

SECTION 5

MEASURING INFLUENCE IN THE POLITICAL BLOGOSPHERE

WHO'S WINNING AND HOW CAN WE TELL?

BY DAVE KARPFF

Abstract

Previous studies of the elite political blogosphere have sought to distinguish a population of “A-list” blogs by culling the top 100 rankings from multiple sources. This methodology is no longer adequate for keeping pace with the quickly expanding universe of blogs, in fact from a practical standpoint; those sources have largely become defunct or developed systematic flaws. This article presents the Blogosphere Authority Index (BAI), a new system for tracking online influence. The index combines four measures of authority to produce comparative rankings of the elite blogs. The BAI is then applied to the progressive and conservative blog networks, resulting in an analysis of their comparative positions. It is evident that the progressive blogosphere has a clear advantage over its conservative counterpart, and that this advantage is based in a small number of hub sites – the elite of the elite. My conclusions run counter to previous scholarship and are due to the BAI’s more comprehensive aggregation of available data.

Introduction

In the past six years, the political blogosphere has grown by leaps and bounds, moving from an interesting curiosity to an enduring feature of the political landscape. Political bloggers are now regularly featured in the mainstream media. The 2007 Yearlykos convention, a gathering of the progressive “netroots”, drew all but one Democratic Presidential candidate while evoking the outrage of Fox News Channel’s Bill O’Reilly. Lenhart and Fox find that the blogosphere has become a major source of informa-

tion for the American public, with 39% of adult internet users turning to the blogosphere as of July, 2006.³⁷

As the phenomenon of blogging has grown, scholarly researchers have had trouble keeping pace with this rapid development. Methods developed just three years ago, when the total size of the blogosphere was estimated around 2 million, are functionally incapacitated by a blogosphere that is now more than 112 million-strong.³⁸ The rise of the blogosphere’s importance has been coupled with a declining capacity for measuring and understanding what, exactly, is going on.

Meanwhile, the most heavily-cited recent article on the blogosphere – Lada Adamic and Natalie Glance’s 2005 piece, “The Political Blogosphere and the 2004 US Election: Divided They Blog” – depicts the conservative bloggers as more heavily linked and thus relatively stronger than their progressive counterparts.³⁹ Ackland’s follow-up analysis supports this finding, concluding that 9 of the 10 most prominent political bloggers come from the con-

37 Amanda Lenhart and Susan Fox, “Bloggers: A portrait of the internet’s new storytellers,” Pew Internet and American Life Project. (July 2006). Available at www.pewinternet.org

38 As measured by technorati.com. Lada Adamic and Natalie Glance, “The Political Blogosphere and the 2004 US Election: Divided They Blog.” Presented at Conference on Knowledge Discovery in Data. 2005. Available at <http://portal.acm.org/citation.cfm?doid=1134271.1134277>

39 Adamic and Glance, “The Political Blogosphere and the 2004 US Election: Divided They Blog.” Pg 14.

servative blogosphere.⁴⁰ This stands at odds with the prevailing media narrative, which portrays resurgent liberals as a dominant force in the emerging medium.⁴¹

This article introduces the Blogosphere Authority Index (BAI), a system for measuring online influence that takes into account four distinct measures of strength and combines them to offer a comprehensive picture of the elite blogosphere. After the failings of previous methodologies, the BAI represents a new technique for comparative ranking of political blog sites. The resulting application takes a fresh look at the comparative strength of the American Left and Right in the new media landscape.

Literature Review

Only a handful of scholars have devoted significant

- 40 Robert Ackland, "Mapping the U.S. Political Blogosphere: Are Conservative Bloggers More Prominent?" Submission to BlogTalk 2005. Accessed online at <http://incsub.org/blogtalk/images/robertackland.pdf>
- 41 Daniel Shulman, "Politics 2.0: Fight Different," *Mother Jones*, July/August 2007.

attention to the political blogosphere. It appears that scholarly interest has particularly focused on the 2004 elections, with the bulk of papers that study the elite bloggers choosing this time period for their data collection.⁴² One common methodological thread running through these conference papers, dissertation manuscripts, and journal articles is that they all identify the population of elite political blogs by combining the publicly-available rankings of a few sites. Below are lists of the publicly-available research to date, along with the type of analysis that was conducted and the sites used to create these population lists.

Today, these data sources are either defunct, systematically flawed, or overbroad.

- 42 More recent scholarship has branched out to look at other types of political bloggers. This includes Pole's work on black bloggers (2005), Congressional bloggers (2006), and Bloom and Kerbel's work on "Blog for America" (2005) and political campaign blogs (2006). There is, however, no available work that analyzes or contrasts the elite political blogosphere(s) post-2004.

FIG 5.1

AUTHORS, TITLE	TYPE OF ANALYSIS	DATA SOURCE(S)
McKenna and Pole (2004), "Do Blogs Matter? Weblogs in American Politics"	Elite population study culled from multiple rankings	Blogstreet Truth Laid Bear Ecosystem Technorati Truth Laid Bear EcoTraffic
Drezner and Farrell (2004), "The Power and Politics of Blogs"	Detailed overview of the political blogosphere	Technorati Truth Laid Bear EcoTraffic Blogstreet
Gill (2004), "How Can We Measure the Influence of the Blogosphere?"	Methodological overview of influence tracking mechanisms	Blogosphere.us Blogrunner Blogstreet Technorati
Hindman (2005), Voice, Equality, and the Internet [doctoral dissertation]	Blogger Social Survey based on traffic rankings	Truth Laid Bear EcoTraffic
Adamic and Glance (2005), "The Political Blogosphere and the 2004 US Election: Divided They Blog."	Study of link patterns in political blogosphere	Blogpulse, with comparisons to Technorati Truth Laid Bear EcoTraffic Truth Laid Bear Ecosystem
Ackland (2005), "Mapping the U.S. Political Blogosphere: Are Conservative Bloggers More Prominent?"	Network analysis of A-list blogosphere using uberlink_ research software	Adamic and Glance
Wallsten (2007), "Political Blogs: Transmission Belts, Soapboxes, Mobilizers or Conversation Starters." [doctoral dissertation chapter]	Elite population study culled from multiple rankings	Blogstreet Truth Laid Bear Ecosystem Technorati Truth Laid Bear EcoTraffic

Defunct:

- Blogstreet.com is now the home address of Blogstreet India. Formerly host to the reputable “Blogstreet Top 100,” it now tracks a little over 6,000 blogs, all of them India-focused.
- Blogosphere.us no longer exists
- Blogrunner.com no longer provides a ranking system.

Systematically flawed:

- Truth Laid Bear Ecosystem ranks blogs based on hyperlink patterns. It demonstrates a strong bias in favor of conservative blogs that are considered insignificant in size according to other sites that track link density. It also appears to be infrequently updated. The purveyor of this site promotes his conservative web consulting agency prominently on the ranking page.
- Truth Laid Bear EcoTraffic aggregates traffic ranking data from sitemeter.com. This data is only available for sites that choose to add the tracking software and publicly report their results. This excludes close to 40% of the top progressive blogs. Additionally, the “top 100” feature is muddled by misreports from sports blogs and economics blogs⁴³. This crowds out the EcoTraffic top 100, though aggregation of sitemeter data is nonetheless the best available means of comparing site traffic.
- Blogpulse.com still exists, but fails to include a number of the major political blogs in its tracking system.

Overbroad:

- Technorati has emerged as the most reputable blog-tracking site on the web. For researchers, this success comes with a price, however. In 2004, technorati tracked 2 million blogs. A large chunk of their top 100 was occupied by political blogs. At the time of this writing, technorati tracks 112.8 million blogs. As the total blogosphere has expanded, much of the political blogosphere has been crowded out of the overall top 100. Only 13 political

blogs are listed in the technorati top 100. All political blogs are tracked, but the value of technorati is now derived from site-specific search rather than from culling a top 100 list.

Culling the population of “A-list” blogs from multiple top 100 lists is no longer methodologically viable. A suitable set of top 100 lists is no longer available. In addition, relying on third-party top 100 listings bypasses the theoretically challenging question of what, exactly, should be treated as a measure of influence.

What Is Influence?

Consider the following example: blogger A posts infrequently on her personal site. This results in a small reader base, and comparatively few hypertext links from around the blogosphere. Years ago, blogger A was a mentor to bloggers B, C, and D, and she now holds a key position within her party’s establishment. The few people who frequent her blog are highly influential, either in the blogosphere or in more traditional political institutions. Blogger B posts once or twice a day on his individual blog, which was picked up by a major online news magazine last year. He has a journalistic background and specializes in developing new arguments or breaking new stories. He chooses to be a blogger because he likes to set his own deadlines, operate without an editor looking over his shoulder, and publish instantaneously. He often relies on blogger A for insights and tidbits that he researches and turns into original articles. He is among the most often-cited bloggers online, by liberals and conservatives alike. Blogger C posts 15-20 times per day. She rarely publishes original content, instead pouring over other blogs and writing short, pithy posts that tell her reader base about something interesting elsewhere on the web. She acts as a gatekeeper for her gigantic readership, who use her site as a roadmap to the rest of the Internet. Blogger D is the purveyor of one of the most active community blogs in the country. He posts 8 times per day, with some original content and some “open threads” so that his community can keep their own discussion going. This community also publishes their own diaries, often 50 or so in a day. Political endorsements from this site mean dollars in a candidate’s pocket. The membership recently spun off two new sites to support activity around universal health care and global warming, and the policy proposals from these sites have been adopted into Congressional legislation.

Which blogger is most influential?

This example is an illustration of four distinct areas of influence: *network centrality*, *link density*, *site traffic*, and *community activity*. To create a comprehensive ranking system, this paper identifies the best-available proxy for each of these types of influence, converts them to ordinal

43 As an example, Bulletsforever.com is a community fan site for the Washington Wizards basketball team. It’s self-reported traffic-data lists 674 visits/day. EcoTraffic mistakenly lists bulletsforever and 48 other sports-related fan sites as part of a 49-way tie for the 37th highest-trafficked blog in the world, with 129,277 visits/day! Similar errors occur for a cluster of 69 stock market-tracking blogs.

rankings, and then combines them into a single index of authority.

The Blogosphere Authority Index

The Blogosphere Authority Index (BAI) combines data from four measures of online influence into a single ranking system. Construction of the BAI is a two-stage process. The first stage involves collection of raw data on each of the influence measures. Some of this data is directly gathered through content analysis of archived site information. Some is gathered through publicly available tracking systems such as Technorati and Sitemeter. For future researchers, it is important to note that as the blogosphere continues to expand, these tracking systems may eventually become obsolete. Raw data should be collected from the best-available data source. The rest of the BAI's structure remains unaffected by long-term changes in the source of proxy data on any of the four metrics.

In the second stage, we convert this raw data into ordinal rankings. These rankings can be within the progressive or conservative blog "neighborhoods," or the raw scores of both neighborhoods can be combined to form a combined BAI. We then convert these individual ordinal rankings into a comprehensive ranking, leading to a final authority index. Data collection and conversion methods are described in greater detail below.

Stage 1:

The Authority Index combines four measures of influence: the Network Centrality Score, Hyperlink Authority Score, Site Traffic Score, and Community Activity Score.

Network Centrality Score

The *Network Centrality Score* (NCS) is an applied sociometric variable that is culled from blogrolls, which are self-reports of recommended or approved blogs. Blogrolls are similar in nature to self-reports of friendship networks, a common tool in social network analysis. Deletions from blogrolls are rare and usually accompanied by an uproar.⁴⁴ This biases the centrality score in favor of those blogs/bloggers who have been around for the longest. This is a bias which *should* be captured, though. Jerome Armstrong (founder of MyDD) is described by his peer bloggers as "the true blogfather" because of the importance his site had when many were getting their start. This elder statesman-status is a type of power that doesn't track to current visits/activity. It also gives a hint as to *who* might be visiting his site. A site that receives 30 visits/day could be hugely important if those 30 visitors are frontpage authors on the 30 largest sites, for instance. The centrality score, therefore, is meant as an indicator of accrued reputation over time. Beginning with network cen-

trality also solves the problem of coding blogs as political/apolitical or liberal/independent/conservative. Adamic and Glance have demonstrated that the liberal and conservative blogospheres form largely independent neighborhoods.⁴⁵ If an elite blog self-identifies as conservative, but other conservative bloggers refuse to acknowledge it as such, are they part of the conservative blogosphere? This method treats the expansive blogosphere as consisting of network neighborhoods. The blogroll of a site is a means of self-reporting who one's "neighbors" are. In a world of 118 million+ blogs, this is a necessary simplifying assumption for scholars interested in identifying elite blogs within a particular category. To construct the NCS, we begin with a seed site. This is the site that we have good reason to believe lies in or around the center of the blog network of interest. The selection of a seed site will have an effect on our findings, so this is an important decision. Conveniently, this hub will often be a well-known site. For this study, I chose to begin with dailykos.com and michellemalkin.com. These two sites are listed as the progressive and conservative frontrunners by both technorati.com and truthlaidbear.com.

The NCS aggregates blogroll data within the network neighborhood. The data begins with the seed site blogroll. All listed blogs are then visited and their blogrolls are added to the dataset. A first-round tally is then created. Which sites are mentioned most often? Any sites in the top 25 that have not already had their blogrolls added are then visited in a second round of data collection. This continues for a third round if needed. This then produces a final set of tallies. *Network Centrality Score* is constructed by normalizing these tallies on a 0 – to –1 scale. Add 1 to the tally of any site that provided a blogroll (some elite sites do not include a blogroll) to obtain N. This sets the highest possible score at 1.0, removing a bias against contributing sites (source sites do not blogroll themselves). *Network Centrality Score* is then equal to N/P , where P is the total population of blogs included in the coding scheme.⁴⁶ The *Network Centrality Score* provides a population list for data collection on the other 3 measures. This is a necessary initial step, as sitemeter and Technorati data are gathered by searching for sites by name. For the purposes of the other three data sources, I gathered data on the top 50 progressive and top 50 conservative blogs.

Hyperlink Authority Score

The *Hyperlink Authority Score* (HAS) is derived directly from Technorati.com's authority tracking system. Technorati tracks the global blogosphere, using sophisticated web crawlers to measure link patterns and definitively describe which blogs are the most often-cited. As previously noted, Technorati's top 100 tells us less *within* the political blogosphere than it used to because Technorati now tracks so many apolitical sites. It is some-

44 For an illustrative example of "blogroll purges," see <http://www.boomantribune.com/story/2007/3/21/91449/7914>

45 Adamic and Glance, "The Political Blogosphere and the 2004 US Election: Divided They Blog." Pg 14.

46 P can vary drastically depending on the blogroll size of the seed site.

what akin to a public opinion poll that asks citizens which they prefer: Barack Obama, the Detroit Tigers, celebrity gossip, or puppies. With that said, Technorati measures link patterns at a depth that cannot be replicated, so the within-group measure is still a valuable indicator of importance. To construct the HAS, search for the rankings of each blog in the population list (as determined by NCS rankings) in Technorati's overall system. These rankings must be individual accessed by entering [http://www.technorati.com/blogs/\[webaddress\]](http://www.technorati.com/blogs/[webaddress]).

Site Traffic Score

The *Site Traffic Score* (STS) is the most obviously relevant and also the most challenging to accurately measure. Sitemeter.com directly measures the number of unique visitors a website receives every day. A high-traffic site should, *ceteris paribus*, be more important than a low-traffic site. Some bloggers, wary of the various link-tracking systems, rely exclusively on Sitemeter data to measure their success.⁴⁷ The problem, however, is that Sitemeter is an opt-in system, and over 25% of the sites in the dataset do not choose to include it. Previous scholars have used Sitemeter in their ranking systems, relying on the Truth Laid Bear traffic rankings, which aggregate sitemeter data, and simply make note that this is an incomplete system. These scholars only try to determine the population of elite political blogs without attempting to rank-order them. This study attempts to fill the gaps so we can use traffic rankings to gain greater analytic leverage.

To augment the standard Sitemeter data, I used another traffic-ranking system, Alexa.com. Alexa tracks visits to all web sites, rather than just blogs. It measures unique visits/month (a different datum than Sitemeter's visits/day) and offers higher-quality information on larger sites rather than smaller ones. I considered switching to Alexa rankings as a primary traffic ranking system, but many blogs in the study, particularly the lower-traffic ones, have low-quality or no data recorded by Alexa. Instead, I rank-ordered the sites with available Sitemeter data and then checked these sites against their comparative Alexa rankings to see if there was relative similarity between the two. Though the fit was not perfect, it was quite close, with the 9th-ranked site in Sitemeter being the 9th-largest Alexa site +/- 1 in the rank order (the error signature gets larger as we go down the list, as the raw score gaps between the sites quickly shrinks). This fit was the best available indicator, so I then performed "Alexa fills" for all sites that did not offer Sitemeter data. With an Alexa fill, we have no relevant raw score for the site, but we can estimate that it is larger than sites with lower Alexa scores and smaller than sites with higher ones. This was important for estimating the size of a few major sites, including Huffingtonpost.com, thinkprogress.org, and talkingpointsmemo.com. I would characterize this technique as making the best of a

bad-data situation.⁴⁸

The STS metric begins by recording Sitemeter data from all participating sites. This directly replicates the information available from TruthLaidBear EcoTraffic. Rank these sites and attempt to personally contact any bloggers who do not provide Sitemeter data to see if they will share it. Any remaining sites should receive an Alexa fill, triangulating their position in the rankings in comparison to Sitemeter-ranked blogs.

Community Activity Score

The fourth and final data source is the *Community Activity Score* (CAS). Nearly all of the blogs in this study (88%) include a mechanism for readers to write comments in response to the author's post. This is the most basic form of user-generated content on the blogosphere, and it has been lauded as a major difference between blogging and traditional journalism. To date, no scholarly study has used comments as a data source, though it has been noted that someone eventually should.⁴⁹

The particular value of the CAS is that it is a measure of engagement. Presumably, a reader who leaves a comment is likely to spend more time on the site and is more likely to become part of that site's community. The CAS balances out the historical bias of the NCS by giving additional weight to the blogs where active, vibrant conversation occurs. To construct the CAS dataset, take a one-week snapshot of all postings for sites in the population set. For this study, I chose the week of 11/4-11/11. Record the total number of reader-generated comments on these posts. Total Comments/Week serves as a site's CAS.

Stage 2:

At this point, we have four types of raw data recorded for the top 50 conservative and top 50 progressive sites. Each of these should be converted to "top 50" ordinal rankings. In stage 2, these rankings are combined into an aggregate score, adding the three best scores together and dropping the fourth from the study. This is to avoid unfairly biasing the study against sites whose architecture does not allow for reader comments and to minimize outlier effects that come from flaws within any of

48 Additionally, there were four blogs for which an Alexa fill did not work. Alexa tracks traffic to an entire site domain name. This means that Glenn Greenwald's blog, which is featured on slate.com, is recorded as receiving all slate.com traffic, despite slate being a major online news organization which directs only a small portion of its traffic to Greenwald. This was also a problem for Matthew Yglesias's blog on the Atlantic Monthly, James Wolcott's blog on Vanity Fair, and The Corner, a group blog hosted by the National Review.

49 Kevin Wallsten, "Political Blogs: Transmission Belts, Soapboxes, Mobilizers or Conversation Starters?" Doctoral dissertation chapter, University of California, Berkeley

47 Duncan Black, interview, Philadelphia, PA, October 20, 2007.

the four measures employed. Additionally, it minimizes the penalty against institutional sites like HuffingtonPost.com, and Newsbusters.com. Sites of this type receive depressed Network Centrality Scores because some bloggers only list traditional individually-based sites in their blog-rolls. The final ranking equation is $\text{RankFinal} = \text{Rank1} + \text{Rank2} + \text{Rank3} + \text{Rank4} - \text{WorstRank}$. The best possible score is 3, indicating that a blog was first-ranked in three categories. The worst possible score is 150, indicating that a blog was last-ranked in all categories. This purpose of this study is to produce a list of the top 25 progressive and conservative blogs. Data on the top 50's was gathered in order to ensure that all potential members of the top 25 would be captured. The final output is a Blogosphere Authority Index of the two top 25 lists. A combined list was also constructed by combining the raw score data for the two network neighborhoods, constructing combined ordinal rankings, and then applying the same final ranking equation listed above.

Findings

Within the progressive and conservative neighborhoods, the BAI produces the following top 25 rankings circa November 2007 (individual rankings for each of the four authority scores are listed in parentheses):

Progressive Blogosphere

1. Daily Kos (1,2,2,1)
2. Huffington Post (34,1,1,6)
3. Atrios (2,13,6,2)
4. Crooks and Liars (7,4,3,4)
5. Talking Points Memo (2,5,5,12)
6. Firedoglake (4,8,7,3)
7. Think Progress (20,3,4,7)
8. America Blog (14,10,8,5)
9. MyDD (5,12,13,8)
10. Washington Monthly (9,7,9,9)
11. Digby (5,11,11,15)
12. Glenn Greenwald (13,6,10,14)
13. Juan Cole (11,16,14,45)
14. Carpetbagger Report (24,15,12,13)
15. Feministing (47,9,14,18)
16. Pandagon (8,31,18,16)
17. Matthew Yglesias (16,16,21,10)
18. Talk Left (11,17,19,17)
19. Sadly, No! (24,n/a,30,31)
20. Ezra Klein (14,19,23,19)
21. Crooked Timber (29,18,22,23)
22. TBogg (16,40,25,24)
23. Open Left (37,20,20,28)
24. Lawyers, Guns, and Money (20.30.27,21)
25. Brad DeLong (24,25,26,33)

Conservative Blogosphere

1. Michelle Malkin (2,1,2,6)
2. Instapundit (1,5,1,n/a)
3. Little Green Footballs (6,6,4,1)
4. Hot Air (50,4,3,3)
5. Town Hall (34,2,11,2)
6. Newsbusters (73,3,5,7)
7. Powerline (3,8,6,19)
8. Ace of Spades HQ (9,17,10,5)
9. Red State (15,11,9,4)
10. Captain's Quarters (4,13,8,15)
11. Hugh Hewitt (5,9,13,16)
12. Right Wing News (7,16,20,10)
13. Wizbang (9,20,7,18)
14. The Corner (16,7,12,n/a)
15. Volokh Conspiracy (42,12,14,11)
16. Outside the Beltway (16,14,15,n/a)
17. Patterico's Pontifications (11,29,22,14)
18. Protein Wisdom (12,30,23,13)
19. Polipundit (16,44,24,8)
20. The Jawa Report (13,21,17,22)
21. Ann Althouse (37,25,18,9)
22. Tim Blair (23,34,21,12)
23. Black Five (24,15,19,26)
24. Dean Esmay (16,41,25,23)
25. Gateway Pundit (49,18,16,33)

When the four metrics are combined into a single index of authority, a picture of the comparative strength of the progressive and conservative blogospheres emerges (**progressive sites are in bold**, *conservative sites are in italics*):

1. **Dailykos**
2. **Huffington Post**
3. *Michelle Malkin*
4. **Crooks and Liars**
5. *Instapundit*
6. **Atrios**
7. **Think Progress**
8. **Talking Points Memo**
9. **FireDogLake**
10. *Little Green Footballs*
11. *Hot Air*
12. *Town Hall*
13. *Newsbusters*
14. *Powerline*
15. **America Blog**
16. **MyDD**
17. **Washington Monthly**
18. *Hugh Hewitt*
19. **Glenn Greenwald**
20. *Red State*
21. **Digby**

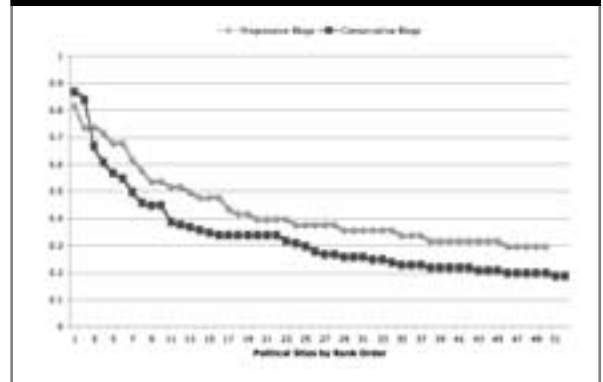
22. *Captain's Quarters*
23. *Ace of Spades HQ*
- 24. Talk Left**
25. *Volokh Conspiracy*
- 26. Juan Cole**
27. *Right Wing News*
- 28. Matthew Yglesias**
- 29. Pandagon**
30. *Wizbang*
- 31. Carpetbagger Report**
- 32. Feministing**
33. *The Corner*
- 34. Sadly, No!**
- 35. Ezra Klein**
36. *Ann Althouse*
37. *Outside the Beltway*
38. *Patterico's Pontifications*
39. *Protein Wisdom*
40. *Polipundit*
41. *The Jawa Report*
- 42. TBogg**
- 43. Crooked Timber**
- 44. Lawyers, Guns, and Money**
45. *Black Five*
46. *Tim Blair*
- 47. Open Left**
- 48. Brad DeLong**
49. *Gateway Pundit*
50. *Dean Esmay*

Discussion

Seven of the top 10 political blogs are progressive in nature, indicating a clear advantage to the progressive bloggers. Additionally, previous researchers find that the distribution of hyperlinks follows a power law, or scale-free pattern⁵⁰. This distribution is characterized by a steeply-declining curve, distinguishing it from the normal curve distributions we are used to studying in social science research. Scale-free distributions are often termed “winner-take all” or “rich-get-richer” distributions because the largest site receives roughly 5 times the attention of the 5th largest site, 10 times the attention of the 10th largest site, etc. Below, the raw scores of each individual metric are presented graphically to see if power law tendencies emerge:

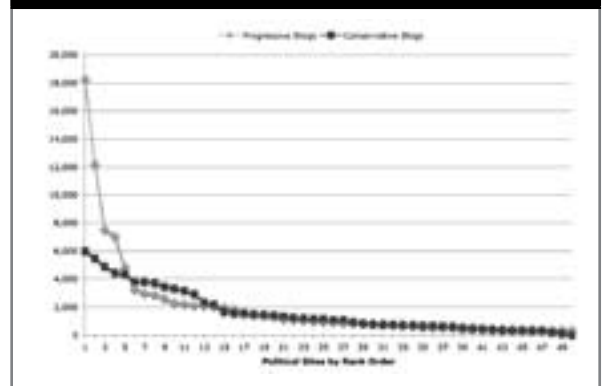
50 See Matthew Hindman, Kostas Tsioutsoulis, and Judy Johnson, “‘Googlearchy’: How a Few Heavily-Linked Sites Dominate Politics on the Web.” Paper presented at the Annual Meeting of the Midwest Political Science Association (2003). Also Clay Shirky, “Power Laws, Weblogs, and Inequality,” reprinted in Ratcliffe and Lebkowsky, eds. 2005. *Extreme Democracy*. Stanford, CA: Creative Commons. <http://extremedemocracy.com>

FIG 5.2: COMPARATIVE NETWORK CENTRALITY SCORES



Recall that NCS accounts for the number of times a site is mentioned in other sites’ blogrolls, normalized on a zero-to-one scale. The sites included in this study all have a relatively high NCS, as constituted the screening process for further data collection. What is particularly interesting is that there is evidence that Adamic and Glance’s initial finding remains correct with relation to blogrolls. The average conservative blogroll in this study had 129 links, while the average progressive blog had 72. Conservative bloggers are more generous with their links, it would seem, but this generosity is diffuse. Only 7 conservative blogs are listed in over 50% of all included blogrolls, which is surprising given the outstandingly large number of average links. How is it that a major site like powerlineblog.com can be ignored in so many blogrolls, despite its longstanding authority within the conservative blogosphere?

FIG 5.3: COMPARATIVE HYPERLINK AUTHORITY SCORES

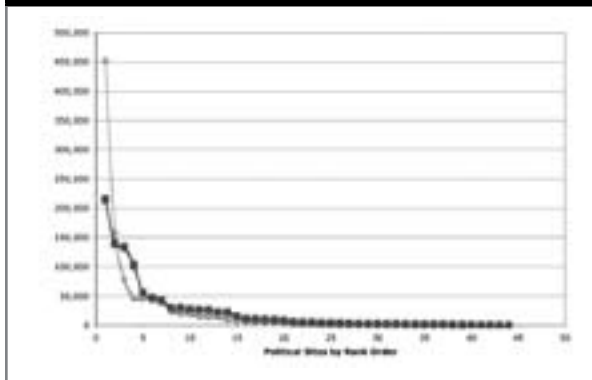


Within the progressive blogosphere, the trademark steeply-descending curve of the power law distribution is present in hyperlink patterns. Two additional interesting observations emerge from this graph. First, this rich-get-richer pattern does not seem to hold true for the conservative blogosphere. The top sites in the conservative blogosphere act less as central hubs than do the top sites in the progressive blogosphere. It is likely that this is tied

to a disparity between the architecture of the elite blogs. It appears as though “community blogging” architecture is more heavily distributed through the progressive than conservative blogosphere. Dailykos.com includes over 1,000 user-generated “diaries” per week, providing far more content for potential hyperlinking. Huffingtonpost.com receives submissions from hundreds of well-known authors, including many elite bloggers who also maintain their own personal sites. The top two conservative sites – michellemalkin.com and instapundit.com – employ a simpler “individual blogging” architecture.

Second, notice that after the top four progressive sites – Huffington Post, Dailykos, Think Progress and Crooks and Liars – there is convergence between the progressive and conservative authority scores. A random sampling of elite progressive and conservative blogs would be likely to find relative equality between the two blogospheres. The difference between the liberal and conservative blogospheres lies in the hub sites – the elite of the elite.

FIG 5.4: COMPARATIVE SITE TRAFFIC SCORES



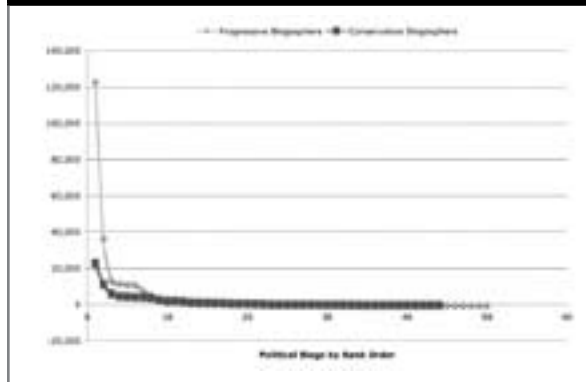
It bears noting that the data problems endemic to measurement of site traffic skew this graph. The sites that required an “Alexa fill” are not included. That includes three of the top 5 progressive sites in the STS rankings. Huffington Post appears to be 3 times as large as Dailykos, represented here as the top liberal site.⁵¹ Think Progress and Talking Points Memo are also both major site traffic hubs, but neither provides sitemeter tracking data. If anything, this suggests that the progressive blogosphere is even further influential than the conservative blogosphere that the graph above suggests.

Nonetheless, again we see the same two patterns as in the HAS influence graph. Once more, there is a clear power law distribution in the liberal blogosphere and a less-distinct (but clearly present in this case) power law

51 This statement is itself problematic. Huffington Post is a news aggregator site with major topical headings much like online newspaper sites. It includes politics, media, business, entertainment, and living sections. As such, it is not possible to discern the percentage of visits/day that are directed to the site’s political section per se.

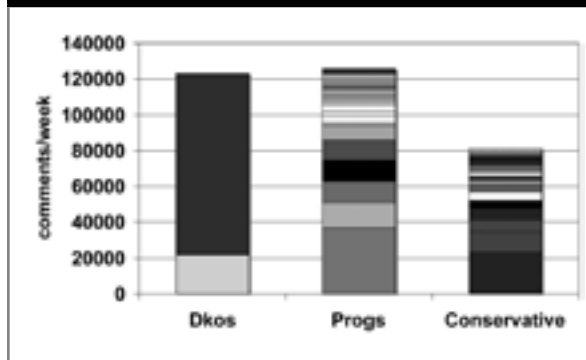
distribution in the conservative blogosphere. The progressive blogosphere is much larger than the conservative blogosphere, but only among the top 5 or so hub sites. Outside of the hubs, there is parity between the two.

FIG 5.5: COMPARATIVE COMMUNITY ACTIVITY SCORES



The Community Activity Score is the one area of comparison that has never before been used in scholarly analysis of blog authority. The results were so stark that they are better presented in a different form:

FIG 5.6: COMMUNITY ACTIVITY SCORES



This graphic compares the community activity on Dailykos.com with the community activity on all other progressive sites and all conservative sites in the top 50 dataset. Each slice of the bar is the CAS of a site in that category.

In terms of user generated content production, when we combine comments on Dailykos’s frontpage with comments in Dailykos diaries, that individual site is nearly as large as all of the other progressive blogs *combined* and is roughly 50% larger than the *entire* conservative blogosphere. This starkly underlines the importance of the central hub sites in comparing elite political blog networks. The community activity on Dailykos dwarfs that of any competing site. As such, it is unsurprising that this site has become known for successful mobilization of political activists. Dailykos has more in common with traditional interest groups than with individual opinion

columnists or journalists. There is simply no comparable entity in the conservative political universe.

Conclusion

It is clear that the progressive blogosphere is indeed larger than the conservative blogosphere. Conservatives link to each other through their blogrolls at a higher average rate, but beyond that the progressive neighborhood receives more traffic, hyperlinks, and community activity. This is concentrated within a few hub sites that stand out as winners in the power law topology of the Internet. This provides a cautionary methodological warning for future blogosphere research. Previous scholarship has made a distinction between the blogging "A-list" and the large mass of non A-list blogs. This population-sampling approach runs the risk of missing some of the most important variance within the blogosphere. All A-list blogs are not created equal. Indeed, the progressive advantage is housed almost entirely in its top four or five sites.

This illuminates the value of the Blogosphere Authority Index. By combining four separate measures of authority, the BAI provides for greater analytic leverage through comparative ranking of sites. This has not previously been used by social scientists in this field. The BAI is simple to construct, relying on the best-available proxies for each measure of influence. Moving forward, it has the potential to provide snapshots of the changing scope of influence among political blogs.

References

- Ackland, Robert. 2005. "Mapping the U.S. Political Blogosphere: Are Conservative Bloggers More Prominent?" Submission to BlogTalk 2005. Accessed online at <http://incsub.org/blogtalk/images/robertackland.pdf>
- Adamic, Lada and Natalie Glance. 2005. "The Political Blogosphere and the 2004 US Election: Divided They Blog." Presented at Conference on Knowledge Discovery in Data. Accessed online at <http://portal.acm.org/citation.cfm?doid=1134271.1134277>
- Bloom, Joel David and Matthew Kerbel. 2006. "Campaign Blogs in 2004 and Beyond: The Care, Feeding and Harvesting of Online Communities of Supporters" Paper presentation at the 2006 annual meeting of the American Political Science Association
- Drezner, Daniel and Henry Farrell. 2004. "The Power and Politics of Blogs." Paper presented at the annual meeting of the American Political Science Association
- Gill, Kathy. 2004. "How Can We Measure the Influence of the Blogosphere?" http://faculty.washington.edu/kegill/pub/www2004_blogosphere_gill.pdf
- Hindman, Matthew, Kostas Tsioutsoulis, and Judy Johnson. 2003. "'Googlearchy': How a Few Heavily-Linked Sites Dominate Politics on the Web." Paper presented at the Annual Meeting of the Midwest Political Science Association.
- Hindman, Matthew. 2005. **Voice, Equality, and the Internet**. Doctoral dissertation, Princeton University
- Kerbel, Matthew and Joel David Bloom. 2005. "Blog for America and Civic Involvement" **Harvard International Journal of Press/Politics**. volume10; issue 3. Sage Publications.
- Lenhart, Amanda and Susan Fox. 2006. "Bloggers: A portrait of the internet's new storytellers," Pew Internet and American Life Project. Available at www.pewinternet.org
- McKenna, Laura and Antoinette Pole. 2004. "Do Blogs Matter? Weblogs in American Politics." Paper presented at the annual meeting of the American Political Science Association
- Pole, Antoinette. 2006. "Congressional Blogging: Advertising, Credit Claiming, & Position Taking." Paper presentation at the 2006 Annual Meeting of the American Political Science Association.
- Shirky, Clay. 2003. "Power Laws, Weblogs, and Inequality," reprinted in Ratcliffe and Lebkowsky, eds. 2005. **Extreme Democracy**. Stanford, CA: Creative Commons. <http://extremedemocracy.com>
- Shulman, Daniel, "Politics 2.0: Fight Different," **Mother Jones**, July/August 2007.
- Wallsten, Kevin. 2007. "Political Blogs: Transmission Belts, Soapboxes, Mobilizers or Conversation Starters?" Doctoral dissertation chapter, University of California, Berkeley