The Externalities of Crowded Trades

JESSE BLOCHER*

March 15, 2013

ABSTRACT

This paper documents fund flow externalities across mutual funds associated by similar asset holdings. With a network specification of embedded instrumental variables to control for correlated shocks to associated funds, I find that mutual fund managers who ignore these spillover effects may underestimate fund flows by approximately 20%. Peer Flows, (flows to and from other mutual funds funds with similar holdings) account for 2% of mutual fund quarterly return after controlling for various factor models, which is subsequently and completely reversed in the following year. I provide evidence that this overshoot is the result of spillover among connected mutual funds. This effect seems to be the result of crowded trades since similarity is transient, concentrated holdings drive the mutual fund similarity measure, and the initial overshoot lasts only one quarter. Fund similarity may be measurable ahead of time and thus be of interest to fund managers and regulators alike.

*This paper is based on my dissertation from the UNC - Chapel Hill's Kenan-Flagler Business School and was previously circulated as "Contagious Capital: A Network Analysis of Interconnected Intermediaries". Special thanks to my committee - Jennifer Conrad, Adam Reed, Joey Engelberg, Greg Brown, Pab Jotikasthira and Peter Mucha - for timely and helpful advice on this project. I am also grateful to Rick Sias (FRA discussant), Ethan Cohen-Cole, Tim Conley, Matt Bothner, Eric Ghysels, Chris Lundblad, Ed Van Wesep and Matt Ringgenberg for useful comments. Thanks also to seminar participants at University of Michigan, Boston College, Vanderbilt University, University of Georgia, Case Western Reserve University, Tulane University, Federal Reserve Bank-New York, and Federal Reserve Bank-Boston for useful feedback, as well as conference attendees at the 2011 Washington University at St. Louis Corporate Finance Conference PhD poster session and the 2011 Financial Research Association meeting in Las Vegas, NV. Blocher is from the Owen Graduate School of Management, Vanderbilt University, 401 21st Avenue South, Nashville, TN 37203, Contact info: jesse.blocher@owen.vanderbilt.edu or www.owen.vanderbilt.edu/blocher.

In his presidential address, Stein (2009) develops a theoretical model of “crowding” among sophisticated investors, and shows how externalities generated by crowded trades may adversely affect market efficiency. This paper provides empirical evidence of the externality created by fund flows among mutual fund managers who hold similar securities in their portfolios, that is, “crowded trades.”

Crowded trade externalities can be illustrated through a real estate example. Suppose some investors buy homes in a well-located, but dilapidated neighborhood. Suppose further that more investors do the same, and soon commercial real estate developers follow, buying old lots and replacing them with new restaurants and retail stores. Each subsequent investor generates positive externalities for previous investors due to similar location. As a result, real estate prices quickly appreciate. Furthermore, once this reinforcing investment cycle has begun, there is no natural ceiling apart from potential buyer’s beliefs that the neighborhood has become too expensive.

From this real estate example, we can see how an externality transmits across parcels of land that are close geographically. In financial markets, the externality transmits across mutual funds that are close in security space (i.e., funds that hold similar securities in their portfolios). Suppose investors deposit cash into a mutual fund, and the mutual fund uses that cash primarily to purchase more of the securities already held in its portfolio (Lou, 2012). These purchases induce price pressure on those securities (Coval and Stafford, 2007, Jotikasthira, Lundblad, and Ramadorai, 2012), benefiting other mutual funds holding the same securities. As a result of positive returns, the other funds are likely to receive inflows (Chevalier and Ellison, 1997, Sirri and Tufano, 1998), and purchase those securities again, repeating the cycle.

Stein (2009) notes that this crowded trade externality can only exist when a trading strategy has no natural limit which is the case for most long-only investment strategies. As a contrast, in pairs trading or spread trading prices converge and thus the arbitrage naturally disappears as more investors exploit it. But when a stock experiences post-earnings-announcement-drift, for example, investors recognize the price may drift upward after an earnings surprise, but by how much?[1] If more traders attempt to exploit the drift, the price will simply rise with the increased demand as long as enough traders continue to purchase the stock. The only countervailing force to this price appreciation is short sellers subject to limits to arbitrage (Shleifer and Vishny, 1997). Given the short selling and portfolio constraints of mutual funds, as well as the likelihood that mutual funds engage in herding behavior (e.g. Nofsinger and Sias, 1999, Sias, 2004, Choi and Sias, 2009), mutual funds are a natural first place to investigate externalities among crowded trades.[2,3]

Compared to the analysis of disconnected portfolio managers common in the literature, I find that coefficient estimates of common predictors of fund flows increase by approximately 90% due to fund flow externalities when estimated within my network specification.[4] The economic impact of this spillover externality is 20% of average fund flow for a one standard deviation move in *Peer Flow* (flows

[1]Many strategies have this same property. Examples include buying the stocks of firms with low values of accruals (Sloan, 1996), equity issues (Daniel and Titman, 2006), or asset growth (Cooper, Gulen, and Schill, 2008) as well as buying stocks that are expected to be added to a widely tracked index (Petajisto, 2011). Stein (2009) describes these strategies as having demand functions which are non-decreasing in price.

[2]Herding and crowded trades are related but not the same. Herding is when agents follow each other's behavior. A crowded trade is the result of herding in financial markets where participants end up with similar positions. Rationally, managers may herd on correlated private information (Froot, Scharfstein, and Stein, 1992). Other explanations include reputation-based herding (Scharfstein and Stein, 1990) and information cascades (Bikhchandani, Hirshleifer, and Welch, 1992). I only investigate the impact of the choices without taking a stand on how or why they get into these positions.

[3]Hedge funds have been singled out in the popular press for crowded trades but Reca, Sias, and Turtle (2011) find that hedge funds herd much less than mutual funds.

[4]By not accounting for interconnections, existing literature predicting fund flows implicitly assumes each fund to be independent (e.g. Sirri and Tufano, 1998).

to and from neighboring mutual funds with similar holdings). Peer Flows are additionally associated with 2% quarterly fund return on average after controlling for either a Carhart four-factor model (Carhart, 1997) or a DGTW return (Daniel, Grinblatt, Titman, and Wermers, 1997). This gain is completely reversed over the subsequent four quarters, indicating that Peer Flows are uninformative and only temporarily influence fund returns.

Measuring the magnitude of externalities among market participants has remained elusive thus far due to difficulties in identification. There are two main identification issues. First, a specification should isolate a spillover process distinct from a correlated shock to mutual funds already established as similar and thus susceptible to correlated shocks. Second, one must identify the channel of spillover; that is, what is the network that connects market participants? To address the first issue, I analyze the entire network of mutual funds in an econometric framework employing a mutual fund's two-step neighbor (his neighbor's neighbor) as an instrument for the neighbor's effect on that fund (Bramoullé, Djebbari, and Fortin, 2009).

I address the second identification problem by investigating the effect of Peer Flows on the holdings that I propose create the network linkages. When I isolate the top 10 holdings (by dollar value) of fund with high Peer Outflows, the excess DGTW returns of the portfolio of stocks is -1.42% per quarter compared to a baseline group with zero net Peer Flows. Similarly, a portfolio of the top 10 holdings of funds with high Peer Inflows returns 0.86% more per quarter than the baseline group. Both differences are statistically significant at the 1% level, as is their difference. This intermediate finding validates the first part of the reinforcing loop described above, that inflows leading to purchases of existing holdings induce upward price pressure on those holdings which increases the returns of 'peer' funds holding the same securities.

To determine these 'peer' or 'neighbor' funds, I employ a similarity measure almost identical to that used in Cohen, Coval, and Pastor (2005) and frequently employed in ecology and social network analysis (Legendre and Legendre, 2012, Wasserman and Faust, 1994) whereby two mutual funds are more similar to the extent that they have similar portfolio weightings, emphasizing concentrated holdings. A panel auto-regression shows that this similarity measure features a one quarter lag coefficient of just 0.4 with no relation after two years, which is consistent with my result finding a one quarter return effect which is subsequently reversed. This low autocorrelation also indicates that the connectivity among mutual funds is not related to style investing, common benchmarks, or index funds, as those factors should show more persistence.[5] The transience indicated by a low autocorrelation coefficient instead indicates short-term connectivity (i.e., crowded trades) is of primary importance.

My results emphasize an important aspect of managing mutual funds: what does a portfolio manager do with inflows in light of potentially diminishing investment opportunities due to constraints?[6] Mutual fund managers obtain inflows which they invest, at least partly, in their existing portfolio *even if they no longer believe those stocks to be undervalued.* If other managers do the same, all of those holding similar securities benefit from this behavior and appear to be savvy investors – as others buy the same stocks, prices rise and returns are positive, reinforcing the behavior even though it may not be supported by fundamentals. But with no clear indicator of fundamental value, there is no natural ceiling or floor to

[5] I also separately control for mutual fund style explicitly, this is just indirect evidence that style investing is not driving my result.

[6] Almazan, Brown, Carlson, and Chapman (2004) examine mutual fund manager constraints in light of optimal contract theory and performance, but do not address this question. Mutual funds are prohibited from short selling and also constrained to choose only equity securities in a certain category or strategy (large cap, technology, etc.) as defined in the fund's prospectus.

price movements. This mechanism of rational behavior inducing price overshoot is a similar but distinct mechanism from the rational herding modeled in Froot, Scharfstein, and Stein (1992), who show that rational investors will coordinate on sometimes unrelated market information. My paper can be seen as measuring the externalities which allow herding of any type to produce a potentially positive payoff in the short term.

Since the fund flow externality described may induce mutual fund managers to continue buying their portfolio even after they think the stocks are correctly priced, it may contribute to price overshoot and reversal as measured in Dasgupta, Prat, and Verardo (2011a) and Pareek (2009) as well as the "smart" and "dumb" money effect (Zheng (1999) and Frazzini and Lamont (2008), respectively).[7] All of these studies focus on the effect of a mutual fund's own flows predicting its own future returns, whereas my contribution is to consider a mutual fund's peers in the prediction of returns and fund flows through the channel of common stock holdings.

This paper proceeds as follows. Section I establishes my hypotheses in the context of the existing literature. Section II describes my empirical approach to measuring capital flow externalities, detailing network similarity measures, methodology and identification strategy. Section III discusses results, including the interpretation of network coefficients and their economic significance as well as cross-sectional implications. Section IV concludes.

[7]Dasgupta, Prat, and Verardo (2011b) model price overshoot and reversal due to herding with reputational concerns. In their model, managers receive enough reputational payoff to purchase negative expected return securities, thus inducing them to overshoot fundamentals. My innovation is to show that the return may be positive due to flow externalities and thus does not rely on a potentially non-pecuniary reputation effect.

I. Hypotheses and Background

I am proposing that the flows into and out of a mutual fund's peers create spillover effects which artificially amplify a fund's return, illustrated in the three panels of Table I. These three tables sort portfolios by a mutual fund's own flow and its Peer Flow. Since I am claiming that there are reinforcing fund flow loops among connected peer funds, returns should be highest among positive flow loops (high peer inflow combined with high own inflow – the lower right corner) and lowest among negative flow loops (high peer outflow combined with high own outflow – upper left corner). Using three different return measures, DGTW return (Daniel, Grinblatt, Titman, and Wermers, 1997), Carhart alpha (Carhart, 1997), and a simple market excess return, we see this pattern quite strongly. All portfolio 1 minus portfolio 5 differences are statistically significant and the return effects are economically significant as well, ranging from -4.4% to positive 4.2% per quarter.

If spillover induces overshoot, then there should be a reversal. This is depicted in Figure 1 with the DGTW excess return. At time 0, we see the initial starting point tabulated in the upper and lower row of Table I (this figure only accounts for Peer Flow dimension). In quarters 1-4, the subsequent year, we see the effect completely reverses.

While these results are suggestive, they are not rigorous. The remainder of this section establishes the paper's main empirical hypotheses related to the prediction of both mutual fund returns and mutual fund flows. First, to establish the hypotheses, I discuss the motivation for the similarity measure – how mutual fund managers are connected in security space. Second, I detail my hypothesis related to the prediction of portfolio returns which helps establish portfolio overlaps as the mechanism for spillover. Third, I develop two subsequent hypotheses related to the spillover effects of managers' fund flows.

A. Mutual Fund Similarity Measure

The mutual fund similarity measure, rigorously derived in Section II.B, uses the commonality of raw holdings in the form of a vector of portfolio weights to determine portfolio similarity. There are, however, other potential choices to determine similarity with manager's portfolios. First, I could use holdings in excess of a fund's benchmark (e.g. the Active Share measure of Cremers and Petajisto, 2009), or second, changes in holdings (shares at t minus shares at $t-1$) could be appropriate.

I choose to use raw holdings over excess holdings because Del Guercio and Tkac (2002) find that mutual fund investors seem to care more about raw returns than returns above a benchmark, and therefore raw holdings seem more appropriate. Changes in holdings are intuitively appealing and frequently specified in the herding literature (e.g. Sias, 2004). However, since Lou (2012) finds that managers sell dollar for dollar, changes in holdings and raw holdings should yield identical results for outflows. Using changes in portfolio holdings as the basis for a similarity measure requires action on the part of both parties, since two portfolio managers would only be similar to the extent that they are buying and selling the same stocks in the same quarter. I am trying primarily to capture passive effects – i.e. when a peer manager buys stocks in another's portfolio, the other manager benefits even having done nothing. Capturing passive effects requires raw holdings, not changes in holdings.[8]

[8]Wahal and Wang (2011) use a measure of holdings overlap as a measure of competition among similar mutual funds. Their study primarily looks at funds within the same category whereas I attempt to control for category effects while looking across categories. Ultimately, this is an empirical question: the relation should be negative if funds are primarily competing for funds (flows into one fund correlate with flows out of a similar fund). It may be that this relation is negative if I restrict my analysis to funds within the same Morningstar category, but

Others have studied the effect of common owners on financial securities. Kyle and Xiong (2001) model convergence traders spanning disparate markets inducing co-movement in the securities they hold, and more recently Antón and Polk (2010) measure stock co-movement related to the number of common owners. Greenwood and Thesmar (2011) measure the effect of correlated flows on stock price volatility. My innovation is to additionally consider the spillover effects on to other managers holding the same securities.

Cohen, Coval, and Pastor (2005) use a closely related similarity measure as a benchmark for portfolio manager performance. Their argument is that fund managers should be judged based on their relative performance compared to others holding similar portfolios. While the research question is very different, their similarity measure is almost identical. Their motivation is that this similarity measure captures the real 'competition' in a sense and thus reinforces it as a good choice for my investigation of peer effects among mutual fund managers.[9]

B. The Effect of Peer Flows on Portfolio Returns

It is known that a portfolio manager's fund flows affect asset prices. Coval and Stafford (2007) show that stocks with significant buying or selling pressure (due to fund flows) experience subsequent positive and negative returns, respectively. Jotikasthira, Lundblad, and Ramadorai (2012) confirm this internationally. Lou (2012) addresses fund flow driven price pressure including all fund flows, not just extreme positive and negative flows, and shows that this effect is still significant but

that is beyond the scope of this paper.

[9]Technically, Cohen, Coval, and Pastor use a doubly-stochastic weighting matrix (both rows and columns sum to one) while mine is only row-stochastic (rows sum to one) but the underlying data prior to these transformations is the same. See the appendix for more details on how I compute the weighting matrix, noting that Cohen, Coval, and Pastor additionally require columns to sum to one. In unreported results, I can show that the two measures are correlated 0.98.

asymmetric – he estimates that one dollar of inflows correlates with purchasing 0.6 dollars of the existing portfolio, while one dollar of outflows corresponds to selling one dollar of the existing portfolio.

Chen, Goldstein, and Jiang (2010) show that mutual fund investors behave like demand depositors and are susceptible to bank-run-like behavior, where outflows coincide with poor returns in a reinforcing cycle. While their analysis focuses on multiple investors in a single mutual fund, I consider the entire network of mutual funds, connected to the extent that they hold common securities. If investors in a single fund can experience strategic complementarities, then investors in other funds holding similar securities should be included in the analysis since the flows from peer funds may reinforce the effect.

To validate that my Peer Flow measure does, indeed, operate through the channel of common holdings, I investigate the return effect that peer flows have on the holdings of mutual funds experiencing large inflows and outflows, respectively. Because my similarity measure emphasizes concentrated holdings, discussed later in Section II.C, I focus on the top ten holdings of each mutual fund. If the flows into peer mutual funds affect a mutual fund through similar holdings, we should expect to see an effect on its most important holdings. Specifically,

HYPOTHESIS 1: *Stocks present in the top ten holdings of mutual funds with high peer inflows should experience positive returns, and those with high peer outflows negative returns.*

To test this hypothesis, I identify a portfolio of stocks who are exclusively held as one of the top 10 holdings of funds in one of three Peer Flow terciles (Outflow, No Flow, Inflow) and investigate the corresponding return effect.[10]

[10]Note that stocks not in the top 10 holdings of any fund are excluded, as well as top 10 stocks held simultaneously by mutual funds in differing terciles.

Intuitively, sophisticated arbitrageurs should attempt to exploit price anomalies, making any fund-flow driven mispricing small or short term. There are many reasons why this may not be so, however. Given that the source of price pressure may be hidden (e.g. Kyle, 1985), arbitrageurs may not identify price movement as a deviation from fundamentals, and thus not act to correct it. Arbitrageurs also face synchronization risk (Abreu and Brunnermeier, 2002, 2003), since multiple arbitrageurs may be necessary to absorb the price pressure, as well as other limits to arbitrage (e.g. Shleifer and Vishny, 1997). Indeed, rather than immediately arbitraging an over- or under-pricing, these sophisticated investors may even exacerbate the problem to increase the mispricing in a predatory manner and thus the profitability of a subsequent reversal trade (Brunnermeier and Pedersen, 2005).

To identify common portfolio holdings as a channel of contagion, I hypothesize that the fund flows of a manager's connected neighbors predict portfolio returns through the buying and selling of commonly held securities. Formally,

HYPOTHESIS 2: *The fund flows of neighbors connected by common asset holdings positively predict a mutual fund manager's portfolio return.*

To test this hypothesis, I compute my Peer Flow measure of connected-neighbor fund flows weighted by portfolio similarity, then estimate its impact on portfolio returns. To determine baseline and control variables, I draw known predictor variables of mutual fund returns from the existing literature. These include the two main factor models employed in the literature, DGTW (Daniel, Grinblatt, Titman, and Wermers, 1997) and the Carhart (1997) four factor model, in addition to the market return. I also include past flows to account for the flow-performance relation established in Chevalier and Ellison (1997). Since contemporaneous fund flows and portfolio returns may suffer from endogeneity, I instrument peer flows in a GMM framework, discussed in detail in Section II.D.

Another important factor is style investing (Barberis and Shleifer, 2003). To explicitly control for style flows, I take the average flow to a given mutual fund style as defined by the assigned Morningstar Category for a fund. This includes such styles as Large Cap, Small Cap, Technology, Health Care, Growth, and Value. There is evidence that a 'flow' factor exists in the prediction of mutual fund returns (Sirri and Tufano, 1998, Lou, 2012), so I show that, of the various candidate variables for flow factors, Peer Flow and Style Flow are the two that matter the most (see Appendix A). Thus, I include Style Flow in all specifications moving forward as a control for a style investing effect.[11]

Finally, if this flow effect is indeed an abnormality which induces overshoot, there should be a measurable reversal as observed in Figure 1. Thus, I propose to test the following hypothesis:

HYPOTHESIS 3: *The fund flows of neighbors connected by common asset holdings predict a return reversal in the subsequent year.*

The time frame of the following four quarters is obtained from visual inspection of Figure 1 which seems to indicate reversing returns in the following year.

C. The Effect of Peer Flows on Fund Flows

Having identified common portfolios as the channel of influence and shown that peer flows affect fund performance, the next step is to investigate the effect of peer fund flows on a manager's own fund flows – the fund-flow externality result which generates the spillover effect. Chevalier and Ellison (1997) identify a flow-performance relation such that positive past returns predict future inflows and

[11]This is not surprising since Lou (2012) notes that his flow factor accounts for market-wide flows whereas my measure focuses on closely related funds, thus providing more precision.

poor past returns predict future outflows. While Chevalier and Ellison measure these effects through lagged returns, the outflows could be contemporaneous, since a sophisticated manager, seeing his poor returns, may sell in anticipation of future outflows. Until now, investigation of this relation has been challenging due to the endogeneity problem between contemporaneous flows and returns, a problem I solve with an instrumental variables specification.

This connection between the fund flows of neighboring managers suggests two related hypotheses:

Hypothesis 4: *The fund flows of neighbors connected by common asset holdings positively predict a manager's own fund flows.*

Hypothesis 5: *Fund flow externalities from each manager on to each other manager are greater than zero.*

To test both hypotheses, I employ a network specification which allows a contemporaneous estimation of spillover effects across a network of connected agents. In this network specification, I include my measure of connected-neighbor fund flows as a predictor variable (Peer Flow), instrumented by the two-step neighbor fund flows along with other common predictor variables. A positive and significant coefficient on Peer Flow confirms Hypothesis 4: fund flows generate externalities through interconnected portfolios.

A positive and significant relation establishes the existence of an externality, but obtaining evidence for Hypothesis 5 requires interpretation of the coefficient estimate. Indeed, the richness of information available from this network specification constitutes a primary advantage over a standard linear regression model. My specification behaves like an auto-regression, but in the cross-section: fund flows at time t show up both as dependent and independent variables, and as such the estimated coefficient on connected-neighbor flows affects all other coefficient estimates in steady-state, similar to a temporal auto-regression framework.[12] When the model is rearranged such that flows are isolated on the left-hand side, the coefficient on each right-hand side variable becomes a matrix specifying the effect each portfolio manager has on each other manager in equilibrium.[13] This matrix coefficient compares to the scalar coefficient estimating the average effect in most other specifications. The average of the matrix coefficient off-diagonal values measures the spillover effects, while the average of the diagonal in excess of the non-networked linear coefficient measures feedback effects. Positive off diagonals in this matrix coefficient provides evidence of Hypothesis 5.

II. Network Methodology

Having motivated my research design in the context of the existing literature, I next turn to the description of the exact implementation of my network methods and variables. I first describe my Data in Section A, then develop my mutual fund similarity measure in detail in Section B. I next discuss how this measure relates to crowded trades in Section C before proceeding to descriptions of the GMM estimation approach, network instrument, and full specification in Section D.

A. Data

My primary dataset is from Morningstar and contains the flows, returns, and full portfolio holdings of U.S. Open Ended funds from 1998 to 2009.[14] Flows of

[12]Specifically, this model is a Spatial Auto-Regression (SAR), popular in spatial econometrics, and developed fully in LeSage and Pace (2009). It is typically used to measure geographic externalities like traffic congestion or environmental spillovers.

[13]I develop this more rigorously in Section III.C.

[14]Elton, Gruber, Blake, Krasny, and Ozelge (2010) perform a thorough comparison of the Morningstar holdings data with the more commonly used data from Thomson Reuters and find it to be very similar.

funds are a simple dollar value per fund, per month or quarter and are net flows summarizing all subscriptions and redemptions across the relevant time period. Note that the data includes reported values for both fund flows and portfolio returns, whereas other studies typically compute fund flows from returns and changes in total net assets. To focus on equity funds as is common in the literature, I only keep funds with at least 75% of their holdings in equities, but my results are robust to this choice.[15] I combine this data with CRSP by CUSIP when necessary to obtain stock characteristics. DGTW returns are computed as in Daniel, Grinblatt, Titman, and Wermers (1997) and Wermers (2003), downloaded from Russ Wermers' website.

I include quarterly observations of each fund's cash holdings rather than the annual measures reported in the CRSP Mutual Fund database typically used because the Morningstar data contains the entire portfolio holdings of each open-ended fund. Going forward, *Flow* is defined as fund flow divided by total net assets as in Coval and Stafford (2007) and *Size* is the log of total net assets. *Cash* is defined as currency, treasuries, and other cash-like holdings, also divided by total net assets. Summary statistics for these and other variables of interest are available in Table II.

B. Portfolio Similarity Measure – The Network

I construct the similarity between two portfolios i and j, denoted s_{ij}, as the dot product between the security holding weight vectors of each portfolio manager i and j, divided by the product of the Euclidean norm of each vector.[16] Figure

[15] In previous drafts of the paper, I keep any fund with nonzero equities and get similar results.

[16] Note that this similarity measure is the same as the cosine of the angle between the two vectors in security space, and for centered data is identical to the computed correlation between the two vectors. See more discussion of cosine similarity vs correlation computations in Reca, Sias, and Turtle (2011). Detailed definitions of the norm and a full derivation is available in Appendix B. This notion of portfolio distance is intuitively and mathematically similar to that

2 plots percentiles of the distribution of this portfolio distance measure through time and shows time variation in the measure but no real time trend.

This similarity measure is a primary component of the key variable *Peer Flow*, which quantifies the impact of related neighbor's fund flows. To construct *Peer Flow* for each manager i, I compute a weight vector which is the similarity measure s_{ij} for each other manager j divided by the sum over all similarities, setting self-similarity s_{ii} to 0. For example, consider a portfolio manager with three neighbors with similarities of 0.1, 0.2, and 0.1, such that the weights are .25, .50, and .25, respectively. If those neighbor's flows (divided by total net assets) are 0.01, 0.05, and 0.10, respectively, then *Peer Flow* is $(.25 * .01) + (.50 * .05) + (.25 * .10) = 0.0525$.

Note that I also compute other peer variables such as peer size (total net assets) and peer cash (divided by total net assets) in the same way.[17] The primary advantage of using this peer weighting procedure is that peer statistics are very similarly distributed to the corresponding 'own' statistics, as seen in Table II.

C. Measuring Crowded Trades

The nature of portfolio connections is important for academics to understand the dynamics of the externalities involved but also market participants or regulators who may want to prevent spillover among connected mutual funds. We can think about the nature of these connections in two dimensions. First, consider the driver of the similarity: does it primarily rely on a few concentrated holdings or require a broad, dispersed similarity? Second, consider the temporal aspect of the connection, is it transient or persistent?

---

of social distance as in Conley and Topa (2002).

[17]All of the *Peer* variables are the same as structural equivalence variables in sociology originally defined by Burt (1987) and used more recently by Bothner (2003). The primary difference is that in sociology, these variables are typically lagged to provide identification, whereas herein I use an IV/GMM specification for better identification with my data.

The driver of the similarity, dispersed or concentrated, can be determined by choice of similarity measure. An absolute difference measure would treat each holding equally such that \$100 difference in a small position is the same as \$100 difference in a large position. This gets problematic when you consider that a \$100 difference each in 25 small positions would equate with a \$2,500 difference in a single large position. It is not clear that those should be treated equally.

Motivated by this (and other) problems with equal weighted measures as well as literature indicating that concentrated positions are most important, I choose to emphasize concentrated positions.[18] Kacperczyk, Sialm, and Zheng (2005) show that mutual fund managers are able to make abnormal returns primarily due to their concentrated bets, which indicates that outperformance lies in those holdings. This is corroborated by (Cohen, Polk, and Silli, 2010) and (Pomorski, 2009) who also find that the large holdings are the ones that matter to mutual fund managers and fund performance.

Figure 3 shows how my normalized dot product similarity measure increases in two dimensions, emphasizing concentrated positions. First, it is increasing in portfolio overlap, its primary purpose. As the percentage of portfolio overlap increases, the similarity between them increases (the distance between two managers in security space decreases). The mutual fund similarity measure is also increasing in the concentration of those holdings due to the fact that portfolio weight vectors are multiplied together, creating a quadratic relation. Holding total portfolio overlap constant, a single concentrated position gives twice as much similarity as two overlapping holdings of equal proportion. This property of my portfolio distance measure indicates that concentrated positions give rise to more interconnected-

[18]Another problem using the absolute difference is that it equates -\$100 with \$100, yet without the absolute value, summing a sequence of signed differences creates canceling problems. Counting number of overlaps is another method of equal weighting, used for overlapping analyst coverage in Israelsen (2010) but it is less precise.

ness. Accordingly, crowding or overweighting in a specific set of securities may induce more connectedness among those managers than they may realize.

Next, consider the temporal dimension of portfolio similarity, which is an empirical question, not a modeling choice. In terms of identification, if portfolio similarity is persistent, then the underlying driver of similarity itself must be persistent. This would indicate a role for common mutual fund benchmarks or index investing as driver of similarity. However, if portfolio similarity is transient, then the period-by-period choices of mutual fund managers must play a substantial role as they dynamically allocate funds to their portfolio.

Between the two, transient, or hard-to-observe portfolio interconnections pose the greater risk to portfolio managers and regulators alike since a hidden externality is more likely to generate unexpected negative shocks. Transient portfolio interconnections may arise due to crowded trades, which occur when portfolio managers take temporary concentrated or overweighted positions in a small set of stocks. Due to lags in mandatory disclosures, these portfolio positions may not be detectable to market participants until many months after the trades are established. Thus, with no knowledge of network connections, negative flow shocks across portfolio connections will be unanticipated and likely produce greater negative consequences than shocks which are at least partially anticipated.

Table III presents the results of a panel auto-regression on the mutual fund similarity measure, a design similar to the main specification in Antón and Polk (2010). This specification takes the symmetric $N \times N$ network of relations between all of the portfolio managers at time t and puts them in a $N(N-1)/2 \times 1$ vector as the dependent variable. The same network of relations at times $t-1$, $t-2$, $\ldots$, $t-p$ enter as independent variables, vectorized. I then run this regression for each time t and summarize the p coefficients (for $p = 8$ quarters) across time in a

Fama-MacBeth framework.

The marginal effects of the lags diminish to statistical insignificance after three lags, but still show some auto-regressive properties. The mutual fund similarity measure correlation lagged one quarter is 0.41, which indicates some short-term persistence. To estimate the auto-correlation two quarters previous, I compute $0.41^2 + 0.25 = 0.42$, showing that the persistence extends to six months. But the correlation between the network distance measure and that of three quarters past is $0.41^3 + 0.25^2 + 0.765 = 0.21$, a significant drop off, and then one year past is $0.41^4 + 0.25^3 + 0.08^2 + 0.0 = 0.05$ if I treat the insignificant fourth lag as zero, or 0.13 if I retain it. After two years, retaining the first four coefficients, the correlation is $0.41^8+0.25^7+0.08^6+0.08^5 = 0.0009$, or almost zero.[19] Since portfolio objectives likely persist longer than two years, these results suggests that there is some transience to my measure of interconnectedness and thus that crowded trades or herding among institutional managers plays a role in fund flow externalities.

D. Network Structure as Instrument

I employ an instrument to identify influence rather than just correlation, since cross-sectional fund flows and returns of each portfolio manager are endogenous.[20] Without instrumentation, a correlation between two portfolio manager's fund flows is not sufficient evidence of one's influence on the other.

Following Bramoullé, Djebbari, and Fortin (2009), I employ a network-structure based instrument to address this endogeneity based on "intransitive triads" which are often present in a network. An intransitive triad is present if A connects to B and B to C, but A is not connected to C. Thus, A can instrument for B's influence

[19] This analysis of time series coefficients comes from Hamilton (1994), Chapter 1.

[20] Since the diagonal of weighting matrix W is set to zero, $Flow_i$ is never on both sides of the same specification, so there is no mechanical collinearity, only endogeneity.

on C since any influence A has on C must be through the common relation with B. In network terminology, A and C are *Two-Step* neighbors, so my instrument is *Two Step Peer Flow*, computed the same as *Peer Flow* but with the two-step neighbor.[21]

For instance, a U.S. technology fund may be connected to a mid-cap fund through common mid-cap technology holdings, and that mid-cap fund may also be connected to a Latin American fund through mid-cap Latin American holdings. Thus, the flows of the Latin American fund can instrument for the mid-cap fund's influence on the U.S. technology fund since they are only connected through their common mid-cap neighbor. However, while two portfolio managers may not be directly connected, they both likely maintain some set relation to market-wide movements – i.e. they have a 'beta' in some sense. I address this by including time and fund fixed effects as well as market returns as control variables (or simply using excess returns or alphas as the dependent variable).[22]

Not all two-step neighbors form intransitive triads, however. Two-step neighbors can only serve as an instrument if they satisfy the exclusion restriction – that the instrument is only correlated with the dependent variable through the endogenous regressor. To address these concerns, Bramoullé, Djebbari, and Fortin (2009) specify a rank test which establishes that peer influence effects are identified through two-step neighbors, which my network satisfies.[23] To ensure overidenti-

[21] Details are in the Appendix.

[22] One may argue that there may be a time-varying connection between otherwise unconnected funds as just described. For this to be a problem, three things must be true. First, the time varying connection must vary significantly over time, otherwise my fund fixed effects control for it. Second, the time varying connection must be consistently in one direction - i.e. always positively correlated or always negatively correlated. If the time-varying nature of the connection varies between positive and negative correlation, then this just adds noise to the process and works against me finding a result. Finally, the time-varying connection must not be correlated with market-wide events, otherwise the time fixed effects will control for it.

[23] This is Proposition 5 and is simply a rank test of the matrix of network connections. I take the identity matrix, my Peer Weight matrix, and my Two-Step Peer Weight matrix (see the appendix for details), reorder them to be in column vectors, and show that the resulting matrix

fication, I include not just $TwoStepPeerFlow$ but also $TwoStepPeerFlow^2$ as excluded instruments, which is common practice in an IV specification.

To test my first hypothesis, I instrument for peer fund flows as described above, but place mutual fund returns as my dependent variable. Specifically, I estimate:

$$PeerFlow_{it} = TwoStepPeerFlow_{it} + TwoStepPeerFlow_{it}^2 \tag{1}$$

$$\begin{aligned} Return_{it} = \widehat{PeerFlow}_{it} + Flow_{t-p} + Return_{t-p} \\ + FundSize_{it} + CashPct_{it} + Amihud_{it} + StyleFlow_{jt} \end{aligned} \tag{2}$$

where *Return* is either the excess DGTW return or raw return with the mutual fund DGTW return included as an independent variable. Excess DGTW return is defined as the mutual fund gross return less mutual fund DGTW return, which is the portfolio weighted individual equity DGTW returns.[24] If $PeerFlow$ is a positive predictor of portfolio returns, then it seems highly likely that commonly held securities are the channel of influence and thus Peer Flows may induce mutual fund return overshoot.

A good GMM/IV specification uses instruments which are correlated with endogenous regressors but orthogonal to the error term. As a first test that my instruments are correlated with the endogenous regressors, I present the results of the first stage regression in Table IV. The R^2 is 0.87 which indicates a high degree of correlation. In addition, for each subsequent GMM specification, I include four key statistics which validate the instrument specification. The *KP LM Stat* is the Kleinbergen-Paap Wald Statistic which tests for weak instruments. Weak instruments are not correlated 'enough' with the endogenous regressors and can lead to

has at least rank 3. More details of this computation available on request. Bramoullé, Djebbari, and Fortin (2009) also require that the network variable be 'centered' which I satisfy since fund flows are already mean zero.

[24]In unreported results, I also use Carhart four-factor alpha as the dependent variable with similar results.

biased results. This test should reject the null hypothesis so strong instruments should show a large test statistic and low p-value (less than 0.05). The *Hansen J statistic*, also known as the test of overidentifying restrictions, tests the validity of the overidentification of the model, which is correctly specified when we cannot reject the null hypothesis. This is a test of whether the instruments are correlated with the error term of the second stage regression, which should not be so. Thus, a very small J statistics and correspondingly large p-value which fail to reject the null indicate a properly overidentified GMM specification.[25]

To test my second hypothesis that fund flows spillover across similar mutual funds, I incorporate my network measure in addition to common predictor variables in a specification with fund flows as the dependent variable. Coval and Stafford (2007) employ both lagged flows and lagged returns as predictors, but I instead follow (Lou, 2012) who uses lagged Carhart four-factor alpha.[26] Alpha is computed using a 12 month rolling average regression and so accounts for alpha over the past 12 months, so only a single lag is necessary. Sirri and Tufano (1998) shows that Style Flow (Barberis and Shleifer, 2003) and fund size (measured as log of total net assets) are important determinants of flows given investors' non-zero search costs.[27] Since temporary asset price movements may be stronger for illiquid securities, I include a portfolio-wide *Amihud* measure, which is simply the weighted average of the Amihud liquidity measure computed for each individual equity holding (Amihud, 2002).[28]

[25]For more information on the derivations of these test statistics, see Hayashi (2000) or some other comparable statistics text.

[26]Lagged raw returns show little significance in predicting flows in my specification, but do not change any of the primary results.

[27]Style Flows are the average flows to reported Morningstar categories. In unreported results, I control for common flows to mutual fund families instead of fund category average flows with no material difference in outcome (Elton, Gruber, and Green, 2007). I do the same with average market-wide flows as well to account for market share effects (Spiegel and Zhang, 2013), with no difference. I cannot include them simultaneously due to collinearity problems.

[28]I also computed a full portfolio Amihud measure including cash and non-equity, non-cash

I also include *Peer Size* as a control variable since fund size is an important predictor of flows. The peer size control is important in a network specification because if flows primarily go to larger funds (Sirri and Tufano, 1998), then funds that are both large and connected may simply experience correlated flows without any mutual influence.

A portfolio manager's cash holdings provide a vital cushion against unexpected redemptions, and as such likely influence the prediction of inflows and outflows. Most studies exclude cash holdings because data is unavailable, not because cash holdings are unimportant.[29] I do have cash holdings data, and therefore include it for both the manager and connected neighbors (*PeerCash*), since a manager connected to cash-poor neighbors may be more susceptible to flow externalities.

In sum, I estimate the following set of equations in a GMM specification:

$$PeerFlow_{it} = TwoStepPeerFlow_{it} + TwoStepPeerFlow_{it}^2 \tag{3}$$

$$\begin{aligned} Flow_{it} = {} & \widehat{PeerFlow}_{it} + Flow_{t-p} + Alpha_{t-1} \\ & + Size_{it} + Cash_{it} + Amihud_{it} + PeerSize_{it} \\ & + PeerCash_{it} + StyleFlow_{jt} \end{aligned} \tag{4}$$

in which $Fund_i \in Style_j$, 4 time lags are included ($p = 4$), *Alpha* is computed as the 12 month rolling average Carhart four-factor alpha and $\widehat{PeerFlow}_{it}$ is the

holdings at the minimum and maximum Amihud measure, respectively, with similar results. Other liquidity measures such as spreads (bid minus ask over midpoint) and turnover (average daily volume divided by shares outstanding) also gave similar results, available on request.

[29]A notable exception is Simutin (2009) who measures mutual fund cash holdings and finds a positive relation between cash holdings and fund performance. My findings corroborate his results.

fitted values from equation (3).[30,31]

III. Results

In this section, I provide evidence confirming the hypotheses proposed. First, I establish that common holdings is a primary channel of spillover, after which I report the primary regression results. Next, I interpret those regression results in the context of my network specification and finally provide some robustness checks.

A. Overlapping Holdings as Spillover Channel

To show that overlapping holdings are the correct network specification, I show that neighbor's fund flows – Peer Flows – are positively correlated with a mutual fund's top holdings. The focus on top holdings is because the portfolio similarity measure emphasizes these top, concentrated holdings such that peer effects should be most dominant among the largest holdings of a mutual fund.

Results are presented in Table V. Stocks are sorted into portfolios based on their *Peer Flow* tercile, where portfolio 1 is Peer Outflows, portfolio 3 is Peer Inflows and portfolio 2 has Peer Flow very close to zero. Stocks in each portfolio must be in the top ten holdings of a mutual fund, and exclusively be held (in the top 10) by mutual funds with Peer Flows in the given tercile. Stocks that would otherwise have been assigned multiple portfolios are dropped.

[30]Here I use Carhart Alpha for parsimony since a single lag can account for the past year's performance. Including four lags of DGTW excess return yields similar results but sometimes give singularity problems in the variance-covariance matrix with so many explanatory variables. All results go through with either DGTW returns or Carhart Alpha.

[31]Note that the exact specification of equation (3) includes all control variables in equation (4). To use strict GMM terminology, *Peer Flow* is the endogenous regressor, $TwoStepPeerFlow$ and $TwoStepPeerFlow^2$ are excluded instruments, and the rest of equation (4) are included instruments.

The results confirm Hypothesis 1, both with DGTW returns as well as raw returns. Those stocks held by the outflow portfolio have a DGTW return of -0.457% and raw return of -0.267%, with differences of -1.42% and -3.20%, respectively, when compared to either the middle (net zero flow) tercile. Both differences are also statistically significant. Stocks held by the inflow portfolio have DGTW returns of 1.82% and raw returns of 6.91%, and are also significantly different from the middle tercile portfolio with no Peer Flow effect. That Peer Flow is positively correlated with a return effect on the top holdings of the mutual funds affected indicates that the trade channel through common holdings is an important channel of spillover among mutual funds.

B. Regression Results

The baseline fund flow specification follows Sirri and Tufano (1998) and Coval and Stafford (2007), and is included as Model 1 in Table VIII. I find it necessary to include both time and fund fixed effects and further cluster my standard errors in both time and mutual fund dimensions, though most of the literature thus far has used Fama-MacBeth or an OLS specification with fewer corrections.[32]

I regress mutual fund portfolio returns on my networked and instrumented *Peer Flow* variable as evidence that portfolio overlaps are driving a spillover effect. As shown in Table VI, there is a positive and significant coefficient on *Peer Flow* across all specifications, wether excess returns or raw returns controlling for the fund's

[32]A Breusch-Pagan test and an F test on RSS of regressions with and without time and fund fixed effects show that it is necessary to include some type of fixed or random effects. A Hausman test verifies that fixed effects are necessary over random effects (Kennedy, 2003). Clustering standard errors in both time and manager dimensions produces large changes in standard errors indicating that this is a necessary step (Petersen, 2009). With the same pooled OLS and Fama-MacBeth framework as Coval and Stafford (2007), I get results qualitatively similar to them and others who have investigated this relation such as Lou (2012) and Ferreira, Keswani, Miguel, and Ramos (2011). Results from these tests as well as a table comparing the varying differences in specification are available upon request. Recall that my dataset is different from the other studies cited and as such these test results may or may not extend to their specifications.

DGTW return. A Carhart four-factor alpha specification gives the same result, untabulated. *Peer Flow* is also economically significant: a one standard deviation move in *Peer Flow* corresponds to a 2% move in mutual fund return per quarter, which is quite large. The fact that fund flows from neighboring portfolio managers positively predict returns is solid evidence that portfolio interconnections are the channel for this influence, and are consistent with Hypothesis 2.

Next, we turn to reversals. If this is indeed a spillover effect which induces overshoot, we should see this effect reverse. Indeed, in Table VII we see exactly that. With the same specification, the only thing that changes is the dependent variable, this time the leading annual holding period return. Here, the sign on *Peer Flow* reverses and is statistically significant. Again, the economic significance is 2% return, this time annually, indicating a full reversal of the effect over the subsequent year. This confirms Hypothesis 3.[33]

I propose that these effects are due to spillover effects among connected mutual funds. To address that, I move to unique a specification which allows the measurement of spillovers taking into account adjacent peers. It is called a Spatial Auto-Regression (SAR) and is typically used in a geographic setting to measure spillovers from traffic flow or environmental emissions. The main SAR flow specification is in Table VIII. Here, *Flow* is the dependent variable with *Peer Flow* as independent variable alongside other control variables. Again, *Peer Flow* enters in positively and significantly with slight decreases in other predictor variables, indicating a flow externality consistent with Hypothesis 4. However, since *Flow* enters into the specification both as a dependent and independent variable, I must transform the equation similar to a temporal auto-regression specification to fully

[33]Note that the specifications include lags of flow and returns are not included in Table VII. They give a similar result (coefficient on Peer Flow is negative and significant), but require the lagged coefficients to be partialled out for variance-covariance matrix to be full rank (Baum, Schaffer, and Stillman, 2007) and so are not included here.

interpret this coefficient.

C. Network Coefficient Interpretation

The coefficients in Model 5 of Table VIII require further analysis since this is a Spatial Auto-Regression, similar to a time series autoregression specification (Hamilton, 1994). Ultimately, the coefficient on each explanatory variable is a matrix, not a scalar. Thus, there is not an 'average' effect per se, but rather a separate coefficient for the effect of each mutual fund on each other mutual fund. It is these matrix coefficients which allow the direct measurement of spillover effects.

To interpret these matrix coefficients, I begin by rewriting my specification in Equation 4 in matrix form, without the instrumentation:[34]

$$F_t = \rho_s W_t F_t + \rho_t F_{t-1} + X_t \beta + \epsilon \tag{5}$$

in which F_t is the $N \times 1$ vector of fund flows at time t. W_t is a row-stochastic transformation of $N \times N$ portfolio similarity matrix S at time t, such that $PeerFlow_t = W_t \cdot F_t$. X_t represents all other control and explanatory variables for simplicity. The result is

$$(I_N - \rho_s W)\, F = X\beta + \epsilon \tag{6}$$

$$F = X_1 \tilde{\beta}_1 + X_2 \tilde{\beta}_2 + \ldots + X_L \tilde{\beta}_L + \epsilon \tag{7}$$

[34]For this analysis, I simply use the endogenous *Peer Flow* rather than the predicted value from the first stage regression, which simplifies the exposition and likely is a good approximation since the R^2 of the first stage regression is 0.87. I still use the coefficient estimates from the instrumented specification, however. Matrix algebra showing how this relation is the same as my empirical specification can be found in the Appendix.

for each $l = 1 \ldots L$ explanatory variables. Each actual estimated coefficient is

$$\tilde{\beta}_{l,N\times N} = (I_N - \rho_s W)^{-1} \beta_l \tag{8}$$

which is an $N \times N$ matrix. Without my network specification, the comparable coefficient would be the scalar coefficient estimate times an $N \times N$ identity matrix.

To interpret the network coefficients, I divide each matrix coefficient into feedback effects, represented by the diagonal, and spillover effects which reside on the off-diagonal. The results in Table IX, noting that the peer effects from size and cash are included in those coefficients. The first column is the scalar coefficient estimate, β, without the network transformation. Next are the direct effects, computed as the average of the diagonal of the matrix coefficient. Intuitively, these are the cumulative effects that each mutual fund has on itself (the traditional analysis) except additionally including any feedback effects from spillover to others which is then propagated back to the originator. Next, spillover effects are computed as the average of all off-diagonal entries in the matrix coefficient. This captures the average effect that a shock to a mutual fund has on other mutual funds.[35]

Table IX shows how the network specification measures feedback and spillover effects, significantly increasing the estimates of each variable. Specifically, for most variables in Panel IXa, the spillover effects are about 90% of the direct effects, providing solid evidence for Hypothesis 5. Panels IXb and IXc additionally show results for Models 4 and 3, respectively, giving a range for spillover coefficients of 48% of direct effects in Panel IXb up to 137% of direct effects in Panel IXc. This is quite important because the primary driver of spillover effects is the coefficient on Peer Flow, yet for small movements (0.33 and 0.58), we see very large swings in estimates of spillover effects, which remain statistically significant in most cases.

[35]This entire exposition follows LeSage and Pace (2009) and is standard for an SAR model.

Of note in Table IX is the size effect, which is negative. This means that larger sized funds experience lower spillover effects, and larger peer funds produce larger spillover effects. To see this, we return to Table VIII to investigate the original coefficients. The coefficient on Fund Size is positive and significant, indicating that larger funds see more inflows and fewer outflows. Peer Fund Size, however, is negative, indicating a 'competitive' effect which means that larger peer funds are associated with more outflows, and smaller peer funds with more inflows. This is consistent with the above interpretation of size effects.

I further explore this by re-running the specification in Table VIII with a new variable, *Relative Fund Size* which is Peer Fund Size divided by Fund Size (and removing Fund Size and Peer Fund Size due to collinearity). A larger Relative Fund Size means either larger Peer Fund Size or smaller Fund Size, and vice versa. The results are in Table X show a negative coefficient on Relative Fund Size. Thus, as Relative Size gets bigger, which means that Peer Fund Size is bigger and/or a fund's own size is smaller, we see more outflows. Inversely, as Relative Fund Size gets smaller, which means Peer Fund Size is smaller and/or a fund's own size is bigger, we see more inflows. This further confirms the effect of size in this regression as well as past results that larger mutual funds see more inflows (Sirri and Tufano, 1998).

To further investigate the economic effects spillover, I simulate a shock to a random subset of fund managers in the sample and measure the impact to the others. Each variable listed gets a negative one standard deviation shock, and Table XI records the spillover effects on predicted fund flows. For most variables, the ratio of spillover to direct effects is about 20%. The economic effects of spillover are not trivial, raging from 12% of mean predicted flow for a shock to lagged fund Alpha to 18% of mean flow for lagged fund flows (the mean and standard deviation for fund flows is available in Table II). A shock to Fund Size gives a very large result, but a one standard deviation shock to fund size is unlikely in a short amount of time, so I discount that result. While these effects may seem modest, recall that these effects are assumed to be zero in non-networked specifications, the whole-sample coefficient estimate of 0.47 in Model 5 of Table VIII seems to be lower than most subsets, and localized effects may be much larger.

D. Robustness Checks

To more fully identify crowded trade externalities as a unique phenomenon, I perform several robustness checks. I re-run my main specification removing all sector funds from the dataset.[36] Results with and without sector funds are presented in Table XII. Model 1 is reproduced from Model 5 of Table VIII for reference, Model 2 omits sector funds and Model 3 has only sector funds. Sector funds are funds restricted to holding stocks based on industry classifications like Technology or Health Care. The results show consistent parameter estimates for all three models, none statistically different from any of the others, indicating no special role for sector funds, again reinforcing that sector rotation, a form of style investing, is not driving the result.

Since financial crises induce correlations across disparate asset groups, it is possible that the results are simply arising from the recent financial crisis. Accordingly, I re-run my specification omitting the financial crisis, ending the analysis in the second quarters of 2007 and 2008, respectively, as well as beginning in 2001 to omit the tech bubble of the late 1990's with results presented in Table XIII. Again, we see consistently similar parameter estimates indicating that these large, market-wide events are not driving the result.

[36] Sector funds are those labeled Technology, Utilities, Financials, etc. corresponding to equities held in a specific industry.

IV. Conclusion

In the wake of the recent financial crisis, the interconnection of market participants has become an important new area of research. Employing a network-based specification, I show that interconnected mutual funds exhibit externalities in fund flows, exposing them to feedback and spillover effects resulting in estimates 20% greater than non-networked coefficients. To incorporate these network connections, I contemporaneously estimate the influence of each portfolio manager's fund flows on each other manager by exploiting the network structure as an instrument. The fund-flow spillover effect result in mutual fund return overshoot of 2% per quarter, completely reversed in the subsequent year, indicating that they are clearly non-fundamental in nature.

I also show evidence that this externality is the result of crowded trades – short-term, popular market positions – since portfolio connections exhibit only a small amount of short-term persistence and the effect quickly reverses. Furthermore, I have illustrated how distances between portfolios in security space emphasize concentrated positions, such that active managers overweighting portions of their portfolio may unintentionally increase their dependence on similar neighbors.

While my analysis focuses on the equity holdings of open ended funds, it also has implications for collateralized financing. Financial intermediaries who rely on collateralized financing to fund their investments are growing in market share (Adrian and Shin, 2010). It may be that my results imply a broader collateral externality similar to the geographic model of Kiyotaki and Moore (1997), which may have played a role in recent runs on repo financing (Gorton and Metrick, 2011).[37] Since even interbank lending is becoming more collateralized, the canon-

[37]Kiyotaki and Moore (1997) develop a model in which the collateral for a loan is also the productive asset, in their case farmland. This metaphor quickly extends to financial markets

ical model of interbank financial contagion from Allen and Gale (2000) may be further amplified by collateral connections.[38]

This work also provides motivation for the collection of more detailed holdings data from market participants, as the results described herein can be characterized as a negative network externality which could merit government regulation. Indeed, Brunnermeier, Hansen, Kashyap, Krishnamurthy, and Lo (2011) recently responded to an AEA/NSF call for proposals on "grand challenge questions" for research in the next ten years by advocating the collection of additional data and development of network models in the pursuit of quantifying systemic financial risk. While immediate public disclosure may have unintended predatory trading effects (Brunnermeier and Pedersen, 2005), this problem seems surmountable (Abbe, Khandani, and Lo, 2012). Confidential disclosure to regulatory bodies and/or delayed public disclosure are likely to be beneficial and could be the purview of the newly formed Office of Financial Research established by the Dodd-Frank Act.

Network methods are becoming more popular in corporate finance (e.g. Hochberg, Ljungqvist, and Lu, 2007, Cohen, Frazzini, and Malloy, 2008, Ahern and Harford, 2010, Lewellen, 2012) and market microstructure (e.g. Cohen-Cole, Kirilenko, and Patacchini, 2010), although little has been done to apply network methods to securities markets. My network approach allows a steady-state analysis of this peer influence process in the cross-section, bringing structure to cross-sectional analysis previously only available in the time-series. While I have applied it to portfolio interconnections, it may also have broad applicability to other areas

where the productive securities of an intermediary's portfolio are collateral for repo financing.

[38] In November 2009, the ECB (Heider and Hoerova, 2009) reported that interest rates for collateralized lending in the interbank market since 2007 were significantly lower than unsecured rates - a historical divergence - and a more recent report from the Financial Times indicates that interbank unsecured lending in Europe has essentially disappeared. (http://ftalphaville.ft.com/blog/2010/08/16/315556/euribor-has-been-vaporised/)

such as interbank lending (Cohen-Cole, Patacchini, and Zenou, 2011) and stock market volatility (Greenwood and Thesmar, 2011). In a time when bailouts are motivated not because of too-big-to-fail but because of too-interconnected-to-fail, understanding and quantifying the interconnections among market participants is a vital pursuit.

References

Abbe, Emmanuel A, Amir E Khandani, and Andrew W Lo, 2012, Privacy-preserving methods for sharing financial risk exposures, *The American Economic Review* 102, 65–70.

Abreu, Dilip, and Markus K Brunnermeier, 2002, Synchronization risk and delayed arbitrage, *Journal of Financial Economics* 66, 341–360.

———, 2003, Bubbles and crashes, *Econometrica* 71, 173–204.

Adrian, Tobias, and Hyun Song Shin, 2010, The changing nature of financial intermediation and the financial crisis of 2007–2009, *Annual Review of Economics* 2, 603–618.

Ahern, Kenneth R, and Jarrad Harford, 2010, The importance of industry links in merger waves, *Unpublished working paper*.

Allen, Franklin, and Douglas Gale, 2000, Financial contagion, *Journal of Political Economy* 108, 1–33.

Almazan, Andres, Keith C Brown, Murray Carlson, and David A Chapman, 2004, Why constrain your mutual fund manager?, *Journal of Financial Economics* 73, 289–321.

Amihud, Yakov, 2002, Illiquidity and stock returns: Cross-section and time-series effects, *Journal of Financial Markets* 5, 31–56.

Antón, Miguel, and Christopher Polk, 2010, Connected stocks, *Unpublished working paper* pp. 1–56.

Barberis, N, and Andrei Shleifer, 2003, Style investing, *Journal of Financial Economics* 68, 161–199.

Bartholomew, David J, Martin Knott, and Irini Moustaki, 2011, *Latent Variable Models and Factor Analysis* . A Unified Approach (Wiley).

Baum, Christopher F, Mark E Schaffer, and Steven Stillman, 2007, Enhanced routines for instrumental variables/GMM estimation and testing., *Unpublished working paper*.

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch, 1992, A theory of fads, fashion, custom, and cultural change as informational cascades, *Journal of Political Economy* 100, 992–1026.

Bothner, Matthew S, 2003, Competition and social influence: The diffusion of the sixth-generation processor in the global computer industry, *The American Journal of Sociology* 108, 1175–1210.

Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin, 2009, Identification of peer effects through social networks, *Journal of Econometrics* 150, 41–55.

Brunnermeier, Markus K, Lars Peter Hansen, Anil K Kashyap, Arvind Krishnamurthy, and Andrew W Lo, 2011, Modeling and measuring systemic risk, *SSRN Electronic Journal*.

Brunnermeier, Markus K, and Lasse H Pedersen, 2005, Predatory trading, *The Journal of Finance* 60, 1825–1863.

Burt, Ronald S, 1987, Social contagion and innovation: Cohesion versus structural equivalence, *The American Journal of Sociology* 92, 1287–1335.

Carhart, Mark M, 1997, On persistence in mutual fund performance, *The Journal of Finance* 52, 57–82.

Chen, Qi, Itay Goldstein, and Wei Jiang, 2010, Payoff complementarities and financial fragility: Evidence from mutual fund outflows, *Journal of Financial Economics* 97, 239–262.

Chevalier, Judith, and Glenn Ellison, 1997, Risk taking by mutual funds as a response to incentives, *Journal of Political Economy* 105, 1167–1200.

Choi, Nicole, and Richard W Sias, 2009, Institutional industry herding, *Journal of Financial Economics* 94, 469–491.

Cohen, Lauren, Andrea Frazzini, and Christopher Malloy, 2008, The small world of investing: Board connections and mutual fund returns, *Journal of Political Economy* 116, 951–979.

Cohen, Randolph, Joshua D Coval, and Lubos Pastor, 2005, Judging fund managers by the company they keep, *The Journal of Finance* 60, 1057–1096.

Cohen, Randolph, Christopher Polk, and Bernhard Silli, 2010, Best ideas, *SSRN Electronic Journal*.

Cohen-Cole, Ethan, Andrei Kirilenko, and Eleonora Patacchini, 2010, Are networks priced? network topology and systemic risk in high liquidity markets, *Unpublished working paper*.

Cohen-Cole, Ethan, Eleonora Patacchini, and Yves Zenou, 2011, Systemic risk and network formation in the interbank market, *Carefin Working Paper*.

Conley, Timothy G, 1999, GMM estimation with cross sectional dependence, *Journal of Econometrics* 92, 1–45.

———, and Giorgio Topa, 2002, Socio-economic distance and spatial patterns in unemployment, *Journal of Applied Econometrics* 17, 303–327.

Cooper, Michael J, Huseyin Gulen, and Michael J Schill, 2008, Asset growth and the cross-section of stock returns, *The Journal of Finance* 63, 1609–1651.

Coval, Joshua D, and Erik Stafford, 2007, Asset fire sales (and purchases) in equity markets, *Journal of Financial Economics* 86, 479–512.

Cremers, K. J. Martijn, and Antti Petajisto, 2009, How active is your fund manager? A new measure that predicts performance, *Review of Financial Studies* 22, 3329–3365.

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

Daniel, Kent, and Sheridan Titman, 2006, Market reactions to tangible and intangible information, *The Journal of Finance* 61, 1605–1643.

Dasgupta, Amil, Andrea Prat, and Michela Verardo, 2011a, Institutional trade persistence and long-term equity returns, *The Journal of Finance* 66, 635–653.

——— , 2011b, The price impact of institutional herding, *Review of Financial Studies* 24, 892–925.

Del Guercio, Diane, and Paula A. Tkac, 2002, The determinants of the flow of funds of managed portfolios: Mutual funds vs. pension funds, *Journal of Financial and Quantitative Analysis* 37, 523–557.

Elhorst, J Paul, 2010, Spatial panel data models, *Handbook of Applied Spatial Analysis* pp. 377–407.

Elton, Edwin J, Martin J Gruber, Christopher R Blake, Yoel Krasny, and Sadi O Ozelge, 2010, The effect of holdings data frequency on conclusions about mutual fund behavior, *Journal of Banking and Finance* 34, 912–922.

Elton, Edwin J, Martin J Gruber, and T Clifton Green, 2007, The impact of mutual fund family membership on investor risk, *Journal of Financial and Quantitative Analysis* 42, 257–277.

Ferreira, Miguel A, Aneel Keswani, Antonio F Miguel, and Sofia Brito Ramos, 2011, The flow-performance relationship around the world, *Unpublished working paper.*

Frazzini, Andrea, and Owen A Lamont, 2008, Dumb money: Mutual fund flows and the cross-section of stock returns, *Journal of Financial Economics* 88, 299–322.

Froot, Kenneth A, David S Scharfstein, and Jeremy C Stein, 1992, Herd on the street: Informational inefficiencies in a market with short-term speculation, *The Journal of Finance* 47, 1461–1484.

Gorton, Gary B, and Andrew Metrick, 2011, Securitized banking and the run on repo, *Journal of Financial Economics*.

Greenwood, Robin, and David Thesmar, 2011, Stock price fragility, *Journal of Financial Economics* 102, 471–490.

Hamilton, James, 1994, *Time Series Analysis* (Princeton University Press: Princeton, NJ).

Hayashi, Fumio, 2000, *Econometrics* (Princeton University Press).

Heider, Florian, and Marie Hoerova, 2009, Interbank lending, credit-risk premia, and collateral, *International Journal of Central Banking* 5, 5–43.

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

Israelsen, Ryan, 2010, Does correlated analyst coverage explain excess comovement?, *Unpublished working paper* pp. 1–46.

Jotikasthira, Chotibhak, Christian T Lundblad, and Tarun Ramadorai, 2012, Asset fire sales and purchases and the international transmission of financial shocks, *The Journal of Finance*.

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

Kelejian, Harry H., and Ingmar R. Prucha, 2002, 2SLS and OLS in a spatial autoregressive model with equal spatial weights, *Regional Science and Urban Economics* 32, 691–707.

———, and Yevgeny Yuzefovich, 2006, Estimation problems in models with spatial weighting matrices which have blocks of equal elements, *Journal of Regional Science* 46, 507–515.

Kennedy, Peter, 2003, *Guide to Econometrics* (The MIT Press) 5 edn.

Kiyotaki, Nobuhiro, and John Moore, 1997, Credit cycles, *Journal of Political Economy* 105, 211–248.

Korajczyk, Robert A, and Ronnie Sadka, 2008, Pricing the commonality across alternative measures of liquidity, *Journal of Financial Economics* 87, 45–72.

Kyle, Albert S., 1985, Continuous auctions and insider trading, *Econometrica* 53, 1315–1335.

———, and Wei Xiong, 2001, Contagion as a wealth effect, *The Journal of Finance* 56, 1401–1440.

Legendre, Pierre, and Louis Legendre, 2012, *Numerical Ecology* (Elsevier) 3 edn.

LeSage, James, and R Kelley Pace, 2009, *Introduction to Spatial Econometrics* (Chapman and Hall/CRC: Boca Raton, FL).

Lewellen, Stefan, 2012, Executive compensation and peer effects, *Unpublished working paper* pp. 1–62.

Lou, Dong, 2012, A flow-based explanation for return predictability, *Review of Financial Studies* 25, 3457–3489.

Manski, Charles F, 1993, Identification of endogenous social effects: The reflection problem, *The Review of Economic Studies* 60, 531–542.

Nofsinger, John R, and Richard W Sias, 1999, Herding and feedback trading by institutional and individual investors, *The Journal of Finance* 54, 2263–2295.

Pareek, Ankur, 2009, Information networks: Implications for mutual fund trading behavior and stock returns, *Unpublished working paper*.

Petajisto, Antti, 2011, The index premium and its hidden cost for index funds, *Journal of Empirical Finance* 18, 271–288.

Petersen, Mitchell A, 2009, Estimating standard errors in finance panel data sets: Comparing approaches, *Review of Financial Studies* 22, 435–480.

Pomorski, Lukasz, 2009, Acting on the most valuable information: "best idea" trades of mutual fund managers, *Unpublished working paper* pp. 1–44.

Reca, Blerina, Richard W Sias, and Harry J Turtle, 2011, Hedge fund herding: The apologists' evidence, *SSRN Electronic Journal*.

Scharfstein, David S, and Jeremy C Stein, 1990, Herd behavior and investment, *The American Economic Review* 80, 465–479.

Shleifer, Andrei, and Robert W Vishny, 1997, The limits of arbitrage, *The Journal of Finance* 52, 35–55.

Sias, Richard W, 2004, Institutional herding, *Review of Financial Studies* 17, 165–206.

Simutin, Mikhail, 2009, Excess cash and mutual fund performance, *SSRN Electronic Journal*.

Sirri, Erik R, and Peter Tufano, 1998, Costly search and mutual fund flows, *The Journal of Finance* 53, 1589–1622.

Sloan, Richard G, 1996, Do stock prices fully reflect information in accruals and cash flows about future earnings?, *The Accounting Review* 71, 289–315.

Spiegel, Matthew I, and Hong Zhang, 2013, Mutual fund risk and market share-adjusted fund flows, *Journal of Financial Economics*.

Stein, Jeremy C, 2009, Presidential address: Sophisticated investors and market efficiency, *The Journal of Finance* 64, 1517–1548.

Wahal, Sunil, and Albert (Yan) Wang, 2011, Competition among mutual funds, *Journal of Financial Economics* 99, 40–59.

Wasserman, Stanley, and Katherine Faust, 1994, *Social Network Analysis: Methods and Applications* (Cambridge University Press).

Wermers, Russ, 2003, Is money really "smart"? New evidence on the relation between mutual fund flows, manager behavior, and performance persistence, *Unpublished working paper*.

Zheng, Lu, 1999, Is money smart? A study of mutual fund investors' fund selection ability, *The Journal of Finance* 54, 901–933.

V. Appendix

A. Exploratory Factor Analysis Results

As stated in the main paper, there is evidence of a 'flow' type factor which is predictive of mutual fund returns. Sirri and Tufano (1998) first proposed using average flows of a fund's category as predictive; more recently, Lou (2012) has developed a measure of flow-induced trading to account for price pressure on underlying securities based on mutual fund flows, though he neglects to account for Style Flows as I do here and Sirri and Tufano suggest.

I am proposing a new flow-based measure which should predict mutual fund returns, so in addition I provide some additional analysis showing A. Flow-based factors provide new information above and beyond existing factor models such as the Carhart four-factor model or DGTW model and B. Peer Flow, my flow factor, is the best proxy for a flow factor in predicting mutual fund returns.

First, a description of the candidate flow factors. Style Flow is simply the average flow of all mutual funds in a given style as stated in the text. Peer Flow is also as described in the text and later in this appendix.

Lou (2012) computes a flow-induced trading (FIT) measure for mutual funds. This measure accumulates the total effect of "flow-induced trading" across all mutual funds on their holdings to compute a stock-level measure which he then applies to mutual funds using portfolio weights. Because contemporaneous fund flows are mechanically related to fund returns in his study (though not in my data), he computes an expected flow-induced trading (E[FIT]) measure where expected flows predicted solely using lagged Carhart four-factor alpha because this captures the flow-performance relationship without capturing too much of the time series dynamics of fund flows. Indeed, Lou even says, "An intuitive way to interpret this measure is that if we think of the entire mutual fund industry as one giant fund, FIT then captures the magnitude of flow-induced trading by this aggregate fund." My goal is to be more precise by focusing on the fund flows of a funds close peers rather than aggregating across all mutual funds, and so I expect my Peer Flow measure to more cleanly predict mutual fund performance.

Thus, I analyze four flow factors: Peer Flow, Style Flow, Lagged FIT, and E[FIT]. To investigate the commonality among these flow measures, I turn to Exploratory Factor Analysis (EFA). EFA attempts to reduce multiple observed factors in to a reduced set of unobservable latent factors by investigating common variation. I use EFA to choose the correct number of factors and then assign the multiple observed factors to each 'core' factor, thus assisting in identifying the underlying latent variable of interest. EFA then provides for each variable a score of "commonality" which is the variable's overlap with the common factor and "uniqueness" which is a score of remaining variability not explained by the common factor. [39]

[39] A full discussion of factor analysis is beyond the scope of this article, for more information

Initial results are in Table A1, which show that using the Carhart four-factor model only, we need two factors, but when we include the set of mutual fund flow variables (Peer Flow, Style Flow, Expected Flow-Induced Trading and Lagged Flow-Induced Trading) we need an additional factor as indicated by the additional eigenvector above 1, with the breakpoint moving downward.

Table A1: Eigenvectors from Factor Analysis of Mutual Fund Return predictor variables

	With Flow Variables	Carhart Four Factors Only
Factor1	1.604983	1.458651
Factor2	1.490384	1.138413
Factor3	1.226093	.0002345
Factor4	.0728658	-.000289
Factor5	.040284	
Factor6	.0045403	
Factor7	-.0466907	
Factor8	-.0710479	

Next, we investigate the loadings of each predictor variable on the factors, starting with just the Carhart four-factor model in Table A2. Note that factors less than 0.3 not displayed to aid in identifying meaningful factor loadings. Factor loadings in this table are similar to a correlation or Beta between the variable and the latent factor. Uniqueness is the amount of unique variation remaining not explained by the factors. Here, the first factor is dominated by the book to market factor, the second by the market return. The size and momentum factors both load significantly on factor 1 and the momentum factor also loads on factor 2 but both have large unexplained uniqueness.

see Bartholomew, Knott, and Moustaki (2011). It is similar to what Korajczyk and Sadka (2008) do to explore commonality in liquidity, though less rigorous. EFA is also distinct from Principal Component Analysis which seeks to explain variance but does not take into account variance due to fundamentals vs measurement error. It also gives greater weight to high variance processes, thus potentially emphasizing variables with higher variance due to higher measurement error.

Table A2: Factor Loadings from factor analysis on Carhart four factor model variables only.

	Factor1	Factor2	Uniqueness
Excess Mkt Return		-.8098018	.2572426
SMB Factor	.5016491		.7415112
HML Factor	-.996438		.0030609
UMD Factor	.355363	.6874562	.4011212

Finally, in Table A3 we have the factor analysis including flow factors. We see clearly that there is a distinct 'Flow' factor for which my Peer Flow measure has the highest loading (and lowest uniqueness) indicating a high degree of overlap. Style Flow also has a large loading as does the measured, lagged FIT measure. The Expected FIT measure (predicted with lagged Carhart Alpha) does not load significantly on the flow factor (or any factor). This does not indicate that it is not useful, only that it is less correlated with a broad measure of fund flows. In total, this analysis indicates the need for a 'Flow' factor in the prediction of mutual fund performance and that my Peer Flow measure is an excellent candidate.

Table A3: Factor Loadings from factor analysis on Carhart four factor model variables including mutual fund flow variables.

	Factor1	Factor2	Factor3	Uniqueness
Excess Mkt Return		-.3522781	-.6641007	.4187848
SMB Factor		-.4532773		.7785095
HML Factor		1.044188		-.0948019
UMD Factor			.8761278	.1746024
Peer Flow	.9846845			.027369
Lagged FIT Measure	.4558635			.780215
E[FIT*] Measure				.9477101
Style Flow	.5942485			.6461514

Since EFA is not a standard from of analysis in financial economics, I present Table A4 as well.

Table A4: The effect of different flow-based variables on mutual fund returns.

The dependent variable is the excess return over the DGTW return. DGTW return is the quarterly return of the portfolio weighted stock-level DGTW returns, weighted by mutual fund stock holdings. Peer Flow is the weighted average flow of mutaul fund peers defined by a common holdings network. Network relation for peers is the normalized dot product, and peer effects are the weighted average of peer characteristics. Flow is the fund flow divided by total net assets. E[FIT] with Alpha is the E[FIT] measure computed as in (Lou, 2012) (flows predicted using lagged Carhart four-factor alpha). FIT is also computed as in (Lou, 2012) using measured flows. Data is quarterly from 1998 to 2009, each panel variable is any open ended fund holding a nonzero equity position. Style Flow is the average of all reported fund flows by Morningstar category. Time and Fund Fixed Effects included. Hansen J stat is a test of overidentification for which the null hypothesis is that instruments are uncorrelated with stage 2 regression, KP LM stat tests the null of weak instruments. T statistics are in parentheses and significance is denoted at the 1, 5, and 10% level.

	(1) Ex DGTW Ret	(2) Ex DGTW Ret	(3) Ex DGTW Ret	(4) Ex DGTW Ret	(5) Ex DGTW Ret
Peer Flow	0.9400***			0.9730***	1.0303***
	(4.61)			(4.52)	(4.44)
Style Flow	0.2229***	0.3561***	0.3611***	0.2209***	0.2213***
	(6.17)	(6.63)	(6.63)	(6.07)	(6.20)
E[FIT] with Alpha		-0.3055		-0.7396	
		(-0.41)		(-0.96)	
Lag1 Fund FIT			-0.3382		-0.5480
			(-0.97)		(-1.47)
Fund Size	-0.0033***	-0.0023*	-0.0020	-0.0033***	-0.0030**
	(-2.68)	(-1.75)	(-1.47)	(-2.73)	(-2.35)
Cash Pct	0.0439**	0.0503**	0.0529**	0.0443**	0.0470**
	(2.08)	(2.22)	(2.43)	(2.11)	(2.32)
Amihud Illiq	0.0003	0.0002	0.0002	0.0003	0.0004

Continued on next page...

	(1) Ex DGTW Ret	(2) Ex DGTW Ret	(3) Ex DGTW Ret	(4) Ex DGTW Ret	(5) Ex DGTW Ret
	(0.30)	(0.18)	(0.18)	(0.29)	(0.32)
Lag1 Flow	-0.0005	-0.0005	0.0022	-0.0005	0.0039
	(-0.19)	(-0.16)	(0.66)	(-0.20)	(1.17)
Lag2 Flow	-0.0010	-0.0019	-0.0013	-0.0011	-0.0000
	(-0.29)	(-0.56)	(-0.37)	(-0.35)	(-0.00)
Lag3 Flow	-0.0013	-0.0021	-0.0013	-0.0014	-0.0002
	(-0.59)	(-0.83)	(-0.54)	(-0.65)	(-0.10)
Lag4 Flow	-0.0063**	-0.0056**	-0.0051**	-0.0063**	-0.0056**
	(-2.40)	(-2.16)	(-2.06)	(-2.42)	(-2.28)
Lag1 Return	-0.0237	0.0349	0.0445	-0.0218	-0.0082
	(-0.47)	(0.62)	(0.72)	(-0.43)	(-0.15)
Lag2 Return	-0.0646	-0.0301	-0.0195	-0.0616	-0.0469
	(-1.50)	(-0.68)	(-0.40)	(-1.49)	(-1.07)
Lag3 Return	-0.0759*	-0.0508	-0.0457	-0.0719	-0.0656
	(-1.75)	(-1.11)	(-0.96)	(-1.64)	(-1.46)
Lag4 Return	-0.1299***	-0.1104**	-0.1068**	-0.1257***	-0.1230***
	(-2.83)	(-2.31)	(-2.20)	(-2.76)	(-2.67)
Observations	67193	67193	67184	67193	67184
R Squared	0.10	0.06	0.06	0.10	0.11
Fund clusters	4,185	4,185	4,184	4,185	4,184
Time clusters	44	44	44	44	44

I only present results here for Excess DGTW return, but the results for raw return controlling for Excess DGTW return are similar. These confirm the above factor analysis showing that Peer Flow and Style Flow are the two important flow factors, and the FIT measure, whether measured or expected does not enter significantly into the regression.

B. Details on how to compute Similarity Measure

As noted in the text, similarity between two portfolios i and j, denoted s_{ij}, as the dot product between the security holding weight vectors of each portfolio manager i and j, divided by the product of the Euclidean norm of each vector. s_{ij} is computed at each time t, but I suppress the time subscript for expositional ease. Specifically, where h_i is a vector of portfolio weights for manager i, the similarity between managers i and j is defined as

$$s_{ij} = \frac{h_i \cdot h_j}{|h_i|\,|h_j|} \tag{9}$$

For each manager i, the Euclidean norm is defined across M securities as

$$|h_i| = \sqrt{\sum_{m=1}^{M} h_{im}^2} \tag{10}$$

Deriving this same measure in matrix form, let H be the $M \times N$ holdings matrix, with portfolio managers as each column, and each row consisting of the weight between 0 and 1 each manager places on that security. My portfolio similarity matrix is then

$$S = \frac{H^T H}{|H| \cdot |H|} \tag{11}$$

in which each s_{ij} already defined above is an element of symmetric similarity matrix $S_{N\times N}$ and h_i are the columns of H. The norm of the matrix H is a Euclidean column norm, such that for each column j, the norm of H_j is defined as

$$|H_j| = \sqrt{\sum_{m=1}^{M} h_{jm}^2} \tag{12}$$

I then compute *Peer Flow* as the dot product of the weight vector and the corresponding vector of fund flows for each manager. Formally, peer weights are computed as

$$PeerWeight_{ij} = \frac{s_{ij}}{\sum_k s_{ik}}, k \neq i \tag{13}$$

and *Peer Flow* is thus

$$PeerFlow_i = \sum_k PeerWeight_{ik} Flow_k \tag{14}$$

In matrix form, if W is a row-stochastic transformation of S, such that each row sums to 1, then $PeerFlow = W \cdot Flow$ in which both $PeerFlow$ and $Flow$ are $N \times 1$ vectors and W is an $N \times N$ matrix at time t.

Two-step neighbors are computed as $B = S^2$, with matrix multiplication (as opposed to element-by-element) where the diagonal of S has already been set to 0 to avoid duplicating one-step and two-step neighbors.[40] In summation notation, the equivalent product is

$$b_{ij} = \sum_{q=1}^{N} s_{iq} s_{qj},\ q \neq i, j \tag{15}$$

with the diagonal of B also set to zero such that a manager cannot be his own two-step neighbor.[41] If $\widetilde{W}$ is the row-stochastic, $N \times N$, two-step weighting matrix derived from B, then $TwoStepPeerFlow = \widetilde{W} \cdot Flow$ or as a summation:

$$\widetilde{w}_{ji} = \frac{b_{ji}}{\sum_k b_{jk}} \tag{16}$$

$$TwoStepPeerFlow_j = \sum_k \widetilde{w}_{jk} Flow_k \tag{17}$$

C. Identification and Estimation of a Network Influence Process

This discussion addresses identification of peer effects using the language common to the peer effects literature. Because this language is not intuitive to a general finance audience, I discussed it in terms more broadly understood in the main text, but this more technical discussion is available for those more steeped in the complications involved in identifying peer effects.

Inference on networks is complicated by some unique identification problems, notably discussed in Manski (1993).[42] According to Manski, identifying an endogenous social influence process requires controlling for two other potential confounding effects: correlated effects and contextual effects.[43]

[40]A nonzero diagonal indicates a 'self-loop.' If S has a nonzero diagonal, a 'two-step' neighbor could be i connecting to i (a self loop) and then i connecting to j, which therefore duplicates a one-step neighbor. This is a common adjustment in network analysis.

[41]The diagonal of B must now be set to 0 because for every one-step neighbor, a manager is his own two-step neighbor. For instance, i connects to j, but then j also connects back to i, such that for every connection like this i is his own two-step neighbor.

[42]Thanks to Ethan Cohen-Cole for explaining the nuances of these different effects to me.

[43]Bramoullé, Djebbari, and Fortin (2009) also note that these controls are a necessary prerequisite for their instrumentation approach.

Correlated effects can be conceived as a co-integrated relation when a relatively fixed relation among two neighbors induces a proportional response to exogenous events. Correlated effects simultaneously affect two connected managers due to common, but time-invariant, characteristics. For example, two mutual funds, one half the size of the other, may find that on average the smaller fund receives half the fund flows of the large one. Since there may be a similar relation due to cash holdings, I include *Peer Size* and *Peer Cash* to control for these potentially common fund characteristics which may drive correlated flows.[44]

Contextual effects can be viewed as a network version of industry effects, in which market-wide trends affect all members of the group equally, but may change through time. For instance, a sector rotation strategy which suggests buying utilities and health care stocks in a declining market represents a wider shift in investor behavior, possibly inducing a spurious correlation among related mutual funds. I control for Manski's contextual effects by including *StyleFlows*, which represents the average flow for the Morningstar category to which each open-ended fund belongs.

A further identification problem may arise due to network density, as noted by Kelejian, Prucha, and Yuzefovich (2006). If a network is very dense or "complete" such that each agent is equally connected, then each agent would have exactly the same *Peer Flow* measure. For example, assume that each portfolio manager is connected to each other manager with a weight of exactly 1. This would make *Peer Flow* equal to the average market-wide flow, since the weight on each flow variable would be $\frac{1}{N}$ for every manager, and therefore no longer display cross-sectional variation. Given that my weighted density is less than 5%, this is unlikely to be a problem, but as a further robustness check, I have run my specification thresholding my network at the 80^{th} percentile, thus obtaining an unweighted density of 10% with no material change in results.[45]

Finally, I produce estimates using Generalized Method of Moments, whereas most specifications of this type in the spatial econometrics literature estimate models via Maximum Likelihood. Conley (1999) notes that maximum likelihood specifications in which spatial dependence is measured with error are misspecified. While measurement error is unlikely to be a problem with geographical measures of distance typical of the spatial econometrics literature, my measure of distance in security space may be much less precise. Fortunately, Kelejian and Prucha (2002) show that with panel data, both OLS and GMM estimators are consistent, and thus represent the appropriate estimation approach. Elhorst (2010) includes

[44] *Peer Size* and *Peer Cash* are obviously not time-invariant, but are quite persistent. Typical traits used to control for correlated effects are things like gender and race, which are time-invariant but do not have a clear equivalent among mutual funds.

[45] Weighted density is the sum of all network connections in the network divided by the sum of all possible network connections set to 1, or N^2. Unweighted density is the same, but sets any weighted network link to 1 first. On average, the unweighted density of one of my un-thresholded networks is approximately 80%, which is quite high and would be a problem if my analysis ignored the weights.

a short discussion on MLE vs IV/GMM estimators, noting that while the use of IV/GMM is promising, it is still new to the spatial econometrics literature and needs further research.[46]

[46]Spatial Econometrics primarily uses MLE because they have only one network – geography – and thus no panel data. In these cases, only MLE is appropriate (Elhorst, 2010).

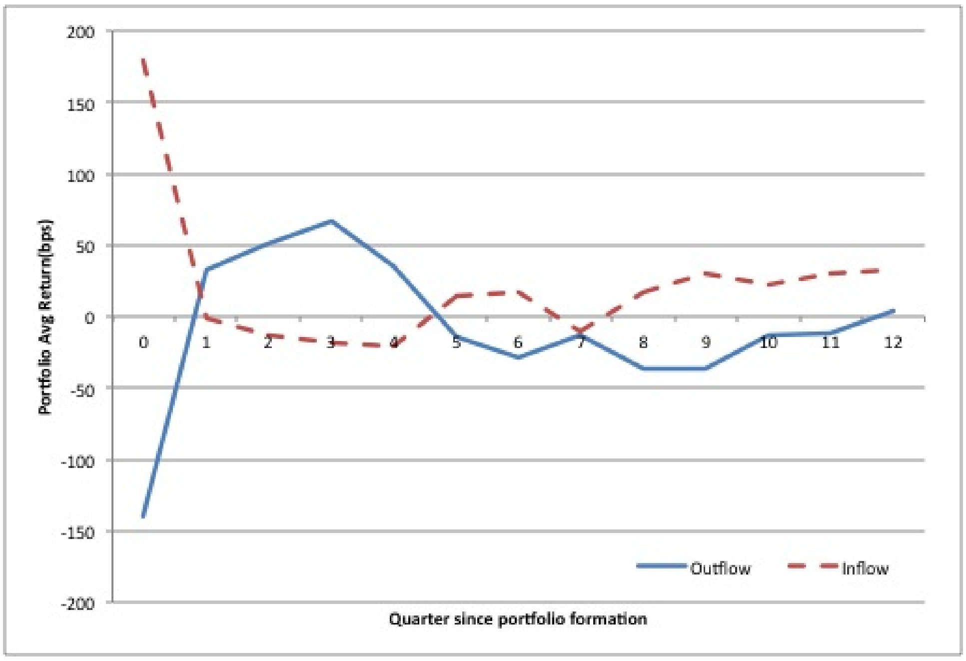

Figure 1: Performance of Peer Flow portfolios through time. Portfolios based on Peer Flow are formed at time 0. Outflow is defined as the lower two quintiles, Inflow the upper two quintiles. Portfolios held for the subsequent 12 quarters. Return is Excess DGTW return in basis points. Data is quarterly from 1998 to 2009, each panel variable is any open ended fund holding a nonzero equity position.

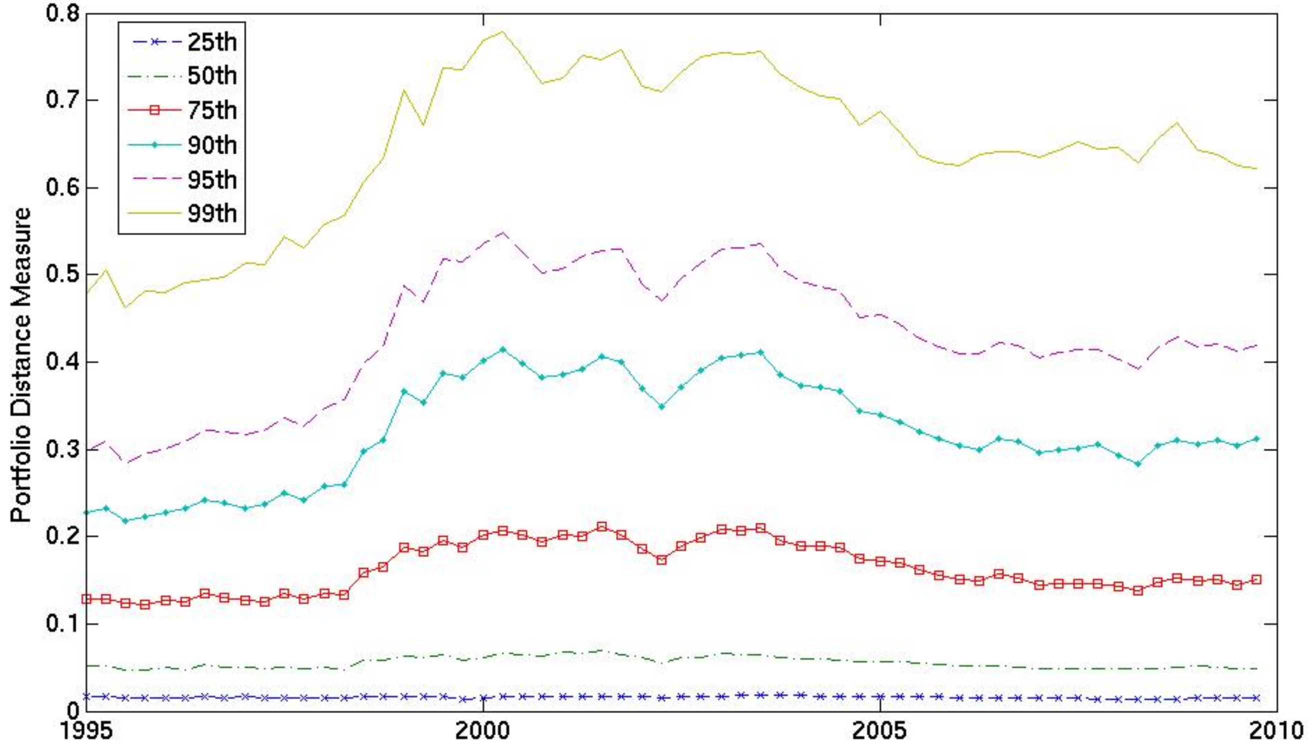

Figure 2: Percentiles of portfolio distance measure through time. Each line represents the time series of a specific percentile of the cross-sectional distribution of the normalized dot product portfolio distance measure. The top is the 99^{th} percentile, then the 95^{th}, 90^{th}, 75^{th}, 50^{th}, and 25^{th}.

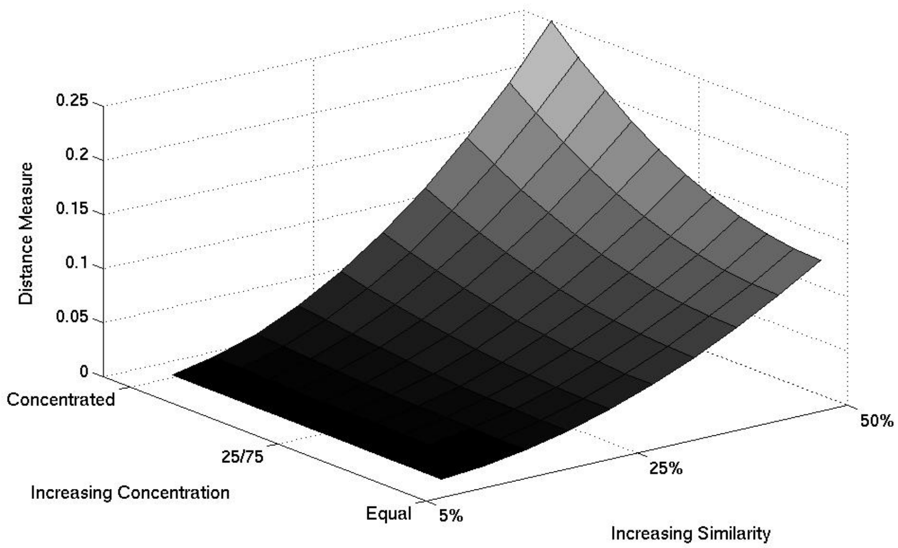

Figure 3: Two dimensions of the portfolio distance measure. This plot of simulated data demonstrates how interconnectedness as measured by the normalized dot product between two portfolios of two securities increases in two different dimensions. Plotted is the distance between two managers who each hold the same two asset in common and a third asset not in common. The axis on the right is increasing in portfolio overlap, i.e. the percentage of the portfolio that overlaps. The axis on the left is increasing in the concentration of that position, holding the percentage of overlap constant. So a 'concentrated' position is where the entire portfolio overlap is in one security, and an 'equal' position is 50% each of the two securities. More concentrated positions are thus closer in security space, holding overlap constant.

Table I:

Porfolio sorts of Mutual Fund Return across Peer Flow and Flow quintiles. Each quarter, mutual funds are independently sorted into quintiles by Flow (dollar flow divided by total net assets) and Peer Flow. The average mutual fund return is summarized below for each Flow/Peer Flow quintile portfolio, as well as the average across portfolios. Each 1-5 quintile difference is statistically significant at the 1% level. Returns are quarterly in basis points. Excess DGTW return is computed as the mutual fund return less the portfolio weighted stock-level DGTW return. Carhart alpha is the intercept of a rolling 12 month regression of mutual fund returns on a four factor model. Excess market return is computed as the mutual fund return less the market return.

(a) Excess DGTW Return

	Fund Flow Portfolios					
	Outflow				Inflow	
Peer Flow Portfolios	1	2	3	4	5	All
Outflow 1	-438.6	-284.6	-221.0	-229.8	-278.1	-314.5
2	-155.9	-71.6	-39.6	-39.4	-44.1	-75.7
3	-2.0	80.4	115.7	110.2	143.0	91.7
4	-12.1	70.9	124.9	133.5	175.4	108.6
Inflow 5	36.6	130.5	169.5	221.9	418.2	241.6
All	-179.2	-37.5	35.5	66.4	141.0	

(b) Carhart Alpha

	Outflow				Inflow	
Peer Flow Portfolios	1	2	3	4	5	All
Outflow 1	-43.6	-24.3	-23.3	-16.4	-5.9	-27.4
2	-28.8	-12.3	-4.6	-1.4	6.5	-10.5
3	-35.6	-18.6	-10.8	-5.9	2.7	-13.8
4	-30.1	-19.0	-11.2	-2.2	13.4	-8.3
Inflow 5	-0.6	0.4	16.3	32.5	61.9	30.1
All	-31.9	-16.1	-6.8	4.2	23.4	

(c) Excess Market Return

	Outflow				Inflow	
Peer Flow Portfolios	1	2	3	4	5	All
Outflow 1	-272.2	-115.2	-96.8	-101.8	-105.8	-160.9
2	-161.5	-86.4	-47.9	-44.8	-29.0	-81.1
3	-84.5	-30.5	-3.3	13.2	42.8	-11.7
4	-22.4	39.8	66.3	81.7	148.4	70.1
Inflow 5	112.2	178.4	216.8	257.0	441.4	282.2
All	-133.5	-24.5	27.6	66.9	162.5	

Table II: Summary statistics

Data is quarterly from 1998 to 2009, each panel variable is any open ended fund holding a nonzero equity position. Funds with Total Net Assets less than \$1M are discarded. Flow is the fund flow divided by total net assets. Return is the mutual fund return (gross). Carhart alpha is the intercept of a rolling 12 month regression of mutual fund returns on a four factor model. DGTW return is the quarterly return of the portfolio weighted stock-level DGTW returns, weighted by mutual fund stock holdings. Cash Pct is cash holdings divided by total net assets. Amihud Illiq is the portfolio weighted sum of equity holdings' Amihud measures computed over the previous quarter, logged. Style Flow is the average of all reported fund flows by Morningstar category. E[FIT*] (Alpha) is the E[FIT*] measure computed as in (Lou, 2012) (flows predicted using lagged Carhart four-factor alpha). E[FIT*] (Peer Flow) is the E[FIT*] measure from (Lou, 2012) except with flows predicted using Peer Flow. FIT Measure is as computed as in (Lou, 2012) with realized fund flows. Peer Flow is the weighted average flow of mutaul fund peers defined by a common holdings network. Network relation for peers is the normalized dot product, and peer effects are the weighted average of peer characteristics. Pr Fund Size is the peer weighted total net assets, logged. Pr Cash Pct is the peer weighted cash holdings divided by total net assets.

Variable	Mean	Std. Dev.	Min.	Max.	N
Flow	0.009	0.166	-1	0.735	124,640
Return	0.017	0.118	-0.99	1.648	135,687
Carhart Alpha	-0.001	0.011	-0.14	0.137	129,551
DGTW Return	0.015	0.079	-2.974	9.912	123,397
Total Net Assets ($M)	1,043	4,470	1.008	193,453	138,073
Cash Pct	0.035	0.041	-0.003	0.535	138,073
Amihud Illiq	-13.458	2.035	-30.567	-4.675	130,540
Style Flow	0.01	0.043	-1	0.735	128,418
E[FIT*] (PeerFlow)	0.001	0.013	-0.466	0.772	126,990
E[FIT*] (Alpha)	-0.001	0.004	-0.395	0.109	126,990
FIT Measure	-0.005	0.012	-0.506	0.49	126,995
Peer Flow	0.007	0.021	-0.488	0.484	128,440
Pr Fund Size	20.86	0.487	12.435	22.814	138,073
Pr Cash Pct	0.042	0.009	0	0.351	138,073

Table III: Persistence of Network Distance Relation

Network relation is the normalized dot product, and is the dependent variable. Results shown from Fama-MacBeth regression of eight lags of network connectivity. Data is quarterly from 1998 to 2009. Significance is denoted at the 1, 5, and 10% level.

	Coeff Estimate	Std Dev	T statistic
Lag 1	0.4138***	0.1255	3.2972
Lag 2	0.2500***	0.0830	3.0112
Lag 3	0.0765*	0.0454	1.6835
Lag 4	0.0851	0.0535	1.5918
Lag 5	0.0189	0.0395	0.4785
Lag 6	0.0407	0.0495	0.8222
Lag 7	0.0110	0.0392	0.2802
Lag 8	0.0460	0.0488	0.9418

Table IV: First Stage Regression of GMM/IV Specification

First stage regressions with endogenous regressors as dependent variables. Peer Flow is the weighted average of peer connected flow, and Two Step Peer Flow is the weighted average of their neighbor's neighbors flow, used as instruments. Flow is dollar flows divided by total net assets and Cash Pct is cash holdings divided by total net assets. Size is log of total net assets. Amihud is a portfolio weighted measure of the Amihud values of equity holdings, logged. Style Flow is the average flow within each Morningstar category. Data is quarterly from 1995 to 2009. Time and Fund Fixed Effects included. T statistics are in parentheses and significance is denoted at the 1, 5, and 10% level.

	(1) Peer Flow	(2) Peer Flow
Two Step Peer Flow	1.4053***	1.3791***
	(60.30)	(50.79)
Two Step Peer Flow^2	1.3262	1.0457
	(1.31)	(1.03)
Lag1 Alpha	0.0319***	0.0326***
	(4.13)	(4.30)
Lag1 Flow	0.0018***	0.0017***
	(6.49)	(6.52)
Lag2 Flow	0.0009***	0.0008***
	(3.36)	(3.14)
Lag3 Flow	0.0003	0.0002
	(1.23)	(0.91)
Lag4 Flow	0.0003*	0.0002
	(1.75)	(0.97)
Fund Size	0.0002*	0.0002*
	(1.65)	(1.81)
Cash Pct	0.0069***	0.0037***
	(4.64)	(3.08)
Amihud Illiq	0.0002***	0.0002***
	(4.31)	(4.13)
Style Flow	0.0465***	0.0457***
	(9.26)	(9.35)
Pr Fund Size		0.0009
		(0.83)
Pr Cash Pct		0.2678***
		(6.13)
Observations	73319	73319
R Squared	0.87	0.87
Fund clusters	4,428	4,428
Time clusters	44	44

Table V: Performance of the top 10 holdings of mutual funds in different Peer Flow Terciles. Each tercile contains the top 10 stocks, by weight, held exclusively by mutual funds in that tercile. Thus, stocks not in the top 10 holdings of any fund are excluded, as well as top 10 stocks held by mutual funds in differing terciles so not all stocks are included. Peer Flow Tercile 1 is the bottom tercile of peer flow, meaning the fund experienced significant outflow. Peer Flow Tercile 3 is the top tercile of peer flow, meaning the fund experienced significant inflow. Peer Flow Tercile 2 is the middle tercile of peer flow, meaning the fund experienced average net flow of zero. Returns are in percent, quarterly. T statistics are in parentheses and significance is denoted at the 1, 5, and 10% level

Peer Flow Tercile	Excess DGTW Return	Raw Return
1 (Outflow)	-0.457%	-0.267%
2	0.958%	2.93%
3 (Inflow)	1.82%	6.91%
1 minus 2	-1.42%***	-3.20%***
	(-4.45)	(-9.14)
1 minus 3	-2.28%***	-7.18%***
	(-6.63)	(-19.14)
2 minus 3	-0.862%***	-3.99%***
	(-2.73)	(-11.47)

Table VI: Return regression: effect of Peer Flow on Mutual Fund Returns

In models 1-5, the dependent variable is mutual fund raw return, in models 6-10, it is the excess return over the DGTW return. DGTW return is the quarterly return of the portfolio weighted stock-level DGTW returns, weighted by mutual fund stock holdings. Peer Flow is the weighted average flow of mutaul fund peers defined by a common holdings network. Network relation for peers is the normalized dot product, and peer effects are the weighted average of peer characteristics. Flow is the fund flow divided by total net assets. Lagged flows are the mutual fund's own flows, lagged quarterly. Lagged returns are raw mutual fund returns, also quarterly. Data is quarterly from 1998 to 2009, each panel variable is any open ended fund holding a nonzero equity position. Funds with Total Net Assets less than $1M are discarded. Fund Size is the log of total net assets. Cash Pct is cash holdings divided by total net assets. Amihud is the portfolio weighted sum of equity holdings' Amihud measures computed over the previous quarter, logged. Style Flow is the average of all reported fund flows by Morningstar category. Time and Fund Fixed Effects included. Hansen J stat is a test of overidentification for which the null hypothesis is that instruments are uncorrelated with stage 2 regression, KP LM stat tests the null of weak instruments. T statistics are in parentheses and significance is denoted at the 1, 5, and 10% level.

	(1) Excess DGTW Ret	(2) Excess DGTW Ret	(3) Excess DGTW Ret	(4) Excess DGTW Ret	(5) Excess DGTW Ret	(6) Return	(7) Return	(8) Return	(9) Return	(10) Return
Peer Flow	0.8646***	0.8991***	0.6583***	0.6225***	0.9400***	1.3056***	1.3180***	1.0636***	1.0510***	1.3451***
	(3.77)	(3.85)	(3.00)	(2.90)	(4.61)	(4.67)	(4.90)	(4.35)	(4.14)	(5.80)
DGTW Ret						0.2479**	0.2542**	0.2534**	0.2473**	0.4400***
						(1.96)	(1.98)	(1.98)	(1.97)	(3.63)
Fund Size		-0.0027**	-0.0030**		-0.0033***		-0.0040**	-0.0044***		-0.0038***
		(-2.02)	(-2.28)		(-2.68)		(-2.48)	(-2.72)		(-2.61)
Cash Pct		0.0024	-0.0034		0.0439**		0.0043	-0.0010		0.0365**
		(0.11)	(-0.17)		(2.08)		(0.30)	(-0.07)		(2.14)
Amihud Illiq		-0.0006	-0.0006		0.0003		0.0014**	0.0015**		0.0013*
		(-0.51)	(-0.46)		(0.30)		(1.99)	(2.09)		(1.92)
Style Flow			0.2012***	0.2051***	0.2229***			0.1952***	0.1982***	0.2267***
			(6.31)	(6.23)	(6.17)			(5.36)	(5.32)	(5.92)
Lag1 Flow					-0.0005					0.0000

Continued on next page...

	(1) Excess DGTW Ret	(2) Excess DGTW Ret	(3) Excess DGTW Ret	(4) Excess DGTW Ret	(5) Excess DGTW Ret	(6) Return	(7) Return	(8) Return	(9) Return	(10) Return
					(-0.19)					(0.01)
Lag2 Flow					-0.0010					-0.0007
					(-0.29)					(-0.22)
Lag3 Flow					-0.0013					-0.0026
					(-0.59)					(-0.94)
Lag4 Flow					-0.0063**					-0.0075***
					(-2.40)					(-2.64)
Lag1 Return					-0.0237					-0.0505
					(-0.47)					(-0.84)
Lag2 Return					-0.0646					-0.0912*
					(-1.50)					(-1.79)
Lag3 Return					-0.0759*					-0.0780
					(-1.75)					(-1.64)
Lag4 Return					-0.1299***					-0.1423***
					(-2.83)					(-2.58)
Observations	111920	111405	111392	111907	67193	111920	111405	111392	111907	67193
R Squared	0.03	0.03	0.04	0.04	0.10	0.12	0.13	0.14	0.13	0.22
Fund clusters	5,308	5,273	5,273	5,308	4,185	5,308	5,273	5,273	5,308	4,185
Time clusters	48	48	48	48	44	48	48	48	48	44
Est Method	GMM	GMM	GMM	GMM	GMM	GMM	GMM	GMM	GMM	GMM
Hansen J stat	0.26	0.03	0.00	0.12	0.62	0.04	0.12	0.28	0.00	0.30
J p value	0.6075	0.8644	0.9577	0.7273	0.4312	0.8327	0.7280	0.5969	0.9698	0.5826
KP LM Stat	28.29	29.11	31.06	30.45	32.29	28.65	29.58	31.45	30.75	32.72
KP LM p value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table VII: Return reversal of peer flow effect in the following year.

In models 1-4, the dependent variable is excess return over the DGTW return, in models 5-8, it is the mutual fund raw return. In all cases, the dependent variable is the leading annual return, defined as the holding period return for the subsequent four quarters after time t. Leading DGTW ret is the annual return of the portfolio weighted stock-level DGTW returns, weighted by mutual fund stock holdings, computed the same way. All other variables are quarterly and defined as in Table VI.

	(1) Excess DGTW Ret	(2) Excess DGTW Ret	(3) Excess DGTW Ret	(4) Excess DGTW Ret	(5) Return	(6) Return	(7) Return	(8) Return
Peer Flow	-1.1465***	-0.8309***	-0.8083***	-1.0923***	-1.2292***	-0.9077***	-0.8597***	-1.1458***
	(-3.83)	(-2.91)	(-2.88)	(-3.82)	(-3.95)	(-3.07)	(-3.13)	(-4.07)
Leading DGTW Ret					0.6596***	0.6439***	0.6447***	0.6595***
					(2.93)	(2.86)	(2.86)	(2.92)
Fund Size		-0.0334***	-0.0333***			-0.0352***	-0.0352***	
		(-13.37)	(-13.42)			(-11.86)	(-11.96)	
Cash Pct		0.0231	0.0232			0.0245	0.0250	
		(0.55)	(0.56)			(0.61)	(0.63)	
Amihud Illiq		0.0048*	0.0048*			0.0039*	0.0039*	
		(1.87)	(1.87)			(1.82)	(1.81)	
Style Flow			-0.0166	-0.0416			-0.0356	-0.0631
			(-0.30)	(-0.72)			(-0.55)	(-0.93)
Observations	89727	89498	89490	89719	89727	89498	89490	89719
R Squared	0.01	0.04	0.04	0.01	0.23	0.26	0.26	0.23
Fund clusters	4,521	4,496	4,496	4,521	4,521	4,496	4,496	4,521
Time clusters	44	44	44	44	44	44	44	44
Est Method	GMM	GMM	GMM	GMM	GMM	GMM	GMM	GMM
Hansen J stat	0.29	0.60	0.60	0.30	0.28	0.57	0.58	0.31
J p value	0.5921	0.4399	0.4371	0.5821	0.5949	0.4501	0.4444	0.5796
KP LM Stat	26.36	26.73	28.40	28.24	26.50	26.84	28.47	28.33
KP LM p value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table VIII: Flow regression to measure spillover externalities.
Flow is the dependent variable and is the fund flow divided by total net assets. Data is quarterly from 1998 to 2009, each panel variable is any open ended fund holding a nonzero equity position. Funds with Total Net Assets less than $1M are discarded. Peer Flow is the weighted average flow of mutaul fund peers defined by a common holdings network. Network relation for peers is the normalized dot product, and peer effects are the weighted average of peer characteristics. Fund Size is the log of total net assets. Cash Pct is cash holdings divided by total net assets. Amihud is the portfolio weighted sum of equity holdings' Amihud measures computed over the previous quarter, logged. Style Flow is the average of all reported fund flows by Morningstar category. Time and Fund Fixed Effects included. Hansen J stat is a test of overidentification for which the null hypothesis is that instruments are uncorrelated with stage 2 regression, KP LM stat tests the null of weak instruments. T statistics are in parentheses and significance is denoted at the 1, 5, and 10% level.

	(1) Flow	(2) Flow	(3) Flow	(4) Flow	(5) Flow
Peer Flow			0.5785***	0.3268**	0.4740***
			(4.10)	(2.43)	(3.28)
Lag1 Flow	0.0856***	0.0594***	0.0641***	0.0580***	0.0580***
	(4.83)	(3.43)	(3.75)	(3.40)	(3.39)
Lag2 Flow	0.0919***	0.0771***	0.0814***	0.0765***	0.0769***
	(5.47)	(5.10)	(5.76)	(5.24)	(5.31)
Lag3 Flow	0.0370***	0.0272***	0.0303***	0.0270***	0.0274***
	(3.47)	(2.63)	(3.08)	(2.71)	(2.75)
Lag4 Flow	0.0114	0.0040	0.0039	0.0038	0.0040
	(1.06)	(0.36)	(0.35)	(0.34)	(0.36)
Lag1 Alpha	1.1627***	0.9102***		0.8572***	0.8623***
	(7.25)	(7.34)		(7.13)	(7.24)
Fund Size		0.0154***	0.0148***	0.0150***	0.0156***
		(5.85)	(5.50)	(5.64)	(5.85)
Cash Pct		0.5843***	0.5822***	0.5796***	0.5821***
		(17.44)	(17.38)	(17.36)	(17.44)
Amihud Illiq		0.0024***	0.0016**	0.0027***	0.0019***
		(3.71)	(2.58)	(4.18)	(3.00)
Style Flow		0.7990***	0.7406***	0.7417***	0.7344***
		(9.98)	(7.34)	(7.23)	(7.13)
Pr Fund Size			-0.0290***		-0.0300***
			(-4.45)		(-4.53)
Pr Cash Pct			-0.2855		-0.2152
			(-0.96)		(-0.68)
Observations	76600	73252	73487	73252	73252
R Squared	0.03	0.08	0.08	0.08	0.09

Continued on next page...

	(1) Flow	(2) Flow	(3) Flow	(4) Flow	(5) Flow
Fund clusters	4,526	4,423	4,442	4,423	4,423
Time clusters	44	44	44	44	44
Est Method	OLS	OLS	GMM	GMM	GMM
Hansen J stat			0.29	0.00	0.00
J p value			0.5880	0.9665	0.9930
KP LM Stat			30.63	30.89	30.66
KP LM p value			0.0000	0.0000	0.0000

Table IX: Spillover effects from measured coefficient estimates. Coefficients are transformed from Models 3,4, and 5, Table VIII by moving all contemporaneous flow variables to the LHS and dividing all estimated coefficients by one minus weighting matrix and spatial lag parameter. Coefficients taken directly from Table VIII for comparison, and variables are defined the same as in Table VIII. Direct Effect is the average of the diagonal of the matrix coefficient and measures the average direct effect. Spillover Effect is the average of the off-diagonal of the matrix coefficient and measures the average spillover effect the each fund has on its peers. Std Err of Coeff is the standard error of each coefficient estimate from Model 5. T Stat of Spillover is the T statistic computed by testing the Spillover Effect against the null of 0 using the coefficient standard error estimate. Significance is denoted at the 1, 5, and 10% level

(a) Model 5 (Main Result)

	Coeff Estimate	Direct Effect	Spillover Effect	Std Err of Coeff	Spillover T Stat
Lag1 Flow	0.0580	0.0580	0.0522***	0.0171	3.0543
Lag2 Flow	0.0769	0.0770	0.0692***	0.0145	4.7763
Lag3 Flow	0.0274	0.0274	0.0247**	0.0100	2.4735
Lag4 Flow	0.0040	0.0040	0.0036	0.0111	0.3242
Lag Alpha	0.8623	0.8629	0.7762***	0.1190	6.5210
Fund Size	0.0156	0.0156	-0.0429***	0.0027	-16.0791
Cash Pct	0.5821	0.5822	0.1152***	0.0334	3.4523
Style Flow	0.7344	0.7349	0.6611***	0.1030	6.4211
Amihud Illiq	0.0019	0.0019	0.0017***	0.0006	2.7018

(b) Model 4 (No Peer Controls)

	Coeff Estimate	Direct Effect	Spillover Effect	Std Err of Coeff	Spillover T Stat
Lag1 Flow	0.0580	0.0580	0.0281	0.0171	1.6487
Lag2 Flow	0.0765	0.0769	0.0373**	0.0146	2.5543
Lag3 Flow	0.0270	0.0274	0.0133	0.0100	1.3303
Lag4 Flow	0.0038	0.0040	0.0019	0.0111	0.1743
Lag Alpha	0.8572	0.8625	0.4183***	0.1202	3.4803
Fund Size	0.0150	0.0156	0.0076***	0.0027	2.8440
Cash Pct	0.5796	0.5822	0.2823***	0.0334	8.4567
Style Flow	0.7417	0.7346	0.3562***	0.1026	3.4735
Amihud Illiq	0.0027	0.0019	0.0009	0.0006	1.4282

(c) Model 3 (No Lag Alpha)

	Coeff Estimate	Direct Effect	Spillover Effect	Std Err of Coeff	Spillover T Stat
Lag1 Flow	0.0641	0.0641	0.0878***	0.0171	5.1390
Lag2 Flow	0.0814	0.0815	0.1117***	0.0141	7.8995
Lag3 Flow	0.0303	0.0304	0.0416***	0.0098	4.2250
Lag4 Flow	0.0039	0.0039	0.0053	0.0111	0.4819
Fund Size	0.0148	0.0148	-0.0483***	0.0027	-17.9173
Cash Pct	0.5822	0.5823	0.1215***	0.0335	3.6262
Style Flow	0.7406	0.7415	1.0155***	0.1009	10.0601
Amihud Illiq	0.0016	0.0016	0.0022***	0.0006	3.5313

Table X: Flow regression with relative size.

Relative Fund Size is the weighted average of peer total net assets divided by total net assets, logged. All other variables are as defined in Table VIII

	(1) Flow	(2) Flow	(3) Flow
Peer Flow	0.5125***	0.3938***	0.4087***
	(3.78)	(2.98)	(2.96)
Lag1 Flow	0.0640***	0.0579***	0.0578***
	(3.75)	(3.39)	(3.39)
Lag2 Flow	0.0809***	0.0766***	0.0764***
	(5.74)	(5.28)	(5.28)
Lag3 Flow	0.0299***	0.0271***	0.0270***
	(3.04)	(2.70)	(2.70)
Lag4 Flow	0.0037	0.0037	0.0038
	(0.33)	(0.33)	(0.34)
Lag1 Alpha		0.8614***	0.8607***
		(7.19)	(7.20)
Relative Fund Size	-0.0150***	-0.0158***	-0.0158***
	(-5.58)	(-5.93)	(-5.94)
Cash Pct	0.5820***	0.5795***	0.5819***
	(17.40)	(17.33)	(17.46)
Amihud Illiq	0.0020***	0.0023***	0.0023***
	(3.25)	(3.61)	(3.64)
Style Flow	0.7437***	0.7378***	0.7375***
	(7.39)	(7.19)	(7.18)
Pr Cash Pct	-0.2687		-0.2022
	(-0.90)		(-0.64)
Observations	73487	73252	73252
R Squared	0.08	0.09	0.09
Fund clusters	4,442	4,423	4,423
Time clusters	44	44	44
Est Method	GMM	GMM	GMM
Hansen J stat	0.23	0.00	0.00
J p value	0.6295	0.9926	0.9592
KP LM Stat	30.84	30.86	30.91
KP LM p value	0.0000	0.0000	0.0000

Table XI: Economic significance - shock to random nodes.

Economic significance of coefficients in Table IXa. A random third of mutual funds are shocked by one negative standard deviation event to each variable. The effect listed is the change in Flow, the dependent variable in Model 5, Table VIII, which is fund flow divided by total net assets. Mean flow is 0.013, Standard deviation of flow is 0.13. Mean spillover is the average effect on mutual funds who received no shock. Mean direct effect is the average effect on mutual funds who were shocked. The third column is the ratio of the two. The fourth and fifth columns display the mean spillover effect divided by mean flow and standard deviation of flow, respectively. Flow is the fund flow divided by total net assets. Fund Size is the log of total net assets. Cash Pct is cash holdings divided by total net assets. Amihud Illiq is the portfolio weighted sum of equity holdings' Amihud measures computed over the previous quarter, logged. Style Flow is the average of all reported fund flows by Morningstar category.

	Mean Spillover	Mean Direct Effect	Spillover/ Direct	Spillover/ Mean Flow	Spillover/ Std Flow
Lag1 Flow	-0.0018	-0.0097	0.1870	-0.1333	-0.0134
Lag2 Flow L2	-0.0024	-0.0128	0.1870	-0.1769	-0.0178
Lag3 Flow L3	-0.0009	-0.0046	0.1870	-0.0630	-0.0063
Lag4 Flow L4	-0.0001	-0.0007	0.1870	-0.0092	-0.0009
Lag1 Alpha L1	-0.0016	-0.0085	0.1892	-0.1184	-0.0119
Fund Size	0.0089	-0.0015	-5.7844	0.6557	0.0660
Cash Pct	-0.0010	-0.0178	0.0534	-0.0703	-0.0071
Style Flow	-0.0055	-0.0281	0.1972	-0.4091	-0.0412
Amihud Illiq	-0.0005	-0.0025	0.2211	-0.0406	-0.0041

Table XII: Flow regression isolating Sector Funds.
Flow is the dependent variable and is the fund flow divided by total net assets. Model 1 is the baseline, taken from Model 5 of Table VIII. Model 2 is the same, but with sector funds omitted from the analysis. Model 3 includes only sector funds. Sector funds are mutual funds with an industry-specific category, such as Technology or Health Care. All other variables are as defined in Table VIII

	(1) Flow	(2) Flow	(3) Flow
Peer Flow	0.4740***	0.5857***	0.5910**
	(3.28)	(3.40)	(2.05)
Lag1 Alpha	0.8623***	1.4310***	0.0273
	(7.24)	(9.47)	(0.24)
Lag1 Flow	0.0580***	0.0949***	-0.1763***
	(3.39)	(5.30)	(-4.22)
Lag2 Flow	0.0769***	0.0761***	-0.0060
	(5.31)	(5.12)	(-0.20)
Lag3 Flow	0.0274***	0.0266**	-0.0106
	(2.75)	(2.56)	(-0.51)
Lag4 Flow	0.0040	0.0082	-0.0314
	(0.36)	(0.84)	(-1.10)
Fund Size	0.0156***	0.0139***	0.0443***
	(5.85)	(5.55)	(4.10)
Cash Pct	0.5821***	0.5838***	0.4701***
	(17.44)	(17.13)	(5.61)
Amihud Illiq	0.0019***	0.0019***	0.0014
	(3.00)	(3.09)	(0.55)
Style Flow	0.7344***	0.6637***	0.7598***
	(7.13)	(4.85)	(5.14)
Pr Fund Size	-0.0300***	-0.0339***	-0.0507**
	(-4.53)	(-4.76)	(-2.26)
Pr Cash Pct	-0.2152	-0.1674	-0.4577
	(-0.68)	(-0.48)	(-0.92)
Observations	73252	65716	7536
R Squared	0.09	0.08	0.17
Fund clusters	4,423	4,005	418
Time clusters	44	44	44
Hansen J stat	0.00	0.02	0.76
J p value	0.9930	0.8777	0.3836
KP LM Stat	30.66	27.35	23.61
KP LM p value	0.0000	0.0000	0.0000

Table XIII: Flow regression isolating the Financial Crisis and Tech Bubble. Flow is the dependent variable and is the fund flow divided by total net assets. Model 1 is the baseline, taken from Model 5 of Table VIII from 1998 to 2009. Model 2 includes quarters from 2001 through 2009, thus excluding the tech bubble of the late 1990's. Model 3 includes quarters from 1998 through the second quarter of 2007. Model 4 extends Model 3 through the second quarter of 2008. Model 5 includes only 2001 through the second quarter of 2007, excluding both the tech bubble and the financial crisis. All other variables are as defined in Table VIII

	(1) Flow	(2) Flow	(3) Flow	(4) Flow	(5) Flow
Peer Flow	0.4740***	0.4305**	0.5348***	0.4958***	0.4869**
	(3.28)	(2.50)	(3.28)	(3.27)	(2.32)
Lag1 Alpha	0.8623***	1.0210***	0.8517***	0.9001***	1.0778***
	(7.24)	(6.81)	(6.62)	(6.75)	(6.69)
Lag1 Flow	0.0580***	0.0349**	0.0220	0.0469**	-0.0191
	(3.39)	(1.98)	(1.03)	(2.39)	(-0.84)
Lag2 Flow	0.0769***	0.0623***	0.0648***	0.0777***	0.0384***
	(5.31)	(4.27)	(4.01)	(4.61)	(2.62)
Lag3 Flow	0.0274***	0.0277***	0.0107	0.0183*	0.0052
	(2.75)	(2.59)	(0.80)	(1.68)	(0.33)
Lag4 Flow	0.0040	-0.0046	-0.0152	-0.0055	-0.0326**
	(0.36)	(-0.40)	(-1.04)	(-0.41)	(-2.04)
Fund Size	0.0156***	0.0205***	0.0221***	0.0176***	0.0321***
	(5.85)	(5.54)	(4.72)	(5.03)	(4.32)
Cash Pct	0.5821***	0.6078***	0.5848***	0.5788***	0.6026***
	(17.44)	(16.08)	(16.06)	(17.29)	(13.50)
Amihud Illiq	0.0019***	0.0020***	0.0018**	0.0015**	0.0021**
	(3.00)	(2.98)	(2.36)	(2.21)	(2.37)
Style Flow	0.7344***	0.8021***	0.7138***	0.7239***	0.7904***
	(7.13)	(6.77)	(6.56)	(6.99)	(6.04)
Pr Fund Size	-0.0300***	-0.0204***	-0.0337***	-0.0333***	-0.0231**
	(-4.53)	(-2.67)	(-4.31)	(-4.74)	(-2.46)
Pr Cash Pct	-0.2152	0.2992	-0.4138	-0.0159	0.2668
	(-0.68)	(0.80)	(-1.23)	(-0.05)	(0.67)
Observations	73252	64438	51012	60816	42202
R Squared	0.09	0.08	0.09	0.09	0.07
Fund clusters	4,423	4,263	3,831	4,161	3,663
Time clusters	44	32	35	39	23
Hansen J stat	0.00	3.22	0.81	0.10	1.44
J p value	0.9930	0.0727	0.3679	0.7577	0.2295
KP LM Stat	30.66	23.20	25.27	27.59	17.20
KP LM p value	0.0000	0.0000	0.0000	0.0000	0.0002