COVERAGE AND INDEPENDENCE
Defining Quality in Web Search Results

Panagiotis Metaxas and Lilia Ivanova
Computer Science Department, Wellesley College, U.S.A.

Keywords: Web search, adversarial information retrieval, web spam.

Abstract: Web search results enjoy an increasingly greater importance in our daily lives. But what can be said about their quality, especially when querying a controversial issue? The traditional information retrieval metrics of precision and recall do not provide much insight in the case of the web. In this paper we examine new ways of evaluating quality in search results: coverage and independence. We give examples on how these new metrics can be calculated and what their values reveal regarding the two major search engines, Google and Yahoo.

1 INTRODUCTION

The web has changed the way millions of people are being informed and make decisions. Most of them use search engines to access web information. Since people use search engines daily to make all kinds of financial, medical, political or religious decisions, quality of search results is of great importance. In the last ten years the two major search engines, Google and Yahoo, have gained the lion's share in the search market (Moran and Hunt, 2006).

But does higher market share implies higher search quality? Performance of information retrieval methods is traditionally measured in terms of precision (fraction of results that are relevant to a query) and recall (fraction of relevant items included in the results) (Manning et al., 2008). It is well known, however, that web searchers rarely look past the top-10 results (Silverstein et al., 1999). The web has enormous size. More that 10 billion pages are indexed by search engines and this represents a small portion of the static web. At the same time, search results on important issues are being heavily spammed (Gyuongyi and Garcia-Molina, 2005; Metaxas and Destefano, 2005). Therefore, high precision is easy to achieve but does not convey useful information, while recall cannot be computed accurately because of the enormous size of the web.

The problem of measuring search result quality becomes more interesting when searching for controversial issues. A controversial issue is one that has several possible relevant "answers", depending on one's point of view. Are the results users receive characterized by a reasonably comprehensive *coverage*? In other words, are the various opinions equally represented in the search results?

While search engines are trying to provide unbiased results, Search Engine Optimization (SEO) companies and web spammers are actively trying to force a search engine to list their own sites high on its search results. They do so using a variety of techniques, such as creating "link farms" (Gyuongyi and Garcia-Molina, 2005; Moran and Hunt, 2006). How *independent* are the top-10 results? For example, is it possible for a successful group of spammer to claim, not only the top spot in the top-10 search results, but a large group of them?

In this paper we take on the problem of defining coverage and independence of web search results. As far as we know, even though several papers have tried to define search quality (e.g., (Amento et al., 2000)) the metrics we introduce have never been addressed in the past. In the process we study the structure and density of the web neighborhood that supports each of the web search results according to each search engine, and we observe some interesting characteristics of these neighborhoods and of the way the search engines operate.

2 WEB SEARCH RESULTS OF CONTROVERSIAL ISSUES

To address these questions we decided to do a sequence of web searches on highly contested issues us-

Metaxas P. and Ivanova L. (2008).
COVERAGE AND INDEPENDENCE - Defining Quality in Web Search Results.
In *Proceedings of the Fourth International Conference on Web Information Systems and Technologies*, pages 106-113
DOI: 10.5220/0001529201060113

Table 1: Top-10 results of the Google search engine when given the query "HGH benefits" for August, 2007 and September, 2007. For each entry we have calculated the size of the backGraph as $(|V|,|E|)$ revealed by the Google API and the change between these two dates.

#	Google results	August, 2007	September, 2007	Change (%)	Notes
1	http://www.i-care.net/hgh-benefits.html	(415,426)	(346,353)	-16.6	G1=Y8
2	http://www.alwaysyoung.com/hgh/benefits/benefits.html	(161,167)	(133,137)	-17.4	G2=Y1
3	http://www.ghsales.com/ghsales2/hgh_growth_hormone _benefits.html	(1245,1314)	(981,1027)	-21.2	G3=G4
4	http://www.humangrowthhormonesales.com/ghsales2/index.html	(1245,1314)	(981,1027)	-21.2	(r)
5	http://www.hgh-human-growth-hormone.org/	(71,77)	(71,77)	0.0	G5=G6
6	http://www.hgh-human-growth-hormone.org/benefits-of-hgh.htm	(71,77)	(71,77)	0.0	(u)
7	http://www.csmngt.com/human_growth_hormone.htm	(1,0)	(1,0)	0.0	filter
8	http://www.associatedcontent.com/article/38893/human _growth_hormone_hgh_benefits_risks.html	(1749,2086)	(2905,3386)	+66.1	
9	http://www.hgharticles.com/	(1388, 1585)	(463, 506)	-66.7	G9=Y2
10	http://www.godswaynutrition.com/products/growthhormone.html	(502 ,506)	(419 ,423)	-16.6	

Table 2: Top-10 results of the Yahoo search engine when given the query "HGH benefits" for August, 2007 and September, 2007. For each entry we have calculated the size of the backGraph as $(|V|,|E|)$ revealed by the Yahoo API and the change between these two dates.

#	Yahoo results	August, 2007	September, 2007	Change (%)	Notes
1	http://www.alwaysyoung.com/hgh/benefits/benefits.html	(13151,16690)	(7829,9294)	-40.5	Y1=G2
2	http://www.hgharticles.com/hgh_benefits.html	(2933,3871)	(3380,4634)	+15.2	Y2=G9
3	http://www.hgh-pro.com/pro-blenhgh.html	(9587,11741)	(6376,7727)	-33.5	Y3=Y9
4	http://www.hghhomeopathic.com/HGH.html	(2137,2402)	(3125,3551)	+46.2	
5	http://www.hgharticles.com/	(2933,3871)	(3380, 4634)	+15.2	(u)
6	http://www.hghnstuff.com/faq-benefits-hgh.htm	(3063, 3444)	(5575, 6877)	+82.0	
7	http://linkspiders.com/HGH/benefits%20of%20hgh.htm	(3194, 3665)	(491, 495)	-84.6	
8	http://eyecare.freeyellow.com/hgh-benefits.html	(1041, 1204)	(1990, 2295)	+91.2	Y8=G1
9	http://www.hgh-pro.com/homeopathichgh.html	(6376,7727)	(6376,7727)	0.0	(u)
10	http://www.hgh.com/Descriptions/sec.aspx	(5471, 6460)	(7418, 9114)	+35.6	

ing the two most popular search engines, Google and Yahoo. For each of the queries we selected, one can expect that there are at least three possible answers: a "pro", a "con" and a "bal" (short, for "balanced") answer.

We argue that for a controversial issue, equal and comprehensive coverage in the top-N results is to have an equal number of pro, con and bal results. We will simply refer to this quality as "coverage" and we will define it below. For example, in the top-10 search results that search engines are giving back by default, equal and comprehensive coverage would be to have 3-4 results (or, on average, 3.3 results) from each category.

Let's assume we have k different categories for a complete coverage (above we have $k = 3$) and we have N results in the top-N slots. Let's further assume that category i received r_i results. We define as *coverage bias* the following quantity B:

$$B = \sum_{1 \le i \le k} |r_i - \frac{N}{k}| \qquad (1)$$

In other words, B is the distance of r_i from the expected number of results N/k. We note that bias B is bounded by:

$$B_{min} = 0 \le B \le N + (k-2)N/k = B_{max}$$

Specifically for top-10 search results, when we have comprehensive and equal coverage, we have minimum bias $B_{min} = 0$. At the other end, when one category takes all top-10 spots, the bias is maximized at $B_{max} = 13.3$.

The further bias B is from 0, the worst the coverage is. Therefore we can define as **coverage** C, the lack of bias, that is,

$$C = \frac{B_{max} - B}{B_{max}} \qquad (2)$$

Coverage C, therefore, has a value between 0 (one-sided coverage) and 1 (equal and comprehensive coverage).

The second metric we introduce is search result *independence*. To define independence we need first to examine the various ways in which search results can be dependent. We see three kinds of dependent results:

- *URL Dependency* is the situation when multiple entries in the top-N results are actually coming

Table 3: Top-10 results of the Google search engine when given the query ”Is ADHD a real disease” (August and September, 2007).

#	Google results	August, 2007	September, 2007	Change (%)	Notes
1	http://www.spiritofmaat.com/archive/oct1/drfred.htm	(1928,2135)	(1423,1614)	-25.2	G1=Y6
2	http://www.clickpress.com/releases/Detailed/2728005cp.shtml	(1223,1297)	(912,991)	-25.4	G2=Y9
3	http://www.wildestcolts.com/mentalhealth/stimulants.html	(496,515)	(873,925)	+76.0	G3=G4
4	http://www.wildestcolts.com/safeEducation/real.html	(496,515)	(873,925)	+76.0	(u)
5	http://web4health.info/en/answers/adhd-real-disorder.htm	(2545,2759)	(1688,1790)	-33.7	pro
6	http://www.adhdfraud.org/	(3280,3912)	(3955,4791)	+20.58	G6=Y3
7	http://www.adhdfraud.org/commentary/5-27-01-1.htm	(3280,3912)	(3955,4791)	+20.58	(u)
8	http://www.mykidsdeservebetter.com/adhd/disease.asp	(1590,1708)	(1557,1848)	-2.1	G8=Y2
9	http://www.virtualvienna.net/community/modules.php?name=News &file=article&sid=295	(1,0)	(1,0)	0.0	(c)
10	http://www.escolar.com/Escolar-Parenting_Articles/Escolar-is-adhd-a-real-disease.php	(5062,5764)	(incomplete)	-99.9	(c)

Table 4: Top-10 results of the Yahoo search engine when given the query ”Is ADHD a real disease” (August and September, 2007).

#	Yahoo results	August, 2007	September, 2007	Change (%)	Notes
1	http://www.mental-health-matters.com/articles/article.php?artID=849	(10263,13422)	(12321,16184)	-16.7	
2	http://www.mykidsdeservebetter.com/adhd/disease.asp	(7610,9556)	(5208,6202)	+46.12	Y2=G8
3	http://www.adhdfraud.org/	(7567,9093)	(9791,12232)	-22.7	Y3=G6
4	http://www.healthstatus.com/articles/Is_ADHD_A_Real_Disease.html	(13631,17032)	(12590,15645)	+8.3	
5	http://www.healthtalk.com/adhd/diseasebasics.cfm	(7011,8488)	(6704,7993)	+4.6	pro
6	http://www.spiritofmaat.com/archive/oct1/drfred.htm	(6425,8651)	(4242,5148)	+51.5	Y6=G1
7	http://www.nexusmagazine.com/articles/ADHDisbogus.html	(8424,10808)	(8266,10184)	+2.0	(l)
8	http://www.advancingwomen.com/diabetes/is_adhd_a_real_disease.php	(19291,25469)	(21150,27145)	-8.8	(c)
9	http://www.clickpress.com/releases/Detailed/2728005cp.shtml	(5675,6644)	(4920,5775)	+15.37	Y9=G2
10	http://ritalindeath.com/Against-ADHD-Diagnosis.htm	(2840,3440)	(2359,2604)	+20.4	(l)

from the same site URL, e.g., they correspond to different pages of the very same site, as it is defined by the domain URL. For example, in Table 1, results numbered 5 and 6 have URL dependency. In the Tables of this paper we mark one of the two URL dependencies with a (u).

- *Redirection Dependency* is the situation when two different site URLs resolve onto the very same location. Redirection is often used by web spammers who try to increase the visibility of a target site by creating many other sites that will point to the target site (Gyuongyi and Garcia-Molina, 2005). For example, in Table 1, results numbered 3 and 4 have redirection dependency. In the Tables we mark one of the two redirection dependencies with an (r).
- *Content Dependency* is the situation when the contents of two or more pages included in the search results are essentially the same. The contents may be surrounded by images and menus that are different and are stored on different web sites. This is also a trick used by spammers who try to increase visibility to a target site by creating entries in blogs or “news” sites that lack their own content (Gyuongyi and Garcia-Molina, 2005). For example, in Table 4, results numbered 4 and 8 have content dependency. In the Tables we mark one of the two content dependencies with a (c).
- *Link Dependency* is the situation when the supporting link structure in the web graph is substantially similar in two or more sites. Link dependency reveals “link farms’ of spammers (Gyuongyi and Garcia-Molina, 2005). This type of dependency has been studied extensively in the literature, and it is considered a major tool that the Search Engine Optimization industry is using to acquire high PageRank (Brin and Page, 1998). In this paper we will focus on the most basic similarity structure, namely the *Circular Link Dependency* between two or more sites. For example, in Table 4, results numbered 3 and 10 have circular link dependency, as do results numbered 6 and 7. See Figure 1. In the Tables we mark one of the two link dependencies with an (l).

Let’s consider the search results of some controversial query. Let’s assume that out of the top-N results, *u* results are URL dependent, *r* results are redi-

Table 5: Top-10 results of the Google search engine when given the query ”Morality of abortion” (August and September, 2007).

#	Google results	August, 2007	September, 2007	Change (%)	Notes
1	http://www.efn.org/ bsharvy/abortion.html	(5942,6710)	(5597,6350)	-5.9	G1=Y1
2	http://atheism.about.com/od/abortioncontraception/p/Religions.htm	(3922,4610)	(3429,4138)	-12.6	G2=G3
3	http://atheism.about.com/od/abortioncontraception/p/AtheistsAbort.htm	(3922,4610)	(3429,4138)	-12.6	(u)
4	http://ethics.sandiego.edu/Applied/Abortion/index.asp	(12045,14659)	(10433,13102)	-13.38	G4=Y8
5	http://rwor.org/a/038/morality-right-to-abortion.htm	(1667,2095)	(2687,3213)	+61.19	
6	http://www.answers.com/topic/abortion-debate	(7440,8657)	(5187,6276)	-30.28	G2 = Y6
7	http://ocw.mit.edu/NR/rdonlyres/054E18A6-DC9A-460E-826E-9EEC31A573E1/0/abortion.pdf	(3006,3373)	(2902,3211)	-3.5	
8	http://www.nrlc.org/news/2002/NRL06/pres.html	(6422,8119)	(5810,7295)	-9.5	G4 = Y8
9	http://www.manitowoc.uwc.edu/staff/awhite/mark_b97.htm	(2079,2353)	(1331,1560)	-36.0	
10	http://www.keele.ac.uk/depts/la/ documents/rfletcherFlagsubamd.pdf	(incomplete)	(incomplete)	0.0	

Table 6: Top-10 results of the Yahoo search engine when given the query ”Morality of abortion” (August and September, 2007).

#	Yahoo results	August, 2007	September, 2007	Change (%)	Notes
1	http://www.efn.org/ bsharvy/abortion.html	(17224,21732)	(15367,19843)	-10.8	Y1=G1
2	http://jbe.gold.ac.uk/5/barnh981.htm	(17491,22553)	(16137,21632)	-7.74	
3	http://www.rit.org/editorials/abortion/moralwar.html	(13434,17522)	(13752,17656)	+2.37	Y3=Y9 (u)
4	http://en.wikipedia.org/wiki/Morality_and_legality_of_abortion	(25672,35167)	(25672,35167)	0.0	
5	http://www.abort73.com/HTML/I-H-2-morality.html	(10282,14156)	(12243,16284)	+19.1	
6	http://atheism.about.com/od/abortioncontraception/p/Religions.htm	(6274, 8169)	(6274, 8169)	0.0	Y6=G2
7	http://www.ashby2004.com/abortion.html	(7370,9919)	(7370,9919)	0.0	
8	http://ethics.sandiego.edu/Applied/Abortion/	(13877,18248)	(13877,18248)	0.0	Y8-G4
9	http://www.rit.org/editorials/abortion/morality.html	(13752,17656)	(13752,17656)	0.0	Y9 = Y3
10	http://gospelway.com/morality/index.php	(2195,2721)	(incomplete)	0.0	

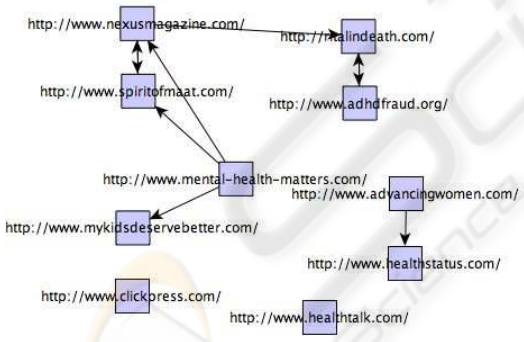

Figure 1: Link dependencies between the top-10 sites of Table 4.

rection dependent, c results are content dependent, and l results are link dependent. We define as **independence** of the search results the ratio of the non-dependent sites over N, or

$$I = \frac{N-(u+r+c+l)}{N} \quad (3)$$

Independence essentially measures the percentage of independent results in a collection of top-N search results. Note that in the formula above we penalize URL dependencies for all but one (their “representative”) of the identical URLs. We do the same for all other types of dependencies. Of course, we do not penalize results having multiple dependencies more than once.

Like coverage, independence is also a number between 0 (lack of independence) and 1 (all independent results). As with coverage, the higher the independence, the higher the quality of search results.

We should clarify that, even though it is a very important issue, we are not examining here the correctness or factual accuracy of search results. It is known that misinformation is a serious problem on the internet (Vedder, 2001; Berenson, 2000; Graham and Metaxas, 2003). But evaluating correctness or accuracy require significant amounts of time by qualified experts and there is no easy way to be done automatically by computer. We are mainly interested in exploring quality measures that can be computed using automatic or semi-automatic algorithms that will help search engines increase the quality of their search results.

3 EXPERIMENTAL RESULTS

An examination of a daily newspaper will reveal many controversial issues for which web users may search, and our colleagues have recommended many others. We chose to follow search results to three queries coming from the commercial, medical and political arena:

- Q1 (Commercial): *Human growth hormone (HGH) benefits*
- Q2 (Medical): *Is ADHD a real disease ?*
- Q3 (Political): *Morality of abortions*

Previous research (Moran and Hunt, 2006) suggests that Q1 is a highly contested query by small (mostly) companies selling steroids on the internet. Q2 is a highly contested query by activists claiming that ADHD is the result of conspiracy by psychiatrists. Finally, Q3 is a perennially important issue in the USA on which every politician must take a position and an issue elections might be decided on.

The question we wanted to address is: Do we get good coverage and independent responses when searching on these issues using the two major search engines, Google and Yahoo?

To evaluate our hypothesis we first conducted the three searches using the Google and Yahoo search engines in August, 2007. We analyzed the search results and computed the coverage and independence of each search.

It has been shown (Ntoulas et al., 2004) that the link structure of pages is evolving at a very fast pace, faster than the page contents themselves. We wanted to check this observation with regards to the web site back links. Do they also change fast? To do that, we followed the link structure of the supporting web graph twice, in August and September 2007, with our results reported here. We continue monitoring on monthly basis.

We examined several questions regarding the comparison of the results between the two search engines. The results of our work appear in the subsections below. First, we explain how we computed the supporting web graphs of each search result.

We define as the *supporting web subgraph G* of depth d for some URL U according to search engine S the graph that is computed using S's back links for d iterations.

To compute the supporting web subgraphs of each search result, we created a java program, `backGraph`, that, given a particular URL U as input, it first collects the set of links $L(U)$ of sites pointing to U according to search engine S. Since we cannot search the whole web from scratch to determine these links, we used the Search APIs provided by Yahoo (Yahoo, 2006) and Google (Google, 2003). This collection corresponds to back link depth $d = 1$.

We continued computing the sets $L^2(U) = L(L(U))$ for depth $d = 2$ and $L^3(U)$ for depth $d = 3$ using each of the two APIs. We stop at depth $d = 3$ following (Metaxas and Destefano, 2005) that shows is to be sufficient for calculation of the supporting web graph. Going further would strengthen our results but would also require significantly larger computational time.

More formally, the algorithm and the parameters we used are as follows:

```
Input:
   s[i] = each of top-10 results' URL
   d = Depth of back link search (d=3)
   B = Number of backlinks to record (B=100)
   SE = {YahooAPI, GoogleAPI}

Algorithm:
  S = {s[i]}
  Using depth-first-search for depth d do:
    Find the set U of sites linking to sites in S
       using the SE for up to B backlinks/site
    S = S + U

Output:
   Graph recorded in S for each API
```

One may expect that $L(L(L(U)))$ would create a tree of sites pointing to U in no more than 3 links. It turns out that the graph created is not a tree but a directed graph G with a bi-connected core (called BCC in (Metaxas and Destefano, 2005)). It has been shown that this graph reveals the deliberate link support that activists and spammers are using to promote a particular web page so that this page scores high on a search engine's query results. We then evaluated the size of each graph $|G|$ by calculating the number of nodes (sites) $|V|$ and edges (links) $|E|$ as well as the size of its BCC. In this paper, due to space considerations, we do not present the BCC data.

3.1 Overall Results

Table 7 shows the coverage and independence results of the three queries. We observe that in the commercially important Q1 and the medically important Q2, coverage from both search engines is very low. Results from both search engines show also low independence. On the other hand, for the politically important Q3, coverage and independence is high for both search engines, with Yahoo scoring a little higher in coverage than Google.

In the late 1990's search engine algorithms used to differ significantly. It has been argued that these days,

Table 7: Results for coverage and independence metrics. $C(G)$ and $C(Y)$ are the coverage scores in the Google and the Yahoo results, respectively. $I(G)$ and $I(Y)$ are the independence scores for the Google and Yahoo results, respectively.

Query	bal	pro	con	$C(G)$	bal	pro	con	$C(Y)$	$I(G)$	$I(Y)$
Q1: HGH benefits	0	10	0	0.0	0	10	0	0.0	0.8	0.7
Q2: ADHD real disease	0	1	9	0.1	0	1	9	0.1	0.7	0.7
Q3: Morality of Abortion	5	3	1	0.7	3	2	4	0.8	0.9	0.9

most search engines use very similar search algorithms. Our results provide evidence for this. About a third or more of the results are shared for each query, providing evidence that there is significant overlap in the results reported by the search engines.

The size of the support web subgraph is one area where the search results differ significantly between the search engines. Yahoo's supporting web graph size is significantly larger than Google's. Given the popularity and reputation of Google, one might have guessed the opposite.

We believe that the reported sizes of the supporting web graphs by Google are affected by some filtering of the back links. This is strongly evidenced by the fact that several supporting web graphs for Google have trivial sizes (e.g., result 7 in Q1 and result 9 in Q2), and this can be explained with the existence of result filtering.

Next, we discuss results for each query in some detail.

3.1.1 Q1 Results

We observe the following:

The size of support web subgraphs reported by Yahoo is 7 times greater than that reported by Google (total of 49886 nodes vs 6848 nodes). In the period of the two months we monitored the sizes of the supporting web subgraphs, we saw small-to-medium percentage variation in the size of the Google graphs and wide range for the Yahoo graphs. Despite that, they seem to almost agree on the top result (G2 = Y1) as well as in other results (e.g., G1=Y8, G9=Y2).

All of the top-10 results for both Google and Yahoo seem to be coming from companies that sell steroids online (the "pro" case). There are no results from medical authorities or research articles that refer to the medical or legal problems from the use of steroids (the "con" case). There are also no "balanced" views represented in the top-10 results. Coverage, therefore, is very low. We believe that these results reveal the work of very successful commercial SEOs who have dedicated lots of resources to gain from the lucrative HGH industry.

3.1.2 Q2 Results

We observe the following:

As we mentioned, Q2 coverage results are largely similar to Q1 results for both search engines. The size of support web subgraphs reported by Yahoo is, again, far greater than that reported by Google (total of 88737 nodes vs 19901 nodes, or 4 times greater).

These results reveal low coverage, with only one of the top-10 entries differing from the overall "con" direction of the results. There are no "balanced" results included. Interestingly, for both search engines, the "pro" result occupies position 5! The results for Yahoo reveal low link independence as four of the top-10 results are forming two circular link farms (Figure 1).

We believe that when a controversial issue is below the horizon of current news awareness, such as the ADHD issue at the time of the search, activists can be successful in getting the top spots in the relevant queries. The situation seems to be a bit different for issues that have enormous visibility, such as the next query.

3.1.3 Q3 Results

We observe the following:

Yahoo's sizes of support web subgraphs is roughly equal to Google's, (total of 125376 nodes vs 113463 nodes for the first nine results) in big contrast with the results in Q1 and Q2. Interestingly, both engines also agree on the top spot (G1=Y1), which represents a "balanced" opinion of the query. They both include "pro-choice" and "pro-life" results, while devoting the remaining top-10 entries on opinion gatherers (such as about.com and wikipedia.org).

Over the two month period we observe much smaller variation of the sizes of the supporting web graph, especially for Yahoo. It is not the case that this was due to lack of news for the issue of abortion at the time of the searches. The abortion issue is always at the top of the political agendas of the US candidates for office, and has high visibility on the news. We conjecture that for such highly sensitive issues, the search engines "tune" their results so that they will present wider coverage of opinions. We have seen this happening in the past in a variety of well publicized queries (such as "miserable failure").

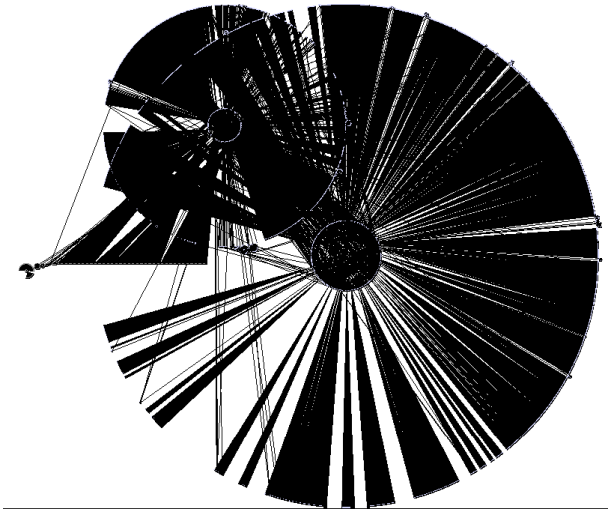

Figure 2: The graph induced by the backlinks of Google's 1st result and Yahoo's 1st result for Q3 (G1=Y1). The graph has been drawn to emphasize the two BCCs. The upper left group is composed by the majority of Google's 5942 sites, the lower right group is composed by most of Yahoo's 17224 sites, while the group in the middle left side (appears as a horizontal line) is mostly composed of 470 sites in the intersection of the two groups.

3.2 Graph Overlaps

Another question we are studying is how much overlap exists in the backGraphs produced by the two search engines on the same result. Given that Yahoo reports many more links than Google, one might expect that there is a major overlap between their backlink graphs. We have found this not to be the case. See Figure 2 for a graphical representation of a typical example. The two search engines seem to report a largely different set of links in their results: Between G1's 5942 sites and Y1's 17224 sites, there is a mere overlap of 470 sites!

4 CONCLUSIONS & FUTURE WORK

With this paper we have started an effort to evaluate quality in search results in terms of two important metrics, coverage and independence. We have found that when searching controversial issues, both of these quality metrics can be low. However, in high-visibility queries the search engines may tune the results manually, as they have been criticized in the past for not offering a balanced view.

We also found that Yahoo reports a much greater number of back links than Google. Does it mean that Yahoo sees a much larger portion of the web graph in the neighborhood we were searching? This is rather unlikely since the sites in the web neighborhoods we examined seem to have a great interest to be seen. On the other hand, there is a high degree of overlapping results in the top-10 results in all queries, suggesting the employment of very similar algorithms and heuristics by the two search engines. One possible explanation is that Google is filtering the back links it is reporting. More research is needed to test these conclusions, and we are currently working on it.

Queries on controversial issues will continue to play an important role in web search. Search engines need to be more sensitive in providing high quality results that include high coverage and high independence between results. Evolution of search results over time is also important as it may reveal information about the part of the web that is being manipulated by spammers, activists and SEOs. Future research will hopefully shed some light in this area.

ACKNOWLEDGEMENTS

Part of this work was supported by a Brachman-Hoffman grant. We would like to thank Sociology Professor Rosanna Hertz for providing expert evaluation on the coverage of Q3.

REFERENCES

Amento, B., Terveen, L., and Hill, W. (2000). Does authority mean quality? Predicting expert quality ratings of web documents. In *Proceedings of the Twenty-Third Annual International ACM SIGIR Conference*

on Research and Development in Information Retrieval. ACM.

Berenson, A. (2000). On hair-trigger wall street, a stock plunges on fake news. New York Times.

Brin, S. and Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 30(1–7):107–117.

Google (2003). The Google API, google, inc. http://code.google.com/apis/.

Graham, L. and Metaxas, P. T. (2003). "Of course it's true; i saw it on the internet!": Critical thinking in the internet era. *Commun. ACM*, 46(5):70–75.

Gyuongyi, Z. and Garcia-Molina, H. (2005). Web spam taxonomy. In *Proceedings of the First International Workshop on Adversarial Information Retrieval on the Web*, Chiba, Japan.

Manning, C., Raghavan, P., and Schultze, H. (2008). *Introduction to Information Retrieval*. Cambridge Press, Cambridge, UK, (forthcoming) edition.

Metaxas, P. T. and Destefano, J. (2005). Web spam, propaganda and trust. In *Proceedings of the First International Workshop on Adversarial Information Retrieval on the Web*, Chiba, Japan.

Moran, M. and Hunt, B. (2006). *Search Engine Marketing*. IBM Press, New Jersey, USA.

Ntoulas, A., Cho, J., and Olston, C. (2004). What's new on the web? the evolution of the web from a search engine perspective. In *Proceedings of the WWW 2004 Conference*, New York, NY.

Silverstein, C., Marais, H., Henzinger, M., and Moricz, M. (1999). Analysis of a very large web search engine query log. *SIGIR Forum*, 33(1):6–12.

Vedder, A. (2001). *Misinformation through the internet: Epistemology and ethics*. Intersentia, Antwerpen, Gronigen, Oxford.

Yahoo (2006). The Yahoo search API, yahoo, inc. http://developer.yahoo.com/search/.