

TABLE OF CONTENTS

- I. ABSTRACT**
- II. NARRATIVE**
- III. ENVIRONEMNTAL SCAN**
- IV. WORK PLAN**
- V. REQUIRED PERSONEL**
- VI. EVALUATION PERIOD**
- VII. PROOF OF CONCENT**
- VIII. BUDGET**

NEH Application for Digital Humanities Start-Up Grant for NETHUM

I. Abstract:

We seek a grant to create NETHUM (Networked Humanities), an archive of data and visualizations of social network communications in humanities research. The NETHUM website will include data for visualizations, visualizations of social media communications during natural and man-made disasters, and technical documentation for scholars who want to create their own visualizations.

Our project first seeks to visualize communication during the Boston Marathon bombings related to Sunil Tripathi, the man falsely identified as one of the bombing suspects via social media. In creating visualizations of communications during the Boston marathon bombings, we can more easily make visualizations of real-time natural or man-made disasters in the future, informing the general public, as well as humanities researchers, on resources and news related to tragedies. By creating an archive of data and visualizations, NETHUM will provide scholars with an open-access resource of data for their own work.

Statement of Innovation:

While many digital humanities now use Mathematica and Gephi to create their own visualizations, there is no central database that aids collaborative research focused on creating visualizations. Not every scholar needs to create his or her own data sets if there is a central database that can help provide the needed information. NETHUM will be a current, dedicated archive of data that helps humanities scholars produce innovative research using visualizations.

II. Narrative:

We seek a \$20,170 grant to create NETHUM, an archive of data and visualizations of social network communications in humanities research. While the Twitter API recently changed, limiting the affordance of collecting data from hashtags, it will soon resolve this issue. Our proposal demonstrates the use of visualizing social networks based on what is currently available using Twitter API, and advocates for the ways that social media visualizations are useful for humanities researchers.

Our project first seeks to visualize communication during the Boston Marathon bombings related to Sunil Tripathi, the man falsely identified via social media as one of the bombing suspects. By visualizing how conversations about Tripathi unfolded, we can learn how information was miscommunicated to the general public.

As Twitter and other micro-blogging platforms become more widely used, they increasingly serve as a valuable source of information for reporting disasters or other time-sensitive circumstances. After the Boston marathon bombings in April of 2013, the FBI called on the public to participate in the search for the perpetrators of the crime. Tragically, however, the use of social media led to the false accusation that twenty-two year old Tripathi, a Brown university student, was the terrorist who bombed Boston. Reddit reported Tripathi as a suspect and, shortly after this report, his name reached the Twitter top trends list. His body was later found dead by the members of the Brown university crew team. While social media cannot be blamed for the communication that seemingly link to Tripathi's death, the use and interpretation

of social media posts in this tragedy demonstrate the significance of understanding social media posts related to events.

Yet, while social media has become increasingly well used, there are relatively few visualizations that demonstrate how communication is spread across social networks. Such information is crucial to understanding how users communicate and to bettering the structure of social media sites. Shortly after Tripathi's name was identified as a suspect on Reddit, his name reached the Twitter "Top Trends" list, which further spread his name across social media networks. It is important to understand how the structure of Twitter, in placing his name on the "Top Trends" list impacted how the story of Tripathi as a suspect spread to millions of viewers.

While users may not trust Reddit, the spread of news on Twitter made it difficult for users to trace the source of information declaring Tripathi as a suspect. Journalists from leading news organizations, including the *New York Times*, retweeted the news informing the public on Tripathi. While reputable news organizations have largely revised their social media use guidelines since the bombings, during the Boston tragedy there was no other news source providing information about the source of information on suspects. The ability to quickly make visualizations that demonstrate sources of information behind the retweets of Tripathi's name would have been hugely significant in helping users determine the accuracy of information.

Users are still posting on hashtags linked to the Boston marathon bombings, such as #freejahaar, which emphasizes the continued relevance of the event, but most importantly visualizations of the bombings can serve as a model for how visualizations can impact future events. Our project will offer technical documentation for how researchers can easily use Gephi and Mathematica to produce visualizations of social media posts as well as visualizations of literary characters. While Gephi and Mathematica are continually used in Engineering and Computer Science sectors, their use is significant in the humanities as well. What prevents humanities researchers from creating visualizations are the lack of models of relevant research and the documentation to advise researchers on how to make their own.

NETHUM will provide documentation to make visualizations easier, but it will also archive valuable data and create significant visualizations of real-time events. The archived data will become a free, public resource for scholars and the general public.

III. Environmental Scan:

The significance of this project overlaps with the following types of projects:

- a) **Scholarly Social Networks:** By making visualizations of communication and relationships between people easier, our project intersects with online resources such as UCSB's RoSE, Academia.edu, and WorldCat Identities Network.¹ UCSB's RoSE project "shapes bibliographical resources into a social-computing model presenting the past and present as one living 'social network'" (UCSB RoSE Proposal). Academia.edu offers public profiles of academic scholars, and affords analytics of profiles and papers. WorldCat Identities Network allows users to explore the

¹ RoSE: <http://rose.english.ucsb.edu>, Academia.edu: <http://www.academia.edu/>, WorldCat Identities Network: <http://www.oclc.org/research/news/2011/07-28.html>

interconnectivity and relationships between WorldCat Identities. While RoSE, Academia.edu, and WorldCat Identities Network show the interrelation of humanities knowledge in different forms, our project will help other scholars study the interrelation of communications between scholars and the general public on social networks.

- b) Social Media Archiving:** NETHUM archives social media hashtags from significant events and makes the data public access. Currently, there are resources such as ArchiveSocial and Webmonkey that allow individuals to archive social media messages for business or individual purposes, but they do not offer public data or models of visualizations for humanities projects.² The Library of Congress is also currently archiving tweets, which will serve as an invaluable resource for NETHUM. NETHUM offers interested parties a public resource of data, visualizations of that data, and documentation to help scholars produce their own data.

IV. Work plan

Our plan includes four primary tasks:

1. **Data-mine and export data** from Twitter, Facebook, and any other social network that proves important to archiving social media data and visualizations of natural or man-made disasters.
2. **Create an archive of data-mined data** on the open-access NETHUM website.
3. **Create visualizations and NETHUM website.** Visualizations will be created on events that were significantly impacted by social media and made in Mathematica and Gephi.
4. **Use Scenario Studies/Evaluation.** We plan to run usability tests that include introducing the site a) to undergraduate classes b) in professional conferences c) in graduate student dissertation work. The use scenario studies will involve participate interviews, questionnaires, and analysis of user-created visualizations from the data and visualization archive.

V. Required Personnel

Staff: Ashley Champagne (PI, 30% effort) has experience creating visualizations in Gephi and Mathematica and managing collaborative projects. Her current dissertation topic involves studying science research and writing for the public, with particular emphasis on social media use. **RA's** will be recruited from graduate students in the UCSB Media Arts & Technology Program and English Dept. (and possibly other humanities dept.). They will scan works that are not already in electronic form, and export the large amounts of data for the NETHUM website.

VI. Evaluation Period

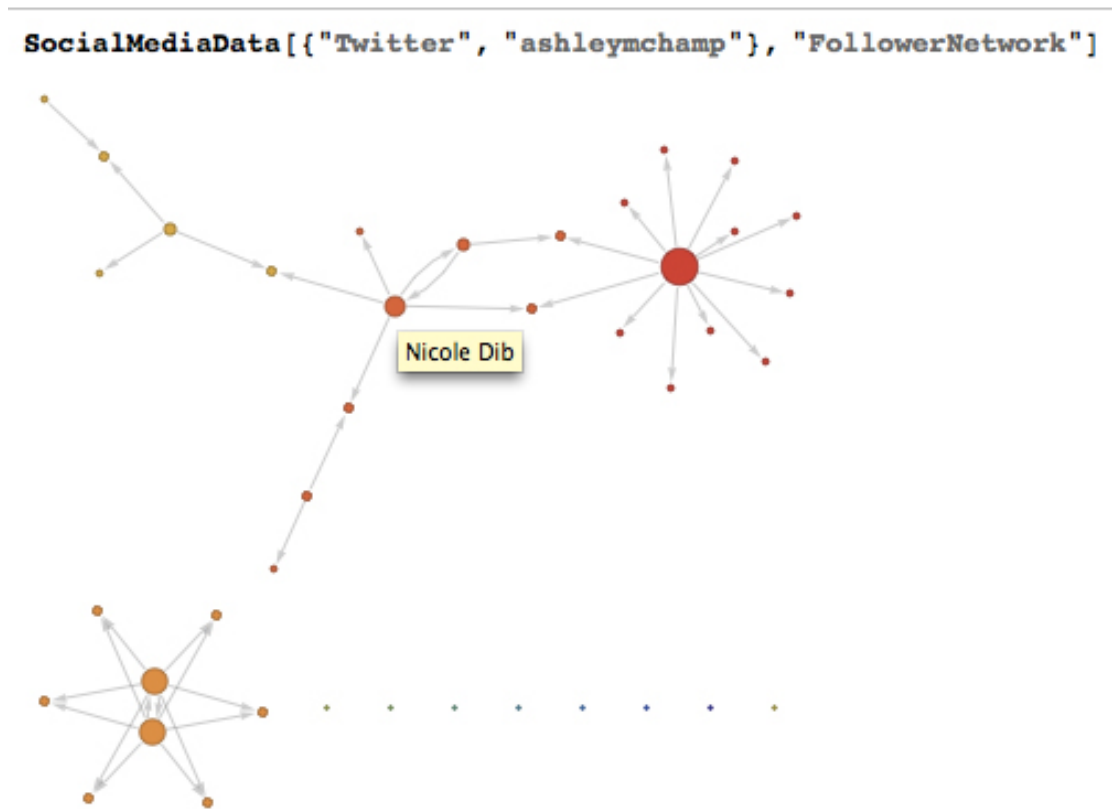
After gathering a significant amount of data, making necessary visualizations, creating the first draft of the website, and writing some of the technical documentation on creating visualizations, we will evaluate the product.

² Archive Social: <http://archivesocial.com/>, Webmonkey: <http://www.webmonkey.com/>

We will distribute the link to the website to a) undergraduates b) graduate students working in the digital humanities c) digital humanities scholars. We will ask them standard usability questions, such as: 1) How easy is it to use the site? 2) How relevant are the visualizations and data to your own work? 3) Is it easier for you to create visualizations from the site based on the models? 4) What are your suggestions for making the site easier to use?

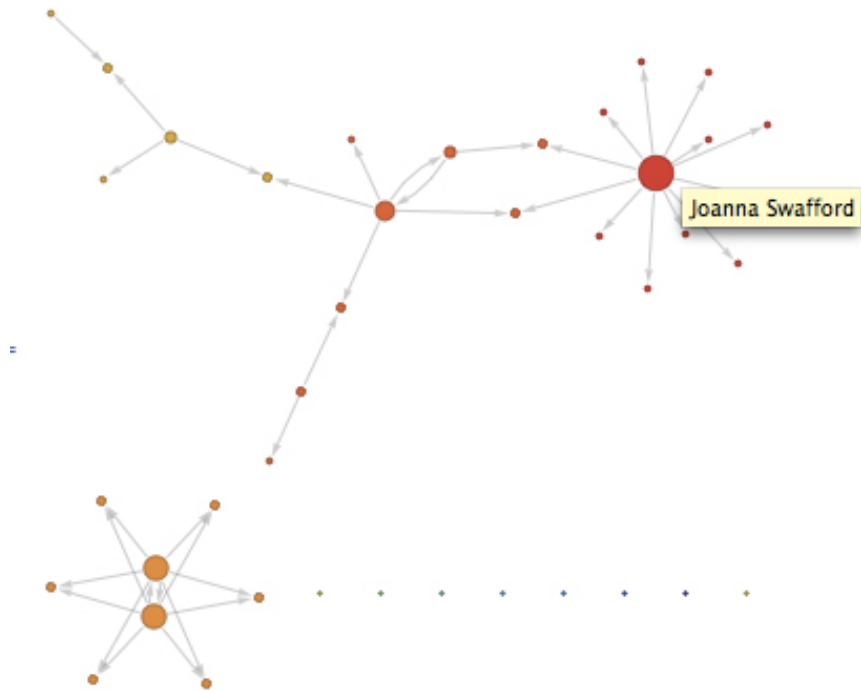
VII. Proof of Concept

Due to the Twitter API changes this year, our Proof of Concept is limited. The following images are made in Mathematica from Twitter data:



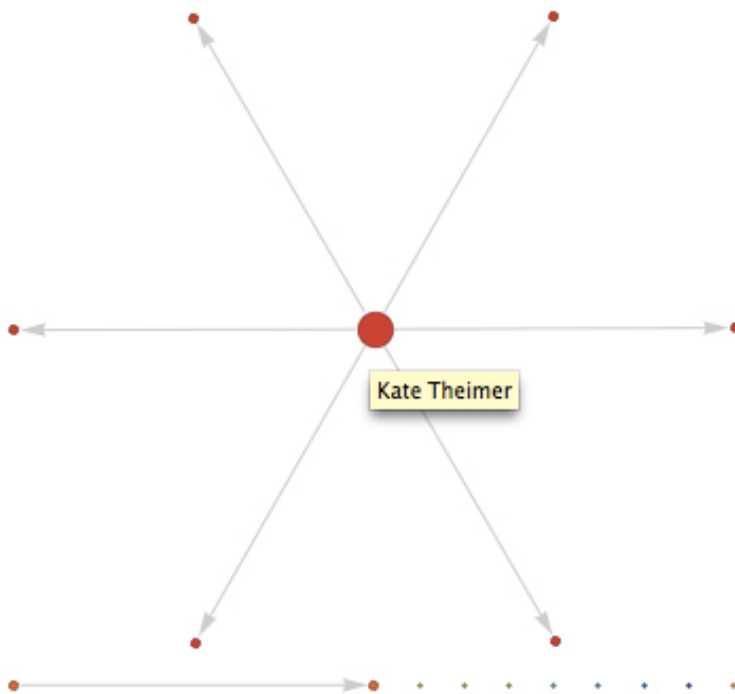
Visualization 1: The following is a visualization of Ashley Champagne's Twitter Follower Network. Each follower is a node on the visualization; the larger the node, the more connected within the network each follower is. Nicole Dib, for example, is well connected within this network. By moving the cursor over each node, Mathematica shows the user the name the node represents.

```
SocialMediaData[{"Twitter", "ashleymchamp"}, "FollowerNetwork"]
```



Visualization II: The following is another visualization of Ashley Champagne's Twitter Follower Network that demonstrates that Joanna Swafford is the most connected follower in the network.

```
SocialMediaData[{"Twitter", "alanyliu"}, "FriendNetwork"]
```



Visualization III: While current Twitter API limits the user's ability to visualize other people's networks, it will soon resolve this problem. Mathematica, however, does currently allow for visualizations of the "Friend Networks" of Twitter users. Below is Alan Liu's Friend Network, where Kate Theimer is the most connected friend in the network.

After the Twitter API resumes the affordances it had in the past, we will be able to create visualizations like the above ones that demonstrate how communications are spread among users and how users are connected.

VIII. Budget



OMB No 3136-0134
Expires 7/31/2015

[click for Budget Instructions](#)

	Computational Details/Notes	Hourly Pay	Project Total
1. Salaries & Wages			
RA for Social Media Data Collecting Phase	160 hours (4 weeks work at 40/week)	\$15	\$2,400
RA for Archiving Phase	80 hours (2 weeks work at 40/week)	\$15	\$1,200
RA for Visualization and Website Design Phase	540 hours (14 weeks work at 40/week)	\$25	\$13,500
RA for Scenario Studies/Evaluation Phase	80 hours (2 weeks work at 40/week)	\$15	\$1,200
2. Fringe Benefits			
RA for Social Media Data Collecting Phase	0.015		\$36
RA for Archiving Phase	0.015		\$18
RA for Visualization Phase	0.015		\$202.50
RA for Scenario Studies/Evaluation Phase	0.015		\$18
5. Supplies & Materials			
Mathematica License	Standard License		\$1,095
6. Services			

7. Other Costs			
DIY Scanning Equipment			\$500
8. Total Direct Costs	Per Year		\$20,169
10. Total Project Costs	(Direct and Indirect costs for entire project)		\$20,169
11. Project Funding	a. Requested from		
	NEH		\$20,169
			\$0
			\$20,169
12. Total Project Funding			\$20,169