Online Network Centrality and Stock Returns

Logan P. Emery*

April 5, 2019

Abstract

This project investigates a network of firm competition and the asset pricing implications of firms' positions in this network. Networks provide a setting to analyze higher order interactive effects, and with the rise of increased computing power and larger data sets, more complex analysis has become more feasible. Using a unique data set of the degree to which website audiences overlap, I construct a network of firm websites with linkages between firms corresponding to competition for website audiences. I then seek to use this network to investigate the information content of these connections, how shocks propagate through the network determining risk exposure, and how asset prices reflect the network structure. My objective is to better understand the role networks play in determining risk exposure and asset prices.

The economy is a network. The interactions of investors, firms, and consumers constitute linkages that reveal the channels of economic activity. Studying these linkages is important to our understanding of how these agents respond to and eventually generate aggregate economic fluctuations. Financial markets are one setting where this economic network structure is of keen interest. A firm's equity is priced by its exposure to systematic risks, and its exposure is closely tied to its position within the economic network. Therefore, to fully understand the pricing of risk exposure, we need to understand the structure of firm networks. Thus far, however, research into the relationship between risk exposure and network structure has been primarily theoretical.

In this project, I empirically study the asset pricing implications of connections between firms based on their website audience overlap, which I argue reflects their competition for demand. I use these overlap scores to form a network of firm competition, and investigate three questions. First, what is the information content of these connections compared to other relatedness measures (e.g., industries)? Second, given the information content, how does the network structure affect the propagation of shocks from one firm to another? Third, how does a firm's position in the network affect its exposure to shocks and what are the asset pricing implications?

My analysis begins with a unique data set of web traffic, including estimates of audience overlap. The data come from a web traffic analytics company and describe the degree to which the audiences of two websites overlap.[1] I obtain approximately 150 million unique website-pair overlap scores, relating to nearly 10 million unique websites including more than 5,000 firm websites that cover over 90% of the market. The overlap scores are estimated from a large panel of millions of internet users, gathered from individuals and websites that agree to share their data with the web traffic analytics company. The company also uses search engine results to identify websites that share similar keywords, complementing the web traffic data. Many website pairs do not have a reported overlap score, but overlap through other websites. To capture this, I calculate an effective

*Doctoral Candidate, Krannert School of Management, Purdue University. Email: lemerype@purdue.edu

[1]This overlap score is calculated as the Jaccard similarity coefficient of estimated website audiences.

overlap score between each website pair that is not directly connected. I start by constructing a network of firm websites, which can be seen in Figure 1. I then use Dijkstra's (1959) algorithm to find the shortest path between two previously unconnected websites and define the effective overlap score as the product of the overlap scores along the path. All calculations and manipulations of the network are done in Python.

Networks, and more broadly measures of inter-firm connections, provide a powerful tool for examining deep economic relationship and have been used in a number of areas.[2] One particular strand of research has focused on how these connections can be used to identify peer firms or create alternative industry classifications. For example, Hoberg and Phillips (2016) use textual analysis of 10-K's to identify which other firms a firm thinks it is more related to.[3] The network of firm websites proposed here, however, is built from the web traffic data of everyday internet users, which are predominantly consumers of one type or another. In this vein, I first plan to investigate what the connections between firm websites, and the consumers who create those connections, reveal about firm relatedness.[4] This is an important step in understanding the relevance of the network connections.

Next, I intend to examine how the network structure affects the propagation of economic shocks. There is a literature attempting to understand how idiosyncratic shocks propagate through a network to create aggregate shocks. Firm-specific shocks have been argued to have a negligible effect on economy-wide fluctuations as the shocks are averaged out in aggregate (Lucas 1977). Gabaix (2011), however, argues that this "diversification" phenomenon is importantly a function of the size distribution of firms. If the distribution is fat-tailed, i.e., the largest firms constitute a sufficiently large portion of the economy, then idiosyncratic shocks to these large firms can generate significant aggregate fluctuations. Acemoglu, et al. (2012) model a network where some firms are significantly more central than others, and find that idiosyncratic shocks to these central firms can propagate through the network and become systematic. Using the network of firm websites, I plan to empirically investigate the propagation of shocks through the network.[5] My focus is on understanding how shocks propagate, how they contribute to aggregate volatility, and how a firm's position in the network affects its exposure to these shocks.

Finally, I want to explore the asset pricing implications of firms' risk exposure due to their position in the network. I measure a firm's position by calculating its network centrality. I then plan to investigate whether a firm's position in the network generates a significant risk premium.[6] Further analysis could include examining stock returns and centrality within sub-networks, such as clusters of related firms extracted from the larger network. This would provide direct evidence of the importance of network structure for financial markets.

[2]Networks capturing information flow have been examined with regard to investors (Ozsoylev et al. 2013), CEOs (El-Khatib, Fogel, and Jandik 2015), underwriters (Bajo et al. 2016), and investment managers (Rossi et al. 2018).

[3]Other examples include Lee, Ma, and Wang (2015), who use investor searches for firms on the Securities and Exchange Commission website, and Kaustia and Rantala (2018), who use analyst co-coverage.

[4]The traditional methodology for comparing relatedness measures, such as industry classifications, follows Bhojarj, Lee, and Oler (2003).

[5]While these theoretical papers model a network of input-output linkages, the network proposed here connects firms via their competition for website audiences. Another important part of this project will be to bridge the gap between the input-output network predictions and the competition network.

[6]In a related paper, Ahern (2013) examines firm-level stock returns and sector-level centrality as calculated from a network of input-output linkages.

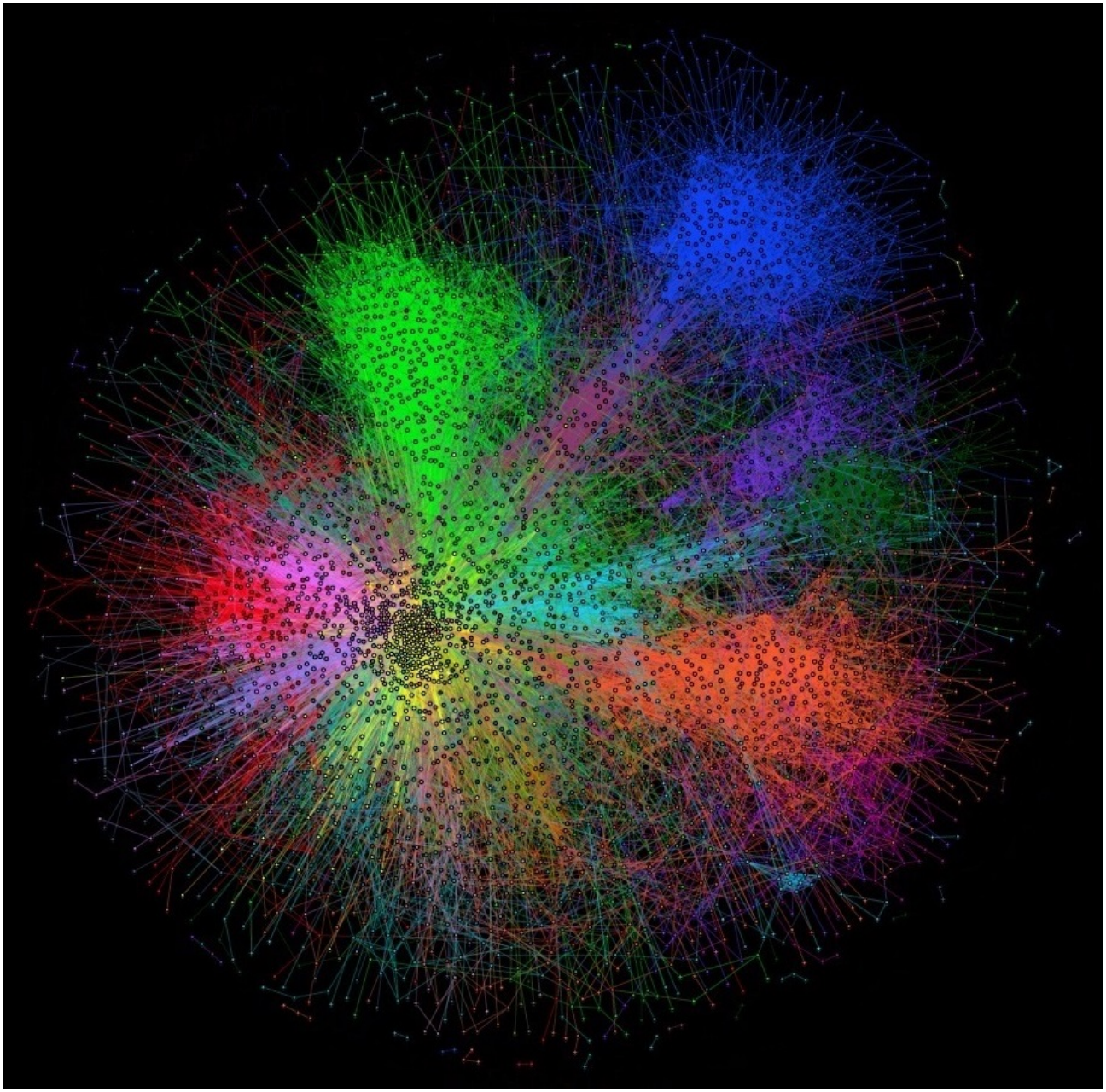


Figure 1
Visual representation of the network of firm websites This figure presents a visual representation of the network of firm websites. A Fruchterman-Reingold (1991) algorithm was run to move websites toward those they are more connected to and away from those to which they are less connected. Clusters extracted from the network using a modularity optimizing algorithm are identified by different colors.

References

Acemoglu, D., V. M. Carvalho, A. Ozdaglar, and A. Tahbaz-Salehi (2012). The network origins of aggregate fluctuations. *Econometrica 80*, 1977–2016.

Ahern, K. R. (2013). Network centrality and the cross section of stock returns. *Available on SSRN*.

Bajo, E., T. J. Chemmanur, K. Simonyan, and H. Tehranian (2016). Underwriter networks, investor attention, and initial public offerings. *Journal of Financial Economics 122*, 376–408.

Bhojraj, S., C. M. C. Lee, and D. K. Oler (2003). What's my line? a comparison of industry classification schemes for capital market research. *Journal of Accounting Research 41*, 745–774.

Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numerische Mathematik 1*, 269–271.

El-Khatib, R., K. Fogel, and T. Jandik (2015). Ceo network centrality and merger performance. *Journal of Financial Economics 116*, 349–382.

Fruchterman, T. M. J. and E. M. Reingold (1991). Graph drawing by force-directed placement. *Journal of Software: Practice and Experience 21*, 1129–1164.

Gabaix, X. (2011). The granular origins of aggregate fluctuations. *Econometrica 79*, 733–772.

Hoberg, G. and G. Phillips (2016). Text-based network industries and endogenous product differentiation. *Journal of Political Economy 124*, 1423–1465.

Kaustia, M. and V. Rantala (2013). Common analyst-based method for defining peer firms. *Available at SSRN*, 1–45.

Lee, C. M. C., P. Ma, and C. Wang (2015). Search-based peer firms: Aggregating investor perception through internet co-searches. *Journal of Financial Economics 2015*, 410–431.

Lucas, R. E. (1977). Understanding business cycles. *Carnegie-Rochester Conference Series on Public Policy 5*, 7–29.

Ozsoylev, H. N., J. Walden, M. D. Yavuz, and R. Bildik (2013). Investor networks in the stock market. *Review of Financial Studies 27*, 1323–1366.

Rossi, A. G., D. Blake, A. Timmermann, I. Tonks, and R. Wermers (2018). Network centrality and delegated investment performance. *Journal of Financial Economics 128*, 183–206.