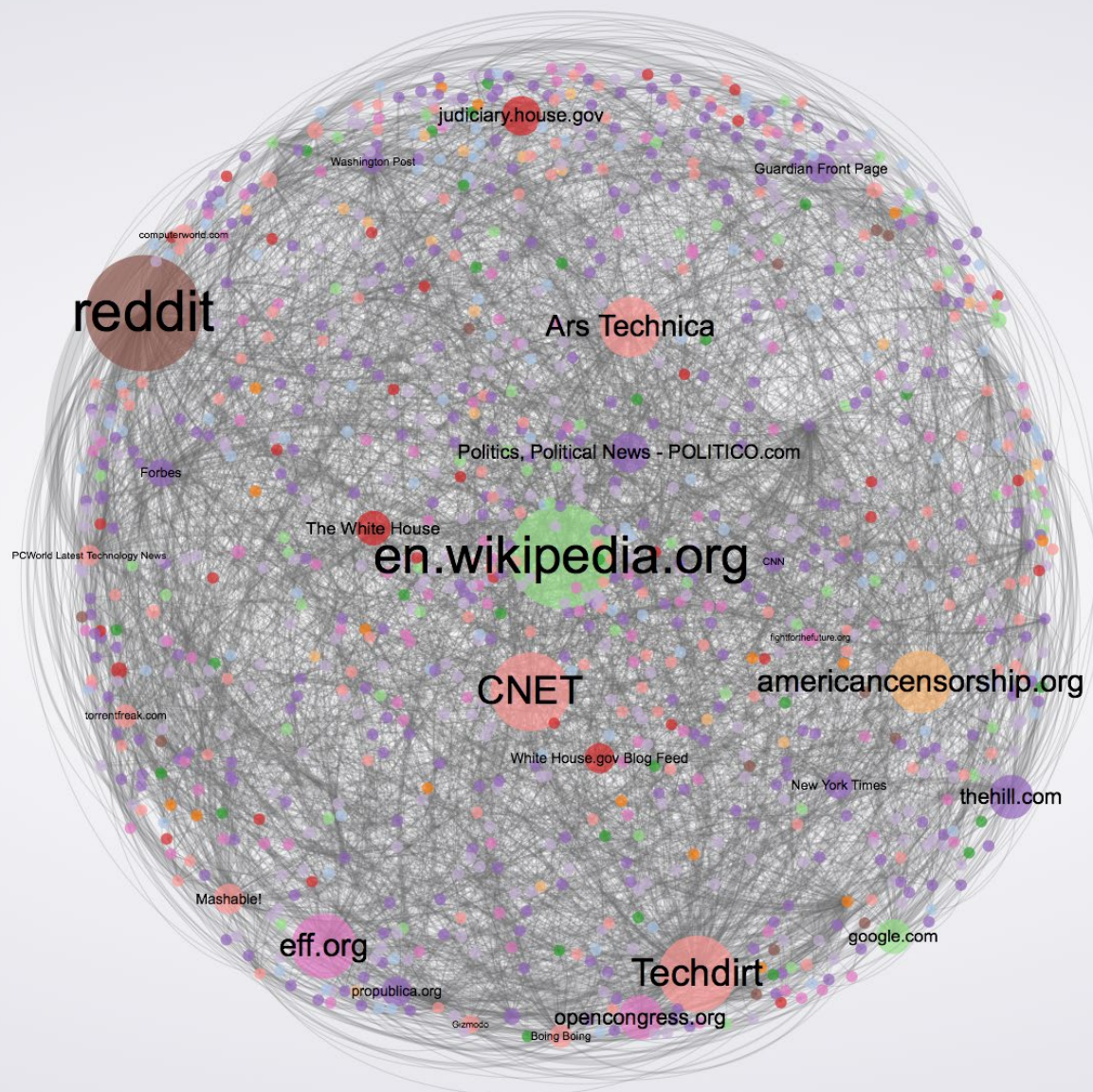


Social Mobilization and the Networked Public Sphere: Mapping the SOPA-PIPA Debate

By Yochai Benkler, Hal Roberts, Robert Faris, Alicia Solow-Niederman, Bruce Etling
July 2013



Berkman

The Berkman Center for Internet & Society
at Harvard University

Download the electronic version of this paper:

http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2295953

Acknowledgements

This paper would not have been possible without the help and input of many others. Amar Ashar coordinated the many pieces of this research project, helped to keep us on track, and contributed to the production of the paper and web resources. Mayo Fuster Morell generously provided deep substantive feedback on the paper and helped us to develop, refine, and better describe the methods employed for this study. Our friends and collaborators at the Center for Civic Media, led by Ethan Zuckerman, have played a foundational role in working with the Media Cloud team at the Berkman Center to develop the tools and methods used for this project. Their active participation and support over the past two years was instrumental in producing this study. David Larochelle helped to build the Media Cloud platform and continues to develop and maintain its technical infrastructure. Jennifer Jubinville supported the technical team. Justin Clark is responsible for creating the web-based visualizations of the network maps that accompany this paper. Zoe Fraade-Blanc contributed to the design of the online tool. Olivia Conetta provided editorial help. Jessie Schanzle, Malavika Jagannathan, Marianna Mao, Cale Weissman, and Melody Zhang contributed research assistance. The Media Cloud team is grateful to the leadership of the Berkman Center, in particular Urs Gasser, Colin Maclay, Caroline Nolan, and the faculty board of directors, for their continued support and guidance.

We thank the many people—far too many to mention—who shared their perspectives and knowledge on the debate, engaged in conversations that informed our understanding and analysis of the controversy, and participated in talks and workshops devoted to this topic.

We are grateful to the Ford Foundation and the Open Society Foundation for their generous support of this research and of the development of the Media Cloud platform.

About Media Cloud



Media Cloud, a joint project of the Berkman Center for Internet & Society at Harvard University and the Center for Civic Media at MIT, is an open source, open data platform that allows researchers to answer complex quantitative and qualitative questions about the content of online media. Using Media Cloud, academic researchers, journalism critics, and interested citizens can examine what media sources cover which stories, what language different media outlets use in conjunction with different stories, and how stories spread from one media outlet to another.

<http://www.mediacloud.org/>

Abstract

This paper uses a new set of online research tools to develop a detailed study of the public debate over proposed legislation in the United States designed to give prosecutors and copyright holders new tools to pursue suspected online copyright violations. For this study, we compiled, mapped, and analyzed a set of 9,757 stories relevant to the COICA-SOPA-PIPA debate from September 2010 through the end of January 2012 using Media Cloud, an open source tool created at the Berkman Center to allow quantitative analysis of a large number of online media sources. This study applies a mixed-methods approach by combining text and link analysis with human coding and informal interviews to map the evolution of the controversy over time and to analyze the mobilization, roles, and interactions of various actors.

This novel, data-driven perspective on the dynamics of the networked public sphere supports an optimistic view of the potential for networked democratic participation, and offers a view of a vibrant, diverse, and decentralized networked public sphere that exhibited broad participation, leveraged topical expertise, and focused public sentiment to shape national public policy. We find that the fourth estate function was fulfilled by a network of small-scale commercial tech media, standing non-media NGOs, and individuals, whose work was then amplified by traditional media. Mobilization was effective, and involved substantial experimentation and rapid development. We observe the rise to public awareness of an agenda originating in the networked public sphere and its framing in the teeth of substantial sums of money spent to shape the mass media narrative in favor of the legislation. Moreover, we witness what we call an attention backbone, in which more trafficked sites amplify less-visible individual voices on specific subjects. Some aspects of the events suggest that they may be particularly susceptible to these kinds of democratic features, and may not be generalizable. Nonetheless, the data suggest that, at least in this case, the networked public sphere enabled a dynamic public discourse that involved both individual and organizational participants and offered substantive discussion of complex issues contributing to affirmative political action.

Interactive versions of this paper can be found at

http://cyber.law.harvard.edu/research/mediacloud/2013/mapping_sopa_pipa/

Introduction

On September 20, 2010, the *Hill*, a daily newspaper that covers the US Congress, reported the introduction of a “bipartisan bill that would make it easier for the Justice Department to shut down Web sites that traffic pirated music, movies, and counterfeit goods.”¹ The bill, introduced by Senators Patrick Leahy and Orrin Hatch, described in technocratic terms that it “would create an expedited process for the DOJ to shut down Web sites providing pirated materials”, and was accompanied by Senator Leahy’s confident statement that “protecting intellectual property is not uniquely a Democratic or Republican priority—it is a bipartisan priority.” Seventeen months later, on January 18, 2012, Wikipedia was blacked out for a day to protest a successor bill. Its front page read: “Imagine a World Without Free Knowledge. For over a decade, we have spent millions of hours building the largest encyclopedia in human history. Right now, the US Congress is considering legislation that could fatally damage the free and open Internet.” That day, several million people phoned or emailed Congress to protest the bill. This unprecedented surge of mobilization forced Congress to retreat from proposed legislation that started out with bipartisan support and the backing of some of the most powerful lobbies in the United States, including Hollywood, the recording industry, and, most significantly, the US Chamber of Commerce. The work that follows analyzes the dynamics of this debate.

This paper uses a new set of online research tools to develop a detailed study of the progression of the public debate over what began as the Combating Online Infringement and Counterfeits Act (COICA) and ultimately failed as the Stop Online Piracy Act (SOPA) in the House and the PROTECT IP Act (PIPA) in the Senate. It combines text and link analysis with human coding and informal interviews to map the controversy over the relevant 17 months and thereby offers an analysis of the shape of the networked public sphere engaged in this issue. The data suggest that, at least in this case, the networked public sphere enabled a dynamic and diverse discourse that involved both individual and organizational participants and offered substantive discussion of complex issues contributing to affirmative political action. This story depicts a depth and range of activity that is more consequential than most discussions of the networked public sphere in the last decade would predict. Instead of fragmentation and polarization, there was widespread attention across partisan and substantive divides, spanning Tea Party Patriots and libertarians along with traditional liberal and conservative factions. Tech media played a critical role, but game sites and political blogs were also significant. Non-governmental organizations (NGOs) and venture capitalists all showed up at different stages of the debate, and sites created specifically for this campaign served to aggregate and redirect attention at policy makers. Mainstream media played a role, though not a central one. And a varied set of sites collectively formed an attention backbone, linking together different clusters in the network and providing a boost to less visible sites to reach broader audiences. As we describe in this paper, the SOPA-PIPA² debate offers a view of a vibrant and diverse networked public sphere that exhibited broad participation, leveraged topical expertise, and focused public sentiment to shape national public policy. Because the controversial topic was technology-centric and thus intensely interesting to technologically capable individu-

1 Gautham Nagesh, “Bipartisan Bill Would Ramp Up Anti-Piracy Enforcement Online,” *Hill*, September 20, 2010.

2 In this paper we use the term “SOPA-PIPA” as shorthand for the debate that commenced with and includes the first version of this legislation, COICA.

als, and the effective action required was blocking a discrete legislative proposal in a veto-rich environment, it is not clear that the SOPA-PIPA dynamic will be replicated in other public policy debates and become a more generalized form of civic discussion and engagement. Nonetheless, this case study offers a rich example of how a mobilized and effective networked public sphere can function.

The Networked Public Sphere

Facilitated by the spread of digital communication technologies, the networked public sphere has emerged over the past two decades as an important venue for discussion and debate over matters of public interest. The networked public sphere is an alternative arena for public discourse and political debate, an arena that is less dominated by large media entities, less subject to government control, and more open to wider participation. The networked public sphere is manifest as a complex ecosystem of communication channels that collectively offer an environment that is conducive for communication and the creation of diverse organizational forms. This digital space provides an alternative structure for citizen voices and minority viewpoints as well as highlights stories and sources based on relevance and credibility.³

A robust ongoing debate over the Internet's impact on democracy and the democratic character of the networked public sphere has evolved over time in tandem with the development and adoption of digital technologies. This debate began with a rather utopian early stage in the 1990s. The US Supreme Court in *Reno v. ACLU* captured the spirit of the times:⁴

Any person or organization with a computer connected to the Internet can “publish” information. [...] Through the use of chat rooms, any person with a phone line can become a town crier with a voice that resonates farther than it could from any soapbox. Through the use of Web pages, mail exploders, and newsgroups, the same individual can become a pamphleteer.

Nicholas Negroponte emphasized the highly tailored information and knowledge we could acquire to become better-informed citizens and consumers, using the term “the Daily Me” to describe his optimistic vision.⁵ Yochai Benkler argued that the increasing importance of the commons as a factor of information production would weaken the power of the state and of incumbent media to shape public debate and that radically decentralized, commons-based production by once passive consumers would enhance participation and diversity of views.⁶ By 2002, however,

3 Yochai Benkler, *The Wealth of Networks: How Social Production Transforms Markets and Freedom* (New Haven, CT: Yale University Press, 2007).

4 *Reno vs. American Civil Liberties Union*, 521 U.S. 844 (1997).

5 Nicholas Negroponte, *Being Digital* (New York: Vintage Books, 1995).

6 Yochai Benkler, “Communications Infrastructure Regulation and the Distribution of Control over Content,” *Telecommunications Policy* 22, no. 3 (1998a): 183-196; Yochai Benkler, “The Commons as Neglected Factor of Information Production” (lecture, Telecommunications Policy Research Conference, September 1998b); Yochai Benkler, “Free as the Air to Common Use: First Amendment Constraints on Enclosure of the Public Domain,” *New York University Law Review* 74 (1999): 354-446; Yochai Benkler, “From Consumers to Users: Shifting the Deeper Structures of Regulation towards Sustainable Commons and User Access,” *Federal Communications Law Journal* 52 (2000): 561-597.

that early wave had given way to more skeptical writing. Cass Sunstein set the tone that would mark much of the second wave, arguing that “the Daily Me” stood not for refined knowledge, but rather for fragmentation, polarization, and the destruction of the possibility of common discourse in the public sphere.⁷

This first generation of arguments was based largely on anecdotal evidence. By 2001–2002, however, scholars began to apply network analysis to study the shape of participation and deliberation online. The most important and consistent finding was not fragmentation but rather concentration: researchers observed that linking patterns on the Web tend to follow a power law distribution,⁸ implying that speaking on the Internet is less like everyone being a town crier so much as everyone having the freedom to sing in the shower. Barabási and later Hindman claimed, in effect, that you can talk, but no one will hear you unless you are at the top of the link and attention economy; only a very small number of sites at the very top of the power law distribution would be seen.⁹ Interpreting then-published link analysis data, Benkler argued that participation nonetheless increased to the extent that individuals could contribute to debates directly or through someone they know directly. By contributing to blogs that are part of tightly clustered communities of interest, the argument was that less-known individuals could attract attention from ever larger attention clusters and communities, relying on mutual linking and the power law distribution as an attention backbone along which statements found to be interesting within a given cluster could travel and be observed outside that cluster.¹⁰

Drezner and Farrell argued that political bloggers could exert influence over the public because they were read by mass media,¹¹ an argument supported by Wallsten’s analysis of agenda setting and the blogosphere during the 2004 campaign.¹² Hindman countered these arguments with empirical claims that the overall size of the political public sphere was negligible, and that the leading voices in the blogosphere were as elite as those of the most exclusive editorial pages of the country’s newspapers.¹³ Wallsten’s work supported the latter portion of this claim as well.

At this point, Sunstein’s argument that the Internet increased polarization gained support from Adamic and Glance’s finding that only one in six links at the top of the left and right blogospheres linked across the ideological divide.¹⁴ Benkler disputed whether linking across the divide in one out of six cases should be interpreted as evidence of polarization and fragmentation as opposed

7 Cass Sunstein, *Republic.com* (Princeton, NJ: Princeton University Press, 2002).

8 Albert-László Barabási and Réka Albert, “Emergence of Scaling in Random Networks,” *Science* 286 (1999); “Power Laws, Weblogs, and Inequality,” Clay Shirky, last modified February 10, 2003, http://www.shirky.com/writings/powerlaw_weblog.html.

9 Albert-László Barabási, *Linked: How Everything Is Connected to Everything Else and What It Means for Business, Science, and Everyday Life* (New York: Penguin, 2003); Matthew Hindman, *The Myth of Digital Democracy* (Princeton, NJ: Princeton University Press, 2008).

10 Benkler, *The Wealth of Networks*.

11 Daniel Drezner and Henry Farrell, “The Power and Politics of Blogs,” *Public Choice* 134 (2008): 15-30.

12 Kevin Wallsten, “Agenda Setting and the Blogosphere: An Analysis of the Relationship between Mainstream Media and Political Blogs,” *Review of Policy Research* 24, no. 6 (2007): 567-587.

13 Hindman, *The Myth of Digital Democracy*.

14 Lada Adamic and Natalie Glance. “The Political Blogosphere and the 2004 US Election: Divided They Blog.” In *Proceedings of the 3rd International Workshop on Link Discovery*. Chiba, Japan: ACM Press, 2005.

to a normal allocation of attention to debates within one's political milieu and across the divide. Hargittai and her collaborators, in an early study combining link analysis with content analysis, showed that many of the links across the divide involved substantive argument, and that the two sides of the blogosphere did not exhibit greater insularity or polarization over time.¹⁵ Similarly, Gentzkow and Shapiro disputed the argument that online readers are subject to greater polarization and fragmentation in their media consumption patterns. They presented evidence that people online are exposed to a wider range of views than they are in their offline lives and that even those that are entrenched in one side of the political divide are exposed to opposing views.¹⁶ In a 2010 study, Lawrence, Sides, and Farrell observed that blog readers are particularly "activated," reporting high degrees of political participation in surveys, but that the most politically engaged were also the most polarized.¹⁷ Recent data-driven work on the shape of the political blogosphere has generally focused on the different practices on the left and right,¹⁸ in particular emphasizing that the left tends to adopt technologies and organizational practices that are more discursive and participatory, whereas the right tends to adopt technologies and practices that emphasize more hierarchical, one-way communications models.¹⁹

An important and complex set of questions in the field relate to the impact of digital technologies on civic engagement, social mobilization, and politics. Farrell highlights three causal mechanisms that shape the relationship of the Internet and politics: declining transaction costs for organizing collective action, homophilous sorting, and preference falsification.²⁰ Homophilous sorting—the tendency for individuals with common interests and shared views to form groups—is aided by the lower cost of finding like-minded individuals on the Internet and by the emergence of key nodes on the Internet that serve as meeting points for people with similar perspectives or interest in a particular issue. In a context where expressing one's political views may entail risks or social opprobrium, the Internet may offer a safer venue for individuals to express their true views, reducing the incidence of preference falsification and leading to a more accurate rendering of underlying political sentiment and to an environment more conducive for collective action.

The questions related to digitally mediated organizing for collective action stem from a long and rich literature on social movements.²¹ Many observers have voiced skepticism over the potential impact of digitally mediated collective action on political change. One view is that the Internet has enabled a new set of tactics that are useful for social movements, yet questions whether the

15 Eszter Hargittai, Jason Gallo, and Matthew Kane, "Cross-Ideological Discussions among Conservative and Liberal Bloggers," *Public Choice* 134 (2008): 67-86.

16 Matthew Gentzkow and Jesse Shapiro, "Ideological Segregation Online and Offline," *The Quarterly Journal of Economics* 126, no. 4 (2011): 1799-839.

17 Eric Lawrence, John Sides, and Henry Farrell, "Self-Segregation or Deliberation? Blog Readership, Participation, and Polarization in American Politics," *Perspectives on Politics* 8, no. 1 (2010): 141-157.

18 David Karpf, "Understanding Blogspace," *Journal of Information Technology and Politics* 5, no. 4 (2008): 369-385; Kevin Wallsten, "Political Blogs: Transmission Belts, Soapboxes, Mobilizers, or Conversation Starters?," *Journal of Information Technology & Politics* 4, no. 3 (2008): 19-40.

19 Aaron Shaw and Yochai Benkler, "A Tale of Two Blogospheres: Discursive Practices on the Left and Right," *American Behavioral Scientist* 56, no. 4 (2012): 459-487.

20 Henry Farrell, "The Consequences of the Internet for Politics," *Annual Review of Political Science* 15 (2012): 35-52.

21 See, for example, the writing of Sidney Tarrow, Charles Tilly, and Doug McAdam.

Internet is able to create the stable ties required for sustained collective action.²² Building on that general argument, but based less in research and more on general observation and speculation, Gladwell argued that online ties are too weak to convert into effective political action.²³ This assertion garnered much attention, most of which was critical. Tufekci, for example, pointed out that Gladwell underestimated the power of weak ties and did not appreciate that strong ties and weak ties associated with social movements should not be seen as crowding out one another, but rather are often strong complements that work in conjunction to expand the reach and impact of activist communities.²⁴

The experiences of social movements over the past several years have helped to temper, if not eradicate, some of the more extreme views on this subject and fueled a surge in popular and academic interest on the topic. Many of the studies conducted in the run-up to and in the wake of the Arab Spring suggest that networked communications have played a significant role in creating networks of activists and routing around state media to deliver videos that helped fan the flames, which concentrated global and national attention on the uprisings and ultimately sustained action.²⁵ Studies coming out of the *Indignados* movement in Spain and Occupy Wall Street have drawn similar conclusions, though in markedly different social and political contexts.²⁶

The questions in this field go beyond the number of citizens engaged on issues of collective interest to ask whether digitally mediated activism can draw in a more diverse slice of society or whether we are seeing a reinforcement of existing inequalities in participation and access to political systems. This line of inquiry connects back to a longstanding interest in equity and participation often framed as the digital divide.²⁷

A related dynamic is the changing nature of organizational forms in the digital age. Benkler's depiction of commons-based peer-production—frequently taking on distributed organizational forms and falling outside of the prevailing organizational structures—is as applicable to politics and social movements as it is to economic and cultural production, even before considering the scope for overlap and mutual support between these realms. Bimber, Flanagin, and Stohl describe a similar shift from organizations with formal, hierarchical structures to those that allow

22 Jeroen Van Laer and Peter Van Aelst, "Internet and Social Movement Action Repertoires: Opportunities and limitations," *Information, Communication & Society* 13 (2010): 1146-1171.

23 Malcolm Gladwell, "Small Change: Why the Revolution Will Not Be Tweeted," *New Yorker*, October 4, 2010, http://www.newyorker.com/reporting/2010/10/04/101004fa_fact_gladwell#ixzz2Q4hO6Uar.

24 Zeynep Tufekci, "What Gladwell Gets Wrong: The Real Problem is Scale Mismatch (Plus, Weak and Strong Ties are Complementary and Supportive)," *Technosociology* (blog), September 27, 2010, <http://technosociology.org/?p=178>.

25 See for example: Philip Howard and Muzammil Hussain, "The Upheavals in Egypt and Tunisia: The Role of Digital Media," *Journal of Democracy* 22 (2011): 35-48; Philip Howard, *The Digital Origins of Dictatorship and Democracy: Information Technology and Political Islam* (New York: Oxford University Press, 2010); Zeynep Tufekci and Christopher, "Social Media and the Decision to Participate in Collective Action: Observations from Tahrir Square," *Journal of Communication* 62 (2012): 363-379; Marc Lynch, "After Egypt: The Limits and Promise of Online Challenges to the Authoritarian Arab State," *Perspectives on Politics* 9 (2011): 301-10.

26 See for example the set of articles in: "Occupy!," ed. Jenny Pickerill and John Krinsky, special issue, *Social Movement Studies: Journal of Social, Cultural and Political Protest* 11 (2012): 3-4.

27 Pippa Norris, *Digital Divide: Civic Engagement, Information Poverty, and the Internet Worldwide* (Cambridge, UK: Cambridge University Press, 2001).

for greater individual agency.²⁸ Shirky offers a view of online organizing in which the role of traditional organizational structure is diminished if not rendered unnecessary: organizing without organizations.²⁹ By contrast, Karpf argues that, in the United States, organizations still play a key role in intermediating collective action between citizens and government, but that new types of organizations are emerging, such as MoveOn, that are able to take advantage of low-cost digital tools.³⁰ Etling, Faris, and Palfrey have described the strengths and limitations of organizations across varying levels of hierarchy and the challenges inherent in applying the power of digital organizations to government and governance.³¹ In their many manifestations, digitally mediated organizations are increasingly recognized as providing alternatives to existing intermediaries in political processes and opening new avenues for social movements, political campaigns, and public policy advocacy as well as threatening traditional institutions.

Furthermore, debates over the shape and meaning of the networked public sphere have added to concerns over the decline or future of the fourth estate function. Paul Starr, as well as Robert McChesney and John Nicols, have raised concerns that the decline of the independent, advertising-supported local newspaper will undermine the watchdog role historically fulfilled by the fourth estate in democratic society.³² These and other authors seek solutions to the crisis of journalism in public or nonprofit support; others propose changes to intellectual property law, aimed at making competition from non-traditional media harder and allowing newspapers to retain sufficient rents to fund their operations.³³ By contrast, Benkler has argued that the networked public sphere is emerging as a combination of a smaller number of survivors of the major media outlets, possessing larger reach and integrating online contributions; small-scale for-profit online media like Snopes.com; nonprofit supported professional journalism like ProPublica; volunteer-driven new “party presses,” like Daily Kos; newly effective nonprofits, like the Electronic Frontier Foundation, Public Knowledge, or the Sunlight Foundation; and individuals in networks of mutual linking and attention.³⁴

Our study of the SOPA-PIPA debate provides a novel, data-driven perspective on the dynamics of the networked public sphere that tends to support the more optimistic view of the potential

28 Bruce Bimber, Andrew Flanagin, and Cynthia Stohl, *Collective Action in Organizations: Interaction and Engagement in an Era of Technological Change* (Cambridge, UK: Cambridge University Press, 2012).

29 Clay Shirky, *Here Comes Everybody: The Power of Organizing without Organizations* (New York: Penguin Books, 2008).

30 David Karpf, *The MoveOn Effect: The Unexpected Transformation of American Political Advocacy* (Oxford: Oxford University Press, 2012).

31 Bruce Etling, Robert Faris, and John Palfrey, “Political Change in the Digital Age: The Fragility and Promise of Online Organizing,” *SAIS Review* 30, no. 2 (2010): 37-49.

32 Paul Starr, “Goodbye to the Age of Newspapers (Hello to a New Era of Corruption),” *The New Republic*, March 4, 2009, <http://www.tnr.com/article/goodbye-the-age-newspapers-hello-new-era-corruption?page=1>; Robert McChesney and John Nichols, *The Death and Life of American Journalism: The Media Revolution that Will Begin the World Again* (New York: Nation Books, 2010); Leonard Downie, Jr., and Michael Schudson, “The Reconstruction of American Journalism,” *Columbia Journalism Review* (2009).

33 “2009 FTC Workshop: News Media Workshop,” Federal Trade Commission, <http://www.ftc.gov/opp/workshops/news/index.shtml>; Federal Trade Commission, *Federal Trade Commission Staff Discussion Draft: Potential Policy Recommendations to Support the Reinvention of Journalism*, www.ftc.gov/opp/workshops/news/jun15/docs/new-staff-discussion.pdf.

34 Yochai Benkler, “Giving the Networked Public Sphere Time to Develop,” in *Will the Last Reporter Turn Out the Lights: The Collapse of Journalism and What Can Be Done to Fix It*, ed. Robert McChesney and Victor Pickard (New York: The New Press, 2011); Yochai Benkler, “A Free Irresponsible Press: Wikileaks and the Battle over the Soul of the Networked Fourth Estate,” *Harvard Civil Rights-Civil Liberties Law Review* 46, no. 2 (2011): 311-397.

of networked democratic participation. We find that the fourth estate function was fulfilled by a network of small-scale commercial tech media, standing non-media NGOs, and individuals, whose work was then amplified by traditional media. Mobilization was effective and involved substantial experimentation and rapid development for collective action. We also observe the rise to public awareness of an agenda that intensely concerns a specific set of users and its framing in the teeth of substantial sums of money spent to shape the mass media narrative in favor of the legislation. Experts were able to participate both in framing the debate and in assessing the effects of legislative proposals as they developed over the course of the controversy. Moreover, we witness what we call an attention backbone, in which more trafficked sites amplify less visible individual voices, expert or lay, on specific subjects; for instance, a single Reddit user's idea (the Go Daddy boycott of December 2011) produced mobilization for action that resulted in large-scale political influence. And throughout the debate, we see cross-sector political attention and cooperation.

Compared to the mass-mediated public sphere and many computationally instantiated descriptions of the networked public sphere, the progression of this story suggests a networked public sphere that is more dynamic, diverse, decentralized, and effective. Our paper offers only a single case study, and a case study under the most favorable conditions. It therefore cannot be seen as decisive in the debates over the networked public sphere. Nonetheless, our findings here provide both evidence of and texture for the most prominent, discrete legislative success that fits the ideal model of the networked public sphere. Moreover, the methods we develop here offer a new tool set that will allow us to undertake more such case studies in the future and provide a richer evidence-based analysis of the present and future of the networked public sphere. If what we observed here is not a lone event, the future seems brighter than the skeptics of the past decade predicted. It need not become the standard form of politics to exert influence on other modes of political contestation; it merely needs to become a well-understood option in public contestation, both for those who were empowered by the mass-mediated public sphere and must now contend with new forms of expression and organization, and for those who were excluded from it and now have alternative pathways to participation and mobilization.

Digital Copyright: Background

Copyright law and policy have long been set by a lopsided political economy: incumbents in the copyright-dependent industries come to the legislature asking, "Please, sir, may I have some more?" Historically, users have been largely absent from the lobbying efforts, and through a process of negotiations among the then-powerful producer lobbies, law has been passed to codify the prevailing power balance in the industry.³⁵ The process repeated itself when digital technologies threatened to disrupt the settlement codified by the 1976 Copyright Act, and COICA, PIPA, and SOPA were merely the most recent iteration of a 17-year-long drive by the US copyright industries to use law and public policy to enlist various intermediaries and service providers of networked facilities to enforce their rights. Beginning in a Clinton White House white paper in

³⁵ Jessica Litman, "Copyright, Compromise, and Legislative History," 72 Cornell L. Rev. 857 (1987): 869-79; James Boyle, *The Public Domain: Enclosing the Commons of the Mind* (New Haven, CT: Yale University Press, 2008).

1995³⁶ and culminating in 1998, when the Digital Millennium Copyright Act (DMCA), the No Electronic Theft (NET) Act, and the Copyright Term Extension Act (Sonny Bono) were passed, the industries sought to create a set of liabilities that would lead Internet service providers and Web-hosting companies to remove infringing materials. The safe harbor notice and takedown procedures adopted in the DMCA represented the settlement of the first half-decade of policy-making in this field. Under these provisions, pure telecommunications carriers were excluded from the requirements of policing content. Providers of caching, Web hosting, and search engines and Web directories were required to have a procedure for receiving notices regarding specific offending materials and taking down those specified offending materials; however, they were not required to search them out themselves and were not required to block whole sites. These legislative changes were complemented by a series of litigation strategies intended to extend and tighten the legal control of copyright owners over use of their works in an effort to make the Net into a “celestial jukebox,” where everyone could buy anything they wanted, but get nothing for free.³⁷

The decade following 1998 represented a stalemate. On one hand, content industries sought to expand control over materials on the Net to preserve and expand their revenues; on the other, an emerging coalition of businesses in computers, software, and communications that profited from the free flow of information and cultural goods online joined together with civil society organizations motivated to preserve a space for a cultural commons and concerned that the efforts to impose controls would hamper the open, creative, participatory structure of the networked environment. This conflict led to extended legislative deadlock. While Republicans seem to have been less open to pressures from Hollywood, beginning in 2006, Democrats controlling the Senate began to push through a slate of laws aimed at implementing the longstanding agenda of the Motion Picture Association of America (MPAA). These included most prominently and pertinently the PRO-IP Act of 2008, which created an IP czar in the White House and funded additional resources for criminal copyright enforcement,³⁸ and provisions in the Higher Education Opportunity Act of 2008 that required colleges to redesign their networks and develop offerings that would enforce the interests of Hollywood and the recording industry against their students.³⁹ These laws began to expand both elements of what became the centerpiece of SOPA and PIPA: increased involvement of criminal enforcement authorities in what was traditionally an area of private commercial law and increased use of state power to harness private platform providers to enforce the interests of the copyright industries.

Furthermore, in the international arena, the US Trade Representative had been negotiating a series of bilateral and multilateral agreements. These agreements tried to do three things: (a) export the current political settlement of American copyright law to other countries; (b) create a ratchet effect, so that anything that was passed in Congress and set in an international agreement could no longer be reversed without placing the United States in violation of its international trade obliga-

36 Information Infrastructure Task Force, *Report of the Working Group on Intellectual Property Rights* (1995).

37 Jessica Litman, *Digital Copyright* (Amherst, NY: Prometheus Books, 2001); Boyle, *The Public Domain*; Siva Vaidyanathan, *Copyrights and Copywrongs: The Rise of Intellectual Property and How it Threatens Creativity* (New York: New York University Press, 2002).

38 Prioritizing Resources and Organization for Intellectual Property Act, Pub. L. 110-403.

39 The Higher Education Opportunity Act, Public Law 110-315.

tions; and (c) set new standards, more aggressive than those stipulated by US law, that would then form the basis for expanding domestic protection as well. The iteration of this dynamic that preceded the introduction of COICA (the direct antecedent to PIPA) was the Anti-Counterfeiting Trade Agreement (ACTA), under which a set of like-minded nations led by the United States, Japan, and the European Union developed a treaty intended to set new, higher international standards for digital copyright.

What came to be known as the SOPA-PIPA battle therefore surfaced after waves of successes by copyright incumbents to pass laws that enhanced their capacity to extract value from their work. The introduction of COICA as a bipartisan, uncontroversial, and technocratic statute fits with this overall pattern, and the original text included the core memes that the US Chamber of Commerce and Hollywood used to frame the narrative: “The Chamber estimates copyright piracy from music, movies, software and video games costs the US economy \$58 billion in total output every year. The Chamber’s projected cost of lost output is more than 370,000 domestic jobs, \$16.3 billion in earnings, and \$2.6 billion in tax revenue for state, local, and federal governments.”⁴⁰

Methods

This study of the evolution of the public controversy surrounding COICA, SOPA, and PIPA was conducted by collecting digital resources related to this debate, including digital news stories, blog posts, and Web pages, and analyzing the resulting corpus of digital media using a combination of quantitative and qualitative methods. Social network analysis served as the foundation for the quantitative analysis and provided the basis for creating weekly and monthly network maps, starting in September 2010 and continuing through January 2012. These maps were then used as a guide for a qualitative assessment of the debate that included reading many of the articles that were generated by the media sources that appear in the maps and also involved discussions with key players in the debate.

The quantitative mapping element of this study uses social network analysis of linking patterns that have been developed and employed in previous studies.⁴¹ This approach is based on the premise that links between Web pages are a reflection of human value judgments and that the number of links to Web pages is therefore indicative of the collective valuation of these sources and a good measure of their relative importance. A link to another site may or may not signal agreement with the views expressed there, and the links need not represent an implicit endorsement of the story or Web page in order to draw meaningful inferences using this method, but as a minimum threshold, links indicate that linked-to sites are deemed worthy of citation.

This study employed the media analysis platform Media Cloud, an open source tool created at the Berkman Center to allow quantitative analysis of a large number of online media sources. Media

40 Nagesh, “Bipartisan Bill Would Ramp Up Anti-Piracy Enforcement Online.”

41 See, for example, Lada Adamic and Natalie Glance. “The Political Blogosphere and the 2004 US Election: Divided They Blog.” In *Proceedings of the 3rd International Workshop on Link Discovery*. Chiba, Japan: ACM Press, 2005; Eszter Hargittai, Jason Gallo, Matthew Kane, “Cross-Ideological Discussions among Conservative and Liberal Bloggers.” *Public Choice*. (2008) 134:67-86.

Cloud allowed us to compile a corpus of stories and Web pages that cover this issue and to create the controversy maps upon which we base much of the analysis described in this paper. Using these tools, we were able to map who is saying what and who is citing whom in public-facing Web media at any point in the emerging public conversation.⁴² This paper provides an analysis over time of the mobilization and interaction of various actors in this controversy and the roles and centrality of different actors in the debate.

For this study, we compiled a set of 9,757 stories relevant to the COICA-SOPA-PIPA debate from September 2010 through the end of January 2012. To gather these stories, we started by mining a seed set of stories related to this topic from the Media Cloud collection of blogs and digital media sets, searching for content by matching a set of regular expressions based on the names and acronyms of the bills put forward during this time: COICA, PIPA, and SOPA. After a manual review of this story set to verify relevance to this debate, we ran a spider that followed all the links in these stories to locate other relevant stories. The spider then iterated over the new stories discovered by this process, extracting their links, downloading the linked URLs, extracting the text from those Web pages, and seeking text that matched the COICA/PIPA/SOPA pattern. We repeated this process until the spider found no new stories. We then supplemented this initial set with Google searches for the relevant text patterns and added fewer than 50 stories that our initial sweep missed.

We used the stories, media sources, and links described above to generate the maps found in the remainder of this paper, which inform and facilitate the analysis of the debate. In the maps, each node represents a different media source, and each edge, or line connecting a pair of nodes, represents one or more links between stories in the respective media sources. The size of each node is proportional to the number of incoming links during a given time period, offering a visual representation of the prominence of the different media sources at different junctures in the debate. For any given time period, the map includes any media source that either has a story that was published during that period or a story that was linked to from another story that was published during that period. Links in the map, and the weight they give individual nodes and relations among nodes, include only the links from stories published in the time period mapped.

This map layout clusters together the groups of nodes that link most heavily to one another. Generally, nodes that are most heavily connected to the rest of the network appear toward the center of the map; thus, nodes that appear towards the edges of the map are typically the least heavily connected to the rest of the network. We used the ForceAtlas2 layout of the Gephi network visualization tool to obtain this layout.⁴³ Most of the maps that appear in this paper are sized to depict the full landscape of the discussion for a given time period. Other maps, where noted, appear at a higher level of magnification for illustration purposes only.

To facilitate the interpretation of the maps, we partition the network by assigning each media

42 A more thorough description of Media Cloud and the quantitative methods used in this study is available in the appendix.

43 Mathieu Jacomy et al., "ForceAtlas2, A Graph Layout Algorithm for Handy Network Visualization." Preprint, submitted August 29, 2011. http://webatlas.fr/tempshare/ForceAtlas2_Paper.pdf. (ForceAtlas2 is a force-directed algorithm that determines the position of each node in the network by simulating a repulsive force between nodes and an attractive force by the links between nodes.)

source to one of several categories of media type: bloggers, campaign-specific site, gaming site, general online news media, government, independent group, news aggregator, private sector, social linking site, tech media, and user-generated content platform. These categories were created to highlight and distinguish the roles and level of activity of a diverse set of groups that participated in the debate.⁴⁴

MEDIA TYPE	NUMBER OF SITES	NUMBER OF STORIES	COLOR
Bloggers	300	789	Light purple
General online news media	242	2399	Dark purple
Tech media	128	3272	Peach
Private Sector	113	255	Blue
Independent groups	111	416	Pink
User-generated content platform	61	371	Light green
Government	38	66	Red
Gaming site	30	219	Dark green
SOPA/PIPA/COICA-specific campaign	21	38	Light yellow
News aggregator	18	184	Yellow
Social linking site	12	257	Brown

Table 1: Summary of Media Types

Our approach focuses on mapping the public-facing online portion of the networked public sphere, combining several elements to produce a novel study approach. First, unlike most prior studies, we understand the relevant communicative sphere not in terms of a stable, broad category of sites that are “blogosphere” or “political blogs,” but rather in terms of discrete “controversies.” By “controversy” we mean a set of communications and actions around a core set of connected issues, irrespective of whether they originate in blogs or mainstream media, Web sites, or even the customer-service discussion boards of gaming companies.⁴⁵ Controversies have linguistic markers, a temporal dimension, and a political-economy valence or potential outcome. We use textual cues, here “SOPA,” “PIPA,” and “COICA,” as ways of filtering a broader range of blogs, online media, organizational, and personal sites and drawing in sites that addressed the controversy.

Second, we emphasize the time dimension. We understand controversies as having a beginning, an end, and internal dynamics that can shift and evolve in revealing ways. We take as given that

⁴⁴ Four coders manually coded each media source in the data set by media type. Inter-coder reliability was determined by a test in which two coders double-coded 147 Web sites, resulting in a high Krippendorff’s alpha of .9. We asked the coders to answer this question about each media source: “Which category below best describes the individual(s) or entity that publish(es) this source?” Coders could select from the 11 possible media types listed in Table 1, which were identified based on a pre-coding review by the research team of the different media sources found in the data set.

⁴⁵ Bruno Latour, *Science in Action: How to Follow Scientists and Engineers Through Society* (Cambridge, MA: Harvard University Press, 1987). We borrow this framing from Bruno Latour’s conception of scientific controversies, but apply it here to political communication and action.

public debates happen over time, as a series of moves and countermoves, and posit that temporally sensitive tools offer better insight into the shape of influence, framing, and action than do tools that capture broad, time-independent states. As a result, we find a public sphere that is more diverse and dynamic than has generally been portrayed by prior computationally instantiated analyses of the networked public sphere.

Third, we combine the text and link analysis with detailed human analysis, including interviews, desk research, and coding, particularly around highly visible stories and sites that emerge as significant from the data-driven network analysis. These efforts allow us to characterize the phenomena we observe in the data in terms of the unfolding political economy and the discursive structure of the controversy over time. In this way, our work preserves more of the richness and complexity of historical and sociological analysis of social movements, using computation to create a corpus of objective data that both guides our selection of particular interventions and organizations for more detailed analysis and highlights relations among sites and interventions at given moments over the course of the controversy.

The decision to adopt a mixed methods approach for this study is motivated not only by recognition of the value of detailed qualitative analysis in understanding the dynamics of public policy debates, but also by the belief that this choice permits us to test and better understand the strengths and limitations of the quantitative methods developed for this study. Accordingly, we use qualitative analysis both to complement the insights derived from the link-based methods and to offer an alternative means to assess the validity of the quantitative methods.

The narratives that accompany the maps in later sections of this paper start with a simple description of the most prominent nodes that appear for that time period—that is, the media sources that received the greatest number of in-links. Layered on top of this, we add qualitative interpretations that are based on reading the content of the stories, observing the links between sources and the context of the links within a given story, and drawing on more general knowledge of the entities engaged in the debate. The overall conclusions that we derive from this study follow the same general process and are informed by a mix of quantitative and qualitative methods. Additionally, we conducted informal interviews and discussions with some of the key actors in the debate to inform the qualitative aspects of the study and to offer another point of reference and validity check for the quantitative mapping.

This study does not attempt to gauge audience, exposure, or readership of the various media that participate in the debate. Instead, we focus on the relative influence among the participants in the debate as manifested through linking behavior. The research and data gathering includes the portion of the networked public sphere comprised of digital media and blogs but does not cover activity in social network sites such as Facebook, microblogging platforms like Twitter, and bulletin boards. The study also does not include analysis of private and semi-private communication channels, such as email and listservs, thus leaving out much of the behind-the-scenes maneuvering. These choices are driven in part by resource constraints—covering all possible digital platforms would be more time-consuming and expensive—and in part by simple data availability: many private listservs and email are not available for analysis. An important question that remains, which we hope to address in future work, is the extent to which omitting certain platforms

from the analysis misses important elements of the debate or understates the role of certain actors who concentrate their efforts on a given platform. Our assumption here is that the activities, ideas, and influence of the most prominent actors will be reasonably reflected in the digital media included in this study, both because of the tendency of digital media organizations, activists, and the private sector to engage in multiple platforms and due to the strength of cross-media linking. We therefore believe that despite the data limitations, our approach provides significant insight into the dynamics of a given controversy. For this particular controversy mapping, our observations based on the data largely cohere with what we learned from interviews and desk research into the dynamics of organization against COICA-SOPA-PIPA. What emerged from our study of over 9,000 articles, Web pages, and blog posts that discussed SOPA, PIPA, or COICA over a period of 17 months is a network map that reveals quite a different game from what we had seen in the traditional, mass-mediated public sphere.

When COICA-SOPA-PIPA Met the Networked Public Sphere

We begin tracking this debate in September 2010 and follow the debate until the demise of the legislation in January 2012. The debate is punctuated by several key events:

SEPTEMBER 2010	Introduction of COICA
NOVEMBER 2010	COICA passes the Senate Judiciary Committee; Senator Ron Wyden places a hold on the bill
MAY 2011	Introduction of PIPA; Senator Wyden again places a hold on the bill
SEPTEMBER 2011	Introduction of SOPA
NOVEMBER 2011	American Censorship Day
DECEMBER 2011	Reddit boycott of Go Daddy
JANUARY 2012	Online protests and blackouts

These key events coincide with spikes in online activity. It is notable that activity started off at a low level and included many weeks of quiet before becoming a phenomenon of national reach in early 2012, as shown in Figure 1. The initial stages of the debate are tiny in comparison to the peak of activity. This late-breaking crest means that the controversy was restricted to a modest number of highly engaged individuals and organizations for more than a year before a period of rapid expansion starting in November 2011 and building to the subsequent eruption in January 2012.

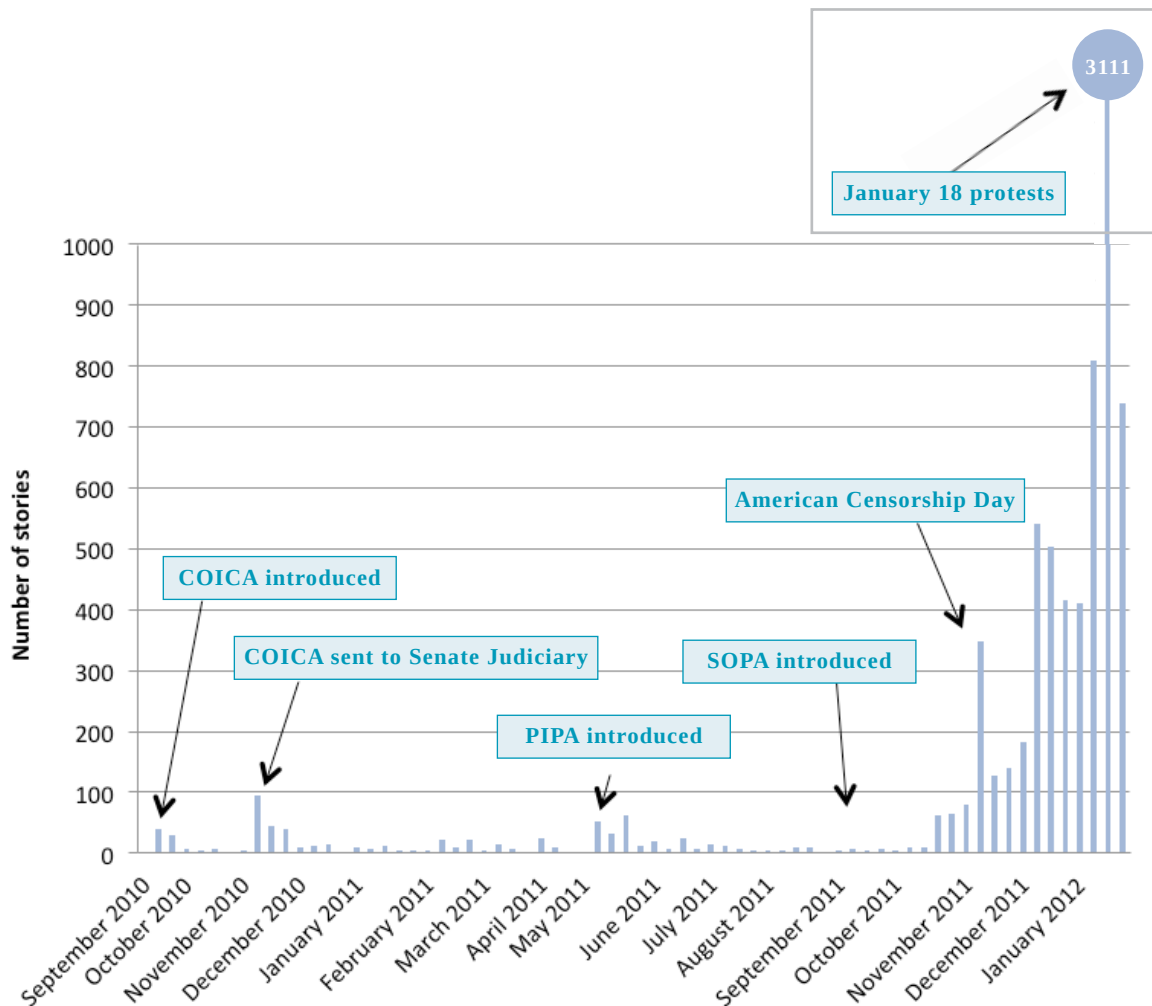


Figure 1: Number of stories published each week related to COICA, SOPA, or PIPA

This vast difference in the level of activity between the first 15 months and the period from mid-November to mid-January suggests that we observe two quite distinct phenomena: the first is how networked communications enable highly engaged actors to inform themselves and create and sustain a degree of political power in the absence of widespread mobilization; the second, covering the last two months, is the stage at which this long-term, gradually growing kernel of engaged actors and organizations creates, informs, directs, and engages with a surge of interest and mobilized activation on a much larger and broader scale.

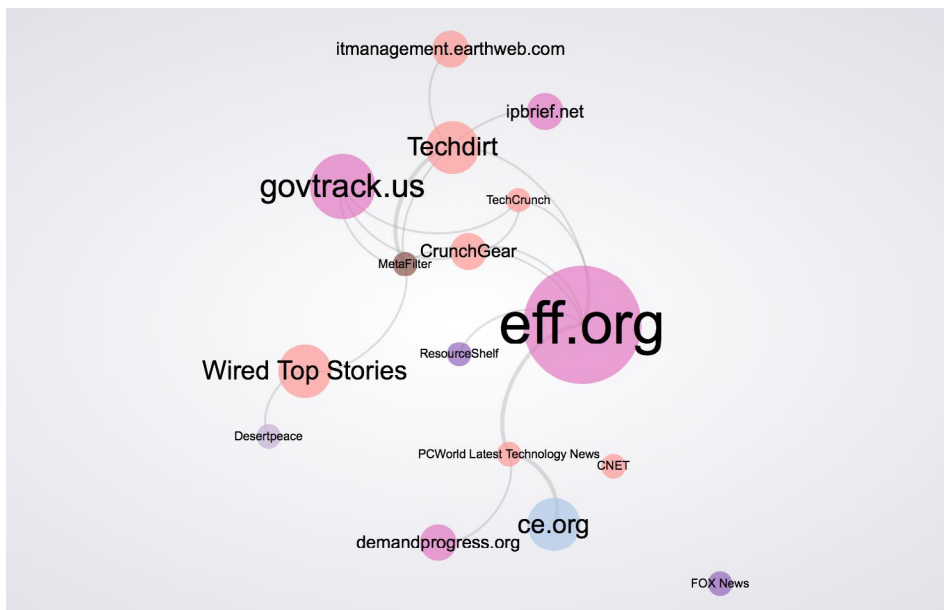


Figure 3: September 27–October 4, 2010

By the following week, as shown in Figure 3, action had shifted somewhat from tech media to NGOs, including the Electronic Frontier Foundation (EFF), GovTrack, and—albeit to a lesser extent at this stage—the left-leaning Demand Progress. We also see efforts by industry—in this case CE.org, the consumer electronics industry, opposing the bill. In this week, EFF played two roles that it sustained during the first wave of the controversy. First, it served as an information clearinghouse for events in the legislative arena. And second, it played a role that others like Public Knowledge or Techdirt would play in future iterations of the debate: it amplified peripheral voices and made them visible throughout the network engaged in the controversy. The network was still quite small at this stage; in the month of September, 30 different entities entered the debate.

In this week, acting as an amplifier, EFF posted a letter by 87 Internet engineers who wrote to the Senate Judiciary Committee that the legislation “will risk fragmenting the Internet’s global domain name system (DNS), create an environment of tremendous fear and uncertainty for technological innovation, and seriously harm the credibility of the United States in its role as a steward of key Internet infrastructure.”⁴⁹ By posting this letter on a site of major visibility, the EFF here, and others later, formed an attention backbone along which an otherwise peripheral intervention could travel to the attention of many more participants than the initial speakers could have reached given the visibility of the outlets directly under their control.

Demand Progress took a step into the debate in this period and would eventually take a leading role in channeling online outrage into political action by initiating a petition drive that collected over 300,000 total signatures. This model of translating online debate into Congress-focused communications would ultimately become a core force of the efforts to block the laws.

49 Peter Eckersley, “An Open Letter from Internet Engineers to the Senate Judiciary Committee,” *Deep Links* (blog), *Electronic Frontier Foundation*, September 28, 2010, <https://www.eff.org/deeplinks/2010/09/open-letter>.

As noted in the methods section, the size of a given node in the map is a reflection of the number of in-links to that media source during that time period and as such offers a measure of the relative importance of that site within the ecosystem of entities engaged with the topic. At this nascent stage in the debate, the ecosystem is still relatively sparse, and the number of links is small. *Wired* has six in-links during the week of September 20, and EFF was linked to seven times in the week of September 27. When the number of links are sparse in the network, as in these early weeks of the debate, sites that receive just a handful of links can appear as large nodes in the figures, and we caution against assigning too much importance to large nodes during these weeks without further investigation. We base our conclusions during these weeks with few links on a substantive review of every article published in this period to ensure that we do not assign more importance to a given node than is deserved.

Following a quiet October, there was a burst of activity in mid-November as the Senate Judiciary Committee considered and approved COICA (Figure 4). Several features were notable. First, while Fox News, the *Los Angeles Times*, and other mainstream media outlets began to take notice, online tech media, such as CNET and Techdirt, joined by TorrentFreak, continued to anchor the flow of information within the controversy. Second, the blogosphere on both sides of the political aisle emerged as a prominent actor in the debate.

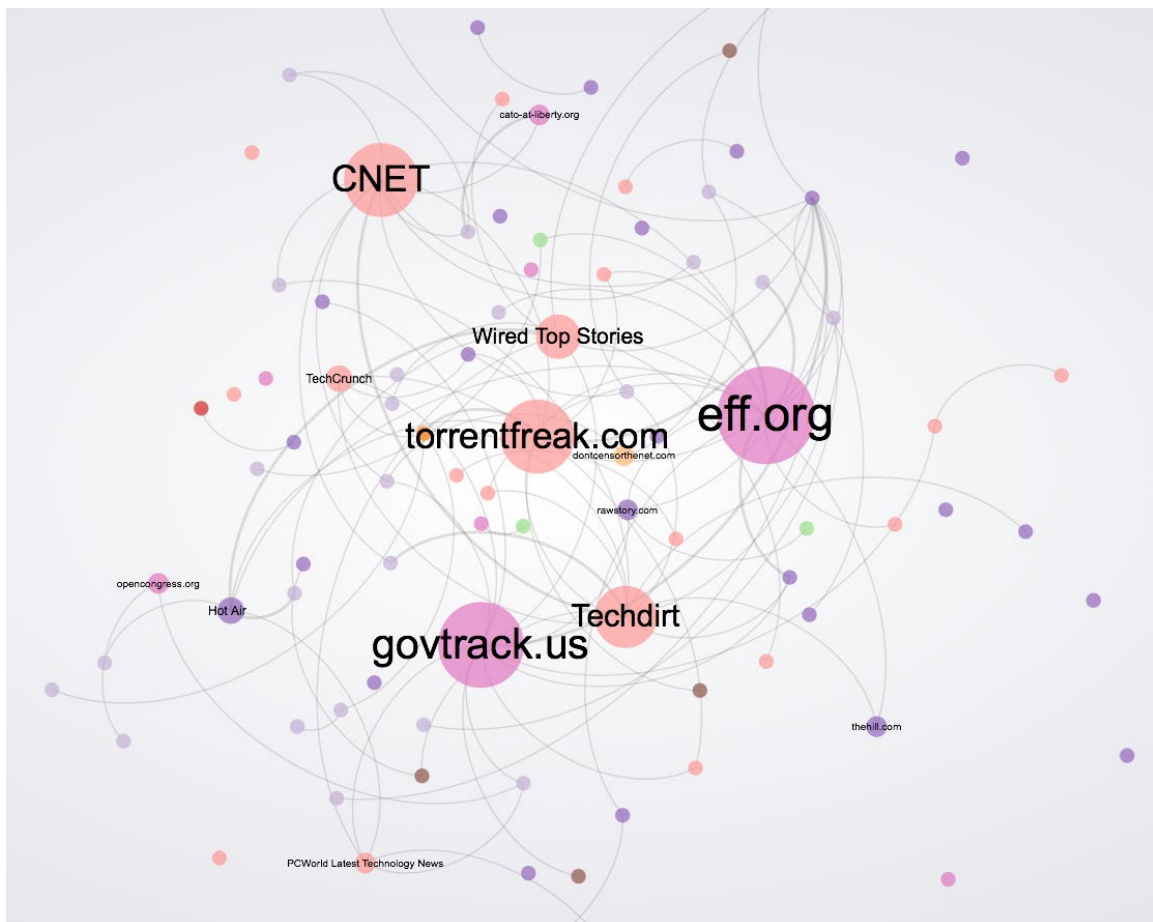


Figure 4: November 2010

The network of individuals and organizations that discussed COICA grew in November. By the end of the month, nearly 100 entities had entered the discussion, including 20 tech media sites, an equal number of general media organizations, and several dozen bloggers.⁵⁰ The node with the largest number of in-links is EFF with 15, followed by GovTrack with 13.

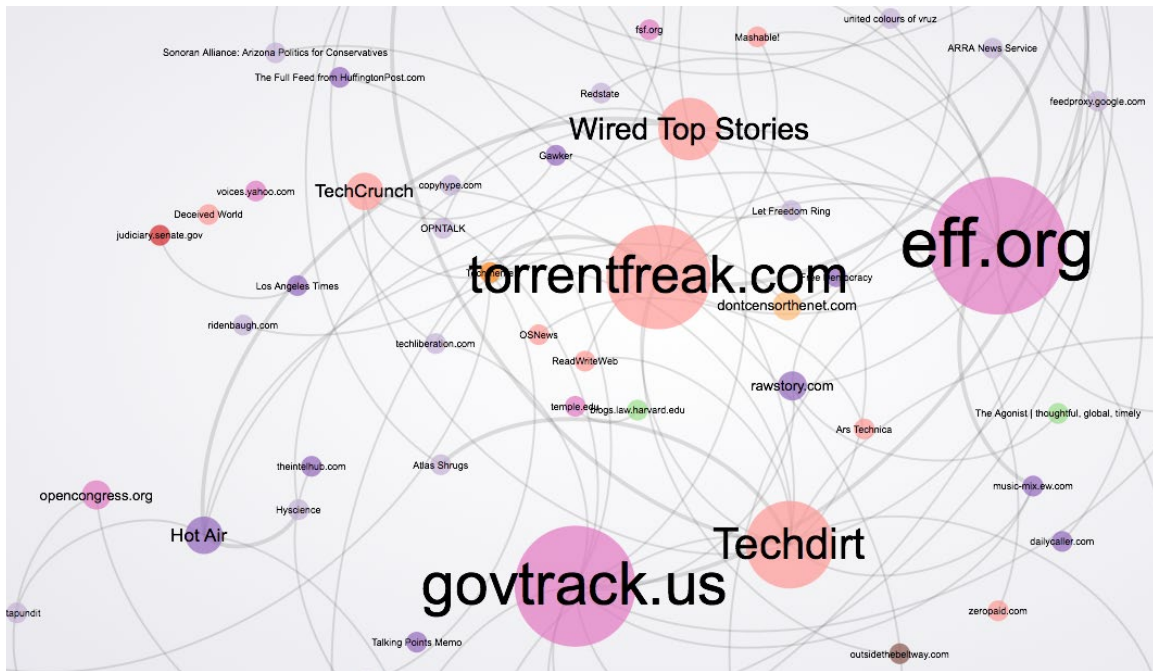


Figure 5: November 2010 (detailed window)

At this point, the right wing of the blogosphere began to take up the resistance to COICA, marking the emergence of the left-right online coalition that will continue through the controversy. This phenomenon is evident in the range of libertarian blogs and organizations that voiced opposition to the bill, such as the Cato Institute and Atlas Shrugs, and via the appearance of prominent right wing political blogs, such as Hot Air, Instapundit, and Red State. Individually, these blogs are not highly visible in the map as they received a limited number of in-links, but the very fact that this arm of the blogosphere entered the conversation is significant. The right-wing resistance was facilitated by Patrick Ruffini of Engage LLC, a consultancy that specializes in building online campaigns for the political right and which launched Don't Censor the Net, a special purpose online advocacy organization created to oppose the pending legislation. Epitomizing the cross-partisan nature of the opposition to COICA (and later, SOPA and PIPA), Ruffini continued to collaborate with David Moon and David Segal of the left-leaning Demand Progress throughout the campaign, although Demand Progress does not appear as an active force in this month. The month rounds out as Senator Wyden blocked lame duck passage of COICA in late November 2010, setting the stage for the reemergence of the controversy in the spring of 2011, when Senator Leahy introduced the successor bill, PROTECT-IP, or PIPA.

⁵⁰ The maps shown in this paper are sized to depict the full landscape of the discussion for a given time period, making it impractical to label each of the nodes. Interactive versions of all of the maps, which allow users to explore each of the nodes in detail, are available at: http://cyber.law.harvard.edu/research/mediacloud/2013/mapping_sopa_pipa/.

Phase II: PIPA

The second phase of the controversy began with the introduction of PIPA in the Senate on May 12, 2011. As seen in Figure 6, four sets of actors took center stage in the debate in this month, which featured 76 distinct media sources and over 150 stories. Congress played a significant role, represented by Senator Wyden, a vocal opponent of this bill, and by Senator Leahy, who introduced the bill on the Senate floor. Senator Wyden placed a hold on the bill on May 26, 2011, raising a general alarm that he might not be able to hold back the flood.⁵¹ Tech media led the media coverage, most notably Ars Technica, Techdirt, and CNET. Independent organizations offered subject-area leadership, including EFF on the West Coast and Public Knowledge and the Center for Democracy and Technology (CDT) on the East Coast. Don't Censor the Net offered an online petition against PIPA, paralleling the role that Demand Progress played in the fall of 2010.⁵²

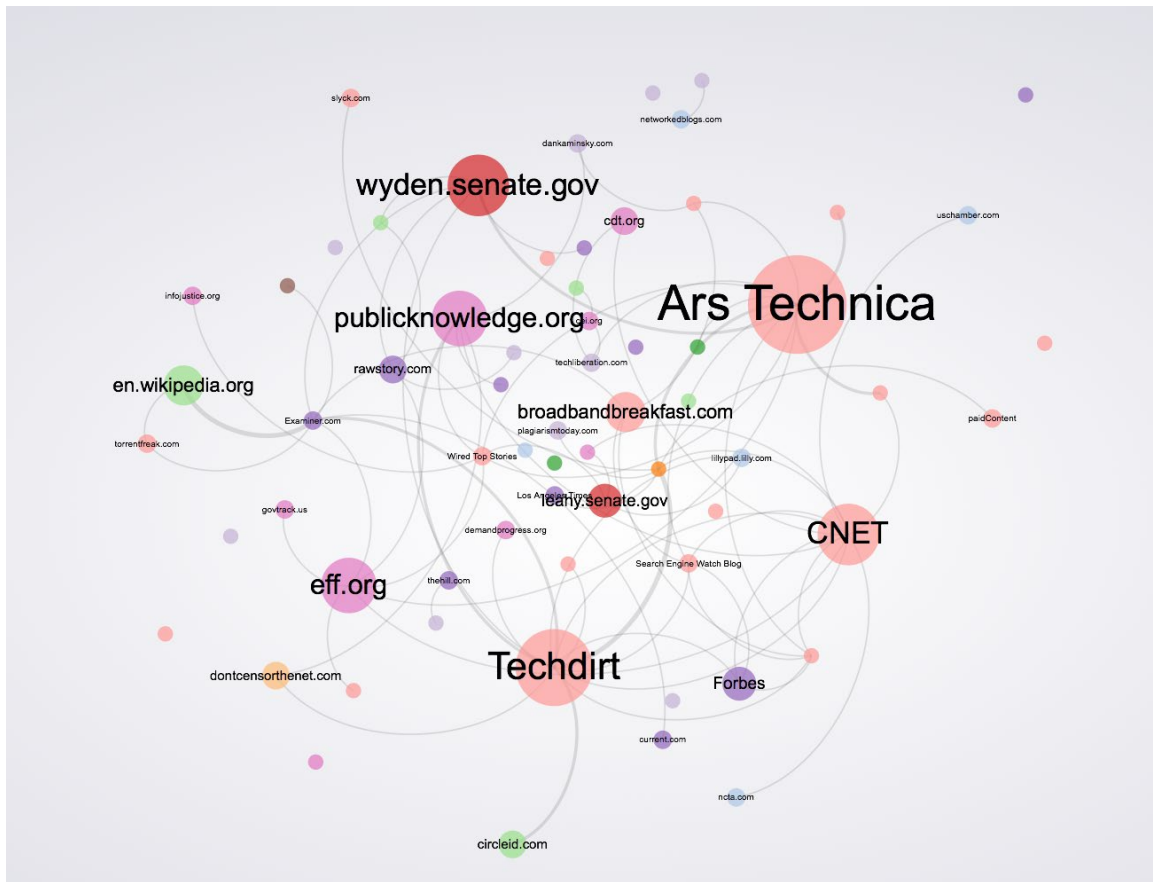


Figure 6: May 2011

51 "Wyden Places Hold on Protect IP Act," Senator Ron Wyden press release, May 26, 2011, on the senator's website, <http://www.wyden.senate.gov/news/press-releases/wyden-places-hold-on-protect-ip-act>.

52 "Full Text of the PROTECT IP Act of 2011," Don't Censor the Net, <http://www.dontcensorthenet.com/full-text-of-the-protect-ip-act-of-2011>.

Alongside these recurring nodes, the three most important “newcomers” to the maps during these weeks were Wikipedia, Fight for the Future (FftF), and OpenCongress. At this stage, Wikipedia played a purely informational role. The protest that would emerge almost three months later had not yet been hatched, and the links into Wikipedia were to the articles that provided descriptions of SOPA and PIPA, not to mobilization or talk pages.⁵⁹ FftF and OpenCongress, both co-founded by Tiffiny Cheng and Holmes Wilson, began to play a central role in the controversy that would continue until the end of the campaign. In this week, FftF launched the first popular use of video in the anti-SOPA-PIPA campaign, presenting a video that was in part informational and in part polemical and that urged viewers to take action by contacting legislators.⁶⁰ FftF also provided a platform for artists and musicians to sign an open letter to Congress.⁶¹ OpenCongress offered a complementary model of access to the written materials on the act, again with a point of contact and an option to “vote” publicly on the bill.⁶² The Screen Actors Guild (sag.org) became active in this period, offering a joint statement of industry member in support of the act.⁶³ This week marked one of the only periods in the debate when pro-SOPA perspectives received in-links comparable to those garnered by opposition voices, receiving the same number of in-links as EFF’s Wired for Change action alert, although each only receives three in-links.⁶⁴

The following week, October 31 to November 7, is notable as the week that appears closest to the prevailing conception of the blogosphere or networked public sphere (Figure 9). The visibility of major online media sources, including the Huffington Post, Politico, and Daily Kos, and traditional media such as the *Los Angeles Times*, the *San Francisco Chronicle*, and the *Hill*, was comparable to the more prominent tech media and independent organizations. In addition, Free Press (savetheinternet.com and freepress.net) and the Screen Actors Guild (SAG) appeared. This week also exhibited the potential for mobilization via established institutional channels in addition to purpose-built campaign sites. For example, a White House blog petition, Stop the E-Parasite Act, appeared as a result of links from Mashable to the petition page.⁶⁵ The White House response to the petition would come in January 2012.

59 Wikipedia contributors, “Stop Online Piracy Act,” *Wikipedia, The Free Encyclopedia*, http://en.wikipedia.org/wiki/Stop_Online_Piracy_Act; Wikipedia contributors, “PROTECT IP Act,” *Wikipedia, The Free Encyclopedia*, http://en.wikipedia.org/wiki/Protect_IP_Act.

60 “PROTECT IP / SOPA Breaks the Internet,” Fight for the Future, <http://fightforthefuture.org/pipa/>.

61 “Artists opposing the PROTECT-IP / SOPA Act,” Fight for the Future, <http://www.fightforthefuture.org/pipa/artists>.

62 “S.968 – PIPA,” OpenCongress, <http://www.opencongress.org/bill/112-s968>.

63 “Joint Statement from SAG, AFM, AFTRA, DGA, IATSE, and IBT Regarding Stop Online Piracy Act (HR 3261),” October 26, 2011, <http://www.sagaftra.org/joint-statement-sag-afm-aftra-dga-iatse-and-ibt-regarding-stop-online-piracy-act-hr-3261>.

64 “Strike Against Censorship,” Electronic Frontier Foundation, https://wfc2.wiredforchange.com/o/9042/p/dia/action/public/?action_KEY=8173.

65 “Stop the E-PARASITE Act,” We the People, <https://petitions.whitehouse.gov/petition/stop-e-parasite-act/SWBYXX55>.

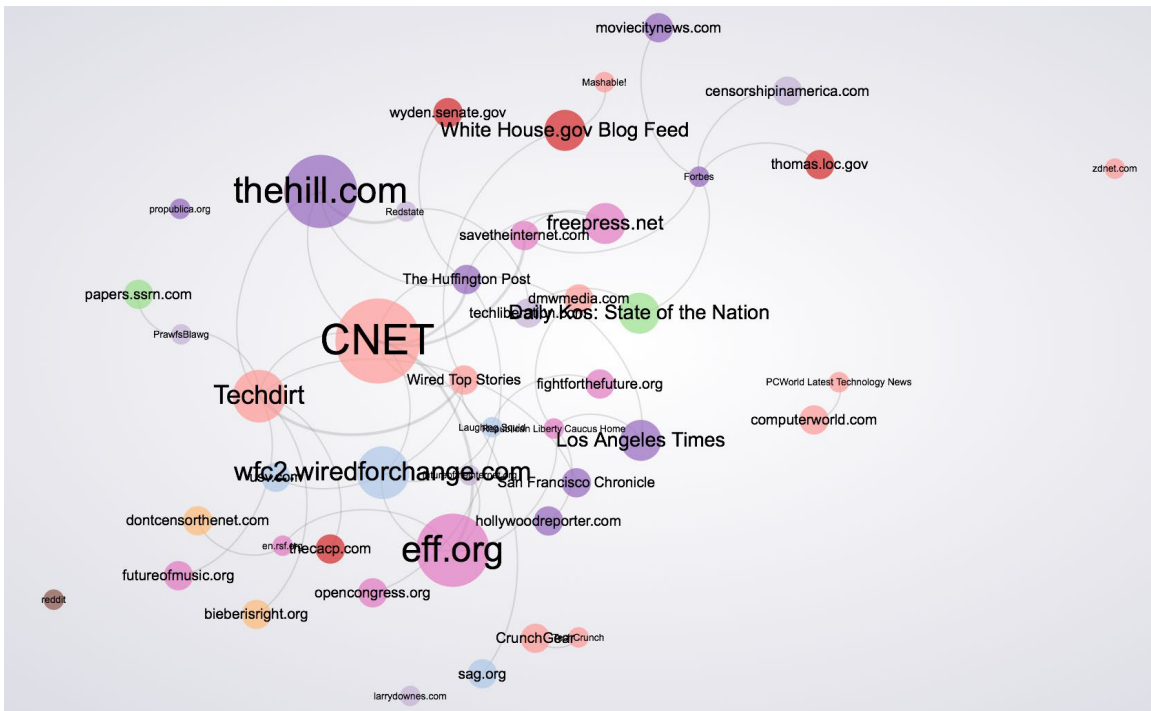


Figure 9: October 31–November 7, 2011

The following two weeks (Figures 10 and 11) showed a marked increase in online activism against SOPA and PIPA, culminating in American Censorship Day on November 16, 2011, an initiative that encouraged users or Web site owners to “censor” their own sites and pointed users to a contact page for their congressional representatives. The number of entities posting content related to SOPA and PIPA hovered at around 40 for the three previous weeks; more than 160 contributed stories in the third week of November (Figure 11). American Censorship became a center of activity in this week as well, receiving 25 in-links. It would maintain this prominent position through the end of the controversy as the site, which was initiated by the co-founders of Fight for the Future, Participatory Politics Foundation, and Demand Progress with Public Knowledge, the EFF, and support from the Mozilla Foundation, became a major point for coalescence on action, and ultimately the model for the January 18, 2012, protests.

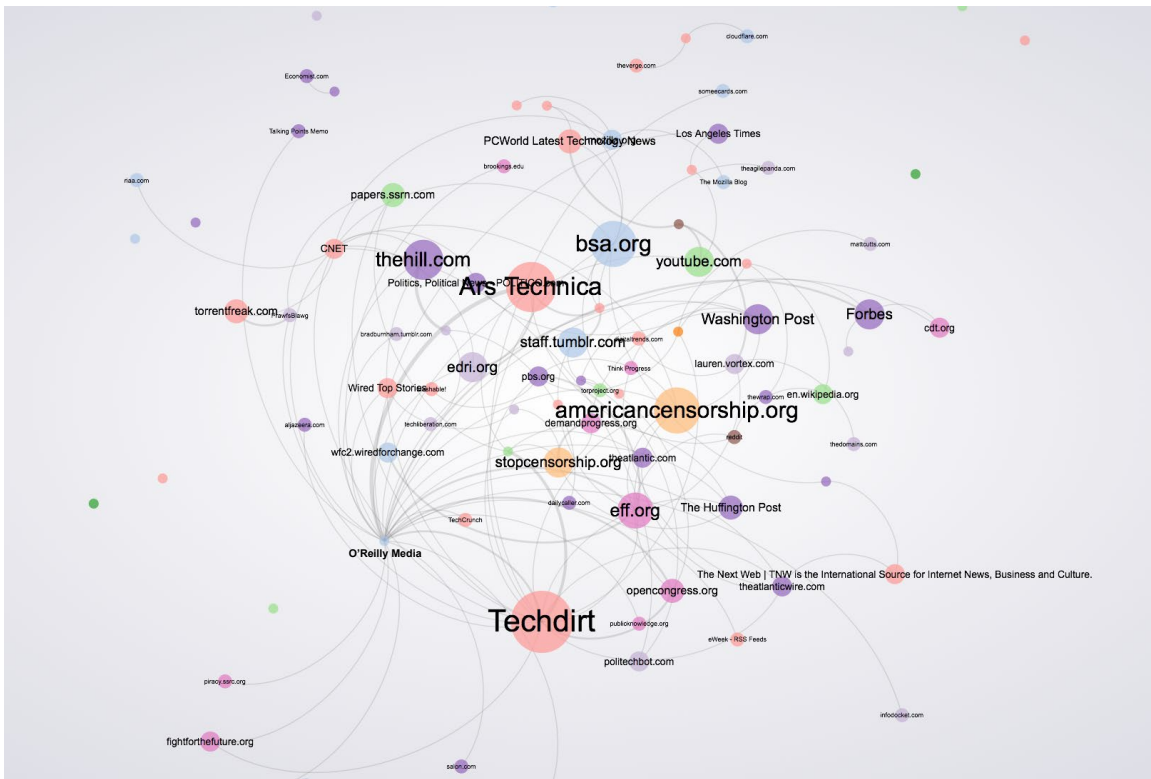


Figure 13: November 21–28, 2011

In the fourth week of November, which featured over 100 distinct media sources, Techdirt and Ars Technica, as well as American Censorship, carried forward with the momentum of the prior week (Figure 13). A notable new dimension of the controversy occurred as the Business Software Alliance (BSA) received much attention when it announced on November 21 that it had reversed its position and would oppose SOPA. Several research papers appeared during this period, including Allan Friedman’s analysis of the effects of SOPA and PIPA on cybersecurity, published on [brookings.edu](http://www.brookings.edu/research/papers/2011/11/15-cybersecurity-friedman);⁷¹ Derek Bambauer’s paper on the Social Science Research Network (SSRN);⁷² and survey results of research conducted by Joe Karaganis at the Social Sciences Research Council (SSRC) suggesting that the practices targeted by SOPA and PIPA are rare and that public opinion supports a certain level of “copy culture.”⁷³ In a long blog post that described the state of the debate at the time, Alex Howard of O’Reilly Media linked to more than two dozen other stories and thereby occupied a central location in the network map.⁷⁴

In December 2011 (Figures 14 and 15), some of the action shifted back to DC as the House Judiciary Committee held its first markup session on SOPA on December 15 and Representative

71 Allan Friedman, “Cybersecurity in the Balance: Weighing the Risks of the PROTECT IP Act and the Stop Online Piracy Act,” Brookings Institution, November 15, 2011, <http://www.brookings.edu/research/papers/2011/11/15-cybersecurity-friedman>.

72 Derek Baumbauer, “Orwell’s Armchair,” 79 U. Chi. L. Rev. 863 (2012).

73 Joe Karaganis, “The Copy Culture Survey: Infringement and Enforcement in the US,” November 15, 2011, <http://piracy.americanassembly.org/the-copy-culture-survey-infringement-and-enforcement-in-the-us/>.

74 Alex Howard, “Congress Considers Anti-Piracy Bills that Could Cripple Internet Industries,” *O’Reilly Radar*, November 22, 2011, <http://radar.oreilly.com/2011/11/sopa-protectip.html#acta>.

Darrell Issa and Senator Wyden introduced the Online Protection and Enforcement of Digital Trade (OPEN) Act in both the House and the Senate as an alternative to SOPA and PIPA. Along with these developments, the Sunlight Foundation offered a multipart review of contributions to members of Congress that highlighted the ties between SOPA co-sponsors and legacy media entities.⁷⁵ This period also revealed internal partisan divisions as the right wing tried to reconcile its members who supported SOPA with those who opposed it; the National Review published an editorial entitled “Defending SOPA” as the conservative-backed group Don’t Censor the Net continued to advocate against the bill.⁷⁶

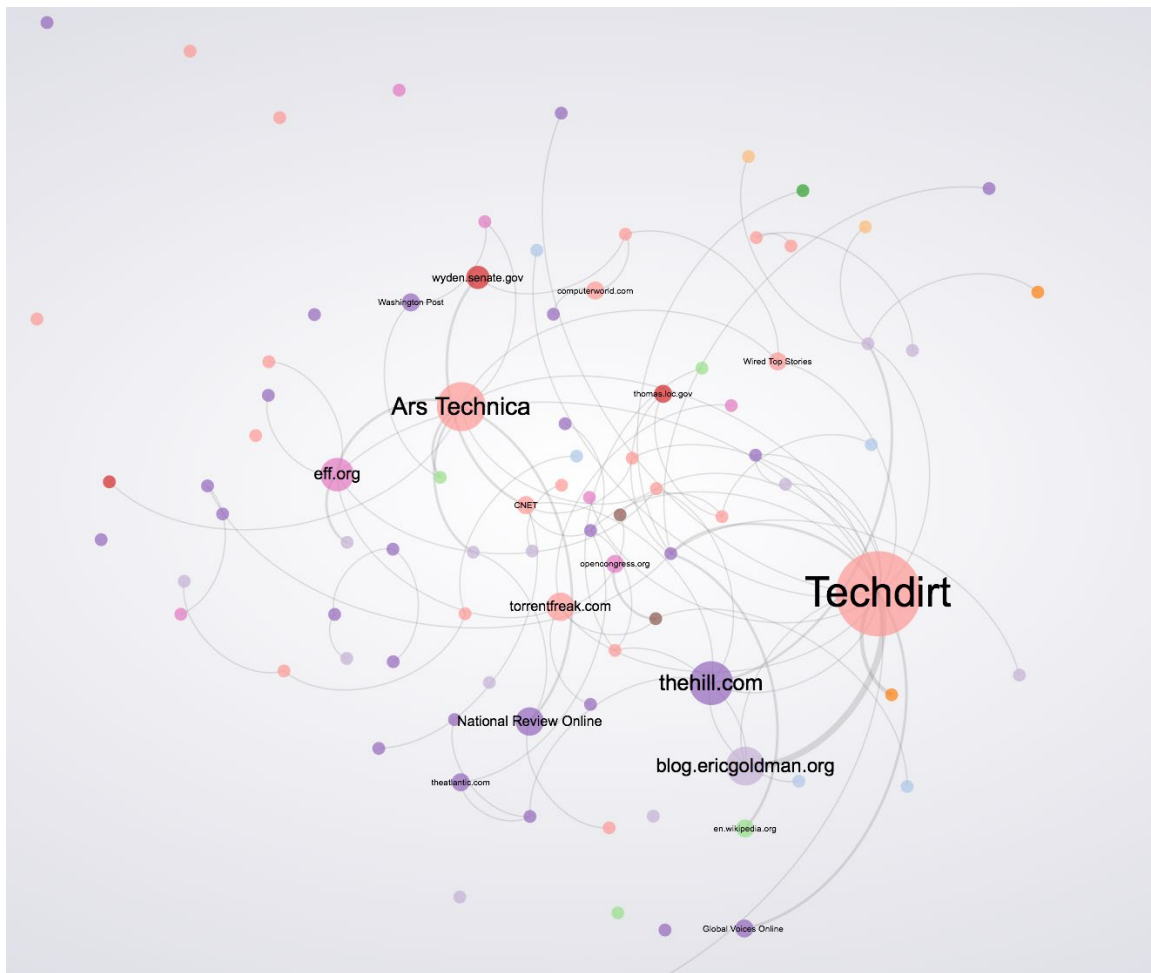


Figure 14: December 5–12, 2011

75 Tessa Muggeridge, “Legacy Media Bankrolling Campaigns of SOPA Cosponsors,” *Sunlight Foundation Reporting Group Blog*, Sunlight Foundation, December 1, 2011, <http://reporting.sunlightfoundation.com/2011/legacy-media-bankrolling-campaigns-of-sopa-cosponsors/>.

76 Lamar Smith, “Defending SOPA,” *National Review*, December 1, 2011, <http://www.nationalreview.com/articles/284535/defending-sopa-lamar-smith>.

In addition, the voices of individual law professors played a notable role in this period; in the week of December 5, 2011, Eric Goldman's blog received attention for his posts addressing recent copyright infringement litigation,⁷⁷ assessing the newly proposed OPEN Act,⁷⁸ providing a compilation of his past anti-SOPA posts,⁷⁹ and explaining why he opposed SOPA.⁸⁰

After the week of December 5, which included approximately 180 stories published by 100 discrete media sources, there was a marked increase in activity during the latter half of December. Each week, between 150 and 200 media sources weighed in on the debate, publishing over 500 stories per week.

On December 21, 2011, the House Judiciary Committee released a long list of corporate supporters of SOPA, hoping to bolster the claim that SOPA was good for business and innovation. The difference between the map before (Figure 15) and after that event (Figure 16) is stark, and provides one of the clearest examples we have of the dynamism of the networked public sphere and the possibility of converting discourse into action. It marks a major online mobilization to boycott registrar Go Daddy for its support of SOPA and PIPA, an instance of activism initiated by a single Reddit user on December 22.⁸¹ Go Daddy quickly retreated and abandoned its support for the legislation as consumers began to change the domain name registrar they used.⁸²

77 Eric Goldman, "Ad Network Avoids Contributory Copyright Infringement for Serving Ads to a Rogue Website--Elsevier v. Chitika," *Technology & Marketing Law Blog*, December 5, 2011, http://blog.ericgoldman.org/archives/2011/12/ad_network_didn.htm.

78 Eric Goldman, "The OPEN Act: Significantly Flawed but More Salvageable than SOPA/PROTECT-IP," *Technology & Marketing Law Blog*, December 10, 2011, http://blog.ericgoldman.org/archives/2011/12/the_open_act_de.htm.

79 Eric Goldman, "I Don't Heart SOPA or PROTECT-IP: A Linkwrap," *Technology & Marketing Law Blog*, December 7, 2011, http://blog.ericgoldman.org/archives/2011/12/linkwrap_on_sopa.htm.

80 Eric Goldman, "Why I Oppose the Stop Online Piracy Act (SOPA)/E-Parasites Act," *Technology & Marketing Law Blog*, November 15, 2011, http://blog.ericgoldman.org/archives/2011/11/stop_online_pir.htm.

81 "GoDaddy Supports SOPA, I'm Transferring 51 Domains & Suggesting a Move Your Domain Day," December 22, 2011, http://www.reddit.com/r/politics/comments/nmnie/godaddy_supports_sopa_im_transferring_51_domains/.

82 "Go Daddy's Position on SOPA," Go Daddy, <http://support.godaddy.com/godaddy/go-daddys-position-on-sopa/?isc=smtwsup>; "Go Daddy No Longer Supports SOPA," Go Daddy press release, December 23, 2011, on the Go Daddy website, <http://www.godaddy.com/news/article/go-daddy-no-longer-supports-sopa.aspx>.

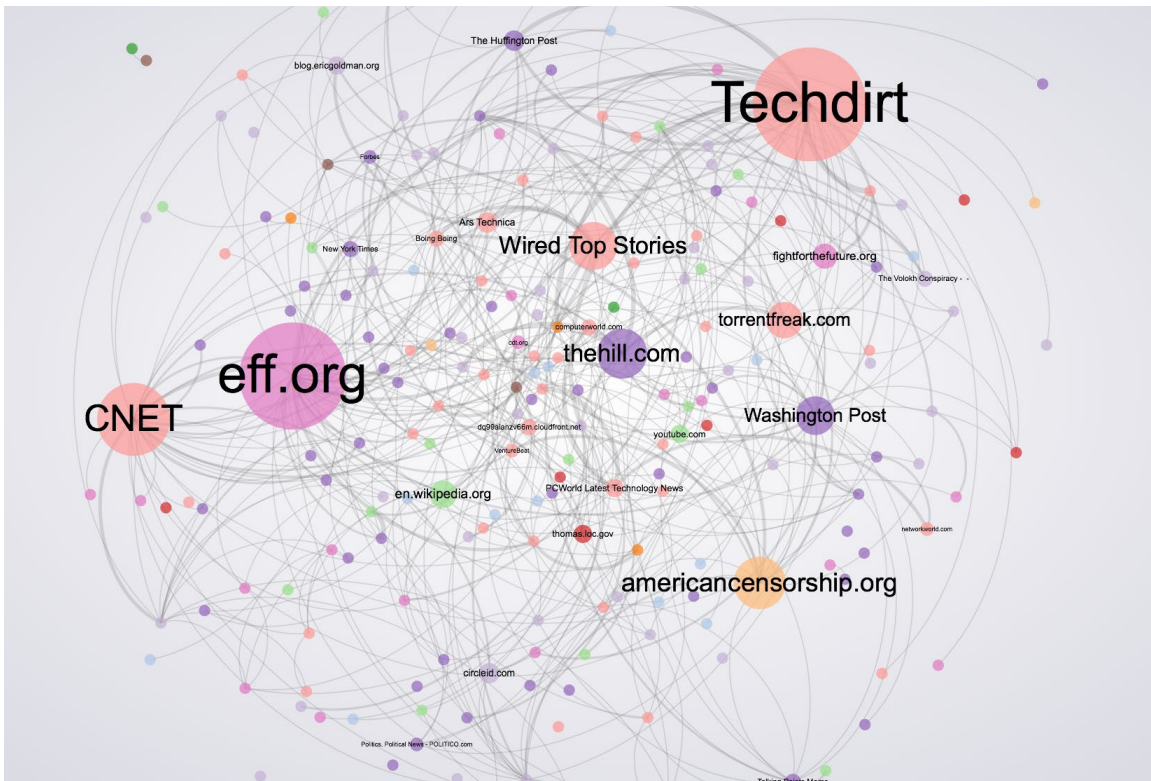


Figure 15: December 12–19, 2011

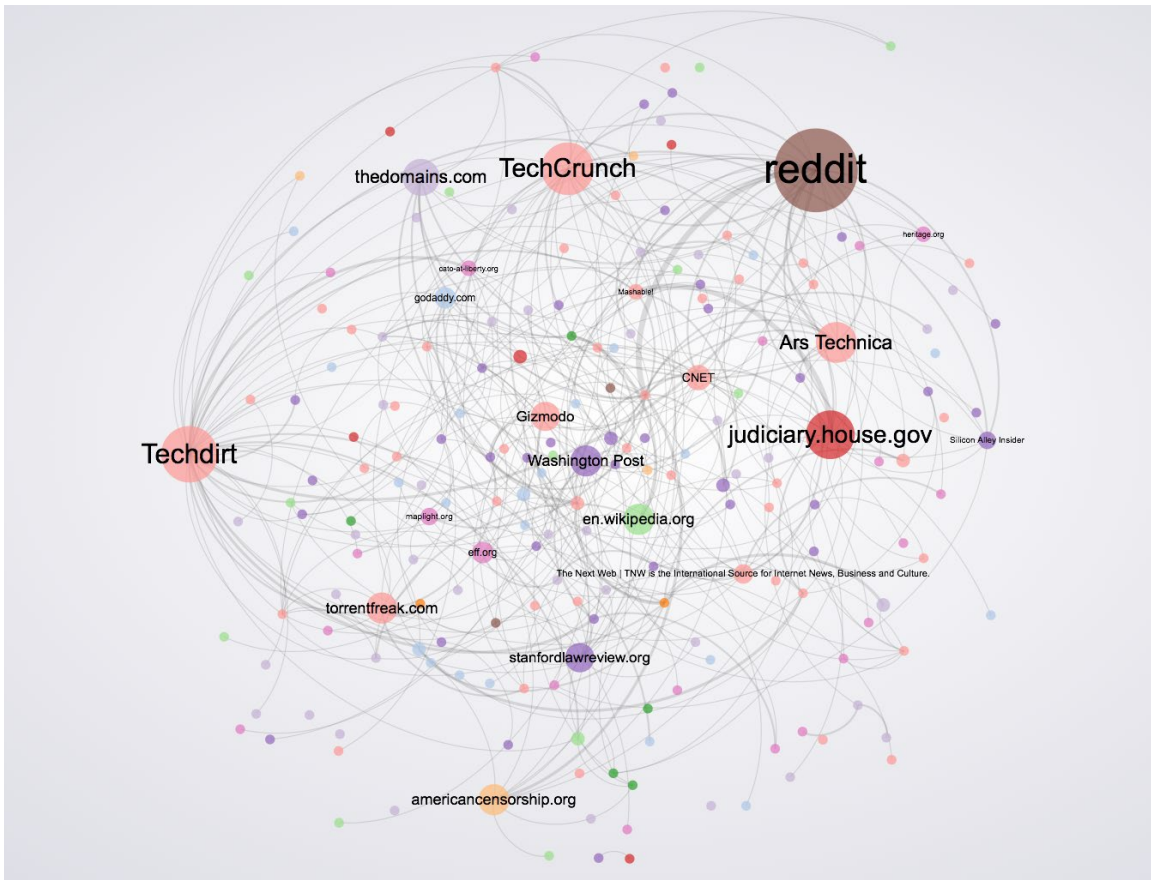


Figure 16: December 19–26, 2011

By the beginning of January 2012 (Figure 18), the action came back to the center, spearheaded by many of the now-familiar nodes in the debate, including Techdirt, CNET, TorrentFreak, EFF, and American Censorship, which stood out as the leading site campaigning against the act. Many continued to link to the list of SOPA supporters at the House Judiciary site⁸⁷ and to the Wikipedia pages describing SOPA and PIPA.

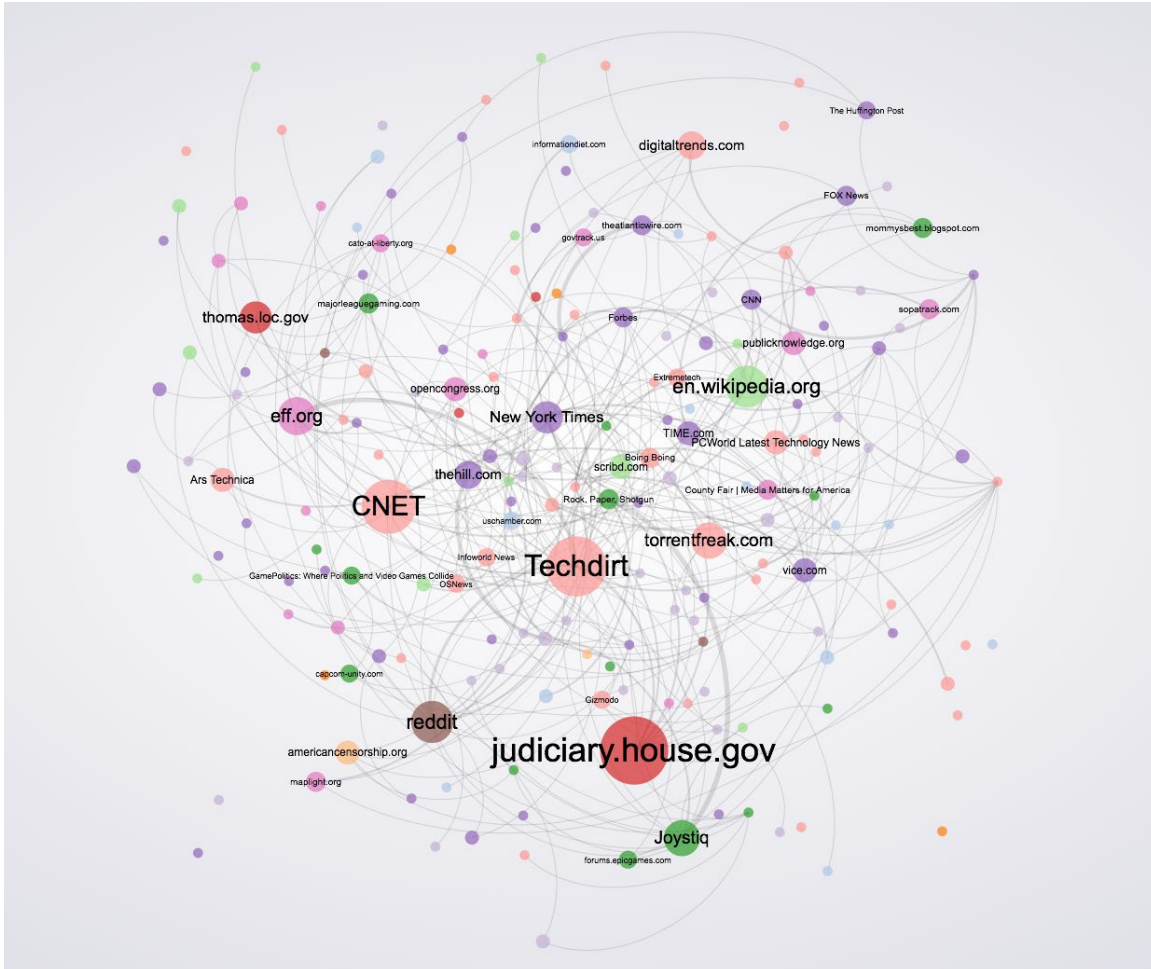


Figure 18: January 2–9, 2012

⁸⁷ This link is not provided because it has been updated and no longer reflects the same content that was significant in January 2012.

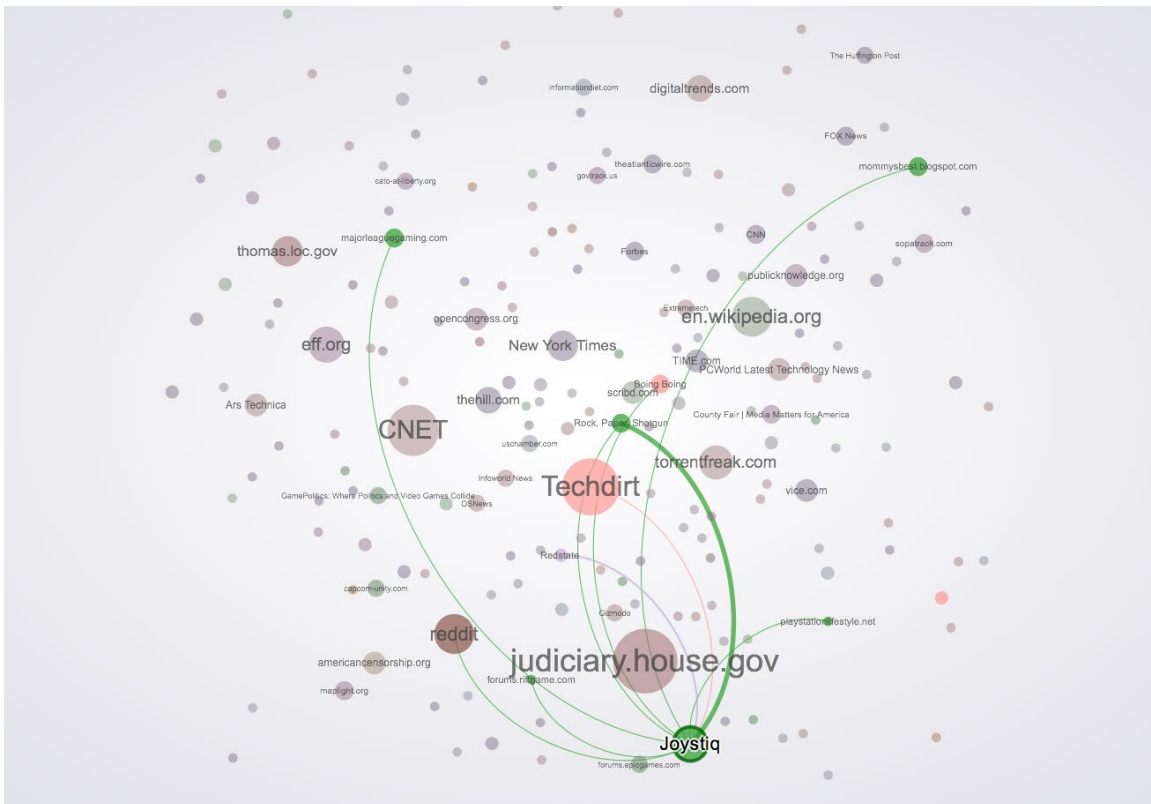


Figure 19: January 2–9, 2012 (detailed window, highlighting Joystiq and other gaming sites)

Along with the ongoing substantial role of these recurring actors, several new dynamics emerged. For instance, this week also featured increased activity by gaming sites such as Joystiq, mommy'sbest.blogspot.com, majorleaguegaming.com, and Rock, Paper, Shotgun, among others (Figure 19). Gamers played a distinct role at this stage, acting as a discrete group of individuals and organizations that mobilized against the laws. In particular, Joystiq's actions epitomize this dynamic: the site pointed to a post by game developer Nathan Fouts at Mommy's Best Games in which he encouraged game developers to determine their bosses' stance on SOPA and urged the gaming community to lobby the Entertainment Software Association (ESA) to cease its support of SOPA.⁸⁸ Building from successful earlier activation and protest, this mobilization effort included a clear call to action, complete with draft email text and email contact information for personnel at the ESA.⁸⁹

This week also showcased the potential power of an attention backbone as it called attention to the contributions of smaller sites via links from more trafficked sites. An often-repeated meme from the earliest stages of the debate had been that the cost of piracy to the United States is \$58 billion each year. This meme was repeated across many stories, including across mainstream

88 JC Fletcher, "Mommy's Best Games Asks ESA Members to Renounce SOPA," *Joystiq*, January 3, 2012, <http://www.joystiq.com/2012/01/03/mommys-best-games-asks-esa-members-to-renounce-sopa/>.

89 Nathan Fouts, "Convince the ESA to Drop SOPA Support," *Mommy's Best Devlog* (blog), January 2, 2012, <http://www.mommysbest.blogspot.com/2012/01/convince-esa-to-drop-sopa-support.html>.

media outlets.⁹⁰ On January 3, 2012, Julian Sanchez published a piece in *Cato @ Liberty* in which he dismantled this \$58 billion figure.⁹¹ Although few actors linked directly to Cato, and the Cato site appeared as a small node in this week, the story itself received attention due to an attention backbone, as Techdirt and EFF directed readers to Cato, and Reddit⁹² and others directed users to the Techdirt article⁹³ in conjunction with a link to the Sanchez article.

Activity begins to ramp up rapidly the week of January 9, 2012, with a doubling of the number of stories from the prior week, increasing from about 400 to over 800 (Figure 20). Reddit occupied the most prominent node in the network with links to a January 10 call to action by its admins (“Stopped they must be; on this all depends”) that announced their decision to black out Reddit on January 18, 2012, and included links to the various campaign sites. The American Censorship petition site continued to grow in size this week.

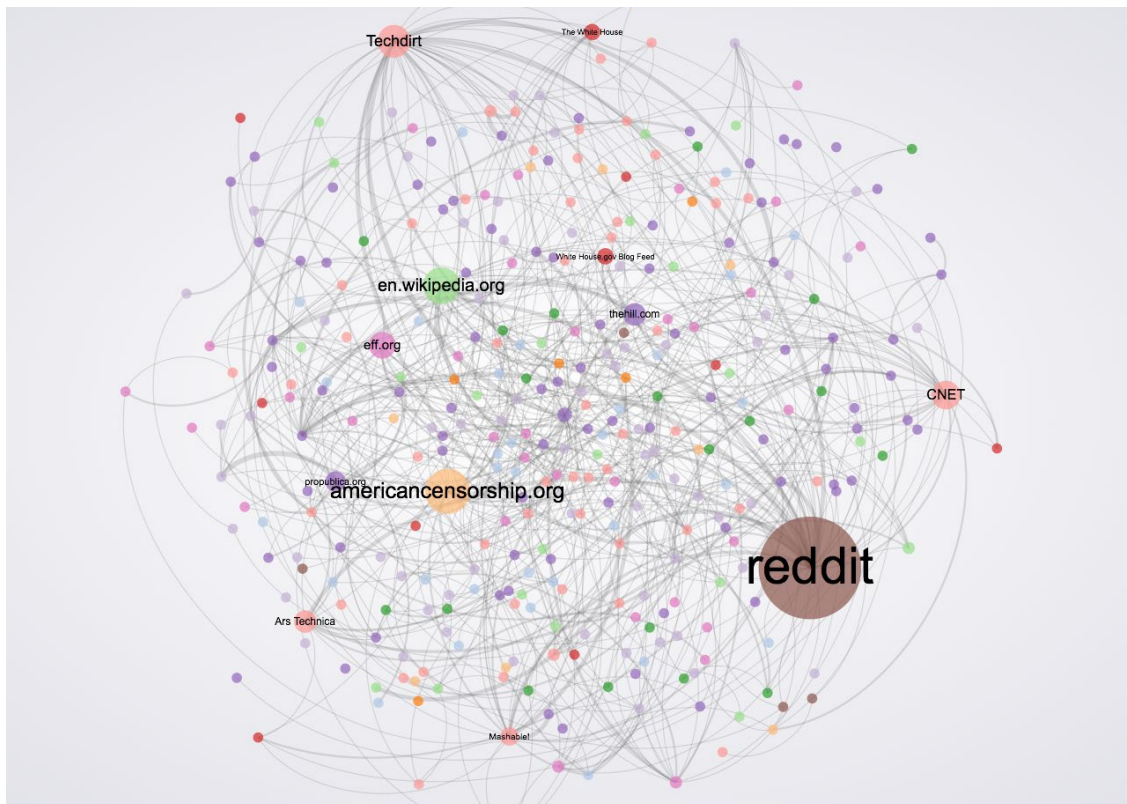


Figure 20: January 9–16, 2012

90 See, for example, David Carr’s otherwise very critical piece on SOPA in the *New York Times*: David Carr, “The Danger of an Attack on Piracy Online,” *New York Times*, January 1, 2012, <http://www.nytimes.com/2012/01/02/business/media/the-danger-of-an-attack-on-piracy-online.html>.

91 Julian Sanchez, “How Copyright Industries Con Congress,” *Cato At Liberty* (blog), January 3, 2012, <http://www.cato-at-liberty.org/how-copyright-industries-con-congress/>.

92 “Cato Institute Digs into MPAA’s own Research to Show that SOPA Wouldn’t Save a Single Net Job,” January 4, 2012, http://www.reddit.com/r/movies/comments/o2wx3/cato_institute_digs_into_mpaas_own_research_to/.

93 Mike Masnick, “Cato Institute Digs into MPAA’s Own Research to Show that SOPA Wouldn’t Save a Single Net Job,” *Overhype* (blog), *Techdirt*, January 4, 2012, <http://www.techdirt.com/articles/20120104/04545217274/cato-institute-digs-into-mpaas-own-research-to-show-that-sopa-wouldnt-save-single-net-job.shtml>.

By this point, Wikipedia served as both a source of information and a platform for action. Links to the Wikipedia pages describing SOPA and PIPA continued to be popular along with links to the talk pages on which Wikipedia contributors, led by Wikipedia founder Jimbo Wales, discussed the initiatives and debated the role that the online encyclopedia should play; these talk pages included deliberation as to whether SOPA is so dangerous to the open Internet that Wikipedia should shut down in protest for a day.⁹⁴ The debate within the Wikipedia community, including over 2,000 participants in the decision, was itself a fascinating instance of direct democracy—in this case, within the community of contributors to one of the world’s most visited Web sites.⁹⁵

The White House garnered attention this week with a blog entry entitled “Combating Online Piracy while Protecting an Open and Innovative Internet,”⁹⁶ in which senior White House officials responded to the online “We the People” petitions against SOPA and related bills and announced that the White House, while sensitive to the problems posed by online piracy, did not support the proposed legislation.⁹⁷ The petition drive had garnered over 50,000 signatures using the petition.whitehouse.gov platform, prompting the official White House response. Tech media sites, including Techdirt, CNET, and Ars Technica, continued to be key sources, along with the nonprofit advocates.

The following week (Figure 21) marked the culmination of the debate, with an explosion of action and attention on January 18, 2012, when thousands of sites were blacked out including Wikipedia, Reddit, and BoingBoing. While Google’s landing page remained operable, it offered a link to its “End piracy, not liberty” petition page. Millions of people signed on. In the wake of this massive outpouring of opposition, both the House and Senate versions of the bill were shelved.

94 “User talk:Jimbo Wales,” http://en.wikipedia.org/wiki/User_talk:Jimbo_Wales.

95 Ayelet Oz, “Legitimacy and Efficacy: The Blackout of Wikipedia,” *First Monday* 17, no. 12 (2012), <http://firstmonday.org/ojs/index.php/fm/article/view/4043>.

96 Macon Phillips, “Obama Administration Responds to We the People Petitions on SOPA and Online Piracy,” *White House Blog*, January 14, 2012, <http://www.whitehouse.gov/blog/2012/01/14/obama-administration-responds-we-people-petitions-sopa-and-online-piracy>.

97 “Veto the SOPA Bill and Any Other Future Bills that Threaten to Diminish the Free Flow of Information,” We the People, December 18, 2011, <https://petitions.whitehouse.gov/petition/veto-sopa-bill-and-any-other-future-bills-threaten-diminish-free-flow-information/g3W1BscR>.

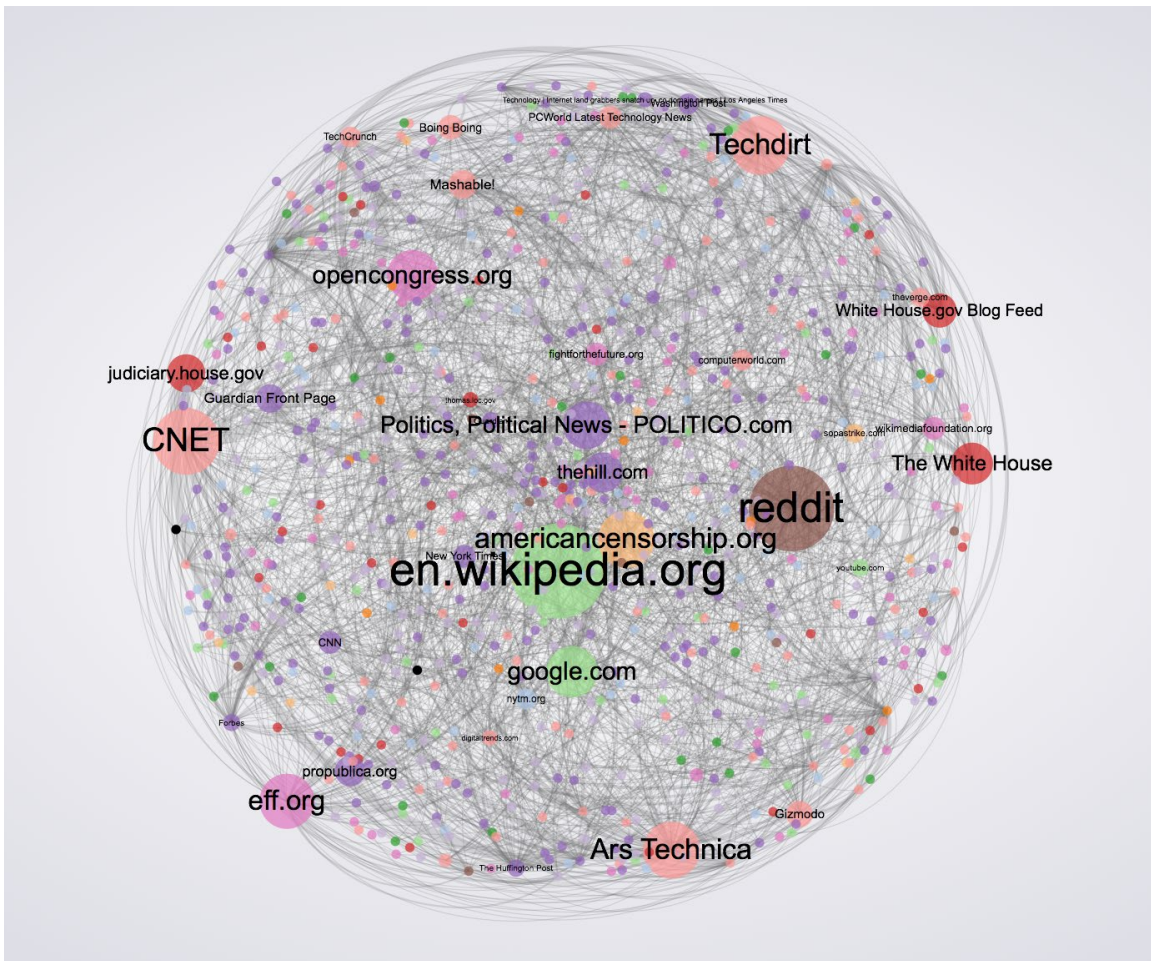


Figure 21: January 16–23, 2012

The week of January 16–23 alone saw over 3,000 stories on SOPA and PIPA, about 40% of the total number of stories from the introduction of SOPA to its defeat, and about one-third of all stories from the 17-month period we studied. In this particular moment of massive mobilization, several of the longstanding core nodes remained visible: Techdirt, CNET, and Ars Technica; Wikipedia; the EFF; and OpenCongress and Fight for the Future.

Along with the significant role played by Wikipedia, Google, and Reddit in this portion of the debate, it is important to note the remarkable scale and diversity of voices present this week. Mainstream media occupied important positions alongside individual blogs, including experts who spoke out and rose to positions of widespread public attention via an attention backbone. ProPublica played an interesting role as a member of the networked press by compiling and hosting a list of members of the House and Senate, identifying supporters and opponents of the bills based on their public statements, and making the list sortable by party and, most interestingly, by levels of contribution received by the competing corporate sponsors and opponents of SOPA and PIPA.⁹⁸

⁹⁸ See Dan Nguyen, “SOPA Opera,” *ProPublica*, <http://projects.propublica.org/sopa/>.

Discussion and Key Lessons

The major flip in support in the House and Senate between January 18 and 19, 2012, clearly indicates that the protest of January 18 closed the deal.⁹⁹ Following the defeat of SOPA and PIPA, two conflicting narratives developed to describe the events. The politics-as-usual narrative interpreted the events as “Google and Facebook have come to town;” the new major industry players had become new players in the same old lobbying game. The more radical narrative was that the networked public sphere had come into its own; the events reflected a new model of political organization and democratic participation. The game itself had changed, not merely its players.

The debate and subsequent mobilization looked very different at various points in time. Our data suggest that the events unfolded in three distinct stages. The first stage of the controversy, which took place over a period of more than a year, involved a relatively small number of individuals and organizations in an online debate of modest proportions. The principal participants included tech media and independent organizations, joined by general media, private organizations, targeted campaigns, individuals, and bloggers. The second stage saw the entry of larger players such as the online communities at Reddit and Wikipedia along with Google, Mozilla, and other technology companies. This second stage started to ramp up in mid-November 2011 and lasted until January 2012. The third stage was marked by the engagement of millions of individuals in the week of January 18. Throughout this period we see a highly committed group of actors that engaged early in the debate and continued to play a leadership role throughout the controversy. Although it is impossible to clearly establish the degree of influence, the entry of larger players in the second stage, who in turn were able to reach a national audience, is likely to stem in part from the efforts and persistence of the core actors during the first stage. These core actors developed the frames that were used to engage the larger public and helped to organize and reveal the broadly manifest cross-sectoral opposition to the legislation, thereby changing the calculus of legislators. A potentially productive area for future research is the degree of commonality and variation in the character and evolution of movements that have been facilitated through digital communication and organizing.

Nothing in our analysis suggests that the networked public sphere is immune to influence and manipulation by powerful special interests. In this story, the role of Google and other technology companies was seen as benign by civil society participants opposed to SOPA-PIPA. Certainly, the narrative that arises from this case study is one of broad-based public sentiment coalescing around discussion and activity in the networked public sphere and delivering a decisive blow to ill-conceived legislation, propelled and aided by the actions of large nonprofit and for-profit organizations: in essence, public interest overcoming the efforts of well-funded special interests. In that sense, our analysis supports the more radical interpretation of the events. In this case, the MPAA, the RIAA, and other backers of the legislation whose economic interests would have been served by its passage gained little traction in the online discussion, while those technology

⁹⁹ See Dan Nguyen, “SOPA Opera Update: Opposition Surges,” *ProPublica*, January 19, 2012, <http://www.propublica.org/nerds/item/sopa-opera-update>. (80 supporters of SOPA PIPA in the House and Senate and 31 opponents on January 18, 65 supporters and 101 opponents on January 19.)

companies whose views were aligned with those of a broad civil society coalition were able to ride that alignment to legislative victory. But nothing in our findings precludes less benign future interactions, where parties—be they private commercial interests, government agencies, or political parties—seek to leverage online mobilization techniques in ways that are merely extensions of the media and astroturf campaigns of yore.

Although our analysis does not therefore provide any assurance of future benevolent cooperation, it seems impossible to understand the events of January 18 without also understanding the discourse, framing, and organizing dynamics of the preceding 17 months. This period, as we saw, was comprised of a highly dynamic, decentralized, and experimentation-rich public sphere, where different actors played diverse roles in diagnosing the problems with the acts, reframing the public debate from “piracy that costs millions of jobs” to “Internet censorship” and organizing for action. Moreover, while Google’s role portends the risk of manipulation, the role of Wikipedia suggests a nested, iterative model of democracy. Although Wikipedia is a major distinct player, the potential of a nonprofit, self-governing community of users adopting action after extensive public deliberation is itself an instance of democratic governance in a way that the political activity of a for-profit corporation with distinct commercial interests in the outcome is not.

We identify 10 core findings from our analysis:

First, the networked public sphere is much more dynamic than many previous descriptions. Whether looking at in-links over the 17-month period, either to media sources (Figure 22) or to individual stories or Web pages (Figure 23), the distribution of in-links roughly follows the contours of the familiar power law distribution curves. Visual inspection of weekly or monthly periods reveals the same distributional pattern, but as we look at discrete time slices, the curves are comprised of a more diverse set of nodes: a major node like Wikipedia may be secondary, while an otherwise minor node, such as the blog of a law professor commenting on an amendment or a technical paper on DNS security, may be more important. The dynamic nature of attention in controversies over time means that prior claims regarding a re-concentration of the ability to shape discourse miss vital fluctuations in influence and visibility. Perspective, opinions, and actions are developed and undertaken over time. Fluctuations in attention given progressive development of arguments and frames over time allow for greater diversity of opportunity to participate in setting and changing the agenda early in the debate compared to the prevailing understanding of the power law structure of attention in digital media. This dynamic also likely provides more pathways for participation than were available in the mass-mediated public sphere. This core set of findings squarely supports the networked public sphere model and suggests a substantial limitation of prior empirical claims about the relatively static and highly hierarchical structure of online discourse based on images of link structures on the blogosphere.

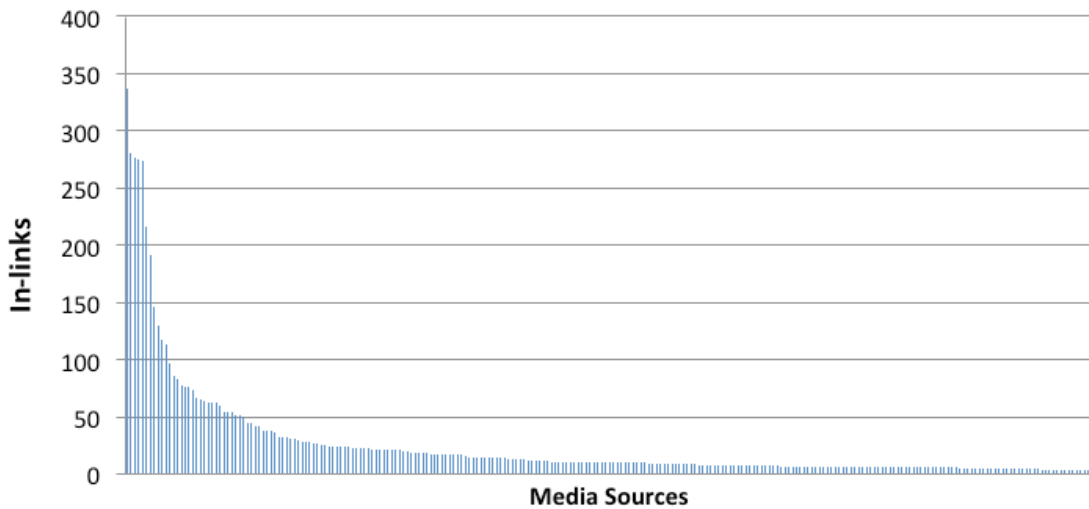


Figure 22: Total in-links to media sources over 17-month period

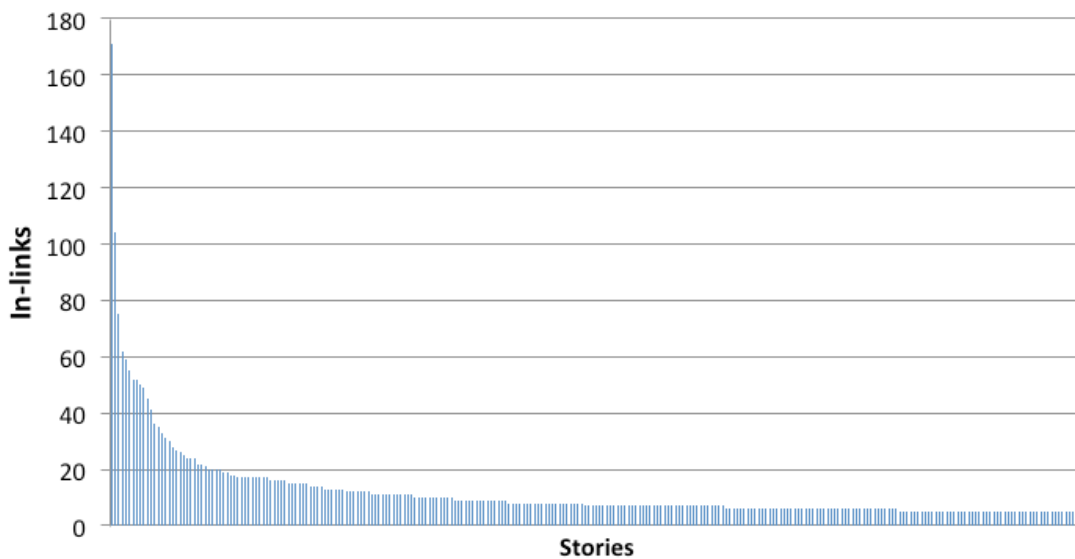


Figure 23: Total in-links to stories over 17-month period

Second, subject-area, professional media, in this case tech media, played a much larger role in shaping the political debate than the traditional major outlets. Techdirt, CNET, Ars Technica, and *Wired* carried the burden of media coverage throughout the period. As seen in Table 2, using in-links as a measure of prominence, tech media occupy three of the top six positions in the network. Tech media initiated the reporting of this issue and continued to lead media coverage throughout the 17-month period.

MEDIA SOURCE	IN-LINKS
Techdirt	337
EFF	315
Reddit	281
Wikipedia	275
CNET	274
Ars Technica	216
American Censorship	192
The Hill	146
House of Representatives Judiciary Committee	130
White House	128
OpenCongress	118
TorrentFreak	114
Politico	97
Washington Post	86
Fight for the Future	83
TechCrunch	83
New York Times	78
Mashable!	77
Wired	77
Forbes	74
Boing Boing	66
Google	64
Guardian	63
PCWorld	60
ProPublica	54
The Huffington Post	54
Public Knowledge	54
Gizmodo	51
YouTube	51
The Library of Congress - Thomas	50

Table 2: Media sources with the most in-links

Third, traditional non-governmental organizations like the Electronic Frontier Foundation and Public Knowledge played a critical role as information centers and as core amplifiers in the attention backbone (discussed more below) that transmits the voices of various, more peripheral players to the wider community. On several occasions, various letters written and posted by experts found a larger audience after being highlighted by the EFF or Public Knowledge. These organizations also proved essential in informing the network about changes and upcoming legislative events.

URL	MEDIA SOURCE	IN-LINKS
http://americancensorship.org/	American Censorship	171
http://en.wikipedia.org/wiki/Stop_Online_Piracy_Act	Wikipedia	104
http://www.whitehouse.gov/blog/2012/01/14/obama-administration-responds-we-people-petitions-sopa-and-online-piracy	The White House	85
http://blog.reddit.com/2012/01/stopped-they-must-be-on-this-all.html	Reddit	75
http://en.wikipedia.org/wiki/Protect_IP_Act	Wikipedia	62
https://wfc2.wiredforchange.com/o/9042/p/dia/action/public/?action_KEY=8173	EFF	56
http://judiciary.house.gov/issues/Rogue%20Websites/List%20of%20SOPA%20Supporters.pdf	Judiciary Committee	55
http://www.reddit.com/r/politics/comments/nmnie/godaddy_supports_sopa_im_transferring_51_domains	Reddit	52
https://www.google.com/landing/takeaction/	Google	52
http://thomas.loc.gov/cgi-bin/query/z?c112:H.R.3261:	Library of Congress	50
http://www.reddit.com/	Reddit	49
https://www.eff.org/deeplinks/2011/12/internet-inventors-warn-against-sopa-and-pipa	EFF	45
http://fightforthefuture.org/	Fight for the Future	41
http://www.opencongress.org/bill/112-h3261/show	OpenCongress	35
http://www.opencongress.org/bill/112-s968/	OpenCongress	33
http://projects.propublica.org/sopa/	ProPublica	31
http://en.wikipedia.org/wiki/Anti-Counterfeiting_Trade_Agreement	Wikipedia	30
http://www.govtrack.us/congress/bill.xpd?bill=s111-3804	GovTrack	28
https://www.eff.org/deeplinks/2012/01/how-pipa-and-sopa-violate-white-house-principles-supporting-free-speech	EFF	27
http://nytm.org/sos/	NYTechMeetup	25
http://www.stanfordlawreview.org/online/dont-break-internet	Stanford Law Review	24
http://thehill.com/blogs/hillicon-valley/technology/204167-sopa-shelved-until-consensus-is-found	The Hill	24
http://judiciary.house.gov/news/01202012.html?scp=2&sq=lamar%20smith&st=cse	Judiciary Committee	24
http://boingboing.net/2011/11/16/internet-giants-place-full-pag.html	Boing Boing	22
http://staff.tumblr.com/post/12930076128/a-historic-thing	Tumblr	22

Table 3: URLs with the most in-links¹⁰⁰

¹⁰⁰ The most linked-to URLs include both the main pages of Web sites (e.g., reddit.com) and specific stories or Web pages (e.g., <http://blog.reddit.com/2012/01/stopped-they-must-be-on-this-all.html>).

Fourth, the widespread experimentation carried out by new and special-purpose sites facilitated the conversion of discussion into action. Several different organizations and individuals experimented with dozens of special-purpose sites and mobilization drives, some of which succeeded in garnering attention and mobilizing effectively via, for example, emails or phone calls to Congress, the symbolic strike of January 18, 2012, and consumer boycotts. Among these, Demand Progress was an early player against COICA, Don't Censor the Net played a large role around the introduction of PIPA, and Fight for the Future emerged as a force around the introduction of SOPA. Each of these players instituted successful efforts prior to the ultimate Wikipedia boycott. Similarly, the Reddit community boycott of Go Daddy was a transformative moment in the campaign targeting corporate support of the bill. The widespread experimentation in these sites was a critical feature. It replicated the same model of innovation observed in the context of Internet innovation more generally: rapid experimentation and prototyping, cheap failure, adaptation, and ultimately rapid adoption of successful models, although in this case channeling that innovative approach towards social mobilization and political action. As seen in Table 2, the top 10 URLs for this period are either informational—the Wikipedia articles on the bills or the bill text itself—or tightly linked with diverse successful mobilization efforts: action sites like American Censorship or Wired for Change; calls to action with discrete instructions, like the two Reddit posts about the Go Daddy boycott and the January 18 blackout; or markers of such mobilization, like the White House response to the petition drive and the House Judiciary Committee list of corporate sponsors that served as a target list for boycotts to change corporate positions.

Fifth, highly visible sites within the controversy network were able to provide an attention backbone for less visible sites or speakers, overcoming the widely perceived effect of the power law distribution of links. In this debate, we see many instances in which posts get picked up by increasingly more visible sites, and are then themselves amplified by yet-more visible sites. For example, Fight for the Future benefited from links from more established sites, such as the Mozilla front page, and as discussed earlier, Julian Sanchez's debunking of the \$58 billion meme benefited from being linked to by Techdirt, which in turn was linked to by both Reddit and the EFF, further amplifying this critique.

Sixth, individuals play a much larger role than was feasible for all but a handful of major mainstream media in the past. A single post on Reddit, by one user, launched the Go Daddy boycott; this is the clearest such example in our narrative. But we also see individuals embedded in organizations that in the past would have been peripheral, who are now able to play prominent roles. Notably, Mike Masnick propelled Techdirt into the single most important professional media site over the entire period, overshadowing the more established media. Individual blogs by academics were able to rise at various moments, like the visible role that law professor Eric Goldman's blog posts played in early December 2011.

Seventh, the network was highly effective at mobilizing and amplifying expertise to produce a counter-narrative to the one provided by proponents of the law. Technologists, law professors, and entrepreneurs emerged at various stages of the controversy to challenge proponents and make expert assertions that went to the core of the debate: the meaning of changes in various drafts, the effects of the laws on DNS security or innovation, or the constitutionality of the bills.

Eighth, consumer boycotts and pressure facilitated by online communities played a key role in shaping business support and opposition. The two most visible instances were the Reddit boycott against Go Daddy and the pressure gamers put on game companies to oppose SOPA-PIPA, which bore fruit in late 2011 and early 2012.

Ninth, at least on questions of intellectual property, the long-decried fragmentation and polarization of the Net was nowhere to be seen. Political activism crossed the left-right divide throughout the period; the opposition was every bit as bipartisan as was congressional support. Demand Progress and Don't Censor the Net are the two most obvious nodes in this bipartisan effort, but we also see more traditional left and right political blogs, like Daily Kos and HotAir, joining in the fight on the same side.

Tenth, the narrative and online actions that are observable in the digital record are highly consistent with the description of events that we took away from interviews and personal knowledge. This congruence lends support to the proposition that the methods we developed for this study offer a reliable rendition of the series of events and the public roles that various actors played in the controversy. We nonetheless must acknowledge that this version, seen only through publicly visible interventions in the networked public sphere, omits the strategic planning and coalition building that occurred in face-to-face meetings, telephone calls, and email¹⁰¹ and that it will require more such studies to refine and validate these methods.

Inferences and implications: locating the SOPA-PIPA debate in a larger context

By the end of the 17 months under study, a diverse network of actors, for-profit and nonprofit, media and non-media, individuals and collectives, left, right, and politically agnostic, had come together. They fundamentally shifted the frame of the debate, experimented with diverse approaches and strategies of communication and action, and ultimately blocked legislation that had started life as a bipartisan, lobby-backed, legislative juggernaut. While it is certainly possible that behind-the-scenes maneuvering was more important and not susceptible to capture by our methods, what is clear is that by ProPublica's tally, before January 18, 2012, SOPA-PIPA had 80 publicly declared supporters and 31 opponents, but by the next day, the bills had 65 supporters and 101 opponents. The January 18 online protest campaign and its anchor, the Wikipedia blackout, were the core interventions that blocked the acts. But our study suggests that this day's events cannot be understood in terms of lobbying or backroom deals; rather, this outcome represents the fruits of the online discourse and campaign by many voices and organizations, most of which are not traditional sources of power in shaping public policy in the United States.

101 A description of some of the behind-the-scenes efforts to coordinate action: Grant Gross, "Who Really Was behind the SOPA Protests?," *Macworld*, February 6, 2012, http://www.macworld.com/article/1165221/who_really_was_behind_the_sopa_protests.html; Mike Masnick, "People Realizing that it Wasn't Google Lobbying that Stopped PIPA/SOPA," *Politics* (blog), *Techdirt*, February 8, 2012, <http://www.techdirt.com/articles/20120207/03304417679/people-realizing-that-it-wasnt-google-lobbying-that-stopped-pipasopa.shtml>. Also see: Susan Sell, "Revenge of the 'Nerds': Collective Action against Intellectual Property Maximalism in the Global Information Age" *International Studies Review* (2013) 15, no. 1, 67-85.

In the longstanding academic debates we described at the beginning of the paper, the SOPA-PIPA debate lends support to the practical feasibility of the models of the networked public sphere and networked fourth estate. It also lends support to the feasibility of effective online mobilization providing sufficiently targeted action to achieve real political results. Perhaps the SOPA-PIPA dynamic will not recur. Perhaps the high engagement of young, net-savvy individuals is only available for the politics of technology; perhaps copyright alone is sufficiently orthogonal to traditional party lines to traverse the left-right divide; perhaps Go Daddy is too easy a target for low-cost boycotts; perhaps all this will be easy to copy in the next cyber-astroturf campaign. Perhaps.

But perhaps SOPA-PIPA follows William Gibson's "the future is already here—it's just not very evenly distributed." Perhaps, just as was the case with free software that preceded widespread adoption of peer production, the geeks are five years ahead of a curve that everyone else will follow. If so, then SOPA-PIPA provides us with a richly detailed window into a more decentralized democratic future, where citizens can come together to overcome some of the best-funded, best-connected lobbies in Washington, DC.

Appendix: Controversy Mapping using Media Cloud

This paper uses Media Cloud, an open source tool created at the Berkman Center to allow quantitative analysis of online media and to create the controversy maps that we use to analyze the SOPA-PIPA controversy. Media Cloud was developed to allow researchers to perform quantitative analysis of the online media ecology without having to incur the cost of discovering and collecting new content themselves; the tool publishes both the code that performs the discovery, crawling, and extraction of text from online sources and the data that the project has crawled from English- and Russian-language mainstream news sources and blogs. Using Media Cloud for this research allows us to answer questions about the structure of controversies like SOPA-PIPA in the networked public sphere, which present thousands of sources instead of dozens. Media Cloud is particularly suited for this work because it allows us to use the same methods to ask the same questions about a variety of different media types—including mainstream media, blogs, advocacy groups, technology media, and so on—rather than just evaluating social media, an essential feature to understand how controversies operate in the diverse online media ecology.

We describe here in some detail how the Media Cloud code works to generate the results in this paper. We provide as much of the data collected by Media Cloud as is legally permissible through data dumps at the Media Cloud website.¹⁰² We encourage those interested in further detail to look at the code for a full description of how Media Cloud and its controversy mapping work.¹⁰³

Media Cloud consists of two related systems: an agenda mapping system that collects and analyzes the content of online media, and a controversy mapping system that mines and analyzes the link networks of online media. The agenda mapping system collects, processes, and analyzes all of the content published by tens of thousands of online media sources. The controversy mapping system mines that content for links that it uses to generate the link networks that are the basis of the controversy maps described in this paper.

The Media Cloud agenda mapping system performs five functions: media set discovery, crawling, text extraction, word vectoring, and analysis. First, Media Cloud defines the set of media sources to collect and discovers the RSS feeds associated with each media source (in the case of many newspapers, there may be hundreds of feeds; in the case of blogs, there is often just one feed). Second, Media Cloud crawls each of those feeds several times a day to discover any new stories published by each feed, then downloads the HTML of each new story. Third, the system extracts each story's substantive content from every downloaded HTML page, filtering out ads, navigation, and any extraneous text that does not represent the primary substantive content of the page. Fourth, the system divides that substantive text into a set of word counts that are broken down to the level of individual sentences. And finally, the project has a number of tools it uses to perform quantitative analysis of online agendas using those sentence-level word counts. As described in more detail below, the multiple functions of the Media Cloud agenda mapping system make our SOPA-PIPA controversy mapping work possible.

102 "Data," Media Cloud, http://www.mediacloud.org/dashboard/data_dumps. The data that underlies the COICA-SOPA-PIPA controversy mapping analysis described in this paper is available here: http://cyber.law.harvard.edu/publications/2013/social_mobilization_and_the_networked_public_sphere.

103 The code for Media Cloud is available at: <https://mediacloud.svn.sourceforge.net/viewvc/mediacloud>.

The controversy mapping work discussed in this paper uses this online content collection system to find the initial set of stories with which to seed the controversy spider. For each media source collected by Media Cloud, the system associates a set of syndication feeds (RSS, RDF, or Atom) that ideally include all stories published by that media source. To discover the set of feeds for each media source, Media Cloud runs a feed discovery spider in the site's main URL. Humans then manually approve the set of feeds (RSS/RDF/Atom) found by the spider. If the spider does not find any feeds for a given media source, humans manually search for a feed associated with that source. For media sources with multiple feeds (mostly big newspapers, some of which include hundreds of feeds), the system includes all of the feeds associated with that media source to try to capture all of the source's content.

Once feeds have been associated with every media source in the system, Media Cloud downloads the feed(s) for each media source about once every four hours. For each item in each feed, Media Cloud first checks whether a story with the item's RSS URL or GUID already exists in the database for the given media source. If not, the system checks whether a story with the same title exists in the database for the given media source within the past week. If it does not, the system then adds the item to the database as a story and queues the story for download. The crawler downloads the URL for each queued story, usually within 15 minutes. Each story is downloaded only once, which means that updates after the first download are not captured. The crawler also tries to discover any additional pages for each story and downloads any such pages as well.

Next, Media Cloud uses a text extractor to pull only the substantive text from each page of HTML downloaded by the crawler. The HTML pages downloaded by the crawler contain not only the substantive text of each story, but also all of the surrounding HTML necessary for formatting, navigation, ads, and other content ancillary to the core story. The navigational content can be especially harmful for analysis of the content because it may include text that meets the criteria for inclusion in a study, even though the text of the associated article contains no relevant text. For example, if the system searched all of the content on a web page of a *New York Times* story for the pattern 'SOPA,' it might find some pages that only mention 'SOPA' in navigational parts of the page, but are not about SOPA in any meaningful sense and which therefore should be excluded from the substantive story text. The Media Cloud text extractor uses the HTML density (the ratio of the number of characters in HTML markup tags to the number of characters in plain text) of each line as the primary signal to determine whether the line should be extracted as part of the substantive text of a story or thrown away.¹⁰⁴ A variety of other signals are used to further tune the decision of the extractor, including total number of characters in a line, location within "clickprint" and other tags that indicate the printable content on some pages, the distance from the last extracted line, the number of comment-related tags before the given line, and the similarity of the text to the feed title and description of the story.

Finally, the Media Cloud agenda mapping system includes code for breaking story texts into sentence-level word counts that can be used for quantitative analysis to explore the relationships between different parts of the online media ecology. Since we do not engage in this sort of analysis in this paper, we will not describe it in detail here.

104 Alexjc [pseud.], "The Easy Way to Extract Useful Text from Arbitrary HTML," *AI Depot*, April 5, 2007, <http://ai-depot.com/articles/the-easy-way-to-extract-useful-text-from-arbitrary-html/>.

Once content is collected as described above, the Media Cloud controversy mapping tool searches this content for a seed set of stories relevant to a given topic. It mines those seed stories for links to stories also relevant to this topic and iteratively repeats this discovery and mining process until it has searched the set of sites linked to the initial set of seed stories. The first step of this process is to search the Media Cloud-collected content for stories that belong to a given set of media sources, fall within a given date range, and match a given regular expression.¹⁰⁵ Media Cloud groups media sources into larger media sets defined by language and media segment. For the SOPA-PIPA controversy, we searched the following media sets: US Popular Blogs, US Political Blogs, and US Top 25 Mainstream Media.¹⁰⁶

Even when searching only within relevant media sets, not all of the stories collected in the seed set are necessarily relevant; in this case, many of the stories were not actually related to the SOPA-PIPA controversy. For instance, some of the stories captured were in fact articles in Spanish about food (*sopa*), entries on a Chinese instrument (*pipa*), pages for the Coordinator of the Indigenous Organisations of the Amazon Basin (COICA), and so on. To remove such noise from our results, we manually reviewed every story from the seed set to verify that it was relevant to the SOPA-PIPA controversy. We used a minimal definition of relevancy: at least one mention of the SOPA, PIPA, or COICA bill within the body of a given story. The result was a seed set of 4,942 stories from the Media Cloud content that were relevant to the COICA/SOPA/PIPA controversy.

Based on this seed set of relevant stories, we extracted the links from the substantive portion of the each story. For each of those links, we downloaded the URL referenced by the link, ran it through the text extractor, and tried to match the extracted text against the above pattern. The spidered stories that matched the pattern were added to the set of controversy stories for the SOPA-PIPA controversy, and those that did not match were dropped. The spider then iterated over those new controversy stories discovered by this spidering process, extracting links, downloading the linked URLs, extracting the substantive text from those web pages, and trying to match that text to the SOPA-PIPA pattern. We continued to iterate through this process until the spider found no new stories. The spider ultimately iterated through 10 generations of stories, finding the following number of stories in each iteration:¹⁰⁷

105 For this paper, we used the following regular expression to search for stories that mentioned any of the SOPA, PIPA, or COICA bills by their acronyms or full titles: `[[:<:]]sopa[[:>:]]stop[[:space:]]+online[[:space:]]+piracy[[:space:]]+act[[:<:]]acta[[:>:]]anti-counterfeiting[[:space:]]+trade[[:space:]]+agreement[[:<:]]coica[[:>:]]combating[[:space:]]+online[[:space:]]+infringement[[:space:]]+and[[:space:]]+counterfeits[[:space:]]+act[[:<:]]pipa[[:>:]]protect[[:space:]]+ip[[:space:]]+act`

106 See http://www.mediacloud.org/dashboard/media_sets/1 for a description of how each of these media sets was generated and a full list of the members of each media set.

107 1,462 stories were merged as duplicates but found in different iterations and so are not included in this list.

0: 4663
1: 2649
2: 639
3: 193
4: 73
5: 36
6: 23
7: 10
8: 4
9: 4
10: 1

For each story found by spidering, we first attempted to match it to an existing story within the database by finding any story with the same URL or redirect URL of the newly spidered story, or by finding any story with a matching title of at least 16 characters within the same media source as the given story. Upon discovering a link with the hypothetical URL “<http://sopa.blog/fight-sopa>,” we would search for any stories already within the database with that URL or with a matching title of at least 16 characters within the same media source.

Within the spidered story set, we did not systematically review stories for relevance to the SOPA-PIPA controversy, but we did remove any stories that we noticed during analysis of the data not to be relevant to the controversy. We also removed any stories that we discovered, through a combination of manual search and clustering, to be written in any language other than English.¹⁰⁸

Much of the analysis in this paper centers around the media sources publishing the stories rather than merely on the stories. For stories already within the core Media Cloud content, we used the existing Media Cloud media source associated with each story. For a spidered URL that did not match a story already within our database, we created a new media source in the database with a URL that matched the host name of the story URL. For example, if our hypothetical URL “<http://sopa.blog/fight-sopa>” did not match any existing story in the database, we looked for any existing media source within our database with the URL “<http://sopa.blog/>”. If no such media source was found, we created one. This approach resulted in many split media sources that should be combined into a single media source. For instance, we might find stories that start with ‘<http://www.sopa.blog/>’, ‘<http://sopa.blog/>’, and ‘<http://news.sopa.blog/>’. After running the spider to completion, we ran a script to group media sources created by the spider according to their domain names and then manually reviewed each such group to determine whether its members should be merged into a single media source.

¹⁰⁸ In this study, we focused exclusively on discussion conducted in English, under the assumption that this would capture the core and most influential nodes of the US policy debate. There is a notable international dimension to this controversy that links the US debate over SOPA-PIPA with the international debate over ACTA.

In addition to the set of seed stories found by searching through the existing Media Cloud content, we also added a set of URLs found by manually searching on Google for the top 100 stories matching each of the following search terms: [SOPA], [Stop Online Piracy Act], [COICA], [Combating Online Infringement and Counterfeits Act], [PIPA], [Protect IP Act]. We then filtered those search results for relevancy, imported the remaining 500 URLs as controversy stories for SOPA/PIPA, and ran the spider to spider out from those stories running the same code we had used to spider from the Media Cloud seed set. In addition to the 114 new seed stories added, we discovered fewer than 50 additional stories by spidering out from the Google seed set.

Adding a fine-grained temporal dimension to our analysis of controversies is a central strength of our approach. Dating stories accurately is, however, a significant challenge, and one caution in reading the results of these studies is that our solutions lead us to a high degree of confidence in our dates, but the solutions we adopted do not provide perfect dating for all sources. For stories within the core Media Cloud content set, we used the date associated with the story in its RSS feed. Of the 9,757 stories in the data set, 4,942 were discovered via their syndication feeds and dated either with the publication in the feed or with the current time (we download all stories within 24 hours of publication in a feed and most within a few hours). We take the automated dating of stories discovered this way to be conclusive. For stories discovered through spidering, we had to guess the date of each story by parsing the URL or text of the story; to do so, we used nine different methods that included looking for specific structured data in the HTML and looking for a URL in the form of “<http://sopa.blog/2011/11/01/fight-sopa>.”¹⁰⁹ As a last resort, when all other dating methods failed, we set the date of publication based on any date located within the text of the story. For a random sample of spidered stories, we found dating stories using these methods to be accurate to the same day 87% of the time. Note that this 87% number is our *lowest* level of confidence in an automatically generated date. For stories in which no date was found in the story text, we assigned the date based on the publication date of the story associated with the earliest link that we discovered to the story in question. Note that to understand the significance of a story in time, the first link is an appropriate date for marking the moment at which the story linked to had any measurable influence.

During analysis, as we happened upon stories that were misdated, we fixed their dates based on a manual review of the publication date offered in the story. We manually reviewed all the highly linked stories, suggesting that at least for all the stories of significance, we performed this manual backup check to verify date. We also ran a query to discover “future links” (links from stories in the past to stories in the future, which is obviously not possible). We manually reviewed each of the 355 cases in which there were more than two such “post-dated” links to a given story, and

109 The default date for each spidered story was the date of the story first found to link to it. The date guessing module then tried to find a more accurate date in the text or url of the story and overwrote the default date with that more accurate date if found. The module first looked for dates in the HTML of each story in the following forms: [`<meta name="DC.date.issued" content="2012-01-17T12:00:00-05:00" />`], [`<li property="dc:date dc:created" content="2012-01-17T12:00:00-05:00" datatype="xsd:dateTime" class="created">`], [`<meta name="item-publish-date" content="Tue, 17 Jan 2012 12:00:00 EST" />`], [`<p class="storydate">Tue, Jan 17th 2012</p>`], [``], [`<time date time="2012-01-17" pubdate>`]. The module then looked for a date in the URL of the story in the form [`http://sopa.blog/2011/11/01/fight-sopa`]. If no date was found in those forms, the module looked for any date anywhere in the text of the story within 14 days of the default date. The full code for this date guessing module is available at: [`http://sourceforge.net/p/mediacloud/code/5070/tree/trunk/lib/MediaWords/CM/GuessDate.pm`].

corrected their dates as necessary. In 43 instances, it was not possible to accurately date a story because the story was associated with a category page that is inherently undateable, such as a search results page, a blog posts archive page, a Wikipedia page, or a web site front page. For these stories, we tried to assign a best guess date that would least disrupt the data, usually the date of the story first linking to that story.

We use the stories, media sources, and links described above to generate the maps found in the body of this paper. For example, below is the map of the SOPA-PIPA controversy during the week of May 23, 2011.

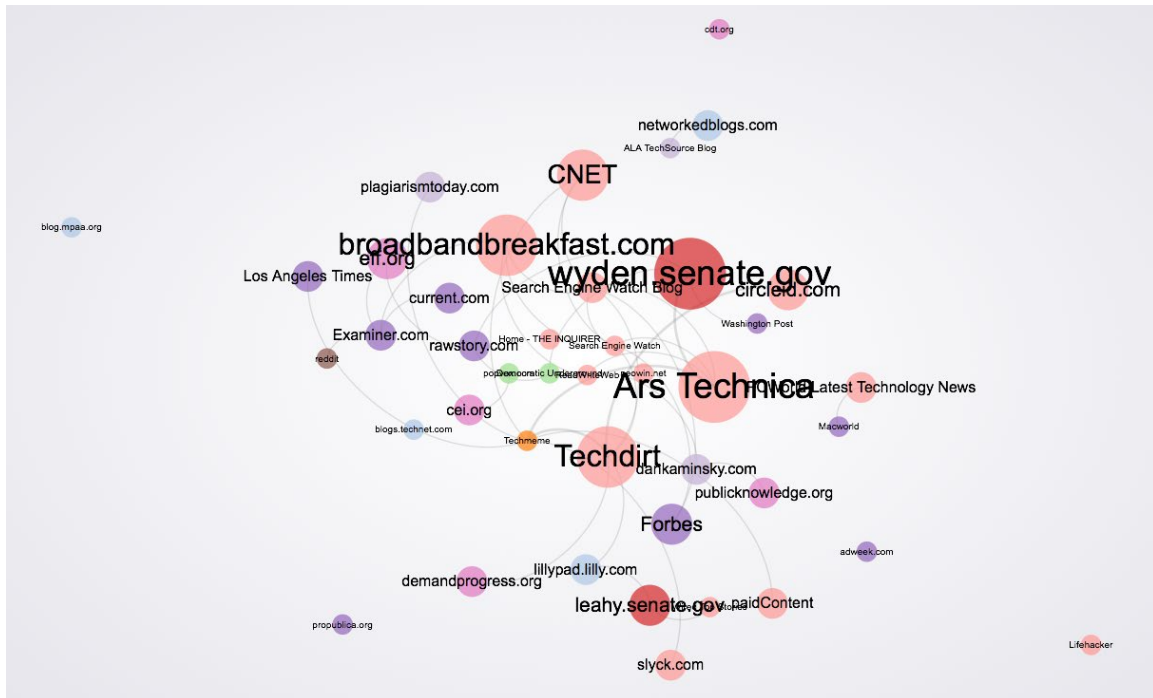


Figure 24: May 23–30, 2011

In this map, each node represents a different media source, and each line between a pair of nodes represents one or more links between stories in the respective media sources. For example, the link between Ars Technica and wyden.senate.gov represents the existence of a link from a story in one of those two sources to a story in the other source. The size of each dot in the map is proportional to the number of incoming links to that media source—the number of links *to* (but not *from*) stories within that media source during the given week. In the above map, Ars Technica, Techdirt, wyden.senate.gov, and broadbandbreakfast.com are the biggest nodes because they have the most incoming links during the week of May 23, 2011.

For any given week, the map includes any media source that either has a story that was published that week or was linked to from a story that was published during that week. A media source that is linked to by another media source may or may not have published a story during that week. Individual media sources that have no lines connecting them to the rest of the network represent stories that were published during the given week, but that were not linked to during the week by any other media source.

To determine the position of each node on the map, we used the ForceAtlas2 layout of the Gephi network visualization tool.¹¹⁰ ForceAtlas2 is a force-directed algorithm that determines the position of each node in the network by simulating a repulsive force between nodes themselves and an attractive force by the links between the nodes. This algorithm produces maps in which groups of nodes that link most heavily to one another are clustered together. Generally, nodes that are the most heavily connected to the rest of the network appear toward the center of the map. Although the centrality of nodes on the map is significant, the physical position of a given node on the map is not; for instance, in Figure 20 above, the *Los Angeles Times* appears on the left side of the map, yet this location is not meaningful beyond the fact that this media source is on the periphery.

To create these maps, we assign a weight to the attractive force between two media sources that is equal to the total number of story-to-story links between the pair of media sources. For instance, in the map above, there are two links between stories in wyden.senate.gov and [Ars Technica](http://www.ars-technica.com) during this week, so the weight of the attraction between those media sources is two. We also use a feature of the ForceAtlas2 layout called “dissuade hubs” that pushes to the periphery nodes that have a high proportion of outgoing to incoming links; consequently, media sources that are connected to the network primarily by outgoing links rather than because they receive incoming links are pushed to the sides of the map. In the above map, [examiner.com](http://www.examiner.com) is a good example of one such node. It has one incoming and three outgoing links, so even though it is relatively well connected for this small map, it gets pushed toward the outside of the network.

¹¹⁰ Jacomy et al., “ForceAtlas2, A Graph Layout Algorithm for Handy Network Visualization.”