New Methods for Representing and Interacting with Qualitative Geographic Information

Contract #: W912HZ-12-P-0334

Contract Period: January 1, 2014 – June 30, 2014

Principal Investigators:

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

Dr. Prasenjit Mitra, IST & 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 3: Social-focused use case

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

< maceachren, savelyev, wul132, spezanowski, karimzadeh, arobinson >@psu.edu; pmitra@ist.psu.edu

GeoVISTA Center, Department of Geography, The Pennsylvania State University
Submitted, June, 30, 2014

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
06-16-2014	Final	Jan. 1, 2014 – June 30, 2014

4. TITLE AND SUBTITLE

Report on New Methods for Representing and Interacting with Qualitative Geographic Information, Stage 2: *Task Group 3 – Social-focused use case*

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

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Alan M. MacEachren, Alexander Savelyev, Scott Pezanowski, Anthony C. Robinson, and Prasenjit Mitra

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)
Topographic Engineering Center (TEC)
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 Pennsylvania State University's (PSU) research on place-focused analysis of microblogs, specifically Twitter. The first section of the report introduces the goals for this task and briefly reviews both SensePlace 2 (the web application in which we have implemented these new methods) and prior work on Tasks 1 & 2. The second section of the report (detailing the primary results of this component of research) focuses on our approach for enabling query and visual exploration for social and other relationships in Twitter data and on the development and implementation of tools to support this task; these include a Group Builder, a Force-directed Graph tool, and a Hive Plot that are dynamically connected to existing SensePlace 2 components. The third section of the report presents two scenarios of use to more fully illustrate the potential of the methods and tools implemented. Section four briefly summarizes adaptations made to our system architecture to support adding a social focus to the existing tweet and tweeter focus of our tools and to support scaling up to increasingly large data volumes. Finally, we summarize our progress 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			19b. TELEPHONE NUMBER *(include area code)*
SAR	SAR	SAR	SAR	28	

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

Abstract

This report documents Pennsylvania State University's (PSU) research on visual tools for query and exploration of data from microblogs, with a specific focus on Twitter as a primary data source. Our emphasis is on support for revealing where, when, what, and who appear in microblog data to support tasks such as situational awareness for natural disasters and other emergency events. This report summarizes research and outcomes for Task 3 in this research, which is focused on adding capabilities for social-focused queries and visual analysis to augment the existing place-focused and tweeter-focused methods and tools completed for Tasks 1 & 2, respectively.

The first section of the report introduces the goals for this task and briefly reviews SensePlace 2, the web application in which we have implemented these new methods and prior work on Tasks 1 & 2. The second section of the report (detailing the primary results of this component of research) focuses on our approach for enabling query and visual exploration for social and other relationships in Twitter data and on the development and implementation of tools to support this task. The third section of the report presents two scenarios of use to more fully illustrate the potential of the methods and tools implemented. Section four briefly summarizes adaptations made to our system architecture to support adding a social focus to the existing tweet and tweeter focus of our tools and to support scaling up to increasingly large data volumes. Finally, we summarize our progress and outline some future research challenges.

1 Introduction

This report details progress on a multi-stage research project focused on leveraging place-relevant data from microblogs through methods that support queries and visual exploration of the unstructured data they produce. All methods under development are implemented in SensePlace 2, a visual analytics web application that collects Twitter data for selected keywords, processes and indexes those data, and provides interactive visual tools that support queries focused on where, when, who, and what can be found, and on the exploration of interrelationships among these information components.

The main SensePlace 2 application is designed to leverage data from Twitter (or other text streams with short messages) to support information foraging and situation awareness. The most novel feature of SensePlace 2 is that, in addition to depicting location that tweets are from (the 1.5% on average for which the user has enabled geolocation via either GPS coordinates, IP address geolocation, or user-supplied place names), it adds locations listed in profiles of the Tweeters (when present) and the places tweeters are talking about. For the latter, the application uses named-entity extraction and geoparsing methods to identify and geolocate the place references in tweets. Early development work on SensePlace 2 is presented in MacEachren, et al (2011) and a detailed explanation of the core components of the interface is available as a downloadable user guide (which does not yet include the new features introduced here): http://www.geovista.psu.edu/SensePlace 2/SensePlace 2_Interface_Mini_UserGuide.htm. Other information can be found in: (a) Savelyev (2013), which introduces the client-side component coordination mechanism implemented to support our multi-view approach to foraging for and sensemaking with information that contains place, time, and attribute components and (b) Savelyev and MacEachren (accepted), which introduces the application of

heterogeneous network modeling to interactive web-based visual analytics tools for exploring relationships in geo-located social media data.

Task 1 focused on a Place-focused Use Case (MacEachren et al., 2013b). Task 2 built upon and complemented this with a Tweeter-focused Use Case (MacEachren et al., 2013a). Here, our focus is on extending the environment to explore activities of individuals and networks of individuals, grounded in the space-time context that SensePlace 2 provides. More specifically, this use case is directed to the entity extraction, overview, drill-down, and dynamic connection advances implemented to support analysis focused on understanding place-related social networks.

2 Techniques Developed & Extended

Three new components have been implemented in SensePlace 2 to support the social-focused use case. The first component is a group building and managing tool – Gbuilder, which allows users to specify and edit groups of Twitter IDs to follow. The other two new components support interrogation of relationships among entities, which can include Twitter users as well as the places associated with those users (places that they talk about, tweet from, or cite as the place in their profile). These are: (a) a force-directed node-link graph tool – ForceNet and (b) a Hive Plot, which is essentially a radially organized parallel coordinates plot in which connections among entities (ordered on axes representing different analytical variables) are depicted with links. Each of these three components is detailed below.

2.1 GBuilder: Creating and managing groups to follow and explore

This component complements the TTable (Tweeter Table) developed in Task 2. The GBuilder (Groups Builder) component enables authorized SensePlace 2 users to create and edit membership in multiple ad hoc groups (Figure 1). Once a group is created, "following" of the group can be turned on or off. Multiple groups can be created and multiple groups can be followed at the same time (subject to Twitter API rate limits). When the SensePlace 2 user turns following on, this initiates systematic queries to Twitter. These queries return all tweets by those users (up to a maximum of 5000 per user ID queried upon) and with them the full tweet metadata, which contains fields detailing a range of tweeter characteristics that include: profile location; time zone; counts of followers, friends, favorites, statuses; language. See System Extensions, Section 4.1 below for more details on the data collection process.

For any group created, SensePlace 2 users can initiate a query on tweets by group members that are in the system index (by selecting a full group or subset of members, then clicking the "Query on Selected" button). If the "Reset Query" button in the main interface is used first, the system returned up to 1000 of the most recent tweets (stored in the index) by any of the users in the group. If "Reset Query" is not used, the system returns up to 1000 of the most recent tweets (stored in the index) by any of the users in the group that also match the current query parameters (thus the current time range, query terms, and any spatial constraints). If such a query is submitted shortly after creating a group, the number of tweets returned will be based on prior query-based data collection from Twitter; thus the 1% sample limits of Twitter's keyword query and/or the use of relevant keywords within tweets may result in zero tweets returned from the SensePlace 2 index for some or all group members. But, over time, as tweets by the specified users are collected, more tweets by group members will be returned. Since

SensePlace 2 focuses on place-based tweets, only tweets by group members that have one of three kinds of place information will be indexed (and thus be available for query): from locations (based on coordinates or place specifications), about locations (based on place names in text that are recognized and geocoded by the SensePlace 2 geoparsing facility), or Profile locations (again, when they can be recognized and geocoded by the SensePlace 2 geoparsing facility). More details on the process of "following" group members and the limitations imposed by Twitter are detailed in Section 4.1.

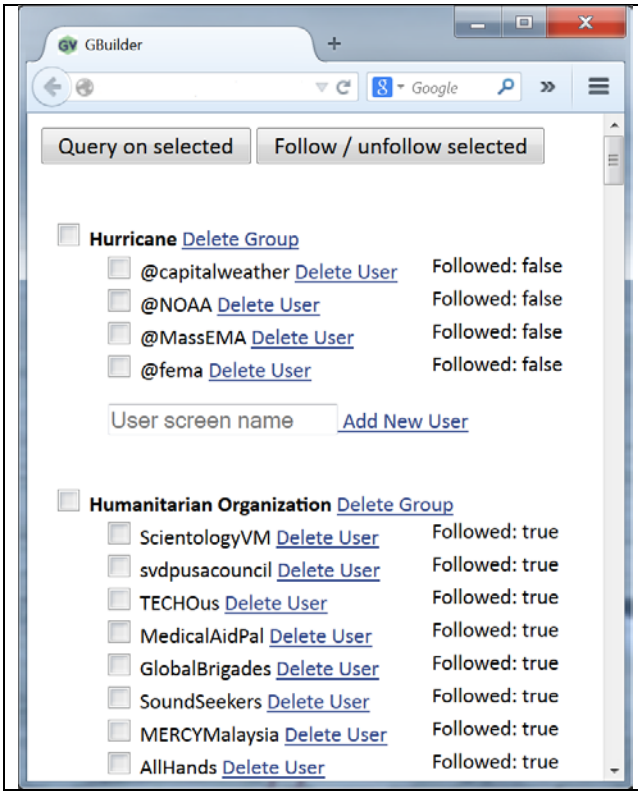

Figure 1 Gbuilder interface for building and monitoring groups' tweets. The interface enables users to add and delete users in any group. It also enables adding new groups and deleting existing groups. Not shown is the toggle to turn on and off the following capability and the control that enables a SensePlace 2 user to launch a query on the index to retrieve all tweets by any of the users in the group.

2.2 ForceNet: Exploring connections among users, places, and topics

The goal of the ForceNet tool is to understand how linked entities relate; more specifically in relation to analysis of Twitter posts, to understand the social connections among Twitter users and how these connections relate to the places and topics posted from and about. ForceNet implements a force-directed node-link diagram layout that reconfigures automatically to depict any nodes and links associated with a particular SensePlace 2 user query. The tool is adapted from the open-source d3 library: d3 API. Here is the link: https://github.com/mbostock/d3/wiki/Force-Layout; we have customized the basic graph layout tool to coordinate with other SensePlace 2 components and to recognize multiple node types representing entities of interest in the Twitter data: Twitter user, places, hashtags, and user mentions. The base d3 graph tool implements a force-directed layout, introduced by Dwyer (2009), that combines physical simulation and iterative constraint relaxation for stable graph layout. The initial implementation generates a dynamic graph layout in which users can interactively move nodes to explore connections and shift nodes of interest to make them more visible. Currently, when the graph is redrawn, the node position changes. We plan to provide an option to fix node position to support comparisons (e.g., across time).

When using ForceNet, the SensePlace 2 user has a range of controls on the ForceNet view:

- Users can resize the nodes according to different attributes: tweeter user nodes can be resized according to the number of their followers or followees, and all other nodes can be resized according to the number of times they are mentioned in the tweet list. Those different attributes also pop up when users mouse over those nodes.
- Users can zoom in/out, and pan the force-directed layout with the mouse scroll wheel.
- The network layout can scale with the window size accordingly.
- The coordination among this view and all other views works well in terms of highlight and selection.

As an example of functionality, the query term of "attack" was used to access the 1000 most relevant tweets with the term (Figure 2). The ForceNet tool was launched (shown in the dynamically linked view at right); it depicts 4 node types: users (red), mentioned locations (green), mentioned users (yellow), and hashtags (blue). As seen, for tweets mentioning "attack" during this 2-week period in June, 2014, three place-based clusters of tweets are apparent (it can be seen that they are place-focused clusters because the node at the center of each is green). From upper right to lower left, these tweets focus on Karachi, Benghazi, and Iraq. In this example, one tweet is highlighted (in the TweetList). These corresponding entities in the ForceNet are also highlighted and linked in red. This includes the Twitter username (red circle) and four features (the place of Benghazi in green and three hashtags in blue, for: #Benghazi, #BoweBergdahl, and #RedEye). Linking works in the other direction as well, of course.

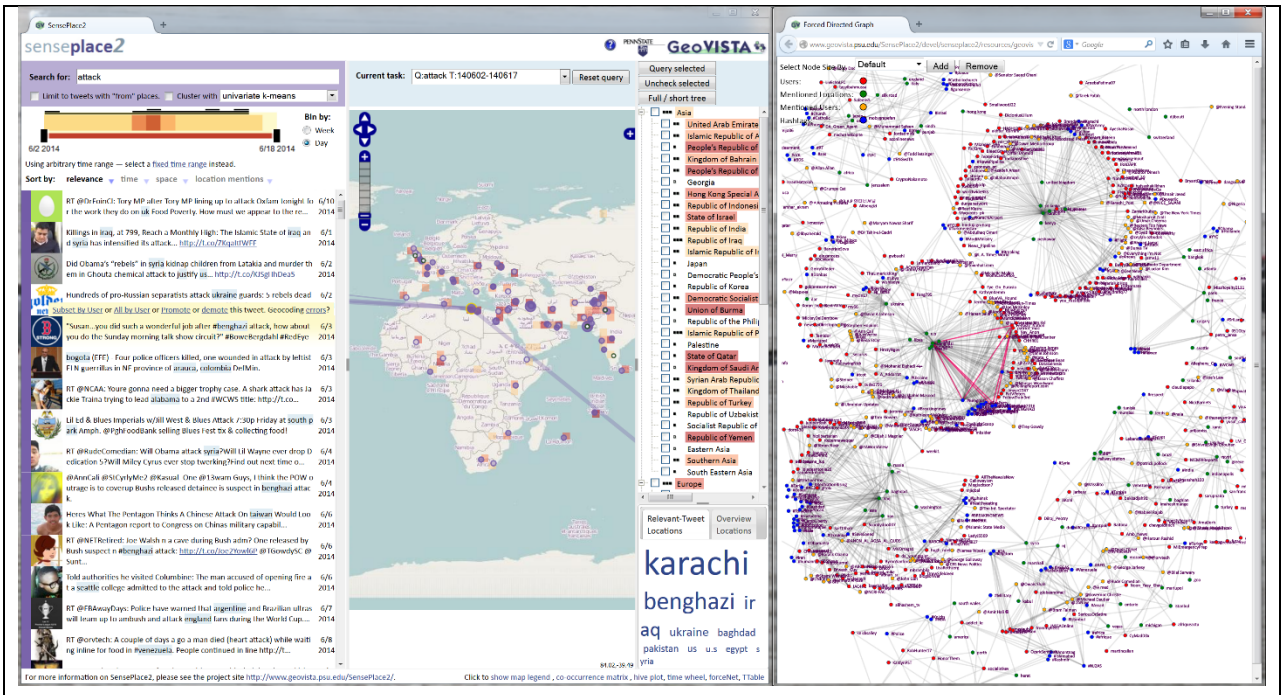

Figure 2. A query on "attack" in the first half of June, 2014. The Tweet List (left view) depicts the 1000 most relevant tweets, initially sorted by relevance. The map depicts the overview of tweet frequency (with grid cells for which darker red means more) and details of the number of mentions for specific places with graduated circles in red and the number of tweets from specific places with graduated circles in green. The right column of the main (left) view displays the place-tree hierarchy and a place-focused tag cloud. The ForceNet view depicting relationships among all Tweeters, place mentions, user mentions, and hashtags in the top 1000 tweets is depicted at the right. Dynamic linking between views is illustrated.

One of the clusters, for Benghazi, is selected in the Place Tag Cloud and that prompts the ForceNet view to re-configure to depict only nodes from the tweets that mention Benghazi (Figure 3). In this view, the option to signify the number of tweets represented by each node is turned on and depicted with graduated circles. The place of Benghazi is the largest since it was the feature selected on. Other prominent mentions include the U.S. and the hashtag of #Benghazi.

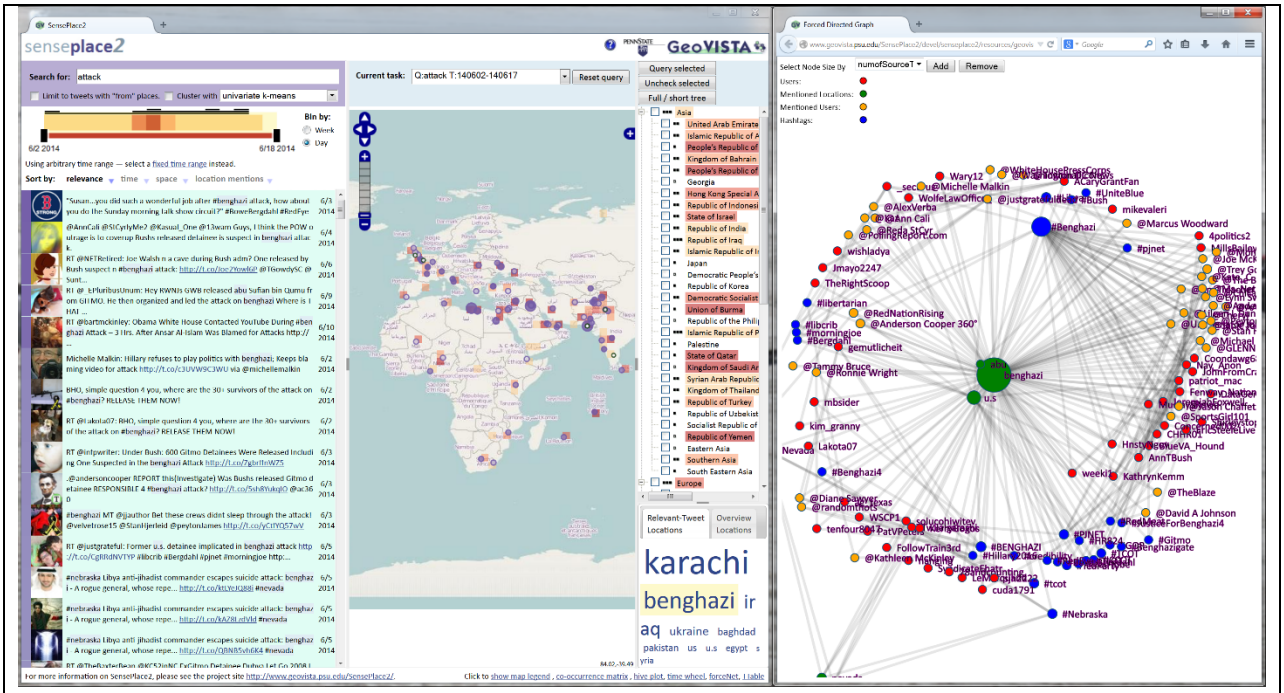

Figure 3. The result of a selection on Benghazi initiated by clicking on its name in the place tag cloud. This set of views provides quick access to all tweets in the time period that mention Benghazi (in the scrolling list at the left) and all other places, plus users, user mentions, and hashtags associated with those tweets (in the ForceNet view).

The view below (Figure 4) illustrates more interactivity. Specifically, the #Benghazi node in ForceNet has been moused over (right view). This highlights the places and people and hashtags mentioned in conjunction with the #Benghazi hashtag within tweets. The visible matching tweets are highlighted in the Tweet List as are the linked places (on the map). Clicking on the node would promote all associated tweets to the top of the Tweet List and would refocus the ForceNet on all tweets that have the #Benghazi hashtag rather than all that mention the place (the mentions include the hashtag references plus other references without hashtags).

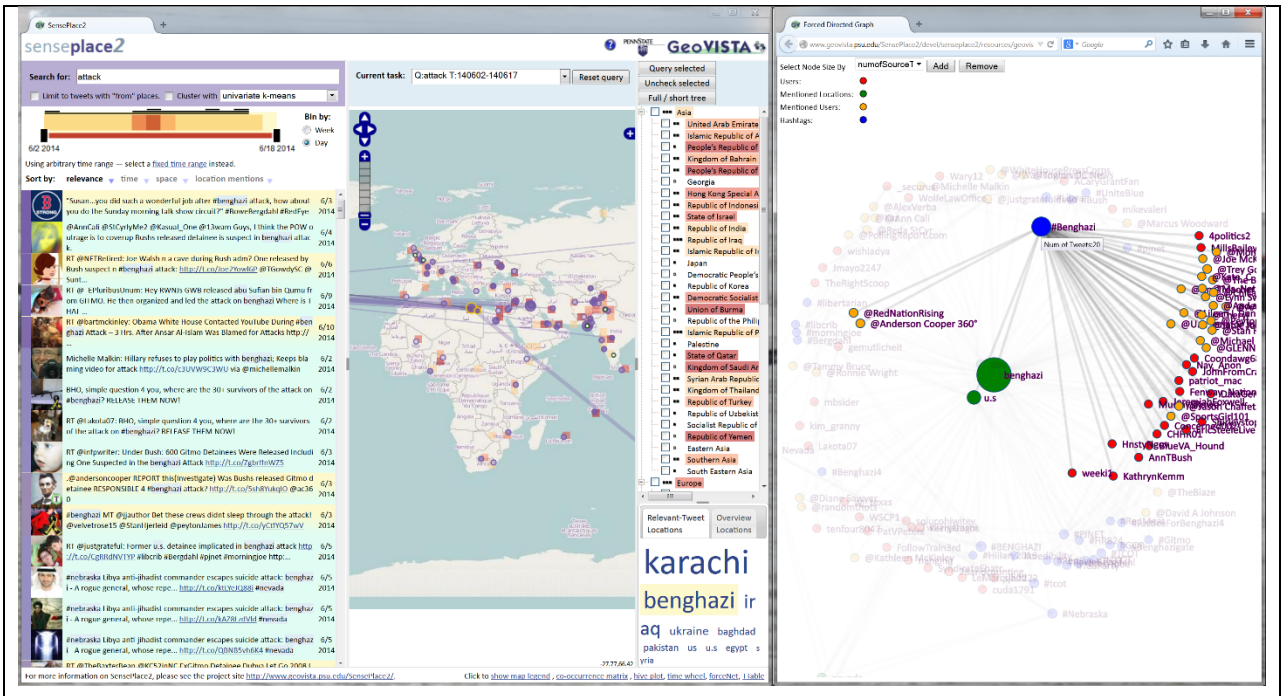

Figure 4. A mouse-over highlight of the subset of tweets containing a place reference to Benghazi that specifically includes the hashtag of #benghazi. This view emphasizes the Tweeters who posted with this hashtag (in red) as well as the individuals who the referenced with "@" references (in yellow).

Most of the discussion above centers on working with a single node or tweet (or a small number of them) and looking at the connections. However, the combination of the GBuilder tools with the ForceNet offers potential to explore groups as a whole. The GBuilder supports both: (a) situations in which an analyst pre-builds a group to follow and/or (b) situations in which the analyst identifies some tweeters with interesting connections through initial exploration. We provide a simple example of the former in Scenario 1 below (section 3.1) and an example of the latter (using the HivePlot tool discussed next rather than the ForceNet) in Scenario 2 below (section 3.2).

The group building process implemented thus far is manual, depending on the analyst to know who to follow or to discover users to follow using visual exploration. In the future, a useful extension to the methods would be to implement computational community identification methods to extract groups based on tweet content and other parameters, such as the methods introduced in (Phuvipadawat and Murata, 2011) and integrate these detection methods into a visual analytics workflow that enables iterative refinement and analysis of groups.

2.2.1 Discussion

The ForceNet implements the widely used force-directed layout method for generating graphs with many nodes on the fly from node-link data. We illustrate more completely how the ForceNet tool can be used in an analysis, through a hypothetical scenario focused on leveraging social media for crisis management situation monitoring. Here, we provide brief insights on the pros and cons of a node-link

tool using the force-directed layout approach for use in analysis of place-based data from Tweets. The most obvious advantage of ForceNet is to identify local clusters easily without prior knowledge of the network itself. For example, it is easy to identify a cluster within which the most popular mentioned locations and hashtags are in the middle linked by multiple users or other attribute nodes (i.e., mentioned locations, hashtags, mentioned users). Through dynamic connections with other views in SensePlace 2, the place, time, and concept characteristics of the cluster can be explored.

Initial work with the ForceNet has identified two primary disadvantages of the graph method, as implemented thus far.

1. The method does not scale well to large numbers of nodes and edges. There are two components to this. The first is that large graphs take several seconds to render and direct manipulation with them can be awkward because the graph will not keep pace with user interaction. The typical force-directed algorithms are in general considered to have a running time equivalent to O(n3), where n is the number of nodes of the input graph. If we visualize the animation process with the calculation of node positions at every step, interaction will not be smooth with an interface containing 2000 nodes and 9000 links (typical with 1000 relevant tweets returned). The solution we use here is to hide the animation process (around 10 seconds) and to present the static network layout after the calculation. In this way, it would not cause any interaction issues after the calculation is done. The second scaling problem is more challenging. With a large graph, there will be substantial overlap of nodes. In addition to making some nodes hard (or impossible) to see, overlap also makes it difficult to interact with those nodes that fall in the background. There are several possible solutions to this that we will experiment with in subsequent work: (a) allow users to promote one of the 4 node types to the foreground by clicking on the legend nodes (without otherwise changing the graph); (b) allow users to hide any particular node type completely; (c) allow the user to click on a legend node and pop up a scrollable list of all of the current entities of that type and let the user mouse over and select from that list; and (d) somehow allow the user to move or demote whatever is in the way, repeatedly if needed, until the object of interest is selectable (without redrawing the entire graph).
2. The technique implemented prescribes a *layout algorithm* to be used when generating a visualization. This means that users have limited input (limited to parameters of the algorithm used) on the visual structure of the final visualization. Additionally, the forced directed layout is unstable – small changes in the initial dataset can trigger significant changes in the visual structure of the final visualization, making it hard to compare related or evolving networks. We plan to address both issues by providing variations on the force-directed layout method that enable the user to constrain the layout in various ways. One idea we plan to implement is to allow the user to specify a specific location for selected important nodes, letting other nodes be positioned by the algorithm in relation to these fixed positions (e.g., anchoring 4 key places to locations at the corner of a rectangle and then enabling visual analysis of the other nodes that group with those places).

Beyond issues with the ForceNet as a network visualization tool detailed above, a more important limitation on using the ForceNet tool, or any other network visualization tool, to explore data from Twitter is the limitations that Twitter imposes on access to data. Network statistics, based on the follower-followee relationships of Twitter users, can indicate potentially key individuals (e.g., with high

betweenness). But, the calculation of those statistics is based on the follower-followee network and because of the Twitter API rate limitation, only a partial follower-followee network with a small percentage of the actual connections can be constructed. Costenbader and Valente (2003) show that network statistics (e.g., betweenness, closeness) cannot be calculated reliably with such partial proportions of the whole network. One proposed alternative is to calculate those network statistics based on our current network model among different entities (e.g., users, hashtags, mentioned locations). Those network statistics can imply the potentially key entities during a critical event (e.g., disease outbreak). For example, if one user mentions a lot of hashtags, locations, or users that have also been mentioned by many other users, the user can be identified through network statistics as a key user.

2.3 Hive Plot: Drill-down on connections among users, places, and topics

2.3.1 What is a hive plot

There is a large range of network visualization techniques that vary from simple geometric layouts (e.g. circular or parallel linear layout) to complex, dynamic, data-driven layouts (e.g. force-directed layouts as in the ForceNet tool above or inverted self-organized map layout). These techniques have their advantages, but also share a range of limitations, two of which were outlined above.

Hive plots (http://www.hiveplot.net/) attempt to address the issues described above by providing users with full control over the *visual structure* of the network layout. This allows for comparison across different networks or across time (for evolving networks), and might make the visualization easier to interpret, as its layout always matches the analyst's request.

2.3.2 Hive plots in SensePlace 2

SensePlace 2 makes use of a complex network data model that captures the multitude of entity types that exist in the project's data (users, user mentions, place mentions, hashtags, etc.) and the relationships between them. This network data model allows for flexible queries on the underlying data, making the kind of analysis done by the co-occurrence matrix component possible. Two different approaches to visualizing the network data model *itself* are made in SensePlace 2, with one using a force-directed network visualization layout (described previously) and the second one using a hive plot. An example hive plot is shown in Figure 5.

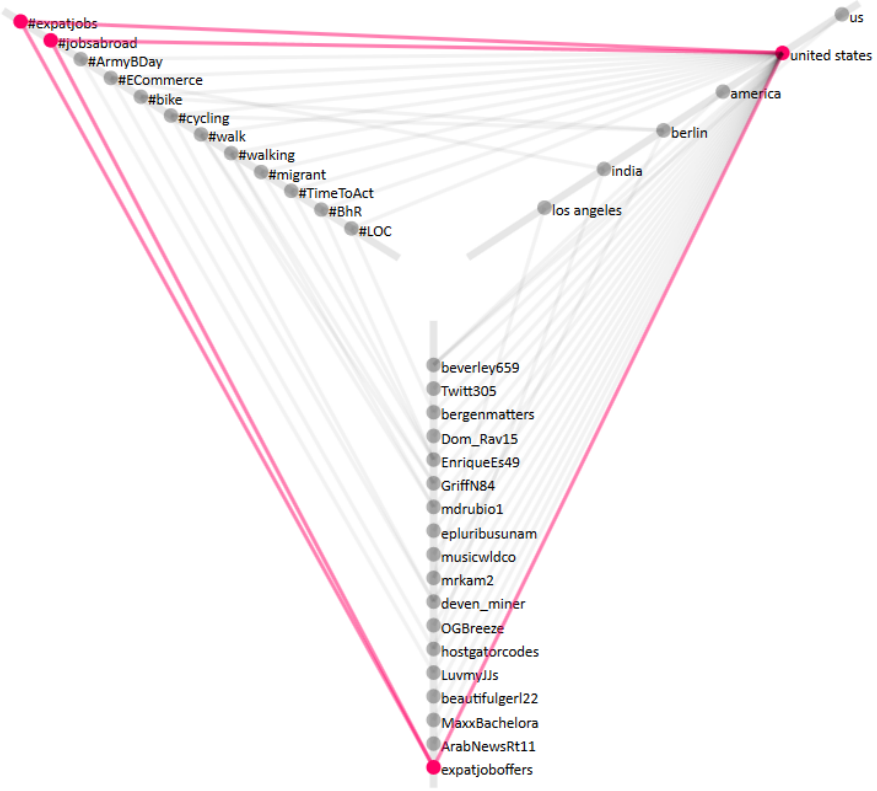

Figure 1. This Hive Plot has three axes, thus depicts three kinds of information stored in the SensePlace 2 tweet index. From upper right clockwise are axes depicting "about" places (place references in the text of the tweet), user screen names, and hashtags. The bottom user (expatjoboffers) is clicked on and we can see that this user mentioned "united states" and two hashtags: #expatjobs and #jobsabroad.

Hive plots are built by having the analyst specify a number of hive plot *axes* that they would like to add to the layout. In the illustration above, the axes seen are the place mention (about) axis (top right), user screen name axis (bottom), and hashtag axis (top left). Once the analyst adds an axis to the plot and specifies its type, it is automatically populated by the hive plot *nodes*. The nodes in the hive plot are the representation of *individual entities* found in the data model. That is, the *hashtag axis* is showing all of the *hashtags* found in the current query as its nodes, the *place mention axis* is showing all of the *place mentions* found in the current query as its nodes, etc. The thin lines drawn between the nodes are the hive plot *links* – a visual representation of a connection between these nodes that was established in the SensePlace 2 network data model. In the illustration above, the user "expatjoboffers" is highlighted on the user (bottom) axis, causing the hashtags and place mentions found in that user's tweets to be highlighted as well (upper left and upper right axis, respectively).

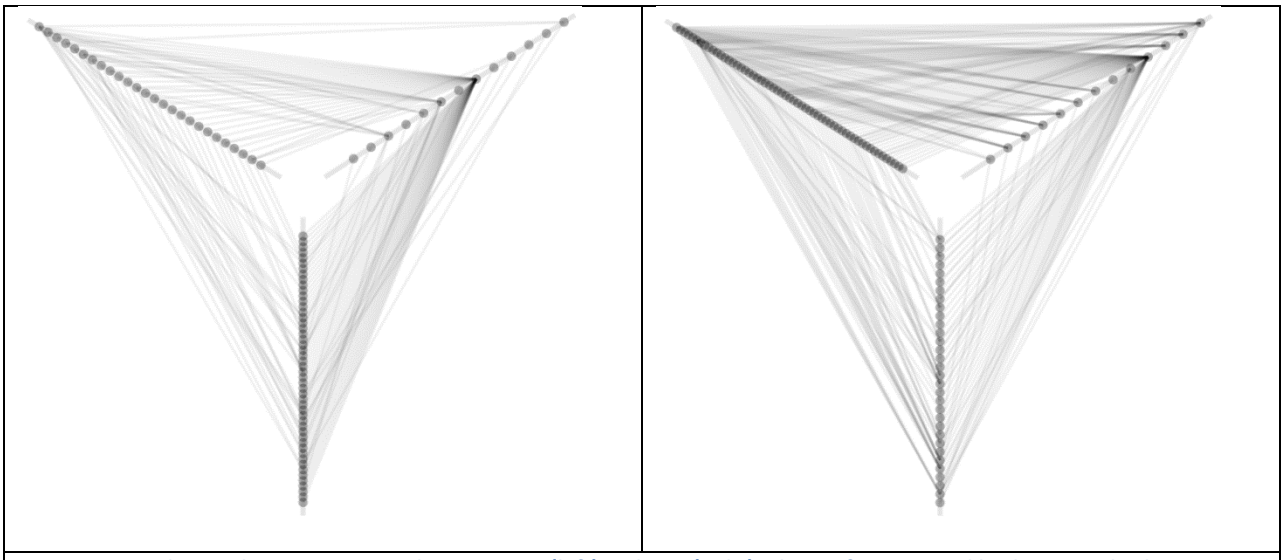

Figure 2. Paired Hive Plots to compare places: Britain (left) and Iran (right). Place references and hashtags are both more frequent in the tweets that mention or are otherwise associated with Iran.

Shown in Figure 6 are two networks that were built using the same structure outlined above (with labels turned off for link legibility) – hashtag axis in the upper-left corner, location axis in the upper-right corner, and user axis at the bottom. The network on the left is based on a query on Britain, whereas the network on the right is based on a query on Iran. The difference in the structure of the information networks describing these two conversations is obvious – the use of hashtags is much heavier in the conversation on Iran (despite a somewhat smaller number of users participating in the conversation), and the associations between place mentions and hashtags are stronger as well.

The Hive plot is fully integrated with the rest of SensePlace 2. It supports the full range of entities used in SensePlace 2 and is fully linked with the coordination framework, making it possible to both refine the hive plot by filtering the data through word cloud or a co-occurrence matrix, and to drive other visualization components by selections made in the hive plot (e.g. clicking on a node will bring up the tweets associated with that node and will cause the co-occurrence matrix to re-draw, highlighting relationships in the selected subset of the data).

The Hive plot has a very simple user interface, shown in the illustration below (Figure 7). When focusing on a particular axis, the user interface expands to accommodate axis controls, as shown.

Click on existing axes to change their properties, use scroll wheel or left / right arrows to move selected axis around.
Add a new axis? Turn labels on or off? Reset all selections? Hide or show organization nodes?
Axis type: hashtags , Axis order: alphabetic , remove axis?

Figure 3. The simple, text-based interface to the Hive plot. It includes options to add and remove axes and assign variables to those axes, turn on and off labels, and move axes into any desired order.

2.3.3 Hive plot and Task 4

The Hive plot in its current implementation is geared towards Task 3 goals – exploring connections between entities with emphasis on user connections and place mentions. The software development

work undertaken so far is, however, generic, and will be re-used in Task 4 as part of information propagation and network structure analysis.

2.3.4 Understanding groups within the networks

Hive plots can also be used in two alternative ways, each one having its own strengths and weaknesses. The first use case is to explore small networks of entities and the relationships between them. In this use case, hive plots can show labels for individual nodes, and tracking relationships between specific entities is somewhat straightforward. Label overplotting makes this use case impractical for large networks. The second use case is to explore the overall structure of large networks. In this use case, hive plots do not show labels for individual nodes, but instead make use of axes selection and node clustering / sorting to display the density of connections across different segments of the network in an attempt to identify nodes and groups of nodes with unusually high or low connectivity. In this use case, hive plots are not the most convenient tool to explore the relationships of any individual node to the rest of the nodes on the network, but, coupled with grouping / ordering algorithms, are a potent tool for showing trends (and aberrations) in strength of relationship between types of nodes.

Any network visualization (and visualization in general) introduces a certain (limited) perspective on the data it is trying to visualize. The Hive plot attempts to introduce structure to the problem of network visualization, but, beyond the simple principle of plotting nodes along rotating axes, much of the structure is left for the user to specify. The usefulness of the hive plot in relation to the specific task at hand depends on the selection of axes, their order, and the order of nodes along the axes. In our current hive plot examples, that are generated to illustrate the methods, all of these are most rudimentary -- axes correspond to entity types and do not take into account the role each node plays in the network data model (e.g. we could have an axis with nodes that have mostly incoming connections, nodes that have mostly outgoing connections, and nodes that act as intermediates, regardless of their entity types), nodes are ordered along the axes in alphabetical order, there is no clustering of nodes along the axes, etc. Thus, we have illustrated default choices for axes to depict and for ordering on those axes. The power of the method is its flexibility to adapt to particular analytical questions and problem domains.

Given the task of visually identifying clusters of entities related to a particular topic of interest, the following extensions to the method as implemented have the potential to make it easier to identify / understand groups of related entities (applied in any sort of combination):

1. Using network properties of individual nodes to assign them to axes;
2. Positioning individual nodes along the axes using their network statistics;
3. Clustering of nodes within an axis based on an external clustering algorithm (text clustering for tweet nodes, follower-followee similarity for user nodes, chronological arrangement - linear and cyclical - for nodes of any type, etc).

2.3.5 Discussion

As noted above, an advantage of Hive Plots over a standard force-directed node-link graph is that they provide the analysts with full control over the *visual structure* of the network layout. One potential advantage of this control of layout is that it supports comparison between/among places and for different times. The main weakness of the hive plot is that it enforces a partial perspective on the network data. The selection of axes types determines which nodes will be shown, and the arrangement of axes dictates what kinds of connections will be visualized.

In follow up research, we intend to make use of network properties of the data that we have (including statistics like betweenness and centrality and more qualitative characteristics, such as source / sink classification) to enhance the capacity of the Hive Plot and other network visualization tools that we have included in SensePlace 2. A caveat with using network statistics is that it is important to avoid the small sample problem, which would require the infrastructure to perform these calculations on the entire dataset at hand, and even that dataset is a small sample of Twitter over all, since we can collect a maximum of 1% of all tweets and we currently index only tweets that contain some form of location information.

3 Use-Case Scenarios to Illustrate the Methods and Their Application

The development of the Task 3 tools was driven by a formal scenario-based design approach in which use case scenarios representative of different kinds of analysis problems using different subsets of tools were developed to guide both tool design and evaluation. The scenarios include "claims" about the advantages of implementing specific features in specific ways and also list the potential disadvantages. For an overview of scenario-based design, see (Rosson and Carroll, 2002) and for an example application to development of a web-based exploratory geovisualization tool see (MacEachren et al., 2008).

3.1 Scenario 1: Emergency Management Data Center Analyst

ForceNet allows users to explore the follower-followee relationships among members of a group. As an example, we collected the Twitter user IDs for 194 humanitarian organizations and derived the Twitter relationships among the organizations (using a partially manual process described in Section 4.2 that is independent of SensePlace 2 thus far). The network of organizational connections derived was imported into SensePlace 2 and used to support this scenario.

3.1.1 Analysis of Twitter utilization by international aid organizations

Sally works as a data analyst at an emergency management center in an NGO. One of her major tasks is to help decision makers understand how key stakeholder organizations are reacting and marshalling resources with social media on any relief/aid work related to important events (disease, natural disasters, war, etc). She needs to understand how those organizations connect/collaborate with each other in social media and geographical space over time, as well as how those organizations distribute information to each other and to the public via social media.

Using the *ForceNet* tool in SensePlace 2, Sally starts with a pre-created group of organizations in anticipation of emergency management. The pre-created group includes 194 social media accounts represented as a group in the Gbuilder component. As with any group, Sally can direct SensePlace 2 to collect the group's social network information (whether they are friends or not, followers and followees, mentioned locations and users, and hashtags) and *ForceNet* can support analysis of such information from a social network perspective (Figure 8). Based on the follower-and-followee relationships among the organizations, Sally can tell that most organizations are highly interconnected among each other; this should facilitate communication among those organizations. It is also easy to identify some hub organizations within our 194 samples, such as Oxfam and UNICEF; these are hubs based on their follower-and-followee relationships, which may or may not result in being communication hubs in terms of being retweeted frequently or being mentioned frequently by other organizations[1].

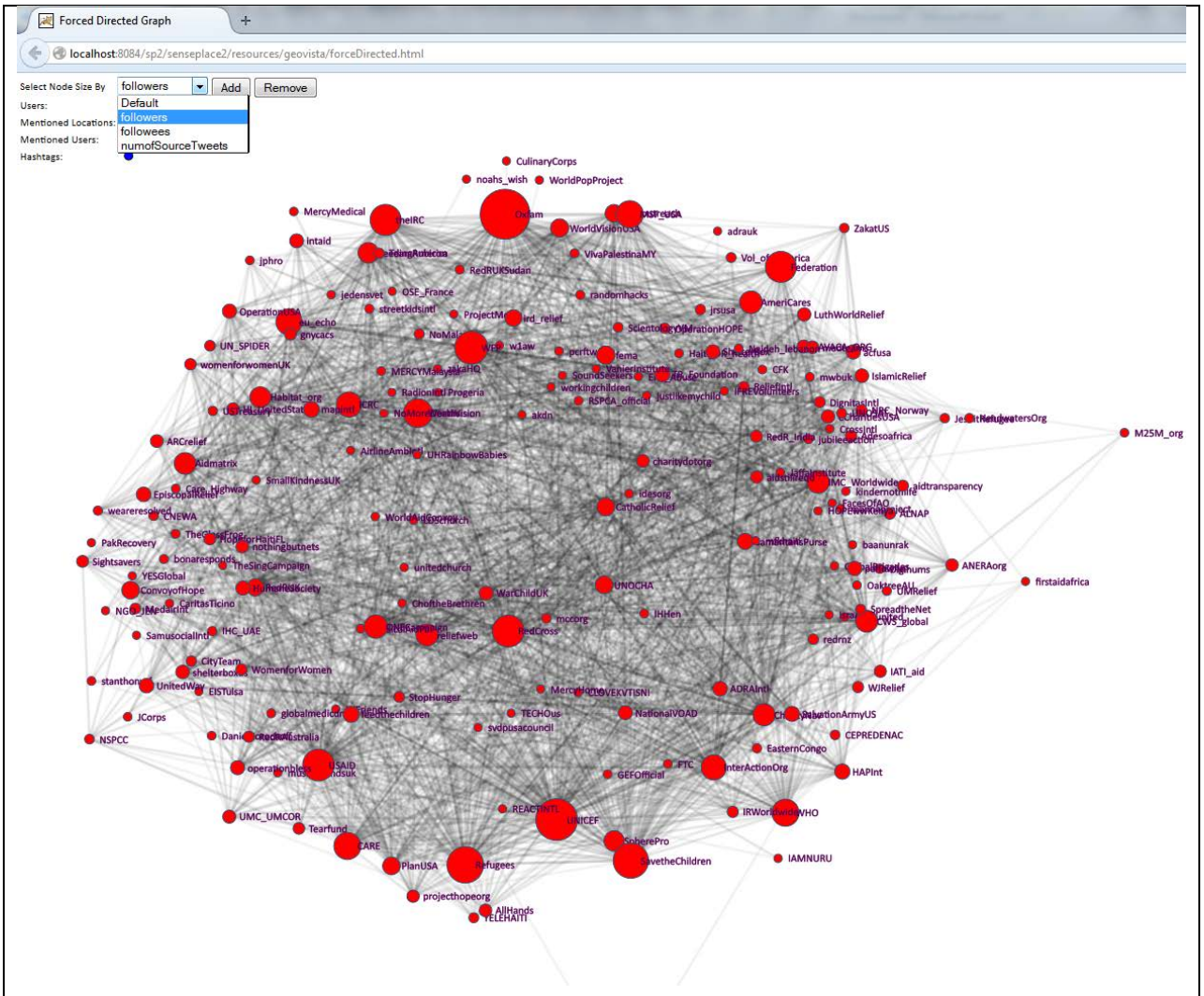

Figure 4. The follower-and-followee relationships among all humanitarian organizations in the pre-created group. The size of nodes represents the number of followers for each organization in this figure. This tool also allows users to resize those nodes by the number of their followees from the drop down menu at the top left.

[1] The analysis of message connectedness will be a focus on the next stage of this research.

Sally wants to know how the organizations in her group are reacting and marshalling resources with Twitter on current important events. From the main view, Sally executes a query on all the user IDs for organizations in the group. The main interface result is shown in Figure 9.

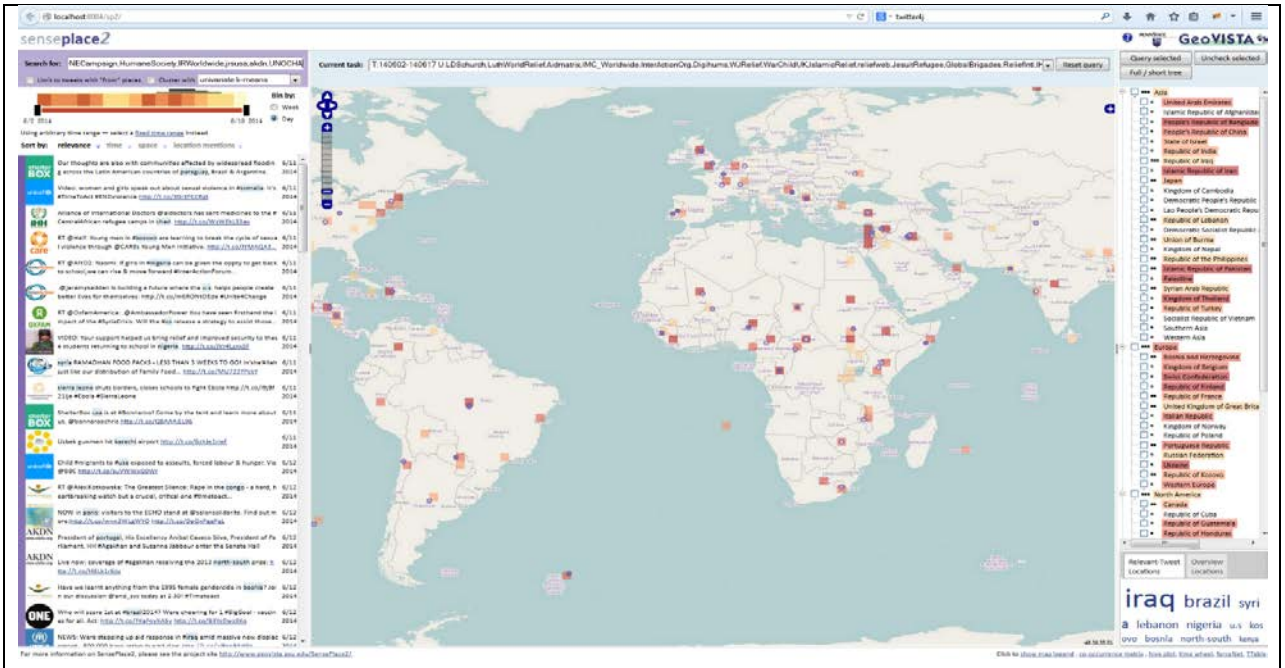

Figure 5. Query results for search by all organizations

The relationship view is depicted in ForceNet (Figure 10). In the latter view, Sally can see four nodes for organizations (red), mentioned locations (green), mentioned users (yellow), and hashtags (blue). Node size is scaled to depict frequency of each. The links among those nodes indicate that they co-exist in the same tweets. From the ForceNet View, Sally notices one mentioned location node: Brazil in green with a larger size. She clicks on the node (Figure 11), the tweets that mention Brazil go to the top of the tweet list (Figure 9).

Sally realizes (Figure 12) that due to the World Cup, many organizations mentioned Brazil in their recent tweets (seen by size of the place node in the graph), but those organizations have different foci (seen through examination of the hashtags used). Then, by selecting any node (promoting individual tweets in the TweetList, Sally examines more specifically what groups of organizations are tweeting about. For a subset of organizations focused on disease (specifically one campaign, PAHO/WHO, USAID), Sally sees that they are warning travelers to Brazil about the availability of and need for vaccines. For UNICEF, an organization that aims to promote the rights and wellbeing of children, the focus is on appeals to protect children in Brazil from child labor as the World Cup begins. The Salvation Army USA is using their posts more generally as an opportunity to support the World Cup and boost the Salvation Army's mission. Sally is also able to use the mentioned nodes (in yellow) to identify organizations that messages are directed to, with those that are larger indicating more communication directed at them.[2]

[2] Message traffic is challenging to monitor in Twitter due to the 1% sample limitations for those without access to the firehose. Strategies to monitor and analyze message traffic will be a focus of subsequent research.

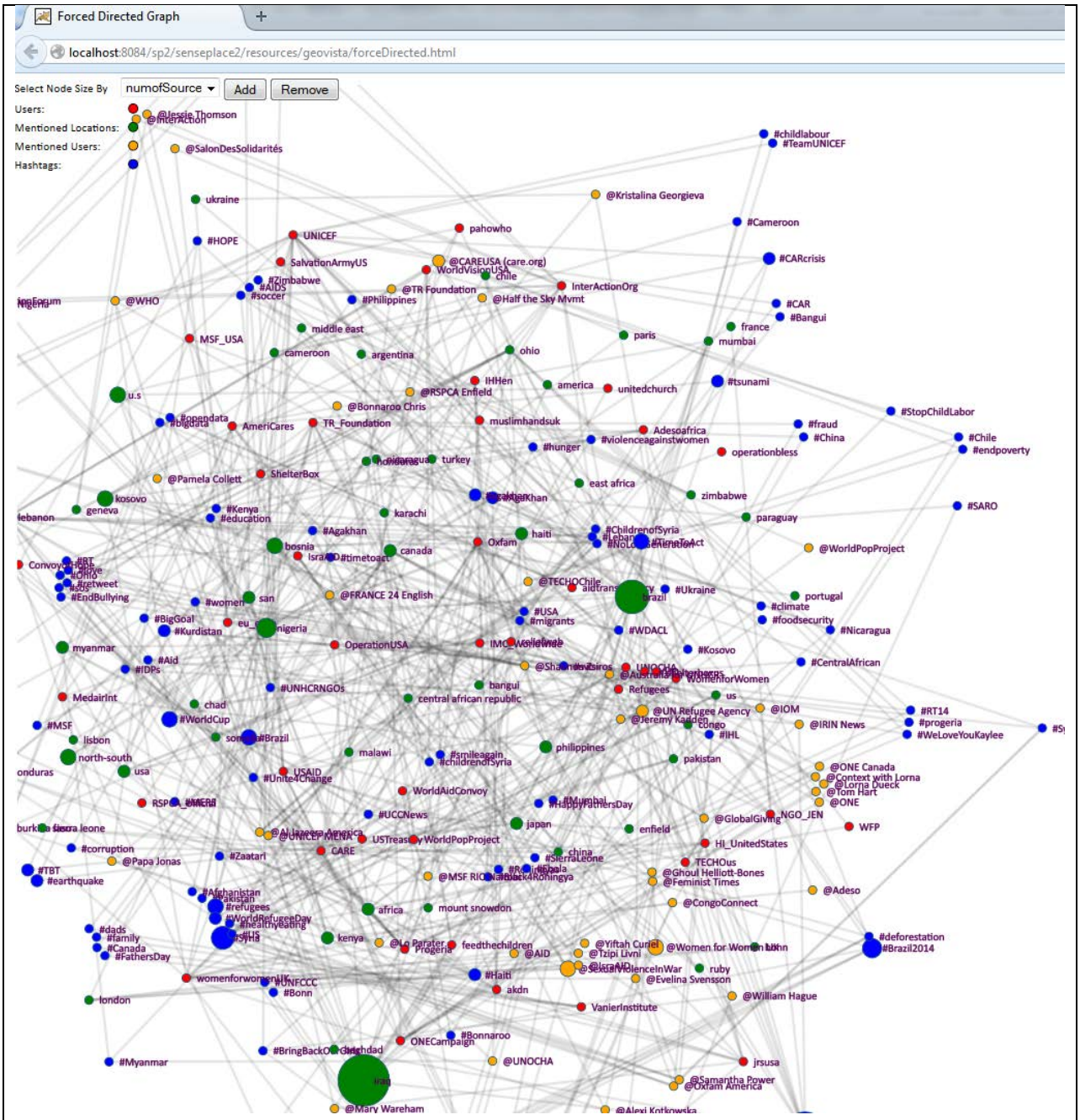

Figure 6. ForceNet with the query results for search by all organizations. The node size depicts the number of tweets in the set that refer to the places (in green), hashtags (in blue), and other users (in yellow). Iraq, Brazil, and Nigeria stand out as places mentioned in tweets by these international organizations. The hashtags of #Brazil2014 and #Syria are most frequently mentioned; the latter is unexpected in tweets that refer to the World Cup, but examination of the tweets shows that they appear to be an attempt to draw attention of the millions of individuals what are following the World Cup to the plight of refugees, particularly children. The fact that the top @ mentions are to screen names of WomenforWomenInter and SexualViolenceInWar also illustrate how international organizations are using a major international sports event to draw attention to global social issues.

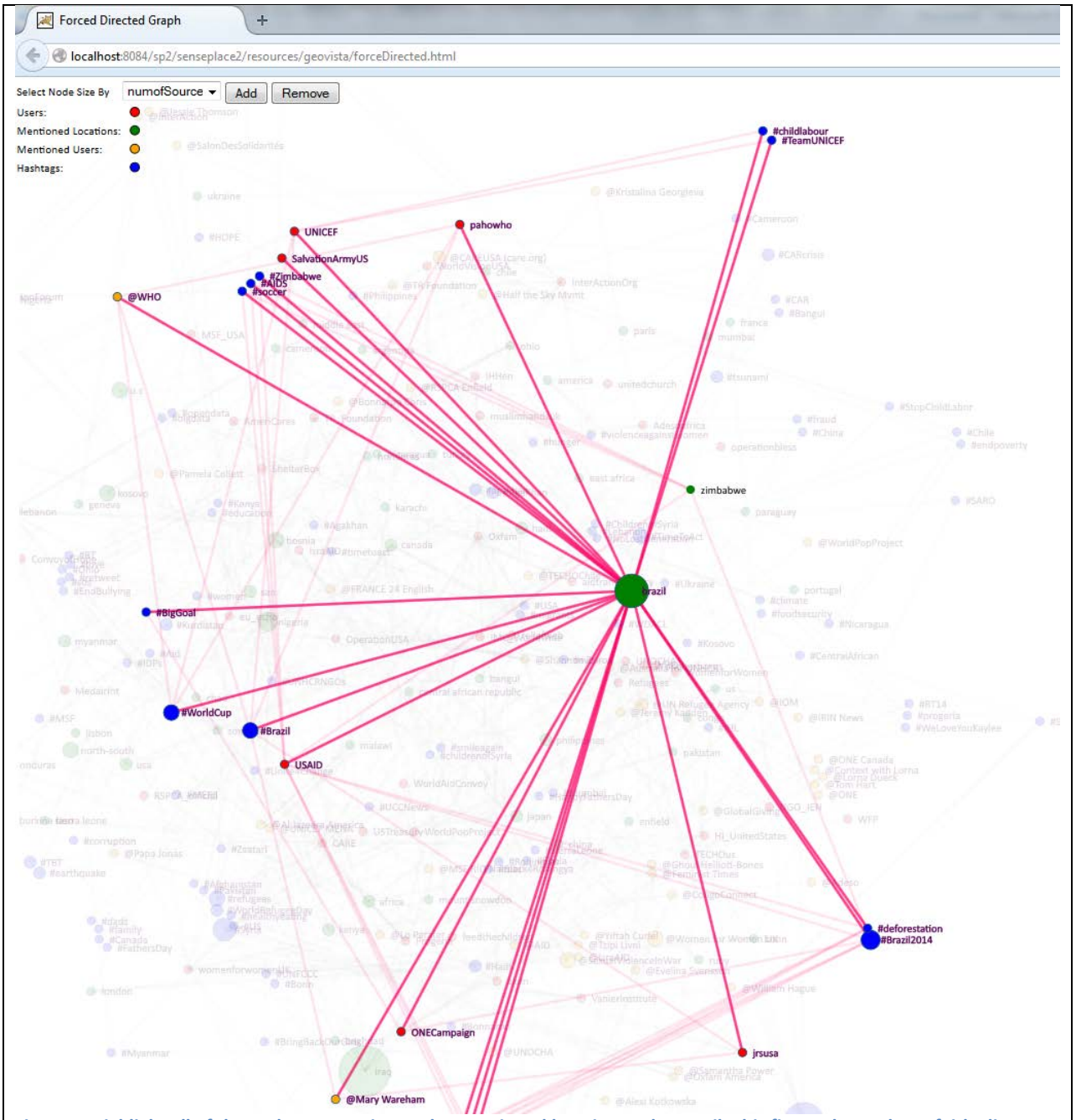

Figure 7. Highlight all of the nodes connecting to the mentioned location node: Brazil. This figure shows that a fairly diverse set of our international organization group members mentioned Brazil specifically in tweets; these include USAID, UNICEF, Salvation ArmyUS, and pahowho (the Pan American Health Organization). In addition to the direct mentions of the World Cup and its location in Brazil, the organizations are again drawing attention to non-sports issues of concern that they are apparently attempting to prompt some of the very large World Cup audience to attend to. Issues include deforestation, child labor, and AIDS.

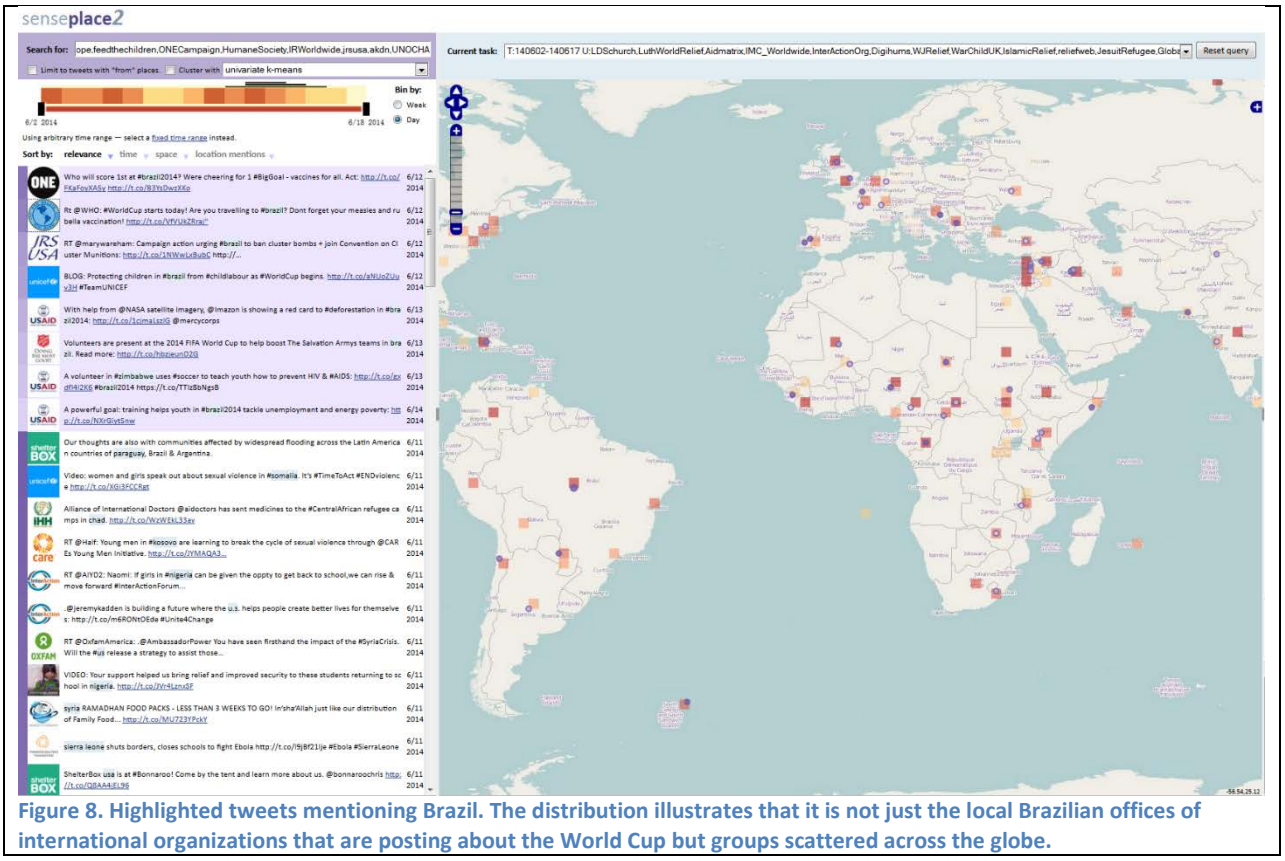

Figure 8. Highlighted tweets mentioning Brazil. The distribution illustrates that it is not just the local Brazilian offices of international organizations that are posting about the World Cup but groups scattered across the globe.

3.1.2 Lessons learned and future developments

The most important lesson learned through exploring current data with the ForceNet tool is that the tool (as implemented thus far) is able to give analysts a big picture that provides context to answer questions such as: what are the most important events, where do those events occur, who tweets about those events, and who has been communicated to because of those events. After checking the content of tweets for the scenario above, we identified a few patterns: many organization users mention themselves; the organizations do communicate with each other through mentioning other organizations; different organizations pay attention to different aspects of events according to their missions (e.g., WHO worries about potential disease transmission during the World Cup period).

We also found that some limitations mentioned above for the basic force-directed layout graph implemented limit analysis. For example, if the analyst is interested in understanding differences in perspective between organizations that mention Brazil alone versus Brazil plus other places, there is no easy way to make the comparison. Similarly, if the objective is to track conversations about an event over time, the fact that the ForceNet currently has no way to fix node positions makes such comparison very difficult. Research we have initiated on linked data should help to address the first of these issues and strategies to allow users to impose constraints on the graph layout process are intended to enhance the ability to make comparisons.

3.2 Scenario 2: Rapid Build-up of Situational Awareness

Scenario 2 illustrates the application of the Hive Plot to supporting situation awareness during an ongoing event that happens over a few days.

John is an editor in a major newspaper. He is often assigned with running "live" columns, collecting and organizing information from multiple sources in an attempt to maintain an up-to-date picture of unfolding events. Some of his major challenges include fact-checking with limited ground truth available and detection of new developments that are not yet picked up by the rest of the media community.

John starts with a couple of queries ("protest", "riot") related to the recent events, and he runs the queries both in SensePlace 2 and in a number of web search engines. John quickly finds some relevant information in SensePlace 2 results. He clicks on the user icon to bring up their full Twitter timeline in a separate window, and, having confirmed that this user appears to focus closely on the event of interest, clicks the "follow" button next to the user's name in the SensePlace 2 interface.

While browsing the web, John discovers that a number of freelance reporters tweeting about the recent events are not represented in the SensePlace 2 tweet list. John brings up the GBuilder tool and drops in the Twitter user IDs of those users, clicks "save" and continues browsing through SensePlace 2 results and the web.

Having explored SensePlace 2 and the web for a few hours, John now has built a sizeable list of users to follow (about 80 total). SensePlace 2 starts following all of the users' tweets (including retweets of their tweets by other users) the moment John clicks the "follow" button. By now some of their more recent tweets are popping up in the list as John runs a new query.[3]

Overall, the data is noisy, as is true with any social media. A few detailed reports are hidden in the mixture of high-level political forecasts and individual thought pieces and ramblings on the latest event. Iterative exploration using the SensePlace 2 dynamically coordinated views, as described in (Savelyev, 2013), makes it possible to sift fairly quickly through the noise to find a signal.

Done with building the first draft of his information network, John starts making his first write-up for the day, using the place tag cloud and co-occurrence matrix to focus on specific topics and their intersections. When reporting based on Twitter posts from individuals, John weights their contributions by whether they report from the ground (using a map of "from" locations), the illustrations they include, as well as whether he can find other users reporting the same story that are not affiliated with each other. Having worked until the early AM hours, John writes a summary piece of the events that occurred so far and heads to bed.

Human-guided construction of information network:	
+	A lot of leeway for determining what constitutes a "relevant" source of information.
+	Flexible approach that is well-fit for rapid assessment of situation.
-	Analyst can get trapped in a few sources of information that mostly cross-reference each other.

[3] Once you subscribe to a user stream, you get all of their tweets from then on, including retweets and @ replies to their tweets.

Early in the morning, John logs into the SensePlace 2 again and repeats one of his queries from the previous night ("protest"). Overnight, a sizeable volume of new tweets accumulated, with a lot of new names – John's users started interacting with a number of new sources, retweeting and sending @replies to people John doesn't recognize from the day before. John checks Twitter for the posts by the people who are being alerted to information by the individuals in his current group and if their own posts look relevant, he adds those individuals to his group and "follows" them using the GBuilder interface. He also reruns the query with these individuals in the group to forage for additional insights. Looking for a quick summary of the user activity, John brings up the *co-occurrence matrix* tool (Figure 13).

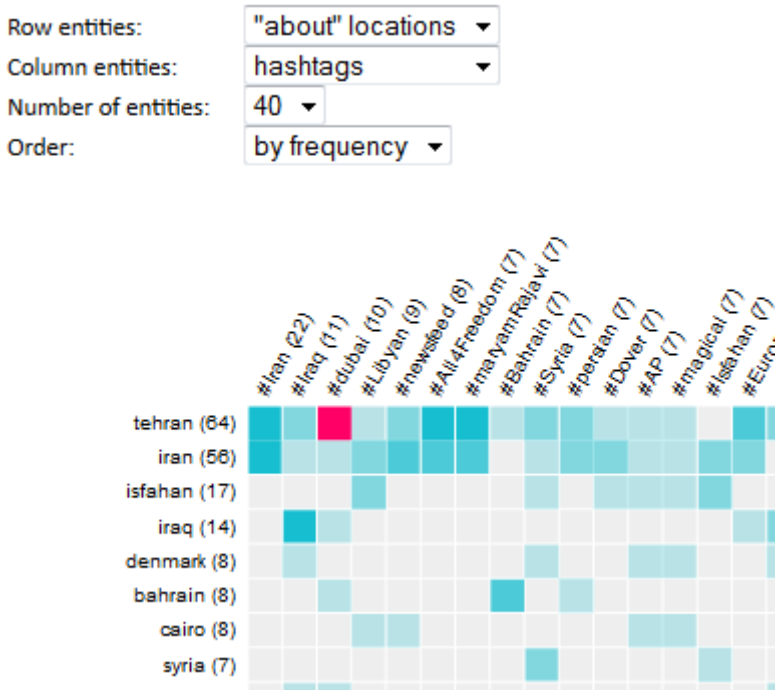

Figure 9. A section of a co-occurrence matrix showing the top 8 mentioned places and the top 15 mentioned hashtags. One cell is selected, which represents (and is dynamically linked to) 10 tweets that include mentions of Tehran together with the #Dubai hashtag.

Having explored the co-occurrence relationships between place mentions and hashtags, John is now curious to learn more about the conversation concerning two specific entities – a reference to Tehran and a hashtag #dubai. He brings up the *Hive Plot* tool to explore other dimensions of this conversation, and clicks on the cell in the co-occurrence matrix to subset the data in the Hive Plot (and other views) to tweets related to these two entities specifically along with the other entities, people, and places mentioned in tweets that contain both a reference to Tehran and the #Dubai hashtag (Figure 14). The Hive Plot offers, potentially, a much more complete picture of the conversation around this combination of entities than is apparent in the Matrix.

In the Hive Plot tool, John adds a few axes and assigns them to visualize hashtags (top and bottom axes), place mentions (upper-right axis), users (bottom right axis) and user mentions (far left axis).

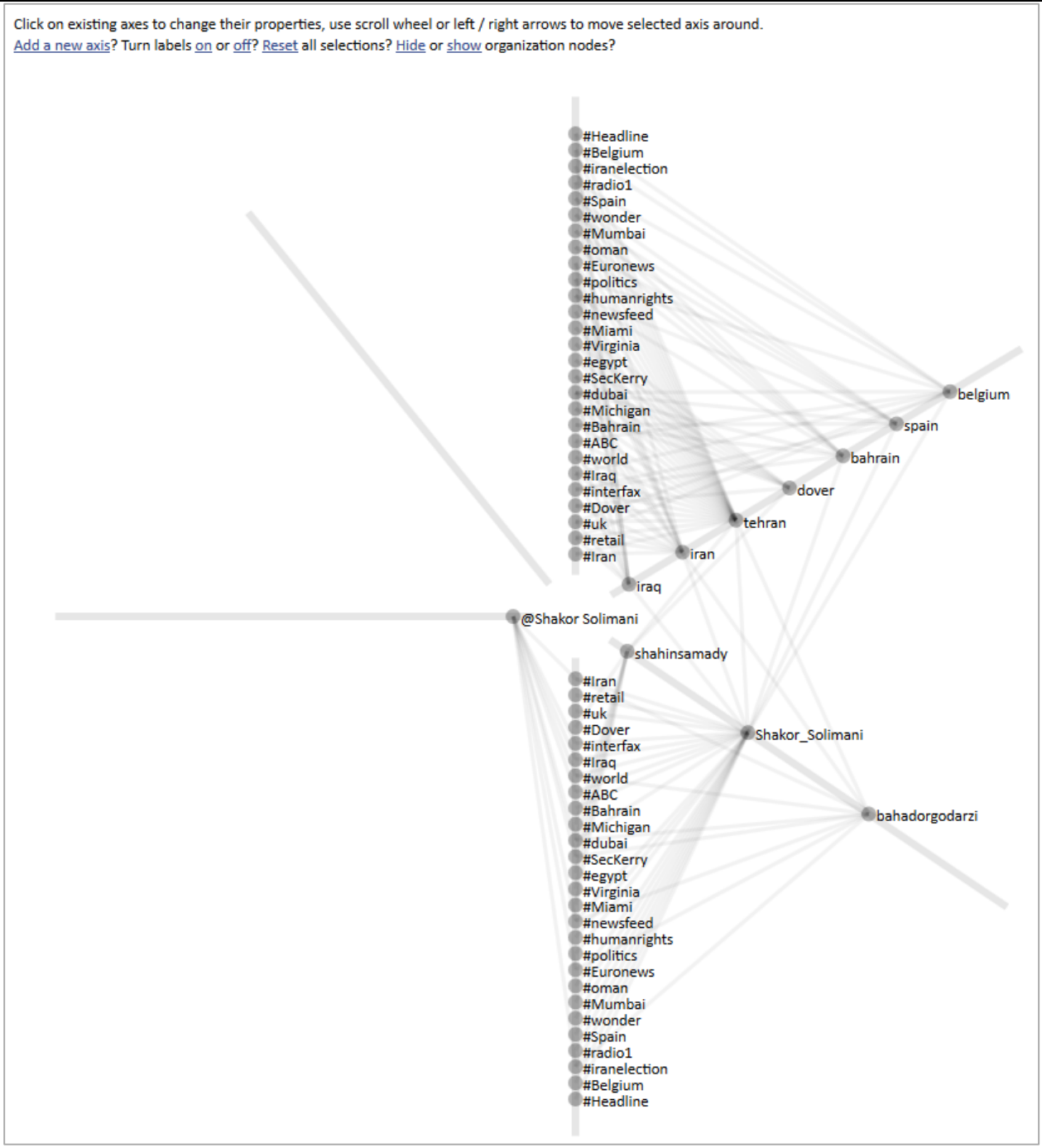

Figure 10. This Hive plot depicts the result of slecting the Tehran - # Dubai cell in the Matrix, which represents 10 tweets. The Hive Plot depicts the three Tweeters responsible for those 10 tweets, the 7 places that appear in one or more of the tweets, the 27 hashtags that are present, and the one individual Twitter account that received an "@" mention.

The links between these axes are now split into a number of segments. Links in the top right segment show co-occurrence relationship between hashtags and place mentions, similar to what the co-

occurrence matrix does. Links in the right and bottom right segments show place mentions and hashtags that are brought up in this conversation on a *per-user* basis – a more detailed characterization of the conversation than what the co-occurrence matrix can provide. The bottom left segment shows user mentions in relationship to the topics that are brought up in this conversation – John added this segment to explore if there is a lot of between-user cross-talk in his information network.

John can now see that, despite a flurry of messages in his tweet list, there are only three users that are generating all of the tweets in this conversation. It also appears that different users cover slightly different topics based on their selection of hashtags – John decides to investigate further (Figure 15). He hovers his mouse over the each of the users on the user axis in order.

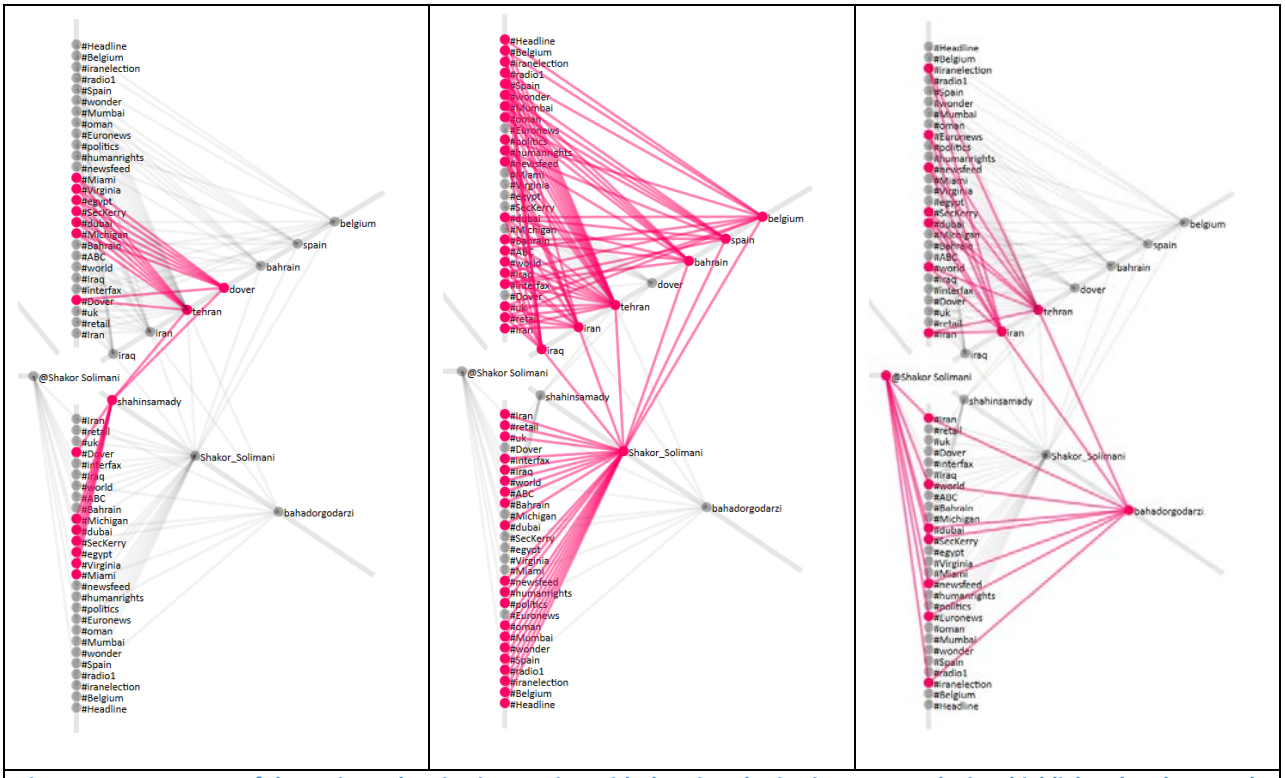

Figure 11. A sequence of three views showing interaction with the Hive Plot in Figure 14. Each view highlights the places and hashtags of one Tweeter; the last view shows that this tweeter is the only one who uses an @ mention.

Seeing the sub-networks associated with each of the users highlighted in the Hive Plot display, John comes to the conclusion that two of his users are covering a somewhat disjoint set of topics, with the second user more prolific than the first one (left and middle illustrations).

The third user (the illustration on right) appears to be mirroring a subset of tweets generated by the second user – all of his tweets contain a mention of the second user (shown on far left axis). John clicks on the third user to explore his tweets in the tweet list and, seeing that they are plain re-tweets with no commentary added, brings up the User Group tool and removes this user.

Hive Plot as a network analysis tool:	
+	Complex relationships between all types of entities present in SensePlace 2 UI can be explored.
+	Cross-component coordination makes drill-down based on observed patterns quite easy.
+	Makes analysis of information dissemination possible without relying on inaccurate text mining algorithms.
-	Very dense networks can be hard to visualize as a result of label overplotting.
-	No overview of network structure as a whole is provided, and analyst can get trapped in the few hypotheses they have of their own.

3.2.1 Discussion

The Hive Plot provides a very flexible tool for exploring direct as well as multi-step links among Tweeters, places, hashtags, people mentions, and other entities. The scenario above introduces the potential of the Hive Plot used alone or with minor input from other tools. The real power of this tool, along with others in SensePlace 2 is likely to derive from coordinated use of several tools at once.

In the scenario above, the Hive Plot was used to help an analyst carry out a largely unstructured exploration process in which it was possible to identify individuals to organize into an analytical group, to explore the discussions that members of the group are engaged in, and to add (or potentially subtract) users from the group over time. The Hive Plot can also support analysis of groups built in any other way. For example, the analyst might start with one Tweeter of interest and add their followers to create a group. Alternatively, a group might be extracted from the data computationally by identifying individuals with the greatest similarity in hashtags used, places mentioned, and/or individuals mentioned. Network metrics (e.g., centrality, betweeness) could also be integrated as a factor in group building (e.g., identifying the individuals with greatest betweeness who has tweeted about a particular topic such as protests and then building a group that consists of their most recent followers).

One potential use of hive plots that is not addressed in our scenario is the visual comparison of network structure for *high density* networks. Possible analytical approaches include building hive plots in a panel arrangement for the same query term across multiple time periods (highlighting the incremental change) or for multiple query terms in the same time range (highlighting shared nodes and just eyeballing the difference in structure, if any). Using hive plots in this fashion will allow us to get closer to identifying patterns in user behavior without becoming too reliant on clustering algorithms, which might be quite tricky to set up given our high-dimensional spatio-temporal dataset.

3.3 Summary: Understanding Networks

Overall, the initial steps toward adding network analysis capabilities to SensePlace 2 are encouraging from the perspective of highlighting the potential information about connections that is embedded in microbogs. But the initial work also illustrates the challenge of extracting useful information from massive data sources such as Twitter. Part of the challenge is the lack of methods for

integrating geographic analysis with social network analysis. Geo-social analytics is a nascent research domain with more open questions than answers at this point; for an overview of the state of the art, see the review by Luo and MacEachren (2014).

Based upon our initial tool development of the ForceNet and HivePlot tools and their application to the scenarios detailed above, our assessment is that network visualization methods, when dynamically linked to other views that depict spatial and temporal characteristics of the data, have considerable potential to help analysts confirm expected place-based connections and uncover unexpected place-based connections. But, to deal with large volumes of data such as that generated by microblog services, these interactive visual tools will need to have underlying computational processing that support real-time interaction for the data volumes involved and will need to be closely coupled with computational methods that help to identify the most promising places, times, people, things, or events to direct attention to. Beyond leveraging computational methods to support interaction and help in finding patterns, new data models beyond the traditional relational models that underlie current GIS need to be investigated. We propose that adopting linked data methods and heterogeneous network modelling approaches will provide the potential to explore connections that go well beyond standard social network analysis. We have begun to develop such approaches and anticipate being able to implement them as part of work on the next task (Savelyev and MacEachren, accepted).

4 System Extensions: Enabling Group-based Analysis & Scaling Analysis

4.1 Gbuilder: group following

The Gbuilder interface outlined above allows users to initiate queries to Twitter's Stream API with user filters (by selecting the "follow" option). This API has a rate limitation of 5000 userids per account doing queries. The stream of tweets returned will contain, for each user ID submitted: (a) tweets created by the user; (b) tweets which are retweeted by the user; (c) replies to any Tweet created by the user; (d) retweets of any Tweet created by the user; and (e) manual replies, created without pressing a reply button (e.g. "@twitterapi I agree"). The stream will not contain: (a) tweets mentioning the user (e.g. "Hello @twitterapi!"); (b) manual Retweets created without pressing a Retweet button (e.g. "RT @twitterapi The API is great"); or (c) tweets by protected users.

In addition to collecting tweets from the streaming API, we carried out a one-time collection of follower-followee information for the group of international organizations used in Scenario 1. To do this, we used the REST API. The rate limit is 180 calls within 15 mins. It required a couple of days to manually collect all of follower-followee relationships among 194 organizations. The function we use from the Twitter4j API is showFriendship (sourceId, targetId), so the input is Twitter user ids. Returned results are:


```
{
  "relationship": {
    "target": {
      "id": 12148,
      "screen_name": "ernie",
      "following": false,
      "followed_by": false
```

```
8.     },
9.     "source": {
10.      "id": 8649302,
11.      "screen_name": "bert",
12.      "following": false,
13.      "followed_by": false,
14.    }
15.  }
16.}
```

We plan to investigate the potential to build this capability into the Gbuilder component in a way that enables either one-time construction of connections or that sets up automated updates (e.g., once per month, or week, or day depending on how quickly changes are anticipated). Doing this in an automated way for multiple groups creates system scaling issues. It would be necessary to build a multi-computer crawler, each of which collected some of the information.

4.2 Scaling capabilities with distributed computing

Since we can collect up to a 1% sample of the 500 million (or more) tweets posted per day, it is a challenge to geoparse and index the tweets quickly enough to keep the system current. To address the scaling challenges, we have deployed our Solr system on the Research Computing and Cyberinfrastructure (RCC) high performance infrastructure at Penn State (https://rcc.its.psu.edu/). The system is implemented using SolrCloud, which is embedded in a Cloudera Hadoop cluster with 5 computational nodes. This configuration will improve query performance, increase scalability, and increase fault tolerance. We are in the early stages of testing this implementation. But, we anticipate that it will improve the usability and discovery capabilities of SensePlace 2. Simple search times are in the range of 100-500ms. (compared to multiple seconds previously). Our faceted on-the-fly aggregation searches have decreased from 10-20s to 2-3s. Our Twitter Table search, requiring a Solr search combining hundreds of OR combined parameters (userA OR userB OR userC OR userD...), has decreased from over 1m to 4-5s. We also now use SolrSpatial4's improved spatial indexing algorithm (using their SpatialRecursivePrefixTree field type) for faster spatial searches.

5 Conclusions and Future Work

This report presents the Task 3 outcomes from our larger research directed to supporting information foraging and sensemaking with open media. Outcomes reported focused on social-based analysis grounded in place. More generally, the research presented demonstrates the potential of linking place-focused, tweeter-focused, and social-focused analysis.

The use of scenario-based design guided development of a new ForceNet tool to explore connections among tweeters and the places and topics they reference, a new Group Builder that enables users to create user groups to follow and edit the group membership over time, and a new Hive Plot component designed for detailed analysis of relationships among tweeters and the places and topics talked about. Underpinning this work was development and implementation of a novel approach to establishing flexible links among heterogeneous entities (people, places, organizations, things) that integrated ideas from heterogeneous network modeling into a visual analytics-based interface.

A key challenge for the future, to enable methods developed to be effective for real-time analysis, is scaling. The goal is to leverage the growing availability of social media that provides insights about ongoing events by those who often have unique knowledge. But, to find the insights requires foraging through massive volumes of information quickly. Our approach to address this scaling challenge is to migrate the methods to a distributed computing environment that supports parallel processing of computations that must be repeated on millions of entities (as is the case for geoparsing place references within tweets).

A second challenge is to support coordination across many diverse tools that are required to address the complex questions involving where, when, what, and who and the relationships among them. To address this, we are continuing to develop enhancements to our coordination mechanisms and have begun to implement a novel linked data model to underpin complex queries. This work will continue through the conclusion of Task 4 focused on information propagation.

6 References

Costenbader, E. and Valente, T.W. 2003: The stability of centrality measures when networks are sampled. *Social Networks* 25, 283-307.

Dwyer, T. 2009: Scalable, Versatile and Simple Constrained Graph Layout. *Computer Graphics Forum* 28, 991-998.

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

MacEachren, A.M., Crawford, S., Akella, M. and Lengerich, G. 2008: Design and Implementation of a Model, Web-based, GIS-Enabled Cancer Atlas. *Cartographic Journal* 45, 246-260.

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., 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.

Phuvipadawat, S. and Murata, T. 2011: Detecting a multi-level content similarity from microblogs based on community structures and named entities. *Journal of Emerging Technologies in Web Intelligence* 3, 11-19.

Rosson, M.B. and Carroll, J.M. 2002: Scenario-based design. In Jacko, J. and Sears, A., editors, *Chapter 53 in The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications*: Lawrence Erlbaum Associates, 1032-1050.

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. accepted: 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.