Coolhunting for Trends on the Web

Peter A. Gloor
MIT Center for Collective Intelligence
pgloor@mit.edu

Invited Paper

ABSTRACT
This paper introduces a new way of measuring the popularity of brand names and famous people such as movie stars, politicians, and business executives. It is based upon the premise that in today's Internet economy the Web displays a mirror of the real world. Our system uses TeCFlow, a social networking tool developed for the last four years at MIT, to measure popularity and influence of brands and stars by looking at their relative position on the Web. It is based on the simple insight: "You are who links to you". It applies the Social Network Analysis (SNA) metric of "betweenness centrality" to the Web, looking at the linking structure of Web sites to find how Web pages discussing brands and stars are connected. It uses high-betweenness Web sites returned to a search engine query for a brand or star name as a proxy for the significance of this brand or star.

KEYWORDS: Coolhunting, degree-of-separation search, TeCFlow, online metrics

1. INTRODUCTION
The well-know saying "on the Internet nobody knows that you are a dog" alludes to the perception that the Internet offers anonymity to its users. In reality, the opposite is true. The Internet has become a major communication channel for late-breaking news and to disclose innermost secrets. For example, when CBS published documents about George W. Bush's behavior during his military service, Republican bloggers quickly identified weak spots in the authenticity of the documents. This questionable evidence regarding George Bush's potential evasion of military service during the Vietnam War era ultimately lead to the early retirement of CBS news anchor Dan Rather. This incident is just one of many illustrating that today's news are made and disseminated on the Web and in the Blogosphere. This paper introduces a new Web mining approach which we call "Web Coolhunting" [27] making use of the fact that the Web has become a mirror of the real world, breaking latest news through active participation of millions of volunteers on Web sites such as Wikipedia, and political blogs such as dailykos and instapundit.

Large-scale phone polling through surveys to track popularity of politicians has been used for a long time to gauge public opinion. Our approach offers an automated and much cheaper way than polling people over the phone to achieve similar goals by analyzing the linking structure of Web sites and blogs. Using the Web as a mirror of the real word permits to automatically measure and track the popularity and attributes of brands and stars. It offers an efficient way to trace fame of brands and stars in the real world.

2. RELATED WORK
Popularized by Barabazi in his book "Linked" [2], there is a rich body of research on how the linking structure of the Web influences accessibility of Web pages [9, 13,15] and their ranking in search engines.

Visualization of Web structure and contents has been an active area of research since the creation of the Web. There are numerous systems for the static visualization and analysis of the link structure of the Web [5,6]. Inxight [18], Visual Insight [21], Touchgraph [20], Grokster [17], and Mooter [19] are all systems for the visualization of the linking structure of the Web, sometimes also offering a visual front end for search results.

In a related stream of work, researchers have been trying to predict the hidden linking structure based on known links [1,10]. Additionally, by looking at contents of Web sites, subspaces of the Web have been clustered by topics [4,11,12]. Combining these two lines of research, community Web sites have been mined to discover trends and trendsetters for viral marketing [14].

Our research focuses on a similar application – tracking the strengths of brands over time. For our analysis we are using the TeCFlow system [8] originally developed to mine e-mail networks to automatically generate dynamic social network movies.

3. DEGREE-OF-SEPARATION SEARCH
Our Web datamining approach combines two ideas: measuring betweenness centrality of Web sites as defined in social network theory, and doing degree-of-separation search, explained in the subsequent paragraph.

Betweenness centrality has originally been defined in the context of social network analysis [16]. It measures the knowledge flow in a social network as a function of the shortest paths, that is it looks at the percentages of all shortest paths in a network that go through a given node. Betweenness is a measure of the centrality of a node in a network. It may be characterized loosely as the number of times that a node needs a given node to reach another node. It is usually calculated as the fraction of shortest paths between node pairs that pass through the node of interest. It is defined as

$$b_k = \sum_{i,j} \frac{g_{ikj}}{g_{ij}}$$

where g_{ij} is the number of shortest paths from node i to node j, and g_{ikj} is the number of shortest paths from i to j that pass through k. Betweenness ranges from 0, for nodes that are totally peripheral, to 1, for nodes which are on all shortest paths.

Degree-of-separation search works by building a network map displaying the linking structure of a list of Web sites returned in response to a Google query. For example, a search to get the betweenness of "Hillary Clinton" works as follows:

1. Start by entering the search string "Hillary Clinton" into Google.
2. Take the top N (N is a small number, for example 10), of Web sites returned to query "Hillary Clinton".
3. Get the top N Web sites pointing to each of the returned Web sites in step 2 by executing a "link:URL" query, where URL is one of the top N Web sites returned in step 2. The Google "link" query returns the "significant" Web sites linking to a specific URL[1].
4. Get the top N Web sites pointing to each of the returned Web sites in step 3. Repeat step 4 up to the desired degree of separation from the original top N Web sites collected in step 2. Usually it is sufficient, however, to run step 4 just once.

Figure 1 illustrates the network map returned to the query "Hillary Clinton". The level-1 nodes are the ones connected directly to the query, i.e. the original search results. Level-2 nodes are the most highly ranked search results returned by the "link" query, to each of the top N level-1 nodes. Level-3 nodes are the most highly Google-ranked nodes returned by the "link" queries of each of the level-2 nodes.

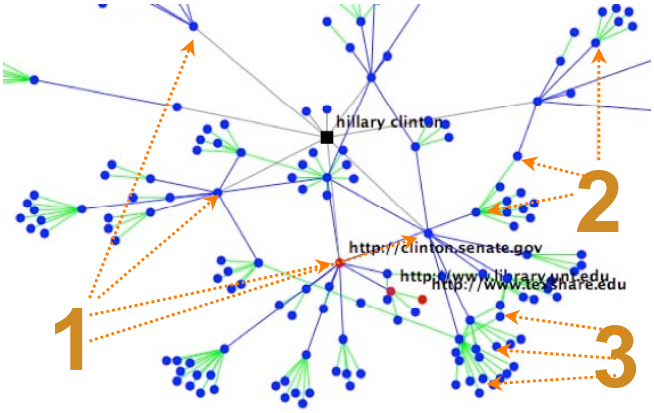

Figure 1. Degree-of-separation search for "Hillary Clinton"

Figure 1 already gives a visual overview of the betweenness of each of the level-1 and level-2 nodes. The more links a node has pointing to it, the more between it is. For example the node labeled http://clinton.senate.gov is linked by a group of level 2 nodes which themselves are linked by groups of level-3 nodes. This indicates that the node http://clinton.senate.gov will have fairly high betweenness itself.

The most between node in figure 2 is the search query "Hillary Clinton" itself, with a value of 0.61. The second most between node is indeed, as figure 2 illustrates, http://clinton.senate.gov with a betweenness value of 0.36. Some other high-betweenness nodes are www.ovaloffice2008.com and www.hillaryclinton.com.

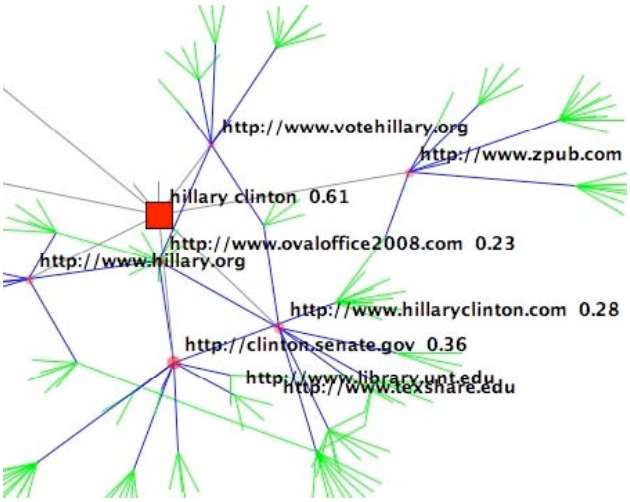

Figure 2. Betweenness of Web Sites indicated by size of square or circle

[1] We have not been able to find a precise definition of what makes a Web site "significant" for Google, but it seems that the linking Web sites themselves are linked to by other Web sites with a page rank larger than 0.

Top Google search results do not necessarily have highest betweenness centrality. Google sorts search results by the "Page Rank" algorithm [3], which looks at what Web pages link back to a particular page. It also weights the links to the page by the page rank of the originating page. In terms of social network analysis Google measures the in-degree of a page, that is the number of incoming links. Page Rank looks at the nearest neighbors of the page it is measuring. It includes page-rank of the neighbors, weighting incoming links higher from sites that themselves have a high page rank. Jon Kleinberg's HITS algorithm [25] is similar to Page Rank in that it also looks at static linking structure, but computing two local parameters "authority" and "hub" per page. Our approach based on betweenness, on the other hand, is basically a dynamic concept, because it looks at all the shortest paths within the local context network that are going through a particular node. A node, which has a high page rank therefore does not necessarily also have to exhibit high betweenness centrality.

4. COOLHUNTING ON THE WEB FOR STARS

Doing a degree-of-separation search is a quick way to find the most influential nodes in a relevant subset of the Web. Combining multiple datasets, each containing the degree-of-separation Web sites collected through querying a search engine for the name of one search term (the "star") permits to find the most central star in a group of stars by comparing the betweenness scores of the different stars. By combining the nodes returned by different degree-of-separation searches, we can compare betweenness of different stars, identifying the ones with the highest betweenness values. What this means is that they are the most linked, or most "talked about" on the Web or in the Blogosphere. This process will now subsequently be explained through tracking the trends of political candidates.

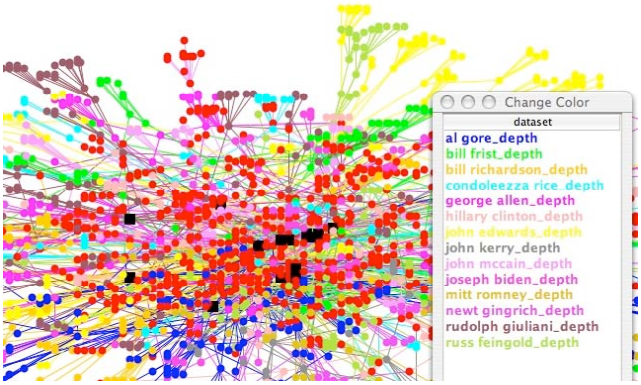

Figure 3. Combined degree-of-separation search results for 14 US Presidential hopefuls

Figure 3 displays Google query results of doing a degree-of-separation search for the leading 7 republican and 7 democratic contenders to become the next US President, as of end of August 2006. Each of the colors identifies the set of nodes returned by the Google queries for one of the presidential candidates, e.g. the Web sites returned to "Al Gore" are shown in blue. The red nodes are the Web sites returned by more than one query. Table 1 lists the results of the two most recent presidential polls as of end of August 2006 and compares them with the betweenness values of the candidates on the Web as displayed in figure 3.

Table 1. Results of US presidential polls Aug 2006 (source Wikipedia) and Web betweenness values

	Pew	*Am. Polling*	*Betweenness*
Democrats	*Aug 9-13*	*June 13-16*	*Web Aug 26*
Hillary Clinton	40%	36%	0.05
Al Gore	18%	-	0.1
John Edwards	11%	15%	0.1
John Kerry	11%	13%	0.05
Joseph Biden	6%	4%	0.02
Bill Richardson	4%	5%	0.06
Russ Feingold	2%	6%	0.01
Republicans			
Rudolph Giuliani	24%	21%	0.09
Condoleezza Rice	21%	30%	0.04
John McCain	20%	20%	0.03
Newt Gingrich	9%	8%	0.05
Mitt Romney	4%	7%	0.02
George Allen	-	5%	0.03
Bill Frist	3%	2%	0.06

Based on the poll values in table 1, we would expect Hillary Clinton and Rudy Giuliani to be the most between actors in our Web coolhunting analysis. The result is slightly different, however. While there are no surprises for Rudy Giuliani, Hillary is not really the top ranked democratic candidate by betweenness. This honor falls to Al Gore and John Edwards, who are tied for first place. The reason for non-candidate Al Gore's surprising popularity were the recent launch of his new movie "An Inconvenient Truth" about global warming, generating buzz for Al Gore not only as a politician, but also as a movie actor and environmentalist. Al Gore therefore connects different Web communities, or in the language of social networks, he bridges structural holes, leading to high betweenness[16].

The same evaluation by betweenness also permits to find the most relevant Web sites discussing presidential candidates. These Web sites also double up as

"kingmakers." Kingmakers are the Web sites that, through linking to a query, increase the betweenness of the original query through their own high betweenness. In our presidential polling analysis, en.wikipedia.org and www.ovaloffice2008.com are the most between Web sites. While it is not surprising that Wikipedia is very central, as all candidates take care to get their profiles entered and updated there, the central position of ovaloffice2008 comes as somewhat of a surprise. Looking at the individual query results for each candidate, Wikipedia, the candidates' own Web sites, and the sites of national newspapers such as the New York Times or the Washington Post rank higher. If the Web sites returned to the 14 different degree-of-separation queries are combined, however, a different picture emerges, with Wikipedia and ovaloffice2008 by far having the highest centralities. While the Google page rank of Wikipedia is 9 (out of 10), ovaloffice2008's Google page rank in August 2006 was only 5. Its betweenness in the context of presidential elections, however, is the second highest of all Web sites included in this analysis of presidential hopefuls. Ovaloffice2008 also includes a very active forum where citizens of different inclinations and party colors discuss strengths, weaknesses, and chances of the various candidates, further adding to the centrality of this Web site.

Collecting the contents of the Web sites returned to a degree-of-separation query and using the content retrieval and semantic mapping feature of TeCFlow [24] permits to obtain the top ranked terms describing the most relevant topics for each "star". Figure 4 shows the most significant terms on the Web sites collected for query "Hillary Clinton."

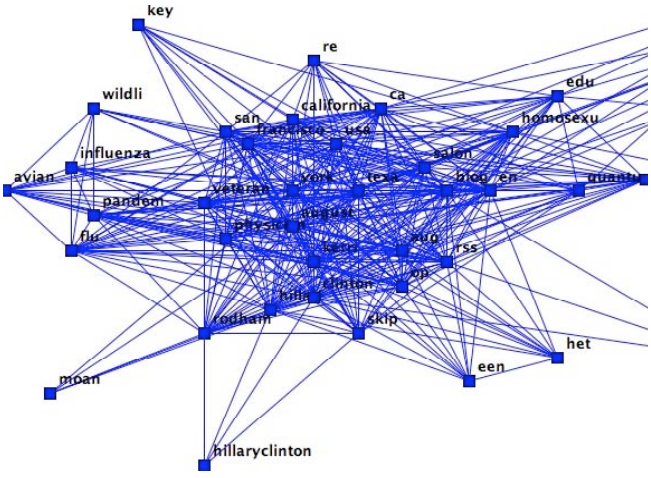

Figure 4 Key terms in the context of Hillary Clinton

As figure 4 illustrates, "California", "veterans", "John Kerry", "New York", "Texas", and "avian flu" are the key topics coming up when analyzing the context of Hillary Clinton in August 2006.

Repeating degree-of-separation searches over longer periods of time permits to measure changes in betweenness of the different "stars" to detect trends early.

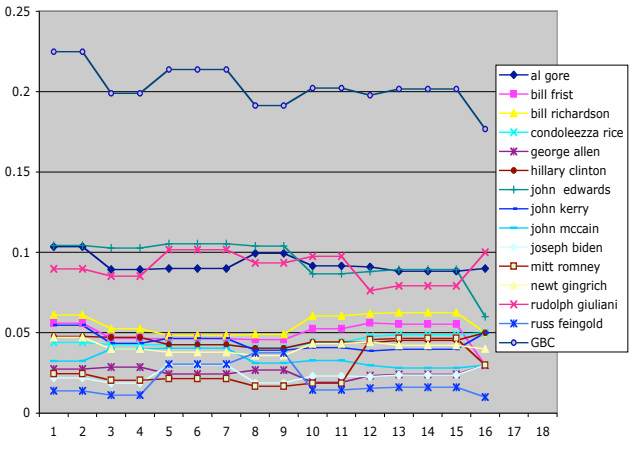

Figure 5. Changes over time in betweenness of 14 presidential contenders (x-axis is time in days, y-axis is betweenness) (GBC=Group Betweenness Centrality)

Figure 5 illustrates the changes in betweenness of the 14 presidential contenders over 16 days in August 2006. As the blue line shows, non-competing candidate Al Gore's lead is growing, while other leading democratic candidate John Edward's fortunes are declining. The big winner of the first week is candidate Russ Feingold, whose absolute betweenness and thus Web popularity is more than doubling before going down again in the second week. Leading republican candidate Rudy Giuliani is keeping his lead, in a neck-on-neck race with Al Gore. The overall centrality of the combined group analysis is slightly diminishing over the time period, indicating that there is no clear leader emerging thus far.

To see how incumbent but rather unpopular President George W. Bush fits in, we did two coolhunting queries on Sept 1, 2006, one comparing President Bush with Vice President Dick Cheney, and another against his 2000 Presidential adversary, Al Gore. As was to be expected, President George W. Bush generates more buzz on the Web than his Vice President Dick Cheney, with betweenness of 0.33 against betweenness of 0.25 of Dick Cheney. Bush's 2000 competitor, former Vice President Al Gore, however, would easily beat them both, at least on the Web, by scoring 0.42 against George W. Bush's 0.38 in a one-to-one comparison.

5. FINDING TRENDS AND TRENDSETTERS IN ONLINE FORUMS

So far, we have described coolhunting by degree-of-separation search on the Web using Google queries. The

same approach can also be used to find and track tends and trendsetters in online forums and blogs. The basic process consists of first finding online forums in the domain where trends should be tracked, and second parsing the forum and loading its social networking structure into TeCFlow.

In the first step degree-of-separation search is employed to find the online forums discussing a particular subject that the coolhunter is interested in. Because posts in online forums usually are heavily interlinked, online forums come up high by betweenness in degree-of-separation searches, even if the initial Google search did not contain the correct search terms. The discussion threads in the forums can then further be analyzed in TeCFlow to find the trendsetters and the trends they are discussing in their posts.

In the following example, latest trends in loudspeaker development are identified and tracked. Initially, a degree-of-separation search for "loudspeakers audio forum" is run, returning the "ecoustics forum" among its top ranked sites.

The "ecoustic.com" forum is then selected for an in-depth analysis. Different attributes of interaction can easily be extracted automatically from each forum thread, such as the thread name, the initiator of the thread, the nickname of each poster as well as her or his username, the timestamp of the post, and the contents. A social relationship between two people is constructed if a poster responds to the previous post.

Parsing all threads about "speakers" permits to create a social network of all participants to find influencers and gatekeepers as well as analyze the contents of their discussion about loudspeakers.

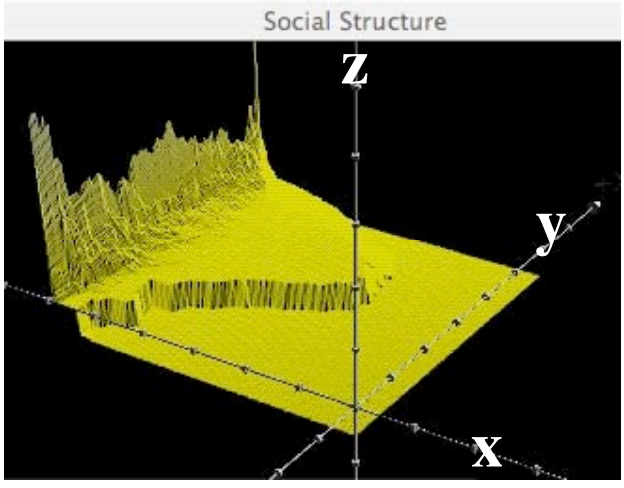

Figure 6. TeCFlow Social Structure of 4 months of discussion of speaker thread of ecoustics forum (x-axis=actors, y-axis=time, z-axis=actor betweenness)

Figure 6 shows the social structure [7] of four months worth of interaction data in the thread about loudspeakers in the ecoustics forum. It is easy to see that there is a small group of very central people (the peaks at left with high betweenness) dominating the discussion, and that the bulk of the participants are joining and leaving the discussion in the third month.

Observing the movie of the social network in the third month and filtering for the actors with highest betweenness identifies Nuck as the most central and influential person, and Jan Vigne as an important contributor. Once the most influential actors have been identified their influence can by tracked over time by looking at how their betweenness changes over time. Figure 7 illustrates when in time community leader Nuck and influencer Jan Vigne were most central.

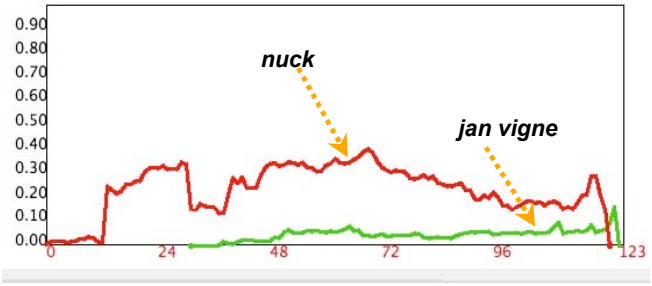

Figure 7. Evolution of betweenness over time for influentials Nuck and Jan Vigne

Besides tracking when who is most influential, we can also track when those influencers talk about what products – when is the buzz about a certain product the greatest?

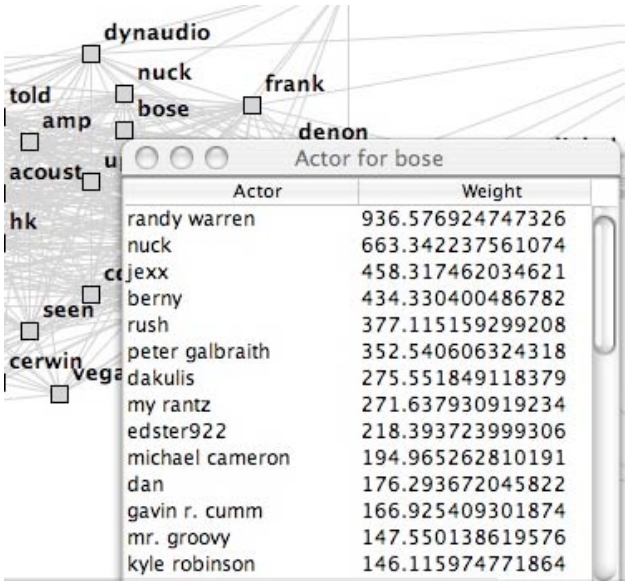

Figure 8. Concept map (term view), most significant actors on term "Bose" shown

Based on standard information retrieval procedures (tfidf) [24] TeCFlow computes a concept map of the most significant terms used in the online forum, and the relationships between those terms. Figure 8 displays a concept map for the 50 most significant terms over the entire 4 months period. In figure 8, the user has selected the term "Bose", showing that "Randy Warren" is the most influential person talking about Bose, with "Nuck" ranking second.

Using the TeCFlow taxonomy feature permits to identify the most significant messages about the brands we are most interested in. The taxonomy:

```
<level0>loudspeaker brand
<level1>Bose
<level1>Klipsch
<level1>Denon
<level1>Marantz
<level1>Polk
```

will categorize documents about the loudspeaker brands Bose, Klipsch, Denon, Marantz, and Polk into 5 clusters. Figure 9 displays the result of clustering the documents into these 5 categories. As figure 13 shows, the largest document clusters are for terms Bose or Polk.

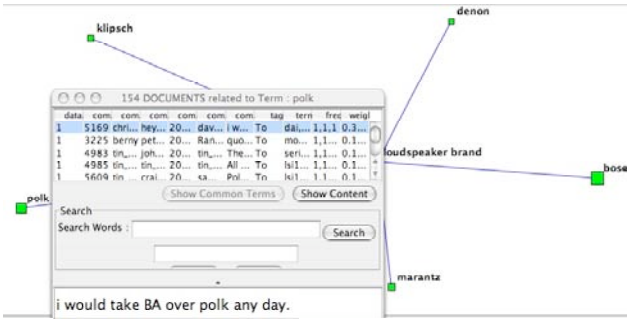

Figure 9. Auto-categorization by taxonomy, and most significant document about "Polk"

Figure 9 also shows a pop-up window brought up by the user, listing the documents about "Polk", sorted by significance. The most highly ranked document tells that Chris "would take BA over Polk any day", providing useful information to loudspeaker manufacturer Polk to get back to Chris and ask him why he got this unfavorable opinion about their product.

Finally, figure 10 shows when in time what brand was the most spoken about in terms of betweenness. For example, discussion about Paradigm speakers peaked around day 40. This means that at this point in time group betweenness centrality was high (the light yellow line), as was betweenness centrality of term "paradigm". In other words: all the discussion was about "paradigm." Discussion about Denon, on the other hand, was relatively flat for the first ninety days, but became a dominant issue in days 92 to 120. At this time other discussion topics were still ongoing, because group betweenness centrality was not as high as in the first 40 days.

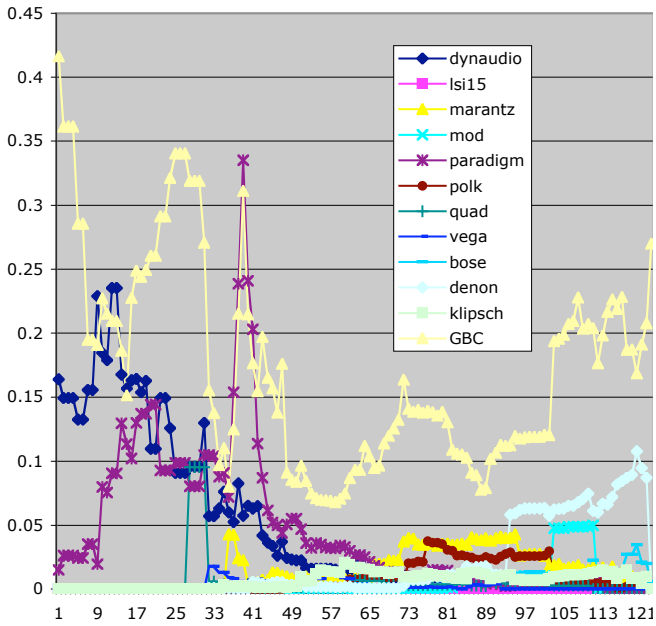

Figure 10. Evolution of betweenness over time of speaker brands in ecoustics forum, identifying trends (x-axis=time in days; y-axis=betweenness) (GBC=Group Betweenness Centrality)

6. RANKING BLOGS

To compare rankings by betweenness with conventionally obtained polling results, we compared the ten most popular blogs as listed on Technorati by their "favorite" ranking with their betweenness obtained through degree-of-separation search. These blogs have been manually nominated by Technorati users as their favorites. The more "favorite" nominations a blog gets, the higher is its ranking.

Figure 1 displays the results of coolhunting for Technorati's 10 most popular blogs as of Aug 20, 2006. The first amazing result is that the top three blogs obtained by the degree-of-separation search are not on Technorati's top ten list. They came up in the degree-of-separation search as having higher betweenness values than the ten blogs on the Technorati list that were put into TeCFlow as search input. Google.blognewschannel.com, www.techmeme.com and slashdot.org are all more central while not even being on the Technorati favorite list. Those three Web sites are all so-called Meme-trackers, trying to discover and rank the most popular new Blog posts, either

by displaying manually submitted posts in the case of Slashdot, or by user voting in the case of blognewschannel, or fully automatic ranking by techmeme.

Rank	Betw.	Blog
1	***0.1***	***http://www.boingboing.net***
2	0.15	http://www.43folders.com
3	***0.08***	***http://www.techcrunch.com***
4	0.12	http://www.lifehacker.com
5	0.11	http://www.cybernetnews.com
6	0.12	http://www.engadget.com
7	0.1	http://www.micropersuasion.com
8	0.1	http://www.alistapart.com
9	***0.05***	***http://scobleizer.wordpress.com***
10	0.1	http://googleblog.blogspot.com

Underperformers by Betweenness

Overperformers by Betweenness
"**Memetrackers beat 'em all**
0.33google.blognewschannel.com
0.17www.techmeme.com
0.17slashdot.org

Figure 11. Most popular blogs by coolhunting

As figure 11 shows, the ranking of Technorati and degree-of-separation search produce quite similar results, with a few notable exceptions. BoingBoing, TechCrunch, and Scobleizer came all up by at least 4 ranks lower in degree-of-separation search. While those three blogs contain interesting gossip about technology and gadgets, the more highly ranked Web sites 43folders, lifehacker, cybernetnews, and engadget all contain tips and tricks to make digital life easier, by posting novelty product reviews and recommendations about the latest technology and gadgets, with much less gossip than the three underperformers by betweenness. It seems that useful tips and tricks get more Web linkage and therefore higher betweenness than tech gossip. Overall, however, human based ranking and coolhunting by degree-of-separation search led to very similar results.

We finally used TeCFlow to collect blog posts for analyzing the contents of the three top ranked tech gossip blogs Slashdot, TechCrunch, and Digg. A TeCFlow concept map was generated by extracting the top-ranked terms for each blog (figure 12). Slashdot, TechCrunch, and Digg show surprising similarity.

Because these three blogs focus on the discussion of the latest technology trends, terms such as "Yahoo", "Google", "Microsoft", "Apple", and "Linux" occupy a central position. There are differences in the discussion of the most central gadgets, however. On the day when the blogs' content was collected, Slashdot focused on iPod, Digg talked about the Cybershot, and TechCrunch was discussing the new Chumby and Wablet tools.

Monitoring the central terms on these three blogs therefore offers yet another way of tracking the latest trends in high tech. The same principles could also be applied to politics, e.g. tracking leading US political blogs dailykos (for Democrats) and instapundit (for Republicans).

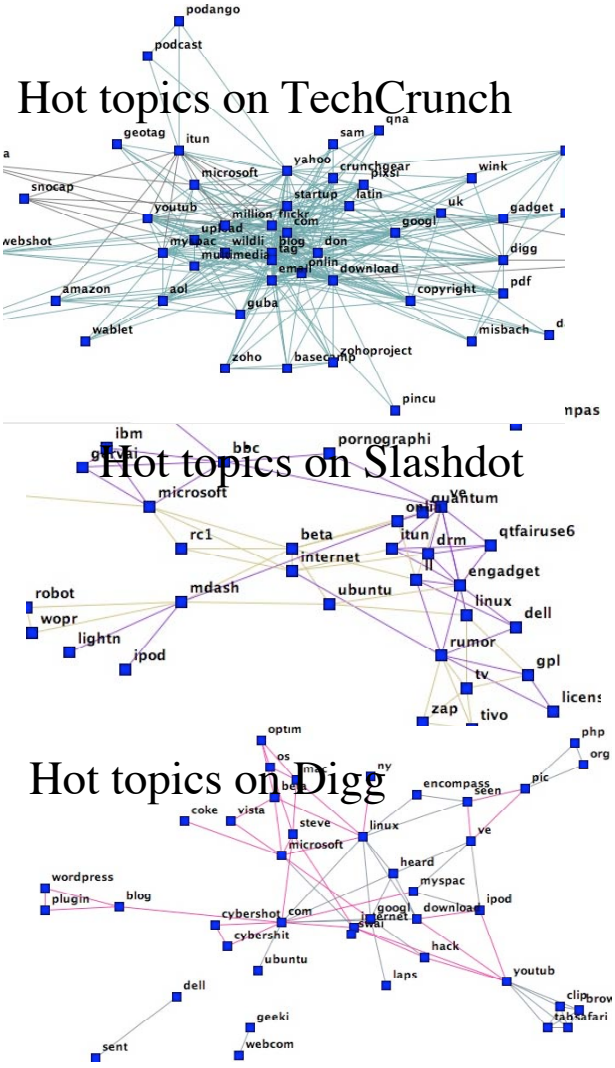

Figure 12. Most significant terms on Slashdot, TechCrunch, and Digg

7. CONCLUSIONS

Coolhunting by degree-of-separation search offers a novel way to search for trends and trendsetters. As illustrated in this paper, significance in the real world and significance on the Web as measured by coolhunting correlate quite well. If there are differences, they can be explained with characteristics of the virtual world and with social networking theory. Al Gore has higher ratings on the Web than in conventional polls because he bridges the different communities of politics, movies, and environmentalists. However such results correspond with behavior in the real world and insights gained in social network analysis, where it has been shown that people bridging different communities are high performers [26].

More systematic evaluations need to be done, comparing political campaign and poll results, Nielsen ratings, and

other external metrics with betweenness centrality parameters gained by coolhunting. We are also working on extending TeCFlow so that it will be able to track if a certain brand is spoken about in positive or negative terms. This will permit us to better understand the cases where degree-of-separation search and real-world metrics differ. Our first results are encouraging, opening up a new way to understand and measure how trends are created and how they disseminate on the Web.

ACKNOWLEDGEMENTS

I am grateful for many inspiring discussions about coolhunting and trend tracking with my colleagues Brent Cohen, Scott Cooper, Marius Cramer, Matt Guilford, Glen Kushner, Rob Laubacher, John Quimby and Detlef Schoder, and to Tom Allen and Tom Malone for their support and encouragement.

REFERENCES

[1] Al Hasan, Mohammad, Chaoji, Vineet, Salem, Saeed, & Mohammed Zaki. 2006. Link Prediction using Supervised Learning, Proc 2006 Workshop on Link Analysis, Counterterrorism and Security.

[2] Barabasi, L. Linked: How Everything Is Connected to Everything Else and What It Means. Plume, 2003

[3] Brin, S. Page, L. The anatomy of a large-scale hypertextual Web search engine. In Proceedings of the Seventh International World Wide Web Conference, Brisbane, Australia, 1998. Elsevier.

[4] Chakrabarti, S. Joshi, M, Kunal, P. Pennok, D. The Structure of Broad Topics on the Web. Proc. WWW 2002, Hawaii, 2002.

[5] Dodge, M. Kitchin, R. Atlas of Cyberspace, Pearson Education. 2002.

[6] Dodge, M. Kitchin, R. Mapping Cyberspace. Routledge, 2000.

[7] Gloor, P. Capturing Team Dynamics Through Temporal Social Surfaces, Proceedings of 9th International Conference on Information Visualisation IV05, London, 6-8 July 2005.

[8] Gloor, P. Zhao, Y. TeCFlow - A Temporal Communication Flow Visualizer for Social Networks Analysis, ACM CSCW Workshop on Social Networks. ACM CSCW Conference, Chicago, Nov. 6. 2004.

[9] Kleinberg, J. Authoritative sources in a hyperlinked environment. Proceedings of the Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, pages 668-677, Baltimore, MD, 1998. ACM Press.

[10] Liben-Nowell, David, & Jon Kleinberg. 2003. The Link Prediction Problem for Social Networks. In Proceedings of CIKM'03.

[11] Liu, B. Zhao, K. Yi, L. Visualizing Web Site Comparisons, WWW 2002, May 7-11, 2002, Honolulu, Hawaii, USA. 2002.

[12] Mukherjea, S. Organizing topic-specific Web information. Proc. Eleventh ACM Conference on Hypertext and Hypermedia, 2000.

[13] Pennock, D.M. Flake . G.W. Lawrence, S. Glover, E.J. Giles, C.L. Winners don't take all: Characterizing the competition for links on the web. Proceedings of the National Academy of Sciences, Volume 99, Issue 8, pp. 5207–5211, April, 2002.

[14] Richardson, M. Domingos, P. Mining Knowledge Sharing Sites for Viral Marketing. Proc. ACM SIGKDD, 2002.

[15] Smith, M. Fiore, A. Visualization components for persistent conversations, Proc ACM CHI. 2001.

[16] Wasserman, S. Faust, K. Social Network Analysis. Cambridge University Press, 1994.

[17] www.groxis.com

[18] www.inxight.com

[19] www.mooter.com

[20] www.touchgraph.com

[21] www.visualinsights.com

[22] Adar, E. Zhang, L. Adamic, L. Lukose, R. Implicit Structure and the Dynamics of Blogspace. Workshop on the Weblogging Ecosystem, 13th International World Wide Web Conference, May 18th, 2004

[23] Adamic, L. Adar, E. Friends and Neighbors on the Web. First Monday, 8(6), 2003.

[24] Gloor, P. Zhao, Y. Analyzing Actors and Their Discussion Topics by Semantic Social Network Analysis, Proceedings of 10th IEEE International Conference on Information Visualisation IV06, London, 5-7 July 2006

[25] Kleinberg, J. Authoritative sources in a hyperlinked environment. Proc. 9th ACM-SIAM Symposium on Discrete Algorithms, 1998.

[26] Gloor, P. Swarm Creativity: Competitive Advantage through Collaborative Innovation Networks. Oxford University Press, 2006.

[27] Gloor, P. Cooper, S. Coolhunting, Chasing Down The Next Big Thing. AMACOM, New York, 2007.