

Proposal for New Investment in Faculty Hiring in Computational Social Science

Andrew McCallum* Jane Fountain* David Jensen* Michael Lavine+ Stuart Shulman* Donald Tomaskovic-Devey•

* Computer Science, ★ Political Science, • Sociology, + Mathematics & Statistics

UMass Amherst has the opportunity to capture a leading position in the nascent but rapidly expanding field of *computational social science*—a scientific and technological intersection that we and others¹ believe is poised to revolutionize the social sciences, expand the reach and impact of computer science, and enable our nation to understand the complex systems and social interactions that we must manage in order to address fundamental national and international challenges in the economy, energy, biosustainability, health care, education and crime.

This emerging science is driven by new sources of data (from the web, sensors, governments, universities, NGOs and corporations) as well as by recent advances in computational modeling, machine learning, statistics and social network analysis. The *Chronicle of Higher Education* encourages social and computer scientists to reach for national social impact, saying,² “Facebook and Wikipedia are just the beginning. The real power of social networks will be showcased by projects that unite far-flung participants to help track disease outbreaks, revolutionize neighborhood-watch programs, encourage energy conservation, and other civic- and community-oriented objectives.” Microsoft Chairman Bill Gates, in his keynote at the 2009 *International Conference on Information and Communication Technologies and Development*, spoke about third-world development—focusing on the point that “when you engage in a goal-oriented activity, such as development, progress can only be made when you measure the impact of your efforts.” He advocated statistics for measuring and modeling development efforts, and then changing the world through the understanding provided by these models.³ Harvard Professor and Institute Director Gary King says in his letter of support for our proposal, “[Computational social science] will comprise a large part of the future of the social sciences in the coming years.”

The UMass Amherst departments of Computer Science, Mathematics & Statistics, Political Science and Sociology propose to hire four new junior faculty, one in each department, in order to connect and build on our existing, disjoint strengths, forming a world-class cluster of excellence in this area. The new hires will catalyze interdisciplinary cooperation, fill holes in our proposed new educational programs, and multiply funding opportunities beyond our already outstanding track record of grants, contracts and awards. Our proposed interdisciplinary research and teaching collaboration spans the campus, connecting departments that have not historically collaborated, but have a strong desire to work together now.

Computational social science is particularly important now, as the nation strives to reduce energy use, improve health care, revive the economy, and strengthen our educational system. Addressing each of these challenges will require significant improvements in our understanding of the interactions of people, institutions, markets, and the other components that make up these complex systems. Without substantial innovation in the study of social, political, and economic systems, policy-makers and the public will be “flying blind” as we move through the dramatic changes likely to be wrought in the next several decades from new social forces, including climate change, globalization, and the increasing rate of technological innovation.

This computational revolution in social science research mirrors similar changes in the life sciences. Immense volumes of data about social and group behavior are now gathered and available, providing the social sciences with information that was inconceivable even just ten years ago. Furthermore, consider the following recent events and trends: The political campaign that led to the election of President Obama signaled a sea change in the use of data and digital tools in political campaigns. Digital social networks are key components of social change throughout the world, such as Twitter.com in Iran. The U.S. federal government has embarked on building a new level of information transparency and interactivity with extraordinary implications for governance and society. UMass students, whose have been immersed in a digital world since birth, readily recognize the social significance of computing in their own lives; and students in the social sciences increasingly rely on digital information and digital tools to form the core of their education.

¹ Program managers at NIH and DARPA, Directors of Research at Google and Raytheon, Institute directors and professors at Harvard, Duke, and other research leaders; see letters of support in Appendix F. See also *Science*, Vol. 323, p721, 6/2/09.

² [Researchers Call for National Strategy to Adapt Social Networks to Public Good](#), *Chronicle of Higher Ed.*, 5/04/2009

³ [Bill Gates to Development Researchers: Create and Share Statistics](#), Victoria Stodden’s paraphrase of Gates.

As further evidence of the rising importance of academic research at the intersection of computation and the social sciences, consider the remarkable impact and recognition achieved by the list of “prototype” individuals appearing in Appendix A, representing more mature versions of the junior faculty we propose to hire. Their accolades and achievements include a MacArthur Fellow “Genius” Award, a Sloan Fellowship, membership in the National Academy of Arts and Sciences, authorship of highly prominent books and articles in the journal *Science*, and training PhD students who became faculty at Stanford and Princeton. Motivations and descriptions of proposed new faculty hires are also in Appendix A.

There are many examples of relevant individual researchers at other institutions, but UMass has an opportunity to be distinguished by building a significant multi-person center of excellence in computational social science. There are relatively few other academic institutions that are well-placed to drive a coordinated multi-disciplinary effort. Harvard’s *Institute for Quantitative Social Sciences* and UCLA’s *Center for Human Complex Systems* have statisticians, but not strong connections to computer science; (see Gary King’s letter). Carnegie Mellon University’s *Center for Computational Analysis of Social and Organizational Systems* has an NSF-IGERT-funded PhD program and a good connection to computer science, but surprisingly weak ties to statistics and machine learning and the methodological innovation they bring. UC Irvine, UMichigan, and Northwestern are all beginning efforts in this direction, but, as Harvard’s David Lazar says in his letter of support, “The field is wide open.” We can find no university offering significant undergraduate training in this area.

As described below, UMass has remarkable strength in several related but disconnected islands of expertise. The proposed faculty hiring will bridge gaps between the social sciences, computer science and statistics, creating new interdisciplinary opportunities, and the high potential to build an international reputation for UMass in this area of growing national importance, practical impact, educational interest, and funding; (see letters from former DARPA PMs, Google’s Director of Research, and others.) We would then be well-equipped to study transnational civil society, advocacy group behavior, campaign and electoral processes (voting blocks, lobbying, funding), as well as public opinion, participation and policy-making (for climate, energy, health, crime, human rights, etc). We would be primed to research emerging methods of collaboration and consensus-building, including crowd-sourcing (Wikipedia and digital libraries), group dynamics & decision-making, distributed intelligence, privacy, open-source software, social networks of researchers, scientific collaboration, expert-finding, and team-building, as well as educational management, student admissions, student academic success and instructional methods. Hiring goals and the gaps they fill are further discussed in Appendix A.

Building on Excellence

This proposal will bridge four previously-separated islands of strength, each of which have a distinguished record of excellence in peer-reviewed publication, awards, and sponsored research. A list of departments and the faculty affiliated with this proposal, as well as a sampling of their extensive achievements appears in Appendix B.

The Department of Computer Science is ranked by the most recent Computing Research Association Survey in the top 10 nationally in artificial intelligence (the sub-area most relevant to this proposal). Its faculty have received over 50 international, national and UMass awards for research and teaching. The most recent departmental AQAD report states “Computer Science at UMass Amherst Computer Science is one of the strongest departments in the college and, in fact, one of the strongest departments at UMass. Both the college and the university could greatly benefit from building on this strength and highlighting interdisciplinary initiatives led by computer science.” Among the 7 CS faculty that would join this cluster are two UMass Distinguished Professors, two recent NSF CAREER award winners, multiple PIs with research funding ~\$1m/year, a researcher with over 33,000 citations. **The Department of Mathematics and Statistics** includes several faculty with strong expertise in Bayesian and computational statistics, which align closely with computational social science, including a Fellow of the American Statistical Association, and an Associate Editor of the Journal of the American Statistical Association. **The Department of Political Science** has the National Center for Digital Government, and several faculty with international reputations on connections between IT and politics, including the Editor-in-Chief of the *Journal of Information Technology and Politics*. **The Department of Sociology** graduate program is ranked 31st in nation, making it one of the highest ranked departments on campus. They have significant foundational strength in the study of social networks and dynamic modeling in the presence of relational complexity using a wide variety of source data including interview, historical, archival, survey, official records, and simulations. The faculty affiliated with this proposal include elected members of the Sociological Research Society and a Fellow of the American Statistical Association.

Opportunities in Scholarship and Research

What third-party evidence indicates strong or emerging demand for the proposed program? Scholarly publication at our proposed intersection is rising across all four disciplines, and includes recent articles in the journal *Science* and *Communications of the ACM*, and a special issue on computational social science in the top-ranked political science journal, *Political Analysis*. NSF-funded workshops in the area are also on the rise, and include “Coding across the Disciplines,” chaired by Assist. Prof. Stuart Shulman. In the past two months, Profs. Fountain and McCallum have been featured speakers at nationally-recognized workshops “Studying Society in a Digital World” and “New Directions in Text Analysis for Social Scientists.” See also the following sections and our letters of support from funders, industry leaders and fellow academics (Appendix G).

What is trajectory of the field, and what will be necessary to sustain leadership? The Obama Administration is focusing intently on the use of computational tools and applications to transform governance, public engagement and policy making. As a consequence, the major national funding agencies, NSF and NIH, and the IT industry are focusing their strategic priorities on computational methods for social analysis and use. Scientifically, the field is following a similar trajectory to that of biology, linguistics, and other computationally-revolutionized disciplines: access to massive data and the computational/statistical models to understand them is transformative; low-hanging fruit comes from applying off-the-shelf methodology, but the ultimate, sustained leadership will go to those who can cohesively innovate in the domain and the methodology—necessitating close collaboration between the social sciences, computer science and statistics among faculty whose motivations are aligned.

What are the best, worst, and most likely case scenarios for the success of the proposal? Best: This cluster establishes UMass Amherst as an international leader in an emergent field, garnering scholarly reputation, a unique undergraduate program, and increased external funding, as well as catalyzing new types of cross-campus collaboration. As digitized data with social and political relevance proliferates and threatens to overwhelm the ability of researchers and policy makers to analyze the information, UMass presents research-based solutions to a wide array of basic and applied social research problems. Worst: Production of tools with trivial applications, scholarship that is not published or cited, and failure to attract students or outstanding young faculty through our hiring initiative. Bureaucratic red tape inhibits cross-departmental research, and detracts from our mission. We do not anticipate these scenarios and have the experience in emerging research areas to avoid such pitfalls. Likely: Something close to “Best,” with degree of success depending on the quality of our new hires, as well as existing PI energy and dedication.

What is the strategy for increasing our competitiveness, and how would the requested positions figure into that strategy? Distinguish ourselves from other universities by building a coherent, collaborating cluster of excellence across all four necessary sub-disciplines, using new hires to bridge our considerable (currently disjoint) foundational strength. We would engender good esprit de corps among the faculty in the cluster—building ties across campus; (the camaraderie among the PIs is already very good). Together develop external funding proposal for large-scale centers—both NSF and NIH are eager to identify and fund leaders in this domain who will build the national (and global) research capacity and intellectual infrastructure for the country. New hires in targeted areas will increase our credibility for these proposals.

Opportunities in Teaching and Attracting Out-of-State Students

We propose to develop a Certificate in computational social science to complement the current bachelor’s degree programs in computer science, math and across the social sciences. This certificate would combine theoretical and applied courses across the core departments. Current faculty can teach existing relevant courses in data mining, natural language processing, information retrieval, introductory programming (computer science), elections, globalization, legislative process, international political economy, public policy (political science), social movements, corporate crime, work & society, the family (sociology), introduction to statistics, linear algebra, Bayesian latent variable modeling (statistics). What we need are new faculty with the expertise to teach “bridging” courses such as computational social network analysis, networked behavior, computational game theory (computer science), quantitative modeling of public policy, opinion and legislative bodies (political science), behavioral social network analysis, quantitative modeling of social phenomena (sociology). These courses would both (a) fill holes in a well-designed coherent undergraduate education in this area, and (b) provide bridges that would prepare and encourage students beginning in one sub-discipline to take courses in another. The courses would also create opportunities for team-teaching—creating more vibrant courses and further opportunities for faculty collaboration.

This certificate will help attract to UMass top, motivated undergraduate students. Beyond general undergraduate fascination with Facebook and the sociological impact of the web, there is more academic evidence of dramatic demand for education in this area: A stand-alone computational social network course taught at UPenn had the fastest-growing enrollment of any course in its computer science department's history. A related course at Cornell has been successful enough that its instructors are publishing a textbook. Unlike these stand-alone courses, we are proposing a multi-course, integrated, cross-disciplinary undergraduate program—to our knowledge, the first of its kind in the country. The most comparable alternatives we could find are the *Program in Social Complexity* at George Mason University and Carnegie Mellon University's program in *Computation, Organization and Society*, but both of these are Ph.D.-level programs.) Both undergraduates and their parents are likely to note the rising demand in these areas from employers; (see letter of support from Raytheon's Director of Contract R&D).

Opportunities in Fund-Raising

The faculty proposing this cluster have excellent records of obtaining research funding. However, even larger opportunities await a coordinated and cross-disciplinary effort in computational social science. Appendix D provides an extensive listing of relevant funding programs at seven federal agencies.

For example, the NSF has demonstrated clear interest computational social science through multiple programs in recent years including: *Social-Computational Systems* (\$18m, "seeks to capitalize upon the collaborative knowledge and research methods of investigators in the computational and human sciences"), *Human and Social Dynamics*, (\$65m, "the convergence of ... the cognitive and social sciences, mathematical and physical sciences, and computer and information sciences allows the development of basic knowledge about the evolution of individual, group, and organizational behavior with a rigor never before possible," now closed), *Science of Science and Innovation Policy* (\$31m, "models, analytical tools, data and metrics that can be applied in the science policy decision making process... analyzing innovation in organizations"), and *Cyber-Enabled Discovery and Innovation Program* (\$40m in 2008 alone, "exploit computation as a means of achieving deeper understanding in the natural and social sciences"). NSF program managers have also indicated that center-level funding in this area is likely. Appendix D lists many relevant NSF programs.

The interest in computational social science goes far beyond NSF. As noted by Peter Brooks in his letter of support, **DARPA** is "pursuing several initiatives in computational social science." Current DARPA Program Manager Sean O'Brien noted in a recent public speech that "the capabilities that [DARPA's Information Processing Techniques Office] is envisioning may allow us to subject [social science] theories to rigorous empirical evaluations in ways that could revolutionize the manner in which we anticipate and respond to threats from abroad... and we may revolutionize the social sciences along the way." (O'Brien 2007). This interest is reflected in several recent funding initiatives described in the Appendix. As noted by several letters of support, DARPA has also recently requested white papers on "Technologies for the Applications of Social Computing" to explore "the development of new technologies to rapidly create theoretically-informed, data-driven models of complex human, social, cultural, and behavioral dynamics."

A range of related initiatives are being supported by other federal funders of research. In the past year alone, the **National Institutes of Health** has issued two calls for proposals directly calling for broad-based research that combines social science and computational methods to better understand research collaboration and health outcomes; (see also NIH Program Manager Edmund Talley's letter of support). The **Office of Naval Research**, the **Air Force Office of Scientific Research**, and the Intelligence Advanced Research Projects Activity (**IARPA**) are all pursuing to programs directly related to computational social science.

Finally, several highly successful corporations have come to see large-scale analysis of data on social relationships and behavior as crucial to their businesses. **Companies** such as Google, Yahoo!, Facebook, LinkedIn, and others have discovered the value of applying the latest results from academic research to better understand the complex social interactions that increasingly take place on their networks. Yahoo Research, for example, recently announced a list of key scientific challenges whose solution is vital to their business, including: (1) understanding the structure and evolution of social groups and networks; (2) decision making, social influence, diffusion, and collective decisions; and (3) networking and collaborative problem solving; and (4) modeling massive social networks. Companies such as Google and Yahoo have historically been a significant source of UMass sponsorship (more than \$800k/year), research support, and job opportunities for graduates. In addition, graduates in computational social science have strong potential to start their own companies and later become strong supporters of the University when those companies are successful.

UMass Internal and External Support

The strength and interest in this proposal is indicated by the high level of support that has been committed, should this proposal be accepted. Beyond internal support, Raytheon Inc has already committed \$5000.

The proposal PIs: The PIs have committed to providing \$40,000 in RTF to help defray faculty new-hire start-up costs. Collectively the affiliated faculty are already funding over 40 graduate research assistantships in related research areas, and anticipate funding even more in the future thanks to the synergistic opportunities envisioned by this proposal. The PIs have sufficient lab space to house current and future research assistants and activities. PI McCallum has committed to leading the creation of a new Institute at UMass that will form the cultural center of the proposed collaboration, and provide a structure for resources and governance in order to ensure its maintenance into the future. The PIs will contribute to the management of this effort in addition to their existing service duties. Governance and assessment are covered further in Appendix B. The PIs will have self-funded a one-day workshop in July 2009 in which they and other related faculty will make research presentations and discussed potential collaboration.

Existing Institutes and Centers: The National Center for Digital Government has committed \$5000 in support of the proposed cluster. The Center for Intelligent Information Retrieval (CIIR) would provide \$5000 to facilitate workshops and other coordinating activities. The CIIR has also committed the availability of computational resources, including a 40-CPU cluster computer and over 5 terabytes of disk space.

Department Chairs: The Chair of Computer Science has committed \$125,000 in start-up funds. The CS department has also committed to support the development of the undergraduate Certificate described above, as well as the creation of a new, targeted computer science major and minor “track” in Computational Social Sciences. The Chairs of Political Science and Sociology have committed to \$5000 each in start-up, a two-course teaching release per hire, and development of the Certificate. The Chair of Math & Statistics has offered \$15,000-\$20,000 in start-up. All chairs also commit to sufficient space and ongoing administrative support.

Deans: The Dean of NSM has committed to \$100,000 in start-up funds. In addition, the Dean of NSM has committed to fully funding the hiring of an additional junior-level tenure-track faculty member in the distinct, but supporting research area of machine learning to replace a recent death in the department; this faculty member would not substitute for a computer scientist working in computational social science, but would provide assistive, synergistic expertise in computational modeling. The Dean of SBS has committed \$40,000 in startup contributions, plus an additional \$15,000 over five years toward supporting seminars and workshops.

Facilities and Return on Investment

Appendix C contains a budget for this proposal, including cash flow and return on investment. As described below, ***we will require no new facilities or renovations***. Proposed new faculty hires can be accommodated in existing buildings and office space. Our research work will require data, computation, workshops, travel and RAships, all of which can be gathered through grant funding and other support we will raise.

Total startup funds of \$275k are entirely covered by contributions from the PIs, departments and colleges, ***requiring no startup contribution from the chancellor***. Salary costs are estimated based on recent average starting salaries in each department (or higher values judged sufficient to attract top talent), with 3% raises each year. The calculations are based on hiring in CS and Political Science in the first year, and then Sociology and Math the following year; but many different schedules are possible.

Research income estimates are justified as follows: Historically **Computer Science** new-hire faculty funding averages \$200k/year by the 3rd year, and rises to a sustained \$350k/year thereafter. However, these figures are averaged across all CS sub-areas; within the sub-areas relevant to this proposal, the figure is significantly higher—averaging above \$1m/year; (see Appendix F). Therefore, we claim that our 7th-year estimate of \$650k/year is actually conservative. Historically new **Sociology** faculty do not receive substantial external research grants until after promotion. Recent exceptions, with profiles more similar to the proposed hires, have averaged \$50k in early years and \$150k after four years; however because of the synergies produced by the cluster hire we expect this candidate to do somewhat better. We similarly expect our **Statistics** new hires to obtain higher funding than average also; for example Andrew Gelman (Appendix A) has been awarded > \$1.2m/8 years from NSF alone, and he additionally obtains funding from NIH, Yahoo and elsewhere. **Political Science** faculty in sub-areas related to or collaborating with CS have been remarkable fund-raisers, *e.g.* Fountain >\$2.7m/7 years, Shulman >\$3m/9 years in NSF funding alone.

According to our calculations, with respect to salaries, our proposal becomes ***cash-flow-positive after 4 years***, and almost completely recovers (93%) cumulative salary costs by the 8th year, after which we become a cumulative revenue generator.

Appendix A: Prototype New Faculty Members

As evidence of the rising importance of academic research at the intersection of computation and the social sciences, consider the remarkable impact and recognition achieved by the following list of "prototype" individuals who represent more mature versions of the junior faculty we propose to hire.

- Jon Kleinberg (**Computer Science**, Cornell, age 38): Research in social and information networks, on the web and elsewhere. MacArthur Fellow "Genius" Award, Sloan Foundation Fellowship, ACM InfoSys Award, ONR Young Investigator Award, NSF CAREER Award, Guggenheim Fellow, National Academy of Arts and Sciences, National Academy of Engineering. Citation "H-index" of 44.
- Duncan Watts (**Sociology**, Columbia, now Yahoo Research, age 38): Research in social network dynamics, with publications in *Annual Review of Sociology*, *Science*, *Nature*, *Journal of Theoretical Biology*, and *Physical Review Letters*. Created the "Collective Dynamics Group" at Columbia, which applies mathematical and computational techniques to problems relevant to the social sciences. Recent Sociology Ph.Ds from this group got faculty positions at Stanford (Paolo Parigi) and Princeton (Matt Salganik).
- David Lazer (**Political Science and Public Policy**, Harvard, age 43): Research in complexity and social networks to examine global regulatory regimes, networks of trade, and social influence processes in groups. Created the Program on Networked Governance at Harvard, which fosters research at the intersection of governance, information, technology and networks (social and hard-wired). Lead author of the *Science* article, "Computational Social Science"; research featured on PBS, "60 Minutes," NPR, Wall St. Journal, the Economist, and in *Nature*.
- Andrew Gelman (**Statistics**, Columbia, age 45): Research in mathematical models of voting, social behavior, and social network structure. Outstanding Statistical Application award from the American Statistical Association, Council of Presidents of Statistical Societies award for outstanding contributions by a person under the age of 40, award for best article published in the *American Political Science Review*. His books include *Bayesian Data Analysis* (with Carlin, Stern, and Rubin), and *Red State, Blue State, Rich State, Poor State: Why Americans Vote the Way They Do* (with Park, Shor, Bafumi, and Cortina).

New Faculty Needs at UMass

While existing UMass faculty committed to supporting growth in this area are excellent, they are inadequate to address many of the growing list of research and teaching opportunities. First, we lack a critical mass of faculty whose principal commitment is to research in computational social science. These faculty will form the bridges that tie together our existing strengths. Second, most of the existing faculty committed to this proposal are fairly senior, and any broad-scale endeavor at UMass toward computational social science also needs the input of new skills and connections to fuel the next generation of discoveries and methodological innovations.

Finally, there are key gaps in the expertise and focus represented by existing faculty. For example, in computer science, we need faculty with focus on social science applications who simultaneously have expertise in areas including large-scale probabilistic representation and reasoning for social networks, structure discovery and causal discovery in complex systems, distributed decision-making, and sensor networks for sociological data. In political science, we need bridging faculty with expertise in quantitative and probabilistic modeling of legislative behavior, public opinion, public participation, policy-making, and government data, as well as quantitative networked governance and cyber-infrastructure. In sociology, we need computationally-leaning expertise in large-scale social network analysis and quantitative modeling of social behavior. Finally, in math and statistics, we need faculty with expertise in Bayesian statistical modeling of networks and hierarchical and multi-level modeling who also focus on social science applications.

Appendix B: Building on Excellence

Computer Science

The Department of Computer Science is ranked by the most recent Computing Research Association Survey in the top 10 nationally in artificial intelligence (the sub-area most relevant to this proposal). Its faculty have received numerous national and international research and teaching awards, and include 9 ACM Fellows, 5 IEEE Fellows, 7 AAAI Fellows, 7 international lifetime achievement awardees, 11 UMass Lilly Teaching Fellows, 5 NSM Distinguished Research awardees, 5 NSM Outstanding Teacher awardees, and 15 NSF CAREER Award recipients. The department raised \$16.2m in individual, collaborative, and interdisciplinary grants in fiscal year 2008 from NSF, DARPA, ARL, ONR, other agencies, industry and private foundations.

It is a departmental priority to demonstrate how CS can revolutionize neighboring fields, and to use these new application areas to inspire new broadly-applicable CS research. The most recent departmental AQAD report states “Computer Science at UMass Amherst Computer Science is one of the strongest departments in the college and, in fact, one of the strongest departments at UMass. Both the college and the university could greatly benefit from building on this strength and highlighting interdisciplinary initiatives led by computer science.” Another high departmental priority is to attract more undergraduate majors (especially women and minorities) and it sees socially-relevant application areas as a major component of that strategy.

Computer Science faculty who would be affiliated with this cluster include

- *Professor Andrew McCallum*: Research in information extraction, network analysis of textual data, bibliometrics, machine learning. AAAI Fellow, UMass NSM Distinguished Research Award, Program chair of 2008 International Conference on Machine Learning, UMass Lilly Teaching Fellow. IBM Faculty Partnership Award. Google Research Award. Director, Information Extraction and Synthesis Laboratory. Over 120 publications, 13,000 citations. Averages ~\$2m/year in research funding raised.
- *Associate Professor David Jensen*: Research in relational data mining, knowledge discovery and automated causal discovery. Former congressional staffer, leader of DARPA ISAT study. Editorial boards of the *Journal of Artificial Intelligence Research*, *Journal of Privacy Technology*. Co-Chair, IJCAI 2003 Workshop on Learning Statistical Models from Relational Data; Co-Chair, AAAI 2000 Workshop on Learning Statistical Models from Relational Data; Co-Chair, 1998 AAAI Fall Symposium on AI & Link Analysis. Director, Knowledge Discovery Laboratory.
- *Distinguished University Professor Bruce Croft*: Research in many areas of information retrieval, including retrieval models, representation, Web search, query processing, cross-lingual retrieval, and search architectures. Over 200 research papers and multiple books. ACM Fellow. Research Award from the American Society for Information Science and Technology. Gerard Salton (lifetime achievement) Award from the ACM Special Interest Group in Information Retrieval (SIGIR). Member of the National Research Council Computer Science and Telecommunications Board, 2000-2003. Editor-in-Chief of ACM Transactions on Information Systems, 1995-2002. Director, Center for Intelligent Information Retrieval (NSF S/IUCRC).
- *Distinguished Professor Victor Lesser*: Research in multi-agent networked systems, control and organization of complex artificial intelligence systems, plan recognition, and computer-supported cooperative work. AAAI Fellow. IEEE Fellow. IJCAI (lifetime achievement) Award for Research Excellence. UMass NSM Distinguished Research Award. UMass Chancellor’s Award for Outstanding Accomplishments in Research and Creative Activity. Founding President of the International Foundation of Autonomous Agents and Multi-Agent Systems (IFAAMAS). Director, Multi-Agent Systems Laboratory.
- *Professor James Allan*: Research in interactive information retrieval and organization, topic detection and tracking, automatic information organization. UMass Lilly Fellow.
- *Assistant Professor Deepak Ganesan*: Research in sensor networks, networking and data management issues in sensor networks. UMass Junior Faculty Fellow. UMass Lilly Fellow. NSF CAREER Award. IBM Faculty Award.

- *Assistant Professor Jerome Miklau*: Research in databases, including privacy, confidentiality and web-based data exchange, confidentiality and authenticity for exchanged data, database forensics. Web Chair for SIGMOD/PODS 2009. NSF CAREER Award.

These faculty form a foundation of related expertise to build on, but none of them have the background or research agendas to overlap with the motivations of social scientists.

Political Science

The Department of Political Science is in the midst of a transformation from a more traditional department built around sub-fields to one that is motivated by problem-driven and cross-boundary research.” (J. Hird, letter of support.)

Political Science faculty who would be affiliated with this cluster include

- *Professor Jane Fountain*: Research in organization theory; information technology and politics; economic sociology; public policy; political sociology; gender; science, technology and society. Director of the *Science, Technology, and Society Initiative*; the *National Center for Digital Government*; and the *Women in the Information Age Project*. Member of the World Economic Forum Global Advisory Council on the Future of Government. Member of American Bar Association blue ribbon commission on the Future of e-Rulemaking and a member of Science, Technology and Society program review committee at the National Science Foundation. Served on several advisory bodies for organizations including the Social Science Research Council, the Internet Policy Institute, the National Science Foundation. Given invited lectures and keynote addresses and worked with governments and research institutions including the World Bank, the European Commission, Knowledge Management Asia Pacific, Japan, Portugal, the Netherlands, Nicaragua, Chile, Estonia, Hungary, Slovenia, Saudi Arabia, and the United Arab Emirates.
- *Associate Professor Ray La Raja*: Research interests include political parties, interest groups, elections, campaign finance, political participation, state politics, public policy and political reform. Managing editor of the *The Forum: A Journal of Applied Research in Contemporary Politics*. Academic Advisory Board of the *Campaign Finance Institute*. Outstanding Teaching Award, UMass, Social and Behavioral Sciences, 2005-2006.
- *Assistant Professor Stuart Shulman*: Research in electronic rulemaking, human language technologies, manual annotation, digital citizenship, and service-learning efforts in the United States. He is editor of the *Journal of Information Technology and Politics* and directs the *Qualitative Data Analysis program*. Strong record of funding from the computer science-oriented CISE Directorate at NSF (nine awards in nine years). Organizer and chair of 2009 UMass workshop on “YouTube and the 2008 Election Cycle in the United States.”

Sociology

The Department of Sociology graduate program was ranked 31st in nation in the most recent US News & World Report ratings, making it one of the highest ranked departments on campus. Its faculty have significant foundational strength in the study of social networks and dynamic modeling in the presence of relational complexity using a wide variety of source data including interview, historical, archival, survey, official records, and simulations.

Sociology faculty who would be affiliated with this cluster include

- *Professor Tomaskovic-Devey*: Research in processes of workplace inequality, particularly discrimination and segregation, as well as topics in economic sociology more generally. His long-term agenda is to move sociological studies of inequality to a more fully relational and organizational stance. He is advancing this agenda through studies of jobs and workplaces, as well as social relationships between jobs within workplaces and the social relationships that link organizations to each other. He has expertise in dynamic modeling,

including hierarchical and multiple membership models. Elected members of the Sociological Research Society (Sociology's research honorary society).

- *Professor Anderton*: Research in historical and mathematical population studies, statistics, environmental health and policy. Elected members of the Sociological Research Society. Director of the Social and Demographic Research Institute. Sociological Research Association, elected 2008. Social Science History, Editor 2006-Current. NIH SSPS Study Section, 2004-2008. Expertise in dynamic modeling. Fellow of the American Statistical Association.
- *Professor Faulkner*: Expertise in business and professions, markets, careers, corporate crime, qualitative methods.
- *Assistant Professor Erikson*: Expertise in comparative historical sociology, social networks, economic sociology, complex organization, and foreign trade. Center for International Business Education and Research Grant (CIBER), Chazen Institute, Columbia Business School. Runner up for Best Graduate Student Paper, ASA Economic Sociology Section.
- *Assistant Professor Papachristos*: Uses social network analysis and dynamic models to examine the social structures and group processes at the heart of interpersonal crime, violence and delinquency. He is currently involved in the evaluation of the *Project Safe Neighborhoods* program in Chicago and has recently completed data collection on a four-neighborhood study of how illegal and pro-social networks of probationers and parolees influence offending patterns, interpersonal violence, gun markets, and perceptions of neighborhood social order. Harvard University, Kennedy School of Government – Visiting Scholar. Professor of the Year, 2008. Student Sociological Council. University of Massachusetts, Amherst. Department of Sociology. Hans Mattick Award for Best Member Publication 2002, Illinois Academy of Criminology.

Sociology is seeking a faculty member to bridge the gap between its existing strengths and computationally intensive modeling and data collection opportunities in computer science. Furthermore, Erikson has an offer from elsewhere and may be leaving the department.

Mathematics and Statistics

The Department of Mathematics and Statistics includes a group of seven tenure-track statisticians, several of whom have expertise in Bayesian and computational statistics, which align closely with computational social science.

Mathematics & Statistics faculty who would be affiliated with this cluster include

- *Professor Michael Lavine*: Research in Bayesian robustness, Bayesian nonparametrics, spatial statistics, Markov random fields, dynamic state space models. Applications to ecology and the environment and neurobiology. Fellow of the American Statistical Association.
- *Associate Professor John Staudenmayer*: Expertise in Bayesian data analysis, measurement error, smoothing, mixed models, with applications to toxicology, environmental health. Associate Editor of the Journal of the American Statistical Association.
- *Associate Professor Erin Conlon*: Research in Bayesian data analysis for the analysis of genomic data and comparative genomics, microarray and DNA sequence analysis. Visiting Scholar, Functional Genomics, Institute for Pure and Applied Mathematics (IPAM), University of California, Los Angeles.

A significant priority of the department is to form better ties to Computer Science. The department has a small faculty community of Bayesians, but is lacking someone focusing on applications to the social sciences.

Appendix C: Governance and Assessment

Governance

PI McCallum has committed to leading the creation of a new institute at UMass that will form the cultural center of the proposed collaboration, and provide a structure for resources and governance in order to ensure its maintenance into the future. The institute will have a Director, as well as three Assistant Directors to represent the multiple departments in this proposal. Tentatively, those positions would be filled as follows: Andrew McCallum (Director), Jane Fountain, Donald Tomaskovic-Devey and Michael Lavine (Assistant Directors).

These four Directors, as well as additional personnel they appoint, will form a Steering Committee, which will discuss, collaborate, (strive for consensus, and vote as necessary) in order to:

- find and suggest additional UMass faculty to join the effort, consider requests to join the effort, and generally manage the membership roster of the institute;
- discuss world-wide research trends and where our effort fits into those trends; brainstorm about faculty hiring candidates;
- help new and affiliated faculty obtain resources and facilities to conduct their research; identify new collaborative possibilities; find grant proposal opportunities;
- create and manage the new undergraduate Certificate in computational social science; coordinate with the teaching duties with the departments involved;
- report to home departments evaluations of participants' contributions to the effort in research and teaching;
- informally arbitrate disagreements and complaints concerning this effort.

The Steering Committee will meet as necessary, most likely 2-3 times per semester. Directors and other Steering Committee members will perform their duties in addition to existing service duties, without release time.

Furthermore, the institute will have an Advisory Council consisting of the chairs (or their designees) of the four departments involved.

The institute will use committed funds to sponsor a regular research seminar series, with speakers both external and internal to UMass. This will help maintain an intellectual and social bond among its participants.

Existing Evidence of Compatibility and Likelihood to Collaborate

The PIs will have self-funded a one-day workshop July 10th in which they and other related faculty will make research presentations and discuss potential collaborations.

There are several examples of preliminary collaboration: (1) Schulman's Qualitative Data Analysis Program laboratory will be housed in computer science department space, under an MoU researched between SBS and CS. (2) Sociology and Political science have a long history of fruitful collaboration, the *Science, Technology and Society* initiative being the most germane to this proposal. Tomaskovic-Deby and Yucel (formerly of Biostatistics/Public Health) have an ongoing NSF/NIH project. (3) McCallum, Allan and Erikson are Co-PIs on a UMass Nanotechnology Center NSF CDI proposal, working on bibliometrics, expert-finding, social networks among nanotechnology researchers. (4) Jensen and Papachristos have had some discussions about joint research on building statistical models of organized crime networks. (5) McCallum has given a research presentation in a seminar series at the National Center for Digital Government. (6) There is successful history of collaboration between Math and CS, but not in this research area.

Evaluation

The performance of the cluster should be measured in terms of (a) research funding, (b) publications and citations, (c) number of students advised and their publications, awards and other contributions to research and teaching, (d) placement of affiliated undergraduates and graduate students in relevant graduate programs, faculty positions, or industry, (e) workshops organized.

Appendix D: Return on Investment

	Relevant rates	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8	Total to 8 years
Computer Science (CS)										
CS Startup (amortized)		(\$50,000)	(\$50,000)	(\$50,000)	(\$50,000)	(\$50,000)	\$0	\$0	\$0	(\$250,000)
CS Salary	3%	(\$88,000)	(\$90,640)	(\$93,359)	(\$96,160)	(\$99,045)	(\$102,016)	(\$105,077)	(\$113,777)	(\$674,297)
CS Research Income		\$0	\$150,000	\$200,000	\$350,000	\$450,000	\$550,000	\$650,000	\$750,000	\$2,350,000
CS Research Overhead Return	57%	\$0	\$54,459	\$72,611	\$127,070	\$163,376	\$199,682	\$235,987	\$272,293	\$853,185
Mathematics and Statistics (MS)										
MS Startup (amortized)		\$0	(\$15,000)	(\$15,000)	\$0	\$0	\$0	\$0	\$0	(\$30,000)
MS Salary	3%	\$0	(\$70,000)	(\$72,100)	(\$74,263)	(\$76,491)	(\$78,786)	(\$81,149)	(\$83,584)	(\$452,789)
MS Research Income		\$0	\$0	\$90,000	\$150,000	\$200,000	\$200,000	\$250,000	\$250,000	\$890,000
MS Research Overhead Return	57%	\$0	\$0	\$32,675	\$54,459	\$72,611	\$72,611	\$90,764	\$90,764	\$323,121
Political Science (PS)										
PS Startup (amortized)		(\$12,500)	(\$12,500)	\$0	\$0	\$0	\$0	\$0	\$0	(\$25,000)
PS Salary	3%	(\$75,000)	(\$77,250)	(\$79,568)	(\$81,955)	(\$84,413)	(\$86,946)	(\$89,554)	(\$92,254)	(\$574,685)
PS Research Income		\$0	\$50,000	\$100,000	\$200,000	\$300,000	\$300,000	\$350,000	\$350,000	\$1,300,000
PS Research Overhead Return	57%	\$0	\$18,153	\$36,306	\$72,611	\$108,917	\$108,917	\$127,070	\$127,070	\$471,975
Sociology (SO)										
SO Startup (amortized)		\$0	(\$12,500)	(\$12,500)	\$0	\$0	\$0	\$0	\$0	(\$25,000)
SO Salary	3%	\$0	(\$75,000)	(\$77,250)	(\$79,568)	(\$81,955)	(\$84,413)	(\$86,946)	(\$89,554)	(\$485,131)
SO Research Income		\$0	\$0	\$50,000	\$100,000	\$100,000	\$150,000	\$150,000	\$200,000	\$550,000
SO Research Overhead Return	57%	\$0	\$0	\$18,153	\$36,306	\$36,306	\$54,459	\$54,459	\$72,611	\$199,682
PI, Dept and College contributions										
Proposal PIs contribution to startup		\$8,000	\$8,000	\$8,000	\$8,000	\$8,000	\$0	\$0	\$0	\$40,000
CS Chair contribution to startup		\$25,000	\$25,000	\$25,000	\$25,000	\$25,000	\$0	\$0	\$0	\$125,000
MS Chair contribution to startup		\$0	\$7,500	\$7,500	\$0	\$0	\$0	\$0	\$0	\$15,000
PS Chair contribution to startup		\$2,500	\$2,500	\$0	\$0	\$0	\$0	\$0	\$0	\$5,000
SO Chair contribution to startup		\$0	\$2,500	\$2,500	\$0	\$0	\$0	\$0	\$0	\$5,000
SBS Dean contribution to startup		\$10,000	\$20,000	\$10,000	\$0	\$0	\$0	\$0	\$0	\$40,000
NSM Dean contribution to startup		\$17,000	\$24,500	\$24,500	\$17,000	\$17,000	\$0	\$0	\$0	\$100,000
Chancellor contributions										
Chancellor contribution to startup		\$0	\$0	\$0	\$0	\$0	\$0	\$0	\$0	\$0
Total salaries, yearly		(\$163,000)	(\$312,890)	(\$322,277)	(\$331,945)	(\$341,903)	(\$352,160)	(\$362,725)	(\$385,168)	
Running total		(\$163,000)	(\$475,890)	(\$798,167)	(\$1,130,112)	(\$1,472,015)	(\$1,824,176)	(\$2,186,901)	(\$2,572,069)	
Overhead return, summed over departments		\$0	\$72,611	\$159,745	\$290,446	\$381,210	\$435,669	\$508,280	\$562,739	
Overhead return, running total		\$0	\$72,611	\$232,357	\$522,803	\$904,013	\$1,339,682	\$1,847,962	\$2,410,701	
<small>Cash flow = (Overhead return - Chancellor startup - Total salaries)</small>										
Cash flow, yearly		(\$163,000)	(\$240,279)	(\$162,531)	(\$41,499)	\$39,307	\$83,508	\$145,555	\$177,571	
Cash flow, running total		(\$163,000)	(\$403,279)	(\$565,810)	(\$607,309)	(\$568,002)	(\$484,494)	(\$338,939)	(\$161,368)	
<small>Return = Overhead return / (Chancellor startup + Total salaries)</small>										
Return on Investment, annual		0%	23%	50%	87%	111%	124%	140%	146%	
Return on Investment, total		0%	15%	29%	46%	61%	73%	85%	94%	

Appendix E: Funding Opportunities

Below are details of recent funding solicitations in computational social science.

NOTE: All text is taken from agency websites and solicitations. Awards or overall budget numbers are provided where available.

National Science Foundation (NSF)

Cyber-Enabled Discovery and Innovation (NSF 08-604)

Synopsis — Cyber-Enabled Discovery and Innovation (CDI) is NSF's bold five-year initiative to create revolutionary science and engineering research outcomes made possible by innovations and advances in computational thinking. Computational thinking is defined comprehensively to encompass computational concepts, methods, models, algorithms, and tools. Applied in challenging science and engineering research and education contexts, computational thinking promises a profound impact on the Nation's ability to generate and apply new knowledge. Collectively, CDI research outcomes are expected to produce paradigm shifts in our understanding of a wide range of science and engineering phenomena and socio-technical innovations that create new wealth and enhance the national quality of life.

Relevant text — "Congruent with the three thematic areas, CDI projects will enable transformative discovery to identify patterns and structures in massive datasets; exploit computation as a means of achieving deeper understanding in the natural and social sciences and engineering; abstract, model, simulate and predict complex stochastic or chaotic systems; explore and model nature's interactions, connections, complex relations, and interdependencies, scaling from sub-particles to galactic, from subcellular to biosphere, and from the individual to the societal; train future generations of scientists and engineers to enhance and use cyber resources; and facilitate creative, cyber-enabled boundary-crossing collaborations, including those with industrial and international dimensions, to advance the frontiers of science and engineering and broaden participation in STEM fields."

2008 Awards — \$40 million

Social-Computational Systems (NSF 09-559)

Synopsis — The Social-Computational Systems (SoCS) program seeks to reveal new understanding about the properties that systems of people and computers together possess, and to develop a practical understanding of the purposeful design of systems to facilitate socially intelligent computing. By better characterizing, understanding, and eventually designing for desired behaviors arising from computationally mediated groups of people at all scales, new forms of knowledge creation, new models of computation, new forms of culture, and new types of interaction will result. Further, the investigation of such systems and their emergent behaviors and desired properties will inform the design of future systems. The SoCS program will support research in socially intelligent computing arising from human-computer partnerships that range in scale from a single person and computer to an Internet-scale array of machines and people. The program seeks to create new knowledge about the capabilities these partnerships can demonstrate - new affordances and new emergent behaviors, as well as unanticipated consequences and fundamental limits. The program also seeks to foster new ideas that support even greater capabilities for socially intelligent computing, such as the design and development of systems reflecting explicit knowledge about people's cognitive and social abilities, new models of collective, social, and participatory computing, and new algorithms that leverage the specific abilities of massive numbers of human participants.

Relevant text — "The SoCS program seeks to capitalize upon the collaborative knowledge and research methods of investigators in the computational and human sciences, recognizing that researchers in computer science and

related disciplines often focus on the limits and capabilities of computation in isolation from the people that use computation, while researchers in the social sciences often focus on the use of technology or the capabilities of people with limited impact on how such knowledge can influence the design of new technologies. Proposals that reflect collaborative efforts spanning computational and human centered approaches and perspectives are specifically encouraged."

2001-2008 Awards — \$18.1 million

Next Generation Cyberinfrastructure Tools With Applications to Complex Behavior of Organizations and Individuals (NSF 05-563)

Synopsis — Researchers in the social and behavioral sciences and computer and information sciences have many important synergistic relationships. One way in which this is manifest is in the development and utilization of data. On the one hand, social and behavioral scientists find new ways to create and analyze data in their endeavors to describe human and organizational behavior. On the other hand, computer and information scientists conduct research that yields new ways to improve both domain-specific and general-purpose tools to analyze and visualize scientific data -- such as improving processing power, enhanced interoperability of data from different sources, data mining, data integration, information indexing and data confidentiality protection - or what we have termed cyberinfrastructure tools.

Relevant text — "Breakthroughs in the social and behavioral sciences have been enabled by and increasingly depend on advances in cybertechnologies. The impact of social and behavioral research has proven to be substantial – leading, for example, to new insights into core national challenges such as the creation of wealth and jobs, as well as to fundamental understanding of behavioral phenomena at scales ranging from the molecular to the social. The special technical needs of the social and behavioral scientists have also posed numerous research challenges to cyberinfrastructure experts and contributed to substantial advances in the cybertechnologies of other sciences: for example, geographic information systems applications used in engineering, the geosciences and biology, and statistical software packages developed for social and behavioral science applications but now widely used across the health sciences. Yet, the range of opportunities and the full benefits that can be derived from the application and use of large-scale scientific infrastructure, currently called "cyberinfrastructure" at NSF, have yet to be fully explored and realized by social and behavioral scientists."

Awards — (unknown)

Science of Science and Innovation Policy

Synopsis — The Science of Science & Innovation Policy (SciSIP) program supports research designed to advance the scientific basis of science and innovation policy. Research funded by the program thus develops, improves and expands models, analytical tools, data and metrics that can be applied in the science policy decision making process. For example, research proposals may develop behavioral and analytical conceptualizations, frameworks or models that have applications across a broad array of SciSIP challenges, including the relationship between broader participation and innovation or creativity. Proposals may also develop methodologies to analyze science and technology data, and to convey the information to a variety of audiences. Researchers are also encouraged to create or improve science and engineering data, metrics and indicators reflecting current discovery, particularly proposals that demonstrate the viability of collecting and analyzing data on knowledge generation and innovation in organizations.

2006-2009 Awards — \$30.8 million

Human and Social Dynamics (NSF 08-508)

Synopsis — The Human and Social Dynamics (HSD) priority area fosters breakthroughs in understanding the dynamics of human action and development, as well as knowledge about organizational, cultural, and societal adaptation and change. HSD aims to increase our collective ability to (1) understand the complexities of change;

(2) understand the dynamics of human and social behavior at all levels, including that of the human mind; (3) understand the cognitive and social structures that create, define, and result from change; and (4) manage profound or rapid change, and make decisions in the face of changing risks and uncertainty. Accomplishing these goals requires multidisciplinary research teams and comprehensive, interdisciplinary approaches across the sciences, engineering, education, and humanities, as appropriate.

Relevant text — "...the convergence of biology, engineering, geosciences, the cognitive and social sciences, mathematical and physical sciences, and computer and information sciences allows the development of basic knowledge about the evolution of individual, group, and organizational behavior with a rigor never before possible. At the same time, the convergence offers scientists, engineers, and educators more realistic models with which to test and transform organizational and classroom practice. Examination of the dynamic interactions of science, engineering, and technology with persons and social institutions requires integrative research efforts with broad and deep theoretical and methodological sophistication. Geographic information systems and related technologies, together with advances in the multilevel modeling of complex systems and network analysis from engineering, the geosciences, and the mathematical and physical sciences, open new frontiers for understanding such diverse subjects as crime, environmental management, epidemics, linguistic behaviors, societal innovation and adaptation, appropriations of new technologies, and the interplay of biological and social aspects of behavior within our species."

This program has ended, but similar programs may be created in the future.

2005-2008 Awards — \$65 million

Accelerating Discovery in Science and Engineering Through Petascale Simulations and Analysis (Petaapps) (NSF 08-592)

Synopsis — This solicitation seeks proposals to develop the future simulation and analysis tools that can use petascale computing to advance the frontiers of scientific and engineering research. Proposals are sought from researchers aiming to capitalize on emerging petascale computing architectures, catalyzing potentially transformative research. NSF's emphasis is on implementation and exploitation of forefront techniques. Proposers must be prepared to demonstrate that they have a research problem that requires and can exploit petascale computing capabilities. Proposals from or including junior researchers are encouraged as one of the goals of this solicitation is to build a community capable of using petascale computing.

Relevant text — "New modeling techniques enable researchers to take a more comprehensive view of physical social and behavioral phenomena stretching across disparate length and time scales, including the capability to match details accurately at the scale boundaries."

2008-2009 Awards — \$27.8 million

Explosives and Related Threats: Frontiers in Prediction and Detection (NSF 07-528)

Synopsis — In FY 2007, NSF will invest in leading edge, frontier research on sensors and other areas, including social and behavioral sciences, that are potentially relevant to the prediction and detection of explosives and related threats. This is an NSF-wide effort, in coordination with the efforts of other agencies, which seeks to advance fundamental knowledge in new technologies for sensors and sensor networks, and in the use of sensor data in control and decision making, particularly in relation to the prediction and detection of explosives and related threats. This research is seen as critical to our nation's ability to deploy effective homeland security measures, and to protect civilians and our military forces throughout the world.

Relevant text — "Research toward this goal might include mathematical, behavioral, and cognitive modeling, including algorithms and sensor systems that may predict the possible assembly and placement of explosive devices. This may involve the recognition of emplacement patterns, behavioral pattern recognition from video and other innovative sensing systems, human intelligence and social network analysis of terrorist networks, analysis of communications, and knowledge-management systems."

2007-2008 Awards — \$19.8 million

Defense Advanced Research Projects Agency (DARPA)

Conflict Modeling, Planning, and Outcomes Experimentation (Current Program)

Synopsis — DARPA's Conflict Modeling, Planning, and Outcomes Experimentation (COMPOEX) program is developing a suite of tools that will help military commanders and their civilian counterparts to plan, analyze and conduct complex campaigns. "Complex" here refers to those operations - often of long duration and large scale - that require careful consideration of not only traditional military, but also political, social, economic actions and ramifications.

2007-2009 Funding — \$16.1 million

Integrated Crisis Early Warning System (BAA 07-10)

Synopsis — The Integrated Crisis Early Warning System (ICEWS) program seeks to develop a comprehensive, integrated, automated, generalizable, and validated system to monitor, assess, and forecast national, sub-national, and international crises in a way that supports decisions on how to allocate resources to mitigate them. ICEWS will provide Combatant Commanders (COCOMs) with a powerful, systematic capability to anticipate and respond to stability challenges in the Area of Responsibility (AOR); allocate resources efficiently in accordance to the risks they are designed to mitigate; and track and measure the effectiveness of resource allocations toward end-state stability objectives, in near-real time.

2007-2009 Funding — \$30.4 million

Technologies for the Applications of Social Computing (TASC) (RFI DARPA-SN-09-20)

Synopsis — Information Processing Techniques Office (IPTO), invites white paper responses from all qualified vendors interested in exploring the development of new technologies to rapidly create theoretically-informed, data-driven models of complex human, social, cultural, and behavioral dynamics that are instantiated in near-real-time simulations. These technologies would leverage the entire social science community and provide a rich test bed for establishing the empirical validity of alternative theories, and identifying gaps in knowledge that cannot be accounted for by the current body of social science theory. Other important technologies of interest include the formalization and semantic representation of social science theories, the semantic integration of disparate types of social science data, techniques for analyzing these data, and efficient computational techniques for rapid data processing. DARPA refers to this range of technologies as "Technologies for the Applications of Social Computing (TASC)." DARPA anticipates all these technologies would be integrated to develop a flexible, modular social simulation system that integrates sound social science theory with real world data, that facilitates a wide spectrum of military and intelligence applications, and that supports reliable, real-world decisions at multiple levels of analysis.

Funding — (not yet known)

Office of the Secretary of Defense (OSD)

Human, Social, Cultural & Behavioral (HSCB) – Decision Support Tools Technology Focus Area (SBIR FY2009.2)

Synopsis — Current military operations need and future operations will demand the capability to understand the social and cultural terrain and the various dimensions of human behavior within those terrains. Behaviors in the social and cultural terrain context extend across the spectrum, from adversaries to our joint U.S. forces, with our coalition partners, and with government and non-government organizations. For operational, strategic and tactical warfighters, there is a significant need for socio-cultural models that provide predictive capabilities with regard to the behavior of adversaries and contested populations.

Office of Naval Research (ONR)

Human, Social, Cultural and Behavioral Sciences (ONR BAA 08-025)

Synopsis — The overall goal of this solicitation is to foster new developments in Science and Technology which may ultimately lead to future operational capabilities beyond those represented by current acquisition programs and requirements. As such, it is anticipated that successful proposals would ultimately contribute to the scientific and technological underpinning from which future Naval Expeditionary and Combating Terrorism warfighting requirements and capabilities may become possible.

Sciences Addressing Asymmetric Explosive Threats (ONR BAA 09-022)

Synopsis — The Office of Naval Research (ONR) is interested in receiving proposals describing innovative scientific concepts that will form the foundation for future technologies that may be developed and implemented to efficiently and effectively address future asymmetric threats, resulting in explosive events. This BAA is part of a basic research program, and as such the knowledge and technologies developed here may lead to future applied research; however, research under this program is limited to fundamentally new theories, data, concepts, and principles.

Relevant text — This topic seeks a better understanding of, or the ability to predict, human networks that carry out asymmetric activities without an understanding of human resources, motivations, logistics capabilities, environment and resources available to the network. Research is needed into how to determine what the relevant parameters are that enable human networks to perform asymmetric acts. This includes research to develop techniques to effectively aggregate and model disparate data sources and types. Anticipated deliverables would be computational models and techniques and methods.

Air Force Office of Scientific Research (AFOSR)

Research Interests of the Air Force Office of Scientific Research (AFOSR BAA 2009-1)

Listed topics include:

Collective Behavior and Socio-Cultural Modeling — We are interested in developing a basic research foundation for using computational and modeling approaches to study behavior of group and communities. This program seeks fundamental understanding of the interactions between demographic groups both to create understanding for technology developments for enhanced cooperation, such as operational decision making with coalition partners, and to explain and predict outcomes between competing factions within geographic regions. This program encourages collaboration between social, behavioral, cognitive, and biological scientists with computational researchers in disciplines such as mathematics, computer science, modeling, artificial intelligence, control theory, and adaptive systems.

Socio-Cultural Modeling of Effective Influence — We are interested in developing a basic research foundation for incorporating an understanding of factors underlying socio- cultural/population variability into effects based

operations suitable for application in a variety of domains. Areas of interest include the modeling of the effects of population variables in groups and communities to situations that might occur in military and related situations. Such situations might include weapons effects, the behavior of individuals and groups, including friendly forces and populations, to non-lethal weapons, culturally conditioned responses to natural cataclysmic events and other disasters, cyber-related effects, etc...Innovative approaches to data collection and analysis in this domain are needed such as adaptation of anthropology, bioinformatics, computer science, dynamic systems, economics, epidemiology, international relations, marketing, mathematics, psychology, political science, sociology, operations research, etc.

Intelligence Advanced Research Projects Activity (IARPA)

Reynard (IARPA BAA 09-05)

Synopsis — Events in today’s world move at an accelerating pace, and are often shaped by information (whether correct or incorrect) that is communicated by diverse and increasingly novel means. A nuanced picture of events and attitudes is required to inform policy developers and other decision-makers in this complex and fast-paced environment. IARPA sponsors leading-edge research to help the Intelligence Community keep abreast of the many sources of information needed to inform our nation’s leadership. A new channel for information exchange and social interaction is emerging with the growing popularity of Virtual Worlds (VWs) and Massive Multiplayer Online Games (MMOGs). Industry experts predict the number of VW users will exceed a half billion by the end of 2009. As participation in these VWs broadens and deepens, the need for understanding the nature of interactions in these worlds arises. Starting from the premise that Real World (RW) characteristics are reflected in Virtual World (VW) behavior, the Reynard program seeks to determine behavioral indicators in the realm of Massive Multiplayer Online Games (MMOGs) and VWs that are predictive of RW characteristics of the users.

Computational Cognitive Models of Sensemaking Program (IARPA RFI 09-02)

Synopsis — Intelligence analysts are tasked with identifying relevant information within massive data and synthesizing these fragments into a coherent understanding of the entities, events, and relational networks that characterize a data space. This process of adaptively filtering and interpreting data has been termed "sensemaking" in the literature. However, prevailing models of sensemaking are largely qualitative in nature and fail to reflect emerging insights from cognitive science and neuroscience research regarding the architecture of human information processing. Current models thus provide a shaky foundation upon which to build a deep understanding of analyst cognition. IARPA is soliciting submissions on either one or both of the following topics aimed at addressing these challenges: 1) Cognitive models of sensemaking...2) Test framework for evaluating sensemaking...

National Institutes of Health (NIH)

Using Systems Science Methodologies to Protect and Improve Population Health (PAR-08-224)

Synopsis — This FOA solicits Exploratory/Developmental (R21) applications from institutions/organizations that propose to apply one or more specific system science methodologies to “policy resistant” public health problems and contribute knowledge that will enhance effective decision making around the development of and prioritization of policies, interventions, and programs to improve population health in the U.S. and abroad, especially where resources are limited and only a limited number of programs/policies/interventions can be implemented. Systems science methodologies, discussed in more detail below, are specific methodologies that have been developed to understand connections between a system’s structure and its behavior over time; these

methods are often capable of illuminating a range of “policy resistant problems” where interventions, programs, or policies typically have unexpected, unintended, or counter-intuitive consequences. A system, in this context, refers to the particular configuration of all relevant entities, resources, and processes that together adequately characterize the problem space under study (i.e., a system is defined the boundaries that stakeholders use to determine which facts/observations are relevant for their inquiry as well as the interpretations/judgments that they use to guide decisions or actions) (Ulrich, 2002).

Relevant text — "Specific examples of systems science methodologies include, but are not limited to: systems dynamics modeling (Sterman, 2000), agent based modeling (Epstein, 2006), discrete event simulation (Banks et al., 2005), network analysis (Wasserman & Faust, 1994; Scott, 2000), dynamic microsimulation modeling (e.g., Mitton, Sutherland & Weeks, 2000), and Markov modeling (Sonnenberg & Beck, 1993)."

Methodology and Measurement in the Behavioral and Social Sciences (PAR-08-212)

Synopsis — The behavioral and social sciences offer significant, fundamental insights into the comprehensive understanding of human health, including disease etiology and treatment, and the promotion of health and well-being. To encourage the investigation of the impact of social and behavioral factors on health and disease, the participating Institutes and Centers (ICs) invite qualified researchers to submit research grant applications on methodology and measurement in the behavioral and social sciences. Methodology and measurement encompass research design, data collection techniques, measurement, and data analysis techniques. The goal of this program announcement is to encourage research that will improve the quality and scientific power of data collected in the behavioral and social sciences, using humans or animals, relevant to the missions of the NIH ICs. Research that addresses methodology and measurement issues in diverse populations, issues in studying sensitive behaviors, issues of ethics in research, issues related to confidential data and the protection of research subjects, and issues in developing interdisciplinary, multimethod, and multilevel approaches to behavioral and social science research is particularly encouraged, as are approaches that integrate behavioral and social science research with biomedical, physical, or computational science research or engineering. Because the focus of this program announcement is developing methodology, rather than addressing a single, health-related research question, applicants are encouraged to propose approaches that would be broadly applicable to basic or applied behavioral and social sciences research related to health.

Appendix F: Funding Record of PIs

A partial list of funding activities for the PIs.

PI	Funder	Project Title	Amount	Start	End
A. McCallum, D. Smith	DARPA/SRI	A Universal Machine Reading System	\$2,572,052	May-09	May-14
A. McCallum	Google	Reseach Award	\$90,000	May-09	
J. Fountain	NSF	Developing a Beta Site for a National Digital Library for Ethics in Science and Engineering	\$249,999	May-09	Oct-10
J.Allan, W. B. Croft	NSF	Learning Word Relationships Using TupleFlow	\$450,000	Feb-09	Jan-11
M. Lavine	NSF	An Innovative Model for Workforce Development in Statistics	\$600,000	Jan-09	Jun-14
J. Allan	IBM	Open Collaboration Faculty Award	\$75,000	Dec-08	
W. B. Croft	Yahoo	Research Award	\$25,000	Dec-08	
J. Allan	IBM	UIMA Faculty Award	\$19,000	Nov-08	
A. McCallum	NSF	Dynamically-Structured Conditional Random Fields For Natural Language Processing	\$449,992	Aug-08	Jul-11
J. Allan	Microsoft	Research Award	\$5,000	Apr-08	
J. Allan	UpToDate	Research Award	\$77,710	Apr-08	
D. Jensen	IARPA	Causality by Design: Learning Causal Models of Social Networks.	\$150,000	Apr-08	Sep-08
D. Jensen	MIT Lincoln Lab	Relational Learning Methods and Knowledge Discovery for Social Networks and Intent Recognition.	\$83,700	Apr-08	Sep-08
W. B. Croft	NSF	Supporting User Data, Privacy, and Evaluation in the Lemur Toolkit	\$771,955	Sep-07	Aug-10
J. Fountain	NSF	International Dimensions of Ethics Education in Science and Engineering	\$299,233	Sep-07	Aug-10

PI	Funder	Project Title	Amount	Start	End
A. McCallum	AFRL/Lockheed	Information Integration Seedling	\$290,000	Sep-07	Apr-09
S. Shulman	NSF	Collaborative Research III-COR: From a Pile of Documents to a Collection of Information: A Framework for Multi-Dimensional Text Analysis	\$1,000,000	Aug-07	
W. B. Croft	NSF	Searching Archives of Community Knowledge	\$404,132	Aug-07	Jul-10
A. McCallum	IBM	Faculty Award	\$30,000	Jul-07	
D. Jensen	IARPA	Learning Compositional Simulation Models.	\$2,033,716	Jun-07	Dec-12
A. McCallum, C. Potts	Army/UPenn	Situation Understanding Bot through Language and Environment	\$1,074,997	May-07	04/31/12
D. Tomaskovic-Devey	NSF	Equal Employment Opportunity Progress in Private Sector Workplaces since 1966: Dynamic Models of Organizational Change	\$151,000	Apr-07	Apr-08
A. McCallum	DARPA/White Oak	KDD Entity Enhancement Study	\$19,655	Mar-07	Jul-07
J. Allan	Google	Research Award	\$65,000	Jan-07	
W. B. Croft	Google	Research Award	\$55,000	Jan-07	
D. Jensen	Comcast	Relational learning for Media Indexing	\$110,000	Jan-07	Nov-07
R. Manmatha, J. Allan, et	NSF	Cluster Acquisition for Computational Research into Large Scale Data Rich Problems	\$350,000	Sep-06	Aug-08
A. McCallum	DARPA/NGA	Resource-Bounded Information Gathering for Efficient Entity Resolution and Link Analysis using Probabilistic Reasoning and Discovery	\$239,579	Aug-06	Aug-08
W. B. Croft	IBM	Eclipse Faculty Award	\$22,000	Aug-06	
S. Shulman	NSF	Coding Across the Disciplines: A Project-Based Workshop on Manual Text Annotation Techniques	\$25,000	Jul-06	
W. B. Croft	Microsoft	Research Award	\$42,000	Jun-06	

PI	Funder	Project Title	Amount	Start	End
D. Jensen	NASD	Detecting Securities Fraud with Network Analysis	\$150,000	Jun-06	Sep-06
W. B. Croft	NSF	Text Reuse and Information Flow	\$398,000	Apr-06	Mar-10
A. McCallum	NSF	Improving Experimental Computer Science with a Searchable Web Portal for Data Sets	\$200,000	Mar-06	Feb-10
J. Allan, W. B. Croft, A. M	DARPA/SRI	Supporting Distillation for GALE	\$2,304,519	Sep-05	Jun-09
J. Fountain	NSF	Developing a Basic Research Program for Digital Government	\$26,766	Sep-05	Aug-06
J. Fountain	NSF	A National Center for Digital Government: Integrating Information and Institutions	\$348,021	Sep-05	Sep-08
D. Tomaskovic-Devey	Russell Sage Four	Documenting Desegregation: Equal Opportunity in Private Sector Employment Since the Civil Rights Act of 1964	\$35,000	Sep-05	Aug-07
A. McCallum	DARPA/IBM	The KDD Evaluation Problems: Conditional Random Field, Topic Models and Relational Methods	\$67,622	Jul-05	Dec-05
A. McCallum	NSF/ITIC	Aiding Collaboration through Probabilistic Methods on Text and Semi-Structured Data	\$920,700	Jul-05	Aug-10
W. B. Croft	NSF/CMU	Developing the Lemur Toolkit into a Community Resource	\$392,538	Jul-05	Jun-08
J. Allan R. Manmatha, A.	DARPA	Shared Computing Resource for Language and Multimedia Technologies	\$117,321	May-05	Apr-06
W. B. Croft	NSF	Breaking the Keyword Bottleneck: Towards More Effective Access of Government Information	\$99,903	Apr-05	Mar-06
D. Jensen	LLNL	Relational Pathfinding" Lawrence Livermore National Laboratory	\$125,000	Jan-05	Apr-06
A. McCallum	NSF	Machine Learning for Sequences and Structured Data: Tools for Non-Experts	\$332,999	Sep-04	Aug-08
A. McCallum	DoI/BBNT	Statistical Models for Information Extraction for REFLEX	\$250,603	Sep-04	Aug-07
S. Shulman	NSF	Collaborative Research: Language Processing Technology for Electronic Rulemaking	\$1,225,000	Aug-04	

PI	Funder	Project Title	Amount	Start	End
A. McCallum	AFOSR/Aptima	Interactive Text Classification of Usability Reports	\$105,000	Jun-04	May-06
A. McCallum, D. Jensen	NSF	Unified Graphical Models of Information Extraction and Data Mining with Application to Social Network Analysis	\$2,944,718	Sep-03	Aug-10
A. McCallum	AFOSR/Aptima	Text Classification of Usability Reports	\$29,988	Sep-03	May-04
W. B. Croft, J. Allan	NSF	The Analytic Web	\$712,144	Sep-03	Aug-07
S. Shulman	NSF	Democracy and E-Rulemaking: Comparing Traditional vs. Electronic Comment from a Discursive Democratic Framework.	\$260,000	Jul-03	
A. McCallum, W. B. Croft	DARPA/SRI	CALO: Cognitive Assistant that Learns and Observes	\$2,189,197	May-03	Mar-09
S. Shulman	NSF	SGER COLLABORATIVE: A Testbed for eRulemaking Data.	\$210,000	Apr-03	
D. Jensen	DARPA	Collective Inference with Learned and Engineered Knowledge.	\$1,500,000	Feb-03	May-06
J. Allan, W.B. Croft	NSF/UCAR	Leading NSDL toward Long-Term Success	\$523,086	Oct-02	Sep-04
W. B. Croft	NSF	Question Triage for Experts and Documents: Expanding the Information Retrieval Function of the NSDL	\$502,309	Sep-02	Feb-05
D. Tomaskovic-Devey	NSF	Establishment Change in Gender and Race/Ethnic Segregation and Access to Managerial and Craft Occupations, 1966-1999	\$91,065	Jun-02	Jun-04
J. Fountain	NSF	A National Center for Digital Government: Integrating Information and Institutions	\$1,695,598	May-02	Aug-06
W.B. Croft, J. Allan	DARPA	Formal Frame works and Empirical Evaluations for Information Organization	\$2,650,389	Apr-02	Apr-05
D. Jensen	NSF	Computational Statistics for Intelligence Analysis.	\$650,000	Mar-02	Jun-04
J. Fountain	NSF	Developing a Basic Research Program for Digital Government	\$99,936	Feb-02	Jul-06
J. Fountain	PriceWaterhouseC	Women in the Information Age	\$100,000	Jan-02	

PI	Funder	Project Title	Amount	Start	End
S. Shulman	NSF	ITR/PE:Digital Citizenship: Expanding Information Technology Literacy with a Service-Learning Approach.	\$440,000	Sep-01	
D. Jensen	DARPA	Transformative Pattern Learning.	\$1,570,357	Sep-01	Dec-05
D. Tomaskovic-Devey	NSF	Racial and Spatial Influences on the Labor Market Opportunities of Low Skilled Workers	\$136,054	Jul-01	Jun-03
W.B. Croft, J. Allan	NSF	Supporting Effective Access through User- and Topic-Based Language Models	\$351,964	Sep-00	Sep-04
S. Shulman	NSF	Workshop: YouTube and the 2008 Election Cycle in the United States, April 3-4, 2009	\$75,000	Aug-00	
S. Shulman	NSF	Digital Government: SGER: Citizen Agenda-Setting in the Regulatory Process: Electronic Collection and Synthesis of Public Commentary.	\$33,498	Aug-00	
W. B. Croft	NSF	Digital Government: A Language Modeling Approach for Cross-Database Linkage and Search	\$455,484	Jul-00	Jun-04
D. Jensen	DARPA/AFRL	Analytical Tools for Investigating and Modeling Agent-Based Systems.	\$1,170,000	Jun-00	Dec-05
D. Hochman, M. Lavine	NIH	Neocortical hemodynamics during epileptic activity in primates and humans	\$1,948,000	-TBD-	