Trading Networks

Lada Adamic, Celso Brunetti, Jeffrey H. Harris, and Andrei Kirilenko*

Abstract

In this paper we analyze the time series of 12,000+ networks of traders in the e-mini S&P 500 stock index futures contract and empirically link network variables with financial variables more commonly used to describe market conditions. We show that network variables lead trading volume, intertrade duration, effective spreads, trade imbalances and other market liquidity measures. Network variables reflect information, information asymmetry and market liquidity and significantly presage future market conditions prior to volume or liquidity measures. We also find two-way Granger-causality between network variables and both returns and volatility, highlighting strong feedback between market conditions and trading behavior.

JEL Classifications: D85 Network Formation and Analysis, G12 Trading Volume,

G17 Financial Forecasting and Simulation, L14 Networks

*Adamic is with the University of Michigan, Brunetti is with the Board of Governors at the Federal Reserve Bank, Harris is with American University, and Kirilenko is with Imperial College. We are grateful to Paul Tsyhura for invaluable assistance with the retrieval, organization, and processing of transaction-level data. We thank Kirsten Anderson, Ana Babus, Matt Elliott, Pat Fishe, Ben Golub, Matt Harding, Matt Jackson, Pete Kyle, Shawn Mankad, Antonio Mele, Pam Moulton, Anna Obizhaeva, Han Ozsoylev, Kester Tong, Kumar Venkataraman and presentation participants at the American University, Cambridge University, the Chicago Mercantile Exchange, the Commodities Futures Trading Commission, 2009 Complexity Conference at Northwestern University, 2009 Econometric Society Summer Meetings in Barcelona, the Federal Reserve Board of Governors, George Washington University, NASDAQ, 2009 NBER-NSF Time Series Conference, Oxford, the Securities and Exchange Commission, Southern Methodist University, Stanford University, Syracuse University, the University of Maryland, the University of Michigan, the University of Missouri—Columbia, the University of Tennessee, the University of South Florida, Villanova University, and 2009 Workshop on Information in Networks at New York University for very helpful comments and suggestions. Some earlier drafts of the paper were circulated under the title "On the Informational Properties of Trading Networks." The views expressed in this paper are our own and do not constitute an official position of the Commodity Futures Trading Commission, its Commissioners or staff.

Financial markets bring buyers and sellers together, aggregating information for price discovery and providing liquidity for uninformed orders. Indeed, numerous models show that trading outcomes—prices, volume, volatility, and liquidity—emerge from the complex mix of information flows and liquidity demands of individual buyers and sellers.[1] More recent work directly connect network metrics to information flows (Babus and Kondor (2016)), dealer structure (Li and Schürhoff (2014)), order shredding/inventory management (Kyle, Obizhaeva, and Wang (2016)), and information percolation/market connectivity (Duffie, Malamud, and Manso (2015)).[2] Each of these works serve to motivate an empirical examination of financial markets through the lens of network technology.

In this paper we use established network analysis tools to characterize the time series dimensions of information and liquidity flows in the e-Mini S&P stock index futures market. We find strong and significant contemporaneous correlations between network statistics and financial variables, with network statistics significantly leading (Granger-causing) intertrade duration, trading volume, and other liquidity metrics. The fact that network statistics consistently lead these more traditional market metrics suggests that network statistics serve as primitive measures of market information and liquidity, lending support to the various

[1] See, for instance, Copeland and Galai (1983), Glosten and Milgrom (1985), Kyle (1985), Admati and Pfleiderer (1988), Foster and Viswanathan (1990, 1993, 1994). The theoretical literature on limit order markets includes Parlour (1998), Foucault (1999), Biais, Martimort and Rochet (2000), Parlour and Seppi (2003), Foucault, Kadan and Kandel (2005), Back and Baruch (2007), Goettler, Parlour, and Rajan (2005, 2009), Large (2009), Rosu (2009), and Biais and Weill (2009), among others.
[2] Other related works by DeMarzo, Vayanos, and Zwiebel (2003), Gale and Kariv (2003), Acemoglu, Dahleh, Lobel, and Ozdaglar (2008), and Golub and Jackson (2010) explore learning in social networks (rather than the financial networks we study here).

theoretical models of information flows, trading strategies and information percolation in financial markets. Moreover, network statistics change prior to volume and duration changes, metrics which have been heretofore ascribed to capture information in financial markets.[3]

We further explore whether measurement noise helps to explain the fact that network variables significantly lead more traditional market metrics. We find that, while the distributions of network statistics are largely unrelated to the noise-to-signal ratio in the market, duration, effective spread and Herfindahl trade measures are all (at least marginally) affected by market noise.

Importantly, the e-Mini S&P stock index futures market matches buyers and sellers electronically and anonymously, with time-price priority. In this regard, our findings are not driven by personal, business, or other social ties more typically explored with network theory, but rather represent a clean test of spontaneous order theory, where markets exhibit order brought about by mutual adjustments to market prices (see Ferguson (1767) and Hayek (1948)).[4] Moreover, this spontaneous order is consistent with models by Babus and Kondor (2016), Kyle et al. (2016), and Duffie et al. (2015) where information manifest in trading networks leads to changes in market information and liquidity.

Our findings also support models of information percolation in networks. For example, Golub and Jackson (2010) show that beliefs of network agents converge to

---

[3] See Epps and Epps (1976) and Engle and Russell (1998).

[4] Ferguson (1787) notes that people "stumble upon establishments, which are indeed the result of human action, but not the execution of any human design." To describe this order Hayek (1948) coins the term *catallaxy*, meaning 'to admit in the community' or 'to make friends' (notions familiar to social networks).

the truth if and only if the influence of the most central agents diminishes as the network grows. We show that a small group of market-makers extract mark-ups and deviations from the fundamental price are higher when the influence of the market-makers is also higher. Likewise, Duffie et al. (2015) show that information percolates more effectively in markets with more transactions (higher volume) or in markets where traders are highly connected if duration is sufficiently large.

While network analysis is data intensive, the fact that network statistics lead volume, duration and liquidity metrics suggests that there is value in monitoring network statistics from a regulatory and policy perspective. In Cohen-Cole, Patacchini and Zenou (2015) the propagation of incentives or strategic trading behavior in the interbank network generates systemic risk. Billio, Getmansky, Lo and Pelizzon (2012) and Brunetti, Harris, Mankad and Michailidis (2016) also use network analysis to identify and quantify financial crisis periods.

We also document that network statistics capture complex feedback mechanisms in the time series with network variables exhibiting bi-directional Granger-causality with returns and volatility, suggesting that trading strategies evolve dynamically in response to changing market conditions. To highlight this dynamic, we simulate an agent-based trading model which replicates the contemporaneous correlations between financial and network variables, but exhibits no Granger-causality. Overall, we reject the null hypothesis of random trading patterns in transactions among market participants and support trading models motivated by information flows (Babus and Kondor (2016)), information percolation

(Duffie et al. (2015)), and trading strategies (Kyle et al. (2016)), which predict persistent trading patterns.

Our results are robust to different equity index futures markets (e-mini Dow Jones and Nasdaq 100 futures), different observation periods (May and August 2008), and different network sampling frequencies (240 and 600 transactions). In the same E-mini S&P500 futures market that we study, Cohen-Cole, Kirilenko and Patacchini (2015) demonstrate how much a market shock is amplified by the network of traders and how widely it is transmitted across the network. They demonstrate that network pattern of trades captures the relations between behavior in the market and returns, showing that network spillovers explain as much as 90% of the individual variation in returns.

Our study using the time-series properties of network statistics complements other recent work exploring networks in different settings. In addition to Billio et al. (2012) and Brunetti et al. (2016), Leitner (2005) and Babus (2009) model bank networks to explore contagion issues. Similarly, Allen and Gale (2000) and Upper (2006) highlight how common asset holdings can drive interconnectedness within bank networks. Braverman and Minca (2014) describe how common holdings can transmit financial distress in bank networks with the severity of contagion depending on both the level and liquidity of common holdings. Similarly, Lagunoff and Schreft (1998) develop a model which shows that a high level of interconnectedness may increase financial fragility. Cabrales and Gottardi (2014) note a trade-off between risk-sharing and contagion within networks, while Acemoglu, Ozdaglar and Tahbaz-Salehi (2015) find that financial contagion is a function of the network structure.[5]

The paper proceeds as follows. In Section 1, we describe our unique high frequency data and our network and financial variables. In Section 2, we show examples of different trading networks and use these examples to illustrate how our network statistics relate to financial variables. In Section 3, we present contemporaneous correlations and lead-lag (Granger-causality) tests among network and financial variables. Section 4 explores the information content of trading networks. We conclude with Section 5.

1. High Frequency Data, Network Metrics, and Financial Variables

To build our trading networks, we use audit trail, transaction-level data for over 7.2 million regular transactions (executed between 9:30 a.m. ET and 4:00 p.m. ET) in the September 2009 e-mini S&P 500 futures contract during August 2009.[6] The e-mini S&P 500 futures contract is a highly liquid, fully electronic and cash-settled. The data contain the date, time (up to the second), unique transaction identifier (which enables us to order transactions sequentially within each second),

[5] See also De Vries (2005) and Acharya and Yorulmazer (2008). Roukny, Battiston, and Stiglitz (2016) show credit market networks can affect regulators' capacity to assess systemic risk. Bank networks are also connected to systemic risk (see e,g, Elsinger, Lehar and Summer (2006), Cifuentes, Ferrucci and Shin (2005), Allen and Babus (2010) and Allen, Babus and Carletti (2012), Caccioli, Farmer, Foti and Rockmore (2013) and Roukny, Bersini, Pirotte, Caldarelli and Battiston (2013).

[6] We also replicate our results with trades in the e-mini Nasdaq 100 and e-mini Dow Jones stock index futures contracts and in e-mini S&P 500 futures data from May 2008, August 2008 and August 2009. For brevity, we report only August 2009 e-mini S&P 500 results.

executing and opposite trading accounts, and opposite brokers, buy/sell flag, price, and quantity;[7] 31,585 unique accounts trade during our August 2009 sample.

Since our goal is to explore links between financial variables and network statistics, we first determine a sampling frequency that ensures our financial variables are not contaminated by market microstructure noise. We apply both the Andersen, Bollerslev, Diebold and Labys (2000) volatility signature plot and Bandi and Russell (2006) technique and find that the smallest acceptable sampling window must contain at least 50 transactions. Cognizant that networks constructed over such a small number of transactions can be too sparsely connected to adequately capture the dynamic nature of the informational and liquidity forces that drive markets, we use 600 transactions to construct meaningful trading networks.

1.1 Networks and Network Metrics

Over the entire month of trading, we construct 12,032 sequential (non-overlapping) trading networks, each comprised of 600 consecutive trades.[8] Following network terminology, a network consists of nodes and edges. We define a node as an individual trader (trading account) and an edge that connects a pair of nodes as a trade between two traders. In our work the edges are directed, with buys representing edges pointed toward a trader and sells representing edges pointed away from a trader. Multiple transactions between traders are represented by a single directed edge pointed toward the buyer and away from the seller, even if the prices and

[7] We applying standard filters designed to look for recording errors and outliers in the price and quantity series (see Hansen and Lunde (2006)) and find no irregularities in the data.

[8] Since the number of trades during the trading hours is seldom precisely divisible by 600, we complete the final network for each day with the trades reported after the close. Our results are robust to excluding this final network of each day.

quantities of each transaction differ. Additionally, from network terminology the degree of a node is the number of edges connected to it, with indegree (outdegree) representing the number of edges pointing toward (away from) the node. For clarity, we use the terms buydegree and selldegree to represent indegree and outdegree, respectively.

Quantitative analysis of networks employs a set of standard metrics. We focus on network metrics that represent the properties of information and liquidity in markets. Centrality, for instance, quantifies the importance of a specific trader in a network. Although several centrality measures exist, we use degree centrality to characterize how "central" a particular trader is in terms of its trading, utilizing the buy/sell indicators in our data to construct buydegree and selldegree centrality—representing the number of bilateral purchases and sales, respectively, for each trader in the network.[9]

We then compose aggregate centralization from individual trader centrality measures that characterizes the inequality in degree among all traders in the network. Specifically, we compute a centralization Gini ($G_{B,S}$) for buyers (B) and sellers (S) defined as:

$$G_{B,S} = \frac{\sum_1^N (2r_i - N - 1)k_i}{N * E}$$

[9] Note that buydegree and selldegree do not fully capture the role of a trader in the network. For instance, a trader that does not execute a large number of trades but serves to connect otherwise disconnected traders may be very central, but will have relatively low buydegree and selldegree measures.

summing over all N traders, where k_i is the i_{th} trader's buydegree and selldegree, r_i, $i=1,\ldots,N$, is each trader's rank order number[10] and E is the number of unique connections between traders present among the 600 trades in each network. By construction, G_B and G_S are 0 if every trader has the same number of buy and sell connections, and positive with increasing degree inequality. Maximum centralization occurs when one trader does all the buying or selling in the network.

Using the above formula, we compute network-wide centralization as the difference between G_B and G_S. Intuitively, centralization can be interpreted as the presence of a dominant buyer (close to 1) or a dominant seller (close to -1) within the 600-trade network. The absolute value of centralization (|centralization|) therefore represents the presence of a dominant trader on either side of the market. In economic terms, centralization is a measure of informed trading or strong net demand for liquidity in the market.

Table I presents summary statistics for the network variables across all 12,032 networks in our sample.[11] Centralization is approximately symmetric and ranges from -0.74 to +0.65 with a mean near zero and standard deviation of 0.17. |Centralization| ranges from 0.0 to 0.74 with mean 0.14 and standard deviation of 0.10. The mass of this distribution is more than one standard deviation away from zero, which can be interpreted as inequality in the number of buy *versus* sell matches per trader.

[10] The trader with the highest centrality ranks highest, the trader with the second highest centrality ranks second, and so on. We rank buydegree and selldegree centrality separately.

[11] Each network variable is stationary and exhibit persistent autocorrelations. Jarque-Bera (1980) tests reject the null of normality at standard significance levels.

Assortativity in networks represents the tendency of "like to be connected with like" for any node (Newman (2002)). For our purposes, we use each trader's buydegree and selldegree to assess assortativity within our networks, examining the propensity of highly connected buyers to trade with highly connected sellers, for instance. In an assortative network, buyers/sellers with many edges (high buydegree/selldegree nodes) are more likely to connect to similar buyers/sellers. In a disassortative network, buyers/sellers with many edges tend to connect to buyers/sellers with few edges. In real markets, block trades between two large traders represent disassortative networks, while large orders "walking the book" against many small counterparties represents assortative networks. As in Golosov, Lorenzoni and Tsyvinski (2009), small uninformed traders seeking to elicit information about asset values by making small offers may encounter large, informed traders (generating assortativity) or other small traders (generating a disassortative market).

We measure assortativity by the Pearson correlations between buydegree and selldegree for all edges present in the network. Using the notation above, over all edges $E_{i.j}$ we calculate four pairwise correlations $\rho(k_i^{buy}, k_j^{buy})$, $\rho(k_i^{buy}, k_j^{sell})$, $\rho(k_i^{sell}, k_j^{buy})$, and $\rho(k_i^{sell}, k_j^{sell})$ corresponding to the four conditional degree distributions (buydegree $k_{i,j}^{buy}$ and selldegree $k_{i,,j}^{sell}$) for each connected trader pair *i* and *j*. Intuitively, the coefficient $\rho(k_i^{sell}, k_j^{buy})$ measures the correlation between the number of unique buyers to whom a seller is selling to and the number of unique sellers that those buyers are buying from. A negative correlation indicates that when a seller matches with many buyers, those buyers are buying from few or no other sellers.

From these four correlations, we construct an aggregate assortativity index (AI) for each network

$$AI = \frac{1}{4}\left(\rho\left(k_i^{buy}, k_j^{buy}\right) - \rho\left(k_i^{buy}, k_j^{sell}\right) - \rho\left(k_i^{sell}, k_j^{buy}\right) + \rho\left(k_i^{sell}, k_j^{sell}\right)\right)$$

computed over all edges, where the scaling factor, ¼, assures that the assortativity index falls between -1 and 1. By construction, the assortativity index captures patterns in networks that feature large degree dispersion. For example, the assortativity index is high in a network that contains dominant traders or intermediaries—in either case, trades occur between parties that differ in both connectivity and liquidity provision/removal.

Intuitively, assortativity represents a measure of asymmetric information or liquidity imbalance in a trading network.[12] For instance, one large buyer matched with a number of small sellers exhibits high assortativity while a large buyer matched with a single large seller exhibits low assortativity. Indeed, if large orders are "shred" into smaller trades (perhaps stemming from inventory concerns as in Kyle et al. (2016)), assortativity will be high. Likewise, the presence of informed traders who have a greater number of counterparties will increase assortativity as in Babus and Kondor (2016).

As Table I shows, the assortativity index in our sample ranges from -0.06 to +0.34, with a mean of 0.04 and standard deviation of 0.04. This positive mean stems in part from the skewed degree distributions. Most buyers have low buydegree and,

[12] Glosten (1987), Stoll (1989), George, Kaul and Nimalendran (1991) and Madhavan, Richardson and Roomans (1997) each model asymmetric information components in quoted bid-ask spreads.

on average, buy from a seller with high selldegree. Similarly, most sellers have low selldegree and, on average, sell to buyers with high buydegree. This pattern is consistent with a market populated by a small number of highly connected intermediaries who trade with many liquidity-demanding (or informed) traders.

Clustering is a measure of transitivity in the network, i.e. if i trades with j, and j trades with k, clustering measures whether i also trades directly with k. We quantify clustering using the global clustering coefficient (CC) (Newman (2002)):

$$CC = \frac{3 * \ T_{closed}}{T}$$

where T represents the total number of "connected triples" of three traders (i, j and k) and T_{closed} represents the number of "closed triples" where i trades with j, j trades with k, and i also trades directly with k.[13]

Economically, the clustering coefficient represents liquidity in the market. Connected triples represent the presence of at least one short-term liquidity provider. Intuitively, large clustering coefficients represent greater liquidity. In the extreme case where a single trader provides liquidity, no closed triples exist and the clustering coefficient is zero. At the other extreme, where all triples involve three traders connected as a "closed triple", the clustering coefficient is 3. Higher clustering levels (connectivity) are also linked to higher levels of information in Duffie et al. (2015).

As shown in Table I, the clustering coefficient in our sample ranges from 0.0 to 0.23, with an average of 0.05 and standard deviation of 0.03. On average, there is

[13] We treat the edges as undirected in the clustering coefficient since intermediation/liquidity involves both buying and selling. Fagiolo (2007) discusses directed clustering coefficients.

only a slight tendency for traders to cluster in this market. Indeed, the average clustering coefficient we observe is nearly identical to the coefficient we would expect by randomly connecting traders.

Dispersion of information or liquidity in the market may be measured using connected components. A connected component is the set of all traders connected with each other through bilateral trading. The largest strongly connected component (LSCC) is the maximum number of traders that can be reached from any other trader by following directed edges.

We compute the largest strongly connected component as:

$$LSCC = \frac{LSCC_{Max}}{N}$$

where $LSCC_{Max}$ is the raw count of traders in the LSCC and N is the total number of traders in the market. This ratio ranges between 1/N (one trader connects all other traders) and 1 (all traders are connected to all other traders).

A larger LSCC forms when many traders are both buying and selling, that is when the supply of liquidity is dispersed among many traders. For example, a large strongly connected component is likely to emerge as a result of a large number of limit orders rather than one large market order. Since the LSCC is scaled by the total number of traders in the market, the arrival of concentrated net demand for liquidity will manifest itself in a smaller LSCC. Similar to clustering, a larger LSCC indicates greater connectivity, which is linked to higher levels of information in Duffie et al. (2015). Golosov et al. (2009) focus on the time-dimension of information diffusion either between differentially informed agents, or from homogeneously informed to uninformed agents. As Table I indicates, the portion of the network occupied by the largest strongly connected component varies significantly, ranging from 0.00 to 0.52, with a mean of 0.09 and standard deviation of 0.06. While both clustering and LSCC represent liquidity/market information, the mean and variability in LSCC are significantly larger, suggesting that the two metrics are not equivalent.

1.2 Financial Variables

For each of the 12,032 network sampling periods, we also compute market returns, volatility, intertrade duration, trading volume, and more standard liquidity measures including effective bid-ask spreads, and signed volume, buy/sell trade imbalances. These variables typically describe financial market conditions. Lastly, given that clustering and centralization may be related to market concentration, we also construct a Herfindahl measure of buy/sell trade concentration during each interval. Descriptive statistics for these financial variables are shown in Table II.[14]

Market returns contain valuable information about the true underlying price formation process, but may also contain market microstructure noise, measurement errors, and seasonal patterns (Engle (2000)). We compute open-to-close market returns as the log difference between the last and the first transaction price for each network period. We remove the predictable intraday seasonal component from raw returns by regressing returns on a constant and a sequence of dummy variables for each half-hour during the trading day and use the unexplained term as our measure

[14] As with our network variables, all financial variables in our sample are stationary and (other than returns) highly persistent, with significant autocorrelation coefficients at 1, 5, and 10 lags.

of market returns.[15] As shown in Table II, returns range from -0.15 to +0.20, are centered on zero with standard deviation of 0.04 and are slightly negatively skewed.

We use three measures to estimate volatility during each sampling interval period: absolute returns, squared returns, and the log difference between the maximum and minimum prices (price range). For the results reported below, we use price range as an estimate of volatility.[16] Importantly, however, our main results are not affected by the choice of volatility estimator. As shown in Table II, across 12,032 sampling intervals, volatility averages about 0.06 percent, corresponding to an annual volatility of 23.56 percent. Volatility ranges from 0.02 to 0.29 percent, with a standard deviation of just 0.02 percent.

Trading volume contains valuable information about the underlying price formation process, because volume together with observed transaction prices may be driven by a common latent factor.[17] We compute trading volume as the number of contracts both bought and sold during each network interval. Volume ranges from 1095 to 7706 with an average of 2629 and standard deviation of 636 contracts.

Intertrade duration can be interpreted as a proxy for the arrival of new information or liquidity to the market (Engle and Russell (1998) and Engle (2000)). While the concept of duration in Duffie et al. (2015) differs somewhat from intertrade duration, they show that markets with more transactions (or with greater

---

[15] We apply the same technique and use a Fourier flexible form to remove seasonality from all variables and examine alternative close-to-close returns with similar results.

[16] Range-based volatility estimators are more efficient than return-based volatility estimators (see, e.g., Parkinson (1980), Garman and Klass (1980), Beckers (1983), and Brunetti and Lildtholdt (2006)). Christensen and Podolskij (2007) use the price range to compute realized volatility in high frequency data.

[17] The vast literature on the properties of trading volume includes Clark (1973), Epps and Epps (1976), Tauchen and Pitts (1983), Admati and Pfleiderer (1988), Easley and O'Hara (1992), and Andersen (1996), among others.

connectivity) have higher information if duration is sufficiently low. We compute three measures of duration (total period duration, volume-weighted period duration, and the average of the 599 durations) and report results for total period duration—the time elapsed during the network interval. Our main results are not affected by the choice of duration estimator. As shown in Table II, duration ranges from zero to 344 seconds with 600 transactions occurring every 40.8 seconds, on average.

Liquidity reflects the ease with which a security can be bought or sold without a significant price change. Unlike trading volume and duration, liquidity is not directly observable and has multiple dimensions. Indeed, our network variables capture some of these dimensions. We also compute more traditional liquidity metrics, like effective spreads, signed volume and trade imbalances.[18] The effective spread captures the implied cost of trading and is equal to twice the square root of the first order autocovariance of returns over each interval (Stoll (1978)). Signed volume (buy minus sell volume) and trade imbalances (number of buys minus sells) capture the propensity for buy and sell orders to match up during the interval.

The e-mini S&P 500 contract is extremely liquid. Table II shows that the average effective spread is less than 0.8 cents, with a maximum effective spread of just over two cents during our sample period. The raw Amihud illiquidity measure (unreported) also has mean, median, and standard deviation very close to zero and indicates it takes about 21 contracts (or over $1 million) to move prices by one tick (0.25 index points). While mean and median signed volume and trade imbalances are

[18] We also compute Amihud illiquidity measures using both contract volume and dollar trading volume with nearly identical results.

very small, the market can be imbalanced, with maximum signed volume exceeding 5000 contracts and maximum buy imbalance at 13 trades. These extreme cases suggest that market liquidity can be stressed during the short intervals we study.

Table II also displays a Herfindahl measure computed over each trading interval to compare with the clustering and centralization metrics we apply from network statistics. As shown, the Herfindahl trades statistic is less variable than most other variables, with a minimum of 0.0002, maximum of 0.0013 and standard deviation of just 0.0001.[19]

2. Illustrative Examples

Prior to conducting the time-series statistical analysis for the network and financial variables, we present statistics from representative trading networks to illustrate the key concepts and the intuition behind our approach. Figure 1 presents three actual trading networks from our data, accompanied by network statistics and financial variables. The three trading networks are chosen to display relatively high values of centralization (left column), assortativity (middle column), and both clustering and the largest strongly connected component (right column).

The left column of Figure 1 presents a network with one dominant seller (perhaps an informed trader or trader with strong liquidity demand) matched with many small buyers. As shown in Panel A, this trading network has a large negative centralization coefficient, intermediate levels of assortativity and LSCC, and a small

[19] In unreported results we also construct a Herfindahl volume metric with nearly identical results.

clustering coefficient. As shown in Panel B, this network is also associated with a large negative market return, high volatility, intermediate trading volume, and low duration. Liquidity, volatility, duration and return statistics suggest that strong liquidity demand drains a significant portion of market liquidity and made a considerable price impact. Indeed, the signed volume and trade imbalance are accompanied by a large negative return, volatility, and short intertrade duration suggesting a large market order which quickly "walks the book."

The center column of Figure 1 presents a trading network with a relatively large buyer trading with several moderately-sized sellers through a number of intermediaries. As shown in Panel A, this network is characterized by intermediate centralization and clustering levels, suggesting that the net demand for liquidity is relatively balanced, with slightly negative signed volume but positive trade imbalance. Moreover, high assortativity means that large traders are mostly matched with many small traders (rather than with each other); and the small LSCC suggests that just a few intermediaries provide liquidity. As shown in Panel B, this network is associated with good liquidity and relatively low effective spreads. Furthermore, this network is characterized by high duration and relatively low volume—these 600 trades are relatively small. This network is associated with intermediate positive market returns and volatility. Overall, these market conditions represent a handful of intermediaries supplying liquidity to smaller traders.

The right column of Figure 1 presents a trading network with greater dispersion among buyers and sellers of various sizes, reflected in zero centralization (no dominant buyer or seller) and low assortativity (like traders trade with like). High levels of clustering and LSCC, suggest that the net supply of liquidity is dispersed among many traders. Indeed, this trading network exhibits zero returns accompanied by good liquidity—low signed volume imbalance, zero trade imbalance and relatively large trading volume.

These three examples illustrate our conjecture that network statistics and standard financial variables capture different dimensions of market conditions. Indeed, since information and liquidity appear to drive both network and financial variables, the two sets of variables are likely to be statistically interrelated as well. We formally examine the statistical relation between network metrics and traditional financial variables below.

3. The Statistical Relation between Network and Financial Variables

We first examine contemporaneous correlations among network and financial variables and then test for lead-lag relations between and among these variables. Panel A in Table III reports contemporaneous correlations among network variables. Although centralization is not correlated with other network variables, the other network variables are often strongly correlated with each other. |Centralization| is negatively correlated with both the clustering coefficient and the LSCC. That is, when a dominant trader increases net demand for market liquidity, counterparties to that dominant trader are less likely to trade with each other. Conversely, the high positive correlation between clustering and LSCC suggests that dispersed net liquidity supply among many traders is often accompanied by strong liquidity demand. Assortativity is relatively uncorrelated with centralization and clustering, but is strongly negatively correlated with the LSCC—both higher assortativity and lower LSCC reflect greater liquidity imbalances.

Panel B in Table III examines this economic intuition more closely with contemporaneous correlations between financial and network variables. Centralization is strongly correlated with returns, signed volume and trade imbalances. While |centralization| is positively correlated with volume and volatility, it is negatively correlated with duration, effective spreads and trade concentration. Intuitively, a large order (with high |centralization|) that walks the limit order book results in higher volatility and volume with lower duration.

Higher asymmetric information or greater liquidity imbalance, as represented by assortativity, is negatively correlated with volume and volatility, but largely unrelated to more traditional market metrics. The former suggests that higher asymmetric information or liquidity imbalance is accompanied by a reduction in both volatility and trading volume. The latter suggests that asymmetric information, as represented by assortativity, is not related to signed volume or trade imbalances.

Clustering is positively correlated with volume and negatively correlated with effective spreads and volatility. Intuitively, since clustering captures short-term market making activities, greater intermediation improves liquidity, smooths volatility and enhances trading volume. Clustering is also highly negatively correlated with trade concentration (Herfindahl trades). The LSCC, the dispersion of net liquidity supply, is positively correlated with both volume and volatility. The LSCC, however, is not significantly correlated with other more traditional financial market variables.

While the contemporaneous correlations suggest a link between network and financial variables, we now examine whether this link reflects a redundancy in the data or whether network variables add incremental value to more traditional market quality metrics. We use the time series of data to examine the lead-lag relations among network and financial variables and conjecture that network metrics serve as primitive measures of information and liquidity.

We apply standard Granger causality tests in vector autoregressive (VAR) models among network and financial variables.[20] Table IV provides the p-values of Granger-non-causality tests among the five network variables. Centralization neither Granger-causes nor is Granger-caused by other network variables (p-values equal to 0.65 and 0.37, respectively). As a system, however, the remaining network variables are both jointly Granger-caused by (and jointly Granger-cause) each other (p-values between 0.00 and 0.05), indicating strong feedback effects within our network statistics. Indeed, most pair-wise tests are also significant.

Table V presents p-values for the Granger-non-causality tests between network variables and each financial variable both jointly and independently. In Panel A, we find that returns and volatility are jointly both Granger-caused by and

[20] Since the variables exhibit heteroskedasticity and serial correlation, we estimate VAR models using generalized method of moments (GMM) and Newey-West robust standard errors. Standard tests show that the VAR model must include all five network variables. The Akaike and Schwartz Information Criteria indicate an optimal lag-length of twelve.

Granger-cause network variables (p-values of 0.00 and 0.01 for returns and 0.00 and 0.00 for volatility, respectively). Centralization drives the result for returns, with strong bi-directional Granger-causality. For volatility, we find strong bi-directional Granger-causality (feedback effects) between volatility and each network variable.

Panel B in Table V reports test results for intertrade duration and volume. We find strong evidence of one-way causality—duration and volume are both Granger-caused by each network variable, both individually and jointly (joint p-value = 0.00). While duration leads the LSCC and volume leads assortativity, neither duration nor volume lead network variables jointly (p-values = 0.20 and 0.16, respectively). These results indicate that network statistics significantly lead both duration and trading volume. To the extent that duration and volume proxy for information arrival, network statistics reflect forthcoming changes in market conditions.

Lastly, Panels C and D show that network variables jointly Granger-cause signed volume, trade concentration, effective spreads and trade imbalances. These strong results suggest that the demand and supply of liquidity are reflected in network variables *prior to* emerging in more traditional liquidity measures. While individual pair-wise tests vary significantly, we find no feedback effects here. Taken as a whole, these network statistics lead short-term changes in both liquidity and market concentration as well.

For comparison, Table VI displays bi-variate VAR Granger-causality tests between traditional liquidity metrics and the traditional market statistics of returns, volatility, volume and duration. Contrary to the tests with network statistics, we find bi-directional causality among many of these variables—they are jointly determined contemporaneously. Effective spreads, in particular, exhibit bi-directional relations with volatility, volume and duration. Trade imbalances (Panel B) are unrelated to these other market statistics. Interestingly, we find no leads or lags between these traditional liquidity measures and returns, in stark contrast to the strong bi-directional effects between network statistics and returns. These results confirm that network metrics capture market conditions that are not reflected in traditional market descriptors. In fact, network metrics significantly presage future market conditions at very short horizons.

4. Information, Noise and Network Formation

4.1 Information

For greater perspective on real world trading networks, we examine the time series properties in both re-wired networks and simulated networks. We first re-wire our networks using each active trader while preserving the degree distribution: *e.g.* if trader i in a given network sells (buys) three times, in the re-wired network node i still sells (buys) three times but to randomly assigned counterparties in the same network. We re-compute all the network variables in the re-wired networks to compare with the actual trading networks. If the trading networks are randomly formed, there will be no difference between the trading and re-wired networks.

In contrast to the high correlations found between trading network variables and financial variables, we find almost no correlations with statistics generated from re-wired networks.[21] Moreover, the results from the Granger-non-causality tests do not hold for the re-wired network. These findings support the notion that the trading networks do not form randomly, but rather reflect information flows and/or information percolation.

Secondly, we explore the source of the bi-directional Granger-causality that we find between network and financial variables by constructing an agent-based simulation model of a limit order market that is devoid of a feedback mechanism. The simulated networks allow for heterogeneous beliefs about the price process, but impart no intentionality or memory upon the traders. Consequently, we might expect to find significant correlations among network and financial variables, but not necessarily evidence of Granger-causality since feedback effects are not included.[22]

Generally, we find that the simulated executions generate contemporaneous correlations similar to, but less significant than, those from real market data. The correlations between network variables and between network and financial variables are displayed in Table VII. As with the market data, we find strong correlations between returns and centralization. The simulation also sheds light on the possible sources of this high correlation. We find that the mechanics of the impatient order submission strategy raises the correlation: when a large buy order at a high price is matched against several existing sell orders network centralization is high. Matching a large number of sell orders to a single buy order likely reflects the large buy order walking up the book, so contemporaneous returns are also high.

[21] To conserve space, we do not report the results.
[22] Details on the simulation are included in the Appendix

Generally, correlations between volatility and |centralization|, and between the Hefindhal index and |centralization| are also relatively high in the simulated networks. These results suggest that the order matching process can generate contemporaneous correlations between financial and network variables in the absence of any economic motivation (other than the prescribed urgency to trade built in to simulate perishable information).

Granger-causality tests among simulated variables lack the dynamic structure found in live market data, as might be expected.[23] Indeed, Granger-causality tests among network and financial variables are generally insignificant and yield few feedback effects, indicating a very poor fit. This suggests that the Granger-causality results that we find in the actual market data arise as a result of the strategic behavior of traders (as in Cohen-Cole et al. (2015)) or information flows or percolation (as in Babus and Kondor (2016) and Duffie et al. (2015)) and are not simply artifacts of the order matching process.

4.2 Noise

We also explore whether measurement error (known to affect returns) affects our inferences about network statistics.[24] Both financial and network variables may be noisy estimates of other latent variables in the market. We conjecture that networks measure information and market interactions among agents with less error

[23] Reflecting the lack of dynamics in the simulated data, the AIC (SIC) selects an optimal lag-length of order one (zero) in the VAR specification. We use lag-length of order one.

[24] We thank an anonymous referee for moving the paper in this direction.

than traditional financial statistics. Borrowing from the high frequency econometrics literature[25] we may write the observed (log) trading price as

$$p_t = p_t^F + \varepsilon_t$$

where p_t^F is the fundamental (true) price and ε_t represents the market microstructure noise stemming from bid-ask bounce, discrete prices, etc. The effects of noise on observed prices depends on the properties of the noise itself. In particular, if we compute returns as

$$\Delta p_t = \Delta p_t^F + \Delta\varepsilon_t$$

and if ε_t is autocorrelated and also correlated to p_t^F then Δp_t is a moving average process and the effects of the error term can be substantial. While we select the optimal sampling frequency to minimize the effects of the error when computing daily realized volatilities, other financial variables might also be affected by noise. Conversely, our network variables utilize only the direction of trades and do not include information about prices and/or quantities, so we hypothesize that network variables are less affected by microstructure noise, ε_t.

To test this hypothesis, we first compute the ratio of high-low (range) volatility estimates to open-to-close returns as a simple measure of the noise-to-signal ratio in the market. For instance, if a network interval has a large bid-offer spread (large value of ε_t) and the fundamental price (p_t^F) does not change much, the range mainly captures noise. Additionally, information that moves the price captures the signal, represented in the denominator, and the ratio of the range to the open-to-close return

[25] See Andersen et al. (2000), Barndorff-Nielsen, Hansen, Lunde and Shephard (2009), and Zhang, Mykland and Ait-Sahalia (2005).

is a proxy for the noise-to-signal ratio (at least for the extreme cases—i.e. when large values of the range are associated with low returns). In fact, we concentrate on the right tail of the distribution of our proxy of the noise-to-signal ratio.

Figure 2 depicts the noise-to-signal ratios for the top decile of our network intervals. As shown, 60 intervals (representing 0.5% of the total) contain the largest noise-to-signal ratios, with the ratio stabilizing around the 400th observation (3% of our sample). In the analysis that follows we consider both the 97th percentile and the 95th percentile of the noise to signal ratio.

Figure 3 depicts the distribution of the network statistics, segmenting networks in the 95th percentile by noise-to-signal ratio from all others. The observations that correspond to top quintile of the noise-to-signal ratio are in red, while all other observations are in green.[26] Figure 3 shows that there is no material difference between the network variables corresponding to the top quintile of the noise-to-signal ratio and all other observations. Moreover, the tails of these distributions of network statistics appear to be more populated by observations not related to high noise-to-signal ratios.

In Table VIII we present formal tests of whether the means of the network variables corresponding to the top quintile of the noise-to-signal ratio and the mean of the rest of the observations are equal. These tests clearly fail to reject the null of different population means—when the signal-to-noise ratio is high, network variables are not materially different.

[26] Results are similar for the 97th percentile.

We run the same test for the financial variables and marginally reject (at 10% level) the null of equal means for volume, duration, effective spread and Herfindahl trades when comparing the top quintile of the noise-to-signal ratio to the rest of the sample.[27] These results provide some evidence in support of our hypothesis that network variables contain less noise than financial variables.

4.3 Network Formation

Given that patterns emerge in network statistics, we also formally test the mark-up hypothesis of Li and Schürhoff (2014)—that more central traders can estimate their own contemporaneous influence in the market and can then use this advantage to more accurately predict future market returns. To do so we first isolate the top seven traders (by trading volume) and compute the following over each network interval. Let Q_{ijt} be the quantity transacted between accounts i and j in period t. Then, $buys_t = \sum_{ij} Q_{ijt}\,,\ j \in Top7$, $sells_t = \sum_{ij} Q_{ijt}\,,\ i \in Top7$, and $k_t = buys_t - sells_t$ and k_t is a centrality measure for the top seven traders, analogous to centrality computed across all traders individually. We then run the following regression:

$$r_{i,t} = \alpha_0 + \sum_{j=0}^{3} \beta_j\, k_{i,t} + \gamma_0 + h_0 + m_0 + s_0 + u_{i,t}$$

where γ_0, h_0, m_0, s_0 are the day of the week, hour, minute, and second fixed effects. The mark-up hypothesis implies that central traders make markets so that $\beta_1 < 0$, and learn from this central position and can forecast future returns, so that $\beta_2 > 0$ and $\beta_3 > 0$.

[27] Note we do not run the test for volatility and returns since both measures are used in constructing our noise-to-signal ratios.

The first column in Table IX shows that contemporaneous returns are negatively related to more central traders (k_t) suggesting these traders serve an informal market making role. Moreover, returns also contain information from more central traders at up to two lags. The second column in Table IX shows that lagged returns are strongly positively related to more central traders for at least three lags. These results are consistent with the mark-up hypothesis of Li and Schürhoff (2014).

5. Conclusions

We use audit trail data that identifies trading accounts on both sides of every trade to construct the links between buyers and sellers in the e-mini S&P 500 stock index futures contract, the price discovery venue for the U.S. equity market. We construct a time series of more than 12,000 trading networks each comprised of 600 sequential trades and use established network analysis tools to empirically link network variables with financial variables that describe market conditions. We find that network statistics are contemporaneously correlated with these financial variables, suggesting that the underlying economics that drive financial variables also drive network parameters. Importantly, network variables capture information, information asymmetry and liquidity characteristics in the e-mini S&P500 futures market that we study consistent with models of information flows or percolation (e.g. Babus and Kondor (2016) and Duffie et al. (2015))

Moreover, network variables are primitive determinants of market conditions, emerging prior to more common measures of information and liquidity, due in part to the fact that network variables are less contaminated by market noise. For example, we find that network statistics Granger-cause trading volume, intertrade duration, effective spreads, trade imbalances and signed volume imbalances. We also document bi-directional Granger-causality between network variables and both returns and volatility, suggesting that network variables also capture complex feedback mechanisms that underlie trading strategies in financial markets.

We conjecture that market-wide patterns of order execution are informative about the price formation process in order driven markets and these dynamics are manifest in both financial and network variables. Given the fact that network variables presage volume, duration and other market liquidity measures, we conclude that trading network analysis offers a fruitful mechanism for assessing trading strategies that drive price formation and liquidity supply in financial markets (as in Cohen-Cole et al. (2015) and Kyle et al. (2016)).

Indeed, we form our trading networks based on common trader attributes—a buyer and a seller agree on a precise price and quantity at a given time and therefore trade. Like Billio et al. (2012), we show that networks formed not on the basis of personal relationships, but rather on economic considerations, reflect market conditions prior to more traditional volume and volatility metrics. Moreover, we find support for the mark-up hypothesis of Li and Schürhoff (2014)—that more central traders can use their information advantage to predict short horizon market returns.

Given the strong contemporaneous link between financial and network variables, we believe that network analysis tools also offer new insights into the behavior of financial markets. We adopt confidential data. However, financial networks can be formed by public data as shown in Billio et al. (2012), Diebold and Yilmaz (2014) and Brunetti et al. (2016).

While the data we employ here differs from the typical social network setting where connections are personal, we show that networks formed by economic connections also reveal information about market conditions, despite the lack of social connections between our traders. Whether our results hold in settings where orders are executed via preferencing arrangements, across competitive trade execution venues (like the U.S. equity markets), in floor-based exchanges, or in markets where internalization or payments/rebates for order flow are possible remain questions for further research.

References

Acemoglu, Daron, Munzer A. Dahleh, Ilan Lobel, and Asuman Ozdaglar, 2008, Bayesian learning in social networks, NBER working paper #14040.

Acemoglu, Daron, Asuman Ozdaglar, and Alireza Tahbaz-Salehi, 2015, Systematic risk and stability in financial networks, *American Economic Review* 105, 564-608.

Acharya, Viral V., and Tanju Yorulmazer, 2008, Information contagion and bank herding, *Journal of Money, Credit, and Banking* 40, 215-231.

Admati, Anat R., and Paul Pfleiderer, 1988, A theory of intraday patterns: Volume and price variability, *Review of Financial Studies* 1, 3-40.

Allen, Franklin, and Ana Babus, 2009, Networks in finance. The Network Challenge, Edited by Paul Kleindorfer and Jerry Wind, Wharton School Publishing, 367-382.

Allen, Franklin, Ana Babus, and Elena Carletti, 2012, Asset commonality, debt maturity and systemic risk, *Journal of Financial Economics* 104, 519-534.

Andersen, Torben G., 1996, Return volatility and trading volume: An information flow interpretation of stochastic volatility, *Journal of Finance* 51, 169-204.

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Paul Labys, 2000, Great realizations, *Risk* 13, 105-108.

Babus, Ana, 2009, The formation of financial networks, University of Cambridge working paper.

Babus, Ana, and Péter Kondor, 2016, Trading and information diffusion in over-the-counter markets, CERP discussion paper #9271.

Back, Kerry, and Shmuel Baruch, 2007, Working orders in limit-order markets and floor exchanges, *Journal of Finance* 61, 1589-1621.

Bandi, Federico M., and Jeffrey R. Russell, 2006, Separating microstructure noise from volatility, *Journal of Financial Economics* 79, 655-692.

Barndorff-Nielsen, Ole Eiler, P. Reinhard Hansen, Asger Lunde, and Neil Shephard, 2009, Realized kernels in practice: trades and quotes, *Econometrics Journal*, 12, C1-C32.

Beckers, Stan, 1983, Variance of security price returns based on high, low and closing prices, *Journal of Business* 56, 97-112.

Biais, Bruno, David Martimort, and Jean-Charles Rochet, 2000, Competing mechanisms in a common value environment, *Econometrica* 68, 799–837.

Biais, Bruno and Pierre-Olivier Weill, 2009, Liquidity shocks and order book dynamics, NBER working paper #15009.

Billio, Monica, Mila Getmansky, Andrew W. Lo and Loriana Pelizzon, 2012, Econometric Measures of Connectedness and Systemic Risk in the Finance and Insurance Sectors, *Journal of Financial Economics* 104, 536-559.

Braverman, A. and A. Minca, 2014, Networks of common asset holdings: aggregation and measures of vulnerability, Available at SSRN: http://ssrn.com/abstract=2379669 or http://dx.doi.org/10.2139/ssrn.2379669

Brunetti, Celso, Jeffrey H. Harris, Shawn Mankad and George Michailidis, 2016, Interconnectedness in the Interbank Market, Finance and Economics Discussion Series 2015-090. Washington: Board of Governors of the Federal Reserve System, http://dx.doi.org/10.17016/FEDS.2015.090.

Brunetti, Celso, and Peter M. Lildholdt, 2006, Relative efficiency of return- and range-based volatility estimators, Mimeo.

Cabrales, Antonio, and Piero Gottardi, 2014, Markets for information: Of inefficient firewalls and efficient monopolies*, Games and Economic Behavior* 83, 24-44.

Caccioli, Fabio, J. Doyne Farmer, Nick Foti, and Daniel Rockmore, 2013, How interbank lending amplifies overlapping portfolio contagion: A case study of the Austrian banking network. arXiv preprint, arXiv:1306.3704.

Christensen, Kim, and Mark Podolskij, 2007, Realised range-based estimation of integrated variance. *Journal of Econometrics* 141, 323-349.

Cifuentes, Rodrigo, Gianluigi Ferrucci and Hyun Song Shin, 2005, Liquidity Risk and Contagion, *Journal of European Economic Association* 3, 556-566.

Clark, Peter K., 1973, A subordinated stochastic process model with finite variance for speculative prices, *Econometrica* 41, 135-155.

Cohen-Cole, Ethan, Andrei A. Kirilenko and Eleonora Patacchini, 2014, Trading Networks and Liquidity Provision*, Journal of Financial Economics* 113, 235-251.

Cohen-Cole, Ethan, Eleonora Patacchini and Yves Zenou, 2015, Static and Dynamic Networks in Interbank Markets, *Network Science* 3, 98-123.

Copeland, Thomas E. and Dan Galai, 1983, Information effects on the bid-ask spread, *Journal of Finance* 38, 1457-1469.

Duffie, Durrell, Semyon Malamud and Gustavo Manso, 2015, Information Percolation in Segmented Markets, *Journal of Economic Theory*, 157, 1130-1158.

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

De Vries, Casper G., 2005, The simple economics of bank fragility, *Journal of Banking and Finance* 29, 803-825.

Diebold, Frank X., and Yilmaz Kamil, 2014, On the Network Topology of Variance Decompositions: Measuring the Connectedness of Financial Firms, *Journal of Econometrics*, 182, 119-134.

Easley, David and Maureen O'Hara, 1992, Adverse selection and large trade volume: the implications for market efficiency, *Journal of Financial and Quantitative Analysis* 27, 185-208.

Elsinger, Helmut, Alfred Lehar, and Martin Summer, 2006. Risk assessment for banking systems. *Management Science* 52, 1301-1314.

Engle, Robert F., 2000, The econometrics of ultra-high-frequency data, *Econometrica* 68, 1-22.

Engle, Robert F., and Jeffrey R. Russell, 1998, Autoregressive conditional duration: A new model for irregularly spaced transaction data, *Econometrica* 66, 1127-1162.

Epps, Thomas W., and Mary Lee Epps, 1976, The stochastic dependence of security price changes and transaction volumes: Implications for the mixture-of-distribution hypothesis, *Econometrica* 44, 305-321.

Fagiolo, Giorgio, 2007, Clustering in complex directed networks, *Physical Review E* 76, 26107.

Ferguson, Adam, 1767, An Essay on the History of Civil Society.

Foster, F. Douglas, and S. Viswanathan, 1990, A theory of interday variations in volumes, variances and trading costs in securities markets, *Review of Financial Studies* 4, 595-624.

Foster, F. Douglas, and S. Viswanathan, 1993, Variations in trading volume, return volatility and trading costs: Evidence on recent price formation models, *Journal of Finance* 48, 187-211.

Foster, F. Douglas, and S. Viswanathan, 1994, Strategic trading with asymmetrically informed traders and long-lived information, *Journal of Finance and Quantitative Analysis* 29, 499-518.

Foucault, Thierry, 1999, Order flow composition and trading costs in a dynamic limit order market, *Journal of Financial Markets* 2, 99–134.

Foucault, Thierry, Ohad Kadan, and Eugene Kandel, 2005, The limit order book as a market for liquidity, *Review of Financial Studies* 18, 1171–1217.

Garman, Mark B., and Michael J. Klass, 1980, On the estimation of security price volatilities from historical data, *Journal of Business* 53, 67-78.

Gale, Douglas, and Shachar Kariv, 2003, Bayesian learning in social networks, *Games and Economic Behavior* 45, 329-346.

George, Thomas J., Gautam Kaul and M. Nimalendran, 1991, Estimation of the bid-ask spread and its components: A new approach, *Review of Financial Studies* 4, 623-656.

Glosten, Lawrence R., 1987, Components of the bid-ask spread and the statistical properties of transaction prices, *Journal of Finance* 42, 1293-1307.

Glosten, Lawrence R., and Paul R. Milgrom, 1985, Bid, ask and transaction prices in a specialist market with heterogeneously informed traders, *Journal of Financial Economics* 14, 71-100.

Goettler, Ronald L., Christine A. Parlour, and Uday Rajan, 2005, Equilibrium in a dynamic limit order market, *Journal of Finance* 60, 2149–2192.

Goettler, Ronald L., Christine A. Parlour, and Uday Rajan, 2009, Informed traders and limit order markets, *Journal of Financial Economics* 93, 67-87.

Golub, Benjamin, and Matthew O. Jackson, 2010, Naïve learning in social networks and the wisdom of crowds, *American Economic Journal: Microeconomics* 2, 112-149.

Golosov, Mikhail, Guido Lorenzoni, and Aleh Tsyvinski, 2009, Decentralized trading with private information, NBER working paper #w15513.

Hansen, P. Reinhard, and Asger Lunde, 2006, Realized variance and market microstructure noise, *Journal of Business and Economic Statistics* 24,127-218.

Hayek, Freidrich A., 1948, Individualism and Economic Order, University of Chicago Press.

Jarque, Carlos M., and Anil K. Bera, 1980, Efficient tests for normality, homoscedasticity and serial independence of regression residuals, *Economics Letters* 6, 255–259.

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

Kyle, Albert S., Anna Obizhaeva, and Yajun Wang, 2016, Smooth trading with overconfidence and market power, Robert S. Smith School research paper #RHS 2423207.

Lagunoff, Roger, and Stacey L. Schreft, 2001, A model of financial fragility, *Journal of Economic Theory* 99: 220-264.

Large, Jeremy, 2009, A market-clearing role for inefficiency on a limit order book, *Journal of Financial Economics* 91, 102-117.

Leitner, Yaron, 2005, Financial networks: Contagion, commitment, and private sector bailouts, *Journal of Finance* 60, 2925-2953.

Li, Dan, and Norman Schürhoff, 2014, Dealer networks, Finance and Economics Discussion Series 95, Board of the Governors of the Federal Reserve System.

Madhavan, Ananth, Matthew Richardson and Mark Roomans, 1997, Why do security prices change? A transaction-level analysis of NYSE stocks, *Review of Financial Studies* 10, 1035-1064.

Newman, Mark E. J., 2002, Assortative mixing in networks, *Physical Review Letters* 89, 208701.

Parkinson, Michael, 1980, The extreme value method for estimating the variance of rate of return, *Journal of Business* 53, 61-65.

Parlour, Christine A., 1998, Price dynamics in limit order markets, *Review of Financial Studies* 11, 789-816.

Parlour, Christine A., and Duane J. Seppi, 2003, Liquidity-based competition for order flow, *Review of Financial Studies* 16, 301-343.

Rosu, Ioanid, 2009, A dynamic model of the limit order book, *Review of Financial Studies* 22, 4601-4641.

Roukny, Tarik, Stefano Battiston, and Joseph E. Stiglitz, 2016, Interconnectedness as a source of uncertainty in systemic risk, Columbia Business School Research Paper #16-14.

Roukny, Tarik, Hugues Bersini, Hugues Pirotte, Guido Caldarelli, and Stefano Battiston, 2013, Default cascades in complex networks: topology and systematic risk, *Scientific Reports* 3, #2759.

Scheinkman, Jose A., and Wei Xiong, 2003, Overconfidence and speculative bubbles, *Journal of Political Economy* 111, 1183-1219.

Stoll, H., 1978, The supply of dealer services in securities markets, *Journal of Finance*, 33, 1133-1151.

Stoll, Hans R., 1989, Inferring the components of the bid-ask spread: Theory and empirical tests, *Journal of Finance* 44, 115-134.

Tauchen, George E., and Mark Pitts, 1983, The price variability-volume relationship on speculative markets, *Econometrica* 51, 485-505.

Upper, C., 2006, Contagion due to interbank credit exposures: what do we know, why do we know it, and what should we know? Working paper, Bank for International Settlements.

Zhang, Lan, Per A. Mykland and Yacine Ait-Sahalia, 2005, A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data, *Journal of the American Statistical Association*, 100, 1394-1411.

Appendix—Re-wired and Simulation Results

We start the simulation by endowing a fixed number of traders with orders which arrive with a Poisson distributed inter-arrival time and are assigned at random. The quantity for each order is drawn from a lognormal distribution and with an equal probability of being a buy or a sell order. For a sell (buy) order the price is set a small fixed number above (below) the last transaction price plus a log-normally distributed, mean zero random variable. The (log-normally distributed, zero mean) random variable added to the order price is consistent with an equilibrium price function under the assumption of heterogeneous beliefs about the true price process (see Scheinkman and Xiong (2003)).

As our simulation starts, orders begin to populate the limit order book. Each incoming order is compared to previously placed orders by price and time priority. If a match between a buy and sell order is made, a transaction takes place. If an order is only partially filled, we attempt to match the remaining quantity against other orders on the book, and if no match is found, the order remains in the limit order book. In order to avoid stale limit orders, each order expires and is withdrawn from the market after 100 subsequent transactions are executed.

We simulate orders until we attain 7.2 million transactions, segment the data into 12,000 networks of 600 consecutive transactions, and compute simulated network and financial variables for each sampling interval using the same methods applied to the empirically observed data. We then analyze both correlations and lead/lag relations among these variables to compare with actual market results.

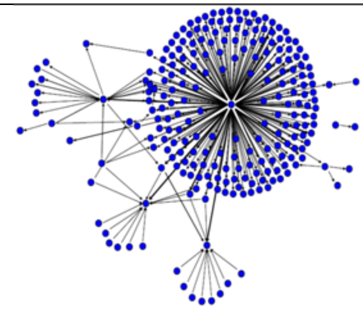
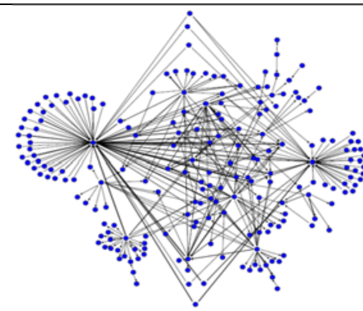
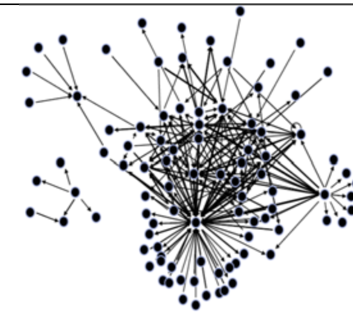
			
Panel A: Network Metrics			
Centralization	-0.736	0.044	0.000
|Centralization|	0.736	0.044	0.000
Assortativity	0.097	0.297	0.011
Clustering	0.003	0.009	0.170
LSCC	0.013	0.005	0.333
Panel B: Traditional Financial Variables			
Returns	-0.100	0.051	0.000
Volatility	0.100	0.051	0.051
Volume	2325	2304	3832
Duration	10	19	14
Effective Spread	0.007	0.005	0.009
Signed Volume	-2151	-557	-170
Trade Imbalance	-4.0	2.0	0.0
Herfindahl Trades	7.50	7.94	9.56

Figure 1. Trading Network Statistics This figure displays three examples of trading networks comprised of 600 sequential trades in the nearby e-mini S&P 500 futures contract on the CME Globex platform during August 2009. *Centralization*, |*Centralization*|, *assortativity*, *clustering* and the *largest strongly connected component* (*LSCC*) are computed as defined in Section I. *Returns* are open-to-close returns computed as differences between log ending and beginning prices during each interval. *Volatility* is the range between high and low prices, *volume* is the total number of contracts traded and duration is the time (in seconds) elapsed between the start and end of the network interval. *Effective Spread* is equal to twice the square root of the first order autocovariance of returns over each interval (x10^{-3}). *Signed Volume* is equal to the buy volume minus the sell volume. *Trade Imbalance* is equal to the number of buy trades less the number of sell trades in each interval. *Herfindahl Trades* is the Herfindahl index computed using the number of trades traded by each trader over each interval (x10^{-3}).

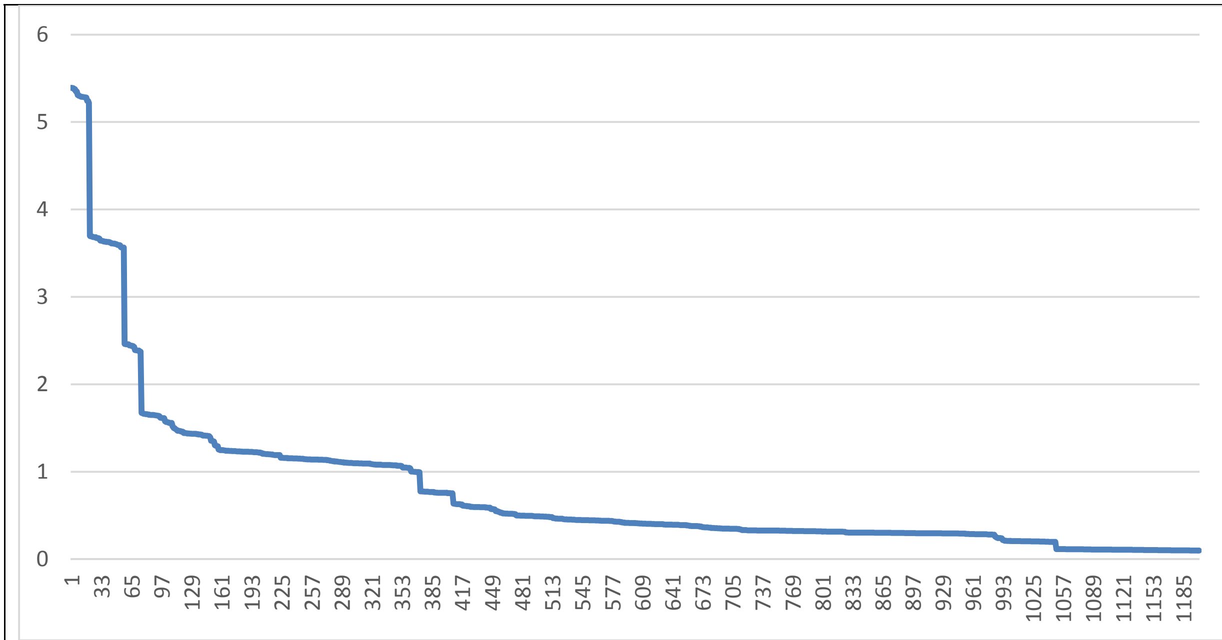


Figure 2: Noise-to-signal ratio. This figure displays the ratio between the range (high-low prices) and the open-to-close returns for the top decile of our network intervals.

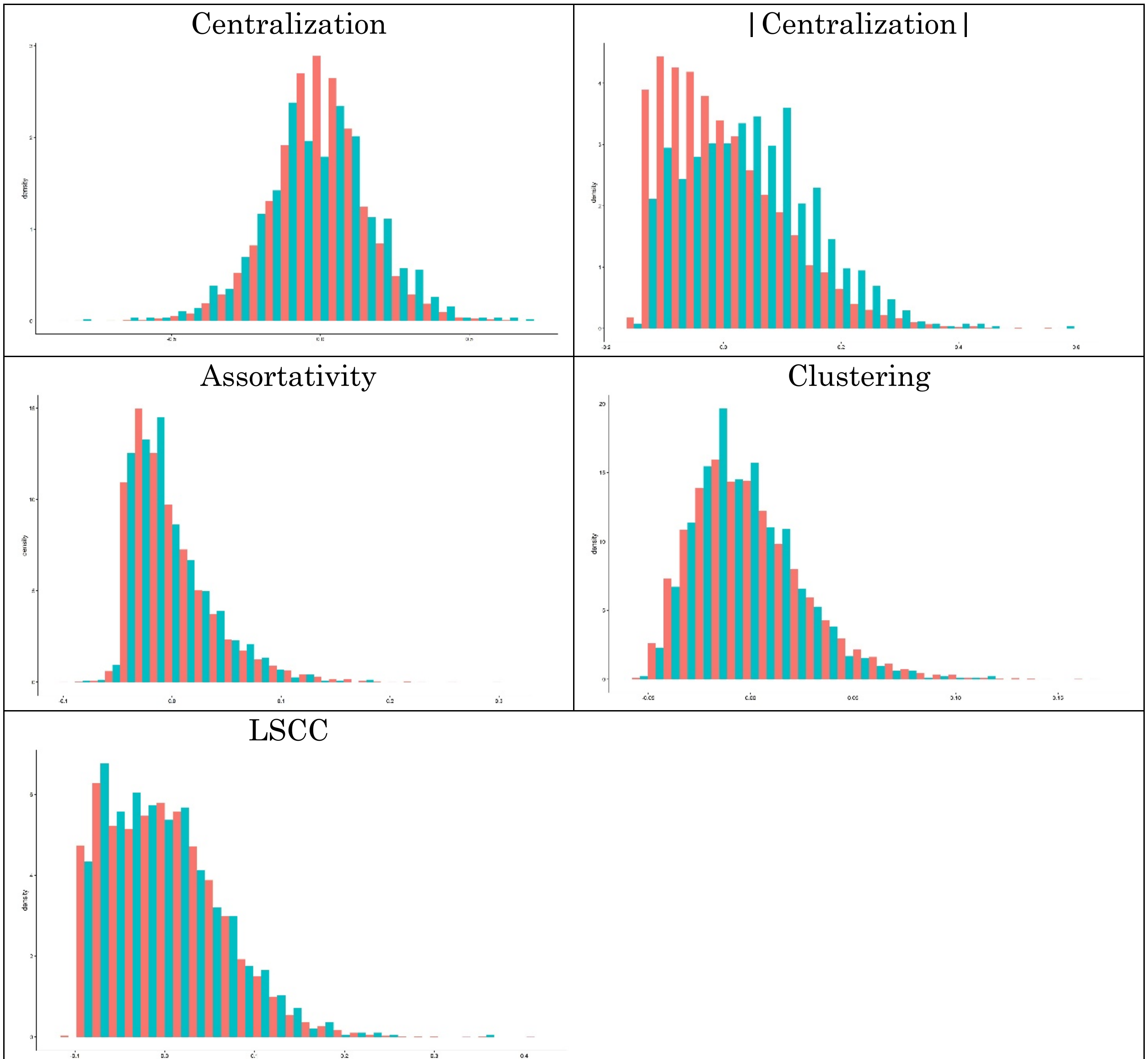


Figure 3: Histograms for the network variables. Observations corresponding to the 95th percentile of the noise-to-signal ratio are in red; all other observations are in green.

Table I: Network Variables: Summary Statistics

The sample contains summary statistics for 12,032 networks each comprised of 600 sequential trades in the nearby e-mini S&P 500 futures contract on the CME Globex platform during August 2009. Centralization, |centralization|, assortativity, clustering and the largest strongly connected component (LSCC) are network variables computed as defined in Section I. ADF probability refers to the p-value of the ADF test for the null of unit root. Terms in [brackets] below autocorrelation coefficients refer to p-values of the Portmanteau Q-test for no serial correlation at 1, 5, and 10 lags.

	Centralization	|Centralization|	Assortativity	Clustering	LSCC
Mean	0.0022	0.1373	0.0424	0.0524	0.0912
Median	0.0037	0.1177	0.0321	0.0483	0.0857
Maximum	0.6499	0.7356	0.3385	0.2250	0.5217
Minimum	-0.7356	0.0000	-0.0557	0.0004	0.0024
Std. Dev.	0.1711	0.1021	0.0391	0.0282	0.0641
Skewness	-0.0200	0.9944	1.5562	0.9605	0.7500
Kurtosis	2.9827	4.0422	6.2788	4.4588	3.6643
ADF probability	0.0000	0.0001	0.0000	0.0000	0.0000
Autocorrelations:	0.007	0.003	0.120	0.197	0.273
Lag 1	[0.462]	[0.779]	[0.000]	[0.000]	[0.000]
	0.079	0.013	0.074	0.085	0.132
Lag 5	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	0.060	0.029	0.058	0.055	0.078
Lag 10	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]

Table II: Financial Variables: Summary Statistics

The sample contains summary statistics for 12,032 networks each comprised of 600 sequential trades in the nearby e-mini S&P 500 futures contract on the CME Globex platform during August 2009. *Returns* are open-to-close network returns computed as differences between log ending and beginning prices during each interval. *Volatility* is the range between high and low prices, *volume* is the total number of contracts traded and *duration* is the time (in seconds) elapsed between the start and end of the network interval. *Effective Spread* is equal to twice the square root of the first order autocovariance of returns, Stoll (1978), over each interval ($x10^{-3}$). *Signed Volume* is equal to the buy volume minus the sell volume. When the price is constant (zero returns), buy/sell volume is equal to that of the previous interval. *Trade Imbalance* is equal to the number of buy trades less the number of sell trades in each interval. *Herfindahl Trades* is the Herfindahl index computed using the number of trades traded by each trader over each interval ($x10^{-3}$). ADF probability refers to the p-value of the ADF test for the null of unit root. Terms in [brackets] below autocorrelation coefficients refer to p-values of the Portmanteau Q-test for no serial correlation at 1, 5, and 10 lags.

	Returns	Volatility	Volume	Duration	Effective Spread	Signed Volume	Trade Imbalance	Herfindahl Trades
Mean	0.0002	0.0621	2628.7	40.775	7.9369	13.804	0.0145	0.0842
Median	0.0000	0.0507	2521.0	28.000	7.9046	18.000	0.0000	0.0836
Maximum	0.1992	0.2901	7706.0	344.00	21.256	5171.0	13.000	0.1306
Minimum	-0.1493	0.0241	1095.0	0.0000	0.0011	-4592.0	-8.0000	0.0242
Std. Dev	0.0383	0.0191	636.06	37.926	2.1513	1219.3	1.5679	0.0126
Skewness	-0.0116	0.9046	1.9566	1.9630	0.1778	0.0462	0.0615	0.1456
Kurtosis	2.7977	7.0270	11.016	8.1385	3.7118	2.9539	3.5352	3.1088
ADF probability	0.0001	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001	0.0000
Autocorrelations:	-0.0121	0.1747	0.5211	0.5483	0.1867	0.1985	-0.0113	0.3796
Lag 1	[0.185]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.213]	[0.000]
Lag 5	-0.0136	0.1392	0.3378	0.3805	0.1263	0.0364	-0.0080	0.1840
	[0.363]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.073]	[0.000]
Lag 10	0.0027	0.1076	0.2351	0.3299	0.0992	0.0064	0.0054	0.1249
	[0.289]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.115]	[0.000]

Table III: Correlations between Financial and Network Variables

This table contains pairwise correlations calculated from 12,032 networks of 600 sequential trades in the nearby e-mini S&P 500 futures contract on the CME Globex platform during August 2009. Centralization, |centralization|, assortativity, clustering and the largest strongly connected component (LSCC) are network variables computed as defined in Section I. Returns are open-to-close network returns computed as differences between log ending and beginning prices during each interval. Volatility is the range between high and low prices, volume is the total number of contracts traded and duration is the time (in seconds) elapsed between the start and end of the network interval. Effective Spread is equal to twice the square root of the first order autocovariance of returns, Stoll (1978), over each interval ($x10^{-3}$). Signed Volume is equal to the buy volume minus the sell volume. Trade Imbalance is equal to the number of buy trades less the number of sell trades in each interval. Herfindahl Trades is the Herfindahl index computed using the number of trades traded by each trader over each interval. * indicates statistical significance at the 5 percent level.

	Central -ization	|Central- ization|	Assorta- tivity	Cluster- ing	LSCC
Panel A:					
|Centralization|	0.008				
Assortativity	0.001	0.085*			
Clustering	-0.001	-0.235*	-0.038*		
LSCC	-0.002	-0.142*	-0.550*	0.541*	
Panel B:					
Returns	0.675*	-0.011	0.013	0.011	0.006
Volatility	-0.025*	0.170*	-0.072*	-0.034*	0.024*
Volume	0.018	0.093*	-0.052*	0.101*	0.153*
Duration	-0.006	-0.193*	-0.011	0.003	-0.015
Effective Spread	-0.003	-0.157*	0.033	-0.250*	-0.318
Signed Volume	0.637*	0.007	-0.004	-0.009	-0.004
Trade Imbalance	0.669*	-0.013	0.015	0.013	0.002
Herfindahl Trades	-0.004	-0.027*	0.067*	-0.621*	-0.559

Table IV: Network Variables: p-values for the Null Hypothesis of Granger Non-causality

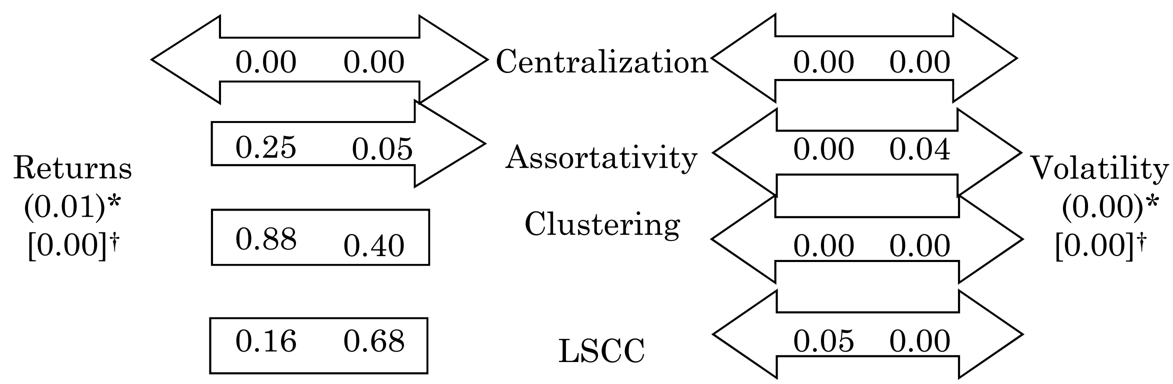
The sample contains p-values for the null hypothesis of Granger non-causality for vector autoregressions (VARs) using the time series of 12,032 network variables each comprised of 600 sequential trades in the nearby e-mini S&P 500 futures contract on the CME Globex platform during August 2009. Centralization, |centralization|, assortativity, clustering and the largest strongly connected component (LSCC) are network variables computed as defined in Section I. The VARs are estimated using generalized method of moments (GMM) with HAC robust standard errors and with optimal lag-length of 12 selected using Akaike Information Criterion. The upper right quadrant displays p-values for pairwise tests of the first column variable Granger-causing the first row variable, with Total representing the p-value for the first column variable jointly Granger-causing all variables in the first row. The lower left quadrant displays p-values for pairwise tests of the first row variable Granger-causing the first column variable, with All representing the p-value for the first row variable jointly Granger-causing all variables in the first column.

	Centralization	|Centralization|	Assortativity	Clustering	LSCC	Total
Centralization		0.59	0.75	0.99	0.65	0.65
|Centralization|	0.18		0.11	0.01*	0.04*	0.00*
Assortativity	0.69	0.09		0.03*	0.00*	0.00*
Clustering	0.54	0.06	0.00*		0.00*	0.00*
LSCC	0.98	0.15	0.00*	0.06		0.00*
All	0.37	0.05*	0.00*	0.01*	0.00*	

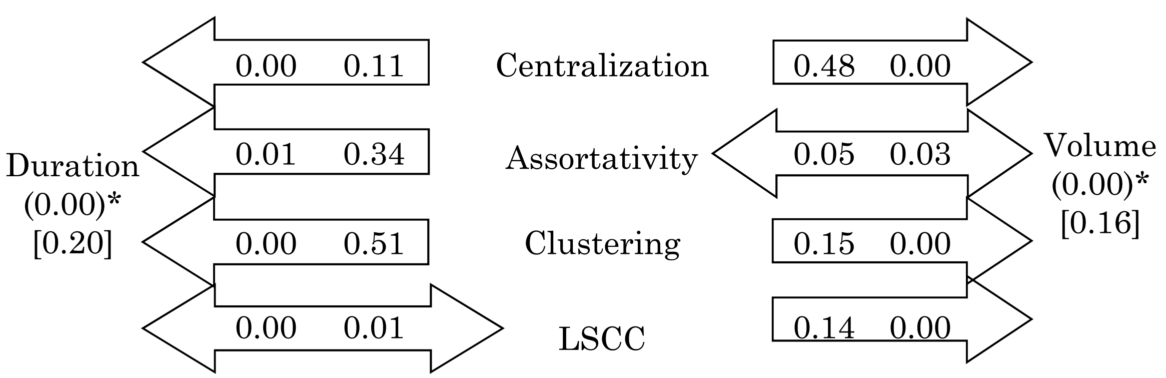
Table V: Financial and Network Variables: p-values for the Null Hypothesis of Granger Non-causality

The sample contains p-values for the null hypothesis of Granger non-causality for VARs using the time series of 12,032 network variables each comprised of 600 sequential trades in the nearby e-mini S&P 500 futures contract on the CME Globex platform during August 2009. *Returns*, *Volatility*, *Duration*, *Volume*, *Effective Spreads*, *Signed Volume*, *Trade Imbalance*, and *Herfindahl Trades* are computed over each interval. *Centralization*, *Assortativity*, *Clustering* and the *Largest Strongly Connected Component* (*LSCC*) are network variables computed as defined in Section I. The VARs are estimated using GMM with HAC robust standard errors and with optimal lag-length of 12 selected using AIC. Each number represents the p-value for pairwise tests between variables in each column, with arrows indicating significance at the 5% level. The p-values in parentheses denote that these four Network Variables jointly Granger-cause Financial Variables with significance at the 5 percent level labeled *. The p-values in brackets denote that Financial Variables jointly Granger-cause the four Network Variables with significance at the 5 percent level labeled †. Centralization is used in the VAR for returns while |centralization| is used in the VARs for all other financial variables.

Panel A:



Panel B:



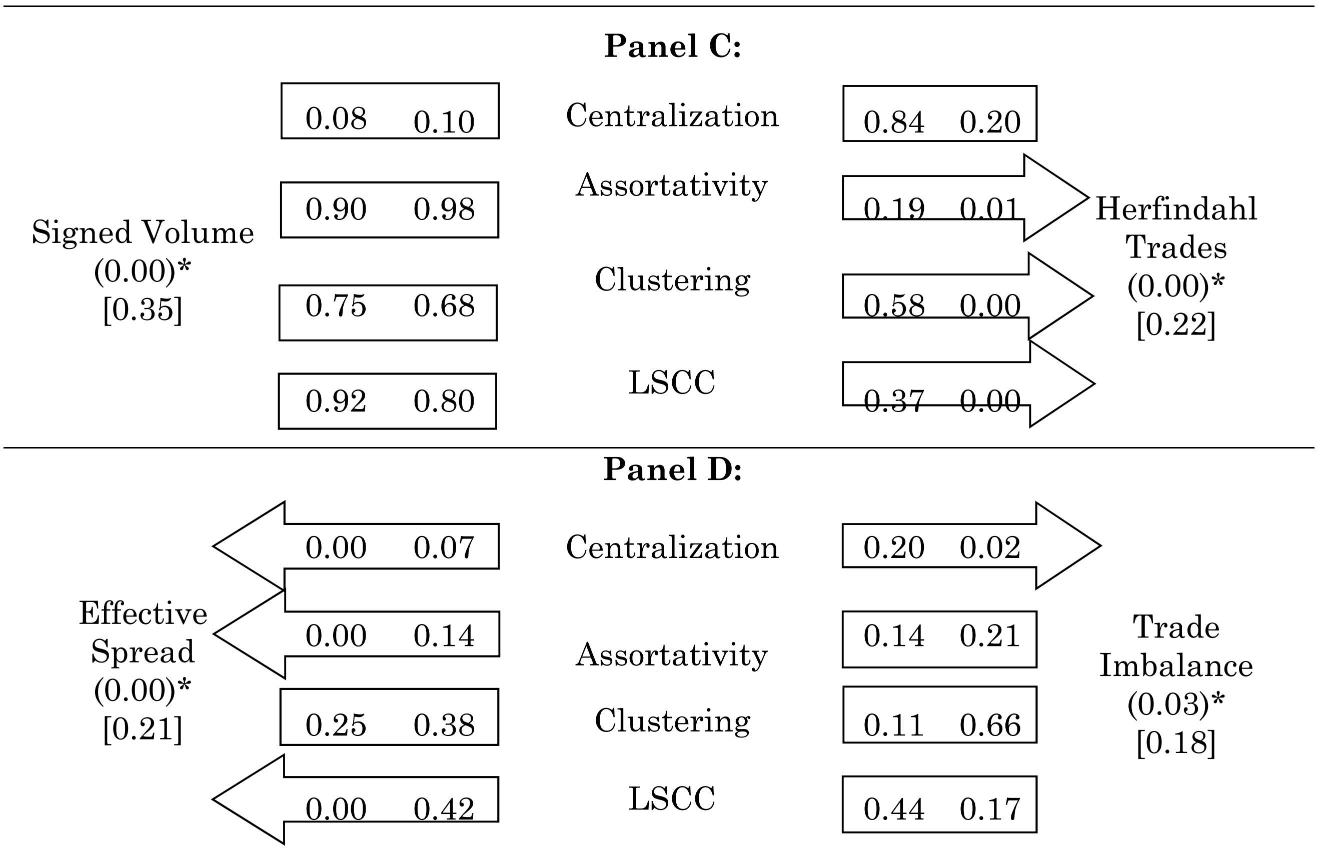

Panel C:
Signed Volume
(0.00)*
[0.35]
0.08 0.10
0.90 0.98
0.75 0.68
0.92 0.80
Centralization
Assortativity
Clustering
LSCC
0.84 0.20
0.19 0.01
0.58 0.00
0.37 0.00
Herfindahl
Trades
(0.00)*
[0.22]
Panel D:
Effective
Spread
(0.00)*
[0.21]
0.00 0.07
0.00 0.14
0.25 0.38
0.00 0.42
Centralization
Assortativity
Clustering
LSCC
0.20 0.02
0.14 0.21
0.11 0.66
0.44 0.17
Trade
Imbalance
(0.03)*
[0.18]

Table VI: Financial and Liquidity Measures: p-values for the Null Hypothesis of Granger Non-causality

The sample contains p-values for the null hypothesis of Granger non-causality for bi-variate VARs estimated with optimal lag-length 6 selected using the AIC over 12,032 sequential networks each comprised of 600 sequential trades in the nearby e-mini S&P 500 futures contract on the CME Globex platform during August 2009. Each number represents the p-value for pairwise tests between variables in each column, with arrows indicating significance at the 5% level. *Returns* are open-to-close network returns computed as differences between log ending and beginning prices during each interval. *Volatility* is the range between high and low prices, *Volume* is the total number of contracts traded and *Duration* is the time (in seconds) elapsed between the start and end of the network interval. *Effective Spread* is equal to twice the square root of the first order autocovariance of returns, Stoll (1978), over each interval ($\times 10^{-3}$). *Signed Volume* is equal to the buy volume minus the sell volume. *Trade Imbalance* is equal to the number of buy trades less the number of sell trades in each interval. *Herfindahl Trades* is the Herfindahl index computed using the number of trades traded by each trader over each interval.

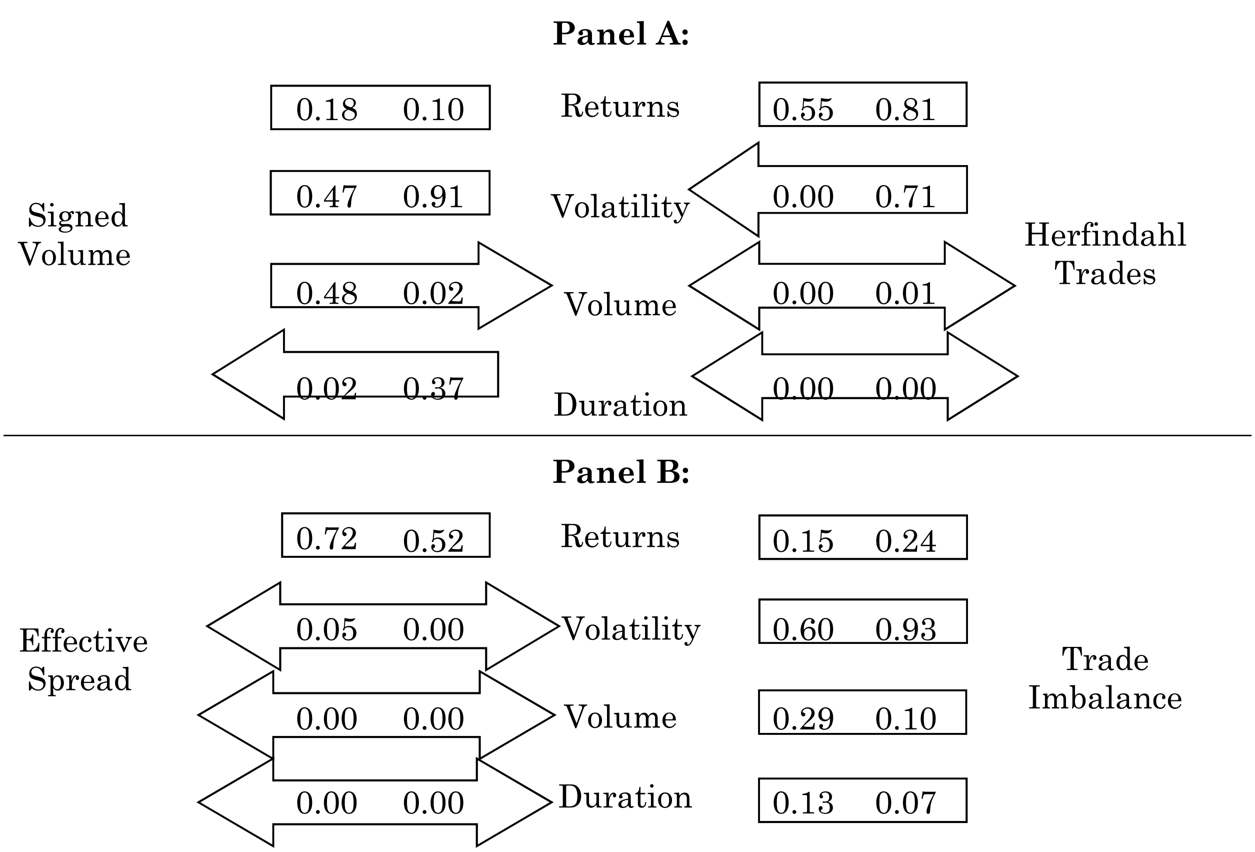


Table VII: Simulation Results - Correlations between Financial and Network Variables

This table contains pairwise correlations between network variables each calculated from 12,000 simulated networks of 600 sequential trades as described in Section IV. Centralization, |centralization|, assortativity, clustering and the largest strongly connected component (LSCC) are network variables computed as defined in Section I. Returns are open-to-close network returns computed as differences between log ending and beginning prices during each interval. Volatility is the range between high and low prices, volume is the total number of contracts traded and duration is the time (in seconds) elapsed between the start and end of the network interval. Effective Spread is equal to twice the square root of the first order autocovariance of returns, Stoll (1978), over each interval ($x10^{-3}$). Signed Volume is equal to the buy volume minus the sell volume. Imbalance is equal to the number of buy trades less the number of sell trades in each interval. The Herfindahl Trades measure is the Herfindahl index computed using the number of trades traded by each trader over each interval. * indicates statistical significance at the 5 percent level.

	Central -ization	|Central- ization|	Assorta- tivity	Cluster- ing	LSCC
Returns	0.605*	-0.024*	0.003	-0.000	0.003
Volatility	0.008	0.219*	0.013	0.004	-0.059*
Volume	-0.004	0.067*	0.119*	0.042*	-0.028*
Duration	0.016	0.003	0.012	-0.002	0.003
Effective Spread	0.052*	0.039*	0.031*	0.007	-0.010
Signed Volume	0.299*	0.009	0.043*	0.021*	-0.016
Imbalance	0.722*	0.052*	-0.011	0.000	-0.022*
Herf. Trades	-0.001	-0.138*	0.001	-0.050*	-0.818*

Table VIII: t-tests for Differences in Means by Noise-to-Signal Ratio
Mean (0) and Std. Dev. (0) refer to the means and standard deviations of the network statistics calculated from the network intervals *not* belonging to the largest quintile of the noise-to-signal ratio. Mean (1) and Std. Dev. (1) refer to means and standard deviations of the network statistics calculated from the network intervals within the largest quintile of the noise-to-signal ratio. *t-tests* relate to the differences between means.

	Centralization	|Centralization|	Assortativity	Clustering	LSCC
Mean (0)	0.0000	0.0012	-0.0020	0.0010	0.0101
Std. Dev. (0)	0.1591	0.1025	0.0389	0.0279	0.1592
Mean (1)	0.0116	-0.0372	0.0056	0.0057	0.0117
Std. Dev. (1)	0.1331	0.0733	0.0441	0.0319	0.1331
t-test	0.041	-0.218	0.070	0.099	0.024

Table IX: Testing the Li and Schürhoff (2014) Mark-up Hypothesis
This table presents results from the regression:

$$r_{i,t} = \alpha_0 + \sum_{j=0}^{3} \beta_j \, k_{i,t} + \gamma_0 + h_0 + m_0 + s_0 + u_{i,t}$$

Where k_t is a centrality measure for the top seven traders, by net volume, r_i, is the return over the network measurement interval and γ_0, h_0, m_0, s_0 are the day of the week, hour, minute, second fixed effects. *, ** and *** indicate significance at the 10, 5 and 1 percent level. The mark-up hypothesis implies that $\beta_1 < 0$, $\beta_2 > 0$and $\beta_3 > 0$.

	Dependent variable	
Independent Variable	r_t	k_t
k_t	-0.0002***	
	(0.00001)	
k_{t-1}	0.00003***	
	(0.00001)	
k_{t-2}	0.00002**	
	(0.00001)	
k_{t-3}	0.00001	
	(0.00001)	
r_{t-1}		77.882***
		(8.195)
r_{t-2}		52.083***
		(8.199)
r_{t-3}		24.678***
		(8.208)
Observations	12,043	12,043
Adjusted R^2	0.032	0.011
F Statistic	3.968*** (df = 133; 11909)	1.975*** (df = 132; 11910)