complex relations and hierarchical structures among its members. In this section we provide some sample analytical processes, in order to showcase the type of geointelligence information that can be harvested through Twitter feed analysis. The experiments we present are from a system we developed in-house to collect Twitter feeds using thematic and geographic queries, and store them in a local database where they are further analyzed.

5.1 Hotspot emergence
Hotspots emerge in Twitter traffic through notable increases in the references to specific terms. This is comparable to the identification of trends in Web search engines. For example, Google is tracking through its Google Trends facility the number of searches for various terms, and analyzes the data to identify spikes in these searches. Figure 3 shows the trends in searches for the term “Tahrir Square” in 2011. The horizontal axis shows time (over a period of 12 months) and the vertical shows frequency of searches for a specific term. As we can see in this simple chart, we can...
identify peaks, which indicate strong interest among the Google user community about Tahrir Square, most notably at the end of January 2011 when the anti-government Egyptian protesters started gathering there, and then again in November 2011 when pro-democracy demonstrations started again.

Trends analysis of Internet searches has been used in epidemiology to detect the outbreak of diseases (Polgreen et al., 2008; Carneiro and Mylonakis, 2009) and was also extended to address blog and Twitter content and news articles. Google has even set up a corresponding website to monitor flu outbreaks through an analysis of search word patterns. A comparable analysis can be applied to Twitter data, this time searching for references to particular keywords. In Figure 4 we show, for example, an analysis of Twitter feeds originating from Cairo (based on Twitter’s API location filtering) over

Figure 4
Twitter data with a Tahrir Square label, 2/25/11-3/09/11, overlaid on a map of the square. Hashtag statistics over time show how off the spike (marked by a circle) is from normal.
Source: Stefanidis et al., 2011.


identify peaks, which indicate strong interest among the Google user community about Tahrir Square, most notably at the end of January 2011 when the anti-government Egyptian protesters started gathering there, and then again in November 2011 when pro-democracy demonstrations started again.

Trends analysis of Internet searches has been used in epidemiology to detect the outbreak of diseases (Polgreen et al., 2008; Carneiro and Mylonakis, 2009) and was also extended to address blog and Twitter content and news articles. Google has even set up a corresponding website to monitor flu outbreaks through an analysis of search word patterns. A comparable analysis can be applied to Twitter data, this time searching for references to particular keywords. In Figure 4 we show, for example, an analysis of Twitter feeds originating from Cairo (based on Twitter’s API location filtering) over

Figure 4
Twitter data with a Tahrir Square label, 2/25/11–3/09/11, overlaid on a map of the square. Hashtag statistics over time show how off the spike (marked by a circle) is from normal.
Source: Stefanidis et al., 2011.


the period 2/25/11–3/09/11, to identify ones that were labeled with the hashtag “Tahrir Square.” The data are shown on the upper right-hand side of the figure: a chart shows the number of tweets per hour within a 10km radius from Tahrir Square (vertical axis) over the period 2/25/11–3/09/11 (horizontal axis). Tweets are grouped hourly, and we can identify a peak (marked by a red circle) corresponding to the period 07:00-08:00 local time (UTC+2) of 3/4/11. This is actually the morning of the day when the new prime minister of Egypt eventually addressed his people at the square. Thus, Twitter traffic analysis offers an effective means to receive advance notice of this event, as people started discussing it as soon as it became known.

5.2 Tracing information dissemination avenues and social network structure

While hotspot detection allows us to identify locations of interest by analyzing trends in references to them, social network analysis allows us to recognize the structure of complex social groups: who is connected to whom, either directly or via common links, and how persons are clustered in groups sharing common interests. This also leads to the identification of leaders and followers in social networks, and to mapping the manner in which information is disseminated within them.

In order to visualize this complex process, we collected Twitter data relating to the devastating Sendai (Tohoku-Fukushima) earthquake in Japan (3/11/11). The structure of social groups is manifesting itself through retweets and direct references. By aggregating such activities over a period of 14 hours (3/11/11, from 05:00 to 19:00) we see in Figure 5 the structure of this network. We can see some major disseminators of information identified through their Twitter names (e.g. NHK_PR, asahi_tokyo, etc.). The lines within the graph link users and show retweets: every time someone retweets (i.e., retweets) another user’s post, this person is added to the original user’s cluster. This is the typical pattern of information dissemination seen within social networks.

As is common in many complex networks, this network is highly skewed in the sense that the majority of nodes have a low degree of connectivity (star-like shapes in the graph) while there are a small number of nodes which have a high degree of connectivity: these can be considered as hubs of information dispersal and to some extent key actors in the social media sphere. For example, in Figures 5 and 6 we can identify NHK_PR, asahi_tokyo, and TobikoMX. The first is a national news organization, while the other two tweet mostly about local information in general, such as schools and metro services. This behavior follows power law patterns: a large number of tweeters only tweet infrequently, while a small number of tweeters tweet a lot. This behavior is consistent with observed blogosphere characteristics and comparable behavioral patterns observed in online forums or file sharing sites.

The network structure displayed in Figure 5 can be particularly informative in crisis situations, as it allows us to identify information dissemination routes and

23. Hashtags represent a bottom-up, user-generated convention for adding content (in a sense, metadata) about a specific topic, by identifying keywords to describe content. Thus they allow easy searching of tweets and trends. Sites such as http://hashtag.org/ monitor such trends from tweets and provide relevant statistics, but only over short periods of time.

24. Simply stated social network analysis (SNA) allows us to explore how different parts of a social system (e.g. people, organizations) are linked together. Moreover, it allows one to define the system’s structure and evolution over time (e.g. kinship or role-based networks). SNA is a quantitative methodology using mathematical graphs to represent people or organizations, where each person is a node, and nodes are connected to others via links (edges). Such links can be directed or undirected (e.g. friendship networks don’t have to be reciprocal).


affected communities, supporting management and response. It is also important for monitoring the society, as its variations over time reveal the evolution of the human landscape, with clusters formed and broken in response to various external factors.

Furthermore, by using the geolocation information of these tweets we can map the spatial footprint of the social clusters in our area of interest, moving from the nebulous social space of Figures 5 and 6 to the geographic one as shown in Figure 7. Figure 7 shows the spatial footprint of tweet-retweet pairs in Tokyo, captured on a random instance (6/13/2011 at 8 a.m. in this case). We can see the location of the original tweeter (marked as a sitting bird next to the "source author" tag), and a link to the location of the person retweeting the original message (marked as the flying bird next to the "reTwitter" tag). These links are indicators of sociocultural kinship between their start and end node. Therefore they can be considered as sample points for the identification of socioculturally similar neighborhoods, as the places around these nodes are reasonably expected to be crowded with people who live, think and vote like the node-corresponding tweeters (Bishop, 2008). This can lead to the establishment of a human landscape reference baseline, and the monitoring of its variations over time.
Figure 6
The social network of Figure 5 at different instances during a 12-hour period after the Fukushima earthquake, as information is broadcast from the major providers to its members. At the top we see how information disseminated primarily from NHK_PR is disseminated, while at the bottom we see activity related to feeds contributed by other key nodes.
Source: Stefanidis et al. 2011.
5.3 Impact area assessment following a natural disaster

In order to assess the value and use of Twitter data during natural disasters, we collected Twitter streams immediately after the 5.8 magnitude Mineral, VA earthquake of 8/23/2011. Figure 8 shows a plot of the origin locations of geolocated tweets referring to the earthquake during the first 60 minutes after the event. As we see the tweets are heavily clustered inside the impact area. It is worth mentioning that first reports of the event appeared on Twitter less than a minute after it happened, and we already had 1,000 Twitter reports of this earthquake within 5 minutes of the event. As these data were collected from a random 1 percent sample of Twitter content, one could reasonably anticipate that as many as 100,000 Twitter reports were made for this event within these first 5 minutes.\(^2\) Actually, as early as within 5-10 minutes we have observed a good delineation of the impact area using Twitter content, with the formation of identifiable dense clusters of geolocated tweets within it.

We observe, therefore, that Twitter content serves as a timely and fairly accurate reporting system for natural disasters, with humans acting as sensors that collect and report this information. Considering the growing importance of disaster response in the geointelligence community, this is a critical advantage of social media content analysis.

---

5.4 Real-time event monitoring
Social media can also aid with real-time monitoring of events, for example in Figure 9 we show geolocated tweets collected on 11/17/2011 referring to Occupy Wall Street events in New York City. This was the planned “Day of Action,” with a march across the Brooklyn Bridge. We show two different instances of these data. The top view in Figure 9, is from the afternoon of that date showing various tweets contributed from Manhattan as the protesters were marching from midtown towards downtown. The bottom view in Figure 9 shows geolocated tweets harvested later in the evening as the protesters were crossing the Brooklyn Bridge.29

Figure 9 vividly demonstrates two facts, which are rather crucial observations regarding the use of social media feeds to gather geospatial information. Firstly, we observe that Twitter is being used to provide real-time in-situ reports from events. In this particular situation we see that people (either protesters or bystanders) are using their cell phones or other mobile devices to tweet during the march. This is consistent with the observation made in Section 5.3 where we reported the robust use of Twitter to report the impact of an earthquake, but now applies to an event of longer duration than the quasi-instantaneous earthquake. Secondly, we observe that by harvesting this information we get an excellent overview of the activities in the ground, without deploying any local sensors. With locals acting as sensors and providing steady feeds in the form of tweets we can gain remotely valuable situational awareness.

6. Outlook
Social media platforms have provided the general public with an effective and irreversible mechanism to broadcast in real-time a variety of information, ranging from personal observations to commentaries on events of broader interest. With an already substantial and steadily increasing membership, platforms such as Twitter and YouTube serve as conduits of massive amounts of information, thus rapidly becoming essential components of open-source intelligence. The information communicated through such feeds conveys opinions, interests and links within its user community, thus revealing the complex structure of social networks.

However, this information is only partially exploited if one does not consider its

---

29. A video capturing these events and selected tweet content is available at [http://youtu.be/TPlOM5eXjk](http://youtu.be/TPlOM5eXjk). Additional examples are also provided at the website: [http://www.geosocial.gmu.edu](http://www.geosocial.gmu.edu).
Figure 9
Geolocated tweets referring to the “Day of Action” events of the Occupy Wall Street movement march to the Brooklyn Bridge on 11/17/2011. Top: the march of the protesters from Union Square towards downtown. Bottom: tweets as the protesters are moving across the bridge.
geographical aspect, such as the location from where a particular feed was contributed, or a tweet reference to a specific sociocultural hotspot. By harvesting geographic content from social media feeds we can transfer the extracted knowledge from the amorphous cyberspace to the geographic space, and gain a unique understanding of the human landscape, its structure and organization, and its evolution over time. This newfound opportunity signals the emergence of open-source geospatial intelligence, whereby social media contributions can be analyzed and mined to gain unparalleled situational awareness.

In this paper we presented a number of sample applications that demonstrate the geospatial intelligence value of information harvested from social media feeds. Twitter content reveals the emergence of sociocultural hotspots, and provides advanced warning of forthcoming events, as was the case with the Tahrir Square references during the Arab Spring events of spring 2011. It also offers a mechanism to obtain a rapid assessment of the impact area of natural disasters as demonstrated by data collected during the Virginia earthquake of August 2011. It provides unparalleled situational awareness by supporting the monitoring of evolving events, as was the case with the Occupy Brooklyn Bridge experiment. Furthermore, the Sendai experiment demonstrated how Twitter data analysis allows us to identify information dissemination routes, knowledge that can be very crucial when designing emergency response plans.

These examples demonstrate the fact that humans act as hybrid sensors when using social media platforms. Unlike typical sensors that always operate on specific bands of the spectrum, or collect specific types of measurements, humans operate across a wide range of the sociocultural spectrum, commenting in one message on a natural phenomenon, and in the next on a political issue. Thus the information they provide when blogging or posting a picture has substantial intelligence value.

One could argue though that the most important information collected by harvesting and analyzing social media content is the structure and spatial distribution of social networks, and the manner in which they evolve over time, reacting to news and events and adapting to the state of the world.

With the ability to collect such information in real-time we are now presented with an unprecedented opportunity to redefine the concept of human landscape or its various synonyms (e.g. human geography or human terrain). Our traditional approach to the problem of human terrain data collection was very static: collecting human geography data at distinct time instances through census-like campaigns, and then trying to interpret these datasets in order to identify clusters of similarly-behaving people. By harvesting information from social media feeds we are actually able to identify these relationships directly and to monitor them continuously.

Accordingly, harvesting open-source geospatial intelligence represents a transformation of our traditional operations that substantially improves our ability to analyze and understand sociocultural dynamics, and allows us to examine the human landscape as the living and evolving organism that it is.
Anthony Stefanidis is Professor in the Department of Geography and Geoinformation Science (GGS) at George Mason University (GMU), and director of Mason's Center for Geospatial Intelligence. He earned his PhD from The Ohio State University. His areas of expertise include image analysis, sensor networks and harvesting geospatial information from social media.

Andrew Crooks is Assistant Professor with the Department of Computational Social Science at GMU. Andrew's areas of expertise include GIS, creating and analyzing large spatial data sets and finding patterns and insights. He is a recognized expert on the integration of GIS and agent-based modeling.

Arie Croitoru is Assistant Professor in the GGS Department at GMU and member of Mason's Center for Geospatial Intelligence. He holds a PhD in Geoinformatics from Technion University (Israel). His research includes three interrelated research streams: geosocial analysis and location-based mass collaboration systems (LB-MCS), spatio-temporal data mining and digital image processing.

Jacek Radzikowski is Senior Researcher with Mason's Center for Geospatial Intelligence. He holds degrees in Computer Science from Warsaw University of Technology and Computational Science from Mason. Jacek's areas of expertise are database management and development, software design and system implementation.

Matthew Rice is Assistant Professor with the Department of Geography and Geoinformation Science at George Mason University. He conducts research in geospatial crowdsourcing, assistive geotechnology and geovisualization. Dr. Rice is a leader and contributor to the Association of American Geographers Cartography Specialty Group and the International Cartographic Association.
Social Media and the Emergence of Open-Source Geospatial Intelligence

Anthony Stefanidis\textsuperscript{1,3}, Andrew Crooks\textsuperscript{2}, Arie Croitoru\textsuperscript{1,3}, Jacek Radzikowski\textsuperscript{1}, and Matt Rice\textsuperscript{3}

\textsuperscript{1}Center for Geospatial Intelligence
\textsuperscript{2}Center for Social Complexity
\textsuperscript{3}Department of Geography and Geoinformation Science
George Mason University
4400 University Drive, MS 6C3
Fairfax, VA 22030
\{astefani; acrooks2; acroitor; jradziko; mrice4\}@gmu.edu

Abstract

The emergence of social media has provided the public with an effective and irrepressible real-time mechanism to broadcast information. The great popularity of platforms such as Twitter and YouTube, and the substantial amount of content that is communicated through them are making social media an essential component of open-source intelligence. The information communicated through such feeds conveys the interests and opinions of individuals, and reveals links and the complex structure of social networks. However, this information is only partially exploited if one does not consider its geographical aspect. Indeed, social media feeds more often than not have some sort of geographic content, as they may communicate the location from where a particular report is contributed, the geolocation of an image, or they may refer to a specific sociocultural hotspot. By harvesting this geographic content from social media feeds we can transfer the extracted knowledge from the amorphous cyberspace to the geographic space, and gain a unique understanding of the human landscape, its structure and organization, and its evolution over time. This newfound opportunity signals the emergence of open-source geospatial intelligence, whereby social media contributions can be analyzed and mined to gain unparalleled situational awareness. In this paper we showcase a number of sample applications that highlight the capabilities of harvesting geospatial intelligence from social media feeds, focusing particularly on Twitter as a representative data source.
1: Introduction

The power of social media to serve as an effective and irrepressible information dissemination mechanism was demonstrated quite vividly during the Arab Spring events, across North Africa and the Middle East, in early 2011. Platforms like twitter, facebook, and YouTube were instrumental not only on reporting news from these events (New York Times, 2011), but also supported the organization and coordination of activities that were part of these events (Pollock, 2011). While this is widely considered to be a watershed moment for the use of social media in geopolitical events, it was not the first time that this happened. Twenty months before Arab Spring, in June 2009, social media platforms were used to broadcast to the world real-time information from the clashes in the streets of Tehran following the rigged Iranian presidential election, bypassing the state-imposed crackdown on crisis coverage (Newsweek, 2009). The Tehran unrest eventually fizzled out, but this experience served notice to the general public in the Arab world. It made them fully aware of the power of social media to bypass state-controlled news channels in order to bring their message to the attention of the rest of their countrymen, and to engage the world community, leading to the organized and purposeful use of social media during Arab Spring. The user-generated and web-delivered content of social media is complementing established information sources to support open-source intelligence analysis.

The emergence of social media feeds as an intelligence source is presenting interesting challenges and opportunities to the geospatial intelligence community in particular. These feeds more often than not have some sort of geographic content, for example communicating the location from which a particular report is contributed, the geolocation of an image, or making a reference to a specific sociocultural hotspot. It is very interesting to observe for example that Fukushima and Cairo were among the top 10 trending topics of discussion in twitter during the first six months of 2011\(^1\), while events were unfolding at the Japanese nuclear plant and the Egyptian uprising.

The geographic content of social media feeds represents a new type of geographic information. It does not fall under the established geospatial community definitions of crowdsourcing (Fritz et al., 2009) or volunteered geographic information (Goodchild, 2007a), as it is not the product of a process through which citizens explicitly and purposefully contribute geographic information to update or expand geographic databases. Instead, the type of geographic information that can be harvested from social media feeds can be referred to as Ambient Geographic Information (AGI; Stefanidis et al., 2013): it is embedded in the content of these feeds, often across the content of numerous entries rather than within a single one, and has to be somehow extracted. Nevertheless, it is of great importance as it communicates in real-time information about emerging issues, and also provides an unparalleled view of the complex social networking and cultural dynamics within a society, capturing the temporal evolution of the human landscape. In this paper we present emerging analysis techniques to harvest geospatial intelligence from social media feeds, focusing particularly on twitter, as a representative data source.

\(^1\) http://blog.twitter.com/2011/06/200-million-tweets-per-day.html
The article is organized as follows. In Section 2 we trace the emergence of social media, followed in Section 3 by a discussion on their geolocation content, and discuss harvesting social media feeds in Section 4. We follow with a discussion of case studies in Section 5, showcasing novel types of geospatial analysis that can be performed using ambient geospatial information. Finally, in Section 6 we offer our outlook assessment.

2: The Emergence of Social Media

During the last few years and aided by the growth and evolution of Web 2.0 technologies (O'Reilly, 2005), social media applications have emerged to facilitate interactive information sharing, interoperability, user-centered design, and collaboration on the World Wide Web. The term social media is typically used to refer to services like facebook, twitter, flickr and YouTube, which enable the general public to communicate with their peers, sharing information with them instantly and constantly in an effortless and intuitive way. By bypassing the need for advanced computing skills to participate, and by enabling practically everybody to contribute, social media has revolutionized information contribution and dissemination through the Internet. This allowed the public at large, who were primarily information consumers in the Web 1.0 world, to become active contributors and disseminators of information and thus bypassing traditional media outlets (such as news organizations).

Today, social media applications thrive. 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 that twitter has at least 100 million active users, logging in at least once a month, and 50 million users who do so daily. A year later, in October 2012, twitter CEO Dick Costolo announced that these numbers had doubled, with the number of accounts reaching 400-500 millions, with 200 millions among them considered regular users. As a measure of reference, if the online twitter community were to be viewed as a country, a population of 200 million would make it the 5th most populous country in the world, on par with Brazil. However, it is not just twitter that has a large user community: facebook had reached 1.1 billion users in 2013, with 665 million of them using it daily. A population of 1.1 billion would have facebook in a close race with China and India to be the world's most populous community. Extending beyond the English speaking word, QQ is a Chinese service for instant messaging, with over 800 million accounts, while Sina Weibo (a Chinese micro-blogging service) reported over 500 million users in 2012.

In addition to constantly increasing user communities, the amount of data released through social media applications is also increasing at very impressive rates. Seven years after its 2004 launch, flickr was hosting 6 billion photos uploaded by its user community in 2011, with over 3,000 photos uploaded to flickr every minute and a 20%

---

1 http://www.telegraph.co.uk/technology/twitter/9945505/Twitter-in-numbers.html
2 http://investor.fb.com/releasedetail.cfm?ReleaseID=761090
3 http://blogs.wsj.com/chinarealtime/2013/03/12/how-many-people-really-use-sina-weibo/
annual rate increase over the past few years. And while 6 billion photos is a very impressive number, it only reflects the estimated number of photos uploaded monthly to Facebook by its user community, bringing the total number of photos hosted by Facebook to nearly 100 billion. At the same time, 100 million active users are uploading daily an estimated 40 millions of images in Instagram, while every minute, Flickr users upload in excess of 3,000 images (Sapiro, 2011), and YouTube users upload approximately 72 hours of video.

While their content ranges in format from SMS-like messages limited to 140 characters (Twitter) to images (Flickr) and video (YouTube), these social media feeds share a common nature: they are real-time published expressions of a society’s cultural and societal interests. Thus harvesting and analyzing their content can offer unparalleled insight on sociocultural dynamics. For example, they allow us to:

- map the manner in which ideas and information propagate in a society, information that can be used for example to identify appropriate strategies for information dissemination in crisis response;
- map people’s opinions and reaction on specific topics and current events, thus improving our ability to collect precise cultural, political, economic and health data, and to do so at near real-time rates; and
- identify emerging socio-cultural hotspots.

This represents an evolution of the manner in which human landscape data can be collected and analyzed. While traditional human geography approaches are rather static (primarily due to the cumbersome strategies used to collect the necessary data, for example censuses), this emerging opportunity allows us to treat the human landscape as the living, breathing organism that it is: we can witness the explosion-like dissemination of information within a society, or the clusters of individuals who share common opinions or attitudes, and map the locations of these clusters. This is an unprecedented development that broadens drastically our current geointelligence capabilities. In the remainder of this paper we focus on Twitter as a representative social media feed and discuss its potential use for geointelligence applications.

3: Geolocating Twitter Feeds

Geolocation information in tweets can be provided directly by the contributing bloggers, if they decide to make this information available, or it can be deduced from IP addresses using any of the IP geolocation solutions (see Eriksson et al., 2010; Poese et al., 2011). In this paper we are focusing 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 Figures 1 and 2, or in a descriptive manner (e.g. listing only a city name). It is typically harvested from Twitter using the capabilities provided by the communication protocol linking the client to it. For example,

---

7 http://www.allfacebook.com/infographic-facebook-will-have-100-billion-photos-this-summer-2011-02
8 http://instagram.com/press/
9 http://www.youtube.com/yt/press/statistics.html
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, see Doty and Wilde, 2010). 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.

Figure 1: Example of a geolocated tweet (screenshot from the twitter site) following the anti-government demonstrations of December 10, 2011 in Moscow. The message likens the events to the Decembrist anti-tsarist (anti-government) movement of 1825, reading: ‘Today the term was born Decembrists’. A photo accompanies the tweet, and the location from which this tweet was originated is shown on a map below.
The reported percentage of geolocated tweets varies, from as high as two thirds (Hecht et al., 2011) or half of the tweets having with 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 et al., 2010). In our own study earlier we had approximately 16% of our feeds with detailed coordinate location information (Stefanidis et al., 2011), with another substantial percentage of the tweets having some geolocation information at coarser granularity (e.g. the city level).

These variations in the availability of geolocation information tend to be geographic in nature as providing precise geolocational information is common practice when using mobile devices to post tweets. Accordingly, one can reasonably argue that precise geolocation information will be more frequently available in areas where the latest mobile technology is more easily and rapidly adopted.

Furthermore, it is worth mentioning, that in the absence of other information, geolocation data can be obtained from IP addresses using any of the IP geolocation solutions (see e.g. Eriksson et al., 2010). The accuracy of this geolocated information may range from building level all the way to broader neighborhood (see Poese et al., 2011).
4: Harvesting 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.

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), thus creating a local mirror of the content of the original server for the entries specified by our queries. Depending on the subject, the queries may be periodic, or may be 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.

5: Case studies: Turning Twitter Content into Geospatial Intelligence

Tweet content is a prototypical description of human landscape dynamics, as it includes references to geographic entities that are sociocultural hotspots at the time of the
reference. It can also be analyzed to identify social dynamics within the twitter community, in the form of complex relations and hierarchical structures among its members. In this section we provide some sample analytical processes, in order to showcase the type of geointelligence information that can be harvested through twitter feed analysis. The experiments we present are from a system we developed in-house to collect twitter feeds using thematic and geographic queries, and store them in a local database where they are further analyzed.

5.1 Hotspot Emergence
Hotspots emerge in twitter traffic through notable increases in the references to specific terms. This is comparable to the identification of trends in web search engines. For example, Google is tracking through its Google Trends facility the number of searches for various terms, and analyzes the data to identify spikes in these searches. Figure 3 shows the trends in searches for the term ‘Tahrir Square’ in 2011. The horizontal axis shows time (over a period of 12 months) and the vertical shows frequency of searches for a specific term. As we can see in this simple chart, we can identify peaks, which indicate strong interest among the Google user community about Tahrir Square, most notably at the end of January 2011, when the anti-government Egyptian protesters started gathering there, and then again in November, 2011, when pro-democracy demonstrations started again.

![Figure 3: Google Trends: a plot of the frequency of searches for the term ‘Tahrir Square’ in the first 11 months of 2011.](http://www.google.org/flutrends/)
Twitter feeds originating from Cairo (based on Twitter’s API location filtering) over the period 2/25/11 – 3/09/11, to identify ones that were labeled with the hashtag 11 ‘Tahrir Square’. The data are shown on the upper right-hand side of the figure: a chart shows the number of tweets per hour within a 10km radius from Tahrir square (vertical axis) over the period 2/25/11 - 3/09/11 (horizontal axis). Tweets are grouped hourly, and we can identify a peak (marked by red circle) corresponding to the period 07:00-08:00 local time (UTC+2) of 3/4/11. This is actually the morning of the day when the new prime minister of Egypt eventually addressed his people at the square. Thus, Twitter traffic analysis offers an effective means to receive advance notice of this event, as people started discussing it as soon as it became known.

Figure 4 serves as an example of how spikes in references to specific terms correlate to events taking place there. The data shown in this figure correspond to approximately 38,000 tweets with a Tahrir Square hashtag reference in that 2 week period, selected from among a total of 684,000 tweets from 40,000 persons with a Cairo label over that period. We have advance notice of this event, as people started discussing it as soon as it became known.

![Figure 4: Twitter data with a Tahrir Square label, 2/25/11–3/09/11, overlaid on a map of the square. Hashtag statistics over time show how off the spike (circled in red) is from normal (Stefanidis et al., 2011).](image)

---

11 Hashtags represent a bottom up, user-generated convention for adding content (in a sense, metadata) about a specific topic, by identifying keywords to describe content. Thus they allow easy searching of tweets and trends. Sites such as [http://hashtags.org/](http://hashtags.org/) monitor such trends from tweets and provide relevant statistics, but only over short periods of times.
5.2 Tracing Information Dissemination Avenues and Social Network Structure

While hotspot detection allows us to identify locations of interest by analyzing trends in references to them, social network analysis\textsuperscript{12} allows us to recognize the structure of complex social groups: who is connected to whom, either directly or via common links, and how persons are clustered in groups sharing common interests. This also leads to the identification of leaders and followers in social networks, and to mapping the manner in which information is disseminated within them.

\textbf{Figure 5:} Network cores and users (nodes) retweeting. The color of the edges matches the outgoing node (retweet), which is colored according to degree of importance (from blue for lowest, to green, yellow and red for highest) (Stefanidis et al., 2011).

\textsuperscript{12}Simply stated social network analysis (SNA) allows us to explore how different parts of a social system (e.g. people, organizations) are linked together. Moreover, it allows one to define the systems’ structure and evolution over time (e.g. kinship or role-based networks). SNA is a quantitative methodology using mathematical graphs to represent people or organizations, where each person is a node, and nodes are connected to others via links (edges). Such links can be directed or undirected (e.g. friendship networks don’t have to be reciprocal).
In order to visualize this complex process we collected *twitter* data relating to the devastating Sendai (Tohoku – Fukushima) earthquake in Japan (3/11/11). The structure of social groups is manifesting itself through retweets and direct references. By aggregating such of activities over a period of 14 hours (3/11/11, from 05:00 to 19:00) we see in Figure 5 the structure of this network. We can see some major disseminators of information identified through their *twitter* names (e.g. NHK_PR, asahi_tokyo, etc). The lines within the graph link users, and show retweets: every-time someone retweets (i.e. rebroadcasts) another user’s post, this person is added to the original user’s cluster. This is the typical pattern of information dissemination seen within social networks.

As is common in many complex networks, this network is highly-skewed (Barabasi and Albert, 1999), in the sense that the majority of nodes have a low degree of connectivity (blue star-like shapes in the graph) while there are a small number of nodes which have a high degree of connectivity: these can be considered as hubs of information dispersal and to some extent key actors in the social media sphere (see Asur and Huberman, 2010). For example, in Figures 5 and 6 we can identify NHK_PR, asahi_tokyo, and TokyoMX. The first is a national news organization, while other two tweet mostly about local information in general, such as schools and metro services. This behavior follows power law patterns (Newman, 2005): a large number of tweeters only tweet infrequently, while a small number of tweeters tweet a lot. This behavior is consistent with observed blogosphere characteristics (see Shi et al., 2007) and comparable behavioral patterns observed in online forums (see Zhang et al., 2007) or file sharing sites (see Adar and Huberman, 2000).

The network structure displayed in Figure 5 can be particularly informative in crisis situations, as it allows us to identify information dissemination routes and affected communities, supporting management and response. It is also important for monitoring the society, as its variations over time reveal the evolution of the human landscape, with clusters formed and broken in response to various external factors.

Furthermore, by using the geolocation information of these tweets we can map the spatial footprint of the social clusters in our area of interest, moving from the nebulous social space of Figures 5 and 6 to the geographic one as shown in Figure 7. Figure 7 shows the spatial footprint of tweet-retweet pairs in Tokyo, captured on a random instance (6/13/2011 at 8am in this case). We can see the location of the original tweeter (marked as a sitting bird next to the ‘source author’ tag), and a link to the location of the person retweeting the original message (marked as the flying bird next to the ‘retwitter’ tag). These links are indicators of sociocultural kinship between their start and end node. Therefore they can be considered as sample points for the identification of socioculturally similar neighborhoods, as the places around these nodes are reasonably expected to be crowded with people who live, think, and vote like the node-corresponding tweeters (Bishop, 2008). This can lead to the establishment of a human landscape reference baseline, and the monitoring of its variations over time.
Figure 6: The social network of Figure 5 at different instances during a 12-hour period after the Fukushima earthquake, as information is broadcasted from the major providers to its members. At the top we see how information disseminated primarily from NHK_PR is disseminated, while at the bottom we see activity related to feeds contributed by other key nodes.
5.3 Impact Area Assessment Following a Natural Disaster

In order to assess the value and use of twitter data during natural disasters we collected twitter streams immediately after the 5.8 magnitude Mineral, VA earthquake of 8/23/2011. Figure 8 shows a plot of the origin locations of geolocated tweets referring to the earthquake during the first 60 minutes after the event. As we see the tweets are heavily clustered inside the impact area. It is worth mentioning that first reports of the event in twitter appeared less than a minute after it, and we already had 1,000 twitter reports of this earthquake within 5 minutes of the event. As these data were collected from a random 1% sample of twitter content one could reasonably anticipate that as many as 100,000 twitter reports were made for this event within these first 5 minutes (Crooks et al., 2011). Actually, as early as within 5-10 minutes we have observed a good delineation of the impact area using twitter content, with the formation of identifiable dense clusters of geolocated tweets within it.
Figure 8: Geolocated tweets referring to the 5.8 magnitude Virginia earthquake of August, 2011 collected during the first 60 minutes after the earthquake. The star marks the epicenter. The blue circle outlines the perceived earthquake impact area.

We observe, therefore, that twitter content serves as a timely and fairly accurate reporting system for natural disasters, with humans acting as sensors that collect and report this information. Considering the growing importance of disaster response in the geointelligence community, this is a critical advantage of social media content analysis.

5.4 Real-Time Event Monitoring

Social media can also aid with real-time monitoring of events, for example in Figure 9 we show geolocated tweets collected on 11/17/2011 referring to Occupy Wall Street events in New York city. This was the planned ‘Day of Action’, with a march across Brooklyn Bridge. We show two different instances of these data. The first as can be seen to the left of Figure 9, is from the afternoon of that date showing various tweets contributed from Manhattan as the protesters were marching from midtown towards downtown. While on the right of Figure 9, we show geolocated tweets harvested later in the evening as the protesters were crossing Brooklyn Bridge.\(^\text{13}\)

\(^{13}\) A video capturing these events and selected tweet content is available at http://youtu.be/TarlOM6eXJk
Figure 9: Geolocated tweets referring to the ‘Day of Action’ events of the Occupy Wall Street movement march to the Brooklyn Bridge on 11/17/2011. Left: the march of the protesters from Union Square towards downtown. Right: tweets as the protesters are moving across the bridge.

Figure 9 demonstrates vividly two facts, which are rather crucial observations regarding the use of social media feeds to gather geospatial information. Firstly, we observe that twitter is being used to provide real-time in-situ reports from events. In this particular situation we see that people (either protesters or bystanders) are using their cell phones or other mobile devices to tweet during the march. This is consistent with the observation made in Section 5.3 where we reported the robust use of twitter to report the impact of an earthquake, but now applies to an event of longer duration than the quasi-instantaneous earthquake. Secondly, we observe that by harvesting this information we get an excellent overview of the activities in the ground, without deploying any local sensors. With locals acting as sensors and providing steady feeds in the form of tweets we can gain remotely valuable situational awareness.

5. Outlook
Social media platforms have provided the general public with an effective and irrepressible mechanism to broadcast in real-time a variety of information, ranging from personal observations to commentaries on events of broader interest. With an already substantial and steadily increasing membership, platforms such as twitter and YouTube, serve as conduits of massive amounts of information, thus rapidly becoming essential
components of open-source intelligence. The information communicated through such feeds conveys opinions, interests, and links within its user community, thus revealing the complex structure of social networks.

However, this information is only partially exploited if one does not consider its geographical aspect, such as the location from where a particular feed was contributed, or a tweet reference to a specific sociocultural hotspot. By harvesting geographic content from social media feeds we can transfer the extracted knowledge from the amorphous cyberspace to the geographic space, and gain a unique understanding of the human landscape, its structure and organization, and its evolution over time. This newfound opportunity signals the emergence of open-source geospatial intelligence, whereby social media contributions can be analyzed and mined to gain unparalleled situational awareness.

In this paper we presented a number of sample applications that demonstrate the geospatial intelligence value of information harvested from social media feeds. Twitter content reveals the emergence of sociocultural hotspots, and provides advanced warning of forthcoming events, as was the case with the Tahrir Square references during the Arab Spring events of spring 2011. It also offers a mechanism to obtain a rapid assessment of the impact area of natural disasters as demonstrated by data collected during the Virginia earthquake of August 2011. It provides unparalleled situational awareness by supporting the monitoring of evolving events, as was the case with the Occupy Brooklyn Bridge experiment. Furthermore, the Sendai experiment demonstrated how twitter data analysis allows us to identify information dissemination routes, knowledge which can be very crucial when designing emergency response plans. These examples demonstrate the fact that humans act as hybrid sensors when using social media platforms. Unlike typical sensors that always operate on specific bands of the spectrum, or collect specific types of measurements, humans operate across a wide range of the socio-cultural spectrum, commenting in one message on a natural phenomenon, and in the next on a political issue. Thus the information they provide when blogging or posting a picture has substantial intelligence value.

One could argue though that the most important information collected by harvesting and analyzing social media content is the structure and spatial distribution of social networks, and the manner in which they evolve over time, reacting to news and events, and adapting to the state of the world. With the ability to collect such information in real-time we are now presented with an unprecedented opportunity to redefine the concept of human landscape or its various synonyms (e.g. human geography or human terrain). Our traditional approach to the problem of human terrain data collection was very static: collecting human geography data at distinct time instances through census-like campaigns, and then trying to interpret these datasets in order to identify clusters of similarly-behaving people. By harvesting information from social media feeds we are actually able to identify these relationships directly and to monitor them continuously. Accordingly, harvesting open-source geospatial intelligence represents a transformation of our traditional operations that substantially improves our ability to analyze and understand sociocultural dynamics, and allows us to examine the human landscape as the living and evolving organism that it is.
References


Financial Times (2010), Coke Sees 'Phenomenal' Result from Twitter Ads, Financial Times (June, 25th), Available at http://www.ft.com/cms/s/0/6726ef4e-805a-11df-8b9e-00144feabdc0.html#axzz1OoiRiNWO [Accessed on 27th, June, 2011].


TomTom (2011), This is What We Really Do with Your Data, Available at http://www.tomtom.com/page/facts [Accessed on 27th June, 2011].

