Information Diffusion in Institutional Investor Networks

Edwin Hu*

November 15, 2015

Job Market Paper[1]

Abstract

This paper shows that network information flows generate a substantial information advantage for "central" institutional investors. I develop a new measure, Information Diffusion Centrality, which captures an investor's access to novel information as it arrives and diffuses through the network. I show that the abnormal interim trading performance of central investors is on average 32 basis points higher than that of peripheral investors in the following quarter. Central investors also have superior access to information about mergers. Central investors' round-trip trading performance in target stock is on average 150 basis points higher than that of peripheral investors. Centrality is distinct from common sources of information due to connections to merger advisors, is unrelated to industry expertise, and does not spuriously predict trading performance around sudden deaths of board members and key executives. My results suggest that institutional investors regularly share valuable information with one another.

*Jesse H. Jones School of Business at Rice University. I thank Jefferson Duarte, Alex Butler, Kerry Back and seminar participants at Rice University for valuable comments and feedback. Email: eddyhu@gmail.com.

[1]Latest version available at `http://edwinhu.github.io/idc.pdf`

In this paper I examine the hypothesis that investors share valuable information *through the grapevine*—e.g., via word-of-mouth communication—by studying network flows. Many authors in sociology and economics argue that the way in which people communicate affects the diffusion of information (e.g. Granovetter, 1973, 1985, 2005; Ellison and Fudenberg, 1995; Shiller, 1995). A recent theoretical literature demonstrates that the structure of social or information networks shapes the efficiency of financial markets.[2] Yet there is surprisingly little evidence that investors share valuable information with one another. One important exception, Shiller and Pound (1989), finds in a survey of 30 institutional investors that the majority purchased stocks based on conversations with other investment professionals. The result is surprising because, as Stein (2008) points out, professional investors have strong incentives to compete for—rather than share—information.[3] Given the implications for stock market efficiency, as well as for our understanding of investor behavior, it is fair to say that the empirical case for the importance of grapevine communication is underdeveloped.

In order to provide large-sample evidence that investors share valuable information, I examine more than a decade of order-level institutional trades. My empirical approach rests on the assumption that information sharing between investors is reflected in correlated trading activity. As a motivating example, we can think of investors sharing information through an online messaging platform. If two investors frequently chat with one another and share information about specific stocks, then we should expect their trades to have a positive contemporaneous correlation. A *Weighted Trading Correlation Network*, represented by nodes (investors) and edges (trading correlations), is an empirical estimate of the underlying network of conversations.[4]

[2] Colla and Mele (2010), Ozsoylev and Walden (2011), Han and Yang (2013), and Walden (2014)

[3] As for why informed investors would share information, Stein (2008) argues that in a repeated game it is optimal to share truthfully so that new information may be produced and shared in return. Also, it may be valuable for a fund manager to be able take a position and credibly communicate this to other managers so that their subsequent trades will move the stock price in the right direction.

[4] Instant Bloomberg, founded in 1982, is often described as an private network for investors to share and develop investment research ideas—predating current online social network platforms and even the widespread use of e-mail (see Edgecliffe-Johnson, Andrew; Alloway, Tracy; and Philip Stafford. "Instant

As information flows through the network, some investors will be better positioned than others to receive information—generating predictable information asymmetry between investors. I develop a new measure, *Information Diffusion Centrality*, to capture the expected information advantage provided by an investor's position in the network. Information Diffusion Centrality has two basic features. First, information diffuses along the paths of symmetric random walks on the network—information sharing is bilateral. Second, information heard through the grapevine is discounted, such that second-hand information is more valuable than third-hand information. The main prediction is that central investors have a higher probability of receiving novel information, whereas peripheral investors have less valuable commonly-held information.

An investor's information advantage is reflected in the optimal timing of purchases and sales of individual stocks. Implied trading performance based on quarterly changes in holdings is more common in the finance literature, but using changes in holdings assumes that all trades occur at the end of the quarter—ignoring the timing of the majority of intraquarter trades. Kacperczyk, Sialm, and Zheng (2008) and Puckett and Yan (2011) show that there is persistent outperformance that is not reflected in quarterly changes in holdings, which both papers attribute to superior information. For my main analysis, I compute quarterly trading performance using actual trades following Puckett and Yan (2011) in order to test the hypothesis that central investors have superior information relative to peripheral investors.

While my main analysis is most naturally motivated by access to information, it does not rule out the possibility that central investors' information advantage is due to superior information processing. To address this issue, I examine a setting in which access to information is directly related to trading performance—merger and acquisition announcements. For example, Ahern (2015) examines 239 instances of illegal insider trading around mergers

Bloomberg not going to fade away yet" *Financial Times* 21 May 2013). In addition, Saavedra, Hagerty, and Uzzi (2011) shows that simultaneous trading corresponds with instant messaging activity of 66 day traders.

and acquisitions and finds evidence that individuals in "tipping chains" located closer to the original source of information earn higher returns. Specifically, I examine round-trip trading performance—e.g. buys and subsequent sells—in target stock around merger announcements to verify that centrality measures access to information.[5]

I lack a good instrument for centrality that would allow me to cleanly identify the effects on trading performance. As a result, I rely on predictive regressions to measure the expected information advantage from being central. Centrality predicts subsequent trading performance if it measures an investor's information advantage, and equally importantly, if an investor's position in the network is stable over time. Stein (2008) shows that if information sharing is mutually beneficial, then there exists a sustainable "conversational equilibrium." I verify that my network estimates are highly persistent, which is consistent with such an equilibrium. I also take considerable care to account for the fact that centrality and the subset of investors who trade around events is not randomly assigned.

I conduct three additional empirical tests to address concerns that centrality proxies for omitted variables that predict subsequent trading performance.

First, Jegadeesh and Tang (2011) find that institutions connected to brokerage houses that serve as merger advisors are net buyers of target stocks and trade profitabily ahead of merger announcements. I name-match brokers to merger advisors and estimate investors' connections to advisors based on the share volume executed by brokerage houses in the previous quarter. I use merger advisor connections to control for common sources of information around merger announcements.

Second, I examine round-trip trading around new product announcements where product market expertise is likely to predict superior trading performance (e.g. Kacperczyk, Sialm, and Zheng, 2005). I construct a measure of investors' expertise in announcing firms' products

[5]Round-trip trading is consistent with investors trading on short-lived information—investors may want to reverse their positions in order to lock in gains or take advantage of overreactions (Hirshleifer, Subrahmanyam, and Titman, 1994; Brunnermeier, 2005).

based on the concentration of their trades in the previous year in the stocks of firms with similar product descriptions.[6] I use the concentration of trades to control for industry or product market expertise.

Third, as a falsification test, I examine sudden deaths of board directors and key executives to see if centrality spuriously predicts trading performance when there is no information. Sudden deaths are used as unpredictable shocks to corporate governance (Nguyen and Nielsen, 2010), hence there should be no returns to being central in this setting.[7]

My results are as follows. In predictive regressions of interim trading performance on Information Diffusion Centrality, I find that central investors—those with above median Information Diffusion Centrality—have 0.32% (0.48%) higher principal-weighted (equal-weighted) average abnormal trading performance in the following quarter. To put this estimate in perspective, the median investor's principal-weighted (equal-weighted) average trading performance in my sample from 1999Q1 to 2011Q3 is 0.35% (0.47%). Becoming central would nearly double the median investor's expected trading performance. The results hold when controlling for measures of trading activity and past performance, and are robust in a propensity score matched sample of central and peripheral investors.

Central investors also trade profitably ahead of merger announcements. Central investors have 1.50% (1.53%) higher principal-weighted (equal-weighted) average abnormal round-trip trading performance. To put this estimate in perspective, the median investor has a 1.27% (1.24%) principal-weighted (equal-weighted) trading performance around merger announcements. Becoming central would more than double the median investor's expected trading performance. Again, the results are robust in a matched sample, and in addition the results are nearly identical when I jointly estimate the propensity to trade around mergers

[6]Similarity scores come from the Hoberg and Phillips (2010, 2015) text-based industry network classifications.

[7]I identify sudden deaths by searching through news article headlines for keywords such as `death of`, `demise of`, and `passing of`, and using the article text I filter out deaths of individuals over the age of 65, deaths related to cancer, suicide, disease, illness, or complications related to prior conditions.

along with the returns to being central using a Heckman (1976) selection model.

Common sources of information due to merger and acquisition (M&A) advisor connections do not explain the returns to being central. Investors connected to M&A advisors have 0.91% (0.87%) higher principal-weighted (equal-weighted) average abnormal round-trip trading performance. Controlling for M&A advisor connections, central investors still have 1.40% (1.44%) higher principal-weighted (equal-weighted) trading performance.

Centrality is distinct from expertise, and both predict round-trip trading performance around new product announcements. In a multiple regression, experts have 0.51% (0.49%) higher principal-weighted (equal-weighted) average abnormal trading performance, and central investors have 0.66% (0.66%) higher trading performance. Omitting expertise, central investors have 0.65% (0.65%) higher principal-weighted (equal-weighted) round-trip trading performance.

Finally, when there is no information to share, there is no information advantage to being central. Central investors' trading performance around sudden deaths, relative to peripheral investors, is statistically indistinguishable from zero.

This paper makes two contributions. First, I find evidence that institutional investors regularly share valuable information. Information Diffusion Centrality predicts trading performance, and is also very persistent, which suggests that the consistent outperformance observed in the literature, e.g. Kacperczyk, Sialm, and Zheng (2008) and Puckett and Yan (2011), is due to a conversational equilibrium in the spirit of Stein (2008). The second contribution is methodological. Most network analysis in economics and finance assumes that economic interactions are binary, and ignores network flows. Novel information often flows through "weak ties" (Bakshy, Rosenn, Marlow, and Adamic, 2012)—hence Weighted Trading Correlation Networks are a more accurate representation of the underlying information structure than binary representations. Understanding centrality and the implicit network flows is important because existing measures of centrality in the finance literature—borrowed from the sociology literature—are measures of the diffusion of influence. Borgatti (2005) shows via simulation that measures of centrality are only applicable to the specific network flows they are designed for and give the "wrong" answer when used as proxies for other flows. Information Diffusion Centrality is the first measure of centrality in the finance literature—to my knowledge—that is specifically designed for information flows. I show that using measures of influence as proxies for information diffusion can lead to different empirical conclusions.

My results support the view that the way in which economic agents converse—i.e., the structure of social and information networks—have important implications for financial markets. Shiller (1995) argues that many behavioral biases may be explained by patterns of conversations, and Hirshleifer (2015) argues that the next step in behavioral finance is to study *social* finance—in other words, to incorporate theories of sociology into our understanding of financial markets. Herding, home bias, and underreaction, three of the most well-documented behavioral biases, are all arguably related to the localization of information embedded within the community structure of networks.

The remainder of this paper is organized as follows. Section 1 provides a brief review of the literature. Section 2 describes the network framework. In Section 3 I discuss the sample selection process. Section 4 demonstrates that central investors have superior trading performance on average, and Section 5 shows that central investors have superior trading performance around merger announcements. In Section 6, I rule out alternative hypotheses. Section 7 concludes.

1 Background and Related Literature

The empirical literature on "community effects" is often interpreted as a evidence of word-of-mouth communication. Community effects appear to be an important factor in individuals' stock market participation (Brown, Ivković, Smith, and Weisbenner, 2008), stock purchasing decisions (Ivković and Weisbenner, 2007), and local portfolio performance (Ivković and Weisbenner, 2005). Institutional investors who live in the same city and same neighborhood also have similar holdings (Hong, Kubik, and Stein, 2005; Pool, Stoffman, and Yonker, 2015). Pool, Stoffman, and Yonker (2015) also shows that "neighborhood" portfolios earn abnormal returns which suggests that there is an information advantage to being local, consistent with Coval and Moskowitz (1999, 2001).

With several significant differences, my paper is most closely related to Ozsoylev, Walden, Yavuz, and Bildik (2014) (OWYB). First, OWYB estimates networks based on overlapping trades, and use an ad-hoc measure called Rescaled Centrality to show that more central investors earn higher returns. I estimate networks based on correlated trading, and I develop Information Diffusion Centrality based on network information flows. Second, I extend their analysis to rule out alternative hypotheses that could explain the returns to being central including superior information processing, common sources of information, expertise, and spurious return predictability. Third, OWYB use data from the Istanbul Stock Exchange during 2005 which is comprised of 99.9% individual investors, whereas I focus on institutional investors in the US from 1999Q1–2011Q3.

It is not surprising that individuals rely on word of mouth if they are uninformed investors. But institutions are informed, have price impact, and compete for information.[8] Furthermore institutions make up the majority of US stock trading volume (Boehmer and Kelley, 2009), which suggests that word of mouth among institutions may have a larger economic impact than word of mouth among individual investors.

[8] E.g. Hendershott, Livdan, and Schürhoff (2015); Sias and Starks (1997); Akins, Ng, and Verdi (2011)

2 Framework

I begin with a brief introduction to networks in Section 2.1, and describe the empirical estimation of Weighted Trading Correlation Networks in Section 2.2, before defining Information Diffusion Centrality in Section 2.3. I describe the measures of abnormal trading performance in Section 2.4.

2.1 Networks Primer

2.1.1 Weighted Networks

Most network analysis focuses on collections of nodes joined by edges in a binary fashion. These networks are typically represented by (0,1) "adjacency" matrices:

$$A_{ij} = \begin{cases} 1 & \text{if } i \text{ is connected to } j \\ 0 & \text{otherwise} \end{cases} . \tag{1}$$

However, many settings, especially social and economic ones, include inherently "stronger" links and "weaker" links (Granovetter, 1973). Both strong and weak links are important, and focusing on only binary adjacency matrices ignores potentially important data. Therefore, we can also represent networks in terms of "weighted" adjacency matrices, $A_{ij} = w_{ij}$, where w_{ij} is the weight or strength of the connection between i and j.

For example, we can think of a simple triangle network with three nodes A, B, and C represented in geometric and adjacency matrix form below in Figure 1. The connection between A and B is strongest with a weight of 1, followed by the link between B and C with a weight of 0.5. The connection between A and C has the lowest weight of 0.25. In Figure 1 Panel A (below), edge thickness corresponds to the strength of the connections. The same data is encoded in the off-diagonal terms of the adjacency matrix in Panel B.

Insert Figure 1

2.1.2 Degree Centrality and Random Walks on Networks

The simplest measure of network centrality is Degree Centrality which is the sum of a node's direct connections. The Degree Centrality for node i is defined as the row sum of the adjacency matrix $\boldsymbol{A}$:

$$d_i = \sum_j A_{ij}. \tag{2}$$

The diagonal matrix, $D_{ii} = d_i$, is often used to define a random walk on a network $\boldsymbol{W} = \boldsymbol{D}^{-1}\boldsymbol{A}$. The typical off-diagonal element, $W_{ij} = \frac{1}{d_i}$, describes the probability of a random walker jumping from node i to node j. Note that the probability of a jump from i to any of its immediate neighbors is identical, and the row-sums of $\boldsymbol{W}$ are one. A symmetric random walk is defined as $\boldsymbol{\mathcal{W}} = \boldsymbol{D}^{-1/2}\boldsymbol{A}\boldsymbol{D}^{-1/2}$. The typical off-diagonal element of the symmetric random walk matrix is $\mathcal{W}_{ij} = \frac{w_{ij}}{\sqrt{d_i}\sqrt{d_j}}$. Unlike the "standard" random walk, the probability of a jump from i to j is the same as the probability of a jump from j to i.

As a simple example, consider an un-weighted network of five nodes A–E in Figure 2 below. A standard random walker (Panel A) from A jumps to B with probability $\frac{1}{4}$, and with probability one from B to A. The symmetric random walker jumps from A to B, and vice versa, with probability $\frac{1}{2}$. As we will see in Section 2.3, the symmetric random walk matrix will be useful in describing the dynamics of bilateral information sharing.

Insert Figure 2

2.2 Estimating Weighted Trading Correlation Networks

We often do not know the strength of the connections between investors and hence the adjacency matrix must be estimated from trading data. I estimate the strength of the connections between investors using pairwise trading correlations. More frequent information sharing is reflected in higher trading correlations and stronger links.

An investor's buying activity of n stocks over T periods can be represented by a (0,1) nT-vector stacked by stock, $\boldsymbol{b}_i = [\boldsymbol{b}_{i1}, \boldsymbol{b}_{i2}, \boldsymbol{b}_{i3}, \cdots, \boldsymbol{b}_{in}]'$. $\boldsymbol{b}_{in}$ is a T-vector which takes on values of one if if investor i bought stock n in period t and is zero otherwise. The buy correlation between investor i and j is the Pearson correlation, $\rho^b_{ij} = \frac{cov(\boldsymbol{b}_i, \boldsymbol{b}_j)}{\sigma_{\boldsymbol{b}_i} \sigma_{\boldsymbol{b}_j}}$, and the sell correlation, ρ^s_{ij}, is defined analogously for the vector of selling activity.

I estimate the weight of the connections between investors by taking the average of buy and sell trading correlations: $\hat{w}_{ij} = \frac{\rho^b_{ij} + \rho^s_{ij}}{2}$. I call the resulting adjacency matrix with elements $\hat{A}_{ij} \in [0, 1]$ the Weighted Trading Correlation Network (WTCN).[9]

Weighting based on trading correlations mitigates the influence of active liquidity traders on network estimates. For example, a related *un*-weighted procedure, Empirical Investor Networks (Ozsoylev, Walden, Yavuz, and Bildik, 2014), assigns links in a binary fashion based on the presence of overlapping trades. If two investors trade the same stock on the same side in the same period then they are connected. However, in the presence of active liquidity traders this methodology tends to identify many non-information linkages and can assign undue centrality to un-informed investors.

As a hypothetical example suppose the true information network is the star network in Figure 3 (below) of four investors who all share information with one central investor. The five informed investors are represented by dark nodes (●). Now suppose that these five investors only trade based on shared information. As a result, they will have perfectly correlated trades as indicated by the dark lines connecting them. There is also one very active liquidity trader (○) who does not have or share any information but trades frequently. Because the liquidity trader has overlapping trades with all five informed investors we observe

[9]Weighted networks based on pairwise correlations are common in the genomics and systems biology literature (see, Horvath, 2011, for example).

that the liquidity trader is part of the network. However, because the liquidity trader trades in periods when the informed traders are not trading, the overall trading correlation with the informed investors is low, and the liquidity trader's centrality is also low. If we ignore the trading correlations and connect investors based only on the existence of overlapping trades the liquidity trader appears to be most well-connected, with five connections when the bona fide central informed investor has only four connections. Trading correlation weighting correctly relegates active liquidity traders to the periphery of information networks.

Insert Figure 3

2.3 Information Diffusion Centrality

The ideal measure of an investors' information advantage should capture not only how much information an investor receives, but how delayed the information is as it diffuses through the network. Many measures of centrality exist in the finance literature but the question of what measure is "correct" is difficult to answer empirically because most measures of centrality are highly correlated with one another (see for example Valente, Coronges, Lakon, and Costenbader, 2008).[10] Therefore, the best way to distinguish between measures of centrality is to understand the implicit assumptions that each makes about network flows. For a more detailed discussion of centrality and network flows see Borgatti (2005).

Degree Centrality (Equation 2) measures flows that involve no indirect connections. In the context of information networks, Degree corresponds to information linkages generated by correlated signals. In practice, information linkages may be due to correlated signals or information sharing, both of which would generate correlated trading (Colla and Mele,

[10]For instance Bonacich, Katz, and Eigenvector Centrality are popular in the asset pricing literature (Cohen-Cole, Kirilenko, and Patacchini, 2014; Ozsoylev, Walden, Yavuz, and Bildik, 2014; Ahern, 2013). Betweenness, Closeness, and Degree Centrality are popular in the venture capital, corporate finance, and accounting literatures (Hochberg, Ljungqvist, and Lu, 2007, 2010; Hochberg, Lindsey, and Westerfield, 2015; El-Khatib, Fogel, and Jandik, 2015; Engelberg, Gao, and Parsons, 2013; Larcker, So, and Wang, 2013).

2010). If information linkages are primarily due to correlated signals, then investors with high Degree, i.e. many signals, should have an information advantage. However if information sharing is important, then Degree Centrality will be unable to capture information "heard through the grapevine" as it ignores access to information through indirect links.

Eigenvector Centrality measures flows of influence or importance (Bonacich, 1972a,b). Eigenvector Centrality is defined recursively in vector form:

$$\boldsymbol{\varepsilon} = \frac{1}{\lambda}\boldsymbol{A}\boldsymbol{\varepsilon}. \tag{3}$$

Bonacich and Katz Centrality are generalizations of Eigenvector Centrality, all of which define investors' centrality based on connections to other central investors. Eigenvector-like measures are useful because they take into account indirect links and are easy to compute. Ozsoylev, Walden, Yavuz, and Bildik (2014) use Eigenvector Centrality as a proxy for information diffusion, although it is more appropriate as a measure of social influence on opinion formation in the context of DeMarzo, Vayanos, and Zwiebel (2003). In short, if I persuade an influencer to share my opinion, then I am important.

I define Information Diffusion Centrality as:[11]

$$\boldsymbol{c} = \boldsymbol{\mathcal{S}}_t\boldsymbol{p}. \tag{4}$$

The heart of Information Diffusion Centrality is the symmetric information sharing matrix $\boldsymbol{\mathcal{S}}_t = e^{-(\boldsymbol{I}-\boldsymbol{\mathcal{W}})t}$, where $\boldsymbol{I}$ is the identity matrix and $\boldsymbol{\mathcal{W}}$ is the symmetric random walk matrix. The information sharing matrix describes the dynamics of continuous bilateral information sharing based on a stylized model of "social learning" in which agents take averages of the

[11]It is worth mentioning that Information Diffusion Centrality is distinct from the Diffusion Centrality defined from Banerjee, Chandrasekhar, Duflo, and Jackson (2013) which is essentially a finite sum version of Katz Centrality. In Appendix 8.2 I show that Katz Centrality can be derived as a special case of a proxy for Information Diffusion Centrality, hence Diffusion Centrality can be derived in a similar fashion.

differences between their signals and the signals of their neighbors (see Appendix 8.1).[12] Information Diffusion Centrality is parameterized by t and $\boldsymbol{p}$. $t > 0$ describes the time scale of information diffusion, such that when $t \to \infty$ information diffuses completely. $\boldsymbol{p}$ represents the probability that an investor receives new information. In empirical applications, I assume that $t = 1$ and p is uniform, which assumes information arrives with equal probability to each investor.[13]

Information Diffusion Centrality can be interpreted as a measure of the *expected discounted information advantage* as information diffuses through the network following the paths of symmetric random walks on the network. To see this interpretation more clearly, we can rewrite Information Diffusion Centrality as an exponential sum of symmetric random walks $\boldsymbol{\mathcal{W}}$:

$$\boldsymbol{c} = e^{-t} \sum_{k=0}^{\infty} \frac{t^k}{k!} \boldsymbol{\mathcal{W}}^k \boldsymbol{p}, \tag{5}$$

which follows from the definition of a matrix exponential. The ijth element of the k-order symmetric random walk matrix $\boldsymbol{\mathcal{W}}^k$ describes the probability of a random walker jumping from node i to node j in k hops. The ith element of $\boldsymbol{\mathcal{W}}^k \boldsymbol{p}$ is the sum of the probabilities of information arriving at node $\forall j \neq i$ and diffusing to node i in k hops. The contribution to centrality from all k-order random walks is "discounted" by $\frac{t^k}{k!}$. Naturally, information heard second-hand is more valuable than information heard third-hand, especially when opportunities for profitable trading are short-lived.

The symmetric random walk interpretation implies that connections to low degree nodes can actually provide more centrality than high degree nodes. If i's neighbor j has few

[12] In other words, agents only update based on new or different information. According to Kahneman (2011), "our mind has a useful capability to focus on whatever is odd, different or unusual."

[13] The assumption of $t = 1$ is innocuous. Assuming $\boldsymbol{p} \propto \boldsymbol{d}$, i.e. better connected investors receive information from outside the network more frequently, makes Information Diffusion Centrality more correlated with Degree and Eigenvector, which actually reduces its ability to predict trading performance. Hence, the uniform assumption seems to be a better description of the arrival of information to the network.

connections, i.e. low degree d_j, then the probability of a random walk jumping from j to i is relatively high. Eigenvector Centrality, by its recursive construction, predicts the opposite—connections to higher degree nodes defines importance. Combining the discounting and symmetric random walk features, we can interpret Information Diffusion Centrality as discounting commonly-held information in favor of novel information.

To visualize the difference between Eigenvector Centrality and Information Diffusion Centrality we can look at a simple example with 13 nodes in Figure 4. The most central node C has six connections, with access to novel information from five unique connections and one connection to the cluster on the bottom left. The cluster is fully connected such that any information that arrives to any of the seven nodes is shared with all seven members. As a result, any information in the fully connected cluster is commonly held. Node size is proportional to the number of connections and node color correlates with Information Diffusion Centrality on the left and Eigenvector Centrality on the right. Information Diffusion Centrality correctly identifies C as the most central node because it has access to novel information from five nodes, while nodes in the fully connected cluster have low centrality despite each node having many connections. According to Eigenvector Centrality, C is peripheral while all of the members of the fully connected cluster are central because each of them is connected to other well-connected nodes.

Insert Figure 4

2.4 Trading Performance

In the main analysis I compute quarterly interim trading performance by tracking the abnormal performance of all stocks bought and sold by a fund from the execution date, using the execution price, until the end of the quarter. In the event-study portion of the paper I compute round-trip trading performance around specific events. For example, a round-trip trade could consist of a fund buying target stock prior to the public announcement of a merger and selling the same stock following the announcement.

Trading performance computed using actual trades is distinct from implied performance computed using quarterly changes in portfolio holdings. Kacperczyk, Sialm, and Zheng (2008) find a large and persistent return gap between reported fund performance and the return on a portfolio based on changes in holdings. The authors argue that the return gap can be explained by fund managers' informational advantage in the optimal timing of purchases and sales of individual stocks. Puckett and Yan (2011) use actual trades to show that there exists persistent trading performance among the top performers, whereas using implied quarterly changes produces no persistent performance. The discrepancy arises because changes in holdings lack the exact timing and execution prices of actual trades. Following Puckett and Yan (2011) I compute a fund's quarterly interim trading performance as the difference between the weighted-average of buy and sell trading performances.

Puckett and Yan (2011) also show that changes in quarterly holdings do not capture round-trip trades. Round-trip trades are important if private information is short-lived and trading opportunities dissipate quickly. Investors trading on short-lived information may want to reverse their positions in order to lock in gains or exploit overreactions (Hirshleifer, Subrahmanyam, and Titman, 1994; Brunnermeier, 2005). To measure round-trip trading around an event I restrict my sample to trades made 60 days before and 30 days after the announcement date. I define a round-trip trade as a buy (sell) trade in the [-60,-1] window followed by a subsequent sell (buy) in the [0,30] window. A fund's round-trip trading performance for a particular event is computed as the principal- or equal-weighted average of all signed holding period returns using the actual execution prices and volume traded. I compute a fund's quarterly round-trip trading performance as the average round-trip trading performance across events that occur in a given quarter in which a fund trades.

For both interim and round-trip trading performance I adjust prices and share volume for stock splits. For interim trading performance I also cumulate dividends over the interim holding period and include them in the return calculations. I also compute "excess" or abnormal trading performance using Daniel, Grinblatt, Titman, and Wermers 1997 (DGTW) size, book-to-market, and momentum characteristic-matched value weighted portfolio returns computed over the corresponding (interim or round-trip) holding period and adjusted for delistings.

3 Data

To estimate an investor's centrality and trading performance I use high-frequency institutional trading data from a proprietary database provided by ANcerno Ltd. (a.k.a. Abel/Noser Solutions Ltd.). The ANcerno data allows me to observe the exact date, price, direction (buy or sell), and shares traded for all funds in the database from 1999Q1 to 2011Q3. My sample stops in 2011Q3 because ANcerno removed fund identifiers in 2012, preventing me from tracking subsequent trading activity. Furthermore, 2011Q4 and 2012Q1 have only one-third of the number of funds compared to 2011Q3 which results in unreliable network estimates.

I restrict my sample to all trades made in the US, in US currency, of common stock listed on the NYSE, AMEX, and NASDAQ exchanges, for which I am able to identify the ANcerno fund, and for which I am able to match the ANCerno provided point-in-time CUSIP to a CRSP `PERMNO`. The 10,355 funds in my sample trade a total of 8,555 unique common stocks. Overall, the funds make 141.85 million trades of 1.11 trillion shares valued at USD 34.26 trillion dollars. Puckett and Yan (2011) estimate that the ANcerno institutions account for approximately 8% of the total dollar value of CRSP trading volume between 1999 and 2005. Table 8 in Appendix 8.3 provides yearly summary statistics on the trading activity in my sample.

Because ANcerno timestamps are incomplete (see Anand, Irvine, Puckett, and Venkataraman, 2013) I estimate Weighted Trading Correlation Networks based on daily-level trading activity. To mitigate concerns that overlapping trades are coincidental, I restrict my sample of overlapping trades to "time-sensitive" overlapping trades. To identify time-sensitive trades, I exploit the fact that ANcerno tracks the number of days over which a stock was traded as part of an order ticket. I define time-sensitive trades as trades of stocks executed within a single day. If a fund manager receives short-lived information he or she would have a strong incentive to trade immediately, versus splitting up the order execution over multiple days, consequently sacrificing price impact for immediacy. On average, time-sensitive trades make up 62% of order executions. The resulting quarterly network snapshots contain on average 1,192 funds, with a minimum of 780 in 2011Q3 and a maximum of 1,639 in 2002Q1.

In the event-study portion of the paper I use merger and acquisition (M&A) announcements, new product announcements, and news announcements of the sudden deaths of board directors and key managers (CEO, CFO, VPs, etc.). The data on mergers comes from the Thomson Reuters SDC Platinum database (SDC). I use SDC's detailed information on M&As to identify announcement dates, the identities of the target firms, and the names of the merger advisors.

The data on new product announcements and sudden deaths comes from the S&P Capital IQ (CapIQ) Key Developments database (codes 41; 16, 101, and 102 respectively). The CapIQ data includes headlines and full-text articles of over 100 types of significant corporate events. To identify sudden deaths I search through headlines and full-text articles for keywords such as `death of`, `demise of`, and `passing of`. In addition, I filter out deaths of individuals over the age of 65, deaths related to cancer, suicide, disease, illness, or complications as reported in the articles.

3.1 Fund Summary Statistics

Table 1 Panel A summarizes the main variables of interest. The performance measures are the principal-weighted average abnormal interim trading performance (PW) and the equal-weighted performance (EW). The measures of centrality are Information Diffusion Centrality (Centrality), defined in Eq. 4, Eigenvector Centrality (Eigenvector) defined in Eq. 3 and Degree Centrality (Degree) defined in Eq. 2. I also report the average share volume (Volume), and number of trades (# Trades).

The average fund in my sample has a principal-weighted (equal-weighted) average abnormal interim trading performance of 0.52% (0.65%). For comparison, Puckett and Yan (2011) report that the average fund between 1999 and 2005 has a principal-weighted average abnormal interim trading performance of 0.57%. The average fund in my sample, over the same period, has a principal-weighted abnormal interim trading performance of 0.56%.

Because I am only concerned with a fund's relative centrality within a quarter, I normalize Information Diffusion and Eigenvector Centrality, without loss of generality, such that a fund's centrality can be interpreted as the percentage of the "total centrality" in a given quarter. Normalizing also facilitates comparing the two centrality measures. The median fund has an Information Diffusion Centrality of 2.9% and an Eigenvector Centrality of 0.07%.

The median fund makes 171 trades (Trades), exchanges 669,819 million shares (Volume), and has a Degree of four. The Degree is much smaller than the number of trades because trading correlations tend to be low. The average edge weight is about 2.6%. If every trade made by the median fund overlaps with one fund then the Degree is roughly $171 \times 0.026 \approx 4.5$.

Table 1 Panel B reports Spearman correlations for the same variables and includes auto-correlations on the diagonal. The two performance measures have a 77% correlation with one another, and are negatively correlated with the measures trading activity and Eigenvector and Degree Centrality. Information Diffusion Centrality has a positive correlation of 21 (36) basis points for principal-weighted (equal-weighted) average abnormal trading performance.

The centrality measures are positively correlated with one another and the trading measures, and are also highly autocorrelated. Degree and Information Diffusion Centrality have autocorrelations of 91%, which is consistent with the notion that the networks are stable over time and that there exists a sustainable "conversational equilibrium" (Stein, 2008).

Insert Table 1

Table 1 Panel A shows that many of the variables are heavy-tailed with high kurtosis, and the centrality and trading measures, apart from Information Diffusion Centrality, are positively skewed. Most real-world networks are characterized by power-law degree distributions (Barabási and Albert, 1999), and as a result most centrality measures have heavy right tails. Heavy tails tend to bias certain statistics which is why, one, I focus on the medians instead of means, two, I use Spearman correlations instead of Pearson correlations, and three, I take logs of the centrality and trading measures before using them in subsequent regressions. Information Diffusion Centrality and trading performance are slightly negatively skewed. The negative skew is due to the trading correlation edge weighting, which mitigates the influence of active liquidity traders which would otherwise have many connections and have high centrality.

To visualize the effect of edge weighting more clearly, Figure 5 plots the distribution of Information Diffusion Centrality over the sample period at the fund-quarter level using both unweighted (—) and trading correlation weighted edges (—). The solid lines are the median values of Information Diffusion Centrality unweighted and weighted, which are highly correlated over time. The shaded regions show the distribution between the 5 and 95 percentiles of Information Diffusion Centrality. The lighter shaded region corresponds to unweighted Information Diffusion Centrality, which is heavily right-skewed and has high variance much like Degree and Eigenvector Centrality.[14] In contrast, the darker shaded region, which corre-

[14] For example, the Pearson correlation between unweighted Information Diffusion, Eigenvector, and Degree Centrality is around 90% which is mostly driven by the outliers in the right tail.

sponds to trading-correlation weighted Information Diffusion Centrality, is not right-skewed and has a noticeably smaller variance.

Insert Figure 5

Figure 6 plots an example of a Weighted Trading Correlation Network in 2002Q2. Nodes are arranged using a force-directed layout with edge weights corresponding to the estimated trading correlations. Adjacent nodes are "attracted" to one another, while all nodes have "repulsive" forces. Lighter edges indicate lower trading correlations and darker edges indicate higher trading correlations. Node size corresponds to the number of connections (Degree) and node color corresponds to Information Diffusion Centrality. The most central node (○) using the force-directed layout appears on the right side because it has many distinct connections which provide it access to novel information.

Insert Figure 6

3.2 Event Summary Statistics

Table 2 Panel A reports the number of M&As, new product announcements, and director and manager deaths, as well as the average number of funds trading in the median event per year. On average there are 695 mergers, 7,749 new product announcements, and 26 deaths per year for which at one fund in my sample makes a round-trip trade. On average, 13 funds trade around the typical merger in a given year, 21 trade around new product announcements, and 8 trade around director and manager deaths. The CapIQ data starts in 2001, which is why the sample of new products and deaths is sparse in the early sample, and the trades data end in 2011Q3 which is why 2011 has fewer events.

Panel B includes average cumulative abnormal returns (CARs) from -30 to +30 based on a one-factor CAPM model as well as average principal-weighted (PW) and equal-weighted

(EW) average abnormal round-trip trading performance. The mean CAR for target firms around merger announcements is 6.49%, the mean CAR around new product announcements is -0.74%, and the mean CAR around director and manager deaths is 2.41%.[15] The mean principal-weighted (equal-weighted) average abnormal round-trip trading performance around merger announcements is 1.29% (1.27%) and is statistically different from zero, consistent with the observation that there are information leakages prior to merger announcements (see Betton, Eckbo, and Thorburn, 2008). The mean principal-weighted (equal-weighted) average trading performance around new product announcements is -0.03% (-0.05%) and is statistically significant but economically small. The mean principal-weighted (equal-weighted) average trading performance around deaths is -0.10% (-0.14%) and is statistically indistinguishable from zero, consistent with the notion that sudden deaths are unpredictable (e.g. Nguyen and Nielsen, 2010).

Insert Table 2

4 Centrality and Trading Performance

If valuable information diffuses through the network, then there should be predictable information asymmetry between investors due to the structure of the network. Central investors, in particular, should consistently outperform peripheral investors.

As formal test of this hypothesis, I estimate predictive regressions of the form:

$$r_{i,t} = \alpha_t + \boldsymbol{c}_{i,t-1}\boldsymbol{\beta}_0 + \boldsymbol{X}_{i,t-1}\boldsymbol{\beta}_1 + \epsilon_{i,t}. \tag{6}$$

$r_{i,t}$ is the average abnormal interim trading performance (in logs) computed using DGTW-

[15] The one-factor model is estimated over a year of data (252 days) with a 30-day gap between the estimation window and the event window. Other factor models produce similar results since most of the return is due to the announcement effect.

adjusted returns at the fund-quarter level, denoted by the subscripts i and t respectively. I include a quarter fixed effect, α_t, because I am interested in the cross-sectional variation in trading performance. $\boldsymbol{c}_{i,t-1}$ is the vector of centralities including Information Diffusion, Degree, and Eigenvector Centrality. I say that a fund is central if it has above median centrality in a given quarter. Finally, $\boldsymbol{X}_{i,t-1}$ is the vector of controls: trading volume, the number of trades, and trading performance all from the previous quarter and in logs.

Table 3 presents the main results of the paper. Panel A (B) of Table 3 shows OLS estimates from regressions of principal-weighted (equal-weighted) average abnormal interim trading performance on measures of centrality. In Column 1 of Panel A (B), funds with above median Information Diffusion Centrality have, on average, 0.32% (0.48%) higher principal-weighted (equal-weighted) average abnormal trading performance in the following quarter. The effect is statistically different from zero at the one percent level, based on standard errors clustered by fund and quarter. For comparison, the median fund in my sample has a 0.35% (0.47%) principal-weighted (equal-weighted) average abnormal interim trading performance. In other words, becoming central would nearly double the median fund's expected trading performance.

In contrast, Columns 2 and 3 of Table 3 show that funds with above median Degree and Eigenvector Centrality have lower trading performance. Central funds according to Degree Centrality have 0.11% (0.04%) lower trading performance, and central funds according to Eigenvector Centrality have 0.37% (0.42%) lower principal-weighted (equal-weighted) average abnormal interim trading performance. The parameter estimates for Degree are statistically indistinguishable from zero, and the estimates for Eigenvector are large and negative.

Column 4 reports estimates for a multiple regression specification with all three measures of centrality and controls for last quarter's trading volume, number of trades, and interim trading performance. The results remain unchanged, central funds according to Information Diffusion Centrality have 0.41% (0.52%) higher trading performance, and central funds according to Degree and Eigenvector Centrality have lower trading performance.

Column 5 repeats the multiple regression from Column 4 within a sample of central funds matched to peripheral funds based on Degree, Eigenvector, and the same set of controls. I require that the difference in propensity scores does not exceed 0.1 basis points in absolute value. Central funds, according to Information Diffusion Centrality, have 0.49% (0.57%) higher principal-weighted (equal-weighted) average abnormal interim trading performance. Within the smaller sample of 10,196 propensity score matched fund-quarter observations, the only observable difference between central and peripheral funds, according to Information Diffusion Centrality, is that central funds have higher Degree Centrality.[16] Funds with higher Degree tend to have lower trading performance, which biases against the finding that funds with higher Information Diffusion Centrality have higher trading performance as the two measures of centrality are positively correlated.

Insert Table 3

Overall, the results in Table 3 suggest that central funds according to Information Diffusion Centrality have a substantial information advantage. Using a measure of influence—e.g., Eigenvector Centrality—to proxy for information diffusion negatively predicts trading performance. Intuitively, connections to well-connected individuals provides commonly-held information which is not valuable. In contrast, Information Diffusion Centrality measures an investors' access to novel information which is why it predicts trading performance. Degree has a zero or negative coefficient because the relationship between trading performance and Degree is concave.

[16]Normalized differences are reported in Table 9 in the Appendix.

5 Evidence from Merger Announcements

In this section I restrict my attention to round-trip trading around merger announcements where I expect access to information to be the primary determinant of trading performance. For instance, Ahern (2015) examines 239 instances of illegal insider trading around M&As and finds evidence that individuals closer to the original source of information earn higher returns. Hence, if central investors are well-positioned to receive information on average, then they might also have superior access to information about mergers.

To test this hypothesis I estimate the same regression as in Equation 6, where the dependent variable is now the average abnormal round-trip trading performance based on trades made in target stock around merger announcements. As a robustness test, I restrict my sample to a set of propensity score matched funds as in Section 4. Because only a subset of funds trade around mergers, I jointly estimate the propensity to trade and the returns to being central using a Heckman (1976) selection model. The model can be written:

$$\begin{aligned} &r_{i,t} = \alpha_t + \boldsymbol{c}_{i,t-1}\boldsymbol{\beta}_0 + \boldsymbol{X}_{i,t-1}\boldsymbol{\beta}_1 + \epsilon^1_{i,t} && \text{(main regression)} \\ &\boldsymbol{Z}_{i,t-1}\boldsymbol{\gamma} + \epsilon^2_{i,t} && \text{(selection equation)} \end{aligned} \tag{7}$$

where $\epsilon^1 \sim N(0,\sigma)$, $\epsilon^2 \sim N(0,1)$, and $\text{corr}(\epsilon^1,\epsilon^2) = \rho$. The assumption is that $r_{i,t}$ is observed, i.e. a fund trades in mergers, if $\boldsymbol{Z}_{i,t-1}\boldsymbol{\gamma} + \epsilon^2_{i,t} > 0$. I include an indicator for whether a fund's initial Centrality, Degree, and Eigenvector is above median as additional instruments in Z.

Table 4 Columns 1–3 (4–6) present estimates from regressions of principal-weighted (equal-weighted) average abnormal round-trip trading performance on measures of centrality and trading activity. In Column 1 (4), central funds have 1.50% (1.53%) higher principal-weighted (equal-weighted) average abnormal round-trip trading performance. To put the magnitude of the estimate in perspective, the median fund has a 1.27% (1.24%) principal-weighted (equal-weighted) average abnormal round-trip trading performance around merger announcements. Becoming central would more than double the median fund's expected trading performance. Funds with above median Degree Centrality have 1.39% (1.47%) lower trading performance, and funds with above median Eigenvector Centrality have 3.9% (3.8%) lower trading performance.

In Column 2 (5), within the sample of central funds matched to peripheral funds, central funds have 2.46% (2.45%) higher principal-weighted (equal-weighted) average abnormal round-trip trading performance. Funds with above median Degree Centrality have 1.13% (1.17%) lower trading performance, and funds with above median Eigenvector Centrality have 3.16% (3.12%) lower trading performance.

The model in Column 3 (6) controls for the propensity to trade around mergers within the matched sample. As before, central funds have 2.55% (2.58%) higher principal-weighted (equal-weighted) average abnormal round-trip trading performance. Funds with above median Degree Centrality have 1.33% (1.42%) lower trading performance, and funds with above median Eigenvector Centrality have 2.99% (2.95%) lower trading performance.

Insert Table 4

Taken together, the above evidence suggests that central funds appear to be informed about mergers before they are publicly announced. Central funds have more than three times higher round-trip trading performance relative to the average fund within the matched sample and controlling for the likelihood of trading around mergers. Degree and Eigenvector Centrality are unable to predict superior trading performance because they do not capture the diffusion of information.

6 Ruling Out Alternative Hypotheses

In this section I show that centrality is unrelated to common sources of information, industry or product market expertise, and does not predict trading performance spuriously when there is no private information.

6.1 M&A Advisor "Tipping"

Large broker-dealers frequently act as merger advisors to acquiror and target firms and potentially "tip off" their institutional clients to these deals. Jegadeesh and Tang (2011) find that institutions connected to merger advisors are net buyers of target stocks—they also make profitable trades in aggregate. Therefore, brokerage house connections may be an important source of common information about mergers.

I begin by name-matching ANcerno brokers to SDC merger advisors on both the target and acquiror side. In most M&A deals there are multiple advisory firms on each side, so there are potentially many sources of informative tips. In order to try and isolate the relevant connections, I compute the total shares traded by each fund with each of their brokers for each quarter. I say that a fund is connected to an M&A advisor if in the previous quarter the advisor was one of the top-quintile brokers for that fund.[17] I then repeat the round-trip trading analysis with an indicator for whether a fund has any connections to an M&A advisor (Tip=1) in the previous quarter.

Table 5 Column 1 (4) shows that funds connected to M&A advisors have 0.91% (0.87%) higher principal-weighted (equal-weighted) average abnormal round-trip trading performance. Central funds, those with higher than median Information Diffusion Centrality, still have 1.40% (1.44%) higher principal-weighted (equal-weighted) trading performance. Both estimates are statistically significant at the one percent level.

[17] Jegadeesh and Tang (2011) focus on top decile brokers at the yearly level.

In Column 2 (5), within the sample of central funds matched to peripheral funds, funds connected to M&A advisors have 0.63% (0.78%) higher trading performance although the estimate is not statistically significant. Central funds have 2.45% (2.43%) higher principal-weighted (equal-weighted) average abnormal round-trip trading performance.

The model in Column 3 (6) controls for the propensity to trade around mergers within the matched sample. Funds connected to M&A advisors have 0.39% (0.54%) higher trading performance and the estimate is not statistically significant. As before, central funds have 2.54% (2.57%) higher principal-weighted (equal-weighted) average abnormal round-trip trading performance.

Insert Table 5

Controlling for brokerage house connections does not attenuate the returns to being central by much. The returns to being central in Column 1 (3) are only ten basis points lower than are reported in Table 4. In Columns 2–3 (5–6) the returns to being central are only one basis point lower. These results suggest that centrality is largely unrelated to brokerage house connections.

6.2 Expertise versus Information Diffusion

Trading performance may also be related to expertise in a particular industry or product market (Kacperczyk, Sialm, and Zheng, 2005). I examine new product announcements as a setting in which I expect product market expertise to be distinguishable from centrality as predictors of trading performance.

To measure product market expertise I rely on the text-based network industry classifications (`TNICs`) provided by Hoberg and Phillips (2010, 2015). Hoberg and Phillips (2010, 2015) classify firms into `TNICs` industries using the product descriptions in 10-K filings. Each product announcement in my sample is assigned to an industry based on yearly `GVKEY-TNIC` pairs. I compute a fund's expertise in the announcing firm's product market as the trade-weighted average cosine similarity between the announcing firm's product description and the product descriptions of all of the other firms in which the fund trades. The resulting measure is bounded between zero and one, where larger values indicate that a fund's trading is concentrated in the announcing firm's industry. Intuitively, if a fund exclusively trades in technology stocks such as `AAPL`, `MSFT`, and `IBM` then the fund is classified as an expert in the `TNIC` technology industry.[18] I say that a fund is an Expert = 1 if it has above median average expertise, measured across all round-trip trades in announcing firm stock in a given quarter.

In Table 5 Column 1 (5), central funds, those with higher than median Information Diffusion Centrality, have 0.65% (0.65%) higher principal-weighted (equal-weighted) average abnormal round-trip trading performance. Controlling for expertise, the returns to being central appear to be higher. In Column 2 (6), Experts have 0.51% (0.49%) higher principal-weighted (equal-weighted) average trading performance, and central funds have 0.66% (0.66%) higher trading performance.

In Column 3 (7) I construct a sample of Experts matched to non-Experts.[19] In the matched sample Experts have 0.75% (0.72%) higher principal-weighted (equal-weighted) average trading performance, and central funds have 0.80% (0.80%) higher trading performance.

The model in Column 4 (8) controls for the propensity to trade around product announcements within the matched sample. Experts have 0.55% (0.53%) higher principal-weighted (equal-weighted) average trading performance, and central funds have 0.75% (0.75%) higher

[18] `TNIC` definitions change over time as product categories change. Because I am only concerned with whether a fund trades stocks similar to the announcing firm within a given year, the time series variation in `TNIC`s does not affect my definition of expertise. See Hoberg and Phillips (2010, 2015) for more details.

[19] I form matches based on log standardized Information Diffusion Centrality and the same covariates reported in Section 4. A matched sample of central and peripheral investors produces similar magnitude estimates.

trading performance.

Insert Table 6

The results in Table 6 indicate that both Centrality and Expertise are significant predictors of round-trip trading performance around new product announcements. More importantly, centrality is distinct from expertise.

6.3 (No) Timing Sudden Director and Manager Deaths

An important concern is whether central fund managers have genuine information or whether centrality predicts trading performance due to pseudo market timing. For example, Butler, Grullon, and Weston (2005) finds that managers pseudo market time equity markets with their share issuances and appear to be able to forecast unanticipated events such as the Japanese attack of Pearl Harbor.

As a falsification test, I look at round-trip trading around sudden deaths of directors and managers. Because sudden deaths are unpredictable—centrality—as a proxy for access to information, should play no role in the performance of trades made around these events.

Consistent with this intuition, Table 7 Column 1 (4) shows that central funds do not have statistically significant higher trading performance relative to the average fund.

In Column 2 (5) I repeat the regression in Column 1 (4) within a sample of central funds matched to peripheral funds based on Degree, Eigenvector, and the same set of controls. I require that the difference in propensity scores does not exceed five basis points in absolute value.[20] Again central funds do not have statistically significant higher trading performance relative to the average fund.

The model in Column 3 (6) controls for the propensity to trade around mergers within

[20]Using a caliper of 0.1 basis points reduces the sample size to roughly 200 fund-quarter observations which produces unreliable parameter and standard error estimates.

the matched sample. As before, central funds do not have statistically significant higher trading performance relative to the average fund.

Insert Table 7

When there is no information to share, there is no information advantage to being central.

7 Conclusion

Overall my findings suggest that institutional investors regularly share valuable information with one another. Information Diffusion Centrality predicts trading performance, and is also very persistent, which suggests that the consistent outperformance observed in the literature, e.g. Kacperczyk, Sialm, and Zheng (2008) and Puckett and Yan (2011), is due to a conversational equilibrium in the spirit of Stein (2008). My findings further indicate that word-of-mouth information sharing may be a stylized fact of investor behavior which is not limited to individual investors (Ivković and Weisbenner, 2007) or local institutions (Hong, Kubik, and Stein, 2005; Pool, Stoffman, and Yonker, 2015).

My results support the view that the way in which economic agents converse—i.e., the structure of social and information networks—have important implications for financial markets. Shiller (1995) argues that many behavioral biases may be explained by patterns of conversations, and Hirshleifer (2015) argues that the next step in behavioral finance is to study *social* finance—in other words, to incorporate theories of sociology into our understanding of financial markets. Herding, home bias, and underreaction, three of the most well-documented behavioral biases, are all arguably related to the localization of information embedded within the community structure of networks. The sociology literature has long recognized that economic actions such as learning may be "embedded" in social networks (Granovetter, 1985), but network analysis is relatively new to the finance literature. A more detailed subsequent analysis of network structure, beyond centrality, may provide a fuller understanding of the distribution and dispersion of information in financial markets.

Novel information often flows through "weak ties" (Bakshy, Rosenn, Marlow, and Adamic, 2012)—hence empirical work should take advantage of the continuous nature of social interactions.

Finally, I find that using measures of influence as proxies for information diffusion can lead to different conclusions. Empirical researchers in finance should be careful in choosing measures of centrality and understanding the implicit network flows.

References

Ahern, Kenneth R, 2013, Network centrality and the cross section of stock returns, *Available at SSRN 2197370.*

——— , 2015, Information networks: Evidence from illegal insider trading tips, *Available at SSRN.*

Akins, Brian K, Jeffrey Ng, and Rodrigo S Verdi, 2011, Investor competition over information and the pricing of information asymmetry, *The Accounting Review* 87, 35–58.

Anand, Amber, Paul Irvine, Andy Puckett, and Kumar Venkataraman, 2013, Institutional trading and stock resiliency: Evidence from the 2007–2009 financial crisis, *Journal of Financial Economics* 108, 773–797.

Bakshy, Eytan, Itamar Rosenn, Cameron Marlow, and Lada Adamic, 2012, The role of social networks in information diffusion, in *Proceedings of the 21st international conference on World Wide Web* pp. 519–528. ACM.

Banerjee, Abhijit, Arun G Chandrasekhar, Esther Duflo, and Matthew O Jackson, 2013, The diffusion of microfinance, *Science* 341, 1236498.

Barabási, Albert-László, and Réka Albert, 1999, Emergence of scaling in random networks, *Science* 286, 509–512.

Betton, Sandra, B. Espen Eckbo, and Karin S. Thorburn, 2008, Chapter 15 - corporate takeovers, in *Handbook of Empirical Corporate Finance*, ed. by B. Espen Eckbo, vol. 2 of *Handbooks in Finance* pp. 291–429.

Boehmer, Ekkehart, and Eric K Kelley, 2009, Institutional investors and the informational efficiency of prices, *Review of Financial Studies* 22, 3563–3594.

Bonacich, Phillip, 1972a, Factoring and weighting approaches to status scores and clique identification, *Journal of Mathematical Sociology* 2, 113–120.

——— , 1972b, Technique for analyzing overlapping memberships, *Sociological Methodology* 4, 176–185.

Borgatti, Stephen P, 2005, Centrality and network flow, *Social Networks* 27, 55–71.

Brin, Sergey, and Lawrence Page, 1998, The anatomy of a large-scale hypertextual web search engine, *Computer Networks and ISDN Systems* 30.

Brown, Jeffrey R, Zoran Ivković, Paul A Smith, and Scott Weisbenner, 2008, Neighbors matter: Causal community effects and stock market participation, *The Journal of Finance* 63, 1509–1531.

Brunnermeier, Markus K, 2005, Information leakage and market efficiency, *Review of Financial Studies* 18, 417–457.

Butler, Alexander W, Gustavo Grullon, and James P Weston, 2005, Can managers forecast aggregate market returns?, *The Journal of Finance* 60, 963–986.

Chung, Fan, 2007, The heat kernel as the pagerank of a graph, *Proceedings of the National Academy of Sciences* 104, 19735–19740.

——— , and Shing-Tung Yau, 1998, Coverings, heat kernels and spanning trees, *Journal of Combinatorics* 6, 163–184.

Cohen-Cole, Ethan, Andrei Kirilenko, and Eleonora Patacchini, 2014, Trading networks and liquidity provision, *Journal of Financial Economics* 113, 235–251.

Colla, Paolo, and Antonio Mele, 2010, Information linkages and correlated trading, *Review of Financial Studies* 23, 203–246.

Coval, Joshua D, and Tobias J Moskowitz, 1999, Home bias at home: Local equity preference in domestic portfolios, *Journal of Finance* 54, 2045–2073.

——— , 2001, The geography of investment: Informed trading and asset prices, *Journal of Political Economy* 109, 811–841.

Daniel, Kent, Mark Grinblatt, Sheridan Titman, and Russ Wermers, 1997, Measuring mutual fund performance with characteristic-based benchmarks, *The Journal of Finance* 52, 1035–1058.

DeMarzo, Peter M, Dimitri Vayanos, and Jeffrey Zwiebel, 2003, Persuasion bias, social influence, and unidimensional opinions, *The Quarterly Journal of Economics* 118, 909–968.

El-Khatib, Rwan, Kathy Fogel, and Tomas Jandik, 2015, Ceo network centrality and merger performance, *Journal of Financial Economics* 116, 349–382.

Ellison, Glenn, and Drew Fudenberg, 1995, Word-of-mouth communication and social learning, *The Quarterly Journal of Economics* pp. 93–125.

Engelberg, Joseph, Pengjie Gao, and Christopher A Parsons, 2013, The price of a ceo's rolodex, *Review of Financial Studies* 26, 79–114.

Granovetter, Mark, 1985, Economic action and social structure: the problem of embeddedness, *American Journal of Sociology* 91, 481–510.

——— , 2005, The impact of social structure on economic outcomes, *Journal of Economic Perspectives* pp. 33–50.

Granovetter, Mark S, 1973, The strength of weak ties, *American Journal of Sociology* 78, 1360–1380.

Han, Bing, and Liyan Yang, 2013, Social networks, information acquisition, and asset prices, *Management Science* 59, 1444–1457.

Heckman, James J, 1976, The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models, in *Annals of Economic and Social Measurement*vol. 5 pp. 475–492.

Hendershott, Terrence, Dmitry Livdan, and Norman Schürhoff, 2015, Are institutions informed about news?, *Journal of Financial Economics* 112, 249–287.

Hirshleifer, David, 2015, Behavioral finance, *Annual Review of Financial Economics* 7.

——— , Avanidhar Subrahmanyam, and Sheridan Titman, 1994, Security analysis and trading patterns when some investors receive information before others, *Journal of Finance* 49, 1665–1698.

Hoberg, Gerard, and Gordon Phillips, 2010, Product market synergies and competition in mergers and acquisitions: A text-based analysis, *Review of Financial Studies* 23, 3773–3811.

Hoberg, Gerard, and Gordon M Phillips, 2015, Text-based network industries and endogenous product differentiation, *Journal of Political Economy, Forthcoming.*

Hochberg, Yael V, Laura Anne Lindsey, and Mark M Westerfield, 2015, Resource accumulation through economic ties: Evidence from venture capital, *Journal of Financial Economics (JFE), Forthcoming.*

Hochberg, Yael V, Alexander Ljungqvist, and Yang Lu, 2007, Whom you know matters: Venture capital networks and investment performance, *The Journal of Finance* 62, 251–301.

———, 2010, Networking as a barrier to entry and the competitive supply of venture capital, *The Journal of Finance* 65, 829–859.

Hong, Harrison, Jeffrey D Kubik, and Jeremy C Stein, 2005, Thy neighbor's portfolio: Word-of-mouth effects in the holdings and trades of money managers, *The Journal of Finance* 60, 2801–2824.

Horvath, Steve, 2011, *Weighted Network Analysis: Applications in Genomics and Systems Biology* (Springer Science & Business Media).

Imbens, Guido W., and Jeffrey M. Wooldridge, 2009, Recent developments in the econometrics of program evaluation, *Journal of Economic Literature* 47, 5–86.

Ivković, Zoran, and Scott Weisbenner, 2005, Local does as local is: Information content of the geography of individual investors' common stock investments, *The Journal of Finance* 60, 267–306.

———, 2007, Information diffusion effects in individual investors' common stock purchases: Covet thy neighbors' investment choices, *Review of Financial Studies* 20, 1327–1357.

Jegadeesh, Narasimhan, and Yue Tang, 2011, Institutional trades around takeover announcements: Skill vs. inside information, *SSRN Working Paper Series*.

Kacperczyk, Marcin, Clemens Sialm, and Lu Zheng, 2005, On the industry concentration of actively managed equity mutual funds, *The Journal of Finance* 60, 1983–2011.

———, 2008, Unobserved actions of mutual funds, *Review of Financial Studies* 21, 2379–2416.

Kahneman, Daniel, 2011, *Thinking, fast and slow* (Macmillan).

Larcker, David F, Eric C So, and Charles CY Wang, 2013, Boardroom centrality and firm performance, *Journal of Accounting and Economics* 55, 225–250.

Nguyen, Bang Dang, and Kasper Meisner Nielsen, 2010, The value of independent directors: Evidence from sudden deaths, *Journal of Financial Economics* 98, 550–567.

Ozsoylev, Han N, and Johan Walden, 2011, Asset pricing in large information networks, *Journal of Economic Theory* 146, 2252–2280.

——— , M Deniz Yavuz, and Recep Bildik, 2014, Investor networks in the stock market, *Review of Financial Studies* 27, 1323–1366.

Page, Lawrence, Sergey Brin, Rajeev Motwani, and Terry Winograd, 1999, The pagerank citation ranking: bringing order to the web., .

Pool, Veronika K, Noah Stoffman, and Scott E Yonker, 2015, The people in your neighborhood: Social interactions and mutual fund portfolios, *forthcoming in The Journal of Finance.*

Puckett, Andy, and Xuemin Sterling Yan, 2011, The interim trading skills of institutional investors, *The Journal of Finance* 66, 601–633.

Saavedra, Serguei, Kathleen Hagerty, and Brian Uzzi, 2011, Synchronicity, instant messaging, and performance among financial traders, *Proceedings of the National Academy of Sciences* 108, 5296–5301.

Shiller, Robert J, 1995, Conversation, information, and herd behavior, *The American Economic Review* 85, 181–185.

——— , and John Pound, 1989, Survey evidence on diffusion of interest and information among investors, *Journal of Economic Behavior & Organization* 12, 47–66.

Sias, Richard W, and Laura T Starks, 1997, Return autocorrelation and institutional investors, *Journal of Financial economics* 46, 103–131.

Stein, Jeremy C, 2008, Conversations among competitors, *The American Economic Review* 98, 2150–2162.

Valente, Thomas W, Kathryn Coronges, Cynthia Lakon, and Elizabeth Costenbader, 2008, How correlated are network centrality measures?, *Connections* 28, 16.

Walden, Johan, 2014, Trading, profits, and volatility in a dynamic information network model, *Available at SSRN 2561055.*

Table 1: **Fund Summary Statistics 1999Q1–2011Q3.** Interim trading performance is measured quarterly as the difference between a fund's principal- or equal-weighted average excess buy return minus its average excess sell return. Returns are computed as the simple return from the actual execution price to the end of quarter closing price including dividends and are also DGTW-adjusted over the same holding period. The principal weight is computed using the actual price $\times$ volume traded. Prices and volume are adjusted for stock splits, and benchmark returns are adjusted for delistings. Centrality is Information Diffusion Centrality defined in Eq. (4), and Eigenvector is Eigenvector Centrality defined in Eq. (3). Both measures are computed based on quarterly trades and reported as percentages of the "total" centrality. Degree is the sum of a fund's weighted edges (trading correlations) in a given quarter. Volume refers to fund-quarter share volume traded, and # of Trades refers to the actual number of transactions. Correlations in Panel B are Spearman correlations and the diagonal entries are autocorrelations.

(a) Summary Statistics

	N	Mean	Med.	Std	Skew	Kurt
PW	54,796	0.52%	0.32%	6.3%	−0.01	5.55
EW	54,796	0.65%	0.46%	5.9%	−0.19	6.00
Centrality	54,796	2.9%	2.9%	0.60%	−1.86	9.68
Eigenvector	54,796	0.75%	0.07%	2.9%	7.27	59.55
Degree	54,796	4.89	4.09	3.04	2.19	7.09
Volume	54,796	19.03M	669,819	188.40M	33.53	2,287.08
# of Trades	54,796	2,401	171	20,802	25.15	959.94

(b) Correlations

	PW	EW	Cent	Eig	Deg	Vol	Trades
PW	0.048	.	.	.	.	.	.
EW	0.77	0.065	.	.	.	.	.
Cent	0.0021	0.0036	0.91	.	.	.	.
Eig	-0.051	-0.070	0.52	0.72	.	.	.
Deg	-0.026	-0.021	0.23	0.55	0.91	.	.
Vol	-0.030	-0.023	0.14	0.20	0.21	0.84	.
Trades	-0.00048	0.014	0.31	0.25	0.37	0.66	0.87

Table 2: **Event Summary Statistics.** Panel A shows the number of merger announcements, new product announcements, and director and manager deaths per year, as well as the average number of funds trading around the median event. Panel B shows the average cumulative abnormal returns (CARs) and round-trip trading performance around events. Merger announcements come from the Thomson Reuters SDC Platinum database and director and manager sudden deaths and new product announcements come from the S&P Capital IQ Key Developments Database (event types 101, 102, 16 for deaths and 41 for new products). CARs are measured from [-30, +30] using a one-factor CAPM model estimated over a 252-day window with a 30-day gap between the estimation window and the event window. Round-trip trading performance is measured as the principal- or equal-weighted average of all signed simple excess returns of buy (sell) trades initiated between [-60, -1] and reversed between [0, +30]. t-statistics are based on un-adjusted standard errors.

(a) Distribution of Events

	Mergers		New Products		Deaths	
Year	# Events	# Funds	# Events	# Funds	# Events	# Funds
1999	824	6	.	.	.	.
2000	986	7	.	.	.	.
2001	668	7	125	11	1	2
2002	597	11	7178	16	6	5
2003	688	11	7384	18	11	10
2004	743	15	8835	19	20	9
2005	803	12	9533	16	37	5
2006	817	13	9168	20	35	10
2007	829	15	9084	23	42	7
2008	697	14	9860	25	48	11
2009	554	16	9174	29	39	14
2010	507	16	8865	27	36	10
2011	324	20	6036	23	17	6

(b) Event Returns

Event	# Events	# Trades	CAR	t	PW	t	EW	t
M&A Targets	9,242	171,968	6.49%	19.52	1.29%	32.97	1.27%	34.34
New Products	84,465	2,770,048	−0.74%	−8.50	−0.03%	−4.91	−0.05%	−8.78
Deaths	219	4,274	2.41%	1.18	−0.10%	−0.55	−0.14%	−0.90

Table 3: **Interim Trading Performance and Centrality.** In Panel A (B) the dependent variable is the quarterly fund-level log principal-weighted (equal-weighted) average abnormal interim trading performance. Centrality, Degree, and Eigenvector are indicator variables which are one if the fund has above median Information Diffusion, Degree, or Eigenvector Centrality in the previous quarter $t-1$. Control variables include the previous quarter's trading volume, number of trades, and trading performance, all of which are in logs and are standardized. Column 5 contains central and peripheral funds matched on Degree and Eigenvector (standardized), and the same control variables where the difference in propensity scores does not exceed 0.1 basis points in absolute value. Coefficients and R^2 values are reported in percentages. t-statistics are reported in parentheses, with standard errors clustered by fund and quarter. Stars indicate significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

(a) Principal-Weighted

	(1)	(2)	(3)	(4)	(5)
Centrality_{t-1}	0.322***			0.414***	0.486***
	(4.64)			(4.83)	(2.83)
Degree_{t-1}		-0.109		-0.162	-0.408**
		(-1.08)		(-1.37)	(-2.23)
Eigenvector_{t-1}			-0.365***	-0.378***	-0.238
			(-3.42)	(-3.54)	(-1.19)
Qtr FE	Yes	Yes	Yes	Yes	Yes
Controls	No	No	No	Yes	Yes
R^2(%)	0.64	0.59	0.66	1.00	1.88
Obs.	54,796	54,796	54,796	54,782	11,045

(b) Equal-Weighted

	(1)	(2)	(3)	(4)	(5)
Centrality_{t-1}	0.475***			0.517***	0.565***
	(6.71)			(6.29)	(3.33)
Degree_{t-1}		-0.0417		-0.112	-0.328*
		(-0.49)		(-1.14)	(-1.87)
Eigenvector_{t-1}			-0.419***	-0.515***	-0.360*
			(-4.57)	(-5.23)	(-1.99)
Qtr FE	Yes	Yes	Yes	Yes	Yes
Controls	No	No	No	Yes	Yes
R^2(%)	0.64	0.49	0.61	1.26	1.80
Obs.	54,796	54,796	54,796	54,782	11,045

Table 4: **Round-trip Trading around M&A Announcements — Targets.** Merger announcements come from the Thomson Reuters SDC Platinum Database. The dependent variable is the quarterly fund-level principal-weighted (equal-weighted) average abnormal round-trip trading performance in target stock around merger announcements. Centrality, Degree, and Eigenvector are indicator variables which are one if the fund has above median Information Diffusion, Degree, or Eigenvector Centrality in the previous quarter $t-1$. Control variables include the previous quarter's trading volume, number of trades, and trading performance, all of which are in logs and are standardized. Columns 2 and 5 contain central and peripheral funds matched on Degree and Eigenvector (standardized), and the same control variables where the difference in propensity scores does not exceed 0.1 basis points in absolute value. In Columns 3 and 6 I include centrality indicators based on the first fund-quarter observation as additional instruments in a Heckman selection model (Equation 7) estimated using maximum likelihood. λ refers to the coefficient on the inverse Mills ratio, and standard errors are clustered at the quarterly level. Coefficients and R^2 values are reported in percentages. OLS t-statistics are reported in parentheses, with standard errors clustered by fund and quarter. Stars indicate significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

	Principal Weighted			Equal Weighted		
	(1)	(2)	(3)	(4)	(5)	(6)
Centrality_{t-1}	1.495***	2.463***	2.546***	1.531***	2.448***	2.576***
	(4.29)	(3.22)	(3.47)	(4.29)	(3.25)	(3.51)
Degree_{t-1}	-1.389**	-1.132	-1.335	-1.465**	-1.171	-1.420
	(-2.34)	(-1.11)	(-1.29)	(-2.48)	(-1.15)	(-1.36)
Eigenvector_{t-1}	-3.907***	-3.159***	-2.985***	-3.839***	-3.128***	-2.953***
	(-6.60)	(-3.04)	(-2.94)	(-6.54)	(-2.98)	(-2.90)
λ(%)			0.34			0.31
$s.e.(\lambda)$			0.36			0.36
Qtr FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
R^2(%)	2.40	4.69		2.50	4.73	
Obs.	21,658	4,171	11,544	21,658	4,171	11,544

Table 5: **Tipping around M&A Announcements — Targets.** A fund is connected to a brokerage house (Tip=1) if the brokerage house was one of the funds' main brokers in the last quarter. A fund's brokers are ranked according to share volume executed in each quarter, and top quintile brokers are identified as main brokers. Control variables include the previous quarter's trading volume, number of trades, and trading performance, all of which are in logs and are standardized. Coefficients and R^2 values are reported in percentages. OLS t-statistics are reported in parentheses, with standard errors clustered by fund and quarter. Stars indicate significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

	Principal Weighted			Equal Weighted		
	(1)	(2)	(3)	(4)	(5)	(6)
Centrality$_{t-1}$	1.400***	2.448***	2.540***	1.440***	2.430***	2.567***
	(4.02)	(3.20)	(3.45)	(4.06)	(3.23)	(3.48)
Degree$_{t-1}$	-1.491**	-1.219	-1.395	-1.562**	-1.276	-1.500
	(-2.52)	(-1.20)	(-1.37)	(-2.65)	(-1.26)	(-1.46)
Eigenvector$_{t-1}$	-3.851***	-3.106***	-2.938***	-3.787***	-3.064***	-2.890***
	(-6.56)	(-3.02)	(-2.94)	(-6.50)	(-2.95)	(-2.88)
Tip	0.906***	0.632	0.392	0.869***	0.776	0.538
	(3.93)	(0.95)	(0.53)	(3.72)	(1.16)	(0.72)
λ(%)			0.37			0.35
$s.e.(\lambda)$			0.37			0.37
Qtr FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
R^2(%)	2.48	4.73		2.57	4.78	
Obs.	21,658	4,171	11,544	21,658	4,171	11,544

Table 6: **Round-trip Trading around New Product Announcements.** New product events are identified in the S&P Capital IQ Key Events Database as event type 41. Hoberg and Phillips (2010, 2015) classify firms into industries (`TNICs`) using the product descriptions found in 10-K filings. Each product announcement is assigned to an industry based on yearly `GVKEY-TNIC` pairs. A fund's product market expertise is measured as the trade-weighted average cosine similarity between the announcing firm's product description and the product descriptions of all of the other firms in which the fund traded stocks. A fund is an Expert = 1 if it has above median average expertise in a given quarter. Columns 3 and 7 contain Expert and non-Expert funds matched on on Degree and Eigenvector (standardized), and controls for trading activity. I require that the difference in propensity scores does not exceed 0.1 basis points in absolute value. In Columns 4 and 8 I include centrality indicators based on the first fund-quarter observation as additional instruments in a Heckman selection model (Equation 7) estimated using maximum likelihood. λ refers to the coefficient on the inverse Mills ratio, and standard errors are clustered at the quarterly level. Coefficients and R^2 values are reported in percentages. OLS t-statistics are reported in parentheses, with standard errors clustered by fund and quarter. Stars indicate significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

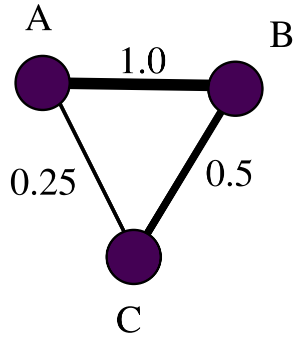
	Principal Weighted				Equal Weighted			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Centrality$_{t-1}$	0.680***	0.688***	0.832***	0.781***	0.652***	0.660***	0.803***	0.751***
	(3.08)	(3.14)	(2.98)	(3.02)	(3.00)	(3.05)	(2.96)	(2.99)
Degree$_{t-1}$	-0.414	-0.406	-0.281	-0.275	-0.431	-0.423	-0.286	-0.283
	(-1.22)	(-1.19)	(-0.64)	(-0.66)	(-1.29)	(-1.26)	(-0.66)	(-0.68)
Eigenvector$_{t-1}$	-0.746**	-0.756**	-0.817**	-0.792**	-0.685**	-0.695**	-0.754**	-0.726**
	(-2.21)	(-2.24)	(-2.22)	(-2.43)	(-2.06)	(-2.08)	(-2.04)	(-2.20)
Expert		0.522***	0.741***	0.550***		0.491***	0.722**	0.521**
		(2.86)	(2.73)	(2.59)		(2.72)	(2.70)	(2.48)
λ(%)				0.017				0.024
$s.e.(\lambda)$				0.14				0.13
Qtr FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2(%)	1.15	1.19	1.62		1.33	1.37	1.78	
Obs.	27,219	27,219	17,662	35,008	27,219	27,219	17,662	35,008

Table 7: **Round-trip Trading around Sudden Director & Manager Deaths.** Director or manager changes are identified in the S&P Capital IQ Key Events Database as event types 101 (CEO), 102 (CFO), and 16 (Board and Other Executive). I identify sudden deaths by searching the headlines for key phrases including `death of`, `demise of`, and `passing of`. I filter out deaths of individuals over the age of 65, deaths related to cancer, suicide, disease, illness, or complications as reported in the full text articles. Columns 2 and 5 contain central and peripheral funds matched on Degree and Eigenvector (standardized), and controls for trading activity. I require that the difference in propensity scores does not exceed five basis points in absolute value. In Columns 3 and 6 I include centrality indicators based on the first fund-quarter observation as additional instruments in a Heckman selection model (Equation 7) estimated using maximum likelihood. λ refers to the coefficient on the inverse Mills ratio, and standard errors are clustered at the quarterly level. Coefficients and R^2 values are reported in percentages. OLS t-statistics are reported in parentheses, with standard errors clustered by fund and quarter. Stars indicate significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

	Principal Weighted			Equal Weighted		
	(1)	(2)	(3)	(4)	(5)	(6)
Centrality_{t-1}	1.059	0.236	0.223	0.884	-0.222	-0.160
	(0.81)	(0.23)	(0.22)	(0.74)	(-0.25)	(-0.19)
Degree_{t-1}	-1.910	-0.309	-0.381	-1.722	0.0904	0.00639
	(-1.48)	(-0.22)	(-0.28)	(-1.45)	(0.066)	(0.0050)
Eigenvector_{t-1}	1.001	0.191	0.216	0.874	0.0635	0.181
	(0.75)	(0.11)	(0.13)	(0.62)	(0.034)	(0.11)
λ(%)			-0.14			0.087
$s.e.(\lambda)$			0.46			0.26
Qtr FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
R^2(%)	0.84	0.64		0.76	0.44	
Obs.	2,277	1,880	48,362	2,277	1,880	48,362

Figure 1: **Triangle Network.** A network is defined as a collection of nodes (●) and edges (▬). Panel A shows the geometric representation of the network and Panel B shows the (weighted) adjacency matrix representation. The adjacency matrix is symmetric, and the edge weights are on the off-diagonal entries.

(a) Geometric Representation



(b) Adjacency Matrix

$$A = \begin{bmatrix} 0 & 1 & 0.25 \\ 1 & 0 & 0.5 \\ 0.25 & 0.5 & 0 \end{bmatrix}$$

Figure 2: **Random Walks on a Network.** The standard random walk in Panel A assumes that a random walker from A jumps with equal probability $\frac{1}{4}$ to any one of the nodes B–E, while a random walker from B jumps with probability 1 to A. The symmetric random walk in Panel B assumes that a random walker from A jumps to B with the same probability that a random walker from B jumps to A, which in this example is $\frac{1}{2}$.

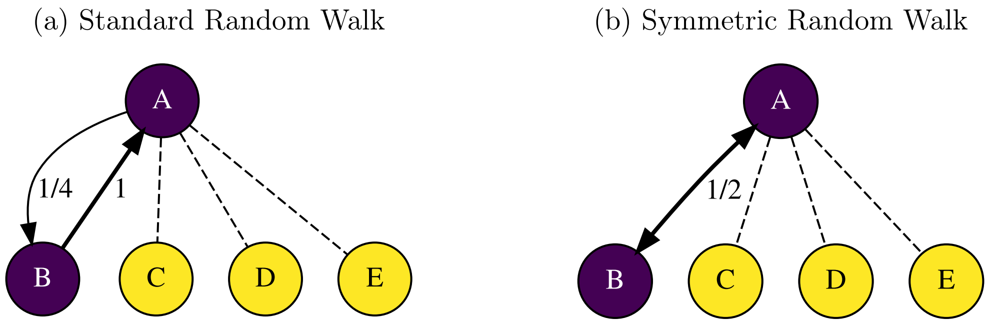


Figure 3: **Weighted Trading Correlation Network with an active liquidity trader.** The "true" star network consists of five informed investors (●) connected to one another via black lines (▬). Adding an active liquidity trader (●) results in many un-informative linkages to the informed investors. Using Weighted Trading Correlation Network estimation attenuates the strength of the un-informed connections (▬). Node size corresponds to the sum of the weighted edges such that the liquidity trader is "unimportant" and on the periphery.

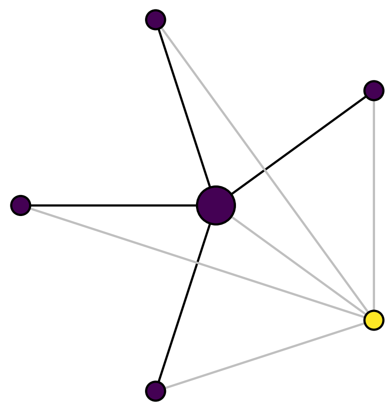


Figure 4: **Information Diffusion vs. Eigenvector Centrality.** The central node C has six connections, five unique connections, and one to a fully connected cluster of seven nodes. Node size is proportional to the number of connections and node color is proportional to Information Diffusion Centrality on the left and Eigenvector Centrality on the right. Information Diffusion Centrality identifies C as the most central node, whereas the nodes in the fully connected cluster are not central despite the fact that each of them has many connections. Eigenvector Centrality identifies C as being peripheral because most of its connections are not well-connected.

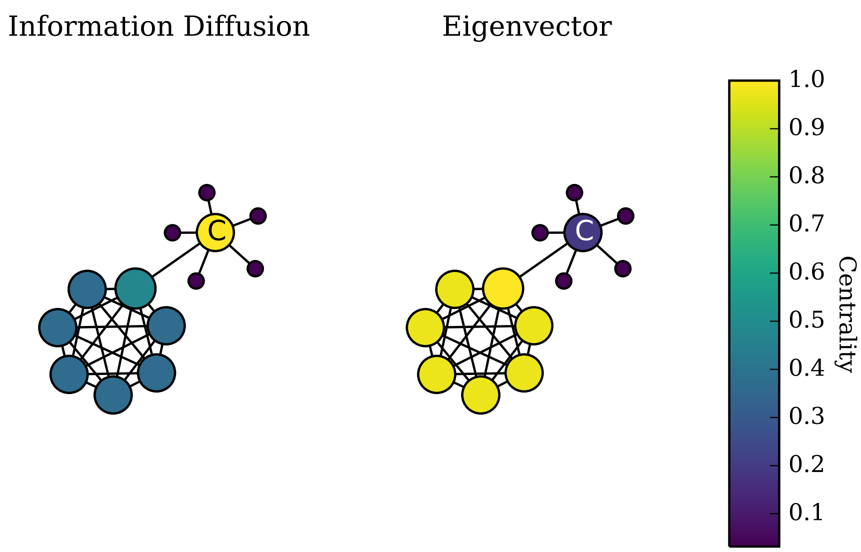


Figure 5: **Centrality 1999Q1–2011Q3.** This Figure plots the distribution of Information Diffusion Centrality (Eq. 4) over the sample period. The solid lines represent the median, and the shaded regions represent the distribution between the 5 and 95 percentiles of Information Diffusion Centrality. The Figure includes Information Diffusion Centrality based on unweighted (—), and weighted (—) edges.

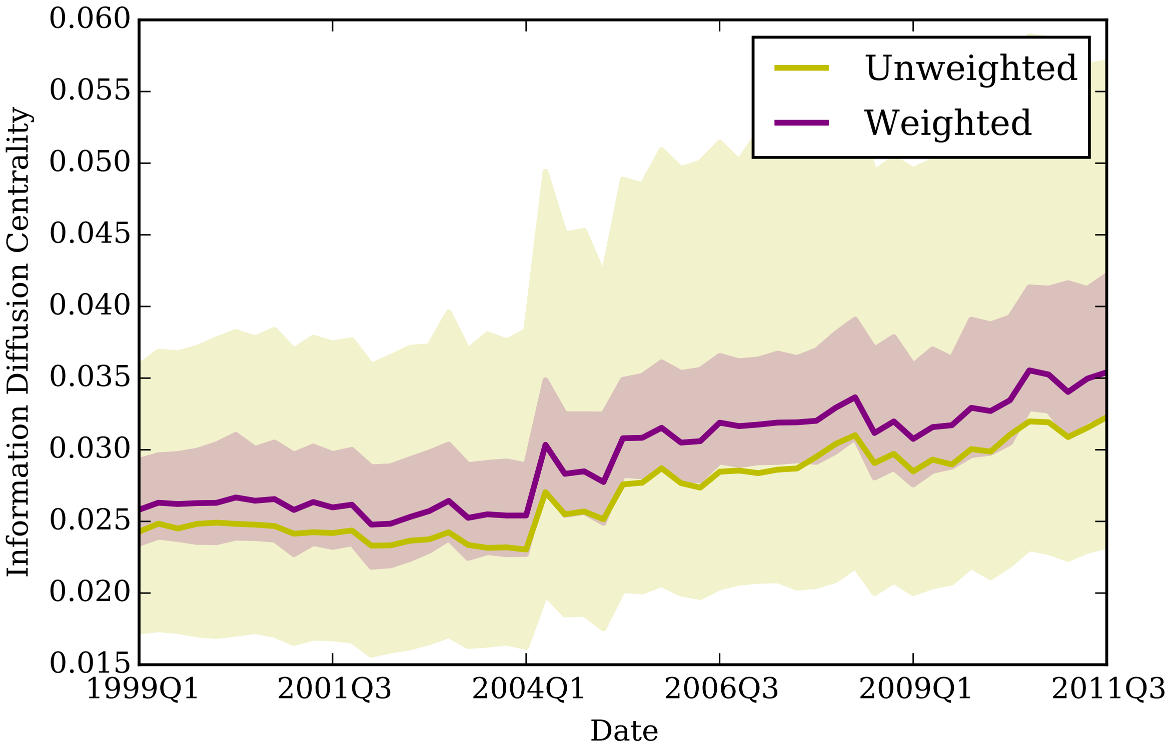
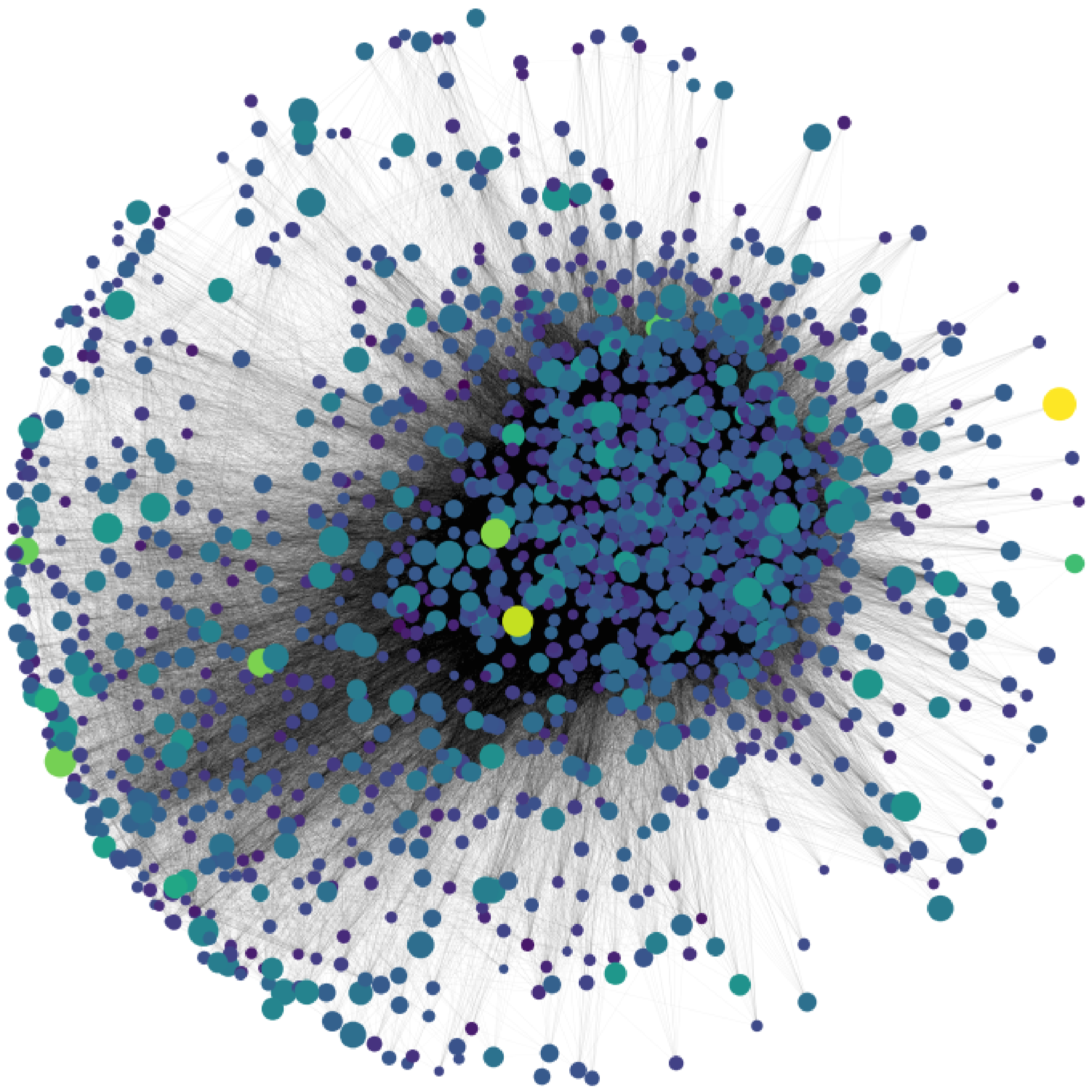


Figure 6: **Weighted Trading Correlation Network (2002Q2).** Nodes represent funds and edges represent trading correlations. Nodes are placed using a force-directed layout with edge weights corresponding to trading correlations. Lighter edges indicate lower trading correlations and darker edges indicate higher trading correlations. Node size is proportional to the number of edges, and node color is proportional to Information Diffusion Centrality. Lighter colored nodes (○) are "central" while darker colored nodes (●) are "peripheral."



8 Appendix

8.1 Information Sharing Matrix

Suppose that each agent $i = 1, \ldots, N$ is endowed with a private signal s_i, and each agent updates their signal by taking weighted averages of the differences between their current signal and the signals of their immediate neighbors such that the evolution of agent i's signal is given by:

$$\begin{aligned}\frac{ds_i}{dt} &= -\sum_j \frac{A_{ij}}{\sqrt{d_i}\sqrt{d_j}}(s_j - s_i), \\ &= -s_i + \sum_j \frac{A_{ij}}{\sqrt{d_i}\sqrt{d_j}} s_j.\end{aligned} \tag{8}$$

The second equality implies agent i "sends" his signal s_i to his neighbors and receives $\frac{A_{ij}}{\sqrt{d_i}\sqrt{d_j}} s_j$ from each of his neighbors. Sharing is bilateral because $\frac{A_{ij}}{\sqrt{d_i}\sqrt{d_j}}$ is symmetric.

In vector form,

$$\frac{d\boldsymbol{s}}{dt} = -(\boldsymbol{I} - \boldsymbol{D}^{-1/2}\boldsymbol{A}\boldsymbol{D}^{-1/2})\boldsymbol{s} = -\boldsymbol{\mathcal{L}}\boldsymbol{s}, \tag{9}$$

where $\boldsymbol{I}$ is the identity matrix, $\boldsymbol{D}$ is the diagonal matrix of each agents' degrees, $\boldsymbol{A}$ is the adjacency matrix, and $\boldsymbol{\mathcal{L}}$ is the symmetric graph Laplacian. The differential equation has the solution:

$$\boldsymbol{s}_t = e^{-\boldsymbol{\mathcal{L}}t}\boldsymbol{s}_0, \tag{10}$$

given an initial distribution of signals $\boldsymbol{s}_0$. I call $\boldsymbol{\mathcal{S}}_t = e^{-\boldsymbol{\mathcal{L}}t}$ the symmetric information "sharing" matrix. The non-symmetric version $\boldsymbol{S}_t$ is also called the "heat kernel" in the spectral graph theory literature. Chung and Yau (1998) introduces the symmetric heat kernel, and

Chung (2007) develops a centrality measure based on the non-symmetric version $e^{-\boldsymbol{L}t}$.

8.2 Eigenvector centralities

One measure which has not been popular in the finance literature, Personalized PageRank, is useful because it is a close proxy for Information Diffusion Centrality and can be used to derive the "eigenvector centralities" popular in the finance literature (Bonacich, Katz, and Eigenvector). Recall that Information Diffusion Centrality can be written as the exponential sum of symmetric random walks (4). Analogously, Personalized PageRank can be written as the *geometric* sum of random walks:

$$\boldsymbol{\varepsilon}(\alpha, \beta) = \beta \sum_{k=0}^{\infty} \alpha^k \boldsymbol{W}^k \boldsymbol{p}. \tag{11}$$

Or in vector form, $\boldsymbol{\varepsilon} = (\boldsymbol{I} - \alpha \boldsymbol{W})^{-1} \beta \boldsymbol{p}$. Here, $\boldsymbol{p}$ represents the "preference" vector or probability that a random walker will start from any given node which is analogous to the arrival of information above. α acts as a damping factor in the geometric series, and β is a positive constant.[21] The standard random walk matrix $\boldsymbol{W}$ can also be replaced with its symmetric counterpart to be consistent with the bilateral exchange of information.

PageRank, named after Google co-founder Larry Page, is designed to measure the importance of websites based on the notion that a website is important if other important websites link to it. As we have shown, it can also be expressed in terms of a "random (web) surfer" clicking links on pages such that important sites tend to be on the paths of many random walks (see Brin and Page, 1998; Page, Brin, Motwani, and Winograd, 1999). However it would be a mistake to call PageRank a measure of information diffusion as PageRank contains no description of the underlying information sharing process. Hence the canonical form of PageRank (with $\boldsymbol{p} = \boldsymbol{1}$) is not the random walk form shown above, but the recursive

[21] β is sometimes set to $(1 - \alpha)$ with $\alpha < 1$ such that the above random walk can be interpreted as a "teleporting random walk."

form:

$$\boldsymbol{\varepsilon} = \alpha \boldsymbol{W} \boldsymbol{\varepsilon} + \beta \mathbf{1}. \tag{12}$$

Bonacich, Katz, and Eigenvector centrality are also defined in recursive form, and can in fact be derived as special cases of PageRank. To see this, we only need to examine the typical element of the PageRank vector:

$$\varepsilon_i = \alpha \sum_j A_{ij} \frac{\varepsilon_j}{d_j} + \beta. \tag{13}$$

PageRank can be viewed as a scaled or random walk version of Bonacich centrality. When ε_j is not scaled by the degree d_j inside the summation (i.e. if $d_j = 1$ above) PageRank becomes Bonacich centrality. Katz centrality is a special case of Bonacich centrality when $\beta = 1$, and the standard Eigenvector centrality is a special case of Bonacich centrality with $\alpha = 1$ and $\beta = 0$.

Although these "eigenvector centralities" may be positively correlated with Information Diffusion Centrality, only PageRank has the random walk interpretation, and even then PageRank lacks the information sharing foundations of information diffusion.

8.3 ANCerno Data

Each observation in the ANcerno dataset consists of the execution of a trade along with the corresponding order-level identifiers, and trade-day liquidity measures.

Stocks are identified by `CUSIP` at the time of execution, ANcerno clients are identified by *clientcode*, their money managers by *managercode*, and the broker-dealer responsible for each execution by *brokercode*. Following Puckett and Yan (2011) I define a fund as a client-manager pair. Using the ANcerno provided cross-reference metadata, I identify 10,355 unique funds, which are composed of 1,358 clients and 1,007 money managers. A sample of the cleaned data is shown below.

Obs	Order	Date	Fund	Broker	Stock	Price	Volume	Side	Days
1	0	2002-01-23	1	5	75577	22.529	8100	-1	1
2	784	2002-01-23	5	7486	77472	38.591	3550	-1	1
3	1138	2002-01-23	5	7486	83597	35.960	3645	-1	1
4	1660	2002-01-29	5	6826	77662	50.960	5775	-1	1
5	521	2002-02-01	5	7486	79133	31.955	2600	1	3
6	521	2002-02-05	5	7486	79133	31.560	10700	1	3
7	1547	2002-02-19	5	13	12490	100.153	2600	-1	1
8	621	2002-03-07	5	161	43772	57.025	1765	-1	1
9	621	2002-03-07	5	26438	43772	57.025	1765	-1	1
10	1996	2002-03-12	5	50	10104	14.313	5800	1	1

Order identifies an order ticket, Date indicates when the trade was made, Fund is the composite *clientcode-brokercode* identifier, Broker is the broker identifier, Stock is the `PERMNO` corresponding to the provided `CUSIP`, Price and Volume are at the trade level, Side is 1 if the trade was a buy and -1 if the trade was a sell, and Days refers to the number of days over which the stock was traded as part of the Order.

The funds in my sample trade a total of 8,555 unique common stocks listed across the NYSE, NASDAQ, and AMEX exchanges. Overall, the funds make 141.85 million trades of 1.11 trillion shares valued at 34.26 trillion dollars USD. Puckett and Yan (2011) estimate that the ANcerno institutions account for approximately 8% of the total dollar value of CRSP trading volume between 1999 and 2005. More detailed summary statistics are provided in Table 8 below.

Insert Table 8

Puckett and Yan (2011) obtains a sample of 64 ANcerno client institution names which include pension plan sponsors such as CalPERS, the Commonwealth of Virginia, the YMCA retirement fund, money managers such as Massachusetts Financial Services, Putnam Investments, Lazard Asset Management, and several broker-dealers. Although ANcerno does not report the identities of their clients, the dataset does include the names of their clients' money managers and brokers. These money managers include investment management companies such as AllianceBernstein, BlackRock, and Wellington as well as investment banks such as Goldman Sachs, JP Morgan, and Deutsche Bank.[22] Broker-dealers include independent broker-dealers such as LPL Financial, Raymond James Financial Services, Ameriprise Financial Services as well as dealer banks such as Bank of America, Citigroup, and Credit Suisse.

[22]The database also includes the Carlyle Group, KKR & Co., and Blackstone Group LP which were formerly external managers investing on behalf of CalPERS, an ANcerno client. Source: "Calpers to Cut External Money Managers by Half." WSJ. Web. 21 July 2015.

Table 8: **ANCerno Data (1999Q1–2012Q1).** A fund is defined as a unique *clientcode–managercode* pair. The sample of stocks traded includes all common stock listed on the NYSE, AMEX, and NASDAQ exchanges which I am able to match to CRSP `PERMNO`s using CUSIPs and tickers. Brokers are identified by *brokercode*. Prices and volume are adjusted for stock splits using adjustment factors from the CRSP database.

Year	# of Funds	# Stocks Traded	# Brokers	# of Trades	Volume Traded	Dollar Value Traded
1999	1,992	4,921	651	1.98M	18.17B	904.88B
2000	1,901	4,814	641	3.15M	32.67B	1.72T
2001	1,996	4,557	693	3.59M	40.97B	1.32T
2002	2,105	4,348	747	3.57M	39.81B	1.02T
2003	2,068	4,315	729	3.77M	34.45B	859.37B
2004	1,950	4,256	671	4.42M	33.64B	985.22B
2005	1,612	4,057	687	4.39M	31.42B	1.02T
2006	1,404	3,984	665	7.32M	47.10B	1.55T
2007	1,296	3,928	638	8.64M	50.87B	1.84T
2008	1,285	3,771	668	7.32M	57.82B	1.71T
2009	1,304	3,712	648	6.49M	58.47B	1.28T
2010	1,220	3,350	652	5.19M	29.57B	870.19B
2011	1,223	3,387	659	4.80M	17.47B	561.19B
2012	238	2,657	186	506,477	1.45B	50.49B

Table 9: **Normalized Differences.** The propensity score matched sample of central and peripheral funds is based on the previous quarter's Degree, Eigenvector, number of trades, share volume traded, and principal-weighted (equal-weighted) average abnormal interim trading performance. I require that the difference in propensity scores does not exceed 0.1 basis points in absolute value to ensure covariate balance in the matched sample. Imbens and Wooldridge (2009) suggest a rule of thumb threshold for normalized differences of 0.25, above which linear regression methods tend to be affected by covariate balance.

	Peripheral	Central	Norm. Diff.
Degree	3.73	4.16	0.24
Eigenvector	0.0025	0.0027	0.021
Trades	382.6	872.7	0.079
Volume	2970581.3	6701747.0	0.052
PW	0.0033	0.0079	0.048
EW	0.0038	0.0090	0.059
N	5267	4982	.