Institutional Brokerage Networks: Facilitating Liquidity Provision*

Munhee Han   Sanghyun (Hugh) Kim   Vikram K. Nanda

This Draft: December 13, 2019

Abstract

We argue that institutional brokerage networks facilitate liquidity provision and mitigate price impact of large non-information motivated trades. We use commission payments to map trading networks of mutual funds and brokers as affiliation networks. We find that central funds outperform peripheral funds, especially as measured by return gap. The outperformance is more pronounced when trading is primarily liquidity driven to accommodate large redemptions. This centrality premium is further strengthened by brokers' incentives to generate greater revenues and by repeated interactions between brokers and funds. By merging daily transactions with quarterly holdings, we confirm that the centrality premium is indeed driven by reduced trading costs, rather than higher interim (intra-quarter) trading performance or profitable information flows from the brokers.

JEL classification: G14, G23, G24

Keywords: Institutional brokerage networks, mutual funds, return gap, trading costs, liquidity provision

*Han (munhee.han@utdallas.edu), Kim (hugh.kim@utdallas.edu), and Nanda (vikram.nanda@utdallas.edu) are at the University of Texas at Dallas, 800 West Campbell Road, Richardson, TX 75080. We thank Steven Xiao, Kelsey Wei, Paul Irvine, Stacey Jacobsen (discussant), Junghoon Lee, Charles Trzcinka, Talis Putnins (discussant), Johan Sulaeman (discussant), Kumar Venkataraman, Andy Puckett, Terry Hendershott (discussant), Duane Seppi, Xi Dong (discussant), Abhiroop Mukherjee (discussant), Francesco Franzoni, Kenneth Ahern (discussant), seminar participants at the University of Texas at Dallas, Texas Christian University, Florida International University, University of Texas at San Antonio, and Emory University, and conference participants at the Conference on Financial Economics and Accounting (CFEA) 2018, Australasian Finance and Banking Conference 2018, Society for Financial Studies (SFS) Cavalcade Asia-Pacific 2018, Midwest Finance Association (MFA) Annual Meeting 2019, NYU Stern Microstructure Meeting 2019, FMA European Conference 2019, FMA Asia-Pacific Conference 2019, China International Conference in Finance (CICF) 2019, ESMT Asset Management Conference 2019, and AIM Investment Conference 2019 for helpful comments.

Institutional Brokerage Networks: Facilitating Liquidity Provision

Abstract

We argue that institutional brokerage networks facilitate liquidity provision and mitigate price impact of large non-information motivated trades. We use commission payments to map trading networks of mutual funds and brokers as affiliation networks. We find that central funds outperform peripheral funds, especially as measured by return gap. The outperformance is more pronounced when trading is primarily liquidity driven to accommodate large redemptions. This centrality premium is further strengthened by brokers' incentives to generate greater revenues and by repeated interactions between brokers and funds. By merging daily transactions with quarterly holdings, we confirm that the centrality premium is indeed driven by reduced trading costs, rather than higher interim (intra-quarter) trading performance or profitable information flows from the brokers.

JEL classification: G14, G23, G24

Keywords: Institutional brokerage networks, mutual funds, return gap, trading costs, liquidity provision

1 Introduction

Brokers play a vital role in institutional trading in equity markets. When executing large client orders, brokers can mitigate price impact by actively searching for potential counterparties across various trading venues. Despite the shift toward electronic trading (trading algorithms, dark pools, and electronic market makers) over the last 10-15 years, non-electronic "high-touch" trading (single stock trades routed to a sales trader and communicated typically via telephone) still represents the largest execution channel even in the U.S. equity markets. (Greenwich Associates (2017)). Thus, institutional trading tends to be broker-intermediated, with its efficacy closely tied to the trading networks of institutional investors and their brokers. In this paper, we argue that institutional brokerage networks facilitate liquidity provision and mitigate price impact for non-information driven trades.

Using brokerage commission payments, we map trading networks of mutual funds and their brokers as affiliation networks in which mutual funds are connected to other mutual funds through their overlapping brokerage relationships. In these networks, mutual funds that trade through brokers that are also heavily used by other funds will tend to be central. A key finding of the paper is that central funds in institutional brokerage networks outperform peripheral funds, especially as measured by their trading performance. In order to shed light on the specific mechanisms driving the positive relation between mutual funds' brokerage network centrality and their trading performance i.e., *centrality premium*, we propose a liquidity provision hypothesis.

Our notion is that centrality in brokerage networks is especially valuable when mutual funds are forced to trade for liquidity reasons. As is well-recognized, open-end mutual funds incur substantial trading costs due to the adverse market impact of their trades when they liquidate holdings in response to investor redemptions (e.g., Edelen (1999)).[1] Thus, as in Admati and Pfleiderer (1991), liquidity traders that are transacting large quantities for non-informational reasons have an incentive to make their trading intentions known (i.e., engage in "sunshine trading") to distinguish themselves from informed traders and

[1] In market microstructure models (such as Glosten and Milgrom (1985), Kyle (1985)), risk-neutral market makers are unable to identify trading motives. In these models, market makers set market prices and expect to lose to informed traders, while breaking even with gains from uninformed, liquidity traders.

attract more traders to provide liquidity. In practice, however, the announcement of trading intentions could make a fund vulnerable to "predatory trading".[2] Nevertheless, while large liquidity traders may be unable to signal their trading motives directly to market participants, our view is that they might achieve the desired outcome by relying on their brokerage networks and upstairs block trading.[3]

We contend that institutions trading for liquidity reasons may be able to credibly convey their trading motives to brokers with whom they have well-established relationships. The notion is that misrepresentation of trading motives could be costly to institutions in terms of a loss of reputation and damage to the broker-institution relationship. Central funds, connected to a larger network of brokers and funds have more at risk in terms of a potential loss of reputation and relationships and, hence, may have greater credibility in conveying their reasons for trading. The fund's brokers, in turn, could certify their clients' liquidity motives and execute trades at better prices (Seppi (1990)).[4] Additionally, upstairs brokers can expand the available liquidity pool using information about their clients' latent trading interests and reaching out to a wider set of potential counterparties to lower trading costs (Grossman (1992)).[5] Thus, even though all funds may have similar access to the available pool of expressed liquidity, for instance, through an electronic limit order book in the downstairs market, central funds will be better positioned to tap into larger pools of unexpressed liquidity through their brokers, especially when submitting large blocks of

[2] The concern is that strategic traders that become aware of, say, the liquidation of a large holding could engage in "predatory trading", an argument advanced in Brunnermeier and Pedersen (2005). The notion is that strategic traders could sell the asset prior to or simultaneously with the liquidating trader; these traders could subsequently reverse their trades to profit from the depressed asset price at the expense of the liquidating trader.

[3] "'Upstairs' Trading Draws More Big Investors," by Bradley Hope, the *Wall Street Journal*, December 8, 2013. The article reports that even in today's markets with dark pools quickly replacing off-exchange trading, large institutional investors still choose to trade in blocks upstairs where a block broker facilitates the trading process by locating counterparties to the trade. The article quotes a trader as stating that "It's like trying to fill up your gas tank, but you have to go to 15 gas stations. By the time you get to the 15th one, they've increased the price because they've heard you were coming. Wouldn't someone rather go to two or three stations and fill up the tank in blocks?"

[4] An upstairs market is an off-exchange market where a block broker facilitates the trading process by locating counterparties to the trade. Madhavan and Cheng (1997), Smith, Turnbull, and White (2001), and Booth et al. (2002) present evidence consistent with the Seppi (1990) hypothesis that upstairs market makers effectively screen out information-motivated orders and execute large liquidity-motivated orders at a lower cost than the downstairs market in the New York Stock Exchange (NYSE), the Toronto Stock Exchange (TSE), and the Helsinki Stock Exchange (HSE), respectively.

[5] Bessembinder and Venkataraman (2004) present direct evidence in support of the Grossman (1992) prediction that upstairs brokers lower execution costs by tapping into unexpressed liquidity. The authors find that execution costs for upstairs trades on the Paris Bourse are much lower than would be expected if the trades were executed against the expressed (displayed and hidden) liquidity in the downstairs limit order book.

liquidity-motivated orders.

However, liquidity traders may have their own concerns about revealing their trading intentions to brokers. In the context of outflow-driven fire sales, Barbon et al. (2019) document that institutional brokers appear to foster predatory trading by leaking their clients' order flow information about impending fire sales to other important clients. These clients then sell the stocks being liquidated – only to buy them back later at lower prices. Our view is that while brokers may occasionally disclose client trades, they are unlikely to do so against their important clients, if it puts their trading relationships in jeopardy (e.g., Carlin, Lobo, and Viswanathan (2007)). Hence, institutional investors are likely to share trading intentions only if brokerage firms have valuable relationship and reputation capital: capital that could be lost if brokers do not act in their clients' interests. The brokers used by central funds are apt to have greater relationship/reputation capital as indicated, for instance, by their well-established relationships to many other funds (and greater costs to being seen as untrustworthy). Hence, central funds are likely to benefit from lower costs for their liquidity motivated trades.[6]

To test our liquidity provision hypothesis, we exploit a unique dataset on brokerage commissions for a comprehensive sample of mutual funds from Form N-SAR semi-annual reports filed with the Securities and Exchange Commission (SEC). Using techniques from graph theory, we map the connections between mutual funds and their brokers as affiliation networks represented by weighted bi-partite graphs.[7] The weight of the bi-partite graph represents the strength of connection between a given fund-broker pair and is calculated as (normalized) brokerage commissions paid to the given broker. Further, to measure mutual funds' brokerage network centrality, we reduce this bi-partite graph of funds and brokers into a mono-partite graph in which fund-to-fund links are operationalized through their overlapping broker ties (see

[6] Our paper is complementary to Barbon et al. (2019) in the sense of Carlin, Lobo, and Viswanathan (2007), who present a multi-period model of trading based on liquidity needs. In their model, traders cooperate most of the time through repeated interaction, providing liquidity to one another. However, "episodically" this cooperation breaks down when the stakes are high enough, leading to predatory trading.

[7] Affiliation networks can be represented by bi-partite graphs in which there are two types of nodes (say, brokers and mutual funds) such that the nodes of one type are only connected to nodes of the other type. In our case, a mutual fund is directly connected only to its brokers, while brokers are directly connected only to funds. Hence, any pair of mutual funds can be connected only indirectly through their overlapping brokerage connections. The connection between two funds is stronger if the extent to which their brokerage connections overlap is greater.

e.g. Faust (1997)). We then use degree centrality and eigenvector centrality to quantify the importance of a given fund's position in the network.

Mutual funds that trade through many brokers that many other funds also trade through tend to be central in the network. Goldstein et al. (2009) note that smaller institutions concentrate their order flows with a small number of brokers in order to become their important clients, whereas large institutions can more easily obtain premium status with several brokers. Consistent with this observation, we find that funds that are large or belong to large fund families tend to be more central in the network, as they can afford to trade through a larger network of brokers that are themselves central in the network. We also find that mutual funds' brokerage network centrality is highly persistent, reflecting the persistence in the underlying brokerage relationships.

We begin our empirical analysis by showing that mutual funds' brokerage network centrality positively predicts their trading performance. Since we do not directly observe trading activities of mutual funds, we use the return gap as our measure of trading performance. The return gap is calculated as the difference between the reported fund return (gross of expenses) and the return on a hypothetical portfolio that invests in the previously disclosed fund holdings (Grinblatt and Titman (1989), Kacperczyk, Sialm, and Zheng (2008)).[8] We find that mutual funds' brokerage network centrality positively predicts future return gap. Under our liquidity provision hypothesis, the economic magnitude of the centrality premium is quite meaningful. One standard deviation increase in centrality is associated with an increase in return gap by about 13 basis points per year.[9] In addition, we show in our later analysis that the centrality premium more than doubles when funds are forced to trade to accommodate large investor redemptions.

One might concern that there might exist an unobservable (to the econometrician) factor that is

[8] The return gap was originally proposed by Grinblatt and Titman (1989) as a measure of total transactions costs for mutual funds. Grinblatt and Titman (1989), however, point out that the return gap may be affected by interim trades within a quarter (Puckett and Yan (2011)) and possibly by window-dressing activities. Kacperczyk, Sialm, and Zheng (2008) further note that skilled fund managers can use their informational advantage to time the trades of individual stocks optimally and show that the past return gap helps predict fund performance.

[9] To put the numbers in perspective, Edelen (1999) estimates that liquidity-driven mutual fund trades lead to losses against informed traders on the order of approximately 140 basis points per year. In terms of an alternative hypothesis, Puckett and Yan (2011) estimate that interim trading performance contributes between 20 and 26 basis points to the annual abnormal returns of mutual funds.

correlated with both brokerage network centrality and return gap. For example, Puckett and Yan (2011) find that institutional investors earn significant abnormal returns on their interim (intra-quarter) trades and that interim trading performance is persistent. In addition, Anand et al. (2012) show that trading costs are closely linked to trading desks' execution skills over and above selecting better brokers. In order to mitigate those potential confounding factors, we use fund or family fixed-effects to control for unobserved heterogeneity. Our results remain quantitatively and qualitatively similar after controlling for fund fixed-effects, although the relation between centrality and return gap becomes somewhat weaker when we control for family fixed-effects. The latter finding makes sense because, since funds belonging to the same fund family tend to have similar brokerage connections, there may be little within-family cross-sectional variation in brokerage network centrality.

Next we turn to testing key predictions of our liquidity provision hypothesis. The primary prediction of our hypothesis is that the centrality premium should be more pronounced when funds' trading activities are largely driven by liquidity motives and funds can credibly signal this to their brokers. We use large outflow events to identify such periods of liquidity-motivated trading. When a mutual fund is experiencing severe redemptions, the fund is forced to liquidate a large fraction of its holdings in several stocks and their selling is, to a large extent, uninformed (e.g., Coval and Stafford (2007)). In addition, such forced liquidations are likely to send a particularly strong signal to brokers that the fund's sell orders are liquidity driven, rather than information motivated, thus helping the brokers communicate more credibly with other institutional clients to take the other end of the trades. Consistent with this prediction, we find that the centrality premium is more pronounced when funds are forced to liquidate their holdings to accommodate large outflows.

We conduct sub-sample analysis by splitting our sample based on the fund family size as measured by the number of funds within the fund family in order to address concerns that our results could be driven by within-family cross-subsidization (e.g., Bhattacharya, Lee, and Pool (2013)).[10] On the contrary to this

[10] A potential concern is that the above results could be consistent with cross-subsidization within a fund family: when a fund is suffering severe redemptions, another fund in the same family could step in to provide liquidity. For instance, Bhattacharya, Lee, and Pool (2013) show that affiliated funds of mutual funds that invest only in other funds within the family provide an insurance pool against temporary liquidity shocks to other funds in the family.

alternative family network hypothesis, however, we find that the conditional centrality premium is *greater* among funds that belong to *smaller* fund families. This makes sense because large fund families tend to be capable of internalizing temporary liquidity shocks to their member funds and, as a result, funds belonging to large fund families do not have to rely on their brokerage networks to mitigate liquidity-driven trading costs and externalities arising from forced liquidations (fire sales).

Our liquidity provision hypothesis requires an active role on the part of brokers, such as in discerning their clients' trading motives and communicating with other institutional clients. As made clear in Carlin, Lobo, and Viswanathan (2007), whether brokers facilitate liquidity provision or foster predatory trading is likely to hinge on the incentives they face and the strength of repeated interactions with their clients. To the extent that brokers seek to maximize the expected value of future revenues, we expect funds that are central and have greater revenue potential to be more likely to benefit from broker-facilitated liquidity provision. Consistent with this prediction, we find that the conditional centrality premium is more pronounced for funds that are more valuable to brokers.

Similarly, our hypothesis relies on the repeated nature of interactions between institutional clients and their brokers. Institutional investors need a reputation for being truthful in order to credibly signal to brokers when they are trading for liquidity reasons. The brokers, in turn, must develop their reputation/relationship capital for being discreet when handling their clients' orders. Thus, the signalling and certification of non-information motivated trading is likely to be most effective if funds have well-established trading relationships with their brokers. Consistent with this prediction, we find that the conditional centrality premium is larger for the clients that have stronger existing trading relationships with their brokers.

We argue that the centrality premium is driven by lower trading costs. The return gap, however, reflects not only trading costs such as bid-ask spread, price impact, and brokerage commissions, but also a host of other factors including interim (intra-quarter) trading performance (Kacperczyk, Sialm, and Zheng (2008), Puckett and Yan (2011)). Central funds might be better positioned to acquire short-

One might concern that central funds tend to belong to large fund families and large fund families are likely better equipped to provide such within-family cross-subsidization.

lived information through their strong brokerage connections and trade on it, thus contributing to interim trading performance. For instance, Irvine, Lipson, and Puckett (2007) document a positive relation between brokerage connections and early access to sell-side research ("tipping") (see also Goldstein et al. (2009), Xie (2014)). In addition, Di Maggio et al. (2019) find that central brokers can extrapolate large informed trades from order flows and selectively leak this information to their more important clients.

In order to corroborate our liquidity provision hypothesis and rule out alternative explanations for the centrality premium, we employ direct measures of trading costs and interim returns by restricting our sample to fund-quarters in which we can reliably merge ANcerno (also known as Abel Noser) daily transactions data with Thomson s12 quarterly holdings data at the *fund* level. For each trade, we compute dollar implementation shortfall as the difference between the execution price and the benchmark price prevailing at the time when the broker receives the order (e.g., Anand et al. (2012)), multiplied by the number of shares executed. Trading costs based on implementation shortfall capture *implicit* trading costs such as bid-ask spread and price impact. We aggregate trade-level implicit trading costs across all trades to the fund level. ANcerno also reports for each trade commissions and taxes plus fees paid to the executing brokers, which constitute *explicit* trading costs. We aggregate trade-level explicit trading costs in a similar way across all trades to the fund level.

Consistent with our liquidity provision hypothesis, we find that brokerage network centrality negatively predicts trading costs, especially implicit trading costs. One standard deviation increase in centrality is associated with a decrease in implicit and explicit trading costs by about 7 and 2 basis points per year, respectively. The finding that central funds tend to incur slightly lower explicit trading costs casts doubt on the alternative information flow hypothesis. For instance, if anything, central funds are *less* likely to pay soft dollars that are typically associated with profitable analyst recommendations (e.g., Irvine, Lipson, and Puckett (2007). Goldstein et al. (2009)). Lending further support to our liquidity provision hypothesis, we find that the negative relation between centrality and trading costs becomes stronger when funds are forced to trade to accommodate large investor redemptions. Conditional on funds receiving large outflows, one standard deviation increase in centrality is associated with a decrease in implicit and explicit trading costs by 17 and 8 basis points per year, respectively.

Last, we compute mutual fund returns from interim trading (Puckett and Yan (2011)). Specifically, for each trade executed between adjacent portfolio report dates, we track the performance of the trade from the execution date (using the execution price) to the following portfolio report date. We then aggregate dollar returns from each interim trade to the fund level. Even though we do find a significant positive relation between brokerage network centrality and return gap in our much restricted sample, we find an insignificant relation between brokerage network centrality and interim returns.[11] Combined with a significant negative relation between centrality and trading costs, an insignificant relation between centrality and interim trading performance suggests that the return gap premium associated with brokerage network centrality (the centrality premium) is indeed driven by lower trading costs, rather than higher interim performance.

The remainder of this paper is organized as follows. In the next section, we discuss our paper in the context of related literature. Section 3 introduces our data and describes how we construct our samples and variables. We report our results in Sections 4 and 5. Section 6 concludes.

2 Related Literature

Our paper uncovers novel network effects in equity markets by documenting the return gap premium associated with mutual funds' brokerage network centrality (simply the centrality premium). In addition, we show that the centrality premium is driven by reduced trading costs, especially when funds are forced to liquidate holdings to accommodate large investor redemptions. We contribute to a growing literature on broker-dealer networks in financial markets by shedding light on the unique role of institutional brokers in facilitating liquidity provision. While there is a large literature on dealer networks in over-the-counter (OTC) markets (see Section V. D of Bessembinder, Spatt, and Venkataraman (Forthcoming) for a com-

[11] This is not to say that the return gap itself is not driven by interim trading performance. In fact, the return gap is positively correlated with interim returns (correlation coefficient = 0.34). Instead, our results suggest that the centrality premium is driven by the trading cost component of the return gap, rather than the interim performance component of the return gap.

prehensive survey), studies on broker networks in the equity markets have been relatively scant and our paper attempts to fill this gap. In addition, it is *a priori* unclear and thus an empirical question whether the network effects found in decentralized OTC markets would carry over to centralized equity markets.[12]

In a recent paper, Di Maggio et al. (2019) shows that central brokers can extrapolate large informed trades from order flows and selectively leak this information to their more important clients, thereby facilitating "back-running" as described by Yang and Zhu (2019). Given such rent-extraction behavior, it is thus unclear whether central brokers can obtain "best execution" for their institutional clients. Our paper shows that central funds that trade through a *larger* network of central brokers can earn the centrality premium by effectively leveraging their strong brokerage connections to mitigate trading costs associated with adverse selection.[13] Our paper is consistent with a related literature on client-dealer networks in the OTC corporate bond market. Hendershott et al. (2017) shows that many insurers use only one dealer, but execution costs decrease as a non-monotone function of the network size until it reaches 20 dealers, consistent with insurers trading off the benefits of relationship trading against dealer competition.

Our paper is related to, but differs from recent studies that document evidence of information flows or leakages from some clients to others through brokers. Chung and Kang (2016) shows strong return comovement among hedge funds sharing the same prime broker and argue that the prime broker provides profitable information to its hedge fund clients. As potential sources of such profitable information, Kumar et al. (Forthcoming) points to privileged information on corporate borrowers from the affiliated banking division of an investment bank with prime brokerage business and Di Maggio et al. (2019) hints at client order flow information about large informed trades by hedge funds or activist investors right before 13D

[12] Our main finding that the centrality premium is driven by reduced trading costs is largely consistent with findings in dealer networks. In inter-dealer networks in the OTC municipal bond market, Li and Schürhoff (2019) find that dealers that are more central in the networks have better access to clients and more information about which securities are available and who wants to buy or sell, which results in shorter "intermediation chains," i.e., that fewer dealers are involved before a bond is transferred to another customer. Also consistent are our results that the conditional centrality premium is further strengthened by brokers' incentives to generate greater revenues and by repeated interactions between brokers and funds. For instance, Di Maggio, Kermani, and Song (2017) show that prior trading relationships are especially valuable in turbulent times in the OTC corporate bond market.

[13] As noted in Bonacich (1991), centrality in affiliation networks involves the "duality" of groups and individuals. A broker gets its central position from the trading patterns of its client funds and a central fund should be one that trades through a larger network of central brokers.

filings. Our paper, however, differs substantially from these papers in that our focus is on information flows regarding large liquidity-motivated trades, rather than private information about company fundamentals. Our paper also differs from the papers that shows how institutional investors can gain informational advantage through their brokerage connections. Examples of such information channels include early access to sell-side research or "tipping" (Irvine, Lipson, and Puckett (2007)) and invitations to broker-hosted investor conferences (Green et al. (2014)).

Our paper is closely related to Barbon et al. (2019) which documents that institutional brokers can foster predatory trading by leaking their clients' order flow information about impending fire sales to other important clients, such as prime brokerage hedge fund clients. The clients then sell the stocks being liquidated by distressed funds only to buy them back later at much lower prices. Our paper can be seen as complementary to Barbon et al. (2019). While we contend that brokers can mitigate trading costs associated with adverse selection, this is contingent on brokerage firms valuing client relationships and being deterred by the potential reputation costs of being seen as exploiting client information. Our results are broadly consistent with those in client-dealer networks. In this regard, Carlin, Lobo, and Viswanathan (2007) presents a multi-period model of trading in which both cooperation and conflict can exist. In their model, traders cooperate most of the time through repeated interaction, providing liquidity to one another. However, "episodically" this cooperation breaks down when the stakes are high enough, leading to predatory trading.[14]

3 Data and variable construction

Sections 3.1 and 3.2 provide the details on our sample construction merging four different databases: (i) CRSP Mutual Fund Database, (ii) Thomson-Reuter Ownership Database (s12), (iii) Form N–SAR

[14] Some investment banks generate a substantial amount of fee revenues from hedge funds that use their prime brokerage services, such as securities lending, margin financing, and risk management. Kumar et al. (Forthcoming) finds strong evidence that investment banks sometimes leak privileged information about their corporate borrowers to their prime brokerage hedge fund clients who subsequently trade on and profit from it, whereas Griffin, Shu, and Topaloglu (2012) finds little evidence of such information-based trading by the average brokerage house client of investment banks.

filings, and (iv) ANcerno (also known as Abel Noser) institutional daily transactions database. Section 3.3 describes our data on brokerage commissions and explains how we construct fund-level control variables. Section 3.4 explains how we construct institutional brokerage networks and centrality measures, and discusses the characteristics of the network.

3.1 CRSP–Thomson–NSAR merged sample

Our primary data comes from the Form N–SAR reports, which we combine with other data sets. We obtain data on mutual fund monthly returns, total net assets (TNA), and fund expenses from the Center for Research in Security Prices (CRSP) Survivor-Bias-Free Mutual Fund Database. The returns are net of fees, expenses, and brokerage commissions, but before any front-end or back-end loads. The stock holdings of mutual funds are from Thomson-Reuter Ownership Database (Thomson s12). We use the MFLINKS files available through Wharton Research Data Services (WRDS) to merge CRSP and Thomson data sets. For funds with multiple share classes in CRSP, we aggregate share-class-level variables at the fund-level by computing the sum of total net assets and the value-weighted average of returns and expenses.

Under the Investment Company Act of 1940, all registered investment companies are required to file Form N–SAR with the Securities and Exchange Commission (SEC) on a semi-annual basis. N–SAR reports are filed at the investment company level.[15] N–SAR filings disclose information about fund operations and financials under 133 numbered items with alphabetized sub-items. Many items are reported at the fund level, but some of the items such as brokerage commissions are aggregated and reported at the investment company level. We extract all N-SAR reports filed between 1994 and 2016 available through the SEC's Electronic Data Gathering, Analysis, and Retrieval (EDGAR) system.

Since our focus is on U.S. domestic equity funds, we exclude N-SAR funds that are not equity-oriented (Item 66.A), international funds (Item 68.B), and the funds with percentage of TNA invested in common stocks (Item 74.F divided by Item 74.T) below 80% or above 105%. We also exclude N-SAR

[15] An investment company (or a registrant) can consist of multiple funds or be part of a fund family. For instance, according to our N–SAR data, Fidelity reported its 466 mutual funds with about $1.5 trillion assets under management using 82 separate N–SAR filings during the first half of 2016.

reports where aggregate brokerage commissions paid (Item 21) are reported as zero or missing.[16] From the CRSP–Thomson merged data set, we eliminate international, municipal, bonds and preferred, and metals funds using the investment objective code from Thomson (*ioc*) and screen for U.S. domestic equity funds using the investment objective code from CRSP (*crsp_obj_cd*). We also exclude all observations where the fund's TNA does not exceed $5 million or the number of stocks held does not exceed 10.

After the above data screens, we automatically match N–SAR fund names with CRSP fund names after removing share-class identifiers using the generalized Levenshtein (1966) edit distance while exploiting the typical structure of CRSP fund names (*INVESTMENT COMPANY NAME: FUND NAME; SHARE CLASS*). In the automated name matching procedure, we require that the monthly average net assets (TNA) during the reporting period (Item 75.B) and the corresponding TNA value computed from CRSP be within the 5% range from each other. Finally, we augment the automated name matching procedure by manually verifying the matches based on fund names reported in CRSP and NSAR data sets. The total number and aggregate TNA of our CRSP–Thomson–NSAR matched sample funds are reported in Panel A of Table A1 in the Appendix.

3.2 CRSP–Thomson–NSAR–ANcerno merged sample

For some of our later analysis, we use daily transactions obtained from ANcerno (also known as Abel Noser) institutional daily transactions database to directly estimate mutual fund trading costs and interim (intra-quarter) returns for the period from January 1999 to March 2011. Using a matching procedure similar to the one employed by Agarwal, Tang, and Yang (2012) and Busse et al. (2019), we merge ANcerno daily transactions data with Thomson s12 mutual fund quarterly holdings data at the *fund* level. Specifically, for each reporting period between two adjacent portfolio report dates (as identified by *rdate*) for each fund F (as identified by *fundno* and *wficn*) in the Thomson s12 and MFLINKs data sets, we compute split-adjusted changes in holdings for fund F and for each client manager M (as identified by *clientmgrcode*) in

[16] Reuter (2006) reports that in his sample, approximately 82% of the N-SAR filings that report paying no brokerage commissions are from investment companies that consist solely of bond funds, which do not pay explicit brokerage commissions on their transactions.

the ANcerno database in each stock during the reporting period. Then we compare split-adjusted changes in holdings by F and M for each stock and, if they match *exactly* with each other, then we call this stock a *matched* stock between F and M for that reporting period.

Next, we compute (i) the ratio of the number of matched stocks to the number of stocks traded by Thomson s12 fund F (denoted by *MATCH_S12_PCT*) and (ii) the ratio of the number of matched stocks to the number of stocks traded by ANcerno client manager M (denoted by *MATCH_AN_PCT*). Following Agarwal, Tang, and Yang (2012), we call a reporting period for fund F a *matched* period with client manager M (i) if there are at least 5 matched stocks, (ii) *MATCH_S12_PCT* is at least 10%, and (iii) *MATCH_AN_PCT* is at least 10%. If there are multiple matches for a reporting period for fund F, we choose the best matched period with client manager M by *MATCH_S12_PCT* and, if there is a tie, then by *MATCH_AN_PCT*. Last, we augment the above matching procedure based on changes in holdings, by manually verifying the matches using fund names from CRSP and Thomson and a list of client manager names released by ANcerno in 2011. We emphasize that, critical to our analysis at the fund level, our matching procedures based on changes in holdings and names are performed at the mutual fund level (as identified by *clientmgrcode* in ANcerno), rather than at the institution or fund family level (as identified by *managercode* in ANcerno). We also note that a single Thomson fund F can be matched with multiple ANcerno client manager Ms for different reporting periods because, unlike Wharton Financial Institution Center Number (*wficn*), ANcerno's *clientmgrcode* does not track funds consistently over time. The total number and aggregate TNA of our CRSP–Thomson–NSAR–ANcerno matched sample funds are reported in Panel B of Table A1 in the Appendix.

3.3 Brokerage commissions and other fund-level variables

Of particular interest to our study from N–SAR reports are brokerage commissions paid to the ten brokers that received the largest amount from the investment company during the reporting period and the identities of those brokers (Item 20). Table 1 provides an example of brokerage commission payments along with some descriptive statistics.

[Insert Table 1]

We recognize that N–SAR filings do not report all brokerage firms to which the investment company paid commissions and, as a result, we are likely to miss some of the less important brokerage connections. For example, Panel A of Table 1 reports brokerage commissions that T. Rowe Price Blue Chip Growth Fund paid to its top ten brokers and the aggregate commissions paid to all brokers during the first half of 2016. As is typically the case, the sum of brokerage commissions do not add up to the aggregate commissions, suggesting that the fund employed more than ten brokers. In general, as shown in Panel B of Table 1, brokerage commissions are highly concentrated with a few important brokers, but the top ten brokers reported in N–SAR filings on average account for only 72.45% (or 71.62% at the median) of the aggregate brokerage commissions that the investment company paid to *all* brokers. Nevertheless, data trunctation issues are unlikely to bias our results because centrality calculated in the reduced network is highly correlated with full-network centrality (Ozsoylev et al. (2014)).[17] Furthermore, we present evidence in Section 4.1 that, if anything, the incompleteness of our networks is likely to bias *against* our findings.

Next, we describe how we construct fund-level control variables. We obtain mutual fund returns, total net assets (TNA), and expenses from CRSP. We sum the fund TNAs (Item 74.T) across all funds reported in N–SAR filings that belong to the same fund family using the fund family code reported in N–SAR (Item 19.C) to compute the fund family TNA on a semi-annual basis. We estimate the fund's quarterly dollar volume in two steps: we first compute the ratio of the sum of purchases (Item 71.A) and sales (Item 71.B) to monthly average value of portfolio (Item 71.C). We then multiply this ratio by the fund's most recent quarterly TNA and divide it by two (four) if the reporting period is for 6 (12) months. We take the fund's turnover from N–SAR (Item 71.D), which equals the min of purchases (Item 71.A) and sales (Item 71.B) divided by monthly average value of portfolio (Item 71.C), and annualize it by multiplying the figures by two if the reporting period is for 6 months. For each fund-quarter, we compute the value-weighted averages of log of market capitalization, book-to-market ratio, and 12-month

[17] For example, in simulations Ozsoylev et al. (2014) show that even when a reduced network represents only 10% of the links in the full network, the correlation between true centrality and centrality calculated in the reduced network is about 0.5.

cumulative returns (skipping the most recent month) of the stock holdings from the most recent holdings report.

Finally, following the literature (e.g., Coval and Stafford (2007)), we estimate monthly net flows for each fund share class i during month t as follows:

$$FLOW_{i,t} = TNA_{i,t} - TNA_{i,t-1} \times (1 + R_{i,t}) \tag{1}$$

where $FLOW_{i,t}$ is the dollar value of fund flow (net new issues and redemptions), $TNA_{i,t}$ is the total net asset, and $R_{i,t}$ is the monthly return. To compute the monthly fund flow for the fund, we sum monthly fund flows for all share classes belonging to the same fund as identified by MFLINKS. Monthly fund flows are summed over the quarter to calculate the quarterly fund flow. For the percentage figures, we divide the dollar value of fund flows by the beginning-of-period TNA. The summary statistics are reported in Table 2.

[Insert Table 2]

3.4 Institutional brokerage networks

Using brokerage commission payments, we map trading networks of mutual funds and their brokers as affiliation networks represented by weighted bi-partite graphs. In a graph, agents can be represented by nodes and connections (ties) between agents by edges. In a bi-partite graph, nodes can be partitioned into two types and nodes of one type can only be connected to the nodes of a different type, not with the ones of the same type. Like any graph, a bi-partite graph can be represented by an adjacency matrix, denoted G, where rows index investment companies (simply referred to as funds) and columns index brokers. Each element $g_{i,k}$ of G represents the strength of connection between fund i and broker k and is defined as brokerage commissions paid to broker k, normalized by the length of brokerage commissions paid to all brokers. That is, column vector g_i is equal to $\frac{c_i}{\|c_i\|}$ where c_i is fund i's brokerage commission vector. If broker k does not appear as one of the top ten brokers, then $g_{i,k}$ is assumed zero. We illustrate how we construct institutional brokerage networks and calculate brokerage network centrality step-by-step using a simple example network consisting of ten funds and four brokers in Figure 1.

[Insert Figure 1]

To measure a mutual fund's connections to all the other mutual funds through their overlapping brokerage connections, we reduce the bi-partite graph of mutual funds and brokers into a mono-partite graph of mutual funds only by defining its adjacency matrix A as $A = GG^T$.[18] By construction, the strength of connection between any pair of funds is simply measured as the cosine similarity between brokerage commission vectors for that pair, i.e., $a_{i,j} = g_i \cdot g_j = \frac{c_i \cdot c_j}{\|c_i\| \|c_j\|}$. Intuitively, $a_{i,j}$ measures the extent to which two funds' brokerage connections overlap.

We borrow techniques from graph theory and social network literature to quantify the importance of a mutual fund's position in the network. The importance of a node in a network is typically measured by its centrality and we use degree centrality (Freeman (1979)) and eigenvector centrality (Bonacich (1972, 1987)).[19] Degree centrality is defined as the sum of each row in the adjacency matrix, A, defining the network, scaled by the number of rows. Eigenvector centrality is defined as the principal eigenvector of the adjacency matrix defining the network. That is,

$$\lambda v = Av \tag{2}$$

where A is the adjacency matrix of the graph, λ is a constant (the eigenvalue), and v is the eigenvector.

In order to line up with the semi-annual N–SAR reporting frequency, we construct networks every half-year at the end of June (December) for N–SAR filings with reporting period ending in January to

[18]We wish to thank Kenneth Ahern (AIM Investment Conference discussant) for his detailed discussions on the network construction.

[19] Many different measures of centrality have been proposed and among the most commonly used measures of centrality are degree, closeness, betweenness, and eigenvector centrality. When choosing the most appropriate measure, one must consider the implicit assumptions underlying these centrality measures. As laid out in Borgatti (2005), closeness centrality and betweenness centrality are built upon an implicit assumption that traffic flows along the shortest paths until it reaches a pre-determined destination like the package delivery process. In our networks, traffic is likely to freely flow from one fund (the fund submitting a trade order) to another (a potential fund that could absorb the submitted trade order) through the broker intermediating the trade. Since this type of traffic must flow through unrestricted walks, rather than via geodesics, degree and eigenvector centrality are more suitable than closeness and betweenness centrality.

June (July to December) from the first half of 1994 to the first half of 2016. Since brokerage commission payments are only reported at the investment company level, we assume that all funds within the same investment company inherit the same network structure. Figure 2 shows institutional brokerage networks constructed using our N–SAR data for the first half of 2016. In general, funds that trade through the brokers that many other funds also trade through tend to be central in the network.[20]

[Insert Figure 2]

We next examine the characteristics of mutual funds' brokerage network centrality by estimating the following linear regression model:

$$Centrality_{i,t} = \gamma \times Covariates_{i,t} + \alpha_i + \varepsilon_{i,t} \tag{3}$$

where i indexes mutual funds and t indexes time in quarters. The dependent variable, $Centrality_{i,t}$, is fund i's brokerage network centrality measured at the end of quarter t. We use two measures of centrality: degree centrality and eigenvector centrality. $Covariates_{i,t-1}$ are a vector of fund and portfolio characteristics that include log of fund TNA, log of family TNA, turnover, expense ratio, and value-weighted averages of log of market capitalization, book-to-market ratio, and 12-month cumulative return (excluding the most recent month) of stock holdings, all measured at the end of quarter t. α_i denotes fund or family fixed-effects and standard errors are double-clustered by fund and time.

Table 3 presents the regression results. The dependent variable is degree centrality in columns (1) through (5) and eigenvector centrality in columns (6) through (10). Overall, we find that funds that are large or belong to large fund families tend to be more central in the network. This result is consistent with Goldstein et al. (2009) which notes that most institutions concentrate their order flows with a small number of brokers in order to become their important clients, whereas larger institutions can more easily obtain a premium status with several brokers.

[20] As noted in Bonacich (1991), centrality in affiliation networks involves the "duality" of groups and individuals. A broker gets its central position from the trading patterns of its client funds and a central fund should be one that trades through a larger network of central brokers (see Faust (1997) for more discussions on centrality in affiliation networks).

[Insert Table 3]

As can be seen in columns (1) and (6), fund and family sizes alone can explain 19% and 28% of the variation in degree centrality and eigenvector centrality, respectively. Adding other fund characteristics in columns (2) and (7) only marginally improves explanatory power, raising adjusted R^2 to 26% and 31% for degree centrality and eigenvector centrality, respectively. In contrast, fixed-effects, especially fund fixed-effects, account for a large amount of variation in brokerage network centrality, suggesting network effects that are orthogonal to size effects. Adding time and fund fixed-effects in columns (5) and (10) raises adjusted R^2 to 74% and 72% for degree centrality and eigenvector centrality, respectively. This result also implies that brokerage network centrality is highly persistent, reflecting the persistence in the underlying brokerage relationships.

4 The centrality premium

In Section 4.1, we begin our empirical analysis by showing that mutual funds' brokerage network centrality predicts their trading performance as measured by return gap, referred to as the centrality premium. In Section 4.2, we turn to inspecting the specific mechanisms driving the centrality premium.

4.1 The return gap premium associated with brokerage network centrality

Despite extensive disclosure requirements, mutual funds are only required to disclose their holdings on a quarterly basis and their trading activities are generally unobservable (Kacperczyk, Sialm, and Zheng (2008)). In order to examine how institutional brokerage networks affect mutual fund trading performance, we use the return gap as our measure of trading performance. The return gap is calculated as the difference between the reported fund return (gross of expenses) and the return on a hypothetical portfolio that invests in the previously disclosed fund holdings (Grinblatt and Titman (1989), Kacperczyk, Sialm, and Zheng

(2008)):

$$Return\ Gap_{i,t} = RET_{i,t} + EXP_{i,t} - HRET_{i,t} \tag{4}$$

where $RET_{i,t}$, is the fund i's reported return (net of expenses) during month t, $EXP_{i,t}$, is the expense ratio for fund i reported prior to month t, and $HRET_{i,t}$ is the fund i's holdings return during month t, which is defined as:

$$HRET_{i,t} = \sum_k w_{i,k,t-1} R_{k,t} \tag{5}$$

where $w_{i,k,t-1}$ is the fund i's portfolio weight on stock k at the end of month $t-1$ and $R_{k,t}$ is the return on stock k during month t. We use quarterly averages of the monthly return gap in our analysis and annualize the return gap by multiplying by twelve.

In order to understand the specific mechanisms through which brokerage network centrality affects the return gap, it is important to recognize key factors affecting the return gap. The return gap was originally proposed by Grinblatt and Titman (1989) as a measure of total transactions costs for mutual funds. Grinblatt and Titman (1989), however, point out that the return gap may be affected by interim trades within a quarter and, possibly, by window-dressing activities. Kacperczyk, Sialm, and Zheng (2008) further note that skilled fund managers can use their informational advantage to time the trades of individual stocks optimally and show that the past return gap helps predict fund performance.

We also recognize that network formation is likely endogenous. For instance, the marginal benefits of brokerage network centrality could be greater for more skilled fund managers that might self-select into central positions in the network. There might also exist an unobservable (to the econometrician) factor that is correlated with both mutual funds' brokerage network centrality and their trading performance. For example, Puckett and Yan (2011) find that institutional investors earn significant abnormal returns on their interim (intra-quarter) trades and that interim trading performance is persistent. In addition, Anand et al. (2012) show that trading costs are closely linked to trading desks' execution skills over and above selecting better brokers.

In order to mitigate these confounding factors, we use panel regressions with various fixed-effects to control for unobserved heterogeneity along with observable fund and portfolio characteristics. Specifically, we estimate the following linear regression model:

$$\textit{Return Gap}_{i,t} = \beta \times \textit{Centrality}_{i,t-1} + \gamma \times \textit{Covariates}_{i,t-1}\ (+\alpha_i) + \theta_t + \varepsilon_{i,t} \tag{6}$$

where i indexes mutual funds and t indexes time in quarters. The dependent variable, *Return Gap*$_{i,t}$, is fund i's average return gap during quarter t. *Centrality*$_{i,t-1}$ is fund i's brokerage network centrality measured prior to quarter t. We use two measures of centrality: degree centrality and eigenvector centrality. *Covariates*$_{i,t-1}$ are a vector of fund and portfolio characteristics that include log of fund TNA, log of family TNA, log of dollar volume, turnover, expense ratio, and value-weighted averages of log of market capitalization, book-to-market ratio, and 12-month cumulative return (excluding the most recent month) of stock holdings, all measured at the end of quarter $t-1$. In some specifications, the regression includes fund or family fixed-effects (α_i). All regressions include time fixed-effects (θ_t) and standard errors are double-clustered by fund and time.

We present the regression results in Table 4. We use *Degree Centrality* in columns (1) through (3) and *Eigenvector Centrality* in columns (4) and (6) as our measure of centrality. Columns (1) and (4) report our baseline specification including fund and portfolio characteristics and time-fixed effects. The estimate $\hat{\beta}$ is all positive and statistically significant at 1% levels, regardless of centrality measures used. One standard deviation increase in *Degree Centrality* is associated with an increase in return gap by 12.7 basis points per year ($0.127 = 1.27 \times 0.10$). Under our liquidity provision hypothesis, the economic magnitude is also meaningful. To put the numbers in perspective, Edelen (1999) estimates that liquidity-driven mutual fund trades lead to losses against informed traders on the order of approximately 140 basis points per year. In terms of an alternative hypothesis, Puckett and Yan (2011) estimate that interim trading performance contributes between 20 and 26 basis points to the annual abnormal returns of mutual funds.

Next, we add fund fixed-effects to our baseline specification to mitigate concerns about endogenous network formation. The identification relies on within-fund time-series variation in brokerage network centrality and largely comes from the aggregation of all small changes in brokerage connections of all the other funds, given the persistent nature of brokerage relationships. In columns (2) and (5), our results remain robust to controlling for fund fixed-effects. That is, mutual funds' brokerage network centrality positively predicts their trading performance in terms of return gap. In the remaining columns in (3) and (6), we replace fund fixed-effects with fund family fixed-effects. The estimate $\hat{\beta}$ is still positive, but becomes slightly weaker. This makes sense because, since funds belonging to the same fund family tend to have similar brokerage connections, there may be little within-family cross-sectional variation in brokerage network centrality. Overall, our results suggest that brokerage network centrality positively predicts mutual fund return gap, which we refer to as the centrality premium.

[Insert Table 4]

Before moving on, we conduct some robustness checks on our network construction. We recognize that our networks are incomplete because Item 20 of N–SAR filings only report brokerage commissions paid to the 10 brokers that received the largest amount from the investment company. A vast majority of mutual funds trade through more than 10 brokers and, as a result, we are likely to miss some of the less important brokerage connections outside the top 10 brokers. In order to understand how this data truncation may bias our results, we re-construct our networks by truncating our data even further to top 7, 5, or 3 brokers, re-calculate centrality measures, and re-estimate the baseline model in Equation (6). The results are reported in Table A2 in the Appendix. As we truncate our data further, the results remain qualitatively similar, but become quantitatively weaker, especially when we use *Eigenvector Centrality*. For instance, as we move from networks constructed through top 10 brokers to networks constructed through top 3 brokers, the coefficient on *Degree Centrality* decreases from 1.31 to 1.18 and the coefficient on *Eigenvector Centrality* decreases from 0.44 to 0.23. Thus, if anything, the incompleteness of our networks is likely to bias *downwards* the effects of brokerage network centrality on return gap.

4.2 Inspecting the mechanism

In this subsection, we present evidence in support of our liquidity provision hypothesis that institutional brokerage networks facilitate liquidity provision and mitigate price impact of large liquidity motivated trades. Specifically, we show that the centrality premium is more pronounced when funds are forced to liquidate a large fraction of their holdings to accommodate large investor redemptions, which we refer to as the *conditional* centrality premium. In addition, we show that the conditional centrality premium is strengthened by brokers' incentives to generate greater revenues and repeated interactions between funds and their brokers.

4.2.1 The centrality premium when funds are forced to liquidate

The primary prediction that we can derive from our hypothesis is that the centrality premium should be more pronounced when funds' trading activities are largely driven by liquidity motives and funds can credibly signal this to their brokers. We use large outflow events to identify such periods of liquidity-motivated trading. When a mutual fund is experiencing severe redemptions, the fund is forced to liquidate a large fraction of its holdings in several stocks and their selling is, to a large extent, uninformed (see, e.g., Edelen (1999), Coval and Stafford (2007)). In addition, such forced liquidations are likely to send a particularly strong signal to brokers that the fund's sell orders are liquidity driven, rather than information motivated, thus helping the brokers communicate more credibly with other institutional clients to take the other end of the trades.

In order to test this prediction, we estimate the following linear regression model:

$$\begin{aligned} \textit{Return Gap}_{i,t} = {} & \delta \times \textit{Centrality}_{i,t-1} \times \mathbb{1}(\textit{Outflow}_{i,t} > 5\%) + \beta \times \textit{Centrality}_{i,t-1} \\ & + \rho \times \mathbb{1}(\textit{Outflow}_{i,t} > 5\%) + \gamma \times \textit{Covariates}_{i,t-1} \ (+\alpha_i) + \theta_t + \varepsilon_{i,t} \end{aligned} \tag{7}$$

where i indexes mutual funds and t indexes time in quarters. The dependent variable, $\textit{Return Gap}_{i,t}$, is fund i's average return gap during quarter t. The interaction term, $\mathbb{1}(\textit{Outflow}_{i,t} > 5\%)$, is an indicator variable that is equal to 1 if fund i's percentage fund flow during quarter t falls below -5%. $\textit{Centrality}_{i,t-1}$ is fund i's brokerage network centrality measured prior to quarter t. We use two measures of centrality: degree centrality and eigenvector centrality. $\textit{Covariates}_{i,t-1}$ are the same vector of fund and portfolio characteristics as in Equation (6). In some specifications, the regression includes fund or family fixed-effects (α_i). All regressions include time fixed-effects (θ_t) and standard errors are double-clustered by fund and time.

We present the regression results in Table 5. We use *Degree Centrality* in columns (1) through (3) and *Eigenvector Centrality* in columns (4) and (6) as our measure of centrality. In the baseline specification in columns (1) and (4), the estimates $\hat{\delta}$ and $\hat{\beta}$ are all positive and statistically significant at conventional levels. This result suggests that central funds tend to outperform peripheral funds in terms of return gap during normal times, but the centrality premium is greater when funds are forced to liquidate their holdings to satisfy large investor redemptions. One standard deviation increase in *Degree Centrality* is associated with an increase in return gap by 30.7 basis points per year when funds are forced to liquidate to meet large redemptions ($0.307 = (2.25 + 0.82) \times 0.10$). This *conditional* centrality premium is almost three times as large as the *unconditional* centrality premium, consistent with our liquidity provision hypothesis that institutional brokerage networks mitigate price impact of large liquidity-motivated trades.

In the remaining columns, we add fund or family fixed-effects to our baseline specification to control for unobserved heterogeneity such as trading skills of fund managers and execution skills of trading desks. By exploiting within-fund or within-family variation in fund flows, we continue to find that the centrality premium is greater when funds experience large outflows. For robustness checks, we experiment with different cutoffs (7% or 3%) to define large outflow events and obtain qualitatively similar results, as reported in Table A3 in the Appendix.

[Insert Table 5]

Since central funds tend to belong to large fund families, a potential concern could be that the above results might be driven by cross-subsidization within a fund family: when a fund suffers severe redemptions, another fund in the same family could step in to provide liquidity. For instance, Bhattacharya, Lee, and Pool (2013) show that affiliated funds of mutual funds that invest only in other funds within the family provide an insurance pool against temporary liquidity shocks to other funds in the family. In order to examine whether family networks can explain away our results, we split our sample into three sub-samples by terciles based on the number of funds within the fund family and re-estimate the baseline specification in Equation (7). In order to be consistent with the alternative family network hypothesis, the estimate $\hat{\delta}$ should be larger among funds that belong to larger fund families that are better equipped to provide such within-family cross-subsidization.

The evidence reported in Table 6, however, is at odds with the alternative family network hypothesis. The conditional centrality premium is, in fact, *greater* among funds that belong to *smaller* families. This makes sense to the extent that large fund families should be capable of internalizing temporary liquidity shocks to their member funds and, as a result, those funds do not have to rely on their brokerage networks to mitigate liquidity-driven trading costs and externalities arising from forced liquidations (fire sales).

[Insert Table 6]

4.2.2 The conditional centrality premium for valuable clients

Our liquidity provision hypothesis requires an active role on the part of brokers, such as in discerning their clients' liquidity motives for trading and communicating with other institutional clients. As we have discussed, whether brokers facilitate liquidity provision or foster predatory trading is likely to hinge on the incentives they face and the strength of repeated interactions between brokers and their clients (Carlin et al. (2007)). To the extent that brokers are incentivized to maximize the expected value of future revenues from their clients, central funds with greater revenue generating potential for brokers may be better positioned to benefit from liquidity provision facilitated by brokers, especially when forced to liquidate in order to accommodate severe redemptions.

In order to test this prediction, we estimate the following linear regression model:

$$\begin{aligned}
\textit{Return Gap}_{i,t} = {} & \delta \times \textit{Centrality}_{i,t-1} \times \mathbb{1}(\textit{Outflow}_{i,t} > 5\%) \times \mathbb{1}(\textit{High Value Client}_{i,t-1}) \\
& + \beta_1 \times \textit{Centrality}_{i,t-1} \times \mathbb{1}(\textit{Outflow}_{i,t} > 5\%) + \beta_2 \times \textit{Centrality}_{i,t-1} \times \mathbb{1}(\textit{High Value Client}_{i,t-1}) \\
& + \beta_3 \times \mathbb{1}(\textit{Outflow}_{i,t} > 5\%) \times \mathbb{1}(\textit{High Value Client}_{i,t-1}) + \rho_1 \times \textit{Centrality}_{i,t-1} \\
& + \rho_2 \times \mathbb{1}(\textit{Outflow}_{i,t} > 5\%) + \rho_3 \times \mathbb{1}(\textit{High Value Client}_{i,t-1}) + \gamma \times \textit{Covariates}_{i,t-1}\ (+\alpha_i) + \theta_t + \varepsilon_{i,t}
\end{aligned} \tag{8}$$

where i indexes mutual funds and t indexes time in quarters. The triple interaction term, $\mathbb{1}(High\ Value\ Client_{i,t-1})$ is an indicator variable that takes the value of one if fund i's dollar volume during quarter $t-1$ belong to a top quartile. The rest of the variables are the same as in Equation (7). In some specifications, the regression includes fund or family fixed-effects (α_i). All regressions include time fixed-effects (θ_t) and standard errors are double-clustered by fund and time.

We present the regression results in Table 7. We use *Degree Centrality* in columns (1) through (3) and *Eigenvector Centrality* in columns (4) and (6) as our measure of centrality. In the baseline specification in columns (1) and (4), the estimates $\hat{\delta}$ and $\hat{\beta}_1$ are all positive and statistically significant at conventional levels. Even among low value clients, central funds earn a small, but positive centrality premium when forced to liquidate to meet large investor redemptions. The conditional centrality premium, however, is much greater for high value clients that have greater revenue generating potential for brokers. In the remaining columns, we control for fund or family fixed-effects and continue to find that the conditional centrality premium is greater among client funds that are highly valuable for brokers.

[Insert Table 7]

4.2.3 The conditional centrality premium for relationship clients

Carrying on our discussion in the previous subsection, our hypothesis also relies on the repeated nature of interaction between institutional clients and their brokers. Institutional investors must build reputation for being truthful in order to credibly signal liquidity motives for their uninformed orders to their brokers. The brokers, in turn, must develop their reputation capital for being discreet when handling their clients' orders. Thus, the signaling and certification of uninformed trading motives are likely most effective if funds have established trading relationships with their brokers.

In order to test this prediction, we estimate the following linear regression model:

$$
\begin{aligned}
\textit{Return Gap}_{i,t} = {} & \delta \times \textit{Centrality}_{i,t-1} \times \mathbb{1}(\textit{Outflow}_{i,t} > 5\%) \times \textit{Repeated Interaction}_{i,t-1} \\
& + \beta_1 \times \textit{Centrality}_{i,t-1} \times \mathbb{1}(\textit{Outflow}_{i,t} > 5\%) + \beta_2 \times \textit{Centrality}_{i,t-1} \times \textit{Repeated Interaction}_{i,t-1} \\
& + \beta_3 \times \mathbb{1}(\textit{Outflow}_{i,t} > 5\%) \times \textit{Repeated Interaction}_{i,t-1} + \rho_1 \times \textit{Centrality}_{i,t-1} \\
& + \rho_2 \times \mathbb{1}(\textit{Outflow}_{i,t} > 5\%) + \rho_3 \times \textit{Repeated Interaction}_{i,t-1} + \gamma \times \textit{Covariates}_{i,t-1}\ (+\alpha_i) + \theta_t + \varepsilon_{i,t}
\end{aligned}
\tag{9}
$$

where i indexes mutual funds and t indexes time in quarters. The triple interaction term, $\textit{Repeated Interaction}_{i,t-1}$, is measured by *cosine* similarity between fund i's brokerage commission vector available prior to quarter t and its brokerage commission vector one year prior to that. $\textit{Repeated Interaction}_{i,t-1}$ will be higher if fund i tends to trade through the brokers through which the fund used to trade in the past. The rest of the variables are the same as in Equation (7). In some specifications, the regression includes fund or family fixed-effects (α_i). All regressions include time fixed-effects (θ_t) and standard errors are double-clustered by fund and time.

We present the regression results in Table 8. We use *Degree Centrality* in columns (1) through (3) and *Eigenvector Centrality* in columns (4) and (6) as our measure of centrality. In the baseline specification in columns (1) and (4), the estimate $\hat{\delta}$ is all positive and statistically significant at 5% levels. Consistent with our argument that whether brokers facilitate liquidity provision is likely to hinge on the strength of repeated interactions between brokers and their clients, we find that the conditional centrality premium is greater for relationship clients. In the remaining columns, we find that our results remain robust to controlling for fund or family fixed-effects. Overall, our results in this and previous subsections suggest that the (conditional) centrality premium is further driven up by brokers' incentives to generate greater revenues and by repeated interactions between brokers and funds. These results are consistent with our hypothesis that institutional brokerage networks facilitate liquidity provision and mitigate price impact of large non-information motivated trades.

[Insert Table 8]

5 Further evidence from daily transactions

Under our liquidity provision hypothesis, the centrality premium is driven by lower trading costs. The return gap, however, reflects not only trading costs such as bid-ask spread, price impact, and brokerage commissions, but also a host of other factors including interim (intra-quarter) trading performance (Kacperczyk, Sialm, and Zheng (2008), Puckett and Yan (2011)). In addition, Irvine, Lipson, and Puckett (2007) document the relation between brokerage connections and early access to sell-side research ("tipping") (see also Goldstein et al. (2009), Xie (2014)). In order to corroborate our liquidity provision hypothesis and rule out alternative explanations for the centrality premium, we employ direct measures of trading costs and interim returns by restricting our sample to fund-quarters in which we can reliably merge daily transactions with quarterly holdings at the mutual fund level.

5.1 The trading costs channel

In this subsection, we present evidence that trading costs are the main driver of the centrality premium. For each trade, we compute dollar implementation shortfall as the difference between the trade execution price and the price prevailing at the time the broker receives the order (e.g., Anand et al. (2012)), multiplied by the number of shares executed:

$$\textit{Dollar Implementation Shortfall} = D \times (\textit{Price} - \textit{Benchmark Price}) \times \textit{Volume} \tag{10}$$

where D denotes the trade direction, taking a value of 1 for a buy and -1 for a sell, *Price* is the execution price of a trade, *Benchmark Price* is the price prevailing at the time the broker receives the order, and *Volume* is the number of shares executed.

To measure the fund-level trading cost component of the return gap, we aggregate trade-level trading costs in Equation (10) by summing over all trades for a given fund-month and then dividing by the fund's lagged TNA. Fund trading costs based on implementation shortfall capture *implicit* trading costs such as bid-ask spread and price impact. ANcerno also reports for each trade commissions and taxes plus fees paid to the executing brokers, which constitute *explicit* trading costs. We aggregate trade-level explicit trading costs in a similar way over all trades for a given fund-month. Total trading costs are the sum of implicit and explicit trading costs, expressed as percentage of lagged TNA. We use quarterly averages of monthly trading costs in our analysis and annualize trading costs by multiplying by twelve.

In order to corroborate that the centrality premium is indeed driven by lower trading costs incurred by central funds, we estimate the following linear regression model:

$$\textit{Trading Costs}_{i,t} = \beta \times \textit{Centrality}_{i,t-1} + \gamma \times \textit{Covariates}_{i,t-1} + \theta_t + \varepsilon_{i,t} \tag{11}$$

where i indexes mutual funds and t indexes time in quarters. The dependent variable, $\textit{Trading Costs}_{i,t}$, is fund i's average trading costs during quarter t. We use three measures of trading costs: implicit, explicit, and total trading costs. $\textit{Centrality}_{i,t-1}$ is fund i's brokerage network centrality measured prior to quarter t. We use two measures of centrality: degree centrality and eigenvector centrality. $\textit{Covariates}_{i,t-1}$ are the same vector of fund and portfolio characteristics as in Equation (6). All regressions include time fixed-effects (θ_t) and standard errors are double-clustered by fund and time.

We present the regression results in Table 9. We use *Degree Centrality* in columns (1) through (3) as our measure of centrality. In column (1), the dependent variable is *Implicit Trading Costs*, which capture bid-ask spread and price impact. The estimate $\hat{\beta}$ is negative and statistically significant at the 5% level, consistent with our liquidity provision hypothesis. One standard deviation increase in *Degree Centrality* is associated with a decrease in implicit trading costs by 7.4 basis points per year ($-0.074 \approx -0.82 \times 0.09$). In column (2), the dependent variable is *Explicit Trading Costs*, which primarily capture brokerage commissions. The estimate $\hat{\beta}$ is negative and statistically significant at the 10% level, which casts doubt on the alternative "tipping" channel as a driver of the centrality premium. The finding that central funds tend to incur lower explicit trading costs suggests that, if anything, central funds are *less* likely to pay soft dollars that are associated with profitable analyst recommendations (e.g., Irvine, Lipson, and Puckett (2007), Goldstein et al. (2009)). In column (3), the dependent variable is *Total Trading Costs*, which is the sum of *Implicit Trading Costs* and *Explicit Trading Costs*. One standard deviation increase in *Degree Centrality* is associated with a decrease in total trading costs by 10 basis points per year ($-0.10 \approx -1.11 \times 0.09$). When we replace *Degree Centrality* with *Eigenvector Centrality* in columns (4) through (6), we obtain qualitatively similar results that central funds tend to incur lower trading costs.

[Insert Table 9]

In order to lend further support to our liquidity provision hypothesis, we estimate the following linear regression model:

$$\begin{aligned} \textit{Trading Costs}_{i,t} &= \delta \times \textit{Centrality}_{i,t-1} \times \mathbb{1}(\textit{Outflow}_{i,t} > 5\%) + \beta \times \textit{Centrality}_{i,t-1} \\ &+ \rho \times \mathbb{1}(\textit{Outflow}_{i,t} > 5\%) + \gamma \times \textit{Covariates}_{i,t-1} + \theta_t + \varepsilon_{i,t} \end{aligned} \tag{12}$$

where i indexes mutual funds and t indexes time in quarters. The interaction term, $\mathbb{1}(\textit{Outflow}_{i,t} > 5\%)$, is an indicator variable that is equal to 1 if fund i's percentage fund flow during quarter t falls below -5%. The rest of the variables are the same as in Equation (11). All regressions include time fixed-effects (θ_t) and standard errors are double-clustered by fund and time.

We present the regression results in Table 10. We use *Degree Centrality* as our centrality measure in columns (1) through (3). In column (1), the dependent variable is *Implicit Trading Costs*. The estimate $\hat{\delta}$ is negative and statistically significant at the 5% level, consistent with our hypothesis that institutional brokerage networks mitigate price impact of large liquidity-motivated trades. One standard deviation increase in *Degree Centrality* is associated with a decrease in *Implicit Trading Costs* by 17.6 basis points per year when funds are forced to liquidate holdings to accommodate large redemptions ($-0.176 \approx (-1.31 - 0.65) \times 0.09$). In column (2), the dependent variable is *Explicit Trading Costs*. The negative estimate $\hat{\delta}$ suggests that central funds incur smaller brokerage commissions when forced to trade, which further reduces trading costs associated with liquidity-motivated trades. When we consider *Total Trading Costs* in column (3), we find that one standard deviation increase in *Degree Centrality* is associated with a decrease in *Total Trading Costs* by 26.1 basis points per year when funds are forced to trade ($-0.261 = (-2.06 - 0.84) \times 0.09$). When we replace *Degree Centrality* with *Eigenvector Centrality* in columns (4) through (6), we obtain qualitatively similar results that brokerage network centrality helps reduce trading costs associated with large liquidity-motivated trades.

[Insert Table 10]

5.2 The interim trading skills channel

In this subsection, we address concerns that interim (intra-quarter) trading performance (Kacperczyk, Sialm, and Zheng (2008), Puckett and Yan (2011)) could be another factor driving the centrality premium. To explore this alternative channel, we compute the mutual fund returns from interim trading (e.g., Puckett and Yan (2011)). Specifically, for each trade executed between adjacent portfolio report dates, we track the performance of the trade from the execution date (using the execution price) to the following portfolio report date and compute dollar interim returns of the trade each month during that time period. To measure the fund-level interim trading performance component of the return gap, we aggregate trade-level dollar interim returns by summing over all trades executed between adjacent portfolio report dates for each fund-month and then dividing by the fund's lagged TNA. We use quarterly averages of monthly interim returns in our analysis and annualize interim returns by multiplying by twelve.

In order to examine whether the centrality premium is driven by higher interim trading performance of central funds, we estimate the following linear regression model:

$$\textit{Interim Return}_{i,t}\ (\textit{Return Gap}_{i,t}) = \beta \times \textit{Centrality}_{i,t-1} + \gamma \times \textit{Covariates}_{i,t-1} + \theta_t + \varepsilon_{i,t} \tag{13}$$

where i indexes mutual funds and t indexes time in quarters. The dependent variable, $\textit{Interim Return}_{i,t}$, is fund i's average interim return during quarter t. For a comparison, we use $\textit{Return Gap}_{i,t}$ as the dependent variable side-by-side. $\textit{Centrality}_{i,t-1}$ is fund i's brokerage network centrality measured prior to quarter t. We use two measures of centrality: degree centrality and eigenvector centrality. $\textit{Covariates}_{i,t-1}$ are the same vector of fund and portfolio characteristics as in Equation (6). All regressions include time fixed-effects (θ_t) and standard errors are double-clustered by fund and time.

We present the regression results in Table 11. We use *Degree Centrality* in columns (1) and (2) as our measure of centrality. In column (1), the dependent variable is *Return Gap*. The estimate $\hat{\beta}$ is positive and statistically significant at the 10% level, which is reassuring in that our return gap results found in the comprehensive CRSP–Thomson–NSAR merged sample largely carry over to the much restricted CRSP–Thomson–NSAR–ANcerno merged sample. In column (2), the dependent variable is *Interim Return*. The negative, albeit insignificant, estimate $\hat{\beta}$ suggests that central funds do not exhibit higher interim trading performance, casting doubt on interim trading performance as a driver of the centrality premium. When we replace *Degree Centrality* with *Eigenvector Centrality* in columns (3) and (4), we obtain qualitatively similar results that interim trading performance can be safely ruled out an alternative explanation for the centrality premium.

[Insert Table 11]

6 Conclusion

Using a unique dataset on brokerage commission payments for a comprehensive sample of mutual funds, we map trading networks of mutual funds and their brokers as affiliation networks in which mutual funds are connected through their overlapping brokerage relationships. Mutual funds that trade through a large number of brokers that many other funds also trade through tend to be central in the network. We find that central funds outperform peripheral ones, especially in terms of return gap. In order to shed light on the specific mechanisms behind this centrality premium return gap, we propose a liquidity provision hypothesis.

Our contention is that central funds will tend to receive better price execution when they trade for liquidity reasons. Consider, for instance, a central fund subject to an extreme fund outflow. We would expect the fund to trade through brokers with whom it has strong trading relationships. These ongoing relationships could allow the fund to credibly communicate its liquidity motives for the trades. Brokers will also be less willing to exploit the information communicated by a central client, if it could lead to a loss of its relationship/reputation capital. Brokers are likely to turn to other institutional clients with whom they have strong relationships to absorb the orders, while communicating the likely liquidity motives for the trades.

Consistent with our liquidity provision hypothesis, we find that the centrality premium is more pronounced when funds' trading activities are largely driven by liquidity motives, such as to accommodate large fund outflows. We find that this conditional centrality premium is further strengthened by brokers' incentives to generate greater revenues and by repeated interactions between brokers and funds. By merging daily transactions with quarterly holdings, we confirm that the centrality premium is indeed driven by reduced trading costs, rather than higher interim (intra-quarter) trading performance or profitable information flows from the brokers.

References

Admati, A. R., and P. Pfleiderer. 1991. Sunshine Trading and Financial Market Equilibrium. *Review of Financial Studies* 4:443–481.

Agarwal, V., Y. Tang, and B. Yang. 2012. Do Mutual Funds Have Market Timing Ability? Evidence from Mutual Fund Trades. *Working Paper, Georgia State University* .

Anand, A., P. Irvine, A. Puckett, and K. Venkataraman. 2012. Performance of Institutional Trading Desks: An Analysis of Persistence in Trading Costs. *Review of Financial Studies* 25:557–598.

Barbon, A., M. Di Maggio, F. A. Franzoni, and A. Landier. 2019. Brokers and Order Flow Leakage: Evidence from Fire Sales. *The Journal of Finance* 74:2707–2749.

Bessembinder, H., C. S. Spatt, and K. Venkataraman. Forthcoming. A Survey of the Microstructure of Fixed-Income Markets. *Journal of Financial and Quantitative Analysis* .

Bessembinder, H., and K. Venkataraman. 2004. Does an Electronic Stock Exchange Need an Upstairs Market? *Journal of Financial Economics* 73:3–36.

Bhattacharya, U., J. H. Lee, and V. K. Pool. 2013. Conflicting Family Values in Mutual Fund Families. *Journal of Finance* 68:173–200.

Bonacich, P. 1972. Factoring and Weighting Approaches to Status Scores and Clique Identification. *The Journal of Mathematical Sociology* 2:113–120.

Bonacich, P. 1987. Power and Centrality: A Family of Measures. *American Journal of Sociology* 92:1170–1182.

Bonacich, P. 1991. Simultaneous group and individual centralities. *Social Networks* 13:155–168.

Booth, G. G., J.-C. Lin, T. Martikainen, and Y. Tse. 2002. Trading and Pricing in Upstairs and Downstairs Stock Markets. *Review of Financial Studies* 15:1111–1135.

Borgatti, S. P. 2005. Centrality and Network Flow. *Social Networks* 27:55–71.

Brunnermeier, M. K., and L. H. Pedersen. 2005. Predatory Trading. *Journal of Finance* 60:1825–1863.

Busse, J. A., T. Chordia, L. Jiang, and Y. Tang. 2019. Transaction Costs, Portfolio Characteristics, and Mutual Fund Performance. *Working Paper, Emory University* .

Carlin, B. I., M. S. Lobo, and S. Viswanathan. 2007. Episodic Liquidity Crises: Cooperative and Predatory Trading. *Journal of Finance* 62:2235–2274.

Chung, J.-W., and B. U. Kang. 2016. Prime Broker-level Comovement in Hedge Fund Returns: Information or Contagion? *Review of Financial Studies* 29:3321–3353.

Coval, J., and E. Stafford. 2007. Asset Fire Sales (and Purchases) in Equity Markets. *Journal of Financial Economics* 86:479–512.

Di Maggio, M., F. A. Franzoni, A. Kermani, and C. Sommavilla. 2019. The relevance of broker networks for information diffusion in the stock market. *Journal of Financial Economics* 134:419 – 446.

Di Maggio, M., A. Kermani, and Z. Song. 2017. The value of trading relations in turbulent times. *Journal of Financial Economics* 124:266–284.

Edelen, R. M. 1999. Investor Flows and the Assessed Performance of Open-end Mutual Funds. *Journal of Financial Economics* 53:439–466.

Faust, K. 1997. Centrality in affiliation networks. *Social Networks* 19:157–191.

Freeman, L. C. 1979. Centrality in Social Networks Conceptual Clarification. *Social Networks* 1:215–239.

Glosten, L. R., and P. R. Milgrom. 1985. Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders. *Journal of Financial Economics* 14:71–100.

Goldstein, M. A., P. Irvine, E. Kandel, and Z. Wiener. 2009. Brokerage Commissions and Institutional Trading Patterns. *Review of Financial Studies* 22:5175–5212.

Green, T. C., R. Jame, S. Markov, and M. Subasi. 2014. Broker-hosted Investor Conferences. *Journal of Accounting and Economics* 58:142–166.

Greenwich Associates. 2017. The Voice of Trading .

Griffin, J. M., T. Shu, and S. Topaloglu. 2012. Examining the Dark Side of Financial Markets: Do Institutions Trade on Information from Investment Bank Connections? *Review of Financial Studies* 25:2155–2188.

Grinblatt, M., and S. Titman. 1989. Mutual Fund Performance: An Analysis of Quarterly Portfolio Holding. *Journal of Business* 62:393–416.

Grossman, S. J. 1992. The Informational Role of Upstairs and Downstairs Trading. *Journal of Business* 65:509–528.

Hendershott, T., D. Li, D. Livdan, and N. Schürhoff. 2017. Relationship Trading in OTC Markets. *Swiss Finance Institute Research Paper No. 17-30* .

Irvine, P., M. Lipson, and A. Puckett. 2007. Tipping. *Review of Financial Studies* 20:741–768.

Kacperczyk, M., C. Sialm, and L. Zheng. 2008. Unobserved Actions of Mutual Funds. *Review of Financial Studies* 21:2379–2416.

Kumar, N., K. Mullally, S. Ray, and Y. Tang. Forthcoming. Prime (Information) Brokerage. *Journal of Financial Economics* .

Kyle, A. S. 1985. Continuous Auctions and Insider Trading. *Econometrica* 53:1315–1335.

Levenshtein, V. I. 1966. Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady* 10:707–701.

Li, D., and N. Schürhoff. 2019. Dealer Networks. *Journal of Finance* 74:91–144.

Madhavan, A., and M. Cheng. 1997. In Search of Liquidity: Block Trades in the Upstairs and and Downstairs. *Review of Financial Studies* 10:175–203.

Ozsoylev, H. N., J. Walden, M. D. Yavuz, and R. Bildik. 2014. Investor Networks in the Stock Market. *Review of Financial Studies* 27:1323–1366.

Puckett, A., and X. S. Yan. 2011. The Interim Trading Skills of Institutional Investors. *Journal of Finance* 66:601–633.

Reuter, J. 2006. Are IPO Allocations for Sale? Evidence from Mutual Funds. *Journal of Finance* 61:2289–2324.

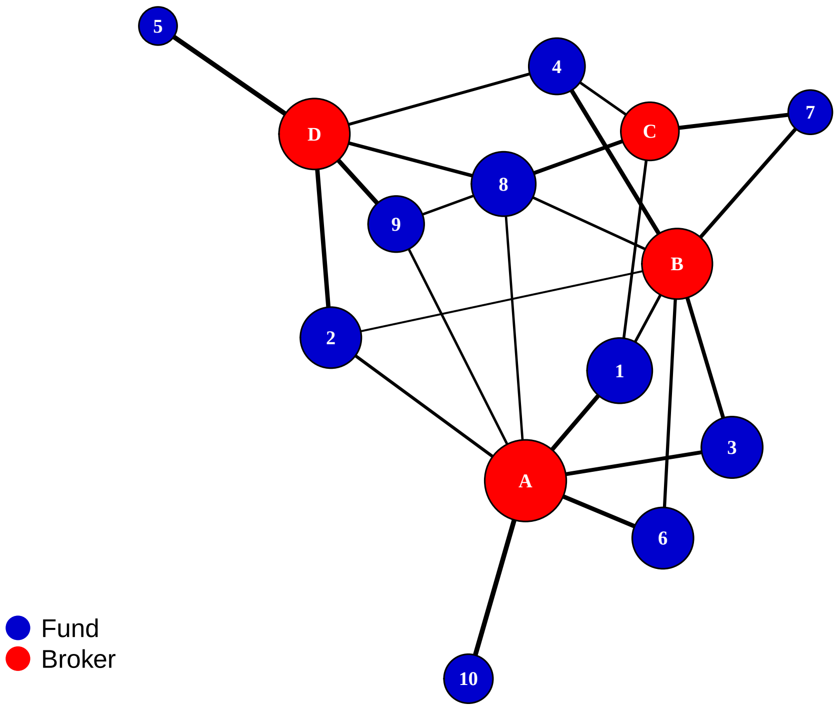
Seppi, D. J. 1990. Equilibrium Block Trading and Asymmetric Information. *Journal of Finance* 45:73–94.

Smith, B. F., D. A. S. Turnbull, and R. W. White. 2001. Upstairs Market for Principal and Agency Trades: Analysis of Adverse Information and Price Effects. *Journal of Finance* 56:1723 –1746.

Xie, L. 2014. Above and Beyond Recommendations: How Analysts Add Value to Their Fund Clients. *Working paper* .

Yang, L., and H. Zhu. 2019. Back-Running: Seeking and Hiding Fundamental Information in Order Flows. *The Review of Financial Studies* .

Panel A: Graphical representation



Panel B: Fund-to-broker direct connections

$$
\begin{array}{rc}
 & \text{Fund} \\
 & \begin{array}{cccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \end{array} \\
G^T = \quad \text{Broker} \begin{array}{c} A \\ B \\ C \\ D \end{array} & \begin{bmatrix}
0.851 & 0.483 & 0.707 & 0.000 & 0.000 & 0.868 & 0.000 & 0.238 & 0.280 & 1.000 \\
0.375 & 0.109 & 0.707 & 0.852 & 0.000 & 0.496 & 0.660 & 0.317 & 0.000 & 0.000 \\
0.368 & 0.000 & 0.000 & 0.328 & 0.000 & 0.000 & 0.751 & 0.693 & 0.287 & 0.000 \\
0.000 & 0.869 & 0.000 & 0.409 & 1.000 & 0.000 & 0.000 & 0.602 & 0.916 & 0.000
\end{bmatrix}
\end{array}
$$

Panel C: Fund-to-fund indirect connections

$$A = \text{Fund}\;\begin{array}{c} \\ 1\\2\\3\\4\\5\\6\\7\\8\\9\\10 \end{array}\!\!\begin{array}{c} \text{Fund} \\ \begin{array}{cccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \end{array} \\ \begin{bmatrix} 1.000 & 0.452 & 0.867 & 0.440 & 0.000 & 0.925 & 0.524 & 0.577 & 0.344 & 0.851 \\ 0.452 & 1.000 & 0.418 & 0.448 & 0.869 & 0.473 & 0.072 & 0.673 & 0.931 & 0.483 \\ 0.867 & 0.418 & 1.000 & 0.602 & 0.000 & 0.965 & 0.467 & 0.393 & 0.198 & 0.707 \\ 0.440 & 0.448 & 0.602 & 1.000 & 0.409 & 0.423 & 0.809 & 0.744 & 0.469 & 0.000 \\ 0.000 & 0.869 & 0.000 & 0.409 & 1.000 & 0.000 & 0.000 & 0.602 & 0.916 & 0.000 \\ 0.925 & 0.473 & 0.965 & 0.423 & 0.000 & 1.000 & 0.328 & 0.364 & 0.243 & 0.868 \\ 0.524 & 0.072 & 0.467 & 0.809 & 0.000 & 0.328 & 1.000 & 0.730 & 0.216 & 0.000 \\ 0.577 & 0.673 & 0.393 & 0.744 & 0.602 & 0.364 & 0.730 & 1.000 & 0.817 & 0.238 \\ 0.344 & 0.931 & 0.198 & 0.469 & 0.916 & 0.243 & 0.216 & 0.817 & 1.000 & 0.280 \\ 0.851 & 0.483 & 0.707 & 0.000 & 0.000 & 0.868 & 0.000 & 0.238 & 0.280 & 1.000 \end{bmatrix} \end{array}$$

Panel D: Brokerage network centrality

Fund	1	2	3	4	5	6	7	8	9	10
Degree Centrality	0.598	0.582	0.562	0.534	0.380	0.559	0.414	0.614	0.541	0.443
Eigenvector Centrality	1.000	0.935	0.941	0.863	0.586	0.939	0.677	0.987	0.859	0.750

Figure 1: Institutional brokerage networks: an illustration

This figure illustrates how we construct institutional brokerage networks and calculate brokerage network centrality using a simple example network consisting of ten funds and four brokers. Its graphical representation and adjacency matrix, denoted G, are presented in Panel A and Panel B, respectively. In Panel A, the size of a node is proportional to its eigenvector centrality. In Panel B, each element ($g_{i,k}$) represents the strength of connection between fund i and broker k, as measured by commissions paid to broker k by fund i, normalized so that g_i is a unit vector. That is, g_i is equal to $\frac{c_i}{\|c_i\|}$ where c_i is a vector of fund i's commission payments. To measure a mutual fund's connections to all the other mutual funds through their overlapping brokerage connections, we reduce the bi-partite graph of mutual funds and brokers into a mono-partite graph of mutual funds only by defining its adjacency matrix A as $A = GG^T$ (see, e.g., Faust (1997)). Panel C presents A where each element ($a_{i,j}$) represents the strength of connection between fund i and fund j, as measured by the cosine similarity between brokerage commission vectors for that pair, i.e., $a_{i,j} = g_i \cdot g_j = \frac{c_i \cdot c_j}{\|c_i\|\|c_j\|}$. Panel D presents brokerage network centrality, as measured by degree centrality (Freeman (1979)) and eigenvector centrality (Bonacich (1972, 1987)). Degree centrality is defined as the sum of all elements in each row of the adjacency matrix A, scaled by the number of rows. Eigenvector centrality is defined as the principal eigenvector associated with the adjacency matrix A, scaled so that the largest element is equal to unity. That is, $\lambda v = Av$ where A is the adjacency matrix of the graph, λ is a constant (the eigenvalue), and v is the eigenvector.

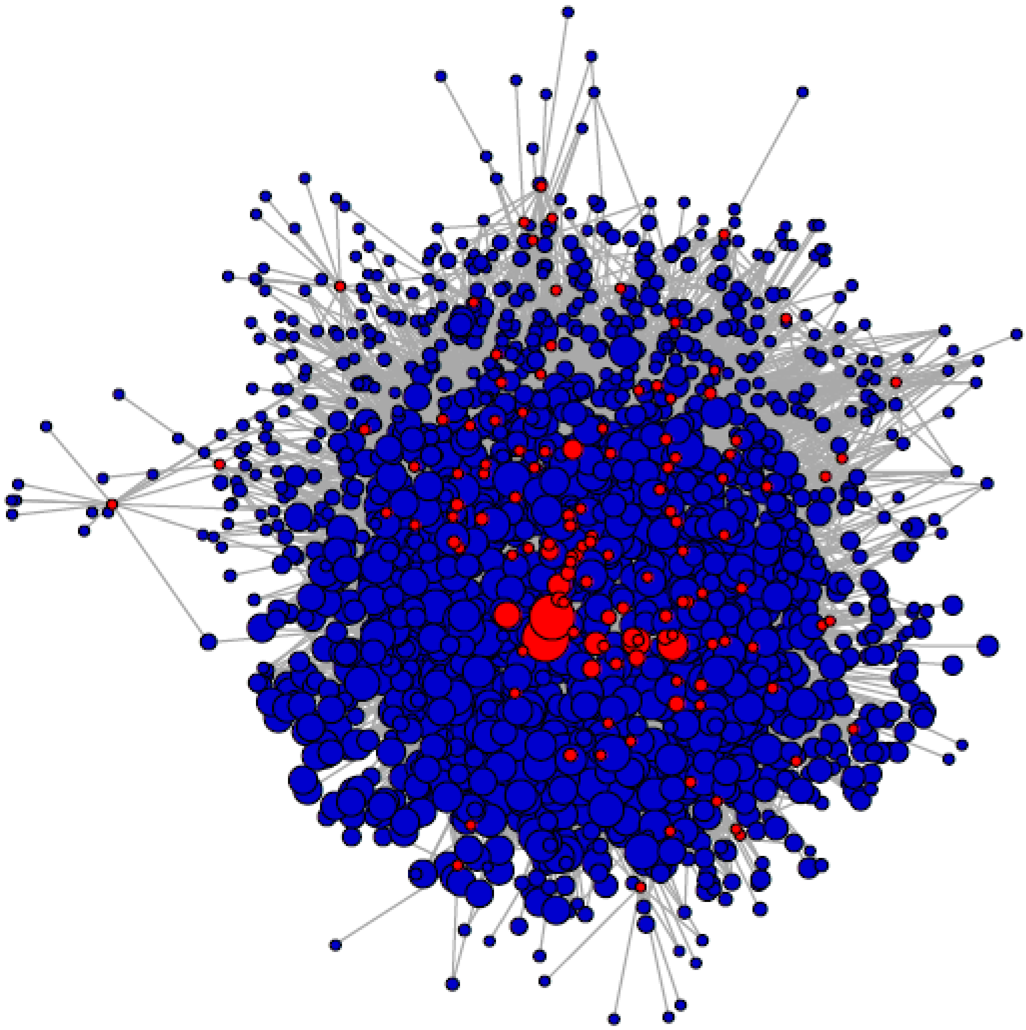


Figure 2: Institutional brokerage networks: a graphical representation

This figure shows a snapshot of institutional brokerage networks constructed using our N–SAR dataset at the end of June 2016. Blue nodes represent mutual funds, red nodes represent brokers, and lines represent connections between mutual funds and their brokers. The size of a node is proportional to its eigenvector centrality.

Table 1: Brokerage commission payments: example and descriptive statistics

This table provides an example of and some descriptive statistics on brokerage commission payments. N-SAR filings report brokerage commissions paid to the 10 brokers that received the largest amount (Item 20) from the investment company and the aggregate brokerage commission payments (Item 21). Panel A provides an example for T. Rowe Price Blue Chip Growth Fund for the period ending in June 30, 2016. Panel B reports the commissions paid to top n brokers, as percentage of commissions paid to *all* brokers. Panel C reports the transition matrix of year-to-year changes in the broker rankings for a given investment company by the amount of commission payments.

Panel A: Example: T ROWE PRICE BLUE CHIP GROWTH FUND (CIK = 902259), June 30, 2016

Item 20	Name of Broker	IRS Number	Commisions ($000)
1	BANK OF AMERICA MERRILL LYNCH	13-5674085	415
2	JPMORGAN CHASE	13-4994650	292
3	MORGAN STANLEY CO INC	13-2655998	252
4	DEUTSCHE BANK SECURITIES	13-2730828	207
5	RBC CAPITAL MARKETS	41-1416330	159
6	CITIGROUP GLOBAL MARKETS INC	11-2418191	157
7	CS FIRST BOSTON	13-5659485	153
8	BAIRD ROBERT W	39-6037917	148
9	GOLDMAN SACHS	13-5108880	144
10	SANFORD C BERNSTEIN	13-2625874	115
Item 21	Aggregate Brokerage Commissions ($000)		3107

Panel B: Concentration of brokerage commissions

	Commissions paid to top n brokers as % of commissions paid to *all* brokers				
	Mean	St. Dev.	Q_1	Median	Q_3
Top 1 Broker	25.65	22.21	11.54	16.88	30.00
Top 3 Brokers	45.24	23.95	27.59	37.44	56.76
Top 5 Brokers	56.60	22.53	39.26	51.17	71.13
Top 7 Brokers	64.47	20.91	48.25	61.32	80.63
Top 10 Brokers	72.45	18.91	58.08	71.62	88.89

Table 1–*Continued*

Panel C: Persistence in Brokerage Relationship (Transition Matrix)

Probability (%)	Next Year									
Current Year	Top 1	Top 2	Top 3	Top 4	Top 5	Top 6	Top 7	Top 8	Top 9	Top 10
Top 1	46.74	20.55	13.35	10.06	7.57	6.44	5.41	4.99	4.41	4.00
Top 2	17.37	23.69	17.29	13.03	10.96	8.89	7.58	6.63	6.30	5.65
Top 3	10.71	15.83	17.64	14.52	12.34	10.93	9.61	8.53	7.14	7.62
Top 4	7.17	11.24	13.33	15.12	13.31	11.82	10.10	9.59	9.42	8.47
Top 5	5.31	8.09	10.53	12.73	13.83	12.84	12.16	10.54	10.10	9.67
Top 6	4.00	6.47	8.87	10.49	11.84	13.00	12.91	12.12	10.78	10.67
Top 7	3.12	5.30	6.65	8.18	10.38	11.67	12.77	13.11	12.71	11.21
Top 8	2.41	3.60	5.00	6.56	8.02	9.93	11.72	13.73	13.70	13.53
Top 9	1.85	2.86	4.06	5.12	6.15	8.22	10.22	11.28	14.08	13.93
Top 10	1.32	2.36	3.28	4.19	5.60	6.25	7.52	9.48	11.35	15.26

Table 2: Summary statistics

This table reports the summary statistics on the return gap (Grinblatt and Titman (1989), Kacperczyk, Sialm, and Zheng (2008)), degree centrality (Freeman (1979)), eigenvector centrality (Bonacich (1972, 1987)), and other fund and portfolio characteristics that include log of fund total net assets (TNA), log of fund family TNA, turnover, expense ratio, and value-weighted averages of log of market capitalization, book-to-market ratio, and 12-month cumulative return (skipping the most recent month) of the stock holdings. We calculate monthly net flows for each fund share class i during month t as follows: $FLOW_{i,t} = TNA_{i,t} - TNA_{i,t-1} \times (1 + R_{i,t})$ where $FLOW_{i,t}$ is the dollar value of fund flow (net new issues and redemptions), $TNA_{i,t}$ is the total net asset, and $R_{i,t}$ is the monthly return. To compute the monthly fund flow for the fund, we sum monthly fund flows for all share classes belonging to the same fund as identified by MFLINKS. Monthly fund flows are summed over the quarter to calculate the quarterly fund flow. We scale the quarterly fund flow by the beginning-of-period TNA. $\mathbb{1}(\text{Outflow} > 5\%)$ is an indicator variable that takes the value of one if the quarterly percentage fund flow is less than -5% and zero otherwise. The CRSP–Thomson–NSAR merged sample in Panel A is for the period from January 1994 to December 2016. For the CRSP–Thomson–NSAR–ANcerno merged sample in Panel B, we also report the summary statistics on implicit, explicit, and total trading costs, and interim return for the period from January 1999 to March 2011. The unit of observation is fund-quarters. Panel C reports the differences in average fund-level variables between the two samples. Standard errors are double-clustered by fund and time, and the resulting t-statistics are reported in parentheses. Statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

Panel A: CRSP–Thomson–NSAR merged sample

Variable	Obs.	Mean	St. Dev.	Q_1	Median	Q_3
Return Gap (%)	123,565	−0.34	6.80	−3.13	−0.22	2.43
Degree Centrality	123,565	0.19	0.10	0.11	0.20	0.27
Eigenvector Centrality	123,565	0.50	0.27	0.28	0.54	0.73
$\mathbb{1}$(Outflow > 5%)	123,565	0.20	0.40	0	0	0
log(Fund TNA)	123,565	5.63	1.83	4.29	5.59	6.93
log(Family TNA)	123,565	9.48	2.49	7.82	9.76	11.23
Turnover (%)	123,565	78.32	73.62	29	58	102
Expense Ratio (%)	123,565	1.14	0.43	0.93	1.13	1.37
log(Market Capitalization)	123,565	8.84	1.28	7.79	9.23	9.97
Book-to-market	123,565	0.47	0.18	0.34	0.45	0.58
Momentum (%)	123,565	20.07	24.50	5.75	17.81	31.92

Table 2–*Continued*

Panel B: CRSP–Thomson–NSAR–ANcerno merged sample

Variable	Obs.	Mean	St. Dev.	Q_1	Median	Q_3
Return Gap (%)	7,791	0.58	6.85	−2.36	0.40	3.36
Implicit Trading Costs (%)	7,791	0.45	1.15	0.02	0.17	0.48
Explicit Trading Costs (%)	7,791	0.27	0.55	0.06	0.14	0.28
Total Trading Costs (%)	7,791	0.72	1.55	0.09	0.32	0.76
Interim Return (%)	7,791	0.41	7.07	−1.49	0.08	1.89
Degree Centrality	7,791	0.27	0.09	0.22	0.27	0.33
Eigenvector Centrality	7,791	0.65	0.21	0.54	0.71	0.82
$\mathbb{1}$(Outflow > 5%)	7,791	0.20	0.40	0	0	0
log(Fund TNA)	7,791	6.37	1.95	4.95	6.28	7.83
log(Family TNA)	7,791	11.95	2.18	10.78	13.17	13.62
Turnover (%)	7,791	96.01	76.66	42	78	128
Expense Ratio (%)	7,791	1.10	0.47	0.86	1.10	1.35
log(Market Capitalization)	7,791	9.01	1.10	8.38	9.38	9.92
Book-to-market	7,791	0.47	0.18	0.33	0.44	0.56
Momentum (%)	7,791	17.02	26.05	2.05	14.99	29.96

Panel C: Differences in fund-level variables between the two samples

	CRSP–Thomson–		Difference	
Variable	NSAR	NSAR–ANcerno	(2) − (1)	(t-stat)
Return Gap (%)	−0.34	0.58	0.98***	(4.98)
Degree Centrality	0.19	0.27	0.08***	(13.94)
Eigenvector Centrality	0.50	0.65	0.16***	(14.54)
$\mathbb{1}$(Outflow > 5%)	0.20	0.20	−0.002	(−0.17)
log(Fund TNA)	5.63	6.37	0.79***	(6.56)
log(Family TNA)	9.48	11.95	2.63***	(18.51)
Turnover (%)	78.32	96.01	18.88***	(4.78)
Expense Ratio (%)	1.14	1.10	−0.04	(−1.46)
log(Market Capitalization)	8.84	9.01	0.18***	(3.00)
Book-to-market	0.47	0.47	−0.007	(−0.55)
Momentum (%)	20.07	17.02	−3.26*	(−1.88)

Table 3: Determinants of mutual funds' brokerage network centrality

This table presents the results of the following linear regression model:

$$Centrality_{i,t} = \gamma \times Covariates_{i,t} \ (+\alpha_i) + \varepsilon_{i,t}$$

where i indexes mutual funds and t indexes time in quarters. The dependent variable, $Centrality_{i,t}$, is fund i's brokerage network centrality measured at the end of quarter t. We use two measures of centrality: degree centrality and eigenvector centrality. $Covariates_{i,t}$ are a vector of fund and portfolio characteristics that include log of fund TNA, log of family TNA, turnover, expense ratio, and value-weighted averages of log of market capitalization, book-to-market ratio, and 12-month cumulative return (excluding the most recent month) of stock holdings, all measured at the end of quarter t. α_i denotes fund or family fixed-effects. Standard errors are double-clustered by fund and time, and the resulting t-statistics are reported in parentheses. Statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

Dependent variable:	Degree Centrality ×100						Eigenvector Centrality ×100					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
log(Fund TNA)	1.38***		0.05	0.20**	0.21*	0.25***	4.40***		0.48**	0.69***	0.87***	0.78***
	(18.03)		(0.51)	(2.44)	(1.88)	(4.60)	(21.77)		(2.15)	(3.19)	(3.33)	(5.76)
log(Family TNA)		1.75***	1.73***	1.76***	0.16	0.05		5.30***	5.10***	5.07***	1.10***	1.84***
		(27.52)	(22.85)	(23.15)	(1.16)	(0.18)		(33.02)	(25.93)	(24.62)	(4.52)	(4.35)
Turnover (%)				0.01***	0.01***	0.01***				0.03***	0.01**	0.01***
				(7.26)	(4.41)	(5.89)				(7.96)	(2.48)	(5.61)
Expense Ratio (%)				1.60***	3.01***	1.82***				1.57*	0.46	0.95
				(4.19)	(4.99)	(5.77)				(1.71)	(0.37)	(1.24)
log(Market Capitalization)				0.36***	0.64	0.21***				1.30***	1.04	0.62***
				(2.94)	(1.43)	(2.94)				(4.48)	(1.50)	(4.12)
Book-to-market				−5.96***	−5.44***	−3.21***				−9.41***	−4.91***	−2.92***
				(−5.71)	(−3.69)	(−4.59)				(−4.93)	(−2.62)	(−2.59)
Momentum (%)				−0.02	−0.02	−0.02*				−0.02	−0.02	−0.02*
				(−1.31)	(−1.42)	(−1.77)				(−1.03)	(−1.49)	(−1.90)
Constant	11.36***	2.49***	2.41***	−1.62			25.44***	−0.08	−0.89	−12.69***		
	(17.86)	(4.02)	(3.72)	(−0.93)			(17.94)	(−0.05)	(−0.50)	(−3.14)		
Fund Fixed-effects	No	No	No	No	Yes	No	No	No	No	No	Yes	No
Family Fixed-effects	No	No	No	No	No	Yes	No	No	No	No	Yes	No
Observations	123,565	123,565	123,565	123,565	123,565	123,565	123,565	123,565	123,565	123,565	123,565	123,565
Adjusted R^2	0.06	0.19	0.19	0.22	0.59	0.58	0.09	0.24	0.25	0.26	0.65	0.62

Table 4: The centrality premium

This table presents the results of the following linear regression model:

$$Return\ Gap_{i,t} = \beta \times Centrality_{i,t-1} + \gamma \times Covariates_{i,t-1}\ (+\alpha_i) + \theta_t + \varepsilon_{i,t}$$

where i indexes mutual funds and t indexes time in quarters. The dependent variable, $Return\ Gap_{i,t}$, is fund i's average return gap (in % per year) during quarter t. $Centrality_{i,t-1}$ is fund i's brokerage network centrality measured prior to quarter t. We use two measures of centrality: degree centrality and eigenvector centrality. $Covariates_{i,t-1}$ are a vector of fund and portfolio characteristics that include log of fund TNA, log of family TNA, turnover, expense ratio, and value-weighted averages of log of market capitalization, book-to-market ratio, and 12-month cumulative return (excluding the most recent month) of stock holdings, all measured at the end of quarter $t-1$. In some specifications, the regression includes fund or family fixed-effects (α_i). All regressions include time fixed-effects (θ_t). Standard errors are double-clustered by fund and time, and the resulting t-statistics are reported in parentheses. Statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

Dependent variable:	Return Gap (%)					
Centrality measure:	Degree Centrality			Eigenvector Centrality		
	(1)	(2)	(3)	(4)	(5)	(6)
Centrality	1.30***	1.07**	0.59	0.48***	0.43***	0.24*
	(3.98)	(2.46)	(1.57)	(4.16)	(2.81)	(1.86)
log(Fund TNA)	−0.11***	−0.39***	−0.14***	−0.11***	−0.39***	−0.14***
	(−3.74)	(−7.58)	(−5.42)	(−3.75)	(−7.60)	(−5.43)
log(Family TNA)	0.11***	0.06	0.01	0.11***	0.06	0.01
	(6.28)	(1.44)	(0.19)	(6.20)	(1.43)	(0.18)
Turnover (%)	0.0004	0.002**	0.0004	0.0004	0.002**	0.0003
	(0.41)	(2.53)	(0.42)	(0.41)	(2.54)	(0.42)
Expense Ratio (%)	0.01	0.38**	0.17	0.02	0.38**	0.17
	(0.08)	(2.37)	(1.18)	(0.09)	(2.41)	(1.20)
log(Market Capitalization)	−0.12	0.04	−0.07	−0.12	0.04	−0.07
	(−1.28)	(0.19)	(−0.76)	(−1.29)	(0.19)	(−0.76)
Book-to-market	−1.62***	−1.26	−1.35***	−1.62***	−1.27	−1.35***
	(−3.88)	(−1.57)	(−3.18)	(−3.89)	(−1.58)	(−3.18)
Momentum (%)	0.001	0.004	0.002	0.001	0.004	0.002
	(0.07)	(0.36)	(0.22)	(0.07)	(0.36)	(0.22)
Time Fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes
Fund Fixed-effects	No	Yes	No	No	Yex	No
Family Fixed-effects	No	No	Yes	No	No	Yes
Observations	123,565	123,565	123,565	123,565	123,565	123,565
Adjusted R^2	0.08	0.10	0.09	0.08	0.10	0.09

Table 5: The *conditional* centrality premium when funds are forced to trade

This table presents the results of the following linear regression model:

$$Return\ Gap_{i,t} = \delta \times Centrality_{i,t-1} \times \mathbb{1}(Outflow_{i,t} > 5\%) + \beta \times Centrality_{i,t-1} + \rho \times \mathbb{1}(Outflow_{i,t} > 5\%) + \gamma \times Covariates_{i,t-1}\ (+\alpha_i) + \theta_t + \varepsilon_{i,t}$$

where i indexes mutual funds and t indexes time in quarters. The dependent variable, $Return\ Gap_{i,t}$, is fund i's average return gap (in % per year) during quarter t. $Centrality_{i,t-1}$ is fund i's brokerage network centrality measured prior to quarter t. We use two measures of centrality: degree centrality and eigenvector centrality. $\mathbb{1}(Outflow_{i,t} > 5\%)$ is an indicator variable that takes the value of one if fund i's percentage fund flow during quarter t falls below -5% and zero otherwise. $Covariates_{i,t-1}$ are a vector of fund and portfolio characteristics that include log of fund TNA, log of family TNA, turnover, expense ratio, and value-weighted averages of log of market capitalization, book-to-market ratio, and 12-month cumulative return (excluding the most recent month) of stock holdings, all measured at the end of quarter $t-1$. In some specifications, the regression includes fund or family fixed-effects (α_i). All regressions include time fixed-effects (θ_t). Standard errors are double-clustered by fund and time, and the resulting t-statistics are reported in parentheses. Statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

Dependent variable:	Return Gap (%)					
Centrality measure:	Degree Centrality			Eigenvector Centrality		
	(1)	(2)	(3)	(4)	(5)	(6)
Centrality × $\mathbb{1}$(Outflow > 5%)	2.24***	2.15***	1.84***	0.88***	0.79***	0.74***
	(3.43)	(3.07)	(2.84)	(3.83)	(3.16)	(3.21)
Centrality	0.87***	0.65	0.23	0.31***	0.27*	0.10
	(2.63)	(1.51)	(0.61)	(2.69)	(1.82)	(0.76)
$\mathbb{1}$(Outflow > 5%)	−0.51***	−0.44***	−0.45***	−0.52***	−0.43***	−0.47***
	(−3.20)	(−2.79)	(−2.90)	(−3.33)	(−2.74)	(−3.04)
log(Fund TNA)	−0.11***	−0.39***	−0.14***	−0.11***	−0.39***	−0.14***
	(−3.79)	(−7.58)	(−5.47)	(−3.78)	(−7.58)	(−5.47)
log(Family TNA)	0.11***	0.06	0.01	0.11***	0.06	0.01
	(6.25)	(1.44)	(0.21)	(6.18)	(1.44)	(0.20)
Turnover (%)	0.0004	0.002**	0.0004	0.0004	0.002***	0.0004
	(0.48)	(2.57)	(0.47)	(0.48)	(2.58)	(0.47)
Expense Ratio (%)	0.01	0.38**	0.17	0.02	0.39**	0.17
	(0.09)	(2.41)	(1.19)	(0.11)	(2.45)	(1.21)
log(Market Capitalization)	−0.12	0.04	−0.07	−0.12	0.04	−0.07
	(−1.31)	(0.19)	(−0.78)	(−1.32)	(0.19)	(−0.78)
Book-to-market	−1.62***	−1.27	−1.35***	−1.62***	−1.28	−1.35***
	(−3.88)	(−1.58)	(−3.18)	(−3.90)	(−1.59)	(−3.19)
Momentum (%)	0.001	0.004	0.002	0.001	0.004	0.002
	(0.06)	(0.35)	(0.20)	(0.05)	(0.35)	(0.20)
Time Fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes
Fund Fixed-effects	No	Yes	No	No	Yex	No
Family Fixed-effects	No	No	Yes	No	No	Yes
Observations	123,565	123,565	123,565	123,565	123,565	123,565
Adjusted R^2	0.08	0.10	0.09	0.08	0.10	0.09

Table 6: The conditional centrality premium – fund family networks?

This table presents the results of the following linear regression model:

$$Return\ Gap_{i,t} = \delta \times Centrality_{i,t-1} \times \mathbb{1}(Outflow_{i,t} > 5\%) + \beta \times Centrality_{i,t-1} + \rho \times \mathbb{1}(Outflow_{i,t} > 5\%) + \gamma \times Covariates_{i,t-1} + \theta_t + \varepsilon_{i,t}$$

where i indexes mutual funds and t indexes time in quarters. We split our sample into three sub-samples by terciles based on the number of funds within the fund family. The dependent variable, $Return\ Gap_{i,t}$, is fund i's average return gap (in % per year) during quarter t. $Centrality_{i,t-1}$ is fund i's brokerage network centrality measured prior to quarter t. We use two measures of centrality: degree centrality and eigenvector centrality. $\mathbb{1}(Outflow_{i,t} > 5\%)$ is an indicator variable that takes the value of one if fund i's percentage fund flow during quarter t falls below −5% and zero otherwise. $Covariates_{i,t-1}$ are a vector of fund and portfolio characteristics that include log of fund TNA, log of family TNA, turnover, expense ratio, and value-weighted averages of log of market capitalization, book-to-market ratio, and 12-month cumulative return (excluding the most recent month) of stock holdings, all measured at the end of quarter $t-1$. All regressions include time fixed-effects (θ_t). Standard errors are double-clustered by fund and time, and the resulting t-statistics are reported in parentheses. Statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

Dependent variable:	Return Gap (%)					
Centrality measure:	Degree Centrality			Eigenvector Centrality		
Fund family size	Small	Mid	Large	Small	Mid	Large
	(1)	(2)	(3)	(4)	(5)	(6)
Centrality × $\mathbb{1}$(Outflow > 5%)	3.01**	1.86*	0.85	1.10**	0.91**	0.12
	(2.43)	(1.78)	(0.76)	(2.35)	(2.30)	(0.30)
Centrality	1.38**	1.17*	−0.59	0.49**	0.40*	−0.24
	(2.05)	(1.95)	(−1.09)	(2.00)	(1.91)	(−1.30)
$\mathbb{1}$(Outflow > 5%)	−0.74***	−0.39*	−0.06	−0.71***	−0.48**	0.06
	(−3.44)	(−1.74)	(−0.21)	(−3.41)	(−2.14)	(0.19)
log(Fund TNA)	−0.15***	−0.13***	−0.09***	−0.15***	−0.13***	−0.09***
	(−2.99)	(−4.12)	(−2.83)	(−3.01)	(−4.12)	(−2.79)
log(Family TNA)	0.18***	0.14***	0.26***	0.18***	0.14***	0.26***
	(4.39)	(2.77)	(5.61)	(4.43)	(2.68)	(5.74)
Turnover (%)	0.0003	0.001	0.0003	0.0003	0.001	0.0003
	(0.17)	(0.93)	(0.43)	(0.17)	(0.92)	(0.43)
Expense Ratio (%)	0.11	0.04	0.10	0.11	0.04	0.11
	(0.52)	(0.19)	(0.50)	(0.51)	(0.21)	(0.55)
log(Market Capitalization)	−0.23**	−0.14	0.06	−0.24**	−0.14	0.06
	(−2.32)	(−1.44)	(0.57)	(−2.33)	(−1.45)	(0.58)
Book-to-market	−2.00***	−1.73***	−1.17**	−2.01***	−1.74***	−1.17**
	(−3.76)	(−3.45)	(−2.38)	(−3.77)	(−3.46)	(−2.39)
Momentum (%)	−0.0001	0.001	−0.0001	−0.0002	0.001	−0.0000
	(−0.01)	(0.08)	(−0.01)	(−0.02)	(0.08)	(−0.005)
Time Fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	41,188	41,188	41,189	41,188	41,188	41,189
Adjusted R^2	0.08	0.08	0.08	0.08	0.08	0.08

Table 7: The conditional centrality premium for valuable clients

This table presents the results of the following linear regression model:

$$\begin{aligned} \mathit{Return\ Gap}_{i,t} &= \delta \times \mathit{Centrality}_{i,t-1} \times \mathbb{1}(\mathit{Outflow}_{i,t} > 5\%) \times \mathbb{1}(\mathit{High\ Value\ Client}_{i,t-1}) \\ &+ \beta_1 \times \mathit{Centrality}_{i,t-1} \times \mathbb{1}(\mathit{Outflow}_{i,t} > 5\%) + \beta_2 \times \mathit{Centrality}_{i,t-1} \times \mathbb{1}(\mathit{High\ Value\ Client}_{i,t-1}) \\ &+ \beta_3 \times \mathbb{1}(\mathit{Outflow}_{i,t} > 5\%) \times \mathbb{1}(\mathit{High\ Value\ Client}_{i,t-1}) + \rho_1 \times \mathit{Centrality}_{i,t-1} \\ &+ \rho_2 \times \mathbb{1}(\mathit{Outflow}_{i,t} > 5\%) + \rho_3 \times \mathbb{1}(\mathit{High\ Value\ Client}_{i,t-1}) + \gamma \times \mathit{Covariates}_{i,t-1}\ (+\alpha_i) + \theta_t + \varepsilon_{i,t} \end{aligned}$$

where i indexes mutual funds and t indexes time in quarters. The dependent variable, $\mathit{Return\ Gap}_{i,t}$, is fund i's average return gap (in % per year) during quarter t. $\mathit{Centrality}_{i,t-1}$ is fund i's brokerage network centrality measured prior to quarter t. We use two measures of centrality: degree centrality and eigenvector centrality. $\mathbb{1}(\mathit{Outflow}_{i,t} > 5\%)$ is an indicator variable that takes the value of one if fund i's percentage fund flow during quarter t falls below -5% and zero otherwise. $\mathbb{1}(\mathit{High\ Value\ Client}_{i,t-1})$ is an indicator variable that takes the value of one if fund i's estimated dollar volume during quarter $t-1$ belong to a top quartile and zero otherwise. $\mathit{Covariates}_{i,t-1}$ are a vector of fund and portfolio characteristics that include log of fund TNA, log of family TNA, turnover, expense ratio, and value-weighted averages of log of market capitalization, book-to-market ratio, and 12-month cumulative return (excluding the most recent month) of stock holdings, all measured at the end of quarter $t-1$. In some specifications, the regression includes fund or family fixed-effects (α_i). All regressions include time fixed-effects (θ_t). Standard errors are double-clustered by fund and time, and the resulting t-statistics are reported in parentheses. Statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

Dependent variable:	Return Gap (%)					
Centrality measure:	Degree Centrality			Eigenvector Centrality		
	(1)	(2)	(3)	(4)	(5)	(6)
Centrality × $\mathbb{1}$(Outflow > 5%)× $\mathbb{1}$(High Value Client)	2.53*	3.29**	2.97**	1.11*	1.32**	1.22**
	(1.75)	(2.18)	(2.06)	(1.87)	(2.12)	(2.05)
Centrality × $\mathbb{1}$(Outflow > 5%)	1.68**	1.41**	1.17*	0.65***	0.50*	0.47**
	(2.49)	(1.98)	(1.81)	(2.66)	(1.92)	(2.00)
Centrality × $\mathbb{1}$(High Value Client)	0.84	1.06	0.57	0.13	0.07	−0.05
	(1.11)	(1.24)	(0.81)	(0.47)	(0.21)	(−0.19)
$\mathbb{1}$(Outflow > 5%)× $\mathbb{1}$(High Value Client)	−0.44	−0.59	−0.51	−0.54	−0.64	−0.58
	(−1.21)	(−1.56)	(−1.42)	(−1.40)	(−1.58)	(−1.49)
Centrality	0.71**	0.37	0.10	0.28**	0.25	0.12
	(2.10)	(0.82)	(0.25)	(2.40)	(1.52)	(0.85)
$\mathbb{1}$(Outflow > 5%)	−0.43***	−0.34**	−0.36**	−0.44***	−0.32**	−0.37**
	(−2.75)	(−2.21)	(−2.39)	(−2.80)	(−2.10)	(−2.46)
$\mathbb{1}$(High Value Client)	−0.27	−0.20	−0.21	−0.16	−0.01	−0.05
	(−1.42)	(−0.96)	(−1.18)	(−0.86)	(−0.05)	(−0.31)
log(Fund TNA)	−0.10***	−0.39***	−0.14***	−0.10***	−0.39***	−0.13***
	(−3.23)	(−7.60)	(−4.70)	(−3.21)	(−7.59)	(−4.67)
log(Family TNA)	0.11***	0.06	0.01	0.11***	0.06	0.01
	(6.34)	(1.53)	(0.23)	(6.22)	(1.50)	(0.21)
Turnover (%)	0.001	0.002**	0.0005	0.001	0.002**	0.0005
	(0.59)	(2.49)	(0.55)	(0.60)	(2.50)	(0.56)
Expense Ratio (%)	0.01	0.38**	0.17	0.02	0.39**	0.18
	(0.08)	(2.42)	(1.22)	(0.11)	(2.46)	(1.25)
log(Market Capitalization)	−0.12	0.04	−0.07	−0.12	0.04	−0.07
	(−1.31)	(0.18)	(−0.77)	(−1.32)	(0.18)	(−0.77)
Book-to-market	−1.61***	−1.28	−1.34***	−1.62***	−1.28	−1.35***
	(−3.88)	(−1.59)	(−3.17)	(−3.90)	(−1.60)	(−3.20)
Momentum (%)	0.001	0.004	0.002	0.001	0.004	0.002
	(0.06)	(0.37)	(0.21)	(0.06)	(0.36)	(0.20)
Time Fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes
Fund Fixed-effects	No	Yes	No	No	Yex	No
Family Fixed-effects	No	No	Yes	No	No	Yes
Observations	123,565	123,565	123,565	123,565	123,565	123,565
Adjusted R^2	0.08	0.10	0.09	0.08	0.10	0.09

Table 8: The conditional centrality premium for relationship clients

This table presents the results of the following linear regression model:

$$
\begin{aligned}
\textit{Return Gap}_{i,t} = {} & \delta \times \textit{Centrality}_{i,t-1} \times \mathbb{1}(\textit{Outflow}_{i,t} > 5\%) \times \textit{Repeated Interaction}_{i,t-1} \\
& + \beta_1 \times \textit{Centrality}_{i,t-1} \times \mathbb{1}(\textit{Outflow}_{i,t} > 5\%) + \beta_2 \times \textit{Centrality}_{i,t-1} \times \textit{Repeated Interaction}_{i,t-1} \\
& + \beta_3 \times \mathbb{1}(\textit{Outflow}_{i,t} > 5\%) \times \textit{Repeated Interaction}_{i,t-1} + \rho_1 \times \textit{Centrality}_{i,t-1} \\
& + \rho_2 \times \mathbb{1}(\textit{Outflow}_{i,t} > 5\%) + \rho_3 \times \textit{Repeated Interaction}_{i,t-1} + \gamma \times \textit{Covariates}_{i,t-1} \; (+\alpha_i) + \theta_t + \varepsilon_{i,t}
\end{aligned}
$$

where i indexes mutual funds and t indexes time in quarters. The dependent variable, $\textit{Return Gap}_{i,t}$, is fund i's average return gap (in % per year) during quarter t. $\textit{Centrality}_{i,t-1}$ is fund i's brokerage network centrality measured prior to quarter t. We use two measures of centrality: degree centrality and eigenvector centrality. $\mathbb{1}(\textit{Outflow}_{i,t} > 5\%)$ is an indicator variable that takes the value of one if fund i's percentage fund flow during quarter t falls below -5%. $\textit{Repeated Interaction}_{i,t-1}$ is measured by the *cosine* similarity between fund i's brokerage commission vector available prior to quarter t and its brokerage commission vector one year prior to that. $\textit{Covariates}_{i,t-1}$ are a vector of fund and portfolio characteristics that include log of fund TNA, log of family TNA, turnover, expense ratio, and value-weighted averages of log of market capitalization, book-to-market ratio, and 12-month cumulative return (excluding the most recent month) of stock holdings, all measured at the end of quarter $t-1$. In some specifications, the regression includes fund or family fixed-effects (α_i). All regressions include time fixed-effects (θ_t). Standard errors are double-clustered by fund and time, and the resulting t-statistics are reported in parentheses. Statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

Dependent variable:	Return Gap (%)					
Centrality measure:	Degree Centrality			Eigenvector Centrality		
	(1)	(2)	(3)	(4)	(5)	(6)
Centrality × $\mathbb{1}$(Outflow > 5%)× Repeated Interaction	4.04**	4.63**	4.83***	1.41**	1.58**	1.68***
	(2.28)	(2.35)	(2.71)	(2.15)	(2.22)	(2.58)
Centrality × $\mathbb{1}$(Outflow > 5%)	−0.29	−0.73	−1.20	0.005	−0.19	−0.32
	(−0.21)	(−0.51)	(−0.88)	(0.01)	(−0.35)	(−0.63)
Centrality × Repeated Interaction	−0.25	−0.21	−0.15	−0.04	−0.004	−0.001
	(−0.31)	(−0.24)	(−0.17)	(−0.11)	(−0.01)	(−0.004)
$\mathbb{1}$(Outflow > 5%)× Repeated Interaction	−0.87**	−1.03***	−1.01***	−0.83**	−0.97***	−0.96***
	(−2.45)	(−2.69)	(−2.84)	(−2.39)	(−2.63)	(−2.81)
Centrality	0.99*	0.79	0.36	0.32	0.28	0.12
	(1.68)	(1.31)	(0.61)	(1.42)	(1.26)	(0.52)
$\mathbb{1}$(Outflow > 5%)	0.02	0.18	0.16	−0.02	0.15	0.11
	(0.07)	(0.64)	(0.61)	(−0.07)	(0.55)	(0.43)
Repeated Interaction	0.16	0.06	0.01	0.14	0.04	−0.01
	(0.92)	(0.36)	(0.06)	(0.81)	(0.21)	(−0.05)
log(Fund TNA)	−0.11***	−0.39***	−0.14***	−0.11***	−0.39***	−0.14***
	(−3.80)	(−7.57)	(−5.33)	(−3.79)	(−7.57)	(−5.33)
log(Family TNA)	0.11***	0.06	0.01	0.11***	0.06	0.01
	(6.24)	(1.44)	(0.20)	(6.17)	(1.44)	(0.20)
Turnover (%)	0.0004	0.002***	0.0004	0.0004	0.002***	0.0004
	(0.47)	(2.58)	(0.48)	(0.47)	(2.59)	(0.47)
Expense Ratio (%)	0.01	0.38**	0.17	0.02	0.39**	0.17
	(0.07)	(2.40)	(1.18)	(0.10)	(2.45)	(1.20)
log(Market Capitalization)	−0.12	0.04	−0.07	−0.12	0.04	−0.07
	(−1.31)	(0.19)	(−0.78)	(−1.32)	(0.19)	(−0.79)
Book-to-market	−1.62***	−1.28	−1.34***	−1.63***	−1.28	−1.35***
	(−3.89)	(−1.59)	(−3.18)	(−3.91)	(−1.60)	(−3.19)
Momentum (%)	0.001	0.004	0.002	0.001	0.004	0.002
	(0.06)	(0.35)	(0.20)	(0.05)	(0.35)	(0.20)
Time Fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes
Fund Fixed-effects	No	Yes	No	No	Yex	No
Family Fixed-effects	No	No	Yes	No	No	Yes
Observations	123,565	123,565	123,565	123,565	123,565	123,565
Adjusted R^2	0.08	0.10	0.09	0.08	0.10	0.09

Table 9: The centrality premium – trading costs channel

This table presents the results of the following linear regression model:

$$Trading\ Costs_{i,t} = \beta \times Centrality_{i,t-1} + \gamma \times Covariates_{i,t-1} + \theta_t + \varepsilon_{i,t}$$

where i indexes mutual funds and t indexes time in quarters. The dependent variable, $Trading\ Costs_{i,t}$, is fund i's average trading costs (in % per year) during quarter t. We use three measures of trading costs: implicit, explicit, and total trading costs. $Centrality_{i,t-1}$ is fund i's brokerage network centrality measured prior to quarter t. We use two measures of centrality: degree centrality and eigenvector centrality. $Covariates_{i,t-1}$ are a vector of fund and portfolio characteristics that include log of fund TNA, log of family TNA, turnover, expense ratio, and value-weighted averages of log of market capitalization, book-to-market ratio, and 12-month cumulative return (excluding the most recent month) of stock holdings, all measured at the end of quarter $t-1$. All regressions include time fixed-effects (θ_t). Standard errors are double-clustered by fund and time, and the resulting t-statistics are reported in parentheses. Statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

Centrality measure:	Degree Centrality			Eigenvector Centrality		
Dependent variable:	Trading Costs (%)			Trading Costs (%)		
	Implicit	Explicit	Total	Implicit	Explicit	Total
	(1)	(2)	(3)	(4)	(5)	(6)
Centrality	-0.77^{*}	-0.26	-1.04^{*}	-0.29^{*}	-0.10	-0.39^{*}
	(-1.91)	(-1.51)	(-1.89)	(-1.88)	(-1.47)	(-1.86)
log(Fund TNA)	-0.02	-0.02^{**}	-0.04	-0.02	-0.02^{**}	-0.04
	(-0.84)	(-2.19)	(-1.33)	(-0.84)	(-2.20)	(-1.33)
log(Family TNA)	-0.04^{***}	-0.02^{**}	-0.06^{***}	-0.04^{***}	-0.02^{**}	-0.06^{***}
	(-2.87)	(-2.48)	(-3.07)	(-2.92)	(-2.52)	(-3.13)
Turnover (%)	0.004^{***}	0.002^{***}	0.01^{***}	0.004^{***}	0.002^{***}	0.01^{***}
	(5.30)	(6.63)	(5.85)	(5.29)	(6.63)	(5.84)
Expense Ratio (%)	0.08	0.06^{*}	0.14	0.07	0.06^{*}	0.13
	(1.12)	(1.79)	(1.46)	(1.08)	(1.77)	(1.42)
log(Market Capitalization)	-0.09^{**}	-0.05^{***}	-0.13^{***}	-0.09^{**}	-0.05^{***}	-0.13^{***}
	(-2.33)	(-3.27)	(-2.73)	(-2.32)	(-3.26)	(-2.72)
Book-to-market	-0.61^{***}	-0.06	-0.66^{***}	-0.60^{***}	-0.06	-0.66^{***}
	(-3.33)	(-0.71)	(-2.79)	(-3.31)	(-0.69)	(-2.77)
Momentum (%)	0.01^{***}	-0.001	0.01^{**}	0.01^{***}	-0.001	0.01^{**}
	(2.99)	(-0.89)	(1.98)	(3.00)	(-0.88)	(1.98)
Time Fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7,791	7,791	7,791	7,791	7,791	7,791
Adjusted R^2	0.12	0.13	0.13	0.12	0.13	0.13

Table 10: The *conditional* centrality premium – trading costs channel

This table presents the results of the following linear regression model:

$$\begin{aligned} \textit{Trading Costs}_{i,t} &= \delta \times \textit{Centrality}_{i,t-1} \times \mathbb{1}(\textit{Outflow}_{i,t} > 5\%) + \beta \times \textit{Centrality}_{i,t-1} \\ &+ \rho \times \mathbb{1}(\textit{Outflow}_{i,t} > 5\%) + \gamma \times \textit{Covariates}_{i,t-1} + \theta_t + \varepsilon_{i,t} \end{aligned}$$

where i indexes mutual funds and t indexes time in quarters. The dependent variable, $\textit{Trading Costs}_{i,t}$, is fund i's average trading costs (in % per year) during quarter t. We use three measures of trading costs: implicit, explicit, and total trading costs. $\textit{Centrality}_{i,t-1}$ is fund i's brokerage network centrality measured prior to quarter t. We use two measures of centrality: degree centrality and eigenvector centrality. $\mathbb{1}(\textit{Outflow}_{i,t} > 5\%)$ is an indicator variable that takes the value of one if fund i's percentage fund flow during quarter t falls below -5% and zero otherwise. $\textit{Covariates}_{i,t-1}$ are a vector of fund and portfolio characteristics that include log of fund TNA, log of family TNA, turnover, expense ratio, and value-weighted averages of log of market capitalization, book-to-market ratio, and 12-month cumulative return (excluding the most recent month) of stock holdings, all measured at the end of quarter $t-1$. All regressions include time fixed-effects (θ_t). Standard errors are double-clustered by fund and time, and the resulting t-statistics are reported in parentheses. Statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

Centrality measure:	Degree Centrality			Eigenvector Centrality		
Dependent variable:	Trading Costs (%)			Trading Costs (%)		
	Implicit	Explicit	Total	Implicit	Explicit	Total
	(1)	(2)	(3)	(4)	(5)	(6)
Centrality × $\mathbb{1}$(Outflow > 5%)	-1.30^{**}	-0.74^{***}	-2.04^{**}	-0.63^{**}	-0.35^{***}	-0.99^{***}
	(-2.26)	(-2.61)	(-2.53)	(-2.44)	(-2.78)	(-2.70)
Centrality	-0.58	-0.15	-0.73	-0.20	-0.04	-0.24
	(-1.41)	(-0.85)	(-1.31)	(-1.26)	(-0.62)	(-1.13)
$\mathbb{1}$(Outflow > 5%)	0.41^{**}	0.23^{***}	0.64^{***}	0.48^{**}	0.27^{***}	0.74^{***}
	(2.34)	(2.58)	(2.58)	(2.44)	(2.75)	(2.69)
log(Fund TNA)	-0.02	-0.02^{**}	-0.04	-0.02	-0.02^{**}	-0.04
	(-0.80)	(-2.17)	(-1.29)	(-0.79)	(-2.16)	(-1.28)
log(Family TNA)	-0.04^{***}	-0.02^{**}	-0.06^{***}	-0.04^{***}	-0.02^{**}	-0.06^{***}
	(-2.84)	(-2.44)	(-3.03)	(-2.91)	(-2.49)	(-3.12)
Turnover (%)	0.004^{***}	0.002^{***}	0.01^{***}	0.004^{***}	0.002^{***}	0.01^{***}
	(5.29)	(6.62)	(5.84)	(5.29)	(6.64)	(5.85)
Expense Ratio (%)	0.07	0.06^{*}	0.13	0.07	0.05^{*}	0.12
	(1.05)	(1.70)	(1.38)	(1.00)	(1.66)	(1.33)
log(Market Capitalization)	-0.08^{**}	-0.05^{***}	-0.13^{***}	-0.08^{**}	-0.05^{***}	-0.13^{***}
	(-2.29)	(-3.23)	(-2.69)	(-2.28)	(-3.21)	(-2.68)
Book-to-market	-0.61^{***}	-0.06	-0.67^{***}	-0.60^{***}	-0.06	-0.66^{***}
	(-3.33)	(-0.71)	(-2.79)	(-3.29)	(-0.68)	(-2.75)
Momentum (%)	0.01^{***}	-0.001	0.01^{**}	0.01^{***}	-0.001	0.01^{**}
	(3.11)	(-0.74)	(2.12)	(3.13)	(-0.70)	(2.15)
Time Fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7,791	7,791	7,791	7,791	7,791	7,791
Adjusted R^2	0.12	0.13	0.14	0.12	0.13	0.14

Table 11: The Centrality Premium – Interim Trading Performance Channel

This table presents the results of the following linear regression model:

$$Interim\ Return_{i,t}\ (Return\ Gap_{i,t}) = \beta \times Centrality_{i,t-1} + \gamma \times Covariates_{i,t-1} + \theta_t + \varepsilon_{i,t}$$

where i indexes mutual funds and t indexes time in quarters. The dependent variable, $Interim\ Return_{i,t}$, is fund i's average interim return (in % per year) during quarter t. For a comparison, we use $Return\ Gap_{i,t}$ as the dependent variable side-by-side. $Centrality_{i,t-1}$ is fund i's brokerage network centrality measured prior to quarter t. We use two measures of centrality: degree centrality and eigenvector centrality. $Covariates_{i,t-1}$ are a vector of fund and portfolio characteristics that include log of fund TNA, log of family TNA, turnover, expense ratio, and value-weighted averages of log of market capitalization, book-to-market ratio, and 12-month cumulative return (excluding the most recent month) of stock holdings, all measured at the end of quarter $t-1$. All regressions include time fixed-effects (θ_t). Standard errors are double-clustered by fund and time, and the resulting t-statistics are reported in parentheses. Statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

Centrality measure:	Degree Centrality		Eigenvector Centrality	
Dependent variable:	Return Gap (%)	Interim Return (%)	Return Gap (%)	Interim Return (%)
	(1)	(2)	(3)	(4)
Centrality	2.53**	−0.21	1.18**	0.13
	(2.00)	(−0.21)	(2.31)	(0.31)
log(Fund TNA)	−0.07	−0.08	−0.07	−0.08
	(−0.93)	(−1.19)	(−0.96)	(−1.21)
log(Family TNA)	0.14**	0.16***	0.13**	0.15***
	(2.51)	(2.98)	(2.35)	(2.85)
Turnover (%)	0.002	0.001	0.002	0.001
	(0.99)	(0.32)	(0.95)	(0.29)
Expense Ratio (%)	0.06	−0.09	0.04	−0.12
	(0.21)	(−0.37)	(0.12)	(−0.49)
log(Market Capitalization)	−0.01	−0.07	−0.01	−0.07
	(−0.04)	(−0.61)	(−0.08)	(−0.65)
Book-to-market	0.06	−0.21	0.06	−0.20
	(0.06)	(−0.23)	(0.06)	(−0.22)
Momentum	−0.001	−0.01	−0.001	−0.01
	(−0.08)	(−0.53)	(−0.08)	(−0.51)
Time Fixed-effects	Yes	Yes	Yes	Yes
Observations	7,791	7,791	7,791	7,791
Adjusted R^2	0.07	0.01	0.07	0.01

Appendix

Table A1: Sample of CRSP-Thomson-NSAR Matched Funds

This table reports the total number and aggregate total net assets (TNA) of our CRSP–Thomson–NSAR matched funds in Panel A and the same statistics for our CRSP–Thomson–NSAR–ANcerno matched funds in Panel B at the end of June and December of each year.

Panel A: CRSP–Thomson–NSAR matched sample

Year	June		December	
	Total Number of Funds	Aggregate TNA ($ billion)	Total Number of Funds	Aggregate TNA ($ billion)
1993			128	80.3
1994	337	216.0	404	264.1
1995	521	413.6	608	511.7
1996	569	559.2	759	729.7
1997	836	831.0	835	980.1
1998	894	1,191.8	899	1,226.9
1999	1,002	1,383.8	1,069	1,511.2
2000	1,229	1,836.3	1,163	1,487.4
2001	1,323	1,555.1	1,373	1,539.4
2002	1,435	1,363.0	1,467	1,275.9
2003	1,574	1,562.9	1,649	1,789.8
2004	1,638	1,932.0	1,678	2,091.1
2005	1,684	2,083.8	1,695	2,228.3
2006	1,672	2,297.0	1,616	2,394.5
2007	1,747	2,628.8	1,780	2,560.7
2008	1,817	2,264.9	1,961	1,679.7
2009	1,687	1,732.2	1,862	2,149.9
2010	1,801	1,953.8	1,652	2,275.9
2011	1,680	2,439.9	1,713	2,285.9
2012	1,691	2,398.7	1,622	2,494.3
2013	1,591	2,711.9	1,560	3,220.1
2014	1,536	3,321.8	1,506	3,285.9
2015	1,449	3,276.3	1,415	3,100.2
2016	1,365	3,003.9		

Table A1–*Continued*

Panel B: CRSP–Thomson–NSAR–ANcerno matched sample

Year	Second Quarter		Fourth Quarter	
	Total Number of Funds	Aggregate TNA ($ billion)	Total Number of Funds	Aggregate TNA ($ billion)
1998			55	161.5
1999	79	280.1	68	228.4
2000	98	367.9	103	282.0
2001	118	287.0	138	352.1
2002	124	242.4	158	406.4
2003	191	478.0	187	525.2
2004	200	538.7	228	637.9
2005	205	553.8	228	700.7
2006	230	698.9	199	684.3
2007	169	719.0	192	724.8
2008	185	617.7	193	459.3
2009	180	520.3	202	587.4
2010	112	311.9	99	295.7

Table A2: The centrality premium – data truncation bias?

This table presents the results of the following linear regression model:

$$Return\ Gap_{i,t} = \beta \times Centrality_{i,t-1} + \gamma \times Covariates_{i,t-1} + \theta_t + \varepsilon_{i,t}$$

where i indexes mutual funds and t indexes time in quarters. The dependent variable, $Return\ Gap_{i,t}$, is fund i's average return gap (in % per year) during quarter t. $Centrality_{i,t-1}$ is fund i's brokerage network centrality measured prior to quarter t. We use two measures of centrality: degree centrality and eigenvector centrality. Networks are constructed through top 7, 5, or 3 brokers, instead of the top 10 brokers as in our main analysis. $Covariates_{i,t-1}$ are a vector of fund and portfolio characteristics that include log of fund TNA, log of family TNA, turnover, expense ratio, and value-weighted averages of log of market capitalization, book-to-market ratio, and 12-month cumulative return (excluding the most recent month) of stock holdings, all measured at the end of quarter $t-1$. All regressions include time fixed-effects (θ_t). Standard errors are double-clustered by fund and time, and the resulting t-statistics are reported in parentheses. Statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

Dependent variable:	Return Gap (%)					
Centrality measure:	Degree Centrality			Eigenvector Centrality		
Top n broker connections	Top 7	Top 5	Top 3	Top 7	Top 5	Top 3
	(1)	(2)	(3)	(4)	(5)	(6)
Centrality	1.34***	1.37***	1.22***	0.45***	0.38***	0.24***
	(3.86)	(3.72)	(2.94)	(3.97)	(3.72)	(2.63)
log(Fund TNA)	−0.11***	−0.11***	−0.11***	−0.11***	−0.11***	−0.11***
	(−3.73)	(−3.71)	(−3.69)	(−3.73)	(−3.71)	(−3.69)
log(Family TNA)	0.11***	0.11***	0.12***	0.11***	0.11***	0.12***
	(6.45)	(6.61)	(6.91)	(6.40)	(6.63)	(6.97)
Turnover (%)	0.0004	0.0004	0.0004	0.0004	0.0004	0.0005
	(0.42)	(0.45)	(0.49)	(0.43)	(0.46)	(0.50)
Expense Ratio (%)	0.01	0.01	0.01	0.02	0.02	0.01
	(0.08)	(0.09)	(0.08)	(0.09)	(0.09)	(0.08)
log(Market Capitalization)	−0.12	−0.12	−0.11	−0.12	−0.12	−0.11
	(−1.28)	(−1.27)	(−1.25)	(−1.29)	(−1.28)	(−1.25)
Book-to-market	−1.62***	−1.63***	−1.63***	−1.63***	−1.63***	−1.64***
	(−3.89)	(−3.90)	(−3.92)	(−3.90)	(−3.91)	(−3.93)
Momentum (%)	0.001	0.001	0.001	0.001	0.001	0.001
	(0.07)	(0.07)	(0.06)	(0.06)	(0.06)	(0.06)
Time Fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	123,565	123,565	123,565	123,565	123,565	123,565
Adjusted R^2	0.08	0.08	0.08	0.08	0.08	0.08

Table A3: The conditional centrality premium – robustness checks

This table presents the results of the following linear regression model:

$$\begin{aligned} \textit{Return Gap}_{i,t} &= \delta \times \textit{Centrality}_{i,t-1} \times \mathbb{1}(\textit{Outflow}_{i,t} > x\%) + \beta \times \textit{Centrality}_{i,t-1} \\ &+ \rho \times \mathbb{1}(\textit{Outflow}_{i,t} > x\%) + \gamma \times \textit{Covariates}_{i,t-1} \ (+\alpha_i) + \theta_t + \varepsilon_{i,t} \end{aligned}$$

where i indexes mutual funds and t indexes time in quarters. The dependent variable, $\textit{Return Gap}_{i,t}$, is fund i's average return gap (in % per year) during quarter t. The interaction term, $\mathbb{1}(\textit{Outflow}_{i,t} > x\%)$, is an indicator variable that takes the value of one if fund i's percentage fund flow during quarter t falls below -7% in Panel A (or -3% in Panel B) and zero otherwise. $\textit{Centrality}_{i,t-1}$ is fund i's brokerage network centrality measured prior to quarter t. We use two measures of centrality: degree centrality and eigenvector centrality. $\textit{Covariates}_{i,t-1}$ are a vector of fund and portfolio characteristics that include log of fund TNA, log of family TNA, turnover, expense ratio, and value-weighted averages of log of market capitalization, book-to-market ratio, and 12-month cumulative return (excluding the most recent month) of stock holdings, all measured at the end of quarter $t-1$. In some specifications, the regression includes fund or family fixed-effects (α_i). All regressions include time fixed-effects (θ_t). Standard errors are double-clustered by fund and time, and the resulting t-statistics are reported in parentheses. Statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

Panel A: A higher cutoff to define large outflow events

Dependent variable:	Return Gap (%)					
Centrality measure:	Degree Centrality			Eigenvector Centrality		
	(1)	(2)	(3)	(4)	(5)	(6)
Centrality × 𝟙(Outflow > 7%)	2.66***	2.52***	2.24***	1.04***	0.95***	0.89***
	(3.39)	(3.10)	(2.86)	(3.77)	(3.23)	(3.20)
Centrality	0.97***	0.76*	0.32	0.35***	0.31**	0.14
	(3.02)	(1.81)	(0.87)	(3.14)	(2.13)	(1.07)
𝟙(Outflow > 7%)	−0.57***	−0.49***	−0.51***	−0.57***	−0.48***	−0.53***
	(−3.20)	(−2.80)	(−2.94)	(−3.32)	(−2.79)	(−3.09)
log(Fund TNA)	−0.11***	−0.39***	−0.14***	−0.11***	−0.39***	−0.14***
	(−3.79)	(−7.57)	(−5.48)	(−3.79)	(−7.57)	(−5.49)
log(Family TNA)	0.11***	0.06	0.01	0.11***	0.06	0.01
	(6.23)	(1.42)	(0.20)	(6.17)	(1.42)	(0.19)
Turnover (%)	0.0004	0.002**	0.0004	0.0004	0.002**	0.0004
	(0.46)	(2.57)	(0.45)	(0.47)	(2.57)	(0.45)
Expense Ratio (%)	0.01	0.38**	0.17	0.02	0.39**	0.17
	(0.09)	(2.41)	(1.19)	(0.11)	(2.45)	(1.21)
log(Market Capitalization)	−0.12	0.04	−0.07	−0.12	0.04	−0.07
	(−1.30)	(0.19)	(−0.77)	(−1.31)	(0.19)	(−0.78)
Book-to-market	−1.62***	−1.27	−1.35***	−1.62***	−1.27	−1.35***
	(−3.89)	(−1.58)	(−3.18)	(−3.90)	(−1.59)	(−3.19)
Momentum (%)	0.001	0.004	0.002	0.001	0.004	0.002
	(0.06)	(0.36)	(0.21)	(0.06)	(0.35)	(0.20)
Time Fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes
Fund Fixed-effects	No	Yes	No	No	Yex	No
Family Fixed-effects	No	No	Yes	No	No	Yes
Observations	123,565	123,565	123,565	123,565	123,565	123,565
Adjusted R^2	0.08	0.10	0.09	0.08	0.10	0.09

Table A3–*Continued*

Panel B: A lower cutoff to define large outflow events

Dependent variable:	Return Gap (%)					
Centrality measure:	Degree Centrality			Eigenvector Centrality		
	(1)	(2)	(3)	(4)	(5)	(6)
Centrality × $\mathbb{1}$(Outflow > 3%)	1.37**	1.50**	1.10*	0.53**	0.52**	0.42*
	(2.25)	(2.26)	(1.76)	(2.46)	(2.21)	(1.92)
Centrality	0.88**	0.59	0.24	0.31***	0.26*	0.11
	(2.46)	(1.34)	(0.59)	(2.59)	(1.72)	(0.80)
$\mathbb{1}$(Outflow > 3%)	−0.32**	−0.31**	−0.28*	−0.32**	−0.28*	−0.28**
	(−2.15)	(−2.04)	(−1.92)	(−2.23)	(−1.92)	(−1.98)
log(Fund TNA)	−0.11***	−0.39***	−0.14***	−0.11***	−0.39***	−0.14***
	(−3.75)	(−7.61)	(−5.42)	(−3.74)	(−7.60)	(−5.42)
log(Family TNA)	0.11***	0.06	0.01	0.11***	0.06	0.01
	(6.26)	(1.44)	(0.19)	(6.19)	(1.44)	(0.19)
Turnover (%)	0.0004	0.002**	0.0004	0.0004	0.002**	0.0004
	(0.47)	(2.55)	(0.46)	(0.47)	(2.56)	(0.46)
Expense Ratio (%)	0.02	0.38**	0.17	0.02	0.39**	0.17
	(0.10)	(2.39)	(1.21)	(0.12)	(2.43)	(1.23)
log(Market Capitalization)	−0.12	0.04	−0.07	−0.12	0.04	−0.07
	(−1.30)	(0.19)	(−0.77)	(−1.31)	(0.19)	(−0.77)
Book-to-market	−1.62***	−1.26	−1.35***	−1.63***	−1.27	−1.35***
	(−3.89)	(−1.58)	(−3.18)	(−3.91)	(−1.59)	(−3.19)
Momentum (%)	0.001	0.004	0.002	0.0005	0.004	0.002
	(0.06)	(0.36)	(0.20)	(0.05)	(0.35)	(0.20)
Time Fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes
Fund Fixed-effects	No	Yes	No	No	Yex	No
Family Fixed-effects	No	No	Yes	No	No	Yes
Observations	123,565	123,565	123,565	123,565	123,565	123,565
Adjusted R^2	0.08	0.10	0.09	0.08	0.10	0.09