

Agathe Benichou  
CS200- Computers and Society  
Professor Dahl  
Due: April 3rd 2017

### The Influence of PageRank, Google's infamous algorithm

Within the past decade, Google has become the most popular and utilized search engine. Google's topnotch query performance combined with its effortless usage makes it superior compared to other search engines. Designed to explore and index the billions of web pages using an applicable number of keywords, Google is able to produce millions of accurate queries per day. As the growth of the web and the amount of amateur web users expands exponentially, Google has had to scale up its technology to keep up with the increasing demand. While other search engines have developed different methods to rank web pages, the quality of Google queries is based on PageRank which is a search algorithm developed by Google. PageRank has drastically altered how we as a society produce, share, consume and interact with data and has driven Google to be as successful as it is today.

In the 1990s, search engine technology was growing quickly as a result of an increase in regular people gaining access to computers as well as the amount of documents on the web. Sergey Brin and Larry Page had realized that they were going to need faster technology for their Stanford based search engine, Google. A majority of search engine companies at the time sorted web pages using keywords typed in by the user but this was a very poor way of figuring out which page was the most relevant for a given set of words. Most users will often only look through the first couple of search results, so it is imperative that these results include only the best and most relevant documents. Knowing this, Brin and Page had to improve their search

engine so that users would be satisfied within the first couple of query results. In a 1997 paper called *The Anatomy of a Large-Scale Hypertextual Web Search Engine*, Brin and Page describe their search engine: “The Google search engine has two important features that help it produce high precision results. First, it makes use of the link structure of the Web to calculate a quality ranking for each web page, this ranking is called PageRank. Second, Google utilizes link to improve search results” (Brin and Page). Brin and Page utilized link popularity to develop their PageRank algorithm which calculates and assigns a rank to each document.

The rank of each document depends on the rank of other documents on the web pointing to it. Essentially, “Page had realized that packed into the linked structure of the Web was a lot more data than most search engines made use of. In fact, a Web page linked to another page could be considered a ‘vote’ for that page. It wasn’t just which pages linked to which that Brin and Page were interested in - the position of a link on the page, the size of the link, the age of the page- all of these factors mattered”(Pariser, 31-32). Professor Eric Roberts at Carnegie Mellon University described PageRank as an “idealized random web surfer which rates web pages objectively and mechanically, measuring the human interest and attention devoted to them” (Roberts). Web pages want to be linked to other web pages with a high rank in order to increase their rank, or popularity, among the web. Web page developers produce data knowing that a random, unconnected web page will not have any relevance on the web. For this reason, companies often pay to have their site advertised on frequently used websites. This recursive algorithm is known as the core of Google and has been attributed as the “algorithm that drove the company to worldwide dominance” (Pariser, 31).

As time went on and Google continued to drive the search engine market, Brin and Page continued to strive to give their users the best and most relevant pages. An important part of knowing which web page would best interest a user is knowing the user, their interests and their attitude. As Eli Pariser puts it in his book, *The Filter Bubble*: “The challenge was getting enough data to figure out what is personally relevant to each user. You have to get to know a person’s behavior over a sustained period of time.” (Pariser, 33). Brin and Page had realized that the solution to relevance was more data, and they were determined to preserve every item of data; no matter how small or seemingly futile it was. As a result, Google started providing products and services which required users to make an account and log in. Such services include Gmail, Google Photos, Google Calendar, Google Drive but these are just a few of many. Gmail is a very popular email service and it was the first of Google's products. “By getting users to log in, Google got its hands on an enormous pile of data: the hundreds of millions of emails Gmail users send and receive each day. Google can use this to cross reference each user's email and behavior with the links they clicked in the Google search engine.” (Pariser, 33). At a time where information was mainly passed on through email and telephone, Google used Gmail to learn as much as it could about its users. If a user was subscribed to a Republican email list, Google knows what kind of politically-leaning sites the user might be interested in. For users, this data is what helps provide relevant and personalized news and results. Today, Google Chrome is the most widely used Internet browser. Its sleek design makes it appealing to users but for Google, it’s ability to store cookies is an enormous tool. A user writing an essay on Google Docs via their Gmail account on the influence of Google’s PageRank Algorithm, for example, will be able to receive useful webpages with uncanny relevance.

There are several interesting academic journals that use Google's PageRank algorithm to model social networks and user interactions. In *Identifying Key Users in Online Social Networks: A PageRank Based Approach*, Julia Heidemann states that PageRank can be used to evaluate social influence on popular social media websites. Social media has not only become a part of everyday activity, it has also become a major news source. The popular online social networks have "evolved into a global mainstream medium that generate an increasing social and economic impact." (Heidemann, 1). These sites realized that the identification of users online is important for retaining users as well as for efficient advertising strategies. If Facebook users want their story to become popular, they share their story and hope that other people do the same. Daniel Romero states in his article *Influence and Passivity in Social Media*: "The amount of information flowing through social media forces the members of networks to compete for attention and influence by relying on other people to spread their message" (Romero, 1). This has given people the opportunity to create and share content on a large scale without having to develop web sites. However, the popularity of a user is key to their influence on a site. The more followers or friends users have, the more attention they receive, the more their content propagates through the network and the more popular they become. Biao Xiang's article, *PageRank with Priors: An Influence Propagation Perspective*, discusses how if the participation in social networks continues to expand, companies and governments will have to increase their effort in getting their ideas, policies and products to users. In reality, the large scale of these networks makes it difficult for anyone to receive attention. As a result, "the data in these networks provide unparalleled opportunities for researchers to understand the human world." (Xiang, 3).

While PageRank is known to be a revolutionary algorithm, critics of Google have described the company as a monopoly that uses the algorithm as a tool in invasive information retrieval. Matteo Pasquinellis' article, *Google's PageRank Algorithm: A diagram of cognitive capitalism and the rentier of the common intellect* heavily criticizes Google's expanding dominance and compares Google to Big Brother. He states "The battle against the accumulation of data operated by PageRank and Google reminds of the social struggles against the traditional forms of monopoly and accumulation of capitals"(Pasquinelli, 3). Pasquinelli argues that Google has such a firm grasp on the global search engine market that its competitors do not stand a chance. He states that this widespread dominance over the market comes from the "data surveillance made possible only thanks to a monopoly of data that are previously accumulated through the PageRank algorithm" (Pasquinelli,1). PageRanks aides Google in providing its users with more relevant searches but at a cost considering Google has more information about us than we would probably be comfortable with. However, most are not worried about this since the well known motto of Google is "Don't be evil", as stated in their code of conduct, which was first introduced in 2000.

Today, Google monitors every piece of data about us it can get. It is now an all encompassing Internet giant which offers a Google assistant, mapping service, cell phones, wifi, music as well as over 50 other products and services. It can store your pictures, your contacts and essentially your entire digital life. Google does not just provide all these services to make more money, it is spreading its influence and storing an alarming amount of information about almost

everyone who has access to the Internet. Google's large servers contain a real-time copy of the Web which it sifts through for every query. However, it seems like the majority of average Internet users do not mind being constantly tracked with their personal data being stored. The convenience of a fast and applicable web browsing experience seems to dominate the fact that Google knows more about the average person than some of their friends. "That Google has moved from PageRank - which was just a small part of the project - to a mix of PageRank and personalization represents a shift in how Google understands relevance, meaning and the future of the Web." (Pariser, 33).

## Sources

- Andersson, Erik, and Per-Anders Ekstrom. "Investigating Google's PageRank Algorithm." (2009): n. pag. Uppsala Universitat. Web. 19 Mar. 2017.
- Brin, Sergey, and Larry Page. "The Anatomy of a Large-Scale Hypertextual Web Search Engine." *The Anatomy of a Search Engine*. Stanford University, n.d. Web. 19 Mar. 2017.
- Heidemann, Julia, Mathias Killer, and Florian Probst. "Identifying Key Users in Online Social Networks: A PageRank Based Approach." (n.d.): n. pag. Web. 19 Mar. 2017.
- Pasquinelli, Matteo. "Google's PageRank Algorithm: A Diagram of Cognitive Capitalism and the Rentier of the Common Intellect." *Deep Search*. N.p., n.d. Web. 19 Mar. 2017.
- Pariser, Eli. *The Filter Bubble: What the Internet Is Hiding from You*. London, England: Penguin, 2012. Print.
- Roberts, Eric. "Google's Pagerank Algorithm." : n. pag. Carnegie Mellon University. Web. 19 Mar. 2017.
- Rogers, Ian. "Pagerank Explained Correctly with Examples." *Princeton University*. The Trustees of Princeton University, n.d. Web. 19 Mar. 2017.
- Romero, Daniel M., Wojciech Galuba, Sitaram Asur, and Bernardo A. Huberman. "Influence and Passivity in Social Media." *SSRN Electronic Journal* (n.d.): n. pag. HP Labs. Web. 19 Mar. 2017.
- Sobek, Markus. "Google PageRank - Introduction." *Google PageRank - Introduction*. N.p., n.d. Web. 19 Mar. 2017.
- "Page Rank Algorithm Explained - Search Engine Optimization." *Page Rank Algorithm Explained - Search Engine Optimization*. N.p., n.d. Web. 19 Mar. 2017.
- Xiang, Biao, Qi Liu, Enhong Chen, Hui Xiong, and Yi Zheng. "PageRank with Priors: An Influence Propagation Perspective." *The 23rd International Joint Conference on Artificial Intelligence* (2015): n. pag. Web. 19 Mar. 2017.