The Center for Visual and Decision Informatics (CVDI) works in partnership with government, industry and academia to develop the next-generation visual and decision support tools and techniques that enable decision makers to significantly improve the way their organization’s information is interpreted and exploited. Business analysts have typically worked to a requirements-based model, answering clearly defined business questions. Big data, however, demands a different approach, using opportunistic analytics and exploring answers to ill-formed or nonexistent questions. Through a unique partnership with our Industry Advisory Board members, our researchers are identifying and addressing state-of-the-art challenges in Big Data, including event emergence in networked data, mining and visualization of patterns in data streams and cross-enterprise data integration.

CVDI is excited to present its accomplishments for 2013-2014, its second year of operation. With the support of our IAB members, CVDI researchers completed five projects in Year 2. These projects covered topics prioritized by our industry members (see pages 6-10). This report also highlights the work with undergraduate students and the NSF fundamental research project awarded to CVDI (see pages 11-12). CVDI has also attracted several NSF research awards to complement the research funded by our industry affiliates, namely:

- The Early Concept Grants for Exploratory Research (EAGER) - A Collaborative Exploration in Networked VR Environments and Application to Remotely-Guided Classroom (UL Lafayette);
- Major Research Instrumentation (MRI) - Development: A Distributed Visual Analytics Sandbox for High Volume Data Streams (UL Lafayette);
- Collaborative Opportunity for Research Between I/UCRCs (CORBI) - Modeling, Visualization, and Understanding of Large Data Sets (Drexel);
- Fundamental Research: Visualization-based Gap Analysis and Link Prediction (UL Lafayette and Drexel);
- CVDI I/UCRC Phase I Supplement (Drexel).

We have new domestic and international partnerships. These will provide CVDI IAB members access to a wider and deeper array of research projects, greater ability to tap into technical expertise of new faculty, access to potentially new funding sources, and an opportunity to collaborate with international industry members and students. The University of North Carolina at Chapel Hill (UNC) is actively collaborating with CVDI to join as the third domestic academic research site. This partnership is moving forward and UNC received an I/UCRC planning grant in April 2014. The grant enables UNC to hold a workshop (in 2015) involving potential industry partners, CVDI, UNC and NSF to develop a research projects and an operations plan for the site.

In addition, we are closely working together with DIGILE, one of Finland’s Strategic Centers for Science, Technology and Innovation, to establish an international research site. DIGILE is devoted to increasing the pace of development of Finnish Internet Economy Competencies to enable global business opportunities. DIGILE is excited to collaborate with CVDI and has spearheaded the creation of a research site at Tampere University of Technology (TUT), Finland. Our leadership attended the official launch event in Finland in September, and CVDI is excited to welcome DIGILE and TUT guests at the fall meeting in October. While TUT will be the lead academic site in Finland, several other prominent Finnish universities will contribute further research capacity to the Finnish research site. The DIGILE team is currently working with their potential list of IAB members and possible research topics that align with the research themes of CVDI.

We look forward to the continued growth of the organization and are excited to join the elite club of programs with international partnerships. If you have any questions, please contact CVDI for more information.
CVDI delivers innovation, excellence and leadership in using data-driven paradigms for visual and decision informatics. The Center’s mission is to research and develop next generation visual and decision support tools and techniques to enable decision makers in government and industry to fundamentally improve the way their organization’s information is interpreted and analyzed.

By uniting analytic, visual and predictive techniques, CVDI is advancing state-of-the-art research in the fields of Information Visualization, Visual Analytics, Data Mining and Automated Data Analysis. This research will be supported by advanced computing and visualization facilities to create Decision-Making Environments (DME) — a framework that will enable users to explore and customize information streams in a variety of modalities to gain better insight to that information.

Research Areas
In consultation with our IAB members and strategic discussions with thought leaders in the field, CVDI will focus on three primary research areas (1) Visual Analytics, (2) Predictive Analytics and (3) Deep Pattern Analytics.

Visual Analytics
Visualization is an essential mechanism for delivering the meanings of data to users. It allows users to notice and perceive what is not clear from numbers alone. The growing information overload limits the ability of what a human can see and perceive with just data visualization. Visual analytics empower users with analytical reasoning for visual discovery with interactive interfaces. In order to address the emerging information overload challenges posed by Big Data, it is more important than ever to develop a novel class of visualization metaphors, algorithms, methodologies and solutions. Visual Analytics research in CVDI targets to improve the ability of users to interpret more information than what is possible today with existing commercial visualization tools. Some of the research that will be pursued includes: real-time visualization of extracted core data, methods to support interactive visualization of large and often complex data sets, real-time data processing and
analysis in memory, visualization of mashed up data, effective visualization over mobile devices, n-dimensional visualization, interactive exploration of data with dimension reduction and abstraction, visual gap analytics for data exploration, visual discovery, high-resolution displays and visualization adaptable to user preferences.

Predictive Analytics
Predictive analytics encompasses a variety of advanced techniques to make predictions about future events by discovering patterns and trends from current and historical facts. Techniques used include statistics, modeling, machine learning, text mining, social media analytics and data mining. By using predictive analytics, organizations can move beyond simple reporting or displaying of historical data on static dashboards, to improve their knowledge about their business, competitors, services and customers. Ultimately, this can help to reduce risks and aid in better business decisions. CVDI can contribute significantly in this area by integrating advanced techniques with real-time challenges faced by our industry members to develop better techniques and models to make more accurate predictions.

Deep Pattern Analytics
Once the initial analytic work is concluded, researchers can further identify the deep underlying patterns in the data. This type of work is being initiated by the Center, and it is a natural progression based on the previous two research areas. Additionally, as very few organizations are exploring Deep Pattern Analytics, all of CVDI’s future contributions to the research will increase its recognition as a leader in this field. Examples of research in this area include highly efficient and scalable approaches based on statistics, semantics and machine learning algorithms; deep analytics automation techniques and frameworks; advanced HCI that facilitates multidisciplinary approaches; temporal, graph, sense-making and context-based approaches; and data-based model refinement approaches.

Through an internal survey process, IAB members identified several research areas that were most important to them for obtaining value from Big Data.
RETURN ON INVESTMENT

Industry Advisory Board (IAB) members contributed $860,000 towards the first two years of CVDI research projects. IAB member contributions directly support research, workforce development, and overhead for these efforts. Based on a $30,000 membership contribution for each of the two years, this represents a 13:1 return on investment. Considering revenue from all sources, IAB members benefited from a 31:1 return on investment.

Revenue

<table>
<thead>
<tr>
<th>Revenue Source</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>NSF Center Award</td>
<td>$430,000.00</td>
</tr>
<tr>
<td>NSF Supplements</td>
<td>$63,585.00</td>
</tr>
<tr>
<td>IAB Memberships</td>
<td>$860,000.00</td>
</tr>
<tr>
<td>Enhancement Funding</td>
<td>$254,773.00</td>
</tr>
<tr>
<td>University Contributions</td>
<td>$282,080.00</td>
</tr>
<tr>
<td><strong>Total Income</strong></td>
<td><strong>$1,890,438.00</strong></td>
</tr>
</tbody>
</table>

Expenditures

<table>
<thead>
<tr>
<th>Expenditure Type</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student Support</td>
<td>$692,951.00</td>
</tr>
<tr>
<td>Other Project Expenses</td>
<td>$127,810.00</td>
</tr>
<tr>
<td>Center Operating Expenses</td>
<td>$414,865.00</td>
</tr>
<tr>
<td>Indirect Costs</td>
<td>$182,253.00</td>
</tr>
<tr>
<td><strong>Total Expenses</strong></td>
<td><strong>$1,417,879.00</strong></td>
</tr>
<tr>
<td><strong>Remaining Balance</strong></td>
<td><strong>$472,559.00</strong></td>
</tr>
</tbody>
</table>

IAB Projects Direct Operating Expenses

<table>
<thead>
<tr>
<th>Project Details</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS 12-1 Benton/C. Chen</td>
<td>$62,628.00</td>
</tr>
<tr>
<td>CS 12-2 Yang</td>
<td>$53,412.00</td>
</tr>
<tr>
<td>CS 12-3 An/Lin/Totaro</td>
<td>$119,416.00</td>
</tr>
<tr>
<td>CS 12-4 Cruz-Neira</td>
<td>$18,076.00</td>
</tr>
<tr>
<td>CS 12-5 Gottumukkala/Borst</td>
<td>$90,506.00</td>
</tr>
<tr>
<td>CS 13-1 An</td>
<td>$93,200.00</td>
</tr>
<tr>
<td>CS 13-2 Hu/Benton</td>
<td>$91,593.00</td>
</tr>
<tr>
<td>CS 13-3 Gottumukkala/Borst</td>
<td>$55,189.00</td>
</tr>
<tr>
<td>CS 13-5 J. Chen</td>
<td>$63,929.00</td>
</tr>
<tr>
<td>CS 13-6 C. Chen</td>
<td>$69,653.00</td>
</tr>
<tr>
<td><strong>Total IAB Projects Direct Operating Expenses</strong></td>
<td><strong>$717,602.00</strong></td>
</tr>
</tbody>
</table>

NOTES TO THE FINANCIAL STATEMENT:

CVDI revenue is recognized on an accrual basis and expenditures are recorded on a cash basis. NSF supplemental revenue includes all supplemental funds awarded to CVDI Phase 1. Enhancement funding includes all other revenue awarded to CVDI such as U.S. federal grants. Remaining balance indicates current funds available for ongoing projects, IAB semi-annual meetings, and Center administration. Direct project expenses do not include any student support charged to non-IAB funded projects. Indirect costs are calculated at the rate of 10% for CVDI projects, 43% modified total direct costs (MTDC) for UL Lafayette federal grants and 54.5% MTDC for Drexel federal grants.
Semantic Information Extraction, Integration, and Visualization for Big Data Analytics

Personnel

**Principal Investigator:** Yuan An (Drexel)

**Graduate Students:** Mengwen Liu (Drexel), Yuan Ling (Drexel), Yizhou Zang (Drexel)

Summary/Abstract

Free text documents such as scientific literature contain abundant knowledge about relationships among concepts or entities. Unfortunately, this type of knowledge is expressed in natural language, where different types of relationships are not explicitly categorized.

In this project, the team developed techniques for extracting structured knowledge from unstructured data through weak supervision over existing sources of knowledge. In particular, a novel relation extraction approach was developed that integrates distant supervision earning with open information extraction techniques, which, unlike state-of-the-art models, does not need manually-labeled examples of relations.

In addition, the model incorporates a grouping strategy to take into consideration the interdependency among entities occurring in one sentence, which has been largely ignored in previous studies.

To verify the method, the team used Elsevier’s document corpus containing about 1 million articles in the neuroscience domain as the test bed for the project. The goal was to extract gene expression relationships between genes and brain regions. The experiments, which used manually annotated “gold standards,” demonstrated that these methods can achieve better performance than baselines.

The work shows that it is possible to develop general distant supervision approaches for relation extraction from free text. With the success in the biomedical domain, the team looks forward to expanding the techniques to a broader range of applications.
Large-scale Social Media Analytical Tools with Application to Detecting Emerging Events

Personnel
Principal Investigators: Xiaohua Tony Hu (Drexel), Ryan G. Benton (UL Lafayette)
Co-Investigator: Weimao Ke (Drexel)
Graduate Students: Satya S. Katragadda (UL Lafayette), Yue Shang (Drexel), Xiaoli Song (Drexel), Dhanalakshmi Veeramachaneni (UL Lafayette)

Summary/Abstract
With the increasing usage of social media, there is an ever-growing need to extract useful, actionable information from the posts, blogs, and reviews provided by the public. The objective of this project is to create and enhance techniques for detecting and analyzing events and topics, ranging from conducting real-time analysis to conducting longer, more in depth analysis on detected topics. In particular, the project focused on four major developments.

First, a sentiment analysis method was developed that can not only find out how people feel about a subject, but determine which parts (aspects) of a policy or consumer object people like and dislike. As a bonus, the new methods outperform the current state-of-the-art sentiment detection methods.

Second, the project extended the real-time, Event Detection on Onset (EDO) method, developed last year to be able to incorporate feeds from any number of social media sites; this improves the ability to detect new emerging events.

Third, two improved topic evolution models were developed, which better captures topics and their evolution over time. One version is more appropriate for real-time/near real-time event analysis; the other, more powerful version is better suited for detecting topic changes on an hourly and/or daily basis.

Finally, the project developed a high-speed, distributed clustering algorithm, called Pruned AP method, which significantly reduces time for data processing and network communication while improving the qualitative clustering results for event/topic detection.
Personnel

Principal Investigator: Raju Gottumukkala (UL Lafayette)
Co-Investigator: Christoph Borst (UL Lafayette)
Graduate Students: Siva Rama Krishna Venna (UL Lafayette), Nicholas Lipari (UL Lafayette)

Summary/Abstract

One unavoidable fact in the pursuit of creating techniques that can easily handle big data is that many problems can naturally be represented via large, heterogeneous, time-varying graphs. While there is a plethora of batch-oriented graph mining algorithms, there are fewer that naturally handle evolving graphs informed by streaming data; there is also few if any methods that allow the user to easily interact with the large scale graphs, both for exploration and for analysis.

This project continues CVDI’s quest to create a Real-time Graph-based Visual Analytic Environment and builds upon the results/techniques developed in last year’s Real-Time Analysis and Visualization of Multi-Dimensional Sensor Data project.

In this quest, three objectives were pursued. The first was the creation of a methodology to construct a dependency graph using standard association analysis techniques to understand relationships between various entities. The methodology has been validated using an IAB member’s sales transaction data, where the goals were to understand product purchase patterns and identify at-risk customers.

Second, a prediction model was created to predict event trends from evolutionary (or temporal) graphs, where individual nodes have non-stationary correlations. Testing was accomplished by using Google Flu Trends data, as well as various environmental conditions (temperature, sun exposure) and influenza history. The tests demonstrated the prediction model outperformed existing time series models found in the literature.

Third, a number of visual analytic interfaces were developed for a variety of emerging interaction and display devices. The interfaces included a standard web-based visualization to show the results of the predictive analytics and the development of novel multi-touch graph interaction interfaces/displays for manipulation of graphs. Based on the results obtained, the IAB has approved the continued development of the Real-time Graph-based Visual Analytic Environment.
A Spatio-Temporal Data Mining Approach for Fraud Detection

Personnel

Principal Investigator: Jian Chen (UL Lafayette)
Co-Investigators: Ryan Benton (UL Lafayette), Raju Gottumukkala (UL Lafayette)
Graduate Students: Shaaban Abbady (UL Lafayette), Maria Bala Duggimpudi (UL Lafayette)

Summary/Abstract

Detecting fraud, such as credit card, healthcare, and tax (to name a few), is an important function in many organizations. Unfortunately, automating the process can be difficult, as this typically requires collection of a diverse amount of information, and the encoding on the knowledge of experts either to (a) define normal activity, (b) define fraudulent activity or (c) provide use cases for learning systems.

The objective of this project is to develop a scalable, autonomous spatio-temporal data mining framework for fraud detection that avoids the need to encode expert knowledge. To accomplish this, the project (a) developed new, unsupervised algorithms to detect anomalies (outliers) based on spatio-temporal context, (b) adapted sophisticated partitioning methods for parallelization to achieve scalability and (c) applied pruning strategy to improve efficiency. A simple visualization system to display the results has also been developed.

The proposed unsupervised approach was able to detect anomalies according to spatio-temporal context and locality considerations with excellent recall and satisfactory precision. Datasets included synthetic data and Medicare data. Experiments on both synthetic and real data sets demonstrated the scalability of the framework to big data sets. Moreover, the applicability of the system to detect anomalies for non-fraud problems was also demonstrated using buoy/hurricane data.
The project aims to develop a set of tools for gap analytics so that organizations can better track their performance with reference to peer organizations to make better-informed decisions in strategic planning and assessments of their operations. To support this, the team developed a computational and visual analytic architecture to support gap analytics through the notion of three-dimensional fitness landscapes. Fitness landscapes provide intuitive representations of potentially complex and dynamic situations concerning organizations. These are designed in such a way that navigation over the landscapes is well defined with reference to possible courses of action and qualitative assessments of corresponding outcomes of a course of action.

This project has successfully laid down the foundation and has proven the fundamental idea that generating fitness landscapes from non-geospatial data is feasible and meaningful.

To validate the approach, the group conducted a gap analysis of U.S. patents as one of the representative application domains. This involved constructing a local database containing over 5 million patents and over 30 million patent co-citations. The team conducted experiments to identify hot areas of patenting activities, with burst detection of patent citations at three levels: the individual patent level, the international patent classification (IPC) level, and the U.S. patent classification level.

The work can be further improved within the established framework. In particular, the team envisions additional follow-up work of the project to reach a more practical stage to support gap analytics. Moreover, the group has begun to explore procedural modeling at the level of individual users of an analytic system such that novice users can learn from expert users.
Visualization-Based Gap Analysis and Link Prediction

Both implicitly and explicitly, graph analysis and visualization techniques have been utilized in multiple CVDI projects. Given this importance, CVDI initiated a two year project to conduct fundamental research within the graph analysis/visualization space. The two focus areas of this project, which began in August 2013, are Visualization-Based Gap Analysis and Link Prediction. The Visualization-Based Gap Analysis effort aims at providing an intuitive visualization and analysis techniques to provide analysts with the ability to understand what has happened within a domain, comprehend its current status and operations, and explore the impact of changes to the system.

The Link Prediction effort seeks to automatically predict what relations arise in the future between objects. In particular, the proposed research intends to reformulate the link discovery problem to include predicting the strength of the link as well as create techniques to detect and predict special categories of future links.

By establishing more rigorous capabilities for the Visualization-Based Gap Analysis and Link Prediction, it is expected that industry will be better able to understand what has happened, comprehend what is the current state and obtain some grasp of what is likely to happen. This, in turn, should lead to improved ability to resolve current problems and exploit future opportunities.

This project is funded, in part, by the National Science Foundation program Fundamental Research Program for Industry/University Cooperative Research Centers.
One goal of CVDI is to create a future workforce by educating a diverse body of students on the interdisciplinary field of visual and decision informatics. The following research projects were supported, in part, by two NSF programs: Research Experience for Undergraduates (REU) and Research Experiences for Veterans (REV).

### Exploration in Data Mining/Data Integration

**Personnel**

**Undergraduates:** Sheik Hassan (Drexel), Pattamapon Thongsima (Drexel)

**Veteran:** Kristian Linares (Drexel)

**Advisors:** Yuan An (Drexel), Xiaohua Tony Hu (Drexel), Weimao Ke (Drexel)

**Summary/Abstract**

One goal of CVDI is to create a future workforce by educating a diverse body of students on the interdisciplinary field of visual and decision informatics. The purpose of this project is to provide research opportunities for three students (two undergraduate and one veteran) who are interested in data mining and data integration. The students engaged in research, under the supervision of three advisors, designed to improve their understanding of the impacts of data mining and data integration upon visualization and decision making.

### Enhancing Social Media Event Detection

**Personnel**

**Undergraduates:** Justin Garrett (UL Lafayette), Nakul Regmi (UL Lafayette)

**Advisors:** Ryan G. Benton (UL Lafayette), Christoph Borst (UL Lafayette), Satya Katraggada (UL Lafayette, graduate student)

**Summary/Abstract**

CVDI engaged two undergraduate students in research efforts aimed at providing a better understanding of how to analyze and visualize events detected from social media data. Garrett implemented a system that would (a) acquire event data from Twitter and determine its importance and (b) visualize the data in a meaningful and aesthetically pleasing way. Regmi sought to enhance a previously developed CVDI system, Event Detection on Onset, by incorporating term hierarchies and introducing event types.
The National Science Foundation (NSF)’s Industry/University Cooperative Research Center (I/UCRC) program is a sound and well-tested model for companies to acquire external technologies through open innovation by partnering with universities. An I/UCRC is supported through membership fees from the Industry, grant money from NSF and commitments of participating universities.

The program currently supports 67 centers in 42 states and four countries (Belgium, Germany, Russia and India). In 2013, I/UCRC’s were responsible for more than 1,400 publications, and have spun off half a dozen start-up companies.

More than 900 faculty members, along with some 1,500 graduate students and 300 undergraduate students, carry out the research at these Centers annually, which encompass almost the entire spectrum of current technological fields.

To industry, government and other organizations with research needs, the NSF I/UCRC program provides the means to leverage research and development (R&D) investments with multi-university centers renowned for innovative research capabilities.

The Center for Visual and Decision Informatics (CVDI) is a multi-university – industry research center established as a NSF I/UCRC in February 2012. CVDI is part of the NSF’s Directorate for Computer and Information Science and Engineering.

CVDI’s value proposition is to offer a low-cost, low-risk venue for its members from industry and government agencies to validate their early stage innovations with the involvement of university faculty, post-doctoral scholars, researchers and students.

Members of the CVDI Industry Advisory Board (IAB) have access to leading-edge developments in visual and decision informatics. IAB members guide the direction of the Center’s research projects. Members will also have access to the interdisciplinary knowledge base from academic researchers and engage with other industry advisory board members to explore potential opportunities. The following are highlights of membership benefits:

- Opportunity to network and collaborate with other Industry Advisory Board members to explore joint ventures;
- One-stop access to interdisciplinary university professors, researchers and graduate students from the partnering universities;
- The ability to investigate research topics that may otherwise be too expensive to investigate with your existing staff;
- Convenient access to world-class facilities, including cloud computing, supercomputing and immersive visualizations.

**Small Business Membership**

- $20,000 annual membership for small businesses as defined by the U.S. Small Business Administration; Small business members benefit from a greater ROI and can qualify for additional funding and subsidized fees through Small Business Innovation Research and Small Business Technology Transfer (SBIR/STTR) grants.

**Members Supporting Year 2 Projects**

- The Children’s Hospital of Philadelphia
- Elsevier
- IMS Health Incorporated
- Johnson & Johnson
- Lockheed Martin
- Louisiana Department of Health and Hospitals (2)
- Louisiana Department of Revenue
- Louisiana Health Care Quality Forum
- Louisiana Immersive Technologies Enterprise
- Microsoft Research
- Stuller, Inc.
- SunGard Availability Services LP
- The SI Organization, Inc.
- Thomson Reuters

CVDI is grateful to The Institute of Museum and Library Services (IMLS) and the Water Institute of the Gulf for their Center support.
September 26, 2014

Dear Members,

I served as an Industry-Advisory Board for the Center for Visual and Decision Informatics (CVDI) for the past two years. This past August, I transitioned to be the Chairman of the Board. I consider myself a veteran, representing the Louisiana Health Care Quality Forum since the CVDI’s inception.

This past year has marked a year of tremendous growth for CVDI. The organization has made solid progress on many fronts, including closely aligning research that meets the industry needs, collaborating amongst researchers, expanding our research resources through academic and international partnerships, and adjusting the annual meeting format to better serve our member’s interests.

We continue to build CVDI by adding new members and broadening our research pool as well as resources. The organization strives to foster good working relationships as industry and research come together to solve big data challenges. The leadership is strong and works very hard to facilitate the needs and administration of the organization.

It is an honor to serve in this role, please contact me if you have any questions.

Sincerely,

Brian C. Richmond
Chief Technology Officer
Louisiana Health Care Quality Forum
Office: (225) 334-9299
brichmond@lhcqf.org
THANK YOU
The Center for Visual and Decision Informatics (CVDI) would like to thank our Industry Advisory Board (IAB) members, member companies, the Institute for Museum and Library Services (IMLS), the Water Institute of the Gulf and the National Science Foundation (NSF) for their generous support and insight.

Our work benefits greatly from academic, corporate and government collaborators who provide leadership and thoughtful guidance as the organization engages these critical, emergent issues. Many thanks as well to the University of Louisiana at Lafayette and Drexel University for providing operational support and a dynamic body of students, professional staff, and faculty who make this research possible.

University of Louisiana at Lafayette Site
Abdalla Hall
635 Cajundome Blvd.
Lafayette, LA 70506
Phone: 337-482-0600

Drexel University Site
3141 Chestnut Street
Philadelphia, PA 19104
Phone: 215-895-5884

Visit www.nsfcvdi.org for more information.