**Table of Contents**

- Introduction ........................................................................................................... 3
- Key Ideas .................................................................................................................. 6
- Next Steps ............................................................................................................... 9
- Short-Term .............................................................................................................. 9
- Long-Term .............................................................................................................. 9
  - Discussion Points I: Project Strengths ................................................................. 9
  - Discussion Points II: Issues/Challenges ............................................................... 10
  - Discussion Points III: Suggestions .................................................................. 10
  - Resources & Further References ..................................................................... 11
- Conclusion ............................................................................................................... 12
- Report Prepared by: .............................................................................................. 12
- Participants ............................................................................................................. 13
  - National Visiting Committee (NVC) ................................................................. 13
  - Participants (Specialists) ................................................................................ 13
  - San Diego Participants (Specialists) ................................................................. 14
  - SDSU Research Team ...................................................................................... 15
  - Student Assistants ............................................................................................ 16
  - Agenda ............................................................................................................... 17
**INTRODUCTION**

This annual specialist meeting (workshop) was funded by an NSF-CDI project (#1028177, *Mapping Cyberspace to Realspace: Visualizing and Understanding the Spatiotemporal Dynamics of Global Diffusion of Ideas and the Semantic Web, [http://mappingideas.sdsu.edu](http://mappingideas.sdsu.edu)*). The goal of this workshop was to foster the multidisciplinary collaboration in related research disciplines, including geography, linguistics, computer science, political science, and communication. The two-day workshop (August 7 and August 8, 2013), organized by San Diego State University, brought together twelve specialists drawn from the many disciplines with interests in these issues. The workshop assessed the current state of the art, identified and prioritized a research agenda, and continued to develop a research community of collaborating scholars working on these issues.

![Figure 1. The Knowledge Discovery in Cyberspace Research Framework.](image)

The main theme of the 2013 workshop is "**Knowledge Discovery in Cyberspace and Big Data**". To analyze and visualize social media messages and web content effectively, the CDI project research team introduce a new research framework, called Knowledge Discovery in Cyberspace (KDC) (Tsou and Leitner, 2013). The purpose of KDC is to use various tools (machine learning, computational linguistics, GI/Science, spatial statistics, geovisualization, data/information mining, etc.) to scale up our research capability of handling millions of records and information items available in social media or web pages. By developing highly scalable information mining algorithms, visualization tools, and statistical methods, scientists and researchers can discover new patterns and new
knowledge from very large numbers of message records and human communication networks (social networks). One unique characteristics of KDC is a triangular knowledgebase consisting of place, time, and messages (Figure 1). The three components of KDC are interdependent and human-centered, which means they are originally defined or created by human beings (messengers).

Twelve specialists and scholars were invited by the NSF project team, consisting of PI Ming-Hsiang Tsou (Geography) and Co-PIs Dipak Gupta (Political Science), Jean Mark Gawron (Linguistics), Brian Spitzberg (Communication), and Li An (Geography). The meeting included plenary presentations by invited experts, and ample time for small-group discussion of the issues related to the main theme, “Knowledge Discovery in Cyberspace and Big Data”. Specific research questions addressed in the workshop included:

1. What are the characteristics of big data and social media messages? Are social media messages a unique type of big data?
2. What type of big data research frameworks can help knowledge discovery in cyberspace?
3. Should the data speak for itself in “Big Data” analyses? How accurate can we be in understanding the mood of the society? How can we improve our understanding?
4. Do (can) big data drive theory, or should (can) theory drive big data?
5. Is meme theory falsifiable (or, are any theories in big data scholarship falsifiable), and if so, how should scholarship proceed?
6. Given time-series of spatial data (big data), what statistical methods or measures can be used to “dig out” the space-time pattern(s) in the data, and explain/predict such patterns?
7. If we have point data collected over Internet, how to create meaningful continuous information landscape(s)?
8. Can sentiment analysis in its present state of advancement find meaningful changes in group sentiment and propensity to action over large data sets spread out over time?

This report highlights key ideas addressed in the specialist meeting, discussions and opinions from the participants, and the plenary suggestions for the next steps of this CDI
project, including short-term actions and long-term strategies. Summarized discussion points and reference resources are also listed at the end of this report.

Sincerely,

Ming-Hsiang (Ming) Tsou
PI of the NSF-CDI Project, Mapping ideas from cyberspace to realspace.
Professor, Department of Geography.
San Diego State University.
**KEY IDEAS**

**Establish a New Trans-disciplinary Research Framework**

Many scientific research domains and disciplines are in the paradigm shift towards a data intensive, also known as big data, era. The data and information that we are working with comes from various sources and requires wisdoms from domain experts to reveal insights and knowledge to solve our research questions. Our CDI research team has been developing a new trans-disciplinary research framework called “Visualizing Information Space in Ontological Networks (VISION)” (Figure 2). This research framework incorporates pieces from research question grounding, data collection, data visualization, analysis, and knowledge discovery. This promoted framework can serve as the bridge of bringing domain expertise from different disciplines together to a trans-disciplinary research project.

*Figure 2, the VISION research framework*

This core idea behind the framework emphasized in Knowledge Discovery in Cyberspace (KDC) and its three components: time, place, and messages (see Figure 1). The three components of KDC are interdependent and human-centered, meaning that they are originally defined or created by human beings (messengers). For example, a social media message may include a place component: San Diego (geo-tagged coordinates by built-in GPS devices or from user profile), a time component: 08:13:15 AM, 08/11/2013 (when the message was posted or forwarded), and a message component: “San Diego old town is so beautiful”. All three components are highly interdependent and should be analyzed together.
Create Effective Visualization and Analysis Methods for Representing Dynamic Changes of Information in Cyberspace

The CDI project has been adopting the current big data trend and data-intensive scientific discovery theme is many of the research works in the last few years. However, at the same time the unprecedented volume and velocity of how data is generated by user in cyberspace makes visualizing and analyzing information in cyberspace a great challenge. Thus, how to represent the dynamic change of information in cyberspace was one of the key ideas during the two-day workshop. To address the possibility, several invited specialists presented some of their works as examples of a web-based dynamic visualization tool, for example, the CyberGIS Gateway introduced by Shaowen Wang (Figure 3, left) and the HealthMap presented by Sumiko Mekaru (Figure 3, right). These two visualization applications illustrated an important aspect of creating effective visualization and analysis methods for representing dynamic changes of information in cyberspace: (1) the use of spatial analysis methods with high-performance computing environments, and (2) the design of intuitive web-based application that allow fast and easy ways for users to visualize the studied phenomenon.

Figure 3, two examples of web-based information visualization applications (from Dr. Shaowen Wang’s and Dr. Sumiko R. Mekaru’s presentations).

Build Domain Ontology and (Provocative) Event Knowledge Bases for Social Media and Cyberspace Activities

As described earlier when introducing the three components of the Knowledge Discovery in Cyberspace (KDC) concept, message is one of the important components and the meaning of message content may vary depends on the author, which media channel, and under what conversation context, which all lead to the dynamic of message and the challenge of decoding information in cyberspace. Several discussants in the workshop discussed the need of building domain ontology for analyzing the meaning of social media and web contents. For example, hashtags (#) have been widely used to understand the diffusion of certain ideas and memes through the social network on Twitter, however, the meaning of each hashtag may change because of the change of context on the topics. Furthermore, the same hashtag can be used by pro-topic user as a tag of identification or used by con-topic users in a sarcastic fashion, which results in huge challenge for linguistics analysis on information in cyberspace. Thus, how to build domain ontology
and (provocative) event knowledge bases was one of the major key topics of the workshop discussion. Without deep understanding of the norm of studied event and use of terms, the processes of decoding web content and cyberspace activity may lead to incorrect analysis results and explanations.

**Develop theoretical model(s) for the spatiotemporal diffusion of ideas**

One of the most important problems of formulating theory related to memes is bridging the micro to the macro features of society. How to develop the theory that can adequately explain how micro structures of individual-to-individual message exchange influence or translate to macro societal structures remains the major challenge. Marco societal structure includes but not limited to social status, social movements, institutions, policy formation, and culture. A Multilevel Model of Meme Diffusion (M³D) is proposed by the project to incorporate a flexible, scalable, and potentially predictive model of memetic fitness (Figure 3). In the setting of this model, six major nested levels of variables likely to influence or reflect meme diffusion: (1) meme(s), (2) the individual sources’ competence, (3) the various social network structures through which the meme is diffused, (4) the societal contextual factors and (5) geospatial/technical factors, and (6) the criteria of meme fitness and the conceptual criteria indicating the practical outcomes of meme diffusion, which feed back into the original system levels. This proposed M³D model is expected to be the analysis basis for CDI projects research on information diffusion in cyberspace, and also can contribute to the broader scientific discipline in understanding the fitness and diffusion of memes.

![Multilevel Model of Meme Diffusion](image)

*Figure 3, The Multilevel Model of Meme Diffusion (M³D).*
NEXT STEPS

Our specialists identified a few short-term actions (within one year) and long-term strategies (3-4 years) for related social media, communication and information diffusion, and inter-disciplinary research.

SHORT-TERM

1. Apply the proposed VISION framework to web contents and social media research.
2. Developing a discrete-time stepwise regression and consistency index to determine the relationship between real and cyberspace.
3. Apply linguistics and natural language processing methods to the flu and disease tracking works.
4. Test the proposed M3D model with case study.
5. Examine and compare different visualization methods for web contents and social media information.
6. Find collaboration opportunity with other related projects.

LONG-TERM

A. Create the big data knowledge discovery initiative for new research centers
B. Collaboration with other research groups or institutes.
C. Apply the proposed M3D communication to future research.
D. Determine the relationship between cyberspace events and real space events.

Discussion Points I: Project Strengths

- The combination of geography, communication, linguistic, and social science expertise.
- Developing a discrete-time stepwise regression and consistency index to determine the relationship between real and cyberspace.
- The proposed trans-disciplinary framework.
- The public health surveillance can possibly be a good predictor for disease outbreak.
- Diversity of the group provides different aspects from different projects.
- The participants provided excellent presentations and discussions.
- Domain expertise from communication (Spitzberg) and social science (Gupta) add qualitative research aspects to explain the project findings.
- The workshop presented a great coverage of theoretical and application approaches to working with big data.
Discussion Points II: Issues/Challenges

- Big Data and Social Media
  - Big Data has problem of unmeasured biases and may not be generalizable.
  - Big Data has unprecedented data sets; but requires platforms, which can be efficiently and correctly organized.
  - Keyword search may not be optimal because language is highly dynamic.
  - Dealing with noise in the data, which is now based on keywords
  - Noise and robot tweets are hard to detect.
  - Privacy issue and the data redistribution policy with social media sources.
  - Determining value of Big Data

- Data Visualization
  - Representing activities of users and interactions in a way that can be analyzed and visualized.
  - Making the process more automatic or connected to other necessary tools for analysis.
  - Translate the information shown in static maps into online interactive applications.
  - Building web-based visualizations that implements high-performance computing to support scaling

- Sentiment and Spatio-Temporal Analysis
  - Sentiment Analysis is very difficult in short text
  - Privacy issues, and understanding issues.
  - How to determine the relationship between real space and cyberspace.
  - Negative emotions seemed specific while positive sentiment very fuzzy.
  - Traditional models cannot be applied for social media data.

- M³D Communication Model
  - Linguistics analysis should be added to the model.
  - How can the model be quantitatively formed into formula.
  - How to design experiments to test the performance of the model

- Others
  - Research models differ widely due to mission of the funders.

Discussion Points III: Suggestions

- Big Data
  - Determine the value of using big data or looking into small data thoroughly.
  - Refine how data will be collected.
  - Ground the research question before collecting data.

- Building new models and methods
- Mapping movement of individuals spatiotemporally and compare their activities and relationships over various factors (such as socioeconomic status).
- Developing flow-mapping that may help predict disease breakout.
- Examining the roles and influence of different professional groups in the context of a political campaign.
- Developing a discrete-time stepwise regression and consistency index to determine the relationship between real and cyberspace.
- Look into areas about understanding behavior mechanisms.
- Applying GeoMasking to the analysis process.

- Sentiment Analysis
  - Categorize keywords before collecting tweets
  - The meaning of keywords may change so need to
  - Applying weight to words expedites analysis and reveals new questions

- M3D Communication Model
  - Use hashtags in Twitter to collect sample as possible memes to test the model.
  - Build social networks (friend networks) and apply the M3D model to study meme diffusion.

---

**Resources & Further References**

- **Tools:**
  - Web Apps for SNA from Dalhousie U: [http://socialmedialab.ca/?page_id=3213](http://socialmedialab.ca/?page_id=3213)

- **Data centers/Contacts:**
  - SDSU, HDMA Center ([http://humandynamics.sdsu.edu/](http://humandynamics.sdsu.edu/))
  - Social Media Lab, Dalhousie University ()

- **Publications/Sources:**
  - Anatoily – “Social Networks” and a new journal “Big Data and Society”
  - Shaw – special issue on Time Geography
  - Haegerstrand and space-time concepts
  - Conference –

- **Other Projects:**
  - Shaw – project website: [http://web.utk.edu/~sshaw/NSF-Project-Website/default.htm](http://web.utk.edu/~sshaw/NSF-Project-Website/default.htm)
  - Shaowen – project website: [http://www.cigi.illinois.edu/](http://www.cigi.illinois.edu/)
CONCLUSION

All participants and panelists agree that this workshop is very informative and successful. Many specialists suggested that we should continue our discussion and collaboration either through personal contacts or through the project or a future conference to continue the dialogue between different disciplines. We will continue to organize the CDI Specialist meeting in summer 2014 with the support from the NSF-CDI funding.

Our research methods will enable scientists as detectives to investigate spatial patterns and spatial fingerprints among web information landscapes. However, many key technologies and social media may be changed and evolved very quickly. How can we create a well-articulated research framework and sustainable system/tools will be our major challenges in this project.

REPORT PREPARED BY:

Ming-Hsiang Tsou (PI), Jiue-An Yang (Graduate Assistant, Ph.D. student), Christopher Allen (Graduate Assistant, Ph.D. student).
## PARTICIPANTS

### National Visiting Committee (NVC)

<table>
<thead>
<tr>
<th>First Name/ Last Name</th>
<th>Title</th>
<th>Organization</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shih-Lung Shaw</td>
<td>Professor</td>
<td>Department of Geography at the University of Tennessee</td>
<td>[<a href="http://web.utk.edu/~sshaw/Personal">http://web.utk.edu/~sshaw/Personal</a> Homepage/default.htm](<a href="http://web.utk.edu/~sshaw/Personal">http://web.utk.edu/~sshaw/Personal</a> Homepage/default.htm)</td>
</tr>
<tr>
<td>Edna Reid</td>
<td>Associate Professor</td>
<td>James Madison University</td>
<td></td>
</tr>
</tbody>
</table>

### Participants (Specialists)

<p>| Shaowen Wang          | Associate Professor, Director of CyberInfrastructure and Geospatial Information Laboratory | Department of Geography and Geographic Information Science, University of Illinois at Urbana-Champaign; CyberInfrastructure and Geospatial Information Laboratory. | <a href="http://www.cigi.illinois.edu/shaowen/index.htm">http://www.cigi.illinois.edu/shaowen/index.htm</a> |</p>
<table>
<thead>
<tr>
<th>Name</th>
<th>Title/Position</th>
<th>Organization</th>
<th>Website</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marc Armstrong</td>
<td>Professor &amp; CLAS Collegiate Fellow</td>
<td>Department of Geography, The University of Iowa</td>
<td><a href="http://www.uiowa.edu/~geog/faculty/armstrong/index.shtml">http://www.uiowa.edu/~geog/faculty/armstrong/index.shtml</a></td>
</tr>
<tr>
<td>Les Servi</td>
<td>Group Leader</td>
<td>The MITRE Corporation</td>
<td><a href="http://mitre.org/">http://mitre.org/</a></td>
</tr>
<tr>
<td>Feng Chen</td>
<td>Post Doc.</td>
<td>Civil and Environmental Engineering, Carnegie Mellon University</td>
<td><a href="http://www.ce.cmu.edu/people/students/index.html">http://www.ce.cmu.edu/people/students/index.html</a></td>
</tr>
<tr>
<td>Anatoliy Gruzd</td>
<td>Director, Associate Professor</td>
<td>Social Media Lab, Dalhousie University</td>
<td><a href="http://anatoliygruzd.com/">http://anatoliygruzd.com/</a> <a href="http://socialmedialab.ca/">http://socialmedialab.ca/</a></td>
</tr>
<tr>
<td>Sumiko Mekaru</td>
<td>MPVM, DVM, Sc.D candidate at Boston University, Surveillance Epidemiologist</td>
<td>Children's Hospital Boston, HealthMap</td>
<td><a href="http://www.linkedin.com/pub/sumiko-mekaru/33/ba8/471">http://www.linkedin.com/pub/sumiko-mekaru/33/ba8/471</a></td>
</tr>
<tr>
<td>Jaime Arredondo</td>
<td>Graduate Student Researcher-Intelligence Advanced Research Projects Activity (IARPA)</td>
<td>Center for Iberian and Latin American Studies, University of California, San Diego</td>
<td><a href="http://cilas.ucsd.edu/people/staff.html">http://cilas.ucsd.edu/people/staff.html</a></td>
</tr>
<tr>
<td>Pranav Anand</td>
<td>Assistant Professor</td>
<td>Department of Linguistics, University of California- Santa Cruz</td>
<td><a href="http://people.ucsc.edu/~panand/">http://people.ucsc.edu/~panand/</a></td>
</tr>
<tr>
<td>Bruce Appleyard</td>
<td>Assistant Professor</td>
<td>School of Public Affair San Diego State University</td>
<td></td>
</tr>
</tbody>
</table>

**San Diego Participants (Specialists)**
<table>
<thead>
<tr>
<th>Name</th>
<th>Title</th>
<th>Affiliation</th>
<th>Website</th>
</tr>
</thead>
<tbody>
<tr>
<td>Michael K. Peddecord</td>
<td>Professor Emeritus</td>
<td>Graduate School of Public Health, San Diego State University</td>
<td><a href="http://publichealth.sdsu.edu/facultydetail.php?ID=66">http://publichealth.sdsu.edu/facultydetail.php?ID=66</a></td>
</tr>
<tr>
<td>Suzanne Lindsay</td>
<td>Associate Professor, Executive Director</td>
<td>Division of Epidemiology and Biostatistics, Graduate School of Public Health, San Diego State University; Institute for Public Health (IPH)</td>
<td><a href="http://publichealth.sdsu.edu/facultydetail.php?ID=181">http://publichealth.sdsu.edu/facultydetail.php?ID=181</a></td>
</tr>
<tr>
<td>James H. Fowler</td>
<td>Professor</td>
<td>Medical Genetics Division Department of Medicine, Division of Social Sciences Political Science Department; UC San Diego</td>
<td><a href="http://fowler.ucsd.edu">http://fowler.ucsd.edu</a></td>
</tr>
<tr>
<td>John W. Ayners</td>
<td>Research Professor, Graduate School of Public Health</td>
<td>San Diego State University</td>
<td></td>
</tr>
</tbody>
</table>

**SDSU Research Team**

<table>
<thead>
<tr>
<th>Name</th>
<th>Title</th>
<th>Affiliation</th>
<th>Website</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ming-Hsiang Tsou</td>
<td>Professor</td>
<td>Department of Geography, San Diego State University</td>
<td><a href="http://geography.sdsu.edu/People/Pages/tsou/index.html">http://geography.sdsu.edu/People/Pages/tsou/index.html</a></td>
</tr>
<tr>
<td>Dipak Gupta</td>
<td>Professor</td>
<td>Fred J. Jansen Professor of Peace Studies Distinguished Professor in Political Science Program Chair of International Security and Conflict Resolution (ISCOR) San Diego State University</td>
<td><a href="http://www-rohan.sdsu.edu/~dgupta/index.html">http://www-rohan.sdsu.edu/~dgupta/index.html</a></td>
</tr>
<tr>
<td>Brian Spitzberg</td>
<td>Professor</td>
<td>Senate Distinguished Professor School of Communication San Diego State University</td>
<td><a href="http://communication.sdsu.edu/pages/faculty.html#spitzberg">http://communication.sdsu.edu/pages/faculty.html#spitzberg</a></td>
</tr>
</tbody>
</table>
### Student Assistants

<table>
<thead>
<tr>
<th>Student Name</th>
<th>Level</th>
<th>University</th>
<th>Department</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zachary Toll</td>
<td>Undergraduate</td>
<td>UC at Santa Barbara</td>
<td>Geography</td>
</tr>
<tr>
<td>Alejandra Coronado</td>
<td>Undergraduate</td>
<td>San Diego State University</td>
<td>Geography</td>
</tr>
<tr>
<td>Evan Casey</td>
<td>Undergraduate</td>
<td>San Diego State University</td>
<td>Geography</td>
</tr>
<tr>
<td>Elias Issac</td>
<td>Undergraduate</td>
<td>San Diego State University</td>
<td>Geography</td>
</tr>
<tr>
<td>Kathleen Burke</td>
<td>Undergraduate</td>
<td>San Diego State University</td>
<td>Linguistics</td>
</tr>
<tr>
<td>Anoshé Aslam</td>
<td>Masters</td>
<td>San Diego State University</td>
<td>Public Health</td>
</tr>
<tr>
<td>Alex Dodge</td>
<td>Masters</td>
<td>San Diego State University</td>
<td>Linguistics</td>
</tr>
<tr>
<td>Su Yeon Han</td>
<td>PhD</td>
<td>San Diego State University</td>
<td>Geography</td>
</tr>
<tr>
<td>Christopher Allen</td>
<td>PhD</td>
<td>San Diego State University</td>
<td>Geography</td>
</tr>
<tr>
<td>Jiue-An Yang</td>
<td>PhD</td>
<td>San Diego State University</td>
<td>Geography</td>
</tr>
</tbody>
</table>
## AGENDA

San Diego State University, San Diego, California  
*(Fairfield Inn & Suites San Diego Old Town)*  
August 7, 8 (Wednesday, Thursday), 2013

### Wednesday, August 7

<table>
<thead>
<tr>
<th>Time</th>
<th>Activity</th>
<th>Presenter(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>8:00</td>
<td>Free Breakfast at Fairfield Inn &amp; Suites San Diego Old Town (hotel).</td>
<td></td>
</tr>
<tr>
<td>9:00</td>
<td>Welcome and Introductions (PI) (Logistics, each person orders sandwiches for the lunch break).</td>
<td>Ming-Hsiang Tsou</td>
</tr>
</tbody>
</table>
| 9:10   | Background and Introduction to Meeting Goals (Co-PIs) (5 minutes for each Co-PI)  
Participant self introduction (1 minute per person) | Dipak Gupta, Mark Gawron, Brian Spitzberg, Li An. Participants |
| 9:40   | Overview of CDI project (Year 3) Research Task Progress *(1B,2A)*                                 | Ming-Hsiang Tsou              |
| 10:00  | Establish a New Trans-disciplinary Research Framework for Analyzing Big Data and Cyberspace (Project Goal-1) *(Chaired by Ming Tsou)*: |                               |
|        | Automatic Event Detection and Storytelling in Social Media (15 mins presentation + 5 mins Q&A)       | Feng Chen                     |
|        | Infectious Disease Surveillance, Big Data, and the HealthMap Platform (15+5 mins)                    | Sumiko R MeKaru               |
| 10:40  | Break 10 minutes [coffee and cookies]                                                                 |                               |
| 10:50  | Discussant Comments (regarding the previous three presentations)                                    | James Fowler                  |
| 11:00  | Create Effective Visualization and Analysis Methods for Representing Dynamic Changes of Information in Cyberspace (Project Goal-2) *(Chaired by An Li)*: |                               |
|        | Space, Time, and Human Activities in Virtual and Physical Spaces (15+5 mins)                        | Shih-Lung Shaw                |
|        | A CyberGIS Environment for Near-Real-Time Spatial Analysis of Social Media Data (15+5 mins)         | Shaowen Wang                  |
|        | Investigating Political Polarization on Twitter: A Canadian Perspective (15+5 mins)                 | Anatoliy Gruzd                |
|        | CDI Project report in task 2B (space-time analysis) 10 minutes                                      | An Li                        |
| 12:10  | Group Discussion (10 mins) – open to all participants.                                              |                               |
12:20  **Lunch** - Sandwiches from the famous Deli and soft drinks and water from the Hotel.

1:30  **Build domain ontology and (provocative) event knowledge bases** for social media and cyberspace activities (Project Goal-3) *(chaired by Mark Gawron).*

- **Attitude beyond the word level** (15+5 mins)  
  Pranav Anand

- **Geo-location and text wrangling with Twitter data** (15+5 mins)  
  Clay Fink

- **Big data and social media messages: the discourse of social unrest in twitter** (15+5 mins)  
  Jaime Arredondo and Dipak Gupta

- **CDI Project report in task 3A/3B (knowledge bases) 10 minutes**  
  Mark Gawron

2:40  **Group Discussion** (10 mins) – open to all participants.

2:50  **Break 10 minutes [coffee and cookies]**

3:00  **Focus Group Session ONE:** (3:00-3:50)

Room-A (Mission Room – conference room): Linguistic Research Opportunity:
(led by Mark Gawron).

1. **Should the data speak for itself in "Big Data" analyses? How accurate can we be in understanding the mood of the society? How can we improve our understanding?**
2. **Can sentiment analysis in its present state of advancement find meaningful changes in group sentiment and propensity to action over large data sets spread out over time?**

Room-B (Old-Town Room – the other room): Spatial-Temporal Research Opportunity:
(led by Li An).

1. **Given time-series of spatial data (big data), what statistical methods or measures can be used to "dig out" the space-time pattern(s) in the data, and explain/predict such patterns?**
2. **If we have point data collected over Internet, how to create meaningful continuous information landscape(s)?**

Each group will be asked to summarize their discussion (15 minutes) and report them in the next session (report persons should be one of the participants, excluding CDI faculty).

3:50 – 4:30  **Focus Group Report (15 mins + 5 mins Q&A) x 2 groups = 40 minutes**

4:30 – 4:35  **Short Break**

4:35 – 4:50  **Novel Monitoring of Population Health Concerns During the United States’ Great Recession (15 minutes)**  
  John W. Ayers

4:50 – 5:40  **DAY-1 Conclusion and Discussion (Feedbacks from Participants + Your Vision about the future research in Big Data). 2 minutes from each person.**

5:40  **Walk to the Dinner Restaurant (The Cosmopolitan)**
Thursday, August 8

8:00  Free Breakfast at Fairfield Inn & Suites San Diego Old Town (hotel).

9:00  Review of the Day-1 Discussion (3 minutes from each person – 50 minutes) Important! Each person orders sandwiches for the lunch break

9:50  Mining and Mine field: a Revolution in Social Science Research  Dipak Gupta

10:10  Geo-Privacy in Data-Rich Social Media Environments (15+5 mins)  Mark Armstrong

10:30  Break [coffee and snacks]

10:40  Develop theoretical model(s) for the spatiotemporal diffusion of ideas (Project Goal – 4) (Chaired by Brian Spitzberg).

Stickiness in AQAP Inspire magazine or detecting gamification in online communications. (15+5 mins)  Edna Ried

Knowledge Discovery in Cyberspace and Big Data: A look at moods in social media (15+5 mins)  Les Servi

Vaccine Information and Sentiment Over Space and Time. (15+5 mins)  Anna Nagel and Michael Peddecord

Big Social Media, Civic Operations, and Urban Planning Research (15+5 mins)  Bruce Appleyard

CDI Project report in task 4A/4B (communication model) 10 minutes  Brian Spitzberg

Group Discussion (10 minutes) – open to all participants

12:20  Lunch - Sandwiches from the famous Deli and soft drinks and water from the Hotel

1:30  Focus Group Session TWO: (1:30-2:20)

Room-A (Mission Room – conference room): Big Data and Social Media Research Opportunity: (led by Ming Tsou).

1. What are the characteristics of big data and social media messages? Are social media messages a unique type of big data?
2. What type of big data research frameworks can help knowledge discovery in cyberspace?

Room-B (Old-Town Room – the other room): Communication Theory Research Opportunity: (led by Brian Spitzberg).

1. Do (can) big data drive theory, or should (can) theory drive big data?
2. Is meme theory falsifiable (or, are any theories in big data scholarship falsifiable), and if so, how should scholarship proceed?

Each group will be asked to summarize their discussion (15 minutes) and report them in the next session (report persons should be one of the participants, excluding CDI faculty).
2:20 – 3:00: Focus Group Report (15 mins + 5 mins Q&A) x 2 = 40 minutes

3:00 Break [coffee and snacks]

3:20 Research Collaboration Opportunities.
   1. Human Dynamic in the Mobile Age at SDSU (Ming and Brian) (10 mins)
   2. Social Media Research Consortium and Data Center (10 mins)
   3. Academic Research Centers Collaborations (all participants) (10 mins)
   4. CDI project examples for NSF STEM educational objectives (10 mins)
   5. NIH Big Data Center of Excellence CFP (10 mins)

4:10 Plenary Discussion: Next Steps. (3 minutes from each participant)

4:50 Project and Workshop evaluation questionnaires (Spitzberg)

5:30 Car Pool from the Hotel to the Restaurant (China Max)