Computational Social Science

biogeographic patterns. Their study, too, is centered on a large database, but in this case it is entirely of living organisms, the marine bivalves. Over 28,000 records of bivalve genera and subgenera from 322 locations around the world have now been compiled by these authors, giving a global record of some 854 genera and subgenera and 5132 species. No fossils are included in the database, but because bivalves have a good fossil record, it is possible to estimate accurately the age of origin of almost all extant genera. It is then possible to plot a backward survivorship curve (*8*) for each of the 27 global bivalve provinces (*9*).

On the basis of these curves, Krug *et al.* find that origination rates of marine bivalves increased significantly almost everywhere immediately after the K-Pg mass extinction event. The highest K-Pg origination rates all occurred in tropical and warm-temperate regions. A distinct pulse of bivalve diversification in the early Cenozoic was concentrated mainly in tropical and subtropical regions (see the figure).

The steepest part of the global backward survivorship curve for bivalves lies between 65 and 50 million years ago, pointing to a major biodiversification event in the Paleogene (65 to 23 million years ago) that is perhaps not yet captured in Alroy *et al.*'s database (*5*, *7*). The jury is still out on what may have caused this event. But we should not lose sight of the fact that the steep rise to prominence of many modern floral and faunal groups in the Cenozoic may bear no simple relationship to climate or any other type of environmental change (*10*, *11*).

References

1. G. G. Mittelbach *et al.*, *Ecol. Lett.* **10**, 315 (2007).
2. A. Z. Krug, D. Jablonski, J. W. Valentine, *Science* **323**, 767 (2009).
3. P. W. Signor, *Annu. Rev. Ecol. Syst.* **21**, 509 (1990).
4. R. K. Bambach, *Geobios* **32**, 131 (1999).
5. J. Alroy *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 6261 (2001).
6. A.M. Bush *et al.*, *Paleobiology* **30**, 666 (2004).
7. J. Alroy *et al.*, *Science* **321**, 97 (2008).
8. M. Foote, in *Evolutionary Patterns*, J. B. C. Jackson *et al.*, Eds. (Univ. of Chicago Press, Chicago, IL, 2001), vol. 245, pp. 245–295.
9. M. D. Spalding *et al.*, *Bioscience* **57**, 573 (2007).
10. S. M. Stanley, *Paleobiology* **33**, 1 (2007).
11. M. J. Benton, B. C. Emerson, *Palaeontology* **50**, 23 (2007).

10.1126/science.1169410

SOCIAL SCIENCE

Computational Social Science

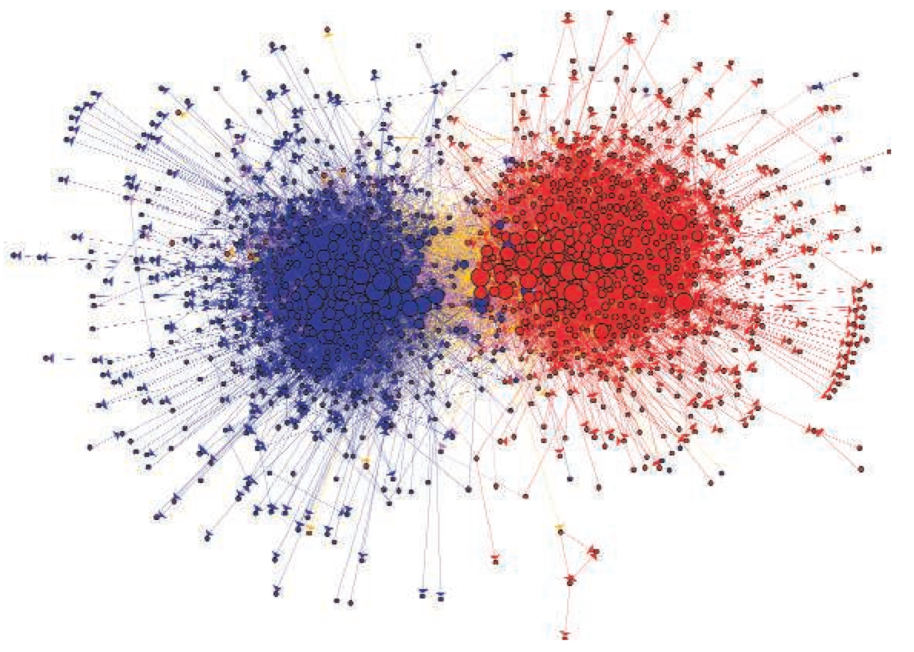
A field is emerging that leverages the capacity to collect and analyze data at a scale that may reveal patterns of individual and group behaviors.

David Lazer,[1] Alex Pentland,[2] Lada Adamic,[3] Sinan Aral,[2,4] Albert-László Barabási,[5] Devon Brewer,[6] Nicholas Christakis,[1] Noshir Contractor,[7] James Fowler,[8] Myron Gutmann,[3] Tony Jebara,[9] Gary King,[1] Michael Macy,[10] Deb Roy,[2] Marshall Van Alstyne[2,11]

We live life in the network. We check our e-mails regularly, make mobile phone calls from almost any location, swipe transit cards to use public transportation, and make purchases with credit cards. Our movements in public places may be captured by video cameras, and our medical records stored as digital files. We may post blog entries accessible to anyone, or maintain friendships through online social networks. Each of these transactions leaves digital traces that can be compiled into comprehensive pictures of both individual and group behavior, with the potential to transform our understanding of our lives, organizations, and societies.

The capacity to collect and analyze massive amounts of data has transformed such fields as biology and physics. But the emergence of a data-driven "computational social science" has been much slower. Leading journals in economics, sociology, and political science show little evidence of this field. But computational social science is occurring—in Internet companies such as Google and Yahoo, and in government agencies such as the U.S. National Security Agency. Computational social science could become the exclusive domain of private companies and government agencies. Alternatively, there might emerge a privileged set of academic researchers presiding over private data from which they produce papers that cannot be critiqued or replicated. Neither scenario will serve the long-term public interest of accumulating, verifying, and disseminating knowledge.

What value might a computational social science—based in an open academic environment—offer society, by enhancing understanding of individuals and collectives? What are the



Data from the blogosphere. Shown is a link structure within a community of political blogs (from 2004), where red nodes indicate conservative blogs, and blue liberal. Orange links go from liberal to conservative, and purple ones from conservative to liberal. The size of each blog reflects the number of other blogs that link to it. [Reproduced from (*8*) with permission from the Association for Computing Machinery]

[1]Harvard University, Cambridge, MA, USA. [2]Massachusetts Institute of Technology, Cambridge, MA, USA. [3]University of Michigan, Ann Arbor, MI, USA. [4]New York University, New York, NY, USA. [5]Northeastern University, Boston, MA, USA. [6]Interdisciplinary Scientific Research, Seattle, WA, USA. [7]Northwestern University, Evanston, IL, USA. [8]University of California–San Diego, La Jolla, CA, USA. [9]Columbia University, New York, NY, USA [10]Cornell University, Ithaca, NY, USA. [11]Boston University, Boston, MA, USA. E-mail: david_lazer@harvard.edu. Complete affiliations are listed in the supporting online material.

obstacles that prevent the emergence of a computational social science?

To date, research on human interactions has relied mainly on one-time, self-reported data on relationships. New technologies, such as video surveillance (*1*), e-mail, and "smart" name badges, offer a moment-by-moment picture of interactions over extended periods of time, providing information about both the structure and content of relationships. For example, group interactions could be examined through e-mail data, and questions about the temporal dynamics of human communications could be addressed: Do work groups reach a stasis with little change, or do they dramatically change over time (*2*)? What interaction patterns predict highly productive groups and individuals? Can the diversity of news and content we receive predict our power or performance (*3*)? Face-to-face group interactions could be assessed over time with "sociometers." Such electronic devices could be worn to capture physical proximity, location, movement, and other facets of individual behavior and collective interactions. The data could raise interesting questions about, for example, patterns of proximity and communication within an organization, and flow patterns associated with high individual and group performance (*4*).

We can also learn what a "macro" social network of society looks like (*5*), and how it evolves over time. Phone companies have records of call patterns among their customers extending over multiple years, and e-Commerce portals such as Google and Yahoo collect instant messaging data on global communication. Do these data paint a comprehensive picture of societal-level communication patterns? In what ways do these interactions affect economic productivity or public health? It is also increasingly easy to track the movements of people (*6*). Mobile phones allow the large-scale tracing of people's movements and physical proximities over time (*7*). Such data may provide useful epidemiological insights: How might a pathogen, such as influenza, driven by physical proximity, spread through a population?

The Internet offers an entirely different channel for understanding what people are saying, and how they are connecting (*8*). Consider, for example, this past political season, tracing the spread of arguments, rumors, or positions about political and other issues in the blogosphere (*9*), as well as the behavior of individuals "surfing" the Internet (*10*), where the concerns of an electorate become visible in the searches they conduct. Virtual worlds, which by their nature capture a complete record of individual behavior, offer ample opportunities for research—experimentation that would otherwise be impossible or unacceptable (*11*). Similarly, social network Web sites offer a unique opportunity to understand the impact of a person's position in the network on everything from their tastes to their moods to their health (*12*), whereas Natural Language Processing offers increased capacity to organize and analyze the vast amounts of text from the Internet and other sources (*13*).

In short, a computational social science is emerging that leverages the capacity to collect and analyze data with an unprecedented breadth and depth and scale. Substantial barriers, however, might limit progress. Existing ways of conceiving human behavior were developed without access to terabytes of data describing minute-by-minute interactions and locations of entire populations of individuals. For example, what does existing sociological network theory, built mostly on a foundation of one-time "snapshot" data, typically with only dozens of people, tell us about massively longitudinal data sets of millions of people, including location, financial transactions, and communications? These vast, emerging data sets on how people interact surely offer qualitatively new perspectives on collective human behavior, but our current paradigms may not be receptive.

There are also enormous institutional obstacles to advancing a computational social science. In terms of approach, the subjects of inquiry in physics and biology present different challenges to observation and intervention. Quarks and cells neither mind when we discover their secrets nor protest if we alter their environments during the discovery process. As for infrastructure, the leap from social science to a computational social science is larger than from biology to a computational biology, largely due to the requirements of distributed monitoring, permission seeking, and encryption. There are fewer resources available in the social sciences, and even the physical (and administrative) distance between social science departments and engineering or computer science departments tends to be greater than for the other sciences.

Perhaps the thorniest challenges exist on the data side, with respect to access and privacy. Much of these data are proprietary (e.g., mobile phone and financial transactional information). The debacle following AOL's public release of "anonymized" search records of many of its customers highlights the potential risk to individuals and corporations in the sharing of personal data by private companies (*14*). Robust models of collaboration and data sharing between industry and academia are needed to facilitate research and safeguard consumer privacy and provide liability protection for corporations. More generally, properly managing privacy issues is essential. As the recent U.S. National Research Council's report on geographical information system data highlights, it is often possible to pull individual profiles out of even carefully anonymized data (*15*). Last year, the U.S. National Institutes of Health and the Wellcome Trust abruptly removed a number of genetic databases from online access (*16*). These databases were seemingly anonymized, simply reporting the aggregate frequency of particular genetic markers. However, research revealed the potential for de-anonymization, based on the statistical power of the sheer quantity of data collected from each individual in the database (*17*).

Because a single dramatic incident involving a breach of privacy could produce rules and statutes that stifle the nascent field of computational social science, a self-regulatory regime of procedures, technologies, and rules is needed that reduces this risk but preserves research potential. As a cornerstone of such a self-regulatory regime, U.S. Institutional Review Boards (IRBs) must increase their technical knowledge to understand the potential for intrusion and individual harm because new possibilities do not fit their current paradigms for harm. Many IRBs would be poorly equipped to evaluate the possibility that complex data could be de-anonymized. Further, it may be necessary for IRBs to oversee the creation of a secure, centralized data infrastructure. Currently, existing data sets are scattered among many groups, with uneven skills and understanding of data security and widely varying protocols. Researchers themselves must develop technologies that protect privacy while preserving data essential for research. These systems, in turn, may prove useful for industry in managing customer privacy and data security (*18*).

Finally, the emergence of a computational social science shares with other nascent interdisciplinary fields (e.g., sustainability science) the need to develop a paradigm for training new scholars. Tenure committees and editorial boards need to understand and reward the effort to publish across disciplines. Initially, computational social science needs to be the work of teams of social and computer scientists. In the long run, the question will be whether academia should nurture computational social scientists, or teams of computationally literate social scientists and socially literate computer scientists. The emergence of cognitive science offers a powerful model for the development of a computational social science. Cognitive science has involved fields ranging from neurobiology to philosophy to computer science. It has attracted the investment of substantial resources to create a common field, and created enormous progress for public good in the last generation. We would argue that a computational social science has a similar potential, and is worthy of similar investments.

References and Notes

1. D. Roy *et al.*, "The Human Speech Project," Proceedings of the 28th Annual Conference of Cognitive Science Society, Vancouver, BC, Canada, 26 to 29 July 2009.
2. J. P. Eckmann *et al. Proc. Natl. Acad. Sci. U.S.A.* **101**, 14333 (2004).
3. S. Aral, M. Van Alstyne, "Network Structure & Information Advantage," Proceedings of the Academy of Management Conference, Philadelphia, PA, 3 to 8 August 2007.
4. A. Pentland, *Honest Signals: How They Shape Our World* (MIT Press, Cambridge, MA, 2008).
5. J.-P. Onnela *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**,7332 (2007).
6. T. Jebara, Y. Song, K. Thadani, "Spectral Clustering and Embedding with Hidden Markov Models," Proceedings of the European Conference on Machine Learning, Philadelphia, PA, 3 to 6 December 2007.
7. M. C. González *et al.*, *Nature* **453**, 779 (2008).
8. D. Watts, *Nature* **445**, 489 (2007).
9. L. Adamic, N. Glance, in Proceedings of the 3rd International Workshop on Link Discovery (LINKDD 2005), pp. 36–43; http://doi.acm.org/10.1145/1134271.1134277.
10. J. Teevan, *ACM Trans. Inform. Syst.* **26**, 1 (2008).
11. W. S. Bainbridge, *Science* **317**, 472 (2007).
12. K. Lewis *et al.*, *Social Networks* **30**, 330 (2008).
13. C. Cardie, J. Wilkerson, *J. Inf. Technol. Polit.* **5**, 1 (2008).
14. M. Barbarao, T. Zeller Jr., "A face is exposed for AOL searcher No. 4417749," *New York Times*, 9 August 2006, p. A1.
15. National Research Council, *Putting People on the Map: Protecting Confidentiality with Linked Social-Spatial Data*, M. P. Gutmann, P. Stern, Eds. (National Academy Press, Washington, DC, 2007).
16. J. Felch. "DNA databases blocked from the public," *Los Angeles Times*, 29 August 2008, p. A31.
17. N. Homer, S. Szelinger, M. Redman, D. Duggan, W. Tembe, *PLoS Genet.* **4**, e1000167 (2008).
18. M.V.A. has applied for a patent on an algorithm for protecting privacy of communication content.
19. Additional resources in computational social science can be found in the supporting online material.

Supporting Online Material
www.sciencemag.org/cgi/content/full/323/5915/721/DC1

10.1126/science.1167742

CELL BIOLOGY

Moonlighting in Mitochondria

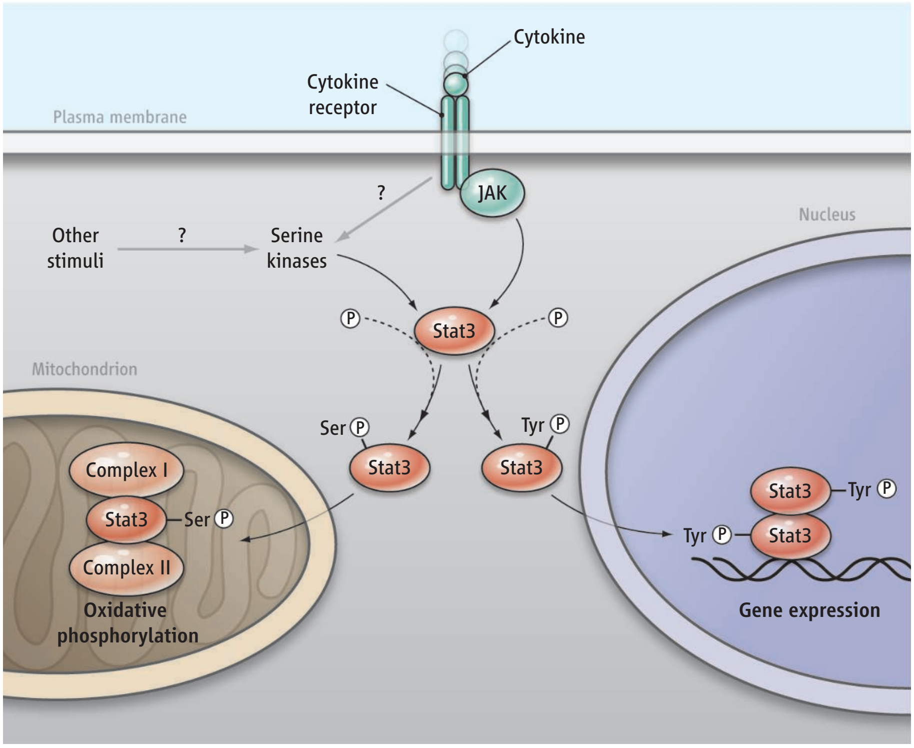
A cellular signaling pathway that responds to cytokines may coordinately control energy production by mitochondria.

Martin G. Myers Jr.

Molecules known as signal transducers and activators of transcription (STATs) regulate gene expression in the nucleus in response to cell surface receptors that are activated by cytokines. On page 793 of this issue, Wegrzyn *et al*. (*1*) reveal that the isoform Stat3 also functions in another organelle—the mitochondria—to control cell respiration and metabolism. This finding not only reveals a new role for Stat3, but implies its potential role in linking cellular signaling pathways to energy production.

Stat3 proteins represent the canonical mediators of signals elicited by type I cytokine receptors at the cell surface (*2*). For instance, the adipocytokine leptin activates Stat3 in hypothalamic neurons to promote the expression of the catabolic neuropeptide, propiomelanocortin, thereby regulating whole-body energy intake and metabolism (*3*). The binding of a cytokine to its receptor triggers an intracellular cascade of events, beginning with the activation of an enzyme, Jak kinase, which is associated with the receptor's cytoplasmic domain. The activated receptor-Jak complex then recruits and phosphorylates a tyrosine residue in cognate STAT proteins. This modification causes the STAT protein to relocate to the nucleus, where, as a dimer, it binds to specific DNA sequences and promotes gene expression (see the figure). Thus, the well-understood job of STAT proteins is to transmit a transcriptional signal from the cell surface to the nucleus. The phosphorylation of some STAT proteins on a specific serine residue may also contribute to their regulation (*2*).

Wegrzyn *et al*. have now identified another crucial role for Stat3, the isoform that responds to cytokines of the interleukin-6 and -10 families (including leptin). These cytokines act in the immune system and many other organ systems to regulate diverse cellular processes, including differentiation, proliferation, and apoptosis (*2*). Noting that GRIM-19, a mitochondrial protein, interacts with Stat3 and inhibits Stat3 transcriptional activity (*4*–*7*), the authors investigated the potential mitochondrial location of Stat3, revealing that a fraction of cellular Stat3 resides within the mitochondria of mouse myocytes and hepatocytes. Here, Stat3 associates with GRIM-19–containing



Dual deployment. The activation of a cytokine receptor at the cell surface promotes the tyrosine phosphorylation (Tyr-P) of Stat3, which dimerizes and moves to the nucleus to control gene expression. Serine phosphorylation (Ser-P) of Stat3 appears to be required for its action in mitochondria, where it promotes increased oxidative phosphorylation. Because many stimuli promote the serine phosphorylation of Stat3, many signaling pathways could regulate mitochondrial respiration via Stat3.

CREDIT: C. BICKEL/SCIENCE

Division of Metabolism, Endocrinology and Diabetes, Department of Internal Medicine, and Department of Molecular and Integrative Physiology, University of Michigan, Ann Arbor, MI 48109, USA. E-mail: mgmyers@umich.edu

www.sciencemag.org/cgi/content/full/323/5915/721/DC1

Supporting Online Material for

Computational Social Science

David Lazer,* Alex Pentland, Lada Adamic, Sinan Aral, Albert-László Barabási, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann, Tony Jebara, Gary King, Michael Macy, Deb Roy, Marshall Van Alstyne

*E-mail: david_lazer@harvard.edu

Published 6 February 2009, *Science* **323**, 721 (2009)
DOI: 10.1126/science.1167742

This PDF file includes:

Complete list of author affiliations
References
Online resources

Supplementary Online Materials

"Computational Social Science" *Science* **323**: 721-723 (6 February 2009)

Contact Information

David Lazer
Director, Program on Networked Governance
Associate Professor of Public Policy
Harvard Kennedy School
Harvard University
Cambridge, MA 02138
david_lazer@harvard.edu
http://www.davidlazer.com

Alex Pentland
Professor
MIT Media Laboratory
20 Ames Street
Cambridge, MA 02142, USA
pentland@media.mit.edu
http://web.media.mit.edu/~sandy/

Lada Adamic
Assistant Professor
School of Information and Center for the Study of Complex Systems
University of Michigan
1085 S. University Ave.
Ann Arbor, MI 48109
ladamic@umich.edu
http://www-personal.umich.edu/~ladamic/

Sinan Aral
Assistant Professor
NYU Stern School of Business
44 West 4 Street Room 8-81
New York, NY, 10012
Research Affiliate
MIT Sloan School of Management.
saral@nyu.edu
Personal Webpage: http://pages.stern.nyu.edu/~saral
SSRN Page: http://ssrn.com/author=110270

Albert-László Barabási
Professor

Center for Complex Network Research
Departments of Physics, Biology and Computer Science
Northeastern University, Boston, MA 02115
Department of Medicine
Harvard Medical School
Center for Cancer Systems Biology
Dana Farber Cancer Institute, Boston, MA 02115
alb@neu.edu
http://www.barabasi.com/

Devon D. Brewer
Interdisciplinary Scientific Research
P.O. Box 15110
Seattle, WA 98115
interscientific@yahoo.com
http://www.interscientific.net

Nicholas A. Christakis
Professor
Department of Health Care Policy, Harvard Medical School
Department of Sociology, Harvard University
180 Longwood Avenue
Boston, MA 02115-5899
christak@hcp.med.harvard.edu
http://christakis.med.harvard.edu/

Noshir Contractor
Jane S. & William J. White Professor of Behavioral Sciences, Northwestern University
Professor of Industrial Engineering & Management Science, McCormick School of Engineering
Professor of Communication Studies, School of Communication
Professor of Management & Organizations, Kellogg School of Management
Director, Science of Networks in Communities (SONIC)
Research Laboratory Affiliate, National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign
2145 Sheridan Road, TECH D241
Evanston, Illinois 60208-3119
nosh@northwestern.edu
http://nosh.northwestern.edu

James H. Fowler
Associate Professor
Political Science Department
University of California, San Diego
Social Science Building 383
9500 Gilman Drive #0521
La Jolla, CA 92093-0521

jhfowler@ucsd.edu
http://jhfowler.ucsd.edu

Myron P. Gutmann
Professor of History and Information
Director, Inter-university Consortium for Political and Social Research
University of Michigan
PO Box 1248, Ann Arbor, MI 48106-1248
gutmann@umich.edu
http://www.icpsr.org/ICPSR/staff/gutmann.html

Tony Jebara
Associate Professor
Department of Computer Science
Columbia University
1214 Amsterdam Avenue, Mail Code 0401
New York, NY 10027
jebara@cs.columbia.edu
http://www1.cs.columbia.edu/~jebara/

Gary King
Professor
Director, Institute of Quantitative Social Science
Harvard University
Cambridge, MA 02138
king@harvard.edu
http://gking.harvard.edu/

Michael Macy
Professor
Social Dynamics Laboratory
372 Uris Hall
Cornell University
Ithaca, New York
14853-7601 Cornell University
Mwm14@cornell.edu
http://www.people.cornell.edu/pages/mwm14/

Deb Roy
Associate Professor of Media Arts and Sciences
MIT Media Laboratory
20 Ames Street
Cambridge, MA 02142, USA
dkroy@media.mit.edu
http://web.media.mit.edu/~dkroy/

Marshall Van Alstyne
Associate Professor
Boston University School of Management
595 Commonwealth Avenue
Boston, MA 02215
Research Associate
MIT Center for E-Business
mva@bu.edu
http://smgnet.bu.edu/mgmt_new/profiles/VanAlstyneMarshall.html

Acknowledgments
This paper is an outgrowth of a December 7, 2007 conference by the same title. Thanks go to the Institute for Quantitative Social Science and the Program on Networked Governance at Harvard University (the latter supported by National Science Foundation grant # 0131923), and the MIT Legatum Center for sponsorship of that conference. Video of this conference is available at http://www.hks.harvard.edu/netgov/html/computational_social_science.html.

S. Aral acknowledges generous funding support from the Marketing Sciences Institute and the Institute for Innovation and Information Productivity. A.-L.B. is supported by James S. McDonnell Foundation 21st Century Initiative in Studying Complex Systems, the National Science Foundation within the DDDAS (CNS-0540348), ITR (DMR-0426737) and IIS-0513650 programs, the US Office of Naval Research Award N00014-07-C, The Office of Naval Research (ONR), Grant: N00014-07-C-0227 and by the Defense Threat Reduction Agency.

The viewpoints expressed in this article are the authors' alone.

Additional References

Adamic, L.A., Adar, E. (2005) "How to search a social network." *Social Networks* 27 (3):187-203

Adamic, L.A., Buyukkokten, O., Adar, E. (2003) "A Social Network Caught in the Web." *First Monday* 8 (6)

Adamic, L.A., Zhang, J., Bakshy, E., Ackerman, M.S. (2008) "Knowledge sharing and yahoo answers: everyone knows something." *Proceeding of the 17th international Conference on World Wide Web* (Beijing, China, April 21 - 25, 2008). WWW '08. ACM, New York, NY, 665-674. DOI= http://doi.acm.org/10.1145/1367497.1367587

Aral, S., Brynjolfsson, E., Van Alstyne, M. (2006) "Information, Technology & Information Worker Productivity: Task Level Evidence." *Proceedings of the 27th Annual International Conference on Information Systems*, December 10-13, Milwaukee, Wisconsin. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=942310

Aral, S., Brynjolfsson, E., Van Alstyne, M. (2007) "Productivity Effects of Information Diffusion in Networks." *Proceedings of the 28th Annual International Conference on Information Systems*, December 9-12, Montreal, CA. http://ssrn.com/abstract=987499

Aral, S., M. Van Alstyne (2007) "Network Structure & Information Advantage." *Proceedings of the Academy of Management Conference*, August 3-8, Philadelphia, PA. http://ssrn.com/abstract=958158

Aral, S., Muchnik, L., Sundararajan, A. (2008). "Economic Influence in Massive Online Social Networks." Workshop on Information Systems Economics, December 13-14, Paris, France

Bulkley, N., Van Alstyne, M. (2004) "Why Information Should Influence Productivity." in *The Network Society: A Global Perspective*. Ed. Castells, M. Edward Elgar Publishers, 145-173. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=518242

Burke, D.S., Epstein, J.M., Cummings, D.A.T., Parker, J.I., Cline, K.C., Singa, R., Chakravarty, S. "Individual-based Computational Modeling of Smallpox Epidemic Control Strategies." *Academic Emergency Medicine* 13 (11): 1142-1149 (2006)

Christakis, N.A., Fowler, J.H. (2007) "The Spread of Obesity in a Large Social Network Over 32 Years." *New England Journal of Medicine* 357 (4): 370-379 (26 July) http://dx.doi.org/10.1056/NEJMsa066082

Christakis, N.A., Fowler, J.H. (2008). "The Collective Dynamics of Smoking in a Large Social Network." *New England Journal of Medicine* 358 (21): 2249-58 (22 May) http://dx.doi.org/10.1056/NEJMsa0706154

Colizza, V., Barrat, A., Barthelemy, M., Vespignani, A. (2006) "Prediction and predictability of global epidemics: the role of the airline transportation network." *Proceedings of the National Academy of Sciences* 103: 2015-2020. http://alexves.googlepages.com/0510525103v1.pdf

Contractor, N., Wasserman, S., Faust, K. (2006). "Testing multi-theoretical multilevel hypotheses about organizational networks: An analytic framework and empirical example." *Academy of Management Review* 31: 681-703

Doyle, P., Lane, J.I., Theeuwes, J.J.M., Zayatz, L.V., eds. (2001) *Confidentiality, disclosure, and data access: Theory and practical applications for statistical agencies*. New York: Elsevier

Eagle, N. (2005), "Machine Perception and Learning of Complex Social Systems." Ph.D. Thesis, Program in Media Arts and Sciences, Massachusetts Institute of Technology, (June)

Eagle, N., Pentland, A. (2006) "Reality Mining: Sensing Complex Social Systems." *Personal and Ubiquitous Computing* 10 (4): 255-268

Epstein, J.M., Goedecke, M., Yu, F., Morris, J., Wagener, D., Bobashev, G. (2007) "Controlling Pandemic Influenza: The Role of International Travel Restrictions." *Public Library of Science One*. (May 12)

Fowler, J.H. (2006) "Legislative Cosponsorship Networks in the U.S. House and Senate." *Social Networks* 28 (4): 454-465 (October) http://dx.doi.org/10.1016/j.socnet.2005.11.003

Fowler, J.H. (2006) Connecting the Congress: A Study of Cosponsorship Networks." *Political Analysis* 14 (4): 456-487 (Fall) http://dx.doi.org/10.1093/pan/mpl002

Fowler, J.H., Christakis, N.A. (2008) "Dynamic Spread of Happiness in a Large Social Network: Longitudinal Analysis Over 20 Years in the Framingham Heart Study." *British Medical Journal* 337: a2338 (4 December) http://dx.doi.org/10.1136/bmj.a2338

Fowler, J.H., Dawes, C.T., Christakis, N.A. "Model of Genetic Variation in Human Social Networks." *Proceedings of the National Academy of Sciences*, http://dx.doi.org/10.1073/pnas.0806746106

Gutmann, M.P., Witkowski, K., Colyer, C., O'Rourke, J.M., McNally, J. (2008) "Providing Spatial Data for Secondary Analysis. Issues and Current Practices relating to Confidentiality." *Population Research and Policy Review* 27: 639-665. http://dx.doi.org/10.1007/s11113-008-9095-4

Huang, B., Jebara, T. (2007) "Loopy Belief Propagation for Bipartite Maximum Weight b-Matching." *Proceedings of the 2007 Conference on Artificial Intelligence and Statistics*, AISTATS, (March)

Leskovec, J., Adamic, L.A., Huberman, B.A. (2006) "The dynamics of viral marketing." *Proceedings of the 7th ACM Conference on Electronic Commerce* (Ann Arbor, Michigan, USA, June 11 - 15, 2006). EC '06. ACM, New York, NY, 228-237. http://doi.acm.org/10.1145/1134707.1134732

Liben-Nowell, D., Novak, J., Kumar, R., Raghavan P., Tomkins, A. (2005) "Geographic routing in social networks." *Proceedings of the National Academy of Sciences* 102. (August)

Monge, P.R., Contractor, N. (2003). *Theories of communication networks*. New York: Oxford University Press

Nathan Eagle (2008), "Behavioral Inference Across Cultures: Using Telephones as a Cultural Lens." *IEEE Intelligent Systems* 23: 4, 62-64

National Research Council. (2007) *Putting People on the Map: Protecting Confidentiality with Linked Social-Spatial Data*. Ed. Gutmann, M.P., Stern, P. Washington: National Academy Press

National Research Council. (2004) *Protecting Participants and Facilitating Social and Behavioral Science Research*. Washington: National Academies Press

Reynolds, M., Van Alstyne, M. Aral, S. (2009) "Functions that Preserve Privacy but Permit Analysis of Text." MIT Center for Digital Business Working Paper # 246, Cambridge, MA

Roy, D., Patel, R., DeCamp, P., Kubat, R., Fleischman, M., Roy, B., Mavridis, N., Tellex, S., Salata, A., Guiness, J., Levit, M., Gorniak, P. (2006) "The Human Speechome Project." *Proceedings of the Twenty-eighth Annual Meeting of the Cognitive Science Society*

Salganik, M.J., Dodds, P.S., Watts, D.J. (2006) "Experimental study of inequality and unpredictability in an artificial cultural market." *Science* 311:854–56

Shaw, B., Jebara, T. (2007) "Minimum Volume Embedding." *Proceedings of the 2007 Conference on Artificial Intelligence and Statistics,* AISTATS, (March)

Van Wey, L., Rindfuss, R., Gutmann, M.P., Entwisle, B., Balk, D. (2005) "Confidentiality and Spatially Explicit Data: Concerns and Challenges." *Proceedings of the National Academy of Science* 102(43): 15337-15342

Wu, F., Huberman, B.A. (2007) "Novelty and collective attention." *Proceedings of the National Academy of Sciences* 104:17599–17601

Wu, L., Waber, B., Aral, S., Brynjolfsson, E., Pentland, S. (2008). "Mining Face to Face Interaction Networks Using Sociometric Badges: Predicting Productivity in an IT Configuration Task." *Proceedings of the 29th Annual International Conference on Information Systems*, December 15-18, Paris, France. http://ssrn.com/abstract=1130251

Online resources

http://socdynamics.org
http://www.hks.harvard.edu/netgov