

privacy. How can the government ensure the proper usage and control of citizen-generated contents? How can the government safeguard online transaction and information sharing among citizens? As the Internet becomes more pervasive, the inherent danger of information leaks, information misuse, corruption and manipulation, and system intrusion has become more imminent. Careful cybersecurity and information assurance research must address these issues.

**T**his issue's Trends and Controversies department includes five essays on e-government and politics 2.0 from distinguished experts. Each essay presents a unique, innovative research framework, computational methods, and selected results and examples.

In "Blogsphere Research: A Mixed-Methods Approach to Rapidly Changing Systems," David Karpf discusses the growth of the political blogsphere and presents a few lessons that may be applicable to IT and political research. Karpf suggests that the rapidly changing nature of information technologies and the sheer abundance of available data are two of the most pressing challenges for researchers. In "OntoCop: Constructing Ontologies for Public Comments," Hui Yang and Jamie Callan present a system, designed to help organize online public comments from citizens, that works interactively with a person to organize a set of concepts into an ontology. Experimental results show that interactive learning produces useful ontologies and saves time and human effort. In "Enabling the Dialogue—Scientist<>ResourceManager<>Stakeholder: Visual Analytics as Boundary Objects," Judith B. Cushing and colleagues report on scientists' use of visual

analytics to repurpose ecology research for government resource managers. The authors believe that visual analytics can act as effective boundary objects to communicate and improve understanding when working across linguistic, conceptual, disciplinary, or interest group boundaries. In "Moving Toward 'Intelligent' Policy Development?" Ann Macintosh presents selected augmentation support tools for assisting effective policy development. Social network analysis, argumentation mining, argumentation analysis, and argumentation visualization are some of the promising areas of research she suggests. In the fifth and last essay, "E-Government 2.0 in Asia: Trends, Opportunities, and Challenges," Paul Jen-Hwa Hu and colleagues describe representative e-government projects in Asia and present several more in-depth case studies of e-government development in Taiwan. The authors suggest knowledge mapping, social media scanning, and social network analysis as promising research directions.

### Acknowledgments

I greatly appreciate some of the materials presented and discussions that occurred in my recent Knowledge Management class at the University of Arizona. Special thanks to students Adam Aronoff, Ben Gray, Preston Troy, Pooja Bhandari, and Clayton Badeaux. I also acknowledge the previous funding support provided by the NSF Digital Government program for my COPLINK research ("COPLINK Center: Information and Knowledge Management for Law Enforcement," EAI-9983304, 2003–2008).

### References

1. S.L. Stirland, "Propelled by Internet, Barack Obama Wins Presidency," *Wired*, 4 Nov. 2008; <http://blog.wired.com/27bstroke6/2008/11/propelled-by-in.html>.
2. N. Scola, "Obama and Politics 2.0," TechPresident, Personal Democracy Forum, 17 July 2008; <http://techpresident.com/blog-entry/obama-and-politics-20-documenting-history-real-time>.

3. J. Green, "The Amazing Money Machine," *The Atlantic*, vol. 301, no. 5, June 2008; [www.theatlantic.com/doc/200806/obama-finance](http://www.theatlantic.com/doc/200806/obama-finance).
4. H. Chen et al., eds., *Digital Government: E-Government Research, Case Studies, and Implementation*, Springer, 2007.
5. T. O'Reilly, "What Is Web 2.0? Design Patterns and Business Models for the Next Generation of Software," presented at O'Reilly Media Web 2.0 Conf., 2004; <http://oreilly.com/web2/archive/what-is-web-20.html>.
6. C. O'Hara, "Virtual Learning Gets Second Wind from Second Life," *Federal Computer Week*, vol. 23, no. 10, 4 May 2009, pp. 26–31.

---

**Hsinchun Chen** is the McClelland Professor of Management Information Systems at the University of Arizona and director of the Artificial Intelligence Lab. He received his PhD in information systems from New York University. Contact him at [hchen@eller.arizona.edu](mailto:hchen@eller.arizona.edu).

### Blogsphere Research: A Mixed-Methods Approach to Rapidly Changing Systems

David Karpf, *Brown University*

"Politics 2.0" can be understood as the harnessing of the Internet's lowered transaction costs and condition of information abundance toward the goal of building more participatory, interactive political institutions. As with any such term, the definition's boundaries are fuzzy, and the topic lends itself both to technologically deterministic prognostication and hastily constructed rebuttal. Among the various elements of politics 2.0, the political blogsphere has attracted

the greatest early scrutiny. This essay reviews political science research on the blogosphere, noting in particular how the medium itself has continued to change and evolve, undermining the assumptions in our research designs. Only through a combination of quantitative and qualitative methods have we been able to accurately depict the size, scope, and usage of political blogs in American politics today. A few lessons from the brief history of blog research may be applicable to the study of technology and politics 2.0 more generally.

### **Political Blogging, Take 1**

The first generation of blog research emerged in 2003 and 2004. Although blog software had been widely available since Pyra Labs introduced its Blogger platform in 1999, blogging didn't receive attention as a significant tool for political engagement until 2002, when political bloggers played a key role in pressuring US Senator Trent Lott to step down from his position as Majority Leader (after making racially controversial statements at the birthday party of retiring Senator Strom Thurmond). Early bloggers were a mix of "citizen journalists" and amateur political strategists, relying upon the near-costless publishing platform to express their ideas and opinions, and creating nascent communities through blogrolls and within-text hyperlinks. Political bloggers were also influential in supporting Howard Dean's Presidential candidacy and in exposing forged documents presented by Dan Rather on *60 Minutes*.

Early academic research on blogging focused on the potential of the low-cost tool for radical increases in mass participation and as a challenge to elite media operations. Anecdotal examples of blogger effectiveness, coupled with empirical evidence of rapid

blogosphere growth, suggested that the new medium could give every motivated individual the ability to reach millions and influence public policy.

By those measures, blogging proved to have a rather disappointing track record. As Matthew Hindman demonstrates in his 2008 book,<sup>1</sup> political blogs have attracted far less attention than humor and entertainment blogs, and the maturation of the blogosphere has featured the growth of a limited number of "hub" sites that attract exponentially more hyperlinks and site visits than the average blog. Blog software may give anyone

---

**Blog software may give anyone a megaphone, but with the information abundance of the Internet, only the elite few can attract a large audience.**

---

a megaphone, but with the information abundance of the Internet, only the elite few can attract a large audience. Given that the lion's share of top political bloggers held advanced graduate degrees, the development of a new Internet elite seemed to mirror the same demographic disparities prevalent in offline society. Meanwhile, though blogs provided an alternate venue to the mainstream media, major media outlets began hiring bloggers and incorporating blogs into their set of offerings. Although newspapers are clearly experiencing financial distress today, blogs are hardly

the main culprit. Rather, it was Craigslist that undermined the market for classified ads, and Google that altered how we search for and obtain new information. Blogs have served as an alternate venue for political reporting and engagement, but not one that necessarily replaces or challenges existing elite structures.

### **Political Blogging, Take 2**

If the blogosphere failed to live up to the hopes of first-generation proponents and researchers, however, that isn't to say that it had no substantive impacts. Though the introduction of online self-publishing tools failed to radically recast elite political institutions, the blogosphere has nonetheless yielded a new set of elites and, arguably, increasingly porous networks of influence within those institutions. The introduction of these new technologies of political engagement has yielded opportunities for the formation of novel political associations and a host of expanded tactical repertoires that change how politically active citizens engage in politics. Indeed, one of the chief problems the initial generation of Internet researchers faced was that, as the blogosphere expanded, it was adopted in new and unexpected manners that didn't fit within the boundaries of our research program.

Consider DailyKos.com, the largest political blogging community in America today. Daily Kos was initially the personal blogging home of Markos "Kos" Moulitsas, an outspoken left-wing Democrat. In October 2003, Moulitsas adopted a new "community blogging" software platform operated by "Scoop." The community blogging platform allowed all registered users to post their own blog entries as "diaries" on the site. The site quickly soared in popularity as progressive bloggers settled upon it

as a hub space for their community of interest. As several longtime bloggers revealed in interviews, “the difference between Kos and everyone else wasn’t that he was so much more talented. It was that he adopted the community platform first.”

The introduction of community blogging software is one dimension of the blogosphere’s shifting terrain. These platforms enable blogs to function as coordination points for online communities of interest, the largest of them functionally indistinguishable from political interest groups. The Daily Kos community establishes campaign priorities, endorses political candidates, fundraises and volunteers for them, and even holds an annual in-person convention. While the first generation of Internet researchers were busy exploring the degree to which individual elite bloggers were distinct from other political elites, the bloggers themselves were refashioning their sites to provide greater voice and mobility to their communities.

A second dimension of innovation involves the adoption of blogging platforms into the Web offerings of existing institutions. While early proponents of the blogosphere envisioned the small core of counter-institutional bloggers expanding ever outward and reshaping American politics, the actual diffusion process looked quite different. Once blogging gained enough notoriety to be taken seriously, existing media and political institutions adopted the technology themselves. Major media institutions hired bloggers to work full time for them, offering content on their sites. Political campaigns, businesses, and interest groups all added blogs to their own sites, and new content management systems made blogging a basic feature of Web site redesign. These “institutional blogs,” of course, adapt the technology

**Table 1. Blogosphere authority index overview.**

Score	Definition
NCS	The <i>network centrality score</i> gathers applied sociometric data based on appearance in progressive blogrolls.
HAS	The <i>hyperlink authority score</i> rates a site’s authority among political bloggers using link patterns, as measured by Technorati.
STS	The <i>site traffic score</i> rates visits-per-day statistics as measured by a combination of Sitemeter and Alexa traffic rankings.
CAS	The <i>community activity score</i> rates interactive participation as measured by total blog comments.
BAI	The final <i>blogosphere authority index</i> score combines the raw NCS, HAS, STS, and CAS scores. The three best scores are added, and the fourth is dropped; the ranking of the cumulative scores gives the index.

to their own needs and goals. As technology writer Clay Shirky predicted in 2003, “At some point (probably one we’ve already passed), weblog technology will be seen as a platform for so many forms of publishing, filtering, aggregation, and syndication that blogging will stop referring to any particularly coherent activity.”<sup>2</sup> The central challenge for the research community, then, lies in accurately measuring and describing a phenomenon that is itself rapidly changing in unpredictable ways. I treat these two types of blog innovation as dimensions in a typology of blogspace, allowing for categorization and comparison of differing blog formats.<sup>3</sup>

An additional challenge in this regard comes with the issue of how we measure influence in the blogosphere. Hyperlink analysis and site traffic are the two most common measures, but each has its particular flaws. Hyperlinks give an accurate map of clustering and communities within the larger blogosphere, but the rise of “splogs” (spam blogs) muddies the waters. Also, as community sites like Daily Kos become more self-referential, a number of the outbound hyperlinks go to the traditional news organizations that employ full-time journalists. The relationship between links and site traffic heavily fluctuates. Site traffic, meanwhile, is notoriously difficult to measure, and the best data is kept behind proprietary firewalls. In my own research, I convert four independent

measures of blog influence (network centrality, hyperlinks, site traffic, and total comment volume) into ordinal rankings, then merge those rankings to form a composite ranking system for the political blogosphere. Table 1 describes the individual measures I use. Critically, this methodology is the only currently available that produces rankings without relying solely on a single, flawed metric.<sup>3</sup>

Combining the 2D blogspace map with the composite rankings of the elite blogosphere yields several key findings about the partisan makeup of the American political blogosphere. The community blogging platform has been heavily used by progressive bloggers (to great financial benefit for their favored political candidates), while conservative attempts to replicate Daily Kos and the other large community sites have largely foundered. The most successful conservative bloggers operate closed-authorship individual or institutional blogs, limiting their capacity for collective action in the blogosphere. As blog traffic increased in the 2008 election season, the progressive blog network substantially increased its advantage over its conservative counterpart.<sup>4</sup> These findings only emerge when detailed qualitative analysis techniques are combined with large-scale data collection. The research literature ignored variability in blog type for years because qualitative insights on

the use of blogs were left apart from quantitative research design choices.

### Lessons for Future Researchers

The two greatest challenges for researchers interested in e-government and politics 2.0 are the rapidly changing nature of the technologies under study and the sheer abundance of available data, much of which is of questionable quality. Link-based analysis in particular has become highly popular recently, as complex data visualization packages claim to offer “maps” of Web communities and their most influential members. Although there is some substantial value in these tools, we should recall the longstanding slogan, “garbage in, garbage out”: a data analysis program can never far exceed the quality of its inputs. And in particular, we lack an empirically based understanding of what a link actually signifies. Increased traffic? More or less. Ideological homophily? Occasionally. Avenues for diffusion? Possibly. What’s more, it’s entirely possible that a link between sites at time  $x$ , where  $x$  occurs in the lead-adopter phase of diffusion, will mean something quite different from a link between sites at time  $x + 5$ , where  $x + 5$  occurs during the early-majority phase of adoption.

This points to the importance of the diffusion of innovation literature<sup>5</sup> for the research community. As with the blogosphere, it seems evident that most Web-based political tools are first used by a small core of highly tech-savvy “lead adopters” who cocreate and alter the medium. This group is demographically and ideologically distinct from the “early adopter” and “early majority” classes, each of which is a good deal larger than the initial class. As a technology scales up, it adapts to the interests of these larger segments of

the populace, and in turn is modified and adopted by existing institutions that show little interest in the technology until it has moved beyond the tiny core of lead adopters. That these changes in the nature and scope of a social technology will occur is both predictable and model-able, but the resultant direction of such changes requires close observation of the emerging large-scale communities.

**T**he condition of online data abundance is both a blessing and a curse to the Web sciences research community. Torrential data is available for analysis; what it signifies is far less clear. Large-scale data analysis must be wedded to close observation of the changing boundaries of the phenomena, communities, and platforms under study, so that we may remain attuned to changes in usage patterns, form realistic hypotheses, and separate high-quality data from the noise.

### References

1. M. Hindman, *The Myth of Digital Democracy*, Princeton Univ. Press, 2008.
2. C. Shirky, “Power Laws, Weblogs, and Inequality,” 2003; [www.shirky.com/writings/powerlaw\\_weblog.html](http://www.shirky.com/writings/powerlaw_weblog.html).
3. D. Karpf, “Understanding Blogspace,” *J. Information Technology and Politics*, vol. 5, no. 4, 2008, pp. 369–385.
4. D. Karpf, “Stability and Change in the Political Blogosphere in the 2008 Election,” paper presentation, Midwest Political Science Assoc. Ann. Meeting, Chicago, 2009.
5. E. Rogers, *Diffusion of Innovation*, 5th ed., Free Press, 2003.

**David Karpf** is a postdoctoral research associate at the Taubman Center for Public Policy at Brown University. He holds a PhD in political science from the University of Pennsylvania. Contact him at [davekarpf@gmail.com](mailto:davekarpf@gmail.com).

## OntoCop: Constructing Ontologies for Public Comments

Hui Yang and Jamie Callan, *Carnegie Mellon University*

US law defines a process known as *Notice and Comment Rulemaking* that requires regulatory agencies to seek comment from the public before establishing new regulations. Regulatory agencies are also expected to demonstrate that the final regulation addresses all substantive issues raised by the public. Most proposed regulations attract few comments from the public, but each year a few regulations attract tens of thousands or hundreds of thousands of comments. When comment volume is high, most comments are form letters and modified form letters, which are not difficult to process,<sup>1</sup> but there are still tens of thousands of unique comments that address a variety of issues. There may also be political or public pressure for the agency to issue regulations quickly. In these cases, regulatory agencies and other interested parties need tools that help them to quickly make sense of public comments.

Browsing hierarchies such as the Yahoo Directory are a popular method of quickly discovering the “lay of the land” in a large text corpus. By associating documents with topics and concepts in a hierarchy, we can structure and partition the information space into smaller spaces that are easy to understand and navigate. Regulation-specific browsing hierarchies make it easier to understand the range of issues that the public has raised; they also let agencies and policymakers drill down into comments that discuss particular issues, which lets them be more responsive to the public’s concerns. Information analysis tools that allow quick and efficient