A STATISTICAL ANALYSIS OF HIGH-TRAFFIC WEBSITES

by

JOSHUA MADDEN

B.S., Kansas State University, 2011
M.S.I.S., Baylor University, 2012

A THESIS

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

A.Q. Miller School of Journalism and Mass Communications
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2014

Approved by:

Major Professor
Steven Smethers, Ph.D.

Copyright

Abstract

Although scholars have increasingly recognized the important role of the Internet within the field of mass communications, little research has been done analyzing the behavior of individuals online. The success or failure of a site is often dependent on the number of visitors it receives (often called “traffic”) and this includes newspapers that are attempting to direct larger audiences to their websites. Theoretical arguments have been made for certain factors (region, social media presence, backlinks, etc.) having a positive correlation with traffic, but few, if any, statistical analyses have been done on traffic patterns. This study looks at a sample of approximately 300 high-traffic websites and forms several regression models in order to analyze which factors are most highly correlated with Internet traffic and what the nature of that correlation is.

Table of Contents

List of Tables v
Acknowledgements vi
Chapter 1 - Introduction 1
Applications for Mass Communications 4
Advertising 4
Newspapers 6
Summary 6
Chapter 2 - Literature Review 8
Regional and Geographic Factors 9
Demographic and Psychological Factors 14
Technical and Infrastructural Factors 17
Unique Contributions of Study 20
Chapter 3 - Methodology 21
Chapter 4 - Findings 26
Statistical Findings 27
Chapter 5 - Conclusion 45
Answers for Research Questions 47
Research Question 1 47
Research Question 2 48
Research Question 3 49
Research Question 4 49
Points for Future Research 50
Bibliography 52
Introduction 52
Literature Review 53
Conclusion 56

List of Tables

Table 4.1 – Chi-square summary table for United States Quantcast rankings/Alexa global rankings 27
Table 4.2 - Chi-Square results for Chinese Quantcast rankings/Alexa global rankings 28
Table 4.3 - Chi-square results for Saudi Arabia Quantcast rankings/Alexa global rankings 28
Table 4.4 - Regression model for U.S. traffic 30
Table 4.5 - Coefficient table for U.S. only traffic regression 31
Table 4.6 - Regression summary table for China-only traffic 32
Table 4.7 - Coefficient table for China-only traffic regression 33
Table 4.8 - Regression summary table for Saudi Arabia-only traffic 34
Table 4.9 - Coefficient table for Saudi Arabia-only traffic regression 35
Table 4.10 - Regression summary table for Saudi/Chinese combined traffic 36
Table 4.11 - Coefficient summary table for China/Saudi Arabia-combined traffic model 37
Table 4.12 - Regression summary for Saudi Arabia/United States combined traffic model 38
Table 4.13 - Coefficient table for United States/Saudi Arabia-combined traffic regression 39
Table 4.14 - Regression summary table for United States/China-combined traffic model 40
Table 4.15 - Coefficient summary table for United States/China-combined regression 41
Table 4.16 - Regression table summary for combined United States/China/Saudi Arabia traffic model 42
Table 4.17 - Coefficient summary table for United States/China/Saudi Arabia combined traffic regression 43

Acknowledgements

First and most importantly, I would like to thank the faculty at Kansas State University, specifically the faculty at the A.Q. Miller School of Journalism and Mass Communications, for their help with this study. I appreciate all of the contributions and suggestions from the members of my committee, Steve Smethers, Tom Hallaq and Curtis Matthews. The committee has offered a tremendous amount of feedback in a very timely fashion and for that I am extremely thankful. I would also like to thank Nancy Muturi, Angela Powers, Wes Wise and Sue Zschoche for their help with finding articles and meeting deadlines. Secondly, I would also like to thank the faculty at Baylor University, where I received my M.S.I.S., for their continued help and support after graduation. Specifically, I would like to thank Jonathan Trower and Randall Vaughn for their help with this project and others. Finally, I would like to thank many of my colleagues, friends and family, including Lance Cahill, Mortiz Cleve, Ambrosia Franklin, Brad Forsythe, Elizabeth Grilliot, Andrew Korkmas, Jacob Lubeck, Mark Line, Mark Lyle, Joe and Tammy Madden, Levi Madden, Kelsey Madden, Laura Monisse, Brent Mordhorst, Eric Nehm, Ron and Kayla Phillips, Daniel Pope, Bethany Quesnell, Noel Rizzuti, Feiyang Sun, Imani Thanji, Charles and Betha Regan, Tom Reust, Bryan Roth, Jeanette Valdez and Stephanie and Kevin Waddell, for their support and feedback throughout this project. Many of the people listed above (specifically Cleve, Grilliot, Korkmas, Monise, Nehm, Quesnell and Reust) have taken an increased role in other projects in order to enable me to get this study done on time. For their selflessness and support, I am extremely thankful. I would like to conclude this acknowledgement section by saying that although I am truly thankful for the contributions of everyone listed above, any errors or faults in this study are mine entirely.

Chapter 1 - Introduction

The potential for exponential growth in traffic for newly developed websites truly allows for millions of dollars to be made almost overnight. Facebook is perhaps the most famous example of this phenomenon – it was started as a project in 2004 by college roommates.[1] Yet fewer than 10 years later, the site is the second most-viewed website in the world.[2]

However, criteria traditionally used to measure a site's reliability seems to indicate that Facebook should not enjoy the popularity that it does. For example, Alexa's measurements list the site as "Very Slow" with an average load time of 3.195 seconds per page and emphasizes that 81% of sites are faster.[3] This is significantly slower than their competitor, Orkut.com, a social network run by Google with high popularity rankings in Brazil. Yet despite the fact that Okrut's speed is listed as "Average" with a significantly lower load time of 1.932 seconds per page and only 58% of sites being faster, Okrut.com is ranked as 2,067th globally, meaning that there are 2,066 more popular sites worldwide.[4]

The load time for a given page is generally used as an indicator of a website's quality and is a key factor in the site's placement on search engines such as Google. Google makes no effort to hide the fact that this is a key metric they use to analyze a site, writing on a public developer page that "fast and optimized pages lead to higher visitor engagement, retention, and

---

[1] Carlson, Nicholas. "At Last -- The Full Story Of How Facebook Was Founded." *Business Insider*. Business Insider, 5 Mar. 2010. Web. 24 Oct. 2013.

[2] "Facebook.com." *Facebook.com Site Info*. Alexa, n.d. Web. 24 Oct. 2013.

[3] Ibid.

[4] "Orkut.com." *Orkut.com Site Info*. Alexa, n.d. Web. 24 Oct. 2013.

conversions."[5] Considering only this metric for estimating traffic, it would not be unreasonable to assume that Orkut.com should see higher traffic than Facebook. The two sites have a similar format and target demographic and Orkut.com is actually the faster site in terms of load times. Yet Facebook dwarfs Orkut in terms of traffic. Clearly something else is impacting the traffic of the two sites.

In fact, in an article for BusinessWeek, Arik Hesseldahl finds strong evidence that access to high Internet connection service is not expanding nearly as quickly as one might assume.[6] He finds that a great deal of consumers, even in developed nations such as the United States, still do not have access to a broadband Internet connection.[7]

There are few who would dispute that these individuals are actually using the Internet in large numbers – in fact, Hesseldahl finds that users often still access the Internet through lower speed connections such as dial-up connections through phone lines – and these individuals are undoubtedly decreasing the average load time for majorly popular sites such as Facebook. This idea is supported by the research of political scientists Norris and Inglehart, who find that the connection speeds in other nations are seldom faster than those in the United States.[8]

---

[5] "Make the Web Faster." — Google Developers. Accessed October 24, 2013. https://developers.google.com/speed/pagespeed/.

[6] Hesseldahl, Arik. "Bringing Broadband to the Urban Poor." *BusinessWeek*, December 31, 2008. http://www.businessweek.com/stories/2008-12-31/bringing-broadband-to-the-urban-poorbusinessweek-business-news-stock-market-and-financial-advice.

[7] In a conversation with the author of this study, Hesseldahl, he explained that, to his surprise, annual measurements for broadband access rates do not exist within the United States.

[8] Norris, Pippa, and Ronald Inglehart. *Cosmopolitan Communications*. New York: Cambridge University Press, 2009.

The problem here is the user's connection speed and not any major infrastructural issue for the site itself. So why is load time speed so frequently used as a predictor in traditional traffic estimation models? It is apparent that a more modern approach to understanding what drives traffic is worth researching.

One of the key aspects of exponential traffic growth online is the idea that any site has the potential to go "viral." In his book *Growth Hacker Marketing: A Primer on the Future of PR, Marketing, and Advertising*, online marketer Ryan Holiday quotes Jonah Berger, who writes that "virality isn't luck. It's not magic. And it's not random. There's a science behind why people talk and share. A recipe. A formula, even."[9] In essence, this project is an attempt to model that formula. It seeks to understand how such explosive growth in traffic happens and why it happens on some websites and not others.

Interestingly, part of the answer may lie with social media websites like Facebook, which allow users to post links to other websites, easily driving potentially substantial traffic to sites by sharing these links with friends and family who might not see them otherwise. Many political science researchers have found that social media platforms can be valuable in mobilizing support at the polls during election years.[10] Other researchers have found that presence online can be used to predict turnout at protests as well.[11] It seems unlikely that these sites do not have a similar impact on driving traffic between websites. In fact, other political scientists have found

---

[9] Holiday, Ryan. *Growth Hacker Marketing: A Primer on the Future of PR, Marketing, and Advertising*. Portfolio, 2013.

[10] Cardenal, Ana S. "Why mobilize support online? The paradox of party behaviour online." *Party Politics* 19, no. 1 (2013): 83-103.

[11] Bell, Sam, David R. Cingranelli, Amanda Murdie, and Alper Caglayan. 2013. "Coercion, Capacity, and Coordination: Predictors of Political Violence." *Conflict Management and Peace Science*, forthcoming.

evidence to support this very phenomenon. Hafner-Burton finds that individuals are able to use social media to send links and information to key figures in government (particularly judges) in order to lobby for outcomes they wish to see on human rights issues.[12] This means that by sending links to other individuals, users of social networks like Twitter are able to increase the traffic directed towards those sites – often in substantial numbers.

All of this combines to suggest that a study is needed to answer some important research questions. It is important find out what factors are positively or negatively correlated with Internet traffic, whether these factors strongly or weakly correlated with traffic, and whether or not factors like Internet presence – such as social media presence and backlinks – will have a strong and positive correlation with traffic.

Applications for Mass Communications

Although this topic is relevant to information systems research, the relevance to the field of mass communications is perhaps somewhat less obvious. In truth, the relevance of the subject is two-fold. Firstly, advertisers are paying a great deal of attention to the behavior of individuals online, and the traffic of a website has a strong relationship with what people will be willing to pay to advertise on a specific site. Secondly, as newspapers are shifting more and more of their focus online, it is important to recognize whether or not people are actually viewing the information online.

Advertising

This is perhaps a relatively clear point, but it warrants inclusion nonetheless: advertisers are indeed paying a great deal of attention to the behavior of individuals online, particularly with

[12] Hafner-Burton, Emilie. *Making Human Rights a Reality*. Princeton, NJ: Princeton University Press, 2013.

regards to social media. As with most forms of mass communications, the reach of a website has a strong relationship with what people will be willing to pay to advertise. This relationship has been studied quite a bit previously for other forms of media. Radio advertisers are willing to pay more based on the reach of the stations in question. The relationship between advertising and ratings for a radio station has been long acknowledged within the mass communications field.[13] The same is true for television and newspapers. Advertising rates for television stations are strongly correlated with the station's rating.[14] A similar relationship can likely be theorized for websites and the advertisements on the site. It seems likely that the amount an advertiser is willing to pay to place an advertisement on a specific website is positively correlated with the traffic that website receives.

The relationship between advertisers and media has even been extended to the study of those who own multiple types of media outlets.[15] As it is not unusual for a radio or television station to also operate a website on which they charge for advertising, this study fits in well with the existing literature on the subject.

Previous studies have found that advertisers are also interested in the reach and effectiveness of existing forms of online advertising, such as sponsored links.[16] Sponsored links are a common form of advertising online, particularly for search engines, which are often some of the highest trafficked sites online. As this study focuses exclusively on high-traffic websites, it

---

[13] Soley, Lawrence C., Jesse E. Teel, and Leonard N. Reid. "A Comparison of Influences on Fixed and Grid Radio Advertising Rates." *Journal of Advertising* 9, no. 4 (1980): 15-19.

[14] French, Warren A., and J. Timothy Mcbrayer. "Arriving at Television Advertising Rates." *Journal of Advertising* 8, no. 1 (1979): 15-18.

[15] Ferguson, James M. "Daily Newspaper Advertising Rates, Local Media Cross-Ownership, Newspaper Chains, and Media Competition." *The Journal of Law and Economics* 26, no. 3 (1983): 635-54.

[16] Chen, Jianqing, De Liu, and Andrew B. Whinston. "Auctioning Keywords in Online Search." *Journal of Marketing* 73, no. 4 (2009): 125-41.

may yield important results with regards to this phenomenon as well as the relationship of advertising and audiences more generally.

Newspapers

As newspapers are shifting more and more of their focus online, it is important to recognize whether or not people are actually reading such websites. Newspapers are increasingly placing more and more of an emphasis on their online publications. In fact, the online versions of newspapers have been seen as one of the primary competitors for traditional print outlets.[17]

The online versions of newspapers can focus on different types of content, such as music streaming and video, which is subsequently attracting a different form of advertiser to the online versions of traditional papers.[18] This shift to more digital oriented forms of media has even impacted more traditional forms of media contained in newspapers, such as political cartoons, which are increasingly being animated using Flash and other software.[19] All of this, however, is somewhat of theoretical situation unless users are actually accessing the websites in large numbers, which is the focus of this study.

Summary

The importance of this research for the study of both newspapers and advertising combines to suggest that factors affecting website access is an important topic from a mass

---

[17] Gentzkow, Matthew. "Valuing New Goods in a Model with Complementarity: Online Newspapers." *The American Economic Review* 97, no. 3 (June 01, 2007): 713-44. Accessed March 24, 2014. JSTOR.

[18] Clemons, Eric K. "Newly Vulnerable Markets in an Age of Pure Information Products: An Analysis of Online Music and Online News." *Journal of Management Information Systems* 19, no. 3 (December 01, 2002): 17-41. Accessed March 24, 2014. JSTOR.

[19] Baumgartner, Jody C. "Polls and Elections: Editorial Cartoons 2.0: The Effects of Digital Political Satire on Presidential Candidate Evaluations." *Presidential Studies Quarterly* 38, no. 4 (December 01, 2008): 735-58. Accessed March 24, 2014. JSTOR.

communications perspective. The relationship between audience and advertising rates has long been acknowledged for most forms of mass media, but the study of advertising in relation to online traffic is still a relatively unexplored area within mass communications. Likewise, although many studies have shown that newspapers are shifting more and more of their focus online, few have actually shown that individuals are actually viewing these sites. These two areas could benefit from further exploration and study, which is one of the primary goals of this research.

Chapter 2 - Literature Review

Systems theory is the most promising theoretical approach for studying online traffic growth. Previous researchers (particularly in the field of law) have taken the arguments of famed systems theorist Jurgen Habermas and attempted to apply them to the empirical study of the Internet.[20] This theoretical approach to the study of mass communications shows potential specifically within the context of cyberspace. Ansari and Mela have already done promising research in this area, showing that by viewing the interaction between users, websites and email subscriptions as a system, they were able to much more accurately predict potential click-through rates (CTR) than more traditional models were able to do.[21] This is a measurement that Chinese inventors have used to help tailor search results and it is strongly correlated with traffic rates.[22]

In fact, Ansair and Mela find that in the context of their research, "if the likelihood of observing i's data is relatively low when user i's random effect is used for user i, the more likely it is that the i's random effect is a new value."[23] This means by viewing each user as unique – and subsequently, with unique viewing habits – they are able to more accurately predict that user's click-through rates. This seems to support the idea that as more data becomes available, scholars will be able to predict traffic with higher degrees of accuracy. Ansari and Mela's findings[24] seem

---

[20] Froomkin, A. Michael. "Habermas@Discourse. Net: Toward a Critical Theory of Cyberspace." *The Harvard Law Review Association* 116, no. 3 (January 2003): 749-873.

[21] Ansari, Asim, and Carl F. Mela. "E-Customization." *Journal of Marketing Research* 40, no. 2 (2003): 131-45.

[22] "WIPO PUBLISHES PATENT OF ALIBABA GROUP HOLDING FOR "RANKING SEARCH RESULTS BASED ON CLICK THROUGH RATES" (CHINESE INVENTORS)." *US Fed News Service, Including US State News,* Dec 27, 2013. http://search.proquest.com.er.lib.k-state.edu/docview/1470893859?accountid=11789.

[23] Ansari, Asim, and Carl F. Mela. "E-Customization." *Journal of Marketing Research* 40, no. 2 (2003): 131-45.

[24] Ansari, Asim, and Carl F. Mela. "E-Customization." *Journal of Marketing Research* 40, no. 2 (2003): 131-45.

to suggest that the application of an approach grounded in systems theory could yield important findings within the study of newer forms of mass communications.

In fact, the general trend within the literature tends to include more complex models, with less emphasis on more simplistic exemplars. This is certainly the case in industry, where more complicated statistical frameworks are constantly being developed.[25] Although there is a relative beauty in simplistic explanations, such efforts have found to be lacking in the context of researching Internet growth.

Regional and Geographic Factors

As mentioned earlier, Norris and Inglehart found that there are a great deal of factors impacting an individual's behavior online – often factors such as Internet connection speeds that are beyond their control.[26] These speeds my vary based on where people reside – for example, Hong Kong currently boasts the fastest average connection speeds and has now for several years.[27] It seems likely that these speeds will allow users to visit more sites, subsequently meaning that a site targeted towards an audience in Hong Kong might see a higher number of page views than a site with a targeted audience living in a different region.

Within the field of electrical engineering, researchers have found that technology limits what people can do online and how many websites they can access within a given timeframe.[28]

---

[25] "WIPO PUBLISHES PATENT OF ALIBABA GROUP HOLDING FOR "RANKING SEARCH RESULTS BASED ON CLICK THROUGH RATES" (CHINESE INVENTORS)." *US Fed News Service, Including US State News,* Dec 27, 2013. http://search.proquest.com.er.lib.k-state.edu/docview/1470893859?accountid=11789.

[26] Norris, Pippa, and Ronald Inglehart. *Cosmopolitan Communications*. New York: Cambridge University Press, 2009.

[27] "HK Still Leads by Average Broadband Speeds." *Telecom Asia (Online)* (Oct 25, 2013). http://search.proquest.com.er.lib.k-state.edu/docview/1445008998?accountid=11789.

[28] Cao, Jin, William S. Cleveland and David X. Sun. 2004. Bandwidth Estimation for Best-Effort Internet Traffic. *Statistical Science*, *19*(3), 518-543.

Other researchers in economics have found that even factors like the language of a keyboard can impact an individual's behavior online – as keyboards are often built around the Latin alphabet, which is not the predominant alphabet for written communications in areas such as the Arabic world.[29]

Although factors like these undoubtedly have an impact on the traffic of certain websites – particularly those marketed towards regional areas without stable Internet connections – often no variable is used to attempt to measure the impact of such limitations, despite the existence of a variety of indicators that actually could measure these characteristics. As previously mentioned, measuring Internet usage in the Arab world due to the linguistic differences is problematic, yet few studies have actually included a variable to account for these linguistic factors. Such a variable might, in fact, explain a substantial portion of the difference between the Arab region and the rest of the world.

In addition, geographers have found that Internet access rates are rapidly expanding across the globe. Not only that, but access rates have been growing consistently for the last decade.[30] While Internet access may have become somewhat of a standard to be expected in the developed world, these access rates are rapidly increasing elsewhere. Bargh and McKenna found that in 2001, only 1 in 250 individuals in Africa used the Internet, while the access rates in North America and Europe at the same time were roughly 1 in 3.[31] Likewise, a study in rural Scotland found that even those living in rural areas with no interest in broadband access might be willing

---

[29] Warf, Barney and Peter Vincent. 2007. Multiple geographies of the Arab Internet. *Area*, *39*(1), 83-96.

[30] Malecki, Edward J. 2002. "The Economic Geography of the Internet's Infrastructure." *Economic Geography* 78:399–424

[31] Bargh, John A., and Katelyn Y. A. McKenna. "The Internet and Social Life." *Annual Review of Psychology* 55, no. 1 (2004): 573-90.

to adopt the technology if they were the subject of targeted initiatives.[32] Thus, even groups that are seemingly resistant to broadband technology now can be seen as potential users in the future. This finding seems to be further supported by similar research conducted three years later.[33]

While there may someday be concrete technical limits on Internet growth (a point that will be elaborated on later) for the time being, Internet access rates appear on track to continue exponentially. To put it simply, the increase in Internet access should lead to increase in demand for content online. One would expect that this alone could explain a large portion of the exponential growth of certain sites online, yet this factor is almost never considered in traditional models that look at Internet traffic growth. It logically follows from these findings that this should translate to larger numbers of users on all websites. Thus, while the number of unique users might actually be increasing for the vast majority of Internet sites, the percentage of Internet users may not be increasing at the same rate.

Conversely, with such rapid expansion of Internet access, sites may be seeing an increase in traffic while the percentage of Internet users who visit their site may actually be decreasing. This theoretical argument is echoed in the findings of Bargh and McKenna, who find that users tend to gravitate towards websites dedicated towards regional concerns.[34] To put it simply, rapid expansion of Internet access rates in Africa might translate to a small increase in usage of American websites, but it will more likely result in a large increase in usage of African websites.

---

[32] Howick, S., and J. Whalley. "Understanding the Drivers of Broadband Adoption: The Case of Rural and Remote Scotland." *Journal of the Operational Research Society* 59, no. 10 (2007): 1299-311.

[33] Johnson, David Lloyd. "Re-Architecting Internet Access and Wireless Networks for Rural Developing Regions." Order No. 3559800, University of California, Santa Barbara, 2013. http://search.proquest.com.er.lib.k-state.edu/docview/1354493124?accountid=11789.

[34] Bargh, John A., and Katelyn Y. A. McKenna. "The Internet and Social Life." *Annual Review of Psychology* 55, no. 1 (2004): 573-90.

This disparity between regions seems to counter some of the ideas proposed within a subset of systems theory called world-systems theory, which argues that the trend in economics is towards a more globalized economic system.[35] While a trend towards a more globalized economic system may be true for the larger economy – and such a question is well-beyond the scope of this study – it appears that with specific websites, there is evidence to suggest that the trend may be more towards regional concerns instead of global ones. In fact, researchers have found that nationalist and xenophobic sites within China have flourished[36] while resistance to censorship has been minimal.[37] These regional factors combine to suggest that as people are drawn more towards regional sites, they are unlikely to be as concerned when their access to international sites (which are often the ones censored by the Chinese government) is limited. This is just one of many examples of how regional and demographic concerns of users can impact large-scale traffic trends. Perhaps unsurprisingly, the vast majority of studies done on Internet adoption rates have focused on Europe and the United States. Studies focusing on Asian countries, such as the academic articles referenced above, are much more the exception than the rule with regards to this research. Yet researchers have found differences in behavior between residents of Asia and the Western world and that these differences are important and arguably

---

[35] Straussfogel, Debra. "World-Systems Theory: Toward a Heuristic and Pedagogic Conceptual Tool." *Economic Geography* 73, no. 1 (1997): 118-30.

[36] MacKinnon, Rebecca. "China's Internet: Let A Thousand Filters Bloom." *China's Internet: Let A Thousand Filters Bloom*, June 28, 2005. http://yaleglobal.yale.edu/content/chinas-Internet-let-thousand-filters-bloom.

[37] MacKinnon, Rebecca. 2007. "Flatter World and Thicker Walls? Blogs, Censorship And civic Discourse in China." *Public Choice* 134, no. 1-2: 31-46.

understudied.[38] This is despite the fact that recent studies have shown many regions within Asia – such as Hong Kong and South Korea – are consistently showing the highest average connection speeds in the world.[39]

Another example of the importance of factoring in regional analysis is the implementation of Internet access in public libraries. Studies have been done analyzing the impact of such investments on the behavior of the individuals who use the Internet in libraries and other public access locations.[40] Library access has become somewhat of a standard feature of such institutions at this point, particularly in the United States, where the inclusion of Internet connections is nearly universal. In this sense, whether or not an individual lives in a region where public libraries offer Internet connections could impact that person's odds of using the Internet on a frequent basis, particularly if that individual lives at or below the poverty line. This is perhaps most clearly important in the United States where Internet equipped public libraries are relatively common.

Before moving on from the focus on regional concerns, it is worth noting that a common misconception among non-scholars is that the United States has some of the most advanced Internet infrastructure in the world. In fact, this is not necessarily true. In 2005, an important early piece of research found that the U.S. is 13th in terms of global rankings for Internet access rates and speeds.[41] Among developed nations, particularly with the advanced economic levels of

---

[38] Tan, Margaret, and Thompson S.H. Teo. "Factors Influencing the Adoption of the Internet." *International Journal of Electronic Commerce* 2, no. 3 (1998): 5-18.

[39] "HK Still Leads by Average Broadband Speeds." *Telecom Asia (Online)* (Oct 25, 2013). http://search.proquest.com.er.lib.k-state.edu/docview/1445008998?accountid=11789.

[40] Bertot, John Carlo, Charles R. McClure, and Paul T. Jaeger. "The Impacts of Free Public Internet Access on Public Library Patrons and Communities." *The Library Quarterly* 78, no. 3 (2008): 285-301.

[41] Bleha, Thomas. "Down to the Wire." *Foreign Affairs* 84, no. 3 (June 2005): 111-24.

development present in the United States, this makes the United States somewhat of an anomaly worth studying more in-depth. In fact, as pointed out in the study referenced above, as of 2005, the United States was the only industrialized nation that did not have "an explicit national policy for promoting broadband."[42]

According to a more recent study from Akamai Technologies in 2013, "the U.S. placed 11th on the list of the twenty areas with the fastest Internet connections. The country's large landmass, which makes it difficult to fully cover with high-speed fiber, is partly to blame."[43] So while the United States may have slightly risen in these rankings with regards to access speeds, it is still well behind other developed nations.

Demographic and Psychological Factors

On a related note, Madden has found that there are a significant number of variables that can impact a person's attitude towards the Internet, including demographic information, political views and, specifically, education within the field of information systems.[44] This attitude can impact an individual's behavior online, such as the sites they view, the amount of time they are likely to spend on each website and how many unique pages they will view within a single domain. This echoes earlier research, which suggests that the unique psychology of an individual (and even a person's genetic traits) can have a profound impact on how a consumer interacts

---

[42] Ibid.

[43] Milian, Mark. "Top 20: Where to Find the World's Fastest Internet: The Download." Bloomberg.com. July 22, 2013. Accessed April 16, 2014. http://www.bloomberg.com/slideshow/2013-07-23/top-20-where-to-find-the-world-s-fastest-internet.html.

[44] Madden, Joshua E. *Information Leaks and Information Systems*. Thesis, Baylor University, 2012.

with computers and the Internet.[45] Information systems theorists have found that, within the workplace in the United States and Finland, widespread implementation of Internet access within major firms has translated into major changes in behavior among members of the workplace.[46] This evidence would seem to suggest that the answer to what drives traffic online is, in fact, a very complex one and likely contains a great deal of statistically significant variables, including ones that have not been traditionally studied in this context.

This research into individualist behavior online combines to suggest that the role of human-computer interaction (often considered a subset of systems theory) in the study of Internet usage and behavior could be an important one. While this claim seems somewhat intuitive, relatively few studies have taken human-computer interaction findings and theories and applied them towards predicting future Internet traffic. The findings have often been somewhat more general and not applied specifically to traffic rates of individual websites. For example, an older study of Internet adoption rates within companies in Singapore found that organizational and technological factors do play a major role in terms of Internet adoption, yet these findings have not been applied to actual traffic measurement models.[47]

This echoes previous research on other behavioral factors, which has shown that changes within the structure of a company is perhaps best-described as a two-way street in that the changes to the infrastructure often impact the behavior of the employees, but the behavior of the employees can also impact and shape a company's infrastructure. For example, in a classic work

---

[45] Kock, Ned. "The Psychobiological Model: Towards a New Theory of Computer-Mediated Communication Based on Darwinian Evolution." *Organization Science* 15, no. 3 (2004): 327-48.

[46] Lyytinen, Katie, and Gregory M. Rose. "The Disruptive Nature of Information Technology Innovations: The Case of Internet Computing in Systems Development Organizations." *MIS Quarterly* 27, no. 4 (December 2003): 557-96.

[47] Teo, Thompson S.H., Magaret Tan, and Wong Kok Buk. "A Contingency Model of Internet Adoption in Singapore." *International Journal of Electronic Commerce* 2, no. 2 (Winter 1997/1998): 95-118.

on business management practices, business researcher William Ouchi has found that management can have a profound impact on how employees act within a business setting and that, in turn, how the employees act impacts the structure of management.[48]

He also finds strong evidence to suggest that demographic information – such as the individual's nationality – can have profound impacts on this relationship.[49] It seems relatively likely that widespread implementation of Internet access, which is arguably one of the most profound infrastructural changes a company can make from a technical perspective, should show similar effects. This research was done before Internet usage was as widespread as it is today, which is part of the reason this study is necessary. Many of Ouchi's concepts and theories may still be applicable, but they have not as yet been applied to Internet-related research.

The study of behavior online, however, is hardly limited to political scientists and information systems theorists. Economists Goolsbee, Lovenheim, and Slemrod found that access to Internet, and how it is used, has wide ranging impacts on various economic factors in both the public and private sector.[50] They find that individuals often use the Internet to spread information on cigarette prices across state lines. This, too, seems to support the idea that it is the sharing of links that actually could explain the majority of traffic for websites. Instead of looking at the internal infrastructure of a website to understand its traffic rankings, one must include external measurements such as backlinks and social media presence.

In essence, Goolsbee, Lovenheim, and Slemrod point to a variety of interesting possibilities, but one of the most notable ones is that people are using the Internet to discuss

---

[48] Ouchi, William G. *Theory Z: How American Business Can Meet the Japanese Challenge*. Reading, MA: Addison-Wesley, 1981. Print.

[49] Ibid.

[50] Goolsbee, Austan, Michael F. Lovenheim, and Joel Slemrod. "Playing With Fire: Cigarettes, Taxes, and Competition from the Internet." *American Economic Journal: Economic Policy* 2, no. 1 (2010): 131-54.

political events. As mentioned previously, many researchers within political science have found that social media platforms can be valuable in mobilizing support at the polls during election years[51] and other researchers have found that presence online can be used to predict turnout at protests as well.[52] If the Internet is impacting political behavior, could not the reverse be true as well? It seems likely that political beliefs and actions could very reasonably be impacting online behavior and browsing habits.

Technical and Infrastructural Factors

The claim that technical factors could impact traffic is hardly unique to this paper, but it has not been explored to a great degree within academic research. Previous researchers have shown that backlinks are included in a wide variety of measures to estimate Internet traffic, including Google's famous PageRank algorithm.[53] Few researchers, however, have taken this conclusion further and actually studied the direct correlation between the number of backlinks to a website and the amount of traffic that website receives. This is despite the fact that various players in industry are continuing to explore the importance of backlinks in terms of a relationship with Internet traffic patterns and have made recent developments on this front.[54] Although researchers have shown that the number of backlinks can increase a website's

---

[51] Cardenal, Ana S. "Why mobilize support online? The paradox of party behaviour online." *Party Politics* 19, no. 1 (2013): 83-103.

[52] Bell, Sam, David R. Cingranelli, Amanda Murdie, and Alper Caglayan. 2013. "Coercion, Capacity, and Coordination: Predictors of Political Violence." *Conflict Management and Peace Science*, forthcoming.

[53] Bryan, Kurt, and Tanya Leise. "The $25,000,000,000 Eigenvector: The Linear Algebra behind Google*." *SIAM Review* 48, no. 3 (September 2006): 569-81.

[54] "Brightedge Technologies, Inc.; "Detection and Analysis of Backlink Activity" in Patent Application Approval Process." *Marketing Weekly News* (Oct 20, 2012): 414. http://search.proquest.com.er.lib.k-state.edu/docview/1095678281?accountid=11789.

PageRank, this study seeks to show that this variable will be positively correlated with overall traffic as well.

As backlinks are such an important source of traffic for websites due to their presence in search engine algorithms, it would seem logical to include this as a variable in any kind of substantial traffic analysis. Yet, surprisingly, academic literature as a whole has paid very little attention to backlinks and seldom includes it as a variable in any form of analysis. In fact, the vast majority of academic studies that even reference the concept do so from a design perspective and, even then, such references are often only done in passing. For example, Goldschmidt and Weil mention the concept of backlinks within a larger paper on design reasoning, but even they use the concept of backlinks in a very general sense and do not even limit it to the technical definition, defining it instead as links that "go back in time."[55] One would think that this concept would have been studied more among academics as the specific Internet application of the concept developed more thoroughly, but this has yet to happen in a notable way. Given how much the Internet has developed since Goldschmidt and Weil's piece was first written in 1998, it seems appropriate for research to be done on the more modern application of backlinks with regards to Internet traffic.

It is also important to remember the research mentioned earlier emphasizing within the field of electrical engineering, researchers have found technology limits what people can do online and how many websites they can access within a given timeframe.[56] Subsequently, from a technical perspective, one can conclude there must be a way to project the maximum possible

---

[55] Goldschmidt, Gabriela, and Maya Weil. "Contents and Structure in Design Reasoning." *Design Issues* 14, no. 3 (1998): 85-100. Pg. 90.

[56] Cao, Jin, William S. Cleveland and David X. Sun. 2004. Bandwidth Estimation for Best-Effort Internet Traffic. *Statistical Science*, *19*(3), 518-543.

growth of Internet access rates and to use that growth within a model to predict traffic, although this has apparently not yet been done by other scholars.

The arguments pointing towards a relationship between internet access rates and traffic is a finding echoed by the economic literature on Internet infrastructure, which has found that as broadband speeds increase (and subsequently become more expensive to operate in many cases) and Internet service providers are consolidated, the field of providers in a given area can be decreased.[57] In this sense, the market factors can impact the technology available and, again, one can conclude that there must be a way to project the maximum possible growth of Internet access rates and to use these market factors within a model that predicts traffic, although this study can only do so in the most superficial sense.

Even with all of the information we do actually know about the technological factors in question, there are still public policy factors impacting the technological possibilities. In a study of municipal wireless Internet access in Philadelphia, scholars have found that political institutions can have a profound impact on the implementation of wireless Internet access.[58]

This intersection between the technical and the social factors related to Internet access highlights the need to study the technological factors alongside the demographics of the sites and their target audiences. In many senses, the two form a system that is perhaps best studied from a systems theoretical viewpoint. Such an approach would mark a remarkable shift from current approaches towards studying Internet traffic patterns.

---

[57] Rubinfeld, Daniel L., and Hal J. Singer. "Vertical Foreclosure in Broadband Access?" *The Journal of Industrial Economics* 49, no. 3 (2001): 299-318.

[58] Jain, Abhijit, Munir Mandviwalla, and Rajiv D. Banker. "Government as Catalyst: Can It Work Again with Wireless Internet Access?" *Public Administration Review* 67, no. 6 (2007): 993-1005.

Unique Contributions of Study

Ultimately, all of this translates to a situation in which there are notable gaps within the current literature that will hopefully be filled by this study and others. Attempts to predict future traffic on specific sites could be best described as minimal. Part of this is due to the inability of traditional models to actually predict traffic instead of measure existing reliability. While measuring the speed of a given website may help to determine whether or not this website is stable or even whether or not it will appear high in search rankings, it is, due to the technical nature of Internet access, a theoretically poor predictor of traffic.

The existing models on sites such as Alexa, while useful in terms of predicting the stability of a website, are often unable to predict whether or not a site will draw high traffic due to the exclusion of important variables that could be used to measure linguistic and regional differences, demographic information and the general upward trends in terms of expanding Internet access rates. In addition, the relatively new application of systems theory towards Internet traffic research could yield fruitful results in this area.

As will be elaborated on in within the methodology section later on in this study, this work proposes an examination of factors that are actually correlated with Internet traffic rates by doing a regression analysis including many of the variables mentioned thus far. This will include binary variables to indicate if a site is based out of Saudi Arabia (as a proxy to represent at least part of the Arabic region) as well as one to indicate whether or not it is based out of the United States, variables to control for the general upward trend in terms of expanding Internet access and more traditional measurements as well. Such an approach, rooted in the ideas behind systems theory and, more specifically, human-computer interaction, will hopefully yield more accurate results than traditional models.

Chapter 3 - Methodology

Because of what is present in the literature, this study is designed to answer the following research questions:

Research Question 1: What factors are correlated with higher traffic on high-traffic websites?

Research Question 2: Does a website's target audience limit how much traffic it can eventually receive?

Research Question 3: Is a website's social media presence correlated with higher traffic?

Research Question 4: Finally, are purely technical factors correlated with higher traffic as well?

In an attempt to understand what factors are correlated strongly or weakly with Internet traffic, this study utilizes a regression analysis on a relevant dataset. Using publically available data from various groups that measure Internet traffic, an analysis will be done using Internet traffic as the dependent variable and running it against several independent variables, including dummy variables used to indicate the domain type and the targeted region or origin for the websites in question. The dependent variable is the unique monthly visitors to a website between January and February of 2014 according to Quantcast's measures.

The unit of analysis, in this case, is the domain name of the website instead of the brand or company that operates the website. For example, this means that the unit of analysis would be "Google.com" instead of just Google and would allow for inclusion of multiple sites from the same owners that meet the criteria for the study. The sites were chosen using the regional rankings from quantcast.com for the United States, China and Saudi Arabia. The top 100 most

visited sites from each nation were included in the study when data was available. Approximately 300 websites ended up being included in the study, excluding the ones for which reliable data was not displayed[59].

The focus was on high-traffic websites. Using the Alexa and monthly Quantcast rankings mentioned earlier in this paper, it is a relatively straight-forward process to determine which sites are the most visited. However, Alexa's publicly available ratings are focused only on global rankings. A few of the variables included will need little introduction. Perhaps most obvious, dummy variables will be included to indicate if the website is based on a .com, .net, .org, .gov or .edu domain. The .com domain top level domain (TLD) is largely the standard online and is generally the one preferred by web developers whenever it is available. Thus, it should be expected that .com domains will generally be more positively correlated with traffic than other domain types.

Other factors, however, are less traditional for this type of analysis and are unique to this research. An exhaustive study did not isolate any previous research showing otherwise. As mentioned previously, there is strong evidence within the literature on this topic to suggest that the website's region of origin and target audience should have a strong correlation with traffic. Specifically, the behavior of residents of the United States, China and Saudi Arabia appear to be markedly different than other regions. Thus, a dummy variable for each of these regions is included in the analysis. This regional variable was determined by using Alexa's indicator for the majority of traffic; in other words, if the largest share of a site's traffic was coming from the United States, it was classified as a United States-based site for the purposes of this study.

[59] Some website owners might request that their traffic be excluded from public estimates and others might present unique challenges in terms of being measured and are subsequently not included in Quantcast's monthly measurements. Subsequently, that means that these sites were not included in the regression models.

Likewise, sites with the largest share of traffic coming from China were classified as Chinese and sites with largest share of traffic coming from Saudi Arabia were classified as Saudi.

To be clear, sites deriving the majority of their traffic from outside of these three countries/regions were included in the dataset in order to make sure that there is something to compare against, but there is very little in the literature to suggest any other region should be markedly different in terms of traffic. For example, European-centered sites will be included in this analysis, but a binary variable will not be included to indicate whether or not the site is based out of Europe or whether or not the majority of its traffic comes from European audiences.

In terms of more technical factors, variables will be included that measure a website's presence online. These technical factors will include the number of backlinks directed towards a website. Indicators of the site's presence on social media will also be included. More specifically, a variable will be included measuring a site's presence on each of the following social media networks: Facebook, Twitter and Google+. This social media presence was measured by looking at the number of links on Twitter, the number of "+1"s on Google Plus and the number of "likes" and shares on Facebook. While it would perhaps be preferable to include measures for other social media networks, these measures for other social media networks are not easily available and, unfortunately, even when it is available, it is often unreliable.

All of these technical and/or social media-related indicators can be found on a variety of websites, including Alexa.com, ahrefs.com and quantcast.com. These websites are used by web administrators and developers to measure traffic and the other selected indicator. Other than a few possible exceptions (for example, Alexa does not measure data on itself) the data was compiled using these three sites.

Selecting the sites to be analyzed, however, is much more problematic than the selection of the independent variables. The selection is problematic for regression analysis because, to put it simply, the traffic distribution online is not normal. The vast majority of sites on the Internet receive little to no traffic and measuring all of them would be nearly impossible from a technical perspective. In general, traffic measurements are only reliable for some of the most popular websites online, which is misleading due to the large number of sites that receive little traffic.

Subsequently, it seems reasonable to expect the distribution for the dependent variable for this dataset will be strongly skewed to the right despite the fact that the general trends for websites would actually be skewed towards the left.[60] However, it should generally be assumed that the sites would all fall within an even enough distribution to still yield meaningful results. In order to combat this inherent problem with measuring Internet traffic, this dataset will include a substantial number of observations. This large number of observations should help the distribution to be as close to normal as possible as well to help verify any findings of statistical significance.

In conclusion, the regression analysis will be using Internet traffic as the dependent variable and it will be analyzed against this study's selection of thirteen independent variables. As mentioned previously, this large number of variables is in part an attempt to recognize the general trend towards more complex models within this literature. Although thirteen independent variables may seem like a significant amount, eight of them are binary variables. Also, the vast

[60] A large number of websites, potentially even a majority, are assumed to get very little traffic, often with less than ten visitors a month. This presents a unique problem in that most of the methodologies used to measure traffic are dependent on the site actually hitting a minimum threshold of measurable traffic. Thus, while this is just speculative on the part of the author of this study, it seems reasonable to assume that the general trend for websites would be towards a low level of traffic on average if all sites could be accurately measured.

majority of this literature has found support for more complex models, so this model is reasonable and will result in a useful contribution to this literature.

Chapter 4 - Findings

Ultimately, this study included two primary forms of analysis on the subject of traffic patterns on high-traffic websites. There were three chi-square tables created and analyzed – running each of the Quantcast national rankings (one for the American ranking, one for the Chinese ranking and one for the Arabic ranking) against the Alexa global rankings. This was done for two reasons. The first was to verify the measures are, at least to some extent, reliable against one another. The second reason was to see which of the three selected nations was most strongly correlated with the global rankings.

Following these three chi-square analyses, there were seven regression models run, using different dependent variables but always including the same thirteen independent variables.[61] These seven regression models can be further divided into three subgroups. The first subgroup is the individual countries, meaning that the independent variables were regressed against the traffic for the United States, China and Saudi Arabia. The second subgroup is the combination of two countries, meaning that the independent variables were regressed against the combined traffic for the United States and China, the United States and Saudi Arabia, and then Saudi Arabia and China. Finally, the last regression model was the traffic for all countries combined,

---

[61] In most cases, the dummy variable for an .edu domain type and a .gov domain type were dropped from the models as none of the sites fit the criteria necessary to justify their inclusion in the models. Although it is still possible that these domain types might have unique impacts on traffic patterns, for high-traffic websites, the number of sites that even had these domain-types was too limited to be analyzed.

Only one .gov site, irs.gov, was in the top 100 websites for a country (in this case, it was in the top 100 for the United States) and none of the .edu sites were in the top 100 most viewed sites for any of the three countries selected for the study. It should be noted, however, that this study did not divide up backlinks into these subcategories based on domain type and that this might be an area for further research.

meaning that the independent variables were regressed against the combined traffic for the United States, China and Saudi Arabia.

Statistical Findings

All of the analysis was calculated using SPSS.[62] The first chi-square was the Quantcast rankings for the United States run against Alexa's global rankings. The SPSS results were as follows:

Table 4.1 – Chi-square summary table for United States Quantcast rankings/Alexa global rankings

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	29406.000[a]	29237	.242
Likelihood Ratio	1784.261	29237	1.000
Linear-by-Linear Association	61.003	1	.000
N of Valid Cases	174		

a. 29580 cells (100.0%) have expected count less than 5. The minimum expected count is .01.

This shows, somewhat unsurprisingly, that the United States' Quantcast rankings are highly correlated with Alexa's global rankings. Likewise, a chi-square analysis was done of the Quantcast rankings for China against Alexa's global rankings. The results were as follows:

[62] The full chi-square results are not published here due to their sheer size. As the range of rankings ran in the hundreds of thousands, each full chi-square table would have well over a million individual cells if it was to include all of the cells with a 0% frequency.

Table 4.2 - Chi-Square results for Chinese Quantcast rankings/Alexa global rankings

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	15488.000[a]	15375	.259
Likelihood Ratio	1225.484	15375	1.000
Linear-by-Linear Association	1.453	1	.228
N of Valid Cases	128		

a. 15624 cells (100.0%) have expected count less than 5. The minimum expected count is .01.

Again, the Chinese Quantcast rankings are correlated with Alexa's global ratings, but not nearly as strongly as the placement on the United States' list. There were a much lower number of observations for the Chinese rankings than in the ratings for the United States, largely just due to the inherent problems with finding accurate measures for websites based somewhere outside of the United States, which resulted in slightly less conclusive findings. Finally, a chi-square analysis was performed using the Quantcast placement for sites in Saudi Arabia against Alexa's global rankings. The results were as follows:

Table 4.3 - Chi-square results for Saudi Arabia Quantcast rankings/Alexa global rankings

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	14182.500[a]	14042	.201
Likelihood Ratio	1158.318	14042	1.000
Linear-by-Linear Association	1.112	1	.292
N of Valid Cases	122		

a. 14280 cells (100.0%) have expected count less than 5. The minimum expected count is .01.

The Saudi Quantcast rankings are correlated Alexa's global rankings, but not nearly as strongly as the United States' or the Chinese rankings. Even when compared with the Chinese rankings, there were a lower number of observations for the Saudi rankings, which, again, resulted in slightly less conclusive findings. In each of these cases, the linear-by-linear association, chi-square value and likelihood ratios all went down alongside the number of observations.

As mentioned previously, this chi-square analysis was done for two reasons. The first was to verify that the measures are, at least to some extent, reliable against one another. The second was to see which of the three selected nations was most strongly correlated with the global rankings.

With regards to the verifying measures, the findings show no reason to reject either the Quantcast or Alexa rankings. There is certainly some variation between the measures, which is to be expected, but nothing problematic for the purposes of this study. Whenever there are enough observations available to determine statistical significance, it is clear that the relationship between the two is not merely random and that at least some degree of similarity between the rankings exists.

With regards to correlation with global rankings for including the chi-square analysis, it is clear that Quantcast's placement of sites on the United States' list is the most strongly correlated with the global rankings. This suggests that the habits of Internet users in the United States are having a disproportionate impact on global ratings.[63]

---

[63] Some of this may be due to an over-representation of American audiences in the data used to generate the Alexa rankings, but without access to all of Alexa's data, which is not entirely publically available, this would be impossible to determine.

For the first subgroup of regression models, the traffic for the individual countries was analyzed. The SPSS results using traffic in the United States as the dependent variable were as follows:

Table 4.4 - Regression model for U.S. traffic

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.805[a]	.649	.622	18643545.111

a. Predictors: (Constant), facebook_links, gov_domain, net_domain, org_domain, twitter, arabic, chinese, google_plus, american, com_domain, backlinks, facebook_likes

Ultimately, the r-squared values for this model were promising and show a fairly high degree of predictive power. An adjusted r-square value of 0.622 indicates that roughly 62% of the traffic can be predicted using the independent variables within the model. For a further breakdown, it is worth analyzing the correlation of each of the variables:

Table 4.5 - Coefficient table for U.S. only traffic regression

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-10209580.362	6741267.336		-1.514	.132
	arabic	1862984.061	10318034.118	.009	.181	.857
	american	17156289.480	4473840.155	.215	3.835	.000
	chinese	1782337.357	10318078.020	.009	.173	.863
	com_domain	12789703.386	5968702.217	.146	2.143	.034
	org_domain	18052916.946	9634750.104	.109	1.874	.063
	net_domain	8311469.571	9560124.915	.050	.869	.386
	gov_domain	7948080.346	19542758.488	.020	.407	.685
	backlinks	.003	.002	.197	1.332	.185
	google_plus	-.919	1.385	-.033	-.664	.508
	twitter	-.457	1.681	-.024	-.272	.786
	facebook_likes	3.837	20.923	.032	.183	.855
	facebook_links	32.519	5.678	.606	5.727	.000

a. Dependent Variable: us_traffic

Unfortunately, most of the variables were not statistically significant. Only three of the variables showed statistical significance at a 0.05 level or less. The number of Facebook links and whether or not a site's target audience was based out of the United States were both significant at a < 0.001 level. Whether or not a site was a .com domain was significant at a < 0.05 level and whether or not a site was a .org domain was significant at a similar level, with a significance at 0.063.

Interestingly, all of the variables except for the number of "+1"s on Google+ and the number of links on Twitter were positively correlated with traffic. Surprisingly, even the dummy variables for a Chinese and Saudi audience were positively correlated with the United States'

traffic ratings. However, because these coefficients showed such weak statistical significance, it would be unwise to draw any major conclusions for this model with regards to those variables.

The results for the regression against Chinese traffic were slightly different. The SPSS results using Chinese traffic as the dependent variable were as follows:

Table 4.6 - Regression summary table for China-only traffic

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.594[a]	.353	.289	385786.030

a. Predictors: (Constant), facebook_links, backlinks, google_plus, arabic, twitter, org_domain, chinese, net_domain, american, facebook_likes, com_domain

Ultimately, the r-squared values for this model were less promising and show a much lower level of predictive power. An adjusted r-square value of 0.289 indicates that roughly 28.9% of the traffic can be predicted using the independent variables within the model. For a further breakdown, it is worth analyzing the correlation of each of the variables:

Table 4.7 - Coefficient table for China-only traffic regression

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	165896.361	150405.621		1.103	.272
	arabic	-249984.088	247660.698	-.085	-1.009	.315
	american	-276895.742	99996.023	-.255	-2.769	.007
	chinese	-20123.301	217108.794	-.008	-.093	.926
	com_domain	164500.644	129260.782	.136	1.273	.206
	org_domain	120429.022	237413.708	.047	.507	.613
	net_domain	87096.918	198772.149	.041	.438	.662
	backlinks	.001	.000	.510	6.651	.000
	google_plus	.000	.030	-.001	-.016	.987
	twitter	-.014	.041	-.026	-.345	.731
	facebook_likes	-.599	1.383	-.040	-.433	.666
	facebook_links	1.941	.831	.226	2.335	.021

a. Dependent Variable: china_traffic

For this Chinese-only model as well, most of the variables were not statistically significant. In this case, only one of the variables showed statistical significance at a 0.001 level or less. The number of backlinks for a site was significant, with findings significant at even a < 0.001 level. Again, the dummy variable for whether or not a site's target audience was based out of the United States was statistically significant, although it was less significant in this model with a significance at 0.007 level. The number of Facebook shares was also still statistically significant, although again it was less significant in this model with a significance at 0.021 level.

The results for the regression against Saudi traffic were also slightly different. The SPSS results using Saudi traffic as the dependent variable were as follows:

Table 4.8 - Regression summary table for Saudi Arabia-only traffic

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.798[a]	.638	.602	213090.158

a. Predictors: (Constant), facebook_links, backlinks, google_plus, chinese, twitter, org_domain, arabic, net_domain, american, facebook_likes, com_domain

Ultimately, the r-squared values for this model were more promising than the Chinese model and show a higher level of predictive power, although it showed a slightly lower level of predictive power than the United States-only model. An adjusted r-square value of 0.602 indicates that roughly 60.2% of the traffic can be predicted using the independent variables within the model. For a further breakdown, it is worth analyzing the correlation of each of the variables:

Table 4.9 - Coefficient table for Saudi Arabia-only traffic regression

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	162188.022	84471.565		1.920	.057
	arabic	119155.671	119378.005	.063	.998	.320
	american	-146465.620	53986.560	-.184	-2.713	.008
	chinese	-89434.846	134554.769	-.041	-.665	.508
	com_domain	-10272.234	73661.698	-.011	-.139	.889
	org_domain	-95319.252	132184.056	-.050	-.721	.472
	net_domain	-199144.066	113291.431	-.127	-1.758	.082
	backlinks	.000	.000	.108	1.896	.061
	google_plus	-.015	.017	-.053	-.909	.365
	twitter	.004	.023	.011	.187	.852
	facebook_likes	-.847	.764	-.077	-1.109	.270
	facebook_links	5.066	.459	.797	11.034	.000

a. Dependent Variable: arabia_traffic

Like the other two models, most of the variables were not statistically significant. In this case, only one of the variables showed statistical significance at a 0.001 level or less. The number of Facebook links for a site was significant, with findings significant at even a < 0.001 level. Again, the dummy variable for whether or not a site's target audience was in the United States was statistically significant, although it was less significant in this model than even the Chinese model, with a significance at 0.008 level. Whether or not a site was a .net domain was significant at close to a 0.05 level, with a significance at 0.082 and the number of backlinks had a significance at 0.061.

Interestingly, unlike the United States-only model, nearly all of the variables were negatively correlated with traffic. Only the dummy variable for a Saudi target audience, the

number of Twitter links and the number of Facebook shares were positively were positively correlated with traffic.

The number of backlinks showed an extremely low effect per backlink, with a coefficient of roughly 0.000. Given the significance of the coefficient, however, it seems safe to assume that this effect is still positive, but likely just an increase of < 0.001 visitors per backlink. Considering that many of the sites selected for the study have millions (and even billions, in some rare cases) of backlinks, this is actually a much larger effect than it initially may seem.

Ultimately, all of the models showed at least some degree of significance for both the number of backlinks and the number of links or shares on Facebook. All models also showed a degree of significance for whether or not the target audience was based out of the United States. Both the United States' and Saudi Arabian models showed some strong significance for at least one type of domain, although the Chinese model showed nothing of the sort.

For the second subgroup of regression models was more complicated in that it includes the traffic for two countries as part of the dependent variable, the traffic for the individual countries was no longer analyzed, but included the data for two of the countries from the study. The SPSS results using the combined Saudi and Chinese traffic as the dependent variable were as follows:

Table 4.10 - Regression summary table for Saudi/Chinese combined traffic

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.705[a]	.497	.447	498184.81232

a. Predictors: (Constant), facebook_links, backlinks, google_plus, twitter, arabic, chinese, org_domain, net_domain, american, facebook_likes, com_domain

The adjusted r-squared value for this regression model was 0.447, which means it had less predictive power than the Saudi-only traffic model but more predictive power than the Chinese model, which seems reasonable as it is essentially a combination of those two models. This would seem to suggest that the traffic patterns in China are slightly less predictable than the patterns from Saudi Arabia, but in order to verify this, a further breakdown of the variables is necessary. The SPSS output was as follows:

Table 4.11 - Coefficient summary table for China/Saudi Arabia-combined traffic model

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	330961.924	199553.396		1.659	.100
	arabic	-53980.302	319913.426	-.013	-.169	.866
	american	-430670.217	129136.795	-.268	-3.335	.001
	chinese	-92691.641	315672.342	-.022	-.294	.770
	com_domain	159867.239	173442.089	.089	.922	.359
	org_domain	24059.454	309931.710	.006	.078	.938
	net_domain	-128232.979	265171.946	-.042	-.484	.630
	backlinks	.002	.000	.405	5.957	.000
	google_plus	-.016	.039	-.028	-.408	.684
	twitter	-.010	.054	-.013	-.189	.850
	facebook_likes	-1.461	1.786	-.068	-.818	.415
	facebook_links	7.014	1.075	.560	6.525	.000

a. Dependent Variable: arabia_china_traffic

As with the rest of the models shown thus far, the number of backlinks, Facebook shares and whether or not a site's target audience was composed of Americans all had a high degree of

statistical significance. However, unlike the United States-only and Saudi-only models, none of the dummy variables for the domain type showed anything close to statistical significance.

It is worth noting, however, that the degree of significance for whether or not a site's target audience was Saudi or Chinese was actually lower than in the Saudi Arabia-only model.[64] This seems somewhat counter-intuitive, as one would think that at least the dummy variable for a Chinese target audience would show a higher degree of significance for a model that includes Chinese traffic measures in the dependent variable, but this was not the case.

The two models including the United States as one of the two countries in the dependent variable showed slightly different results. The SPSS results using the combined Saudi Arabian and United States-based traffic as the dependent variable were as follows:

Table 4.12 - Regression summary for Saudi Arabia/United States combined traffic model

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.477[a]	.228	.151	12945301.22664

a. Predictors: (Constant), facebook_links, backlinks, google_plus, twitter, chinese, org_domain, arabic, net_domain, american, facebook_likes, com_domain

In this case, the model shows almost no predictive power whatsoever. In fact, with an adjusted r-squared value of 0.151, this model had the lowest predictive power of any of the

[64] Also interestingly, the coefficients of each of the specific country variables was negative. This may be in part due to the large number of sites that were in the top 100 for either Saudi Arabia or China but were actually targeting another area. For example, many sites in the sample actually drew more traffic from India than any other country and it is possible that a dummy variable for India or some other country might have shown a positive effect in this model.

models analyzed thus far. There is little to suggest that this model has much practical use in terms of predicting traffic, but it is still worth analyzing the coefficients to see which ones actually showed statistical significance. The SPSS output was as follows:

Table 4.13 - Coefficient table for United States/Saudi Arabia-combined traffic regression

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-6369547.695	5169780.174		-1.232	.221
	arabic	446608.907	7281754.216	.006	.061	.951
	american	10039641.720	3355046.260	.301	2.992	.003
	chinese	-1945368.429	8201247.494	-.021	-.237	.813
	com_domain	8668690.503	4476757.820	.229	1.936	.055
	org_domain	6503669.118	8042764.851	.082	.809	.420
	net_domain	7753462.496	6883852.361	.119	1.126	.262
	backlinks	.021	.007	.264	3.156	.002
	google_plus	.233	1.021	.019	.228	.820
	twitter	-.298	1.390	-.018	-.214	.831
	facebook_likes	18.373	46.397	.040	.396	.693
	facebook_links	29.580	27.911	.112	1.060	.292

a. Dependent Variable: arabia_us_traffic

In this model, only the dummy variable for the United States' audience and the number of backlinks showed statistical significance at a level < 0.01. Interestingly, this was one of the few models that did not show a high degree of significance for the number of links or shares on Facebook.

However, it did show a similar degree of significance (with regards to the United States-only traffic model) in terms of the indicator for a .com domain type. Although it was slightly less

significant at 0.055 instead of 0.034, this difference is marginal and would still seem to indicate that, at least for sites targeting the United States, whether or not a site has a .com domain does seem to impact the traffic in some form.

The second of the two models including the United States as one of the two countries in the dependent variable showed very similar results. The SPSS results using the combined traffic for China and the United States as the dependent variable were as follows:

Table 4.14 - Regression summary table for United States/China-combined traffic model

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.481[a]	.232	.156	12922706.36896

a. Predictors: (Constant), facebook_links, backlinks, google_plus, arabic, twitter, org_domain, chinese, net_domain, american, facebook_likes, com_domain

Again, in this case, the model shows almost no predictive power whatsoever. With an adjusted r-squared value of 0.156, this model had the lowest predictive power of any of the models analyzed thus far with the exception of the United States-Saudi Arabian model, which had an adjusted r-squared value of 0.151. The difference here is, of course, borderline non-existent between the two models. Again, there is little to suggest that this model has much practical use in terms of predicting traffic, but it is still worth analyzing the coefficients to see which ones actually showed statistical significance. The SPSS output was as follows:

Table 4.15 - Coefficient summary table for United States/China-combined regression

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-5819344.284	5038149.445		-1.155	.251
	arabic	870740.753	8295910.774	.010	.105	.917
	american	9882533.446	3349575.015	.296	2.950	.004
	chinese	196868.144	7272511.121	.002	.027	.978
	com_domain	8298844.988	4329859.051	.223	1.917	.058
	org_domain	6179979.171	7952666.491	.078	.777	.439
	net_domain	7040733.049	6658287.026	.108	1.057	.293
	backlinks	.023	.007	.278	3.327	.001
	google_plus	.256	1.019	.021	.251	.802
	twitter	-.316	1.388	-.019	-.227	.821
	facebook_likes	18.566	46.321	.041	.401	.689
	facebook_links	25.955	27.848	.098	.932	.353

a. Dependent Variable: china_us_traffic

The results with regards to the coefficients and their statistical significance were very similar. Again, in this model, only the dummy variable for an audience in the United States and the number of backlinks showed statistical significance at a level < 0.01. Interestingly, this was also one of the few models that did not show a very high degree of significance for the number of links or shares on Facebook.

The dummy variable for whether or not the site featured a .com domain was still one of the most interesting ones in terms of the results. This model also showed a similar degree of significance in terms of the indicator for a .com domain type with regards to the United States-only traffic model and the combined United States-Saudi Arabian model. Although it was slightly less significant at 0.058 instead of 0.055, this difference is essentially negligible between

the two models. Again, this does suggest that at least for the sites targeting the United States, whether or not a site uses a .com domain does seem to have at least some impact on traffic.

The final subgroup of the calculations is only one model, but it is by far the most complicated of any of the regressions performed in this study. This analysis uses the combined traffic of all three targeted countries (the United States, China and Saudi Arabia) in order to form an approximate estimate of the traffic patterns for these three nations. The SPSS results using the combined traffic of all countries in the study as the dependent variable were as follows:

Table 4.16 - Regression table summary for combined United States/China/Saudi Arabia traffic model

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.485[a]	.236	.160	12897235.25767

a. Predictors: (Constant), facebook_links, backlinks, google_plus, twitter, org_domain, arabic, chinese, net_domain, american, facebook_likes, com_domain

Again, the predictive power of the regression was relatively low. With an adjusted r-squared value of 0.160, it was slightly higher in terms of predictive power than the joint United States-Chinese model and the joint United States-Saudi Arabian analysis, but it was lower than all four of the other previous models presented. This would seem to suggest that something about American traffic patterns is very different than that of the other two countries. This can be verified by analyzing the coefficients to see which ones actually showed statistical significance. The SPSS output was as follows:

Table 4.17 - Coefficient summary table for United States/China/Saudi Arabia combined traffic regression

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-5572813.306	5013697.872		-1.112	.269
	arabic	100255.427	7253348.673	.001	.014	.989
	american	9744102.708	3342399.471	.294	2.915	.004
	chinese	36992.844	7255014.384	.000	.005	.996
	com_domain	8184612.656	4292397.605	.219	1.907	.059
	org_domain	5998434.726	7927067.306	.076	.757	.451
	net_domain	6921192.782	6637060.006	.106	1.043	.299
	backlinks	.023	.007	.279	3.366	.001
	google_plus	.244	1.017	.020	.240	.811
	twitter	-.310	1.385	-.019	-.224	.823
	facebook_likes	17.880	46.224	.039	.387	.700
	facebook_links	30.793	27.774	.116	1.109	.270

a. Dependent Variable: total_traffic

As expected, the statistical significance for whether or not a site's target audience is in the United States is remarkably different than its counterpart for either Saudi Arabia or China. In fact, out of all of the models included in the study, this one showed the least amount of significance for the Saudi and Chinese dummy variables. This is somewhat surprising given that this dependent variable includes the traffic measures for all three countries, but nonetheless confirms the suspicion that something remarkably different is going on with regards traffic coming out of the United States.

As with all of the rest of the models, the indicator for whether or not a site's target audience was in the United States and the number of backlinks showed the highest degree of

statistical significance. In this case, the dummy variable for a .com domain showed some significance with a value of 0.059, but it was still lower than any of the other models that have included the measures for the United States in the dependent variable. Also, the number of Facebook links/shares showed very little statistical significance in this model, indicating that perhaps this significance disappears at the international level.

Chapter 5 - Conclusion

Ultimately, all of this combines to suggest a few key points. Firstly and perhaps most importantly, there can be little doubt that backlinks are having a profound impact on the traffic that each individual website is receiving. In nearly every model, backlinks were statistically significant at a very high level. In many of the models, the number of Facebook links/shares, which is essentially just a backlink count limited to Facebook, was also statistically significant.

More importantly from a theoretical perspective, however, is that in nearly every case, the coefficient was positive, meaning that these links are indeed increasing the amount of traffic to the websites in question. In some cases, the effect per backlink is very small. In one model, the effect is an increase in < 0.001 visitor per backlink, but many of these sites have millions (or even billions) of backlinks directed towards the domain, meaning that the effect is still incredibly important overall. There can be little doubt that given the amount of backlinks in existence, there is a positive correlation between the number of backlinks and the number of visitors to a site.

The second important point is that audiences in the United States clearly are the primary drivers of traffic at both a global and local level. In all cases that included a measure for audiences in the United States at the dependent level, there was a positive impact of nearly ten million visitors to the site if the site was targeted towards audiences in the United States. For the models that did not include a measure for audiences in the United States at the dependent level, there was a negative impact to a site if it was targeted towards audiences in the United States. In nearly all of these cases, the effects were statistically significant.

It is clear that at a global level and for United States-only traffic, targeting the United States is positively correlated with the number of visitors. However, it seems that this has a

reverse effect for other countries. In these cases, it seems that targeting the United States actually has a negative impact on traffic.

Third, the domain-type does appear to have an impact on traffic as well. Again, with the exceptions of the two models focused only on China and Saudi Arabia as individual countries, having one of the major domain-types (.com, .net or .org) was positively correlated with the number of visitors going to the site. In many of these cases, the .com variable was statistically significant at a high level. Although there are some notable exceptions, it is clear from the results of this study that the dominant domain names online are the ones that use one of the three major domain-types.

Finally, these findings suggest that the potential impact of social media on traffic may, perhaps, be overblown. Aside from the number of links or shares on Facebook, the social media measures were never statistically significant within the context of this study. It is possible that there is an effect here still, but the effect seems to be less important than the number of backlinks directed at the site.

The two most important findings from this study are clear: backlinks are positively (and significantly) correlated with traffic and there is clearly some kind of regional effect on where people spend their time online. Interestingly enough, both of these points seem to go against the grain with regards to the current literature on the subject.

On the first point, it seems clear that the literature in mass communications and information systems needs to be paying more attention to the impact that backlinks are having on internet traffic patterns, particularly for high traffic websites. Within the field of mass communications, essentially no technical research has been done looking at the impact of

backlinks on traffic patterns. Yet, in the context of this study, backlinks are clearly correlated with traffic in a positive fashion.

As for the second point, while the literature in mass communications has suggested that people might be more likely to go to sites based out of their own regions, the conventional thinking has still been to look at the Internet as a part of global systems theory[65] and to suggest that the processes are globalized to the point of no longer being limited to specific country-based effects.

That simply does not seem to be the case here – although users are going to a select few sites regardless of where they are from, most people are going to the websites that target their regions. With the exception of the model dedicated exclusively to Chinese traffic patterns, each model showed a positive impact for its own targeted country. People in Saudi Arabia are the most likely to go to Saudi sites and residents of the United States are the most likely to go to sites targeting the United States. This is reminiscent of the early the findings of Bargh and McKenna, who find that users tend to gravitate towards websites dedicated towards regional concerns.[66]

Answers for Research Questions

Research Question 1

Research Question 1 asked what factors are correlated with higher traffic on high-traffic websites. This was arguably the most over-arching question contained in this study and so the answer is complex, but the three factors that consistently were correlated with higher traffic were

---

[65] Straussfogel, Debra. "World-Systems Theory: Toward a Heuristic and Pedagogic Conceptual Tool." *Economic Geography* 73, no. 1 (1997): 118-30.

[66] Bargh, John A., and Katelyn Y. A. McKenna. "The Internet and Social Life." *Annual Review of Psychology* 55, no. 1 (2004): 573-90.

the number of backlinks, the number of Facebook links/shares and whether or not the site is targeting an audience in the United States.

In some regional models, the type of domain could sometimes have a positive impact on traffic, with the binary variables for the domain type occasionally having positive coefficients with statistical significance. Very little support was found for the claim that a large presence on social media is positively correlated with traffic. None of the variables representing social media presence (the number of shares on Twitter, the number of "+1"s on Google+ and the number of likes on Facebook) aside from the number of Facebook links/shares was statistically significant in any of the models.

Research Question 2

Research Question 2 asked if a website's target audience limit how much traffic it can eventually receive. The answer to this question seems to depend on the specific target audience, but in general, this study found support for the idea that a website's target audience limits how much traffic it can eventually receive. As evidenced by the chi-square analyses done in this study, the traffic patterns within the United States are the most influential on global traffic patterns.

Given the sites listed in the rankings used for this study and the findings of the analyses, it seems clear that the most viewed sites in the world generally target audiences in the United States and that not doing so has a negative impact on the potential traffic levels. There is also apparently a regional effect on traffic, which is to be expected due to the linguistic factors mentioned throughout this study. Sites that target audiences in the United States tend to do best in the United States, sites that target Chinese audiences tend to do best in China, sites that target

Saudi Arabian audiences tend to do best in Saudi Arabia, etc. Further research should be done to determine whether or not this effect holds true for other countries not included in this study.

Research Question 3

Research Question 3 asked if a website's social media presence correlated with higher traffic. Little support was found for the existence of this correlation. As mentioned previously, very little support was found for the claim that a large presence on social media is positively correlated with traffic. None of the variables representing social media presence (the number of shares on Twitter, the number of "+1"s on Google+ and the number of likes on Facebook) aside from the number of Facebook links/shares were positively correlated with traffic in a statistically significant way.

The coefficients for the social media presence measures were mixed with regards to the nature of the relationship. In some models, the coefficients were positive and in others they were negative. In none of them were they actually statistically significant, aside from the notable exception of the number of Facebook links/shares. This makes sense in that the number of Facebook links/shares is essentially just a secondary form of backlink, which was nearly always statistically significant and positively correlated with traffic.

Research Question 4

Finally, Research Question 4 asked if purely technical factors correlated with higher traffic as well. The answer to this question is quite clearly yes. In nearly all of the models, the number of backlinks to a given website was positively correlated with traffic in a statistically significant way.

Of all the factors included in this study, the number of backlinks was arguably the most technical and also arguably the most strongly and consistently correlated with traffic. The causal

nature of this relationship, however, remains in question. There are some theoretical arguments suggesting that an increase in traffic increases the number of visits to a webpage, but there are also theoretical arguments suggesting that the larger a site's audience, the more likely it is to be linked to on other sites. This study makes no claim one way or the other and would suggest that this would be a fruitful area for further research.

Points for Future Research

In addition to the points listed above, one of the key things suggested by this study is that scholars should increase their emphasis on some of these alternative variables. Much of the technical literature has suggested that .edu and .gov domains are treated differently, but none of these sites is actually gaining a high-level of traffic. In most of these models, there was not a single domain based on these domain types in the study. While many scholars have suggested that these domain-types might have a positive impact on traffic – for different theoretical reasons – this study finds little evidence to support such a claim.[67]

It is also clear that some of the theoretical impacts of social media on Internet browsing patterns may be overstated. While there was no support for the idea of backlinks having an impact on traffic within the mass communications literature, it seemed clear that the literature was relatively unified on the argument that a large social media presence is positively correlated with traffic. This study would suggest otherwise.

Obviously a limitation of this study – particularly from a mass communications perspective – is that it did not look at brand awareness or specific points with regards to the companies that owned the sites. Instead, this study only looked at the relationship to the domain

[67] One could make the theoretical argument that while the domain type is important for technical reasons, it is not actually correlated with the resulting traffic. Further research is warranted to understand this issue more in-depth.

itself. It is still very possible that a large presence on social media might translate into "real world" sales boosts or a more positive perception of the company itself, but it does not appear that a large number of links on Twitter or a large number of "+1"s on Google+ is having any actual impact on browsing behavior. When all else was held constant, a site with a higher presence on Twitter or Google+ was no more likely to attract visitors than one that did not have such a presence.

Finally, it seems that pursuing the study of online behavior in India (in addition the United States, China and Saudi Arabia) would be a fruitful area for future study. Given the large population of India, it seems likely that including the nation in a study of high-traffic websites would be a promising are for future research. Many of the sites included in this study were targeting Indian audiences, despite the fact that they were highly viewed in other countries. This presence would seem to suggest that India has a role to play with regards to high-traffic Internet patterns and should be included in future study.

All of these above factors – the technical and cultural – combine to suggest that the literature in mass communications with regards to online media should begin exploring more alternative variables with regards to predicting and measuring traffic. Social media and lesser known domain-types, such as .edu and .gov, are apparently not having any discernable impact on traffic behavior. However, the number of backlinks and the website's targeted regions are having some kind of impact on the number of visitors to a website. This study is hopefully the first of many, but it would suggest that there is still a significant amount of work to be done with regards to this topic.

Bibliography

Introduction

Baumgartner, Jody C. "Polls and Elections: Editorial Cartoons 2.0: The Effects of Digital Political Satire on Presidential Candidate Evaluations." *Presidential Studies Quarterly* 38, no. 4 (December 01, 2008): 735-58. Accessed March 24, 2014. JSTOR.

Bell, Sam, David R. Cingranelli, Amanda Murdie, and Alper Caglayan. 2013. "Coercion, Capacity, and Coordination: Predictors of Political Violence." *Conflict Management and Peace Science*, forthcoming.

Cardenal, Ana S. "Why mobilize support online? The paradox of party behaviour online." *Party Politics* 19, no. 1 (2013): 83-103.

Carlson, Nicholas. "At Last -- The Full Story Of How Facebook Was Founded." *Business Insider*. Business Insider, 5 Mar. 2010. Web. 24 Oct. 2013.

Clemons, Eric K. "Newly Vulnerable Markets in an Age of Pure Information Products: An Analysis of Online Music and Online News." *Journal of Management Information Systems* 19, no. 3 (December 01, 2002): 17-41. Accessed March 24, 2014. JSTOR.

"Facebook.com." *Facebook.com Site Info*. Alexa, n.d. Web. 24 Oct. 2013.

Ferguson, James M. "Daily Newspaper Advertising Rates, Local Media Cross-Ownership, Newspaper Chains, and Media Competition." *The Journal of Law and Economics* 26, no. 3 (1983): 635-54.

French, Warren A., and J. Timothy Mcbrayer. "Arriving at Television Advertising Rates." *Journal of Advertising* 8, no. 1 (1979): 15-18.

Gentzkow, Matthew. "Valuing New Goods in a Model with Complementarity: Online Newspapers." *The American Economic Review* 97, no. 3 (June 01, 2007): 713-44. Accessed March 24, 2014. JSTOR.

Hafner-Burton, Emilie. *Making Human Rights a Reality*. Princeton, NJ: Princeton University Press, 2013.

Hesseldahl, Arik. "Bringing Broadband to the Urban Poor." *BusinessWeek*, December 31, 2008. http://www.businessweek.com/stories/2008-12-31/bringing-broadband-to-the-urban-poorbusinessweek-business-news-stock-market-and-financial-advice.

Holiday, Ryan. *Growth Hacker Marketing: A Primer on the Future of PR, Marketing, and Advertising*. Portfolio, 2013.

"Make the Web Faster." — Google Developers. Accessed October 24, 2013. https://developers.google.com/speed/pagespeed/.

Norris, Pippa, and Ronald Inglehart. *Cosmopolitan Communications*. New York: Cambridge University Press, 2009.

"Orkut.com." *Orkut.com Site Info*. Alexa, n.d. Web. 24 Oct. 2013.

Soley, Lawrence C., Jesse E. Teel, and Leonard N. Reid. "A Comparison of Influences on Fixed and Grid Radio Advertising Rates." *Journal of Advertising* 9, no. 4 (1980): 15-19.

Literature Review

Ansari, Asim, and Carl F. Mela. "E-Customization." *Journal of Marketing Research* 40, no. 2 (2003): 131-45.

Bargh, John A., and Katelyn Y. A. McKenna. "The Internet and Social Life." *Annual Review of Psychology* 55, no. 1 (2004): 573-90.

Bertot, John Carlo, Charles R. McClure, and Paul T. Jaeger. "The Impacts of Free Public Internet Access on Public Library Patrons and Communities." *The Library Quarterly* 78, no. 3 (2008): 285-301.

Bleha, Thomas. "Down to the Wire." *Foreign Affairs* 84, no. 3 (June 2005): 111-24.

"Brightedge Technologies, Inc.; "Detection and Analysis of Backlink Activity" in Patent Application Approval Process." *Marketing Weekly News* (Oct 20, 2012): 414. http://search.proquest.com.er.lib.k-state.edu/docview/1095678281?accountid=11789.

Bryan, Kurt, and Tanya Leise. "The $25,000,000,000 Eigenvector: The Linear Algebra behind Google*." *SIAM Review* 48, no. 3 (September 2006): 569-81.

Cao, Jin, William S. Cleveland and David X. Sun. 2004. Bandwidth Estimation for Best-Effort Internet Traffic. *Statistical Science*, *19*(3), 518-543.

Froomkin, A. Michael. "Habermas@Discourse. Net: Toward a Critical Theory of Cyberspace." The Harvard Law Review Association 116, no. 3 (January 2003): 749-873.

Goldschmidt, Gabriela, and Maya Weil. "Contents and Structure in Design Reasoning." *Design Issues* 14, no. 3 (1998): 85-100.

"HK Still Leads by Average Broadband Speeds." *Telecom Asia (Online)* (Oct 25, 2013). http://search.proquest.com.er.lib.k-state.edu/docview/1445008998?accountid=11789.

Howick, S., and J. Whalley. "Understanding the Drivers of Broadband Adoption: The Case of Rural and Remote Scotland." *Journal of the Operational Research Society* 59, no. 10 (2007): 1299-311.

Jain, Abhijit, Munir Mandviwalla, and Rajiv D. Banker. "Government as Catalyst: Can It Work Again with Wireless Internet Access?" *Public Administration Review* 67, no. 6 (2007): 993-1005.

Johnson, David Lloyd. "Re-Architecting Internet Access and Wireless Networks for Rural Developing Regions." Order No. 3559800, University of California, Santa Barbara, 2013. http://search.proquest.com.er.lib.k-state.edu/docview/1354493124?accountid=11789.

Kock, Ned. "The Psychobiological Model: Towards a New Theory of Computer-Mediated Communication Based on Darwinian Evolution." *Organization Science* 15, no. 3 (2004): 327-48.

Lyytinen, Katie, and Gregory M. Rose. "The Disruptive Nature of Information Technology Innovations: The Case of Internet Computing in Systems Development Organizations." *MIS Quarterly* 27, no. 4 (December 2003): 557-96.

MacKinnon, Rebecca. "China's Internet: Let A Thousand Filters Bloom." *China's Internet: Let A Thousand Filters Bloom*, June 28, 2005. http://yaleglobal.yale.edu/content/chinas-Internet-let-thousand-filters-bloom.

MacKinnon, Rebecca. 2007. "Flatter World and Thicker Walls? Blogs, Censorship And civic Discourse in China." *Public Choice* 134, no. 1-2: 31-46.

Madden, Joshua E. *Information Leaks and Information Systems*. Thesis, Baylor University, 2012.

Milian, Mark. "Top 20: Where to Find the World's Fastest Internet: The Download." Bloomberg.com. July 22, 2013. Accessed April 16, 2014. http://www.bloomberg.com/slideshow/2013-07-23/top-20-where-to-find-the-world-s-fastest-internet.html.

Norris, Pippa, and Ronald Inglehart. *Cosmopolitan Communications*. New York: Cambridge University Press, 2009.

Rubinfeld, Daniel L., and Hal J. Singer. "Vertical Foreclosure in Broadband Access?" *The Journal of Industrial Economics* 49, no. 3 (2001): 299-318.

Straussfogel, Debra. "World-Systems Theory: Toward a Heuristic and Pedagogic Conceptual Tool." *Economic Geography* 73, no. 1 (1997): 118-30.

Tan, Margaret, and Thompson S.H. Teo. "Factors Influencing the Adoption of the Internet." *International Journal of Electronic Commerce* 2, no. 3 (1998): 5-18.

Teo, Thompson S.H., Magaret Tan, and Wong Kok Buk. "A Contingency Model of Internet Adoption in Singapore." *International Journal of Electronic Commerce* 2, no. 2 (Winter 1997/1998): 95-118.

"WIPO PUBLISHES PATENT OF ALIBABA GROUP HOLDING FOR "RANKING SEARCH RESULTS BASED ON CLICK THROUGH RATES" (CHINESE INVENTORS)." *US Fed News Service, Including US State News,* Dec 27, 2013. http://search.proquest.com.er.lib.k-state.edu/docview/1470893859?accountid=11789.

Conclusion

Bargh, John A., and Katelyn Y. A. McKenna. "The Internet and Social Life." *Annual Review of Psychology* 55, no. 1 (2004): 573-90.

Straussfogel, Debra. "World-Systems Theory: Toward a Heuristic and Pedagogic Conceptual Tool." *Economic Geography* 73, no. 1 (1997): 118-30.