New Methods for Representing and Interacting with Qualitative Geographic Information

Contract #: W912HZ-12-P-0334

Contract Period: July 1, 2014 – December 31, 2014

Principal Investigators:

Dr. Alan M. MacEachren, GeoVISTA Center, Penn State University

Dr. Anthony Robinson, GeoVISTA Center, Penn State University

Report on New Methods for Representing and Interacting with Qualitative Geographic Information, Stage 2: Task Group 4: Message-focused use case

Alexander Savelyev, Alan M. MacEachren, Scott Pezanowski, Morteza Karimzadeh, Wei Luo, Jonathan Nelson, and Anthony C. Robinson

<maceachren, savelyev, wul132, spezanowski, karimzadeh, jkn128, arobinson>@psu.edu

GeoVISTA Center, Department of Geography, The Pennsylvania State University
Submitted December 17, 2014

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
12-17-2014	Final	July. 1, 2014 – December 31, 2014

4. TITLE AND SUBTITLE

Report on New Methods for Representing and Interacting with Qualitative Geographic Information, Stage 2: *Task Group 4 – Message-focused use case*

5a. CONTRACT NUMBER:
W912HZ-12-P-0334

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Alexander Savelyev, Alan M. MacEachren, Scott Pezanowski, Morteza Karimzadeh, Wei Luo, Jonathan Nelson, and Anthony C. Robinson

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

PENNSYLVANIA STATE UNIVERSITY , THE
408 OLD MAIN
UNIVERSITY PARK PA 16802-1505

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

US Army Engineer Research and Development Center (ERDC)
Geospatial Research Laboratory
7701 Telegraph Road
Alexandria, VA 22135-3864

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; Distribution is unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This report documents research in the GeoVISTA Center at The Pennsylvania State University (PSU) to develop visual analytics tools for leveraging data from microblogs, with a specific focus on Twitter as a primary data source. The tools are implemented within SensePlace2, a web-based application that provides a visual interface and display methods for query and exploration of large repositories of microblog data. Our emphasis is on revealing the where, when, what, and who components of microblog data to support situational awareness in natural disasters and other emergency events. This report summarizes project research to date and focuses specifically on outcomes for Task 4 in this research, which is directed to adding capabilities for message-focused queries and visual analysis of the query results. This augments the existing place-, tweeter-, and social-focused methods and tools completed for Tasks 1, 2 & 3, respectively. The report begins with a brief overview of work completed for Tasks 1, 2, and 3 to highlight achievements of this project to date and set the context for adding message-focused capabilities. This overview is followed by a comprehensive review of work completed in Task 4, divided into four sections that report on: (a) the sources of spatial information that can be used to map spatial aspects of Twitter information diffusion, (b) system extensions to collect data on Twitter-related message propagation and to store and enable quick queries of this information, (c) description of new SensePlace2 functionality for depicting retweet connections, and (d) advances in system architecture developed and implemented to support quick access to the growing volumes of microblog data. Finally, we present conclusions and outline some future research challenges.

15. SUBJECT TERMS

geovisualization, visual analytics, social media, microblogs, cartography, qualitative geographic information, text analytics

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Douglas R. Caldwell
SAR	SAR	SAR	SAR	31	19b. TELEPHONE NUMBER *(include area code)* 703-428-3594

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

Abstract

This report documents research in the GeoVISTA Center at The Pennsylvania State University (PSU) to develop visual analytics tools for leveraging data from microblogs, with a specific focus on Twitter as a primary data source. The tools are implemented within SensePlace2, a web-based application that provides a visual interface and display methods for query and exploration of large repositories of microblog data. Our emphasis is on revealing the where, when, what, and who components of microblog data to support situational awareness in natural disasters and other emergency events. This report summarizes project research to date and focuses specifically on outcomes for Task 4 in this research, which is directed to adding capabilities for message-focused queries and visual analysis of the query results. This augments the existing place-, tweeter-, and social-focused methods and tools completed for Tasks 1, 2 & 3, respectively.

The report begins with a brief overview of work completed for Tasks 1, 2, and 3 to highlight achievements of this project to date and set the context for adding message-focused capabilities. This overview is followed by a comprehensive review of work completed in Task 4, divided into four sections that report on: (a) the sources of spatial information that can be used to map spatial aspects of Twitter information diffusion, (b) system extensions to collect data on Twitter-related message propagation and to store and enable quick queries of this information, (c) description of new SensePlace2 functionality for depicting retweet connections, and (d) advances in system architecture developed and implemented to support quick access to the growing volumes of microblog data. Finally, we present conclusions and outline some future research challenges.

SensePlace2: Geovisual Analytics for Microblog Data

This report details progress on a multi-stage visual analytics research project focused on leveraging place-relevant data from microblogs through methods that support queries and visual exploration of the unstructured data they produce. The methods are implemented in SensePlace2, a web application that collects Twitter data for selected keywords and Twitter IDs; processes and indexes those data; and provides interactive visual tools that support queries focused on where, when, who, and what, as well as interrelationships among this information.

SensePlace2 is designed to help users leverage data from Twitter or other short message sources to support information foraging and situation awareness. A key contribution of SensePlace2 is its focus on revealing multiple aspects of geography, including the locations where tweets originate, the locations listed in profiles of the tweeters, and the places mentioned in the messages themselves. For the latter two, SensePlace2 uses named-entity recognition and geoparsing methods to identify and geolocate the place references, leveraging the GeoTxt application we have developed under separate funding (Karimzadeh et al., 2013; Wallgrün et al., 2014).

Early development work on SensePlace2 is presented in MacEachren, et al (2011). Further advances are outlined in Savelyev (2013), which introduces the client-side component coordination mechanism we have implemented to support a multi-view approach to foraging for and sensemaking of information that contains place, time, and attribute components. In Savelyev and MacEachren (2014), we introduce heterogeneous network modeling methods in SensePlace2 for exploring relationships in

geo-located social media data. A public version of the system, which includes a subset of the initial capabilities, is available (http://www.geovista.psu.edu/SensePlace2/public/) and a detailed explanation of the core components of the interface is available as a downloadable user guide (http://www.geovista.psu.edu/SensePlace 2/SensePlace 2_Interface_Mini_UserGuide.htm).

To provide context for our presentation of the Task 4 research, we begin with a brief overview of outcomes from Tasks 1-3. Task 1 focused on a place-focused use case to develop new methods for highlighting the places people talk about in tweets and linkages between place name mentions (MacEachren et al., 2013b). Figure 1 illustrates the basic visual information foraging interface developed during work on this task (and available at the public URL listed above).

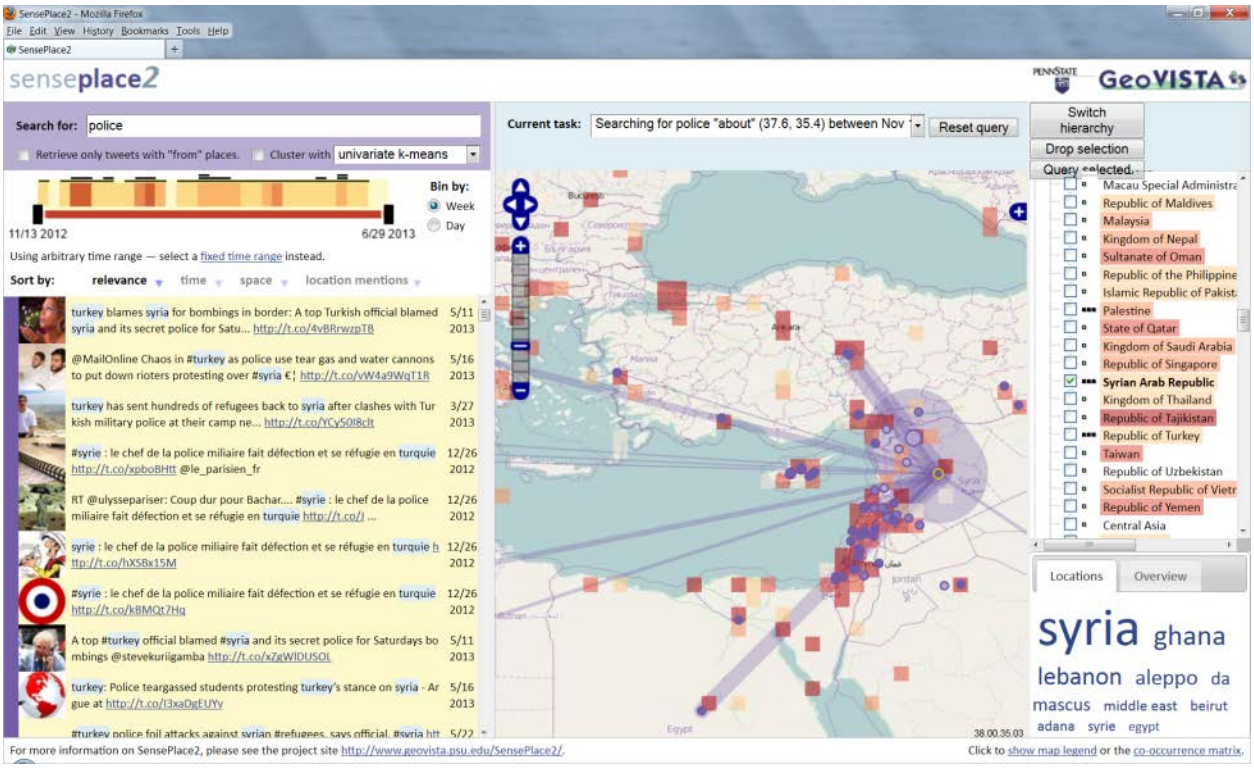

Figure 1. An example of the co-reference function on the SensePlace2 map as implemented for Task 1. For any tweet that mentions more than one place, this function links the places mentioned. The thickness of the connecting link represents the relative number of tweets that mention the connected places within the same tweet, out of the 1000 most relevant tweets that match the query. In this case, results shown are based on a spatial point location query for "police" with the point in Syria; the query retrieves the 1000 tweets mentioning police that are closest to the point specified by the user.

Task 2 investigated the tweeter-focused use case (MacEachren et al., 2013a). Techniques were developed for SensePlace2 analysts to explore individual Twitter users and the extent to which those users talk about or tweet from particular places. The primary addition to SensePlace2 in this phase of research was a sortable view that depicts multiple attributes associated with Twitter IDs. This view enables users to sort on the number of followers, places mentioned, geolocated tweets posted, hashtags, people mentioned, and entities (e.g., organization names) for Twitter users. This view is dynamically linked to the primary SensePlace2 interface and it can be used to launch new queries for all

tweets by Twitter ID(s) of interest. Figure 2 shows a representative use case through which a locally-produced photo is found from a user talking about the aftermath of Typhoon Yolanda in the Philippines.

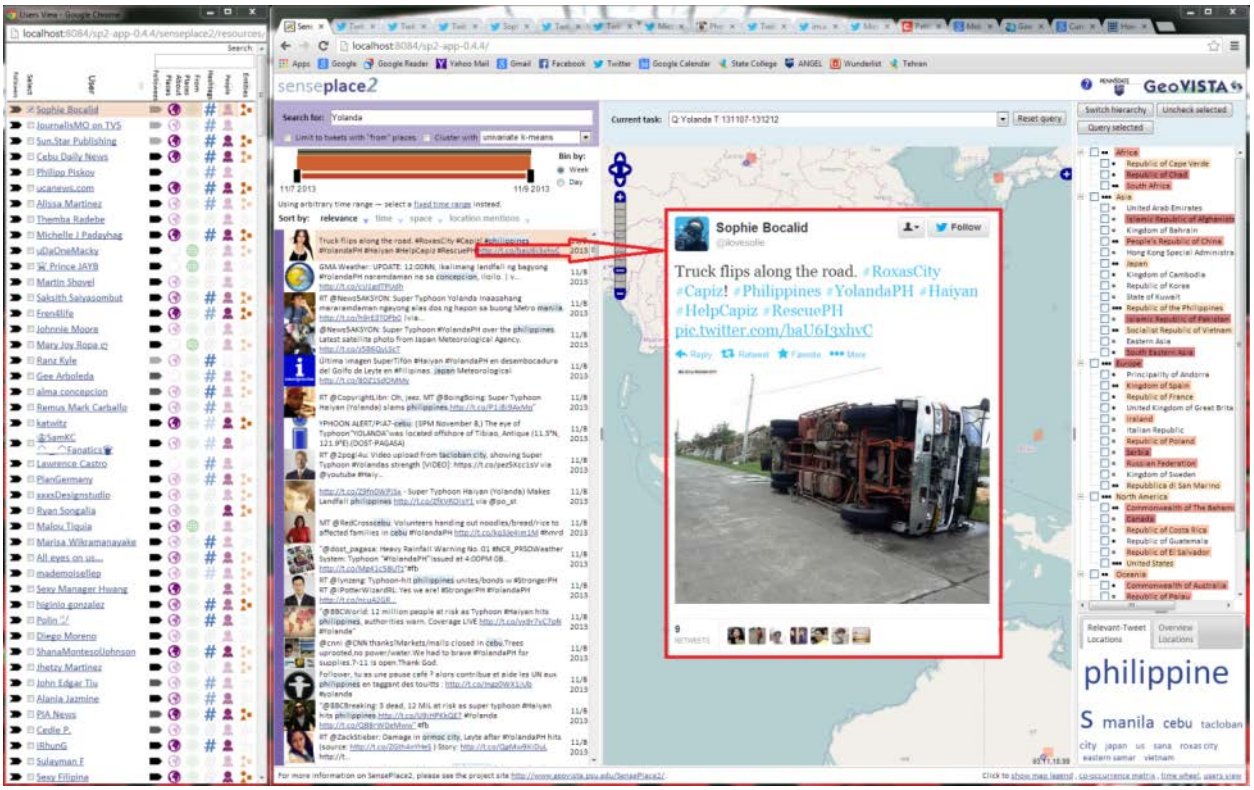

Figure 2. The main view depicts the initial result of a query for "Yolanda." Tweeters with a small number of followers and friends who mention many locations, people, entities and hashtags may pass along useful information that is missing from official news media. The TTable is used here to identify some of these tweeters. The particular Twitter ID selected is one with a lower number of followers, but with a picture and hashtags. Collecting a set of such locally-produced images can provide important situational awareness related to local impacts and reactions.

Task 3 extended SensePlace2 to explore a social-focused use case to forage for information generated by individual Twitter users and networks of users in their space-time context (MacEachren et al., 2014). Three components were added to the system: (a) GBuilder, a tool to build groups of Twitter IDs for which both tweets and interconnections among IDs are retrieved from the Twitter API, (b) ForceNet, a node-link graph, using a force-directed layout algorithm to distribute the nodes, that enables exploration of connections among nodes representing IDs, mentioned IDs, mentioned places, and hashtags, and (c) HivePlot, an alternative view for depicting connections that is conceptually like a parallel coordinate plot with radial rather than parallel axes. Figure 3 shows the application of the ForceNet tool to visualize connections among Twitter IDs, ID mentions, place mentions, and hashtags. These attributes are associated with a query on "attack," filtered to focus on a specific place (Benghazi, Libya).

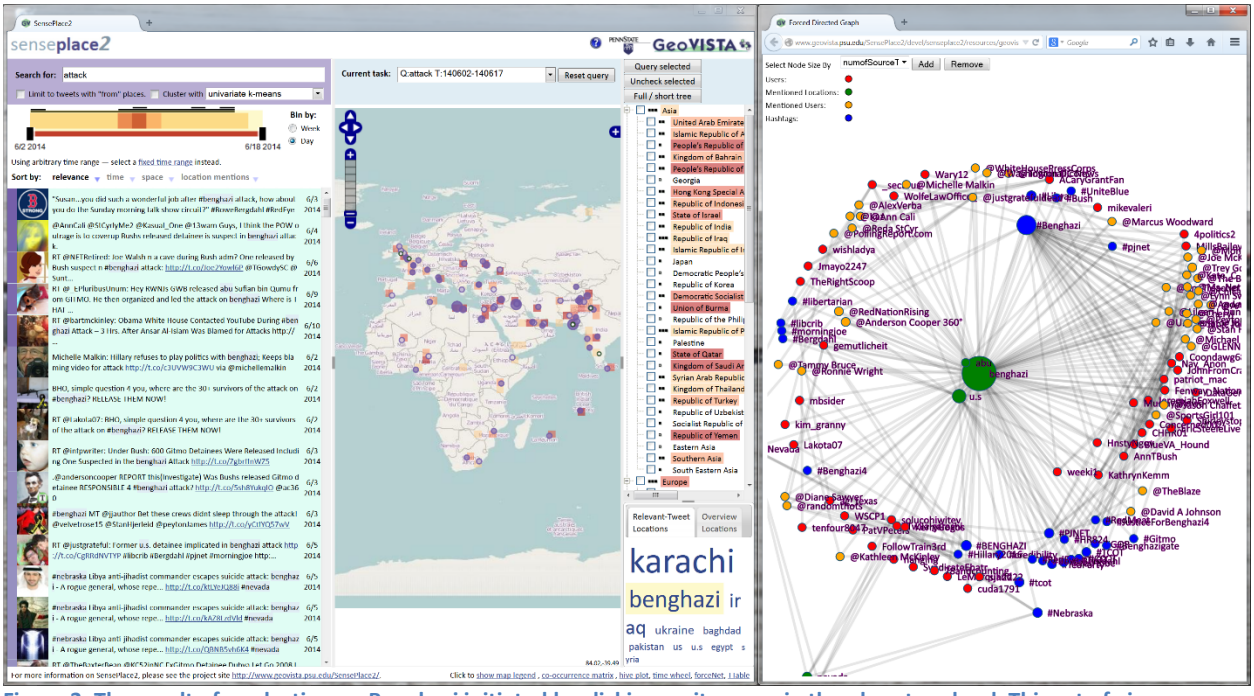

Figure 3. The result of a selection on Benghazi initiated by clicking on its name in the place tag cloud. This set of views provides quick access to all tweets in the time period that mention Benghazi (in the scrolling list at the left) and depicts links of Benghazi mentions with all other places, users, user mentions, and hashtags associated with those tweets (in the ForceNet view). Clicking on any node drills-down to the subset of nodes it is linked to (by fading all other nodes and links into the background where they provide context for interpretation.

1 Introduction to Task 4: The Message-focused use case

Here, our focus is on extending SensePlace2 to explore geographically-grounded message connections in Twitter. The initial focus of our work in Task 4 was on exploring retweets. Twitter data, however, contains traces of at least two other modes of information diffusion – @mentions of individual users and @replies to individual tweets. The @mentions consist of a Twitter user name prefixed by an @ sign and can be inserted anywhere in the body of the tweet. The @replies appear similar to @mentions in that they contain a reference to a Twitter user prefixed by an @ sign, however, they follow a more rigid format with the user mention always positioned at the beginning of the tweet. Importantly, Twitter keeps track of which specific tweet a given reply is directed toward.

Given these three potential modes of information diffusion, we envision three core scenarios of information diffusion on Twitter: a retweet of an existing tweet, an @mention of a user in a tweet, and an @reply to a tweet. Each scenario has a different semantic meaning and a different set of potential spatial properties that we can reveal and explore.

From a semantic point of view, retweets are primarily used to share an existing tweet with one's followers. Mentions, on the other hand, can be used for a number of purposes (e.g. to forward a piece of information to someone, initiate a conversation with a person, invite a person to an on-going conversation, etc.). Mentions are a very rich source of information, as they are likely to add additional content that goes well beyond a simple retweet indication that someone finds a tweet by someone else

interesting. Replies are primarily used to maintain a multi-way conversation, with the potential for multiple branches in the conversation forming, often featuring users joining and leaving the thread at different times.

It is possible that any information diffusion process will include a combination of these modes (e.g. a series of back and forth @replies discussing a particular retweet, with new people added to the conversation through @mentions). The spatial properties of the information diffusion process, which are different for each of these modes, are described in **Section 2**. The Twitter API capabilities and limitations are described in **Section 3.** Our prototype tool, which is designed to help analysts explore the information diffusion process, is described in **Sections 4 and 5**. **Section 6** contains conclusions, lessons learned, and ideas for future research.

2 Spatial properties of information diffusion on Twitter

Three sources of spatial data currently available for mapping the spatial aspects of the Twitter information diffusion process are the Twitter user profile locations, "about" locations (place mentions within the tweet text), and "from" locations (coordinates indicating the place from which the tweet was posted).

Profile locations are derived on a per-tweet basis. Each tweet contains all the Twitter user metadata along with the tweet metadata. The user metadata contains the self-reported "home location" in the user's profile. If a user changes their profile location at any point, this will be recorded in their next tweet, processed, and stored. This allows us to link any particular tweet or retweet to the profile location that existed at the time of the tweet as well as to capture a temporal record of the user's profile location if it changes.

The home location in the user's profile may be free-form text (some users insert a coordinate pair). Therefore, we use GeoTxt first as an entity extractor to find and geocode possible locations. If multiple locations are found, the first location is used. This location name and its coordinates are stored as added metadata to the tweet in both the database storage and within our Solr index. For our sample of tweets, which were collected from Twitter using event-relevant query terms (e.g., terms associated with natural disasters, airline incidents, protests, etc), approximately 57% of users have a non-null home location. However, only a subset of these locations represents real place names that are possible to geolocate. Overall, approximately 31% of users in our sample have a home location in which a geographic place name can be found and geocoded.

As with profile locations, "about" locations are derived on a per-tweet basis through the process of entity extraction and geocoding using GeoTxt. "From" locations are reported by Twitter directly and come from multiple sources – GPS coordinates embedded when sending a tweet from a GPS-enabled device, user-reported locations from a fixed set of places provided by the Twitter Place API, and other less common sources (e.g. satellite coordinates for weather-related posts from NOAA, pre-determined locations of fixed flood gauges, etc).

The first two sources of spatial information, user profile locations and "about" place mentions, have substantially higher location uncertainty than "from" locations. Part of this uncertainty is explained

by the limited success in extracting and geocoding the place names in both the profile location fields and in the body of the tweet; part stems from the fact that neither of the two sources is regulated in terms of spatial resolution and veracity (i.e., users may use a spatial reference as vague as "US" or as specific as "123 Main St, Chicago, IL"; there is nothing that prohibits them from listing a false location; and the location may not reflect the "from" location); and part stems from the uncertainty in temporal relevance i.e., the profile location may not be current.

"From" locations are a reliable source of information about the location of an individual at the time a tweet was posted, but they are only present with a small proportion of tweets, generally estimated at around 1.5%. Plus, the location of an individual at a particular instant may or may not be relevant to the content of their tweet. In our current dataset, the prevalence of tweets having "from" locations varies depending on the topic of interest. Based on a sample of more than 200 million tweets, we found a mean of 2.3% with a "from" location; but these varied by tweet topic, from 0.35% for tweets containing the term "protest", through 1.3 and 1.8% for tweets containing the terms "fire" and "tornado", respectively, to 4.5% and 7.6% for tweets containing the terms "flood" and "earthquake," respectively (MacEachren et al., 2013b).

The three sources described above are not equally prominent in different scenarios of information diffusion. Retweets are a poor source of "about" and "from" locations. This is because they are identical to each other in textual content and Twitter only associates the posting location with the original tweet if the tweeter has location enabled. Twitter does not associate the posting location with the retweets regardless of whether or not the user has location enabled. Retweets can be geographically analyzed on the basis of the Tweeter's profile location when one is present.

In contrast, @mentions can be a rich source of all three types of locations, while @replies can contain all three types of locations and also provide context to link the three types together as part of a geo-centric conversation. This potentially increases the accuracy of place mention extraction and geocoding with @replies and makes them a semantically-rich source of spatial information. Although SensePlace2 makes use of all three sources of spatial information, the process to fuse these sources has not yet been automated. (See Section 6.3 for further discussion.)

3 Twitter data collection and representation in SensePlace2

This section provides details on the extensions implemented in SensePlace2 to collect data on Twitter-related message propagation and to store and enable quick queries of this information. We also provide statistics on the relative prevalence of different forms of message propagation.

3.1 Data collection

The SensePlace2 extensions implemented to meet Task 4 objectives make extensive use of the "follow" endpoint of the Twitter Streaming API to focus on a number of individual users designated through the SensePlace2 Gbuilder Tool. This Streaming API endpoint provides all tweets from specified users (including retweets initiated by those users and @replies they made), as well as all retweets of their tweets and all @replies their tweets receive (both automatic and manual, i.e. tweets manually

prefaced with @mention of that user). Manual retweets (tweets manually prefaced with an "RT @" string) as well as arbitrary @mentions are *not* returned by the Twitter Streaming API.

There are two key limitations of the "follow" endpoint of the Twitter Streaming API. First, a maximum of 5000 users can be followed through a single connection to the Twitter API. Second, only first-degree retweets are returned as part of the query results (i.e. retweets of retweets will not be returned). The overall rate of the data collected is limited in the same way the rest of the Streaming API endpoints are limited – as long as the volume of query results is less than 1% of total Twitter traffic, all data matching the request parameters will be returned.

Overall, the conditions described above put us in an excellent position for data collection in small, pre-defined communities of organizations or individuals. A system that would track diffusion of information across ill-defined, evolving or very large groups of users would be challenging to develop. These challenges are further elaborated on in **Section 6.3**.

Task 4 also makes use of the Twitter REST API to collect information on the networks of followee-follower relationships for small groups of users. After a SensePlace2 analyst picks a number of user accounts through our Gbuilder Tool, we can also acquire corresponding Twitter User IDs, and their follower-followee relationships. These follower-followee relationships provide us with another way to collect information diffusion information, because followees are likely to be retweeted, replied to, and mentioned by their followers. Given the Twitter API rate limitation (of 180 ID queries within 15 minutes), we can only collect such information for small and well-defined communities.

We have compiled a range of statistics that describe the availability of information diffusion information, as made available by Twitter and our processing tools. A representative sample of the statistics is displayed below, describing the month of July 2014. In this table, the cell count values represent the number of tweets that intersect the column and row descriptions. The "All" column and row report overall counts for the rows or columns and the other columns represent subsets of these totals (or their union in the case of columns with a "+"). For example, the first row shows all of the tweets with retweets and the first column is all of the tweets with an "about" location. Therefore, there are 1,669,700 tweets out of 41,908,640 total that are both a retweet and have an "about" location. The second part of the table depicts percentages of the relevant totals. The row and column labels are:

RT = retweet
@ = user mention
# = hashtag
About = tweet "about" location
From = tweet "from" location
Profile = user profile location that is not empty.

	About	From	Profile	A+F	A+P	F+P	All	
RT	1,669,700		8,942,577		1,070,195		15,199,829	
@	2,161,376	277,784	13,400,386	23,988	1,421,055	200,954	22,092,746	
#	1,231,055	131,404	5,423,746	21,333	827,741	105,083	8,707,015	
RT+@	1,669,697		8,942,525		1,070,058		15,199,739	
RT+#	573,927		2,457,122		367,701		3,964,430	
@+#	672,709	33,043	3,149,152	3,856	440,573	26,701	4,971,056	
All	3,801,795	966,547	23,892,093	84,534	2,520,415	664,889	41,908,640	total

	About	From	Profile	A+F	A+P	F+P	All
RT	4.0%		21.3%		2.6%		36.3%
@	5.2%	0.7%	32.0%	0.1%	3.4%	0.5%	52.7%
#	2.9%	0.3%	12.9%	0.1%	2.0%	0.3%	20.8%
RT+@	4.0%		21.3%		2.6%		36.3%
RT+#	1.4%		5.9%		0.9%		9.5%
@+#	1.6%	0.1%	7.5%	#REF!	1.1%	0.1%	11.9%
All	9.1%	2.3%	57.0%	0.2%	6.0%	1.6%	100.0%

Table 1. Statistics for retweets (RT), place mentions (About), @ mentions, their pairs, and overall tweets in a sample for July 2014. The lower half of the table depicts percentages of the relevant totals.

3.2 Data representation in the system

The data used by searches initiated through the SensePlace2 web application are stored in an Apache Solr index. One of the more powerful functions Solr provides is a faceted (filtered) query. This enables analysts to systematically drill down on subsets of data and on subsets within these subsets. For example, we can select all tweets that mention police, then identify within these all that have been retweeted 20 or more times, and finally filter these to show those that contain hashtags.

SensePlace2 searches the Solr index through HTTP requests. The faceted query HTTP request parameter is used to drill-down through the results. Solr can accept as many filter query parameters as needed. For example, a SensePlace2 user may search for all tweets containing @mentions. Next, they filter by all tweets containing @mentions from User A. Next, they select all tweets containing @mentions from User A that mention a place. Finally, they go beyond those tweets that mention any place to focus on the places that are located within a bounding box drawn on the map.

For Task 4, the updated client data model incorporates all three modes of information diffusion – retweets, @mentions and @replies. This is accomplished by extending the network data model built as part of Task 3 (Savelyev and MacEachren, 2014) to include relationships of types "*retweets / retweetOf*" and "*replies / replyToTweet*" between tweets and "*userMentions / mentionedIn*" between tweets and users. The new relationship types are fully integrated with previous node and relationship types. Further details on the implementation and capabilities of the network data model in the SensePlace2 client can be found in Savelyev and MacEachren, 2014.

4 Geovisualization of information diffusion in SensePlace2

This section presents some of the SensePlace2 functionality for depicting retweet connections. Capabilities have been implemented in the SensePlace2 interface to enable analysts to select either tweets or Twitter IDs (for individuals and groups) and see the geographical 'footprints' for the result. The footprint depicts the profile location of the original tweeter and those for retweeters who have a geolocatable profile location.

4.1 Overall framework

In order to make spatial context available on the map, we have designed a prototype information diffusion visualization tool (Figure 4) that builds on the existing SensePlace2 infrastructure to help reveal connections between locations relevant to retweets.

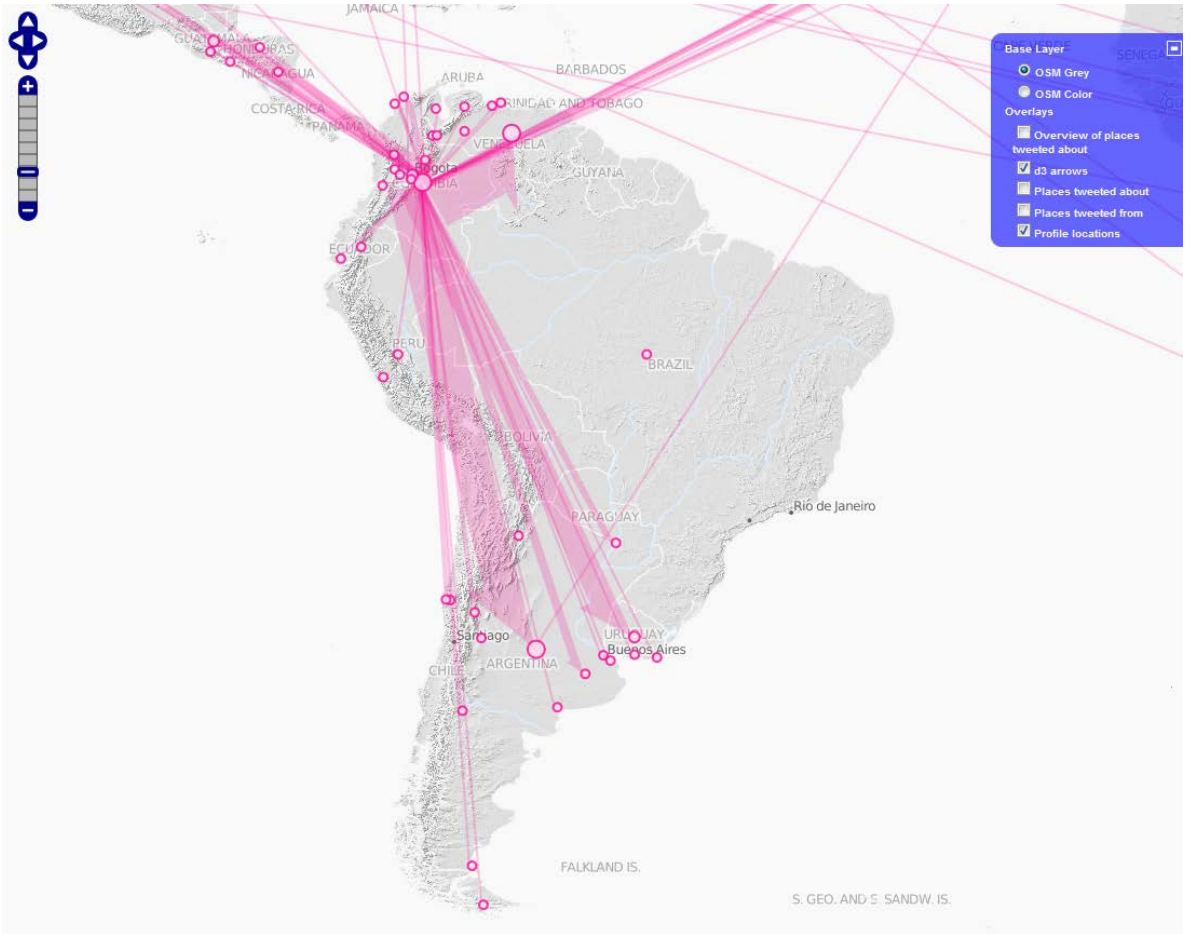

Figure 4. Information propagation among reported profile locations for a sample of 100 frequently retweeted tweets.

Figure 4 shows the spatial information diffusion process through user profile locations, shown as point symbols of varying size (in deep pink), and displays the intensity of information propagation through arrows of varying widths that connect related user profile locations. Corresponding with the use of symbol size to depict counts in earlier versions of SensePlace2, point symbols are scaled according to

the number of people reporting that particular place as their profile location, and arrows are scaled to display the total volume of retweets, @mentions, and @replies exchanged between all profile locations. The direction of arrows depicts the direction of information transmission. For retweets, the arrows are drawn from the profile location of the original tweet to the profile location of the Twitter ID that retweeted it.

The total count of connections between profile locations is calculated and the widths of the arrows are resized accordingly. The current ratio is 1px of width for every retweet. Arrows can be drawn both ways, between every pair of profile locations shown in the "Profile locations" layer, if information is flowing in both directions. If there is only one arrow pointing to or away from a profile location, this indicates that the diffusion flows in only that direction. If the source tweet and its retweets share the same profile location, no lines will be drawn. A limitation of the current version of the information diffusion tool is that the system does not yet include symbolization to depict these place-recursive messages.

The specific information diffusion map in Figure 4 was generated using a test dataset compiled in the following fashion. From our entire database of tweets, the 100 most retweeted tweets that come from users with a geocoded profile location were selected. For each of those tweets, the most recent 10 retweets that were initiated by users with geocoded profile locations were selected (in this example, we limit to 10 of each to keep the retrieved result manageable).

Figure 4 also serves as a direct visual representation of a slice of the SensePlace2 network data model – the model used to store data in the SensePlace2 User Interface (UI). The data model functions by creating links from one profile location to others. To do this, we first identify all users that specify a given profile location as their home. Second, we look at all the tweets sent by those users, and at all the retweets of those tweets. Third, we identify the authors of the retweets and compile the list of profile locations they report as their home. Finally, we draw connecting arrows between the given profile location and the profile locations identified in the previous step. The process described above is easy to extend to visualize other aspects of the information diffusion process. For example, we can focus on mapping the locations of followers instead of users who retweet a particular message. SensePlace2, in its current iteration, does not rely on the follower/followee information as part of the diffusion visualization process due to the limited size of the follower/followee networks that can be constructed given the rate limitations of the Twitter Search API (described in **Section 3.1**).

To highlight the temporal aspect of the information diffusion process, the display shown in Figure 4 can be animated through the use of a simple animation control located at the bottom right corner of the SensePlace2 interface. When applied, the animation will visually represent the information diffusion process over time. First, the diffusion layer is cleared. Next, arrows are added according to the order and the relative timing of the diffusion process. For example, if a tweet from Chicago was retweeted in New York, and then in San Francisco, the arrow from Chicago to New York will be drawn first, and the arrow from Chicago to San Francisco will be added later (with delay proportional to the original time difference between those retweets). The thickness of the arrows is animated over time as well. For example, if a tweet from Chicago was retweeted in New York, then in San Francisco, and then in New York again, the

arrow from Chicago to New York will be first drawn with a thickness of 1 pixel, then expanded to 2 pixels to signify another retweet in New York.

4.2 Scenario 1: spatial information diffusion footprint of an individual tweet

Spatial information diffusion can be rendered using the new map tool described in **Section 4.1** above. As a simple, case study analytical example, Figure 5 depicts profile locations (in deep pink) of Twitter IDs that retweeted a message by *The Economist* (top tweet in TweetList) about potential bias of the International Criminal Court in relation to cases in Africa.

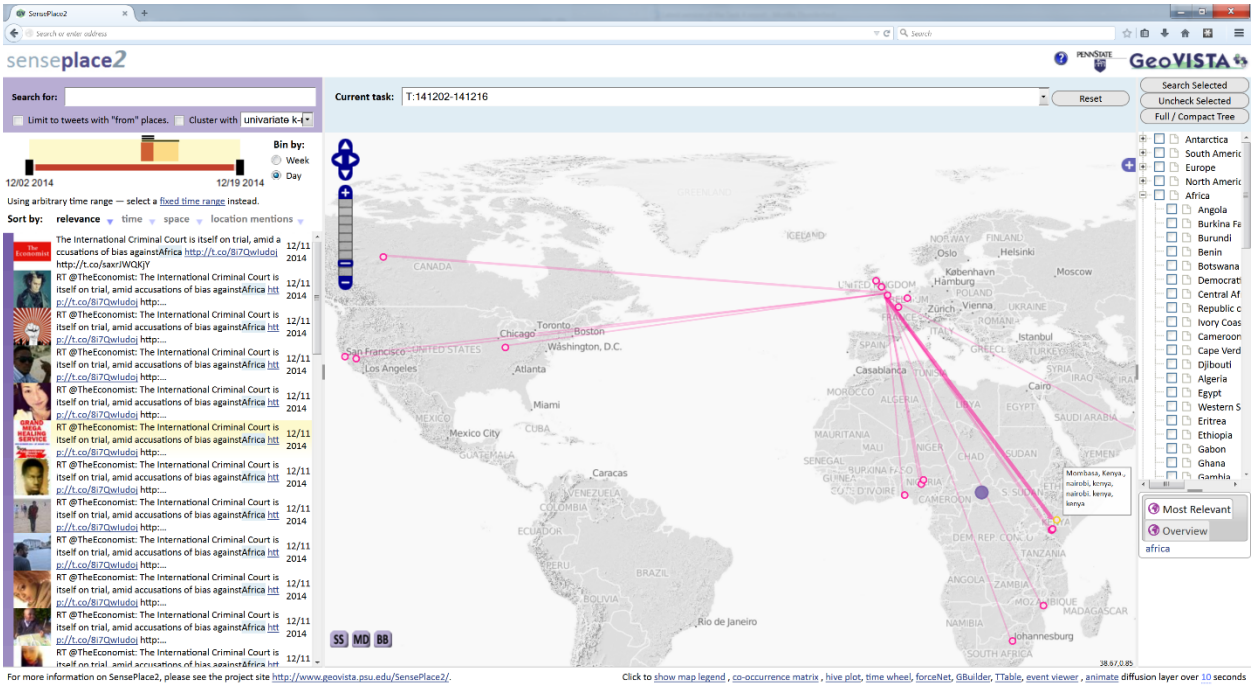

Figure 5. This figure depicts the result of an analyst identifying one tweet of interest and utilizing the SensePlace2 capability to launch a query for all retweets of that particular tweet. This results in up to the 1000 most recent retweets of that tweet appearing in the TweetList along with a depiction on the map of both the place(s) mentioned in the tweets (purple) and the profile locations of the retweeters (pink, for those who have a geolocatable profile location). If multiple retweets have the same profile location, then the connecting arrow depicts the number of retweets with its width.

4.3 Scenario 2: spatial information diffusion footprint of a user group

In addition to supporting analysis of the geographic reach of individual tweets, as depicted above, we have added the capability in SensePlace2 to retrieve the retweets of all tweets by any individual or any small set of Twitter IDs. For individuals, the analyst can initiate the query through the TweetList as above; thus when a tweet of interest is identified, a mouse-over brings up options and one of these is to access "Retweets from User". Alternatively, the analysts can identify a Twitter ID of interest in the TTable and query for "Show retweets from" there. The TTable, however, adds the capability to select multiple Twitter IDs and query for all retweets of all tweets by the set of IDs.

Some Twitter IDs are seldom retweeted, of course. Thus, to quickly generate likely candidates IDs of interest, a new query term has been added to the main SensePlace2 interface, "RTC". This query retrieves only tweets that have been retweeted. The query accepts one parameter, an integer

representing the threshold of retweets that must be met before a tweet and its retweets are returned. With no integer added, the query returns the 1000 most recent tweets that have been retweeted. In Figure 6 below, the query "RTC:15 Sydney" returns up to 1000 of the most recent tweets that include the term "Sydney" and that have been retweeted at least 15 times. These were initially depicted in the TweetList. The Twitter IDs responsible for these tweets populate the TTable. There may be fewer IDs in the TTable than tweets in the TweetList as some IDs will have multiple matching, frequently retweeted tweets.

In this example, the analyst sorted the TTable twice to find individuals who mentioned people or hashtags in their tweets. A set of 14 Twitter IDs are selected and the "Show retweets from" is issued. To make the returned results tractable, constraints are put on what is retrieved. At present, the constraints are fixed, but it would be possible for future releases of SensePlace2 to expose these as user settings. This process works as follows: 1) when the query is issued, the system first searches for the most recent tweets from each user in the list for which there are N or more retweets (N = the number specified in RTC: query, in this case 15); 2) next the system takes those 15 most recent tweets (that reach the retweet threshold) and searches for the 15 most recent retweets of each of those tweets (as above, the integer inserted in the query defines the number retrieved here, with the current system retrieving the same number of retweets per tweet); and 3) the system returns one or more tweets from each selected Twitter ID along with its associated retweets. You should see the 20 tweets along with 20 retweets of those tweets in the tweet list. It is often far fewer since we are working with a small data set.

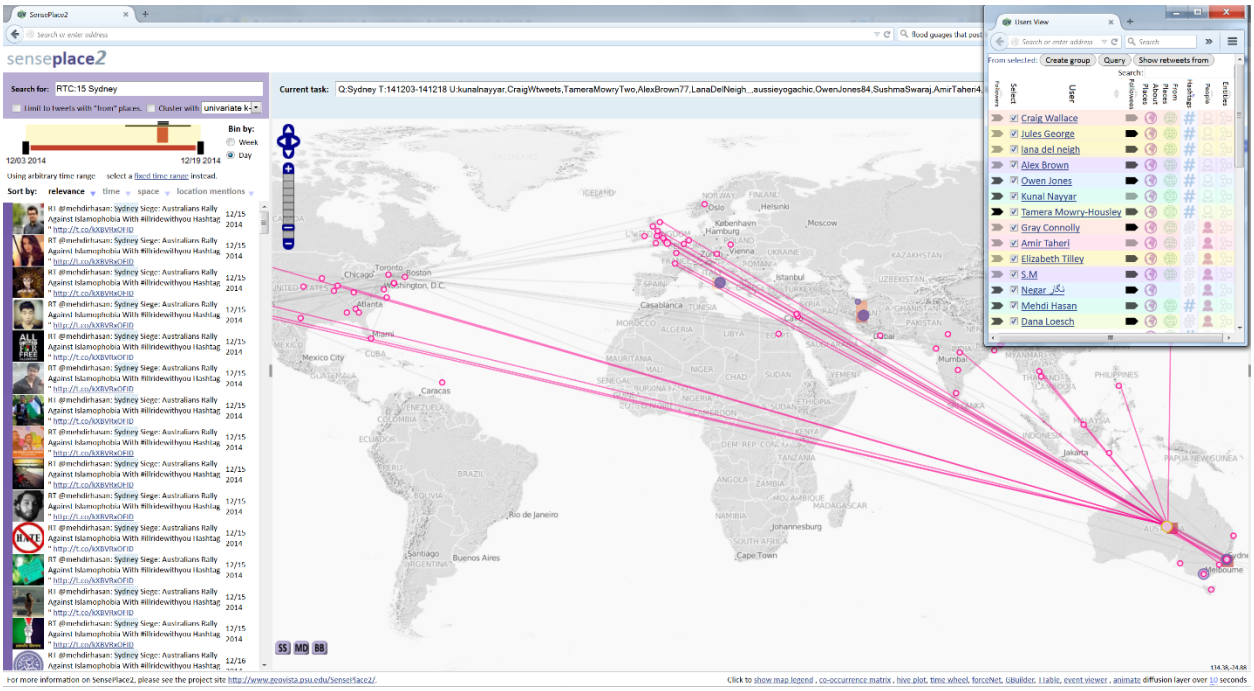

Figure 6. This figure depicts the results of a query for tweets mentioning Sydney that were retweeted frequently. The query was issued on Dec. 16, 2014, the day that the hostage siege in a Sydney, Australia café ended. Four profile locations at which tweets from Sydney were retweeted have been clicked on by the analyst and those links are highlighted in brighter pink with the retweets moved to the top of the TweetList and highlighted. The TTable view shows the set of Twitter IDs used as the query for retweets.

A second example is focused on the geographic reach of news media from different countries (Figure 7). Here, the analyst queried Twitter IDs with fairly large numbers of retweets for messages mentioning Sydney and then used the TTable to find the two that mentioned the most additional places, ABC news (top) and BBC Breaking News (bottom). Not surprisingly, ABC News dominates the U.S., with most of their retweeters in the U.S. while BBC Breaking News has a more global footprint reaching into Europe and having a balance across the rest of the globe. Neither source made an impact in Russia/Asia.

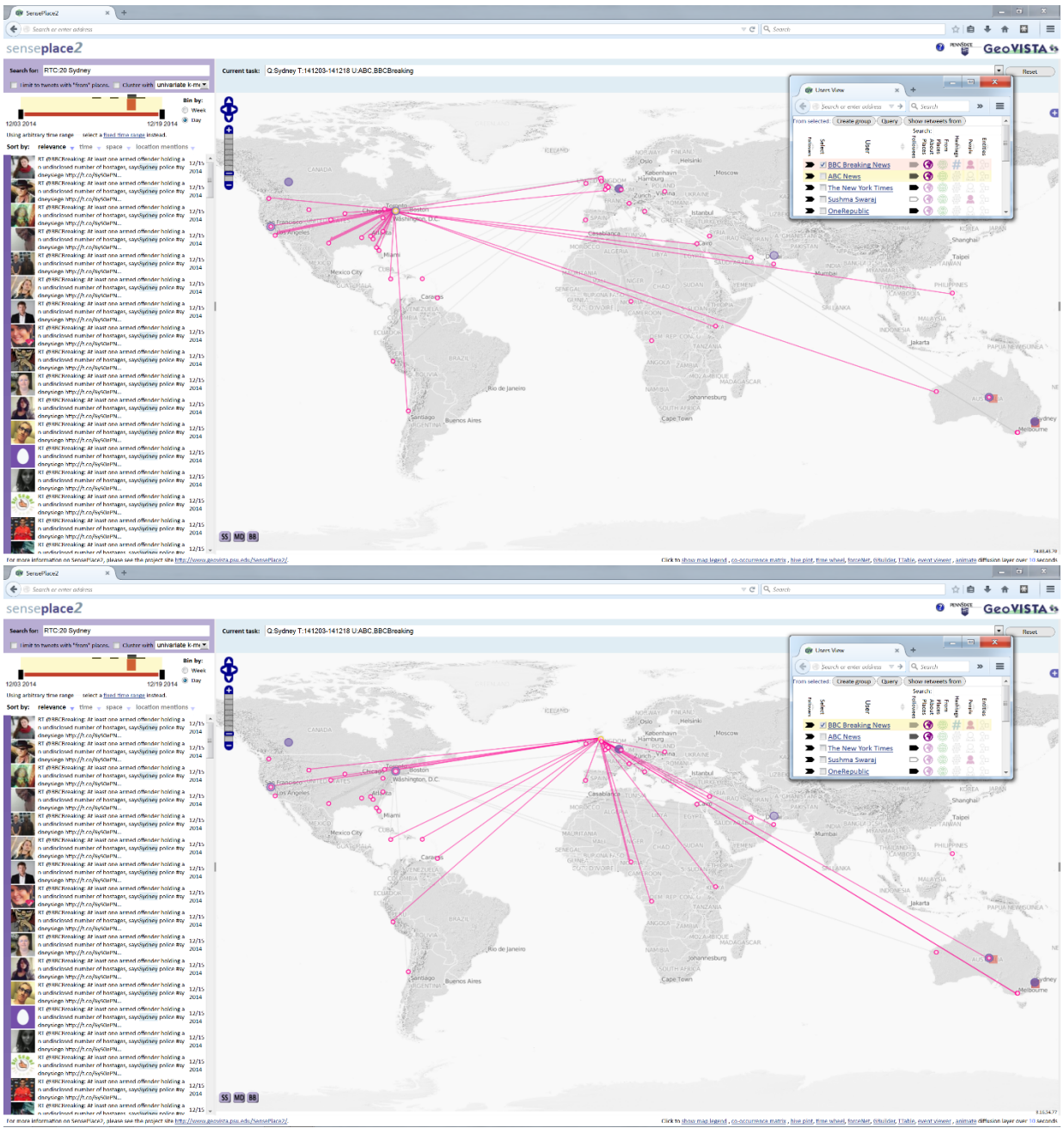

Figure 7. The two screen captures depict results for a retweet query that retrieved tweets mentioning Sydney that were retweeted 20 or more times. The analyst sorted the TTable to find those Twitter IDs that mention the most places in their tweets (thus tweets that not only mention Sydney but other places).

5 Drill-down through dynamic connections in SensePlace2

5.1 Incremental build-up of analytical capacity across Tasks 1, 2, 3 and 4

In the initial SensePlace2 web application, searches were run as SQL queries against a PostgreSQL database that stored tweet information. PostgreSQL does have support for a free-form text search such as the one a SensePlace2 user would expect when searching for tweets containing a term. However, as more data from Twitter was harvested, processed, and expected to be analyzed, it quickly became apparent that PostgreSQL and SQL queries (or any relational database with SQL query) would not return results in a reasonable amount of time to allow a SensePlace2 user to derive actionable information.

SensePlace2 user searches are difficult to predict and therefore searches need to be done in real-time versus preprocessing the data set. Early prototype SensePlace2 searches also required re-ordering the results and aggregated counts. SQL query slowness in these early prototypes prompted us to begin storing tweet information in a Solr text based index. Solr implements an inverse index on all fields determined to be of interest; this strategy supports orders of magnitude faster query than a full database scan used in a typical relational database. Previously, some of the slower SensePlace2 searches would take ten or more seconds, which is unacceptable for an analyst. With Solr, similar searches, including aggregate counts, now have mostly sub-second performance. PostgreSQL is still used in SensePlace2 for archived storage and limited pre-processing of tweets.

In addition to moving away from SQL as the mechanism to support live SensePlace2 searches, tweet harvesting, archived storage, processing of tweets, and Solr storage for the web application was also simplified. First, tweets are harvested from Twitter's streaming API through the use of the Twitter4J Java API. Tweets coming from the streaming API are written to a PostgreSQL database using a JSON field type. Because tweets are already provided as JSON, this only requires us to insert a record. We do not process tweets at this stage in order to ensure that the harvester is able to keep up with the full volume of incoming tweets. An intermediate step at this point is to download and locally save the Twitter user's profile image for use in the web application, as the links to these images often change, resulting in broken links.

After tweets are stored in the PostgreSQL database, the tweets are then retrieved for processing. Both the tweet text and user profile location are run through our GeoTxt API (GeoTxt.org, (Karimzadeh et al., 2013)) to extract location and other entities, and geocode the locations. The locations are also indexed to map overview grid cells at the various zoom levels. Solr also supports point within polygon calculations. This would allow for the incorporation of unpredicted geographic polygon features as a data aggregation layer. Performance times of these aggregate counts have not been tested and would depend on many factors such as the number of tweets matching a search and the complexity and number of geographic polygons. A minor amount of other processing is done on the tweets such as pre-calculating the day of the week for the tweet.

After the geographic processing is finished, every Tweet ID associated with tweets in the repository is queried against a table to find a record of that user's Twitter followers and followees (known as "friends" in Twitter). The follower and followee information supports the generation of the SensePlace2 orceNet visualization as well as enabling exploration of message diffusion through existing networks.

The user's followers and followees have been previously acquired through a separate process using the Twitter Users API.

The last steps of processing involve storing the processed data. The processed data is stored as SolrJ document attributes and JSON data. A SolrJ document is built by flattening all of the tweet's JSON attributes and appending the processed attributes. The processed attributes are also converted to JSON data which is appended to the original JSON data provided by Twitter. The SolrJ document is then written to the Solr index for use in the web application and the JSON data is written back to the PostgreSQL database into an additional JSON field type to support archiving the data.

Even with dramatic performance improvements in SensePlace2 searches due to the use of Solr, web application users have come to expect search results immediately – even a few seconds is far too long for most to tolerate. Since some SensePlace2 Solr searches take a couple of seconds to complete, users may notice delays, such as slow map interactions. Zooming and panning the map triggers a redraw of the map overview layer. Web application users will not accept a redraw time of a few seconds each time the map is panned.

Given the need of even greater search performance, we are collaborating with the Penn State Institute of CyberScience (ICS) to run the SensePlace2 web application searches against a SolrCloud deployment on five powerful servers. SolrCloud is designed to be distributed across multiple computers to improve search times. Deployment of the SensePlace2 data into this SolrCloud has once again improved search times over a single instance Solr. Most Solr searches, including aggregate counts, ordering of results, and grouping of similar tweets, return results in under one second. This collaboration with ICS has increased the usability of SensePlace2 and therefore the ability to derive more meaningful analysis.

5.2 Enhancements of the coordination mechanism

As part of the progress made in Task 3, limitations were identified in the coordination framework used by the SensePlace2 application. In Task 4, a number of those limitations have been addressed. This effort can be broadly classified into three related categories that build on each other: data propagation through the coordination mechanism, replay of existing events, and the introduction of an event viewer component.

Data propagation is one of the most important tasks in the SensePlace2 interface engineering and also one of the more challenging ones. As the number of components used in the interface grew, maintaining manually-written code for updating data between queries throughout the entire SensePlace2 interface became infeasible. A new event type, *dataUpdate*, was then introduced to the SensePlace2 coordination framework, and all of the components were eventually moved to using the coordination mechanism as the way to refresh data between queries, thereby simplifying project management and making components less dependent on each other.

Replay of existing events became necessary with the introduction of new SensePlace2 components that appear in pop-up windows. One of the challenges of the multi-window setup in a browser environment is in keeping windows coordinated with the rest of the application in terms of data contents, data selections, and highlights. A "fast forward" component was introduced to the

SensePlace2 architecture that synchronizes pop-up windows with the main application in a transparent fashion that does not require any engagement from the user. This functionality is currently enabled in the co-occurrence matrix for beta-testing.

An *Event viewer component* was devised to experiment with guided coordination, that is, to let the analyst interfere with the default decisions made by the coordination mechanism. The original use case was to allow for a smooth reset of the coordination state of any given component, and this use case was expanded into a visual interface for flexible management of events across components. An analyst is currently able to disable any of the existing events and later turn it back on, if needed, in order to bring the system to a desired state. The event viewer is currently in early testing phase with the co-occurrence matrix and the hive plot components.

5.3 Representing information diffusion through co-occurrence matrix

The co-occurrence matrix tool has been expanded to take full advantage of the new "retweets / retweet of" relationship added to the SP2 UI data model. The following examples illustrate a number of scenarios in which the co-occurrence matrix complements existing tools and provides for fine-grain exploration of the information diffusion patterns.

The first example deals with the recent news event (hostage situation) originating from Sydney, Australia (December, 2014). With the objective of building an understanding of the current diffusion patterns, the analyst retrieves the latest 1000 retweets dealing with the topic of Sydney and opens the co-occurrence matrix tool. The analysts then adjusts the co-occurrence matrix to show the authors of the most popular messages along the vertical axis and the profile locations of people that retweeted their messages along the horizontal axis (shown in the Figure 8). It is clear that in the current dataset, it is just a few messages that acquired significant retweet popularity.

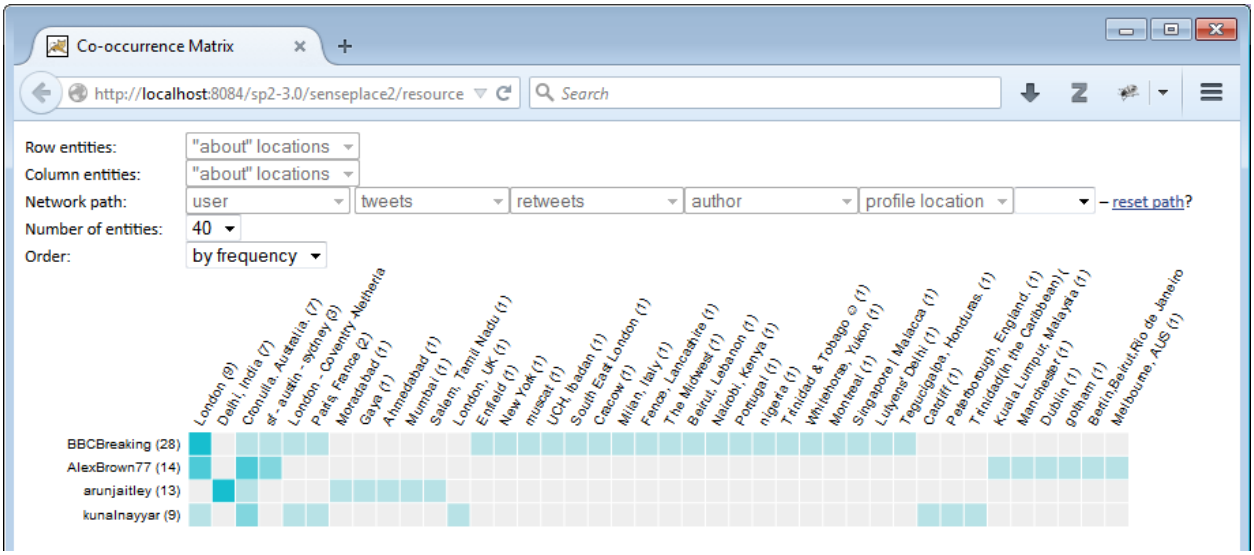

Figure 8. Co-occurrence matrix depicting a metapath from Tweet IDs to profile locations of other Tweet IDs who retweeted their posts. The results are for a query on "Sydney" on December 17, 2014, a day after the hostage situation in Sydney, Australia ended. Rows are the Tweet IDs posting the most retweeted tweets and the columns are the profile locations for the Tweet IDs that issued the most recent retweets.

The analyst then readjusts the co-occurrence matrix to look at the spatial aspect of this diffusion (Figure 9). The co-occurrence matrix now shows the profile locations of the source tweets along with the profile locations of the people retweeting them. It is clear that there are four distinct places with fairly disjoint spatial information diffusion footprints.

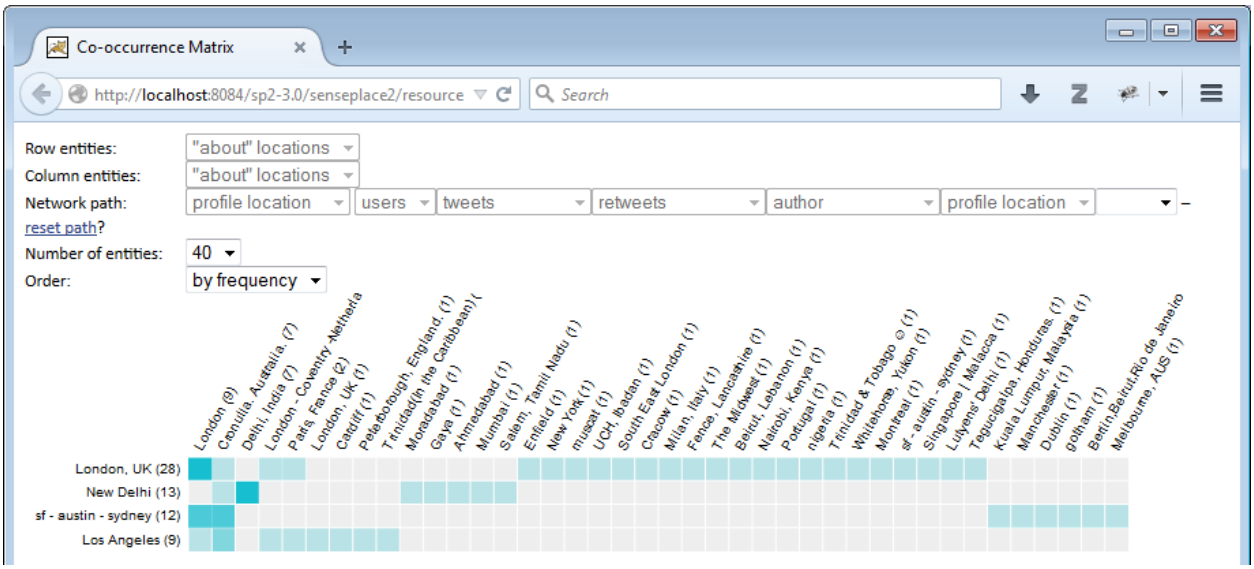

Figure 9. Co-occurrence matrix showing profile locations of key source tweets (rows) matched with profile locations of retweeters (columns).

The analyst clicks on a couple cells in the first row ("London") to see where the messages from London end up traveling. The results appear on the map of the main SensePlace2 view (Figure 10).

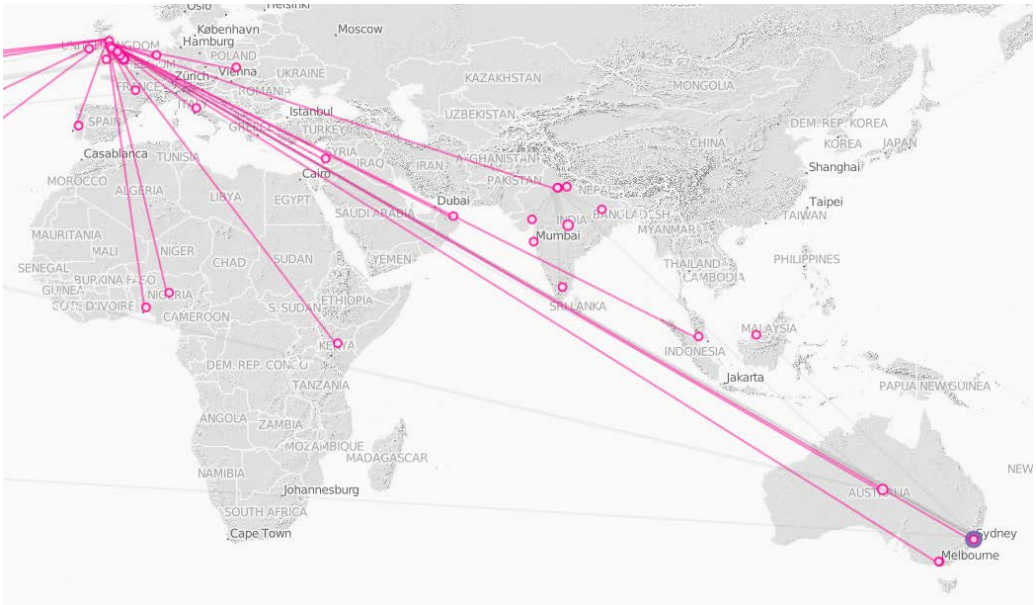

Figure 10. Map view from analysis session showing the geographic footprint (based upon profile locations of retweeters) for tweets posted in London.

The analyst then generates addition screenshots (Figure 11), to produce diffusion maps for two additional tweet sources, New Delhi then for Sydney. As shown, in contrast to the relatively global reach of the tweet from London, the one from New Dehli is retweeted primarily within India and the one from Sydney shows strong UK connections.

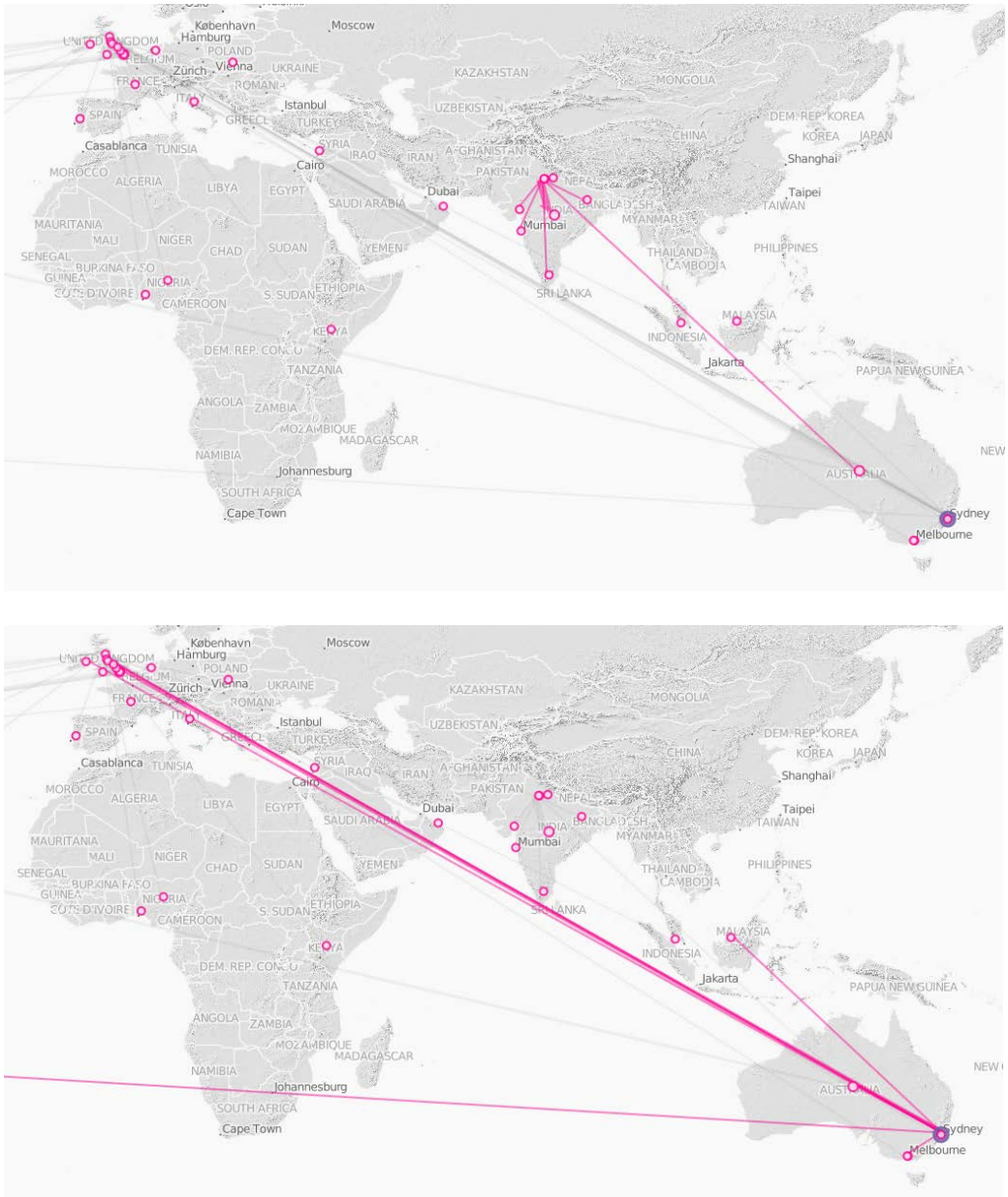

Figure 11. Geographic footprints of retweets for tweets from Delhi, India (top) and for Sydney, Australia (bottom).

The second example deals with a structured overview and drill-down of the information diffusion pattern for a group of users of interest. The analyst starts the query by opening the GBuilder tool, selecting a number of users he/she is interested in, and fetching the latest retweets originating from them. The resulting diffusion display is complex, with many connections between multiple places (Figure 12).

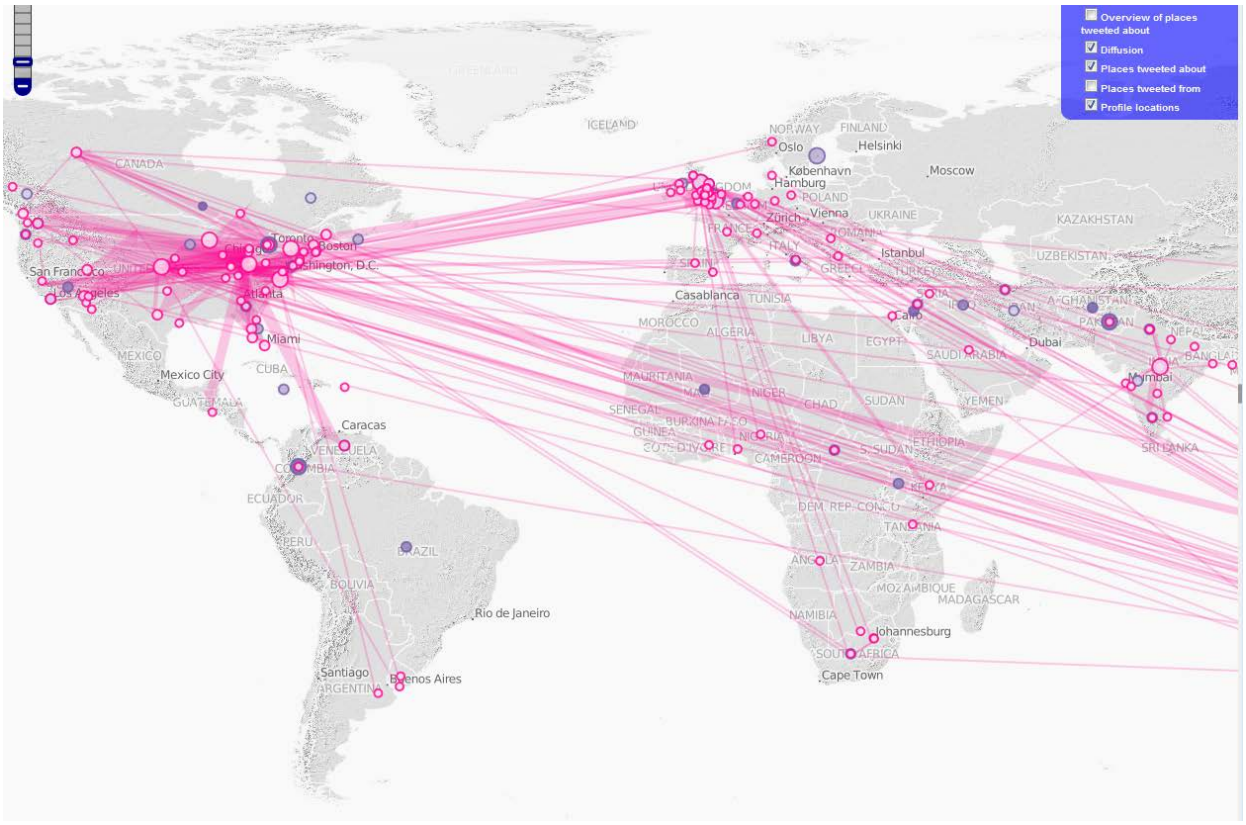

Figure 12. SensePlace2 map view resulting from a query on retweets from a group of Tweet IDs of interest.

To explore this complexity, the analyst then opens the co-occurrence matrix tool and re-adjusts it to see the diffusion of different hashtags across the globe (Figure 13). It is clear that some of the most recent conversations are dominated by the Sydney event. However, there are a number of other trending retweets that look unfamiliar.

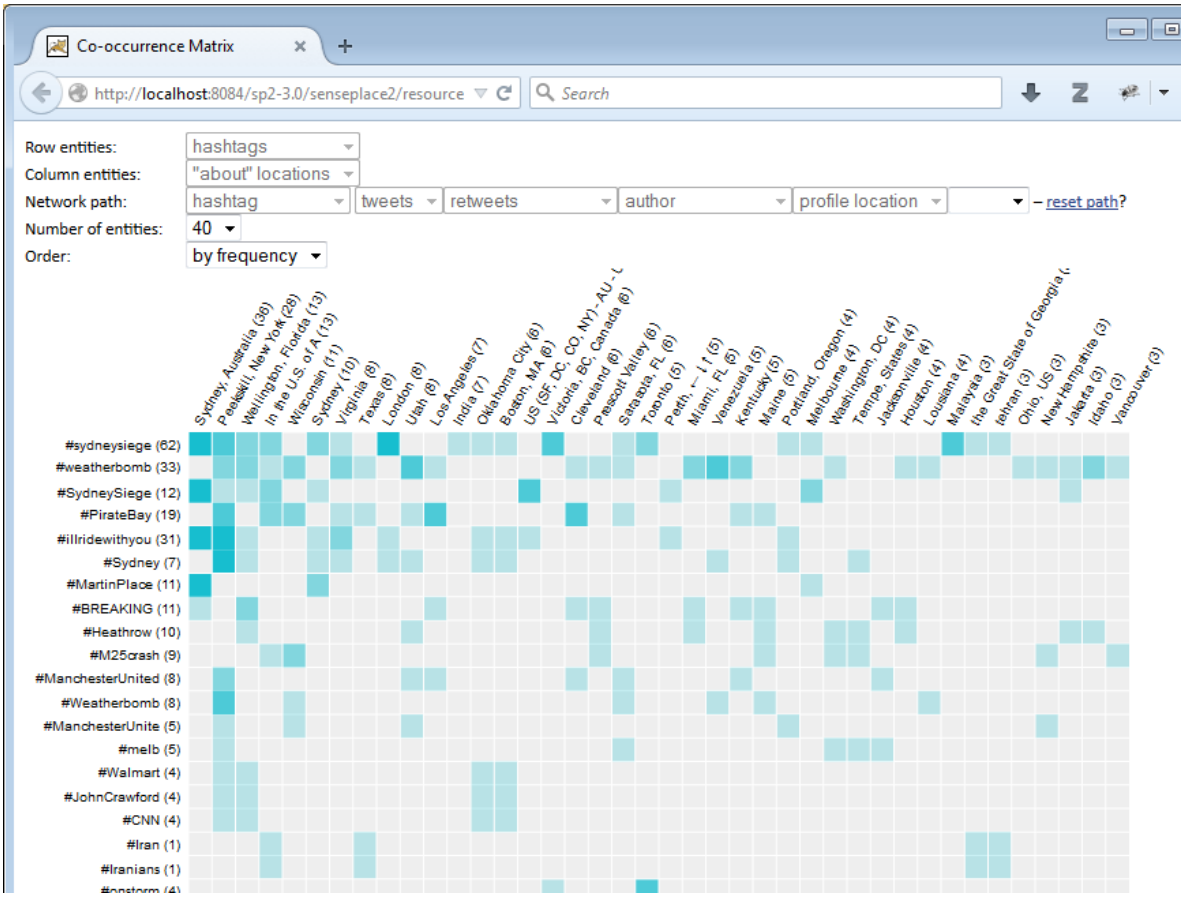

Figure 13. Co-occurrence matrix depicting hashtags in retweeted tweets matched with profile locations of the Tweet IDs for retweeters.

The analyst clicks on a few cells next to the "weatherbomb" hashtag. This causes the main map to display the image shown in Figure 14.

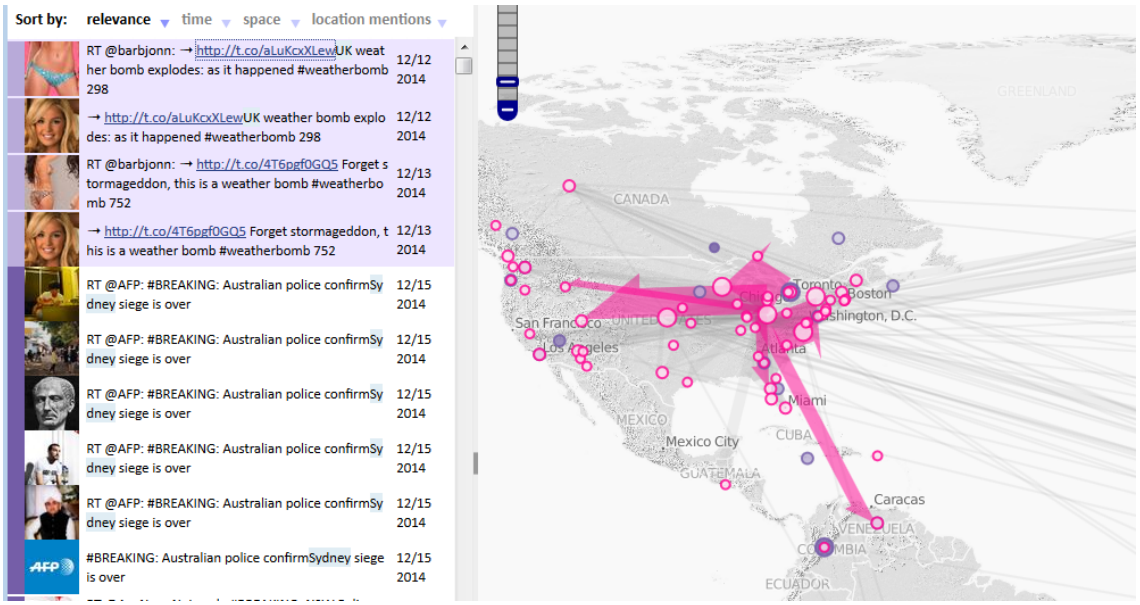

Figure 14. Display showing retweet connections (between profile location of Tweet ID of poster and Tweet ID of retweeter) for a subset of tweets that include the #weatherbomb hashtag.

Apparently, this was a new hashtag associated with severe weather anomalies, and it gained certain traction, although its spread has been mostly local (within the U.S. with an emphasis on the eastern U.S. where the tweet originated). The analyst then clicks on cells next to another unfamiliar hashtag – "pirate bay". The resulting display is shown in Figure 15.

It appears that this snippet of news is related to the closure of the Pirate Bay website is generating some local traction as well. Pirate Bay is an online index of digital content including films, TV shows, and music {http://www.engadget.com/2014/12/16/pirate-bay-shutdown-explainer/ } that went off-line on Dec. 9, 2014 after a raid by Swedish police; http://www.bbc.com/news/technology-30477678).

Finally, the analyst generates a daily snapshot of the location awareness statistic, looking to map media exposure that different places get on Twitter. The analyst re-adjusts the co-occurrence matrix, generating the screenshot depicted in Figure 16, showing connections between 'about' locations mentioned in tweets and the profile locations of retweeters. Not surprisingly, most profile locations that are retweeted from frequently are geographically associated with the mentioned places (e.g., 'Sydney, Australia' has the most retweets for tweets with 'Sydney' and 'Hyderabad, India' has the most retweets for tweets that mention 'India.' But, many other places mention Sydney frequently (e.g., Indonesia, Kuala Lumpur, Malaysia, England) and the most frequent mentions of 'Iran' are from 'England' and 'London'.

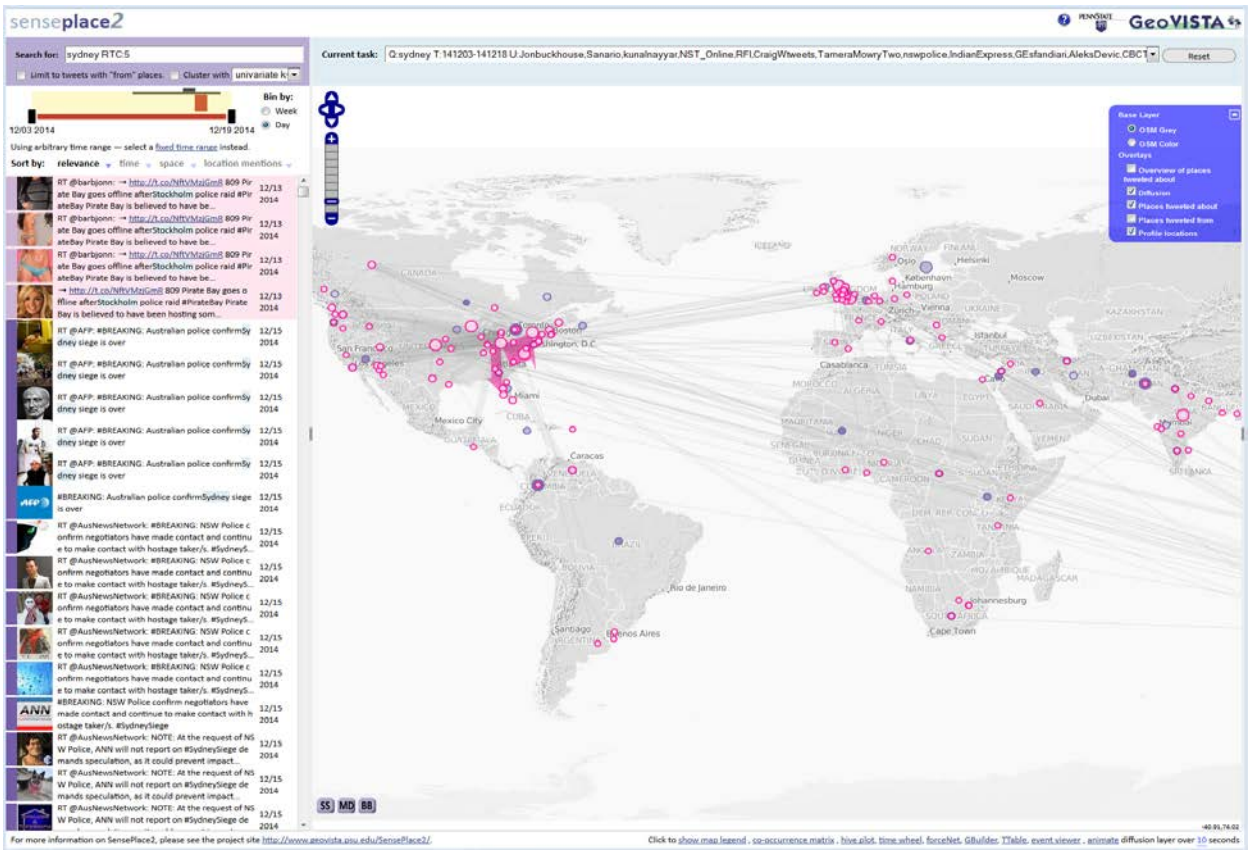

Figure 15. Depiction of retweets for the "#PirateBay" hashtag.

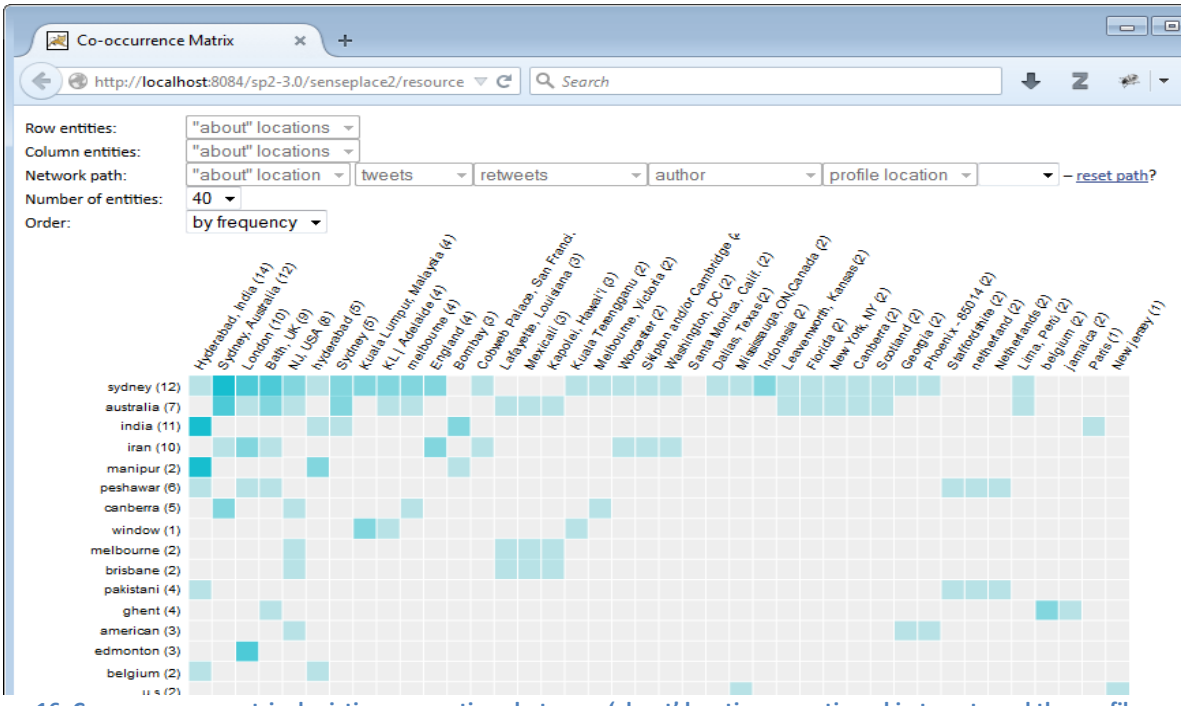

Figure 16. Co-occurrence matrix depicting connections between 'about' locations mentioned in tweets and the profile locations of retweeters.

6 Conclusions and Future Work

6.1 Conclusions and Lessons Learned

This report presents the Task 4 outcomes from our larger research directed to supporting place-based information foraging and sensemaking with open media. Outcomes reported in this report focused on message-based analysis grounded in place. More generally, the research demonstrates the potential of linking place-focused, tweeter-focused, and social-focused analysis. The research demonstrated both the potential to leverage social media to support information foraging and sensemaking while also identifying some of the limitations related to characteristics of the data and its accessibility.

The resulting extensions to SensePlace2 demonstrate the potential of Twitter and other related data as a source of information about message propagation in social media. We have designed and implemented both new data models and visualization tools to enable flexible exploration and analysis of such data and we have taken advantage of Solr indexing and high performance computing to support processing of and quick access to large volumes of streaming data. However, there are also some key lessons learned about limitations of Twitter as a data source, particularly when access is limited to the public 1% data stream. In particular, a system that would track diffusion of information across ill-defined, evolving or very large groups of users will be challenging to create for Twitter, as well as for other social media data sources. It would require both extended hardware capabilities to carry out iterative queries to Twitter that follow links and/or access to the entire Twitter fire hose of data. Conceptually, this problem poses a new research challenge that could leverage recent developments in GeoSocial Visual Analytics (Luo et al., 2014; Luo and MacEachren, 2014).

Beyond the limitations of accessible Twitter data, a second lesson learned relates to existing methods for visualization of data that includes spatial, temporal, and network components. While there have been some advances on methods for structured space-time-network data associated with data sources related to transportation and GPS tracked movement (e.g., (Andrienko et al., 2011; Andrienko et al., 2014)), much less attention has been given to less structured data extracted from text and related sources. In the next section we briefly review recent research that may provide a platform for developing new visualization strategies to depict and explore such data.

6.2 Literature review of promising visualization techniques

A large proportion of the literature has focused on building the computational methods to collect and process data from social media. Innovative approaches developed for dynamically-linked multi-view displays on the web provide both overviews and detail-on-demand. Our literature review also identified a need for innovation in visualization methods designed to cope with place-time referenced data extracted from unstructured sources. Below, we provide a brief overview of some promising developments in information visualization and related domains upon which new advances could be built.

The network graph is a common visualization technique used to represent information propagation. SensePlace2 already takes advantage of the network graph technique in the ForceNet plot.

The plot allows analysts to explore connections between Twitter users, place mentions, and hashtags. The ForceNet plot could be extended to represent more aspects of information propagation in the Twitterverse, including the kinds of heterogeneous connections that can be specified through the Co-occurrence matrix as illustrated above (e.g., from the about locations of tweets through their authors, to the retweets, and on to the profile locations of the retweeter). Thus an analyst could use the ForceNet to identify connections among places talked about by a TwitterID and the profile locations where those place mentions are retweeted.

TweetXplorer, referenced in Morstatter et al., 2013, is a system for visualizing big social media data that uses a simple network graph to visualize retweets (Figure 17). Nodes represent users and edges between nodes represent retweet activity. The outer color of the nodes encodes keyword group associations, which are determined by the majority of the group of user's individual tweets. The darkness of the inner color encodes the number of times a user was retweeted. Drawing on these ideas, careful use of color in SensePlace 2's ForceNet plot could be an effective strategy for distinguishing among analyst identified categories (e.g., local versus national place mentions, mentions of people, places, or organizations) or to depict the proportion of retweeters who include or do not include a geo-locatable place in their profile.

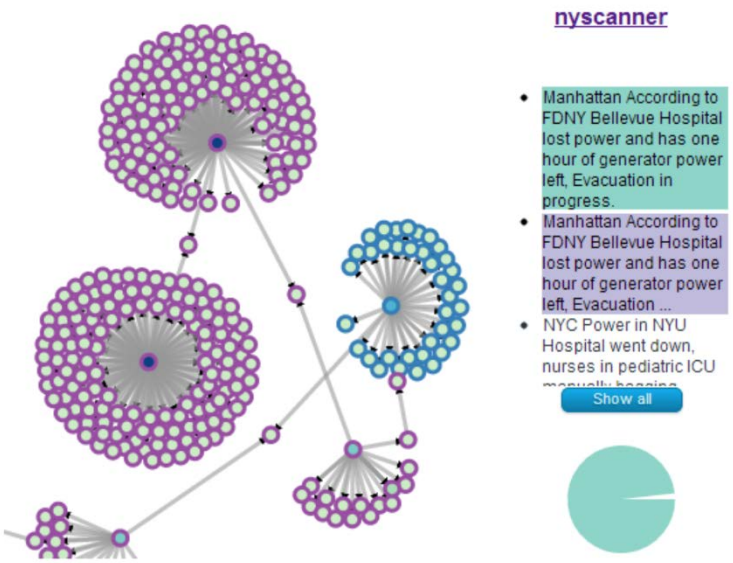

Figure 17. Adapted from (Morstatter et al., 2013). This figure depicts the network graph and tweet panel components of TweetXplorer.

From an analytical perspective, Ota et al. (2012) have shown how visualizing the overlap between retweet propagation paths in a network graph can provide new insights on connections between Twitter users who share similar interests but do not necessarily follow one another (Figure 18). Their approach aims to better understand relationships between central users and retweet root users. Thus, the visualization allows users to explore the ways in which their interests on Twitter overlap with other users' interests through overlaying retweet network paths.

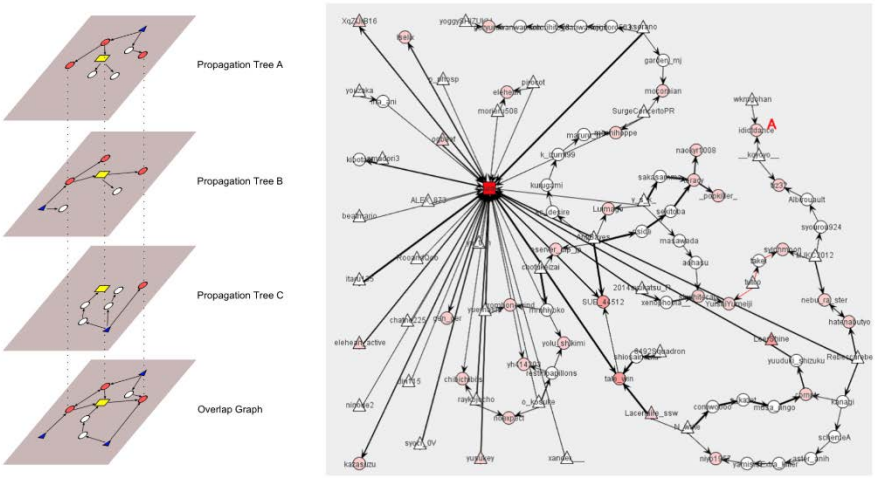

Figure 18. Adapted from (Ota et al., 2012). This figure illustrates the overlaying of retweet propagation paths to inform connections between users who not necessarily follow one another.

However, a major challenge in using network graphs to visualize information propagation is interpretability. If, for example, each node in the graph represents a single user and the number of potential users is unbound, the graph can become highly complex, and links between nodes become indiscernible. Aside from cognitive overload, graph size and complexity also affect query time and interaction responsiveness. Allowing users to zoom to areas of interest can alleviate complexity in network graphs. SensePlace2's ForceNet plot already provides traditional zoom functionality, with focus+context zoom functionality via a "fisheye" lens under development. Such functionality would allow analysts to see areas of interest in the network graph in full detail, while also providing an overview of the surrounding area for context.

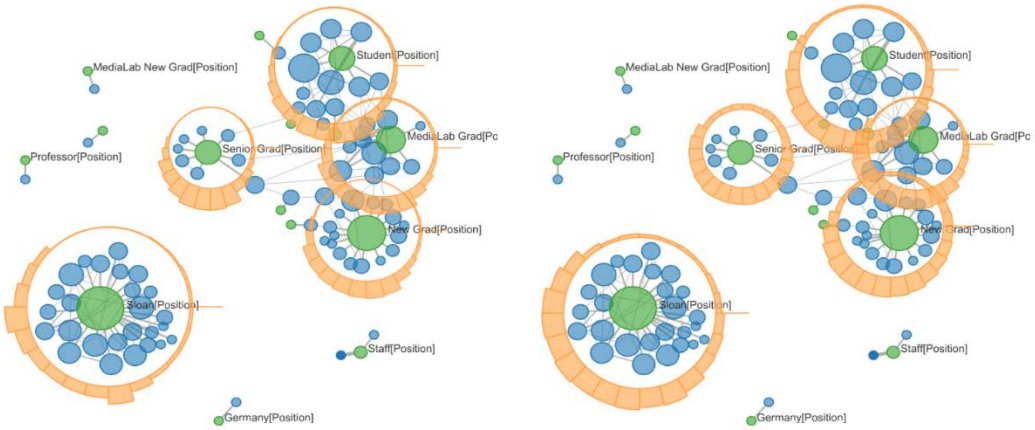

Figure 19. Adapted from (Tinati et al., 2012). This figure illustrates the use of "behavior rings" in the network graph to encode social, spatial, and temporal characteristics of users and their connections.

Another approach to reducing complexity in a network graph is to aggregate individual nodes into meaningful groups. Tinati et al. (2012), for example, developed a network graph of retweet activity between groups of users with similar communication behavior instead of mapping individual user activity. More specifically, the Tinati group applied a "topology of influence" to classify users into idea starters, amplifiers, curators, commentators, and viewers.

Similarly, Shen and Ma (2008) developed MobiVis, a visual analytics tool that allows analysts to interactively group nodes in a heterogeneous network visualization to help them make sense of the connectivity between social-spatial-temporal data captured by mobile phone devices. To support comparisons between groups of users, MobiVis plots "behavior rings", or radial plots, around aggregate nodes, which encode temporal and behavioral attributes about user groups (Figure 19). Aggregating users and designing attribute rings could again be useful strategies in relaying information propagation in future SensePlace2 iterations.

Currently in SensePlace2, coxcomb plots, a specific type of radial plot, are being developed to more intuitively compare tweet and retweet distributions, starting at the continent scale (Figure 20), and eventually being extended to finer spatial resolutions.

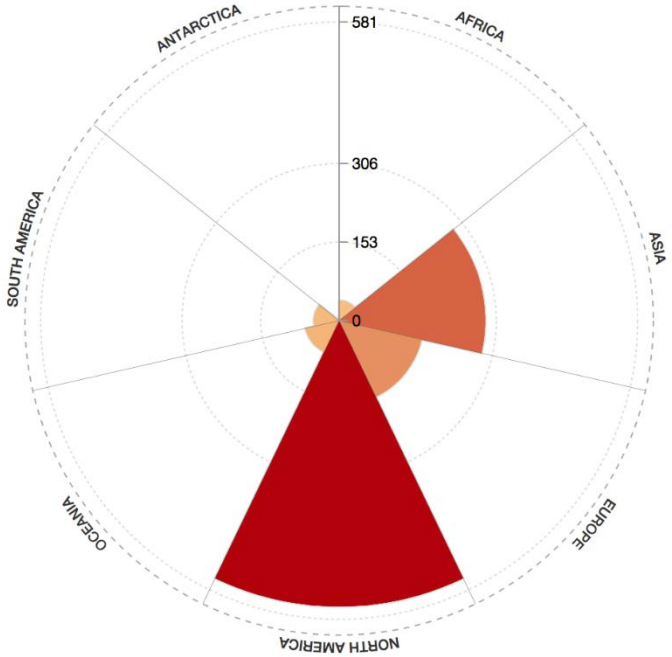

Figure 20. This radial plot depicts the distribution of 1000 sample tweets by continent.

Lastly, because visualizing and making sense of information propagation in Twitter data is so challenging, entire visual analytics systems can be designed for this sole purpose. For example, Whisper is a visual analytics tool that uses a sunflower metaphor to visualize when, where, and how an idea is dispersed (Cao et al., 2012). By default, original tweets are placed in the center of the visualization, and paths to user groups by location move outward (Figure 21). Inactive or non-retweeted tweets eventually fade out from the center to reduce clutter and focus on active tweets, which are those being retweeted. Color encodes sentiment, opacity conveys activeness, and symbol shape relays user types (square = media outlet, journalist, or organization representative; circle = user that does not belong to previous categories; and arrow = retweet). The visualization can also be inverted from a diffusion mode to a convergence mode if the analyst is interested in seeing the origin of the source tweet. The tool incorporates a simple but useful linear timeline that relies on knowing who retweeted a retweet.

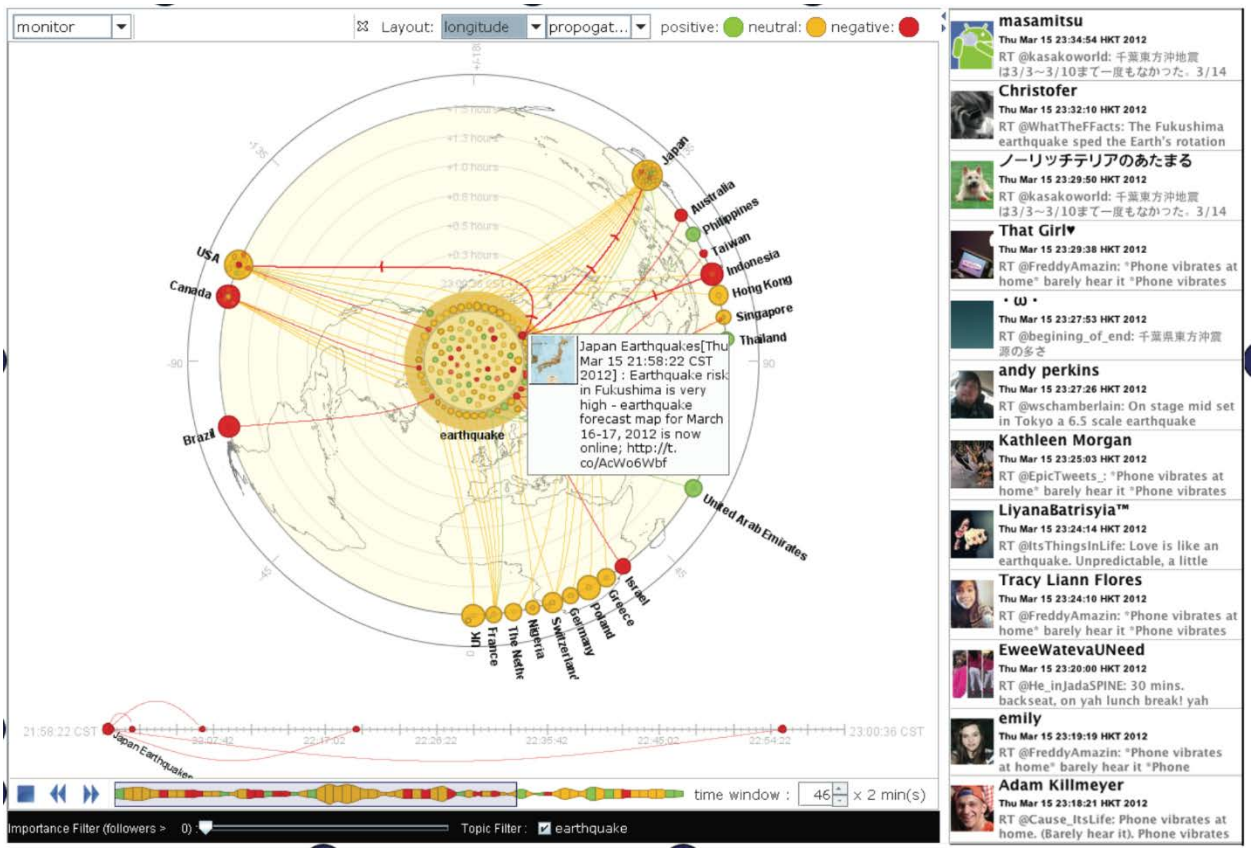

Figure 21. Adapted from (Cao et al., 2012). This figure depicts the Whisper interface. Original tweets are placed in the center of the visualization and retweets propagate outward in real-time.

6.3 Future directions for the SensePlace2 project

We have identified multiple challenges for future research. A primary one is to extend SensePlace2 capabilities to leverage @mention and @reply information. In our research thus far, we determined that when a tweet is provided by Twitter's streaming API and that tweet is a retweet, the original tweet is also provided in full. This means a visualization of retweets is currently possible in SensePlace2.

When a user is mentioned or a tweet is replied to, Twitter's streaming API provides less information. The mentioned or replied to tweet is not provided; only its tweet ID is available. This creates a challenge for visualizing user @mention's and @replies. The Twitter streaming API provides a fraction of all tweets. Therefore, given that only the tweet ID of the mentioned or replied to tweet is provided, currently there would only be a small chance that the tweet it is associated with is also collected. Tweets are not acquired specifically by tweet ID, but by search terms or Twitter user ID.

In the future, it would be possible to design a system that began collecting tweets from a user when that user is mentioned in a tweet. In this way, SensePlace2 could visualize tweets from a user along with the future tweets of a mentioned user. In addition, Twitter provides an API to retrieve a tweet given the tweet's ID. If a tweet is replied to, the tweet that replied can be obtained in this manner. This API is rate limited, which would make this collection difficult.

The development of algorithms for building comprehensive overviews of existing user communities and their informational footprint that would operate effectively within the limitations of Twitter Streaming API is an additional challenge. One possible solution lies in building networks of intelligent Twitter crawlers capable of sharing the task of community mapping through real-time interaction.

Another interesting challenge lies in exploring feasible scenarios for fusion of spatial data from multiple sources (profile locations, "about" and "from" locations) in the context of social interactions between Twitter users. Individual user profiles, friendship networks, as well as individual conversations provide unique contextual information that could make fine-grained location disambiguation possible.

The process for fusing these sources of spatial data remains a significant issue that is not yet automated. An example problem where this fusion is necessary is the need to reconcile an out-of-date profile location, a set of "from" locations with multiple spatial clusters, and a range of "about" locations mentioned in a Twitter stream of a given user. This problem is further exacerbated when we attempt to characterize communities of users that communicate with each other and share interests.

Finally, as outlined in **Section 6.2**, a range of interesting challenges lie in the domain of visual representation of complex geosocial interactions between individual users in the networks.

7 References

Andrienko, G., Andrienko, N., Hurter, C., Rinzivillo, S. and Wrobel, S. 2011: From movement tracks through events to places: Extracting and characterizing significant places from mobility data. *Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on*, 161-170.

Andrienko, G., Andrienko, N., Schumann, H. and Tominski, C. 2014: Visualization of Trajectory Attributes in Space–Time Cube and Trajectory Wall. *Cartography from Pole to Pole*: Springer, 157-163.

Cao, N., Lin, Y.-R., Sun, X., Lazer, D., Liu, S. and Qu, H. 2012: Whisper: Tracing the spatiotemporal process of information diffusion in real time. *Visualization and Computer Graphics, IEEE Transactions on* 18, 2649-2658.

Karimzadeh, M., Huang, W., Banerjee, S., Wallgrün, J.O., Hardisty, F., Pezanowski, S., Mitra, P. and MacEachren, A.M. 2013: GeoTxt: A Web API to Leverage Place References in Text *7th ACM SIGSPATIAL Workshop on Geographic Information Retrieval,* Orlando, FL: ACM.

Luo, W., Di, Q., Yin, P., Hardisty, F. and MacEachren, A.M. 2014: A Geovisual Analytic Approach to Understanding Geo-Social Relationships in the International Trade Network. *PLoS ONE* 9, e88666.

Luo, W. and MacEachren, A.M. 2014: Geo-Social Visual Analytics. *Journal of Spatial Information Science* 8, 27-66.

MacEachren, A.M., Jaiswal, A., Robinson, A.C., Pezanowski, S., Savelyev, A., Mitra, P., Zhang, X. and Blanford, J. 2011: SensePlace2: GeoTwitter Analytics Support for Situational Awareness. In Miksch, S. and Ward, M., editors, *IEEE Conference on Visual Analytics Science and Technology,* Providence, RI: IEEE, 181 - 190.

MacEachren, A.M., Karimzadeh, M., Banerjee, S., Luo, W., Savelyev, A., Pezanowski, S., Robinson, A.C. and Mitra, P. 2013a: Report on New Methods for Representing and Interacting with Qualitative Geographic Information, Stage 2: Task Group 2: Tweeter-focused use case. *Report on New Methods for Representing and Interacting with Qualitative Geographic Information, Contract #: W912HZ-12-P-0334,* University Park, PA: GeoVISTA Center, Department of Geography, The Pennsylvania State University.

MacEachren, A.M., Savelyev, A., Luo, W., Pezanowski, S., Karimzadeh, M., Robinson, A.C. and Mitra, P. 2014: Report on New Methods for Representing and Interacting with Qualitative Geographic Information, Stage 2: Task Group 3: Social-focused use case. *Report on New Methods for Representing and Interacting with Qualitative Geographic Information, Contract #: W912HZ-12-P-0334,* University Park, PA: GeoVISTA Center, Department of Geography, The Pennsylvania State University.

MacEachren, A.M., Savelyev, A., Pezanowski, S., Robinson, A.C. and Mitra, P. 2013b: Report on New Methods for Representing and Interacting with Qualitative Geographic Information, Stage 2: Task Group 1 – Core Re-engineering and Place-based Use Case. *Report on New Methods for Representing and Interacting with Qualitative Geographic Information, Contract #: W912HZ-12-P-0334,* University Park, PA: GeoVISTA Center, Department of Geography, The Pennsylvania State University.

Morstatter, F., Kumar, S., Liu, H. and Maciejewski, R. 2013: Understanding Twitter Data with TweetXplorer. *KDD '13,* Chicago, Illinois, 1482-1485

Ota, Y., Maruyama, K. and Terada, M. 2012: Discovery of interesting users in Twitter by overlapping propagation paths of retweets. *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2012 IEEE/WIC/ACM International Conferences on*: IEEE, 274-279.

Savelyev, A. 2013: Multiview User Interface Coordination in Browser-Based Geovisualization Environments (Demo Paper). *The 1st ACM SIGSPATIAL Workshop on Map Interaction, in conjunction ACM SIGSPATIAL,* Orlando, FL.

Savelyev, A. and MacEachren, A.M. 2014: Interactive, Browser-based Information Foraging in Heterogeneous Space-Centric Networks. In Andrienko, G., Andrienko, N., Dykes, J., Kraak, M.-J., Robinson, A. and Schumann, H., editors, *Workshop on GeoVisual Analytics: Interactivity, Dynamics, and Scale, in conjunction with GIScience 2014,* Vienna, Austria.

Shen, Z. and Ma, K.L. 2008: MobiVis: A Visualization System for Exploring Mobile Data. 175-182.

Tinati, R., Carr, L., Hall, W. and Bentwood, J. 2012: Identifying communicator roles in twitter. *Proceedings of the 21st international conference companion on World Wide Web*: ACM, 1161-1168.

Wallgrün, J.O., Karimzadeh, M., MacEachren, A.M., Hardisty, F., Pezanowski, S. and Ju, Y. 2014: Construction and First Analysis of a Corpus for the Evaluation and Training of Microblog/Twitter Geoparsers. In Purves, R. and Jones, C., editors, *GIR'14: 8th ACM SIGSPATIAL Workshop on Geographic Information Retrieval,* Dallas, TX: ACM.