Project Summary

Disasters are events of an "accidental or uncontrollable" nature (Drabek, 1968) that substantially disrupt social processes (Fritz, 1961). Such events trigger efforts by organizations to minimize losses and restore normal conditions in the affected region. This response process typically involves joint interventions by multiple organizations focused on different problem domains (e.g., search and rescue, infrastructure deployment). In order to avoid fragmentation, gaps in service delivery, and unnecessary duplication of service, interorganizational coordination of activities is therefore essential (Gillespie, 1991).

A recent review of research on interorganizational coordination in disasters by Drabek and McEntire (2002) found that multiorganizational coordination is a key factor in the success of disaster response operations. If vital activities such as search and rescue, medical care, traffic control, and resource allocation are carried out in an unstructured manner, serious consequences such as insufficient resource allocation, omission of tasks, and even counterproductive activities can result (Drabek, 1968, Auf der Heide, 1989). In practice, coordination is achieved largely through interorganizational interaction *networks*, which serve as conduits for directives, information, and resources (Drabek, 1985). The unfolding of these networks over time is thus a critical element of the response process, and the networks themselves provide insight into the nature of interorganizational coordination during disasters. Moreover, from a public policy standpoint, since interorganizational activities are generally regulated, interorganizational coordination provides a particularly attractive target for intervention.

This project seeks to advance our understanding of interorganizational coordination in disaster response. More specifically, it will help us realize the potential for real-time intervention in emergent multiorganizational networks (EMONs) such as those formed during the response to Hurricane Katrina. The proposed collaborative project leverages the theoretical, methodological and computational expertise being developed in two ongoing NSF-ITR funded projects on emergency response: RESCUE (*Responding to Crises and Unexpected Events*) and CP2R (*Collaboration Framework for Preparing Against, Responding to, and Recovering from Disasters Involving Critical Physical Infrastructures*). Using novel computational and statistical methods, the research collaboration between PIs on the two ongoing ITR projects will capture, validate, and integrate data from news reports, official documents, and information sources such as weblogs to produce estimates of interorganizational interaction over time. This approach extends in scale, richness, and accuracy the pioneering methodology for mapping EMONs developed by Tierney (a consultant on this proposed project) and her colleagues in the wake of 9/11. Automated and augmented network extraction, using NCSA's D2K technology, will be evaluated and validated based on human coding of source documents and secondary information. We will use the EMONs data to empirically validate, at a scale not anticipated, the theoretical models developed in the two NSF-ITR projects.

Intellectual Merit: The proposed project is a pioneering effort at collecting large scale interorganizational networks. Our efforts are based on a suite of web-crawling, data-mining and text-extraction tools, coupled with sophisticated statistical methods for inferring network links based on error-prone sources, and assessing their structures in comparison to theoretically hypothesized structural tendencies. *First*, the proposed project will provide a systematic assessment of the relative viability and validity of manual, automated, and computer-augmented techniques for coding networks based on a large corpus of digital data. *Second*, the proposed project will provide an opportunity to scale up our ongoing methodological investigations using Bayesian statistical methods to infer network ties from error-prone data and exponential random graph models to test theoretically derived hypotheses about network structures. *Third*, the improvements in the quality of data and the statistical techniques applied to them will provide an unprecedented opportunity to uncover theoretical explanations for coordination among emergent multi-organizational networks in disaster response. *Fourth*, the data set produced by this research will be a key resource for social scientists, disaster researchers, information technologists, and policy analysts studying problems related to the Katrina response.

Broader Impact: The findings, tools, and methodologies derived from the proposed research will be immediately generalizable to a wide range of disaster response situations. Our ongoing partnerships with government and non-government agencies in the US and internationally as part of our current NSF ITR projects make us exceptionally well-equipped to rapidly incorporate our findings, tools, and methodologies into our ongoing education, training, and outreach workshops with response organizations. Our research will produce an account of interorganizational coordination in the Katrina response which is unprecedented in scope and detail, and which will itself serve as a tool for emergency managers and others seeking to understand what transpired during the operation. The intertemporal data set produced by this research, as well as the new tools for the automated extraction, measurement, and analysis of emergent multiorganizational networks will be made publicly available via NCSA's SONIC (Science of Networks in Communities) portal for use by disaster researchers and other social scientists, policy analysts, and emergency managers.

Project Description
Collaborative Research: Mapping and Analysis of Emergent Multiorganizational Networks in the Hurricane Katrina Response

1. Background and Overview

Disasters are events of an "accidental or uncontrollable" nature (Drabek, 1968) that substantially disrupt social processes (Fritz, 1961). Such events trigger efforts by organizations to minimize losses and restore normal conditions in the affected region. This response process typically involves joint interventions by multiple organizations focused on different problem domains (e.g., search and rescue, infrastructure deployment). In order to avoid fragmentation, gaps in service delivery, and unnecessary duplication of service, interorganizational coordination of activities is essential (Gillespie, 1991).

Interorganizational coordination has been widely studied, both in the context of disaster per se (e.g., Drabek, 1968; Auf der Heide, 1989; Roberts, 1994, Topper and Carley, 1999) and in more general settings (e.g. Williamson, 1975; Granovetter, 1985; Powell, 1990; Baker and Faulkner, 1993; Powell et al., 1996). A recent review of research in this area by Drabek and McEntire (2002) supports the position that multiorganizational coordination is a key factor in the success of disaster response operations. Conversely, if vital response activities such as search and rescue, medical care, traffic control, and resource allocation are carried out in an unstructured manner, insufficient resource allocation, omission of tasks and counterproductive activities can result. (Drabek, 1968, Auf der Heide, 1989). In practice, coordination is achieved largely through the emergence of interorganizational interaction networks. These networks serve as conduits for directives, information, and resources (Drabek, 1985). Over time, the unfolding of these networks is a critical element of the response process; therefore, the networks themselves provide insight into the nature of interorganizational coordination during disasters. Because interorganizational activities are generally regulated, interorganizational coordination provides a particularly attractive target for intervention from a public policy standpoint.

This project seeks to advance our understanding of interorganizational coordination in disaster response. More specifically, our efforts will help to realize the potential for real-time intervention in emergent multiorganizational networks (EMONs) such as those formed during the response to Hurricane Katrina and its successors in the 2005 Atlantic hurricane season (e.g., Rita and Ophelia). The proposed collaborative research project brings together and leverages the theoretical, methodological, and computational expertise being developed in two major ongoing NSF-ITR funded projects on emergency response: RESCUE (*Responding to Crises and Unexpected Events*) and CP2R (*Collaboration Framework for Preparing Against, Responding to, and Recovering from Disasters Involving Critical Physical Infrastructures*). Using novel computational and statistical methods, the research collaboration between the three PIs (in communication, sociology, and civil engineering) on the two ongoing ITR projects at University of California at Irvine (UCI) and University of Illinois at Urbana-Champaign (UIUC), will capture, validate, and integrate data from news reports, official documents, and information sources such as weblogs, to produce estimates of interorganizational interaction over time. This approach extends in scale, richness, and accuracy the pioneering methodology for mapping EMONs developed by Tierney (a consultant on this proposed project) and her colleagues in the wake of 9/11. Automated and augmented network extraction, using NCSA's D2K technology, will be evaluated and validated based on human coding of source documents.

The EMON data to be collected in the aftermath of Hurricane Katrina and its successors provides an opportunity to theoretically extend, empirically validate, and computationally advance the research proposed in the two NSF-ITR projects. The size of these networks, their dynamic nature, and the necessity to generate real time (or near real time) analysis for intervention will serve as an ideal test bed to leverage high-end distributed computing resources within the prototype Social Network Analysis Cyberinfrastructure (SNAC) being developed by the Science of Network in Communities (SONIC) research group at the University of Illinois, Urbana Champaign's National Center for Supercomputing Applications (NCSA). This remarkable data set, as well as the new tools for the automated extraction, measurement and analysis of emergent multiorganizational networks will be made publicly available via NCSA's SONIC (Science of Networks in Communities) portal for use by disaster researchers and other social scientists, policy analysts, and emergency managers.

2. Detailed Research Plan

2.1. Initial Data Collection

Data collection will focus on publicly available, textual sources. While these sources provide a rich resource for reconstructing the EMONs formed during the Katrina response, they are potentially ephemeral. The first priority of

this project will be to contemporaneously archive documents before they are altered or destroyed. Information sources to be utilized for this project fall into two broad categories: official documentation produced by response organizations themselves, and third-party accounts of interorganizational activities. We here consider each in turn.

2.1.1 Documentation from Official Sources

Response activities during disaster are coordinated in large part via the use of situation reports, memos, advisories, and other documents which are produced by and for members of responding organizations. The UCI team will establish a corpus of such official documents regarding the Katrina response, utilizing publicly available sources (e.g., EOC releases, departmental web sites). In preparation for subsequent coding, documents will be tagged by attributes such as organizational source, publication date, and document type. This effort will extend ongoing data collection activities by the UCI team, which has amassed over one thousand documents from approximately 60 public and private entities at the state, local, and federal levels. For a partial list of sources, see:
https://webfiles.uci.edu/buttsc/distribution/RESCUE/sources.list.html?ticket=t_RXEuuOAy

2.1.2 Third-party Sources

In addition to primary sources, information regarding the interorganizational interaction to the Katrina response will be gleaned from web sites, weblogs, and Wikis created by third-party sources such as media outlets, interest groups, and individuals. The UIUC team will apply automated web-crawling tools to identify and construct a corpus which will include such third-party documents regarding the Katrina response. The approach taken will utilize components in the D2K (Data to Knowledge application environment) developed at the National Center for Supercomputing Applications (NCSA) at UIUC for traversing the web automatically, in parallel and in cooperation with many instances running on different machines. The components of this crawler include the following parts: data collection, collection guidance, data normalization/analysis, data indexing, and data retrieval/search. By applying these technologies to a range of online data sources, we will accumulate a wide array of secondary information on organizational activities. As with the official sources, documents acquired through this system will be tagged by source, date, and type to facilitate coding and extraction.

2.2 Network Data Extraction

Once the corpus of source documents has been assembled, the next phase of this project consists of the extraction of EMON information from each document in the corpus. This process will be conducted via a combination of manual, automated, and computer-augmented coding procedures; automated coding will dramatically enlarge the size of the corpus which can be considered, with the smaller manually coded sample utilized as both an initial analysis target and training data for the automated system. Finally, computer-augmented coding procedures will leverage the learning algorithms in D2K to improve the quality of automated coding.

2.2.1 Identification of Named Entities

In order to reconstruct the Hurricane Katrina EMONs, it is first necessary to identify the various organizations and other entities involved in the response. This is a well-known problem in computer science (named entity extraction), for which fairly good algorithms are available. Here, a combination of manual (UCI) and automated (UIUC) procedures will be employed to identify the full set of named entities for the Katrina response. Manual tagging of named entities from a random document sample will be used to calibrate the behavior of the automated tagging routine; input from domain experts in both teams will also be employed to verify the credibility of the entity set.

2.2.2 Human Coding of Relational Information

The "gold standard" for extraction of relational information from text continues to be tagging by human coders. To this end, the UCI team will manually code interorganizational information from the corpus of official source documents regarding the Katrina response. The network data set produced by this activity will serve two purposes: it will be analyzed directly to gain insight into the Katrina response, and it will be used to validate the automated network extraction techniques discussed below. Once validated, these tools can be applied to the full corpus of third-party response data (a corpus which will be too large to for manual coding to be feasible).

The coding procedure to be followed here draws upon prior work such as that of Tierney and Trainor (2004) regarding the measurement of interorganizational networks in disasters. An initial pass will be made by a team of undergraduate research assistants, who will identify all cases of apparent organizational interaction within each document. (Multiple coders will be used for each document, so as to allow inter-rater reliability estimation.) All identified interactions will then be independently examined by a pair of graduate student researchers, who will classify the interaction by type (e.g., activity) and/or reject it as ineligible. Disagreements by the second-pass coding

team will be resolved by the UCI lead researcher (Butts), with aid from consultant Kathleen Tierney (a nationally recognized expert in organizational response to disasters). The resulting data set will consist of the validated interorganizational ties described in each source document, classified by type. Reliability estimates from first and second pass coding will also be retained for use in the cross-validation activity (see below).

2.2.3 Automated Network Extraction

Information extraction techniques analyze unstructured text in order to extract information about pre-specified types of events, entities or relationships. Information extraction (IE) gathers facts out of documents to support subsequent analysis. Information extraction is different from traditional information retrieval techniques in that it is not a search and retrieval process; that is, IE recovers facts from documents regarding pre-specified events, entities, or relationships. As a staring point for this research, we plan to extract Agents, Locations, and Organizations. These entities can then be entered into a database for examination of patterns for social network analysis. ALG has developed the D2K - Data to Knowledge application environment for data mining (Cai et al., 2004; Kooper et al., 2005). D2K is a rapid, flexible data mining and machine learning system that integrates analytical data mining methods for prediction, discovery, and deviation detection, with data and information visualization tools. It offers a visual programming environment that allows users to connect programming modules together to build data mining applications, and it supplies a core set of modules, application templates, and a standard API for software component development. T2K-Text to Knowledge is a set of components that provides text mining and analysis capabilities that have been specially designed to operate in and capitalize upon the complexity of rich natural language domains of very large stores of text and multimedia documents. We have also integrated T2K with GATE from the University of Sheffield for leveraging their information extraction techniques. The D2K environment with the set of T2K components will be leveraged to extract facts from the Hurricane Katrina corpus.

2.2.4 Cross-validation of Automated Extraction Methods

After extracting relational structures for the Katrina EMONs, edges gathered by automated extraction will be compared against those produced by human coders (for documents on which both approaches were used). Measures of accuracy (including both false positive and false negative rates) will be computed for the automated system, taking the manually coded solution as the criterion set. Identified errors will be "spot-checked" by human coders for verification, and (where possible) to identify the source of the error. This information, in turn, will be used to further refine the automated extraction system. A second round of error assessment on a random sample of additional documents will be used to produce error estimates for the refined system, and these will be utilized as base error rates for the network inference task (see 2.3.1).

2.3 Analysis of Katrina EMONs

This section describes how the novel data collection techniques proposed above will extend the statistical modeling and theoretical efforts of the two ongoing NSF-ITR funded projects (RESCUE and CP2R).

2.3.1 Intertemporal Network Inference

While validated extraction technologies can offer an accurate account of the organizational interactions described in a given source document, they cannot provide insight into whether the interactions so identified actually exist. Similarly, such methods cannot, in and of themselves, suggest the presence of ties which are not mentioned by particular sources, nor reconcile disagreements between sources. To reconstruct the actual time-history of the Katrina EMONs, then, we require a modeling approach which can infer network ties from multiple, error-prone sources. Using Bayesian statistical methods, Butts (2003, 2004) has developed models for performing such inference in the presence of uncertain, possibly complex error processes. These models will be applied by the UCI team to the Hurricane Katrina data, to estimate the detailed intertemporal structure of the Katrina response EMONs. This analysis will not only supply point estimates of the interorganizational networks over time, but will also provide measures of uncertainty surrounding those estimates -- this will allow us to appropriately weight more poorly measured aspects of the network in subsequent analyses. As an additional benefit, the models to be employed will provide estimates of the accuracy associated with different types of information sources. This information can be leveraged for future studies of interorganizational networks.

2.3.2 Testing Multi-theoretical Multilevel (MTML) Models for Katrina EMONs

In prior and ongoing NSF-funded research, Contractor and his colleagues have developed a model to explain organization's motivations to create, maintain, dissolve, and reconstitute inter-organizational links

(Contractor & Monge, 2002; Contractor & Monge, 2003; Contractor, Wasserman & Faust, in press; Monge & Contractor, 2003). The MTML model included seven families of theoretical mechanisms: (1) *theories of self-interest*, (2) *theories of mutual interest and collective action* (3) *contagion theories* (4) *cognitive theories* (5) *exchange and dependency theories* (6) *homophily and proximity theories* and (7) *coevolutionary theory*. Each theoretical mechanism implies a specific structural "signature" in the observed interorganizational networks. Since organizations' motivations to forge their networks are driven by multiple theories at multiple levels (MTML), untangling the multiple structural signatures in the observed network structure is a non-trivial statistical endeavor. In our prior research, we empirically tested MTML predictions in over three dozen knowledge networks, using recent advances in exponential random graph modeling (ERGM) techniques (Contractor, Wasserman, & Faust, in press). Mathematical background can be found in Wasserman and Faust (1994) and more detailed descriptions of the *random graph models* (often referred to as the *p** family) can be found in the chapters of Carrington, Scott, and Wasserman (2005).

The proposed project extends our prior research effort on *two* dimensions. *First*, in our prior research we collected self-report data from relatively small networks. In the proposed study, the considerably larger networks generated by the automated network extraction procedures will pose new theoretical as well as methodological challenges. However, they will also provide much greater statistical power in terms of making theoretical inferences. *Second*, unlike our prior research effort, we will use our estimates from the ERGMs to generate theoretically based diagnostics on the state of the EMONs (e.g., "islands" of organizations not coordinating with other islands). By engaging with domain experts in the area of emergency response and critical physical infrastructure, we will benchmark the observed networks against theoretically deduced normative networks for first responders. This benchmarking will offer us the opportunity to prototype the development of real time (or near real time) dashboard-style decision support tools that will enable first responders to better monitor and manage their EMONs.

2.4 Plan for Completion of Research Objectives with Start Date of 1/11/05

Research Task	**Team**	**Completion Target**
Initial Data Collection		
Documentation from Official Sources	UCI	1/06
Third-party Documentation	UIUC	1/06
Network Data Extraction and Validation		
Identification of Named Entities	UCI/UIUC	2/06
Human Coding of Relational Information	UCI	5/06
Automated Network Extraction	UIUC	5/06
Cross-validation of Automated Extraction Methods	UCI/UIUC	7/06
Analysis of Katrina EMONs		
Intertemporal Network Inference	UCI	10/06
Testing MTML Models for Katrina EMONs	UIUC	10/06

3. Merit and Impact

3.1 Intellectual Merit

The proposed project is a pioneering effort at collecting large scale interorganizational networks. Collection will be based on a suite of web-crawling, data-mining, and text-extraction tools, coupled with sophisticated statistical methods for both inferring network links based on error-prone sources, and for assessing their structures in comparison to theoretically hypothesized structural tendencies. *First*, the proposed project will provide a systematic assessment of the relative viability and validity of manual, automated, and computer-augmented techniques for coding networks based on a large corpus of digital data. *Second*, the proposed project will provide an opportunity to scale up our ongoing methodological investigations using Bayesian statistical methods to infer network ties from error-prone data and exponential random graph models to test theoretically derived hypotheses about network structures. *Third*, the improvements in the quality of data and the statistical techniques applied to them will provide an unprecedented opportunity to uncover theoretical explanations for coordination among emergent multi-organizational networks in disaster response. *Fourth*, the data set produced by this research will be a key resource for social scientists, disaster researchers, information technologists, and policy analysts studying problems related to the Katrina response.

3.2 Broader Impact

The findings, tools, and methodologies derived from the proposed research will be immediately generalizable to a wide range of disaster response situations. Our ongoing partnerships with local, state and federal government as well as non-government agencies in the US and internationally as part of our current NSF ITR projects make us exceptionally well-equipped to rapidly incorporate our findings, tools, and methodologies into our ongoing education, training, and outreach workshops with response organizations. Our research will produce an account of interorganizational coordination in the Katrina response which is unprecedented in scope and detail, and which will itself serve as a tool for emergency managers and others seeking to understand what transpired during the operation. The intertemporal data set produced by this research, as well as the new tools for the automated extraction, measurement, and analysis of emergent multiorganizational networks will be made publicly available via NCSA's SONIC (Science of Networks in Communities) portal for use by disaster researchers and other social scientists, policy analysts, and emergency managers.

3.3 Diversity

Both teams are committed to recruiting research group participants from diverse backgrounds (underrepresented minorities, women, and other historically disadvantaged groups). At UIUC, the Office of Minority Affairs have excellent track records for helping us recruit and mentor women and minority students. Further, we will continue to sponsor McNair Program students. The UCI team will work with the Louis Stokes California Alliance for Minority Participation and groups such as Women In Computer Science to recruit talented undergraduates from a range of academic and personal backgrounds who may not otherwise have considered engaging in social science research.

3.4 Synergy with Current Work

As indicated, the proposed project extensively leverages and integrates across the theoretical and methodological insights developed in our two ongoing ITR projects investigating emergency response efforts (RESCUE and CP2R). In addition, the network data gathered, as part of the proposed project, from this unforeseen event, will significantly enhance our ability to test models that we had theoretically developed as part of our ITR projects, but had not anticipated being able to empirically validate in field settings.

4. Prior NSF Supported Research

Butts has received prior NSF research support for work in the areas of modeling, social networks, and crisis response. He is a co-PI on **ITR IIS-0331707**: *Collaborative Research: Responding to the Unexpected*, a large, interdisciplinary research effort focused on improving the use of information in response to crises and unexpected events. In its first two years, this award has lead to over 200 publications, working papers, and conference presentations, as well as ongoing connections with response organizations (including the LA County Office of Emergency Management, the California Governor's Office of Emergency Services, the cities of Irvine, San Diego, and Los Angeles, and the UCI Environmental Health and Safety Department). He was co-PI on **SES-0100999**: *Dissertation Improvement: Spatial Models of Large-Scale Interpersonal Networks*, which supported work on the stochastic modeling of geographically embedded interpersonal networks.

Contractor has received major NSF funding continuously for the past eleven years. His efforts have demonstrated the intellectual viability of an innovative analytic framework, coupling social science theory construction, computational modeling, web-based network mapping, and network analysis. He is a co-PI on **ITR CMS-0427089**: *IT-Based Collaboration Framework for Preparing Against, Responding to, and Recovering from Disasters Involving Critical Physical Infrastructures*, ($2,370,000.; 9/04-9/09) and PI on **IIS-0233449** (9/ 02 – 8/ 03): *Knowledge Networks and Emerging Heterarchies in Rebuilding New York*. He organized **IIS-0341928** (9/03 – 9/04): *Workshop to address IT challenges of 21st century society* and **SCI-0533892** (8/05- 8/06): Workshop on *The Role of Social Network Research in Enabling Cyberinfrastructure and the Role of Cyberinfrastructure in Enabling Social Network*. The research resulted in over 50 publications and involved over 3 dozen graduate students.

Peña-Mora's NSF research has focused on dynamic project management, collaborative environments, dispute resolution, and collaborative disaster response. The research outcomes from the prior NSF Awards **NSF/CONACyT IRIS-9630021**, **NSF CMS-9626315**, ***NSF PECASE/CAREER CMS-9875557***, **NSF CMS-0324501**, and **NSF/ITR CMS-0427089** are (1) multiple device collaborative environments and real time analysis systems for project management and disaster response; and (2) the initiation of the *Interaction Space* theory for geographically distributed teams involved in large-scale engineering projects and disaster relief. The research under these grants has resulted in 25 journal publications, 20 conference articles, 4 PhD thesis, 15 Master's thesis, 4 pending patents, and 2 textbooks.